Entropic Hardness of Module-LWE from Module-NTRU

Katharina Boudgoust[1], Corentin Jeudy[2,3], Adeline Roux-Langlois[2], and Weiqiang Wen[4]

katharina.boudgoust@cs.au.dk, corentin.jeudy@irisa.fr, adeline.roux-langlois@irisa.fr, weiqiang.wen@telecom-paris.fr

[1] Dept. Computer Science, Aarhus University, Aarhus, Denmark
[2] Univ Rennes, CNRS, IRISA, Rennes, France
[3] Orange Labs, Applied Crypto Group, Cesson-Sévigné, France
[4] LTCI, Telecom Paris, Institut Polytechnique de Paris, Paris, France

Abstract. The Module Learning With Errors problem (M-LWE) has gained popularity in recent years for its security-efficiency balance, and its hardness has been established for a number of variants. In this paper, we focus on proving the hardness of (search) M-LWE for general secret distributions, provided they carry sufficient min-entropy. This is called entropic hardness of M-LWE. First, we adapt the line of proof of Brakerski and Döttling on R-LWE (TCC'20) to prove that the existence of certain distributions implies the entropic hardness of M-LWE. Then, we provide one such distribution whose required properties rely on the hardness of the decisional Module-NTRU problem.

Keywords: Lattice-Based Cryptography · Module Learning With Errors · Entropic Hardness · Module-NTRU

1 Introduction

The *Learning With Errors* (LWE) [Reg05] and NTRU [HPS98] problems are the most widespread computational assumptions for designing lattice-based cryptosystems. The LWE problem asks to find a secret $\mathbf{s} \in \mathbb{Z}_q^d$ given the noisy system $(\mathbf{A}, \mathbf{b} = \mathbf{A}\mathbf{s} + \mathbf{e} \bmod q)$ for $\mathbf{A} \in \mathbb{Z}_q^{m \times d}$ uniformly random and $\mathbf{e}$ drawn from ψ^m, where ψ is an error distribution over $\mathbb{Z}$ (or $\mathbb{R}$). In the decisional variant, one has to distinguish such $\mathbf{b}$ from a uniform vector $\mathbf{u}$ over $\mathbb{Z}_q$ (or $\mathbb{T}_q = \mathbb{R}/q\mathbb{Z}$). Although the error distribution can be arbitrary, most theoretical proofs use a Gaussian distribution. LWE benefits from strong hardness guarantees, as its average-case formulation is proven to be at least as hard as worst-case lattice problems. The parameter d is known as the LWE dimension, and is often seen as the security parameter of the problem as it is linked to the dimension of the underlying lattice. The NTRU problem is defined over a more algebraic setting, using algebraic integers instead of rational ones. In a number field K, consider R to be the ring of algebraic integers in K. (Decisional) NTRU asks to distinguish between $gf_q^{-1} \bmod q$ and u, where f and g are short elements of $R_q = R/qR$, f_q^{-1} is the R_q-inverse of f and u is uniform over R_q. The search version, consisting in finding f and g given $gf_q^{-1} \bmod q$, has recently be linked to standard ideal lattice problems [PS21]. Although very few theoretical hardness results are known for NTRU, it has been widely studied for more than two decades from a cryptanalytic standpoint. Unless for overstretched parameter sets [ABD16,CJL16,KF17], it is believed to be a reliable hardness assumption to design public-key cryptosystems, as is currently done by two candidates [CDH$^+$20,BBC$^+$20] in the NIST competition [NIS].

Although LWE allows for designing provably secure cryptosystems, the resulting schemes are usually not practical enough to be used in real-world systems. With this perspective in mind, several algebraically structured variants of LWE [SSTX09,LPR10,BGV12,LS15] were introduced to improve efficiency in terms of computation and storage, while maintaining strong enough hardness guarantees from problems over structured lattices. The underlying algebraic framework is the same as the one in the NTRU problem. In this paper, we focus on the *Module Learning With Errors* (M-LWE) problem which is similar to the original LWE formulation. The set of integers $\mathbb{Z}$ is replaced by a ring of algebraic integers R, as above. The problem is formulated over the free R-module R^d, where d is called the rank of the module. Then, for a modulus q, and an error distribution ψ over R (or $K_\mathbb{R} = K \otimes_\mathbb{Q} \mathbb{R}$), and given $(\mathbf{A}, \mathbf{b} = \mathbf{A}\mathbf{s} + \mathbf{e} \mod q)$ with $\mathbf{A}$ uniform in $R_q^{m \times d}$, $\mathbf{s} \in R_q^d$ and $\mathbf{e}$ drawn from ψ^m, the goal is to recover $\mathbf{s}$. The decisional version asks to distinguish such a $\mathbf{b}$ from a uniformly distributed vector $\mathbf{u}$. When $d = 1$, we call it *Ring* LWE (R-LWE). A module version of NTRU was also recently studied by Chuengsatiansup et al. [CPS$^+$20]. It consists in distinguishing $\mathbf{G}\mathbf{F}_q^{-1} \bmod q$, with $(\mathbf{F}, \mathbf{G})$ short matrices in $R^{d \times d} \times R^{m \times d}$ and $\mathbf{F}_q^{-1}$ the R_q-inverse[5] of $\mathbf{F}$, from a uniform matrix $\mathbf{U} \in R_q^{m \times d}$. The module versions M-LWE and M-NTRU give more flexibility to adjust the balance between efficiency and security.

The average-case formulation of M-LWE is parameterized by a distribution $\mathcal{S}$ on the secret $\mathbf{s}$. The original definition uses $\mathcal{S}$ to be the uniform distribution over R_q^d. Even though this choice is the most natural one for theoretical results, practical schemes vary from it. For efficiency reasons, it is advantageous to choose secret distributions that lead to small-norm secrets, as is done in the M-LWE-based schemes Kyber [BDK$^+$18] and Dilithium [DKL$^+$18] currently participating in the NIST competition. The study of LWE with small-norm secrets [GKPV10,BLP$^+$13,Mic18] has recently been extended to the module setting by Boudgoust et al. [BJRW20,BJRW21]. The second reason to deviate from the uniform distribution stems from the situation where the key is not sampled from the prescribed distribution but from an imperfect one. Braskerski and Döttling conducted a study of LWE with general entropic secret distributions [BD20a], which they afterwards extended to the ring setting ($d = 1$) [BD20b]. It was left open to thoroughly generalize this proof method

[5] As we use both the R_q-inverse and the K-inverse, we insist on differentiating them as $\mathbf{F}_q^{-1}$ and $\mathbf{F}^{-1}$ respectively.

to larger ranks $d > 1$. Hardness results encapsulating imperfect distributions provide theoretical insights on the resistance to key leakage, like cold boot attacks for instance. These attacks leverage the physical properties of the hardware to recover remanent information from the memory. In the representative cold boot attack on M-LWE based schemes [ADP18], the adversary can manage to obtain a faulty version of the (long-term) secret key $\mathbf{s}$ stored in memory. From the faulty key $\widetilde{\mathbf{s}} = \mathbf{s} + \Delta\mathbf{s}$, the adversary can now recover the full secret $\mathbf{s}$ by targeting a new M-LWE instance $(\mathbf{A}, \mathbf{A}\widetilde{\mathbf{s}} - \mathbf{b} \bmod q) = (\mathbf{A}, \mathbf{A}\Delta\mathbf{s} - \mathbf{e} \bmod q)$, where the new secret $\Delta\mathbf{s}$ is only promised to have certain entropy. The motivation is therefore to prove the hardness of M-LWE when some secret key information is leaked to the attacker, i.e., the remaining entropy in $\Delta\mathbf{s}$ is smaller than that of $\mathbf{s}$.

Our contributions. In this paper, we extend the line of work of Brakerski and Döttling on the entropic hardness to the module setting, i.e., $d \geq 1$, which has gained popularity over its preceding variants. Our main contribution is given in the following informal theorem. For a complete statement, refer to Theorem 5.2.

Theorem 1.1 (Informal). *Assuming decisional* M-NTRU *is hard, the problem search* M-LWE *with secret distribution $\mathcal{S}$ is hard for a sufficiently large Gaussian noise and provided that $\mathcal{S}$ has large enough min-entropy.*

The first step of our proof is to translate the M-NTRU problem from its algebraic form to a non-algebraic one, which we call Structured NTRU (S-NTRU). This simply consists in embedding the algebraic setting into vectors and matrices over the integers or the reals. We similarly define Structured LWE (S-LWE), generalizing the notion given in [BD20b]. The core of our work is then to prove the hardness of S-LWE from that of S-NTRU.

To do so, we construct a *sometimes lossy pseudorandom distribution*, as introduced in [BD20b]. The pseudorandomness property, proven under the hardness of S-NTRU, essentially allows us to replace the uniform matrix $\mathbf{A}$ from the S-LWE instance by the S-NTRU matrix $\mathbf{G}\mathbf{F}_q^{-1} \bmod q$, where $\mathbf{F}_q^{-1}$ is the R_q-inverse of $\mathbf{F}$. Then, the sometimes lossiness property translates the fact that going from $\mathbf{s}$ to $\mathbf{G}\mathbf{F}_q^{-1}\mathbf{s} + \mathbf{e} \bmod q$ loses enough information on $\mathbf{s}$ so that it is hard to recover it. The sometimes lossiness thus entails the hardness of S-LWE when $\mathbf{A} = \mathbf{G}\mathbf{F}_q^{-1} \bmod q$. By the pseudorandomness, it then yields the hardness of (search) S-LWE for a uniform $\mathbf{A}$. Section 4 gives explicit conditions on the distributions of $\mathbf{F}$ and $\mathbf{G}$ for the sometimes lossiness property to be satisfied. It requires $\mathbf{F}$ and $\mathbf{G}$ to be somewhat well-behaved in terms of invertibility and spectral bounds. More precisely, $\mathbf{F}$ must be invertible in R_q but also in K. Additionally, the distributions on $\mathbf{F}$ and $\mathbf{G}$ should both minimize their largest singular value, and the distribution of $\mathbf{F}$ should also maximize its smallest singular value. These conditions are obtained by generalizing the approach from [BD20b] in the case of rings ($d = 1$). It consists in introducing the extra dimension d associated to the module rank. This has the effect of trading ring elements for matrices and vectors of ring elements, but the theoretical arguments remain valid.

We then exhibit in Section 5 a distribution on $\mathbf{F}$ and $\mathbf{G}$ that satisfies the required conditions, which is our main contribution. The invertibility in R_q has been studied in [CPS+20], and we show that the invertibility in K is achieved with overwhelming probability in our context. The spectral analysis is more tricky and is the object of Sections 5.1 and 5.2. We give technical results on the smallest (and largest) singular values of discrete Gaussian matrices over the ring of integers of a number field, which might be of independent interest. In addition to dealing with matrices, our techniques differ from that of [BD20b] as we adopt a different, and inherently simpler, distribution on $\mathbf{F}$ and $\mathbf{G}$. This provides another valid way to construct sometimes lossy pseudorandom distributions from module-based assumptions. For our distribution, call it *way 1*, the pseudorandomness property is obtained by assuming that $\mathbf{G}\mathbf{F}_q^{-1} \bmod q \in R_q^{m \times d}$ is indistinguishable from uniform, hence the M-NTRU assumption for a rectangular matrix $\mathbf{G}$. The approach in [BD20b], call it *way 2*, is adapted to the module setting in Appendix A. It argues that $\tilde{\mathbf{G}}\mathbf{F}_q^{-1} \bmod q$ is indistinguishable from uniform, where $\tilde{\mathbf{G}}$ is obtained by randomizing square matrices $\mathbf{F}, \mathbf{G}$. More concretely, $\tilde{\mathbf{G}} = \mathbf{E}\mathbf{G} + \mathbf{E}'\mathbf{F}$ with $\mathbf{E}, \mathbf{E}'$ random matrices of size $m \times d$. Then, $\tilde{\mathbf{G}}\mathbf{F}_q^{-1} = \mathbf{E}\mathbf{G}\mathbf{F}_q^{-1} + \mathbf{E}'$. Arguing that $\tilde{\mathbf{G}}\mathbf{F}_q^{-1} \bmod q \approx \mathbf{A}$ now requires to first replace the square matrix $\mathbf{G}\mathbf{F}_q^{-1}$ by a uniform matrix $\mathbf{U}$ (M-NTRU for a square matrix $\mathbf{G}$) and then to replace $\mathbf{E}\mathbf{U} + \mathbf{E}'$ by $\mathbf{A}$ (Hermite Normal Form M-LWE[6]). The high-level picture of the entire proof is summarized in Figure 1.1.

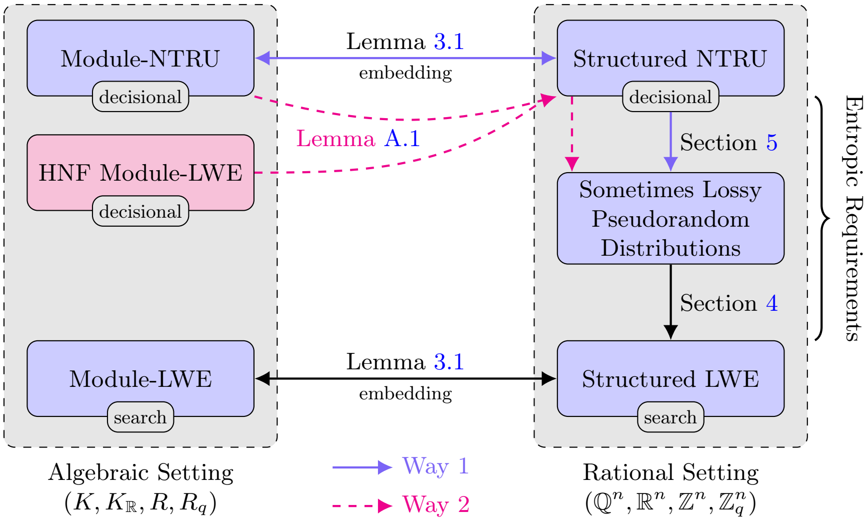

Fig. 1.1. High-level summary of the contributions of this work, proving Theorem 1.1.

[6] The Hermite Normal Form of M-LWE denotes a specific class of instances where the secret is chosen from the same distribution as the error, i.e., $\mathcal{S} = \psi^d \bmod q$.

One advantage of removing the HNF-M-LWE assumption is that our proof can be seen as a reduction from decisional M-NTRU to search M-LWE. This generalizes the observation of Peikert [Pei16] that decisional NTRU reduces to search R-LWE. Another upside in only assuming M-NTRU is that we directly recover the expected statistical hardness of M-LWE from that of M-NTRU whenever the parameters are large enough. We discuss briefly this statistical result in Section 5.3 but note that it seems hard to build cryptosystems purely based on the statistical hardness of M-LWE in these parameter regimes. Our reduction is also rank-preserving between M-NTRU and M-LWE. Although it prevents us from reaching unusually short secrets, it can still lead to better parameters compared to (close to) rank-preserving reductions based on M-LWE assumptions, as discussed in Section 6. As explained, we rely on a rectangular formulation of M-NTRU which is similar to the multiple public keys version in the case of NTRU. For non-overstretched parameters, we have no reason to believe that this multiple public keys version, or rectangular M-NTRU in our case, is a substantially weaker assumption than square M-NTRU.

Lin et al. [LWW20] recently adapted the proof method from [BD20a] on LWE to modules, which uses a sensibly different approach from the one we described above. Their proof is based on an M-LWE hardness assumption, while it is based on M-NTRU for us. This assumption allows them to tweak the starting rank k of the module to reach smaller or larger secrets and noise, depending on which one to optimize. However, their result does not provide a rank-preserving reduction as $k < d$. Additionally, when k is close to d, our proof can lead to better parameters, at the expense of trading the underlying assumption. We discuss their result and compare it to ours in more details in Section 6.

Open Questions. Our proof only shows the entropic hardness of search M-LWE, and we leave it as an open problem to extend it to the decisional variant. One possibility (of more general interest) would be to find a search-to-decision reduction for M-LWE that preserves the (non-uniform) secret distribution. Additionally, it would be interesting to have a worst-case to average-case reduction from module lattice problems to decision M-NTRU in order to gain confidence in this rather new assumption.

Organization. In Section 2, we introduce the notions and results that are needed in our proof. In Section 3, we provide an equivalent formulation of M-LWE and M-NTRU called *Structured LWE* and *Structured NTRU*, which generalize the notions defined in [BD20b]. In Section 4, we adapt the line of proof from [BD20b] to our more general setting to give sufficient conditions for Structured LWE to be (mildly) hard. Section 5 is then dedicated to instantianting this hardness result for M-LWE parameters. Finally, in Section 6, we discuss in more details how our contribution places itself in the landscape of existing entropic hardness results of M-LWE.

2 Preliminaries

In this paper, q denotes a positive integer, $\mathbb{Z}$ the set of integers, and $\mathbb{Z}_q = \mathbb{Z}/q\mathbb{Z}$ the set of integers reduced modulo q. For a positive integer n, we use $[n]$ to denote $\{1, \ldots, n\}$. The vectors are written in bold lowercase letters $\mathbf{a}$ while the matrices are in bold uppercase letters $\mathbf{A}$. The transpose and Hermitian operators for vectors and matrices are denoted with the superscript T and $\dagger$ respectively. The identity matrix of size $n \times n$ is denoted by $\mathbf{I}_n$. For a vector $\mathbf{a} \in \mathbb{C}^n$, we define its Euclidean norm as $\|\mathbf{a}\|_2 = (\sum_{i\in[n]} |a_i|^2)^{1/2}$, and its infinity norm as $\|\mathbf{a}\|_\infty = \max_{i\in[n]} |a_i|$. Further, we denote by $\operatorname{diag}(\mathbf{a})$ the diagonal matrix whose diagonal entries are the entries of $\mathbf{a}$. For a matrix $\mathbf{A} = [a_{ij}]_{i\in[n], j\in[m]} \in \mathbb{C}^{n\times m}$, we define its spectral norm as $\|\mathbf{A}\|_2 = \max_{\mathbf{x}\in\mathbb{C}^m\setminus\{\mathbf{0}\}} \|\mathbf{A}\mathbf{x}\|_2 / \|\mathbf{x}\|_2$, which corresponds to its *largest singular value*. The *smallest singular value* of $\mathbf{A}$ is denoted by $s_{\min}(\mathbf{A})$. For a ring R and integers k, q, we denote $GL_k(R, q)$ the set of matrices of $R^{k\times k}$ that are invertible in $R_q = R/qR$. The uniform distribution over a finite set S is denoted by $U(S)$, and the statistical distance between two discrete distributions P and Q over a countable set S is defined as $\Delta(P, Q) = \frac{1}{2}\sum_{x\in S} |P(x) - Q(x)|$. Finally, the action of sampling $x \in S$ from a distribution P is denoted by $x \hookleftarrow P$.

2.1 Algebraic number theory

Although we focus on the problem once embedded into the integers, we provide here the minimal background to understand the underlying ring structure. An *algebraic number* ζ is a root of a polynomial over $\mathbb{Q}$. The monic polynomial f of minimal degree such that $f(\zeta) = 0$ is called the *minimal polynomial* of ζ and is unique. If f has coefficients in $\mathbb{Z}$, then ζ is called an *algebraic integer*. A number field $K = \mathbb{Q}(\zeta)$ is a finite field extension of $\mathbb{Q}$ by adjoining the algebraic number ζ. The degree of the field is that of f, i.e., $[K : \mathbb{Q}] = \deg(f)$. We denote by R the ring of algebraic integers in K. We also define the tensor field $K_\mathbb{R} = K \otimes_\mathbb{Q} \mathbb{R}$. Finally, we define $\mathbb{T}_q = K_\mathbb{R}/qR$. When concretely instantiating the results in Section 5, we need to restrict the choice of fields to cyclotomic fields, or even a subclass of them. A cyclotomic field is a number field $K = \mathbb{Q}(\zeta)$, where ζ is a primitive ν-th root of unity, for an integer ν. The minimal polynomial is then $f = \Phi_\nu = \prod_{j\in[n]}(x - \alpha_j)$, where the α_j are the primitive ν-th roots of unity. Its degree is $n = \varphi(\nu)$, where φ is Euler's totient function. A common subclass is the one composed of power-of-two cyclotomic fields, which consists in having $\nu = 2^{\ell+1}$, in which case $n = 2^\ell$, and $f = x^n + 1$.

Space H. We use t_1 to denote the number of real roots of f, and t_2 the number of pairs of complex conjugate roots. We then have $n = t_1 + 2t_2$. For cyclotomic fields, f has no real roots and therefore $t_1 = 0$ and $t_2 = n/2$. The space H is then defined by $H = \{\mathbf{x} \in \mathbb{R}^{t_1} \times \mathbb{C}^{2t_2} \colon \forall j \in [t_2], x_{t_1+t_2+j} = \overline{x_{t_1+j}}\}$. H turns out to be a real vector space of dimension n with the following orthonormal basis

$$\mathbf{U}_H = \frac{1}{\sqrt{2}} \begin{bmatrix} \sqrt{2}\mathbf{I}_{t_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{t_2} & i\mathbf{I}_{t_2} \\ \mathbf{0} & \mathbf{I}_{t_2} & -i\mathbf{I}_{t_2} \end{bmatrix}.$$

We can therefore define the isomorphism $\Theta : \mathbf{x} \in H \mapsto \mathbf{U}_H^\dagger \mathbf{x} \in \mathbb{R}^n$ between H and $\mathbb{R}^n$.

Embeddings. A number field $K = \mathbb{Q}(\zeta)$ of degree n can be seen as a vector space of dimension n over the rationals with basis $\{1, \zeta, \ldots, \zeta^{n-1}\}$, meaning that each element $x \in K$ can be written as $x = \sum_{0 \leq j \leq n-1} x_j \zeta^j$ with $x_j \in \mathbb{Q}$. The *coefficient embedding* is the isomorphism τ between K and $\mathbb{Q}^n$ that maps every $x \in K$ to its coefficient vector $\tau(x) = [x_0, \ldots, x_{n-1}]^T$. For simplicity, we use $\tau_k(x)$ to denote x_k. Additionally, τ can be extended to $K_\mathbb{R}$, mapping it to $\mathbb{R}^n$.

Another way to embed K is to use the *canonical embedding.* K has exactly n field homomorphisms $\sigma_1, \ldots, \sigma_n$, which are characterized by the fact that they map ζ to one of the distinct roots of f. We order them so that $\sigma_1, \ldots, \sigma_{t_1}$ map to one of the real roots, and $\sigma_{t_1+1}, \ldots, \sigma_{t_1+2t_2}$ map to one of the complex roots. The *canonical embedding* is defined as the field homomorphism from K to $\mathbb{C}^n$ defined by $\sigma(x) = [\sigma_1(x), \ldots, \sigma_n(x)]^T$, and the addition and multiplication is done component-wise. As f has rational coefficients, it holds that the complex embeddings come in conjugate pairs, and therefore the range of σ is a subset of H. Using Θ, we can therefore map K to $\mathbb{R}^n$ with $\sigma_H = \Theta \circ \sigma$.

The two embeddings are linked by the linear relation

$$\sigma(x) = \mathbf{V}\tau(x) \text{ for all } x \in K, \text{ where } \mathbf{V} = \begin{bmatrix} 1 & \alpha_1 & - & \alpha_1^{n-1} \\ 1 & \alpha_2 & - & \alpha_2^{n-1} \\ | & | & & | \\ 1 & \alpha_n & - & \alpha_n^{n-1} \end{bmatrix},$$

is the Vandermonde matrix defined by the roots of f. We define the field trace of K using the canonical embedding σ as $\mathrm{Tr}(x) = \sum_{k \in [n]} \sigma_k(x)$ for all $x \in K$. The discriminant of K denoted by Δ_K is defined by $\det\left([\mathrm{Tr}(b_i \cdot b_j)]_{(i,j) \in [n]}\right)$, where $(b_i)_i$ is a $\mathbb{Z}$-basis of R. In fields where the power basis $1, \zeta, \ldots, \zeta^{n-1}$ is such a basis, we simply have $\Delta_K = \det(\mathbf{V})^2$. In cyclotomic fields, we have the bound $\Delta_K \leq n^n$. From the trace, we can also define the dual of R by $R^\vee = \{x \in K : \mathrm{Tr}(xR) \subseteq \mathbb{Z}\}$. Further, we define the field norm as $N(x) = \prod_{k \in [n]} \sigma_k(x)$ for all $x \in K$. The canonical embedding, as well as the trace and norm, can be extended to $K_\mathbb{R}$ too.

Multiplication matrices. The multiplication in K (or $K_\mathbb{R}$) translates into a matrix-vector multiplication once embedded with either τ or σ. In the canonical embedding, the multiplication matrix can be easily expressed. For all x and a in K, we have $\sigma(x \cdot a) = \sigma(x) \odot \sigma(a) = \mathrm{diag}(\sigma(x)) \cdot \sigma(a)$, where $\odot$ denotes the coefficient-wise product or Hadamard product. Therefore, the multiplication matrix is

$$M_\sigma(x) = \mathrm{diag}(\sigma(x)).$$

We can then express the multiplication matrix with respect to σ_H as

$$M_{\sigma_H}(x) = \mathbf{U}_H^\dagger M_\sigma(x) \mathbf{U}_H = \mathbf{U}_H^\dagger \mathrm{diag}(\sigma(x)) \mathbf{U}_H.$$

In the coefficient embedding, we can still express $\tau(x \cdot a)$ as $M_\tau(x) \cdot \tau(a)$, but the expression of $M_\tau(x)$ is more involved as we have

$$M_\tau(x) = \sum_{k=0}^{n-1} \tau_k(x)\mathbf{C}^k, \text{ with } \mathbf{C} = \begin{bmatrix} 0 & \cdots & 0 & -f_0 \\ & & & -f_1 \\ & \mathbf{I}_{n-1} & & \vdots \\ & & & -f_{n-1} \end{bmatrix} \tag{1}$$

the companion matrix of the minimal polynomial $f = \sum_{k=0}^{n-1} f_k x^k + x^n$. We did not find a reference for this identity. Although it is a straightforward calculation relying on properties of companion matrices, we provide a proof in Appendix B.2 for completeness. We extend the embeddings to vectors in K^d in the natural way by concatenating the embedding vectors of each coefficient, i.e., $\tau(\mathbf{x}) = [\tau(x_1)^T, \ldots, \tau(x_d)^T]^T$ and similarly for σ. We can therefore translate the matrix-vector multiplication in K^d to a matrix-vector multiplication in $\mathbb{R}^{nd}$ by extending the multiplication matrix maps M_σ, M_{σ_H} and M_τ to a matrix in $K^{m \times d}$. More precisely, for a matrix $\mathbf{F} = [f_{ij}]_{(i,j)} \in K^{m \times d}$, we define the block matrix $M_\sigma(\mathbf{F}) = [M_\sigma(f_{ij})]_{(i,j)}$. We define $M_{\sigma_H}(\mathbf{F})$ and $M_\tau(\mathbf{F})$ the same way.

Ideals. An *ideal* $\mathcal{I} \subseteq R$ is an additive subgroup of R that is closed under multiplication by R. An ideal $\mathfrak{p}$ is prime if for all $a, b \in R$, $ab \in \mathfrak{p}$ implies that either $a \in \mathfrak{p}$ or $b \in \mathfrak{p}$. Any ideal of R can be factored into a product of prime ideals. The integer q is said unramified if the prime factors of qR are all distinct. The field norm can also be extended to an ideal $\mathcal{I}$ of R by $N(\mathcal{I}) = |R/\mathcal{I}|$.

2.2 Lattices

A *lattice* is a discrete subgroup of $\mathbb{R}^n$. Each lattice Λ can be represented by a basis $\mathbf{B} = [\mathbf{b}_i]_{i \in [r]} \in \mathbb{R}^{n \times r}$ as the set of its integer linear combinations, i.e., $\Lambda = \sum_{i \in [r]} \mathbb{Z} \cdot \mathbf{b}_i$. We use $\Lambda(\mathbf{B})$ to denote the lattice generated by the basis $\mathbf{B}$. The parameter r is called the rank of the lattice, but in this work we only consider *full-rank* lattices where the rank matches the dimension n. The *dual lattice* of a lattice Λ is defined by $\Lambda^* = \{\mathbf{x} \in \mathrm{Span}(\Lambda) : \forall \mathbf{y} \in \Lambda, \langle \mathbf{x}, \mathbf{y} \rangle \in \mathbb{Z}\}$.

Ideal lattices. Any ideal $\mathcal{I}$ of R embeds into a lattice of $\mathbb{R}^n$ via τ or σ_H. These lattices are called *ideal lattices*. In the rest of the paper, we denote by $\mathbf{L}_R$ a basis of the lattice $\sigma_H(R)$, and $\mathbf{B}_R = \mathbf{L}_R^{-1}$. Since each lattice element can be represented as $\mathbf{L}_R\mathbf{x}$ for some $\mathbf{x} \in \mathbb{Z}^n$, we can map R to $\mathbb{Z}^n$ by $\mathbf{B}_R \circ \sigma_H$. We denote by $M_R(x)$ the associated multiplication matrix, i.e.,

$$M_R(x) = \mathbf{B}_R M_{\sigma_H}(x) \mathbf{B}_R^{-1} = (\mathbf{B}_R \mathbf{U}_H^\dagger) \mathrm{diag}(\sigma(x)) (\mathbf{B}_R \mathbf{U}_H^\dagger)^{-1}.$$

Notice that when the power basis is a $\mathbb{Z}$-basis of R, e.g. cyclotomics, we can choose $\mathbf{L}_R = [\sigma_H(1) \mid \ldots \mid \sigma_H(\zeta^{n-1})] = \mathbf{U}_H^\dagger \mathbf{V}$. In that case, we have $\mathbf{B}_R \circ \sigma_H = \tau$ and also $M_R(\cdot) = M_\tau(\cdot)$. We keep the notations without assuming how the basis $\mathbf{L}_R$ is chosen. We only use this specific choice of basis in Section 5.

Module lattices. More generally, for an R-module $M \subseteq K^d$, M embeds into a *module lattice* of $\mathbb{R}^{nd}$ via τ or σ_H. We define $\mathbf{L}_{R^d} = \mathbf{I}_d \otimes \mathbf{L}_R$, which is a basis of the module lattice $\sigma_H(R^d)$, and we define $\mathbf{B}_{R^d} = \mathbf{L}_{R^d}^{-1} = \mathbf{I}_d \otimes \mathbf{B}_R$. We can then extend the map M_R to matrices by applying the map coefficient-wise as before. It maps (matrices of) ring elements to (block matrices of) structured matrices, as depicted in Figure 2.1. Further, just like M_τ, M_σ and M_{σ_H}, M_R is a ring homomorphism.

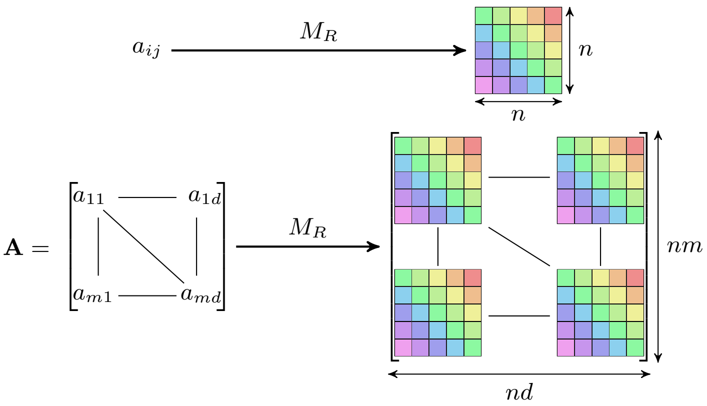

Fig. 2.1. Illustration of the (structured) multiplication matrix map M_R

2.3 Probabilities

We first recall a trivial lemma that bounds the maximal singular value of a random matrix $\mathbf{Z}$ by bounding the singular values of its *blocks*. The result can be made deterministic for $\delta = 0$. The proof is provided in Appendix B.2.

Lemma 2.1. *Let a, b, c, d be integers. Let $\mathbf{Z} = [\mathbf{Z}_{ij}]_{(i,j)\in[a]\times[b]} \in \mathbb{R}^{ac\times bd}$ be a random block matrix where each $\mathbf{Z}_{ij} \in \mathbb{R}^{c\times d}$. Assume it holds for all (i, j) that $\mathbb{P}[\|\mathbf{Z}_{ij}\|_2 \geq \gamma] \leq \delta$, for $0 \leq \delta \leq 1$. Then, we have $\mathbb{P}[\|\mathbf{Z}\|_2 \geq \sqrt{ab}\cdot\gamma] \leq ab\cdot\delta$.*

Gaussians. For a full-rank matrix $\mathbf{S} \in \mathbb{R}^{m\times n}$ ($m \geq n$), and a center $\mathbf{c} \in \mathbb{R}^n$ we define the Gaussian function by $\rho_{\mathbf{c},\mathbf{S}}(\mathbf{x}) = \exp(-\pi(\mathbf{x}-\mathbf{c})^T(\mathbf{S}^T\mathbf{S})^{-1}(\mathbf{x}-\mathbf{c}))$. We can then define the continuous Gaussian distribution $D_{\mathbf{c},\mathbf{S}}$ whose density is $\det(\mathbf{S})^{-1}\rho_{\mathbf{c},\mathbf{S}}$. If $\mathbf{S} = s\mathbf{I}_n$, then we simply write $D_{\mathbf{c},s}$, and we omit $\mathbf{c}$ if it is $\mathbf{0}$. We then define the discrete Gaussian distribution $\mathcal{D}_{\Lambda,\mathbf{c},\mathbf{S}}$ over a lattice Λ by conditioning on $\mathbf{x}$ being in the lattice, i.e., for all $\mathbf{x} \in \Lambda$, $\mathcal{D}_{\Lambda,\mathbf{c},\mathbf{S}}(\mathbf{x}) = D_{\mathbf{c},\mathbf{S}}(\mathbf{x})/D_{\mathbf{c},\mathbf{S}}(\Lambda)$, where $D_{\mathbf{c},\mathbf{S}}(\Lambda) = \sum_{\mathbf{y}\in\Lambda} D_{\mathbf{c},\mathbf{S}}(\mathbf{y})$. We use the simplified notation $\mathcal{D}_{R,\mathbf{c},\mathbf{S}}$ to denote $\mathcal{D}_{\sigma_H(R),\mathbf{c},\mathbf{S}}$. Additionally, by abuse of notation, we also use $D_{\mathbf{c},\mathbf{S}}$ to denote the distribution obtained by sampling $\mathbf{x}$ from $D_{\mathbf{c},\mathbf{S}}$ and outputting $\sigma_H^{-1}(\mathbf{x}) \in K_{\mathbb{R}}$.

Definition 2.1 (Sub-Gaussian Distribution). *Let n be a positive integer, and $\mathbf{x}$ a (discrete or continuous) random vector over $\mathbb{R}^n$. We say that $\mathbf{x}$ is sub-Gaussian with sub-Gaussian moment s, if for all unit vector $\mathbf{u} \in \mathbb{R}^n$, and all $t \in \mathbb{R}$, we have $\mathbb{E}[\exp(2\pi t\langle \mathbf{x}, \mathbf{u}\rangle)] \leq e^{\pi s^2 t^2}$.*

A standard calculation shows that the discrete Gaussian distribution $\mathcal{D}_{\Lambda,s}$ is sub-Gaussian with sub-Gaussian moment s [MP12, Lem. 2.8], for any lattice Λ and $s > 0$.

Smoothing. The *smoothing parameter* of a lattice Λ introduced by Micciancio and Regev [MR07], and denoted by $\eta_\varepsilon(\Lambda)$ for some $\varepsilon > 0$, is the smallest $s > 0$ such that $\rho_{1/s}(\Lambda^* \setminus \{\mathbf{0}\}) \leq \varepsilon$. It represents the threshold above which the discrete Gaussian distribution behaves like a continuous one.

Min-Entropy. Let $\mathbf{x}$ follow a discrete distribution on a set X, and $\mathbf{z}$ follow a (possibly continuous) distribution on a (measurable) set Z. The *average conditional min-entropy*[7] of $\mathbf{x}$ given $\mathbf{z}$ is defined by

$$\widetilde{H}_\infty(\mathbf{x}|\mathbf{z}) = -\log_2\left(\mathbb{E}_{\mathbf{z}'}\left[\max_{\mathbf{x}' \in X} \mathbb{P}[\mathbf{x} = \mathbf{x}' | \mathbf{z} = \mathbf{z}']\right]\right)$$

If $\mathbf{z}$ is deterministic, we obtain the definition of the *min-entropy* of $\mathbf{x}$ as $H_\infty(\mathbf{x}) = -\log_2(\max_{\mathbf{x}' \in X} \mathbb{P}[\mathbf{x} = \mathbf{x}'])$. For $\varepsilon > 0$, we also define the ε-smooth average conditional min-entropy by

$$\widetilde{H}_\infty^\varepsilon(\mathbf{x}|\mathbf{z}) = \max\{\widetilde{H}_\infty(\mathbf{x}'|\mathbf{z}') : \Delta((\mathbf{x}', \mathbf{z}'), (\mathbf{x}, \mathbf{z})) \leq \varepsilon\}.$$

For convenience, we simply refer to all these notions as min-entropy instead of their full name when it is clear from the context or notations. But it should be noted that the notions are distinct.

2.4 Noise Lossiness

We now recall the notion of *noise lossiness* of a distribution $\mathcal{S}$ of secrets as introduced in [BD20a,BD20b]. It quantifies how much information is lost about a secret from $\mathcal{S}$ when perturbed by a Gaussian noise. As we are in the module setting, we highlight the dimension as nd where n is the ring degree, and d the module rank.

Definition 2.2 (Noise Lossiness). *Let n, d, q be integers and $s > 0$ be a Gaussian parameter. Let $\mathbf{B}$ be a non-singular matrix in $\mathbb{R}^{nd \times nd}$. Let $\mathcal{S}$ be a distribution of secrets over $\mathbb{Z}_q^{nd}$. The noise lossiness $\nu_{s\mathbf{B}}(\mathcal{S})$ is defined by*

$$\nu_{s\mathbf{B}}(\mathcal{S}) = \widetilde{H}_\infty(\mathbf{s}|\mathbf{s} + \mathbf{e}),$$

where $\mathbf{s} \hookleftarrow \mathcal{S}$, and $\mathbf{e} \hookleftarrow D_{s\mathbf{B}}$.

[7] The (non-average) conditional min-entropy of $\mathbf{x}$ given $\mathbf{z}$ is denoted by $H_\infty(\mathbf{x}|\mathbf{z})$ instead of $\widetilde{H}_\infty(\mathbf{x}|\mathbf{z})$, and given by $H_\infty(\mathbf{x}|\mathbf{z}) = -\log_2(\max_{\mathbf{z}' \in Z} \max_{\mathbf{x}' \in X} \mathbb{P}[\mathbf{x} = \mathbf{x}' | \mathbf{z} = \mathbf{z}'])$.

We also recall the bounds on the noise lossiness derived in [BD20a] in the case of general distributions, as well as that of distributions over bounded secrets.

Lemma 2.2 ([BD20a], Lem. 5.2). *Let n, d, q be integers, and a Gaussian parameter s such that $0 < s \leq q\sqrt{\pi/\ln(4nd)}$. Let $\mathcal{S}$ be any distribution over $\mathbb{Z}_q^{nd}$. Then it holds that*

$$\nu_s(\mathcal{S}) \geq H_\infty(\mathcal{S}) - nd \cdot \log_2(q/s) - 1.$$

Lemma 2.3 ([BD20a], Lem. 5.4). *Let n, d, q be integers and $s > 0$ be a Gaussian parameter. Let $\mathcal{S}$ be a r-bounded distribution (for the Euclidean norm) over $\mathbb{Z}_q^{nd}$. Then it holds that*

$$\nu_s(\mathcal{S}) \geq H_\infty(\mathcal{S}) - \sqrt{2\pi nd}\log_2(e) \cdot \tfrac{r}{s}.$$

2.5 Module Learning With Errors

In this work, we deal with *Module Learning With Errors* (M-LWE) over the (primal) ring of integers R of a number field K. Additionally, we do not limit the secret distribution to be uniform, and we thus define M-LWE for an arbitrary distribution of secrets $\mathcal{S}$.

Definition 2.3. *Let K be a number field of degree n, and R its ring of integers. Let d, q, m be positive integers. Finally, let $\mathcal{S}$ be a secret distribution supported on R_q^d, and Υ be a distribution over error distributions on $K_\mathbb{R}$. The search $\text{M-LWE}_{n,d,q,m,\Upsilon,\mathcal{S}}$ problem is to find the secret $\mathbf{s}$ given $(\mathbf{A}, \mathbf{b}) = (\mathbf{A}, \mathbf{A}\mathbf{s} + \mathbf{e} \bmod qR)$ for $\mathbf{A} \hookleftarrow U(R_q^{m\times d})$, $\mathbf{s} \hookleftarrow \mathcal{S}$, and $\mathbf{e} \hookleftarrow \psi^m$ for $\psi \hookleftarrow \Upsilon$. The decisional version consists in deciding whether such $(\mathbf{A}, \mathbf{b})$ is distributed as above or if it is uniform over $R_q^{m\times d} \times \mathbb{T}_q^m$.*

In most cases, we consider Υ to be the deterministic distribution over $\{\psi\}$ for some (non-deterministic) error distribution ψ over $K_\mathbb{R}$. In this case, we use ψ instead of Υ in the notation. In Section 6, we compare our result with existing ones which requires to use the dual $R^\vee$ by considering $\mathbf{s} \in (R_q^\vee)^d$ and $\mathbf{e} \in (R^\vee)^m$. We may also use discrete error distributions over R instead of $K_\mathbb{R}$. The standard formulation of M-LWE corresponds to $\mathcal{S} = U(R_q^d)$, for which we omit the $\mathcal{S}$ in the notation. For arbitrary secret distributions $\mathcal{S}$, we usually analyze the hardness of the problem based on some requirement on the entropy of $\mathcal{S}$. This is why these cases are also referred to as *entropic* M-LWE or *entropic hardness* of M-LWE.

2.6 Module-NTRU

We also recall the Module-NTRU (M-NTRU) problem, as defined by Chuengsatiansup et al. [CPS+20]. Although we use discrete Gaussian distributions as in [CPS+20], it can be formulated using an arbitrary distribution ψ as follows.

Definition 2.4 (Module-NTRU (M-NTRU)). *Let R be the ring of integers of a number field K and let q be a modulus. Let m, d be positive integers, and ψ be a distribution on R. Let $\mathbf{G} \hookleftarrow \psi^{m\times d}$ and $\mathbf{F} \hookleftarrow \psi^{d\times d}$ conditioned on $\mathbf{F} \in GL_d(R, q)$. Let $\mathbf{F}_q^{-1}$ be the R_q-inverse of $\mathbf{F}$. The $\text{M-NTRU}_{n,d,q,m,\psi}$ problem asks for distinguishing $\mathbf{G}\mathbf{F}_q^{-1} \in R_q^{m\times d}$ from a uniformly random $\mathbf{A} \hookleftarrow U(R_q^{m\times d})$.*

In the rest of the paper, we consider $\psi = \mathcal{D}_{R,\gamma}$ for some Gaussian parameter $\gamma > 0$. In this case, we simply denote it as M-NTRU$_{n,d,q,m,\gamma}$.

3 Structured LWE

In this section, we provide another formulation of LWE and NTRU. We then prove that these new problems are equivalent to M-LWE and M-NTRU respectively in a specific setting. Finally, we recall the notion of *mild hardness* presented in [BD20b], and how it relates to the standard definition of hardness.

3.1 Structured LWE and Structured NTRU

In the following, we define a version of LWE that generalizes the *Structured LWE* problem from [BD20b]. We also define the *Structured NTRU* problem which is a generalization of the DSR problem from [BD20b]. We use the name Structured NTRU only because it is to M-NTRU what Structured LWE is to M-LWE as we explain in Section 3.2. NTRU already being a structured problem, it should not be interpreted as if there exists an *unstructured* version of NTRU.

The only difference stems in introducing the extra dimension of the module rank d. Instead of considering vectors of m matrices of size $n \times n$, we consider block matrices of size $m \times d$ with $n \times n$ blocks. We only define the search variant of Structured LWE and the decisional variant of Structured NTRU as they are the only one needed in this paper, but one could define the other versions in the natural way. The main motivation for working with these problems is the simpler analysis due to its formulation over $\mathbb{Z}$ instead of R. Furthermore, both S-LWE and S-NTRU can be instantiated with distributions that are not directly linked to M-LWE and M-NTRU.

The additional dimension in Structured LWE does not introduce much complications compared to the case $d = 1$ from [BD20b]. This is why most of the results in Section 3 and 4 are mere reformulations encompassing larger ranks. The technical difficulties of dealing with block matrices arise when concretely instantiating the hardness result to derive the hardness of M-LWE for arbitrary secret distributions later in Section 4 and 5.

Definition 3.1 (Structured LWE, [BD20b, Def. 3.1] adapted). *Let n, d, q, and m be positive integers. Let $\mathcal{M}$ be a distribution of matrices on $\mathbb{Z}_q^{n\times n}$, and Υ be a distribution of error-distributions on $\mathbb{R}^n$. Furthermore, let $\mathcal{S}$ be a distribution on $\mathbb{Z}_q^{nd}$. The goal of the* S-LWE$_{n,d,q,m,\mathcal{M},\Upsilon,\mathcal{S}}$ *problem is to find the secret $\mathbf{s} \hookleftarrow \mathcal{S}$ given $(\mathbf{A}, \mathbf{y}) = (\mathbf{A}, \mathbf{A}\mathbf{s} + \mathbf{e} \bmod q)$, with $\mathbf{A} \hookleftarrow \mathcal{M}^{m\times d}$, and $\mathbf{e} \hookleftarrow \psi^m$ where $\psi \hookleftarrow \Upsilon$.*

If m is not specified, it means we consider the samples one by one. A single sample follows the same definition for $m = 1$.

Definition 3.2 (Structured NTRU [BD20b, Def. 5.1] adapted). *Let n, d, q, and m be positive integers. Let $\mathcal{M}$ be a distribution of matrices on $\mathbb{Z}_q^{n\times n}$, and Ψ*

a distribution on $GL_{nd}(\mathbb{Z}, q) \times \mathbb{Z}^{nm\times nd}$. *The* $\text{S-NTRU}_{n,d,q,m,\mathcal{M},\Psi}$ *problem is to distinguish the two following distributions*

1. $\mathbf{G} \cdot \mathbf{F}_q^{-1} \bmod q$, *where* $(\mathbf{F}, \mathbf{G}) \hookleftarrow \Psi$, *and* $\mathbf{F}_q^{-1}$ *is the* $\mathbb{Z}_q$*-inverse of* $\mathbf{F} \bmod q$.
2. $\mathbf{U} \hookleftarrow \mathcal{M}^{m\times d}$.

Notice that these formulations are very similar to Definition 2.3 and 2.4 but where the ring R is embedded as $\mathbb{Z}^n$. In Section 3.2, we prove that they are equivalent up to carefully chosen mappings between the different distributions, as done in [BD20b] for the case of R-LWE. Further, we sometimes decompose the vectors (resp. matrices) in blocks of size n (resp. $n \times n$) to show the correspondence between R and $\mathbb{Z}^n$. We then have $\mathbf{A} = [\mathbf{A}_{ij}]_{(i,j)\in[m]\times[d]}$ with $\mathbf{A}_{ij} \in \mathbb{Z}_q^{n\times n}$, and $\mathbf{x} = [\mathbf{x}_1^T, \ldots, \mathbf{x}_k^T]^T$ for a vector $\mathbf{x}$ of size nk, and $\mathbf{x}_i \in \mathbb{Z}^n$ (or $\mathbb{R}^n$ depending on the context).

3.2 Module Problems as Structured Problems

For completeness, we detail here how to transform a (primal) Module-LWE (resp. Module-NTRU) instance into an instance of Structured LWE (resp. Structured NTRU) by choosing the correct distributions. The transformation consists in embedding the ring elements either as vectors or as multiplication matrices. M_R maps a ring element to a structured matrix as illustrated in Figure 2.1, thus motivating the names Structure LWE and Structured NTRU. Although the proof is trivial, we provide it for completeness.

Lemma 3.1. *Let* $K = \mathbb{Q}(\zeta)$ *be a number field of degree* n*, and* R *its ring of integers. Let* d, q, m *be positive integers. We set* $\mathcal{M} = M_R(U(R_q))$ *the distribution over* $\mathbb{Z}_q^{n\times n}$*, where* M_R *is defined in Section 2.2.*

LWE. *Let* ψ' *be a distribution over* $K_\mathbb{R}$*, and* $\mathcal{S}'$ *a distribution over* R_q^d*. Define* $\psi = (\mathbf{B}_R \circ \sigma_H)(\psi')$*, and* $\mathcal{S} = (\mathbf{B}_R \circ \sigma_H)(\mathcal{S}')$*, where* $\mathbf{B}_R$ *is defined in Section 2.2. Then, the two problems* $\text{M-LWE}_{n,d,q,m,\psi',\mathcal{S}'}$ *and* $\text{S-LWE}_{n,d,q,m,\mathcal{M},\psi,\mathcal{S}}$ *are equivalent.*

NTRU. *Now, let* ψ' *be a distribution over* R*. We define* Ψ *to be the distribution over* $GL_{nd}(\mathbb{Z}, q) \times \mathbb{Z}^{nm\times nd}$ *obtained by drawing* $\mathbf{F}$ *from* $(\psi')^{d\times d}$ *conditioned on being in* $GL_d(R, q)$*,* $\mathbf{G}$ *from* $(\psi')^{m\times d}$ *and outputting* $(M_R(\mathbf{F}), M_R(\mathbf{G}))$*. Then, the two problems* $\text{M-NTRU}_{n,d,q,m,\psi'}$ *and* $\text{S-NTRU}_{n,d,q,m,\mathcal{M},\Psi}$ *are equivalent.*

Proof. Since all the distributions are defined via an invertible mapping (either M_R or $\mathbf{B}_R \circ \sigma_H$), we only prove one direction. The other simply consists in using the inverse mappings.

LWE. Let $(\mathbf{A}, \mathbf{A}\mathbf{s} + \mathbf{e} \bmod qR) \in R_q^{m\times d} \times \mathbb{T}_q^m$ where $\mathbf{A} \hookleftarrow U(R_q^{m\times d})$, $\mathbf{s} \hookleftarrow \mathcal{S}'$ with $\mathcal{S}'$ is a distribution on R_q^d, and $\mathbf{e} \hookleftarrow (\psi')^m$. The transformation consists in applying $(M_R, \mathbf{B}_R \circ \sigma_H)$ to the instance $(\mathbf{A}, \mathbf{A}\mathbf{s} + \mathbf{e} \bmod qR)$. We now verify that the result is correctly distributed with respect to $\mathcal{M}$, $\mathcal{S}$ and ψ. For clarity, we denote $\mathbf{B}_R \circ \sigma_H$ by ϕ. Note that ϕ is linear. For all $j \in [m]$, it holds

$$\phi([\mathbf{A}\mathbf{s} + \mathbf{e}]_i) = \sum_{j\in[d]} \phi(a_{ij} \cdot s_j) + \phi(e_i) = \sum_{j\in[d]} M_R(a_{ij}) \cdot \phi(s_j) + \phi(e_i)$$
$$= [M_R(\mathbf{A}) \cdot \phi(\mathbf{s}) + \phi(\mathbf{e})]_i.$$

Here we handle vectors and block matrices but the computation still holds. Simply note that $[\phi(\mathbf{s})]_j$ corresponds to the j-th block of $\phi(\mathbf{s})$ which is column vector of size n, and $M_R(a_{ij})$ is the (i, j) block of size $n \times n$ of the matrix $M_R(\mathbf{A})$ by definition. The way we defined, $\mathcal{M}$, $\mathcal{S}$ and ψ from $M_R(.)$, $\mathcal{S}'$ and ψ', it is clear that $(M_R(\mathbf{A}), M_R(\mathbf{A})\phi(\mathbf{s}) + \phi(\mathbf{e}))$ is distributed according to S-LWE$_{n,d,q,m,\mathcal{M},\psi,\mathcal{S}}$. The oracle for S-LWE$_{n,d,q,m,\mathcal{M},\psi,\mathcal{S}}$ thus returns $\phi(\mathbf{s})$, from which we can recover $\mathbf{s}$ by inverting ϕ.

NTRU. Let $\mathbf{A}$ be in $R_q^{m\times d}$. First assume that $\mathbf{A}$ is uniform over $R_q^{m\times d}$. Then, it holds that $M_R(\mathbf{A})$ is distributed according to $\mathcal{M}$. Next, assume that $\mathbf{A} = \mathbf{G}\mathbf{F}_q^{-1} \bmod qR$, where $\mathbf{F} \hookleftarrow (\psi')^{d\times d}$ conditioned on being in $GL_d(R, q)$, $\mathbf{G} \hookleftarrow (\psi')^{m\times d}$, and $\mathbf{F}_q^{-1}$ is the R_q-inverse of $\mathbf{F}$. As M_R is a ring homomorphism, it holds that $M_R(\mathbf{A}) = M_R(\mathbf{G})M_R(\mathbf{F}_q^{-1}) \bmod q\mathbb{Z}$. Since $M_R(\mathbf{F}_q^{-1})$ is the $\mathbb{Z}_q$-inverse of $M_R(\mathbf{F})$, the definition of Ψ gives that $M_R(\mathbf{A})$ is correctly distributed. A distinguisher for S-NTRU$_{n,d,q,m,\mathcal{M},\Psi}$ applied to $M_R(\mathbf{A})$ thus gives a distinguisher for M-NTRU$_{n,d,q,m,\psi'}$. □

Recall that in cyclotomic fields for example, we can choose $\mathbf{B}_R = (\mathbf{U}_H^\dagger \mathbf{V})^{-1}$, which leads to $M_R = M_\tau$ and $\phi = \tau$. In this case, it simply uses the coefficient embedding to embed M-LWE (resp. M-NTRU) into S-LWE (resp. S-NTRU). Note that the hardness of S-NTRU can be established for other distributions Ψ, but based on different assumptions than M-NTRU. We discuss it in Section 5.2. The distribution of the blocks is chosen to be $\mathcal{M} = M_R(U(R_q))$. The reader can keep this choice in mind, but we point out that the results of Section 3.3, 3.4 and 4 hold for arbitrary distributions $\mathcal{M}$, $\mathcal{S}$ and ψ.

3.3 (Mild) Hardness

We consider the two notions of hardness for S-LWE as is done in [BD20b], namely standard hardness and the weaker notion of *mild hardness*. We show that standard hardness naturally implies mild hardness, while the converse require an a priori unbounded number of samples in order to use a success amplification argument. All the proofs are provided in Appendix B.3 for completeness.

Definition 3.3 (Standard and Mild Hardness). *Let n, d, q be positive integers. Let $\mathcal{M}$ be a distribution over $\mathbb{Z}_q^{n\times n}$, and Υ a distribution of distributions over $\mathbb{R}^n$. Finally, let $\mathcal{S}$ be a distribution on $\mathbb{Z}_q^{nd}$. For any $(\mathbf{s}, \psi)$ sampled from $(\mathcal{S}, \Upsilon)$, we denote by $\mathcal{O}_{\mathbf{s},\psi}$ the (randomized) oracle that, when called, returns $(\mathbf{A}_i, \mathbf{A}_i\mathbf{s} + \mathbf{e}_i \bmod q)$, where $\mathbf{A}_i \hookleftarrow \mathcal{M}^{1\times d}$ and $\mathbf{e}_i \hookleftarrow \psi$.*

The S-LWE$_{n,d,q,\mathcal{M},\Upsilon,\mathcal{S}}$ *problem is standard hard, if for every* PPT *adversary $\mathcal{A}$ and every non-negligible function ε, there exists a negligible function ν*

such that

$$\mathbb{P}_{\substack{\mathbf{s} \hookleftarrow \mathcal{S} \\ \psi \hookleftarrow \Upsilon}} \left[\mathbb{P}_{\mathcal{A}, \mathcal{O}_{\mathbf{s},\psi}} \left[\mathcal{A}^{\mathcal{O}_{\mathbf{s},\psi}}(1^\lambda) = \mathbf{s} \right] \geq \varepsilon(\lambda) \right] \leq \nu(\lambda),$$

where $\mathcal{A}^{\mathcal{O}_{\mathbf{s},\psi}}$ *means that the adversary has access to* $\mathcal{O}_{\mathbf{s},\psi}$ *as a black-box and can thus query it as many times as they want. The internal probability is over the random coins of* $\mathcal{A}$ *and the random coins of* $\mathcal{O}_{\mathbf{s},\psi}$ *(meaning over the randomness of the* $(\mathbf{A}_i, \mathbf{e}_i)$*).*

We now say that the S-LWE$_{n,d,q,\mathcal{M},\Upsilon,\mathcal{S}}$ *problem is* mildly *hard, if for every* PPT *adversary* $\mathcal{A}$ *and every negligible function* μ*, there exists a negligible function* ν *such that*

$$\mathbb{P}_{\substack{\mathbf{s} \hookleftarrow \mathcal{S} \\ \psi \hookleftarrow \Upsilon}} \left[\mathbb{P}_{\mathcal{A}, \mathcal{O}_{\mathbf{s},\psi}} \left[\mathcal{A}^{\mathcal{O}_{\mathbf{s},\psi}}(1^\lambda) = \mathbf{s} \right] \geq 1 - \mu(\lambda) \right] \leq \nu(\lambda).$$

When the number of available samples m *is fixed a priori, we use the same definitions except that* $\mathcal{A}$ *is only allowed at most* m *queries to the oracle. The samples can be written in matrix form as* $(\mathbf{A}, \mathbf{A}\mathbf{s} + \mathbf{e} \bmod q)$ *with* $\mathbf{A} \hookleftarrow \mathcal{M}^{m \times d}$ *and* $\mathbf{e} \hookleftarrow \psi^m$*.*

Lemma 3.2. *Let* n, d, q *be positive integers. Let* $\mathcal{M}$ *be a distribution over* $\mathbb{Z}_q^{n \times n}$*, and* Υ *a distribution of distributions over* $\mathbb{R}^n$*. Finally, let* $\mathcal{S}$ *be a distribution on* $\mathbb{Z}_q^{nd}$*. If* S-LWE$_{n,d,q,\mathcal{M},\Upsilon,\mathcal{S}}$ *is standard hard, then it is also mildly hard. The same result holds when the number of available samples* m *is fixed.*

Lemma 3.3 ([BD20b, Lem. 3.4] adapted). *Let* n, d, q *be positive integers. Let* $\mathcal{M}$ *be a distribution over* $\mathbb{Z}_q^{n \times n}$*, and* Υ *a distribution of distributions over* $\mathbb{R}^n$*. Finally, let* $\mathcal{S}$ *be a distribution on* $\mathbb{Z}_q^{nd}$*. If* S-LWE$_{n,d,q,\mathcal{M},\Upsilon,\mathcal{S}}$ *is mildly hard, then it is also standard hard.*

3.4 Rerandomization

Lemma 3.3 holds for an unbounded number of samples, which is not always realistic. In order to generate new samples from a fixed number, we recall the rerandomization lemma provided in [BD20b]. This procedure creates a new sample by adding up a random subset of the available samples. It results in a different error distribution which might not be easy to work with. The proof of the following is given in Appendix B.3.

Lemma 3.4 ([BD20b, Lem. 3.5] adapted). *Let* n, d, q *be positive integers. Let* $\mathcal{M}$ *be a distribution over* $\mathbb{Z}_q^{n \times n}$*, and* $\mathcal{S}$ *be a distribution on* $\mathbb{Z}_q^{nd}$*. Let* m *be a positive integer such that* $m \geq n^2 d \log_2 q + n \log_2 q + \omega(\log_2 \lambda)$*. Let* Φ *be an error distribution on* $\mathbb{Z}^n$*. The distribution of error-distributions* $\Upsilon_{\Phi,bin}$ *is defined as follows: A distribution* $\psi \hookleftarrow \Upsilon_{\Phi,bin}$ *is determined by* m *elements* $\mathbf{e}_1, \ldots, \mathbf{e}_m \in \mathbb{Z}^n$ *chosen from* Φ*. To sample from the distribution* ψ*, choose* $\mathbf{x} \hookleftarrow \{0,1\}^m$ *uniformly at random and output* $\sum_i x_i \mathbf{e}_i$*. If* S-LWE$_{n,d,q,m,\mathcal{M},\Phi,\mathcal{S}}$ *is mildly hard, then* S-LWE$_{n,d,q,\mathcal{M},\Upsilon_{\Phi,bin},\mathcal{S}}$ *is also mildly hard.*

4 Entropic Hardness of Structured LWE

In this section, we adapt the notion of *sometimes lossy pseudorandom distribution* from [BD20b] to our more general version of Structured LWE. They gather two main properties, namely pseudorandomness and sometimes lossiness, which are essential in proving the entropic hardness of S-LWE. Section 4.1 formalizes this idea that if there exists a sometimes lossy pseudorandom distribution, then S-LWE is mildy hard. Then, Section 4.2 gives sufficient conditions to construct such distributions. As the proofs of this section do not require technical novelty compared to the original ones, we defer them to Appendix B.4. Also, note that all the matrices in this section are over $\mathbb{Z}$, $\mathbb{Z}_q$, $\mathbb{Q}$ or $\mathbb{R}$, and not R, R_q, K or $K_\mathbb{R}$.

Definition 4.1 ([BD20b, Def. 4.1] adapted). *Let n, d, q, m be positive integers. Let $\mathcal{X}$ be a distribution on $\mathbb{Z}_q^{nm \times nd}$, $\mathcal{M}$ a distribution on $\mathbb{Z}_q^{n \times n}$, $\mathcal{S}$ a distribution on $\mathbb{Z}_q^{nd}$ and ψ an error distribution on $\mathbb{R}^n$. We say that $\mathcal{X}$ is a* sometimes lossy pseudorandom distribution *for $(\mathcal{S}, \mathcal{M}, \psi)$ if there exists a negligible function ε, a $\kappa = \omega(\log_2 \lambda)$ and a $\delta \geq 1/\mathsf{poly}(\lambda)$ such that the following properties hold.*

Pseudorandomness: *$\mathcal{X}$ is computationally indistinguishable from $\mathcal{M}^{m \times d}$*

Sometimes Lossiness: $\mathbb{P}_{\mathbf{A} \hookleftarrow \mathcal{X}}[\widetilde{H}_\infty^\varepsilon(\mathbf{s}|\mathbf{A}, \mathbf{A}\mathbf{s} + \mathbf{e} \bmod q) \geq \kappa] \geq \delta$*, where $\mathbf{s} \hookleftarrow \mathcal{S}$ and $\mathbf{e} \hookleftarrow \psi^m$.*

4.1 From Sometimes Lossiness to the Entropic Hardness of Structured LWE

The following theorem adapted from [BD20b] states that the existence of a sometimes lossy pseudorandom distribution implies the mild hardness of Structured LWE. The proof can be found in Appendix B.4. The pseudorandomness property essentially allows us to trade the uniform matrix $\mathbf{A}$ from the S-LWE instance for the matrix $\mathbf{G}\mathbf{F}_q^{-1} \bmod q$ (where $\mathbf{F}_q^{-1}$ is the $\mathbb{Z}_q$-inverse of a short matrix $\mathbf{F}$) as in Definition 2.4. Then, the sometimes lossiness property translates the fact that going from $\mathbf{s}$ to $\mathbf{G}\mathbf{F}_q^{-1}\mathbf{s} + \mathbf{e} \bmod q$ loses enough information on $\mathbf{s}$ (with non-negligible probability over the choice of $\mathbf{F}$ and $\mathbf{G}$) so that it is hard to recover $\mathbf{s}$. The sometimes lossiness thus entails the hardness of S-LWE when $\mathbf{A} = \mathbf{G}\mathbf{F}_q^{-1}$ instead of being uniform. By the pseudorandomness, it then yields the hardness of S-LWE.

Theorem 4.1 ([BD20b, Thm. 4.2] adapted). *Let n, d, q, m be positive integers. Let $\mathcal{X}$ be a distribution on $\mathbb{Z}_q^{nm \times nd}$, $\mathcal{M}$ a distribution on $\mathbb{Z}_q^{n \times n}$, $\mathcal{S}$ a distribution on $\mathbb{Z}_q^{nd}$ and ψ an error distribution on $\mathbb{R}^n$. We assume that all the distributions are efficiently sampleable. If the distribution $\mathcal{X}$ is a sometimes lossy pseudorandom distribution for $(\mathcal{M}, \mathcal{S}, \psi)$, then* $\text{S-LWE}_{n,d,q,m,\mathcal{M},\psi,\mathcal{S}}$ *is mildly hard.*

4.2 Construction of Sometimes Lossy Pseudorandom Distributions

We now provide the generalization of [BD20b] to our new problem S-NTRU in order to give sufficient conditions to construct sometimes lossy pseudorandom distributions, and therefore get the mild hardness of S-LWE$_{n,d,q,m,\mathcal{M},\psi,\mathcal{S}}$ by Theorem 4.1. We first provide a few technical lemmata before the main result of this subsection. The proofs follow the same structure as the one from [BD20b], which is why we defer them to Appendix B.4. The goal is to prove that $\mathbf{s}$ has sufficient min-entropy left, even if $\mathbf{G}\mathbf{F}_q^{-1}\mathbf{s}+\mathbf{e}_0$ is known. Recall that $\mathbf{B}_R \in \mathbb{R}^{n\times n}$ and for $\ell \geq 1$, $\mathbf{B}_{R^\ell} = \mathbf{I}_\ell \otimes \mathbf{B}_R$, as defined in Section 2.2. Note that in the following, we need both the inverse modulo q ($\mathbf{F}_q^{-1}$) and the rational inverse ($\mathbf{F}^{-1}$) of a matrix $\mathbf{F} \in \mathbb{Z}_q^{nd\times nd}$. The invertibility modulo q implies that the determinant of $\mathbf{F}$ is a unit in $\mathbb{Z}_q$, and is therefore non-zero when seen as an element of $\mathbb{Q}$, which in turns implies the rational invertibility.

Lemma 4.1 ([BD20b, Lem. 5.4] adapted). *Let n, d, and m be positive integers. Let $\mathbf{Z} = [\mathbf{Z}_{i,j}]_{(i,j)\in[m]\times[d]} \in \mathbb{R}^{nm\times nd}$, where for all (i,j), $\mathbf{Z}_{i,j} \in \mathbb{R}^{n\times n}$ and set $\mathbf{Z}' = \mathbf{B}_{R^m}^{-1} \cdot \mathbf{Z} \cdot \mathbf{B}_{R^d}$. Let $s, s' > 0$ be such that $s \geq \|\mathbf{Z}'\|_2 \cdot s'$. There exists a distribution Ψ on $\mathbb{R}^{nm}$, such that if $\mathbf{e}' \hookleftarrow D_{s'\cdot\mathbf{B}_{R^d}} = D_{s'\cdot\mathbf{B}_R}^d$ and $\mathbf{e}'' \hookleftarrow \Psi$ are independent, then $\mathbf{e} = \mathbf{Z}\mathbf{e}' + \mathbf{e}''$ is distributed according to $D_{s\cdot\mathbf{B}_{R^m}} = D_{s\cdot\mathbf{B}_R}^m$.*

Lemma 4.2 ([BD20b, Lem. 5.5] adapted). *Let n, d, q be positive integers. Let $\mathbf{F} \in GL_{nd}(\mathbb{Z}, q)$, and $\mathbf{F}' = \mathbf{B}_{R^d}^{-1}\mathbf{F}\mathbf{B}_{R^d}$. Let $s > \sqrt{2}\|\mathbf{F}'\|_2 \eta_\varepsilon(\mathbf{B}_{R^d}^{-1})$. Let $\mathbf{f} \hookleftarrow D_{\sqrt{2}\cdot s\mathbf{B}_{R^d}}$ and $\mathbf{e} \hookleftarrow \mathcal{D}_{\varLambda(\mathbf{F}), s\mathbf{B}_{R^d}}$. Let $\mathbf{s}$ be a random variable supported on $\mathbb{Z}_q^{nd}$. Then it holds that*

$$\widetilde{H}_\infty^{2\varepsilon}(\mathbf{s}|\mathbf{F}_q^{-1}\mathbf{s} + \mathbf{F}^{-1}\mathbf{f} \bmod q) \geq \widetilde{H}_\infty(\mathbf{s}|\mathbf{F}_q^{-1}\mathbf{s} + \mathbf{F}^{-1}\mathbf{e} \bmod q),$$

with $\mathbf{F}_q^{-1}$ denoting its $\mathbb{Z}_q$-inverse, and $\mathbf{F}^{-1}$ its $\mathbb{Q}$-inverse.

Lemma 4.3 ([BD20b, Lem. 5.6] adapted). *Let n, d be positive integers, and $\mathbf{s}$ a random variable supported on $\mathbb{Z}_q^{nd}$. Let $s \geq \eta_\varepsilon(\varLambda(\mathbf{B}_{R^d}^{-1}))$, $\mathbf{f} \hookleftarrow \mathcal{D}_{\mathbb{Z}^{nd},\sqrt{2}\cdot s\mathbf{B}_{R^d}}$ and $\mathbf{e} \hookleftarrow D_{s\mathbf{B}_{R^d}}$, then it holds that $\widetilde{H}_\infty^{8\varepsilon}(\mathbf{s}|\mathbf{s}+\mathbf{f}) \geq \widetilde{H}_\infty(\mathbf{s}|\mathbf{s}+\mathbf{e})$.*

Altogether, Lemmata 4.1, 4.2, and 4.3 lead to the following theorem. It gives the necessary link between the hardness of recovering $\mathbf{s}$ given $\mathbf{G}\mathbf{F}_q^{-1}\mathbf{s}+\mathbf{e}_0$, and the noise losiness of the secret distribution. We first use Lemma 4.1 to decompose the noise $\mathbf{e}_0$ into $\mathbf{G}\mathbf{F}^{-1}\mathbf{e}_1 + \mathbf{e}_1'$ along the matrix $\mathbf{G}\mathbf{F}^{-1}$, where $\mathbf{F}^{-1}$ is the $\mathbb{Q}$-inverse of $\mathbf{F}$. To avoid a noise blow-up, the spectral norm of $\mathbf{G}\mathbf{F}^{-1}$ needs to be as small as possible, which is done by maximizing the smallest singular value of $\mathbf{F}$ and minimizing the largest singular value of $\mathbf{G}$. We then use Lemma 4.2 to go from a continuous to a discrete Gaussian distribution. We argue that the entropy of $\mathbf{s}$ knowing $\mathbf{G}\mathbf{F}_q^{-1}\mathbf{s}+(\mathbf{G}\mathbf{F}^{-1}\mathbf{e}_1+\mathbf{e}_1')$ mod q is larger than that of $\mathbf{s}$ knowing only some (reduced) coset $\mathbf{s}+\mathbf{e}_2$ mod q where $\mathbf{e}_2$ is sampled on the lattice generated by $\mathbf{F}$. If the latter entropy is large enough, which requires to minimize the largest singular value of $\mathbf{F}$, then so is the former. Finally, Lemma 4.3 leads back to a continuous noise, to express this condition in terms of the noise lossiness of $\mathcal{S}$.

Theorem 4.2 ([BD20b, Thm. 5.7] adapted). *Let n, d, q, m be positive integers, and $\mathcal{S}$ a distribution over $\mathbb{Z}_q^{nd}$. Then, let $\mathbf{F} \in GL_{nd}(\mathbb{Z}, q)$. For $(i, j) \in [m] \times [d]$, let $\mathbf{G}_{ij} \in \mathbb{Z}^{n \times n}$ be matrices and $\mathbf{G} = [\mathbf{G}_{ij}]_{(i,j)\in[m]\times[d]}$ be the block matrix of the $\mathbf{G}_{ij}$. Further let $\mathbf{F}^{-1} \in \mathbb{Q}^{nd \times nd}$ be the $\mathbb{Q}$-inverse of $\mathbf{F}$ and $\mathbf{F}_q^{-1} \in \mathbb{Z}_q^{nd\times nd}$ be the $\mathbb{Z}_q$-inverse of $\mathbf{F} \bmod q$. Define the matrix $\mathbf{F}' = \mathbf{B}_{R^d}^{-1}\mathbf{F}\mathbf{B}_{R^d}$ and $\mathbf{G}' = \mathbf{B}_{R^m}^{-1}\mathbf{G}\mathbf{B}_{R^d}$. For $s > \|\mathbf{F}'\|_2 \cdot \eta_\varepsilon(\Lambda(\mathbf{B}_{R^d}^{-1}))$, let $s_0 \geq 2^{3/2}s\left\|\mathbf{G}'(\mathbf{F}')^{-1}\right\|_2$. Then it holds that*

$$\widetilde{H}_\infty^{18\varepsilon}(\mathbf{s}|\mathbf{G}\mathbf{F}_q^{-1}\mathbf{s} + \mathbf{e}_0 \bmod q) \geq \nu_{s\mathbf{B}_{R^d}}(\mathcal{S}) - nd\log_2(\|\mathbf{F}'\|_2),$$

where $\mathbf{e}_0 \hookleftarrow D_{s_0\mathbf{B}_{R^m}}$, and $\mathbf{s} \hookleftarrow \mathcal{S}$.

By combining the previous results, we obtain the main theorem that gives explicit conditions to construct sometimes lossy pseudorandom distributions. The proof can be found in Appendix B.4.

Theorem 4.3 ([BD20b, Thm. 5.8] adapted). *Let n, d, m, q be positive integers, and $\beta_1, \beta_2 > 0$. Let Ψ be a distribution on $GL_{nd}(\mathbb{Z}, q) \times \mathbb{Z}^{nm\times nd}$, $\mathcal{M}$ a distribution on $\mathbb{Z}_q^{n\times n}$, and $\mathcal{S}$ a distribution on $\mathbb{Z}_q^{nd}$. Assume the hardness of* $\text{S-NTRU}_{n,d,q,m,\mathcal{M},\Psi}$. *Additionally, assume that if $(\mathbf{F}, \mathbf{G}) \hookleftarrow \Psi$ then*

- $\left\|\mathbf{B}_{R^m}^{-1}\mathbf{G}\mathbf{F}^{-1}\mathbf{B}_{R^d}\right\|_2 \leq \beta_1$ *where $\mathbf{F}^{-1}$ is the rational inverse of $\mathbf{F}$.*
- $\left\|\mathbf{B}_{R^d}^{-1}\mathbf{F}\mathbf{B}_{R^d}\right\|_2 \leq \beta_2$

with probability at least $\delta \geq 1/\mathsf{poly}(\lambda)$ over the choice of $(\mathbf{F}, \mathbf{G})$. Define the distribution $\mathcal{X}$ on $\mathbb{Z}_q^{nm\times nd}$ by $\mathbf{G}\mathbf{F}_q^{-1}$, where $(\mathbf{F}, \mathbf{G}) \hookleftarrow \Psi$ and $\mathbf{F}_q^{-1} \in \mathbb{Z}_q^{nd\times nd}$ is the $\mathbb{Z}_q$-inverse of $\mathbf{F}$. Let $s > \beta_2\eta_\varepsilon(\Lambda(\mathbf{B}_{R^d}^{-1}))$ and $s_0 > 2^{3/2}\beta_1 s$. Further assume that $\nu_{s\mathbf{B}_{R^d}}(\mathcal{S}) \geq nd\log_2(\beta_2) + \omega(\log_2(\lambda))$.
Then $\mathcal{X}$ is a sometimes lossy pseudorandom distribution for $(\mathcal{S}, \mathcal{M}, D_{s_0\mathbf{B}_{R^m}})$.

Therefore, Theorems 4.3 and 4.1 together yield the following immediate corollary.

Corollary 4.1. *Assume that the conditions of Theorem 4.3 are satisfied. Then the problem* $\text{S-LWE}_{n,d,q,m,\mathcal{M},\psi,\mathcal{S}}$ *is mildly hard.*

5 Instantiation for M-LWE

This section constitute our main contribution, which consists in concretely exhibiting a sometimes lossy pseudorandom distribution that implies the entropic hardness of M-LWE. We thus set the parameters so that it fits the requirements of both Section 3.2 and 4. As seen in the latter in Theorem 4.3, the S-NTRU problem must be hard for this distribution, and the distribution also needs to be somewhat well behaved in terms of its spectral properties. Lemma 3.1 gives the equivalence between M-NTRU and S-NTRU, which allows for expressing the entire result in the more algebraic module setting. The more technical aspect of this section comes from Section 5.1, in which we study the spectral properties that we need. In particular, we derive a lower bound on the smallest singular value of discrete Gaussian matrices over R (once embedded via M_{σ_H}). Then, in Section 5.2, we define this distribution and verify that it indeed leads to a sometimes lossy pseudorandom distribution. Combining it with Corollary 4.1 and the equivalence between M-LWE and S-LWE of Lemma 3.1, we then obtain the entropic (mild) hardness of M-LWE. All the results in this section hold for arbitrary number fields at the exception of Corollary 5.2 which is stated for cyclotomic fields.

5.1 Invertibility and Singular Values of Discrete Gaussian Matrices

We now recall [CPS+20, Thm. A.5] that gives the density of square discrete Gaussian matrices over R_q that are invertible modulo qR. This theorem gives concrete conditions so that $\mathcal{D}_{GL_d(R,q),\gamma}$ is efficiently sampleable. The proofs of [CPS+20, Thm. A.5, Lem. A.6] depend on the embedding that is chosen to represent R as a lattice. The paper uses Gaussian distributions in the coefficient embedding over the lattice $\tau(R)$ which differ from our context. As such we need to adapt the proofs for Gaussian distributions in the canonical embedding over the lattice $\sigma_H(R)$. The changes are mostly limited to volume arguments as the volume of the lattice R depends on the embedding. Also, note that the proofs of [CPS+20, Thm. A.5, Lem. A.6] still hold in any number field and for any splitting behaviour of q provided that it is unramified, no matter the size of the norm of its prime factors.

Theorem 5.1 ([CPS+20, Thm. A.5] adapted). *Let K be a number field of degree n and R its ring of integers. Let $d \geq 1$ and $q > 2n$ be an unramified prime. We define $N_{\max} = \max_{\mathfrak{p}|qR, \mathfrak{p} \text{ prime}} N(\mathfrak{p})$ and $N_{\min} = \min_{\mathfrak{p}|qR, \mathfrak{p} \text{ prime}} N(\mathfrak{p})$. Assume that $\gamma \geq 2^{1/(2d-1)} \cdot (|\Delta_K| \cdot N_{\max}^{(d-1)/(2d-1)})^{1/n}$. Then*

$$\begin{aligned}\rho_\gamma(R^{d\times d} \setminus GL_d(R,q)) &\leq \frac{2r}{N_{\min}} \cdot \frac{\gamma^{nd^2}}{|\Delta_K|^{d^2/2}} \cdot (1 + 8d^2 2^{-n}) \\ &\leq \frac{2r}{N_{\min}} \cdot (1 + 8d^2 2^{-n}) \cdot \rho_\gamma(R^{d\times d}),\end{aligned}$$

where r is the number of distinct prime factors of qR.

Remark 5.1. Consider $q = 1 \bmod \nu$ over the ν-th cyclotomic field. Then, qR fully splits into n distinct prime ideals, each of norm q ($N_{\min} = N_{\max}$). Thus, if $\gamma \geq 2^{1/(2d-1)}(|\Delta_K| \cdot q^{(d-1)/(2d-1)})^{1/n}$ which is roughly $\Omega(n)$, then we have that

$$\mathbb{P}_{\mathbf{F} \hookleftarrow \mathcal{D}_{R,\gamma}^{d\times d}}[\mathbf{F} \notin GL_d(R,q)] \leq \frac{2n}{q} + \mathsf{negl}(n) \tag{2}$$

Note that if $q \leq 2n$, the inequality is vacuous. However, in practice q is usually much larger. For example, Kyber [BDK+18] uses $q \geq 13 \cdot n$ while the signature Dilithium [DKL+18] uses $q \geq n^{5/2} \gg 2n$. This yields a probability of invertibility that is sufficient for this work, while allowing for reducing the parameter γ as much as possible. More precisely, it allows for taking $\gamma = \Omega(n)$ with a constant close to 1. Also, as mentioned before, the invertibility in R_q implies that the determinant is a unit of R_q. As such, it is a non-zero element when seen in K which implies the K-invertibility.

Although it seems folklore, we weren't able to find a Gaussian tail bound on $\sigma(x)$ in the infinity norm for $x \hookleftarrow \mathcal{D}_{R,\gamma}$. We therefore provide the following lemma, whose proof is mostly based on [Pei08, Cor. 5.3], and which proves that $\|\sigma(x)\|_\infty \leq \gamma \log_2 n$ with overwhelming probability. Most of the tail bounds are with respect to the Euclidean norm and thus require an extra $\sqrt{n}$ factor. Here, we are only interested in the infinity norm. The proof is provided in Appendix B.5.

Lemma 5.1. *Let R be a ring of integers of degree n. Then for any $\gamma > 0$ and any $t \geq 0$, it holds that*

$$\mathbb{P}_{f \hookleftarrow \mathcal{D}_{R,\gamma}}[\|\sigma(f)\|_\infty \leq \gamma t] \geq 1 - 2ne^{-\pi t^2}.$$

Choosing $t = \log_2 n$ gives $\|\sigma(f)\|_\infty \leq \gamma \log_2 n$ with overwhelming probability.

The main challenge in instantiating Theorem 4.3 is to provide a decent bound for $\left\|\mathbf{G}'(\mathbf{F}')^{-1}\right\|_2$. It seems to require knowledge on the smallest singular value of $\mathbf{F}'$, which in our case is taken from a discrete Gaussian distribution. We now provide a lower bound on the smallest singular value of discrete Gaussian matrices. This automatically gives an upper bound on $\left\|(\mathbf{F}')^{-1}\right\|_2$, as $\left\|(\mathbf{F}')^{-1}\right\|_2 = 1/s_{\min}(\mathbf{F}')$.

Lemma 5.2. *Let $K = \mathbb{Q}(\zeta)$ be a number field of degree n, and R its ring of integers. Let $\mathcal{I}$ be any ideal of R. Let $\gamma > 0$ be a Gaussian parameter. Then, for all $\delta \geq 0$, it holds that*

$$\mathbb{P}_{\mathbf{F} \hookleftarrow \mathcal{D}_{\mathcal{I},\gamma}^{d \times d}}\left[s_{\min}(M_{\sigma_H}(\mathbf{F})) \leq \frac{\delta}{\sqrt{d}}\right] \leq nC_\gamma \delta + nc_\gamma^d,$$

with $C_\gamma > 0$ and $c_\gamma \in (0,1)$ parameters depending on γ.

Proof. **Spectral analysis.** For convenience, we define $S(\mathbf{A})$ to be the set of singular values of any complex matrix $\mathbf{A}$. First of all, note that $M_{\sigma_H}(\mathbf{F}) = (\mathbf{I}_d \otimes \mathbf{U}_H^\dagger)M_\sigma(\mathbf{F})(\mathbf{I}_d \otimes \mathbf{U}_H)$. Since $\mathbf{U}_H$ is unitary, we have $S(M_{\sigma_H}(\mathbf{F})) = S(M_\sigma(\mathbf{F}))$. Recall that $M_\sigma(\mathbf{F})$ is the block matrix of size $nd \times nd$ whose block $(i,j) \in [d]^2$ is $\operatorname{diag}(\sigma(f_{ij}))$. The matrix can therefore be seen as a $d \times d$ matrix with blocks of size $n \times n$. The idea is now to permute the rows and columns of $M_\sigma(\mathbf{F})$ to end up with a matrix of size $n \times n$ with blocks of size $d \times d$ only on the diagonal, as noticed in e.g. [DM14]. For that, we define the following permutation π of $[nd]$. For all $i \in [nd]$, write $i - 1 = k_1^{(i)} + nk_2^{(i)}$, with $k_1^{(i)} \in \{0, \ldots, n-1\}$ and $k_2^{(i)} \in \{0, \ldots, d-1\}$. Then, define $\pi(i) = 1 + k_2^{(i)} + dk_1^{(i)}$. This is a well-defined permutation based on the uniqueness of the Euclidean division. We can then define the associated permutation matrix $\mathbf{P}_\pi = [\delta_{i,\pi(j)}]_{(i,j)\in[nd]^2} \in \mathbb{R}^{nd\times nd}$. Then, it holds that

$$\mathbf{P}_\pi M_\sigma(\mathbf{F})\mathbf{P}_\pi^T = \begin{bmatrix} \sigma_1(\mathbf{F}) & & \\ & \ddots & \\ & & \sigma_n(\mathbf{F}) \end{bmatrix}.$$

Since $\mathbf{P}_\pi$ is a permutation matrix, it is unitary and therefore $S(M_\sigma(\mathbf{F})) = S(\mathbf{P}_\pi M_\sigma(\mathbf{F})\mathbf{P}_\pi^T)$. As $\mathbf{P}_\pi M_\sigma(\mathbf{F})\mathbf{P}_\pi^T$ is block-diagonal, we directly get the singular values by $S(\mathbf{P}_\pi M_\sigma(\mathbf{F})\mathbf{P}_\pi^T) = \cup_{k\in[n]} S(\sigma_k(\mathbf{F}))$. This proves that $S(M_{\sigma_H}(\mathbf{F})) = \cup_{k\in[n]} S(\sigma_k(\mathbf{F}))$. In particular, taking the minimum of the sets yields

$$s_{\min}(M_{\sigma_H}(\mathbf{F})) = \min_{k\in[n]} s_{\min}(\sigma_k(\mathbf{F})).$$

Random matrix theory. By [MP12, Lem. 2.8], for all $i, j \in [d]$ and unit vector $\mathbf{u} \in \mathbb{C}^n$, $\langle \sigma_H(f_{ij}), \mathbf{u}\rangle$ is sub-Gaussian with sub-Gaussian moment γ. Hence, since the rows of $\mathbf{U}_H$ are unit vectors of $\mathbb{C}^n$, it holds that for all $k \in [n]$ and $i, j \in [d]$, $\sigma_k(f_{ij})$ is sub-Gaussian with moment γ. Thus, for all $k \in [n]$, $\sigma_k(\mathbf{F})$ has independent and identically distributed sub-Gaussian entries. A result from random matrix theory by Rudelson and Vershynin [RV08, Thm 1.2] yields

$$\mathbb{P}\left[s_{\min}(\sigma_k(\mathbf{F})) \le \frac{\delta}{\sqrt{d}}\right] \le C_\gamma \cdot \delta + c_\gamma^d, \tag{3}$$

for some parameters $C_\gamma > 0$ and $c_\gamma \in (0,1)$ that only depend on the sub-Gaussian moment γ, for all $k \in [n]$. A union-bound gives that

$$\mathbb{P}\left[s_{\min}(M_{\sigma_H}(\mathbf{F})) \le \frac{\delta}{\sqrt{d}}\right] \le nC_\gamma\delta + nc_\gamma^d,$$

thus concluding the proof. □

An immediate application of this lemma is for $\mathcal{I} = R$, which gives a probabilistic bound depending on the constants C_γ and c_γ, which in turns depend on γ. Aside from experimentally noticing that C_γ decreases polynomially with γ and that c_γ seems to decrease exponentially with γ, we do not have closed-form expression of these constants. By scaling the distribution by γ and applying Lemma 5.2 for $\mathcal{I} = \gamma^{-1}R$, we obtain

$$\mathbb{P}_{\mathbf{F}\hookleftarrow\mathcal{D}_{R,\gamma}^{d\times d}}\left[s_{\min}(M_{\sigma_H}(\mathbf{F})) \le \frac{\gamma\delta}{\sqrt{d}}\right] \le nC\delta + nc^d,$$

with $C > 0$ and $c \in (0,1)$ no longer depending on γ. In this case, we can mitigate the union bound blow-up by choosing $\delta = n^{-3/2}$ for example, yielding $s_{\min}(M_{\sigma_H}(\mathbf{F})) \ge \gamma/n\sqrt{nd}$ with non-negligible probability. In what follows, we thus use the following corollary. We discuss in Section 6 how we can experimentally expect a better bound.

Corollary 5.1. *Let $K = \mathbb{Q}(\zeta)$ be a number field of degree n, and R its ring of integers. Let $\gamma > 0$ be a Gaussian parameter. It holds that*

$$\mathbb{P}_{\mathbf{F}\hookleftarrow\mathcal{D}_{R,\gamma}^{d\times d}}\left[s_{\min}(M_{\sigma_H}(\mathbf{F})) \le \frac{\gamma}{n\sqrt{nd}}\right] \le O(n^{-1/2}).$$

5.2 Instantiation

We now define a distribution Ψ over $GL_{nd}(\mathbb{Z}, q) \times \mathbb{Z}^{nm \times nd}$ and prove it verifies the conditions of Theorem 4.3 under a careful choice of parameters. This distribution is actually a direct application of Definition 2.4 for a Gaussian distribution ψ. As opposed to what is done in [BD20b], we no longer need to assume the hardness of Hermite Normal Form M-LWE and we solely rely on the M-NTRU assumption for *rectangular* matrices. Nonetheless, the distribution proposed in [BD20b, Sec. 6] can be adapted to the module setting as shown in Appendix A for completeness. In this case, the hardness of S-NTRU is proven under that of M-NTRU for *square* matrices and of HNF-M-LWE. It thus uses a more standard formulation of M-NTRU but at the expense of an additional assumption, and also slightly worse parameters. This highlights a trade-off between the underlying hardness assumptions and the parameters.

Definition 5.1. *Let K be a number field of degree n, and R its ring of integers. Let d, q, m be positive integers. Let $\gamma > 0$ be a Gaussian parameter. We define the distribution Ψ as follows:*

- *Choose $\mathbf{F} \hookleftarrow \mathcal{D}_{R,\gamma}^{d\times d}$ such that $\mathbf{F} \in GL_d(R, q)$;*
- *Choose $\mathbf{G} \hookleftarrow \mathcal{D}_{R,\gamma}^{m\times d}$.*
- *Output $(M_R(\mathbf{F}), M_R(\mathbf{G}))$.*

Note that by Theorem 5.1, Ψ is efficiently sampleable if γ is sufficiently large, depending on the splitting behaviour of q. In particular, as stated in Remark 5.1, one can choose $\gamma = \Theta(n)$ and have a non-negligible probability that a sample from $\mathcal{D}_{R^{d\times d},\gamma}$ is also in $GL_d(R, q)$ for a fully splitted prime q in a cyclotomic field. Also, since M_R is a ring homomorphism, if $\mathbf{F}$ is invertible modulo q, then so is $M_R(\mathbf{F})$, and vice-versa. By Lemma 3.1, M-NTRU and S-NTRU, are equivalent for a specific connection between the distributions. We specifically defined Ψ so that it matches with M-NTRU$_{n,d,q,m,\gamma}$. Hence, assuming the hardness of S-NTRU for this distribution Ψ is equivalent to assuming the hardness of M-NTRU$_{n,d,q,m,\gamma}$. We now show that the distribution Ψ leads to a sometimes lossy pseudorandom distribution. By Theorem 4.3 it suffices to bound the maximal singular values of $\mathbf{B}_{R^d}^{-1} M_R(\mathbf{F})\mathbf{B}_{R^d}$, and $\mathbf{B}_{R^m}^{-1} M_R(\mathbf{G}) M_R(\mathbf{F})^{-1}\mathbf{B}_{R^d}$, which is the object of the following lemma.

Lemma 5.3. *Let K be a number field of degree n, and R its ring of integers. Let d, q, m be positive integers, and $\gamma > 0$. Let $(M_R(\mathbf{F}), M_R(\mathbf{G})) \hookleftarrow \Psi$. It holds that*

1. $\left\|\mathbf{B}_{R^d}^{-1} M_R(\mathbf{F})\mathbf{B}_{R^d}\right\|_2 \leq d\gamma \log_2 n$
2. $\left\|\mathbf{B}_{R^m}^{-1} M_R(\mathbf{G}) M_R(\mathbf{F})^{-1}\mathbf{B}_{R^d}\right\|_2 \leq nd\sqrt{nm}\log_2 n$

except with probability at most $O(n^{-1/2}) + 2d(d+m)ne^{-\pi \log_2^2 n}$ over the choice of $(M_R(\mathbf{F}), M_R(\mathbf{G}))$. When $d, m = \mathsf{poly}(n)$, this probability is simply $O(n^{-1/2})$.

Proof. **1.** It holds that $\mathbf{B}_{R^d}^{-1} M_R(\mathbf{F}) \mathbf{B}_{R^d} = M_{\sigma_H}(\mathbf{F})$. Let (i,j) be in $[d]^2$. As $\mathbf{U}_H$ is unitary, it preserves the largest singular value, so it holds that

$$\left\| \mathbf{U}_H^{\dagger} \mathrm{diag}(\sigma(f_{ij})) \mathbf{U}_H \right\|_2 = \|\mathrm{diag}(\sigma(f_{ij}))\|_2 = \|\sigma(f_{ij})\|_\infty.$$

As $f_{ij} \hookleftarrow \mathcal{D}_{R,\gamma}$, Lemma 5.1 gives that $\|\sigma(f_{ij})\|_\infty \leq \gamma \log_2 n$ except with a probability of at most $2ne^{-\pi \log_2^2 n}$. Lemma 2.1 yields $\left\| \mathbf{B}_{R^d}^{-1} M_R(\mathbf{F}) \mathbf{B}_{R^d} \right\|_2 \leq d\gamma \log_2 n$ except with a probability of at most $d^2 \cdot 2ne^{-\pi \log_2^2 n}$ which is negligible for $d = \mathsf{poly}(n)$.

2. We now bound $\left\| \mathbf{B}_{R^m}^{-1} M_R(\mathbf{G}) M_R(\mathbf{F})^{-1} \mathbf{B}_{R^d} \right\|_2$ from above. Note that we have

$$\begin{aligned} \mathbf{B}_{R^m}^{-1} M_R(\mathbf{G}) M_R(\mathbf{F})^{-1} \mathbf{B}_{R^d} &= \mathbf{B}_{R^m}^{-1} M_R(\mathbf{G}) \mathbf{B}_{R^d} \mathbf{B}_{R^d}^{-1} M_R(\mathbf{F})^{-1} \mathbf{B}_{R^d} \\ &= M_{\sigma_H}(\mathbf{G}) M_{\sigma_H}(\mathbf{F})^{-1}. \end{aligned}$$

As $\|.\|_2$ is sub-multiplicative, we simply bound $\|M_{\sigma_H}(\mathbf{G})\|_2$ and $\left\|M_{\sigma_H}(\mathbf{F})^{-1}\right\|_2$ from above seperately. Using Lemma 5.1 and Lemma 2.1 once again yields

$$\|M_{\sigma_H}(\mathbf{G})\|_2 \leq \sqrt{md}\gamma \log_2 n,$$

except with probability at most $md \cdot 2ne^{-\pi \log_2^2 n}$. Finally, by Corollary 5.1 we have that $\left\|M_{\sigma_H}(\mathbf{F})^{-1}\right\|_2 \leq n\sqrt{nd}/\gamma$ except with a probability of at most $O(n^{-1/2})$. Hence

$$\left\| \mathbf{B}_{R^m}^{-1} M_R(\mathbf{G}) M_R(\mathbf{F})^{-1} \mathbf{B}_{R^d} \right\|_2 \leq nd\sqrt{nm} \log_2 n,$$

except with probability at most $O(n^{-1/2}) + md \cdot 2ne^{-\pi \log_2^2 n}$. □

We can now summarize the results of this section in our main theorem by combining Lemma 5.3 with Corollary 4.1. Using the equivalence of Lemma 3.1 between the module and structured formulations, we have the following.

Theorem 5.2. *Let K be a number field of degree n, and R its ring of integers. Let m, d be positive integers. Let q be a positive integer and $\gamma > 0$ be such that $\mathcal{D}_{GL_d(R,q),\gamma}$ is efficiently sampleable. Let $\mathcal{S}$ be a distribution on R_q^d and $\mathcal{S}' = \mathbf{B}_R \circ \sigma_H(\mathcal{S})$. Assume the hardness of $\text{M-NTRU}_{n,d,q,m,\mathcal{D}_{R,\gamma}}$, and that $\nu_s(\mathcal{S}') \geq nd \log_2(d\gamma \log_2 n) + \omega(\log_2 \lambda)$ for some $s > \gamma d \log_2(n) \cdot \eta_\varepsilon(R)$. Then, for $s_0 \geq 2^{3/2} \cdot s \cdot n\sqrt{nmd} \log_2 n$, we have that $\text{M-LWE}_{n,d,q,m,D_{s_0},\mathcal{S}}$ is mildly hard.*

What characterizes the secret distribution is the noise lossiness condition of Theorem 5.2. By using Lemma 2.2 and 2.3, we can have concrete conditions on the entropy of the secret distribution.

Corollary 5.2. *Let K be the ν-th cyclotomic field of degree $n = \varphi(\nu)$, and R its ring of integers. Let m, d be positive integers. Let $q > 2n$ be a prime such that $q = 1 \bmod \nu$, and $\gamma > 2^{1/(2d-1)} n q^{(d-1)/(n(2d-1))}$. Let $\mathcal{S}$ be a distribution on R_q^d*

and $\mathcal{S}' = \tau(\mathcal{S})$ (see Section 3.2). Assume the hardness of $\text{M-NTRU}_{n,d,q,m,\mathcal{D}_{R,\gamma}}$*. Also, assume that for some $\gamma d \log_2(n) \cdot \eta_\varepsilon(R) < s < q\sqrt{\pi \ln(4nd)}$ it holds that*

$$H_\infty(\mathcal{S}) \geq nd \log_2(d\gamma \log_2 n) + nd \log_2(q/s) + \omega(\log_2 \lambda).$$

Then, for $s_0 \geq 2^{3/2} \cdot s \cdot n\sqrt{nmd} \log_2 n$, we have that $\text{M-LWE}_{n,d,q,m,D_{s_0},\mathcal{S}}$ *is mildly hard.*

If $\mathcal{S}$ is supported on R_η^d for some positive integer $\eta \geq 2$, and for some $s > \gamma d \log_2(n) \cdot \eta_\varepsilon(R)$ it holds that

$$H_\infty(\mathcal{S}) \geq nd \log_2(d\gamma \log_2 n) + \sqrt{2\pi} \log_2 e \cdot nd \cdot (\eta - 1)/s + \omega(\log_2 \lambda),$$

then the conclusion still holds.

Proof (of Corollary 5.2). Note that as τ is a bijection, $\mathcal{S}$ and $\mathcal{S}'$ have the same entropy. Hence, the first statement is obtained simply by combining Theorem 5.2 with Lemma 2.2. For the second statement, we take a bounded distribution. In practical uses of M-LWE, the bounds are considered on the coefficients of the polynomials, which is why we consider a bound on the infinity norm of the coefficient embedding of the secrets. To use Lemma 2.3, we simply have to translate this bound into a bound on the secrets from $\mathcal{S}'$ in the Euclidean norm. Since K is a cyclotomic field, we can choose $\mathbf{B}_R = (\mathbf{U}_H^\dagger \mathbf{V})^{-1}$ as discussed in Section 2.2, which yields $\mathbf{B}_R \circ \sigma_H = \tau$. Hence, we indeed have $\mathcal{S}' = \mathbf{B}_R \circ \sigma_H(\mathcal{S})$ as required by Theorem 5.2. Now, as $\tau(\mathcal{S}) = \mathcal{S}'$ is supported on $\{0, \ldots, \eta - 1\}^{nd}$, for $\mathbf{s} \hookleftarrow \mathcal{S}$ it holds that

$$\|\mathbf{s}'\|_2 = \|\mathbf{B}_R \circ \sigma_H(\mathbf{s})\|_2 = \|\tau(\mathbf{s})\|_2 \leq (\eta - 1)\sqrt{nd}.$$

Applying Lemma 2.3 thus yields the second statement. □

Remark 5.2. The distribution from Definition 5.1 and the subsequent results can also be instantiated for modules of rank 1. The parameters can be optimized using the technical lemmata from [BD20b, Sec. 6] in this case. It provides an alternative distribution to the one proposed in [BD20b], which improves the parameters and discard the Hermite Normal Form R-LWE hardness assumption. It would then give a detailed proof of the argument made by Peikert [Pei16] that decisional NTRU reduces to search R-LWE. Our result can therefore be seen as a generalization of this argument to modules.

5.3 On the Statistical Entropic Hardness of M-LWE

Chuengsatiansup et al. [CPS$^+$20] show that if the Gaussian $\psi = \mathcal{D}_{R,\gamma}$ is sufficiently wide, and where q does not split too much, then the $\text{M-NTRU}_{n,d,q,m,\gamma}$ problem is statistically hard, as we restate below. Note that the original statement requires $d \geq m$, which is actually not needed in the proof. The final proof of Theorem 5.3 uses Theorem 5.1, and also the volume of R but in a ratio, which does not affect the result when changing the embedding.

Theorem 5.3 ([CPS+20, Thm. A.1] adapted). *Let K be a number field of degree n and R its ring of integers. Let $m, d \geq 1$ and q be an unramified prime such that $\min_{\mathfrak{p}|qR, \mathfrak{p} \text{ } prime} N(\mathfrak{p}) = 2^{\Omega(n)}$. We define $N_{\max} = \max_{\mathfrak{p}|qR, \mathfrak{p} \text{ } prime} N(\mathfrak{p})$. Let $\gamma \geq \max\left(2nq^{m/(d+m)+2/(n(d+m))}, 2^{1/(2d-1)}|\Delta_K|^{1/n} N_{\max}^{(d-1)/(n(2d-1))}\right)$. Then, let $\mathcal{X}_\gamma$ be the distribution of $\mathbf{G}\mathbf{F}_q^{-1} \bmod qR$ where $\mathbf{F} \hookleftarrow \mathcal{D}_{R,\gamma}^{d \times d}$ such that $\mathbf{F} \in GL_d(R, q)$, $\mathbf{F}_q^{-1}$ the R_q-inverse of $\mathbf{F}$ and $\mathbf{G} \hookleftarrow \mathcal{D}_{R,\gamma}^{m \times d}$. Then, we have*

$$\varDelta(\mathcal{X}_\gamma, \mathcal{U}(R_q^{m \times d})) \leq 2^{-\Omega(n)}.$$

Our hardness assumption for the mild hardness of Entropic M-LWE is statistically thus proven by Theorem 5.3 for wide Gaussian distributions and a modulus q that does not split into too many factors. In this case, our result introduces non-trivial lower bounds on the entropy of the secret and the size of the noise such that the M-LWE becomes statistically hard. By Theorem 5.2, this provides the mild hardness of M-LWE$_{n,d,q,m,D_{s_0},\mathcal{S}}$ with no computational assumption whatsoever. Nevertheless, the parameter γ required by Theorem 5.3 is roughly $2n\sqrt{q}$ and hence makes the parameters of M-LWE not usable in practice. In particular, the entropy of the secret distribution must be very large. It then requires that both the size of the secret and the size of the masking noise s must be of the order of at least $\sqrt{q}$, making it hard to build usable cryptosystems purely based on it. As such, the statistical result should only be seen as an interesting byproduct of the theoretical proof, but not as a groundbreaking result for practical applications.

Alternatively, one can also try to prove this result by showing that the distribution of the given M-LWE samples is statistically close to the uniform distribution. However, this would require a leftover hash lemma over modules where the input of the hash is only promised to have certain entropy. To the best of our knowledge, there is no such entropic variant of the leftover hash lemma over modules and we leave the investigation of it as a future work.

6 Related Work

In this section, we detail our main result of Theorem 5.2 and how it places with respect to existing hardness results for entropic M-LWE. To simplify the concrete comparison of parameters, we use the case of power-of-two cyclotomic fields and use the formulation of Corollary 5.2. The reasoning can be extended to more general number fields but requires the derivation of field-dependent values such as $\varDelta_K$ and $\|\tilde{\mathbf{L}}_{R^\vee}\| = \max_{i \in [n]} \|\tilde{\mathbf{L}}_{R^\vee} \mathbf{e}_i\|_2$. First, note that our reduction is rank-preserving in the sense that the module rank from our M-NTRU assumption equals the final module rank for entropic M-LWE. It can be advantageous for the concrete hardness analysis, but it also gives less room to tweak the parameters in order to achieve small secrets.

Constant secrets. In the case of uniform binary secrets as studied by Boudgoust et al. [BJRW20,BJRW21], the entropy of the secret distribution is nd. We can see that the entropy condition of Corollary 5.2 for $\mathcal{S} = U(R_2^d)$ ($\eta = 2$) is then never met. Therefore, our hardness proof does not encompass unusually small secrets. This is in part due to the rank-preserving nature of our proof, which does not enable us to increase the rank by $\log_2 q$ to attain constant secrets.

General secrets. We now focus on the case where the secret is from a general distribution $\mathcal{S}$. The first statement of Corollary 5.2 requires the entropy of $\mathcal{S}$ to be at least $nd \log_2(d\gamma \log_2(n) \cdot \sqrt{\ln(4nd)/\pi})$, when the masking noise s is at its upper-bound of $q\sqrt{\pi}/\sqrt{\ln 4nd}$. For this choice of s, it leads to $s_0 = n\sqrt{8\pi} \cdot q \cdot \sqrt{nm/\ln(4nd)} \cdot d \log_2 n$. This noise parameter s_0 is quite large because the masking noise is close to q. If we choose s to be much smaller, this impacts the minimal entropy that we can achieve. There is therefore a trade-off between minimizing the entropy of the secret distribution or the noise parameter s_0 to preserve the same hardness result, as expected.

Remark 6.1. Note that this result can be instantiated for secret distribution over secret with bounded norms. Hence a way of minimizing the entropy of the secret distribution can be to minimize the size of the secret. So the trade-off is now between the size of the secret and the size of the error.

Bounded secrets. In the more specific case of secrets with bounded norm, the second statement of Corollary 5.2 can improve the parameters slightly. For clarity, we define $C' = \sqrt{2\pi} \log_2 e$. Now, consider a parameter $\alpha > 0$. Assume we set $C'nd(\eta - 1)/s \leq \alpha n \log_2(d\gamma \log_2 n)$. This simplifies the entropy condition to $H_\infty(\mathcal{S}) \geq n(d + \alpha) \log_2(d\gamma \log_2 n) + \omega(\log_2 \lambda)$, while requiring $s \geq C'd(\eta - 1)/(\alpha \log_2(d\gamma \log_2 n))$. This condition on s is likely subsumed by the condition on s in the corollary statement, as long as α is not too small. For example, $\alpha = 1/n$ allows to satisfy both conditions on s. For the uniform distribution over R_η^d (maximal entropy), this translates to $\eta \geq (d\gamma \log_2 n)^{1+\alpha/d}$. One can thus tweak the parameter α to either achieve a smaller masking noise (and therefore a smaller noise s_0), or smaller secrets.

Concrete parameters. Our reduction provides theoretical insights and gives confidence in the fact that M-LWE-based cryptosystems are resilient even if the secret distribution present a certain amount of leakage. We insist on the fact that our proof does not encompass practical parameters, but can still be instantiated with concrete parameters that satisfy all the conditions. We give for example one such set of parameters.

The final noise s_0 can be improved by a factor of roughly n experimentally. Indeed, the bound on the smallest singular value of Lemma 5.2 can be applied for $\mathcal{I} = R$ directly which introduces the constant C_γ. Experimentally, it seems like $C_\gamma = O(1/\gamma^\delta)$ for $\delta \in (3/2, 2)$. Hence, for $\gamma = \Omega(n)$, it would yield $s_{\min}(M_\sigma(\mathbf{F})) \geq \gamma/\sqrt{nd}$ with non-negligible probability. This would thus save a factor of n in the expression of s_0. We have studied the lower bound with a heuristics for cyclotomic fields, and with γ under the condition of Theorem 5.1. It yields a lower bound of $\gamma/\sqrt{nd}$ on the smallest singular value with (experimental) probability at least $3/4$ (and going to 1 with polynomial speed as n grows). A bound of $\gamma/(10\sqrt{nd})$ is verified with (experimental) probability of at

n	d	m	q	γ	ε	s	η	s_0
256	4	4	281474976729601	299	2^{-100}	12005359	9649	8901435781154

Table 6.1. Example parameter sets verifying the following conditions: q prime with $q = 1 \bmod 2n$, $\gamma \geq 2^{\frac{1}{2d-1}} nq^{\frac{d-1}{n(2d-1)}}$, $s > d\gamma \log_2 n \cdot \eta_\varepsilon(R)$, $s > \sqrt{2\pi} \log_2 e \cdot \frac{d(\eta-1)}{\alpha \log_2(d\gamma \log_2 n)}$, $\eta \geq (d\gamma \log_2 n)^{1+\alpha/d}$, $\alpha = \frac{1}{n}$, $s_0 \geq \sqrt{8} \cdot s \cdot nd\sqrt{nm} \log_2 n$. Note that for a given $\varepsilon > 0$, one has $\eta_\varepsilon(R) \leq \sqrt{\ln(2n(1+\varepsilon^{-1}))/\pi} \frac{1}{\lambda_1^\infty(R^\vee)}$ and $\lambda_1^\infty(R^\vee) \geq N(R^\vee)^{1/n} = |\Delta_K|^{-1/n}$ which gives $\lambda_1^\infty(R^\vee) \geq 1/n$ for a power-of-two cyclotomic field.

least 99/100. For completeness, we provide a simulation script in Sage [Sag20] as Appendix alongside this paper. It simply generates such matrices and compares their smallest singular value to the bound $\gamma/(10\sqrt{nd})$. The bound seems coherent with the extensive research around spectral estimations of random matrices. The smallest singular value of random matrices has been widely studied in order to prove the Von Neumann & Goldstine conjecture [vNG47] that for a random matrix $\mathbf{A}$ of size $N \times N$, $s_{\min}(\mathbf{A})$ is asymptotically equivalent to $1/\sqrt{N}$ with high probability. This conjecture has been proven for specific distributions satisfying various conditions on the entries, which our matrix $M_{\sigma_H}(\mathbf{F})$ for $\mathbf{F} \hookleftarrow \mathcal{D}_{R,\gamma}^{d\times d}$ unfortunately does not verify. The bound seems however to hold heuristically.

6.1 Existing Hardness of Entropic M-LWE

As mentioned in the introduction, Lin et al. [LWW20][8] adapted the lossy argument approach of [BD20a] to the module setting. In order to compare with our hardness result, we need to make some modifications to their result to adapt it to the primal ring. We also need to adjust it to our definition of M-LWE which differs by a factor of q, i.e., considering $\mathbf{As}+\mathbf{e} \bmod qR$ instead of $q^{-1}\mathbf{As}+\mathbf{e}' \bmod R$ with $\mathbf{e}'$ having a Gaussian parameter in $(0, 1)$. An advantage of their hardness proof is that it holds for all number fields, but we limit our comparison to the case of power-of-two cyclotomic fields. Also, note that the following only concerns the search variant of entropic M-LWE. We then obtain the following.

Theorem 6.1 (Adapted from [LWW20]). *Let K be a power-of-two cyclotomic field of degree n, and R its ring of integers. Let d, k, q, m be positive integers such that $m > d > k \geq 1$, and $m = \omega(\log_2 \lambda)$. Assume the hardness of (primal decisional)* $\text{M-LWE}_{n,k,q,m,\mathcal{D}_{R,s_1}}$*. Let $\mathcal{S}$ be a distribution on R_q^d. Assume that for some $0 < s < q\|\tilde{\mathbf{L}}_{R^\vee}\|^{-1}\sqrt{\pi/\ln(4nd)}$ it holds that*

$$H_\infty(\mathcal{S}) \geq nk \log_2 q + nd \log_2(q|\Delta_K|^{1/2n}/s) + \omega(\log_2 \lambda).$$

Then, for $s_0 \geq O(s \cdot s_1\sqrt{m})$, we have that $\text{M-LWE}_{n,d,q,m,D_{s_0},\mathcal{S}}$ *is hard.*

[8] Note that at the time of writing, the paper by Lin et al. is only accessible on ePrint and has not yet been peer-reviewed.

If $\mathcal{S}$ is supported on R_η^d for a positive integer $\eta \geq 2$, and for some $s > 0$ it holds that

$$H_\infty(\mathcal{S}) \geq nk \log_2 q + \sqrt{2\pi} \log_2 e \cdot nd \cdot n(\eta - 1)/s + \omega(\log_2 \lambda),$$

then the conclusion still holds.

Moving to the primal ring simply has consequences on the smoothing parameter and therefore on the Gaussian parameters, hence the presence of $\|\tilde{\mathbf{L}}_{R^\vee}\|$ which is the norm of a basis of the dual of $\sigma_H(R)$ (up to conjugation). In the case of power-of-two cyclotomics, $\mathbf{L}_{R^\vee} = (1/n) \cdot \sqrt{n}\mathbf{U}$ for a unitary matrix $\mathbf{U}$. So $\|\tilde{\mathbf{L}}_{R^\vee}\| = 1/\sqrt{n}$. The other consequence comes from the volume of the module lattice $\sigma_H(R^d)$ which is $|\Delta_K|^d$ larger than the volume of the dual lattice $\sigma_H((R^\vee)^d)$. In power-of-two cyclotomic fields, we also have $|\Delta_K| = n^n$. Note that the masking noise with parameter s covers $\mathbf{s}$ and not $\mathbf{s}/q$ as in [LWW20]. This is the result of multiplying by q in the definition of M-LWE they use.

Constant secrets. As the hardness proof is not rank-preserving, the result of [LWW20] can be instantiated using a weaker hardness assumption (low k) to achieve constant secrets. In particular, Theorem 6.1 can be instantiated for binary secrets, requiring a rank $d \gtrsim (k+1)\log_2 q$.

General secrets. The first statement of Theorem 6.1 requires a large enough min-entropy, namely $H_\infty(\mathcal{S}) \geq nk\log_2 q + nd \log_2 \sqrt{\ln(4nd)/\pi} + \omega(\log_2 \lambda)$, when choosing $s = q\sqrt{n}/\sqrt{\ln(4nd)}$, leading to $s_0 = O(qs_1\sqrt{nm}/\sqrt{\ln(4nd)})$. Therefore, when k is much smaller than d, the entropy requirement becomes better than ours. To compare the noise parameter, we need to know the order of magnitude of s_1. By combining [LS15, Thm. 4.7] and [BJRW20, Lem. 13], one can obtain the hardness of decisional M-LWE for $s_1 \geq 2\sqrt{d}\omega((\log_2 n)^{3/2})$ based on module lattice problems. From this restriction on s_1, the noise parameter s_0 obtained by [LWW20] is slightly bigger than ours (for d small compared to n) by a factor of roughly $\omega(\sqrt{\log_2 n})/\sqrt{d}$. Also, recall that in our case m can also be a small constant, while [LWW20] requires $m \geq \max(d, \omega(\log_2 n))$.

Bounded secrets. In the case of secrets with bounded norm, the second statement can once again improve the parameters slightly. Consider a parameter $\alpha > 0$. Assume we set $C'nd \cdot n(\eta - 1)/s \leq \alpha n \log_2 q$. This simplifies the entropy condition to $H_\infty(\mathcal{S}) \geq n(k+\alpha)\log_2 q + \omega(\log_2 \lambda)$, while requiring $s \geq C'nd(\eta - 1)/(\alpha \log_2 q)$. For the uniform distribution over R_η^d which reaches maximal entropy, this now gives $\eta \geq q^{(k+\alpha)/d}$.

Conclusion. We summarize the discussion for secrets with bounded norm with maximal entropy in Table 6.2. Note that $k + \alpha$ cannot be larger than d otherwise $\eta_{\min} > q$. We trade the underlying hardness assumption for differences in the parameters. In particular, the main difference comes from comparing $k \log_2 q$ with $d \log_2(nd \log_2 n)$, as γ can be as low as n. In some parameter regimes, our proof method thus leads to slightly improved parameters. For example, for $n = 256$, $q = n^3$, and k close to d (close to rank-preserving reduction), we achieve better parameters in terms of noise and secret size.

	[LWW20]	**Corollary 5.2**
Number fields	Arbitrary	Arbitrary[a]
Constant secrets	Yes	No
Rank-preserving	No	Yes
Hardness assumption	M-LWE$_{k,\mathcal{D}_{R,s_1}}$	M-NTRU$_{d,\gamma=\Omega(n)}$
General Distribution $\mathcal{S}$		
Minimal Entropy $H_\infty(\mathcal{S}) - \omega(\log_2 \lambda)$	$nk \log_2 q$ $+nd \log_2(\frac{q\sqrt{n}}{s})$	$nd \log_2(d\gamma \log_2 n) + nd \log_2(\frac{q}{s})$
Maximal Masking Noise s	$s < q\sqrt{n}\sqrt{\frac{\pi}{\ln(4nd)}}$	$d\gamma \log_2(n)\eta_\varepsilon(R) < s < q\sqrt{\frac{\pi}{\ln(4nd)}}$
Bounded Distribution $\mathcal{S}$ (over R_η^d)		
Minimal Entropy $H_\infty(\mathcal{S}) - \omega(\log_2 \lambda)$	$nk \log_2 q$ $+\frac{C'nd \cdot n(\eta - 1)}{s}$	$nd \log_2(d\gamma \log_2 n) + \frac{C'nd \cdot (\eta - 1)}{s}$
Minimal secret bound $\eta_{\min}$	$q^{\frac{k+1/n}{d}}$	$(nd \log_2 n)^{1+\frac{1}{nd}}$
Minimal noise s_0	$s_1 \cdot O\left(\frac{\sqrt{m}n^2 d}{\log_2 q} q^{\frac{k+1/n}{d}}\right)$	$\sqrt{8}(nd \log_2 n)^2 \sqrt{nm} \cdot \eta_\varepsilon(R)$

Table 6.2. Comparison of [LWW20] in the primal ring (Theorem 6.1) and Corollary 5.2 for the hardness of M-LWE$_{n,d,q,m,D_{s_0},\mathcal{S}}$ over power-of-two cyclotomic fields of degree n, with module rank d and secret distribution $\mathcal{S}$. For clarity, we have $C' = \sqrt{2\pi} \log_2 e$. The minimal secret bound $\eta_{\min}$ and minimal noise s_0 are obtained by fixing $\alpha = 1/n$.

[a]: The result for arbitrary number fields is that of Theorem 5.2, but Corollary 5.2 is stated for cyclotomic fields.

Acknowledgments

This work was supported by the European Union PROMETHEUS project (Horizon 2020 Research and Innovation Program, grant 780701). Katharina Boudgoust was funded by the Direction Générale de l'Armement (Pôle de Recherche CYBER). We thank Alexandre Wallet and Damien Stehlé for helpful discussions. We also thank our anonymous referees of Eurocrypt 2022 for their thorough proof reading and constructive feedback.

References

ABD16. M. R. Albrecht, S Bai, and L. Ducas. A subfield lattice attack on overstretched NTRU assumptions - cryptanalysis of some FHE and graded encoding schemes. In *CRYPTO (1)*, volume 9814 of *Lecture Notes in Computer Science*, pages 153–178. Springer, 2016.

ADP18. M. R. Albrecht, A. Deo, and K. G. Paterson. Cold boot attacks on ring and module LWE keys under the NTT. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2018(3):173–213, 2018.

BBC$^+$20. D. J. Bernstein, Brumley B. B., M.-S. Chen, C. Chuengsatiansup, T. Lange, A. Marotzke, B.-Y. Peng, N. Tuveri, C. van Vredendaal, and B.-Y. Yang. *NTRU Prime round-3 candidate to the NIST post-quantum cryptography standardisation project*, 2020.

BD20a. Z. Brakerski and N. Döttling. Hardness of LWE on general entropic distributions. In *EUROCRYPT (2)*, volume 12106 of *Lecture Notes in Computer Science*, pages 551–575. Springer, 2020.

BD20b. Z. Brakerski and N. Döttling. Lossiness and entropic hardness for ring-lwe. In *TCC (1)*, volume 12550 of *Lecture Notes in Computer Science*, pages 1–27. Springer, 2020.

BDK$^+$18. J. W. Bos, L. Ducas, E. Kiltz, T. Lepoint, V. Lyubashevsky, J. M. Schanck, P. Schwabe, G. Seiler, and D. Stehlé. CRYSTALS - kyber: A cca-secure module-lattice-based KEM. In *EuroS&P*, pages 353–367. IEEE, 2018.

BGV12. Z. Brakerski, C. Gentry, and V. Vaikuntanathan. (leveled) fully homomorphic encryption without bootstrapping. In *ITCS*, pages 309–325. ACM, 2012.

BJRW20. K. Boudgoust, C. Jeudy, A. Roux-Langlois, and W. Wen. Towards classical hardness of module-lwe: The linear rank case. In *ASIACRYPT (2)*, volume 12492 of *Lecture Notes in Computer Science*, pages 289–317. Springer, 2020.

BJRW21. K. Boudgoust, C. Jeudy, A. Roux-Langlois, and W. Wen. On the hardness of module-lwe with binary secret. In *CT-RSA*, volume 12704 of *Lecture Notes in Computer Science*, pages 503–526. Springer, 2021.

BLP$^+$13. Z. Brakerski, A. Langlois, C. Peikert, O. Regev, and D. Stehlé. Classical hardness of learning with errors. In *STOC*, pages 575–584. ACM, 2013.

CDH$^+$20. C. Chen, O. Danba, J. Hoffstein, A. Hülsing, J. Rijneveld, T. Saito, J. M. Schank, P. Schwabe, W. Whyte, K. Xagawa, T. Yamakawa, and Z. Zhang. *NTRU round-3 candidate to the NIST post-quantum cryptography standardisation project*, 2020.

CJL16. J. H. Cheon, J. Jeong, and C. Lee. An algorithm for NTRU problems and cryptanalysis of the GGH multilinear map without an encoding of zero. *IACR Cryptol. ePrint Arch.*, page 139, 2016.

CPS$^+$20. C. Chuengsatiansup, T. Prest, D. Stehlé, A. Wallet, and K. Xagawa. Modfalcon: Compact signatures based on module-ntru lattices. In *AsiaCCS*, pages 853–866. ACM, 2020.

DKL$^+$18. L. Ducas, E. Kiltz, T. Lepoint, V. Lyubashevsky, P. Schwabe, G. Seiler, and D. Stehlé. Crystals-dilithium: A lattice-based digital signature scheme. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2018(1):238–268, 2018.

DM14. Léo Ducas and Daniele Micciancio. Improved short lattice signatures in the standard model. In *CRYPTO (1)*, volume 8616 of *Lecture Notes in Computer Science*, pages 335–352. Springer, 2014.

DORS08. Y. Dodis, R. Ostrovsky, L. Reyzin, and A. D. Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. *SIAM J. Comput.*, 38(1):97–139, 2008.

GKPV10. S. Goldwasser, Y. Tauman Kalai, C. Peikert, and V. Vaikuntanathan. Robustness of the learning with errors assumption. In *ICS*, pages 230–240. Tsinghua University Press, 2010.

HPS98. J. Hoffstein, J. Pipher, and J. H. Silverman. NTRU: A ring-based public key cryptosystem. In *ANTS*, volume 1423 of *Lecture Notes in Computer Science*, pages 267–288. Springer, 1998.

KF17. P. Kirchner and P.-A. Fouque. Revisiting lattice attacks on overstretched NTRU parameters. In *EUROCRYPT (1)*, volume 10210 of *Lecture Notes in Computer Science*, pages 3–26, 2017.

LPR10. V. Lyubashevsky, C. Peikert, and O. Regev. On ideal lattices and learning with errors over rings. In *EUROCRYPT*, volume 6110 of *Lecture Notes in Computer Science*, pages 1–23. Springer, 2010.

LS15. A. Langlois and D. Stehlé. Worst-case to average-case reductions for module lattices. *Des. Codes Cryptogr.*, 75(3):565–599, 2015.

LWW20. H. Lin, Y. Wang, and M. Wang. Hardness of module-lwe and ring-lwe on general entropic distributions. *IACR Cryptol. ePrint Arch.*, page 1238, 2020.

Mic18. D. Micciancio. On the hardness of learning with errors with binary secrets. *Theory Comput.*, 14(1):1–17, 2018.

MP12. D. Micciancio and C. Peikert. Trapdoors for lattices: Simpler, tighter, faster, smaller. In *EUROCRYPT*, volume 7237 of *Lecture Notes in Computer Science*, pages 700–718. Springer, 2012.

MR07. D. Micciancio and O. Regev. Worst-case to average-case reductions based on gaussian measures. *SIAM J. Comput.*, 37(1):267–302, 2007.

NIS. NIST. Post-quantum cryptography standardization. https://csrc.nist.gov/Projects/Post-Quantum-Cryptography/Post-Quantum-Cryptography-Standardization.

Pei08. C. Peikert. Limits on the hardness of lattice problems in l_p norms. *Comput. Complex.*, 17(2):300–351, 2008.

Pei10. C. Peikert. An efficient and parallel gaussian sampler for lattices. In *CRYPTO*, volume 6223 of *Lecture Notes in Computer Science*, pages 80–97. Springer, 2010.

Pei16. C. Peikert. A decade of lattice cryptography. *Found. Trends Theor. Comput. Sci.*, 10(4):283–424, 2016.

PS21. A. Pellet-Mary and D. Stehlé. On the hardness of the NTRU problem. *IACR Cryptol. ePrint Arch.*, page 821, 2021.

Reg05. O. Regev. On lattices, learning with errors, random linear codes, and cryptography. In *STOC*, pages 84–93. ACM, 2005.

RV08. M. Rudelson and R. Vershynin. The littlewood-offord problem and invertibility of random matrices. *Advances in Mathematics*, 218:600–633, 2008.

Sag20. Sage Developers. *SageMath, the Sage Mathematics Software System (Version 9.2)*, 2020. https://www.sagemath.org.

SSTX09. D. Stehlé, R. Steinfeld, K. Tanaka, and K. Xagawa. Efficient public key encryption based on ideal lattices. In *ASIACRYPT*, volume 5912 of *Lecture Notes in Computer Science*, pages 617–635. Springer, 2009.

vNG47. J. von Neumann and H. H. Goldstine. Numerical inverting of matrices of high order. *Bull. Amer. Math. Soc.*, 53:1021–1099, 1947.

A Alternative Distribution

In this section, we give an different distribution than the one from Definition 5.1 to highlight the trade-off between the underlying hardness assumptions and the parameters.

Definition A.1. *Let K be a number field of degree n, and R its ring of integers. Let d, q, m be positive integers. Let $\gamma > 0$ be a Gaussian parameter, and ψ be a a distribution on R such that $\mathbb{P}_{e \hookleftarrow \psi}[\|\sigma(e)\|_\infty > B] \leq \delta$, for $B \geq 0$, and $\delta \in [0, 1)$. We define the distribution $\varPsi$ as follows:*

- *Choose $\mathbf{F}, \mathbf{G} \hookleftarrow \mathcal{D}_{R,\gamma}^{d \times d}$ such that $\mathbf{F} \in GL_d(R, q)$;*
- *Choose $\mathbf{e}_1, \ldots, \mathbf{e}_m, \mathbf{e}'_1, \ldots, \mathbf{e}'_m \hookleftarrow \psi^d$; Define the matrix $\mathbf{E} \in R^{m \times d}$ whose rows are the $\mathbf{e}_i^T$, and $\mathbf{E}' \in R^{m \times d}$ whose rows are the $(\mathbf{e}'_i)^T$.*
- *Define $\tilde{\mathbf{G}} = \mathbf{E}\mathbf{G} + \mathbf{E}'\mathbf{F}$.*
- *Output $(M_R(\mathbf{F}), M_R(\tilde{\mathbf{G}}))$.*

Note that if $\psi = \mathcal{D}_{R,s_1}$, then by Lemma 5.1 one can set $B = s_1 \log_2 n$ and $\delta = \mathsf{negl}(n)$. As discussed, conditioning $\mathbf{F}$ to be invertible modulo q can be done by taking γ sufficiently large depending on the splitting of q. We show that this distribution now relies on a different M-NTRU assumption that may be considered more standard, but at the expense of requiring the hardness of HNF-M-LWE. More precisely, there is a reduction from M-NTRU and HNF-M-LWE to S-NTRU for this distribution $\varPsi$.

Lemma A.1. *Let K be a number field of degree n, and R its ring of integers. Let d, q, m be positive integers. Let $\gamma > 0$ be a Gaussian parameter, and ψ be a a distribution on R. Define $\mathcal{M} = M_R(U(R_q))$ as in Section 2.1. Assuming the hardness of (decisional)* $\text{HNF-M-LWE}_{n,d,q,d,\psi}$ *and* $\text{M-NTRU}_{n,d,q,d,\mathcal{D}_{R,\gamma}}$*, it holds that* $\text{S-NTRU}_{n,d,q,m,\mathcal{M},\varPsi}$ *is hard.*

Proof. Let $(M_R(\mathbf{F}), M_R(\tilde{\mathbf{G}}))$ be sampled from $\varPsi$. Let $\mathbf{F}_q^{-1}$ be the R_q-inverse of $\mathbf{F}$. We have that $\mathbf{Y} = \tilde{\mathbf{G}}\mathbf{F}_q^{-1} \bmod q$ equals $\mathbf{E}\mathbf{G}\mathbf{F}_q^{-1} + \mathbf{E}' \bmod q$. Under the assumption that (decisional) $\text{M-NTRU}_{n,d,q,d,\mathcal{D}_{R,\gamma}}$ is hard, it holds that $\mathbf{G}\mathbf{F}_q^{-1} \bmod q$ is indistinguishable from a uniformly random matrix $\mathbf{A}^T \in R_q^{d \times d}$. Then, assuming the hardness of $\text{HNF-M-LWE}_{n,d,q,d,\psi}$ (up to transpose), a hybrid argument (over m) yields that $\mathbf{Y} = \mathbf{E}\mathbf{A}^T + \mathbf{E}'$ is indistinguishable from a uniform matrix $\mathbf{U}$ over $R_q^{m \times d}$. Applying the ring homomorphism M_R yields $M_R(\mathbf{Y}) = M_R(\tilde{\mathbf{G}})M_R(\mathbf{F}_q^{-1}) \bmod q$ is indistinguishable from $M_R(\mathbf{U})$. Since $M_R(\mathbf{F}_q^{-1})$ is the $\mathbb{Z}_q$-inverse of $M_R(\mathbf{F})$, and $M_R(\mathbf{U})$ is exactly distributed according to $\mathcal{M}^{m \times d}$, it indeed proves that the problem $\text{S-NTRU}_{n,d,q,m,\mathcal{M},\varPsi}$ is hard. □

For completeness, we derive the equivalent statement of Lemma 5.3 for this distribution. By Theorem 4.3, it leads to a sometimes lossy pseudorandom distribution thus proving the entropic hardness of M-LWE for different parameters.

Lemma A.2. *Let K be the ν-th cyclotomic field of degree $n = \varphi(\nu)$, and R its ring of integers. Let d, q, m be positive integers. Let $\gamma > 0$ be a Gaussian parameter. Let ψ be a a distribution on R such that $\mathbb{P}_{e \hookleftarrow \psi}[\|\sigma(e)\|_\infty > B] \leq \delta$, for $B \geq 0$, and $\delta \in [0, 1)$. Let $(M_R(\mathbf{F}), M_R(\tilde{\mathbf{G}})) \hookleftarrow \Psi$. It holds that*

1. $\left\|\mathbf{B}_{R^d}^{-1} M_R(\mathbf{F}) \mathbf{B}_{R^d}\right\|_2 \leq d\gamma \log_2 n$
2. $\left\|\mathbf{B}_{R^m}^{-1} M_R(\tilde{\mathbf{G}}) M_R(\mathbf{F})^{-1} \mathbf{B}_{R^d}\right\|_2 \leq O(Bn\sqrt{nmd}^2 \log_2 n)$

except with probability at most $O(n^{-1/2}) + 2md\delta + 4d^2 n e^{-\pi \log_2^2 n}$ over the randomness of $(M_R(\mathbf{F}), M_R(\tilde{\mathbf{G}}))$. When $d = \mathsf{poly}(n)$, this probability is $O(n^{-1/2}) + 2md\delta + \mathsf{negl}(n)$.

Proof. **1.** Note that $M_R(\mathbf{F})$ is distributed the same way as in Definition 5.1. The argument for 1. from the proof of Lemma 5.3 is therefore identical, yielding $\mathbb{P}_{\mathbf{F} \hookleftarrow \mathcal{D}_{R,\gamma}^{d \times d}}[\left\|\mathbf{B}_{R^d}^{-1} M_R(\mathbf{F}) \mathbf{B}_{R^d}\right\|_2 > d\gamma \log_2 n] \leq 2d^2 n e^{-\pi \log_2^2 n}$.

2. We now bound $\left\|\mathbf{B}_{R^m}^{-1} M_R(\tilde{\mathbf{G}}) M_R(\mathbf{F})^{-1} \mathbf{B}_{R^d}\right\|_2$ from above. First, it holds that $\tilde{\mathbf{G}}\mathbf{F}^{-1} = \mathbf{E}\mathbf{G}\mathbf{F}^{-1} + \mathbf{E}'$. Yet we also have

$$\begin{aligned}\mathbf{B}_{R^m}^{-1} M_R(\tilde{\mathbf{G}}) M_R(\mathbf{F})^{-1} \mathbf{B}_{R^d} &= \mathbf{B}_{R^m}^{-1} M_R(\tilde{\mathbf{G}}\mathbf{F}^{-1}) \mathbf{B}_{R^d} \\ &= M_{\sigma_H}(\tilde{\mathbf{G}}\mathbf{F}^{-1}).\end{aligned}$$

Due to the homomorphic properties of M_{σ_H}, and the fact that $\|.\|_2$ is a submultiplicative norm, it holds

$$\left\|M_{\sigma_H}(\tilde{\mathbf{G}}\mathbf{F}^{-1})\right\|_2 \leq \|M_{\sigma_H}(\mathbf{E})\|_2 \cdot \|M_{\sigma_H}(\mathbf{G})\|_2 \cdot \left\|M_{\sigma_H}(\mathbf{F}^{-1})\right\|_2 + \|M_{\sigma_H}(\mathbf{E}')\|_2$$

As $\mathbf{U}_H$ is unitary and $\|M_\sigma(e_{ij})\|_2 = \|\sigma(e_{ij})\|_\infty$, each block of $M_{\sigma_H}(\mathbf{E})$ has a spectral norm of $\|\sigma(e_{ij})\|_\infty$ respectively. Then, using Lemma 2.1 and the fact that the entries of $\mathbf{E}$ are i.i.d. and B-bounded except with probability at most δ, we have

$$\mathbb{P}[\|M_{\sigma_H}(\mathbf{E})\|_2 \geq B\sqrt{md}] \leq md \cdot \delta$$

The same reasoning holds for $\mathbf{E}'$. Then, the same argument as in 1. yields

$$\|M_{\sigma_H}(\mathbf{G})\|_2 \leq d\gamma \log_2 n,$$

except with probability at most $2d^2 n e^{-\pi \log_2^2 n}$.

Finally, Corollary 5.1 yields $\left\|M_{\sigma_H}(\mathbf{F}^{-1})\right\|_2 \leq n\sqrt{nd}/\gamma$ except with probability of at most $O(n^{-1/2})$. Combining it all gives

$$\begin{aligned}\left\|\mathbf{B}_{R^m}^{-1} M_R(\tilde{\mathbf{G}}) M_R(\mathbf{F})^{-1} \mathbf{B}_{R^d}\right\|_2 &\leq Bn\sqrt{nmd}^2 \log_2 n + B\sqrt{md} \\ &= O(B\sqrt{nmd}^2 \log_2 n),\end{aligned}$$

except with probability at most $O(n^{-1/2}) + 2md\delta + 2d^2 n e^{-\pi \log_2^2 n}$. □

The noise lossiness condition is therefore unchanged, meaning that the entropic argument remains exactly the same. However, using a more standard M-NTRU assumption (for square matrices) requires an extra step in the reduction depending on HNF-M-LWE. This additional step further deteriorates the Gaussian noise parameter achieved by the reduction. It indeed leads to

$$\begin{aligned} s_0 &\geq O(s \cdot Bn\sqrt{nm}d^2 \log_2 n) \\ &= O(s \cdot s_1 n\sqrt{nm}d^2 \log_2^2 n), \end{aligned}$$

which is roughly $s_1 \cdot d \log_2 n$ larger than for the other distribution we proposed.

B Missing Proofs

B.1 Preliminaries

Lemma B.1 ([DORS08,Reg05]). *Let $\mathbb{G}$ be a finite Abelian group, and Y be a finite set. Let $\ell \geq \log|\mathbb{G}| + \log|Y| + \omega(\log \lambda)$ be an integer. Let $\mathbf{g} \hookleftarrow U(\mathbb{G}^\ell)$, and $\mathbf{x} \hookleftarrow U(\{0,1\}^\ell)$. Let y be a random variable supported on Y which is possibly correlated with $\mathbf{x}$ but independent of the g_i. Then it holds that $(\mathbf{g}, \mathbf{x}^T\mathbf{g}, y)$ is statistically close to $(\mathbf{g}, u, y)$, where $u \hookleftarrow U(\mathbb{G})$.*

Lemma B.2 ([BD20b], Claim 1). *Let $\Lambda \subseteq \mathbb{R}^N$ and $\mathbf{F}' \in \mathbb{R}^{N\times N}$. It holds that $\eta_\varepsilon(\mathbf{F}' \cdot \Lambda) \leq \|\mathbf{F}'\|_2 \cdot \eta_\varepsilon(\Lambda)$.*

Proof. We consider full-rank lattices so we can assume that $\mathbf{G}$ is invertible. First note that $(\mathbf{F}' \cdot \Lambda)^* = (\mathbf{F}')^{-T} \cdot \Lambda^*$. So for $s > 0$ it holds that

$$\begin{aligned} \rho_{1/s}((\mathbf{F}'\Lambda)^*) &= \sum_{\mathbf{x}\in\Lambda^*} \exp(-\pi s^2 \mathbf{x}^T (\mathbf{F}')^{-1}(\mathbf{F}')^{-T}\mathbf{x}) \\ &\leq \sum_{\mathbf{x}\in\Lambda^*} \exp(-\pi (s/\|\mathbf{F}'\|_2)^2 \|\mathbf{x}\|_2^2) \\ &= \rho_{\|\mathbf{F}'\|_2/s}(\Lambda^*). \end{aligned}$$

Now take $s = \|\mathbf{F}'\|_2 \eta_\varepsilon(\Lambda)$. It thus yields $\rho_{\|\mathbf{F}'\|_2/s}(\Lambda^* \setminus \{\mathbf{0}\}) \leq \varepsilon$. Hence

$$\rho_{1/s}((\mathbf{F}'\Lambda)^* \setminus \{\mathbf{0}\}) \leq \rho_{\|\mathbf{F}'\|_2/s}(\Lambda^* \setminus \{\mathbf{0}\}) \leq \varepsilon,$$

which gives $s \geq \eta_\varepsilon(\mathbf{F}'\Lambda)$ by definition, thus concluding the proof. □

We then recall a smoothing result that gives the distribution of the sum of a discrete Gaussian and a continuous one. We also need the convolution theorem by Peikert [Pei10].

Lemma B.3 ([Reg05], Claim 3.9). *Let $\Lambda \subseteq \mathbb{R}^N$ be a lattice and let $\sigma \geq \sqrt{2}\eta_\varepsilon(\Lambda)$. Let $\mathbf{e} \sim \mathcal{D}_{\Lambda,\sigma}$ be a discrete gaussian and $\mathbf{e}' \sim D_{\mathbb{R}^N,\sigma}$ be a continuous gaussian. Then $\mathbf{e} + \mathbf{e}'$ is 2ε-close to $D_{\mathbb{R}^N,\sqrt{2}\sigma}$.*

Theorem B.1 ([Pei10], Theorem 3.1). *Let $\mathbf{\Sigma}_1, \mathbf{\Sigma}_2 > 0$ be two positive definite matrices such that $\mathbf{\Sigma} = \mathbf{\Sigma}_1 + \mathbf{\Sigma}_2 > 0$ and $\mathbf{\Sigma}_1^{-1} + \mathbf{\Sigma}_2^{-1} > 0$. Let Λ_1, Λ_2 be two lattices such that $\sqrt{\mathbf{\Sigma}_1} \geq \eta_\varepsilon(\Lambda_1)$ and $\sqrt{\mathbf{\Sigma}_2} \geq \eta_\varepsilon(\Lambda_2)$ for some $\varepsilon > 0$. Let $\mathbf{c}_1, \mathbf{c}_2 \in \mathbb{R}^N$ be arbitrary. Consider the following sampling procedure for $\mathbf{x} \in \Lambda_2 + \mathbf{c}_2$.*

- *Choose $\mathbf{x}_1 \hookleftarrow \mathcal{D}_{\Lambda_1+\mathbf{c}_1, \sqrt{\mathbf{\Sigma}_1}}$.*
- *Choose $\mathbf{x} \hookleftarrow \mathbf{x}_1 + \mathcal{D}_{\Lambda_2+\mathbf{c}_2-\mathbf{x}_1, \sqrt{\mathbf{\Sigma}_2}}$.*

Then it holds that the marginal distribution of $\mathbf{x}$ is within statistical distance 8ε to $\mathcal{D}_{\Lambda_2+\mathbf{c}_2, \sqrt{\mathbf{\Sigma}}}$. It still holds if $\mathbf{x}_1$ is sampled from the continuous Gaussian $D_{\sqrt{\mathbf{\Sigma}_1}}$.

We were not able to find a result in the literature that bounds the ε-smooth average conditional min-entropy by the (non-smooth) average conditional min-entropy. We thus provide the following lemma along with a straightforward proof showing that, for a sufficiently small ε, the ε-smooth average conditional min-entropy cannot be much larger than the average conditional min-entropy. This fact seems intuitive but our bound seems quite loose and restrictive on the value of ε. We leave it as an interesting open problem to improve it.

Lemma B.4. *Let $(\mathbf{x}, \mathbf{z})$ be two discrete random variables with values in $X \times Z$. For any $\varepsilon \geq \varepsilon' \geq 0$, we have $\widetilde{H}_\infty^{\varepsilon'}(\mathbf{x}|\mathbf{z}) \leq \widetilde{H}_\infty^{\varepsilon}(\mathbf{x}|\mathbf{z})$. Additionally, for all $\varepsilon < |Z|^{-1} 2^{-\widetilde{H}_\infty(\mathbf{x}|\mathbf{z})}$, it holds*

$$\widetilde{H}_\infty^{\varepsilon}(\mathbf{x}|\mathbf{z}) \leq \widetilde{H}_\infty(\mathbf{x}|\mathbf{z}) - \log_2 \left(1 - \varepsilon |Z| 2^{\widetilde{H}_\infty(\mathbf{x}|\mathbf{z})}\right).$$

In particular, if $\varepsilon = (1 - p(\lambda)^{-1}) |Z|^{-1} 2^{-\widetilde{H}_\infty(\mathbf{x}|\mathbf{z})}$, for a function $p(\lambda) > 1$, then

$$\widetilde{H}_\infty^{\varepsilon}(\mathbf{x}|\mathbf{z}) \leq \widetilde{H}_\infty(\mathbf{x}|\mathbf{z}) + \log_2 p(\lambda).$$

Proof. Let $\varepsilon \geq \varepsilon' \geq 0$. Consider $(\mathbf{a}, \mathbf{b})$ random variables over $X \times Z$ such that $\Delta((\mathbf{a}, \mathbf{b}), (\mathbf{x}, \mathbf{z})) \leq \varepsilon'$. Then, we have that $\Delta((\mathbf{a}, \mathbf{b}), (\mathbf{x}, \mathbf{z})) \leq \varepsilon$. It then gives $\widetilde{H}_\infty(\mathbf{a}|\mathbf{b}) \leq \widetilde{H}_\infty^{\varepsilon}(\mathbf{x}|\mathbf{z})$. Maximizing over such $(\mathbf{a}, \mathbf{b})$ yields the first inequality.

Now, let $\varepsilon < |Z|^{-1} 2^{-\widetilde{H}_\infty(\mathbf{x}|\mathbf{z})}$. Consider $(\mathbf{a}, \mathbf{b})$ random variables over $X \times Z$ such that $\Delta((\mathbf{a}, \mathbf{b}), (\mathbf{x}, \mathbf{z})) \leq \varepsilon$. First, notice that

$$\begin{aligned}
\widetilde{H}_\infty(\mathbf{a}|\mathbf{b}) &= -\log_2 \mathbb{E}_{\mathbf{b}} \left[\max_{\mathbf{a}' \in X} \mathbb{P}[\mathbf{a} = \mathbf{a}' | \mathbf{b}]\right] \\
&= -\log_2 \sum_{\mathbf{b}' \in Z} \mathbb{P}[\mathbf{b} = \mathbf{b}'] \cdot \max_{\mathbf{a}' \in X} \mathbb{P}[\mathbf{a} = \mathbf{a}' | \mathbf{b} = \mathbf{b}'] \\
&= -\log_2 \sum_{\mathbf{b}' \in Z} \max_{\mathbf{a}' \in X} \mathbb{P}[(\mathbf{a}, \mathbf{b}) = (\mathbf{a}', \mathbf{b}')].
\end{aligned}$$

For all $(\mathbf{a}', \mathbf{b}') \in X \times Z$, we have $|\mathbb{P}[(\mathbf{a}, \mathbf{b}) = (\mathbf{a}', \mathbf{b}')] - \mathbb{P}[(\mathbf{x}, \mathbf{z}) = (\mathbf{a}', \mathbf{b}')]| \leq \Delta((\mathbf{a}, \mathbf{b}), (\mathbf{x}, \mathbf{z})) \leq \varepsilon$. Hence, maximizing over $\mathbf{a}'$ and summing over $\mathbf{b}'$ gives

$$\mathbb{E}_{\mathbf{b}} \left[\max_{\mathbf{x}' \in X} \mathbb{P}[\mathbf{x} = \mathbf{x}' | \mathbf{z}]\right] - \varepsilon |Z| \leq \mathbb{E}_{\mathbf{b}} \left[\max_{\mathbf{a}' \in X} \mathbb{P}[\mathbf{a} = \mathbf{a}' | \mathbf{b}]\right]$$

By our condition on ε, we can obtain

$$\begin{aligned}\widetilde{H}_\infty(\mathbf{a}|\mathbf{b}) &\le -\log_2\left(\mathbb{E}_{\mathbf{b}}\left[\max_{\mathbf{x}'\in X}\mathbb{P}[\mathbf{x}=\mathbf{x}'|\mathbf{z}]\right] - \varepsilon|Z|\right)\\ &= -\log_2\left(2^{-\widetilde{H}_\infty(\mathbf{x}|\mathbf{y})} - \varepsilon|Z|\right).\end{aligned}$$

Finally, by maximizing over all $(\mathbf{a},\mathbf{b})$ that are ε-close to $(\mathbf{x},\mathbf{z})$, we get

$$\widetilde{H}^{\varepsilon}_\infty(\mathbf{x}|\mathbf{z}) \le -\log_2\left(2^{-\widetilde{H}_\infty(\mathbf{x}|\mathbf{y})} - \varepsilon|Z|\right) = \widetilde{H}_\infty(\mathbf{x}|\mathbf{z}) - \log_2\left(1-\varepsilon|Z|2^{\widetilde{H}_\infty(\mathbf{x}|\mathbf{z})}\right),$$

as claimed. □

B.2 Proofs of Section 2

Equation (1)

Proof. Let $f = x^n + \sum_{k=0}^{n-1} f_k x^k$ denote the minimal polynomial of ζ, and $K = \mathbb{Q}(\zeta)$. Let $\mathbf{C}$ denote the companion matrix of f, as in Equation 1. It is well known that the characteristic (and minimal) polynomial of the companion matrix of f is f itself. This entails that $\mathbf{C}$ has the roots of f for eigenvalues, which we denote by $\alpha_1, \ldots, \alpha_n$. Recall that the field embeddings are such that $\sigma_i(\zeta) = \alpha_i$ for all $i \in [n]$. Since the roots of f are distinct, it means that $\mathbf{C}$ is diagonalizable. More precisely, it holds that $\mathbf{C} = \mathbf{V}^{-1}\mathrm{diag}(\alpha_1, \ldots, \alpha_n)\mathbf{V} = \mathbf{V}^{-1}\mathrm{diag}(\sigma(\zeta))\mathbf{V}$. Now let x be in K. We have

$$\forall y \in K,\ \tau(xy) = \mathbf{V}^{-1}\sigma(xy) = \mathbf{V}^{-1}\mathrm{diag}(\sigma(x))\sigma(y) = \mathbf{V}^{-1}\mathrm{diag}(\sigma(x))\mathbf{V}\tau(y),$$

thus proving that $M_\tau(x) = \mathbf{V}^{-1}\sigma(x)\mathbf{V}$. We can then rewrite this expression in terms of the τ_k and $\mathbf{C}$ as follows.

$$\begin{aligned}\mathbf{V}^{-1}\mathrm{diag}(\sigma(x))\mathbf{V} &= \mathbf{V}^{-1}\mathrm{diag}\left(\sigma_1\left(\sum_{k=0}^{n-1}\tau_k(x)\zeta^k\right), \ldots, \sigma_n\left(\sum_{k=0}^{n-1}\tau_k(x)\zeta^k\right)\right)\mathbf{V}\\ &= \sum_{k=0}^{n-1}\tau_k(x)\mathbf{V}^{-1}\mathrm{diag}(\sigma_1(\zeta)^k, \ldots, \sigma_n(\zeta)^k)\mathbf{V}\\ &= \sum_{k=0}^{n-1}\tau_k(x)\mathbf{V}^{-1}\mathrm{diag}(\sigma(\zeta))^k\mathbf{V}\\ &= \sum_{k=0}^{n-1}\tau_k(x)\mathbf{C}^k,\end{aligned}$$

concluding the proof. □

Lemma 2.1

Proof. Fix any vector $\mathbf{x} = (\mathbf{x}_1^T, \ldots, \mathbf{x}_b^T) \in \mathbb{R}^{db}$, where the $\mathbf{x}_i$ are in $\mathbb{R}^d$. Then it holds that $\|\mathbf{Z}\mathbf{x}\|_2^2 = \sum_{i\in[a]} \left\| \sum_{j\in[b]} \mathbf{Z}_{ij}\mathbf{x}_j \right\|_2^2$. Yet

$$\left\| \sum_{j\in[b]} \mathbf{Z}_{ij}\mathbf{x}_j \right\|_2 \leq \sum_{i\in[b]} \|\mathbf{Z}_{ij}\mathbf{x}_j\|_2 \leq \sum_{j\in[b]} \|\mathbf{Z}_{ij}\|_2 \|\mathbf{x}_j\|_2 \leq \sqrt{\sum_{j\in[b]} \|\mathbf{Z}_{ij}\|_2^2} \sqrt{\sum_{j\in[b]} \|\mathbf{x}_j\|_2^2}$$

$$\leq \sqrt{\sum_{j\in[b]} \|\mathbf{Z}_{ij}\|_2^2} \cdot \|\mathbf{x}\|_2,$$

So by a union bound over the $(\mathbf{Z}_{ij})_j$, it holds that

$$\mathbb{P}\left[\left\| \sum_{j\in[b]} \mathbf{Z}_{ij}\mathbf{x}_j \right\|_2 \geq \gamma\sqrt{b}\|\mathbf{x}\|_2 \right] \leq b\delta.$$

Another union bound over $i \in [a]$ yields $\|\mathbf{Z}\mathbf{x}\|_2^2 \leq \gamma^2 ab\|x\|_2^2$, hence proving that $\|\mathbf{Z}\|_2 \leq \gamma\sqrt{ab}$, except with probability $ab\delta$ over the choice of $\mathbf{Z}$. □

B.3 Proofs of Section 3

Lemma 3.2

Proof. Assume that S-LWE$_{n,d,q,m,\mathcal{M},\Upsilon,\mathcal{S}}$ is standard hard. Let $\mathcal{A}$ be a PPT adversary and μ a negligible function. Then, $\varepsilon = 1 - \mu$ is a non-negligible function. By our assumption, there exists a negligible function ν such that

$$\mathbb{P}_{\substack{\mathbf{s}\hookleftarrow\mathcal{S} \\ \psi\hookleftarrow\Upsilon}} \left[\mathbb{P}_{\mathcal{A},\mathcal{O}_{\mathbf{s},\psi}} \left[\mathcal{A}^{\mathcal{O}_{\mathbf{s},\psi}}(1^\lambda) = \mathbf{s} \right] \geq \varepsilon(\lambda) \right] \leq \nu(\lambda).$$

This means that for all PPT adversary $\mathcal{A}$ and every negligible function μ, there exists a negligible function ν that verifies

$$\mathbb{P}_{\substack{\mathbf{s}\hookleftarrow\mathcal{S} \\ \psi\hookleftarrow\Upsilon}} \left[\mathbb{P}_{\mathcal{A},\mathcal{O}_{\mathbf{s},\psi}} \left[\mathcal{A}^{\mathcal{O}_{\mathbf{s},\psi}}(1^\lambda) = \mathbf{s} \right] \geq 1 - \mu(\lambda) \right] \leq \nu(\lambda),$$

thus proving the mild hardness of S-LWE$_{n,d,q,m,\mathcal{M},\Upsilon,\mathcal{S}}$. □

Lemma 3.3

Proof. Assume towards contradiction there was a PPT adversary $\mathcal{A}$ breaks the standard hardness of search S-LWE$_{n,d,q,\mathcal{M},\Upsilon,\mathcal{S}}$. It means that there is a non-negligible function $\varepsilon(\lambda) = 1/\mathsf{poly}(\lambda)$ such that for all negligible function ν we have

$$\mathbb{P}_{\substack{\mathbf{s}\hookleftarrow\mathcal{S} \\ \psi\hookleftarrow\Upsilon}} \left[\mathbb{P}_{\mathcal{A},\mathcal{O}_{\mathbf{s},\psi}} \left[\mathcal{A}^{\mathcal{O}_{\mathbf{s},\psi}}(1^\lambda) = \mathbf{s} \right] \geq \varepsilon(\lambda) \right] > \nu(\lambda).$$

We define $\kappa(\lambda) = \lambda/\varepsilon(\lambda)$. Note that $\kappa(\lambda)$ is polynomial in λ, and we can assume that $\kappa(\lambda) \geq \lambda$, as we have $\varepsilon(\lambda) \leq 1$ without loss of generality. For any $(\mathbf{s}, \psi)$ sampled from $\mathcal{S} \times \Upsilon$, we say that $(\mathbf{s}, \psi)$ is *good* if it holds that

$$\mathbb{P}_{\mathcal{A},\mathcal{O}_{\mathbf{s},\psi}} \left[\mathcal{A}^{\mathcal{O}_{\mathbf{s},\psi}}(1^\lambda) = \mathbf{s} \right] \geq \varepsilon(\lambda).$$

We now construct the adversary $\mathcal{B}$, which is given access to an oracle $\mathcal{O}_{\mathbf{s},\psi}$ for some $(\mathbf{s},\psi) \hookleftarrow \mathcal{S} \times \varUpsilon$.

Algorithm $\mathcal{B}^{\mathcal{O}_{\mathbf{s},\psi}}$

- Repeat $\kappa(\lambda)$ times:
  - Compute $\mathbf{s}' \leftarrow \mathcal{A}^{\mathcal{O}_{\mathbf{s},\psi}}(1^\lambda)$
  - Query λ additional samples and test whether $\mathbf{s}'$ is a valid solution. If so output $\mathbf{s} \leftarrow \mathbf{s}'$
- If none of the iterations stopped the algorithm by returning a value, output $\perp$.

For convenience, we denote by $\mathbf{s}_i$ the value of $\mathbf{s}'$ at the i-th iteration. We now analyze the advantage of $\mathcal{B}$. For that, fix a *good* $(\mathbf{s},\psi)$ as defined above. We show that the probability that none of the repetitions returns $\mathbf{s}$ is negligible. For a given $(\mathbf{s},\psi)$, all of the repetitions are independent, which means thatthe outcomes are independent. As such, it holds that

$$\begin{aligned}
\mathbb{P}[\forall i \in [\kappa(\lambda)] : \mathbf{s}_i \neq \mathbf{s}] = &\prod_{i \in [\kappa(\lambda)]} \mathbb{P}[\mathbf{s}_i \neq \mathbf{s}] \\
&\prod_{i \in [\kappa(\lambda)]} \mathbb{P}[\mathcal{A}^{\mathcal{O}_{\mathbf{s},\psi}}(1^\lambda) \neq \mathbf{s}] \\
&\leq (1 - \varepsilon(\lambda))^{\kappa(\lambda)} \qquad (4) \\
&= (1 - \lambda/\kappa(\lambda))^{\kappa(\lambda)} \\
&\leq e^{-\lambda}, \qquad (5)
\end{aligned}$$

where inequality (4) follows from the assumption that $(\mathbf{s},\psi)$ is *good*, and inequality (5) stems from the fact that for all $y \geq x > 0$, $(1 - x/y)^y \leq e^{-x}$. Hence, for $(\mathbf{s},\psi)$ *good*, we have

$$\mathbb{P}_{\mathcal{B},\mathcal{O}_{\mathbf{s},\psi}}\left[\mathcal{B}^{\mathcal{O}_{\mathbf{s},\psi}}(1^\lambda) = \mathbf{s}\right] = 1 - \mathbb{P}_{\mathcal{B},\mathcal{O}_{\mathbf{s},\psi}}[\forall i \in [\kappa(\lambda)] : \mathbf{s}_i \neq \mathbf{s}] \geq 1 - e^{-\lambda}.$$

It thus holds that

$$\mathbb{P}_{\substack{\mathbf{s} \hookleftarrow \mathcal{S} \\ \psi \hookleftarrow \varUpsilon}}\left[\mathbb{P}_{\mathcal{B},\mathcal{O}_{\mathbf{s},\psi}}\left[\mathcal{B}^{\mathcal{O}_{\mathbf{s},\psi}}(1^\lambda) = \mathbf{s}\right] \geq 1 - e^{-\lambda}\right] \geq \mathbb{P}_{\substack{\mathbf{s} \hookleftarrow \mathcal{S} \\ \psi \hookleftarrow \varUpsilon}}\left[(\mathbf{s},\psi) \text{ is } \textit{good}\right].$$

Yet by assumption, for all negligible functions ν, we have that

$$\mathbb{P}_{\substack{\mathbf{s} \hookleftarrow \mathcal{S} \\ \psi \hookleftarrow \varUpsilon}}[(\mathbf{s},\psi) \text{ is } \textit{good}] = \mathbb{P}_{\substack{\mathbf{s} \hookleftarrow \mathcal{S} \\ \psi \hookleftarrow \varUpsilon}}\left[\mathbb{P}_{\mathcal{A},\mathcal{O}_{\mathbf{s},\psi}}\left[\mathcal{A}^{\mathcal{O}_{\mathbf{s},\psi}}(1^\lambda) = \mathbf{s}\right] \geq \varepsilon(\lambda)\right] > \nu(\lambda).$$

By defining $\mu(\lambda) = e^{-\lambda}$, which is a negligible function, we have proven that there exists an adversary $\mathcal{B}$ that runs in polynomial time (as $\kappa(\lambda) = \mathsf{poly}(\lambda)$) and a negligible function μ, such that for all negligible function ν

$$\mathbb{P}_{\substack{\mathbf{s} \hookleftarrow \mathcal{S} \\ \psi \hookleftarrow \varUpsilon}}\left[\mathbb{P}_{\mathcal{B},\mathcal{O}_{\mathbf{s},\psi}}\left[\mathcal{B}^{\mathcal{O}_{\mathbf{s},\psi}}(1^\lambda) = \mathbf{s}\right] \geq 1 - \mu(\lambda)\right] > \nu(\lambda).$$

This means that $\mathcal{B}$ breaks the mild hardness of S-LWE$_{n,d,q,\mathcal{M},\psi,\mathcal{S}}$, thus yielding a contradiction. □

Remark B.1. The reason why the proof does not hold for a bounded number of samples m is because if $\mathcal{A}$ is allowed m samples, then $\mathcal{B}$ would make $\kappa(\lambda)(m+\lambda)$ queries to the oracle. Since $\kappa(\lambda)$ depends on the obtained advantage of $\mathcal{A}$, the number of samples needed by $\mathcal{B}$ cannot be set a priori.

Lemma 3.4

Proof. Assume towards contradiction that there is an adversary $\mathcal{A}$ that breaks the mild hardness of S-LWE$_{n,d,q,\mathcal{M},\Upsilon_{\Phi,bin},\mathcal{S}}$. So there is a negligible function μ such that for all negligible functions ν,

$$\mathbb{P}_{\substack{\mathbf{s}\hookleftarrow\mathcal{S}\\ \psi\hookleftarrow\Upsilon}}\left[\mathbb{P}_{\mathcal{A},\mathcal{O}_{\mathbf{s},\psi}}\left[\mathcal{A}^{\mathcal{O}_{\mathbf{s},\psi}}(1^\lambda)=\mathbf{s}\right]\geq 1-\mu(\lambda)\right]>\nu(\lambda).$$

We construct an adversary $\mathcal{B}$ against the mild hardness of S-LWE$_{n,d,q,m,\mathcal{M},\Phi,\mathcal{S}}$. Note that in this case, the *distribution of distributions* Υ can be seen as deterministic outputting Φ with probability 1. The adversary proceeds as follows.
Algorithm $\mathcal{B}$

- Input: m samples $(\mathbf{A}_1,\mathbf{y}_1),\ldots,(\mathbf{A}_m,\mathbf{y}_m)$.
- Setup an oracle $\mathcal{O}$, which when queried chooses a uniformly random $\mathbf{x}\in\{0,1\}^m$ and outputs $(\sum_i x_i\mathbf{A}_i \bmod q, \sum_i x_i\mathbf{y}_i \bmod q)$.
- Compute and output $\mathbf{s}\leftarrow\mathcal{A}^{\mathcal{O}}(1^\lambda)$.

We now show that $\mathcal{B}$ simulates the oracle $\mathcal{O}$ of the problem S-LWE$_{n,d,q,\mathcal{M},\Upsilon_{\Phi,bin},\mathcal{S}}$. We define $\mathbf{y}_i=\mathbf{A}_i\mathbf{s}+\mathbf{e}_i \bmod q$. Then the rerandomized sample

$$\left(\sum_i x_i\mathbf{A}_i \bmod q, \sum_i x_i\mathbf{y}_i=(\sum_i x_i\mathbf{A}_i)\mathbf{s}+\sum_i x_i\mathbf{e}_i \bmod q\right),$$

has an error term $\mathbf{e}^*=\sum_i x_i\mathbf{e}_i$ which follows a distribution ψ of $\Upsilon_{\Phi,bin}$, where ψ is defined by $\mathbf{e}_1,\ldots,\mathbf{e}_m\in\mathbb{Z}^n$ that are sampled from Φ. By Lemma B.1, the distribution of $\sum_i x_i\mathbf{A}_i$ is statistically close to the uniform distribution over $\mathbb{G}=\text{support}(\mathcal{M}^d)$ given the $\mathbf{e}^*$. This is because $|\mathbb{G}|\leq\left|\mathbb{Z}_q^{n\times nd}\right|=q^{n^2d}$. Also, as the hint $\sum_i x_i\mathbf{y}_i$ is supported on $\mathbb{Z}_q^n$, $|\mathcal{Y}|\leq q^n$. Since we have $m\geq n^2d\log_2 q+n\log_2 q+\omega(\log_2\lambda)\geq\log_2|\mathbb{G}|+\log_2|\mathcal{Y}|+\omega(\log_2\lambda)$, we can also apply the leftover hash lemma in our setting. We conclude that the distribution of the samples generated by $\mathcal{O}$ is statistically close to the correct distribution. Therefore, by defining $\mu'(\lambda)=\mu(\lambda)+\Delta(((\mathbf{A}_i)_i,\sum_i x_i\mathbf{A}_i \bmod q,\mathbf{e}^*),((\mathbf{A}_i)_i,U(\mathbb{G}),\mathbf{e}^*))$, which is still negligible as explained above, we have that for all negligible functions ν

$$\begin{aligned}\mathbb{P}_{\substack{\mathbf{s}\hookleftarrow\mathcal{S}\\ \Phi\hookleftarrow\{\Phi\}}}\left[\mathbb{P}_{\mathcal{B},\mathcal{O}_{\mathbf{s},\Phi}}\left[\mathcal{B}^{\mathcal{O}_{\mathbf{s},\Phi}}(1^\lambda)\right.\right. &= \left.\mathbf{s}\right]\geq 1-\mu'(\lambda)] \\ &\geq \mathbb{P}_{\substack{\mathbf{s}\hookleftarrow\mathcal{S}\\ \psi\hookleftarrow\Upsilon_{\Phi,bin}}}\left[\mathbb{P}_{\mathcal{A},\mathcal{O}_{\mathbf{s},\psi}}\left[\mathcal{A}^{\mathcal{O}_{\mathbf{s},\psi}}(1^\lambda)=\mathbf{s}\right]\geq 1-\mu(\lambda)\right] \\ &> \nu(\lambda),\end{aligned}$$

which concludes the proof. □

B.4 Proofs of Section 4

Theorem 4.1

Proof. Let $\delta = 1/\mathsf{poly}(\lambda)$ be as in Definition 4.1, and define $\ell = \lambda/\delta = \mathsf{poly}(\lambda)$. By a standard hybrid argument, it holds that

$$(\mathbf{A}^{(i)})_{i\in[\ell]} \approx_c (\mathbf{U}^{(i)})_{i\in[\ell]},$$

where $\mathbf{A}^{(i)} \hookleftarrow \mathcal{X}$ and $\mathbf{U}^{(i)} \hookleftarrow \mathcal{M}^{m\times d}$ for all $i = 1,\ldots,\ell$. The argument then makes use of the fact that by our choice of ℓ, there is only a negligible probability that all of the $\mathbf{A}^{(i)}$ are not lossy.
Assume towards contradiction that S-LWE$_{n,d,q,m,\mathcal{M},\psi,\mathcal{S}}$ is not mildly hard, meaning there exists a PPT adversary $\mathcal{A}$ and a negligible function μ such that for all negligible functions ν

$$\mathbb{P}_{\substack{\mathbf{s}\hookleftarrow\mathcal{S}\\ \psi\hookleftarrow\{\psi\}}}\left[\mathbb{P}_{\substack{\mathbf{A}\hookleftarrow\mathcal{X}\\ \mathbf{e}\hookleftarrow\psi^m}}\left[\mathcal{A}(\mathbf{A},\mathbf{A}\mathbf{s}+\mathbf{e} \bmod q)=\mathbf{s}\right] \geq 1-\mu(\lambda)\right] > \nu(\lambda).$$

We omit the randomness over ψ as it comes from the deterministic distribution $\{\psi\}$ Since the inequality is strict and over all negligible functions, it holds that there exists a non-negligible function $\varepsilon' = 1/\mathsf{poly}(\lambda)$ such that

$$\mathbb{P}_{\mathbf{s}\hookleftarrow\mathcal{S}}\left[\mathbb{P}_{\substack{\mathbf{A}\hookleftarrow\mathcal{X}\\ \mathbf{e}\hookleftarrow\psi^m}}\left[\mathcal{A}(\mathbf{A},\mathbf{A}\mathbf{s}+\mathbf{e} \bmod q)=\mathbf{s}\right] \geq 1-\mu(\lambda)\right] = \varepsilon'(\lambda).$$

We use $\mathcal{A}$ to construct a distinguisher $\mathcal{D}$ which distinguishes the $(\mathbf{A}^{(i)})_{i\in[\ell]}$ and the $(\mathbf{U}^{(i)})_{i\in[\ell]}$ with non-negligible advantage. Let $N = \lambda/\varepsilon' = \mathsf{poly}(\lambda)$. The distinguisher $\mathcal{D}$ works as follows.

```
$\mathcal{D}((\mathbf{A}^{(i)})_{i\in[\ell]})$:
 – For $i = 1,\ldots,\ell$:
     – For $j = 1,\ldots,N$:
         – Choose $\mathbf{s}_{i,j} \hookleftarrow \mathcal{S}$ and $\mathbf{e}_{i,j} \hookleftarrow \psi$
         – Compute $\mathbf{s}'_{i,j} \leftarrow \mathcal{A}(\mathbf{A}^{(i)}, \mathbf{A}^{(i)}\mathbf{s}_{i,j} + \mathbf{e}_{i,j} \bmod q)$
     – If for all $j \in [N]$ it holds that $\mathbf{s}'_{i,j} \neq \mathbf{s}_{i,j}$, abort and output 1.
 – Output 0.
```

We now analyze the distinguishing advantage of $\mathcal{D}$.

1. First assume that $\mathcal{D}$'s input is $(\mathbf{A}^{(i)})_{i\in[\ell]}$ distributed according to $\mathcal{X}^\ell$. Since the $\mathbf{A}^{(i)}$ are all independent and $\mathcal{X}$ is sometimes lossy for $\mathcal{S}$ and ψ, recalling that $\ell = \lambda/\delta$ it holds that

$$\begin{aligned}\mathbb{P}_{(\mathbf{A}^{(i)})_{i\in[\ell]}\hookleftarrow\mathcal{X}^\ell}&[\forall i \in [\ell] : \widetilde{H}^{\varepsilon}_{\infty}(\mathbf{s}|\mathbf{A}^{(i)}\mathbf{s}+\mathbf{e} \bmod q) < \kappa]\\ &= \prod_{i=1}^{\ell}\mathbb{P}_{\mathbf{A}^{(i)}\hookleftarrow\mathcal{X}}[\widetilde{H}^{\varepsilon}_{\infty}(\mathbf{s}|\mathbf{A}^{(i)}\mathbf{s}+\mathbf{e} \bmod q) < \kappa]\\ &\leq (1-\delta)^\ell \leq e^{-\delta\ell} = e^{-\lambda},\end{aligned}$$

which is negligible. Consequently, there must exist an index $i \in [\ell]$ such that $\widetilde{H}^{\varepsilon}_{\infty}(\mathbf{s}|\mathbf{A}^{(i)}\mathbf{s}+\mathbf{e} \bmod q) \geq \kappa$, except with negligible probability over the choice of $(\mathbf{A}^{(i)})_{i\in[\ell]}$. Thus, fix $(\mathbf{A}^{(i)})_{i\in[\ell]}$ for which there exists an $i^* \in [\ell]$ with $\widetilde{H}^{\varepsilon}_{\infty}(\mathbf{s}|\mathbf{A}^{(i^*)}\mathbf{s}+\mathbf{e} \bmod q) \geq \kappa$. Now, since $(\mathbf{s}_{i^*,j})_{j\in[N]}$ is distributed according to $\mathcal{S}^N$, it holds by a union-bound that

$$\begin{aligned}
\mathbb{P}[\exists j \in [N] : \mathcal{A}(\mathbf{A}^{(i^*)},&\mathbf{A}^{(i^*)}\mathbf{s}_{i^*,j} + \mathbf{e}_{i^*,j} \bmod q) = \mathbf{s}_{i^*,j}] \\
&\leq N \cdot \mathbb{P}[\mathcal{A}(\mathbf{A}^{(i^*)}, \mathbf{A}^{(i^*)}\mathbf{s} + \mathbf{e} \bmod q) = \mathbf{s}] \\
&\leq N \cdot 2^{-\widetilde{H}_{\infty}(\mathbf{s}|\mathbf{A}^{(i^*)},\mathbf{A}^{(i^*)}\mathbf{s}+\mathbf{e} \bmod q)} \\
&\leq N \cdot 2^{-\widetilde{H}^{\varepsilon}_{\infty}(\mathbf{s}|\mathbf{A}^{(i^*)},\mathbf{A}^{(i^*)}\mathbf{s}+\mathbf{e} \bmod q)+\log_2 \mathsf{poly}(\lambda)} \qquad (6) \\
&\leq N \cdot \mathsf{poly}(\lambda) \cdot 2^{-\kappa},
\end{aligned}$$

where $\mathbf{s} \hookleftarrow \mathcal{S}$ and $\mathbf{e} \hookleftarrow \psi$. Inequality (6) follows by Lemma B.4 if ε is sufficiently small. Then, since we have $N = \mathsf{poly}(\lambda)$ and $\kappa = \omega(\log_2 \lambda)$, the bound is indeed negligible. It follows that in the computation of $\mathcal{D}$ for input $(\mathbf{A}^{(i)})_{i\in[\ell]}$, in the i^*-th iteration of the outer loop it will hold that $\mathbf{s}'_{i^*,j} \neq \mathbf{s}_{i^*,j}$ for all $j \in [N]$, except with negligible probability over the choice of $(\mathbf{s}_{i^*,j})_{j\in[N]}$ and $(\mathbf{e}_{i^*,j})_{j\in[N]}$. The distinguisher therefore outputs $\mathcal{D}((\mathbf{A}^{(i)})_{i\in[\ell]}) = 1$, except with probability at most $e^{-\lambda} + (1-e^{-\lambda}) \cdot N \cdot 2^{-\kappa}$ over the choice of $(\mathbf{A}^{(i)})_{i\in[\ell]}$ and the random coins of $\mathcal{D}$.

2. Now assume that $\mathcal{A}$'s input is $(\mathbf{U}^{(i)})_{i\in[\ell]}$, where each $\mathbf{U}^{(i)}$ is distributed according to $\mathcal{M}^{m\times d}$. We show that with high probability over the choice of $(\mathbf{U}^{(i)})_{i\in[\ell]}$ and the random coins of $\mathcal{D}$, for every iteration i there will be an index j such that $\mathbf{s}'_{i,j} = \mathbf{s}_{i,j}$, which leads to $\mathcal{D}((\mathbf{U}^{(i)})_{i\in[\ell]}) = 0$.

   Fix an $i^* \in [\ell]$. Define the event $\mathsf{BAD}(\mathbf{s})$ by

$$\mathsf{BAD}(\mathbf{s}) :\Leftrightarrow \mathbb{P}_{\mathbf{U},\mathbf{e}}[\mathcal{A}(\mathbf{U}, \mathbf{U}\mathbf{s} + \mathbf{e} \bmod q) = \mathbf{s}] < 1 - \mu,$$

   where $\mathbf{U} \hookleftarrow \mathcal{M}^{m\times d}$. Recall that since we assume that $\mathcal{A}$ breaks mild hardness it holds that $\mathbb{P}_{\mathbf{s}}[\mathsf{BAD}(\mathbf{s})] \leq 1-\varepsilon'$. We now bound the probability that the $\mathbf{s}_{i^*,j}$ are bad for all $j \in [N]$. Since the $\mathbf{s}_{i^*,j}$ are independent, it holds that

$$\begin{aligned}
\mathbb{P}_{(\mathbf{s}_{i^*,j})_j}[\forall j \in [N] : \mathsf{BAD}(\mathbf{s}_{i^*,j})] &= \prod_{j\in[N]} \mathbb{P}_{\mathbf{s}_{i^*,j}}[\mathsf{BAD}(\mathbf{s}_{i^*,j})] \\
&\leq (1-\varepsilon')^N \leq e^{-\varepsilon'\cdot N} = e^{-\lambda},
\end{aligned}$$

   recalling that $N\varepsilon' = \lambda$. Then, with overwhelming probability $1 - e^{-\lambda}$, it holds that at least one $\mathbf{s}_{i^*,j}$ is not bad. Fix $(\mathbf{s}_{i^*,j})_{j\in[N]}$ such that there must exist an index j^* such that $(\mathbf{s}_{i^*,j^*})$ is not bad, i.e., $\mathbb{P}_{\mathbf{U},\mathbf{e}_{i^*,j^*}}[\mathcal{A}(\mathbf{U}, \mathbf{U}\mathbf{s}_{i^*,j^*} +$

$\mathbf{e}_{i^*,j^*} \bmod q) = \mathbf{s}_{i^*,j^*}] \geq 1 - \mu$. It follows that

$$\begin{aligned} &\mathbb{P}_{\mathbf{U}^{(i^*)},(\mathbf{e}_{i^*,j})_j}[\exists j \in [N] : \mathcal{A}(\mathbf{U}^{(i^*)}, \mathbf{U}^{(i^*)}\mathbf{s}_{i^*,j} + \mathbf{e}_{i^*,j} \bmod q) = \mathbf{s}_{i^*,j}] \\ &\qquad \geq \mathbb{P}_{\mathbf{U}^{(i^*)},\mathbf{e}_{i^*,j^*}}[\mathcal{A}(\mathbf{U}^{(i^*)}, \mathbf{U}^{(i^*)}\mathbf{s}_{i^*,j^*} + \mathbf{e}_{i^*,j^*} \bmod q) = \mathbf{s}_{i^*,j^*}] \\ &\qquad \geq 1 - \mu, \end{aligned}$$

which is overwhelming. So the i^*-th iteration of the outer loop does not abort (with output 1) with at most negligible probability over the choice of the $(\mathbf{s}_{i^*,j})_{j\in[N]}$, $(\mathbf{e}_{i^*,j})_{j\in[N]}$ and $\mathbf{U}^{(i^*)}$.

A union-bound over all $i^* \in [\ell]$ yields that with at most negligible probability over the choice of the $(\mathbf{U}^{(i)})_{i\in[\ell]}$ and the randomness of $\mathcal{D}$ that in the computation of $\mathcal{D}((\mathbf{U}^{(i)})_{i\in[\ell]})$ any of the ℓ iterations of the outer loop results in an abort with output 1. By construction of $\mathcal{D}$, this means that $\mathcal{D}((\mathbf{U}^{(i)})_{i\in[\ell]}) = 0$ with overwhelming probability.

Putting everything together, we conclude that

$$\begin{aligned} &\mathbb{P}[\mathcal{D}((\mathbf{A}^{(i)})_{i\in[\ell]}) = 1] - \mathbb{P}[\mathcal{D}((\mathbf{U}^{(i)})_{i\in[\ell]}) = 1] \\ &\qquad = \mathbb{P}[\mathcal{D}((\mathbf{A}^{(i)})_{i\in[\ell]}) = 1] + \mathbb{P}[\mathcal{D}((\mathbf{U}^{(i)})_{i\in[\ell]}) = 0] - 1 \\ &\qquad = 1 - \mathsf{negl}(\lambda). \end{aligned}$$

Thus, $\mathcal{D}$ distinguishes $\mathcal{X}$ and $\mathcal{M}^{m\times d}$ with advantage close to 1, which contradicts the assumption that $\mathcal{X}$ and $\mathcal{M}^{m\times d}$ are computationally indistinguishable. This concludes the proof. □

Lemma 4.1

Proof. Let $\mathbf{S} = \sqrt{s^2\mathbf{I}_{nm} - (s')^2\mathbf{Z}'(\mathbf{Z}')^T}$. Let $\mathbf{f}' \sim D_{s'\mathbf{I}_{nd}} = D_{s'}^{nd}$ and $\mathbf{f}'' \sim D_{\mathbf{S}}$. It thus holds that $\mathbf{f} = \mathbf{Z}'\mathbf{f}' + \mathbf{f}''$ is distributed according to $D_s^{nm} = D_{s\mathbf{I}_{nm}}$. Multypling on the left by $\mathbf{B}_{R^m}$ yields $\mathbf{B}_{R^m}\mathbf{f} = \mathbf{Z}\mathbf{B}_{R^d}\mathbf{f}' + \mathbf{B}_{R^m}\mathbf{f}''$.

Now notice that $\mathbf{B}_{R^d}\mathbf{f}'$ is distributed according to $D_{s'\cdot\mathbf{B}_{R^d}}$ and that $\mathbf{B}_{R^m}\mathbf{f}$ is distributed according to $D_{s\cdot\mathbf{B}_{R^m}}$. Hence $\mathbf{e}'$ and $\mathbf{B}_{R^d}\mathbf{f}'$ are identically distributed, and $\mathbf{e}$ and $\mathbf{B}_{R^m}\mathbf{f}$ are too. Setting Ψ to be the distribution of $\mathbf{B}_{R^m}\mathbf{f}''$, the result follows. □

Lemma 4.2

Proof. Let $\tilde{\mathbf{e}}'$ be distributed according to the continuous Gaussian $D_{\sigma\mathbf{I}_{nd}}$, and let $\tilde{\mathbf{e}}$ according to the discrete Gaussian $\mathcal{D}_{\Lambda(\mathbf{F}'\mathbf{B}_{R^d}^{-1}),\sigma\mathbf{I}_{nd}}$. Let $\tilde{\mathbf{f}} \hookleftarrow D_{\sqrt{2}\sigma\mathbf{I}_{nd}}$. By Lemma B.2, it holds that $\|\mathbf{F}'\|_2 \cdot \eta_\varepsilon(\mathbf{B}_{R^d}^{-1}) \geq \eta_\varepsilon(\Lambda(\mathbf{F}' \cdot \mathbf{B}_{R^d}^{-1}))$. Lemma B.3 yields $\Delta(\tilde{\mathbf{f}}, \tilde{\mathbf{e}} + \tilde{\mathbf{e}}') \leq 2\varepsilon$.

Now note that by the definition of $\mathbf{e}$ we have that $\mathbf{e}$ and $\mathbf{B}_{R^d}\tilde{\mathbf{e}}$ are identically distributed. Similarly, $\mathbf{f}$ and $\mathbf{B}_{R^d}\tilde{\mathbf{f}}$ are identically distributed. By setting $\mathbf{e}' = \mathbf{B}_{R^d}\tilde{\mathbf{e}}'$, we then obtain that $\Delta(\mathbf{F}^{-1}\mathbf{f},\ \mathbf{F}^{-1}\mathbf{e} + \mathbf{F}^{-1}\mathbf{e}') \leq 2\varepsilon$. We finally conclude

$$\begin{aligned}\widetilde{H}_\infty(\mathbf{s}|\mathbf{F}_q^{-1}\mathbf{s}+\mathbf{F}^{-1}\mathbf{e} \bmod q) &= \widetilde{H}_\infty(\mathbf{s}|\mathbf{F}_q^{-1}\mathbf{s}+\mathbf{F}^{-1}\mathbf{e} \bmod q, \mathbf{e}')\\ &\le \widetilde{H}_\infty(\mathbf{s}|\mathbf{F}_q^{-1}\mathbf{s}+\mathbf{F}^{-1}\mathbf{e}+\mathbf{F}^{-1}\mathbf{e}' \bmod q)\\ &\le \widetilde{H}_\infty^{2\varepsilon}(\mathbf{s}|\mathbf{F}_q^{-1}\mathbf{s}+\mathbf{F}^{-1}\mathbf{f} \bmod q),\end{aligned}$$

which yields the result. □

Lemma 4.3

Proof. Let $\mathbf{e}'$ be distributed according to $\mathcal{D}_{\mathbb{Z}^{nd}-\mathbf{e},s\mathbf{B}_{R^d}}$. Then, by Theorem B.1 it holds that $\Delta(\mathbf{e}+\mathbf{e}',\mathbf{f}) \le 8\varepsilon$. A chain of inequalities on the min-entropy similar to the one from the previous proof yields the result. □

Theorem 4.2

Proof. Fix a distribution of secrets $\mathcal{S}$ and let $\mathbf{s} \hookleftarrow \mathcal{S}$. Let $s_1 = s_0/\left\|\mathbf{G}'(\mathbf{F}')^{-1}\right\|_2 \ge 2^{3/2}s$. As $\mathbf{G}'(\mathbf{F}')^{-1} = \mathbf{B}_{R^m}^{-1}(\mathbf{G}\mathbf{F}^{-1})\mathbf{B}_{R^d}$, Lemma 4.1 states that there exists a distribution Ψ over $\mathbb{R}^{nm}$ such that we can equivalently sample $\mathbf{e}_0$ by $\mathbf{e}_0 = \mathbf{G}\mathbf{F}^{-1}\mathbf{e}_1 + \mathbf{e}_1'$, where $\mathbf{e}_1 \hookleftarrow D_{\sigma_1\mathbf{B}_{R^d}}$ and $\mathbf{e}_1' \hookleftarrow \Psi$. Consequently, we can write

$$\mathbf{y} = \mathbf{G}\mathbf{F}_q^{-1}\mathbf{s}+\mathbf{e}_0 = \mathbf{G}\mathbf{F}_q^{-1}\mathbf{s}+\mathbf{G}\mathbf{F}^{-1}\mathbf{e}_1+\mathbf{e}_1' = \mathbf{G}(\mathbf{F}_q^{-1}\mathbf{s}+\mathbf{F}^{-1}\mathbf{e}_1)+\mathbf{e}_1'.$$

Thus, since $\mathbf{y} \bmod q$ can be computed from $\mathbf{F}_q^{-1}\mathbf{s}+\mathbf{F}^{-1}\mathbf{e}_1 \bmod q$ and $\mathbf{e}_1'$ it follows that

$$\begin{aligned}\widetilde{H}_\infty(\mathbf{s}|\mathbf{G}\mathbf{F}_q^{-1}\mathbf{s}+\mathbf{e}_0 \bmod q) &= \widetilde{H}_\infty(\mathbf{s}|\mathbf{F}_q^{-1}\mathbf{s}+\mathbf{F}^{-1}\mathbf{e}_1 \bmod q, \mathbf{e}_1')\\ &= \widetilde{H}_\infty(\mathbf{s}|\mathbf{F}_q^{-1}\mathbf{s}+\mathbf{F}^{-1}\mathbf{e}_1 \bmod q),\end{aligned} \tag{7}$$

where the second equality follows as $\mathbf{e}_1'$ is independent from $\mathbf{s}$ and $\mathbf{e}_1$. Now let $s_2 = s_1/\sqrt{2} \ge 2s$ and let $\mathbf{e}_2 \hookleftarrow \mathcal{D}_{\Lambda(\mathbf{F}),s_2\mathbf{B}_{R^d}}$ be a discrete Gaussian. By Lemma 4.2 it holds that

$$\widetilde{H}_\infty^{2\varepsilon}(\mathbf{s}|\mathbf{F}_q^{-1}\mathbf{s}+\mathbf{F}^{-1}\mathbf{e}_1 \bmod q) \ge \widetilde{H}_\infty(\mathbf{s}|\mathbf{F}_q^{-1}\mathbf{s}+\mathbf{F}^{-1}\mathbf{e}_2 \bmod q). \tag{8}$$

Now, since $\mathbf{F}_q^{-1}$ is the $\mathbb{Z}_q$-inverse of $\mathbf{F} \bmod q$, multiplying $\mathbf{F}_q^{-1}\mathbf{s}+\mathbf{F}^{-1}\mathbf{e}_2$ by $\mathbf{F}$ preserves the entropy and yields

$$\widetilde{H}_\infty(\mathbf{s}|\mathbf{F}_q^{-1}\mathbf{s}+\mathbf{F}^{-1}\mathbf{e}_2 \bmod q) = \widetilde{H}_\infty(\mathbf{s}|\mathbf{s}+\mathbf{e}_2 \bmod q). \tag{9}$$

Now let $s_3 = s_2/\sqrt{2} \ge \sqrt{2}s$, $\mathbf{e}_3 \hookleftarrow \mathcal{D}_{\mathbb{Z}^{nd},s_3\mathbf{B}_{R^d}}$ and $\mathbf{e}_3' \hookleftarrow \mathcal{D}_{\Lambda(\mathbf{F})-\mathbf{e}_3,s_3\mathbf{B}_{R^d}}$. Setting $\Lambda_2 = \mathbb{Z}^{nd}$ and $\Lambda_1 = \Lambda(\mathbf{F})$ in Theorem B.1 and noting that $s_3 > s > \eta_\varepsilon(\Lambda(\mathbf{B}_{R^d}^{-1}))$ we obtain that $\Delta(\mathbf{e}_2, \mathbf{e}_3+\mathbf{e}_3') \le 8\varepsilon$. It follows that

$$\begin{aligned}\widetilde{H}_\infty^{8\varepsilon}(\mathbf{s}|\mathbf{s}+\mathbf{e}_2 \bmod q) &\ge \widetilde{H}_\infty(\mathbf{s}|\mathbf{s}+\mathbf{e}_3+\mathbf{e}_3' \bmod q)\\ &\ge \widetilde{H}_\infty(\mathbf{s}|\mathbf{s}+\mathbf{e}_3 \bmod q, \mathbf{e}_3').\end{aligned} \tag{10}$$

Since $\mathbf{e}_3'$ is distributed according to $\mathcal{D}_{\Lambda(\mathbf{F})-\mathbf{e}_3, s_3\mathbf{B}_{R^d}}$, it only depends on the lattice coset $\mathbf{e}_3 \bmod \Lambda(\mathbf{F})$. Thus

$$\begin{aligned}\widetilde{H}_\infty(\mathbf{s}|\mathbf{s}+\mathbf{e}_3 \bmod q, \mathbf{e}_3') &\geq \widetilde{H}_\infty(\mathbf{s}|\mathbf{s}+\mathbf{e}_3 \bmod q) - H_0(\mathbf{e}_3') \\ &\geq \widetilde{H}_\infty(\mathbf{s}|\mathbf{s}+\mathbf{e}_3 \bmod q) - nd \cdot \log_2(\|\mathbf{F}'\|_2),\end{aligned} \tag{11}$$

as $\left|\mathbb{Z}^{nd}/\Lambda(\mathbf{F})\right| = |\det(\mathbf{F})| = |\det \mathbf{F}'| \leq \|\mathbf{F}'\|_2^{nd}$. Finally, we have that $s_3/\sqrt{2} = s > \eta_\varepsilon(\Lambda(\mathbf{B}_{R^d}^{-1}))$, Lemma 4.3 provides with the bound

$$\widetilde{H}_\infty^{8\varepsilon}(\mathbf{s}|\mathbf{s}+\mathbf{e}_3 \bmod q) \geq \widetilde{H}_\infty(\mathbf{s}|\mathbf{s}+\mathbf{e} \bmod q), \tag{12}$$

where $\mathbf{e} \hookleftarrow D_{s\mathbf{B}_{R^d}}$. Putting everything together, we obtain that

$$\begin{aligned}&\widetilde{H}_\infty(\mathbf{s}|\mathbf{s}+\mathbf{e} \bmod q) - nd\log_2(\|\mathbf{F}'\|_2) \\ &\qquad\leq \widetilde{H}_\infty^{8\varepsilon}(\mathbf{s}|\mathbf{s}+\mathbf{e}_3 \bmod q) - nd\log_2(\|\mathbf{F}'\|_2) && (12) \\ &\qquad\leq \widetilde{H}_\infty^{8\varepsilon}(\mathbf{s}|\mathbf{s}+\mathbf{e}_3 \bmod q, \mathbf{e}_3') && (11) \\ &\qquad\leq \widetilde{H}_\infty^{16\varepsilon}(\mathbf{s}|\mathbf{s}+\mathbf{e}_2 \bmod q) && (10) \\ &\qquad= \widetilde{H}_\infty^{16\varepsilon}(\mathbf{s}|\mathbf{F}_q^{-1}\mathbf{s}+\mathbf{F}^{-1}\mathbf{e}_2 \bmod q) && (9) \\ &\qquad\leq \widetilde{H}_\infty^{18\varepsilon}(\mathbf{s}|\mathbf{F}_q^{-1}\mathbf{s}+\mathbf{F}^{-1}\mathbf{e}_1 \bmod q) && (8) \\ &\qquad= \widetilde{H}_\infty^{18\varepsilon}(\mathbf{s}|\mathbf{GF}_q^{-1}\mathbf{s}+\mathbf{e}_0 \bmod q), && (7)\end{aligned}$$

concluding the proof. □

Theorem 4.3

Proof. Directly by the S-NTRU assumption it holds that $\mathcal{X}$ is computationally indistinguishable from $\mathcal{M}^{m\times d}$ and thus verifies pseudorandomness.

For $\mathbf{s} \hookleftarrow \mathcal{S}$ and $\mathbf{e}_0 \hookleftarrow \psi$, we have

$$\mathbb{P}_{\mathbf{GF}_q^{-1}\hookleftarrow\mathcal{X}}[\widetilde{H}_\infty^{18\varepsilon}(\mathbf{s}|\mathbf{GF}_q^{-1}\mathbf{s}+\mathbf{e}_0 \bmod q, \mathbf{GF}_q^{-1}) \geq \omega(\log_2\lambda)] \geq \delta.$$

Indeed, take $\mathbf{GF}_q^{-1} \hookleftarrow \mathcal{X}$. With probability at least δ, the conditions of Theorem 4.2 are verified and thus

$$\begin{aligned}\widetilde{H}_\infty^{18\varepsilon}(\mathbf{s}|\mathbf{GF}_q^{-1}\mathbf{s}+\mathbf{e}_0 \bmod q, \mathbf{GF}_q^{-1}) &= \widetilde{H}_\infty^{18\varepsilon}(\mathbf{s}|\mathbf{GF}_q^{-1}\mathbf{s}+\mathbf{e}_0 \bmod q) \\ &\geq \widetilde{H}_\infty(\mathbf{s}|\mathbf{s}+\mathbf{e} \bmod q) - nd\log_2(\beta_2),\end{aligned}$$

where $\mathbf{e} \hookleftarrow D_{s\mathbf{B}_{R^d}}$. Yet $\nu_{s\mathbf{B}_{R^d}}(\mathcal{S}) = \widetilde{H}_\infty(\mathbf{s}|\mathbf{s}+\mathbf{e} \bmod q) \geq nd\log_2(\beta_2)+\omega(\log_2(\lambda))$. As a result, it holds that $\widetilde{H}_\infty^{18\varepsilon}(\mathbf{s}|\mathbf{GF}_q^{-1}\mathbf{s}+\mathbf{e}_0 \bmod q, \mathbf{GF}_q^{-1}) \geq \omega(\log_2\lambda)$ with probability at least δ, thus proving the sometimes lossiness. □

B.5 Proofs of Section 5

Lemma 5.1

Proof. Let $a \in R$ be sampled from $\mathcal{D}_{R,\gamma}$. Then $\sigma_H(a) = \mathbf{U}_H^\dagger \sigma(a)$ is distributed according to $\mathcal{D}_{\Lambda,\gamma}$ where $\Lambda = \sigma_H(R)$. So $\|\sigma(a)\|_\infty = \|\mathbf{U}_H \sigma_H(a)\|_\infty \leq \|\sigma_H(a)\|_\infty$. We briefly explain the last inequality. For clarity, we define $\mathbf{a} = \sigma_H(a)$. By decomposing $\mathbf{a} = [\mathbf{a}_1^T | \mathbf{a}_2^T | \mathbf{a}_3^T]^T$, with $\mathbf{a}_1 \in \mathbb{R}^{t_1}$ and $\mathbf{a}_2, \mathbf{a}_3 \in \mathbb{R}^{t_2}$, a standard calculation gives

$$\mathbf{U}_H \mathbf{a} = \frac{1}{\sqrt{2}} \begin{bmatrix} \sqrt{2}\mathbf{a}_1 \\ \mathbf{a}_2 + i\mathbf{a}_3 \\ \mathbf{a}_2 - i\mathbf{a}_3 \end{bmatrix}.$$

Hence $\|\mathbf{U}_H \mathbf{a}\|_\infty \leq \|\mathbf{a}\|_\infty$.

By the second part of [Pei08, Cor. 5.3] for $m = 1$, $\mathbf{z} = 1$ and $\mathbf{c} = \mathbf{0}$, it holds that for all $t \geq 0$

$$\mathbb{P}_{\mathbf{a} \hookleftarrow \mathcal{D}_{\Lambda,\gamma}}[\|\mathbf{a}\|_\infty \geq \gamma t] \leq 2ne^{-\pi t^2}.$$

Note that in the case where $\mathbf{c} = \mathbf{0}$, the restriction of $\gamma \geq \eta_\varepsilon(\Lambda)$ for some $\varepsilon \leq 1/(2m+1)$ is not necessary, and the calculation of the bound on the probability saves a factor of e for that reason. With the observation that $\|\sigma(a)\|_\infty \leq \|\sigma_H(a)\|_\infty$ it holds

$$\begin{aligned} \mathbb{P}_{a \hookleftarrow \mathcal{D}_{R,\gamma}}[\|\sigma(a)\|_\infty \leq \gamma t] &\geq \mathbb{P}_{a \hookleftarrow \mathcal{D}_{R,\gamma}}[\|\sigma_H(a)\|_\infty \leq \gamma t] \\ &\geq 1 - 2ne^{-\pi t^2} \end{aligned}$$

Choosing $t = \log_2 n$ gives an overwhelming probability, concluding the proof. □