LWE Without Modular Reduction and Improved Side-Channel Attacks Against BLISS

Jonathan Bootle[1], Claire Delaplace[2,3], Thomas Espitau[4], Pierre-Alain Fouque[2], and Mehdi Tibouchi[5]

[1] University College London
jonathan.bootle.14@ucl.ac.uk
[2] Univ Rennes
{claire.delaplace,pierre-alain.fouque}@univ-rennes1.fr
[3] Univ Lille
[4] Sorbonne Université
thomas.espitau@lip6.fr
[5] NTT Secure Platform Laboratories
tibouchi.mehdi@lab.ntt.co.jp

Abstract. This paper is devoted to analyzing the variant of Regev's learning with errors (LWE) problem in which modular reduction is omitted: namely, the problem (ILWE) of recovering a vector $\mathbf{s} \in \mathbb{Z}^n$ given polynomially many samples of the form $(\mathbf{a}, \langle \mathbf{a}, \mathbf{s} \rangle + e) \in \mathbb{Z}^{n+1}$ where $\mathbf{a}$ and e follow fixed distributions. Unsurprisingly, this problem is much easier than LWE: under mild conditions on the distributions, we show that the problem can be solved efficiently as long as the variance of e is not superpolynomially larger than that of $\mathbf{a}$. We also provide almost tight bounds on the number of samples needed to recover $\mathbf{s}$.

Our interest in studying this problem stems from the side-channel attack against the BLISS lattice-based signature scheme described by Espitau et al. at CCS 2017. The attack targets a *quadratic* function of the secret that leaks in the rejection sampling step of BLISS. The same part of the algorithm also suffers from a *linear* leakage, but the authors claimed that this leakage could not be exploited due to signature compression: the linear system arising from it turns out to be *noisy*, and hence key recovery amounts to solving a high-dimensional problem analogous to LWE, which seemed infeasible. However, this noisy linear algebra problem does not involve any modular reduction: it is essentially an instance of ILWE, and can therefore be solved efficiently using our techniques. This allows us to obtain an improved side-channel attack on BLISS, which applies to 100% of secret keys (as opposed to $\approx 7\%$ in the CCS paper), and is also considerably faster.

1 Introduction

Learning with errors. Regev's *learning with errors* problem (LWE) is the problem of recovering a uniformly random vector $\mathbf{s} \in (\mathbb{Z}/q\mathbb{Z})^n$ given polynomially many samples of the form $(\mathbf{a}, \langle \mathbf{a}, \mathbf{s} \rangle + e \bmod q)$, with $\mathbf{a}$ uniform in $(\mathbb{Z}/q\mathbb{Z})^n$,

and e sampled according to a fixed distribution over $\mathbb{Z}/q\mathbb{Z}$ (typically a discrete Gaussian). Regev showed [43] that for suitable parameters, this problem is as hard as worst-case lattice problems, and is polynomial-time equivalent to its decision version, which asks to distinguish the distribution of tuples $(\mathbf{a}, \langle \mathbf{a}, \mathbf{s} \rangle + e \bmod q)$ as above from the uniform distribution over $(\mathbb{Z}/q\mathbb{Z})^{n+1}$. These results are a cornerstone of modern lattice-based cryptography, which is to a large extent based on LWE and related problems.

Many variants of the LWE problem have been introduced in the literature, mostly with the goal of improving the efficiency of lattice-based cryptography. For example, papers have been devoted to the analysis of LWE when the error e has a non-Gaussian distribution and/or is very small [6, 38, 16], when the secret $\mathbf{s}$ is sampled from a non-uniform distribution [5, 12, 3, 7, 2], or when the vectors $\mathbf{a}$ are non-uniform [20, 23]. A long line of research has considered variants of LWE in which auxiliary information is provided about the secret $\mathbf{s}$ [21, 15, 12, 31]. Extensions of LWE over more general rings have also been extensively studied, starting from the introduction of the Ring-LWE problem [36, 46, 37, 29]. Yet another notable variant of LWE is the learning with rounding (LWR) problem [8, 4, 9], in which the scalar product $\langle \mathbf{a}, \mathbf{s} \rangle$ is partly hidden not by adding some noise e, but by disclosing only its most significant bits.

Recently, further exotic variants have emerged in association with schemes submitted to the NIST postquantum cryptography standardization process. One can mention for example Compact-LWE [33, 34], which has been broken [11, 48, 30]; learning with truncation, considered in pqNTRUSign [24]; and Mersenne variants of Ring-LWE, introduced for ThreeBears [22] and Mersenne–756839 [1].

The ILWE problem. In this paper, we introduce a simpler variant of LWE in which computations are carried out over $\mathbb{Z}$ rather than $\mathbb{Z}/q\mathbb{Z}$, i.e. without modular reduction. More precisely, we consider the problem which we call ILWE ("integer LWE") of finding a vector $\mathbf{s} \in \mathbb{Z}^n$ given polynomially many samples of the form $(\mathbf{a}, \langle \mathbf{a}, \mathbf{s} \rangle + e) \in \mathbb{Z}^{n+1}$, where $\mathbf{a}$ and e follow fixed distributions on $\mathbb{Z}$.

This problem may occur more naturally in statistical learning theory or numerical analysis than it does in cryptography; indeed, contrary to LWE, it is usually not hard. It can even be solved efficiently when the error e is *much larger* than the inner product $\langle \mathbf{a}, \mathbf{s} \rangle$ (but not superpolynomially larger), under relatively mild conditions on the distributions involved.

The fact that standard learning techniques like least squares regression should apply to this problem can be regarded as folklore, and is occasionally mentioned in special cases in the cryptographic literature (see e.g. [20, §7.6]). The main purpose of this work is to give a completely rigorous treatment of this question, and in particular to analyze the number of samples needed to solve ILWE both in an information-theoretic sense and using concrete algorithms.

ILWE and side-channel attacks on BLISS. Our main motivation for studying the ILWE problem is a side-channel attack against the BLISS lattice-based signature scheme described by Espitau et al. at CCS 2017 [19].

BLISS [17] is one of the most prominent, efficient and widely implemented lattice-based signature schemes, and it has received significant attention in terms of side-channel analysis. Several papers [13, 40, 19] have pointed out that, in available implementations, certain parts of the signing algorithm can leak sensitive information about the secret key via various side-channels like cache timing, electromagnetic emanations and secret-dependent branches. They have shown that this leakage can be exploited for key recovery.

We are in particular interested in the leakage that occurs in the rejection sampling step of BLISS signature generation. Rejection sampling is an essential element of the construction of BLISS and other lattice-based signatures following Lyubashevsky's "Fiat–Shamir with aborts" framework [35]. Implementing it efficiently in a scheme using Gaussian distributions, as is the case for BLISS, is not an easy task, however, and as observed by Espitau et al., the optimization used in BLISS turns out to leak two functions of the secret key via side-channels: an *exact*, *quadratic* function, as well as a *noisy*, *linear* function.

The attack proposed by Espitau et al. relies only on the quadratic leakage, and as a result uses very complex and computationally costly techniques from algorithmic number theory (a generalization of the Howgrave-Graham–Szydlo algorithm for solving norm equations). In particular, not only does the main, polynomial-time part of their algorithm takes over a CPU month for standard BLISS parameters, technical reasons related to the hardness of factoring make their attack only applicable to a small fraction of BLISS secret key (around 7%; these are keys satisfying a certain smoothness condition). They note that using the *linear* leakage instead would be much simpler if the linear function was exactly known, but cannot be done due to its noisy nature: recovering the key then become a high-dimensional noisy linear algebra problem analogous to LWE, which should therefore be hard.

However, the authors missed an important difference between that linear algebra problem and LWE: the absence of modular reduction. The problem can essentially be seen as an instance of ILWE instead, and our analysis thus shows that it is easy to solve. This results in a much more computationally efficient attack taking advantage of the leakage in BLISS rejection sampling, which moreover applies to *all* secret keys.

Our contributions. We propose a detailed theoretical analysis of the ILWE problem and show how it can be applied to the side-channel attack on BLISS. We also provide numerical simulations showing that our proposed algorithms behave in a way consistent with the theoretical predictions.

On the theoretical side, our first contribution is to prove that, in an information-theoretic sense, solving the ILWE problem requires at least $m = \Omega\big((\sigma_e/\sigma_a)\big)^2$ samples from the ILWE distribution when the error e has standard deviation σ_e, and the coefficients of the vectors $\mathbf{a}$ in samples have standard deviation σ_a. We show this by estimating the statistical distance between the distributions arising from two distinct secret vectors $\mathbf{s}$ and $\mathbf{s}'$. In particular, the ILWE problem is

hard when σ_e is superpolynomially larger than σ_a, but can be easy otherwise, including when σ_e exceeds σ_a by a large polynomial factor.

We then provide and analyze concrete algorithms for solving the problem in that case. Our main focus is least squares regression followed by rounding. Roughly speaking, we show that this approach solves the ILWE problem with m samples when $m \geq C \cdot (\sigma_e/\sigma_a)^2 \log n$ for some constant C (and is also a constant factor larger than n, to ensure that the noise-free version of the corresponding linear algebra problem has a unique solution, and that the covariance matrix of the vectors $\mathbf{a}$ is well-controlled). Our result applies to a very large class of distributions for $\mathbf{a}$ and e including bounded distributions and discrete Gaussians. It relies on subgaussian concentration inequalities.

Interestingly, ILWE can be interpreted as a bounded distance decoding problem in a certain lattice in $\mathbb{Z}^n$ (which is very far from random), and the least squares approach coincides with Babai's rounding algorithm for the approximate closest vector problem (CVP) when seen through that lens. As a side contribution, we also show that even with a much stronger CVP algorithm (including an exact CVP oracle), one cannot improve the number of samples necessary to recover $\mathbf{s}$ by more than a constant factor. And on another side note, we also consider alternate algorithms to least squares when very few samples are available (so that the underlying linear algebra system is not even full-rank), but the secret vector is known to be sparse. In that case, compressed sensing techniques using linear programming [14] can solve the problem efficiently.

After this theoretical analysis, we concretely examine the noisy linear algebra problem arising from the linear part of the BLISS rejection sampling leakage, and show that is strongly resembles an ILWE problem, which allows us to estimate the number of side-channel traces needed to recover the secret key.

Simulation results both for the vanilla ILWE problem and the BLISS attack are consistent with the theoretical predictions (only with better constants). In particular, we obtain a much more efficient attack on BLISS than the one in [19], which moreover applies to 100% of possible secret keys. The only drawback is that our attack requires a larger number of traces (around 20000 compared to 512 in [19] for BLISS–I parameters), and even that is to a large extent counterbalanced by the fact that we can easily handle errors in the values read off from side-channel traces, whereas Espitau et al. need all their leakage values to be exact.

2 Preliminaries

2.1 Notation

For $r \in \mathbb{R}$, we denote by $\lceil r \rfloor$ the nearest integer to r (rounding down for half-integers), and by $\lfloor r \rfloor$ the largest integer less or equal to r. For a vector $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, the p-norm $\|\mathbf{x}\|_p$ of $\mathbf{x}$, $p \in [1, \infty)$, is given by $\|\mathbf{x}\|_p = \left(|x_1|^p + \cdots + |x_n|^p\right)^{1/p}$, and the infinity norm by $\|\mathbf{x}\|_\infty = \max\left(|x_1|, \ldots, |x_n|\right)$. For a matrix $A \in \mathbb{R}^{m \times n}$, the operator norm $\|A\|_p^{\mathrm{op}}$ of A with respect to the p-norm,

$p \in [1, \infty]$, is given by:

$$\|A\|_p^{\mathrm{op}} = \sup_{\mathbf{x} \in \mathbb{R}^n \setminus \{0\}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p} = \sup_{\|\mathbf{x}\|_p = 1} \|A\mathbf{x}\|_p.$$

For any random variable X, we denote by $\mathbb{E}[X]$ its expectation and by $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ its variance. We write $X \sim \chi$ to denote that X follows the distribution χ. If χ is a discrete distribution over some set S, then for any $s \in S$, we denote by $\chi(s)$ the probability that a sample from χ is equal to s. In particular, if $f \colon S \to \mathbb{R}$ is any function and $X \sim \chi$, we have:

$$\mathbb{E}[f(s)] = \sum_{s \in S} f(s) \cdot \chi(s).$$

Similarly, the statistical distance $\Delta(\chi, \chi')$ of two distributions χ, χ' over the set S is:

$$\Delta(\chi, \chi') = \frac{1}{2} \sum_{s \in S} \big|\chi(s) - \chi(s')\big|.$$

Let $\rho(x) = \exp(-\pi x^2)$ for all $x \in \mathbb{R}$. We define $\rho_{c,\sigma}(x) = \rho\big((x-c)/\sigma\big)$ the Gaussian function of parameters c, σ. For any subset $S \subset \mathbb{R}$ such that the sum converges, we let:

$$\rho_{c,\sigma}(S) = \sum_{s \in S} \rho_{c,\sigma}(s).$$

The discrete Gaussian distribution $D_{c,\sigma}$ centered at c and of parameter σ is the distribution on $\mathbb{Z}$ defined by

$$D_{c,\sigma}(x) = \frac{\rho_{c,\sigma}(x)}{\rho_{c,\sigma}(\mathbb{Z})} = \frac{\exp\big(-\pi(x-c)^2/\sigma^2\big)}{\rho_{c,\sigma}(\mathbb{Z})}$$

for all $x \in \mathbb{Z}$. We omit the subscript c in $\rho_{c,\sigma}$ and $D_{c,\sigma}$ when $c = 0$.

2.2 LWE over the Integers

It is possible to define a variant of the LWE problem "over the integers", i.e. without modular reduction. We call this problem ILWE("integer-LWE"), and define it as follows. The problem arising from the scalar product leakage in the BLISS rejection sampling is essentially of that form.

Definition 2.1 (ILWE Distribution). *For any vector $\mathbf{s} \in \mathbb{Z}^n$ and any two probability distributions χ_a, χ_e over $\mathbb{Z}$, the* ILWE distribution *$\mathscr{D}_{\mathbf{s},\chi_a,\chi_e}$ associated with those parameters (which we will simply denote $\mathscr{D}_{\mathbf{s}}$ for short when χ_a, χ_e are clear) is the probability distribution over $\mathbb{Z}^n \times \mathbb{Z}$ defined as follows: samples from $\mathscr{D}_{\mathbf{s},\chi_a,\chi_e}$ are of the form*

$$(\mathbf{a}, b) = \big(\mathbf{a}, \langle \mathbf{a}, \mathbf{s} \rangle + e\big) \quad \textit{with} \quad \mathbf{a} \leftarrow \chi_a^n \textit{ and } e \leftarrow \chi_e.$$

Definition 2.2 (ILWE Problem). *The* ILWE problem *is the computational problem parametrized by n, m, χ_a, χ_e in which, given m samples $\{(\mathbf{a}_i, b_i)\}_{1 \le i \le m}$ from a distribution of the form $\mathscr{D}_{\mathbf{s},\chi_a,\chi_e}$ for some $\mathbf{s} \in \mathbb{Z}^n$, one is asked to recover the vector $\mathbf{s}$.*

2.3 Subgaussian Probability Distributions

In this paper, the distributions χ_a, χ_e we will consider will usually be of mean zero and rapidly decreasing. More precisely, we will assume that those distributions are *subgaussian*. The notion of a subgaussian distribution was introduced by Kahane in [27], and can be defined as follows.

Definition 2.3. *A random variable X over $\mathbb{R}$ is said to be* τ-subgaussian *for some $\tau > 0$ if the following bound holds for all $s \in \mathbb{R}$:*

$$\mathbb{E}\big[\exp(sX)\big] \leq \exp\Big(\frac{\tau^2 s^2}{2}\Big). \tag{2.1}$$

A τ-subgaussian probability distribution is defined in the same way.

This section collects useful facts about subgaussian random variables; most of them are well-known, and presented mostly in the interest of a self-contained and consistent presentation (as definitions of related notions tend to vary slightly from one reference to the next).

For a subgaussian random variable X, there is a minimal τ such that X is τ-subgaussian. This τ is sometimes called the *subgaussian moment* of the random variable (or of its distribution).

As expressed in the next lemma, subgaussian distributions always have mean zero, and their variance is bounded by τ^2.

Lemma 2.4. *A τ-subgaussian random variable X satisfies:*

$$\mathbb{E}[X] = 0 \quad \textit{and} \quad \mathbb{E}[X^2] \leq \tau^2.$$

Proof. For s around zero, we have:

$$\mathbb{E}[\exp(sX)] = 1 + s\mathbb{E}[X] + \frac{s^2}{2}\mathbb{E}[X^2] + o(s^2).$$

Since, on the other hand, $\exp(s^2\tau^2/2) = 1 + \frac{s^2}{2}\tau^2 + o(s^2)$, the result follows immediately from (2.1). □

Many usual distributions over $\mathbb{Z}$ or $\mathbb{R}$ are subgaussian. This is in particular the case for Gaussian and discrete Gaussian distributions, as well as all *bounded* probability distributions with mean zero.

Lemma 2.5. *The following distributions are subgaussian.*

(*i*) *The centered normal distribution $\mathcal{N}(0, \sigma^2)$ is σ-subgaussian.*
(*ii*) *The centered discrete Gaussian distribution D_σ of parameter σ is $\frac{\sigma}{\sqrt{2\pi}}$-subgaussian for all $\sigma \geq 0.283$.*
(*iii*) *The uniform distribution $\mathscr{U}_\alpha$ over the integer interval $[-\alpha, \alpha] \cap \mathbb{Z}$ is $\frac{\alpha}{\sqrt{2}}$-subgaussian for $\alpha \geq 3$.*
(*iv*) *More generally, any distribution over $\mathbb{R}$ of mean zero and supported over a bounded interval $[a, b]$ is $\big(\frac{b-a}{2}\big)$-subgaussian.*

Moreover, in the cases (i)–(iii) above, the quotient $\tau \geq 1$ between the subgaussian moment and the standard deviation satisfies:

(i) $\tau = 1$;
(ii) $\tau < \sqrt{2}$ *assuming* $\sigma \geq 1.85$;
(iii) $\tau \leq \sqrt{3/2}$

respectively.

Proof. See the full version of this paper [10]. □

The main property of subgaussian distributions is that they satisfy a very strong tail bound.

Lemma 2.6. *Let X be a τ-subgaussian distribution. For all $t > 0$, we have*

$$\Pr[X > t] \leq \exp\Big(-\frac{t^2}{2\tau^2}\Big). \tag{2.2}$$

Proof. Fix $t > 0$. For all $s \in \mathbb{R}$ we have, by Markov's inequality:

$$\Pr[X > t] = \Pr[\exp(sX) > e^{st}] \leq \frac{\mathbb{E}[\exp(sX)]}{e^{st}}$$

since the exponential is positive. Using the fact that X is τ-subgaussian, we get:

$$\Pr[X > t] \leq \exp\Big(\frac{s^2\tau^2}{2} - st\Big)$$

and the right-hand side is minimal for $s = t/\tau^2$, which exactly gives (2.2). □

The following result states that a linear combination of *independent* subgaussian random variables is again subgaussian.

Lemma 2.7. *Let $X_1, \ldots, X_n$ be independent random variables such that X_i is τ_i-subgaussian. For all $\mu_1, \ldots, \mu_n \in \mathbb{R}$, the random variable $X = \mu_1 X_1 + \cdots + \mu_n X_n$ is τ-subgaussian with:*

$$\tau^2 = \mu_1^2\tau_1^2 + \cdots + \mu_n^2\tau_n^2.$$

Proof. Since the X_i's are independent, we have, for all $s \in \mathbb{R}$:

$$\begin{aligned}\mathbb{E}[\exp(sX)] &= \mathbb{E}\Big[\exp\big(s(\mu_1 X_1 + \cdots + \mu_n X_n)\big)\Big] \\ &= \mathbb{E}\Big[\exp(\mu_1 s X_1)\cdots\exp(\mu_n s X_n)\Big] = \prod_{i=1}^{n}\mathbb{E}\big[\exp(\mu_i s X_i)\big].\end{aligned}$$

Now, since X_i is τ_i-subgaussian, we have

$$\mathbb{E}\big[\exp(\mu_i s X_i)\big] \leq \exp\Big(\frac{s^2(\mu_i\tau_i)^2}{2}\Big)$$

for all i. Therefore:

$$\mathbb{E}[\exp(sX)] \leq \prod_{i=1}^{n} \exp\left(\frac{s^2(\mu_i\tau_i)^2}{2}\right) = \exp\left(\frac{s^2\tau^2}{2}\right)$$

with $\tau^2 = \mu_1^2\tau_1^2 + \cdots + \mu_n^2\tau_n^2$ as required. □

The previous result shows that the notion of a subgaussian random variable has a natural extension to higher dimensions.

Definition 2.8. *A random vector* $\mathbf{x}$ *in* $\mathbb{R}^n$ *is called a* τ-subgaussian random vector *if for all vectors* $\mathbf{u} \in \mathbb{R}^n$ *with* $\|\mathbf{u}\|_2 = 1$*, the inner product* $\langle \mathbf{u}, \mathbf{x} \rangle$ *is a* τ*-subgaussian random variable.*

It clearly follows from Lemma 2.7 that if $X_1, \ldots, X_n$ are independent τ-subgaussian random variables, then the random vector $\mathbf{x} = (X_1, \ldots, X_n)$ is τ-subgaussian. In particular, if χ is a τ-subgaussian distribution, then a random vector $\mathbf{x} \sim \chi^n$ is τ-subgaussian. A nice feature of subgaussian random vectors is that the image of such a random vector under any linear transformation is again subgaussian.

Lemma 2.9. *Let* $\mathbf{x}$ *be a* τ*-subgaussian random vector in* $\mathbb{R}^n$*, and* $A \in \mathbb{R}^{m\times n}$*. Then the random vector* $\mathbf{y} = A\mathbf{x}$ *is* τ'*-subgaussian, with* $\tau' = \|A^T\|_2^{\mathrm{op}} \cdot \tau$*.*

Proof. Fix a unit vector $\mathbf{u}_0 \in \mathbb{R}^m$. We want to show that the random variable $\langle \mathbf{u}_0, \mathbf{y} \rangle$ is τ'-subgaussian. To do so, first observe that:

$$\langle \mathbf{u}_0, \mathbf{y} \rangle = \langle A^T\mathbf{u}_0, \mathbf{x} \rangle = \mu \langle \mathbf{u}, \mathbf{x} \rangle$$

where $\mu = \|A^T\mathbf{u}_0\|_2$, and $\mathbf{u} = \frac{1}{\mu}A^T\mathbf{u}_0$ is a unit vector of $\mathbb{R}^n$. Since $\mathbf{x}$ is τ-subgaussian, we know that the inner product $\langle \mathbf{u}, \mathbf{x} \rangle$ is a τ-subgaussian random variable. As a result, by Lemma 2.7 in the trivial case of a single variable, we obtain that $\langle \mathbf{u}_0, \mathbf{y} \rangle = \mu\langle \mathbf{u}, \mathbf{x} \rangle$ is $\big(|\mu|\tau\big)$-subgaussian. But by definition of the operator norm, $|\mu| \leq \|A^T\|_2^{\mathrm{op}}$, and the result follows. □

3 Information-Theoretic Analysis

A first natural question one can ask regarding the ILWE problem is how hard it is in an information-theoretic sense. In other words, given two vectors $\mathbf{s}, \mathbf{s}' \in \mathbb{Z}^n$, how close are the ILWE distributions $\mathscr{D}_{\mathbf{s}}, \mathscr{D}_{\mathbf{s}'}$ associated to $\mathbf{s}$ and $\mathbf{s}'$, or equivalently, how many samples do we need to distinguish between those distributions?

In this section, we show that, at least when the error distribution χ_e is either uniform or Gaussian, the statistical distance between $\mathscr{D}_{\mathbf{s}}$ and $\mathscr{D}_{\mathbf{s}'}$ admits a bound of the form $O\big(\frac{\sigma_a}{\sigma_e}\|\mathbf{s} - \mathbf{s}'\|\big)$. In particular, distinguishing between those distributions with constant success probability requires

$$\Omega\left(\frac{1}{\|\mathbf{s} - \mathbf{s}'\|^2}\left(\frac{\sigma_e}{\sigma_a}\right)^2\right)$$

samples, and the distributions are statistically indistinguishable when σ_e is superpolynomially larger than σ_a. To see this, we first give a relatively simple expression for the statistical distance.

Lemma 3.1. *The statistical distance between $\mathscr{D}_{\mathbf{s}}$ and $\mathscr{D}_{\mathbf{s}'}$ is given by:*

$$\Delta(\mathscr{D}_{\mathbf{s}}, \mathscr{D}_{\mathbf{s}'}) = \mathbb{E}\big[\Delta(\chi_e, \chi_e - \langle \mathbf{a}, \mathbf{s} - \mathbf{s}' \rangle)\big],$$

where $\chi_e + t$ denotes the translation of χ_e by the constant t, and the expectation is taken over $\mathbf{a} \leftarrow \chi_a^n$.

Proof. By definition of the statistical distance, we have:

$$\Delta(\mathscr{D}_{\mathbf{s}}, \mathscr{D}_{\mathbf{s}'}) = \frac{1}{2} \sum_{(\mathbf{a},b) \in \mathbb{Z}^{n+1}} \left| \Pr\left[(\mathbf{a}, b) \leftarrow \mathscr{D}_{\mathbf{s}}\right] - \Pr\left[(\mathbf{a}, b) \leftarrow \mathscr{D}_{\mathbf{s}'}\right] \right|.$$

Now to sample from $\mathscr{D}_{\mathbf{s}}$, one first samples $\mathbf{a}$ according to χ_a^n, independently sample e according to χ_e, and returns $(\mathbf{a}, b)$ with $b = \langle \mathbf{a}, \mathbf{s} \rangle + e$. Therefore:

$$\Pr\left[(\mathbf{a}, b) \leftarrow \mathscr{D}_{\mathbf{s}}\right] = \chi_a^n(\mathbf{a}) \cdot \chi_e(b - \langle \mathbf{a}, \mathbf{s} \rangle),$$

and similarly for $\mathscr{D}_{\mathbf{s}'}$. Thus, we can write:

$$\begin{aligned}
\Delta(\mathscr{D}_{\mathbf{s}}, \mathscr{D}_{\mathbf{s}'}) &= \frac{1}{2} \sum_{(\mathbf{a},b) \in \mathbb{Z}^{n+1}} \chi_a^n(\mathbf{a}) \cdot |\chi_e(b - \langle \mathbf{a}, \mathbf{s} \rangle) - \chi_e(b - \langle \mathbf{a}, \mathbf{s}' \rangle)| \\
&= \sum_{\mathbf{a} \in \mathbb{Z}^n} \chi_a^n(\mathbf{a}) \cdot \frac{1}{2} \sum_{b \in \mathbb{Z}} |\chi_e(b - \langle \mathbf{a}, \mathbf{s} \rangle) - \chi_e(b - \langle \mathbf{a}, \mathbf{s}' \rangle)| \\
&= \sum_{\mathbf{a} \in \mathbb{Z}^n} \chi_a^n(\mathbf{a}) \cdot \frac{1}{2} \sum_{x \in \mathbb{Z}} |\chi_e(x) - \chi_e(x + \langle \mathbf{a}, \mathbf{s} - \mathbf{s}' \rangle)|
\end{aligned}$$

where the last equality is obtained with the change of variables $x = b - \langle \mathbf{a}, \mathbf{s} \rangle$. We now observe that the expression

$$\frac{1}{2} \sum_{x \in \mathbb{Z}} |\chi_e(x) - \chi_e(x + \langle \mathbf{a}, \mathbf{s} - \mathbf{s}' \rangle)|$$

is exactly the statistical distance $\Delta(\chi_e, \chi_e - \langle \mathbf{a}, \mathbf{s} - \mathbf{s}' \rangle)$, and therefore we do obtain:

$$\Delta(\mathscr{D}_{\mathbf{s}}, \mathscr{D}_{\mathbf{s}'}) = \mathbb{E}\big[\Delta(\chi_e, \chi_e - \langle \mathbf{a}, \mathbf{s} - \mathbf{s}' \rangle)\big]$$

as required. □

Thus, we can bound the statistical distance $\Delta(\mathscr{D}_{\mathbf{s}}, \mathscr{D}_{\mathbf{s}'})$ using a bound on the statistical distance between χ_e and a translated distribution $\chi_e + t$. We provide such a bound when χ_e is either uniform in a centered integer interval, or a discrete Gaussian distribution.

Lemma 3.2. *Suppose that χ_e is either the uniform distribution $\mathscr{U}_\alpha$ in $[-\alpha, \alpha] \cap \mathbb{Z}$ for some positive integer α, or the centered discrete Gaussian distribution D_σ with parameter $\sigma \geq 1.60$. In either case, let $\sigma_e = \sqrt{\mathbb{E}[\chi_e^2]}$ be the standard deviation of χ_e. We then have the following bound for all $t \in \mathbb{Z}$:*

$$\Delta(\chi_e, \chi_e + t) \leq C \cdot |t| / \sigma_e$$

where $C = 1/\sqrt{12}$ in the uniform case and $C = 1/\sqrt{2}$ in the discrete Gaussian case.

Proof. See the full version of this paper [10]. □

Combining Lemma 3.1 and Lemma 3.2, we obtain a bound of the form announced at the beginning of this section.

Theorem 3.3. *Suppose that χ_e is as in the statement of Lemma 3.2. Then, for any two vectors $\mathbf{s}, \mathbf{s}' \in \mathbb{Z}^n$, the statistical distance between $\mathscr{D}_\mathbf{s}$ and $\mathscr{D}_{\mathbf{s}'}$ is bounded as:*

$$\Delta(\mathscr{D}_\mathbf{s}, \mathscr{D}_{\mathbf{s}'}) \le C \cdot \frac{\sigma_a}{\sigma_e} \|\mathbf{s} - \mathbf{s}'\|_2,$$

where C is the constant appearing in Lemma 3.2.

Proof. Lemma 3.1 gives:

$$\Delta(\mathscr{D}_\mathbf{s}, \mathscr{D}_{\mathbf{s}'}) = \mathbb{E}\big[\Delta(\chi_e, \chi_e - \langle \mathbf{a}, \mathbf{s} - \mathbf{s}' \rangle)\big],$$

and according to Lemma 3.2, the statistical distance on the right-hand side is bounded as:

$$\Delta(\chi_e, \chi_e + \langle \mathbf{a}, \mathbf{s} - \mathbf{s}' \rangle) \le \frac{C}{\sigma_e} \cdot \big|\langle \mathbf{a}, \mathbf{s} - \mathbf{s}' \rangle\big|.$$

It follows that:

$$\Delta(\mathscr{D}_\mathbf{s}, \mathscr{D}_{\mathbf{s}'}) \le \frac{C}{\sigma_e} \cdot \mathbb{E}\Big[\big|\langle \mathbf{a}, \mathbf{s} - \mathbf{s}' \rangle\big|\Big] \le \frac{C}{\sigma_e} \sqrt{\mathbb{E}\Big[\langle \mathbf{a}, \mathbf{s} - \mathbf{s}' \rangle^2\Big]}$$

where the second inequality is a consequence of the Cauchy–Schwarz inequality. Now, for any $\mathbf{u} \in \mathbb{Z}^n$, we can write:

$$\mathbb{E}\Big[\langle \mathbf{a}, \mathbf{u} \rangle^2\Big] = \mathbb{E}\Big[\sum_{1 \le i,j \le n} u_i u_j a_i a_j\Big] = \sum_{1 \le i,j \le n} u_i u_j \mathbb{E}[a_i a_j] = \sigma_a^2 \|\mathbf{u}\|_2^2$$

since $\mathbb{E}[a_i a_j] = \sigma_a^2 \delta_{ij}$. As a result:

$$\Delta(\mathscr{D}_\mathbf{s}, \mathscr{D}_{\mathbf{s}'}) \le C \cdot \frac{\sigma_a}{\sigma_e} \|\mathbf{s} - \mathbf{s}'\|_2$$

as required. □

As discussed in the beginning of this section, this shows that distinguishing between $\mathscr{D}_\mathbf{s}$ and $\mathscr{D}_{\mathbf{s}'}$ requires $\Omega\left(\frac{1}{\|\mathbf{s}-\mathbf{s}'\|^2}\left(\frac{\sigma_e}{\sigma_a}\right)^2\right)$ samples. In particular, recovering $\mathbf{s}$ (which implies distinguishing $\mathscr{D}_\mathbf{s}$ from all $\mathscr{D}_{\mathbf{s}'}$ for $\mathbf{s}' \ne \mathbf{s}$) requires

$$m = \Omega\big((\sigma_e/\sigma_a)^2\big) \tag{3.1}$$

samples. In what follows, we will describe efficient algorithms that actually recover $\mathbf{s}$ from only slightly more samples than this lower bound.

Remark 3.4. Contrary to the results of the next section, which will apply to arbitrary subgaussian distributions, we cannot establish an analogue of Lemma 3.2 using only a bound on the tail of the distribution χ_e. For example, if χ_e is supported over $2\mathbb{Z}$, then $\Delta(\chi_e, \chi_e + t) = 1$ for any odd t! One would presumably need an assumption of the small-scale regularity of χ_e to extend the result.

4 Solving the ILWE Problem

We now turn to describing efficient algorithms to solve the ILWE problem. We are given m samples $(\mathbf{a}_i, b_i)$ from the ILWE distribution $\mathscr{D}_{\mathbf{s}}$, and try to recover $\mathbf{s} \in \mathbb{Z}^n$. Since $\mathbf{s}$ can a priori be any vector, we, of course, need at least n samples to recover it; indeed, even without any noise, fewer samples can at best reveal an affine subspace on which $\mathbf{s}$ lies, but not its actual value. We are thus interested in the regime when $m \geq n$.

The equation for $\mathbf{s}$ can then be written in matrix form:

$$\mathbf{b} = A\mathbf{s} + \mathbf{e} \tag{4.1}$$

where $A \in \mathbb{Z}^{m \times n}$ is distributed according to $\chi_a^{m \times n}$, $\mathbf{e} \in \mathbb{Z}^m$ is distributed as χ_e^m, $A, \mathbf{b}$ are known and $\mathbf{e}$ is unknown.

The idea to find $\mathbf{s}$ will be to use simple statistical inference techniques to find an approximate solution $\tilde{\mathbf{s}} \in \mathbb{R}^n$ of the noisy linear system (4.1) and to simply round that solution coefficient by coefficient to get a candidate $\lceil \tilde{\mathbf{s}} \rfloor = (\lceil \tilde{s}_1 \rfloor, \ldots, \lceil \tilde{s}_n \rfloor)$ for $\mathbf{s}$. If we can establish the bound:

$$\|\mathbf{s} - \tilde{\mathbf{s}}\|_\infty < 1/2 \tag{4.2}$$

or, a fortiori, the stronger bound $\|\mathbf{s} - \tilde{\mathbf{s}}\|_2 < 1/2$, then it follows that $\lceil \tilde{\mathbf{s}} \rfloor = \mathbf{s}$ and the ILWE problem is solved.

The main technique we propose to use is least squares regression. Under the mild assumption that both χ_a and χ_e are subgaussian distributions, we will show that the corresponding $\tilde{\mathbf{s}}$ satisfies the bound (4.2) in the linear programming setting with high probability when m is sufficiently large. Moreover, the number m of samples necessary to establish those bounds, and hence solve ILWE, is only a $\log n$ factor larger than the information-theoretic minimum given in (3.1) (with the additional constraint that m should be a constant factor larger than n, to ensure that A is invertible and has well-controlled singular values).

We also briefly discuss lattice reduction as well as compressed sensing techniques based on linear programming. We show that even an exact-CVP oracle cannot significantly improve upon the $\log n$ factor of the least squares method. On the other hand, if the secret is known to be very sparse, compressed sensing techniques can recover the secret even in cases when $m < n$, where the least squares method is not applicable.

4.1 Least Squares Method

The first approach we consider to obtain an estimator $\tilde{\mathbf{s}}$ of $\mathbf{s}$ is the linear, unconstrained least squares method: $\tilde{\mathbf{s}}$ is chosen as a vector in $\mathbb{R}^n$ minimizing the squared Euclidean norm $\|\mathbf{b} - A\tilde{\mathbf{s}}\|_2^2$. In particular, the gradient vanishes at $\tilde{\mathbf{s}}$, which means that $\tilde{\mathbf{s}}$ is simply a solution to the linear system:

$$A^T A\tilde{\mathbf{s}} = A^T \mathbf{b}.$$

As a result, we can compute $\tilde{\mathbf{s}}$ in polynomial time (at most $O(mn^2)$) and it is uniquely defined if and only if A^TA is invertible.

It is intuitively clear that A^TA should be invertible when m is large. Indeed, one can write that matrix as:

$$A^TA = \sum_{i=1}^{m} \mathbf{a}_i \mathbf{a}_i^T$$

where the $\mathbf{a}_i$'s are the independent identically distributed rows of A, so the law of large numbers shows that $\frac{1}{m}A^TA$ converges almost surely to $\mathbb{E}\big[\mathbf{a}\mathbf{a}^T\big]$ as $m \to +\infty$, where $\mathbf{a}$ is a random variable in $\mathbb{Z}^n$ sampled from χ_a^n. We have:

$$\mathbb{E}\big[(\mathbf{a}\mathbf{a}^T)_{ij}\big] = \mathbb{E}[a_i a_j] = \delta_{ij}\sigma_a^2,$$

and therefore we expect A^TA to be close to $m\sigma_a^2 I_n$ for large m.

Making this heuristic argument rigorous is not entirely straightforward, however. Assuming some tail bounds on the distribution χ_a, concentration of measure results can be used to prove that, with high probability, the smallest eigenvalue $\lambda_{\min}(A^TA)$ is not much smaller than $m\sigma_a^2$ (and in particular A^TA is invertible) for m sufficiently large, with a concrete bound on m. This type of bound on the smallest eigenvalue is exactly what we will need in the rest of our analysis.

More precisely, when χ_a is bounded, one can apply a form of the so-called Matrix Chernoff inequality, such as [47, Cor. 5.2]. However, we would prefer a result that applies to e.g. discrete Gaussian distributions as well, so we only assume a subgaussian tail bound for χ_a. Such result can be derived from the following lemma due to Hsu et al. [26, Lemma 2] (for simplicity, we specialize their statement to $\epsilon_0 = 1/4$ and to the case of jointly independent vectors).

Lemma 4.1. *Let χ be a τ-subgaussian distribution of variance 1 over $\mathbb{R}$, and consider m random vectors $\mathbf{x}_1, \ldots, \mathbf{x}_m$ in $\mathbb{R}^n$ sampled independently according to χ^m. For any $\delta \in (0,1)$, we have:*

$$\Pr\left[\lambda_{\min}\Big(\frac{1}{m}\sum_{i=1}^{m}\mathbf{x}_i\mathbf{x}_i^T\Big) < 1 - \varepsilon(\delta,m) \text{ or } \lambda_{\max}\Big(\frac{1}{m}\sum_{i=1}^{m}\mathbf{x}_i\mathbf{x}_i^T\Big) > 1 + \varepsilon(\delta,m)\right] < \delta$$

where the error bound $\varepsilon(\delta,m)$ is given by:

$$\varepsilon(\delta,m) = 4\tau^2\left(\sqrt{\frac{8\log 9 \cdot n + 8\log(2/\delta)}{m}} + \frac{\log 9 \cdot n + \log(2/\delta)}{m}\right).$$

Using this lemma, one can indeed show that for χ_a subgaussian, $\lambda_{\min}(A^TA)$ is within an arbitrarily small factor of $m\sigma_a^2$ with probability $1 - 2^{-\eta}$ for $m = \Omega(n+\eta)$ (and similarly for $\lambda_{\max}$).

Theorem 4.2. *Suppose that χ_a is τ_a-subgaussian, and let $\tau = \tau_a/\sigma_a$. Let A be an $m \times n$ random matrix sampled from $\chi_a^{m\times n}$. There exist constants C_1, C_2 such that for all $\alpha \in (0,1)$ and $\eta \geq 1$, if $m \geq (C_1 n + C_2\eta)\cdot(\tau^4/\alpha^2)$ then*

$$\Pr\left[\lambda_{\min}\big(A^TA\big) < (1-\alpha)\cdot m\sigma_a^2 \text{ or } \lambda_{\max}\big(A^TA\big) > (1+\alpha)\cdot m\sigma_a^2\right] < 2^{-\eta}. \quad (4.3)$$

Furthermore, one can choose $C_1 = 2^8 \log 9$ *and* $C_2 = 2^9 \log 2$.

Proof. Let $\mathbf{a}_i$ be the i-th row of A, and $\mathbf{x}_i = \frac{1}{\sigma_a}\mathbf{a}_i$. Then the coefficients of $\mathbf{x}_i$ follow a τ-subgaussian distribution of variance 1, and every coefficient of any of the $\mathbf{x}_i$ is independent from all the others, so the $\mathbf{x}_i$'s satisfy the hypotheses of Lemma 4.1. Now:

$$\frac{1}{m}\sum_{i=1}^{m} \mathbf{x}_i \mathbf{x}_i^T = \frac{1}{m\sigma_a^2}\sum_{i=1}^{m} \mathbf{a}_i \mathbf{a}_i^T = \frac{1}{m\sigma_a^2} A^T A.$$

Therefore, Lemma 4.1 shows that:

$$\Pr\left[\lambda_{\min}\big(A^T A\big) < \big(1-\varepsilon(2^{-\eta}, m)\big)\cdot m\sigma_a^2 \text{ or } \lambda_{\max}\big(A^T A\big) > \big(1+\varepsilon(2^{-\eta}, m)\big)\cdot m\sigma_a^2\right] < 2^{-\eta}$$

with $\varepsilon(\delta, m)$ defined as above. Thus, to obtain (4.3), it suffices to take m such that $\varepsilon(2^{-\eta}, m) \leq \alpha$.

The value $\varepsilon(\delta, m)$ can be written as $4\tau^2 \cdot (\sqrt{8\rho} + \rho)$ where $\rho = \big(\log 9 \cdot n + \log(2/\delta)\big)/m$. For the choice of m in the statement of the theorem, we necessarily have $\rho < 1$ since $\sigma_a \leq \tau_a$, and hence $\tau^4 \geq 1$. As a result, $\varepsilon(\delta, m) \leq 16\tau^2 \cdot \sqrt{\rho}$. Thus, to obtain the announced result, it suffices to choose:

$$m \geq \frac{2^8 \tau^4}{\alpha^2}\Big(\log 9 \cdot n + \log 2^{1+\eta}\Big),$$

which concludes the proof. □

Remark 4.3. The ratio τ between the subgaussian moment τ_a of χ_a and the actual standard deviation σ_a is typically small (e.g. 1 for Gaussians, $\sqrt{3}$ for uniform distributions in a centered interval, etc.), so it isn't the important factor in the theorem.

The asymptotic bound saying that $m = \Omega\big((n+\eta)/\alpha^2\big)$ suffices to ensure that $\lambda_{\min}(A^T A)$ is within a factor α of the limit $m\sigma_a^2$ is a satisfactory result, but the implied constant in our theorem is admittedly rather large. This is an artifact of our reliance on Hsu et al.'s lemma. A more refined analysis is carried out by Litvak et al. in [32], and can in principle be used to reduce the constant C_1 in our theorem to $1 + o(1)$ for sufficiently large n. The authors omit concrete constants, however, and making [32, Th. 3.1] explicit is nontrivial.

From now on, let us suppose that the assumptions of Theorem 4.2 are satisfied for some $\alpha \in (0, 1)$, and η equal to the "security parameter". In particular, $A^T A$ is invertible with overwhelming probability, and we can thus write:

$$\tilde{\mathbf{s}} = (A^T A)^{-1} \cdot A^T \mathbf{b}.$$

As discussed in the beginning of this section, we would like to bound the distance between the estimator $\tilde{\mathbf{s}}$ and the actual solution $\mathbf{s}$ of the ILWE problem in the infinity norm, so as to obtain an inequality of the form (4.2). Since by definition $\mathbf{b} = A\mathbf{s} + \mathbf{e}$, we have:

$$\tilde{\mathbf{s}} - \mathbf{s} = (A^T A)^{-1} \cdot A^T\big(A\mathbf{s} + \mathbf{e}\big) - \mathbf{s} = (A^T A)^{-1} \cdot A^T \mathbf{e} = M\mathbf{e},$$

where M is the matrix $(A^TA)^{-1} \cdot A^T$. Now suppose that all the coefficients of $\mathbf{e}$ are τ_e-subgaussian. Since they are also independent, the vector $\mathbf{e}$ is a τ_e-subgaussian random vector in the sense of Definition 2.8. Therefore, it follows from Lemma 2.9 that $\tilde{\mathbf{s}} - \mathbf{s} = M\mathbf{e}$ is $\tilde{\tau}$-subgaussian, where:

$$\begin{aligned}\tilde{\tau} &= \|M^T\|_2^{\mathrm{op}} \cdot \tau_e = \tau_e\sqrt{\lambda_{\max}(MM^T)} = \tau_e\sqrt{\lambda_{\max}\big((A^TA)^{-1}A^T \cdot A(A^TA)^{-1}\big)} \\ &= \tau_e\sqrt{\lambda_{\max}\big((A^TA)^{-1}\big)} = \frac{\tau_e}{\sqrt{\lambda_{\min}(A^TA)}}.\end{aligned}$$

As a result, under the hypotheses of Theorem 4.2, $\tilde{\mathbf{s}}-\mathbf{s}$ is a $\frac{\tau_e}{\sigma_a\sqrt{(1-\alpha)m}}$-subgaussian random vector, except with probability at most $2^{-\eta}$ on the randomness of the matrix A.

This bound on the subgaussian moment can be used to derive a bound with high probability on the infinity norm as follows.

Lemma 4.4. *Let $\mathbf{v}$ be a τ-subgaussian random vector in $\mathbb{R}^n$. Then:*

$$\Pr\left[\|\mathbf{v}\|_\infty > t\right] \leq 2n \cdot \exp\Big(-\frac{t^2}{2\tau^2}\Big).$$

Proof. If we write $\mathbf{v} = (v_1, \ldots, v_n)$, we have $\|\mathbf{v}\|_\infty = \max(v_1, \ldots, v_n, -v_1, \ldots, -v_n)$. Therefore, the union bound shows that:

$$\Pr\left[\|\mathbf{v}\|_\infty > t\right] \leq \sum_{i=1}^{n} \Pr[v_i > t] + \Pr[-v_i > t]. \tag{4.4}$$

Now each of the random variables $v_1, \ldots, v_n, -v_1, \ldots, -v_n$ can be written as the scalar product of $\mathbf{v}$ with a unit vector of $\mathbb{R}^n$. Therefore, they are all τ-subgaussian. If X is one of them, the subgaussian tail bound of Lemma 2.6 shows that $\Pr[X > t] \leq \exp\big(-\frac{t^2}{2\tau^2}\big)$. Combined with (4.4), this gives the desired result. □

This is all we need to establish a sufficient condition for the least squares approach to return the correct solution to the ILWE problem with good probability.

Theorem 4.5. *Suppose that χ_a is τ_a-subgaussian and χ_e is τ_e-subgaussian, and let $(A, \mathbf{b} = A\mathbf{s}+\mathbf{e})$ the data constructed from m samples of the ILWE distribution $\mathscr{D}_{\mathbf{s},\chi_a,\chi_e}$, for some $\mathbf{s} \in \mathbb{Z}^n$. There exist constants $C_1, C_2 > 0$ (the same as in the hypotheses of Theorem 4.2) such that for all $\eta \geq 1$, if:*

$$m \geq 4\frac{\tau_a^4}{\sigma_a^4}(C_1 n + C_2\eta) \quad \textit{and} \quad m \geq 32\frac{\tau_e^2}{\sigma_a^2}\log(2n)$$

then the least squares estimator $\tilde{\mathbf{s}} = (A^TA)^{-1}A^T\mathbf{b}$ satisfies $\|\mathbf{s}-\tilde{\mathbf{s}}\|_\infty < 1/2$, and hence $\lceil\tilde{\mathbf{s}}\rfloor = \mathbf{s}$, with probability at least $1 - \frac{1}{2n} - 2^{-\eta}$.

Proof. Applying Theorem 4.2 with $\alpha = 1/2$ and the same constants C_1, C_2 as introduced in the statement of that theorem, we obtain that for $m \geq \frac{\tau_a^4}{\sigma_a^4}(4C_1 n + 4C_2\eta)$, we have

$$\Pr\left[\lambda_{\min}\left(A^T A\right) < m\sigma_a^2/2\right] < 2^{-\eta}. \tag{4.5}$$

Therefore, except with probability at most $2^{-\eta}$, we have $\lambda_{\min}\left(A^T A\right) \geq m\sigma_a^2/2$. We now assume that this condition is satisfied.

We have shown above that $\tilde{\mathbf{s}} - \mathbf{s}$ is a $\tilde{\tau}$-subgaussian random vector with $\tilde{\tau} = \tau_e/\sqrt{\lambda_{\min}(A^T A)}$. Applying Lemma 4.4 with $t = 1/2$, we therefore have:

$$\begin{aligned}\Pr\left[\|\tilde{\mathbf{s}} - \mathbf{s}\|_\infty > \frac{1}{2}\right] &\leq 2n \cdot \exp\Big(-\frac{1}{8\tilde{\tau}^2}\Big) \leq 2n \cdot \exp\Big(-\frac{\lambda_{\min}(A^T A)}{8\tau_e^2}\Big) \\ &\leq \exp\Big(\log(2n) - \frac{m\sigma_a^2}{16\tau_e^2}\Big).\end{aligned}$$

Thus, if we assume that $m \geq 32\frac{\tau_e^2}{\sigma_a^2}\log(2n)$, it follows that:

$$\Pr\left[\|\tilde{\mathbf{s}} - \mathbf{s}\|_\infty > \frac{1}{2}\right] \leq \exp\big(\log(2n) - 2\log(2n)\big) = \frac{1}{2n}.$$

This concludes the proof. □

In the typical case when τ_a and τ_e are no more than a constant factor larger than σ_a and σ_e, Theorem 4.5 with $\eta = \log(2n)$ says that there are constants C, C' such that whenever

$$m \geq Cn \qquad \text{and} \qquad m \geq C' \cdot \frac{\sigma_e^2}{\sigma_a^2}\log n \tag{4.6}$$

one can solve the ILWE problem with m samples with probability at least $1 - 1/n$ by rounding the least squares estimator. The first condition ensures that $A^T A$ is invertible and to control its eigenvalues: a condition of that form is clearly unavoidable to have a well-defined least squares estimator. On the other hand, the second condition gives a lower bound of the form (3.1) on the required number of samples; we see that this bound is only a factor $\log n$ worse than the information-theoretic lower bound, which is quite satisfactory.

We also note that the cost of this approach is equal to the complexity of computing $(A^T A)^{-1} A^T \mathbf{b}$, hence at most $O(n^2 \cdot m)$. This is quite efficient in practice (see §6 for concrete timings). In practice, arithmetic operations can be implemented using standard floating point instructions, since the almost scalar nature of $A^T A$ ensures that the computations are numerically very stable.

4.2 An Exact-CVP Oracle Will Not Help

One can interpret this approach to solving ILWE by computing a least squares estimator and rounding it as an application of Babai's *rounding algorithm* for the closest vector problem (CVP).

More precisely, consider the sublattice $L = A^T A \cdot \mathbb{Z}^n$ of $\mathbb{Z}^n$, which is full-rank when $A^T A$ is invertible (i.e. m large enough). Then, the ILWE problem can be seen as the problem of recovering the lattice vector $\mathbf{v} = A^T A\mathbf{s} \in L$ given the close vector $A^T\mathbf{b} = \mathbf{v} + A^T\mathbf{e}$ (which is essentially an instance of bounded distance decoding in L). Closeness in this setting is best measured in terms of the infinity norm. Now, since for large m, the matrix $A^T A$ is almost scalar, and hence the corresponding lattice basis of L is somehow already reduced, one can try to solve this problem by applying a CVP algorithm like Babai rounding directly on this basis. It is easy to see that this approach is identical to our least squares approach.

One could ask whether applying another CVP algorithm such as Babai's *nearest plane* algorithm could allow solving the problem with asymptotically fewer samples (e.g. reduce the $\log n$ factor in (4.6)). The answer is no. In fact, a much stronger result holds: one cannot improve Condition (4.6) using that strategy even given access to an *exact*-CVP oracle for any p-norm, $p \in [2, \infty]$. Given such an oracle, the secret vector $\mathbf{v}$ can be recovered uniquely if and only if the vector of noise $A^T\mathbf{e}$ lies in a ball centered on $\mathbf{v}$ and of radius half the first minimum of L in the p-norm, $\lambda_1^{(p)}(L) = \min_{x \in L} \|x\|_p$, that is:

$$\|A^T\mathbf{e}\|_p \leq \frac{\lambda_1^{(p)}(L)}{2}. \tag{4.7}$$

To take advantage of this condition, we need to get sufficiently precise estimates of both sides.

Estimation of the first minimum. Due to the quasi-scalar shape of the matrix $A^T A$, one can estimate accurately the $\lambda_1^{(p)}(L)$. Indeed, $A^T A$ has a low orthogonality defect, so that it is in a sense already reduced. Hence, the shortest vector of this basis constitutes a very good approximation of the shortest vector of L.

Lemma 4.6. *Suppose that χ_a is τ_a-subgaussian, and let $\tau = \tau_a/\sigma_a$. Let A be an $m \times n$ random matrix sampled from $\chi_a^{m \times n}$. Let L be the lattice generated by the rows of the matrix $A^T A$. There exist constants C_1, C_2 (the same as in Theorem 4.2) such that for all $\alpha \in (0,1)$, $p \geq 2$ and $\eta \geq 1$, if $m \geq (C_1 n + C_2\eta) \cdot (\tau^4/\alpha^2)$ then*

$$\Pr\left[\lambda_1^{(p)}(L)\big(A^T A\big) > m\sigma_a^2(1+\alpha)\right] \leq 2^{-\eta}. \tag{4.8}$$

Proof. Remark first that by norm equivalence in finite dimension, $\mathbf{x} \in \mathbb{R}^n$ we have $\|\mathbf{x}\|_p \leq \|\mathbf{x}\|_2$ so that $\lambda_1^{(p)}(L) \leq \lambda_1^{(2)}(L)$, this bound being actually sharp. Without loss of generality it then suffices to prove the result in 2-norm. From Theorem 4.2, we can assert that except with probability at most $2^{-\eta}$, $\|A^T A\|_2^{\mathrm{op}} \leq m\sigma_a^2(1+\alpha)$; for any integral vector $\mathbf{x} \in \mathbb{Z}^n$ we therefore have by definition of the operator norm:

$$\|A^T A\mathbf{x}\|_2 \leq m\sigma_a^2 \|\mathbf{x}\|_2 (1+\alpha).$$

In particular, for any $\mathbf{x} \in \mathbb{Z}^n$ of unit 2-norm, $\lambda_1^{(2)}(L) \leq \|A^T A\mathbf{x}\|_2 \leq (1+\alpha)m\sigma_a^2$. □

Estimation of the p-norm of $A^T\mathbf{e}$. Suppose that χ_e is a centered Gaussian distribution of standard deviation σ_e. The distribution of $A^T\mathbf{e}$ for $\mathbf{e} \sim \chi_e^n$ is then a Gaussian distribution of covariance matrix $\sigma_e^2 A^T A \approx m\sigma_a^2\sigma_e^2 I_n$. We deal with the cases $p = \infty$ and $p \leq \infty$ separately.

Case $p < \infty$: The expected p-th power of the p-norm of $A^T\mathbf{e}$ satisfies:

$$\mathbb{E}\Big[\|A^T\mathbf{e}\|_p^p\Big] = n\mathbb{E}[x^p] = n(2m)^{p/2}\sigma_e^p\sigma_a^p \cdot \frac{\Gamma\left(\frac{p}{2}+\frac{1}{2}\right)}{\sqrt{\pi}},$$

where x is drawn under the centered gaussian distribution of variance $m\sigma_e^2\sigma_a^2$, and Γ is classically the Euler's Gamma function. But by the partial converse of Jensen's inequality for norms of Stadje [44] we have:

$$\mathbb{E}\Big[\|A^T\mathbf{e}\|_p^p\Big] \leq 2^p\Gamma\left(\frac{p}{2}+\frac{1}{2}\right)\sqrt{\pi}^{(p-1)}\mathbb{E}\Big[\|A^T\mathbf{e}\|_p\Big]^p$$

so that:

$$n^{1/p}\sigma_e\sigma_a\sqrt{\frac{m}{2\pi}} \leq \mathbb{E}\Big[\|A^T\mathbf{e}\|_p\Big]$$

Case $p = \infty$: The estimate is obtained by the order statistic theory of Gaussian distributions (see e.g. [42]):

$$C_\infty\sigma_e\sigma_a\sqrt{m\log n} \leq \mathbb{E}\Big[\|A^T\mathbf{e}\|_\infty\Big],$$

where $C_\infty = \frac{3}{2}\left(1-\frac{1}{e}\right) - \frac{1}{\sqrt{2\pi}} \approx 0.23$

Now that we have access to the expected value of the random variable $\|A^T\mathbf{e}\|_p$, we are going to use the concentration of its distribution around its expected value. Explicitly by the random version of Dvoretzky's theorem proven in [39], there exist absolute constants $K, c > 0$ such that for any $0 < \varepsilon < 1$:

$$\Pr\left[\left|A^T\mathbf{e} - \mathbb{E}\Big[\|A^T\mathbf{e}\|_p\Big]\right| > \varepsilon\mathbb{E}\Big[\|A^T\mathbf{e}\|_p\Big] \leq Ke^{-c\beta(n,p,\varepsilon)}\right. \tag{4.9}$$

with

$$\beta(m,p,\varepsilon) = \begin{cases} \varepsilon^2 n & \text{if } 1 < p \leq 2 \\ \max(\min(2^{-p}\varepsilon^2 n, (\varepsilon n)^{2/p}), \varepsilon p n^{2/p}) & \text{if } 2 < p \leq c_0\log n \\ \varepsilon\log n & \text{if } p > c_0\log n \end{cases},$$

for $0 < c_0 < 1$ a fixed absolute constant.

Summing up. Taking $\varepsilon = 1/2$ in (4.9) ensures that, except with probability $Ke^{-c\beta(n,p,1/2)}$,

$$\frac{1}{2}\mathbb{E}\Big[\|A^T\mathbf{e}\|_p\Big] \leq \|A^T e\|_p \leq \frac{3}{2}\mathbb{E}\Big[\|A^T\mathbf{e}\|_p\Big]. \tag{4.10}$$

For any fixed p, the probability can be made as small as desired for large enough n. We can therefore assume that (4.10) occurs with probability at least $1-\delta$ for some small $\delta > 0$.

In that case, Condition (4.7) asserts that if $\mathbb{E}\Big[\|A^T\mathbf{e}\|_p\Big] > \lambda_1^{(p)}(L)$ then $\mathbf{s}$ can't be decoded uniquely in L. Now using the result of Lemma 4.6 with $\alpha = 1/2$ and the previous estimates, we know that this is the case when:

$$n^{1/p}\sigma_e\sigma_a\sqrt{\frac{m}{2\pi}} > \frac{3}{2}m\sigma_a^2, \qquad \text{that is, } m < \left(\frac{\sigma_e}{\sigma_a}\right)^2 \frac{2n^{2/p}}{9\pi},$$

when $p < \infty$, and

$$0.23\sigma_e\sigma_a\sqrt{m\log n} > \frac{3}{2}m\sigma_a^2, \qquad \text{that is, } m < 0.02\left(\frac{\sigma_e}{\sigma_a}\right)^2 \log n,$$

otherwise. In both cases, it follows that we must have $m = \Omega\big((\sigma_e/\sigma_a)^2\log n\big)$ for the CVP algorithm to output the correct secret with probability $> \delta$. Thus, this approach cannot improve upon the least squares bound 4.5 by more than a constant factor.

4.3 Sparse Secret and Compressed Sensing

Up until this point, we have supposed that the number m of samples we have access to is greater than the dimension n. Indeed, without additional information on the secret $\mathbf{s}$, this condition is necessary to get a well-defined solution to the ILWE problem *even without noise*.

Suppose however that the secret $\mathbf{s}$ is known to be *sparse*, with only a small number $S \ll n$ of non zero coefficients. Even if the positions of these non zero coefficients are not known, knowledge of the sparsity S may help in determining the secret, possibly even with fewer samples than the ambient dimension n with the sole additional knowledge of its sparsity (though of course more than S samples are necessary!). Such a recovery is made possible by compressed sensing techniques, epitomized by the results of Candes and Tao in [14]. The idea is once again to find an estimator $\tilde{\mathbf{s}}$ such that the infinity norm $\|\tilde{\mathbf{s}}-\mathbf{s}\|_\infty$ is small enough to fully recover the secret $\mathbf{s}$ from it. This can be done with the Dantzig selector introduced in [14], and efficiently computable as a solution $\tilde{\mathbf{s}} = (\tilde{s}_1, \ldots, \tilde{s}_n)$ of the following linear program with $2n$ unknowns $\tilde{s}_i, \tilde{u}_i$, $1 \leq i \leq n$:

$$\begin{aligned} &\min \sum_{i=1}^{n} u_i \qquad \text{such that} \qquad -u_i \leq \tilde{s}_i \leq u_i \qquad \text{and} \\ &-\sigma_e\sigma_a\sqrt{2m\log n} \leq \big[AA^T(A^T\mathbf{b} - A^TA\tilde{\mathbf{s}})\big]_i \leq \sigma_e\sigma_a\sqrt{2m\log n}. \end{aligned} \tag{4.11}$$

Table 1: Maximum value of the ratio σ_e/σ_a to recover a S sparse secret in dimension n with the Dantzig selector

n \ (S/n)	0.1	0.3	0.5	0.7	0.9
128	16.2	9.4	7.3	6.1	5.4
256	15.2	8.8	6.8	5.7	5.0
512	14.3	8.3	6.4	5.4	4.8
1024	13.6	7.8	6.0	5.1	4.5
2048	13.0	7.5	5.8	4.9	4.3

In the case when the distributions χ_e and χ_a are Gaussian distributions of respective standard deviations σ_e and σ_e, the quality of the output of the program defined by (4.11) is quantified as follows.

Theorem 4.7 (adapted from [14]). *Suppose* $\mathbf{s} \in \mathbb{Z}^n$ *is any* S*-sparse vector so that* $\log(m\sigma_a^2/n)S \leq m$ *Then with large probability,* $\tilde{\mathbf{s}}$ *obeys the relation*

$$\|\tilde{\mathbf{s}} - \mathbf{s}\|_2^2 \leq 2C_1^2 S \log n \left(\frac{\sigma_e}{\sqrt{m}\sigma_a}\right)^2 \tag{4.12}$$

for some constant $C_1 \approx 4$.

Hence as before, if $\|\tilde{\mathbf{s}} - \mathbf{s}\|_2^2 \leq 1/4$, we have $\|\tilde{\mathbf{s}} - \mathbf{s}\|_\infty \leq 1/2$ and one can then decode the coefficients of $\mathbf{s}$ by rounding $\tilde{\mathbf{s}}$. This is satisfied with high probability as soon as:

$$2C_1^2 \frac{S \log n}{m} \left(\frac{\sigma_e}{\sigma_a}\right)^2 \leq \frac{1}{4}.$$

Since we aim at solving the ILWE problem in parsimonious sample setting, where $m < n$ we deduce that the compressed sensing methodology can be successfully applied when

$$S \leq \frac{n}{8C_1^2 \log n} \left(\frac{\sigma_a}{\sigma_e}\right)^2. \tag{4.13}$$

Let us discuss the practicality of this approach with regards to the parameters of the ILWE problem. First of all, note that in order to make Condition (4.13) non-vacuous, one needs σ_e and σ_a to satisfy:

$$2C_1\sqrt{\frac{2\log n}{n}} \leq \frac{\sigma_a}{\sigma_e} \leq 2C_1\sqrt{2\log n},$$

where the lower bound follows from the fact that S is a positive integer, and the upper bound from the observation that the right-hand side of (4.13) must

be smaller than n to be of any interest compared to the trivial bound $S \leq n$. Practically speaking, this means that this approach is only interesting when the ratio σ_e/σ_a is relatively small; concrete bounds are provided in Table 1 various sparsity levels and dimensions ranging from 128 to 2048.

We note that the required sparsity is much higher than proposed parameters for BLISS, for example. Moreover, the complexity of this linear programming based approach is worse than least squares regression. However, only this method is applicable when only $m < n$ samples are available.

5 Application to the Side-channel Attack of BLISS

5.1 BLISS Signatures and Rejection Sampling Leakage

The BLISS signature scheme [17] is a lattice-based signature scheme based on the Ring-Learning With Error (RLWE) assumption. Its signing algorithm is recalled in Figure 1.

The rejection sampling. The BLISS signature scheme follows the "Fiat–Shamir with aborts" paradigm of Lyubashevsky [35]. In particular, signature generation involves a *rejection sampling* step (Step 8 of function Sign in Figure 1) which is essential for security: in order to ensure that the distribution of signatures is independent of the secret key $\mathbf{s} = (\mathbf{s}_1, \mathbf{s}_2)$, a signature candidate $\big(\mathbf{z} = (\mathbf{z}_1, \mathbf{z}_2), \mathbf{c}\big)$ should be kept with probability

$$1 \Bigg/ \left(M \exp\left(-\frac{\|\mathbf{sc}\|^2}{2\sigma^2} \right) \cosh\left(\frac{\langle \mathbf{z}, \mathbf{sc} \rangle}{\sigma^2} \right) \right).$$

Since it would be impractical to directly compute this expression involving transcendental functions with sufficient precision, all existing implementations of BLISS [18, 41, 45] rely instead on the iterated Bernoulli trials technique described in [17, §6]. A signature $(\mathbf{z}, \mathbf{c})$ is kept if the function calls SampleBernExp($x_{\exp}$) and SampleBernCosh($x_{\cosh}$) both return 1, where functions SampleBernExp and SampleBernCosh are described in Figure 2 and the values $x_{\exp}, x_{\cosh}$ are given respectively by $x_{\exp} = \log M - \|\mathbf{sc}\|^2$ and $x_{\cosh} = 2 \cdot \langle \mathbf{z}, \mathbf{sc} \rangle$.

Side-channel leakage of the rejection sampling. Based on their description in Figure 2, it is clear that SampleBernExp and SampleBernCosh do not run in constant time. In fact, they iterate over the bits of their input, and part of the code is executed when the bit is 1 and skipped over when the bit is 0. As a result, as observed by Espitau et al. [19, §3], the inputs $x_{\exp}, x_{\cosh}$ of these functions can be read off directly on a trace of power consumption or electromagnetic emanations, in much the same way as naive square-and-multiply implementations of RSA leak the secret exponent via simple power analysis [28, §3.1]. As a result, side-channel analysis allows to reliably recover the squared

Fig. 1: BLISS signing algorithm. The hash function H is modeled as a RO with values in the set of polynomials in $\mathcal{R}$ with 0/1-coefficient and Hamming weight κ. See [17] for details regarding notation like ζ, $\lceil\cdot\rfloor_d$ and p not discussed in this paper.

function SIGN($\mu, pk = \mathbf{v}_1$, $sk = \mathbf{s} = (\mathbf{s}_1, \mathbf{s}_2)$)
  $\mathbf{y}_1, \mathbf{y}_2 \leftarrow D_\sigma^{\bar{n}}$
  $\mathbf{u} = \zeta \cdot \mathbf{v}_1 \cdot \mathbf{y}_1 + \mathbf{y}_2 \mod 2q$
  $\mathbf{c} \leftarrow H(\lceil \mathbf{u} \rfloor_d \mod p, \mu)$
  choose a random bit b
  $\mathbf{z}_1 \leftarrow \mathbf{y}_1 + (-1)^b \mathbf{s}_1 \mathbf{c}$
  $\mathbf{z}_2 \leftarrow \mathbf{y}_2 + (-1)^b \mathbf{s}_2 \mathbf{c}$
  continue with probability $1/\big(M \exp(-\|\mathbf{sc}\|^2/(2\sigma^2)) \cosh(\langle \mathbf{z}, \mathbf{sc}\rangle/\sigma^2\big)$; otherwise **restart**
  $\mathbf{z}_2^\dagger \leftarrow (\lceil \mathbf{u} \rfloor_d - \lceil \mathbf{u} - \mathbf{z}_2 \rfloor_d) \mod p$
  return $(\mathbf{z}_1, \mathbf{z}_2^\dagger, \mathbf{c})$

norm $\|\mathbf{sc}\|^2 = \|\mathbf{s}_1\mathbf{c}\|^2 + \|\mathbf{s}_2\mathbf{c}\|^2$ and the scalar product $\langle \mathbf{z}, \mathbf{sc}\rangle = \langle \mathbf{z}_1, \mathbf{s}_1\mathbf{c}\rangle + \langle \mathbf{z}_2, \mathbf{s}_2\mathbf{c}\rangle$ from generated signatures.

Espitau et al. show that the norm leakage can be leveraged in practice to recover the secret key from a little over $\bar{n}$ signature traces, where $\bar{n}$ is the extension degree of the ring $\mathcal{R}$ ($\bar{n} = 512$ for the most common parameters). However, the recovery technique is mathematically quite involved and computationally costly (it is based on the Howgrave-Graham–Szydlo solution to cyclotomic norm equations [25], and takes over a month of CPU time for typical parameters). More importantly, it has the major drawback of relying on the ability to factor this norm and thus only applying to "weak" signing keys satisfying a certain semismoothness condition (around 7% of BLISS secret keys).

It is natural to think that the scalar product leakage, which is linear rather than quadratic in the secret key, is a more attractive target to attack. And indeed, Espitau et al. point out that in a simplified version of BLISS where $\mathbf{z}_2$ is returned in full as part of signatures, it is very easy to recover the secret key from about $2\bar{n}$ side-channel traces using elementary linear algebra. However, in the actual BLISS scheme, the element $\mathbf{z}_2$ is returned in a compressed form $\mathbf{z}_2^\dagger$, so that the linear system arising from scalar product leakage is noisy. Solving this linear system amounts to solving a problem analogous to LWE [43] in dimension about $2\bar{n}$, which leads Espitau et al. to conclude that this approach is unlikely to be helpful. In doing so, however, they overlook a crucial difference between standard LWE and the problem that actually arises in this way, namely the *lack of modular reduction*.

5.2 Description of the Attack

As we have mentioned already, recovering the secret $\mathbf{s} \in \mathbb{Z}^{2\bar{n}} = \mathbb{Z}^n$ from the linear leakage $\langle \mathbf{z}, \mathbf{sc}\rangle$ essentially amounts to an instance of the ILWE problem. We now describe more precisely in what sense. To do so, we need to write this inner product in terms of the known ring elements $(\mathbf{c}, \mathbf{z}_1, \mathbf{z}_2^\dagger)$ that appear in the

Fig. 2: Sampling algorithms for the distributions $\mathcal{B}_{\exp(-x/2\sigma^2)}$ and $\mathcal{B}_{1/\cosh(x/\sigma^2)}$. The values $c_i = 2^i/f$ precomputed, and the x_i's are the bits in the binary expansion of $x = \sum_{i=0}^{\ell-1} 2^i x_i$. BLISS uses $x = K - \|\mathbf{sc}\|^2$ for the input to the exponential sampler, and $x = 2\langle \mathbf{z}, \mathbf{sc}\rangle$ for the input to the cosh sampler.

```
1: function SampleBernExp(x)
2:     for i = 0 to ℓ − 1 do
3:         if x_i = 1 then
4:             Sample a ← B_{c_i}
5:             if a = 0 then return 0
6:     return 1
```

```
1: function SampleBernCosh(x)
2:     if x < 0 then x ← −x
3:     Sample a ← B_{exp(−x/f)}
4:     if a = 1 then return 1
5:     Sample b ← B_{1/2}
6:     if b = 1 then restart
7:     Sample c ← B_{exp(−x/f)}
8:     if c = 1 then restart
9:     return 0
```

signature on the one hand, and unknown elements on the other hand. This can be done as follows:

$$\begin{aligned}\langle \mathbf{z}, \mathbf{sc}\rangle &= \langle \mathbf{z}_1, \mathbf{s}_1\mathbf{c}\rangle + \langle \mathbf{z}_2, \mathbf{s}_2\mathbf{c}\rangle = \langle \mathbf{z}_1\mathbf{c}^*, \mathbf{s}_1\rangle + \langle 2^d\mathbf{z}_2^\dagger, \mathbf{s}_2\mathbf{c}\rangle + \langle \mathbf{z}_2 - 2^d\mathbf{z}_2^\dagger, \mathbf{s}_2\mathbf{c}\rangle \\ &= \langle \mathbf{z}_1\mathbf{c}^*, \mathbf{s}_1\rangle + \langle 2^d\mathbf{z}_2^\dagger\mathbf{c}^*, \mathbf{s}_2\rangle + e = \langle \mathbf{a}, \mathbf{s}\rangle + e,\end{aligned}$$

where we let:

$$\mathbf{a} = (\mathbf{z}_1\mathbf{c}^*, 2^d\mathbf{z}_2^\dagger\mathbf{c}^*) \in \mathbb{Z}^{2\bar{n}} = \mathbb{Z}^n \quad \text{and} \quad e = \langle \mathbf{z}_2 - 2^d\mathbf{z}_2^\dagger, \mathbf{s}_2\mathbf{c}\rangle.$$

The vector $\mathbf{a}$ can be computed from the signature, and is therefore known to the side-channel attacker, whereas e is some unknown value. In these expressions, $\mathbf{c}^*$ is the conjugate of $\mathbf{c}$ with respect to the inner product (i.e. the matrix of multiplication by $\mathbf{c}$ in the polynomial basis of $\mathbb{Z}[x]/(x^{\bar{n}}+1)$ is the transpose of that of $\mathbf{c}$).

Now the rejection sampling ensures that the coefficients of $\mathbf{z}_1$ are independent and distributed according to a discrete Gaussian D of standard deviation σ. On the other hand, $\mathbf{c}$ is a random vector with coefficients in $\{0, 1\}$ and exactly κ non zero coefficients; thus, $\mathbf{c}^*$ has a similar shape possibly up to the sign of coefficients. It follows that the coefficients of $\mathbf{z}_1\mathbf{c}^*$ are all linear combinations with ± 1 coefficients of exactly κ independent samples from D and the signs clearly do not affect the resulting distribution.

Therefore, if we denote by χ_a the distribution $D^{*\kappa}$ obtained by summing κ independent samples from D, the coefficients of $\mathbf{z}_1\mathbf{c}^*$ follow χ_a. It is not exactly correct that $\mathbf{z}_1\mathbf{c}^*$ as a whole follows $\chi_a^{\bar{n}}$ (as its coefficients are not rigorously independent), but we will heuristically ignore that subtlety and pretend it does. Note that χ_a is a distribution of variance:

$$\sigma_a^2 = \mathrm{Var}\left(D^{*\kappa}\right) = \kappa \cdot \mathrm{Var}(D) = \kappa\sigma^2.$$

We have not precisely described how the BLISS signature compression works, but roughly speaking, $\mathbf{z}_2^\dagger$ is essentially obtained by keeping the $(\log q - d)$ most significant bits of $\mathbf{z}_2$, and therefore the distribution of $2^d\mathbf{z}_2^\dagger$ is close to that of $\mathbf{z}_2$. The distributions cannot coincide exactly, since all the coefficients of $2^d\mathbf{z}_2^\dagger$ are multiples of 2^d while this normally does not happen for $\mathbf{z}_2$, but the difference will not matter much for our purposes, and we will therefore heuristically assume that the entire vector $\mathbf{a}$ is distributed as χ_a^n.

We now turn our attention to the noise value e, which we write as $\langle \mathbf{w}, \mathbf{u} \rangle$ with $\mathbf{w} = \mathbf{z}_2 - 2^d\mathbf{z}_2^\dagger$ and $\mathbf{u} = \mathbf{s}_2\mathbf{c}$. Now, $\mathbf{w}$ is obtained as the difference between $\mathbf{z}_2$ and $2^d\mathbf{z}_2^\dagger$, where again the latter is roughly speaking obtained by zeroing out the d least significant bits of $\mathbf{z}_2$ in a centered way. We can therefore heuristically expect that the coefficients of $\mathbf{w}$ are distributed uniformly in $[-2^{d-1}, 2^{d-1}] \cap \mathbb{Z}$, i.e. $\mathbf{w} \sim \mathscr{U}_\alpha^n$ with $\alpha = 2^{d-1}$. In particular, these coefficients have variance $\alpha(\alpha+1)/3 \approx 2^{2d}/12$.

As for $\mathbf{u}$, its coefficients are obtained as sums of κ coefficients of $\mathbf{s}_2$. Now $\mathbf{s}_2$ itself (ignoring the constant coefficient, which is shifted by 1) is obtained as a random vector with $\delta_1\bar{n}$ coefficients equal to ± 2, $\delta_2\bar{n}$ coefficients equal to ± 4 and all its other coefficients equal to zero. This is a somewhat complicated distribution to describe, but we do not make a large approximation by pretending that all the coefficients are sampled independently in the set $\{-4, -2, 0, 2, 4\}$ with probabilities $\delta_2/2, \delta_1/2, (1-\delta_1-\delta_2), \delta_1/2$ and $\delta_2/2$ respectively. Making that approximation, it follows that the coefficients of $\mathbf{u}$ have variance $\kappa \cdot (4\delta_1 + 16\delta_2)$.

Write $\mathbf{u} = (u_1, \ldots, u_{\bar{n}})$ and $\mathbf{w} = (w_1, \ldots, w_{\bar{n}})$. Under the heuristic approximations above, since $\mathbf{w}$ and $\mathbf{u}$ are independent and their coefficients have mean zero, the error e follows a certain bounded distribution χ_e of variance σ_e^2 given by:

$$\sigma_e^2 = \mathbb{E}[e^2] = \mathbb{E}\Big[\Big(\sum_{i=1}^{\bar{n}} w_i u_i\Big)^2\Big] = \mathbb{E}\Big[\sum_{i,j} w_i w_j u_i u_j\Big] = \mathbb{E}\Big[\sum_{i=1}^{\bar{n}} w_i^2 u_i^2\Big]$$
$$= \sum_{i=1}^{\bar{n}} \mathbb{E}[w_i^2] \cdot \mathbb{E}[u_i^2] = \bar{n} \cdot \mathrm{Var}\left(\mathscr{U}_\alpha\right) \cdot \kappa(4\delta_1 + 16\delta_2) \approx \frac{2^{2d}}{3}(\delta_1 + 4\delta_2)\bar{n}\kappa.$$

With these various approximations, recovering $\mathbf{s}$ from the leakage exactly becomes an ILWE problem with distributions χ_a and χ_e, where each side-channel trace provides a sample. It should therefore be feasible to recover the full secret key with least squares regression using $m = O\big((\sigma_e/\sigma_a)^2 \log n\big)$ traces.

5.3 Experimental Distributions

The description of the previous section made a number of heuristic approximations which we know cannot be precisely satisfied in practice. In order to validate those approximations nonetheless, we have carried out numerical simulations comparing in particular our estimates for the standard deviations σ_a and σ_e of the distributions of $\mathbf{a}$ and e with the standard deviations obtained from the actual rejection sampling leakage in BLISS.

Table 2: Parameter estimation for ILWE instances arising from the side channel attack

	BLISS–0	BLISS–I	BLISS–II	BLISS–III	BLISS–IV
$n = 2\bar{n}$	512	1024	1024	1024	1024
σ_a (theory)	346	1031	513	1369	1692
σ_e (theory)	1553	49695	49695	38073	24535
$\sigma_{\mathbf{a}_1}$ (exp.)	347	1031	513	1370	1691
$\sigma_{\mathbf{a}_2}$ (exp.)	349	2009	1418	1782	1814
σ_e (exp.)	1532	42170	32319	38627	23926

These simulations were carried out in Python using the numpy package. We used 10000 ILWE samples arising from side channel leaks for each BLISS parameter set. Results are collected in Table 2; experimental values for σ_a are provided separately for the two halves $(\mathbf{a}_1, \mathbf{a}_2)$ of the vector $\mathbf{a}$, which we have seen are computed differently. As we can see, the experimental values match the heuristic estimates quite closely overall.

6 Numerical Simulations

In this section, we present simulation results for recovering ILWE secrets using linear regression, first for normal ILWE instances, and then for ILWE instances arising from BLISS side-channel leakage, as described in §5.2, leading to BLISS secret key recovery. These results are based on simulated leakage data rather than actual side-channel traces. However, we note that the leakage scenario for BLISS is essentially identical to the one described in [19] (namely, a SPA/SEMA setting where each trace reveals the exact value of a certain function of the secret key—in our case, the linear function given by the inner product), and was therefore experimentally validated in that paper.

6.1 Plain ILWE

Recall that the ILWE problem is parametrized by $n, m \in \mathbb{Z}$ and probability distributions χ_a and χ_e. Samples are computed as $\mathbf{b} = A\mathbf{s} + \mathbf{e}$, where $\mathbf{s} \in \mathbb{Z}^n$, $\mathbf{b} \in \mathbb{Z}^m$, $A \in \mathbb{Z}^{m \times n}$ with entries drawn from χ_a, and $e \in \mathbb{Z}^m$ with entries drawn from χ_e. Choosing χ_a and χ_e as discrete gaussian distributions with standard deviations σ_a and σ_e respectively, we investigated the number of samples, m required to recover ILWE secret vectors $\mathbf{s} \in \mathbb{Z}^n$ for various concrete values of n, σ_a and σ_e. We sampled sparse secret vectors $\mathbf{s}$ uniformly at random from the set of vectors with $\lceil 0.15n \rceil$ entries set to ± 1, $\lceil 0.15n \rceil$ entries set to ± 2, and the rest zero.

We present two types of experimental results for plain ILWE. In our first experiment, we began by estimating the number of samples m required to recover

the secret perfectly with good probability, for different values of n, σ_a, and σ_e. Then, fixing m, we measured the probability of recovering $\mathbf{s}$ over the random choices of $\mathbf{s}$, A and e. Our results are displayed in Table 3.

In our second experiment, we investigated the distribution of the minimum value of m required to recover the secret perfectly, over the random choices of $\mathbf{s}$, A, and $\mathbf{e}$, when the linear regression method was run to completion. In other words, for fixed n, σ_a, and σ_e, we generated more and more samples until the secret could be perfectly recovered. Our results for $\sigma_e = 2000$ are plotted in Figure 3. Additional results and some additional notes may be found in the full version of this paper [10]. Each figure plots the dimension n against the mean number of samples m required to recover the secret, for $\sigma_a = 100$, 200, and 500. Here, 'mean' refers to the interquartile mean number of samples. The error bars show the upper and lower quartiles for the number of samples required.

The results of our second experiment are consistent with the theoretical results given in §4.1. According to (4.6), we require

$$m \geq C' \cdot \frac{\sigma_e^2}{\sigma_a^2} \log n$$

samples in order to recover the secret correctly. The dimension n on the horizontal axis of each graph is plotted on a logarithmic scale. Therefore, theory predicts that we should observe a straight line, which the graphs confirm.

The gradient of the graph corresponds to the constant C' giving the number of samples required for secret-recovery in practice. Note that in this case, where χ_a and χ_e follow the discrete Gaussian distribution, Theorem 4.5 gives $C' = 32$ for a small failure probability of $\frac{1}{2n}$. However, in this experiment, we are likely to succeed much sooner, with a smaller number of samples. For example, in any particular trial, as soon as m is such that the failure probability is at least one half, we are likely to recover the secret. This explains why the gradient is much lower than given by Theorem 4.5. Computing the gradients of the lines of best fit and dividing by $(\sigma_e/\sigma_a)^2$ gives an estimate for the observed value of the constant C'. See the full version of this paper [10] for details.

6.2 BLISS Side-Channel Attack

Having obtained an instance of the ILWE problem from BLISS side-channel leakage as described in §5.2, we used linear regression to recover BLISS secret keys. We performed several trials. For each trial, we generated ILWE samples using side-channel leakage until we could recover the secret key. For BLISS–0, we simply used regression to recover the entire secret key. For BLISS–I and BLISS–II, we usually ran into memory issues before being able to successfully recover the entire secret key. However, we noticed that in practice, we could recover the first half of the secret key correctly using far fewer samples. Since the two halves of the secret key are related by the public key, this is sufficient to compute the entire secret key. Therefore, for BLISS–I and BLISS–II, we stopped generating samples as soon as the least-squares estimator correctly recovered the first half of the secret.

Table 3: Practical results of the experiments on ILWE

n	σ_a	σ_e	m	Success
128	100	1000	3300	6/10
	100	2000	11500	6/10
	100	5000	65000	4/10
	200	1000	900	5/10
	200	2000	4000	7/10
	200	5000	17000	4/10
	300	1000	550	10/10
	300	2000	1890	8/10
	300	5000	9000	7/10
	400	1000	350	8/10
	400	2000	800	5/10
	400	5000	5750	7/10
	500	1000	350	10/10
	500	2000	700	6/10
	500	5000	3300	4/10
256	100	1000	5600	9/10
	100	2000	14500	6/10
	100	5000	95000	7/10
	200	1000	1300	6/10
	200	2000	4700	8/10
	200	5000	23000	6/10
	300	1000	900	9/10
	300	2000	1800	5/10
	300	5000	12000	8/10
	400	1000	550	10/10

n	σ_a	σ_e	m	Success
256	400	5000	6000	5/10
	500	1000	450	7/10
	500	2000	950	8/10
	500	5000	4200	5/10
512	100	1000	5100	7/10
	100	2000	16000	4/10
	200	1000	1600	9/10
	200	2000	5200	7/10
	300	1000	1000	8/10
	300	2000	2600	8/10
	400	1000	900	10/10
	400	2000	1500	4/10
	500	1000	800	10/10
	500	2000	1250	8/10
1024	100	1000	5950	10/10
	100	2000	19000	5/10
	200	1000	2250	6/10
	200	2000	5900	6/10
	300	1000	1550	7/10
	300	2000	3350	6/10
	400	1000	1350	9/10
	400	2000	2300	7/10
	500	1000	1500	10/10
	500	2000	1900	8/10

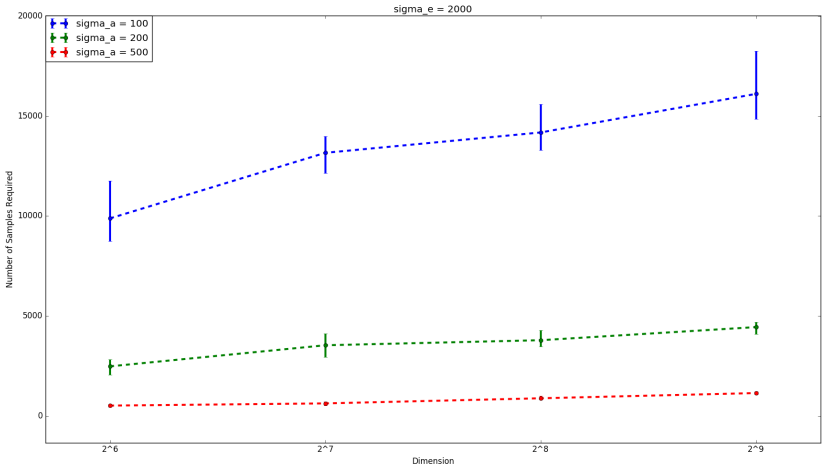

Fig. 3: Results for $\sigma_e = 2000$

Table 4: Number of samples required to recover the secret key (minimum, lower quartile, interquartile mean, upper quartile, maxium)

	# Trials	Min	LQ	IQM	UQ	Max
BLISS–0	12	1203	1254	1359.5	1515	1641
BLISS–I	12	14795	18648	20382.9	21789	24210
BLISS–II	8	19173	20447	22250.3	24482	29800

For these two different scenarios, we obtain the results displayed on Table 4, which gives information on the range, quartiles, and interquartile mean of the number of samples required. Typical timings for the side-channel attacks, using SAGEMath, on a laptop with 2.60GHz processor, are displayed in Table 5. Timings are in the orders of minutes and seconds. By comparison, some of the attacks from [19] may take hours, or even days, of CPU time.

Acknowledgments This work has been supported in part by the European Union's H2020 Programme under grant agreement number ERC-669891.

References

1. Aggarwal, D., Joux, A., Prakash, A., Santha, M.: A new public-key cryptosystem via Mersenne numbers. Cryptology ePrint Archive, Report 2017/481 (2017), http://eprint.iacr.org/2017/481

Table 5: Typical timings for secret key recovery

	Typical ILWE sample gen.	Typical time for regression
BLISS–0	$\approx$ 2m	$\approx$ 5s
BLISS–I	$\approx$ 10m	$\approx$ 2m
BLISS–II	$\approx$ 10m	$\approx$ 2m

2. Albrecht, M.R.: On dual lattice attacks against small-secret LWE and parameter choices in HElib and SEAL. In: Coron, J., Nielsen, J.B. (eds.) EUROCRYPT 2017, Part II. LNCS, vol. 10211, pp. 103–129. Springer, Heidelberg (Apr / May 2017)
3. Albrecht, M.R., Faugère, J.C., Fitzpatrick, R., Perret, L.: Lazy modulus switching for the BKW algorithm on LWE. In: Krawczyk, H. (ed.) PKC 2014. LNCS, vol. 8383, pp. 429–445. Springer, Heidelberg (Mar 2014).
4. Alwen, J., Krenn, S., Pietrzak, K., Wichs, D.: Learning with rounding, revisited - new reduction, properties and applications. In: Canetti, R., Garay, J.A. (eds.) CRYPTO 2013, Part I. LNCS, vol. 8042, pp. 57–74. Springer, Heidelberg (Aug 2013). doi:10.1007/978-3-642-40041-4_4
5. Applebaum, B., Cash, D., Peikert, C., Sahai, A.: Fast cryptographic primitives and circular-secure encryption based on hard learning problems. In: Halevi, S. (ed.) CRYPTO 2009. LNCS, vol. 5677, pp. 595–618. Springer, Heidelberg (Aug 2009)
6. Arora, S., Ge, R.: New algorithms for learning in presence of errors. In: Aceto, L., Henzinger, M., Sgall, J. (eds.) ICALP 2011, Part I. LNCS, vol. 6755, pp. 403–415. Springer, Heidelberg (Jul 2011)
7. Bai, S., Galbraith, S.D.: Lattice decoding attacks on binary LWE. In: Susilo, W., Mu, Y. (eds.) ACISP 14. LNCS, vol. 8544, pp. 322–337. Springer, Heidelberg (Jul 2014).
8. Banerjee, A., Peikert, C., Rosen, A.: Pseudorandom functions and lattices. In: Pointcheval, D., Johansson, T. (eds.) EUROCRYPT 2012. LNCS, vol. 7237, pp. 719–737. Springer, Heidelberg (Apr 2012)
9. Bogdanov, A., Guo, S., Masny, D., Richelson, S., Rosen, A.: On the hardness of learning with rounding over small modulus. In: Kushilevitz, E., Malkin, T. (eds.) TCC 2016-A, Part I. LNCS, vol. 9562, pp. 209–224. Springer, Heidelberg (Jan 2016).
10. Bootle, J., Delaplace, C., Espitau, T., Fouque, P.A., Tibouchi, M.: LWE without modular reduction and improved side-channel attacks against BLISS. Cryptology ePrint Archive, Report 2018/822 (2018), http://eprint.iacr.org/2018/822. Full version of this paper
11. Bootle, J., Tibouchi, M., Xagawa, K.: Cryptanalysis of Compact-LWE. In: Smart, N.P. (ed.) CT-RSA. LNCS, vol. 10808, pp. 80–97. Springer (2018)
12. Brakerski, Z., Langlois, A., Peikert, C., Regev, O., Stehlé, D.: Classical hardness of learning with errors. In: Boneh, D., Roughgarden, T., Feigenbaum, J. (eds.) 45th ACM STOC. pp. 575–584. ACM Press (Jun 2013)
13. Bruinderink, L.G., Hülsing, A., Lange, T., Yarom, Y.: Flush, gauss, and reload - A cache attack on the BLISS lattice-based signature scheme. In: Gierlichs, B., Poschmann, A.Y. (eds.) CHES 2016. LNCS, vol. 9813, pp. 323–345. Springer, Heidelberg (Aug 2016).

14. Candes, E., Tao, T.: The Dantzig selector: Statistical estimation when p is much larger than n. Ann. Statist. **35**(6), 2313–2351 (dec 2007)
15. Dodis, Y., Goldwasser, S., Kalai, Y.T., Peikert, C., Vaikuntanathan, V.: Public-key encryption schemes with auxiliary inputs. In: Micciancio, D. (ed.) TCC 2010. LNCS, vol. 5978, pp. 361–381. Springer, Heidelberg (Feb 2010)
16. Döttling, N., Müller-Quade, J.: Lossy codes and a new variant of the learning-with-errors problem. In: Johansson, T., Nguyen, P.Q. (eds.) EUROCRYPT 2013. LNCS, vol. 7881, pp. 18–34. Springer, Heidelberg (May 2013).
17. Ducas, L., Durmus, A., Lepoint, T., Lyubashevsky, V.: Lattice signatures and bimodal Gaussians. In: Canetti, R., Garay, J.A. (eds.) CRYPTO 2013, Part I. LNCS, vol. 8042, pp. 40–56. Springer, Heidelberg (Aug 2013).
18. Ducas, L., Lepoint, T.: BLISS: Bimodal lattice signature schemes (Jun 2013), http://bliss.di.ens.fr/bliss-06-13-2013.zip, (proof-of-concept implementation)
19. Espitau, T., Fouque, P.A., Gérard, B., Tibouchi, M.: Side-channel attacks on BLISS lattice-based signatures: Exploiting branch tracing against strongSwan and electromagnetic emanations in microcontrollers. In: Thuraisingham, B.M., Evans, D., Malkin, T., Xu, D. (eds.) ACM CCS 17. pp. 1857–1874. ACM Press (Oct / Nov 2017)
20. Galbraith, S.D.: Space-efficient variants of cryptosystems based on learning with errors. On-Line (2012), https://www.math.auckland.ac.nz/~sgal018/compact-LWE.pdf
21. Goldwasser, S., Kalai, Y.T., Peikert, C., Vaikuntanathan, V.: Robustness of the learning with errors assumption. In: Yao, A.C.C. (ed.) ICS 2010. pp. 230–240. Tsinghua University Press (Jan 2010)
22. Hamburg, M.: Post-quantum cryptography proposal: ThreeBears (2017), https://csrc.nist.gov/projects/post-quantum-cryptography/round-1-submissions
23. Herold, G., May, A.: LP solutions of vectorial integer subset sums — cryptanalysis of Galbraith's binary matrix LWE. In: Fehr, S. (ed.) PKC 2017, Part I. LNCS, vol. 10174, pp. 3–15. Springer, Heidelberg (Mar 2017)
24. Hoffstein, J., Pipher, J., Whyte, W., Zhang, Z.: A signature scheme from learning with truncation. Cryptology ePrint Archive, Report 2017/995 (2017), http://eprint.iacr.org/2017/995
25. Howgrave-Graham, N., Szydlo, M.: A method to solve cyclotomic norm equations. In: Buell, D.A. (ed.) ANTS. LNCS, vol. 3076, pp. 272–279. Springer (2004)
26. Hsu, D., Kakade, S., Zhang, T.: Tail inequalities for sums of random matrices that depend on the intrinsic dimension. Electron. Commun. Probab.
27. Kahane, J.P.: Propriétés locales des fonctions à séries de Fourier aléatoires. Stu. Math. **19**, 1–25 (1960)
28. Kocher, P.C., Jaffe, J., Jun, B., Rohatgi, P.: Introduction to differential power analysis. J. Cryptographic Engineering **1**(1), 5–27 (2011)
29. Langlois, A., Stehlé, D.: Worst-case to average-case reductions for module lattices. Designs, Codes and Cryptography **75**(3), 565–599 (2015)
30. Li, H., Liu, R., Pan, Y., Xie, T.: Cryptanalysis of Compact-LWE submitted to NIST PQC project. Cryptology ePrint Archive, Report 2018/020 (2018), https://eprint.iacr.org/2018/020
31. Ling, S., Phan, D.H., Stehlé, D., Steinfeld, R.: Hardness of k-LWE and applications in traitor tracing. In: Garay, J.A., Gennaro, R. (eds.) CRYPTO 2014, Part I. LNCS, vol. 8616, pp. 315–334. Springer, Heidelberg
32. Litvak, A., Pajor, A., Rudelson, M., Tomczak-Jaegermann, N.: Smallest singular value of random matrices and geometry of random polytopes. Advances in Mathematics **195**(2), 491–523 (2005)

33. Liu, D.: Compact-LWE for lightweight public key encryption and leveled IoT authentication. In: Pierpzyk, J., Suriadi, S. (eds.) ACISP 17, Part I. LNCS, vol. 10342, p. xvi. Springer, Heidelberg (Jul 2017).
34. Liu, D., Li, N., Kim, J., Nepal, S.: Compact-LWE (2017), https://csrc.nist.gov/projects/post-quantum-cryptography/round-1-submissions
35. Lyubashevsky, V.: Fiat-Shamir with aborts: Applications to lattice and factoring-based signatures. In: Matsui, M. (ed.) ASIACRYPT 2009. LNCS, vol. 5912, pp. 598–616. Springer, Heidelberg (Dec 2009)
36. Lyubashevsky, V., Peikert, C., Regev, O.: On ideal lattices and learning with errors over rings. In: Gilbert, H. (ed.) EUROCRYPT 2010. LNCS, vol. 6110, pp. 1–23. Springer, Heidelberg (May / Jun 2010)
37. Lyubashevsky, V., Peikert, C., Regev, O.: A toolkit for ring-LWE cryptography. In: Johansson, T., Nguyen, P.Q. (eds.) EUROCRYPT 2013. LNCS, vol. 7881, pp. 35–54. Springer, Heidelberg (May 2013).
38. Micciancio, D., Peikert, C.: Hardness of SIS and LWE with small parameters. In: Canetti, R., Garay, J.A. (eds.) CRYPTO 2013, Part I. LNCS, vol. 8042, pp. 21–39. Springer, Heidelberg (Aug 2013).
39. Paouris, G., Valettas, P., Zinn, J.: Random version of Dvoretzky's theorem in ℓ_p^n. Stochastic Processes and their Applications **127**(10), 3187–3227 (2017)
40. Pessl, P., Bruinderink, L.G., Yarom, Y.: To BLISS-B or not to be: Attacking strongSwan's implementation of post-quantum signatures. In: Thuraisingham, B.M., Evans, D., Malkin, T., Xu, D. (eds.) ACM CCS 17. pp. 1843–1855. ACM Press (Oct / Nov 2017)
41. Pöppelmann, T., Oder, T., Güneysu, T.: High-performance ideal lattice-based cryptography on 8-bit ATxmega microcontrollers. In: Lauter, K.E., Rodríguez-Henríquez, F. (eds.) LATINCRYPT 2015. LNCS, vol. 9230, pp. 346–365. Springer, Heidelberg (Aug 2015).
42. Ramon van Handel: Probability in high dimension. Tech. rep., Princeton University (2014)
43. Regev, O.: On lattices, learning with errors, random linear codes, and cryptography. In: Gabow, H.N., Fagin, R. (eds.) 37th ACM STOC. pp. 84–93. ACM Press (May 2005)
44. Stadje, W.: An inequality for ℓ_p-norms with respect to the multivariate normal distribution. Journal of Mathematical Analysis and Applications **102**(1), 149 – 155 (1984)
45. Steffen, A., et al.: strongSwan: the open source IPsec-based VPN solution (version 5.5.2) (Mar 2017), https://www.strongswan.org/
46. Stehlé, D., Steinfeld, R., Tanaka, K., Xagawa, K.: Efficient public key encryption based on ideal lattices. In: Matsui, M. (ed.) ASIACRYPT 2009. LNCS, vol. 5912, pp. 617–635. Springer, Heidelberg (Dec 2009)
47. Tropp, J.A.: User-friendly tail bounds for sums of random matrices. Foundations of Computational Mathematics **12**(4), 389–434 (Aug 2012).
48. Xiao, D., Yu, Y.: Cryptanalysis of Compact-LWE and related lightweight public key encryption. Security and Communication Networks **2018** (2018)