IACR Transactions on Cryptographic Hardware and Embedded Systems
ISSN 2569-2925, Vol. 2026, No. 4, pp. 397–422. DOI:10.46586/tches.v2026.i4.397-422

Refined Approx-SVP Rank Reduction Conditions and Adaptive Lattice Reduction for MSIS Security Estimation

Xiaohan Zhang, Zijian Zhou and Longjiang Qu(✉)

College of Science, National University of Defense Technology, Changsha 410073, Hunan, China
ljqu_happy@hotmail.com

Abstract. The security of lattice-based cryptography relies critically on the concrete hardness of the approximate shortest vector problem (Approx-SVP). For cryptographic-sized instances, existing Approx-SVP rank reduction conditions may be overly aggressive, as they implicitly assume access to a large number of extremely short lattice vectors. In this work, we systematize and refine Approx-SVP rank reduction conditions from a feasibility perspective. We identify that, in the context of the dimension-for-free (D4f) technique, the existence of a single sufficiently short vector is the essential requirement, and we derive two refined and compact rank reduction conditions accordingly. The first condition is based on geometric properties of lattice sieving, while the second incorporates a basis-quality-dependent probabilistic bound. These conditions are evaluated on high-dimensional q-ary lattice challenges, where the compact condition yields finite-dimensional speedups of up to a factor of 60 relative to the practical rank reduction condition in dimensions 850 and 925. To realize these conditions in high dimensions, we present APBKZ, an adaptive Pump-based lattice reduction strategy that dynamically selects the blocksize and D4f parameters according to the evolving Gram–Schmidt profile. We further introduce HeadAPBKZ, a head-focused execution mode that restricts reduction to a critical prefix once the rank reduction condition is satisfied. Combining these advances, we develop a refined concrete security estimation framework for the MSIS problem. Applied to Dilithium, our model-dependent estimates indicate that, when compact rank reduction behavior is integrated with the D4f technique, the estimated concrete security is reduced by 3.65–6.09 bits relative to the conservative Core-SVP baseline. We also derive an analytic approximation whose closed-form estimates remain close to the numerical estimation, with discrepancies of only about 1.1 bits.

Keywords: Approximate Shortest Vector Problem · Rank Reduction · Lattice Reduction · BKZ Algorithms · MSIS Security Estimation · Post-Quantum Cryptography

1 Introduction

Lattice-based cryptography is a cornerstone of post-quantum cryptography, with security mainly based on the hardness of problems such as the Learning with Errors (LWE) problem [Reg09] and the Short Integer Solution (SIS) problem [Ajt99]. In concrete security evaluations, these problems are typically reduced to instances of the approximate shortest vector problem (Approx-SVP). As a result, the reliability of cryptographic parameter selection critically depends on how accurately the concrete feasibility of Approx-SVP can be assessed under realistic algorithmic assumptions.

Licensed under Creative Commons License CC-BY 4.0.
Received: 2026-04-15 Accepted: 2026-07-15 Published: 2026-09-11

Rank reduction for Approx-SVP. Modern lattice attacks do not attempt to solve Approx-SVP in the full lattice dimension. Instead, they rely on *rank reduction*: after sufficient basis reduction, the attacker searches for a short vector in a lower-dimensional prefix sublattice [LM18, ADH+19, DSvW21, ZDY25]. The feasibility of such attacks is governed by a *rank reduction condition*, which specifies how small the prefix dimension k can be while still enabling Approx-SVP with a given approximation factor.

A prominent rank reduction condition for Approx-SVP was proposed by Laarhoven and Mariano [LM18] in the context of progressive lattice sieving. However, this condition implicitly assumes that the sieving procedure can reliably access an exponentially large set of vectors with norms close to the Gaussian Heuristic. This is a strong geometric assumption for practical high-dimensional lattice reduction, where feasibility is determined by the existence of only a small number of sufficiently short vectors.

From idealized rank reduction to practical feasibility. In practice, lattice sieving typically returns vectors whose norms are bounded by a constant factor above the Gaussian Heuristic, most commonly within $\sqrt{4/3}$. The above analysis builds on the geometric modeling of lattice sieving introduced by Nguyen and Vidick [NV08] and further refined in subsequent works such as [MV10]. Under this model, short lattice vectors are treated as samples from a thin spherical shell, and the success of sieving is linked to the existence of sufficiently many vector pairs with favorable angular relations. This framework has become standard in the analysis of modern sieving-based SVP algorithms [WLTB11, ZPH13, BGJ14, Laa15, BGJ15, LdW15, BL16, BDGL16, LM18, ADH+19, DSvW21, ZDY25].

In particular, when the dimension-for-free (D4f) technique is employed to reduce the cost of sieving [Duc18], the sieving does not require access to all short vectors, but only to the existence of *at least one* vector below a certain length threshold [BDK+21].

This gap between idealized rank reduction and practical feasibility motivates a refined formulation of Approx-SVP rank reduction conditions for cryptographic-sized lattices. In this work, we develop two such conditions that explicitly distinguish what is required for *feasibility* from what is merely sufficient in theory.

Lattice reduction in high dimensions. Accurately realizing refined rank reduction conditions in high dimensions requires lattice reduction algorithms that behave robustly under practical constraints. Existing BKZ variants, such as PnJBKZ within the G6K framework, rely on simulator-driven parameter choices, fixed jump schedules and aggressive D4f estimates [ADH+19]. In high dimensions, these strategies frequently lead to unsaturated reduction and unstable basis quality, undermining the feasibility assumptions required for rank reduction.

Motivated by these limitations, we develop *adaptive lattice reduction strategies* that dynamically adjust their behavior based on the evolving quality of the basis. Our proposed algorithm, APBKZ (Adaptive Pump BKZ), adaptively selects blocksizes and dimension-for-free parameters without relying on simulator predictions or jump scheduling. Building on this approach, we further introduce HeadAPBKZ, a head-focused execution mode that restricts reduction to a critical prefix once a compact rank reduction condition is satisfied, reflecting empirical observations that late-stage tail reduction has negligible impact on feasibility [ZD23].

In this framework, APBKZ is of independent interest as a general-purpose adaptive lattice reduction strategy that does not rely on any rank reduction condition. It therefore applies beyond the specific feasibility conditions, while HeadAPBKZ should be viewed as a refinement of APBKZ rather than a replacement. Specifically, HeadAPBKZ becomes useful only after the compact rank reduction condition is satisfied, at which point the reduction scope can be restricted to the head region to obtain a more favorable cost–quality trade-off.

Security estimation of MSIS. A refined understanding of Approx-SVP rank reduction has direct implications for the concrete security estimation of SIS and MSIS-based cryptographic schemes. Notably, the Dilithium specification [BDK+21] highlights the lack of security estimates that accurately account for modern reduction techniques such as D4f and progressive sieving. Existing estimates based on the conservative Core-SVP model implicitly assume access to vectors at the Gaussian Heuristic bound, which may not reflect realistic attack feasibility.

By combining the compact rank reduction condition with adaptive lattice reduction strategies, we develop a refined security estimation framework for MSIS. The resulting Dilithium estimates should be understood as model-dependent concrete cost estimates that complement, rather than replace, conservative Core-SVP estimates.

This work does not prove new worst-case or asymptotic rank reduction theorems. Rather, it studies practical feasibility thresholds for sieve-based reduction in cryptographic parameter regimes, separating the geometric feasibility conditions, the adaptive reduction procedures used to realize them, and the MSIS cost estimates derived from the resulting model.

Our contributions. The main contributions of this work are summarized as follows:

- We systematize and refine Approx-SVP rank reduction conditions by replacing idealized assumptions with feasibility-oriented criteria that reflect the probabilistic behavior of lattice sieving. In particular, we introduce a practical rank reduction condition derived from geometric sieving bounds, and a compact rank reduction condition that incorporates basis-quality-dependent sieving guarantees.
- We introduce APBKZ, an adaptive lattice reduction strategy designed to robustly realize rank reduction feasibility in high dimensions, and HeadAPBKZ, a head-focused execution mode that exploits the compact rank reduction condition, mitigating early-dimension ineffectiveness and unsaturated reduction.
- We empirically validate the proposed rank reduction conditions and the robustness of the adaptive reduction strategies on high-dimensional q-ary lattices from the TU Darmstadt Lattice Challenge. The compact condition (Condition 2, see Section 3.4 for details) yields finite-dimensional speedups of up to a factor of 60 relative to the practical rank reduction condition (Condition 1, see Section 3.3 for details) in dimensions 850 and 925 under the tested parameters, as detailed in Table 1 and Section 5.
- Building on these results, we integrate the compact condition into a refined security estimation framework for MSIS and apply it to Dilithium. Compared to the conservative Core-SVP model, the resulting model-dependent estimates decrease by 3.65–6.09 bits. We also derive a continuous-shape analytic approximation whose final Dilithium estimates differ from the numerical estimation by only about 1.1 bits.

Table 1: Finite-dimensional speedups obtained by the compact rank reduction condition relative to the practical rank reduction condition on lattice challenges (wall times reported in days (d), hours (h), and minutes (m)).

Dim	n	Baseline wall time by Condition 1	Baseline wall time by Condition 2	Speedup factor
850	230	$1.40d$	$31.2m$	**64.6**×
925	295	$2.07h$	$2m$	**62.1**×
1025	330	$1.22d$	$1.89h$	15.5×

† The speedup factors compare Condition 2 with Condition 1 under the same experimental framework.

Organization. The paper is organized as follows. Section 2 provides the necessary preliminaries of lattice reduction and hard problems. Section 3 revisits Approx-SVP rank reduction and introduces refined feasibility conditions. Section 4 presents adaptive lattice reduction strategies, including APBKZ and HeadAPBKZ. Section 5 reports experimental results. Section 6 applies our framework to MSIS security estimation and evaluates Dilithium.

2 Preliminaries

2.1 Notations and Basic Definitions

In this paper, column vectors are denoted by bold lowercase letters (e.g., $\mathbf{b}$), and matrices are denoted by bold capital letters (e.g., $\mathbf{B}$). We write a matrix $\mathbf{B}$ as $\mathbf{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$, where $\mathbf{b}_i$ denotes the i-th column vector of $\mathbf{B}$. If $\mathbf{B} \in \mathbb{R}^{m \times n}$ has full column rank n, the lattice generated by $\mathbf{B}$ is denoted by $\mathcal{L}(\mathbf{B}) = \{\mathbf{Bx} \mid \mathbf{x} \in \mathbb{Z}^n\}$. Let $(\mathbf{b}_1^*, \ldots, \mathbf{b}_n^*)$ denote the Gram-Schmidt orthogonalization of the basis $(\mathbf{b}_1, \ldots, \mathbf{b}_n)$. For $i \in \{1, \ldots, n\}$, we denote by π_i the orthogonal projection onto the orthogonal complement of $\mathrm{span}(\mathbf{b}_1, \ldots, \mathbf{b}_{i-1})$. For $1 \leq i < j \leq n$, we define $\mathbf{B}_{[i:j]} = (\mathbf{b}_i, \ldots, \mathbf{b}_j)$ and denote by $\mathcal{L}_{[i:j]} = \mathcal{L}(\mathbf{B}_{[i:j]})$ the sublattice of $\mathcal{L}(\mathbf{B})$ generated by these basis vectors. In contrast, we define the locally projected block $\mathbf{B}_{\pi[i:j]} = (\pi_i(\mathbf{b}_i), \ldots, \pi_i(\mathbf{b}_j))$ and denote by $\mathcal{L}_{\pi[i:j]} = \mathcal{L}(\mathbf{B}_{\pi[i:j]})$ the lattice generated by the projected vectors. For convenience, we use $\mathbf{B}_{[i]}$ and $\mathcal{L}_{[i]}$ as shorthands for $\mathbf{B}_{[i:n]}$ and $\mathcal{L}_{[i:n]}$, respectively, and similarly $\mathbf{B}_{\pi[i]}$ and $\mathcal{L}_{\pi[i]}$ as shorthands for $\mathbf{B}_{\pi[i:n]}$ and $\mathcal{L}_{\pi[i:n]}$.

Let BKZ-β denote one tour of BKZ with blocksize β. A basis is said to be BKZ-β reduced if it is obtained after sufficiently many tours of BKZ-β. For simplicity, we use β to denote the quality of a BKZ-β reduced basis. The Euclidean norm of a vector $\mathbf{v}$ is denoted by $\|\mathbf{v}\|$. The volume of a lattice $\mathcal{L}(\mathbf{B})$ is $\mathrm{Vol}(\mathcal{L}(\mathbf{B})) = \prod_i \|\mathbf{b}_i^*\|$, which is an invariant of the lattice. The first minimum of a lattice $\mathcal{L}$ is the length of a shortest non-zero vector, denoted by $\lambda_1(\mathcal{L})$. We use the abbreviations $\mathrm{Vol}(\mathbf{B}) = \mathrm{Vol}(\mathcal{L}(\mathbf{B}))$ and $\lambda_1(\mathbf{B}) = \lambda_1(\mathcal{L}(\mathbf{B}))$.

Definition 1 (Gaussian Heuristic [Ngu10])**.** Let $S \subset \mathbb{R}^n$ be a measurable body. Then the number of lattice points in S is approximately equal to $\mathrm{Vol}(S)/\mathrm{Vol}(\mathcal{L})$.

We denote by $|S \cap \mathcal{L}|$ the number of lattice points in S. Note that the number of lattice points predicted by the Gaussian Heuristic is exactly the expected number of lattice points in a random translation of S. When applying the Gaussian Heuristic to an n-dimensional ball of volume $\mathrm{Vol}(\mathcal{L})$ we obtain that the minimal length $\lambda_1(\mathcal{L})$ is approximately the radius of this ball, which asymptotically means that $\lambda_1(\mathcal{L}) \approx \sqrt{n/(2\pi e)} \cdot \mathrm{Vol}(\mathcal{L})^{1/n}$. For a lattice $\mathcal{L} \subset \mathbb{R}^n$ we denote this radius by $\mathrm{gh}(\mathcal{L})$, and we also denote $\mathrm{gh}(n)$ for $\mathrm{gh}(\mathcal{L})$ of any n-dimensional lattice $\mathcal{L}$ of volume 1: $\mathrm{gh}(n) = \sqrt{n/2\pi e}$. Note that a ball of radius $\omega \cdot \mathrm{gh}(\mathcal{L})$ contains an exponential number of ω^n lattice vectors not much longer than the minimal length. We say that a list of lattice vectors saturates a volume S if it contains some significant ratio [Duc18], say 50%, of the lattice vectors in $\mathcal{L} \cap S$ as predicted by the Gaussian Heuristic. Here the saturation ratio is defined as $p = N_{\mathrm{list}}/N_{\mathrm{GH}}$, with $N_{\mathrm{list}} = |\mathbf{v} \in \mathrm{list} : \mathbf{v} \in S|$ and $N_{\mathrm{GH}} = \mathrm{Vol}(S)/\mathrm{Vol}(\mathcal{L})$. And we call the list p-saturated when p reaches a predefined threshold.

Definition 2 (Root Hermite Factor [Che13])**.** Let $\mathbf{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ be a reduced basis for an n-dimensional lattice $\mathcal{L}$. The root Hermite factor is defined by

$$\delta = \left(\frac{\|\mathbf{b}_1\|}{\mathrm{Vol}(\mathcal{L})^{\frac{1}{n}}} \right)^{\frac{1}{n}}.$$

Definition 3 (HKZ reduction and BKZ reduction [SE94])**.** The basis $\mathbf{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ of a lattice $\mathcal{L}$ is said to be HKZ reduced if $\|\mathbf{b}_i^*\| = \lambda_1\left(\mathcal{L}\left(\mathbf{B}_{\pi[i:n]}\right)\right)$ for all $i \leq n$. It is said to be BKZ reduced with blocksize β (for short, BKZ-β reduced) if $\|\mathbf{b}_i^*\| = \lambda_1\left(\mathcal{L}\left(\mathbf{B}_{\pi[i:\min(i+\beta-1,n)]}\right)\right)$ for all $i \leq n$.

Based on Definition 3, the Geometric Series Assumption provides a simplified model for BKZ-reduced bases:

Definition 4 (Geometric Series Assumption (GSA) [Sch03])**.** Let $\mathbf{B}$ be a BKZ-β reduced basis of a lattice $\mathcal{L}$. The GSA states that:

$$\|\mathbf{b}_i^*\| = \|\mathbf{b}_1^*\| \cdot \alpha_\beta^{i-1},$$

where $\frac{\sqrt{3}}{2} \leq \alpha_\beta < 1$ and α_β is called GSA constant. This model is reasonably accurate in practice for $\beta > 50$ and $\beta \ll n$.

Our analysis also relies on fundamental probability distributions that characterize random lattice properties:

Definition 5 (Exponential Distribution)**.** For $\lambda > 0$, a continuous random variable X follows the exponential distribution Expo(λ) if its probability density function is given by

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

The cumulative distribution function is $F_X(x) = 1 - e^{-\lambda x}$ for $x \geq 0$.

Definition 6 (Gaussian distribution)**.** For $\mu \in \mathbb{R}$ and $\sigma > 0$, let $\mathcal{N}(\mu, \sigma^2)$ denote the Gaussian distribution with mean μ and variance σ^2, whose density is

$$\rho_{\mathcal{N}(\mu,\sigma^2)}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad x \in \mathbb{R}.$$

Its cumulative distribution function is

$$F_X(x) = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right),$$

where

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt.$$

These distributions play a crucial role in understanding the statistical properties of lattice parameters.

2.2 Lattice Hard Problems

We now introduce the fundamental computational problems that form the foundation of lattice-based cryptography. These problems are believed to be hard even for quantum computers, making them central to post-quantum cryptography.

Definition 7 (Shortest Vector Problem (SVP) [GN08])**.** The SVP is defined as: given a basis $\mathbf{B}$ of a lattice $\mathcal{L}$, finding the shortest non-zero vector $\mathbf{v}$ in $\mathcal{L}$.

Definition 8 (Approximate Shortest Vector Problem (Approx-SVP) [GN08])**.** The Approx-SVP is defined as: let $\gamma \geq 1$, given a basis $\mathbf{B}$ of a lattice $\mathcal{L}$, finding a non-zero vector $\mathbf{v} \in \mathcal{L}$ such that $\|\mathbf{v}\| \leq \gamma \cdot \lambda_1(\mathcal{L})$.

While SVP and Approx-SVP are fundamental lattice problems, modern cryptographic constructions typically rely on structured variants that offer better efficiency and security properties:

Definition 9 (Module-LWE Problem (MLWE) [BDK+21])**.** For integers m, k, and a probability distribution $D : \mathbb{R}_q \to [0, 1]$, we say that the advantage of algorithm $\mathcal{A}$ in solving the decisional $\mathrm{MLWE}_{m,k,D}$ problem over the ring $\mathbb{R}_q$ is

$$\begin{aligned}\mathrm{Adv}^{\mathrm{MLWE}}_{m,k,D}(\mathcal{A}) :=& \left| \Pr\left[b = 1 \mid \mathbf{A} \leftarrow \mathbb{R}_q^{m\times k}; \mathbf{t} \leftarrow \mathbb{R}_q^{m}; b \leftarrow \ \mathcal{A}(\mathbf{A}, \mathbf{t})\right]\right. \\ &\left. - \Pr\left[b = 1 \mid \mathbf{A} \leftarrow \mathbb{R}_q^{m\times k}; \mathbf{s}_1 \leftarrow D^k; \mathbf{s}_2 \leftarrow D^m; b \leftarrow \ \mathcal{A}\left(\mathbf{A}, \mathbf{A}\mathbf{s}_1 + \mathbf{s}_2\right)\right] \right| .\end{aligned}$$

Definition 10 (Short Integer Solution Problem (SIS) [BDK+21])**.** To an algorithm $\mathcal{A}$ we associate the advantage function $\mathrm{Adv}^{\mathrm{SIS}}_{m,k,\gamma}$, to solve the (Hermite Normal Form) $\mathrm{SIS}_{m,k,\gamma}$ problem over the ring $\mathbb{Z}_q$ as

$$\mathrm{Adv}^{\mathrm{SIS}}_{m,k,\gamma}(\mathcal{A}) := \Pr\left[0 < \|\mathbf{y}\|_\infty \le \gamma \wedge [\mathbf{I} \mid \mathbf{A}] \cdot \mathbf{y} = \mathbf{0} \mid \mathbf{A} \leftarrow \mathbb{Z}_q^{m\times k}; \mathbf{y} \leftarrow \mathcal{A}(\mathbf{A})\right] .$$

For vectors over $\mathbb{R}_q$, we interpret the infinity norm coefficientwise: each polynomial entry is represented by its centered integer coefficients, and $\|\mathbf{y}\|_\infty$ denotes the maximum absolute value among all coefficients of all entries.

Definition 11 (Module-SIS Problem (MSIS) [BDK+21])**.** To an algorithm $\mathcal{A}$ we associate the advantage function $\mathrm{Adv}^{\mathrm{MSIS}}_{m,k,\gamma}$, to solve the (Hermite Normal Form) $\mathrm{MSIS}_{m,k,\gamma}$ problem over the ring $\mathbb{R}_q$ as

$$\mathrm{Adv}^{\mathrm{MSIS}}_{m,k,\gamma}(\mathcal{A}) := \Pr\left[0 < \|\mathbf{y}\|_\infty \le \gamma \wedge [\mathbf{I} \mid \mathbf{A}] \cdot \mathbf{y} = \mathbf{0} \mid \mathbf{A} \leftarrow \mathbb{R}_q^{m\times k}; \mathbf{y} \leftarrow \mathcal{A}(\mathbf{A})\right] .$$

2.3 Lattice Sieving and Dimension for Free

Lattice sieving generates short lattice vectors by repeatedly combining vectors from a large list and retaining shorter reductions; heuristically, a sieve in dimension n uses lists of size about $(4/3)^{n/2+o(n)}$ and returns vectors close to the Gaussian-Heuristic scale [NV08, BDGL16].

The dimension-for-free (D4f) technique [Duc18] reduces the effective sieving dimension by projecting away the first f coordinates. If a target vector $\mathbf{v} \in \mathcal{L}$ has a projection $\pi_{f+1}(\mathbf{v})$ that is short enough to appear in the sieve output over the projected lattice $\mathcal{L}_{\pi[f+1:n]}$, then one may sieve in dimension $n - f$ and lift the projected vector back to a candidate vector in the original lattice. Thus D4f is justified only when the projected lattice still contains a sufficiently short vector that can be lifted to a valid solution. This is why the rank reduction conditions below are formulated as single-vector feasibility conditions rather than as assumptions about full list saturation.

3 Refined Rank Reduction for Approx-SVP

In this section, we revisit the Approx-SVP rank reduction condition proposed by Laarhoven and Mariano [LM18], which is commonly used as a feasibility criterion in concrete security analyses, but may be overly aggressive for cryptographic-sized lattice instances. In particular, this classical condition is based on single-vector geometric bounds and does not explicitly account for the probabilistic output behavior of modern sieve-based SVP oracles. The goal is not to prove a new asymptotic reduction theorem, but to isolate the finite-dimensional feasibility assumptions that underlie sieve-based rank reduction estimates in cryptographic parameter regimes.

Motivated by this limitation, we introduce two feasibility-oriented rank reduction conditions for Approx-SVP that reflect different levels of geometric modeling fidelity. The *practical* rank reduction condition is derived from the standard geometric model of lattice

sieving introduced by Nguyen and Vidick [NV08], and captures probabilistic feasibility thresholds based on single-vector geometric bounds within this framework. The *compact* rank reduction condition follows the same prefix-volume viewpoint, but further incorporates the dependence of sieve output quality on the reduction quality of the basis.

Throughout the remainder of the paper, these two conditions play distinct but complementary roles. The practical condition is primarily used for experimental analysis and validation, while the compact condition serves as the operative criterion for security estimation and algorithmic optimization in the subsequent sections.

3.1 Where Rank Reduction Enters BKZ Reduction

We begin with a high-level description of how rank reduction arises in the BKZ setting and how it interacts with the dimension-for-free technique.

BKZ reduction progressively improves the Gram–Schmidt profile of the basis by applying an SVP oracle to projected local blocks. As this process continues, a prefix sublattice $\mathcal{L}_{[1:k]}$ may already reach a quality at which a target Approx-SVP vector appears within this prefix.

At that stage, solving Approx-SVP no longer requires searching in the full lattice dimension, but only in the rank-k prefix sublattice. This is where rank reduction enters the BKZ algorithm. When dimension-for-free technique is used, the relevant cost is governed by the effective sieving dimension after projection, so feasibility depends on whether the reduced prefix already contains a sufficiently short recoverable vector.

The next subsection derives a practical rank reduction condition from the geometric sieve model, and the following subsection refines it by incorporating the dependence on basis quality.

3.2 Revisiting the Original Rank Reduction Condition

In [LM18], Laarhoven and Mariano provided a preliminary estimation of Approx-SVP rank reduction.

Under the Geometric Series Assumption, a BKZ-reduced lattice basis $(\mathbf{b}_1, \ldots, \mathbf{b}_n)$ satisfies $\|\mathbf{b}_i^*\| \approx \delta^{n-2i+2} \cdot \mathrm{Vol}(\mathcal{L})^{1/n}$ for all $i = 1, \ldots, n$. The first k basis vectors form a rank-k sublattice $\mathcal{L}_{[1:k]} = \mathcal{L}(\mathbf{b}_1, \ldots, \mathbf{b}_k) \subset \mathcal{L}$ of volume:

$$\mathrm{Vol}\left(\mathcal{L}_{[1:k]}\right) = \prod_{i=1}^{k} \|\mathbf{b}_i^*\| = \mathrm{Vol}(\mathcal{L})^{k/n} \cdot \delta^{k(n-k+1)}. \tag{1}$$

By the Gaussian Heuristic, the shortest vector in $\mathcal{L}_{[1:k]}$ has relative norm:

$$\mathrm{gh}(\mathcal{L}_{[1:k]}) \approx \delta^{(n-k+1)} \sqrt{\frac{k}{n}} \cdot \mathrm{gh}(\mathcal{L}). \tag{2}$$

To solve Approx-SVP with approximation factor γ, we require $\delta^{(n-k+1)}\sqrt{k/n} \approx \gamma$. This indicates how many dimensions can be effectively "skipped" when seeking approximate solutions.

3.3 Geometric Basis for Sieving Output Lengths

The original Approx-SVP rank reduction condition in [LM18] is based on the assumption that lattice sieving can reliably find vectors with norms close to the Gaussian Heuristic estimate $\mathrm{gh}(\mathcal{L})$. However, this represents a theoretical lower bound that is rarely achieved in practical high-dimensional lattice reduction. In practice, the success of rank reduction depends on the probabilistic behavior of lattice sieving rather than on achieving the

Gaussian Heuristic bound exactly. To capture this probabilistic behavior, we adopt a geometric perspective on lattice sieving to derive a more practical rank reduction condition.

Importantly, the geometric properties used in our analysis arise from the standard geometric model employed in the analysis of lattice sieving algorithms. In this model, the analysis of lattice sieving treats short vectors encountered during the algorithm as if they are sampled independently and uniformly from a thin spherical shell, a heuristic introduced by Nguyen and Vidick [NV08]. Within this geometric framework, Nguyen and Vidick showed that the expected number of vector pairs whose mutual angle is at most $\pi/3$ grows exponentially with the dimension. Equivalently, this implies that the number of vectors of norm at most $\sqrt{4/3} \cdot \mathrm{gh}(\mathcal{L}_{[1:k]})$ produced by lattice sieving is exponential in k, provided that the basis of $\mathcal{L}_{[1:k]}$ is sufficiently reduced.

In practice, D4f-enabled sieving does not benefit from enumerating all short vectors returned by the sieve. Instead, feasibility hinges on the existence of a single sufficiently short vector that can initiate the projection-and-lifting procedure. Once such a vector is available, additional short vectors do not further reduce the asymptotic cost [BDK+21]. Consequently, rank reduction feasibility should be formulated in terms of a single-vector requirement rather than full list saturation.

Accordingly, we require that lattice sieving on the rank-k prefix sublattice $\mathcal{L}_{[1:k]}$ produces at least one nonzero vector $\mathbf{x}$ satisfying

$$\|\mathbf{x}\| \;\leq\; \sqrt{\frac{4}{3}} \cdot \mathrm{gh}(\mathcal{L}_{[1:k]}).$$

Under the rank reduction viewpoint, we consider the regime where a valid Approx-SVP witness already occurs in the chosen prefix sublattice $\mathcal{L}_{[1:k]} \subseteq \mathcal{L}$. In this case, feasibility is ensured whenever

$$\gamma \cdot \mathrm{gh}(\mathcal{L}) \leq \sqrt{\frac{4}{3}} \cdot \mathrm{gh}\left(\mathcal{L}_{[1:k]}\right). \tag{3}$$

Combining the Gaussian Heuristic with inequalities (2) and (3), we derive:

$$\gamma \cdot \sqrt{\frac{n}{2\pi e}} \cdot \mathrm{Vol}(\mathcal{L})^{1/n} \leq \sqrt{\frac{4}{3}} \cdot \sqrt{\frac{k}{2\pi e}} \cdot \mathrm{Vol}(\mathcal{L})^{1/n} \cdot \delta^{(n-k+1)}, \tag{4}$$

The existence of such a vector in $\mathcal{L}_{[1:k]}$ therefore requires:

$$\gamma \leq \sqrt{\frac{4k}{3n}} \cdot \delta^{(n-k+1)}. \tag{5}$$

Condition 1 (Practical Approx-SVP Rank Reduction Condition)**.** *Let $\mathcal{L}$ be a lattice of dimension n with a BKZ-β reduced basis and root Hermite factor δ. Let γ denote the approximation factor of the target vector relative to* $\mathrm{gh}(\mathcal{L})$*. A rank-k prefix sublattice $\mathcal{L}_{[1:k]}$ is said to satisfy the* practical Approx-SVP rank reduction condition *if Inequality* (5) *holds.*

After sufficient BKZ reduction satisfying GSA, the desired short vector can be found by searching only within the first k dimensions. For example, in an 850-dimensional q-ary lattice with root Hermite factor 1.668906, a short vector with approximation factor 1.007991 can be obtained by searching only the rank-164 sublattice.

3.4 Basis-Quality-Dependent Sieving Output Assumption

In this subsection, we formalize a probabilistic model for the output behavior of lattice sieving. Rather than assuming that a sieve deterministically returns a shortest vector, we model the sieve as a sampling procedure over short lattice vectors, whose output distribution depends on the quality of the reduced basis.

Concretely, we view the sieve as inducing a probability distribution over lattice vectors in $\mathcal{L}_{[1:k]}$, supported on vectors whose norms are bounded by a small constant multiple of the Gaussian Heuristic. Conditioned on the basis quality, a single invocation of the sieve outputs a vector $\mathbf{v} \in \mathcal{L}_{[1:k]}$ satisfying $\|\mathbf{v}\| \leq B \cdot \mathrm{gh}(\mathcal{L}_{[1:k]})$ with non-negligible probability, where the bound B depends only on the reduction quality and the blocksize.

Importantly, our analysis does not assume that the sieve outputs all short vectors or succeeds deterministically. Instead, it relies solely on the existence of at least one sufficiently short output per sieve invocation. As a result, the rank reduction condition introduced below depends only on a non-negligible success probability and is independent of implementation-specific details of the sieve.

Lemma 1 ([Che13])**.** *Let $\Gamma_n = \{\mathcal{L} \in \mathbb{R}^n \mid \mathrm{Vol}(\mathcal{L}) = 1\}$ be the set of all full-rank lattices of rank n with unit volume. For a random lattice $\mathcal{L}$ uniformly sampled from the set Γ_n of n-dimensional lattices with unit volume, the scaled first minimum $v_n \cdot \lambda_1(\mathcal{L})^n$ converges in distribution to the exponential distribution* $\mathrm{Expo}(1/2)$ *as the dimension $n \to \infty$, where v_n denotes the volume of the n-dimensional unit ball.*

Heuristic 1 (Basis-Quality-Dependent Sieve Output)**.** *Given an n-dimensional q-ary random lattice $\mathcal{L}$ with a BKZ-β reduced basis, lattice sieving produces, with very high probability, a nonzero vector $\mathbf{x} \in \mathcal{L}$ satisfying*

$$\|\mathbf{x}\| \leq (8\ln(10))^{1/\beta} \cdot \mathrm{gh}(\mathcal{L}),$$

for blocksize $\beta > 50$, where the Gaussian Heuristic accurately captures lattice geometry.

Here, the basis quality is characterized by the BKZ blocksize β. A larger β corresponds to higher basis quality and yields a smaller exponential factor $(8\ln(10))^{1/\beta}$ in the norm upper bound, resulting in shorter output vectors from lattice sieving. This heuristic thus captures how the output quality of sieving depends on the preprocessing quality of the lattice basis in the context of concrete cryptographic instances. We explicitly note that this probabilistic bound is intended to model the behavior of sieving in practice and does not imply a polynomial-time algorithm for Approx-SVP in the asymptotic sense, as the required blocksize β for validity grows with n. In practical settings considered here, solving Approx-SVP via sieving still incurs exponential computational cost.

We now justify the bound given in Heuristic 1. For a β-dimensional random lattice $\mathcal{L}$ of unit volume, Lemma 1 shows that the first minimum satisfies:

$$v_\beta \cdot \lambda_1(\mathcal{L})^\beta \sim \mathrm{Expo}(1/2), \tag{6}$$

where v_β denotes the volume of the β-dimensional unit ball.

For the exponential distribution $\mathrm{Expo}(1/2)$, the cumulative distribution function gives:

$$P\left(v_\beta \cdot \lambda_1(\mathcal{L})^\beta \leq t\right) = 1 - e^{-t/2}. \tag{7}$$

Setting a high success probability (e.g., above 0.9999), we obtain:

$$1 - e^{-t/2} = 0.9999 \Rightarrow t = -2\ln(0.0001) = 8\ln(10). \tag{8}$$

Thus, the probabilistic upper bound satisfies:

$$v_\beta \cdot (\lambda_1^{\max})^\beta = 8\ln(10) \Rightarrow \lambda_1^{\max} = \left(\frac{8\ln(10)}{v_\beta}\right)^{1/\beta}. \tag{9}$$

Note that the Gaussian Heuristic indicates $\mathrm{gh}(\mathcal{L}) = v_\beta^{-1/\beta}$ for a unit-volume lattice, we have

$$\lambda_1^{\max} = (8\ln(10))^{1/\beta} \cdot \mathrm{gh}(\mathcal{L}). \tag{10}$$

By applying this bound to each local block of dimension β encountered during BKZ-β reduction, we obtain, with high confidence, a sieving capability ensuring the existence of a nonzero vector $\mathbf{x}$ satisfying

$$\|\mathbf{x}\| \leq (8\ln(10))^{1/\beta} \cdot \mathrm{gh}(\mathcal{L}). \tag{11}$$

In particular, for a rank-k prefix sublattice $\mathcal{L}_{[1:k]}$, applying the same sieve-output heuristic to the local sublattice yields the existence of a vector $\mathbf{x} \in \mathcal{L}_{[1:k]}$ such that

$$\|\mathbf{x}\| \leq (8\ln(10))^{1/\beta} \cdot \mathrm{gh}(\mathcal{L}_{[1:k]}). \tag{12}$$

Expressing the Gaussian Heuristic of $\mathcal{L}_{[1:k]}$ in terms of the Gram–Schmidt profile of a BKZ-β reduced basis yields the feasibility inequality

$$\gamma \leq (8\ln(10))^{1/\beta} \cdot \sqrt{\frac{k}{n}} \cdot \delta^{(n-k+1)}. \tag{13}$$

Condition 2 (Compact Approx-SVP Rank Reduction Condition)**.** *Let $\mathcal{L}$ be a q-ary lattice of dimension n, let β be the BKZ blocksize with root Hermite factor δ, and let γ be the target approximation factor. A rank-k prefix sublattice is said to satisfy the* compact Approx-SVP rank reduction condition *if Inequality* (13) *holds.*

The analysis above is therefore primarily theoretical. In practice, lattice reduction implementations often employ additional implementation-level heuristics to improve empirical performance, which may lead to successful reductions beyond the predicted feasibility thresholds.

Where the q-ary structure enters. The rank reduction conditions themselves are derived from prefix-volume estimates, the Gaussian Heuristic, and a probabilistic model of sieve outputs; as such, the formal inequalities are not intrinsically restricted to q-ary lattices. In this work, the q-ary structure enters through the concrete reduction regime used to calibrate and validate the conditions. A q-ary basis contains initial q-vectors of norm q, and this structured head region strongly influences the early Gram–Schmidt profile, the reliability of local Gaussian-Heuristic estimates, and the behavior of adaptive blocksize and D4f selection.

Consequently, our empirical constants and adaptive rules are justified for the q-ary SIS-style challenge instances studied here. The same methodology may extend to non-q-ary lattices, including NTRU-type lattices, but only after a separate calibration of the corresponding basis profiles, early structured regions, and sieve-output behavior. Such a study is left for future work.

Remark 1. Re-randomization refers to multiplying the current lattice basis by a random unimodular matrix, typically instantiated as a random permutation matrix, followed by light reduction. This technique is widely adopted in practical lattice reduction algorithms such as BKZ 2.0 [CN11] and PBKZ [AWHT16], and it also appears in concrete security evaluations such as Dilithium and recent practical sieving studies [BDK+21, ZDY25].

While a formal proof remains open, our experiments on cryptographic-sized lattices indicate that re-randomization aids in escaping local minima and can improve the prefix shape seen by later reductions; heuristically, this translates into a smaller prefix volume $\mathrm{Vol}(\mathcal{L}_{[1:k]})$ and therefore a more favorable rank reduction condition. We therefore treat this improvement as an empirical heuristic rather than as a formal guarantee.

Which rank reduction condition is used later. Throughout the remainder of this paper, we explicitly distinguish between the roles of the two rank reduction conditions introduced in Section 3.

The *practical* Approx-SVP rank reduction condition (Condition 1) is used to interpret reduction behavior through the geometric lens of lattice sieving, as predicted by the standard spherical-shell model underlying modern sieve analyses. It captures feasibility at the level of single-vector geometric bounds, without explicitly accounting for basis-quality-dependent amplification, as validated in Section 5.2.

The *compact* Approx-SVP rank reduction condition (Condition 2) is adopted as the operative feasibility criterion for theoretical modeling, security estimation, and optimization procedures in the remainder of this work, including the experimental analysis in Section 5.2 and 5.3.2, as well as the MSIS security evaluation in Section 6. This choice reflects its conservative nature and its derivation from basis-quality-dependent sieving guarantees.

4 Adaptive Lattice Reduction Strategies in High Dimensions

In this section, we present adaptive lattice reduction strategies aimed at improving the stability and robustness of lattice reduction in high dimensions. Unlike classical BKZ variants that rely on simulator-driven parameter choices and fixed reduction schedules, the strategies introduced here adapt their behavior dynamically based on the evolving quality of the lattice basis.

In particular, we introduce APBKZ as a general-purpose adaptive reduction strategy, and a head-focused variant, HeadAPBKZ, which restricts the reduction scope once a compact rank reduction condition has been satisfied. Together, these strategies form a flexible reduction framework suitable for high-dimensional lattice reduction and subsequent security evaluation.

4.1 APBKZ: Adaptive Pump BKZ for Stable Lattice Reduction

The PnJBKZ algorithm is among the most efficient lattice reduction methods in high dimensions. Its performance is governed by three key parameters: the blocksize β, the dimension-for-free parameter f, and the jump parameter j. These parameters determine how G6K invokes the core reduction primitive $\mathrm{Pump}_{\kappa,f,\beta}$ [ADH+19]: β fixes the local block $[\kappa : \kappa + \beta - 1]$, f specifies how many projected coordinates are skipped by the D4f technique, and j controls the shift between consecutive pump calls. In practice, β and j are typically selected using a BKZ simulator based on the Gaussian Heuristic and predicted Gram-Schmidt profiles, while f is chosen according to D4f estimates that aim to recover vectors close to $\mathrm{gh}(\mathcal{L})$.

However, the effectiveness of PnJBKZ in high dimensions critically depends on simulator-driven parameter choices and static reduction schedules, both of which may become unreliable when the actual basis quality deviates from Gaussian Heuristic predictions. In particular, inaccurate blocksize selection, overly optimistic dimension-for-free estimates, and aggressive jump strategies often lead to unsaturated reduction, especially in early blocks of high-dimensional q-ary lattices.

Compared to classical PnJBKZ, APBKZ differs in three fundamental aspects. First, it eliminates simulator-driven blocksize prediction and instead derives local blocksize directly from the evolving Gram–Schmidt profile of the basis. Second, it replaces static dimension-for-free estimates with basis-quality-dependent rules that adapt to the current reduction state, thereby avoiding reliance on idealized HKZ-reduced assumptions. Third, it enforces full coverage of early blocks during each reduction tour, preventing unsaturated reduction caused by aggressive jump scheduling.

These design choices reflect the central philosophy of APBKZ: lattice reduction in high dimensions should be guided by the actual quality of the current basis, rather than by fixed schedules or simulator-based predictions. In particular, APBKZ does not depend on

any rank reduction condition. This is why it is introduced and evaluated separately from HeadAPBKZ, which is activated only after the compact condition becomes relevant.

APBKZ is governed by three adaptive design rules, which jointly determine its reduction behavior.

Rule 1: Adaptive blocksize selection. For a given index i, APBKZ selects the local blocksize β_i as the smallest value such that $\mathrm{gh}(\mathcal{L}_{\pi[i:i+\beta_i-1]}) < \|\mathbf{b}_i^*\|$. This condition ensures that the current block is sufficiently reduced to allow further improvement, while avoiding unnecessarily large blocksizes that incur excessive computational cost.

Rule 2: Adaptive dimension-for-free estimation. Given the selected blocksize β_i, the dimension-for-free parameter f_i is chosen as the largest value satisfying

$$\sqrt{\frac{4}{3}} \cdot \mathrm{gh}(\mathcal{L}_{\pi[i+f_i:i+\beta_i-1]}) \;\geq\; \|\mathbf{b}_i^*\|.$$

Unlike classical D4f rules that target vectors close to $\mathrm{gh}(\mathcal{L})$, this criterion only requires outperforming the current Gram–Schmidt norm. As a result, multiple vectors may qualify in practice, which mitigates unsaturated reduction. Traditional D4f estimates instead assume access to vectors close to $\mathrm{gh}(\mathcal{L})$, which is appropriate only when the local basis is close to the idealized reduction state assumed by the model. In high-dimensional finite reductions, this assumption may fail in early blocks. The adaptive criterion used here is therefore tied to the observed Gram–Schmidt profile, and should be understood as a finite-dimensional feasibility rule rather than as a replacement for the asymptotic D4f model.

Rule 3: Scope control via a runtime upper bound. Although the D4f value f_i is selected adaptively and may vary across blocks, it is always bounded by the local blocksize. APBKZ therefore maintains a runtime upper bound $f_{\max} = \max_i f_i$ over a reduction tour and restricts subsequent reduction to indices $i \leq n - f_{\max} - 50$. This constraint is empirically motivated to ensure robust saturation behavior. Restricting the pump scope to indices $i \leq n - f_{\max} - 50$ ensures that the underlying reduction subroutines operate on projected sublattices of dimension at least 50. In this regime, the behavior of lattice reduction is better aligned with GSA, whereas in lower dimensions boundary effects may render GSA-based predictions unreliable [CN11]. We conducted a sensitivity study for this scope constraint. Tightening it to $i \leq n - \beta + f_{\max}$ consistently reduced saturation and stability in our parameter regime, while relaxing it produced no noticeable benefit. We therefore retain $i \leq n - f_{\max} - 50$ as a conservative empirical compromise between stability and efficiency in the present setting.

Algorithm 1 summarizes the APBKZ algorithm and explicitly implements the above adaptive rules. Notably, APBKZ does not rely on any rank reduction condition and can be applied as a general-purpose adaptive reduction strategy for Approx-SVP.

4.2 HeadAPBKZ: Head-Focused Reduction under Rank Reduction Condition

Once the compact Approx-SVP rank reduction condition (Condition 2) is satisfied for a prefix of length k, further reduction of the lattice tail has negligible impact on feasibility. This observation is consistent with prior empirical findings [ZD23], which indicate that during the final stages of BKZ-style reduction, applying SVP subroutines to the first few indices is typically sufficient.

Our rank reduction framework formalizes this phenomenon by providing an explicit criterion on the required prefix length. Under this condition, it suffices for the lattice basis

Algorithm 1 APBKZ: Adaptive Pump BKZ

Input: Lattice basis $\mathbf{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$, maximum blocksize $\beta_{\max}$, maximum number of tours $T_{\max}$.
Output: Reduced lattice basis $\mathbf{B}$.

```
tour ← 0
f_max ← 0
while tour < T_max do
    Compute Gram–Schmidt norms ‖b_i^*‖ for all i
    Determine reduction bound i_max ← n − f_max − 50                ▷ Rule 3
    for i = 1 to i_max do
        Select adaptive blocksize β_i ≤ β_max                       ▷ Rule 1
        Select adaptive dimension-for-free value f_i                ▷ Rule 2
        f_max ← max(f_max, f_i)
        B ← Pump_{i,f_i,β_i}(B)
    end for
    tour ← tour + 1
    if no further improvement detected then
        break
    end if
end while
return B
```

to satisfy the Geometric Series Assumption within the rank-k sublattice $\mathcal{L}_{[1:k]}$. Restricting reduction to this prefix does not affect correctness, as sufficiently short vectors found in early blocks propagate through subsequent size reduction and basis updates.

HeadAPBKZ exploits this property as a restricted execution mode of APBKZ. It inherits the same adaptive blocksize and dimension-for-free selection rules, but limits the reduction scope to indices

$$i \leq \min\big(k,\ n - f_{\max} - 50\big),$$

where $f_{\max}$ denotes a conservative upper bound on the dimension-for-free parameter encountered during reduction. This joint constraint reflects both the rank reduction condition, which determines the relevant prefix length, and the Gaussian Heuristic, which ensures that the final Pump operation acts on a sublattice of sufficiently large dimension.

In our experiments, the transition from APBKZ to HeadAPBKZ is triggered immediately once the compact rank reduction condition (Condition 2) is met. This condition-driven switch concentrates computational effort on the most impactful part of the basis, while avoiding unnecessary reduction of the tail that would only increase computational cost without improving feasibility.

5 Experiments and Analysis

This section presents experimental results and analysis aimed at validating the rank reduction conditions developed in Section 3 and evaluating the performance of the adaptive lattice reduction algorithms proposed in Section 4. Rather than pursuing new performance records, our experiments are designed to assess whether these algorithms can reliably realize the feasibility conditions required by modern Approx-SVP rank reduction analyses on cryptographic-sized lattices.

5.1 Experimental Setup

All reduction experiments in this work are carried out on a server equipped with four NVIDIA RTX 4090 GPUs, two Intel Xeon Platinum 8336C CPUs (2.0 GHz), and 1.5 TB of RAM. This hardware specification is reported to fix the experimental context for the finite-regime validation below; the feasibility conditions and MSIS estimates are model-based and do not rely on this particular platform. We consider high-dimensional q-ary random lattices from the TU Darmstadt Lattice Challenge [BLR08a], which serves as a standard benchmark for evaluating the hardness of SIS instances. For a lattice of dimension n, the challenge consists of finding a non-zero vector of norm smaller than the modulus q, corresponding to solving SIS in a q-ary random lattice.

Following common practice for high-dimensional q-ary lattices, we apply the column deletion technique introduced by Micciancio and Regev [MR09] and refined in [SWCQ24] to reduce the lattice dimension while preserving the existence of short vectors. All subsequent reduction procedures are applied to these column-deleted sublattices. The 1050-dimensional instances used in the reduction experiments should not be interpreted as full Dilithium instances. They are generic q-ary benchmarks whose dimension is close to the 1- - level Dilithium Challenge and Extended settings [BDK+21]. This comparison is relevant because mainstream MSIS estimates commonly reduce module-SIS instances to generic q-ary SIS instances without exploiting module structure. Thus the experiments provide finite-regime validation in a comparable q-ary setting, while the Dilithium security estimates themselves are kept separate in Section 6.

5.2 Experimental Validation of Approx-SVP Rank Reduction Conditions

This subsection experimentally validates the feasibility thresholds predicted by the rank reduction conditions introduced in Section 3. In particular, we evaluate both the practical rank reduction condition (Condition 1) and the compact Approx-SVP rank reduction condition (Condition 2) by comparing their predicted thresholds with observed attack feasibility in practice.

Reading guide for rank reduction experiments. In the following tables, "MSD" denotes the maximal sieving dimension under the rank reduction condition stated in the caption. "Theory" is the predicted feasibility threshold, and "Experiments" is the largest dimension at which the attack succeeded in practice. "[LM18] rank" refers to the previous state-of-the-art rank reduction condition proposed in [LM18], while "Our rank" refers to the refined condition used in the table. In Table 2, "[LM18] wall time" denotes the time required under the original rank reduction condition, and "Condition 1 wall time" denotes the corresponding practical-rank baseline when it outputs the target short vector. In Table 3, "Condition 1 wall time" denotes the time required under the practical condition, while "Condition 2 wall time" denotes the compact-rank baseline under the same experimental framework. These wall times are internal comparisons between rank reduction conditions, not external performance records. Wall times are reported using the abbreviations d (days), h (hours), and m (minutes).

Table 2 validates the practical Approx-SVP rank reduction condition (Condition 1) on 850-, 925-, and 1025-dimensional q-ary lattice challenges. Across all instances, the observed maximal sieving dimensions match the predicted feasibility thresholds with relative deviations below 1%, confirming that the geometric sieving model captures the practical feasibility boundary accurately. The original condition proposed in [LM18] does not output the target short vector for these instances, as indicated by the "-" entries in the corresponding wall-time column.

Table 3 validates the compact Approx-SVP rank reduction condition (Condition 2) on the same challenge instances. The observed maximal sieving dimensions and achieved

Table 2: Experimental validation of the feasibility threshold predicted by Condition 1.

Dim	n	γ	δ	β	[LM18] rank	Our rank	MSD	
							Theory	Experiments
850	230	1.668906	1.007991	133	130	164	**132**	133
925	295	2.777266	1.008458	120	121	153	**123**	123
1025	330	2.891528	1.007766	132	135	170	**137**	137

Dim	n	Target norm	Experimental norm	[LM18] wall time	Condition 1 wall time
850	230	102.40	100.00	-	$1.40d$
925	295	137.56	133.36	-	$2.07h$
1025	330	164.99	163.43	-	$1.22d$

Table 3: Experimental validation of the feasibility threshold predicted by Condition 2.

Dim	n	γ	δ	β	[LM18] rank	Our rank	MSD	
							Theory	Experiments
850	230	1.668906	1.007991	133	130	149	**119**	118
925	295	2.777266	1.008458	120	121	134	**107**	107
1025	330	2.891528	1.007766	132	135	162	**130**	130

Dim	n	Target norm	Experimental norm	Condition 1 wall time	Condition 2 wall time
850	230	102.40	102.20	$1.40d$	**31.2m**
925	295	137.56	136.20	$2.07h$	**2m**
1025	330	164.99	163.01	$1.22d$	**1.89h**

solution norms both closely track the theoretical predictions, with discrepancies below 2%. This confirms that the compact condition provides a conservative yet accurate feasibility threshold in high dimensions.

The smaller predicted ranks translate into finite-dimensional speedups of approximately 64.6×, 62.1×, and 15.5× for dimensions 850, 925, and 1025, respectively, relative to Condition 1 under the tested parameters. The original condition proposed in [LM18] does not output the target short vector in these instances, whereas both proposed conditions do. These results support the relevance of accurate rank estimation for feasibility prediction and practical attack cost, without implying an asymptotic improvement.

5.3 Experiments for Lattice Basis Reduction

This subsection validates the reduction behavior predicted by the adaptive strategies introduced in Section 4. In particular, we evaluate whether APBKZ can reliably avoid early-dimension ineffectiveness and unsaturated reduction, and whether HeadAPBKZ preserves reduction feasibility once a compact rank reduction condition (Condition 2) is satisfied.

On the use of fixed blocksizes in the experiments. In the 1050-dimensional q-ary instance considered here, the initial prefix is dominated by q-vectors of norm q, and the corresponding sublattices do not yet follow Gaussian-Heuristic behavior. In this phase, adaptive blocksize selection based on local Gaussian-Heuristic estimates is therefore unreliable. Accordingly, before sufficiently short vectors are recovered, APBKZ is run with a fixed blocksize aligned with the comparison algorithm PnJBKZ, and adaptive blocksize

selection is enabled only after q-vectors no longer dominate the reduction.

5.3.1 Validation of Adaptive Reduction Behavior

We first evaluate APBKZ on a 340-dimensional sublattice derived from the 1050-dimensional q-ary instance of the TU Darmstadt Lattice Challenge. Table 4 compares classical PnJBKZ with APBKZ under comparable parameter settings. All comparisons are conducted using identical input bases, and differences in outcomes are solely attributable to the reduction strategy.

Table 4: Validation of adaptive reduction behavior on a 340-dimensional q-ary sublattice

Algorithm	(β, J)	Initial norm	Output norm	Saturation Ratio	Wall time
PnJBKZ	(118,2)	275.33	267.54	0.1%	6.37d
PnJBKZ	(121,2)	267.54	262.14	0%	8.43d
APBKZ	(118,1)	275.33	254.92	**46.2%**	3.77h
APBKZ	(121,1)	267.54	247.94	22.1%	7.02h

The comparison is against PnJBKZ, a widely adopted state-of-the-art strategy in the G6K framework. The results confirm that APBKZ consistently avoids early-dimension ineffectiveness. While PnJBKZ can still output valid short vectors, it does so with much lower saturation and weaker reduction efficiency, yielding longer output norms under comparable parameter settings. By contrast, APBKZ maintains substantially higher saturation and achieves shorter output norms.

These observations also support the use of the extended reduction scope $i \leq n - f_{\max} - 50$, which keeps early blocks in a regime where reduction remains effective.

5.3.2 Validation of HeadAPBKZ Mode Switching

We next evaluate the head-focused execution mode HeadAPBKZ introduced in Section 4.2. In these experiments, HeadAPBKZ is activated only after the compact Approx-SVP rank reduction condition (Condition 2) is satisfied for a prefix of dimension k, at which point further reduction of the lattice tail is expected to have negligible impact on feasibility. This behavior provides an algorithmic justification for treating Condition 2 as a stopping criterion rather than merely a theoretical bound.

Table 5: Validation of head-focused reduction after satisfying the compact rank reduction condition

Algorithm	(β, J)	Initial norm	Output norm	Saturation Ratio	Wall time
APBKZ	(118,1)	275.33	254.92	**46.2%**	3.77h
APBKZ	(121,1)	267.54	247.94	22.1%	7.02h
HeadAPBKZ	(118,1)	275.33	255.34	**41.8%**	3.11h
HeadAPBKZ	(121,1)	267.54	250.57	23.5%	4.64h

Table 5 reports reduction outcomes when restricting the reduction scope to the prefix sublattice $\mathcal{L}_{[1:k]}$ with $k = 161$. All HeadAPBKZ runs inherit the adaptive parameter selection rules of APBKZ and differ only in the reduction scope.

The results show that once the compact rank reduction condition is met, restricting reduction to the prefix sublattice preserves reduction quality while substantially reducing computational cost. In particular, HeadAPBKZ achieves reduction quality comparable to full APBKZ, as reflected by similar output norms and saturation behavior, while avoiding unnecessary tail reduction.

These observations validate the design of HeadAPBKZ described in Section 4.2. The transition from APBKZ to HeadAPBKZ is triggered by the satisfaction of the compact rank

reduction condition, which indicates that the attack has entered a feasible regime. Beyond this point, further reduction of the tail basis vectors can still yield shorter output norms; however, the marginal improvement per unit of computational cost becomes significantly smaller. HeadAPBKZ therefore restricts reduction to the head region in order to achieve a more favorable cost–quality trade-off, leading to faster norm decrease under comparable computational budgets.

5.4 Scalability and Limitations of the Experiments

The experiments above validate the feasibility conditions and adaptive reduction rules on the tested high-dimensional q-ary instances. They do not establish a scaling law for substantially larger dimensions. Within the hardware setting reported in Section 5.1, dimension 1050 is close to the largest regime in which we can complete controlled APBKZ/PnJBKZ comparisons under matched input bases, blocksize constraints, and success criteria. Table 4 already shows that q-vector-dominated prefixes exhibit reduced saturation in this range, so larger-dimensional experiments must control this effect before timing comparisons can be interpreted reliably.

The same experimental methodology can in principle be applied to higher dimensional q-ary instances, such as the 1100-dimensional case. We have started such experiments, but the preprocessing stage has not yet completed. The estimated remaining time is more than 30 days when both the APBKZ/HeadAPBKZ runs and the PnJBKZ baseline are included. This time is needed because the comparison requires two matched experiment families: one for APBKZ/HeadAPBKZ and one for the PnJBKZ baseline, using the same input bases, blocksize constraints, and success criteria.

The hardware resources also differ from those used in the public 1100-dimensional q-ary challenge record [BLR08b]. That record used two machines, one with 1.5 TB RAM and four NVIDIA RTX 4090 GPUs and another with 2 TB RAM and three NVIDIA RTX 4090D GPUs, whereas our experiments use the single machine described in Section 5.1. With fewer machines and less aggregate GPU/RAM capacity than the public-record setup, the matched 1100-dimensional comparison cannot be completed within the present revision window.

6 Refined Security Estimation of MSIS

In this section, we revisit the concrete security estimation of the Module Short Integer Solution (MSIS) problem underlying Dilithium. Our goal is not to propose new attacks, but to refine the security evaluation by instantiating the compact Approx-SVP rank reduction condition (Condition 2) and lattice reduction behavior developed in Sections 3 and 4 within a concrete estimation framework. The estimates reported below are model-dependent refinements of the standard Core-SVP evaluation framework accounting for finite-dimensional rank reduction effects. They should not be read as new conservative security margins for Dilithium.

6.1 Baseline MSIS Security Estimation

Bai et al. [BDK+21] estimate the hardness of MSIS within the Core-SVP framework by modeling BKZ reduction with a sieving-based SVP oracle on projected sublattices. In this baseline model, the oracle is assumed to return vectors whose lengths are close to the Gaussian Heuristic, and the oracle success is captured through the expected number of valid short vectors produced by sieving.

Concretely, the **BKZShapeSim** procedure (Algorithm 2) is used to generate a reduced Gram-Schmidt profile, from which a head region and a middle region are identified. The

Algorithm 2 BKZShapeSim(q, h, w, β)

Input: q, h, w, β.
Output: $i, j, \{\ell_t\}_{1 \le t \le w}$

$\log_2(\mathrm{V}) \leftarrow h \cdot \log_2(q)$
$slope \leftarrow 2 \cdot \log_2(\mathsf{RHF}(\beta))$
$V_{bkz} \leftarrow 0$
$S_{len} \leftarrow w$
for $t = 1$ to w **do**
  $V_{bkz} \leftarrow V_{bkz} + t \cdot slope$
  if $V_{bkz} > \log_2(\mathrm{V})$ **then**
    $S_{len} \leftarrow t$
    break
  end if
end for
for $t = 1$ to S_{len} **do**
  $\ell_t \leftarrow (S_{len} - t + 1) \cdot slope$
end for
for $t = S_{len} + 1$ to w **do**
  $\ell_t \leftarrow 0$
end for
$V_{diff} \leftarrow V_{bkz} - \log_2(\mathrm{V})$
for $t = 1$ to S_{len} **do**
  $\ell_t \leftarrow \ell_t - \frac{V_{diff}}{S_{len}}$
end for
$i \leftarrow \min\{t \in [1, w] : \ell_t < \log_2(q)\}$ ▷ head: $1, \ldots, i-1$
$j \leftarrow \max\{t \in [1, w] : \ell_t > 0\}$ ▷ middle: $i, \ldots, j$
return $i, j, \{\ell_t\}_{1 \le t \le w}$

Algorithm 3 MSISCost $(i, j, \zeta', \ell_i, q, \beta)$

Input: $i, j, \zeta', \ell_i, q, \beta$ with $1 \le i \le j$.
Output: B, p.

$\sigma \leftarrow \exp(\ell_i)/\sqrt{j - i + 1}$
$p_{mid} \leftarrow \Pr\left[x \leftarrow N\left(0, \sigma^2\right) : -\zeta' \le x \le \zeta'\right]$
$p_{head} \leftarrow \frac{2\zeta'}{q}$
$p_{single} \leftarrow {p_{mid}}^{j-i+1} \cdot {p_{head}}^{i-1}$
$N \leftarrow 2^{0.2075\beta}$
$p\left(\|s\|_\infty < \zeta'\right) \leftarrow N \cdot p_{single}$
$B \leftarrow \frac{1}{p(\|s\|_\infty < \zeta')} \cdot 2^{0.292\beta}$
return $B, p\left(\|s\|_\infty < \zeta'\right)$

per-vector success probability is then evaluated under a mixed distribution model, and combined with the expected sieve output count to derive the resulting attack cost, as summarized in Algorithm 3. We refer to [BDK+21] for the full details of this baseline estimator.

6.2 Refined Security Estimation Framework and an Analytic Approximation

We present the refined MSIS estimation together with a closed-form analytic approximation. The analytic approximation explains the numerical objective in closed form by using a continuous BKZ-shape model, an analytic rank approximation, and direct analytic approximations of the two D4f parameters.

The estimator is organized into the following steps. First, the MSIS instance is embedded into a generic q-ary lattice. Second, a BKZ-shape model provides the Gram–Schmidt profile used to identify the head and middle regions. Third, the compact rank condition determines the feasible prefix rank. Fourth, a probability-aware sieve model

estimates the success probability of obtaining a valid short vector. Finally, the two-stage cost model aggregates preprocessing and final sieving costs. This decomposition separates the parts supported by the experiments in Section 5 from the modelling assumptions needed to extrapolate to Dilithium-scale parameters.

Lattice model and Gaussian heuristic. As in the baseline estimator, solving the target $\mathrm{MSIS}_{k,\ell,\zeta'}$ instance is reduced to Approx-SVP on the associated q-ary lattice

$$\mathcal{L}' = \mathcal{L}\left(\begin{bmatrix} q\mathbf{I} & M(\mathbf{A}) \\ \mathbf{0} & -\mathbf{I} \end{bmatrix}\right),$$

with $h = nk$ and $w = nk + n\ell$, where $M(\mathbf{A})$ is the standard integer embedding of the public MSIS matrix $\mathbf{A} \in \mathbb{R}_q^{k\times\ell}$. Under the Gaussian Heuristic, $\mathrm{gh}(\mathcal{L}') \approx \sqrt{w/(2\pi e)}\, q^{h/w}$, and we write $\delta_\beta := \mathsf{RHF}(\beta) = \left(\frac{\beta}{2\pi e}(\pi\beta)^{1/\beta}\right)^{1/(2(\beta-1))}$ for the root Hermite factor of a BKZ-β reduced basis.

Continuous BKZ-shape approximation. To approximate Algorithm 2 without an explicit loop, we use a closed-form profile with the same slope, total log-volume, and uniform correction. Set $s_\beta := 2\ln\delta_\beta$, $\mathrm{glv} := h\ln q$, and

$$B_\beta := \left\lfloor \frac{\sqrt{1 + 8\,\mathrm{glv}/s_\beta} - 1}{2} \right\rfloor, \qquad \Delta_\beta := \frac{s_\beta B_\beta(B_\beta+1)/2 - \mathrm{glv}}{B_\beta},$$

where B_β is the discrete solution of $s_\beta B_\beta(B_\beta+1)/2 \approx \mathrm{glv}$. We then approximate the profile by

$$L_t^{(\beta)} \approx \begin{cases} (B_\beta - t + 1)s_\beta - \Delta_\beta, & 1 \le t \le B_\beta, \\ 0, & t > B_\beta, \end{cases}$$

so that $i_\beta = 1$, $j_\beta = B_\beta$, $\ell_\beta := B_\beta s_\beta - \Delta_\beta$, and $\alpha_\beta := e^{\ell_\beta}/\mathrm{gh}(\mathcal{L}')$.

Analytic rank from the compact condition. For a preprocessing blocksize β, the compact Approx-SVP rank reduction condition (Condition 2) is

$$\alpha_\beta \ge \xi_\beta \sqrt{\frac{r}{w}}\, \delta_\beta^{w-r+1}, \qquad \xi_\beta := (8\ln 10)^{1/\beta},$$

where r denotes the candidate prefix rank. Replacing the inequality with equality gives

$$F_\beta(r) := \ln\alpha_\beta - \ln\xi_\beta - \frac{1}{2}\ln\frac{r}{w} - (w - r + 1)\ln\delta_\beta = 0.$$

Dropping the logarithmic correction term first yields

$$r_\beta^{(0)} := w + 1 - \frac{\ln\alpha_\beta - \ln\xi_\beta}{\ln\delta_\beta}.$$

Substituting this into F_β gives

$$F_\beta(r_\beta^{(0)}) = -\frac{1}{2}\ln\frac{r_\beta^{(0)}}{w}.$$

Under typical parameters, $r_\beta^{(0)} < w$, hence $F_\beta(r_\beta^{(0)}) > 0$. Moreover,

$$F_\beta'(r) = -\frac{1}{2r} + \ln\delta_\beta, \qquad F_\beta''(r) = \frac{1}{2r^2} > 0.$$

Thus F_β is convex and locally increasing near the continuous root. Applying one Newton step gives

$$\widetilde{r}_\beta = r_\beta^{(0)} - \frac{F_\beta(r_\beta^{(0)})}{F'_\beta(r_\beta^{(0)})} \approx w + 1 - \frac{\ln \alpha_\beta - \ln \xi_\beta - \frac{1}{2}\ln(r_\beta^{(0)}/w)}{\ln \delta_\beta}, \tag{14}$$

and we set $r_\beta := \lfloor \widetilde{r}_\beta \rceil$.

Single-vector success probability. We adopt the same head–middle model as in the baseline estimator $\mathbf{MSISCost}(i, j, \zeta', \ell_i, q, \beta)$ (Algorithm 3) from Section 6.1, following [BDK+21]. Let $\sigma_\beta := e^{\ell_\beta}/\sqrt{B_\beta}$. Then the middle coordinates are modeled by a Gaussian law with standard deviation σ_β, giving

$$p_{\mathrm{mid}}(\beta) = \Pr(|X| \le \zeta') = \operatorname{erf}\left(\frac{\zeta'}{\sqrt{2}\sigma_\beta}\right), \qquad X \sim \mathcal{N}(0, \sigma_\beta^2),$$

while the head coordinates are modeled uniformly modulo q, so $p_{\mathrm{head}} = 2\zeta'/q$. Hence

$$p_{\mathrm{single}}(\beta) = p_{\mathrm{mid}}(\beta)^{j_\beta - i_\beta + 1} p_{\mathrm{head}}^{i_\beta - 1}.$$

Under the present continuous approximation, $i_\beta = 1$ and $j_\beta = B_\beta$, so this simplifies to

$$p_\beta := p_{\mathrm{single}}(\beta) = p_{\mathrm{mid}}(\beta)^{B_\beta} = \left(\operatorname{erf}\left(\frac{\zeta'}{\sqrt{2}\sigma_\beta}\right)\right)^{B_\beta}. \tag{15}$$

Direct analytic D4f parameters. We now derive an analytic approximation for the optimal D4f parameter at a fixed stage dimension $\kappa \in \{\beta, r_\beta\}$, where β is the preprocessing blocksize and r_β is the analytic prefix rank obtained above. For each such κ, define

$$\delta_\kappa := \mathsf{RHF}(\kappa), \qquad A_\kappa := \left(\sqrt{\frac{4}{3}}\right)^\kappa, \qquad a_\kappa := \kappa \frac{w}{w-1} \ln \delta_\kappa, \qquad b := 0.292 \ln 2. \tag{16}$$

We use the refined model of usable short vectors under dimension-for-free from [WXG25]:

$$N_\kappa(f) \approx \left(\sqrt{\frac{4}{3}}\, \delta_\kappa^{-fw/(w-1)}\right)^\kappa = A_\kappa e^{-a_\kappa f},$$

and model the stage success probability, stage cost, and local expected cost by

$$P_\kappa(f) = 1 - (1 - p_\beta)^{N_\kappa(f)}, \qquad T_\kappa(f) = 2^{0.292(\kappa - f)}, \qquad C_\kappa(f) := T_\kappa(f)/P_\kappa(f).$$

Since $\ln(\cdot)$ is monotone, we minimize $\ln C_\kappa(f)$ with respect to f instead of $C_\kappa(f)$. For small p_β, we use the first-order expansion $\ln(1 - p_\beta) \approx -p_\beta$, together with the Poisson approximation

$$P_\kappa(f) = 1 - (1 - p_\beta)^{N_\kappa(f)} \approx 1 - e^{-N_\kappa(f) p_\beta}.$$

Introducing $u_\kappa(f) := N_\kappa(f)\, p_\beta = A_\kappa p_\beta e^{-a_\kappa f}$, this yields the following continuous approximation to $\ln C_\kappa(f)$

$$\ln C_\kappa(f) \approx b(\kappa - f) - \ln(1 - e^{-u_\kappa(f)}).$$

Differentiating with respect to f gives the stationarity condition

$$b = a_\kappa \frac{u_\kappa(f)}{e^{u_\kappa(f)} - 1}.$$

Setting $u := u_\kappa(f)$ and $c_\kappa := b/a_\kappa$, the stationarity condition becomes

$$c_\kappa = \frac{u}{e^u - 1},$$

whose unique positive solution, denoted by u^*_κ, is[1]

$$u^*_\kappa = -W_{-1}\big(-c_\kappa e^{-c_\kappa}\big) - c_\kappa.$$

Since $u_\kappa(f) = A_\kappa p_\beta e^{-a_\kappa f}$, solving for f gives

$$f^*_\kappa = \frac{\ln(A_\kappa p_\beta) - \ln u^*_\kappa}{a_\kappa}, \tag{17}$$

and we set $f^{\mathrm{ana}}_\kappa := \max\{0, \lfloor f^*_\kappa \rceil\}$. For convenience, write $f_\beta := f^{\mathrm{ana}}_\beta$ and $f_{r_\beta} := f^{\mathrm{ana}}_{r_\beta}$.

One-dimensional analytic approximation. Substituting the analytic rank and analytic D4f parameters into the two-stage model gives

$$T^{\mathrm{ana}}_\beta = 2^{0.292(\beta - f_\beta)}, \qquad T^{\mathrm{ana}}_{r_\beta} = 2^{0.292(r_\beta - f_{r_\beta})},$$

$$P^{\mathrm{ana}}_\beta = 1 - (1-p_\beta)^{A_\beta e^{-a_\beta f_\beta}}, \qquad P^{\mathrm{ana}}_{r_\beta} = 1 - (1-p_\beta)^{A_{r_\beta} e^{-a_{r_\beta} f_{r_\beta}}},$$

and hence

$$P^{\mathrm{ana}}_{\mathrm{tot}}(\beta) = P^{\mathrm{ana}}_\beta + (1 - P^{\mathrm{ana}}_\beta) P^{\mathrm{ana}}_{r_\beta}.$$

Here p_β is given by (15), while A_κ and a_κ are defined in (16) for $\kappa \in \{\beta, r_\beta\}$. Therefore

$$\widehat{\lambda}^{\mathrm{ana}}_{\mathrm{MSIS}}(q, h, w, \zeta') = \min_{\beta \geq 50} \log_2 \left(\frac{2^{0.292(\beta - f_\beta)} + 2^{0.292(r_\beta - f_{r_\beta})}}{P^{\mathrm{ana}}_\beta + (1 - P^{\mathrm{ana}}_\beta) P^{\mathrm{ana}}_{r_\beta}} \right). \tag{18}$$

The original four-parameter optimization is over (β, r, f_β, f_r). Using the analytic approximations above, the prefix rank r and the two D4f parameters f_β and f_r are expressed as functions of β. As a result, the problem is approximated by a one-dimensional search over β, with the corresponding analytic choices

$$r = r_\beta := \lfloor \widetilde{r}_\beta \rceil, \qquad \widetilde{r}_\beta \text{ is given by (14)},$$

$$f_\kappa := \max\{0, \lfloor f^*_\kappa \rceil\}, \qquad f^*_\kappa \text{ is given by (17)}, \qquad \kappa \in \{\beta, r_\beta\},$$

so that, in particular, the final-stage parameter is $f_r = f_{r_\beta}$.

Numerical refined estimation. The numerical refined estimates are obtained by direct discrete optimization of the original refined MSIS cost model over the parameter tuple (β, r, f_β, f_r), namely by searching over the admissible values of the preprocessing blocksize, the prefix rank, and the two D4f parameters, and selecting the combination that minimizes the estimated attack cost. By contrast, (18) is the one-dimensional analytic approximation obtained after substituting the analytic rank and analytic D4f parameters into the two-stage model.

[1]Here W_{-1} denotes the (-1)-branch of the Lambert W function, defined by $W(z)e^{W(z)} = z$.

Table 6: MSIS security estimates for Dilithium under the Core-SVP baseline, the numerical refined estimation, and the analytic approximation [BDK+21]

	Dilithium-II	Dilithium-III	Dilithium-V
NIST Level	2	3	5
Lattice dimension w	2048	2816	3840
Core-SVP β	417	602	868
Core-SVP (bits)	121.96	176.37	253.87
β_{num}	429	619	892
r_{num}	431	621	894
$f_{\beta_{\text{num}}}$	25	32	45
$f_{r_{\text{num}}}$	38	49	64
Numerical refined estimation (bits)	118.31	171.75	247.78
Numerical change (bits)	-3.65	-4.62	-6.09
β_{ana}	430	619	892
r_{ana}	432	620	893
$f_{\beta_{\text{ana}}}$	26	32	45
$f_{r_{\text{ana}}}$	26	32	45
Analytic approximation (bits)	119.48	172.84	248.89
Analytic change (bits)	-2.48	-3.53	-4.98
Analytic gap (bits)	+1.17	+1.09	+1.11

Relationship with the numerical estimation. The analytic approximation is derived from the continuous BKZ-shape approximation, the analytic rank approximation, and the direct analytic D4f parameters developed above. For Dilithium-II, Dilithium-III, and Dilithium-V, it yields 119.48, 172.84, and 248.89 bits, respectively, compared with the numerical refined estimates 118.31, 171.75, and 247.78 bits, giving discrepancies of about 1.17, 1.09, and 1.11 bits. This shows that the analytic approximation closely tracks the numerical estimation. The equality of the two analytic D4f values in Table 6 is an integer-rounding effect after two independent continuous computations. The quantity f_β^{ana} is computed for the BKZ preprocessing stage, whereas f_r^{ana} is computed for the final rank-reduced sieving stage. Although β and r are different parameters, the rounded integer values coincide for the reported instances; this equality is not an imposed constraint of the model.

Resulting estimates. Table 6 reports the optimized parameters and resulting security levels for the Dilithium parameter sets under the refined MSIS estimation, together with the Core-SVP baseline. Compared with the Core-SVP baseline, the refined estimation yields a reduction of approximately 3.65–6.09 bits, reflecting the tighter feasibility constraints introduced by the compact rank reduction condition and the refined probability model.

6.3 Interpreting the Experiments for MSIS Estimation

The experiments in Section 5 support the estimator at the level of finite-dimensional q-ary reduction behavior. In particular, they test whether the proposed rank conditions predict feasible prefix dimensions, whether low saturation occurs under less adaptive reduction schedules, and whether head-focused reduction preserves the relevant prefix quality once the compact condition is met. These are precisely the phenomena that motivate the refined two-stage estimator.

At the same time, the Dilithium estimates in Table 6 remain model-dependent. The estimator expands the MSIS instance to a generic q-ary SIS setting and does not exploit module structure. Moreover, extrapolating from the tested finite-dimensional challenges to larger dimensions relies on the continued validity of the BKZ-shape approximation, the

sieve-output probability model, and the D4f success model. The estimates should therefore be understood as refined concrete cost estimates under these assumptions, not as a direct experimental attack on full Dilithium instances.

7 Conclusion

In this work, we revisited Approx-SVP rank reduction from a feasibility-driven perspective and identified fundamental limitations in existing conditions when applied to cryptographic-sized lattices. By focusing on the existence of a single sufficiently short vector, we derived refined and compact rank reduction conditions that better reflect the requirements of modern D4f-enabled attacks.

We further proposed APBKZ and HeadAPBKZ, two robust lattice reduction variants whose purpose is to ensure that the basis profiles assumed in theoretical analyses can be reliably realized in high dimensions. Experimental results confirm that these strategies significantly improve stability and efficiency without relying on simulator-driven parameter selection.

By integrating refined rank reduction theory with robust reduction behavior, we developed a concrete security estimation framework for MSIS and applied it to Dilithium. Under the modelling assumptions made in this work, accounting for compact rank reduction behavior leads to a decrease of 3.65–6.09 bits relative to the conservative Core-SVP baseline when D4f techniques are considered. We further instantiated an analytic approximation whose estimates stay within about 1.1 bits of the numerical estimation, indicating that the closed-form approximation remains close to the numerical baseline.

Future work. Several limitations remain. The rank reduction conditions rely on geometric and probabilistic sieve models; the experiments are restricted to finite-dimensional q-ary regimes and do not cover arbitrary lattice families; and the MSIS estimator expands module-SIS to generic q-ary SIS without exploiting module structure. Regarding a systematic scalability study in larger dimensions, once the ongoing 1100-dimensional experiments are completed and checked, we will promptly update the ePrint version with the additional results. More broadly, a dedicated study of structured lattices such as NTRU and Module-LWE/SIS-based schemes, and a rigorous theoretical understanding of practical heuristics such as re-randomization are important directions for future work.

Availability

The implementation and experiment scripts are currently being refined and will be made publicly available upon the publication of the final camera-ready version.

References

[ADH+19] Martin R. Albrecht, Léo Ducas, Gottfried Herold, Elena Kirshanova, Eamonn W. Postlethwaite, and Marc Stevens. The general sieve kernel and new records in lattice reduction. In *Advances in Cryptology - EUROCRYPT 2019 - 38th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Darmstadt, Germany, May 19-23, 2019, Proceedings, Part II*, volume 11477 of *Lecture Notes in Computer Science*, pages 717–746. Springer, 2019.

[Ajt99] Miklós Ajtai. Generating hard instances of the short basis problem. In *Automata, Languages and Programming, 26th International Colloquium,*

ICALP'99, Prague, Czech Republic, July 11-15, 1999, Proceedings, volume 1644 of *Lecture Notes in Computer Science*, pages 1–9. Springer, 1999.

[AWHT16] Yoshinori Aono, Yuntao Wang, Takuya Hayashi, and Tsuyoshi Takagi. Improved progressive BKZ algorithms and their precise cost estimation by sharp simulator. In *Advances in Cryptology - EUROCRYPT 2016 - 35th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Vienna, Austria, May 8-12, 2016, Proceedings, Part I*, volume 9665 of *Lecture Notes in Computer Science*, pages 789–819. Springer, 2016.

[BDGL16] Anja Becker, Léo Ducas, Nicolas Gama, and Thijs Laarhoven. New directions in nearest neighbor searching with applications to lattice sieving. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 10–24. SIAM, 2016.

[BDK+21] Shi Bai, Léo Ducas, Eike Kiltz, Tancrède Lepoint, Vadim Lyubashevsky, Peter Schwabe, Gregor Seiler, and Damien Stehlé. CRYSTALS-Dilithium – algorithm specifications and supporting documentation (version 3.1). Submission to the NIST Post-Quantum Cryptography Standardization Project, 2021.

[BGJ14] Anja Becker, Nicolas Gama, and Antoine Joux. A sieve algorithm based on overlattices. *LMS J. Comput. Math.*, 17(Theory):49–70, 2014.

[BGJ15] Anja Becker, Nicolas Gama, and Antoine Joux. Speeding-up lattice sieving without increasing the memory, using sub-quadratic nearest neighbor search. *IACR Cryptol. ePrint Arch.*, 2015:522, 2015.

[BL16] Anja Becker and Thijs Laarhoven. Efficient (ideal) lattice sieving using cross-polytope LSH. In *Progress in Cryptology - AFRICACRYPT 2016 - 8th International Conference on Cryptology in Africa, Fes, Morocco, April 13-15, 2016, Proceedings*, volume 9646 of *Lecture Notes in Computer Science*, pages 3–23. Springer, 2016.

[BLR08a] Johannes Buchmann, Richard Lindner, and Markus Rückert. Explicit hard instances of the shortest vector problem. In *Post-Quantum Cryptography, Second International Workshop, PQCrypto 2008, Cincinnati, OH, USA, October 17-19, 2008, Proceedings*, volume 5299 of *Lecture Notes in Computer Science*, pages 79–94. Springer, 2008.

[BLR08b] Johannes Buchmann, Richard Lindner, and Markus Rückert. Explicit hard instances of the shortest vector problem. In *Post-Quantum Cryptography, Second International Workshop, PQCrypto 2008, Cincinnati, OH, USA, October 17-19, 2008, Proceedings*, volume 5299 of *Lecture Notes in Computer Science*, pages 79–94. Springer, 2008.

[Che13] Yuanmi Chen. *Réduction de réseau et sécurité concrète du chiffrement complètement homomorphe*. PhD thesis, Université Paris Diderot – Paris 7, 2013.

[CN11] Yuanmi Chen and Phong Q. Nguyen. BKZ 2.0: Better lattice security estimates. In *Advances in Cryptology - ASIACRYPT 2011 - 17th International Conference on the Theory and Application of Cryptology and Information Security, Seoul, South Korea, December 4-8, 2011. Proceedings*, volume 7073 of *Lecture Notes in Computer Science*, pages 1–20. Springer, 2011.

[DSvW21] Léo Ducas, Marc Stevens, and Wessel P. J. van Woerden. Advanced lattice sieving on GPUs, with tensor cores. In *Advances in Cryptology - EUROCRYPT 2021 - 40th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Zagreb, Croatia, October 17-21, 2021, Proceedings, Part II*, volume 12697 of *Lecture Notes in Computer Science*, pages 249–279. Springer, 2021.

[Duc18] Léo Ducas. Shortest vector from lattice sieving: A few dimensions for free. In *Advances in Cryptology - EUROCRYPT 2018 - 37th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tel Aviv, Israel, April 29 - May 3, 2018 Proceedings, Part I*, volume 10820 of *Lecture Notes in Computer Science*, pages 125–145. Springer, 2018.

[GN08] Nicolas Gama and Phong Q. Nguyen. Predicting lattice reduction. In *Advances in Cryptology - EUROCRYPT 2008, 27th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Istanbul, Turkey, April 13-17, 2008. Proceedings*, volume 4965 of *Lecture Notes in Computer Science*, pages 31–51. Springer, 2008.

[Laa15] Thijs Laarhoven. Sieving for shortest vectors in lattices using angular locality-sensitive hashing. In *Advances in Cryptology - CRYPTO 2015 - 35th Annual Cryptology Conference, Santa Barbara, CA, USA, August 16-20, 2015, Proceedings, Part I*, volume 9215 of *Lecture Notes in Computer Science*, pages 3–22. Springer, 2015.

[LdW15] Thijs Laarhoven and Benne de Weger. Faster sieving for shortest lattice vectors using spherical locality-sensitive hashing. In *Progress in Cryptology - LATINCRYPT 2015 - 4th International Conference on Cryptology and Information Security in Latin America, Guadalajara, Mexico, August 23-26, 2015, Proceedings*, volume 9230 of *Lecture Notes in Computer Science*, pages 101–118. Springer, 2015.

[LM18] Thijs Laarhoven and Artur Mariano. Progressive lattice sieving. In *Post-Quantum Cryptography - 9th International Conference, PQCrypto 2018, Fort Lauderdale, FL, USA, April 9-11, 2018, Proceedings*, volume 10786 of *Lecture Notes in Computer Science*, pages 292–311. Springer, 2018.

[MR09] Daniele Micciancio and Oded Regev. Lattice-based cryptography. In *Post-Quantum Cryptography*, pages 147–191. Springer Berlin Heidelberg, 2009.

[MV10] Daniele Micciancio and Panagiotis Voulgaris. A deterministic single exponential time algorithm for most lattice problems based on Voronoi cell computations. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, pages 351–358. ACM, 2010.

[Ngu10] Phong Q. Nguyen. Hermite's constant and lattice algorithms. In *The LLL Algorithm - Survey and Applications*, Information Security and Cryptography, pages 19–69. Springer, 2010.

[NV08] Phong Q. Nguyen and Thomas Vidick. Sieve algorithms for the shortest vector problem are practical. *J. Math. Cryptol.*, 2(2):181–207, 2008.

[Reg09] Oded Regev. On lattices, learning with errors, random linear codes, and cryptography. *J. ACM*, 56(6):34:1–34:40, 2009.

[Sch03] Claus-Peter Schnorr. Lattice reduction by random sampling and birthday methods. In *STACS 2003, 20th Annual Symposium on Theoretical Aspects of Computer Science, Berlin, Germany, February 27 - March 1, 2003, Proceedings*, volume 2607 of *Lecture Notes in Computer Science*, pages 145–156. Springer, 2003.

[SE94] Claus-Peter Schnorr and M. Euchner. Lattice basis reduction: Improved practical algorithms and solving subset sum problems. *Math. Program.*, 66:181–199, 1994.

[SWCQ24] Minghao Sun, Shixiong Wang, Hao Chen, and Longjiang Qu. A new random sampling method and its application in improving progressive BKZ algorithm. *J. Syst. Sci. Complex.*, 37(5):2262–2292, 2024.

[WLTB11] Xiaoyun Wang, Mingjie Liu, Chengliang Tian, and Jingguo Bi. Improved Nguyen-Vidick heuristic sieve algorithm for shortest vector problem. In *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security, ASIACCS 2011, Hong Kong, China, March 22-24, 2011*, pages 1–9. ACM, 2011.

[WXG25] Geng Wang, Wenwen Xia, and Dawu Gu. Heuristic algorithm for solving restricted SVP and its applications. In *Post-Quantum Cryptography - 16th International Workshop, PQCrypto 2025, Taipei, Taiwan, April 8-10, 2025, Proceedings, Part II*, volume 15578 of *Lecture Notes in Computer Science*, pages 119–152. Springer, 2025.

[ZD23] Ziyu Zhao and Jintai Ding. Practical improvements on BKZ algorithm. In *Cyber Security, Cryptology, and Machine Learning - 7th International Symposium, CSCML 2023, Be'er Sheva, Israel, June 29-30, 2023, Proceedings*, volume 13914 of *Lecture Notes in Computer Science*, pages 273–284. Springer, 2023.

[ZDY25] Ziyu Zhao, Jintai Ding, and Bo-Yin Yang. Sieving with streaming memory access. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2025(2):362–384, 2025.

[ZPH13] Feng Zhang, Yanbin Pan, and Gengran Hu. A three-level sieve algorithm for the shortest vector problem. In *Selected Areas in Cryptography - SAC 2013 - 20th International Conference, Burnaby, BC, Canada, August 14-16, 2013, Revised Selected Papers*, volume 8282 of *Lecture Notes in Computer Science*, pages 29–47. Springer, 2013.