# Gaussian Difference Dynamics: Breaking the $2^{n/2}$ Barrier for the Shortest Vector Problem and Discrete Gaussian Sampling

Yiming Gao[1] Xuyuan Han[1] Honggang Hu[1,2]

[1]School of Cyber Science and Technology,
University of Science and Technology of China
qw1234567@mail.ustc.edu.cn, hanxuyuan@mail.ustc.edu.cn

[2]Hefei National Laboratory, Hefei, China
hghu2005@ustc.edu.cn

**Abstract**

We give classical randomized algorithms in a word RAM model that break the $2^{n/2}$ barrier for both discrete Gaussian sampling (DGS) and the Shortest Vector Problem (SVP). For every fixed $\alpha > 1$, shifted DGS on arbitrary lattices at width $\sigma \geq \alpha\eta_{1/2}(L)$ takes

$$(3/2)^{n/2+o(n)} \text{ time} \qquad \text{and} \qquad (4/3)^{n/2+o(n)} \text{ space.}$$

These bounds match the Becker–Ducas–Gama–Laarhoven (BDGL) heuristic sieving exponents. At widths at least $3\eta_{1/2}(L)/2$, the same total cost suffices for an entire batch of $\lfloor (4/3)^{n/2} \rfloor$ samples jointly close to independent target Gaussians.

For SVP on arbitrary lattices, our algorithm returns a shortest nonzero vector with high probability in

$$2^{0.4130n+o(n)} \text{ time} \qquad \text{and} \qquad 2^{0.3138n+o(n)} \text{ space.}$$

As a corollary, we obtain centered DGS at every width with the same exponents, answering a question of Stephens-Davidowitz (SODA 2016).

We introduce Gaussian difference dynamics (GDD), a framework for tracking expected counts and pairwise clustering of vectors through repeated differencing. A closed recursion for the first two counting moments controls these quantities relative to a Gaussian reference, without requiring independence within intermediate lists. Accurate moment bounds let us extract a Gaussian sample by weighted selection, while the weaker bounds needed to establish a shortest vector representation allow us to retain more differences. We perform these transitions on thin spherical shells, where the Gaussian rejection weights can be renormalized at subexponential cost. For batch sampling, we combine this geometry with the double rejection technique of Doulgerakis and Laarhoven to control the joint distribution of outputs while reusing vectors.

# Contents


**1 Introduction** **4**
- 1.1 Gaussian difference dynamics . . . 5
- 1.2 Main results . . . 6

**2 Technique overview** **8**
- 2.1 A weighted Gaussian difference . . . 8
- 2.2 Supplying the next round via restriction to thin shells . . . 9
- 2.3 Iterating the Gaussian transition . . . 11
- 2.4 Controlling correlations by isolated pairing . . . 12
- 2.5 One Gaussian sample by weighted readout . . . 14
- 2.6 Representations of a shortest vector . . . 15
- 2.7 Independent Gaussian batches . . . 17

**3 Preliminaries and lattice tools** **17**
- 3.1 Ideal analysis and finite realization . . . 18
- 3.2 Counting measures and conditioning . . . 18
- 3.3 Gaussian scales and complexity parameters . . . 18
- 3.4 Theta estimates . . . 19
- 3.5 Basis reduction and initial Gaussian samples . . . 20
- 3.6 Lattice point counts and random superlattices . . . 21

**4 Counting moments under differencing** **22**
- 4.1 The first two moments of one merge . . . 22
- 4.2 Affine shifts and full differencing . . . 23
- 4.3 Isolation for accurate cooling . . . 25
- 4.4 From moment bounds to outputs . . . 26

**5 Gaussian cooling by isolated pairing** **27**
- 5.1 Shells and pair weights for repeated cooling . . . 27
- 5.2 Sparse activation controls the moment error . . . 28
- 5.3 Initializing and repeating the cooling step . . . 29

**6 Finding pairs with BDGL product codes** **31**
- 6.1 Counting filter insertions and pair reports . . . 31
- 6.2 Expected search costs for the generated lists . . . 32
- 6.3 Time and space bounds . . . 33
- 6.4 From ideal operations to finite algorithms . . . 33

**7 Gaussian sampling by weighted readout** **34**
- 7.1 Choosing the readout weight . . . 34
- 7.2 Paying for the smoothing extension . . . 35
- 7.3 Reaching the BDGL exponents . . . 36

**8 Gaussian shells and full differencing** **38**
- 8.1 Counting representations by their midpoints . . . 38
- 8.2 Moment bounds for repeated full differencing . . . 39

**9 Shortest vector search in a random coset** **41**
9.1 A random common coset . . . 41
9.2 Counting representations in the coset . . . 42
9.3 A target difference in the two lists . . . 43
9.4 Cost of the final pair search . . . 43

**10 Solving the shortest vector problem** **45**
10.1 Choosing the list sizes from the target . . . 45
10.2 Constructing the two final lists . . . 46
10.3 The target representation count . . . 47
10.4 The cost of finding the target . . . 47
10.5 The algorithm and success amplification . . . 49
10.6 Other choices of the four layer energies . . . 49

**11 Independent Gaussian batches** **51**
11.1 Two rejections against a fixed bank . . . 51
11.2 Exponential batches with a fixed bank . . . 52

**12 Reductions and algorithmic consequences** **58**
12.1 Accuracy and composition . . . 58
12.2 Reusable decoding from Gaussian batches . . . 58
12.3 Shortest vectors and gap tests from decoding . . . 59
12.4 Centered Gaussian sampling from SVP . . . 61
12.5 Gaussian samples for CVP and successive minima . . . 62
12.6 Short independent vectors near smoothing . . . 63
12.7 Polynomial-factor approximations by block reduction . . . 63

**13 Conclusion and further discussion** **64**

**Appendices** **65**

**A Lattice and Gaussian tools** **65**

**B Search and application estimates** **69**

**C Finite realization** **80**

**D Centered DGS by a mixture over radii** **89**

**E Proofs of the further consequences** **98**

**F Reusable decoding and promise problem reductions** **102**

**AI methodology** **106**

**References** **107**

# 1 Introduction

The Shortest Vector Problem (SVP) and discrete Gaussian sampling (DGS) are basic problems in lattice algorithms, complexity, and cryptography. For a full rank lattice $L \subset \mathbb{R}^n$, SVP asks for a shortest nonzero vector, while DGS samples lattice points with probabilities proportional to $e^{-\pi\|x\|^2/r^2}$ at a prescribed width $r$. Discrete Gaussians connect sampling with the search for short vectors, and Gaussian identities give quantitative control over the distributions produced by repeatedly combining vectors. Geometric sieving exploits the fact that nearby vectors have a shorter difference, suggesting algorithms that combine Gaussian distributional control with the efficiency of geometric search.

**The $2^{n/2}$ barrier in provable algorithms.** Aggarwal, Dadush, Regev, and Stephens-Davidowitz (ADRS) gave an SVP algorithm using $2^{n+o(n)}$ time and space via Gaussian sampling at arbitrary width [ADRS15], and Aggarwal and Stephens-Davidowitz later simplified the algorithm by pairing and averaging [AS18]. ADRS also give a faster sampler above smoothing that generates $2^{n/2}$ independent samples in $2^{n/2+o(n)}$ time and space. Building on sampling above smoothing, our earlier work uses random superlattices to reach the shortest vector scale, first by direct sampling in time $2^{0.7314n+o(n)}$ [GFH26a] and then by geometric pair search in time $2^{0.5596n+o(n)}$ [GFH26b]. More recently, Hhan's $q$-ary coset difference tree brings arbitrary lattice SVP to $2^{n/2+o(n)}$ time and space [Hhan26], while Kim obtains one centered Gaussian sample at any width in expected $2^{n/2+o(n)}$ time [Kim26].

The fast ADRS sampler combines pairs $x, y$ by averaging, $(x+y)/2$. Before rescaling, it pairs vectors from the same coset of $L_{i+1} \subset L_i$ and retains their sum, with $2L_i \subseteq L_{i+1}$. Its lattice tower satisfies

$$L_{i+2} \subseteq 2L_i \quad \Longrightarrow \quad [L_i : L_{i+1}]\,[L_{i+1} : L_{i+2}] = [L_i : L_{i+2}] \geq 2^n.$$

Balancing the adjacent indices gives $2^{n/2+o(n)}$ cosets per level, which sets the sampler's list scale [ADRS15, Proposition 5.7, Definition 5.8, and Theorem 5.11]. Algorithms that use this batch interface inherit its generation cost, while Hhan's half exponent arises from a balance between Gaussian signal, estimation error, and coset table size [Hhan26, Section 1.1 and Remark 1.2]. Improving these bounds calls for revisiting how intermediate vectors are generated and how much supply the final task needs. Stephens-Davidowitz asks for a faster centered sampler, possibly producing only one sample, and writes that "entirely new techniques would be needed" to go below $2^{n/2}$ [SD16, Section 1.4, p. 7].

**Making geometric sieving provable on arbitrary lattices.** Sieving began with the algorithm of Ajtai, Kumar, and Sivakumar [AKS01] and developed through both provable and heuristic analyses [NV08, PS09, BGJ15]. Under heuristic assumptions on the successive lists, the batched BDGL sieve takes $(3/2)^{n/2+o(n)}$ time and $(4/3)^{n/2+o(n)}$ space [BDGL16]. Recent work of Doulgerakis and Laarhoven [DL26] and Zhou, Wang, and Wang [ZWW26] proves these SVP exponents for Haar-random lattices, using distributional estimates for random lattice geometry and the generated lists. Doulgerakis and Laarhoven also construct Gaussian batches at every width.

Extending this efficiency to arbitrary input lattices requires control of both the geometry and the dependence of the generated vectors. Repeated differences share inputs and ancestors, while local lattice geometry governs the numbers of partners of an input and representations of an output, especially at widths below smoothing. A rigorous iteration must therefore follow the supply at the intended locations together with the correlations that affect its availability.

We combine Gaussian distributional control with geometric sieving to obtain algorithms on every lattice. Above smoothing, Gaussian sampling attains the BDGL time and space exponents, while SVP and centered sampling at arbitrary width take $2^{0.4130n+o(n)}$ time and $2^{0.3138n+o(n)}$ space. We first describe the counting framework behind the single sample and shortest vector constructions, then state their guarantees and the stronger batch result.

### 1.1 Gaussian difference dynamics

We start with wide Gaussian samples and repeatedly form weighted short differences to reach a smaller Gaussian width. Once the vectors reach a suitable scale, we use the resulting lists to produce the desired output. For DGS above smoothing, one weighted selection gives a Gaussian sample. For SVP, we first build additional representations and then use BDGL filters to search for short differences between two terminal lists. The final task thus determines what guarantees the intermediate lists must provide.

The main difficulty is to control the lists produced by this iteration. Differences sharing an input are correlated, and later rounds inherit dependence through common ancestors. The heuristic BDGL sieve analyzes successive lists under assumptions on their angular distribution [BDGL16]. The ADRS Gaussian combiner instead maintains a joint distribution close to independent Gaussian samples through successive rounds [ADRS15]. Retaining many differences reuses the same inputs, so an analysis of this iteration must account for the resulting dependence.

We track two counting moments relative to a Gaussian reference: the expected number of records in a region and the expected number of ordered pairs of distinct records in two regions. The first describes where records appear, while the second controls their clustering. At each merge, conditional independence between the two separately generated child lists gives a closed recursion for these moments, even though either list may be internally correlated. These bounds connect expected counts to the probability of producing useful vectors and to the work needed to find them. We call this evolution *Gaussian difference dynamics*.[1]

Thin spherical shells let us renormalize Gaussian rejection weights at subexponential cost. On a fixed superlattice smooth at all midpoint widths, the Gaussian midpoint identity gives the reference counts for each cooling round. Geometric search finds the pairs needed for the next list.

The final task determines the required moment accuracy. For Gaussian sampling, weighted selection divides by the list's random total weight. Accurate moments control this normalization when the reference total weight is sufficiently large. Sparse activation followed by isolation preserves this precision, retaining a pair only when no other activated pair shares either input.

For SVP, it suffices that a shortest vector is represented with probability $e^{-o(n)}$, since repetition then costs only a subexponential factor. This weaker requirement permits a fixed number of full differencing layers that retain every eligible pair, building more representations before the final search. These layers keep the reference width fixed and lower the shell energy. We compare two terminal lists in a common random affine coset and select differences that belong to the original lattice.

An exponential batch additionally requires the joint output law to be close to a product of target Gaussians. We combine the same shell geometry with the two rejection corrections of Doulgerakis and Laarhoven [DL26]. At each merge, these corrections make the complete trial law, including failure, independent of the reused partner list. Uniform estimates for these lists and mixing across radial bins then give the batch guarantee. Section 2 develops these constructions.

The counting framework has also been applied to $\ell_\infty$-SVP with uniform reference counts on cubes [Gao26].

[1] Dynamical systems have also been used in the analysis of lattice basis reduction [HPS11, LN25, Wang24].

## 1.2 Main results

For a rational lattice $L = B\mathbb{Z}^n$, write

$$\rho_\sigma(x) = e^{-\pi\|x\|^2/\sigma^2}, \qquad D_{a+L,\sigma}(x) = \frac{\rho_\sigma(x)}{\rho_\sigma(a+L)}.$$

The smoothing parameter is defined by $\rho_{1/\eta_\epsilon(L)}(L^* \setminus \{0\}) = \epsilon$. We write $\mathcal{L}(X)$ for the law of $X$ and $d_{\mathrm{TV}}$ for total variation distance. Write

$$\alpha_{\mathrm{G}} = \tfrac{1}{2}\log_2(3/2), \qquad \beta_{\mathrm{G}} = \tfrac{1}{2}\log_2(4/3), \qquad J(z) = \frac{z - 1 - \log z}{2\log 2}.$$

Our algorithms use a randomized word RAM, and their time and bit space bounds hold on every execution, including random bit generation and finite arithmetic. We suppress polynomial factors in the binary input length $\mathcal{B}$ and the accuracy parameter $f$, when present. The computational model and finite implementation are given in Appendix C.1.1 and Proposition 6.3.

Our first application of GDD is a Gaussian sampler. Controlling the first two counting moments of the correlated list suffices to obtain one Gaussian sample by weighted selection, at the BDGL time and space exponents.

**Theorem 1.1** (Single sample DGS above smoothing)**.** *Fix $\alpha > 1$ and $\xi > 0$. There is a classical randomized algorithm that, given a full rank rational lattice $L$, a rational shift $a$, and a positive rational squared width $\sigma^2$ with $\sigma \geq \alpha\eta_{1/2}(L)$, outputs $X \in a + L$ with*

$$d_{\mathrm{TV}}(\mathcal{L}(X), D_{a+L,\sigma}) \leq \exp[-\Omega(n/\log^{2+\xi}(n+2))].$$

*Its time is $2^{\alpha_{\mathrm{G}} n + o(n)}$ and its bit space is $2^{\beta_{\mathrm{G}} n + o(n)}$.*

Section 7 proves this result. The constants may depend on $\alpha$ and $\xi$.

To obtain an exponential batch within the time and space bounds of Theorem 1.1, we combine the Gaussian geometry above with the double rejection technique of Doulgerakis and Laarhoven [DL26]. This gives the stronger joint sampling guarantee below.

**Theorem 1.2** (Batch DGS above smoothing)**.** *Fix $\alpha > 1$ and $0 < b < \min\{\beta_{\mathrm{G}}, J(\alpha^{-2})\}$. Given a full rank rational lattice $L = B_L\mathbb{Z}^n$, a rational shift $a$, and a positive rational squared width $\sigma^2$ satisfying $\sigma \geq \alpha\eta_{1/2}(L)$, an algorithm outputs $B = \lfloor 2^{bn} \rfloor$ points in $a + L$ with*

$$d_{\mathrm{TV}}(\mathcal{L}(X_1, \ldots, X_B), D_{a+L,\sigma}^{\otimes B}) \leq \exp[-\Omega(n/\log^2(n+2))].$$

*Its total time is $2^{\alpha_{\mathrm{G}} n + o(n)}$ and its bit space, including the output, is $2^{\beta_{\mathrm{G}} n + o(n)}$. If $\alpha \geq 3/2$, one may take the full batch $B = \lfloor 2^{\beta_{\mathrm{G}} n} \rfloor$.*

The single sample construction gives a direct application of the counting moment framework, while the batch guarantee requires additional control of the joint output law. Retaining the first batch output also gives a single sample with this stronger error bound and the same BDGL exponents $\alpha_{\mathrm{G}} \approx 0.2925$ and $\beta_{\mathrm{G}} \approx 0.2075$.

In the full batch case, shared preprocessing gives average time $2^{\frac{1}{2}\log_2(9/8)n + o(n)}$ per output. Section 11.2 proves the joint law and total resource bounds.

**Theorem 1.3** (SVP)**.** *Given a full rank rational lattice $L = B\mathbb{Z}^n$, there is a classical randomized algorithm that returns a shortest nonzero vector with probability at least $1 - 2^{-\Omega(n \log n)}$, in*

$$T = 2^{0.4130n+o(n)} \quad \textit{time}, \qquad S = 2^{0.3138n+o(n)} \quad \textit{space}.$$

*Every output is a nonzero vector of $L$.*

Stephens-Davidowitz [SD16, Theorem 4.6] gave a reduction from centered DGS at arbitrary width to SVP, with polynomial expected overhead and no increase in dimension. Applying this reduction to Theorem 1.3 gives a single Gaussian sample with the same time and space exponents, answering his question about sampling below $2^{n/2}$.

We use his notion of accuracy: a law is $(\gamma, \varepsilon)$-close to the target if it is within total variation distance $\varepsilon$ of a law whose pointwise probabilities differ from the target by at most a factor $\gamma$ (Section 12.4).

**Corollary 1.4** (Centered Gaussian sampling at arbitrary width)**.** *Let $f : \mathbb{N} \to \mathbb{Z}_{\geq 1}$ be efficiently computable, with $1 \leq f(n) \leq \mathrm{poly}(n)$. For a full rank rational lattice $L = B\mathbb{Z}^n$ and any positive rational width $\sigma$, one can sample $X \in L$ whose law is $(1 + 1/f(n), 2^{-f(n)})$-close to $D_{L,\sigma}$. The algorithm uses*

$$T = 2^{0.4130n+o(n)} \quad \textit{time}, \qquad S = 2^{0.3138n+o(n)} \quad \textit{space}$$

*per sample.*

The total variation error is $O(1/f(n)) + 2^{-f(n)}$, permitting any prescribed inverse polynomial accuracy without changing the leading time or space exponent.

The comparison of direction weights in the proof of [SD16, Theorem 4.6] does not go through as stated for the additive grid of squared radii. We obtain the required control using a multiplicative grid and complete the proof within the original reduction framework. Appendix D gives the counterexample and the repaired argument.

| Problem | Guarantee | Time exponent | Space exponent |
|---|---|---|---|
| Single Gaussian sample | $\alpha > 1$ | $\alpha_{\mathrm{G}}$ | $\beta_{\mathrm{G}}$ |
| Full Gaussian batch | $\alpha \geq 3/2$ | $\alpha_{\mathrm{G}}$ | $\beta_{\mathrm{G}}$ |
| Reusable BDD | fixed $r < 0.3157071$ | $P : \alpha_{\mathrm{G}},\ D : \beta_{\mathrm{G}}$ | $\beta_{\mathrm{G}}$ |
| Exact 3-uSVP | $\lambda_2 \geq 3\lambda_1$ | $\alpha_{\mathrm{G}}$ | $\beta_{\mathrm{G}}$ |
| 4.15-GapSVP | promise decision | $\alpha_{\mathrm{G}}$ | $\beta_{\mathrm{G}}$ |
| SVP | arbitrary lattice | 0.4130 | 0.3138 |
| Centered DGS | every width | 0.4130 | 0.3138 |
| 4-CVP, SMP, SIVP | constant approximation | 0.4611 | 0.3138 |
| $\gamma$-CVP, SMP, SIVP | $\gamma(n) = \omega(1)$ | 0.4130 | 0.3138 |
| $O(\sqrt{n \log n})$-SIVP | near smoothing sampling | $\alpha_{\mathrm{G}}$ | $\beta_{\mathrm{G}}$ |

Table 1: Algorithmic consequences on arbitrary rational lattices. Time and bit space entries mean $2^{cn+o(n)}$, with polynomial input factors suppressed. $\alpha_{\mathrm{G}} \approx 0.2925$ and $\beta_{\mathrm{G}} \approx 0.2075$. BDD has preprocessing $P$ and per-target work $D$.

| Reference | Input model | Guarantee | Time | Space |
|---|---|---|---|---|
| Ajtai–Kumar–Sivakumar [AKS01] | Arbitrary | Randomized | $O(1)$ | $O(1)$ |
| Pujol–Stehlé [PS09] | Arbitrary | Randomized | 2.465 | 1.233 |
| Micciancio–Voulgaris [MV13] | Arbitrary | Deterministic | 2 | 1 |
| Aggarwal–Dadush–Regev–Stephens-Davidowitz [ADRS15] | Arbitrary | Randomized | 1 | 1 |
| Gao–Feng–Hu [GFH26a] | Arbitrary | Randomized | 0.7314 | 0.5 |
| Hhan (Hessian) [Hhan26a] | Arbitrary | Randomized | 0.6039 | 0.5 |
| Gao–Feng–Hu [GFH26b] | Arbitrary | Randomized | 0.5596 | 0.5 |
| Hhan (coset tree) [Hhan26] | Arbitrary | Randomized | 0.5 | 0.5 |
| This work | Arbitrary | Randomized | 0.4130 | 0.3138 |
| Pouly–Shen [PS26] | Haar-random | Randomized | 0.6327 | 0.5 |
| Doulgerakis–Laarhoven [DL26] | Haar-random | Randomized | 0.2925 | 0.2075 |
| Zhou–Wang–Wang [ZWW26] | Haar-random | Randomized | 0.2925 | 0.2075 |
| Nguyen–Vidick [NV08] | Arbitrary | Heuristic | 0.4151 | 0.2075 |
| Becker–Ducas–Gama–Laarhoven [BDGL16] | Arbitrary | Heuristic | 0.2925 | 0.2075 |

Table 2: Classical SVP bounds, with time and space exponents in $2^{\alpha n+o(n)}$.

Gaussian batches on the dual lattice give a reusable BDD data structure whose advice is correct simultaneously for every promised target. Accounting separately for its preprocessing and query costs gives exact 3-uSVP and 4.15-GapSVP within the BDGL time and space, with the latter using the smaller batch region of Theorem 1.2.

Centered sampling at every width gives a second route. Gaussian small ball estimates improve the explicit sampling-call exponent in the Euclidean reduction of Aggarwal, Chen, Kumar, Li, and Stephens-Davidowitz [ACKLS21], cover every fixed approximation factor greater than one, and give direct SMP and SIVP algorithms in dimension $n$. Here $\gamma$-SMP returns independent vectors $v_i$ with $\|v_i\| \le \gamma\lambda_i(L)$, while $\gamma$-SIVP bounds their maximum length by $\gamma\lambda_n(L)$. Section 12 develops each reduction together with its algorithmic consequences.

**Organization.** Section 2 develops the algorithms. After the preliminaries, Sections 4 and 5 establish the counting framework and Gaussian cooling, and Section 6 supplies geometric pair search. Section 7 completes the single-sample application. Sections 8 to 10 then develop full differencing, terminal coset search, and SVP. Section 11 establishes independent Gaussian batches, followed by the reductions in Section 12. The appendices give the geometric estimates, finite implementations, and proofs of the reductions.

We used Codex to assist with algorithm design, mathematical analysis, and manuscript preparation. The AI methodology statement at the end of the paper describes its role.

# 2 Technique overview

## 2.1 A weighted Gaussian difference

Write $\rho_r(x) = e^{-\pi\|x\|^2/r^2}$, and fix a desired contraction $0 < \gamma < 1$. For two input vectors $x, y$, set $z = x - y$ and $m = (x + y)/2$. Completion of squares separates their Gaussian weights:

$$\rho_r(x)\rho_r(y) = \rho_{r/\sqrt{2}}(m)\rho_{\sqrt{2}r}(z), \tag{1}$$

$$\rho_r(x)\rho_r(y)e^{-\pi(\gamma^{-2}-1/2)\|z\|^2/r^2} = \rho_{r/\sqrt{2}}(m)\rho_{\gamma r}(z). \tag{2}$$

The exponential factor in the second identity gives an acceptance rule that changes the weight of the difference from width $\sqrt{2}r$ to width $\gamma r$.

To see the operation directly, fix a lattice $\Gamma$ that is smooth at the midpoint width $r/\sqrt{2}$. Consider the following trial:

1. Draw $x, y$ independently from $D_{\Gamma,r}$.
2. Compute $z = x - y$.
3. Accept $z$ with probability $p_r(z) = e^{-\pi(\gamma^{-2}-1/2)\|z\|^2/r^2}$.

For a fixed $z \in \Gamma$, we multiply each input pair's probability mass by $p_r(z)$ and sum over all pairs with difference $z$. Parametrizing these pairs by their midpoints gives

$$\Pr[\text{trial accepts } z] = \frac{\rho_{\gamma r}(z)}{\rho_r(\Gamma)^2} \sum_{m \in z/2+\Gamma} \rho_{r/\sqrt{2}}(m).$$

Smoothing makes the midpoint sum nearly independent of its shift. Consequently, conditional on acceptance, the output is close to $D_{\Gamma,\gamma r}$. The midpoint records the different ways to produce $z$.

To turn this single pair operation into an algorithm, we would like to maintain a list of vectors and repeatedly contract its width, using the output list at width $\gamma r$ as the input to the next round. This suggests a layered sieve in the spirit of BDGL [BDGL16]. We next choose the list sizes and acceptance rule so that each round supplies enough vectors for this iteration.

### 2.2 Supplying the next round via restriction to thin shells

Before rejection, the difference has Gaussian width $\sqrt{2}r$, while the accepted difference has width $\gamma r$. At our smoothing scales, Gaussian mass is proportional to the $n$th power of the width, so the acceptance probability is

$$p_{\text{raw}} \approx \frac{\rho_{\gamma r}(\Gamma)}{\rho_{\sqrt{2}r}(\Gamma)} \approx \left(\frac{\gamma}{\sqrt{2}}\right)^n.$$

Among $N$ vectors there are order $N^2$ candidate pairs. Supplying another list of size $N$ therefore calls for

$$N \gtrsim p_{\text{raw}}^{-1} \approx \left(\frac{\sqrt{2}}{\gamma}\right)^n \geq 2^{n/2}.$$

Thus simple Gaussian rejection cannot break the $2^{n/2}$ barrier.

Our earlier work [GFH26b, Sections 2.3 and 3.3] restricts Gaussian samples to thin shells before a single BDGL decoding layer, giving $2^{0.5596n+o(n)}$ SVP time. Nearly equal endpoint lengths turn the search for short differences into an angular problem, while Gaussian shell estimates control the retained list size. To eliminate this exponential rejection loss, we restrict the inputs and outputs to thin norm shells on which the Gaussian weight factor is nearly constant and can be normalized away. We choose these shells so that the resulting differences supply the next round at a smaller Gaussian width.

Measure squared lengths relative to $nr^2/(2\pi)$ by the normalized energy

$$\mathcal{E}_r(x) = \frac{2\pi\|x\|^2}{nr^2}, \qquad \mathcal{S}_r(a) = \{x : \mathcal{E}_r(x) \in [a(1-w_n), a(1+w_n)]\},$$

where $w_n \to 0$ is the relative shell width. A Gaussian of width $r$ has typical energy near 1 at our smoothing scales, so $a$ specifies the shell's squared radius relative to that typical value. Let $R_-$ be

the inner radius of the desired output shell and write

$$R_-^2 = \frac{n(\gamma r)^2 a(1-w_n)}{2\pi}, \qquad A_r = \frac{\pi(\gamma^{-2}-1/2)}{r^2}.$$

Consider the following trial:

1. Draw $x, y$ independently from $D_{\Gamma,r} \mid \mathcal{S}_r(a)$.
2. Compute $z = x - y$. If $z \notin \mathcal{S}_{\gamma r}(a)$, return failure.
3. Output $z$ with probability
$$\omega_r(z) = \frac{p_r(z)}{p_r(R_-)} = e^{-A_r(\|z\|^2 - R_-^2)},$$
and return failure otherwise.

Here $p_r(R_-) = e^{-A_r R_-^2}$ is the value of the radial acceptance weight at the inner boundary. It is the largest value on the output shell, so $0 < \omega_r(z) \le 1$ there. Dividing by it preserves the relative Gaussian weights and makes $\omega_r(z) = 2^{-O(nw_n)} = 2^{-o(n)}$ throughout that shell.

We can now read off the leading acceptance rate from the shell geometry. Write $u = 1 - \gamma^2/2$. The input lengths are nearly equal, and the output shell forces

$$\frac{\langle x, y\rangle}{\|x\|\|y\|} = 1 - \frac{\gamma^2}{2} + O(w_n) = u + O(w_n).$$

For a fixed endpoint direction, the partner therefore lies in a thin spherical band of correlation $u$. Its transverse radius is $\sqrt{1-u^2}$, giving angular measure $(1-u^2)^{n/2}2^{o(n)}$. At our smoothing scales the shell counts have this same leading rate, and the normalized Gaussian correction adds only a subexponential factor. Thus

$$p_{\text{shell}} = (1-u^2)^{n/2}\, 2^{o(n)}.$$

Two lists of size $N$ offer expected accepted supply of scale $N^2 p_{\text{shell}}$. Supplying another list with the same leading size calls for $N$ slightly above $1/p_{\text{shell}}$. Since contraction requires $u > 1/2$, the smallest supply exponent is approached as $u \downarrow 1/2$. This gives the sampling list size $(4/3)^{n/2+o(n)}$, with exponent about 0.2075. This is exactly the space exponent of the heuristic BDGL sieve!

To iterate the trial, we also need its accepted distribution to have the next Gaussian profile. For $z$ in the output shell, (2) gives

$$\Pr[\text{trial returns } z] \propto \rho_{\gamma r}(z) \sum_{\substack{m \in z/2+\Gamma \\ m+z/2, m-z/2 \in \mathcal{S}_r(a)}} \rho_{r/\sqrt{2}}(m).$$

The endpoint and output shells put these midpoints at energy $\mathcal{E}_r(m) \approx a(1+u)/2$. The midpoint Gaussian in the sum has typical energy $1/2$, which suggests the choice

$$a = a_0 = \frac{1}{1+u}.$$

Figure 1 illustrates the input pair and its midpoint region $\mathcal{W}_r(z) = \{m : m + z/2, m - z/2 \in \mathcal{S}_r(a_0)\}$. With this choice of $a_0$, the region contains almost all of the midpoint Gaussian mass. Smoothing then makes the sum nearly independent of $z$, so the accepted law is close to $D_{\Gamma,\gamma r} \mid \mathcal{S}_{\gamma r}(a_0)$. The same shell parameter can therefore be used in the next round. Section 5.1 proves this Gaussian closure using only the input and output shells.

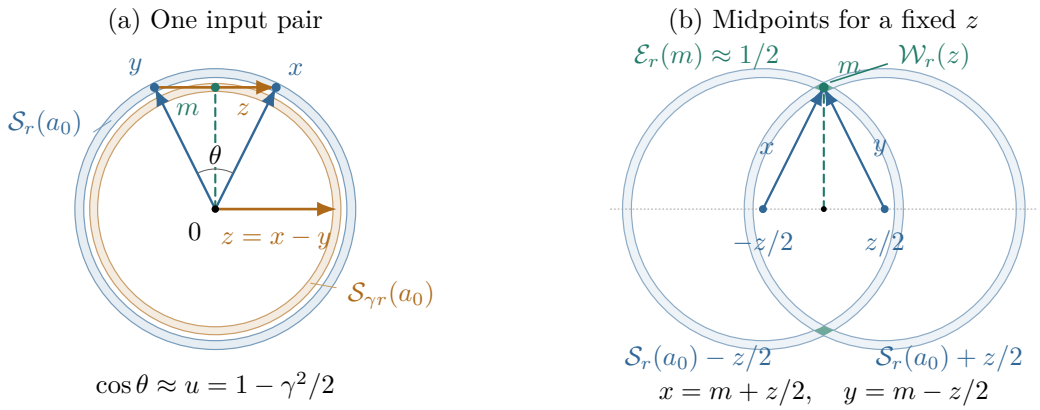



Figure 1: Shell geometry in a planar section, with thickness enlarged for visibility. Left: two input vectors and their shorter difference. Right: the shaded region $\mathcal{W}_r(z)$ contains the admissible midpoints for a fixed difference $z$.

### 2.3 Iterating the Gaussian transition

The preceding calculation starts from two independent inputs. A list of differences soon loses this independence: $x - y$ and $x - y'$ share $x$, and later rounds can share entire chains of ancestors. Even if each output has the desired Gaussian law, pairing these outputs does not automatically repeat the single trial calculation.

We therefore organize the contractions as a binary cooling tree. The two child lists at each merge use disjoint leaf pools and independent private randomness. Conditional on the common lattice and prescribed child cosets, the lists are therefore independent and their expected counts can be multiplied. Each list may still be internally correlated, which the counting argument below will control.

Generating fresh child lists has an affordable cost because cooling needs only a shallow tree. If the leaves start at width $R$ and the root has width $r_0$, then

$$t \approx \frac{\log(R/r_0)}{-\log\gamma}, \qquad \text{number of leaf pools} = 2^t.$$

Basis preprocessing makes $R/r_0$ polynomial in $n$, giving $t = O(\log n)$ for fixed $\gamma$, as used for SVP. Near the BDGL endpoint, the sampling schedule lets $\gamma$ approach 1 slowly enough that $t = O(\log n \log\log n)$ (Lemma 7.2). Thus all the independent leaf pools together add only a factor $2^{o(n)}$ to the cost of one pool.

The random superlattice construction of [GFH26a] supplies $\Gamma \supset L$ smooth at the smallest midpoint width and hence at every wider level. Independent Klein/GPV proposals, which generate lattice vectors at a sufficiently wide Gaussian scale, are restricted to the leaf shells to initialize the tree [Klein00, GPV08]. Random child cosets preserve the prescribed parent difference coset while averaging midpoint locations for the later representation and search estimates (Sections 4 and 5.1). In Figure 2, offsets are assigned from the root toward the leaves, and records are generated in the reverse direction.

We now have independent left and right supply at every merge without changing the leading exponent. To make the resulting list useful at the next merge, we must also control how its records cluster.

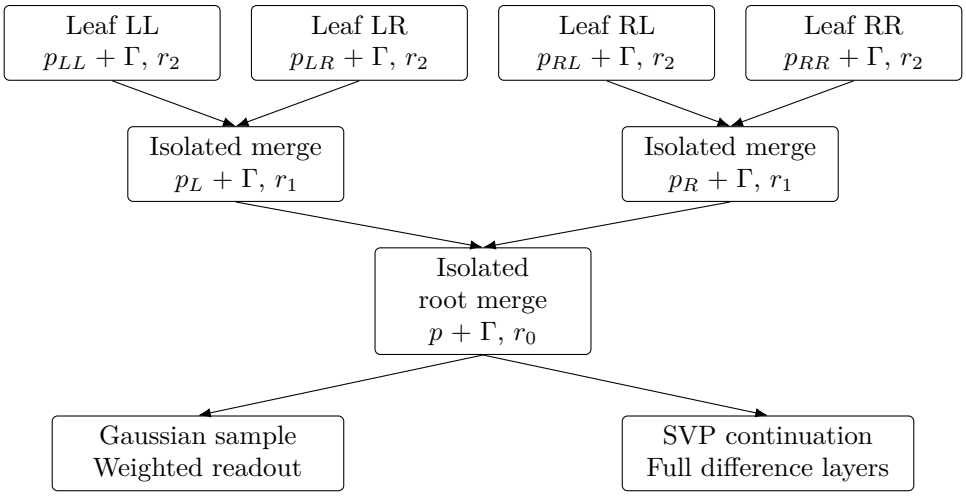



Figure 2: A depth-two isolated cooling tree, with $r_0 = \gamma r_1 = \gamma^2 r_2$. Under the respective parameter schedules, the root supplies weighted Gaussian readout or the full differencing layers used for SVP.

### 2.4 Controlling correlations by isolated pairing

The tree makes the two input lists independent, so their expected counts determine the expected supply of differences. A large expected supply can still come from very few executions. If a region $A$ has Gaussian probability $p$, both $N$ independent samples and $N$ copies of one sample give $\mathbb{E}[N_A] = Np$. Yet their probabilities of hitting $A$ are $1 - (1-p)^N$ and $p$, respectively. A large expected count can therefore come entirely from rare lists containing many copies. A second count distinguishes these behaviors.

Let $G_j$ be the reference Gaussian law on the shell and coset at round $j$, and let $M_j$ be its reference list size. Write $N_A$ for the number of records in $A$ and $N_{A,B}^{(2)}$ for the number of ordered pairs of distinct records in $A \times B$. Different occurrences of the same vector remain distinct records. With the lattice and prescribed affine data fixed, we seek

$$\begin{aligned} e^{-\varepsilon_j} M_j G_j(A) \le \mathbb{E}[N_A] \le e^{\varepsilon_j} M_j G_j(A), \\ \mathbb{E}[N_{A,B}^{(2)}] \le e^{\varepsilon_j} M_j^2 G_j(A) G_j(B). \end{aligned} \tag{3}$$

The first bound describes where the supply should appear. The second limits how often two records appear together, relative to two independent reference records, giving a quantitative bound on clustering.

The ancestry of two output records explains why these two counts are enough. Their input pairs either use four distinct child records, share a left record, or share a right record. In the first case, independence of the child lists multiplies their two factorial counts. A shared left record instead uses a first count on the left and a factorial count on the right, with the roles reversed for a shared right record. All three contributions thus depend only on the first two child moments. Repeated differencing therefore admits a closed recursion for just these two counting moments. This is what lets us follow an entire tree of correlated lists without describing their joint laws. We call this evolution of the reference counts and their moment bounds *Gaussian difference dynamics* (GDD).

For the small errors needed through many cooling rounds, we can simplify the recursion by removing the two shared input cases. View eligible pairs as edges of a bipartite graph on the child records. Keeping only edges that share neither endpoint with another edge leaves exactly the four distinct input case. To keep this operation from discarding most of the supply, we first activate

edges sparsely. Multiplying the Gaussian pair weight by a common factor $c_r$ reduces the density without changing its reference profile. These observations give the following ideal CoolingMerge at input width $r$:

1. List all input record pairs $(x, y)$ whose difference lies in $\mathcal{S}_{\gamma r}(a_0)$.
2. Activate each pair with probability $c_r \omega_r(x - y)$, using the weight from Section 2.2.
3. In the completed activated graph, keep only edges whose two endpoints both have degree one, and output their differences.

As Figure 3 illustrates, the surviving pair count is bounded by the child factorial moments, including dependence inherited from earlier ancestors.

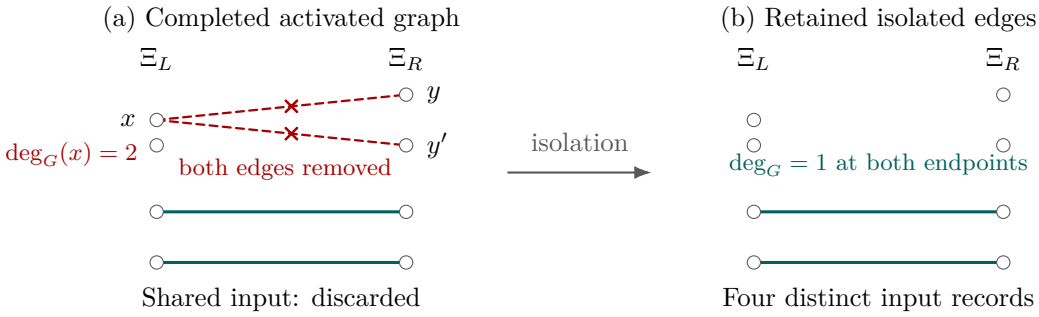



Figure 3: Isolation in CoolingMerge. Keep exactly the edges whose endpoints both have degree one in the completed activated graph. Any two surviving edges use four distinct input records, eliminating the shared input terms in the second factorial moment at this merge.

The first moment survives for a different reason. Choose $c_r$ so that the activated weighted reference degrees have a small budget $\theta$. Every deleted edge has an activated competitor, and the same child moment bounds count such competing pairs. Charging lost edges to these pairs bounds the expected isolation loss by an $O(\theta)$ fraction of the activated supply. Thus isolation gives a sharper second moment while changing the Gaussian first moment by only a small relative error.

The choice of $\theta$ balances moment accuracy against the number of retained records. Moment errors roughly double at each merge, while smaller activation probabilities reduce the expected output count. The main requirements are

$$2^t \theta = o(1), \qquad t \log_2(1/\theta) = o(n).$$

The first keeps the accumulated isolation error small. The second allows us to compensate for thinning by increasing the initial list size by only a subexponential factor. Both requirements can be met at depth $t = O(\log n \log\log n)$. Section 5 combines this choice with the geometric error budget.

We find eligible pairs with BDGL product codes [BDGL16], a structured family of spherical cap filters, using the geometric search interface of [GFH26b]. A cap of threshold $u$ has measure $(1-u^2)^{n/2+o(n)}$, so a list of size $N = (1-u^2)^{-n/2+o(n)}$ has subexponential expected cap occupancy. A target pair of correlation $u$ meets in a common cap with probability $\left(1 - 2u^2/(1+u)\right)^{n/2+o(n)}$, whose reciprocal gives

$$T_{\text{search}}(u) = \left(1 - \frac{2u^2}{1+u}\right)^{-n/2+o(n)}.$$

Inserting the lists into the filters also fits this rate, and Section 6 bounds the candidate reports for our correlated lists at the same rate. The search recovers the activated competitors needed for isolation. As $u \downarrow 1/2$, the time approaches $(3/2)^{n/2+o(n)}$.

### 2.5 One Gaussian sample by weighted readout

Cooling gives a correlated list whose counting moments follow a Gaussian reference. To extract one sample, with a possible coset restriction or width change, assign each record a weight $b(x) \in [0,1]$. For root reference measure $\nu$, the target law is

$$\pi = \frac{b\nu}{M}, \qquad M = \nu(b) > 0.$$

We call $M$ the *effective mass*. Given the completed list $\Xi = (X_i)$, let $W = \sum_i b(X_i)$ and choose record $i$ with probability $b(X_i)/W$, returning a prescribed fallback point if $W = 0$. The difficulty is this random normalization. For a region $A$, write

$$Z_A = \sum_i b(X_i)\mathbf{1}_{X_i \in A}.$$

The output probability involves $\mathbb{E}[Z_A/W;\, W > 0]$, rather than the ratio of expected counts. The numerator and denominator come from the same correlated list, so we use the two maintained moments to control their ratio directly.

Writing $\varepsilon$ for the root moment error, those bounds give

$$\mathbb{E}[Z_A] \geq e^{-\varepsilon} M\pi(A), \qquad \mathbb{E}[Z_A W] \leq e^{\varepsilon} M\pi(A)(M+1).$$

The factorial bound controls distinct records in $Z_A W$, while $b^2 \leq b$ controls its diagonal terms, producing the extra 1. Cauchy–Schwarz then gives

$$\mathbb{E}[Z_A/W;\, W > 0] \geq \frac{(\mathbb{E}[Z_A])^2}{\mathbb{E}[Z_A W]} \geq e^{-3\varepsilon} \frac{M}{M+1}\, \pi(A).$$

Since this holds for every event $A$ and the fallback adds only nonnegative mass, Theorem 4.7 gives

$$d_{\mathrm{TV}}(\mathcal{L}(X_{\mathrm{out}}), \pi) \leq 3\varepsilon + \frac{1}{M+1}.$$

Thus moment accuracy and effective mass determine the readout error, without requiring independence within the list.

For the requested Gaussian on $a + L$ at width $\sigma$, choose terminal cooling width $r = \sigma/\sqrt{a_0}$, so the root shell is centered at the typical target radius. On a root record $x \in a + \Gamma$, use

$$b_{\mathrm{out}}(x) = \mathbf{1}_{x-a \in L} e^{-\pi(\sigma^{-2} - r^{-2})(\|x\|^2 - R_-^2)},$$

where $R_-^2$ is the shell's inner squared radius. This weight restricts to $a + L$ and changes width $r$ to $\sigma$. Anchoring at $R_-^2$ keeps it in $[0,1]$, with normalized law equal to the target Gaussian conditioned on the shell.

It remains to check that these corrections leave enough effective mass. The radial weight varies by only $2^{o(n)}$ on the thin shell, while restriction to $L$ costs a factor of $[\Gamma : L]$. With cooling list exponent $P(u) = -\frac{1}{2}\log_2(1-u^2)$, this gives

$$\nu(b_{\mathrm{out}}) = \frac{\nu(\mathbb{R}^n)}{[\Gamma : L]}\, 2^{o(n)} = 2^{(P(u)-\iota)n + o(n)}.$$

For the width promise $\sigma \geq \alpha\eta_{1/2}(L)$, our extension has index rate $\iota = \max\{0, \frac{1}{2}\log_2(2a_0/\alpha^2)\}$, and

$$P(u) - \iota \geq \min\{\beta_{\mathrm{G}}, \log_2 \alpha\} > 0.$$

This is the role of $\alpha > 1$: after paying for the smoothing extension, an exponential reserve remains for the weighted readout. Gaussian concentration controls the omitted shell tail. Letting $u \downarrow 1/2$ gives time $(3/2)^{n/2+o(n)}$ and space $(4/3)^{n/2+o(n)}$ (Section 7). We next use the weaker accuracy requirement of target existence to build more representations for SVP.

### 2.6 Representations of a shortest vector

At the shortest vector scale, restricting a cooled list back to $L$ can leave too few records. We instead search for a shortest vector among differences of two terminal lists. For the analysis at the correct scale, fix a shortest vector $v$. Success depends on both the number of records at each endpoint location and the number of admissible positions $(x, x-v)$ at which the target can be represented.

Let $A$ be the reference expected number of records at each endpoint location and $\mu$ the average number of admissible positions. The two lists are generated independently conditional on their common lattice and affine data, so the reference target count is

$$\Lambda = A^2\mu.$$

This factorization guides the construction. Full differencing builds the endpoint intensity $A$, while the random affine geometry controls the number of positions counted by $\mu$ and its second moment.

Place both terminal lists on a common random affine coset $C$ of the superlattice $\Gamma$. If $N_{\pm}(x)$ count occurrences in the two lists, the actual target count is

$$Z_v = \sum_x N_+(x)N_-(x-v).$$

Quotient labels restrict differences to the original lattice: $\mathrm{lab}(x) = \mathrm{lab}(y)$ exactly when $x - y \in L$. One ideal search trial therefore has the following form:

1. Draw the superlattice $\Gamma$ and common affine coset $C$.
2. Conditional on this environment, independently generate two terminal lists on the prescribed shell in $C$.
3. Return the shortest nonzero difference between equally labeled records, or failure if there is none.

To increase $A$, we use the weaker accuracy needed for target existence. Sampling required logarithmic moment error $o(1)$, whereas target existence allows error $o(n)$ while retaining success probability $e^{-o(n)}$, provided the reference target count is large enough. On a thin shell, Gaussian weights vary by only $e^{O(nw_n)} = e^{o(n)}$, so replacing the Gaussian reference by a constant intensity fits this error allowance. After isolated cooling, a fixed number of full differencing layers retain every eligible pair, with no sparse activation or isolation. Each layer keeps the width $r$ fixed and lowers the shell energy. If $B$ is the reference number of midpoint representations for one output, then

$$A_{\mathrm{new}} = A_{\mathrm{old}}^2 B.$$

The shared input contributions from Section 4 have relative size of order $1/(A_{\mathrm{old}}B)$. Thus sufficiently large input counts both increase the endpoint intensity and control the additional moment error caused by reuse.

The geometric factor $\mu$ comes from the same midpoint picture as in Section 2.2. For endpoints of energy $a$ and a difference of energy $b$, the parallelogram identity gives

$$\mathcal{E}_r(m) \approx a - \frac{b}{4} = \frac{E}{4}, \qquad E = 4a - b.$$

The midpoint region therefore has volume $r^n(eE/4)^{n/2+o(n)}$. In an internal full layer, two such counts share the same random half grid shift, and propagating their moment bound requires $E > 1$ (Section 8).

For the terminal target, write $c = \mathcal{E}_r(v)$ and $E = 4a - c$. Here a random common affine coset lets us average over a denser ambient lattice, extending the affine slicing method of [GFH26b,

Section 6]. This average counts the midpoint volume at density $1/\det\Gamma$. Our chosen scale has $r^n/\det\Gamma = 2^{n/2+o(n)}$, so

$$\mu = \frac{r^n}{\det\Gamma}(eE/4)^{n/2+o(n)} = (eE/2)^{n/2+o(n)}.$$

Thus $E > 2/e$ already supplies exponentially many positions on average. The terminal step can use smaller endpoint shells than an internal layer, reducing the final search cost.

To make this average useful, we must also control collisions of representation positions modulo $L$. The Kabatiansky–Levenshtein bound gives at most $2^{\kappa_{\mathrm{pack}}n+o(n)}R^n$ lattice points within radius $R\lambda_1(L)$ for fixed $R \geq 1$, where $\kappa_{\mathrm{pack}} = 0.401$ (Lemma 3.4). Applied to these collisions, it gives the leading factor $2^{(\kappa_{\mathrm{pack}}-1/2)n}(ec)^{-n/2}$. Choosing

$$c = \frac{2^{2\kappa_{\mathrm{pack}}-1}}{e}$$

balances these factors. The resulting geometric second moment, combined with the list moments, gives $Z_v > 0$ with probability $e^{-o(n)}$ whenever $\Lambda \geq e^{-o(n)}$ (Section 9).

The required target count now determines the terminal list size. Write $2^{L_{\mathrm{term}}n+o(n)}$ for that size and $u = 1 - c/(2a)$ for the target correlation. Expressing $A$ in terms of the list size and substituting the geometric count $\mu$ gives

$$\frac{1}{n}\log_2\Lambda = 2L_{\mathrm{term}} - P(u) - \kappa_{\mathrm{pack}} + o(1).$$

We set the leading rate to zero and enlarge the lists by a subexponential factor to absorb the moment and geometric losses. This fixes the list size needed to represent a shortest vector.

Geometric filters then replace the exhaustive terminal comparison. We choose their cap threshold $\tau$ by $P(\tau) = L_{\mathrm{term}}$, so each filter has subexponential expected occupancy. As in the cooling search, the reciprocal common cap probability gives time

$$T_{\mathrm{term}} = \left(1 - \frac{2\tau^2}{1+u}\right)^{-n/2+o(n)}.$$

Thus the target count determines the list size, and the list size determines the filter threshold. With these two balances, letting $E$ approach $2/e$ gives time exponent

$$-\frac{1}{2}\log_2\left(1 - 2^{2\kappa_{\mathrm{pack}}-2}\right) = 0.412961878\ldots.$$

Proposition 9.3 derives the terminal curve. Taking $E$ just above $2/e$ puts the time exponent below 0.4130.

Earlier list sizes follow backward from this terminal requirement. The pair supply calculation of Section 2.2 gives

$$L_j = 2L_{j-1} - P(u_j), \qquad L_{j-1} = \frac{L_j + P(u_j)}{2}, \qquad u_j = 1 - \frac{a_j}{2a_{j-1}}.$$

This is the same balance that sustained supply through cooling, now used to work backward from a target. A constant number of full layers connects the cooling shell to the terminal shell, with all internal searches below the terminal cost. The largest intermediate list determines space. Section 10 gives parameters realizing $2^{0.4130n+o(n)}$ time and $2^{0.3138n+o(n)}$ space, including subexponential repetition. Finally, the radius mixture in Section 12.4 converts SVP into centered Gaussian sampling at every width with the same exponents.

### 2.7 Independent Gaussian batches

The single sample construction reads once from a correlated list, whereas a batch requires joint control of all outputs. We retain the Gaussian shell geometry and replace isolated pairing by the fixed bank corrections of Doulgerakis and Laarhoven [DL26, Sections 3 and 5]. At each merge, reuse the right child stream as a *bank* $C = (y_i)$ and use fresh left trials, independent of the bank and of one another.

Write $q(x)$ for the probability that a left trial returns $x$, and $K(x, y) = \mathbf{1}_{\text{Eligible}}\omega_r(x - y)$ for the pair weight. For the fixed bank, put

$$A_C(x) = \sum_i K(x, y_i), \qquad B_C(z) = \sum_i q(z + y_i)K(z + y_i, y_i).$$

These are the total partner weight for an input $x$ and the total representation weight for a difference $z$. The target weight is $h(z) = \rho_{\gamma r}(z)$ on the prescribed output shell and coset, and zero elsewhere. Choose common constants $D, c > 0$ so that every good bank satisfies $A_C(x) \le D$ and $B_C(z) \ge ch(z)$ on the relevant supports. For a successful left trial $x$, continue with probability $A_C(x)/D$, choose bank record $i$ proportionally to $K(x, y_i)$, and accept $z = x - y_i$ with probability $ch(z)/B_C(z)$. Any failed trial or rejection returns failure. For one input pair with $z = x - y_i$, the returned probability mass is

$$q(x)\frac{A_C(x)}{D}\frac{K(x, y_i)}{A_C(x)}\frac{ch(z)}{B_C(z)} = \frac{q(x)K(x, y_i)}{D}\frac{ch(z)}{B_C(z)}.$$

The first rejection cancels the partner normalization $A_C(x)$. Summing over all representations of the same $z$ then cancels $B_C(z)$ in the second rejection:

$$\Pr[\text{trial returns } z \mid C] = \frac{ch(z)}{DB_C(z)}\sum_i q(z + y_i)K(z + y_i, y_i) = \frac{c}{D}h(z).$$

The remaining failure probability, $1 - (c/D)\sum_z h(z)$, is also bank independent. Thus fresh left trials and private choices give conditionally independent outputs with the same complete law on every good bank. Averaging over good banks preserves this product law (Algorithm 5).

Smoothing and uniform concentration establish the bank bounds with common $D, c$. An iid reference tree uses this trial law at every merge, and coupling until the first bad bank transfers its product law to the construction (Sections 11.1 and 11.2).

For $B$ outputs, a per-sample tail $\delta$ can cost $B\delta$ in joint error, so we cover a wider annulus with polynomially many thin bins. Each bin uses its own smoothing extension, followed by restriction to $a + L$ and correction to width $\sigma$. Mixing with probabilities inversely proportional to the known coefficients gives the Gaussian weight throughout the annulus. An additional terminal merge supplies the inner bins for a full batch at $\sigma \ge 3\eta_{1/2}(L)/2$.

The list size, correction, and search schedules give total time $(3/2)^{n/2+o(n)}$ and space $(4/3)^{n/2+o(n)}$. At the $3/2$ width threshold they supply $\lfloor(4/3)^{n/2}\rfloor$ outputs jointly close to independent target Gaussians. Every fixed width factor $\alpha > 1$ supports the batch rate in Theorem 1.2.

## 3 Preliminaries and lattice tools

Gaussian identities propagate mass through the differences, random superlattices supply smoothing, and lattice point bounds control both terminal representations and dual Gaussian mass. Basis preprocessing provides the initial samples. All logarithms without a subscript are natural. Constants in asymptotic notation are uniform over the compact parameter ranges specified below.

### 3.1 Ideal analysis and finite realization

We analyze ideal sampling primitives, exact geometric tests, and exact probability weights. Lists retain multiplicities, so different records may store the same vector. The ideal lists are defined independently of the search codes, allowing their counting moments to control both target representations and the work needed to find them.

The finite realization result in Section 6.4 transfers these list size and search estimates to bounded word RAM executions, with the stated errors and unchanged leading exponents. Record encodings, data structures, and exceptional execution rules are collected in Appendix C. We first introduce the counting measures and Gaussian tools used by the analysis.

### 3.2 Counting measures and conditioning

For nonnegative measures on a countable set, write $\mu(f) = \sum_x f(x)\mu(x)$, $(b\mu)(x) = b(x)\mu(x)$, and $(\mu \otimes \nu)(x, y) = \mu(x)\nu(y)$. Measure inequalities are pointwise, and $T_*$ denotes pushforward. For $\Delta(x) = (x, x)$, write $\Delta_*\mu$ for the diagonal measure. A reference $\nu = NG$ has intended list size $N$ and probability profile $G$. We use $\mathcal{L}(X)$ for the law of $X$ and $d_{\mathrm{TV}}$ for total variation distance.

A record consists of an ID and a vector value. For a random finite ID set $I$ and values $(X_i)_{i\in I}$, the list's counting measure is $\Xi = \sum_{i\in I} \delta_{X_i}$. Thus $\Xi(A)$ counts records in $A$, including multiplicities: IDs remain distinct even when their values coincide. Its *intensity* and *second factorial measure* are

$$m_\Xi = \mathbb{E}[\Xi], \qquad S_\Xi = \mathbb{E}\left[\sum_{\substack{i,j\in I\\ i\neq j}} \delta_{(X_i,X_j)}\right]. \tag{4}$$

The exclusion is by ID, so for $N_x = \Xi(\{x\})$, $S_\Xi(x, x) = \mathbb{E}[N_x(N_x - 1)]$ can be positive. The ordinary second moment restores the diagonal:

$$\mathbb{E}[\Xi^{\otimes 2}] = S_\Xi + \Delta_* m_\Xi. \tag{5}$$

Consequently, for nonnegative weights,

$$\mathbb{E}[\Xi(f)\Xi(g)] = S_\Xi(f \otimes g) + m_\Xi(fg). \tag{6}$$

This correction enters both target counts and weighted selection.

**Conditioning.** Conditional moments replace the expectations in (4) by $\mathbb{E}[\cdot \mid \mathcal{H}]$, where $\mathcal{H}$ is the retained information. We form products of child moments with their common environment fixed, then average that environment. This order preserves the dependence introduced by shared shifts. Table 4 specifies the retained information for each use.

### 3.3 Gaussian scales and complexity parameters

For an affine lattice $C$, let $\rho_t(C) = \sum_{x\in C} \rho_t(x)$. It is convenient to measure squared lengths in units of the Gaussian scale. We use the normalized energy and shells

$$V_r = \frac{r^n}{\det \Gamma}, \qquad \mathcal{E}_r(x) = \frac{2\pi\|x\|^2}{nr^2}, \qquad \mathcal{S}_r(a) = \{x : \mathcal{E}_r(x) \in a(1 \pm w_n)\}. \tag{7}$$

The notation $a(1 \pm w_n)$ denotes the closed interval $[a(1 - w_n), a(1 + w_n)]$.

Three functions describe the leading costs along the sequence of shells:

$$J(z) = \frac{z - 1 - \log z}{2 \log 2}, \qquad P(u) = -\frac{1}{2} \log_2(1 - u^2), \qquad W(z) = \frac{1}{2} \log_2(2ez). \tag{8}$$

Here $J$ is the Gaussian shell cost, $P$ the angular pair and cap cost, and $W$ the shell volume exponent when $V_r = 2^{n/2+o(n)}$. The rate for two directions to enter a common cap of threshold $\tau$ is

$$K_\tau(u) = -\frac{1}{2} \log_2 \left(1 - \frac{2\tau^2}{1+u}\right) \tag{9}$$

whenever the logarithm has positive argument. If $b = 2a(1-u)$ and $E = 4a - b$, then

$$\frac{1}{2} \log_2(eE/2) = 2W(a) - P(u) - W(b). \tag{10}$$

The identity counts midpoint representations from the input and output list rates. The shell width and logarithmic error in shell counts are

$$w_n = \lceil \log_2(n+2) \rceil^{-4}, \qquad \ell_n = n w_n + \lceil \log_2(n+2) \rceil.$$

These are the default widths used for SVP and batches. The direct single sample algorithm uses the wider shells of Appendix C.3.2. In either case the local Gaussian estimates use $w_n \to 0$, $n w_n^2 \to \infty$, and smoothing and list size allowances larger than the logarithmic shell errors. The default error and list size allowances of Equation (163) satisfy

$$\ell_n \ll h_n \ll \chi_n \ll b_n \ll n.$$

At the scheduled cooling depths, $t(h_n + \ell_n) = o(b_n)$, so the extra initial records can absorb the losses from all rounds. The Gaussian tail scale is $n w_n^2$. Table 3 collects the quantities used across sections. Local normalization constants are introduced with their respective kernels. Explicit budget choices and their compatibility are proved in Appendix C.1.2.

### 3.4 Theta estimates

Fix a full rank lattice $\Gamma \subset \mathbb{R}^n$ and a scale $s > 0$ with

$$\rho_{1/s}(\Gamma^* \setminus \{0\}) \le \varepsilon_n < 1. \tag{11}$$

This is the usual smoothing condition $s \ge \eta_{\varepsilon_n}(\Gamma)$ of [MR07]. We use the Euclidean norm and the dual defined by integer pairings, $\Gamma^* = \{y \in \mathbb{R}^n : \langle y, x \rangle \in \mathbb{Z} \text{ for all } x \in \Gamma\}$.

**Lemma 3.1.** *For every $t \ge s$ and every shift $b \in \mathbb{R}^n$,*

$$(1 - \varepsilon_n) \frac{t^n}{\det \Gamma} \le \rho_t(\Gamma + b) \le (1 + \varepsilon_n) \frac{t^n}{\det \Gamma}. \tag{12}$$

*For $\Gamma/t_0$, with positive integer $t_0$, the same estimate holds at parameters at least $s/t_0$. For any lattice $K$ and any $t > 0$, $\rho_t(K + b) \le \rho_t(K)$.*

These standard bounds follow from Poisson summation [MR07].

| Quantity | Meaning and use |
|---|---|
| $\Xi,\ m_\Xi,\ S_\Xi$ | Counting measure of the record list, expected record counts, and expected ordered counts of distinct records (Equation (4)). |
| $\nu = NG$ | Reference expected counts. $N$ is its total mass and $G$ its normalized shape. |
| $V_r,\ \mathcal{E}_r,\ \mathcal{S}_r(a)$ | Gaussian volume factor, normalized squared length, and thin shell (Equation (7)). |
| $J(z),\ P(u),\ W(z)$ | Gaussian shell cost, angular pair cost, and shell volume rate (Equation (8)). |
| $K_\tau(u)$ | Common cap cost for threshold $\tau$ and correlation $u$ (Equation (9)). |
| $w_n,\ \ell_n$ | Relative shell width and logarithmic error in shell counts. |
| $h_n,\ \chi_n,\ b_n$ | Smoothing and thinning budget, SVP repetition budget, and extra initial list exponent. |
| $\Upsilon_j$ | Coefficient of the Gaussian reference in Equation (40). The shell mass converts it to an expected list size (Equation (41)). |
| $A_j,\ L_j$ | Expected record count per location in a full layer, up to moment error, and its list size exponent (Equations (79) and (104)). |
| $a_j,\ E_j$ | Shell energy and midpoint bridge parameter $E_j = 4a_{j-1} - a_j$ (Equation (71)). |

Table 3: Notation shared by the counting, geometry, and algorithm sections. The number of sampling trials, the number of stored records, and the number of representations of a target are counted separately.

### 3.5 Basis reduction and initial Gaussian samples

The following proposition relates the block dimension to the cost and quality of the initial basis.

**Proposition 3.2.** *Given a full rank rational basis $B_0$ of a rank-$n$ lattice $\Gamma$ and an integer $2 \le k < n$, put*

$$H(n,k) = 2k^{(n-1)/(2(k-1))}.$$

*One can deterministically compute a size-reduced basis $B = (b_1, \dots, b_n)$ satisfying*

$$\max_i \|b_i^*\| \le H(n,k)\lambda_n(\Gamma).$$

*The algorithm makes polynomially many exact Euclidean SVP calls of rank at most $k$. Its time and workspace are $4^{k+o(k)}\,\mathrm{poly}(\mathrm{size}(B_0))$. All projected bases, lifted vectors, and integer coordinates have polynomial encoding length in the original input.*

For fixed $0 < \vartheta < 1$, the choice $k = \lfloor \vartheta n \rfloor$ gives $H(n,k) = O_\vartheta(n^{1/(2\vartheta)})$. Both Gaussian applications use $k = \lfloor n/10 \rfloor$, giving preprocessing exponent 0.2 and a Gram–Schmidt factor $O(n^5)$. Appendix A.3 proves these claims by applying the DBKZ theorem of [ALNS20, Theorem 2.3] in successive projections, with exact SVP supplied by [MV13]. Fixed small dimensions use exact SVP.

**Estimating the shortest length.** For the SVP scale grid, [ALNS20, Corollary 4.4] with an exact oracle gives, for every fixed integer $c \ge 1$, a nonzero $v \in L$ with

$$\lambda_1(L) \le \|v\| \le D_c(n)\lambda_1(L), \qquad D_c(n) = K_c(n+1)^c, \tag{13}$$

where $K_c$ is an effective constant. With the oracle of [MV13], the cost is $4^{n/(c+1)+o(n)}\,\mathrm{poly}(\mathrm{size}(B_0))$, using polynomially many calls of rank at most $\lfloor n/(c+1) \rfloor$ for large $n$ and polynomial encoding length. We use $c = 9$, so this estimate also has preprocessing exponent 0.2.

The ideal leaf pools use independent proposals followed by a correction whose accepted mass has a known Gaussian coefficient. The following transference estimate relates this starting width to the terminal smoothing scale.

A shortest dual vector and its negative give $2e^{-\pi s^2\lambda_1(\Gamma^*)^2} \le \varepsilon_n$. Euclidean transference $\lambda_n(\Gamma)\lambda_1(\Gamma^*) \le n$ [Ban93, Theorem 2.1] therefore yields

$$\lambda_n(\Gamma) \le n\sqrt{\frac{\pi}{\log(2/\varepsilon_n)}}\, s \le n\sqrt{\frac{\pi}{\log 2}}\, s. \tag{14}$$

For $G = \max_i \|b_i^*\|$, this gives $G \le H(n,k)n\sqrt{\pi/\log 2}\, s$, uniformly for $\varepsilon_n < 1$, which supplies a polynomially wide initial sampler.

**Lemma 3.3.** *Let $B$ be a basis of $\Gamma$, let $p+\Gamma$ be an affine coset, and let $\mathcal{S}$ be a prescribed shell. If*

$$R \ge G\sqrt{\log(8n)/\pi}, \qquad G = \max_i \|b_i^*\|,$$

*one independent ideal trial returns a point or a failure symbol $\dagger$, with probability*

$$q_{R,p}(x) = \frac{\rho_R(x)}{2V_R}\mathbf{1}_{x\in(p+\Gamma)\cap\mathcal{S}} \tag{15}$$

*of returning each vector $x$, and the remaining probability at $\dagger$. The trial uses a sequential Klein/GPV proposal and a rejection probability strictly below one. With rational input and squared width, it has a finite implementation of any prescribed TV accuracy $\delta$ in polynomial time and space in the input length and $\log(1/\delta)$.*

The correction cancels the proposal's scalar normalizers, so the coefficient in (15) is known without computing the total Gaussian mass on the coset. Appendix A.8 proves the law and its counting moments.

### 3.6 Lattice point counts and random superlattices

For SVP, a lattice point count and a random superlattice supply smoothing at the scale of a shortest vector. Put

$$\kappa_{\mathrm{pack}} = \frac{401}{1000}, \qquad A_{\mathrm{G}} = \frac{e}{2^{2\kappa_{\mathrm{pack}}}}, \qquad c = \frac{1}{2A_{\mathrm{G}}} = \frac{2^{2\kappa_{\mathrm{pack}}-1}}{e}. \tag{16}$$

In particular, $W(c) = \kappa_{\mathrm{pack}}$.

**Lemma 3.4.** *Write $\beta_{\mathrm{KL}} = 2^{\kappa_{\mathrm{pack}}} = 2^{0.401}$. For every fixed $R_0 \ge 1$, there are constants $C_{R_0}$ and $n_{R_0}$ such that, for every $n \ge n_{R_0}$ and every full rank lattice $L \subset \mathbb{R}^n$, simultaneously for all $1 \le R \le R_0$,*

$$\#\{x \in L : \|x\| \le R\lambda_1(L)\} \le C_{R_0}(n+1)\beta_{\mathrm{KL}}^n R^n.$$

The underlying spherical-code bound is due to Kabatiansky–Levenshtein [KL78]. The bounded radius lattice form used here is [GFH26a, Lemma A.1], also [GFH26b, Lemma 3.8]. Its polynomial factor contributes $O(\log n)$ to the rates. We will also use the following weighted form, proved by integration in Appendix A.2.

**Lemma 3.5.** *For fixed $R_0 \ge 1$, set $\xi_w = \|w\|^2/\lambda_1(L)^2$. Uniformly for positive $H, c$ with $H/c \le R_0^2$,*

$$\sum_{w\in L\setminus\{0\}} (H - c\xi_w)_+^{n/2} \le \mathrm{poly}(n)\, 2^{\kappa_{\mathrm{pack}} n}\left(\frac{H^2}{4c}\right)^{n/2}.$$

*The polynomial depends only on $R_0$.*

We combine it with the GFH random superlattice [GFH26a, Sections 3–5], [GFH26b, Proposition 3.10].

**Proposition 3.6.** *Suppose a squared scale $Q$ satisfies $Q \le \lambda_1(L)^2 < (1+1/n)Q$. Choose a rational approximation to*

$$s^2 = \frac{2\pi A_{\mathrm{G}}}{n} Q\,(1 + O(1/n)).$$

*There is a prime $p = \Theta(2^{h_n} \det L/s^n)$ of polynomial bit length such that, for $K = p^{-1}L$ and a uniformly random one-dimensional subspace $U \subset K/L \simeq \mathbb{F}_p^n$, its inverse image $\Gamma$ has*

$$[\Gamma : L] = p, \qquad \frac{s^n}{\det \Gamma} = 2^{h_n + O(1)}, \qquad \mathbb{P}\left[\rho_{1/s}(\Gamma^* \setminus \{0\}) > 2^{-h_n/3}\right] \le 2^{-2h_n/3 + O(\log n)}. \tag{17}$$

*For every permitted scale and prime, the ambient lattice $K$ satisfies the following unconditional bound, before the random line is chosen. For every fixed $a_* > 0$, uniformly over shifts $b$ and $t \ge a_* s$,*

$$\rho_t(b + K) = \frac{t^n}{\det K}(1 + O(2^{-\Omega(h_n)})). \tag{18}$$

The line representation is $\Gamma = L + \mathbb{Z}Bv/p$ for uniform nonzero $v \in \mathbb{F}_p^n$, with scale and prime fixed before the draw. Call $\Gamma$ good when $\rho_{1/s}(\Gamma^* \setminus \{0\}) \le 2^{-h_n/3}$, and on this event put $r_0 = \sqrt{2}s$. For any shortest vector $v$, the actual normalized energy satisfies

$$\mathcal{E}_{r_0}(v) = c_n = c + O(1/n), \qquad V_{r_0} = 2^{n/2 + h_n + O(1)}. \tag{19}$$

The estimates below are uniform in this perturbation. At the affine terminal step we first use the unconditional estimate on the dense ambient lattice, then restrict to good superlattices.

# 4 Counting moments under differencing

Two differences $x - y$ and $x - y'$ share an input. To control the resulting dependence, we classify two distinct output records by their ancestry at the current merge: they use four distinct child records, share a left record, or share a right record. Conditional independence of the two child lists expresses each case through their first two factorial moments. The classification therefore closes without tracking the joint law of an entire list.

Full differencing keeps all three contributions and increases the number of target representations. Isolation removes the shared endpoint cases at the current merge, giving the sharper control needed through many cooling rounds. We first establish these counting identities for a general pair map and activation rule. Sections 5 and 8 supply the Gaussian estimates that make each operation useful.

## 4.1 The first two moments of one merge

Consider a merge of two record lists under fixed information that makes them independent, using the counting measures and conditional moments of Section 3.2. For each ordered pair of child IDs $(i, j)$, let $B_{ij} \in \{0, 1\}$ indicate whether the pair is activated. The pair produces the value $T(X_i, Y_j)$ when $B_{ij} = 1$, with

$$\mathbb{E}[B_{ij} \mid \Xi_L, \Xi_R] = q(X_i, Y_j) \in [0, 1].$$

The fixed information is suppressed here. Distinct pair coins are conditionally pairwise independent given the completed child lists. The full merge retains one record for every activated pair. Different

input record pairs remain distinct even when they produce the same vector. This convention makes the following ancestry classification exact.

The positive maps

$$\Phi_q(f,g)(z) = \sum_{T(x,y)=z} q(x,y)f(x)g(y), \tag{20}$$

$$\Psi_q(U,V)(z,z') = \sum_{\substack{T(x,y)=z\\ T(x',y')=z'}} q(x,y)q(x',y')U(x,x')V(y,y') \tag{21}$$

transport one and two copies of the input counts.

**Lemma 4.1** (Exact factorial recursion)**.** *Under the preceding conditional independence and activation law, the full merge satisfies*

$$m_a = \Phi_q(m_L, m_R), \tag{22}$$

$$S_a^{\mathrm{act}} = \Psi_q(S_L^{\mathrm{act}}, S_R^{\mathrm{act}}) + \Psi_q(\Delta_* m_L, S_R^{\mathrm{act}}) + \Psi_q(S_L^{\mathrm{act}}, \Delta_* m_R). \tag{23}$$

*All moments are conditional on the fixed information of the merge.*

*Proof.* Linearity and independence of the children give the first identity. For the second, classify two distinct parent IDs according to whether their child IDs are distinct on both sides, shared only on the left, or shared only on the right. Conditional pairwise independence of their coins gives the three terms. Sharing both IDs would give the same parent record, which the factorial measure excludes. Equal vector values can still belong to different IDs and remain included. □

We compare the actual moments to a reference measure $\nu$, whose value at each point specifies the expected number of records there. Its total mass is the intended list size, and one error parameter controls both these expected counts and pairs of distinct records.

**Definition 4.2.** Let $\mathcal{H}$ be retained information, and let the nonnegative measure $\nu$ and number $\varepsilon \geq 0$ be $\mathcal{H}$-measurable. Write $\Xi \in \mathcal{C}_{\mathcal{H}}(\nu, \varepsilon)$ if, almost surely, its conditional moments satisfy

$$e^{-\varepsilon}\nu \leq m_\Xi \leq e^{\varepsilon}\nu, \qquad S_\Xi \leq e^{\varepsilon}\nu^{\otimes 2}. \tag{24}$$

The comparisons are pointwise on the countable underlying space. We omit $\mathcal{H}$ when conditioning on a fixed complete external environment $\Omega$.

A reference on a prescribed coset and shell is extended by zero outside that support. The first upper bound then ensures that all records lie in the same support almost surely. Nonnegative summation also gives the corresponding bounds for weighted counts and pairs. The parameter $\varepsilon$ measures logarithmic distortion: sampling will require $\varepsilon = o(1)$, whereas target extraction can allow $\varepsilon = o(n)$.

## 4.2 Affine shifts and full differencing

The two children share an affine shift, so their references vary together. For a parent coset $p+\Gamma$, the local shift $d_a$ gives child cosets $p/2 + d_a + \Gamma$ and $-p/2 + d_a + \Gamma$. Once $d_a$ is fixed, the children use independent private shifts, leaf pools, and coin seeds. We can therefore multiply their conditional moments, then average this product over $d_a$. Averaging the references separately would lose their shared geometric dependence.

Before drawing $d_a$, let $\mathcal{H}_a$ retain the lattice, scales, the node's prescribed affine offset, and ancestor assignments. After the draw, $\mathcal{A}_a = \mathcal{H}_a \vee \sigma(d_a)$ also fixes the child offsets. The child moments are conditional on $\mathcal{A}_a$, and averaging $d_a$ gives the parent moments conditional on $\mathcal{H}_a$. Bounds uniform over child offsets let us repeat this calculation throughout the tree.

Table 4 records what is fixed and what is averaged in each application. Cooling bounds hold with all external shifts fixed. Full merges average the local shift, while search work also averages the ancestor shifts that determine the support density.

| Use | Retained information | Order of averaging |
|---|---|---|
| Cooling | $\Omega$: lattice, scales, and all external affine shifts | Average the samples and pair coins within each independent subtree. |
| Full merge | $\mathcal{H}_a$: lattice, scales, prescribed node offset, and ancestor assignments | Reveal $d_a$. Form products of child moments conditional on $\mathcal{A}_a$. Then average $d_a$. |
| Search work | $\mathcal{B}_v$: lattice, scales, and prescribed affine data independent of the shifts being averaged | Reveal local and ancestor shifts. Multiply conditional child intensities. Then average those shifts (Section 6). |

Table 4: Conditioning in the moment and work estimates. Fix the shared shifts, multiply the child moments, then average the indicated shifts.

For references $\nu_{L,d}, \nu_{R,d}$ at the induced child offsets, put

$$\lambda_d = \Phi_{q_d}(\nu_{L,d}, \nu_{R,d}), \tag{25}$$

$$Q_{L,d} = \Psi_{q_d}(\Delta_*\nu_{L,d}, \nu_{R,d}^{\otimes 2}), \qquad Q_{R,d} = \Psi_{q_d}(\nu_{L,d}^{\otimes 2}, \Delta_*\nu_{R,d}). \tag{26}$$

The two $Q$ terms bound the contribution from reusing a child record. The next theorem controls the merge error using the child moment bounds and two geometric estimates for the parent reference.

**Theorem 4.3** (Full differencing)**.** *Consider a full merge under the preceding activation law and affine offset construction.*

Child bounds. *Suppose the two children satisfy $\mathcal{C}_{\mathcal{A}_a}(\nu_{L,d}, \varepsilon_L)$ and $\mathcal{C}_{\mathcal{A}_a}(\nu_{R,d}, \varepsilon_R)$, where $\varepsilon_L, \varepsilon_R$ are $\mathcal{H}_a$-measurable bounds uniform over $d$. Set $D = \varepsilon_L + \varepsilon_R$.*

Reference bounds. *For a reference $\nu'$ and numbers $\eta, \rho \geq 0$, all $\mathcal{H}_a$-measurable, assume*

$$e^{-\eta}\nu' \leq \mathbb{E}_d[\lambda_d] \leq e^{\eta}\nu', \tag{27}$$

$$\mathbb{E}_d[\lambda_d^{\otimes 2} + Q_{L,d} + Q_{R,d}] \leq e^{\eta}(1+\rho)(\nu')^{\otimes 2}. \tag{28}$$

*Here and below $\mathbb{E}_d[\cdot]$ keeps $\mathcal{H}_a$ fixed.*

Parent bound. *Then the full output satisfies*

$$\Xi_a \in \mathcal{C}_{\mathcal{H}_a}(\nu', D + \eta + \log(1+\rho)). \tag{29}$$

*Proof.* First condition on $\mathcal{A}_a$. Positivity and Lemma 4.1 give

$$e^{-D}\lambda_d \leq m_a \leq e^{D}\lambda_d,$$
$$S_a^{\mathrm{act}} \leq e^{D}(\lambda_d^{\otimes 2} + Q_{L,d} + Q_{R,d}).$$

Each term uses one moment from each child, including the terms with a shared record, so all have the same factor $e^D$. Take conditional expectations over $d$ and apply the two closure bounds. □

On a finite binary tree, the update adds the leaf errors $\varepsilon_i$ and the local losses $c_a = \eta_a + \log(1+\rho_a)$:

$$\varepsilon_{\mathrm{root}} = \sum_{i \text{ leaf}} \varepsilon_i + \sum_{a \text{ internal}} c_a. \tag{30}$$

Here the bounds must be uniform over the offsets being averaged and determined by the retained root information.

The theorem separates the counting argument from the geometry of a particular kernel. To use it, one supplies a reference intensity, bounds its variation under the shared shift, and bounds the two shared input terms.

### 4.3 Isolation for accurate cooling

Sampling needs closer control than target existence. Retain an activated pair only if no other activated pair uses either endpoint. Distinct surviving pairs then use four distinct child IDs. The lost intensity is bounded by the expected number of activated competitors. Isolation removes the two shared input contributions in (23), leaving a product bound whose small error can pass through the growing cooling depth. Full differencing retains those terms and controls them with larger input lists when the geometry permits it.

**Proposition 4.4.** *Fix information under which the input lists are independent and satisfy factorial moment bounds with references $\nu_L, \nu_R$ and errors $\varepsilon_L, \varepsilon_R$. Set $D = \varepsilon_L + \varepsilon_R$. Use the preceding activation law with $q = cK \le 1$, and keep only isolated edges. Put*

$$\lambda = \Phi_q(\nu_L, \nu_R), \qquad \theta = c(D_+ + D_-),$$
$$D_+ = \sup_x \sum_y K(x,y)\nu_R(y), \qquad D_- = \sup_y \sum_x K(x,y)\nu_L(x).$$

*Then the conditional moments of the output obey*

$$(e^{-D} - e^D\theta)\lambda \le m_{\mathrm{iso}} \le e^D\lambda, \tag{31}$$
$$S_{\mathrm{iso}}^{\mathrm{act}} \le e^D\lambda^{\otimes 2}. \tag{32}$$

*Proof.* Before isolation, the intensity is between $e^{-D}\lambda$ and $e^D\lambda$. Charge a deleted edge to an activated competitor. A shared left ID uses the first moment on the left and the second factorial measure on the right, contributing at most $e^D cD_+\lambda$. A shared right ID contributes at most $e^D cD_-\lambda$. A union bound proves (31). Two surviving edges have distinct IDs on both sides. Dropping their isolation indicators and using the first term of (23) proves (32). □

**Corollary 4.5.** *Assume the input hypotheses of Proposition 4.4 conditional on $\mathcal{A}_a$, with $\mathcal{H}_a$-measurable $D$ and a degree bound $\theta$ uniform over $d$. Let $\nu'$ and $0 \le \eta < 1$ also be $\mathcal{H}_a$-measurable, and suppose*

$$(1-\eta)\nu' \le \mathbb{E}_d[\lambda_d] \le (1+\eta)\nu', \qquad \mathbb{E}_d[\lambda_d^{\otimes 2}] \le (1+\eta)^2(\nu')^{\otimes 2}. \tag{33}$$

*If $D \le 1/10$ and $\eta, \theta \le 1/4$, the output satisfies $\mathcal{C}_{\mathcal{H}_a}(\nu', \varepsilon')$ with*

$$\varepsilon' = D + 4(\eta + \theta). \tag{34}$$

*Pointwise closure $(1-\eta)\nu' \le \lambda_d \le (1+\eta)\nu'$ gives the same conclusion in each complete environment in which the input moment, independence, activation, and degree assumptions hold.*

Averaging (31)–(32) gives factors $e^{-D}(1-e^{2D}\theta)(1-\eta)$ and $e^{D}(1+\eta)^2$. Elementary logarithmic bounds yield (34).

**Corollary 4.6.** *Consider a balanced isolated tree of depth $t$, with leaves at level $t$ and root at level 0. Suppose its activation probabilities are valid, its conditioning follows the preceding construction, and every level satisfies the closure and degree assumptions of Corollary 4.5 with uniform $\eta_j, \theta_j \in [0, 1/4]$. If the leaf error is at most $\varepsilon_t$ and*

$$B_t = 2^t \varepsilon_t + 4 \sum_{j=1}^{t} 2^{j-1}(\eta_j + \theta_j) \le 1/20, \tag{35}$$

*then the root satisfies the factorial moment bounds with error at most $B_t$. One may condition on the prescribed node offset or on the complete environment, using the same convention throughout the tree.*

*Proof.* Iterating $\bar{\varepsilon}_{j-1} = 2\bar{\varepsilon}_j + 4(\eta_j + \theta_j)$ gives $\bar{\varepsilon}_0 = B_t$. Every intermediate envelope is at most $B_t \le 1/20$, so (34) applies throughout the induction. □

### 4.4 From moment bounds to outputs

Weighted selection normalizes by the actual list's random total weight. To control this normalization, we compare the weighted count in an event with its mixed second moment against the total weight. Cauchy–Schwarz turns these two bounds into a lower bound for the selection probability of every event. The resulting comparison with the weighted reference has error linear in the moment distortion and inverse effective mass.

**Theorem 4.7** (Weighted readout)**.** *Let $\Xi \in \mathcal{C}_{\mathcal{H}}(\nu, \varepsilon)$ be finite. For an $\mathcal{H}$-measurable weight $b : \mathcal{X} \to [0, 1]$, put*

$$0 < M = \nu(b) < \infty, \qquad W = \sum_i b(X_i), \qquad \pi_{\mathcal{H}} = \frac{b\nu}{M}.$$

*Choose record $i$ with probability $b(X_i)/W$ when $W > 0$, and return a fixed fallback point when $W = 0$. Then, almost surely,*

$$d_{\mathrm{TV}}\big(\mathcal{L}(X_{\mathrm{out}} \mid \mathcal{H}), \pi_{\mathcal{H}}\big) \le 1 - e^{-3\varepsilon}\frac{M}{M+1} \le 3\varepsilon + \frac{1}{M+1}, \tag{36}$$

$$\mathbb{P}[W = 0 \mid \mathcal{H}] \le 1 - e^{-3\varepsilon}\frac{M}{M+1}. \tag{37}$$

*Proof.* Fix $\mathcal{H}$ and write $\pi = b\nu/M$. For an event $A$, put $Z_A = \Xi(b\mathbf{1}_A)$ and $\nu_A = \nu(b\mathbf{1}_A)$. The moment bounds and $b^2 \le b$ imply

$$\mathbb{E}[Z_A] \ge e^{-\varepsilon}\nu_A, \qquad \mathbb{E}[Z_A W] = S_\Xi((b\mathbf{1}_A) \otimes b) + m_\Xi(b^2\mathbf{1}_A) \le e^{\varepsilon}\nu_A(M+1).$$

On $W = 0$ we have $Z_A = 0$, and define $Z_A/W = 0$ there. If $\nu_A > 0$, Cauchy–Schwarz applied to $\sqrt{Z_A/W}$ and $\sqrt{Z_A W}$ gives

$$\mathbb{E}[Z_A/W; W > 0] \ge \frac{(\mathbb{E}[Z_A])^2}{\mathbb{E}[Z_A W]} \ge e^{-3\varepsilon}\frac{M}{M+1}\pi(A).$$

For $\nu_A = 0$ this lower bound is immediate. The fallback contributes nonnegative event mass. Thus the output law dominates $\alpha\pi$, where $\alpha = e^{-3\varepsilon}M/(M+1) < 1$. It can be written as $\alpha\pi + (1-\alpha)R$ for a probability law $R$, proving the first inequality in (36). The second follows from $1 - e^{-3\varepsilon} \le 3\varepsilon$.

Finally $\mathbb{E}[W] \ge e^{-\varepsilon}M$ and $\mathbb{E}[W^2] \le e^{\varepsilon}(M^2 + M)$. The second moment inequality gives $\mathbb{P}[W > 0] \ge \alpha$, proving (37). □

**Corollary 4.8.** *If $\Xi \in \mathcal{C}_{\mathcal{H}}(\nu, \varepsilon)$ and $F = \nu(v) > 0$, then*

$$\mathbb{P}[\Xi(\{v\}) > 0 \mid \mathcal{H}] \geq \frac{e^{-3\varepsilon}}{1 + 1/F}. \tag{38}$$

*In particular, $\varepsilon = o(n)$ and $F \geq e^{-o(n)}$ give probability $e^{-o(n)}$.*

This is (37) with $b = \mathbf{1}_{\{v\}}$.

**Averaging good environments.** Suppose each good environment has the same normalized target $\pi$, moment error at most $\varepsilon_*$, and effective mass at least $M_* > 0$. Convexity of TV then gives the unconditional bound

$$\varepsilon_{\mathrm{env}} + \varepsilon_{\mathrm{impl}} + 1 - e^{-3\varepsilon_*} \frac{M_*}{M_* + 1},$$

where the first two terms pay for bad environments and implementation disagreement. A weight randomized independently of the list conditional on retained information is handled by additionally conditioning on that weight, provided these uniform hypotheses still hold.

The weighted choice can be performed by one scan using constant additional record space. Its finite realization is described in Appendix C.1.8.

# 5 Gaussian cooling by isolated pairing

A Gaussian difference naturally has a larger width than either input. To obtain a narrower profile, we select pairs at a suitable angle and correct the weight of their difference. Thin shells make this correction nearly constant on the exponential scale, while sparse activation and isolation preserve enough moment precision for another round. This section specifies the shell, ELIGIBLE, and pair weight interfaces of Algorithm 1, then proves its levelwise moment guarantee.

## 5.1 Shells and pair weights for repeated cooling

We use the shell geometry of Section 2.2. At input width $r$, write $z = x - y$, $M = (x + y)/2$, and $A_r = \pi(\gamma^{-2} - 1/2)/r^2$. The correction $e^{-A_r\|z\|^2}$ in (2) gives the desired Gaussian factor $\rho_{\gamma r}(z)$.

For $u = u_n \in (1/2, 31/50]$, fix

$$a_0 = \frac{1}{1+u}, \qquad a_1 = 4a_0 - 2, \qquad u_1 = u, \qquad \gamma^2 = 2(1-u) = \frac{a_1}{a_0}, \qquad \beta = a_0(1-u^2) = 1 - u = \frac{\gamma^2}{2}. \tag{39}$$

The identities $a_0 = 1/2 + a_1/4$ and $a_1 = \gamma^2 a_0$ put the midpoint at its typical energy and preserve the endpoint shell parameter after contraction. SVP uses $u = 47/80$. For DGS, $u$ approaches $1/2$ according to the schedule in Lemma 7.2.

Fix $\Gamma, s$ with $\epsilon_\Gamma = \rho_{1/s}(\Gamma^* \setminus \{0\}) \leq 2^{-h_n/3}$, and terminal width $r_0 = \sqrt{2}s$. This smoothing condition makes the midpoint masses almost independent of their affine shifts. The DGS application sets $s = s_0$ from the requested width. At scale $r$, use the reference

$$\nu_{p,r,\Upsilon}(x) = \frac{\Upsilon}{V_r} \rho_r(x) \mathbf{1}_{x \in (p+\Gamma) \cap \mathcal{S}_r(a_0)}. \tag{40}$$

The coefficient controls the expected number of records, while the remaining factors fix their Gaussian profile. Uniformly over the scheduled scales,

$$\nu_{p,r,\Upsilon}(\mathbb{R}^n) = \Upsilon\, 2^{-J(a_0)n + O(\ell_n)}. \tag{41}$$

Indeed, tilting to width $r\sqrt{a_0} > s$ concentrates the mass in this shell and gives the rate $J(a_0)$. The quantitative calculation appears in Appendix A.5.

For output coset $p + \Gamma$, draw a uniform $d \in (\Gamma/2)/\Gamma$ and use child cosets $p/2 + d + \Gamma$ and $-p/2 + d + \Gamma$. The child lists are supported on $\mathcal{S}_r(a_0)$. For $z = x - y$, the eligibility test is

$$\textsc{Eligible}_r(z) \iff \mathcal{E}_r(z) \in a_1(1 \pm w_n). \tag{42}$$

Thus the output lies in $\mathcal{S}_{\gamma r}(a_0)$ and has the same shell parameter at the next width. For a fixed output $z$, its admissible midpoint region is

$$\mathcal{W}_r(z) = \{M : M + z/2, M - z/2 \in \mathcal{S}_r(a_0)\}. \tag{43}$$

The endpoint identities

$$\mathcal{E}_r(M) = \frac{\mathcal{E}_r(x) + \mathcal{E}_r(y)}{2} - \frac{\mathcal{E}_r(z)}{4}, \qquad \frac{2\pi\langle M, z\rangle}{nr^2} = \frac{\mathcal{E}_r(x) - \mathcal{E}_r(y)}{2}$$

place the midpoint near energy $1/2$ and bound its normalized inner product with $z$ by $O(w_n)$. The region $\mathcal{W}_r(z)$ contains almost all midpoint Gaussian mass, as proved below.

Reweight eligible pairs by

$$R_-^2 = \frac{nr^2 a_1(1 - w_n)}{2\pi}, \qquad \omega_r(z) = e^{-A_r(\|z\|^2 - R_-^2)}.$$

On the output shell, $2^{-O(nw_n)} \le \omega_r \le 1$.

### 5.2 Sparse activation controls the moment error

The Gaussian weight sets the new profile, while a common thinning factor controls how often activated pairs share an endpoint. Normalize by an upper bound on the weighted degree, so a typical activated edge has only a small chance of acquiring a competitor.

For input coefficient $\Upsilon_j = 2^{\varrho n + \sigma_j}$, set

$$\mathsf{B}_j = e^{A_r R_-^2}\left(\frac{\gamma}{\sqrt{2}}\right)^n, \quad b_{\deg} = \lceil 10\ell_n \rceil, \quad D_j^+ = \Upsilon_j \mathsf{B}_j 2^{b_{\deg}}. \tag{44}$$

Activate a pair with probability

$$q_j(x, y) = \mathbf{1}_{x,y \in \mathcal{S}_r(a_0)} \mathbf{1}_{\mathcal{E}_r(x-y) \in a_1(1 \pm w_n)} \frac{2^{-h_n}}{D_j^+} \omega_r(x - y), \tag{45}$$

and retain edges isolated at both endpoints. Thus the algorithm uses $c_j = 2^{-h_n}/D_j^+$ when its children are at level $j$, producing level $j-1$ at width $\gamma r_j = r_{j-1}$. With this normalization, the degree budget is $\theta \le 2^{1-h_n}$, controlling the isolation loss in Proposition 4.4. The next lemma supplies the closure and degree estimates behind these choices.

**Lemma 5.1** (Gaussian closure and weighted degrees)**.** *Uniformly over (39) with $\gamma^2 \le 1 - w_n/4$, for all sufficiently large $n$,*

$$\mathsf{B}_j = 2^{-(J(a_0) + P(u_1))n + O(nw_n)}. \tag{46}$$

*For every scheduled input $r \ge r_0/\gamma$, shift, and eligible output $z = p + g$,*

$$\sum_{M \in d - g/2 + \Gamma} \rho_{r/\sqrt{2}}(M) \mathbf{1}_{M \in \mathcal{W}_r(z)} = V_{r/\sqrt{2}}(1 \pm \eta_{\mathrm{mid}}), \quad \eta_{\mathrm{mid}} = O(\epsilon_\Gamma) + e^{-\Omega(nw_n^2)}. \tag{47}$$

*Every eligible endpoint has weighted reference degree at most $D_j^+$ before the common factor $2^{-h_n}/D_j^+$.*

*Proof.* The coefficient follows from $A_r R_-^2 = n(1-a_0)(1-w_n)/2$. Radial and linear Gaussian concentration give a lower bound on an inner region of $\mathcal{W}_r(z)$, while the full shifted Gaussian mass gives the upper bound. For degrees, the endpoint and output shells bound the exponent after completion of squares. A full coset sum then gives the bound. The uniform estimates, including the validity of every tilt, are proved in Appendix A.6. □

### 5.3 Initializing and repeating the cooling step

The corrected proposals of Lemma 3.3 initialize the desired intensity exactly. Basis preprocessing gives $G \le \mathrm{poly}(n) r_0$, so a schedule $R = r_0 \gamma^{-t}$ with $t = O(\log n)$ for fixed $u$ reaches $R \ge G\sqrt{\log(8n)/\pi}$.

For fixed $C_a > 0$ and a bounded raw exponent $\varrho \ge P(u_1) + J(a_0)$, set

$$M_{\mathrm{leaf}} = \left\lceil 2^{\varrho n + C_a b_n} \right\rceil \tag{48}$$

and run $2M_{\mathrm{leaf}}$ independent corrected trials at width $R$, with shell $\mathcal{S}_R(a_0)$. Store only nonfailure records. By (137), their reference coefficient is $M_{\mathrm{leaf}}$ and their initial moment error is zero.

Number levels from the root, with leaves at level $t$, and set $r_j = r_0 \gamma^{-j}$. Fix $\Upsilon_t = M_{\mathrm{leaf}}$ and $\Upsilon_{j-1} = \Upsilon_j 2^{-h_n - b_{\deg}}$ before the construction. The complete ideal procedure is Algorithm 1, called as BuildIdeal$(p, 0)$. Conditional on the completed children, distinct pair coins are pairwise independent, with the probabilities in (45). The all-pairs loop defines the ideal law. Pair search in Section 6 recovers this graph with the stated coverage and incidence bounds. Figure 2 illustrates the affine routing and Gaussian scales in a tree of depth two.

**Algorithm 1:** Ideal cooling tree: BuildIdeal$(p, j)$

**Input:** Offset $p$, level $j$, and the fixed cooling schedule.
**Output:** A record list in $(p + \Gamma) \cap \mathcal{S}_{r_j}(a_0)$.

**if** $j = t$ **then**
  $\Xi \leftarrow$ empty record list
  **for** $i = 1, \dots, 2M_{\mathrm{leaf}}$ **do**
    Run a corrected trial at $(p, r_t, \mathcal{S}_{r_t}(a_0))$ using Lemma 3.3
    **if** the trial returns $X \neq \dagger$ **then**
      Append a record with value $X$ to $\Xi$
  **return** $\Xi$
Draw $d \sim \mathrm{Unif}((\Gamma/2)/\Gamma)$
$\Xi_L \leftarrow$ BuildIdeal$(p/2 + d, j + 1)$
$\Xi_R \leftarrow$ BuildIdeal$(-p/2 + d, j + 1)$ using independent private randomness
$r \leftarrow r_{j+1}$
$G \leftarrow$ the empty bipartite graph on the child records
**for** each record pair $(x, y) \in \Xi_L \times \Xi_R$ **do**
  $z \leftarrow x - y$
  **if** Eligible$_r(z)$ **then**
    Activate edge $(x, y)$ in $G$ with probability $c_{j+1}\omega_r(z)$, using pairwise independent coins
Keep the edges whose two endpoints both have degree one in the completed graph $G$
**return** one difference record per retained edge

Each round reduces the coefficient by $2^{-h_n - b_{\deg}}$. The extra initial records absorb these losses, while isolation keeps the accumulated moment error small. Isolation uses the completed graph so that every activated competitor enters the degree test. Appendix C.2.3 implements this test within

the stated space. The following theorem states both recursions, including the growing depth case used at the DGS endpoint.

**Theorem 5.2** (Gaussian cooling)**.** Parameters. *Choose $u = u_n$ from (39), a bounded raw sampling rate $\varrho = \varrho_n \geq P(u_1) + J(a_0)$, and depth $t = t_n$ satisfying*

$$\gamma^2 \leq 1 - w_n/4, \qquad t(h_n + \ell_n) = o(b_n).$$

*For either schedule in Appendices C.1.2 and C.3.2, we also have $t = o(nw_n^2)$ and $nw_n^2 = o(h_n)$, which ensure that the Gaussian tail error remains small through the entire tree.* Construction. *Fix $\Gamma, s$ as above, $C_a > 0$, and the external offsets. Construct the ideal isolated forest at starting width $R = r_0\gamma^{-t}$ satisfying Lemma 3.3, with disjoint pools of $2M_{\mathrm{leaf}}$ corrected trials and $M_{\mathrm{leaf}}$ given by (48).*

Root guarantee. *For all sufficiently large $n$, every root satisfies*

$$\Xi_0 \in \mathcal{C}(\nu_{p,r_0,\Upsilon_0}, \varepsilon_0), \qquad \varepsilon_0 \leq e^{-\Omega(nw_n^2)}.$$

*Its reference coefficient and expected list size obey*

$$\Upsilon_0 = 2^{\varrho n + \sigma_0}, \qquad \sigma_0 = C_a b_n - \sum_{j=1}^{t}(h_n + b_{\deg}) + O(1) = \Theta(b_n). \tag{49}$$

*In particular, the expected size is $2^{[\varrho - J(a_0)]n + \Theta(b_n)}$. All pair probabilities are at most one. Independently constructed roots are independent conditional on their common public parameters and external offsets.*

*Proof.* Set $\Upsilon_t = M_{\mathrm{leaf}}$ and $\Upsilon_{j-1} = \Upsilon_j 2^{-h_n - b_{\deg}}$. We first verify that these coefficients define valid pair probabilities. At every level, (46) gives

$$\log_2(\Upsilon_j \mathsf{B}_j) = [\varrho - J(a_0) - P(u_1)]n + C_a b_n - \sum_{k=j+1}^{t}(h_n + b_{\deg}) + O(nw_n + 1) \geq C_a b_n/2.$$

Thus $D_j^+ \geq 2^{-h_n}$, making (45) a valid probability.

Next, completion of squares and (47) identify the activated reference in each complete environment as $\kappa_d = (1 \pm \eta_{\mathrm{mid}})\nu_{p,\gamma r,\Upsilon_{j-1}}$. Indeed, writing $c_j = 2^{-h_n}/D_j^+$, its coefficient is

$$c_j \Upsilon_j^2 e^{A_r R_-^2} \frac{V_{r/\sqrt{2}} V_{\gamma r}}{V_r^2} = c_j \Upsilon_j^2 \mathsf{B}_j = \Upsilon_j 2^{-h_n - b_{\deg}}.$$

The weighted degree bound gives $\theta_j \leq 2^{1-h_n}$. By (137), the initial moment error is zero.

The error budget in Corollary 4.6 is therefore

$$\bar{\varepsilon}_0 = 4(2^t - 1)(\eta_{\mathrm{mid}} + 2^{1-h_n}) = e^{-\Omega(nw_n^2)}. \tag{50}$$

Here $t = o(nw_n^2)$ and $nw_n^2 = o(h_n)$ make the budget small enough for the local estimates at every level. The bounds hold with all external offsets fixed, and averaging private offsets preserves them because the root reference depends only on its own offset.

Summing the coefficient losses proves (49). Equation (41) gives the expected size. Independence of the private random sources gives the final conditional independence claim. □

# 6 Finding pairs with BDGL product codes

The cooling construction specifies useful differences through its counting moments. Finding them efficiently requires a search that preserves the transitions just analyzed. Spherical filters make correlated pairs meet in common buckets. Their cost separates into point incidences for insertion and traversal, candidate reports for pair testing, and true outputs for dictionary storage.

We first bound these counts on fixed lists and then average using the intensities and shell geometry of the lists generated by the algorithm. This order connects the search cost to the same lists that supply the output. Query and bank sizes are kept separate when they differ, as in the terminal batch merge.

## 6.1 Counting filter insertions and pair reports

Write $C_{d_{\mathrm{pc}}}(\tau)$ for spherical cap measure and $W_{d_{\mathrm{pc}}}(\tau, u)$ for common cap measure at correlation $u$. The product code construction uses dimension $d_{\mathrm{pc}} = n + O(\log n)$. The spherical cap and wedge estimates of [BDGL16, Lemmas 2.1–2.2] give the following rates, in the notation of [GFH26b, Section 4.1]. For $\tau \in (0,1)$ and $u \in (-1,1)$ in fixed compact ranges with $1 - 2\tau^2/(1+u)$ bounded below by a positive constant,

$$\log_2 C_{d_{\mathrm{pc}}}(\tau) = -P(\tau)n + O(\log n), \qquad \log_2 W_{d_{\mathrm{pc}}}(\tau, u) = -K_\tau(u)n + O(\log n). \tag{51}$$

The product code and decoder originate in [BDGL16, Lemma 5.1 and Theorem 5.1]. We use the following uniform coverage and restricted traversal interface from [GFH26b, Theorem 4.2, Proposition 4.3, and Appendix D], which avoids storing all filters explicitly.

**Proposition 6.1.** *Let $A, B$ be fixed finite, internally deduplicated lists of nonzero vectors with equality keys, as in [GFH26b, Algorithm 1]. Assume $\log_2 M_f = O(n)$.*

Expected counts. *For a fresh code of $M_f$ filters,*

$$\mathbb{E}[Z_A] = |A| M_f C_{d_{\mathrm{pc}}}(\tau), \qquad \mathbb{E}[Z_B] = |B| M_f C_{d_{\mathrm{pc}}}(\tau), \tag{52}$$

$$\mathbb{E}[P_{\mathrm{bucket}}] = M_f \sum_{\substack{x \in A, y \in B \\ \mathrm{key}(x) = \mathrm{key}(y)}} W_{d_{\mathrm{pc}}}(\tau, \langle \widehat{x}, \widehat{y} \rangle). \tag{53}$$

*Here $Z_A, Z_B$ count point and filter incidences, and $P_{\mathrm{bucket}}$ counts reports of equally keyed pairs in common filters.*

Coverage. *Fix $\delta > 0$ and suppose*

$$\delta \le \tau \le 1 - \delta, \qquad |u| \le 1 - \delta, \qquad \tau(1-u) \ge \delta, \qquad 1 - \frac{2\tau^2}{1+u} \ge \delta.$$

*The last two inequalities keep both cap constraints active and their intersection nondegenerate. Uniformly on this range, a fixed target pair is missed with probability $\exp[-\Omega(\log^2 n)]$ if $M_f W_{d_{\mathrm{pc}}}(\tau, u) \ge 2^{H_{d_{\mathrm{pc}}}}$, where $H_{d_{\mathrm{pc}}} = O(\sqrt{n}\ \mathrm{polylog}\, n)$.*

Resources. *For live insertion budget $S \ge |A|$, restricted Cartesian traversal uses space*

$$O(S + |A| + |B| + \mathrm{poly}(n) B_{\mathrm{pc}}) \tag{54}$$

*and, up to polynomial factors, time at most*

$$Z_A + Z_B + P_{\mathrm{bucket}} + \left(1 + \frac{nZ_A}{S}\right)(|A| + |B|) B_{\mathrm{pc}}, \qquad \log_2 B_{\mathrm{pc}} = O(n/\log n). \tag{55}$$

*Here $B_{\mathrm{pc}}$ is the block code size. Reports can be streamed, so candidate enumeration does not require their simultaneous storage.*

For target correlation $u_0$, choose the filter count

$$\log_2 M_f = K_\tau(u_0)n + C_{\mathrm{code}} b_n + O(\log n). \tag{56}$$

The padding is chosen to meet the product code coverage threshold in Proposition 6.1. Independent repetitions then reduce the probability of missing a target pair. Internal calls give every record the same key. The SVP root instead uses quotient labels, so equally keyed pairs have differences in the input lattice.

The interface extends to record lists, counting each record and pair with its multiplicity (Appendix C.2.2). During cooling, the search must recover every eligible activated edge to identify all competitors and decide isolation correctly. Complete coverage couples the implemented merge to the ideal transition analyzed by the moments. At the SVP root, it suffices to recover one target pair: the analysis fixes a witness from the completed ideal lists before drawing the terminal codes.

## 6.2 Expected search costs for the generated lists

The affine offsets determine which lattice points can occur in each input list. Averaging these offsets gives the support density used in the candidate count.

Each routing step chooses a uniform half grid coset. Thus a height-$j$ subtree below a fixed root of depth $d$ has offset uniform over $\pm p_{\mathrm{root}}/t + \Gamma/t$ modulo $\Gamma$, where $t = 2^{d-j}$ and the path fixes the sign. For every nonnegative $f$,

$$\mathbb{E}_{\mathrm{ancestors}}\left[\sum_{z \in p_j + \Gamma} f(z)\right] = t^{-n} \sum_{z \in \pm p_{\mathrm{root}}/t + \Gamma/t} f(z). \tag{57}$$

For a call at scale $r$, put $g = r^2/s^2$. A pair of endpoint energy $a$ and correlation $u$ has difference energy $b = 2a(1-u)$ and midpoint bridge energy $E = 2a(1+u)$. Counting its midpoint representations uses a Gaussian of width $r\sqrt{E}/2$ on $\Gamma/2$, and hence requires only

$$gE > 1. \tag{58}$$

At $r = r_0$, this is $2E > 1$. The stronger internal requirement $E > 1$ comes from the second moment transfer, rather than this work estimate.

To bound the search work, let $N_{I_n(u)}$ count candidate record pairs whose correlation lies in a bin $I_n(u) \subseteq [u - 1/n, u + 1/n]$, before the activation coin is applied. The next estimate separates the angular pair cost from the additional cost of counting differences below smoothing.

The work estimate follows the same order as the moment recursion: fix the shifts to multiply child intensities, then average the resulting geometric count. Here we average both the local shift $d_v$ and the ancestor shifts $\mathbf{d}_{\uparrow v}$, whose support density is given by (57). At node $v$, let $\mathcal{B}_v$ retain the lattice, scale, and prescribed affine data independent of these shifts, as in Table 4. Revealing the shifts gives $\mathcal{A}_v = \mathcal{B}_v \vee \sigma(d_v, \mathbf{d}_{\uparrow v})$, which fixes the node and child offsets and leaves the two child subtrees independent. When $t = 1$, only the local shift is averaged, so the prescribed node offset can remain in $\mathcal{B}_v$.

**Lemma 6.2** (Candidate pairs in a correlation interval)**.** *At input energy $a$ and scale $r$, suppose the two child lists are independent conditional on $\mathcal{A}_v$, with intensities at most $Ae^{O(b_n)}$ on their input shells, uniformly in the shifts being averaged. Assume the local half grid average and (57) apply with denominator $t$, and write*

$$AV_r(ea)^{n/2} = 2^{\beta n + O(b_n)}, \qquad c_0 = 2t^2 ga.$$

*For $u$ in a fixed compact correlation interval with $gE > 1$ by a fixed margin, the conditional candidate count satisfies*

$$\mathbb{E}[N_{I_n(u)} \mid \mathcal{B}_v] \le 2^{[2\beta - P(u)]n + O(b_n)} \max\{1, [c_0(1-u)]^{-n/2}\}. \tag{59}$$

*Proof.* With the shifts fixed, independence gives a product of child intensities inside the tower identity

$$\mathbb{E}[N_{I_n(u)} \mid \mathcal{B}_v] = \mathbb{E}_{d_v, \mathbf{d}_{\uparrow v}}\Big[\mathbb{E}_{\text{private}}[N_{I_n(u)} \mid \mathcal{A}_v] \,\Big|\, \mathcal{B}_v\Big].$$

For a fixed output, averaging the local half grid shift contributes $V_r(eE/4)^{n/2}e^{O(\ell_n)}$ midpoint representations. The free output shell contributes the bound (138) proved in Appendix B.1. The support density cancels its covolume, and $bE/4 = a^2(1-u^2)$ gives the stated rate. The last factor is the penalty when the output Gaussian lies below smoothing. Shell perturbations and bin width contribute $O(\ell_n)$. The moment bounds for cooling and full differencing supply the required conditional intensity estimate in each application. □

A lower bound $g_0 \le r^2/s^2$ may replace $g$ in the smoothing factors and in $c_0$, while $V_r$ retains the actual scale. For cooling, we use $g_0 = 2/\gamma^2$ to make the correlation cutoff independent of the level.

### 6.3 Time and space bounds

Let $\log_2 M_f = K_{\text{code}} n + O(b_n)$. Multiplying the spectrum by common cap probabilities controls the expected bucket reports. In the region $c_0(1-u) \ge 1$, the identity

$$P(u) + K_\tau(u) \ge 2P(\tau), \qquad (1-\tau^2)^2 - (1-u)(1+u-2\tau^2) = (u-\tau^2)^2 \tag{60}$$

gives rate $K_{\text{code}} + 2(\beta - P(\tau))$. In the subcritical region, the rate is $K_{\text{code}} + 2\beta + \frac{1}{2}\log_2((1+u-2\tau^2)/c_0)$, which increases with $u$.

For the calls used here, $\beta < 1/2$. With the strict smoothing margins of Lemma 6.2 and the cutoff conditions in Appendix B.2, the expected pair work has rate

$$T_{\text{pair}} \le \max\left\{K_{\text{code}} + 2(\beta - P(\tau)), \quad K_{\text{code}} + 2\beta + \frac{1}{2}\log_2 \frac{2(1-\tau^2)}{c_0}\right\}. \tag{61}$$

The same rate extends across the remaining correlations by the boundary estimates in Appendix B.2.

For input lists of size at most $2^{\beta n + O(b_n)}$, choose a live insertion budget $S = 2^{\beta n + C_S b_n}$. The point incidence rate is $K_{\text{code}} + \beta - P(\tau)$. Equation (55) shows that traversal adds at most this rate and the list rate $\beta$. Its live storage is $2^{\beta n + O(b_n)}$, including the implicit block codes.

Depth first evaluation stores the current lists and traversal state. The resulting list size and incidence bounds are the inputs to Proposition 6.3. Its application preserves the leading rates while accounting for incomplete search and exceptional resource use.

### 6.4 From ideal operations to finite algorithms

The incidence estimates specify what the search must recover and how much work it may use. Fixed resource limits and finite arithmetic turn these ideal constructions into algorithms whose costs are bounded on every execution. The following result collects the transfer guarantee. The common cap rule is given in Appendix C.3, with execution schedules in Appendices C.3.2, C.3.3 and C.3.5. Appendix C.2.4 proves the transfer guarantee.

**Proposition 6.3.** *Fix one of the single sample, batch, or SVP constructions below with these resource limits. Let $T_0$ be its total work cap and let $S_0$ be its peak storage cap, counting ideal operations and stored data items before the finite arithmetic overhead. These budgets include preprocessing, search workspace, and all prescribed repetitions. They have leading time and space exponents $\alpha_{\mathrm{G}}, \beta_{\mathrm{G}}$ for the sampling constructions and $0.4130, 0.3138$ for SVP. For every requested $\lambda \geq 1$, the chosen construction admits a finite word RAM implementation with*

$$T_{\mathrm{word}} \leq T_0\, 2^{o(n)}\, \mathrm{poly}(\mathcal{B}, n, \lambda), \qquad S_{\mathrm{bits}} \leq S_0\, 2^{o(n)}\, \mathrm{poly}(\mathcal{B}, n, \lambda)$$

*on every execution. Once the search and resource limits have been fixed, the finite and exact-arithmetic executions can be coupled so that their discrete transcripts of record lists, pair decisions, and sampling decisions differ with probability at most $2^{-\lambda}$. Let $\mathcal{G}$ be the good environment event used in the application. On the ideal construction, let $\delta_{\mathrm{search}}$ bound the probability that $\mathcal{G}$ holds but the required search coverage fails, and let $\delta_{\mathrm{resource}}$ bound the probability that $\mathcal{G}$ holds but a resource cap is exceeded. For sampling, the ideal and finite outputs admit a coupling with*

$$\mathbb{P}[\mathcal{G} \text{ and } X_{\mathrm{finite}} \neq X_{\mathrm{ideal}}] \leq \delta_{\mathrm{search}} + \delta_{\mathrm{resource}} + 2^{-\lambda}.$$

*For SVP, fix a shortest vector $v$ and let $E_v$ be the event that the ideal terminal lists contain a representation of $v$. Then*

$$\mathbb{P}[\text{the implemented trial finds a shortest vector}] \geq \mathbb{P}[\mathcal{G} \cap E_v] - \delta_{\mathrm{search}} - \delta_{\mathrm{resource}} - 2^{-\lambda}.$$

*The search and resource allowances are specified in Appendix C.2.4. For sampling they fit within the stated statistical errors. For SVP their per-trial sum is smaller than the ideal target probability by a factor tending to zero. Bad environment probabilities are accounted for in the respective applications. The batch allowance applies to its complete output law.*

# 7 Gaussian sampling by weighted readout

Cooling gives a correlated Gaussian list on a smoothing superlattice. To sample from the requested coset of the original lattice, we choose one record with a weight that both restricts its coset and adjusts its Gaussian width. The weighted sampling theorem reduces the analysis to showing that the expected total selection weight remains large after these two corrections.

## 7.1 Choosing the readout weight

Fix $\alpha > 1$ and rational input $(L, a, \sigma^2)$ with

$$\sigma \geq \alpha \eta_{1/2}(L), \tag{62}$$

and write $V_r(H) = r^n / \det H$. For $1/2 < u \leq 3/5$, use the shell parameters in (39), so $a_0 = 1/(1+u)$. Choose the terminal cooling width

$$r = \sigma/\sqrt{a_0}, \qquad \mathcal{S} = \mathcal{S}_r(a_0).$$

Then $\mathcal{S}$ is centered at the typical target radius $\|x\|^2 = n\sigma^2/(2\pi)$.

Suppose cooling on $\Gamma \supset L$ supplies a root $\Xi \in \mathcal{C}(\nu, \varepsilon_{\mathrm{cool}})$, where

$$\nu(x) = \frac{\Upsilon}{V_r(\Gamma)} \rho_r(x) \mathbf{1}_{x \in (a+\Gamma) \cap \mathcal{S}}.$$

The profile is wider than the target Gaussian, so restricting to $a+L$ and multiplying by $\exp[-\pi(\sigma^{-2}-r^{-2})\|x\|^2]$ gives the desired relative weights. To normalize this factor on the shell, anchor it at the inner squared radius $R_-^2$ and set

$$b_{\mathrm{out}}(x) = \mathbf{1}_{x-a\in L} e^{-\pi(\sigma^{-2}-r^{-2})(\|x\|^2-R_-^2)}. \tag{63}$$

Thus $b_{\mathrm{out}} \in [0,1]$ on the root support, and multiplication into $\nu$ gives

$$\frac{b_{\mathrm{out}}\nu}{\nu(b_{\mathrm{out}})} = D_{a+L,\sigma}(\cdot \mid \mathcal{S}). \tag{64}$$

One weighted selection therefore has distance at most

$$3\varepsilon_{\mathrm{cool}} + \frac{1}{M_{\mathrm{eff}}+1}, \qquad M_{\mathrm{eff}} = \nu(b_{\mathrm{out}}),$$

from this conditional target by Theorem 4.7. The Gaussian tail outside $\mathcal{S}$ is $e^{-\Omega(nw_n^2)}$ by Appendix B.8.2. It remains to choose the smoothing extension and the cooling list size so that $M_{\mathrm{eff}}$ is large.

## 7.2 Paying for the smoothing extension

At terminal width $r$, the cooling estimates use smoothing at $s = r/\sqrt{2}$. A denser superlattice provides it, while restriction back to $L$ costs a factor equal to its index. The following construction quantifies this tradeoff and will also be used for radial bins in Section 11.

**Lemma 7.1.** *Let $s/\sigma$ be bounded below by a fixed positive constant and put $\iota = \max\{0, \log_2(\sigma/(\alpha s))\}$. Choose a prime with $\log_2 p = \iota n + 3h_n + O(1)$. For a uniform nonzero $v \in \mathbb{F}_p^n$, set*

$$\Gamma = L + \mathbb{Z}Bv/p, \qquad L = B\mathbb{Z}^n. \tag{65}$$

*Then $[\Gamma : L] = p$, and*

$$\mathbb{P}[\rho_{1/s}(\Gamma^* \setminus \{0\}) > 2^{-h_n}] \le 2^{-2h_n+O(1)}. \tag{66}$$

*Proof.* The input promise and theta monotonicity bound the dual mass at $s$ by $(3/2)2^{\iota n}$. A random modular constraint retains each unforced dual vector with probability at most $1/p$, while the mass on $pL^*$ is $2^{-\Omega(p^2)}$. Markov's inequality gives the claim. Appendix B.8.1 supplies the mass calculation. □

Call the extension good when $\rho_{1/s}(\Gamma^* \setminus \{0\}) \le 2^{-h_n}$. At our chosen width $s = \sigma/\sqrt{2a_0}$, the index rate is

$$\iota = \max\left\{0, \frac{1}{2}\log_2 \frac{2a_0}{\alpha^2}\right\}.$$

On a good extension, the effective mass can be computed directly from (63):

$$M_{\mathrm{eff}} = \frac{\Upsilon}{p} a_0^{n/2} e^{n(1-a_0)(1-w_n)/2} \frac{\rho_\sigma((a+L)\cap\mathcal{S})}{V_\sigma(L)}. \tag{67}$$

The last ratio lies between positive constants for large $n$ by Appendix B.8.2, and the middle factors equal $2^{-J(a_0)n+O(nw_n)}$. Hence the readout spends the shell rate $J(a_0)$ on width correction and the index rate $\iota$ on returning to the original lattice.

To pay these costs, we choose the following list, raw sampling, and forward search rates:

$$\ell(u) = P(u), \qquad m(u) = P(u) + J(a_0), \tag{68}$$
$$K(u) = K_u(u) = \frac{1}{2}\log_2 \frac{1+u}{(1-u)(1+2u)}.$$

With raw trial rate $m(u)$, Theorem 5.2 gives root coefficient $\log_2 \Upsilon = m(u)n + \Theta(b_n)$ and expected list size $2^{P(u)n+\Theta(b_n)}$, provided the depth and padding absorb the accumulated cooling loss. We choose such a schedule below. Substituting this coefficient into (67), using $nw_n + h_n = o(b_n)$, cancels the shell cost:

$$\log_2 M_{\text{eff}} = (P(u) - \iota)n + \Theta(b_n). \tag{69}$$

If $\iota = 0$, then $P(u) \geq \beta_{\text{G}}$. Otherwise, $2a_0(1-u^2) = \gamma^2 < 1$ gives

$$P(u) - \iota = \log_2 \alpha - \frac{1}{2}\log_2 \gamma^2 \geq \log_2 \alpha.$$

In either case,

$$P(u) - \iota \geq \min\{\beta_{\text{G}}, \log_2 \alpha\} > 0.$$

Thus exponentially large effective mass survives the return to $L$. This positive margin is what makes every fixed $\alpha > 1$ sufficient: the smoothing extension enables the cooling estimates while leaving enough total selection weight for one Gaussian readout.

### 7.3 Reaching the BDGL exponents

The forward search rate $K(u)$ determines time, while the list rate $P(u)$ determines space. To approach the BDGL exponents, let $u$ decrease toward $1/2$. This slows the contraction, so we choose the schedule below to keep the overhead from the additional cooling levels subexponential.

**Lemma 7.2.** *Set*

$$k_n = \max\{10, \lceil \log_2 \log_2(n+4) \rceil\}, \qquad u_n = \frac{1}{2} + \frac{1}{k_n}, \qquad \gamma_n^2 = 1 - \frac{2}{k_n},$$

*and use $a_0, a_1$ from (39). Starting from a basis with $G \leq \text{poly}(n)s$ and an endpoint width $r = \Theta(s)$, a depth*

$$t = O(k_n \log n) = O(\log n \log\log n) \tag{70}$$

*reaches the proposal threshold of Lemma 3.3. For either schedule in Appendices C.1.2 and C.3.2,*

$$t(h_n + b_{\text{deg}} + 2) = o(b_n), \qquad K(u_n) \to \alpha_{\text{G}}, \quad P(u_n) \to \beta_{\text{G}}, \quad m(u_n) \to \frac{1}{2} - \frac{1}{6\log 2} < \alpha_{\text{G}}.$$

*Proof.* Since $-\log \gamma_n \geq 1/k_n$, the scheduled width $r\gamma_n^{-t}$ reaches any prescribed polynomial multiple of $s$ within the stated depth. The default budgets (163) give total loss $O(n \log\log n / \log n) = o(b_n)$, and (167) gives the same conclusion for the single sample schedule. Substitution yields the rate limits. The energies remain in compact positive ranges and $k_n w_n \to 0$, so the shell and search estimates hold uniformly along the schedule. □

For the accuracy exponent $\xi > 0$ in Theorem 1.1, choose a fixed rational $0 < \delta \leq \min\{1, \xi\}$ and shell width $w_n \asymp (\log n)^{-1-\delta/2}$. The budgets of Appendix C.3.2 absorb the cooling loss above and give $nw_n^2 = \Theta(n/\log^{2+\delta} n)$. Algorithm 2 combines the construction with one weighted selection. Its bounded implementation is given in Appendix C.3.2.

**Algorithm 2:** One Gaussian sample: GaussianReadout

**Input:** A rational lattice $L$, shift $a$, squared width $\sigma^2$, and fixed $\alpha > 1$, $\xi > 0$ as in Theorem 1.1.
**Output:** One point of $a + L$.

Choose the endpoint parameters of Lemma 7.2 and set $r = \sigma/\sqrt{a_0}$, $s = r/\sqrt{2}$
Construct the smoothing extension $\Gamma$ at scale $s$ and preprocess its basis
Generate a cooling root $\Xi$ in $a + \Gamma$ at width $r$ using BuildIdeal$(a, 0)$
Select one record from $\Xi$ with weights $b_{\mathrm{out}}(x)$ in (63), using the readout convention of Theorem 4.7
**return** its value

*Proof of Theorem 1.1.* On each good extension, cooling and the effective mass estimate (69) give $\varepsilon_{\mathrm{cool}} \le e^{-\Omega(nw_n^2)}$ and $M_{\mathrm{eff}} = 2^{\Omega_\alpha(n)}$. The weighted sampling theorem therefore bounds the distance from (64) by $3\varepsilon_{\mathrm{cool}} + 1/(M_{\mathrm{eff}} + 1)$. Because the normalized target is the same on every good extension, this comparison survives averaging over extensions. Adding the shell tail and bad extension probability gives

$$d_{\mathrm{TV}}(\mathcal{L}(X_{\mathrm{ideal}}), D_{a+L,\sigma}) \le e^{-\Omega(nw_n^2)} + 2^{-2h_n+O(1)}.$$

The implementation in Appendix C.3.2 adds search, resource, and numerical loss at most $2^{-\Omega(n)} + 2^{-\Omega(b_n)} + 2^{-Cn}$. These errors and the bad extension term are smaller than the shell and moment error under (167), so the final distance is at most

$$e^{-\Omega(nw_n^2)} \le \exp[-\Omega(n/\log^{2+\xi}(n+2))].$$

For time, forward pair search is the dominant operation. Its expected work at a node is $2^{K(u_n)n+O(b_n)}$ by Appendix B.8.3, and the $2^{O(t)}$ nodes multiply this by only $2^{o(n)}$. The other operations fit this rate: raw sampling has rate $m(u_n) \to 1/2 - 1/(6\log 2) < \alpha_{\mathrm{G}}$, basis preprocessing has rate $1/5$, and the final weighted scan has the smaller list rate $P(u_n)$. The work caps in Appendix C.3.2 therefore give time $2^{K(u_n)n+o(n)} = 2^{\alpha_{\mathrm{G}}n+o(n)}$.

For space, depth first evaluation retains the current lists and search workspace, while leaf proposals and candidate reports are streamed. Each list has rate $P(u_n)$, and the weighted scan needs only one additional record. With the caps of Appendix C.3.2, the bit space on every execution is $2^{P(u_n)n+o(n)} = 2^{\beta_{\mathrm{G}}n+o(n)}$. □

**Remark.** Our recent work [GH26] identifies a route to improving the BDGL sieving exponent through faster matrix multiplication. We find that this approach also applies to our provable single sample DGS algorithm. Specifically, let $\omega$ denote the matrix multiplication exponent. If $\omega < 9/4$, then for every fixed $\alpha > 1$, a single sample can be obtained with the same width range and accuracy as in Theorem 1.1 in $(3/2 - \delta)^{n/2+o(n)}$ time for some constant $\delta = \delta(\alpha, \omega) > 0$. The result can be obtained by replacing pair enumeration within the Gaussian cooling buckets with compressed correlation search and enlarging the buckets to reduce the number of filters. The surplus supply available for a single output allows us to strengthen smoothing, which controls the recovery cost of highly correlated pairs. The fixed-list coverage guarantee preserves the ideal cooling transitions, while GDD correlation estimates bound the search work after averaging the generated lists.

Polynomially many independent executions give a joint Gaussian approximation with the same leading resource exponents. To obtain an exponential number of outputs within the same total cost, Section 11 reuses partner lists and corrects each trial's complete output law.

# 8 Gaussian shells and full differencing

At the shortest vector scale, restricting to the input lattice can consume many records in the cooled list. To create more representations before this restriction, the SVP algorithm keeps every eligible difference through a fixed number of layers. Each layer squares the expected count at a location and multiplies it by the number of possible midpoints. This gain permits the larger moment error allowed when the goal is to find one target vector.

For each output $z$, its *midpoint lens* consists of the midpoints whose two endpoints lie in the input shells. We must control both the size of this lens under a shared affine shift and the effect of reusing an input record. The former requires bridge energy $E > 1$, while a sufficiently large input list controls the latter. The terminal coset argument in Section 9 will allow the smaller threshold $E > 2/e$ because no further moment propagation is needed there.

Here we keep the width $r = \sqrt{2}s$ fixed and vary the shell energy from one layer to the next. Thus smaller energies give shorter vectors at the same reference width. Assume (11) with $\varepsilon_n \to 0$ and set $w = w_n$.

## 8.1 Counting representations by their midpoints

For a transition from child energy $a$ to output energy $b$, assume

$$E = 4a - b \geq 1 + \kappa \tag{71}$$

with fixed $\kappa > 0$. At parent offset $p$, draw $d \in (\Gamma/2)/\Gamma$ uniformly and use

$$C_+ = p/2 + d + \Gamma, \qquad C_- = -p/2 + d + \Gamma.$$

For a point $z = p + g$ in the prescribed output coset and shell, with $e_z = \mathcal{E}_r(z) \in b(1 \pm w)$, let $J_d(z)$ count pairs $z = x - y$ with $x \in C_+ \cap \mathcal{S}_r(a)$ and $y \in C_- \cap \mathcal{S}_r(a)$. Put

$$\beta_z = a - e_z/4, \quad E_z = 4\beta_z, \quad \sigma_z = r\sqrt{\beta_z}, \quad \mathcal{B}_z = V_r(e\beta_z)^{n/2}, \quad \mathcal{B} = V_r(eE/4)^{n/2}. \tag{72}$$

In midpoint coordinates the normalized energy is approximately $E/4$. The parameter $E$ makes the second moment smoothing condition take the simple form $E > 1$. Writing $M = (x+y)/2 \in d - g/2 + \Gamma$, the shell conditions are exactly

$$|\mathcal{E}_r(M) - \beta_z| + \left|\frac{2\pi\langle M, z\rangle}{nr^2}\right| \leq aw. \tag{73}$$

The inner rectangle has both summands at most $aw/2$, and the outer radial window has $|\mathcal{E}_r(M) - \beta_z| \leq aw$. Averaging the local shift counts on a finer lattice:

$$\mathbb{E}_d\left[\sum_{M \in d - g/2 + \Gamma} f(M)\right] = 2^{-n} \sum_{M \in \Gamma/2} f(M), \qquad f \geq 0. \tag{74}$$

**Theorem 8.1** (Midpoint lens moments)**.** *Let $(a,b)$ range over a fixed compact subset of $(0,\infty)^2$ satisfying (71) for a fixed $\kappa>0$. Under (11) with $\varepsilon_n\to 0$, the following estimates hold uniformly over this range for all sufficiently large $n$. There is a constant $C$ depending only on the range and $\kappa$ such that, with $\eta_n = C(nw+\log(n+2))$,*

$$e^{-\eta_n}\mathcal{B} \le \mathbb{E}_d[J_d(z)] \le e^{\eta_n}\mathcal{B}, \tag{75}$$

$$\mathbb{E}_d[J_d(z)J_d(z')] \le e^{\eta_n}\mathcal{B}^2. \tag{76}$$

*The second estimate uses the same $d$ in both copies and is uniform over $z,z'$ in the same parent coset and output shell, including $z=z'$.*

*Proof.* Apply the window count of Lemma A.1 on $\Gamma/2$ at width $\sigma_z$, using the inner rectangle and outer radial window from (73). Its radial tilts have a fixed smoothing margin since $4\sigma_z^2/s^2 = 2E_z > 1$. The factor $2^{-n}$ in (74) cancels the finer lattice's covolume, and $\beta_z = E/4 + O(w)$ gives (75).

For the second moment, write $G_d(t) = \rho_t(d+\Gamma)$. The half grid has a useful square-sum identity:

$$\mathbb{E}_d[G_d(t)^2] = 2^{-n}\rho_{\sqrt{2}t}(\Gamma)^2. \tag{77}$$

Indeed, the expanded sum runs over $x,y\in\Gamma/2$ with $x-y\in\Gamma$. The map $(x,y)\mapsto(x+y,x-y)$ is a bijection onto $\Gamma^2$, with inverse $(u,v)\mapsto((u+v)/2,(u-v)/2)$. The parallelogram identity separates its Gaussian weight into $\rho_{\sqrt{2}t}(u)\rho_{\sqrt{2}t}(v)$.

The outer window gives $J_d(z) \le e^{n/2+O(nw)}G_{d-g/2}(\sigma_z)$. Since $g\in\Gamma$, translating $d$ by $g/2$ permutes its possible values. Also $2\sigma_z^2 = E_z s^2 > s^2$, so Lemma 3.1 yields

$$\mathbb{E}_d[J_d(z)^2] \le e^{n+O(nw)}2^{-n}\left((1+\varepsilon_n)\frac{(\sqrt{2}\sigma_z)^n}{\det\Gamma}\right)^2 \le e^{O(nw)}\mathcal{B}_z^2.$$

Thus the threshold $E>1$ comes from smoothing at width $\sqrt{2}\sigma_z$. Cauchy–Schwarz now bounds $\mathbb{E}_d[J_d(z)J_d(z')]$ by the geometric mean of the two square moments. Using $E_z, E_{z'} = E + O(w)$ proves (76) for the same local shift in both copies. □

## 8.2 Moment bounds for repeated full differencing

Consider a balanced tree of fixed depth $d$. At height $j$, retain pairs with kernel

$$q_j(x,y) = \mathbf{1}_{x,y\in\mathcal{S}_r(a_{j-1})}\mathbf{1}_{x-y\in\mathcal{S}_r(a_j)},$$

using the preceding local environment at every node. Cooling supplies each leaf at offset $p$ with Gaussian reference

$$\nu_{0,p}(x) = \frac{\Upsilon_0}{V_r}\rho_r(x)\mathbf{1}_{x\in(p+\Gamma)\cap\mathcal{S}_r(a_0)}. \tag{78}$$

The recursion is made explicit in Algorithm 3. Here height counts upward from the cooled leaves, whereas the level argument of BuildIdeal counts downward from a cooling root. Each call at height zero starts a fresh cooling tree with the prescribed offset. Consequently a root of height four uses sixteen cooling pools.

The all-pairs loop specifies the ideal transition. The searches of Section 6 recover these eligible pairs with the coverage and work guarantees used below. Every eligible record pair contributes a difference, preserving the multiplicities needed for the intensity recursion. The shared input terms remain in its second moment and will be bounded using the available number of records.

**Algorithm 3:** Ideal full difference tree: BUILDFULLTREE$(p, j)$

**Input:** Offset $p$, height $j$, fixed width $r$, energies $(a_i)$, and a cooling schedule ending at width $r$.
**Output:** A record list in $(p+\Gamma)\cap\mathcal{S}_r(a_j)$.

**if** $j = 0$ **then**
  **return** BUILDIDEAL$(p, 0)$ with a fresh leaf pool and private randomness
Draw $d \sim \mathrm{Unif}((\Gamma/2)/\Gamma)$
$\Xi_L \leftarrow$ BUILDFULLTREE$(p/2 + d, j-1)$
$\Xi_R \leftarrow$ BUILDFULLTREE$(-p/2 + d, j-1)$ with independent private randomness
$\Xi \leftarrow$ empty record list
**for** each record pair $(x, y) \in \Xi_L \times \Xi_R$ **do**
  **if** $x - y \in \mathcal{S}_r(a_j)$ **then**
    Append a record with value $x - y$ to $\Xi$
**return** $\Xi$

The nominal intensity per atom is

$$A_0 = \frac{\Upsilon_0}{V_r} e^{-na_0/2}, \qquad E_j = 4a_{j-1} - a_j, \qquad \mathcal{B}_j = V_r(eE_j/4)^{n/2}, \qquad A_j = A_{j-1}^2 \mathcal{B}_j. \tag{79}$$

The list size also depends on the number of output atoms. Three quantities govern the iteration. Bridge energy controls the common environment, $A_{j-1}\mathcal{B}_j$ controls the relative error from reusing an input, and $A_d$ sets the final target count. We express the needed moment bounds directly relative to the constant intensity $A_j$ on each shell, reducing their propagation to two scalar recursions.

At a leaf, the Gaussian reference differs from $A_0$ by at most $e^{\pm b_0^*}$, where $b_0^* = na_0 w_n/2$. Replacing this reference by $A_0$ costs $b_0^*$ in the first moment and $2b_0^*$ in the second. Thus cooling error $\varepsilon_{\mathrm{cool}}$ gives leaf error

$$\varepsilon_0 = \varepsilon_{\mathrm{cool}} + 2b_0^* = O(nw_n) + e^{-\Omega(nw_n^2)} = o(n). \tag{80}$$

**Theorem 8.2** (Full difference trees)**.** Geometry and expected counts. *Let $d$ be fixed, with energies in a fixed compact positive range, under the smoothing assumption of Theorem 8.1. Suppose constants $\kappa, \zeta > 0$ satisfy, at every height,*

$$E_j \geq 1 + \kappa, \qquad A_{j-1}\mathcal{B}_j \geq 2^{\zeta n}. \tag{81}$$

Leaf bounds. *Assume each leaf satisfies $\mathcal{C}(A_0 \mathbf{1}_{(p+\Gamma)\cap\mathcal{S}_r(a_0)}, \varepsilon_0)$ uniformly in its prescribed offset, with $\varepsilon_0 = o(n)$. Use independent private randomness in disjoint subtrees.*

Propagation. *For sufficiently large $n$, each height-$j$ list, conditional on its prescribed offset and ancestor assignments, satisfies*

$$e^{-\varepsilon_j} A_j \leq m_{j,p}(z) \leq e^{\varepsilon_j} A_j, \qquad S_{j,p}^{\mathrm{act}}(z, z') \leq e^{\varepsilon_j} A_j^2 \tag{82}$$

*for $z, z'$ in its prescribed coset and shell, where*

$$\varepsilon_j = 2\varepsilon_{j-1} + \eta_n + \log\left(1 + \frac{2}{A_{j-1}\mathcal{B}_j}\right), \qquad \eta_n = C(nw_n + \log(n+2)). \tag{83}$$

*The root error obeys*

$$\varepsilon_d \leq 2^d \varepsilon_0 + (2^d - 1)\eta_n + 2(2^d - 1)2^{-\zeta n} = o(n). \tag{84}$$

*Proof.* Fix the prescribed node offset and reveal its local shift $d_a$. Conditional on this information the children are independent. Write $A = A_{j-1}$, $B = \mathcal{B}_j$ and take $\nu' = A_j \mathbf{1}_{\text{support}}$. The reference intensity is $\lambda_{d_a}(z) = A^2 J_{d_a}(z)$. By Theorem 8.1,

$$e^{-\eta_n} A_j \le \mathbb{E}_{d_a}[\lambda_{d_a}(z)] \le e^{\eta_n} A_j, \qquad \mathbb{E}_{d_a}[\lambda_{d_a}(z)\lambda_{d_a}(z')] \le e^{\eta_n} A_j^2.$$

For a shared left value $x$, the right values are determined by $z, z'$. The admissible triples are a subset of those counted by $J_{d_a}(z)$. Consequently

$$\mathbb{E}_{d_a}[Q_{L,d_a}(z,z')] \le A^3 \mathbb{E}_{d_a}[J_{d_a}(z)] \le e^{\eta_n} \frac{A_j^2}{AB}.$$

The right shared input term has the same bound. Thus Theorem 4.3 applies with $\eta = \eta_n$ and $\rho = 2/(AB)$, proving (83). Now apply (30) and $\log(1 + 2/(AB)) \le 2^{1-\zeta_n}$ to obtain (84). □

For a specified root point $v$ in the output support, Corollary 4.8 gives occurrence probability $e^{-o(n)}$ whenever $A_d \ge e^{-o(n)}$. For SVP we instead apply the target bound for two lists in Lemma 9.2. In both cases, merging equal values at the terminal search preserves every represented target.

# 9 Shortest vector search in a random coset

An internal layer must support further propagation. The root only needs one shortest difference. A random affine coset lets us use this simpler requirement to lower the bridge parameter from the internal threshold $E > 1$ to $E > 2/e$. We extend the affine slicing argument of [GFH26b, Section 6, especially Proposition 6.4] from one pairing layer to the lists generated by our dynamics.

Fix a scale $Q$ satisfying $Q \le \lambda_1(L)^2 < (1 + 1/n)Q$, and take $p, s$ from Proposition 3.6. Choose a uniform one-dimensional subspace $U \subset p^{-1}L/L \simeq \mathbb{F}_p^n$ and let $\Gamma$ be its inverse image. For a fixed shortest vector $v$ used in the analysis, write

$$K = p^{-1}L, \qquad r = \sqrt{2}s, \qquad V = \frac{r^n}{\det \Gamma} = 2^{n/2 + h_n + O(1)}, \qquad c_n = \mathcal{E}_r(v) = c + O(1/n). \tag{85}$$

Here $\kappa_{\text{pack}} = 401/1000$, $ec = 2^{2\kappa_{\text{pack}} - 1}$, and $W(c) = \kappa_{\text{pack}}$. Recall $\ell_n = nw_n + \lceil \log_2(n+2) \rceil = o(h_n)$. We also have $nw_n^2 \to \infty$. The dense lattice clause of Proposition 3.6 gives shifted theta flatness on $K$ at every fixed positive multiple of $s$. The scale grid in Section 10.2 ensures that the algorithm has a branch with the required $Q$.

## 9.1 A random common coset

Independently of the random line, choose $z$ uniformly in $K/L$ and $\delta$ uniformly in $(L/2)/L$. Both root lists are constructed in

$$C = z + \delta + \Gamma. \tag{86}$$

Sampling $z$ uses $n$ field elements. Since $p$ is odd, $L \cap 2\Gamma = 2L$, so $L/2L \longrightarrow \Gamma/2\Gamma$ is an isomorphism. Thus $\delta$ is also uniform modulo $\Gamma$ on the half grid, supplying the ancestor averages for internal work.

Conditional on $(\Gamma, z, \delta)$, use disjoint raw pools, independent internal shifts, and independent pair coins for the two subtrees. Their output lists are independent, and dependence within each list is controlled by the factorial moment bounds. Store $x = z + \delta + g_x$, with $g_x \in \Gamma$, and label it by $g_x + L$. Then

$$\text{lab}(x) = \text{lab}(y) \quad \Longleftrightarrow \quad x - y \in L. \tag{87}$$

## 9.2 Counting representations in the coset

For positive child energy $a$, define

$$E = 4a - c, \qquad E_n = 4a - c_n, \qquad f_v(x) = \mathbf{1}_{\mathcal{S}_r(a)}(x)\mathbf{1}_{\mathcal{S}_r(a)}(x - v), \qquad J_C(v) = \sum_{x \in C} f_v(x). \tag{88}$$

Throughout, $a$ ranges over a fixed compact set with $E - 2/e$ bounded below by a positive constant. All estimates are uniform on this range.

**Theorem 9.1** (Affine target representations)**.** *Suppose $E > 2/e$ with a fixed positive margin. For each fixed $\delta$, let $\mu_\delta = \mathbb{E}_{\Gamma,z}[J_C(v)]$. Uniformly in $\delta$,*

$$\mu_\delta = V(eE/4)^{n/2} \exp(\pm O(\ell_n)), \tag{89}$$

$$\frac{\operatorname{Var}_{\Gamma,z} J_C(v)}{\mu_\delta^2} \le \frac{1}{\mu_\delta} + 2^{-h_n + O(\ell_n)}. \tag{90}$$

*For every fixed $\Gamma$, the conditional mean $\mathbb{E}_z[J_C(v) \mid \Gamma]$ is also $\mu_\delta$.*

*Proof.* Flatness on the dense ambient lattice counts the lens even when its midpoint width is below $s$. The Gaussian estimate (143), proved in Appendix B.4.1, gives its ambient mass.

For $t \in K/L$, put $F_t = \sum_{x \in \delta + t + L} f_v(x)$ and $\theta = p^{1-n}$. A uniform affine line contains each residue with probability $\theta$. Distinct residues have nonpositive covariance because their joint probability is $\theta(p-1)/(p^n - 1) \le \theta^2$. Hence

$$\mathbb{E}[J_C(v)] = \theta \sum_t F_t, \qquad \operatorname{Var} J_C(v) \le \theta \sum_t F_t^2. \tag{91}$$

Since $\theta r^n / \det K = V$, the ambient mass proves the mean estimate. The second sum counts collisions modulo the original lattice:

$$\sum_t F_t^2 = \sum_{w \in L} \sum_{x \in \delta + K} f_v(x) f_v(x + w). \tag{92}$$

Its diagonal contributes $\mu_\delta$ after multiplication by $\theta$. For nonzero $w$, put $\xi = \|w\|^2/\|v\|^2$. The two midpoint balls intersect inside a ball of normalized squared radius at most $(E - c\xi)/4 + O(w_n)$. The KL bound counts the possible $w$, while this shrinking intersection bounds the midpoint locations for each one. Maximizing $\xi(E - c\xi) \le E^2/(4c)$ and dividing by the squared mean gives

$$V^{-1} 2^{\kappa_{\mathrm{pack}} n} (ec)^{-n/2} e^{O(\ell_n)} = 2^{-h_n + O(\ell_n)}.$$

The choice $ec = 2^{2\kappa_{\mathrm{pack}} - 1}$ makes the packing and volume factors cancel. Appendix B.4.1 supplies the weighted radial estimate and its uniform boundary bound, proving the variance claim.

For every fixed direction $\Gamma$, averaging $z$ includes each ambient residue with the same probability $\theta$, proving the last claim. Also $\log_2 \mu_\delta = \frac{n}{2} \log_2(eE/2) + h_n + O(\ell_n)$, so $E > 2/e$ gives exponentially many representations on average. □

The lower threshold comes from averaging on the dense lattice $K$ and controlling collisions in $L$. Internal layers instead require uniform transfer over their shared half grid environments, which is where the stronger condition $E > 1$ enters.

### 9.3 A target difference in the two lists

The geometric count supplies possible endpoint locations. The following lemma turns these locations into a target occurrence using only the first two factorial measures.

**Lemma 9.2** (Target success)**.** *Fix the scale, prime, and $\delta$ above. Let $\mathcal{G}$ depend only on $\Gamma$, with $g = \mathbb{P}[\mathcal{G}] > 0$, and put $\mu = \mu_\delta$ and $\xi = 2^{-h_n+O(\ell_n)}$ as in Theorem 9.1, with $\mu \geq 1$ and $\xi \leq 1$. Conditional on each environment $(\Gamma, z, \delta)$ in $\mathcal{G}$, suppose the two record lists are independent, supported on $C \cap \mathcal{S}_r(a)$, and satisfy the following bounds with uniform $A > 0$ and $D \geq 0$:*

$$e^{-D}A \leq m_\pm(x) \leq e^D A, \qquad S_\pm(x,x') \leq e^D A^2 \quad (x, x' \in C \cap \mathcal{S}_r(a)). \tag{93}$$

*Let $Z_v$ count record pairs with difference $v$, and put $Y = \mathbf{1}_{\mathcal{G}} Z_v$ and $\Lambda = A^2\mu$. Then*

$$\mathbb{E}[Y] \geq g e^{-2D}\Lambda, \tag{94}$$

$$\mathbb{E}[Y^2] \leq e^{2D}(4\Lambda^2 + 2\Lambda). \tag{95}$$

*Consequently*

$$\mathbb{P}[\mathcal{G},\ Z_v > 0] \geq \frac{g^2}{6} e^{-6D} \min\{1, \Lambda\}. \tag{96}$$

*If $g = 1 - o(1)$, $\Lambda \geq e^{-o(n)}$, and $D = o(n)$, this is $e^{-o(n)}$.*

*Proof.* First fix a good environment. Let $\mathcal{R}_v = \{x \in C : x, x - v \in \mathcal{S}_r(a)\}$, $J = |\mathcal{R}_v|$, and write $Z_v = \sum_{x \in \mathcal{R}_v} N^+_x N^-_{x-v}$. Independence gives conditional mean at least $e^{-2D}A^2 J$. For $x \neq x'$, the product of counts within either list has expectation at most $e^D A^2$. For $x = x'$, the diagonal correction gives at most $e^D(A^2 + A)$. Consequently,

$$\mathbb{E}[Z_v^2 \mid \Gamma, z, \delta] \leq e^{2D}(J(J-1)A^4 + J(A^2 + A)^2) = e^{2D}(A^4J^2 + 2A^3J + A^2J).$$

The coset average in Theorem 9.1 equals $\mu$ for every fixed $\Gamma$. Since $\mathcal{G}$ depends only on $\Gamma$, it follows that $\mathbb{E}[\mathbf{1}_{\mathcal{G}} J] = g\mu$, proving the first bound. For the second, extend the nonnegative terms to all environments and use $\mathbb{E}[J^2] \leq (1 + \xi)\mu^2 + \mu$. The inequalities $\mu \geq 1$, $\xi \leq 1$, and $2A \leq A^2\mu + 1$ give (95). The second moment method now yields $\mathbb{P}[Y > 0] \geq g^2 e^{-6D}\Lambda/(4\Lambda + 2)$, which implies (96). □

The success estimate averages the root environment, with $\delta$ fixed, and is uniform in $\delta$. The algorithm samples every environment without testing $\mathcal{G}$. Keeping one representative of each vector value preserves target existence, so the final search can use distinct values while this analysis counts occurrences.

### 9.4 Cost of the final pair search

The affine average also bounds all other reports from the decoder. If equally labeled endpoints differ by $w \in L$, membership of one endpoint in $C$ already ensures membership of the other. Such a pair therefore has one affine membership factor $\theta$, rather than two.

The search parameters balance the terminal list size against the number of filters. We choose the list rate so that the expected number of target representations has leading exponent zero, then choose the cap threshold so that point incidence work has the same rate as the filter count. For an energy $a$ satisfying Theorem 9.1, set

$$u = 1 - \frac{c}{2a}, \qquad \mathsf{L} = \frac{P(u) + \kappa_{\mathrm{pack}}}{2}, \qquad P(\tau) = \mathsf{L}. \tag{97}$$

We merge equal vector values within each list and use quotient labels to restrict the search to differences in $L$. The filter count is (56), so $\log_2 M_f = K_\tau(u)n + O(b_n)$.

**Proposition 9.3.** *Fix $0 < \kappa < 1/2$ and $c = 2^{2\kappa-1}/e$. Under the balance (97) with $\kappa_{\mathrm{pack}} = \kappa$ and $E = 4a - c > 2/e$, the terminal work rate is*

$$K_{\mathrm{term}}(E) = -\frac{1}{2}\log_2\left(2^{1-\kappa}\sqrt{c/E} - c/E\right). \tag{98}$$

*Its infimum on this curve is*

$$\inf_{E>2/e} K_{\mathrm{term}}(E) = -\frac{1}{2}\log_2(1 - 2^{2\kappa-2}). \tag{99}$$

*Proof.* Put $x = \sqrt{c/E}$. The identities $1 - u^2 = cE/(4a^2)$ and $1 - \tau^2 = 2^{-2\mathsf{L}}$ give $1 - 2\tau^2/(1+u) = 2^{1-\kappa}x - x^2$. The energy condition is equivalent to $0 < x < 2^{\kappa-1}$. Since $\kappa < 1/2$, the quadratic is increasing throughout this interval. Letting $x \uparrow 2^{\kappa-1}$ proves the infimum. □

At $\kappa_{\mathrm{pack}} = 0.401$, the infimum is $0.412961878\ldots$. Our parameters approach this value while keeping a fixed positive geometric margin. Earlier energies can reduce space, whereas a substantial time improvement requires a different terminal balance or geometric estimate. Table 5 records the costs and the geometric margin.

**Proposition 9.4.** *Use the preceding search parameters, with $1 - 2\tau^2/(1+u) > 0$ by a fixed margin. Conditional on each root environment $(\Gamma, z, \delta)$ with $\Gamma$ good, suppose the two lists are independent, supported on $C \cap \mathcal{S}_r(a)$, and satisfy $m_\pm(x) \le e^{D_n} A_n$ uniformly, where*

$$D_n = o(n), \qquad \log_2 A_n = (\mathsf{L} - W(a))n + o(n). \tag{100}$$

*Then the following hold for the search of Proposition 6.1.*

*(i) Averaged over the root environment conditional on a good superlattice, the expected list sizes are at most $2^{\mathsf{L}n+o(n)}$, and the expected point incidences and bucket reports are at most*

$$2^{K_\tau(u)n+o(n)}. \tag{101}$$

*If $D_n$, the amplitude remainder, and code padding are $O(b_n)$, the logarithmic remainders in both bounds are $O(b_n)$.*

*(ii) The code padding can be chosen so that every fixed shortest vector representation has the coverage guarantee of Proposition 6.1.*

*Proof.* First apply the uniform intensity bounds on the good event. The remaining geometric sums are nonnegative, so extend them to all environments and divide by the good event probability $1 - o(1)$. The affine average then has its original uniform direction.

The average list size is at most $A_n V(ea)^{n/2} e^{D_n + O(\ell_n)}(1 + o(1))$. Its rate is $\mathsf{L}$, and the point incidence rate is $\mathsf{L} + K_\tau(u) - P(\tau) = K_\tau(u)$. For nonzero differences, the conditional intensities multiply, followed by the single affine membership factor. Appendix B.4.2 combines the midpoint volume with common cap probability, uniformly up to the cap boundary.

The choice of parameters balances this work against the target count. Put $\mathcal{B}_E = V(eE/4)^{n/2}$ and $Q = 4a(1 - \tau^2)$. Then

$$A_n^2 \mathcal{B}_E = 2^{o(n)}, \qquad \frac{Q^2}{4cE} = 2^{-2\kappa_{\mathrm{pack}}}. \tag{102}$$

These follow from $1 - u^2 = cE/(4a^2)$ and $1 - \tau^2 = 2^{-2\mathsf{L}}$. The first sets the target mass. The second cancels the factor from Lemma 3.4 when summing the filtered difference counts, giving rate $K_\tau(u)$. The appendix supplies the uniform bin estimates.

After deduplication, each zero pair in a filter is charged to one point incidence. A shortest vector representation has correlation $u + O(w_n)$ with the stated active constraint margin, giving coverage. All other logarithmic losses are $O(\ell_n + \log n)$, which also proves the quantitative $O(b_n)$ assertion. □

## 10 Solving the shortest vector problem

The SVP algorithm cools thirty-two pools, combines them through four full difference layers, and searches the resulting two lists on a common affine coset. We choose the list sizes backward from the mass needed for one shortest difference. This makes the target condition, rather than the initial sampler, determine the required list sizes.

Each cooled pool is a tree of the form in Figure 2 and Algorithm 1. The thirty-two pools form the leaves of two full difference trees of depth four, followed by the terminal affine pair search. Cooling retains only isolated activated edges, whereas the four full layers retain every eligible pair. Their analysis uses the conditioning on prescribed offsets in Table 4: average private child randomness, form the conditional products, and only then average the shared local shift. This order lets the same two moment bounds pass through the four full layers.

### 10.1 Choosing the list sizes from the target

Use the rates in (8) and the scale constants in (16), so $W(c) = \kappa_{\mathrm{pack}} = 401/1000$. Let $a_0, \dots, a_4$ be the energies of the generated lists and $a_5 = c$ the target energy. For each transition, write

$$u_j = 1 - \frac{a_j}{2a_{j-1}}, \qquad E_j = 4a_{j-1} - a_j. \tag{103}$$

A terminal list of rate $L_4$ has nominal intensity rate $L_4 - W(a_4)$ per location. Combining two such intensities with the midpoint count, (10) gives target mass rate

$$2L_4 - P(u_5) - W(c).$$

We set this rate to zero and add sublinear padding to make the actual mass large. Earlier full layers obey the list size balance $L_j = 2L_{j-1} - P(u_j)$, so all required list rates follow backward:

$$L_4 = \frac{P(u_5) + W(c)}{2}, \qquad L_{j-1} = \frac{L_j + P(u_j)}{2} \quad (1 \le j \le 4). \tag{104}$$

The terminal geometry anchors the energy choice. Proposition 9.3 shows that choosing $E_5 = 4a_4 - c$ just above $2/e$ puts the terminal work rate near 0.412962. It also fixes $a_4 \approx 0.264120$ and $L_4 \approx 0.260971$. The four internal layers connect this endpoint to the cooling shell $a_0 = 1/(1 + u_{\mathrm{s}})$. Their energies must meet three goals:

(i) leave a positive cooling list margin, internal bridge margins $E_j > 1$, and abundance margins for the moment transfer,
(ii) keep cooling and internal search work below the terminal rate,
(iii) reduce the largest intermediate list, which determines space.

The backward recurrence (104) couples these choices: changing an internal energy changes both adjacent correlations and the list size required at every earlier level. Scheme A below keeps cooling and the largest internal search close to the terminal rate while reducing peak storage. The following rational parameters realize this tradeoff with fixed positive margins.

$$u_{\mathrm{s}} = \frac{47}{80}, \qquad z_{\mathrm{s}} = \frac{80}{127}, \qquad \gamma_{\mathrm{s}}^2 = \frac{33}{40}, \qquad q_{\mathrm{s}} = \frac{66}{127}. \tag{105}$$

$$(a_0, a_1, a_2, a_3, a_4, a_5) = \left(\frac{80}{127}, \frac{5201}{10000}, \frac{851}{2000}, \frac{179}{500}, \frac{6603}{25000}, c\right). \tag{106}$$

Here $a_0 = z_{\mathrm{s}}$, $z_{\mathrm{s}}(1 + u_{\mathrm{s}}) = 1$, and $q_{\mathrm{s}} = 4z_{\mathrm{s}} - 2 = \gamma_{\mathrm{s}}^2 z_{\mathrm{s}}$, as required by Theorem 5.2. The initial padding pays for cooling and leaves the target mass needed by the final second moment bound. Cooling needs raw sampling rate

$$m_{\mathrm{raw}} = L_0 + J(z_{\mathrm{s}}). \tag{107}$$

The resulting parameters are

| Height | Energy | List rate |
|---|---|---|
| 0 | 80/127 | 0.306032 . . . |
| 1 | 5201/10000 | 0.307101 . . . |
| 2 | 851/2000 | 0.304330 . . . |
| 3 | 179/500 | 0.313711 . . . |
| 4 | 6603/25000 | 0.260970 . . . |

In particular, $L_0 - P(u_{\mathrm{s}}) > 0.000643$ and $m_{\mathrm{raw}} < 0.372456$, so the cooling list exceeds its critical size.

### 10.2 Constructing the two final lists

The length estimate (13) gives a polynomial-factor estimate of $\lambda_1(L)$. The multiplicative scale grid in Appendix B.9.1 has $O(n \log n)$ branches, including one with $Q_i \le \lambda_1(L)^2 < (1 + 1/n)Q_i$. Thus one branch places every shortest vector at the target energy $c + O(1/n)$.

On each branch, choose the scale and prime by the bounded initialization in Appendix B.9.1. A complete trial samples a fresh uniform one dimensional subspace $U = \Gamma/L$ in $K/L \simeq \mathbb{F}_p^n$, where $K = p^{-1}L$. Put $r = \sqrt{2}s$ and $V = r^n/\det\Gamma$. On the correct branch, the good smoothing event $\mathcal{G}$ has probability $1 - 2^{-\Omega(h_n)}$, and

$$\log_2 V = n/2 + v_n, \qquad v_n = h_n + O(1), \qquad \frac{2\pi\|v\|^2}{nr^2} = c_n = c + O(1/n) \tag{108}$$

for every fixed shortest vector $v$. The index is always $p$, so $V$ is fixed before the line is drawn.

Draw the common coset as in Equation (86), independently of the line. The two terminal lists have support

$$C = z + \delta + \Gamma. \tag{109}$$

The half grid property established there supplies the ancestor averages. Build two trees of depth four with root offset $z + \delta$, using BuildFullTree from Algorithm 3. At their thirty-two leaves, use independent cooling forests with the parameters above, terminal scale $r$, and raw rate $m_{\mathrm{raw}}$.

The wide starting scale is polynomially larger than $r$. A fixed logarithmic cooling depth, with raw rate $m_{\mathrm{raw}}$, therefore supplies each initial pool. The schedule is given in Appendix B.9.1. The reference from Theorem 5.2 is

$$\nu_{0,b}(x) = \frac{\Upsilon_0}{V}\rho_r(x)\mathbf{1}_{x \in (b+\Gamma)\cap\mathcal{S}_r(a_0)}, \qquad \log_2 \Upsilon_0 = m_{\mathrm{raw}} n + \sigma_0, \tag{110}$$

and factorial error $\varepsilon_{\mathrm{cool}} \le \exp[-\Omega(nw_n^2)]$, where

$$\sigma_0 = C_{\mathrm{leaf}} b_n - O((h_n + \ell_n)\log n) = \Theta(b_n).$$

The amplitudes are fixed before the affine environment is chosen. Appendix B.9.1 verifies these scale and list size choices.

### 10.3 The target representation count

At height $j = 1, \dots, 4$, retain every record pair whose difference lies in $\mathcal{S}_r(a_j)$. Apply (79) with $d = 4$ and $V_r = V$. Writing $k_j = L_j - W(a_j)$ gives

$$\log_2 A_j = k_j n + 2^j \sigma_0 - v_n, \qquad 0 \le j \le 4. \tag{111}$$

Substitution of (106) gives $\min_{1 \le j \le 4} E_j > 1.16$ and an abundance exponent greater than 0.13 for the four internal layers. Thus Theorem 8.2 applies with bridge margin 0.16 and abundance parameter 0.12.

Retain $\mathcal{H}_{\mathrm{term}} = \sigma(\Gamma, z, \delta)$ and average the private environments, samples, and coins of each tree. The two trees remain independent given this retained information. On the input shell, replace the Gaussian reference by its central intensity $A_0$. By (80), the resulting leaf error is

$$\varepsilon_0 = \varepsilon_{\mathrm{cool}} + n a_0 w_n = O(n w_n).$$

For target extraction we only need logarithmic distortion $o(n)$. Applying Theorem 8.2 now gives

$$\varepsilon_4 \le 16\varepsilon_0 + 15\eta_n + 30 \cdot 2^{-0.12n} = O(\ell_n) + e^{-\Omega(n w_n^2)}.$$

Choose the uniform upper bound $D_n = C_D(h_n + \ell_n)$ with $C_D$ sufficiently large. The terminal moments then satisfy

$$e^{-D_n} A_4 \le m_{\pm}(x) \le e^{D_n} A_4, \qquad S_{\pm}(x, x') \le e^{D_n} A_4^2, \qquad D_n \le C_D(h_n + \ell_n), \tag{112}$$

on $C \cap \mathcal{S}_r(a_4)$ for every good $\Gamma$. The second factorial measure counts distinct IDs, including equal values.

The terminal bridge has energy

$$E_5 = 4a_4 - c = 0.735778066\ldots > 2/e + 10^{-5}, \tag{113}$$

so Theorem 9.1 applies to $J_C(v) = \#\{x \in C : x, x - v \in \mathcal{S}_r(a_4)\}$. For every fixed $\delta$, its mean satisfies

$$\mu_\delta = V(eE_5/4)^{n/2} e^{\pm O(\ell_n)}.$$

The actual energy $c_n = c + O(1/n)$ changes its logarithm by $O(1)$. The terminal list size balance now gives

$$A_4^2 \mu_\delta \ge 2^{32\sigma_0 - v_n - O(\ell_n)} = 2^{\Omega(b_n)}. \tag{114}$$

Apply Lemma 9.2 with $A = A_4$ and $D = D_n$. The good superlattice probability tends to one, and the estimate is uniform in $\delta$. Averaging $\delta$ gives

$$\mathbb{P}[\mathcal{G} \text{ and some terminal pair has difference } v] \ge 2^{-B_n}, \qquad B_n = C_B(h_n + \ell_n) = o(\chi_n). \tag{115}$$

This proves target existence. Deduplicating terminal values preserves it.

### 10.4 The cost of finding the target

The moments also control the search work. Averaging the extra half grid shift and the ancestor shifts gives

$$\mathbb{E}[|\Xi_j| \mid \Gamma \text{ good}, z] \le 2^{L_j n + O(b_n)}, \qquad 1 \le j \le 4. \tag{116}$$

The cooling lists satisfy the corresponding bound at rate $L_0$. Appendix B.9.2 derives these bounds and the candidate counts by first fixing the common environment and then multiplying the conditional first intensities of the independent children.

Cooling uses $\tau_{\mathrm{s}} = u_{\mathrm{s}}$ and $K_{\mathrm{s}} = K_{u_{\mathrm{s}}}(u_{\mathrm{s}})$. At heights $j = 1, \dots, 5$, use

$$P(\tau_j) = L_{j-1}, \qquad K_j = K_{\tau_j}(u_j). \tag{117}$$

The cooling pair work has rate

$$T_{\mathrm{s}} = K_{\mathrm{s}} + 2(L_0 - P(u_{\mathrm{s}})) = 0.412924647\ldots. \tag{118}$$

For each internal full layer, the ordinary correlation term in (61) gives rate $K_j$ and dominates the contribution of nearly parallel pairs. The strict smoothing and wedge margins needed for this comparison are checked in Appendix B.9.2.

At the terminal call, write $x = z + \delta + g_x$ and assign $\mathsf{lab}(x) = g_x + L$. Then

$$\mathsf{lab}(x) = \mathsf{lab}(y) \quad \Longleftrightarrow \quad x - y \in L. \tag{119}$$

The affine search bound Proposition 9.4, applied to (112), gives rate $K_5$ after deduplication. Its expectation uses the same unconditioned coset average as the target bound. The chosen parameters give

$$\max\{0.2, m_{\mathrm{raw}}, T_{\mathrm{s}}, K_1, \dots, K_4\} = K_3 < 0.412960 < K_5 < 0.412965 < 0.413.$$

Thus the terminal search determines time and is already close to the minimum of its balance curve (Proposition 9.3). The earlier energies mainly determine peak storage.

The largest retained list determines space. The input rate of the last full layer reflects its angular pair loss:

$$u_4 = 0.631117\ldots, \qquad P(u_4) = 0.366452\ldots, \qquad L_3 = \frac{L_4 + P(u_4)}{2}.$$

This layer needs a larger input list to supply the smaller terminal list after its angular restriction. Changing $a_3$ changes this loss and also the preceding correlation $u_3$, so reducing storage must respect the earlier search budget. For the chosen energies, set

$$L_{\max} = L_3 < 0.313712, \qquad s_{\mathrm{live}} = 0.31375, \qquad S = 2^{s_{\mathrm{live}} n + C_S b_n}, \tag{120}$$

with $C_S$ sufficiently large. Every input rate is at most $L_{\max} < s_{\mathrm{live}}$, so the restricted traversal of Proposition 6.1 fits this storage with the same leading time rate. Raw samples are streamed, and depth first evaluation keeps only the current lists and traversal state. These estimates supply the work and storage inputs to Proposition 6.3.

Table 5 collects the costs and the terminal geometric margin for this parameter choice.

| Quantity | Value or bound | Role |
|---|---|---|
| Cooling work $T_{\mathrm{s}}$ | $0.4129246\ldots$ | Initial list generation |
| Internal work $K_3$ | $0.4129598\ldots$ | Largest nonterminal cost |
| Terminal work $K_5$ | $0.412964 < K_5 < 0.412965$ | Time bottleneck |
| List rate $L_3$ | $< 0.313712$ | Space bottleneck |
| Terminal margin | $19 \cdot 10^{-6} < E_5 - 2/e < 20 \cdot 10^{-6}$ | Strict affine lens margin |

Table 5: The main costs and terminal margin of the SVP construction.

## 10.5 The algorithm and success amplification

Algorithm 4 assembles the construction. The scale grid has a branch at the shortest length, and independent trials on that branch turn the target probability $2^{-B_n}$ into constant success. A final polynomial number of complete executions gives the stated high probability. Keeping the shortest valid vector makes it unnecessary to identify which branch or trial succeeded.

The recursive calls in the algorithm use the finite implementation of Proposition 6.3. Initialization checks, resource limits, and exceptional execution rules are specified in Appendix C.3.5.

**Algorithm 4:** Shortest vectors: SVP

**Input:** A rational basis $B$ of $L$ and the fixed parameters of Section 10.1.
**Output:** A nonzero vector of $L$, shortest with the probability in Theorem 1.3.

$v_{\mathrm{best}} \leftarrow$ a nonzero column of $B$
Construct the polynomial scale grid $(Q_i)$ of Equation (160)
**for** $e = 1, \dots, \lceil n \log_2(n+2) \rceil$ **do**
  **for** each scale $Q_i$ **do**
    Initialize its squared width and prime $p$ as in Appendix B.9.1, skipping the scale on failure
    **for** $t = 1, \dots, \lceil 2^{\chi_n} \rceil$ **do**
      Draw a fresh uniform line defining $\Gamma$ as in Proposition 3.6
      Independently draw $z \sim \mathrm{Unif}((p^{-1}L)/L)$ and $\delta \sim \mathrm{Unif}((L/2)/L)$
      Preprocess $\Gamma$ for cooling
      $b \leftarrow z + \delta$
      $\Xi_+ \leftarrow \textsc{BuildFullTree}(b, 4)$
      $\Xi_- \leftarrow \textsc{BuildFullTree}(b, 4)$ with independent private randomness
      Merge equal values within each terminal list
      Label $x = b + g_x$ by $g_x + L$ and draw fresh terminal search codes
      **for** each equally labeled pair $(x, y)$ reported by the terminal search **do**
        **if** $x \neq y$ and $\|x - y\| < \|v_{\mathrm{best}}\|$ **then**
          $v_{\mathrm{best}} \leftarrow x - y$
**return** $v_{\mathrm{best}}$

On the correct branch, (115) and Proposition 6.3 give success at least $2^{-B_n - 1}$ per trial, conditional on a valid public prime. Since $B_n = o(\chi_n)$, the scheduled $2^{\chi_n}$ trials give constant success even after the public prime and sampling errors. Taking the shortest output of $\lceil n \log_2(n+2) \rceil$ independent complete executions reduces failure to $2^{-\Omega(n \log n)}$, with only $2^{o(n)}$ extra time and the same workspace. Appendix C.3.5 accounts for the implementation error over this amplified execution, proving Theorem 1.3.

## 10.6 Other choices of the four layer energies

The terminal energy fixes the final search cost, while the earlier energies determine the list sizes needed to reach it. Varying these energies therefore changes peak storage and may also make an internal search the dominant cost. We describe this tradeoff by the largest list rate and the largest work rate, then state the conditions under which the preceding analysis applies.

Fix rational $u_{\mathrm{s}} \in (1/2, 31/50]$ and positive rational $a_0, \dots, a_4$ with $a_0 = 1/(1+u_{\mathrm{s}})$. Keep $\kappa_{\mathrm{pack}} = 401/1000$, $a_5 = c$, and define $u_j, E_j, L_j, \tau_j, K_j$ by (103), (104), and (117). Put

$$\begin{aligned} \Delta_{\mathrm{s}} &= L_0 - P(u_{\mathrm{s}}), \\ s(\mathbf{a}) &= \max_{0 \le j \le 4} L_j, \\ t(\mathbf{a}) &= \max\{1/5, L_0 + J(a_0), K_{u_{\mathrm{s}}}(u_{\mathrm{s}}) + 2\Delta_{\mathrm{s}}, K_1, \dots, K_5\}. \end{aligned}$$

Here $\Delta_{\mathrm{s}}$ is the extra list exponent available for cooling, $s(\mathbf{a})$ sets peak storage, and $t(\mathbf{a})$ counts preprocessing, raw sampling, cooling search, and the five later searches. Choose a rational storage rate $s_* > s(\mathbf{a})$.

The construction works when the following conditions hold with positive margins independent of $n$.

(i) Cooling has enough records: $\Delta_{\mathrm{s}} > 0$.

(ii) Internal layers preserve the moment bounds. For $1 \le j \le 4$,

$$E_j > 1, \qquad \mathfrak{a}_j := L_{j-1} - W(a_{j-1}) + \frac{1}{2} + \frac{1}{2}\log_2(eE_j/4) > 0.$$

The first inequality controls the shared midpoint shift, and the second bounds the relative contribution from reusing an input.

(iii) Pair search satisfies the cap and smoothing estimates. For $1 \le j \le 5$, require $0 < u_j < 1$, $0 < \tau_j^2 < 1$, and $1 - 2\tau_j^2/(1+u_j) > 0$. For the remaining comparisons put $t_j = 2^{5-j}$, $c_{0,j} = 4t_j^2 a_{j-1}$ for $1 \le j \le 4$, $c_{0,\mathrm{s}} = 2/(1-u_{\mathrm{s}}^2)$, and $\Delta = 1/2000$. The smoothing and nearly parallel pair bounds require, for $1 \le j \le 4$,

$$\begin{gathered} 2t_j^2 a_j > 1, \qquad 8a_{j-1}\tau_j^2 > 1, \qquad c_{0,j} > 4, \\ L_{j-1} - \tfrac{1}{2}\log_2(c_{0,j}/2) < 0, \\ c_{0,j}\Delta < 1, \qquad 4a_{j-1}(2-\Delta) > 1. \end{gathered}$$

For cooling the corresponding checks are $4u_{\mathrm{s}}^2/(1-u_{\mathrm{s}}^2) > 1$, $c_{0,\mathrm{s}}\Delta < 1$, and $c_{0,\mathrm{s}}(2-\Delta) > 1$.

(iv) The terminal affine count and search apply: $E_5 > 2/e$. With $\zeta = \frac{1}{2}\log_2(eE_5/2)$, $\omega = L_4 - W(a_4)$, and $Q = 4a_4(1-\tau_5^2)$, require $\zeta > 0$, $\omega + \zeta > 0$, and $Q/(2c) > 1$.

(v) Storage permits restricted traversal: $1/5 < s(\mathbf{a}) < \min\{s_*, 1\}$. Use the depth and list size schedule of Appendix B.9.1 with this storage rate.

**Proposition 10.1.** *Every parameter choice satisfying conditions (i)–(v) above gives an SVP algorithm with the success probability of Theorem 1.3 and, on every execution,*

$$T \le 2^{t(\mathbf{a})n + o(n)}, \qquad S \le 2^{s_* n + o(n)}.$$

*Proof.* Conditions (i) and (ii) give valid cooling coefficients through Theorem 5.2 and propagate the moment bounds through four full layers by Theorem 8.2. The backward recurrence sets the leading target count exponent to zero, and padding makes that count $2^{\Omega(b_n)}$. Condition (iv) permits the affine count and work estimates, so Lemma 9.2 gives target probability $e^{-o(n)}$ per trial.

The ancestor denominators remain $16, 8, 4, 2$. Condition (iii) gives the list smoothing, active cap, nearly parallel pair, and terminal cap comparisons used in Appendix B.9.2. The bound $L_j < 1$

also makes the contribution with common cap factor at most $1/16$ smaller than the filter rate. The cooling point rate $K_{u_{\mathrm{s}}}(u_{\mathrm{s}}) + \Delta_{\mathrm{s}}$ lies between $L_0$ and the cooling pair rate, while $K_j > L_{j-1}$ at every later search. Thus (v) fits restricted traversal into the stated work and storage. The amplification in Section 10.5 and Proposition 6.3 give the success and resource bounds. □

The main instance above is scheme A. A second feasible instance, scheme B, keeps $u_{\mathrm{s}}, a_0, a_4, c$ unchanged and takes

$$(a_1, a_2, a_3) = \left( \frac{6501}{12500}, \frac{42549}{100000}, \frac{35777}{100000} \right).$$

Its largest list rate is $L_3 < 0.313533$, so $s_* = 0.31355$ gives a displayed space bound $2^{0.3136n+o(n)}$. Its largest work rate is the internal rate $K_2 < 0.412980 < 0.4130$. Thus it offers less space at a slightly higher time exponent than A before rounding. Both have the same terminal geometry, and direct substitution verifies the admissibility conditions for both. Schemes A and B give two operating points in the four layer parameter region: increasing the overall work rate by about $1.6 \cdot 10^{-5}$ lowers the displayed space exponent from 0.3138 to 0.3136.

# 11 Independent Gaussian batches

The single sample construction reads once from a correlated list. To obtain an exponential number of outputs within the same total cost, we instead reuse a fixed partner list, called a bank, at each merge. Two rejections make every trial's complete output law independent of that list, allowing conditional product laws to pass through the recursion.

We retain the Gaussian shell geometry, smoothing extensions, and endpoint schedule of Section 7, using the default budgets (163). The new ingredients are uniform control of partner and representation weights, a correction against each fixed bank, and radial coverage sufficient for the entire batch.

## 11.1 Two rejections against a fixed bank

Selecting a partner from a bank introduces a normalization that depends on the left record. Even after this dependence is removed, a difference with more representations is proposed more often. We correct these two effects in order. Accepting the left record with probability $A_C(x)/D$ cancels the denominator used to select its partner. A second rejection divides by the total proposal weight of the difference and inserts its desired target weight.

The correction has two consequences. First, the complete trial law, including failure, is the same for every good bank, so independent trials remain independent after averaging over banks. Second, a matching upper bound on representation weights bounds the proposal law before the last rejection by a constant times the output law at each fixed bank. Applying this comparison before averaging the bank-dependent query cost lets the bank's original first moment control the reverse search. The lemma records both properties.

**Lemma 11.1** (Correction against a fixed bank)**.** *Fix public information $H$. Conditional on $H$, let the left trials be iid on $\mathcal{X} \cup \{\dagger\}$, with their entire stream independent of a finite bank $C = (y_i)$. For each vector $x \in \mathcal{X}$, write $q(x) = \mathbb{P}[X = x \mid H]$. Thus $\sum_x q(x) \le 1$, with the remaining probability assigned to failure $\dagger$. For $K \ge 0$, define*

$$A_C(x) = \sum_i K(x, y_i), \qquad B_C(z) = \sum_i q(z + y_i) K(z + y_i, y_i).$$

*Let $D, c > 0$ and a summable target weight $h \geq 0$ depend only on $H$. Call a bank good if $A_C(x) \leq D$ for every $x$ with $q(x) > 0$, and $B_C(z) \geq ch(z)$ for every $z$ with $h(z) > 0$. Apply Algorithm 5 to each left trial, using independent private randomness for its choices. On a good bank the output trials are iid, with probability*

$$q_o(z) = ch(z)/D \tag{121}$$

*of returning each vector $z$ and the remaining probability at †. Both parts of this distribution are independent of $C$. If also $B_C(z) \leq C_0 ch(z)$ for every $z$, the proposal law before the second rejection is at most $C_0 q_o(z)$.*

*Proof.* Fix a good bank. For $h(z) > 0$, summing over representations gives

$$\mathbb{P}[Z = z \mid H, C] = \sum_i q(z + y_i) \frac{K(z + y_i, y_i)}{D} \frac{ch(z)}{B_C(z)} = \frac{ch(z)}{D}.$$

Outputs with $h(z) = 0$ are rejected, leaving failure mass $1 - (c/D) \sum_z h(z)$. Thus the complete trial law, including failure, is the same for every good bank. Conditional independence of the left trials and their private choices gives a product of this common law, which remains a product after averaging the bank. Before the second rejection the mass is $B_C(z)/D$, giving the proposal bound. □

**Algorithm 5:** Two rejections: FixedBankTrial$(X, C)$

**Input:** A left trial $X \in \mathrm{supp}(q) \cup \{\dagger\}$, fixed bank $C = (y_i)$, and $q, K, D, c, h$ from Lemma 11.1.
**Output:** A difference $z$ or †. The bank is retained for later trials.

```
if X = † then
    return †
x ← X, A ← Σ_i K(x, y_i)
if a fresh Bernoulli trial of probability min{1, A/D} fails then
    return †
Choose a bank record i with probabilities K(x, y_i)/A
z ← x − y_i
B ← Σ_k q(z + y_k) K(z + y_k, y_k)
if a fresh Bernoulli trial of probability min{1, ch(z)/B} fails then
    return †
return z
```

The bank is retained across trials, so the complete law in (121) applies to every reuse of it. Clipping defines the procedure on arbitrary banks. The implementation of the two sums, including the positive selected contribution to the reverse sum, is given in Appendix C.3.3.

For a recursive construction, define a reference tree that replaces the whole output of each bad bank by fresh iid trials of (121). The common complete law proves the product property by induction. Coupling the actual and reference trees until their first bad bank transfers it to the algorithm. Appendix B.6 applies the fixed bank proposal bound to the reverse search cost.

## 11.2 Exponential batches with a fixed bank

We apply Lemma 11.1 recursively with the Gaussian kernel. The construction has three parts: produce independent trials on thin shells, combine radial bins using their known coefficients, and add one terminal merge for the inner bins needed at width 3/2. We use the fixed bank strategy

of [DL26]. On an arbitrary lattice, the main task is to control both the partner weight of every possible left record and the representation weight of every possible output. Gaussian identities estimate their means on a smoothing superlattice, and uniform concentration bounds make these estimates hold simultaneously for the realized bank. The two rejections then remove the bank from each trial's complete output law. Each radial bin uses its own smoothing extension so that its index cost reflects the lengths it must supply.

Use the endpoint parameters $u_n, \gamma, a_0, a_1$ of Lemma 7.2. Put

$$b_{\deg} = \lceil 10\ell_n \rceil, \quad M = \left\lceil 2^{m(u)n + C_a b_n} \right\rceil, \quad R(u) = \tfrac{1}{2}\log_2(2a_0).$$

A stream consists of $M$ trials, each returning a vector or †. Including failed trials keeps the Gaussian coefficient explicit throughout the recursion and radial mixing. Conditioning on success would instead introduce the unknown Gaussian mass of the shell. Only successful records occupy vector storage, as described in Appendix C.3.3.

### 11.2.1 A smoothing extension for each radial bin

A single shell of relative width $w_n$ discards Gaussian mass $\exp[-\Omega(nw_n^2)]$. This error is small for one output, but truncating an independent $B$-product can cost up to $B$ times that mass. We therefore cover an annulus of constant radial width whose large deviation tail is exponentially smaller than $B^{-1}$. Product code search still uses thin shells, so we partition the annulus into polynomially many bins, build a separate tree on each bin, and balance the root streams. The choice below makes the endpoint tail rates $J(z_-), J(z_+)$ exceed the batch rate.

Choose fixed rational $\alpha^{-2} < z_- < 1 < z_+$. Partition $[z_-, z_+]$ into disjoint half-open bins $I_j$ of width at most $cz_-w_n$, including the final endpoint in the last bin. Let $z_j$ be their rational centers and set

$$A_j = \{x : 2\pi\|x\|^2/(n\sigma^2) \in I_j\}, \quad r_j^2 = \sigma^2 z_j/a_0, \quad s_j^2 = \sigma^2 z_j/(2a_0).$$

The number $Q$ of bins is polynomial and $A_j \subseteq \mathcal{S}_{r_j}(a_0)$ for small fixed $c$. For each bin choose an independent $\Gamma_j \supset L$ using Lemma 7.1 at scale $s_j$, with

$$\iota_j = \max\{0, \tfrac{1}{2}\log_2(2a_0/(\alpha^2 z_j))\}, \qquad \log_2 p_j = \iota_j n + 3h_n + O(1).$$

Except with probability $2^{-2h_n+O(1)}$ its dual mass at $s_j$ is at most $2^{-h_n}$. A union bound pays for the polynomially many public choices. Basis preprocessing has rate 0.2 and gives $G_j \le \mathrm{poly}(n)s_j$. Within each tree suppress the bin index on its lattice and write $V_r = r^n/\det\Gamma_j$.

Use one common depth $t$ for all bins so that each leaf width $R_j = r_j\gamma^{-t}$ satisfies $(R_j/G_j)^2 \ge \log(8n)/\pi$. Then $t = O(\log n \log\log n)$ and $t(b_{\deg}+2) = o(b_n)$. Generate each of the $M$ leaf labels by Lemma 3.3. Its probability mass on vector outputs is $\lambda_t\rho_{R_j}(x)/V_{R_j}$ on the prescribed shell and coset, with $\lambda_t = 1/2$.

### 11.2.2 Two rejections and independent output trials

At a node of input width $r$, both child streams assign probability

$$q(x) = \frac{\lambda}{V_r}\rho_r(x)\mathbf{1}_{x\in\text{prescribed coset}\cap\mathcal{S}_r(a_0)}$$

to each vector $x$, with the remaining mass at †. Route the half grid shift as in Section 5.1. Keep the right child vectors as a fixed bank $C = (y_i)$. Selecting a partner leaves every bank record available

for later trials. Use $K(x,y) = \mathbf{1}_{\text{Eligible}}\omega_r(x-y)$ on the source shells, extended by zero elsewhere, in Lemma 11.1, with partner and representation weights $A_C(x)$ and $B_C(z)$ defined there. The deterministic normalizers are

$$\mathcal{B}_r = e^{A_r R_-^2}(\gamma/\sqrt{2})^n, \quad A_{\max} = 2M\lambda\mathcal{B}_r 2^{b_{\deg}}, \quad b_* = \frac{M\lambda^2}{V_r^2} e^{A_r R_-^2} V_{r/\sqrt{2}}.$$

Lemma B.1 proves, with failure $\exp[-2^{\Omega(b_n)}]$ uniformly in the input width and determinant, the simultaneous inequalities

$$A_C(x) \le A_{\max}, \qquad \tfrac{1}{2} b_* \rho_{\gamma r}(z) \le B_C(z) \le 2 b_* \rho_{\gamma r}(z) \tag{122}$$

for every admissible source or output query.

Apply Lemma 11.1 with $D = A_{\max}$, $c = b_*/2$, and $h(z) = \rho_{\gamma r}(z)$ on the prescribed output shell and coset, and zero elsewhere. The bank bounds give $C_0 = 4$. The output trial therefore assigns mass

$$q_{\text{out}}(z) = \frac{b_*/2}{A_{\max}} \rho_{\gamma r}(z) = \frac{\lambda 2^{-(b_{\deg}+2)}}{V_{\gamma r}} \rho_{\gamma r}(z) \mathbf{1}_{z \in \text{output coset} \cap \mathcal{S}_{\gamma r}(a_0)} \tag{123}$$

to each vector $z$, with the remaining probability at †. The actual algorithm clips probabilities on bad branches. Use the reference replacement of Section 11.1 for the analysis. Independent child streams satisfy the lemma's hypotheses at each node, so induction gives iid root labels and

$$\lambda_{j-1} = \lambda_j 2^{-(b_{\deg}+2)}, \qquad \lambda_0 = 2^{-1-t(b_{\deg}+2)} = 2^{-o(b_n)}.$$

Use disjoint private randomness for different bins. Conditional on the public environment their root streams are independent. The coupling charges the first bad bank in the whole forest.

Thus the recursive procedure uses the same affine routing as cooling, with the two rejections replacing activation and isolation.

### 11.2.3 Balancing radial bins and extracting a product law

Restrict root $j$ to $A_j \cap (a+L)$. Correct its width to $\sigma$ by radial rejection, anchored at the bin's least squared norm if $r_j \ge \sigma$ and at its greatest squared norm otherwise. Let $\zeta_j$ be that bin endpoint in normalized target energy. It is rational and $\zeta_j = z_j + O(w_n)$. The surviving probability mass on vector outputs is

$$\widetilde{q}_j(x) = \kappa_j \rho_\sigma(x) \mathbf{1}_{A_j \cap (a+L)}, \qquad \kappa_j = \frac{\lambda_0}{V_{r_j}} e^{n\zeta_j(1-a_0/z_j)/2}.$$

Each coefficient uses its own determinant and is explicitly known. Choose bin $j$ with probability proportional to $\kappa_j^{-1}$, and put

$$\kappa_{\text{mix}} = \left( \sum_{j=1}^{Q} \kappa_j^{-1} \right)^{-1}.$$

The bin probability $\kappa_{\text{mix}}/\kappa_j$ cancels its coefficient, giving the same Gaussian weight throughout the annulus.

In each of $M$ mixture trials, independently choose a bin, consume its next unused label, and apply the radial correction. Each bin supplies $M$ labels, and the conditional law of its unused stream gives iid mixture trials with

$$q_{\text{mix}}(x) = \kappa_{\text{mix}} \rho_\sigma(x) \mathbf{1}_{\{x \in a+L:\, \mathcal{E}_\sigma(x) \in [z_-, z_+]\}}.$$

Return the first $B$ successful values, or the fallback tuple if fewer occur. For iid trials the success indicators are independent of the successful values, so conditioning on enough successful trials leaves exactly the truncated Gaussian product law. This law is independent of the public extensions and therefore survives their average.

The identity

$$\kappa_j \frac{\sigma^n}{\det L} = \frac{\lambda_0}{p_j} 2^{-J(a_0)n + J(z_j)n + O(nw_n)}$$

becomes, after multiplying by the number $M$ of trial labels,

$$M\kappa_j V_\sigma(L) = 2^{[P(u) - \iota_j + J(z_j)]n + \Theta(b_n)}.$$

This expresses the expected successful count after normalization by the target Gaussian volume. Its limiting exponent is

$$F_E(z) = \min\left\{\beta_{\mathrm{G}} + J(z),\ \log_2 \alpha + \frac{z-1}{2\log 2}\right\}. \tag{124}$$

Since $\kappa_{\mathrm{mix}} \geq Q^{-1} \min_j \kappa_j$ and $Q$ is polynomial, the smallest count exponent among bins also bounds the mixture count. The radial rarity and the expansion loss are therefore charged within the same bin. The following theorem records the resulting range.

**Theorem 11.2** (Batches with independently chosen extensions)**.** *Fix $\alpha > 1$ and*

$$0 < b < \min\{\beta_{\mathrm{G}}, J(\alpha^{-2})\}. \tag{125}$$

*On the inputs of Theorem 1.2, there is an algorithm producing $\lfloor 2^{bn} \rfloor$ points with joint statistical distance at most $\exp[-\Omega(n/\log^2(n+2))]$ from the target product distribution, in time $2^{\alpha_{\mathrm{G}} n + o(n)}$ and bit space $2^{\beta_{\mathrm{G}} n + o(n)}$.*

*Proof.* Choose $\alpha^{-2} < z_- < 1 < z_+$ with $J(z_-), J(z_+) > b$. The minimum of (124) on this annulus exceeds $b$, because

$$\log_2 \alpha + \frac{z_- - 1}{2\log 2} - J(z_-) = \log_2(\alpha\sqrt{z_-}) > 0.$$

The expected mixture success count exceeds $2^{bn}$ by a fixed exponential factor, so a Chernoff bound makes a shortfall doubly exponentially unlikely. The tail estimate below and the search and realization bounds at the end of this subsection complete the proof. □

The radial exponential moment of Appendix A.5, optimized at the two bin endpoints, gives

$$\delta_{\mathrm{tail}} \leq 3\,2^{-J(z_-)n} + 3\,2^{-J(z_+)n}.$$

The condition $z_- > \alpha^{-2}$ keeps the smaller tilted width $\sigma\sqrt{z_-}$ above $\eta_{1/2}(L)$. Truncating a $B$-product costs at most $B\delta_{\mathrm{tail}}$.

### 11.2.4 A terminal merge for full batches at width 3/2

At $\alpha = 3/2$, the general theorem allows rates only below $J(4/9) \approx 0.1842$, whereas a full batch has rate $\beta_{\mathrm{G}} \approx 0.2075$. Reaching that rate requires more radial coverage. We supply the extra inner bins with one additional merge and estimate their lower tail using a tilt that stays at smoothing. For $z \in [5/12, 1/2]$ the merge uses

$$u_T = 3/5, \qquad c_T = 5/8, \qquad d_T = 1/2, \qquad \gamma_T^2 = 4/5.$$

Here $c_T$ and $d_T$ are the input and difference energies at the same reference width. Since $d_T/\gamma_T^2 = c_T$, rescaling the output returns it to the input energy. For each inner bin, first generate two independent streams using the ordinary batch construction above. Choose their squared width $R^2 = (c_T/a_0)r^2$, so that their stored vectors already lie in the shell of energy $c_T$ at width $r$. A radial rejection changes their Gaussian profile while preserving this support. We then apply one additional fixed bank merge with correlation $u_T$. The two input lists still have size exponent $\beta_{\mathrm{G}}$. Equations (149) and (150) give the width and coefficient calculations.

The forward residual Gaussian can lie below smoothing, introducing the factor $H_n$ in the theta bound (151). We absorb this factor into the degree normalizer to keep rejection valid. The chosen parameters cancel its leading exponent against the degree estimate, leaving a $2^{o(n)}$ normalizer and allowing the same two rejections. The output count has rate $q_T = 2\beta_{\mathrm{G}} - P(3/5) = \log_2(16/15)$. The inner bins have exponentially small target mass, so this smaller count suffices for their share of a full batch. After restricting to $a + L$ and correcting the width, the normalized output count exponent is

$$F_T(z) = q_T - \max\{0, -\tfrac{1}{2}\log_2(135z/64)\} + J(z). \tag{126}$$

Its minimum on $[5/12, 1/2]$ exceeds $\beta_{\mathrm{G}}$. The outer bins $[1/2, 2]$ use (124). Their limiting minimum is $\beta_{\mathrm{G}}$, and the finite endpoint parameters and positive padding provide the extra subexponential number of records needed to extract $\lfloor 2^{\beta_{\mathrm{G}} n} \rfloor$ samples.

**Theorem 11.3** (Full batches at width 3/2)**.** *If $\sigma \geq 3\eta_{1/2}(L)/2$, the algorithm outputs $\lfloor 2^{\beta_{\mathrm{G}} n} \rfloor$ points jointly within $\exp[-\Omega(n/\log^2(n+2))]$ of $D_{a+L,\sigma}^{\otimes \lfloor 2^{\beta_{\mathrm{G}} n} \rfloor}$, in time $2^{\alpha_{\mathrm{G}} n + o(n)}$ and bit space $2^{\beta_{\mathrm{G}} n + o(n)}$.*

*Proof.* Appendix B.7 proves the terminal law, output count, and search bounds. The ordinary tilt to $\sigma\sqrt{5/12}$ may be below smoothing. Tilting only to $\sigma/\alpha$ instead gives the lower-tail rate

$$H_{3/2}(5/12) = \frac{\log(9/4) - 25/48}{2\log 2} > \beta_{\mathrm{G}}, \qquad J(2) > \beta_{\mathrm{G}}.$$

The bounds $\log(9/5) > 7/12$ and $\log(27/16) > 25/48$ verify these inequalities and $\min F_T > \beta_{\mathrm{G}}$. Consequently the full product truncation costs $2^{-\Omega(n)}$. Mixing the known coefficients gives the same normalized target on every good public environment. The remaining joint errors are those below and in Equation (155). □

The general region in Theorem 11.2 remains useful: smaller batches allow a smaller width and yield the GapSVP application. Theorem 11.3 is a separate strengthening at full size. Algorithm 6 collects the two cases into one procedure. In the full batch case we use the width promise with $\alpha = 3/2$, even if a stronger promise is available, and split bins at $z = 1/2$. Resource limits, stream representation, and exceptional execution rules are given in Appendix C.3.3 under Proposition 6.3.

Every mixture trial consumes a new label, including when that label fails. Each bin has $M$ labels, so none can be exhausted by the $M$ mixture trials. The reference sequence continues to $M$ trials for the analysis, while execution stops when the output list fills. The coefficient calculation uses determinants and Gaussian weights, without computing the Gaussian mass of any bin.

### 11.2.5 Search, space, and the joint error

The sums used by FixedBankTrial require two searches. Forward queries enumerate all eligible pairs of left and bank records, giving the partner weights $A_C(x)$. Reverse search uses fresh codes on the complete ideal forward proposal list and computes $B_C(z)$, counting each bank record once. The second search must recover all representations of each proposed difference, because its rejection

**Algorithm 6:** Gaussian sampling by radial bins: GaussianBatch

**Input:** $L, a, \sigma^2$ and feasible batch parameters from Theorem 11.2 or Theorem 11.3.
**Output:** $B$ points in $a + L$, or failure.

Fix the bins $(I_j)_{j=1}^{Q}$, trial count $M$, and depths of this section
**for** each bin $j$ **do**
  Draw its independent extension $\Gamma_j$ and preprocess its basis
  **if** full batch mode and $I_j \subseteq [5/12, 1/2]$ **then**
    Generate two independent ordinary streams on the prescribed child cosets
    Correct their widths and apply the additional fixed bank merge of Appendix B.7
  **else**
    Generate its root stream from corrected leaves, using FixedBankTrial at each merge
  Compute $\kappa_j$, the known coefficient after restriction to $A_j \cap (a + L)$ and radial correction
Initialize an empty output list $\mathcal{O}$
**for** $m = 1, \dots, M$ **do**
  Choose bin $j$ with probabilities proportional to $\kappa_j^{-1}$ and consume its next unused trial $X$
  Restrict $X$ to $A_j \cap (a + L)$ and apply bin $j$'s radial rejection, preserving failures
  **if** $X \neq \dagger$ **then**
    Append $X$ to $\mathcal{O}$
    **if** $|\mathcal{O}| = B$ **then**
      **return** $\mathcal{O}$
**return** failure

probability divides by their total weight. To find those bank records, query the bank with $-z$. Indeed, writing $x = z + y_i$, the norm identity gives

$$2\langle -z, y_i \rangle = \|z\|^2 + \|y_i\|^2 - \|x\|^2.$$

Both endpoints have energy $a_0$, while $z$ has energy $a_1 = 2a_0(1-u)$. Their thin shells therefore force

$$\frac{\langle -z, y_i \rangle}{\|z\|\,\|y_i\|} = \frac{\|z\|}{2\|y_i\|} + O(w_n) = \sqrt{\frac{a_1}{4a_0}} + O(w_n) = \sqrt{\frac{1-u}{2}} + O(w_n).$$

Thus reverse search uses target correlation $v_{\mathrm{rev}} = \sqrt{(1-u)/2}$ and code rate

$$K_{\mathrm{rev}}(u) = -\tfrac{1}{2}\log_2\left(1 - \frac{2u^2}{1 + \sqrt{(1-u)/2}}\right).$$

Appendix B.6 proves the reverse work bound while retaining the bank dependence of these proposals. The forward and reverse dictionaries contribute storage rates $P(u)$ and $2P(u) - R(u)$, respectively. Streams store only vector records, while processing all $M$ trial labels contributes work at the raw label rate. Depth first evaluation and storing the polynomial number of root bins give

$$K(u_n), K_{\mathrm{rev}}(u_n) \longrightarrow K_*, \qquad P(u_n),\ 2P(u_n) - R(u_n) \longrightarrow \beta_{\mathrm{G}},$$

where $K_* = \frac{1}{2}\log_2(3/2)$. The exponential rate of the total number of trial labels tends to $1/2 - 1/(6\log 2) < K_*$. All forest sizes, code repetitions, and bin counts are $2^{o(n)}$. These list size and work bounds allow Proposition 6.3 to transfer the ideal batch law to a finite algorithm. Combining its whole batch allowance with the public extension failure, bad bank probabilities, and annulus tail gives

$$d_{\mathrm{TV}}(\mathcal{L}(\mathrm{batch}), D_{a+L,\sigma}^{\otimes B}) \le \epsilon_{\mathrm{global}} + B\delta_{\mathrm{tail}},$$

$$\epsilon_{\mathrm{global}} \le 2^{-\Omega(h_n)} + 2^{-\Omega(b_n)} + 2^{-\Omega(n)} + \exp[-2^{\Omega(b_n)}].$$

In this combiner, midpoint error verifies the good bank inequalities. The actual reverse count cancels it from (123). Summing bad bank probabilities over the forest accounts for this error jointly for the entire batch. The public extension failure is the leading term, yielding $\exp[-\Omega(n/\log^2(n+2))]$. These rates prove Theorem 11.2. Together with the terminal analysis they prove Theorem 11.3 and hence Theorem 1.2.

# 12 Reductions and algorithmic consequences

The sampling results lead to two kinds of applications. A Gaussian batch on the dual lattice builds advice that can answer many decoding queries. SVP instead supplies centered Gaussian samples at every width, making rare useful vectors accessible by repeated sampling. We develop each reduction together with its resulting algorithm.

## 12.1 Accuracy and composition

We use the accuracy convention of [SD16, Definition 2.10]. A law $\mu$ is $(\gamma, \varepsilon)$-close to $D$ if some law $\widetilde{\mu}$ satisfies

$$d_{\mathrm{TV}}(\mu, \widetilde{\mu}) \le \varepsilon, \qquad \gamma^{-1} D(x) \le \widetilde{\mu}(x) \le \gamma D(x) \quad \text{for every } x.$$

Its total variation error is $O(\gamma - 1) + \varepsilon$. The pointwise comparison is essential for our sampling reductions: it preserves useful events whose probability is exponentially smaller than this total variation error.

If $\mu$ is $(1+\delta, \varepsilon)$-close to $D$, then every event $E$ satisfies

$$\frac{D(E)}{1+\delta} - \varepsilon \le \mu(E) \le (1+\delta) D(E) + \varepsilon. \tag{127}$$

If $D(E) \ge p$, $\delta \le 1$, and $\varepsilon \le p/4$, this gives success probability at least $p/4$, which determines the number of sampling calls in each application.

For a joint batch within TV distance $\varepsilon_B$ of $D^{\otimes B}$,

$$\Pr[\exists i : X_i \in E] \ge 1 - (1 - D(E))^B - \varepsilon_B.$$

If advice is simultaneously correct on all promised targets except with probability $\varepsilon_A$, and the conditional search failure on every good advice is at most $\zeta$, total failure is at most $\varepsilon_A + \zeta$.

This advice event is charged once across all queries. For an event of probability $n^{-O(1)} 2^{-\kappa n}$, use $Q = \mathrm{poly}(n) 2^{\kappa n}$ queries and count time as $\mathrm{poly}(n, \mathcal{B})(P + QD)$, with one preprocessing cost $P$ per lattice and query cost $D$. Sequential calls reuse the workspace. Fixed call budgets and exact membership and length checks let us retain the best valid candidate over all scales.

## 12.2 Reusable decoding from Gaussian batches

The preprocessing of a periodic Gaussian decoder consists of samples on the dual lattice [DRS14]. Our batch interface supplies these samples at a total cost much larger than a single decoding query. The same advice can therefore serve an exponential number of attempts.

**Theorem 12.1** (Decoding from a feasible Gaussian batch)**.** *Let* $(\alpha, b)$ *satisfy either Equation (125) or* $(\alpha, b) = (3/2, \beta_{\mathrm{G}})$*, and put* $A_{\mathrm{KL}} = 2^{0.401}$*. For every fixed*

$$0 < r < r(\alpha, b) := \frac{\sqrt{e\, b \log 2}}{\alpha A_{\mathrm{KL}}}, \tag{128}$$

*there is an* $r$*-BDD data structure on every rational lattice, with*

$$P = 2^{\alpha_{\mathrm{G}} n + o(n)}, \quad D = 2^{bn + o(n)}, \quad S = 2^{\beta_{\mathrm{G}} n + o(n)}.$$

*With probability* $1 - o(1)$ *the same advice determines a function that returns the exact nearest lattice point for every rational target at distance at most* $r\lambda_1(L)$*. All resource bounds hold on every execution, with polynomial input and target encoding factors suppressed.*

*Proof outline.* Set $\varepsilon = 2^{-2bn}$. The KL counting estimate gives a dual sampling width just above $\alpha A_{\mathrm{KL}} \sqrt{n/(2\pi e)}/\lambda_1(L)$. The feasible parameters obey $\alpha > 2^{2b}$, so this width both satisfies the batch promise and makes $\rho_{1/s}(L \setminus \{0\}) \le \varepsilon$. The periodic Gaussian decoder needs $N = O(n \log(1/\varepsilon)/\sqrt{\varepsilon}) = \mathrm{poly}(n) 2^{bn}$ samples and decodes to any fixed fraction below $\sqrt{\log(1/\varepsilon)/\pi}/(2s)$. Substitution gives (128). Appendix F.1 proves the version using a known mass upper bound, the uniform empirical estimates, and the unknown length grid. Polynomially many independent batches contribute only $o(1)$ joint TV error to the advice. Its global good event is used once. □

### 12.3 Shortest vectors and gap tests from decoding

Assume an $r$-BDD interface with $0 < r < 1/2$, preprocessing $P$, query work $D$, and space $S$. With probability at least $9/10$ its advice determines a function correct on all promised rational targets. The function has a fixed resource bound even outside the promise. The reductions below preserve dimension and polynomial encoding length.

A sphere perturbation reaches a specified decoding ball with a probability determined by a spherical cap. If $\|v\| = d > a > 0$ and $U$ is uniform on $S^{n-1}$, then

$$\Pr[\|v + \sqrt{d^2 - a^2} U\| \le a] = \Theta_{a/d}(n^{-1/2} (a/d)^n). \tag{129}$$

Enumerating the unknown error length lets us use this optimal perturbation radius for any fixed distance promise.

**Proposition 12.2.** *For every fixed* $R > r$*, exact CVP under the promise* $\mathrm{dist}(t, L) \le R\lambda_1(L)$ *can be solved with constant success probability in time*

$$\mathrm{poly}(n, \mathcal{B}_t)[P + (R/r)^n D]$$

*and space* $S + \mathrm{poly}(n, \mathcal{B}_t)$*. For* $R < 1/2$ *this is* $R$*-BDD. The statement also permits* $R \ge 1/2$ *and multiple closest points.*

**Theorem 12.3** (Direct unique-SVP tradeoff)**.** *For every fixed* $g \ge 1$*, exact* $g$*-uSVP, with promise* $\lambda_2(L) \ge g\lambda_1(L)$*, is solvable with probability at least* $2/3$ *in time and space*

$$T \le \mathrm{poly}(n, \mathcal{B})[P + \max\{1, (gr)^{-n}\} D], \qquad S + \mathrm{poly}(n, \mathcal{B}). \tag{130}$$

*At* $g = 1$ *the problem is ordinary SVP.*

*Proof outline.* A shortest vector $v$ is primitive. Choose a fixed prime $p > g$ and a uniform nonzero linear function on basis coordinates modulo $p$. Its kernel defines an index-$p$ sublattice $L'$. With probability $p^{n-1}(p-1)/(p^n-1) > 1 - 1/p$, it excludes $v$, and then $\lambda_1(L') \geq g\|v\|$. Among its $p-1$ nonzero coset representatives $x_j$ there is one with $x_j - v \in L'$. Preprocess $L'$ once. Direct decoding suffices if $gr \geq 1$. Otherwise perturb $x_j$ and use (129) to hit the decoding ball of $x_j - v$ in $\mathrm{poly}(n)(gr)^{-n}$ attempts. For each legal reply $y \in L'$ keep the unperturbed residual $x_j - y \in L \setminus \{0\}$, and return the shortest candidate. Unknown scales, rational perturbations, and candidate checks are proved in Appendix F.2. □

This uses the sublattice idea of Lyubashevsky and Micciancio [LM09]. At $r = 1/g$ it recovers the polynomial query relation. Below that radius the cap probability gives an explicit exponential query cost, while retaining a single preprocessing cost for each sublattice.

**Theorem 12.4** (GapSVP from reusable decoding)**.** *Fix $\gamma > 1$ with $r\gamma > 1/2$, and put*

$$c(r,\gamma) = -\tfrac{1}{2}\log_2\left(1 - \frac{1}{4r^2\gamma^2}\right). \tag{131}$$

*There is a randomized algorithm distinguishing $\lambda_1(L) \leq d$ from $\lambda_1(L) \geq \gamma d$ with success probability at least $2/3$, time $\mathrm{poly}(n,\mathcal{B})[P + 2^{c(r,\gamma)n}D]$, and space $S + \mathrm{poly}(n,\mathcal{B})$.*

*Proof outline.* Fix the advice as a deterministic function $F$ on a half-open fundamental parallelepiped. Draw $e$ from a ball of radius $R = r\gamma d$, decompose $e = x + y$ with $y \in L$, and check $F(x) = -y$. In the NO case good advice passes every check, since $\|e\| \leq r\lambda_1(L)$.

In the YES case take $0 \neq v \in L$ with $\|v\| \leq d$. Whenever both $e$ and $e + v$ lie in the ball, they have the same representative $x$ but demand different answers. Every fixed $F$ therefore fails on at least half the overlap, a fraction at least

$$n^{-O(1)}\left(1 - \frac{1}{4r^2\gamma^2}\right)^{n/2} = n^{-O(1)}2^{-c(r,\gamma)n}.$$

Repeated checks detect a conflict with constant probability. Appendix F.3 proves this argument on a finite rational grid containing $L$ and implements ball sampling with polynomial overhead. The YES argument holds for every advice function, including one obtained outside a sampling promise. □

The two geometries yield different query exponents. With $P = 2^{an+o(n)}$ and $D = 2^{bn+o(n)}$, the total exponents are

$$\max\{a, b + [\log_2(1/(gr))]_+\}, \qquad \max\{a, b + c(r,\gamma)\},$$

respectively. The following choices fit both query costs into the preprocessing time.

**Algorithms at the BDGL exponents.** The full batch gives the radius supremum

$$r_* = \frac{\sqrt{e\beta_{\mathrm{G}}\log 2}}{(3/2)2^{0.401}} = 0.315707104\ldots. \tag{132}$$

Every fixed $r < r_*$ has preprocessing exponent $\alpha_{\mathrm{G}}$, query exponent $\beta_{\mathrm{G}}$, and space exponent $\beta_{\mathrm{G}}$ by Theorem 12.1. The query savings leave room for $2^{(\alpha_{\mathrm{G}}-\beta_{\mathrm{G}})n}$ attempts during the preprocessing time.

**Corollary 12.5.** *Exact 3-uSVP and 4.15-GapSVP admit classical randomized algorithms with success probability at least $2/3$, time $2^{\alpha_{\mathrm{G}}n+o(n)}$, and bit space $2^{\beta_{\mathrm{G}}n+o(n)}$ on every execution. More generally, the same bounds hold for exact g-uSVP for every fixed $g > 2.986340909\ldots$.*

*Proof.* For uSVP, the ratio of preprocessing time to query time is $2^{(\alpha_G-\beta_G)n} = (3/\sqrt{8})^n$. Hence Theorem 12.3 fits the budget whenever $g > 1/(r_* 2^{\alpha_G-\beta_G}) = 2.986340909\ldots$.

For the explicit factors, use the parameters in the table below. For 3-uSVP take prime 5 and geometric hit radius 0.3155. Then $\log_2(1/(3 \cdot 0.3155)) < 0.084$. For 4.15-GapSVP take geometric radius 0.2635. The smaller batch is feasible since $b < J(\alpha^{-2})$ and $0.26355 < r(1.382, 0.1228) = 0.263595555\ldots$. Its call exponent is $c(0.2635, 4.15) < 0.1694$. In each row, this exponent plus $b$ is below $\alpha_G$. The gaps between the base and geometric radii pay for finite perturbations. □

| Application | $\alpha$ | $b$ | Base radius | Query rate bound |
|---|---|---|---|---|
| 3-uSVP | 1.5 | $\beta_G$ | 0.3156 | 0.084 |
| 4.15-GapSVP | 1.382 | 0.1228 | 0.26355 | 0.1694 |

The query rate column counts calls, while each call has work exponent $b$. In both rows their sum is below $\alpha_G$. The full batch serves uSVP, while the smaller batch permits the narrower width used by the GapSVP test.

## 12.4 Centered Gaussian sampling from SVP

The shortest vector algorithm also gives centered Gaussian samples at every width. Following the primitive direction sparsification approach of Stephens-Davidowitz [SD16, Section 4], we reduce sampling to SVP calls in the same dimension.

Each primitive direction $y$ carries the Gaussian mass of all its nonzero multiples,

$$h_\sigma(\|y\|) = 2\sum_{k\geq 1} e^{-\pi k^2\|y\|^2/\sigma^2}.$$

Sparsification gives approximately uniform directions inside a ball. If the ball of radius $r_i$ contains $N_i$ directions, select it with weight $N_i(h_\sigma(r_i) - h_\sigma(r_{i+1}))$, then draw a direction nearly uniformly inside it. The factor $N_i$ cancels, and summing over balls telescopes to the desired direction weight. A one-dimensional Gaussian then selects the multiple of that direction.

Near zero, $h_\sigma(r)$ varies like $1/r$, so a uniform additive grid can miss a large change in direction weight. A sufficiently fine multiplicative grid controls adjacent weights throughout the truncated range and still has polynomially many radii. Appendix D proves this comparison and treats balls with too few primitive directions. Algorithm 7 displays the complete mixture, including the zero vector and the separate treatment of those sparse layers. This yields the following oracle reduction.

**Theorem 12.6** (Centered sampling from SVP)**.** *Let $L = B\mathbb{Z}^n$ be a full rank rational lattice, $\sigma^2 > 0$ rational, and $f \geq 1$ an integer. There is an oracle algorithm whose output law is $(1 + 1/f, 2^{-f})$-close to $D_{L,\sigma}$ in the convention of Section 12.1. In the declared computational model, its work, counting each SVP call as one operation, is bounded by a polynomial in $\mathrm{size}(B, \sigma^2)$ and $f$ on every execution. The same is true of its workspace and number of oracle calls. Every query is a full rank sublattice of $L$ with polynomial encoding length. The oracle may choose among shortest vectors using any rule that depends on the current and previous queries.*

*Proof of Corollary 1.4.* Apply Theorem 12.6 with parameter $f = f(n)$, replacing each oracle query by amplified calls to Theorem 1.3. The queries have rank $n$ and polynomial encoding length, so their sequential execution and the reduction's polynomial overhead preserve the time exponent 0.4130 and space exponent 0.3138. The budget in Appendix D.8 includes all oracle failures, giving additive error below $2^{-f(n)}$ and multiplicative error at most $1 + 1/f(n)$, with resources bounded on every execution. □

## 12.5 Gaussian samples for CVP and successive minima

For $\gamma > 1$, $\gamma$-CVP asks, given a rational target $t$, for $v \in L$ with $\|v - t\| \le \gamma \operatorname{dist}(t, L)$. The problem $\gamma$-SMP asks for independent $v_1, \dots, v_n \in L$ with $\|v_i\| \le \gamma \lambda_i(L)$ for every $i$. This also solves $\gamma$-SIVP. Define

$$\Phi(\Delta) = \begin{cases} 1 + \log(1/\Delta), & 0 < \Delta \le 1, \\ 1/\Delta, & \Delta \ge 1. \end{cases} \tag{133}$$

Here an effective constant admits polynomial-time rational approximation to any requested precision. An effective factor depending on $n$ is polynomial-time computable with polynomial encoding length.

Time–approximation tradeoffs from CVP to SVP were established by Aggarwal, Chen, Kumar, Li, and Stephens-Davidowitz [ACKLS21]. For the Euclidean norm and $\gamma \ge 4$, their Theorem 5.1 gives reduction cost $\widetilde{O}(\exp(16n/\gamma^2))$, using SVP instances in dimension $n + 1$. Their Theorem 5.2 isolates the sampling step. We refine the explicit sampling-call exponent, cover every fixed $\gamma > 1$, and obtain direct SMP and SIVP bounds in dimension $n$ from the same estimates. We first separate these geometric arguments from the cost of our sampler.

**Proposition 12.7.** Sampler assumption. *Suppose a sampler accepts every full rank rational lattice $K \subset \mathbb{R}^N$ and every positive rational width $s$, and permits $(1 + 1/f(N), 2^{-f(N)})$-closeness to $D_{K,s}$ for every efficiently computable polynomially bounded precision parameter $f$.*

Approximation algorithms. *For every fixed effective $\gamma > 1$, there are randomized algorithms for $\gamma$-SMP, $\gamma$-SIVP, and $\gamma$-CVP with success probability at least $2/3$, making at most*

$$Q_\gamma(n, I) = \operatorname{poly}(|I|, n) \exp\left(\frac{n}{2}\Phi(\gamma^2 - 1) + O(1)\right) \tag{134}$$

*sampler calls. Here $I$ is the rational input. All calls for SMP and SIVP have dimension $n$. All calls for CVP have dimension $n + 1$.*

Resources and growing factors. *Their inputs have polynomial encoding length. The number of calls is fixed before sampling, the additional work per call is polynomial, and the additional storage is polynomial.*

*For every effective $\gamma(n) = \omega(1)$, the construction is uniform and uses $2^{o(n)} \operatorname{poly}(|I|, n)$ calls. In the asymptotic range the precision choice is $f(N) = N^2$.*

Appendix E proves the proposition, including rational widths, unknown scales, candidate verification, and predetermined call budgets. Sequential calls reuse the sampler workspace.

The common geometric observation is a choice of sign. In each pair $\{x, -x\}$, choose the sign with $\langle u, \pm x \rangle \le 0$. Then

$$\|u \pm x\|^2 \le \|u\|^2 + \|x\|^2, \qquad \rho_s(u \pm x) \ge \rho_s(u)\rho_s(x).$$

A centered small ball estimate supplies the mass of the possible $x$. For a closest displacement this yields Lemma E.1. For a vector $u$ outside a fixed subspace it yields Lemma E.2. Both give, up to polynomial factors,

$$p_\gamma \ge \beta^{n/2} \exp[-n\beta/(2\Delta) - O(1)], \qquad \Delta = \gamma^2 - 1, \quad 0 < \beta \le 1.$$

Minimizing $\beta/\Delta + \log(1/\beta)$ gives $\beta = \min\{1, \Delta\}$ and (133). The event guarantee (127), with $f(N) = N^2$, preserves this exponentially small success probability. For SMP, fix the current span and draw fresh samples until a short vector outside it appears. For CVP, the rational embedding $\{(v - kt, ks) : v \in L, k \in \mathbb{Z}\}$ uses centered sampling in dimension $n + 1$. Its layers $k = \pm 1$ contain the desired candidates.

In the Euclidean setting and the common range $\gamma \geq 4$, the sampling overhead expressed with natural exponentials decreases from $16/\gamma^2$ in [ACKLS21, Theorems 5.1–5.2] to $1/[2(\gamma^2-1)]$. At $\gamma = 4$ the coefficients are 1 and $1/30$. The logarithmic branch of $\Phi$ also covers every fixed $1 < \gamma < \sqrt{2}$. SMP and SIVP retain dimension $n$.

**Instantiating the sampler.** Write $\alpha_{\mathrm{SVP}} = 0.4130$ and $\beta_{\mathrm{SVP}} = 0.3138$ for the displayed time and space exponents of Corollary 1.4. Sequential calls add the sampling-call exponent to $\alpha_{\mathrm{SVP}}$ and reuse the same workspace.

**Corollary 12.8.** *For every fixed effective $\gamma > 1$, there are classical randomized algorithms for $\gamma$-CVP, $\gamma$-SMP, and $\gamma$-SIVP with success probability at least $2/3$ and resource bounds*

$$T \leq 2^{\left(\alpha_{\mathrm{SVP}} + \Phi(\gamma^2-1)/(2\log 2)\right)n + o(n)}, \qquad S \leq 2^{\beta_{\mathrm{SVP}} n + o(n)}.$$

*For every effective $\gamma(n) > 1$ with $\gamma(n) = \omega(1)$, these problems admit time $2^{\alpha_{\mathrm{SVP}} n + o(n)}$ and space $2^{\beta_{\mathrm{SVP}} n + o(n)}$. Both resource bounds hold on every execution.*

For example, the time exponents are below 0.4968 at $\gamma = 3.1$ and 0.4611 at $\gamma = 4$, with space exponent 0.3138 in both cases.

The extra time pays for rare useful samples. As $\gamma(n) \to \infty$, their call exponent tends to zero.

## 12.6 Short independent vectors near smoothing

Polynomially many samples above smoothing contain $n$ linearly independent vectors with high probability, and their lengths are controlled by the last successive minimum. The standard reduction of [ADRS15, Lemma 7.5 and Corollary 7.6] therefore gives the following consequence.

**Corollary 12.9.** *Given a full rank rational lattice $L \subset \mathbb{R}^n$, there is a classical randomized algorithm that returns $n$ linearly independent vectors $v_1, \ldots, v_n \in L$ and, with probability at least $2/3$, satisfies*

$$\max_i \|v_i\| \leq C\sqrt{n \log(n+2)}\, \lambda_n(L)$$

*for an absolute constant $C$. Its time and space are*

$$T = (3/2)^{n/2 + o(n)}, \qquad S = (4/3)^{n/2 + o(n)}.$$

The proof in Appendix E.4 uses independent executions of the weighted readout sampler in Theorem 1.1. Polynomially many sequential calls preserve its leading time and space exponents.

## 12.7 Polynomial-factor approximations by block reduction

Block reduction uses SVP in a smaller dimension at the cost of a larger approximation factor. Applying Theorem 1.3 in [ALNS20] gives the bounds below. Hermite SVP (HSVP) measures length against $\det(L)^{1/n}$, and SVP against $\lambda_1(L)$.

**Corollary 12.10.** *For every fixed $c$ in the indicated range, there is a classical randomized algorithm that, on input a full rank rational lattice $L \subseteq \mathbb{R}^n$, returns a nonzero $v \in L$ with probability at least $2/3$ of satisfying the following bound:*

| *Problem and range* | *Output length* | *Time* | *Space* |
|---|---|---|---|
| *SVP*, $c \geq 1$ | $\|v\| \leq O_c(n^c)\lambda_1(L)$ | $2^{\alpha_{\mathrm{SVP}} n/(c+1)+o(n)}$ | $2^{\beta_{\mathrm{SVP}} n/(c+1)+o(n)}$ |
| *SVP*, $1/2 < c < 1$ | $\|v\| \leq O_c(n^c)\lambda_1(L)$ | $2^{\alpha_{\mathrm{SVP}} n/(2c)+o(n)}$ | $2^{\beta_{\mathrm{SVP}} n/(2c)+o(n)}$ |
| *HSVP*, $c \geq 1/2$ | $\|v\| \leq O_c(n^c)\det(L)^{1/n}$ | $2^{\alpha_{\mathrm{SVP}} n/(2c)+o(n)}$ | $2^{\beta_{\mathrm{SVP}} n/(2c)+o(n)}$ |

For example, $O(n)$-SVP takes $2^{0.2065n+o(n)}$ time and $2^{0.1569n+o(n)}$ space. The SVP rows use [ALNS20, Corollaries 4.4 and 3.4] with block sizes $\lfloor n/(c+1) \rfloor$ and $\lceil n/(2c) \rceil$, respectively. For HSVP, use [ALNS20, Theorem 2.3] with $k = \lceil n/(2c) \rceil$. At $c = 1/2$, SVP and Minkowski's bound suffice. Only constant-factor block SVP is needed, so rational Gram-matrix approximation permits our oracle on projected lattices. Appendix E.3 gives the interface and resource analysis.

## 13 Conclusion and further discussion

Gaussian difference dynamics controls the expected counts and dependence of lattice vectors through their first two counting moments. Accurate moment bounds yield a Gaussian sample, while weaker bounds allow more differences to be retained for shortest vector search. This gives provable DGS above smoothing at the BDGL heuristic exponents and exact SVP on arbitrary lattices in $2^{0.4130n+o(n)}$ time.

**SVP at the BDGL exponent.** Can exact SVP on arbitrary lattices be solved in $(3/2)^{n/2+o(n)}$ time, matching the BDGL heuristic exponent? Our sampler attains this exponent above smoothing, while the shortest vector construction requires additional representations at the terminal scale. Under the terminal balance of Proposition 9.3, the optimized time exponent is approximately 0.412962. Changing the intermediate shell energies may improve space usage, but does not lower this terminal cost.

**Exact CVP below $2^n$.** Another central question is whether exact CVP on arbitrary lattices admits time $2^{(1-\epsilon)n+o(n)}$ for some fixed $\epsilon > 0$. The equivalence between CVP and DGS at arbitrary shifts and widths gives a sampling formulation of this question [SD16, Theorem 3.6 and Section 7]. An extension of fast sampling to arbitrary shifts below smoothing would therefore also give faster exact CVP. Our two sampling routes suggest a starting point: direct sampling handles arbitrary shifts above smoothing, and the SVP reduction handles centered sampling at every width.

**Practical and heuristic algorithms.** A further goal is to implement algorithms based on DGS and bring their practical performance closer to that of heuristic sieving. For SVP, can their time and memory costs be made comparable to those of heuristic sieves in dimensions of practical interest? Alongside implementations of the provable constructions, heuristic variants of DGS algorithms are also worth exploring.

**Extending GDD to other norms.** The counting framework provides a common way to control dependence, while each construction supplies its own reference profile and geometric estimates. The cube construction of [Gao26] illustrates this flexibility beyond Gaussian profiles. Extending the approach to other $\ell_p$ norms calls for suitable reference profiles, difference rules, and efficient pair

searches. Can these ingredients improve the provable time and space bounds for SVP, and how do the resulting exponents depend on $p$?

# A Lattice and Gaussian tools

## A.1 Centered Gaussian moments and scale comparison

For every full rank lattice $L \subset \mathbb{R}^n$, $s > 0$, and $0 < \beta \le 1$,

$$\mathbb{E}_{X\sim D_{L,s}}[\|X\|^2] \le \frac{ns^2}{2\pi}, \qquad \rho_{\sqrt{\beta}s}(L) \ge \beta^{n/2}\rho_s(L), \qquad \rho_s(L-t) \le \rho_s(L). \tag{135}$$

These standard consequences of Poisson summation follow by taking a logarithmic derivative, comparing widths, and using positivity of the Fourier coefficients, respectively.

## A.2 Weighted radial counts

*Proof of Lemma 3.5.* Put $a = n/2$ and $N(x) = \#\{w \neq 0 : \xi_w \le x\}$. For $x < 1$, $N(x) = 0$, and Lemma 3.4 bounds it by $C_{R_0}(n+1)2^{\kappa_{\mathrm{pack}}n}x^a$ on the remaining integration range. Stieltjes integration by parts gives

$$\sum_{w\neq 0}(H - c\xi_w)_+^a = ac\int_0^{H/c} N(x)(H-cx)^{a-1}\,dx.$$

With $x = (H/c)t$, the integral after inserting the count bound, apart from its factor $C_{R_0}(n+1)2^{\kappa_{\mathrm{pack}}n}$, is

$$H^{2a}c^{-a}\frac{\Gamma(a+1)^2}{\Gamma(2a+1)} \le \mathrm{poly}(n)\left(\frac{H^2}{4c}\right)^a.$$

The last estimate follows from Stirling's inequality. Boundary terms vanish since $N(0) = 0$ and the weight is zero at $H/c$. □

## A.3 Basis preprocessing and length estimates

We derive Proposition 3.2 from DBKZ in successive projections. For exact SVP on a rational projected lattice of rank $d$ in $\mathbb{R}^n$, [MV13, Theorem 5.4 and Corollary 5.5] gives $4^{d+o(d)}$ time and $2^{d+o(d)}$ space, with polynomial input overhead.

**Projected calls.** Fix $2 \le k < n$. Suppose a prefix $F = (b_1, \dots, b_i)$ has been fixed as part of a basis $(F, D)$ of $\Gamma$. Its orthogonal projection and remaining lattice are

$$\pi = I - F(F^{\mathsf{T}}F)^{-1}F^{\mathsf{T}}, \qquad \Lambda = (\pi D)\mathbb{Z}^d = \pi\Gamma, \qquad d = n - i,$$

with $\pi = I$ when $i = 0$. Clear one common positive denominator in $\pi D$ to obtain the integer input required by ALNS20. This preserves length ratios and coefficients: an output with coefficient vector $z$ lifts to $Dz$. Use the same conversion for rational queries to the exact oracle.

If $d > k$, apply [ALNS20, Theorem 2.3] with block dimension $k$, exact oracle factor $\delta = 1$, and slack $\varepsilon = 1$. It returns a nonzero projected vector $v$ with

$$\|v\| \le 2\gamma_k^{(d-1)/(2(k-1))}\det(\Lambda)^{1/d} \le 2k^{(d-1)/(2(k-1))}\det(\Lambda)^{1/d},$$

using polynomially many SVP calls of rank $k$. Here $\gamma_k \le k$ is the Euclidean Hermite bound, and the determinant is taken within the span of $\Lambda$. If $d \le k$, use exact SVP on $\Lambda$.

**From projected vectors to a basis.** Take the primitive core of the returned coefficient vector, complete it unimodularly, and size reduce the lifted column against the fixed prefix. LLL reduction of the remaining projected suffix, followed by size reduction in the original space, prepares the next round while preserving the chosen Gram–Schmidt vectors [LLL82, MV13].

The projections of $n$ independent vectors of length at most $\lambda_n(\Gamma)$ span the remaining space. Choose $d$ independent projections. Their integer span is a sublattice of $\Lambda$, so Hadamard's inequality gives

$$\det(\Lambda)^{1/d} \le \lambda_d(\Lambda) \le \lambda_n(\Gamma).$$

For $d > k$, the selected column therefore satisfies

$$\|b^*_{i+1}\| \le 2k^{(d-1)/(2(k-1))}\lambda_n(\Gamma) \le H(n,k)\lambda_n(\Gamma).$$

For $d \le k$, exact SVP gives the stronger bound $\|b^*_{i+1}\| \le \lambda_n(\Gamma)$. Later operations preserve each selected Gram–Schmidt vector, proving the geometric conclusion of Proposition 3.2.

**Representation and resource bounds.** The usual LLL bound gives Gram–Schmidt lengths at most $2^{n/2}\lambda_n(\Gamma)$ in each remaining suffix. After size reduction, every current column therefore has length at most

$$\sqrt{1+n/4}\,\max\{H(n,k), 2^{n/2}\}\lambda_n(\Gamma).$$

Its denominator divides the original common denominator, and $\lambda_n(\Gamma) \le \max_j \|b_{0,j}\|$. Thus each round begins with a polynomial description bound in terms of $B_0$, preventing encoding growth across projections. Rational linear algebra and the intermediate bounds of [ALNS20, Section 2.3] give the same control within a round. The $n$ rounds make polynomially many calls of rank at most $k$, proving the resource bound. For $k = \lfloor \vartheta n \rfloor$, direct substitution gives the stated factor $O_\vartheta(n^{1/(2\vartheta)})$ and exponent $2\vartheta$. Fixed small ranks use exact SVP.

## A.4 Rational projected basis interfaces

The block reductions query projected lattices and their duals, which have rational rectangular bases. The following construction puts these queries in the square basis format of Theorem 1.3.

Let $A \in \mathbb{Q}^{m\times d}$ have full column rank. In polynomial time in its rational input length, one can compute an invertible $C \in \mathbb{Q}^{d\times d}$, of polynomial encoding length, such that

$$\|Az\|^2 \le \|Cz\|^2 \le \tfrac{3}{2}\|Az\|^2 \qquad \text{for every } z \in \mathbb{R}^d. \tag{136}$$

To obtain $C$, compute the rational factorization $A^{\mathsf{T}}A = U^{\mathsf{T}}\operatorname{diag}(d_i)U$ with $d_i > 0$. Choose rational $r_i$ with $d_i \le r_i^2 \le 3d_i/2$ and set $C = \operatorname{diag}(r_i)U$. The Gram matrix inequalities give (136). A shortest vector $Cz$ maps back to $Az$, whose length is at most $\sqrt{3/2}\lambda_1(A\mathbb{Z}^d)$. This supplies the 2-approximate SVP oracle needed by the block reductions without changing the rank.

Orthogonal projections and dual bases within rational spans are rational, so the construction applies to both types of block query. Lift the returned integer coordinates in the original block, taking their primitive core when needed. Clearing denominators allows the use of the integer-lattice statements in [ALNS20] without changing approximation factors.

## A.5 Radial and linear tails

The following estimates support Gaussian cooling and midpoint lenses, uniformly on the compact parameter ranges in Sections 5 and 8.

Let $X$ be a discrete Gaussian of width $t$ on a shifted lattice. Suppose the shifted theta estimates hold with error $\varepsilon$ at the widths used below. Put $U = \pi\|X\|^2/t^2$ and $R_\varepsilon = (1+\varepsilon)/(1-\varepsilon)$. Ratios of theta masses and completion of squares give

$$\mathbb{E}[e^{aU}] \le R_\varepsilon (1-a)^{-n/2}, \qquad \mathbb{E}[e^{b\langle X,z\rangle}] \le R_\varepsilon e^{b^2 t^2 \|z\|^2/(4\pi)}.$$

The first requires $a < 1$ and theta control at $t/\sqrt{1-a}$. The second holds for every real $b$, since the tilt translates the coset. Optimizing gives, for $0 < \delta \le 1/2$ and, in the last bound, $z \ne 0$ and $h > 0$,

$$\begin{aligned}
\mathbb{P}[U \ge (1+\delta)n/2] &\le R_\varepsilon e^{-n[\delta - \log(1+\delta)]/2}, \\
\mathbb{P}[U \le (1-\delta)n/2] &\le R_\varepsilon e^{-n[-\delta - \log(1-\delta)]/2}, \\
\mathbb{P}[|\langle X, z\rangle| \ge h] &\le 2R_\varepsilon e^{-\pi h^2/(t^2\|z\|^2)}.
\end{aligned}$$

For (41), tilt the reference Gaussian to width $r\sqrt{a_0}$. The shell density factor is $e^{-n(a_0-1)/2+O(nw_n)}$, and the volume factor is $a_0^{n/2}$. The tilted width has a fixed smoothing margin $r\sqrt{a_0}/s \ge \sqrt{2a_0} > 1$. The radial bound puts $1-o(1)$ of the tilted mass in a narrower shell. Multiplying the density and volume factors gives $\Upsilon 2^{-J(a_0)n+O(\ell_n)}$.

### A.6 Gaussian profile and pairing degree bounds

For the midpoint in Lemma 5.1, take

$$t = r/\sqrt{2}, \qquad U = n\mathcal{E}_r(M), \qquad L_z = \frac{2\pi\langle M, z\rangle}{nr^2}, \qquad R_\epsilon = \frac{1+\epsilon_\Gamma}{1-\epsilon_\Gamma}.$$

For each eligible $z$, the region $\mathcal{W}_r(z)$ contains

$$\mathcal{T}_r(z) = \{M : |\mathcal{E}_r(M) - 1/2| \le w_n/8,\ |L_z| \le w_n/8\}.$$

Indeed, on this region,

$$|\mathcal{E}_r(M \pm z/2) - a_0| \le w_n/8 + a_1 w_n/4 + w_n/8 = (a_0 - 1/4)w_n < a_0 w_n.$$

The preceding bounds give

$$\begin{aligned}
\mathbb{P}[|U - n/2| > nw_n/8] &\le 2R_\epsilon e^{-nw_n^2/128}, \\
\mathbb{P}[|L_z| > w_n/8] &\le 2R_\epsilon \exp[-nw_n^2/(64\mathcal{E}_r(z))].
\end{aligned}$$

Every radial tilt is valid because $t\sqrt{1-w_n/4} \ge (s/\gamma)\sqrt{1-w_n/4} \ge s$. The linear tilt uses shifted theta flatness at $t$. Consequently the mass on $\mathcal{T}_r(z)$ is at least $V_{r/\sqrt{2}}(1 - O(\epsilon_\Gamma) - e^{-\Omega(nw_n^2)})$. The mass on $\mathcal{W}_r(z)$ lies between this lower bound and the full shifted mass, at most $(1+\epsilon_\Gamma)V_{r/\sqrt{2}}$. This proves (47).

For the coefficient in (46), use

$$e^{n(1-a_0)/2}\beta^{n/2} = 2^{-(J(a_0)+P(u_1))n}, \qquad A_r R_-^2 = n(1-a_0)(1-w_n)/2.$$

To bound a weighted degree, fix an eligible endpoint $x$. Completion of squares gives

$$\rho_r(y) = \rho_{r\sqrt{\beta}}(y - ux)\exp\left\{\frac{n}{2}\left[-u\mathcal{E}_r(x) + \frac{u}{1-u}\mathcal{E}_r(x-y)\right]\right\}.$$

The bounds $\mathcal{E}_r(x) \ge a_0(1-w_n)$ and $\mathcal{E}_r(x-y) \le a_1(1+w_n)$ put the bracket at most $(1-a_0)(1+3w_n)$. Since $r\sqrt{\beta} \ge s$ and $\omega_r \le 1$, summing over the full coset bounds the degree by

$$\Upsilon_j \beta^{n/2} e^{n(1-a_0)(1+3w_n)/2}(1+\epsilon_\Gamma) = \Upsilon_j \mathsf{B}_j e^{2(1-a_0)nw_n}(1+\epsilon_\Gamma) \le D_j^+.$$

The last inequality uses $b_{\deg} = \lceil 10\ell_n \rceil$. Fixing the right endpoint gives the same bound.

## A.7 A common Gaussian window count

**Lemma A.1.** *Let $K+b$ be a full rank affine lattice and let $\beta$ range over a fixed compact positive interval. Put $t = r\sqrt{\beta}$ and suppose shifted theta flatness holds with error at most $1/4$ at $t$ and its relative $O(w_n)$ radial tilts. Let $\|z\|^2 = O(nr^2)$. Suppose a window $W$ contains*

$$|\mathcal{E}_r(x) - \beta| \le c_1 w_n, \qquad |2\pi\langle x, z\rangle/(nr^2)| \le c_1 w_n$$

*and is contained in $|\mathcal{E}_r(x) - \beta| \le c_2 w_n$, where $0 < c_1 < c_2$ are fixed. Then*

$$\#((K+b)\cap W) = \frac{r^n}{\det K}(e\beta)^{n/2} e^{\pm O(nw_n + \log n)}.$$

*Proof.* The radial and linear exponential moments above put $1 - O(e^{-\Omega(nw_n^2)})$ of $D_{K+b,t}$ in the inner rectangle. On either radial window, $\rho_t(x) = e^{-n/2+O(nw_n)}$. Multiply the Gaussian mass of the inner rectangle and of the outer window by the corresponding inverse weight bounds. Theta flatness gives $\rho_t(K+b) = (1 \pm 1/4)t^n/\det K$ and proves both inequalities. □

## A.8 Corrected Gaussian proposals and leaf moments

*Proof of Lemma 3.3.* The Klein/GPV proposal [Klein00, GPV08] starts with $v_n = -p$. For $i = n, \dots, 1$, set

$$h_i = \frac{\langle v_i, b_i^*\rangle}{\|b_i^*\|^2}, \qquad s_i = \frac{R}{\|b_i^*\|}, \qquad z_i \sim D_{\mathbb{Z}, s_i, h_i}, \qquad v_{i-1} = v_i - z_i b_i.$$

It proposes $x = p + \sum_i z_i b_i$. Orthogonal decomposition makes the product of the Gaussian numerators equal to $\rho_R(x)$, hence

$$Q_R(x) = \frac{\rho_R(x)}{\prod_i \theta_{s_i}(h_i)}, \qquad \theta_s(h) = \sum_{k\in\mathbb{Z}} e^{-\pi(k-h)^2/s^2}.$$

Poisson summation and $e^{-\pi s_i^2} \le 1/(8n)$ give $|\theta_{s_i}(h_i)/s_i - 1| \le 2/(8n-1) \le 1/(3n)$. Since $\prod_i s_i = V_R$, accept with probability

$$a(x) = \frac{1}{2V_R}\prod_i \theta_{s_i}(h_i) = \frac{1}{2}\prod_i \frac{\theta_{s_i}(h_i)}{s_i} \le \frac{1}{2}e^{1/3} < 1.$$

Then reject points outside $\mathcal{S}$. The product $Q_R(x)a(x)$ is exactly (15).

For finite implementation, use the uniform scalar sampler of Appendix C.1.7 with error $\delta/(2n)$ at each coordinate. After reducing $h_i$ modulo $\mathbb{Z}$, the Poisson series for $\theta_{s_i}(h_i)/s_i$ converges geometrically and is bounded away from zero. Certified intervals evaluate the product and its acceptance coin to error $\delta/2$ using Appendix C.1.6. The scalar guarantee and sequential coupling give total error at most $\delta$ within deterministic polynomial budgets. □

Make $2M$ independent trials of Lemma 3.3, where $M \ge 1$ is an integer, and retain their nonfailure records with their trial IDs. For

$$\nu(x) = \frac{M}{V_R}\rho_R(x)\mathbf{1}_{x\in(p+\Gamma)\cap\mathcal{S}},$$

independence gives the leaf moments

$$m = \nu, \qquad S^{\mathrm{act}} = \left(1 - \frac{1}{2M}\right)\nu^{\otimes 2} \le \nu^{\otimes 2}. \tag{137}$$

Thus the retained list belongs to $\mathcal{C}(\nu, 0)$.

Cooling uses these $2M$ trials to initialize coefficient $\Upsilon_t = M$. Batch streams instead keep $M$ trial labels and use the same single trial law, with coefficient $\lambda_t = 1/2$. Thus the fixed trial counts determine known coefficients in both constructions, while only nonfailure records occupy vector storage.

# B Search and application estimates

## B.1 Midpoint and output counts

We justify the geometric factors in Lemma 6.2. At bridge energy $E$, the midpoint Gaussian has width $r\sqrt{E}/2$ on $\Gamma/2$. The first moment argument of Theorem 8.1 therefore applies whenever $gE > 1$ with a fixed margin, giving $V_r(eE/4)^{n/2} e^{O(\ell_n)}$ representations per output.

After the ancestor average, the output support is $p_* + \Gamma/t$ with density $t^{-n}$. For an output shell of fixed positive energy $b$, put $p_- = t^2 g b$. If $p_- \geq 1$, the Gaussian of width $r\sqrt{b}$ is above smoothing on $\Gamma/t$. If $p_- < 1$, the shifted maximum and monotonicity give

$$\rho_{r\sqrt{b}}(p_* + \Gamma/t) \leq \rho_{s/t}(\Gamma/t) \leq (1 + \varepsilon_n) \frac{(r\sqrt{b})^n}{\det(\Gamma/t)} p_-^{-n/2}.$$

The usual Gaussian majorant on a thin shell consequently yields

$$t^{-n} \#((p_* + \Gamma/t) \cap \mathcal{S}_r(b)) \leq V_r(eb)^{n/2} e^{O(n w_n + \log n)} \max\{1, p_-^{-n/2}\}. \tag{138}$$

All estimates are uniform on fixed compact positive energy ranges. Multiplication with the midpoint count gives $V_r^2 (ea)^n (1 - u^2)^{n/2}$ and $p_- = c_0(1 - u)$, proving the rate in Lemma 6.2. Endpoint perturbations and bins of width at most $2/n$ cost $O(\ell_n)$ in the logarithm. A fixed compact interval uses $O(n)$ bins.

## B.2 Nearly parallel pairs and equal values

Choose a fixed $\Delta \in (0, 1)$ with $c_0 \Delta < 1$. Put $u_* = 1 - \Delta > 0$ and require the midpoint smoothing factor $2ga(1 + u_*)$ to exceed one by a fixed margin. For correlations $u \geq u_*$, the difference energy is at most $b_* = 2a\Delta + O(w_n)$. Write $E_* = 4a - b_*$ at the central parameters. The applications choose fixed cutoffs satisfying these conditions.

Use Gaussian widths $r\sqrt{E_*}/2$ for the midpoint and $r\sqrt{b_*}$ for the output. On the endpoint shells, $\mathcal{E}_r(M) = a - \mathcal{E}_r(z)/4 + O(w_n)$, and

$$\frac{\mathcal{E}_r(M)}{E_*/4} + \frac{\mathcal{E}_r(z)}{b_*} = \frac{4a}{E_*} + \mathcal{E}_r(z) \left( \frac{1}{b_*} - \frac{1}{E_*} \right) + O(w_n) \leq 2 + O(w_n).$$

The coefficient of $\mathcal{E}_r(z)$ is positive because $u_* > 0$. Thus $e^{n + O(n w_n)}$ times this product Gaussian bounds the counting weight throughout the region, including $z = 0$. The local and ancestor averages separate the midpoint and output sums. Their smoothing factors are $gE_* > 1$ and $t^2 g b_* = c_0 \Delta < 1$, respectively. The theta and monotonicity estimates above give candidate mass

$$2^{[2\beta - P(1-\Delta)]n + O(b_n)} (c_0 \Delta)^{-n/2}. \tag{139}$$

A common cap is contained in either single cap. Multiplying by $M_f C_{d_{\mathrm{pc}}}(\tau)$ therefore gives work rate

$$K_{\mathrm{code}} + 2\beta + \frac{1}{2} \log_2 \frac{(2 - \Delta)(1 - \tau^2)}{c_0}. \tag{140}$$

Replacing $2 - \Delta$ by 2 yields the second term of (61). Every occurrence is counted with its multiplicity.

### B.3 Pairs near the common cap threshold

The exact geometric estimate from [GFH26b, Lemma 4.1] is

$$W_{d_{\mathrm{pc}}}(\tau, u) \le \left(1 - \frac{2\tau^2}{1+u}\right)^{(d_{\mathrm{pc}}-2)/2} \tag{141}$$

for $-1 < u < 1$ and positive slack, with zero measure otherwise. If the slack is at most $1/16$, input size limits $|A|, |B| \le 2^{\beta n + O(b_n)}$ give expected bucket work at most $2^{(K_{\mathrm{code}}+2\beta-2)n+O(b_n)}$. The forward calls have $\beta < 1/2$ and $\beta \ge P(\tau)$, so this is below the first rate in (61). The remaining correlation range has a fixed positive slack and is covered by the compact bins and (60). Nonnegativity permits the same full reference sums to bound capped sums.

The traversal bound also holds with these work estimates. Choose $S = 2^{\beta n + C_S b_n}$ after the list caps, with $C_S$ larger than their insertion slack constants. In (55), the term without $Z_A/S$ has rate $\beta + o(1)$, and the other has at most the point incidence rate, up to $O(b_n)$ in the exponent. Polynomially many sequential codes preserve the stated time and space rates.

### B.4 Representation counts in random cosets

Use the notation of Section 9. Fix an upper bound $a_{\max}$ for the compact energy range and choose $R_0 \ge 1$ with $R_0^2 > 16a_{\max}/c + 1$. For large $n$, $c_n \ge c/2$ and $w_n \le 1$. This radius bound covers all uses of Lemmas 3.4 and 3.5 below.

#### B.4.1 The ambient lens and residue collisions

We supply the geometric estimates used in Theorem 9.1. Put $M = x - v/2$ and $\beta_n = E_n/4$. The endpoint shell conditions become

$$|\mathcal{E}_r(M) - \beta_n| + \left|\frac{2\pi\langle M, v\rangle}{nr^2}\right| \le aw_n. \tag{142}$$

On $\delta - v/2 + K$, use a Gaussian of width $t = r\sqrt{\beta_n}$. Dense ambient flatness remains uniform at $t$ and its relative $O(w_n)$ radial tilts, including when $t < s$. Applying Lemma A.1 to the inner rectangle with both window constants $aw_n/3$ and the outer radial window in (142) gives

$$\sum_{x \in \delta + K} f_v(x) = \frac{r^n}{\det K}(eE_n/4)^{n/2} e^{\pm O(\ell_n)}. \tag{143}$$

Replacing $E_n$ by $E$ changes this by $e^{O(1)}$.

For the collision sum (92), take $w \ne 0$ and put $\xi = \|w\|^2/\|v\|^2 \ge 1$. A nonzero summand puts both $M$ and $M + w$ in the midpoint ball. The parallelogram identity gives

$$\mathcal{E}_r(M + w/2) \le \frac{E_n - c_n\xi}{4} + aw_n. \tag{144}$$

It follows that $c_n\xi \le E_n + 4aw_n$ and hence $\xi \le 16a_{\max}/c < R_0^2 - 1$, so all contributing difference radii are below $R_0$.

Choose a fixed $0 < b_0 < E/32$ throughout the range. When $(E_n - c_n\xi)/4 \ge b_0$, the ball majorant on $K$ is uniform. Put $H_n = E_n + 4aw_n$. The contribution of each such difference to the nonzero part of $\theta \sum_t F_t^2$ is at most

$$V\left(\frac{e(H_n - c_n\xi)}{4}\right)^{n/2} e^{O(\ell_n)}.$$

Extend this nonnegative majorant to all nonzero differences. Lemma 3.5 bounds their sum by

$$V\,2^{\kappa_{\mathrm{pack}}n}\left(\frac{eH_n^2}{16c_n}\right)^{n/2}e^{O(\ell_n)}.$$

Since $H_n = E + O(w_n + 1/n)$, dividing by the mean squared gives

$$V^{-1}2^{\kappa_{\mathrm{pack}}n}(ec_n)^{-n/2}e^{O(\ell_n)} = 2^{-h_n+O(\ell_n)}.$$

Here $ec = 2^{2\kappa_{\mathrm{pack}}-1}$ and $c_n = c + O(1/n)$ give the cancellation.

For the remaining boundary range, use one ball majorant of normalized squared radius $b_0 + aw_n$ and count all differences with $\xi \le E_n/c_n + O(w_n)$. Before division by the mean squared, its exponential base is at most $eb_0E_n/c_n + O(w_n)$. The interior bound has base $eE_n^2/(16c_n)$, so the choice of $b_0$ leaves a fixed exponential margin. This proves the nonzero collision bound used in (90).

### B.4.2 Filtered differences and the affine work bound

Continue the proof of Proposition 9.4 after applying the uniform intensity bounds on the good event $\mathcal{G}$. Write $g = \mathbb{P}[\mathcal{G}] = 1 - o(1)$. Nonnegativity extends the remaining geometric sums to the original uniform affine experiment at a factor $g^{-1}$. For any $w \in L \setminus \{0\}$, the affine membership average is exactly $\theta$ times the ambient sum over $\delta + K$.

Put $\xi = \|w\|^2/\|v\|^2$ and $d = c_n\xi$. For endpoint energies $R, S \in a(1 \pm w_n)$, their correlation is $t = (R + S - d)/(2\sqrt{RS})$. A positive common cap requires $1 + t > 2\tau^2$, hence $d \le Q + O(w_n)$ and $4a - d \ge 4a\tau^2 - O(w_n) > 0$. The midpoint Gaussian majorant is therefore uniform on the whole contributing range. Also $\xi \ge 1$ bounds the correlation away from one, so the Gram determinant of the two endpoint directions has a fixed positive margin.

Let $B_\pm = 4a(1 \pm w_n)$. The common cap radius factor satisfies

$$1 - \frac{2\tau^2}{1+t} = \frac{(\sqrt{R}+\sqrt{S})^2 - d - 4\tau^2\sqrt{RS}}{(\sqrt{R}+\sqrt{S})^2 - d} \le \frac{Q + 4a(1+\tau^2)w_n - d}{B_- - d}.$$

The midpoint ball has normalized squared radius at most $(B_+ - d)/4$, and throughout the contributing range,

$$B_- - d \ge 4a\tau^2 - O(w_n) > 0, \qquad \frac{B_+ - d}{B_- - d} = 1 + O(w_n).$$

Choose $Q_n = Q + Cw_n + 1/n$, with fixed $C > 4a(1+\tau^2) + 1$ uniformly over the energy range. The inflated cap numerator is at least $1/n$ on the support, and its squared radius factor is uniformly below one. The spherical wedge estimate in dimension $n + O(\log n)$ can therefore use exponent $n/2$ at cost $e^{O(\ell_n)}$.

Multiplying by the midpoint ball estimate cancels the denominator up to $(1 + O(w_n))^{n/2} = e^{O(nw_n)}$, and conditional independence of the lists gives

$$\mathbb{E}[\text{reports of difference } w \mid \mathcal{G}] \le g^{-1}e^{2D_n}M_fA_n^2V\left(\frac{e(Q_n - c_n\xi)}{4}\right)^{n/2}e^{O(\ell_n)}. \tag{145}$$

Its support has $Q_n - c_n\xi > 0$. The endpoint shells also imply $c_n\xi \le 4a(1 + w_n) \le 8a_{\max}$, so $\xi \le 16a_{\max}/c$ and the fixed $R_0$ covers the counting bins. Flatness on $K$ is the smoothing input here, replacing the internal condition $8a\tau^2 > 1$.

Apply Lemma 3.5 with $H = Q_n$ and $c = c_n$ to (145). Together with the target normalization in (102), it gives

$$2^{\kappa_{\mathrm{pack}} n} M_f \left( \frac{Q_n^2}{4 c_n E} \right)^{n/2} 2^{o(n)}$$

as an upper bound for the sum over all nonzero differences. Since $Q_n = Q + O(w_n)$ and $c_n = c + O(1/n)$, the second identity in (102) cancels $2^{\kappa_{\mathrm{pack}} n}$. The total rate is therefore $K_\tau(u)$. If all amplitude and code remainders are $O(b_n)$, the same calculation keeps every logarithmic remainder $O(b_n)$.

For target coverage, choose the code padding to dominate $O(\ell_n)$ and $H_{d_{\mathrm{pc}}}$ from Proposition 6.1. The target correlation $u + O(w_n)$ stays inside the prescribed active constraint range. Equal values have already been charged to point incidences in the main proof, completing the work and coverage bounds.

## B.5 Uniform concentration for fixed banks

The relative approximation theorem [HS11, Theorem 2.11(ii)] controls all queries without enumerating their lattice supports. Let $P_M$ be the empirical law of $M$ iid draws from $P$, and let the positive superlevel sets of a class $\mathcal{F}$ of $[0,1]$ functions have VC dimension $d$. For $0 < \epsilon \le 1/2$, $0 < p \le 1$, and $0 < \delta < 1$, the sample size condition

$$M \ge \frac{C}{\epsilon^2 p} \left( d \log \frac{2}{p} + \log \frac{2}{\delta} \right)$$

gives $|P_M A - PA| \le \epsilon \max\{PA, p\}$ simultaneously on these sets, with failure at most $\delta$. The finite-set formulation extends to our finitely supported laws by approximating their masses by rational frequencies and replicating the atoms. As the replicated population grows, uniform sampling without replacement converges to the required iid law. Integrating over $\{f > t\}$, $0 < t < 1$, yields

$$|P_M f - Pf| \le \epsilon (Pf + p) \qquad (f \in \mathcal{F}). \tag{146}$$

The failure symbol belongs to no positive superlevel set.

**A common test for the two bank queries.** For a forward family $0 \le f_x \le 1$ with $Pf_x \le H$ and a reverse family $0 \le g_z \le 1$ with $Pg_z = \bar{\mu}(1 \pm \eta_{\mathrm{mid}})$, $\eta_{\mathrm{mid}} \le 1/4$, apply (146) at error $1/16$ and floor $p_0 = \min\{1, H, 3\bar{\mu}/4\}$. In both applications, $\log(1/p_0) = O(n)$ and the shatter coefficients satisfy $S(N) \le (CN)^{Cn^2}$, giving $d = O(n^2 \log(n+2))$. Thus $Mp_0 \ge 2^{\Omega(b_n)}$ permits $\delta = \exp[-cMp_0] = \exp[-2^{\Omega(b_n)}]$ for a small absolute $c > 0$. The forward sum is at most $2MH$. If $B_C(z) = h(z) U_r \sum_i g_z(y_i)$, the reverse sum lies between $\frac{1}{2} M U_r \bar{\mu} h(z)$ and $2 M U_r \bar{\mu} h(z)$. The two kernels below supply $H, U_r, \bar{\mu}$ and verify these conditions.

**Lemma B.1** (Simultaneously good banks)**.** *For the node parameters in Section 11.2, the bank satisfies (122) for all admissible queries except with probability* $\exp[-2^{\Omega(b_n)}]$. *This holds uniformly in the input width, lattice determinant, and affine shift.*

*Proof.* Condition on the node's lattice, scales, and routing shift. The full right trial law is $q(y) = \lambda \rho_r(y)/V_r$ on its source shell, with the remaining mass at †. For forward queries use $f_x(y) = K(x,y)$ and the mean envelope $Pf_x \le H_r = \lambda \mathcal{B}_r 2^{b_{\mathrm{deg}}}$ from Lemma 5.1. For reverse queries let

$$t_z(y) = \frac{q(z+y) K(z+y, y)}{\rho_{\gamma r}(z)}, \qquad g_z(y) = t_z(y)/U_r, \quad U_r = (\lambda/V_r) e^{n w_n}.$$

Completion of squares gives

$$t_z(y) = \frac{\lambda}{V_r} e^{A_r R_-^2 - n\mathcal{E}_r(z+y)/2 + n\mathcal{E}_r(z)/4}.$$

The source and output shells put the exponent after $\lambda/V_r$ at most $na_1 w_n/2 \le nw_n$, so $g_z \in [0,1]$. The midpoint identity gives

$$Pg_z = \bar{\mu}_r(1 \pm \eta_{\mathrm{mid}}), \qquad \bar{\mu}_r = \lambda e^{A_r R_-^2 - nw_n} 2^{-n/2} = \lambda 2^{-[J(a_0)+R(u)]n - O(nw_n)}.$$

For either family, a positive superlevel set is given by a constant number of quadratic inequalities in $y$ (take logarithms of the positive weight). Lifting to degree-two monomials gives $S(N) \le (CN)^{Cn^2}$. Queries are restricted to prescribed cosets: then $z+y$ automatically lies in the source coset. Varying coset membership is not included among the quadratic predicates. The failure symbol contributes to no positive superlevel.

Use Appendix B.5 with $H = H_r$ and the displayed reverse mean and envelope. The recursion gives $\lambda \ge 2^{-o(b_n)}$, and

$$P(u) - R(u) = -\tfrac{1}{2}\log_2 \gamma^2 \ge 0, \qquad M \min\{H_r, 3\bar{\mu}_r/4\} \ge 2^{C_a b_n - o(b_n)}.$$

Since $MU_r\bar{\mu}_r = b_*$, the common test proves (122). □

## B.6 Reverse-query work with a shared bank

Conditional on bank $C$, the proposal before the second rejection has mass $B_C(z)/A_{\max}$. On a good bank this is at most four times the prescribed output subprobability in (123). Applying this domination at fixed $C$ removes the bank-dependent proposal weight before averaging the query cost $g(z) + C(H_z)$. We can then drop the good bank indicator and bound the bank contribution by its original first moment. The sum retains terms involving the selected record itself, so it also pays for inserting that record before reverse search.

Write $q = -z$, so query and bank have energies $a_1, a_0$. After averaging the local half grid shift, sum first over the bank and then over the queries. On a correlation bin near $v$, the triangular change

$$(q, y) \longmapsto \left(q, y - v\sqrt{a_0/a_1}\, q\right)$$

admits Gaussian widths $r\sqrt{a_1}$ on $\Gamma$ and $r\sqrt{a_0(1-v^2)}$ on $\Gamma/2$. Their squared normalized radii add to $2 + O(w_n)$ on the shells. Shifted theta upper bounds apply for each fixed $q$ even though this change need not preserve the lattice. Writing $g = r^2/s^2$, we have

$$ga_1 \ge 2a_0 > 1, \qquad C_{\mathrm{rev}} = 4ga_0 \ge C_0 := 4/(1-u^2).$$

The half grid factor $2^{-n}$ cancels its covolume. Converting Gaussian densities to query and bank list sizes $N_Q, N_C$ uses $e^{(1-a_0)n} a_0^n = 2^{-2J(a_0)n}$. Consequently the angular spectrum is

$$N_Q N_C\, 2^{-P(v)n + O(nw_n + \log n)} \max\{1, [C_{\mathrm{rev}}(1-v^2)]^{-n/2}\}. \tag{147}$$

For uniform control near parallel directions, fix $v_0 < 1$ with $C_0(1-v_0^2) < 1$. Throughout $v \ge v_0$, use the single residual $y - v_0\sqrt{a_0/a_1}\, q$, of energy at most $a_0(1-v_0^2) + O(w_n)$. Sequential Gaussian sums bound that region by the angular factor

$$\max\{(1-v_0^2)^{n/2}, C_{\mathrm{rev}}^{-n/2}\} \le C_0^{-n/2}$$

times $2^{O(nw_n+\log n)}$. This includes parallel directions and repeated values. It uses the fixed lower bound $C_0$, rather than requiring a cutoff subcritical for arbitrarily large actual $C_{\mathrm{rev}}$. Away from this region, replace the penalty in (147) by the weaker $C_0$ if needed.

Set $\tau = u$ and use code rate $K_{\mathrm{rev}}(u)$ from Section 11.2. The identity $P(v) + K_\tau(v) \geq 2P(\tau)$ gives a report rate in the ordinary region at most $K_{\mathrm{rev}}$, since both list size exponents are $P(u)$. In the subcritical and near-parallel regions the single-cap bound gives

$$K_{\mathrm{rev}} + 2P(u) - \tfrac{1}{2}\log_2 C_0 - P(u) = K_{\mathrm{rev}} - 1.$$

For common cap slack at most $1/16$, (141) instead gives $K_{\mathrm{rev}} + 2P(u) - 2$, also smaller. The other correlations lie in fixed compact bins. Point incidences and restricted traversal have rate at most $K_{\mathrm{rev}}$ by Proposition 6.1. Independent fresh reverse codes are introduced only after ideal proposals have been fixed, so complete coverage and Lemma C.1 apply.

True forward dictionaries have rate $P(u)$. For reverse dictionaries evaluate the spectrum at $v_{\mathrm{rev}}$, where $P(v_{\mathrm{rev}}) = R(u)$, giving $2P(u) - R(u)$. The full bank charging argument above includes contributions from repeated IDs. Sort by canonical pair IDs before adding weights, and use independent coins for left trials. Depth-first list storage, sparse labels, dictionaries, and code descriptors therefore fit the limits in Section 11.2.

## B.7 Independent extensions and the terminal batch merge

We prove the additional statements used in Theorems 11.2 and 11.3. Throughout, write $u_E = 1/2 + 1/k_n$, $c_E = (1+u_E)^{-1}$, $\ell_E = P(u_E)$, and $R_E = \frac{1}{2}\log_2(2c_E)$, with $k_n$ as in Lemma 7.2. Thus $c_E \to 2/3$ and $\ell_E, R_E \to \beta_{\mathrm{G}}$. The existing endpoint forest gives $M$ iid reference labels with law on vector outputs

$$q_E(x) = \frac{\lambda_E}{V_R(\Gamma)}\rho_R(x)\mathbf{1}_{\mathcal{E}_R(x)\in c_E(1\pm w_n)}, \qquad M = \left\lceil 2^{[\ell_E + J(c_E)]n + C_M b_n} \right\rceil, \quad \lambda_E = 2^{-1-t(b_{\deg}+2)}. \tag{148}$$

The affine support is the prescribed coset. The forest has common depth $t = O(k_n \log n)$, and $-\log_2 \lambda_E = o(b_n)$. The reference coupling is the one in Section 11.1.

### B.7.1 Extensions and exact mixing coefficients

Use the independent extensions of Lemma 7.1, with $\iota_j = \max\{0, \log_2(\sigma/(\alpha s_j))\}$ and exact index $p_j$. Here $s_j/\sigma$ stays in a fixed compact positive range, so the lemma's failure bound applies to every bin. Algorithm 6 specifies the balancing and label consumption. The remaining calculation keeps the finite rates needed at full size.

For ordinary bins take $R_j^2 = \sigma^2 z_j / c_E$ and $s_j^2 = R_j^2/2$. Radial correction anchored at endpoint $\zeta_j = z_j + O(w_n)$ gives

$$\kappa_j = \frac{\lambda_E}{V_{R_j}(\Gamma_j)} e^{n\zeta_j(1-c_E/z_j)/2}, \qquad M\kappa_j V_\sigma(L) = 2^{[\ell_E - \iota_j + J(z_j)]n + C_M b_n - o(b_n) + O(nw_n) - 3h_n}.$$

Taking the endpoint limit gives (124). In particular, using the innermost index for every bin would lose the gain. On $[1/2, 2]$ with $\alpha = 3/2$, the second branch of that minimum exceeds $\beta_{\mathrm{G}}$, since $\log(27/16) > 1/2$. The first branch attains its minimum at $z = 1$, where the exact finite rate $\ell_E > \beta_{\mathrm{G}}$ and padding give the required surplus.

### B.7.2 The new terminal kernel and its degree

Fix $u = 3/5$, $c = 5/8$, $d = 1/2$, and $\gamma^2 = 4/5$. For an inner bin $z_j \in [5/12, 1/2]$, set

$$r_o^2 = \sigma^2 z_j/c, \qquad r^2 = r_o^2/\gamma^2 = 2\sigma^2 z_j, \qquad R^2 = (c/c_E)r^2, \qquad s_j^2 = R^2/2. \tag{149}$$

For output coset $a + \Gamma_j$, route a uniform $e \in (\Gamma_j/2)/\Gamma_j$ to the child cosets $a/2 + e + \Gamma_j$ and $-a/2 + e + \Gamma_j$. Their endpoint shells are precisely $\mathcal{E}_r(x) \in c(1 \pm w_n)$. For sufficiently large $n$, $R < r$. If $U$ is the maximum squared norm on this shell, a radial acceptance $\exp[\pi(\|x\|^2 - U)(R^{-2} - r^{-2})]$ converts each child to

$$q(x) = \frac{\bar{\lambda}}{V_r(\Gamma_j)} \rho_r(x) \mathbf{1}_{\mathcal{E}_r(x) \in c(1 \pm w_n)}, \qquad \bar{\lambda} = \lambda_E (r/R)^n e^{-\pi U(R^{-2} - r^{-2})}. \tag{150}$$

Thus $\log_2 \bar{\lambda} = \log_2 \lambda_E + [J(c) - J(c_E)]n + O(nw_n)$. The coefficient can exceed one, since it multiplies a restricted Gaussian density. The acceptance probability above already guarantees a subprobability. With $A = M\bar{\lambda}$, $\log_2 A = [\ell_E + J(c)]n + O(b_n)$, and the stored child list still has rate $\beta_{\mathrm{G}}$.

For $z = x - y$, the eligibility test is

$$\mathcal{E}_r(z) \in d(1 \pm w_n).$$

Put $a_r = \pi(\gamma^{-2} - 1/2)/r^2$, $L_-^2 = nr^2 d(1 - w_n)/(2\pi)$, and $K(x, y) = \mathbf{1}_{\mathrm{Eligible}} e^{-a_r(\|x-y\|^2 - L_-^2)}$. Writing $m = (x + y)/2$, completion of squares gives

$$\rho_r(x)\rho_r(y)K(x, y) = e^{a_r L_-^2} \rho_{r/\sqrt{2}}(m) \rho_{\gamma r}(z) \mathbf{1}_{\mathrm{Eligible}}.$$

For the fixed right bank define $A_C, B_C$ as in Section 11.2, using the child law (150). The forward residual width $r\sqrt{1-u}$ can be smaller than $s_j$. Its cost is

$$H_n = \max\{1, (s_j/(r\sqrt{1-u}))^n\}, \quad \mathcal{B}_r = e^{a_r L_-^2} (\gamma/\sqrt{2})^n, \quad D = 8A\mathcal{B}_r H_n 2^{C_D(nw_n + \log(n+2))}. \tag{151}$$

Indeed for $t < s_j$, shift maximality and monotonicity give

$$\rho_t(b + \Gamma_j) \le \rho_t(\Gamma_j) \le \rho_{s_j}(\Gamma_j) \le (1 + 2^{-h_n}) V_t(\Gamma_j)(s_j/t)^n.$$

This yields $\mathbb{E}[A_C(x)] \le 2A\mathcal{B}_r e^{O(nw_n)} H_n$. Now $s_j^2/r^2 \to 15/32$, so

$$n^{-1} \log_2 H_n \to \tfrac{1}{2} \log_2(75/64) = P(3/5) - \beta_{\mathrm{G}}, \qquad \mathcal{B}_r = 2^{-[J(c) + P(3/5)]n + O(nw_n)}.$$

The fixed degree exponent cancels. At finite $n$ the remaining term is

$$\log_2 D = (\ell_E - R_E)n + O(b_n) = o(n). \tag{152}$$

In particular $O(n/k_n)$ need not be $O(b_n)$ and is retained in the work budgets.

### B.7.3 Uniform banks and the complete output law

Set

$$b_0 = \frac{M\bar{\lambda}^2}{V_r^2} e^{a_r L_-^2} V_{r/\sqrt{2}}.$$

We claim that, except with probability $\exp[-2^{\Omega(b_n)}]$,

$$A_C(x) \le D, \qquad \tfrac{1}{2} b_0 \rho_{\gamma r}(z) \le B_C(z) \le 2b_0 \rho_{\gamma r}(z) \tag{153}$$

simultaneously for all admissible queries. For the forward family take $H = D/(4M)$ in Appendix B.5. The expected sum is at most $D/4$, and positive padding gives $MH = D/4 \geq 2^{\Omega(b_n)}$. For the reverse class use

$$t_z(y) = \frac{q(z+y)K(z+y,y)}{\rho_{\gamma r}(z)}, \quad g_z = t_z/U_r, \quad U_r = (\bar{\lambda}/V_r)e^{Cnw_n}.$$

On the eligible support,

$$t_z(y) = \frac{\bar{\lambda}}{V_r} e^{a_r L_-^2 - 2\pi\|y+z/2\|^2/r^2 + \pi\|y\|^2/r^2}.$$

Equivalently the exponent is $a_r L_-^2 - n\mathcal{E}_r(z+y)/2 + n\mathcal{E}_r(z)/4$. The source and output shells bound it by $ndw_n/2$, so $0 \leq g_z \leq 1$ for fixed large $C$. The midpoint width satisfies $(r^2/2)/s_j^2 \to 16/15 > 1$. Since $c = 1/2 + d/4$, the endpoint region contains the same inner midpoint window as in Appendix A.6. Radial and linear tilts retain smoothing, and comparison with the full midpoint mass shows

$$Pg_z = \bar{\mu}_r(1 \pm \eta_{\mathrm{mid}}), \quad \eta_{\mathrm{mid}} = O(2^{-h_n}) + e^{-\Omega(nw_n^2)}, \quad M\bar{\mu}_r = 2^{(\ell_E - R_T)n + O(b_n)}, \quad R_T = \tfrac{1}{2}\log_2(5/4).$$

Here $\beta_{\mathrm{G}} - R_T > 0.0465$. Both positive superlevel classes use a constant number of quadratic predicates, so $S(N) \leq (CN)^{Cn^2}$. The forward and reverse sample floors satisfy Appendix B.5, with $MU_r\bar{\mu}_r = b_0$, proving (153).

Apply Lemma 11.1 with the normalizer $D$ in (151), $c = b_0/2$, and the output Gaussian weight restricted to its shell and coset. The bank bounds give $C_0 = 4$ and the following probability mass on vector outputs

$$q_o(z) = \frac{b_0}{2D}\rho_{\gamma r}(z)\mathbf{1}_{z \in a+\Gamma_j,\ \mathcal{E}_r(z) \in d(1\pm w_n)}. \tag{154}$$

The reference replacement in Section 11.1 gives iid labels. The known output coefficient is $\lambda_o = b_0 V_{\gamma r}/(2D)$.

### B.7.4 Output counts at finite dimension

The live count rate tends to $q_T = 2\beta_{\mathrm{G}} - P(3/5) = \log_2(16/15)$. More precisely, putting $d_E = n^{-1}\log_2 H_n = P(3/5) - R_E$ gives

$$\log_2(M\lambda_o) - J(c)n = [\ell_E + R_E - P(3/5)]n + \Theta(b_n).$$

The padding is positive after choosing $C_M$ large enough. The function $P(u) + \frac{1}{2}\log_2(2/(1+u))$ is increasing for $u \geq 1/2$, so this finite rate is at least $q_T$. The actual extension rate is

$$\iota_E(z) = \max\{0, \tfrac{1}{2}\log_2(c_E/(\alpha^2 cz))\}.$$

At $\alpha = 3/2$ it is no larger than its limit, since $c_E \leq 2/3$. Restricting to $a + L$ and changing width to $\sigma$ consequently gives (126), with errors $O(1/k_n) + O(b_n/n)$. Its minimum is at an endpoint: it increases up to $64/135$ and then decreases up to $1/2$. At those endpoints,

$$F_T(5/12) = J(5/12) > \beta_{\mathrm{G}}, \qquad F_T(1/2) = \log_2(16/15) + J(1/2) > \beta_{\mathrm{G}}.$$

These strict margins absorb all finite parameter changes.

### B.7.5 Forward and reverse search

The two stored child lists have size exponent $\beta_{\mathrm{G}}$. Write $g = r^2/s_j^2 \to 32/15$ and $c_0 = 2gc \to 8/3$. The angular spectrum from Section 6 is

$$2^{[2\beta_{\mathrm{G}} - P(v)]n + o(n)} \max\{1, [c_0(1-v)]^{-n/2}\}.$$

Use forward filter threshold $\tau = 1/2$, target correlation $3/5$, and define $K_\tau(v) = -\frac{1}{2}\log_2(1 - 2\tau^2/(1+v))$. The point and the two report bounds all have rate at most

$$T_F = K_{1/2}(3/5) = \tfrac{1}{2}\log_2(16/11) < 0.2703 < \alpha_{\mathrm{G}}.$$

The forward true edge dictionary has rate $q_T < \beta_{\mathrm{G}}$. The common cap compact region has midpoint smoothing margin $g\,2c(1+v) > 4/3 + o(1)$ whenever $v > -1/2$. The fixed cutoff and single-cap bounds in Appendix B.6 handle its parallel and degenerate limits.

The reverse query depends on the bank. Conditional on a good bank, its proposal density is $B_C(z)/D \le 4q_o(z)$. For any nonnegative cost, apply this domination before averaging over the bank and removing its good event indicator. The right side does not depend on the bank, so its original first moment controls the bank cost, including contributions from the same record. Fresh reverse codes are sampled only after ideal proposals are fixed.

The reverse target correlation is $v_T = 1/\sqrt{5}$. Its query list has size exponent $q_T$ and the bank has size exponent $\beta_{\mathrm{G}}$. The triangular Gaussian transformation in Appendix B.6, with different query and bank sizes retained, gives

$$N_Q N_C 2^{-P(v)n + o(n)} \max\{1, [C_{\mathrm{rev}}(1-v^2)]^{-n/2}\}, \qquad C_{\mathrm{rev}} = 4gc \to 16/3.$$

The query width has squared ratio $gd \to 16/15 > 1$ to smoothing. The residual sum is on $\Gamma_j/2$, which explains the factor four. Use reverse threshold $1/3$ and put $K_R = K_{1/3}(1/\sqrt{5})$. Point, ordinary report, and subcritical report rates are at most

$$\begin{aligned} K_R + \beta_{\mathrm{G}} - P(1/3) &< 0.2429, \\ K_R + q_T + \beta_{\mathrm{G}} - 2P(1/3) &< 0.2510, \\ K_R + q_T + \beta_{\mathrm{G}} - \tfrac{1}{2}\log_2(16/3) - P(1/3) &< -0.8715. \end{aligned}$$

The ordinary bound uses $P(v) + K_\tau(v) \ge 2P(\tau)$. The single-cap bound controls the subcritical and parallel regions. The exact wedge estimate at very small common cap slack gives the smaller rate $K_R + q_T + \beta_{\mathrm{G}} - 2$. Thus every reverse rate is below $\alpha_{\mathrm{G}}$ with a fixed margin. The reverse dictionary rate is $q_T + \beta_{\mathrm{G}} - R_T < 0.1397 < \beta_{\mathrm{G}}$. The raw label rate tends to $\beta_{\mathrm{G}} + J(2/3) = 1/2 - 1/(6\log 2) < 0.2596 < \alpha_{\mathrm{G}}$. Sparse labels, record IDs, both dictionaries, and depth first forest storage fit $2^{\beta_{\mathrm{G}} n + o(n)}$ bits. Restricted Cartesian traversal adds no larger rate. Polynomially many bins and $2^{o(n)}$ forest nodes preserve these bounds.

### B.7.6 Tails, joint error, and finite realization

The lower tail cannot be tilted to width $\sigma\sqrt{5/12}$, which may be below smoothing. Instead,

$$\mathbb{E}_{D_{a+L,\sigma}}[e^{-(\alpha^2-1)n\mathcal{E}_\sigma(X)/2}] = \frac{\rho_{\sigma/\alpha}(a+L)}{\rho_\sigma(a+L)} \le 3\alpha^{-n}.$$

At $\alpha = 3/2$ this gives lower-tail exponent $H_{3/2}(5/12) = [\log(9/4) - 25/48]/(2\log 2)$. Direct substitution gives $J(5/12), H_{3/2}(5/12), J(2) > \beta_{\mathrm{G}}$ and $F_T(1/2) > \beta_{\mathrm{G}}$, with fixed positive margins.

Thus the annulus tail $\delta_{\mathrm{ann}}$ obeys $2^{\beta_{\mathrm{G}} n}\delta_{\mathrm{ann}} = 2^{-\Omega(n)}$. The expected mixture success count is at least $2^{\beta_{\mathrm{G}} n+\Omega(b_n)}$ and its shortfall is doubly exponentially unlikely. The full joint error is bounded by

$$\mathrm{poly}(n)2^{-2h_n+O(1)} + 2^{o(n)}e^{-2^{\Omega(b_n)}} + 2^{-\Omega(b_n)} + 2^{-\Omega(n)} + \lfloor 2^{\beta_{\mathrm{G}} n}\rfloor\delta_{\mathrm{ann}}. \tag{155}$$

Here $\log(\text{forest size}) = O(k_n \log n) = o(b_n)$. The leading error is $\exp[-\Omega(n/\log^2(n+2))]$.

For finite execution use rational squared radii and half-open bins, with normalized endpoints within $2^{-Cn}$ of the stated values. Use exact shell comparisons, including at boundary lattice points, and the exact finite rates, in particular $(\ell_E - R_E)n = O(n/k_n)$, before enlarging work and storage limits by $2^{Cb_n}$. First-overflow bounds cost $2^{-\Omega(b_n)}$. The positive coefficient formulas and whole batch numerical budget are implemented in Appendix C.3.4. Their loss is charged once in (155). Fixed-dimensional sampling handles the remaining small ranks.

## B.8 Sampling estimates

We verify the scale and search margins used in Sections 7 and 10.

### B.8.1 The smoothing extension for DGS

Put $\eta = \eta_{1/2}(L)$ and $A(t) = \sum_{y\in L^*} e^{-\pi t^2\|y\|^2}$. For $t \ge \eta$, $A(t) \le 3/2$. For $t < \eta$, Poisson summation and $\rho_t(L) \le \rho_\eta(L)$ give $A(t) \le (\eta/t)^n A(\eta)$. Hence

$$A(s) \le \tfrac{3}{2}\, 2^{\iota n}. \tag{156}$$

The extension has index $p$. Every dual vector outside $pL^*$ survives its random modular constraint with probability $(p^{n-1}-1)/(p^n-1) \le 1/p$. For $a_y = e^{-\pi\eta^2\|y\|^2}$, $y \ne 0$, one has $\sum_y a_y \le 1/2$, so the forced mass is at most $\sum_y a_y^T \le 2^{-T}$, where $T = (ps/\eta)^2$. Since $s/\eta$ has a fixed positive lower bound,

$$\mathbb{E}_\Gamma[\rho_{1/s}(\Gamma^* \setminus \{0\})] \le A(s)/p + 2^{-T} \le 2^{-3h_n+O(1)}.$$

Markov's inequality proves Lemma 7.1.

### B.8.2 The target shell tail

For $Y \sim D_{a+L,\sigma}$, put $T = 2\pi\|Y\|^2/\sigma^2$. At $|\theta| = O(w_n)$, both widths below are above smoothing, and

$$\mathbb{E}[e^{\theta T}] = \frac{\rho_{\sigma/\sqrt{1-2\theta}}(a+L)}{\rho_\sigma(a+L)} \le 3(1-2\theta)^{-n/2}.$$

Chernoff bounds with $\theta = \pm\Theta(w_n)$ give $\mathbb{P}[|T/n - 1| > w_n] \le 6e^{-c_\alpha n w_n^2}$. The target shell mass divided by $V_\sigma(L)$ is consequently between two positive constants for large $n$.

### B.8.3 Ordinary forward search

For the ordinary shell geometry, the smallest input scale has $r^2/s^2 = 2/\gamma^2$ and free output denominator one. The two smoothing factors are

$$c_0(1-v), \quad c_0(1+v), \qquad c_0 = \frac{2}{1-u^2}. \tag{157}$$

A nonempty common cap has $v > 2u^2 - 1$, hence midpoint factor at least $4u^2/(1-u^2) \ge 4/3$. Multiplying child intensities before averaging the shift gives, by Lemma 6.2, the candidate count

$$2^{[2\ell(u)-P(v)]n+O(b_n)} \max\{1, [c_0(1-v)]^{-n/2}\}. \tag{158}$$

Earlier scales improve these factors. Substituting $\ell(u) = P(u)$ and $\tau = u$ into (61) makes both report rates equal to $K(u)$, as is the point incidence rate. The endpoint bins also fit: with $\Delta = 1/100$, $c_0\Delta < 1$ and the terminal cap rate is $K(u) + \frac{1}{2}\log_2(1-\Delta/2) < K(u)$, including equal values. When $1 - 2u^2/(1+v) \le 1/16$, (141) gives $K(u) + 2\ell(u) - 2 < K(u)$. The remaining $O(n)$ bins are compact. Restricted traversal costs at most $2N_{\mathrm{cap}}B_{\mathrm{pc}} + 2nZ_AB_{\mathrm{pc}}$. The bound $\log_2 B_{\mathrm{pc}} = o(b_n)$ therefore gives, per call within the prescribed limits,

$$\mathbb{E}[\text{forward search work}] \le 2^{K(u)n+O(b_n)}. \tag{159}$$

## B.9 Shortest-vector scales and search margins

### B.9.1 Scale selection and initial pools

Apply (13) to $L$ with $c = 9$, and set $\widehat{Q} = \|v\|^2$ and $\Delta_n = D_9(n)$. Then $\widehat{Q}/\Delta_n^2 \le \lambda_1(L)^2 \le \widehat{Q}$. The grid

$$Q_i = \frac{\widehat{Q}}{\Delta_n^2}(1+1/n)^i \tag{160}$$

up to $\widehat{Q}$ has $O(n\log n)$ branches, one with $Q_i \le \lambda_1(L)^2 < (1+1/n)Q_i$. On each branch choose the rational squared width of Proposition 3.6, put $Y = \max\{100, \lceil 2^{h_n}\det L/s^n\rceil\}$, and search for a prime in $[Y, 2Y]$ using Appendix C.1.6. Skip the scale if the search exhausts its budget. On the correct branch, the Gaussian mass bound underlying Proposition 3.6 gives $2^{h_n}\det L/s^n \ge 2^{h_n}/\operatorname{poly}(n) > 100$ for large $n$, so the chosen prime has the required scale.

Within a trial, apply Proposition 3.2 to $\Gamma$ with $k = \lfloor n/10\rfloor$, at cost $2^{0.2n+o(n)}$. On $\mathcal{G}$, (14) gives $G \le \operatorname{poly}(n)s$. A fixed depth $t_{\mathrm{s}} = \lceil C_t\log(n+2)\rceil$ therefore makes $R = r\gamma_{\mathrm{s}}^{-t_{\mathrm{s}}}$ a valid starting width. Each leaf uses $2M$ corrected trials, where $M = \lceil 2^{m_{\mathrm{raw}}n + C_{\mathrm{leaf}}b_n}\rceil$, and $b_{\mathrm{deg}} = \lceil 10\ell_n\rceil$. Its root padding is

$$\sigma_0 = \log_2 M - m_{\mathrm{raw}}n - t_{\mathrm{s}}(h_n + b_{\mathrm{deg}}) = C_{\mathrm{leaf}}b_n - O((h_n+\ell_n)\log n) = \Theta(b_n).$$

These predetermined schedules supply the cooling calls in Algorithms 3 and 4, with the conditional independence specified in Section 10.2.

### B.9.2 List sizes and pair search

At height $j = 1, \dots, 4$, put $t_j = 2^{5-j}$. The ancestor average (57) becomes

$$\mathbb{E}_{\delta,\text{ancestors}}\left[\sum_{x\in b_j+\Gamma} f(x)\right] = t_j^{-n}\sum_{x\in\varepsilon z/2^{4-j}+\Gamma/t_j} f(x), \qquad f \ge 0, \tag{161}$$

where $\varepsilon \in \{1,-1\}$ depends on the path. Fixing $\Gamma, z$, the path, and scale leaves the local and ancestor shifts to be averaged. Conditional on those shifts, child intensities are each at most $A_{j-1}e^{O(b_n)}$ by Theorem 8.2, and multiply by private independence. Cooling gives the analogous Gaussian bounds. The shell majorants on $\Gamma/t_j$ have squared smoothing factors $266.2912, 54.464, 11.456, 2.11296$. Their covolumes cancel $t_j^{-n}$, proving (116).

For input rate $\beta$, threshold $\tau$, free denominator $t$, and $g = r_{\mathrm{in}}^2/s^2$, let $c_0 = 2t^2ga$. Equation (61) gives

$$K_{\mathrm{code}} + \max\left\{2(\beta - P(\tau)), 2\beta + \tfrac{1}{2}\log_2\frac{2(1-\tau^2)}{c_0}\right\}. \tag{162}$$

In cooling, $g \ge \bar{g} := 2/\gamma_{\mathrm{s}}^2$ and $t = 1$. Use the fixed comparison factor $c_{0,\mathrm{s}} = 2\bar{g}a_0 = 2/(1-u_{\mathrm{s}}^2)$ throughout. Equivalently, at input width $r_{\mathrm{in}}$ use the auxiliary smoothing scale $s_{\mathrm{eff}} = r_{\mathrm{in}}/\sqrt{\bar{g}} \ge s$.

The lattice is smooth there by monotonicity, so this weaker comparison gives uniform upper bounds at every cooling depth. Substitution of $\tau = u_{\mathrm{s}}$ gives (118). Point work is smaller, and the comparison midpoint factor exceeds 2.10 wherever the common cap is nonempty.

At full height $j$, use $g = 2$, $a = a_{j-1}$, $t = t_j$, and $\beta = L_{j-1}$. Substituting (106) gives $8a_{j-1}\tau_j^2 > 1.008$ and $c_0 \geq 716/125 > 4$. Since $P(\tau_j) = L_{j-1}$, both point and pair rates are at most $K_j$. For full layers use $c_{0,j} = 4t_j^2 a_{j-1}$, and for cooling use $c_{0,\mathrm{s}}$ above. With these comparison factors, $\Delta = 1/2000$ gives $c_{0,j}\Delta, c_{0,\mathrm{s}}\Delta \leq 1024/3175 < 1$ and midpoint factor at correlation $1 - \Delta$ at least $715821/250000 > 1$. The common cap region of measure factor at most $1/16$ costs $K_{\mathrm{code}} + 2\beta - 2 < K_{\mathrm{code}}$. The other $O(n)$ bins are compact. Window perturbations cost $O(b_n)$.

The terminal bound is Proposition 9.4 with Equations (112) and (119), averaging the common coset. Deduplicated equal values cost point work. Restricted traversal has rate at most $\max\{\beta, \beta + T_{\mathrm{point}} - s_{\mathrm{live}}\}$. Every input has $\beta \leq L_{\max} < s_{\mathrm{live}}$ with gap greater than $38 \cdot 10^{-6}$, so traversal fits the point rate. The factor $B_{\mathrm{pc}}$ is subexponential, and both preprocessing stages have exponent $0.2 < s_{\mathrm{live}}$.

# C Finite realization

We first give the common data representations, random primitives, and search coupling, then choose the budgets for single samples, batches, and SVP. The application analyses supply their list and work rates. Predetermined caps turn these estimates into bounds on every execution, and one coupling accounts for all implementation errors.

## C.1 Finite data and random tools

### C.1.1 Computational model

We use a randomized word RAM with independent uniform bits and word size $w = \Theta(n + \log(\mathcal{B} + f + 2))$, where $\mathcal{B}$ is the binary input length and $f$ is an accuracy parameter when present. Time counts word operations, including multiword arithmetic and numerical evaluation. Space counts bits. Polynomial input factors are suppressed, and polynomial factors in $f(n) \leq \mathrm{poly}(n)$ are included in $2^{o(n)}$. The same leading exponents hold with bit costs for random-access operations.

All resource bounds hold on every execution. Lists, work, and random trials have predetermined caps. On failure, shifted DGS returns its affine offset, centered DGS returns zero, and SVP retains its best nonzero lattice vector. A failed batch returns copies of its fallback.

### C.1.2 Sublinear budgets

The default schedules for SVP and batches are

$$\begin{aligned} L_{\log} &= \lceil \log_2(n+2) \rceil, & w_n &= L_{\log}^{-4}, \\ h_n &= \lceil n/L_{\log}^2 \rceil, & \ell_n &= nw_n + L_{\log}, \\ \chi_n &= \lceil n/L_{\log}^{3/4} \rceil, & b_n &= n/\sqrt{L_{\log}}. \end{aligned} \tag{163}$$

Round resource budgets upward to integers. These schedules satisfy $\ell_n \ll h_n \ll \chi_n \ll b_n \ll n$ and $(h_n + \ell_n)\log n = o(\chi_n)$. Extra records and resource allowances use a fixed multiple of $b_n$, while $\chi_n$ controls SVP repetitions and $nw_n^2$ is the Gaussian tail exponent.

### C.1.3 Choosing the constants

Choose the fixed padding constants in the order: reference list sizes and depths, moment bounds, coverage, storage, then work and call caps. Each choice may depend on earlier choices. In particular, the ideal reference sequence is fixed before the implementation budgets.

### C.1.4 Fixed dimensions

For each fixed rank under the smoothing promise, an LLL basis has $G \le C_n \sigma$ by transference. The GPV proposal can then be corrected by rejection with probability $\prod_i \theta_{s_i}(h_i)/\theta_{s_i}(0)$, where $s_i = \sigma/\|b_i^*\| \ge 1/C_n$. This produces the target Gaussian and has acceptance bounded below by a positive rank-dependent constant. Standard theta evaluation and Appendix C.1.7 give polynomial bit cost, and logarithmically many attempts achieve any prescribed error. Batch size is constant at fixed rank, and fixed-dimensional SVP uses [MV13]. Thus the finitely many ranks excluded by the asymptotic estimates affect only constants.

### C.1.5 Finite representations

Bases, lattice coordinates, quotient labels, and affine offsets are stored as exact rational data. Shell supports bound vector norms, and each difference increases coordinate length by only $O(1)$ bits. Together with the polynomial basis bounds in Proposition 3.2, this gives polynomial encoding length throughout the forest.

Each record has a node ID and a local index. Leaves use their trial indices, isolated merges use left child order, and full merges use lexicographic order of the distinct child ID pairs. Equal vector values from different pairs remain distinct records. These choices make the output IDs independent of report order, so complete coverage gives the same IDs and subsequent pair coins in the ideal and implemented forests. Batch labels continue to denote positions in the original trial stream.

For DGS bin $j$, the squared scale $r_j^2 = \sigma^2 z_j/a_0$ is rational, as are all later scales obtained from the exact ratio $\gamma^2$. For SVP, approximate the initial $s^2$ with relative error $O(n^{-2})$ and then use the same exact ratios. Thus the Gaussian identities remain exact along each sequence of widths. Membership and squared norms use rational comparisons, while the shell predicates and spherical search are treated in Appendix C.2.1.

### C.1.6 Numerical and random primitives

Elementary functions and Bernoulli and categorical draws use standard polynomial bit algorithms. After canceling common factors, positive weights are evaluated to relative error $\epsilon < 1$, giving normalized TV error at most $\epsilon/(1-\epsilon)$. A known lower bound on total mass also permits absolute weight error, as used in the radius mixture.

For rational $x \ge 1$, write $\operatorname{clog}_2 x = \min\{k \in \mathbb{Z}_{\ge 0} : 2^k \ge x\}$. Uniform integers use binary rejection. Primes in $[Y, 2Y]$, $Y \ge 100$, use uniform candidates, primality testing [AKS04], and the density bound of [RS62, (3.8)]. For at most $N$ requests, fixed polynomial trial budgets give total failure at most $\delta$, with logarithmic dependence on $N/\delta$ and polynomial dependence on the input length.

For adaptive calls we use sequential coupling: conditional errors $\delta_i$ at each common preceding history give complete-execution error at most $\sum_i \delta_i$. Resource caps and the fallbacks of Appendix C.1.1 keep the same budgets on every path.

### C.1.7 Scalar Gaussian sampling

Let $D_{S,a,c}$ have mass proportional to $e^{-\pi(k-c)^2/a^2}$ on $S \subseteq \mathbb{Z}$. For rational $a^2 > 0$, $c$, and $0 < \delta \leq 1$, bounded interval rejection samples within TV distance $\delta$ of $D_{\mathbb{Z},a,c}$ and $D_{\mathbb{Z}\setminus\{0\},a,0}$. Normalize rejection by a point of maximum weight. With $T = \Theta(1 + \log(1/\delta))$, an interval of radius $\Theta(\max\{1, a\}T)$ has discarded mass $e^{-\Omega(T^2)}$ and acceptance probability $\Omega(1/T)$, uniformly in the width. For the centered nonzero law, omit zero and normalize by the weight at 1. A cap of $O(T^2)$ trials and numerical error at most $\delta/2$ give polynomial work, space, and random bit use in the input length and $\log(1/\delta)$ on every execution.

For GPV, apply the uniform guarantee to the rational squared widths and centers supplied by Gram–Schmidt. Sequential coupling sums the scalar errors over the prescribed call bound.

### C.1.8 Weighted selection

Standard weighted reservoir sampling implements Theorem 4.7 in one pass with one stored sample: at each positive-weight record $k$, replace the current sample with probability $b(X_k)/\sum_{i\leq k} b(X_i)$. Return the prescribed fallback if the total weight is zero. The finite positive-weight computations use Appendix C.1.6 and share the execution's numerical error allowance.

## C.2 Search and the execution coupling

### C.2.1 Finite precision spherical search

We prove a finite implementation of the fixed list product code interface in Proposition 6.1. Lattice operations and eligibility tests remain exact, so the implemented shell supports agree with those of the ideal process throughout the coupling.

Shell predicates reduce to comparisons $A\pi \leq B$ with rational $A, B$. The irrationality measure bound $\mu(\pi) < 8$ of [ZZ20] gives a separation of $2^{-O(L)}$ at input length $L$, so interval evaluation decides them in polynomial bit time. For the initial Gaussian proposal, the sufficient condition $R^2/G^2 \geq \operatorname{clog}_2(8n)$ is rational.

Write the ideal code as $\mathcal{C} = Q(\mathcal{C}_1 \times \cdots \times \mathcal{C}_m)$, with Haar $Q$ and block vectors of norm $m^{-1/2}$. For nonzero $x$ let $s_{j,k}(x) = \langle (Q^T\widehat{x})_j, c_{j,k}\rangle$. Choose dyadic upper scores and a dyadic lower threshold satisfying

$$s_{j,k} \leq \overline{s}_{j,k} \leq s_{j,k} + \eta/m, \qquad \tau - \eta \leq \underline{\tau} \leq \tau.$$

Decoding the dyadic scores with threshold $\underline{\tau}$ gives

$$\mathcal{I}_\tau(x) \subseteq \widetilde{\mathcal{I}}(x) \subseteq \mathcal{I}_{\tau-2\eta}(x), \tag{164}$$

because the total score error is at most $\eta$. The Cartesian decoder of [BDGL16, Section 5] and its restricted traversal [GFH26b, Appendix D] apply directly to these additive scores, with the resource bounds of Proposition 6.1.

**Lemma C.1** (Finite code coupling)**.** *Consider finitely many calls whose input lists obey prescribed size limits and are defined independently of their fresh codes. If call $a$ uses $R_a$ codes with $M_a$ implicit filters and list caps $N_{a,L}, N_{a,R}$, put*

$$U_{\mathrm{cap}} = \sum_a R_a M_a (N_{a,L} + N_{a,R}).$$

*For requested failure $\delta$, finite bit code descriptions and scores can be generated so that all incidence sets on these lists agree with the ideal ones except with probability at most $\delta$. Their precision and*

*the cost per explicitly generated scalar are polynomial in the rational input length,* $n$, $\log(1/\delta)$, *and* $\log(U_{\mathrm{cap}}+2)$.

*Proof.* For $d \geq 3$, a coordinate of a uniform point on $S^{d-1}$ has density

$$\frac{\Gamma(d/2)}{\sqrt{\pi}\,\Gamma((d-1)/2)}(1-t^2)^{(d-3)/2} \leq C\sqrt{d}.$$

Each fixed product code word has this uniform marginal under the outer Haar rotation. Therefore a width $2\eta$ boundary band has probability $O(\sqrt{n}\,\eta)$. By (164), a union bound over all potential point-filter incidences costs at most $C\sqrt{n}\,U_{\mathrm{cap}}\eta$. Choose

$$\log_2(1/\eta) \geq \log_2(1/\delta) + \log_2(U_{\mathrm{cap}}+2) + C\log_2(n+2) + C.$$

Generate the rotation by Gaussian QR and the block points by normalizing Gaussian vectors. Couple their finite approximations in distance to the continuous code. For $D$ generated scalars, Gaussian tails and the bound $\Pr[\chi_k < \zeta] = O(\zeta)$ for $k \geq 1$ exclude large coordinates and small normalization denominators at total cost $\delta/4$. On the remaining event, standard perturbation bounds show that

$$O(\log(1/\eta) + n\log(nD/\delta))$$

working bits, in addition to input encodings, suffice for score error $\eta$. Each code uses one fixed description throughout its traversal. Combining generation and boundary errors proves the claim. □

**Finite canonical pair coins.** Let $U_{\mathrm{pair}} = \sum_a N_{a,L}N_{a,R}$ bound potential tagged pairs. Use a standard pairwise independent family of uniform $p$-bit integers $J_e$, indexed by the ordered child ID pairs. Its seed has polynomial length and is drawn after the children are completed, independently of the search codes. Repeated reports reuse the same $J_e$. For the coupling, add independent $V_e \sim U[0,1)$ and put $U_e = (J_e + V_e)/2^p$. This gives the ideal pairwise independent uniform family. Approximating the activation thresholds to $p$ bits therefore costs $O(U_{\mathrm{pair}}2^{-p})$ over all potential pairs.

### C.2.2 Product code search on lists with repeated values

Write $C_{d_{\mathrm{pc}}}$ and $W_{d_{\mathrm{pc}}}$ as in Section 6. The source construction uses zero padding to dimension $d_{\mathrm{pc}} = n + O(\log n) = m_{\mathrm{pc}}b_{\mathrm{pc}}$, where $m_{\mathrm{pc}} = \lceil \log_2 d_{\mathrm{pc}} \rceil$. An integer block code size $B_{\mathrm{pc}}$ gives $M_f = B_{\mathrm{pc}}^{m_{\mathrm{pc}}}$ filters. The fixed list work, space, and coverage interface is Proposition 6.1. We extend it to records with repeated values and explain how complete coverage reconstructs the merges used in the dynamics.

Filter membership depends only on vector values, so the incidence and resource bounds of Proposition 6.1 count records and pairs with their multiplicities. Each eligible tagged pair with unequal values has the coverage probability of its values. Thus the same search interface applies to record lists.

For target correlation $u_0$, choose the least integer $B_{\mathrm{pc}} \geq 2$ with $m_{\mathrm{pc}}\log_2 B_{\mathrm{pc}} \geq K_\tau(u_0)n + C_{\mathrm{code}}b_n$. This gives (56). The padding dominates the coverage threshold and shell perturbations. All eligible correlations $u_0 + O(w_n)$ have uniform active constraint margins. In particular, one code misses any fixed eligible record edge with probability at most $1/2$ for large $n$.

### C.2.3 Canonical pairs and complete coverage

A record has a node ID and a local integer index as in Appendix C.1.5. At a cooling node, conditional on the completed children and their IDs, draw a fresh seed for a pairwise independent activation family indexed by ordered child ID pairs. This seed is independent of all earlier nodes and search codes. Repeated reports reuse the same coin. Distinct coins have the pairwise independence required by the dynamics, and the activated graph is independent of the search codes. A compact seed recovers the same decision on every repeated report without storing the rejected pairs. Resampling on each report would make the activation probability depend on how often search finds the pair.

A cooling call records each endpoint's unique activated neighbor, or marks it as having several distinct neighbors. Repeated reports leave these states unchanged. A final scan in left ID order emits the mutually unique pairs, using $O(N_L + N_R)$ space and work beyond the reports. Complete coverage is essential: it recovers every competitor and makes mutual uniqueness agree with isolation in the ideal graph.

A full call deduplicates eligible pair IDs in a dictionary and emits their differences in lexicographic ID order. The dictionary uses space proportional to the output count. Distinct pairs with equal output values remain distinct records.

Integer block sizes change $\log_2 M_f$ by only $O(\log n)$, which is absorbed by the padding. The active constraints, including the unequal-threshold case of [GFH26b, Appendix D], hold uniformly over the parameter ranges used here.

Run $r_{\mathrm{cov}} = \lceil C_{\mathrm{cov}} n \rceil$ independent codes per intermediate call with $C_{\mathrm{cov}} = 10$. All codes share the canonical state and, at a cooling node, the pair coin seed. Their implicit descriptors are generated and traversed sequentially, using $\mathrm{poly}(n) B_{\mathrm{pc}}$ auxiliary space. The miss probability of a fixed eligible activated edge is at most $2^{-C_{\mathrm{cov}} n}$. Charge code generation, state updates, dictionaries, and the final endpoint traversal to the operation cap. At the SVP root, deduplicate values and stream reports without a pair dictionary. If a target exists, choose one canonical witness from the ideal lists before drawing top codes.

**Lemma C.2.** *Let an ideal forest on a prescribed finite tree and an event $\mathcal{G}$ be independent of all search codes. Couple it to an implementation with list limits $N_v$, work caps $W_a^{\mathrm{cap}}$, and $R_a$ codes at call $a$. Conditional on ideal inputs, shared pair coin seeds, and the environment independent of the codes, these codes are independent with their original law. For each fixed input within its list size limits, suppose one code misses a fixed eligible edge with probability at most $q_a < 1$.*

*Let $\epsilon_{\mathrm{num}}$ bound the additional finite arithmetic and random bit coupling loss on histories within the resource limits, with zero in the exact arithmetic case. Let $\Xi_v$ be the ideal lists and $W_a$ the work of a complete search on ideal inputs, set to zero when an input exceeds its limit. Work includes code generation, the report and endpoint scans, canonicalization, state updates, and dictionary operations. Then*

$$\mathbb{P}[\mathcal{G} \text{ and disagreement}] \le \epsilon_{\mathrm{leaf}} + \epsilon_{\mathrm{num}} + \sum_v \frac{\mathbb{E}[\mathbf{1}_{\mathcal{G}} |\Xi_v|]}{N_v} + \sum_a \frac{\mathbb{E}[\mathbf{1}_{\mathcal{G}} W_a]}{W_a^{\mathrm{cap}}} + \sum_{a \text{ internal}} N_{a,L} N_{a,R} q_a^{R_a} + \epsilon_{\mathrm{top}}. \tag{165}$$

*Here disagreement means a mismatch of the reconstructed internal lists or sampling outputs, or failure to report the chosen terminal witness when it exists. The first error bounds leaf coupling. The numerical allowance includes code generation, boundary incidences, finite coins, and readout for the word RAM implementation, as in Proposition 6.3. For a terminal witness fixed from the ideal lists before the top codes, take $\epsilon_{\mathrm{top}} = q_{\mathrm{top}}^{R_{\mathrm{top}}}$, and include missing that witness in disagreement.*

*Otherwise set* $\epsilon_{\mathrm{top}} = 0$. *For at most* $N_{\mathrm{raw}}$ *adaptive raw calls of uniform statistical error* $\epsilon_{\mathrm{raw}}$, *one may use* $\epsilon_{\mathrm{leaf}} = N_{\mathrm{raw}}\epsilon_{\mathrm{raw}}$.

*Proof.* Share leaves, shifts, and canonical pair coins, and expose calls from the leaves upward. Until the first disagreement, the inputs coincide. For inputs within the list size limits, a union bound gives incomplete coverage probability $N_{a,L}N_{a,R}q_a^{R_a}$. Complete coverage reconstructs the competitor graph and hence the ideal output. At the top, apply the bound to the canonical witness alone.

Markov's inequality on ideal list sizes and complete search work, with the indicator of $\mathcal{G}$, gives the two expectation sums. Sequential coupling bounds the adaptive raw calls uniformly over their histories. All moments are taken on the ideal process before conditioning on coverage, resource success, or existence of a terminal witness. □

#### C.2.4 Proof of the finite realization interface

The key comparison is with a forest defined independently of every search code. Complete intermediate coverage recovers its activated graphs, while terminal SVP needs only one fixed target witness. By Lemma C.2 and the budgets in Appendix C.3, search and resource failures cost

$$\delta_{\mathrm{search}} + \delta_{\mathrm{resource}} \le 2^{-\Omega(n)} + 2^{-\Omega(b_n)}$$

per core execution. The single sample budgets are given in Appendix C.3.2. For SVP, Appendix C.3.5 makes this loss $o(2^{-B_n})$, where $2^{-B_n}$ is the ideal success probability.

For the finite arithmetic, let $N$ bound all scalar decisions and let $U_{\mathrm{cap}}, U_{\mathrm{pair}}$ bound potential incidences and pairs. Their logarithms are $O(n) + o(n) + O(\log(\mathcal{B}+2))$ per core execution, with the logarithmic repetition count added for amplification. Allocate total error $2^{-\lambda}$ over these decisions using Appendix C.1.6 and Lemma C.1. The positive-weight formulas in Appendix C.3.4 and the scalar sampler require polynomial precision in $\mathcal{B}, n, \lambda, \log N$. The stable record IDs then allow sequential coupling of the record lists, pair decisions, and sampling decisions over the forest. For sampling, this gives

$$\mathbb{P}[\mathcal{G} \text{ and } X_{\mathrm{finite}} \neq X_{\mathrm{ideal}}] \le \delta_{\mathrm{search}} + \delta_{\mathrm{resource}} + 2^{-\lambda}.$$

Here $\mathcal{G}$ is the application's good environment event. For SVP, on $\mathcal{G} \cap E_v$ choose one representation of the fixed shortest vector $v$ before drawing the terminal codes. Unless one of the charged failures occurs, the implementation reports this witness and hence finds a shortest vector. This gives the SVP success bound in Proposition 6.3. Polynomial arithmetic costs and the decoder's $\mathrm{poly}(n)B_{\mathrm{pc}}$ workspace preserve the leading exponents.

Take $\lambda = Cn$ for a single sample or an entire Gaussian batch. Combining the implementation loss with the analytic error bounds proves the corresponding claims of Proposition 6.3. For SVP, first subtract the per-trial loss from $2^{-B_n}$ and amplify as in Section 10.5, then couple the complete execution with $\lambda = Cn\log_2(n+2)$. The centered reduction uses the query and accuracy bounds of Appendix D.8.

### C.3 Application budgets

The applications use one common cap rule. For an expected list size or work bound $2^{rn+O(b_n)}$, choose a cap $2^{rn+Cb_n}$ with a larger fixed $C$, following Appendix C.1.3. Use the finite rates in these caps, including any $O(n/k_n)$ correction, before adding the padding. The table gives the rates and storage needed by each construction.

| Construction | Leaf trials | Live and dictionary rates | Search rates |
|---|---|---|---|
| Single-sample cooling | $2M_{\mathrm{leaf}}$ | $P(u_n)$ | $K(u_n)$ |
| SVP cooling and full layers | $2M_{\mathrm{leaf}}$ | $\max_j L_j < s_{\mathrm{live}}$ | $T_{\mathrm{s}}, K_1, \dots, K_5$ |
| Ordinary batch | $M$ | $P(u)$, $2P(u) - R(u)$ | $K(u), K_{\mathrm{rev}}(u)$ |
| Terminal batch | $M$ | $\beta_{\mathrm{G}}$, $q_T + \beta_{\mathrm{G}} - R_T$ | $T_F$ and the reverse rates of Appendix B.7 |

The SVP quantities are defined in Section 10, ordinary batch rates are in Section 11.2, and terminal rates in Appendix B.7. Full-layer dictionaries contain output records. Batch dictionaries contain true eligible pairs. Single-sample rates are defined in Section 7. Its cooling tree and final weighted scan use the separate schedule and coupling in Appendix C.3.2.

For a prescribed forest, the number of raw trials is at most its number of leaves times the largest trial count in the table. The number of leaves is $2^{O(t)}$ per bin, with polynomially many bins and $t = O(\log n \log\log n)$. Include all SVP repetitions and scale branches in the global call bound $N_{\mathrm{trial}}$. Each trial uses $n$ scalar draws and one acceptance coin. Assign their total error from the requested whole-execution allowance.

Evaluate forests depth first, streaming leaf proposals and candidate reports. Retain only the active lists, endpoint states, dictionaries, and search workspace described in the table. Use $\lceil 10n \rceil$ independent codes per merge, processed sequentially. The one-code miss probability is at most $1/2$ on inputs within the prescribed list limits, and all list rates are below $1/2$. A union over the potential pairs and forest nodes therefore gives probability of incomplete coverage $2^{-\Omega(n)}$. For isolation, the final endpoint traversal of Appendix C.2.3 recovers the output after the report scan.

Markov's inequality on ideal list sizes and complete search work gives resource loss $2^{-\Omega(b_n)}$ over a core forest. The coupling in Lemma C.2 combines this loss with coverage and finite arithmetic. Failures use the returns specified in Appendix C.1.1. Subsequent SVP trials retain their fresh randomness and fixed budgets.

### C.3.1 Cooling schedules and their implementation

Apply Proposition 3.2 with $k = \lfloor n/10 \rfloor$ to obtain a basis with $G \le \mathrm{poly}(n)s$. Choose the first scheduled width $R = r_0\gamma^{-t}$ satisfying Lemma 3.3. The depth is $O(\log n)$ for fixed $u$, and is bounded by Lemma 7.2 otherwise. Enforce the depth cap before constructing a forest.

Implement Algorithms 1 and 3 using the report scan and endpoint traversal of Appendix C.2.3 for isolation, and its ordered pair dictionary for full merges. These procedures preserve the ideal local IDs on complete coverage and use the common call budgets of Appendix C.3.

### C.3.2 Single-sample schedules and execution

**Shell width and sublinear budgets.** Fix the rational $0 < \delta \le \min\{1, \xi\}$ used in Section 7, and set $L_{\log} = \lceil \log_2(n+2) \rceil$. For this algorithm use

$$w_n = \frac{1}{\lceil L_{\log}^{1+\delta/2} \rceil}, \qquad h_n = \left\lceil \frac{n}{L_{\log}^{1+\delta/3}} \right\rceil, \qquad b_n = \left\lceil \frac{n}{L_{\log}^{\delta/6}} \right\rceil, \qquad \ell_n = nw_n + L_{\log}. \tag{166}$$

For $t = O(L_{\log} \log L_{\log})$ and $b_{\deg} = \lceil 10\ell_n \rceil$, these choices give

$$\ell_n = o(h_n), \quad t(h_n + b_{\deg} + 2) = o(b_n), \quad t = o(nw_n^2), \quad nw_n^2 = o(h_n), \quad \frac{n}{\log n} = o(b_n), \quad b_n = o(n). \tag{167}$$

Also $k_n w_n \to 0$, so the endpoint contraction and shell margins hold uniformly. The Gaussian estimates use only shrinking shells, diverging $nw_n^2$, and the fixed smoothing margins of the relevant tilts. Their logarithmic shell losses are $O(\ell_n)$ for these widths as well. The cooling error is therefore $e^{-\Omega(nw_n^2)} = \exp[-\Omega(n/\log^{2+\delta} n)]$.

**Initialization and expected counts.** Choose the prime and extension by Lemma 7.1, using the bounded prime procedure of Appendix C.1.6, with total prime and uniform-integer failure allowance $2^{-Cn}$. On a good extension, basis preprocessing with block size $\lfloor n/10 \rfloor$ gives $G \le \mathrm{poly}(n)s$ within rate $1/5$. Choose a deterministic depth $t = O(k_n \log n)$ large enough for this bound and check the resulting starting width before generating records. The good-extension bound ensures that the check succeeds there. Give every leaf $2M_{\mathrm{leaf}}$ corrected trials with $\varrho = m(u_n)$ in (48). The total number of leaves is $2^t$, and (167) makes the root coefficient $2^{m(u_n)n+\Theta(b_n)}$ as required in (69).

**Work, space, and coupling.** Applying the common cap rule of Appendix C.3, choose at each cooling node

$$N_{\mathrm{cap}} = \left\lceil 2^{P(u_n)n + C_{\mathrm{list}} b_n} \right\rceil, \qquad W_{\mathrm{cap}} = \left\lceil 2^{K(u_n)n + C_{\mathrm{work}} b_n} \right\rceil.$$

The list first moment and (159), averaged over shifts on each good extension, supply the expectations in Lemma C.2. The common schedules of Appendix C.3 give resource and coverage losses $2^{-\Omega(b_n)}$ and $2^{-\Omega(n)}$. Lists and endpoint states use rate $P(u_n)$, and code descriptors fit because $\log_2 B_{\mathrm{pc}} = O(n/\log n) = o(b_n)$. Keep the exact finite rates $P(u_n)$, $m(u_n)$, and $K(u_n)$ in every cap before adding padding.

Finish with the reservoir scan of Appendix C.1.8 using (63). On the retained shell, every positive weight lies in $[e^{-O(nw_n)}, 1]$, so the relative weight computations of Appendix C.1.6 use polynomial precision. There are $2^{O(t)} 2M_{\mathrm{leaf}}$ leaf proposals, giving logarithmic decision count $O(n) + o(n)$ and total numerical error $2^{-Cn}$ under the common finite realization.

### C.3.3 Finite batch execution

Implement Algorithms 5 and 6 with the forward and reverse searches of Section 11.2, using the finite decoder of Appendix C.2.1. Define the complete ideal forward proposal list before drawing the independent reverse codes.

Dictionaries store true eligible pairs keyed by record IDs, counting each pair once and using ID order for accumulation and partner selection. Each reverse query includes its selected forward partner, which has positive weight. Thus its denominator is positive even if search does not report that partner again. The forward and reverse dictionary rates are $P(u)$ and $2P(u) - R(u)$, with the smaller terminal rates in Appendix B.7. Store trial streams as sparse lists of live records and labels, so scanning all $M$ labels takes $O(M)$ work without an array of $M$ failure symbols.

With the common budgets of Appendix C.3, Appendix C.2.4 couples the entire batch with numerical error $2^{-Cn}$. This is the implementation allowance used in the joint error bound of Section 11.2.

### C.3.4 Evaluation of the batch weights

The leaf correction uses the Poisson series for $\theta_s(h)/s$, whose widths have a fixed positive lower bound. The forward weights $K$ lie in $[e^{-O(nw_n)}, 1]$ when nonzero. For reverse sums, cancel the

factors $\lambda/V_r$ and $\rho_{\gamma r}(z)$ before evaluation:

$$\frac{B_C(z)}{(\lambda/V_r)\rho_{\gamma r}(z)} = \sum_{i:\,\text{eligible}} e^{A_r R_-^2 - 2\pi\|y_i+z/2\|^2/r^2+\pi\|y_i\|^2/r^2}.$$

Every nonzero summand lies in $e^{\pm O(nw_n)}$, and the selected forward partner supplies a positive term. The same cancellation applies to $b_*\rho_{\gamma r}/B_C$, so relative approximation of positive sums suffices for both reverse operations.

For bin selection, put $\vartheta_j = \kappa_j V_\sigma(L)$, using each bin's own determinant in $\kappa_j$. The common volume cancels, leaving weights $\vartheta_j^{-1}$. Since $\log\vartheta_j = O(n) + o(n)$, these weights also admit relative approximation with polynomial precision. The terminal branch uses the same cancellation and positive sums. All these decisions share the complete execution allowance in Appendix C.2.4.

### C.3.5 Finite shortest vector execution

We implement the fixed trial schedule of Algorithm 4. On each scale branch use the bounded initialization of Appendix B.9.1. On the correct branch, prime search exhausts its budget with probability at most $2^{-\Omega(n)}$. Conditional on successful prime selection, the starting width and depth checks pass on the good smoothing event $\mathcal{G}$.

**Resource loss compared with target success.** The implementation loss must be smaller than the ideal target probability $2^{-B_n}$. Use the list rates and storage of (120) with the common caps in Appendix C.3. The ideal list and search bounds in Section 10.4, averaged over the common coset on good superlattices, give the per-trial Markov bound

$$\mathbb{P}[\mathcal{G} \text{ and some list or work cap fails}] \le 2^{-\Omega(b_n)}. \tag{168}$$

The per-trial coverage and terminal witness losses are $2^{-\Omega(n)}$.

Preprocessing uses a $2^{0.2n+O(b_n)}$ cap. Use the deterministic raw call budget

$$N_{\text{call}} \le \text{poly}(n)2^{\chi_n}2^{m_{\text{raw}}n+C_N b_n}$$

and error $2^{-3n}/N_{\text{call}}$ per corrected trial. Sequential coupling changes the complete transcript by at most $2^{-3n}$. Consider first trials with ideal corrected leaves. Their resource and coverage losses on good environments sum to

$$\delta_{\text{trial}} \le 2^{-\Omega(b_n)} + 2^{-\Omega(n)} = o(2^{-B_n}).$$

Subtracting this loss leaves success at least $2^{-B_n-1}$ per trial, uniformly on the correct branch. Independent repetition therefore amplifies success before the Gaussian call coupling adds its $2^{-3n}$ error over the complete execution.

**Complete execution.** The uniform trial guarantee permits the amplification in Section 10.5. Prime exhaustion and corrected-leaf coupling cost $2^{-\Omega(n)} + 2^{-3n}$ per complete execution, leaving success at least $2/3$ for large $n$. After the final polynomial repetitions, finite realization with $\lambda = Cn\log_2(n+2)$ preserves failure $2^{-\Omega(n\log n)}$ and both resource exponents.

# D Centered DGS by a mixture over radii

We follow the primitive direction sparsification approach of Stephens-Davidowitz [SD16, Section 4]. A multiplicative radius grid controls the change in Gaussian direction weight throughout the truncated range. Treating layers with few primitive directions separately allows the same mixture to work at every width.

We first establish the direction weight comparison, then use the sparsification facts of [SD16] to construct the radius mixture. The proof separates the multiplicative comparison of direction weights from the additive errors due to truncation and failed draws. These steps prove Theorem 12.6. We then replace its exact SVP oracle by Theorem 1.3 to obtain Corollary 1.4. All polynomial bounds include the numerical accuracy parameters and depend on $\sigma/\lambda_1(L)$ through its logarithm and the binary input length. Prime and uniform integer draws are first treated as ideal primitives, with their finite implementation and numerical allowances given in Appendices D.7 and D.8.

## D.1 Gaussian direction weights and the radius grid

Let $\mathcal{P}(L)$ contain one representative of each primitive direction modulo sign. For $r > 0$, put

$$h_\sigma(r) = 2\sum_{k\geq 1} e^{-\pi k^2 r^2/\sigma^2}.$$

Every nonzero vector is uniquely $zy$, with $y \in \mathcal{P}(L)$ and $z \in \mathbb{Z}\setminus\{0\}$. Therefore

$$\rho_\sigma(L) = 1 + \sum_{y\in\mathcal{P}(L)} h_\sigma(\|y\|). \tag{169}$$

After selecting a direction with this weight, the conditional coefficient law is $D_{\mathbb{Z}\setminus\{0\},\sigma/\|y\|,0}$.

**Why the grid must control relative changes.** The mixture uses samplers for directions inside successive balls. For exact counts and uniform sampling, weighting the sampler at $r_i$ by its count times $h_\sigma(r_i) - h_\sigma(r_{i+1})$ cancels the count. Give the last layer its residual weight $h_\sigma(r_{\text{last}})$. A direction first included at $r_j$ then receives total weight $h_\sigma(r_j)$ by telescoping. Thus the grid must make $h_\sigma(r_j)$ close to the desired $h_\sigma(\|y\|)$ multiplicatively.

Normalize $\sigma = 1$ and write $h = h_1$. The proof of [SD16, Theorem 4.6] uses $r_i^2 = \lambda_1(L)^2 + i\Delta$, where $\Delta = 1/(100nf)$, and claims

$$h(r_i) \leq h(\|y\|) \leq (1+1/f)^{1/10} h(r_i) \quad (r_{i-1} \leq \|y\| \leq r_i). \tag{170}$$

The difficulty is the behavior near zero. Poisson summation gives $h(r) = r^{-1}(1+h(1/r)) - 1$, so $h(r) \sim r^{-1}$ as $r \downarrow 0$. An additive step in $r^2$ can therefore cause an arbitrarily large relative change in direction weight. For example, if $\lambda_1(L) = a$ and a primitive direction has length $2a$, then for $a^2 \ll \Delta$ it first appears at $r_1 = \sqrt{a^2+\Delta}$, while $h(2a)/h(r_1) \to \infty$ as $a \to 0$. The required comparison concerns the mass of all nonzero multiples of the direction. Short directions have many multiples with appreciable Gaussian weight, which produces this reciprocal growth.

**A numerical counterexample.** Take $\sigma = 1$, $n = f = 20$, and

$$a = 2^{-20}, \qquad L = \operatorname{diag}(a, 2a, 32, \ldots, 32)\mathbb{Z}^{20}, \qquad y = 2ae_2.$$

Here $\lambda_1(L) = a$ and $y$ is primitive, with $r_0 = a < \|y\| = 2a < r_1 = \sqrt{a^2 + 1/40000}$. Poisson summation and comparison with the Gaussian integral give $r^{-1} - 1 \le h(r) \le r^{-1}$, hence

$$h(\|y\|) \ge 2^{19} - 1 = 524287, \qquad (1 + 1/f)^{1/10} h(r_1) < (21/20)^{1/10} \cdot 200 < 201.$$

Thus (170) fails by more than a factor of 2600, even for this rational diagonal lattice. Even with exact counts and uniform direction sampling, the mixture assigns this direction the telescoped weight $h(r_1)$. The missing control therefore lies in the radius discretization.

**The multiplicative grid.** We use multiplicative steps in squared radius instead. They control the ratio of adjacent radii near zero, while a step size chosen for the largest radius also controls the Gaussian decay at the other end of the range. Lemma D.1 proves the resulting weight comparison. Sparsification then supplies the counts and direction samplers, and a thin interval estimate bounds the contribution of layers with too few directions. After selecting a direction, a scalar Gaussian chooses its nonzero multiple.

Fix

$$\begin{aligned} F &= \max\{f, 2\}, \qquad \varepsilon = 2^{-f}, \qquad \eta = \frac{1}{1000F}, \qquad \kappa = \varepsilon/4, \\ Q &= n + F + 20, \qquad G = \lceil 10^4 FQ \rceil, \qquad T = \sigma\sqrt{Q}, \qquad R_{\mathrm{g}} = \frac{T}{\sqrt{1 - G^{-2}}}. \end{aligned} \tag{171}$$

The radius $T$ separates the region with a multiplicative guarantee from the Gaussian tail. The grid extends to $R_{\mathrm{g}}$, and its entire range is available to the sampler.

The centered case of [SD16, Lemma 2.5], at radius $T = \sigma\sqrt{Q}$, gives

$$\Pr_{X \sim D_{L,\sigma}} [\|X\| \ge T] < (2\pi e Q/n)^{n/2} e^{-\pi Q} < e^{-Q} < \varepsilon/64. \tag{172}$$

This also bounds the mass of points whose primitive direction is longer than $T$.

For $n = 1$ use Appendix C.1.7 on the input generator, with error $\kappa/8$. Suppose henceforth $n \ge 2$. One SVP call gives a shortest direction $y_0$ and the rational number $q_0 = \|y_0\|^2/\sigma^2$. If $q_0 \ge Q$, output zero. Equation (172) proves the required guarantee for this branch. Otherwise define $D = 16000FQ$ and

$$q_{i+1} = \min\{(1 + 1/D) q_i, q_{\mathrm{g}}\}, \qquad q_{\mathrm{g}} = \frac{Q}{1 - G^{-2}}. \tag{173}$$

Stop at the first $J$ with $q_J = q_{\mathrm{g}}$, and write $r_i = \sigma\sqrt{q_i}$. The implementation stores only rational squared radii. All length comparisons are squared comparisons.

**Lemma D.1** (Adjacent direction weights)**.** *The grid is strictly increasing and satisfies*

$$1 \le \frac{h_\sigma(r_i)}{h_\sigma(r_{i+1})} \le 1 + \eta, \qquad J \le 1 + (D + 1) \log(q_{\mathrm{g}}/q_0).$$

*Its points have polynomial encoding length in the input and $f$.*

*Proof.* For $H_0(t) = 2\sum_{k \ge 1} e^{-\pi k^2 t^2}$,

$$M(t) := -\frac{d \log H_0(t)}{d \log t} = 2\pi t^2 \frac{\sum_{k \ge 1} k^2 e^{-\pi k^2 t^2}}{\sum_{k \ge 1} e^{-\pi k^2 t^2}} \le 13 \max\{1, t^2\}. \tag{174}$$

Indeed, for $t \le 1/2$, Poisson summation for $1 + H_0$ gives $-tH_0'(t) \le 1 + H_0(t)$ and $H_0(t) \ge t^{-1} - 1 \ge 1$. For $t \ge 1/2$, divide the numerator by its first term and use $k^2 \le 4^{k-1}$ and $k^2 - 1 \ge 3(k - 1)$. The

resulting geometric series is at most $(1-4e^{-3\pi/4})^{-1}<2$, proving the bound. Throughout the grid, $q\le q_{\mathrm{g}}\le 4Q/3$ and $q_{\mathrm{g}}>1$. Integrating the derivative across any step, including the truncated last step, gives

$$\log\frac{h_\sigma(r_i)}{h_\sigma(r_{i+1})}\le\frac{13q_{\mathrm{g}}}{2}\log(1+1/D)\le\frac{13}{24000F}<\log(1+\eta).$$

The step count and polynomial encoding length follow directly from the geometric progression (173). □

## D.2 Primitive directions and sparsification

We use the primitive direction sparsification method of Stephens-Davidowitz [SD16, Sections 2.6 and 4]. The statements below give the counting and sampling interfaces needed for the radius mixture, including their behavior at equal lengths and under arbitrary causal choices among shortest vectors. For a basis $B$ of $L$, put

$$\mathcal{P}(L)=L_{\mathrm{prim}}/\{\pm1\},\qquad \mathcal{Y}_r=\{y\in\mathcal{P}(L):\|y\|\le r\},\qquad C_L(r)=|\mathcal{Y}_r|.$$

Choose one canonical representative of each direction. Taking the $L$-primitive core of a nonzero vector means dividing its coordinates in $B$ by their greatest common divisor, and then choosing this sign. It preserves the real direction and cannot increase length.

We use three sparsification facts from [SD16]. Its counting function $\xi(L,r)$ is exactly $C_L(r)$, with closed balls and opposite vectors identified. First, [SD16, Lemma 2.18] gives the following collision criterion: if $p\ge101$ is prime and distinct $x,y\in\mathcal{Y}_r$ satisfy $B^{-1}x\equiv aB^{-1}y\pmod p$ for some $a\in\mathbb{F}_p^\times$, then

$$C_L(r)>\frac{p}{20\log p}. \tag{175}$$

For uniform $z\in\mathbb{F}_p^n$, define the full rank sublattice

$$L_z=\{Bx:x\in\mathbb{Z}^n,\ \langle z,x\rangle=0\pmod p\}. \tag{176}$$

For $z=0$ use $B$. Otherwise choose $j$ with $z_j\neq0$, let $c_k\equiv-z_kz_j^{-1}\pmod p$ with $0\le c_k<p$, and use

$$b_k+c_kb_j\quad(k\neq j),\qquad pb_j. \tag{177}$$

These columns generate the kernel, with determinant $p$ in $B$-coordinates. Constructing each query afresh from $B$ gives encoding length polynomial in size$(B)$ and $\log p$.

**Lemma D.2.** *Suppose $r\ge\lambda_1(L)$, $p\ge101$ is prime, and, with $C=C_L(r)$,*

$$p\lambda_1(L)>r,\qquad C\le\frac{p}{20\log p}. \tag{178}$$

*For every SVP oracle, including any choice among tied shortest vectors, the one trial probability $a_y$ of returning direction $y\in\mathcal{Y}_r$ satisfies*

$$\frac1p-\frac{C-1}{p^2}\le a_y\le\frac1p. \tag{179}$$

*Consequently,*

$$\frac{C}{p}-\frac{C(C-1)}{p^2}\le\Pr[\lambda_1(L_z)\le r]\le\frac{C}{p}. \tag{180}$$

*Proof.* By [SD16, Theorem 4.1 and Proposition 4.2], the event that $y$ survives and all competing directions of length at most $\|y\|$ do not survive has probability at least $1/p - (C-1)/p^2$. The condition $p\lambda_1(L) > r$ excludes a short nonprimitive multiple whose primitive core does not survive. Hence this event forces every correct oracle to return $\pm y$, regardless of ties or previous replies. The upper bound is the survival probability $1/p$, and summing the direction bounds gives (180). Fresh $z$ makes the same argument valid conditional on any prior oracle history. □

### D.3 A counting and sampling interface

At each radius we need an approximate number of primitive directions and a procedure that samples them nearly uniformly. The next lemma either supplies these two objects or identifies a radius with few directions, whose contribution is treated separately in the mixture.

Loop bounds use the integer logarithm $\operatorname{clog}_2$ defined in Appendix C.1.6. As above, prime and integer draws are analytic until the finite implementation in Appendix D.8.

**Lemma D.3** (Counting primitive directions)**.** *Given $L$, a positive rational squared radius $r^2$, rational $0 < \eta \le 1/10$, integer $G \ge 2$, and $0 < \zeta_{\mathrm{pre}} < 1/4$, an analytic algorithm using SVP returns a tagged result* GOOD *or* DEGENERATE*. It makes a deterministically bounded polynomial number of queries. There is an event $E_{\mathrm{pre}}(r)$ of probability at least $1 - \zeta_{\mathrm{pre}}$, depending only on preprocessing, such that on every transcript in this event:*

*(i) A* GOOD *result supplies $N > 0$, an integer scale $M$, a prime $p$, an integer $A$, and a sampling procedure, with*

$$(1-\eta)C_L(r) \le N \le (1+\eta)C_L(r), \qquad C_L(r) \le M, \quad AM \le p \le 2AM.$$

*For any subsequent history and any $0 < \zeta_{\mathrm{s}} < 1/4$, the procedure makes at most*

$$R_{\mathrm{s}} = 7A \operatorname{clog}_2(1/\zeta_{\mathrm{s}}) \tag{181}$$

*queries and returns either a direction in $\mathcal{Y}_r$ or* TIMEOUT*. The timeout probability is at most $\zeta_{\mathrm{s}}$. Conditional on success, each direction has probability in $[(1-\eta)/C_L(r), (1+\eta)/C_L(r)]$.*

*(ii) A* DEGENERATE *result guarantees*

$$C_L(r) < \frac{r}{G\lambda_1(L)}. \tag{182}$$

*The probability bounds hold for every rule for choosing a shortest vector based on the current and previous queries. The query caps hold on every execution.*

*Construction and counting guarantee.* An initial SVP query gives $s_1 = \lambda_1(L)^2$, if it is not already available. If $r^2 < s_1$, return DEGENERATE. Otherwise, using rational squared comparisons, compute

$$R = \left\lceil \sqrt{r^2/s_1} \right\rceil, \quad m = \max\left\{1, \left\lceil \sqrt{r^2/(G^2 s_1)} \right\rceil\right\}, \quad U = (1+2R)^n.$$

Packing gives $C := C_L(r) \le U$. Put

$$\begin{aligned} V_0 &= \max\{U, m, 2\}, \qquad S = 101 + G + \eta^{-1} + \operatorname{clog}_2 V_0, \\ A &= \lceil 10^4 S \rceil, \qquad k = \min\{j \ge 0 : m2^j \ge U\}, \qquad K = k+1, \\ e &= \frac{\eta}{1000A}, \qquad s = \left\lceil \frac{\operatorname{clog}_2(2K/\zeta_{\mathrm{pre}})}{2e^2} \right\rceil. \end{aligned} \tag{183}$$

Scan $M = m2^k, m2^{k-1}, \ldots, m$ in this order. At each scale generate a prime $p \in [AM, 2AM]$ by independent integer trials with deterministic primality testing. The prime-density estimate above supplies the geometric tail used by the finite budget. For this fixed $p$, make $s$ independent kernel lattice queries, recording the indicators $X = \mathbf{1}[\lambda_1(L_z) \le r]$. In this analytic interface, coordinates are exactly uniform in $\mathbb{F}_p$. Their implemented draws use Appendix C.1.6. Let $\widehat{q}$ be their average. At the first scale satisfying

$$\widehat{D} := p\widehat{q} \ge M/3, \tag{184}$$

return GOOD with $N = \widehat{D}$ and retain $M, p, A$. If no scale triggers, return DEGENERATE.

All generated primes satisfy $p \le 4AV_0$. Since $\log A < S$ and $\log V_0 \le S \log 2$, the chosen constants imply

$$A > G, \qquad 40 \log p < 80S < A. \tag{185}$$

In particular, at any scale $M \ge C$, $p\lambda_1 > r$ and $C \le M < p/(20 \log p)$, so Lemma D.2 applies.

Conditional on all previous history and the stage prime, the hit indicators are independent Bernoulli variables: their values depend on shortest lengths, not on the oracle's choice of vector. For $q_p = \Pr[\lambda_1(L_z) \le r]$, Hoeffding's inequality and a conditional union bound over at most $K$ stages give

$$\Pr[E_{\mathrm{pre}}(r)] \ge 1 - \zeta_{\mathrm{pre}}, \qquad E_{\mathrm{pre}}(r) = \bigcap_{M \text{ reached}} \{|\widehat{q} - q_{p(M)}| \le e\}.$$

Fix a transcript in this event. If $C \ge m$, let $M_*$ be the smallest candidate scale at least $C$, so $C \le M_* < 2C$. All scales preceding $M_*$ are at least $C$. Unless an earlier scale has triggered, Lemma D.2 therefore applies at $M_*$ and gives

$$\widehat{D} \ge C(1 - C/p) - pe \ge \frac{M_*}{2}(1 - 1/A) - \frac{\eta M_*}{500} > M_*/3.$$

Thus the first trigger occurs at $M \ge M_* \ge C$. If $C < m$, every candidate is already larger than $C$. In both cases, a first trigger always lies in the range where the isolation bounds are valid.

If no scale triggers, $C < m$. Since $C \ge 1$ and $m = \max\{1, \lceil r/(G\lambda_1) \rceil\}$, integrality gives $C < r/(G\lambda_1)$. At a trigger, the upper bound $\widehat{D} \le C + pe$ gives

$$C \ge (1/3 - \eta/500)M. \tag{186}$$

Consequently,

$$|N - C| \le C^2/p + pe \le C/A + \eta M/500 < 0.01\eta C,$$

using $A \ge 10^4 \eta^{-1}$. This proves the required count estimate. There are at most $Ks + 1$ queries. The quantities $K, A, s, \log p$ are polynomial in the input length, $G$, $\eta^{-1}$, and $\log(1/\zeta_{\mathrm{pre}})$, on every preprocessing path. The sampling assertion follows from the next lemma. □

**Lemma D.4.** *Fix a* GOOD *transcript in $E_{\mathrm{pre}}(r)$ and retain its prime $p$ and integer $A$. Independently sample successive $z \in \mathbb{F}_p^n$, query SVP on $L_z$, and stop at the first reply of length at most $r$, returning its canonical direction. Stop with* TIMEOUT *after $R_{\mathrm{s}}$ trials as in (181). This procedure has the timeout and conditional-success guarantees of Lemma D.3, uniformly over every prior history and every causal oracle.*

*Proof.* Write $C = C_L(r)$, $\alpha = 1/p - (C-1)/p^2$, and $\beta = 1/p$. Condition on any history just before a new trial. Its fresh $z$ is still uniform, so Lemma D.2 gives, for the success event $S_t$ and its returned direction $Y_t$,

$$\alpha \le \Pr[S_t, Y_t = y \mid \text{history}] \le \beta.$$

Summing and conditioning on $S_t$ gives

$$\frac{1-(C-1)/p}{C} \le \Pr[Y_t = y \mid S_t, \text{history}] \le \frac{1}{C(1-(C-1)/p)}. \tag{187}$$

Conditional on a hit before the cap, the final direction law is a convex combination of these laws over the termination time and its preceding history. Hence the same bounds hold. Since $C/p \le 1/A$ and $A \ge 10^4\eta^{-1}$, they imply the claimed $(1 \pm \eta)/C$ interval.

By (186) and $p \le 2AM$, every history has conditional hit probability at least

$$C\alpha \ge \frac{1/3 - \eta/500}{2A}(1 - 1/A) > \frac{1}{7A}.$$

Iterated conditioning bounds the probability of $R_{\mathrm{s}}$ failures by $e^{-R_{\mathrm{s}}/(7A)} \le \zeta_{\mathrm{s}}$. No independence assumption on the oracle's choices is used. The cap is imposed even when preprocessing was incorrect. □

### D.4 Bounding the radii with too few primitive directions

Finally, [SD16, Lemma 2.19] gives, for $n \ge 2$ and $r \ge \lambda_2(L)$,

$$C_L(r) > \frac{\sqrt{r^2 - \lambda_2(L)^2}}{\lambda_1(L)} + \left\lfloor \frac{r - \lambda_2(L)}{\lambda_1(L)} \right\rfloor. \tag{188}$$

In both counting results, taking the $L$-primitive cores of the constructed vectors preserves their distinct real directions and does not increase their lengths.

**Lemma D.5.** *For $n \ge 2$ and $G \ge 2$, put $r_* := \lambda_2(L)/\sqrt{1 - G^{-2}}$. Then*

$$\{r \ge \lambda_1 : C_L(r) > 1, C_L(r) < r/(G\lambda_1)\} \subseteq [\lambda_2, r_*). \tag{189}$$

*In particular, on valid preprocessing transcripts every layer with radius at least $r_*$ is* GOOD. *For the global parameters $Q, G, T, R_{\mathrm{g}}$ and $\eta$ of this reduction, if $\lambda_2 \le s \le T$ and $s < r_*$, then*

$$r_* \le R_{\mathrm{g}}, \qquad h_\sigma(s) \le e^{\eta} h_\sigma(r_*). \tag{190}$$

*Proof.* More than one primitive direction in the ball implies $r \ge \lambda_2$. By (188), $C_L(r) > \sqrt{r^2 - \lambda_2^2}/\lambda_1$. For $r \ge r_*$ this exceeds or equals the degenerate threshold $r/(G\lambda_1)$, proving (189) and the assertion about valid tags.

For the last claim, integrate the logarithmic-derivative bound of Equation (174) to obtain

$$\log \frac{h_\sigma(s)}{h_\sigma(r_*)} \le 13 \max\{1, r_*^2/\sigma^2\} \left[-\tfrac{1}{2}\log(1 - G^{-2})\right] \le \frac{104Q}{9G^2} < \eta.$$

Here $s \le T$ implies $r_* \le R_{\mathrm{g}}$ and $r_*^2/\sigma^2 \le 4Q/3$, while $-\frac{1}{2}\log(1 - G^{-2}) \le 2/(3G^2)$. The last inequality follows from $G \ge 10^4 FQ$ and $\eta = 1/(1000F)$. □

The case $n = 1$ uses a one-dimensional sampler directly. For $n \ge 2$, a degenerate layer with $C_L(r) = 1$ contains only the chosen shortest direction and can be represented exactly by its point mass. A degenerate layer omitting any other direction must lie in (189). All statements use closed balls, so equal length directions and grid endpoints need no perturbation.

## D.5 Sampling directions by a mixture over radii

At each radius $r_i$ run Lemma D.3 with the parameters in (171) and

$$\zeta_{\mathrm{pre}} = \frac{\kappa}{8(J+1)}.$$

Reuse the initial shortest length in these calls. For a GOOD layer retain its $N_i$ and sampling procedure. For a DEGENERATE layer set $N_i = 1$ and use the point mass on $y_0$. Define

$$w_i = h_\sigma(r_i) - h_\sigma(r_{i+1}) \quad (i < J), \qquad w_J = h_\sigma(r_J). \tag{191}$$

These weights turn the retained direction samplers into the desired Gaussian mixture. Give the zero vector weight one and set the direction failure allowance to $\zeta_{\mathrm{s}} = \kappa/8$. After choosing a layer and drawing its direction, sample a nonzero integer multiple as in Algorithm 7. Finite draws use Appendix D.7, with zero returned on timeout or a failed check.

**Algorithm 7:** Centered Gaussian sampling by a mixture over radii

**Input:** A rational lattice $L$, squared width $\sigma^2 > 0$, accuracy $f \geq 1$, and an SVP oracle.
**Output:** A vector in $L$, with the accuracy of Theorem 12.6.

Set the parameters of Equation (171) and the finite budgets of Appendix D.7
Set $\zeta_{\mathrm{s}} \leftarrow \kappa/8$
**if** $n = 1$ **then**
  **return** a scalar Gaussian multiple of the input generator using Appendix C.1.7
Obtain a shortest direction $y_0$ by SVP and set $q_0 \leftarrow \|y_0\|^2/\sigma^2$
**if** $q_0 \geq Q$ **then**
  **return** 0
Construct the multiplicative squared-radius grid of Equation (173)
**for** $i = 0, \ldots, J$ **do**
  Run the counting interface at $r_i^2$ with $\zeta_{\mathrm{pre}} = \kappa/(8(J+1))$
  **if** the finite counting procedure times out **then**
    **return** 0
  **if** the returned tag is DEGENERATE **then**
    Set $N_i \leftarrow 1$ and retain the point mass at $y_0$
  **else**
    Retain the GOOD layer's count $N_i$ and capped direction sampler
Compute the positive weights $w_i$ from Equation (191)
Choose zero or layer $i$ with weights $1, N_0 w_0, \ldots, N_J w_J$
**if** zero is chosen **then**
  **return** 0
Draw $y$ from layer $i$ using its retained procedure and failure allowance $\zeta_{\mathrm{s}}$
**if** the direction sampler returns TIMEOUT **then**
  **return** 0
Draw $z$ from $D_{\mathbb{Z}\setminus\{0\},\sigma/\|y\|,0}$ using the scalar sampler
**return** $zy$

Let $E_{\mathrm{pre}}$ be the event that all layer interfaces satisfy their assertions. The conditional union bound gives $\Pr[E_{\mathrm{pre}}^c] \leq \kappa/8$. Fix a transcript $\omega \in E_{\mathrm{pre}}$. In the analytic mixture, use exact categorical weights and the complete capped direction sampling transcript conditioned on success. Its direction marginal $P_{i,\omega}$ is within $(1 \pm \eta)/C_L(r_i)$ on GOOD layers. On degenerate layers it is the point mass

at $y_0$. Keeping the complete transcript permits subsequent oracle choices to depend on all previous queries and replies. The unnormalized weight of a direction is

$$\widetilde{h}_\omega(y) = \sum_{i:\|y\|\le r_i} N_i P_{i,\omega}(y) w_i.$$

**Lemma D.6.** *Put $u = (1-\eta)^2$, $v = (1+\eta)^2$, $a = u/((1+\eta)e^\eta)$, and $b = v$. For every good preprocessing transcript and every direction of length at most $T$,*

$$a h_\sigma(\|y\|) \le \widetilde{h}_\omega(y) \le b h_\sigma(\|y\|).$$

*For $T < \|y\| \le R_{\mathrm{g}}$ the upper bound still holds.*

*Proof.* On a good layer the multiplier $N_i P_{i,\omega}(y)$ lies in $[u, v]$. A degenerate layer gives multiplier one to $y_0$ and zero to all other directions. Thus $y_0$ has total weight in $[u h_\sigma(\lambda_1), v h_\sigma(\lambda_1)]$. For any other $y$, let $j(y)$ be its first eligible layer. The upper bound follows from nonnegativity and telescoping: $\sum_{i\ge j(y)} w_i = h_\sigma(r_{j(y)}) \le h_\sigma(\|y\|)$. If all eligible layers are good, the same sum and Lemma D.1 give the lower bound with factor $u/(1+\eta)$.

Otherwise an eligible degenerate layer has $C_L(r_i) > 1$. By Lemma D.5, $\lambda_2 \le \|y\| \le r_i < r_* = \lambda_2/\sqrt{1-G^{-2}}$. Since $\|y\| \le T$, we have $r_* \le R_{\mathrm{g}}$. Let $j_*$ be the first grid index with $r_{j_*} \ge r_*$. Every layer from $j_*$ onward is good and contains $y$. The endpoint $j_* = J$ is allowed. Therefore

$$\widetilde{h}_\omega(y) \ge u \sum_{i=j_*}^{J} w_i = u h_\sigma(r_{j_*}) \ge \frac{u}{1+\eta} h_\sigma(r_*) \ge a h_\sigma(\|y\|),$$

using the grid bound and (190). □

## D.6 Normalization and the two accuracy parameters

Let

$$\mathcal{A} = \{0\} \cup \{zy : y \in \mathcal{P}(L),\ \|y\| \le T,\ z \in \mathbb{Z}\setminus\{0\}\}, \qquad \tau = D_{L,\sigma}(\mathcal{A}^c) < \varepsilon/64,$$

and let $D_{\mathcal{A}}$ be the target conditioned on $\mathcal{A}$. For a fixed good transcript $\omega$, let $\widehat{Q}_\omega$ be the analytic mixture law. Its zero weight is one, and the weight of $zy$ is

$$\widetilde{h}_\omega(y) \frac{e^{-\pi z^2 \|y\|^2/\sigma^2}}{h_\sigma(\|y\|)}.$$

Every point in $\mathcal{A}$ has a multiplier in $[a, b]$ relative to its target unnormalized mass, including zero since $a \le 1 \le b$. Thus $Q_\omega = \widehat{Q}_\omega(\cdot \mid \mathcal{A})$ satisfies

$$\frac{a}{b} D_{\mathcal{A}}(x) \le Q_\omega(x) \le \frac{b}{a} D_{\mathcal{A}}(x). \tag{192}$$

Mixing $Q_\omega$ over the good preprocessing transcripts gives $\overline{Q}_{\mathcal{A}}$ with the same bounds. The direction sampler's guarantee permits every causal choice of shortest vectors. Moreover,

$$\log(b/a) = 3\log(1+\eta) - 2\log(1-\eta) + \eta < 7\eta < \log(1+1/F) \le \log(1+1/f).$$

The unnormalized buffer mass is at most $b\rho_\sigma(L)\tau$, while the mass in $\mathcal{A}$ is at least $a\rho_\sigma(L)(1-\tau)$. Since $\eta \le 1/2000$ and $\tau < 1/128$, this gives

$$d_{\mathrm{TV}}(\widehat{Q}_\omega, Q_\omega) = \widehat{Q}_\omega(\mathcal{A}^c) \le \frac{b}{a}\frac{\tau}{1-\tau} < 2\tau < \varepsilon/32 = \kappa/8. \tag{193}$$

This accounts for the buffer without rejecting its directions.

For now leave prime and uniform integer draws analytic, and denote the output law with the other finite budgets by $\mu_0$. On a good preprocessing transcript, couple the finite choices to $\widehat{Q}_\omega$. At the direction invocation retain the actual complete transcript on success. On timeout use a fresh transcript from its success-conditioned kernel for the reference and charge probability at most $\zeta_{\mathrm{s}}$. This preserves oracle histories up to the first disagreement. Then compare $\widehat{Q}_\omega$ to $Q_\omega$ using (193). On a bad preprocessing transcript the reference uses an independent sample of $\overline{Q}_{\mathcal{A}}$ and charges that branch as a disagreement. Together with the finite numerical implementation below, these comparisons give

$$d_{\mathrm{TV}}(\mu_0, \overline{Q}_{\mathcal{A}}) \le \underbrace{\kappa/8}_{\text{preprocessing}} + \underbrace{\kappa/8}_{\text{direction timeout}} + \underbrace{\kappa/8}_{\text{buffer}} + \underbrace{\kappa/8}_{\text{categorical}} + \underbrace{\kappa/8}_{\text{scalar}} < \kappa.$$

Failures return zero and contribute only to this additive error. Their mass is not included in the multiplicative comparison.

Define a full-support witness by

$$P(x) = \begin{cases} (1-\tau)\overline{Q}_{\mathcal{A}}(x), & x \in \mathcal{A}, \\ D_{L,\sigma}(x), & x \notin \mathcal{A}. \end{cases}$$

It is normalized, has the same support as $D_{L,\sigma}$, and satisfies $D_{L,\sigma}/(b/a) \le P \le (b/a)D_{L,\sigma}$. Furthermore $d_{\mathrm{TV}}(\overline{Q}_{\mathcal{A}}, P) = \tau$, so

$$d_{\mathrm{TV}}(\mu_0, P) < \kappa + \varepsilon/64 = 17\varepsilon/64. \tag{194}$$

This proves the accuracy of the analytic interface, with slack for the finite prime and integer budgets and the randomized oracle implementation.

### D.7 Evaluation of the radius mixture

Use the parameters of (171) and write

$$H(q) = 2\sum_{k\ge 1} e^{-\pi k^2 q}, \qquad w_i = H(q_i) - H(q_{i+1}) \ (i < J), \qquad w_J = H(q_J).$$

The zero vector contributes weight one. This lower bound on total mass lets us use absolute accuracy for the differences $w_i$, even when adjacent radii are very close.

For nonnegative rational counts, put $N_* = \max\{1, N_0, \dots, N_J\}$. To achieve error $0 < \delta \le 1$, evaluate each $H(q_i)$ to absolute error $\delta/(8(J+1)N_*)$, take successive differences, and clip them to be nonnegative. The total weighted error is at most $\delta/4$, and the lower bound one on total mass controls normalization.

For $q \ge 1/4$, the Gaussian series converges geometrically. For smaller $q$, Poisson summation gives $H(q) = q^{-1/2}(1+H(1/q))-1$. Thus evaluation has polynomial cost in the grid and count descriptions and $\log(1/\delta)$, uniformly for small $q$ as well. Categorical rounding to $p = \lceil \log_2(4(J+2)/\delta) \rceil$ bits completes a draw within TV distance $\delta$ under a fixed polynomial budget.

Use error $\kappa/8$ for this categorical draw and for the one scalar draw from $D_{\mathbb{Z}\setminus\{0\},\sigma/\|y\|,0}$ supplied by Appendix C.1.7. Their combined error is at most $\kappa/4 = \varepsilon/16$.

### D.8 Query bounds and the randomized SVP implementation

Write $N = \mathrm{size}(B, \sigma^2)$ and choose rational $\ell \le \|B^{-1}\|_F^{-1}$. Initialize every SVP query with a shortest basis column and retain only shorter nonzero lattice vectors. The initial answer has length $\ell \le \widehat{\lambda} \le \min_j \|b_j\|$ on every execution.

The grid bound in Lemma D.1 is therefore polynomial in $N, f$, even with an incorrect initial answer. Packing bounds the logarithm of the number of directions in the largest ball by

$$O(n \log(1 + R_{\mathrm{g}}/\ell)) = \mathrm{poly}(N, f).$$

Substitution in (183) bounds every layer's candidate scales and trial counts by computable polynomials. Thus we can fix a polynomial bound

$$Q_* = \mathrm{poly}(N, f) \tag{195}$$

on all oracle queries before the first call.

The kernel bases in (177) have polynomial encoding length because each query starts from $B$ and uses a prime of polynomial bit length. The length checks give the same bound for answers and primitive cores, uniformly over causal oracle choices. Together with Appendix D.7, these bounds give fixed polynomial work and space outside the oracle. The prime and integer draws use Appendix C.1.6 with total failure at most $\varepsilon/8$. Enforcing these budgets and returning zero on exhaustion adds this allowance to (194).

Replace each exact query by independent executions of Theorem 1.3, retaining the shortest candidate, with

$$\delta_{\mathrm{or}} = \frac{\varepsilon}{8Q_*}, \qquad r = \left\lceil \log_3(\delta_{\mathrm{or}}^{-1}) \right\rceil$$

repetitions. Conditional on any preceding transcript, the amplified reply fails with probability at most $\delta_{\mathrm{or}}$. Sequential coupling over $Q_*$ queries adds at most $\varepsilon/8$ to total variation. Hence the complete additive error is less than

$$17\varepsilon/64 + \varepsilon/8 + \varepsilon/8 = 33\varepsilon/64 < \varepsilon,$$

with the same multiplicative witness as in the analytic proof. The algorithm remains $(1 + 1/f, 2^{-f})$-close to its target.

Every query has dimension $n$, and the repetition count is polynomial in $N, f$. Reusing the workspace gives time $2^{0.4130n+o(n)}$ and space $2^{0.3138n+o(n)}$, up to polynomial factors in the input and accuracy parameters.

## E Proofs of the further consequences

This appendix gives the sampling argument behind the time–approximation tradeoff. The two geometric estimates apply at every Gaussian width. One produces a point near a prescribed shift. The other produces a short vector outside an arbitrary proper subspace. The latter permits a direct SMP algorithm in the original dimension. Appendix E.3 applies the rational projected basis construction of Appendix A.4 to polynomial-factor approximations.

### E.1 Gaussian mass in a small ball

Write $B_R = \{x : \|x\| \le R\}$ and use (135). Markov's inequality at width $\sqrt{\beta}s$, followed by the scale comparison, gives

$$\rho_s(L \cap B_{R_0}) \ge \frac{\tau}{1+\tau}\beta^{n/2}\rho_s(L), \qquad R_0^2 = (1+\tau)\beta n s^2/(2\pi). \tag{196}$$

**Lemma E.1** (Gaussian mass near a shifted closest point)**.** *Let* $r = \operatorname{dist}(t, L)$, $s > 0$, $0 < \beta \le 1$, *and* $\tau > 0$. *Set*

$$R_0^2 = \frac{(1+\tau)\beta n s^2}{2\pi}, \qquad R^2 = r^2 + R_0^2.$$

*Then*

$$\rho_s((L-t) \cap B_R) \ge \frac{\tau}{2(1+\tau)}\, \beta^{n/2} e^{-\pi r^2/s^2} \rho_s(L). \tag{197}$$

*Proof.* Put $C = L \cap B_{R_0}$ and use (196). Choose a closest lattice point $x_0$ and put $u = x_0 - t$. For each unordered pair $\{x, -x\}$ in $C \setminus \{0\}$, choose a sign such that $\langle u, \pm x\rangle \le 0$. Then

$$\|u \pm x\|^2 \le r^2 + \|x\|^2 \le R^2, \qquad \rho_s(u \pm x) \ge e^{-\pi r^2/s^2} \rho_s(x).$$

The selected points belong to $L - t$ and are distinct for distinct pairs. The point $x = 0$ contributes $u$ and satisfies the same weight bound. Consequently the shifted ball has mass at least $\frac{1}{2} e^{-\pi r^2/s^2} \rho_s(C)$, which gives the claim by (196). □

**Lemma E.2** (A short sample outside a subspace)**.** *Let* $W \subsetneq \mathbb{R}^n$ *be a linear subspace and* $v \in L \setminus W$. *For* $s > 0$, $0 < \beta \le 1$, *and* $\tau > 0$, *put*

$$R_0^2 = \frac{(1+\tau)\beta n s^2}{2\pi}, \qquad R^2 = \|v\|^2 + R_0^2.$$

*Then*

$$\mathbb{P}_{X \sim D_{L,s}}[X \notin W,\ \|X\| \le R] \ge \frac{\tau}{3(1+\tau)}\, \beta^{n/2} e^{-\pi \|v\|^2/s^2}. \tag{198}$$

*Proof.* Again put $C = L \cap B_{R_0}$ and let

$$A = \rho_s(C \setminus W), \qquad B = \rho_s(C \cap W), \qquad q = e^{-\pi \|v\|^2/s^2}, \qquad G = (L \setminus W) \cap B_R.$$

Since $C \setminus W \subseteq G$, we have $\rho_s(G) \ge A$. For each pair $\{x, -x\}$ in $C \cap W$, select $v + x$ or $v - x$ with nonpositive cross term. The selected point lies outside $W$, has squared norm at most $\|v\|^2 + \|x\|^2$, and has weight at least $q\rho_s(x)$. Distinct pairs give distinct points, and the weight bound also holds for $x = 0$, so $\rho_s(G) \ge qB/2$. The two sources of points can overlap, but

$$\rho_s(G) \ge \max\{A, qB/2\} \ge \frac{q}{2+q}(A + B) \ge \frac{q}{3}\rho_s(C).$$

Use (196) and divide by $\rho_s(L)$. □

## E.2 A general sampling reduction

We use $\Phi$ from (133). Approximation factors are effective in the sense stated in Section 12.5.

*Proof of Proposition 12.7.* We first treat a rational $\gamma > 1$. Write $\Delta = \gamma^2 - 1$, $\tau = 1/n$, and

$$\beta = \min\{1, \Delta\}. \tag{199}$$

Suppose a squared scale $r_0^2$ is available with

$$\ell/(1 + 1/n) < r_0 \le \ell, \tag{200}$$

where $\ell$ is the unknown distance or successive minimum relevant to the application. Choose a positive rational $s$ satisfying

$$\frac{A_0}{(1+1/n)^2} \le s^2 \le A_0, \qquad A_0 = \frac{2\pi\Delta r_0^2}{(1+1/n)\beta n}. \tag{201}$$

These rational squared scales and a lower approximation to $\pi$ give

$$R_0^2 \le \Delta r_0^2 \le \Delta \ell^2, \qquad \frac{\pi\ell^2}{s^2} \le \frac{n\beta}{2\Delta}(1+1/n)^5 \le \frac{n\beta}{2\Delta} + 16. \tag{202}$$

Thus discretization costs only an absolute constant in the exponent. Optimizing the remaining sampling cost gives

$$\min_{0<\beta\le 1}\left(\frac{\beta}{\Delta} + \log\frac{1}{\beta}\right) = \Phi(\Delta). \tag{203}$$

Differentiating in $\beta$ gives (199).

**Finding independent vectors in dimension $n$.** At stage $i$, keep $W = \mathrm{span}(v_1, \dots, v_{i-1})$ fixed during the entire stage. There exists $v \in L \setminus W$ with $\|v\| \le \lambda_i(L)$, since $i$ independent vectors attain the $i$th successive minimum. At a scale satisfying (200) with $\ell = \lambda_i(L)$, Lemma E.2 and (202) show that

$$\begin{aligned}\mathbb{P}_{X\sim D_{L,s}}[X \notin W,\ \|X\| \le \gamma\lambda_i(L)] \\ \ge \frac{e^{-16}}{3(n+1)}\exp\left[-\frac{n}{2}\Phi(\Delta)\right].\end{aligned} \tag{204}$$

This estimate holds for every history of the preceding stages.

Let $Q$ clear the input denominators and $U^2 = \max_j \|b_j\|^2$. A descending squared scale grid of ratio $(1+1/n)^2$ covering $[Q^{-2}, U^2]$ has polynomial size and satisfies (200) for every $\lambda_i(L)$. At stage $i$, sample at every scale and retain the shortest vector outside $W$, initialized by a qualifying input basis vector. The good scale and (204) give $\|v_i\| \le \gamma\lambda_i(L)$ once a useful sample appears.

**Finding a closest-vector approximation in dimension $n+1$.** Given $t \in \mathbb{Q}^n$, first test $t \in L$ by exact rational linear algebra, and return $t$ if so. Otherwise set $r = \mathrm{dist}(t, L) > 0$. At each candidate width $s$ construct the full rank rational lattice

$$\Lambda_s = \{(v - kt, ks) : v \in L,\ k \in \mathbb{Z}\} = \begin{pmatrix} B & -t \\ 0 & s \end{pmatrix}\mathbb{Z}^{n+1}. \tag{205}$$

Call the centered sampler on $\Lambda_s$ at width $s$. By shift maximality,

$$\rho_s(\Lambda_s) = \sum_{k\in\mathbb{Z}} e^{-\pi k^2}\rho_s(L - kt) \le \rho_1(\mathbb{Z})\rho_s(L). \tag{206}$$

The ideal probability of a sample $(z, s)$ with $z \in L - t$ and $\|z\| \le R$ is therefore at least

$$\frac{e^{-\pi}}{\rho_1(\mathbb{Z})}\frac{\tau}{2(1+\tau)}\beta^{n/2}e^{-\pi r^2/s^2}. \tag{207}$$

Here Lemma E.1 defines $R^2 = r^2 + (1+\tau)\beta n s^2/(2\pi)$. At a scale satisfying (200) with $\ell = r$, (202) gives $R \le \gamma r$. The bound is on the first block of the sample, which is the distance to the target. Letting $Q$ also clear the denominators of $t$, use the same squared scale grid over $[Q^{-2}, \|t\|^2]$. A sample $(z, ks)$ with $k = \pm 1$ yields the candidate $t + kz$. Retaining the closest verified candidate over all scales, initialized by zero, gives the approximation when the good scale supplies a hit.

**Precision and a predetermined repetition count.** Since $\rho_1(\mathbb{Z}) < 2$, both (204) and (207) are bounded below by

$$p_* := \frac{e^{-24}}{n+1} \exp\left[-\frac{n}{2}\Phi(\Delta)\right] \tag{208}$$

at their respective good scales. With $N = n$ for SMP and $N = n+1$ for CVP, choose $f(N) = N^2$. For every good event $E$, the two-parameter guarantee implies

$$\mathbb{P}_{\text{actual}}(E) \geq \frac{\mathbb{P}_{\text{ideal}}(E)}{1 + 1/N^2} - 2^{-N^2} \geq p_*/2 \tag{209}$$

for large $n$, since $p_* = e^{-O(n)}$ for fixed $\gamma > 1$. Increase the precision in the finitely many remaining ranks so that $f \geq 4$ and $2^{-f} \leq p_*/4$. The pointwise multiplicative guarantee is essential here: it preserves an event whose probability is exponentially smaller than the sampler's total variation error.

At every scale use

$$M = \left\lceil \frac{2}{p_*} \log(6n) \right\rceil \tag{210}$$

calls. Standard amplification gives failure at most $1/(6n)$ per stage and at most $1/6$ over all SMP stages. Multiplying $M$ by the polynomial number of scales and stages proves (134).

The rational grids and query bases have polynomial encoding length, so sequential calls give the stated space bound.

**Effective constants and growing approximation factors.** For a fixed nonrational effective $\gamma > 1$, choose a rational $\widehat{\gamma} \leq \gamma$ with $\widehat{\gamma}^2 - 1 = (\gamma^2 - 1)(1 - O(1/n))$. This changes $n\Phi(\gamma^2 - 1)$ by $O(1)$ and yields at least the requested approximation quality. All computations above then use $\widehat{\gamma}$.

For a rational effective $\gamma(n) = \omega(1)$, run the construction with $\widehat{\gamma}(n) = \min\{\gamma(n), n\}$, for $n \geq 2$. Eventually $\widehat{\Delta} = \widehat{\gamma}(n)^2 - 1 \geq 1$, so $\beta = 1$ and

$$\frac{n}{2}\Phi(\widehat{\Delta}) = \frac{n}{2(\widehat{\gamma}(n)^2 - 1)} = o(n).$$

The constant loss in (202) is uniform, and $f(N) = N^2$ still gives (209). Rounding an effective factor downward with relative squared error $O(1/n)$ preserves divergence. Capping at $n$ costs only $O(1/n)$ in $n\Phi$, while keeping the parameter and grids polynomially encoded. Hence the call count is uniformly $2^{o(n)} \operatorname{poly}(|I|, n)$. □

## E.3 Block reduction consequences

The rational projected basis interface of Appendix A.4 supplies a 2-approximate SVP oracle in each block. Exact projected preprocessing uses Appendix A.3.

For every fixed $c \geq 1$, [ALNS20, Corollary 4.4], with oracle factor 2 and slack $1/n$, reduces $O_c(n^c)$-SVP to blocks of rank $k = \lfloor n/(c+1) \rfloor$. For $1/2 < c < 1$, [ALNS20, Corollary 3.4] gives the same factor with $k = \lceil n/(2c) \rceil$.

For HSVP, [ALNS20, Theorem 2.3] gives

$$\|v\| \leq (1 + 1/n)(4\gamma_k)^{(n-1)/(2(k-1))} \det(L)^{1/n},$$

where $\gamma_k \leq k$ is Hermite's constant. Taking $k = \lceil n/(2c) \rceil$ for fixed $c > 1/2$ gives $(n-1)/(2(k-1)) = c + O_c(1/n)$ and hence factor $O_c(n^c)$. This theorem applies for $n > k \geq 2$. At $c = 1/2$, direct SVP and $\lambda_1(L) \leq \sqrt{n} \det(L)^{1/n}$ give the claim.

The polynomial call and encoding bounds of [ALNS20, Section 2.3] remain valid under Appendix A.4. Standard amplification and sequential workspace reuse therefore give total failure at most $1/3$ and time and space

$$2^{\alpha_{\mathrm{SVP}}k+o(n)} \qquad \text{and} \qquad 2^{\beta_{\mathrm{SVP}}k+o(n)}.$$

Substituting the three block sizes proves Corollary 12.10.

### E.4 Short independent vectors

*Proof of Corollary 12.9.* Apply the reduction of [ADRS15, Lemma 7.5] using independent calls to Theorem 1.1, with $\alpha = 2$ and $\xi = 1$. The usual basis bounds on $\eta = \eta_{1/10}(L)$ give a polynomial squared width grid containing $2\eta < \sigma \le 4\eta$. At that width, with probability $1 - o(1)$, $n^2$ samples supply independent vectors of length $O(\sqrt{n}\,\sigma) = O(\sqrt{n \log n})\lambda_n(L)$, while their joint sampling error is $n^2 \exp[-\Omega(n/\log^3(n+2))] = o(1)$. Retain the verified independent tuple with the smallest maximum length over all scales, initialized by the input basis. Sequential sampling preserves the resource exponents. □

## F Reusable decoding and promise problem reductions

This appendix proves the decoding interfaces and the geometric reductions in Sections 12.2 and 12.3. All calls have fixed work, precision, and reply length limits. Illegal replies are discarded before subsequent arithmetic.

### F.1 Uniform advice from dual Gaussian batches

**Lemma F.1** (Uniform Gaussian decoding)**.** *Suppose $s > 0$, $0 < \varepsilon < 1/200$, and*

$$\rho_{1/s}(L \setminus \{0\}) \le \varepsilon = 2^{-\Theta(n)}.$$

*For every fixed $0 < \theta < 1/2$, let $N = \lceil Cn \log(1/\varepsilon)/\sqrt{\varepsilon} \rceil$ for a sufficiently large constant $C$. Then $N$ independent $D_{L^*,s}$ samples determine, with probability $1 - 2^{-\Omega(n)}$, a deterministic function correct simultaneously on all rational targets at distance at most*

$$\frac{\theta}{s}\sqrt{\frac{\log(1/\varepsilon)}{\pi}} \tag{211}$$

*from $L$. Each query takes $N\,\mathrm{poly}(n, \mathcal{B}_t)$ work. A fixed radius margin permits finite arithmetic with the same bound.*

*Proof.* We adapt the periodic Gaussian decoder of [DRS14, Theorem 3.1 and Proposition 3.2]. The two points to verify are that a known mass upper bound determines the sample count and that one sample event works for all targets.

Normalize to $K = sL$, $W_i = w_i/s \sim D_{K^*,1}$, and put

$$f(t) = \rho_1(K + t)/\rho_1(K), \qquad f_N(t) = N^{-1} \sum_i \cos(2\pi\langle W_i, t\rangle).$$

Poisson summation gives $\mathbb{E}[f_N] = f$. On the ball of radius $R_\theta = \theta\sqrt{\log(1/\varepsilon)/\pi}$, symmetric pairing gives $f(t) \ge \rho_1(t) \ge \varepsilon^{\theta^2}$. Completion of squares and shifted theta maximality give

$$\Pr_{Y \sim D_{K,1}}[|\langle Y, v\rangle| > u,\ Y \ne 0] \le \min\{\varepsilon, 2e^{-\pi u^2}\}$$

for every unit $v$. Put $S = \sqrt{\log(2/\varepsilon)/\pi}$. The double tail integral in [DRS14, Lemma 4.2], split at $S$, is at most $C_\theta(1+\log(1/\varepsilon))\varepsilon e^{2\pi S R_\theta}\|t\|$ throughout the ball. Since $\varepsilon e^{2\pi S R_\theta} = \varepsilon^{1-2\theta-o(1)}$, this gives

$$\left\| \frac{\nabla f(t)}{2\pi f(t)} + t \right\| \le C_\theta \varepsilon^{1-2\theta-o(1)}\|t\|. \tag{212}$$

Only the upper bound on nonzero mass enters these estimates.

Here is the uniform empirical estimate. The moment-generating bound $\mathbb{E}[e^{u\langle W,v\rangle}] \le e^{u^2/(4\pi)}$ makes the summands of

$$v^{\mathsf{T}}\nabla^2 f_N(t)v = -\frac{4\pi^2}{N}\sum_i \langle W_i, v\rangle^2 \cos(2\pi\langle W_i, t\rangle)$$

uniformly subexponential. Bernstein's inequality at deviation $\delta = c\varepsilon^{1/4}$ has failure $2e^{-c'N\delta^2}$. A constant-mesh sphere net costs $e^{O(n)}$ points. Gaussian norm tails give $\max_i \|W_i\| = O(\sqrt{n})$ with failure $2^{-\Omega(n)}$, so the empirical and expected Hessians are $O(n^{3/2})$-Lipschitz. A target mesh of spacing $\delta/(Cn^{3/2})$ has logarithmic size $O(n\log(1/\varepsilon))$. Thus $N\varepsilon^{1/2} = \Omega(n\log(1/\varepsilon))$ yields a uniform Hessian error $O(\varepsilon^{1/4})$. Both gradients vanish at zero. Integration, together with bounded cosine concentration on the same mesh, gives

$$|f_N(t) - f(t)| \le C_1\varepsilon^{1/4}, \qquad \|\nabla f_N(t) - \nabla f(t)\| \le C_1\varepsilon^{1/4}\|t\| \tag{213}$$

simultaneously throughout the ball. Since $\theta^2 < 1/4$, these errors and (212) imply the contraction

$$\left\| t + \frac{\nabla f_N(t)}{2\pi f_N(t)} \right\| \le \frac{1}{2}\|t\|.$$

Periodicity gives the same contraction around every lattice point on this single sample event.

After $O(\log n)$ steps the residual is below $1/(4C\sqrt{n})$. Theta derivatives give expected covariance $I/(2\pi) + O(\varepsilon\log(1/\varepsilon))$. Constant-accuracy Bernstein bounds on a sphere net show that $O(n)$ initial samples contain $n$ independent rows of norm at most $C\sqrt{n}$, with failure $2^{-\Omega(n)}$. Round the current point's inner products with these rows to integers, solve the resulting system for the unique lattice point underlying the residual, and verify membership. Rescaling proves (211). □

*Proof of Theorem 12.1.* Put $C_0 = A_{\mathrm{KL}}/\sqrt{2\pi e}$. The KL estimate in [ADRS15, Lemma 6.1, Eq. (9)] gives

$$\lambda_1(L)\eta_{1/2}(L^*) \le (C_0 + o(1))\sqrt{n}, \qquad \lambda_1(L)\eta_\varepsilon(L^*) \le (C_0 + o(1))\sqrt{n}\,\varepsilon^{-1/n}. \tag{214}$$

Take $\varepsilon = 2^{-2bn}$ and ideal width $s = (1+\xi)\alpha C_0\sqrt{n}/\lambda_1(L)$ with small fixed $\xi > 0$. The feasible parameters obey $\alpha > 2^{2b}$. For the full batch this is $3/2 > 4/3$. In the general region, if $\alpha \ge 4/3$ use $b < \beta_{\mathrm{G}}$. For $1 < \alpha < 4/3$, the function $1 - \alpha^{-2} - \log\alpha$ increases from zero, which implies $J(\alpha^{-2}) < \frac{1}{2}\log_2\alpha$. Thus both width requirements in (214) hold.

The decoder needs $\mathrm{poly}(n)2^{bn}$ samples, supplied by polynomially many independent batches. Their total joint TV error is $o(1)$. Substituting their ideal samples into Lemma F.1, and taking $\theta$ close to $1/2$ and $\xi$ small, yields every strict radius in (128). TV transfers the single global good advice event to the actual samples.

A polynomial rational grid of ratio $1+1/n$ between basis bounds contains $\ell \le \lambda_1(L) < (1+1/n)\ell$. Use $\ell$ in the width formula, with rational rounding inside the fixed radius margin, and retain the closest legal reply over all scales. Their fixed budgets preserve the exponents even outside the sampling promise. On the good scale, one advice event guarantees correctness for every target at once, so this selection also preserves the uniform guarantee.

On good advice $f_N \ge \frac{1}{2}\varepsilon^{\theta^2} = 2^{-O(n)}$, so polynomial precision preserves the contraction and the fixed slack in the final rounding. With fixed iteration and precision bounds, including failure replies on other scales, every advice defines a deterministic bounded query function. □

## F.2 Perturbations and larger distance guarantees

We work with a fixed good advice function on the relevant lattice. For $\|v\| = d > a > 0$, rotate $v$ to $de_1$ and put $q = a/d$. The event in (129) becomes $U_1 \le -\sqrt{1-q^2}$. The standard spherical cap estimate [BDGL16, Lemma 2.1] gives probability $\Theta_q(n^{-1/2}q^n)$. When $q$ approaches one, cap monotonicity supplies a uniform lower bound from any fixed smaller ratio. For any other perturbation length $u > 0$, the cap threshold is

$$\frac{d^2 + u^2 - a^2}{2du} \ge \sqrt{1 - a^2/d^2},$$

with equality at $u = \sqrt{d^2 - a^2}$. Thus this radius maximizes the probability of hitting the specified ball with one spherical perturbation. Averaging any direction-independent perturbation distribution over the error direction gives no larger probability.

*Proof of Proposition 12.2.* First suppose $\ell = \lambda_1(L)$ is known. Fix a closest point $y$ to $t$ and put $d = \|t - y\|$. An initial query handles $d \le r\ell$. Otherwise put $\eta = r\ell/n^2$ and $a = r\ell - 2\eta$. Use a distance grid with spacing at most $\eta$ on $[r\ell, R\ell + \eta]$. Some grid value satisfies $d \le d_i \le d + \eta$. Query at $t + \sqrt{d_i^2 - a^2}U$. Extend $t - y$ in its own direction to a vector $\widehat{v}$ of length $d_i$. If $\|\widehat{v} + \sqrt{d_i^2 - a^2}U\| \le a$, then the actual perturbed residual has length at most $a + \eta < r\ell$. The cap ratio is at least

$$\frac{r(1 - 2/n^2)}{R + r/n^2},$$

whose $n$th power is $(r/R)^n e^{-O_{r,R}(1/n)}$. Polynomially many distance guesses and a polynomial multiple of $(R/r)^n$ trials give any fixed conditional success probability.

For unknown $\lambda_1(L)$, a polynomial rational grid between basis bounds contains $\ell \le \lambda_1(L) < (1 + 1/n^3)\ell$. Enlarging the distance endpoint accordingly and using radius $r\ell$ costs only $e^{O_{r,R}(1/n)}$ in cap probability. All scales share the same advice, and retaining the closest verified reply proves the claim. □

*Proof of Theorem 12.3.* Rank one is immediate. Otherwise fix a shortest vector $v = Bz$. It is primitive, so $z$ is nonzero modulo every prime. For a fixed prime $p > g$, a uniform nonzero linear function $h$ on $\mathbb{F}_p^n$ satisfies

$$\Pr[h(z) \ne 0] = \frac{p^{n-1}(p-1)}{p^n - 1} > 1 - 1/p.$$

On this event its kernel lattice $L'$ contains no parallel multiple of $v$ shorter than $p\|v\|$. Every nonparallel vector has length at least $\lambda_2(L) \ge g\|v\|$, so $\lambda_1(L') \ge g\|v\|$. Solve a modular equation for $w \in L$ with $h(B^{-1}w) = 1$, and enumerate $x_j = jw$, $1 \le j < p$. At $j = h(z)$ the point $y = x_j - v$ belongs to $L'$. All representatives and perturbation scales use one preprocessing of this $L'$.

If $gr \ge 1$, direct decoding at this representative returns $y$. For $q = gr < 1$, a perturbation of length $\|v\|\sqrt{1-q^2}$ hits its ball of radius $q\|v\|$ with probability at least $n^{-O(1)}q^n$. For the unknown $\lambda = \|v\|$ use the preceding rational grid. At a good lower scale $\ell$ set $a = q\ell(1 - 2/n^2)$, $d_i = (1 + 1/n^3)\ell$, and perturb by length $\sqrt{d_i^2 - a^2}$. The extended true residual differs by $O(\ell/n^3)$. Thus the extra slack preserves the same main cap exponent. At $gr = 1$, direct decoding uses the closed promise. Shrinking the geometric radius by $1 - O(1/n^2)$ and adding small perturbations also gives polynomial overhead for a strict-distance interface.

For each reply $y' \in L'$, keep the nonzero residual $x_j - y'$. The perturbation is not subtracted from this residual. Every retained candidate belongs to $L$, and at a hit it equals $v$. Keeping the shortest

candidate, initialized by a basis vector, gives constant success from the sublattice, advice, and cap events. The argument applies also to $g \le 2$, including $g = 1$. The kernel basis and representatives have polynomial encoding length, and standard amplification gives success 2/3. □

Generate spherical perturbations by normalizing Gaussian vectors. Standard tail and small-norm bounds permit a distance coupling within $r\ell/(4n^2)$, with total error $\zeta$ over $Q$ queries, using polynomial precision in the input and $\log(Q/\zeta)$. Shrinking the hit radius by $r\ell/n^2$ preserves its exponent and keeps every coupled target inside the decoding promise.

### F.3 The GapSVP conflict test on a finite grid

*Proof of Theorem 12.4.* Fix the advice, iteration limits, precision, and exceptional returns. Replace every illegal reply by $0 \in L$, obtaining a complete deterministic function $F$ on the rational points in $B[0,1)^n$. Write $R = r\gamma d$. For a test point $e$, compute exactly $e = x + y$ with $x \in B[0,1)^n$ and $y \in L$, and check $F(x) = -y$. Declare YES if any check fails and NO otherwise.

In the NO case good advice passes every point with $\|e\| \le R$, since $\|x - (-y)\| \le r\lambda_1(L)$ and $r < 1/2$. The only errors are bad advice and the global implementation error. In the YES case fix $0 \ne v \in L$ with $\|v\| \le d$. Whenever $e, e + v$ both lie in the ball, their representatives are equal but they require the distinct answers $-y, -y - v$. For every fixed $F$, including bad advice, the failure indicators obey $I(e) + I(e + v) \ge 1$.

In the continuous geometry, put $A = B(0,R) \cap (B(0,R) - v)$. Integration bounds the failure fraction below by $\mathrm{vol}(A)/(2\,\mathrm{vol}(B(0,R)))$. With axis along $v$, a cylinder about the perpendicular bisector has half-length $\delta = (R - d/2)/n$ and cross-section radius $b_0 = \sqrt{R^2 - (d/2 + \delta)^2}$. It lies in the overlap. Since $R/d = r\gamma > 1/2$ is fixed,

$$\frac{\mathrm{vol}(A)}{\mathrm{vol}(B(0,R))} \ge \frac{2\delta V_{n-1} b_0^{n-1}}{V_n R^n} \ge n^{-O(1)} \left(1 - \frac{1}{4r^2\gamma^2}\right)^{n/2}.$$

Only the finite counting version below is used for actual queries.

Let $Q_B$ be the product of the input denominators and take $h = (Q_B 2^k)^{-1}$, so $L \subset h\mathbb{Z}^n$. On $\Omega_h = h\mathbb{Z}^n \cap B(0,R)$, translation by $v \in L$ preserves the grid exactly. Summing the indicator conflict gives

$$\Pr_{e \sim \mathrm{Unif}(\Omega_h)}[I(e) = 1] \ge \frac{|\Omega_h \cap (\Omega_h - v)|}{2|\Omega_h|}. \tag{215}$$

Grid cubes of radius $\delta_h = \sqrt{n}h/2$ sandwich ball counts between the volumes at radii $R \pm \delta_h$. The overlap count has error at most two boundary layers, of relative volume $O(n\delta_h/R)$. Choose $k = O(n + \mathcal{B} + \log Q)$ so this error is at most one tenth of the preceding overlap lower bound. A rational radius approximation with relative error $2^{-Cn}$ preserves that estimate and the closed NO radius bound. The finite conflict probability is therefore at least $n^{-O(1)}2^{-c(r,\gamma)n}$.

Efficient grid ball sampling avoids rejection from an enclosing cube. Draw continuously from $B(0, R + \delta_h)$, round to the nearest grid point, and retain the point precisely when it lies in $\Omega_h$. Every accepted grid cube lies entirely in the expanded ball and has volume $h^n$. Thus the conditional reference law is exactly uniform on $\Omega_h$, and its acceptance probability is $1 - O(n\delta_h/R)$. The continuous ball draw is $(R + \delta_h)U_0^{1/n} Z/\|Z\|$, with independent uniform $U_0 \in [0,1]$ and standard Gaussian $Z$. Certified finite approximations control both ball boundaries and rounding cell boundaries. A coordinate error much smaller than $h\zeta/(Q\,\mathrm{poly}(n))$ makes total rounding disagreement at most $\zeta$, using the bounded coordinate density and a union bound over its grid boundary strips. Gaussian tails and the small-norm cutoff are handled as for perturbations. Cap each rejection run at

$O(\log Q + \log(1/\zeta))$ attempts. All generation overflows then cost at most another $\zeta$. The finite program is coupled to a discrete reference distribution, so (215) remains exact.

Take $Q = \mathrm{poly}(n)2^{c(r,\gamma)n}$ trials with a sufficient polynomial factor. On YES, conditional on any fixed advice, the miss probability is at most $1/100$ plus implementation error. On NO it is bounded by the single bad advice probability and implementation error. Constant independent whole run repetitions and majority vote give success at least $2/3$. Only the advice, current test point, decomposition, and counters are stored, giving the stated $P + QD$ time and space. □

When Theorem 12.1 supplies the advice, one may instead construct it at the known scale $\gamma d$ on a GapSVP input. Only NO requires legal Gaussian widths. On YES every bounded preprocessing result still defines a fixed function, and the same conflict proof applies. Keeping the query function deterministic is essential for both formulations.

## AI methodology

Building on our earlier work [GFH26a, GFH26b], the authors used Codex (GPT-5.6-sol, GPT-6.1-sol, and GPT-6-astra) to explore provable algorithms for SVP and DGS below the $2^{n/2}$ barrier. We initially aimed to make BDGL sieving provable and proposed tracking the first two counting moments under repeated differencing as a dynamical system. With assistance from GPT-5.6-sol in algorithm design and mathematical analysis, we obtained an SVP algorithm with running time below $2^{n/2}$ on September 10, 2026. We subsequently named the framework Gaussian difference dynamics (GDD), refined the algorithms, and developed the single sample DGS and SVP results in Theorems 1.1 and 1.3 and their consequences.

After reading Doulgerakis and Laarhoven [DL26], we recognized that their rejection sampling method could extend the single sample construction to Gaussian batches (Theorem 1.2). We worked with GPT-6-astra to design the batch algorithm and develop its proof, then worked with GPT-6.1-sol to organize and draft the manuscript. We thoroughly checked and revised the entire paper and remain responsible for all mathematical claims, proofs, and references.

## References


[ACKLS21] Divesh Aggarwal, Yanlin Chen, Rajendra Kumar, Zeyong Li, and Noah Stephens-Davidowitz. Dimension-preserving reductions between SVP and CVP in different $p$-norms. In Proceedings of SODA, pages 2444–2462, 2021. https://doi.org/10.1137/1.9781611976465.145.

[ADRS15] Divesh Aggarwal, Daniel Dadush, Oded Regev, and Noah Stephens-Davidowitz. Solving the shortest vector problem in $2^n$ time using discrete Gaussian sampling. In Proceedings of STOC, pages 733–742, 2015. https://doi.org/10.1145/2746539.2746606.

[ALNS20] Divesh Aggarwal, Jianwei Li, Phong Q. Nguyen, and Noah Stephens-Davidowitz. Slide reduction revisited—filling the gaps in SVP approximation. In Advances in Cryptology—CRYPTO, pages 274–295, 2020. https://doi.org/10.1007/978-3-030-56880-1_10.

[AS18] Divesh Aggarwal and Noah Stephens-Davidowitz. Just take the average! An embarrassingly simple $2^n$ time algorithm for SVP (and CVP). In Proceedings of SOSA, volume 61 of OASIcs, pages 12:1–12:19, 2018. https://doi.org/10.4230/OASIcs.SOSA.2018.12.

[AKS04] Manindra Agrawal, Neeraj Kayal, and Nitin Saxena. PRIMES is in P. Annals of Mathematics, 160(2):781–793, 2004. https://doi.org/10.4007/annals.2004.160.781.

[AKS01] Miklós Ajtai, Ravi Kumar, and D. Sivakumar. A sieve algorithm for the shortest lattice vector problem. In Proceedings of STOC, pages 601–610, 2001. https://doi.org/10.1145/380752.380857.

[Ban93] Wojciech Banaszczyk. New bounds in some transference theorems in the geometry of numbers. Mathematische Annalen, 296:625–635, 1993. https://doi.org/10.1007/BF01445125.

[BDGL16] Anja Becker, Léo Ducas, Nicolas Gama, and Thijs Laarhoven. New directions in nearest neighbor searching with applications to lattice sieving. In Proceedings of SODA, pages 10–24, 2016. https://doi.org/10.1137/1.9781611974331.ch2.

[BGJ15] Anja Becker, Nicolas Gama, and Antoine Joux. Speeding-up lattice sieving without increasing the memory, using sub-quadratic nearest neighbor search. Cryptology ePrint Archive, Paper 2015/522, 2015. https://eprint.iacr.org/2015/522.

[DRS14] Daniel Dadush, Oded Regev, and Noah Stephens-Davidowitz. On the closest vector problem with a distance guarantee. In Proceedings of CCC, pages 98–109, 2014. https://doi.org/10.1109/CCC.2014.18.

[DL26] Emmanouil Doulgerakis and Thijs Laarhoven. Smooth sailing through spherical shells: provable random-lattice sieving in time $2^{0.292n}$. arXiv:2609.37482, 2026. https://arxiv.org/abs/2609.37482.

[Gao26] Yiming Gao. $\ell_\infty$ norm SVP in $\widetilde{O}(3^n)$. arXiv:2611.00811, 2026. https://arxiv.org/abs/2611.00811.

[GFH26a] Yiming Gao, Yansong Feng, and Honggang Hu. Solving the shortest vector problem in $2^{0.7314n+o(n)}$ time via discrete Gaussian sampling on superlattices. Cryptology ePrint Archive, Paper 2026/1587, 2026. https://eprint.iacr.org/2026/1587.

[GFH26b] Yiming Gao, Yansong Feng, and Honggang Hu. Discrete Gaussian sampling meets BDGL decoding: solving the shortest vector problem in $2^{0.5596n+o(n)}$ time. Cryptology ePrint Archive, Paper 2026/1844, 2026. https://eprint.iacr.org/2026/1844.

[GH26] Yiming Gao and Honggang Hu. Faster matrix multiplication leads to faster lattice sieving. Cryptology ePrint Archive, Paper 2026/2622, 2026. https://eprint.iacr.org/2026/2622.

[GPV08] Craig Gentry, Chris Peikert, and Vinod Vaikuntanathan. Trapdoors for hard lattices and new cryptographic constructions. In Proceedings of STOC, pages 197–206, 2008. https://doi.org/10.1145/1374376.1374407.

[HPS11] Guillaume Hanrot, Xavier Pujol, and Damien Stehlé. Analyzing blockwise lattice algorithms using dynamical systems. In Advances in Cryptology—CRYPTO, pages 447–464, 2011. https://doi.org/10.1007/978-3-642-22792-9_25.

[HS11] Sariel Har-Peled and Micha Sharir. Relative $(p, \varepsilon)$-approximations in geometry. Discrete & Computational Geometry, 45(3):462–496, 2011. https://doi.org/10.1007/s00454-010-9248-1.

[Hhan26a] Minki Hhan. Solving the shortest vector problem in time $2^{0.6039n}$ via mid-point Hessian. arXiv:2608.02478, 2026. https://arxiv.org/abs/2608.02478.

[Hhan26] Minki Hhan. Finding a shortest vector and more in $2^{n/2+o(n)}$ time using $q$-ary coset difference tree. arXiv:2609.02764, 2026. https://arxiv.org/abs/2609.02764.

[KL78] Grigorii A. Kabatiansky and Vladimir I. Levenshtein. On bounds for packings on a sphere and in space. Problems of Information Transmission, 14(1):1–17, 1978.

[Kim26] Jiseung Kim. One discrete Gaussian sample in $2^{n/2+o(n)}$ time. arXiv:2608.03220, 2026. https://arxiv.org/abs/2608.03220.

[Klein00] Philip N. Klein. Finding the closest lattice vector when it is unusually close. In Proceedings of SODA, pages 937–941, 2000.

[LLL82] Arjen K. Lenstra, Hendrik W. Lenstra, Jr., and László Lovász. Factoring polynomials with rational coefficients. Mathematische Annalen, 261(4):515–534, 1982. https://doi.org/10.1007/BF01457454.

[LN25] Jianwei Li and Phong Q. Nguyen. A complete analysis of the BKZ lattice reduction algorithm. Journal of Cryptology, 38(1):12, 2025. https://doi.org/10.1007/s00145-024-09527-0.

[LM09] Vadim Lyubashevsky and Daniele Micciancio. On bounded distance decoding, unique shortest vectors, and the minimum distance problem. In Advances in Cryptology—CRYPTO, pages 577–594, 2009. https://doi.org/10.1007/978-3-642-03356-8_34.

[MR07] Daniele Micciancio and Oded Regev. Worst-case to average-case reductions based on Gaussian measures. SIAM Journal on Computing, 37(1):267–302, 2007. https://doi.org/10.1137/S0097539705447360.

[MV13] Daniele Micciancio and Panagiotis Voulgaris. A deterministic single exponential time algorithm for most lattice problems based on Voronoi cell computations. SIAM Journal on Computing, 42(3):1364–1391, 2013. https://doi.org/10.1137/100811970.

[NV08] Phong Q. Nguyen and Thomas Vidick. Sieve algorithms for the shortest vector problem are practical. Journal of Mathematical Cryptology, 2(2):181–207, 2008. https://doi.org/10.1515/JMC.2008.009.

[PS26] Amaury Pouly and Yixin Shen. Solving the shortest vector problem in $2^{0.63269n+o(n)}$ time on random lattices. In Advances in Cryptology—EUROCRYPT, volume 16544 of LNCS, pages 92–123, 2026. https://doi.org/10.1007/978-3-032-25327-9_4.

[PS09] Xavier Pujol and Damien Stehlé. Solving the shortest lattice vector problem in time $2^{2.465n}$. Cryptology ePrint Archive, Paper 2009/605, 2009. https://eprint.iacr.org/2009/605.

[RS62] J. Barkley Rosser and Lowell Schoenfeld. Approximate formulas for some functions of prime numbers. Illinois Journal of Mathematics, 6(1):64–94, 1962. https://doi.org/10.1215/ijm/1255631807.

[SD16] Noah Stephens-Davidowitz. Discrete Gaussian sampling reduces to CVP and SVP. arXiv:1506.07490, 2016. https://arxiv.org/pdf/1506.07490.

[Wang24] Leizhang Wang. Analyzing Pump and Jump BKZ algorithm using dynamical systems. In Post-Quantum Cryptography—PQCrypto, pages 406–432, 2024. https://doi.org/10.1007/978-3-031-62743-9_14.

[ZZ20] Doron Zeilberger and Wadim Zudilin. The irrationality measure of $\pi$ is at most 7.103205334.... Moscow Journal of Combinatorics and Number Theory, 9(4):407–419, 2020. https://doi.org/10.2140/moscow.2020.9.407.

[ZWW26] Yang Zhou, Zhedong Wang, and Mingsheng Wang. Exact SVP on Haar-random lattices at the heuristic sieving exponents. Cryptology ePrint Archive, Paper 2026/2271, 2026. https://eprint.iacr.org/2026/2271.