

Statistical Decoding 2.0: Reducing Decoding to LPN

Kévin Carrier¹, Thomas Debris-Alazard², Charles Meyer-Hilfiger³, and Jean-Pierre Tillich³

¹ ETIS Laboratory, CY Cergy-Paris University, kevin.carrier@ensea.fr

² Project GRACE, Inria Saclay-Ile de France, thomas.debris@inria.fr

³ Project COSMIQ, Inria de Paris,
charles.meyer-hilfiger@inria.fr, jean-pierre.tillich@inria.fr

Abstract. The security of code-based cryptography relies primarily on the hardness of generic decoding with linear codes. The best generic decoding algorithms are all improvements of an old algorithm due to Prange: they are known under the name of information set decoders (ISD). A while ago, a generic decoding algorithm which does not belong to this family was proposed: statistical decoding. It is a randomized algorithm that requires the computation of a large set of parity-checks of moderate weight, and uses some kind of majority voting on these equations to recover the error. This algorithm was long forgotten because even the best variants of it performed poorly when compared to the simplest ISD algorithm. We revisit this old algorithm by using parity-check equations in a more general way. Here the parity-checks are used to get LPN samples with a secret which is part of the error and the LPN noise is related to the weight of the parity-checks we produce. The corresponding LPN problem is then solved by standard Fourier techniques. By properly choosing the method of producing these low weight equations and the size of the LPN problem, we are able to outperform in this way significantly information set decodings at code rates smaller than 0.3. It gives for the first time after 60 years, a better decoding algorithm for a significant range which does not belong to the ISD family.

1 Introduction

1.1 The Decoding Problem and Code-based Cryptography

Code-based cryptography relies crucially on the hardness of decoding generic linear codes which can be expressed as follows in the binary case

Problem 1.1 (decoding a linear code). *Let \mathcal{C} be a binary linear code over \mathbb{F}_2 of dimension k and length n , i.e. a subspace of \mathbb{F}_2^n of dimension k . We are given $\mathbf{y} \in \mathbb{F}_2^n$, an integer t and want to find a codeword $\mathbf{c} \in \mathcal{C}$ and an error vector $\mathbf{e} \in \mathbb{F}_2^n$ of Hamming weight $|\mathbf{e}| = t$ for which $\mathbf{y} = \mathbf{c} + \mathbf{e}$.*

This terminology stems from information theory, \mathbf{y} is a noisy version of a codeword \mathbf{c} : $\mathbf{y} = \mathbf{c} + \mathbf{e}$ where \mathbf{e} is a vector of weight t and we want to recover the original codeword \mathbf{c} . It can also be viewed as solving an underdetermined linear system with a weight constraint. Indeed, we can associate to a subspace \mathcal{C} of dimension k of \mathbb{F}_2^n a binary $(n-k) \times n$ matrix \mathbf{H} (also called a *parity-check* matrix of the code) whose kernel defines \mathcal{C} , namely $\mathcal{C} = \{\mathbf{x} \in \mathbb{F}_2^n : \mathbf{H}\mathbf{x}^\top = \mathbf{0}\}$. The decoding problem is equivalent to find an \mathbf{e} of Hamming weight t such that $\mathbf{H}\mathbf{e}^\top = \mathbf{s}^\top$ where \mathbf{s} is the *syndrome* of \mathbf{y} with respect to \mathbf{H} , i.e. $\mathbf{s}^\top = \mathbf{H}\mathbf{y}^\top$. This can be verified by observing that if there exists $\mathbf{c} \in \mathcal{C}$ and \mathbf{e} such that $\mathbf{y} = \mathbf{c} + \mathbf{e}$ then $\mathbf{H}\mathbf{y}^\top = \mathbf{H}(\mathbf{c} + \mathbf{e})^\top = \mathbf{H}\mathbf{c}^\top + \mathbf{H}\mathbf{e}^\top = \mathbf{H}\mathbf{e}^\top$.

The decoding problem has been studied for a long time and despite many efforts on this issue [Pra62, Ste88, Dum91, Bar97, FS09, BLP11, MMT11, BJMM12, MO15] the best algorithms [BJMM12, MO15, BM17] are exponential in the number of errors that have to be corrected: correcting t errors in a binary linear code of length n with the aforementioned algorithms has a cost of $2^{\alpha n(1+o(1))}$ where $\alpha = \alpha(R, \tau)$ is a constant depending of the code rate $R \triangleq \frac{k}{n}$, the error rate $\tau \triangleq \frac{t}{n}$ and the algorithm which is used. All the efforts that have been spent on this problem have only managed to decrease slightly this exponent α . Let us emphasize that this exponent is the key for estimating the security level of any code-based cryptosystem. We expect that this problem

is the hardest at the Gilbert-Varshamov relative distance $\tau = \delta_{\text{GV}}$ where $\delta_{\text{GV}} \triangleq h^{-1}(1 - R)$, with h being the binary entropy function $h(x) \triangleq -x \log_2 x - (1 - x) \log_2(1 - x)$ and $h^{-1}(x)$ its inverse ranging over $[0, \frac{1}{2}]$ (see an illustration of this behavior in Figure 1.1). This corresponds in the case of random linear codes to the largest relative weight below which there is typically just one solution of the decoding problem assuming that there is one. Above this bound, the number of solutions becomes exponential (at least as long as $\tau < 1 - \delta_{\text{GV}}$) and this helps to devise more efficient decoders. Furthermore, all the aforementioned algorithms become polynomial in the regime $\frac{1-R}{2} \leq \tau \leq \frac{1+R}{2}$.

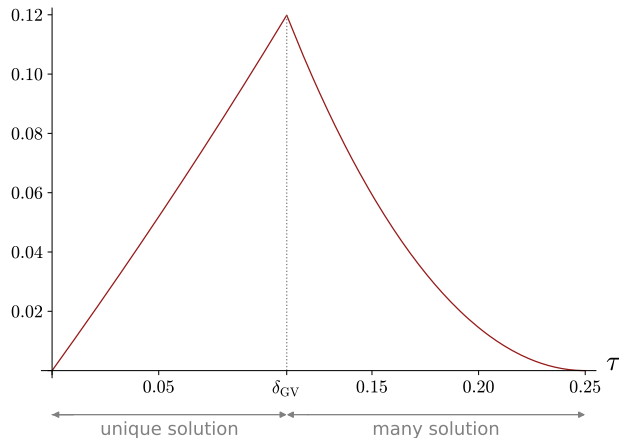


Figure 1.1. Complexity exponent α of the Prange ISD algorithm [Pra62] as a function of the error ratio $\tau \triangleq \frac{t}{n}$ at rate $R = \frac{1}{2}$. The peak corresponds to the normalized Gilbert-Varshamov distance $\delta_{\text{GV}} = h^{-1}(1 - R)$.

There are code-based cryptographic primitives whose security relies precisely on the difficulty of decoding at the Gilbert-Varshamov relative distance (something which is also called *full distance decoding* [MO15, BM17]), for instance the Stern code-based identification schemes or associated signatures schemes [Ste93, GG07, AGS11, FJR21]. In the light of the upcoming NIST second call for new quantum resistant signature algorithms, it is even more important to have a stable and precise assessment of what we may expect about the complexity of solving this problem. For much smaller distances, say sub-linear, which is relevant for cryptosystems like [MTSB13, McE78], the situation seems much more stable/well understood, since the complexity exponent of all the above-mentioned algorithms is the same in this regime [CS16].

1.2 ISD Algorithms and Beyond: Statistical Decoding

All the aforementioned algorithms can be viewed as a refinement of the original Prange algorithm [Pra62] and are actually all referred to as Information Set Decoding (ISD) algorithms. Basically, they all use a common principle, namely making the bet that in a certain set of about k positions (the “information set”) there are only very few errors and using this bet to speed-up decoding. The parameters of virtually all code-based cryptographic algorithms (for the Hamming metric) have been chosen according to the running time of this family of algorithms. Apart from these algorithms, there is one algorithm which is worth mentioning, namely statistical decoding. It was first proposed by Al Jabri in [Jab01] and improved a little bit by Overbeck in [Ove06]. Later on, [FKI07] proposed an iterative version of this algorithm.

It is essentially a two-stage algorithm, the first step consisting in computing an exponentially large number of parity-check equations of the smallest possible weight w , and then from these

parity-check equations the error is recovered by some kind of majority voting based on these parity-check equations. This majority voting is based on the following principle, take a parity-check equation \mathbf{h} for the code \mathcal{C} we want to decode, *i.e.* a binary vector $\mathbf{h} = (h_i)_{1 \leq i \leq n}$ such that $\langle \mathbf{h}, \mathbf{c} \rangle = 0$ for every \mathbf{c} in \mathcal{C} . Assume that the i -bit of the parity-check is 1, then since $\langle \mathbf{h}, \mathbf{y} \rangle = \langle \mathbf{h}, \mathbf{e} \rangle = e_i + \sum_{j \neq i} h_j e_j$, the i -th bit e_i of the error \mathbf{e} we want to recover satisfies

$$e_i + \sum_{j \neq i} h_j e_j = \langle \mathbf{h}, \mathbf{y} \rangle. \quad (1.1)$$

The sum $\sum_{j \neq i} h_j e_j$ is biased, say it is equal to 1 with probability $\frac{1-\varepsilon}{2}$ with a bias ε which is (essentially) a decreasing function of the weight w of the parity-check \mathbf{h} . This allows to recover e_i with about $\Theta(1/\varepsilon^2)$ parity-checks. However the bias is exponentially small in the codelength. An asymptotic analysis of this algorithm was performed in [DT17a] and it turns out that even if we had a way to obtain freely the parity-check equations we need, this kind of algorithm could not even outperform the simplest ISD algorithm: the Prange algorithm. This is done in [DT17a] by showing that there is no loss in generality if we just care about getting the best exponent to restrict ourselves to a single parity-check weight w (see Section 5 in [DT17a]) and then analyze the complexity of such a putative algorithm for a single weight by using the knowledge of the typical number of parity-check equations of a given weight in a random linear code. The complexity exponent we get is a lower bound on the complexity of statistical decoding. We call such a putative statistical decoding algorithm, *genie-aided statistical decoding*: we are assisted by a genie which gives for free all the parity-check equations we require (but of course we can only get as much parity-check equations of some weight w as there exists in the code we want to decode). The analysis of the exponent we obtain with such genie-aided statistical decoding is given in Section 7 of [DT17a] and shows that it is outperformed very significantly by the Prange algorithm (see [DT17a, §7.2, Fig. 6]).

1.3 Contributions

In this work, we modify statistical decoding so that each parity-check yields now an LPN sample which is a noisy linear combination involving part of the error vector. This improves significantly statistical decoding, since the new decoding algorithm outperforms significantly all ISD's for code rates smaller than 0.3. It gives for the first time after 60 years, a better decoding algorithm that does not belong to the ISD family, and this for a very significant range of rates. The only other example where ISD algorithms have been beaten was in 1986, when Dumer introduced his collision technique. This improved the Prange decoder only for rates in the interval $[0.98, 1]$ and interestingly enough it gave birth to all the modern improvements of ISD algorithms starting from Stern's algorithm [Ste88].

A New Approach : Using Parity-Checks to Reduce Decoding to LPN. Our approach for solving the decoding problem reduces it to the so-called Learning Parity with Noise Problem (LPN).

Problem 1.2 (LPN). Let $\mathcal{O}_{\mathbf{s}, \tau}(\cdot)$ be an oracle parametrized by $\mathbf{s} \in \mathbb{F}_2^s$ and $\tau \in [0, 1]$ such that on a call it outputs $(\mathbf{a}, \mathbf{s} \cdot \mathbf{a} + e)$ where $\mathbf{a} \in \mathbb{F}_2^n$ is uniformly distributed and e is distributed according to a Bernoulli of parameter τ . We have access to $\mathcal{O}_{\mathbf{s}, \tau}(\cdot)$ and want to find \mathbf{s} .

(1.1) can be interpreted as an LPN sample with an \mathbf{s} of size 1, namely e_i . However, if instead of splitting the support of the parity-check with one bit on one side and the other ones on the other side, but choose say s positions on the first part (say the s first ones) and $n - s$ on the other, we can write

$$\langle \mathbf{h}, \mathbf{y} \rangle = \underbrace{\sum_{i=1}^s h_i e_i}_{\text{linear comb.}} + \underbrace{\sum_{j>s} h_j e_j}_{\text{LPN noise}}.$$

We may interpret such a scalar product as an LPN sample where the secret is (e_1, \dots, e_s) ; *i.e.* we have a noisy information on a linear combination $\sum_{i=1}^s h_i e_i$ on the s first bits of the error where the noise is given by the term $\sum_{j>s} h_j e_j$ and the information is of the form $\sum_{i=1}^s h_i e_i + \text{noise} = \langle \mathbf{h}, \mathbf{y} \rangle$. Again the second linear combination is biased, say $\mathbb{P}\left(\sum_{j>s} h_j e_j = 1\right) = \frac{1-\varepsilon}{2}$ and information theoretic arguments show that again $\Theta(1/\varepsilon^2)$ samples are enough to determine (e_1, \dots, e_s) . It seemed that we gained nothing here since we still need as many samples as before and it seems that now recovering (e_1, \dots, e_s) is much more complicated than performing majority voting.

However with this new approach, we just need parity-check equations of low weight on $n - s$ positions (those that determine the LPN noise) whereas in statistical decoding algorithm we have to compute parity-check equations of low weight on $n - 1$ positions. This brings us to the main advantage of our new method: the parity-checks we produce have much lower weight on those $n - s$ positions than those we produce for statistical decoding. This implies that the bias ε in the LPN noise is much bigger with the new method and the number $N = \Theta(1/\varepsilon^2)$ of parity-check equations much lower. Second by using the fast Fourier transform, we can recover (e_1, \dots, e_s) in time $O(s2^s)$. Therefore, as long as the number of parity-checks we need is of order $\Omega(2^s)$, there is no exponential extra cost of having to recover (e_1, \dots, e_s) . This new approach will be called from now on *Reduction to LPN decoding* (RLPN).

Subset Sum Techniques and Bet on the Error Distribution. As just outlined, our RLPN decoder needs an exponential number $N = \Theta(1/\varepsilon^2)$ of parity-checks of small weight on $n - s$ positions. This can be achieved efficiently by using collision/subset techniques used in the inner loop of ISD's. Recall that all ISD's proceed in two steps, *(i)* first they pick an augmented information set and *(ii)* then have an inner loop computing low weight codewords of some sort. *(ii)* uses advanced techniques to solve subset-sum problems like birthday paradox [Dum86, Dum91], Wagner algorithm [Wag02], representations techniques [MMT11, BJMM12]. All these techniques can also be used in a natural way in our RLPN decoder to compute the low weight parity-checks we need.

Furthermore, another idea of ISD's can be used in our RLPN decoder. All ISD's are making, in a fundamental way, a bet on the error weight distribution in several zones related to the information set picked up in *(i)*. There are two zones: the potentially augmented information set and the rest of the positions. ISD algorithms assume that the (augmented) information set contains only very few errors. A similar bet can be made in our case. We have two different zones: on one hand the s positions determining s error bits and on the other $n - s$ bits which determine the LPN noise. It is clearly favorable to have an error ratio which is smaller on the second part. The probability that this unlikely event happens is largely outweighed by the gain in the bias of the LPN noise.

Our Results. Using all the aforementioned ingredients results in dramatically improving statistical decoding (see Figure 1.2), especially in the low rate regime ($R \leq \frac{1}{2}$) where ISD algorithms are known to perform slightly worse than in the high rate regime ($R > \frac{1}{2}$). Indeed, the complexity exponent $\alpha(R) \triangleq \alpha(R, \delta_{GV}(R))$ of ISD's for full decoding (*a.k.a* the GV bound decoding) which could be expected to be symmetric in R is actually bigger in the low rate regime than in the high rate regime: $\alpha(R) > \alpha(1 - R)$ for $0 < R < \frac{1}{2}$. This results in an exponent curve which is slightly tilted towards the left, the maximum exponent being always obtained for $R < \frac{1}{2}$. Even worse, the behavior for very small rates (*i.e.* $R \rightarrow 0^+$) is fundamentally different in the very high rate regime ($R \rightarrow 1^-$). The complexity curve behaves like $\alpha(R) \approx R$ in the first case and like $\alpha(R) \approx \frac{1-R}{2}$ in the second (at least for all later improvements of the Prange decoder incorporating collision techniques). This behavior at 0 for full distance decoding has never been changed by *any* decoder. It should be noted that $\alpha(R) = R(1 + o(1))$ around 0 means that the complexity behaves like $2^{\alpha(R)n} = 2^{R(1+o(1))n} = 2^{k(1+o(1))}$, so in essence ISD's are not performing really better than trivial enumeration on all codewords. This fundamental barrier is still unbroken by RLPN decoding, but it turns out that $\alpha(R)$ approaches R much more slowly with RLPN. For instance, for $R = 0.02$

we have $\alpha(R) \approx \frac{R}{2}$. This behavior in the very low regime is instrumental for the improvement we obtain on ISD's. In essence, this improvement is due in this regime to the conjunction of RLPN decoding with a collision search of low weight parity-checks. This method can be viewed as the dual (*i.e.* operating on the dual code) of the collision search performed in advanced ISD's which are successful for lowering the complexity exponent down to $\alpha(R) \approx \frac{1-R}{2}$ in the high rate regime. In some sense, the RLPN strategy allows us to *dualize* advanced ISD techniques for working in the low rate regime.

All in all, using [BJMM12] (one of the most advanced ISD techniques) to compute low weight codewords of some shape we are able to outperform significantly even the latest improvements of ISD algorithms for code rates R smaller than 0.3 as shown in Figure 1.2. This is a breakthrough in this area, given the dominant role that ISD algorithms have played during all those years for assessing the complexity of decoding a linear code.

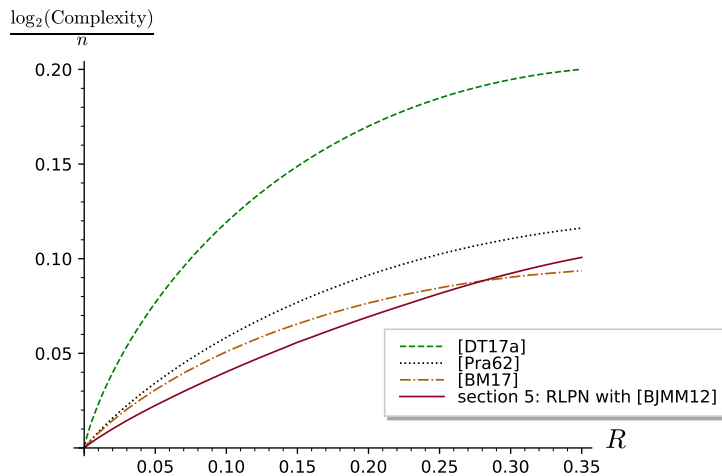


Figure 1.2. Complexity exponent for full distance decoding of genie-aided statistical decoding (recall that this is a *lower bound* on the complexity exponent of statistical decoding), the basic Prange ISD algorithm [Pra62], the best state-of-the-art algorithm [BM17] and our RLPN decoder as a function of R . Note that [BM18] claims a slightly better exponent with an improvement of $\approx 7\%$ for full decoding over [BM17] at the maximum exponent. These results are however incorrect, see the Appendix B for this point.

Proving the Standard assumption of Statistical Decoding. In analyzing the new decoding algorithm, we also put statistical decoding on a much more rigorous foundation by showing that the basic condition that has to be met for both statistical decoding and RLPN decoding, namely that the number N of parity-check equations that are available is at least of order $\Omega(1/\varepsilon^2)$ in the case of statistical decoding and $\Omega(s/\varepsilon^2)$ in the case of RLPN decoding, where ε is the bias of the LPN noise is also essentially the condition which ensures that the bias is well approximated by the standard assumption made for statistical decoding which assumes that

$$\text{bias}(\langle \mathbf{e}, \mathbf{h} \rangle) \approx \text{bias}(\langle \mathbf{e}, \mathbf{h}' \rangle), \quad (1.2)$$

where $\text{bias}(X)$ is defined for a binary random variable as $\text{bias}(X) = \mathbb{P}(X = 0) - \mathbb{P}(X = 1)$, where \mathbf{h} is chosen uniformly at random among the parity-checks of weight w on \mathcal{N} of the code \mathcal{C} we decode whereas \mathbf{h}' is chosen uniformly at random among the words of weight w on \mathcal{N} . We will namely prove that as soon as the parameters are chosen such that $N = \omega(1/\text{bias}(\langle \mathbf{e}, \mathbf{h}' \rangle)^2)$, we have that for all but a proportion $o(1)$ of codes \mathcal{C}

$$\text{bias}(\langle \mathbf{e}, \mathbf{h} \rangle) = (1 + o(1)) \text{bias}(\langle \mathbf{e}, \mathbf{h}' \rangle).$$

This is Proposition 3.8 in Subsection 3.1.

2 Notation and Coding Theory Background

In this section, we introduce notation and coding theoretic background which are used throughout the paper.

Vectors and matrices. Vectors and matrices are respectively denoted in bold letters and bold capital letters such as \mathbf{x} and \mathbf{A} . The entry at index i of the vector \mathbf{x} is denoted by x_i . The canonical scalar product $\sum_{i=1}^n x_i y_i$ between two vectors \mathbf{x} and \mathbf{y} of \mathbb{F}_2^n is denoted by $\langle \mathbf{x}, \mathbf{y} \rangle$. Let \mathcal{J} be a list of indexes. We denote by $\mathbf{x}_{\mathcal{J}}$ the vector $(x_i)_{i \in \mathcal{J}}$. In the same way, we denote by $\mathbf{A}_{\mathcal{J}}$ the sub-matrix made of the columns of \mathbf{A} which are indexed by \mathcal{J} . The concatenation of two vectors \mathbf{x} and \mathbf{y} is denoted by $\mathbf{x} \parallel \mathbf{y}$. The Hamming weight of a vector $\mathbf{x} \in \mathbb{F}_2^n$ is defined as the number of its non-zero coordinates, namely $|\mathbf{x}| \triangleq \#\{i \in \llbracket 1, n \rrbracket : x_i \neq 0\}$ where $\#$ stands for the cardinal and $\llbracket a, b \rrbracket$ stands for the set of the integers between a and b .

Probabilistic notation. For a finite set \mathcal{S} , we write $X \stackrel{\$}{\leftarrow} \mathcal{S}$ when X is an element of \mathcal{S} drawn uniformly at random in it. For a Bernoulli random variable X , denote by $\text{bias}(X)$ the quantity $\text{bias}(X) \triangleq \mathbb{P}(X = 0) - \mathbb{P}(X = 1)$. For a Bernoulli random variable X of parameter $p = \frac{1-\varepsilon}{2}$, i.e. $\mathbb{P}(X = 1) = \frac{1-\varepsilon}{2}$ we have $\text{bias}(X) = \varepsilon$.

Soft-O notation. For real valued functions defined over \mathbb{R} or \mathbb{N} we define $o()$, $O()$, $\Omega()$, $\Theta()$, in the usual way and also use the less common notation $\tilde{O}()$ and $\tilde{\Omega}()$, where $f = \tilde{O}(g)$ means that $f(x) = O(g(x) \log^k g(x))$ and $f = \tilde{\Omega}(g)$ means that $f(x) = \Omega(g(x) \log^k g(x))$ for some k . We will use this for functions which have an exponential behaviour, say $g(x) = e^{\alpha x}$, in which case $f(x) = \tilde{O}(g(x))$ means that $f(x) = O(P(x)g(x))$ where P is a polynomial in x . We also use $f = \omega(g)$ when f dominates g asymptotically; that is when $\lim_{x \rightarrow \infty} \frac{|f(x)|}{g(x)} = \infty$.

Coding theory. A binary linear code \mathcal{C} of length n and dimension k is a subspace of the vector space \mathbb{F}_2^n of dimension k . We say that it has parameters $[n, k]$ or that it is an $[n, k]$ -code. Its *rate* R is defined as $R \triangleq \frac{k}{n}$. A generator matrix \mathbf{G} for \mathcal{C} is a full rank $k \times n$ matrix over \mathbb{F}_2 that is such that

$$\mathcal{C} = \{\mathbf{u}\mathbf{G} : \mathbf{u} \in \mathbb{F}_2^k\}.$$

In other words, the rows of \mathbf{G} form a basis of \mathcal{C} . A parity-check matrix \mathbf{H} for \mathcal{C} is a full-rank $(n - k) \times n$ matrix over \mathbb{F}_2 such that

$$\mathcal{C} = \{\mathbf{c} \in \mathbb{F}_2^n : \mathbf{H}\mathbf{c}^\top = \mathbf{0}\}.$$

In other words, \mathcal{C} is the null space of \mathbf{H} . The code whose generator matrix is the parity-check matrix of \mathcal{C} is called the dual code of \mathcal{C} . It might be seen as the subspace of parity-checks of \mathcal{C} and is defined equivalently as

Definition 2.1 (dual code). *The dual code \mathcal{C}^\perp of an $[n, k]$ code \mathcal{C} is an $[n, n - k]$ code which is defined by*

$$\mathcal{C}^\perp \triangleq \{\mathbf{d} \in \mathbb{F}_2^n : \forall \mathbf{c} \in \mathcal{C}, \langle \mathbf{c}, \mathbf{d} \rangle = 0\}.$$

It will also be very convenient to consider the operation of puncturing a code, i.e. keeping only a subset of entries in a codeword.

Definition 2.2 (punctured code). *For a code \mathcal{C} and a subset \mathcal{J} of code positions, we denote by $\mathcal{C}_{\mathcal{J}}$ the punctured code obtained from \mathcal{C} by keeping only the positions in \mathcal{J} , i.e.*

$$\mathcal{C}_{\mathcal{J}} = \{\mathbf{c}_{\mathcal{J}} : \mathbf{c} \in \mathcal{C}\}.$$

We will also use several times that random binary linear codes achieve the capacity of any binary input symmetric channel whose definition is

Definition 2.3 (binary input memoryless symmetric channel). A binary input memoryless symmetric channel (BIMS) with output a finite alphabet \mathcal{Y} , is an error model on $\{0, 1\}^*$ assuming that when a bit $b \in \{0, 1\}$ is sent, it gets mapped to $y \in \mathcal{Y}$ with probability denoted by $p(y|b)$ (these are the transition probabilities of the channel). Being symmetric means that there is an involution f such that $p(y|0) = p(f(y)|1)$. Being memoryless means that the outputs of the channel are independent conditioned on the inputs: when $b_1 \cdots b_n \in \{0, 1\}^n$ is sent, the probability that the output is $y_1 \cdots y_n$ is given by $p(y_1|b_1) \cdots p(y_n|b_n)$.

The capacity of such a channel is given by

Definition 2.4 (capacity of a BIMS channel). The capacity⁴ C of a BIMS channel with transition probabilities $(p(y|b))_{\substack{y \in \mathcal{Y} \\ b \in \{0,1\}}}$ is given by

$$C \triangleq \sum_{y \in \mathcal{Y}} \sum_{b \in \{0,1\}} \frac{p(y|b)}{2} \log_2 \frac{p(y|b)}{\frac{1}{2}p(y|0) + \frac{1}{2}p(y|1)}.$$

We will consider here the binary symmetric channel (BSC) given by

Definition 2.5 (binary symmetric channel). $BSC(p)$ is a BIMS channel with output alphabet $\mathcal{Y} = \{0, 1\}$ and transition probabilities $p(0|0) = p(1|1) = 1 - p$, $p(1|0) = p(0|1) = p$, where p is the crossover probability of the channel.

In other words, this means that a bit b is transformed into its opposite $1 - b$ with probability p when sent through the channel. It is readily verified that

Definition 2.6. The capacity C of $BSC(p)$ is given by $C = 1 - h(p)$.

We will also talk about maximum likelihood decoding a code (under the assumption that the input codeword is chosen uniformly at random) for a given channel, meaning the following

Definition 2.7 (maximum likelihood decoding). Maximum likelihood decoding of a binary code $\mathcal{C} \subset \{0, 1\}^n$ over a BIMS channel with transitions probabilities $(p(y|b))_{\substack{y \in \mathcal{Y} \\ b \in \{0,1\}}}$ corresponds, given a received word $\mathbf{y} \in \mathcal{Y}^n$, to output the (or one of them if there are several equally likely candidates) codeword \mathbf{x} which maximizes $p(\mathbf{y}|\mathbf{x})$. Here $p(\mathbf{y}|\mathbf{x}) \triangleq p(y_1|x_1) \cdots p(y_n|x_n)$ denotes the probability of receiving \mathbf{y} given that \mathbf{x} was sent.

In a sense, this is the best possible decoding algorithm for a given channel model. There is a variation of Shannon's theorem (see for instance [RU08, Th. 4.68 p. 203]) which says that a family of random binary linear codes $(\mathcal{C}_n)_n$ attain the capacity of a BIMS channel

Theorem 2.8. Consider a BIMS channel of capacity C . Let $\delta > 0$ and consider a family of random binary linear codes \mathcal{C}_n of length n and rate smaller than $(1 - \delta)C$ obtained by choosing their generator matrix uniformly at random. Then under maximum likelihood decoding, the probability of error after decoding goes to 0 as n tends to infinity.

3 Reduction to LPN and the Associated Algorithm

The purpose of this section is (i) to explain in detail the reduction to LPN, (ii) to give a high level description of the algorithm which does not specify the method for finding the dual codewords we need, and then (iii) to give its complexity. We assume from now on that we are given \mathbf{y} which is equal to a sum of a codeword \mathbf{c} of the code \mathcal{C} we want to decode plus an error vector \mathbf{e} of Hamming weight t :

$$\mathbf{y} = \mathbf{c} + \mathbf{e}, \quad \mathbf{c} \in \mathcal{C}, \quad |\mathbf{e}| = t.$$

We will start this section by explaining how we reduce decoding to an LPN problem and also show how the LPN noise can be estimated accurately.

⁴ The formula given here is strictly speaking the symmetric capacity of a channel, but these two notions coincide in the case of a BIMS channel.

3.1 Reduction to LPN

Recall that in RLPN decoding we first randomly select a subset \mathcal{P} of s positions

$$\mathcal{P} \subseteq \llbracket 1, n \rrbracket \quad \text{such that} \quad \#\mathcal{P} = s$$

where s is a parameter that will be chosen later. \mathcal{P} corresponds to the entries of \mathbf{e} we aim to recover and is the secret in the LPN problem. We denote by $\mathcal{N} \triangleq \llbracket 1, n \rrbracket \setminus \mathcal{P}$ the complementary set, with a choice of the letter \mathcal{N} standing for “noise” for reasons that will be clear soon. Given $\mathbf{h} \in \mathcal{C}^\perp$, we compute,

$$\begin{aligned} \langle \mathbf{y}, \mathbf{h} \rangle &= \langle \mathbf{e}, \mathbf{h} \rangle \\ &= \sum_{j \in \mathcal{P}} h_j e_j + \sum_{j \in \mathcal{N}} h_j e_j \\ &= \langle \mathbf{e}_{\mathcal{P}}, \mathbf{h}_{\mathcal{P}} \rangle + \langle \mathbf{e}_{\mathcal{N}}, \mathbf{h}_{\mathcal{N}} \rangle \end{aligned}$$

It gives access to the following LPN sample:

$$(\mathbf{a}, \langle \mathbf{s}, \mathbf{a} \rangle + e) \quad \text{where} \quad \begin{cases} \mathbf{s} \triangleq \mathbf{e}_{\mathcal{P}} \\ \mathbf{a} \triangleq \mathbf{h}_{\mathcal{P}} \\ e \triangleq \langle \mathbf{e}_{\mathcal{N}}, \mathbf{h}_{\mathcal{N}} \rangle \end{cases}$$

Here e follows a Bernoulli distribution that is a function of n , s and u (*resp.* w) the weight of \mathbf{e} (*resp.* \mathbf{h}) restricted to \mathcal{N} , namely

$$u \triangleq |\mathbf{e}_{\mathcal{N}}| \quad \text{and} \quad w \triangleq |\mathbf{h}_{\mathcal{N}}|.$$

The probability that e is equal to 1 is estimated through the following proposition which gives for the first time a rigorous statement for the standard assumption (1.2) made for statistical decoding.

Proposition 3.1. *Assume that the code \mathcal{C} is chosen by picking for it an $(n-k) \times n$ binary parity-check matrix uniformly at random. Let \mathcal{N} be a fixed set of $n-s$ positions in $\llbracket 1, n \rrbracket$ and \mathbf{e} be some error of weight u on \mathcal{N} . Choose \mathbf{h} uniformly at random among the parity-checks of \mathcal{C} of weight w on \mathcal{N} and \mathbf{h}' uniformly at random among the words of weight w on \mathcal{N} . Let $\delta \triangleq \text{bias}(\langle \mathbf{e}, \mathbf{h}' \rangle)$. If the parameters k , s , u , w are chosen as functions on n so that for n going to infinity, the expected number N of parity-checks of \mathcal{C} of weight w on \mathcal{N} satisfies $N = \omega(1/\delta^2)$ then for all but a proportion $o(1)$ of codes we have*

$$\text{bias}(\langle \mathbf{e}, \mathbf{h} \rangle) = (1 + o(1))\delta.$$

Proof. Let us define for $b \in \{0, 1\}$:

$$E_b \triangleq \#\{\mathbf{h} \in \mathcal{C}^\perp : |\mathbf{h}|_{\mathcal{N}} = w, \langle \mathbf{e}, \mathbf{h} \rangle = b\} \quad (3.1)$$

$$E'_b \triangleq \#\{\mathbf{h}' \in \mathbb{F}_2^n : |\mathbf{h}'|_{\mathcal{N}} = w, \langle \mathbf{e}, \mathbf{h}' \rangle = b\} \quad (3.2)$$

By using [Bar97, Lemma 1.1 p.10], we obtain

$$\mathbb{E}(E_b) = \frac{E'_b}{2^k} \quad (3.3)$$

$$\mathbf{Var}(E_b) \leq \frac{E'_b}{2^k} \quad (3.4)$$

By using now the Bienaymé-Tchebychev inequality, we obtain for any function f mapping the positive integers to positive real numbers:

$$\mathbb{P}_{\mathcal{C}} \left(|E_b - \mathbb{E}(E_b)| \geq \sqrt{f(n)\mathbb{E}(E_b)} \right) \leq \frac{1}{f(n)}. \quad (3.5)$$

Since $\text{bias}(\langle \mathbf{e}, \mathbf{h} \rangle) = \frac{E_0 - E_1}{E_0 + E_1}$ we have with probability greater than $1 - \frac{2}{f(n)}$ that

$$\frac{\mu_0 - \mu_1 - \sqrt{2f(n)}\sqrt{\mu_0 + \mu_1}}{\mu_0 + \mu_1 + \sqrt{2f(n)}\sqrt{\mu_0 + \mu_1}} \leq \text{bias}(\langle \mathbf{e}, \mathbf{h} \rangle) \leq \frac{\mu_0 - \mu_1 + \sqrt{2f(n)}\sqrt{\mu_0 + \mu_1}}{\mu_0 + \mu_1 - \sqrt{2f(n)}\sqrt{\mu_0 + \mu_1}} \quad (3.6)$$

where $\mu_i \triangleq \mathbb{E}(E_i)$ and where we used that for all positive x and y , $\sqrt{x} + \sqrt{y} \leq \sqrt{2(x+y)}$. We let $f(n) = \delta\sqrt{N}/2$. Since $N = \mu_0 + \mu_1$ this implies $f(n) = \delta\sqrt{\mu_0 + \mu_1}/2$. By the assumptions made in the proposition, note that $f(n)$ tends to infinity as n tends to infinity. We notice that

$$\begin{aligned} \sqrt{2f(n)}\sqrt{\mu_0 + \mu_1} &= \delta^{1/2}(\mu_0 + \mu_1)^{3/4} \\ &= o(\delta(\mu_0 + \mu_1)) \end{aligned} \quad (3.7)$$

because

$$\begin{aligned} \frac{\delta^{1/2}(\mu_0 + \mu_1)^{3/4}}{\delta(\mu_0 + \mu_1)} &= \frac{1}{\sqrt{\delta\sqrt{\mu_0 + \mu_1}}} \\ &= \frac{1}{\sqrt{2f(n)}} \\ &\rightarrow 0 \text{ as } n \rightarrow \infty. \end{aligned}$$

(3.6) can now be rewritten as

$$\frac{\mu_0 - \mu_1 - o(\delta(\mu_0 + \mu_1))}{\mu_0 + \mu_1 + o(\delta(\mu_0 + \mu_1))} \leq \text{bias}(\langle \mathbf{e}, \mathbf{h} \rangle) \leq \frac{\mu_0 - \mu_1 + o(\delta(\mu_0 + \mu_1))}{\mu_0 + \mu_1 - o(\delta(\mu_0 + \mu_1))} \quad (3.8)$$

Now on the other hand

$$\begin{aligned} \delta &= \text{bias}(\langle \mathbf{e}, \mathbf{h}' \rangle) = \frac{E'_0 - E'_1}{E'_0 + E'_1} = \frac{\frac{E'_0}{2^k} - \frac{E'_1}{2^k}}{\frac{E'_0}{2^k} + \frac{E'_1}{2^k}} \\ &= \frac{\mu_0 - \mu_1}{\mu_0 + \mu_1} \text{ (by (3.1)).} \end{aligned}$$

From this it follows that we can rewrite (3.8) as

$$\frac{\delta}{1 + o(\delta)} - o(\delta) \leq \text{bias}(\langle \mathbf{e}, \mathbf{h} \rangle) \leq \frac{\delta}{1 - o(\delta)} + o(\delta) \quad (3.9)$$

from which it follows immediately that

$$\text{bias}(\langle \mathbf{e}, \mathbf{h} \rangle) = \delta(1 + o(1)).$$

□

Remark 3.2. Note that the condition $N = \Omega(1/\delta^2)$, respectively $N = \Omega(s/\delta^2)$ is the condition we need in order that statistical decoding, respectively RLPN decoding succeed. This means that if we just have slightly more equations than the ratio $\frac{1}{\delta^2}$, then the standard assumption (1.2) made for statistical decoding holds. The point of this assumption is that it allows easily to estimate the bias as the following lemma shows.

Lemma 3.3. *Under the same assumptions made in Proposition 3.8 we have that for all but a proportion $o(1)$ of codes,*

$$\text{bias}(\langle \mathbf{e}_{\mathcal{N}}, \mathbf{h}_{\mathcal{N}} \rangle) = \delta(1 + o(1)) \quad \text{with} \quad \delta \triangleq \frac{K_w^{n-s}(u)}{\binom{n-s}{w}}$$

where $u \triangleq |\mathbf{e}_{\mathcal{N}}|$ and K_w^n stands for the Krawtchouk polynomial of order n and degree $w \in \llbracket 0, n \rrbracket$ which is defined as:

$$K_w^n(X) \triangleq \sum_{j=0}^w (-1)^j \binom{X}{j} \binom{n-X}{w-j}.$$

Proof. By using Proposition 3.8 (and the same notation as the one used there) we have that for all but a proportion $o(1)$ of codes

$$\text{bias}(\langle \mathbf{e}_{\mathcal{N}}, \mathbf{h}_{\mathcal{N}} \rangle) = (1 + o(1)) \text{bias}(\langle \mathbf{e}_{\mathcal{N}}, \mathbf{h}'_{\mathcal{N}} \rangle)$$

Now by definition of u we have

$$\begin{aligned} \text{bias}(\langle \mathbf{e}_{\mathcal{N}}, \mathbf{h}'_{\mathcal{N}} \rangle) &= \frac{1}{\binom{n-s}{w}} \sum_{j \text{ even}} \binom{u}{j} \binom{n-s-u}{w-j} - \frac{1}{\binom{n-s}{w}} \sum_{j \text{ odd}} \binom{u}{j} \binom{n-s-u}{w-j} \\ &= \frac{1}{\binom{n-s}{w}} \sum_j (-1)^j \binom{u}{j} \binom{n-s-u}{w-j} \\ &= \frac{K_w^{n-s}(u)}{\binom{n-s}{w}}. \end{aligned}$$

□

We will now repeatedly denote by *bias* of the LPN sample the quantity ε appearing in the previous lemma and the *estimated bias* the quantity namely

Definition 3.4 (bias of the LPN samples). *The bias ε of the LPN samples is defined by*

$$\varepsilon \triangleq \text{bias}(\langle \mathbf{e}_{\mathcal{N}}, \mathbf{h}_{\mathcal{N}} \rangle)$$

when $\mathbf{e}_{\mathcal{N}}$ has Hamming weight u and \mathbf{h} is drawn uniformly at random among the parity-check equations of weight w restricted on \mathcal{N} . The estimated bias is the quantity δ defined by

$$\delta \triangleq \text{bias}(\langle \mathbf{e}_{\mathcal{N}}, \mathbf{h}'_{\mathcal{N}} \rangle)$$

when $\mathbf{e}_{\mathcal{N}}$ has Hamming weight u and \mathbf{h}' is drawn uniformly at random among the binary words of weight w restricted on \mathcal{N} . This quantity is equal to

$$\delta = \frac{K_w^{n-s}(u)}{\binom{n-s}{w}}.$$

The point of introducing Krawtchouk polynomials is that we can bring in asymptotic expansions of Krawtchouk polynomials. From [IS98] (see also [DT17a, DT17b]) we obtain

Lemma 3.5. *Let n, t and w be three positive integers. We set $\tau \triangleq \frac{t}{n}$, $\omega \triangleq \frac{w}{n}$ and $\omega^\perp \triangleq \frac{1}{2} - \sqrt{\omega(1-\omega)}$.*

We assume $t \leq n/2$. Let $z \triangleq \frac{1-2\tau-\sqrt{D}}{2(1-\omega)}$ where $D \triangleq (1-2\tau)^2 - 4\omega(1-\omega)$. In the case $\tau \in [0, \omega^\perp]$,

$$K_w^n(t) = \sqrt{\frac{1-z^2}{2\pi\sqrt{D}n}} \frac{(1-z)^t(1+z)^{n-t}}{z^{w+\frac{1}{2}}} (1 + O(w^{-1})). \quad (3.10)$$

In the case $\tau \in [\omega^\perp, 1/2]$, D is negative, and

$$K_w^n(t) = \frac{2}{\sqrt{\pi n}} \left(\frac{\tau(1-\tau)}{-\omega(1-\omega)D} \right)^{\frac{1}{4}} 2^{\frac{t}{2}(1+h(\omega)-h(\tau))} \sin(n\theta + \theta_0 + o(1))(1 + o(1)) \quad (3.11)$$

where

$$\begin{aligned} \theta &= (\tau - 1) \cos^{-1} \left(\frac{\frac{3}{2} - \omega - \tau}{\sqrt{2(1-\omega)(1-\tau)}} \right) + \tau \cos^{-1} \left(\frac{\frac{1}{2} - \omega + \tau}{\sqrt{2\tau(1-\omega)}} \right) + \omega \cos^{-1} \left(\frac{1-2\tau}{b(\omega)} \right) \\ \theta_0 &= \frac{1}{2} \cos^{-1} \left(\frac{(1-2\tau)(1-2\omega)}{b(\tau)b(\omega)} \right) + \frac{\pi}{4} \end{aligned}$$

In the zone of zeros of the Krawtchouk polynomials, *i.e.* when $\tau \in [\omega^\perp, 1/2]$, the Krawtchouk polynomial may take 0 values and it is not clear whether by slightly changing w or t , the behavior of $\frac{\log_2(K_w^n(t))}{n}$ would be of the form $\frac{1+h(\omega)-h(\tau)}{2} + o(1)$. This is actually the case, since we can use

Lemma 3.6 ([KS21, Cor. 2.16]). *Between any two consecutive roots of K_w^n , where $1 \leq w \leq \frac{n}{2}$, there exists ℓ such that:*

$$\frac{\log_2(K_w^n(\ell))}{n} = \frac{1+h(\omega)-h(\lambda)}{2} + o(1), \quad \text{where } \omega \triangleq \frac{w}{n} \text{ and } \lambda \triangleq \frac{\ell}{n}.$$

We also know that

Lemma 3.7 ([KS21, §2.2, 5], [Lev95, Eq. (128)]). *For all $w \in \llbracket 1, n \rrbracket$, K_w^n has w distinct zeros, all in $[n\omega^\perp, n(1-\omega^\perp)]$ where $\omega^\perp \triangleq \frac{1}{2} - \sqrt{\omega(1-\omega)}$ and $\omega \triangleq \frac{w}{n}$. The smallest root is at distance $o(n)$ from $n\omega^\perp$, the largest at distance $o(n)$ from $n(1-\omega^\perp)$. The distance between any two consecutive roots is $o(n)$.*

By using these two lemmas, we readily obtain

Proposition 3.8 (exponential behavior of δ^2). *Let τ and ω be two reals in the interval $[0, \frac{1}{2}]$. Let $\omega^\perp \triangleq \frac{1}{2} - \sqrt{\omega(1-\omega)}$ and $z \triangleq \frac{1-2\tau-\sqrt{D}}{2(1-\omega)}$ where $D \triangleq (1-2\tau)^2 - 4\omega(1-\omega)$. There exists a sequence of positive integers $(t_n)_{n \in \mathbb{N}}$ and $(w_n)_{n \in \mathbb{N}}$, such that $\frac{t_n}{n} \xrightarrow{n \rightarrow \infty} \tau$, $\frac{w_n}{n} \xrightarrow{n \rightarrow \infty} \omega$ and $\frac{\log_2(K_{w_n}^n(t_n)^2 / \binom{n}{w_n}^2)}{n}$ has a limit which we denote $\tilde{\delta}(\tau, \omega)$ with*

$$\tilde{\delta}(\tau, \omega) = \begin{cases} 2(\tau \log_2(1-z) + (1-\tau) \log_2(1+z) - \omega \log_2 z - h(\omega)) & \text{if } \tau \in [0, \omega^\perp] \\ 1 - h(\tau) - h(\omega) & \text{otherwise.} \end{cases}$$

Proof. In the case $\tau \in [0, \omega^\perp]$ we just let $t_n = \lceil \tau n \rceil$, $w_n = \lceil \omega n \rceil$ and use directly the asymptotic expansion (3.10). In the case $\tau \in [\omega^\perp, \frac{1}{2}]$ we still define w_n with $w_n \triangleq \lceil \omega n \rceil$ but define t_n differently. For n large enough, we know from Lemma 3.7 that $\lceil \tau n \rceil$ lies between two zeros of the Krawtchouk polynomial and that there exists an integer t_n in this interval such that $\frac{\log_2(K_{w_n}^n(t_n))}{n} = \frac{1+h(\omega)-h(\tau_n)}{2} + o(1)$ where $\tau_n = \frac{t_n}{n}$. Now since the size of this interval is an $o(n)$ we necessarily have $\tau_n = \tau + o(1)$ and therefore $\frac{\log_2(K_{w_n}^n(t_n))}{n} = \frac{1+h(\omega)-h(\tau)}{2} + o(1)$. \square

The point of this proposition is that the term $2 \log_2(K_w^{n-s}(u) / \binom{n-s}{w})$ quantifies the exponential behavior of the square ε^2 of the bias ε (see Lemma 3.3) and $1/\varepsilon^2$ is up to polynomial terms the number of parity-checks we need for having enough information to solve the LPN problem as will be seen. This is because the capacity of the BSC($\frac{1-\varepsilon}{2}$) is $1 - h(\frac{1-\varepsilon}{2}) = \theta(\varepsilon^2)$ and that solving an LPN-problem with a secret of size s and N samples amounts to be able to decode a random linear code of rate $\frac{s}{N}$ over the BSC($\frac{1-\varepsilon}{2}$). It is therefore doable as soon as the rate is below the capacity (see Theorem 2.8). The reason why the Shannon capacity appears here is because of the following heuristic/assumption we will make here:

Assumption 3.9 (LPN modelling). *We will assume that the $\langle \mathbf{e}_{\mathcal{Y}}, \mathbf{h}_{\mathcal{Y}} \rangle$ are i.i.d Bernoulli random variables of parameter $\frac{1-\varepsilon}{2}$.*

Strictly speaking, the corresponding random variables are not independent, however we have performed extensive simulations showing that indeed decoding the corresponding code behaves similarly to this model. Some experiments with the programs that have been used for running them can be found on <https://github.com/tillich/RLPNdecoding>. Note that similar heuristics are also used to analyze a related lattice decoder making use of short dual lattice vectors (they are called dual attacks in the literature).

Fact 3.10. *With Assumption 3.9, the number N of LPN samples is such that $s/N = O(\varepsilon^2)$ for a small enough constant in the O , performing maximum-likelihood decoding of the corresponding $[N, s]$ binary code recovers the secret $\mathbf{e}_{\mathcal{P}}$ with probability $1 - o(1)$.*

Performing maximum likelihood decoding of the corresponding code can be achieved by a fast Fourier transform on a relevant vector. Indeed, for a given received word \mathbf{y} and a set \mathcal{H} of N parity-checks so that their restriction to \mathcal{P} leads to a set \mathcal{H} of N different vectors of \mathbb{F}_2^s , we let for $\mathbf{a} \in \mathcal{H}$, $\tilde{\mathbf{a}}$ be the unique parity-check in \mathcal{H} such that $\tilde{\mathbf{a}}_{\mathcal{P}} = \mathbf{a}$ and define $f_{\mathbf{y}, \mathcal{H}}$ as

$$f_{\mathbf{y}, \mathcal{H}} : \mathbb{F}_2^s \rightarrow \mathbb{R} \\ \mathbf{a} \mapsto \begin{cases} (-1)^{\langle \mathbf{y}, \tilde{\mathbf{a}} \rangle} & \text{if } \mathbf{a} \in \mathcal{H} \\ 0 & \text{otherwise} \end{cases} \quad (3.12)$$

We define the Fourier transform of such a function by

$$\widehat{f}(\mathbf{x}) \triangleq \sum_{\mathbf{u} \in \mathbb{F}_2^s} f(\mathbf{u}) (-1)^{\langle \mathbf{x}, \mathbf{u} \rangle}.$$

The code \mathcal{D} we want to decode (obtained via our LPN samples) is described as

$$\mathcal{D} \triangleq \{ \mathbf{c}_{\mathbf{x}}, \mathbf{x} \in \mathbb{F}_2^s \} \text{ where } \mathbf{c}_{\mathbf{x}} \triangleq (\langle \mathbf{x}, \mathbf{a} \rangle)_{\mathbf{a} \in \mathcal{H}}, \quad (3.13)$$

and the word $\mathbf{u}_{\mathbf{y}, \mathcal{H}}$ we want to decode is given by $\mathbf{u}_{\mathbf{y}, \mathcal{H}} = (\langle \mathbf{y}, \tilde{\mathbf{a}} \rangle)_{\mathbf{a} \in \mathcal{H}}$. It is readily seen that

$$\begin{aligned} \widehat{f_{\mathbf{y}, \mathcal{H}}}(\mathbf{x}) &= \sum_{\mathbf{a} \in \mathbb{F}_2^s} f(\mathbf{a}) (-1)^{\langle \mathbf{x}, \mathbf{a} \rangle} \\ &= \sum_{\mathbf{a} \in \mathcal{H}} (-1)^{\langle \mathbf{x}, \mathbf{a} \rangle + \langle \mathbf{y}, \tilde{\mathbf{a}} \rangle} \\ &= \#\mathcal{H} - 2|\mathbf{u}_{\mathbf{y}, \mathcal{H}} - \mathbf{c}_{\mathbf{x}}|, \end{aligned}$$

In other words, finding the closest codeword to $\mathbf{u}_{\mathbf{y}, \mathcal{H}}$ is nothing but finding the \mathbf{x} which maximizes $\widehat{f_{\mathbf{y}, \mathcal{H}}}(\mathbf{x})$. This is achieved in time $O(s2^s)$ by performing a fast Fourier transform. Notice that an exhaustive search would cost $O(2^{2s})$.

3.2 Sketch of the whole algorithm

Algorithm 3.1. RLPN decoder

Input: \mathbf{y} , t , \mathcal{C} an $[n, k]$ -code

Output: \mathbf{e} such that $|\mathbf{e}| = t$ and $\mathbf{y} - \mathbf{e} \in \mathcal{C}$.

function RLPNDECODE(\mathbf{y} , \mathcal{C} , t)

$s, u \leftarrow \text{OPTIM}(t, k, n)$

$\triangleright s$ and u in order to minimize the complexity of the following procedure.

for i from 1 to N_{iter} **do**

$\triangleright N_{\text{iter}}$ is a certain function of n, s, t and u .

$\mathcal{P} \xleftarrow{\$} \{ \mathcal{J} \subseteq [1, n] : \#\mathcal{J} = s \}$

$\mathcal{N} \leftarrow [1, n] \setminus \mathcal{P}$

$\mathcal{H} \leftarrow \text{CREATE}(N, w, \mathcal{P})$

$\widehat{f_{\mathbf{y}, \mathcal{H}}} \leftarrow \text{FFT}(f_{\mathbf{y}, \mathcal{H}})$

$\mathbf{x}_0 \leftarrow \arg \max f_{\mathbf{y}, \mathcal{H}}$

if $\widehat{f_{\mathbf{y}, \mathcal{H}}}(\mathbf{x}_0) \geq \frac{\delta N}{2}$ **then**

$\triangleright \delta \triangleq K_w^{n-s}(u) / \binom{n-s}{w}$.

return \mathbf{e} such that $\mathbf{e}_{\mathcal{P}} = \mathbf{x}_0$ and $\mathbf{e}_{\mathcal{N}} = \text{RLPNDECODE}(\mathbf{y}_{\mathcal{N}}, \mathcal{C}_{\mathcal{N}}, t - |\mathbf{x}_0|)$

end if

end for

end function

Besides, the fast Fourier transform solving the LPN problem, Algorithm 3.1 uses two other ingredients:

- A routine $\text{CREATE}(N, w, \mathcal{P})$ creating a set \mathcal{H} of N parity-check equations \mathbf{h} such that $|\mathbf{h}_{\mathcal{N}}| = w$ where $\mathcal{N} \triangleq [1, n] \setminus \mathcal{P}$. We will not specify how this function is realized here: this is done in the following sections. This procedure together with an FFT for decoding the code associated to the parity-check equations in \mathcal{H} (see Equation (3.13)) form the inner loop of our algorithm.
- An outer loop making a certain number N_{iter} of calls to the inner procedure, checking each time a new set \mathcal{P} of s positions with the hope of finding an \mathcal{N} containing an unusually low number u of errors in it. The point is that with a right u , the number of times we will have to check a new \mathcal{P} is outweighed by the decrease in N because the bias δ is much higher for such a u .

3.3 Analysis of the RLPN decoder

We need to show now that our RLPN decoder returns what we expect.

Proposition 3.11 (acceptation criteria). *Under Assumption 3.9, by choosing $N_{\text{iter}} = \omega\left(\frac{1}{P_{\text{succ}}}\right)$, $s = \omega(1)$ and $N = \omega\left(\frac{n}{\delta^2}\right)$, we have with probability $1 - o(1)$ that at least one iteration is such that $\mathbf{e}_{\mathcal{P}}$ meets the acceptation criteria $\widehat{f_{\mathbf{y}, \mathcal{H}}}(\mathbf{e}_{\mathcal{P}}) \geq \frac{\delta N}{2}$. Moreover, the probability that there exists $\mathbf{x} \neq \mathbf{e}_{\mathcal{P}}$ which meets this acceptation criteria is $o(1)$.*

Proof. We need to show that two things happen both with probability $1 - o(1)$:

(i) there is at least one iteration in the Algorithm 3.1 for which $\widehat{f_{\mathbf{y}, \mathcal{H}}}(\mathbf{e}_{\mathcal{P}}) \geq \frac{\delta N}{2}$

(ii) for all $\mathbf{x} \in \mathbb{F}_2^s$ different from $\mathbf{e}_{\mathcal{P}}$, we have $\widehat{f_{\mathbf{y}, \mathcal{H}}}(\mathbf{x}) < \frac{\delta N}{2}$ for all iterations.

(i) The first point follows from the fact that by taking $N_{\text{iter}} \triangleq \omega\left(\frac{1}{P_{\text{succ}}}\right)$, we have that one iteration is such that $|\mathbf{e}_{\mathcal{N}}| = u$ and $|\mathbf{e}_{\mathcal{P}}| = t - u$ with probability $1 - o(1)$. For such an iteration, we have from Assumption 3.9 that $|\mathbf{c}_{\mathbf{e}_{\mathcal{P}}} - \mathbf{u}_{\mathbf{y}, \mathcal{H}}| \sim \text{Binomial}\left(N, \frac{1-\varepsilon}{2}\right)$. Thus, by using the Hoeffding inequality, we have:

$$\mathbb{P}\left(\widehat{f_{\mathbf{y}, \mathcal{H}}}(\mathbf{e}_{\mathcal{P}}) \geq \frac{\delta N}{2}\right) = \mathbb{P}\left(|\mathbf{c}_{\mathbf{e}_{\mathcal{P}}} - \mathbf{u}_{\mathbf{y}, \mathcal{H}}| \leq \frac{1 - \frac{\delta}{2}}{2} N\right) \quad (3.14)$$

$$\geq 1 - \exp\left(-\frac{(\varepsilon - \delta/2)^2 N}{2}\right) \quad (3.15)$$

By the choice made on N this quantity is $1 - o(1)$.

(ii) For the second point, consider now an $\mathbf{x} \in \mathbb{F}_2^s$ such that $\mathbf{x} \neq \mathbf{e}_{\mathcal{P}}$. Let $\mathbf{c}' \triangleq \mathbf{c}_{\mathbf{x}} - \mathbf{c}_{\mathbf{e}_{\mathcal{P}}}$, $d \triangleq |\mathbf{c}'|$ and $\mathbf{y}' \triangleq \mathbf{u}_{\mathbf{y}, \mathcal{H}} - \mathbf{c}_{\mathbf{e}_{\mathcal{P}}}$. Then we have :

$$\begin{aligned} \mathbb{P}\left(\widehat{f_{\mathbf{y}, \mathcal{H}}}(\mathbf{x}) \geq \frac{\delta N}{2}\right) &= \mathbb{P}\left(|\mathbf{c}' - \mathbf{y}'| \leq \frac{1 - \frac{\delta}{2}}{2} N\right) \quad (\text{because } |\mathbf{c}_{\mathbf{x}} - \mathbf{u}_{\mathbf{y}, \mathcal{H}}| = |\mathbf{c}' - \mathbf{y}'|) \\ &= \sum_{b=0}^N \sum_{a=0}^{\frac{1-\delta}{2} N} \mathbb{P}\left(|\mathbf{c}' - \mathbf{y}'| = a \mid |\mathbf{y}'| = b\right) \cdot \mathbb{P}(|\mathbf{y}'| = b) \\ &\leq \sum_{b=0}^N \sum_{a=0}^{\frac{1-\delta}{2} N} \mathbb{P}_{\mathbf{y}'}\left(|\mathbf{c}' - \mathbf{y}'| = a \mid |\mathbf{y}'| = b\right) \\ &\leq \sum_{b=0}^N \sum_{a=0}^{\frac{1-\delta}{2} N} \frac{\binom{d+b-a}{2} \binom{N-b}{2}}{\binom{N}{d}}. \end{aligned} \quad (3.16)$$

The last point follows from the fact that $|\mathbf{y}'| \sim \text{Binomial}\left(N, \frac{1-\varepsilon'}{2}\right)$ for some ε' depending on $|\mathbf{e}_{\mathcal{N}}|$ and therefore the conditional distribution of \mathbf{y}' given that $|\mathbf{y}'| = b$ is the uniform distribution of

words of weight b over \mathbb{F}_2^N . This implies

$$\mathbb{P}_{\mathbf{y}'} \left(|\mathbf{c}' - \mathbf{y}'| = a \mid |\mathbf{y}'| = b \right) = \frac{\binom{b}{\frac{d+b-a}{2}} \binom{N-b}{\frac{d-b+a}{2}}}{\binom{N}{d}}.$$

Note that both $\mathbf{c}_{\mathbf{x}}$ and $\mathbf{c}_{\mathbf{e}_{\mathcal{D}}}$ are in the code \mathcal{D} defined by Equation (3.13). We can bound again the (typical) minimum nonzero weight and the maximum weight by using [Bar97, Lemma 1.1, p.10]. Let $\mathcal{B} = \{\mathbf{x} \in \mathbb{F}_2^N : 0 < |\mathbf{x}| \leq Nh^{-1}(1 - 2s/N)\}$. The expected number M of codewords of \mathcal{D} in \mathcal{B} is $\frac{\#\mathcal{B}}{2^{N(1-s/N)}}$. Since $\#\mathcal{B} \leq 2^{Nh(h^{-1}(1-2s/N))} = 2^{N(1-2s/N)}$ we have that $M \leq 2^{-s}$. Since the probability that the minimum distance of \mathcal{D} is less than or equal to $Nh^{-1}(1 - 2s/N)$ is upper-bounded by M , we obtain that the minimum distance of \mathcal{D} is greater than $Nh^{-1}(1 - 2s/N)$ with probability $1 - o(1)$. A similar reasoning can be made for the maximum weight. We therefore obtain that with probability $1 - o(1)$ all the weights d of the non-zero codewords of \mathcal{D} lie in $\llbracket d^-, d^+ \rrbracket$ where

$$\begin{aligned} d^- &\triangleq Nh^{-1}(1 - 2s/N) \\ d^+ &\triangleq N - Nh^{-1}(1 - 2s/N). \end{aligned}$$

In such a case, we always have for n large enough

$$a \leq d \leq N - a \quad (3.17)$$

$$1 - 2s/N \leq h(d/N) \leq 1. \quad (3.18)$$

We observe now that

$$\begin{aligned} \frac{\binom{b}{\frac{d+b-a}{2}} \binom{N-b}{\frac{d-b+a}{2}}}{\binom{N}{d}} &\leq 2^{bh(\frac{1}{2} - \frac{d-a}{2b}) + (N-b)h(\frac{1}{2} - \frac{N-d-a}{2(N-b)}) - Nh(\frac{d}{N})} \quad (\text{because of (3.17)}) \\ &\leq 2^{\min[-b(1-h(\frac{1}{2} - \frac{d-a}{2b})) + 2s, -(N-b)(1-h(\frac{1}{2} - \frac{d-a}{2(N-b)})) + 2s]} \quad (\text{because of (3.18)}) \\ &\leq 2^{-\frac{N}{2}(1-h(\frac{1}{2} - \frac{d-a}{2N})) + 2s} \quad (3.19) \end{aligned}$$

The last inequality comes from the fact that either b or $N - b$ is greater than $\frac{N}{2}$ and that both of them are smaller than N . Since $h^{-1}(1 - u) = \frac{1}{2} - \Theta(\sqrt{u})$ for $u \rightarrow 0^+$ we have that $N/2(1 - \Theta(\sqrt{\frac{s}{N}})) \leq d \leq N/2(1 + \Theta(\sqrt{\frac{s}{N}}))$ and therefore

$$\frac{d-a}{2N} \geq \frac{\delta/2 - \Theta(\sqrt{\frac{s}{N}})}{4} = \Omega(\delta). \quad (3.20)$$

By using that $h(1/2 - u) = 1 - \Theta(u^2)$ for $u \rightarrow 0$ together with (3.20) in (3.19) we obtain

$$\frac{\binom{b}{\frac{d+b-a}{2}} \binom{N-b}{\frac{d-b+a}{2}}}{\binom{N}{d}} \leq 2^{-\Omega(N\delta^2) + 2s}. \quad (3.21)$$

By plugging this inequality in (3.16) we finally obtain

$$\mathbb{P} \left(\widehat{f_{\mathbf{y}, \mathcal{A}}}(\mathbf{x}) \geq \frac{\delta N}{2} \right) \leq N^2 2^{-\Omega(N\delta^2) + 2s}$$

and the probability of the event “there exists an iteration and an $\mathbf{x} \neq \mathbf{e}_{\mathcal{D}}$ such that $\widehat{f_{\mathbf{y}, \mathcal{A}}}(\mathbf{x}) \geq \frac{\delta N}{2}$ ” is upperbounded by

$$N_{\text{iter}} 2^s N^2 2^{-\Omega(N\delta^2) + 2s} = o(1).$$

□

The space and time complexity of this method are readily seen to be given by

Proposition 3.12. *Assume that $\text{CREATE}(N, w, \mathcal{P})$ produces N parity-check equations in space S_{eq} and time T_{eq} . The probability P_{succ} (over the choice of \mathcal{N}) that there are exactly u errors in \mathcal{N} is given by $P_{succ} = \frac{\binom{s}{t-u} \binom{n-s}{u}}{\binom{n}{t}}$. The space complexity S and the time complexity T of the RLPN-decoder are given by*

$$\text{Space: } S = O(S_{eq} + 2^s), \quad \text{Time: } T = \tilde{O}\left(\frac{T_{eq} + 2^s}{P_{succ}}\right).$$

The parameters s , u and w have to meet the following constraints

$$N \leq 2^s \tag{3.22}$$

$$N \leq \frac{\binom{n-s}{w}}{2^{k-s}}. \tag{3.23}$$

The algorithm outputs the correct $\mathbf{e}_{\mathcal{P}}$ with probability $1 - o(1)$ if in addition we choose N and N_{iter} such that

$$N = \omega \left(n \left(\frac{\binom{n-s}{w}}{K_w^{n-s}(u)} \right)^2 \right) \tag{3.24}$$

$$N_{iter} = \omega \left(\frac{1}{P_{succ}} \right). \tag{3.25}$$

Proof. All the points are straightforward here, with the exception of the constraints. The first constraint is that the number of parity-checks should not be bigger than the total number of different LPN samples we can possibly produce. The second one is that the number of parity-checks needed is smaller than the number of available parity-checks. The conditions ensuring the correctness of the algorithm follow immediately from Proposition 3.11. \square

The proof of the correctness of the algorithm relies on the validity of the LPN modeling (Assumption 3.9). We have programmed this algorithm and have verified that for several parameters that it really behaves as predicted. The corresponding experiments with the programs that have been used for running them can be found on <https://github.com/tillich/RLPNdecoding>.

3.4 Using several independent LPN samples $\langle \mathbf{a}, \mathbf{e}_{\mathcal{P}} \rangle + \mathbf{e}$ with the same \mathbf{a}

It can be noticed that when N is large enough, we may actually have several different parity-checks \mathbf{h} which collide on \mathcal{P} , *i.e.* lead to the same $\mathbf{h}_{\mathcal{P}}$. This leads to different LPN samples $\langle \mathbf{a}, \mathbf{e}_{\mathcal{P}} \rangle + \mathbf{e}$ with the same \mathbf{a} . A simple way to deal with this would be simply to keep just one of them. However, there is a much better way to deal with this issue, which is to use them to decrease the LPN noise. It uses the following lemma, which is for instance one of the two main ingredients used for decoding LDPC codes.

Lemma 3.13 ([Gal63]). *Assume that we are given $\mathbf{a} \in \mathbb{F}_2^s$ and M LPN samples $y_i = \langle \mathbf{a}, \mathbf{s} \rangle + e_i$ where the e_i 's are independent and $p_i \triangleq \mathbb{P}(e_i = 1)$. \mathbf{s} is distributed uniformly at random in \mathbb{F}_2^s . Define $p'_i \triangleq p_i$ if $y_i = 0$ and $p'_i \triangleq 1 - p_i$ otherwise. Then*

$$\mathbb{P}(\langle \mathbf{a}, \mathbf{s} \rangle = 1) = \frac{\prod_{i=1}^M p'_i}{\prod_{i=1}^M p'_i + \prod_{i=1}^M (1 - p'_i)}. \tag{3.26}$$

This allows to obtain a finer information on $\langle \mathbf{a}, \mathbf{s} \rangle$. We let $p(\mathbf{a})$ be the probability obtained in (3.26). We define now $f_{\mathbf{y}, \mathcal{H}}$ slightly differently by

$$f_{\mathbf{y}, \mathcal{H}} : \mathbb{F}_2^s \rightarrow \mathbb{R} \\ \mathbf{a} \mapsto \begin{cases} \log \left(\frac{1 - p(\mathbf{a})}{p(\mathbf{a})} \right) & \text{if } \mathbf{a} \in \mathcal{H} \\ 0 & \text{otherwise} \end{cases} \tag{3.27}$$

A straightforward computation leads now to

$$\begin{aligned}
\widehat{f_{\mathbf{y}, \mathcal{H}}}(\mathbf{x}) &= \sum_{\mathbf{a} \in \mathbb{F}_2^s} f(\mathbf{a})(-1)^{\langle \mathbf{x}, \mathbf{a} \rangle} \\
&= \sum_{\mathbf{a} \in \mathcal{H}, \langle \mathbf{a}, \mathbf{x} \rangle = 0} \log \left(\frac{1-p(\mathbf{a})}{p(\mathbf{a})} \right) + \sum_{\mathbf{a} \in \mathcal{H}, \langle \mathbf{a}, \mathbf{x} \rangle = 1} \log \left(\frac{p(\mathbf{a})}{1-p(\mathbf{a})} \right) \\
&= \log \left(\prod_{\mathbf{a} \in \mathcal{H}, \langle \mathbf{a}, \mathbf{x} \rangle = 0} \frac{1-p(\mathbf{a})}{p(\mathbf{a})} \prod_{\mathbf{a} \in \mathcal{H}, \langle \mathbf{a}, \mathbf{x} \rangle = 1} \frac{p(\mathbf{a})}{1-p(\mathbf{a})} \right) \\
&= \log \left(\frac{\left(\prod_{\mathbf{a} \in \mathcal{H}, \langle \mathbf{a}, \mathbf{x} \rangle = 0} (1-p(\mathbf{a})) \prod_{\mathbf{a} \in \mathcal{H}, \langle \mathbf{a}, \mathbf{x} \rangle = 1} p(\mathbf{a}) \right)^2}{\underbrace{\prod_{\mathbf{a} \in \mathcal{H}} p(\mathbf{a})(1-p(\mathbf{a}))}_{\text{constant}}} \right)
\end{aligned}$$

It is readily seen that since the function $x \mapsto \log x^2$ is increasing and that the denominator $\prod_{\mathbf{a} \in \mathcal{H}} p(\mathbf{a})(1-p(\mathbf{a}))$ does not depend on \mathbf{x} , the \mathbf{x} maximizing this quantity is precisely the \mathbf{x} which maximizes the quantity

$$g(\mathbf{x}) \triangleq \prod_{\mathbf{a} \in \mathcal{H}, \langle \mathbf{a}, \mathbf{x} \rangle = 0} (1-p(\mathbf{a})) \prod_{\mathbf{a} \in \mathcal{H}, \langle \mathbf{a}, \mathbf{x} \rangle = 1} p(\mathbf{a})$$

which in turn is clearly the maximum-likelihood candidate. In other words, we just have to replace in Algorithm 3.1 the function $f_{\mathbf{y}, \mathcal{H}}$ with this new definition. It remains to understand how the necessary number of needed parity-check equations evolves. This is done through the following lemma.

Lemma 3.14. *Assume that we choose uniformly at random a bit b and we pass it through M different BSC channels of crossover probability $\frac{1-\varepsilon}{2}$ and get as outputs (y_1, \dots, y_M) . The BIMS channel corresponding to the output (y_1, \dots, y_M) is of capacity $\Omega(M\varepsilon^2)$ as long as $M\varepsilon^2 = O(1)$.*

Proof. We bring in the Bhattacharyya parameter Z of a BIMS channel with transition probabilities $p(y|b)$ by

$$Z = \sum_y \sqrt{p(y|0)p(y|1)}.$$

The Bhattacharyya parameter of a BSC channel of crossover probability $\frac{1-\varepsilon}{2}$ is equal to $Z(\varepsilon) = 2\sqrt{\frac{1-\varepsilon}{2}\frac{1+\varepsilon}{2}} = \sqrt{1-\varepsilon^2}$. By [Kha02, (4.6) p.77] we know that the Bhattacharyya parameter Z of our BIMS channel satisfies

$$Z = Z(\varepsilon)^M = (1-\varepsilon^2)^{M/2} = 1 - \Theta(M\varepsilon^2), \quad (3.28)$$

as long as $M\varepsilon^2 = O(1)$. In the case of BIMS channel the capacity C is lower-bounded by the Bhattacharyya parameter Z as (see [Ari09, prop. 1]⁵)

$$C \geq \log \frac{2}{1+Z}.$$

This quantity is readily seen to be of order $\Theta(M\varepsilon^2)$ by using (3.28). \square

⁵ In this paper, this is actually a lower bound on the *symmetric capacity*, but capacity and symmetric capacity coincide for a BIMS channel.

By putting all these ingredients together, we obtain a slightly more complicated algorithm taking into account multiple copies of LPN samples. The structure of the algorithm is the same, the only difference being how $f_{\mathbf{y}, \mathcal{N}}$ is defined. We now use (3.27) instead of (3.12). This allows to obtain a slight generalization of Proposition 3.12 and remove the constraint (3.22) in it.

Proposition 3.15. *Assume that $\text{CREATE}(N, w, \mathcal{P})$ produces N parity-check equations in space S_{eq} and time T_{eq} . The probability P_{succ} (over the choice of \mathcal{N}) that there are exactly u errors in \mathcal{N} is given by $P_{succ} = \frac{\binom{s}{t-u} \binom{n-s}{u}}{\binom{n}{t}}$. The space complexity S and the time complexity T of the RLPN-decoder are given by*

$$\mathbf{Space: } S = O(S_{eq} + 2^s), \quad \mathbf{Time: } T = \tilde{O}\left(\frac{T_{eq} + 2^s}{P_{succ}}\right).$$

The parameters s , u and w have to meet the constraint $N \leq \frac{\binom{n-s}{w}}{2^{k-s}}$. The algorithm outputs the correct $\mathbf{e}_{\mathcal{P}}$ with probability $1 - o(1)$ if in addition we choose N and N_{iter} such that

$$N = \omega\left(n \left(\frac{\binom{n-s}{w}}{K_w^{n-s}(u)}\right)^2\right) \quad (3.29)$$

$$N_{iter} = \omega\left(\frac{1}{P_{succ}}\right). \quad (3.30)$$

Proof. If $N \leq 2^s$ there is nothing to prove. However, if N is bigger, say $N = M2^s$ with $M > 1$, the length of the code we decode with the FFT is $N' = \Theta(2^s)$ and we use Lemma 3.14 and say that decoding this $[N', s]$ -code is successful if $N' = \omega\left(\frac{n}{M\varepsilon^2}\right)$. By multiplying both terms by M , $N = \omega\left(\frac{n}{\varepsilon^2}\right)$. We replace the bias ε by its expression coming from Lemma 3.3 to finish the proof. \square

4 Collision techniques for finding low weight parity-checks

4.1 Using the [Dum86] method

A way for creating parity-checks with a low weight on \mathcal{N} is simply to use subset-sum/collision techniques [Dum86, Ste88, Dum89]. We start here with the simplest method for performing such a task pioneered by Dumer in [Dum86]. Consider a parity-check matrix \mathbf{H} for the code \mathcal{C} we want to decode and keep only the columns belonging to \mathcal{N} to obtain an $(n-k) \times (n-s)$ matrix $\mathbf{H}_{\mathcal{N}}$. The row-space of $\mathbf{H}_{\mathcal{N}}$ generates the restrictions $\mathbf{h}_{\mathcal{N}}$ to \mathcal{N} of the parity-checks \mathbf{h} of \mathcal{C} . This row-space is nothing but the dual code \mathcal{C}^{\perp} punctured in \mathcal{P} , *i.e.* we keep only the positions in \mathcal{N} . With our notation, this is $\mathcal{C}_{\mathcal{N}}^{\perp}$ and is an $[n-s, n-k]$ -code. Therefore if we want to find parity-checks \mathbf{h} of \mathcal{C} such $|\mathbf{h}_{\mathcal{N}}| = w$, this amounts to find codewords of $\mathcal{C}_{\mathcal{N}}^{\perp}$ of weight w . For this, we compute a parity-check matrix \mathbf{H}' of $\mathcal{C}_{\mathcal{N}}^{\perp}$ *i.e.* a $(k-s) \times (n-s)$ matrix such that

$$\mathcal{C}_{\mathcal{N}}^{\perp} = \{\mathbf{c} \in \mathbb{F}_2^{n-s} : \mathbf{H}'\mathbf{c}^{\top} = 0\}.$$

We split such a matrix in two parts⁶ of the same size $\mathbf{H}' = (\mathbf{H}_1 \mathbf{H}_2)$. We obtain an algorithm of time and space complexity, T and S respectively, producing N codewords of weight w , with

$$N = \frac{\left(\frac{n-s}{2}\right)^2}{2^{k-s}}(1 + o(1)) \quad \text{and} \quad S = T = O\left(\left(\frac{n-s}{2}\right) + N\right).$$

⁶ To simplify the presentation, the cut is explained by taking the first $\frac{n-s}{2}$ positions for the first part and the $\frac{n-s}{2}$ for the second part, but of course in general these positions are randomly chosen.

The algorithm for producing such codewords is to set up two lists,

$$\mathcal{L}_1 \triangleq \left\{ (\mathbf{H}_1 \mathbf{h}_1^\top, \mathbf{h}_1) : |\mathbf{h}_1| = \frac{w}{2}, \mathbf{h}_1 \in \mathbb{F}_2^{\frac{n-s}{2}} \right\}$$

$$\mathcal{L}_2 \triangleq \left\{ (\mathbf{H}_2 \mathbf{h}_2^\top, \mathbf{h}_2) : |\mathbf{h}_2| = \frac{w}{2}, \mathbf{h}_2 \in \mathbb{F}_2^{\frac{n-s}{2}} \right\}$$

and looking for collisions $\mathbf{H}_1 \mathbf{h}_1^\top = \mathbf{H}_2 \mathbf{h}_2^\top$ in the lists. It yields vectors $\mathbf{h}' = \mathbf{h}_1 \parallel \mathbf{h}_2$ of weight w which are in $\mathcal{C}_\mathcal{N}^\perp$ since $\mathbf{H}' \mathbf{h}'^\top = \mathbf{H}_1 \mathbf{h}_1^\top + \mathbf{H}_2 \mathbf{h}_2^\top = \mathbf{0}$. These vectors in \mathbb{F}_2^{n-s} can be completed to give vectors $\mathbf{h} \in \mathbb{F}_2^n$ such that $\mathbf{h}_\mathcal{N} = \mathbf{h}'$. The number of collisions is expected to be of order $\left(\frac{n-s}{\frac{w}{2}}\right)^2 / 2^{k-s}$ since $2^{-(k-s)}$ is the collision probability of two vectors in \mathbb{F}_2^{k-s} . The algorithm for performing this task is given by Algorithm 4.1.

Algorithm 4.1. Creating low weight parity-checks by collisions

Input $\mathcal{C}, w, \mathcal{P}$
Output a list of parity-check equations \mathbf{h} of \mathcal{C} such that $|\mathbf{h}_\mathcal{N}| = w$ where $\mathcal{N} \triangleq [1, n] \setminus \mathcal{P}$.
function CREATE($\mathcal{C}, w, \mathcal{P}$)
 $\mathbf{H} \leftarrow$ PARITY-CHECK-MATRIX($\mathcal{C}^\perp, \mathcal{P}$)
 \triangleright returns a parity-check matrix for \mathcal{C}^\perp with an identity corresponding to the positions in \mathcal{P} :
 $\mathbf{H} = \begin{pmatrix} \mathbf{I} & \mathbf{P} \\ \mathbf{0} & \mathbf{H}' \end{pmatrix}$ where we assume that the first block corresponds to the positions of \mathcal{P} .
 $\mathcal{L}_1 \leftarrow \{(\mathbf{H}_1 \mathbf{h}_1^\top, \mathbf{h}_1) : |\mathbf{h}_1| = w/2, \mathbf{h}_1 \in \mathbb{F}_2^{\frac{n-s}{2}}\}$
 $\mathcal{L}_2 \leftarrow \{(\mathbf{H}_2 \mathbf{h}_2^\top, \mathbf{h}_2) : |\mathbf{h}_2| = w/2, \mathbf{h}_2 \in \mathbb{F}_2^{\frac{n-s}{2}}\}$
 \triangleright We assume $\mathbf{H}' = (\mathbf{H}_1 \ \mathbf{H}_2)$, with \mathbf{H}_1 and \mathbf{H}_2 of the same size.
 $\mathcal{L} \leftarrow \{\mathbf{h}_1 \parallel \mathbf{h}_2 \in \mathcal{L}_1 \times \mathcal{L}_2 : \mathbf{H}_1 \mathbf{h}_1^\top = \mathbf{H}_2 \mathbf{h}_2^\top\}$
return $\{\mathbf{h}' \mathbf{P}^\top \parallel \mathbf{h}' : \mathbf{h}' \in \mathcal{L}\}$
 \triangleright It is straightforward to check that $\mathbf{h}' \mathbf{P}^\top \parallel \mathbf{h}'$ belongs to \mathcal{C}^\perp .
end function

We have represented in Figure 4.1 the form of the parity-checks output by this method, together with the bet we make on the error.

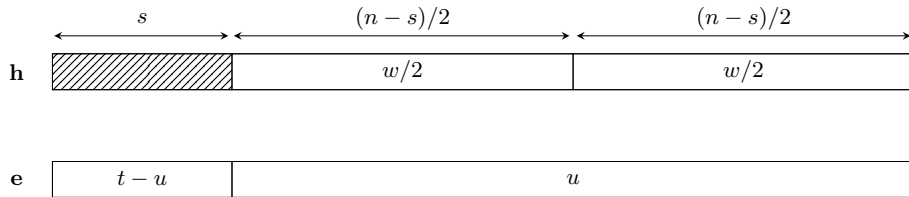


Figure 4.1. The form of the parity-checks produced by this method, *vs.* the bet made on the error. The hatched rectangle of size s for \mathbf{h} indicates that the weight is arbitrary on this part.

The amortized cost for producing a parity-check equation of weight w is $O(1)$ as long as $N \geq \Omega\left(\left(\frac{n-s}{\frac{w}{2}}\right)\right)$. It is insightful to consider the smallest value of w for which $\left(\frac{n-s}{\frac{w}{2}}\right) \leq \left(\frac{n-s}{\frac{w}{2}}\right)^2 / 2^{k-s}$. This is roughly speaking the smallest value (up to negligible terms) of w for which the amortized cost for producing parity-check equations of weight w is $O(1)$ per equation. In such a case, we roughly have

$$N \approx \left(\frac{n-s}{\frac{w}{2}}\right) \approx \frac{\left(\frac{n-s}{\frac{w}{2}}\right)^2}{2^{k-s}} \approx 2^{k-s}.$$

In other words with this choice we have

$$T_{\text{eq}} = O(2^{k-s}).$$

Let us choose now u as the “typical error weight” when restricted to \mathcal{N} , namely $u \approx t \frac{n-s}{n}$ and s such that the decoding complexity of the $[N, s]$ code is also of order the codelength, *i.e.* $N = \tilde{\Theta}(2^s)$. This would imply $2^s \approx 2^{k-s}$, which means that we are going to choose $s = \frac{k}{2}$. By using Proposition 3.12, all these choices would yield a time complexity T_{Dumer86} for decoding \mathcal{C} which would be of order

$$T_{\text{Dumer86}} = \tilde{O}\left(2^{k/2}\right), \quad (4.1)$$

if the constraint $N = \tilde{\Omega}\left(\left(\frac{\binom{n-s}{w}}{K_w^{n-s}(u)}\right)^2\right)$ for successful decoding the $[N, s]$ -code is met. This amounts to $2^{Rn/2} = \tilde{\Omega}\left(\left(\frac{\binom{n(1-R/2)}{w}}{K_w^{n(1-R/2)}(t(1-R/2))}\right)^2\right)$, where R is the code rate, *i.e.* $R = \frac{k}{n}$. By using Proposition 3.8, we can give an asymptotic formula for this constraint. It translates into

$$R/2 \geq 2(1 - R/2)\tilde{\varepsilon}(\tau, \omega/(1 - R/2)),$$

where $\tilde{\varepsilon}$ is the function defined in Proposition 3.8. Amazingly enough this constraint is met up to very small values of R , it is only below $R \approx 0.02$ that this condition is not met anymore. This innocent looking remark has actually very concrete consequences. This means that above the range $R \gtrsim 0.02$ the asymptotic complexity exponent, *i.e.* $\alpha_{\text{Dumer86}} \triangleq \limsup_n \log_2 T_{\text{Dumer86}}/n$ where T_{Dumer86} is the time complexity, satisfies

$$\alpha_{\text{Dumer86}} \leq \frac{R}{2}. \quad (4.2)$$

This is very surprising, since in the vicinity of $R \approx 0$ the asymptotic time complexity of *all known* decoding methods approach quickly R . In other words, in this regime, the complexity is of order $T \approx 2^{Rn} = 2^k$ for full distance (*a.k.a* GV) decoding, meaning that they are not better than exhaustive search. Unfortunately this is also the case for our method. It can namely be proved that even by optimizing on the value of s , w and u we can not do better than this with our method, since $\alpha_{\text{Dumer86}}(R) \sim R$ as R approaches 0. However, as can be guessed from the fact that $\alpha_{\text{Dumer86}} \leq \frac{R}{2}$ for $R \gtrsim 0.02$, the behavior of the complexity is much better for our RLPN decoder. This can be verified in Figure 4.2.

It is worthwhile to recall that ISD algorithms in the regime of the rate *close to 1* precisely use this collision method to find low weight codewords in order to reduce significantly the complexity of decoding. In a sense, we have a dual version of the birthday/collision decoder of [Dum86] with reduced complexity for rates close to 0.

4.2 Improving [Dum86] by puncturing as in [Dum89]

There is a simple way of improving the generation of dual codewords of low weight on \mathcal{N} . It consists in partitioning \mathcal{N} in two sets \mathcal{N}_1 and \mathcal{N}_2 with \mathcal{N}_2 being a subset of positions of size just a little bit above $n - k$ (which is the dimension of the dual code \mathcal{C}^\perp), say $n - k + \ell$ and then to use the collision method to get dual codewords of weight w_2 on \mathcal{N}_2 . The same method is used in the improvement [Dum89] of the simple collision decoder [Dum86] or in a slightly less efficient way in [Ste88]. It just consists in finding codewords in \mathcal{C}^\perp which have weight w_1 on \mathcal{N}_1 and w_2 on \mathcal{N}_2 instead of simply weight w on \mathcal{N} . We have represented in Figure 4.3 the form of the parity-checks we produce with this method. Note that the weight w_1 is expected to be half the size $k - \ell - s$ of \mathcal{N}_1 .

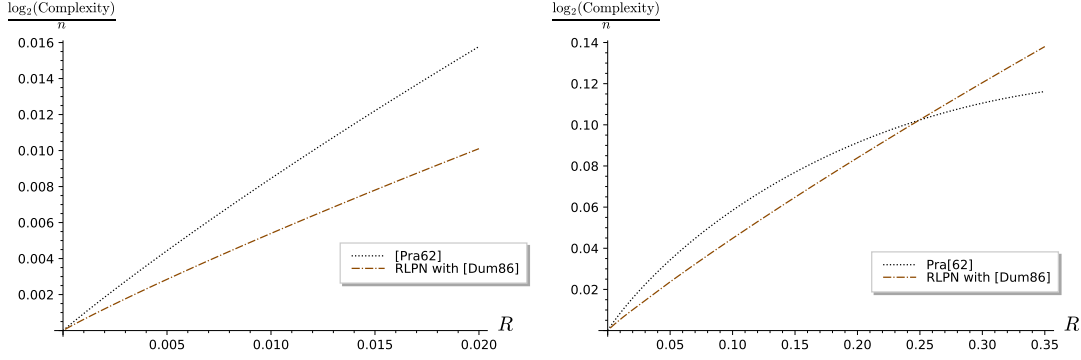


Figure 4.2. The complexity of the RLPN-decoder for very small rates *vs.* the simplest information set decoder, namely the ISD Prange decoder [Pra62]. For small R , there is no much difference between the ISD Prange decoder and much more evolved decoders like [BJMM12, MO15, BM17]. The RLPN-decoder with the very simple [Dum86] technique performs much better for small rates than ISD decoders. It is only outperformed by the Prange decoder for rates above 0.25 approximately.

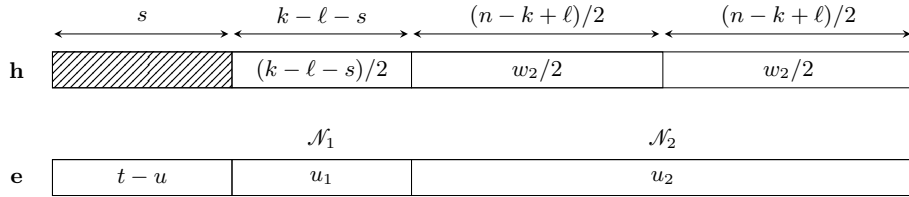


Figure 4.3. The form of the parity-checks produced by this method, *vs.* the bet made on the error. The hatched rectangle of size s for \mathbf{h} indicates that the weight is arbitrary on this part.

We have however to be particularly careful here in filtering parity-checks with the right-weight w_1 on \mathcal{N}_1 . Assume indeed that the weight of the error \mathbf{e} we try to recover is u_1 on \mathcal{N}_1 . If $w_1 = \frac{k-l-s}{2}$ and u_1 is odd, it turns out that the bias of such parity-checks is 0 (this is a consequence of Lemma 4.2 below). To understand the bias we get in this case, the proof of Proposition 3.8 can be readily adapted to yield

Proposition 4.1. *Assume that the code \mathcal{C} is chosen by picking for it an $(n-k) \times n$ binary parity-check matrix uniformly at random. Let \mathcal{N} be a fixed set of $n-s$ positions in $\llbracket 1, n \rrbracket$ which is partitioned in two sets \mathcal{N}_1 and \mathcal{N}_2 and \mathbf{e} be some error of weight u_i on \mathcal{N}_i for $i \in \{1, 2\}$. For $i \in \{1, 2\}$, choose \mathbf{h} uniformly at random among the parity-checks of \mathcal{C} of weight w_i on the \mathcal{N}_i 's and \mathbf{h}' uniformly at random among the words of weight w_i on the \mathcal{N}_i 's. For $i \in \{1, 2\}$, let*

$$\begin{aligned} \delta_i &\triangleq \text{bias}(\langle \mathbf{e}_{\mathcal{N}_i}, \mathbf{h}'_{\mathcal{N}_i} \rangle) \\ \delta &\triangleq \delta_1 \delta_2 \end{aligned}$$

If the parameters k, s, u_i, w_i are chosen as functions on n so that for n going to infinity, the expected number N of parity-checks of \mathcal{C} of respective weight w_i on \mathcal{N}_i for $i \in \{1, 2\}$, satisfies $N = \omega(1/\delta^2)$ then for all but a proportion $o(1)$ of codes we have

$$\text{bias}(\langle \mathbf{e}, \mathbf{h} \rangle) = (1 + o(1))\delta.$$

With the collision method we use, the parity-checks we produce have actually a slightly more specific form, since \mathcal{N}_2 is partitioned in two sets of (almost) the same size on which \mathbf{h} has weight $w_2/2$. It is not difficult to turn such a generation of parity-checks at the cost of a polynomial overhead into a generation of uniformly distributed parity-checks of weight w_2 on \mathcal{N}_2 . We leave out the details for doing this here. Under such an assumption, we have

Lemma 4.2. *With the same assumptions as in Proposition 4.1,*

$$\mathbb{P}_{\mathbf{h}}(\langle \mathbf{e}_{\mathcal{N}}, \mathbf{h}_{\mathcal{N}} = 1 \rangle) = \frac{1 - \varepsilon}{2} \quad \text{where} \quad \varepsilon = \delta_1 \cdot \delta_2 (1 - o(1))$$

$$\text{with } \delta_1 \triangleq \frac{K_{w_1}^{k-\ell-s}(u_1)}{\binom{k-\ell-s}{w_1}}, \quad \delta_2 \triangleq \frac{K_{w_2}^{n-k+\ell}(u_2)}{\binom{n-k+\ell}{w_2}}, \quad u_1 \triangleq |\mathbf{e}_{\mathcal{N}_1}|, \quad u_2 \triangleq |\mathbf{e}_{\mathcal{N}_2}|, \quad w_1 \triangleq |\mathbf{h}_{\mathcal{N}_1}|$$

$$\text{and } w_2 \triangleq |\mathbf{h}_{\mathcal{N}_2}|.$$

Proof. This is a straightforward application of the previous proposition and Lemma 3.3. \square

All these considerations lead to a slight variation of the RLPN decoder given in Algorithm 3.1. Let us make now a bet on the weight u_i of the error restricted to \mathcal{N}_i for $i \in \{1, 2\}$ and use Dumer's [Dum89] collision low-weight codeword generator to produce N parity-checks \mathbf{h} such that $|\mathbf{h}_{\mathcal{N}_i}| = w_i$ for $i \in \{1, 2\}$. We call the associated function $\text{CREATE}(N, w_1, w_2, \mathcal{P})$.

Proposition 4.3. *We assume that $\text{CREATE}(N, w_1, w_2, \mathcal{P})$ produces N parity-check equations in space S_{eq} and time T_{eq} that are of weight w_i on \mathcal{N}_i for $i \in \{1, 2\}$. The probability P_{succ} (over the choice of \mathcal{N}_1 and \mathcal{N}_2) that there are exactly u_1 errors in \mathcal{N}_1 and u_2 errors in \mathcal{N}_2 is given by*

$$P_{succ} = \frac{\binom{s}{t-u_1-u_2} \binom{k-\ell-s}{u_1} \binom{n-k+\ell}{u_2}}{\binom{n}{t}}.$$

The space complexity $S_{Dumer89}$ and time complexity $T_{Dumer89}$ of the RLPN-decoder are given by

$$\text{Space: } S = O(S_{eq} + 2^s), \quad \text{Time: } T = \tilde{O}\left(\frac{T_{eq} + 2^s}{P_{succ}}\right).$$

under the constraint on the parameters s, u_1, u_2, w_1 and w_2 given by

$$N = \tilde{\Omega}\left(\left(\frac{\binom{k-\ell-s}{w_1} \binom{n-k+\ell}{w_2}}{K_{w_1}^{k-\ell-s}(u_1) K_{w_2}^{k-\ell-s}(u_2)}\right)^2\right).$$

From this proposition we readily obtain the following asymptotic form

Proposition 4.4. *The asymptotic complexity exponent of the RLPN decoder based on Dumer's collision low weight dual codeword generators is given by*

$$\alpha_{Dumer89}(R) \triangleq \min_{(\sigma, \nu_1, \nu_2, \lambda, \omega_1, \omega_2) \in \mathcal{R}} \beta(R, \sigma, \nu_1, \nu_2, \lambda, \omega_1, \omega_2) \quad (4.3)$$

where

$$\begin{aligned} \beta(R, \sigma, \nu_1, \nu_2, \lambda, \omega_1, \omega_2) &\triangleq \max(\sigma, \nu(R, \lambda, \omega_2)) + \pi(R, \sigma, \nu_1, \nu_2, \lambda), \\ \nu(R, \lambda, \omega_2) &\triangleq \max\left(\frac{(1-R+\lambda)}{2} h\left(\frac{\omega_2}{1-R+\lambda}\right), (1-R+\lambda) h\left(\frac{\omega_2}{1-R+\lambda}\right) - \lambda\right), \\ \pi(R, \sigma, \nu_1, \nu_2, \lambda) &\triangleq 1-R-\sigma h\left(\frac{\tau-\nu_1-\nu_2}{\sigma}\right) - (R-\lambda-\sigma) h\left(\frac{\nu_1}{R-\lambda-\sigma}\right) - (1-R+\lambda) h\left(\frac{\nu_2}{1-R+\lambda}\right), \\ \tau &\triangleq \delta_{GV}(R) = h^{-1}(1-R) \end{aligned}$$

and the constraint region \mathcal{R} is defined by the subregion of nonnegative tuples $(\sigma, \nu_1, \nu_2, \lambda, \omega_1, \omega_2)$ such that

$$\sigma \leq R - \lambda, \quad \nu_1 \leq R - \lambda - \sigma, \quad \nu_2 \leq 1 - R + \lambda, \quad \tau - \sigma \leq \nu_1 + \nu_2 \leq \tau, \quad \omega_1 = \frac{R - \lambda - \sigma}{2}$$

and

$$\nu(R, \lambda, \omega_2) = -(R - \lambda - \sigma) \tilde{\varepsilon}\left(\frac{\nu_1}{R - \lambda - \sigma}, \frac{\omega_1}{R - \lambda - \sigma}\right) - (1 - R + \lambda) \tilde{\varepsilon}\left(\frac{\nu_2}{1 - R + \lambda}, \frac{\omega_2}{1 - R + \lambda}\right)$$

where $\tilde{\varepsilon}$ is the function defined in Proposition 3.8.

5 Using advanced collision techniques

ISD techniques have evolved [Ste88, Dum89, BLP11, MMT11, BJMM12] by first introducing [Ste88] collision/subset-sum techniques whose purpose is to produce for codes of rate close to 1, all codewords of some small weight, and later on by substantially improving them by using on top of that for instance representation techniques [MMT11]. These algorithms come very handy in our case for devising the function $\text{CREATE}(N, w, \mathcal{P})$ that we need. In the previous section, we have explored what could be achieved by the very first techniques of this type taken from [Dum86, Dum89]. We are going to explain now what can be gained by using [MMT11, BJMM12]. It is convenient here to formalize the basic step used in the previous section which can be explained by the following function

Input: $\mathcal{L}_1 \subseteq \mathbb{F}_2^n$, $\mathcal{L}_2 \subseteq \mathbb{F}_2^n$, $w \in \llbracket 1, n \rrbracket$, $\mathbf{H} \in \mathbb{F}_2^{\ell \times n}$
Output: a list $\mathcal{L} = \{\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2 : \mathbf{x}_i \in \mathcal{L}_i, i \in \{1, 2\}, |\mathbf{x}| = w, \mathbf{H}\mathbf{x}^\top = \mathbf{0}\}$ of elements of the form $\mathbf{x}_1 + \mathbf{x}_2$ with \mathbf{x}_i belonging to \mathcal{L}_i of weight w belonging to the code of parity-check matrix \mathbf{H}

function MERGE($\mathcal{L}_1, \mathcal{L}_2, w, \mathbf{H}$)

```

 $\mathcal{L} \leftarrow \emptyset$ 
for all  $\mathbf{x}_1 \in \mathcal{L}_1$  do
  | Store  $\mathbf{x}_1$  in a hashtable  $\mathcal{T}$  at address  $\mathbf{H}\mathbf{x}_1^\top$ 
end for
for all  $\mathbf{x}_2 \in \mathcal{L}_2$  do
  | if  $\exists \mathbf{x}_1$  in  $\mathcal{T}$  at address  $\mathbf{H}\mathbf{x}_2^\top$  and  $|\mathbf{x}_1 + \mathbf{x}_2| = w$  then
  | | Put  $\mathbf{x}_1 + \mathbf{x}_2$  in  $\mathcal{L}$ 
  | end if
end for
return  $\mathcal{L}$ 
end function

```

It creates codewords of weight w in a code of parity-check matrix \mathbf{H} as sums $\mathbf{x}_1 + \mathbf{x}_2$ of two lists \mathcal{L}_1 and \mathcal{L}_2 with a complexity which is of the form $O\left(\max\left(\#\mathcal{L}_1, \#\mathcal{L}_2, \frac{\#\mathcal{L}_1 \cdot \#\mathcal{L}_2}{2^i}\right)\right)$ if the $\mathbf{H}\mathbf{x}_i^\top$'s are independently distributed uniformly at randomise (we will make this assumption from now on). It is clear that [Dum86] and [Dum89] is more or less a direct application of this method. [MMT11] and [BJMM12] use several layers of this function. [MMT11] starts by partitioning the set of positions of the vectors of \mathbb{F}_2^n which are considered in two sets \mathcal{S}_1 and \mathcal{S}_2 of about the same size. Then it starts with two lists \mathcal{L}_1^0 and \mathcal{L}_2^0 of all elements of weight p_0 and support \mathcal{S}_1 and \mathcal{S}_2 respectively. It merges them in a list \mathcal{L}^1 of elements of weight p_1 in the kernel of a parity-check matrix \mathbf{H}_1 . Since the elements of \mathcal{L}_1^0 and \mathcal{L}_2^0 have disjoint supports by construction, we necessarily have that $p_1 = 2p_0$. \mathcal{L}^1 is then merged with itself to yield elements which are in the kernel of another matrix \mathbf{H}_2 (see Figure 5.1). Since these are sums of elements of \mathcal{L}^1 they are also in the kernel of \mathbf{H}_1 , so that that the elements of the final list are of weight p_2 and belong to the code of parity-check $\mathbf{H} = \begin{pmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \end{pmatrix}$. The size of \mathbf{H}_1 is chosen such that an element \mathbf{x} of weight p_2 and $\mathbf{H}_1\mathbf{x}^\top = \mathbf{0}$ is typically the sum of only two elements of \mathcal{L}^1 (this is the point of the representation technique). [BJMM12] is similar to [MMT11] with one layer which is added. In this case, we create at the end a list of elements of weight p_3 which are in the code of parity-check matrix

$$\mathbf{H} \triangleq \begin{pmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \mathbf{H}_3 \end{pmatrix} \quad (5.1)$$

(see Figure 5.1). The sizes of \mathbf{H}_1 in the [MMT11] case, and of \mathbf{H}_1 and \mathbf{H}_2 in [BJMM12] are chosen to ensure unicity of the representation of an element of a list as the sum of two elements of the lists used for the merge (this is the representation technique).

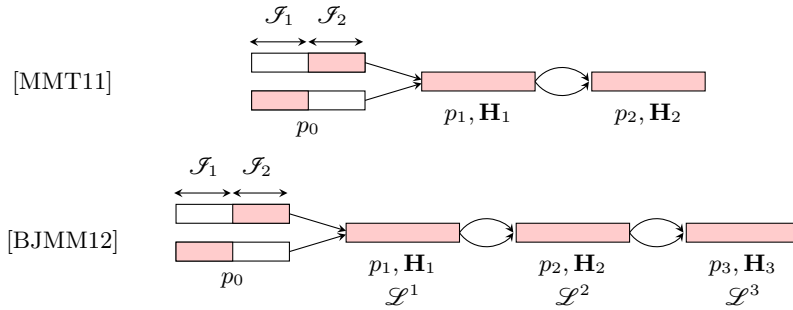


Figure 5.1. This figure represents the successive lists obtained in [MMT11] and [BJMM12]. The support of the elements of the list are represented in pink. Arrows point from the lists which are merged to the result of the merge and if two arrows depart from a list and arrive at another list, this means that the departure list is merged with itself. The weights of the elements are indicated for each level and the matrix \mathbf{H}_i used for the merge is also given at the level of the result of the merge.

We use these two techniques as we used the [Dum86] technique inside the [Dum89] technique, namely to generate codewords of \mathcal{C}^\perp (*i.e.* $\mathbf{H}\mathbf{x}^\top = \mathbf{0}$ for \mathbf{H} given by (5.1)) which are of weight p_3 on a set of indices of size $n - k + \ell$ (see Figure 5.2).

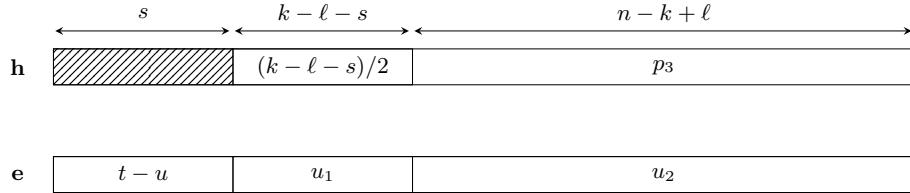


Figure 5.2. The form of the parity-checks produced by this method [BJMM12], *vs.* the bet made on the error. The hatched rectangle of size s for \mathbf{h} indicates that the weight is arbitrary on this part.

If we let ℓ_1 be the number of rows of \mathbf{H}_1 , ℓ_2 be the number of rows of the matrix of $\mathbf{H}_2 \triangleq \begin{pmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \end{pmatrix}$, then the fact that the elements of \mathcal{L}^2 should have a unique representation in terms of a sum of a pair of elements of \mathcal{L}^1 respectively and that they are all elements \mathbf{x} of weight p_1 and p_2 respectively which satisfy $\mathbf{H}_2'\mathbf{x}^\top = \mathbf{0}$ and $\mathbf{H}\mathbf{x}^\top = \mathbf{0}$ respectively, imposes conditions (5.2) which follow. The S_i represent the space complexity of the successive lists (*i.e.* \mathcal{L}^0 , \mathcal{L}^1 , \mathcal{L}^2 and \mathcal{L}^3) used in the [BJMM12] algorithm, whereas the T_i 's denote the complexity of each merge and T_{eq} is the final complexity.

$$2^{\ell_1} = \binom{p_2}{p_2/2} \binom{n - k + \ell - p_2}{p_1 - p_2/2}, \quad 2^{\ell_2} = \binom{p_3}{p_3/2} \binom{n - k + \ell - p_3}{p_2 - p_3/2} \quad (5.2)$$

$$S_0 = \binom{\frac{n-k+\ell}{2}}{\frac{p_1}{2}}, \quad S_1 = \frac{\binom{n-k+\ell}{p_1}}{2^{\ell_1}}, \quad S_2 = \frac{\binom{n-k+\ell}{p_2}}{2^{\ell_2}}, \quad S_3 = \frac{\binom{n-k+\ell}{p_3}}{2^\ell} \quad (5.3)$$

$$T_0 = S_0, \quad T_1 = S_0 + \frac{S_0^2}{2^{\ell_1}}, \quad T_2 = S_1 + \frac{S_1^2}{2^{\ell_2 - \ell_1}}, \quad T_3 = S_2 + \frac{S_2^2}{2^{\ell - \ell_2}}, \quad T_{\text{eq}} = T_0 + T_1 + T_2 + T_3 \quad (5.4)$$

There is a similar proposition as Proposition 4.4 which gives the asymptotic complexity of the RLPN decoder used in conjunction with the [MMT11] or [BJMM12] techniques for producing low weight codewords. For [MMT11] it is given by

Proposition 5.1. *The asymptotic complexity exponent of the RLPN decoder based on [MMT11] is given by*

$$\alpha_{MMT}(R) \triangleq \min_{(\sigma, \nu_1, \nu_2, \lambda, \lambda_1, \omega_1, \omega_2, \pi_1) \in \mathcal{R}} \beta(R, \sigma, \nu_1, \nu_2, \lambda, \lambda_1, \omega_1, \omega_2, \pi_1) \quad (5.5)$$

where

$$\begin{aligned} \beta &\triangleq \max(\sigma, \nu) + \pi, \\ \nu &\triangleq \max(\gamma_1, \gamma_2), \\ \gamma_1 &\triangleq \max\left(\frac{1-R+\lambda}{2}h\left(\frac{\pi_1}{1-R+\lambda}\right), (1-R+\lambda)h\left(\frac{\pi_1}{1-R+\lambda}\right) - \lambda_1\right), \\ \gamma_2 &\triangleq 2(1-R+\lambda)h\left(\frac{\pi_1}{1-R+\lambda}\right) - \lambda_1 - \lambda, \\ \rho &\triangleq 1-R-\sigma h\left(\frac{\tau-\nu_1-\nu_2}{\sigma}\right) - (R-\lambda-\sigma)h\left(\frac{\nu_1}{R-\lambda-\sigma}\right) - (1-R+\lambda)h\left(\frac{\nu_2}{1-R+\lambda}\right), \\ \tau &\triangleq \delta_{GV}(R) = h^{-1}(1-R) \end{aligned}$$

and the constraint region \mathcal{R} is defined by the subregion of nonnegative tuples $(\sigma, \nu_1, \nu_2, \lambda, \lambda_1, \pi, \omega_1, \omega_2)$ such that

$$\sigma \leq R - \lambda, \quad \lambda_1 \leq \lambda, \quad \pi_1 \leq \omega_2, \quad \nu_1 \leq R - \lambda - \sigma, \quad \nu_2 \leq 1 - R + \lambda, \quad \tau - \sigma \leq \nu_1 + \nu_2 \leq \tau,$$

$$\omega_1 = \frac{R - \lambda - \sigma}{2}, \quad \omega_2 < 1 - R + \lambda,$$

$$\frac{\omega_2}{2} < \pi_1 < 1 - R + \lambda, \quad \lambda_1 = \omega_2 + (1 - R + \lambda - \omega_2)h\left(\frac{\pi_1 - \omega_2/2}{1 - R + \lambda - \omega_2}\right),$$

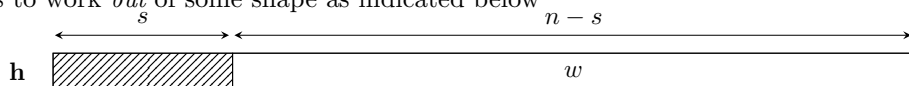
$$\begin{aligned} (1-R+\lambda)h\left(\frac{\omega_2}{1-R+\lambda}\right) - \lambda &= -(R-\lambda-\sigma)\tilde{\varepsilon}\left(\frac{\nu_1}{R-\lambda-\sigma}, \frac{\omega_1}{R-\lambda-\sigma}\right) \\ &\quad - (1-R+\lambda)\tilde{\varepsilon}\left(\frac{\nu_2}{1-R+\lambda}, \frac{\omega_2}{1-R+\lambda}\right), \end{aligned}$$

where $\tilde{\varepsilon}$ is the function defined in Proposition 3.8.

We give a proof in the Appendix A for reference. There is a similar proposition for the asymptotic behavior of RLPN decoding used together with [BJMM12] which is given in the Appendix A. We have used them for producing the complexity curves given in Figure 6.1 which display the various complexities of the RLPN decoders we have presented. Even if there is a tiny improvement by using [BJMM12] instead of [MMT11] the two curves are nearly indistinguishable. A perspective of improvement of our algorithm could be to produce the parity-check equations by using more recent ISD techniques than [BJMM12], in particular [MO15] or [BM17] which both use nearest-neighbor search. Our preliminary results using in particular [MO15] do not provide significant improvement, we have only been able to achieve a very slightly better complexity for rates close to 0.2.

6 A Lower bound on the complexity of RLPN decoders

As pointed out all along the paper, RLPN decoding needs a large number N of parity-check equations to work *but* of some shape as indicated below



where the hatched area indicates that the weight is arbitrary on this part while \mathbf{h} restricted on the other positions needs to have Hamming weight w . The number N of such parity-checks has to verify (see Proposition 3.15)

$$N = \tilde{\Omega} \left(\left(\frac{\binom{n-s}{w}}{K_w^{n-s}(u)} \right)^2 \right) \quad (6.1)$$

in order to be able to solve the underlying LPN problem. It can be verified that the smaller w is (the bigger is the bias ε), the smaller is N and the more efficient is our algorithm. Obviously if w is too small, there are not enough such parity-checks. It can be verified that the expected number of parity-checks of the aforementioned shape is given by $2^s \binom{n-s}{w} / 2^k$ in a random code (which is our assumption). Therefore we need

$$N = O \left(\frac{2^s \binom{n-s}{w}}{2^k} \right). \quad (6.2)$$

Given this picture it is readily seen that the complexity of RLPN decoding is always lower-bounded by N (which is at least the cost to produce N parity-checks) but we can be more accurate on the lower-bound over the complexity. Recall that we first need to solve an underlying LPN problem and that we make a bet on the number of errors u in \mathcal{N} . Therefore, *assuming* that we can compute a parity-check of the aforementioned shape in time $O(1)$, the complexity of our extended statistical decoding is given by

$$\tilde{O} \left(\frac{1}{P_{\text{succ}}} \max(2^s, N) \right) \quad (6.3)$$

where P_{succ} is given in Proposition 3.15. Our only constraints are given by (6.1) and (6.2). By optimizing (6.3) over s, u and w , we can give a lower-bound on the complexity of RLPN decoding. However notice that our lower-bound applies to a partition of parity-checks in two parts (s and $n-s$). We do not consider here finer partitions. This method for lower bounding the complexity of RLPN decoding is very similar to the technique used in [DT17a, §7] to lower bound the complexity of statistical decoding. All in all, we give in Figure 6.1 this lower-bound of the complexity. The optimal parameters computed for each RLPN algorithms can be found on <https://github.com/tillich/RLPNdecoding>. As we see our RLPN decoders meet this lower-bound for small rates and we can hope to outperform significantly ISD's for code rates smaller than ≈ 0.45 .

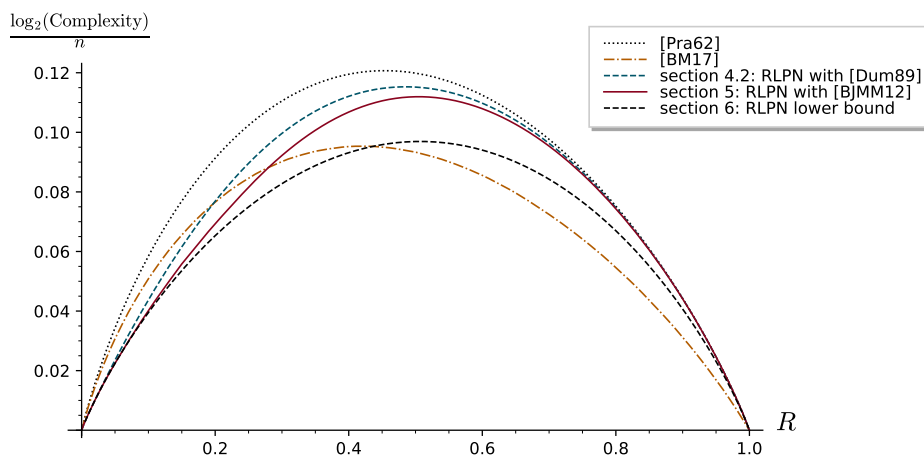


Figure 6.1. Complexity exponents of our different RLPN decoders, ISD's and the genie-aided RLPN algorithm when splitting parity-checks in two parts.

7 Concluding remarks

Since Prange’s seminal work [Pra62] in 1962, ISD algorithms have played a predominant role for assessing the complexity of code-based cryptographic primitives. In the fixed rate regime, they have been beaten only once in [Dum86] with the help of collision techniques, and this only for a tiny code rate range ($R \in (0.98, 1)$) and for a short period of time [Ste88, Dum89] until these collision techniques were merged with the collision techniques to yield modern ISD’s. Surprisingly enough, these improved ISD have resulted in decoding complexity curves tilting more and more to the left (*i.e.* with a maximum which is attained more and more below $\frac{1}{2}$) instead of being symmetric around $\frac{1}{2}$ as it could have been expected. It is precisely for rates below $\frac{1}{2}$ that RLPN decoding is able to outperform the best ISD’s. This seems to point to the fact that it is precisely for this regime of parameters that we should aim for improving them. Interestingly enough, even if there is some room of improvement for RLPN decoding by using better strategies for producing the needed low weight parity-checks, there is a ceiling that this technique can not break (at least if we just split the parity-checks in two parts) and which is extremely close at rate $R = 0.45$ to the best ISD algorithm [BM17]. The RLPN decoding algorithm presented here has not succeeded in changing the landscape for very tiny code rates (R going to 0), since the complexity exponent of RLPN decoding approaches the one of exhaustive search on codewords, but the speed at which this complexity approaches exhaustive search is much smaller than for ISD’s in the full decoding regime (*i.e.* at the GV distance). The success of RLPN decoding for $R < 0.3$ could be traced back precisely to this behavior close to 0. An interesting venue for research could be to try to explore if there are other decoding strategies that would be candidate for beating exhaustive search in the tiny code rate regime.

Acknowledgement

The work of TDA was funded by the French Agence Nationale de la Recherche through ANR JCJC COLA (ANR-21-CE39-0011). We wish also to thank Ilya Dumer for his very insightful thoughts about decoding linear codes in the low rate regime.

References

- [AGS11] Carlos Aguilar, Philippe Gaborit, and Julien Schrek. A new zero-knowledge code based identification scheme with reduced communication. In *Proc. IEEE Inf. Theory Workshop- ITW 2011*, pages 648–652. IEEE, October 2011.
- [Ari09] Erdal Arıkan. Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Trans. Inform. Theory*, 55(7):3051–3073, 2009.
- [Bar97] Alexander Barg. Complexity issues in coding theory. *Electronic Colloquium on Computational Complexity*, October 1997.
- [BJMM12] Anja Becker, Antoine Joux, Alexander May, and Alexander Meurer. Decoding random binary linear codes in $2^{n/20}$: How $1 + 1 = 0$ improves information set decoding. In *Advances in Cryptology - EUROCRYPT 2012*, LNCS. Springer, 2012.
- [BLP11] Daniel J. Bernstein, Tanja Lange, and Christiane Peters. Smaller decoding exponents: ball-collision decoding. In *Advances in Cryptology - CRYPTO 2011*, volume 6841 of LNCS, pages 743–760, 2011.
- [BM17] Leif Both and Alexander May. Optimizing BJMM with Nearest Neighbors: Full Decoding in $2^{2/21n}$ and McEliece Security. In *WCC Workshop on Coding and Cryptography*, September 2017.
- [BM18] Leif Both and Alexander May. Decoding linear codes with high error rate and its impact for LPN security. In Tanja Lange and Rainer Steinwandt, editors, *Post-Quantum Cryptography 2018*, volume 10786 of LNCS, pages 25–46, Fort Lauderdale, FL, USA, April 2018. Springer.
- [CS16] Rodolfo Canto-Torres and Nicolas Sendrier. Analysis of information set decoding for a sub-linear error weight. In *Post-Quantum Cryptography 2016*, LNCS, pages 144–161, Fukuoka, Japan, February 2016.

- [DT17a] Thomas Debris-Alazard and Jean-Pierre Tillich. Statistical decoding. preprint, January 2017. arXiv:1701.07416.
- [DT17b] Thomas Debris-Alazard and Jean-Pierre Tillich. Statistical decoding. In *Proc. IEEE Int. Symposium Inf. Theory - ISIT 2017*, pages 1798–1802, Aachen, Germany, June 2017.
- [Dum86] Ilya Dumer. On syndrome decoding of linear codes. In *Proceedings of the 9th All-Union Symp. on Redundancy in Information Systems, abstracts of papers (in russian), Part 2*, pages 157–159, Leningrad, 1986.
- [Dum89] Ilya Dumer. Two decoding algorithms for linear codes. *Probl. Inf. Transm.*, 25(1):17–23, 1989.
- [Dum91] Ilya Dumer. On minimum distance decoding of linear codes. In *Proc. 5th Joint Soviet-Swedish Int. Workshop Inform. Theory*, pages 50–52, Moscow, 1991.
- [FJR21] Thibault Feneuil, Antoine Joux, and Matthieu Rivain. Shared permutation for syndrome decoding: New zero-knowledge protocol and code-based signature. *IACR Cryptol. ePrint Arch.*, page 1576, 2021.
- [FKI07] Marc P. C. Fossorier, Kazukuni Kobara, and Hideki Imai. Modeling bit flipping decoding based on nonorthogonal check sums with application to iterative decoding attack of McEliece cryptosystem. *IEEE Trans. Inform. Theory*, 53(1):402–411, 2007.
- [FS09] Matthieu Finiasz and Nicolas Sendrier. Security bounds for the design of code-based cryptosystems. In M. Matsui, editor, *Advances in Cryptology - ASIACRYPT 2009*, volume 5912 of *LNCS*, pages 88–105. Springer, 2009.
- [Gal63] Robert G. Gallager. *Low Density Parity Check Codes*. M.I.T. Press, Cambridge, Massachusetts, 1963.
- [GG07] Philippe Gaborit and Marc Girault. Lightweight code-based authentication and signature. In *Proc. IEEE Int. Symposium Inf. Theory - ISIT*, pages 191–195, Nice, France, June 2007.
- [IS98] Mourad E.H. Ismail and Plamen Simeonov. Strong asymptotics for Krawtchouk polynomials. *Journal of Computational and Applied Mathematics*, pages 121–144, 1998.
- [Jab01] Abdulrahman Al Jabri. A statistical decoding algorithm for general linear block codes. In Bahram Honary, editor, *Cryptography and coding. Proceedings of the 8th IMA International Conference*, volume 2260 of *LNCS*, pages 1–8, Cirencester, UK, December 2001. Springer.
- [Kha02] Aamod Khandekar. *Graph-based Codes and Iterative Decoding*. PhD thesis, California Institute of Technology, Pasadena, California, 2002.
- [KS21] Naomi Kirshner and Alex Samorodnitsky. A moment ratio bound for polynomials and some extremal properties of krawchouk polynomials and hamming spheres. *IEEE Trans. Inform. Theory*, 67(6):3509–3541, 2021.
- [Lev95] Vladimir I. Levenshtein. Krawtchouk polynomials and universal bounds for codes and designs in hamming spaces. *IEEE Trans. Inf. Theory*, 41(5):1303–1321, 1995.
- [McE78] Robert J. McEliece. *A Public-Key System Based on Algebraic Coding Theory*, pages 114–116. Jet Propulsion Lab, 1978. DSN Progress Report 44.
- [MMT11] Alexander May, Alexander Meurer, and Enrico Thomae. Decoding random linear codes in $O(2^{0.054n})$. In Dong Hoon Lee and Xiaoyun Wang, editors, *Advances in Cryptology - ASIACRYPT 2011*, volume 7073 of *LNCS*, pages 107–124. Springer, 2011.
- [MO15] Alexander May and Ilya Ozerov. On computing nearest neighbors with applications to decoding of binary linear codes. In E. Oswald and M. Fischlin, editors, *Advances in Cryptology - EUROCRYPT 2015*, volume 9056 of *LNCS*, pages 203–228. Springer, 2015.
- [MTSB13] Rafael Misoczki, Jean-Pierre Tillich, Nicolas Sendrier, and Paulo S. L. M. Barreto. MDPC-McEliece: New McEliece variants from moderate density parity-check codes. In *Proc. IEEE Int. Symposium Inf. Theory - ISIT*, pages 2069–2073, 2013.
- [Ove06] Raphael Overbeck. Statistical decoding revisited. In Reihaneh Safavi-Naini Lynn Batten, editor, *Information security and privacy : 11th Australasian conference, ACISP 2006*, volume 4058 of *LNCS*, pages 283–294. Springer, 2006.
- [Pra62] Eugene Prange. The use of information sets in decoding cyclic codes. *IRE Transactions on Information Theory*, 8(5):5–9, 1962.
- [RU08] Tom Richardson and Ruediger Urbanke. *Modern Coding Theory*. Cambridge University Press, 2008.
- [Ste88] Jacques Stern. A method for finding codewords of small weight. In G. D. Cohen and J. Wolfmann, editors, *Coding Theory and Applications*, volume 388 of *LNCS*, pages 106–113. Springer, 1988.
- [Ste93] Jacques Stern. A new identification scheme based on syndrome decoding. In D.R. Stinson, editor, *Advances in Cryptology - CRYPTO'93*, volume 773 of *LNCS*, pages 13–21. Springer, 1993.

- [Wag02] David Wagner. A generalized birthday problem. In Moti Yung, editor, *Advances in Cryptology - CRYPTO 2002*, volume 2442 of *LNCS*, pages 288–303. Springer, 2002.

Appendices

A Asymptotic complexity of the RLPN decoder used in conjunction with the [MMT11] and [BJMM12] technique

Proof of Proposition 5.1. In the case of the MMT technique we have the following set of equalities

$$2^{\ell_1} = \binom{w_2}{w_2/2} \binom{n-k+\ell-w_2}{p_1-w_2/2} \quad (\text{A.1})$$

$$S_0 = \binom{\frac{n-k+\ell}{2}}{\frac{p}{2}}, \quad S_1 = \frac{\binom{n-k+\ell}{p}}{2^{\ell_1}}, \quad S_2 = \frac{\binom{n-k+\ell}{w_2}}{2^\ell} \quad (\text{A.2})$$

$$T_0 = S_0, \quad T_1 = S_0 + \frac{S_0^2}{2^{\ell_1}}, \quad T_2 = S_1 + \frac{S_1^2}{2^{\ell-\ell_1}}, \quad T_{\text{eq}} = T_0 + T_1 + T_2 \quad (\text{A.3})$$

We let

$$\sigma \triangleq \frac{s}{n}, \quad \lambda \triangleq \frac{\ell}{n}, \quad \lambda_1 \triangleq \frac{\ell_1}{n}, \quad \pi \triangleq \frac{p}{n}, \quad \omega_2 \triangleq \frac{w_2}{n}, \quad \nu_1 \triangleq \frac{v_1}{n}, \quad \nu_2 \triangleq \frac{v_2}{n}.$$

The equality

$$2^{\ell_1} = \binom{w_2}{w_2/2} \binom{n-k+\ell-w_2}{p_1-w_2/2}$$

translates into

$$\lambda_1 = \omega_2 + (1-R+\lambda-\omega_2)h\left(\frac{\pi-\omega_2/2}{1-R+\lambda-\omega_2}\right).$$

In our case, the number N of parity-check which are produced is equal to S_2 and therefore the condition

$$N = \tilde{\Omega} \left(\left(\frac{\binom{k-\ell-s}{w_1} \binom{n-k+\ell}{w_2}}{K_{w_1}^{k-\ell-s}(u_1)} \right)^2 \right).$$

of Proposition 3.15 becomes

$$(1-R+\lambda)h\left(\frac{\omega_2}{1-R+\lambda}\right) - \lambda = -(R-\lambda-\sigma)\tilde{\varepsilon}\left(\frac{\nu_1}{R-\lambda-\sigma}, \frac{\omega_1}{R-\lambda-\sigma}\right) \\ - (1-R+\lambda)\tilde{\varepsilon}\left(\frac{\nu_2}{1-R+\lambda}, \frac{\omega_2}{1-R+\lambda}\right). \quad \square$$

Proposition A.1. *The asymptotic complexity exponent of the RLPN decoder based on [BJMM12] is given by*

$$\alpha_{BJMM}(R) = \min_{(\sigma, \nu_1, \nu_2, \lambda, \lambda_1, \lambda_2, \pi_1, \pi_2, \omega_1, \omega_2) \in \mathcal{R}} \beta(\sigma, \nu_1, \nu_2, \lambda, \lambda_1, \lambda_2, \pi_1, \pi_2, \omega_1, \omega_2) \quad (\text{A.4})$$

where

$$\beta \triangleq \min(\sigma, \nu + \rho) \quad \text{with}$$

$$\nu = \max(\gamma_1, \gamma_2, \gamma_3)$$

$$\gamma_1 = \max\left(\frac{1-R+\lambda}{2}h\left(\frac{\pi_1}{1-R+\lambda}\right), (1-R+\lambda)h\left(\frac{\pi_1}{1-R+\lambda}\right) - \lambda_1\right)$$

$$\gamma_2 = 2(1-R+\lambda)h\left(\frac{\pi_1}{1-R+\lambda}\right) - \lambda_1 - \lambda_2$$

$$\gamma_3 = 2(1-R+\lambda)h\left(\frac{\pi_2}{1-R+\lambda}\right) - \lambda - \lambda_2$$

$$\rho = 1-R-\sigma h\left(\frac{\tau-\nu_1-\nu_2}{\sigma}\right) - (R-\lambda-\sigma)h\left(\frac{\nu_1}{R-\lambda-\sigma}\right) - (1-R+\lambda)h\left(\frac{\nu_2}{1-R+\lambda}\right)$$

$$\tau \triangleq \delta_{GV}(R) = h^{-1}(1-R)$$

and the constraint region \mathcal{R} is defined by the sub-region of nonnegative tuples $(\sigma, \nu_1, \nu_2, \lambda, \lambda_1, \lambda_2, \pi_1, \pi_2, \omega_1, \omega_2)$ such that

$$\begin{aligned} \sigma &\leq R - \lambda, & \lambda_1 &\leq \lambda_2 \leq \lambda, & \pi_1 &\leq \pi_2 \leq \omega_2, & \nu_1 &\leq R - \lambda - \sigma, & \nu_2 &\leq 1 - R + \lambda, & \tau - \sigma &\leq \nu_1 + \nu_2 \leq \tau, \\ \pi_1 &\geq \frac{\pi_2}{2}, & \pi_2 &\geq \frac{\omega_2}{2}, & \pi_2 &< \lambda_1, & \omega_2 &< \lambda_2, & \pi_1 &< 1 - R + \lambda, & \omega_1 &= \frac{R - \lambda - \sigma}{2}, \\ \lambda_1 &= \pi_2 + (1 - R + \lambda - \pi_2)h \left(\frac{\pi_1 - \pi_2/2}{1 - R + \lambda - \pi_2} \right), \\ \lambda_2 &= \omega_2 + (1 - R + \lambda - \omega_2)h \left(\frac{\pi_2 - \omega_2/2}{1 - R + \lambda - \omega_2} \right) \end{aligned}$$

and

$$\begin{aligned} (1 - R + \lambda)h \left(\frac{\omega_2}{1 - R + \lambda} \right) - \lambda &= -(R - \lambda - \sigma)\tilde{\varepsilon} \left(\frac{\nu_1}{R - \lambda - \sigma}, \frac{\omega_1}{R - \lambda - \sigma} \right) \\ &\quad - (1 - R + \lambda)\tilde{\varepsilon} \left(\frac{\nu_2}{1 - R + \lambda}, \frac{\omega_2}{1 - R + \lambda} \right) \end{aligned}$$

where $\tilde{\varepsilon}$ is the function defined in Proposition 3.8.

B About [BM18]

We have compared all along the paper our results to the state-of-the-art for solving the decoding problem. To date, the best complexity is achieved in [BM17] which essentially uses the techniques in [MO15] with an additional stage.

There is another result in the literature which claims better results than [BM17]. Indeed, in [BM18], the same authors pretend that we can get a gain of $\approx 7\%$ for full decoding and at the worst code rate in comparison with [BM17]. However, this result is flawed: there is indeed an error in the analysis of this new decoding algorithm which leads to this result. The decoding algorithm in [BM18] essentially consists in producing lists recursively by using nearest-neighbor searches at each stage; the solution of the decoding problem is then contained in the last list. Thus, an accurate analysis of this algorithm lies in the good estimation of the size of the various lists. In [BM18, Section 4, p. 16], an upper bound of the size of these lists is given. Unfortunately this upper bound does not hold. Let us recall it. It is a bound of the list-size S_i at stage i :

$$\begin{aligned} S_i &\leq \left| \{ \mathbf{x} \in \mathbb{F}_2^k, |\mathbf{x}| = p_i \} \right| \cdot \mathbb{P}_{\mathbf{x} \in \mathbb{F}_2^i} \left(|\mathbf{x}| = \omega_i^{(i)} \right) \\ &\quad \cdot \prod_{h=1}^{i-1} \mathbb{P}_{\mathbf{x}, \mathbf{y} \in \mathbb{F}_2^h} \left(|\mathbf{x} + \mathbf{y}| = \omega_h^{(i)} : |\mathbf{x}| = |\mathbf{y}| = \omega_h^{(i-1)} \right) \end{aligned} \tag{B.1}$$

We use here the notation from [BM18]: at each stage $i \in \{1..m\}$, the algorithm is producing lists of S_i vectors with weight p_i on an information set and weight $\left(\omega_h^{(i)} \right)_{h \in \{1..i\}}$ on redundancy subsets of respective length $(\ell_i)_{h \in \{1..i\}}$. Since the list at the final stage m contains the solution, the decoding distance ω that we aim to achieve is $\omega = p_3 + \sum_{i=1}^m \omega_i$.

We have noticed that there is a problem with (B.1) by running simulations. For instance, let look at the following parameters for depth $m = 3$:

Decoding problem parameters:	$n = 160$	$k = 70$	$\omega = 18$
Size of the redundancy subsets:	$\ell_1 = 40$	$\ell_2 = 14$	$\ell_3 = 36$
Stage 1 parameters:	$p_1 = 4$	$\omega_1^{(1)} = 25$	
Stage 2 parameters:	$p_2 = 4$	$\omega_1^{(2)} = 30$	$\omega_2^{(2)} = 8$
Stage 3 parameters:	$p_3 = 6$	$\omega_1^{(3)} = 2$	$\omega_2^{(3)} = 4$ $\omega_3^{(3)} = 6$

The Equation (B.1) tells us that S_2 should not be bigger than 13.66 whereas in practice it is 33.12. The problem lies in the fact that in (B.1), the probability

$$\mathbf{P} \triangleq \mathbb{P} \left(|\mathbf{y}_i| = \omega_i^{(i)} \right) \cdot \prod_{h=1}^{i-1} \mathbb{P} \left(|\mathbf{y}'_h + \mathbf{y}''_h| = \omega_h^{(i)} \right) \quad (\text{B.2})$$

represents the probability that a pair from the lists at stage $i - 1$ produces an element in the list at stage i , but it is not the probability that a vector of length k and weight p_i is in that list as it is suggested by Equation (B.1). If we do not filter the duplicates, the actual expected size of the lists at stage i is $N'\mathbf{P}$ where

$$N' \triangleq L_{i-1}^2 \cdot \frac{\binom{k-p_{i-1}}{p_i/2} \binom{p_{i-1}}{p_i/2}}{\binom{k}{p_{i-1}}} \quad (\text{B.3})$$

is the number of pairs in a list at stage $i - 1$ which are a representation of a vector of weight p_i . After filtering the duplicates in the resulting lists, we finally obtain

$$S_i = \mathcal{O}(\min(N'\mathbf{P}, L_i)) \quad (\text{B.4})$$

where L_i is the maximal size of the list, obtained when the whole set of vectors with the desired weight distribution is typically produced by the algorithm at the considered stage; thus we have

$$L_i = \binom{k}{p_i} \prod_{h=1}^i \frac{\binom{\ell_h}{w_h^{(i)}}}{2^{\ell_h}} \quad (\text{B.5})$$

Note that if the assumptions in the correctness lemma [BM18, Lemma 2, p. 17] are met, then we actually have $L_i = S_i$. This phenomenon can notably be observed for the input values of parameters given in the paper.

For instance, in [BM18, Theorem 3, Section 5, p. 19] the following parameters are given for the full decoding at the hardest rate $\frac{k}{n} = 0.46$:

$$\begin{array}{cccccc} \frac{\ell_1}{n} = 0.0366 & \frac{\ell_2}{n} = 0.0547 & \frac{\ell_3}{n} = 0.0911 & \frac{\ell_4}{n} = 0.3576 & & \\ \frac{p_1}{n} = 0.00559 & \frac{w_1^{(1)}}{n} = 0.011515 & & & & \\ \frac{p_2}{n} = 0.01073 & \frac{w_1^{(2)}}{n} = 0.023029 & \frac{w_2^{(2)}}{n} = 0.016676 & & & \\ \frac{p_3}{n} = 0.02029 & \frac{w_1^{(3)}}{n} = 0.0232 & \frac{w_2^{(3)}}{n} = 0.033351 & \frac{w_3^{(3)}}{n} = 0.009993 & & \\ \frac{p_4}{n} = 0.03460 & \frac{w_1^{(4)}}{n} = 0.0066 & \frac{w_2^{(4)}}{n} = 0.0099 & \frac{w_3^{(4)}}{n} = 0.0114 & w_4^{(4)} = 0.0612 & \end{array}$$

With these parameters, if we believe the bound (B.1), we should have

$$S_0 \leq 2^{0.02179n} \quad S_1 \leq 2^{0.03987n} \quad S_2 \leq 2^{0.05939n} \quad S_3 \leq 2^{0.05975n}$$

But since the correctness lemma is verified we can use formula (B.5) to compute the expected size of the lists:

$$S_0 = 2^{0.02179n} \quad S_1 = 2^{0.03987n} \quad S_2 = 2^{0.0655n} \quad S_3 = 2^{0.0705n}$$

We can see that S_2 and S_3 exceed their presumed upper bound, which increases the complexity of the algorithm to $C = \mathcal{O}(2^{0.1083n})$ in comparison to the claimed one $C = \mathcal{O}(2^{0.0885n})$.