SMOOTHING CODES AND LATTICES: SYSTEMATIC STUDY AND NEW BOUNDS

THOMAS DEBRIS-ALAZARD[1], LÉO DUCAS[2,3], NICOLAS RESCH[2], AND JEAN-PIERRE TILLICH[1]

Abstract. In this article we revisit smoothing bounds in parallel between lattices *and* codes. Initially introduced by Micciancio and Regev, these bounds were instantiated with Gaussian distributions and were crucial for arguing the security of many lattice-based cryptosystems. Unencumbered by direct application concerns, we provide a systematic study of how these bounds are obtained for both lattices *and* codes, transferring techniques between both areas. We also consider various spherically symmetric noise distributions.

We found that the best strategy for a worst-case bound combines Parseval's Identity, the Cauchy-Schwarz inequality, and the second linear programming bound, and this for both codes and lattices, and for all noise distributions at hand. For an average-case analysis, the linear programming bound can be replaced by a tight average count.

This alone gives optimal results for spherically uniform noise over random codes and random lattices. This also improves previous Gaussian smoothing bound for worst-case lattices, but surprisingly this provides even better results for uniform noise than for Gaussian (or Bernouilli noise for codes).

This counter-intuitive situation can be resolved by adequate decomposition and truncation of Gaussian and Bernouilli distribution into a superposition of uniform noise, giving further improvement for those cases, and putting them on par with the uniform cases.

1. Introduction

1.1. **Smoothing bounds.** In either a code or a lattice, smoothing refers to fact that, as an error distribution grows wider and wider, the associated syndrome distribution tends towards a uniform distribution. In other words, the error distribution, reduced modulo the code or the lattice, becomes essentially flat. This phenomenon is pivotal in arguing security of cryptosystems [MR07, GPV08, DST19]. In information theoretic literature, it is also sometimes referred to as flatness [LLBS14]. Informally, by a "smoothing bound" we are referring to a result which lower bounds the amount of noise which needs to be added so that the smoothed distribution "looks" flat.

To be more concrete, by a "flat distribution", we are referring to a uniform distribution over the ambient space modulo the group of interest. For a (linear) code $\mathcal{C} \subseteq \mathbb{F}_2^n$, this quotient space is $\mathbb{F}_2^n/\mathcal{C}$; for a lattice $\Lambda \subseteq \mathbb{R}^n$, it is $\mathbb{R}^n/\Lambda$. We then consider some "noise" vector $\mathbf{e}$ distributed over the ambient space $\mathbb{F}_2^n$ (respectively, $\mathbb{R}^n$), and attempt to prove that $\mathbf{e} \mod \mathcal{C}$ (respectively, $\mathbf{e} \mod \Lambda$) is "close" to the uniform distribution over the quotient space $\mathbb{F}_2^n/\mathcal{C}$ (respectively, $\mathbb{R}^n/\Lambda$). To quantify "closeness" between distributions, we will use the standard choice of *statistical distance*.

[1] Inria

[2] CWI, Amsterdam, The Netherlands

[3] Mathematical Institute, Leiden University

E-mail addresses: thomas.debris@inria.fr, L.Ducas@cwi.nl, Nicolas.Resch@cwi.nl, jean-pierre.tillich@inria.fr.

The work of TDA and JPT was funded by the French Agence Nationale de la Recherche through ANR JCJC COLA (ANR-21-CE39-0011) for TDA and ANR CBCRYPT (ANR-17-CE39-0007) for JPT. NR is partially supported by ERC H2020 grant No.74079 (ALGSTRONGCRYPTO). LD is supported by an ERC starting Grant 947821 (ARTICULATE).

An important question to be addressed is the choice of distribution for the noise vector $\mathbf{e}$. In lattice-based cryptography (where such smoothing bounds originated [MR07]), the literature ubiquitously uses Gaussian distributions for errors, and smoothness is guaranteed for an error growing as the inverse of the minimum distance of the dual lattice. The original chain [MR07] of argument goes as follows:

- Apply the Poisson summation formula (PSF);
- Bound variations via the triangle inequality (TI) over all non-zero dual lattice points;
- Bound the absolute sum above via the Banaszczyk tail bound [Ban93] for discrete Gaussian (BT).

An intermediate quantity called the smoothing parameter introduced by [MR07] before the last step is also often used in the lattice-based cryptographic literature. Each bounding step is potentially non-tight, and indeed more recent works have replaced the last step by the following [ADRS15]:

- Bound the number of lattice points in balls of a given radius via the Linear Programming bound [Lev79] (LP) and "sum over all radii" (with care).

With this LP strategy, it is in principle possible to also compute a smoothing bound for spherically symmetric distributions of errors other than the Gaussian; however, we are not aware of prior work doing this explicitly. A very natural choice would be uniform distributions over Euclidean balls.

For codes, there are also two natural distributions of errors: Bernoulli noise, *i.e.* flip each bit independently with some probability p (a.k.a. the binary symmetric BSC_p channel), and a uniform noise over a Hamming sphere of a fixed radius. The latter is typically preferred for the design of concrete and practical cryptosystems [McE78, Ale11, MTSB13, DST19], while the former appears more convenient in theoretical works[(1)]. Cryptographic interest for code smoothing has recently arisen [BLVW19, YZ20], but results are so far limited to codes with extreme parameters and specific "balancedness" constraints. However we note that the question is not entirely new in the coding literature (see for instance [Kløve07]). In particular, an understanding of the smoothing properties of Bernoulli noise is intimately connected to the *undetected error probability* of a code transmitted through the BSC_p.

In this light, it is interesting to revisit and systematize our understanding of smoothing bounds, unencumbered by direct application concerns. We find it enlightening to do this exploration in parallel between codes and lattices, transferring techniques back and forth between both areas whenever possible.

Furthermore, we keep our arguments agnostic to the specific choice of error distribution, allowing us to apply them with different error distributions and compare the results. To compare different (symmetric) distributions, we advocate parametrizing them by the expected weight/norm of a vector. That is, we quantify the magnitude of a noise vector $\mathbf{e}$ by $t = \mathbb{E}(|\mathbf{e}|)$ (where $|\cdot|$ denotes either the Hamming weight or the Euclidean norm of the vector). Our smoothing bounds will depend on this parameter, and we consider a smoothing bound to be more effective if for the smoothed distribution to be close to uniform we require a smaller lower-bound on t.

1.2. **Contributions.** In this work, we collect the techniques that have been used for smoothing, both in the code and lattice contexts. We view individual steps as modular components of arguments, and consider all permissible combinations of steps, thereby determining the most effective

[(1)] A third choice of distribution, described as a discrete-time random walk, also made an appearance for a complexity theoretic result [BLVW19]. The expert reader may note that the Bernoulli distribution can also be treated as a continuous-time random walk, and both can be analysed via the heat kernel formalism [Chu97, Chap. 10].

arguments. In the following, we outline our systematization efforts, describing the various proof frameworks that we tried before settling on the most effective argument.

Code smoothing bounds. Given the relative dearth of results concerning code smoothing, it seems natural to start by adapting the first argument (PSF+TI+BT) to codes following the proof techniques of [Ban93, MR07]. And indeed, the whole strategy translates flawlessly, with only one caveat: it leads to a very poor result, barely better than the trivial bound. Namely, smoothness is established only for Bernoulli errors with parameter very close to $p = 1/2$.

The adaptation of Banaszczyk tail bound [Ban93] to codes (where replacing the Gaussian by a Bernoulli distribution) is rather naïve, and it is therefore not very surprising that it leads to a disappointing result. Instead, we can also follow the improved strategy for lattices from [ADRS15], and resort to linear programming bounds for codes [Bas65, MRJW77, ABL01]. Briefly, by an LP bound we are referring to a result that bounds the number of codewords (*resp.* lattice vectors) of a certain weight (*resp.* norm) in terms of the dual distance (*resp.* shortest dual vector) of the code (*resp.* lattice). In both cases, the results are obtained by considering a certain LP relaxation of the combinatorial quantities one wishes to bound, hence the name. Even more, the bounds for codes and lattices are obtained via essentially the same arguments [MRJW77, DL98, CE03]. We therefore find it natural to apply LP bounds in our effort to develop proof techniques which apply to both code- and lattice-smoothing.

The strategy (PSF+TI+LP) turns out to give a significantly better result, but it nevertheless still appears to be far from optimal. We believe that the application of the triangle inequality in the second step to bound the sum of Fourier coefficients given by the Poisson summation formula leads to the unsatisfactory bound. Indeed, a common heuristic when dealing with sums of Fourier coefficients is that, unless there is a good reason otherwise, the sum should have magnitude roughly the square-root of the order of the group (as is the case for random signs): the triangle inequality is far too crude to notice this.

Instead, we turn to another common upper-bound on a sum, namely, the Cauchy-Schwarz (CS) inequality. It is natural to subsequently apply Parseval's Identity (PI). It turns out that this strategy yields very promising results, upon which we now elucidate. The upper-bound is described in terms of the *weight distribution* of a code, *i.e.* the number of codewords of weight w for each $w = 1, \ldots, n$. Unfortunately, it is quite difficult to understand the weight distribution of arbitrary codes, and the bounds that we do have are quite technical.

Random codes. For this reason, we first apply our proof template to *random codes*, as it is quite simple to compute the (expected) weight distribution of a random code. Quite satisfyingly, the simple two steps arguments (PI+CS) already yields *optimal* results for this case, but when the error is sampled uniformly at random from a sphere! That is, we can show that the support size of the error distribution matches the obvious lower bound that applies to *any* distribution that successfully smoothes a code: namely, for a code $\mathcal{C}$ the support size must be at least $\#(\mathbb{F}_2^n/\mathcal{C})$. Using coding-theoretic terminology, the weight of the error vector that we need to smooth is given by the ubiquitous Gilbert-Varshamov bound

$$\omega_{\mathrm{GV}}(R) = h^{-1}(1 - R)$$

which characterizes the trade-off between a random code's rate R and its minimum distance. Here, h^{-1} is the inverse of the binary entropy function.

Moreover, as the argument is versatile enough to apply to essentially all spherical error distributions, we also tried applying it to the Bernoulli distribution, and the random walk distribution of [BLVW19]. Comparing them, we were rather surprised that our argument provided better bounds for the uniform distribution over a Hamming sphere than the other two distributions for the same average Hamming weight.

However, while the (PI+CS) sequence of arguments is more effective when the noise is sampled uniformly on the sphere, we can exploit the fact that the Hamming weight of a Bernoulli-distributed vector is tightly concentrated to recover the same smoothing bound for this distribution. In more detail, we use a "truncated" argument. First, we decompose the Bernoulli distribution into a convex combination of uniform sphere distributions. But, by Chernoff's bound, a Bernoulli distribution is concentrated on vectors whose weight lies in a width εn interval around its expected weight. Therefore, outside of this interval, the contribution of the Bernoulli on the statistical distance is negligible. Then apply the (PI+CS) sequence of arguments to each constituent distribution close to the expected weight. In this way, we are able to demonstrate that Bernoulli distributions also optimally smooth random codes.

Arbitrary codes. Next, we turn our attention to smoothing worst-case codes. Motivated by our success in smoothing random codes, we again follow the (PI+CS) sequence of arguments and combine this with LP bounds to derive smoothing bounds when the dual distance of the code is sufficiently large. Again, the sequence of arguments is most effective when the error is distributed uniformly over the sphere, with one caveat: we are also required to assume that the dual code is *balanced* in the sense that it also does not contain any vectors of too large weight. While this assumption has appeared in other works [BLVW19, YZ20], we find it somewhat unsatisfactory.

Fortunately, this condition is not required if the error is sampled according to the Bernoulli distribution. But then we run into the same issue that we had earlier with random codes: the (PI+CS) argument, followed by LP bounds, natively yields a lesser result when instantiated with Bernoulli noise. Fortunately, we have already seen how to resolve this issue: we pass to the truncated Bernoulli distribution and decompose it into uniform sphere distributions. This yields a best-of-both-worlds result: we obtain the strongest smoothing bound we can in terms of the noise magnitude, while requiring the weakest assumption on the code.

And back to lattices. Having now uncovered this better strategy for codes, we can return to lattices and apply our new proof template. Indeed, as we outline in Section 2.3, the (PI+CS) sequence of arguments can be applied in a very broad context; see, in particular, Corollary 1.

Random lattices. First, just as we set our expectations for code-smoothing by first studying the random case, we analogously start here by considering random lattices. However, defining a random lattice is a non-trivial task. We actually consider two distributions. The first, which is based on the deep Minkowski-Hlwaka-Siegel (MHS) Theorem, we only abstractly describe. Thanks to the MHS Theorem, we can very easily compute the (expected value) of our upper-bound.

For the MHS distribution of lattices, we consider two natural error distributions: the Gaussian distribution (which is used ubiquitously in the literature), as well as the uniform distribution over the Euclidean ball. And again, perhaps surprisingly (although less so now thanks to our experience with the code case), we obtain a better result with the uniform distribution over the Euclidean ball. And moreover, the Euclidean ball result is *optimal* in the same sense that we had for codes: the support volume of the error distribution is exactly equal to the covolume of the lattice.[(2)] We view the value w such that the volume of the n-ball of radius w is equal to the covolume of a lattice as being the lattice-theoretic analogue of the Gilbert-Varshamov quantity, and it is as half quantity that appears in Minkowski bound,

$$w_{\mathrm{M}/2} \stackrel{\text{def}}{=} \frac{\sqrt[n]{\mathrm{vol}(\mathbb{R}^n/\Lambda)\cdot\Gamma(n/2+1)}}{\sqrt{\pi}} \ .$$

However, as Gaussian vectors satisfy many pleasing properties that are often exploited in lattice-theoretic literature, we would like to obtain the same smoothing bound for this error distribution.

[(2)] That is, for a lattice Λ, the volume of the torus $\mathbb{R}^n/\Lambda$.

Distribution	Proof strategy	smoothing factor F	General statement
Gaussian	PSF+TI+BT	$1/(2\pi) \approx 0.15915$	Lemma 3.2 [MR07]
Gaussian	PSF+TI+LP	$C_{\mathrm{KL}}/(2\pi\sqrt{e}) \approx 0.12746$	Lemma 6.1 [ADRS15]
Gaussian	PI+CS+LP	$C_{\mathrm{KL}}/(2\pi\sqrt{2e}) \approx 0.09013$	Theorem 11 (this work)
Unif. Euclidean ball	PI+CS+LP	$C_{\mathrm{KL}}/(2\pi e) \approx 0.07731$	Theorem 10 (this work)
Gaussian	via Unif. + Trunc.	$C_{\mathrm{KL}}/(2\pi e) \approx 0.07731$	Theorem 12 (this work)

TABLE 1. Comparing smoothing bounds for various proof strategies and error distributions. The smoothing constant F is the smallest constant C such that the bounds proves exponential smoothness when the average norm of an error is at least C times the minimal distance of the dual lattice. Here $C_{\mathrm{KL}} \approx 2^{0.401}$ denotes the constant that is involved in the Kabatiansky and Levenshtein bound [KL78].

Fortunately, our experience with codes also tells us how to recover the result for Gaussian noise from the Euclidean ball noise smoothing bound: we decompose the Gaussian distribution appropriately into a convex combination of Euclidean ball distributions. Together with a basic tail, we recover the same smoothing bound for Gaussian noise that we had for the uniform ball noise.

We also study random q-ary lattices, which are more concretely defined: following the traditional lattice-theoretic terminology, they are obtained by applying Construction A to a random code. This does lead to a slight increase in the technicality of the argument – in particular, we need to apply a certain "summing over annuli" trick – but the computations are still relatively elementary. Again, we find that the argument naturally works better when the errors are distributed uniformly over a ball, but we can still transfer the bound to the Gaussian noise.

Arbitrary lattices. Next, we address the challenge of smoothing arbitrary lattices. And again, we follow the (PI+CS) sequence of arguments, and subsequently use the Kabatiansky and Levenshtein bound [KL78] to obtain a smoothing bound in terms of the minimum distance of the dual lattice. The Kabatiansky and Levenshtein bound is the lattice-analogue of the second LP bound from coding theory. We can directly apply the arguments with both of our error distributions of interest, and again, the uniform ball distribution wins. But the decomposition and tail-bound trick again applies to yield the same result for the Gaussian distribution that we had for the uniform ball distribution.

Comparison. We summarize how our work improves on the state of the art in Table 1 for lattices, and in Table 2 and Figure 1 for codes.

In the case of lattices (Table 1), we fix the smoothing bound target to exponentially small, that is we state the minimal value of $F > 0$ such that the bound over the statistical distance implies $\Delta(\mathbf{e} \bmod \Lambda, U(\mathbb{R}^n/\Lambda)) \leq 2^{-\Omega(n)}$ when the error follows the prescribed distribution and of an average Euclidean length of $\mathbb{E}(|\mathbf{e}|_2) = F \cdot n/\lambda_1^*(\Lambda)$.[3]

In the case of codes we also fix the smoothing bound target to negligible,[4] but we compare two cases: smoothing bounds for random codes (in average) and for a fixed code (worst case). In Figure 1 we compare the minimal value $F > 0$ such that $\mathbb{E}_{\mathcal{C}}\left(\Delta(\mathbf{e} \bmod \mathcal{C}, U(\mathbb{F}_2^n/\mathcal{C}))\right) \leq 2^{-\Omega(n)}$ when the error $\mathbf{e}$ follows the prescribed distribution and with an expectation that is taken over codes of rate R. In Table 1 we make the same comparison but to reach $\Delta(\mathbf{e} \bmod \mathcal{C}, U(\mathbb{F}_2^n/\mathcal{C})) \leq 2^{-\Omega(n)}$ for a fixed code $\mathcal{C}$ such that the minimum distance of its dual $\mathcal{C}^*$ is known.

[3] In fact, the values in this table guarantee exponentially small statistical distance from the uniform distribution.

[4] Again, it is the same if we insist the statistical distance to uniform is exponentially small.

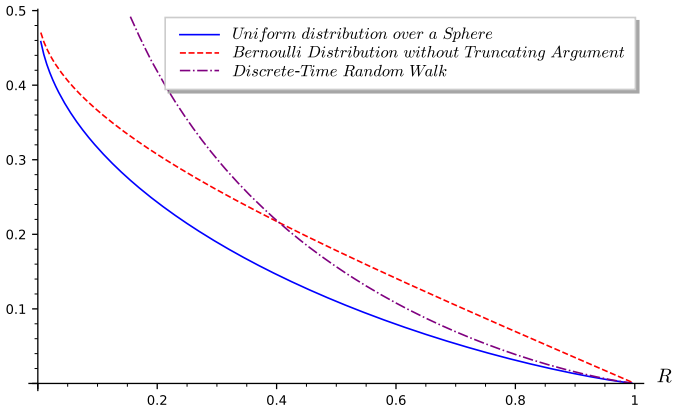

FIGURE 1. Comparing smoothing constants, for *random* codes as function of their rate R, for various error distributions . The smoothing constant is the smallest constant C such that the bounds proves exponential smoothness when the average Hamming weight of an error is at least C times n.

Distribution	smoothing factor F	Balanced-code	General statement
Bernoulli	≈ 0.24	NO	Eq. (18), Prop. 7, 8
Discrete Rand. Walk	≈ 0.27	YES	Theorem 4
Unif. Hamming sphere	≈ 0.17	YES	Theorem 4
Bernoulli + Trunc.	≈ 0.17	NO	Theorem 5

TABLE 2. Comparing smoothing bounds, for a code $\mathcal{C}$ of length n such that its dual $\mathcal{C}^*$ has minimum distance $0.11n$ (which is the typical case for a code of rate $1/2$), for various error distributions . The smoothing constant F is the smallest constant C such that the bounds proves exponential smoothness when the average Hamming weight of an error is at least C times n. Furthermore the balanced-code hypothesis means that we suppose there are no dual codewords $\mathbf{c}^* \in \mathcal{C}^*$ of Hamming weight larger than $(1 - 0.11)n$.

2. Preliminaries: Notations and Fourier Analysis over Locally Compact Abelian Group

2.1. **General Notation.** The notation $x \stackrel{\text{def}}{=} y$ means that x is defined as being equal to the quantity y. Given a set $\mathscr{S}$, its indicator function will be denoted $1_{\mathscr{S}}$. For a finite set $\mathscr{S}$, we will denote by $\#\mathscr{S}$ its cardinality. Vectors will be written with bold letters (such as $\mathbf{x}$).

The statistical distance between two discrete probability distributions f and g over a same space $\mathscr{S}$ is defined as:

$$\Delta(f,g) \stackrel{\text{def}}{=} \frac{1}{2} \sum_{x \in \mathscr{S}} |f(x) - g(x)|.$$

Similarly, for two continuous probability density functions f and g over a same measure space $\mathscr{E}$, the statistical distance is defined as

$$\Delta(f,g) \stackrel{\text{def}}{=} \frac{1}{2} \int_{\mathscr{E}} |f - g|.$$

2.2. **Codes and Lattices.** We give here some basic definitions and notation about linear codes and lattices.

Codes. In the whole paper, we will deal exclusively with binary linear codes, namely subspaces of $\mathbb{F}_2^n$ for some positive integer n. The space $\mathbb{F}_2^n$ will be embedded with the Hamming weight $|\cdot|$, namely

$$\forall \mathbf{x} \in \mathbb{F}_2^n, \quad |\mathbf{x}| \stackrel{\text{def}}{=} \#\{i \in [\![1, n]\!] : x_i \neq 0\}.$$

We will denote by $\mathscr{S}_w(\mathbf{x})$ the sphere with center $\mathbf{x}$ and radius w. For the center $\mathbf{0}$ we will use the notation $\mathscr{S}_w$. The size of any sphere of radius w is given by $\binom{n}{w}$ and we have $\frac{1}{n}\log_2\binom{n}{w} = h(w/n)+o(1)$ where h denotes the binary-entropy, namely $h(x) \stackrel{\text{def}}{=} -x\log_2(x)-(1-x)\log_2(1-x)$. An $[n,k]$-code $\mathcal{C}$ is defined as a dimension k subspace of $\mathbb{F}_2^n$. The rate of $\mathcal{C}$ is $\frac{k}{n}$. Its minimal distance is given by

$$\begin{aligned} d_{\min}(\mathcal{C}) &\stackrel{\text{def}}{=} \min\{|\mathbf{c}-\mathbf{c}'| : \mathbf{c}, \mathbf{c}' \in \mathcal{C} \text{ and } \mathbf{c} \neq \mathbf{c}'\} \\ &= \min\{|\mathbf{c}| : \mathbf{c} \in \mathcal{C} \text{ and } \mathbf{c} \neq \mathbf{0}\}. \end{aligned}$$

The number of codewords of $\mathcal{C}$ of weight t will be denoted by $N_t(\mathcal{C})$, namely:

$$N_t(\mathcal{C}) \stackrel{\text{def}}{=} \#\{\mathbf{c} \in \mathcal{C} \text{ and } |\mathbf{c}| = t\}.$$

The dual of a code $\mathcal{C}$ is defined as $\mathcal{C}^* \stackrel{\text{def}}{=} \{\mathbf{c}^* : \forall \mathbf{c} \in \mathcal{C},\ \mathbf{c}\cdot\mathbf{c}^* = 0\}$ where $\cdot$ denotes the standard inner product on $\mathbb{F}_2^n$.

Lattices. We will consider lattices of $\mathbb{R}^n$ which is embedded with the Euclidean norm $|\cdot|_2$, namely

$$\forall \mathbf{x} \in \mathbb{R}^n, \quad |\mathbf{x}|_2 \stackrel{\text{def}}{=} \sqrt{\sum_{i=1}^{n} x_i^2}.$$

We will denote by $\mathscr{B}_w(\mathbf{x})$ the ball with center $\mathbf{x}$ and radius w. For the center $\mathbf{0}$ we will use the notation $\mathscr{B}_w$. The volume of any ball of radius w is given by:

$$V_n(w) \stackrel{\text{def}}{=} \frac{\pi^{n/2} w^n}{\Gamma(n/2+1)}.$$

An n-dimension lattice Λ is defined as a discrete subgroup of $\mathbb{R}^n$. The covolume $|\Lambda| \stackrel{\text{def}}{=} \mathrm{vol}(\mathbb{R}^n/\Lambda)$ of a lattice Λ is the volume of any fundamental parallelotope. The minimal distance of Λ is given by $\lambda_1(\Lambda) \stackrel{\text{def}}{=} \min\{|\mathbf{x}|_2 : \mathbf{x} \in \Lambda \text{ and } \mathbf{x} \neq \mathbf{0}\}$. The number of lattice points of Λ of weight $\leq t$ will be denoted by $N_{\leq t}(\Lambda)$, namely:

$$N_{\leq t}(\Lambda) \stackrel{\text{def}}{=} \#\{\mathbf{x} \in \Lambda : |\mathbf{x}|_2 \leq t\}.$$

2.3. **Fourier Analysis.** We give here a brief introduction to Fourier analysis over arbitrary locally compact Abelian groups. Our general treatment will allow us to apply directly some basic results in a code and lattice context, obviating the need in each case to introduce essentially the same definitions and to provide the same proofs.

Corollary 1 at the end of this subsection is the starting point of our smoothing bounds: all of our results are obtained by using different facts to bound the right hand side of the inequality.

Groups and Their Duals. In what follows G will denote a locally compact Abelian group. Such a group admits a Haar measure μ. For instance $G = \mathbb{R}$ with μ the Lebesgue measure λ, or $G = \mathbb{F}_2^n$ with μ the counting measure $\#$.

The dual group $\widehat{G}$ is given by the continuous group homomorphisms χ from G into the multiplicative group of complex numbers of absolute value 1, and it is again a locally compact Abelian group. In Figure 2 we give groups, their duals as well as their associated Haar measures that will be considered in this work.

G	μ	$\widehat{G}$	μ
$\mathbb{F}_2^n$	$\frac{1}{2^n}\,\#$		
$\mathbb{F}_2^n/\mathcal{C}$	$\frac{\#\mathcal{C}}{2^n}\,\#$	$\widehat{\mathbb{F}_2^n/\mathcal{C}} \simeq \mathcal{C}^*$	$\#$
$\mathcal{C}$	$\frac{1}{\#\mathcal{C}}\,\#$		
$\mathbb{R}^n$	λ		
$\mathbb{R}^n/\Lambda$	$\frac{1}{\lvert\Lambda\rvert}\,\lambda$	$\widehat{\mathbb{R}^n/\Lambda} \simeq \Lambda^*$	$\#$
Λ	$\#\,\lvert\Lambda\rvert$		

FIGURE 2. Some groups G, their duals $\widehat{G}$ and their associated Haar measures. Here λ denotes the Lebesgue measure and $\#$ the counting measure.

It is important to note that if $H \subseteq G$ is a closed subgroup, then G/H and H are also locally compact groups. Furthermore, G/H has dual group that satisfies the following isomorphism

$$\widehat{G/H} \simeq H^{\perp} \stackrel{\text{def}}{=} \left\{\chi \in \widehat{G} :\ \forall h \in H,\ \chi(h) = 1\right\}.$$

Norms and Fourier Transforms. For any $p \in [1, \infty[$, $L_p(G)$ will denote the space of measurable functions $f : G \to \mathbb{C}$ (up to functions which agree almost everywhere) with finite norm $\|f\|_p$ which is defined as:

$$\|f\|_p^p \stackrel{\text{def}}{=} \int_G |f|^p d\mu.$$

The Fourier transform of $f \in L_1(G)$ is defined as:

$$\widehat{f} : \chi \in \widehat{G} \longmapsto \int_G f\overline{\chi} d\mu.$$

We omitted here the dependence on G. It will be clear from the context.

Theorem 1 (Parseval's Identity)**.** *Let $f \in L_1(G) \cap L_2(G)$, then with appropriate normalization of the Haar measure*

$$\|f\|_2 = \|\widehat{f}\|_2.$$

Poisson Formula. Given $H \subseteq G$ and any function $f : G \to \mathbb{C}$, its restriction over H is defined as $f_{|H} : h \in H \mapsto f(h) \in \mathbb{C}$. We define its periodization as follows.

Definition 1 (Periodization)**.** *Let H be a closed subgroup of G and $f \in L^1(G)$. We define the H-periodization of f as:*

$$f^{|H} : (g + H) \in G/H \longmapsto \int_H f(g + h) d\mu_H(h) \in \mathbb{C}$$

where μ_H denotes any choice of the Haar measure for H.

There always exists a Haar measure $\mu_{G/H}$ such that for any continuous function with compact support $f : G \to \mathbb{C}$ the quotient integral formula holds:

$$\int_{G/H} \left(\int_H f(g + h) d\mu_H(h) \right) d\mu_{G/H}(g + H) = \int_G f(g) d\mu(g). \tag{1}$$

Theorem 2 (Poisson Formula)**.** *Let $H \subseteq G$ be a closed subgroup and $f \in L^1(G)$, then with appropriate normalization of the Haar measures,*

$$\widehat{(f^{|H})} = \left(\widehat{f}\right)_{|\widehat{G/H}}.$$

The following corollary is a simple consequence of the Cauchy-Schwarz inequality, Parseval identity and the Poisson formula. Our results on smoothing bounds are all based on this theorem.

Corollary 1. *Let H be a closed subgroup of G. Let $a : x \in G/H \mapsto 1$ and $f \in L^1(G)$ such that $\int_G f d\mu = \mu_{G/H}(G/H)$. Then with appropriate normalization of the Haar measure*[5]*,*

$$\|a - f^{|H}\|_1 \leq \sqrt{\mu_{G/H}(G/H)} \sqrt{\int_{\widehat{G/H}\setminus\{\chi_{\mathbf{0}}\}} |\widehat{f}|^2 \, d\mu_{\widehat{G/H}}}$$

where $\chi_{\mathbf{0}}$ denotes the identity element of $\widehat{G/H}$.

Proof. We have

$$\begin{aligned}
\|a - f^{|H}\|_1 &= \int_{\widehat{G/H}} |a - f^{|H}| d\mu_{G/H} \\
&\leq \sqrt{\mu_{G/H}(G/H)} \cdot \|a - f\|_2 \quad \text{(By Cauchy-Schwarz)} \\
&= \sqrt{\mu_{G/H}(G/H)} \cdot \|\widehat{a} - \widehat{f}\|_2 \quad \text{(By Parseval)} \\
&= \sqrt{\mu_{G/H}(G/H)} \cdot \sqrt{\int_{\widehat{G/H}\setminus\{\chi_{\mathbf{0}}\}} |\widehat{f^{|H}}|^2 d\mu_{\widehat{G/H}}} \qquad (2) \\
&= \sqrt{\mu_{G/H}(G/H)} \cdot \sqrt{\int_{\widehat{G/H}\setminus\{\chi_{\mathbf{0}}\}} |\widehat{f}|^2 d\mu_{\widehat{G/H}} 2} \quad \text{(By Poisson)}
\end{aligned}$$

where in Equation (2) we used the following equalities:

$$\begin{aligned}
\widehat{f^{|H}}(\chi_{\mathbf{0}}) &= \int_{G/H} f^{|H} \overline{\chi_{\mathbf{0}}} \, d\mu_{G/H} \\
&= \int_{G/H} \left(\int_H f(g+h) d\mu_H(h) \right) d\mu_{G/H}(g+H) \\
&= \int_G f \quad \text{(By Equation (1))} \\
&= \mu_{G/H}(G/H) \quad \text{(By assumption on } f\text{)}
\end{aligned}$$

and

$$\widehat{a}(\chi_{\mathbf{0}}) = \int_{G/H} u\overline{\chi_{\mathbf{0}}} d\mu_{G/H} = \mu_{G/H}(G/H), \quad \forall \chi \in \widehat{G/H} \setminus \{\chi_{\mathbf{0}}\}, \ \widehat{a}(\chi) = \int_{G/H} \overline{\chi} d\mu_{G/H} = 0. \quad \square$$

In this work we will choose $G = \mathbb{R}^n$ and $H = \Lambda$ or $G = \mathbb{F}_2^n$ and $H = \mathcal{C}$. Haar measures associated to $G, G/H$ and $\widehat{G/H}$ for which the corollary holds are given in Figure 2. Furthermore, we will use Fourier transforms over $\widehat{G}$ and $\widehat{G/H}$. We describe in Figure 3 these dual groups that we will consider.

3. Smoothing Bounds: Code Case

Given a binary linear code $\mathcal{C}$ of length n, the aim of smoothing bounds is to quantify at which condition on the noise, $\mathbf{c} + \mathbf{e}$ is statistically close to the uniform distribution over $\mathbb{F}_2^n$ when $\mathbf{c}$ is uniformly drawn from $\mathcal{C}$ and $\mathbf{e}$ picked up according to some noise distribution f. Equivalently, we want to understand when $(\mathbf{e} \mod \mathcal{C}) \in \mathbb{F}_2^n/\mathcal{C}$ is close to the uniform distribution. We will focus on the case where the distribution of $\mathbf{e}$ is radial, meaning that it only depends on the Hamming weight of $\mathbf{e}$. We will use throughout this section the following notation

[5] We choose the Haar measures μ_G, μ_H,$\mu_{G/H}$ and $\widehat{\mu_{G/H}}$ for which both Poisson formula and Parseval hold.

$\mathbb{R}^n$	$\mathbb{F}_2^n$
$\widehat{\mathbb{R}^n} = \left\{\chi_{\mathbf{x}} : \mathbf{y} \in \mathbb{R}^n \mapsto e^{-2i\pi \mathbf{x}\cdot\mathbf{y}},\ \mathbf{x} \in \mathbb{R}^n\right\}$	$\widehat{\mathbb{F}_2^n} = \{\chi_{\mathbf{x}} : \mathbf{y} \in \mathbb{F}_2^n \mapsto (-1)^{\mathbf{x}\cdot\mathbf{y}},\ \mathbf{x} \in \mathbb{F}_2^n\}$
$\widehat{\mathbb{R}^n/\Lambda} = \{\chi_{\mathbf{x}},\ \mathbf{x} \in \Lambda^*\}$	$\widehat{\mathbb{F}_2^n/\mathcal{C}} = \{\chi_{\mathbf{x}},\ \mathbf{x} \in \mathcal{C}^*\}$
$\widehat{f}(\mathbf{x}) = \int_{\mathbb{R}^n} f(\mathbf{y}) e^{2i\pi \mathbf{x}\cdot\mathbf{y}} d\mathbf{y}$	$\widehat{f}(\mathbf{x}) = \frac{1}{2^n}\sum_{\mathbf{y}\in\mathbb{F}_2^n} f(\mathbf{y})(-1)^{\mathbf{x}\cdot\mathbf{y}}$

FIGURE 3. Dual groups and Fourier transforms that we will consider. We will identify $\widehat{f}(\chi_{\mathbf{x}})$ with $\widehat{f}(\mathbf{x})$.

Notation 1.

- *u followed or not by a subscript will always denote the uniform distribution.*
- *We will use mostly the uniform probability distribution over the quotient space $\mathbb{F}_2^n/\mathcal{C}$ and for this reason we just write u in this case. The uniform distribution over the whole space $\mathbb{F}_2^n$ is denoted by u_{full} and the uniform distribution over the codewords of $\mathcal{C}$ is denoted by $u_{\mathcal{C}}$.*
- *We also use the uniform distribution over the Hamming sphere $\mathscr{S}_w$ of radius w centered at 0, which we denote by u_w.*
- *for two probability distributions f and g over $\mathbb{F}_2^n$ we denote by $f \star g$ the convolution over $\mathbb{F}_2^n$: $f \star g(\mathbf{x}) = \sum_{\mathbf{y}\in\mathbb{F}_2^n} f(\mathbf{x}-\mathbf{y})g(\mathbf{y})$.*

It will be more convenient to work in the quotient space and for this we use

Proposition 1. *Let f be a probability distribution over $\mathbb{F}_2^n$ and $\mathcal{C}$ be an $[n,k]$-code. We have:*

$$\begin{aligned} \Delta(u_{\text{full}}, u_{\mathcal{C}} \star f) &= \Delta(u, f^{\mathcal{C}}), \\ \text{where } f^{\mathcal{C}}(\mathbf{x}) &\stackrel{def}{=} 2^k\, f^{|\mathcal{C}}(\mathbf{x}) = \sum_{\mathbf{c}\in\mathcal{C}} f(\mathbf{x}-\mathbf{c}). \end{aligned}$$

Proof.

$$\begin{aligned} \Delta(u_{\text{full}}, u_{\mathcal{C}} \star f) &= \frac{1}{2}\sum_{\mathbf{x}\in\mathbb{F}_2^n}\left|\frac{1}{2^n} - \mathbb{P}_{u_{\mathcal{C}},f}\left(\mathbf{c}+\mathbf{e}=\mathbf{x}\right)\right| \\ &= \frac{1}{2}\sum_{\mathbf{x}\in\mathbb{F}_2^n}\left|\frac{1}{2^n} - \sum_{\mathbf{c}_0\in\mathcal{C}}\mathbb{P}_f(\mathbf{c}+\mathbf{e}=\mathbf{x} \mid \mathbf{c}=\mathbf{c}_0)\frac{1}{2^k}\right| \\ &= \frac{1}{2}\sum_{\mathbf{x}\in\mathbb{F}_2^n}\left|\frac{1}{2^n} - \frac{1}{2^k}\sum_{\mathbf{c}_0\in\mathcal{C}} f(\mathbf{x}-\mathbf{c}_0)\right| \\ &= \frac{1}{2}\sum_{\mathbf{x}\in\mathbb{F}_2^n/\mathcal{C}}\left|\frac{1}{2^{n-k}} - \sum_{\mathbf{c}_0\in\mathcal{C}} f(\mathbf{x}-\mathbf{c}_0)\right| \quad (3) \\ &= \frac{1}{2}\sum_{\mathbf{x}\in\mathbb{F}_2^n/\mathcal{C}}\left|\frac{1}{2^{n-k}} - f^{\mathcal{C}}(\mathbf{x})\right| \quad (4) \end{aligned}$$

where in Equation (3) we used that each term of the sum is constant on $\mathbf{x}+\mathcal{C}$. □

As a rewriting of Corollary 1 we get the following proposition that upper-bounds $\Delta(u, f^{\mathcal{C}})$, namely:

Proposition 2. *Let $\mathcal{C}$ be an $[n,k]$-code and f be a radial distribution on $\mathbb{F}_2^n$. We have,*

$$\Delta\left(u, f^{\mathcal{C}}\right) \leq 2^n \sqrt{\sum_{t=d_{\min}(\mathcal{C}^*)}^{n} N_t(\mathcal{C}^*)|\widehat{f}(t)|^2}$$

where by abuse of notation we denote by $\widehat{f}(t)$ the common value of $\widehat{f}$ on vectors of weight t.

Proof. We have that $\mathcal{C}$ is a closed subgroup of $\mathbb{F}_2^n$ with associated Haar measures:

$$\mu_{\mathbb{F}_2^n} = \frac{1}{2^n}\,\# \quad \text{and} \quad \mu_{\mathbb{F}_2^n/\mathcal{C}} = \frac{2^k}{2^n}\,\#$$

for which we can apply Corollary 1. Let, $a \stackrel{\text{def}}{=} 2^{n-k}u$ and $b \stackrel{\text{def}}{=} 2^n f$. First, it is clear that $a : \mathbf{x} \in \mathbb{F}_2^n/\mathcal{C} \mapsto 1$ and that

$$\int_{\mathbb{F}_2^n} b\, d\mu_{\mathbb{F}_2^n} = \frac{1}{2^n}\sum_{\mathbf{x}\in\mathbb{F}_2^n} 2^n f(\mathbf{x}) = 1 = \mu_{\mathbb{F}_2^n/\mathcal{C}}(\mathbb{F}_2^n/\mathcal{C})$$

where we used that f is a distribution. Therefore we can apply Corollary 1 with functions a and b. Furthermore, $b^{|\mathcal{C}} = 2^n f^{|\mathcal{C}} = 2^{n-k} f^{\mathcal{C}}$ by definition of $f^{\mathcal{C}}$. We get the following computation,

$$\begin{aligned}
\|a - b^{|\mathcal{C}}\|_1 &= \|a - 2^{n-k} f^{\mathcal{C}}\| \\
&= \sum_{\mathbf{x}\in\mathbb{F}_2^n/\mathcal{C}} \left|1 - 2^{n-k} f^{\mathcal{C}}(\mathbf{x})\right| \; \frac{1}{2^{n-k}} \\
&= \sum_{\mathbf{x}\in\mathbb{F}_2^n/\mathcal{C}} \left|\frac{1}{2^{n-k}} - f^{\mathcal{C}}(\mathbf{x})\right| \\
&= 2\,\Delta(u, f^{\mathcal{C}}) \qquad (5)
\end{aligned}$$

To conclude the proof it remains to apply Corollary 1 with Equation (5) and then to use that f is radial and therefore also $\widehat{f}$.

□

Our upper-bound of Proposition 2 involves the weight distribution of the code $\mathcal{C}^*$, namely $(N_t(\mathcal{C}^*))_{t\geq d_{\min}(\mathcal{C}^*)}$. To understand how our bound behaves for a given distribution f, we will start (in the following subsection) with the case of random codes. The expected value for N_t is well known in this case. This will lead us to estimate our bound on almost all codes and gives us some hints about the best distribution to choose for our smoothing bound in the worst case (which is the case that we treat in Subsection 3.2).

3.1. **Smoothing Random Codes.** The probabilistic model $\mathcal{C}_{n,k}$ that we use for our random code of length n is defined by sampling uniformly at random a generator matrix $\mathbf{G} \in \mathbb{F}_2^{k\times n}$ for it, *i.e.*,

$$\mathcal{C} = \left\{\mathbf{mG} : \mathbf{m} \in \mathbb{F}_2^k\right\}.$$

It is straightforward to check the expected number of codewords of weight t in the dual $\mathcal{C}^*$ is given by

Fact 1. *For $\mathcal{C}$ chosen according to $\mathcal{C}_{n,k}$:*

$$\mathbb{E}_{\mathcal{C}}(N_t(\mathcal{C}^*)) = \frac{\binom{n}{t}}{2^k}.$$

This estimation combined with Proposition 2 enables us to bound $\mathbb{E}_{\mathcal{C}}\left(\Delta(u, f^{\mathcal{C}})\right)$:

Proposition 3. *We have:*

$$\mathbb{E}_{\mathcal{C}}\left(\Delta(u, f^{\mathcal{C}})\right) \leq 2^n \sqrt{\sum_{t>0} \frac{\binom{n}{t}}{2^k} |\widehat{f}(t)|^2}. \tag{6}$$

Proof. By using Proposition 2, we obtain:

$$\begin{aligned}
\mathbb{E}_{\mathcal{C}}\left(\Delta(u, f^{\mathcal{C}})\right) &\leq \mathbb{E}_{\mathcal{C}}\left(2^n \sqrt{\sum_{t=d_{\min}(\mathcal{C}^*)}^{n} N_t(\mathcal{C}^*)|\widehat{f}(t)|^2}\right) \\
&\leq 2^n \sqrt{\mathbb{E}_{\mathcal{C}}\left(\sum_{t=d_{\min}(\mathcal{C}^*)}^{n} N_t(\mathcal{C}^*)|\widehat{f}(t)|^2\right)} \quad \text{(Jensen's inequality)} \\
&= 2^n \sqrt{\sum_{t>0} \frac{\binom{n}{t}}{2^k} |\widehat{f}(t)|^2}
\end{aligned}$$

where in the last line we used the linearity of the expectation and Fact 1. □

It remains now to choose the distribution f. A natural choice in code-based cryptography is the uniform distribution u_w over a sphere $\mathscr{S}_w$ of radius w centered around 0.

Uniform Distribution over a Sphere. The Fourier transform of the uniform distribution u_w over the sphere $\mathscr{S}_w$ is intimately connected to Krawtchouk polynomials. The Krawtchouk polynomial of order n and degree $w \in \{0, \ldots, n\}$ is defined as:

$$K_w(X; n) \stackrel{\text{def}}{=} \sum_{j=0}^{w} (-1)^j \binom{X}{j} \binom{n-X}{w-j}$$

To simplify notation, since n is clear here from context, we will drop the dependency on n and simply write $K_w(X)$. The following fact allows to relate K_w with $\widehat{u_w}$ (see for instance [vL99, Lem. 3.5.1, §3.5])

Fact 2. *For any* $\mathbf{y} \in \mathscr{S}_t$,

$$\sum_{\mathbf{e} \in \mathscr{S}_w} (-1)^{\mathbf{y}\cdot\mathbf{e}} = K_w(t). \tag{7}$$

This leads us to

$$\widehat{u_w}(\mathbf{x}) = \frac{1}{2^n} K_w(|\mathbf{x}|) \Big/ \binom{n}{w}.$$

By plugging this in Equation (6) of Proposition 3 we obtain

$$\mathbb{E}_{\mathcal{C}}\left(\Delta(u, u_w^{\mathcal{C}})\right) \leq \sqrt{\sum_{t>0} \frac{\binom{n}{t}}{2^k} \left(\frac{K_w(t)}{\binom{n}{w}}\right)^2}. \tag{8}$$

The above sum can be upper-bounded by observing that $\left(K_w/\sqrt{\binom{n}{w}}\right)_{0\leq w\leq n}$ is an orthonormal basis of functions $f : \{0, 1, \cdots, n\} \to \mathbb{C}$ for the inner product $\langle f, g\rangle_{\text{rad}} \stackrel{\text{def}}{=} \sum_{t=0}^{n} f(t)\overline{g(t)}\binom{n}{t}/2^n$. It can be viewed in a natural as the standard inner product between radial functions over $\mathbb{F}_2^n$. In particular, $\sum_{t=0}^{n} \frac{K_w(t)^2}{\binom{n}{w}} \frac{\binom{n}{t}}{2^n} = 1$ [Lev95, Corollary 2.3]. Therefore, for random codes we obtain the following proposition:

Proposition 4. *We have for random $\mathcal{C}$ chosen according to $\mathcal{C}_{n,k}$:*

$$\mathbb{E}_{\mathcal{C}}\left(\Delta(u, u_w^{\mathcal{C}})\right) \leq \sqrt{2^{n-k} \Big/ \binom{n}{w}}. \tag{9}$$

In other words, if one wants to smooth a random code with target distance $2^{-\Omega(n)}$ via the uniform distribution over a sphere, one has to choose its radius $w \leq n/2$ bigger than some w_0 such that $\binom{n}{w_0} = 2^{-\Omega(n)}\, 2^{n-k}$. It is readily seen that for fixed code rate $R \stackrel{\text{def}}{=} \frac{k}{n}$, choosing $\omega_0 \stackrel{\text{def}}{=} \frac{w_0}{n}$ such that $\omega_0 < \omega_{\mathrm{GV}}(R)$ where $\omega_{\mathrm{GV}}(R)$ corresponds to the asymptotic relative Gilbert-Varshamov (GV) bound

$$\omega_{\mathrm{GV}}(R) \stackrel{\text{def}}{=} h^{-1}(1-R)\ ,$$

with $h^{-1} : [0,1] \to [0,1/2]$ being the inverse of the binary entropy function $h(p) = -p\log_2(p) - (1-p)\log_2(1-p)$. The GV bound $\omega_{\mathrm{GV}}(R)$ appears ubiquitously in the coding-theoretic literature: amongst other contexts, it arises as the (expected) relative minimum distance of a random code of dimension Rn, or as the maximum relative minimum error weight for which decoding can be successful with non vanishing error probability.

This value of radius w_0 is optimal: clearly, the support size of an error distribution smoothing a code $\mathcal{C}$ must exceed $\#\mathbb{F}_2^n/\mathcal{C}$. Thus, we cannot expect to smooth a code $\mathcal{C}$ with errors in the sphere $\mathscr{S}_{w_0}$ if its volume is smaller than $2^{n-k} = \#\mathbb{F}_2^n/\mathcal{C}$.

Therefore the uniform distribution over a sphere is optimal for random codes. By this, we mean that it leads to the smallest amount of possible noise (when it is concentrated on a ball) to smooth a random code. However notice that we obtain such result after applying the chain of arguments Cauchy-Schwarz, Parseval and Poisson to bound the statistical distance.

About the original chain of arguments of Micciancio and Regev. It can be verified that by coming back to the original steps of [MR07, ADRS15], namely the Poisson summation formula and then the triangle inequality, we would obtain:

$$\Delta\left(u, f^{\mathcal{C}}\right) \leq 2^n \sum_{t \geq d_{\min}(\mathcal{C}^*)} N_t(\mathcal{C}^*)|\widehat{f}(t)|. \tag{10}$$

By using that $a^2 + b^2 \leq (a+b)^2$ (when $a, b \geq 0$) we see that our bound (Proposition 2) is sharper. It turns out that our bound is exponentially sharper for random codes (and even in the worst case) when choosing f as the uniform distribution over a sphere of radius w, namely $f = u_w$. In this case the Micciancio-Regev argument yields the following computation:

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}\left(\Delta\left(u, u_w^{\mathcal{C}}\right)\right) &\leq \mathbb{E}_{\mathcal{C}}\left(\sum_{t \geq d_{\min}(\mathcal{C}^*)} N_t(\mathcal{C}^*)\ \frac{|K_w(t)|}{\binom{n}{w}}\right) \\ &= \sum_{t>0} \frac{\binom{n}{t}}{2^k}\ \frac{|K_w(t)|}{\binom{n}{w}}.\end{aligned} \tag{11}$$

To carefully estimate this upper-bound (and to compare with (9)) we are going to use the following proposition, which gives the asymptotic behaviour of K_w (see for instance [IS98, DT17]).

Proposition 5. *Let n, t and w be three positive integers. We set $\tau \stackrel{def}{=} \frac{t}{n}$, $\omega = \frac{w}{n}$ and $\omega^{\perp} \stackrel{def}{=} 1/2 - \sqrt{\omega(1-\omega)}$. We assume $w \leq n/2$. Let $z \stackrel{def}{=} \frac{1-2\tau-\sqrt{D}}{2(1-\omega)}$ where $D \stackrel{def}{=} (1-2\tau)^2 - 4\omega(1-\omega)$. In the case $\tau \in (0, \omega^{\perp})$,*

$$K_w(t) = O\left(2^{n(a(\tau,\omega)+o(1))}\right) \quad \textit{where} \quad a(\tau,\omega) \stackrel{def}{=} \tau\log_2(1-z) + (1-\tau)\log_2(1+z) - \omega\log_2 z.$$

In the case $\tau \in (\omega^\perp, 1/2)$, *$D$ is negative, and*

$$K_w(t) = O\left(2^{n(a(\tau,\omega)+o(1))}\right) \quad \textit{where} \quad a(\tau,\omega) \stackrel{def}{=} \frac{1}{2}(1 + h(\omega) - h(\tau)).$$

We let,

$$\omega_0 \stackrel{\text{def}}{=} \overline{\lim_{n\to\infty}} \left\{ \frac{w}{n} : \sqrt{2^{n(1-R)} \Big/ \binom{n}{w}} \geq 1 \right\},$$

$$\omega_1 \stackrel{\text{def}}{=} \overline{\lim_{n\to\infty}} \left\{ \frac{w}{n} : \sum_{t>0} \frac{\binom{n}{t}}{2^{Rn}} \frac{|K_w(t)|}{\binom{n}{w}} \geq 1 \right\}.$$

In Figure 4 we compare the asymptotic values of ω_0 and ω_1 as functions of R. Notice that $\omega_0 = \omega_{\mathrm{GV}}(R)$. We see that ω_1 is undefined for a rate $R < 1/2$. In other words, it is impossible to reach $\mathbb{E}_{\mathcal{C}}\left(\Delta(u, u_w^{\mathcal{C}})\right) \leq 2^{-\Omega(n)}$ with the standard approach of [MR07, ADRS15]. Furthermore, for larger rates (and sufficiently large n), ω_0 is much smaller than ω_1.

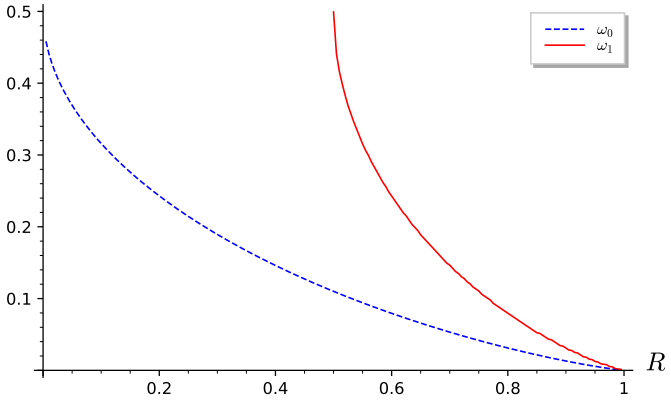

FIGURE 4. ω_0 and ω_1 as functions of $R \stackrel{\text{def}}{=} \frac{k}{n}$.

Bernoulli Distribution. Another natural distribution to consider when dealing with codes is the so-called "Bernoulli" distribution $f_{\mathrm{ber},p}$, which is defined for $p \in [0, 1/2]$ as

$$\forall \mathbf{x} \in \mathbb{F}_2^n, \quad f_{\mathrm{ber},p}(\mathbf{x}) \stackrel{\text{def}}{=} p^{|\mathbf{x}|}(1-p)^{n-|\mathbf{x}|}.$$

This choice leads to simpler computations compared to the uniform distribution over a sphere. For instance we have $\widehat{f_{\mathrm{ber},p}}(\mathbf{x}) = \frac{1}{2^n}(1-2p)^{|\mathbf{x}|}$. By plugging this in Equation (6) of Proposition 3 we obtain

$$\begin{aligned} \mathbb{E}_{\mathcal{C}}\left(\Delta(u, f_{\mathrm{ber},p}^{\mathcal{C}})\right) &\leq \sqrt{\sum_{t>0} \frac{\binom{n}{t}}{2^k}(1-2p)^{2t}} \\ &\leq \sqrt{\frac{1}{2^k}\left(1 + (1-2p)^2\right)^n} \end{aligned} \tag{12}$$

Thus, if one wants to smooth a random code at target distance $2^{-\Omega(n)}$ with the Bernoulli distribution, the above argument says that one has to choose $p > p_0 \stackrel{\text{def}}{=} \frac{1}{2}\left(1 - \sqrt{2^R - 1}\right)$ where $R = k/n$. As $\mathbb{E}_{f_{\mathrm{ber},p}}(|\mathbf{x}|) = pn$, it is meaningful to compare p_0 and ω_0. It is readily seen that

$\omega_0 = \omega_{\mathrm{GV}}(R) = h^{-1}(1-R) < \frac{1}{2}\left(1-\sqrt{2^R-1}\right) = p_0$. In other words, this time the upper-bound given by Proposition 3 does not give what would be optimal, namely the Gilbert-Varshamov relative distance $\omega_{\mathrm{GV}}(R)$ but a quantity which is bigger. However, it is expected that the average amount of noise to smooth a random code is the same in both cases, since a Bernoulli distribution of parameter p is extremely concentrated over words of Hamming weight pn and that therefore $\Delta(u, f^{\mathcal{C}}_{\mathrm{ber},p}) \approx \Delta(u, u^{\mathcal{C}}_{pn})$. This suggests that Proposition 3 is not tight in this case. This is indeed the case, we can indeed prove that we can smooth a random with the Bernoulli noise as soon as $p > \omega_{\mathrm{GV}}(R)$. This follows from the following proposition.

Proposition 6. *Let $\varepsilon > 0$ and $p \in [0, 1/2]$. Then,*

$$\Delta(u, f^{\mathcal{C}}_{\mathrm{ber},p}) \leq \sum_{r=(1-\varepsilon)np}^{(1+\varepsilon)np} \Delta(u, u^{\mathcal{C}}_r) + 2^{-\Omega(n)}.$$

Proof. See Appendix A. □

This proposition shows that if one wants $\Delta(u, f^{\mathcal{C}}_{\mathrm{ber},p}) \leq 2^{-\Omega(n)}$ it is enough to have $\Delta(u, f^{\mathcal{C}}_{\mathrm{unif},r}) \leq 2^{-\Omega(n)}$ for any $r \in [(1-\varepsilon)np, (1+\varepsilon)np]$. This can be achieved by choosing ε and p such that $(1-\varepsilon)p > \omega_{\mathrm{GV}}(R)$.

To summarize this subsection we have the following theorem:

Theorem 3. *Let $\mathcal{C}$ be a random code chosen according to $\mathcal{C}_{n,k}$, $R \stackrel{def}{=} \frac{k}{n}$. Let u (resp. $u_{\lceil pn \rceil}$) be the uniform distribution over $\mathbb{F}_2^n/\mathcal{C}$ (resp. $\mathscr{S}_w$) and $f_{\mathrm{ber},p}$ be the Bernoulli distribution over $\mathbb{F}_2^n$ of parameter p. We have,*

$$\mathbb{E}_{\mathcal{C}}\left(\Delta(u, u^{\mathcal{C}}_{\lceil pn \rceil})\right) \leq 2^{\frac{n}{2}(1-R-h(p)+o(1))} \quad \textit{and} \quad \mathbb{E}_{\mathcal{C}}\left(\Delta(u, f^{\mathcal{C}}_{\mathrm{ber},p})\right) \leq 2^{\frac{n}{2}(1-R-h(p)+o(1))}.$$

3.2. **Smoothing a Fixed Code.** Our upper-bound on $\Delta(u, f^{\mathcal{C}})$ given in Proposition 2 involves the weight distribution of the dual of $\mathcal{C}$, namely the $N_t(\mathcal{C}^*)$'s. To derive smoothing bounds on a fixed code our strategy will simply consist in using the upper-bounds on the $N_t(\mathcal{C}^*)$'s that follow. Roughly speaking, these bounds show that $N_t(\mathcal{C}^*) \leq \binom{n}{t} 2^{-Kn}$ for some constant K which is function of $d_{\min}(\mathcal{C}^*)$.

Notation. Let $\delta \in (0, 1/2)$ and $\delta \leq \tau \leq 1$,

$$b(\delta, \tau) \stackrel{\mathrm{def}}{=} \overline{\lim_{n\to\infty}} \max_{\mathcal{C}} \left\{ \frac{1}{n} \log_2 N_{\lfloor \tau n \rfloor}(\mathcal{C}) \right\} \tag{13}$$

where the maximum is taken over all codes $\mathcal{C}$ of length n and minimum distance $\geq \delta n$.

We recall (or slightly extend) results taken from [ABL01]:

Proposition 7. *Let $\delta \in (0, 1/2)$ and $\delta^\perp \stackrel{def}{=} 1/2 - \sqrt{\delta(1-\delta)}$. For any $\delta \leq \tau \leq 1$:*

$$b(\delta, \tau) \leq c(\delta, \tau) \stackrel{def}{=} \begin{cases} h(\tau) + h\left(\delta^\perp\right) - 1 & \textit{if } \tau \in [\delta, 1-\delta] \\ 2\left(h(\delta^\perp) - a(\tau, \delta^\perp)\right) & \textit{otherwise.} \end{cases} \tag{14}$$

where $a(\cdot,\cdot)$ is defined in Proposition 5.

Proof. See Appendix B □

Proposition 8 ([ABL01, Proposition 4])**.** *Let $\delta_{\mathrm{JSB}} \stackrel{def}{=} \left(1-\sqrt{1-2\delta}\right)/2$ and*

$$\tau_0 \stackrel{def}{=} \operatorname*{argmin}_{\delta_{\mathrm{JSB}} \leq \alpha \leq 1/2} 1 - h(\alpha) + R_1(\alpha, \delta)$$

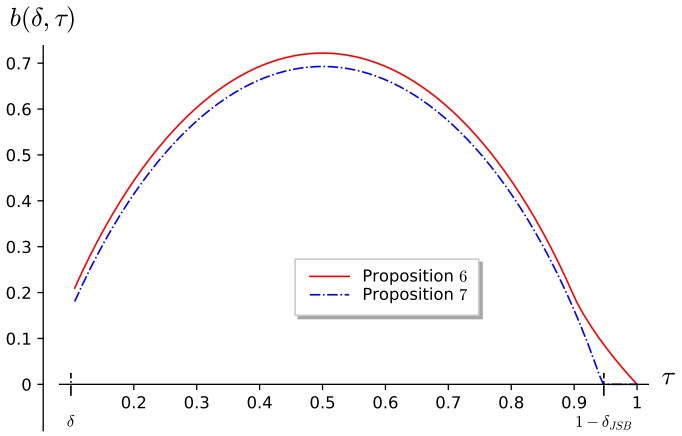

FIGURE 5. Bounds of Propositions 7 and 8 on $b(\delta,\tau)$ as function of $\tau \in [\delta, 1]$ for $\delta = 0.1$.

where

$$R_1(\tau,\delta) \stackrel{def}{=} h\left(\frac{1}{2}\left(1-\sqrt{1-\left(\sqrt{4\tau(1-\tau)-\delta(2-\delta)}-\delta\right)^2}\right)\right)$$

For any $\delta \leq \tau \leq 1$*:*

$$b(\delta,\tau) \leq d(\delta,\tau) \stackrel{def}{=} \begin{cases} h(\tau)-h(\tau_0)+R_1(\tau_0,\delta) & \text{if } \tau \in (\delta_{\text{JSB}}, 1-\delta_{\text{JSB}}) \text{ and } \tau_0 \leq \tau \\ R_1(\tau,\delta) & \text{if } \tau \in (\delta_{\text{JSB}}, 1-\delta_{\text{JSB}}) \text{ and } \tau_0 > \tau \\ 0 & \text{otherwise.} \end{cases} \quad (15)$$

Both of these bounds are derived from "linear programming arguments" which were initially used to upper-bound the size of a code given its minimum distance. Proposition 7 is an extension of [ABL01, Theorem 3] in the case of linear codes, in particular we give an upper-bound for any $\tau \in [\delta, 1]$ (and not for only $\tau \in [\delta, 1/2]$). The proof is in appendix. The second bound is usually called the *the second linear programming bound*. According to δ and τ, Proposition 7 and 8 are among the best (known) upper-bounds on $b(\delta,\tau)$. In the case where $0 \leq \delta \leq 0.273$, Proposition 8 is leading to better smoothing bounds compared to Proposition 7.

Remark 1. *There exist many other bounds on* $b(\delta,\tau)$*, like* [ACKL05, Theorem 8] *which holds only for linear codes or* [ACKL05, Theorem 7]*. However for our smoothing bounds, Propositions 7 and 8 lead to the best results, partly because these are the best bounds on the number of codewords of Hamming weight close to the minimum distance of the code.*

We draw in Figures 5 and 6 the bounds of Propositions 7 and 8 as function of $\tau \in [\delta, 1]$ according to some values of δ.

Equipped with these bounds we are ready to give our smoothing bounds for codes in the worst case, namely for a fixed code. Our study with random codes gave a hint that the choice of the uniform distribution over a sphere could give better results than the Bernoulli distribution. However, as we will show now, the distribution on a sphere is forcing us to suppose that no codewords of large weights belong to the dual $\mathcal{C}^*$ when wanting to smooth $\mathcal{C}$. It corresponds to the hypothesis of balanced-codes made in [BLVW19] to obtain a worst-to-average case reduction. We would like to avoid making this assumption as nothing forbids large weight vectors to belong to

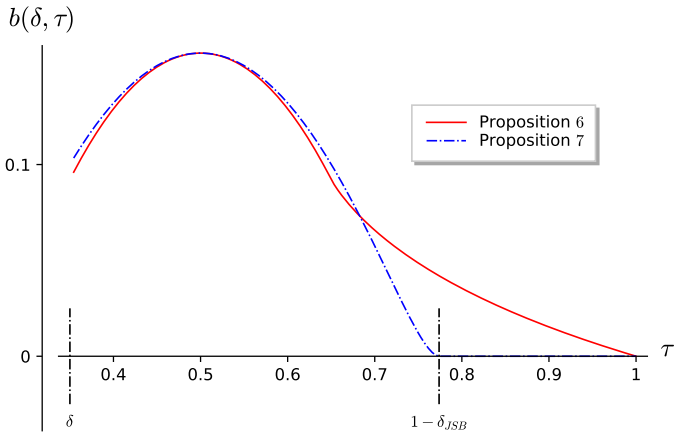

FIGURE 6. Bounds of Propositions 7 and 8 on $b(\delta, \tau)$ as function of $\tau \in [\delta, 1]$ for $\delta = 0.35$.

some fixed code. Fortunately we can avoid making this hypothesis by still keeping the advantages of the distribution over a sphere as we will show later.

Impossibility to smooth with the uniform distribution over a sphere a code whose dual is not balanced. It is readily seen that in the case where the dual code $\mathcal{C}^*$ is not balanced, meaning that it contains the all-one vector (and therefore that the dual weight distribution is symmetric: $N_w(\mathcal{C}^*) = N_{n-w}(\mathcal{C}^*)$ for any $w \in \{0, \cdots, n\}$ when the codelength is n), then it is impossible to smooth it with the uniform distribution u_w over a sphere. Indeed, this implies that all codewords of $\mathcal{C}$ have an even Hamming weight. The parity of the Hamming weights of vectors in a coset (*i.e* in the class of representatives of some element in $\mathbb{F}_2^n/\mathcal{C}$) will be the same. Therefore, half of the cosets cannot be reached when periodizing u_w over $\mathcal{C}$.

Difficulty of using Proposition 2 for proving smoothness of the uniform distribution if the dual has large weight codewords. Even in the case where the dual is balanced, difficulties can arise if we want to use Proposition 2 for proving smoothness of the uniform distribution over a sphere when the dual has large weight codewords. First of all, the fact that it contains the all one codeword also reflects in the upper-bound of Proposition 2. Recall that $\widehat{u_w}(\mathbf{x}) = \frac{1}{2^n} K_w(|\mathbf{x}|)/\binom{n}{w}$ and that we have $K_w(n) = (-1)^w \binom{n}{w}$ (see Fact 2). Therefore, when the full weight vector belongs to $\mathcal{C}^*$, our upper-bound on $\Delta(u, u_w^{\mathcal{C}})$ of Proposition 2 cannot be smaller than 1. Furthermore, even if the dual does not contain the all-one codeword, codewords of weight say $t = n - O(\log n)$ have also a terrible contribution to the upper-bound of Proposition 2 since they give a polynomial $n^{-O(1)}$ contribution to it.

Difficulty of using Proposition 2 for proving smoothness of the "discrete walk distribution" if the dual has large weight codewords. Other meaningful distributions in the cryptographic context display the same problem as the uniform distribution concerning the difficulty of applying Proposition 2 to them if the dual contains large weight codewords. This is the case of the discrete time random walk distribution $f_{\mathrm{RW},t}$ introduced in [BLVW19] for worst-to-average case reductions. The authors were only able to prove smoothness of this distribution if

the dual code has no small *and no large* codewords. This distribution is given by

$$\forall \mathbf{x} \in \mathbb{F}_2^n, \quad f_{\mathrm{RW},w}(\mathbf{x}) \stackrel{\text{def}}{=} \mathbb{P}\left(\sum_{i=1}^{w} \mathbf{e}_{u_i} = \mathbf{x}\right)$$

where the u_i's are independently and uniformly drawn at random in $\{1, \ldots, n\}$ and $\mathbf{e}_j$ denotes the j-th canonical basis vector. Recall that [BLVW19]

$$\widehat{f_{\mathrm{RW},w}}(\mathbf{y}) = \frac{1}{2^n}\left(1 - 2\frac{|\mathbf{y}|}{n}\right)^w.$$

Therefore, $\widehat{f_{\mathrm{RW},w}}(\mathbf{y}) = \frac{1}{2^n}(-1)^w$ when $|\mathbf{y}| = n$, as for the Fourier transform of the uniform distribution over a sphere, showing that $f_{\mathrm{RW},w}$ cannot smooth a code when the full weight vector belongs to its dual. In summary, *a direct application* of Proposition 2 is quite unsatisfactory for these distributions u_w and $f_{\mathrm{RW},w}$. If we accept to also make an assumption on the largest weight of the codeword, then of course a direct application of Proposition 2 is able to provide meaningful smoothing bounds for them as shown in the following theorem (obtained by just combining Propositions 2, 7 and 8).

Theorem 4. *Let $\mathcal{C}$ be a binary linear code of length n and $\omega \in (0,1)$. Suppose that $d_{\min}(\mathcal{C}^*) = \delta^* n$ and that $\mathcal{C}^*$ has no element of Hamming weight $\geq \beta n$ for some $\beta \in (\delta^*, 1)$. We have:*

$$\frac{1}{n}\log_2 \Delta\left(u, u^{\mathcal{C}}_{\omega n}\right) \leq \max_{\delta^* \leq \tau \leq \beta}\left\{\frac{1}{2}\min\left\{c(\delta^*,\tau), d(\delta^*,\tau)\right\} + a(\omega,\tau)\right\} - h(\omega)$$

$$\frac{1}{n}\log_2 \Delta\left(u, f^{\mathcal{C}}_{\mathrm{RW},\omega n}\right) \leq \max_{\delta^* \leq \tau \leq \beta}\left\{\frac{1}{2}\min\left\{c(\delta^*,\tau), d(\delta^*,\tau)\right\} + \omega\log_2(1-2\tau)\right\}$$

where $a(\cdot,\cdot)$, $c(\cdot,\cdot)$ and $d(\cdot,\cdot)$ are defined respectively in Propositions 5, 7 and 8.

Avoiding making an assumption on the largest dual codeword: the case of the Bernoulli distribution. Even if the Bernoulli distribution has some drawbacks when compared to the uniform distribution over a sphere concerning the direct use of Proposition 2 on it as we saw in the case of random codes, it has however a nice property concerning the large weight codewords: the large weight dual codewords have a negligible contribution in the upper-bound of Proposition 2. To see this let us first recall that:

$$\widehat{f_{\mathrm{ber},p}}(\mathbf{x}) = \frac{1}{2^n}(1-2p)^{|\mathbf{x}|}. \tag{16}$$

Therefore, by Proposition 2 we have:

$$\Delta(u, f^{\mathcal{C}}_{\mathrm{ber},p}) \leq \sqrt{\sum_{t=d_{\min}(\mathcal{C}^*)}^{n} N_t(\mathcal{C}^*)(1-2p)^{2t}} \tag{17}$$

On the other hand, we have the following lemma which shows that large weight codewords can only have an exponentialy small contribution to the above upper-bound.

Lemma 1. *Let $\mathcal{C}$ be a linear code of length n and let $t > n - d_{\min}(\mathcal{C})/2$. There is at most one codeword $\mathbf{c}$ of weight t.*

Proof. Suppose by contradiction that there exists two distinct codewords $\mathbf{c}, \mathbf{c}' \in \mathcal{C}$ of Hamming weight t. By using the triangle inequality we obtain (where $\mathbf{1}$ denotes the all one vector)

$$\begin{aligned} |\mathbf{c} - \mathbf{c}'| &\leq |\mathbf{c} - \mathbf{1}| + |\mathbf{1} - \mathbf{c}'| \\ &= 2(n-t) \\ &< d_{\min}(\mathcal{C}) \end{aligned}$$

which contradicts the fact that $\mathcal{C}$ has minimum distance $d_{\min}(\mathcal{C})$. □

Therefore, using Lemma 1 in Equation (17) gives for $p \in (0, 1/2]$,

$$\Delta(u, f^{\mathcal{C}}_{\mathrm{ber},p}) \leq \sqrt{\sum_{t=d_{\min}(\mathcal{C}^*)}^{n-d_{\min}(\mathcal{C}^*)/2} N_t(\mathcal{C}^*)(1-2p)^{2t} + 2^{-\Omega(n)}}. \tag{18}$$

In other words, large weight dual codewords (if they exist) have only an exponentially small contribution to our smoothing bound with the Bernoulli distribution. In principle, we could plug in Equation (18) bounds on the $N_t(\mathcal{C}^*)$'s given in Propositions 7 and 8. We will improve on the bounds obtained in this way by truncating the Bernoulli distribution, then

(i) prove that by appropriately truncating both distributions have the same smoothness property,
(ii) show that the truncated distribution has the same nice properties with respect to large weights,
(iii) show that we can apply Proposition 2 to the truncated distribution and get appropriate smoothness properties.

We obtain in this way

Theorem 5. *Let $\mathcal{C}$ be a binary linear code of length n and $p \in (0, 1/2]$ such that $d_{\min}(\mathcal{C}^*) \geq \delta^* n$ for some $\delta^* \in [0, 1]$. We have asymptotically,*

$$\frac{1}{n}\log_2 \Delta\left(u, f^{\mathcal{C}}_{\mathrm{ber},p}\right) \leq \max_{\delta^* \leq \tau \leq 1-\delta^*/2} \{\frac{1}{2}\min\{c(\delta^*, \tau), d(\delta^*, \tau)\} + \max_{(1-\varepsilon)p \leq \lambda \leq (1+\varepsilon)p} \{\lambda \log_2 p + (1-\lambda)\log_2(1-p) + a(\lambda, \tau)\}\} + O\left(\frac{1}{n}\right)$$

where $a(\cdot,\cdot)$, $c(\cdot,\cdot)$ and $d(\cdot,\cdot)$ are defined respectively in Propositions 5, 7 and 8.

Proof. See Appendix C. □

Let $i \in \{0, 1\}$ and p_i be the smallest $p \in (0, 1/2]$ that enables to reach $\Delta\left(u, f^{\mathcal{C}}_{\mathrm{ber},p}\right) \leq 2^{-\Omega(n)}$ with

- Theorem 5 when $i = 0$,
- Equation (18) and Propositions 7, 8 when $i = 1$.

In Figure 7 we compare the smallest p that enables to reach $\Delta\left(u, f^{\mathcal{C}}_{\mathrm{ber},p}\right) \leq 2^{-\Omega(n)}$ with Equation (18) and thanks to Theorem 5. As we can see Theorem 5 leads to significantly better bounds. Furthermore, it turns out that $p_0 n$ is roughly equal to the smallest radius w such that $\Delta(u, u^{\mathcal{C}}_w) \leq 2^{-\Omega(n)}$ if we had supposed that no codewords of weight $> n - d_{\min}(\mathcal{C}^*)$ belong to $\mathcal{C}^*$. In other words, our proof using the tweak of truncating the Bernoulli enables to obtain a smoothing bound without the hypothesis of no dual codewords of large Hamming weight but as good as with the uniform distribution over a sphere if we had made this assumption.

4. Smoothing Bounds: Lattice Case

Given an n-dimensional lattice Λ the aim of smoothing bounds is to give a non-trivial model of noise $\mathbf{e} \in \mathbb{R}^n$ for $(\mathbf{e} \mod \Lambda) \in \mathbb{R}^n/\Lambda$ (namely the reduction of $\mathbf{e}$ modulo Λ) to be uniformly distributed. Followwing Micciancio and Regev [MR07], the standard choice of noise is given by the Gaussian distribution, defined via

$$\forall \mathbf{x} \in \mathbb{R}^n, \quad D_s(\mathbf{x}) \stackrel{\text{def}}{=} \frac{1}{s^n}\rho_s(\mathbf{x}) \quad \text{where} \quad \rho_s(\mathbf{x}) \stackrel{\text{def}}{=} e^{-\pi(|\mathbf{x}|_2/s)^2}.$$

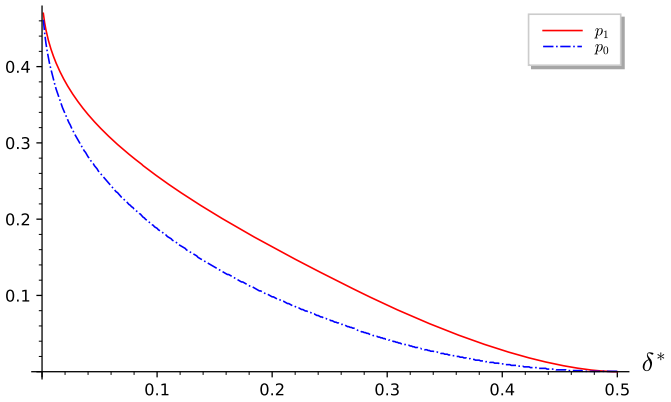

FIGURE 7. Smoothing bounds for a code $\mathcal{C}$ as function of $\delta^* \stackrel{\text{def}}{=} d_{\min}(\mathcal{C}^*)/n$ via Theorem 5 (for $\varepsilon = 10^{-2}$) and Equation (18) .

The parametrization is chosen such that $s/\sqrt{2\pi}$ is the standard deviation of D_s. Micciancio and Regev showed that when $\mathbf{e}$ is distributed according to D_s, choosing s large enough enables $\mathbf{e} \bmod \Lambda$ to be statistically closed to the uniform distribution.

However, following the intuitition from the case of codes we will first analyze the case where $\mathbf{e}$ is sampled uniformly from a ball. Interestingly, just as with codes where our methodology led to stronger bounds when the uniform distribution over a sphere was used to smooth than the Bernoulli distribution, we will obtain better results when we work with the uniform distribution over a ball. Fortunately, using concentration of the Gaussian measure one can translate results from the case where $\mathbf{e}$ is uniformly distributed over a ball to the case that it is sampled according to D_s; see Proposition 13. This is analogous to the translation from results for the uniform distribution over a sphere to the Bernoulli distribution for codes elucidated in Proposition 6.

For either choice of noise, to obtain a smoothing bound we are required to bound the statistical distance between the distribution of $\mathbf{e} \bmod \Lambda$ if $\mathbf{e}$ has density g, and the uniform distribution over $\mathbb{R}^n/\Lambda$. It is readily seen that $\mathbf{e} \bmod \Lambda$ has density $|\Lambda| g^{|\Lambda}$ which is defined as (see Definition 1 with the choice of Haar measures given in Table 2)

$$g^{|\Lambda}(\mathbf{x}) = \frac{1}{|\Lambda|} \sum_{\mathbf{y} \in \Lambda} g(\mathbf{x} + \mathbf{y}).$$

Notation. For any $g : \mathbb{R}^n \to \mathbb{C}$,

$$g^{\Lambda} \stackrel{\text{def}}{=} |\Lambda| \, g^{|\Lambda}.$$

In the following proposition we specialize Corollary 1 to the case of lattices.

Proposition 9. *Let Λ be an n-dimensional lattice. Let g be some density function on $\mathbb{R}^n$ and v be the density of the uniform distribution over $\mathbb{R}^n/\Lambda$. We have*

$$\Delta\left(v, g^{\Lambda}\right) \leq \frac{1}{2} \sqrt{\sum_{\mathbf{x} \in \Lambda^* \setminus \{\mathbf{0}\}} |\widehat{g}(\mathbf{x})|^2} \ .$$

We will restrict our instantiations to functions g whose Fourier transforms are radial, namely $\widehat{g}(\mathbf{x})$ depends only on the Euclidean norm of $\mathbf{x}$.

4.1. **Smoothing Random Lattices.** As with codes, we begin our investigation of smoothing lattices by considering the random case. However, defining a "random lattice" is much more involved than the analogous notion of random codes. Fortunately for us, we can apply the Siegel version of the Minkowski-Hlawka theorem to conclude that there exists a random lattice model which behaves very nicely from the perspective of "test functions". We first state the technical theorem that we require.

Theorem 6 (Minkowski-Hlawka-Siegel)**.** *On the set of all the lattices of covolume M in $\mathbb{R}^n$ there exists a probability measure μ such that, for any Riemann integrable function $g(\mathbf{x})$ which vanishes outside some bounded region,*

$$\mathop{\mathbb{E}}_{\Lambda\sim\mu}\left(\sum_{\mathbf{x}\in\Lambda\setminus\{\mathbf{0}\}} g(\mathbf{x})\right) = \frac{1}{M}\int_{\mathbb{R}^n} g(\mathbf{x})d\mathbf{x}\ .$$

As intuition for the above theorem, consider the case that g is the indicator function for a bounded, measurable subset $S \subseteq \mathbb{R}^n$. Then, Theorem 6 promises that the expected number of lattice points (other than the origin[(6)]) in S is equal to the volume of S over the covolume of the lattice.

Uniform Distribution over a Ball. Let

$$u_{w\mathscr{B}} \stackrel{\text{def}}{=} \frac{1_{\mathscr{B}_w}}{V_n(w)}$$

be the density of the uniform distribution over the Euclidean ball of radius w. Let us recall that $V_n(w)$ denotes the volume of any ball of radius w. From Theorem 6, we may obtain the following proposition. This should be compared with Proposition 4.

Proposition 10. *On the set of all lattices of covolume M in $\mathbb{R}^n$ there exists a probability measure ν such that, for any $w > 0$:*

$$\mathop{\mathbb{E}}_{\Lambda\sim\nu}\left(\Delta(u, u^{\Lambda}_{w\mathscr{B}})\right) \le \frac{1}{2}\sqrt{\frac{M}{V_n(w)}}.$$

In particular, defining

$$w_0 \stackrel{def}{=} \sqrt{n/2\pi e}M^{1/n},$$

if $w > w_0$ we have

$$\mathop{\mathbb{E}}_{\Lambda\sim\nu}\left(\Delta(u, u^{\Lambda}_{w\mathscr{B}})\right) \le O(1)\ \left(\frac{w_0}{w}\right)^{n/2}.$$

Proof. We define ν to be the procedure that samples a lattice according to μ of covolume M^{-1}, then outputs its dual. In the following chain, we first apply Proposition 9; then, Jensen's inequality; then, the Minkowski-Hlawka-Siegel (MHS) Theorem (Theorem 6) to the function $|\widehat{u^{\Lambda}_{w\mathscr{B}}}|^2$; and,

[(6)] Note that as $\mathbf{0} \in \Lambda$ with certainty, there is really no "randomness" for this event.

lastly, Parseval's Identity (Theorem 1). This yields:

$$\begin{aligned}
\mathop{\mathbb{E}}_{\Lambda\sim\nu}\left(2\Delta(u,u^{\Lambda}_{w\mathscr{B}})\right) &\leq \mathop{\mathbb{E}}_{\Lambda^*\sim\mu}\left(\sqrt{\sum_{\mathbf{x}\in\Lambda^*\setminus\{\mathbf{0}\}}|\widehat{u_{w\mathscr{B}}}(\mathbf{x})|^2}\right) \quad \text{(Proposition 9)}\\
&\leq \sqrt{\mathop{\mathbb{E}}_{\Lambda^*\sim\mu}\left(\sum_{\mathbf{x}\in\Lambda^*\setminus\{\mathbf{0}\}}|\widehat{u_{w\mathscr{B}}}(\mathbf{x})|^2\right)} \quad \text{(Jensen's Inequality)}\\
&= \sqrt{\frac{1}{M^{-1}}\cdot\left(\int_{\mathbb{R}^n}|\widehat{u_{w\mathscr{B}}}(\mathbf{x})|^2 d\mathbf{x}\right)} \quad \text{(MHS Theorem)}\\
&= \sqrt{M\int_{\mathbb{R}^n}|u_{w\mathscr{B}}(\mathbf{x})|^2 d\mathbf{x}} \quad \text{(Parseval's Identity)}\\
&= \sqrt{\frac{M}{V_n(w)^2}\int_{\mathbb{R}^n}1_{\mathscr{B}_w}(\mathbf{x})d\mathbf{x}}\\
&= \sqrt{\frac{M}{V_n(w)}}.
\end{aligned}$$

For the "in particular" part of the proposition, we use Stirling's estimate to derive

$$V_n(w) = \frac{\pi^{n/2}w^n}{\Gamma(n/2+1)} = \frac{\pi^{n/2}w^n}{\left(\frac{n}{2e}\right)^{n/2}}(1+o(1))^n$$

from which it follows that if

$$w > w_0 = \sqrt{n/2\pi e}M^{1/n},$$

we have

$$\sqrt{\frac{M}{V_n(w)}} \leq O(1)\left(\frac{w}{w_0}\right)^{n/2}. \qquad \square$$

It is easily verified that the value of w_0 defined in Proposition 11 corresponds to the so-called Gaussian heuristic. We view this condition on $w > w_0$ as the equivalent of the Gilbert-Varshamov bound for codes as we discussed just below Proposition 4. In particular, as we need the support of the noise to have volume at least M if we hope to smooth a lattice of covolume M, we see that the uniform distribution over a ball is optimal for smoothing random lattices, just as the uniform distribution over a sphere was optimal for smoothing random codes.

Gaussian Noise. We now turn to the case of Gaussian noise. Following the proof of Proposition 10 to the point where we apply Parseval's identity, but replacing $u_{w\mathscr{B}}$ by D_s, we obtain that

$$\mathbb{E}\left(\Delta(u,D^{\Lambda}_s)\right) \leq \frac{1}{2}\sqrt{M\int_{\mathbb{R}^n}|D_s(\mathbf{x})|^2 d\mathbf{x}}\,.$$

To conclude, one uses the following routine computation:

$$\int_{\mathbb{R}^n}|D_s(\mathbf{x})|^2 d\mathbf{x} = \frac{1}{s^{2n}}\int_{\mathbb{R}^n}e^{-2\pi\left(\frac{|\mathbf{x}|_2}{s}\right)^2}d\mathbf{x} = \frac{1}{s^{2n}}\int_{\mathbb{R}^n}\rho_{s/\sqrt{2}}(\mathbf{x})d\mathbf{x} = \left(\frac{1}{s\sqrt{2}}\right)^n\,.$$

Thus, we obtain:

Proposition 11. *On the set of all the lattices of covolume M in $\mathbb{R}^n$ there exists a probability measure ν such that, for any $s > 0$:*

$$\mathop{\mathbb{E}}_{\Lambda\sim\nu}\left(\Delta(u,D^{\Lambda}_s)\right) \leq \sqrt{\frac{M}{\left(s\sqrt{2}\right)^n}}.$$

In particular, if $s > s_0 := M^{1/n}/\sqrt{2}$*, we have*

$$\mathop{\mathbb{E}}_{\Lambda\sim\nu}\left(\Delta(u, D_s^\Lambda)\right) \leq \left(\frac{s_0}{s}\right)^{n/2}.$$

To compare Propositions 10 and 11, we note that a random vector sampled according to D_s has an expected Euclidean norm given by $s\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n}{2}\right)} \approx s\sqrt{\frac{n}{2\pi}}$. So, it is fair to compare the effectiveness of smoothing with a parameter s Gaussian distribution and the uniform distribution over a ball of radius $s\sqrt{\frac{n}{2\pi}}$. We note that, if s_0 is as in Proposition 11 and w_0 is the radius of the so-called Gaussian heuristic, then

$$s_0\sqrt{\frac{n}{2\pi}} = \frac{M^{1/n}}{\sqrt{2}}\sqrt{\frac{n}{2\pi}} = w_0 \cdot \sqrt{e/2}.$$

Thus, we conclude that the parameter s_0 from Proposition 11 is larger than what we could hope it to be by a factor of $\sqrt{e/2}$.

4.2. **Connecting Uniform Ball Distribution to Gaussian.** However, recall that in the code-case we argued that, as the Hamming weight of a vector sampled according to the Bernoulli distribution is tightly concetrated, we could obtain the same smoothing bound for the Bernoulli distribution as we did for the uniform sphere distribution, essentially by showing that we can approximate a Bernoulli distribution by a convex combination of uniform sphere distributions. Similarly, we can relate the Gaussian distribution to the uniform distribution over a ball, and thereby remove this additional $\sqrt{e/2}$ factor.

While the intuition for the argument is the same as that which we used in the code-case, the argument is itself a bit more sophisticated. We begin with a lemma decomposing the Gaussian as a convex combination of balls.

Lemma 2. *The Gaussian distribution in dimension* n *of parameter* s *is the following convex combination of uniform distributions over balls:*

$$D_s = \frac{1}{s}\int_0^\infty G_n(w/s) \cdot u_{w\mathscr{B}}\, dw$$

where $G_n(x) := V_n(1) \cdot 2\pi \cdot x^{n+1} \cdot \exp\left(-\pi x^2\right) \geq 0$ *and* $\int_0^\infty G_n = 1$*.*

Proof. Following the probability density function of the Gaussian distribution D_s, one may write

$$D_s = \frac{1}{s^n}\int_0^\infty 1_{w\mathscr{S}} \cdot \exp\left(-\pi w^2/s^2\right) dw$$

where $1_{w\mathscr{S}}$ is the indicator function for the sphere of radius w. We use integration by parts to rewrite the above into an integral of balls rather than spheres:

$$D_s = \frac{1}{s^n}\int_0^\infty 1_{w\mathscr{B}} \cdot \frac{2\pi w}{s^2}\exp\left(-\pi w^2/s^2\right) dw$$

where $1_{w\mathscr{B}}$ is the indicator function for the ball or radius w. By definition, we have that $u_{w\mathscr{B}} = \frac{1}{V(w)}1_{w\mathscr{B}}$, that is, $1_{w\mathscr{B}} = V(1) \cdot w^n \cdot u_{w\mathscr{B}}$. Applying this substitution and reorganizing factors leads to the claim. □

We now quote the following bound, which makes precise the intuition that it is exponentially unlikely that a random Gaussian vector has norm $(1-\eta)$ factor smaller than its expected norm. This result provides the analogy for the Chernoff bound that we used for the code-case.

Proposition 12 ([Wai19, Example 2.5])**.** *Let* $\mathbf{X}$ *be a random Gaussian vector of dimension* n *and parameter* s*. Let* $0 < \eta < 1$*. Then*

$$\mathbb{P}\left(\|\mathbf{X}\|^2 \leq (1-\eta)\frac{ns^2}{2\pi}\right) \leq \exp\left(-n\eta^2/8\right).$$

This proposition allows us to prove the following lemma bounding $\frac{1}{s}\int_0^{\overline{w}} G_n(w/s)dw$ when $\overline{w} < s \cdot \sqrt{n/(2\pi)}$.

Lemma 3. *Let $\eta \in (0,1)$ and $\overline{w} = \sqrt{1-\eta} \cdot s \cdot \sqrt{n/(2\pi)}$. Then*

$$\frac{1}{s}\int_0^{\overline{w}} G_n(w/s)dw \le O(\sqrt{n}/s) \cdot \exp\bigl(-n\eta^2/8\bigr).$$

Proof. Let $\mathbf{X}$ be a random Gaussian vector of dimension n and parameter s. By Proposition 12:

$$\begin{aligned}
\frac{1}{s}\int_0^{\overline{w}} G_n(w/s)dw &= \frac{1}{s}\int_0^{\overline{w}} V_n(1)\left(\frac{w}{s}\right)^{n+1} \cdot 2\pi \cdot \exp\left(-\pi\tfrac{w^2}{s^2}\right) dw \\
&\le \frac{\overline{w}\cdot 2\pi}{s^2} \cdot \frac{1}{s^n}\int_0^{\overline{w}} V_n(w) \cdot \exp\left(-\pi\tfrac{w^2}{s^2}\right) dw \\
&\le O(\sqrt{n}/s) \cdot \frac{1}{s^n}\int_{0\le\|\mathbf{x}\|\le\overline{w}} \exp\left(-\pi\tfrac{\|\mathbf{x}\|^2}{s^2}\right) d\mathbf{x} \\
&= O(\sqrt{n}/s) \cdot \mathbb{P}\left(\|\mathbf{X}\| \le \overline{w}\right) \\
&\le O(\sqrt{n}/s) \cdot \mathbb{P}\left(\|\mathbf{X}\|^2 \le (1-\eta)\frac{ns^2}{2\pi}\right) \\
&\le O(\sqrt{n}/s) \cdot \exp\bigl(-\eta^2 n/8\bigr).
\end{aligned}$$

□

We now state a general proposition that allows us to translate smoothing bounds for the uniform ball distribution to the Gaussian distribution. It guarantees that if the uniform ball distribution smooths whenever $w > w_0$, the Gaussian distribution smoothes whenever $s > w_0 \cdot \sqrt{\frac{2\pi}{n}}$.

Proposition 13. *Let Λ be a random lattice of covolume M and let $u = u_{\mathbb{R}^n/\Lambda}$ be the uniform distribution over its cosets. Suppose that for all $w > w_0$ there is a function $f(n)$ such that*

$$\mathbb{E}_\Lambda\left(\Delta(u, u^\Lambda_{w\mathscr{B}})\right) \le f(n)\left(\frac{w_0}{w}\right)^{n/2}.$$

Let $s_0 = w_0\sqrt{\frac{2\pi}{n}}$. Then, for all $s > s_0$, defining $\eta = 1 - \frac{s_0}{s} \in (0,1)$, we have

$$\mathbb{E}_\Lambda\left(\Delta(u, D^\Lambda_s)\right) \le O(\sqrt{n}/s)\exp\bigl(-\eta^2 n/8\bigr) + f(n)\left(\frac{s}{s_0}\right)^{n/4}.$$

Proof. By Lemma 2 we have a convex combination $D_s = \frac{1}{s}\int_0^\infty G_n(w/s) \cdot u_{w\mathscr{B}}\, dw$. In particular, $G_n(x) \ge 0$ for all x. Therefore

$$\mathbb{E}_\Lambda\left(\Delta(u, D^\Lambda_s)\right) \le \frac{1}{s}\int_0^\infty G_n(w/s)\mathbb{E}_\Lambda\left(\Delta(u, u^\Lambda_{w\mathscr{B}})\right) dw.$$

We split the integral in two parts at radius $\overline{w} = \sqrt{1-\eta} \cdot s \cdot \sqrt{n/(2\pi)}$. For the first part $w \le \overline{w}$, we use the trivial bound $\mathbb{E}_\Lambda\left(\Delta(u, D^\Lambda_s)\right) \le 1$ which gives:

$$\frac{1}{s}\int_0^{\overline{w}} G_n(w/s)\mathbb{E}_\Lambda\left(\Delta(u, D^\Lambda_s)\right) dw \le \frac{1}{s}\int_0^{\overline{w}} G_n(w/s)dw.$$

We then apply Lemma 3, which bounds this part by $O(\sqrt{n}/s) \cdot \exp\bigl(-n\eta^2/8\bigr)$.

For the second part $w \ge \overline{w}$, we use the trivial bound $\frac{1}{s}\int_{\overline{w}}^\infty G_n(w/s)dw \le 1$ and, noting

$$w \ge \overline{w} = \sqrt{1-\eta} \cdot s \cdot \sqrt{n/(2\pi)} = \frac{1}{\sqrt{1-\eta}} \cdot s_0 \cdot \sqrt{n/(2\pi)} > s_0 \cdot \sqrt{n/(2\pi)} = w_0,$$

we may apply the assumption of the proposition, yielding

$$\mathbb{E}_\Lambda\left(\Delta(u, u^\Lambda_{w\mathscr{B}})\right) \le f(n)\left(\frac{w_0}{w}\right)^{n/2} \le f(n)\left(\frac{w_0}{\overline{w}}\right)^{1/2} = f(n)\left(\sqrt{1-\eta}\right)^{n/2} = f(n)\left(\frac{s_0}{s}\right)^{n/4}.$$

Adding these bounds yields the proposition. □

Combining the above proposition with Theorem 6, setting $f(n) = O(1)$, we obtain the following theorem.

Theorem 7. *Let Λ be a random lattice of covolume M sampled according to ν, let $u = u_{\mathbb{R}^n/\Lambda}$ be the uniform distribution over its cosets, and let*

$$s_0 = M^{1/n}/\sqrt{e}.$$

Then, for any $s > s_0$, setting $\eta = 1 - \frac{s_0}{s} \in (0,1)$, we have

$$\mathbb{E}_\Lambda\left(\Delta(u, D_s^\Lambda)\right) \le O(\sqrt{n}/s)\exp\left(-\eta^2 n/8\right) + O(1)\cdot\left(\frac{s_0}{s}\right)^{n/4}.$$

4.3. **Smoothing Random q-ary Lattices.** While the method of sampling lattices promised by the Minkowski-Hlawka-Siegel Theorem (Theorem 6) is indeed very convenient for computations, it does not tell us much about how to explicitly sample from the distribution. Furthermore it is not very relevant if one is interested in the random lattices that are used in cryptography.

For a more concrete sampling procedure that is relevant to cryptography, we can consider the randomized Construction A (or, more precisely, its dual), which gives a very popular random model of lattices which are easily constructed from random codes. Specifically, for a prime q and a linear code $\mathcal{C} \subseteq (\mathbb{Z}/q\mathbb{Z})^n$ we obtain a lattice as follows. First, we "lift" the codewords $\mathbf{c} \in \mathcal{C}$ to vectors in $\mathbb{R}^n$ in the natural way by identifying $\mathbb{Z}/q\mathbb{Z}$ with the set $\{0, 1, \ldots, q-1\}$; denote the lifted vector as $\tilde{\mathbf{c}}$. Then, we can define the following lattice:

$$\Lambda_\mathcal{C} \stackrel{\text{def}}{=} \{\tilde{\mathbf{c}} : \mathbf{c} \in \mathcal{C}\} + q\mathbb{Z}^n.$$

In other words: $\Lambda_\mathcal{C}$ consists of all vectors in the integer lattice $\mathbb{Z}^n$ whose reductions modulo q give an element of $\mathcal{C}$.

Fix integers $1 \le k \le n$, a prime q and a desired covolume M. We sample a random lattice Λ as follows.

- First, sample a random linear code $\mathcal{C} \subseteq (\mathbb{Z}/q\mathbb{Z})^n$ of dimension k (recall this means that we sample a random $k \times n$ matrix $\mathbf{G}$ and define $\mathcal{C} = \{\mathbf{mG} : \mathbf{m} \in (\mathbb{Z}/q\mathbb{Z})^k\}$).
- Then, we scale $\Lambda_\mathcal{C}$ by $\frac{1}{M^{1/n}} \cdot \frac{1}{q^{1-k/n}}$.
- Lastly, we output the dual of $\frac{1}{M^{1/n}} \cdot \frac{1}{q^{1-k/n}}\Lambda_\mathcal{C}$.

Notice that the scaling is chosen so that, as long as $\mathbf{G}$ is of full rank, the lattice Λ we output has the desired covolume M. We denote this procedure of sampling Λ by ν_A (the dependence on q, k and n is left implicit).

The important fact is that, up to an error term (which decreases as q increases), the expected number of lattice points from Λ^* in a Euclidean ball of radius r is roughly $\frac{V_n(r)}{M}$, as one would hope.

Proposition 14 ([Z14, Lemma 7.9.2])**.** *For every $n \ge 2$ and $1 \le k < n$ and prime power q, for $\Lambda \sim \nu_\mathrm{A}$ the expected number of lattice points from Λ^* in a Euclidean ball of radius $w = t\sqrt{n}$ satisfies*

$$\sqrt[n]{\frac{M\ \mathbb{E}_\Lambda(N_{\le r}(\Lambda^*))}{V_n(w)}} = 1 \pm \delta/t,$$

where $\delta = \frac{1}{q^{1-k/n}}$.

We now turn to bounding the expected statistical between u and $u^\Lambda_{w\mathscr{B}}$, where $\Lambda \sim \nu_\mathrm{A}$ and $w > 0$ is the radius of the Euclidean ball from which the noise is uniformly sampled. First, we will state

an explicit formula for the Fourier transform of $1_{\mathscr{B}_w}$, the indicator function of a Euclidean ball of radius w, in terms of *Bessel functions*.

Notation 2. *For a positive real number $\mu > 0$, we denote by $J_\mu : \mathbb{R} \to \mathbb{R}$ the Bessel function of the first kind of order μ.*

The important fact concerning Bessel functions that we will use is the following.

Fact 3. *We have*

$$\widehat{1_{\mathscr{B}_w}}(\mathbf{y}) = \left(\frac{w}{|\mathbf{y}|_2}\right)^{n/2} J_{n/2}(2\pi w|\mathbf{y}|_2). \tag{19}$$

We will refrain from providing an explicit formula for Bessel functions, and instead use the following asymptotic estimate as a black-box.

Proposition 15 ([Kra06])**.** *For any $x \in \mathbb{R}$ we have*

$$|J_{n/2}(x)| \leq |x|^{-1/3}.$$

Using this proposition, we first prove a technical lemma that will be reused when we discuss smoothing arbitrary lattices. In order to state the lemma, we introduce the following auxiliary function.

Notation 3. *For a real $w > 0$, we define $g_w : \mathbb{R} \to \mathbb{R}$ via*

$$g_w(t) \stackrel{def}{=} \frac{1}{V_n(w)} \widehat{1_{\mathscr{B}_w}}(\mathbf{x})^2,$$

where $\mathbf{x}$ is any vector in $\mathbb{R}^n$ of norm t. Note that as $\widehat{1_{\mathscr{B}_w}}(\mathbf{x})$ depends only on $|\mathbf{x}|_2$, this is indeed well-defined.

The following lemma leverages Proposition 15 to upper bound g_w on a closed interval.

Lemma 4. *For any $w > 0$ and any $0 \leq a$ and $b = \left(1 + \frac{1}{n}\right) a$ we have, for some constant $C > 0$:*

$$\max_{a \leq t \leq b} g_w(t) \leq \frac{C}{V_n(b)\, w^{2/3}} \frac{1}{a^{2/3}}.$$

Proof. First, we notice that for all $t \in [a, b]$

$$V_n(t) = \left(\frac{t}{b}\right)^n V_n(b) \geq \left(\frac{a}{b}\right)^n V_n(b) = \left(1 + \frac{1}{n}\right)^{-n} V_n(b) \geq \frac{1}{C'} V_n(b)$$

for some constant $C' > 0$. We now use Proposition 15 to derive

$$\max_{a \leq t \leq b} g_w(t) \leq \frac{C'}{V_n(b)} \max_{a \leq t \leq b} J_{n/2}(2\pi w t)^2 \leq \frac{C}{V_n(b)\, w^{2/3}} \frac{1}{a^{2/3}}$$

for an appropriate constant $C > 0$. □

We now provide the main theorem of this section. It demonstrates that to smooth our ensemble of random q-ary codes (in expectation) with the uniform distribution over the ball of radius w, it still suffices to choose $w > w_0 \stackrel{\text{def}}{=} \sqrt{n2\pi/e} M^{1/n}$, assuming q is not too small.

Theorem 8. *Let $n > 2$ and $1 \leq k < n$. Let q be a prime and set $\gamma = \frac{n^{3/2}}{q^{1-k/n}}$. Let $\Lambda \sim \nu_{\mathrm{A}}$. For some constant $C > 0$, we have*

$$\mathbb{E}_\Lambda\left(\Delta(u, u^\Lambda_{w\mathscr{B}})\right) \leq C \left(\frac{n}{w}\right)^{1/3} e^{\gamma/2} \sqrt{\frac{M}{V_n(w)}}.$$

In particular, if $w > w_0 \stackrel{def}{=} \sqrt{n2\pi/e}M^{1/n}$, *we have*

$$\mathbb{E}_\Lambda\left(\Delta(u, u^\Lambda_{w\mathscr{B}})\right) \leq O\left(\left(\frac{n}{w}\right)^{1/3} e^{\gamma/2}\right) \cdot \left(\frac{w_0}{w}\right)^{n/2}.$$

Proof. Let $t_j \stackrel{\text{def}}{=} \left(1+\frac{1}{n}\right)^j$ for $j \in \mathbb{N}$ and

$$N_j \stackrel{\text{def}}{=} \#\{\mathbf{x}^* \in \Lambda^* : t_j \leq |\mathbf{x}^*|_2 < t_{j+1}\} \quad ; \quad \varphi_j \stackrel{\text{def}}{=} \max_{t_j \leq t \leq t_{j+1}} g_w(t).$$

Now, we apply Proposition 9 and the above definitions to obtain

$$\begin{aligned}
\mathbb{E}_\Lambda\left(2\Delta(u, u^\Lambda_{w\mathscr{B}})\right) &\leq \mathbb{E}_\Lambda\left(\sqrt{\sum_{\mathbf{x}\in\Lambda^*\setminus\{\mathbf{0}\}} |\widehat{u_{w\mathscr{B}}}(\mathbf{x})|^2}\right) \\
&\leq \sqrt{\frac{1}{V_n(w)}\mathbb{E}_\Lambda\left(\sum_{\mathbf{x}\in\Lambda^*\setminus\{\mathbf{0}\}} g_w(\mathbf{x})\right)} \quad \text{(Jensen's inequality)} \\
&\leq \sqrt{\frac{1}{V_n(w)}\mathbb{E}_\Lambda\left(\sum_{j=0}^{\infty} N_j\varphi_j\right)} \\
&\leq \sqrt{\frac{1}{V_n(w)}\sum_{j=0}^{\infty}\mathbb{E}\left(N_{\leq t_{j+1}}(\Lambda^*)\right)\varphi_j}\ .
\end{aligned}$$

By Proposition 14, we may upper bound

$$\mathbb{E}_\Lambda\left(N_{\leq t_{j+1}}(\Lambda^*)\right) \leq M \cdot V_n(t_{j+1})\left(1+\left(\frac{\sqrt{n}}{\left(1+\frac{1}{n}\right)^{jn} q^{1-k/n}}\right)\right)^n. \tag{20}$$

Now, recalling $\gamma = \frac{n^{3/2}}{q^{1-k/n}}$ we have for any $j \geq 0$:

$$\left(1+\left(\frac{\sqrt{n}}{\left(1+\frac{1}{n}\right)^{jn} q^{1-k/n}}\right)\right)^n \leq \left(1+\left(\frac{\sqrt{n}}{q^{1-k/n}}\right)\right)^n \leq e^{n\cdot\frac{\sqrt{n}}{q^{1-k/n}}} = e^\gamma\ .$$

Thus, we conclude

$$\mathbb{E}_\Lambda\left(2\Delta(u, u^\Lambda_{w\mathscr{B}})\right) \leq \sqrt{\frac{e^\gamma M}{V_n(w)}\sum_{j=0}^{\infty} V_n(t_{j+1}))\varphi_j}\ .$$

Now, by Lemma 4 we have $\varphi_j \leq \frac{C_1}{V_n(t_{j+1})w^{2/3}}\frac{1}{t_j^{2/3}}$ for all $j \geq 0$. Hence,

$$\begin{aligned}
\sum_{j=0}^{\infty} V_n(t_{j+1})\,\varphi_j &\leq \frac{C_1}{w^{2/3}}\sum_{j=0}^{\infty}\frac{V_n(t_{j+1})}{V_n(t_{j+1})}\frac{1}{t_j^{2/3}} \\
&= \frac{C_1}{w^{2/3}}\sum_{j=0}^{\infty}\frac{1}{(1+1/n)^{2j/3}} \\
&= \frac{C_1}{w^{2/3}}\frac{1}{1-(1+1/n)^{-2/3}} \\
&\leq \frac{C_2\, n^{2/3}}{w^{2/3}}\ ,
\end{aligned}$$

for an appropriate constant $C_2 > 0$. Thus, putting everything together we derive

$$\mathbb{E}_\Lambda\left(\Delta(u, u^\Lambda_{w\mathscr{B}})\right) \le \sqrt{\frac{e^\gamma M}{2V_n(w)} \cdot \frac{C_2 n^{2/3}}{w^{2/3}}} \le C\left(\frac{n}{w}\right)^{1/3} e^{\gamma/2}\sqrt{\frac{M}{V_n(w)}}$$

for some constant $C > 0$. The "in particular" part of the Theorem follows analogously to the corresponding argumentation used in the proof of Theorem 6. □

Next, turning to Gaussian noise, we could again prove a smoothing bound "directly", but this will lose the same factor of $\sqrt{e/2}$ as we had earlier. Instead, we apply Proposition 13 with the function $f(n) = O\left(\left(\frac{n}{w}\right)^{1/3} e^{\gamma/2}\right)$ to conclude the following.

Theorem 9. *Let $n > 2$ and $1 \le k < n$. Let q be a prime and set $\gamma = \frac{n^{3/2}}{q^{1-k/n}}$. Let Λ be a random q-ary lattice sampled according to ν_A, let $u = u_{\mathbb{R}^n/\Lambda}$ be the uniform distribution over its cosets, and let*

$$s_0 = M^{1/n}/\sqrt{e}.$$

Then, for any $s > s_0$, setting $\eta = 1 - \frac{s_0}{s} \in (0,1)$, we have

$$\mathbb{E}_\Lambda\left(\Delta\left(u, D^\Lambda_s\right)\right) \le O\left(\sqrt{n}/s\right)\exp\left(-\eta^2 n/8\right) + O(1)\cdot(s/s_0)^{n/4}\, e^{\gamma/2}.$$

4.4. **Smoothing Arbitrary Lattices.** We now turn our attention to the task of smoothing arbitrary lattices.

Analogously to how we used the minimum distance of the dual code to give our smoothing bound for worst-case codes, we will use the shortest vector of the dual lattice in order to provide our smoothing bound for worst-case lattices. The lemma that we will apply is the following where

$$C_{\mathrm{KL}} \stackrel{\mathrm{def}}{=} 2^{0.401}.$$

Lemma 5 ([PS09, Lemma 3])**.** *For any n-dimensional lattice Λ,*

$$\forall t \ge \lambda_1(\Lambda), \quad N_{\le t}(\Lambda) \le \frac{V_n(t)}{V_n(\lambda_1(\Lambda))}\, C_{\mathrm{KL}}^{n(1+o(1))}$$

Remark 2. *This lemma is a consequence of the Kabatiansky and Levenshtein' bound* [KL78] *on the size of spherical codes, historically known as the "second linear programming bound". It is why we may refer to the aforementioned bound of Lemma 5 as the second linear programming bound.*

We begin by considering the effectiveness of smoothing with noise uniformly sampled from the ball. The following theorem is proved using similar techniques to those we used for Theorem 8, although instead of using Proposition 14 to bound the $N_{\le t}(\Lambda^*)$'s, we use Lemma 5.

Theorem 10. *Let Λ be an n-dimensional lattice and $u = u_{\mathbb{R}^n/\Lambda}$ be the uniform distrution over its cosets. Then, it holds that*

$$\Delta\left(u, u^\Lambda_{w\mathscr{B}}\right) \le \sqrt{\frac{C_{\mathrm{KL}}^{n(1+o(1))}}{V_n(\lambda_1(\Lambda^*))\ V_n(w)}}.$$

In particular, setting

$$w_0 = n \cdot \frac{C_{\mathrm{KL}}^{1+o(1/n)}}{2\pi \cdot e \cdot \lambda_1(\Lambda^*)}$$

for all $w > w_0$, it holds that

$$\Delta\left(u, u^\Lambda_{w\mathscr{B}}\right) \le O(1)(w_0/w)^{n/2}.$$

Proof. Define

$$t_0 \stackrel{\text{def}}{=} \lambda_1(\Lambda^*),\ t_{j+1} \stackrel{\text{def}}{=} \left(1+\tfrac{1}{n}\right) t_j \ \text{ for } j \geq 1,$$
$$\varphi_j \stackrel{\text{def}}{=} \max_{t_j \leq t \leq t_{j+1}} \{g_w(t)\},$$

where we recall the definition of $g_w(t) = \frac{1}{V_n(w)} \widehat{1_{\mathscr{B}_w}}(\mathbf{x})^2$ with $|\mathbf{x}|_2 = t$ (see Notation 3). We also define

$$N_j \stackrel{\text{def}}{=} \#\{\mathbf{x}^* \in \Lambda^* : t_j \leq |\mathbf{x}^*|_2 \leq t_{j+1}\} \ .$$

With this notation and Proposition 9 we have

$$\begin{aligned}
\Delta\left(u, u^{\Lambda}_{w\mathscr{B}}\right) &\leq \sqrt{\sum_{\mathbf{x}\in\Lambda^*\setminus\{\mathbf{0}\}} |\widehat{u_{w\mathscr{B}}}(\mathbf{x})|^2} \\
&\leq \sqrt{\frac{1}{V_n(w)} \sum_{\mathbf{x}\in\Lambda^*\setminus\{\mathbf{0}\}} g_w(\mathbf{x})} \\
&\leq \sqrt{\frac{1}{V_n(w)} \sum_{j=0}^{\infty} N_j \varphi_j} \\
&\leq \sqrt{\frac{1}{V_n(w)} \sum_{j=0}^{\infty} N_{\leq t_{j+1}}(\Lambda^*) \varphi_j} \ . \qquad (21)
\end{aligned}$$

By Lemma 4, for some constant $C_1 > 0$ we obtain

$$\varphi_j \leq \frac{C_1}{V_n(t_{j+1}) w^{2/3}} \frac{1}{t_j^{2/3}} \ .$$

Combining this with the upper bound on $N_{\leq t_{j+1}}(\Lambda^*)$ provided by Lemma 5 (note that $t_{j+1} \geq \lambda_1(\Lambda^*)$ for all $j \geq 0$), we find

$$\begin{aligned}
\sum_{j=0}^{\infty} N_{\leq t_{j+1}}(\Lambda^*)\varphi_j &\leq \sum_{j=0}^{\infty} \frac{V_n(t_{j+1})}{V_n(\lambda_1(\Lambda^*))} C_{\mathrm{KL}}^{n(1+o(1))} \ \frac{C_1}{V_n(t_{j+1}) w^{2/3}} \ \frac{1}{t_j^{2/3}} \\
&= \frac{C_{\mathrm{KL}}^{n(1+o(1))}}{V_n(\lambda_1(\Lambda^*)) w^{2/3}} \ \sum_{j=0}^{\infty} \frac{1}{t_j^{2/3}} \\
&= \frac{C_{\mathrm{KL}}^{n(1+o(1))}}{V_n(\lambda_1(\Lambda^*)) w^{2/3}} \ \sum_{j=0}^{\infty} \frac{1}{\lambda_1(\Lambda^*)^{2/3} \left(1+\frac{1}{n}\right)^{2j/3}} \\
&\leq \frac{C_{\mathrm{KL}}^{n(1+o(1))}}{V_n(\lambda_1(\Lambda^*)) w^{2/3}} \ \left(\frac{n}{w\lambda_1(\Lambda^*)}\right)^{2/3}
\end{aligned}$$

In the above, all necessary constants were absorbed into the $C_{\mathrm{KL}}^{o(n)}$ term. Combining this with (21), we obtain the first part of the theorem. The "in particular" part again follows using Stirling's approximation. □

Next, we can consider the effectiveness of smoothing with the Gaussian distribution. As usual, we could follow the steps of the proof of Theorem 10 and obtain the same result, but with an additional multiplicative factor of $\sqrt{\frac{e}{2}}$. That is, we obtain:

Theorem 11. *Let Λ be an n-dimensional lattice and $u = u_{\mathbb{R}^n/\Lambda}$ be the uniform distrution over its cosets. Then, it holds:*

$$\Delta\left(u, D_s^\Lambda\right) \leq \sqrt{\frac{C_{\mathrm{KL}}^{n(1+o(1))}}{V_n\left(\lambda_1(\Lambda^*)\right)\ V_n\left(s\sqrt{n/(2\pi)}\right)}\ \left(\frac{e}{2}\right)^{n/2}}$$

In particular, setting

$$s_0 = \sqrt{n} \cdot \frac{C_{\mathrm{KL}}^{1+o(1/n)}}{2\sqrt{\pi e} \cdot \lambda_1(\Lambda^*)},$$

it holds for any $s > s_0$ that $\Delta\left(u, D_s^\Lambda\right) \leq O(1) \cdot (s_0/s)^{n/2}$.

However, as usual it is more effective to combine the bound for the uniform ball distribution and decompose the Gaussian as a convex combination of uniform ball distributions, i.e., to apply Proposition 13. In this way, we can obtain the following theorem, improving the smoothing bound s_0 by another $\sqrt{e/2}$ factor. Below, we are setting $f(n) = O(1)$.

Theorem 12. *Let Λ be an n-dimensional lattice, $u = u_{\mathbb{R}^n/\Lambda}$ the uniform distribution over its cosets, and*

$$s_0 = \sqrt{n}\frac{C_{\mathrm{KL}}^{1+o(1/n)}}{\sqrt{2\pi} \cdot e \cdot \lambda_1(\Lambda^*)}\ .$$

Then, for any $s > s_0$ and letting $\eta = 1 - \frac{s_0}{s} \in (0,1)$, it holds that

$$\Delta\left(u, D_s^\Lambda\right) \leq O(\sqrt{n}/s) \cdot \exp\left(-\frac{n\eta^2}{8}\right) + O(1) \cdot \left(\frac{s_0}{s}\right)^{n/4}.$$

References

[ABL01] Alexei E. Ashikhmin, Alexander Barg, and Simon Litsyn. Estimates of the distance distribution of codes and designs. *Electron. Notes Discret. Math.*, 6:4–14, 2001.

[ACKL05] Alexei E. Ashikhmin, Gérard D. Cohen, Michael Krivelevich, and Simon Litsyn. Bounds on distance distributions in codes of known size. *IEEE Trans. Inf. Theory*, 51(1):250–258, 2005.

[ADRS15] Divesh Aggarwal, Daniel Dadush, Oded Regev, and Noah Stephens-Davidowitz. Solving the shortest vector problem in 2n time using discrete gaussian sampling. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 733–742, 2015.

[Ale11] Michael Alekhnovich. More on average case vs approximation complexity. *Computational Complexity*, 20(4):755–786, 2011.

[Ban93] Wojciech Banaszczyk. New bounds in some transference theorems in the geometry of numbers. *Mathematische Annalen*, 296(1):625–635, 1993.

[Bas65] LA Bassalygo. New upper bounds for codes correcting errors. *Probl. Peredachi Inform*, 1(4):41–44, 1965.

[BLVW19] Zvika Brakerski, Vadim Lyubashevsky, Vinod Vaikuntanathan, and Daniel Wichs. Worst-case hardness for lpn and cryptographic hashing via code smoothing. In *Annual international conference on the theory and applications of cryptographic techniques*, pages 619–635. Springer, 2019.

[CE03] Henry Cohn and Noam Elkies. New upper bounds on sphere packings I. *Ann. of Math*, (157-2):689–714, 2003.

[Chu97] Fan R. K. Chung. *Spectral graph theory*, volume 92 of *CBMS Regional Conference Series in Mathematics*. American Mathematical Society, 1997.

[DL98] Philippe Delsarte and Vladimir Iossifovitch Levenshtein. Association schemes and coding theory. 44(6):2477–2504, 1998.

[DST19] Thomas Debris-Alazard, Nicolas Sendrier, and Jean-Pierre Tillich. Wave: A new family of trapdoor one-way preimage sampleable functions based on codes. In *2019*, Kobe, Japan, December 2019. Springer.

[DT17] Thomas Debris-Alazard and Jean-Pierre Tillich. Statistical decoding. preprint, January 2017. arXiv:1701.07416.

[GPV08] Craig Gentry, Chris Peikert, and Vinod Vaikuntanathan. Trapdoors for hard lattices and new cryptographic constructions. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 197–206, 2008.

[IS98] Mourad E.H. Ismail and Plamen Simeonov. Strong asymptotics for Krawtchouk polynomials. *Journal of Computational and Applied Mathematics*, pages 121–144, 1998.

[KL78] Grigory Kabatiansky and Vladimir I. Levenshtein. Bounds for packings on a sphere and in space. *Problems of Information Transmission*, (14):1–17, 1978.

[Klø07] Torleiv Kløve. *Codes for Error Detection*, volume 2 of *Series on Coding Theory and Cryptology*. World-Scientific, 2007.

[Kra06] Ilia Krasikov. Uniform bounds for bessel functions. *Journal of Applied Analysis*, 12:83–91, 06 2006.

[Lev79] Vladimir I. Levenshtein. On bounds for packings in n-dimensional euclidean space. *Dokl. Akad. Nauk SSSR*, 245:1299–1303, 1979.

[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.

[LLBS14] Cong Ling, Laura Luzzi, Jean-Claude Belfiore, and Damien Stehlé. Semantically secure lattice codes for the gaussian wiretap channel. *IEEE Transactions on Information Theory*, 60(10):6399–6416, 2014.

[McE78] Robert J. McEliece. *A Public-Key System Based on Algebraic Coding Theory*, pages 114–116. Jet Propulsion Lab, 1978. DSN Progress Report 44.

[MR07] Daniele Micciancio and Oded Regev. Worst-case to average-case reductions based on gaussian measures. *SIAM Journal on Computing*, 37(1):267–302, 2007.

[MRJW77] Robert J. McEliece, Eugene R. Rodemich, Howard Rumsey Jr., and Lloyd R. Welch. New upper bounds on the rate of a code via the delsarte-macwilliams inequalities. *IEEE Trans. Inf. Theory*, 23(2):157–166, 1977.

[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. pages 2069–2073, 2013.

[PS09] Xavier Pujol and Damien Stehlé. Solving the shortest lattice vector problem in time $2^{2.465n}$. *IACR Cryptol. ePrint Arch.*, 2009:605, 2009.

[vL99] Jacobus Hendricus van Lint. *Introduction to coding theory*. Graduate texts in mathematics. Springer, 3rd edition edition, 1999.

[Wai19] Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.

[YZ20] Yu Yu and Jiang Zhang. Smoothing out binary linear codes and worst-case sub-exponential hardness for lpn. *IACR Cryptol. ePrint Arch.*, 2020:870, 2020.

[Z14] *Lattice Coding for Signals and Networks: A Structured Coding Approach to Quantization, Modulation and Multiuser Information Theory.*

Appendix A. Proof of Proposition 6

Our aim in this section is to prove the following proposition

Proposition 6. *Let $\varepsilon > 0$ and $p \in [0, 1/2]$. Then,*

$$\Delta(u, f^{\mathcal{C}}_{\mathrm{ber},p}) \leq \sum_{r=(1-\varepsilon)np}^{(1+\varepsilon)np} \Delta(u, u^{\mathcal{C}}_r) + 2^{-\Omega(n)}.$$

Roughly speaking, this proposition is a consequence of the fact that a Bernoulli distribution concentrates Hamming weights over a small number of slices close to the expected weight (here np) and, on each slice the Bernoulli distribution is uniform. Let us introduce the truncated Bernoulli distribution over words of Hamming weight $[(1-\varepsilon)pn, (1+\varepsilon)pn]$ for some $\varepsilon > 0$, namely

$$f_{\mathrm{truncBer},p}(\mathbf{x}) \stackrel{\mathrm{def}}{=} \begin{cases} \frac{1}{Z} f_{\mathrm{ber},p}(\mathbf{x}) & \text{if } |\mathbf{x}| \in [(1-\varepsilon)pn, (1+\varepsilon)pn] \\ 0 & \text{otherwise.} \end{cases} \tag{22}$$

where

$$Z \stackrel{\mathrm{def}}{=} \sum_{|\mathbf{y}|=(1-\varepsilon)np}^{(1+\varepsilon)np} f_{\mathrm{ber},p}(\mathbf{y}) \tag{23}$$

is the probability normalizing constant.

Proposition 6 is a consequence of the following lemmas.

Lemma 6. *Let $\varepsilon > 0$. We have*

$$\Delta\left(f_{\mathrm{ber},p}, f_{\mathrm{truncBer},p}\right) = 2^{-\Omega(n)}.$$

Proof. By Chernoff's bound:

$$0 \leq 1 - Z = \sum_{|\mathbf{y}| \notin [(1-\varepsilon)np, (1+\varepsilon)np]} f_{\mathrm{ber},p}(\mathbf{y}) \leq 2e^{-\varepsilon^2 n} = 2^{-\Omega(n)}. \tag{24}$$

Therefore for any $|\mathbf{x}| \in [(1-\varepsilon)np, (1+\varepsilon)np]$,

$$\begin{aligned} f_{\mathrm{truncBer},p}(\mathbf{x}) &= \frac{1}{1 - 2^{-\Omega(n)}}\ f_{\mathrm{ber},p}(\mathbf{x}) \\ &= \left(1 + 2^{-\Omega(n)}\right)\ f_{\mathrm{ber},p}(\mathbf{x}). \end{aligned} \tag{25}$$

We have now the following computation:

$$\begin{aligned} 2\Delta\left(f_{\mathrm{ber},p}, f_{\mathrm{truncBer},p}\right) &= \sum_{\mathbf{x} \in \mathbb{F}_2^n} \left|f_{\mathrm{ber},p}(\mathbf{x}) - f_{\mathrm{truncBer},p}(\mathbf{x})\right| \\ &= \sum_{|\mathbf{x}| \in [(1-\varepsilon)np, (1+\varepsilon)np]} \left|f_{\mathrm{ber},p}(\mathbf{x}) - f_{\mathrm{truncBer},p}(\mathbf{x})\right| + \sum_{|\mathbf{x}| \notin [(1-\varepsilon)np, (1+\varepsilon)np]} \left|f_{\mathrm{ber},p}(\mathbf{x})\right| \\ &= 2^{-\Omega(n)} \left(\sum_{|\mathbf{x}| \in [(1-\varepsilon)np, (1+\varepsilon)np]} \left|f_{\mathrm{ber},p}(\mathbf{x})\right| \right) + 2^{-\Omega(n)} \quad \text{(Equations (24) and (25))} \\ &= 2^{-\Omega(n)} \end{aligned}$$

where in the last line we used that $f_{\mathrm{ber},p}$ is a probability distribution.

□

Lemma 7. *We have*

$$\Delta\left(u, f^{\mathcal{C}}_{\mathrm{ber},p}\right) \leq \Delta\left(u, f^{\mathcal{C}}_{\mathrm{truncBer},p}\right) + 2^{-\Omega(n)}$$

Proof. By the triangle inequality,

$$\Delta\left(u, f^{\mathcal{C}}_{\mathrm{ber},p}\right) \leq \Delta\left(u, f^{\mathcal{C}}_{\mathrm{truncBer},p}\right) + \Delta\left(f^{\mathcal{C}}_{\mathrm{ber},p}, f^{\mathcal{C}}_{\mathrm{truncBer},p}\right).$$

Focusing on the second term now:

$$\begin{aligned} \Delta\left(f^{\mathcal{C}}_{\mathrm{ber},p}, f^{\mathcal{C}}_{\mathrm{truncBer},p}\right) &= \frac{1}{2} \sum_{\mathbf{y} \in \mathbb{F}_2^n / \mathcal{C}} \left|f^{\mathcal{C}}_{\mathrm{ber},p}(\mathbf{y}) - f^{\mathcal{C}}_{\mathrm{truncBer},p}(\mathbf{y})\right| \\ &= \frac{1}{2} \sum_{\mathbf{y} \in \mathbb{F}_2^n / \mathcal{C}} \left| \sum_{\mathbf{c} \in \mathcal{C}} f_{\mathrm{ber},p}(\mathbf{c}+\mathbf{y}) - \sum_{\mathbf{c} \in \mathcal{C}} f_{\mathrm{truncBer},p}(\mathbf{c}+\mathbf{y}) \right| \\ &\leq \frac{1}{2} \sum_{\mathbf{y} \in \mathbb{F}_2^n / \mathcal{C}} \sum_{\mathbf{c} \in \mathcal{C}} \left|f_{\mathrm{ber},p}(\mathbf{c}+\mathbf{y}) - f_{\mathrm{truncBer},p}(\mathbf{c}+\mathbf{y})\right| \\ &= \Delta\left(f_{\mathrm{ber},p}, f_{\mathrm{truncBer},p}\right). \end{aligned}$$

which concludes the proof by Lemma 6. □

The following lemma is a basic property of the statistical distance.

Lemma 8. *For any distribution f and $(g_i)_{1 \leq i \leq m}$ we have:*

$$\Delta\left(f, \sum_{i=1}^{m} \lambda_i g_i\right) \leq \sum_{i=1}^{m} \lambda_i\ \Delta(f, g_i)$$

where the λ_i's are positive and sum to one.

We are now ready to prove Proposition 6.

Proof of Proposition 6. First, by Lemma 7 we have:

$$\Delta\left(u, f^{\mathcal{C}}_{\mathrm{ber},p}\right) \leq \Delta\left(u, f^{\mathcal{C}}_{\mathrm{truncBer},p}\right) + 2^{-\Omega(n)}. \tag{26}$$

To upper-bound $\Delta\left(u, f^{\mathcal{C}}_{\mathrm{truncBer},p}\right)$ we are going to use Lemma 8. Notice that

$$f_{\mathrm{ber},p} = \sum_{r=0}^{n} \binom{n}{r} p^r (1-p)^{n-r} u_r.$$

Therefore it is readily seen that

$$f_{\mathrm{truncBer},p} = \sum_{r=(1-\varepsilon)np}^{(1+\varepsilon)np} \lambda_r \; u_r \quad \text{where} \quad \lambda_r \stackrel{\text{def}}{=} \frac{1}{Z} \binom{n}{r} p^r (1-p)^{n-r}.$$

By using Lemma 8 we obtain:

$$\begin{aligned} \Delta\left(u, f^{\mathcal{C}}_{\mathrm{truncBer},p}\right) &\leq \sum_{r=(1-\varepsilon)np}^{(1+\varepsilon)np} \lambda_r \, \Delta\left(u, u^{\mathcal{C}}_r\right) \\ &\leq \sum_{r=(1-\varepsilon)np}^{(1+\varepsilon)np} \Delta\left(u, u^{\mathcal{C}}_r\right) \end{aligned} \tag{27}$$

where in the last line we used that the λ_r's are smaller than one. To conclude the proof we plug Equation (27) in (26).

□

Appendix B. Proof of Proposition 7

Our aim in this section is to prove the following proposition which is an extension of [ABL01, Theorem 3] for $\tau \in [\delta, 1]$. ([ABL01, Theorem 3] only applied for $\tau \in [\delta, 1/2]$.)

Proposition 7. *Let $\delta \in (0, 1/2)$ and $\delta^{\perp} \stackrel{def}{=} 1/2 - \sqrt{\delta(1-\delta)}$. For any $\delta \leq \tau \leq 1$:*

$$b(\delta, \tau) \leq c(\delta, \tau) \stackrel{def}{=} \begin{cases} h(\tau) + h\left(\delta^{\perp}\right) - 1 & \text{if } \tau \in [\delta, 1-\delta] \\ 2\left(h(\delta^{\perp}) - a(\tau, \delta^{\perp})\right) & \text{otherwise.} \end{cases} \tag{14}$$

where $a(\cdot, \cdot)$ is defined in Proposition 5.

Our proof is mainly a rewriting of the proof of [ABL01, Theorem 3] which relies on the following proposition.

Proposition 16 ([ABL01, Proposition 2 with $d' = 0$])**.** *Let $\mathcal{C}$ be a binary code of length n such that $d_{\min}(\mathcal{C}) = \Omega(n)$. Let $t \stackrel{def}{=} \frac{n}{2} - \sqrt{d_{\min}(\mathcal{C})(n - d_{\min}(\mathcal{C}))}$ and a be such that*

$$x_1^{(t+1)} < a < x_1^{(t)} \quad \text{and} \quad \frac{K_t(a)}{K_{t+1}(a)} = -1$$

where $x_1^{(\mu)}$ denotes the first root of the Krawtchouk polynomial of order μ, namely K_μ.
When $0 \leq w < t \leq n/2$, we have:

$$\sum_{\mathbf{c} \in \mathcal{C} \setminus \{\mathbf{0}\}} K_w(|\mathbf{c}|)^2 \leq \frac{t+1}{2a} \frac{\binom{n}{w}}{\binom{n}{t}} \left(\binom{n}{t+1} + \binom{n}{t} \right)^2 \tag{28}$$

The approach is to optimize on the choice of w in Proposition 16 to give an upper-bound on $N_\ell(\mathcal{C})$. More precisely we observe that

$$N_\ell(\mathcal{C}) \leq \frac{1}{K_w(\ell)^2} \sum_{\mathbf{c}\in\mathcal{C}\setminus\{\mathbf{0}\}} K_w(|\mathbf{c}|)^2 \leq \frac{1}{K_w(\ell)^2} \frac{t+1}{2} \frac{\binom{n}{w}}{\binom{n}{t}} \left(\binom{n}{t+1} + \binom{n}{t} \right)^2 \tag{29}$$

and then choose w to minimize $\frac{\binom{n}{w}}{K_w(\ell)^2}$.

Proof of Proposition 7. It will be helpful to bring in the following map:

$$x \in [0,1] \mapsto x^\perp \stackrel{\text{def}}{=} \frac{1}{2} - \sqrt{x(1-x)}$$

It can be verified that this application is an involution, is symmetric $(1-x)^\perp = x^\perp$ and decreasing on $[0, \frac{1}{2}]$.

Let $\mathcal{C}$ be a binary code of length n such that $d_{\min}(\mathcal{C}) = \delta n$ where $\delta \in (0, 1/2]$ and t be defined as in Proposition 16. Let $\omega \stackrel{\text{def}}{=} \frac{w}{n}, \lambda \stackrel{\text{def}}{=} \frac{\ell}{n}$ and $\delta^\perp \stackrel{\text{def}}{=} 1/2 - \sqrt{\delta(1-\delta)}$. Then by Proposition 16 we have (see Equation (29))

$$\frac{\log_2 N_\ell(\mathcal{C})}{n} \leq h(\omega) + h(\delta^\perp) - \frac{2\log_2 |K_w(\ell)|}{n} + o(1). \tag{30}$$

Case 1: $\lambda \in [\delta, 1-\delta]$

It is optimal to choose in this case w such that $\omega = \lambda^\perp - \varepsilon$ where $\varepsilon > 0$ and $\varepsilon = o(1)$ as n tends to infinity. Let us first notice that $\lambda \in [\delta, 1-\delta]$ implies that $\lambda^\perp \leq \delta^\perp$ which together with $\omega < \lambda^\perp$ implies that $\omega < \delta^\perp$ which in turn is equivalent to the condition $w < t$ for being able to apply Proposition 16. Moreover $\omega < \lambda^\perp$ also implies $\lambda < \omega^\perp$ and by using Proposition 5 we obtain

$$\frac{2\log_2 |K_w(\ell)|}{n} \leq h(\omega) + 1 - h(\lambda) + o(1).$$

Therefore:

$$\frac{\log_2 N_\ell(\mathcal{C})}{n} \leq h(\omega) + h(\delta^\perp) - h(\omega) - 1 + h(\lambda) + o(1) = h(\delta^\perp) + h(\lambda) - 1 + o(1).$$

Case 2: $\lambda \in (1-\delta, 1]$

In that case, let $\omega = \delta^\perp - \varepsilon$ with $\varepsilon > 0$ and $\varepsilon = o(1)$ as n tends to infinity. Here we can write

$$\frac{2\log_2 |K_w(\ell)|}{n} = \frac{\log_2(K_w(\ell)^2)}{n} = \frac{\log_2(K_w(n-\ell)^2)}{n}.$$

Since $\lambda > 1-\delta$, we have $1-\lambda < \delta$. On the other hand, $\omega < \delta^\perp$ implies $\delta < \omega^\perp$. We deduce from these two inequalities that $1-\lambda < \omega^\perp$. By using Proposition 5 again, we get

$$\frac{\log_2(K_w(n-\ell)^2)}{n} = 2a(1-\lambda, \delta^\perp) + o(1) = 2a(\lambda, \delta^\perp) + o(1).$$

By plugging this estimate in (30) we get

$$\frac{\log_2 N_\ell(\mathcal{C})}{n} \leq 2h(\delta^\perp) - 2a(\lambda, \delta^\perp).$$

This concludes the proof. □

Appendix C. Proof of Theorem 5

Our aim in this appendix is to prove the following theorem.

Theorem 5. *Let $\mathcal{C}$ be a binary linear code of length n and $p \in (0, 1/2]$ such that $d_{\min}(\mathcal{C}^*) \geq \delta^* n$ for some $\delta^* \in [0, 1]$. We have asymptotically,*

$$\frac{1}{n}\log_2 \Delta\left(u, f^{\mathcal{C}}_{\mathrm{ber},p}\right) \leq \max_{\delta^* \leq \tau \leq 1-\delta^*/2} \{\frac{1}{2}\min\{c(\delta^*, \tau), d(\delta^*, \tau)\} + \max_{(1-\varepsilon)p \leq \lambda \leq (1+\varepsilon)p} \{\lambda \log_2 p + (1-\lambda)\log_2(1-p) + a(\lambda, \tau)\}\} + O\left(\frac{1}{n}\right)$$

where $a(\cdot,\cdot)$, $c(\cdot,\cdot)$ and $d(\cdot,\cdot)$ are defined respectively in Propositions 5, 7 and 8.

Sketch of proof. We will use the following proof strategy:

1. By Lemma 7 we know that on one hand

$$\Delta\left(u, f^{\mathcal{C}}_{\mathrm{ber},p}\right) = \Delta\left(u, f^{\mathcal{C}}_{\mathrm{truncBer},p}\right) + 2^{-\Omega(n)}. \tag{31}$$

   This is actually a consequence of Chernov's bound. This argument can also be used to show that the Fourier transforms are also close to each other pointwise:

$$\forall \mathbf{x} \in \mathbb{F}_2^n, \quad 2^n \left| \widehat{f_{\mathrm{truncBer},p}}(\mathbf{x}) - \widehat{f_{\mathrm{ber},p}}(\mathbf{x}) \right| = 2^{-\Omega(n)}. \tag{32}$$

2. Equation (32) together with Lemma 1 are then used to show that:

$$\Delta\left(u, f^{\mathcal{C}}_{\mathrm{truncBer},p}\right) \leq 2^n \sqrt{\sum_{t=d_{\min}(\mathcal{C}^*)}^{n-d_{\min}(\mathcal{C}^*)/2} N_t(\mathcal{C}^*), \widehat{f_{\mathrm{truncBer},p}}(t)^2 + 2^{-\Omega(n)}} \tag{33}$$

3. We use the two previous points to upper-bound $\Delta\left(u, f^{\mathcal{C}}_{\mathrm{ber},p}\right)$ as in the equation above and conclude by using bounds of Propositions 7 and 8.

Proof of Step 1. As we explained above (31) is just Lemma 7. Let us now prove that

Lemma 9. *We have,*

$$\forall \mathbf{x} \in \mathbb{F}_2^n, \quad 2^n \left| \widehat{f_{\mathrm{truncBer},p}}(\mathbf{x}) - \widehat{f_{\mathrm{ber},p}}(\mathbf{x}) \right| = 2^{-\Omega(n)}$$

Proof. Recall that $Z = \sum_{|\mathbf{y}| = (1-\varepsilon)np}^{(1+\varepsilon)np} f_{\mathrm{ber},p}(\mathbf{y})$. By Chernoff's bound, we have

$$Z = 1 - 2^{-\Omega(n)} \tag{34}$$

Notice now that,

$$f_{\mathrm{ber},p} = \sum_{r=0}^{n} \binom{n}{r} p^r (1-p)^{n-r} u_r \quad \text{and} \quad f_{\mathrm{truncBer},p} = \frac{1}{Z} \sum_{r=(1-\varepsilon)pn}^{(1+\varepsilon)pn} \binom{n}{r} p^r (1-p)^{n-r} u_r$$

Let $\mathscr{I} \stackrel{\text{def}}{=} [\![(1-\varepsilon)pn, (1+\varepsilon)pn]\!]$. By linearity of the Fourier transform we obtain the following computation:

$$\begin{aligned}\left|\widehat{f_{\text{truncBer},p}}(\mathbf{x}) - \widehat{f_{\text{ber},p}}(\mathbf{x})\right| &= \left(\frac{1}{Z} - 1\right) \sum_{r \in \mathscr{I}} \binom{n}{r} p^r (1-p)^{n-r} \left|\widehat{u_r}(\mathbf{x})\right| \\ &\qquad + \sum_{r \notin \mathscr{I}} \binom{n}{r} p^r (1-p)^{n-r} \left|\widehat{u_r}(\mathbf{x})\right| \\ &= 2^{-\Omega(n)} \sum_{r \in \mathscr{I}} \binom{n}{r} p^r (1-p)^{n-r} \left|\widehat{u_r}(\mathbf{x})\right| + (1-Z)\left|\widehat{u_r}(\mathbf{x})\right| \qquad (35)\end{aligned}$$

where in the last line we used Equation (34). Recall now that by definition of the Fourier transform for functions over $\mathbb{F}_2^n$ we have:

$$|u_r(\mathbf{x})| = \left|\frac{1}{2^n} \sum_{\mathbf{y}:|\mathbf{y}|=r} \frac{(-1)^{\mathbf{x}\cdot\mathbf{y}}}{\binom{n}{r}}\right| \leq \frac{1}{2^n}.$$

By plugging this in Equation (35) we get

$$\begin{aligned}\left|\widehat{f_{\text{truncBer},p}}(\mathbf{x}) - \widehat{f_{\text{ber},p}}(\mathbf{x})\right| &\leq \frac{2^{-\Omega(n)}}{2^n} \underbrace{\sum_{r \in \mathscr{I}} \binom{n}{r} p^r (1-p)^{n-r}}_{\leq 1} + \frac{1}{2^n}(1-Z) \\ &= \frac{2^{-\Omega(n)}}{2^n}.\end{aligned}$$

□

Proof of Step 2. This corresponds to proving the following lemma.

Lemma 10.

$$\Delta\left(u, f^{\mathcal{C}}_{\text{truncBer},p}\right) \leq 2^n \sqrt{\sum_{t=d_{\min}(\mathcal{C}^*)}^{n-d_{\min}(\mathcal{C}^*)/2} N_t(\mathcal{C}^*), \widehat{f_{\text{truncBer},p}}(t)^2} + 2^{-\Omega(n)}$$

Proof of Lemma 10. By applying Proposition 2 to $f_{\text{truncBer},p}$ we obtain

$$\Delta\left(u, f^{\mathcal{C}}_{\text{truncBer},p}\right) \leq 2^n \sqrt{\sum_{t=d_{\min}(\mathcal{C}^*)}^{n} N_t(\mathcal{C}^*) |\widehat{f_{\text{truncBer},p}}(t)|^2} \qquad (36)$$

where $\widehat{f_{\text{truncBer},p}}(t)$ denotes the common value of the radial function $\widehat{f_{\text{truncBer},p}}$ on vectors of Hamming weight t. Recall now that $\widehat{f_{\text{ber},p}}(\mathbf{x}) = \frac{1}{2^n}(1-2p)^{|\mathbf{x}|}$ and by Lemma 9 that $2^n \left|\widehat{f_{\text{truncBer},p}}(\mathbf{x}) - \widehat{f_{\text{ber},p}}(\mathbf{x})\right| = 2^{-\Omega(n)}$. Therefore,

$$\forall \mathbf{x} \in \mathbb{F}_2^n,\ |\mathbf{x}| \geq n - \frac{d_{\min}(\mathcal{C}^*)}{2} \quad : \quad 2^n \left|\widehat{f_{\text{truncBer},p}}(\mathbf{x})\right| = 2^{-\Omega(n)}$$

By plugging this in Equation (36) we obtain (as there is at most one dual codeword of weight ℓ for each $\ell > n - d_{\min}(\mathcal{C}^*)/2$, see Lemma 1):

$$\Delta\left(u, f^{\mathcal{C}}_{\text{truncBer},p}\right) \leq 2^n \sqrt{\sum_{t=d_{\min}(\mathcal{C}^*)}^{n-d_{\min}(\mathcal{C}^*)/2} N_t(\mathcal{C}^*) |\widehat{f_{\text{truncBer},p}}(t)|^2} + 2^{-\Omega(n)} \qquad (37)$$

□

Proof of Step 3. We finish the proof of Theorem 5 by noticing that

$$f_{\mathrm{truncBer},p} = \frac{1}{Z} \sum_{\ell=(1-\varepsilon)pn}^{(1+\varepsilon)pn} \binom{n}{\ell} p^\ell (1-p)^{n-\ell} u_\ell$$

where $Z \stackrel{\text{def}}{=} \sum_{|\mathbf{y}|=(1-\varepsilon)np}^{(1+\varepsilon)np} f_{\mathrm{ber},p}(\mathbf{y}) = 1 - 2^{-\Omega(n)}$ by Chernoff's bound. Therefore,

$$\widehat{f_{\mathrm{truncBer},p}} = \left(1 + 2^{-\Omega(n)}\right) \sum_{\ell=(1-\varepsilon)pn}^{(1+\varepsilon)pn} \binom{n}{\ell} p^\ell (1-p)^{n-\ell} \widehat{u_\ell}$$

By plugging this in Equation (37) and using $\widehat{u_\ell} = \frac{1}{2^n} \frac{K_\ell}{\binom{n}{\ell}}$ we obtain:

$$\Delta\left(u, f^{\mathcal{C}}_{\mathrm{truncBer},p}\right) \leq \left(1 + 2^{-\Omega(n)}\right) \sqrt{\sum_{t=d_{\min}(\mathcal{C}^*)}^{n-d_{\min}(\mathcal{C}^*)/2} N_t(\mathcal{C}^*) \left(\sum_{\ell=(1-\varepsilon)pn}^{(1+\varepsilon)pn} p^\ell (1-p)^{n-\ell} K_\ell(t) \right)^2} + 2^{-\Omega(n)}$$

We then use in the righthand term, Propositions 7, 8 which give bounds on the $\frac{1}{n} \log_2 N_\ell(\mathcal{C}^*)$'s (where $d_{\min}(\mathcal{C}^*) \geq \delta^* n$) and Proposition 5 which gives an asymptotic expansion of Krawtchouk polynomials to upper bound $\Delta\left(u, f^{\mathcal{C}}_{\mathrm{truncBer},p}\right)$. We finish the proof of the theorem by using this upper-bound in the righthand term of (31).