On the Independence Assumption in Quasi-Cyclic Code-Based Cryptography

Maxime Bombar[1], Nicolas Resch[2], and Emiel Wiedijk[2]

[1]Cryptology Group, CWI, Amsterdam, The Netherlands
Institut de Mathématiques de Bordeaux, France
maxime.bombar@math.u-bordeaux.fr

[2]University of Amsterdam, The Netherlands
{n.a.resch,e.wiedijk}@uva.nl

January 5, 2025

Abstract

Cryptography based on the presumed hardness of decoding codes – i.e., code-based cryptography – has recently seen increased interest due to its plausible security against quantum attackers. Notably, of the four proposals for the NIST post-quantum standardization process that were advanced to their fourth round for further review, two were code-based. The most efficient proposals – including HQC and BIKE, the NIST submissions alluded to above – in fact rely on the presumed hardness of decoding *structured* codes. Of particular relevance to our work, HQC is based on *quasi-cyclic codes*, which are codes generated by matrices consisting of two cyclic blocks.

In particular, the security analysis of HQC requires a precise understanding of the Decryption Failure Rate (DFR), whose analysis relies on the following heuristic: given random "sparse" vectors e_1, e_2 (say, each coordinate is i.i.d. Bernoulli) multiplied by fixed "sparse" quasi-cyclic matrices A_1, A_2, the weight of resulting vector $e_1A_1 + e_2A_2$ is very concentrated around its expectation. In the documentation, the authors model the distribution of $e_1A_1 + e_2A_2$ as a vector with *independent* coordinates (and correct marginal distribution). However, we uncover cases where this modeling fails. While this does not invalidate the (empirically verified) heuristic that the weight of $e_1A_1 + e_2A_2$ is concentrated, it does suggest that the behavior of the noise is a bit more subtle than previously predicted. Lastly, we also discuss implications of our result for potential worst-case to average-case reductions for quasi-cyclic codes.

1 Introduction

In light of recent calls for post-quantum secure cryptography – i.e., cryptography that is secure in a world with quantum computers – code-based cryptography has recently seen a growth in interest as a prominent candidate for quantum-safe cryptography. In particular, all three remaining finalists in the 4th round of NIST competition are code-based [Agu+22a; Agu+22b; Alb+22]. Informally, code-based cryptographic schemes are those whose security can be reduced to the conjectured hardness of decoding linear codes under the Hamming metric.

More concretely, the quintessential hard problem for code-based cryptography is the decoding problem (also sometimes called *learning parity with noise (LPN)*), which asks one to recover $\boldsymbol{s}$ from the input $(\boldsymbol{A}, \boldsymbol{sA} + \boldsymbol{e})$,[1] where $\boldsymbol{A} \leftarrow \mathbb{F}_2^{n\times \boldsymbol{m}}$ and $\boldsymbol{s} \leftarrow \mathbb{F}_2^n$ are uniformly distributed, and $\boldsymbol{e} \in \mathbb{F}_2^m$ is a *noise* vector where each e_i is an independent Bernoulli variable (i.e., it is 1 with probability p and 0 with probability $1 - p$).[2]

This problem inspired the closely related *learning with errors (LWE)* problem, which is at the core of lattice-based cryptography. Here, the noise $\boldsymbol{e}$ is sampled differently, typically as a (rounding of) a Gaussian random variable (one is also required to work over a large field).

When constructing public-key cryptography from either LPN or LWE, the matrix $\boldsymbol{A}$ always forms (a part of) the public key. Thus, one is required to publish at least nk field elements: this quadratic lower bound on the public-key size often renders these schemes uncompetitive in terms of efficiency. To remedy this situation, it has been proposed to instead sample *structured* matrices $\boldsymbol{A}$: for such matrices, it is still plausible (and widely believed) that (quantum) algorithms cannot efficiently solve the relevant decoding problem; however, their structure allows for a much more succinct representation, ideally with only n field elements. This is precisely the approach taken by many NIST submissions [Agu+22a; Agu+22b]. In particular, these schemes are based on *quasi-cyclic codes*, which we now introduce.

Quasi-cyclic codes. A quasi-cyclic code is a code that is generated by a matrix composed of multiple blocks of circulant submatrices, *i.e.,* matrices such that each row is a circular shift of its first row. Consequently, each

[1] Technically, we are describing the search version of LPN. For cryptographic purposes a decision variant is often required, which states that given $\boldsymbol{A}$ distinguishing $\boldsymbol{sA} + \boldsymbol{e}$ from a uniformly random vector is hard. However, due to a search-to-decision reduction [FS96], they are polynomially equivalent.

[2] In fact, for technical reasons, it is often easier to consider LPN as an *oracle* problem, as we do later. The complexity of the two variants are polynomially related, so they are interchangeable for our purposes.

submatrix can be represented by storing only its first row. For instance, in quasi-cyclic codes of rate 1/2, only $2n$ field elements need to be stored.

One important advantage of quasi-cyclic codes in cryptographic applications is their polynomial representation. Specifically, let $\mathcal{R} \overset{\text{def}}{=} \mathbb{F}_q[X]/(X^n - 1)$. A circulant matrix of the form

$$\mathbf{M}_a \overset{\text{def}}{=} \begin{pmatrix} a_0 & a_1 & \cdots & \cdots & a_{n-1} \\ a_{n-1} & a_0 & \cdots & \cdots & a_{n-2} \\ \vdots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ a_1 & a_2 & \cdots & a_{n-1} & a_0 \end{pmatrix} \in \mathbb{F}_q^{n \times n}$$

represents the endomorphism $P(X) \in \mathcal{R} \mapsto a(X) \cdot P(X) \in \mathcal{R}$ in the monomial basis, where $a(X) = \sum_{i=0}^{n-1} a_i X^i$.

For example, an instance $(\boldsymbol{A}, \boldsymbol{s}\boldsymbol{A} + \boldsymbol{e})$ where $\boldsymbol{A}$ is of the form

$$\mathbf{A} \overset{\text{def}}{=} \begin{pmatrix} \mathbf{M}_{a_{1,1}} & \cdots & \mathbf{M}_{a_{1,r}} \\ \vdots & \ddots & \vdots \\ \mathbf{M}_{a_{\ell,1}} & \cdots & \mathbf{M}_{a_{\ell,r}} \end{pmatrix}$$

can be compactly represented by a collection of r samples of the form $(\mathbf{a}, \langle \mathbf{s}, \mathbf{a} \rangle + \mathbf{e})$, where $\mathbf{a}$ is a vector of ℓ polynomials in $\mathcal{R}$, and $\langle \mathbf{s}, \mathbf{a} \rangle = \sum_{i=0}^{\ell-1} s_i(X) \cdot a_i(X)$. For cryptographic applications, it is common to consider the case where $\ell = 1$, resulting in samples of the form $(a, a \cdot s + e)$, where $a \in \mathcal{R}$.

Such quasi-cyclic codes are employed in the NIST submissions HQC [Agu+22b] and BIKE [Agu+22a], both of which have advanced to the fourth round of the post-quantum cryptography standardization process.

Analysis of noise In the analysis of error vector distribution of HQC [Agu+22b, par. 2.4], one requires an understanding of the product of polynomials $t(X)$ and $R(X)$, where $t(X) \in \mathbb{F}_2[X]/(X^n - 1)$ is a fixed polynomial and $R(X)$ is a polynomial whose coefficients are independently Bernoulli distributed. In fact, they require the analysis of two independent copies of such products $e(X) \overset{\text{def}}{=} t_1(X)R_1(X) + t_2(X)R_2(X)$. To make the analysis tracetable, the authors make the simplifying assumption that the coefficients of this $e(X)$ are independent.

In this work we reconsider this assumption. To set up our result, we quickly introduce some notation. We write $X \leftarrow \mathrm{Ber}(\omega)$ to denote a $\mathbb{F}_2$-valued random variable such that $\Pr[X = b] = \frac{1+(-1)^b 2^{-\omega}}{2}$.[3]

[3] Below, we justify this parametrization for the Bernoulli random variable.

Theorem 1.1 (Main Theorem (Informal); see Theorem 3.4). Let $t(X) \in \mathbb{F}_2[X]/(X^n - 1)$ be a fixed polynomial with τ nonzero coefficients, and let $R(X) = \sum_{i=0}^{n-1} R_i X^i$ be such that each $R_i \leftarrow \mathrm{Ber}(\omega)$ independently. Let $I(X) = \sum_{i=0}^{n-1} I_i X^i$ where each $I_i \leftarrow \mathrm{Ber}(\tau\omega)$ independently. Assuming $\omega \geq \Omega(\log n)$, the statistical distance between $t(X)R(X)$ and $I(X)$ is $\Omega(\sqrt{n}2^{-2\omega})$.

We note that, by an application of the Piling-up lemma (Lemma 2.1), it follows that $t(X)R(X)$ and $I(X)$ share the same marginal distribution for each coordinate: *i.e.,*, for each $0 \leq i \leq n-1$ and $b \in \mathbb{F}_2$ we have $\Pr[(tR)_i = b] = \frac{1+(-1)^b 2^{-\omega}}{2} = \Pr[I_i = b]$. Hence, the "source" of the statistical distance is the lack of independence between the coordinates of $t(X)R(X)$. Furthermore, we remark that [Agu+22b, par 2.4] considers (in our parametrization) $\omega = \Theta(\log n)$, i.e., the lower bound is indeed met.

Now, note that this does not directly invalidate the modelling of HQC: there, they consider two *independent* polynomials $t_1(X)R_1(X)$ and $t_2(X)R_2(X)$, and then model their *sum*. While we cannot invalidate this modelling (and in fact, we suspect it might in most cases be valid), we do point out some cases where the modelling fails:

- Suppose $\langle t_1, t_2 \rangle \subsetneq \mathbb{F}_2[X]/(X^n - 1)$, that is the ideal generated by the noise does not span the entire space. Then $t_1(X)R_1(X) + t_2(X)R_2(X)$ can never be statistically close to a Cartesian product of Bernoulli polynomials, as they do span the entire space. Under some reasonable assumptions polynomials of odd weight are invertible, so in practice this case is easy to avoid.

- As an extreme case, if $t_1(X) = t_2(X) =: t(X)$, then $t_1(X)R_1(X) + t_2(X)R_2(X) = t(X)(R_1(X) + R_2(X))$, and since $R_1 + R_2$ can again be modelled by an independent Bernoulli polynomial, Theorem 3.4 applies.

- Suppose now that the support sets of t_1 and t_2 (i.e., the indices of the nonzero coefficients) are both in *arithmetic progressions* – i.e., sets of the form $\{ax + b \pmod{n} : x \in \{0, 1, \ldots, \tau - 1\}\}$ – with the same common difference a. Then we can still show a nontrivial lower bound on the statistical distance.

For less "structured" cases of t_1, t_2 (or even, say, $t_1, \ldots, t_s$ with $s \geq 2$) we conjecture that such a gap in the statistical distance does not persist.

Related independence heuristic in cryptography The question of the independence of marginals of the distribution of a product of polynomials is not restricted to the analysis of code-based cryptosystems. In particular, a similar assumption has also been made in lattice-based cryptography to analyse the growth of the noise in the context of *fully hommomorphic encryption* (FHE) [Chi+20, Assumption 3.11]. Nevertheless, recent works have began to suggest that this did not actually hold [Bia+23; MP24], which led to underestimating this noise growth. Our results align with those observation.

1.1 Overview of Techniques

In order to lower bound the statistical distance, we in fact find it easier to work with the *Kullback-Leibler (KL) divergence* between the two distributions. Pinsker's inequality shows that these two quantities are intimately related; however, as we are seeking a *lower bound* on the statistical distance, this inequality is not directly applicable for our purposes. Fortunately, under mild "regularity" conditions (namely, the ratio between the two considered probability distributions is never too large nor too small), we can apply a "reverse" Pinsker's inequality [Bin19]. As these regularity conditions hold for our distributions of interest (see Section 3.2), we can focus on the KL divergence.

We begin with a convenient lemma that may be folklore, but for lack of a suitable reference (and because we believe it might be of independent interest) we provide a proof (Lemma 3.4): namely, that if Q and P are distributions over an n-fold Cartesian product with matching marginal distributions and Q follows a product distribution (*i.e.,* its coordinates are independent), then the KL-divergence $D(P\|Q) = H(Q) - H(P)$, where $H(\cdot)$ is the *(Shannon) entropy* of the distributions. For our purposes, P denotes the distribution of $t(X)R(X)$ and Q denotes the distribution of $I(X)$, where we identify their support $\mathbb{F}_2[X]/(X^n-1)$ with $\mathbb{F}_2^n$ via the natural isomorphism. Hence, as $H(Q)$ is easy to compute (being a Cartesian product of Bernoulli distributions), we focus on upper bounding $H(P)$.

Here, we can consider two cases. Firstly, if $t(X)$ happens to not be invertible, then already $t(X)R(X) \in \langle t(X)\rangle \subsetneq \mathbb{F}_2[X]/(X^n-1)$, where $\langle t(X)\rangle = \{t(X)a(X) : a(X) \in \mathbb{F}_2[X]/(X^n-1)\}$ is the ideal generated by $t(X)$. That is, P is distributed over a strict subset of $\mathbb{F}_2[X]/(X^n-1)$ size at most 2^{n-1}; this already guarantees a $H(P) \leq n-1 \ll H(Q)$ for our parameters of interest.

Otherwise, $t(X)$ is invertible. Then, it naturally follows that $H(P) = H(R(X))$, i.e., just the distribution of $R(X)$. Indeed, multiplying by $t(X)$ is then a bijection from $\mathbb{F}_2[X]/(X^n-1)$ to itself, so it does not affect the entropy. And we can again easily compute $H(R(X))$: it is again a Cartesian product of Bernoulli distributions! That is, we can conclude

$$H(Q) - H(P) = n(\tilde{h}(\tau\omega) - \tilde{h}(\omega)) ,$$

where $\tilde{h}(x)$ is the entropy of a $\mathrm{Ber}(x)$ random variable, and we recall τ is the number of nonzero coefficients of $t(X)$. To conclude our desired theorem, it suffices to lower bound $\tilde{h}(\tau\omega) - \tilde{h}(\omega)$, which we do by expanding the Taylor series representation of $\tilde{h}$.

Next, we consider cases where we can understand the entropy of $t_1R_1 + \cdots + t_sR_s$, where $t_1, \ldots, t_s$ are fixed polynomials and $R_1, \ldots, R_s$ are independent Bernoulli polynomials (i.e., their coefficients are sampled independently). To make progress in this case, we write $t_1R_1 + \cdots + t_sR_s = \sum_i C_iX^i$ and bound the entropy via a sum over roughly $n/2$ pairwise entropies $H(C_i, C_j)$. The formula for the joint distribution of C_i and C_j is not too difficult to obtain (and in fact has been obtained by prior work [PGS16]), and one can observe that the joint entropies $H(C_i, C_j)$ are small if for many of the t_ℓ's, many of its nonzero coefficients overlap with many nonzero coefficients of $X^{j-i}t_\ell$. If $t_1, \ldots, t_s$ are all of the form $\sum_{i=0}^{\tau-1} X^{a \cdot i+b}$ where $a \in \mathbb{Z}_n^*$ and $b \in \mathbb{Z}_n$ (i.e., the nonzero coefficients form an *arithmetic progression*) we can show that the entropy bound will indeed be quite small.

1.2 Future Directions

We conclude the introduction with some directions that we leave open for future work.

Concentration of noise weight. In this work, we provided an analysis of $t(X)R(X)$ – the product of a fixed polynomial and an i.i.d. Bernoulli polynomial – and showed that it is "far" from the distribution of an i.i.d. Bernoulli polynomial. As discussed above, this has implications for code-based cryptosystems such as HQC, where in order to allow for successful decoding it is important that the weight of the noise $t_1(X)R_1(X) + t_2(X)R_2(X)$ be tightly concentrated around its expected value. While such a concentration naturally follows if the coordinates were indeed independent, but as we showed here in some cases that does not hold. However, this does not itself disprove the assertion that the weight is concentrated, and indeed empirical evidence suggests that the weight is sufficiently concentrated. Additionally, prior work [Kaw24] already gave some concentration bounds (in this case, via Chebyshev's inequality). We leave it as an open problem to provide further theoretical evidence for the concentration of this weight.

Potential for worst-case to average-case reduction? For cryptographic purposes, it is of course vital that the hardness assumptions hold for *average-case problems*: namely, it is hard to solve some computational problem (such as the decoding problem) when the instances are sampled randomly. However, from a complexity-theoretic perspective we have a much firmer theory of the hardness of *worst-case problems.* That is, we have a more mature theory of which problems are hard when the instances for a given algorithm are chosen *adversarially.*

In the case of LWE, Regev [Reg05; Bra+13] famously showed that the average-case LWE problem can be reduced to certain worst-case problems on lattices. Inspired by this, Brakerski et al. [Bra+19] recently introduced a worst-case to average-case reduction for codes: namely, a reduction from the classic worst-case decoding problem where $\boldsymbol{A}$, $\boldsymbol{s}$ and $\boldsymbol{e}$ are adversarially chosen, with the promise that $\boldsymbol{e}$ has Hamming weight at most τ. One can imagine generalizing this to the quasi-cyclic case, as was successfully done in the case of lattices (the analogous problem there is typically termed Ring-LWE). Here, if one works with a rate $1/s$ quasi-cyclic code the natural reduction strategy takes as input a noisy codeword $(a_1(X)m(X) + t_1(X), \ldots, a_s(X)m(X) + t_s(X))$ (with the sum of the weights of the noise vectors t_i being at most τ), and then produces "Ring-LPN" like samples by sampling a "smoothing" vector $(R_1(X), \ldots, R_s(X))$ and considering

$$\left(\sum_{i=1}^{s} a_i(X)R_i(X), \left(\sum_{i=1}^{s} a_i(X)R_i(X)\right) m(X) + \sum_{i=1}^{s} t_i(X)R_i(X)\right) .$$

At the very least, this requires us to analyze the distribution $\sum_{i=1}^{s} t_i(X)R_i(X)$, as we undertake in this work.[4] Furthermore, for the *standard* Ring-LPN assumption, one must have $\sum_{i=1}^{s} t_i(X)R_i(X)$ close to an independent Bernoulli polynomial. Our work points out that at least some structural assumptions must be made on the vectors $t_1, \ldots, t_s$: for example, if they are all equal, then this reduction is doomed to fail as $\sum_{i=1}^{s} t_i(X)R_i(X)$ is necessarily far from an independent Bernoulli polynomial. Furthermore, if each vector $t_1, \ldots, t_s$ form an *arithmetic progression* with the same common difference, we can also show that $\sum_{i=1}^{s} t_i(X)R_i(X)$ is far from an independent Bernoulli polynomial.

When is statistical distance small? We suspect that, given appropriate assumptions, $\sum_{i=1}^{s} t_i(X)R_i(X)$ is statistically close to an independent Bernoulli polynomial. Recall these assumptions are equivalent to the conditions such that the entropy

$$H\left(\sum_{i=1}^{s} t_i(X)R_i(X)\right)$$

is sufficiently high. While we have found conditions on the noise that ensure that this entropy is low, we currently do not know any conditions that yield a high entropy. In general, computing the entropy of the sum independent random variables is hard, refer for example to [Tao10; GMT24]. To the best of our knowledge, computing the entropy of this polynomial in general is an open problem.

[4] More precisely, we make a step assuming one is choosing the Bernoulli distribution to smooth. Other choices could be made, but we view this as a natural first step.

The conditions for which the entropy $H(\sum_{i=1}^{s} t_i(X)R_i(X))$ is known to be small have an important caveat: the relevant worst case decoding problem

$$(a_1(X)m(X) + t_1(X), \ldots, a_s(X)m(X) + t_s(X))$$

is in fact *easy*. If all noise vector vectors are identical, then we can decode $(a_1(X)m(X) + t(X), a_2(X)m(X) + t(X))$. By computing the difference $((a_1(X) + a_2(X))m(X))$ it is easy to decode to $m(X)$, if $a_1(X) + a_2(X)$ is invertible. Decoding is also easy when the noise is guaranteed to be an arithmetic progression. As the number of arithmetic progressing noise vectors is polynomial (when s is fixed), it is easy to brute force all possible noise vectors. We remain especially interested in finding the entropy $H(\sum_{i=1}^{s} t_i(X)R_i(X))$ for cases where the worst-case problem is assumed to be hard.

2 Preliminaries

General notation. For positive integer n, we write $[n] = \{1, 2, \ldots, n\}$.

We choose a (somewhat) nonstandard definition for the Bernoulli random variable: for $\omega \in [0, +\infty]$, we say $x \leftarrow \mathrm{Ber}(\omega)$ if $x \in \mathbb{F}_2$ and

$$\Pr[x = b] = \begin{cases} \frac{1-2^{-\omega}}{2} & b = 1 \\ \frac{1+2^{-\omega}}{2} & b = 0 \end{cases}$$

In other words, ω is the log of the bias of the Bernoulli. For positive integer n we let $\mathrm{Ber}(\omega)^{\otimes n}$ to denote a vector $(x_1, \ldots, x_n)$, where each $x_i \leftarrow \mathrm{Ber}(\omega)$ independently. We sometimes abuse notation and write $R(X) \leftarrow \mathrm{Ber}(\omega)^{\otimes n}$ to mean that $R(X) = \sum_{i=0}^{n-1} R_i X^i$ and each $R_i \leftarrow \mathrm{Ber}(\omega)$. We choose this parametrization as the statement of the *piling-up lemma* – which determines the distribution of the sum modulo 2 of Bernoulli random variables – becomes very simple.

Lemma 2.1 (Piling-up lemma). *Let $X \leftarrow \mathrm{Ber}(\omega_x)$ and $Y \leftarrow \mathrm{Ber}(\omega_y)$ be independent random variables. Then $X + Y \leftarrow \mathrm{Ber}(\omega_x + \omega_y)$.*

For distributions P and Q over Ω, we use the following notation for information-theoretic quantities:

$$H(P) \overset{\text{def}}{=} \sum_{x\in\Omega} P(x)\log(1/P(x)) \qquad \text{(entropy)}$$

$$D(P\|Q) \overset{\text{def}}{=} \sum_{x\in\omega} P(x)\log(P(x)/Q(x)) \qquad \text{(Kullback-Leibler divergence)}$$

$$\Delta_{\mathrm{tv}}(P,Q) \overset{\text{def}}{=} \frac{1}{2}\sum_{x\in\omega} |P(x) - Q(x)| \qquad \text{(statistical distance)}$$

For convenience, we abuse notation by writing random variables instead of the distribution of these random variables. For example, when $X \leftarrow P$, we write $H(X)$ instead of $H(P)$.

For the binary entropy we write $h(x) \overset{\text{def}}{=} -x\log(x) - (1-x)\log(1-x)$ for the binary entropy. We additionally write $\tilde{h}(\omega)$ for the entropy of a $\mathrm{Ber}(\omega)$ random variable, so

$$\tilde{h}(\omega) \overset{\text{def}}{=} h\left(\frac{1-2^{-\omega}}{2}\right) .$$

Furthermore, we write

$$p(\omega) \overset{\text{def}}{=} \frac{1-2^{-\omega}}{2},$$

for the probability of sampling 1 in the distribution $\mathrm{Ber}(\omega)$.

3 Analysis

The general goal of this article is to analyze the distribution of $t(X)R(X)$. Here $t(X)$ is some fixed polynomial in $\mathbb{F}_2[X]/(X^n-1)$, and $R(X) \leftarrow \mathrm{Ber}(\omega)^{\otimes n}$. Specifically, we want to know when the coefficients of this polynomial are close to independent. First, we give the marginal distribution of these coefficients, which is certainly folklore, but we state as a lemma for future convenience.

Lemma 3.1. Let $P_n := \mathbb{F}_2[X]/(X^n-1)$ be the polynomial quotient ring. Let $t \in P_n$ be a fixed polynomial, let $R \leftarrow \mathrm{Ber}(\omega)^{\otimes n}$ be a random variable.

Write $(tR)_i$ for the coefficient before X^i so

$$tR = \sum_{i=0}^{n-1} (tR)_i X^i.$$

Then for all k in $\{0, \ldots, n-1\}$

$$(tR)_k = \sum_{\substack{j \in \{0,\ldots,n-1\} \\ t_{k-j}=1}}^{n-1} R_j,$$

where for convenience, computations in the indices are modulo n.

Proof. The lemma follows from a simple rewriting of the polynomial

$$
\begin{aligned}
tR &= \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} t_i R_j X^{i+j} \\
&= \sum_{k=0}^{n-1}(\sum_{j=0}^{n-1} t_{k-j} R_j) X^k && \text{(let } k = i+j\text{)} \\
&= \sum_{k=0}^{n-1}\left(\sum_{\substack{j\in\{0,\ldots,n-1\} \\ t_{k-j}=1}}^{n-1} R_j \right) X^k. && \square
\end{aligned}
$$

From this lemma we can easily compute the marginal distributions of each coefficient.

Lemma 3.2. Let R and t as in Lemma 3.1. Then

$$(tR)_k \leftarrow \mathrm{Ber}(\|t\|\omega)$$

Proof. From Lemma 3.1 it follows for all $k \in \{0, \ldots, n-1\}$ that

$$(tR)_k = \sum_{\substack{j\in\{0,\ldots,n-1\} \\ t_{k-j}=1}}^{n-1} R_j.$$

Therefore, $(tR)_k$ is the sum mod 2 of $\|t\|$ independent Bernoulli variables. The piling-up lemma (Lemma 2.1) gives us that $(tR)_k \leftarrow \mathrm{Ber}(\|t\|\omega)$. □

Now Lemma 3.2 perfectly characterizes the marginal distribution of the coefficients Rt. However, this lemma does not imply that $Rt \leftarrow \mathrm{Ber}(\|t\|\omega)^{\otimes n}$. This statement would hold if all the coefficients of Rt were independent. Unfortunately, the coefficients are not independent. Different coefficients of Rt depend on the same coefficients in R. Specifically the intersection between

$$\{j \in \{0, \ldots, n-1\} \mid t_{k-j} = 1\} \cap \big\{j \in \{0, \ldots, n-1\} \mid t_{k'-j} = 1\big\}$$

coefficients in R may very well be non-empty creating a dependence between $(Rt)_k$ and $(Rt)_{k'}$.

One property we can immediately infer from the marginal distribution is the expectation.

Lemma 3.3 (Expectation of $|tR|$). Let $t, R \in \mathbb{F}_2[X]/(X^n-1)$ with t fixed and $R \leftarrow \mathrm{Ber}(\omega)^{\otimes n}$, then

$$\mathbb{E}[|tR|] = n \cdot \frac{1 - 2^{-\|t\|\omega}}{2}$$

Proof. The weight $|tR|$ is the sum of the coefficients of tR. Then we can use the linearity of the expectation to compute the sum of the expectations in $\mathbb{Z}$. Note the addition here is defined over $\mathbb{R}$, different from the previous computations where addition was defined over $\mathbb{F}_2$. Because every $(tR)_i \leftarrow \mathrm{Ber}(\|t\|\omega)$

$$\begin{aligned}\mathbb{E}[|tR|] &= \mathbb{E}\left[\sum_{i=0}^{n-1}(tR)_i\right] \\ &= \sum_{i=0}^{n-1}\mathbb{E}[(tR)_i] \\ &= n \cdot \frac{1 - 2^{-\|t\|\omega}}{2}.\end{aligned}$$

□

Because the coefficients are dependent, many common methods to analyze the probability that the weight $|tR|$ is close to the expectation do not apply. In [Kaw24] an analysis is provided showing that the $|tR|$ is indeed somewhat concentrated around its mean (essentially by analyzing the variance of $|tR|$ and then applying Chebyshev's inequality).

However, this dependence between the coefficients $(tR)_i$ turns out to be significant, at least in the sense that the statistical distance between tR and a Bernoulli distribution will be non-negligible. First let us give some necessary conditions for tR to look like a Bernoulli distribution. A Bernoulli distribution will reach every polynomial with non-zero probability. Specifically there is a non-zero probability that $(tR)(X) = 1$, so t must be invertible in P_n.

So it is necessary for $t(X)$ to be invertible, if $t(X)R(X)$ should look like a Bernoulli distribution. If $R(X)$ is unbiased enough, then $t(X)R(X)$ will look like a Bernoulli distribution. In the extreme case: if $R(X)$ is uniform over $\mathbb{F}_2[X]/(X^n - 1)$, then $t(X)R(X)$ will also be uniform. In Theorem 3.1 below, we discuss how low the bias of $R(X)$ can be for $t(X)R(X)$ to look like a Bernoulli distribution.

In order to analyze the statistical distance between $t(X)R(X)$ and a independent Bernoulli polynomial, as mentioned in the introduction we prefer to analyze the KL-divergence, which the following lemma states has a relatively simple form. This result is quite likely folklore, but for lack of a good reference, we provide a proof.

Lemma 3.4. Let $P : \mathcal{X} \to [0, 1]$ be a discrete distribution over $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_n$, with $P_1, \ldots P_n$ the marginal distributions. Define

$$Q \overset{\text{def}}{=} P_1 \otimes \ldots \otimes P_n$$

the distribution, such that the marginal distributions of Q and P agree (so $Q_i = P_i$), and the marginal distributions of Q are independent. Then

$$D(P\|Q) = H(Q) - H(P).$$

Proof. It is a well-known fact that we can write the KL-divergence as

$$\begin{aligned}
D(P\|Q) &= \sum_{x\in\mathcal{X}} P(x)\log(P(x)/Q(x)) \\
&= \sum_{x\in\mathcal{X}} P(x)\log(P(x)) - \sum_{x\in\mathcal{X}} P(x)\log(Q(x)) \\
&= H(P:Q) - H(P).
\end{aligned}$$

Where $H(P:Q)$ is defined to be the *cross-entropy* $-\sum_{x\in\mathcal{X}} P(x)\log(Q(x))$. We only need to show the equality $H(P:Q) = H(Q)$. To demonstrate this, it is convenient to introduce a random vector $X = (X_1, \ldots, X_n)$ such that $\Pr[X = x] = P(x)$ for all $x \in \mathcal{X}$. We also write $\mathcal{X}_{\neq k} := \mathcal{X}_1 \times \cdots \times \mathcal{X}_{k-1} \times \mathcal{X}_{k+1} \times \cdots \times \mathcal{X}_n$, i.e., it is the cartesian product of all the $\mathcal{X}_i$'s *except* $\mathcal{X}_k$. We then define the notation

$$P_{\neq k|k}(x_{\neq k}|x_k) := \Pr[X_{\neq k} = x_{\neq k}|X_k = x_k]$$

where $X_{\neq k} = (X_1, \ldots, X_{k-1}, X_{k+1}, \ldots, X_n)$, $x_{\neq k} \in \mathcal{X}_{\neq k}$ and $x_k \in \mathcal{X}_k$.

We can now derive

$$\begin{aligned}
& H(P:Q) \\
&= \sum_{x\in\mathcal{X}} P(x_1, \ldots, x_n)\log(1/Q(x_1, \ldots x_n)) \\
&= \sum_{x\in\mathcal{X}} P(x_1, \ldots, x_n)\sum_{k=1}^{n}\log(1/Q_k(x_k)) && \text{(as } Q_1, \ldots Q_n \text{ independent)} \\
&= \sum_{k=1}^{n}\sum_{x_k\in\mathcal{X}_k} P_k(x_k)\log(1/Q_k(x_k)) \sum_{x_{\neq k}\in\mathcal{X}_{\neq k}} P_{\neq k|k}(x_{\neq k} \mid x_k) && \text{(def. conditional prob.)} \\
&= \sum_{k=1}^{n}\sum_{x_k\in\mathcal{X}} P_k(x_k)\log(1/Q_k(x_k)) && \text{(sum prob. is 1)} \\
&= \sum_{k=1}^{n}\sum_{x_k\in\mathcal{X}} Q_k(x_k)\log(1/Q_k(x_k)) && \text{(as } P_k = Q_k\text{)} \\
&= \sum_{k=1}^{n} H(Q_k) && \text{(def. entropy)} \\
&= H(Q) && \text{(as } Q_1, \ldots Q_n \text{ independent)}
\end{aligned}$$

□

The following theorem is now a simple consequence of the above lemma.

Theorem 3.1 (Dependence of tR). Let $t(X), R(X) \in \mathbb{F}_2[X]/(X^n - 1)$ with an invertible t fixed and $R \leftarrow \mathrm{Ber}(\omega)^{\otimes n}$. Define $I \in \mathbb{F}_q[X]/(X^n - 1)$ with $I \leftarrow \mathrm{Ber}(\|t\|\omega)$. In other words the coefficients of I are independent, and have the same marginal distribution as the coefficients of $(tR)(X)$. Then

$$D(tR\|I) = n\Big(\tilde{h}(\|t\|\omega) - \tilde{h}(\omega)\Big).$$

Proof. This statement is a simple consequence of Lemma 3.4. The entropy $H(\mathrm{Ber}(\omega)) = \tilde{h}(\omega)$. So, the entropy of copies is $H(I) = n\tilde{h}(\|t\|\omega)$. As t is invertible we now that $H(tR) = H(R) = n\tilde{h}(\omega)$. In total

$$D(tR\|I) = H(I) - H(tR). \qquad \square$$

This theorem gives an exact analysis of the KL-divergence between tR and n independent Bernoulli variables. Recall that for the purpose of our reduction tR needs to look like a Bernoulli distribution for some parameters. So the main question is: what parameters can we pick such that this divergence $D(tR\|I)$ is negligible?

As a sanity-check, we discuss the application of this theorem to some simple cases. Suppose $\omega \to \infty$, then the distribution R will converge to the uniform distribution. Furthermore, tR will also be uniform, so all the coefficients will be independent. In the theorem we will have $\tilde{h}(\omega) \approx \tilde{h}(\|t\|\omega) \approx 1$. So

$$D(tR\|I) = n(\tilde{h}(\|t\|\omega) - \tilde{h}(\omega)) \approx 0.$$

Another extreme case is $\|t\| = 1$, so $t = X^k$ for some $k \in \{0, \dots, n-1\}$. Then multiplying by t would be equivalent to shifting the coefficients. The shift of independent Bernoulli variables still results in independent Bernoulli variables. As $R \leftarrow \mathrm{Ber}(\omega)$ then $tR \leftarrow (\mathrm{Ber}(\omega))$ with all the $(tR)_i$ coefficients completely independent. In the theorem $D(tR\|I) = n(\tilde{h}(1\omega) - \tilde{h}(\omega)) = 0$.

When $\|t\|$ is small or ω is big, the approximation $Rt \leftarrow \mathrm{Ber}(\|t\|\omega)$ is reasonable. On the other hand when ω is small but $\|t\|$ is quite big then $\tilde{h}(\|t\|\omega) - \tilde{h}(\omega)$ becomes big. This case is exactly when there is not enough entropy in R for all the $(tR)_i$ coefficients to be independent.

In practice, the noise is the sum of s copies of tR, so we need to analyze s copies of this product

$$t_1R_1 + \ldots + t_sR_s,$$

with $R_1, \dots, R_s \leftarrow \mathrm{Ber}(\omega)^n$ independently. Here the total error weight relevant for the decoding problem is

$$\tau = \|t_1\| + \ldots + \|t_s\|.$$

For simplicity, we will first focus on the analysis of one copy, $D(tR\|I)$.

3.1 Approximation for divergence of one product

To give a stricter bound on when the divergence $D(tR\|I)$ is small we need a lemma to approximate $\tilde{h}(\omega)$. This approximation follows from a standard use of Taylor's theorem and may be folklore, but for lack of a suitable citation we provide a proof.

Lemma 3.5 (Approximation of $\tilde{h}(\omega)$). Let $\omega > 0$ we have that

$$\tilde{h}(\omega) = 1 - \frac{2^{-2\omega}}{\ln(2)} + \mathcal{O}(2^{-4\omega}).$$

Proof. The proof follows from expanding the definition of $\tilde{h}$, and performing a Taylor expansion on the logarithm.

$$\begin{aligned}
\tilde{h}(\omega) &= h\left(\frac{1-2^{-\omega}}{2}\right) \\
&= -\frac{1-2^{-\omega}}{2}\log\left(\frac{1-2^{-\omega}}{2}\right) - \frac{1+2^{-\omega}}{2}\log\left(\frac{1+2^{-\omega}}{2}\right) \\
&= -\frac{1-2^{-\omega}}{2}(\log(1-2^{-\omega})-1) - \frac{1+2^{-\omega}}{2}(\log(1+2^{-\omega})-1) \\
&= 1 - \frac{1}{2}\Big((1-2^{-\omega})\log(1-2^{-\omega}) - (1+2^{-\omega})\log(1+2^{-\omega})\Big).
\end{aligned}$$

We compute the Taylor expansion of $(1 \pm x)\log(1 \pm x)$. Because of the convention of using log base 2, we get an additional factor of $1/\ln(2)$ in front of the usual Taylor series of the natural logarithm.

$$\begin{aligned}
(1+x)\log(1+x) &= \frac{1}{\ln(2)}\left(+x + \frac{x^2}{2} - \frac{x^3}{6}\right) + \mathcal{O}(x^4) \\
(1-x)\log(1-x) &= \frac{1}{\ln(2)}\left(-x + \frac{x^2}{2} + \frac{x^3}{6}\right) + \mathcal{O}(x^4)
\end{aligned}$$

Rather than just the asymptotic behavior, we would also like to get an explicit lower bound on $\tilde{h}(\omega)$. Using Taylor's theorem we can compute an explicit formula for approximation error.

Note that the fourth derivatives of $(1+x)\log(1+x)$ and $(1-x)\log(1-x)$ are

$$\begin{aligned}
\frac{d^4}{dx^4}(1+x)\log(1+x) &= \frac{2}{\ln(2)(x+1)^3} \\
\frac{d^4}{dx^4}(1-x)\log(1-x) &= -\frac{2}{\ln(2)(x-1)^3}
\end{aligned}$$

Filling in Taylor's theorem gives that for some $\xi_+, \xi_- \in [0, 2^{-\omega}]$, the errors are of form

$$\begin{aligned}
\varepsilon_+ &:= (1+x)\log(1+x) - \frac{1}{\ln(2)}\left(+x + \frac{x^2}{2} - \frac{x^3}{6}\right) \\
&= \frac{\xi_+^4}{12\ln(2)(1+\xi_+)^3} = \mathcal{O}(2^{-4\omega}) \\
\varepsilon_- &:= (1-x)\log(1-x) - \frac{1}{\ln(2)}\left(-x + \frac{x^2}{2} + \frac{x^3}{6}\right) \\
&= -\frac{\xi_-^4}{12\ln(2)(1-\xi_-)^3} = \mathcal{O}(2^{-4\omega}).
\end{aligned}$$

Filling in the Taylor expansion yields

$$\begin{aligned}
\tilde{h}(\omega) = 1 - \frac{1}{2\ln(2)}\Bigg(&- 2^{-\omega} + \frac{2^{-2\omega}}{2} + \frac{2^{-3\omega}}{6} + \ln(2)\varepsilon_- \\
&+ 2^{-\omega} + \frac{2^{-2\omega}}{2} - \frac{2^{-3\omega}}{6} + \ln(2)\varepsilon_+\Bigg) \\
= 1 - \frac{2^{-2\omega}}{2\ln 2} &+ \varepsilon_+/2 + \varepsilon_-/2.
\end{aligned}$$

Using $\varepsilon_+, \varepsilon_- \in \mathcal{O}(2^{-4\omega})$, we can immediately conclude

$$\tilde{h}(\omega) = 1 - \frac{2^{-2\omega}}{2\ln(2)} + \mathcal{O}(2^{-4\omega})$$

as required. □

Filling in the approximation tells us when $D(tR\|I)$ is negligible.

Corollary 3.1. Let $t, R, I \in \mathbb{F}_2[X]/(X^n - 1)$, with t fixed and invertible and $R \leftarrow \mathrm{Ber}(\omega)^{\otimes n}$, and $I \leftarrow \mathrm{Ber}(\|t\|\omega)^{\otimes n}$. Then

$$D(tR\|I) = n\Theta(2^{-2\omega}).$$

Proof. From Theorem 3.1 we get $D(tR\|I) = n(\tilde{h}(\|t\|\omega) - \tilde{h}(\omega))$. Filling in Lemma 3.5 gives

$$\begin{aligned}
D(tR\|I) &= n\,(\tilde{h}(\|t\|\omega) - \tilde{h}(\omega)) \\
&= n\left(1 - \frac{2^{-2\|t\|\omega}}{2\ln(2)} + \mathcal{O}(2^{-4\omega}) - (1 - \frac{2^{-2\omega}}{2\ln(2)} + \mathcal{O}(2^{-4\omega}))\right) \\
&= n \cdot \Theta(2^{-2\omega})
\end{aligned}$$

as required. □

When ω is small Corollary 3.1 does not tell us much about the quantity. However, even assuming that $\tilde{h}(\|t\|\omega)-\tilde{h}(\omega)$ remains constant, $D(tR\|I) \to \infty$ when $n \to \infty$. Notably, if we want $D(tR\|I) \to 0$, then we need $\omega \to \infty$.

This theorem is also relevant for the post-quantum scheme HQC. [Agu+22b] In Proposition 2.4.2, the noise vector e' has a very similar structure as we analyzed here, of a product of a polynomial with a Bernoulli distribution with some other polynomial. In the paper e' is analyzed as a vector of independent entries. Theorem 3.1 and Corollary 3.1 seem to suggest that making this independence assumption is too optimistic. On the other hand the required properties which should have followed from independence, can still hold. For example, for partial analysis on the Hamming weight of e', refer to [Kaw24].

3.2 Statistical distance

Theorem 3.1 and Corollary 3.1 tell us when the Kullback-Leibler divergence is small. Still, we would also like to give a bound on the statistical distance. To get an upper bound on the statistical distance, we can use the well-known Pinsker's inequality [Kem69, Sec. 6].

Theorem 3.2 (Pinsker's inequality). For P and Q distributions

$$\Delta_{\text{tv}}(P, Q) \leq \sqrt{\frac{1}{2} D(P\|Q)}.$$

Pinsker's inequality immediately tells us that

$$\Delta_{\text{tv}}(tR, I) \leq \sqrt{\frac{1}{2}\, D(tR\|I)} = \sqrt{\frac{n}{2} \cdot \left(\tilde{h}(\|t\|\omega) - \tilde{h}(\omega)\right)},$$

and alternatively

$$\Delta_{\text{tv}}(tR, I) \leq \sqrt{n} \cdot \mathcal{O}(2^{-\omega}).$$

Still we would also like to get a lower bound on the statistical distance. This other direction is harder, because given a certain statistical distance $\delta = \Delta_{\text{tv}}(P, Q)$ the Kullback-Leibler divergence $D(P\|Q)$ can be infinitely large, specifically when there is an x such that $P(x) > 0$, and $Q(x) = 0$. As $P(x)$ can be arbitrarily small, a lower bound on the statistical distance based on the Kullback-Leibler divergence is not possible in general.

In our case, we are working with Bernoulli variables that span the entire outcome space $\mathbb{F}_2[X]/(X^n - 1)$. Therefore, we know that this extreme case cannot occur. Therefore, we can use the results in [Bin19] to get a range on the statistical distance. The paper was about arbitrary random variables. Because we just deal with discrete random variables, we produce their results for the discrete case.

Theorem 3.3 (Reverse Pinsker Inequalities [Bin19]). Let $P, Q : \mathcal{X} \to [0,1]$ be distinct discrete probability distributions over $\mathcal{X}$. Let

$$m = \min_{x \in X} \frac{P(x)}{Q(x)},\ M = \max_{x \in X} \frac{P(x)}{Q(x)}.$$

Then

$$D(P \| Q) \leq \Delta_{\mathrm{tv}}(P, Q)\, \frac{m \log(m)}{1-m} + \frac{M \log(M)}{M-1}.$$

Proof. Filling in $f(x) = x \log(x)$ in [Bin19, Th. 1] gives

$$D(P \| Q) \leq \Delta_{\mathrm{tv}}(P, Q) \left(\frac{m \log(m)}{1-m} + \frac{M \log(M)}{M-1} \right). \qquad \square$$

Now we can apply this bound to $D(tR \| I)$.

Theorem 3.4. For $n \geq 3$ and $\omega \geq \log(n)$, the statistical distance has a lower bound

$$\Delta_{\mathrm{tv}}(tR \| I) \geq 1/3 \cdot D(tR \| I).$$

Theorem 3.4 is a special case of Theorem 3.5 where $s = 1$. We defer the proof to the generic theorem.

The constant 1/3 could have been bigger. First the approximation $\|t\|\omega \geq 2\omega$ is very rough. Furthermore, picking a value $n > 3$ would have yielded a slightly bigger constant than 1/3. However, we mostly care about the linear relation between the statistical distance and the divergence, so the current bound is good enough.

Corollary 3.2. Assuming $\omega \geq \log(n)$ the statistical distance $\Delta_{\mathrm{tv}}(tR, I)$ can be bounded to the range

$$\Delta_{\mathrm{tv}}(tR, I) \in \left[n \cdot \Theta(2^{-2\omega}), \sqrt{n} \cdot \Theta(2^{-\omega}) \right].$$

3.3 Divergence for sum of products

As discussed in practice we would like to analyze the independence of the coefficients of $t_1 R_1 + \ldots + t_s R_s$ for $s \geq 2$. If we want to apply Lemma 3.4, we need to compute the entropy $H(t_1 R_1 + \ldots + t_s R_s)$. Unfortunately, the entropy of this sum is quite hard to compute in general. For some related work on the entropy of sums see [GMT24].

We can easily reason about one special case though. Suppose $t \overset{\mathrm{def}}{=} t_1 = \ldots = t_s$. Then $t_1 R_1 + \ldots + t_s R_s = t(R_1 + \ldots + R_s)$, is invertible. As $R_1 + \ldots + R_s$ $\mathrm{Ber}(s\omega)^{\otimes n}$, we return to our analysis of tR. Now the coefficients of $t(R_1 + \ldots + R_s)$ are marginally distributed according to $\mathrm{Ber}(\|t\| s\omega)$. On the other hand $H(t(R_1+\ldots+R_s)) = n\tilde{h}(s\omega)$. Therefore, for $I_s \leftarrow \mathrm{Ber}(\|t\|s\omega)^{\otimes n}$, we have

$$D(t(R_1+\ldots+R_s)\|I_s) = n(\tilde{h}(\|t\|s\omega) - \tilde{h}(s\omega)).$$

For the general case, it seems much harder to compute this entropy. We can use the bound

$$H(t_1R_1+\ldots+t_sR_s) \leq H(R_1)+\ldots+H(R_s) = sn\tilde{h}(\omega).$$

Unfortunately, this bound is (for parameters of interest) weaker than the trivial upper bound

$$H(t_1R_1+\ldots+t_sR_s) \leq n,$$

as we have a distribution over n bits.

Instead, we can analyze the entropy based on the distribution of the coefficients. If we write

$$C_0 + C_1X + \ldots + C_{n-1}X^{n-1} \overset{\text{def}}{=} t_1R_1+\ldots+t_sR_s,$$

then $H(t_1R_1+\ldots+t_sR_s) = H(C_0, C_1, \ldots, C_{n-1})$. Recall that for $\tau \overset{\text{def}}{=} \|t_1\|+\ldots+\|t_s\|$ the marginal distributions are $C_0, \ldots, C_{n-1} \leftarrow \mathrm{Ber}(\tau\omega)$. Applying again the subadditivity of entropy to these variables only yields the trivial upper bound

$$H(C_0, \ldots C_{n-1}) \leq H(C_0)+\ldots+H(C_{n-1}) = n\tilde{h}(\tau\omega).$$

Recall that we are explicitly trying to compare the entropy of $H(t_1R_1+\cdots+t_sR_s)$ to this value, so this bound yields nothing of interest.

Another possible approach is to compute the upper bound

$$H(C_0, \ldots C_{n-1}) \leq H(C_0, C_1)+H(C_2, C_3)+\cdots+H(C_{n-3}, C_{n-2})+H(C_{n-1}).$$

Recall, that we choose n to be an odd prime, so "splitting up" the entropy this way requires us to deal with $H(C_{n-1})$ separately. Note that for each $1 \leq i < j \leq n-2$, (C_i, C_j) is distributed over 2 bits, and – while this is a bit cumbersome – we can explicitly compute this distribution. We will base our analysis on [PGS16]. This paper is about the weight distribution of the syndrome under Bernoulli noise, i.e. the distribution of RT for $T \in \mathbb{F}_2^{sn\times n}$ a parity-check matrix and $R \leftarrow \mathrm{Ber}(\omega)^{\otimes sn}$. Write $T_i \in \mathbb{F}_2^{n\times n}$ for the cyclic matrix that represents multiplying with $t_i(X)$. Then define

$$T \overset{\text{def}}{=} \left(\begin{array}{c|c|c} \quad T_1 \quad & \quad \ldots \quad & \quad T_s \quad \end{array}\right)$$

which one can view as a parity-check matrix. Furthermore, define R for the concatenation of $R_1, \ldots, R_s$

$$R = (R_1 | \ldots | R_s).$$

Then

$$RT = t_1(X)R_1(X) + \ldots + t_s(X)R_s(X) = (C_0, \ldots, C_{n-1}).$$

It will now be important for us to view elements of $\mathbb{F}_2^n$, or P_n, as elements of $\mathbb{R}^n$, where we naturally map $0 \in \mathbb{F}_2$ to $0 \in \mathbb{R}$ and $1 \in \mathbb{F}_2$ to $1 \in \mathbb{R}$. For a vector $v \in \mathbb{F}_2^n$ (respectively, a polynomial $P \in P_n$), we denote by $\hat{v} \in \mathbb{R}^n$ (respectively, $\hat{P} \in \mathbb{R}^n$) the resulting values in Euclidean space. We also use the same notation for matrices.

The distribution of (C_i, C_j) depends on the symmetric matrix $\Lambda \in \mathbb{R}^{n \times n}$ defined as

$$\Lambda \overset{\text{def}}{=} \hat{T}^\top \hat{T}$$

Above, we emphasize that the multiplication is defined over the **real numbers**. Write λ_{ij} for the coefficients of Λ. Based on the polynomials $t_1, \ldots, t_s$, the value of λ_{ij} is easy to compute. By definition, λ_{ij} is the inner product (over the reals) of the i'th and j'th row of the matrix T. Equivalently

$$\lambda_{ij} = \left\langle \widehat{X^i \cdot t_1}, \widehat{X^j \cdot t_1} \right\rangle_{\mathbb{R}} + \ldots + \left\langle \widehat{X^i \cdot t_s}, \widehat{X^j \cdot t_s} \right\rangle_{\mathbb{R}},$$

where $\langle \cdot, \cdot \rangle_{\mathbb{R}}$ is defined as the usual inner product as vector over the reals, and recall further that the hat notation means that we view the elements as lying in $\{0,1\}^n \subseteq \mathbb{R}^n$. As the multiplication by X^i is just a shift of the vector by i positions, we can replace two shifts by just one shift, so

$$\lambda_{ij} = \left\langle \hat{t}_1, \widehat{X^{i-j} \cdot t_1} \right\rangle_{\mathbb{R}} + \ldots + \left\langle \hat{t}_s, \widehat{X^{i-j} \cdot t_s} \right\rangle_{\mathbb{R}}. \quad (1)$$

Notably, the value of λ_{ij} only depends on the *difference* between the coefficients $i - j$.

Alternatively, one can note that for each $\ell \in [s]$,

$$\left\langle \hat{t}_\ell, \widehat{X^{i-j} \cdot t_\ell} \right\rangle_{\mathbb{R}} = \left| \mathrm{supp}(t_\ell) \cap \mathrm{supp}(X^{i-j} t_\ell) \right| ,$$

where for a polynomial $P(X) = \sum_{i=0}^{n-1} P_i X^i \in \mathbb{F}_2[X]/(X^n - 1)$ we have defined $\mathrm{supp}(P) = \{i \in [n] : P_i = 1\}$.

Recall that $p(\omega) = (1 - 2^{-\omega})/2$. Following [PGS16, Eq. (14)], we can obtain the following joint distribution table for each C_i and C_j:

$C_i \setminus C_j$	0	1
0	$1 - p(\tau\omega) - \frac{1}{2}p(2(\tau - \lambda_{ij})\omega)$	$\frac{1}{2}p(2(\tau - \lambda_{ij})\omega)$
1	$\frac{1}{2}p(2(\tau - \lambda_{ij})\omega)$	$p(\tau\omega) - \frac{1}{2}p(2(\tau - \lambda_{ij})\omega)$

Table 1: Distribution for C_i, C_j

To give a sense of how these quantities can be determined, we provide an example computation. Recall that $(C_1, \ldots, C_n)$ are determined by i.i.d. $\mathrm{Ber}(\omega)$ random variables $R_{u,t}$ for $u \in [n]$ and $t \in [s]$. Fix sets S_i and S_j such that $C_i = \sum_{(u,t)\in S_i} R_{u,t}$ (read modulo 2) and $C_j = \sum_{(u,t)\in S_j} R_{u,t}$. Then $|S_i| = |S_j| = \tau$ and $S_{ij} := S_i \cap S_j$ has $|S_{ij}| = \lambda_{ij}$. Denote by E_{ij} the event that $\sum_{(u,t)\in S_{ij}} R_{u,t} \equiv 0 \pmod 2$, and note that conditioned on E we have that $C_i = 0$ iff $\sum_{(u,t)\in S_i \setminus S_j} R_{u,t} = 0 \pmod 2$, and similarly for the probability $C_j = 1$. Similarly, conditioned on $E_{ij} = 1$ we consider the event that this sum modulo 2 takes on the opposite value. Due to Lemma 2.1, we therefore have

$$\begin{aligned}\Pr[C_i = 0 \wedge C_j = 1] &= \Pr[C_i = 0 \wedge C_j = 1 | E_{ij}] \Pr[E_{ij}] \\ &\quad + \Pr[C_i = 0 \wedge C_j = 1 | \neg E_{ij}] \Pr[\neg E_{ij}] \\ &= (1 - p(\omega(\tau - \lambda_{ij})))p(\omega(t - \lambda_{ij}))(1 - p(\omega\lambda_{ij})) \\ &\quad + p(\omega(\tau - \lambda_{ij}))(1 - p(\omega(\tau - \lambda_{ij})))p(\omega\lambda_{ij}) \\ &= \frac{1}{2}p(2\omega(\tau - \lambda_{ij})),\end{aligned}$$

where in the last line we used the identity $p(x)(1 - p(x)) = \frac{1}{2}p(2x)$.

This distribution (Table 1) yields the entropy

$$\begin{aligned}H(C_i, C_j) &= H(C_i) + H(C_j \mid C_i) \\ &= \tilde{h}(\tau\omega) + p(\tau\omega)H(C_j \mid C_i = 1) + (1 - p(\tau\omega))H(C_j \mid C_i = 0) \\ &= \tilde{h}(\tau\omega) + p(\tau\omega)h\left(\frac{p(2(\tau - \lambda_{ij})\omega)}{2p(\tau\omega)}\right) + (1 - p(\tau\omega))h\left(\frac{p(2(\tau - \lambda_{ij})\omega)}{2(1 - p(\tau\omega))}\right). \\ &\leq \tilde{h}(\tau\omega) + h\left(p(\tau\omega)\frac{p(2(\tau - \lambda_{ij})\omega)}{2p(\tau\omega)} + (1 - p(\tau\omega))\frac{p(2(\tau - \lambda_{ij})\omega)}{2(1 - p(\tau\omega))}\right). \\ &= \tilde{h}(\tau\omega) + \tilde{h}(2(\tau - \lambda_{ij})\omega)\ ,\end{aligned}$$

where the inequality applies the fact that $h(\cdot)$ is concave. Let's compare this to the entropy of $H(C_i) + H(C_j) = 2\tilde{h}(\tau\omega)$ for the idealized case where C_i and C_j are independent. Note that if $\lambda_{ij} > \tau/2$, then the above establishes

$$H(C_i, C_j) \leq \tilde{h}(\tau\omega) + \tilde{h}(2(\tau - \lambda_{ij})\omega) < 2\tilde{h}(\tau\omega) = H(C_i) + H(C_j).$$

Thus, to make thisgap large we would like λ_{ij} to be large. Recalling (1), to have λ_{ij} large we need pairs i, j for which shifting by $j - i$ leads to another coefficient vector with large overlap. We now consider a specific case where we have such large overlap.

Arithmetic progression If the support of the noise vectors is sufficiently structured, then we can find cases such that the λ_{ij}'s are big. By big we mean $\lambda_{ij} \approx \tau$. Suppose that the support of the noise vector is an arithmetic progression: so

$$\mathrm{supp}(t_k) = \{ax + b_k \mid x \in \{0, \ldots, \|t_k\| - 1\}\}$$

for all $k \in [s]$. Then

$$\left\langle \hat{t}_k, \widehat{X^a t_k} \right\rangle_{\mathbb{R}} = |\mathrm{supp}(t_k) \cap \mathrm{supp}(X^a\, t_k)| \geq \|t_k\| - 1.$$

Thus, for each i, we have

$$\lambda_{i,i+a} = \left\langle \hat{t}_1, \widehat{X^a t_1} \right\rangle_{\mathbb{R}} + \ldots + \left\langle \hat{t}_s, \widehat{X^a t_s} \right\rangle_{\mathbb{R}} \geq \tau - s,$$

yielding

$$H(C_i, C_{i+a}) \leq \tilde{h}(\tau\omega) + \tilde{h}(2s\omega).$$

If n is odd, and a and n are coprime, then we can upper bound the total entropy by

$$\begin{aligned} H(C_0, \ldots, C_{n-1}) &\leq H(C_0, C_a) + H(C_{2a}, C_{3a}) + \ldots + H(C_{(n-3)a}, C_{(n-2)a}) + H(C_{(n-1)a}) \\ &\leq \frac{n-1}{2}\Big(\tilde{h}(\tau\omega) + \tilde{h}(2s\omega)\Big) + \tilde{h}(\tau\omega) \end{aligned}$$

where the indices i in C_i are read modulo n, and using that $0, a, \ldots, (n-1)a$ are distinct values mod n when a and n are coprime.

In total using Lemma 3.4 we get

$$\begin{aligned} D(t_1R_1 + \ldots + t_sR_s \| \operatorname{Ber}(\tau\omega)^{\otimes n}) &= H(\operatorname{Ber}(\tau\omega)^{\otimes n}) - H(C_0, \ldots, C_{n-1}) \\ &\geq \frac{n-1}{2}\Big(\tilde{h}(\tau\omega) - \tilde{h}(2s\omega)\Big). \end{aligned}$$

Using the approximation for $\tilde{h}(\omega)$, we can characterize the divergence asymptotically.

Corollary 3.3. Let $\omega \leq \log(n)$, and $\tau \geq 2s$, and $t_1(X), \ldots, t_s(X)$ be vectors that have an arithmetic progression with the same common difference. Then the divergence to $I \stackrel{\text{def}}{=} \operatorname{Ber}(\tau\omega)^{\otimes n}$ can be bounded to the range

$$D(t_1R_1 + \ldots + t_sR_s, I) = \Theta(n2^{-4s\omega}).$$

Notably, this value is negligible when $s\omega \geq \log(n)^2$.

Proof. The proof is a generalization of Corollary 3.1. Using the approximation $\tilde{h} = 1 - 2^{-2\omega} + \mathcal{O}(2^{-4\omega})$ (see Lemma 3.5), we know that

$$\begin{aligned} D(t_1R_1 + \ldots + t_sR_s, I) &= H(I) - H(t_1R_1 + \ldots + t_sR_s) \\ &\geq \frac{n-1}{2}(\tilde{h}(\tau\omega) - \tilde{h}(2s\omega)) \\ &= \frac{n-1}{2}\big(\Theta(2^{-4s\omega}) - \Theta(2^{-\tau\omega})\big) \\ &= \Theta(n2^{-4s\omega}), \end{aligned}$$

as required. □

Statistical distance Similar to the case $s = 1$ (see Section 3.2), we can use the reverse Pinsker inequality to relate the KL-divergence to the statistical distance. To achieve the bound we proof the generalization of Theorem 3.4.

Theorem 3.5. Let $t_1, \ldots, t_s \in \mathbb{F}_2[X]/(X^n - 1)$ polynomials span the entire space, i.e. $\langle t_1, \ldots, t_s \rangle = \mathbb{F}_2[X]/(X^n - 1)$, and $R_1, \ldots, R_s \leftarrow \mathrm{Ber}(\omega)^{\otimes n}$ i.i.d. Furthermore, let $I \leftarrow \mathrm{Ber}(\tau\omega)^{\otimes n}$. For $n \geq 3$ and $\omega \geq \log(n)$, the statistical distance has a lower bound

$$\Delta_{\mathrm{tv}}(t_1R_1 + \ldots + t_sR_s \| I) \geq 1/(3s) \cdot D(t_1R_1 + \cdots + t_sR_s \| I).$$

Proof. In the proof we bound $\Pr[t_1R_1 + \ldots + t_sR_s = x]/\Pr[I = x] \in [1/8^s, 8^s]$. Then, we apply Theorem 3.3. We can achieve these bounds, finding lower/upper bound on the enumerator and the denominator separately. Both probabilities can be bounded using the distributions $\mathrm{Ber}(\omega)$, and $\mathrm{Ber}(\|t\|\omega)$. For any $x \in \mathbb{F}_2[X]/(X^n - 1)$ there are at exactly 2^{sn-n} possible tuples $(r_1, \ldots, r_s)$ such that $t_1R_1 + \ldots + t_sR_s = x$. Then, we can bound the probability of each tuple $\Pr[(R_1, \ldots, R_s) = (r_1, \ldots, r_s)] \in [((1 - 2^{-\omega})/2)^{sn}, ((1 + 2^{-\omega})/2)^{sn}]$.

$$\begin{aligned} \frac{\Pr[t_1R_1 + \ldots + t_sR_s = x]}{\Pr[I = x]} &\leq \frac{\max_{x \in P_n} \Pr[t_1R_1 + \ldots + t_sR_s = x]}{\min_{x \in P_n} \Pr[I = x]} \\ &\leq \frac{2^{sn-n}((1 + 2^{-\omega})/2)^{sn}}{((1 - 2^{-\|t\|\omega})/2)^n} \\ &= \frac{(1 + 2^{-\omega})^{sn}}{(1 - 2^{-\|t\|\omega})^n} \\ &\leq \frac{(1 + 2^{-\omega})^{sn}}{(1 - 2^{-\|t\|\omega})^{sn}} \\ &\leq \frac{(1 + 2^{-\omega})^{sn}}{(1 - 2^{-\omega})^{sn}} && \text{(using } \|t\| \geq 1) \\ &\leq \left(\frac{1 + 1/n}{1 - 1/n}\right)^{ns} && \text{(using } \omega \geq \log(n)), \end{aligned}$$

where we use that $\omega \geq \log(n)$. Define $u_n \overset{\text{def}}{=} \frac{(1+1/n)^n}{(1-1/n)^n}$. The upper bound is to this ratio is equal to $(u_n)^s$. We claim that u_n is decreasing, and $u_3 \leq 8$, which yields an upper bound of 8^s.

To prove this claim we compute the derivative of u_n. As $\ln(\cdot)$ is monotonically increasing, if $f(x) = \ln\left(\left(\frac{1+1/x}{1-1/x}\right)^x\right)$ has a negative derivative, then u_n is indeed decreasing. For $x > 1$ we can give an upper bound for derivative by

$$f'(x) = \ln\left(\frac{x+1}{x-1}\right) - \frac{2x}{x^2-1} < \ln\left(1 + \frac{1}{x}\right) - \frac{2}{x} \leq \frac{1}{x} - \frac{2}{x} = -\frac{1}{x} < 0$$

which is indeed negative.

Using that u_n is decreasing, filling in a value for $n = 3$ gives $u_3 = 8$ so

$$\frac{\Pr[t_1R_1 + \ldots + t_sR_s = x]}{\Pr[I = x]} \leq (u_n)^s \leq (u_3)^s = 8^s.$$

Notably, the upper bound on this ratio independent of n.

For a lower bound on the probability ratio we can apply a similar approach

$$\begin{aligned}
\frac{\Pr[t_1R_1 + \ldots + t_sR_s = x]}{\Pr[I = x]} &\geq \frac{\min_{x\in P_s} \Pr[t_1R_1 + \ldots + t_sR_s = x]}{\max_{x\in P_n} \Pr[I = x]} \\
&\geq \frac{2^{sn-n}((1-2^{-\omega})/2)^{sn}}{((1+2^{-\|t\|\omega})/2)^n} \\
&= \frac{(1-2^{-\omega})^{sn}}{((1+2^{-\|t\|\omega}))^n} \\
&\geq \frac{((1-2^{-\omega}))^{sn}}{((1+2^{-\|t\|\omega}))^{sn}} \\
&\geq \left(\frac{(1-1/n)^n}{(1+1/n)^n}\right)^s && \text{(using } \omega \geq \log(n)\text{)} \\
&= (1/u_n)^s \\
&\geq (1/8)^s
\end{aligned}$$

Where we use that $1/u_n$ is monotonically increasing. In total, we achieve the required bounds

$$(1/8)^s \leq \frac{\Pr[t_1R_1 + \ldots + t_sR_s = x]}{\Pr[I = x]} \leq 8^s.$$

To finish the proof we need to apply Theorem 3.3. Note that for $x > 0$

$$\frac{d}{dx}\left(\frac{x\log(x)}{x-1}\right) = \ln(2) \cdot \frac{x - \log(x) - 1}{(x-1)^2} \geq 0$$

Because the derivative is strictly positive, $x\log(x)/(x-1)$ is monotonically increasing. On the other hand $x\log(x)/(1-x)$ is monotonically decreasing. Therefore, using our bound $M \le 8$, and $m \ge 1/8$, we can give an upper bound on quantity from Theorem 3.3

$$\begin{aligned}\frac{m\log(m)}{1-m} + \frac{M\log(M)}{M-1} &\le \frac{(1/8^s)\cdot\log(1/8^s)}{1-(1/8^s)} + \frac{8^s\log(8^s)}{8^s-1}\\ &= 3s\Big(\frac{8^s}{8^s-1} - \frac{8^{-s}}{1-8^{-s}}\Big)\\ &= 3s.\end{aligned}$$

We can conclude by Theorem 3.3 that

$$D(t_1R_1 + \ldots + t_sR_s \| I) \le 3s \cdot \Delta_{\mathrm{tv}}(t_1R_1 + \ldots + t_sR_s, I),$$

so

$$\Delta_{\mathrm{tv}}(t_1R_1 + \ldots + t_sR_s, I) \ge 1/(3s) \cdot D(t_1R_1 + \ldots + t_sR_s \| I).$$

The theorem follows. □

Acknowledgements. The authors would like to thank Jeroen Zuiddam and Christian Schaffner for helpful feedback. MB's work was supported by NWO Gravitation Project QSC. NR is supported by an NWO (Dutch Research Council) grant with number C.2324.0590. This work was done in part while NR was visiting the Simons Institute for the Theory of Computing, supported by DOE grant # DE-SC0024124.

References

[Agu+22a] C. Aguilar Melchor, N. Aragon, P. Barreto, et al. *BIKE*. Round 4 Submission to the NIST Post-Quantum Cryptography Call, v. 5.1. Version 5.1. Oct. 2022. URL: `https://bikesuite.org`.

[Agu+22b] C. Aguilar Melchor, N. Aragon, S. Bettaieb, et al. *HQC*. Round 4 Submission to the NIST Post-Quantum Cryptography Call. `https://pqc-hqc.org/`. Oct. 2022.

[Alb+22] M. Albrecht, D. J. Bernstein, T. Chou, et al. *Classic McEliece (merger of Classic McEliece and NTS-KEM)*. `https://classic.mceliece.org`. Fourth round finalist of the NIST post-quantum cryptography call. Nov. 2022.

[Bia+23] B. Biasioli, C. Marcolla, M. Calderini, and J. Mono. *Improving and Automating BFV Parameters Selection: An Average-Case Approach*. Cryptology ePrint Archive, Report 2023/600. 2023. URL: `https://eprint.iacr.org/2023/600`.

[Bin19] O. Binette. "A Note on Reverse Pinsker Inequalities". In: *IEEE Trans. Inf. Theory* 65.7 (July 2019), pp. 4094–4096. DOI: `10.1109/TIT.2019.2896192`.

[Bra+13] Z. Brakerski, A. Langlois, C. Peikert, et al. "Classical Hardness of Learning with Errors". In: *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing.* STOC'13: Symposium on Theory of Computing. ACM, June 2013, pp. 575–584. DOI: 10.1145/2488608.2488680.

[Bra+19] Z. Brakerski, V. Lyubashevsky, V. Vaikuntanathan, and D. Wichs. "Worst-Case Hardness for LPN and Cryptographic Hashing via Code Smoothing". In: *Advances in Cryptology – EUROCRYPT 2019.* Springer International Publishing, 2019, pp. 619–635. DOI: 10.1007/978-3-030-17659-4_21.

[Chi+20] I. Chillotti, N. Gama, M. Georgieva, and M. Izabachène. "TFHE: Fast Fully Homomorphic Encryption Over the Torus". In: 33.1 (Jan. 2020), pp. 34–91. DOI: 10.1007/s00145-019-09319-x.

[FS96] J.-B. Fischer and J. Stern. "An efficient pseudo-random generator provably as secure as syndrome decoding". In: *Advances in Cryptology - EUROCRYPT'96.* Vol. 1070. Springer, 1996, pp. 245–255. ISBN: ISBN 978-3-540-61186-8.

[GMT24] B. Green, F. Manners, and T. Tao. "Sumsets and Entropy Revisited". In: *Random Struct. Algorithms* n/a.n/a (July 31, 2024). DOI: 10.1002/rsa.21252.

[Kaw24] A. Kawachi. "Hamming Weight of Product of Random Sparse Polynomials". In: *Int. Symp. Inf. Theory Its Appl.* (24–Oct. 27, 2020).

[Kem69] J. H. B. Kemperman. "On the Optimum Rate of Transmitting Information". In: *Probability and Information Theory.* Springer, 1969, pp. 126–169. DOI: 10.1007/BFb0079123.

[MP24] S. Murphy and R. Player. "A Central Limit Approach for Ring-LWE Noise Analysis". In: *IACR Commun. Cryptol.* 1.2 (2024), p. 7. DOI: 10.62056/AY76C0KR. URL: https://eprint.iacr.org/2019/452.pdf.

[PGS16] C. Pacher, P. Grabenweger, and D. E. Simos. "Weight Distribution of the Syndrome of Linear Codes and Connections to Combinatorial Designs". In: *2016 IEEE International Symposium on Information Theory (ISIT).* 2016 IEEE International Symposium on Information Theory (ISIT). July 2016, pp. 3038–3042. DOI: 10.1109/ISIT.2016.7541857.

[Reg05] O. Regev. "On Lattices, Learning with Errors, Random Linear Codes, and Cryptography". In: *Proceedings of the Thirty-Seventh Annual ACM Symposium on Theory of Computing.* Association for Computing Machinery, May 22, 2005, pp. 84–93. DOI: 10.1145/1060590.1060603.

[Tao10] T. Tao. "Sumset and Inverse Sumset Theory for Shannon Entropy". In: *Comb. Probab. Comput.* 19.4 (July 2010), pp. 603–639. DOI: 10.1017/S0963548309990642.