Lattice cryptography: from cryptanalysis to New Foundations

Woerden, W.P.J. van

CHAPTER 7

Overstretched NTRU

This chapter is based on the joint work 'NTRU Fatigue: How Stretched is Overstretched', with Léo Ducas, published at Asiacrypt 2021.

7.1 Introduction

One could view lattice reduction, or more specifically the BKZ algorithm, as a way to solve the approximate Shortest Vector Problem, or to create a good basis. However, in practice BKZ goes beyond that and can also recover (hidden) geometric structures of a lattice. For example if a rank d lattice contains an unusually short vector, much shorter than the Gaussian Heuristic prescribes, then the BKZ algorithm recovers this shortest vector with a much smaller blocksize $\beta \ll d$.

Heuristically, this behaviour can be fully explained, and after a line of works [GN08b; AFG13; ADPS16; DDGR20; PV21] we can predict accurately for which blocksize β an unusual short vector is recovered. For most lattice-based cryptosystems, key recovery or decryption can be reduced to finding such an unusually short vector, and

thus these predictions directly give an upper bound on the security of such schemes. In fact, for most schemes and common parameters this is (close to) the best known attack.

One such cryptosystems is the NTRU cryptosystem of Hoffstein, Pipher and Silverman [HPS98; Che+20]. The NTRU secret key consists of polynomials $\mathbf{f}, \mathbf{g} \in \mathbb{Z}[X]/(X^n - 1)$ with n prime, and with small, e.g., ternary, coefficients, and the public key is given by $\mathbf{h} := \mathbf{g}/\mathbf{f} \bmod q$ for some modulus q. The public key and the modulus q allow one to define an 'NTRU lattice' of dimension $d = 2n$, which contains an unusually short vector related to the secret key. Recovery of this vector immediately leads to full key-recovery, and thus the estimated security, for example that of the NIST 3rd round finalist NTRU [Che+20], is directly based on the predictions of how the BKZ algorithms recovers this unusually short vector in the NTRU lattice.

However, only recently, it was discovered that the security of NTRU is in fact more subtle than the problem of finding a single unusually short vector in a lattice; the NTRU lattice contains more structure. The first dent in this status quo came in 2016, from two concurrent works of Albrecht et al., and Cheon et al. [ABD16; CJL16], which exploit the specific algebraic structure of the NTRU lattice to improve upon pure lattice reduction attacks[1]. This approach was shown to be applicable when the modulus q is large enough (say, super-polynomial), a regime coined "overstretched".

Shortly thereafter Kirchner and Fouque [KF17] showed that this improved complexity does *not* require any algebraic structure, and is instead rooted in the purely geometrical fact that the NTRU lattice contains an unusually dense sublattice of large rank, *i.e.* a sublattice of small determinant.[2] Due to this dense sublattice lattice reduction attacks perform much better than initially expected. They also go further in their analysis, and conclude that moduli q as small as $n^{2.783+o(1)}$ already belong to the overstretched regime —for random ternary secrets. In particular, for q larger than this bound, the security of NTRU is significantly less than that of Learning With Errors [Reg04] and of

[1] Though the idea had been inconclusively considered already in 2002 by Gentry, Jonsson, Nguyen Stern and Szydlo as reported in [GS02, Sec. 6].

[2] Note that one may associate a short vector to a dense sublattice of dimension 1.

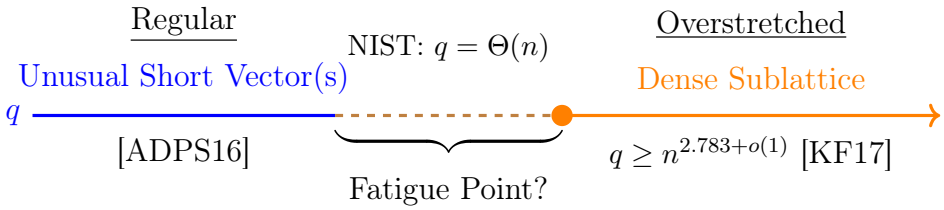

Figure 7.1: A sketch of the regular versus the overstretched regime, and the unknown fatigue point in between.

its Ring variant [SSTX09; LPR13] using similar parameters.[3]

However, it is not so clear from the analysis of Kirchner and Fouque whether this asymptotic quantity $n^{2.783+o(1)}$ is an estimate or merely an upper bound on the *fatigue point*, that is the value of q separating the standard regime from the overstretched regime. Their analysis is based on a lemma of Pataki and Tural [PT08], that constraints the shape of lattice basis in terms of the volume of their sublattices. While it allows to conclude that the dense sublattice must be discovered after reducing the lattice basis beyond these constraints, it does not really explain *how* lattice reduction ends up discovering the dense sublattice, nor does it exclude that the discovery could happen earlier.

So far, it has been generally considered that only advanced schemes —requiring very large q— such as NTRU-based Homomorphic Encryption [BLLN13] or cryptographic multi-linear maps [GGH13] could be affected by this overstretched regime. Yet, because the analysis of Kirchner and Fouque is only asymptotic, and because it may only provide an upper bound on the fatigue point, there is at the moment little documented evidence that the overstretched regime may not in fact extend further down, maybe down to the NTRU encryption scheme itself [HPS98; Che+20]! Admittedly, this seems like a far fetched concern: asymptotically this scheme chooses $q = O(n)$, with a hidden constant between 4 and 5 in practice. However, this scheme being now a finalist of the NIST standardisation process for post-quantum cryptography, it appears rather imperious to refine our understanding of the phenomenon, and to finally close this pending question.

[3] In fact, the presence of n rotations of the secret key already implies a minor security degradation compared to (Ring)-LWE already in the standard regime [MS01; DDGR20].

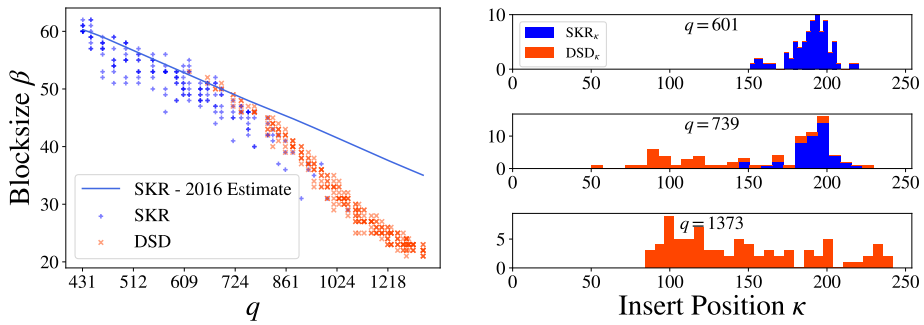

Figure 7.2: Progressive BKZ with 8 tours per blocksize on matrix NTRU instances with parameters $n = 127, \sigma^2 = \frac{2}{3}$ for several moduli q. **Left:** the first blocksize β at which Progressive BKZ detects the Secret Key Recovery (SKR$_\kappa$) or Dense Sublattice Discovery (DSD$_\kappa$) event. We did 10 runs per modulus q. For the 2016-estimates, we use the geometric series assumption (GSA) for the shape of the basis and a probabilistic model for the discovery of the secret vector (see Section 7.2.3). **Right:** the positions κ at which a secret key or dense sublattice vector are detected over 80 runs per modulus.

We found further motivation to go down this rabbit hole by measuring the concrete value of the fatigue point experimentally. Until now, all documented experiments on the overstretched regime [ABD16; KF17; LW20] have focused on rather large values of q, and only used weak lattice reduction (LLL [LLL82], BKZ with blocksize 20): their goal was to demonstrate the claimed general behaviour when parameters are far in the overstretched regime. On the contrary, we focus the attention to the fatigue point for this preliminary experiment. That is, we ran strong reduction (progressive-BKZ [Sch87; AWHT16] up to blocksize 60) until a vector related to the secret key appeared for a range of moduli q. We distinguished the standard regime from the overstretched regime by classifying according to which event occurs first

- Secret Key Recovery (SKR$_\kappa$): a vector as short as a secret key vector is inserted in the basis at any given position κ.
- Dense Sublattice Discovery (DSD$_\kappa$): a vector strictly longer than

the secret key but belonging to the dense sublattice generated by the secret key is inserted in the basis at any given position κ.

The result (Fig. 7.2) is rather striking: for $n = 127$, we start seeing a deviation from the standard regime for q as small as 700, while a naive interpretation of the prediction by Kirchner and Fouque [KF17] would suggest a fatigue point at $q \approx n^{2.783} \approx 700\,000$. We can conclude either that the asymptotic bound is not tight, or that the hidden asymptotic term (the $o(1)$ in $n^{2.783+o(1)}$) is significantly negative in practice. In any case, the bound of Kirchner and Fouque does not seem to provide accurate concrete predictions.

Remark

At this point, we should clarify why the DSD event should essentially be considered a successful attack. First, for q not too much larger than the fatigue point, an SKR event typically quickly follows after the DSD event; what happens is that DSD events cascade, until the full dense sublattice has been extracted: the first half of the reduced basis precisely generates the dense sublattice. Lattice reduction will happen independently on each half of the basis, meaning that the dimension of the search space for the secret key has effectively been halved, and therefore making the problem much easier.

However, as q increases, DSD becomes easier and easier, to the point that it becomes even easier than secret key recovery within the dense sublattice. In other terms, there is a *superstretched* regime for larger q, where DSD does not directly lead to SKR.

Nevertheless, we argue —essentially rephrasing [ABD16]— that the DSD event is typically sufficient for an attack. First, the dense sublattice vector discovered is of length significantly lower than q; in an FHE scheme such as [BLLN13] it is sufficient to decrypt fresh ciphertexts.[4] Secondly, in the case of cyclotomic or circulant NTRU, it is possible to recover the secret key from the dense sublattice by other means than pure lattice reduction; in particular the recent line of work on the principal ideal-SVP [EHKS14; CDPR16; Bia+17] showed that this can be done classically in subexponential time $\exp(\tilde{O}(\sqrt{n}))$ and quantumly in polynomial time.

[4] The secret key being shorter is only required to deal with ciphertexts obtained by homomorphic computation.

7.1.1 Contributions

Having identified precisely what event distinguishes the standard regime of NTRU from its overstretched regime, we may now proceed to a refined analysis, and determine precisely both the fatigue point and the precise cost[5] of attacks in the overstretched regime. The refined analysis in this work diverges from the one of Kirchner and Fouque [KF17] on the following points:

1. we exploit the fact that BKZ runs SVP on large blocks ($\beta \geq 2$) not only to deduce the shape of the basis, but also to actually discover dense sublattice vectors,

2. we do not solely focus on the behaviour at position $\kappa = n - \beta + 1$ out of $d = 2n$ dimensions, but instead predict the most relevant position,

3. we propose an average-case analysis of volumes of the relevant lattices and sublattices, leading to a concrete prediction rather than a worst-case bound,

4. we also validate the intermediate and final predictions quantitatively with extensive experiments.

We note that contributions 1 and 2 alone already give us an important asymptotic result: the fatigue point of NTRU is indeed lower than predicted by Kirchner and Fouque, namely, it should happen at $q = n^{2.484+o(1)}$ instead of $n^{2.783+o(1)}$.

Furthermore, for the concrete average case analysis we differentiates between the *circulant* version of NTRU [HPS98] and its *matrix* version [CG05; Gen+19]. We note minor deviations in the concrete analysis of volumes of relevant sublattices, that on average slightly favours the attacker in the matrix case, but also shows a larger variance in the concrete hardness of the circulant case. The concrete analysis in this work is versatile, as one only has to estimate the volume of the dense sublattice to analyse a different variant.

[5] In this work, we only measure cost of lattice reduction in terms of the required BKZ blocksize; the computational cost of BKZ is essentially an orthogonal question.

In summary: we achieve an explicative and predictive model for the fatigue of NTRU, with concrete predictions confirmed in practice. In particular, the fatigue point is estimated to be at $q \approx 0.004 \cdot n^{2.484}$ for $n > 100$ and ternary errors. All artefacts for experiments and predictions are open-source[6], and are based on the FPLLL and FPyLLL libraries [tea21a; tea21b].

Impact

We wish to clarify that this work *does not* contradict the concrete security of the NTRU candidate to the NIST competition [Che+20]; on the contrary, we close a pending question regarding a potential vulnerability.

Limitation: the lucky-lifts

During the experiments, we also noted rare occurrence of DSD events that qualitatively differ from what we expected. Namely, the vector from the dense sublattice was found at positions κ quite larger than what was predicted by the model, as shown in Fig. 7.11 (a). More remarkable, these vectors were extremely unbalanced: their $2n - \kappa$ last (Gram-Schmidt) coordinates were much smaller than the κ first coordinates (Fig. 7.11 (b)). We call these DSD events lucky-lifts (DSD-LL), while the one we model and mostly observe are called after the Pataki-Tural Lemma (DSD-PT). Despite those two phenomena being very distinct, they nevertheless occured for the same BKZ blocksizes β, at least in the range of parameters we could experiment with.

It could very well be that these rare DSD-LL events are just artefacts of the modest parameters of the experiments and that these events vanish as the dimension grows. Yet, as they seem of a very different nature, a definitive conclusion would require a dedicated study.

7.1.2 Organisation

We introduce some preliminaries, the NTRU lattice, and the state-of-the-art estimates in Section 7.2. In Section 7.3 we introduce the new DSD-PT estimate and give an asymptotic analysis. In Section 7.4 we

[6] Available at: `https://github.com/WvanWoerden/NTRUFatigue`

give an average-case analysis to construct a concrete estimator. In the final Section 7.5 we compare the estimate with experiments.

7.2 The NTRU lattice and estimates

7.2.1 NTRU and lattice attacks

We start with the historical definition of NTRU.

Definition 99 (NTRU). *Let n be prime, q a positive integer and let $\mathbf{f}, \mathbf{g} \in (\mathbb{Z}/q\mathbb{Z})[X]$ be polynomials of degree n with small coefficients sampled from some distribution χ under the condition that $\mathbf{f}$ is invertible in $\mathcal{R}_q := (\mathbb{Z}/q\mathbb{Z})[X]/(X^n - 1)$. The pair $(\mathbf{f}, \mathbf{g})$ forms the secret key, and the public key is defined as $\mathbf{h} := \mathbf{g}/\mathbf{f} \bmod q$. The* NTRU problem *is to recover any rotation $(X^i \cdot \mathbf{f}, X^i \cdot \mathbf{g})$ of the secret key from $\mathbf{h}$.*

For *NTRUencrypt* [HPS98; Che+20] $\mathbf{f}$ and $\mathbf{g}$ have ternary coefficients, with a fixed number of about $n/3$ of each value in $\{-1, 0, 1\}$. For the analysis we consider the case where each coefficient is sampled from a discrete Gaussian over $\mathbb{Z}$ with some variance $\sigma^2 > 0$. For simplicity the ternary case is treated as a discrete Gaussian with variance $\sigma^2 = \frac{2}{3}$.

More generally we consider a matrix description of NTRU where the polynomials are replaced by matrices $\mathbf{F}, \mathbf{G}, \mathbf{H} \in \mathbb{Z}^{n \times n}$ such that $\mathbf{H} := \mathbf{G} \cdot \mathbf{F}^{-1} \bmod q$ [CG05; Gen+19]. Variants of NTRU, e.g., based on different algebraic rings [Ber+20], can be encoded in the structure of the matrices. For example, the original problem can be encoded by setting $\mathbf{F}_{i,j} := f_{(i+j \bmod n)}$ where $\mathbf{f} = \sum_{i=0}^{n-1} f_i X^i$, for each polynomial respectively. We call the original variant *circulant NTRU*, based on the resulting shape of the matrices $\mathbf{F}, \mathbf{G}$, and we treat $\mathbf{f}, \mathbf{g}$ as n-dimensional vectors. We also consider the variant, called *matrix NTRU*, where the matrices $\mathbf{F}, \mathbf{G}$ have no extra structure and the coefficients are independently sampled from a discrete Gaussian.

To reduce the NTRU problem to a lattice problem we define the *NTRU lattice*, which contains a particularly *dense* sublattice generated by the secret key.

Definition 100. *Let $(n, q, \mathbf{F}, \mathbf{G}, \mathbf{H})$ be an NTRU instance. We define the NTRU lattice as*

$$\mathcal{L}^{\mathbf{H},q} := \begin{pmatrix} q\mathbf{I}_n & \mathbf{H} \\ \mathbf{0} & \mathbf{I}_n \end{pmatrix} \cdot \mathbb{Z}^{2n},$$

and its (secret) dense sublattice of rank n by:

$$\mathcal{L}^{\mathbf{GF}} := \mathbf{B}^{\mathbf{GF}} \cdot \mathbb{Z}^n \subset \mathcal{L}^{\mathbf{H},q}, \text{ where } \mathbf{B}^{\mathbf{GF}} := \begin{pmatrix} \mathbf{G} \\ \mathbf{F} \end{pmatrix}.$$

Solving the NTRU problem is equivalent to recovering the dense sublattice basis $\mathbf{B}^{\mathbf{GF}} = [\mathbf{G}; \mathbf{F}]$ up to some permutation of the columns. For uniformity of notation we will denote such a column by $(\mathbf{g}; \mathbf{f})$. These column vectors have a length of about $\|(\mathbf{g}; \mathbf{f})\| \approx \sqrt{2n\sigma^2}$, which for common parameters is much shorter than the expected minimal length $\mathrm{gh}(\mathcal{L}^{\mathbf{H},q}) \approx \sqrt{nq/(\pi e)}$ of the full lattice $\mathcal{L}^{\mathbf{H},q}$ for a truly uniform random $\mathbf{H} \in (\mathbb{Z}/q\mathbb{Z})^{n \times n}$. To recover the secret key we thus have to find these exceptionally short vectors in the full lattice $\mathcal{L}^{\mathbf{H},q}$.

In [CS97] Coppersmith and Shamir showed that we can slightly relax the problem as any small vector from the dense sublattice $\mathcal{L}^{\mathbf{GF}}$ is enough to decode a message. We therefore focus the analysis on the recovery of elements from $\mathcal{L}^{\mathbf{GF}}$, and not (directly) on the full secret basis $\mathbf{B}^{\mathbf{GF}}$. To recover short vectors we resort to lattice reduction.

7.2.2 Lattice reduction on q-ary lattices

While by now the behaviour of BKZ on random lattices is reasonably understood, this is less the case for q-ary lattices (for certain parameters) such as the NTRU lattice $\mathcal{L}^{\mathbf{H},q}$.

Definition 101 (q-ary lattices). *A lattice $\mathcal{L}$ of dimension d is said to be q-ary if for some $q > 0$ we have*

$$q\mathbb{Z}^d \subset \mathcal{L} \subset \mathbb{Z}^d.$$

Note that the first n basis vectors of $\mathcal{L}^{\mathbf{H},q}$ are orthogonal q-vectors $(q, 0, \ldots, 0)$, $(0, q, 0, \ldots, 0), \ldots$, and so the initial basis profile starts with $\|\tilde{\mathbf{b}}_0\| = \cdots = \|\tilde{\mathbf{b}}_{n-1}\| = q$. Additionally after projecting away from these q-vectors, the remaining basis vectors are again orthogonal

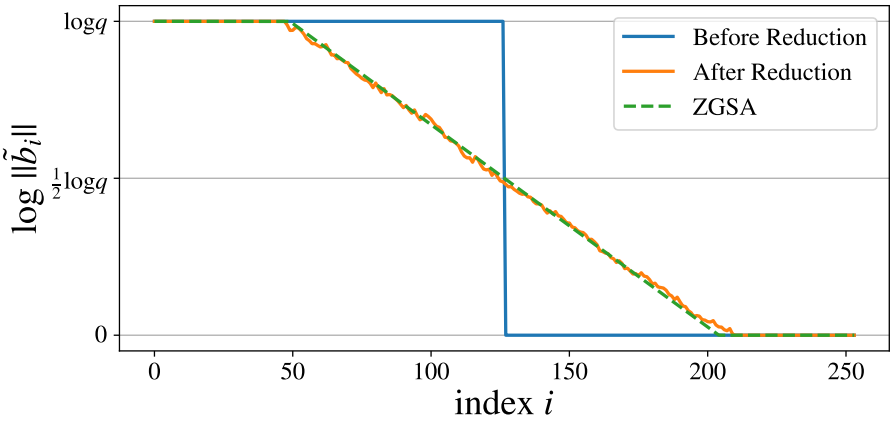

Figure 7.3: The Z-shape of an NTRU-lattice basis with parameters $q = 1031, n = 127$ before and after LLL reduction, versus the ZGSA.

with length 1, and thus we have $\|\tilde{\mathbf{b}}_n\| = \cdots = \|\tilde{\mathbf{b}}_{d-1}\| = 1$. Note that in the BKZ algorithm the length of $\mathbf{b}_0$ can not increase, and is thus always at most q. Also $\mathbf{b}_1$ can not increase in length if $\mathbf{b}_0$ remains unchanged, and so on. For dual-BKZ or the self-dual LLL the profile lengths can not drop below 1 anywhere by the same reasoning. Still LLL and BKZ guarantee that the profile slope in the middle is not too steep. So after LLL reduction the profile must be flat at the start and end, and have a sloped part in the middle, we call this a Z-shape [AD21; AL22]. Because BKZ is not self-dual we do not have any guarantee that the last profile elements do not drop below 1, however we could for example run BKZ only on an appropriate middle context $\mathcal{L}_{[n-m:n+m)}$ to force this behaviour. With this description one would expect the middle part to follow the GSA, leading to an alternative heuristic for q-ary lattices.

Heuristic 102 (Z-shape Geometric Series Assumption (ZGSA)). *Let $\mathbf{B}$ be a basis of a $2n$-dimensional q-ary lattice $\mathcal{L}$ with n q-vectors. After BKZ-β reduction the profile has the following shape:*

$$\|\tilde{\mathbf{b}}_i\| = \begin{cases} q & \text{if } i \leq n-m, \\ \sqrt{q} \cdot \alpha_\beta^{\frac{2n-1-2i}{2}}, & \text{if } n-m < i < n+m-1, \\ 1, & \text{if } i \geq n+m-1, \end{cases}$$

where $\alpha_\beta = \mathrm{gh}(\beta)^{2/(\beta-1)}$, *and* $m = \frac{1}{2} + \frac{\log(q)}{2\log(\alpha_\beta)}$.

Similar to the regular GSA this gives us a good first order estimate, as can be observed in Figure 7.3. Asymptotically setting $\beta = \mathcal{B} \cdot n$ and $q = n^{\mathcal{Q}}$, we obtain $\log(\alpha_\beta) = \frac{\log(n)}{\mathcal{B} \cdot n} + O\left(n^{-1}\right)$, and $m = \frac{1}{2}\mathcal{Q}\mathcal{B} \cdot n + O\left(\frac{n}{\log(n)}\right)$.

7.2.3 Estimates

The main question of this work is to better understand how BKZ recovers the dense sublattice $\mathcal{L}^{\mathbf{GF}}$ from an NTRU lattice $\mathcal{L}^{\mathbf{H},q}$. Several works exist that give estimates on the blocksize β for which BKZ successfully recovers the secret key $(\mathbf{g}, \mathbf{f})$, or more generally a vector from the dense sublattice. We discuss the state-of-the-art estimates, one known as the *2016 Estimate* [ADPS16] with further refinements [DDGR20; PV21], and one by Kirchner and Fouque [KF17].

While the 2016 Estimate already gives a clear explanation *how* BKZ recovers a suitable vector, the Kirchner and Fouque estimate is only based on an impossibility result. To be more precise about what we mean with recovery we define the following two events.

Definition 103 (BKZ Events). *For a BKZ run on an NTRU lattice* $\mathcal{L}$ *with dense sublattice* $\mathcal{L}^{\mathbf{GF}}$ *we define two events:*

1. ***Secret Key Recovery (SKR)****: The first time one the secret keys* $(\mathbf{g}; \mathbf{f})$ *is inserted.*
2. ***Dense Sublattice Discovery (DSD)****: The first time a dense lattice vector* $\mathbf{v} \in \mathcal{L}^{\mathbf{GF}}$ *strictly longer than the secret key(s) is inserted.*

We further specify SKR_κ *and* DSD_κ *when the insertion takes place at position* κ *in the basis.*

2016 Estimate [ADPS16] for SKR

The 2016 Estimate is aimed at the more general problem of detecting an unusually short vector in a lattice. To obtain an estimate for the NTRU problem, and more specifically the SKR event, we apply it to the unusually short vector $(\mathbf{g}; \mathbf{f}) \in \mathcal{L}^{\mathbf{H},q}$.

Heuristic claim 104 (SKR – 2016 Estimate). *Let $\mathcal{L}$ be a lattice of dimension d and let $\mathbf{v} \in \mathcal{L}$ be a unusually short vector $\|\mathbf{v}\| \ll \mathrm{gh}(\mathcal{L})$. Then under the Geometric Series Assumption BKZ recovers $\mathbf{v}$ if*

$$\sqrt{\beta/d} \cdot \|\mathbf{v}\| < \sqrt{\alpha_\beta}^{2\beta-d-1} \cdot \mathrm{vol}(\mathcal{L})^{1/d},$$

where $\alpha_\beta = \mathrm{gh}(\beta)^{2/(\beta-1)}$.

Justification. The left hand side of the inequality is an estimate for $\|\pi_{d-\beta}(\mathbf{v})\|$, while the right hand size is the expected norm of $\tilde{\mathbf{b}}_{d-\beta}$ under the GSA. When the inequality is satisfied we expect that the shortest vector in $\mathcal{L}_{[d-\beta:d)}$ is in fact (a projection of) the unusually short vector, and thus it is inserted by BKZ at position $d-\beta$. See Figure 7.4 for an illustration.

Except for very small blocksizes β, the unusually short vector $\mathbf{v}$ is recovered from its projection $\pi_{d-\beta}(\mathbf{v})$ with high probability, either directly by Babai's nearest plane algorithm, or by later BKZ tours on the blocks $\mathcal{L}_{[d-2\beta+1:d-\beta+1)}, \mathcal{L}_{[d-3\beta+2:d-2\beta+2)}, \ldots$; lifting the vector block by block. △

For q-ary lattices we can easily change the estimate to make use of the ZGSA instead, although for successful blocksizes $\tilde{\mathbf{b}}_{d-\beta}$ will not lie on the flat tail-part, and thus this will not change anything. Additionally for q-ary lattices it can be beneficial to apply the estimate not to the full lattice but on some projected sublattice $\mathcal{L}_{[i:d)}$ for $i \leq n$; the left hand side of the equation is expected to remain unchanged, while the right hand side might increase as $\mathrm{vol}(\mathcal{L})$ loses only a factor q^i and the dimension decreases by i. Note that we do not necessarily have to explicitly let BKZ act on this projected sublattice, as BKZ already does this naturally.

Asymptotics. Consider the NTRU lattice $\mathcal{L}^{\mathbf{H},q}$ and suppose that $q = \Theta(n^{\mathcal{Q}})$, $\|\mathbf{v}\| = \|(\mathbf{g};\mathbf{f})\| = \Theta(n^{\mathcal{S}})$ and $\beta = (\mathcal{B} + o(1))n$. Applying the 2016 Estimate the right hand side of the inequality is minimised when only keeping $k = \min\left((\sqrt{2\mathcal{B}\mathcal{Q}} - 1)n, n\right)$ of the q-vectors, so by applying the estimate to the projected sublattice $\mathcal{L}^{\mathbf{H},q}_{[n-k:2n)}$. For $\mathcal{S} \geq 1$ we have $k = n$, and solving the equation gives $\mathcal{B} = \frac{2}{\mathcal{Q}+2-2\mathcal{S}}$. For $\mathcal{S} < 1$ we have $k = (\sqrt{2\mathcal{B}\mathcal{Q}} - 1)n$, and solving gives $\mathcal{B} = \frac{2\mathcal{Q}}{(\mathcal{Q}+1-\mathcal{S})^2}$. Note in particular that in terms of q we require a blocksize of $\beta = \tilde{\Theta}\left(n/\log(q)\right)$.

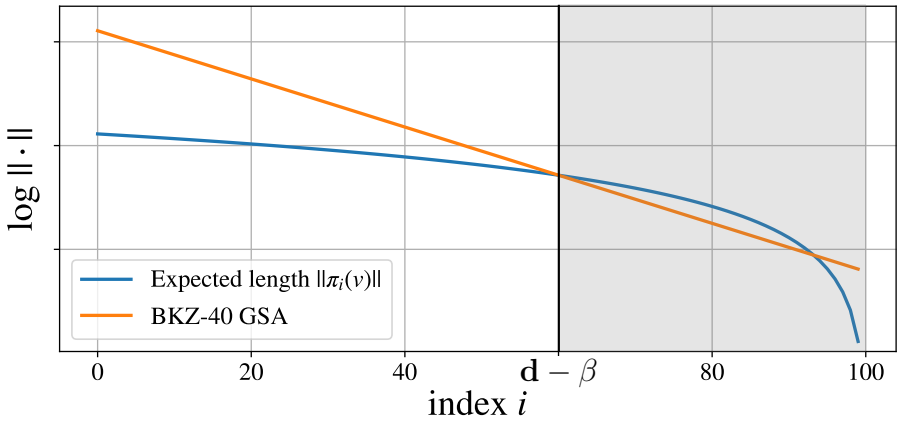

Figure 7.4: Illustration of the 2016 Estimate on a 100-dimensional lattice with an unusually short vector $\mathbf{v}$. Around $\beta = 40$ the projection $\pi_{100-\beta}(\mathbf{v})$ is expected to be the shortest vector in the projected sublattice $\mathcal{L}_{[d-\beta:d)}$, which is thus recovered by the SVP call on this block inside BKZ-β.

Refinements. The 2016 Estimate gives a clear explanation on how and where the secret vector is recovered, namely its projection is found in the last BKZ block. This also allows to further refine the estimate and give concrete predictions. For example by using a BKZ-simulator instead of the GSA, and by accounting for the probability that after the projection $\|\pi_{d-\beta}(\mathbf{v})\|$ has been found, it is successfully lifted to the full vector $\mathbf{v}$. Also instead of working with the expected length of the projection, we can directly model the probability distribution under the assumption that $\mathbf{v}$ is distributed as a Gaussian vector. Such refinements were applied in [DDGR20; PV21], and the resulting concrete predictions match with experiments to recover an unusually short vector. Current simulators for the behaviour of the BKZ algorithm do not account correctly for the Z-shape [AD21]. Therefore, in this work, we will rely on the (Z)GSA for the basis shape, and we adjust the slope to account for the speed of convergence using experimentally determined values. We also use the advanced probabilistic model for the detection and lifting of the short vector.

For NTRU there is not just a single unusually short vector, but there are $n = d/2$ of them, which makes it more likely that at least one

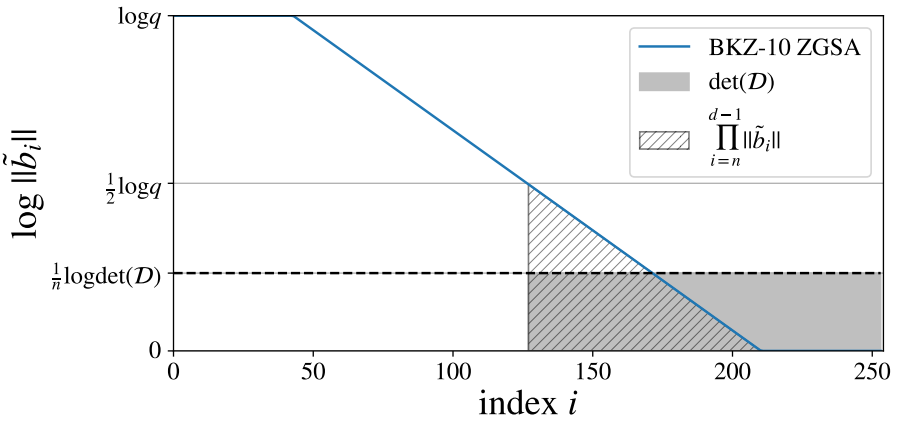

Figure 7.5: Illustration of both sides of the equation when applying Lemma 105 to $\mathcal{D} = \mathcal{L}^{\mathbf{GF}} \subset \mathcal{L}^{\mathbf{H},q}$ and $J = \{n, \ldots, d-1\}$. Following the Pataki and Tural Lemma, the striped area can be at most as large as the solid area. For larger blocksize β, the BKZ-β ZGSA contradicts this, and thus the basis must have deviated from it (and thereby have revealed some information).

of them is recovered. Because the refined concrete estimator already works with a probability distribution, we can easily take multiple vectors into account. The resulting predictions for the SKR event match the experiments reasonably well for smallish q as can be seen in Figure 7.2. For large q, the so-called *overstretched* regime, the estimate is however too pessimistic.

Kirchner–Fouque estimate [KF17] for DSD

In 2016 Albrecht, Bai and Ducas [ABD16] showed that for very large values of q one can mount an algebraic *subfield attack* on the cyclotomic NTRU problem with subexponential or even polynomial complexity. This allowed them to break several homomorphic encryption schemes that relied on NTRU in the overstretched regime.

However soon after, Kirchner–Fouque [KF17] showed that this elaborate algebraic attack was unnecessary: (dual-)BKZ already behaves much better in this regime than the 2016 Estimate predicts,

leading to the same asymptotic improvements. The key idea behind their analysis is that in the overstretched regime the NTRU lattice $\mathcal{L}^{\mathbf{H},q}$ contains an exceptionally dense sublattice $\mathcal{L}^{\mathbf{GF}}$ of low volume. This gives a constraint on the basis profile via the following lemma by Pataki and Tural.

Lemma 105 (Pataki and Tural [PT08]). *Let $\mathcal{L}$ be a d-dimensional lattice with basis $\mathbf{b}_0, \ldots, \mathbf{b}_{d-1}$. For any k-dimensional sublattice $\mathcal{L}' \subset \mathcal{L}$ we have*

$$\mathrm{vol}(\mathcal{L}') \geq \min_J \prod_{j \in J} \|\tilde{\mathbf{b}}_j\|,$$

where J ranges over the k-size subsets of $\{0, \ldots, d-1\}$.

Applying Lemma 105 to the n-dimensional sublattice $\mathcal{L}^{\mathbf{GF}} \subset \mathcal{L}^{\mathbf{H},q}$, and assuming a non-increasing profile, we obtain an upper bound on the volume of $\mathcal{L}^{\mathbf{H},q}_{[n:2n)}$. See Figure 7.5 for an illustration of these volumes. Assuming the ZGSA the latter volume increases when running BKZ-β for increasing blocksizes, eventually contradicting the upper bound. This allows us to detect if a q-ary lattice is in fact an NTRU lattice, but additionally Kirchner–Fouque argue that BKZ must *somehow* have detected the dense sublattice after this point. Based on this impossibility argument they introduced the following estimate.

Heuristic claim 106 (DSD – Kirchner–Fouque Estimate). *Let $\mathcal{L}^{\mathbf{H},q}$ be an NTRU lattice of dimension $2n$, with dense sublattice $\mathcal{L}^{\mathbf{GF}} \subset \mathcal{L}^{\mathbf{H},q}$. Under the Z-shape Geometric Series Assumption BKZ-β triggers the DSD event if*

$$\mathrm{vol}(\mathcal{L}^{\mathbf{GF}}) < q^{\frac{m-1}{2}} \cdot \alpha_\beta^{-\frac{1}{2}(m-1)^2},$$

where $\alpha_\beta = \mathrm{gh}(\beta)^{2/(\beta-1)}$, and $m = \frac{1}{2} + \frac{\log(q)}{2\log(\alpha_\beta)}$.

To apply this estimate we can bound $\mathrm{vol}(\mathcal{L}^{\mathbf{GF}})$ using the Hadamard inequality by $\|(\mathbf{g}; \mathbf{f})\|^n$. As a first approximation this is reasonably tight because the secret basis $\mathbf{B}^{\mathbf{GF}}$ is close to orthogonal.

Asymptotics. Consider the NTRU lattice $\mathcal{L}^{\mathbf{H},q}$ and suppose that $q = \Theta(n^{\mathcal{Q}})$, $\|(\mathbf{g};\mathbf{f})\| = \Theta(n^{\mathcal{S}})$ and $\beta = (\mathcal{B}+o(1))n$. We apply the Kirchner–Fouque Estimate using that $m \approx \frac{\mathcal{B}\mathcal{Q}}{2}n$ and $\alpha_\beta \approx (\mathcal{B}n)^{1/(\mathcal{B}n)}$.The left hand side of the inequality is bounded by $n^{n\mathcal{S}+o(n)}$ and the right hand side equals $n^{\frac{\mathcal{B}\mathcal{Q}^2}{8}n+o(n)}$; solving gives $\mathcal{B} \geq \frac{8\mathcal{S}}{\mathcal{Q}^2}$. Note that in terms of n and q we require a blocksize of $\beta = \tilde{\Theta}\left(n/\log^2(q)\right)$, improving upon the 2016 Estimate by a factor $\log(q)$. So for large enough q the Kirchner–Fouque Estimate predicts a lower successful blocksize than the 2016 Estimate. We call the value of q for which BKZ starts to behave better than predicted by the 2016 Estimate the *fatigue point*. For the common situation that $\mathcal{S} = \frac{1}{2}$, e.g., when each secret coefficient has standard deviation $\sigma = \Theta(1)$, the Kirchner–Fouque Estimate predicts that the fatigue point lies at some $q \leq n^{2.783+o(1)}$.

7.3 A new dense sublattice discovery estimate

7.3.1 Preliminary experiments

Both the 2016 Estimate and the Kirchner–Fouque Estimate analyse an event that leads to successful recovery of a vector of the dense NTRU sublattice. This only gives an upper bound on the hardness; a different event leading to the recovery might happen at a lower blocksize. Additionally the Kirchner–Fouque Estimate is only based on an impossibility result and gives no explanation as to how BKZ actually recovers a vector from the dense sublattice. In order to derive a tight estimate we first run experiments to track down at which point a dense sublattice vector is actually found during the BKZ tours, i.e., when the DSD_κ event is triggered and at what position. Then we model this event in order to hopefully derive a tight estimate.

We run progressive BKZ on NTRU lattices $\mathcal{L}^{\mathbf{H},q}$ for fixed parameters $n = 127, \sigma^2 = \frac{2}{3}$, and several moduli q. For each BKZ insertion at position κ we check if the inserted vector belongs to the dense sublattice $\mathcal{L}^{\mathbf{GF}}$, and thereby if the SKR_κ or DSD_κ event takes place, after which we stop.

The results are shown in Figure 7.2. We take a closer look at the observed SKR_κ and DSD_κ events and where they are triggered. We

can group the observations in three typical circumstances.

- **SKR-2016**. The SKR_κ event is mostly triggered for small values of q, and this mostly happens at the position $\kappa = 2n - \beta$, so in the last block $[2n - \beta : 2n)$, or slightly earlier. This coincides exactly with the $\mathrm{SKR}_{2n-\beta}$ event as predicted by the 2016 Estimate [ADPS16] with further refinements [AGVW17; DDGR20; PV21].

- **DSD-PT**. The DSD_κ event is mostly triggered at positions $\kappa = n + k - \beta$ for $0 < k \ll n$. The inserted dense vector $\mathbf{v}$ is often significantly longer than the secret key but still shorter than the q-vectors. On closer inspection the projected length $\|\pi_{n+k-\beta}(\mathbf{v})\|$ is close to the expected length $\sqrt{\frac{\beta}{n+k}}\|\mathbf{v}\|$ for all instances, more specifically the length of $\mathbf{v}$ is well balanced over the Gram-Schmidt directions $\tilde{\mathbf{b}}_0, \ldots, \tilde{\mathbf{b}}_{n+k-1}$. We name these events after the Pataki–Tural Lemma (DSD-PT).

- **DSD-LL**. For a few instances the DSD_κ event is triggered at large positions κ, up to $2n - \beta$. The inserted dense vector $\mathbf{v}$ is again significantly longer than the secret key, but it has an unexpectedly short projection $\pi_\kappa(\mathbf{v})$ on the BKZ block $[\kappa : \kappa + \beta)$. We call these events lucky-lifts (DSD-LL).

The DSD-LL event could potentially be explained by the relatively large amount of shortish vectors in the close to orthogonal dense sublattice $\mathcal{L}^{\mathbf{GF}}$ compared to what one would expect based on the Gaussian Heuristic. These many vectors might compensate for the low probability event that: (1) such a long vector has such a short projection, and (2) the projected vector is correctly lifted by Babai's nearest plane algorithm (thus a *lucky lift*). The DSD-LL event remains rare for all parameters we used in the experiments, and the successful blocksizes do not seem to deviate from the DSD-PT events. Although we think this circumstance deserves further analysis we therefore base the estimate on the more common DSD-PT event.

For the DSD-PT event the projected length $\|\pi_{n+k-\beta}(\mathbf{v})\|$ is close to $\sqrt{\frac{\beta}{n+k}}\|\mathbf{v}\|$, and thus the inserted dense vector $\mathbf{v}$ must in fact be (close to) a shortest vector of the intersected sublattice $\mathcal{L}^{\mathbf{H},q}_{[0:n+k)} \cap \mathcal{L}^{\mathbf{GF}}$. If not,

the shortest vector would typically have an even smaller projection and would thus be inserted instead. For ease of analysis we therefore assume that $\mathbf{v}$ is a shortest vector of $\mathcal{L}^{\mathbf{H},q}_{[0:n+k)} \cap \mathcal{L}^{\mathbf{GF}}$. In short the new estimate can be described as follows.

Heuristic claim 107 (DSD-PT estimate). *A tour of BKZ-β triggers the DSD event if*

$$\pi_{n+k-\beta}(\mathbf{v}) < \|\tilde{\mathbf{b}}_{n+k-\beta}\|,$$

where $\mathbf{v}$ is a shortest nonzero vector of $\mathcal{L}^{\mathbf{H},q}_{[0:n+k)} \cap \mathcal{L}^{\mathbf{GF}}$ for some $0 < k \leq n$.

7.3.2 Asymptotic analysis

For $0 \leq r \leq 2n$ denote the (possibly trivial) intersected sublattice by $\mathcal{L}^{\mathbf{GF}}_{\cap[0:r)} := \mathcal{L}^{\mathbf{H},q}_{[0:r)} \cap \mathcal{L}^{\mathbf{GF}}$. To directly apply Claim 107 we are interested in the length of $\mathbf{v}$, and thus the value of $\lambda_1\big(\mathcal{L}^{\mathbf{GF}}_{\cap[0:n+k)}\big)$. We break down the analysis into several steps. In order to obtain a bound on the first minimum we first compute a bound on the volume of the intersection $\mathcal{L}^{\mathbf{GF}}_{\cap[0:n+k)}$ in terms of the basis profile and the volume of $\mathcal{L}^{\mathbf{GF}}$. Together with the ZGSA and a simple bound for $\mathrm{vol}(\mathcal{L}^{\mathbf{GF}})$ we can then apply Minkowski's bound on the first minimum. By optimising $\kappa = n+k-\beta$ we obtain a new asymptotic estimate.

Intersection.

To understand the behaviour of the volume of the intersected lattice we first need a small technical Lemma.

Lemma 108 ([DDGR20]). *Given a lattice $\mathcal{L}$ with volume $\mathrm{vol}(\mathcal{L})$, and a primitive dual vector $\mathbf{v} \in \mathcal{L}^*$. Let $\mathbf{v}^\perp$ denote the subspace orthogonal to $\mathbf{v}$. Then $\mathcal{L} \cap \mathbf{v}^\perp$ is a (possibly trivial) lattice with volume $\mathrm{vol}(\mathcal{L} \cap \mathbf{v}^\perp) = \|\mathbf{v}\| \cdot \mathrm{vol}(\mathcal{L})$.*

Recall that we defined $\mathrm{vol}(\{0\}) = 1$ for the trivial lattice. The following Lemma generalises the Pataki–Tural Lemma on which the estimate of Kirchner–Fouque is based. More specifically the Pataki–Tural Lemma only considers the case where the intersection is always trivial ($s = 0$).

Lemma 109 (Generalisation of [PT08]). *Let $\mathcal{L}$ be a d-dimensional lattice with basis $\mathbf{b}_0, \ldots, \mathbf{b}_{d-1}$, and consider the (possibly trivial) sublattice $\mathcal{L}_{[0:s)}$ for $0 \leq s \leq 2n$. For any n-dimensional sublattice $\mathcal{L}' \subset \mathcal{L}$ we have*

$$\mathrm{vol}(\mathcal{L}_{[0:s)} \cap \mathcal{L}') \leq \mathrm{vol}(\mathcal{L}') \cdot \left(\min_J \prod_{j \in J} \|\tilde{\mathbf{b}}_j\| \right)^{-1},$$

where $k := \dim(\mathcal{L}_{[0:s)} \cap \mathcal{L}')$ and J ranges over the $(n-k)$-size subsets of $\{s, \ldots, d-1\}$.

Proof. We write $\mathcal{L}'_{\cap[0:r)} := \mathcal{L}_{[0:r)} \cap \mathcal{L}'$. For $j = k, \ldots, n$ we define $s_j \in \{s, \ldots, d\}$ as the maximal index such that $\dim\left(\mathcal{L}'_{\cap[0:s_j)}\right) = j$, i.e., we obtain the following strict chain of sublattices:

$$\mathcal{L}'_{\cap[0:s)} = \mathcal{L}'_{\cap[0:s_k)} \subsetneq \mathcal{L}'_{\cap[0:s_{k+1})} \subsetneq \cdots \subsetneq \mathcal{L}'_{\cap[0:s_n)} = \mathcal{L}'.$$

Fix $j \in \{k, \ldots, n-1\}$. Because the basis vectors $\mathbf{b}_0, \ldots, \mathbf{b}_{d-1}$ are linearly independent we have that $\mathcal{L}'_{\cap[0:s_{(j+1)})} = \mathcal{L}'_{\cap[0:s_j+1)}$. This allows us to focus on the volume decrease from index $s_j + 1$ to s_j, for which we know that

$$\mathcal{L}'_{\cap[0:s_j)} = \mathcal{L}'_{\cap[0:s_j+1)} \cap (\tilde{\mathbf{b}}_{s_j})^{\perp},$$

where $(\tilde{\mathbf{b}}_{s_j})^{\perp}$ denotes the subspace orthogonal to $\tilde{\mathbf{b}}_{s_j}$. The corresponding dual basis of $\mathbf{b}_0, \ldots, \mathbf{b}_{s_j}$ contains a dual vector $\mathbf{d} \in \mathcal{L}^*_{[0:s_j+1)}$ of length $\left\|\tilde{\mathbf{b}}_{s_j}\right\|^{-1}$ with $\mathrm{span}(\mathbf{d}) = \mathrm{span}(\tilde{\mathbf{b}}_{s_j})$. Let π be the orthogonal projection onto $\mathrm{span}\left(\mathcal{L}'_{\cap[0:s_j+1)}\right)$, then $\pi(\mathbf{d}) \in \left(\mathcal{L}'_{\cap[0:s_j+1)}\right)^*$, and

$$\mathcal{L}'_{\cap[0:s_j)} = \mathcal{L}'_{\cap[0:s_j+1)} \cap \mathbf{d}^{\perp} = \mathcal{L}'_{\cap[0:s_j+1)} \cap \pi(\mathbf{d})^{\perp}.$$

Let $m \in \mathbb{Z}_{\geq 1}$ be such that $\pi(\mathbf{d})/m$ is primitive w.r.t. $\left(\mathcal{L}'_{\cap[0:s_j+1)}\right)^*$, then by Lemma 108 we obtain:

$$\begin{aligned} \mathrm{vol}\left(\mathcal{L}'_{\cap[0:s_j)}\right) &= \mathrm{vol}\left(\mathcal{L}'_{\cap[0:s_j+1)}\right) \cdot \|\pi(\mathbf{d})/m\| \\ &\leq \mathrm{vol}\left(\mathcal{L}'_{\cap[0:s_{j+1})}\right) \cdot \|\tilde{\mathbf{b}}_{s_j}\|^{-1}. \end{aligned}$$

We conclude by chaining the above inequality for $j = k, \ldots, n-1$. □

Before recovering a dense lattice vector we heuristically assume that there is no special relation between the current lattice basis and the dense sublattice. More specific we can consider that the span of $\mathbf{b}_0, \ldots, \mathbf{b}_{n-1}$ and that of $\mathcal{L}^{\mathbf{GF}}$ behave like random n-dimensional subspaces, and thus they have a trivial intersection with high probability in the $2n$-dimensional space. As a direct result we expect that $\dim\left(\mathcal{L}^{\mathbf{GF}}_{\cap[0:n+k)}\right) = k$ for $k = 0, \ldots, n$. Applying this to Lemma 109 we obtain the following corollary.

Corollary 110. *Let $\mathcal{L}^{\mathbf{H},q}$ be an NTRU lattice with dense sublattice $\mathcal{L}^{\mathbf{GF}}$ of dimension n, if $\dim\left(\mathcal{L}^{\mathbf{GF}}_{\cap[0:n+k)}\right) = k$ for some $k \geq 0$, then*

$$\operatorname{vol}\left(\mathcal{L}^{\mathbf{GF}}_{\cap[0:n+k)}\right) \leq \operatorname{vol}\left(\mathcal{L}^{\mathbf{GF}}\right) \cdot \left(\prod_{j=n+k}^{d-1} \|\tilde{\mathbf{b}}_j\|\right)^{-1}.$$

Note that Corollary 110 already shows that the new estimate can not be worse than the Kirchner–Fouque Estimate. Namely if the Kirchner–Fouque Estimate is triggered, then for intersection dimension $k = 1$ the right hand side is smaller than $\|\tilde{\mathbf{b}}_n\|$. Assuming a non-decreasing profile we then have $\lambda_1\left(\mathcal{L}^{\mathbf{GF}}_{\cap[0:n+1)}\right) = \operatorname{vol}\left(\mathcal{L}^{\mathbf{GF}}_{\cap[0:n+1)}\right) \leq \|\tilde{\mathbf{b}}_n\| \leq \left\|\tilde{\mathbf{b}}_{n+1-\beta}\right\|$, which implies that BKZ-β would find a dense sublattice vector in the block $\mathcal{L}_{[n+1-\beta:n+1)}$ (or earlier).

Volume dense sublattice.

To use Corollary 110 we also need to bound the volume of the dense sublattice $\mathcal{L}^{\mathbf{GF}}$. Because the secret basis is close to orthogonal the Hadamard Inequality $\operatorname{vol}(\mathcal{L}^{\mathbf{GF}}) \leq \|(\mathbf{g}; \mathbf{f})\|^n$ is sufficient as a first order approximation.

Conclusion.

To obtain a heuristic asymptotic estimate we will assume that before finding a dense lattice vector the basis follows the ZGSA shape.

Heuristic claim 111. *The BKZ algorithm with blocksize $\beta = \mathcal{B}n$ applied to an NTRU instance with parameters $q = \Theta(n^{\mathcal{Q}})$, $\|(\mathbf{g}; \mathbf{f})\| =$*

$O(n^{\mathcal{S}})$ triggers the DSD event if

$$\mathcal{B} = \frac{8\mathcal{S}}{\mathcal{Q}^2 + 1} + o(1).$$

Justification. By the Hadamard Inequality we have $\log(\mathrm{vol}(\mathcal{L}^{\mathbf{GF}})) \leq \mathcal{S}n\log(n) + O(n)$. Let $k := \mathcal{K}n > 0$ for some constant $0 < \mathcal{K} \leq 1$. Heuristically we expect that $\dim(\mathcal{L}^{\mathbf{H},q}_{[0:n+k)} \cap \mathcal{L}^{\mathbf{GF}}) = k$, and thus by Corollary 110 and by assuming the ZGSA we obtain a bound on the volume of the intersected sublattice:

$$\begin{aligned}
&\log(\mathrm{vol}(\mathcal{L}^{\mathbf{H},q}_{[0:n+k)} \cap \mathcal{L}^{\mathbf{GF}})) \\
&\leq \mathcal{S}n\log(n) - \frac{1}{2}\sum_{i=n+k}^{n+m-1}\left(Q + \frac{2n-1-2i}{\mathcal{B}n}\right)\log(n) + O(n) \\
&= \mathcal{S}n\log(n) - \frac{(\mathcal{B}\mathcal{Q}-2\mathcal{K})^2}{8\mathcal{B}}n\log(n) + O(n).
\end{aligned}$$

By Minkowski's bound we bound the first minimum using the above volume

$$\begin{aligned}
\log(\lambda_1(\mathcal{L}^{\mathbf{H},q}_{[0:n+k)} \cap \mathcal{L}^{\mathbf{GF}})) &\leq \frac{1}{2}\log(\mathcal{K}n) \\
&\quad + \frac{\log(\mathrm{vol}(\mathcal{L}^{\mathbf{H},q}_{[0:n+k)} \cap \mathcal{L}^{\mathbf{GF}}))}{\mathcal{K}n} + O(1) \\
&\leq \left(-\frac{(\mathcal{B}\mathcal{Q}-2\mathcal{K})^2}{8\mathcal{B}\mathcal{K}} + \frac{\mathcal{S}}{\mathcal{K}} + \frac{1}{2}\right)\log(n) + O(1).
\end{aligned}$$

After projecting onto the block $[n+k-\beta : n+k)$ the above short vector does not increase in length.[7] BKZ detects the projected dense lattice vector in this block if the length is less than $\left\|\tilde{\mathbf{b}}_{n+k-\beta}\right\| = \left(\frac{1}{2}\mathcal{Q} + \frac{\mathcal{B}-\mathcal{K}}{\mathcal{B}}\right)\log(n) + O(1)$. Solving for $\mathcal{B}$ shows that this is the case when

$$\mathcal{B} \geq \frac{2\sqrt{(2\mathcal{S}-\mathcal{K})^2 + \mathcal{K}^2\mathcal{Q}^2} + 2(2\mathcal{S}-\mathcal{K})}{\mathcal{Q}^2} + o(1).$$

[7] One may also be concerned that the short vector would collapse to the zero vector $\mathbf{0}$ after projection onto the block $[n+k-\beta : n+k)$, but this becomes increasingly unlikely as the dimension β of the block grows.

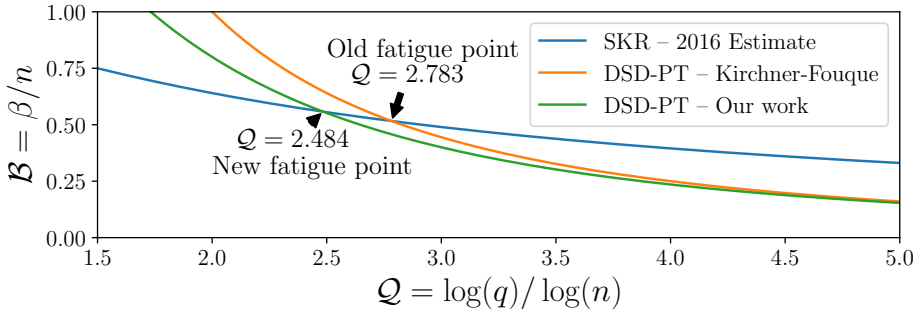

Figure 7.6: Comparison of asymptotic estimates and new fatigue point for $n \to \infty$ when the secret key coefficients have standard deviation $\sigma = \Theta(1)$.

When $\mathcal{K} = \frac{4\mathcal{S}}{\mathcal{Q}^2+1}$ the right hand side is minimised and we obtain that BKZ detects the projected dense lattice vector when $\mathcal{B} \geq \frac{8\mathcal{S}}{\mathcal{Q}^2+1}$, which concludes the claim. This routine computation can be verified symbolically via the sage notebook `claim3_5.ipynb`[8]. △

The new estimate gives an asymptotic improvement over the estimate by Kirchner and Fouque ($\frac{8\mathcal{S}}{\mathcal{Q}^2}$). Asymptotically the optimal position is at $\kappa = n + k - \beta \approx n - \frac{1}{2}\beta$. Interestingly, if we do not optimize k and only consider $k = O(1)$ we obtain the same asymptotic estimate as Kirchner and Fouque, which again emphasizes that we generalised their analysis.

For the fatigue point we compare the relative blocksize of $\frac{8\mathcal{S}}{\mathcal{Q}^2+1}$ to that of the 2016 Estimate given by $\frac{2\mathcal{Q}}{(\mathcal{Q}+1-\mathcal{S})^2}$ for $\mathcal{S} < 1$ and by $\frac{2}{\mathcal{Q}+2-2\mathcal{S}}$ for $\mathcal{S} \geq 1$. For ternary secrets ($\mathcal{S} = \frac{1}{2}$) this narrows down the fatigue point from $q \leq n^{2.783+o(1)}$ to $q = n^{2.484+o(1)}$ compared to the Kirchner–Fouque Estimate. This is still far above the (sub)linear parameters used for NTRU encryption schemes, and thus asymptotically we can close the pending question if these parameters fall in the weaker overstretched regime or not. In practice however we do observe fatigue points that are significantly lower than the naive value of $q = n^{2.484}$, which motivates a concrete analysis with concrete predictions.

[8] Available at: https://github.com/WvanWoerden/NTRUfatigue

7.4 A concrete average-case analysis

In this section we consider a concrete analysis of the new DSD-PT estimate, based on simple heuristics, to better predict the behaviour in practice, and to show that the analysis matches experiments and is thus likely to be tight. The first order asymptotics shown in Section 7.3.2 will remain unchanged, but the differences are significant for practical parameters. Again we split the analysis into several steps, but now derive heuristic expectations instead of loose upper bounds.

We assume that lattice vectors we encounter follow the Gaussian heuristic, and thus in particular that vectors are spherically distributed after normalisation. When projecting such vectors to a lower dimension they become shorter. The following Lemma shows how much shorter we expect them to become.

Lemma 112. *Let $\mathbf{x} \in \mathcal{S}^{d-1}$ follow a spherical distribution, and let $\pi_V : \mathbb{R}^d \to V$ be a projection to some subspace $V \subset \mathbb{R}^d$ of rank $k > 0$, then*

$$\mathbb{E}[\log(\|\pi_V(\mathbf{x})\|)] = \frac{1}{2}(\psi(k/2) - \psi(d/2)),$$

where ψ is the digamma function.

Proof. Let $X_0, \ldots, X_{d-1}$ be standard normal random variables, then the vector $\mathbf{x} = (x_0, \ldots, x_{d-1})$, with $x_j = X_j / \sqrt{\sum_{i=0}^{d-1} X_i^2}$, is spherically distributed. Without loss of generality we can assume that π_V projects onto the first k-coordinates. Then we conclude by Lemma 53 that

$$\begin{aligned}
\mathbb{E}[\log(\|\pi_V(\mathbf{x})\|)] &= \frac{1}{2}\mathbb{E}\left[\log\left(\frac{\sum_{i=0}^{k-1} X_i^2}{\sum_{i=0}^{d-1} X_i^2}\right)\right] \\
&= \frac{1}{2}\mathbb{E}\left[\log\left(\sum_{i=0}^{k-1} X_i^2\right)\right] - \frac{1}{2}\mathbb{E}\left[\log\left(\sum_{i=0}^{d-1} X_i^2\right)\right] \\
&= \frac{1}{2}(\psi(k/2) - \psi(d/2)).
\end{aligned}$$

□

7.4.1 Intersection

We start by giving a concrete average-case estimate for the intersection volumes. Assuming that projections behave as random we obtain the following concrete estimate.

Heuristic claim 113. *Let $\mathcal{L}$ be a $2n$-dimensional NTRU lattice with dense sublattice $\mathcal{L}^{\mathbf{GF}}$, before the DSD event is triggered we have for $k = 1, \ldots, n$ that* $\dim\left(\mathcal{L}^{\mathbf{GF}}_{\cap[0:n+k)}\right) = k$, *and*

$$\mathbb{E}[\log \mathrm{vol}(\mathcal{L}^{\mathbf{GF}}_{\cap[0:n+k)})] = \log \mathrm{vol}(\mathcal{L}^{\mathbf{GF}}) - \left(\sum_{j=n+k}^{2n-1} \log\|\tilde{\mathbf{b}}_j\|\right) + \sum_{l=k+1}^{n} \psi\left(\frac{l}{2}\right) - \psi\left(\frac{n+l}{2}\right) + \frac{\zeta'(l)}{\zeta(l)},$$

where $\zeta(l) := \sum_{m=1}^{\infty} \frac{1}{m^l}$ is the Riemann zeta function and $\zeta'(l) := \sum_{m=1}^{\infty} \frac{\log(m)}{m^l}$ its derivative.

Justification. We follow the proof of Lemma 109. It is tight except for the length decrease from $\|\mathbf{d}\|$ to the projected and primitive vector $\|\pi(\mathbf{d})\|/m$. Note that when obtaining $\log \mathrm{vol}\left(\mathcal{L}^{\mathbf{GF}}_{\cap[0:n+l-1)}\right)$ from $\log \mathrm{vol}\left(\mathcal{L}^{\mathbf{GF}}_{\cap[0:n+l)}\right)$ for some $l = k+1, \ldots, n$, the dual vector $\mathbf{d}$ lives in a $(n+l)$-dimensional space and is projected to an l-dimensional space. Heuristically we assume that the normalisation of $\mathbf{d}$ is spherically distributed (or that π projects to a random l-dimensional subspace). By Lemma 112 the log-expected decrease in length from this projection then equals

$$\mathbb{E}\left[\log(\pi(\|\mathbf{d}\|)) - \log(\|\mathbf{d}\|)\right] = \psi\left(\frac{l}{2}\right) - \psi\left(\frac{n+l}{2}\right).$$

To conclude we also have to include the primitivity of $\pi(\mathbf{d})$ and thus the log-expectation of $m \geq 1$ such that $\pi(\mathbf{d})/m$ is primitive. For any basis $\mathbf{d}_0, \ldots, \mathbf{d}_{l-1}$ and $\pi(\mathbf{d}) = \sum_{i=0}^{l-1} x_i \mathbf{d}_i$, the scaling is given by $m = \gcd(x_0, \ldots, x_{l-1})$. Heuristically we assume that the absolute coefficients $|x_0|, \ldots, |x_{l-1}|$ are random integers in the interval $\{1, \ldots, B\}$

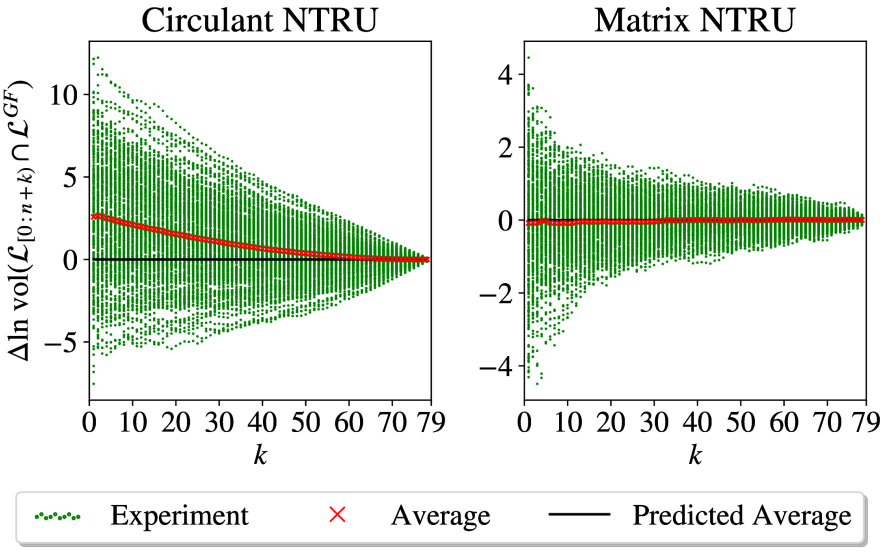

Figure 7.7: Experimental values of $\log \mathrm{vol}\left(\mathcal{L}^{\mathbf{GF}}_{\cap[0:n+k)}\right)$ versus Claim 113 for circulant and matrix NTRU respectively. For each variant we used 256 LLL reduced NTRU lattices with parameters $q = 257, n = 79, \sigma^2 = \frac{2}{3}$ and computed the intersection for each k.

and we let $B \to \infty$. For $l \geq 2$ we have (see e.g., [DE+04])

$$\mathbb{P}_{\mathbf{x}\in\{1,\ldots,B\}^l}\left[\gcd(x_0,\ldots,x_{l-1}) = m\right] = \frac{1}{\zeta(l)} \cdot \frac{1}{m^l} + O(\log(B)/(Bm^{l-1})),$$

where the Riemann zeta function $\zeta(l) = \sum_{m=1}^{\infty} \frac{1}{m^l}$ is just the normalisation factor. From this we conclude that

$$\begin{aligned} &\lim_{B\to\infty} \mathbb{E}_{\mathbf{x}\in\{1,\ldots,B\}^l}[\log \gcd(x_0,\ldots,x_{l-1})] \\ &= \lim_{B\to\infty} \frac{1}{\zeta(l)} \sum_{m=1}^{B} \left[\frac{\log(m)}{m^l} + O\left(\frac{\log(m)\log(B)}{Bm^{l-1}}\right)\right] = -\frac{\zeta'(l)}{\zeta(l)} \end{aligned}$$

for $l \geq k+1 \geq 2$. △

Validation.

To validate Claim 113 we computed the actual intersection volumes $\mathrm{vol}\left(\mathcal{L}^{\mathbf{GF}}_{\cap[0:n+k)}\right)$ for LLL reduced NTRU instances. We observed here, and also in further experiments, that the assumption on the dimension $\dim\left(\mathcal{L}^{\mathbf{GF}}_{\cap[0:n+k)}\right) = k$ holds before we get close to triggering the DSD event. Figure 7.7 shows that the prediction perfectly matches the experiments for matrix NTRU. For circulant NTRU we see both that the expectation is slightly off and that the variance is much higher. The higher variance can be explained from the fact that the projections are very much dependent due to the circulant structure; in fact a closer inspection shows that for k close to n the differences with the prediction are highly correlated. We were not able to explain the error in the predicted expectation, but it seems to be caused by the circulant structure in combination with the Z-shape: the error decreased and eventually disappeared for large values of q and σ, for which the Z-shape disappeared (and before the DSD event was triggered). A maximal log-error of 2.5 is reached at $k = 1$. Note that a log-error of ϵ on $\mathrm{vol}\left(\mathcal{L}^{\mathbf{GF}}_{\cap[0:n+k)}\right)$ translate into a factor of $e^{\epsilon/k}$ on the predicted length for the shortest vector. Except for very small k, this error appears benign.

7.4.2 Dense sublattice

In this section we give a concrete estimate for the expected volume of the dense NTRU sublatice $\mathcal{L}^{\mathbf{GF}}$. Directly from the construction we obtain a basis $[\mathbf{G}; \mathbf{F}]$ of $\mathcal{L}^{\mathbf{GF}}$, with $\mathbf{F}$ invertible. We consider two cases, that of regular NTRU, where $\mathbf{F}$ and $\mathbf{G}$ are circulant matrices, and that of matrix NTRU, where all entries are independently sampled. For both constructions the entries are sampled from independent discrete Gaussians over $\mathbb{Z}$, with some standard deviation $\sigma > 0$. As the only heuristic we assume that the individual entries in fact follow a *continuous* Gaussian instead of the discrete one.

Matrix NTRU.

We start with matrix NTRU, where we heuristically assume that all $2n \times n$ coefficients of the basis $[\mathbf{G}; \mathbf{F}]$ are sampled according to inde-

pendent continuous Gaussians with standard deviation σ. Under this heuristic we can derive an exact expression for the expected log-volume of the dense sublattice.

Lemma 114. *Let* $[\mathbf{G}; \mathbf{F}]$ *be a basis of the lattice* $\mathcal{L}^{\mathbf{GF}}$ *where all sampled entries are i.i.d. continous Gaussians with standard deviation* $\sigma > 0$*, then*

$$\mathbb{E}[\log(\mathrm{vol}(\mathcal{L}^{\mathbf{GF}}))] = \frac{1}{2}n\left(\log(2\sigma^2) - \psi(n)\right) + \sum_{i=0}^{n-1}\psi\left(\frac{2n-i}{2}\right).$$

Proof. By Lemma 53 the log-expectation of the norm of each basis element equals $(\ln(2\sigma^2) + \psi(n))/2$. Note that the i-th Gram-Schmidt vector $\tilde{\mathbf{b}}_i$ is obtained after projecting the i-th basis vector orthogonally away from an i-dimensional subspace, and thus onto a $2n-i$ dimensional subspace. However after normalisation the basis vectors follow a spherical distribution and thus by Lemma 112 we have

$$\begin{aligned}\mathbb{E}[\log\|\tilde{\mathbf{b}}_i\|] &= (\ln(2\sigma^2) + \psi(n))/2 + \psi\left(\frac{2n-i}{2}\right) - \psi(n)\\ &= (\ln(2\sigma^2) - \psi(n))/2 + \psi\left(\frac{2n-i}{2}\right).\end{aligned}$$

We conclude by noting that $\mathbb{E}[\log(\mathrm{vol}(\mathcal{L}^{\mathbf{GF}}))] = \sum_{i=0}^{n-1}\mathbb{E}[\log\|\tilde{\mathbf{b}}_i\|]$. □

Circulant NTRU.

For circulant NTRU both $\mathbf{G}$ and $\mathbf{F}$ in the basis $[\mathbf{G}; \mathbf{F}]$ are circulant matrices. Again we replace discrete with continuous Gaussians. The eigenvalues and eigenvectors of a circulant matrix are well known and we use this to obtain an exact expression for the expected volume of the dense sublattice.

Lemma 115. *Let* $[\mathbf{G}; \mathbf{F}]$ *be a basis of the lattice* $\mathcal{L}^{\mathbf{GF}}$ *where* $\mathbf{G}, \mathbf{F}$ *are circulant and all sampled entries are i.i.d. continous Gaussians with with standard deviation* $\sigma > 0$*, then*

$$\mathbb{E}[\log(\mathrm{vol}(\mathcal{L}^{\mathbf{GF}}))] = \frac{1}{2}n\left(\log(2n\sigma^2) + \psi(1)\right) + \frac{1}{2}(n-1)(1-\log(2)).$$

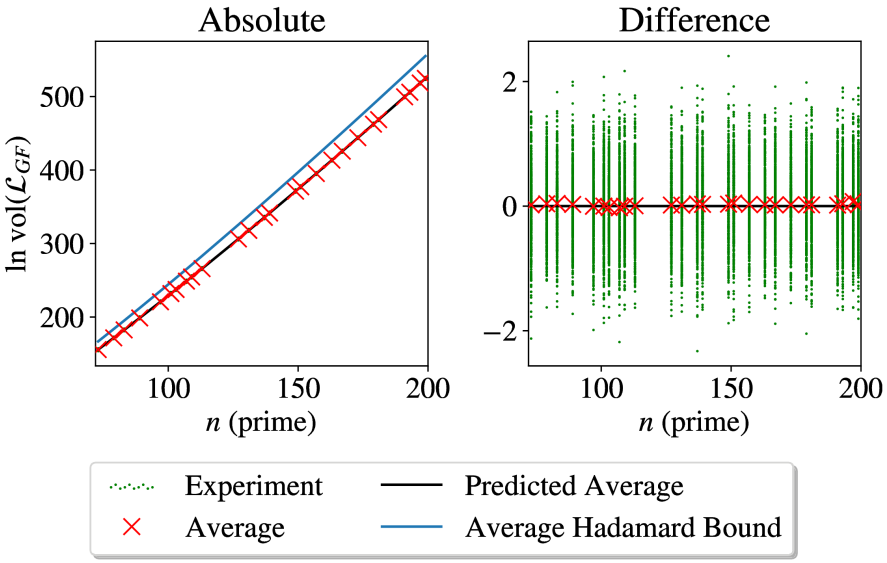

Figure 7.8: Experimental values of $\log(\mathrm{vol}(\mathcal{L}^{\mathbf{GF}}))$ versus Lemma 114 for matrix NTRU with discrete Gaussians and variance $\sigma^2 = \frac{2}{3}$. For each parameter n we generated 512 instances.

Proof. For $n \times n$ circulant matrices $\mathbf{G}, \mathbf{F}$ the eigenvectors are identical and given by $\mathbf{v}_j := (1, \omega^j, \omega^{2j}, \omega^{(n-1)j})$ for $j = 0, \ldots, n-1$, where $\omega := e^{2\pi i/n} \in \mathbb{C}$ is a primitive n-th root of unity. Suppose that the circulant matrix $\mathbf{G}$ is generated by the vector $\mathbf{c} = (c_0, \ldots, c_{n-1})$, then the corresponding eigenvalues are given by the DFT coefficients of $\mathbf{c}$, namely $\lambda_j := c_0 + c_{n-1}\omega^j + \ldots + c_1\omega^{(n-1)j}$. We have that $\lambda_0 = \sum_{j=0}^{n-1} c_j$, and thus λ_0 follows a Gaussian distribution with variance $n\sigma^2$, and in particular $\lambda_0^2 \sim \chi^2_{1,n\sigma^2}$. Additionally for $j = 1, \ldots, n-1$ we can write $\lambda_j = X + i \cdot Y \in \mathbb{C}$ where $X, Y \in \mathbb{R}$ are both linear combinations of the c_i's and thus (X, Y) follows a jointly Gaussion distribution. A simple computation shows that X and Y both have variance $n\sigma^2/2$ and that they are uncorrelated, which for Gaussians implies that they are independent [PP02, p. 212]. So $|\lambda_j|^2 = X^2 + Y^2 \sim \chi^2_{2,n\sigma^2/2}$. Note that all circulant matrices have the same eigenvectors and thus the squared singular values of the concatenation of two circulant matrices are the sum of the squared absolute eigenvalues. So $[\mathbf{G}; \mathbf{F}]$ has one squared singular value s_0^2 distributed as $\chi^2_{1,n\sigma^2} + \chi^2_{1,n\sigma^2} = \chi^2_{2,n\sigma^2}$, and $n-1$

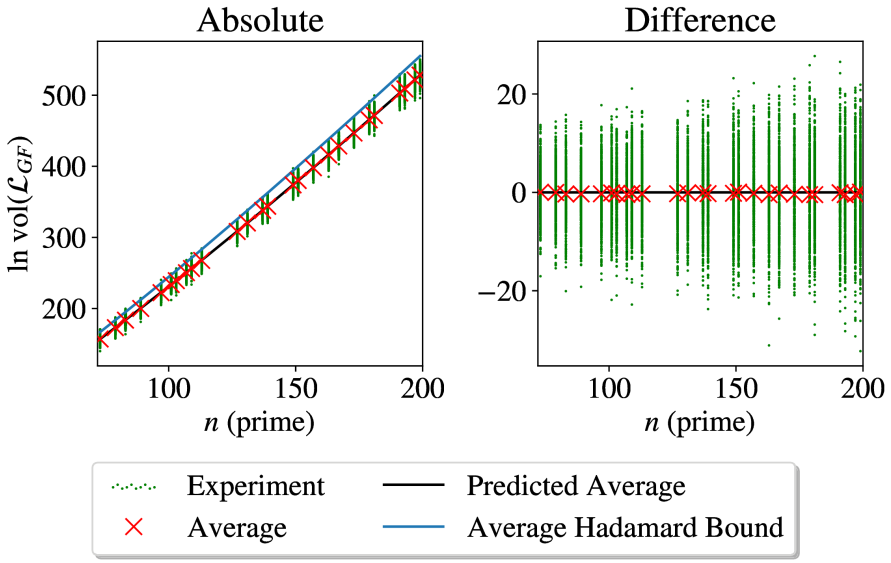

Figure 7.9: Experimental values of $\log(\mathrm{vol}(\mathcal{L}^{\mathbf{GF}}))$ versus Lemma 115 for circulant NTRU with discrete Gaussians and variance $\sigma^2 = \frac{2}{3}$. For each parameter n we generated 512 instances.

squared singular values $s_1^2, \ldots, s_{n-1}^2$ distributed as $\chi^2_{2,n\sigma^2/2} + \chi^2_{2,n\sigma^2/2} = \chi^2_{4,n\sigma^2/2}$. By Lemma 53 they have a log-expectation of

$$\mathbb{E}[\log s_0^2] = \log(2n\sigma^2) + \psi(1), \text{ and } \mathbb{E}[\log s_j^2] = \log(n\sigma^2) + \psi(2)$$

for $j = 1, \ldots, n-1$. We conclude by noting that

$$\mathbb{E}[\log(\mathrm{vol}(\mathcal{L}^{\mathbf{GF}}))] = \frac{1}{2}\sum_{i=0}^{n-1} \log(s_i^2).$$

□

Validation.

To validate the concrete estimate for $\mathrm{vol}(\mathcal{L}^{\mathbf{GF}})$ we generated the NTRU sublattice for several dimensions and computed its volume. We sample the secret coefficients following a discrete Gaussian with variance $\sigma^2 = \frac{2}{3}$ and ran experiments for both matrix NTRU and circulant

NTRU. In Figures 7.8 and 7.9 we see that the predictions from Lemmas 114 and 115 perfectly fit the observed volumes in all dimensions. We do note that the variance is quite significant for the circulant case, but it can be fully explained by the computed eigenvalue distributions in the proof of Lemma 115. We will later see that this high variance has a large influence on the successful BKZ blocksize (Section 7.5.1, Figure 7.11).

7.4.3 Further refinements

We discuss some further refinements, some of which were already successfully applied to the 2016 Estimate [AGVW17; DDGR20; PV21].

Gaussian Heuristic. For the asymptotic analysis we used Minkowski's bound to estimate the length $\lambda_1\left(\mathcal{L}^{\mathbf{GF}}_{\cap[0:n+k)}\right)$ in terms of the volume $\mathrm{vol}\left(\mathcal{L}^{\mathbf{GF}}_{\cap[0:n+k)}\right)$. A natural way to obtain a concrete estimate for the expected minimal length is by assuming that the intersection $\mathcal{L}^{\mathbf{GF}}_{\cap[0:n+k)}$ follows the Gaussian Heuristic and thus for the prediction we assume that

$$\lambda_1\left(\mathcal{L}^{\mathbf{GF}}_{\cap[0:n+k)}\right) = \mathrm{gh}(\mathcal{L}^{\mathbf{GF}}_{\cap[0:n+k)}) \approx \sqrt{k/(2\pi e)} \cdot \mathrm{vol}\left(\mathcal{L}^{\mathbf{GF}}_{\cap[0:n+k)}\right)^{1/k}.$$

We should however be careful with this assumption, as in fact it is false for $k = n$. E.g., the above predicts that $\lambda_1(\mathcal{L}^{\mathbf{GF}}) \approx \sqrt{n/(2\pi e)} \cdot \sqrt{2n\sigma^2}$, while we know that $\lambda_1(\mathcal{L}^{\mathbf{GF}}) = \|(\mathbf{g};\mathbf{f})\| \approx \sqrt{2n\sigma^2}$, a factor $\Theta(\sqrt{n})$ shorter than predicted. The reason for this is that the dense sublattice is up to rotation and scaling very similar to the orthogonal lattice $\mathbb{Z}^n$, precisely the lattice for which it is well known that the Gaussian Heuristic is false. For small $k \ll n$ we do observe that the intersected lattice $\mathcal{L}^{\mathbf{GF}}_{\cap[0:s)}$ follows the Gaussian Heuristic; the orthogonal structure seems to be broken by the intersection. However we do not have a clear idea how large k can become before the orthogonal structure returns and the minimal length stops following the prediction from the Gaussian Heuristic. We think this behaviour deserves some further investigation, e.g., if the transition is very sudden or not, and we leave it as an open problem. This near-orthogonality of $\mathcal{L}^{\mathbf{GF}}_{\cap[0:s)}$ may be critical to model the DSD-LL events.

Probabilities. So far we have only considered expectations of volumes and projections. While this is enough to give a rough concrete estimate we want to be more precise. Success probabilities can accumulate up over multiple BKZ blocks and (progressive) tours, possibly leading to success at much lower blocksizes than the rough estimate. We continue using the expected values for the volume of the dense sublattice and the intersection volumes to obtain the expected length $\lambda_1\big(\mathcal{L}^{\mathbf{GF}}_{\cap[0:s)}\big)$ of the dense sublattice vector via the Gaussian Heuristic. However we then model the short dense sublattice vector $\mathbf{v} \in \mathcal{L}^{\mathbf{GF}}_{\cap[0:s)}$ as an s-dimensional Gaussian vector with the same expected length; allowing us to compute the exact probability that $\big\|\pi_{s-\beta}(\mathbf{v})\big\| \leq \big\|\tilde{\mathbf{b}}_{s-\beta}\big\|$ using the CDF of the chi-square distribution with β degrees of freedom.

Up to now we have ignored the probability that after $\pi_{s-\beta}(\mathbf{v})$ is inserted, it is also correctly lifted to the full vector $\mathbf{v}$ by later BKZ tours. While this almost always happens for higher blocksizes, it is not so likely for lower blocksizes, and ignoring this leads to overly optimistic predictions. For BKZ-β to successfully lift or *eventually* pull the vector $\mathbf{v}$ to the front it should also satisfy $\|\pi_i(\mathbf{v})\| \leq \|\tilde{\mathbf{b}}_i\|$ for all $i = s-2\beta+1, s-3\beta+2, \ldots$. These conditions are not independent which makes them hard to compute exactly. We simplify the computation by only considering the dependence for consecutive positions $i, i-\beta+1$ as done in [DDGR20]. We iteratively run the estimator for progressive $\beta = 2, 3, \ldots$ and take account of all probabilities assuming that all tours behave completely independently. The new concrete estimate will be the expected successful blocksize. Additionally this allows us to combine both the (probabilistic) SKR 2016 Estimate and the new DSD-PT estimate in a single estimator. With some more administration we can also predict the distribution of the successful location κ, and predict the probability that the SKR event happens before the DSD event.

BKZ shape for low blocksizes. While the formulas for the (Z)GSA slope α_β and the expected first minimum $\mathrm{gh}(\beta)$ convert to the experimental values for large blocksizes of say $\beta \geq 50$, they are not as accurate for small β. As expected the convergence is worse for progressive BKZ when we only use a few tours of each blocksize. We ran some experiment on random low dimensional q-ary lattices to obtain practical estimates for $\mathrm{gh}(\beta)$ with $\beta \leq 50$. Earlier works about the

2016 Estimate resorted to BKZ simulators to predict the BKZ shape, which account for the number of tours and also the special shape of the head and tail that do not perfectly follow the GSA shape. Together with the earlier mentioned refinements this resulted in very precise predictions [DDGR20; PV21]. However how BKZ acts on a Z-shaped basis is much less understood [AD21] and as of yet there are no accurate BKZ simulators. Understanding the behaviour and creating an accurate simulator would be very interesting, but is out of the scope of this work. We continue using the ZGSA, but we resort to experimental values for α_β obtained by running BKZ on random q-ary lattices for large q. To remain consistent we also do not use a simulator for the GSA shape, and accept the small discrepancy between the predictions and practical experiments.

7.5 Experimental validation

7.5.1 Successful blocksize

We start with comparing the concrete predictions to the preliminary experiment from Section 7.3.1. We ran progressive BKZ with 8 tours on matrix NTRU instances with parameters $n = 127, \sigma^2 = \frac{2}{3}$ for several moduli q. In Figure 7.10 we show the blocksizes at which the SKR or DSD event is first detected, and compare them to the concrete estimator. We ran the estimator three times for each modulus q: only accounting for SKR, only accounting for DSD-PT, and accounting for both. Note that the combined estimate can be strictly lower than both the first two because the probabilities to succeed accumulate over both events. We calibrated the values of α_β by running the same BKZ routine on $(2 \cdot 127)$-dimensional q-ary lattices with $q \approx 2^{20}$.

We observe that the experiments match the estimates reasonably well, with an average blocksize error of less than 2 for the DSD events and less than 3 for the SKR events. We shortly discuss potential sources of the small errors error.

- We do not actually run the classical BKZ algorithm, but the BKZ 2.0 algorithm as it is more feasible to run for large blocksizes. One part of the latter algorithm is that in each BKZ block $[\kappa : \kappa + \beta)$ the last $\beta - 1$ vectors are randomised before finding a

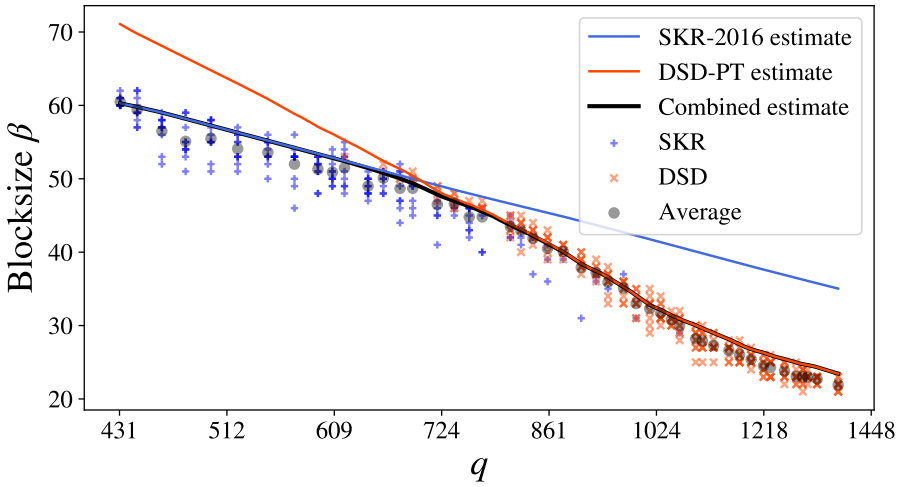

Figure 7.10: Experiment versus prediction for progressive BKZ with 8 tours on matrix NTRU instances with parameters $n = 127, \sigma^2 = \frac{2}{3}$ for several moduli q. We did 10 runs per modulus q.

short projected vector. This temporarily breaks the GSA shape and results a small 'bump' in the profile that is pushed to the right during a tour. On average we measured at the SKR events a log-increase of 0.048 on the value of $\|\tilde{\mathbf{b}}_\kappa\|$ compared to the GSA (while the rest of the basis matches very closely). Although anecdotal, adjusting the estimator with this offset of 0.048 resulted in very close predictions for the SKR events.

- For small blocksizes $\beta \leq 30$ we see that the DSD-PT estimate is slightly pessimistic compared to the experiments. However the successful profile slope α_β (computed from the profile at the moment of detection) does closely match the predicted slope $\alpha_{\beta_{pred}}$, pointing to a wrong calibration of the slope parameter for very low blocksizes. Note that the non-flat part of the Z-shape in the experiments has size less than the $2 \cdot 127$ dimensional lattice used for calibration, which plausibly explain why the slope converges quicker than expected.

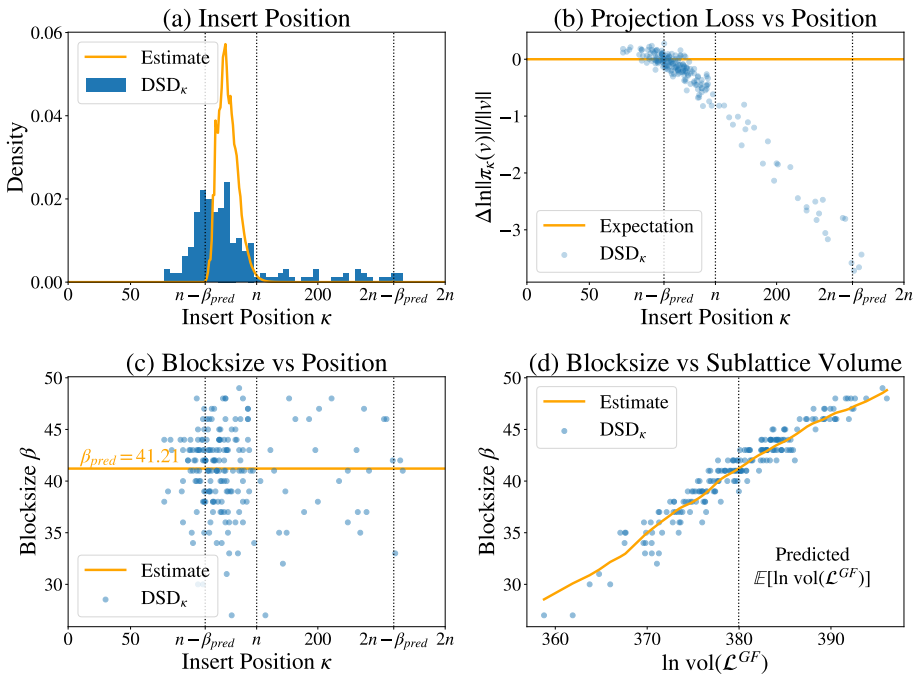

Figure 7.11: Results from running progressive BKZ with 8 tours on 200 circulant NTRU instances in the overstretched regime with parameters $n = 151, q = 2003$ and $\sigma^2 = \frac{2}{3}$. (a) Distribution of κ at which the DSD$_\kappa$ event is triggered. (b) The log-length decrease from $\|\mathbf{v}\|$ to $\|\pi_\kappa(\mathbf{v})\|$ normalised by the expected decrease (Lemma 112). (c) Successful blocksize β versus the position κ, and (d) versus the volume $\mathrm{vol}(\mathcal{L}^{\mathbf{GF}})$.

7.5.2 Detailed behaviour

Because the estimator computes actual probabilities we can compare the predictions to the experiments on a deeper level. We ran 200 experiments on circulant NTRU instances with parameters $n = 151, q = 2003$ and $\sigma^2 = \frac{2}{3}$. These parameters fall into the overstretched regime. The results, compared to the estimator, are shown in Figure 7.11. The average successful blocksize over all 200 instances is $\beta_{\mathrm{avg}} = 40.99$, which is close to the estimated value of $\beta_{\mathrm{pred}} = 41.21$. We note however that the average squared error is as large as 16.9; the successful blocksizes range from 27 to 49 for identical initial parameters. Looking at Figure 7.11(d) we see that this can mostly be

explained by the large variance of the volume of the dense sublattice $\mathrm{vol}(\mathcal{L}^{\mathbf{GF}})$, as earlier noticed in Figure 7.9. Adjusting the estimate with the concrete volume (instead of the average case prediction) decreases the average squared error all the way down to 0.83. E.g., the large average squared error is not an imprecision of the estimator, but an inherent high variance in the hardness of circulant NTRU instances. In many schemes this variance is significantly reduced by sampling the secret keys $\mathbf{f}, \mathbf{g}$ under some fixed constraints on the lengths $\|\mathbf{f}\|, \|\mathbf{g}\|$ or even stronger restrictions.

We now take a look at the positions κ at which the DSD_κ events took place. In Figure 7.11(a) we see that most events (169/200) happen for $\kappa \leq n$, as predicted by the DSD-PT estimate. We see in Figure 7.11(b) that for these instances the projection $\pi_\kappa(\mathbf{v})$ of the dense sublattice vector $\mathbf{v}$ is also not much shorter than expected, which matches the DSD-PT event as defined in Section 7.3.1. For the remaining events (31/200) with $\kappa > n$ we note that κ seems to be evenly distributed over $n+1, \ldots, 2n-\beta$, while the projected length becomes increasingly smaller than expected for increasing κ.

These DSD-LL events do not seem properly predicted by the model. Indeed, looking at κ around $2n-\beta$ the model predicts that the probability that a random vector has such a short projection is as small as 10^{-55}, and the probability that such a vector would correctly be lifted by Babai's nearest plane algorithm is even smaller (see script `lucky_lift.sage`). Interestingly however is that the transition between DSD-PT and DSD-LL events is rather continuous, and it is not so clear where to put the threshold between these two events.

In Figure 7.11(c) we see no clear difference between the successful blocksizes at positions $\kappa \leq n$ versus those at $\kappa > n$, and their average of 41.1 and 40.4 respectively. This suggests that, at least for these parameters, the DSD-LL events do not significantly contribute to making the attack cheaper. This was also the case for all other experiments we ran with different parameters.

In Figure 7.11(a) we also show the predicted distribution of the event positions κ. We correctly predicted the peak around $\kappa \approx n - \frac{1}{2}\beta_{\mathrm{pred}}$, and that $\kappa \leq n$ for the DSD-PT events. However the prediction is much more concentrated than the experimentally observed events, and neglects to notice the events at $\kappa \leq n - \frac{1}{2}\beta_{\mathrm{pred}}$. We give two possible explanations for this.

- The estimate assumes an average-case dense sublattice volume, ignoring the high variance of the dense sublattice volume. Adjusting for this using the real volumes makes the prediction less concentrated and closer to the observed events. However this still not captures all events at $\kappa \leq n - \frac{1}{2}\beta_{\text{pred}}$.
- The experiments and the estimator inherently measure two different things. The experiment detects a DSD_κ event if the projection $\pi_\kappa(\mathbf{v})$ is inserted and is lifted to a dense sublattice vector $\mathbf{v}$ by Babai's nearest plane algorithm. The estimator however estimates the probability that the projection is short enough and that the projection is *eventually* lifted to the dense sublattice vector $\mathbf{v}$ by later tours of the BKZ algorithm. Therefore a DSD-PT event predicted at position κ might experimentally only be detected one tour later at $\kappa - \beta + 1$ or at an even lower position. This could potentially be resolved by also taking into account the direct lift of Babai's nearest plane in the estimator, but note that for the eventual goal of estimating the successful blocksize this will not matter.

7.5.3 Fatigue point

The concrete estimator follows the experiments reasonably well and thus we can use it to estimate the concrete fatigue point for dimensions that are not feasible in practice. To verify the estimate of the fatigue point we also did some experiments in dimensions that are still feasible. For this we ran a *soft* binary search, only decreasing the interval length by 3/4 so as not view a probabilistic result as a definitive answer. More specifically, starting with a range of $[q_{\min}, q_{\max}]$ we ran an experiment for a prime $q \approx (q_{\min} + q_{\max})/2$. If it succeeds with an SKR event we update $q_{\min}$ to $(q_{\min} + q)/2 + 1$, if it succeeds with a DSD event we update $q_{\max}$ to $(q_{\max} + q)/2 - 1$. We repeat this until the interval does not contain any prime and we return $(q_{\min} + q_{\max})/2$ as a rough estimate of the fatigue point. We averaged this over 20 experiments for each parameter n. We chose for matrix NTRU because of the lower variance in the hardness of these instances.

We compared this to the prediction. Because the estimator accounts for probabilities of events, we can predict for which value of q about 50% of the instances succeeds with a DSD event. Because it

would be unreasonable to calibrate the low blocksize slope values α_β for each dimension we reused those of the $2 \cdot 127$ dimensional q-ary lattice from an earlier experiment. This might make the estimates a bit less precise for $n \ll 127$, and $n \gg 127$ if the successful blocksize is small around the fatigue point.

The results are shown in Figure 7.12 and plotted against $Cn^{2.484}$ for several constants C. Remarkably the experiments and concrete predictions closely follow the asymptotics already for reasonably small values of n. A loglog-linear regression of the 50% DSD-PT estimate over all primes $199, \ldots, 499$ gives $0.0034 \cdot n^{2.506}$. Restricting the exponent to 2.484 gives $0.0038 \cdot n^{2.484}$ with a log-standard deviation of only 0.006.

The experimental average appears slightly higher than the estimator prediction for 50% DSD - 50% SKR. The main reason for this seems to be that the estimator is slightly pessimistic for detecting the SKR event, as already observed and explained in Section 7.5.1. Another small detail is that the binary search is slightly biased to higher values of q because at each iteration we pick the *next* prime after

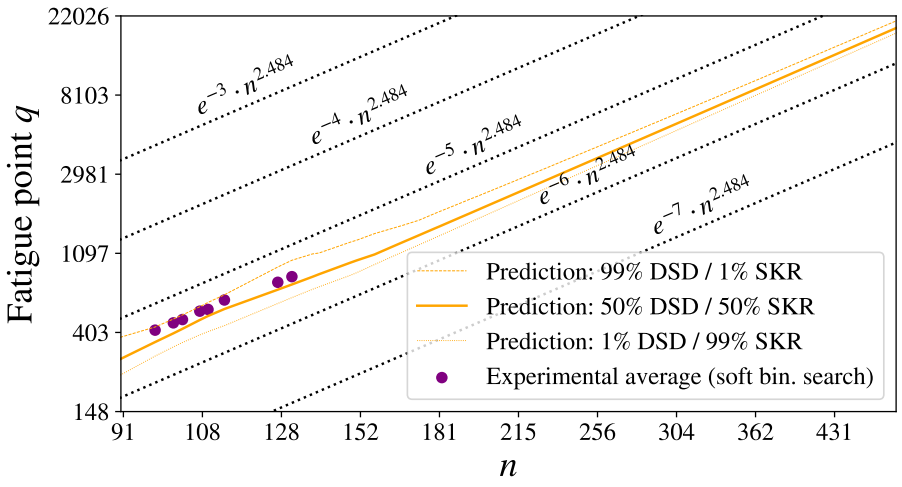

Figure 7.12: Concrete fatigue point versus asymptotics using progressive BKZ with 8 tours on matrix NTRU instances with variance $\sigma^2 = \frac{2}{3}$. The 0.5 percentile line shows for which q we estimate that the DSD event is triggered before the SKR event for about 50% of the instances.

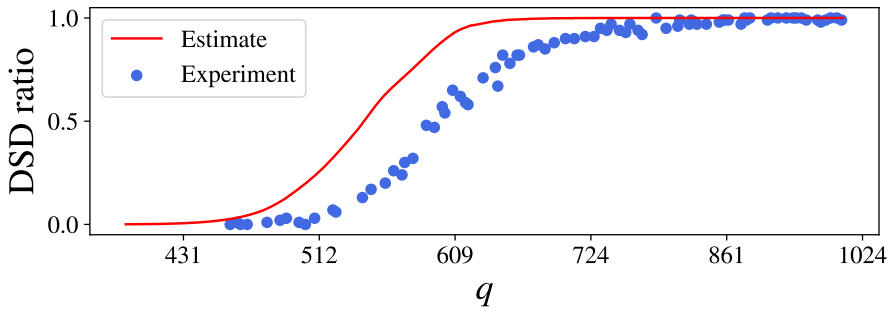

Figure 7.13: Experiment versus prediction for progressive BKZ with 8 tours on matrix NTRU instances with parameters $n = 113, \sigma^2 = \frac{2}{3}$ for several moduli q. We did 100 runs per modulus q and the plot shows the ratio of these runs succeeding with a DSD event (before an SKR event).

$(q_{\min} + q_{\max})/2$.

7.5.4 Zoom on the fatigue point: a smooth probabilistic transition

We take a closer look at the transition from the non-overstretched to the overstretched regime. For this we ran several experiments on matrix NTRU instances with parameters $n = 113, \sigma^2 = \frac{2}{3}$ for several moduli q, with 100 runs each. We compare the DSD success ratio with the probabilistic concrete estimate. The results are shown in Figure 7.13. Just as in Figure 7.12 we see a shift between the experiment and prediction, which can again be explained by the SKR estimator being too pessimistic. Note however that while the discrepancy looks significant in this zoomed plot, it only emphasises a small error of about 2 block sizes between the experiments and the predictions. Ignoring this shift the shape of the predicted transition matches the experiments very well.