Algebraic aspects of solving Ring-LWE, including ring-based improvements in the Blum-Kalai-Wasserman algorithm

KATHERINE E. STANGE

Abstract. We provide several reductions of Ring-LWE problems to smaller Ring-LWE problems in the presence of samples of a restricted form (i.e. (a, b) such that a is restricted to a subring, or multiplicative coset of a subfield of one CRT factor). To create and exploit such restricted samples, we propose Ring-BKW, a version of the Blum-Kalai-Wasserman algorithm which respects the ring structure. It has several key advantages based on the ring structure, including smaller tables, reduced or eliminated back-substitution, and a new opportunity for parallelization. We focus on two-power cyclotomic Ring-LWE with parameters proposed for practical use, with the exception that many splitting types are considered. The orthogonality of the lattice for two-power cyclotomics is exploited. In general, higher residue degree is an advantage to attacks.

1. Introduction

Ring Learning with Errors (Ring-LWE) [17] [18], and Learning with Errors (LWE) [20] more generally, are leading candidates for post-quantum cryptography. The cryptographic hard problem (*Search Ring-LWE*) is formally similar to discrete logarithm problems, so that protocols can be transferred from the latter context to the former. But it also allows for new applications, such as homomorphic encryption [6]. Ring-LWE is also blessed with security reductions to other lattice problems.

Ring-LWE is distinguished from Learning with Errors (LWE) by the use of ideal lattices. This injection of number-theoretical structure leads to performance improvements, but may add vulnerabilities. So far, the number-theoretical structure has been only weakly exploited for attacks. The ring structure plays a role in security when the error distribution is skewed [8] [9] [10] [12] [13], or the secret is chosen from a subring or other ring-related non-uniform distribution [5]. However, the best known attacks on parameters suggested for implementation are still generic attacks for LWE, e.g. [2]. The Blum-Kalai-Wasserman (BKW) algorithm is one such attack, which

Date: February 19, 2019.

2010 *Mathematics Subject Classification.* Primary: 94A60, 11T71, 11R18.

Key words and phrases. Ring learning with errors, Ring-LWE, Blum-Kalai-Wasserman, post-quantum cryptography, cyclotomic field.

This research was supported by NSF-CAREER CNS-1652238 and NSF EAGER DMS-1643552.

proceeds (in the first phase) combinatorially to create new samples in a linear subspace of the original problem, while controlling error expansion [3]. For performance analysis and recent improvements, see [1] [14] [15] [16].

This paper focuses on two-power-cyclotomic unital (but equivalently, dual [11] [19]) Ring-LWE, with prime $q \equiv 1 \pmod 4$ having various splitting behaviours. The only deviation from recommended cryptographic parameters is that, in such recommended parameters, typically q has residue degree 1. The methods in this paper, although applicable to a wide range of residue degrees, appear most suitable for higher residue degree. Modulus switching implies that the security of primes of large residue degree is very relevant [4] [7]. However, modulus switching incurs a cost in the error width which generally increases the runtime of attacks.

The core of the paper is a suite of reductions from larger Ring-LWE problems with samples of a restricted form, to smaller Ring-LWE problems with the same error width. These are Theorems 5.1, 6.1 and 6.3. The restricted form is as follows: samples (a, b) such that a lies in a subring or subfield, or a multiplicative coset of a subfield of one CRT factor. In the context of these theorems, it is natural to ask about creating samples of this restricted form using a ring variant of the Blum-Kalai-Wasserman algorithm.

There are several other key observations:

(1) If the Ring-LWE problem can be usefully projected to a lower dimension, then in many cases the extra symmetry of the ring structure allows it to be projected to many different independent lower-dimensional problems simultaneously with little extra effort, and solved in all of these simultaneously to reconstruct the full secret. In the case of BKW, for example, this implies a parallelization opportunity, and eliminates the need for the back-substitution step. See Theorem 5.1. This applies more strongly for higher residue degree.

(2) There is a significant constant factor reduction in the number of samples which must be used to fill the BKW tables, and the table size, by exploiting ring symmetry. See Section 9.

(3) If not employing BKW or lattice reduction (for example, after reducing to a smaller dimension via BKW reduction), ring-based methods can offer square-root speedups over exhaustive search. See Corollary 5.2.

In Section 8, we propose a version of BKW that respects the ring structure, and applies when the residue degree exceeds the block size. In Section 11, we argue that these improvements, taken together, are likely to imply a practical improvement in BKW runtimes on Ring-LWE as compared to generic LWE. Specifically, for Ring-BKW as proposed in this paper, the most important factors are:

(1) The optimal block size in both cases is the same.

(2) Symmetry reduces the samples needed by a significant constant factor (and reduces table sizes exponentially).
(3) The back-substitution step is reduced or even eliminated.
(4) There is the possibility of parallelization in the reconstruction of the secret after the reduction step.
(5) There are potential speedups in the hypothesis testing step.

Together, these improvements imply a very likely practical performance enhancement. Unfortunately, a full, detailed runtime analysis is beyond the scope of this paper, which concerns itself with laying out the theoretical foundations for these potential improvements.

The key theoretical properties which are potentially advantageous (to an attacker) of Ring-LWE vs. plain LWE, are:

(1) Ring homomorphisms into finite field instances of the problem (the main tool of [9] [10] [12] [13]).
(2) The ability to *rotate* samples, e.g. replacing (a, b) with $(\zeta a, \zeta b)$ or $(a, \zeta b)$, which are different but related Ring-LWE samples (see notation in Section 2); these represent symmetries of the lattice.
(3) The existence of subfields as linear subspaces (which is important in [5]).
(4) More generally, the multiplicative structure of certain linear subspaces.
(5) In the case of 2-power cyclotomics, the orthogonality of the lattice.

The discreteness of the error distribution also has a role to play (see Section 4.5). For us, all five of these attributes play an important role. It is a secondary purpose of this paper to lay out these advantages in a clear manner, to facilitate future analysis of the security of ring aspects of Ring-LWE. See Section 4.

Finally, it is also a secondary purpose of this paper to provide a treatment of the Ring-LWE problem which is inviting to the mathematical community.

Code demonstrating the algorithms is available at:
http://math.colorado.edu/~kstange/ring-bkw.html.

Acknowledgements. I would like to thank my mother, Ursula Stange, and my husband, Jonathan Wise, without whose childcare help in the face of snowstorms, viruses, cancellations and fender-benders, this paper simply would not have been completed. To mathematician moms (and dads) everywhere: take heart.

2. Background and Setup for Ring-LWE

It is typical to set notation for Ring-LWE as in, for example, [5]; here we briefly review this notation in our context, and define the Ring-LWE problems.

2.1. **Number field K and ring R.** Let $K = \mathbb{Q}(\zeta)$ be a number field generated by ζ over the rationals, of degree n. Then K is equipped with a bilinear form given by the trace pairing,

$$\langle \alpha, \beta \rangle = \mathrm{Tr}_{\mathbb{Q}}^{K}(\alpha\beta).$$

This gives an isomorphism of $K_{\mathbb{R}} := \mathbb{R} \otimes_{\mathbb{Q}} K$ with $\mathbb{R}^n$, taking the trace form to the standard inner product. We can also denote the norm by $||\mathbf{x}|| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$. One can also access this norm using the Minkowski or canonical embedding.

The ring of integers R of K forms a lattice in $K_{\mathbb{R}}$.

2.2. **Gaussian distribution.** Having geometry (in particular a norm $|| \cdot ||^2$) on $K_{\mathbb{R}}$ allows us to define Gaussian distributions. For any *Gaussian parameter* r, we write

$$\rho_r : K_{\mathbb{R}} \to (0, 1], \quad \rho_r(\mathbf{x}) = \exp(-\pi ||\mathbf{x}||^2 / r^2).$$

Normalizing this to obtain a probability distribution function $r^{-n}\rho_r$, we obtain the *continuous Gaussian probability distribution of width* r, denoted D_r.

Note that, when considered with respect to an orthonormal basis, such a distribution is the sum of independent distributions in each coordinate, each having the same width. In this paper, we are concerned exclusively with this case, and we henceforth assume it.

With this normalization, the variance is $r^2/2\pi$, and one standard deviation is $r/\sqrt{2\pi}$. It is a sum of independent Gaussians in each coordinate for which the range $[-r, r]$ corresponds to $\sqrt{\pi/2} \sim 1.25\ldots$ standard deviations. Albrecht et al. [1] use the notation α determined by $r/\sqrt{2\pi} = \alpha q$, as a measure of the 'width' of a Gaussian. We will use that notation here also.

In practice, the tails of the Gaussian may be cut off, so that the number of possible values in each coordinate is finite. In this case, the number of possible values may be, for example, in the case of two standard deviations, $4\alpha q$.

2.3. **Choice of ring.** Let $\mathfrak{a}$ be an ideal above a rational prime q. The fundamental setting of the Ring-LWE problem is the ring $R/\mathfrak{a}$.

Note that this defines Ring-LWE in slightly greater generality; most literature concerns only the case $\mathfrak{a} = qR$ (where we write $R_q = R/qR$). Note that $\mathfrak{a}$ is some divisor of qR, i.e. contains it, so that $R/\mathfrak{a}$ is a quotient of R_q. More precisely, letting $q = \mathfrak{q}_1 \cdots \mathfrak{q}_g$ be the decomposition of q into prime ideals in K, the Chinese remainder theorem gives

$$R_q \cong \bigoplus_{i=1}^{g} R/\mathfrak{q}_i.$$

Then

$$R/\mathfrak{a} \cong \bigoplus{}' R/\mathfrak{q}_i.$$

where the $'$ denotes that the sum is taken over some subset of the $\mathfrak{q}_i$.

2.4. **Ring-LWE distributions.** For any $s \in R/\mathfrak{a}$ (the *secret*), and any distribution ψ over $R/\mathfrak{a}$ (the *error distribution*), we write $A_{s,\psi}$ for the associated *Ring-LWE distribution for secret* s over $R/\mathfrak{a} \times R/\mathfrak{a}$, given by sampling a uniformly over $R/\mathfrak{a}$, sampling e from ψ, and outputting $(a, b := as + e)$.

Such outputs (a, b) are called *samples*, and in a crytographic application, these are observed publicly, while the secret is not meant to be exposed.

It is typical to choose for the error distribution a discretized version of the Gaussian distribution described above, and this is the context in which security reductions apply. In implementations, it is sometimes suggested to approximate this by a uniform distribution on a box around the origin.

2.5. **Ring-LWE Problems.** The two fundamental Ring-LWE problems are (a) *search*: to compute the secret, upon observing sufficiently many samples; or (b) *decision*: to determine if the sample are hiding a secret at all, as opposed to being random noise. We state them more formally as follows.

Definition 2.1. The *search Ring-LWE problem*, for error distribution ψ and secret distribution φ, is as follows: Given an error distribution ψ over $R/\mathfrak{a}$ and a secret distribution φ over $R/\mathfrak{a}$, and some number of samples drawn from the distribution $A_{s,\psi}$ for some fixed s drawn from φ, compute s.

Definition 2.2. The *decisional Ring-LWE problem*, for error distribution ψ and secret distribution φ, is as follows: Given an error distribution ψ over $R/\mathfrak{a}$ and a secret distribution φ over $R/\mathfrak{a}$, distinguish samples drawn from the distribution $A_{s,\psi}$ for some fixed s drawn from φ, and samples drawn uniformly from $R/\mathfrak{a} \times R/\mathfrak{a}$.

We remark that Ring-LWE is frequently stated in terms of the dual $R^\vee$ (the different ideal), but in the case that K is a two-power cyclotomic field, we have $R^\vee \cong R$ (i.e. the different is principal), so we can interchange with the simpler 'unital' version [11] [19].

Search-to-decision reductions are known in a variety of contexts [19]. This paper concerns both problems, but especially the search problem.

The Ring-LWE problem is formally similar to the discrete logarithm problem, which could be phrased in terms of *samples* (a, a^s) in a finite field: given (a, a^s), find s. In the ring $R/\mathfrak{a}$, solving for s given (a, as) can be accomplished using linear algebra (Gaussian elimination). By introducing a small error e, so we have $(a, as + e)$, Gaussian elimination becomes useless, as it amplifies the errors to the point of washing out all useful information. From another perspective, the security stems from the fact that addition of an error value is somehow unpredictably mixing with respect to multiplicative cosets.

Another consequence of this setup is that given just one sample (a, b), one has as many solutions s to $b = as + e$ as there are possible values for e. In fact, the problem only has a unique solution once we have enough samples. If the samples are not Ring-LWE samples at all, then with sufficiently many samples, it becomes overwhelmingly likely that there are no values of s so that $b_i - a_i s$ is in the support of the error distribution for all samples s. If the samples are Ring-LWE, this is the point at which the true secret is the only solution, with overwhelming probability.

3. Specializing to 2-power cyclotomic Ring-LWE

We now specialize to the following situation, fixing the variables

$$R, q, R_q, m, m_0, n, n_0, \zeta_m, k, r, \chi$$

for the remainder of the paper.

3.1. **Ring R.** Let $m = 2^{m_0}$ for $m_0 \geq 2$, and write $n = 2^{n_0} = \varphi(m) = 2^{m_0 - 1}$ (this is the Euler φ function). We start with the cyclotomic ring of integers of dimension n, generated by the m-th roots of unity, which can be presented as

$$R = \mathbb{Z}[\zeta_m] = \mathbb{Z}[x]/(x^n + 1).$$

We will use the notation ζ_m for a primitive m-th root of unity in R *and for its image in quotients of this ring.* A basis for R is

$$1, \zeta_m, \zeta_m^2, \ldots, \zeta_m^{n-1}.$$

This will be called the ζ-*basis*. We have the relation $\zeta_m^n + 1 = 0$.

3.2. **Prime q.** Let $q \equiv 1 \pmod 4$ be a prime.

3.3. **Ring $R/\mathfrak{a}$.** We consider the quotient ring

$$R_q = R/qR \cong (\mathbb{Z}/q\mathbb{Z})[x]/(x^n + 1),$$

which is an $\mathbb{F}_q$-vector space of dimension n. We may use the same ζ-basis for this ring. We may also consider further quotients $R/\mathfrak{a}$ for $\mathfrak{a} \mid qR$. We may also use a ζ-basis for these rings, although it may be smaller (fewer powers of ζ_m required).

3.4. **Embedding degree k and the variable r.** Let k be the embedding degree of the m-th roots of unity. Equivalently, k is the smallest power of two such that $2^{m_0} \mid q^k - 1$. Then, $k \mid n$ and

$$R_q \cong \mathbb{F}_{q^k} \times \mathbb{F}_{q^k} \times \cdots \times \mathbb{F}_{q^k} = (\mathbb{F}_{q^k})^{n/k},$$

Write $r = \operatorname{ord}_2(q-1)$, meaning that $2^r \mid q - 1$ but $2^{r+1} \nmid q - 1$. Since $q \equiv 1 \pmod 4$, we know $r \geq 2$, and therefore

$$k = 2^{m_0 - r}. \tag{1}$$

To see this latter fact, note that $q^i + 1 \equiv 2 \pmod 4$ for all i, so that $\operatorname{ord}_2(q^2 - 1) = \operatorname{ord}_2((q-1)(q+1)) = r + 1$, and similarly $\operatorname{ord}_2(q^4 - 1) = r + 2$ and so on.

With the setup above,

$$1, \zeta_m, \zeta_m^2, \ldots, \zeta_m^{k-1}$$

forms an $\mathbb{F}_q$-basis for $\mathbb{F}_{q^k}$ and we call it the ζ-basis of $\mathbb{F}_{q^k}$.

3.5. **Error distribution χ and the coefficient support E_χ.** We will consider Ring-LWE in general rings $R/\mathfrak{a}$, and denote the error distribution by χ. If this error distribution is formed using independent identically distributed coefficients on the ζ-basis, with coefficient distribution χ_0 supported on a subset $E_{\chi_0} \subseteq \mathbb{F}_q$, then we say that χ is *formed on a ζ-basis with coefficients distributed according to* χ_0. This is true, for example, of a discrete Gaussian distribution on two-power cyclotomics, or a distribution formed by choosing coefficients uniformly from some subset of $\mathbb{F}_q$. For this paper, we will concern ourselves exclusively with this case.

3.6. **Secret distribution.** We will not make any particular assumption on the secret distribution. It may be taken to be uniform on R_q.

4. Key theoretical properties

In this section we highlight several key aspects of Ring-LWE absent in LWE.

4.1. **Ring homomorphisms.** If a Ring-LWE problem is presented in R_q, for any $\mathfrak{a} \mid qR$, we have a ring homomorphism

$$\rho : R_q \to R/\mathfrak{a}$$

This transports samples distributed according to $A_{s,\chi}$ to samples distributed according to $A_{\rho(s),\rho(\chi)}$.

Proposition 4.1. *Suppose $R/\mathfrak{a} \cong \mathbb{F}_{q^k}$. If, in R_q, the error distribution χ is formed on the ζ-basis in R_q with coefficients drawn from χ_0 on $\mathbb{F}_q$, then $\chi' := \rho(\chi)$ in $\mathbb{F}_{q^k}$ is formed on the ζ-basis in $\mathbb{F}_{q^k}$ with coefficients drawn from χ_0', where $\rho(\zeta_m^k) \in \mathbb{F}_q$ and*

$$\chi_0' = \sum_{i=0}^{n/k-1} \rho(\zeta_m^k)^i \chi_0.$$

Proof. This follows from the fact that $1, \zeta_m, \ldots, \zeta_m^{k-1}$ is an $\mathbb{F}_q$-basis of $\mathbb{F}_{q^k}$, that $\rho(\zeta_m^k) \in \mathbb{F}_q$ and that for $0 \leq j < k$ and $0 \leq i < n/k$, we have

$$\rho(\zeta_m^{ik+j}) = \rho(\zeta_m^k)^i \rho(\zeta_m^j) = \rho(\zeta_m^k)^i \zeta_m^j.$$

$\square$ $\square$

For example, in the case that $k = n/2$, we obtain

$$\chi' = \chi + \rho(\zeta_m^k)\chi.$$

This means the coefficients of χ' are chosen from a sum of two Gaussian distributions with different coefficients. This is worse than twice a single Gaussian. For, the latter is simply a wider Gaussian, and the size of its support grows by approximately $\sqrt{2}$. However, here the size of the support $E_{\chi'}$ is approximately the square of the size of E_χ. This is a symptom of the protective property of these ring homomorphisms: they transform the error badly. In fact, very quickly the image of a Gaussian error approaches uniform in the image ring as the dimension of the image ring decreases.

4.2. **Rotating samples.** The ring structure allows us to generate new (but not independent) samples from old. The following proposition is an immediate consequence of the fact that the error distribution is invariant under multiplication by powers of ζ^m.

Proposition 4.2. *If (a, b) is distributed according to $A_{s,\chi}$, then for any $\zeta = \zeta_m^i$,*

(1) $(\zeta a, \zeta b)$ is also distributed according to $A_{s,\chi}$,
(2) $(a, \zeta b)$ is distributed according to $A_{\zeta s,\chi}$.

We call these *rotated samples*. One could also rotate by other values, e.g. $1 + \zeta_m$, at a small cost in changing the error distribution.

4.3. **Subfields.** If considering Ring-LWE in $\mathbb{F}_{q^k}$, then certain linear subspaces are actually subfields of $\mathbb{F}_{q^k}$, and $\mathbb{F}_{q^k}$ has a vector space structure over these subfields.

4.4. **Multiplicative cosets.** Consider a subfield $\mathbb{F}_{q^d} \subseteq \mathbb{F}_{q^k}$. Then certain linear subspaces of $\mathbb{F}_{q^k}$ are multiplicative cosets $a\mathbb{F}_{q^d}$ for some $a \in \mathbb{F}_{q^k}^*$. (Strictly speaking, their non-zero elements are multiplicative cosets of $\mathbb{F}_{q^d}^*$ in $\mathbb{F}_{q^k}^*$.) For a fixed d, the collection of such linear subspaces intersect only at 0 and have as their union the full space. Rotation of samples (a, b) respects the multiplicative cosets of a and s in some fashion.

Proposition 4.3. *If (a, b) is a sample with $a \in a_0\mathbb{F}_{q^d}^*$ and $s \in s_0\mathbb{F}_{q^d}^*$, then*

(1) $(\zeta a, \zeta b)$ is a sample with $a \in \zeta a_0\mathbb{F}_{q^d}^*$ and secret $\in s_0\mathbb{F}_{q^d}^*$,
(2) $(a, \zeta b)$ is a sample with $a \in a_0\mathbb{F}_{q^d}^*$ and secret $\in \zeta s_0\mathbb{F}_{q^d}^*$.

The multiplicative coset structure gives rise to another type of sample reduction, beyond ring homomorphism. We have

Proposition 4.4. *Suppose $s \in \mathbb{F}_{q^k}^*$ is fixed. Define $T := Tr_{\mathbb{F}_{q^d}}^{\mathbb{F}_{q^k}}$, the trace map. Consider a collection of samples distributed according to $A_{s,\chi}$ but having the further restriction that $a \in a_0\mathbb{F}_{q^d}^*$. In other words, a is chosen randomly from $a_0\mathbb{F}_{q^d}^*$ and e is chosen according to χ. Suppose $T(a_0) \neq 0$. Then T maps such samples to samples distributed according to $A_{s',\chi'}$ in $\mathbb{F}_{q^d}$, where*

$$s' = \frac{T(a_0 s)}{T(a_0)}$$

and $\chi' = T(\chi)$.

Proof. For $a = a_0 a' \in a_0\mathbb{F}_{q^d}$, we have

$$T(as) = a'T(a_0 s).$$

This implies that

$$(T(a), T(as+e)) = \left(a'T(a_0), a'T(a_0)\left(\frac{T(a_0 s)}{T(a_0)}\right) + T(e)\right)$$

This proves the proposition. □ □

4.5. **Orthogonality of 2-power cyclotomics and the trace map.** The final piece to the puzzle is the behaviour of the trace map T in the previous section. In the case of the 2-power cyclotomics with $q \equiv 1 \pmod 4$, the trace map is particularly well-behaved in terms of its effect on the error distribution. In fact, it takes very many of the basis elements ζ_m to zero. This is a feature of the orthogonality of the basis $1, \zeta_m, \ldots, \zeta_m^{n-1}$, and it may be proved with reference to basic algebraic number theory.

The subgroups of $\mathbb{F}_{q^k}^*$ are exactly μ_ℓ for $\ell \mid q^k - 1$. The subgroup μ_ℓ has index ℓ and is of size $(q^k - 1)/\ell$. The following proposition is the most general statement behind the good behaviour of the trace map. We momentarily suspend the notational conventions of Section 3 and give a general result about finite fields.

Proposition 4.5. *Let q be any prime and k any positive integer. Let $\ell \mid q^k - 1$. Suppose $k = \kappa d$, and $\ell = \kappa \ell'$, where $\kappa, d, \ell' \in \mathbb{Z}$, κ is prime and $\ell' = gcd(\ell, q^d - 1)$. Then all ℓ-th roots of unity lying in $\mathbb{F}_{q^k} \smallsetminus \mathbb{F}_{q^d}$, i.e. elements of $\mu_\ell \smallsetminus \mu_{\ell'}$, have trace 0 down to $\mathbb{F}_{q^d}$.*

Proof. Under the indicated hypotheses, the primitive ℓ-th roots of unity are κ-th roots of elements of $\mathbb{F}_{q^d}$, defined over an extension of degree κ. Therefore their minimal polynomials are of the form $x^\kappa - c$ for some $c \in \mathbb{F}_{q^d}$ and so have trace 0 to $\mathbb{F}_{q^d}$ (and hence to $\mathbb{F}_q$ also). □ □

A particularly useful corollary is the following. We return to our notational conventions.

Proposition 4.6. *Let m, n, q, k, r be as in Section 3. Let $d \mid k$, so that $\mathbb{F}_{q^d} \subseteq \mathbb{F}_{q^k}$. Let $T = Tr_{\mathbb{F}_{q^d}}^{\mathbb{F}_{q^k}}$. Then $\mu_m \cap \mathbb{F}_{q^d}^* = \mu_{md/k}$ and $\mu_m \smallsetminus \mu_{md/k}$ is in the kernel of T. In particular, for $d = 1$, all m-th roots of unity besides μ_{2^r} have trace 0 down to $\mathbb{F}_q$.*

Proof. Since $\mathrm{ord}_2(q-1) = r$, then $q + 1 \equiv 2 \pmod 4$, hence $\mathrm{ord}_2(q^2 - 1) = r + 1$. By the same logic, $\mathrm{ord}_2(q^4 - 1) = r + 2$, $\mathrm{ord}_2(q^8 - 1) = r + 3$, etc. In general, $\mathrm{ord}_2(q^{2^{\ell - r}} - 1) = \ell$. In particular, $\mathrm{ord}_2(q^d - 1) = \frac{md}{k}$ (recall that $m = 2^r k$ from (1)). Hence $\mu_m \cap \mathbb{F}_{q^d}^* = \mu_{md/k}$. For the statement about traces, we apply Proposition 4.5 to $\mathbb{F}_{q^k}$ and $\mathbb{F}_{q^d}$. We conclude that every ℓ-th root of unity save those in $\mathbb{F}_{q^d}$ have trace zero down to $\mathbb{F}_{q^d}$. □ □

The following corollary of Proposition 4.5 tells us that, in the case of interest to us, for all but one multiplicative coset of $\mathbb{F}_{q^{k/2}}^* < \mathbb{F}_{q^k}^*$, the traces down to $\mathbb{F}_q$ are distributed evenly, while for one special coset, the traces are all zero. For the following proposition, we also momentarily suspend our notational conventions and consider any even k.

Proposition 4.7. *Let k be even. Write $d' = q^{k/2} - 1$, so that $\mu_{d'} = \mathbb{F}_{q^{k/2}}^*$. Write $d = 2d'$. Let us write ζ_d for a primitive d-th root of unity in $\mathbb{F}_{q^k}$. Then, for $a_0 \in \mathbb{F}_{q^k}$, the distribution of traces of the multiplicative coset $a_0\mathbb{F}_{q^{k/2}}^*$ down to $\mathbb{F}_q$ is exactly:*

$$Pr\left(\mathrm{Tr}_{\mathbb{F}_{q^d}}^{\mathbb{F}_{q^k}} = x \mid x \in a_0\mathbb{F}_{q^{k/2}}^*\right) = \begin{cases} q^{k/2} - 1 & a_0 = \zeta_d, x = 0 \\ 0 & a_0 = \zeta_d, x \neq 0 \\ q^{k/2-1} - 1 & a_0 \neq \zeta_d, x = 0 \\ q^{k/2-1} & a_0 \neq \zeta_d, x \neq 0 \end{cases}$$

Proof. Note that $2d' \mid q^k - 1$, so that we are in the situation of Proposition 4.5, with $k_0 = 2$.

The traces of $\mathbb{F}_{q^{k/2}}^*$ distribute evenly amongst the values of $\mathbb{F}_q$ (except for one fewer 0). That is, their distribution is

$$Pr(\mathrm{Tr}_{\mathbb{F}_{q^d}}^{\mathbb{F}_{q^k}} = x) = \begin{cases} q^{k/2-1} - 1 & x = 0 \\ q^{k/2-1} & x \neq 0 \end{cases}$$

Suppose $Tr_{\mathbb{F}_{q^{k/2}}}^{\mathbb{F}_{q^k}}(a_0) \neq 0$, and consider the coset $a_0\mathbb{F}_{q^{k/2}}^*$. The trace of a_0x, $x \in \mathbb{F}_{q^{k/2}}$ is just $Tr(a_0)x$. Hence if the x are distributed as in the previous display, then $Tr(a_0)x$ are distributed in the same way.

Now, suppose $Tr_{\mathbb{F}_{q^{k/2}}}^{\mathbb{F}_{q^k}}(a_0) = 0$. This is equivalent to a_0 being a square root of something in $\mathbb{F}_{q^{k/2}}$, i.e. $a_0 \in \mu_d \setminus \mu d' = \zeta_d\mathbb{F}_{q^{k/2}}^*$. But this set has trace all zero, by Proposition 4.5. □ □

Proposition 4.7 has the following interesting consequence worth a brief remark. That is, the quadratic residues of $\mathbb{F}_{q^k}$ lie entirely outside of $\zeta\mathbb{F}_{q^{k/2}}^*$. Therefore the distribution of traces of residues differs from the distribution of traces of non-residues: the occurence of 0 is significantly diminished. This suggest the following attack on Ring-LWE: partition samples (a, b) according to whether a is a quadratic residue or not, and collect statistics on whether b is in the support of the trace of the error distribution to $\mathbb{F}_{q^{k/2}}$. This statistic will be different according as whether s is a quadratic residue or not, or the samples are uniform instead of Ring-LWE. Unfortunately, this approach does not outperform the other approaches in this paper.

The trace map is much less problematic when applied to error distributions than a ring homomorphism is.

Proposition 4.8. *Define $T := Tr_{\mathbb{F}_{q^d}}^{\mathbb{F}_{q^k}}$, the trace map. Suppose that χ is an error distribution formed on the ζ-basis of $\mathbb{F}_{q^k}$ with coefficients chosen according to χ_0. Then $T(\chi)$ is the error distribution formed on the ζ-basis of $\mathbb{F}_{q^d}$ with coefficients from $\frac{k}{d}\chi_0$.*

Proof. This is a consequence of Proposition 4.6 above. For, the trace acts the following way upon the ζ basis of $\mathbb{F}_{q^k}$:

$$T(\zeta_m^i) = \begin{cases} 0 & i \not\equiv 0 \pmod{k/d} \\ \frac{k}{d}\zeta_m^i & \text{otherwise.} \end{cases}$$

□ □

The efficacy of the trace map with respect to the error distribution is not merely an effect of the linear functional corresponding to a short vector of the dual lattice, as discussed in [19]. The efficacy clearly depends upon the orthogonality of the lattice of the two-power cyclotomic field, since so many basis vectors have trace zero. But there is an additional, discrete effect. For, if the error were continuous, the factor of k/d in Proposition 4.8 would be quite problematic. A discrete distribution may have the property that, although the overall width of the error spreads out, the values are restricted, so that the support is still finite, of the same size (though now spread out).

5. Multiplicative coset reduction

Consider a Ring-LWE problem in a finite field. We demonstrate that if one can find sufficiently many samples whose a values are restricted to any fixed multiplicative coset of a subfield, then we can reduce the Ring-LWE problem to multiple independent Ring-LWE instances in the subfield, without error inflation.

Theorem 5.1. *Consider a Ring-LWE instance in $\mathbb{F}_{q^k}$ with error distribution χ. Suppose one obtains N samples (a_i, b_i) where a_i is chosen from a fixed multiplicative coset $a_0\mathbb{F}_{q^d}^*$ of $\mathbb{F}_{q^k}^*/\mathbb{F}_{q^d}^*$, for some fixed $a_0 \neq 0$. Let $T := \mathrm{Tr}_{\mathbb{F}_{q^d}}^{\mathbb{F}_{q^k}}$. Then in time linear in N, and polynomial in k and $\log q$, one can reduce the original Ring-LWE instance to k/d Ring-LWE problems in $\mathbb{F}_{q^d}$ with N samples and error distribution $T(\chi)$.*

In particular, by Proposition 4.8, the support of the coefficient distributions of χ and $T(\chi)$ are of the same size; it is in this sense that the errors do not inflate.

Proof. **Correctness.** Fix an integer $0 \leq j < k/d$. Multiplying the second coordinate of the sample by ζ^j and taking the trace T to the subfield $\mathbb{F}_{q^d} \subseteq \mathbb{F}_{q^k}$, we obtain as in Proposition 4.4, N samples

$$\begin{aligned} &(T(a_i), T(a_i\zeta^j s + \zeta^j e_i)) \\ &= \left(a_i' T(a_0), a_i' T(a_0) \cdot \left(\frac{T(a_0\zeta^j s)}{T(a_0)} \right) + T(\zeta^j e_i) \right), \end{aligned}$$

where $a_i' := a_i/a_0$. These are of the form of a finite field Ring-LWE problem with parameters d, q and $T(\chi)$, since the error distribution $\chi_{\sigma,h}$ on $\mathbb{F}_{q^k}$ is not altered by multiplication by ζ^j. That is, of course, provided that $T(a_0) \neq 0$. The latter can be assured by simply replacing the samples

(a, b) with $(a\zeta, b\zeta)$ if necessary, and/or choosing another representative of the multiplicative coset, thereby altering a_0.

Solving this Ring-LWE problem means finding the secret

$$c_j := \frac{T(a_0\zeta^j s)}{T(a_0)}.$$

Collecting all the values c_j, we have a linear system of k/d equations over $\mathbb{F}_{q^d}$, in indeterminates the coefficients of s (expressed in terms of a basis for $\mathbb{F}_{q^k}$ over $\mathbb{F}_{q^d}$), of the form

$$T(a_0\zeta^j s) = c_j T(a_0), \quad 0 \le j < k/d.$$

The linear equations are independent provided that $\{a_0\zeta^j\}$ is a set of $\mathbb{F}_{q^d}$-independent vectors in $\mathbb{F}_{q^k}$. Note that $\{\zeta^j\}$ is a set of coset representatives for $\mu_m/\mu_{m/d}$, and in fact is a basis for $\mathbb{F}_{q^k}$ over $\mathbb{F}_{q^d}$. Thus independence is guaranteed by the fact that $a_0 \neq 0$. Therefore the system can be solved by Gaussian elimination to recover s. Note that we can consider this instead to be k independent linear equations over $\mathbb{F}_q$.

Runtime. All the field operations concerned are polynomial in $\log q$ and k. We must apply the trace to N samples k/d times, and we must solve Gaussian elimination of dimension k over $\mathbb{F}_q$, which is polynomial in k and $\log q$. □ □

As a corollary, note that in any small Ring-LWE situation where exhaustive search may apply, it is equally possible to use the above for a square-root speedup, provided many samples are available.

Corollary 5.2. *Consider a Ring-LWE problem in $\mathbb{F}_{q^k}$ with error distribution χ formed on a ζ-basis with coefficients $E_\chi \neq \mathbb{F}_q$. There is an algorithm to solve this problem, with success probability $1/2$, in time and number of samples $q^{k/2}$ times factors polynomial in $k \log q$, using space polynomial in $k \log q$.*

Proof. Collect samples, discarding all but those with $a \in \mathbb{F}_{q^{k/2}}$. In time $O(Nq^{k/2})$ we can accumulate N samples with $a \in \mathbb{F}_{q^{k/2}}$. Apply Theorem 5.1 to reduce to two Ring-LWE problems in $\mathbb{F}_{q^{k/2}}$ with N samples each. The error distribution χ on $\mathbb{F}_{q^k}$ gives an error distribution $T(\chi)$ on $\mathbb{F}_{q^{k/2}}$. If χ is formed on a ζ-basis with coefficients $E_\chi \neq \mathbb{F}_q$, then $T(\chi)$ is formed on a ζ-basis with coefficients $2E_\chi \neq \mathbb{F}_q$. Therefore, if the number of samples is sufficient, the reduced Ring-LWE problems are solvable using exhaustive search through possible s values.

In our case, we need N large enough so that a Ring-LWE problem in $\mathbb{F}_{q^{k/2}}$ with N samples has a unique solutions with probability $1/\sqrt{2}$. Although N depends upon $|E_\chi|$, for the worst case $|E_\chi| = q - 1$, N is still polynomial in $k \log q$. Solve the reduced problems by exhaustive search, which takes time $O(q^{k/2})$ and each succeeds with probability $1/\sqrt{2}$. □ □

6. CRT reduction

In the previous section, we showed that, in a finite field, finding samples with a restricted to a multiplicative coset reduced the Ring-LWE problem to numerous smaller Ring-LWE problems without error growth. In this section, we do something similar for the Chinese remainder theorem (CRT) decomposition. To be specific, write a prime decomposition $\mathfrak{a} = \mathfrak{q}_1 \cdots \mathfrak{q}_f$ and consider the CRT decomposition

$$R/\mathfrak{a} \cong \bigoplus_{j=1}^{f} R/\mathfrak{q}_j \cong (\mathbb{F}_{q^k})^f.$$

Write

$$\rho_j : R/\mathfrak{a} \to R/\mathfrak{q}_j \cong \mathbb{F}_{q^k}$$

for the individual projections onto each factor. Then, we show that finding samples with a having $\rho_j(a) = 0$ for all but a fixed $j = j_0$ reduces the problem of finding one CRT coordinate of the secret in the Ring-LWE problem in $R/\mathfrak{a}$ to the problem of solving a Ring-LWE problem in $R/\mathfrak{q}_i$, without error growth.

Theorem 6.1. *Consider a Ring-LWE problem posed in $R/\mathfrak{a}$ where $\mathfrak{a} \mid qR$ and error distribution χ which is formed on the ζ-basis over $\mathbb{F}_{q^k}$ with coefficients chosen from χ_0, a distribution on $\mathbb{F}_{q^k}$. Suppose one obtains N samples (a_i, b_i) where $\rho_j(a_i) = 0$ for all $j \neq j_0$. Then, in time linear in N, and polynomial in k and $\log q$, one can produce fN samples of a Ring-LWE instance in $R/\mathfrak{q}_{j_0} \cong \mathbb{F}_{q^k}$ with secret $\rho_{j_0}(s)$ and error distribution χ_0.*

Furthermore, the set of a_i in the new fN samples is the Cartesian product of the list of $\rho_{j_0}(a_i)$ and a fixed list of size f in $\mathbb{F}_{q^k}$ dependent only on the ring. Furthermore, the latter list is not all zero and can be computed explicitly.

Proof. **Correctness.** Let $(a_i, b_i) = (a_i, a_i s + e_i)$ be one of the N samples with $\rho_j(a_i) = 0$ for all $j \neq j_0$. The sample $(\rho_{j_0}(a_i), \rho_{j_0}(b_i))$ is a Ring-LWE sample in $R/\mathfrak{q}_{j_0}$ with secret $\rho_{j_0}(s)$, but it does not have the error distribution stated. We will describe how to modify it. Write $M : (\mathbb{F}_{q^k})^f \to (\mathbb{F}_{q^k})^f$ for the linear transformation taking $x \in R/\mathfrak{a}$ with respect to the ζ-basis over $\mathbb{F}_{q^k}$ to $x \in (\mathbb{F}_{q^k})^f$ in the CRT basis (i.e. $(\rho_i(x))_{i=1}^f$). The transformation M is invertible. Let $e_{i,w}$ be the coefficient of ζ^w in the ζ-basis representation of e_i on $R/\mathfrak{a}$. Then, using M^{-1}, there is a linear combination

$$\sum_{j=0}^{f} \alpha_{j,w} \rho_j(e_i) = e_{i,w}.$$

Since $\rho_j(a_i) = 0$ for $j \neq j_0$, we have

$$\rho_j(b_i) = \rho_j(e_i) \text{ for all } j \neq j_0.$$

Therefore, taking the corresponding linear combination of the $\rho_j(b_i)$, we obtain

$$\sum_{j=0}^{f} \alpha_{j,w}\rho_j(b_i) = \sum_{j=0}^{f} \alpha_{j,w}\rho_j(e_i) + \alpha_{j_0,w}\rho_{j_0}(a_i)\rho_{j_0}(s) = e_{i,w} + \alpha_{j_0,w}\rho_j(a_i)\rho_j(s)$$

Therefore the sample

$$(\alpha_{j_0,w}\rho_{j_0}(a_i), \sum_{j=0}^{f} \alpha_{j,w}\rho_j(b_i))$$

has the properties described in the theorem. As w ranges from 1 to f, we obtain a total of f samples.

Runtime. The computation of the new sample involves evaluating the CRT map, followed by some linear combinations. If the matrix M^{-1} is pre-computed, this is just linear algebra and finite field operations. □ □

The special form of the samples allows one to solve decision-Ring LWE immediately.

Corollary 6.2. *Consider a Ring-LWE problem posed in $R/\mathfrak{a}$ where $\mathfrak{a} \mid qR$ and error distribution χ which is formed on the ζ-basis with coefficients chosen from χ_0, a distribution on $\mathbb{F}_{q^k}$. Suppose that $|E_{\chi_0}|^2 < q$. Suppose one obtains $N >> q$ samples (a_i, b_i) where $\rho_j(a_i) = 0$ for all $j \neq j_0$. Then, in time linear in N, and polynomial in k and $\log q$, one can solve decision Ring-LWE with non-negligable advantage.*

Proof. Use Theorem 6.1 to obtain fN samples (a_i, b_i) in $\mathbb{F}_{q^k}$ with a_i of the form

$$\{\alpha_k\beta_j : 1 \leq k \leq f, 1 \leq j \leq N\}$$

for some lists $\alpha_1, \ldots, \alpha_f$ and $\beta_1, \ldots, \beta_N$. We may assume that at least one of the α_k, say α_ℓ, is non-zero. Then given a pair

$$(a_1, b_1) = (\alpha_k\beta_j, b_1), \quad (a_2, b_2) = (\alpha_\ell\beta_j, b_1),$$

we can compute

$$(a_1, b_1) - \alpha_k\alpha_\ell^{-1}(a_2, b_2)$$

which will be a sample of the form $(0, e)$ with e drawn on the ζ-basis using coefficients from $\chi_0 - \alpha_k\alpha_\ell^{-1}\chi_0$. We can compute N such samples. If the support of χ_0 satisfies $|E_{\chi_0}|^2 < q$, then this distribution is recognizably non-uniform and a chi-square test can detect this using q bins with $N >> q$ samples. □ □

See [9, §4.1-4.2] and [1, §3.2] for information on using such statistical tests. This result could be tightened, and could apply, with appropriate statistical tests, to situations beyond $|E_{\chi_0}|^2 < q$. One is also tempted to combine this result with BKW, to see if the BKW reduction step could be performed in smaller dimension, followed by a statistical distinguishing step of this type. The problem here is that the distinguishing step requires a much narrower error, likely negating the gain in BKW dimension; this analysis would look somewhat like that in Section 10.1, and is not promising. As this result is not the main result of the paper, we do not pursue the details further at this time.

The next theorem describes how, having found samples whose a have support only on one factor, and having solved the resulting Ring-LWE problem in that CRT factor, we may set up and solve a Ring-LWE problem in the next CRT factor, and so on. This corresponds to the back-substitution step of classical BKW.

The following theorem differs from Theorem 6.1 only in that the hypothesis is loosened so that for $j \neq j_0$, $\rho_j(a_i) = 0$ or $\rho_j(s)$ is known.

Theorem 6.3. *Consider a Ring-LWE problem posed in $R/\mathfrak{a}$ where $\mathfrak{a} \mid qR$ and error distribution χ which is formed on the ζ-basis over $\mathbb{F}_{q^k}$ with coefficients chosen from χ_0, a distribution on $\mathbb{F}_{q^k}$. Suppose one obtains N samples (a_i, b_i) where for each $j \neq j_0$, either $\rho_j(a_i) = 0$ or $\rho_j(s)$ is known. Then, in time linear in N, and polynomial in k and $\log q$, one can produce fN samples of a Ring-LWE instance in $R/\mathfrak{q}_{j_0} \cong \mathbb{F}_{q^k}$ with secret $\rho_{j_0}(s)$ and error distribution χ_0.*

Furthermore, the set of a_i in the new fN samples is the Cartesian product of the list of $\rho_{j_0}(a_i)$ and a fixed list of size f in $\mathbb{F}_{q^k}$ dependent only on the ring. Furthermore, the latter list is not all zero and can be computed explicitly.

Proof. The only difference by comparison to Theorem 6.1 is that for some $j \neq j_0$ we may not have $\rho_j(a_i) = 0$, but instead $\rho_j(s)$ is known. Carry over notation from that proof. The only difference is that we form the sample

$$(\alpha_{j_0}\rho_{j_0}(a_i), \sum_{j=0}^{f} \alpha_j \rho_j(b_i) - \sum_{j \in J} \alpha_j \rho_j(a_i)\rho_j(s))$$

where J is the set of j such that $\rho_j(s)$ is known. This sample has the properties required. □ □

7. Background on the Blum-Kalai-Wasserman algorithm

First, we will give a very brief overview of the BKW algorithm in the context of LWE. It is a combinatorial algorithm in which samples are collected and stored so as to facilitate the iterative creation of new samples, as repeated sums and differences of established ones. The goal is to create new samples for which a is restricted to a linear subspace. This is the *reduction phase* of the full BKW algorithm.

In BKW, after reduction, there is a hypothesis testing phase, in which one solves a smaller Ring-LWE problem by exhaustive search over possible secrets. And then there is a back-substitution phase, where the small piece of the secret recovered in hypothesis testing is used to rework the problem to prepare the next small piece for hypothesis testing. Theorem 6.3 will play the role of of back-substitution in Ring-BKW.

During the reduction phase, only the a-value of a sample matters, considered as a vector in a vector space V, and the goal is to create samples with $a \in W$, a linear subspace of V. Suppose, for the sake of explanation, that W is defined by the first r coefficients of its vectors being 0. One generates an ordered list of the first r entries of all the vectors a which are observed. Whenever a new vector a is observed, it is compared to the ordered list. If it is not already present, it is added. Otherwise, we have discovered two samples (a, b) and (a', b') for which $(a - a', b - b')$ is a new sample for which $a - a'$ lies in W. The penalty is that the error distribution of these new samples is widened. We begin a new table of such vectors as they are generated. In this way, we produce a large number of samples in a smaller subspace at the cost of inflating the error widths.

Instead of performing this reduction all at once, one chooses an appropriate *block size* B for BKW, which is to say, the codimension of W as a subspace of V. Once we have produced enough samples in W, we can use these to perform another BKW reduction to a subspace $W' \subseteq W$ of codimension B in W. The cost of a reduction step is exponential in B, so we keep B as small as possible. We perform block reductions until the samples are all taken from a small enough subspace to run an exhaustive search or other strategy to finish off the problem. The limiting factor on shrinking B is an upper limit on the number of blocks used overall. Each reduction into codimension B has a cost in error-inflation. We have a limit on the total error inflation (because hypothesis testing will fail if the error is so inflated as to appear uniform), which limits the total number of blocks.

The BKW algorithm has been improved in recent years, including using coding theory to reduce the number of values that need to be stored and compared; see [1] [14] [15] [16].

8. Ring-BKW

In this section, we address the problem of finding sufficiently many samples (a, b) having a from a multiplicative coset of a subfield of a CRT factor, or are supported on only one CRT factor, so that Theorems 5.1, 6.1, and 6.3 will apply. For this, we use a BKW algorithm adapted to the ring situation.

8.1. **Reduction Phase.** In light of Coset and CRT Reduction (Theorems 5.1 and 6.1), we first consider the reduction phase of BKW, as our goal is simply to create samples whose a values live in a certain subring or multiplicative coset within a subfield.

Suppose we have Ring-LWE on $R/\mathfrak{a}$ with error χ formed on the ζ-basis of $R/\mathfrak{a}$ with coefficients from χ_0. Suppose that $R/\mathfrak{a}$ has dimension k over $\mathbb{F}_q$ (e.g. if $\mathfrak{a}$ is prime, then $R/\mathfrak{a} \cong \mathbb{F}_{q^k}$). The BKW algorithm is as follows.

(1) Choose an integer parameter $B \mid k$, the *block size*.
(2) Create k/B ordered dictionaries $D_1, \ldots, D_{k/B}$ (the *tables*).

(3) For each new sample (a_i, b_i),
   (a) For each ordered dictionary D_j until the sample is stored,
      (i) Create a *key* for the sample, which is discussed in more detail later, but which is a linear transformation of a_i into $\mathbb{F}_q^B$.
      (ii) Compare the key to the dictionary D_j.
      (iii) If the key is not found, add the sample to the dictionary under that key.
      (iv) If the key is found, replace the sample with its difference with the sample stored under that key, and proceed to the next dictionary.
   (b) If we reach the final dictionary, store the sample in a list corresponding to the key, regardless of whether the key already exists (so a dictionary key references a list of samples, not just one).

After sufficiently many samples, the final dictionary will be populated with samples. They will be samples whose keys for the prior dictionaries are all 0. They will also be samples which are formed as linear combinations, with coefficients ±1, of $2^{k/B-1}$ original samples.

We now discuss naïve keying. First, suppose $\mathfrak{a}$ is prime so that we are in a finite field. Then we replace the usual ζ-basis for the field, namely,

$$1, \zeta, \zeta^2, \ldots, \zeta^{k-1},$$

with a *prioritized basis*, which is a reordering of the ζ-basis above. The prioritized basis is defined by two properties:

(1) if one of ζ^i and ζ^j generates a strictly smaller subfield of $\mathbb{F}_{q^k}$ than the other, then it comes later than the other
(2) if they generate the same field, their relative ordering matches that of the usual basis

The first item is what matters; the second can be replaced with any convenient tie-breaking convention.

Now, for dictionary D_j, the corresponding key is the ordered list of coefficients of basis elements $(j-1)r$ through $jr-1$ in the prioritized basis. In this way, final dictionary $D_{k/B}$ will consider the coefficients of the portion of the basis living in the subfield of size q^B (under the simplifying assumption that B divides k). Therefore samples which are added to the final dictionary have $a \in \mathbb{F}_{q^B}$. If B does not divide k, then the final samples live in a somewhat larger subfield.

Now suppose that $\mathfrak{a}$ is not prime, but has a CRT decomposition as a product of finite fields. Then for $a \in \mathfrak{a}$, we present it as a vector whose first segment of coordinates are $\rho_1(a)$ in the prioritized basis for $\rho_1(R/\mathfrak{a})$, followed by $\rho_2(a)$ in the prioritized basis for $\rho_2(R/\mathfrak{a})$ and so on. In this case, as we move through the dictionaries, we gradually build samples whose a values satisfy $\rho_i(a) = 0$ for more and more values of i. Provided the block size is smaller than the field size of each CRT factor, we eventually find ourselves with samples supported only on the last CRT factor, and begin moving into subfields of that finite field, as above.

8.2. **The Ring-BKW algorithm.** To solve Ring-LWE in $R/\mathfrak{a} \cong (\mathbb{F}_{q^d})^{k/d}$ with secret s, we use the following *Ring-BKW* algorithm. Importantly, this algorithm only applies if the block size does not exceed the dimension of the CRT factors, i.e. $d \geq B$. However, the room for error growth puts a minimum on the block size which may not satisfy this requirement. In this case, the theorem is not applicable in its current form (it is likely that it can be adapted, but we leave this for future work).

We also make the simplifying assumption that $B \mid k$ so that the samples in the final dictionary live in a subfield of size q^B. This assumption is not essential, but is the most efficient situation. Obvious modifications allow the algorithm to work without this assumption.

Finally, in our algorithm and analysis, for convenience, we assume a simplified version of the phase normally called hypothesis testing (that is, the phase of exhaustively searching for s in a small instance), in which we assume the error width is kept less than q. This necessitates relatively few samples for hypothesis testing. For a proper analysis of more general hypothesis testing, see [1].

Ring-BKW algorithm.

(1) Run BKW reduction (as in Section 8.1 above) with a block size of B until all samples (a, b) have a living in the final CRT factor $\rho_f(R/\mathfrak{a})$, and furthermore, in a subfield of size q^B.

(2) Use Theorem 6.1 to create samples from a Ring-LWE problem in $\rho_f(R/\mathfrak{a})$ with secret $\rho_f(s)$. From Step (1) and the proof of Theorem 6.1, by fixing one w in the proof, we can also guarantee that the samples (a, b) thus produced have a in one specific multiplicative coset of the subfield of size q^B.

(3) Solve this Ring-LWE problem as follows:
   (a) Use Theorem 5.1 to create a series of Ring-LWE problems of dimension at most B.
   (b) Use exhaustive search (or another method) to solve these.
   (c) Patch together the resulting information to obtain $\rho_f(s)$ according to the method of Theorem 5.1.

(4) Use Theorem 6.3, and the solution $\rho_f(s)$ in $\rho_f(R/\mathfrak{a})$, using samples from the BKW tables, to create Ring-LWE samples in $\rho_{f-1}(R/\mathfrak{a})$ with secret $\rho_{f-1}(s)$. One can do this using samples which satisfy $\rho_{f-1}(a)$ in a subfield of size q^B of $\rho_{f-1}(R/\mathfrak{a})$, by using the appropriate part of the BKW tables. From the proof of Theorem 6.3, this means we can guarantee that the resulting samples are in a multiplicative coset of that subfield.

(5) Solve this Ring-LWE problem to recover $\rho_{f-1}(s)$ as in Step (3).

(6) Repeat steps (4)-(5) until all $\rho_i(s)$ have been recovered.
(7) Use Chinese remainder theorem to compute s.

Theorem 8.1. *The above algorithm has runtime $O\left(q^B\right)$ times factors polynomial in d, f, B, and $\log q$.*

Proof. We consider the runtime of each step. Suppose we use a BKW block size of B.

(1) This has runtime $O(q^B)$ times polynomial factors, since the number of blocks needed is df/B.
(2) This step is polynomial.
(3) We break this into stages:
    (a) This is polynomial.
    (b) This has runtime $O(q^B)$ times polynomial factors for each problem.
    (c) This is polynomial.

    Therefore this step has total runtime $O(q^B)$ times polynomial factors.
(4) The number of samples which must be doctored is only as many as are needed for exhaustive search to succeed, looking ahead to the following step. Therefore the runtime is polynomial.
(5) As above, this step is $O(q^B)$ times polynomial factors.
(6) We repeat the $O(q^B)$ times polynomial factors work f times in total.
(7) CRT is polynomial.

Combining the above, we expect runtime as described. □ □

Note that our estimates depend on the assumption that each dictionary is filled before samples begin flowing to the next (this assumption is pessimistic from the perspective of the attacker and results in a higher runtime estimate).

The hard work lies in the original BKW reduction, and then in the exhaustive search for the smaller Ring-LWE problems. Note that it would be possible to use block size B but reduce to a rather smaller exhaustive Ring-LWE problem; this would improve the runtime so that it is entirely controlled by the BKW reduction step (1).

A detailed runtime analysis is beyond the scope of this paper.

9. Advanced Keying Speedup

9.1. **Advanced Keying.** Instead of the naïve keying described in the previous section, it can be advantageous to use a more advanced method of keying. To compute the key for a sample, one considers the k rotated samples

$$(a, b), (\zeta a, \zeta b), \ldots, (\zeta^{k-1} a, \zeta^{k-1} b)$$

and chooses the one for which the relevant coefficient window has *smallest maximum value*. That is, if the coefficient list of the key is

$$(a_1, \dots, a_r), a_i \in \mathbb{F}_q$$

then we choose the sample for which $\max a_i$ is minimized (identifying $\mathbb{F}_q$ with $\{0, 1, \dots, q-1\}$).

The purpose of this keying is, first, to find matches between rotations of samples, not just samples themselves, and to reduce the likely number of keys observed before a collision. In particular, if $\max a_i \leq a$ with high probability, then a dictionary need only contain approximately a^B keys instead of q^B keys. However, a number of samples must be thrown away. In the next section we analyse the advantage.

9.2. **Expected Speedup.** The method of keying proposed is not unlike the underlying idea of coded BKW. By reducing the number of keys we are searching through, matches are found faster. In the case of coded BKW, a key is replaced with its nearest codeword, so that a collision is actually only a 'near match' of vectors, which has an error-inflating effect. In our case, a match is an exact match, so there is no cost in error growth. The keying speedup is also independent of the coding speedup, so that both could be applied simultaneously.

We now analyse the expected reduction in table size and number of samples of using advanced keying. The following analyses the reduction on the first block, if we are performing Ring-LWE in a finite field or, equivalently, working in the first CRT block.

The proposition is stated in terms of a threshold for keeping a sample; as this threshold lowers, the table size decreases, but a larger proportion of samples are thrown away. Therefore the key measure of the advantage of advanced keying is the number of samples needed to complete the table.

Proposition 9.1. *Performing a BKW reduction on a first block of size B within a finite field of dimension k, but keeping only samples for which the key entries are below ρq for some $0 < \rho < 1$, we expect a table-size reduction from q^B samples to $(\rho q)^B$ samples, and we expect the number of samples needed to fill the table to be reduced (from q^B) by a multiplicative factor which is asymptotically B/k (as B and k grow in fixed ratio).*

We will see that the speed of approach to this factor k/B is very fast for even moderate r.

Proof. We begin with the first block, of size B within dimension k.

Let (a, b) be a sample and for $i = 0, \dots, k-1$, denote the rotated sample by

$$(a_i, b_i) = (\zeta^i a, \zeta^i b).$$

Note that $a_i \in \mathbb{F}_q$. We will identify $\mathbb{F}_q$ with $\{0, 1, \dots, q-1\}$ for the purposes of the following.

Suppose that the key size is B and $B \mid k$. Write

$$a_{i,1}, a_{i,2}, a_{i,3}, \ldots, a_{i,B}$$

for the key corresponding to (a_i, b_i).

Let $0 \leq \rho \leq 1$. Then assuming the a are uniformly distributed,

$$\text{Prob}\left(\max_{1 \leq j \leq B} a_{i,j} < \rho q\right) = \rho^B.$$

Next, we claim that the probabilities above, for $0 \leq i < k/B$, are independent. This is because these k/B keys represent disjoint subsets of the full list of coefficients of a. This is a consequence of comparing the cyclic permutation on the full list of coefficients induced by multiplication by ζ, with the definition of the prioritized basis given above (here both defining properties of the basis are important).

Using the independence, we obtain

$$\begin{aligned}
&\text{Prob}\left(\min_{0 \leq i < k/B} \max_{1 \leq j \leq B} a_{i,j} < \rho q\right) \\
&= \text{Prob}\left(\max_{1 \leq j \leq B} a_{i,j} < \rho q \text{ for at least one } 0 \leq i < k/B\right) \\
&= 1 - (1 - \rho^B)^{k/B}.
\end{aligned}$$

Call this quantity η. Then, if we store only samples for which

$$\min_{0 \leq i < k/B} \max_{1 \leq j \leq B} a_{i,j} < \rho q,$$

by using that "small" key, we need only store $(\rho q)^B$ elements in the dictionary, but we use up on the order of $(\rho q)^B/\eta$ samples to obtain those elements. Therefore the number of samples needed to fill the table for one reduction block is

$$(\rho q)^B/\eta = \frac{\rho^B}{1 - (1 - \rho^B)^{k/B}} q^B.$$

The function

$$\frac{\rho^B}{1 - (1 - \rho^B)^{k/B}}$$

is very close to B/k in a wide range of the interval $(0, 1)$. More precisely, the Taylor expansion of $x/(1 - (1 - x)^t)$ around $x = 0$ is

$$\frac{1}{t} + \frac{t - 1}{2t} x + \frac{t^2 - 1}{12t} x^2 + \cdots.$$

As k and B grow in tandem, $x = \rho^B$ approaches 0 and this approaches $1/t = B/k$ very quickly. □ □

The problem with applying advanced keying to later blocks is that rotating engineered samples does not necessarily preserve the fact that their previous keys are zero. Sometimes some rotations do preserve this (e.g. when we have reached a subfield of $\mathbb{F}_{q^k}$ or the beginning of another CRT factor). Therefore advanced keying has a diminished effect. We do not analyse the exact overall effect here.

10. Block size

The runtime of one block reduction is exponential in the block size B, so that the block size is all important. In turn, block size is limited by the number of blocks, which determines the error inflation. If the errors are allowed to inflate too much, the final samples will be indistinguishable from uniform and no information is gained.

Since the samples passed from one block to the next are formed as two-term sums or differences of samples from the former block, we have the following standard result for Ring-BKW just as for LWE BKW.

Proposition 10.1. *Suppose one has Ring-LWE samples distributed according to $A_{s,\chi}$. Under one BKW reduction step (one block) in the Ring-BKW algorithm, the samples (a, b) which are produced have a chosen uniformly amongst a satisfying $\mathrm{key}(a) = 0$, and $b = as + e$ where e is distributed according to 2χ. In general, after c block reductions, the error distribution of the resulting samples is $2^c\chi$.*

Therefore, the optimal block sizes are essentially the same for Ring-BKW as for BKW. It is possible that Ring-BKW can use a very slightly smaller block size if the final step can be slightly larger (e.g. using a square-root speedup as in Corollary 5.2). We do not follow this up here.

10.1. Using a homomorphism to avoid some of the reduction.

It is tempting to consider using a CRT homomorphism to transfer samples to one CRT factor, and then start the BKW reduction there. This results in a smaller dimensional problem upon which to use BKW, but it incurs a penalty in error inflation. It also eliminats back-substitution. In this section we examine the tradeoff, and find that it is not advantageous unless possibly if the error width in the original problem is very small.

In the situation where we begin with a Ring-LWE problem in R_q with a discrete Gaussian distribution on the ζ-basis, we can determine the minimum block size under standard LWE BKW, and under applying Ring-BKW on a residue field. For the purposes of a simple comparison, let us suppose the discrete Gaussian tails are cut off at two standard deviations, so that $|E_\chi| = 4\alpha q$ in the notation of Section 2. In this case we have the following. For a more nuanced analysis, see [1].

Proposition 10.2. *Consider a Ring-LWE problem in R_q with discrete Gaussian distribution having parameter α, with tails cut off at two standard deviations. Suppose that R_q has residue field $\mathbb{F}_{q^k}$ where $k = n/\omega$. Then regular LWE BKW will succeed with block size at least*

$$b = \lceil n/(2\log_2(\alpha^{-1}) - 4)\rceil.$$

Provided that ω satisfies

$$8\alpha q^{\frac{\omega-1}{\omega}} < 1.$$

then Ring-BKW will succeed with block size at least

$$b = \lceil n/(2\omega \log_2(\alpha^{-1}) - 4\omega - 2(\omega-1)\log_2(q))\rceil.$$

Proof. We begin with regular LWE BKW. The fundamental restriction is that the error support be strictly smaller than all of $\mathbb{F}_q$ on the final samples. Using the fact that $|E_2\chi| = \sqrt{2}|E_\chi|$ for Gaussians, this requirement becomes

$$\sqrt{2}^c \cdot 4\alpha q < q,$$

This is equivalent to $\alpha^2 < 2^{-c-4}$, or

$$c < -2\log_2(\alpha) - 4.$$

With total dimension n, this is equivalent to

$$b = n/c = \lceil n/(2\log_2(\alpha^{-1}) - 4)\rceil.$$

For comparison, let us run the same naïve analysis on the Ring-BKW algorithm, applied to a finite residue field of degree $k = n/\omega$. Suppose χ_0 is a discrete Gaussian with tails cut at two standard deviations, as above (we use this notation on R_q and on $\mathbb{F}_{q^k}$, since in both cases the coefficient distributions are the same). Let $\rho : R_q \to \mathbb{F}_{q^k}$ be a ring homomorphism. If we take $\chi = \rho(\chi_0)$ on $\mathbb{F}_{q^k}$, then according to Proposition 4.1,

$$\chi = \sum_{i=0}^{\omega-1} \rho(\zeta_m^k)^i \chi_0$$

In particular,

$$2^c\chi = 2^c \sum_{i=0}^{\omega-1} \rho(\zeta_m^k)^i \chi_0$$

The support of this satisfies

$$|E_{2^c\chi}| \leq \prod_{i=0}^{\omega-1} |E_{2^c\rho(\zeta_m^k)^i\chi_0}| = 2^{c\omega/2}|E_{\chi_0}|^\omega.$$

Intuitively, we incur quite a high cost for moving to the finite field, since a sum of two different Gaussians behaves poorly under doubling the distribution, compared to a single Gaussian.

Continuing the analysis, our requirement is

$$2^{c\omega/2}(4\alpha q)^\omega < q$$

This is equivalent to

$$c < -2\log_2(\alpha) - 2\frac{\omega-1}{\omega}\log_2(q) - 4.$$

We need this to hold for $c = 2$, since otherwise BKW won't apply. (For example, if ω is too big, we have already spread the error out so much that its support is all of $\mathbb{F}_q$). That is the requirement that

$$0 < -2\log_2(\alpha) - 2\frac{\omega-1}{\omega}\log_2(q) - 6$$

or

$$8\alpha q^{\frac{\omega-1}{\omega}} < 1.$$

Assuming that is the case, we continue the analysis. We have total dimension $k = n/\omega$, so this requirement becomes

$$b = n/(\omega c) = \lceil n/(2\omega\log_2(\alpha^{-1}) - 4\omega - 2(\omega-1)\log_2(q))\rceil.$$

□ □

Therefore, Ring-BKW will have a smaller block size when

$$\omega\log_2(\alpha^{-1}) - 2\omega - (\omega-1)\log_2(q) > \log_2(\alpha^{-1}) - 2$$

i.e.

$$(\omega-1)\log_2(\alpha^{-1}) > (\omega-1)\log_2(q) + 2(\omega-1)$$

or, when $\omega > 1$,

$$1 > 4\alpha q.$$

This, of course, means that the original Gaussian is so narrow as to have coefficients which are mostly 0. This is certainly not in the range of security reductions. However, given that the block size must be an integer, it is possible that in the case of binary or trinary error, the block sizes may be equal. In this case performing BKW after a CRT homomorphism may be considered.

11. Practical Runtimes

It is evident that the runtime of Ring-BKW is similar to that of standard BKW, as both depend overwhelmingly on the block size B. However, we present some evidence that the practical runtime (e.g. constants) may be significantly improved in Ring-BKW. The salient points are:

(1) The block size of Ring-BKW is the same as for BKW.
(2) Back-substitution is eliminated or reduced. Instead, we use Theorem 6.3, which need only be performed on polynomially many samples, once per CRT factor.
(3) The smaller Ring-LWE problems posed in each CRT factor after BKW reduction can be performed in parallel.
(4) The square root speedup over exhaustive search given in Corollary 5.2 may decrease the block size slightly, or the number of blocks needed, since one need only reduce to a somewhat larger subfield.
(5) Advanced keying provides a significant table size reduction for the first block of each CRT factor and certain subsequent blocks.
(6) Speedups for classical BKW known in the literature, such as coded-BKW, appear to be applicable to Ring-BKW also.

Unfortunately, it is beyond the scope of this paper to verify that a speedup is in fact obtained.

The Ring-LWE Challenges [11] are in the form of *Tweaked Ring-LWE*, which refers to dual Ring-LWE transfered to the unital version (see [11, §2.3]), so that the parameter assumptions in this paper apply to the two-power cyclotomic challenges included therein. It would be very interesting to test these algorithms on those parameters. Note that, although the authors of the challenges include primes of a variety of arithmetic forms, none of the primes proposed for two-power cyclotomics appear to have residue degree above $n/2^7$.

References

[1] Albrecht, M.R., Cid, C., Faugère, J.C., Fitzpatrick, R., Perret, L.: On the complexity of the BKW algorithm on LWE. Des. Codes Cryptogr. 74(2), 325–354 (2015), https://doi.org/10.1007/s10623-013-9864-x

[2] Alkim, E., Ducas, L., Pöppelmann, T., Schwabe, P.: Post-quantum key exchange—a new hope. In: 25th USENIX Security Symposium (USENIX Security 16). pp. 327–343. USENIX Association, Austin, TX (2016), https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/alkim

[3] Blum, A., Kalai, A., Wasserman, H.: Noise-tolerant learning, the parity problem, and the statistical query model. J. ACM 50(4), 506–519 (2003), https://doi.org/10.1145/792538.792543

[4] Brakerski, Z., Langlois, A., Peikert, C., Regev, O., Stehlé, D.: Classical hardness of learning with errors (extended abstract). In: STOC'13—Proceedings of the 2013 ACM Symposium on Theory of Computing, pp. 575–584. ACM, New York (2013), https://doi.org/10.1145/2488608.2488680

[5] Brakerski, Z., Perlman, R.: Order-lwe and the hardness of ring-lwe with entropic secrets. Cryptology ePrint Archive, Report 2018/494 (2018), https://eprint.iacr.org/2018/494

[6] Brakerski, Z., Vaikuntanathan, V.: Fully homomorphic encryption from ring-LWE and security for key dependent messages. In: Advances in cryptology—CRYPTO 2011, Lecture Notes in Comput. Sci., vol. 6841, pp. 505–524. Springer, Heidelberg (2011), https://doi.org/10.1007/978-3-642-22792-9_29

[7] Brakerski, Z., Vaikuntanathan, V.: Efficient fully homomorphic encryption from (standard) LWE. SIAM J. Comput. 43(2), 831–871 (2014), https://doi.org/10.1137/120868669

[8] Castryck, W., Iliashenko, I., Vercauteren, F.: Provably weak instances of ring-LWE revisited. In: Advances in cryptology—EUROCRYPT 2016. Part I, Lecture Notes in Comput. Sci., vol. 9665, pp. 147–167. Springer, Berlin (2016), https://doi.org/10.1007/978-3-662-49890-3_6

[9] Chen, H., Lauter, K., Stange, K.E.: Attacks on the search RLWE problem with small errors. SIAM J. Appl. Algebra Geom. 1(1), 665–682 (2017), https://doi.org/10.1137/16M1096566

[10] Chen, H., Lauter, K., Stange, K.E.: Security considerations for Galois non-dual RLWE families. In: Selected areas in cryptography—SAC 2016, Lecture Notes in Comput. Sci., vol. 10532, pp. 443–462. Springer, Cham (2017)

[11] Crockett, E., Peikert, C.: Challenges for ring-lwe. Cryptology ePrint Archive, Report 2016/782 (2016), https://eprint.iacr.org/2016/782

[12] Eisenträger, K., Hallgren, S., Lauter, K.: Weak instances of plwe. In: Selected Areas in Cryptography–SAC 2014, pp. 183–194. Springer (2014)

[13] Elias, Y., Lauter, K., Ozman, E., Stange, K.: Provably weak instances of ring-lwe. In: Advances in Cryptology – CRYPTO 2015, Lecture Notes in Comput. Sci., vol. 9215, pp. 63–92. Springer, Heidelberg (2015)

[14] Guo, Q., Johansson, T., Må rtensson, E., Stankovski, P.: Coded-BKW with sieving. In: Advances in cryptology—ASIACRYPT 2017. Part I, Lecture Notes in Comput. Sci., vol. 10624, pp. 323–346. Springer, Cham (2017)

[15] Guo, Q., Johansson, T., Stankovski, P.: Coded-BKW: solving LWE using lattice codes. In: Advances in cryptology—CRYPTO 2015. Part I, Lecture Notes in Comput. Sci., vol. 9215, pp. 23–42. Springer, Heidelberg (2015), https://doi.org/10.1007/978-3-662-47989-6_2

[16] Kirchner, P., Fouque, P.A.: An improved BKW algorithm for LWE with applications to cryptography and lattices. In: Advances in cryptology—CRYPTO 2015. Part I, Lecture Notes in Comput. Sci., vol. 9215, pp. 43–62. Springer, Heidelberg (2015), https://doi.org/10.1007/978-3-662-47989-6_3

[17] Lyubashevsky, V., Peikert, C., Regev, O.: On ideal lattices and learning with errors over rings. In: Advances in cryptology—EUROCRYPT 2010, Lecture Notes in Comput. Sci., vol. 6110, pp. 1–23. Springer, Berlin (2010), https://doi.org/10.1007/978-3-642-13190-5_1

[18] Lyubashevsky, V., Peikert, C., Regev, O.: A toolkit for ring-LWE cryptography. In: Advances in cryptology—EUROCRYPT 2013, Lecture Notes in Comput. Sci., vol. 7881, pp. 35–54. Springer, Heidelberg (2013), https://doi.org/10.1007/978-3-642-38348-9_3

[19] Peikert, C.: How (not) to instantiate ring-LWE. In: Security and cryptography for networks, Lecture Notes in Comput. Sci., vol. 9841, pp. 411–430. Springer, [Cham] (2016), https://doi.org/10.1007/978-3-319-44618-9_22

[20] Regev, O.: On lattices, learning with errors, random linear codes, and cryptography. In: STOC'05: Proceedings of the 37th Annual ACM Symposium on Theory of Computing, pp. 84–93. ACM, New York (2005), https://doi.org/10.1145/1060590.1060603

Department of Mathematics, University of Colorado, Campux Box 395, Boulder, Colorado 80309-0395

E-mail address: kstange@math.colorado.edu