entropy

MDPI

Article

Smoothing of Binary Codes, Uniform Distributions, and Applications

Madhura Pathegama and Alexander Barg *

Department of ECE and Institute for Systems Research, University of Maryland, College Park, MD 20742, USA; pankajap@umd.edu
* Correspondence: abarg@umd.edu

Abstract: The action of a noise operator on a code transforms it into a distribution on the respective space. Some common examples from information theory include Bernoulli noise acting on a code in the Hamming space and Gaussian noise acting on a lattice in the Euclidean space. We aim to characterize the cases when the output distribution is close to the uniform distribution on the space, as measured by the Rényi divergence of order $\alpha \in (1, \infty]$. A version of this question is known as the channel resolvability problem in information theory, and it has implications for security guarantees in wiretap channels, error correction, discrepancy, worst-to-average case complexity reductions, and many other problems. Our work quantifies the requirements for asymptotic uniformity (perfect smoothing) and identifies explicit code families that achieve it under the action of the Bernoulli and ball noise operators on the code. We derive expressions for the minimum rate of codes required to attain asymptotically perfect smoothing. In proving our results, we leverage recent results from harmonic analysis of functions on the Hamming space. Another result pertains to the use of code families in Wyner's transmission scheme on the binary wiretap channel. We identify explicit families that guarantee strong secrecy when applied in this scheme, showing that nested Reed–Muller codes can transmit messages reliably and securely over a binary symmetric wiretap channel with a positive rate. Finally, we establish a connection between smoothing and error correction in the binary symmetric channel.

Keywords: noise operator; uniform distribution; Renyi divergence; wiretap channel

check for updates

Citation: Pathegama, M.; Barg, A. Smoothing of Binary Codes, Uniform Distributions, and Applications. *Entropy* **2023**, *25*, 1515. https://doi.org/10.3390/e25111515

Academic Editor: Igal Sason

Received: 26 August 2023
Revised: 30 October 2023
Accepted: 1 November 2023
Published: 5 November 2023

1. Introduction

Many problems of information theory involve the action of a noise operator on a code distribution, transforming it into some other distribution. For instance, one can think of Bernoulli noise acting on a code in the Hamming space or Gaussian noise acting on a lattice in the Euclidean space. We are interested in characterizing the cases when the output distribution is close to the uniform distribution on the space. Versions of this problem have been considered under different names, including resolvability [1–3], smoothing [4,5], discrepancy [6,7], and the entropy of noisy functions [8–10]. Direct applications of smoothing include secrecy guarantees in both the binary symmetric wiretap channel [2,3,11] and the Gaussian wiretap channel [12,13], error correction in the binary symmetric channel (BSC) [14,15], converse coding theorems of information theory [1,16–18], strong coordination [11,19–22], secret key generation [13,23], and worst-to-average case reductions in cryptography [5,24]. Some aspects of this problem also touch upon approximation problems in statistics and machine learning [25–27].

Our main results are formulated for the smoothing in the binary Hamming space $\mathcal{H}_n$. For $r : \mathcal{H}_n \to \mathbb{R}_0^+$, and $f : \mathcal{H}_n \to \mathbb{R}$ define

$$T_r f(x) = (r * f)(x) := \sum_{z \in \mathcal{H}_n} r(z) f(x - z)$$

as the action of r on the functions on the space. We set r to be a probability mass function (pmf) and call the function $T_r f$ the *noisy version* of f with respect to r, and refer to r and T_r as a *noise kernel* and a *noise operator* , respectively. By *smoothing* f with respect to r, we mean applying the noise kernel r to f. We often assume that $r(x)$ is a radial kernel, i.e., its value on the argument $x \in \mathcal{H}_n$ depends only on the Hamming weight of x.

There are several ways to view the smoothing operation. Interpreting it as a shift-invariant linear operator, we note that, from Young's inequality, $\|T_r f\|_\alpha = \|f * r\|_\alpha \leq \|f\|_\alpha, 1 \leq \alpha \leq \infty$, so smoothing contracts the α-norm. Upon applying T_r, the noisy version of f becomes "flatter"; hence, the designation "smoothing". Note that if f is a pmf, then $T_r f$ is also a pmf, and so this view allows us to model the effect of communication channels with additive noise.

The class of functions that we consider are (normalized) indicators of subsets (codes) in $\mathcal{H}_n$. A code $\mathscr{C} \subset \mathcal{H}_n$ defines a pmf $f_{\mathscr{C}} = \frac{\mathbb{1}_{\mathscr{C}}}{|\mathscr{C}|}$, and, thus, $T_r f_{\mathscr{C}}$ can be viewed as a noisy version of the code (we also sometimes call it a *noisy distribution*) with respect to the kernel r. The main question of interest for us is the proximity of this distribution to U_n, or the "smoothness" of the noisy code distributions. To quantify closeness to U_n, we use the Kullback–Leibler (KL) and Rényi divergences (equivalently, L_α norms), and the smoothness measured in $D_\alpha(\cdot\|\cdot)$ is termed the D_α-smoothness (L_α-smoothness).

We say that a code is *perfectly smoothable* with respect to the noise kernel r if the resultant noisy distribution becomes uniform. Our main emphasis is on the asymptotic version of perfect smoothing and its implications for some of the basic information-theoretic problems. A sequence of codes $(\mathscr{C}_n)_n$ is asymptotically smoothed by the kernel sequence r_n if the distance between $(T_{r_n} f_{\mathscr{C}_n})$ and U_n approaches 0 as n increases. This property is closely related to the more general problem of *channel resolvability* introduced by Han and Verdú in [1]. Given a discrete memoryless channel $\mathscr{W}(Y|X)$ and a distribution P_X, we observe a distribution P_Y on the output of the channel. The task of channel resolvability is to find P_X supported on a subset $\mathscr{C} \subset \mathcal{H}_n$ that approximates P_Y with respect to the KL divergence. As shown in [1], there exists a threshold value of the rate such that it is impossible to approximate P_Y using codes of lower rate, while any output process can be approximated by a well-chosen code of a rate larger than the threshold. Other proximity measures between distributions were considered for this problem in [3,28,29]. Following the setting in [3], we consider Rényi divergences for measuring the closeness to uniformity. We call the minimum rate required to achieve perfect asymptotic smoothing the D_α*-smoothing capacity* of the noise kernels $(r_n)_n$, where the proximity to uniformity is measured by the α-Rényi divergence. In this work, we characterize the D_α-smoothing capacity of the sequence $(r_n)_n$ using its Rényi entropy rate.

Asymptotic smoothing. We will limit ourselves to studying smoothing bounds under the action of the Bernoulli noise or ball noise kernels, defined formally below. A common approach to deriving bounds on the norm of a noisy function is through hypercontractivity inequalities [30–32]. In its basic version, given a code $\mathscr{C}$ of size M, it yields the estimate

$$\|T_\delta f_{\mathscr{C}}\|_\alpha \leq \|f_{\mathscr{C}}\|_{\alpha'} = M^{\frac{1-\alpha'}{\alpha'}} 2^{-\frac{n}{\alpha'}},$$

where T_δ is the Bernoulli kernel (see Section 2 for formal definitions) and $\alpha' = 1 + (1 - 2\delta)^2(\alpha - 1)$. This upper bound does not differentiate codes yielding higher or lower smoothness, which in many situations may not be sufficiently informative. Note that other tools, such as "Mrs. Gerber's lemma" [30,33] or strong data-processing inequalities, also suffer from the same limitation.

A new perspective of the bounds for smoothing has recently been introduced in the works of Samorodnitsky [8–10]. Essentially, his results imply that codes satisfying certain regularity conditions have good smoothing properties. Their efficiency is highlighted in recent papers [14,34], which leveraged results for code performance on the binary erasure channel (BEC) to prove strong claims about the error correction capabilities of the codes when used on the BSC. Using Samorodnitsky's inequalities, we show that the duals of some

BEC capacity-achieving codes achieve D_α-smoothing capacity for $\alpha \in \{2,3,\dots,\infty\}$ with respect to the Bernoulli noise. This includes the duals of polar codes and doubly transitive codes, such as the Reed–Muller (RM) codes.

Smoothing and the wiretap channel. Wyner's wiretap channel [35] models communication in the presence of an eavesdropper. Code design for this channel pursues reliable communication between the legitimate parties, while at the same time leaking as little information as possible about the transmitted messages to the eavesdropper. The connection between secrecy in wiretap channels and resolvability was first mentioned by Csiszár [36] and later developed by Hayashi [2]. It rests on the observation that to achieve secrecy it suffices to make the distribution of an eavesdropper's observations conditioned on the transmitted message nearly independent of the message. The idea of characterizing secrecy based on smoothness works irrespective of the measure of secrecy [2,3,11], and it was also employed for nested lattice codes used over the Gaussian wiretap channel in [12].

Secrecy on the wiretap channel can be defined in two ways, measured by the information gained by the eavesdropper, and it depends on whether this quantity is normalized to the number of channel uses (weak secrecy) or not (strong secrecy). This distinction was first highlighted by Maurer [37], and it has been adopted widely in the recent literature. Early papers devoted to code design for the wiretap channel relied on random codes, but, for simple channel models such as BSC or BEC, this has changed with the advent of explicit capacity-approaching code families. Weak secrecy results based on LDPC codes were presented in [38], but initial attempts to attain strong secrecy encountered some obstacles. To circumvent this, the first works on code construction [39,40] had to assume that the main channel is noiseless. The problem of combining strong secrecy and reliability for general wiretap channels was resolved in [41], but that work had to assume that the two communicating parties share a small number of random bits unavailable to the eavesdropper. Apart from the polar coding scheme of [41], explicit code families that support reliable communication with positive rate and strong secrecy have not previously appeared in the literature. In this work, we show that nested RM codes perform well in binary symmetric wiretap channels based on their smoothing properties. While our work falls short of proving that nested RM codes achieve capacity, we show that they can transmit messages reliably and secretly at rates close to capacity.

Ball noise and decoding error. Ball-noise smoothing provides a tool for estimating the error probability of decoding on the BSC. We derive impossibility and achievability bounds for the D_α-smoothness of noisy distributions with respect to the ball noise. Smoothing of a code with respect to the L_2 norm plays a special role because, in this case, the second norm (the variance) of the resulting distribution can be expressed via the pairwise distance between codewords, enabling one to rely on tools from Fourier analysis. The recent paper by Debris-Alazard et al. [4] established universal bounds for the smoothing of codes or lattices, with cryptographic reductions in mind. The paper by Sprumont and Rao [15] addressed bounds for error probability of list decoding at rates above BSC capacity. A paper by one of the present authors [42] studied the variance of the number of codewords in balls of different radii (a quantity known as the quadratic discrepancy [43,44]).

The main contributions of this paper are the following:

1. Characterizing the D_α-smoothing capacities of noise operators on the Hamming space for $\alpha \in (1,\infty]$.
2. Identifying some explicit code families that attain a smoothing capacity of the Bernoulli noise for $\alpha \in \{2,3,\dots,\infty\}$;
3. Obtaining rate estimates for the RM codes used on the BSC wiretap channel under the strong secrecy condition;
4. Showing that codes possessing sufficiently good smoothing properties are suitable for error correction.

In Section 2, we set up the notation and introduce the relevant basic concepts. Then, in Section 3, we derive expressions for the D_α-smoothing capacities for $\alpha \in (1,\infty]$, and in Section 4, we use these results to analyze the smoothing of code families under the action of the Bernoulli noise. Section 5 is devoted to the application of these results for the binary symmetric wiretap channel. In particular, we show that RM codes can achieve rates close to the capacity of the BSC wiretap channel, while at the same time guaranteeing strong secrecy. In Section 6, we establish threshold rates for smoothing under ball noise, and derive bounds for the error probability of decoding on the BSC, including the list case, based on the distance distribution. Concluding the paper, Section 7 briefly points out that the well-known class of uniformly packed codes are perfectly smoothable with respect to "small" noise kernels.

2. Preliminaries

2.1. Notation

Throughout this paper, $\mathcal{H}_n$ is the binary n-dimensional Hamming space

Balls and spheres. Denote by $B(x,t) := \{y \in \mathcal{H}_n : |y-x| \le t\}$ the metric ball of radius t in $\mathcal{H}_n$ with center at x, and denote by $S(x,t) := \{y \in \mathcal{H}_n : |y-x| = t\}$ the sphere of radius t. Let $V_t = |B(x,t)|$ be the volume of the ball, and let $\mu_t(i)$ be the intersection volume of two balls of radius t whose centers are distance i apart:

$$\mu_t(i) = |B(0,t) \cap B(x,t)|, \quad \text{where } |x| = i. \tag{1}$$

Codes and distributions. A code $\mathscr{C}$ is a subset in $\mathcal{H}_n$. The rate and distance of the code are denoted by $R(\mathscr{C}) := \log|\mathscr{C}|/n$ and $d(\mathscr{C})$, respectively. Let

$$A_i = \frac{1}{|\mathscr{C}|}|\{(x,y) \in \mathscr{C}^2 : d(x,y) = i\}| \tag{2}$$

and let $(A_i, i = 0,\dots,n)$ be the distance distribution of the code. If the code $\mathscr{C}$ forms an $\mathbb{F}_2$-linear subspace in $\mathcal{H}_n$, we denote by $\mathscr{C}^\perp := \{y \in \mathcal{H}_n : \sum_i x_i y_i = 0 \text{ for all } x \in \mathscr{C}\}$ its dual code.

The function $\mathbb{1}_{\mathscr{C}}$ denotes the indicator of a subset $\mathscr{C} \subset \mathcal{H}_n$, and $f_{\mathscr{C}} = \frac{\mathbb{1}_{\mathscr{C}}}{|\mathscr{C}|}$ is the corresponding pmf denoting the uniform distribution over the set, calling it a *code distribution*. Let b_t denote the uniform distribution on the ball $B_{0,t}$, given by $b_t(x) = \frac{\mathbb{1}_{B_{0,t}}}{V_t}$. In the context of noise operators, we refer to T_{b_t} as the *ball noise*. Finally, β_δ is the binomial distribution on $\mathcal{H}_n$, given by

$$\beta_\delta(x) = \beta_\delta^{(n)}(x) = \delta^{|x|}(1-\delta)^{n-|x|}, \tag{3}$$

and U_n is the uniform distribution, given by $U_n(x) = 2^{-n}$ for all x.

Entropies and norms. For a function $f : \mathcal{H}_n \to \mathbb{R}$, we define its α-norm as follows.

$$\|f\|_\alpha = \Big(\frac{1}{2^n}\sum_{x \in \mathcal{H}_n} |f(x)|^\alpha\Big)^{1/\alpha} \text{ for } \alpha \in (0,\infty)$$
$$\|f\|_\infty = \max_{x \in \mathcal{H}_n} |f(x)|.$$

Given a pmf P, let

$$H(P) = -\sum_i P_i \log P_i, \tag{4}$$

$$H_\alpha(P) = \frac{1}{1-\alpha}\log\Big(\sum_i P_i^\alpha\Big) \tag{5}$$

denote its Shannon entropy and Rényi entropy of order α, respectively. If P is supported on two points, we write $h(P)$ and $h_\alpha(P)$ instead (all logarithms are to the base 2). The limiting cases of $\alpha = 0, 1, \infty$ are well-defined; in particular, for $\alpha = 1$, $H_\alpha(P)$ reduces to $H(P)$.

For two discrete probability distributions P and Q, the *α-Rényi divergence* (or simply the α-divergence) is defined as follows:

$$D_\alpha(P\|Q) = \begin{cases} -\log Q(\{i : P_i > 0\}) & \text{if } \alpha = 0 \\ \frac{1}{\alpha-1}\log\sum_i P_i^\alpha Q_i^{-(\alpha-1)} & \text{if } \alpha \in (0,1)\cup(1,\infty) \\ \sum_i P_i \log\frac{P_i}{Q_i} & \text{if } \alpha = 1 \\ \max_i \log\frac{P_i}{Q_i} & \text{if } \alpha = \infty \end{cases}. \tag{6}$$

The divergence $D_\alpha(P\|Q)$ is a continuous function of α for $\alpha \in [0,\infty]$. For a pmf f on $\mathcal{H}_n$

$$D_\alpha(f\|U_n) = \frac{\alpha}{\alpha-1}\log\|2^n f\|_\alpha, \quad \alpha \in (0,1)\cap(1,\infty) \tag{7}$$

$$D_\infty(f\|U_n) = \log\|2^n f\|_\infty. \tag{8}$$

Note that $D_\alpha(f\|U_n) = n - H_\alpha(f)$ for all $0 \le \alpha \le \infty$.

Channels. In this paper, a channel is a conditional probability distribution $\mathscr{W} : \{0,1\} \to \mathscr{Y}$, where $\mathscr{Y}$ is a finite set, so that $\mathscr{W}(y|x)$ is the conditional probability of the output y for the input x. We frequently consider the binary symmetric channel with crossover probability δ and the binary erasure channel with erasure probability λ, abbreviating them as BSC(δ) and BEC(λ), respectively. We are often interested in the n-fold channel $\mathscr{W}^{(n)}$, i.e., the conditional probability distribution corresponding to n-uses of the channel. For the input X, let $Y_{(X,\mathscr{W})}$ be the random output of the channel $\mathscr{W}^{(n)}$. If the input sequences are chosen from a uniform distribution on a code $\mathscr{C}$, we denote the input by $X_{\mathscr{C}}$. Since the number of uses of the channel is usually clear from the context, we suppress the dependency on n from the notation for channels and sequences.

Let $\mathscr{C}$ be a code of length n. For a channel $\mathscr{W}$ and input $X_{\mathscr{C}}$, the block-MAP decoder is defined as

$$\hat{x}(y) = \operatorname*{argmax}_{x\in\mathscr{C}} \Pr(x|y).$$

For a given code and channel, denote the error probability of the block-MAP decoding by

$$P_B(\mathscr{W},\mathscr{C}) = \Pr(X_{\mathscr{C}} \neq \hat{X}(Y_{(X_{\mathscr{C}},\mathscr{W})})).$$

2.2. D_α- and L_α-Smoothness

Recall that in the introduction, we expressed the smoothness of a distribution as its proximity to uniformity. Here, we formalize this notion based on two (equivalent) proximity measures.

Let g be a pmf on $\mathcal{H}_n$. A natural measure of the uniformity of g is $D_\alpha(g\|U_n)$ ($\alpha \in [0,\infty]$). We call this the D_α-smoothness of g. Observe that

$$\|2^n g\|_\alpha = \frac{\|g\|_\alpha}{\|g\|_1} \ge 1 \quad \text{for } \alpha \in (1,\infty], \text{ and} \tag{9}$$

$$\|2^n g\|_\alpha = \frac{\|g\|_\alpha}{\|g\|_1} \le 1 \quad \text{for } \alpha \in (0,1) \tag{10}$$

with equality iff $g = U_n$. Thus, the better the pmf g approximates uniformity, the closer is $\|2^n g\|_\alpha$ to 1 (the denominator is simply a normalization quantity that allows dimension-agnostic analysis). Therefore, $\|2^n g\|_\alpha$ ($\alpha \in (0,1)\cup(1,\infty]$) can be considered as another measure of proximity. We call $\|2^n g\|_\alpha$ the L_α-smoothness of g. From (7) and (8), it follows that the D_α-smoothness and L_α-smoothness are equivalent.

Remark 1. *It is easily seen that $D_\alpha(g\|U_n) = n - H_\alpha(g)$; hence, $D_\alpha(g\|U_n)$ is an increasing function of α.*

Recall that for a given code $\mathscr{C}$, and a noise kernel r, $T_r f_{\mathscr{C}} = r * f_{\mathscr{C}}$ is the noisy distribution of code $\mathscr{C}$ with respect to r. We intend to study the smoothing properties of such noisy distributions of codes. In particular, we characterize the necessary conditions for $D_\alpha(T_r f_{\mathscr{C}}\|U_n)$ to be close to zero (equivalently, for $\|2^n T_r f_{\mathscr{C}}\|_\alpha$ close to one). In Section 3, we quantify these requirements in the asymptotic setting.

2.3. Resolvability

The problem of channel resolvability was introduced by Han and Verdú [1] under the name of approximating the output statistics of the channel. The objective of channel resolvability is to approximate the output distribution of a given input by the output distribution of a code with a smaller support size. In this work, we are interested in code families whose noisy distributions approximate uniformity. Resolvability characterizes the necessary conditions for this to happen in terms of the rate of the code.

Let $\mathscr{W}$ be a (discrete memoryless) channel whose input alphabet is $\mathcal{X}$ and whose output alphabet is $\mathscr{Y}$. Let $\boldsymbol{X} = \{X_n\}_{n=1}^\infty$ be a discrete-time random process where the RVs X_n take values in $\mathcal{X}$. Denote by Y_n the random output of $\mathscr{W}$ with input X_n and let $\boldsymbol{Y} = \{Y_n\}_{n=1}^\infty$. Denote by $P_{\boldsymbol{Y}}$ the distribution of $\boldsymbol{Y}$ and let $P_{Y^{(n)}}$ be the pmf of the n-tuple $Y^{(n)} := \{Y_1, Y_2, \dots, Y_n\}$.

For a legitimate (realizable) output process $\boldsymbol{Y}$, define

$$J^{(\Delta)}(\mathscr{W}, P_{\boldsymbol{Y}}) = \inf_{\mathscr{C}_n \subset \mathcal{X}^n} \{\liminf_n R(\mathscr{C}_n) : \Delta(f_{\mathscr{C}_n}, P_{Y^{(n)}}) \to 0\}, \tag{11}$$

where Δ is a measure of closeness of a pair of probability distributions. In words, we look for sequences of distributions $(f_{\mathscr{C}_n})_n$ of the smallest possible rate that approximates $P_{\boldsymbol{Y}}$ on the output of $\mathscr{W}$.

The original problem as formulated by Han and Verdú in [1] seeks to find the *resolvability* of the channel, defined as

$$C_r^{(\Delta)}(\mathscr{W}) = \inf_{P_{\boldsymbol{Y}}} \{J^{(\Delta)}(\mathscr{W}, P_{\boldsymbol{Y}}) : \boldsymbol{Y} \text{ is an output process over } \mathscr{W}\}. \tag{12}$$

where Δ is either the variational distance or the normalized KL divergence $\frac{1}{n}D(\cdot\|\cdot)$. Hayashi [2] considered the same problem where the proximity was measured by the unnormalized KL divergence. In each case, the resolvability equals the Shannon capacity of the channel $\mathscr{W}$.

Theorem 1 ([1,2])**.** *Let $\mathscr{W}$ be a discrete memoryless channel. Suppose that Δ is either the KL divergence (normalized or not) or the variational distance; then, the resolvability is given by*

$$C_r^{(\Delta)}(\mathscr{W}) = C(\mathscr{W}),$$

where $C(\mathscr{W})$ is the Shannon capacity of the channel.

The authors of [1] proved this result under the additional assumption that the channel $\mathscr{W}$ satisfies strong converse, and Hayashi [2] later showed that this assumption is unessential.

In addition to the proximity measures considered in Theorem 1, the papers [3,28,29] considered other possibilities. In particular, Yu and Tan [3] studied the resolvability problem for a specific target distribution $P_{\boldsymbol{Y}}$ and for the Rényi divergence $\Delta = D_\alpha$ (6). Their main result is as follows.

Theorem 2 ([3], Theorem 2)**.** *Let $\mathscr{W}$ be a channel and $P_{\boldsymbol{Y}}$ be an output distribution. then*

$$J^{(D_\alpha)}(\mathscr{W}, P_{\boldsymbol{Y}}) = \begin{cases} \min_{P_X \in \mathcal{P}(\mathscr{W}, P_Y)} \sum_x P_X(x) D_\alpha(\mathscr{W}(\cdot|x) \| P_{\boldsymbol{Y}}) & \textit{if } \alpha \in (1,2] \cup \{\infty\} \\ \min_{P_X \in \mathcal{P}(\mathscr{W}, P_Y)} D(\mathscr{W} \| P_{\boldsymbol{Y}} | P_X) & \textit{if } \alpha \in (0,1] \\ 0 & \textit{if } \alpha = 0, \end{cases}$$

where $\mathcal{P}(\mathscr{W}, P_{\boldsymbol{Y}})$ is the set of distributions $P_{\boldsymbol{X}}$ consistent with the output $P_{\boldsymbol{Y}}$.

A direct corollary of Theorem 2 is the following:

Corollary 1 ([3], Equation (55))**.** *Let $\boldsymbol{Y}^*$ be the output process where for each n, $Y_n^* \sim \mathrm{Ber}(1/2)$. Then,*

$$J^{(D_\alpha)}(\mathrm{BSC}(\delta), P_{\boldsymbol{Y}^*}) = \begin{cases} 1 - h_\alpha(\delta) & \textit{if } \alpha \in (1,2] \cup \{\infty\} \\ 1 - h(\delta) & \textit{if } \alpha \in (0,1] \\ 0 & \textit{if } \alpha = 0. \end{cases}$$

This corollary gives necessary conditions for the rate of codes that can approximate the uniform distribution via smoothing. We will connect this result to the problem of finding smoothing thresholds in Section 4.

3. Perfect Smoothing—The Asymptotic Case

For a given family of noise kernels $(T_{r_n})_n$, there exists a threshold rate such that it is impossible to approximate uniformity with codes of rate below the threshold irrespective of the chosen code, while at the same time, there exist families of codes with a rate above the threshold that allows perfect approximation in the limit of infinite length. For instance, for the Bernoulli(δ) noise applied to a code $\mathscr{C}$, the smoothed distribution is nonuniform unless $\mathscr{C} = \mathcal{H}_n$ or $\delta = 1/2$. At the same time, it is possible to approach the uniform distribution asymptotically for large n once the code sequence satisfies certain conditions. Intuitively, it is clear that, for a fixed noise kernel, it is easier to approximate uniformity if the code rate is sufficiently high. In this section, we characterize the threshold rate for (asymptotically) perfect smoothing. Of course, the threshold also depends on the proximity measure Δ that we are using. In this section, we use perfect smoothing to mean "asymptotically perfect". If the proximity measure Δ for smoothing is not specified, this means that we are using the KL divergence. We obtain the threshold rates for perfect smoothing measured with respect to the α-divergence for several values of α. In the subsequent sections, we work out the details for the Bernoulli and ball noise operators, which also have some implications for communication problems.

Definition 1. *Let $(\mathscr{C}_n)_n$ be a sequence of codes of increasing length n and let $0 \le \alpha \le \infty$. We say that the sequence $\mathscr{C}_n$ is asymptotically perfectly D_α-smoothable with respect to the noise kernels r_n if*

$$\lim_{n\to\infty} D_\alpha(T_{r_n} f_{\mathscr{C}_n} \| U_n) = 0,$$

or equivalently (7) *and* (8) *if*

$$\lim_{n\to\infty} \|2^n T_{r_n} f_{\mathscr{C}_n}\|_\alpha = 1 \quad (\alpha \neq 0, 1).$$

One can also define a dimensionless measure for perfect asymptotic smoothing by considering the limiting process

$$\frac{\|T_{r_n} f_{\mathscr{C}_n} - U_n\|_\alpha}{\|T_{r_n} f_{\mathscr{C}_n}\|_1} = 2^n \|T_{r_n} f_{\mathscr{C}_n} - U_n\|_\alpha \to 0. \tag{13}$$

Proposition 1. *Convergence in* (13) *implies perfect smoothing for all* $1 < \alpha \le \infty$ *and is equivalent to it for* $\alpha \neq \infty$.

Proof. Let $\mathscr{C} = \mathscr{C}_n \subset \mathcal{H}_n$ for some fixed n. Since by the triangle inequality,

$$2^n \|T_r f_{\mathscr{C}}\|_\alpha - 1 \le 2^n \|T_r f_{\mathscr{C}} - U_n\|_\alpha,$$

(13) is not weaker than the mode of convergence in Definition 1 for all $\alpha \in [1, \infty]$. For $\alpha \neq 1, \infty$, we use Clarkson's inequalities ([45], p. 388). Their form depends on α; namely, for $2 \le \alpha < \infty$, we have

$$\begin{aligned} 1 + \left\| \frac{2^n T_r f_{\mathscr{C}} - 1}{2} \right\|_\alpha^\alpha &\le \left\| \frac{2^n T_r f_{\mathscr{C}} + 1}{2} \right\|_\alpha^\alpha + \left\| \frac{2^n T_r f_{\mathscr{C}} - 1}{2} \right\|_\alpha^\alpha \\ &\le \frac{1}{2}(\|2^n T_r f_{\mathscr{C}}\|_\alpha^\alpha + 1). \end{aligned}$$

For $1 < \alpha < 2$, the inequality has the form

$$\begin{aligned} 1 + \left\| \frac{2^n T_r f_{\mathscr{C}} - 1}{2} \right\|_\alpha^{\alpha'} &\le \left\| \frac{2^n T_r f_{\mathscr{C}} + 1}{2} \right\|_\alpha^{\alpha'} + \left\| \frac{2^n T_r f_{\mathscr{C}} - 1}{2} \right\|_\alpha^{\alpha'} \\ &\le \left(\frac{1}{2}(\|2^n T_r f_{\mathscr{C}}\|_\alpha^\alpha + 1) \right)^{\alpha'/\alpha}, \end{aligned}$$

where $\alpha' = \frac{\alpha}{\alpha - 1}$ is the Hölder conjugate. These equations show that, for $\alpha \in (1, \infty)$, $\|2^n T_{r_n} f_{\mathscr{C}_n}\|_\alpha \to 1$ implies $\|2^n T_{r_n} f_{\mathscr{C}_n} - 1\|_\alpha \to 0$, establishing the claimed equivalence. □

Definition 2. *Let* $(r_n)_n$ *be a sequence of noise kernels. We say that the rate* R *is achievable for perfect* D_α*-smoothing if there exists a sequence of codes* $(\mathscr{C}_n)_n$ *such that* $R(\mathscr{C}_n) \to R$ *as* $n \to \infty$ *and* $(\mathscr{C}_n)_n$ *is perfectly* D_α*-smoothable.*

Note that if R_1 is achievable, then any rate $1 \ge R_2 > R_1$ is also achievable. Indeed, consider a (linear) code $\mathscr{C}_1$ of rate R_1 that has good smoothing properties. Construct $\mathscr{C}_2$ by taking the union of $2^{n(R_2 - R_1)}$ non-overlapping shifts of $\mathscr{C}_1$. Then the rate of $\mathscr{C}_2$ is R_2, and since each shift has good smoothing properties, the same is true for $\mathscr{C}_2$. Therefore, let us define the main concept of this section.

Definition 3. *Given a sequence of kernels* $r = (r_n)_n$*, define the* D_α*-smoothing capacity as*

$$S_\alpha^r := \inf_{(\mathscr{C}_n)_n} \{\liminf_{n \to \infty} R(\mathscr{C}_n) : \lim_{n \to \infty} D_\alpha(T_{r_n} f_{\mathscr{C}_n} \| U_n) = 0\}. \tag{14}$$

Note that this quantity is closely related to the resolvability: if, rather than optimizing on the output process in (12), we set the output distribution to uniform and take $\Delta = D_\alpha$, then S_α^r equals $J^{(D_\alpha)}(\mathscr{W}, P_y)$ for the channel $\mathscr{W}$ given by the noise kernel r. To avoid future confusion, we refer to the capacity of reliable transmission as Shannon's capacity.

The following lemma provides a lower bound for D_α-smoothness. It follows from Lemma 2 in [3], and we give a direct proof for completeness.

Lemma 1. *Let* $\mathscr{C} \subset \mathcal{H}_n$ *be a code of size* $M = 2^{nR}$ *and let* r *be a noise kernel. Then, for* $\alpha \in [0, \infty]$

$$D_\alpha(T_r f_{\mathscr{C}} \| U_n) \ge n(1 - R) - H_\alpha(r).$$

Proof. We will first prove that $\|2^n T_r f_{\mathscr{C}}\|_\alpha^\alpha \geq 2^{(\alpha-1)[n(1-R)-H_\alpha(r)]}$ for $\alpha \in (1,\infty)$:

$$\begin{aligned}
\|2^n T_r f_{\mathscr{C}}\|_\alpha^\alpha &= \frac{2^{n\alpha}}{2^n} \sum_{x\in\mathcal{H}_n} T_r f_{\mathscr{C}}(x)^\alpha \\
&= 2^{n(\alpha-1)} \sum_{x\in\mathcal{H}_n} [\sum_{y\in\mathcal{H}_n} r(y) f_{\mathscr{C}}(x-y)]^\alpha \\
&\geq 2^{n(\alpha-1)} \sum_{x\in\mathcal{H}_n} \sum_{y\in\mathcal{H}_n} [r(y) f_{\mathscr{C}}(x-y)]^\alpha \\
&= 2^{n(\alpha-1)} \sum_{y\in\mathcal{H}_n} r(y)^\alpha \sum_{x\in\mathcal{H}_n} f_{\mathscr{C}}(x-y)^\alpha \\
&= \frac{2^{n\alpha}}{|\mathscr{C}|^{(\alpha-1)}} \|r\|_\alpha^\alpha \\
&= 2^{(\alpha-1)[n(1-R)-H_\alpha(r)]}.
\end{aligned}$$

Together with (7), this implies that the claimed inequality holds for $\alpha \in (1,\infty)$.

A similar calculation shows that for $\alpha \in (0,1)$, $\|2^n T_r f_{\mathscr{C}}\|_\alpha^\alpha \leq 2^{(\alpha-1)[n(1-R)-H_\alpha(r)]}$, yielding the claim for $\alpha \in (0,1)$. The limiting cases $\alpha = 0$, $\alpha = 1$, and $\alpha = \infty$ follow by continuity of D_α and H_α for all $\alpha \geq 0$. □

Define

$$\pi(\alpha) = \liminf_{n\to\infty} \frac{H_\alpha(r_n)}{n} \tag{15}$$

Lemma 1 shows that it is impossible to achieve perfect D_α-smoothing if $R < 1 - \pi(\alpha)$. A question of interest is whether there exist sequences of codes of $R > 1 - \pi(\alpha)$ that achieve perfect D_α-smoothing. The next theorem shows that this is the case for $\alpha \in (1,\infty]$.

Theorem 3. *Let $r = (r_n)_n$ be a sequence of noise kernels and let $\alpha \in (1,\infty]$. Then,*

$$S_\alpha^r = 1 - \pi(\alpha). \tag{16}$$

The proof relies on a random coding argument and is given in Appendix B. This result will be used below to characterize the smoothing capacity of the Bernoulli and ball noise operators.

Remark 2. *Equality (16) does not hold in the case $\alpha \in [0,1]$. From Theorem 4 below, the Bernoulli noise does not satisfy (16) for $\alpha \in [0,1)$. To construct a counterexample for $\alpha = 1$, consider the noise kernel that is almost uniform except for one distinguished point, for instance, $r_n(x) = 2^{-(n+1)}$ for $x \neq 0$ and $r_n(0) = \frac{1}{2} + \frac{1}{2^{n+1}}$. Performing the calculations, we then obtain that $S_1^r = 1$ while $\pi(1) = \frac{1}{2}$.*

Remark 3. *It is worth noting that $\pi(\alpha)$ is a decreasing function of α for $0 \leq \alpha \leq \infty$.*

4. Bernoulli Noise

In this section, we characterize the value $S_\alpha^{\beta_\delta}$ for a range of values of α. Then, we provide explicit code families that attain the D_α-smoothing capacities.

As already mentioned, the resolvability for β_δ with respect to α-divergence was considered by Yu and Tan [3]. Their results, stated in Corollary 1, yield an expression for $S_\alpha^{\beta_\delta}$ for $\alpha \in [0,2] \cup \{\infty\}$. The next theorem summarizes the current knowledge about $S_\alpha^{\beta_\delta}$, where the claims for $2 < \alpha < \infty$ form new results.

Theorem 4.

$$S_\alpha^{\beta_\delta} = \begin{cases} 0 & \textit{if } \alpha = 0 \\ 1 - h(\delta) & \textit{if } \alpha \in (0,1] \\ 1 - h_\alpha(\delta) & \textit{if } \alpha \in (1,\infty]. \end{cases} \tag{17}$$

Proof. The claims for $\alpha \in [0,1]$ follow from Corollary 1. The results for $\alpha = (1,\infty]$ follow from Theorem 3 since $\frac{H_\alpha(\beta_\delta)}{n} = h_\alpha(\delta)$. □

Having quantified the smoothing capacities, let us examine the code families with strong smoothing properties. Since the D_1-smoothing capacity and the Shannon capacity coincide, it is natural to speculate that codes that achieve the Shannon capacity when used on the BSC(δ) would also attain the D_1-smoothing capacity. However, the following result demonstrates that the capacity-achieving codes do not yield perfect smoothing. For typographical reasons, we abbreviate T_{β_δ} by T_δ from this section onward.

Proposition 2. *Let $\mathscr{C}_n$ be a sequence of codes achieving a capacity of* BSC(δ)*. Then,*

$$D(T_\delta f_{\mathscr{C}_n} \| U_n) \to \infty, \quad D(T_\delta f_{\mathscr{C}_n} \| U_n) = o(n).$$

Proof. The second part of the statement is Theorem 2 in [46]. The first part is obtained as follows: Let $\mathscr{C}_n$ be a capacity-achieving sequence of codes in BSC(δ). Then, from [47] (Theorem 49), there exists a constant $K > 0$ such that $nR(\mathscr{C}_n) \le n(1 - h(\delta)) - K\sqrt{n}$ for large n. Therefore,

$$0 \le H(X_{\mathscr{C}_n} | Y_{\mathrm{BSC}(\delta),X}) = n(R(\mathscr{C}_n) + h(\delta) - 1) + D(T_\delta f_{\mathscr{C}_n} \| U_n),$$

which implies $D(T_\delta f_{\mathscr{C}_n} \| U_n) \ge K\sqrt{n}$. □

Apart from random codes, only polar codes are known to achieve D_1-smoothing capacity. Before stating the formal result, recall that polar codes are formed by applying several iterations of a linear transformation to the input, which results in creating virtual channels for individual bits with Shannon's capacity close to zero or to one, plus a vanishing proportion of intermediate-capacity channels. While by Proposition 2, that polar codes that achieve the BSC capacity cannot achieve D_1-smoothing capacity, adding some intermediate-bit channels to the set of data bits makes this possible. This idea was first introduced in [39] and expressed in terms of resolvability in [48].

Theorem 5 ([48], Proposition 1)**.** *Let $\mathscr{W}$ be the* BSC(δ) *channel and $\mathscr{W}_n^{(i)}$ be the virtual channels formed after applying n steps of the polarization procedure. For $\gamma \in (0,1/2)$, define $\mathcal{G}_n = \{i \in \{1,\dots,n\} : C(\mathscr{W}_n^{(i)}) \ge 2^{-n^\gamma}\}$. Let $\mathscr{C}_n$ be the polar code corresponding to the virtual channels $\mathcal{G}_n$. Then, $D(T_\delta f_{\mathscr{C}_n} \| U_n) \to 0$.*

Note that $\lim_{n\to\infty} R(\mathscr{C}_n) = \lim_{n\to\infty} \frac{|\mathcal{G}_n|}{n} = 1 - h(\delta)$. Hence, the polar code construction presented above achieves the perfect smoothing threshold with respect to the KL divergence. Furthermore, since the convergence in the α divergence for $\alpha < 1$ is weaker than the convergence in $\alpha = 1$, the same polar code sequence is perfectly D_α-smoothable for $\alpha < 1$. Noting that the smoothing threshold for $\alpha < 1$ is $1 - h(\delta)$ by Theorem 4, we conclude that the above polar code sequence achieves smoothing capacity in α-divergence for $\alpha < 1$.

As mentioned earlier, the smoothing properties of code families other than random codes and polar codes have not been extensively studied. We show that the duals of capacity-achieving codes in the BEC exhibit good smoothing properties using the tools developed in [10]. As the first step, we establish a connection between the smoothing of a generic linear code and the erasure correction performance of its dual code.

Lemma 2. *Let $\mathscr{C}$ be a linear code and let $X_{\mathscr{C}^\perp}$ be a random uniform codeword of $\mathscr{C}^\perp$. Let $Y_{X_{\mathscr{C}^\perp},\mathrm{BEC}(\lambda)}$ be the output of the erasure channel* $\mathrm{BEC}(\lambda)$ *for the input $X_{\mathscr{C}^\perp}$. Then,*

$$D_\alpha(T_\delta f_{\mathscr{C}} \| U_n) \le H(X_{\mathscr{C}^\perp} | Y_{X_{\mathscr{C}^\perp},\mathrm{BEC}(\lambda)}), \tag{18}$$

where $\lambda = (1-2\delta)^2$ for $\alpha = 1$ and $\lambda = 1 - h_\alpha(\delta)$ for $\alpha \in \{2, 3, \dots, \infty\}$.

The proof is given in Appendix D.

Using this lemma, we show that the duals of the BEC capacity-achieving codes (with growing distance) exhibit good smoothing properties. In particular, they achieve D_α-smoothing capacities for $\alpha \in \{2, 3, \dots, \infty\}$.

Theorem 6. *Let $(\mathscr{C}_n)_n$ be a sequence of linear codes with rate $R_n \to R$. Suppose that the dual sequence $(\mathscr{C}_n^\perp)_n$ achieves Shannon's capacity of the BEC(λ) with $\lambda = R$, and assume that $d(\mathscr{C}_n^\perp) = \omega(\log n)$. If $R > (1-2\delta)^2$, then,*

$$\lim_{n\to\infty} D(T_\delta f_{\mathscr{C}_n} \| U_n) = 0.$$

Additionally, for $\alpha \in \{2, 3, \dots, \infty\}$, if $R > 1 - h_\alpha(\delta)$, then,

$$\lim_{n\to\infty} D_\alpha(T_\delta f_{\mathscr{C}_n} \| U_n) = 0.$$

In particular, the sequence $\mathscr{C}_n$ achieves D_α-smoothing capacity $S_\alpha^{\beta_\delta}$ for $\alpha \in \{2, 3, \dots, \infty\}$.

Proof. Since the dual codes achieve the capacity of the BEC, it follows from ([49], Theorem 5.2) that, if their distance grows with n, then their decoding error probability vanishes. In particular, if $d(\mathscr{C}_n^\perp) = \omega(\log(n))$, then, $P_B(\mathrm{BEC}(R-\epsilon), \mathscr{C}_n^\perp) = o(\frac{1}{n})$ for all $\epsilon \in (0, R]$. Hence, from Fano's inequality,

$$\lim_{n\to\infty} H(X_{\mathscr{C}_n^\perp} | Y_{X_{\mathscr{C}_n^\perp},\mathrm{BEC}(R-\epsilon)}) = 0.$$

Now, if $R > (1-2\delta)^2$, then there exists ϵ_0 such that $R - \epsilon_0 = (1-2\delta)^2$. Therefore, from Lemma 2,

$$\lim_{n\to\infty} D(T_\delta f_{\mathscr{C}_n} \| U_n) \le \lim_{n\to\infty} H(X_{\mathscr{C}_n^\perp} | Y_{X_{\mathscr{C}_n^\perp},\mathrm{BEC}(R-\epsilon_0)}) = 0.$$

Similarly, if $R > 1 - h_\alpha(\delta)$ for $\alpha \in \{2, 3, \dots, \infty\}$, then, $\lim_{n\to\infty} D_\alpha(T_\delta f_{\mathscr{C}_n} \| U_n) = 0$.

Together with Theorem 4, we have now proved the final claim. □

The known code families that achieve the capacity of the BEC include polar codes, LDPC codes, and doubly transitive codes, such as constant-rate RM codes. LDPC codes do not fit the assumptions because of low dual distance, but the other codes do. This yields explicit families of codes that achieve the D_α-smoothing capacity.

We illustrate the results of this section in Figure 1, where the curves show the achievability and impossibility rates for perfect smoothing with respect to the Bernoulli noise. Given a code (sequence) of rate R, putting it through a noise β_δ below the Shannon capacity cannot achieve perfect smoothing. The sequence of polar codes from [39], cited in Theorem 5, is smoothable at rates equal to the Shannon capacity, although these codes do not provide a decoding guarantee at that noise level. At the second curve from the bottom, the duals of the codes that achieve Shannon's capacity in BEC achieve perfect D_1-smoothing; at the third (fourth) curve, these codes are perfectly D_2- (or D_∞-) smoothable, and they achieve the corresponding smoothing capacity.

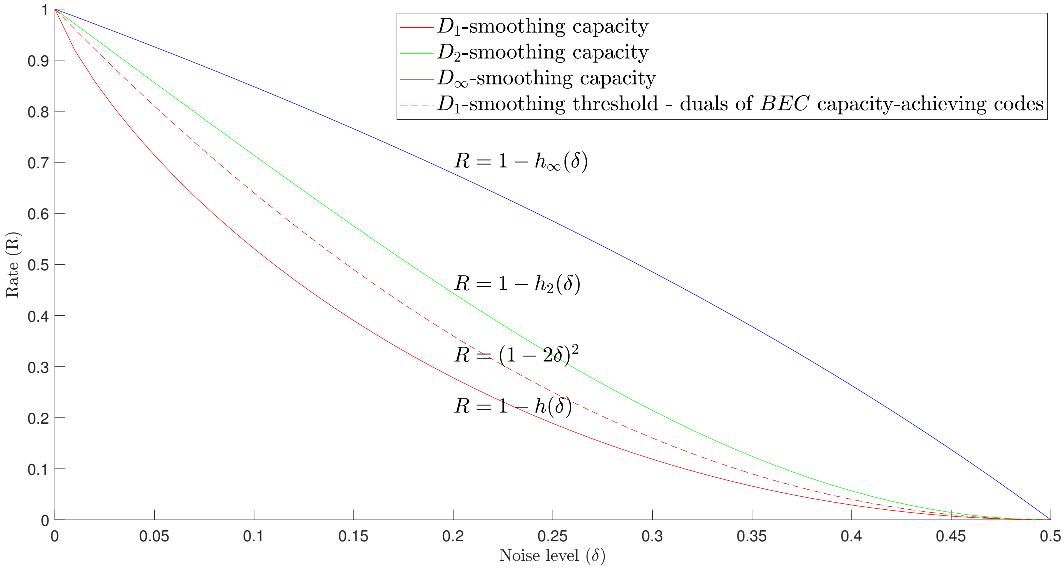

Figure 1. Capacities and achievable rates for perfect smoothing. The lowermost curve gives the Shannon capacity of the BSC(δ), the second curve from the bottom is the smoothing threshold for the duals of the BEC capacity-achieving codes, the third one is $S_2^{\beta_\delta}$ and the top one is $S_\infty^{\beta_\delta}$.

Remark 4. *Observe that the strong converse of the channel coding theorem does not imply perfect smoothing. To give a quick example, consider a code $\mathscr{C}_n = B(0, \delta' n)$ formed of all the vectors in the ball. Let $0 < \delta < 1/2$ and let us use this code on a BSC(δ), where $h(\delta) + h(\delta') > 1$ and $\delta < 1/2$. From the choice of the parameters, the rate of $\mathscr{C}_n$ is above capacity, and, therefore, $P_B(\mathrm{BSC}(\delta), \mathscr{C}_n) \approx 1$ from the strong converse. At the same time,*

$$D(T_\delta f_{\mathscr{C}_n} \| U_n) = n - H(b_{\delta' n} * \beta_\delta) = n - H(\beta_{\delta'} * \beta_\delta) + O(\sqrt{n})$$
$$= n(1 - h(\delta'(1-\delta) + \delta(1-\delta'))) + O(\sqrt{n}).$$

where the transition from the ball noise to the Bernoulli noise (the second equality) is shown in [30]. Since $\delta'(1-\delta) + \delta(1-\delta')) < 1/2$ for all $\delta < 1/2, \delta' < 1/2$, we conclude that $D(T_\delta f_{\mathscr{C}_n} \| U_n) \nrightarrow 0$.

Remark 5. *In this paper, we mostly study the trade-off between the rate of codes and the level of the noise needed to achieve perfect smoothing. A recent work of Debris-Alazard et al. [4] considered guarantees for smoothing derived from the distance distribution of codes and their dual distance (earlier, similar calculations were performed in [42,50]). Our approach enables us to find the conditions for perfect smoothing similar to [4] but relying on fewer assumptions.*

Proposition 3. *Let $\mathscr{C}_n$ be a sequence of codes whose dual distance $d(\mathscr{C}_n^\perp) \geq \partial^\perp n$ where $\partial^\perp \in (0,1)$. If $\partial^\perp > (1-2\delta)^2$, then,*

$$\lim_{n\to\infty} D(T_\delta f_{\mathscr{C}_n} \| U_n) = 0.$$

Proof. Notice that $\lim_{n\to\infty} H(X_{\mathscr{C}_n^\perp} | Y_{X_{\mathscr{C}^\perp}, \mathrm{BEC}(\lambda)}) = 0$ if $\partial^\perp > \lambda$. With this, the proof is a straightforward application of Lemma 2. □

Compared to [4], this claim removes the restrictions on the support of the dual distance distribution of the codes $\mathscr{C}_n$.

5. Binary Symmetric Wiretap Channels

In this section, we discuss applications of perfect smoothing to the BSC wiretap channel. Wyner's wiretap channel model $\mathscr{V}$ [35] for the case of BSCs is defined as follows: The system is formed of three terminals, A, B, and E. Terminal A communicates with B by sending messages M chosen from a finite set $\mathcal{M}$. Communication from A to B occurs over a BSC $\mathscr{W}_b$ with crossover probability δ_b, and it is observed by the eavesdropper E via another BSC $\mathscr{W}_e$ with crossover probability $\delta_e > \delta_b$. A message $M \in \mathcal{M}$ is encoded into a bit sequence $X \in \mathcal{H}_n$ and sent from A to B in n uses of the channel $\mathscr{W}_b$. Terminal B observes the sequence $Y = X + W_b$, where $W_b \sim \text{Bin}(n, \delta_b)$ is the noise vector, while terminal E observes the sequence $Z = X + W_e$ with $W_e \sim \text{Bin}(n, \delta_e)$. We assume that the messages are encoded into a subset of $\mathcal{H}_n$, which imposes some probability distribution on the input of the channels. The goal of the encoding is to ensure reliability and secrecy of communication. The reliability requirement amounts to the condition $\Pr(M \neq \hat{M}) \to 0$ as $n \to \infty$, where $\hat{M}$ is the estimate of M made by B. To ensure secrecy, we require the *strong secrecy condition* $I(M; Z) \to 0$. This is in contrast to the condition $\frac{1}{n} I(M; Z) \to 0$ studied in the early works on the wiretap channel, which is now called weak secrecy. Denote by $R = \frac{1}{n} \log |\mathcal{M}|$ the transmission rate. The *secrecy capacity* $C_s(\mathscr{V})$ is defined as the supremum of the rates that permit reliable transmission, which also conforms to the secrecy condition.

The nested coding scheme, proposed by Wyner [35], has been the principal tool of constructing well-performing transmission protocols for the wiretap channel [38,39,41]. To describe it, let $\mathscr{C}_e$ and $\mathscr{C}_b$ be two linear codes such that $\mathscr{C}_e \subset \mathscr{C}_b$ and $|\mathcal{M}| = \frac{|\mathscr{C}_b|}{|\mathscr{C}_e|}$. We assign each message m to a unique coset of $\mathscr{C}_e$ in $\mathscr{C}_b$. The sequence transmitted by A is a uniform random vector from the coset. As long as the rate of the code $\mathscr{C}_b$ is below the capacity of $\mathscr{W}_b$, we can ensure the reliability of communication from A to B.

Strong secrecy can be achieved relying on perfect smoothing. Denote by c_m a leader of the coset that corresponds to the message m. The basic idea is that if $P_{Z|M=m} = (T_\delta f_{\mathscr{C}_e})(\cdot + c_m)$ is close to a uniform distribution U_n for all m, these conditional pmfs are almost indistinguishable from each other, and terminal E has no means of inferring the transmitted message from the observed bit string Z.

As mentioned earlier, the weak secrecy results for the wiretap channel based on LDPC codes and on polar codes were presented in [38,39], respectively. The problem that these schemes faced, highlighted in Theorems 2 and 5, is that code sequences that achieve BSC capacity have a rate gap of at least $1/\sqrt{n}$ to the capacity value. At the same time, the rate of perfectly smoothable codes must exceed the capacity by a similar quantity [51]. For this reason, the authors of [39] included the intermediate virtual channels in their polar coding scheme, which gave them strong secrecy, but interfered with transmission reliability. A similar general issue arose earlier in attempting to use LDPC codes for the wiretap channel [40].

Contributing to the line of work connecting smoothing and thewiretap channel [2,3,11], we show that nested coding schemes $\mathscr{C}_e \subset \mathscr{C}_b$, where the code $\mathscr{C}_b$ is good for error correction in BSC(δ_b) and $\mathscr{C}_e$ is perfectly smoothable with respect to β_{δ_b}, attain strong secrecy and reliability for a BSC wiretap channel (δ_b, δ_e). As observed in Lemma 2, the duals of the good erasure-correcting codes are perfectly smoothable for certain noise levels and, hence, they form a good choice for $\mathscr{C}_e$ in this scenario.

The following lemma establishes a connection between the smoothness of a noisy distribution of a code and strong secrecy.

Lemma 3. *Consider the nested coding scheme for the BSC wiretap channel introduced above. If $D(T_{\delta_e} f_{\mathscr{C}_e} \| U_n) < \epsilon$, then $I(M; Z) < \epsilon$.*

Proof. We have

$$\begin{aligned} D(P_{Z|M}\|U_n|P_M) &= \sum_{\substack{m\in\mathcal{M}\\ z\in\mathcal{H}_n}} P_{MZ}(m,z)\log\frac{P_{Z|M}(z|m)}{U_n(z)} \\ &= I(M;Z) + D(P_Z\|U_n). \end{aligned}$$

Now, note that $P_{Z|M=m}(z) = (T_{\delta_e} f_{\mathscr{C}_e})(z+c_m) = P_{Z|M=m'}(z+c_{m'}+c_m)$, so $D(P_{Z|M=m}\|U_n)$ is independent of m. Therefore, for all $m\in\mathcal{M}$

$$\begin{aligned} D(P_{Z|M=m}\|U_n) &= D(P_{Z|M}\|U_n|P_M) \\ &= I(M;Z) + D(P_Z\|U_n) \geq I(M;Z). \quad \square \end{aligned}$$

This lemma enables us to formulate conditions for reliable communication while guaranteeing the strong secrecy condition. Namely, it suffices to take a pair (a sequence of pairs) of nested codes $\mathscr{C}_e \subset \mathscr{C}_b$ such that $D(T_{\delta_e} f_{\mathscr{C}_e}\|U_n) \to 0$ as $n\to\infty$. If at the same time the code $\mathscr{C}_b$ corrects errors on a BSC(δ_b), then the scheme fulfills both the reliability and strong secrecy requirements under noise levels δ_b and δ_e for channels $\mathscr{W}_b$ and $\mathscr{W}_e$, respectively, supporting transmission from A to B at rate $R_b - R_e$. Together with the results established earlier, we can now make this claim more specific.

Theorem 7. *Let $((\mathscr{C}_e^n)^{\perp})_n$ and $(\mathscr{C}_b^n)_n$ be sequences of linear codes that achieve the capacity of the BEC for their respective rates. Suppose that $\mathscr{C}_e^n \subset \mathscr{C}_b^n$ and*

1. $d((\mathscr{C}_e^n)^{\perp}) = \omega(\log n), R(\mathscr{C}_e^n) \to R_e;$
2. $d(\mathscr{C}_b^n) = \omega(\log n), R(\mathscr{C}_b^n) \to R_b.$

If $R_b < 1 - \log(1 + 2\sqrt{\delta_b(1-\delta_b)})$ and $R_e > 4\delta_e(1-\delta_e)$, then the nested coding scheme based on $\mathscr{C}_e^n$ and $\mathscr{C}_b^n$ can transmit messages with rate $R_b - R_e$ from A to B, satisfying the reliability and strong secrecy conditions.

Proof. From Corollary A1, the conditions $d(\mathscr{C}_b^{(n)}) = \omega(\log n)$ and $R_b < 1 - \log(1 + 2\sqrt{\delta_b(1-\delta_b)})$ guarantee transmission reliability. Furthermore, by Theorem 6, the conditions $d((\mathscr{C}_e^n)^{\perp}) = \omega(\log n)$ and $R_e > 4\delta_e(1-\delta_e)$ imply that $D(T_{\delta_e} f_{\mathscr{C}_e}\|U_n) \to 0$, which in its turn implies strong secrecy by Lemma 3. □

To give an example of a code family that satisfies the assumptions of this theorem, consider the RM codes of constant rate. Namely, let $\mathscr{C}_e^n \subset \mathscr{C}_b^n$ be two sequences of RM codes whose rates converge to R_e and R_b, respectively. Note that the duals of the RM codes are themselves RM codes. By a well-known result [52], the RM codes achieve the capacity of the BEC, and for any sequence of constant-rate RM codes, the distance scales as $2^{\Theta(\sqrt{n})}$. Therefore, the RM codes satisfy the assumptions of Theorem 7.

Note that for the RM codes, we can obtain a stronger result, based on their error correction properties on the BSC. Involving this additional argument brings them closer to the secrecy capacity under the strong secrecy assumption.

Theorem 8. *Let $\mathscr{C}_e^n$ and $\mathscr{C}_b^n$ be two sequences of RM codes satisfying $\mathscr{C}_e^n \subset \mathscr{C}_b^n$ whose rates approach $R_e > 0$ and $R_b > 0$, respectively. If $R_b < 1 - h(\delta_b)$ and $R_e > 4\delta_e(1-\delta_e)$, then the nested coding scheme based on $\mathscr{C}_e^n$ and $\mathscr{C}_b^n$ supports transmission on a BSC wiretap channel (δ_b, δ_e) with rate $R_b - R_e$, guaranteeing communication reliability and strong secrecy.*

Proof. Very recently, Abbe and Sandon [53], building upon the work of Reeves and Pfister [54], proved that RM codes achieve capacity in symmetric channels. Therefore, the condition $R_b < 1 - h(\delta_b)$ guarantees reliability. The rest of the proof is similar to that of Theorem 7. □

Theorems 7 and 8 stop short of constructing codes that attain the secrecy capacity of the channel (this is similar to the results of [14] for the transmission problem over the BSC). To quantify the gap to capacity, we plot the smoothing and decodability rate bounds in Figure 2.

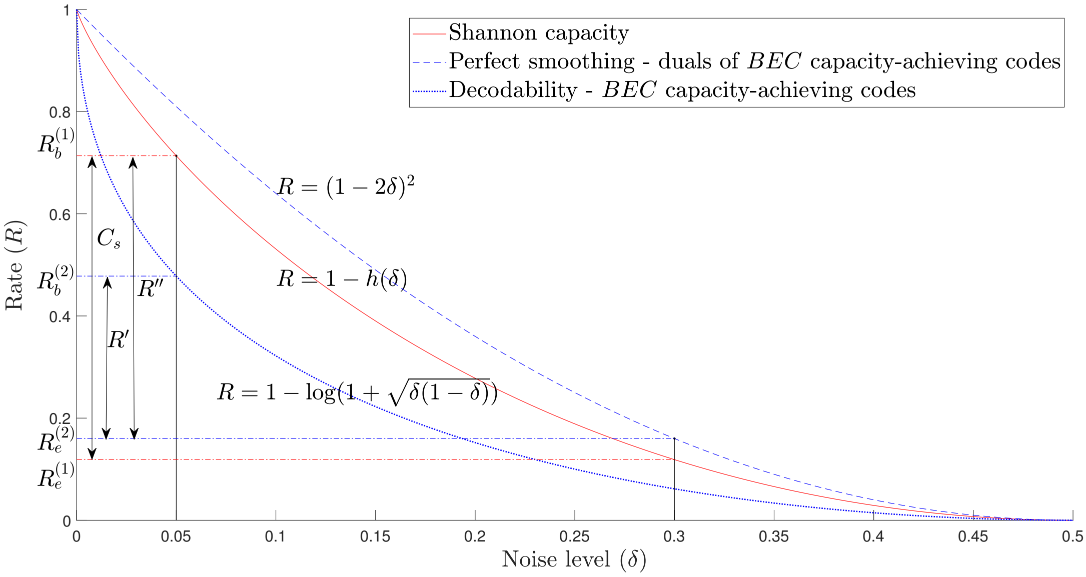

Figure 2. Achievable rates in the BSC wiretap channel with BEC capacity-achieving codes. The bottom curve is the lower bound on the code rate that guarantees decodability on a BSC(δ). The middle curve shows Shannon's capacity and the top one is the D_1-smoothing threshold for the Bernoulli noise T_δ.

As an example, let us set the noise parameters $\delta_b = 0.05$ and $\delta_e = 0.3$ and denote the corresponding secrecy capacity by C_s. Suppose that we use a BEC capacity-achieving code as code $\mathscr{C}_b$ and a dual of a BEC capacity-achieving code as code $\mathscr{C}_e$ in the nested scheme. The value R' is the largest rate at which we can guarantee both reliability and strong secrecy. In the example in Figure 2, $C_s = R_b^{(1)} - R_e^{(1)} = 0.5949$ and $R' = R_b^{(2)} - R_e^{(2)} = 0.3181$. The only assumption required here is that the codes $\mathscr{C}_e^\perp$ and $\mathscr{C}_b$ have good erasure correction properties.

As noted, generally, the RM codes support a higher communication rate than the R'. Let R'' be their achievable rate. For the same noise parameters as above, we obtain $R'' = R_b^{(1)} - R_e^{(2)} = 0.5536$, which is closer to C_s than R'.

Remark 6. *The fact that the RM codes achieve capacity in symmetric channels immediately implies that nested RM codes achieve the secrecy capacity in the BSC wiretap channel under weak secrecy. While it is tempting to assume that, coupled with the channel duality theorems of [55,56], this result also implies that RM codes fulfil the strong secrecy requirement on the BSC wiretap channel, an immediate proof looks out of reach [57].*

Secrecy from α-Divergence

Classically, the (strong) secrecy in the wiretap channel is measured by $I(M,Z)$. In [11], slightly weaker secrecy measures were considered besides the mutual information. However, more stringent secrecy measures may be required in certain scenarios; α-divergence-based secrecy measures were introduced by Yu and Tan [3] as a solution to this problem.

Observe that the secrecy measured by $D_\alpha(P_{Z|M} \| U_n | M)$ for $\alpha \geq 1$ is stronger than the mutual-information-based secrecy. This is because for $\alpha \geq 1$

$$I(M;Z) \leq D(P_{Z|M} \| U_n | P_M) \leq D_\alpha(P_{Z|M} \| U_n | P_M).$$

Given a wiretap channel with an encoding-decoding scheme, we say the α-secrecy is satisfied if

$$\lim_{n\to\infty} D_\alpha(P_{Z|M}\|U_n|P_M) = 0.$$

The following theorem establishes that it is possible to achieve the rate $C(\delta_b) - S_\alpha^{\beta_{\delta_e}} = h_\alpha(\delta_e) - h(\delta_b)$ with RM codes for $\alpha \in \{2, 3, \ldots, \infty\}$.

Theorem 9. *Let $\alpha \in \{2, 3, \ldots, \infty\}$. Let $\mathscr{C}_e^n$ and $\mathscr{C}_b^n$ be two sequences of RM codes satisfying $\mathscr{C}_e^n \subset \mathscr{C}_b^n$ whose rates approach $R_e > 0$ and $R_b > 0$, respectively. If $R_b < 1 - h(\delta_b)$ and $R_e > 1 - h_\alpha(\delta_e)$, then the nested coding scheme based on $\mathscr{C}_e^n$ and $\mathscr{C}_b^n$ supports transmission on a BSC wiretap channel (δ_b, δ_e) guaranteeing α-secrecy with rate $R_b - R_e$, provided that $h_\alpha(\delta_e) - h(\delta_b) > 0$.*

Evidently, to achieve a stringent version of secrecy, it is necessary to reduce the rate of the message. The capacity of the (δ_b, δ_e)-wiretap channel is $h(\delta_e) - h(\delta_b)$, while the known highest rate that assures α-secrecy and reliability is $h_\alpha(\delta_e) - h(\delta_b)$. Hence, to achieve α-secrecy, we must give up $h(\delta_e) - h_\alpha(\delta_e)$ of the attainable rate.

6. Ball Noise and Error Probability of Decoding

This section focuses on achieving the best possible smoothing with respect to the ball noise. As an application, we show that codes that possess good smoothing properties with respect to the ball noise are suitable for error correction in the BSC.

6.1. Ball Noise

Recall that the perfect smoothing of a sequence of codes is only possible if the rate is greater than the corresponding D_α-smoothing capacity. In addition to characterizing the D_α-smoothing capacities of the ball noise, we quantify the best smoothing one can expect with rates below the D_α-smoothing capacity. We will use these results in the upcoming subsection when we derive upper bounds for the decoding error probability on a BSC. The next theorem summarizes our main result on smoothing with respect to the ball noise.

Theorem 10. *Let $(b_{\delta n})_n$ be the sequence of ball noise operators, where δn is the radius of the ball. Let $\delta \in [0, 1/2], \alpha \in [0, \infty]$. Let $\mathscr{C}_n$ be a code of length n and rate R_n. Then, we have the following bounds:*

$$D_\alpha(T_{b_{\delta n}} f_{\mathscr{C}_n}\|U_n) \geq 0 \tag{19}$$

$$\frac{1}{n} D_\alpha(T_{b_{\delta n}} f_{\mathscr{C}_n}\|U_n) \geq 1 - R_n - h(\delta). \tag{20}$$

There exist sequences of codes of rate $R_n \to R$ that achieve asymptotic equality in (19) for all $R > 1 - h(\delta)$. At the same time, if $R < 1 - h(\delta)$, then there exist sequences of codes achieving asymptotic equality in (20).

Proof. The inequality in (19) is trivial. Let us prove that asymptotically it can be achieved with equality. From Theorem 3, there exists a sequence of codes $(\mathscr{C}_n)_n$ such that $D_\infty(T_{b_{\delta n}} f_{\mathscr{C}_n}\|U_n) = o(1)$ given that $R > 1 - h(\delta)$. Hence, for $\alpha \in [0, \infty]$

$$0 \leq D_\alpha(T_{b_{\delta n}} f_{\mathscr{C}_n}\|U_n) \leq D_\infty(T_{b_{\delta n}} f_{\mathscr{C}_n}\|U_n) = o(1).$$

Hence, the equality case in (19) is achievable for all $\alpha \in [0, \infty]$.

Let us prove (20). From Lemma 1, we have

$$D_\alpha(T_{b_{\delta n}} f_{\mathscr{C}_n}\|U_n) \geq n(1 - R_n) - H_\alpha(b_{\delta n}) \geq n(1 - R_n - h(\delta))$$

because $\frac{1}{n} H_\alpha(b_{\delta n}) = \frac{1}{n} \log V_{\delta n} \leq h(\delta)$.

We are left to show that for $R < 1 - h(\delta)$, (20) can be achieved with equality in the limit of large n. We use a random coding argument to prove this. Let $\mathscr{C}_n$ be an $(n, 2^{nR_n})$ code whose codewords are chosen independently and uniformly. In Equation (A6), Appendix B, we define the expected norm of the noisy function. Here, we use this quantity for the ball noise kernel. For $\alpha \in [0, \infty)$, define

$$Q_n(\alpha) = \mathbb{E}_{\mathscr{C}_n} 2^{(\alpha-1)D_\alpha(T_{b_{\delta n}} \| U_n)}.$$

From Lemma A2, for any rational $\alpha \geq 1$,

$$Q_n(\alpha) \leq \sum_{k=0}^{p} \binom{p}{k} 2^{\frac{nk}{q}(1-R_n-\frac{\log V_{\delta n}}{n})} Q_n\Big(\frac{p-k}{q}\Big), \tag{21}$$

for $p, q \in \mathbb{Z}_0^+$ such that $\alpha = 1 + \frac{p}{q}$.

Assume that $R < 1 - h(\delta)$. Let us prove that $Q_n(\alpha) \leq 2^{n(\alpha-1)(1-R-h(\delta)+o(1))}$ for rational values of α using induction. Let $\alpha \in [1, 2]$ be rational and note that $p \leq q$. Since $Q_n(\cdot) \leq 1$ when the argument is less than 1, we can write (21) as follows:

$$Q_n(\alpha) \leq \sum_{k=0}^{p} \binom{p}{k} 2^{\frac{nk}{q}(1-R_n-\frac{\log V_{\delta n}}{n})} = 2^{n(\alpha-1)(1-R-h(\delta)+o(1))}.$$

Now, assume that (21) holds for all rational $\alpha \in [1, m]$ for some integer $m \geq 2$ and prove that, in this case, it holds also for $\alpha \in (m, m+1]$. By the induction hypothesis,

$$\begin{aligned} Q_n(\alpha) &\leq \sum_{0 \leq k \leq p-q} \binom{p}{k} 2^{\frac{nk}{q}(1-R_n-\frac{\log V_{\delta n}}{n})} 2^{n\frac{p-k-q}{q}(1-R-h(\delta)+o(1))} + \sum_{k=p-q}^{p} \binom{p}{k} 2^{\frac{nk}{q}(1-R_n-\frac{\log V_{\delta n}}{n})} \\ &\leq \sum_{0 \leq k \leq p-q} \binom{p}{k} 2^{n(\alpha-2)(1-R-h(\delta)+o(1))} + \sum_{k=p-q}^{p} \binom{p}{k} 2^{n(\alpha-1)(1-R-h(\delta)+o(1))} \\ &= 2^{n(\alpha-1)(1-R-h(\delta)+o(1))}. \end{aligned}$$

Therefore, for every rational $\alpha \in (1, \infty)$, there exists a sequence of codes satisfying

$$D_\alpha(T_{b_{\delta n}} f_{\mathscr{C}_n} \| U_n) = n(1 - R - h(\delta) + o(1)), \tag{22}$$

which is equivalent to the equality in (20).

Let us extend this result to non-negative reals. Let $\alpha \in [0, \infty)$ and let us choose a rational $\alpha' \in (1, \infty)$ such that $\alpha < \alpha'$. We know that there exists a sequence of codes satisfying

$$D_{\alpha'}(T_{b_{\delta n}} f_{\mathscr{C}_n} \| U_n) = n(1 - R - h(\delta) + o(1)).$$

From (20) and from Remark 1,

$$n(1 - R_n - h(\delta)) \leq D_\alpha(T_{b_{\delta n}} f_{\mathscr{C}_n} \| U_n) \leq D_{\alpha'}(T_{b_{\delta n}} f_{\mathscr{C}_n} \| U_n) = n(1 - R - h(\delta) + o(1)).$$

Hence, the asymptotic equality in (20) is achievable for all $\alpha \in [0, \infty)$. □

The above theorem characterizes the D_α-smoothing capacities with respect to ball noise.

Corollary 2. *Let $\delta \in [0, 1/2]$. Let $b(\delta) = (b_{\delta n})_n$ be a sequence of ball noise operators, where δn is the radius corresponding to the n-th kernel. Then,*

$$S_\alpha^{b(\delta)} = 1 - h(\delta) \quad \textit{for } \alpha \in [0, \infty].$$

The norms of $T_{b_t} f_{\mathscr{C}}$ can be used to bound the decoding error probability on a BSC. While estimating these norms for a given code is generally complicated, the second norm affords a compact expression based on the distance distribution of the code. In the next section, we bound the decoding error probability using the second norm of $T_{b_t} f_{\mathscr{C}}$. The following proposition provides closed-form expressions for $\|2^n T_{b_t} f_{\mathscr{C}}\|_2^2$.

Proposition 4.

$$\|2^n T_{b_t} f_{\mathscr{C}}\|_2^2 = \frac{2^n}{|\mathscr{C}|V_t^2} \sum_{i=0}^{n} \mu_t(i) A_i = \frac{1}{V_t^2} \sum_{k=0}^{n} L_t(k)^2 A_k^{\perp}.$$

where $\mu_t(i)$ is defined in (1) *and L_t is the Lloyd polynomial of degree t* (A2).

The proof is immediate from Proposition A1 in combination with (A3) and (A4).

6.2. Probability of Decoding Error on a BSC(δ)

The idea that the smoothing of codes under some conditions implies good decoding performance has appeared in a number of papers using different language. The smoothing of capacity-achieving codes was considered in [18,46]. Hązła et al. [14] showed that if a code (sequence) is perfectly smoothable with respect to the Bernoulli noise, then the dual code is good for decoding (see Theorem A4, Corollary A1). Going from smoothing to decodability involves representing the D_2-smoothness of codes with respect to the Bernoulli noise as a potential energy form and comparing it to the Bhattacharyya bound for the dual codes. One limitation of this approach is that it cannot infer decodability for rates $R > 1 - \log(1 + 2\sqrt{\delta(1-\delta)})$ (this is the region above the blue solid curve in Figure 2). Rao and Sprumont [15] and Hązła [34] proved that sufficient smoothing of codes implies the decodability of the codes themselves rather than their duals. However, these results are concerned with list decoding for rates above the Shannon capacity, resulting in an exponential list size, which is arguably less relevant from the perspective of communication.

Except for [15], the cited papers utilize perfect or near-perfect smoothing to infer decodability. For codes whose rates are below the capacity, perfect smoothing is impossible. At the same time, codes that possess sufficiently good smoothing properties are good for decoding. This property is at the root of the results for list decoding in [15]; however, their bounds were insufficient to make conclusions about list decoding below capacity.

Consider a channel where, for the input $X \sim f_{\mathscr{C}}$, the output Y is given by $Y = X + W$ with $W \sim b_t$. Define $F_t(y) = |\mathscr{C} \cap B(y,t)|$ as the number of codewords in the ball $B(y,t)$. Hence, for a received vector y, the possible number of codewords that can yield y is given by $F_t(y)$. Intuitively, the decoding error is small if $F_t(y) \approx 1$ for typical errors. Therefore, F_t is of paramount interest in decoding problems. Since the typical errors for both ball noise and the Bernoulli noise are almost the same, this allows us to obtain a bound for decodability in the BSC channel. Using this approach, we show that the error probability of decoding on a BSC(δ) can be expressed via the second moment of the number of codewords in the ball of radius $t \gtrsim \delta n$.

Assume, without loss of generality, that $\mathscr{C}$ is a linear code and 0^n is used for transmission. Let Y be the random Bernoulli vector of errors, and note that $Y \sim \beta_\delta$. The calculation below does not depend on whether we rely on unique or list decoding within a ball of radius t, so let us assume that the decoder outputs $L \geq 1$ candidate codewords conditioned on the received vector y, which is a realization of Y.

In this case, the list decoding error can be written as

$$P_{L,t}(\mathscr{C}, \text{BSC}(\delta)) = \Pr\{F_t(Y) \geq L + 1 \cup |Y| > t\}. \quad (23)$$

Theorem 11. *Let t and t' be integers such that $0 < t' < t < n$. Then, for any $L \geq 1$,*

$$P_{L,t}(\mathscr{C}, \mathrm{BSC}(\delta)) \leq \frac{\beta_\delta(t')}{L} \sum_{w=1}^{n} \mu_t(w) A_w + \Pr(|Y| \leq t' \cup |Y| \geq t). \tag{24}$$

Proof. Define $S_{t',t} = B(0,t) \setminus B(0,t')$. Clearly,

$$\begin{aligned} P_{L,t}(\mathscr{C}, \mathrm{BSC}(\delta)) &= \Pr\{F_t(Y) \geq L+1 \cup |Y| > t\} \\ &\leq \Pr\{(F_t(Y) \geq L+1) \cap (Y \in S_{t',t})\} + \Pr(Y \notin S_{t',t}). \end{aligned}$$

Let us estimate the first of these probabilities.

$$\begin{aligned} &\Pr\{(F_t(Y) \geq L+1) \cap (Y \in S_{t',t})\} \\ &= \sum_{y \in S_{t',t}} \mathbb{1}_{F_t(y) \geq L+1} \beta_\delta(y) \\ &\leq \sum_{y \in S_{t',t}} \frac{F_t(y) - 1}{L} \beta_\delta(y) \\ &\leq \frac{\beta_\delta(t')}{L} \sum_{y \in S_{t',t}} (F_t(y) - 1) \\ &\leq \frac{\beta_\delta(t')}{L} \sum_{y \in B(0,t)} (F_t(y) - 1) \quad \text{(because for all } y \in B(0,t), F_t(y) \geq 1) \\ &= \frac{\beta_\delta(t')}{L} \Big(\sum_{y \in \mathcal{H}_n} (\mathbb{1}_{\mathscr{C}} * \mathbb{1}_{B(0,t)})(y) \mathbb{1}_{B(0,t)}(y) - V_t \Big) \\ &= \frac{\beta_\delta(t')}{L} \Big(\sum_{c \in \mathscr{C}} (\mathbb{1}_{B(0,t)} * \mathbb{1}_{B(0,t)})(c) - V_t \Big) \\ &= \frac{\beta_\delta(t')}{L} \sum_{i=1}^{n} \mu_t(i) A_i. \quad \square \end{aligned}$$

Remark 7. *In the case of $L = 1$, the bound in (24) can be considered a slightly weaker version of Poltyrev's bound [58], Lemma 1. By allowing this weakening, we obtain a bound in a somewhat more closed form, also connecting the decodability with smoothing. We also prove a simple bound for the error probability of list decoding expressed in terms of the code's distance distribution (and, from (A4), also in terms of the dual distance distribution). The latter result seems not to have appeared in earlier literature.*

The following version of this lemma provides an error bound, which is useful in the asymptotic setting.

Proposition 5. *Let $t = \delta n + n^\theta$, where $\theta \in (1/2, 1)$. Then,*

$$P_{L,t}(\mathscr{C}, \mathrm{BSC}(\delta)) \leq \frac{\sqrt{2n}}{LV_t} \Big(\frac{1-\delta}{\delta} \Big)^{2n^\theta} \sum_{w=1}^{n} \mu_t(w) A_w + 2e^{-n^{2\theta-1}}.$$

In particular,

$$P_{L,t}(\mathscr{C}, \mathrm{BSC}(\delta)) \leq \frac{\sqrt{2n}}{V_t} \Big(\frac{1-\delta}{\delta} \Big)^{2n^\theta} \sum_{w=1}^{n} \mu_t(w) A_w + 2e^{-n^{2\theta-1}}.$$

Proof. Set $t' = \delta n - n^\theta$. A direct calculation shows that

$$\beta_\delta(t') V_t < \sqrt{2n} \Big(\frac{1-\delta}{\delta} \Big)^{2n^\theta}.$$

By the Hoeffding bound,

$$\Pr(|Y| \leq t' \cup |Y| \geq t) \leq 2e^{-n^{2\theta-1}}.$$

Together with Lemma 11, this implies our statements. □

A question of prime importance is whether the right-hand side quantities in Proposition 5 converge to 0. For $R < 1 - h(\delta)$, one can easily see that for random codes, $\sum_{w=1}^{n} \frac{\mu_t(w)}{V_t} A_w = 2^{-\Theta(n)}$, where $t = \delta n + n^{\theta}$, showing that this is, in fact, the case.

From Proposition 4, it is clear that the potential energy $\sum_{w=1}^{n} \mu_t(w) A_w$ is a measure of the smoothness of $T_{b_t} f_{\mathscr{C}}$. This implies that codes that are sufficiently smoothable with respect to b_t are decodable in the BSC with vanishing error probability. In other words, Proposition 5 establishes a connection between the smoothing and the decoding error probability.

7. Perfect Smoothing—The Finite Case

In this section, we briefly overview another form of perfect smoothing, which is historically the earliest application of these ideas in coding theory. It is not immediately related to the information-theoretic problems considered in the other parts.

We are interested in radial kernels that yield perfect smoothing for a given code. We often write $r(i)$ instead of $r(x)$ if $|x| = i$, and call $\rho(r) := \max(i : r(i) \neq 0)$ the *radius* of r. Note that the logarithm of the support size of r (as a function on the space $\mathcal{H}_n$) is exactly the 0-Rényi entropy of r. Therefore, kernels with smaller radii can be perceived as less random, supporting the view of the radius $\rho(r)$ as a general measure of randomness.

Definition 4. *We say a code $\mathscr{C}$ is perfectly smoothable with respect to r if $T_r f_{\mathscr{C}}(x) = \frac{1}{2^n}$ for all $x \in \mathcal{H}_n$, and, in this case, we say that r is a perfectly smoothing kernel for $\mathscr{C}$.*

Intuitively, such a kernel should have a sufficiently large radius. In particular, it should be as large as the *covering radius* of the code $\rho(\mathscr{C})$ or otherwise smoothing does not affect the vectors that are ρ away from the code. To obtain a stronger condition, recall that the external distance of code $\mathscr{C}$ is $\bar{d}(\mathscr{C}) = |\{i \geq 1 : A_i^{\perp} \neq 0\}|$.

Proposition 6. *Let r be a perfectly smoothing kernel of code $\mathscr{C}$. Then, $\rho(r) \geq \bar{d}(\mathscr{C})$.*

Proof. Note that perfect smoothing of $\mathscr{C}$ with respect to r is equivalent to

$$\|2^n T_r f_{\mathscr{C}}\|_2^2 = 1,$$

which by Proposition A1 is equivalent to the following condition:

$$\sum_{i=1}^{n} \hat{r}(i)^2 A_i^{\perp} = 0.$$

Therefore,

$$\bar{d}(\mathscr{C}) = |\{i \geq 1 : A_i^{\perp} \neq 0\}| \leq n - |\{i \geq 1 : \hat{r}(i) \neq 0\}|.$$

By definition,

$$\hat{r} = \frac{1}{2^n} K^{\intercal} r,$$

where $K = (K_i(j))_{i,j=0}^n$ is the Krawtchouk matrix. Define $I_1 = \{j \in \{1,2,\dots,n\} : \hat{r}(j) = 0\}$ and $I_2 = \{i \in \{1,2,\dots,n\} : r(i) \neq 0\}$ then,

$$0 = \hat{r}|_{I_1} = \frac{1}{2^n} K^\intercal|_{(I_1,:)} r = \frac{1}{2^n} K^\intercal|_{(I_1,I_2)} r|_{I_2}.$$

This relation implies that there exists a linear combination of Krawtchouk polynomials of degree at most $\rho(r)$ with $|I_1|$ roots. Therefore, $\bar{d}(\mathscr{C}) \leq n - |\operatorname{supp}(\{\hat{r}(i)\}_{i=1}^n)| = |I_1| \leq \rho(r)$. □

Since $\rho(\mathscr{C}) \leq \bar{d}(\mathscr{C})$, this inequality strengthens the obvious condition $\rho(r) \geq \rho(\mathscr{C})$. At the same time, there are codes that are perfectly smoothable by a radial kernel r such that $\rho(r) = \rho(\mathscr{C})$.

Definition 5 ([59]). *A code $\mathscr{C}$ is uniformly packed in the wide sense if there exists rational numbers $\{\alpha_i\}_{i=0}^{\rho}$ such that*

$$\sum_{i=0}^{\rho(\mathscr{C})} \alpha_i A_i(x) = 1 \quad \text{for all } x \in \mathcal{H}_n,$$

where $A_i(x)$ is the weight distribution of the code $\mathscr{C} - x$.

Our main observation here is that some uniformly packed codes are perfectly smoothable with respect to noise kernels that are minimal in a sense. The following proposition states this more precisely.

Proposition 7. *Let $\mathscr{C}$ be a code that is perfectly smoothable by a radial kernel of radius $\rho(r) = \rho(\mathscr{C})$. Then, $\mathscr{C}$ is uniformly packed in the wide sense with $\alpha_i \geq 0$ for all i.*

Proof. By definition, if $\mathscr{C}$ is perfectly smoothable with respect to r, then $2^n T_r f_{\mathscr{C}} = 1$, which is tantamount to $\sum_{y \in \mathcal{H}_n} \frac{2^n}{|\mathscr{C}|} r(y) \mathbb{1}_{\mathscr{C}}(x - y) = 1$ for all $x \in \mathcal{H}_n$. This condition can be written as $\sum_{i=0}^{\rho} \left(\frac{2^n}{|\mathscr{C}|} r(i)\right) A_i(x) = 1$ for all $x \in \mathcal{H}_n$, completing the proof. □

To illustrate this claim, we list several families of uniformly packed codes ([59–61]) that are perfectly smoothable by a kernel of radius equal to the covering radius of the code.

(i) Perfect codes: $r = b_\rho$, where $\rho = \rho(\mathscr{C})$ is the covering radius.

(ii) 2-error-correcting BCH codes of length $2^{2m+1}, m \geq 2$. The smoothing kernel r is given by

$$r(0) = r(1) = L, r(2) = r(3) = \frac{3L}{n}, r(i) = 0, i \geq 4.$$

(iii) Preparata codes. The smoothing kernel r is given by

$$r(0) = r(1) = L, r(2) = r(3) = \frac{6L}{n-1}, r(i) = 0, i \geq 4.$$

(iv) Binary $(2^m - 1, 2^{2^m - 3m + 2}, 7)$ Goethals-like codes [60]. The smoothing kernel r is given by

$$r(0) = r(1) = L, r(2) = r(3) = \frac{65L}{2n}, r(4) = r(5) = \frac{30L}{n(n-3)}, r(i) = 0, i \geq 4.$$

Here, L is a generic notation for the normalizing factor. More examples are found in a related class of *completely regular codes* [62].

Definition 5 does not include the condition that $\alpha_i \geq 0$, and, in fact, there are codes that are uniformly packed in the wide sense, but some of the α_i's are negative, and, thus, they are not smoothable by a noise kernel of radius $\rho(\mathscr{C})$. One such family is the 3-error-correcting binary BCH codes of length $2^{2m+1}, m \geq 2$ [60].

Author Contributions: Conceptualization, A.B.; Formal analysis, M.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Science Foundation, USA, grant number CCF-2104489, CCF-2110113.

Data Availability Statement: Not applicable.

Acknowledgments: We are grateful to the reviewers for their feedback on our manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. L_2 Smoothing

The Fourier transform of a function $f : \mathcal{H}_n \to \mathbb{R}$ is a function on the dual group $\widehat{\mathcal{H}}_n$, which we identify with $\mathcal{H}_n$:

$$\widehat{f}(y) = \frac{1}{2^n} \sum_{x \in \mathcal{H}_n} f(x)(-1)^{x \cdot y}, \quad y \in \mathcal{H}_n. \tag{A1}$$

The Fourier transform of the indicator function of the sphere is given by $\widehat{\mathbb{1}}_{S(0,t)} = \frac{1}{2^n} K_t$, where $K_t(x) = K_t^{(n)}(x) = \sum_{j=0}^{t} (-1)^j \binom{x}{j}\binom{n-x}{t-j}$ is a Krawtchouk polynomial of degree t. Then, clearly the Fourier transform of the indicator of the ball is

$$\widehat{\mathbb{1}}_{B(0,t)} = \frac{1}{2^n} L_t, \tag{A2}$$

where $L_t(x) := \sum_{i=0}^{t} K_i(x)$ is called the Lloyd polynomial ([63], p. 64). The intersection of balls in (1) can be written as $\mathbb{1}_{B(0,t)} * \mathbb{1}_{B(x,t)}$, which implies the expression ([42], Lemma 4.1)

$$\mu_t(i) = 2^{-n} \sum_{k=0}^{n} L_t(k)^2 K_k(i). \tag{A3}$$

Given a code $\mathscr{C} \subset \mathcal{H}_n$, we define the *dual distance distribution* of $\mathscr{C}$ as the set of numbers $A_j^{\perp} := \frac{1}{|C|} \sum_{i=0}^{n} A_i K_j(i)$, where $(A_i)_{i=0}^n$ is the distance distribution of $\mathscr{C}$ (2). Note that when $\mathscr{C}$ is linear, the set $(A_j^{\perp})_{j=0}^n$ coincides with the distance distribution of its dual code $\mathscr{C}^{\perp}$. For a radial potential V on $\mathcal{H}_n$ and a code $\mathscr{C}$, we have

$$\sum_{i=0}^{n} V(i) A_i = |\mathscr{C}| \sum_{k=0}^{n} \widehat{V}(k) A_k^{\perp}. \tag{A4}$$

The L_2-smoothness of a noisy code distribution can be written in terms of the distance distribution or of the dual distance distribution.

Proposition A1. *Let $\mathscr{C}$ be a code and r be a noise kernel. Then,*

$$\|2^n T_r f_{\mathscr{C}}\|_2^2 = \frac{2^n}{|\mathscr{C}|} \sum_{i=0}^{n} (r * r)(i) A_i = 4^n \sum_{k=0}^{n} \hat{r}(k)^2 A_k^{\perp}.$$

Proof. Let us prove the first equality:

$$\begin{aligned}
\|2^n T_r f_{\mathscr{C}}\|_2^2 &= \frac{1}{2^n} \sum_{x\in\mathcal{H}_n} (2^n T_r f_{\mathscr{C}}(x))^2 \\
&= \frac{2^n}{|\mathscr{C}|^2} \sum_{x\in\mathcal{H}_n} (r * \mathbb{1}_{\mathscr{C}})(x)^2 \\
&= \frac{2^n}{|\mathscr{C}|^2} \sum_{x\in\mathcal{H}_n} \sum_{y\in\mathcal{H}_n} r(x-y)\mathbb{1}_{\mathscr{C}}(y) \sum_{z\in\mathcal{H}_n} r(x-z)\mathbb{1}_{\mathscr{C}}(z) \\
&= \frac{2^n}{|\mathscr{C}|^2} \sum_{y\in\mathscr{C}} \sum_{z\in\mathscr{C}} \sum_{x\in\mathcal{H}_n} r(x-y)r(x-z) \\
&= \frac{2^n}{|\mathscr{C}|^2} \sum_{y\in\mathscr{C}} \sum_{z\in\mathscr{C}} (r*r)(y-z) \\
&= \frac{2^n}{|\mathscr{C}|} \sum_{i=0}^{n} (r*r)(i) A_i.
\end{aligned} \tag{A5}$$

The second equality is immediate by noticing that $\widehat{r*r} = 2^n \hat{r}^2$ and using (A4). □

Appendix B. Proof of Theorem 3

We will first establish Theorem 3 when α is rational, and then use a density argument to extend the proof to all real numbers. The case $\alpha = \infty$ is handled separately at the end of this appendix.

We will use the following technical claim:

Lemma A1. *Let x and y be two non-negative reals. Further, let p and q be positive integers. Then,*

$$(x+y)^{\frac{p}{q}} \le \sum_{k=0}^{p} \binom{p}{k} x^{\frac{k}{q}} y^{\frac{p-k}{q}}.$$

Proof. Clearly $(x+y)^{\frac{1}{q}} \le x^{\frac{1}{q}} + y^{\frac{1}{q}}$. Therefore,

$$(x+y)^{\frac{p}{q}} \le (x^{\frac{1}{q}} + y^{\frac{1}{q}})^p = \sum_{k=0}^{p} \binom{p}{k} x^{\frac{k}{q}} y^{\frac{p-k}{q}}. \quad \square$$

For $M \ge 1$, let $\mathscr{C} = (c_0, c_2, \ldots, c_{M-1})$ be a code whose codewords are chosen randomly and independently from $\mathcal{H}_n$. For $\alpha \in [0, \infty)$, define

$$Q_n(\alpha) = \mathbb{E}_{\mathscr{C}} 2^{(\alpha-1)D_\alpha(T_r f_{\mathscr{C}} \| U_n)}. \tag{A6}$$

For $\alpha > 0$, $Q_n(\alpha) = \|2^n T_r f_{\mathscr{C}}(x)\|_\alpha^\alpha$. Clearly $Q_n(1) = 1$, $Q_n(\alpha) \le 1$ for $\alpha \in [0,1)$, and $Q_n(\alpha) \ge 1$ for $\alpha > 1$.

In the next lemma, we obtain a recursive bound for Q_n. We will then use an induction argument to show the full result.

Lemma A2. *Let $\alpha = \frac{p}{q} + 1$ and let $\mathscr{C} \subset \mathcal{H}_n$ be a random code of size $M = 2^{nR}$. Then,*

$$Q_n(\alpha) \le \sum_{k=0}^{p} \binom{p}{k} 2^{\frac{nk}{q}(1-R-\frac{1}{n}H_{1+k/q}(r))} Q_n\Big(\frac{p-k}{q}\Big). \tag{A7}$$

Proof. In the calculation below, we write $\mathbb{E}$ for $\mathbb{E}_{\mathscr{C}}$. Starting with (A6), we obtain

$$\begin{aligned}
Q_n(\alpha) &= \mathbb{E}\Big[\frac{1}{2^n}\sum_{x\in\mathcal{H}_n}[2^n(r*f_{\mathscr{C}})(x)]^\alpha\Big] \\
&= 2^{n(\alpha-1)}\mathbb{E}\Big[\sum_{x\in\mathcal{H}_n}\Big[\sum_{z\in\mathscr{C}} r(x-z)\frac{1}{M}\Big]^\alpha\Big] \\
&= \frac{2^{n(\alpha-1)}}{M^\alpha}\sum_{x\in\mathcal{H}_n}\mathbb{E}\Big[\sum_{i=0}^{M-1} r(x-c_i)\Big]^\alpha \\
&= \frac{2^{n(\alpha-1)}}{M^\alpha}\sum_{x\in\mathcal{H}_n}\mathbb{E}\Big[\sum_{i=0}^{M-1} r(x-c_i)\Big[\sum_{j=0}^{M-1} r(x-c_j)\Big]^{\alpha-1}\Big] \\
&= \frac{2^{n(\alpha-1)}}{M^\alpha}\sum_{x\in\mathcal{H}_n}\mathbb{E}\Big[\sum_{i=0}^{M-1} r(x-c_i)\Big[r(x-c_i)+\sum_{j=0,j\neq i}^{M-1} r(x-c_j)\Big]^{\frac{p}{q}}\Big] \\
&\le \frac{2^{n(\alpha-1)}}{M^\alpha}\sum_{x\in\mathcal{H}_n}\mathbb{E}\Big[\sum_{i=0}^{M-1} r(x-c_i)\sum_{k=0}^{p}\binom{p}{k} r(x-c_i)^{\frac{k}{q}}\Big[\sum_{j=0,j\neq i}^{M-1} r(x-c_j)\Big]^{\frac{p-k}{q}}\Big] \\
&= \frac{2^{n(\alpha-1)}}{M^\alpha}\sum_{k=0}^{p}\binom{p}{k}\sum_{x\in\mathcal{H}_n}\mathbb{E}\Big[\sum_{i=0}^{M-1} r(x-c_i)^{1+\frac{k}{q}}\Big[\sum_{j=0,j\neq i}^{M-1} r(x-c_j)\Big]^{\frac{p-k}{q}}\Big] \\
&= \frac{2^{n(\alpha-1)}}{M^\alpha}\sum_{k=0}^{p}\binom{p}{k}\sum_{x\in\mathcal{H}_n}\mathbb{E}\Big[\sum_{i=0}^{M-1} r(x-c_i)^{1+\frac{k}{q}}\Big]\,\mathbb{E}\Big[\sum_{j=0,j\neq i}^{M-1} r(x-c_j)\Big]^{\frac{p-k}{q}},
\end{aligned}$$

where $c_i, i=1,\dots,M$ are random codewords in the code $\mathscr{C}$. Recalling that $\mathbb{E}r(x-c_i)^a = \|r\|_a^a$ for any $a>0$, we continue as follows:

$$\begin{aligned}
&\le \frac{2^{n(\alpha-1)}}{M^\alpha}\sum_{k=0}^{p}\binom{p}{k}\sum_{x\in\mathcal{H}_n} M\|r\|_{1+k/q}^{1+k/q}\,\mathbb{E}\Big[\sum_{j=0}^{M-1} r(x-c_j)\Big]^{\frac{p-k}{q}} \\
&= \frac{2^{n(\alpha-1)}}{M^{\alpha-1}}\sum_{k=0}^{p}\binom{p}{k}\|r\|_{1+k/q}^{1+k/q}\,\mathbb{E}\Big[\sum_{x\in\mathcal{H}_n}\Big[\sum_{j=0}^{M-1} r(x-c_j)\Big]^{\frac{p-k}{q}}\Big] \\
&= \frac{2^{np/q}}{M^{p/q}}\sum_{k=0}^{p}\binom{p}{k}\|r\|_{1+k/q}^{1+k/q} Q_n\Big(\frac{p-k}{q}\Big)\frac{M^{(p-k)/q}}{2^{n((p-k)/q-1)}} \\
&= \sum_{k=0}^{p}\binom{p}{k}\frac{2^{n(1+k/q)}}{M^{k/q}}\|r\|_{1+k/q}^{1+k/q} Q_n\Big(\frac{p-k}{q}\Big) \\
&= \sum_{k=0}^{p}\binom{p}{k} 2^{\frac{nk}{q}\left(1-R-\frac{H_{(1+k/q)}(r)}{n}\right)} Q_n\Big(\frac{p-k}{q}\Big),
\end{aligned}$$

where we used (5) and the fact that r is a pmf. □

On account of (14), (A6), and Lemma 1, to prove Theorem 3, we need to prove the following:

Theorem A1. *Consider a sequence of ensembles of random codes of increasing length n and rate $R_n \to R$. If $R > 1-\pi(\alpha)$, where $\pi(\alpha)$ is given by (15), then,*

$$\lim_{n\to\infty} Q_n(\alpha) = 1 \tag{A8}$$

for all $\alpha \in (1,\infty)$.

We start with the case of rational α.

Proposition A2. *Let $\alpha \geq 0$ be rational. If $R > 1 - \pi(\alpha)$, then $\limsup_n Q_n(\alpha) \leq 1$.*

Proof. This statement is true for all $0 \leq \alpha < 1$, so also true for all rational α in [0,1).

Assume that it holds for all rational α in $[0, m)$, where $m \in \mathbb{Z}^+$. Let $\alpha \in [m, m+1)$ and choose $p, q \in \mathbb{Z}_0^+$ such that $\alpha = 1 + \frac{p}{q}$. By Lemma A2,

$$\begin{aligned}\limsup_n Q_n(\alpha) &\leq \limsup_n \sum_{k=0}^{p} \binom{p}{k} 2^{\frac{nk}{q}\left(1 - R_n - \frac{H_{1+k/q}(r_n)}{n}\right)} Q_n\left(\frac{p-k}{q}\right) \\ &\leq \sum_{k=0}^{p} \binom{p}{k} \limsup_n 2^{\frac{nk}{q}\left(1 - R_n - \frac{H_{1+k/q}(r_n)}{n}\right)} \limsup_n Q_n\left(\frac{p-k}{q}\right).\end{aligned}$$

If $R > 1 - \pi(\alpha)$, then evidently, $R > 1 - \pi(1 + k/q)$ for all $k \leq p$. Therefore,

$$\limsup_{n\to\infty} 2^{\frac{nk}{q}\left(1 - R_n - \frac{H_{1+k/q}(r_n)}{n}\right)} = 0$$

for all $k > 0$. Since $\frac{p}{q} < m$, by the induction hypothesis, we have $\limsup_n Q_n\left(\frac{p-k}{q}\right) \leq 1$ for $k = 0, 1, \ldots, p$. Therefore, all the terms except the one with $k = 0$ vanish, yielding $\limsup_n Q_n(\alpha) \leq 1$. □

Since $Q_n(\alpha) \geq 1$ for $\alpha > 1$, this proves Theorem A1 for all rational $\alpha \in (1, \infty)$.

Finally, let us extend this result to all real $\alpha > 1$. As a first step, let us show that $\pi(\alpha)$ is continuous.

Lemma A3. *$\pi(\alpha)$ is continuous for $1 < \alpha < \infty$.*

Proof. From the monotonicity of the Rényi entropies, for $\alpha' > \alpha > 1$,

$$\begin{aligned}0 &\leq \pi(\alpha) - \pi(\alpha') \\ &= \liminf_{n\to\infty} \frac{1}{n} H_\alpha(r_n) - \liminf_{n\to\infty} \frac{1}{n} H_{\alpha'}(r_n).\end{aligned}$$

Now, let us choose a subsequence $(r_{n_k})_k$ such that

$$\lim_{k\to\infty} \frac{1}{n_k} H_{\alpha'}(r_{n_k}) = \liminf_{n\to\infty} \frac{1}{n} H_{\alpha'}(r_n).$$

Therefore,

$$\begin{aligned}\pi(\alpha) - \pi(\alpha') &= \liminf_{n\to\infty} \frac{1}{n} H_\alpha(r_n) - \lim_{k\to\infty} \frac{1}{n_k} H_{\alpha'}(r_{n_k}) \\ &\leq \liminf_{k\to\infty} \frac{1}{n_k} \left(H_\alpha(r_{n_k}) - H_{\alpha'}(r_{n_k})\right).\end{aligned}$$

Note that H_α is a continuous function of the order α for $\alpha > 1$. We use the mean value theorem to claim that there is a value $\gamma_k \in (\alpha, \alpha'$ such that $H_{\alpha'}(r_{n_k}) - H_\alpha(r_{n_k}) = (\alpha' - \alpha)\frac{d}{d\alpha} H_{\gamma_k}(r_{n_k})$. Next, for any probability vector P,

$$-\frac{dH_\alpha(P)}{d\alpha} = \frac{1}{(1-\alpha)^2} D(Z \| P) \leq \frac{\log|\mathrm{supp}(P)|}{(1-\alpha)^2},$$

where $Z_i = \frac{P_i^\alpha}{\sum_j P_j^\alpha}$. Taking these remarks together, we obtain

$$\begin{aligned}\pi(\alpha) - \pi(\alpha') &\leq \liminf_{k\to\infty} \frac{1}{n_k}(\alpha - \alpha')H'_{\gamma_k}(r_{n_k}) \\ &\leq \liminf_{k\to\infty} \frac{1}{n_k}(\alpha' - \alpha)\frac{n_k}{(\gamma_k - 1)^2} \\ &= \frac{\alpha' - \alpha}{(\alpha - 1)^2},\end{aligned}$$

Therefore, $\pi(\alpha)$ is continuous on $(1, \infty)$. □

Now, let $\alpha \in (1, \infty)$ and assume $R > 1 - \pi(\alpha)$. Choose $\alpha' > \alpha$ such that α' is rational and $R > 1 - \pi(\alpha')$. This is possible from the continuity of π. Therefore,

$$1 \leq \limsup_n Q_n(\alpha) \leq \limsup_n Q_n(\alpha') = 1,$$

which proves that (A8) and Theorem 3 hold for all $\alpha \in [1, \infty)$.

It remains to address the case $\alpha = \infty$. We obtain the following upper bound, whose proof follows closely an argument in Appendix E of [3].

Lemma A4. *Let* $\epsilon > 0$. *We have*

$$\mathbb{E}_{\mathscr{C}}\|2^n T_r f_{\mathscr{C}}\|_\infty \leq 1 + \epsilon + 2^{2n - H_\infty(r)} e^{-\frac{3\epsilon^2}{2(3+\epsilon)} 2^{-[n(1-R) - H_\infty(r)]}}.$$

Proof. Let $\epsilon > 0$, then,

$$\begin{aligned}\mathbb{E}_{\mathscr{C}}\|2^n T_r f_{\mathscr{C}}\|_\infty &= \mathbb{E}_{\mathscr{C}}\Big[\|2^n T_r f_{\mathscr{C}}\|_\infty \mathbb{1}_{\|2^n T_r f_{\mathscr{C}}\|_\infty \geq 1+\epsilon\}}\Big] \\ &\quad + \mathbb{E}_{\mathscr{C}}\Big[\|2^n T_r f_{\mathscr{C}}\|_\infty \mathbb{1}_{\|2^n T_r f_{\mathscr{C}}\|_\infty < 1+\epsilon\}}\Big] \\ &\leq \mathbb{E}_{\mathscr{C}}\Big[\|2^n T_r f_{\mathscr{C}}\|_\infty \mathbb{1}_{\{\|2^n T_r f_{\mathscr{C}}\|_\infty \geq 1+\epsilon\}}\Big] + 1 + \epsilon \\ &\leq \mathbb{E}_{\mathscr{C}}\Big[\|2^n r\|_\infty \mathbb{1}_{\{\|2^n T_r f_{\mathscr{C}}\|_\infty \geq 1+\epsilon\}}\Big] + 1 + \epsilon \\ &= \|2^n r\|_\infty \Pr_{\mathscr{C}}\Big(\max_{y \in \mathcal{H}_n} 2^n T_r f_{\mathscr{C}}(y) \geq 1 + \epsilon\Big) + 1 + \epsilon \\ &\leq \|2^n r\|_\infty 2^n \max_{y \in \mathcal{H}_n} \Pr_{\mathscr{C}}\Big(2^n T_r f_{\mathscr{C}}(y) \geq 1 + \epsilon\Big) + 1 + \epsilon. \end{aligned} \tag{A9}$$

For any $y \in \mathcal{H}_n$,

$$\begin{aligned}\Pr_{\mathscr{C}}(2^n T_r f_{\mathscr{C}}(y) \geq 1 + \epsilon) &= \Pr_{\mathscr{C}}\Big(\frac{2^n}{M}\sum_{z \in \mathscr{C}} r(y - z) \geq 1 + \epsilon\Big) \\ &= \Pr_{c_i \sim U_n,\, \text{iid}}\Big(\frac{2^n}{M}\sum_{i=1}^{M} r(y - c_i) \geq 1 + \epsilon\Big) \\ &= \Pr\Big(\frac{2^n}{M}\sum_{i=1}^{M} r(y - c_i) \geq 1 + \epsilon\Big) \\ &= \Pr\Big(\sum_{i=1}^{M}(2^n r(y - c_i) - 1) \geq M\epsilon\Big)\end{aligned} \tag{A10}$$

To bound the last line from above, we use Bernstein's inequality: For independent, zero-mean random variables $X_i, i = 1, \dots, N$ such that $|X_i| \le a$ for all i,

$$P\Big(\sum_i X_i \ge t\Big) \le \exp\Big(-\frac{t^2/2}{\sum_{i=1}^n EX_i^2 + \frac{1}{3}at}\Big).$$

Note that for a random uniform vector c_i, the expectation $E[r(y - c_i)] = 2^{-n}$ since $r(\cdot)$ satisfies $\sum_{x \in \mathcal{H}_n} r(x) = 1$, so this inequality applies for (A10). We obtain

$$\begin{aligned}\Pr_{\mathscr{C}}(2^n T_r f_{\mathscr{C}}(y) \ge 1 + \epsilon) &\le \exp\Big(-\frac{\frac{1}{2}M^2\epsilon^2}{\sum_{i=1}^M \mathrm{Var}(2^n r(y - \cdot)) + \frac{1}{3}\|2^n r\|_\infty M\epsilon}\Big)\\ &\le \exp\Big(-\frac{\frac{1}{2}M^2\epsilon^2}{\sum_{i=1}^M \|2^n r\|_2^2 + \frac{1}{3}\|2^n r\|_\infty M\epsilon}\Big)\\ &\le \exp\Big(-\frac{\frac{1}{2}M^2\epsilon^2}{M\|2^n r\|_1\|2^n r\|_\infty + \frac{1}{3}\|2^n r\|_\infty M\epsilon}\Big)\\ &= \exp\Big(-\frac{3\epsilon^2}{2(3+\epsilon)}2^{-n(1-R)-H_\infty(r)}\Big).\end{aligned}$$

where on the last line, we use the equalities $\|2^n r\|_1 = 1$ and $\|2^n r\|_\infty = 2^{D_\infty(r\|U_n)}$. The proof is concluded by substituting this inequality into (A9). □

Now, let us consider a sequence of (ensembles of) random codes of increasing length n and rate $R_n \to R$. Recalling the definition of $\pi(\cdot)$ in (15), for $n \to \infty$, we obtain

$$\limsup_n \mathbb{E}_{\mathscr{C}_n} \|2^n T_{r_n} f_{\mathscr{C}_n}\|_\infty \le 1 + \epsilon \tag{A11}$$

once $R > 1 - \pi(\infty)$. Since ϵ is arbitrarily small, the left-hand side of (A11) approaches one, and together with (14) this completes the proof of Theorem 3.

Appendix C. Samorodnitsky's Inequalities and Their Implications

Samorodnitsky [8,10] recently proved certain powerful inequalities for α-norms of noisy functions, which permit us to estimate the proximity to uniformity upon action of the Bernoulli noise kernels. We state some of them in this appendix after introducing a few more elements of notation. These results are used in Theorem 7 and in Appendix D, where we prove Lemma 2.

In this proof, we write $[n]$ for $\{1, \dots, n\}$. For a subset $\Gamma \subset [n]$, write $x|_\Gamma$ to denote the coordinate projection of a vector $x \in \mathcal{H}_n$ on Γ. If the subset Γ is formed by random choice with $\Pr(i \in \Gamma) = \lambda$ independently for all $i \in [n]$, we write $\Gamma \sim \lambda$. For a function f on $\mathcal{H}_n$, let

$$\mathbb{E}(f|\Gamma)(x) = \frac{1}{2^{n-|\Gamma|}} \sum_{y: y|_\Gamma = x|_\Gamma} f(y). \tag{A12}$$

Observe that $\mathbb{E}(f|\Gamma) = f * f_{\mathcal{H}_{[n]\setminus\Gamma}}$, where $\mathcal{H}_S = \{x \in \mathcal{H}_n : x|_{[n]\setminus S} = 0\}$. Therefore, $\mathbb{E}(f|\Gamma)(x)$ is the noisy function of f with respect to the pmf given by the indicator function of the subcube $\mathcal{H}_{[n]\setminus\Gamma}$.

The *entropy* of a function $f : \mathcal{H}_n \to \mathbb{R}$ is defined as

$$\mathrm{Ent}[f] = \|f \log f\|_1 - \|f\|_1 \log(\|f\|_1) = \Big\|f \log \frac{f}{\|f\|_1}\Big\|_1. \tag{A13}$$

This quantity can be thought of as the KL divergence between the distribution induced by f on $\mathcal{H}_n$ and the uniform distribution:

$$\mathrm{Ent}[f] = \|f\|_1 D\Big(\frac{f}{\sum f}\|U_n\Big). \tag{A14}$$

If f itself is a pmf, then $D(f\|U_n) = 2^n \,\mathrm{Ent}(f) = \mathrm{Ent}(2^n f)$.

Theorem A2 ([8], Corollary 9)**.** *Let f be a non-negative function on $\mathcal{H}_n$. then,*

$$\mathrm{Ent}[T_\delta f] \le \mathbb{E}_{\Gamma\sim\lambda}\,\mathrm{Ent}[\mathbb{E}(f|\Gamma)].$$

where $\lambda = (1-2\delta)^2$.

Theorem A3 ([10], Theorem 1.1)**.** *Let f be a non-negative function on $\mathcal{H}_n$ and $\alpha \ge 2$ be an integer. Then,*

$$\log\|T_\delta f\|_\alpha \le \mathbb{E}_{\Gamma\sim\lambda}\log\|\mathbb{E}(f|\Gamma)\|_\alpha. \tag{A15}$$

where $\lambda = \lambda(\alpha,\delta) = 1 + \frac{1}{\alpha-1}\log(\delta^\alpha + (1-\delta)^\alpha) = 1 - h_\alpha(\delta)$. Furthermore,

$$\log\|T_\delta f\|_\infty \le \mathbb{E}_{\Gamma\sim\lambda}\log\|\mathbb{E}(f|\Gamma)\|_\infty. \tag{A16}$$

where $\lambda = \lambda(\infty,\delta) = 1 + \log(1-\delta) = 1 - h_\infty(\delta)$

To interpret the inequalities (A15) and (A16), we note that their left-hand side measures the smoothness of the noisy version of f with respect to the noise β_δ. At the same time, the right-hand side is the average smoothness of the noisy versions of f with respect to the sub-cube pmf's.

Hązła et al. [14] used Theorem A3 to great effect, showing that if a code corrects erasures up to a certain noise level in a BEC, then, with high probability, it corrects errors on a BSC channel up to a certain noise level.

Theorem A4 ([14], Corollary 3.4)**.** *Let $(\mathscr{C}_n)_n$ be a sequence of codes whose rate approaches R. Assume that for some $\lambda \in (0, 1-R]$, $P_B(\mathrm{BEC}(\lambda), \mathscr{C}_n) = o(\frac{1}{n})$. Then, $(\mathscr{C}_n)_n$ decodes errors on a $\mathrm{BSC}(\delta)$ for any δ that satisfies $2\sqrt{\delta(1-\delta)} < 2^\lambda - 1$.*

This theorem implies the following corollary:

Corollary A1 ([14])**.** *Let $(\mathscr{C}_n)_n$ be a sequence of codes with rate $R_n \uparrow R$ that recover transmitted messages with high probability on a $\mathrm{BEC}(1-R)$ (i.e., $(\mathscr{C}_n)_n$ is a capacity-achieving sequence for $\mathrm{BEC}(1-R)$). Furthermore, assume that $d(\mathscr{C}_n) = \omega(\log n)$. If $2\sqrt{\delta(1-\delta)} < 2^{1-R} - 1$, then with high probability, the codes $\mathscr{C}_n$ correct errors when used on a $\mathrm{BSC}(\delta)$ channel.*

The authors of [14] then used this result to show that the RM codes of a constant rate correct a non-vanishing proportion of errors on the BSC.

Appendix D. Proof of Lemma 2

We present the proof as a sequence of lemmas.

Let $\Gamma \subset \{1,\dots,n\}$ be a subset of coordinates and for $z \in (0,1\}^n$ let $\mathscr{C}(\Gamma, z) := \{c \in \mathscr{C} : c|_\Gamma = z|_\Gamma\}$ be the set of codewords that fit z in the positions of Γ. In particular, $\mathscr{C}^{\Gamma^c} = \mathscr{C}(\Gamma, 0)|_{\Gamma^c}$ is the shortened code $\mathscr{C}$, i.e., the subcode with zeros in the positions of Γ, projected on Γ^c. Let $F^{(\mathscr{C})}(\Gamma, z) := |\mathscr{C}(\Gamma, z)|$.

Let us obtain expressions for the norms and the entropy of $F^{(\mathscr{C})}(\Gamma, z)$.

Lemma A5. *Let $\mathscr{C}$ be a linear code and let $\Gamma \subset \{1,\ldots,n\}$. Then,*

$$\|F^{(\mathscr{C})}(\Gamma,\cdot)\|_\alpha = \left[\frac{|\mathscr{C}|}{2^{|\Gamma|}}F^{(\mathscr{C})}(\Gamma,0)^{\alpha-1}\right]^{1/\alpha}.$$

Proof. From the linearity of the code,

$$F^{(\mathscr{C})}(\Gamma,z) = \begin{cases} F^{(\mathscr{C})}(\Gamma,0) & \text{if } z|_\Gamma \text{ is a valid non-erasure pattern} \\ 0 & \text{otherwise.} \end{cases}.$$

Furthermore, the number of distinct $z \in \mathcal{H}_n$ for which $\mathscr{C}(\Gamma,z)$ is nonempty equals $2^{n-|\Gamma|}|\mathscr{C}/\mathscr{C}(\Gamma,0)|$. Hence,

$$\begin{aligned}\|F^{(\mathscr{C})}(\Gamma,\cdot)\|_\alpha &= \left[\frac{1}{2^n}\sum_{x\in\mathcal{H}_n} F^{(\mathscr{C})}(\Gamma,x|_\Gamma)^\alpha\right]^{1/\alpha} \\ &= \left[\frac{1}{2^n}\frac{2^n|\mathscr{C}|}{2^{|\Gamma|}}\frac{1}{F^{(\mathscr{C})}(\Gamma,0)}F^{(\mathscr{C})}(\Gamma,0)^\alpha\right]^{1/\alpha} \\ &= \left[\frac{|\mathscr{C}|}{2^{|\Gamma|}}F^{(\mathscr{C})}(\Gamma,0)^{\alpha-1}\right]^{1/\alpha}.\end{aligned}$$

□

Lemma A6. *Let $\mathbb{E}(f|\Gamma)$ be defined as in (A12). Then,*

$$\|\mathbb{E}(2^n f_{\mathscr{C}}|\Gamma)\|_\alpha = \left[\frac{2^{|\Gamma|}}{|\mathscr{C}|}F_T^{(\mathscr{C})}(\Gamma,0)\right]^{(\alpha-1)/\alpha}.$$

Proof. Using $f = 2^n f_{\mathscr{C}}$ in (A12), we obtain

$$\mathbb{E}(2^n f_{\mathscr{C}}|\Gamma)(x) = \frac{2^n}{2^{n-|\Gamma|}|\mathscr{C}|}\sum_{y\in\mathscr{C}:y|_\Gamma = x|_\Gamma} 1 = \frac{2^{|\Gamma|}}{|\mathscr{C}|}F^{(\mathscr{C})}(\Gamma,x|_\Gamma),$$

and, thus, from Lemma A5,

$$\begin{aligned}\|\mathbb{E}(2^n f_{\mathscr{C}}|\Gamma)\|_\alpha &= \frac{2^{|\Gamma|}}{|\mathscr{C}|}\|F^{(\mathscr{C})}(\Gamma,\cdot)\|_\alpha \\ &= \frac{2^{|\Gamma|}}{|\mathscr{C}|}\left[\frac{|\mathscr{C}|}{2^{|\Gamma|}}F^{(\mathscr{C})}(\Gamma,0)^{\alpha-1}\right]^{1/\alpha} \\ &= \left[\frac{2^{|\Gamma|}}{|\mathscr{C}|}F_T^{(\mathscr{C})}(\Gamma,0)\right]^{(\alpha-1)/\alpha}.\end{aligned}$$

□

Lemma A7. *Let $\mathscr{C}$ be a linear code. For $X = X_{\mathscr{C}^\perp}, Y = Y_{(X,\text{BEC}(\lambda))}$,*

$$H(X|Y) = \mathbb{E}_{\Gamma\sim\lambda}\left[\log\left(\frac{2^{|\Gamma|}}{|\mathscr{C}|}F^{\mathscr{C}}(\Gamma,0)\right)\right]. \tag{A17}$$

Proof. Start with taking $X = X_{\mathscr{C}}$ and $Y = Y_{(\mathrm{BEC}(\lambda), X_{\mathscr{C}})}$; then,

$$H(X|Y=y) = \log[F^{(\mathscr{C})}(y)].$$

Therefore,

$$\begin{aligned} H(X|Y) &= \mathbb{E}_Y[\log(F^{(\mathscr{C})}(Y))] \\ &= \mathbb{E}_\Gamma \mathbb{E}_{Z|\Gamma}[\log(F^{(\mathscr{C})}(\Gamma, Z))|\Gamma] \\ &= \mathbb{E}_{\Gamma\sim 1-\lambda}[\log F^{(\mathscr{C})}(\Gamma, 0)]. \end{aligned}$$

By a standard identity about dual matroids ([64], p. 72),

$$\dim(\mathscr{C}^{\Gamma^c}) = \dim(\mathscr{C}) - |\Gamma| + \dim((\mathscr{C}^\perp)^\Gamma),$$

or

$$F^{(\mathscr{C})}(\Gamma, 0) = \frac{|\mathscr{C}|}{2^{|\Gamma|}} F^{\mathscr{C}^\perp}(\Gamma^c, 0),$$

and, thus, we continue as follows:

$$\begin{aligned} H(X|Y) &= \mathbb{E}_{\Gamma\sim 1-\lambda}\Big[\log\Big(\frac{|\mathscr{C}|}{2^{|\Gamma|}} F^{\mathscr{C}^\perp}(\Gamma^c, 0)\Big)\Big] \\ &= \mathbb{E}_{\Gamma^c\sim\lambda}\Big[\log\Big(\frac{2^{|\Gamma^c|}}{|\mathscr{C}^\perp|} F^{\mathscr{C}^\perp}(\Gamma^c, 0)\Big)\Big] \\ &= \mathbb{E}_{\Gamma\sim\lambda}\Big[\log\Big(\frac{2^{|\Gamma|}}{|\mathscr{C}^\perp|} F^{\mathscr{C}^\perp}(\Gamma, 0)\Big)\Big]. \end{aligned}$$

Switching to the dual code and taking $X = X_{\mathscr{C}^\perp}$ and $Y = Y_{(\mathrm{BEC}(\lambda), X_{\mathscr{C}^\perp})}$ now yields (A17). □

Lemma A8.

$$\frac{\alpha}{\alpha-1}\mathbb{E}_{\Gamma\sim\lambda}\log\|\mathbb{E}(2^n f_{\mathscr{C}}|\Gamma)\|_\alpha = \mathbb{E}_{\Gamma\sim\lambda}\,\mathrm{Ent}[\mathbb{E}(2^n f_{\mathscr{C}}|\Gamma)] = H(X_{\mathscr{C}^\perp}|Y_{(\mathrm{BEC}(\lambda), X_{\mathscr{C}^\perp})}).$$

Proof. From Lemmas A6 and A7,

$$\begin{aligned} \frac{\alpha}{\alpha-1}\mathbb{E}_{\Gamma\sim\lambda}\log\|\mathbb{E}(2^n f_{\mathscr{C}}|\Gamma)\|_\alpha &= \mathbb{E}_{\Gamma\sim\lambda}\Big[\log\Big(\frac{2^{|\Gamma|}}{|\mathscr{C}|} F^{\mathscr{C}}(\Gamma, 0)\Big)\Big] \\ &= H(X_{\mathscr{C}^\perp}|Y_{(\mathrm{BEC}(\lambda), X_{\mathscr{C}^\perp})}), \end{aligned}$$

which establishes the equality between the first and the third quantities. Since the second quantity is a limiting case of the first quantity and the value of the first quantity is independent of α, we have equality between the first and the second quantities. □

Now, Lemma 2 follows by combining Lemma A8 with Theorems A2 and A3.

References

1. Han, T.S.; Verdú, S. Approximation theory of output statistics. *IEEE Trans. Inf. Theory* **1993**, *39*, 752–772. [CrossRef]
2. Hayashi, M. General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel. *IEEE Trans. Inf. Theory* **2006**, *52*, 1562–1575. [CrossRef]
3. Yu, L.; Tan, V.Y. Rényi resolvability and its applications to the wiretap channel. *IEEE Trans. Inf. Theory* **2019**, *65*, 1862–1897. [CrossRef]
4. Debris-Alazard, T.; Ducas, L.; Resch, N.; Tillich, J.-P. Smoothing codes and lattices: Systematic study and new bounds. *IEEE Trans. Inf. Theory* **2023**, *69*, 6006–6027. [CrossRef]
5. Micciancio, D.; Regev, O. Worst-case to average-case reductions based on Gaussian measures. *SIam J. Comput.* **2007**, *37*, 267–302. [CrossRef]

6. Chen, W.W.L.; Skriganov, M.M. Explicit constructions in the classical mean squares problem in irregularities of point distribution. *J. Fur Die Reine Und Angew. Math.* **2002**, *545*, 67–95. [CrossRef]
7. Skriganov, M.M. Coding theory and uniform distributions. *Algebra Anal.* **2001**, *13*, 191–239. Translation in *St. Petersburg Math. J.* **2002**, *13*, 301–337 .
8. Samorodnitsky, A. On the entropy of a noisy function. *IEEE Trans. Inf. Theory* **2016**, *62*, 5446–5464. [CrossRef]
9. Samorodnitsky, A. An upper bound on ℓ_q norms of noisy functions. *IEEE Trans. Inf. Theory* **2019**, *66*, 742–748. [CrossRef]
10. Samorodnitsky, A. An improved bound on ℓ_q norms of noisy functions. *arXiv* **2020**, arXiv:2010.02721. [CrossRef]
11. Bloch, M.R.; Laneman, J.N. Strong secrecy from channel resolvability. *IEEE Trans. Inf. Theory* **2013**, *59*, 8077–8098. [CrossRef]
12. Belfiore , J.-C.; Oggier, F. Secrecy gain: A wiretap lattice code design. In Proceedings of the 2010 International Symposium on Information Theory & Its Applications, Taichung, Taiwan, 17–20 October 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 174–178. [CrossRef]
13. Luzzi, L.; Ling, C.; Bloch, M.R. Optimal rate-limited secret key generation from Gaussian sources using lattices. *IEEE Trans. Inf. Theory* **2023**, *69*, 4944–4960. [CrossRef]
14. Hązła, J.H.; Samorodnitsky, A.; Sberlo, O. On codes decoding a constant fraction of errors on the BSC. In Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual, 21–25 June 2021; pp. 1479–1488. [CrossRef]
15. Rao, A.; Sprumont, O. A criterion for decoding on the BSC. *arXiv* **2022**, arXiv:2202.00240. [CrossRef]
16. Arimoto, S. On the converse to the coding theorem for discrete memoryless channels (corresp.). *IEEE Trans. Inf. Theory* **1973**, *19*, 357–359. [CrossRef]
17. Polyanskiy, Y.; Verdú, S. Arimoto channel coding converse and Rényi divergence. In Proceedings of the 2010 48th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, USA, 29 September–1 October 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 1327–1333. [CrossRef]
18. Polyanskiy, Y.; Verdú, S. Empirical distribution of good channel codes with nonvanishing error probability. *IEEE Trans. Inf. Theory* **2013**, *60*, 5–21. [CrossRef]
19. Chou, R.A.; Bloch, M.R.; Kliewer, J. Empirical and strong coordination via soft covering with polar codes. *IEEE Trans. Inf. Theory* **2018**, *64*, 5087–5100. [CrossRef]
20. Cover, T.M.; Permuter, H.H. Capacity of coordinated actions. In Proceedings of the 2007 IEEE International Symposium on Information Theory, Nice, France, 24–29 June 2007; IEEE: Piscataway, NJ, USA, 2007; pp. 2701–2705. [CrossRef]
21. Cuff, P. Distributed channel synthesis. *IEEE Trans. Inf. Theory* **2013**, *59*, 7071–7096. [CrossRef]
22. Cuff, P.W.; Permuter, H.H.; Cover, T.M. Coordination capacity. *IEEE Trans. Inf. Theory* **2010**, *56*, 4181–4206. [CrossRef]
23. Chou, R.A.; Bloch, M.R.; Abbe, E. Polar coding for secret-key generation. *IEEE Trans. Inf. Theory* **2015**, 61, 6213–6237. [CrossRef]
24. Brakerski, Z.; Lyubashevsky, V.; Vaikuntanathan, V.; Wichs, D. Worst-case hardness for LPN and cryptographic hashing via code smoothing. In Proceedings of the Annual International Conference on the Theory and Applications of Cryptographic Techniques, Paris, France, 30 April– 4 May 2017; Springer: Berlin/Heidelberg, Germany, 2019; pp. 619–635. [CrossRef]
25. Goldfeld, Z.; Kato, K.; Nietert, S.; Rioux, G. Limit distribution theory for smooth p-Wasserstein distances. *arXiv* **2022**, arXiv:2203.00159. [CrossRef]
26. Goldfeld, Z.; Kato, K.; Rioux, G.; Sadhu, R. Statistical inference with regularized optimal transport. *arXiv* **2022**, arXiv:2205.04283. [CrossRef]
27. Nietert, S.; Goldfeld, Z.; Kato, K. Smooth p-Wasserstein distance: Structure, empirical approximation, and statistical applications. In Proceedings of the International Conference on Machine Learning, Virtual, 8–24 July 2021; pp. 8172–8183.
28. Liu, J.; Cuff, P.; Verdú, S. E_γ-resolvability. *IEEE Trans. Inf. Theory* **2016**, *63*, 2629–2658. [CrossRef]
29. Steinberg, Y.; Verdú, S. Simulation of random processes and rate-distortion theory. *IEEE Trans. Inf. Theory* **1996**, *42*, 63–86. [CrossRef]
30. Ordentlich, O.; Polyanskiy, Y. Entropy under additive Bernoulli and spherical noises. In Proceedings of the 2018 IEEE International Symposium on Information Theory (ISIT), Vail, CO, USA, 17–22 June 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 521–525. [CrossRef]
31. Polyanskiy, Y. Hypercontractivity of spherical averages in Hamming space. *Siam J. Discret. Math.* **2019**, *33*, 731–754. [CrossRef]
32. Yu, L. Edge-isoperimetric inequalities and ball-noise stability: Linear programming and probabilistic approaches. *J. Comb. Theory Ser. A* **2022**, *188*, 105583. [CrossRef]
33. Wyner, A.; Ziv, J. A theorem on the entropy of certain binary sequences and applications–I. *IEEE Trans. Inf. Theory* **1973**, *19*, 769–772. [CrossRef]
34. Hązła, J.H. Optimal list decoding from noisy entropy inequality. *arXiv* **2022**, arXiv:2212.01443. [CrossRef]
35. Wyner, A.D. The wire-tap channel. *Bell Syst. Tech. J.* **1975**, *54*, 1355–1387. [CrossRef]
36. Csiszár, I. Almost independence and secrecy capacity. *Probl. Peredachi Informatsii* **1996**, *32*, 48–57. English translation in *Probl. Inform. Transm.* **1996**, *32*, 40–47.
37. Maurer, U.M. Secret key agreement by public discussion from common information. *IEEE Trans. Inf. Theory* **1993**, *39*, 733–742. [CrossRef]
38. Thangaraj, A.; Dihidar, S.; Calderbank, A.R.; McLaughlin, S.W.; Merolla, J.-M. Applications of LDPC codes to the wiretap channel. *IEEE Trans. Inf. Theory* **2007**, *53*, 2933–2945. [CrossRef]

39. Mahdavifar, H.; Vardy, A. Achieving the secrecy capacity of wiretap channels using polar codes. *IEEE Trans. Inf. Theory* **2011**, *57*, 6428–6443. [CrossRef]
40. Subramanian, A.; Suresh, A.T.; Raj, S.; Thangaraj, A.; Bloch, M.; McLaughlin, S. Strong and weak secrecy in wiretap channels. In Proceedings of the 2010 6th International Symposium on Turbo Codes & Iterative Information Processing, Brest, France, 6–10 September 2010; pp. 30–34. [CrossRef]
41. Gulcu, T.C.; Barg, A. Achieving secrecy capacity of the wiretap channel and broadcast channel with a confidential component. *IEEE Trans. Inf. Theory* **2016**, *63*, 1311–1324. [CrossRef]
42. Barg, A. Stolarsky's invariance principle for finite metric spaces. *Mathematika* **2021**, *67*, 158–186. [CrossRef]
43. Bilyk, D.; Dai, F.; Matzke, R. The Stolarsky principle and energy optimization on the sphere. *Constr. Approx.* **2018**, *48*, 31–60. [CrossRef]
44. Skriganov, M.M. Point distributions in two-point homogeneous spaces. *Mathematika* **2019**, *65*, 557–587. [CrossRef]
45. Simon, B. *Real Analysis: A Comprehensive Course in Analysis, Part 1*; American Mathematical Society: Providence, RI, USA, 2015. [CrossRef]
46. Shamai, S.; Verdú, S. The empirical distribution of good codes. *IEEE Trans. Inf. Theory* **1997**, *43*, 836–846. [CrossRef]
47. Polyanskiy, Y.; Poor, H.V.; Verdú, S. Channel coding rate in the finite blocklength regime. *IEEE Trans. Inf. Theory* **2010**, *56*, 2307–2359. [CrossRef]
48. Bloch, M.R.; Luzzi, L.; Kliewer, J. Strong coordination with polar codes. In Proceedings of the 2012 50th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, USA, 1–5 October 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 565–571. [CrossRef]
49. Tillich, J.-P.; Zémor, G. Discrete isoperimetric inequalities and the probability of a decoding error. *Comb. Probab. Comput.* **2000**, *9*, 465–479. [CrossRef]
50. Ashikhmin, A.; Barg, A. Bounds on the covering radius of linear codes. *Des. Codes Cryptogr.* **2002**, *27*, 261–269. [CrossRef]
51. Watanabe, S.; Hayashi, M. Strong converse and second-order asymptotics of channel resolvability. In Proceedings of the 2014 IEEE International Symposium on Information Theory, Honolulu, HI, USA, 29 June–4 July 2014; pp. 1882–1886. [CrossRef]
52. Kudekar, S.; Kumar, S.; Mondelli, M.; Pfister, H.D.; Şaşoğlu, E.; Urbanke, R. Reed-Muller codes achieve capacity on erasure channels. In Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing, Cambridge, MA, USA, 19–21 June 2016; pp. 658–669. [CrossRef]
53. Abbe, E.; Sandon, C. A proof that Reed-Muller codes achieve Shannon capacity on symmetric channels. *arXiv* **2023**, arXiv:2304.02509. [CrossRef]
54. Reeves, G.; Pfister, H.D. Reed-Muller codes achieve capacity on BMS channels. *arXiv* **2021**, arXiv:2110.14631. [CrossRef]
55. Renes, J.M. Duality of channels and codes. *IEEE Trans. Inf. Theory* **2018**, *64*, 577–592. [CrossRef]
56. Rengaswamy, N.; Pfister, H.D. On the duality between the BSC and quantum PSC. In Proceedings of the 2021 IEEE International Symposium on Information Theory (ISIT), Melbourne, Australia, 12–20 July 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 2232–2237. [CrossRef]
57. Pfister, H. (Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA); Rengaswamy, N. (Department of Electrical and Computer Engineering, University of Arizona, Tucson, AZ, USA). Personal communication, 2023.
58. Poltyrev, G. Bounds on the decoding error probability of binary linear codes via their spectra. *IEEE Trans. Inf. Theory* **1994**, *40*, 1284–1292. [CrossRef]
59. Semakov, N.; Zinov'ev, V.A.; Zaitsev, G. Uniformly packed codes. *Probl. Peredachi Informatsii* **1971**, *7*, 38–50.
60. Goethals, J.-M.; van Tilborg, H.C.A. Uniformly packed codes. *Philips Res. Rep.* **1975**, *30*, 9–36.
61. Tokareva, N. An upper bound for the number of uniformly packed codes. In Proceedings of the 2007 IEEE International Symposium on Information Theory, Nice, France, 24–29 June 2007; IEEE: Piscataway, NJ, USA, 2007; pp. 346–349. [CrossRef]
62. Borges, J.; Rifà, J.; Zinoviev, V.A. On completely regular codes. *Probl. Inf. Transm.* **2019**, *55*, 1–45. [CrossRef]
63. Delsarte, P. An algebraic approach to the association schemes of coding theory. *Philips Res. Repts. Suppl.* **1973**, *10*, 1973.
64. Oxley, J. *Matroid Theory*; Oxford University Press: Oxford, UK, 1992.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.