A Toolkit for Ring-LWE Cryptography

Vadim Lyubashevsky* Chris Peikert† Oded Regev‡

May 16, 2013

Abstract

Recent advances in lattice cryptography, mainly stemming from the development of ring-based primitives such as ring-LWE, have made it possible to design cryptographic schemes whose efficiency is competitive with that of more traditional number-theoretic ones, along with entirely new applications like fully homomorphic encryption. Unfortunately, realizing the full potential of ring-based cryptography has so far been hindered by a lack of practical algorithms and analytical tools for working in this context. As a result, most previous works have focused on very special classes of rings such as power-of-two cyclotomics, which significantly restricts the possible applications.

We bridge this gap by introducing a toolkit of fast, modular algorithms and analytical techniques that can be used in a wide variety of ring-based cryptographic applications, particularly those built around ring-LWE. Our techniques yield applications that work in *arbitrary* cyclotomic rings, with *no loss* in their underlying worst-case hardness guarantees, and very little loss in computational efficiency, relative to power-of-two cyclotomics. To demonstrate the toolkit's applicability, we develop a few illustrative applications: two variant public-key cryptosystems, and a "somewhat homomorphic" symmetric encryption scheme. Both apply to arbitrary cyclotomics, have tight parameters, and very efficient implementations.

1 Introduction

The past few years have seen many exciting developments in lattice-based cryptography. Two such trends are the development of schemes whose efficiency is competitive with traditional number-theoretic ones (e.g., [Mic02] and follow-ups), and the breakthrough work of Gentry [Gen09b, Gen09a] (followed by others) on fully homomorphic encryption. While these two research threads currently occupy opposite ends of the efficiency spectrum, they are united by their use of algebraically structured *ideal lattices* arising from polynomial rings. The most efficient and advanced systems in both categories rely on the ring-LWE problem [LPR10], an analogue of the standard *learning with errors* problem [Reg05]. Informally (and a bit inaccurately), in a ring $R = \mathbb{Z}[X]/(f(X))$ for monic irreducible $f(X)$ of degree n, and for an integer

*INRIA and École Normale Supérieure, Paris. Part of this work was performed while at Tel Aviv University and also while visiting Georgia Tech. Partially supported by a European Research Council (ERC) Starting Grant.

†School of Computer Science, College of Computing, Georgia Institute of Technology. This material is based upon work supported by the National Science Foundation under CAREER Award CCF-1054495, by DARPA under agreement number FA8750-11-C-0096, and by the Alfred P. Sloan Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, DARPA or the U.S. Government, or the Sloan Foundation. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.

‡Courant Institute, New York University. Supported by a European Research Council (ERC) Starting Grant. Part of the work done while the author was with the CNRS, DI, ENS, Paris.

modulus q defining the quotient ring $R_q := R/qR = \mathbb{Z}_q[X]/(f(X))$, the ring-LWE problem is to distinguish pairs $(a_i, b_i = a_i \cdot s + e_i) \in R_q \times R_q$ from uniformly random pairs, where $s \in R_q$ is a random secret (which stays fixed over all pairs), the $a_i \in R_q$ are uniformly random and independent, and the error (or "noise") terms $e_i \in R$ are independent and "short."

In all applications of ring-LWE, and particularly those related to homomorphic encryption, a main technical challenge is to control the sizes of the noise terms when manipulating ring-LWE samples under addition, multiplication, and other operations. For correct decryption, q must be chosen large enough so that the final accumulated error terms do not "wrap around" modulo q and cause decryption error. On the other hand, the *error rate* (roughly, the ratio of the noise magnitude to the modulus q) of the original published ring-LWE samples and the dimension n trade off to determine the theoretical and concrete hardness of the ring-LWE problem. Tighter control of the noise growth therefore allows for a larger initial error rate, which permits a smaller modulus q and dimension n, which leads to smaller keys and ciphertexts, and faster operations for a given level of security.

Regarding the choice of ring, the class of *cyclotomic* rings $R \cong \mathbb{Z}[X]/\Phi_m(X)$, where $\Phi_m(X)$ is the mth cyclotomic polynomial (which has degree $n = \varphi(m)$ and is monic and irreducible over the rationals), has many attractive features that have proved very useful in cryptography. For example, the search/decision equivalence for ring-LWE in arbitrary cyclotomics [LPR10] relies on their special algebraic properties, as do many recent works that aim for more efficient fully homomorphic encryption schemes (e.g., [SV11, BGV12, GHS12a, GHS12b, GHPS12]). In particular, *power-of-two* cyclotomics, i.e., where the index $m = 2^k$ for some $k \geq 1$, are especially nice to work with, because (among other reasons) $n = m/2$ is also a power of two, $\Phi_m(X) = X^n + 1$ is maximally sparse, and polynomial arithmetic modulo $\Phi_m(X)$ can be performed very efficiently using just a slight tweak of the classical n-dimensional FFT (see, e.g., [LMPR08]). Indeed, power-of-two cyclotomics have become the dominant and preferred class of rings in almost all recent ring-based cryptographic schemes (e.g., [LMPR08, LM08, Lyu09, Gen09b, Gen10, LPR10, SS11, BV11b, BGV12, GHS12a, GHS12b, Lyu12, BPR12, MP12, GLP12, GHPS12]), often to the exclusion of all other rings.

While power-of-two cyclotomic rings are very convenient to use, there are several reasons why it is essential to consider other cyclotomics as well. The most obvious, practical reason is that powers of two are sparsely distributed, and the desired concrete security level for an application may call for a ring dimension much smaller than the next-largest power of two. So restricting to powers of two could lead to key sizes and runtimes that are at least twice as large as necessary. A more fundamental reason is that certain applications, such as the above-mentioned works that aim for more efficient (fully) homomorphic encryption, *require* the use of non-power-of-two cyclotomic rings. This is because power-of-two cyclotomics lack the requisite algebraic properties needed to implement features like SIMD operations on "packed" ciphertexts, or plaintext spaces isomorphic to finite fields of characteristic two (other than $\mathbb{F}_2$ itself). A final important reason is diversification of security assumptions. While some results are known [GHPS12] that relate ring-LWE in cyclotomic rings when one index m divides the other, no other connections appear to be known. So while we might conjecture that ring-LWE and ideal lattice problems are hard in *every* cyclotomic ring (of sufficiently high dimension), some rings might turn out to be significantly easier than others.

Unfortunately, working in non-power-of-two cyclotomics is rather delicate, and the current state of affairs is unsatisfactory in several ways. Unlike the special case where m is a power of two, in general the cyclotomic polynomial $\Phi_m(X)$ can be quite "irregular" and dense, with large coefficients. While in principle, polynomial arithmetic modulo $\Phi_m(X)$ can still be done in $O(n \log n)$ scalar operations (on high-precision complex numbers), the generic algorithms for achieving this are rather complex and hard to implement, with large constants hidden by the $O(\cdot)$ notation.

Geometrically, the non-power-of-two case is even more problematic. If one views $\mathbb{Z}[X]/(\Phi_m(X))$ as the set of polynomial residues of the form $a_0 + a_1 X + \cdots + a_{n-1}X^{n-1}$, and uses the naïve "coefficient embedding" that views them as vectors $(a_0, a_1, \ldots, a_{n-1}) \in \mathbb{Z}^n$ to define geometric quantities like the ℓ_2 norm, then both the concrete and theoretical security of cryptographic schemes depend heavily on the form of $\Phi_m(X)$. This stems directly from the fact that multiplying two polynomials with small norms can result in a polynomial residue having a *much* larger norm. The growth can be quantified by the "expansion factor" [LM06] of $\Phi_m(X)$, which unfortunately can be very large, up to $n^{\Omega(\log n)}$ in the case of highly composite m [Erd46]. Later works [GHS12a] circumvented such large expansion by using tricks like lifting to the larger-dimensional ring $\mathbb{Z}[X]/(X^m - 1)$, but this still involves a significant loss in the tolerable noise rates as compared with the power-of-two case.

In [PR07, LPR10] a different geometric approach was used, which avoided any dependence on the form of the polynomial modulus $\Phi_m(X)$. In these works, the norm of a ring element is instead defined according to its *canonical embedding* into $\mathbb{C}^n$, a classical concept from algebraic number theory. This gives a much better way of analyzing expansion, since both addition and multiplication in the canonical embedding are simply coordinate-wise. Working with the canonical embedding, however, introduces a variety of practical issues, such as how to efficiently generate short noise terms having appropriate distributions over the ring. More generally, the focus of [LPR10] was on giving an abstract mathematical definition of ring-LWE and proving its hardness under worst-case ideal lattice assumptions; in particular, it did not deal with issues related to practical efficiency, bounding noise growth, or designing applications in non-power-of-two cyclotomics.

1.1 Contributions

Our main contribution is a toolkit of modular algorithms and analytical techniques that can be used in a wide variety of ring-based cryptographic applications, particularly those built around ring-LWE. The high-level summary is that using our techniques, one can design applications to work in *arbitrary* cyclotomic rings, with *no loss* in their underlying worst-case hardness guarantees, and very little loss in computational efficiency, relative to the best known techniques in power-of-two cyclotomics. In fact, our analytical techniques even improve the state of the art for the power-of-two case.

In more detail, our toolkit includes fast, specialized algorithms for all the main cryptographic operations in arbitrary cyclotomic rings. Among others, these include: addition, multiplication, and conversions among various useful representations of ring elements; generation of noise terms under probability distributions that guarantee both worst-case and concrete hardness; and decoding of noise terms as needed in decryption and related operations. Our algorithms' efficiency and quality guarantees stem primarily from our use of simple but non-obvious representations of ring elements, which differ from their naïve representations as polynomial residues modulo $\Phi_m(X)$. (See the second part of Section 1.2 for more details.) On the analytical side, we give tools for tightly bounding noise growth under operations like addition, multiplication, and round-off/discretization. (Recall that noise growth is the main factor determining an application's parameters and noise rates, and hence its key sizes, efficiency, and concrete security.)
Some attractive features of the toolkit include:

- All the algorithms for arbitrary cyclotomics are simple, modular, and highly parallel, and work by elementary reductions to the (very simple) prime-index case. In particular, they do not require any polynomial reductions modulo $\Phi_m(X)$ – in fact, they never need to compute $\Phi_m(X)$ at all! The algorithms work entirely on vectors of dimension $n = \varphi(m)$, and run in $O(n \log n)$ or even $O(nd)$ scalar operations (with small hidden constants), where d is the number of distinct primes dividing m. With the exception of continuous noise generation, all scalar operations are low precision, i.e., they

involve small integers. In summary, the algorithms are very amenable to practical implementation. (Indeed, we have implemented all the algorithms from scratch, which will be described in a separate work.)

- Our algorithm for decoding noise, used primarily in decryption, is fast (requiring $O(n \log n)$ or fewer small-integer operations) and correctly recovers from optimally large noise rates. (See the last part of Section 1.2 for details.) This improves upon prior techniques, which in general have worse noise tolerance by anywhere between $m/2$ and super-polynomial $n^{\omega(1)}$ factors, and are computationally slower and more complex due to polynomial reduction modulo $\Phi_m(X)$, among other operations.

- Our bounds on noise growth under ring addition and multiplication are exactly the same in *all* cyclotomic rings; no ring-dependent "expansion factor" is incurred. (For discretizing continuous noise distributions, our bounds are the same up to very small $1 + o(1)$ factors, depending on the primes dividing m.) This allows applications to use essentially the same underlying noise rate as a function of the ring dimension n, and hence be based on the same worst-case approximation factors, for all cyclotomics. Moreover, our bounds improve upon the state of the art even for power-of-two cyclotomics: e.g., our (average-case, high probability) expansion bound for ring multiplication improves upon the (worst-case) expansion-factor bound by almost a $\sqrt{n}$ factor.

To illustrate the toolkit's applicability, in Section 8 we develop the following illustrative applications:

1. A simple adaptation of the "dual" LWE-based public-key cryptosystem of [GPV08], which can serve as a foundation for (hierarchical) identity-based encryption. (See Section 8.1.)

2. An efficient and compact public-key cryptosystem, which is essentially the "two element" system outlined in [LPR10], but generalized to arbitrary cyclotomics, and with tight parameters. (See Section 8.2.)

3. A "somewhat homomorphic" symmetric encryption scheme, which follows the template of the Brakerski-Vaikuntanathan [BV11a] and Brakerski-Gentry-Vaikuntanathan [BGV12] schemes in power-of-two cyclotomics, but generalized to arbitrary cyclotomics and with much tighter noise analysis. This application exercises all the various parts of the toolkit more fully, especially in its modulus-reduction and key-switching procedures. (See Section 8.3.)

A final contribution of independent interest is a new "regularity lemma" for arbitrary cyclotomics, i.e., a bound on the smoothing parameter of random q-ary lattices over the ring. Such a lemma is needed for porting many applications of standard LWE (and the related "short integer solution" SIS problem) to the ring setting, including SIS-based signature schemes [GPV08, CHKP10, Boy10, MP12], the "primal" [Reg05] and "dual" [GPV08] LWE cryptosystems (as in Section 8.1), chosen ciphertext-secure encryption schemes [Pei09, MP12], and (hierarchical) identity-based encryption schemes [GPV08, CHKP10, ABB10]. In terms of generality and parameters, our lemma essentially subsumes a prior one of Micciancio [Mic02] for the ring $\mathbb{Z}[X]/(X^n - 1)$, and an independent one of Stehlé et al. [SSTX09] for power-of-two cyclotomics. See Section 7 for further discussion.

Following the preliminary publication of this work, our toolkit has also been used centrally in the "ring-switching" technique for homomorphic encryption [GHPS12], and to give efficient "bootstrapping" algorithms for fully homomorphic encryption [AP13].

1.2 Techniques

The tools we develop in this work involve several novel applications of classical notions from algebraic number theory. In summary, our results make central use of: (1) the *canonical embedding* of a number field, which endows the field (and its subrings) with a nice and easy-to-analyze geometry; (2) the decomposition of arbitrary cyclotomics into the *tensor product* of prime-power cyclotomics, which yields both simpler and faster algorithms for computing in the field, as well as geometrically nicer bases; and (3) the "*dual*" ideal $R^\vee$ and its "*decoding*" basis $\vec{d}$, for fast noise generation and optimal noise tolerance in decryption and related operations. We elaborate on each of these next.

The canonical embedding. As in the previous works [PR07, LPR10], our analysis relies heavily on using the *canonical embedding* $\sigma\colon K \to \mathbb{C}^n$ (rather than, say, the naïve coefficient embedding) for defining all geometric quantities, such as Euclidean norms and inner products. For example, under the canonical embedding, the "expansion" incurred when multiplying by an element $a \in K$ is characterized exactly by $\|\sigma(a)\|_\infty$, its ℓ_∞ norm under the canonical embedding; no (worst-case) ring-dependent "expansion factor" is needed. So in the average-case setting, where the multiplicands are random elements from natural noise distributions, for each multiplication we get at least a $\tilde{\Omega}(\sqrt{n})$ factor improvement over using the expansion factor in *all* cyclotomics (including those with power-of-two index), and up to a super-polynomial $n^{\omega(1)}$ factor improvement in cyclotomics having highly composite indices. In our analysis of the noise tolerance of decryption, we also get an additional $\tilde{\Omega}(\sqrt{n})$ factor savings over more simplistic analyses that only use norm information, by using the notion of *subgaussian* random variables. These behave under linear transformations in essentially the same way as Gaussians do, and have Gaussian tails. (Prior works that use subgaussianity in lattice cryptography include [AP09, MP12].)

Tensorial decomposition. An important fact at the heart of this work is that the mth cyclotomic number field $K = \mathbb{Q}(\zeta_m) \cong \mathbb{Q}[X]/(\Phi_m(X))$ may instead be viewed as (i.e., is isomorphic to) the *tensor product* of prime-power cyclotomics:

$$K \cong \bigotimes_\ell K_\ell = \mathbb{Q}(\zeta_{m_1}, \zeta_{m_2}, \ldots),$$

where $m = \prod_\ell m_\ell$ is the prime-power factorization of m and $K_\ell = \mathbb{Q}(\zeta_{m_\ell})$. Equivalently, in terms of polynomials we may view K as the multivariate field

$$K \cong \mathbb{Q}[X_1, X_2, \ldots]/(\Phi_{m_1}(X_1), \Phi_{m_2}(X_2), \ldots), \tag{1.1}$$

where there is one indeterminant X_ℓ and modulus $\Phi_{m_\ell}(X_\ell)$ per prime-power divisor of m. Similar decompositions hold for the ring of integers $R \cong \mathbb{Z}[X]/\Phi_m(X)$ and other important objects in K, such as the dual ideal $R^\vee$ (described below).

Adopting the polynomial interpretation of K from Equation (1.1) for concreteness, notice that a natural $\mathbb{Q}$-basis is the set of multinomials $\prod_\ell X_\ell^{j_\ell}$ for each choice of $0 \le j_\ell < \varphi(m_\ell)$. We call this set the "powerful" basis of K (and of R). Interestingly, for non-prime-power m, under the field isomorphism with $\mathbb{Q}[X]/(\Phi_m(X))$ that maps each $X_\ell \to X^{m/m_\ell}$, the powerful basis does *not* coincide with the standard "power" basis $1, X, X^2, \ldots, X^{\varphi(m)-1}$ usually used to represent the univariate field. It turns out that in general, the powerful basis has much nicer computational and geometric properties than the power basis, as we outline next.

Computationally, the tensorial decomposition of K (with the powerful basis) allows us to modularly reduce operations in K (or R, or powers of $R^\vee$) to their counterparts in much simpler prime-power cyclotomics (which themselves easily reduce to the prime-index case). We can therefore completely avoid all the

many algorithmic complications associated with working with polynomials modulo $\Phi_m(X)$. In particular, we obtain novel, simple and fast algorithms, similar to the FFT, for converting between the multivariate "polynomial" representation (i.e., the powerful basis) and the "evaluation" or "Chinese remainder" representation, in which addition and multiplication are essentially linear time. Similarly, we obtain linear-time (or nearly so) algorithms for switching between the polynomial representation and the "decoding" representation used in decryption (described below), and for generating noise terms in the decoding representation. A final advantage of the tensorial representation is that it yields trivial linear-time algorithms for computing the *trace* function to cyclotomic subfields of K.

The tensorial representation also comes with important geometrical advantages. In particular, under the canonical embedding the powerful basis is better-conditioned than the power basis, i.e., the ratio of its maximal and minimal singular values can be much smaller. This turns out to be important when bounding the additional error introduced when discretizing (rounding off) field elements in noise-generation and modulus-reduction algorithms, among others.

The dual ideal $R^\vee$ and its decoding basis. Under the canonical embedding, the cyclotomic ring R of index m embeds as a lattice which, unlike $\mathbb{Z}^n$, is in general not self-dual. Instead, its dual lattice corresponds to a fractional ideal $R^\vee \subset K$ satisfying $R \subseteq R^\vee \subseteq m^{-1}R$, where the latter inclusion is nearly an equality. (In fact, $R^\vee$ is a scaling of R exactly when m is a power of two, in which case $R = (m/2)R^\vee$.) In [LPR10] it is shown that the "right" definition of the ring-LWE distribution, which arises naturally from the worst-case to average-case reduction, involves the dual ideal $R^\vee$: the secret belongs to the quotient $R_q^\vee = R^\vee/qR^\vee$ (or just $R^\vee$), and ring-LWE samples are of the form $(a, b = a \cdot s + e \bmod qR^\vee)$ for uniformly random $a \in R_q$ and error e which is essentially spherical in the canonical embedding.

While it is possible [DD12] to simplify the ring-LWE distribution by replacing every instance of $R^\vee$ with R, while retaining essentially spherical error (but scaled up by about m, corresponding to the approximate ratio of R to $R^\vee$), in this work we show that *it is actually advantageous to retain $R^\vee$ and expose it in applications*.[1] The reason is that in general, $R^\vee$ supports correct bounded-distance decoding—which is the main operation performed in decryption—under a larger error rate than R does.[2] In fact, the error tolerance of $R^\vee$ is *optimal* for the simple, fast lattice decoding algorithm used implicitly in essentially all decryption procedures, namely Babai's "round-off" algorithm [Bab85]. The reason is that when decoding a lattice Λ using some basis $\{\mathbf{b}_i\}$, the error tolerance depends inversely on the Euclidean lengths of the vectors dual to $\{\mathbf{b}_i\}$. For $R^\vee$, there is a particular "*decoding*" basis whose dual basis is optimally short (relative to the determinant of R), whereas for R no such basis exists in general.[3] In fact, the decoding basis of $R^\vee$ is simply the dual of the (conjugate of the) powerful basis described above!

In addition to its optimal error tolerance, we also show that the decoding basis has good computational properties. In particular, there are linear-time (or nearly so) algorithms for converting to the decoding basis from the other bases of $R^\vee$ or $R_q^\vee$ that are more appropriate for other computational tasks. And Gaussian errors, especially spherical ones, can be sampled in essentially linear time in the decoding basis.

[1] This is unless m is a power of two, in which case nothing is lost by simply scaling up by exactly $m/2$ to replace $R^\vee$ with R.

[2] By "error rate" here we mean the ratio of the error (in, say, ℓ_2 norm) to the dimension-normalized determinant $\det(\Lambda)^{1/n}$ of the lattice Λ, so exact scaling has no effect on the error rate.

[3] We note that decoding by "lifting" R to the larger-dimensional ring $\mathbb{Z}[X]/(X^m - 1)$, as done in [GHS12a], still leads to at least an $m/2$ factor loss in error tolerance overall, because some inherent loss is already incurred when replacing $R^\vee$ with R, and a bit more is lost in the lifting procedure.

Notation	Description	See
m, $n = \varphi(m)$, $\hat{m}$	The cyclotomic *index*, a positive integer having prime-power factorization $m = \prod_\ell m_\ell$, so that $n = \prod_\ell \varphi(m_\ell)$. Also, $\hat{m} = m/2$ if m is even, otherwise $\hat{m} = m$.	
$K = \mathbb{Q}(\zeta_m)$ $\cong \mathbb{Q}[X]/(\Phi_m(X))$ $\cong \bigotimes_\ell \mathbb{Q}(\zeta_{m_\ell})$	The mth *cyclotomic number field*, where ζ_m denotes an abstract element having order m over $\mathbb{Q}$. (Here $\Phi_m(X) \in \mathbb{Z}[X]$ is the mth cyclotomic polynomial, the minimal polynomial of ζ_m, which has degree n.) It is best viewed as the tensor product of the cyclotomic subfields $\mathbb{Q}(\zeta_{m_\ell})$.	§2.5.1
$\sigma\colon K \to \mathbb{C}^n$	The *canonical embedding* of K, which endows K with a geometry, e.g., $\|a\|_2 := \|\sigma(a)\|_2$ for $a \in K$. Both addition and multiplication in K correspond to their coordinate-wise counterparts in $\mathbb{C}^n$, yielding tight bounds on "expansion" under ring operations.	§2.5.2
$R = \mathbb{Z}[\zeta_m]$ $\cong \mathbb{Z}[X]/(\Phi_m(X))$ $\cong \bigotimes_\ell \mathbb{Z}[\zeta_{m_\ell}]$	The *ring of integers* of K. It is best viewed as a tensor product of subrings $R_\ell = \mathbb{Z}[\zeta_{m_\ell}]$.	§2.5.3
$R^\vee = \langle t^{-1} \rangle$, $g, t \in R$	The *dual* fractional ideal of R, generated by $t^{-1} = g/\hat{m}$, so $R \subseteq R^\vee \subseteq \hat{m}^{-1}R$. Each of $R^\vee$, g, and t can be seen as the tensor products of their counterparts in the subfields $\mathbb{Q}(\zeta_{m_\ell})$.	§2.5.4
$\vec{p} \subset R$	The "*powerful*" $\mathbb{Z}$-basis of R, defined as the tensor product of the power $\mathbb{Z}$-bases of each $\mathbb{Z}[\zeta_{m_\ell}]$. For non-prime-power m, it differs from the power $\mathbb{Z}$-basis $\{\zeta_m^0, \zeta_m^1, \ldots, \zeta_m^{n-1}\}$ often used to represent $\mathbb{Z}[\zeta_m]$, and has better computational and geometric properties.	§4
$\vec{c} \subset R_q$	The "*Chinese remainder*" (CRT) $\mathbb{Z}_q$-basis of $R_q = R/qR$, for any prime $q = 1 \bmod m$. It yields linear-time addition and multiplication in R_q, and there is an $O(n \log n)$-time algorithm for converting between $\vec{c}$ and $\vec{p}$ (as a $\mathbb{Z}_q$-basis of R_q).	§2.5.5, §5
$\vec{d} \subset R^\vee$	The "*decoding*" $\mathbb{Z}$-basis of $R^\vee$, defined as the dual of the (conjugate of the) powerful basis $\vec{p}$. It is used for optimal decoding of $R^\vee$ and its powers, and for efficiently sampling Gaussians.	§6

Figure 1: *Dramatis Personæ*.

1.3 Organization

We draw the reader's attention to Figure 1, which provides a glossary of the main algebraic objects and notation used in this work, and pointers to further discussion of their properties. The rest of the paper is organized as follows:

Section 2 Covers background on our (unusual, but useful) notation for vectors, matrices and tensors; Gaussian and subgaussian random variables; lattices and basic decoding/discretization algorithms; algebraic number theory; and ring-LWE. For the reader with some background in algebraic number theory, we draw attention to the lesser-known material in Section 2.5.1 on the tensorial decomposition into prime-power cyclotomics, and Section 2.5.4 on duality ($R^\vee$, dual bases, etc.).

Section 3 Recalls a "sparse decomposition" of the discrete Fourier transform (DFT) matrix, and develops a novel sparse decomposition for a closely related one that we call the "Chinese remainder transform," which plays a central role in many of our fast algorithms.

Section 4 Defines the "powerful" $\mathbb{Z}$-basis $\vec{p}$ of R and describes its algebraic and geometric properties.

Section 5 Defines the "Chinese remainder" $\mathbb{Z}_q$-basis $\vec{c}$ of R_q, gives its connection to the powerful basis, and describes how it enables fast ring operations.

Section 6 Defines the "decoding" basis $\vec{d}$ of $R^\vee$, gives its connection to the powerful basis, describes how it is used for decoding with optimal noise tolerance, and shows how to efficiently generate (continuous) Gaussians as represented in the decoding basis.

Section 7 Gives a regularity lemma for random lattices over arbitrary cyclotomics. This is needed for only one of our applications, as well as for adapting prior signature schemes and LWE-based (hierarchical) identity-based encryption schemes to the ring setting.

Section 8 Gives some applications of the toolkit: two basic public-key encryption schemes, and a "somewhat homomorphic" symmetric-key encryption scheme.

Acknowledgments. We thank Markus Püschel for his help with the sparse decomposition of the "Chinese remainder transform," and Damien Stehlé for useful discussions.

2 Preliminaries

For a positive integer k, we let $[k]$ denote the set $\{0, \ldots, k-1\}$. For any $\bar{a} \in \mathbb{R}/\mathbb{Z}$, we let $[\![\bar{a}]\!] \in \mathbb{R}$ denote the unique representative $a \in (\bar{a} + \mathbb{Z}) \cap [-1/2, 1/2)$. Similarly, for $\bar{a} \in \mathbb{Z}_q = \mathbb{Z}/q\mathbb{Z}$ we let $[\![\bar{a}]\!]$ denote the unique representative $a \in (\bar{a} + q\mathbb{Z}) \cap [-q/2, q/2)$. We extend $[\![\cdot]\!]$ entrywise to vectors and matrices. The *radical* of a positive integer m, denoted $\mathrm{rad}(m)$, is the product of all primes dividing m.

For a vector $\mathbf{x}$ over $\mathbb{R}$ or $\mathbb{C}$, define the ℓ_2 norm as $\|\mathbf{x}\|_2 = (\sum_i |x_i|^2)^{1/2}$, and the ℓ_∞ norm as $\|\mathbf{x}\|_\infty = \max_i |x_i|$. For an n-by-n matrix M we denote by $s_1(M)$ its largest singular value (also known as the spectral or operator norm), and by $s_n(M)$ its smallest singular value.

2.1 Vectors, Matrices, and Tensors

Throughout this paper, the entries of a vector over a domain D are always indexed (in no particular order) by some finite set S, and we write D^S to denote the set of all such vectors. When the domain is $\mathbb{Z}_q$ or a subset of the complex numbers, we usually denote vectors using bold lower-case letters (e.g., $\mathbf{a}$), otherwise we use arrow notation (e.g., $\vec{a}$). Similarly, the rows and columns of an "R-by-C matrix" over D are indexed by some finite sets R and C, respectively. We write $D^{R \times C}$ for the set of all such matrices, and typically use upper-case letters to denote individual matrices (e.g., A). The R-by-R identity matrix I_R has 1 as its (i,i)th entry for each $i \in R$, and 0 elsewhere. All the standard matrix and vector operations are defined in the natural way, for objects having compatible domains and index sets.

In particular, the Kronecker (or tensor) product $M = A \otimes B$ of an R_0-by-C_0 matrix A with an R_1-by-C_1 matrix B is the $(R_0 \times R_1)$-by-$(C_0 \times C_1)$ matrix M with entries $M_{(i_0,i_1),(j_0,j_1)} = A_{i_0,j_0} \cdot B_{i_1,j_1}$. The Kronecker product of two vectors, or of a matrix with a vector, is defined similarly. For positive integers n_0, n_1, we often implicitly identify the index set $[n_0] \times [n_1]$ with $[n_0 n_1]$, using the bijective correspondence

$(i_0, i_1) \leftrightarrow i = i_0 n_1 + i_1$; note that this matches the traditional Kronecker product for *ordered* rows and columns. Similarly, when $m = \prod_\ell m_\ell$ for a set of pairwise coprime positive integers m_ℓ, we often identify the index sets $\mathbb{Z}_m^*$ and $\prod_\ell \mathbb{Z}_{m_\ell}^*$ via the bijection induced by the Chinese remainder theorem. In other settings we reindex a set using another correspondence, which will be described in context.

An important fact about the Kronecker product is the *mixed-product property*: $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$. Using the mixed-product property, a tensor product $A = \bigotimes_\ell A_\ell$ of several matrices can be written as

$$A = \prod_\ell (I \otimes \cdots \otimes I \otimes A_\ell \otimes I \otimes \cdots \otimes I), \tag{2.1}$$

where the identity matrices have the appropriate induced index sets. In particular, if each A_ℓ is a square matrix of dimension n_ℓ, then A is square of dimension $n = \prod_\ell n_\ell$, and multiplication by A reduces to n/n_ℓ parallel multiplications by A_ℓ, in sequence for each value of ℓ (in any order).

2.2 The Space H

When working with cyclotomic number fields and ideal lattices under the canonical embedding (see Section 2.5.2 below), it is convenient to use a subspace $H \subseteq \mathbb{C}^{\mathbb{Z}_m^*}$ (for some integer $m \geq 2$), defined as

$$H = \{\mathbf{x} \in \mathbb{C}^{\mathbb{Z}_m^*} \ : \ x_i = \overline{x_{m-i}}, \ \forall \ i \in \mathbb{Z}_m^*\}.$$

Letting $n = \varphi(m)$, it is not difficult to verify that H (with the inner product induced on it by $\mathbb{C}^{\mathbb{Z}_m^*}$) is isomorphic to $\mathbb{R}^{[n]}$ as an inner product space. For $m = 2$ this is trivial, and for $m > 2$ this can be seen via the $\mathbb{Z}_m^*$-by-$[n]$ unitary basis matrix $B = \frac{1}{\sqrt{2}} \begin{pmatrix} I & \sqrt{-1}J \\ J & -\sqrt{-1}I \end{pmatrix}$ of H, where the $\mathbb{Z}_m^*$-indexed rows are shown in increasing order according to their representatives in $\{1, \ldots, m-1\}$, the $[n]$-indexed columns are shown in increasing order by index, I is the identity matrix, and J is the reversal matrix (obtained by reversing the columns of I).

We equip H with the ℓ_2 and ℓ_∞ norms induced on it from $\mathbb{C}^{\mathbb{Z}_m^*}$. Namely, for $\mathbf{x} \in H$ we have $\|\mathbf{x}\|_2 = \sum_i (|x_i|^2)^{1/2} = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$, and $\|\mathbf{x}\|_\infty = \max_i |x_i|$.

Gram-Schmidt orthogonalization. For an ordered set $B = \{\mathbf{b}_j\}_{j \in [n]} \subset H$ of linearly independent vectors, the Gram-Schmidt orthogonalization $\widetilde{B} = \{\widetilde{\mathbf{b}}_j\}$ is defined iteratively as follows: $\widetilde{\mathbf{b}}_0 = \mathbf{b}_0$, and for $j = 1, 2, \ldots, n-1$, $\widetilde{\mathbf{b}}_j$ is the component of $\mathbf{b}_j$ orthogonal to the linear span of $\mathbf{b}_0, \ldots, \mathbf{b}_{j-1}$:

$$\widetilde{\mathbf{b}}_j = \mathbf{b}_j - \sum_{k \in [j]} \widetilde{\mathbf{b}}_k \cdot \langle \mathbf{b}_j, \widetilde{\mathbf{b}}_k \rangle / \langle \widetilde{\mathbf{b}}_k, \widetilde{\mathbf{b}}_k \rangle.$$

Viewing B as a matrix whose columns are the vectors $\mathbf{b}_j$, its orthogonalization corresponds to the unique factorization $B = QDU$, where Q is unitary with columns $\widetilde{\mathbf{b}}_j / \|\widetilde{\mathbf{b}}_j\|_2$; D is real diagonal with positive diagonal entries $\|\widetilde{\mathbf{b}}_j\|_2 > 0$; and U is real upper unitriangular with entries $w_{k,j} = \langle \mathbf{b}_j, \widetilde{\mathbf{b}}_k \rangle / \langle \widetilde{\mathbf{b}}_k, \widetilde{\mathbf{b}}_k \rangle$.[4] The Gram-Schmidt orthogonalization is $\widetilde{B} = QD$, and so $B = \widetilde{B}U$. The real positive definite Gram matrix of B is $B^*B = U^T D^2 U$. Because U is upper unitriangular, this is exactly the Cholesky decomposition of B^*B, which is unique; it therefore determines the matrices D, U in the Gram-Schmidt orthogonalization of B. One can also verify from the definitions that D^2 and U are both rational if the Gram matrix is rational.

[4] This is often referred to as the "QR" factorization, though here we have also factored out the diagonal entries of the upper-triangular matrix R into D, making U unitriangular.

2.3 Gaussians and Subgaussian Random Variables

For $s > 0$, define the Gaussian function $\rho_s \colon H \to (0, 1]$ as $\rho_s(\mathbf{x}) = \exp(-\pi\langle \mathbf{x}, \mathbf{x}\rangle/s^2) = \exp(-\pi\|\mathbf{x}\|_2^2/s^2)$. By normalizing this function we obtain the *continuous* Gaussian probability distribution D_s of parameter s, whose density is given by $s^{-n} \cdot \rho_s(\mathbf{x})$.

For much of our analysis it is convenient to use the standard notion of *subgaussian* random variables, relaxed slightly as in [MP12]. (For further details and full proofs, see, e.g., [Ver11].) For any $\delta \geq 0$, we say that a random variable X (or its distribution) over $\mathbb{R}$ is δ-*subgaussian* with parameter $s > 0$ if for all $t \in \mathbb{R}$, the (scaled) moment-generating function satisfies

$$\mathbb{E}[\exp(2\pi t X)] \leq \exp(\delta) \cdot \exp(\pi s^2 t^2).$$

Notice that the $\exp(\pi s^2 t^2)$ term on the right is exactly the (scaled) moment-generating function of the one-dimensional Gaussian distribution of parameter s over $\mathbb{R}$. It is easy to see that if X is δ-subgaussian with parameter s, then cX is δ-subgaussian with parameter $|c|s$ for any real c. In addition, by Markov's inequality, the tails of X are dominated by those of a Gaussian of parameter s, i.e., for all $t \geq 0$,

$$\Pr[|X| \geq t] \leq 2\exp(\delta - \pi t^2/s^2). \tag{2.2}$$

Using the inequality $\cosh(x) \leq \exp(x^2/2)$, it can be shown that any B-bounded centered random variable X (i.e., $\mathbb{E}[X] = 0$ and $|X| \leq B$ always) is 0-subgaussian with parameter $B\sqrt{2\pi}$.

The sum of independent subgaussian variables is easily seen to be subgaussian. Here we observe that the same holds even in a martingale-like setting.

Claim 2.1. *Let $\delta_i, s_i \geq 0$ and X_i be random variables for $i = 1, \ldots, k$. Suppose that for every i, when conditioning on any values of $X_1, \ldots, X_{i-1}$, the random variable X_i is δ_i-subgaussian with parameter s_i. Then $\sum X_i$ is $(\sum \delta_i)$-subgaussian with parameter $(\sum s_i^2)^{1/2}$.*

Proof. It suffices to prove the claim for $k = 2$; the general case follows by induction, since X_k is subgaussian conditioned on any value of $\sum_{i=1}^{k-1} X_i$. Indeed,

$$\mathbb{E}\big[\exp(2\pi t(X_1 + X_2))\big] = \mathbb{E}_{X_1}\Big[\exp(2\pi t X_1)\,\mathbb{E}_{X_2}\big[\exp(2\pi t X_2) \mid X_1\big]\Big] \leq \exp(\delta_1 + \delta_2)\exp(\pi(s_1^2 + s_2^2)t^2).$$

□

We also have the following bound on the tail of a sum of squares of independent subgaussian variables.

Lemma 2.2. *Let X be a δ-subgaussian random variable with parameter s. Then, for any $t \in (0, 1/(2s^2))$,*

$$\mathbb{E}\big[\exp(2\pi t X^2)\big] \leq 1 + 2\exp(\delta)\left(\frac{1}{2ts^2} - 1\right)^{-1}.$$

Moreover, if $X_1, \ldots, X_k$ are random variables, each of which is δ-subgaussian with parameter s conditioned on any values of the previous ones, then for any $r > k's^2/\pi$ where $k' = 2k\exp(\delta)$ we have that

$$\Pr\Big[\sum_i X_i^2 > r\Big] \leq \exp\Big(k'\Big(2\Big(\frac{\pi r}{k's^2}\Big)^{1/2} - \frac{\pi r}{k's^2} - 1\Big)\Big).$$

In particular, using the inequality $2\alpha^{1/2} - \alpha - 1 \leq -\alpha/4$ valid for all $\alpha \geq 4$, we obtain that for any $r \geq 4k's^2/\pi$,

$$\Pr\Big[\sum_i X_i^2 > r\Big] \leq \exp\Big(-\frac{\pi r}{4s^2}\Big).$$

Proof. Using integration by parts and (2.2),

$$\begin{aligned}\mathbb{E}\big[\exp(2\pi tX^2)\big] &= 1+\int_0^\infty \Pr[|X|\geq r]\cdot 4\pi tr\exp(2\pi tr^2)\mathrm{d}r\\ &\leq 1+8\pi t\exp(\delta)\int_0^\infty r\exp(-\pi r^2/s^2+2\pi tr^2)\mathrm{d}r\\ &= 1+2\exp(\delta)\Big(\frac{1}{2ts^2}-1\Big)^{-1}\\ &\leq \exp\Big(2\exp(\delta)\Big(\frac{1}{2ts^2}-1\Big)^{-1}\Big),\end{aligned}$$

where the last equality uses that for every $a>0$, $\int_0^\infty r\exp(-ar^2)\mathrm{d}r=(2a)^{-1}$. This completes the first part of the lemma. For the second part, notice that by the above, if $X_1,\ldots,X_k$ are as in the statement, we have for any $t\in(0,1/(2s^2))$,

$$\mathbb{E}\Big[\exp(2\pi t\sum_i X_i^2)\Big]\leq \exp\Big(2k\exp(\delta)\Big(\frac{1}{2ts^2}-1\Big)^{-1}\Big),$$

and hence by Markov's inequality, for all $r>0$ and $t\in(0,1/(2s^2))$,

$$\Pr\Big[\sum_i X_i^2>r\Big]\leq \exp\Big(2k\exp(\delta)\Big(\frac{1}{2ts^2}-1\Big)^{-1}-2\pi tr\Big).$$

Letting $x=2s^2t\in(0,1)$ and $A=\pi r/(s^2k')>1$, the expression inside the exponent is

$$2k\exp(\delta)\Big(\Big(\frac{1}{x}-1\Big)^{-1}-Ax\Big).$$

The lemma follows using the fact that for any $A>1$, the minimum over $x\in(0,1)$ of the expression inside the parenthesis is $2\sqrt{A}-A-1$ (obtained at $1-1/\sqrt{A}$). □

We extend the notion of subgaussianity to random vectors in $\mathbb{R}^n$ (or equivalently, in H). Specifically, we say that a random vector X in $\mathbb{R}^n$ is δ-subgaussian with parameter s if for all unit vectors $\mathbf{u}\in\mathbb{R}^n$, the random variable $\langle X,\mathbf{u}\rangle$ is δ-subgaussian with parameter s. It follows from Claim 2.1 that if the coordinates of a random vector in $\mathbb{R}^n$ are independent, and each is δ-subgaussian with parameter s, then the random vector is $n\delta$-subgaussian with the same parameter s.

Sums of subgaussian random vectors are again easily seen to be subgaussian, even in the martingale setting as in Claim 2.1 above. We summarize this in the following corollary, which considers the more general setting in which we apply a (possibly different) linear transformation to each subgaussian random vector.

Corollary 2.3. *Let $\delta_i,s_i\geq 0$ and X_i be random vectors in $\mathbb{R}^n$ (or in H), and let A_i be $n\times n$ matrices for $i=1,\ldots,k$. Suppose that for every i, when conditioning on any values of $X_1,\ldots,X_{i-1}$, the random vector X_i is δ_i-subgaussian with parameter s_i. Then $\sum A_iX_i$ is $(\sum\delta_i)$-subgaussian with parameter $\lambda_{max}(\sum s_i^2A_iA_i^T)^{1/2}$, where λ_{max} denotes the largest eigenvalue.*

Proof. For any vector $\mathbf{u}\in\mathbb{R}^n$,

$$\Big\langle\sum_i A_iX_i,\mathbf{u}\Big\rangle=\sum_i\langle A_iX_i,\mathbf{u}\rangle=\sum_i\langle X_i,A_i^T\mathbf{u}\rangle,$$

which is a sum of random variables satisfying that for each i, the ith variable is δ_i-subgaussian with parameter $s_i\|A_i^T\mathbf{u}\|_2$ conditioned on any value of the previous ones. By Claim 2.1, this sum is $(\sum \delta_i)$-subgaussian with parameter

$$\Big(\sum_i s_i^2\|A_i^T\mathbf{u}\|_2^2\Big)^{1/2} = \Big(\mathbf{u}^T\big(\sum_i s_i^2 A_i A_i^T\big)\mathbf{u}\Big)^{1/2},$$

whose maximum over all unit vectors $\mathbf{u}$ is $\lambda_{\max}(\sum_i s_i^2 A_i A_i^T)^{1/2}$. □

By applying Corollary 2.3 with the linear transformation induced by coordinate-wise multiplication in $H \subset \mathbb{C}^{\mathbb{Z}_m^*}$ we obtain the following.

Claim 2.4. *If X is a δ-subgaussian with parameter s in H, and $\mathbf{z} \in H$ is any element, then the coordinate-wise multiplication $\mathbf{z} \odot X \in H$ is δ-subgaussian with parameter $\|\mathbf{z}\|_\infty \cdot s$. More generally, if $X_j \in H$ are random vectors satisfying the property in Corollary 2.3 for some $\delta_j, s_j \geq 0$ (respectively), then for any $\mathbf{z}_j \in H$, we have that $\sum_j \mathbf{z}_j \odot X_j \in H$ is $(\sum \delta_j)$-subgaussian with parameter $\max_{i\in\mathbb{Z}_m^*}(\sum_j s_j^2|(\mathbf{z}_j)_i|^2)^{1/2}$.*

2.4 Lattice Background

We define a *lattice* as a discrete additive subgroup of H. We deal here exclusively with full-rank lattices, which are generated as the set of all integer linear combinations of some set of n linearly independent *basis* vectors $B = \{\mathbf{b}_j\} \subset H$:

$$\Lambda = \mathcal{L}(B) = \Big\{\sum\nolimits_j z_j\mathbf{b}_j \ : \ z_j \in \mathbb{Z}\Big\}.$$

Two bases B, B' generate the same lattice if and only if there exists a unimodular matrix U (i.e., integer matrix with determinant ± 1) such that $BU = B'$. The *determinant* of a lattice $\mathcal{L}(B)$ is defined as $|\det(B)|$, which is independent of the choice of basis B. The *minimum distance* $\lambda_1(\Lambda)$ of a lattice Λ (in the Euclidean norm) is the length of a shortest nonzero lattice vector: $\lambda_1(\Lambda) = \min_{\mathbf{0}\neq\mathbf{x}\in\Lambda}\|\mathbf{x}\|_2$.

The *dual lattice* of $\Lambda \subset H$ is defined as $\Lambda^\vee = \{\mathbf{y} \in H : \forall\, \mathbf{x} \in \Lambda,\ \langle \mathbf{x}, \overline{\mathbf{y}}\rangle = \sum_i x_i y_i \in \mathbb{Z}\}$. Notice that this is actually the *complex conjugate* of the dual lattice as usually defined in $\mathbb{C}^n$; our definition corresponds more naturally to the notion of duality in algebraic number theory (see Section 2.5.4). All of the properties of the dual lattice that we use also hold for the conjugate dual. In particular, $\det(\Lambda^\vee)$ is $\det(\Lambda)^{-1}$.

It is easy to see that $(\Lambda^\vee)^\vee = \Lambda$. If $B = \{\mathbf{b}_j\} \subset H$ is a set of linearly independent vectors (i.e., an $\mathbb{R}$-basis of H), its *dual basis* $D = \{\mathbf{d}_j\}$ is characterized by $\langle \mathbf{b}_j, \overline{\mathbf{d}_k}\rangle = \delta_{jk}$, where δ_{jk} is the Kronecker delta. It is easy to verify that $\mathcal{L}(D) = \mathcal{L}(B)^\vee$.

Micciancio and Regev [MR04] introduced a lattice quantity called the *smoothing parameter*, and related it to various lattice quantities.

Definition 2.5. *For a lattice Λ and positive real $\varepsilon > 0$, the* smoothing parameter $\eta_\varepsilon(\Lambda)$ *is the smallest s such that $\rho_{1/s}(\Lambda^\vee\backslash\{\mathbf{0}\}) \leq \varepsilon$.*

Lemma 2.6 ([MR04, Lemma 3.2]). *For any n-dimensional lattice Λ, we have $\eta_{2^{-2n}}(\Lambda) \leq \sqrt{n}/\lambda_1(\Lambda^\vee)$.*[5]

Lemma 2.7 ([Reg05, Claim 3.8]). *For any lattice Λ, real $\varepsilon > 0$ and $s \geq \eta_\varepsilon(\Lambda)$, and $\mathbf{c} \in H$, we have $\rho_s(\Lambda + \mathbf{c}) \in [1 \pm \varepsilon] \cdot s^n \det(\Lambda)^{-1}$.*

[5]Note that we are using $\varepsilon = 2^{-2n}$ instead of 2^{-n} as in [MR04], but the stronger bound holds by the same proof.

For a lattice coset $\Lambda + \mathbf{c}$ and real $s > 0$, define the *discrete Gaussian* probability distribution over $\Lambda + \mathbf{c}$ with parameter s as

$$D_{\Lambda+\mathbf{c},s}(\mathbf{x}) = \frac{\rho_s(\mathbf{x})}{\rho_s(\Lambda + \mathbf{c})} \quad \forall\, \mathbf{x} \in \Lambda + \mathbf{c}. \tag{2.3}$$

It is known to satisfy the following concentration bound.

Lemma 2.8 ([Ban93, Lemma 1.5(i)]). *For any n-dimensional lattice Λ and $s > 0$, a point sampled from $D_{\Lambda,s}$ has Euclidean norm at most $s\sqrt{n}$, except with probability at most 2^{-2n}.*

Gentry, Peikert, and Vaikuntanathan [GPV08] showed how to efficiently sample from a discrete Gaussian, using any lattice basis consisting of sufficiently short orthogonalized vectors.

Lemma 2.9 ([GPV08, Theorem 4.1]). *There is an efficient algorithm that samples to within $\mathrm{negl}(n)$ statistical distance of $D_{\Lambda+\mathbf{c},s}$, given $\mathbf{c} \in H$, a basis B of Λ, and a parameter $s \geq \max_j \|\widetilde{\mathbf{b}}_j\| \cdot \omega(\sqrt{\log n})$, where $\widetilde{B} = \{\widetilde{\mathbf{b}}_j\}$ is the Gram-Schmidt orthogonalization of B.*

We make a few remarks on the implementation of the algorithm from Lemma 2.9. It is a randomized variant of Babai's "nearest plane" algorithm [Bab85] (a related variant was also considered by Klein [Kle00] for a different problem). On input $\mathbf{c} \in H$, and a basis B and parameter s satisfying the above constraint, it does the following: for $j = n-1, \ldots, 0$, let $\mathbf{c} \leftarrow \mathbf{c} - z_j \mathbf{b}_j$, where $z_j \leftarrow c'_j + D_{\mathbb{Z}-c'_j,s_j}$ for $c'_j = \langle \mathbf{c}, \widetilde{\mathbf{b}}_j \rangle / \langle \widetilde{\mathbf{b}}_j, \widetilde{\mathbf{b}}_j \rangle$ and $s_j = s/\|\widetilde{\mathbf{b}}_j\|_2$. Output the final value of $\mathbf{c}$.

In practice, the above algorithm is usually invoked on a fixed basis B whose Gram matrix B^*B is rational. It is best implemented by precomputing the rational matrices D^2, U associated with $\widetilde{B}$ and B^*B (see Section 2.2), and by representing the input and intermediate values $\mathbf{c}$ using rational coefficient vectors with respect to B. Then each value $c'_j = \langle \mathbf{c}, \widetilde{\mathbf{b}}_j \rangle / \langle \widetilde{\mathbf{b}}_j, \widetilde{\mathbf{b}}_j \rangle$ can be computed simply as the inner product of $\mathbf{c}$'s coefficient vector with the jth row of U.

2.4.1 Decoding

In many applications we need to perform the following algorithmic task, which is essentially a bounded-distance decoding. Let Λ be a known fixed lattice, and let $\mathbf{x} \in H$ be an unknown short vector. The goal is to recover $\mathbf{x}$, given $\mathbf{t} = \mathbf{x} \bmod \Lambda$. Although there are several possible algorithms for this task, here we focus on a slight extension of the so-called "round-off" algorithm originally due to Babai [Bab85]. This is due to its high efficiency and because for our purposes it performs optimally (or nearly so). The algorithm is very simple: let $\{\mathbf{v}_i\}$ be a fixed set of n linearly independent (and typically short) vectors in the dual lattice $\Lambda^\vee$. Denote the dual basis of $\{\mathbf{v}_i\}$ by $\{\mathbf{b}_i\}$, and let $\Lambda' \supseteq \Lambda$ be the superlattice generated by $\{\mathbf{b}_i\}$. Given an input $\mathbf{t} = \mathbf{x} \bmod \Lambda$, we express $\mathbf{t} \bmod \Lambda'$ in the basis $\{\mathbf{b}_i\}$ as $\sum_i c_i \mathbf{b}_i$, where $c_i \in \mathbb{R}/\mathbb{Z}$ (so $c_i = \langle \mathbf{x}, \overline{\mathbf{v}_i} \rangle \bmod 1$), and output $\sum_i \lfloor c_i \rceil \mathbf{b}_i \in H$.

Claim 2.10. *Let $\Lambda \subset H$ be a lattice, let $\{\mathbf{v}_i\} \subset \Lambda^\vee$ be a set of n linearly independent vectors in its dual, and let $\{\mathbf{b}_i\} \subset \Lambda$ denote the dual basis of $\{\mathbf{v}_i\}$. The above round-off algorithm, given input $\mathbf{x} \bmod \Lambda$, outputs $\mathbf{x}$ if and only if all the coefficients $a_i = \langle \mathbf{x}, \overline{\mathbf{v}_i} \rangle \in \mathbb{R}$ in the expansion $\mathbf{x} = \sum_i a_i \mathbf{b}_i$ are in $[-1/2, 1/2)$.*

We remark that in Babai's round-off algorithm one often assumes that $\{\mathbf{v}_i\}$ is a *basis* of $\Lambda^\vee$ (and hence $\{\mathbf{b}_i\}$ is a basis of Λ), whereas here we consider the more general case where $\{\mathbf{v}_i\}$ can be an arbitrary set of linearly independent vectors in $\Lambda^\vee$. For some lattices (including those appearing in our applications) this can make a big difference. Consider for instance the lattice of all points in $\mathbb{Z}^n$ whose coordinates sum to an even

number. The dual of this lattice is $\mathbb{Z}^n \cup (\mathbb{Z}^n + (1, \ldots, 1)/2)$, and clearly any basis of this dual must contain a vector of length at least $\sqrt{n}/2$. As a result, when limited to using a basis, the round-off algorithm can fail for vectors of length greater than $1/\sqrt{n}$. However, the dual lattice clearly has a set of n linearly independent vectors of length 1, allowing us to decode up to length $1/2$.

2.4.2 Discretization

We now consider another algorithmic task related to the one in the previous subsection. This task shows up in applications, such as when converting a continuous Gaussian into a discrete Gaussian-like distribution. Given a lattice $\Lambda = \mathcal{L}(B)$ represented by a "good" basis $B = \{\mathbf{b}_i\}$, a point $\mathbf{x} \in H$, and a point $\mathbf{c} \in H$ representing a lattice coset $\Lambda + \mathbf{c}$, the goal is to discretize $\mathbf{x}$ to a point $\mathbf{y} \in \Lambda + \mathbf{c}$, written $\mathbf{y} \leftarrow \lfloor \mathbf{x} \rceil_{\Lambda+\mathbf{c}}$, so that the length (or subgaussian parameter) of $\mathbf{y} - \mathbf{x}$ is not too large. To do this, we sample a relatively short offset vector $\mathbf{f}$ from the coset $\Lambda + \mathbf{c}' = \Lambda + (\mathbf{c} - \mathbf{x})$ in one of a few natural ways described below, and output $\mathbf{y} = \mathbf{x} + \mathbf{f}$. We require that the method used to choose $\mathbf{f}$ be efficient and depend only on the desired coset $\Lambda + \mathbf{c}'$, not on the particular representative used to specify it; we call such a procedure (or the induced discretization) *valid*.

Note that for a valid discretization, $\lfloor \mathbf{z} + \mathbf{x} \rceil_{\Lambda+\mathbf{c}}$ and $\mathbf{z} + \lfloor \mathbf{x} \rceil_{\Lambda+\mathbf{c}}$ are identically distributed for any $\mathbf{z} \in \Lambda$. Therefore, for any sublattice $\Lambda' \subseteq \Lambda$, a valid discretization also induces a well-defined discretization from any *coset* $\bar{\mathbf{x}} = \Lambda' + \mathbf{x}$ to $\bar{\mathbf{y}} = \bar{\mathbf{x}} + \mathbf{f} = \Lambda' + \mathbf{y}$, where $\mathbf{y} \in \Lambda + \mathbf{c}$.

There are several valid ways of sampling $\mathbf{f}$, offering tradeoffs between efficiency and output guarantees:

- A particularly simple and efficient method is "coordinate-wise randomized rounding:" given a coset $\Lambda + \mathbf{c}'$, we represent $\mathbf{c}'$ in the basis B as $\mathbf{c}' = \sum_i a_i \mathbf{b}_i \bmod \Lambda$ for some coefficients $a_i \in [0, 1)$, then randomly and independently choose each f_i from $\{a_i - 1, a_i\}$ to have expectation zero, and output $\mathbf{f} = \sum_i f_i \mathbf{b}_i \in \Lambda + \mathbf{c}'$. The validity of this procedure is immediate, since any representative of $\Lambda + \mathbf{c}'$ induces the same a_i values. Because each f_i has expectation zero and is bounded by 1 in magnitude, it is 0-subgaussian with parameter $\sqrt{2\pi}$ (see Section 2.3), and hence so is the entire vector of f_i values. By Corollary 2.3 (applied with just one random vector), we conclude that $\mathbf{f}$ is 0-subgaussian with parameter $\sqrt{2\pi} \cdot s_1(B)$.

- In some settings we can use a *deterministic* version of the above method, where we instead compute coefficients $a_i \in [-1/2, 1/2)$ and simply output $\mathbf{f} = \sum_i a_i \mathbf{b}_i$. When, for example, $\mathbf{x}$ comes from a sufficiently wide continuous Gaussian, this method yields $\mathbf{y} = \mathbf{x} + \mathbf{f}$ having a (very slightly) better subgaussian parameter than the randomized method. However, the analysis is a bit more involved, and we omit it.

- If $\mathbf{x}$ has a continuous or discrete Gaussian distribution, then using more sophisticated rounding methods it is possible to make $\mathbf{y}$ also be distributed according to a true discrete Gaussian (of some particular covariance), which is needed in some applications (though not any we develop in this paper). By [Pei10, Theorem 3.1], under mild conditions it suffices for $\mathbf{f}$ to be distributed as a discrete Gaussian over $\Lambda + \mathbf{c}'$, and the covariance parameter of $\mathbf{y}$ will be the sum of those of $\mathbf{x}$ and $\mathbf{f}$. Using the algorithm from Lemma 2.9, we can sample a discrete Gaussian $\mathbf{f}$ with parameter bounded by $\max_j \|\widetilde{\mathbf{b}_j}\| \cdot \omega(\sqrt{\log n})$. Alternatively, a simpler and more efficient randomized round-off algorithm obtains a parameter bounded by $s_1(B) \cdot \omega(\sqrt{\log n})$ [Pei10]. Both of these methods are easily seen to be valid, though note that they yield slightly worse Gaussian parameters than the two simpler methods described above.

2.5 Algebraic Number Theory Background

Algebraic number theory is the study of *number fields*. Here we review the necessary background, specialized to the case of *cyclotomic* number fields, which are the only kind we use in this work. More background and complete proofs can be found in any introductory book on the subject, e.g., [Ste04, Lan94], and especially the latter reference for material related to the tensorial decomposition.

2.5.1 Cyclotomic Number Fields and Polynomials

For a positive integer m, the mth *cyclotomic number field* is a field extension $K = \mathbb{Q}(\zeta_m)$ obtained by adjoining an element ζ_m of order m (i.e., a primitive mth root of unity) to the rationals. (Note that we view ζ_m as an abstract element, and not, for example, as any particular value in $\mathbb{C}$.) The minimal polynomial of ζ_m is the mth *cyclotomic polynomial*

$$\Phi_m(X) = \prod_{i \in \mathbb{Z}_m^*} (X - \omega_m^i) \in \mathbb{Z}[X], \tag{2.4}$$

where $\omega_m \in \mathbb{C}$ is any primitive mth root of unity in $\mathbb{C}$, e.g., $\omega_m = \exp(2\pi\sqrt{-1}/m)$. Therefore, there is a natural isomorphism between K and $\mathbb{Q}[X]/(\Phi_m(X))$, given by $\zeta_m \mapsto X$. Since $\Phi_m(X)$ has degree $n = |\mathbb{Z}_m^*| = \varphi(m)$, we can view K as a vector space of degree n over $\mathbb{Q}$, which has $(\zeta_m^j)_{j \in [n]} = (1, \zeta_m, \ldots, \zeta_m^{n-1}) \in K^{[n]}$ as a basis. This is called the *power basis* of K.

We recall two useful facts about cyclotomic polynomials, which can be verified by examining the roots of both sides of each equation.

Fact 2.11. *For any m, we have $X^m - 1 = \prod_{d|m} \Phi_d(X)$, where d runs over all the positive divisors of m. In particular, $\Phi_p(X) = 1 + X + X^2 + \cdots + X^{p-1}$ for any prime p.*

Fact 2.12. *For any m, we have $\Phi_m(X) = \Phi_{\mathrm{rad}(m)}(X^{m/\mathrm{rad}(m)})$, where recall that $\mathrm{rad}(m)$ is the product of all distinct primes dividing m. In particular, if m is a power of a prime p, then $\Phi_m(X) = \Phi_p(X^{m/p})$.*

For instance, $\Phi_8(X) = 1 + X^4$ and $\Phi_{25}(X) = 1 + X^5 + X^{10} + X^{15} + X^{20}$.

For any m' dividing m, it is often convenient to view $K' = \mathbb{Q}(\zeta_{m'})$ as a subfield of $K = \mathbb{Q}(\zeta_m)$, by identifying $\zeta_{m'}$ with $\zeta_m^{m/m'}$.

Non-prime-power cyclotomics. Not all cyclotomic polynomials are "regular"-looking or have 0-1 (or even small) coefficients. Generally speaking, the irregularity and range of coefficients grows with the number of prime divisors of m. For example, $\Phi_6(X) = X^2 - X + 1$; $\Phi_{3\cdot5\cdot7}(X)$ has 33 monomials with coefficients $-2, -1$, and 1; and $\Phi_{3\cdot5\cdot7\cdot11\cdot13}(X)$ has coefficients of magnitude up to 22. Fortunately, the form of $\Phi_m(X)$ for non-prime-power m will never be a concern in this work, due to an alternative way of viewing $K = \mathbb{Q}(\zeta_m)$ by reducing to the case of prime-power cyclotomics.

To do this we first need to briefly recall the notion of a *tensor product* of fields. Let K, L be two field extensions of $\mathbb{Q}$. Then the field tensor product $K \otimes L$ is defined as the set of all $\mathbb{Q}$-linear combinations of *pure tensors* $a \otimes b$ for $a \in K, b \in L$, where $\otimes$ is $\mathbb{Q}$-bilinear and satisfies the mixed-product property, i.e.,

$$\begin{aligned}
(a_1 \otimes b) + (a_2 \otimes b) &= (a_1 + a_2) \otimes b \\
(a \otimes b_1) + (a \otimes b_2) &= a \otimes (b_1 + b_2) \\
e(a \otimes b) &= (ea) \otimes b = a \otimes (eb) \\
(a_1 \otimes b_1)(a_2 \otimes b_2) &= (a_1 a_2) \otimes (b_1 b_2)
\end{aligned}$$

for all $e \in \mathbb{Q}$. These properties define addition and multiplication in $K \otimes L$, and though the result is not always a field (because it may lack multiplicative inverses), it will always be one whenever we take the tensor product of two cyclotomic fields in this work. It is straightforward to verify that if A, B are $\mathbb{Q}$-bases of K, L respectively, then the Kronecker product $A \otimes B$ is a $\mathbb{Q}$-basis of $K \otimes L$. Later on we also consider tensor products of rings, or more generally of $\mathbb{Z}$-modules. These are defined in the same way, except that they are made up of only the $\mathbb{Z}$-linear combinations of pure tensors. This always yields a ring or $\mathbb{Z}$-module, respectively, with $\mathbb{Z}$-bases obtained by tensoring $\mathbb{Z}$-bases of the original objects.

A key fact from algebraic number theory is the following.

Proposition 2.13. *Let m have prime-power factorization $m = \prod_\ell m_\ell$, i.e., the m_ℓ are powers of distinct primes. Then $K = \mathbb{Q}(\zeta_m)$ is isomorphic to the tensor product $\bigotimes_\ell K_\ell$ of the fields $K_\ell = \mathbb{Q}(\zeta_{m_\ell})$, via the correspondence $\prod_\ell a_\ell \leftrightarrow (\otimes_\ell a_\ell)$, where on the left we implicitly embed each $a_\ell \in K_\ell$ into K.*

2.5.2 Embeddings and Geometry

Here we describe the *embeddings* of a cyclotomic number field, which induce a 'canonical' geometry on it.

The mth cyclotomic number field $K = \mathbb{Q}(\zeta_m)$ of degree $n = \varphi(m)$ has exactly n ring homomorphisms (embeddings) $\sigma_i \colon K \to \mathbb{C}$ that fix every element of $\mathbb{Q}$. Concretely, for each $i \in \mathbb{Z}_m^*$ there is an embedding σ_i defined by $\sigma_i(\zeta_m) = \omega_m^i$, where $\omega_m \in \mathbb{C}$ is some fixed primitive mth root of unity. Clearly, the embeddings come in pairs of complex conjugates, i.e., $\sigma_i = \overline{\sigma_{m-i}}$. The *canonical embedding* $\sigma \colon K \to \mathbb{C}^{\mathbb{Z}_m^*}$ is defined as

$$\sigma(a) = (\sigma_i(a))_{i \in \mathbb{Z}_m^*}.$$

Due to the conjugate pairs, σ actually maps into $H \subset \mathbb{C}^{\mathbb{Z}_m^*}$, defined in Section 2.2. Note that σ is a ring homomorphism from K to H, where multiplication and addition in H are both component-wise.

By identifying K with its canonical embedding into H, we endow K with a canonical geometry. Recalling that norms on H are just those induced from $\mathbb{C}^{\mathbb{Z}_m^*}$, we see that for any $a \in K$, the ℓ_2 norm of a is simply $\|a\|_2 = \|\sigma(a)\|_2 = (\sum_i |\sigma_i(a)|^2)^{1/2}$, and the ℓ_∞ norm is $\max_i |\sigma_i(a)|$. Because multiplication of embedded elements is component-wise, for any $a, b \in K$ we have

$$\|a \cdot b\| \leq \|a\|_\infty \cdot \|b\|, \tag{2.5}$$

where $\|\cdot\|$ denotes either the ℓ_2 or ℓ_∞ norm (or indeed, any ℓ_p norm). Thus the ℓ_∞ norm acts as an "absolute value" for K that bounds how much an element expands any other by multiplication. For example, note that for any power ζ of ζ_m, each $\sigma_i(\zeta)$ must be a root of unity in $\mathbb{C}$, and hence $\|\zeta\|_2 = \sqrt{n}$ and $\|\zeta\|_\infty = 1$.

The *trace* $\mathrm{Tr} = \mathrm{Tr}_{K/\mathbb{Q}} \colon K \to \mathbb{Q}$ can be defined as the sum of the embeddings: $\mathrm{Tr}(a) = \sum_i \sigma_i(a)$. Clearly, the trace is $\mathbb{Q}$-linear: $\mathrm{Tr}(a + b) = \mathrm{Tr}(a) + \mathrm{Tr}(b)$ and $\mathrm{Tr}(c \cdot a) = c \cdot \mathrm{Tr}(a)$ for all $a, b \in K$ and $c \in \mathbb{Q}$. Also notice that

$$\mathrm{Tr}(a \cdot b) = \sum_i \sigma_i(a)\sigma_i(b) = \langle \sigma(a), \overline{\sigma(b)} \rangle,$$

so $\mathrm{Tr}(a \cdot b)$ is a symmetric bilinear form akin to the inner product of the embeddings of a and b. The (field) *norm* $\mathrm{N} = \mathrm{N}_{K/\mathbb{Q}} \colon K \to \mathbb{Q}$ can be defined as the product of all the embeddings: $\mathrm{N}(a) = \prod_i \sigma_i(a)$. Clearly, the norm is multiplicative: $\mathrm{N}(a \cdot b) = \mathrm{N}(a) \cdot \mathrm{N}(b)$.

When taking $K \cong \bigotimes_\ell K_\ell$ as in Proposition 2.13, it follows directly from the definitions that σ is the tensor product of the canonical embeddings $\sigma^{(\ell)}$ of K_ℓ, i.e.,

$$\sigma(\otimes_\ell a_\ell) = \bigotimes_\ell \sigma^{(\ell)}(a_\ell). \tag{2.6}$$

(Here the index set of σ is $\prod_\ell \mathbb{Z}_{m_\ell}^*$, which corresponds bijectively to $\mathbb{Z}_m^*$ via the Chinese remainder theorem.) This decomposition of σ in turn implies that the trace decomposes as

$$\mathrm{Tr}_{K/\mathbb{Q}}(\otimes_\ell a_\ell) = \prod\nolimits_\ell \mathrm{Tr}_{K_\ell/\mathbb{Q}}(a_\ell). \tag{2.7}$$

Using the canonical embedding also allows us to think of the Gaussian distribution D_r over H as a distribution over K, or more accurately, over the field tensor product $K_\mathbb{R} = K \otimes \mathbb{R}$, which is isomorphic as a real vector space to H via σ. For our purposes it is usually helpful to ignore the distinction between K and $K_\mathbb{R}$, and to approximate the latter by the former using sufficient precision.

2.5.3 Ring of Integers and Its Ideals

Let $R \subset K$ denote the set of all algebraic integers in a number field K. This set forms a ring (under the usual addition and multiplication operations in K), called the *ring of integers* of K. Note that the trace and norm of an algebraic integer are rational integers (i.e., in $\mathbb{Z}$), so we have the induced functions $\mathrm{Tr}, \mathrm{N} \colon R \to \mathbb{Z}$.

For the mth cyclotomic number field $K = \mathbb{Q}(\zeta_m)$ of degree $n = \varphi(m)$, the ring of integers happens to be $R = \mathbb{Z}[\zeta_m] \cong \mathbb{Z}[X]/\Phi_m(X)$, and hence has the power basis $\{\zeta_m^j\}_{j \in [n]}$ as a $\mathbb{Z}$-basis. Alternatively—and this is the view we adopt throughout the paper—we can view $R \cong \bigotimes_\ell R_\ell$ as a tensor product of the rings of integers R_ℓ in $K_\ell = \mathbb{Q}(\zeta_{m_\ell})$, where $m = \prod_\ell m_\ell$ is the prime-power factorization of m.

The (absolute) *discriminant* Δ_K of K is a measure of the geometric sparsity of its ring of integers, defined as $\Delta_K = \det(\sigma(R))^2$, the squared determinant of the lattice $\sigma(R)$.[6] The discriminant of the mth cyclotomic number field is

$$\Delta_K = \left(\frac{m}{\prod\limits_{\text{prime } p \mid m} p^{1/(p-1)}} \right)^n \leq n^n, \tag{2.8}$$

where the product in the denominator runs over all primes p dividing m. The above inequality is tight exactly when m is a power of two.

An *(integral) ideal* $\mathcal{I} \subseteq R$ is a nontrivial (i.e., $I \neq \emptyset$ and $\mathcal{I} \neq \{0\}$) additive subgroup that is closed under multiplication by R, i.e., $r \cdot a \in \mathcal{I}$ for any $r \in R$ and $a \in \mathcal{I}$.[7] A *principal* ideal $\mathcal{I}$ is one that is generated by a single element, i.e., $\mathcal{I} = uR$ for some $u \in R$ which is unique up to multiplication by units in R; we sometimes write $\mathcal{I} = \langle u \rangle$. An ideal $\mathcal{I}$ always has a $\mathbb{Z}$-basis of cardinality n, which is not unique; if $\mathcal{I} = \langle u \rangle$ and B is any $\mathbb{Z}$-basis of R, then uB is a $\mathbb{Z}$-basis of $\mathcal{I}$. A *fractional ideal* $\mathcal{I} \subset K$ is a set such that $d\mathcal{I} \subseteq R$ is an integral ideal for some $d \in R$, and is principal if it equals uR for some $u \in K$. Any fractional ideal $\mathcal{I}$ embeds under σ as a lattice $\sigma(\mathcal{I})$ in H, which we call an *ideal lattice*. We identify $\mathcal{I}$ with this lattice and associate with $\mathcal{I}$ all the usual lattice quantities (determinant, minimum distance, etc.).

The norm of an ideal $\mathcal{I}$ is its index as an additive subgroup of R, i.e., $\mathrm{N}(\mathcal{I}) = |R/\mathcal{I}|$. This notion of norm generalizes the field norm, in that $\mathrm{N}(\langle a \rangle) = |\mathrm{N}(a)|$ for any $a \in R$, and $\mathrm{N}(\mathcal{I}\mathcal{J}) = \mathrm{N}(\mathcal{I})\,\mathrm{N}(\mathcal{J})$. The norm of a fractional ideal $\mathcal{I}$ is defined as $\mathrm{N}(\mathcal{I}) = \mathrm{N}(d\mathcal{I})/|\mathrm{N}(d)|$, where $d \in R$ is such that $d\mathcal{I} \subseteq R$. It follows that the determinant of an ideal lattice $\mathcal{I}$ is

$$\det(\sigma(\mathcal{I})) = \mathrm{N}(\mathcal{I}) \cdot \sqrt{\Delta_K}. \tag{2.9}$$

The following lemma gives upper and lower bounds on the minimum distance of an ideal lattice. The upper bound is an immediate consequence of Minkowski's first theorem; the lower bound follows from the arithmetic mean/geometric mean inequality, and the fact that $|\mathrm{N}(a)| \geq \mathrm{N}(\mathcal{I})$ for any nonzero $a \in \mathcal{I}$.

[6] Some texts define the discriminant as a signed quantity, but in this work we only care about its magnitude.

[7] Some texts also define the trivial set $\{0\}$ as an ideal, but in this work it is more convenient to exclude it.

Lemma 2.14. *For any fractional ideal $\mathcal{I}$ in a number field K of degree n,*

$$\sqrt{n} \cdot \mathrm{N}^{1/n}(\mathcal{I}) \quad \leq \quad \lambda_1(\mathcal{I}) \quad \leq \quad \sqrt{n} \cdot \mathrm{N}^{1/n}(\mathcal{I}) \cdot \sqrt{\Delta_K^{1/n}}.$$

The sum $\mathcal{I} + \mathcal{J}$ of two ideals is the set of all $a + b$ for $a \in \mathcal{I}$, $b \in \mathcal{J}$, and the product ideal $\mathcal{I}\mathcal{J}$ is the set of all finite sums of terms ab for $a \in \mathcal{I}$, $b \in \mathcal{J}$. Multiplication extends to fractional ideals in the obvious way, and the set of fractional ideals forms a group under multiplication; in particular, every fractional ideal $\mathcal{I}$ has a (multiplicative) inverse ideal, written $\mathcal{I}^{-1}$.

Two ideals $\mathcal{I}, \mathcal{J} \subseteq R$ are *coprime* if $\mathcal{I} + \mathcal{J} = R$. An ideal $\mathfrak{p} \subsetneq R$ is *prime* if whenever $ab \in \mathfrak{p}$ for some $a, b \in R$, then $a \in \mathfrak{p}$ or $b \in \mathfrak{p}$ (or both). An ideal $\mathfrak{p}$ is prime if and only if it is *maximal*, i.e., if the only proper superideal of $\mathfrak{p}$ is R itself, which implies that the quotient ring $R/\mathfrak{p}$ is a finite field. The ring R has unique factorization of ideals, i.e., every ideal $\mathcal{I}$ can be expressed uniquely as a product of powers of prime ideals.

2.5.4 Duality

Here we recall the notion of a dual ideal and explain its close connection to both the inverse ideal and the dual lattice. For more details, see [Con09] as an accessible reference.

For any fractional ideal $\mathcal{I}$ in K, its *dual* is defined as

$$\mathcal{I}^\vee = \{a \in K \ : \ \mathrm{Tr}(a\mathcal{I}) \subseteq \mathbb{Z}\}.$$

It is easy to verify that $(\mathcal{I}^\vee)^\vee = \mathcal{I}$, that $\mathcal{I}^\vee$ is a fractional ideal, and that $\mathcal{I}^\vee$ embeds under σ as the (conjugate) dual lattice of $\mathcal{I}$, as defined in Section 2.4.

For any $\mathbb{Q}$-basis $B = \{b_j\}$ of K, we denote its dual basis by $B^\vee = \{b_j^\vee\}$, which is characterized by $\mathrm{Tr}(b_i \cdot b_j^\vee) = \delta_{ij}$, the Kronecker delta. It is immediate that $(B^\vee)^\vee = B$, and if B is a $\mathbb{Z}$-basis of some fractional ideal $\mathcal{I}$, then $B^\vee$ is a $\mathbb{Z}$-basis of its dual ideal $\mathcal{I}^\vee$. An important fact is that if $a = \sum_j a_j \cdot b_j$ for $a_j \in \mathbb{R}$ is the unique representation of $a \in K_\mathbb{R}$ in basis B, then $a_j = \mathrm{Tr}(a \cdot b_j^\vee)$ by linearity of trace.

Suppose that $K \cong \bigotimes_\ell K_\ell$ as in Proposition 2.13. Then by linearity and the tensorial decomposition of the trace (Equation (2.7)), taking the dual commutes with tensoring, i.e., $(\bigotimes_\ell B_\ell)^\vee = \bigotimes_\ell B_\ell^\vee$ for any $\mathbb{Q}$-bases B_ℓ of K_ℓ. In particular, this implies that $(\bigotimes_\ell \mathcal{I}_\ell)^\vee = \bigotimes_\ell \mathcal{I}_\ell^\vee$ for any fractional ideals $\mathcal{I}_\ell$ in K_ℓ.

Except in the trivial number field $K = \mathbb{Q}$, the ring of integers R is not self-dual, nor are an ideal and its inverse dual to each other. However, an ideal and its inverse *are* related by multiplication with the dual ideal $R^\vee$ of the ring: for any fractional ideal $\mathcal{I}$, its dual is $\mathcal{I}^\vee = \mathcal{I}^{-1} \cdot R^\vee$. The factor $R^\vee$ is often called the *codifferent*, and its inverse $(R^\vee)^{-1}$ the *different*, which is in fact an ideal in R. By Equation (2.9) and the fact that $\det(\sigma(R)) = \det(\sigma(R^\vee))^{-1}$, we have

$$\mathrm{N}(R^\vee) = \Delta_K^{-1}. \tag{2.10}$$

The codifferent $R^\vee$ plays an important role in ring-LWE and its applications. The following material shows that $R^\vee$ is a principal ideal with a particularly simple generator, and that $(R^\vee)^{-1} \subseteq R$ is an integral ideal. We include proofs for completeness. We start with a useful lemma characterizing the traces of the powers of ζ_m.

Lemma 2.15. *Let m be a power of a prime p and $m' = m/p$, and j be an integer. Then*

$$\mathrm{Tr}(\zeta_m^j) = \begin{cases} \varphi(p) \cdot m' & \text{if } j = 0 \bmod m \\ -m' & \text{if } j = 0 \bmod m', j \neq 0 \bmod m \\ 0 & \text{otherwise.} \end{cases}$$

Proof. The first case is immediate, since $\zeta_m^j = 1$. Otherwise, let $d = \gcd(j, m)$ and $\tilde{m} = m/d$, so $\mathrm{Tr}(\zeta_m^j) = d \cdot \mathrm{Tr}_{\mathbb{Q}(\zeta_{\tilde{m}})/\mathbb{Q}}(\zeta_{\tilde{m}}^{j/d})$. Because j/d is coprime with $\tilde{m}$, the latter trace is the sum of all complex primitive $\tilde{m}$th roots of unity, which is -1 when $\tilde{m} = p$, and 0 otherwise. □

Lemma 2.16. *Let m be a power of a prime p and $m' = m/p$, and let $g = 1 - \zeta_p \in R = \mathbb{Z}[\zeta_m]$. Then $R^\vee = \langle g/m \rangle$; $p/g \in R$; and $\langle g \rangle$ and $\langle p' \rangle$ are coprime for every prime integer $p' \neq p$.*

Proof. To prove the first claim, we first show that $g/m \in R^\vee$. Since the power basis is a $\mathbb{Z}$-basis of R, it is necessary and sufficient to show that $\mathrm{Tr}(\zeta_m^j \cdot g/m) = \mathrm{Tr}(\zeta_m^j - \zeta_m^{j+m'})/m$ is an integer for every $j \in [\varphi(m)]$. By Lemma 2.15, it is $(\varphi(p) + 1)m'/m = 1$ for $j = 0$, and 0 for all other j. Now to show that $R^\vee = \langle g/m \rangle$, it suffices to show that $\mathrm{N}(g/m) = \mathrm{N}(R^\vee)$, the latter of which is $p^{m/p}/m^{\varphi(m)}$ by Equations (2.10) and (2.8). Now $\mathrm{N}(m) = m^{\varphi(m)}$, and $\mathrm{N}(1 - \zeta_p) = \mathrm{N}_{\mathbb{Q}(\zeta_p)/\mathbb{Q}}(1 - \zeta_p)^{m/p}$. Because the roots of $\Phi_p(X)$ are exactly the complex primitive pth roots of unity, the latter norm is exactly $\Phi_p(1) = p$, as desired.

To prove that $p/g \in R$, using $1 + \zeta_p + \zeta_p^2 + \cdots + \zeta_p^{p-1} = 0$ one may verify that

$$p = (1 - \zeta_p)\big((p-1) + (p-2)\zeta_p + \cdots + \zeta_p^{p-2}\big).$$

To prove the third claim, recall again that the norm of $\langle g \rangle$ is a power of p. Therefore, the norm of $\langle g \rangle + \langle p' \rangle$, being a divisor of both a power of p and of p', must be 1, implying that $\langle g \rangle$ and $\langle p' \rangle$ are coprime.□

Definition 2.17. *For $R = \mathbb{Z}[\zeta_m]$, define $g = \prod_p (1 - \zeta_p) \in R$, where p runs over all* odd *primes dividing m. Also define $t = \hat{m}/g \in R$, where $\hat{m} = m/2$ if m is even, otherwise $\hat{m} = m$.*

Notice that $\hat{m}/g \in R$ because $(1 - \zeta_2) = 2$, so $\hat{m}/g = m/\prod_p (1 - \zeta_p) \in R$, where here p runs over *all* primes dividing m.

Corollary 2.18. *Adopt the notation from Definition 2.17. Then $R^\vee = \langle g/\hat{m} \rangle = \langle t^{-1} \rangle$, and $\langle g \rangle$ is coprime with $\langle p' \rangle$ for every prime integer p' except those* odd *primes dividing m.*

Proof. Letting $m = \prod_\ell m_\ell$ be the prime-power factorization of m, where each m_ℓ is a power of some prime p_ℓ, and using the ring isomorphism $R \cong \bigotimes_\ell R_\ell$ where $R_\ell = \mathbb{Z}[\zeta_{m_\ell}]$, we can equivalently express g as $g = (\hat{m}/m)(\otimes_\ell g_\ell)$, where $g_\ell = (1 - \zeta_{p_\ell})$. Then by Lemma 2.16,

$$\Big(\bigotimes\nolimits_\ell R_\ell\Big)^\vee = \bigotimes\nolimits_\ell (R_\ell^\vee) = \bigotimes\nolimits_\ell (g_\ell/m_\ell) R_\ell = (g/\hat{m}) \cdot \Big(\bigotimes\nolimits_\ell R_\ell\Big),$$

as desired.

For the coprimality claim, the norm of g is a product of powers of the odd primes dividing m, and the claim follows by the same reasoning as in Lemma 2.16. □

2.5.5 Prime Splitting and Chinese Remainder Theorem

For an integer prime $p \in \mathbb{Z}$, the factorization of the principal ideal $\langle p \rangle \subset R = \mathbb{Z}[\zeta_m]$ is as follows. Let $d \geq 0$ be the largest integer such that p^d divides m, let $h = \varphi(p^d)$, and let $f \geq 1$ be the multiplicative order of p modulo m/p^d. Then $\langle p \rangle = \mathfrak{p}_1^h \cdots \mathfrak{p}_g^h$, where $g = n/(hf)$ and the $\mathfrak{p}_i$ are distinct prime ideals each of norm p^f.

A particular case of interest for us is the factorization of an integer prime $q = 1 \bmod m$, and the form of its prime ideal factors. Here the order of q modulo m is 1, and so $\langle q \rangle$ "splits completely" into n distinct prime ideals of norm q. Notice that the field $\mathbb{Z}_q$ has a primitive root of unity ω_m, because the multiplicative

group of $\mathbb{Z}_q$ is cyclic with order $q-1$. Indeed, there are $n = \varphi(m)$ distinct such roots of unity $\omega_m^i \in \mathbb{Z}_q$, for $i \in \mathbb{Z}_m^*$, and the prime ideal factors of $\langle q \rangle$ are simply $\mathfrak{q}_i = \langle q \rangle + \langle \zeta_m - \omega_m^i \rangle$. Therefore, each quotient ring $R/\mathfrak{q}_i$ is isomorphic to the field $\mathbb{Z}_q$, via the map $\zeta_m \mapsto \omega_m^i$.

The Chinese Remainder Theorem says that if $\mathfrak{p}_i$ are pairwise coprime ideals in R, then the natural ring homomorphism from $R/\prod_i \mathfrak{p}_i$ to the product ring $\prod_i (R/\mathfrak{p}_i)$ is in fact an isomorphism. To support efficient operations in $R_q = R/qR$, we will use the following special case, which we use to define a special $\mathbb{Z}_q$-basis of R_q (see Section 5 for details).

Lemma 2.19. *Let $q = 1 \bmod m$ be prime, and let $\omega_m \in \mathbb{Z}_q$ and ideals $\mathfrak{q}_i$ be as above. Then the natural ring homomorphism $R/\langle q \rangle \to \prod_{i \in \mathbb{Z}_m^*} (R/\mathfrak{q}_i) \cong (\mathbb{Z}_q)^n$ is an isomorphism.*

2.6 Ring-LWE

We now provide the formal definition of the ring-LWE problem and describe the worst-case hardness result shown in [LPR10]. We remark that our definition here differs very slightly from the one used in [LPR10]: we scale the b component by a factor of q, so that it is an element of $K_{\mathbb{R}}/qR^\vee$ and not $K_{\mathbb{R}}/R^\vee$ as in [LPR10]. This is done for convenience when later discretizing the b component, and the two definitions are easily seen to be equivalent.

Definition 2.20 (Ring-LWE Distribution). *For a "secret" $s \in R_q^\vee$ (or just $R^\vee$) and a distribution ψ over $K_{\mathbb{R}}$, a sample from the ring-*LWE *distribution $A_{s,\psi}$ over $R_q \times (K_{\mathbb{R}}/qR^\vee)$ is generated by choosing $a \leftarrow R_q$ uniformly at random, choosing $e \leftarrow \psi$, and outputting $(a, b = a \cdot s + e \bmod qR^\vee)$.*

Definition 2.21 (Ring-LWE, Average-Case Decision). *The* average-case decision *version of the ring-*LWE *problem, denoted R-$\mathsf{DLWE}_{q,\psi}$, is to distinguish with non-negligible advantage between independent samples from $A_{s,\psi}$, where $s \leftarrow R_q^\vee$ is uniformly random, and the same number of uniformly random and independent samples from $R_q \times (K_{\mathbb{R}}/qR^\vee)$.*

Theorem 2.22. *Let K be the mth cyclotomic number field having dimension $n = \varphi(m)$ and $R = \mathcal{O}_K$ be its ring of integers. Let $\alpha = \alpha(n) > 0$, and let $q = q(n) \geq 2$, $q = 1 \bmod m$ be a $\mathrm{poly}(n)$-bounded prime such that $\alpha q \geq \omega(\sqrt{\log n})$. Then there is a polynomial-time quantum reduction from $\tilde{O}(\sqrt{n}/\alpha)$-approximate* SIVP *(or* SVP*) on ideal lattices in K to the problem of solving R-$\mathsf{DLWE}_{q,\psi}$ given only ℓ samples, where ψ is the Gaussian distribution $D_{\xi q}$ for $\xi = \alpha \cdot (n\ell/\log(n\ell))^{1/4}$.*

Note that the above worst-case hardness result deteriorates with the number of samples ℓ. Since most applications only require a small (or even a constant) number of samples, this is not a serious issue. In cases where a large number of samples is needed, one can use two alternative hardness theorems proven in [LPR10]. The first assumes hardness of the search problem for spherical Gaussian error, which as yet lacks a reduction from a worst-case problem. The second is a reduction from a worst-case problem, and it allows an arbitrary number of samples without any deterioration in the approximation factor; it does, however, require the error distribution to be non-spherical and chosen in a specific way, which makes it somewhat less convenient in implementations. We refer to [LPR10] for additional information.

In applications it is often useful to work with a version of ring-LWE whose error distribution is discrete. This leads naturally to a definition of $A_{s,\chi}$ for a discrete error distribution χ over $R^\vee$, with b being an element of $R_q^\vee$. We similarly modify Definition 2.21 by letting R-$\mathsf{DLWE}_{q,\chi}$ be the problem of distinguishing between $A_{s,\chi}$ and uniform samples from $R_q \times R_q^\vee$. As we show next, for a wide family of discrete error distributions, the hardness of the discrete version follows from that of the continuous one. In more detail, the lemma

below implies that if R-$\mathsf{DLWE}_{q,\psi}$ is hard with some number ℓ of samples, then so is R-$\mathsf{DLWE}_{q,\chi}$ with the same number of samples, where the error distribution χ is $\lfloor p \cdot \psi \rceil_{w+pR^\vee}$ for some integer p coprime to q, $\lfloor \cdot \rceil$ is any valid discretization to (cosets of) $pR^\vee$, and w is an arbitrary element in $R_p^\vee$ that can vary from sample to sample (even adaptively and adversarially). In particular, for $p = 1$ we get hardness with error distribution $\lfloor \psi \rceil_{R^\vee}$.

Lemma 2.23. *Let p and q be positive coprime integers, and $\lfloor \cdot \rceil$ be a valid discretization to (cosets of) $pR^\vee$. There exists an efficient transformation that on input $w \in R_p^\vee$ and a pair in $(a', b') \in R_q \times K_\mathbb{R}/qR^\vee$, outputs a pair $(a = pa' \bmod qR, b) \in R_q \times R_q^\vee$ with the following guarantees: if the input pair is uniformly distributed then so is the output pair; and if the input pair is distributed according to the ring-LWE distribution $A_{s,\psi}$ for some (unknown) $s \in R^\vee$ and distribution ψ over $K_\mathbb{R}$, then the output pair is distributed according to $A_{s,\chi}$, where $\chi = \lfloor p \cdot \psi \rceil_{w+pR^\vee}$.*

Proof. Given w and a sample $(a', b') \in R_q \times K_\mathbb{R}/qR^\vee$, the transformation discretizes $pb' \in K_\mathbb{R}/pqR^\vee$ to $\lfloor pb' \rceil_{w+pR^\vee} \in (w + pR^\vee) + pqR^\vee$. It then lets $a = pa' \bmod qR$ and $b = \lfloor pb' \rceil_{w+pR^\vee} \bmod qR^\vee$, and outputs the sample $(a, b) \in R_q \times R_q^\vee$.

If the distribution of (a', b') is $A_{s,\psi}$, then $pb' = (pa') \cdot s + pe' \bmod pqR^\vee$ for $e' \leftarrow \psi$. Because $(pa') \cdot s \in pR^\vee/pqR^\vee$, by validity of the discretization we have that $\lfloor pb' \rceil_{w+pR^\vee}$ and $(pa') \cdot s + \lfloor pe' \rceil_{w+pR^\vee}$ are identically distributed. Because p and q are coprime, $a = pa' \bmod qR$ is uniformly random over R_q, so (a, b) has distribution $A_{s,\chi}$.

On the other hand, if (a', b') is uniformly random, then a is uniform over R_q. Moreover, since the uniform distribution over $K_\mathbb{R}/pqR^\vee$ is invariant under shifts by $pR^\vee$, then by validity so is the distribution of $b = \lfloor pb' \rceil_{w+pR^\vee} \bmod qR^\vee$, for any $w \in R^\vee$. Then because p and q are coprime, b is uniformly random over $R_q^\vee$ and independent of a, as desired. □

Finally, another important variant of ring-LWE, known as the "normal form," is the one in which the secret, instead of being uniformly distributed, is chosen from the error distribution (discretized to $R^\vee$, or a coset of $pR^\vee$ as in Lemma 2.23 above). This modification makes the secret short, which is very useful in some applications. We now show that this variant of ring-LWE is as hard as the original one, closely following the technique of [ACPS09].

Lemma 2.24. *Let p and q be positive coprime integers, $\lfloor \cdot \rceil$ be a valid discretization to (cosets of) $pR^\vee$, and w be an arbitrary element in $R_p^\vee$. If R-$\mathsf{DLWE}_{q,\psi}$ is hard given some number ℓ of samples, then so is the variant of R-$\mathsf{DLWE}_{q,\psi}$ in which the secret is sampled from $\chi := \lfloor p \cdot \psi \rceil_{w+pR^\vee}$, given $\ell - 1$ samples.*

Proof. We show how to solve the former problem given an oracle for the latter. Start by drawing one sample from the unknown distribution and apply the transformation from Lemma 2.23 (with p, w, and $\lfloor \cdot \rceil$) to it. Let $(a_0, b_0) \in R_q \times R_q^\vee$ be the result. If a_0 is not in R_q^*, abort and reject. Otherwise, let $a_0^{-1} \in R_q^*$ denote its inverse. Draw $\ell - 1$ additional samples $(a_i, b_i) \in R_q \times K_\mathbb{R}/qR^\vee$ $(i = 1, \ldots, \ell - 1)$ from the unknown distribution, and return the oracle's output when applied to the pairs

$$(a_i' = -a_0^{-1} a_i \ , \ b_i' = b_i + a_i' b_0) \in R_q \times K_\mathbb{R}/qR^\vee.$$

To prove this gives a valid distinguisher, notice first that by Claim 2.25 below, it suffices to show a noticeable distinguishing gap conditioned on a_0 being invertible. Next, observe that if the input distribution is uniform, then so is the distribution of the pairs (a_i', b_i'). Finally, if the input distribution is $A_{s,\psi}$ for

some $s \in R^\vee$, then we have $b_0 = a_0 \cdot s + e_0$ where e_0 is distributed according to χ. Therefore, for each $i = 1, \ldots, \ell - 1$,

$$b'_i = (a_i \cdot s + e_i) - a_0^{-1} a_i (a_0 \cdot s + e_0) = e_i + a'_i e_0,$$

where the e_i are distributed according to ψ, and so the input to the oracle consists of independent samples from $A_{e_0,\psi}$, as required. □

Claim 2.25. *Consider the mth cyclotomic field of degree $n = \varphi(m)$ for some $m \geq 2$. Then for any $q \geq 2$, the fraction of invertible elements in R_q is at least $1/\operatorname{poly}(n, \log q)$.*

When $q = 1 \bmod m$ is a prime (as in Theorem 2.22), we have by Lemma 2.19 that the fraction of invertible elements in R_q is $(1 - 1/q)^n \geq (1 - 1/(n+1))^n \geq e^{-1}$. This uses the inequality $1 - 1/(\alpha + 1) \geq e^{-1/\alpha}$ for $\alpha > 0$, which we will use again in the proof below.

Proof. We first observe that for any integer $r \geq 1$ and prime ideal $\mathfrak{p}$, an element $a \in R$ is invertible modulo $\mathfrak{p}^r$ if and only if $a \neq 0 \bmod \mathfrak{p}$, and therefore the fraction of uninvertible elements in $R/\mathfrak{p}^r$ is $1/\operatorname{N}(\mathfrak{p})$. One direction is obvious: if $a = 0 \bmod \mathfrak{p}$, then so is $a \cdot b$ for any $b \in R$, so a is uninvertible (because $1 \notin \mathfrak{p}$). For the other direction, if $a \neq 0 \bmod \mathfrak{p}$, then $\mathfrak{p} \nmid \langle a \rangle$, and so $\langle a \rangle, \mathfrak{p}^r$ are coprime, i.e., $\langle a \rangle + \mathfrak{p}^r = R$. Therefore, there exists $b \in R$ such that $ab \in 1 + \mathfrak{p}^r$.

Using the factorization of the ideals $\langle p \rangle$ given in Section 2.5.5 and the Chinese remainder theorem, we get that the fraction of invertible elements in R_q is

$$\prod_{\text{prime } p \mid q} (1 - p^{-f_p})^{n/(f_p \varphi(p^{d_p}))} \geq \prod_{\text{prime } p \mid q} (1 - p^{-f_p})^{n/\varphi(p^{d_p})}, \tag{2.11}$$

where d_p is the largest integer such that p^{d_p} divides m and f_p is the multiplicative order of p modulo m/p^{d_p}. For any prime p we clearly have $p^{f_p} > m \geq m/p^{d_p}$, and therefore

$$\begin{aligned}(1 - p^{-f_p})^{n/\varphi(p^{d_p})} &= (1 - p^{-f_p})^{\varphi(m/p^{d_p})} \\ &\geq (1 - p^{-f_p})^{m/p^{d_p}} \geq e^{-1}.\end{aligned}$$

As a result, the product in (2.11), restricted to primes p dividing m, of which there are at most $\log_2 m$, is at least $1/\operatorname{poly}(m)$. It therefore suffices to bound from below the product in (2.11) restricted to primes p not dividing m. For such primes p we have $d_p = 0$, and the expression simplifies to

$$\prod_{p \mid q, p \nmid m} (1 - p^{-f_p})^n, \tag{2.12}$$

where f_p is the multiplicative order of p modulo m. Notice that the values p^{f_p} are distinct for distinct p. Moreover, they are all 1 modulo m. Therefore, since the product in (2.12) includes at most $\log_2 q$ terms, we can bound it from below by

$$\prod_{k=1}^{\log_2 q} \left(1 - \frac{1}{km+1}\right)^n \geq \prod_{k=1}^{\log_2 q} e^{-n/km} \geq \prod_{k=1}^{\log_2 q} e^{-1/k} \geq e^{-1} \prod_{k=2}^{\log_2 q} \left(1 - \frac{1}{k}\right) = (e \cdot \log_2 q)^{-1}.$$

□

3 Sparse Decompositions of DFT and CRT

Here we give structured (or "sparse") decompositions of two important linear transformations, which lead to fast algorithms for applying them. We follow the algebraic framework of [PM08].

Definition 3.1. *Let m be a prime power and let $\mathcal{R}$ denote any commutative ring containing some element ω_m of multiplicative order m, i.e., a primitive mth root of unity.*

- *The* discrete Fourier transform DFT_m *over $\mathcal{R}$ is the $\mathbb{Z}_m$-by-$\mathbb{Z}_m$ matrix whose (i,j)th entry is ω_m^{ij}.*
- *The* Chinese remainder transform CRT_m *over $\mathcal{R}$ is the (square) submatrix of DFT_m obtained by restricting to the rows indexed by $\mathbb{Z}_m^*$ and the columns indexed by $[\varphi(m)]$.*

For an arbitrary positive integer m having prime-power factorization $m = \prod_\ell m_\ell$, where $\mathcal{R}$ has an mth root of unity (and hence has primitive m_ℓth roots of unity for each m_ℓ), the DFT and CRT matrices are

$$\mathrm{DFT}_m = \bigotimes_\ell \mathrm{DFT}_{m_\ell} \quad \textit{and} \quad \mathrm{CRT}_m = \bigotimes_\ell \mathrm{CRT}_{m_\ell}.$$

We identify the matrices DFT_m and CRT_m with the linear transforms they represent.

For a prime power m, applying DFT_m corresponds with evaluating a polynomial in $\mathcal{R}[X]$ of degree less than m (represented by its vector of coefficients in the natural order) at all the mth roots of unity $\omega_m^i \in \mathcal{R}$ for $i \in [m]$. Similarly, CRT_m corresponds with evaluating a polynomial of degree less than $\varphi(m)$ at all the *primitive* mth roots of unity ω_m^i for $i \in \mathbb{Z}_m^*$. (This interpretation, and its connection with Lemma 2.19, explains our choice of the name "Chinese remainder transform.")

For m with prime-power factorization $m = \prod_\ell m_\ell$, it can be shown using the Good-Thomas decomposition that DFT_m again corresponds with polynomial evaluation at all mth roots of unity, but under some permutations of the input and output vectors. For CRT_m, the correspondence with polynomial evaluation is different, because the columns of CRT_m typically do not correspond to powers $0, \ldots, \varphi(m)-1$ of a primitive mth root of unity ω_m. Instead, CRT_m corresponds with evaluation of a multivariate polynomial (with one variable per factor m_ℓ) at all input tuples in which the ℓth element is a primitive m_ℓth root of unity. We adopt the tensorial form of CRT_m because it corresponds directly with the tensorial (or multivariate) decomposition of the mth cyclotomic number field, and admits a finer-grained decomposition and more efficient algorithms than the univariate perspective.

Decomposition of DFT_m. Let m be a power of some prime p, and let $m' = m/p$. Using the Cooley-Tukey decomposition we can express DFT_m in terms of smaller DFTs of dimensions p and m', and by iterating, in terms of DFT_p alone. Reindex the columns of DFT_m by pairs $(j_0, j_1) \in [p] \times [m']$, using the standard correspondence $j = m' j_0 + j_1 \in [m]$. Similarly, reindex the rows by pairs $(i_0, i_1) \in [p] \times [m']$, this time using the (nonstandard) correspondence $i = p i_1 + i_0 \in [m]$.[8] We then have the decomposition

$$\mathrm{DFT}_m = (I_{[p]} \otimes \mathrm{DFT}_{m'}) \cdot T_m \cdot (\mathrm{DFT}_p \otimes I_{[m']}), \tag{3.1}$$

where all three terms are $([p] \times [m'])$-by-$([p] \times [m'])$ matrices, and T_m is the diagonal "twiddle" matrix having entry $\omega_m^{i_0 i_1}$ in its (i_0, i_1)th diagonal entry. Therefore, applying DFT_m reduces to m' parallel applications of DFT_p, followed by m parallel scalar multiplications by twiddle factors, followed by p parallel applications

[8] This relabeling corresponds with the "bit-reversal" or related "stride" output permutation in the standard decimation-in-frequency FFT algorithm. In an implementation, the permutation can be omitted because the output does not need to be in any particular order.

of $\mathrm{DFT}_{m'}$. Of course, each $\mathrm{DFT}_{m'}$ can be further decomposed in the same way, down to the DFT_p base case. Using any of the Rader, Winograd, or Bluestein FFT algorithms, we can apply such base cases in $O(p \log p)$ time, which implies that DFT_m can be applied in $O(n \log n)$ time, where $n = \varphi(m)$.

To verify Equation (3.1), it suffices by linearity to compare the action of both sides on the standard basis vectors. Take any $(j_0, j_1) \in [p] \times [m']$ and consider the vector with 1 in location (j_0, j_1) and zero elsewhere. Applying $\mathrm{DFT}_p \otimes I_{[m']}$ to it yields the vector that is $\omega_p^{i_0 j_0}$ in locations (i_0, j_1) for $i_0 \in [p]$ and zero elsewhere. The matrix T_m changes these nonzero entries to $\omega_p^{i_0 j_0} \omega_m^{i_0 j_1}$, and finally, $I_{[p]} \otimes \mathrm{DFT}_{m'}$ yields the vector with

$$\omega_p^{i_0 j_0} \cdot \omega_m^{i_0 j_1} \cdot \omega_{m'}^{i_1 j_1} = \omega_m^{m' i_0 j_0 + i_0 j_1 + p i_1 j_1} = \omega_m^{(p i_1 + i_0)(m' j_0 + j_1)}$$

in any location $(i_0, i_1) \in [p] \times [m']$, as required.

Decomposition of CRT_m. Letting m, p, and m' be as above, notice that $\varphi(m) = \varphi(p) \cdot m'$. Moreover, with the above reindexing of rows and columns, CRT_m is the submatrix of DFT_m restricted to rows $\mathbb{Z}_p^* \times [m']$ and columns $[\varphi(p)] \times [m']$. By appropriately restricting the matrices in Equation (3.1), we obtain the decomposition (which can be verified in the same way as above)

$$\mathrm{CRT}_m = (I_{\mathbb{Z}_p^*} \otimes \mathrm{DFT}_{m'}) \cdot \hat{T}_m \cdot (\mathrm{CRT}_p \otimes I_{[m']}), \tag{3.2}$$

where $\hat{T}_m$ is the diagonal twiddle matrix T_m from above, restricted to the rows and columns indexed by $\mathbb{Z}_p^* \times [m']$. Applying CRT_m therefore reduces to m' parallel applications of CRT_p, followed by $\varphi(m)$ parallel scalar multiplications by twiddle factors, followed by $\varphi(p)$ parallel applications of $\mathrm{DFT}_{m'}$.

Inversion. Using the inversion rules for matrix multiplication and the Kronecker product, the inverse DFT and CRT decompose as

$$\mathrm{DFT}_m^{-1} = (\mathrm{DFT}_p^{-1} \otimes I_{[m']}) \cdot T_m^{-1} \cdot (I_{[p]} \otimes \mathrm{DFT}_{m'}^{-1}) \tag{3.3}$$

$$\mathrm{CRT}_m^{-1} = (\mathrm{CRT}_p^{-1} \otimes I_{[m']}) \cdot (\hat{T}_m)^{-1} \cdot (I_{\mathbb{Z}_p^*} \otimes \mathrm{DFT}_{m'}^{-1}), \tag{3.4}$$

and can be applied at exactly the same cost as their forward counterparts. Note that the row and column index sets of CRT_m are different (as they are for CRT_p and $\hat{T}_m$ as well), so $\mathrm{CRT}_m^{-1} \cdot \mathrm{CRT}_m$ and $\mathrm{CRT}_m \cdot \mathrm{CRT}_m^{-1}$ are "different" matrices, although they are both still identity matrices over the appropriate index sets.

Arbitrary m. For m that may have more than one prime divisor, the tensorial form of CRT_m leads immediately to a fast algorithm. Specifically, if m has prime-power factorization $m = \prod_\ell m_\ell$, then by the mixed-product property, applying $\mathrm{CRT}_m = \bigotimes_\ell \mathrm{CRT}_{m_\ell}$ reduces to $\varphi(m/m_\ell)$ parallel applications of CRT_{m_ℓ}, in sequence for each ℓ (see the end of Section 2.1). Since each CRT_{m_ℓ} can be applied in $O(m_\ell \log m_\ell)$ time and $O(\log m_\ell)$ parallel depth, the total runtime and parallel depth are $O(m \log m)$ and $O(\log m)$, respectively.

4 The Powerful Basis

In this section (and Section 6) we study certain $\mathbb{Z}$-bases of certain fractional ideals $\mathcal{I}$ in $K = \mathbb{Q}(\zeta_m)$, which are therefore $\mathbb{Z}_q$-bases of the quotients $\mathcal{I}_q = \mathcal{I}/q\mathcal{I}$ for any positive integer q. Fixing such a basis $\vec{b}$ and viewing it as a (column) vector over $\mathcal{I}$, we can represent any $a \in \mathcal{I}$ (respectively, $a \in \mathcal{I}_q$) uniquely as

$a = \langle \vec{b}, \mathbf{a} \rangle = \vec{b}^T \cdot \mathbf{a}$ for some coefficient vector $\mathbf{a}$ over $\mathbb{Z}$ (respectively, $\mathbb{Z}_q$) having the same index set as $\vec{b}$. Our algorithms simply store and operate on these coefficient vectors, while also keeping track of the corresponding basis, which will be one of the few we define below. Notice that by linearity, if we have some $a \in \mathcal{I}$ represented by coefficient vector $\mathbf{a}$ in basis $\vec{b}$, then $\mathbf{a}$ is also the representation of $ra \in r\mathcal{I}$ in the basis $r\vec{b}$ for any $r \in K$, so we can switch between the two values at essentially no cost.

Here we define a certain useful $\mathbb{Z}$-basis of R, and hence $\mathbb{Q}$-basis of K. We call it the "powerful" basis, due to its decomposition in terms of the power bases of prime-power cyclotomics, and the fast algorithms associated with it.[9]

Definition 4.1. *The* powerful *basis $\vec{p}$ of $K = \mathbb{Q}(\zeta_m)$ and $R = \mathbb{Z}[\zeta_m]$ is defined as follows:*

- *For a prime power m, define $\vec{p}$ to be the power basis $(\zeta_m^j)_{j \in [\varphi(m)]}$, treated as a vector over $R \subset K$.*
- *For m having prime-power factorization $m = \prod_\ell m_\ell$, define $\vec{p} = \bigotimes_\ell \vec{p_\ell}$, the tensor product of the power(ful) bases $\vec{p_\ell}$ of each $K_\ell = \mathbb{Q}(\zeta_{m_\ell})$.*

For any power $\mathcal{I} = (R^\vee)^k$ of $R^\vee = \langle t^{-1} \rangle$, define the powerful basis of $\mathcal{I}$ to be $t^{-k} \cdot \vec{p}$.

By definition of the tensor product, $\vec{p}$ is a vector with index set $\prod_\ell [\varphi(m_\ell)]$. So to specify an entry of $\vec{p}$ we need one index $j_\ell \in [\varphi(m_\ell)]$ per prime divisor of m, and the specified entry is $p_{(j_\ell)} = \prod_\ell \zeta_{m_\ell}^{j_\ell}$. Note that because $\zeta_{m_\ell} = \zeta_m^{m/m_\ell} \in K$, it is possible to "flatten" the index set to a size-$\varphi(m)$ subset of $[m]$, where index tuple (j_ℓ) maps to $j = \sum_\ell (m/m_\ell) j_\ell \bmod m$, and $p_j = \zeta_m^j$. We note that unless m is a prime power, the flattened index set is *not* equal to $[\varphi(m)]$, so the powerful basis differs from the power basis, although it still consists of powers of ζ_m. For instance, for $m = 15$ and $\zeta = \zeta_{15}$, the powerful basis consists of $\zeta^0, \zeta^3, \zeta^5, \zeta^6, \zeta^8, \zeta^9, \zeta^{11}$, and ζ^{14}. Because the flattened indices tend to be a somewhat irregular subset of $[m]$, it is usually preferable to maintain the structured index set.

Observe that $\vec{p}^T$ is a row vector (over K) with columns indexed by $\prod_\ell [\varphi(m_\ell)]$. Applying the canonical embedding σ entry-wise to obtain column vectors indexed by $\mathbb{Z}_m^*$ (or equivalently, $\prod_\ell \mathbb{Z}_{m_\ell}^*$), by Equation (2.6) we obtain the complex matrix $\sigma(\vec{p}^T) = \mathrm{CRT}_m$. With this fact in mind, we now prove two basic facts about the geometry of the powerful basis. The first says that all its elements are short (and in fact, by Lemma 2.14 they are shortest nonzero elements of R), and the second statement says essentially that the elements are close to orthogonal.

Claim 4.2. *The length of each element p_j of $\vec{p}$ in ℓ_∞ norm is $\|p_j\|_\infty = 1$, and in ℓ_2 norm is $\|p_j\|_2 = \sqrt{\varphi(m)} = \sqrt{n}$.*

Proof. Each entry in the CRT_m matrix is a root of unity, hence it has magnitude 1, and so the ℓ_∞ and ℓ_2 norms of each column are 1 and $\sqrt{\varphi(m)}$, respectively. □

Lemma 4.3. *The largest singular value of $\sigma(\vec{p}^T)$ (or equivalently, of CRT_m) is $s_1(\vec{p}) = \sqrt{\hat{m}}$, and the smallest singular value is $s_n(\vec{p}) = \sqrt{m/\operatorname{rad}(m)}$.*

Notice that the ratio of $s_1(\vec{p})$ to $\sqrt{\varphi(m)}$ (i.e., the ℓ_2 norm of each basis element) is just $\sqrt{\hat{m}/\varphi(m)} = (\prod_p p/(p-1))^{1/2} = O(\sqrt{\log \log m})$, where the product runs over all odd primes dividing m.

[9] Although we define the powerful basis in a different way, it can be seen that it coincides with what Bosma [Bos90] calls the "canonical" basis of R. Bosma's work is the only one we know of that explicitly considers this basis.

Proof. It suffices to prove the statement when m is a prime power, due to the tensor structure of CRT_m, and the fact that the vector of singular values of $A \otimes B$ is the tensor product of the two vectors of singular values of A and B. So let m be a power of a prime p, and let $m' = m/p$. By Equation (3.2),

$$\mathrm{CRT}_m = (\sqrt{m'}Q) \cdot (\mathrm{CRT}_p \otimes I_{[m']})$$

for some unitary matrix Q, because $\mathrm{DFT}_{m'}/\sqrt{m'}$ is unitary for any m', and so is the twiddle matrix $\hat{T}_m$. The lemma then follows immediately from the fact that the $\varphi(p) = p - 1$ eigenvalues of the Gram matrix

$$\mathrm{CRT}_p^* \cdot \mathrm{CRT}_p = (pI_{[\varphi(p)]} - \mathbf{1} \cdot \mathbf{1}^T) \tag{4.1}$$

are $p, \ldots, p$ ($p-2$ times) and 1, where the asterisk denotes the conjugate transpose, $\mathbf{1} \in \mathbb{R}^{[\varphi(p)]}$ is the all-ones vector, and the equality is by the fact that CRT_p is obtained by removing the all-1s row and one column from DFT_p, which is a unitary matrix scaled up by a $\sqrt{p}$ factor. □

We conclude this section by characterizing the Gram-Schmidt orthogonalization $\widetilde{\mathrm{CRT}_m}$ of the powerful basis (under the canonical embedding), in Lemma 4.4 below. This orthogonalization is used in the nearest-plane [Bab85] and Klein/GPV [GPV08] algorithms (see Lemma 2.9), which we use for sampling from discrete Gaussians over R. The lemma implies that the orthogonalization is structured so that these algorithms can be executed in substantially less time, and using much less precision, than is required for an arbitrary basis. This is because the U matrix associated with the orthogonalization is block diagonal with $m/\operatorname{rad}(m)$ identical square blocks of dimension $\operatorname{rad}(m)$, which allows an implementation to make $m/\operatorname{rad}(m)$ parallel and independent calls to a quadratic-time subroutine on dimension $\operatorname{rad}(m)$, for $O(m \operatorname{rad}(m))$ scalar operations in total. Moreover, each row of U has a small (common) denominator, allowing an implementation to compute inner products with the rows of U using low-precision integers (see the discussion following Lemma 2.9).

Recalling from Section 2.2 the matrix form of the Gram-Schmidt orthogonalization, it follows by the mixed-product property that $\widetilde{A \otimes B} = \widetilde{A} \otimes \widetilde{B}$. By the tensor structure of CRT_m, it therefore suffices to consider the case where m is a prime power.

Lemma 4.4. *Let m be a power of a prime p and $m' = m/p$. Then*

$$\mathrm{CRT}_m = Q_m \cdot (\sqrt{m'}D_p \otimes I_{[m']}) \cdot (U_p \otimes I_{[m']}),$$

where Q_m is unitary, D_p is the real diagonal $[\varphi(p)]$-by-$[\varphi(p)]$ matrix with $\sqrt{(p-1) - j/(p-j)}$ in its jth diagonal entry, and U_p is the upper unitriangular $[\varphi(p)]$-by-$[\varphi(p)]$ matrix with $-1/(p-i-1)$ in its (i,j)th entry, for $0 \le i < j < \varphi(p)$.

Proof. By Equation (3.2) and the fact that $\hat{T}_m$ and $\mathrm{DFT}_{m'}/\sqrt{m'}$ are unitary matrices, we have

$$\mathrm{CRT}_m = \sqrt{m'}Q' \cdot (\mathrm{CRT}_p \otimes I_{[m']})$$

for some unitary Q'. Thus it suffices to show that $\mathrm{CRT}_p = Q_p \cdot D_p \cdot U_p$ for some unitary Q_p.

Let $G = \mathrm{CRT}_p^* \cdot \mathrm{CRT}_p$ be the Gram matrix of CRT_p and recall from Equation (4.1) that G has diagonal entries $p-1$, and -1 entries elsewhere. As discussed in Section 2.2, by the uniqueness of the Cholesky decomposition it suffices to show that

$$G = U_p^T \cdot D_p^2 \cdot U_p.$$

This equality can be verified by an elementary calculation, as follows. For $k \geq 1$, define

$$T(k) := \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \cdots + \frac{1}{(k-1) \cdot k}.$$

It is easy to see (by induction, or by noticing that adding $1/k$ to the above collapses the sum) that $T(k) = 1 - 1/k$. For any $i \in [\varphi(p)]$, the ith diagonal entry in $U_p^T \cdot D_p^2 \cdot U_p$ is

$$p - 1 - \frac{i}{p-i} + \sum_{k=0}^{i-1} \frac{1}{(p-k-1)^2}\Big(p - 1 - \frac{k}{p-k}\Big).$$

The summation in the above expression is

$$p \sum_{k=0}^{i-1} \frac{1}{(p-k)(p-k-1)} = p\big(T(p) - T(p-i)\big) = p\Big(1 - \frac{1}{p} - 1 + \frac{1}{p-i}\Big) = \frac{i}{p-i},$$

and so the ith diagonal entry is $p - 1$, as required. The off-diagonal entries are calculated in essentially the same way. □

5 The Chinese Remainder Basis and Fast Ring Operations

When working in K or R, we can perform ring operations efficiently by representing elements under the canonical embedding σ. Recall that σ is the ring embedding from $K = \mathbb{Q}(\zeta_m)$ into the product ring $H \subset \mathbb{C}^{\mathbb{Z}_m^*}$ that maps ζ_m to each power $\omega_m^i \in \mathbb{C}$ for $i \in \mathbb{Z}_m^*$, where ω_m is a primitive complex mth root of unity. Under the canonical embedding, addition and multiplication simply apply coordinate-wise on each complex coordinate. Converting to the embedding representation from the powerful basis $\vec{p}$ is done simply by multiplying (with sufficient precision) by the complex matrix $\mathrm{CRT}_m = \sigma(\vec{p}^T)$, i.e., if $a = \langle \vec{p}, \mathbf{a} \rangle \in K$ for some rational vector $\mathbf{a}$ then $\sigma(a) = \mathrm{CRT}_m \cdot \mathbf{a}$.

In ring-LWE and its applications, we often work in R_q and $R_q^\vee$, and sometimes in $\mathcal{I}_q$ for $\mathcal{I} = (R^\vee)^k$, where q is a prime integer congruent to 1 modulo m.[10] While using the canonical embedding as above lets us perform ring operations relatively efficiently in these quotients (by using an arbitrary set of representatives), here we describe more efficient and practical algorithms that only use arithmetic in $\mathbb{Z}_q$, rather than on high-precision complex numbers. These algorithms are facilitated by what we call the *Chinese remainder* (CRT) basis for $\mathcal{I}_q$, defined next.

Recalling that $R \cong \bigotimes_\ell R_\ell$ where $m = \prod_\ell m_\ell$ is the prime-power factorization of m and R_ℓ is the m_ℓth cyclotomic ring, it is easy to verify that the quotient ring $R_q \cong \bigotimes_\ell (R_\ell / qR_\ell)$. Therefore we may focus on the case of prime-power m. Also recall from Section 2.5.5 the prime ideal factorization $\langle q \rangle = \prod_{i \in \mathbb{Z}_m^*} \mathfrak{q}_i$ in R, where $\mathfrak{q}_i = \langle q \rangle + \langle \zeta_m - \omega_m^i \rangle$ is prime in R and ω_m is some fixed element of order m in $\mathbb{Z}_q$.

Definition 5.1. *For a positive integer m, the Chinese remainder (or CRT) $\mathbb{Z}_q$-basis $\vec{c}$ of R_q is as follows:*

- *For a prime power m, $\vec{c} = (c_i)_{i \in \mathbb{Z}_m^*}$ is characterized by $c_i = 1 \bmod \mathfrak{q}_i$ and $c_i = 0 \bmod \mathfrak{q}_j$ for $i \neq j$. (Its existence is guaranteed by Lemma 2.19, the Chinese remainder theorem.)*

[10] The modulus q may also be a product of several primes $q_i = 1 \bmod m$, in which case we can use the Chinese Remainder Theorem to decompose R_q into the product of rings R_{q_i}.

- *For m having prime-power factorization $m = \prod_\ell m_\ell$, define $\vec{c} = \bigotimes_i \vec{c_\ell}$, the tensor product of the CRT bases $\vec{c_\ell}$ of each R_ℓ/qR_ℓ.*

For any power $\mathcal{I} = (R^\vee)^k$ of $R^\vee = \langle t^{-1} \rangle$, the CRT $\mathbb{Z}_q$-basis of $\mathcal{I}_q$ is $t^{-k} \cdot \vec{c}$.

Note that $\vec{c}$ is a vector over R_q having as its index set the Cartesian product $\prod_\ell \mathbb{Z}^*_{m_\ell}$, which may be flattened to the set $\mathbb{Z}^*_m$ using the bijective correspondence $(j_\ell) \leftrightarrow j = \sum_\ell (m/m_\ell) \cdot j_\ell \in \mathbb{Z}^*_m$. But for our purposes it is usually more convenient to retain the structured index set.

Working in the CRT basis yields very fast arithmetic operations. Suppose that m is a prime power. Since $c_i^2 = c_i \in R_q$ and $c_i \cdot c_{i'} = 0 \in R_q$ for distinct $i, i' \in \mathbb{Z}^*_m$, the CRT basis has the property that if $a, b \in R_q$ have coefficient vectors $\mathbf{a}, \mathbf{b}$ (respectively) over $\mathbb{Z}_q$ in the CRT basis—i.e., $a = \langle \vec{c}, \mathbf{a} \rangle$ and $b = \langle \vec{c}, \mathbf{b} \rangle$—then the coefficient vector of $a \cdot b \in R_q$ is the componentwise product $\mathbf{a} \odot \mathbf{b}$ over $\mathbb{Z}_q$. (Addition is componentwise as well, by linearity.) Moreover, this extends immediately to powers of $R^\vee$: if $\mathbf{a}, \mathbf{b}$ are the respective coefficient vectors of $a \in (R^\vee)^{k_1}_q$, $b \in (R^\vee)^{k_2}_q$ in the respective CRT bases $t^{-k_1} \cdot \vec{c}$ and $t^{-k_2} \cdot \vec{c}$, then $\mathbf{a} \odot \mathbf{b}$ is the coefficient vector of $a \cdot b \in (R^\vee)^k_q$ in the CRT basis $t^{-k} \cdot \vec{c}$, where $k = k_1 + k_2$.

Still treating m as a prime power, using the field isomorphisms $R/\mathfrak{q}_i \cong \mathbb{Z}_q$ given by $\zeta_m \mapsto \omega_m^i$, we see that the CRT basis $\vec{c}$ and powerful basis $\vec{p} = (\zeta_m^j)_{j \in [\varphi(m)]}$ of R_q are related by

$$\vec{p}^T = \vec{c}^T \cdot \mathrm{CRT}_m, \tag{5.1}$$

where the matrix CRT_m is over $\mathbb{Z}_q$. So if $a \in R_q$ has coefficient vector $\mathbf{a} \in \mathbb{Z}_q^{[\varphi(m)]}$ in the powerful basis—i.e., $a = \langle \vec{p}, \mathbf{a} \rangle$—then its coefficient vector in the CRT basis is $\mathrm{CRT}_m \cdot \mathbf{a} \in \mathbb{Z}_q^{\mathbb{Z}^*_m}$—i.e., $a = \langle \vec{c}, \mathrm{CRT}_m \cdot \mathbf{a} \rangle$—and similarly for $\mathcal{I}_q$ by linearity. Using the sparse decomposition of CRT_m and its inverse from Section 3, we can therefore switch efficiently between the power and Chinese remainder bases.

Finally, for arbitrary m, by the tensorial decomposition of R_q, multiplication is still componentwise in the CRT basis. Moreover, by the definitions of $\vec{p}$, $\vec{c}$, and CRT_m as tensor products and the mixed-product property, it immediately follows that Equation (5.1) holds as well.

6 The Decoding Basis of $R^\vee$

When working with ring-LWE we need to perform a variety of operations over $R^\vee = \langle t^{-1} \rangle$ or $R^\vee_q$. For certain operations it is best to use a certain $\mathbb{Z}$-basis of $R^\vee$ (and $\mathbb{Z}_q$-basis of $R^\vee_q$), defined below.

Let τ be the automorphism (and involution) of K that maps ζ_m to $\zeta_m^{-1} = \zeta_m^{m-1}$. We refer to τ as the conjugation map, since under the canonical embedding it corresponds to complex conjugation: $\sigma(\tau(a)) = \overline{\sigma(a)}$. Notice that for any m' dividing m, τ also maps $\zeta_{m'} = \zeta_m^{m/m'}$ to $\zeta_{m'}^{-1} = \zeta_m^{-m/m'}$. Also note that $\tau(\vec{p})$ is a $\mathbb{Z}$-basis of R, since τ is an automorphism and hence fixes R.

Definition 6.1. *The* decoding basis *of $R^\vee$ is $\vec{d} = \tau(\vec{p})^\vee$, the dual of the conjugate of the powerful basis $\vec{p}$.*[11]

The decoding basis therefore has the same index set as $\vec{p}$. When m is a prime power, $\vec{d}$ is simply the dual of the conjugate power basis $\tau(\vec{p}) = (\zeta_m^{-j})_{j \in [\varphi(m)]}$ of R. For general m, because $\tau(\vec{p})$ is the tensor product

[11] Note that unlike the powerful and CRT bases, we do not define a decoding basis for any other power of $R^\vee$; see Section 6.2 for discussion. Also, there is some flexibility in the choice of $\vec{d}$, and other definitions may be nearly as good, e.g., $\vec{d} = \vec{p}^\vee$ (without conjugation). We adopt the above definition because it corresponds to the adjoint of $\sigma(\vec{p}^T)$, and yields a particularly simple connection between $\vec{d}$ and the powerful basis $t^{-1}\vec{p}$ of $R^\vee$ (see Lemma 6.3).

of the conjugate power bases for prime-power cyclotomics R_ℓ, and $(\vec{a} \otimes \vec{b})^\vee = (\vec{a}^\vee \otimes \vec{b}^\vee)$, it follows that $\vec{d}$ is the tensor product of the decoding bases for each $R_\ell^\vee$.

We start with some basic facts about the decoding basis. Any $a \in K_\mathbb{R}$ can be represented in the decoding basis as $a = \langle \vec{d}, \mathbf{a} \rangle$ for some vector $\mathbf{a}$ of real coefficients, given by

$$a_j = \mathrm{Tr}(a \cdot d_j^\vee) = \mathrm{Tr}(a \cdot \tau(p_j)) = \langle \sigma(a), \sigma(p_j) \rangle \quad \iff \quad \mathbf{a} = \mathrm{CRT}_m^* \cdot \sigma(a). \tag{6.1}$$

Since $\vec{d}$ is the dual basis of $\tau(\vec{p})$, which embeds as $\sigma(\tau(\vec{p}^T)) = \overline{\mathrm{CRT}_m}$ over $\mathbb{C}$, we have that $\vec{d}$ embeds as

$$\sigma(\vec{d}^T) = (\mathrm{CRT}_m^*)^{-1}.$$

Lemma 4.3, and the fact that complex conjugation leaves singular values unchanged, implies the following geometric fact about the decoding basis.

Lemma 6.2. *The spectral norm of $\vec{d}$ is $s_1(\vec{d}) = \sqrt{\mathrm{rad}(m)/m}$.*

We point out that $s_1(\vec{d})$ can be as large as 1 (in the extreme case where m is square free), which, unlike for $\vec{p}$ (see Lemma 4.3), is much larger than the normalized determinant $\det(R^\vee)^{1/n} = \Delta_K^{-1/(2n)} \approx 1/\sqrt{n}$. Fortunately, the decoding basis is still always a good choice for discretizing a continuous ring-LWE error distribution (while increasing the subgaussian parameter only slightly), because the input error distribution needs to have Gaussian parameter at least $\omega(\sqrt{\log n})$ for provable worst-case hardness (see Theorem 2.22). We also point out that if $\vec{d}$ were instead defined as the dual of the *power* basis (or its conjugate), then its spectral norm could be much larger: e.g., for $m = 1155 = 3 \cdot 5 \cdot 7 \cdot 11$ we would have $s_1(\vec{d}) \approx 22.6$.

In the next few subsections, we prove several important and useful properties of the decoding basis, summarized as follows:

- There are very fast linear transformations (requiring $O(nd)$ scalar operations with small hidden constant, where d is the number of prime divisors of m) for converting between the decoding basis $\vec{d}$ and the powerful basis $t^{-1}\vec{p}$ of $R^\vee$ (see Section 6.1).
- Short elements (as always, in the sense of the canonical embedding) of K have optimally small coefficients with respect to $\vec{d}$, making it a best choice for decoding $R^\vee$. Moreover, $\vec{d}$ also yields (nearly) optimal decoding in higher powers of $R^\vee$. (See Section 6.2.)
- Continuous Gaussians (especially spherical ones) as represented in the decoding basis can be sampled very simply and efficiently (see Section 6.3).

The first fact, combined with the fast CRT transformation, means that we can efficiently convert among the decoding, power, and CRT bases of $R^\vee$ (or $R_q^\vee$) as needed. The latter two facts mean that the decoding basis is an excellent choice for generating and decoding error terms (e.g., in encryption and decryption, respectively). By contrast, the power basis and other natural bases of R or $R^\vee$ do not typically enjoy the above properties (except when m is a power of 2), and while they can in principle be used for all the same tasks, it would come at a potentially large loss in tightness and/or computational efficiency.

6.1 Relation to the Powerful Basis

Recall that both $\vec{d}$ and $t^{-1}\vec{p}$ are $\mathbb{Z}$-bases of $R^\vee$, so there is a unimodular transformation that relates them, which is given in the following lemma.

Lemma 6.3. *Let m be a power of a prime p and let $m' = m/p$, so $\varphi(m) = \varphi(p) \cdot m'$. Then*

$$\vec{d}^T = t^{-1}\vec{p}^T \cdot (L_p \otimes I_{[m']}), \tag{6.2}$$

where $L_p \in \mathbb{Z}^{[\varphi(p)]\times[\varphi(p)]}$ is the lower-triangular matrix with 1s throughout its lower-left triangle, i.e., its (i,j)th entry is 1 for $i \geq j$, and 0 otherwise.

Proof. First reindex the conjugate power basis using index set $[\varphi(p)] \times [m']$, as $\tau(p_{(j_0,j_1)}) = \zeta_p^{-j_0} \cdot \zeta_m^{-j_1}$, and reindex $\vec{d}$ similarly. Equation (6.2) may then be rewritten equivalently as

$$d_{(j_0,j_1)} = t^{-1} \cdot \left(\zeta_p^{j_0} + \zeta_p^{j_0+1} + \cdots + \zeta_p^{p-2}\right) \cdot \zeta_m^{j_1} = \tfrac{1}{m}\left(\zeta_p^{j_0} - \zeta_p^{p-1}\right) \cdot \zeta_m^{j_1}, \tag{6.3}$$

where recall from Definition 2.17 that $t^{-1} = (1-\zeta_p)/m$. To verify the above equation, observe that the product of the right-hand expression with $\tau(p_{(j_0',j_1')})$ for any $(j_0', j_1') \in [\varphi(p)] \times [m']$ is

$$\tfrac{1}{m}\left(\zeta_p^{j_0-j_0'} - \zeta_p^{p-1-j_0'}\right) \cdot \zeta_m^{j_1-j_1'}.$$

By Lemma 2.15, the trace of this is 0 if $j_1 \neq j_1'$ (because $j_1 - j_1' \neq 0 \bmod m'$); otherwise it is 0 if $j_0 \neq j_0'$ (because both $j_0 - j_0', p-1-j_0' \neq 0 \bmod p$); otherwise, it is 1, as desired. □

Observe that multiplication by L_p can be done in $O(\varphi(p))$ scalar operations via partial sums, and similarly for L_p^{-1} via successive differences. Therefore, multiplication by $L_m = (L_p \otimes I_{[m']})$ or L_m^{-1} can be done in a linear number of scalar operations. Finally, for arbitrary m having prime-power factorization $m = \prod_\ell m_\ell$, by the definitions of $\vec{p}$, $\vec{d}$, and t as tensor products and the mixed-product property, we also have

$$\vec{d}^T = t^{-1}\vec{p}^T \cdot L_m, \quad \text{where} \quad L_m = \bigotimes\nolimits_\ell L_{m_\ell}. \tag{6.4}$$

By the discussion at the end of Section 2.1, we can therefore multiply by L_m or L_m^{-1} in $O(nd)$ scalar operations, where d is the number of distinct prime divisors of m and $n = \varphi(m)$.

6.2 Decoding $R^\vee$ and Its Powers

Recall from Section 2.4.1 the "round-off" decoding procedure, which uses short linearly independent vectors in a dual lattice $\Lambda^\vee$ to recover a sufficiently short $\mathbf{x}$, given $\mathbf{x} \bmod \Lambda$. To decode $K/R^\vee$, we apply the procedure using the decoding basis $\vec{d}$ of $R^\vee$, whose dual basis in $(R^\vee)^\vee = R$ is the conjugate powerful basis $\tau(\vec{p})$. By Claim 2.10, the distance (or subgaussian parameter) that the procedure successfully decodes from depends inversely on the maximum length of the dual elements, and by Claim 4.2, every p_j in the powerful basis has $\|\tau(p_j)\|_2 = \sqrt{n}$. From this we get corresponding bounds on the decoding operation, as summarized below in Lemmas 6.5 and 6.6. We remark that the decoding basis is an optimal choice here: by Lemma 2.14, every nonzero element of R has length at least $\sqrt{n}$, hence no shorter set of dual elements exists.

In some applications (e.g., homomorphic encryption), we need to solve the more general problem of decoding $K/\mathcal{I}$, where $\mathcal{I} = (R^\vee)^k = \langle t^{-k} \rangle$ for some (usually small) $k \geq 1$. The naïve way to do this would be to apply the round-off procedure with the $\mathbb{Z}$-basis $t^{1-k}\vec{d}$ of $\mathcal{I}$. This, however, turns out to be highly suboptimal for many values of m, because the elements of the dual basis $t^{k-1}\tau(\vec{p})$ might be much longer than the shortest nonzero elements of $\mathcal{I}^\vee = \langle t^{k-1} \rangle$.[12]

[12] This can be seen already when $k = 2$ and m is a moderately large prime: using the equality $t = m/g$ and noticing that some of the embeddings of $g = 1 - \zeta_m$ are very close to zero, we see that the length of t is a rather large $\Omega(m^2)$.

Instead, in the round-off algorithm we use the *scaled decoding basis* $\hat{m}^{1-k}\vec{d}$, which generates the superideal $\mathcal{J} = \hat{m}^{1-k}R^\vee = t^{-k}g^{1-k} \supseteq \mathcal{I}$, and whose dual elements are $\hat{m}^{k-1}\tau(\vec{p}) \subset \mathcal{I}^\vee$. (Recall from Definition 2.17 that $\hat{m} = t \cdot g$ for some $g \in R$, where $\hat{m} = m/2$ if m is even and $\hat{m} = m$ otherwise.) The lengths of the dual elements are therefore $\hat{m}^{k-1}\sqrt{n}$, from which one gets the bounds summarized in Lemma 6.6 below.

We point out that the use of $\hat{m}^{1-k}\vec{d}$ for decoding $K/\mathcal{I}$ is either optimal or nearly so. Indeed, by Lemma 2.14 and Equation (2.10), the minimum distance of $\mathcal{I}^\vee = (R^\vee)^{1-k}$ is at least $\sqrt{n} \cdot \mathrm{N}(R^\vee)^{(1-k)/n} = \sqrt{n} \cdot \Delta_K^{(k-1)/n}$, so by Equation (2.8), the dual elements $\hat{m}^{k-1}\tau(\vec{p}) \subset \mathcal{I}^\vee$ are nearly as short as possible:

$$\frac{\|\hat{m}^{k-1}\tau(\vec{p})\|_2}{\lambda_1(\mathcal{I}^\vee)} = \frac{\hat{m}^{k-1}\sqrt{n}}{\lambda_1(\mathcal{I}^\vee)} \leq \Big(\prod_{\text{odd prime } p|m} p^{1/(p-1)} \Big)^{k-1},$$

which for almost all choices of m and small k is quite small. (For example, the term inside the parentheses is only ≈ 6.73 when taking all odd primes up to 17, which corresponds to $m \geq 255{,}255$.) Moreover, the above lower bound on $\lambda_1(\mathcal{I}^\vee)$ may not be tight; we suspect that in most cases of interest the minimum distance of $\mathcal{I}^\vee$ is exactly $\hat{m}^{k-1}\sqrt{n}$, which would imply that the scaled decoding basis is optimal.

We summarize the above discussion in the following definition and lemmas. As it will be more convenient for applications, we consider a "scaled up and discretized" version of the decoding procedure, where we decode from $\mathcal{I}_q$ to $\mathcal{I}$ for some $q \geq 1$. So the unknown short element is guaranteed to be in $\mathcal{I}$ but is given modulo $q\mathcal{I}$, and the output is also expected to be in $\mathcal{I}$. The only difference this makes (apart from the scaling by q) is that for $k \geq 2$, since the scaled decoding basis $\hat{m}^{1-k}\vec{d}$ may generate a strict superideal $\mathcal{J} \supset \mathcal{I}$, the round-off procedure might output an element that is not in $\mathcal{I}$. In such a case we just consider the output to be undefined. Lemmas 6.5 and 6.6 show that as long as the unknown element in $\mathcal{I}$ is short enough (or has a small enough subgaussian parameter), the decoding procedure correctly outputs it.

Definition 6.4 (Decoding $\mathcal{I}_q$ to $\mathcal{I}$). *Let $\mathcal{I} = (R^\vee)^k$ for some $k \geq 1$, and define the decoding function $[\![\cdot]\!] \colon \mathcal{I}_q \to \mathcal{I}$ as follows. For input $\bar{a} \in \mathcal{I}_q$, write $\bar{a} = \langle \hat{m}^{1-k}\vec{d}, \bar{\mathbf{a}} \rangle \bmod q\mathcal{J}$ for some vector $\bar{\mathbf{a}}$ over $\mathbb{Z}_q$, where $\mathcal{J} = \hat{m}^{1-k}R^\vee \supseteq \mathcal{I}$. Define $[\![\bar{a}]\!] := \langle \hat{m}^{1-k}\vec{d}, [\![\bar{\mathbf{a}}]\!] \rangle$ if this value is in $\mathcal{I}$, otherwise $[\![\bar{a}]\!]$ is undefined.*
(Recall that $[\![\bar{\mathbf{a}}]\!]$ is a vector over $\mathbb{Z}$, as defined in the beginning of Section 2.)

Lemma 6.5. *Let $\mathcal{I} = (R^\vee)^k$ for some $k \geq 1$, let $a \in \mathcal{I}$ and write $a = \langle \hat{m}^{1-k}\vec{d}, \mathbf{a} \rangle$ for some integral coefficient vector $\mathbf{a}$, and let $q \geq 1$ be an integer. If every coefficient $a_j \in [-q/2, q/2)$, then $[\![a \bmod q\mathcal{I}]\!] = a$. In particular, if every a_j is δ-subgaussian with parameter s, then $[\![a \bmod q\mathcal{I}]\!] = a$ except with probability at most $2n\exp(\delta - \pi q^2/(2s)^2)$.*

Proof. The first part is by Claim 2.10. The second part is by the tail bound on subgaussian random variables (Equation (2.2)), and the union bound. □

Lemma 6.6. *Let $\mathcal{I} = (R^\vee)^k$ for some $k \geq 1$, and let $a \in \mathcal{I}$.*

- *Writing $a = \langle \hat{m}^{1-k}\vec{d}, \mathbf{a} \rangle$ for some integral vector $\mathbf{a}$, we have that every $|a_j| \leq \hat{m}^{k-1}\sqrt{n} \cdot \|a\|_2$.*
- *If a is δ-subgaussian with parameter s, and $b \in (R^\vee)^\ell$ for some $\ell \geq 0$ is arbitrary, then writing $a \cdot b = \langle \hat{m}^{1-k-\ell}\vec{d}, \mathbf{c} \rangle$ for some integral vector $\mathbf{c}$, we have that every c_j is δ-subgaussian with parameter $\hat{m}^{k+\ell-1}\|b\|_2 \cdot s$.*

We remark that the second item above gives a bound that is a $\sqrt{n}$ factor tighter than what we would obtain by treating $a \cdot b$ as δ-subgaussian with parameter $s\|b\|_2$. The tighter bound results from using the particular properties of the powerful basis, namely, that all its elements have ℓ_∞ norm 1.

Proof. The dual elements of $\hat{m}^{1-k}\vec{d}$ are $\hat{m}^{k-1}\tau(\vec{p})$, which all have ℓ_2 norm $\hat{m}^{k-1}\sqrt{n}$. The first item then follows by the Cauchy-Schwarz inequality.

For the second item, notice that the coefficient c_j of $a \cdot b$ in the scaled decoding basis $\hat{m}^{1-k-\ell}\vec{d}$ is

$$c_j = \mathrm{Tr}(\hat{m}^{k+\ell-1}\tau(p_j) \cdot ab) = \hat{m}^{k+\ell-1}\,\mathrm{Tr}((\tau(p_j) \cdot b) \cdot a),$$

which by definition and by Claim 4.2 is δ-subgaussian with parameter

$$\hat{m}^{k+\ell-1}\|\tau(p_j) \cdot b\|_2 \cdot s \le \hat{m}^{k+\ell-1}\|\tau(p_j)\|_\infty \cdot \|b\|_2 \cdot s = \hat{m}^{k+\ell-1}\|b\|_2 \cdot s. \qquad \square$$

6.2.1 Implementation Notes

We conclude this subsection by outlining an efficient implementation of the decoding operation from Definition 6.4. As usual, we wish to use only (nearly) linear time operations, and avoid high-precision quantities. Recall that our goal is to recover an unknown element $a \in \mathcal{I}$ given $\bar{a} = a \bmod q\mathcal{I}$, where $\mathcal{I} = (R^\vee)^k$ for some $k \ge 1$. We assume that the input $\bar{a} \in \mathcal{I}_q$ is given in the form of a coefficient vector $\bar{\mathbf{a}}$ over $\mathbb{Z}_q$ satisfying $\bar{a} = \langle t^{1-k}\vec{b}, \bar{\mathbf{a}}\rangle \bmod q\mathcal{I}$, where $\vec{b}$ is some $\mathbb{Z}_q$-basis of $R_q^\vee$. The output will be given as a coefficient vector $\mathbf{a}$ over $\mathbb{Z}$ with respect to the decoding basis $t^{1-k}\vec{d}$ of $\mathcal{I}$.

The case $k = 1$ can be implemented straightforwardly. Suppose the basis $\vec{b}$ used to specify the input $\bar{a} \in R_q^\vee$ is the decoding basis, i.e., $\bar{a} = \langle \vec{d}, \bar{\mathbf{a}}\rangle \bmod qR^\vee$. We then simply output the integer coefficient vector $\mathbf{a} = [\![\bar{\mathbf{a}}]\!]$ also relative to the decoding basis, i.e., $a = \langle \vec{d}, \mathbf{a}\rangle \in R^\vee$. The number of operations is clearly linear. If the input is represented in a different basis $\vec{b}$, we first convert to the decoding basis, which is very efficient for all bases we consider.

The case $k > 1$ is more interesting, and consists of three efficient steps:

1. compute the representation of $\bar{a}' = \bar{a} \bmod q\mathcal{J}$ in the $\mathbb{Z}_q$-basis $\hat{m}^{1-k}\vec{b}$ of $\mathcal{J}_q$ (where recall that $\mathcal{J} = \hat{m}^{1-k}R^\vee \supseteq \mathcal{I}$);
2. decode it as in the case $k = 1$ to an element $a' \in \mathcal{J}$ (which will equal a if decoding was successful);
3. compute the representation of a' in the $\mathbb{Z}$-basis $t^{1-k}\vec{d}$ of $\mathcal{I}$.

We next explain each of the three steps in detail.

The first step, it turns out, is equivalent to multiplication by $g^{k-1} \in R$, where recall from Definition 2.17 that $\hat{m} = g \cdot t$. Indeed, by factoring out g^{k-1} from the modulus and both sides of the equality, we have

$$g^{k-1} \cdot \bar{a} = \langle t^{1-k}\vec{b}, \bar{\mathbf{a}}\rangle \bmod q\mathcal{I} \quad \iff \quad \bar{a} = \langle \hat{m}^{1-k}\vec{b}, \bar{\mathbf{a}}\rangle \bmod q\mathcal{J},$$

i.e., the desired coefficients of $\bar{a} \bmod q\mathcal{J}$ in basis $\hat{m}^{1-k}\vec{b}$ are exactly those of $g^{k-1}\bar{a}$ in basis $t^{1-k}\vec{b}$. Typically the input basis $\vec{b}$ at this stage would be the CRT basis, and for efficiency one could precompute the CRT coefficients of g^{k-1}, making this step linear time. In addition, multiplication by g in the powerful and decoding bases is also (nearly) linear time, as described below.

The second step is essentially identical to the case $k = 1$. Take the output $\bar{a}'$ of the first step, convert it (if needed) to a representation in the scaled decoding basis $\hat{m}^{1-k}\vec{d}$, so that $\bar{a}' = \langle \hat{m}^{1-k}\vec{d}, \bar{\mathbf{a}}'\rangle$ for some $\bar{\mathbf{a}}'$ over $\mathbb{Z}_q$, and then output the coefficient vector $[\![\bar{\mathbf{a}}']\!]$ over $\mathbb{Z}$, which represents the element $a' = \langle \hat{m}^{1-k}\vec{d}, [\![\bar{\mathbf{a}}']\!]\rangle \in \mathcal{J}$. The element a' is exactly the output of the decoding procedure as in Definition 6.4, except that it might not be in $\mathcal{I}$ (in which case decoding failed).

Finally, in the third step, we convert the representation of a' in the $\mathbb{Z}$-basis $\hat{m}^{1-k}\vec{d}$ of $\mathcal{J}$ to a representation in a $\mathbb{Z}$-basis of $\mathcal{I}$, namely $t^{1-k}\vec{d}$. This conversion might be impossible if $a' \notin \mathcal{I}$, which indicates decoding failure. Assuming $a' \in \mathcal{I}$, it is immediate to see that this conversion is equivalent to division by g^{k-1}:

$$g^{1-k} \cdot a' = \langle \hat{m}^{1-k}\vec{d}, \mathbf{a} \rangle \in \mathcal{J} \quad \iff \quad a' = \langle t^{1-k}\vec{d}, \mathbf{a} \rangle \in \mathcal{I},$$

i.e., the desired coefficients of a' in the $\mathbb{Z}$-basis $t^{1-k}\vec{d}$ of $\mathcal{I}$ are exactly those of $g^{1-k} \cdot a'$ in basis $\hat{m}^{1-k}\vec{d}$.

Division by g^{k-1} can be performed somewhat efficiently using the CRT transform over $\mathbb{C}$, but this requires $\Omega(n \log n)$ time and high-precision operations (since in contrast with the first step, here we are working with $\mathbb{Z}$-bases, and not modulo q). A better way follows from noticing that multiplication and division by g have nice forms in the decoding basis, i.e., $g \cdot \vec{d}^T = \vec{d}^T \cdot A$ for some integral matrix A that is efficient to multiply and divide by. By the tensorial decompositions of $\vec{d}$ and of g, it suffices to consider the case where m is a power of a prime p. Using Equation (6.3) and letting $m' = m/p$, one can verify that multiplication and division by $g = 1 - \zeta_p$ in the decoding basis are given, respectively, by the $[n]$-by-$[n]$ matrices

$$A = \begin{pmatrix} 2 & 1 & 1 & 1 & 1 \\ -1 & 1 & & & \\ & -1 & 1 & & \\ & & & \ddots & \\ & & & -1 & 1 \end{pmatrix} \otimes I_{[m']}, \quad A^{-1} = \frac{1}{p} \begin{pmatrix} 1 & 2-p & 3-p & \cdots & -1 \\ 1 & 2 & 3-p & \cdots & -1 \\ 1 & 2 & 3 & \cdots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 2 & 3 & \cdots & p-1 \end{pmatrix} \otimes I_{[m']}.$$

It is easy to see that left-multiplication by A can be performed in time linear in n. Moreover, multiplication by A^{-1} can also be done in linear time, because every row differs from each of its adjacent rows in just one entry. Note that to avoid rational arithmetic, one would actually multiply by the integer matrix pA^{-T} and then evenly divide the result by p. If the latter step is not possible, that indicates decoding failure.

Lastly, we also note that multiplication by g in the powerful basis is given by JA^TJ, where $J = J_{[n]}$ is the $[n]$-by-$[n]$ reversal matrix, obtained by reversing the columns of the identity matrix $I_{[n]}$ (so $J = J^{-1}$ and $J_{[n]} = J_{[\varphi(p)]} \otimes J_{[m']}$). Therefore, in the powerful basis we can also multiply and divide by g in linear time per prime-power divisor of m.

6.3 Sampling Gaussians in the Decoding Basis

We now describe how to efficiently sample continuous Gaussians over $K_{\mathbb{R}}$, as represented in the decoding basis. In order to obtain the real coefficient vector $\mathbf{a}$ of some Gaussian-distributed $a \in K_{\mathbb{R}}$, by Equation (6.1) it suffices to sample $\sigma(a)$ from the continuous Gaussian distribution over H and then left-multiply by CRT^*_m. The latter step is best done using the sparse decomposition given in Section 3. Recalling the definition of H and its unitary basis matrix $B = \frac{1}{\sqrt{2}} \begin{pmatrix} I & \sqrt{-1}J \\ J & -\sqrt{-1}I \end{pmatrix} \in \mathbb{C}^{\mathbb{Z}^*_m \times [\varphi(m)]}$ from Section 2.2, we see that sampling $\sigma(a)$ amounts to sampling n independent real Gaussians used as coefficients for the columns of B, or equivalently, sampling the first $n/2$ complex coordinates as independent complex Gaussians, and completing the remaining $n/2$ coordinates using the conjugate symmetry of H.

While the above is already quite efficient, here we show that a significantly faster algorithm exists when $\mathrm{rad}(m) \ll m$. The basic idea is to notice that multiplication by the matrix CRT^*_m, with its decomposition as in Equation (3.2), starts with multiplication by two scaled unitary matrices: a (typically high-dimensional) DFT tensored with identity, and a twiddle matrix. Since spherical Gaussians are invariant under unitary transformations, we can effectively skip these two multiplications, and we only need to multiply by the (often much lower-dimensional) CRT^*_p matrices for those primes p dividing m. Details follow.

Using Equation (3.2), for any m_ℓ that is a power of a prime p_ℓ, letting $m'_\ell = m_\ell/p_\ell$ we have

$$\mathrm{CRT}^*_{m_\ell} = (\mathrm{CRT}^*_{p_\ell} \otimes I_{[m'_\ell]}) \cdot \sqrt{m'_\ell} \cdot Q_\ell$$

for some unitary Q_ℓ, because the twiddle matrix $\hat{T}_{m_\ell}$ and scaled Fourier matrix $\mathrm{DFT}_{m'_\ell}/\sqrt{m'_\ell}$ are both unitary. Therefore, by the mixed-product property we have

$$\mathrm{CRT}^*_m = \bigotimes\nolimits_\ell \mathrm{CRT}^*_{m_\ell} = \bigotimes\nolimits_\ell (\mathrm{CRT}^*_{p_\ell} \otimes I_{[m'_\ell]}) \cdot \sqrt{m/\operatorname{rad}(m)} \bigotimes\nolimits_\ell Q_\ell.$$

Since $Q = \bigotimes_\ell Q_\ell$ is unitary, it sends a spherical Gaussian distribution over $H \subset \mathbb{C}^{\mathbb{Z}^*_m}$ to a spherical Gaussian distribution (of the same parameter) over the subspace $H' = QH \subset \mathbb{C}^{\mathbb{Z}^*_m}$.[13] Therefore, to sample a continuous Gaussian of parameter s in the decoding basis, it suffices to generate a Gaussian of parameter $s\sqrt{m/\operatorname{rad}(m)}$ over H' and then left-multiply the result by

$$C^* := \bigotimes\nolimits_\ell (\mathrm{CRT}^*_{p_\ell} \otimes I_{[m'_\ell]}) = \mathrm{CRT}^*_{\mathrm{rad}(m)} \otimes I_{[m/\operatorname{rad}(m)]}.$$

The latter requires $n/\varphi(p_\ell)$ parallel applications of $\mathrm{CRT}^*_{p_\ell}$, in sequence for each ℓ, which can be done in a total of $O(n \log(\mathrm{rad}(m)))$ scalar operations.

It remains to explain how to sample a spherical Gaussian from H'. For this it suffices to give a unitary basis matrix B' of H', which allows us to generate a Gaussian over H' as $B'\mathbf{c}$, where $\mathbf{c}$ is real Gaussian. Now, observe that the subspace H' is

$$H' = \{\mathbf{x} \in \mathbb{C}^{\mathbb{Z}^*_m} : C^*\mathbf{x} \in \mathbb{R}^{[\varphi(m)]}\},$$

because H' is a real vector space of dimension n, and $C^*H' = \mathbb{R}^{[\varphi(m)]}$. So it suffices to give a unitary matrix B' such that C^*B' is real. By the mixed-product property, such a matrix is

$$B' = \bigotimes\nolimits_\ell (B'_{p_\ell} \otimes I_{[m'_\ell]}),$$

where $B'_{p_\ell} = \frac{1}{\sqrt{2}} \begin{pmatrix} I & \sqrt{-1}J \\ J & -\sqrt{-1}I \end{pmatrix}$ for $p_\ell > 2$, and is the scalar identity for $p_\ell = 2$. Clearly, multiplication by B' is a simple linear-time operation in the dimension.

Finally, we remark that because the final vector of decoding basis coefficients is $C^*B'\mathbf{c}$ for a real Gaussian $\mathbf{c}$, it is possible to generate these coefficients using just real arithmetic as $D\mathbf{c}$, where $D = \bigotimes_\ell (D_{p_\ell} \otimes I_{[m'_\ell]})$ and $D_{p_\ell} = \mathrm{CRT}^*_{p_\ell} \cdot B'_{p_\ell}$ is a real $\varphi(p_\ell)$-by-$\varphi(p_\ell)$ matrix.

7 Regularity

In this section we prove a certain "regularity lemma" that is useful in cryptographic applications, such as when adapting the "primal" [Reg05] and "dual" [GPV08] LWE-based cryptosystems, and the identity-based versions of the latter scheme, to ring-LWE. (See Section 8.1 for such an adaptation of the dual cryptosystem.) Independently, a closely related statement, specialized to power-of-2 cyclotomics, was recently shown in [SS11] with a different style of proof.

The theorem says the following. Assume we are working with the mth cyclotomic of degree $n = \varphi(m)$, and let $q \geq 1$ be a prime integer. Let $a_1, \ldots, a_{\ell-1}$ be chosen uniformly and independently from R_q. Then,

[13] Here and in what follows, we identify the index set $\mathbb{Z}^*_m$ with the set $\prod_\ell (\mathbb{Z}^*_{p_\ell} \times [m'_\ell])$ as in the decomposition of CRT_m, and similarly identify $[\varphi(m)]$ with $\prod_\ell [\varphi(m_\ell)]$.

with high probability over the choice of the a_i, the distribution of $b_0 + \sum_{i=1}^{\ell-1} b_i a_i$ is within statistical distance $2^{-\Omega(n)}$ of uniform, where the b_i are chosen from a discrete Gaussian distribution on R of width essentially $n \cdot q^{1/\ell}$ (in the canonical embedding). Equivalently, the lemma says that if a_0 is any fixed invertible element of R_q and $a_1, \ldots, a_{\ell-1}$ are uniformly and independently chosen from R_q, then $\sum_{i=0}^{\ell-1} b_i a_i$ is within $2^{-\Omega(n)}$ of uniform, where the b_i are chosen as before. The equivalence follows by simply dividing by a_0. (The lemma we prove is actually more general, and applies to the joint distribution of $k \geq 1$ sums as above; see Theorem 7.4 and Corollary 7.5 for the exact statement.)

This regularity statement is already interesting and non-trivial when ℓ is as small as 2, and is close to being tight: for instance, when m is a power of 2, a width of at least $\sqrt{n} q^{1/\ell}$ is required just for entropy reasons. To see this, recall that R is a rotation of $\sqrt{n}\mathbb{Z}^n$, so roughly speaking, a discrete Gaussian of width t covers $(t/\sqrt{n})^n$ points.

One might wonder about the significance of the b_0 term, and why we do not analyze the regularity of $\sum_{i=0}^{\ell-1} b_i a_i$ when all the a_i are chosen uniformly from R_q. In fact, a regularity lemma for exactly such sums was shown by Micciancio [Mic02]. (His work is specialized to the ring $R = \mathbb{Z}[x]/\langle x^n - 1\rangle$, but can be extended to other rings, as observed in [SSTX09].) Unfortunately, such sums have a much worse regularity property, and in particular require super-constant ℓ to get negligible distance to uniformity. To see why this is the case, assume that q is a prime satisfying $q = 1 \bmod m$, so that $\langle q \rangle$ splits completely into n ideals of norm q each. Letting $\mathfrak{q}$ denote one of these prime factors, notice that with probability $q^{-\ell}$, all the a_i are in $\mathfrak{q}$. In this case, $\sum_{i=1}^{m} b_i a_i$ is in $\mathfrak{q}$ with certainty, and its distribution is therefore very far from uniform. By adding the b_0 term we avoid this "common divisor" problem and get much better regularity, providing exponentially small distance to uniformity already for ℓ as small as 2. It is also worth mentioning that including the b_0 term (or equivalently, requiring a_0 to be uniform) corresponds to the "normal form" of ring-LWE and ring-SIS.

We start with a technical claim on the Gaussian weight on a lattice.

Claim 7.1. *For any n-dimensional lattice Λ and $\varepsilon, r > 0$,*

$$\rho_{1/r}(\Lambda) \leq \max\left(1, \left(\frac{\eta_\varepsilon(\Lambda^\vee)}{r}\right)^n\right)(1+\varepsilon).$$

Proof. For $r \geq \eta_\varepsilon(\Lambda^\vee)$, the claim follows from Definition 2.5. For $r < \eta = \eta_\varepsilon(\Lambda^\vee)$, it follows from the Poisson summation formula (see [MR04, Lemma 2.8]) that

$$\rho_{1/r}(\Lambda) = (\det \Lambda)^{-1} \cdot r^{-n} \cdot \rho_r(\Lambda^\vee) < (\det \Lambda)^{-1} \cdot r^{-n} \cdot \rho_\eta(\Lambda^\vee) = (\eta/r)^n \cdot \rho_{1/\eta}(\Lambda),$$

and the claim follows from the previous case. □

Using Lemma 2.6 and Lemma 2.14 we have

$$\eta_{2^{-2n}}(\mathcal{I}^\vee) \leq \sqrt{n}/\lambda_1(\mathcal{I}) \leq (\mathrm{N}(\mathcal{I}))^{-1/n},$$

which implies the following corollary.

Corollary 7.2. *For any ideal $\mathcal{I}$ and $r > 0$,*

$$\rho_{1/r}(\mathcal{I}) \leq \max\left(1, \mathrm{N}(\mathcal{I})^{-1} r^{-n}\right)(1 + 2^{-2n}).$$

We will also need the following algebraic claim.

Claim 7.3. *In the mth cyclotomic number field of degree n, for any $q, k \geq 1$,*

$$\sum_{\mathcal{J}|\langle q \rangle} \mathrm{N}(\mathcal{J})^k \leq \exp(c) q^{kn} \leq q^{kn+2},$$

where c is the number of distinct prime integer divisors of q.

Proof. The second inequality is clear. For the first inequality, it suffices to consider the case of a prime power $q = p^e$. Indeed, if q_1 and q_2 are coprime then

$$\sum_{\mathcal{J}|\langle q_1 q_2 \rangle} \mathrm{N}(\mathcal{J})^k = \Big(\sum_{\mathcal{J}|\langle q_1 \rangle} \mathrm{N}(\mathcal{J})^k \Big) \Big(\sum_{\mathcal{J}|\langle q_2 \rangle} \mathrm{N}(\mathcal{J})^k \Big).$$

Next, recall from Section 2.5.5 that for any integer prime p, the ideal $\langle p \rangle$ factors as $\mathfrak{p}_1^h \cdots \mathfrak{p}_g^h$ where $h = \varphi(p^d)$, $d \geq 0$ is the largest integer such that p^d divides m, each $\mathfrak{p}_i$ is of norm p^f where $f \geq 1$ is the multiplicative order of p modulo m/p^d, and $g = n/(hf)$. Therefore, $\langle q \rangle = \mathfrak{p}_1^{eh} \cdots \mathfrak{p}_g^{eh}$, and

$$\begin{aligned}
\sum_{\mathcal{J}|\langle q \rangle} \mathrm{N}(\mathcal{J})^k &= \prod_{i=1}^{g} (1 + \mathrm{N}(\mathfrak{p}_i)^k + \cdots + \mathrm{N}(\mathfrak{p}_i)^{ehk}) \\
&= \Big(1 + p^{fk} + \cdots + p^{ehfk}\Big)^g \\
&\leq p^{ehfkg} (1 - p^{-fk})^{-g} \\
&\leq q^{nk} \exp(g \cdot p^{-fk}).
\end{aligned}$$

Next, observe that p^f is greater than m/p^d (since it is greater than 1 and equals 1 modulo m/p^d) and that $g \leq n/\varphi(p^d) = \varphi(m/p^d)$, hence

$$g \cdot p^{-fk} \leq g \cdot p^{-f} \leq 1,$$

which completes the proof. □

The following is the regularity theorem. Here, for a matrix $A \in R_q^{[k] \times [\ell]}$ we define

$$\Lambda^\perp(A) = \{\vec{z} \in R^{[\ell]} : A\vec{z} = 0 \bmod qR\},$$

which we identify with a lattice in H^ℓ. Its dual lattice (which is again a lattice in H^ℓ) is denoted by $\Lambda^\perp(A)^\vee$.

Theorem 7.4. *Let R be the ring of integers in the mth cyclotomic number field K of degree n, and $q \geq 2$ an integer. For positive integers $k \leq \ell \leq \mathrm{poly}(n)$, let $A = [I_{[k]} \mid \bar{A}] \in (R_q)^{[k] \times [\ell]}$, where $I_{[k]} \in (R_q)^{[k] \times [k]}$ is the identity matrix and $\bar{A} \in (R_q)^{[k] \times [\ell - k]}$ is uniformly random. Then for all $r > 2n$,*

$$\mathbb{E}_{\bar{A}} \Big[\rho_{1/r}\big(\Lambda^\perp(A)^\vee\big) \Big] \leq 1 + 2(r/n)^{-n\ell} q^{kn+2} + 2^{-\Omega(n)}.$$

In particular, if $r > 2n \cdot q^{k/\ell + 2/(n\ell)}$ then $\mathbb{E}_{\bar{A}}[\rho_{1/r}(\Lambda^\perp(A)^\vee)] \leq 1 + 2^{-\Omega(n)}$, and so by Markov's inequality, $\eta_{2^{-\Omega(n)}}(\Lambda^\perp(A)) \leq r$ except with probability at most $2^{-\Omega(n)}$.

Using Lemma 2.7, and the fact that A contains an identity submatrix $I_{[k]}$ and so the columns of A generate all of $R_q^{[k]}$, we obtain the following corollary, which is often more useful in applications.

Corollary 7.5. *Let R, n, q, k, and ℓ be as in Theorem 7.4. Assume that $A = [I_{[k]} \mid \bar{A}] \in (R_q)^{[k]\times[\ell]}$ is chosen as in Theorem 7.4. Then, with probability $1 - 2^{-\Omega(n)}$ over the choice of $\bar{A}$, the distribution of $A\vec{x} \in R_q^{[k]}$ where each coordinate of $\vec{x} \in R_q^{[\ell]}$ is chosen from a discrete Gaussian distribution of parameter $r > 2n \cdot q^{k/\ell+2/(n\ell)}$ over R, satisfies that the probability of each of the q^{nk} possible outcomes is in the interval $(1 \pm 2^{-\Omega(n)})q^{-nk}$ (and in particular is within statistical distance $2^{-\Omega(n)}$ of the uniform distribution over $R_q^{[k]}$).*

Proof of Theorem 7.4. Observe that for any $A \in (R_q)^{[k]\times[\ell]}$, the dual lattice of $\Lambda^\perp(A)$ is

$$\Lambda^\perp(A)^\vee = (R^\vee)^{[\ell]} + \left\{\tfrac{1}{q}A^T\vec{s} : \vec{s} \in (R_q^\vee)^{[k]}\right\}.$$

We therefore have

$$\begin{aligned}\mathbb{E}_{\bar{A}}\Big[\rho_{1/r}\big(\Lambda^\perp(A)^\vee\big)\Big] &= \sum_{\vec{s}\in(R_q^\vee)^{[k]}} \mathbb{E}_{\bar{A}}\Big[\rho_{1/r}\Big((R^\vee)^{[\ell]} + \tfrac{1}{q}A^T\vec{s}\Big)\Big] \\ &= \sum_{\vec{s}\in(R_q^\vee)^{[k]}} \rho_{1/r}\Big((R^\vee)^{[k]} + \tfrac{1}{q}\vec{s}\Big) \cdot \mathbb{E}_{\vec{a}}\Big[\rho_{1/r}\Big(R^\vee + \tfrac{1}{q}\langle\vec{a},\vec{s}\rangle\Big)\Big]^{\ell-k}, \qquad (7.1)\end{aligned}$$

where $\vec{a}$ is chosen uniformly from $R_q^{[k]}$. For any $\vec{s} = (s_1, \ldots, s_k) \in (R_q^\vee)^{[k]}$, define the ideal $\mathcal{I}_{\vec{s}} = s_1R + \cdots + s_kR + qR^\vee \subseteq R^\vee$; this is the "greatest common divisor" ideal of all the s_i and $qR^\vee$. Note that $\langle\vec{a},\vec{s}\rangle$ is uniformly random over $\mathcal{I}_{\vec{s}}/qR^\vee$, and so the expectation above is

$$\big|\mathcal{I}_{\vec{s}}/qR^\vee\big|^{-1} \cdot \rho_{1/r}(\tfrac{1}{q}\mathcal{I}_{\vec{s}}).$$

Therefore, if we let T denote the set of all ideals $\mathcal{J}$ satisfying $qR^\vee \subseteq \mathcal{J} \subseteq R^\vee$, we can write (7.1) as

$$\begin{aligned}&\sum_{\mathcal{J}\in T} |\mathcal{J}/qR^\vee|^{-(\ell-k)} \cdot \rho_{1/r}(\tfrac{1}{q}\mathcal{J})^{\ell-k} \sum_{\vec{s}\text{ s.t. }\mathcal{I}_{\vec{s}}=\mathcal{J}} \rho_{1/r}\Big((R^\vee)^{[k]} + \tfrac{1}{q}\vec{s}\Big) \\ &\quad\le \rho_{1/r}(R^\vee)^\ell + \sum_{\mathcal{J}\in T\setminus\{qR^\vee\}} |\mathcal{J}/qR^\vee|^{-(\ell-k)} \cdot \rho_{1/r}(\tfrac{1}{q}\mathcal{J})^{\ell-k} \cdot \Big(\rho_{1/r}(\tfrac{1}{q}\mathcal{J})^k - 1\Big) \\ &\quad\le \rho_{1/r}(R^\vee)^\ell + \sum_{\mathcal{J}\in T\setminus\{qR^\vee\}} |\mathcal{J}/qR^\vee|^{-(\ell-k)} \cdot \Big(\rho_{1/r}(\tfrac{1}{q}\mathcal{J})^\ell - 1\Big) \\ &\quad= 1 + \sum_{\mathcal{J}\in T} |\mathcal{J}/qR^\vee|^{-(\ell-k)} \cdot \Big(\rho_{1/r}(\tfrac{1}{q}\mathcal{J})^\ell - 1\Big), \qquad (7.2)\end{aligned}$$

where in the first inequality we used the fact that for every $\mathcal{J} \in T \setminus \{qR^\vee\}$, the sets $(R^\vee)^{[k]} + \frac{1}{q}\vec{s}$ for all $\vec{s}$ satisfying $\mathcal{I}_{\vec{s}} = \mathcal{J}$ are disjoint, and their union is contained in $(\frac{1}{q}\mathcal{J})^{[k]} \setminus \{\mathbf{0}\}$. Next, using Corollary 7.2, we see that

$$\begin{aligned}\rho_{1/r}(\tfrac{1}{q}\mathcal{J})^\ell &\le \max\big(1, (|\mathcal{J}/qR^\vee| \cdot \Delta_K r^{-n})^\ell\big)(1 + 2^{-2n})^\ell \\ &\le 1 + \ell 2^{1-2n} + 2(|\mathcal{J}/qR^\vee| \cdot \Delta_K r^{-n})^\ell.\end{aligned}$$

This, together with Claim 7.3 and (2.8), allows us to bound (7.2) by

$$\begin{aligned}&1 + 2^{-\Omega(n)} + 2\Delta_K^\ell r^{-n\ell} \sum_{\mathcal{J}\in T} |\mathcal{J}/qR^\vee|^k \\ &\quad\le 1 + 2^{-\Omega(n)} + 2(r/n)^{-n\ell} q^{kn+2},\end{aligned}$$

and the theorem follows. □

8 Cryptosystems

Here we give three example applications of our toolkit, which all work in arbitrary cyclotomic rings:

- In Section 8.1, we give a simple adaptation of the "dual" LWE-based public-key cryptosystem of [GPV08], which uses our regularity lemma of Section 7, and which can serve as a foundation for (hierarchical) identity-based encryption;
- In Section 8.2, we give a public-key cryptosystem with more compact public keys and ciphertexts (of only two ring elements each), analogous to the ones of [LPS10, LP11];
- In Section 8.3, we describe a symmetric-key "somewhat homomorphic" cryptosystem and associated "modulus reduction" and "key switching" algorithms.

We emphasize that throughout this section, the cryptosystems and associated operations are defined almost entirely in an implementation- and basis-independent manner, using just abstract mathematical objects and operations (e.g., ring addition and multiplication, cosets of ideals and probability distributions over them, etc.). All of the operations can be performed very efficiently using the algorithms described earlier in the paper.

In particular, our cryptosystems need to sample from subgaussian distributions over cosets of $R^\vee$ (or a scaling of it). For this purpose we can use any valid discretization $\lfloor \cdot \rceil$ as described in Section 2.4.2, applied to any continuous error distribution ψ over $K_\mathbb{R}$. The choice of discretization affects only the resulting subgaussian parameter of the sample. For example, we can use the "coordinate-wise randomized rounding" method with the decoding basis $\vec{d}$ of $R^\vee$, which gives good subgaussian bounds (see Lemma 6.2).

8.1 Dual-Style Cryptosystem

In this section we present the ring-based variant of what is commonly called "dual" LWE encryption, first introduced in [GPV08] for the purposes of constructing identity-based encryption schemes. (The name "dual" refers to the fact that the system has dual properties to Regev's first LWE-based cryptosystem [Reg05], namely, the public key is statistically close to uniform, whereas ciphertexts are only pseudorandom and have unique encryption randomness.)

Let R denote the mth cyclotomic ring (of degree $n = \varphi(m)$) and let p and q be coprime integers, where p defines the message space R_p and q is the ring-LWE modulus. Let ψ be a continuous LWE error distribution over $K_\mathbb{R}$, and let $\lfloor \cdot \rceil$ denote a valid discretization to (cosets of) $R^\vee$ or $pR^\vee$. In the key-generation algorithm we need to sample from the discrete Gaussian distribution $D_{R,r}$ for some $r \geq \sqrt{n} \cdot \omega(\sqrt{\log n})$; we can do so using the algorithm from Lemma 2.9 with the powerful basis $\vec{p}$ of R, since by Claim 4.2 its (Gram-Schmidt orthogonalized) elements have maximum length $\sqrt{n}$. We also let $\ell \geq 2$ be a parameter.

The cryptosystem is defined as follows.

- Gen: choose $a_0 = -1 \in R_q$ and uniformly random and independent $a_1, \ldots, a_{\ell-1} \in R_q$, and independent $x_0, \ldots, x_{\ell-1} \leftarrow D_{R,r}$. Output $\vec{a} = (a_1, \ldots, a_{\ell-1}, a_\ell = -\sum_{i \in [\ell]} a_i x_i) \in R_q^{\{1,\ldots,\ell\}}$ as the public key, and $\vec{x} = (x_1, \ldots, x_{\ell-1}, x_\ell = 1) \in R^{\{1,\ldots,\ell\}}$ as the secret key. Note that $\langle \vec{a}, \vec{x} \rangle = x_0 \in R_q$, by construction.
- $\mathsf{Enc}_{\vec{a}}(\mu \in R_p)$: choose independent $e_0, e_1, \ldots, e_{\ell-1} \leftarrow \lfloor p \cdot \psi \rceil_{pR^\vee}$, and $e_\ell \leftarrow \lfloor p \cdot \psi \rceil_{t^{-1}\mu + pR^\vee}$. Let $\vec{e} = (e_1, \ldots, e_\ell) \in (R^\vee)^{\{1,\ldots,\ell\}}$. Output ciphertext $\vec{c} = e_0 \cdot \vec{a} + \vec{e} \in (R_q^\vee)^{\{1,\ldots,\ell\}}$.

- $\mathsf{Dec}_{\vec{x}}(\vec{c})$: compute $d = [\![\langle \vec{c}, \vec{x} \rangle]\!] \in R^\vee$ (see Definition 6.4), and output $\mu = t \cdot d \bmod pR$.

Lemma 8.1. *If $r > 2n \cdot q^{1/\ell + 2/(n\ell)}$, then the above cryptosystem is IND-CPA secure assuming the hardness of R-$\mathsf{DLWE}_{q,\psi}$ given $\ell + 1$ samples.*

Proof. By Corollary 7.5 (with $k = 1$), the public key $\vec{a}$ is within statistical distance $2^{-\Omega(n)}$ of the uniform distribution over $R_q^{\{1,\ldots,\ell\}}$. By Lemma 2.23 and Lemma 2.24, it follows that for any message μ (chosen adversarially given $\vec{a}$), the ciphertext $\vec{c} = e_0 \cdot \vec{a} + \vec{e}$ is computationally indistinguishable from uniform and independent of the public key, under the hardness assumption. □

Lemma 8.2. *Suppose that for any $c \in R_p^\vee$, $\lfloor p \cdot \psi \rceil_{c+pR^\vee}$ is δ-subgaussian with parameter s for some $\delta = O(1/\ell)$, and $q \geq s\sqrt{(r^2\ell + 1)n} \cdot \omega(\sqrt{\log n})$. Then decryption is correct with probability $1 - \mathrm{negl}(n)$ over all the randomness of key generation and encryption.*

In particular, if ψ is a continuous Gaussian with parameter $s' \geq 1$, and we use coordinate-wise randomized rounding in the decoding basis for discretization, then by the discussion in Section 2.4.2 and the equality $s_1(\vec{d}) = \sqrt{\mathrm{rad}(m)/m}$ from Lemma 6.2, we have that $\lfloor p \cdot \psi \rceil_{c+pR^\vee}$ is 0-subgaussian with parameter $s = p\sqrt{s'^2 + 2\pi\,\mathrm{rad}(m)/m} = O(ps')$.

Proof. By construction, $\langle \vec{c}, \vec{x} \rangle = e_0 x_0 + \langle \vec{e}, \vec{x} \rangle = \langle \vec{e}', \vec{x}' \rangle \bmod qR^\vee$, where $\vec{e}' = (e_0, e_1, \ldots, e_\ell) \in (R^\vee)^{[\ell+1]}$ and $\vec{x}' = (x_0, x_1, \ldots, x_\ell = 1) \in R^{[\ell+1]}$. Furthermore, $\langle \vec{e}', \vec{x}' \rangle = t^{-1}\mu \bmod pR^\vee$, so decryption is correct as long as $[\![\langle \vec{e}', \vec{x}' \rangle \bmod qR^\vee]\!] = \langle \vec{e}', \vec{x}' \rangle \in R^\vee$. We next show that this holds with probability $1 - \mathrm{negl}(n)$ over the choice of $\vec{e}', \vec{x}'$.

By Lemma 2.8, for each $i \in [\ell]$ we have $\|x_i\|_2 \leq r\sqrt{n}$ except with probability at most $2^{-n} = \mathrm{negl}(n)$, and $\|x_\ell\|_2 = \|1\|_2 = \sqrt{n}$. Then by Item 6.6 of Lemma 6.6 (with $k = 1$, $\ell = 0$), for every $i \in [\ell]$ each coefficient of $e_i x_i$ when represented in the decoding basis is δ-subgaussian with parameter $sr\sqrt{n}$, and each one of $e_\ell x_\ell$ is δ-subgaussian with parameter $s\sqrt{n}$. Since the e_i are mutually independent, each decoding-basis coefficient of $\langle \vec{e}', \vec{x}' \rangle$ is $\delta(\ell + 1)$-subgaussian with parameter $s\sqrt{(r^2\ell + 1)n}$. Since $\delta(\ell + 1) = O(1)$, the claim follows by Lemma 6.5. □

8.2 Compact Public-Key Cryptosystem

As in the previous subsection, let R denote the mth cyclotomic ring and let p, q be coprime integers, where the message space is R_p. We also require q to be coprime with every odd prime dividing m. Also let ψ be a continuous LWE error distribution over $K_\mathbb{R}$, and let $\lfloor \cdot \rceil$ denote a valid discretization to (cosets of) $R^\vee$ or $pR^\vee$. The cryptosystem is defined as follows.

- Gen: choose a uniformly random $a \leftarrow R_q$. Choose $x \leftarrow \lfloor \psi \rceil_{R^\vee}$ and $e \leftarrow \lfloor p \cdot \psi \rceil_{pR^\vee}$.
  Output $(a, b = \hat{m}(a \cdot x + e) \bmod qR) \in R_q \times R_q$ as the public key, and x as the secret key.
  (Note that because $\hat{m} = t \cdot g$, $R^\vee = \langle t^{-1} \rangle$, and $a \cdot x + e \in R^\vee / qR^\vee$, we have $\hat{m}(a \cdot x + e) \in gR/gqR$, which is then reduced mod qR to obtain $b \in R_q$.)

- $\mathsf{Enc}_{(a,b)}(\mu \in R_p)$: choose $z \leftarrow \lfloor \psi \rceil_{R^\vee}$, $e' \leftarrow \lfloor p \cdot \psi \rceil_{pR^\vee}$, and $e'' \leftarrow \lfloor p \cdot \psi \rceil_{t^{-1}\mu + pR^\vee}$.
  Let $u = \hat{m}(z \cdot a + e') \bmod qR$ and $v = z \cdot b + e'' \in R_q^\vee$. Output $(u, v) \in R_q \times R_q^\vee$.

- $\mathsf{Dec}_x(u, v)$: compute $v - u \cdot x = \hat{m}(e \cdot z - e' \cdot x) + e'' \bmod qR^\vee$, and decode it to $d = [\![v - u \cdot x]\!] \in R^\vee$ (see Definition 6.4). Output $\mu = t \cdot d \bmod pR$.

Lemma 8.3. *The above cryptosystem is IND-CPA secure assuming the hardness of R-$\mathsf{DLWE}_{q,\psi}$.*

Proof. The security proof follows from two applications of the ring-LWE assumption in its normal form (see Lemma 2.24), with secret drawn from $\lfloor \psi \rceil_{R^\vee}$. First, we claim that the public key is indistinguishable from uniform. Using the transformation from Lemma 2.23 with $w = 0$, we see that the pair $(a, a \cdot x + e) \in R_q \times R_q^\vee$, where a, x, e are sampled as in the Gen procedure, is indistinguishable from uniform. Now consider the transformation that multiplies the second component by $\hat{m}$ and reduces the result modulo qR. This transformation maps pairs $(a, a \cdot x + e)$ distributed as before, to pairs in $R_q \times R_q$ distributed as the output of the Gen procedure. Moreover, since $\langle g \rangle$ and $\langle q \rangle$ are coprime by Corollary 2.18, and recalling that $\hat{m} R^\vee = gR$, we see that this transformation maps the uniform distribution over $R_q \times R_q^\vee$ to the uniform distribution over $R_q \times R_q$. This completes the proof of the first claim.

It remains to show that if the public key (a, b) is uniformly random in $R_q \times R_q$, then for any message $\mu \in R_p$, the joint distribution of the public key together with $\mathsf{Enc}_{(a,b)}(\mu)$ is computationally indistinguishable from uniform. To see this, consider a reduction that is given access to a distribution over $R_q \times K_{\mathbb{R}}/qR^\vee$ which is either $A_{z,\psi}$ (for $z \leftarrow \lfloor \psi \rceil_{R^\vee}$) or uniform. It obtains two samples (a', u'') and (b', v') from the distribution, and applies the transformation from Lemma 2.23 with $w = 0$ to (a', u'') to obtain (a, u'), and with $w = t^{-1}\mu \in R_p^\vee$ to (b', v') to obtain (b, v). The reduction then outputs (a, b) as the public key, and $(u = \hat{m} u' \bmod qR, v) \in R_q \times R_q^\vee$ as the encryption of μ.

If the unknown distribution was uniform, then it follows that (a, b, u, v) is uniform in $R_q^{[3]} \times R_q^\vee$. (Showing that u is uniform in R_q is done as above, in the proof of the first claim.) On the other hand, if the unknown distribution is $A_{z,\psi}$, then (a, b) has uniform distribution, and it can be verified that (u, v) has the same distribution as generated by $\mathsf{Enc}_{(a,b)}(\mu)$. This completes the proof. □

We finally show that under suitable parameters, decryption is correct with overwhelming probability.

Lemma 8.4. *Suppose that $\lfloor \psi \rceil_{R^\vee}$ outputs elements having ℓ_2 norm bounded by ℓ with $1 - \mathrm{negl}(n)$ probability, that $\lfloor p \cdot \psi \rceil_{c+pR^\vee}$ (for any coset $c + pR^\vee$) is δ-subgaussian with parameter s for some $\delta = O(1)$, and that $q \geq s\sqrt{2(\hat{m}\ell)^2 + n} \cdot \omega(\sqrt{\log n})$. Then decryption is correct with probability $1 - \mathrm{negl}(n)$ over all the randomness of key generation and encryption.*

In particular, and just as in the previous subsection, if ψ is a continuous Gaussian with parameter $s' \geq 1$, and we use coordinate-wise randomized rounding in the decoding basis for discretization, then $\lfloor p \cdot \psi \rceil_{c+pR^\vee}$ is 0-subgaussian with parameter $s = p\sqrt{s'^2 + 2\pi \operatorname{rad}(m)/m} = O(ps')$. Moreover, by the fact that ψ has $1 - 2^{-\Omega(n)}$ of its mass on vectors of length at most $s'\sqrt{n}$, and because discretization increases lengths by at most $s_1(\vec{d})\sqrt{n}$ (by the triangle inequality), we have that $\lfloor \psi \rceil_{R^\vee}$ outputs elements having norm bounded by $\ell := (s' + \sqrt{\operatorname{rad}(m)/m})\sqrt{n} = O(s'\sqrt{n})$, except with $\mathrm{negl}(n)$ probability.

Proof. By construction, $e, e' \in pR^\vee$ and $x, z \in R^\vee$, so $\hat{m}(e \cdot z - e' \cdot x) \in pR^\vee$. Therefore, $E := \hat{m}(e \cdot z - e' \cdot x) + e'' \in R^\vee$ satisfies $E = \mu \bmod pR^\vee$ when e'' is chosen as when encrypting μ, so decryption is correct as long as $[\![E \bmod qR^\vee]\!] = E$. We next show that this holds with probability $1 - \mathrm{negl}(n)$.

By assumption, $\|x\|_2, \|z\|_2 \leq \ell$ with probability $1 - \mathrm{negl}(n)$, and e, e', and e'' are δ-subgaussian with parameter s. Then by Item 2 of Lemma 6.6 (with $k = 1, \ell = 0$), each coefficient of $\hat{m} \cdot ez, \hat{m} \cdot e'x \in R^\vee$ when represented in the decoding basis is δ-subgaussian with parameter $s\hat{m}\ell$, and those of e'' are δ-subgaussian with parameter $s\sqrt{n}$. Since e, e', e'' are mutually independent, each decoding-basis coefficient of E is 3δ-subgaussian with parameter $s\sqrt{2(\hat{m}\ell)^2 + n}$. The claim follows by Lemma 6.5. □

8.3 Symmetric-Key Homomorphic Cryptosystem

Here we define a symmetric-key cryptosystem that is "somewhat homomorphic," i.e., it supports limited additive and multiplicative homomorphic operations. It is essentially the Brakerski-Vaikuntanathan system [BV11b] based on ring-LWE, but with improved parameters and generalized to arbitrary cyclotomics, which introduces several technical challenges. We also describe generalized "key switching" (also known as degree reduction) and "modulus reduction" procedures akin to those first described for standard LWE in [BV11a], and for ring-LWE in power-of-2 cyclotomics in [BGV12]. (The techniques developed here can also be adapted to work with the "scale free" perspective adopted in [Bra12].) The scheme can also be made to support unbounded homomorphic operations using Gentry's "bootstrapping" technique [Gen09b, Gen09a], and also can be efficiently adapted to a public-key system using the regularity lemma from Section 7.

Description of the scheme. Let R denote the mth cyclotomic ring (of degree $n = \varphi(m)$) and let p and q be coprime integers, where p defines the message space R_p and q is the ring-LWE modulus. To support "degree reduction" (see Section 8.3.2 below), we also require $\langle p \rangle, \langle g \rangle \subseteq R$ to be coprime ideals, which is the case if and only if p is coprime with all odd primes dividing m (see Corollary 2.18).

The secret key is a ring element $s \in R$ chosen from a certain distribution (specifically, t times the LWE error distribution over $R^\vee$; see below). We say that a ciphertext of *degree* $k \geq 1$ is a polynomial $c = c(S)$ of degree at most (and usually equal to) k in an indeterminate S, having coefficients in $\mathcal{I}_q$ where $\mathcal{I} = (R^\vee)^k$. (Fresh ciphertexts produced by the encryption algorithm will have degree $k = 1$, whereas those produced by the homomorphic operations may have larger degree.) A ciphertext $c(S)$ encrypting a message $\mu \in R_p$ under secret key $s \in R$ satisfies the relation

$$c(s) = e \bmod q\mathcal{I}$$

for some sufficiently "short" $e \in \mathcal{I}$ such that $e = t^{-k} \cdot \mu \bmod p\mathcal{I}$ (where "short" can refer to the ℓ_2 norm, ℓ_∞ norm, or subgaussian parameter as needed). Therefore, given the secret key $s \in R$ one can compute $e = [\![c(s)]\!] \in \mathcal{I}$ and recover the message as $t^k \cdot e \bmod pR$. We refer to e as the "noise" in the ciphertext, and its subgaussian parameter or ℓ_2 norm determines the size of q needed to ensure correct decryption with high probability, and the underlying hardness assumption. For each operation supported by the system, we give (nearly) tight bounds on the growth or shrinkage of the noise's subgaussian parameter and ℓ_2 norm; these bounds can be combined in a modular way to calculate appropriate parameters for a particular application.

Throughout this subsection, let ψ be a continuous LWE error distribution over $K_\mathbb{R}$, and let $\lfloor \cdot \rceil$ denote any valid discretization to cosets of some scaling of $R^\vee$ (e.g., using the decoding basis $\vec{d}$ of $R^\vee$). The cryptosystem is defined formally as follows.

- Gen: choose $s' \leftarrow \lfloor \psi \rceil_{R^\vee}$, and output $s = t \cdot s' \in R$ as the secret key.
- $\mathsf{Enc}_s(\mu \in R_p)$: choose $e \leftarrow \lfloor p \cdot \psi \rceil_{t^{-1}\mu + pR^\vee}$. Let $c_0 = -c_1 \cdot s + e \in R_q^\vee$ for uniformly random $c_1 \leftarrow R_q^\vee$, and output the ciphertext $c(S) = c_0 + c_1 S$. The "noise" in $c(S)$ is defined to be e.
- $\mathsf{Dec}_s(c(S))$ for c of degree k: compute $c(s) \in (R^\vee)_q^k$, and decode it to $e = [\![c(s)]\!] \in (R^\vee)^k$. Output $\mu = t^k \cdot e \bmod pR$.

The homomorphic operations are defined as follows. For ciphertexts c, c' of arbitrary degrees k, k' (respectively), their homomorphic product is the degree-$(k + k')$ ciphertext $c(S) \boxdot c'(S) = c(S) \cdot c'(S)$ (i.e., standard polynomial multiplication). The noise in the result is defined to be the product of the noise terms of c, c'. Similarly, for ciphertexts c, c' of *equal* degree k, their homomorphic sum is defined as the degree-k ciphertext $c(S) \boxplus c'(S) = c(S) + c'(S)$, and the noise in the resulting ciphertext is the sum of those of c, c'.

(Observe that any degree-k ciphertext resulting from these operations has coefficients in $(R^\vee)_q^k$, as required.) To homomorphically add two ciphertexts of different degrees, we must first homomorphically multiply the one having smaller degree by a fixed public encryption of $1 \in R_p$ enough times to match the larger degree.[14]

It is easy to verify that if the noise terms in all the ciphertexts are correctly decoded by the decryption algorithm, then its output is correct:

$$\begin{aligned}\mathsf{Dec}_s(\mathsf{Enc}_s(\mu)) &= \mu,\\ \mathsf{Dec}_s(c \boxplus c') &= \mathsf{Dec}_s(c) + \mathsf{Dec}_s(c') \bmod pR,\\ \mathsf{Dec}_s(c \boxdot c') &= \mathsf{Dec}_s(c) \cdot \mathsf{Dec}_s(c') \bmod pR.\end{aligned}$$

The following lemma gives a sufficient condition for correct decoding to occur, and follows directly from Lemmas 6.5 and 6.6.

Lemma 8.5. *Suppose the noise e in a degree-k ciphertext c is δ-subgaussian with parameter r for some $\delta = O(1)$, and $q \geq r \cdot \hat{m}^{k-1}\sqrt{n} \cdot \omega(\sqrt{\log n})$. Then $\mathsf{Dec}_s(c)$ correctly recovers e with probability $1 - \mathrm{negl}(n)$. Alternatively, if $q > 2\|e\|_2 \hat{m}^{k-1}\sqrt{n}$, then $\mathsf{Dec}_s(c)$ recovers e with certainty.*

The next two lemmas give (nearly) tight bounds on the subgaussian parameter of the noise under the homomorphic operations. They follow directly from the definition of the noise term, the properties of subgaussian random variables (described in Section 2.3), and the triangle inequality.

Lemma 8.6. *If the noise terms in ciphertexts c_i are independent and δ_i-subgaussian with parameters r_i (respectively), then the noise in the ciphertext $\boxplus_i c_i$ is $(\sum_i \delta_i)$-subgaussian with parameter $(\sum_i r_i^2)^{1/2}$. Moreover, it is always the case that the ℓ_2 and ℓ_∞ norms of the noise terms in $\boxplus_i c_i$ are at most the sums of those in the c_i.*

Lemma 8.7. *Let e, e' be the noise terms in ciphertexts c, c', respectively. Then the noise $e \cdot e'$ in the ciphertext $c \boxdot c'$ satisfies $\|e \cdot e'\| \leq \|e\| \cdot \|e'\|_\infty$, where $\|\cdot\|$ denotes either the ℓ_2 or ℓ_∞ norm. Moreover, if e is δ-subgaussian with parameter r, then the noise $e \cdot e'$ is δ-subgaussian with parameter $r \cdot \|e'\|_\infty$. In particular, if e' is δ-subgaussian with parameter r' and is independent of e, then $e \cdot e'$ is within $\mathrm{negl}(n)$ statistical distance of a δ-subgaussian with parameter $r \cdot r' \cdot \omega(\sqrt{\log n})$.*

Proof. The first claim follows directly from Equation (2.5), and the second one by the first part of Claim 2.4. For the last claim, by subgaussianity we have $\|e'\|_\infty \leq r' \cdot \omega(\sqrt{\log n})$, except with $\mathrm{negl}(n)$ probability. □

Lemma 8.8. *The above cryptosystem is IND-CPA secure assuming the hardness of R-$\mathsf{DLWE}_{q,\psi}$.*

Proof. We describe a reduction that is given access to either an LWE distribution $A_{s',\psi}$ or the uniform distribution over $R_q \times K_\mathbb{R}/qR^\vee$. In the former case we can assume that the distribution is in normal form, i.e., the secret $s' \in R^\vee$ is distributed according to $\lfloor \psi \rceil_{R^\vee}$ (see Lemma 2.24). The reduction simulates an encryption oracle that in the former case implements the encryption algorithm Enc_s for secret key $s = t \cdot s' \in R$ (which is distributed according to the output of Gen), and in the latter case simply returns ciphertexts that are uniformly random and independent of the queried messages. This suffices to prove IND-CPA security.

To respond to an encryption query on message $\mu \in R_p$, the reduction draws a sample $(a', b') \in R_q \times K_\mathbb{R}/qR^\vee$ from the unknown distribution. It then applies the transformation from Lemma 2.23 with

[14] In particular, we can just multiply $c(S)$ by (an appropriate power of) $t^{-1} = g/\hat{m} \in R^\vee$. By definition of g, this element has ℓ_∞ norm $\|t^{-1}\|_\infty \leq 2^\ell/\hat{m} \leq 1$, where ℓ is the number of odd primes dividing m, so multiplication by t^{-1} does not increase the ℓ_2 norm or subgaussian parameter of the noise.

$w = t^{-1}\mu \in R_p^\vee$ to obtain $(a, b) \in R_q \times R_q^\vee$. It lets $c_1 = -t^{-1}a \in R_q^\vee$ and $c_0 = b$, and outputs the ciphertext $c(S) = c_0 + c_1 S$.

Suppose that the unknown distribution is the ring-LWE distribution $A_{s',\psi}$ for $s' \in R^\vee$, and let $s = t \cdot s' \in R$. Then by Lemma 2.23, the pair (a, b) is such that a is uniformly random in R_q, and $b = a \cdot s' + e = (t^{-1}a)s + e \bmod qR^\vee$, where $e \leftarrow \lfloor p \cdot \psi \rceil_{t^{-1}\mu + pR^\vee}$. Therefore, $c_1 = -t^{-1}a$ is uniformly random in $R_q^\vee$, and $c_0 = b = -c_1 \cdot s + e$, so $c(S)$ is distributed exactly according to $\mathsf{Enc}_s(\mu)$.

On the other hand, if the unknown distribution is the uniform distribution, then by Lemma 2.23 the pair (a, b) is uniformly random and independent of μ, and therefore so are the coefficients of the ciphertext $c(S)$.□

8.3.1 Modulus Reduction

The modulus reduction procedure changes the ciphertext modulus from q to some $q' < q$ (where q' is coprime with p), and outputs a ciphertext that encrypts (essentially) the same message, and whose noise term shrinks nearly proportionately. The procedure works best and is simplest to describe in the case of degree-1 ciphertexts, which can always be obtained via the key switching procedure described below in Section 8.3.2.

The following operation is central to the modulus reduction procedure. Let $\mathcal{J}$ be an ideal and let q, q', p be integers with both q and q' coprime to p. Let $v \in \mathbb{Z}_p$ be $v = q' \cdot q^{-1} \bmod p$. Define a randomized function $F_\mathcal{J} : \mathcal{J}_q \to K$ in the following way: given $x \in \mathcal{J}_q$ and some good basis of $\mathcal{J}$, sample a short (subgaussian) element from the coset $(v - q'/q) \cdot x + p\mathcal{J}$ using one of the valid methods described in Section 2.4.2, and let $F_\mathcal{J}(x)$ be the result. Note that the coset $(v - q'/q) \cdot x + p\mathcal{J}$ is well defined because $(v - q'/q)(q\mathcal{J}) = (vq - q')\mathcal{J} \subseteq p\mathcal{J}$. Also observe that for all $x \in \mathcal{J}_q$, we have $(q'/q)x + F_\mathcal{J}(x) \in \mathcal{J}_{q'}$ and $qF_\mathcal{J}(x) \in p\mathcal{J}$ with certainty.

We now describe the modulus reduction procedure. Let $c(S) = c_0 + c_1 S$ be an input ciphertext, with $c_0, c_1 \in R_q^\vee$. Let $f_0 \leftarrow F_{R^\vee}(c_0)$ and $f_1 \leftarrow t^{-1} \cdot F_R(t \cdot c_1)$, where we use coordinate-wise randomized rounding with the decoding basis $\vec{d}$ of $R^\vee$ for the former, and with the powerful basis $\vec{p}$ of R for the latter. The output is the ciphertext $c'(S) = c'_0 + c'_1 S$, where

$$c'_0 = \frac{q'}{q} c_0 + f_0 \bmod q'R^\vee, \qquad c'_1 = \frac{q'}{q} c_1 + f_1 \bmod q'R^\vee.$$

Notice that by the first of the above properties, we have $c'_0, c'_1 \in R_{q'}^\vee$ as required. Notice also that if $s = t \cdot s' \in R$ is the secret key and e is the noise in $c(S)$, so that $c_0 + c_1 s = e \bmod qR^\vee$, then

$$c'_0 + c'_1 s = \frac{q'}{q}(c_0 + c_1 s) + (f_0 + f_1 s) = \frac{q'}{q} e + (f_0 + (tf_1) \cdot s') \bmod q'R^\vee. \tag{8.1}$$

Accordingly, we define the noise in the ciphertext $c'(S)$ to be $e' = (q'/q)e + (f_0 + f_1 s)$, which is in $R^\vee$ because $c'_0, c'_1 \in R_{q'}^\vee$.

The following lemma describes the procedure's effect on the noise and plaintext. It says that the error is scaled by a factor of q'/q, plus a modulus-independent amount that depends only on the ℓ_∞ norm of $s' = t^{-1}s \in R^\vee$ (which was chosen from $\lfloor \psi \rceil_{R^\vee}$ and hence is short). It also shows that the procedure implicitly introduces a factor of $v = q' \cdot q^{-1} \in R_p$ into the message, which can be kept track of and removed upon decryption, because q' is coprime with p by assumption. In general, this extra factor seems inherent to modulus reduction, though it can be avoided by always using $q' = q \bmod p$, which always holds in the common case $p = 2$.

Lemma 8.9. *If the noise in the input ciphertext is $e \in R^\vee$, then the noise $e' \in R^\vee$ in the output ciphertext satisfies $e' = q' \cdot q^{-1} \cdot e \bmod pR^\vee$. Moreover, e' equals $(q'/q)e$ plus a random variable f that, for any value of e, is 0-subgaussian with parameter*

$$p\sqrt{2\pi}\Big(\mathrm{rad}(m)/m + \hat{m}\|t^{-1}s\|_\infty^2\Big)^{1/2},$$

and for which $\|f\|_2 \le p\sqrt{n}\Big(\sqrt{\mathrm{rad}(m)/m} + \sqrt{\hat{m}}\|t^{-1}s\|_\infty\Big)$ always.

In particular, if e is δ-subgaussian then by Claim 2.1 so is e', although it may not be independent of e.

Proof. Since both $e, e' \in R^\vee$ and q is coprime with p, showing that $e' = v \cdot e \bmod pR^\vee$ is equivalent to showing that $qe' - q'e \in pR^\vee$. The latter follows immediately from the definition of e' and the fact that $qF_{\mathcal{J}}(x) \in p\mathcal{J}$ always.

The subgaussianity claim on $e' = (q'/q)e + (f_0 + f_1 s)$ follows by the fact that for any values of c_0, c_1, the terms f_0 and tf_1 are 0-subgaussian with respective parameters $p\sqrt{2\pi}s_1(\vec{d})$ and $p\sqrt{2\pi}s_1(\vec{p})$; the bounds on $s_1(\vec{d})$ and $s_1(\vec{p})$ given in Lemmas 6.2 and 4.3 respectively; and Claim 2.1. Similarly, the claim on $\|f\|_2$ follows from the fact that coordinate-wise randomized rounding to a coset of $pR^\vee$ (respectively, pR) using basis $p \cdot \vec{d}$ (resp., $p \cdot \vec{p}$) always yields an element having ℓ_2 norm bounded by $p\sqrt{n}s_1(\vec{d})$ (resp., $p\sqrt{n}s_1(\vec{p})$); by Equation (2.5); and by the triangle inequality. □

8.3.2 Key Switching/Degree Reduction

The key-switching procedure (also known as "degree reduction") converts any degree-k ciphertext $c(S)$ encrypted under a secret key $s \in R$, to a degree-1 ciphertext $c'(S')$ encrypted under a key $s' \in R$ (which may or may not be the same as s). Notice that when decrypting $c(S)$, the evaluation $c(s)$ is simply a linear function in the powers $s^0, s^1, \ldots, s^k \in R$. The main idea behind the key-switching method introduced in [BV11a] is to homomorphically apply this linear function to suitable *encryptions* (under s') of these powers; we refer to these ciphertexts as the key-switching "hint." Implementing this idea requires some care in our setting, however, due to the different powers of $R^\vee$ involved in the operations and their homomorphic counterparts.

Rewriting the decryption relation. Let $\mathcal{I} = (R^\vee)^k$ and $d = k + 1$, let $\vec{s} = (s^0, \ldots, s^k) \in R^{[d]}$, and let $\vec{c} \in \mathcal{I}_q^{[d]}$ be the coefficient vector of a valid degree-k ciphertext $c(S)$. Then for a degree-k ciphertext c, we have the decryption relation

$$\langle \vec{c}, \vec{s} \rangle = e \bmod q\mathcal{I}$$

for some short (subgaussian) $e \in t^{-k}\mu + p\mathcal{I}$. We first put this relation in a more convenient form, viewing the ciphertext in the slightly "denser" quotient $\hat{m}^{1-k}R_q^\vee$ (because $\hat{m}^{1-k}R^\vee \supseteq \mathcal{I}$), and then scaling it up by a $\hat{m}^{k-1}$ factor.[15] We also multiply and divide $\vec{c}$ and $\vec{s}$ (respectively) by t, yielding

$$\langle \underbrace{t \cdot \hat{m}^{k-1}\vec{c}}_{\vec{y} \in R_q^{[d]}}, t^{-1}\vec{s} \rangle = \hat{m}^{k-1}e \bmod qR^\vee.$$

We write the relation in this way so that $t^{-1}\vec{s}$ is over $R^\vee$, which is the appropriate domain for encrypting it in the key-switching hint, and so that $\vec{y}$ is over R_q, which will be needed for decomposing it into short elements of R as part of the key-switching operation.

[15] This is essentially the same idea used in decoding $\mathcal{I}_q$ to $\mathcal{I}$, as described in Section 6.2.

We finally make one more important change to the decryption relation. Let $\ell = \lceil \log_2 q \rceil$ and define

$$\mathbf{g} = (1, 2, 4, \ldots, 2^{\ell-1}) \in \mathbb{Z}_q^{[\ell]} \quad \text{and} \quad G = I_{[d]} \otimes \mathbf{g}^T \in \mathbb{Z}_q^{[d] \times [d\ell]}. \tag{8.2}$$

Then for any $\vec{x} \in R^{[d\ell]}$ such that $G\vec{x} = \vec{y} \in R_q^{[d]}$, we have

$$\langle \vec{x}\,,\, t^{-1}G^T\vec{s} \rangle = \langle G\vec{x}\,,\, t^{-1}\vec{s} \rangle = \hat{m}^{k-1}e \bmod qR^\vee. \tag{8.3}$$

The hint will consist essentially of an encryption of $t^{-1}G^T\vec{s}$, and the key-switching operation will homomorphically compute its inner product with a *short* (subgaussian) $\vec{x}$ so as to keep the error in the resulting ciphertext small. The need for a short $\vec{x}$ is why we arranged for $\vec{y}$ to be over R_q, because we always have a good basis for R (namely, the powerful basis) that has nearly optimal spectral norm $s_1(\vec{p}) = \sqrt{\hat{m}}$, whereas we do not always have such a good basis of $\mathcal{I} = (R^\vee)^k$ for $k \geq 1$.

Alternative relations. As an optimization, we can actually omit the constant term 1 from $\vec{s}$. This decreases the dimension d by one, thereby reducing the size of the hint and the amount of extra noise introduced by the key-switching procedure. For ciphertext $c(S) = \sum_{i=0}^k c_i S^i$ we then define $\vec{c} = (c_1, \ldots, c_k)$, so that the main decryption relation becomes $c_0 + \langle \vec{c}, \vec{s} \rangle = e \bmod q\mathcal{I}$. The hint-generation and key-switching procedures then work exactly as described below, with the additional step that we add the constant term $\hat{m}^{k-1}c_0 \bmod qR^\vee$ to the output ciphertext $c'(S')$. This works because the key-switching procedure ensures that $c'(s') \approx \hat{m}^{k-1}\langle \vec{c}, \vec{s} \rangle = \hat{m}^{k-1}(e - c_0) \bmod qR^\vee$.

Similarly, when the original and target secret keys are equal, i.e., $s' = s$, we can omit both 1 and s from $\vec{s}$, define $\vec{c} = (c_2, \ldots, c_k)$, and write the decryption relation as $(c_0 + c_1 s) + \langle \vec{c}, \vec{s} \rangle = e \bmod q\mathcal{I}$. We can then apply the procedures below, adding the linear polynomial $\hat{m}^{k-1}(c_0 + c_1 S) \bmod qR^\vee$ to the output ciphertext $c'(S)$ of the key-switching procedure.

Finally, the vector $\mathbf{g}$ need not contain only powers of 2, but may be defined with respect to a larger integer base (thereby decreasing the dimension ℓ), or may even consist of other exponentially increasing sequences. The particular choice of $\mathbf{g}$ mainly affects the length (or subgaussian parameter) of the decomposition $\vec{x} \in R^{[d\ell]}$. See [MP12] for further discussion.

Constructing the hint. The hint is a collection of independent degree-1 ciphertexts $h_i(S')$ for each $i \in [d\ell]$, prepared as

$$h_i(S') \leftarrow \mathsf{Enc}_{s'}(0) + t^{-1}(G^T\vec{s})_i \bmod qR^\vee,$$

i.e., we generate degree-1 encryptions of 0 and simply add entries of $t^{-1}G^T\vec{s}$ to their constant terms. Notice that by construction,

$$h_i(s') = f_i + t^{-1}(G^T\vec{s})_i \bmod qR^\vee$$

for some short (subgaussian) $f_i \in pR^\vee$ having distribution $\lfloor p \cdot \psi \rceil_{pR^\vee}$. Note also that $h_i(S')$ may not actually be a well-formed encryption of any particular message, because $h_i(s')$ may not be congruent modulo $qR^\vee$ to any short enough element of $R^\vee$; however, this does not matter for the key-switching application.

To the vector $\vec{f} = (f_i)_{i \in [d\ell]}$ of noise terms in the hint we associate a measure of quality F, defined as

$$F := \max_{i \in \mathbb{Z}_m^*} \Big(\sum_{j=1}^{d\ell} |\sigma_i(f_j)|^2 \Big)^{1/2}, \tag{8.4}$$

and bound it as follows.

Claim 8.10. *If the entries $f_j \in R^\vee$ of $\vec{f}$ are all δ-subgaussian with parameter s for some $\delta = O(1)$, then*

$$F \le Cs \cdot \max(\sqrt{d\ell}, \omega(\sqrt{\log n}))$$

except with $\mathrm{negl}(n)$ *probability, for some universal constant $C > 0$.*

Proof. Write

$$\begin{aligned}\max_{i \in \mathbb{Z}_m^*}\Big(\sum_{j=1}^{d\ell} |\sigma_i(f_j)|^2\Big) &= \max_{i \in \mathbb{Z}_m^*}\Big(\sum_{j=1}^{d\ell} \Re(\sigma_i(f_j))^2 + \sum_{j=1}^{d\ell} \Im(\sigma_i(f_j))^2\Big) \\ &\le 2 \max_{i \in \mathbb{Z}_m^*} \max\Big\{\sum_{j=1}^{d\ell} \Re(\sigma_i(f_j))^2, \sum_{j=1}^{d\ell} \Im(\sigma_i(f_j))^2\Big\}.\end{aligned}$$

Each of the $2n$ sums is a sum of squares of $d\ell$ independent δ-subgaussian variables with parameter $s/\sqrt{2}$. The claim now follows by applying Lemma 2.2 to each of the sums and applying the union bound. □

The key-switching procedure. The procedure takes as input $\vec{c} \in \mathcal{I}_q^{[d]}$, computes $\vec{y} = t \cdot \hat{m}^{k-1}\vec{c} \in R_q^{[d]}$, and generates, as described below, a short (subgaussian) $\vec{x} \in R^{[d\ell]}$ such that $G\vec{x} = \vec{y}$. It then outputs the degree-1 ciphertext

$$c'(S') = \sum_{i \in [d\ell]} x_i \cdot h_i(S').$$

Notice that by (8.3), evaluating $c'(S')$ at $S' = s'$ gives

$$c'(s') = \sum_{i \in [d\ell]} x_i(f_i + t^{-1}(G^T\vec{s})_i) = \langle \vec{x}, \vec{f}\rangle + \langle \vec{x}, t^{-1}G^T\vec{s}\rangle = \langle \vec{x}, \vec{f}\rangle + \hat{m}^{k-1}e \bmod qR^\vee.$$

Accordingly, we define the noise term in c' to be $e' = \langle \vec{x}, \vec{f}\rangle + \hat{m}^{k-1}e \in R^\vee$. Notice that the noise is congruent to $\hat{m}^{k-1}e$ modulo $pR^\vee$, because each $f_i \in pR^\vee$ by construction of the hint. The noise is also relatively short: the $\hat{m}^{k-1}$ factor of e is exactly offset by switching from modulus $q\mathcal{I} = q(R^\vee)^k$ to $qR^\vee$, and $\langle \vec{x}, \vec{f}\rangle$ is short because both $\vec{x}$ and $\vec{f}$ are. (See Lemma 8.11 for a precise analysis.)

Also note that while decrypting the original ciphertext $c(S)$ would yield the message $t^k e = \mu \bmod pR$, the resulting degree-1 ciphertext $c'(S')$ decrypts to the message $t \cdot \hat{m}^{k-1}e = g^{k-1}\mu \bmod pR$. This means that an implementation must keep track of the "true" underlying degree of each ciphertext (and limit homomorphic additions to ciphertexts of equal "true" degree), even if its degree as a polynomial has been reduced via key switching. Upon final decryption, the extra g^{k-1} factor in the message can be removed as long as g is invertible modulo p, which by Corollary 2.18 is the case because we have assumed that p is coprime with every odd prime dividing m.

The next lemma says that the key-switching procedure introduces into the ciphertext some subgaussian error, proportional to the quality F of the noise vector $\vec{f}$ in the hint.

Lemma 8.11. *Fix an arbitrary vector $\vec{f}$ and let F be as defined in Equation* (8.4)*. Assume that for some $\delta = O(1)$, every entry $x_j \in R$ of $\vec{x}$ is δ-subgaussian with parameter s', conditioned on any values of the ciphertext c and $x_1, \ldots, x_{j-1}$. Then for any value of the original noise term e, the additional noise term $\langle \vec{x}, \vec{f}\rangle$ is $(d\ell)\delta$-subgaussian with parameter Fs'. In particular, if e is δ-subgaussian with parameter s'' then the new noise term $e' = \langle \vec{x}, \vec{f}\rangle + \hat{m}^{k-1}e$ is $(d\ell + 1)\delta$-subgaussian with parameter $\sqrt{(Fs')^2 + (\hat{m}^{k-1}s'')^2}$.*

Proof. The subgaussianity claim on $\langle \vec{x}, \vec{f}\rangle$ follows directly from Claim 2.4. The claim on e' is immediate by Claim 2.1. □

Choosing $\vec{x}$. In the key-switching procedure we need to sample a subgaussian $\vec{x} \in R^{[d\ell]}$ such that $G\vec{x} = \vec{y}$ for a given $\vec{y} \in R_q^{[d]}$, where $G \in \mathbb{Z}_q^{[d]\times[d\ell]}$ is as defined in Equation (8.2). To start, define the R-module

$$\Lambda^\perp(G) = \{\vec{z} \in R^{[d\ell]} : G\vec{z} = \vec{0} \in R_q^{[d]}\},$$

which may also be seen as a lattice in $H^{[d\ell]}$. The set of all solutions $\vec{x}$ to $G\vec{x} = \vec{y}$ is then a coset of this module. (A solution always exists, because G contains the identity $I_{[d]}$ as a submatrix.) Given a high-quality $\mathbb{Z}$-basis of $\Lambda^\perp(G)$, we can use any of the methods described in Section 2.4.2 for subgaussian sampling over the desired lattice coset, e.g., coordinate-wise randomized rounding. As usual, the relevant measures of basis quality are the largest singular value and the maximum length of the Gram-Schmidt orthogonalized vectors.

Lemma 8.12. *There is an efficiently computable $\mathbb{Z}$-basis $Z \in R^{[d\ell]\times[d\ell n]}$ of $\Lambda^\perp(G)$ satisfying the following bounds, where $\|\widetilde{Z}\|_2$ denotes the largest ℓ_2 norm of the Gram-Schmidt orthogonalized vectors $\widetilde{Z}$. If q is a power of 2, then $s_1(Z) \le 3\sqrt{\hat{m}}$ and $\|\widetilde{Z}\|_2 = 2\sqrt{n}$; otherwise, $s_1(Z) \le \sqrt{(9 + \mathrm{wt}_2(q))\hat{m}}$ and $\|\widetilde{Z}\|_2 = \sqrt{5n}$, where $\mathrm{wt}_2(q)$ denotes the number of 1s in the binary expansion of q.*

The remainder of this section is dedicated to a proof of the lemma. We first recall that Micciancio and Peikert [MP12, Section 4] constructed good bases for the *integer* lattices

$$\mathcal{L}^\perp(G) = \{\mathbf{z} \in \mathbb{Z}^{[d\ell]} : G\mathbf{z} = \mathbf{0} \in \mathbb{Z}_q^{[d]}\},$$

which we now briefly summarize (see that work for further details and full proofs). Recalling that $G = I_{[d]} \otimes \mathbf{g}^T \in \mathbb{Z}_q^{[d]\times[d\ell]}$ where $\mathbf{g} = (1, 2, 4, \ldots, 2^{\ell-1}) \in \mathbb{Z}_q^{[\ell]}$ and $\ell = \lceil \log_2 q \rceil$, define $S_g \in \mathbb{Z}^{[\ell]\times[\ell]}$ as

$$S_g = \begin{pmatrix} 2 & & & & \\ -1 & 2 & & & \\ & -1 & \ddots & & \\ & & & 2 & \\ & & & -1 & 2 \end{pmatrix} \text{ if } q = 2^\ell, \text{ otherwise } S_g = \begin{pmatrix} 2 & & & & q_0 \\ -1 & 2 & & & q_1 \\ & -1 & & & q_2 \\ & & \ddots & & \vdots \\ & & & 2 & q_{\ell-2} \\ & & & -1 & q_{\ell-1} \end{pmatrix},$$

where $q = (q_{\ell-1} \cdots q_1 q_0)_2 = \sum_{i\in[\ell]} q_i 2^i$ for $q_i \in \{0, 1\}$ is the binary representation of q. It is clear by inspection that the columns of S_g are all in the lattice $\mathcal{L}^\perp(\mathbf{g}^T)$; moreover, as shown in [MP12], they are indeed a basis of the lattice. (This can be seen by verifying that the determinants of S_g and $\mathcal{L}^\perp(\mathbf{g}^T)$ are equal.) It immediately follows that $S = I_{[d]} \otimes S_g \in \mathbb{Z}^{[d\ell]\times[d\ell]}$ is a basis for the lattice $\mathcal{L}^\perp(G)$.

In [MP12] it is shown that $\|\widetilde{S}\|_2 = \|\widetilde{S_g}\|_2 = 2$ if $q = 2^\ell$ (where we orthogonalize from right to left), and is $\sqrt{5}$ otherwise (where we orthogonalize from left to right). It also directly follows from the triangle inequality and Pythagorean theorem that $s_1(S) = s_1(S_g) \le 3$ if $q = 2^\ell$, and is at most $\sqrt{9 + \mathrm{wt}_2(q)}$ otherwise.

We now claim that

$$Z = S \otimes \vec{p}^T = I_{[d]} \otimes S_g \otimes \vec{p}^T \in R^{[d\ell]\times[d\ell n]}$$

is a $\mathbb{Z}$-basis of $\Lambda^\perp(G)$ satisfying the bounds in Lemma 8.12, where $\vec{p}$ is the powerful basis of R. For the bounds, observe that by Lemma 4.3, the fact that the longest Gram-Schmidt orthogonalized vector of $\sigma(\vec{p}^T) = \mathrm{CRT}_m$ has length $\sqrt{n}$, and the properties of singular values and orthogonalization under the tensor product, we have

$$s_1(Z) = s_1(S) \cdot s_1(\vec{p}) = s_1(S) \cdot \sqrt{\hat{m}} \quad \text{and} \quad \|\widetilde{Z}\|_2 = \|\widetilde{S}\|_2 \cdot \|\widetilde{\mathrm{CRT}_m}\|_2 = \|\widetilde{S}\|_2 \cdot \sqrt{n},$$

which when combined with the above bounds from [MP12] yields the claim. It only remains to show that Z is a $\mathbb{Z}$-basis of $\Lambda^\perp(G)$, which is a consequence of the following simple lemma.

Lemma 8.13. *Let $A \in \mathbb{Z}_q^{[h]\times[k]}$ for some $h, k \geq 1$ be arbitrary. If B is any $\mathbb{Z}$-basis of $\mathcal{L}^\perp(A) \subseteq \mathbb{Z}^{[k]}$ and $\vec{b}$ is any $\mathbb{Z}$-basis of R, then $B \otimes \vec{b}^T$ is a $\mathbb{Z}$-basis of $\Lambda^\perp(A) \subseteq R^{[k]}$.*

Proof. Clearly, every element of $B \otimes \vec{b}^T$ is in $\Lambda^\perp(A)$. To show that it is a basis, let $\vec{z} \in \Lambda^\perp(A)$ be arbitrary, so $A\vec{z} = \vec{0} \in R_q^{[h]}$. Then we can uniquely write $\vec{z} = \sum_j b_j \cdot \mathbf{z}_j$ for some vectors $\mathbf{z}_j \in \mathbb{Z}^{[k]}$. By linearity and uniqueness with respect to $\vec{b}$, this implies that $A\mathbf{z}_j = \mathbf{0} \in \mathbb{Z}_q^{[h]}$ for every j, so each $\mathbf{z}_j \in \mathcal{L}^\perp(A)$ can be written uniquely as a $\mathbb{Z}$-linear combination of elements in B. It follows that $\vec{z}$ can be expressed uniquely as a $\mathbb{Z}$-linear combination of elements in $B \otimes \vec{b}^T$. □

References

[ABB10] S. Agrawal, D. Boneh, and X. Boyen. Efficient lattice (H)IBE in the standard model. In *EUROCRYPT*, pages 553–572. 2010.

[ACPS09] B. Applebaum, D. Cash, C. Peikert, and A. Sahai. Fast cryptographic primitives and circular-secure encryption based on hard learning problems. In *CRYPTO*, pages 595–618. 2009.

[AP09] J. Alwen and C. Peikert. Generating shorter bases for hard random lattices. *Theory of Computing Systems*, 48(3):535–553, April 2011. Preliminary version in STACS 2009.

[AP13] J. Alperin-Sheriff and C. Peikert. Practical bootstrapping with polylogarithmic overhead. In *CRYPTO*. 2013. To appear.

[Bab85] L. Babai. On Lovász' lattice reduction and the nearest lattice point problem. *Combinatorica*, 6(1):1–13, 1986. Preliminary version in STACS 1985.

[Ban93] W. Banaszczyk. New bounds in some transference theorems in the geometry of numbers. *Mathematische Annalen*, 296(4):625–635, 1993.

[BGV12] Z. Brakerski, C. Gentry, and V. Vaikuntanathan. (Leveled) fully homomorphic encryption without bootstrapping. In *ICTS*, pages 309–325. 2012.

[Bos90] W. Bosma. Canonical bases for cyclotomic fields. *Appl. Algebra Eng. Commun. Comput.*, 1:125–134, 1990.

[Boy10] X. Boyen. Lattice mixing and vanishing trapdoors: A framework for fully secure short signatures and more. In *Public Key Cryptography*, pages 499–517. 2010.

[BPR12] A. Banerjee, C. Peikert, and A. Rosen. Pseudorandom functions and lattices. In *EUROCRYPT*, pages 719–737. 2012.

[Bra12] Z. Brakerski. Fully homomorphic encryption without modulus switching from classical GapSVP. In *CRYPTO*, pages 868–886. 2012.

[BV11a] Z. Brakerski and V. Vaikuntanathan. Efficient fully homomorphic encryption from (standard) LWE. In *FOCS*, pages 97–106. 2011.

[BV11b] Z. Brakerski and V. Vaikuntanathan. Fully homomorphic encryption from ring-LWE and security for key dependent messages. In *CRYPTO*, pages 505–524. 2011.

[CHKP10] D. Cash, D. Hofheinz, E. Kiltz, and C. Peikert. Bonsai trees, or how to delegate a lattice basis. *J. Cryptology*, 25(4):601–639, 2012. Preliminary version in Eurocrypt 2010.

[Con09] K. Conrad. The different ideal, 2009. Available at `http://www.math.uconn.edu/~kconrad/blurbs/`, last accessed 12 Oct 2009.

[DD12] L. Ducas and A. Durmus. Ring-LWE in polynomial rings. In *Public Key Cryptography*, pages 34–51. 2012.

[Erd46] P. Erdős. On the coefficients of the cyclotomic polynomial. *Bulletin of the American Mathematical Society*, 52(2):179–184, 1946.

[Gen09a] C. Gentry. *A fully homomorphic encryption scheme*. Ph.D. thesis, Stanford University, 2009. `crypto.stanford.edu/craig`.

[Gen09b] C. Gentry. Fully homomorphic encryption using ideal lattices. In *STOC*, pages 169–178. 2009.

[Gen10] C. Gentry. Toward basing fully homomorphic encryption on worst-case hardness. In *CRYPTO*, pages 116–137. 2010.

[GHPS12] C. Gentry, S. Halevi, C. Peikert, and N. P. Smart. Ring switching in BGV-style homomorphic encryption. In *SCN*, pages 19–37. 2012. Full version at `http://eprint.iacr.org/2012/240`.

[GHS12a] C. Gentry, S. Halevi, and N. P. Smart. Fully homomorphic encryption with polylog overhead. In *EUROCRYPT*, pages 465–482. 2012.

[GHS12b] C. Gentry, S. Halevi, and N. P. Smart. Homomorphic evaluation of the AES circuit. In *CRYPTO*, pages 850–867. 2012.

[GLP12] T. Güneysu, V. Lyubashevsky, and T. Pöppelmann. Practical lattice-based cryptography: A signature scheme for embedded systems. In *CHES*, pages 530–547. 2012.

[GPV08] C. Gentry, C. Peikert, and V. Vaikuntanathan. Trapdoors for hard lattices and new cryptographic constructions. In *STOC*, pages 197–206. 2008.

[Kle00] P. N. Klein. Finding the closest lattice vector when it's unusually close. In *SODA*, pages 937–941. 2000.

[Lan94] S. Lang. *Algebraic number theory*, volume 110 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1994.

[LM06] V. Lyubashevsky and D. Micciancio. Generalized compact knapsacks are collision resistant. In *ICALP (2)*, pages 144–155. 2006.

[LM08] V. Lyubashevsky and D. Micciancio. Asymptotically efficient lattice-based digital signatures. In *TCC*, pages 37–54. 2008.

[LMPR08] V. Lyubashevsky, D. Micciancio, C. Peikert, and A. Rosen. SWIFFT: A modest proposal for FFT hashing. In *FSE*, pages 54–72. 2008.

[LP11] R. Lindner and C. Peikert. Better key sizes (and attacks) for LWE-based encryption. In *CT-RSA*, pages 319–339. 2011.

[LPR10] V. Lyubashevsky, C. Peikert, and O. Regev. On ideal lattices and learning with errors over rings. *J. ACM*, 2013. To appear. Preliminary version in Eurocrypt 2010.

[LPS10] V. Lyubashevsky, A. Palacio, and G. Segev. Public-key cryptographic primitives provably as secure as subset sum. In *TCC*, pages 382–400. 2010.

[Lyu09] V. Lyubashevsky. Fiat-Shamir with aborts: Applications to lattice and factoring-based signatures. In *ASIACRYPT*, pages 598–616. 2009.

[Lyu12] V. Lyubashevsky. Lattice signatures without trapdoors. In *EUROCRYPT*, pages 738–755. 2012.

[Mic02] D. Micciancio. Generalized compact knapsacks, cyclic lattices, and efficient one-way functions. *Computational Complexity*, 16(4):365–411, 2007. Preliminary version in FOCS 2002.

[MP12] D. Micciancio and C. Peikert. Trapdoors for lattices: Simpler, tighter, faster, smaller. In *EUROCRYPT*, pages 700–718. 2012.

[MR04] D. Micciancio and O. Regev. Worst-case to average-case reductions based on Gaussian measures. *SIAM J. Comput.*, 37(1):267–302, 2007. Preliminary version in FOCS 2004.

[Pei09] C. Peikert. Public-key cryptosystems from the worst-case shortest vector problem. In *STOC*, pages 333–342. 2009.

[Pei10] C. Peikert. An efficient and parallel Gaussian sampler for lattices. In *CRYPTO*, pages 80–97. 2010.

[PM08] M. Püschel and J. M. F. Moura. Algebraic signal processing theory: Cooley-Tukey type algorithms for DCTs and DSTs. *IEEE Transactions on Signal Processing*, 56(4):1502–1521, 2008.

[PR07] C. Peikert and A. Rosen. Lattices that admit logarithmic worst-case to average-case connection factors. In *STOC*, pages 478–487. 2007.

[Reg05] O. Regev. On lattices, learning with errors, random linear codes, and cryptography. *J. ACM*, 56(6):1–40, 2009. Preliminary version in STOC 2005.

[SS11] D. Stehlé and R. Steinfeld. Making NTRU as secure as worst-case problems over ideal lattices. In *EUROCRYPT*, pages 27–47. 2011.

[SSTX09] D. Stehlé, R. Steinfeld, K. Tanaka, and K. Xagawa. Efficient public key encryption based on ideal lattices. In *ASIACRYPT*, pages 617–635. 2009.

[Ste04] W. Stein. A brief introduction to classical and adelic algebraic number theory, 2004. Available at `http://modular.math.washington.edu/papers/ant/`, last accessed 12 Oct 2009.

[SV11] N. Smart and F. Vercauteren. Fully homomorphic SIMD operations. Cryptology ePrint Archive, Report 2011/133, 2011. http://eprint.iacr.org/.

[Ver11] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices, January 2011. Available at http://www-personal.umich.edu/~romanv/papers/non-asymptotic-rmt-plain.pdf, last accessed 4 Feb 2011.