When are Continuous-Source Fuzzy Extractors Possible?

Benjamin Fuller* Lowen Peng†

May 21, 2018

Abstract

Fuzzy extractors (Dodis et al., Eurocrypt 2004) convert repeated noisy readings of a high-entropy source into the same uniformly distributed key. The ideal functionality of a fuzzy extractor outputs the key when provided with a value close to the original reading of the source. A necessary condition for security, called fuzzy min-entropy, is that the probability of every ball of values of the noisy source is small.

Noisy sources are measured from physical phenomena many of which best modeled by continuous metric spaces. To build *continuous-source fuzzy extractors*, prior work assumes that the system designer has a good model of the high (fuzzy) entropy distribution (Verbitskiy et al., IEEE TIFS 2010). However, it is impossible to build an accurate model of a high entropy distribution with oracle access to the distribution.

We show that model inaccuracy is a major hurdle to constructing a continuous-source fuzzy extractors. Namely, there exists a family of continuous distributions $\mathcal{W}$ such that each element $W \in \mathcal{W}$ has fuzzy min-entropy but no fuzzy extractor can produce a three bit key for an average element of $\mathcal{W}$. Our family is built from random p-ary lattices.

We also show a stronger negative result for secure sketches, which are used to construct most fuzzy extractors. Our results are for the Euclidean metric and are information-theoretic in nature. To the best of our knowledge all continuous-source fuzzy extractors argue information-theoretic security.

Fuller, Reyzin, and Smith showed negative results for a discrete metric space equipped with the Hamming metric (Asiacrypt 2016). The geometry of Euclidean space necessitates new techniques.

Keywords: Fuzzy extractor, secure sketch, information theory, key derivation.

1 Introduction

Reproducible secret random bits enable cryptographic applications. Many physical processes have entropy but have exhibit noise between readings of the same process [BBR88, BS00, Dau04, EHMS00, GCVDD02, MG09, PRTG02, SD07, TSS+06]. When a secret is read multiple times, readings are close (according to some metric dis) but not identical. To utilize such sources, it is often necessary to remove noise.

The cryptography goal is for two parties to interactively engage in a protocol to derive the same key from nearby secrets. This problem has a rich history starting from Wyner [Wyn75] and Bennett, Brassard, and Robert [BBR88]. These works identify information-reconciliation (removing noise without

*University of Connecticut. Email: benjamin.fuller@uconn.edu
†University of Connecticut. Email: lowen.peng@uconn.edu

leaking information) and privacy amplification (converting an entropic secret to random). In this work we focus on the non-interactive and information-theoretically secure setting.

In the non-interactive setting information reconciliation is performed by a secure sketch which is a pair of algorithms $(\mathsf{SS}, \mathsf{Rec})$ [DORS08]. The sketch algorithm SS takes an initial reading w and produces a nonsecret helper value ss. The recover Rec algorithm takes a subsequent reading w' along with ss and outputs w. The correctness guarantee is that w is reproduced if $\mathsf{dis}(w, w') \leq t$ for metric dis. The security requirement for a secure sketch is that w is hard to predict given ss.

A fuzzy extractor performs both information reconciliation and privacy amplification, producing a uniform but stable key [DORS08]. Fuzzy extractors similarly consist of two algorithms. The generate algorithm (Gen) takes an initial reading w and produces an output key along with a nonsecret helper value pub. The reproduce (Rep) algorithm takes w' and pub to reproduce key. The key should be reproduced when $\mathsf{dis}(w, w') \leq t$. The security requirement for fuzzy extractors is that key is statistically close to uniform knowing pub. In this work we consider sources with continuous values, the primitive is known as a *continuous source* fuzzy extractor [BDHV07]. We consider an n-dimension space equipped Euclidean distance where $\mathsf{dis}(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$.

Prior work on continuous source fuzzy extractors uses one of two approaches [LSM06, BDHV07, VTO+10]: providing security for specific distributions or introducing a quantization step in the definition of security. The first approach is undesirable as we can't model high (fuzzy) entropy distributions. In the second approach, security is measured with respect to some quantization of the input space. That is, the metric space is partitioned into different values of key. Li et al. [LSM06, Theorem 6] showed that the choice of quantization may be far from optimal. In the worst case, the quantization decrease the length of the derived key by a factor linear in the dimension of the space ($\Theta(n)$) compared to the optimal quantization. When the metric space is a bounded subset of $\mathbb{R}^n$ the desired security level is often linear in the dimension. Thus, measuring security only after quantization may obscure whether key derivation is possible.

Fuller, Reyzin, and Smith [FRS16] introduced fuzzy min-entropy. Fuzzy min-entropy codifies the adversary's success when provided with only the *functionality* of a fuzzy extractor (or secure sketch). Suppose w is sampled from a distribution W. The adversary's best strategy is to find w' that maximizes the weight of possible $w \in W$ within distance t. Denote by $B_t(w)$ the closed ball of radius t around w. Fuzzy min-entropy is formally defined as

$$\mathrm{H}^{\mathsf{fuzz}}_{t,\infty}(W) \stackrel{\text{def}}{=} -\log\left(\max_{w'} \Pr[W \in B_t(w')]\right).$$

Fuzzy min-entropy is suitable for continuous distributions including quantization in the definition. Implicitly, the definition takes an integral over a ball of radius $t > 0$ and thus finite on bounded metric spaces in finite dimensions.

In this work, we ask whether continuous source fuzzy extractors can be built for distributions with fuzzy min-entropy. Since fuzzy extractor designers are securing high entropy distributions they must work with models of the distribution. After deployment, the adversary can build a more accurate model. Fuzzy extractors usually secure a family of sources, $\mathcal{W}$. This is called the *distributional uncertainty* setting. We ask a natural question:

Do continuous source fuzzy extractors exist for all families of sources with fuzzy min-entropy?

Fuller, Reyzin, and Smith [FRS16] answered in the negative for discrete Hamming space.

1.1 Our Contribution and Techniques

We show a family of distributions $\mathcal{W}$ where no fuzzy extractor or secure sketch can secure an average element of $\mathcal{W}$ with more than 3 bits of security (Theorems 2 and 3 respectively). This answers the above question negatively.

The secure sketch result is qualitatively stronger as it holds even if the secure sketch is allowed to be wrong a constant fraction of the time. For secure sketches we consider the metric space as $[0,1]^n$. Our results can be extended to other bounded subsets of $\mathbb{R}^n$. A technicality requires our fuzzy extractor result to consider $(\mathbb{R}/\mathbb{Z})^n$ instead of $[0,1]^n$. We discuss this more below.

In both results we consider distributions W that have fuzzy min-entropy $\mathrm{H}^{\mathsf{fuzz}}_{t,\infty}(W) = \Theta(n)$ and algorithms that correct $t = \Theta(\sqrt{n})$ errors. One may be able to avoid these results when less error tolerance is required or more fuzzy min-entropy is present. To illustrate techniques we first focus on secure sketches (which are subject to stronger negative results). We show a family $\mathcal{W}$ that satisfies two properties:

1. When the secure sketch SS sees $w \leftarrow W$ for some $W \in \mathcal{W}$ it is difficult to tell which distribution it originated from. Intuitively this forces SS to try and secure all distributions W simultaneously. For a secure sketch this corresponds to choosing points $x \in [0,1]^n$ such that most points x' where $\mathsf{dis}(x,x') \leq t$, $\mathsf{Rec}(x', ss) = x$. Call of the set of all such points X. The points in X form a Shannon error correcting code. This implies that $\forall x, y \in X$, $\mathsf{dis}(x,y) \geq t/2$. Since X forms a code, one can bound the size of X using packing arguments.
2. Show that the adversary gains additional information by knowing the distribution W. That is, the size of $X \cap W$ is small. This is true if W consists of points that are the output of a universal hash family [CW79]. Intuitively, this is because seeing one point $w \in W$ does not give much information about the rest of W's support.

In the fuzzy extractor setting the first property is more delicate. This is because a continuous region is allowed to map to the same key. Thus, rather than considering consistent points, the fuzzy extractor partitions the metric space based on what key is output by Rep. We call this partition Q_{key}. The functionality of a fuzzy extractor demands that the true w lies in the *interior* of some part in Q_{key}. The adversary can then limit their search to points in the interior.

A major goal of the proof is deriving bounds on the size of the interiors of a partition. Euclidean space is more challenging than Hamming space for two reasons:

1. Volume grows exponentially with the radius of the ball in Euclidean space. This growth is super exponential in Hamming space, growing exponentially in the binary entropy of the radius (see for example [Gur10]). This slower rate of volume growth increases the size of X (resp. the size of the interior of parts of Q_{key}) for secure sketches (resp. for fuzzy extractors).
2. In the fuzzy extractor case, the interior of the parts in Q_{key} is continuous which precludes the use of counting techniques.

In place of counting techniques, we show the volume of the interior is smaller than the "volume" of different distributions. That is, we show that any region V of fixed volume v must not include some distributions in $\mathcal{W}$. To make this argument, our family $\mathcal{W}$ consists of all lattices of a fixed dimension and minimum distance. We consider each W to be a coset of a lattice. For V to contain a point from every coset, v must be as large as the Voronoi cell of the lattice. The fraction of distributions not represented in V is at least the ratio of v to the volume of the Voronoi cell.

The mod cube In the fuzzy extractor case, we use the "mod" cube $(\mathbb{R}/\mathbb{Z})^n$ as the input metric space rather than the unit cube $[0,1]^n$. In the proof we show that the volume of the interior of Q_{key} is smaller than the volume of Q_{key}. Roughly, we show the maximum security drops by a factor proportional to the ratio between these volumes. To get a $o(n)$ bound on key length this ratio must be exponential in the dimension n.

If we consider the metric space $[0,1]^n$ most parts of Q_{key} can be on a boundary of the unit cube. In the worst case these objects can be 1-dimension so their interior volume is only a constant factor smaller than their total volume.

We consider this to be an artifact of working with the unit cube. If a fuzzy extractor only secure points on the boundary then the data does not vary in all n dimensions. Since extraneous dimensions complicate error-correction, a system designer would use dimensionality reduction techniques to find a representation that varies across all dimensions. This transformed distribution would be used for stable key derivation. In the "mod" space there are no boundary points, the entire region is "n-dimensional."

The positive results of [VTO+10] Verbitskiy et al. [VTO+10] construct a continuous-source fuzzy extractor that requires knowledge of the probability distribution of W. This is called the *precisely known distribution or distribution sensitive setting*. The idea is to sample two independent partitions A and B of the input metric space. The input w's part in A is the key while w's part in B is the public value pub. For security, the probability mass in $A_{\mathsf{key}} \cap B_{\mathsf{pub}}$ must be the same for all $\mathsf{key}, \mathsf{pub}$. If so, knowledge of pub gives no information about key. The resulting key length is $|\mathsf{key}| = \log |A|$. For correctness, for any $w \in A_{\mathsf{key}} \cap B_{\mathsf{pub}}$ and $w' \in A_{\mathsf{key}'} \cap B_{\mathsf{pub}}$ it should be that $\mathsf{dis}(w, w') \geq t$.

The value pub leaks information about w so its size should be minimized. Each part B_{pub} is represented in each part of A. Thus, as $|A|$ increases it becomes more difficult to pack each B_{pub} while ensuring the minimum distance condition. Verbitskiy et al. do not describe how to construct A and B when $n > 1$. It is not clear which distributions their construction can secure. In contrast, Woodage et al. describe a construction for discrete distributions that secures each distribution with fuzzy min-entropy [WCD+17]. Woodage et al.'s construction is for Hamming space but extends to other metrics since it only requires evaluation of whether points are within distance t.

Verbitskiy et al. [VTO+10] extend their construction to work in the distributional uncertainty setting. They show security when the statistical distance between the observed distribution $\tilde{W}$ and the actual distribution W is small. To get a meaningful key, W must have an exponential number of disjoint balls with nonzero probability. Thus, there is no reason to expect a bound on the statistical distance between $\tilde{W}$ and W. When the original construction is instantiable it can be extended to the case where W can be estimated.

We view the gap between our results and those of Vertibskiy et al. as demonstrating a family $\mathcal{W}$ that can not be estimated. However, we have not been able to verify Verbitskiy et al.'s construction can secure each element in the family $\mathcal{W}$.

Our results do not preclude using methods other than sampling to create a model such as an expert's understanding of the physical process.

Other models A line of research views w and w' as samples from a correlated pair of random variables [Wyn75, CK78, AC93, Mau93, RW05, TW15, HTW14]. Key length is bounded based on mutual information. We do not consider this model in this work. Some discrete fuzzy extractors provide computational security [FMR13] [CFP+16, HRvD+17, ACEK17, ABC+18, WLH18]. We are not aware of continuous-source fuzzy extractors that argue computational security. Computational security is a natural way to avoid our results.

Organization The remainder of this paper is organized as follows, Section 2 covers basic notation and mathematical prerequisites, Section 3 shows the family of distributions $\mathcal{W}$ that is used in both negative results, Section 4 shows our fuzzy extractor negative result, and Section 5 shows our secure sketch negative result. We focus on the fuzzy extractor negative result as it is more challenging and shows the interesting aspects of the geometry.

2 Preliminaries

Random Variables We use uppercase letters for random variables and corresponding lowercase letters for their samples. Multiple occurrence of the same random variable in an expression signifies the same value of the random variable: for example $(W, \mathsf{SS}(W))$ is a pair of random variables obtained by sampling w according to W and applying the algorithm SS to w. The *statistical distance* between random variables A and B with the same domain is

$$\mathbf{SD}(A, B) = \frac{1}{2}\sum_{a} |\Pr[A = a] - \Pr[B = b]| = \max_{S} \Pr[A \in S] - \Pr[B \in S].$$

Entropy All logarithms in this work are base 2. Let (X, Y) be a pair of random variables. Define *min-entropy* of X as

$$\mathrm{H}_\infty(X) = -\log(\max_x \Pr[X = x]).$$

The *average (conditional)* min-entropy [DORS08, Section 2.4] of X given Y is

$$\tilde{\mathrm{H}}_\infty(X|Y) = -\log(\mathop{\mathbb{E}}_{y\in Y} \max_x \Pr[X = x|Y = y]).$$

Define Hartley entropy $H_0(X)$ to be the logarithm of the size of the support of X, that is $H_0(X) = \log|\{x| \Pr[X = x] > 0\}|$. Define average-case Hartley entropy by averaging the support size:

$$\tilde{H}_0(X|Y) = \log(\mathop{\mathbb{E}}_{y\in Y} |\{y| \Pr[X = x|Y = y] > 0\}|).$$

Metric Spaces and Balls For a metric space $(\mathcal{M}, \mathsf{dis})$, the *(closed) ball of radius t around w* is the set of all points within radius t, that is, $B_t(w) = \{w'|\mathsf{dis}(w, w') \le t\}$. In this work we consider the Euclidean distance (L_2 metric) over vectors defined via $\mathsf{dis}(w, w') = \sqrt{(\sum_{i=1}^{n}(w_i - w_i')^2)}$.

Mod space $(\mathbb{R}/\mathbb{Z})^n$ Our first result considers vectors in $(\mathbb{R}/\mathbb{Z})^n$ in this metric the maximum distance between any two points is $\sqrt{n/2}$ (the distance between $(0, 0, ..., 0)$ and $(1/2, ..., 1/2)$). Volume in this space is

$$|B_t| = \frac{\pi^{n/2} t^n}{\Gamma(n/2 + 1)}. \tag{1}$$

as long as $t \in [0, \sqrt{n}/2]$. Here Γ is the Γ function. For simplicity we restrict our results to $n = 2k$ where $\Gamma(2k/2 + 1) = k!$.

Unit Cube Our second result considers vectors in the unit cube $[0,1]^n$. In this metric the maximum distance between any two points is $\sqrt{n}$ (the distance between $(0,0,...,0)$ and $(1,1,...,1)$). The volume of a ball in this space depends on whether the point is near a "boundary." We use $|B_t|$ to denote the volume of a ball of radius t around an arbitrary point. This volume is bounded by:

Lemma 2.1. *When n is even the volume $|B_t|$ of the ball of radius in $t \in [0, \sqrt{n}]$ in the L_2 metric over $[0,1]^n$ satisfies*

$$\frac{\pi^{n/2} t^n}{(n/2)! 2^n} \leq |B_t| \leq \frac{\pi^{n/2} t^n}{(n/2)!}.$$

2.1 Fuzzy Extractors

Fuzzy extractors derive stable keys from noisy sources.

Definition 2.2. *[DORS08] An $(\mathcal{M}, \mathcal{W}, \kappa, t, \epsilon)$-*fuzzy extractor *is a pair* $(\mathsf{Gen}, \mathsf{Rep})$*.* Gen *on input $w \in \mathcal{M}$ outputs an extracted string* $\mathsf{key} \in \{0,1\}^\kappa$ *and a helper string* $\mathsf{pub} \in \{0,1\}^*$*.* Rep *takes $w' \in \mathcal{M}$ and* $\mathsf{pub} \in \{0,1\}^*$ *as inputs.* $(\mathsf{Gen}, \mathsf{Rep})$ *have the following properties:*

1. Correctness: *if* $\mathsf{dis}(w, w') \leq t$ *and* $(\mathsf{key}, \mathsf{pub}) \leftarrow \mathsf{Gen}(w)$, $\Pr[\mathsf{Rep}(w', \mathsf{pub}) = \mathsf{key}] = 1$.

2. Security: $\forall W \in \mathcal{W}$, *if* $(\mathsf{Key}, \mathsf{Pub}) \leftarrow \mathsf{Gen}(W)$, $\mathbf{SD}((\mathsf{Key}, \mathsf{Pub}), (U_\kappa, \mathsf{Pub})) \leq \epsilon$.

Note: Often designers consider the family containing all distributions with a certain amount of (fuzzy) min-entropy instead of an arbitrary family $\mathcal{W}$. Every element in our family has fuzzy min-entropy so our results hold if the fuzzy extractor secures all distributions with enough fuzzy min-entropy.

Recovering w from w' forms the core of many fuzzy extractor constructions. The primitive that performs just recovery is called a *secure sketch*. We recall the definition from [DORS08, Section 3.1]:

Definition 2.3. *An $(\mathcal{M}, \mathcal{W}, \tilde{m}, t)$-*secure sketch *with error δ is a pair* $(\mathsf{SS}, \mathsf{Rec})$*.* SS *on input $w \in \mathcal{M}$ returns a bit string $ss \in \{0,1\}^*$.* Rec *takes an element $w' \in \mathcal{M}$ and $ss \in \{0,1\}^*$.* $(\mathsf{SS}, \mathsf{Rec})$ *have the following properties:*

1. Correctness*: $\forall w, w' \in \mathcal{M}$ if $\mathsf{dis}(w, w') \leq t$ then $\Pr[\mathsf{Rec}(w', \mathsf{SS}(w)) = w] \geq 1 - \delta$.*

2. Security*: for any distribution $W \in \mathcal{W}$, $\tilde{H}(W | \mathsf{SS}(W)) \geq \tilde{m}$.*

Note: The functionality of secure sketches ensures that when $w_1, w_2 \in W | \mathsf{SS}(W)$ the points w_1 and w_2 cannot be too close to each other. In particular, this implies that the set $W | \mathsf{SS}(W)$ is discrete and thus $\tilde{H}(W | \mathsf{SS}(W))$ is well defined. This is not true for fuzzy extractors where $W|P$ can be continuous.

Fuller, Smith, and Reyzin [FRS16] proposed *fuzzy min-entropy* to measure suitability of a noisy distribution for key extraction. Fuzzy min-entropy captures the adversary's success probability when provided with the functionality of the primitive. Fuzzy min-entropy measures ideal security. We adopt this notion:

Definition 2.4. *The t-fuzzy min-entropy of a distribution W in a metric space $(\mathcal{M}, \mathsf{dis})$ is:*

$$\mathrm{H}^{\mathtt{fuzz}}_{t,\infty}(W) = -\log \left(\max_{w'} \sum_{w \in \mathcal{M} | \mathsf{dis}(w, w') \leq t} \Pr[W = w] \right)$$

Previous work considered a definition of fuzzy extractors where $\mathcal{W}$ is over a continuous space [LSM06, BDHV07, VTO$^+$10]. Usually such definitions consider a quantization of the space (see for example, [VTO$^+$10, Section 3.1]). Such definitions mix the fuzzy extractor construction with the definition of security. Since fuzzy min-entropy measures the maximal probability ball it "discretizes" the space. This discretization only considers the functionality of Rep rather than any construction.

3 The family of distributions $\mathcal{W}$

In this section we describe the family of distributions $\mathcal{W}$ used in our negative results for both fuzzy extractors and secure sketches. We use different properties of this family in the two negative results, however, all of the properties are achieved by the same family of functions. Our negative results are for an average element of $\mathcal{W}$. Thus, instead of thinking of the adversary as receiving the description of W we think of the adversary receiving Z where Z describes the uniform choice of W from $\mathcal{W}$ and we use W_Z to refer to an individual distribution. We define Z to be the restriction of the uniform distribution to points that have a particular output for a specified element of a hash family. Z consists of two components $z = (\mathsf{A}, \mathsf{h})$ and $W_z = \{w | \mathsf{Hash}_{\mathsf{A}}(w) = \mathsf{h}\}$.

The key to both results is showing that it is hard to recover A, h from a single w and that the hash has good geometric properties. The required properties are: 1) universality 2) regularity 3) the set W_z minimum distance and 4) a large volume is required to cover every possible output of the hash family for every fixed A.

The Hash Function The hash function we use is the coset of the input point with respect to a random p-ary lattice with minimum distance $\geq t$. Let $\mathcal{K}$ be the set of lattices of all p-ary lattices $\Lambda_p(\mathsf{A})$ where $\mathsf{A} \in \mathbb{Z}_p^{n\times m}$ with minimum distance t defined by $\Lambda_p(\mathsf{A}) = \{y \in \mathbb{Z}_p^n : y = \mathsf{A}s \mod p \text{ for some } s \in \mathbb{Z}^m\}$. Define $\mathsf{Hash}_{\mathsf{A}\in\mathcal{A}} : (\mathbb{R}/\mathbb{Z})^n \to (\mathbb{R}/p\mathbb{Z})^m/\Lambda_p(\mathsf{A})$ be defined by

$$x \mapsto [px]_{\Lambda_p(\mathsf{A})}$$

where we understand $[px]_{\Lambda_p(\mathsf{A})}$ to be a coset containing px with respect to the lattice. Scaling a random lattice to the unit cube is known as Construction A and is well studied in the lattice packing literature (Conway and Sloane [CS13]). In our presentation we expand the input point rather than compressing the lattice.

Note: The family is stated with respect to the input space $(\mathbb{R}/\mathbb{Z})^n$. This metric space will be used in Section 4. Section 5 uses the metric $[0,1]^n$. We only use the first three properties in Section 5 and these properties carry over to $[0,1]^n$.

Theorem 3.1. *Let p be some prime and let $n, m \in \mathbb{Z}^+$ such that $m = \mu n$ for some $\mu \in (0, 1/2)$. For some matrix $\mathsf{A} \in \mathbb{Z}_p^{n\times m}$ define the lattice $\Lambda_p(\mathsf{A}) = \{Ax | x \in \mathbb{Z}_p^m\}$. Let $\mathcal{A}$ be the set of all lattices with minimum distance $t' = tp = \tau p\sqrt{n}$ where $\tau = \frac{1}{6p^\mu\sqrt{2e}}$. Define $\mathsf{Hash}_{\mathsf{A}\in\mathcal{A}}(w) = [pw]_{\mathsf{A}}$. If $p \geq (3\sqrt{2e})^{1/(1-\mu)}$ the following are simultaneously achieved:*

1. *is 2^{-a}-universal for $a = (n-m)\log p - 1$, that is*

$$\forall v_1 \neq v_2 \in (\mathbb{R}/\mathbb{Z})^n, \Pr_{\mathsf{A}\leftarrow\mathcal{A}}[\mathsf{Hash}_{\mathsf{A}}(v_1) = \mathsf{Hash}_{\mathsf{A}}(v_2)] \leq 2^{(n-m)\log p - 1},$$

2. *is p^m regular, that is*

$$\forall \mathsf{A} \in \mathcal{A}, h \in \mathtt{Range}(\mathsf{Hash}_{\mathsf{A}}), |\mathsf{Hash}_{\mathsf{A}}^{-1}(h)| \geq p^m,$$

3. preimage sets have minimum distance t for $t = \tau\sqrt{n}$, that is

$$\forall \mathsf{A} \in \mathcal{A}, v_1 \neq v_2, \text{ if } \mathsf{Hash}_\mathsf{A}(v_1) = \mathsf{Hash}_\mathsf{A}(v_2) \text{ then } \mathsf{dis}(v_1, v_2) \geq t,$$

4. and has $p^{-\mu n}$-preimage volume, that is $\forall \mathsf{A} \in \mathcal{A}, V \subseteq (\mathbb{R}/\mathbb{Z})^n$,

$$\Pr_{h \overset{\$}{\leftarrow} \mathtt{Range}(\mathsf{Hash}_\mathsf{A})} \left[\mathsf{Hash}_\mathsf{A}^{-1}(h) \cap V \neq \emptyset\right] \leq \frac{\mathsf{Vol}(V)}{p^{-\mu n}}.$$

Proof. Let $\mathcal{A}$ be the the set of all p-ary lattices of rate $m = \mu n$, length n, and minimum distance $t' = tp$. That is,

$$\mathcal{A} = \left\{ \mathsf{A} | \mathsf{A} \in \mathbb{Z}_p^{n \times m}, \dim(A) = m, \min_{k \in K - \{0^n\}} \mathsf{dis}(k, 0^n) > t' \right\}.$$

Universality We show $2p^{-(m-n)}$-universality by first considering a slightly larger hash family. Let $\mathcal{A}'$ be the set of *all* m-dimensional lattices over $[0,1]^n$. That is, consider $\mathcal{A}'$ where

$$\mathcal{A}' = \{\mathsf{A} | \mathsf{A} \in \mathbb{Z}_p^{m \times n}, \dim(\mathsf{A}) = m\}.$$

Define $\mathsf{Hash}_{\mathcal{A}'}$ as a hash function with this larger set of keys where the evaluation is still the coset after multiplication by p. This hash function is universal. Fix $v \neq w$ and write

$$\begin{aligned}
\Pr_{\mathsf{K} \in \mathcal{K}'} [\mathsf{Hash}_\mathsf{K}(v) = \mathsf{Hash}_\mathsf{K}(w)] &= \Pr_{\mathsf{K}} [\mathsf{Hash}_\mathsf{K}(v - w) = 0] \\
&= \Pr_{\mathsf{K}} [v - w \in \ker \mathsf{Hash}_\mathsf{K}] \\
&= \Pr_{\mathsf{K}} [[p(v - w)]_\mathsf{K} = 0] = p^{m-n}
\end{aligned}$$

The last equality follows because the elements of $\mathsf{A} \in \mathcal{A}'$ are all m-dimensional subspaces. Thus, every nonzero point is included in the null space with probability p^{m-n}. We now show that the set $\mathcal{A}'$ is not much bigger than $\mathcal{A}$.

We now show most p-ary lattices of dimension m have minimum distance $> t$. Our theorem is based on the result of Erez et al. which show this construction (known as Construction A) is good for packing in Euclidean space [ELZ05].

Lemma 3.2. *Let n be an even integer. Let p be a prime, let $\mu \in (0, 1/2)$ and let $m = \mu n$. Suppose that*

$$t \leq \frac{\sqrt{n}}{3(2e)^{1/2} p^\mu} - \frac{\sqrt{n}}{2p}$$

Then the defined hash function has minimum distance t with high probability across $\mathcal{A}$. That is,

$$\Pr_{\mathsf{A} \overset{\$}{\leftarrow} \mathcal{A}} [\mathsf{Hash}_\mathsf{A} \text{ has minimum distance } t] \geq 1/2.$$

In particular, when $p \geq (3\sqrt{2e})^{1/(1-\mu)}$ then the conditions are fulfilled when $t = \frac{\sqrt{n}}{6p^\mu \sqrt{2e}}$.

Proof. First note that we consider even n, this restriction is done to simplify calculations with the Γ function but is not key to the result. By the Theorem statement and Stirling's formula,

$$\begin{aligned} t &\leq \frac{\sqrt{n}}{3(2e)^{1/2}p^{\mu}} - \frac{\sqrt{n}}{2p} \\ &\leq \frac{\sqrt[n]{\sqrt{2\pi}(n/2)^{n/2}e^{-(n/2)}}}{3p^{\mu}\pi^{1/2}} - \frac{\sqrt{n}}{2p} \\ &\leq \frac{\sqrt[n]{(n/2)!}}{3p^{\mu}\pi^{1/2}} - \frac{\sqrt{n}}{2p} \\ &= \frac{\sqrt[n]{\Gamma(n/2+1)}}{3p^{\mu}\pi^{1/2}} - \frac{\sqrt{n}}{2p} \end{aligned}$$

Erez et al. [ELZ05, Theorem 1] show that this lattice achieves this packing radius with good probability. In particular, define $r = \frac{\sqrt[n]{\Gamma(n/2+1)}}{p^{\mu}\pi^{1/2}}$ and $d = \sqrt{n}/(2p)$. Our requirement on t implies that $t \leq r/3 - d/2$. Their proof shows that,

$$\begin{aligned} \Pr_{\mathsf{A} \xleftarrow{\$} \mathcal{A}} [\mathsf{Hash}_{\mathsf{A}} \text{ has minimum distance } t] &\geq 1 - \left(\frac{2t+d}{r}\right)^n \\ &\geq 1 - \left(\frac{2(r/3 - d/2) + d}{r}\right)^n \\ &= 1 - (2/3)^n \end{aligned}$$

The fact that this value is at least $1/2$ follows as n is even. We note that when $p^{1-\mu} \geq 3(2e)^{1/2}$ then for the maximum value of t,

$$\begin{aligned} t &= \frac{\sqrt{n}}{3(2e)^{1/2}p^{\mu}} - \frac{\sqrt{n}}{2p} \geq \frac{\sqrt{n}}{3(2e)^{1/2}p^{\mu}} - \frac{\sqrt{n}}{2p^{1-\mu}p^{\mu}} \\ &\geq \frac{\sqrt{n}}{3(2e)^{1/2}p^{\mu}} - \frac{\sqrt{n}}{6(2e)^{1/2}p^{\mu}} \geq \frac{\sqrt{n}}{6(2e)^{1/2}p^{\mu}}. \end{aligned}$$

This completes the proof of Lemma 2. □

Regularity To show p^m-regularity, fix K, h and write

$$|\mathsf{Hash}_{\mathsf{K}}^{-1}(h)| \geq |\ker \mathsf{Hash}_{\mathsf{K}}| = |\{v : \mathsf{Hash}_{\mathsf{K}}(v) = 0\}| \geq p^m$$

Since there are m linearly independent lattice vectors in $\mathbb{Z}_p^m$ and p possible coefficients that produce distinct vectors (by linear independence).

Minimum distance This condition is immediately implied by the minimum distance of the lattice.

Preimage Volume Finally, note that any μn dimensional lattice has $p^{\mu n}$ points. The Voronoi region of every lattice point in $(\mathbb{R}/\mathbb{Z})^n$ is the same. Since the volume of the unit cube is 1 this means each region has volume $p^{-\mu n}$. This completes the proof that a hash function with the required parameters exists and completes the proof of Theorem 1. □

4 No fuzzy extractor can secure $\mathcal{W}$

We now prove it is impossible to build a fuzzy extractor that secures $\mathcal{W}$. As discussed in the introduction we use $(\mathbb{R}/\mathbb{Z})^n$ as the input space equipped with the Euclidean metric.

Theorem 4.1. *Let $\gamma \geq 1$ be a constant. Let $\mathcal{M} = (\mathbb{R}/\mathbb{Z})^n$. Let $\mu \in [0, 1/2)$ be a constant and define $m = \mu n$. Then there exists a family $\mathcal{W}$ such that, for all $W \in \mathcal{W}$, $\mathrm{H}^{\mathsf{fuzz}}_{t,\infty} = \mathrm{H}_\infty(W) \geq m$. Let $(\mathsf{Gen}, \mathsf{Rep})$ be a $(\mathcal{M}, \mathcal{W}, \kappa, t, \epsilon)$-fuzzy-extractor with perfect correctness with* noise rate $\tau \overset{d}{=} t/\sqrt{n}$ *where the following conditions hold:*

1. *Let p be a prime integer parameter such that $p \geq (3\sqrt{2e})^{1/(1-\mu)}$.*
2. *Noise rate $\tau = \frac{1}{6p^\mu\sqrt{2e}}$*
3. *The key length $\kappa \geq 1 + \max\left\{0, \log\gamma + n\left(\log p^\mu - \log(\frac{6\sqrt{e}+\sqrt{\pi}}{6\sqrt{e}})\right)\right\}$.*

Then $\epsilon \geq \frac{1}{2} - \frac{e}{2\sqrt{2\pi\gamma}}$.

Parameter discussion: There are settings of $\mu, \tau, \kappa = \Theta(1)$ such that the statistical distance ϵ is a constant. Taking $\log p^{-\mu} \leq \log(\frac{6\sqrt{e}+\sqrt{\pi}}{6\sqrt{e}}) \approx -.1334$ implies that κ only needs to satisfy $\kappa \geq 1 + \log\gamma$. Substituting $p \geq (3\sqrt{2e})^{1/(1-\mu)}$ and ignoring factors due to finding a prime p this condition holds when $\mu \leq .045$. When $\gamma = 4$ then $\epsilon \geq .35$ (when $\kappa \geq 3$). The full setting of achievable parameter ranges for a constant κ, ϵ are in Figure 1.

Proof. We show the impossibility for an average member of $\mathcal{W}$. Recall that we think of the distribution $W \in \mathcal{W}$ as being described by an auxiliary variable Z that is a pair (A, h) where $W_z = \{w | \mathsf{Hash}_\mathsf{A}(w) = \mathsf{h}\}$. The hash function we use is $\mathsf{Hash}_{\mathsf{A}\in\mathcal{A}} : (\mathbb{R}/\mathbb{Z})^n \to (\mathbb{R}/p\mathbb{Z})^m/\Lambda_p(\mathsf{A})$ be defined by

$$x \mapsto [px]_{\Lambda_p(\mathsf{A})}$$

where $[px]_{\Lambda_p(\mathsf{A})}$ the coset of the input point with respect to A. The conditions of Theorem 2 implies those of Theorem 1 and thus we can use Theorem 1. For this proof we need the regularity, minimum distance, and preimage volume conditions. By the 2^m-regularity and minimum distance properties of Hash, $\forall z \in Z, \mathrm{H}_\infty(W_z) = \mathrm{H}^{\mathsf{fuzz}}_{t,\infty}(W_z) = m$.

We now want to show that for a random $z \leftarrow Z$, if $(\mathsf{key}, \mathsf{pub})$ is the output of $\mathsf{Gen}(W_z)$, then key can be easily distinguished from uniform in the presence of pub and z. The outline for the proof is as follows:

- In the absence of information about z, the value w is uniform.
- pub partitions the key space (since there is perfect correctness).
- Each part is the partition created by pub is bounded in size.
- Valid w can only come from the interior of a part (by correctness of Rep, every candidate input w to Gen must have all of its neighbors w' produce the same output of $\mathsf{Rep}(w', \mathsf{pub})$).
- The volume of the interior of a part is smaller than the volume of a part.
- For many parts have interior volume smaller than the preimage volume of the lattice (the volume of the Voronoi region of the lattice).
- Many elements $W \in \mathcal{W}$ have no point in the interior of the part.

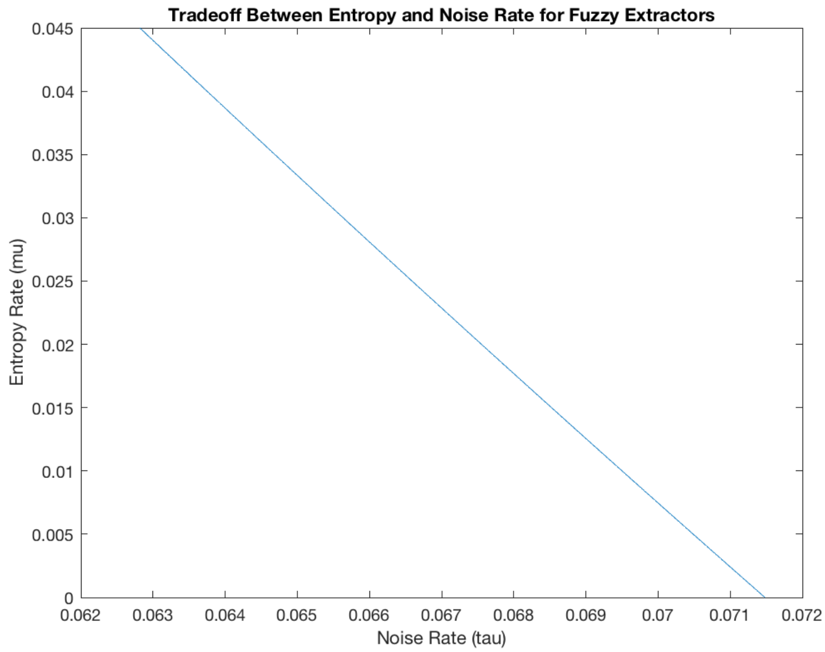

Figure 1: Tradeoff between entropy rate μ and error rate τ for fuzzy extractor. Illustration of parameters in Theorem 2. The figure looks linear due to the fine grained scaled but the relationship between μ and τ is not linear. In this analysis we assume that there always exists a prime of size exactly $3(2e)^{1/(2(1-\mu))}$. The allowed noise rate τ may actually be smaller to accounting for finding such a prime. Recall that Bertrand's postulate state's that a prime exists between n and $2n$ for any integer $n > 1$.

- By averaging across parts, the average distribution W has no points in the interior of many parts.
- It is possible to distinguish a random key from one produced by Gen by checking if it comes from a part whose interior has no preimage in W.

We now proceed with the formal proof. The following lemma bounds the volume of the smallest parts.

Lemma 4.2. *Suppose $\mathcal{M}$ is $(\mathbb{R}/\mathbb{Z})^n$ with the Euclidean metric, $\kappa \geq 2$, $0 \leq t \leq \frac{\sqrt{n}}{2}$, and $\epsilon \geq 0$. Suppose $(\mathsf{Gen}, \mathsf{Rep})$ is a $(\mathcal{M}, \mathcal{W}, \kappa, t, \epsilon)$-fuzzy extractor for distribution family $\mathcal{W}$ over $\mathcal{M}$. For any fixed pub, there is a set $\mathsf{GoodKey}_{\mathsf{pub}} \subset \{0,1\}^\kappa$ of size $2^{\kappa-1}$ such that,*

$$\forall \mathsf{key} \in \mathsf{GoodKey}_{\mathsf{pub}}, \mathsf{Vol}|\{v \in \mathcal{M} | (\mathsf{key}, \mathsf{pub}) \in \mathrm{supp}(\mathsf{Gen}(v))\}| \leq 2^{-\kappa+1}.$$

Proof. Recall that $\mathsf{Vol}([0,1]^n) = 1$. The set $\mathsf{GoodKey}_{\mathsf{pub}}$ consists of the $2^{\kappa-1}$ keys with the smallest volume (breaking ties arbitrarily). Note that for all $\mathsf{key} \in \mathsf{GoodKey}_{\mathsf{pub}}$, $\mathsf{Vol}(\{v \in \mathcal{M} | \mathsf{Rep}(v, \mathsf{pub}) = \mathsf{key})\}) \leq 2^{-\kappa+1}$. If not, then $\cup_{\mathsf{key}} \mathsf{Vol}(Q_{\mathsf{pub},\mathsf{key}}) > 1$ because for every $\mathsf{key} \notin \mathsf{GoodKey}_{\mathsf{pub}}$, $\mathsf{Vol}(\{v \in \mathcal{M} | \mathsf{Rep}(v, \mathsf{pub}) = \mathsf{key})\}) > 2^{-\kappa+1}$. This completes the proof of Lemma 3. □

We now proceed to show on $\mathsf{key} \in \mathsf{GoodKey}_{\mathsf{pub}}$ the size of the interior is bounded. By perfect correctness of Rep, the input w to Gen has the following property: for all w' within distance t of w, $\mathsf{Rep}(w', \mathsf{pub}) = \mathsf{Rep}(w, \mathsf{pub})$. Thus, if we partition $\mathcal{M}$ according to the output of Rep, the true w is t away from the boundary of a part. Interior sets are small, which means the set of possible of w values is small. (Rep has a deterministic output even if the algorithm is randomized, so this partition is well-defined.)

To formalize this intuition, fix pub and partition $\mathcal{M}$ according to the output of $\mathsf{Rep}(\cdot, \mathsf{pub})$ as follows: let $Q_{\mathsf{pub},\mathsf{key}} = \{w' \in \mathcal{M} | \mathsf{Rep}(w', \mathsf{pub}) = \mathsf{key}\}$. Note that there are 2^κ keys and thus 2^κ parts $Q_{\mathsf{pub},\mathsf{key}}$. For the remainder of the proof we focus on elements in $\mathsf{GoodKey}_{\mathsf{pub}}$. As explained above, if w is the input to Gen, then every point w' within distance t of w must be in the same part $Q_{\mathsf{pub},\mathsf{key}}$ as w, by correctness of Rep. Thus, w must come from the interior of some $Q_{\mathsf{pub},\mathsf{key}}$, where interior is defined as

$$\mathsf{Inter}(Q_{\mathsf{pub},\mathsf{key}}) = \{w \in Q_{\mathsf{pub},\mathsf{key}} | \forall w' \text{ s.t. } \mathsf{dis}(w, w') \leq t, w' \in Q_{\mathsf{pub},\mathsf{key}}\}.$$

We now use the isoperimetric inequality to bound the size of $\mathsf{Inter}(Q_{\mathsf{pub},\mathsf{key}})$. The proof for this lemma is in Appendix A.1. The bounds on κ from the theorem statement and the volume of the Voronoi region of the lattices contained in $\mathcal{W}$ are crucial for this proof.

Lemma 4.3. *Define all parameters as in Theorem 2. Then for any fixed*

$$\forall p, \forall \mathsf{key} \in \mathsf{GoodKey}_{\mathsf{pub}}, \mathsf{Vol}(\mathsf{Inter}(Q_{p,\mathsf{key}})) \leq \frac{e}{\sqrt{2\pi} \cdot \gamma \cdot p^{\mu n}}.$$

Lemma 4 implies that for each $\mathsf{key} \in \mathsf{GoodKey}_{\mathsf{pub}}$, the interior $\mathsf{Vol}(\mathsf{Inter}(Q_{p,\mathsf{key}}))$ is smaller than the volume of the Voronoi region by a factor of $\frac{e}{\sqrt{2\pi\gamma}}$. This means that for an average $W \in \mathcal{W}$,

$$\Pr_{W \leftarrow \mathcal{W}, \kappa \xleftarrow{\$} \mathsf{GoodKey}_{\mathsf{pub}}} [\mathsf{Inter}(Q_{\mathsf{pub},\mathsf{key}}) \cap W = \emptyset] \geq 1 - \frac{e}{\sqrt{2\pi\gamma}}.$$

Thus, on average across $z = (k, h)$ a $1 - \frac{e}{\sqrt{2\pi\gamma}}$ fraction of keys in $\mathsf{GoodKey}_{\mathsf{pub}}$ (that is, overall $\frac{1}{2} - \frac{e}{2\sqrt{2\pi\gamma}}$ fraction of keys cannot be produced). Define the set $\mathsf{Implausible} = \{\mathsf{key}, \mathsf{pub}, z | \mathsf{Inter}(Q_{\mathsf{pub},\mathsf{key}}) \cap W_z = \emptyset\}$. Triples drawn by creating a key using the fuzzy extractor never come from the set implausible. However, a uniformly random key will land in this set with probability $\frac{1}{2} - \frac{e}{2\sqrt{2\pi\gamma}}$. Thus, $\epsilon \geq \frac{1}{2} - \frac{e}{2\sqrt{2\pi\gamma}}$. This completes the proof of Theorem 2. □

5 No Secure Sketch can secure $\mathcal{W}$

We now show no secure sketch can be secure for an average member of $\mathcal{W}$. We consider the metric space $[0, 1]^n$ but we can embed other bounded, continuous spaces of finite dimension into the unit cube.

Theorem 5.1. *Let $\mathcal{M} = [0, 1]^n$ with the Euclidean metric dis where n is even positive integer. Let $\mu \in [0, \frac{1}{2})$ be a constant and define $m \stackrel{\text{d}}{=} \mu n$. There is a family $\mathcal{W}$ where for all $W \in \mathcal{W}$, $\mathrm{H}^{\mathsf{fuzz}}_{t,\infty}(W) = \mathrm{H}_\infty(W) \geq m$, such that for any $(\mathcal{M}, \mathcal{W}, \tilde{m}, t)$-secure sketch with error δ, we have $\tilde{m} \leq 3$ provided the following conditions hold:*

1. *$n \geq 2(h_2(2\delta) + \log e)$. Note that $n \geq 6$ suffices.*
2. *Let p be a prime integer parameter such that $p \geq (3\sqrt{2e})^{\frac{1}{(1-\mu)}}$*
3. *Define the* noise rate *$\tau \stackrel{\text{d}}{=} t/\sqrt{n}$ where $\tau = \frac{1}{6p^\mu\sqrt{2e}}$.*
4. *The error parameter δ satisfies*

$$\delta \in \left[0, \frac{1}{2} - \frac{\log(1/\tau)}{2(1-\mu)\log p}\right)$$

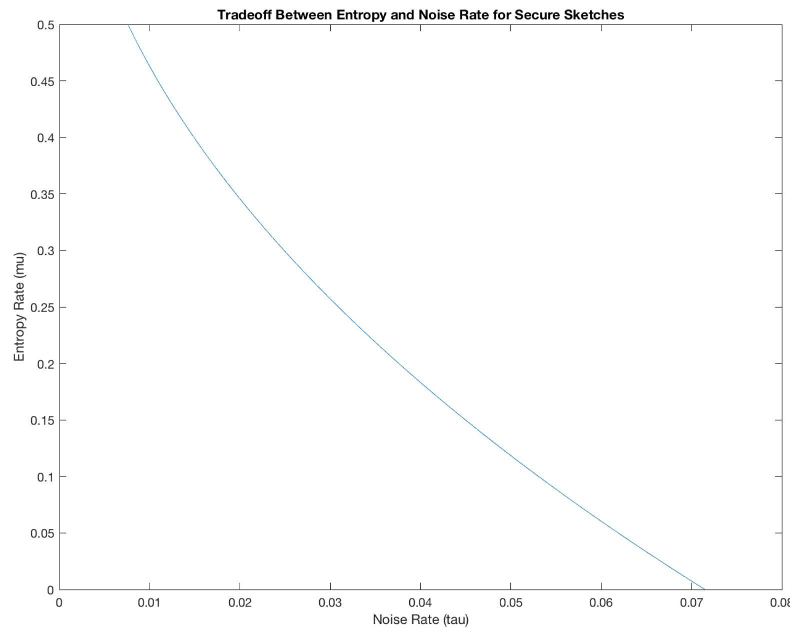

Figure 2: Tradeoff between entropy rate μ and error rate τ for secure sketch. Illustration of parameters in Theorem 3. In this analysis we assume that there always exists a prime of size exactly $3(2e)^{1/(1-\mu)}$. The allowed noise rate τ may be reduced to find such a prime. Recall that Bertrand's postulate state's that a prime exists between n and $2n$ for any integer $n > 1$. The scale here is different than Figure 1.

Parameters: As an example, $\mu = .3$ implies that $p \geq 129$, so if we consider $p = 131$, we have that $\tau \leq .017$ and $\delta \leq .08$. That is, there is a family with constant fuzzy entropy, constant error rate, and constant error where no good secure sketch exists. The trade-off between μ and τ is illustrated in Figure 2.

Interpreting the result: A secure sketch "discretizes" the input space into regions that produce a consistent value. Thus it is not surprising that, like the discrete case, a continuous secure sketch is not always possible. As stated in the introduction, the geometry of the Euclidean metric is more challenging than the Hamming metric due to the slower growth of volume.

Proof Sketch: We think of the interaction between the secure sketch designer and the adversary as follows:

1. The adversary samples $W_z \leftarrow \mathcal{W}$ and learns z.
2. The challenger samples $w \leftarrow W_z$.
3. The challenger runs ss $\leftarrow \mathsf{SS}(w)$.
4. The adversary learns ss and z and guesses w.

The proof uses the family introduced in Section 3. First we show that $W_Z|\mathsf{SS}(W_Z)$ is discrete and forms an error correcting code. We can then use standard arguments to bound the size of $W_Z|\mathsf{SS}(W_Z)$. We then rely on the universality of the lattice to show that $W_Z|(\mathsf{SS}(W_Z), Z)$ is small. That is, that Z provides fresh information that is not contained in $\mathsf{SS}(W_Z)$.

Proof. Recall that we use the family described in Theorem 1, the uniform distribution over the support of a coset h of a random lattice with minimum distance t. In this proof we will use the fact that for each $W \in \mathcal{W}$ the hash family is universal, regular, and has minimum distance.

We will define $Z = (\mathsf{A}, \mathsf{h})$ to be the pair of the lattice and coset and consider $\mathcal{W} = \{W_z\}$. Let n, m, μ, t, τ and p be defined as in the theorem statement. The core of the proof is showing two properties of this hash function:

1. The family restricted to a hash function and output value has fuzzy min-entropy. This is because the hash function is regular and has minimum distance.

2. The adversary learns from sketch value $\mathsf{SS}(W_Z)$ and *new* information by seeing the hash function and value. This is because the hash function is universal.

The first property is immediate, by the $2^{m \log p}$-regularity and minimum distance properties of $\mathsf{Hash}_{\mathcal{A}}$, $\mathrm{H}_\infty(W_z) = \mathrm{H}^{\mathtt{fuzz}}_{t,\infty}(W_z) = m \log p \geq m$. We now proceed to show that some values of sketch that occur with good probability decrease the number of possible input values.

Lemma 5.2. *Let $\mathcal{M}$ denote the Euclidean mod-space $[0,1]^n$ and $|B_t|$ denote the volume of a sphere of radius t in $\mathcal{M}$. Suppose $(\mathsf{SS}, \mathsf{Rec})$ is a $(\mathcal{M}, \mathcal{W}, \tilde{m}, t)$ secure sketch with error δ, for some distribution family $\mathcal{W}$ over $\mathcal{M}$. Then for every $v \in \mathcal{M}$ there exists a set $\mathsf{GoodSketch}_v$ such that $\Pr[\mathsf{SS}(v) \in \mathsf{GoodSketch}_v] \geq 1/2$ and for any fixed* ss*,*

$$\log |\{v \in \mathcal{M} | \mathrm{ss} \in \mathsf{GoodSketch}_v\}| \leq \frac{h_2(2\delta) - \log |B_t|}{1 - 2\delta},$$

and, therefore, for any distribution $D_{\mathcal{M}}$ over $\mathcal{M}$,

$$\mathrm{H}_0(D_{\mathcal{M}} | \mathrm{ss} \in \mathsf{GoodSketch}_{D_{\mathcal{M}}}) \leq \frac{h_2(2\delta) - \log |B_t|}{1 - 2\delta}.$$

Proof. For any $v \in M$, define $\mathsf{Neigh}_t(v)$ be the uniform distribution on the ball of radius t around v and let

$$\mathsf{GoodSketch}_v = \{\mathrm{ss} | \Pr_{v' \leftarrow \mathsf{Neigh}_t(v)}[\mathsf{Rec}(v', ss) \neq v] \leq 2\delta]\}.$$

We prove the lemma by showing two propositions. The first is a simple application of the Markov inequality shown by Fuller et al. [FRS16, Proposition D.2]

Proposition 5.3. *For $v \in \mathcal{M}$, $\Pr[\mathsf{SS}(v) \in \mathsf{GoodSketch}_v] \geq 1/2$.*

To finish the proof of Lemma 5, we show that the set $\{v \in \mathcal{M} | ss \in \mathsf{GoodSketch}_v\}$ forms an error-correcting code and bound the size of the code.

Definition 5.4. *We say that a set C is an (t, δ)-*Shannon code *if there exists a (possibly randomized) function* Decode *such that for all $c \in C$,*

$$\Pr_{c' \leftarrow \mathsf{Neigh}_t(c)}[\mathsf{Decode}(c') \neq c] \leq \delta.$$

The set $\{v \in \mathcal{M} | ss \in \mathsf{GoodSketch}_v\}$ forms $(t, 2\delta)$ Shannon code if we set $\mathsf{Decode}(y) = \mathsf{Rec}(y, ss)$. We now bound the size of such a code.

Lemma 5.5. *If $C \subset [0,1]^n$ is a (t,δ)-Shannon code for any $t > 0$, then $|C| < \infty$ and*

$$\log|C| \le \frac{h_2(\delta) - \log|B_t|}{1-\delta} \tag{2}$$

Proof of Lemma 6. First we show $|C| < \infty$. Suppose for contradiction that $|C| = \infty$. For any $\epsilon > 0$, we can find $p, q \in C$ such that $|p - q| < \epsilon$. Let $B_t(p), B_t(q)$ be the t-radius balls centered on p, q respectively. Then we can choose p, q, ϵ such that $|B_t(p) \cap B_t(q)| \ge 2\delta$ and $|p - q| < \epsilon < t$. Let R be a uniform distribution on $B_t(p) \cap B_t(q)$. Clearly

$$\Pr[\mathsf{dis}(R, p) \le t] \Pr[\mathsf{dis}(R, q) \le t] = 1.$$

If $\Pr[\mathsf{Decode}(R) \neq p] \le 1/2$, we necessarily have $\Pr[\mathsf{Decode}(R) \neq q] \ge 1/2$. Without loss of generality assume that $\Pr[\mathsf{Decode}(R) \neq q] \ge 1/2$. Then we have that

$$\Pr_{c' \leftarrow \mathsf{Neigh}_t(q)}[\mathsf{Decode}(c') \neq c] \ge \Pr_{c' \leftarrow R}[\mathsf{Decode}(c') \neq c] \Pr[c' \in R] > \frac{2\delta}{2} = \delta.$$

By contradiction, we know $|C| < \infty$.

This we can assume that C is finite. Let X be a uniform distribution on C and Y a uniform distribution on the t-radius ball centered on X. Here we use a variant of Fano's inequality [Fan61] when Y is a continuous random variable. This formulation is different from most continuous formulation's of Fano's inequality which point the Euclidean norm of the estimate [CD09, Lemma 2]. We do not prove this formulation, the proof is the same as the standard discrete formulation of Fano's which only relies on $\hat{X}$ and X.

Lemma 5.6. *Let X be a discrete random variable on $\mathcal{M}$, Y be a continuous random variable on $\mathcal{M}$ with $\hat{X}(Y)$ a discrete estimator of X based on Y. Then,*

$$H_1(X|\hat{X}) \le h_2(\delta) + \delta(\log|C|).$$

where $\delta = \Pr[X \neq \hat{X}]$.

In particular, we consider the following estimator $\hat{X}(y)$, if there exists a single $x \in C$ such that $\mathsf{dis}(x, y) \le t$ output x. If there are multiple x_i such that $\forall i, \mathsf{dis}(x_i, y) \le t$ then output a random y. Then we have that

$$H_1(X|\hat{X}) \ge H_0(X|\hat{X}) \ge \log\left(\frac{(|C| * |B_t|)}{|[0,1]^n}\right) = \log|C| + \log|B_t|.$$

Here the second inequality proceeds by noting that that $|C| * |B_t|/|[0,1]^n$ measures the thickness of the space and thus the average number of possible points $x \in C$ within distance t for a uniform point y around a codeword. Thickness is usually used to describe the quality of a covering radius. Here by thickness we simply mean the total volumes of the balls of radius t divided by the size of the space. Combining these two facts yields that Equation 5. This completes the proof of Lemma 6. □

Lemma 5 follows from Lemma 6. □

Since the hash is universal, entropy drops further when the adversary learns $\mathcal{A}, \mathsf{h}$. Let M denote the uniform distribution on $\mathcal{M}$ and K denote the uniform distribution on $\mathcal{K}$. We first recall that a universal hash function reduces entropy of any distribution with small enough support:

Lemma 5.7. *[FRS16, Lemma B.2] Let L be a distribution. Let $\{\mathsf{Hash}_{\mathcal{A}}\}_{\mathcal{A}\in\mathcal{K}}$ be a family of 2^{-a}-universal hash functions on the support of L. Assume $\mathcal{A}$ is uniform in $\mathcal{K}$ and independent of L. Then*

$$\tilde{H}_0(L|\mathcal{A}, \mathsf{Hash}_{\mathcal{A}}(L)) < \log(1 + |\operatorname{supp}(L)| \cdot 2^{-a}) \leq \max(1, 1 + H_0(L) - a)\,.$$

Applying Lemma 8 to Lemma 5, we get that for any ss,

$$\begin{aligned}&\tilde{H}_0(\mathsf{M}|\mathrm{ss} \in \mathsf{GoodSketch}_{\mathsf{M}}, \mathcal{A}, \mathsf{Hash}_{\mathcal{A}}(\mathsf{M})) \\ &\quad < \max\left(1, 2 + \frac{h_2(2\delta) - \log|B_t|}{1-2\delta} - (n-m)\log p\right)\,. \end{aligned} \tag{3}$$

We note that $\tilde{H}_0$ serves as a bound on $\tilde{\mathrm{H}}_\infty$ (see [FRS16, Lemma D.5]). That is,

$$\begin{aligned}&\tilde{H}_\infty(\mathsf{M}|\mathrm{ss} \in \mathsf{GoodSketch}_{\mathsf{M}}, \mathcal{A}, \mathsf{Hash}_{\mathcal{A}}(\mathsf{M})) \\ &\quad < \max\left(1, 2 + \frac{h_2(2\delta) - \log|B_t|}{1-2\delta} - (n-m)\log p\right)\,. \end{aligned}$$

We need just two more lemma technical lemmas:

Lemma 5.8. *[FRS16, Lemma D.6] For any pair of random variables (X, Y) and event η that is a (possibly randomized) function of (X, Y), $\tilde{\mathrm{H}}_\infty(X|\eta, Y) \geq \tilde{\mathrm{H}}_\infty(X|Y) - \log 1/\Pr[\eta]$.*

The second technical lemma bounds the size of ball in the Euclidean space:

Lemma 5.9. *Let p be a prime such that $p \geq (3\sqrt{2e})^{1/(1-\mu)}$ then*

$$\alpha \stackrel{d}{=} \frac{h_2(2\delta) - \log|B_t|}{1-2\delta} - (1-\mu)n\log p \leq 0.$$

Proof. Due to Lemma 1 and Stirling's formula, we have that

$$\begin{aligned}
\alpha &= \frac{h_2(2\delta) - \log|B_t|}{1-2\delta} - (1-\mu)n\log p \\
&\le \frac{h_2(2\delta) + n + \log((n/2)!) - n/2\log\pi - n\log t}{1-2\delta} - (1-\mu)n\log p \\
&\le \frac{h_2(2\delta) + n(1 - \frac{\log\pi}{2}) + \log e(n/2)^{n/2+1/2}e^{-n/2} - n\log t}{1-2\delta} - (1-\mu)n\log p \\
&= \frac{h_2(2\delta) + \log e + n(1 - \frac{\log\pi}{2} - \frac{\log e}{2}) + \log(n/2)^{n/2+1/2} - n\log t}{1-2\delta} - (1-\mu)n\log p \qquad (4)\\
&\le \frac{h_2(2\delta) + \log e + \frac{n}{2}\log\frac{n}{2} - n\log t}{1-2\delta} - (1-\mu)n\log p \qquad (5)\\
&= \frac{h_2(2\delta) + \log e + n\log\frac{(n/2)^{1/2}}{\tau\sqrt{n}}}{1-2\delta} - (1-\mu)n\log p \\
&= \frac{h_2(2\delta) + \log e + n\log\frac{1}{\sqrt{2}\tau}}{1-2\delta} - (1-\mu)n\log p \\
&\le \frac{n\log\frac{1}{\tau}}{1-2\delta} - \frac{(1-2\delta)(1-\mu)n\log p}{1-2\delta} \qquad (6)\\
&\le \frac{n\log\frac{1}{\tau}}{1-2\delta} - \frac{\frac{\log(1/(\tau))}{(1-\mu)\log p}(1-\mu)n\log p}{1-2\delta} \qquad (7)\\
&\le \frac{n\log\frac{1}{\tau}}{1-2\delta} - \frac{n\log\frac{1}{\tau}}{(1-2\delta)} = 0
\end{aligned}$$

Where Equation 8 follows from Equation 7 as $n(1-\log\pi/2-\log e/2)+1/2\log n/2 \le 0$ for all even positive integers. Equation 9 follows by the fact that $n \ge 2(h_2(2\delta)+\log e)$ by assumption. Equation 10 follows by the assumption on δ. □

Putting these facts together allows us to conclude the theorem statement:

$$\begin{aligned}
&\tilde{\mathrm{H}}_\infty(W_Z|Z, \mathsf{SS}(W_Z)) = \tilde{\mathrm{H}}_\infty(\mathsf{M}|\mathsf{SS}(\mathsf{M}), \mathcal{A}, \mathsf{Hash}_{\mathcal{A}}(\mathsf{M}))\\
&\quad\le \log \frac{1}{\Pr[\mathsf{SS}(\mathsf{M}) \in \mathsf{GoodSketch}_{\mathsf{M}}]} + \\
&\qquad \tilde{\mathrm{H}}_\infty(\mathsf{M}|\text{ss s.t. ss} = \mathsf{SS}(\mathsf{M}) \text{ and ss} \in \mathsf{GoodSketch}_{\mathsf{M}}, \mathcal{A}, \mathsf{Hash}_{\mathcal{A}}(\mathsf{M}))\\
&\quad\le \log \frac{1}{\Pr[\mathsf{SS}(\mathsf{M}) \in \mathsf{GoodSketch}_{\mathsf{M}}]} + \\
&\qquad \tilde{H}_0(\mathsf{M}|\text{ss s.t. ss} = \mathsf{SS}(\mathsf{M}) \text{ and ss} \in \mathsf{GoodSketch}_{\mathsf{M}}, \mathcal{A}, \mathsf{Hash}_{\mathcal{A}}(\mathsf{M}))\\
&\quad< \log \frac{1}{\Pr[\mathsf{SS}(\mathsf{M}) \in \mathsf{GoodSketch}_{\mathsf{M}}]} + \max\left(1, 2 + \frac{h_2(2\delta) - \log|B_t|}{1-2\delta} - m\log p\right)\\
&\quad< \log 2 + \max\left(1, 2 + \frac{h_2(2\delta) - \log|B_t|}{1-2\delta} - m\log p\right)\\
&\quad\le \log 2 + \max\left(1, 2 + \frac{h_2(2\delta) - \log|B_t|}{1-2\delta} - (1-\mu)n\log p\right)\\
&\quad\le \log 2 + \max(1, 2)\\
&\quad\le 3
\end{aligned}$$

This completes the proof of Theorem 3. □

Acknowledgements This work was supported by a grant from Comcast Inc. The authors wish to Leonid Reyzin and Alexander Russell for their helpful feedback.

References

[ABC+18] Quentin Alamélou, Paul-Edmond Berthier, Chloé Cachet, Stéphane Cauchie, Benjamin Fuller, Philippe Gaborit, and Sailesh Simhadri. Pseudoentropic isometries: A new framework for fuzzy extractor reusability. In *AsiaCCS*, 2018.

[AC93] Rudolf Ahlswede and Imre Csiszár. Common randomness in information theory and cryptography - I: secret sharing. *IEEE Transactions on Information Theory*, 39(4):1121–1132, 1993.

[ACEK17] Daniel Apon, Chongwon Cho, Karim Eldefrawy, and Jonathan Katz. Efficient, reusable fuzzy extractors from LWE. In *International Conference on Cyber Security Cryptography and Machine Learning*, pages 1–18. Springer, 2017.

[BBR88] Charles H. Bennett, Gilles Brassard, and Jean-Marc Robert. Privacy amplification by public discussion. *SIAM Journal on Computing*, 17(2):210–229, 1988.

[BDHV07] Ileana Buhan, Jeroen Doumen, Pieter Hartel, and Raymond Veldhuis. Fuzzy extractors for continuous distributions. In *Proceedings of the 2nd ACM symposium on Information, computer and communications security*, pages 353–355. ACM, 2007.

[BS00] Sacha Brostoff and M.Angela Sasse. Are passfaces more usable than passwords?: A field trial investigation. *People and Computers*, pages 405–424, 2000.

[CD09] Giacomo Como and Munther Dahleh. Lower bounds on the estimation error in problems of distributed computation. In *Information Theory and Applications Workshop, 2009*, pages 70–76. IEEE, 2009.

[CFP+16] Ran Canetti, Benjamin Fuller, Omer Paneth, Leonid Reyzin, and Adam Smith. Reusable fuzzy extractors for low-entropy distributions. In *Advances in Cryptology–EUROCRYPT*, pages 117–146. Springer, 2016.

[CK78] Imre Csiszár and János Körner. Broadcast channels with confidential messages. *IEEE Transactions on Information Theory*, 24(3):339–348, 1978.

[CS13] John Horton Conway and Neil James Alexander Sloane. *Sphere packings, lattices and groups*, volume 290. Springer Science & Business Media, 2013.

[CW79] Larry Carter and Mark N. Wegman. Universal classes of hash functions. *J. Comput. Syst. Sci.*, 18(2):143–154, 1979.

[Dau04] John Daugman. How iris recognition works. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14(1):21 – 30, January 2004.

[DORS08] Yevgeniy Dodis, Rafail Ostrovsky, Leonid Reyzin, and Adam Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. *SIAM Journal on Computing*, 38(1):97–139, 2008.

[EHMS00] Carl Ellison, Chris Hall, Randy Milbert, and Bruce Schneier. Protecting secret keys with personal entropy. *Future Generation Computer Systems*, 16(4):311–318, 2000.

[ELZ05] Uri Erez, Simon Litsyn, and Ram Zamir. Lattices which are good for (almost) everything. *IEEE Transactions on Information Theory*, 51(10):3401–3416, 2005.

[Fan61] R.M. Fano. *Transmission of Information: A Statistical Theory of Communications*. MIT Press Classics. M.I.T. Press, 1961.

[FMR13] Benjamin Fuller, Xianrui Meng, and Leonid Reyzin. Computational fuzzy extractors. In *Advances in Cryptology-ASIACRYPT*, pages 174–193. Springer, 2013.

[FRS16] Benjamin Fuller, Leonid Reyzin, and Adam Smith. When are fuzzy extractors possible? In *Advances in Cryptology–ASIACRYPT*, pages 277–306. Springer, 2016.

[GCVDD02] Blaise Gassend, Dwaine Clarke, Marten Van Dijk, and Srinivas Devadas. Silicon physical random functions. In *Proceedings of the 9th ACM conference on Computer and communications security*, pages 148–160. ACM, 2002.

[Gur10] Venkatesan Guruswami. Introduction to coding theory - lecture 2: Gilbert-Varshamov bound. University Lecture, 2010.

[HRvD+17] Charles Herder, Ling Ren, Marten van Dijk, Meng-Day Mandel Yu, and Srinivas Devadas. Trapdoor computational fuzzy extractors and stateless cryptographically-secure physical unclonable functions. *IEEE Transactions on Dependable and Secure Computing*, 14(1):65–82, 2017.

[HTW14] Masahito Hayashi, Himanshu Tyagi, and Shun Watanabe. Secret key agreement: General capacity and second-order asymptotics. In *2014 IEEE International Symposium on Information Theory*, pages 1136–1140. IEEE, 2014.

[LSM06] Qiming Li, Yagiz Sutcu, and Nasir Memon. Secure sketch for biometric templates. In *Advances in Cryptology – ASIACRYPT*, pages 99–113. Springer, 2006.

[Mau93] Ueli M. Maurer. Secret key agreement by public discussion from common information. *IEEE Transactions on Information Theory*, 39(3):733–742, 1993.

[MG09] Rene Mayrhofer and Hans Gellersen. Shake well before use: Intuitive and secure pairing of mobile devices. *IEEE Transactions on Mobile Computing*, 8(6):792–806, 2009.

[PRTG02] Ravikanth Pappu, Ben Recht, Jason Taylor, and Neil Gershenfeld. Physical one-way functions. *Science*, 297(5589):2026–2030, 2002.

[RW05] Renato Renner and Stefan Wolf. Simple and tight bounds for information reconciliation and privacy amplification. In Bimal K. Roy, editor, *Advances in Cryptology - ASIACRYPT*, volume 3788 of *LNCS*, pages 199–216. Springer, 2005.

[SD07] G. Edward Suh and Srinivas Devadas. Physical unclonable functions for device authentication and secret key generation. In *Proceedings of the 44th annual Design Automation Conference*, pages 9–14. ACM, 2007.

[TSS+06] Pim Tuyls, Geert Jan Schrijen, Boris Skoric, Jan van Geloven, Nynke Verhaegh, and Rob Wolters. Read-proof hardware from protective coatings. In Louis Goubin and Mitsuru Matsui, editors, *Cryptographic Hardware and Embedded Systems - CHES 2006*, volume 4249 of *Lecture Notes in Computer Science*, pages 369–383. Springer, 2006.

[TW15] Himanshu Tyagi and Shun Watanabe. Converses for secret key agreement and secure computing. *IEEE Transactions on Information Theory*, 61(9):4809–4827, 2015.

[VTO+10] Evgeny A Verbitskiy, Pim Tuyls, Chibuzo Obi, Berry Schoenmakers, and Boris Skoric. Key extraction from general nondiscrete signals. *IEEE Transactions on Information Forensics and Security*, 5(2):269–279, 2010.

[WCD+17] Joanne Woodage, Rahul Chatterjee, Yevgeniy Dodis, Ari Juels, and Thomas Ristenpart. A new distribution-sensitive secure sketch and popularity-proportional hashing. In *Advances in Cryptology – CRYPTO*, pages 682–710. Springer, 2017.

[WLH18] Yunhua Wen, Shengli Liu, and Shuai Han. Reusable fuzzy extractor from the decisional Diffie–Hellman assumption. *Designs, Codes and Cryptography*, pages 1–18, 2018.

[Wyn75] Aaron D Wyner. The wire-tap channel. *Bell System Technical Journal, The*, 54(8):1355–1387, 1975.

A Proofs

A.1 Proof of Lemma 4

Proof. Fix some pub and some $\mathsf{key} \in \mathsf{GoodKey}_{\mathsf{pub}}$ and consider $Q_{p,\mathsf{key}}$. Recall that

$$\kappa \geq 1 + \log(\gamma) + n\left(\log p^{\mu} - \log(\frac{6\sqrt{e}+\sqrt{\pi}}{6\sqrt{e}})\right).$$

By substitution one has,

$$\mathsf{Vol}(Q_{\mathsf{pub},\mathsf{key}}) \leq 2^{-\kappa+1} \leq \frac{1}{(\gamma)\cdot 2^{n\left(\log p^{\mu} - \log(\frac{6\sqrt{e}+\sqrt{\pi}}{6\sqrt{e}})\right)}} = \frac{2^{n\log(\frac{6\sqrt{e}+\sqrt{\pi}}{6\sqrt{e}})}}{(\gamma)\cdot p^{\mu n}.}.$$

Define α to be the radius of a ball with that volume. Using Equation 1 we have that

$$\frac{\pi^{n/2}\alpha^n}{(n/2)!} \leq \frac{2^{n\log(\frac{6\sqrt{e}+\sqrt{\pi}}{6\sqrt{e}})}}{(\gamma)p^{\mu n}}$$

Rearranging terms gives a bound on α:

$$\alpha \leq \frac{\sqrt[n]{(n/2)!}}{\sqrt{\pi}} \cdot \frac{1}{\sqrt[n]{\gamma}\cdot p^{\mu}} \cdot \left(\frac{6\sqrt{e}+\sqrt{\pi}}{6\sqrt{e}}\right)$$

The isoperimetric inequality says that the interior of this part is maximized in the setting when $Q_{\mathsf{pub},\mathsf{key}}$ is a ball. Thus, for all $\mathsf{key} \in \mathsf{GoodKey}_{\mathsf{pub}}$

$$\mathsf{Vol}(\mathsf{Inter}(Q_{p,\mathsf{key}})) \leq \mathsf{Vol}(B_{\alpha-t}).$$

We know the quantity $\alpha - t$ is bounded by:

$$\begin{aligned}
\alpha - t &\leq \frac{\sqrt[n]{(n/2)!}}{\sqrt{\pi}} \cdot \frac{1}{\sqrt[n]{\gamma}\cdot p^{\mu}} \cdot \left(\frac{6\sqrt{e}+\sqrt{\pi}}{6\sqrt{e}}\right) - \frac{1}{6\sqrt{2e}p^{\mu}}\sqrt{n} \\
&\leq \frac{1}{p^{\mu}}\left(\frac{\sqrt[n]{(n/2)!}}{\sqrt[n]{\gamma}\sqrt{\pi}} \cdot (\frac{6\sqrt{e}+\sqrt{\pi}}{6\sqrt{e}}) - \frac{1}{6\sqrt{e}}\sqrt{\frac{n}{2}}\right) &(8)\\
&\leq \frac{1}{p^{\mu}}\left(\frac{(\frac{n}{2})^{(\frac{n+1}{2})}\cdot e^{(\frac{1}{n}-\frac{1}{2})}}{\sqrt[n]{\gamma}} \cdot \left(\frac{1}{\sqrt{\pi}} + \frac{1}{6\sqrt{e}}\right) - \frac{1}{6\sqrt{e}}\sqrt{\frac{n}{2}}\right) &(9)\\
&\leq \frac{\left(\frac{n}{2}\right)^{(\frac{n+1}{2})} e^{\frac{1}{n}-\frac{1}{2}}}{\sqrt{\pi}p^{\mu}\gamma^{1/n}} &(10)
\end{aligned}$$

Here equation 3 follows from 2 by use of Stirling's upper bound that $(n/2)! \leq e(n/2)^{n/2+1/2}\cdot e^{-n/2}$. As $\gamma \geq 1$ and $n \in \mathbb{Z}^{+}$, it is always true that

$$\frac{e^2 n}{2} \leq (\gamma)^2 e^n.$$

Thus, equation 4 follows by noting that the quantity

$$\frac{\left(\frac{n}{2}\right)^{\frac{1}{2n}} e^{\frac{1}{2}-\frac{1}{2n}}}{\gamma^{1/n}} \leq 1$$

This gives us the desired bound on the volume of this interior:

$$\begin{aligned}
\mathsf{Vol}(\mathsf{Inter}(Q_{p,\mathsf{key}})) &\leq \frac{\pi^{n/2}\left(\frac{\left(\frac{n}{2}\right)^{\left(\frac{n+1}{2}\right)} e^{\frac{1}{n}-\frac{1}{2}}}{\sqrt{\pi}p^{\mu}\gamma^{1/n}}\right)^{n}}{(n/2)!} \\
&\leq \frac{\left(\frac{n}{2}\right)^{\frac{n+1}{2}} p^{-\mu n} \cdot e \cdot e^{\frac{-n}{2}} \gamma^{-1}}{\sqrt{2\pi}\left(\frac{n}{2}\right)^{\frac{n+1}{2}} e^{\frac{-n}{2}}} \\
&= \frac{e}{\sqrt{2\pi} \cdot \gamma \cdot p^{\mu n}}.
\end{aligned}$$

Here we use Stirling's lower bound that $(n/2)! \geq \sqrt{2\pi}(n/2)^{n/2+1/2}e^{-n/2}$. This completes the proof of Lemma 4. □