University of Warsaw
Faculty of Mathematics, Informatics and Mechanics

Maciej Skórski

Pseudoentropy

PhD dissertation

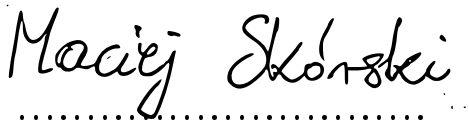
Supervisor

dr hab. Stefan Dziembowski
dr Krzysztof Pietrzak (co-advisor)

Institute of Computer Science
University of Warsaw

May 2019

Author's declaration:
aware of legal responsibility I hereby declare that I have written this dissertation myself and all the contents of the dissertation have been obtained by legal means.



May 14 2019
date

.....
Maciej Skórski

Supervisor's declaration:
the dissertation is ready to be reviewed

May 14 2019
date

.....
dr hab. Stefan Dziembowski
dr Krzysztof Pietrzak (co-advisor)

ACKNOWLEDGMENTS I would like to thank my advisors: Stefan Dziembowski and Krzysztof Pietrzak, for introducing me into the research world and supporting my scientific activities.
I would like to thank several researchers with whom I discussed and cooperated. Particularly to Leo Reyzin for encouraging me when the work was at the very early stage and I was frustrated with the lack of visible progress.
Finally, I would like to thank to my wife. She has been with me all these years, participating in successes and helping in recovery from failures.

Contents

1 Introduction **5**
- 1.1 About the Thesis 5
- 1.2 Pseudoentropy 5
- 1.3 Author's Contribution 8
  - 1.3.1 Equivalence of HILL and Unpredictability Pseudoentropy in High-Entropy Regimes 8
  - 1.3.2 Lower Bounds for Pseudoentropy Chain Rules and Transformations 9
  - 1.3.3 Simulating Auxiliary Information 10
  - 1.3.4 Best Generic Attacks Against Pseudoentropy 11
  - 1.3.5 Geometrical Characterizations of Pseudoentropy 12

2 Equivalence of HILL and Unpredictability Pseudoentropy in High-Entropy Regimes **20**

3 Lower Bounds for Pseudoentropy Chain Rules and Transformations **51**

4 Simulating Auxiliary Information **75**

5 Best Generic Attacks Against Pseudoentropy **98**

6 Geometric Characterization **114**

Chapter 1

Introduction

1.1 About the Thesis

This report presents some results of the author's PhD project, which aimed at systematically studying the theory and applications of *pseudoentropy* - a generalization of the well-established cryptographic notion of pseudorandomness. When the author started his PhD, all technical knowledge about pseudoentropy was essentially summarized in two papers [FR12, BSW03]. Since then our knowledge has grown, in particular with regards to mathematical techniques that can be successfully applied to study this concept, and about applications in cryptography and outside.

This thesis is a compilation of different works (slightly revised conference papers), sharing the novel approach of *bridging pseudoentropy and convex optimization*. The strong conceptual message is that *convex analysis is useful and powerful* as a tool to study pseudoentropy and related problems. This is particularly interesting as these techniques are rather rarely used in crypto, and so far have not been used in this context in a systematic way.

The document starts with the general introduction followed by chapters discussing technical results, each including copies of relevant publications.

1.2 Pseudoentropy

Classical entropy notions and their applications The notion of entropy, introduced by Shannon [Sha01] and extended by Renyi [Ré61], is not only the fundamental concept in information-theory, but has also found many

applications to other research areas. Some applications include statistical learning, image registration, anomaly detection and analysis, reconstructing DNA sequences, password guessability, testing quality of pseudorandom generators or quantifying randomness extraction.

In security applications, entropy is typically used to argue that certain probability distributions (e.g. the distribution of encrypted messages or the randomness source used for key derivation) look random or sufficiently random to external observers (potential attackers).

Computational generalizations The classical notions of entropy (as developed in information theory) are not suitable for applications in modern cryptography, which concerns *computational security*. In accordance to practical requirements, cryptographic tools are designed under the assumption that adversaries have limited computational power. This motivates the need for quantifying the amount of randomness seen by attackers with such limited computational capacity. Such a notion was studied for the first time in works constructing pseudorandom generators (PRGs) from one-way functions (OWSs) [ILL89, HILL99], and was called "pseudoentropy". Currently there are many notions in this spirit, all referred to as *pseudoentropy* or *computational entropy* and followed by specific names to avoid ambiguity.

Applications of computational entropy Loosely speaking, pseuedoentropy lives in the intersection of computational complexity, information-theory and cryptography. It appears as a useful tool in many important results in cryptography and complexity theory, such as

- Cryptography: deterministic encryption [FOR12], showing resilience of cryptographic primitives against leakage [DP08, Pie09], black box separations [GW11], improving conditional constructions of PRGs [VZ12], memory delegation [CKLR11], and key derivation [SGP15].
- Computational complexity: proving computationally efficient versions of the celebrated Green-Tao-Ziegler Dense Model Theorem [RTTV08, Zha11], unifying proofs of hardcore lemmas [Sko15b].

Defining pseudoentropy From a technical point of view, every pseudoentropy notion is defined by quantifying what it means that, for observers

with limited computational resources, it looks as if it has some information-theoretic entropy (which is emphasized by the prefix "pseudo"). The following approaches are most common (see [Rey11] for a survey discussing definitions and fundamental properties):

(a) Indistinguishably - the given distribution is indistinguishable by efficient tests (in a well defined statistical sense) from a distribution with certain entropy. The definition was introduced in [HILL99] and [BSW03]

(b) Unpredictability - a sample from the given distribution cannot be efficiently guessed better than with certain small probability (this mimics the behavior of the information-theoretic entropy notion called min-entropy). The computational variant was first used in [HLR07].

(c) Incompressibility - the given distribution cannot be efficiently "compressed" and "decompressed" back to a distribution of entropy bigger than a certain amount. It was introduced by Yao [Yao82] and further studied in [GS91].

It should be noted that this list is not exhaustive. For example Haitner et al. [HRVW09] study the notion of *inaccessible entropy*, taylored to applications hiding commitments schemes; the purpose of this notion is to argue that in some cases the computationally accesible entropy is *lower* than its real entropy; this paradigm is somewhat in contrast to the intuitive understanding of the word *pseudo*entropy [1].

Concrete pseudoentropy definitions will be given, when necessary, in the chapters discussing technical results.

Amount and quality issues As discussed above, pseudoentropy notions are parametrized not only by the *amount* of entropy, but also by *quality* which quantifies attacker's resources (the larger resources, the stronger security meaning). These resources are typically the running time and the space of an algorithm, or the circuit size (in the non-uniform model of computation).

The reader, especially the one with information-theoretic background, should keep in mind that dealing with pseudoentropy is harder and more tricky compared to classical information-theoretic measures, because one

[1] Citing the authors of [HRVW09]: "Thus, *in contrast* to pseudoentropy, accessible entropy is useful for expressing the idea that the "computational entropy" in a distribution is smaller than its real entropy"

needs to take into account tradeoffs in quality parameters, that arise in reduction proofs. In particular, not all facts that are "natural" for information-theoretic entropy are true in the computational setting. Perhaps the most surprising example is the so called chain rule for conditional min-entropy, which quantifies the entropy decrease due to extra knowledge. It turns out that it fails, even in very relaxed forms, for the popular computational analogue of min-entropy [KPW13].

It may also be that pseudoentropy notions fail to satify obvious "computational" properties. For example for the notion of *metric min-entropy*, which is build on the computational distance, it matters whether attackers are randomized or not [FR12]. As opposed to that for the standard computational distance a coin-fixing argument shows it doesn't matter.

Summing up, pseudoentropy notions are *hybrid* objects that combine both information-theoretic and computational aspects, and therefore may fail to behave as one intuitively expects. This is an inherit issue where only more work on unification can bring more clarification.

The reference notion Another point, briefly discussed in [YL13], is that the choice of the reference information-theoretic entropy is not obvious. While in [HILL99] the indistingusihability-based pseudoentropy was defined based on Shannon entropy, it was (somewhat silently) changed to min-entropy (more relevant to broad cryptography) in [BSW03] and following works. However for some cases min-entropy is an overly conservative measure. The work [YL13] adapts the pseuodentropy notion from [HILL99] to the case of Renyi entropy, which is enough for certain problems in key derivation. One should also mention a recent work [ACHV19] which uses the information-theoretic KL divergence to bridge some pseudoentropy notions.

1.3 Author's Contribution

1.3.1 Equivalence of HILL and Unpredictability Pseudoentropy in High-Entropy Regimes

High Unpredictability and Key Derivation At CRYPTO'14 Dodis at. all [DPW14] showed (continuing the line of research originated in [BDK+11] and [DY13]) that one can directly use keys with small entropy deficiency for a broad class of cryptographic applications, which allows for significant savings

in entropy loss for constructions of *key derivation functions*. Roughly, the sufficient condition for the key to be "good enough" is that unpredictability, measured by the information-theoretic min-entropy, is almost as in the uniform distribution (up to a gap logarithmic in the key length). However no analogue for computational unpredictability was known.

Contribution Motivated by applications in key derivation, the author together with Krzysztof Pietrzak and Alexander Golovnev showed [SGP15] that (in the context of key derivation) *unpredictability pseudoentropy* can be used in place of min-entropy, provided that *entropy deficiency is small*; more precisely, the computational unpredictability of the n bit key, conditioned on any auxiliary information (potentially available to attackers), needs to be at least $n - O(\log n)$. Technically, the result shows computational closeness to a distribution with the same amount of min-entropy.

This result is somewhat unexpected because in general (gaps bigger than logarithmic) unpredictability pseudoentropy is too weak to be a computational substitute for min-entropy. For example, the existence of exponentially hard one-way permutations implies that the input distribution over n bits has unpredictability entropy of $\Omega(n)$ (conditioned on the known output), but it can be easily distinguished from every distribution with few bits of min-entropy (conditioned on the known output).

The result of [SGP15] hence establishes the *equivalence in small-deficiency regimes* for unpredictability and indistinguishability-based computational entropy definitions, and nicely strengthens the observation of Vadhan and Zheng from STOC'12 [VZ12], who proved the equivalence only for logarithmically small domains).

The result is presented in chapter *Equivalence of HILL and Unpredictability Pseudoentropy in High-Entropy Regimes* which includes an article *Condensed Unpredictability*, presented at *ICALP 2015*.

1.3.2 Lower Bounds for Pseudoentropy Chain Rules and Transformations

Chain Rules and Transformations There are two technical tools extremely useful for applications of pseudoentropy in leakage-resilient cryptography. The first one is called the *chain rule*, and concerns by how much pseudoentropy goes down in the presence of auxiliary information (leakage);

concrete bounds are needed for security proofs, where pseudoentropy is used to measure security of a secret state. The second tool are *transformations* between stronger and weaker variants of the indistinguishability-based notion of pseudoentropy (originally due to Barak et. all [BSW03]); the transformations show that a weaker variant (typically easier to work with) implies the stronger one, up to some (acceptable) loss in quality parameters.

Unfortunately best known bounds for both problems incur a *heavy loss in quality* of pseudoentropy (even for small leakages!). Namely, when applying any of these bounds we get only security against much weaker attackers than before. The running time/circuit size goes down by a factor of $\mathrm{poly}(\epsilon^{-1})$, where ϵ is a negligible quantity. Since in concrete applications we require (nowadays) $\epsilon \approx 2^{-100}$ for meaningful security, this loss is huge from a practical point of view.

Contribution It turns out that, unfortunately, the existing bounds are basically optimal (the result establishes first lower bounds for pseudoentropy).

In particular, to prove "resilience" of cryptography constructions using pseudoentropy techniques (for example the first leakage-resilient stream cipher due to Dziembowski and Pietrzak that appeared at FOCS'08 [DP08] and later constructions [Pie09, JP14]) one has to lose a factor $\mathrm{poly}(\epsilon^{-1})$ in the security, comparing to the security in the no-leakage setting. The lower bounds also imply a necessary loss exponential in the key length. Further applications include lower bounds (impossibilities) for the Dense Model Theorem in certain parameter regimes, and the problem of simulating auxiliary information (used in leakage-resilient cryptography and zero-knowledge theory).

Aside from these applications, the lower bounds exhibit the following interesting phenomena: for some notions of pseudoentropy *randomized adversaries might be much more powerful*, even in the non-uniform setting (unlike pseudorandomness, the standard coin-fixing argument does not apply).

The results are discussed in chapter *Lower Bounds for Pseudoentropy Chain Rules and Transformations*, which includes an article (a joint work with Krzysztof Pietrzak) presented at *Theory of Cryptography 2016-B*.

1.3.3 Simulating Auxiliary Information

Simulating leakage from secret states Some security proofs in leakage-resilient cryptography are much easier if we model the leakage as an *efficient*

explicit function of a secret state. The natural question is, whether or not such an assumption is a substantial limitation. This problem was studied in a paper presented at TCC'14 by Pietrzak and Jetchev [JP14], where this restriction was shown not to be significant for *short leakages*. Also some quantitative improvements to leakage-resilient constructions, pseudoentropy chain rules and zero-knowledge protocols where shown. Later a flaw in the paper was identified [Sko15c] which left the result valid only with much weaker bounds than claimed (in particular it doesn't offer better security parameter for the EUROCRYPT'09 stream cipher [Pie09]).

Contribution A generic "simulator" is obtained, which simulates every m-bit leakage correlated to some given distribution (secret state) up to to a statistical error ϵ in time $t = 2^{O(m)}\epsilon^{-2}$, for any ϵ. This result essentially fixes the flaw in [JP14] and for typical cryptographic settings of parameters (negligible ϵ, sub-logarithmic m) is quantitatively better than a more generic simulator due to Vadhan and Zheng [VZ13]. The interesting proof technique is a *descent algorithm* (inspired by subgradient descent techniques from convex analysis) which yields an iterative construction of the simulator. Among other applications, the result gives a *constructive proof* of chain rules for pseudoentropy (indistinguishability-based definitions).

The details are discussed in the chapter *Simulating Auxiliary Information*, which includes the article *Simulating Auxiliary Information, Revisited* presented at *Theory of Cryptography 2016-B* [Skó16] (this paper won the Best Student Paper award). Recently the techniques developed in this paper were used to simplify proofs and slightly improve bounds for Szemeredi Regularity Lemmas (the paper discussing this application is to appear in *Theory and Applications of Models of Computation 2017*, and is not included in this thesis).

It should be noted that the state of art and matching lower bounds were established in the recent work [CCL18], where the optimal exponent constant, here hidden under $O(m)$, was determined. Also a variant of this problem has been studied in the context of interactive protocols for example [CLP15] and recently [HNO+18].

1.3.4 Best Generic Attacks Against Pseudoentropy

Bounds on quality parameters Concrete, good parameters quantifying the quality of pseudoentropy (e.g. adversarial resources and the hardness

of the computational task such as distinguishing probability) are based on computational assumptions. If pseudoentropy of a distribution is related to a well-understood "computational" problem, such as breaking security of a PRG or Diffie-Hellman protocols, one can argue reasonably good quality by a reduction. However at the time of carrying out this research, nothing was known about *generic attacks* against pseudoentropy.

Contribution It turns out that pseudoentropy exhibits a *threshold behavior against adversarial resources*. If the running time of an adversary is much bigger than 2^k, then pseudoentropy is less than k bits. On the other hand, if the time is much less than 2^k, there exist distribution with pseudoentropy more than k (constructed by a non-explicit method, with no computational assumptions). This result appears in *Theory and Applications of Models of Computation 2017* [Skó17b].

In this thesis this result is however updated by a more recent article, where optimal quantitative bounds (up to constants) are obtained. If t denotes the attacker's time then best generic attacks achieve the advantage $\epsilon \approx \sqrt{2^k/t}$ (under this condition the amount is not more than the information-theoretic entropy of roughly k). This result nicely extends the well known *best attacks against pseudorandom generators* shown by De et. all at CRYPTO'10 (see [DTT10]).

Details are discussed in the chapter *Best Generic Attacks Against Pseudoentropy*, which includes the paper *Best Generic Attacks Against Pseudoentropy* written together with Krzysztof Pietrzak, presented at ICALP'17 [PS17].

1.3.5 Geometrical Characterizations of Pseudoentropy

Convex analysis approach to indistinguishability The concept of *computational indistinguishability*, used to define popular pseudoentropy notions, quantifies how close are two probability distributions (more broadly: two sets of distributions) under a class of computational tests (called distinguishers). It turns out that it can be characterized as a problem of *efficiently computing a separating hyperplane* between two certain convex sets (containing probability measures).

Contribution While the fact above may be known in folklore to some researchers, we are not aware of any work using it quantitatively or even

mentioning it. We show that this idea, originated from convex analysis, has powerful applications

(a) Can be used as a key tool in proofs connecting indistinguishability and unpredictability [Sko15a, SGP15] (it also simplifies technical arguments in [VZ12]).

(b) It yields a short proof of the celebrated Dense Model Theorem [Sko15a], achieving optimal constants (originally due to Zhang [Zha11])

(c) Has further applications to key derivation [Sko15a, Sko17a]. Basically, the geometric characterization is used to compute the "shape" of the "weakest" distribution over secret keys, which maximizes the breaking probability.

Details are discussed in the chapter *Geometric Characterizations*, which includes the article [Sko15a], presented at *Information Theoretic Security 2015*.

Bibliography

[ACHV19] Rohit Agrawal, Yi-Hsiu Chen, Thibaut Horel, and Salil P. Vadhan, *Unifying computational entropies via kullback-leibler divergence*, IACR Cryptology ePrint Archive **2019** (2019), 264.

[BDK+11] Boaz Barak, Yevgeniy Dodis, Hugo Krawczyk, Olivier Pereira, Krzysztof Pietrzak, François-Xavier Standaert, and Yu Yu, *Leftover hash lemma, revisited*, Advances in Cryptology - CRYPTO 2011 - 31st Annual Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2011. Proceedings, 2011, pp. 1–20.

[BSW03] Boaz Barak, Ronen Shaltiel, and Avi Wigderson, *Computational analogues of entropy*, Approximation, Randomization, and Combinatorial Optimization: Algorithms and Techniques, 6th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, APPROX 2003 and 7th International Workshop on Randomization and Approximation Techniques in Computer Science, RANDOM 2003, Princeton, NJ, USA, August 24-26, 2003, Proceedings, 2003, pp. 200–215.

[CCL18] Yi-Hsiu Chen, Kai-Min Chung, and Jyun-Jie Liao, *On the complexity of simulating auxiliary input*, Advances in Cryptology - EUROCRYPT 2018 - 37th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tel Aviv, Israel, April 29 - May 3, 2018 Proceedings, Part III, 2018, pp. 371–390.

[CKLR11] Kai-Min Chung, Yael Tauman Kalai, Feng-Hao Liu, and Ran Raz, *Memory delegation*, Advances in Cryptology - CRYPTO 2011 - 31st Annual Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2011. Proceedings, 2011, pp. 151–168.

[CLP15] Kai-Min Chung, Edward Lui, and Rafael Pass, *From weak to strong zero-knowledge and applications*, Theory of Cryptography - 12th Theory of Cryptography Conference, TCC 2015, Warsaw, Poland, March 23-25, 2015, Proceedings, Part I, 2015, pp. 66–92.

[DP08] Stefan Dziembowski and Krzysztof Pietrzak, *Leakage-resilient cryptography*, 49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA, 2008, pp. 293–302.

[DPW14] Yevgeniy Dodis, Krzysztof Pietrzak, and Daniel Wichs, *Key derivation without entropy waste*, Advances in Cryptology - EUROCRYPT 2014 - 33rd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Copenhagen, Denmark, May 11-15, 2014. Proceedings, 2014, pp. 93–110.

[DTT10] Anindya De, Luca Trevisan, and Madhur Tulsiani, *Time space tradeoffs for attacks against one-way functions and prgs*, Advances in Cryptology - CRYPTO 2010, 30th Annual Cryptology Conference, Santa Barbara, CA, USA, August 15-19, 2010. Proceedings, 2010, pp. 649–665.

[DY13] Yevgeniy Dodis and Yu Yu, *Overcoming weak expectations*, TCC, 2013, pp. 1–22.

[FOR12] Benjamin Fuller, Adam O'Neill, and Leonid Reyzin, *A unified approach to deterministic encryption: New constructions and a connection to computational entropy*, Theory of Cryptography - 9th Theory of Cryptography Conference, TCC 2012, Taormina, Sicily, Italy, March 19-21, 2012. Proceedings, 2012, pp. 582–599.

[FR12] Benjamin Fuller and Leonid Reyzin, *Computational entropy and information leakage*, IACR Cryptology ePrint Archive **2012** (2012), 466.

[GS91] Andrew V. Goldberg and Michael Sipser, *Compression and ranking*, SIAM J. Comput. **20** (1991), no. 3, 524–536.

[GW11] Craig Gentry and Daniel Wichs, *Separating succinct non-interactive arguments from all falsifiable assumptions*, Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6-8 June 2011, 2011, pp. 99–108.

[HILL99] Johan Håstad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby, *A pseudorandom generator from any one-way function*, SIAM J. Comput. **28** (1999), no. 4, 1364–1396.

[HLR07] Chun-Yuan Hsiao, Chi-Jen Lu, and Leonid Reyzin, *Conditional computational entropy, or toward separating pseudoentropy from compressibility*, Advances in Cryptology - EUROCRYPT 2007, 26th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Barcelona, Spain, May 20-24, 2007, Proceedings, 2007, pp. 169–186.

[HNO+18] Iftach Haitner, Kobbi Nissim, Eran Omri, Ronen Shaltiel, and Jad Silbak, *Computational two-party correlation: A dichotomy for key-agreement protocols*, 59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018, 2018, pp. 136–147.

[HRVW09] Iftach Haitner, Omer Reingold, Salil P. Vadhan, and Hoeteck Wee, *Inaccessible entropy*, Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009, 2009, pp. 611–620.

[ILL89] R. Impagliazzo, L. A. Levin, and M. Luby, *Pseudo-random generation from one-way functions*, Proceedings of the Twenty-first Annual ACM Symposium on Theory of Computing (New York, NY, USA), STOC '89, ACM, 1989, pp. 12–24.

[JP14] Dimitar Jetchev and Krzysztof Pietrzak, *How to fake auxiliary input*, Theory of Cryptography - 11th Theory of Cryptography Conference, TCC 2014, San Diego, CA, USA, February 24-26, 2014. Proceedings, 2014, pp. 566–590.

[KPW13] Stephan Krenn, Krzysztof Pietrzak, and Akshay Wadia, *A counterexample to the chain rule for conditional HILL entropy - and what deniable encryption has to do with it*, TCC, 2013, pp. 23–39.

[Pie09] Krzysztof Pietrzak, *A leakage-resilient mode of operation*, Advances in Cryptology - EUROCRYPT 2009, 28th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Cologne, Germany, April 26-30, 2009. Proceedings, 2009, pp. 462–482.

[PS17] Krzysztof Pietrzak and Maciej Skorski, *Non-uniform attacks against pseudoentropy*, 44th International Colloquium on Automata, Languages, and Programming, ICALP 2017, July 10-14, 2017, Warsaw, Poland, 2017, pp. 39:1–39:13.

[Ré61] Alfréd Rényi, *On measures of entropy and information*, Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics (Berkeley, Calif.), University of California Press, 1961, pp. 547–561.

[Rey11] Leonid Reyzin, *Some notions of entropy for cryptography - (invited talk)*, Information Theoretic Security - 5th International Conference, ICITS 2011, Amsterdam, The Netherlands, May 21-24, 2011. Proceedings, 2011, pp. 138–142.

[RTTV08] Omer Reingold, Luca Trevisan, Madhur Tulsiani, and Salil P. Vadhan, *Dense subsets of pseudorandom sets*, 49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA, 2008, pp. 76–85.

[SGP15] Maciej Skorski, Alexander Golovnev, and Krzysztof Pietrzak, *Condensed unpredictability*, Automata, Languages, and Programming - 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part I, 2015, pp. 1046–1057.

[Sha01] C. E. Shannon, *A mathematical theory of communication*, SIGMOBILE Mob. Comput. Commun. Rev. **5** (2001), no. 1, 3–55.

[Sko15a] Maciej Skorski, *Metric pseudoentropy: Characterizations, transformations and applications*, Information Theoretic Security - 8th International Conference, ICITS 2015, Lugano, Switzerland, May 2-5, 2015. Proceedings, 2015, pp. 105–122.

[Sko15b] ———, *Nonuniform indistinguishability and unpredictability hardcore lemmas: New proofs and applications to pseudoentropy*, Information Theoretic Security - 8th International Conference, ICITS 2015, Lugano, Switzerland, May 2-5, 2015. Proceedings, 2015, pp. 123–140.

[Sko15c] ———, *On provable security of wprf-based leakage-resilient stream ciphers*, Provable Security - 9th International Conference, ProvSec 2015, Kanazawa, Japan, November 24-26, 2015, Proceedings, 2015, pp. 391–411.

[Skó16] Maciej Skórski, *Simulating auxiliary inputs, revisited*, Theory of Cryptography - 14th International Conference, TCC 2016-B, Beijing, China, October 31 - November 3, 2016, Proceedings, Part I, 2016, pp. 159–179.

[Sko17a] Maciej Skorski, *Lower bounds on key derivation for square-friendly applications*, 34th Symposium on Theoretical Aspects of Computer Science, STACS 2017, March 8-11, 2017, Hannover, Germany, 2017, pp. 57:1–57:12.

[Skó17b] Maciej Skórski, *On the complexity of breaking pseudoentropy*, Theory and Applications of Models of Computation - 14th Annual Conference, TAMC 2017, Bern, Switzerland, April 20-22, 2017, Proceedings, 2017, pp. 600–613.

[VZ12] Salil P. Vadhan and Colin Jia Zheng, *Characterizing pseudoentropy and simplifying pseudorandom generator constructions*, Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012, 2012, pp. 817–836.

[VZ13] ———, *A uniform min-max theorem with applications in cryptography*, Advances in Cryptology - CRYPTO 2013 - 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2013. Proceedings, Part I, 2013, pp. 93–110.

[Yao82] Andrew Chi-Chih Yao, *Theory and applications of trapdoor functions (extended abstract)*, 23rd Annual Symposium on Foundations of Computer Science, Chicago, Illinois, USA, 3-5 November 1982, 1982, pp. 80–91.

[YL13] Yanqing Yao and Zhoujun Li, *Overcoming weak expectations via the rényi entropy and the expanded computational entropy*, Information Theoretic Security - 7th International Conference, IC-ITS 2013, Singapore, November 28-30, 2013, Proceedings, 2013, pp. 162–178.

[Zha11] Jiapeng Zhang, *On the query complexity for showing dense model*, Electronic Colloquium on Computational Complexity (ECCC) **18** (2011), 38.

Chapter 2

Equivalence of HILL and Unpredictability Pseudoentropy in High-Entropy Regimes

Condensed Unpredictability

Maciej Skórski[1⋆], Alexander Golovnev[2], and Krzysztof Pietrzak[3⋆⋆]

[1] University of Warsaw maciej.skorski@gmail.com
[2] New York University alexgolovnev@gmail.com
[3] IST Austria pietrzak@ist.ac.at

Abstract. We consider the task of deriving a key with high HILL entropy (i.e., being computationally indistinguishable from a key with high min-entropy) from an unpredictable source.
Previous to this work, the only known way to transform unpredictability into a key that was ϵ indistinguishable from having min-entropy was via pseudorandomness, for example by Goldreich-Levin (GL) hardcore bits. This approach has the inherent limitation that from a source with k bits of unpredictability entropy one can derive a key of length (and thus HILL entropy) at most $k - 2\log(1/\epsilon)$ bits. In many settings, e.g. when dealing with biometric data, such a $2\log(1/\epsilon)$ bit entropy loss in not an option.
Our main technical contribution is a theorem that states that in the high entropy regime, unpredictability implies HILL entropy. Concretely, any variable K with $|K| - d$ bits of unpredictability entropy has the same amount of so called metric entropy (against real-valued, deterministic distinguishers), which is known to imply the same amount of HILL entropy. The loss in circuit size in this argument is exponential in the entropy gap d, and thus this result only applies for small d (i.e., where the size of distinguishers considered is exponential in d).
To overcome the above restriction, we investigate if it's possible to first "condense" unpredictability entropy and make the entropy gap small. We show that any source with k bits of unpredictability can be condensed into a source of length k with $k - 3$ bits of unpredictability entropy. Our condenser simply "abuses" the GL construction and derives a k bit key from a source with k bits of unpredicatibily. The original GL theorem implies nothing when extracting that many bits, but we show that in this regime, GL still behaves like a "condenser" for unpredictability. This result comes with two caveats (1) the loss in circuit size is exponential in k and (2) we require that the source we start with has *no* HILL entropy (equivalently, one can efficiently check if a guess is correct). We leave it as an intriguing open problem to overcome these restrictions or to prove they're inherent.

1 Introduction

Key-derivation considers the following fundamental problem: Given a joint distribution (X, Z) where $X|Z$ (which is short for "X conditioned on Z") is guaranteed to have some kind of entropy, derive a "good" key $K = h(X, S)$ from

⋆ Research supported by the WELCOME/2010-4/2 grant.
⋆⋆ Research supported by ERC starting grant (259668-PSPC).

X by means of some efficient key-derivation function h, possibly using public randomness S.

In practice, one often uses a cryptographic hash function like SHA3 as the key derivation function $h(.)$ [Kra10, DGH+04], and then simply assumes that $h(.)$ behaves like a random oracle [BR93].

In this paper we continue the investigation of key-derivation with provable security guarantees, where we don't make any computational assumption about $h(.)$. This problem is fairly well understood for sources $X|Z$ that have high min-entropy (we'll formally define all the entropy notions used in 2 below), or are computationally indistinguishable from having so (in this case, we say $X|Z$ has high HILL entropy). In the case where $X|Z$ has k bits of min-entropy, we can either use a strong extractor to derive a $k - 2\log\epsilon^{-1}$ key that is ϵ-close to uniform, or a condenser to get a k bit key which is ϵ-close to a variable with $k - \log\log\epsilon^{-1}$ bits of min-entropy. Using extractors/condensers like this also works for HILL entropy, except that now we only get computational guarantees (pseudorandom/high HILL entropy) on the derived key.

Often one has to derive a key from a source $X|Z$ which has no HILL entropy at all. The weakest assumption we can make on $X|Z$ for any kind of key-derivation to be possible, is that X is hard to predict given Z. This has been formalized in [HLR07] by saying that $X|Z$ has k bits of unpredictability entropy, denoted $H^{\mathsf{unp}}_s(X|Z) \geqslant k$, if no circuit of size s can predict X given Z with advantage $\geqslant 2^{-k}$ (to be more general, we allow an additional parameter $\delta \geqslant 0$, and $H^{\mathsf{unp}}_{\delta,s}(X|Z) \geqslant k$ holds if (X,Z) is δ-close to some distribution (Y,Z) with $H^{\mathsf{unp}}_s(Y|Z) \geqslant k$). We will also consider a more restricted notion, where we say that $X|Z$ has k bits of *list*-unpredictability entropy, denoted $H^{*\mathsf{unp}}_s(X|Z) \geqslant k$, if it has k bits of unpredictability entropy relative to an oracle Eq which can be used to verify the correct guess (Eq outputs 1 on input X, and 0 otherwise).[4] We'll discuss this notion in more detail below. For now, let us just mention that for the important special case where it's easy to verify if a guess for X is correct (say, because we condition on $Z = f(X)$ for some one-way function[5] f), the oracle Eq does not help, and thus unpredictability and list-unpredictability coincide. The results proven in this paper imply that from a source $X|Z$ with k bits of list-unpredictability entropy, it's possible to extract a k bit key with $k-3$ bits of HILL entropy

Proposition 1. *Consider a joint distribution (X,Z) over $\{0,1\}^n \times \{0,1\}^m$ where*

$$H^{*\mathsf{unp}}_{s,\gamma}(X|Z) \geqslant k \tag{1}$$

[4] We chose this name as having access to Eq is equivalent to being allowed to output a list of guesses. This is very similar to the well known concept of list-decoding.

[5] To be precise, this only holds for *injective* one-way functions. One can generalise list-unpredictability and let Eq output 1 on some set $\mathcal{X}$, and the adversary wins if she outputs any $X \in \mathcal{X}$. Our results (in particular Theorem 1) also hold for this more general notion, which captures general one-way functions by letting $\mathcal{X} = f^{-1}(f(X))$ be the set of all preimages of $Z = f(X)$.

Let $S \in \{0,1\}^{n \times k}$ *be uniformly random and* $K = X^T S \in \{0,1\}^k$*, then the unpredictability entropy of* K *is*

$$H^{\mathsf{unp}}_{s/2^{2k}poly(m,n),\gamma}(K|Z,S) \geqslant k-3 \tag{2}$$

and the HILL entropy of K *is*

$$H^{\mathsf{HILL}}_{t,\epsilon+\gamma}(K|Z,S) \geqslant k-3 \tag{3}$$

with[6] $t = s \cdot \frac{\epsilon^7}{2^{2k}poly(m,n)}$.

Proposition 1 follows from two results we prove in this paper.

First, in Section 4 we prove Theorem 1 which shows how to "abuse" the Goldreich-Levin hardcore bits by generating a k bit key $K = X^T S$ from a source $X|Z$ with k bits of list-unpredictability. It is known that the Goldreich-Levin theorem [GL89] can be used as to extract from list-unpredictability (this was used in [HILL99] for the case when Z is a one-way function of X), yet implies nothing about the pseudorandomness of $K|(Z,S)$ when extracting that many bits. Instead, we prove that GL is a good "condenser" for unpredictability entropy: if $X|Z$ has k bits of list-unpredictability entropy, then $K|(Z,S)$ has $k-3$ bits of unpredictability entropy (note that we start with list-unpredictability, but only end up with "normal" unpredictability entropy). This result is used in the first step in Proposition 1, showing that (1) implies (2).

Second, in Section 5 we prove our main result, Theorem 2 which states that any source $X|Z$ which has $|X|-d$ bits of unpredictability entropy, has the same amount of HILL entropy (technically, we show that it implies the same amount of metric entropy against deterministic real-valued distinguishers. This notion implies the same amount of HILL entropy as shown by Barak et al. [BSW03]). The security loss in this argument is exponential in the entropy gap d. Thus, if d is very large, this argument is useless, but if we first condense unpredictability as just explained, we have a gap of only $d = 3$. This result is used in the second step in Proposition 1, showing that (2) implies (3). In the two sections below we discuss two shortcomings of Theorem 1 which we hope can be overcome in future work.[7]

On the dependency on 2^k in Theorem 1. As outlined above, our first result is Theorem 1, which shows how to condense a source with k bits of list-unpredictability into a k bit key having $k-3$ bits of unpredictability entropy.

[6] We denote with $poly(m,n)$ some fixed polynomial in (n,m), but it can denote different polynomial throughout the paper. In particular, the *poly* here is not the same as in (2) as it hides several extra terms.

[7] After announcing this result at a workshop, we learned that Colin Jia Zheng proved a weaker version of this result. Theorem 4.18 in this PhD thesis, which is available via http://dash.harvard.edu/handle/1/11745716 also states that k bits of unpredictability imply k bits of HILL entropy. Like in our case, the loss in circuit size in his proof is polynomial in ϵ^{-1}, but it's also exponential in n (the length of X), whereas our loss is only exponential in the entropy gap $\Delta = n-k$.

The loss in circuit size is $2^{2k}poly(m,n)$, and it's not clear if the dependency on 2^k is necessary here, or if one can replace the dependency on 2^k with a dependency on $poly(\epsilon^{-1})$ at the price of an extra ϵ term in the distinguishing advantage. In many settings $\log(\epsilon^{-1})$ is in the order of k, in which case the above difference is not too important. This is for example the case when considering a k bit key for a symmetric primitive like a block-cipher, where one typically assumes the hardness of the cipher to be exponential in the key-length (and thus, if we want ϵ to be in the same order, we have $\log(\epsilon^{-1}) = \Theta(k)$). In other settings, k can be superlinear in $\log(\epsilon^{-1})$, e.g., if the the high entropy string is used to generate an RSA key.

List vs. normal Unpredictability. Our Theorem 1 shows how to condense a source where $X|Z$ has k bits of *list*-unpredictability entropy into a k bit string with $k-3$ bits unpredictability entropy. It's an open question to which extent it's necessary to assume *list*-unpredictability here, maybe "normal" unpredictability is already sufficient? Note that list-unpredictability is a lower bound for unpredictability as one always can ignore the Eq oracle, i.e., $H^{\mathsf{unp}}_{\epsilon,s}(X|Z) \geqslant H^{*\mathsf{unp}}_{\epsilon,s}(X|Z)$, and in general, list-unpredictability can be much smaller than unpredictability entropy.[8]

Interestingly, we can derive a k bit key with almost k bits of HILL entropy from a source $X|Z$ which k bits unpredictability entropy $H^{\mathsf{unp}}_{\epsilon,s}(X|Z) \geqslant k$ in two extreme cases, namely, if either

1. if $X|Z$ has basically no HILL entropy (even against small circuits).
2. or when $X|Z$ has (almost) k bits of (high quality) HILL entropy.

In case 1. we observe that if $H^{\mathsf{HILL}}_{\epsilon,t}(X|Z) \approx 0$ for some $t \ll s$, or equivalently, given Z we can efficiently distinguish X from any $X' \neq X$, then the Eq oracle used in the definition of list-unpredictability can be efficiently emulated, which means it's redundant, and thus $X|Z$ has the same amount of list-unpredictability and unpredictability entropy, $H^{\mathsf{unp}}_{s,\epsilon}(X|Z) \approx H^{*\mathsf{unp}}_{s',\epsilon'}(X|Z)$ for $(\epsilon', s') \approx (\epsilon, s)$. Thus, we can use Theorem 1 to derive a k bit key with $k - O(1)$ bits of HILL entropy in this case. In case 2., we can simply use any condenser for min-entropy to get a key with HILL entropy $k - \log\log\epsilon^{-1}$ (cf. Figure 2). As condensing almost all the unpredictability entropy into HILL entropy is possible in the two extreme cases where $X|Z$ has either no or a lot of HILL entropy, it seems conceivable that it's also possible in all the in-between cases (i.e., without making any additional assumptions about $X|Z$ at all).

GL vs. Condensing. Let us stress as this point that, because of the issues discussed above, our result does not always allow generate more bits with high HILL entropy than just using the Goldreich-Levin theorem. Out of k bits of unpredictability, our technique gets $k-3$ bits of HILL, whereas GL gives only

[8] E.g., let X by uniform over $\{0,1\}^n$ and Z arbitrary, but independent of X, then for $s = \exp(n)$ we have $H^{\mathsf{unp}}_s(X|Z) = n$ but $H^{*\mathsf{unp}}_s(X|Z) = 0$ as we can simply invoke Eq on all $\{0,1\}^n$ until X is found.

$k - 2\log(1/\epsilon)$; thus we extract more. However our reduction has a quantitatively larger loss in circuit size compared to the GL theorem, so that the guaranteed security (indistinguishability) is against weaker adversaries. In general, the amount of unpredictability (or any other computational) entropy of $X|Z$ can decrease when we consider more powerful adversaries.

2 Entropy Notions

In this section we formally define the different entropy notions considered in this paper. We denote with $\mathcal{D}_s^{rand,\{0,1\}}$ the set of all *probabilistic* circuits of size s with *boolean* output, and $\mathcal{D}_s^{rand,[0,1]}$ denotes the set of all *probabilistic* circuits with *real-valued* output in the range $[0,1]$. The analogous *deterministic* circuits are denoted $\mathcal{D}_s^{det,\{0,1\}}$ and $\mathcal{D}_s^{det,[0,1]}$. We use $X \sim_{\epsilon,s} Y$ to denote computational indistinguishability of variables X and Y, formally[9]

$$X \sim_{\epsilon,s} Y \iff \forall \mathsf{C} \in \mathcal{D}_s^{rand,\{0,1\}} \ : \ |\Pr[\mathsf{C}(X)=1] - \Pr[\mathsf{C}(Y)=1]| \leqslant \epsilon \quad (4)$$

$X \sim_\epsilon Y$ denotes that X and Y have statistical distance ϵ, i.e., $X \sim_{\epsilon,\infty} Y$, and with $X \sim Y$ we denote that they're identically distributed. With U_n we denote the uniform distribution over $\{0,1\}^n$.

Definition 1. *The* **min-entropy** *of a random variable X with support $\mathcal{X}$ is*

$$H_\infty(X) = -\log_2 \max_{x\in\mathcal{X}} \Pr[X=x]$$

For a pair (X,Z) of random variables, the **average min-entropy** *of X conditioned on Z is*

$$\widetilde{H}_\infty(X|Z) \ = \ -\log_2 \mathop{\mathbb{E}}_{z\leftarrow Z} \max_x \Pr[X=x|Z=z] \ = \ -\log_2 \mathop{\mathbb{E}}_{z\leftarrow Z} 2^{-H_\infty(X|Z=z)}$$

HILL entropy is a computational variant of min-entropy, where X (conditioned on Z) has k bits of HILL entropy, if it cannot be distinguished from some Y that (conditioned on Z) has k bits of min-entropy, formally

Definition 2 ([HILL99,BSW03,HLR07]). *A random variable X has* **HILL entropy** *k, denoted by $H^{\mathsf{HILL}}_{\varepsilon,s}(X) \geq k$, if there exists a distribution Y satisfying $H_\infty(Y) \geq k$[10] and $X \sim_{\varepsilon,s} Y$.*

Let (X,Z) be a joint distribution of random variables. Then X has **conditional HILL entropy** *k conditioned on Z, denoted by $H^{\mathsf{HILL}}_{\varepsilon,s}(X|Z) \geq k$, if there exists a joint distribution (Y,Z) such that $\widetilde{H}_\infty(Y|Z) \geq k$ and $(X,Z) \sim_{\varepsilon,s} (Y,Z)$.*

[9] Let us mention that the choice of the distinguisher class in (4) irrelevant (up to a small additive difference in circuit size), we can replace $\mathcal{D}_s^{rand,\{0,1\}}$ with any of the three other distinguisher classes.

[10] The modern definition of HILL entropy is based on min-entropy, following [BSW03] and subsequent works, although [HILL99] formulated it for Shannon entropy.

Barak, Sahaltiel and Wigderson [BSW03] define the notion of metric entropy, which is defined like HILL, but the quantifiers are exchanged. That is, instead of asking for a single distribution (Y, Z) that fools all distinguishers, we only ask that for every distinguisher D, there exists such a distribution. For reasons discussed in Section 2, in the definition below we make the class of distinguishers considered explicit.

Definition 3 ([BSW03], [FR12]). *Let (X, Z) be a joint distribution of random variables. Then X has* **conditional metric entropy** *k conditioned on Z (against probabilistic boolean distinguishers), denoted $H^{\mathsf{Metric},rand,\{0,1\}}_{\varepsilon,s}(X|Z) \geq k$, if for every $\mathsf{D} \in \mathcal{D}^{rand,\{0,1\}}_s$ there exists a joint distribution (Y, Z) such that $\widetilde{H}_\infty(Y|Z) \geq k$ and*

$$|\Pr[\mathsf{D}(X, Z) = 1] - \Pr[\mathsf{D}(Y, Z) = 1]| \leqslant \epsilon$$

More generally, for $class \in \{rand, det\}, range \in \{[0, 1], \{0, 1\}\}$, $H^{\mathsf{Metric},class,range}_{\varepsilon,s}(X|Z) \geq k$ if for every $\mathsf{D} \in \mathcal{D}^{class,range}_s$ such a (Y, Z) exists.

Like HILL entropy, also unpredictability entropy, which we'll define next, can be seen as a computational variant of min-entropy. Here we don't require indistinguishability as for HILL entropy, but only that the variable is hard to predict.

Definition 4 ([HLR07]). *X has* **unpredictability entropy** *k conditioned on Z, denoted by $H^{\mathsf{unp}}_{\epsilon,s}(X|Z) \geq k$, if (X, Z) is (ϵ, s) indistinguishable from some (Y, Z), where no probabilistic circuit of size s can predict Y given Z with probability better than 2^{-k}, i.e.,*

$$H^{\mathsf{unp}}_{s,\epsilon}(X|Z) \geq k \iff \exists (Y, Z), (X, Z) \sim_{\varepsilon,s} (Y, Z)\ \forall \mathsf{C}, |\mathsf{C}| \leqslant s : \Pr_{(y,z)\leftarrow(Y,Z)}[\mathsf{C}(z) = y] \leqslant 2^{-k} \quad (5)$$

We also define a notion called "list-unpredictability", denoted $H^{*\mathsf{unp}}_{\epsilon,s}(X|Z) \geq k$, which holds if $H^{\mathsf{unp}}_{\epsilon,s}(X|Z) \geq k$ as in (5), but where C additionally gets oracle access to a function $\mathsf{Eq}(.)$ which outputs 1 on input y and 0 otherwise. So, C can efficiently test if some candidate guess for y is correct.[11]

Remark 1 (The ϵ parameter). The ϵ parameter in the definition above is not really necessary, following [HLR07], we added it so we can have a "smooth" notion, which is easier to compare to HILL or smooth min-entropy. If $\epsilon = 0$, we'll simply omit it, then the definition simplifies to

$$H^{\mathsf{unp}}_{s}(X|Z) \geq k \iff \Pr_{(x,z)\leftarrow(X,Z)}[\mathsf{C}(z) = x] \leqslant 2^{-k}$$

[11] We name this notion "list-unpredictability" as we get the same notion when instead of giving C oracle access to $\mathsf{Eq}(.)$, we allow $\mathsf{C}(z)$ to output a list of guesses for y, not just one value, and require that $\Pr_{(y,z)\leftarrow(Y,Z)}[y \in \mathsf{C}(z)] \leqslant 2^{-k}$. This notion is inspired by the well known notion of list-decoding.

Let us also mention that unpredictability entropy is only interesting if the conditional part Z is not empty as (already for s that is linear in the length of X) we have $H^{\mathsf{unp}}_s(X) = H_\infty(X)$ which can be seen by considering the circuit C (that gets no input as Z is empty) which simply outputs the constant x maximizing $\Pr[X = x]$.

Metric vs. HILL. We will use a lemma which states that deterministic real-valued metric entropy implies the same amount of HILL entropy (albeit, with some loss in quality). This lemma has been proven by [BSW03] for the unconditional case, i.e., when Z in the lemma below is empty, it has been observed by [FR12, CKLR11] that the proof also holds in the conditional case as stated below

Lemma 1 ([BSW03, FR12, CKLR11]). *For any joint distribution* $(X, Z) \in \{0,1\}^n \times \{0,1\}^m$ *and any* ϵ, δ, k, s

$$H^{\mathsf{Metric},det,[0,1]}_{\epsilon,s}(X|Z) \geqslant k \quad \Rightarrow \quad H^{\mathsf{HILL}}_{\epsilon+\delta, s\cdot\delta^2/(m+n)}(X|Z) \geqslant k$$

Note that in Definition 2 of HILL entropy, we only consider security against probabilistic boolean distinguishers (as $\sim_{\epsilon,s}$ was defined this way), whereas in Definiton 3 of metric entropy we make the class of distinguishers explicit. The reason for this is that in the definition of HILL entropy the class of distinguishers considered is irrelevant (except for a small additive degradation in circuit size, cf. [FR12, Lemma 2.1]).[12] Unlike for HILL, for metric entropy the choice of the distinguisher class does matter. In particular, deterministic boolean metric entropy $H^{\mathsf{Metric},det,\{0,1\}}_{\epsilon,s}(X|Y) \geqslant k$ is only known to imply deterministic real-valued metric entropy $H^{\mathsf{Metric},det,[0,1]}_{\epsilon+\delta,s}(X|Y) \geqslant k - \log(\delta^{-1})$, i.e., we must allow for a $\delta > 0$ loss in distinguishing advantage, and this will at the same time result in a loss of $\log(\delta^{-1})$ in the amount of entropy. For this reason, it is crucial that in Theorem 2 we show that unpredictability entropy implies deterministic *real-valued* metric entropy, so we can then apply Lemma 1 to get the same amount of HILL entropy. Dealing with real-valued distinguishers is the main source of technical difficulty in the proof of the Theorem 2, proving the analogous statement for deterministic *boolean* distinguishers is much simpler.

3 Known Results on Provably Secure Key-Derivation

We say that a cryptographic scheme has security α, if no adversary (from some class of adversaries like all polynomial size circuits) can win some security game with advantage $\geqslant \alpha$ if the scheme is instantiated with a uniformly random string.[13] Below we will distinguish between *unpredictability* applications, where

[12] This easily follows from the fact that in the definition (4) of computational indistinguishability the choice of the distinguisher class is irrelevant.

[13] We'll call this string "key". Though in many settings (in particular when keys are not simply uniform random strings, like in public-key crypto) this string is not used as a key directly, but one rather should think of it as the randomness used to sample the actual keys.

the advantage bounds the probability of winning some security game (a typical example are digital signature schemes, where the game captures the existential unforgeability under chosen message attacks), and *indistinguishability* applications, where the advantage bounds the distinguishing advantage from some ideal object (a typical example is the security definition of pseudorandom generators or functions).

3.1 Key-Derivation from Min-Entropy

Strong Extractors. Let (X, Z) be a source where $\widetilde{H}_\infty(X|Z) \geqslant k$, or equivalently, no adversary can guess X given Z with probability better than 2^{-k} (cf. Def. 1). Consider the case where we want to derive a key $K = h(X, S)$ that is statistically close to uniform given (Z, S). For example, X could be some physical source (like statistics from keystrokes) from which we want to generate almost uniform randomness. Here Z models potential side-information the adversary might have on X. This setting is very well understood, and such a key can be derived using a strong extractor as defined below.

Definition 5 ([WZ93], [DORS08]). *A function* $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^\ell$ *is an average-case* (k, ϵ)*-strong extractor if for every distribution* (X, Z) *over* $\{0,1\}^n \times \{0,1\}^m$ *with* $\widetilde{H}_\infty(X|Z) \geqslant k$ *and* $S \sim U_d$*, the distribution of* $(\mathsf{Ext}(X, S), S, Z)$ *has statistical distance* ϵ *to* (U_m, S, Z)*.*

Extractors Ext as above exist with $\ell = k - 2\log(1/\epsilon)$ [HILL99]. Thus, from any (X, Z) where $\widetilde{H}_\infty(X|Z) \geqslant k$ we can extract a key $K = \mathsf{Ext}(X, S)$ of length $k - 2\log(1/\epsilon)$ that is ϵ close to uniform [HILL99]. The entropy gap $2\log(1/\epsilon)$ is optimal by the so called "RT-bound" [RT00], even if we assume the source is efficiently samplable [DPW14].

If instead of using a uniform ℓ bit key for an α secure scheme, we use a key that is ϵ close to uniform, the scheme will still be at least $\beta = \alpha + \epsilon$ secure. In order to get security β that is of the same order as α, we thus must set $\epsilon \approx \alpha$. When the available amount k of min-entropy is small, for example when dealing with biometric data [DORS08, BDK+05], a loss of $2\log(1/\epsilon)$ bits (that's 160 bits for a typical security level $\epsilon = 2^{-80}$) is often unacceptable.

Condensers. The above bound is basically tight for many *indistinguishability* applications like pseudorandom generators or pseudorandom functions.[14] Fortunately, for many applications a close to uniform key is not necessary, and a key $|K|$ with min-entropy $|K| - \Delta$ for some small Δ is basically as good as a uniform one. This is the case for all *unpredictability* applications, which includes

[14] For example, consider a pseudorandom function $\mathsf{F} : \{0,1\}^k \times \{0,1\}^a \to \{0,1\}$ and a key K that is uniform over all keys where $\mathsf{F}(K, 0) = 0$, this distribution is $\epsilon \approx 1/2$ close to uniform and has min-entropy $\approx |K| - 1$, but the security breaks completely as one can distinguish $\mathsf{F}(U_k, .)$ from $\mathsf{F}(K, .)$ with advantage $\beta \approx 1/2$ (by quering on input 0, and outputting 1 iff the output is 0).

OWFs, digital-signatures and MACs.[15] It's not hard to show that if the scheme is α secure with a uniform key it remains at least $\beta = \alpha 2^{\Delta}$ secure (against the same class of attackers) if instantiated with any key K that has $|K| - \Delta$ bits of min-entropy.[16] Thus, for unpredictability applications we don't have to extract an almost uniform key, but "condensing" X into a key with $|K| - \Delta$ bits of min-entropy for some small Δ is enough.

[DPW14] show that a $(\log \epsilon^{-1} + 1)$-wise independent hash function $\mathsf{Cond} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^\ell$ is a condenser with the following parameters. For any (X, Z) where $\widetilde{H}_\infty(X|Z) \geqslant \ell$, for a random seed S (used to sample a $(\log \epsilon + 1)$-wise independent hash function), the distribution $(\mathsf{Cond}(X, S), S)$ is ϵ close to a distribution (Y, S) where $\widetilde{H}_\infty(Y|Z) \geqslant \ell - \log\log(1/\epsilon)$. Using such an ℓ bit key (condensed from a source with ℓ bits min-entropy) for an unpredictability application that is α secure (when using a uniform ℓ bit key), we get security $\beta \leqslant \alpha 2^{\log\log(1/\epsilon)} + \epsilon$, which setting $\epsilon = \alpha$ gives $\beta \leqslant \alpha(1 + \log(1/\alpha))$ security, thus, security degrades only by a logarithmic factor.

3.2 Key-Derivation from Computational Entropy

The bounds discussed in this section are summarised in Figures 1 and 2 in Appendix A. The last row of Figure 2 is the new result proven in this paper.

HILL Entropy. As already discussed in the introduction, often we want to derive a key from a distribution (X, Z) where there's no "real" min-entropy at all $\widetilde{H}_\infty(X|Z) = 0$. This is for example the case when Z is the transcript (that can be observed by an adversary) of a key-exchange protocol like Diffie-Hellman, where the agreed value $X = g^{ab}$ is determined by the transcript $Z = (g^a, g^b)$ [Kra10, GKR04]. Another setting where this can be the case is in the context of side-channel attacks, where the leakage Z from a device can completely determine its internal state X.

If $X|Z$ has k bits of HILL entropy, i.e., is computationally indistinguishable from having min-entropy k (cf. Def. 2) we can derive keys exactly as described above assuming $X|Z$ had k bits of min-entropy. In particular, if $X|Z$ has $|K| + 2\log(1/\epsilon)$ bits of HILL entropy for some negligible ϵ, we can derive a key K that is pseudorandom, and if $X|Z$ has $|K| + \log\log(1/\epsilon)$ bits of HILL entropy, we can

[15] [DY13] identify an interesting class of applications called "square-friendly", this class contains all unpredictability applications, and some indistinguishability applications like weak PRFs (which are PRFs that can only be queried on random inputs). This class of applications remains somewhat secure even for a small entropy gap Δ: For $\Delta = 1$ the security is $\beta \approx \sqrt{\alpha}$. This is worse that the $\beta = 2\alpha$ for unpredictability applications, but much better than the complete loss of security $\beta \approx 1/2$ required for some indistinguishability apps like (standard) PRFs.

[16] Assume some adversary breaks the scheme, say, forges a signature, with advantage β if the key comes from the distribution K. If we sample a uniform key instead, it will have the same distribution as K conditioned on an event that holds with probability $2^{-\Delta}$, and thus this adversary will still break the scheme with probability $\beta/2^{\Delta}$.

derive a key that is almost as good as a uniform one for any unpredictability application.

Unpredictability Entropy. Clearly, the minimal assumption we must make on a distribution $(X, Z) \in \{0,1\}^n \times \{0,1\}^m$ for any key derivation to be possible at all is that X is hard to compute given Z, that is, $X|Z$ must have some unpredictability entropy as in Definition 4. Goldreich and Levin [GL89] show how to generate pseudorandom bits from such a source. In particular, the Goldreich-Levin theorem implies that if $X|Z$ has at least $2\log\epsilon^{-1}$ bits of list-unpredictability, then the inner product $R^T X$ of X with a random vector R is ϵ indistinguishable from uniformly random (the loss in circuit size is $poly(n,m)/\epsilon^4$). Using the chain rule for unpredictability entropy,[17] we can generate an $\ell = k - 2\log\epsilon^{-1}$ bit long pseudorandom string that is $\ell\epsilon$ indistinguishable (the extra ℓ factor comes from taking the union bound over all bits) from uniform.

Thus, we can turn k bits of list-unpredictability into $k-2\log\epsilon^{-1}$ bits of pseudorandom bits (and thus also that much HILL entropy) with quality roughly ϵ. The question whether it's possible to generate significantly more than $k - 2\log\epsilon^{-1}$ of HILL entropy from a source with k bits of (list-)unpredictability seems to have never been addressed in the literature before. The reason might be that one usually is interested in generating pseudorandom bits (not just HILL entropy), and for this, the $2\log\epsilon^{-1}$ entropy loss is inherent. The observation that for many applications high HILL entropy is basically as good as pseudorandomness is more recent, and recently gained attention by its usefulness in the context of leakage-resilient cryptography [DP08, DY13].

In this paper we prove that it's in fact possible to turn almost all list-unpredictability into HILL entropy.

4 Condensing Unpredictability

Below we state Theorem 1 whose proof is in Appendix B, but first, let us give some intuition. Let $X|Z$ have k bits of list-unpredictability, and assume we start extracting Goldreich-Levin hardcore bits $A_1, A_2, \ldots$ by taking inner products $A_i = R_i^T X$ for random R_i. The first extracted bits $A_1, A_2, \ldots$ will be pseudorandom (given the R_i and Z), but with every extracted bit, the list-unpredictability can also decrease by one bit. As the GL theorem requires at least $2\log\epsilon^{-1}$ bits of list-unpredictability to extract an ϵ secure pseudorandom bit, we must stop after $k-2\log\epsilon^{-1}$ bits. In particular, the more we extract, the worse the pseudorandomness of the extracted string becomes. Unlike the original GL theorem, in our Theorem 1 we only argue about the unpredictability of the extracted string, and unpredictability entropy has the nice property that it can never decrease, i.e., predicting $A_1, \ldots, A_{i+1}$ is always at least as hard as predicting $A_1, \ldots, A_i$.

[17] Which states that if $X|Z$ has k bits of list-unpredictability, then for any (A, R) where R is independent of (X, Z), $X|(Z, A, R)$ has $k - |A|$ bits of list-unpredictability entropy. In particular, extracting ℓ inner product bits, decreases the list-unpredictability by at most ℓ.

Thus, despite the fact that once i approaches k it becomes easier and easier to predict A_i (given $A_1, \ldots, A_{i-1}, Z$ and the R_i's)[18] this hardness will still add up to $k - O(1)$ bits of unpredictability entropy.

The proof is by contradiction, we assume that $A_1, \ldots, A_k$ can be predicted with advantage 2^{-k+3} (i.e., does not have $k-3$ bits of unpredictability), and then use such a predictor to predict X with advantage $> 2^{-k}$, contradicting the k bit list-unpredictability of $X|Z$.

If $A_1, \ldots, A_k$ can be predicted as above, then there must be an index j s.t. A_j can be predicted with good probability conditioned on $A_1, \ldots, A_{j-1}$ being correctly predicted. We then can use the Goldreich-Levin theorem, which tells us how to find X given such a predictor. Unfortunately, j can be close to k, and to apply the GL theorem, we first need to find the right values for $A_1, \ldots, A_{j-1}$ on which we condition, and also can only use the predictor's guess for A_j if it was correct on the first $j-1$ bits. We have no better strategy for this than trying all possible values, and this is the reason why the loss in circuit size in Theorem 1 depends on 2^k.

In our proof, instead of using the Goldreich-Levin theorem, we will actually use a more fine-grained variant due to Hast which allows to distinguish between errors and erasures [Has03] (i.e., cases where we know that we don't have any good guess. As outlined above, this will be the case whenever the predictor's guess for the first $j-1$ inner products was wrong, and thus we can't assume anything about the jth guess being correct). This will give a much better quantitative bound than what seems possible using GL.

Theorem 1 (Condensing Upredictability Entropy). *Consider any distribution (X, Z) over $\{0,1\}^n \times \{0,1\}^m$ where*

$$H^{*\mathsf{unp}}_{\epsilon,s}(X|Z) \geqslant k$$

then for a random $R \leftarrow \{0,1\}^{k \times n}$

$$H^{\mathsf{unp}}_{\epsilon,t}(R.X|Z,R) \geqslant k - \Delta$$

where[19]

$$t = \frac{s}{2^{2k}\,\mathrm{poly}(m,n)} \quad , \quad \Delta = 3$$

5 High Unpredictability implies Metric Entropy

In this section we state our main results, showing that k bits of unpredictability entropy imply the same amount of HILL entropy, with a loss exponential in the "entropy gap". The proof is in Appendix C.

[18] The only thing we know about the last extracted bit A_k is that it cannot be predicted with advantage $\geqslant 0.75$, more generally, A_{k-j} cannot be predicted with advantage $1/2 + 1/2^{j+2}$.

[19] We can set Δ to be any constant > 1 here, but choosing a smaller Δ would imply a smaller t.

Theorem 2 (Unpredictability Entropy Implies HILL Entropy). *For any distribution (X, Z) over $\{0,1\}^n \times \{0,1\}^m$, if $X|Z$ has unpredictability entropy*

$$H^{\mathsf{unp}}_{\gamma,s}(X|Z) \geqslant k \tag{6}$$

then, with $\Delta = n - k$ denoting the entropy gap, $X|Z$ has (real valued, deterministic) metric entropy

$$H^{\mathsf{Metric},det,[0,1]}_{\epsilon+\gamma,t}(X|Z) \geqslant k \quad \textit{for} \quad t = \Omega\left(s \cdot \frac{\epsilon^5}{2^{5\Delta}\log^2\left(2^{\Delta}\epsilon^{-1}\right)}\right) \tag{7}$$

By Lemma 1 this further implies that $X|Z$ has, for any $\delta > 0$, HILL entropy

$$H^{\mathsf{HILL}}_{\epsilon+\delta+\gamma,\Omega(t\delta^2/(n+m))}(X|Z) \geqslant k$$

which for $\epsilon = \delta = \gamma$ is

$$H^{\mathsf{HILL}}_{3\epsilon,\Omega(s\cdot\epsilon^7/2^{5\Delta}(n+m)\log^2(2^{\Delta}\epsilon^{-1}))}(X|Z) \geqslant k$$

References

BDK+05. Xavier Boyen, Yevgeniy Dodis, Jonathan Katz, Rafail Ostrovsky, and Adam D. Smith, *Secure remote authentication using biometric data*, Advances in Cryptology - EUROCRYPT 2005, 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Aarhus, Denmark, May 22-26, 2005, Proceedings, 2005, pp. 147–163.

BR93. Mihir Bellare and Phillip Rogaway, *Random oracles are practical: A paradigm for designing efficient protocols*, CCS '93, Proceedings of the 1st ACM Conference on Computer and Communications Security, Fairfax, Virginia, USA, November 3-5, 1993., 1993, pp. 62–73.

BSW03. Boaz Barak, Ronen Shaltiel, and Avi Wigderson, *Computational analogues of entropy*, Approximation, Randomization, and Combinatorial Optimization: Algorithms and Techniques, 6th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, APPROX 2003 and 7th International Workshop on Randomization and Approximation Techniques in Computer Science, RANDOM 2003, Princeton, NJ, USA, August 24-26, 2003, Proceedings, 2003, pp. 200–215.

CKLR11. Kai-Min Chung, Yael Tauman Kalai, Feng-Hao Liu, and Ran Raz, *Memory delegation*, Advances in Cryptology - CRYPTO 2011 - 31st Annual Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2011. Proceedings, 2011, pp. 151–168.

DGH+04. Yevgeniy Dodis, Rosario Gennaro, Johan Håstad, Hugo Krawczyk, and Tal Rabin, *Randomness extraction and key derivation using the cbc, cascade and HMAC modes*, Advances in Cryptology - CRYPTO 2004, 24th Annual International CryptologyConference, Santa Barbara, California, USA, August 15-19, 2004, Proceedings, 2004, pp. 494–510.

DORS08. Yevgeniy Dodis, Rafail Ostrovsky, Leonid Reyzin, and Adam D. Smith, *Fuzzy extractors: How to generate strong keys from biometrics and other noisy data*, SIAM J. Comput. **38** (2008), no. 1, 97–139.

DP08. Stefan Dziembowski and Krzysztof Pietrzak, *Leakage-resilient cryptography*, 49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA, 2008, pp. 293–302.

DPW14. Yevgeniy Dodis, Krzysztof Pietrzak, and Daniel Wichs, *Key derivation without entropy waste*, Advances in Cryptology - EUROCRYPT 2014 - 33rd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Copenhagen, Denmark, May 11-15, 2014. Proceedings, 2014, pp. 93–110.

DY13. Yevgeniy Dodis and Yu Yu, *Overcoming weak expectations*, Theory of Cryptography - 10th Theory of Cryptography Conference, TCC 2013, Tokyo, Japan, March 3-6, 2013. Proceedings, 2013, pp. 1–22.

FR12. Benjamin Fuller and Leonid Reyzin, *Computational entropy and information leakage*, IACR Cryptology ePrint Archive **2012** (2012), 466.

GKR04. Rosario Gennaro, Hugo Krawczyk, and Tal Rabin, *Secure hashed diffie-hellman over non-ddh groups*, Advances in Cryptology - EUROCRYPT 2004, International Conference on the Theory and Applications of Cryptographic Techniques, Interlaken, Switzerland, May 2-6, 2004, Proceedings, 2004, pp. 361–381.

GL89. Oded Goldreich and Leonid A. Levin, *A hard-core predicate for all one-way functions*, Proceedings of the 21st Annual ACM Symposium on Theory of Computing, May 14-17, 1989, Seattle, Washigton, USA, 1989, pp. 25–32.

Has03. Gustav Hast, *Nearly one-sided tests and the goldreich-levin predicate*, Advances in Cryptology - EUROCRYPT 2003, International Conference on the Theory and Applications of Cryptographic Techniques, Warsaw, Poland, May 4-8, 2003, Proceedings, 2003, pp. 195–210.

HILL99. Johan Håstad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby, *A pseudorandom generator from any one-way function*, SIAM J. Comput. **28** (1999), no. 4, 1364–1396.

HLR07. Chun-Yuan Hsiao, Chi-Jen Lu, and Leonid Reyzin, *Conditional computational entropy, or toward separating pseudoentropy from compressibility*, Advances in Cryptology - EUROCRYPT 2007, 26th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Barcelona, Spain, May 20-24, 2007, Proceedings, 2007, pp. 169–186.

Kra10. Hugo Krawczyk, *Cryptographic extraction and key derivation: The HKDF scheme*, Advances in Cryptology - CRYPTO 2010, 30th Annual Cryptology Conference, Santa Barbara, CA, USA, August 15-19, 2010. Proceedings, 2010, pp. 631–648.

RT00. Jaikumar Radhakrishnan and Amnon Ta-Shma, *Bounds for dispersers, extractors, and depth-two superconcentrators*, SIAM J. Discrete Math. **13** (2000), no. 1, 2–24.

Sko15. Maciej Skorski, *Metric pseudoentropy: Characterizations, transformations and applications*, Information Theoretic Security - 8th International Conference, ICITS 2015, Lugano, Switzerland, May 2-5, 2015. Proceedings, 2015, pp. 105–122.

WZ93. Avi Wigderson and David Zuckerman, *Expanders that beat the eigenvalue bound: explicit construction and applications*, Proceedings of the Twenty-Fifth Annual ACM Symposium on Theory of Computing, May 16-18, 1993, San Diego, CA, USA, 1993, pp. 245–251.

A Figures

Deriving a (pseudo)random key of length $\|K\| = k - 2\log\epsilon^{-1}$ from a source $(X,Z) \in \{0,1\}^n \times \{0,1\}^m$ where $X\|Z$ has k bits (min/HILL/list-unpredictability) entropy			
Entropy type	Entropy quantity and quality of source	Derive key K of length $k - 2\log\epsilon^{-1}$ as	Quality of derived key $H^{\mathsf{HILL}}_{\epsilon',s'}(K\|Z,S) = k - 2\log\epsilon^{-1} = \|K\|$ equivalently $(K,Z,S) \sim_{\epsilon',s'} (U_{\|K\|},Z,S)$
min	$\widetilde{H}_\infty(X\|Z) = k$	$K = \mathsf{Ext}(X,S)$	$\epsilon' = \epsilon \qquad s' = \infty$
HILL	$H^{\mathsf{HILL}}_{\delta,s}(X\|Z) = k$	$K = \mathsf{Ext}(X,S)$	$\epsilon' = \epsilon + \delta \qquad s' \approx s$
Unpredict.	$H^{*\mathsf{unp}}_{\delta,s}(X\|Z) = k$	$K = \mathsf{GL}(X,S) = S^T X$	$\epsilon' = m\epsilon + \delta \qquad s' = s \cdot \epsilon^4/poly(m,n)$

Fig. 1. Bounds on deriving a (pseudo)random key K of length $|K| = k - 2\log\epsilon^{-1}$ bit from a source $X|Z$ with k bits of min, HILL or list-unpredictability entropy. Ext is a strong extractor (e.g. leftover hashing), and GL denotes the Goldreich-Levin construction, which for $X \in \{0,1\}^n$ and $S \in \{0,1\}^{n\times|K|}$ is simply defined as $\mathsf{GL}(X,S) = S^T X$. Leftover hashing requires a seed of length $|S| = 2n$ (extractors with a much shorter seed $|S| = O(\log n + \log\epsilon^{-1})$ that extract $k - 2\log\epsilon^{-1} - O(1)$ bits also exist), whereas Goldreich-Levin requires a longer $|S| = |K|n$ bit seed. The above bound for HILL entropy even holds if $X|Z$ only has k bits of probabilistic boolean metric entropy (a notion implying the same amount of HILL entropy, albeit with a loss in circuit size), as shown in Theorem 2.5 of [FR12]

Deriving k bit key K with high HILL entropy from $X\|Z$ with k bits (min/HILL/list-unpredictability) entropy			
Entropy type	Entropy quantity and quality of soucre	Derive key of length $\|K\| = k$ as	Quantity and quality of HILL entropy of K $H^{\mathsf{HILL}}_{\epsilon',s'}(K\|Z,S) \geqslant k - \Delta$
min	$\widetilde{H}_\infty(X\|Z) = k$	$K = \mathsf{Cond}(X,S)$	$\epsilon' = \epsilon \qquad s' = \infty \qquad \Delta = \log\log\epsilon^{-1}$
HILL	$H^{\mathsf{HILL}}_{\delta,s}(X\|Z) = k$	$K = \mathsf{Cond}(X,S)$	$\epsilon' = \epsilon + \delta \qquad s' \approx s \qquad \Delta = \log\log\epsilon^{-1}$
Unpredict.	$H^{*\mathsf{unp}}_{\delta,s}(X\|Z) = k$	$K = \mathsf{GL}(X,S) = S^T X$	$\epsilon' = \epsilon + \delta \qquad s' = s \cdot \epsilon^7/2^{2k}poly(m,n) \qquad \Delta = 3$

Fig. 2. Bounds on deriving a key of length k with min (or HILL) entropy $k - \Delta$ from a source $X|Z$ with k bits of min, HILL or unpredictability entropy. Cond denotes a $(\log\epsilon + 1)$ wise independent hash function, which is shown to be a good condenser (as stated in the table) for min-entropy in [DPW14]. The bounds for HILL entropy follow directly from the bound for min-entropy. The last row follows from the results in this paper as stated in Proposition 1.

B Proof of Theorem 1

We will use the following theorem due Hast [Has03] on decoding Hadamard code with errors and erasures.

Theorem 3 ([Has03]). *There is an algorithm* LD *that, on input* l *and* n *and with oracle access to a binary Hadamard code of* x *(where* $|x| = n$*) with an* e*-fraction of errors and an* s*-fraction of erasures, can output a list of* 2^l *elements in time* $O(nl2^l)$ *asking* $n2^l$ *oracle queries such that the probability that* x *is contained in the list is at least* 0.8 *if* $l \geqslant \log_2(20n(e+c)/(c-e)^2 + 1)$*, where* $c = 1 - s - e$ *(the fraction of the correct answers from the oracle).*

We'll often consider sequences $v_1, v_2, \ldots$ of values and will use the notation v_a^b to denote $(v_a, \ldots, v_b)$, with $v_a^b = \emptyset$ if $a > b$. v^b is short for $v_1^b = (v_1, \ldots, v_b)$. We will also use the notation $x.y$ for the dot product.

Proof (of Theorem 1). It's sufficient to prove the theorem for $\epsilon = 0$, the general case $\epsilon \geqslant 0$ then follows directly by the definition of unpredictability entropy. To prove the theorem we'll prove its contraposition

$$H_t^{\mathsf{unp}}(R.X|Z,R) < k - \Delta \quad \Rightarrow \quad H_s^{*\mathsf{unp}}(X|Z) < k \tag{8}$$

The left-hand side of (8) means there exists a circuit A of size $|\mathsf{A}| \leqslant t$ such that

$$\Pr_{(x,z)\leftarrow(X,Z),r\leftarrow\{0,1\}^{k\times n}}[\mathsf{A}(z,r) = r.x] \geqslant 2^{-k+\Delta} \tag{9}$$

It will be convenient to assume that A initially flips a coin b, and if $b = 0$ outputs a uniformly random guess. This loses at most a factor 2 in A's advantage, i.e.,

$$\Pr_{(x,z)\leftarrow(X,Z),r\leftarrow\{0,1\}^{k\times n}}[\mathsf{A}(z,r) = r.x] \geqslant 2^{-k+\Delta-1} \tag{10}$$

but now we can assume that for any z, r and $w \in \{0,1\}^k$

$$\Pr[\mathsf{A}(z,r) = w] \geqslant 2^{-k-1} \tag{11}$$

Using the elementary inequality $\Pr[X \geqslant \lambda\mathbb{E}X] \geqslant (1-\lambda)\mathbb{E}X$ valid when $0 \leqslant X \leqslant 1$ and $\mathbb{E}X \geqslant \epsilon$, in eq.(10) with $\lambda = \frac{1}{2}$ gives us

$$\Pr_{(x,z)\leftarrow(X,Z)}\Big[\Pr_{r\leftarrow\{0,1\}^{k\times n}}[\mathsf{A}(z,r) = r.x] \geqslant 2^{-k+\Delta-2}\Big] \geqslant 2^{-k+\Delta-2} \tag{12}$$

We call $(x,z) \in \mathrm{supp}[(X,Z)]$ "good" if

$$(x,z) \text{ is good} \iff \Pr_{r\leftarrow\{0,1\}^{k\times n}}[\mathsf{A}(z,r) = r.x] \geqslant 2^{-k+\Delta-2} \tag{13}$$

Note that by eq.(12), $(z,x) \leftarrow (Z,X)$ is good with probability $\geqslant 2^{-k+\Delta-2}$.

We will use A to construct a new circuit B of size $s = O(t2^{2k}\,\mathrm{poly}(n))$ where

$$\Pr_{(x,z)\leftarrow(X,Z)}[\mathsf{B}(z) = x \mid (x,z) \text{ is good}] > 1/2 \tag{14}$$

Now (14) and (12) further gives

$$\begin{aligned}\Pr_{(x,z)\leftarrow(X,Z)}[\mathsf{B}(z) = x] &= \Pr[\mathsf{B}(z) = x | (x,z) \text{ is good}] \cdot \Pr[(x,z) \text{ is good}] \\ &> 2^{-1} \cdot 2^{-k+\Delta-2} = 2^{-k+\Delta-3}\end{aligned} \tag{15}$$

contradicting the right-hand side of (8), and thus proving the theorem.

We'll now construct B satisfying (14), for this, consider any good (x, z). Let $R = R^k = (R_1, \ldots, R_k)$ be uniformly random and let $A = A^k = (A_1, \ldots, A_k)$ where $A_i = R_i.x$.

Let $\hat{A} \leftarrow \mathsf{A}(z, R)$ and define $\epsilon_i = \Pr_R[\hat{A}_i = A_i | \hat{A}^{i-1} = A^{i-1}]$. Using (13) in the last step

$$\prod_{i=1}^{k} \epsilon_i = \Pr_R[A = \hat{A}] = \Pr_R[\mathsf{A}(z, R) = R.x] \geqslant 2^{-k+\Delta-2}$$

Thus, here exists an i s.t., $\epsilon_i \geqslant 2^{\frac{-k+\Delta-2}{k}} = \frac{1}{2} + \delta$ with $\delta \approx \frac{\Delta-2}{k} \cdot \frac{\ln(2)}{2}$. We fix this i (we don't know which i is good, and later will simply try all of them). Then

$$\mathbb{E}_{R^{i-1}}[\Pr_{R_i, R_{i+1}^k}[\hat{A}_i = A_i \mid \hat{A}^{i-1} = A^{i-1}]] \geqslant 1/2 + \delta$$

Again using $\Pr[X \geqslant \lambda \mathbb{E}X] \geqslant (1 - \lambda)\mathbb{E}X$ vthis time with $\lambda = \frac{0.5+0.5\delta}{0.5+\delta}$ we obtain

$$\Pr_{R^{i-1}}[\Pr_{R_i, R_{i+1}^k}[\hat{A}_i = A_i \mid \hat{A}^{i-1} = A^{i-1}] \geqslant 1/2 + \delta/2] \geqslant \frac{\delta}{2} \tag{16}$$

We call r^{i-1} good if (note that by the previous equation a random r^{i-1} is good with probability $\geqslant \delta/2$)

$$r^{i-1} \text{ is good} \iff \Pr_{R_i, R_{i+1}^k}[\hat{A}_i = A_i \mid \hat{A}^{i-1} = A^{i-1}] \geqslant 1/2 + \delta/2 \tag{17}$$

(note that the dependency on r^{i-1} is in the equation $\hat{A}^{i-1} = A^{i-1}$). From now on, we fix some good r^{i-1} and assume we know $a^{i-1} = r^{i-1}.x$ (later we'll simply try all possible choices for a^{i-1}).

We define a predictor $\mathsf{P}_i(r_i)$ that tries to predict $r_i.x$ given a random r_i (and also knows z, r^{i-1}, a^{i-1} as above) as follows

1. Sample random $r_{i+1}^k \leftarrow R_{i+1}^k$
2. Invoke $\hat{A}^k \leftarrow \mathsf{A}(z, r^{(i)})$. Note that $r^{(i)} := (r^{i-1}, r_i, r_{i+1}^k)$ consists of the fixed r^{i-1}, the input r_i and the randomly sampled r_{i+1}^k.
3. if $\hat{A}^{i-1} = a^{i-1}$ output $\hat{A}_i$, otherwise output $\perp$.

Considering how the output is generated and using (11), which implies $\Pr[\hat{A}^{i-1} = a^{i-1}] \geqslant 2^{-i}$, and (17) we can lower bound P_i's rate and advantage as

$$\Pr_{R_i}[\mathsf{P}_i(R_i) \neq \perp] = \Pr[\hat{A}^{i-1} = a^{i-1}] \geqslant 2^{-i},$$

$$\Pr_{R_i}[\mathsf{P}_i(R_i) = R_i.x] \geqslant \Pr[\hat{A}^{i-1} = a^{i-1}](\frac{1}{2} + \delta/2). \tag{18}$$

In terms of Theorem 3, we have a binary Hadamard code with $e + c = \Pr[\hat{A}^{i-1} = a^{i-1}]$, $c - e \geqslant \delta \cdot \Pr[\hat{A}^{i-1} = a^{i-1}]$, which implies that $(e+c)/(c-e)^2 \leqslant \frac{2^i}{\delta^2}$.

Now Theorem 3 implies that given such a predictor P we can output a list that contains x with probability > 0.8 in time $O(2^i \operatorname{poly}(m,n)) = O(2^k \operatorname{poly}(m,n))$, as we assume access to an oracle Eq with outputs 1 on input x and 0 otherwise, we can find x in this list with the same probability.

Using this, we can now construct an algorithm as claimed in (14) as follows: B will sample $i \in \{1,\dots,k\}$ and then r^{i-1} at random. Then B calls P_i with all possible $a^{i-1} \in \{0,1\}^{i-1}$. We note that with probability $\delta/2k$ (we lose a factor k for the guess of i, and $\delta/2$ is the probability of sampling a good r^{i-1}) the predictor P_i will satisfy (18).

If x is not found, B repeats the above process, but stops if x is not found after $2k/\delta$ iterations. The success probability of B is $\approx (1-1/e)0.8 > 0.5$ as claimed, the overall running time we get is $O(2^{2k} \operatorname{poly}(m,n))$. □

C Proof of Theorem 2

It's sufficient to prove the theorem for $\gamma = 0$, the case $\gamma > 0$ then follows directly by definition of unpredictability entropy. Suppose for the sake of contradiction that (7) does not hold. That is, $H_{t,\epsilon}^{\mathsf{Metric},det,[0,1]}(X|Z) < k$, which means that there exists a distinguisher $\mathsf{D} : \{0,1\}^n \times \{0,1\}^m \to [0,1]$ of size t that satisfies

$$\mathbb{E}\mathsf{D}(X,Z) - \mathbb{E}\mathsf{D}(Y,Z) \geqslant \epsilon \quad \forall (Y,Z) : \ \widetilde{H}_\infty(Y|Z) \geqslant k. \tag{19}$$

We will show how to construct an efficient algorithm that given Z uses D to predict X with probability at least 2^{-k}, contradicting (6). The core of the algorithm is the procedure Predictor described below.

Function PREDICTOR(z, D', ℓ)

Input : $z \leftarrow Z$, $[0,2]$-valued distinguisher D'
Output: $x \in \{0,1\}^n$

$b \leftarrow 1$, $i \leftarrow 1$
while $b \neq 0$ **and** $i < \ell$ **do**
  $x \leftarrow \{0,1\}^n$
  $b \leftarrow$ `BernoulliDistribution(`$\mathsf{D}'(x,z)/2$`)` `/* outputs 1 w.p.` $\mathsf{D}'(x,z)/2$ `*/`
  if $b = 0$ **then**
    $i \leftarrow i + 1$
  else
    return x
  end
end
return $\perp$

Predictor(Z, D, ℓ) samples an element $x \in \{0,1\}^n$ according to some probability distribution. This distribution captures the following intuition: as the

advantage $\mathbb{E}\mathsf{D}(X,Z) - \mathbb{E}\mathsf{D}(Y,Z)$ is positive (as assumed in (19)), we know that x being the correct guess for X is positively correlated with the value $\mathsf{D}(x,Z)$. The probability that Predictor(Z,D,ℓ) returns some particular value x as guess for X will be linear in $\mathsf{D}(x,Z)$.

Predictor(Z,D,ℓ) may also output $\perp$, which means it failed to sample an x according to this distribution. The probability of outputting $\perp$ goes exponentially fast to 0 as ℓ grows.

A toy example: predicting X when Z is empty and D is boolean. Suppose that $\mathbb{E}\mathsf{D}(X) - \mathbb{E}\mathsf{D}(Y) \geqslant \epsilon$ for all Y such that $H_\infty(Y) \geqslant k$. And assume that $\mathsf{D}(.)$ is boolean (not real valued as in our theorem). Then Predictor$(\emptyset,\mathsf{D},\ell)$ will output a guess for X that (if it's not $\perp$) is a random value x satisfying $\mathsf{D}(x) = 1$. The probability that this guess for X is correct equals $\mathbb{E}\mathsf{D}(X)/|D|$ where $|D| = \sum_x \mathsf{D}(x)$. Consider now the distribution Y of min-entropy k that maximizes $\mathbb{E}\mathsf{D}(Y)$. We can assume that Y is flat and supported on those 2^k elements x for which the value $\mathsf{D}(x)$ is the biggest possible. Observe that since $\mathbb{E}\mathsf{D}(X) - \mathbb{E}\mathsf{D}(Y) > 0$, we have $\mathbb{E}\mathsf{D}(Y) < 1$ and since D is boolean, the support of Y contains all the elements x satisfying $\mathsf{D}(x) = 1$. Therefore we obtain $\mathbb{E}\mathsf{D}(Y) = 2^{-k}|D|$. Now we can estimate the predicting probability from below as follows:

$$\Pr[X \text{ is predicted correctly}] = \frac{\mathbb{E}\mathsf{D}(X)}{|D|} \geqslant \frac{\mathbb{E}\mathsf{D}(Y)+\epsilon}{|D|} = 2^{-k} + \frac{\epsilon}{|D|}$$

The above probability holds for $\ell = \infty$, i.e., when predictor never outputs $\perp$. For efficiency reasons, we must use a finite, and not too big ℓ. The predictor will output $\perp$ with probability $(1 - 2^{-n}|D|)^\ell$ and thus

$$\Pr[\text{we predcit } X \text{ in time } \mathcal{O}(\ell \cdot \text{time}(D))] = \left(2^{-k} + \frac{\epsilon}{|D|}\right)\left(1 - \left(1 - 2^{-n}|D|\right)^\ell\right)$$

With a little bit of effort one can prove that setting $\ell = 1 + 2^{n-k}/\epsilon \approx 2^{\Delta}/\epsilon$ yields the success probability 2^{-k} independently of $|D|$.

Proof in general case - important issues Unfortunately, what we have proven above cannot be generalized easily to the case considered in the theorem, there are two obstacles. First, in the theorem we consider a conditional distribution $X|Z$ (i.e., the conditional part Z is not empty as above). Unfortunately we cannot simply make the above argument separately for all possible choices $Z = z$ of the conditional part, as we cannot guarantee that the conditional advantages $\epsilon(z) = \mathbb{E}\mathsf{D}(X|Z=z,z) - \mathbb{E}\mathsf{D}(Y|Z=z,z)$ are *all* positive; we only know that their average $\epsilon = \mathbb{E}_{z\leftarrow Z}\epsilon(z)$ is positive. Second, so far we assumed that D is boolean. This would only prove the theorem where the derived entropy in (7) is against deterministic *boolean* distinguishers, and this is not enough to conclude that we have the same amount of HILL entropy as discussed in Section 2.

Actual proof - preliminaries For real-valued distinguishers in the conditional case, just invoking Predictor(Z,D,ℓ) on a D satisfying (19), will not give a

predictor for X with advantage $> 2^{-k}$ in general. Instead, we first have to transform D into a new distingusiher D' that has the same distinguishing advantage, and for which we can prove that the predictor will work.

The way in which we modify D depends on the distribution $Y|Z$ that minimizes the left-hand side of (19). This distribution can be characterized as follows:

Lemma 2 ([Sko15]). *Given* $\mathsf{D} : \{0,1\}^n \times \{0,1\}^m \to [0,1]$ *and a distribution* $Z \in \{0,1\}^m$ *consider the following optimization problem*

$$\begin{aligned} &\max_{Y|Z} \mathbb{E}\mathsf{D}(Y,Z) \\ &\text{s.t. } \widetilde{H}_\infty(Y|Z) \geqslant k \end{aligned} \tag{20}$$

The distribution $Y|Z = Y^*|Z$ *satisfying* $\widetilde{H}_\infty(Y^*|Z) = k$ *is optimal for (20) iff there exist real numbers* $t(z)$ *and a number* $\lambda \geqslant 0$ *such that for every* z

(a) $\sum_x \max(\mathsf{D}(x,z) - t(z), 0) = \lambda$
(b) If $0 < \mathbf{P}_{Y^*|Z=z}(x) < \max_{x'} \mathbf{P}_{Y^*|Z=z}(x')$ *then* $\mathsf{D}(x,z) = t(z)$.
(c) If $\mathbf{P}_{Y^*|Z=z}(x) = 0$ *then* $\mathsf{D}(x,z) \leqslant t(z)$
(d) If $\mathbf{P}_{Y^*|Z=z}(x) = \max_{x'} \mathbf{P}_{Y^*|Z=z}(x')$ *then* $\mathsf{D}(x,z) \geqslant t(z)$

Proof. The proof is a straightforward application of the Kuhn-Tucker conditions given in Appendix. □

Remark 2. The characterization can be illustrated in an easy and elegant way. First, it says that the area under the graph of $\mathsf{D}(x,z)$ and above the threshold $t(z)$ is the same, no matter what z is (see Figure 3).

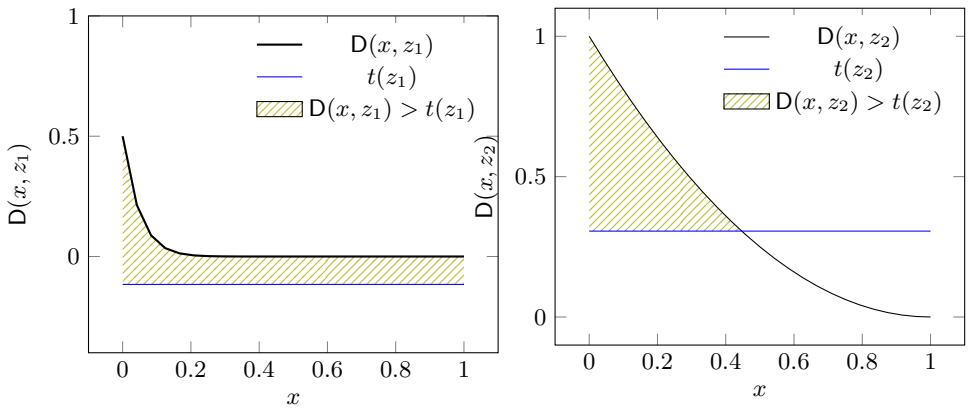


Fig. 3. For every z, the (green) area under $\mathsf{D}(\cdot,z)$ and above $t(z)$ equals λ

Second, for every z the distribution $Y^*|Z = z$ is flat over the set $\{x : \mathsf{D}(x,z) > t(z)\}$ and vanishes for x satisfying $\mathsf{D}(x,z) < t(z)$, see Fig. 4.

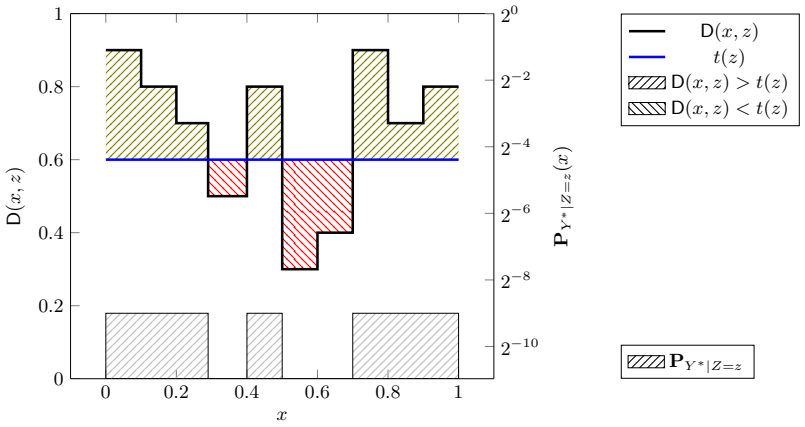


Fig. 4. Relation between distinguisher $\mathsf{D}(x,z)$, threshold $t(z)$ and distribution $Y^*|Z=z$.

Note that because of "freedom" in defining the distribution on elements x satisfying $\mathsf{D}(x,z)=t(z)$ (2, point (b)), there could be many distributions $Y^*|Z$ corresponding to fixed numbers λ and $t(z)$ that satisfy the characterization above, and this way are optimal to (20) with $k=\widetilde{H}_\infty(Y^*|Z)$. For the sake of completeness we characterize bellow the all possible values of k that match to λ and $t(z)$. We note that this fact might be used to modify our nonuniform guessing algorithm into a uniform one.

Corollary 1. *Let* $\mathsf{D}:\{0,1\}^n\times\{0,1\}^m\to[0,1]$ *and* $\lambda\in(0,1)$. *Let* $t(z)=t(\lambda,z)$ *be the unique numbers that satisfy the condition (a) in Lemma 2. Define*

$$k(\lambda)=n-\log\left(\mathbb{E}_{z\leftarrow Z}\left[1/\mathbf{P}(\mathsf{D}(U,z)\geqslant t(z))\right]\right), \tag{21}$$

which is a non-decreasing right-continuous function of λ. *Let also* $k^-(\lambda)=\lim_{\lambda'\to\lambda^-}k(\lambda')$ *and* $k^+(\lambda)=\lim_{\lambda'\to\lambda^+}k(\lambda')=k(\lambda)$ *be the one-sided limits. Then for every* $Y^*|Z$ *of min-entropy* $k=\widetilde{H}_\infty(Y^*|Z)$ *fulfilling (b),(c) and (d) we have* $k^-\leqslant k\leqslant k^+$. *Conversely, if* k *satisfies* $k^-\leqslant k\leqslant k^+$ *then there exists a distribution* $Y^*|Z$ *fulfilling (b),(c) and (d) such that* $\widetilde{H}_\infty(Y^*|Z)=k$.

Predicting given the thresholds $t(z)$. We use the numbers $t(z)$ to modify D and then we call the procedurePredictor on the modified distinguisher. Lemma 3 below shows that we could efficiently predict X from Z, assuming we knew the numbers $t(z)$ for all z in the support of Z (later, we'll show how to efficiently approximate them)

Lemma 3. *Let* $Y^*|Z$ *be the distribution satisfying* $\widetilde{H}_\infty(Y^*|Z)=k$ *and maximizing* $\mathbb{E}\mathsf{D}(Y,Z)$ *over* $\widetilde{H}_\infty(Y|Z)\geqslant k$, *where* $k<n$ *and* D *satisfies* (19). *Let* $t(z)$ *be as in Lemma 2. Define*

$$\mathsf{D}'(x,z)=\max(\mathsf{D}(x,z)-t(z),0) \tag{22}$$

and set $\ell = 2 \cdot 2^{n-k}\epsilon^{-1}$ *in the algorithm* Predictor. *Then we have*

$$\Pr\left(\textsc{Predictor}(Z, \mathsf{D}', \ell) = X\right) \geqslant 2^{-k}\left(1 + 2^{k-n}\epsilon\right) \tag{23}$$

Proof. We start by calculating the probability on the left-hand side of(23)

Claim 1 *For any*[20] D'*, the algorithm* Predictor *outputs* X *given* $Z = z$ *with probability*

$$\Pr_{X,Z}\left(\textsc{Predictor}(Z, \mathsf{D}', \ell) = X \middle| \, Z = z\right) = 2^{-n-1} g\left(\frac{\mathbb{E}\mathsf{D}'(U, z)}{2}\right) \cdot \mathbb{E}\mathsf{D}'(X | Z = z, z) \tag{24}$$

where U *is uniform over* $\{0,1\}^n$ *and* g *is defined by* $g(d) = \frac{1-(1-d)^\ell}{d}$ *(so* $g(d) \approx 1/d$ *for large* ℓ*)*

Proof (of Claim). It is easy to observe that

$$\Pr[\textsc{Predictor}(z, \mathsf{D}', \ell) = x | \, \textsc{Predictor}(z, \mathsf{D}', \ell) \neq \perp] = \frac{\mathsf{D}'(x, z)}{\sum\limits_x \mathsf{D}'(x, z)} \tag{25}$$

In turn, for every round $i = 1, \ldots, \ell$ of the execution, the probability that Predictor stops and outputs x' is equal to $\Pr[U = x']\mathsf{D}'(x', z)/2 = 2^{-n-1}\mathsf{D}'(x', z)$, the probability that it outputs anything (and thus leaves the while loop) is thus $\sum_{x'} \Pr[U = x'] \cdot \left(1 - \frac{\mathsf{D}'(x', z)}{2}\right) = 1 - \frac{\mathbb{E}\mathsf{D}'(U, z)}{2}$. So the probability of not leaving the while loop for ℓ rounds (in this case the output is $\perp$) is

$$\Pr[\textsc{Predictor}(z, \mathsf{D}', \ell) = \perp] = 1 - \left(1 - \frac{\mathbb{E}\mathsf{D}'(U, z)}{2}\right)^\ell \tag{26}$$

Combining the last two formulas we obtain

$$\Pr[\textsc{Predictor}(z, \mathsf{D}') = x] = 2^{-n-1} g(\mathbb{E}\mathsf{D}'(U, z)/2) \cdot \mathsf{D}'(x, z) \tag{27}$$

Hence

$$\begin{aligned}
\Pr[\textsc{Predictor}(z, \mathsf{D}') = X | Z = z] &= \sum_x \Pr[\textsc{Predictor}(z, \mathsf{D}') = x, X = x | Z = z] \\
&= \sum_x \Pr[\textsc{Predictor}(z, \mathsf{D}') = x] \Pr[X = x | Z = z] \\
&= 2^{-n-1} g(\mathbb{E}\mathsf{D}'(U, z)/2) \sum_x \mathsf{D}'(x, z) \Pr[X = x | Z = z] \\
&= 2^{-n-1} g(\mathbb{E}\mathsf{D}'(U, z)/2) \mathbb{E}\mathsf{D}'(X | Z = z, z)
\end{aligned} \tag{28}$$

and the claim follows. □

[20] We will only use the claim for the distinguisher D' as constructed above, but the claim holds in general.

Now we can see why we cannot apply the algorithm Predictor using the distinguisher D satisfying only (19) directly. According to the last formula, the success probability would be an averaged sum of products $g(\mathbb{E}\mathsf{D}(U,z))\cdot\mathbb{E}\mathsf{D}(X|Z=z,z)$ over z. We know the average of the second factors of these products, but in general cannot compare the values of $\mathbb{E}\mathsf{D}(U,z)$ for different z's. The crucial observation is that the distinguisher D' we defined satisfies the same inequality (19) as D (though, D' has the range $[0,2]$ not $[0,1]$ as D). Moreover D' has a special form which allows us to simplify expression (23). The details are given in the next two claims

Claim 2 *We have* $\mathbb{E}\mathsf{D}'(X,Z)-\mathbb{E}\mathsf{D}'(Y,Z)\geqslant\epsilon$ *for all* $Y|Z:\ \widetilde{H}_\infty(Y|Z)\geqslant k$

Proof (of Claim). We argue that (a): $\mathbb{E}\mathsf{D}'(X,Z)-\mathbb{E}\mathsf{D}'(Y^*,Z)\geqslant\mathbb{E}\mathsf{D}(X,Z)-\mathbb{E}\mathsf{D}(Y^*,Z)$ and (b): $Y^*|Z$ maximizes $\mathsf{D}'(Y,Z)$ over $\widetilde{H}_\infty(Y|Z)\geqslant k$. For the proof of (a), observe that by (22) we have $\mathsf{D}'(x,z)\geqslant\mathsf{D}(x,z)-t(z)$ for every x and z. Hence $\mathbb{E}\mathsf{D}'(X,Z)\geqslant\mathbb{E}\mathsf{D}(X,Z)-t(z)$. Moreover, if $\mathsf{D}(x,z)-t(z)<0$ then Lemma 2 implies $\mathbf{P}_{Y^*|Z=z}(x)=0$ and thus $\mathbb{E}\mathsf{D}'(Y^*|Z=z,z)=\mathbb{E}\mathsf{D}(Y^*|Z=z)-t(z)$. Hence, for all z we have

$$\mathbb{E}\mathsf{D}'(X|Z=z)-\mathbb{E}\mathsf{D}'(Y^*|Z=z,z)\geqslant\mathbb{E}\mathsf{D}(X|Z=z,z)-\mathbb{E}\mathsf{D}(Y^*|Z=z,z)$$

The proof of (a) follows now by taking the average over z. The proof of (b) follows by observing that D' satisfies the characterization in (2) with $t(z)=0$ for all z. □

Claim 3 *The exists a number* $\lambda'\in(0,1)$ *such that* $\mathbb{E}\mathsf{D}'(U,z)=\lambda'$ *for every* z.

Proof. Lemma 2 implies $\sum_x\mathsf{D}'(x,z)=\lambda$ for every z. We can define $\lambda'=2^{-n}\lambda$ and then it remains to show $\lambda<2^n$ and $\lambda>0$. Observe that the case $t(z)<0$ in Lemma 2 is possible if and only if $\mathbf{P}_{Y^*|Z=z}(x)=\max_{x'}\mathbf{P}_{Y^*|Z=z}(x')$ for all x, which means $H_\infty(Y^*|Z=z)=n$. Since $k<n$, we have $t(z)\geqslant 0$ for at least one z and then $\lambda=\sum_x\max(\mathsf{D}(x,z)-t(z),0)\leqslant\sum_x\mathsf{D}(x,z)$ which essentially means $\lambda\leqslant 2^n$. Lemma 2 guarantees that $\lambda\geqslant 0$, therefore we need to show that $\lambda\notin\{0,2^n\}$. Observe that if $\lambda=0$ then the condition $\sum_x\mathsf{D}'(x,z)=\lambda$ implies $\mathsf{D}'(x,z)=0$ for all x and z, contradicting to Claim 2 because $\epsilon>0$. In turn, if $\lambda=2^n$ then from Lemma 2 we get $\mathsf{D}(\cdot,z)\equiv 1$ and $t(z)=0$ for all z such that $t(z)\geqslant 0$. This is possible only if $\mathbf{P}_{Y^*|Z=z}(x)=\max_{x'}\mathbf{P}_{Y^*|Z=z}(x')$ for all x which means $H_\infty(Y^*|Z=z)=n$ if $t(z)\geqslant 0$. But then $H_\infty(Y^*|Z=z)=n$ for all z which contradicts $k<n$. □

To calculate the success probability we need one more observation. The following claim shows that support of D' is contained in the support of Y^*.

Claim 4 *For every* z *we have*

$$\mathbb{E}\mathsf{D}'(Y^*|Z=z,z)=\mathbb{E}\mathsf{D}'(U,z)\cdot 2^n\max_{x'}\mathbf{P}_{Y^*|Z=z}(x'). \tag{29}$$

Proof (of Claim). By Lemma 2 we obtain that $\mathsf{D}(x,z) > t(z)$ only if $\mathbf{P}_{Y^*|Z=z}(x) = \max_{x'} \mathbf{P}_{Y^*|Z=z}(x')$ therefore

$$\begin{aligned}\mathbb{E}\mathsf{D}'(Y^*|Z=z,z) &= \sum_x \max(\mathsf{D}(x,z)-t(z),0)\mathbf{P}_{Y^*|Z=z}(x)\\ &= \sum_x \max(\mathsf{D}(x,z)-t(z),0)\max_{x'}\mathbf{P}_{Y^*|Z=z}(x'),\end{aligned}$$

and the claim follows by the definition of D'. □

Now we are ready to prove the main result. From Claim 1 and Claim 3 we obtain

$$\begin{aligned}\Pr\left(\textsc{Predictor}(Z,\mathsf{D}',\ell)=X\right) &= 2^{-n-1}\mathbb{E}_{z\leftarrow Z}\left[g(\lambda'/2)\cdot\mathsf{D}'(X|Z=z,z)\right]\\ &= 2^{-n-1}g(\lambda'/2)\cdot\mathbb{E}\mathsf{D}'(X,Z) \qquad (30)\end{aligned}$$

Claim 2 applied to $Y = Y^*$ yields now the following estimate

$$\Pr\left(\textsc{Predictor}(Z,\mathsf{D}',\ell)=X\right) \geqslant 2^{-n-1}g(\lambda'/2)\cdot\left(\mathbb{E}\mathsf{D}'(Y^*,Z)+\epsilon\right). \qquad (31)$$

Observe that Claim 4, Claim 3, and $\widetilde{H}_\infty(Y^*|Z) = k$ imply

$$\begin{aligned}\mathbb{E}\mathsf{D}'(Y^*,Z) &= \mathbb{E}_{z\leftarrow Z}\left[\mathsf{D}'(Y^*|Z=z,z)\right] = \mathbb{E}_{z\leftarrow Z}\left[\mathbb{E}\mathsf{D}'(U,z)\cdot 2^n\max_{x'}\mathbf{P}_{Y^*|Z=z}(x')\right]\\ &= 2^n\lambda'\cdot\mathbb{E}_{z\leftarrow Z}\left[\max_{x'}\mathbf{P}_{Y^*|Z=z}(x')\right] = 2^{n-k}\lambda' \qquad (32)\end{aligned}$$

Plugging this into (31) we get the following bound

$$\begin{aligned}\Pr\left(\textsc{Predictor}(Z,\mathsf{D}',\ell)=X\right) &\geqslant 2^{-n-1}g(\lambda'/2)\cdot\left(2^{n-k}\lambda'+\epsilon\right)\\ &= 2^{-k}\left(1-(1-\lambda'/2)^\ell\right)\left(1+\frac{2^{k-n-1}\epsilon}{\lambda'/2}\right) \qquad (33)\end{aligned}$$

To give a lower bound on the success probability it remains to minimize the last expression over $\lambda' \in (0,1)$. This is answered below

Claim 5 *Let $h(s) = (1-(1-s)^\ell)(1+as^{-1})$, where $a > 0$ and $\ell \geqslant 1 + a^{-1}$. Then $h(s) \geqslant h(1) = 1 + a$ for all $s \in [0,1]$.*

Proof (of Claim). The proof uses standard calculus and is given in the appendix. □

Computing $t(z)$ from λ So far, we have shown how to construct the predicting algorithm provided that we are given the numbers $t(z)$. Now we will prove that one can compute them *approximately* and use *successfully* in place of the original ones. We start with a few useful facts about the auxiliary function g already introduced in Claim 1 in the proof of Lemma 3. Below we summarize its fundamental properties.

Lemma 4. *For $\ell > 1$ the function $g(d) = \frac{1-(1-d)^\ell}{d}$ on $[0,1]$ satisfies:*

(a) g is continuous at 0 and decreasing
(b) g is convex
(c) for any $d_2 > d_1$ we have $g(d_2) > g(d_1)\left(1 - \frac{\ell}{2} \cdot |d_2 - d_1|\right)$

Proof (of Lemma). The proof uses elementary calculus and is referred to the appendix □

The entire solution is based on the next two lemmas. The first lemma is based on the intuition that replacing D by a distinguisher which approximates it close enough should not affect the success probability of $\textsc{Predictor}(Z, \mathsf{D}, \ell)$ very much. For technical reasons we present this statement assuming *one-sided $\mathcal{L}^1$-approximation*. The second lemma describes an efficient algorithm which obtains λ as a *hint* on its input and computes approximations for $t(z)$ from below, for every z.

Lemma 5. *Let $\mathsf{D}_1, \mathsf{D}_2 : \{0,1\}^n \times \{0,1\}^m \to [0,1]$ be any two functions satisfying*

(a) $\mathsf{D}_2(x,z) \geqslant \mathsf{D}_1(x,z)$ for all x, z
(b) $\mathbb{E}\mathsf{D}_2(U,z) - \mathbb{E}\mathsf{D}_1(U,z) \leqslant \delta$ for all z

Then we have

$$\Pr\left(\textsc{Predictor}(Z, \mathsf{D}_2, \ell) = X\right) \geqslant (1 - \ell\delta/2) \Pr\left(\textsc{Predictor}(Z, \mathsf{D}_1, \ell) = X\right) \tag{34}$$

Proof (of Lemma). We have

$$\begin{aligned}\Pr\left(\textsc{Predictor}(z, \mathsf{D}_2, \ell) = X | Z = z\right) &= g(\mathbb{E}\mathsf{D}_2(U,z))\mathbb{E}\mathsf{D}_2(X|Z=z,z) \\ &\geqslant g(\mathbb{E}\mathsf{D}_2(U,z))\mathbb{E}\mathsf{D}_1(X|Z=z,z), \end{aligned} \tag{35}$$

where the inequality follows from $\mathsf{D}_2 \geqslant \mathsf{D}_1 \geqslant 0$. The assumptions (a) and (b) imply $|\mathbb{E}\mathsf{D}_1(U,z) - \mathbb{E}\mathsf{D}_2(U,z)| \leqslant \delta$ for every z. From property (c) in Lemma 4 it follows that

$$g(\mathbb{E}\mathsf{D}_2(U,z)) \geqslant g(\mathbb{E}\mathsf{D}_1(U,z))(1 - \ell\delta/2)$$

for every z. Combining the last two estimates we get

$$\begin{aligned}&\Pr\left(\textsc{Predictor}(z, \mathsf{D}_2, \ell) = X | Z = z\right) \\ &\qquad \geqslant (1 - \ell\delta/2) \cdot g(\mathbb{E}\mathsf{D}_1(U,z))\mathbb{E}\mathsf{D}_1(X|Z=z,z) \\ &\qquad = (1 - \ell\delta/2) \cdot \Pr\left(\textsc{Predictor}(z, \mathsf{D}_1) = X | Z = z\right) \end{aligned} \tag{36}$$

Taking the average over $z \leftarrow Z$ completes the proof. □

Lemma 6. *Let $\mathsf{D} : \{0,1\}^n \to [0,1]$ be any function computable in time s, let $\lambda \in (0,1)$ and $t \in [0,1]$ be a number such that $\mathbb{E}\max(\mathsf{D}(U) - t, 0) = \lambda$. There exists a probabilistic algorithm $\textsc{FindThreshold}(\mathsf{D}, \lambda, \delta, N)$ that runs in time $\mathcal{O}\left(\log(1/\delta)N \cdot \mathrm{time}(\mathsf{D})\right)$ and with probability at least $1 - 2\log(12/\delta)\mathrm{e}^{-N\delta^2/3}$ outputs a number t' such that $\mathbb{E}\max(\mathsf{D}(U) - t', 0) \in [\lambda, \lambda + \delta]$. In particular, $t' \leqslant t$.*

Function FINDTHRESHOLD$(\mathsf{D}, \lambda, \delta, N)$

Input : $\mathsf{D} : \{0,1\}^n \to [0,1]$, $\lambda \in (0,1)$, parameters δ, N
Output: t' such that $\mathbb{E}\max(\mathsf{D}(U) - t', 0) \in [\lambda, \lambda + \delta]$

$t^- \leftarrow -1, t^+ \leftarrow 1$
repeat
  $t' \leftarrow (t^- + t^+)/2$
  $x_1, \ldots, x_N \leftarrow U$ `/* fresh values every time */`
  $\lambda' \leftarrow N^{-1} \sum_{j=1}^{N} \max(\mathsf{D}(x_j) - t', 0)$ `/*` $\lambda' \approx \mathbb{E}\max(\mathsf{D}(U) - t_i, 0)$ `*/`
  if $\lambda' > \lambda + \frac{2\delta}{3}$ **then**
    $t^- \leftarrow t'$
  else if $\lambda' < \lambda + \frac{\delta}{3}$ **then**
    $t^+ \leftarrow t'$
  else
    return t'
  end
until $t^+ - t^- \leqslant \frac{\delta}{12}$
if $t' < -1 + \frac{\delta}{12}$ **then**
  $t' \leftarrow -1$
return t'

Proof (of Lemma). The idea is pretty simple: given t' we approximate values $\mathbb{E}\max(\mathsf{D}(U) - t', 0)$ by sampling and by comparing the result with λ, we can find the right value of t' using binary search. This corresponds to finding a blue line on Fig. 4 such that the green area above is sufficiently close to λ.
The function $h(t') = \mathbb{E}\max(\mathsf{D}(U) - t', 0)$ is clearly non-increasing with respect to t' and changes from $1 + \mathbb{E}\mathsf{D}(U)$ at $t' = -1$ to 0 for $t = 1$. Moreover, it is strictly decreasing in a small neighborhood of $t' = t$ and for all $t' < t$. Indeed, since $\lambda > 0$ there is at least one x such that $\mathsf{D}(x) > t$. Taking $t' < t'' \leqslant \min_{x:\mathsf{D}(x)>t} \mathsf{D}(x)$ we see that $h(t') - h(t'') \geqslant 2^{-n}(t'' - t') > 0$. Hence, $t' > t$ implies $\mathbb{E}\max(\mathsf{D}(U) - t', 0) < \mathbb{E}\max(\mathsf{D}(U) - t, 0) = \lambda$. This proves the second part of the statement. Denote by $\lambda'_i, t'_i, t^-_i, t^+_i$ the values assigned in round i to λ', t', t^-, t^+ respectively. Observe that by the Chernoff Bound[21] and the union bound over at most $\log(12/\delta)$ rounds of the execution, with probability $p = 1 - 2\log(12/\delta)\exp(-N\delta^2/3)$ we have $|\lambda'_i - h(t_i)| < \frac{\delta}{12}$ for every round i. Note that with the same probability the algorithm satisfies the invariant property: if there is $t_0 \in [t^-_i, t^+_i]$ such that $h(t_0) \in \left[\lambda + \frac{5\delta}{12}, \lambda + \frac{7\delta}{12}\right]$ and the algorithm jumps to round $i+1$ then $t_0 \in \left[t^-_{i+1}, t^+_{i+1}\right]$. Suppose that $h(t_0) \in \left[\lambda + \frac{5\delta}{12}, \lambda + \frac{7\delta}{12}\right]$ for some $t_0 \in [-1, 1]$. Now we have two possibilities: either we terminate with t_i such that $\lambda_i \in \left[\lambda + \frac{\delta}{3}, \lambda + \frac{2\delta}{3}\right]$ which means $h(t_i) \in \left[\lambda + \frac{3\delta}{12}, \lambda + \frac{7\delta}{12}\right]$ and we are done, or we will eventually find such t_0 up to an error $\frac{\delta}{12}$. Since $|h(t_2) - h(t_1)| \leqslant |t_2 - t_1|$ for any t_1, t_2, the returned number t' satisfies $h(t_0) - \frac{\delta}{12} \leqslant h(t') \leqslant h(t_0) + \frac{\delta}{12}$, in particular it satisfies the desired inequality. It remains to consider the case

[21] We use the following version: let $X_1, \ldots, X_N$ be $[0,1]$-valued independent random variables, let $X = \sum_{i=1}^{N} X_i$ and $\mu = \mathbb{E}X$. Then $\Pr(|X - \mu| > \delta\mu) < 2\exp(-\mu\delta^2/3)$

when either $h(t) < \lambda + \frac{5\delta}{12}$ for all t or $h(t) > \lambda + \frac{7\delta}{12}$. Since $h(1) = 0$ the second is clearly impossible. In the first case we have $h(t) \leqslant h(-1) < \lambda + \frac{5\delta}{12}$, which means that in every round i we have $t_i^- = -1$ and either we terminate with t_i such that $\lambda_i' \in \left[\lambda + \frac{\delta}{3}, \lambda + \frac{2\delta}{3}\right]$ which means $h(t_i) \in \left[\lambda + \frac{3\delta}{12}, \lambda + \frac{7\delta}{12}\right]$ and we are done, or in every round i we do the assignment $t_{i+1}^+ = t_i$ which yields $t_i = -1 + 2^{-i+1}$ and the main loop halts with $t_i < -1 + \frac{\delta}{12}$. The algorithm outputs then -1 which satisfies the desired inequality, because of the assumption $h(-1) < \lambda + \frac{5\delta}{12}$ and the trivial inequality $h(-1) \geqslant 1 \geqslant \lambda$. □

Let D' be as in Lemma 3. Let $t'(z) = \textsc{FindThreshold}(\mathsf{D}, \lambda, \delta, N)$, define $D''(x, z) = \max(\mathsf{D}(U, z) - t'(z), 0)$. Let $\Pr[bad]$ be the probability of the event $\mathbb{E}D''(U, z) \notin [\lambda, \lambda + \delta]$ (i.e. failure of the algorithm FindThreshold). If the event *bad* doesn't occur then $\mathsf{D}'' \geqslant \mathsf{D}'$ and $\mathbb{E}\mathsf{D}''(U, z) \leqslant \mathbb{E}\mathsf{D}'(U, z) + \delta$. Applying the last two claims we obtain

$$\Pr\left[\textsc{Predictor}(z, \mathsf{D}'', \ell)\right] \geqslant 2^{-k}\left(1 + 2^{k-n}\epsilon\right) \cdot \left(1 - \frac{\ell\delta}{2}\right) \Pr[\neg bad] \tag{37}$$

By the elementary inequality $(1 + s)(1 - s/4)^2 \geqslant 1$ valid for $s \in [0, 1]$, for this probability to be bigger than 2^{-k} it is enough to require

$$\ell\delta/2 \leqslant 2^{k-n}\epsilon/4 \tag{38}$$

$$2\log(12/\delta)\exp(-N\delta^2)/3) \leqslant 2^{k-n}\epsilon/4 \tag{39}$$

The solution for the first inequality is $\delta = \mathcal{O}(2^{2(k-n)}\epsilon^2)$ which implies $\delta \ll \epsilon$. The second one gives us $N = \Omega\left((1/\delta)^2(\log\log(1/\delta) + n - k + \log(1/\epsilon)\right)$ which can be simplified to $N = \Omega\left((1/\delta)^2(\log(1/\delta)\right)$. The total running time equals (up to a constant factor) the time needed for invoking $\mathcal{O}\left(\ell \cdot N\log(1/\delta)\right) = \mathcal{O}\left((2^{\Delta}/\epsilon)^5 \log^2\left(2^{\Delta}/\epsilon\right)\right)$ times of the distinguisher D .

D Proof of Lemma 2

Proof. Without losing generality we assume that $\mathbf{P}(z) > 0$ for all z (as for (20) only the support of Z is relevant). Consider the following linear optimization program

$$\begin{aligned}
\underset{P_{x,z}, a_z}{\text{maximize}} \quad & \sum_{x,z} \mathsf{D}(x, z) P_{x,z} \\
\text{subject to} \quad & -P_{x,z} \leqslant 0, \ \forall (x, z) \in \{0,1\}^n \times \{0,1\}^m \\
& \sum_x P_{x,z} - \mathbf{P}_Z(z) = 0, \ \forall z \in \{0,1\}^m \\
& P_{x,z} - a_z \leqslant 0, \ z \in \{0,1\}^m \\
& \sum_z a_z - 2^{-k} \leqslant 0
\end{aligned} \tag{40}$$

This problem is equivalent to (20) if we define $\mathbf{P}_{Y,Z}(x,z) = P_{x,z}$ and replace the condition $\sum_z \max_x \mathbf{P}_{Y,Z}(x,z) \leqslant 2^{-k}$ which is equivalent to $\widetilde{H}_\infty(Y|Z) \geqslant k$, by the existence of numbers $\max_x \mathbf{P}_{Y,Z}(x,z) \leqslant a_z$ such that $\sum_z a_z \leqslant 2^{-k}$. The solutions of (40) can be characterized as follows:

Claim 6 *The numbers $(P_{x,z})_{x,z}, (a_z)_z$ are optimal for (40) if and only if there exist numbers $\lambda^1(x,z) \geqslant 0$, $\lambda^2(z) \in \mathbb{R}$, $\lambda^3(x,z) \geqslant 0$, $\lambda^4 \geqslant 0$ such that*

(a) $\mathsf{D}(x,z) = -\lambda^1(x,z) + \lambda^2(z) + \lambda^3(x,z)$ and $0 = -\sum_x \lambda^3(x,z) + \lambda^4$
(b) We have $\lambda^1(x,z) = 0$ if $P_{x,z} > 0$, $\lambda^3(x,z) = 0$ if $P_{x,z} < a_z$, $\lambda^4 = 0$ if $\sum_z a_z < 2^{-k}$.

Proof (of Claim). This is a straightforward application of KKT conditions. □

It remains to apply and simplify the last characterization. Let $(P^*_{x,z})_{x,z}$ and $(a^*_z)_z$ be optimal for (40), where $P^*_{x,z} = \mathbf{P}_{Y^*,Z}(x,z)$ and $\lambda^1(x,z), \lambda^2(z), \lambda^3(x,z), \lambda^4(x)$ are the corresponding multipliers given by the last claim. Define $t(z) = \lambda^2(z)$ and $\lambda = \lambda^4$. Observe that for every z we have $a^*_z \geqslant \max_x P^*(x,z) \geqslant 2^{-n}\mathbf{P}_Z(z) > 0$ and thus for every (x,z) we have

$$\lambda^1(x,z) \cdot \lambda^3(x,z) = 0 \tag{41}$$

If $P^*_{x,z} = 0$ then in particular $P^*_{x,z} < a^*(z)$ and $\lambda^3(x,z) = 0$, hence $\mathsf{D}(x,z) \leqslant t(z)$ which proves (c). If $P^*_{x,z} = \max_{x'} P^*_{x,z}$ then $P^*_{x,z} < 0$ and $\lambda^1(x,z) = 0$ which proves (d). Finally observe that (41) implies

$$\max(\mathsf{D}(x,z) - t(z), 0) = \max(-\lambda^1(x,z) + \lambda^3(x,z), 0) = \lambda^3(x,z)$$

Hence, the assumption $\sum_x \lambda^3(x,z) = \lambda^4 = \lambda$ proves (a).

Suppose now that the characterization given in the Lemma is satisfied. Define $P^*_{x,z} = \mathbf{P}_{Y,Z}(x,z)$ and $a_z = \max_z \mathbf{P}_{Y^*,Z}(x,z)$, let $\lambda^3(x,z) = \max(\mathsf{D}(x,z) - t(z), 0)$, $\lambda^1(x,z) = \max(t(z) - \mathsf{D}(x,z), 0)$ and $\lambda^4 = \lambda$. We will show that these numbers satisfy the conditions described in the last claim. By definition we have $-\lambda^1(x,z) + \lambda^2(z) + \lambda^3(x,z) = \mathsf{D}(x,z)$, by the assumptions we get $\sum_x \lambda^3(x,z) = \lambda = \lambda^4$. This proves part (a). Now we verify the conditions in (b). Note that $\mathsf{D}(x,z) < t(z)$ is possible only if $\mathbf{P}_{Y^*|Z=z}(x) = 0$ and $\mathsf{D}(x,z) > t(z)$ is possible only if $\mathbf{P}_{Y^*|Z=z}(x) = \max_{x'} \mathbf{P}_{Y^*|Z=z}(x')$. Therefore, if $\mathbf{P}_{Y,Z}(x,z) > 0$ then we must have $\mathsf{D}(x,z) \geqslant t(z)$ which means that $\lambda^1(x,z) = 0$. Similarly if $\mathbf{P}_{Y,Z}(x,z) < \max_z \mathbf{P}_{Y^*,Z}(x,z)$ then $\mathsf{D}(x,z) \leqslant t(z)$ and $\lambda^3(x,z) = 0$. Finally, since we assume $\widetilde{H}_\infty(Y^*|Z) = k$ we have $\sum_z a_z = 2^{-k}$ and thus there is no additional restrictions on λ^4. □

E Proof of Corollary 1

Proof (of Corollary). Let $y_{\max}(z) = \max_{x'} \mathbf{P}_{Y|Z=z}(x')$. Consider the function

$$f^\delta_z(x) = \begin{cases} y_{\max}(z) + \delta, & \mathsf{D}'(x,z) > t(z) \\ \frac{1 - \#\{x:\ \mathsf{D}'(x.z) > t(z)\} \cdot (y_{\max} + \delta)}{\#\{x:\ \mathsf{D}'(x,z) = t(z)\}}, & \mathsf{D}'(x,z) = t(z) \\ 0, & \mathsf{D}'(x,z) < t(z) \end{cases} \tag{42}$$

This function defines a distribution that satisfies

$$f_z^\delta(x) \leqslant \max_{x'} f_z^\delta(x') \quad \forall x : \mathsf{D}'(x,z) \leqslant t(z) \tag{43}$$

if and only if δ satisfies

$$\frac{1}{\#\left\{x : \ \mathsf{D}'(x.z) \geqslant t(z)\right\}} \leqslant y_{\max}(z) + \delta \leqslant \frac{1}{\#\left\{x : \ \mathsf{D}'(x.z) > t(z)\right\}} \tag{44}$$

In particular these conditions are satisfied for $\delta = 0$. Suppose now that there are z_i and x_i for $i = 1, 2$ such that $0 < \mathbf{P}_{Y^*|Z=z_i}(x_i) < \max_{x'} \mathbf{P}_{Y^*|Z=z}(x')$. Define δ by

$$\delta = \min\left(y_{\max}(z_1) - \frac{1}{\#\left\{x : \mathsf{D}'(x,z_1) \geqslant t(z_1)\right\}}, \right.$$
$$\left. \frac{1}{\#\left\{x : \mathsf{D}'(x,z_2) > t(z_2)\right\}} - y_{\max}(z_2)\right)$$

By Lemma 2 we immediately obtain that $\delta \geqslant 0$. It follows easily from the definition of δ that the number $-\delta$ satisfies (44) with $z = z_1$ and that δ satisfies (44) for $z = z_2$. We can see now that if we replace the distribution $Y^*|Z = z_1$ by $f_{z_1}^{-\delta}$ and the distribution $Y^*|Z = z_2$ by $f_{z_2}^{\delta}$ then we obtain the distribution $Y'|Z$ satisfying conditions in Lemma 2 and $\widetilde{H}_\infty(Y'|Z) = k$. Finally, observe that $\delta = \frac{1}{\#\{x:\mathsf{D}'(x,z_2)>t(z_2)\}} - y_{\max}(z_2)$ means that the distribution $Y'|Z = z_2$ is uniform on $\{x : \mathsf{D}'(x.z_2) > t(z_2)\}$. In turn, if $\delta = y_{\max}(z_1) - \frac{1}{\#\{x:\mathsf{D}'(x,z_1)\geqslant t(z_1)\}}$ then the distribution $Y'|Z = z_1$ is uniform on $\{x : \mathsf{D}'(x,z_1) \geqslant t(z_1)\}$. □

F Proof of Claim 5, Lemma 3

Proof. We check that $\lim_{s\to 0} h(s) = a\ell$ and thus the function h is continuous on the interval $[0,1]$. This means that h attains its minimum at some point $s = s_0$. There is nothing to prove if $s_0 \in \{0,1\}$. Suppose that $s_0 \in (0,1)$. Then we must have $\frac{\partial h}{\partial s}\big|_{s=s_0} = 0$. The first derivative of the function h is given by the following formula

$$\begin{aligned}\frac{\partial h}{\partial s} &= \frac{s\ell(a+s)(1-s)^{\ell-1} + a\left((1-s)^\ell - 1\right)}{s^2} \\ &= \frac{-a + (1-s)^{\ell-1}\left(a(1-s) + (a+s)\ell s\right)}{s^2}\end{aligned} \tag{45}$$

Therefore for $s = s_0$ we obtain $(1-s_0)^{\ell-1} = \frac{a}{a(1-s_0)+(a+s_0)\ell s_0}$ and hence

$$\begin{aligned}h(s_0) &= \left(1 - (1-s_0)\cdot(1-s_0)^{\ell-1}\right)\left(1 + as_0^{-1}\right) \\ &= \frac{(a+s_0)^2\ell}{a(1-s_0) + (a+s_0)\ell s_0}\end{aligned} \tag{46}$$

Note that the last expression is increasing with respect to ℓ and that from the assumption we have $\ell > \frac{1+a}{a+s_0}$. Using this we obtain

$$h(s_0) \geqslant \frac{(a+s_0)(1+a)}{a(1-s_0)+(1+a)s_0} = 1+a \tag{47}$$

which completes the proof. □

The lemma follows now immediately by combining (33) and the last claim. □

G Proof of Lemma 4

Proof (of Lemma). It is easy to see that $\lim_{d\to 0^+} g(d) = \ell$. We have

$$\frac{\partial g(d)}{\partial d} = \frac{(1-d)^{\ell-1}(d(\ell-1)+1)-1}{d^2} \tag{48}$$

Using the inequality $1-d \leqslant \mathrm{e}^{-d}$ we obtain

$$\frac{\partial g(d)}{\partial d} \leqslant \frac{\mathrm{e}^{-d(\ell-1)}\,(d(\ell-1)+1)-1}{d^2} \leqslant 0$$

Where the second inequality follows from the inequality $\mathrm{e}^s \geqslant 1+s$ applied for $s = d(\ell-1)$. This proves (a). The second derivative is given by

$$\frac{\partial^2 g(d)}{\partial d^2} = -\frac{(1-d)^{\ell-2}\left(2+2d(\ell-2)+d^2((\ell-2)^2+\ell-2)\right)-2}{d^3} \tag{49}$$

Using $1-d \leqslant \mathrm{e}^{-d}$ and applying the inequality $\mathrm{e}^s \geqslant 1+s+\frac{1}{2}s^2$, which holds for $s \geqslant 0$, for $s = d(\ell-1)$ we obtain

$$\begin{aligned}
\frac{\partial^2 g(d)}{\partial d^2} &= -\frac{(1-d)^{\ell-2}\left(2+2d(\ell-2)+d^2((\ell-2)^2+\ell-2)\right)-2}{d^3} \\
&\geqslant -\frac{(1-d)^{\ell-1}\left(2+2d(\ell-1)+d^2(\ell-1)^2\right)-2}{d^3} \\
&\geqslant -\frac{\mathrm{e}^{-d(\ell-1)}\left(2+2d(\ell-1)+d^2(\ell-1)^2\right)-2}{d^3} \\
&\geqslant -\frac{2-2}{d^3} = 0,
\end{aligned} \tag{50}$$

which proves (b). Finally, note that by convexity we have

$$g(d_2)-g(d_1) \geqslant (d_2-d_1)\cdot \left.\frac{\partial g(d)}{\partial d}\right|_{d=d_1}. \tag{51}$$

Since $g(d) > 0$ and $\frac{\partial \ln g(d)}{\partial d} = \frac{\partial g(d)}{\partial d}/g(d)$ we can rewrite this as

$$\frac{g(d_2)-g(d_1)}{g(d_1)} \geqslant (d_2-d_1)\cdot \left.\frac{\partial \ln g(d)}{\partial d}\right|_{d=d_1}. \tag{52}$$

Note that the function $d \to \ln g(d)$ is convex, as the composition of the convex function $g(\cdot)$ and the convex increasing function $\ln(\cdot)$. Therefore,

$$\frac{\partial \ln g(d)}{\partial d} \geqslant \left.\frac{\partial \ln g(d)}{\partial d}\right|_{d=0} = -\frac{\ell - 1}{2} \tag{53}$$

Combining the last two inequalities yields

$$\frac{g(d_2) - g(d_1)}{g(d_1)} > -\frac{\ell}{2} \cdot (d_2 - d_1), \quad d_2 - d_1 > 0. \tag{54}$$

which completes the proof of (c). □

Chapter 3

Lower Bounds for Pseudoentropy Chain Rules and Transformations

Pseudoentropy: Lower-bounds for Chain rules and Transformations

Krzysztof Pietrzak* and Maciej Skórski**

IST Austria and University of Warsaw

Abstract. Computational notions of entropy have recently found many applications, including leakage-resilient cryptography, deterministic encryption or memory delegation. The two main types of results which make computational notions so useful are (1) Chain rules, which quantify by how much the computational entropy of a variable decreases if conditioned on some other variable (2) Transformations, which quantify to which extend one type of entropy implies another.

Such chain rules and transformations typically lose a significant amount in quality of the entropy, and are the reason why applying these results one gets rather weak quantitative security bounds. In this paper we for the first time prove lower bounds in this context, showing that existing results for transformations are, unfortunately, basically optimal for non-adaptive black-box reductions (and it's hard to imagine how non black-box reductions or adaptivity could be useful here).

A variable X has k bits of HILL entropy of quality (ϵ, s) if there exists a variable Y with k bits min-entropy which cannot be distinguished from X with advantage ϵ by distinguishing circuits of size s. A weaker notion is Metric entropy, where we switch quantifiers, and only require that for every distinguisher of size s, such a Y exists.

We first describe our result concerning transformations. By definition, HILL implies Metric without any loss in quality. Metric entropy often comes up in applications, but must be transformed to HILL for meaningful security guarantees. The best known result states that if a variable X has k bits of Metric entropy of quality (ϵ, s), then it has k bits of HILL with quality $(2\epsilon, s \cdot \epsilon^2)$. We show that this loss of a factor $\Omega(\epsilon^{-2})$ in circuit size is necessary. In fact, we show a stronger result that this loss is already necessary when transforming so called deterministic real valued Metric entropy to randomised boolean Metric (both these variants of Metric entropy are implied by HILL without loss in quality).

The chain rule for HILL entropy states that if X has k bits of HILL entropy of quality (ϵ, s), then for any variable Z of length m, X conditioned on Z has $k-m$ bits of HILL entropy with quality $(\epsilon, s \cdot \epsilon^2/2^m)$. We show that a loss of $\Omega(2^m/\epsilon)$ in circuit size necessary here. Note that this still leaves a gap of ϵ between the known bound and our lower bound.

As in related works on query complexity lower bounds for computational indistinguishability problems (Dense Model Theorems, Hardcore Lemmas), the formal proofs are restricted to reductions which query on same inputs. Overcomming this limitation seems challenging.

* Supported by the European Research Council consolidator grant (682815-TOCNeT).

** Supported by the National Science Center, Poland (2015/17/N/ST6/03564)

1 Introduction

There exist various information theoretic notions of entropy that quantify the "uncertainty" of a random variable. A variable X has k bits of Shannon entropy if it cannot be compressed below k bits. In cryptography we mostly consider min-entropy, where we say that X has k bits of min-entropy, denoted $\mathbf{H}_\infty(X) = k$, if for any x, $\Pr[X = x] \leq 2^{-k}$.

In a cryptographic context, we often have to deal with variables that only appear to have high entropy to computationally bounded observers. The most important case is pseudorandomness, where we say that $X \in \{0,1\}^n$ is pseudorandom, if it cannot be distinguished from the uniform distribution over $\{0,1\}^n$.

More generally, we say that $X \in \{0,1\}^n$ has $k \leq n$ bits of HILL pseudoentropy [ILL89, HILL99], denoted $\mathbf{H}^{\mathsf{HILL}}_{\epsilon,s}(X) = k$ if it cannot be distinguished from some Y with $\mathbf{H}_\infty(Y) = k$ by any circuit of size s with advantage $> \epsilon$, note that we get pseudorandomness as a special case for $k = n$. We refer to k as the *quantity* and to (ϵ, s) as the *quality* of the entropy.

A weak notion of pseudoentropy called Metric pseudoentropy [BSW03] often comes up in security proofs. This notion is defined like HILL, but with the quantifiers exchanged: We only require that for every distininguisher there exists a distribution $Y, \mathbf{H}_\infty(Y) = k$ that fools this particular distinguisher (not one such Y to fool them all).

HILL pseudoentropy is named after the authors of the [HILL99] paper where it was introduced as a tool for constructing a pseudorandom generator from any one-way function. Their construction and analysis was subsequently improved in a series of works [Hol06, HRV10, VZ12]. A lower bound on the number of calls to the underlying one-way function was given by [HS12].[1] More recently HILL pseudoentropy has been used in many other applications like leakage-resilient cryptography [NY19, JP14], deterministic encryption [FOR12] and memory delegation [CKLR11].

The two most important types of tools we have to manipulate pseudoentropy are chain rules and transformations from one notion into another. Unfortunately, the known transformations and chain rules lose large factors in the quality of the entropy, which results in poor quantitative security bounds that can be achieved using these tools. In this paper we provide lower bounds, showing that unfortunately, the known results are tight (or almost tight for chain rules), at least when considering non-adaptive black-box reductions. Although black-box impossibility results have been overcome by non black-box constructions in the past [Bar01], we find it hard to imagine how non black-box constructions or adaptivity could help in this setting. We believe that relative to the oracles we construct also adaptive reductions are impossible as adaptivity "obviously" is no of use, but proving this seems hard. Our results are summarized in Figures 1 and 2.

[1] Their $\Omega(n/log(n))$ lower bound matches existing constructions from *regular* one-way functions [GKL93]. For general one-way functions this lower bound is still far of the best construction [VZ12] making $\tilde{\Theta}(n^3)$ calls.

Complexity of the adversary. In order to prove a black-box separation, we will construct an oracle and prove the separation unconditionally relative to this oracle, i.e., assuming all parties have access to it. This then shows that any construction/proof circumventing or separation in the plain model cannot be relativizing, which in particular rules out all black-box constructions [BGS75, IR88].

In the discussion below we measure the complexity of adversaries only in terms of numbers of oracle queries. Of course, in the actual proof we also bound them in terms of circuit size. For our upper bounds the circuits will be of basically the same size as the number of oracle queries (so the number of oracle queries is a good indication of the actual size), whereas for the lower bounds, we can even consider circuits of exponential size, thus making the bounds stronger (basically, we just require that one cannot hard-code a large fraction of the function table of the oracle into the circuit).

Transformations. It is often easy to prove that a variable $X \in \{0,1\}^n$ has

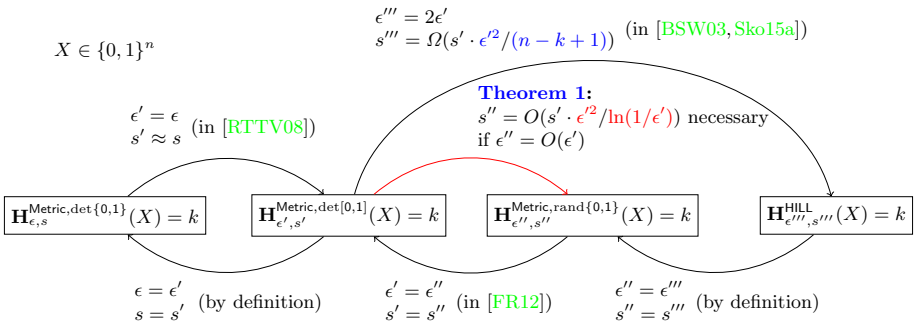


Fig. 1. Transformations: our bound comparing to the state of art. Our Thm. 1, stating that a loss of $\epsilon'^2/\ln(1/\epsilon')$ in circuit size is necessary for black-box reductions that show how deterministic implies randomized metric entropy (if the advantage ϵ' remains in the same order) requires $\epsilon' = 2^{-O(n-k+1)}$ and thus $\ln(1/\epsilon') \in O(n-k+1)$, so there's no contradiction between the transformations from [BSW03, Sko15a] and our lower bound (i.e., the blue term is smaller than the red one).

so called Metric pseudoentropy against deterministic distinguishers, denoted $\mathbf{H}^{\mathsf{Metric},\det\{0,1\}}_{\epsilon,s}(X) = k$. Unfortunately, this notion is usually too weak to be useful, as it only states that for every (deterministic, boolean) distinguisher, there exists some Y with $\mathbf{H}_\infty(Y) = k$ that fools this particular distinguisher, but one usually needs a single Y that fools all (randomised) distinguishers, this is captured by HILL pseudoentropy.

Barak et al. [BSW03] show that any variable $X \in \{0,1\}^n$ that has Metric entropy, also has the same amount of HILL entropy. Their proof uses the min-max theorem, and although it perseveres the amount k of entropy, the quality drops

from (ϵ, s) to $(2\epsilon, \Omega(s\cdot\epsilon^2/n))$. A slightly better bound $\big(2\epsilon, \Omega(s\cdot\epsilon^2/(n+1-k))\big)$ (where again k is the amount of Metric entropy), was given recently in [Sko15a]. The argument uses the min-max theorem and some results on convex approximation in L_p spaces.

In Theorem 1 we show that this is optimal – up to a small factor $\Theta((n-k+1)/\ln(1/\epsilon))$ – as a loss of $\Omega(\ln(1/\epsilon)/\epsilon^2)$ in circuit size is necessary for any black-box reduction. Note that for sufficiently small $\epsilon \in 2^{-\Omega(n-k+1)}$ our bound even matches the positive result up to a small constant factor.

The high-level idea of our separation is as follows; We construct an oracle $\mathcal{O}$ and a variable $X \in \{0,1\}^n$, such that relative to this oracle X can be distinguished from any variable Y with high min-entropy when we can make one randomized query, but for any deterministic distinguisher A, we can find a Y with high min-entropy which A cannot distinguish from X.

To define $\mathcal{O}$, we first choose a uniformly random subset $S \in \{0,1\}^n$ of size $|S| = 2^m$. Moreover we chose a sufficiently large set of boolean functions $D_1(\cdot), \ldots, D_h(\cdot)$ as follows: for every $x \in S$ we set $D_i(x) = 1$ with probability $1/2$ and for every $x \notin S$, $D_i(x) = 1$ with probability $1/2 + \delta$.

Given any x, we can distinguish $x \in S$ from $x \notin S$ with advantage $\approx 2\delta$ by quering $D_i(x)$ *for a random* i. This shows that X cannot have much more than $\log(|S|) = m$ bits of HILL entropy (in fact, even probabilistic Metric entropy) as any variable Y with $\mathbf{H}_\infty(Y) \geqslant m+1$ has at least half of its support outside S, and thus can be distinguished with advantage $\approx 2\delta/2 = \delta$ with one query as just explained. Concretely (recall that in this informal discussion we measure size simply by the number of oracle queries)

$$\mathbf{H}^{\mathsf{Metric},\mathrm{rand}\{0,1\}}_{\delta,1}(X) \leqslant m+1$$

On the other hand, if the adversary is allowed q *deterministic* queries, then intuitively, the best he can do is to query $D_1(x), \ldots, D_q(x)$ and guess that $x \in S$ if less than a $1/2 + \delta/2$ fraction of the outputs is 1. But even if $q = 1/\delta^2$, this strategy will fail with constant probability. Thus, we can choose a Y with large support outside S (and thus also high min-entropy) which will fool this adversary. This shows that X does have large Metric entropy against deterministic distinguishers, even if we allow the adversaries to run in time $1/\delta^2$, concretely, we show that

$$\mathbf{H}^{\mathsf{Metric},\mathrm{det}\{0,1\}}_{\Theta(\delta),O(1/\delta^2)}(X) \geqslant n - O(\log(1/\delta))$$

The adversary. We show impossibility in the non-uniform setting, i.e., for any input length, the distinguisher circuit can depend arbitrarily on the oracle. Like in many non-uniform black-box separation results (including [Sim98, LTW07, RTTV08, Zha11, Wat14]), the type of adversaries for which we can rigorously prove the lower bound is not completely general - but the necessary restrictions seem "obviously" irrelevant. In particular, for a given input x (where the adversary is challenged to distinguish) we only allow the queries on x. This doesn't seem like a real restriction as the distribution of $D_i(x')$ for any $x' \neq x$ is independent of x, and thus seems useless to the adversary. However such queries

can make the success probability of the adversary on different inputs correlated, which makes the overall (average) advantage hard to analyze[2]. Moreover, we assume the adversary makes his queries non-adaptively, i.e., it choses the indices $i_1, \ldots, i_q$ before seeing the outputs of the queries $D_{i_1}(x), \ldots, D_{i_q}(x)$. As all the D_i' are identically distributed, this doesn't seem like a relevant restriction either.

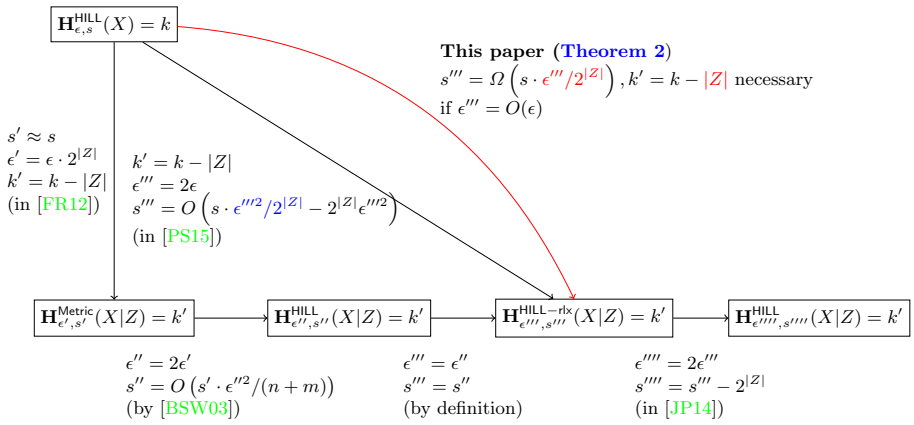


Fig. 2. Chain Rules: our lower bounds comparing to the state of art. In the literature there are basically three approaches to prove a chain rule for HILL entropy. The first one reduces the problem to an efficient version of the dense model theorem [RTTV08], the second one uses the so called auxiliary input simulator [JP14], and the last one is by a convex optimization framework [PS15, Sko16a]. The last approach yields a chain rule with a loss of $\approx 2^m/\epsilon^2$ in circuit size, where m is the length of leakage Z.

Chain Rules. Most (if not all) information theoretic entropy notions $H(.)$ satisfy some kind of chain rule, which states that the entropy of a variable X, when conditioned on another variable Z, can decrease by at most the bitlength $|Z|$ of Z, i.e., $H(X|Z) \geqslant H(X) - |Z|$.

Such a chain rule also holds for some computational notions of entropy. For HILL entropy a chain rule was first proven in [RTTV08, NY19] by a variant of the *dense model theorem*, and was improved by Fuller and Reyzin [FR12]. A different approach using a *simulator* was proposed in [JP14] and later improved by Vadhan and Zheng [VZ13]. A unified approach, based on convex optimization techniques was proposed recently in [PS15, Sko16a] achieving best bounds so far.

The "dense model theorem approach" [FR12] proceeds as follows: one shows that if X has k bits of HILL entropy, then $X|Z$ has $k-m$ (where $Z \in \{0,1\}^m$) bits

[2] This is also the limitation of the mentioned related works on black-box separation, and we don't know the solution for this technical challenge. We state is as an open problem in Section 5.

of Metric entropy. In a second step one applies a Metric to HILL transformation, first proven by Barak et al. [BSW03], to argue that $X|Z$ has also large HILL. The first step loses a factor 2^m in advantage, the second another $2^{2m}\epsilon^2$ in circuit size. Eventually the loss in circuit size is $2^{2m}/\epsilon^2$ and the loss in advantage is 2^m, which in terms of the security ratio size/advantage gives the total loss of $2^m/\epsilon^2$.

A more direct "simulator" approach [VZ13] loses only a multiplicative factor $2^m/\epsilon^2$ in circuit size (there's also an additive $1/\epsilon^2$ term) but there is no loss in advantage. The additive term can be improved to only $2^m\epsilon^2$ as shown in [PS15, Sko16a].

In this paper we show that a loss of $2^m/\epsilon$ is necessary. Note that this still is a factor $1/\epsilon$ away from the positive result. Our result as stated in Theorem 2 is a bit stronger as just outlined, as we show that the loss is necessary even if we only want a bound on the "relaxed" HILL entropy of $X|Z$ (a notion weaker than standard HILL).

To prove our lower bound, we construct an oracle $\mathcal{O}(.)$, together with a joint distribution $(X, Z) \in \{0,1\}^n \times \{0,1\}^m$. We want X to have high HILL entropy relative to $\mathcal{O}(.)$, but when conditioning on Z it should decrease as much as possible (in quantity and quality).

We first consider the case $m = 1$, i.e., the conditional part Z is just one bit. For $n \gg \ell \gg m = 1$ the oracle $\mathcal{O}(.)$ and the distribution (X, Z) is defined as follows. We sample (once and for all) two (disjoint) random subset $\mathcal{X}_0, \mathcal{X}_1 \subseteq \{0,1\}^n$ of size $|\mathcal{X}_0| = |\mathcal{X}_1| = 2^{\ell-1}$, let $\mathcal{X} = \mathcal{X}_0 \cup \mathcal{X}_1$. The oracle $\mathcal{O}(.)$ on input x is defined as follows (below B_p denotes the Bernoulli distribution with parameter p, i.e., $\Pr[b = 1 \ : \ b \leftarrow B_p] = p$).

- If $x \in \mathcal{X}_0$ output a sample of $B_{1/2+\delta}$.
- If $x \in \mathcal{X}_1$ output a sample of $B_{1/2-\delta}$.
- Otherwise, if $x \notin \mathcal{X}$, output a sample of $B_{1/2}$.

Note that our oracle $\mathcal{O}(.)$ is probabilistic, but it can be "derandomized" as we'll explain at the beginning of Section 4. The joint distribution (X, Z) is sampled by first sampling a random bit $Z \leftarrow \{0,1\}$ and then $X \leftarrow \mathcal{X}_Z$.

Given a tuple (V, Z), we can distinguish the case $V = X$ from the case where $V = Y$ for any Y with large support outside of $\mathcal{X}$ (X has min-entropy ℓ, so let's say we take a variable Y with $\mathbf{H}_\infty(Y|Z) \geqslant \ell + 1$ which will have at least half of its support outside $\mathcal{X}$) with advantage $\Theta(\delta)$ by quering $\alpha \leftarrow \mathcal{O}(V, Z)$, and outputting $\beta = \alpha \oplus Z$.

- If $(V, Z) = (X, Z)$ then $\Pr[\beta = 1] = 1/2 + \delta$. To see this, consider the case $Z = 0$, then $\Pr[\beta = 1] = \Pr[\alpha = 1] = \Pr[\mathcal{O}(X) = 1] = 1/2 + \delta$.
- If $(V, Z) = (Y, Z)$ then $\Pr[\beta = 1] = \Pr[Y \notin \mathcal{X}](1/2) + \Pr[Y \in \mathcal{X}](1/2 + \delta) \leq 1/2 + \delta/2$.

Therefore $X|Z$ doesn't have $\ell + 1$ bits of HILL entropy

$$\mathbf{H}^{\mathsf{HILL}}_{\delta/2,1}(X|Z) < \ell + 1$$

On the other hand, we claim that X (without Z but access to $\mathcal{O}(.)$) cannot be distinguished from the uniform distribution over $\{0,1\}^n$ with advantage $\Theta(\delta)$

unless we allow the distinguisher $\Omega(1/\delta)$ oracle queries (the hidden constant in $\Theta(\delta)$ can be made arbitrary large by stetting the hidden constant in $\Omega(1/\delta)$ small enough), i.e.,

$$\mathbf{H}^{\mathsf{HILL}}_{\Theta(\delta),\Omega(1/\delta)}(X) = n \tag{1}$$

To see why (1) holds, we first note that given some V, a single oracle query is useless to tell whether $V = X$ or $V = U_n$: although in the case where $V = X \in \mathcal{X}_Z$ the output $\mathcal{O}(X)$ will have bias δ, one can't decide in which direction the bias goes as Z is (unconditionally) pseudorandom. If we're allowed in the order $1/\delta^2$ queries, we can distinguish X from U_n with constant advantage, as with $1/\delta^2$ samples one can distinguish the distribution $B_{1/2+\delta}$ (or $B_{1/2-\delta}$) from $B_{1/2}$ with constant advantage. If we just want $\Theta(\delta)$ advantage, $\Omega(1/\delta)$ samples are necessary, which proves (1). While it is easy to prove that for the coin with bias δ one needs $O\left(1/\delta^2\right)$ trials to achieve 99% of certainty, finding the number of trials for some confidence level in $o(1)$ as in our case, is more challenging. We solve this problem by a tricky application of *Renyi divergences*[3] The statement of our "coin problem" with precise bounds is given in Lemma 3.

So far, we have only sketched the case $m = 1$. For $m > 1$, we define a random function $\pi : \{0,1\}^n \to \{0,1\}^{m-1}$. The oracle now takes an extra $m-1$ bit string j, and for $x \in \mathcal{X}$, the output of $\mathcal{O}(x, j)$ only has bias δ if $\pi(x) = j$ (and outputs a uniform bit everywhere else). We define the joint distribution (X, Z) by sampling $X \leftarrow \mathcal{X}$, define Z' s.t. $X \in \mathcal{X}_{Z'}$, and set $Z = \pi(X)\|Z'$. Now, given Z, we can make one query $\alpha \leftarrow \mathcal{O}(V, Z[1 \dots m-1])$ and output $\beta = \alpha \oplus Z[m]$, where, as before, getting advantage δ in distinguishing X from any Y with min-entropy $\geq \ell + 1$.

On the other hand, given some V (but no Z) it is now even harder to tell if $V = X$ or $V = Y$. Not only don't we know in which direction the bias goes as before in the case $m = 1$ (this information is encoded in the last bit $Z[m]$ of Z), but we also don't know on which index $\pi(V)$ (in the case $V = X$) we have to query the oracle to observe any bias at all. As there are 2^{m-1} possible choices for $\pi(V)$, this intuitively means we need 2^{m-1} times as many samples as before to observe any bias, which generalises (1) to

$$\mathbf{H}^{\mathsf{HILL}}_{\Theta(\delta),\Omega(2^{m-1}/\delta)}(X) = n$$

1.1 Some implications of our lower bounds

Leakage Resilient Cryptography. The chain rule for HILL entropy is a main technical tool used in several security proofs like the construction of leakage-resilient schemes [NY19, Pie09]. Here, the quantitative bound provided by the chain rule directly translates into the amount of leakage these constructions can tolerate. Our Theorem 2 implies a lower bound on the necessary security degradation for this proof technique. This degradation is, unfortunately, rather

[3] Lower bounds [Zha11, Wat14] also require nontrivial binomial estimates. They were obtained, however by direct and involved calculations.

severe: even if we just leak $m = 1$ bit, we will lose a factor $2^m/\epsilon$, which for a typical security parameter $\epsilon = 2^{-80}$ means a security degradation of "80 bits".

Let us also mention that Theorem 2 answers a question raised by Fuller and Reyzin [FR12], showing that for any chain rule the *simultaneous loss* in quality and quantity is necessary[4].

Faking Auxiliary Inputs. [JP14, VZ13, Skó16b] consider the question how efficiently one can "fake" auxiliary inputs. Concretely, given any joint distribution (X, Z) with $Z \in \{0,1\}^m$, construct an *efficient* simulator h s.t. $(X, h(X))$ is (ϵ, s)-indistinguishable from (X, Z). For example [VZ13] gives a simulator h of complexity $O\left(2^m \epsilon^2 \cdot s\right)$ (plus additive terms independent of s). This result has found many applications in leakage-resilient crypto, complexity theory and zero-knowledge theory. The best known lower bound (assuming exponentially hard OWFs) is $\Omega\left(\max(2^m, 1/\epsilon)\right)$. Since the chain rule for relaxed HILL entropy follows by a simulator argument [JP14] with the same complexity loss, our Theorem 2 yields a better lower bound $\Omega\left(2^m/\epsilon\right)$ on the complexity of simulating auxiliary inputs.

Dense Model Theorem. The computational dense model theorem [RTTV08] says, roughly speaking, that dense subsets of pseudorandom distributions are computationally indistinguishable from true dense distributions. It has found applications including differential privacy, memory delegation, graph decompositions and additive combinatorics. It is well known that the worst-case chain rule for HILL-entropy is equivalent to the dense model theorem, as one can think of dense distributions as uniform distributions X given short leakage Z. For settings with constant density, which correspond to $|Z| = O(1)$, HILL and relaxed HILL entropy are equivalent [JP14]; moreover, the complexity loss in the chain rule is then equal to the cost of transforming Metric Entropy into HILL Entropy. Now our Theorem 1 implies a necessary loss in circuit size $\Omega\left(1/\epsilon^2\right)$ if one wants ϵ-indistinguishability. This way we reprove the tight lower bound due to Zhang [Zha11] for constant densities.

2 Basic Definitions

Let X_1 and X_2 be two distributions over the same finite set. The *statistical distance* of X_1 and X_2 equals $\mathrm{SD}\left(X_1; X_2\right) = \frac{1}{2}\sum_x |\Pr[X_1 = x] - \Pr[X_2 = x]|$.

Definition 1 (Min-Entropy). *A random variable X has* min-entropy *k, denoted by $\mathbf{H}_\infty(X) = k$, if $\max_x \Pr[X = x] \leq 2^{-k}$.*

Definition 2 (Average conditional min-Entropy [DRS04]). *For a pair (X, Z) of random variables, the* average min-entropy *of X conditioned on Z is*

$$\widetilde{\mathbf{H}}_\infty(X|Z) = -\log \mathop{\mathbb{E}}_{z \leftarrow Z}[\max_x \Pr[X = x|Z = z]] = -\log \mathop{\mathbb{E}}_{z \leftarrow Z}[2^{-\mathbf{H}_\infty(X|Z=z)}]$$

[4] Their question was about chain rules bounding the worst-case entropy, that is bounding $\mathbf{H}^{\mathsf{HILL}}(X|Z = z)$ for every z. Our result, stated simply for average entropy $\mathbf{H}^{\mathsf{HILL}}(X|Z)$, is much more general and applies to qualitatively better chain rules obtained by simulator arguments.

Distinguishers. We consider several classes of distinguishers. With $\mathcal{D}_s^{\mathsf{rand},\{0,1\}}$ we denote the class of randomized circuits of size at most s with boolean output (this is the standard non-uniform class of distinguishers considered in cryptographic definitions). The class $\mathcal{D}_s^{\mathsf{rand},[0,1]}$ is defined analogously, but with real valued output in $[0,1]$. $\mathcal{D}_s^{\mathsf{det},\{0,1\}}, \mathcal{D}_s^{\mathsf{det},[0,1]}$ are defined as the corresponding classes for *deterministic* circuits. With $\Delta^D(X;Y) = |\mathbb{E}_X[D(X)] - \mathbb{E}_Y[D(Y)]$ we denote D's advantage in distinguishing X and Y.

Definition 3 (HILL pseudoentropy [HILL99,HLR07]). *A variable X has* HILL entropy *at least k if*

$$\mathbf{H}^{\mathsf{HILL}}_{\epsilon,s}(X) \geq k \iff \exists Y \ , \ \mathbf{H}_\infty(Y) = k \ \forall D \in \mathcal{D}_s^{\mathsf{rand},\{0,1\}} \ : \ \Delta^D(X;Y) \leq \epsilon$$

For a joint distribution (X,Z), we say that X has k bits conditonal Hill entropy *(conditionned on Z) if*

$$\begin{aligned} & \mathbf{H}^{\mathsf{HILL}}_{\epsilon,s}(X|Z) \geq k \\ \iff & \exists (Y,Z), \widetilde{\mathbf{H}}_\infty(Y|Z) = k \ \forall D \in \mathcal{D}_s^{\mathsf{rand},\{0,1\}} : \Delta^D((X,Z);(Y,Z)) \leq \epsilon \end{aligned}$$

Definition 4 (Metric pseudoentropy [BSW03]). *A variable X has* Metric entropy *at least k if*

$$\mathbf{H}^{\mathsf{Metric}}_{\epsilon,s}(X) \geq k \iff \forall D \in \mathcal{D}_s^{\mathsf{rand},\{0,1\}} \exists Y_D \ , \ \mathbf{H}_\infty(Y_D) = k \ : \ \Delta^D(X;Y_D) \leq \epsilon$$

Metric star entropy *is defined analogousely but using deterministic real valued distinguishers*

$$\mathbf{H}^{\mathsf{Metric}*}_{\epsilon,s}(X) \geq k \iff \forall D \in \mathcal{D}_s^{\mathsf{det},[0,1]} \exists Y_D \ , \ \mathbf{H}_\infty(Y_D) = k \ : \ \Delta^D(X;Y_D) \leq \epsilon$$

Relaxed versions of HILL and Metric entropy. A weaker notion of conditional HILL entropy allows the conditional part to be replaced by some computationally indistinguishable variable

Definition 5 (Relaxed HILL pseudoentropy [GW11,Rey11]). *For a joint distribution (X,Z) we say that X has* relaxed HILL entropy *k conditioned on Z if*

$$\begin{aligned} & \mathbf{H}^{\mathsf{HILL}-\mathsf{rlx}}_{\epsilon,s}(X|Z) \geq k \\ \iff & \exists (Y,Z'), \widetilde{\mathbf{H}}_\infty(Y|Z') = k, \forall D \in \mathcal{D}_s^{\mathsf{rand},\{0,1\}} \ , \ : \ \Delta^D((X,Z);(Y,Z')) \leq \epsilon \end{aligned}$$

The above notion of *relaxed* HILL satisfies a chain rule whereas the chain rule for the standard definition of conditional HILL entropy is known to be false [KPW13]. One can analogously define relaxed variants of metric entropy, we won't give these as they will not be required in this paper.

Pseudoentropy against different distinguisher classes. For randomized distinguishers, it's irrelevant if the output is boolean or real values, as we can replace any $D \in \mathcal{D}_s^{\mathsf{rand},[0,1]}$ with a $D' \in \mathcal{D}^{\mathsf{rand},\{0,1\}}$ s.t. $\mathbb{E}[D'(X)] = \mathbb{E}[D(X)]$ by

setting (for any x) $\Pr[D'(x) = 1] = \mathbb{E}[D(x)]$. For HILL entropy (as well as for its relaxed version), it also doesn't matter if we consider randomized or deterministic distinguishers in Definition 3, as we always can "fix" the randomness to an optimal value. This is no longer true for metric entropy,[5] and thus the distinction between metric and metric star entropy is crucial.

3 A Lower Bound on Metric-to-HILL Transformations

Theorem 1. *For every n, k, m and ϵ such that $n \geqslant k + \log(1/\epsilon) + 4$, $\frac{1}{8} > \epsilon$ and $n - 1 \geq m > 6\log(1/\epsilon)$ there exist an oracle $\mathcal{O}$ and a distribution X over $\{0,1\}^n$ such that*

$$\mathbf{H}^{\mathsf{Metric},\mathrm{det}\{0,1\}}_{\epsilon,T}(X) \geqslant k \tag{2}$$

here the complexity T denotes any circuit of size $2^{O(m)}$ that makes at most $\frac{\ln(2/\epsilon)}{216\epsilon^2}$ non-adaptive queries and, simultaneously,

$$\mathbf{H}^{\mathsf{Metric},\mathrm{rand}\{0,1\}}_{2\epsilon,T'}(X) \leqslant m + 1 \tag{3}$$

where the distinguishers size T' is only $O(n)$ and the query complexity is 1.

Let S be a random subset of $\{0,1\}^n$ of size 2^m, where $m \leqslant n - 1$, and let $D_1, \ldots, D_h$ be boolean functions drawn independently from the following distribution D: $D(x) = 1$ on S with probability p if $x \in S$ and $D(x) = 1$ with probability q if $x \in S^c$, where $p > q$ and $p + q = 1$. Denote $X = U_S$. We will argue that the metric entropy against a probabilistic adversary who is allowed one query is roughly m with advantage $\Omega(p-q)$. But the metric entropy against non-adaptive deterministic adversary who can make t queries of the form $D_i(x)$ is much bigger, even if $t = O\left((p-q)^{-2}\right)$. Let us sketch an informal argument before we give the actual proof. We need to prove two facts:

(i) There is a probabilistic adversary A^* such that with high probability over $X, D_1, \ldots, D_h$ we have $\Delta^{\mathsf{A}^*}(X, Y) = \Omega(p - q)$ for all Y with $\mathbf{H}_\infty(Y) \geqslant m + 1$.

(ii) For every deterministic adversary A making at most $t = O\left((p-q)^{-2}\right)$ non-adaptive queries, with high probability over $X, D_1, \ldots, D_h$ we have $\Delta^{\mathsf{A}}(X; Y) = 0$ for some Y with $\mathbf{H}_\infty(Y) = n - \Theta(1)$.

To prove (i) we observe that the probabilistic adversary can distinugish between S and S^c by comparing the bias of ones. We simply let A^* forward its input to D_i for a randomly chosen i, i.e.,

$$\mathsf{A}^*(x) = D_i(x), \quad i \leftarrow [1, \ldots, h]$$

[5] It might be hard to find a high min-entropy distribution Y that fools a randomized distinguisher D, but this task can become easy once D's randomness is fixed.

With extremely high probability we have $\Pr[\mathsf{A}^*(x)=1] \in [p-\delta, p+\delta]$ if $x \in S$ and $\Pr[\mathsf{A}^*(x)=1] \in [q-\delta, q+\delta]$ if $x \notin S$ for some $\delta \ll p-q$ (by a Chernoff bound, δ drops exponentially fast in h, so we just have to set h large enough). We have then $\Pr[\mathsf{A}^*(X)=1] \geqslant p+\delta$ and $\Pr[\mathsf{A}^*(Y)=1] \leqslant 1/2 \cdot (p+q+2\delta)$ for every Y of min-entropy at least $m+1$ (since then $\Pr[Y \in S] \leqslant 1/2$). This yields $\Delta^{\mathsf{A}^*}(X;Y) = (p-q)/2$. In order to prove (ii) one might intuitively argue that the best a t-query deterministic adversary can do to contradict to (ii), is to guess whether some value x has bias p or $q=1-p$, by taking the majority of t samples

$$\mathsf{A}(x) = \mathrm{Maj}(D_1(x), \ldots, D_t(x))$$

But even if $t = \Theta(1/(p-q)^2)$, majority will fail to predict the bias with constant probability. This means there exists a variable Y with min-entropy $n - \Theta(1)$ such that $\Pr[\mathsf{A}(Y)=1] = \Pr[\mathsf{A}(X)=1]$. The full proof gives quantitative forms of (i) and (ii), showing essentially that "majority is best" and appears in Appendix A.

4 Lower Bounds on Chain Rules

For any $n \gg \ell \gg m$, we construct a distribution $(X,Z) \in \{0,1\}^n \times \{0,1\}^m$ and an oracle $\mathcal{O}(.)$ such that relative to this oracle, X has very large HILL entropy but the HILL entropy of $X|Z$ is much lower in quantity and quality: for arbitrary $n \gg \ell \gg m$ (where $|Z|=m$, $X \in \{0,1\}^n$), the quantity drops from n to $\ell - m + 2$ (it particular, by much more than $|Z|=m$), even if we allow for a $2^m/\epsilon$ drop in quality.

Theorem 2 (A lower bound on the chain rule for $\mathbf{H}^{\mathsf{HILL-rlx}}$). *There exists a joint distribution (X,Z) over $\{0,1\}^n \times \{0,1\}^m$, and an oracle $\mathcal{O}$ such that, relative to $\mathcal{O}$, for any (ℓ,δ) such that $\frac{n}{2} - \frac{\log(1/\delta)}{2} > m$ and $\ell > m + 6\log(1/\delta)$, we have*

$$\mathbf{H}^{\mathsf{HILL}}_{\delta/2,T}(X) = n \tag{4}$$

where[6] *$T > c \cdot 2^m/\delta$ with some absolute constant c but*

$$\mathbf{H}^{\mathsf{HILL-rlx}}_{\delta/2,T'}(X|Z) < \ell + 1 \tag{5}$$

where T' captures a circuit of size only $O(n)$ making only 1 oracle query.

Remark 1 (On the technical restrictions). Note that the assumptions on ℓ and δ are automatically satisfied in most interesting settings, as typically we assume $m \ll n$ and $\log(1/\delta) \ll n$.

[6] The class of adversaries here consists of all circuits with the total number of gates, including oracle gates, at most T. Theorem 2 is also true when the circuit size s is much bigger than the total number of oracle gates T (under some assumption on s, ℓ, ϵ). For simplicity, we do not state this version.

Remark 2 (A strict separation). The theorem also holds if we insist on a larger distinguishing advantage after leakage. Concretely, allowing for more than just one oracle query, the $\delta/2$ advantage in (5) can be amplified to $C\delta$ for any constant C assuming δ is small enough to start with (see Remark 4 in the proof).

The full proof appears in Appendix B. The heart of the argument is a lower bound on the query complexity for the corresponding "coin problem": we need to distinguish between T random bits, and the distribution where we sample equally likely T independent bits B_p or T independent bits B_q where $p = \frac{1}{2} + \delta$ and $q = 1 - p$. (see Appendix C for more details). The rest of the proof is based on a standard concentration argument, using extensively Chernoff Bounds.

5 Open Problems

As shown in Figure 2, there remains a gap between the best proofs for the chain-rule, which lose a factor $\epsilon^2/2^{|Z|}$ in circuit size, and the required loss of $\epsilon/2^{|Z|}$ we prove in this paper. Closing this bound by either improving the proof for the chain-rule or give an improved lower bound remains an intriguing open problem.

Our lower bounds are only proven for adversaries that make their queries non-adaptively. Adaptive queries don't seem to help against our oracle, but rigorously proving this fact seems tricky.

Finally, the lower bounds we prove on the loss of circuit size assume that the distinguishing advantage remains roughly the same. There exist results which are not of this form, in particular – as shown in Figure 2 – the HILL to Metric transformation from [FR12] only loses in distinguishing advantage, not in circuit size (i.e., we have $s \approx s'$). Proving lower bounds and giving constructions for different circuit size vs. distinguishing advantage trade-offs leave many challenges for future work.

Although the restricted adversary model - querying on same inputs - seems to be widely accepted, it remains a challenging open problem to rigorously prove that querying on different inputs really doesn't help; not only for the problem discussed here but also in the case of Dense Model Theorems and Hardcore Sets Constructions.

References

Bar01. Boaz Barak, *How to go beyond the black-box simulation barrier*, 42nd Annual Symposium on Foundations of Computer Science, FOCS 2001, 14-17 October 2001, Las Vegas, Nevada, USA, 2001, pp. 106–115.

BGS75. Theodore Baker, John Gill, and Robert Solovay, *Relativizations of the p=?np question*, SIAM Journal on computing **4** (1975), no. 4, 431–442.

BSW03. Boaz Barak, Ronen Shaltiel, and Avi Wigderson, *Computational analogues of entropy*, In 11th International Conference on Random Structures and Algorithms, 2003, pp. 200–215.

CKLR11. Kai-Min Chung, Yael Tauman Kalai, Feng-Hao Liu, and Ran Raz, *Memory delegation*, Advances in Cryptology - CRYPTO 2011 - 31st Annual Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2011. Proceedings, 2011, pp. 151–168.

DRS04. Yevgeniy Dodis, Leonid Reyzin, and Adam D. Smith, *Fuzzy extractors: How to generate strong keys from biometrics and other noisy data*, Advances in Cryptology - EUROCRYPT 2004, International Conference on the Theory and Applications of Cryptographic Techniques, Interlaken, Switzerland, May 2-6, 2004, Proceedings, 2004, pp. 523–540.

FOR12. Benjamin Fuller, Adam O'Neill, and Leonid Reyzin, *A unified approach to deterministic encryption: New constructions and a connection to computational entropy*, Theory of Cryptography - 9th Theory of Cryptography Conference, TCC 2012, Taormina, Sicily, Italy, March 19-21, 2012. Proceedings, 2012, pp. 582–599.

FR12. Benjamin Fuller and Leonid Reyzin, *Computational entropy and information leakage*, Cryptology ePrint Archive, Report 2012/466, 2012, http://eprint.iacr.org/.

GKL93. Oded Goldreich, Hugo Krawczyk, and Michael Luby, *On the existence of pseudorandom generators*, SIAM J. Comput. **22** (1993), no. 6, 1163–1175.

GW11. Craig Gentry and Daniel Wichs, *Separating succinct non-interactive arguments from all falsifiable assumptions*, Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6-8 June 2011, 2011, pp. 99–108.

HILL99. Johan Håstad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby, *A pseudorandom generator from any one-way function*, SIAM J. Comput. **28** (1999), no. 4, 1364–1396.

HLR07. Chun-Yuan Hsiao, Chi-Jen Lu, and Leonid Reyzin, *Conditional computational entropy, or toward separating pseudoentropy from compressibility*, Advances in Cryptology - EUROCRYPT 2007, 26th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Barcelona, Spain, May 20-24, 2007, Proceedings, 2007, pp. 169–186.

Hol06. Thomas Holenstein, *Pseudorandom generators from one-way functions: A simple construction for any hardness*, Theory of Cryptography, Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006, Proceedings, 2006, pp. 443–461.

HRV10. Iftach Haitner, Omer Reingold, and Salil P. Vadhan, *Efficiency improvements in constructing pseudorandom generators from one-way functions*, Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010, 2010, pp. 437–446.

HS12. Thomas Holenstein and Makrand Sinha, *Constructing a pseudorandom generator requires an almost linear number of calls*, 53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012, 2012, pp. 698–707.

ILL89. Russell Impagliazzo, Leonid A. Levin, and Michael Luby, *Pseudo-random generation from one-way functions (extended abstracts)*, Proceedings of the 21st Annual ACM Symposium on Theory of Computing, May 14-17, 1989, Seattle, Washigton, USA, 1989, pp. 12–24.

IR88. Russell Impagliazzo and Steven Rudich, *Limits on the provable consequences of one-way permutations*, Advances in Cryptology - CRYPTO '88, 8th Annual International Cryptology Conference, Santa Barbara, California, USA, August 21-25, 1988, Proceedings, 1988, pp. 8–26.

JP14. Dimitar Jetchev and Krzysztof Pietrzak, *How to fake auxiliary input*, Theory of Cryptography - 11th Theory of Cryptography Conference, TCC 2014, San Diego, CA, USA, February 24-26, 2014. Proceedings, 2014, pp. 566–590.

KPW13. Stephan Krenn, Krzysztof Pietrzak, and Akshay Wadia, *A counterexample to the chain rule for conditional HILL entropy - and what deniable encryption has to do with it*, Theory of Cryptography - 10th Theory of Cryptography Conference, TCC 2013, Tokyo, Japan, March 3-6, 2013. Proceedings, 2013, pp. 23–39.

LTW07. Chi-Jen Lu, Shi-Chun Tsai, and Hsin-Lung Wu, *On the complexity of hard-core set constructions*, Automata, Languages and Programming, 34th International Colloquium, ICALP 2007, Wroclaw, Poland, July 9-13, 2007, Proceedings, 2007, pp. 183–194.

NY19. Ryo Nishimaki and Takashi Yamakawa, *Leakage-resilient identity-based encryption in bounded retrieval model with nearly optimal leakage-ratio*, Public-Key Cryptography - PKC 2019 - 22nd IACR International Conference on Practice and Theory of Public-Key Cryptography, Beijing, China, April 14-17, 2019, Proceedings, Part I, 2019, pp. 466–495.

Pie09. Krzysztof Pietrzak, *A leakage-resilient mode of operation*, Advances in Cryptology - EUROCRYPT 2009, 28th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Cologne, Germany, April 26-30, 2009. Proceedings, 2009, pp. 462–482.

PS15. Krzysztof Pietrzak and Maciej Skorski, *The chain rule for HILL pseudoentropy, revisited*, Progress in Cryptology - LATINCRYPT 2015 - 4th International Conference on Cryptology and Information Security in Latin America, Guadalajara, Mexico, August 23-26, 2015, Proceedings, 2015, pp. 81–98.

Rey11. Leonid Reyzin, *Some notions of entropy for cryptography - (invited talk)*, Information Theoretic Security - 5th International Conference, ICITS 2011, Amsterdam, The Netherlands, May 21-24, 2011. Proceedings, 2011, pp. 138–142.

RTTV08. Omer Reingold, Luca Trevisan, Madhur Tulsiani, and Salil P. Vadhan, *Dense subsets of pseudorandom sets*, 49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA, 2008, pp. 76–85.

Sim98. Daniel R. Simon, *Finding collisions on a one-way street: Can secure hash functions be based on general assumptions?*, Advances in Cryptology - EUROCRYPT '98, International Conference on the Theory and Application of Cryptographic Techniques, Espoo, Finland, May 31 - June 4, 1998, Proceeding, 1998, pp. 334–345.

Sko15a. Maciej Skorski, *Metric pseudoentropy: Characterizations, transformations and applications*, Information Theoretic Security - 8th International Conference, ICITS 2015, Lugano, Switzerland, May 2-5, 2015. Proceedings (Anja Lehmann and Stefan Wolf, eds.), Lecture Notes in Computer Science, vol. 9063, Springer, 2015, pp. 105–122.

Sko15b. ———, *Metric pseudoentropy: Characterizations, transformations and applications*, Information Theoretic Security - 8th International Conference, ICITS 2015, Lugano, Switzerland, May 2-5, 2015. Proceedings, 2015, pp. 105–122.

Sko16a. ———, *A better chain rule for hill pseudoentropy - beyond bounded leakage*, Information Theoretic Security - 9th International Conference, ICITS 2016, 2016.

Skó16b. Maciej Skórski, *Simulating auxiliary inputs, revisited*, Theory of Cryptography - 14th International Conference, TCC 2016-B, Beijing, China, October 31 - November 3, 2016, Proceedings, Part I, 2016, pp. 159–179.

VZ12. Salil P. Vadhan and Colin Jia Zheng, *Characterizing pseudoentropy and simplifying pseudorandom generator constructions*, Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012, 2012, pp. 817–836.

VZ13. ———, *A uniform min-max theorem with applications in cryptography*, Advances in Cryptology - CRYPTO 2013 - 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2013. Proceedings, Part I, 2013, pp. 93–110.

Wat14. Thomas Watson, *Advice lower bounds for the dense model theorem*, TOCT **7** (2014), no. 1, 1.

Zha11. Jiapeng Zhang, *On the query complexity for showing dense model*, Electronic Colloquium on Computational Complexity (ECCC) **18** (2011), 38.

A Proof of Theorem 1

A.1 Majority is best

We prove two statements which are quantitative forms of (i) and (ii) discussed after the statement of Theorem 1. First we show that the probabilistic adversary A^* easily distinguishes X from all Y of high min-entropy.

Claim 1 (Probabilistic Metric Entropy of X is small) *Let A^* be a probabilistic adversary who on input x samples a random $i \in [1,\ldots,h]$, then queries for $D_i(x)$ and outputs the response. Then for any $\delta \leqslant (p-q)/3$ we have*

$$\Pr[\forall Y:\ \mathbf{H}_\infty(Y) \geqslant m+1,\ \Delta^{\mathsf{A}^*}(X;Y) \geqslant (p-q)/3] \geqslant 1-2^{\max(n-1,m+1)}\exp(-h\delta^2). \tag{6}$$

Remark 3 (The complexity of the probabilistic distinguisher). We can chose h in Claim 1 to be 2^n, then A^* is of size $O(n)$ and makes only one query.

Consider now a deterministic adversary A who on input x can make at most t queries learning $D_i(x)$ for t different $i \in [1,\ldots,h]$. We claim that

Claim 2 (Deterministic Metric Entropy is big) *Suppose that we have $n \geqslant k+\log(1/\epsilon)+4$ and $\delta = \frac{\epsilon^2}{2+2\epsilon}$. Then for every nonadaptive adversary A which makes $t \leqslant \frac{\ln(2/\epsilon)}{6(p-q)^2}$ queries we have*

$$\Pr_{X,D_1,\ldots,D_h}\left[\exists Y:\ \mathbf{H}_\infty(Y) \geqslant k,\ \Delta^{\mathsf{A}}(X;Y) \leqslant \epsilon\right] \geqslant 1-4\exp(-2^m\delta^2). \tag{7}$$

Setting $p-q=6\epsilon$ we see that Equation (2) follows from Claim 1 and Equation (3) follows from Equation (7) combined with the union bound over all distinguishers. Note that the right hand side of Equation (7) converges to 1 with the rate *doubly* exponential in m, so we can even afford taking a union bound over all distinguishers of size exponential in m.

Proof (of Claim 1). By a Chernoff bound[7] and the union bound

$$\Pr_{X,D_1,\ldots,D_h}[\forall x \in S^c \ : \ \Pr[\mathsf{A}^*(x)=1] \leqslant q+\delta] \geqslant 1-2^{n-1}\exp(-2\delta^2 h) \tag{8}$$

similarly

$$\Pr_{X,D_1,\ldots,D_h}[\forall x \in S : |\Pr[\mathsf{A}^*(x)=1]-p| \leqslant \delta] \geqslant 1-2^m \cdot 2\exp(-2\delta^2 h). \tag{9}$$

The advantage of A^*, with probability $1-2^{n-1}\exp(-2h\delta^2)$, equals

$$\begin{aligned}\Delta^{\mathsf{A}^*}(X;Y) &\geqslant (p-\delta)-(p+\delta)\Pr[Y\in S]-(q+\delta)\Pr[Y\in S^c]\\ &\geqslant p-q-(p-q)\Pr[Y\in S]-2\delta.\end{aligned}$$

Since by the assumption we have $\Pr[Y\in S]\leqslant\frac{1}{2}$, Equation (6) follows.

Proof (of Claim 2). The adversary A non-adaptively queries for $D_i(x)$ values for t distinct i's and then outputs a bit, this bit is thus computed by a function of the form

$$f\left(x, D_{i_1(x)}(x), \ldots, D_{i_t(x)}(x)\right), \tag{10}$$

for some fixed boolean function $f:\{0,1\}^n\times\{0,1\}^t\to\{0,1\}$. We start by simplifying the event (7) using the following proposition, which gives an alternative characterization of the deterministic metric entropy.

Lemma 1 ([BSW03,Sko15b]). *Let D be a boolean deterministic function on $\{0,1\}^n$. Then there exists Y of min-entropy at least k such that $\Delta^D(X;Y)\leqslant\epsilon$ if and only if*

$$\mathbb{E}\, D'(X) \leqslant 2^{n-k}\,\mathbb{E}\, D'(U)+\epsilon \tag{11}$$

holds for $D'\in\{D,\mathbf{1}-D\}$

Since $|S^c|\geqslant 2^{n-1}$, we have $\mathbb{E}\, D(U)\geqslant \mathbb{E}_{x\leftarrow S^c} D(x)/2$ for any function D. Therefore, by Lemma 1, the inequality (7) will be proved if we show that the following inequality holds:

$$\begin{aligned}\Pr_{X,D_1,\ldots,D_h}\left[\forall \mathsf{A}'\in\{\mathsf{A},\mathbf{1}-\mathsf{A}\}: \ \mathop{\mathbb{E}}_{x\leftarrow S}\mathsf{A}'(x)\leqslant 2^{n-k-1}\mathop{\mathbb{E}}_{x\leftarrow S^c}\mathsf{A}'(x)+\epsilon\right]\\ \geqslant 1-4\exp(-2^m\delta^2)\end{aligned} \tag{12}$$

By the union bound, it is enough to show that for $\mathsf{A}'\in\{\mathsf{A},\mathbf{1}-\mathsf{A}\}$ we have

$$\Pr_{X,D_1,\ldots,D_h}\left[\mathop{\mathbb{E}}_{x\leftarrow S}\mathsf{A}'(x)\leqslant 2^{n-k-1}\mathop{\mathbb{E}}_{x\leftarrow S^c}\mathsf{A}'(x)+\epsilon\right]\geqslant 1-2\exp(-2^m\delta^2) \tag{13}$$

In the next step we simplify the expressions $\mathbb{E}_{x\leftarrow S}\mathsf{A}'(x)$ and $\mathbb{E}_{x\leftarrow S^c}\mathsf{A}'(x)$. The following fact is a direct consequence of the Chernoff bound.

[7] We use the following version: let X_i for $i=1,\ldots,N$ be independent random variables such that $X_i\in[a_i,b_i]$. Then for any positive t we have $\Pr_{X_1,\ldots,X_N}\left[\sum_{i=1}^N X_i-\mathbb{E}\left[\sum_{i=1}^N X_i\right]\geqslant t\right]\leqslant\exp\left(\frac{2t^2}{\sum_{i=1}^N(b_i-a_i)^2}\right)$.

Proposition 1. *For any function* $f \in \{0,1\}^n \times \{0,1\}^t \to [0,1]$ *we have*

$$\left| \mathop{\mathbb{E}}_{x \leftarrow S} f\left(x, D_{i_1(x)}(x), \ldots, D_{i_t(x)}(x)\right) - \mathbb{E}\, f(U_n, B_p^1, \ldots, B_p^t) \right| \leqslant \delta \tag{14}$$

$$\left| \mathop{\mathbb{E}}_{x \leftarrow S^c} f\left(x, D_{i_1(x)}(x), \ldots, D_{i_t(x)}(x)\right) - \mathbb{E}\, f(U_n, B_q^1, \ldots, B_q^t) \right| \leqslant \delta \tag{15}$$

with probability $1 - 2\exp(-2 \cdot 2^m \delta^2)$ *over the choice of* X *and* $D_1, \ldots, D_h$.

For any $\mathbf{r} = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_t) \in [0,1]^t$, and any (deterministic or randomized) function $f \in \{0,1\}^t \to [0,1]$ we denote $\mathbb{E}_{\mathbf{r}} f = \mathbb{E} f(B_{\mathbf{r}_1}, \ldots, B_{\mathbf{r}_t})$. It is enough to show that if $\mathbf{r}, \mathbf{r}'$ are both chosen from $\{p,q\}^t$ then we have

$$\mathbb{E}_{\mathbf{r}} f + \delta \leqslant 2^{n-k-1} \max(\mathbb{E}_{\mathbf{r}'} f - \delta, 0) + \epsilon. \tag{16}$$

This inequality will follow by the following lemma (applied to f in the proposition but considered as a function of $\{0,1\}^t$ randomized with the first n input bits).

Lemma 2. *Suppose that* $p, q > 0$ *are such that* $p + q = 1$. *Let* $f : \{0,1\}^t \to [0,1]$ *be an arbitrary function and let* $\mathbf{r}, \mathbf{r}' \in \{p,q\}^t$. *Then for any* $c > 0$ *we have*

$$\mathbf{E}_{\mathbf{r}} f \leqslant \exp\left(\frac{(c+1)(p-q)^2}{q} \cdot t\right) \cdot \mathbf{E}_{\mathbf{r}'} f + \exp(-2c^2(p-q)^2 t).$$

Proof. The idea of the proof is to show that for most values of z the ratio $\Pr[B_{\mathbf{r}} = z] / \Pr[B_{\mathbf{r}'} = z]$ is bounded. We have

$$\begin{aligned}
&\Pr[B_{\mathbf{r}} = z] / \Pr[B_{\mathbf{r}'} = z] && (17)\\
&= (p/q)^{\#\{i : z_i = 1,\ \mathbf{r}_i > \mathbf{r}'_i\} - \#\{i : z_i = 1,\ \mathbf{r}_i < \mathbf{r}'_i\}} \cdot (q/p)^{\#\{i : z_i = 0,\ \mathbf{r}_i > \mathbf{r}'_i\} - \#\{i : z_i = 0,\ \mathbf{r}_i < \mathbf{r}'_i\}} \\
&= (p/q)^{\#\{i : z_i = 1,\ \mathbf{r}_i > \mathbf{r}'_i\} - \#\{i : z_i = 0,\ \mathbf{r}_i > \mathbf{r}'_i\} - \#\{i : z_i = 1,\ \mathbf{r}_i < \mathbf{r}'_i\} + \#\{i : z_i = 0,\ \mathbf{r}_i < \mathbf{r}'_i\}} \\
&= (p/q)^{\sum_{i=1}^t (2z_i - 1) \cdot \operatorname{sgn}(\mathbf{r}_i - \mathbf{r}'_i)} && (18)
\end{aligned}$$

The random variables $\xi_i = (2z_i - 1) \cdot \operatorname{sgn}(\mathbf{r}_i - \mathbf{r}'_i)$ for $i = 1, \ldots, t$, where z is sampled from $B_{\mathbf{r}}$, are independent with the expectations $\mathbb{E}\xi_i = (2\mathbf{r}_i - 1)\operatorname{sgn}(\mathbf{r}_i - \mathbf{r}'_i) \leqslant p - q$. By the Chernoff bound for any $c > 0$ we get

$$\Pr_{z \leftarrow B_{\mathbf{r}}}\left[\sum_{i=1}^t (2z_i - 1) \cdot \operatorname{sgn}(\mathbf{r}_i - \mathbf{r}'_i) \geqslant (p-q)t + c(p-q)t\right] \leqslant \exp(-2c^2(p-q)^2 t). \tag{19}$$

Therefore,

$$\mathbb{E}_{\mathbf{r}} f \leqslant (p/q)^{(c+1)(p-q)t} \mathbb{E}_{\mathbf{r}'} f + 2\exp(-2c^2(p-q)^2 t) \tag{20}$$

and the claim follows by observing that $p/q = 1 + (p-q)/q \leqslant \exp((p-q)/q)$.

From Lemma 2 it follows that Equation (16) is satisfied with

$$\delta \leqslant \frac{\epsilon}{2\exp\left((c+1)(p-q)^2 \cdot t/q\right) + 2} \tag{21}$$

provided that

$$\exp(-2c^2(p-q)^2 \cdot t) \leqslant \epsilon/2 \tag{22}$$

$$\exp\left((c+1)(p-q)^2 \cdot t/q\right) \leqslant 2^{n-k-1} \tag{23}$$

It is easy to see that Equation (23) and Equation (22) are satisfied if and only if

$$\frac{\ln(2/\epsilon)}{2c^2(p-q)^2} \leqslant t \leqslant (n-k-3)\ln 2 \cdot \frac{q}{(c+1)(p-q)^2}.$$

This inequality can be satisfied if and only if

$$\epsilon \geqslant 2 \cdot 2^{(k-n+3)\cdot\frac{2qc^2}{c+1}}.$$

If we set $t = \frac{\ln(2/\epsilon)}{2c^2(p-q)^2}$ then Equation (21) becomes

$$\delta \leqslant \frac{\epsilon}{(2/\epsilon)^{\frac{c+1}{2qc^2}} + 2}$$

Choosing c so that $\frac{2qc^2}{c+1} = 1$ we see that it is enough to assume $\epsilon \geqslant 2 \cdot 2^{k-n+3}$, any δ such that $\delta \leqslant \frac{\epsilon^2}{2+2\epsilon}$ and $t \approx \frac{\ln(2/\epsilon)}{6(p-q)^2}$ (the constant 6 is sightly bigger than the exact value, but if Claim 2 holds true for some t then also for $t' < t$). This finishes the proof of Claim 2.

B Proof of Theorem 2

A Remark on The Oracle. For convenience, the oracle $\mathcal{O} : \{0,1\}^n \to \{0,1\}$ we use in the proof is probabilistic, in the sense that it flips some random coins before answering a query (in particular, making the same query twice might give different outputs). We remark that, as the adversaries considered are probabilistic, one can replace this oracle with a deterministic one $\mathcal{O}_{\text{det}}$ by assigning to every possible query x a 2^L tuple $(x,r), r \in \{0,1\}^L$ of queries (for some sufficiently large L), where the output for $\mathcal{O}_{\text{det}}((x,r))$ is sampled according to $\mathcal{O}(x)$ for every r. We can emulate the output distribution $\mathcal{O}(x)$ by querying $\mathcal{O}((x,r))$ for a random r. On the other hand, for a random x, even an exponential size distinguisher will not be able to distinguish $\mathcal{O}_{\text{def}}((x,\cdot))$ from an oracle which, when queried on input (x,r) for the first time, samples the output according to the distribution of $\mathcal{O}(x)$.[8]

Proof (of Theorem 2). We first describe how we construct the distribution (X,Z) and the oracle $\mathcal{O}$.

[8] This can be shown along the lines of the proof that a random exponential size subset is unconditionally pseudorandom against exponential size distinguishers, see Goldreich's book "Foundations of Cryptography – Basic Techniques", Proposition 3.2.3.

Construction details. We chose at random two disjoint sets $\mathcal{X}_0, \mathcal{X}_1 \subset \{0,1\}^n$ of size 2^ℓ and define $\mathcal{X} = \mathcal{X}_0 \cup \mathcal{X}_1$. Let $\pi : \{0,1\}^n \to \{0,1\}^{m-1}$ be a random function. The oracle $\mathcal{O}$ on input $(x, j) \in \mathcal{X} \times \{0,1\}^{m-1}$ outputs a sample of $B_{1/2}$ (i.e., a uniformly random bit), except if $x \in \mathcal{X}$ and $\pi(x) = j$, in this case the output bit has bias δ; If $x \in \mathcal{X}_0$, the oracle outputs a sample of $B_{1/2-\delta}$, and otherwise, if $x \in \mathcal{X}_1$, a sample of $B_{1/2+\delta}$. We define the joint distribution (X, Z) by sampling $Z' \leftarrow \{0,1\}, X \leftarrow \mathcal{X}_{Z'}$ and setting $Z = \pi(X) \| Z'$ (note that X is uniform in $\mathcal{X}$)

Adversaries. The adversary on input $x \in \{0,1\}^n$ makes T non-adaptive queries $(x, j_1(x)), \ldots, (x, j_T(x))$ to the oracle. We denote $\mathcal{O}$'s response with $R(x) = \left(R^i(x, j_i(x))\right)_{i=1}^T$. The adversary's final output $f(x, R(x))$ is computed by a boolean function $f : \{0,1\}^n \times \{0,1\}^T \to \{0,1\}$.

Formal proof. Let $R(x) = (R^1(x, j_1(x)), \ldots, R^T(x, j_T(x)))$ be the sequences of the oracle's responses and Let $B(x) = (B^1_{1/2}, \ldots, B^T_{1/2})$ be independent random bits. For every x the number of *useful* responses, that is indexes i such that $R^i(x, j_i(x))$ is biased, is defined to be

$$T(x) = \sum_{i=1}^{T} [j_i(x) = \pi(x)] \tag{24}$$

On average we have $\mathbb{E}_{\mathcal{O}(\cdot)} T(x) = T/2^{m-1}$. We claim that the adversary actually learns basically nothing about $\mathcal{X}$: the sequence of oracle outptus is close to the sequence of unbiased bits. We start by showing that $\mathcal{X}$ is pseudorandom for our adversary.

Claim 3 (X is pseudorandom, even given oracle responses) *For any f and $\epsilon > 0$ we have*

$$\left| \mathop{\mathbb{E}}_{x \leftarrow \mathcal{X}} f(x, R(x)) - \mathop{\mathbb{E}}_{x \leftarrow U_n} f(x, R(x)) \right| \leq \epsilon + O\left(\delta^2 T / 2^m\right) \tag{25}$$

with error probability at most $O\left(\exp\left(-\Omega\left(2^{n-m}\right)\right) + \exp\left(-\Omega\left(2^\ell \epsilon^2\right)\right)\right)$.

Proof. By Lemma 3 and the definition of $\mathcal{O}$, for every $x \in \mathcal{X}$ we obtain

$$|\mathbb{E} f(x, R(x)) - \mathbb{E} f(x, B(x))| = \begin{cases} O\left(T(x)\delta^2\right), & x \in \mathcal{X} \\ 0, & x \notin \mathcal{X} \end{cases} \tag{26}$$

for every boolean function f and some absolute constant hidden under big-Oh. Thus

$$\left| \mathop{\mathbb{E}}_{x \leftarrow \mathcal{X}} f(x, R(x)) - \mathop{\mathbb{E}}_{x \leftarrow \mathcal{X}} f(x, B(x)) \right| = O\left(\mathop{\mathbb{E}}_{x \leftarrow \mathcal{X}} T(x)\delta^2 \right) \tag{27}$$

Note that the random variables $f(x, R(x))$ for different values of x are independent and similarly $f(x, B(x))$ for different values of x are independent. Since the

set $\mathcal{X}$ is chosen at random by the Hoeffding-Chernoff bound we obtain that with probability $1 - 2\exp\left(-\Omega\left(2^\ell\epsilon^2\right)\right)$ over $\mathcal{O}$ the following holds:

$$\left|\mathop{\mathbb{E}}_{x\leftarrow\mathcal{X}} f(x, B(x)) - \mathop{\mathbb{E}}_{x\leftarrow U_n} f(x, B(x))\right| \leqslant \epsilon \tag{28}$$

Combining Equation (27) and Equation (28) we obtain (with probability $1 - 2\exp\left(-\Omega\left(2^\ell\epsilon^2\right)\right)$ over $\mathcal{O}$)

$$\left|\mathop{\mathbb{E}}_{x\leftarrow\mathcal{X}} f(x, R(x)) - \mathop{\mathbb{E}}_{x\leftarrow U_n} f(x, B(x))\right| \leqslant \epsilon + O\left(\mathop{\mathbb{E}}_{x\leftarrow\mathcal{X}} T(x)\delta^2\right) \tag{29}$$

By Equation (26) we have

$$\left|\mathop{\mathbb{E}}_{x\leftarrow U_n} f(x, R(x)) - \mathop{\mathbb{E}}_{x\leftarrow U_n} f(x, B(x))\right| \leqslant O\left(\mathop{\mathbb{E}}_{x\leftarrow U_n} T(x)\delta^2\right). \tag{30}$$

Now Equations (29) and (30) imply

$$\left|\mathop{\mathbb{E}}_{x\leftarrow\mathcal{X}} f(x, R(x)) - \mathop{\mathbb{E}}_{x\leftarrow U_n} f(x, R(x))\right| \leqslant \epsilon + O\left(\mathop{\mathbb{E}}_{x\leftarrow U_n} T(x)\delta^2\right). \tag{31}$$

The random variables $T(x)$ for different x are independent, bounded by T and have the first moment $\mathbb{E}_{\mathcal{O}}(T(x)) = T/2^{m-1}$. By the multiplicative Chernoff bound with probability $1-2\exp\left(-\Omega\left(2^{n-m}\right)\right)$ over $\mathcal{O}$ it holds that $\mathbb{E}_{x\leftarrow U_n} T(x) < 2\cdot T/2^{m-1}$. This implies Equation (25) with error probability at most

$$P_{\text{err}} = O\left(\exp\left(-\Omega\left(2^{n-m}\right)\right) + \exp\left(-\Omega\left(2^\ell\epsilon^2\right)\right)\right).$$

Claim 4 *There exists a distinguisher* $\mathsf{D} : \{0,1\}^n \times \{0,1\}^m \to \{0,1\}$ *which calls the oracle* $\mathcal{O}$ *one time and such that for any joint distribution* Y, Z' *over* $\{0,1\}^n \times \{0,1\}^m$ *with entropy* $\widetilde{\mathbf{H}}_\infty(Y|Z') \geqslant \ell + 1$ *it holds that*

$$\mathbb{E}\,\mathsf{D}(X, Z) - \mathbb{E}\,\mathsf{D}(Y, Z') \geqslant \frac{\delta}{2}$$

with probability $1 - 2\exp(-\Omega\left(2^\ell\delta^2\right))$.

Remark 4 (Amplified distinguisher). Assuming that T is sufficiently large, we can modify D by taking the majority vote over T queries on $\mathcal{O}(x, z)$. This will boost the distinguishing advantage from $\delta/2$ to $C\delta$ where C can be an arbitrary constant (for sufficiently small δ).

Proof (of Claim 4). The distinguisher D simply calls the oracle $\mathcal{O}$ on the pair (x, z). The probability that D outputs 1 on input (Y, Z') is at most (the proba-

bilities below are over the choice of $\mathcal{O}$ and Y, Z')

$$\begin{aligned}
\Pr\left(\mathsf{D}(Y,Z')=1\right) &= \mathop{\mathbb{E}}_{z\leftarrow Z'} \Pr\left(\mathsf{D}(Y|_{Z'=z},z)=1\right)\\
&= \mathop{\mathbb{E}}_{z\leftarrow Z'} \left[\Pr\left(\mathsf{D}(Y,z)=1 \wedge Y\notin\mathcal{X}|\, Z'=z\right)\right] +\\
&\quad + \mathop{\mathbb{E}}_{z\leftarrow Z'} \left[\Pr\left(\mathsf{D}(Y,z)=1 \wedge Y\in\mathcal{X}|\, Z'=z\right)\right]\\
&= \frac{1}{2} + \delta \cdot \mathop{\mathbb{E}}_{z\leftarrow Z'} \left[\Pr\left(Y\in\mathcal{X}|\, Z'=z\right)\right]\\
&\leqslant \frac{1}{2} + \delta \mathop{\mathbb{E}}_{z\leftarrow Z'} \left[|\mathcal{X}|\cdot 2^{-\mathbf{H}_\infty\left(Y|Z'=z\right)}\right]\\
&= \frac{1}{2} + \delta\cdot|\mathcal{X}|\cdot 2^{-\widetilde{\mathbf{H}}_\infty(Y|Z')}
\end{aligned}$$

which is at most $\frac{1}{2}+\frac{\delta}{2}$. On the other hand we have $\Pr(\mathsf{D}(X,Z)=1)=\frac{1}{2}+\delta$. From this we see that the advantage is δ on average - but we need stronger concentration guarantees. Note that $\Pr(\mathsf{D}(X,Z)=1)=\sum_{x\in S}\Pr[X=x]\cdot \mathsf{D}(x,i(x))$ can be viewed as a sum of independent random variables. By the Chernoff-Hoeffding bound we get

$$\Pr_{\mathcal{O}}\left[\Pr(\mathsf{D}(X,Z)=1)\geqslant\frac{1}{2}+\delta-\frac{\delta}{8}\right]\geqslant 1-\exp(-\Omega\left(2^\ell\delta^2\right))$$

Similarly, $\Pr(\mathsf{D}(Y,Z')=1)=\sum_{x,z}\Pr[Y=x,Z'=z]\cdot\mathsf{D}(x,z')$. Since

$$\begin{aligned}
\sum_{x,z}\Pr[Y=x,Z'=z]^2 &= \sum_z\sum x\Pr[Z'=z]^2\Pr[Y=x|Z'=z]^2\\
&\leqslant \sum_z \Pr[Z'=z]2^{-\mathbf{H}_\infty(Y|_{Z'=z})}\\
&\leqslant 2^{-\widetilde{\mathbf{H}}_\infty(Y|Z)},
\end{aligned}$$

the Chernoff-Hoeffding bound implies

$$\Pr_{\mathcal{O}}\left[\Pr(\mathsf{D}(Y',Z)=1)\leqslant\frac{1}{2}+\frac{\delta}{2}+\frac{\delta}{8}\right]\geqslant 1-\exp(-\Omega\left(2^\ell\delta^2\right)) \tag{32}$$

and the result follows.

We set $\epsilon=\frac{\delta}{3}$ and $T=c\cdot 2^m/\epsilon$. Now Claim 4 directly implies Equation (5) whereas Equation (4) follows, when c is sufficiently small, from Claim 3 by a union bound; To see this, note that the right hand side of (32) is doubly exponentially close (in ℓ) to 1, and recall that $\ell > m+6\log(1/\delta)$. So we can take a union bound over all $O(\exp(T))$circuits D of size T and deduce that with high probability the left hand side of (32) hold for all of them.

C Proof of Lemma 3

Lemma 3 (Lower bounds on the coin problem). *Fix $\delta\in(0,1/2)$ and define $p=\frac{1}{2}+\delta$ and $q=1-p$. Consider the following two experiments:*

(a) *We flip a fair coin, and depending on the result we toss T times a biased coin B_p (probability of the head is p) or toss T times a coin B_q (probability of the head is q). The output is the result of these T flips.*

(b) *We flip T times a fair coin and output the results.*

Then one cannot distinguish (a) from (b) better than with advantage $O\left(T\delta^2\right)$.

Remark 5. We give a simple proof based on calculating Renyi divergences. This result can be also derived by more sophisticaed techniques from Fourier analysis (the generalized XOR lemma).

Before we give the proof, let's recall some basic facts about *Pearson Chi-Squared Distance*. For any two distributions P, Q over the same space, their Chi-Squared distance defined by

$$D_{\chi^2}(P \parallel Q) = \sum_x Q(x) \left(\frac{P(x)}{Q(x)} - 1\right)^2 = \sum_x \frac{P(x)^2}{Q(x)^2} - 1 \tag{33}$$

Now let $U_1, \ldots, U_n$ be independent uniform bits, $X_1, \ldots, X_n$ be i.i.d. bits where 1 appears with probability $p = \frac{1}{2} + \delta$ and $Y_1, \ldots, Y_n$ be i.i.d. bits where 1 appears w ith probability $q = 1 - p = \frac{1}{2} - \delta$. We want to estimate the distance between $U = U_1, \ldots, U_n$ and Z distributed as an equally weighted combination of $X = X_1, \ldots, X_n$ and $Y = Y_1, \ldots, Y_n$. We think of δ as a fixed parameter and n as a growing number. Our statement will easily follow by combining the following two claims

Claim 5 *With U and Z as above, and for $n = O\left(\delta^{-2}\right)$, it holds that*

$$D_{\chi^2}(U; Z) = O\left(n^2\delta^4\right) \tag{34}$$

Claim 6 *For any R and uniform U*

$$\mathrm{SD}(R \parallel U) \leqslant \sqrt{D_{\chi^2}(R \parallel U)}, \tag{35}$$

Indeed, combining these claims we obtain $\mathrm{SD}(Z \parallel U) = O(n\delta^2)$ when $n = O\left(\delta^{-2}\right)$. Since the left-hand side is bounded by 1, this is true also when $n > c\delta^{-2}$ for some absolute constant c and the result follows.

Proof (of Claim 5). We have

$$D_{\chi^2}\left(\frac{1}{2}P_{X_1,\ldots,X_n}+\frac{1}{2}P_{Y_1,\ldots,Y_n}\parallel P_{U_1}\cdot\ldots\cdot P_{U_n}\right)=$$
$$2^n\cdot\sum_{z_1,\ldots,z_n}\left(\frac{1}{2}P_{X_1}(z_1)\cdot\ldots\cdot P_{X_n}(z_n)+\frac{1}{2}P_{Y_1}(z_1)\cdot\ldots\cdot P_{Y_n}(z_n)\right)^2-1=$$
$$\frac{1}{4}\cdot 2^n\prod_i\left(\sum_z P_{X_i}(z)^2\right)+\frac{1}{4}\cdot 2\cdot 2^n\prod_i\left(\sum_z P_{X_i}(z)P_{Y_i}(z)\right)+$$
$$+\frac{1}{4}\cdot 2^n\prod_i\left(\sum_z P_{Y_i}(z)^2\right)-1=$$
$$\frac{1}{4}\left((1+4\delta^2)^n+2(1-4\delta^2)^n+(1+4\delta^2)^n-4\right) \tag{36}$$

and the result follows by the Taylor expansion $(1+u)^n=1+nu+O(n^2u^2)$ where $nu=O(1)$ applied to $u=4\delta^2$. The bound is valid as long as $n=O\left(\delta^{-2}\right)$.

Proof (of Claim 6). This inequality follows immediately from the Cauchy-Schwarz inequality and the definition of D_{χ^2}.

Chapter 4

Simulating Auxiliary Information

Simulating Auxiliary Inputs, Revisited *

Maciej Skorski **

maciej.skorski@mimuw.edu.pl
University of Warsaw

Abstract. For any pair (X, Z) of correlated random variables we can think of Z as a randomized function of X. If the domain of Z is small, one can make this function computationally efficient by allowing it to be only approximately correct. In folklore this problem is known as *simulating auxiliary inputs*. This idea of simulating auxiliary information turns out to be a very usefull tool, finding applications in complexity theory, cryptography, pseudorandomness and zero-knowledge. In this paper we revisit this problem, achieving the following results:

(a) We present a novel boosting algorithm for constructing the simulator. This boosting proof is of independent interest, as it shows how to handle "negative mass" issues when constructing probability measures by shifting distinguishers in descent algorithms. Our technique essentially fixes the flaw in the TCC'14 paper "How to Fake Auxiliary Inputs".

(b) The complexity of our simulator is better than in previous works, including results derived from the uniform min-max theorem due to Vadhan and Zheng. To achieve (s, ϵ)-indistinguishability we need the complexity $O\left(s \cdot 2^{5\ell}\epsilon^{-2}\right)$ in time/circuit size, which improve previous bounds by a factor of ϵ^{-2}. In particular, with we get meaningful provable security for the EUROCRYPT'09 leakage-resilient stream cipher instantiated with a standard 256-bit block cipher, like AES256.

Our boosting technique utilizes a two-step approach. In the first step we shift the current result (as in gradient or sub-gradient descent algorithms) and in the separate step we fix the biggest non-negative mass constraint violation (if applicable).

Keywords: simulating auxiliary inputs, boosting, leakage-resilient cryptography, stream ciphers, computational indistinguishability

1 Introduction

1.1 Simulating Correlated Information.

Informal Problem Statement Let $(X, Z) \in \mathcal{X} \times \mathcal{Z}$ be a pair of correlated random variables. We can think of Z as a *randomized* function of X. More

* The full (and updated) version of this paper is available at the Cryptology ePrint archive and the arXiv archive ({http://arxiv.org/abs/1503.00484}).

** Supported by the National Science Center, Poland (2015/17/N/ST6/03564).

precisely, consider the randomized function $h : \mathcal{X} \to \mathcal{Z}$, which for every x outputs z with probability $\Pr[Z = z|X = x]$. By definition it satisfies

$$(X, h(X)) \stackrel{d}{=} (X, Z) \tag{1}$$

however the function h is *inefficient* as we need to hardcode the conditional probability table of $Z|X$. It is natural to ask, if this limitation can be overcome

Q1: Can we represent Z as an *efficient* function of X?

Not surprisingly, it turns out that a positive answer may be given only in computational settings. Note that replacing the equality in Equation (1) by closeness in the total variation distance (allowing the function h to make some mistakes with small probability) is not enough [1]. This discussion leads to the following reformulated question

Q1': Can we *efficiently simulate* Z as a function of X?

Why it matters? Aside from being very foundational, this question is relevant to many areas of computer science. We will not discuss these applications in detail, as they are well explained in [JP14]. Below we only mention where such a generic simulator can be applied, to show that this problem is indeed well-motivated.

(a) Complexity Theory. From the simulator one can derive Dense Model Theorem [RTTV08], Impagliazzo's hardcore lemma [Imp95] and a version of Szemeredis Regularity Lemma [FK99].
(b) Cryptography. The simulator can be applied for settings where Z models short leakage from a secret state X. It provides tools for improving and simplifying proofs in leakage-resilient cryptography, in particular for leakage-resilient stream ciphers [JP14].
(c) Pseudorandomness. Using the simulator one can conclude results called chain rules [GW11], which quantify pseudorandomness in conditioned distributions. They can be also applied to leakage-resilient cryptography.
(d) Zero-knowledge. The simulator can be applied to represent the text exchanged in verifier-prover interactions Z from the common input X [CLP15].

Thus, the simulator may be used as a tool to unify, simplify and improve many results. Having briefly explained the motivation we now turn to answer the posed question, leaving a more detailed discussion of some applications to Section 1.6.

[1] Indeed, consider the simplest case $\mathcal{Z} = \{0, 1\}$, define X to be uniform over $\mathcal{X} = \{0, 1\}^n$, and take $Z = f(X)$ where f is a function which is 0.5-hard to predict by circuits exponential in n, Then $(X, h(X))$ and (X, Z) are at least $\frac{1}{4}$-away in total variation

1.2 Problem Statement

The problem of simulating auxiliary inputs in the computational setting can be defined precisely as follows

> Given a random variables $X \in \{0,1\}^n$ and correlated $Z \in \{0,1\}^\ell$, what is the minimal complexity s_h of a (randomized) function h such that the distributions of $h(X)$ and Z are (ϵ, s)-indistinguishable given X, that is
>
> $$|\,\mathbb{E}\,\mathrm{D}(X, h(X)) - \mathbb{E}\,\mathrm{D}(X, Z)| < \epsilon$$
>
> holds for all (deterministic) circuits D of size s?

The indistinguishability above is understood with respect to deterministic circuits. However it doesn't really matter for distinguishing two distributions, where randomized and deterministic distinguishers are equally powerful[2].

It turns out that it is relatively easy[3] to construct a simulator h with a polynomial blowup in complexity, that is when

$$s_h = \mathrm{poly}\left(s, \epsilon^{-1}, 2^\ell\right).$$

However, more challenging is to minimize the dependency on ϵ^{-1}. This problem is especially important for cryptography, where security definitions require the advantage ϵ to be possibly small. Indeed, for meaningful security $\epsilon = 2^{-80}$ or at least $\epsilon = 2^{-40}$ it makes a difference whether we lose ϵ^{-2} or ϵ^{-4}. We will see later how much inefficient bounds here may affect provable security of stream ciphers.

1.3 Related Works

Original work of Jetchev and Pietrzak (TCC'14) The authors showed that Z can be "approximately" computed from X by an "efficient" function h.

Theorem 1 ([JP14], corrected). *For every distribution (X, Z) on $\{0,1\}^n \times \{0,1\}^\ell$ and every ϵ, s, there exists a "simulator" $h : \{0,1\}^n \to \{0,1\}^\ell$ such that*

(a) $(X, h(X))$ and (X, Z) are (ϵ, s)-indistinguishable
(b) h is of complexity $s_{\mathsf{h}} = O\left(s \cdot 2^{4\ell}\epsilon^{-4}\right)$

The proof uses the standard min-max theorem. In the statement above we correct two flaws. One is a missing factor of 2^ℓ. The second (and more serious) one is the (corrected) factor ϵ^{-4}, claimed incorrectly to be ϵ^{-2}. The flaws are discussed in Appendix A.

[2] If two distributions can be distinguished by a randomized circuit, we can fix a specific choice of coins to achieve at least the same advantage

[3] We briefly sketch the idea of the proof: note first that it is easy to construct a simulator for every single distinguisher. Having realized that, we can use the min-max theorem to switch the quantifiers and get one simulator for all distinguishers.

Vadhan and Zheng (CRYPTO'13) The authors derived a version of Theorem 1 but with incomparable bounds

Theorem 2 ([VZ13]). *For every distribution X, Z on $\{0,1\}^n \times \{0,1\}^\ell$ and every ϵ, s, there exists a "simulator" $h : \{0,1\}^n \to \{0,1\}^\ell$ such that*

(a) $(X, h(X))$ and (X, Z) are (s, ϵ)-indistinguishable
(b) h is of complexity $s_{\mathsf{h}} = O\left(s \cdot 2^\ell \epsilon^{-2} + 2^\ell \epsilon^{-4}\right)$

The proof follows from a general regularity theorem which is based on their uniform min-max theorem. The additive loss of $O\left(2^\ell \epsilon^{-4}\right)$ appears as a consequence of a sophisticated weight-updating procedure. This error is quite large and may dominate the main term for many settings (whenever $s \ll \epsilon^{-2}$).

As we show later, Theorem 2 and Theorem 1 give in fact comparable security bounds when applied to leakage-resilient stream ciphers (see Section 1.6)

1.4 Our Results

We reduce the dependency of the simulator complexity s_h on the advantage ϵ to only a factor of ϵ^{-2}, from the factor of ϵ^{-4}.

Theorem 3 (Our Simulator). *For every distribution X, Z on $\{0,1\}^n \times \{0,1\}^\ell$ and every ϵ, s, there exists a "simulator" $h : \{0,1\}^n \to \{0,1\}^\ell$ such that*

(a) $(X, h(X))$ and (X, Z) are (s, ϵ)-indistinguishable
(b) h is of complexity $s_{\mathsf{h}} = O\left(s \cdot 2^{5\ell} \log(1/\epsilon) \epsilon^{-2}\right)$

Below in Table 1 we compare our result to previous works.

Author	Technique	Advantage	Size	Cost of simulating
[JP14] (Theorem 1)	Min-Max	ϵ	s	$s_{\mathsf{h}} = O\left(s \cdot 2^{4\ell} \epsilon^{-4}\right)$
[VZ13] (Theorem 2)	Complicated Boosting			$s_{\mathsf{h}} = O\left(s \cdot 2^\ell / \epsilon^2 + 2^\ell \epsilon^{-4}\right)$
This paper (Theorem 3)	Simple Boosting			$s_{\mathsf{h}} = O\left(s \cdot 2^{5\ell} \epsilon^{-2}\right)$

Table 1. The complexity of simulating ℓ-bit auxiliary information given required indistinguishability strength, depending on the proof technique. For simplicity, terms polylog$(1/\epsilon)$ are omitted.

Our result is slightly worse in terms of dependency ℓ, but outperforms previous results in terms of dependency on ϵ^{-1}. However, the second dependency is more crucial for cryptographic applications. Note that the typical choice is sub-logarithmic leakage, that is $\ell = o\left(\log \epsilon^{-1}\right)$ is asymptotic settings[4] (see for example [CLP15]). Stated in non-asymptotic settings this assumption translates to $\ell < c \log \epsilon^{-1}$ where c is a small constant (for example $c = \frac{1}{12}$ see [Pie09]). In these settings, we outperform previous results.

[4] This is a direct consequence of the fact that we want ℓ to fit poly-preserving reductions

To illustrate this, suppose we want to achieve security $\epsilon = 2^{-60}$ simulating just one bit from a 256-bit input. As it follows from Table 1, previous bounds are useless as they give the complexity bigger than 2^{256} which is the worst complexity of all boolean functions over the chosen domain. In settings like this, only our bound can be applied to conclude meaningful results. For more concrete examples of settings where our bounds are even only meaningful, we refer to Table 2 in Section 1.6.

1.5 Our Techniques

Our approach utilizes a simple boosting technique: as long as the condition (a) in Theorem 3 fails, we can use the distinguisher to improve the simulator. This makes our algorithm constructive with respect to distinguishers obtained from an oracle[5], similarly to other boosting proofs [JP14,VZ13]. In short, if for a "candidate" solution h there exists D such that

$$\mathbb{E}\,\mathrm{D}(X,Z) - \mathbb{E}\,\mathrm{D}(X,h(X)) > \epsilon$$

then we construct a new solution h' using D and h, according to the equation[6]

$$\Pr[h'(x)=z] = \Pr[h(x)=z] + \gamma \cdot \mathsf{Shift}\,(\mathrm{D}(x,z)) + \mathsf{Corr}(x,z)$$

where

(a) The parameter γ is a *fixed step* chosen in advance (its optimal value depends on ϵ and ℓ and is calculated in the proof.)
(b) $\mathsf{Shift}\,(\mathrm{D}(x,z))$ is a *shifted* version of D, so that $\sum_z \mathsf{Shift}\,(\mathrm{D}(x,z)) = 0$. This restriction correspond to the fact that we want to preserve the constraint $\sum_z h(x,z) = 1$. More precisely, $\mathsf{Shift}\,(\mathrm{D}(x,z)) = \mathrm{D}(x,z) - \mathbb{E}_{z' \leftarrow U_\ell}\,\mathrm{D}(x,z)$
(c) $\mathsf{Corr}(x,z)$ is a *correction term* used to fix (some of) possibly negative weights.

The procedure is being repeated in a loop, over and over again. The main technical difficulty is to show that it eventually stops after not so many iterations.

Note that in every such a step the complexity cost of the shifting term is $O\left(2^\ell \cdot \mathrm{size}(\mathrm{D})\right)$[7]. The correction term, in our approach, does a search over z looking for the biggest negative mass, and redistributes it over the remaining points. Intuitively, it works because the total negative mass is getting smaller with every step. See Algorithm 1 for a pseudo-code description of the algorithm and the rest of Section 3 for a proof.

[5] The oracle evaluates the distance of the given candidate solution and the simulated distribution, answering with a distiguisher if the distance is smaller than required.
[6] As we already mentioned, we can assume that D is deterministic without loss of generality. Then all the terms in the equation are well-defined.
[7] By definition, it requires computing the average of $\mathrm{D}(x,\cdot)$ over 2^ℓ elements

1.6 Applications

Better security for the EUROCRYPT'09 stream cipher. The first construction of leakage-resilient stream cipher was proposed by Dziembowski and Pietrzak in [DP08]. On Figure 1 below we present a simplified version of this cipher [Pie09], based on a weak pseudorandom function (wPRF).

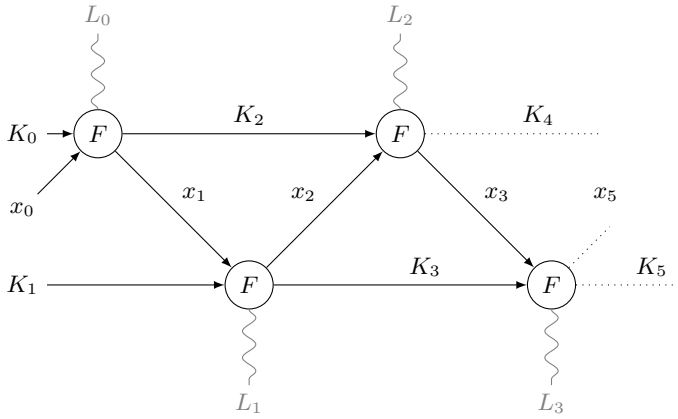


Fig. 1. The EUROCRYPT'09 stream cipher (adaptive leakage). F denotes a weak pseudorandom function. By K_i and x_i we denote, respectively, values of the secret state and keystream bits. Leakages are denotted in gray with L_i.

Jetchev and Pietrzak in [JP14] showed how to use the simulator theorem to simplify the security analysis of the EUROCRYPT'09 cipher. The cipher security depends on the complexity of the simulator as explained in Theorem 1 and Remark 2. We consider the following setting:

- number of rounds $q = 16$,
- F instantiated with AES256 (as in [JP14])
- cipher security we aim for $\epsilon' = 2^{-40}$
- $\lambda = 3$ bits of leakage per round

The concrete bounds for (q, ϵ', s')-security of the cipher (which roughly speaking means that q consecutive outputs is (s', ϵ')-pseudorandom, see Section 2 for a formal definition) are given in Table 2 below. We ommit calculations as they are merely putting parameters from Theorem 1, Theorem 2 and Theorem 3 into Remark 2 and assuming that AES as a weak PRF is (ϵ, s)-secure for any pairs $s/\epsilon \approx 2^k$ (following the similar example in [JP14]).

More generally, we can give the following comparison of security bounds for different wPRF-based stream ciphers, in terms of time-sccess ratio. The bounds in Table 3 follow from the simple lemma in Section 4, which shows how the time-success ratio changes under explicit reduction formulas.

Analysis/Authors	wPRF security	Leakage	Advantage ϵ'	Size s'
[JP14] (Theorem 1)	256	$\lambda = 3$	2^{-40}	0
[VZ13] (Theorem 2)				0
this paper (Theorem 3)				2^{66}

Table 2. The security of the EUROCRYPT'09 stream cipher, instantiated with AES256 as a weak PRF of rouhgly $k = 256$ bits of security. In this settngs only our new bounds provide non-trivial bounds.

Cipher	Analysis	Proof techniques	Security level	Comments
(1)	[Pie09]	Pseudoentropy chain rules	$k' \ll \frac{1}{8}k$	large number of blocks
(1)	[JP14]	Aux. Inputs Simulator (corr.)	$k' \approx \frac{k}{6} - \frac{5}{6}\lambda$	
(1)	[VZ13]	Aux. Inputs Simulator	$k' \approx \frac{k}{6} - \frac{1}{3}\lambda$	
(1)	**This work**	Aux. Inputs Simulator	$k' \approx \frac{k}{4} - \frac{4}{3}\lambda$	
(2)	[FPS12]	Pseudoentropy chain rules	$k' \approx \frac{k}{5} - \frac{3}{5}\lambda$	large public seed
(3)	[YS13]	Square-friendly apps.	$k' \approx \frac{k}{4} - \frac{3}{4}\lambda$	only in minicrypt

Table 3. Different bounds for wPRF-based leakage-resilient stream ciphers. k is the security level of the underlying wPRF. The value k' is the security level for the cipher, understood in terms of time-success ratio. the numbers denote: (1) The EUROCRYPT'09 cipher, (2) The CSS'10/CHESS'12 cipher, (3) The CT-RSA'13 cipher.

1.7 Organization

In Section 2 we discuss basic notions and definitions. The proof of Theorem 3 appears in Section 3.

2 Preliminaries

2.1 Notation

By $\mathbb{E}_{y \leftarrow Y} f(y)$ we denote an expectation of f under y sampled according to the distribution Y.

2.2 Basic Notions

Indistinguishability Let $\mathcal{V}$ be a finite set, and $\mathcal{D}$ be a class of deterministic $[0,1]$-valued functions on $\mathcal{V}$. For any two real functions f_1, f_2 on $\mathcal{V}$, we say that f_1, f_2 are $(\mathcal{D}, \epsilon)$-indistinguishable if

$$\forall \mathrm{D} \in \mathcal{D} : \quad \left| \sum_{x \in \mathcal{V}} \mathrm{D}(x) \cdot f_1(x) - \sum_{x \in \mathcal{V}} \mathrm{D}(x) \cdot f_2(x)) \right| \leqslant \epsilon$$

Note that the domain $\mathcal{V}$ depends on the context. If X_1, X_2 are two probability distributions, we say that they are (s, ϵ)-indistinguishable if their probability

mass functions are indistinguishable, that is when

$$\left| \sum_{x \in V} \mathrm{D}(x) \cdot \Pr[X_1 = x] - \sum_{x \in V} \mathrm{D}(x) \cdot \Pr[X_2 = x] \right| \leqslant \epsilon$$

for all $\mathrm{D} \in \mathcal{D}$. If $\mathcal{D}$ consists of all circuits of size s we say that f_1, f_2 are (s, ϵ)-indistinguishable.

Remark 1. This an extended notion of indistinguishability, borrowed from [TTV09], which captures not only probability measures but also real-valued functions. A good intuition is provided by the following observation [TTV09]: think of functions over $\mathcal{V}$ as $|\mathcal{V}|$-dimensional vectors then $\epsilon \geqslant |\sum_{x \in V} \mathrm{D}(x) \cdot f_1(x) - \sum_{x \in V} \mathrm{D}(x) \cdot f_2(x)| = |\langle f_1 - f_2, \mathrm{D} \rangle|$ means that f_1 and f_2 are *nearly orthogonal* for all test functions in $\mathcal{D}$.

Distinguishers In the definition above we consider deterministic distinguishers, as this is required by our algorithm. However, being randomized doesn't help in distinguishing, as any randomized-distinguisher achieving advantage ϵ when run on two fixed distributions can be converted into a deterministic distinguishers of the same size and advantage (by fixing one choice of coins). Moreover, any real-valued distinguisher can be converted, by a boolean threshold, into a boolean one with at least the same advantage [FR12].

Relative complexity We say that a function h has complexity at most T relative to the set of functions $\mathcal{D}$ if there are functions $\mathrm{D}_1, \ldots, \mathrm{D}_T$ such h can be computed by combining them using at most T of the following operations: (a) multiplication by a constant, (b) application of a boolean threshold function, (c) sum, (d) product.

2.3 Stream ciphers definitions

We start with the definition of weak pseudorandom functions, which are *computationally indistinguishable* from random functions, when queried on random inputs and fed with uniform secret key.

Definition 1 (Weak pseudorandom functions). *A function* $\mathrm{F} : \{0,1\}^k \times \{0,1\}^n \to \{0,1\}^m$ *is an* (ϵ, s, q)*-secure weak PRF if its outputs on* q *random inputs are indistinguishable from random by any distinguisher of size* s*, that is*

$$|\Pr\left[\mathrm{D}\left((X_i)_{i=1}^q, \mathrm{F}((K, X_i)_{i=1}^q) = 1\right] - \Pr\left[\mathrm{D}\left((X_i)_{i=1}^q, (R_i)_{i=1}^q\right) = 1\right]| \leqslant \epsilon$$

where the probability is over the choice of the random $X_i \leftarrow \{0,1\}^n$*, the choice of a random key* $K \leftarrow \{0,1\}^k$ *and* $R_i \leftarrow \{0,1\}^m$ *conditioned on* $R_i = R_j$ *if* $X_i = X_j$ *for some* $j < i$*.*

Stream ciphers generate a keystream in a recursive manner. The security requires the output stream should be indistinguishable from uniform[8].

[8] We note that in a more standard notion the entire stream $X_1, \ldots, X_q$ is indistinguishable from random. This is implied by the notion above by a standard hybrid argument, with a loss of a multiplicative factor of q in the distinguishing advantage.

Definition 2 (Stream ciphers). *A* stream-cipher $\mathsf{SC} : \{0,1\}^k \to \{0,1\}^k \times \{0,1\}^n$ *is a function that, when initialized with a secret state* $S_0 \in \{0,1\}^k$*, produces a sequence of output blocks* $X_1, X_2, ...$ *computed as*

$$(S_i, X_i) := \mathsf{SC}(S_{i-1}).$$

A stream cipher SC *is* (ϵ, s, q)*-secure if for all* $1 \leqslant i \leqslant q$*, the random variable* X_i *is* (s, ϵ)*-pseudorandom given* $X_1, ..., X_{i-1}$ *(the probability is also over the choice of the initial random key* S_0*).*

Now we define leakage resilient stream ciphers, following the "only computation leaks" assumption.

Definition 3 (Leakage-resilient stream ciphers). *A leakage-resilient stream-cipher is* $(\epsilon, s, q, \lambda)$*-secure if it is* (ϵ, s, q)*-secure as defined above, but where the distinguisher in the* j*-th round gets* λ *bits of arbitrary deceptively chosen leakage about the secret state accessed during this round. More precisely, before* $(S_j, X_j) := \mathsf{SC}(S_{j-1})$ *is computed, the distinguisher can choose any leakage function* f_j *with range* $\{0,1\}^\lambda$*, and then not only get* X_j*, but also* $\Lambda_j := f_j(\hat{S}_{j-1})$*, where* $\hat{S}_{j-1}$ *denotes the part of the secret state that was modified (i.e., read and/or overwritten) in the computation* $\mathsf{SC}(S_{j-1})$*.*

2.4 Security of leakage-resilient stream ciphers.

Best provable secure constructions of leakage-resilient stream ciphers are based on so called weak PRFs, primitives which look random when queried on random inputs ([Pie09,FPS12,JP14,DP10,YS13]). The most recent (TCC'14) analysis is based on a version of Theorem 1.

Theorem 4 (Proving Security of Stream Ciphers [JP14]). *If* F *is a* $(\epsilon_F, s_F, 2)$*-secure weak PRF then* SC^F *is a* $(\epsilon', s', q, \lambda)$*-secure leakage resilient stream cipher where*

$$\epsilon' = 4q\sqrt{\epsilon_F 2^\lambda}, \quad s' = \Theta(1) \cdot \frac{s_F \epsilon'^4}{2^{4\lambda}}.$$

Remark 2 (The exact complexity loss). An inspection of the proof in [JP14] shows that s_F equals the complexity of the simulator h in Theorem 1, with circuits of size s' as distingusihers and ϵ replaced by ϵ'.

2.5 Time-Success Ratio

The running time (circuit size) s and success probability ϵ of attacks (practical and theoretical) against a particular primitive or protocol may vary. For this reason Luby [LM94] introduced the time-success ratio $\frac{t}{\epsilon}$ as a universal measure of security. This model is widely used to analyze provable security, cf. [BL13] and related works.

Definition 4 (Security by Time-Success Ratio [LM94]). *A primitive P is said to be 2^k-secure if for* every adversary *with time resources (circuit size in the nonuniform model) s, the success probability in breaking P (advantage) is at most $\epsilon < s \cdot 2^{-k}$. We also say that the time-success ratio of P is 2^k, or that is has k bits of security.*

For example, AES with a 256-bit random key is believed to have 256 bits of security as a *weak* PRF[9].

3 Proof of Theorem 3

For technical convenience, we attempt to efficiently approximate the conditional probability function $g(x,z) = \Pr[Z = z | X = x]$, rather than building the sampler directly. Once we end with building an efficient approximation $h(x,z)$, we transform it into a sampler h_{sim} which on input x outputs z with probability $h(x,z)$ (this transformation yields only a loss of $2^\ell \log(1/\epsilon)$). Let $\mathcal{X} = \{0,1\}^n$ and $\mathcal{Z} = \{0,1\}^\ell$. We are going to prove the following fact

> For every function g on $\mathcal{X} \times \mathcal{Z}$ which is a $\mathcal{X}$-conditional probability mass function over $\mathcal{Z}$ (that is $g(x,z) \geqslant 0$ for all x, z and $\sum_z g(x,z) = 1$ for every x), and for every class $\mathcal{D}$ of bounded real functions on $\mathcal{X} \times \mathcal{Z}$, closed under complements[10] there exists h such that
> (a) h is an $\mathcal{X}$-conditional probability mass function over $\mathcal{Z}$
> (b) h is of complexity $s_h = O(2^{4\ell}\epsilon^{-2})$ with respect to $\mathcal{D}$
> (c) (X, Z) and $(X, h(X))$ are $\mathcal{D}$-indistinguishable, that is
>
> $$\left| \sum_{z \in \mathcal{Z}} \mathop{\mathbb{E}}_{x \sim X} \left[\mathrm{D}(x,z) \cdot (g(x,z) - h(x,z)) \right] \right| \leqslant \epsilon \qquad (2)$$

The sketch of the construction is shown in Algorithm 1. Here we would like to point out two things. First, we stress that we do not produce a strictly positive function; what our algorithm guarantees, is that the total negative mass is *small* but as we will see later this is enough. Second, our algorithm performs essentially same operations for every x, which is why its complexity depends only on $\mathcal{Z}$. Overall, the procedure has a form of the standard *learning algorithm* which iteratively updates the candidate solution; information on how to improve are provided by distinguishers (see Line 3 and after). The algorithm is actually little more technical because of handling constraints on the conditional probability mass function (see Line 7 and after) .

We denote for shortness $\overline{\mathrm{D}}(x,z) = \mathrm{D}(x,z) - \mathbb{E}_{z' \leftarrow U_{\mathcal{Z}}} \mathrm{D}(x,z')$ for any D (the "shift" transformation)

[9] We consider the security of AES256 as a weak PRF, and not a standard PRF, because of non-uniform attacks which show that no PRF with a k-bit key can have $s/\epsilon \approx 2^k$ security [DTT09], at least unless we additionally require $\epsilon \gg 2^{-k/2}$.

[10] This is a standard assumption in indistinguishability proofs. We can always extend the class by adding $-\mathrm{D}$ for every $\mathrm{D} \in \mathcal{D}$, which increases the complexity only by 1.

Algorithm 1: Construct the Auxiliary Inputs Simulator

input : $\mathcal{X}$-conditional probability function $g : \mathcal{X} \times \mathcal{Z} \to [0,1]$, accuracy parameter $\epsilon > 0$, class $\mathcal{D}$ of bounded functions on $\mathcal{X} \times \mathcal{Z}$, step γ

output: Function h which is ϵ-indistinguishable from g under $\mathcal{D}$, add up to 1 for every x, and with total negative mass smaller than $O(\gamma|\mathcal{Z}|^3)$

1 $t \leftarrow 0$

2 $h^0(x,z) \leftarrow \frac{1}{|\mathcal{Z}|}$ for every x and z

/* as long as Equation (2) fails - simulator not good enough */

while *exists* $\mathrm{D} \in \mathcal{D}$ *s.t.* $\mathbb{E}_{x\sim X}\left[\sum_z \overline{\mathrm{D}}(x,z)\cdot\left(g(x,z') - h^t(x,z')\right)\right] \geqslant \epsilon$ **do**

3 $\quad \mathrm{D}^{t+1} \leftarrow \mathrm{D}$

4 $\quad$ **for** $z' \in \mathcal{Z}$ **do** /* modify using the distinguisher */

5 $\quad\quad h^{t+1}(x,z') \leftarrow h^t(x,z') + \gamma \cdot \overline{\mathrm{D}^{t+1}}(x,z')$

6 $\quad t \leftarrow t+1$

7 $\quad m \leftarrow 0$

8 $\quad$ **for** $z' \in \mathcal{Z}$ **do** /* locate the biggest negative point mass */

9 $\quad\quad$ **if** $h^t(x,z') < m$ **then**

10 $\quad\quad\quad m \leftarrow h^t(x,z')$

11 $\quad\quad\quad z^- \leftarrow z'$

12 $\quad h^t(x,z^-) = 0$ /* cut the biggest negative mass */

13 $\quad$ **for** $z' \in \mathcal{Z}$ **do**

14 $\quad\quad h^t(x,z') \leftarrow h^t(x,z') + \frac{m}{|\mathcal{Z}|-1}$ /* redestribute the cut mass */

15 **return** $h^t(x,z)$

Proof. Consider the functions h^t. Define $\tilde{h}^{t+1}(x,z) \stackrel{def}{=} h^t(x,z) + \gamma \cdot \overline{\mathrm{D}}^{t+1}(x,z)$. According to Algorithm 1, we have

$$h^{t+1}(x,z) = h^t(x,z) + \gamma \cdot \overline{\mathrm{D}}^{t+1}(x,z) + \theta^{t+1}(x,z) \tag{3}$$

with the correction term $\theta^t(x,z)$ that be computed recursively as (see Line 12 in Algorithm 1)

$$\theta^t(x,z) = 0$$
$$\theta^t(x,z) = \begin{cases} -\min\left(h^t(x,z) + \gamma \cdot \overline{\mathrm{D}}^{t+1}(x,z), 0\right), & \text{if } z = z^t_{\min}(x) \\ \frac{\min\left(h^t(x,z^t_{\min}(x))) + \gamma \cdot \overline{\mathrm{D}}^{t+1}(x,z^t_{\min}(x)), 0\right)}{\#\mathcal{Z}-1} & \text{if } z \neq z^t_{\min}(x) \end{cases} \qquad t = 0,1,\ldots \tag{4}$$

where $z^t_{\min}(x)$ is one of the points z minimizing $h^t(x,z) + \gamma \cdot \overline{\mathrm{D}}^{t+1}(x,z)$ (chosen and fixed for every t) . In particular

$$h^t(x, z^t_{\min}(x))) + \gamma \cdot \overline{\mathrm{D}}^{t+1}(x, z^t_{\min}(x)) < 0 \iff \exists z : \; h^t(x,z) + \gamma \cdot \overline{\mathrm{D}}^{t+1}(x,z) < 0 \tag{5}$$

Notation: for notational convenience we indenify the functions $\mathrm{D}^t(x,z)$, $\overline{\mathrm{D}}^t(x,z)$, $\theta^t(x,z)$, $\tilde{h}^t(x,z)$ and $h^t(x,z)$ with matrices where x are columns and z are rows.

That is h_x^t denotes the $|\mathcal{Z}|$-dimensional vector with entries $h^t(x,z)$ for $z \in \mathcal{Z}$ and similarly for other functions $\mathrm{D}^t(x,z)$, $\overline{\mathrm{D}}^t(x,z)$, $\theta^t(x,z)$, $\tilde{h}^t(x,z)$.

Claim 1 (Complexity of Algorithm 1). T executions of the "while loop" can be realized with time $O\left(T \cdot |\mathcal{Z}| \cdot \mathrm{size}(\mathcal{D})\right)$ and memory $O(|\mathcal{Z}|)$. [11].

This claim describes precisely resources required to compute the function h^T for every T. In order to bound T, we define the energy function as follows:

Claim 2 (Energy function). Define the auxiliary function

$$\Delta^t = \sum_{i=0}^{t-1} \mathop{\mathbb{E}}_{x\sim X} \left[\overline{\mathrm{D}}_x^{i+1} \cdot \left(g_x - h_x^i\right)\right]. \tag{6}$$

Then we have $\Delta^t = E_1 + E_2$ where

$$\begin{aligned} E_1 &= \tfrac{1}{\gamma}\,\mathbb{E}_{x\sim X}\left[\left(h_x^t - h_x^0\right)\cdot g_x + \tfrac{1}{2}\textstyle\sum_{i=0}^{t-1}\left(h_x^{i+1} - h_x^i\right)^2 - \tfrac{1}{2}\left(\left(h_x^t\right)^2 - \left(h_x^0\right)^2\right)\right] \\ E_2 &= \tfrac{1}{\gamma}\,\mathbb{E}_{x\sim X}\left[-\textstyle\sum_{i=0}^{t-1}\theta_x^{i+1}\cdot\left(g_x - h_x^{i+1}\right) - \sum_{i=0}^{t-1}\theta_x^{i+1}\cdot\left(h_x^{i+1} - h_x^i\right)\right] \end{aligned} \tag{7}$$

Note that all the symbols represent vectors and multiplications, including squares, should be understood as scalar products. The proof is based on simple algebraic manipulations and appears in Appendix B.

Remark 3 (Technical issues and intuitions). To upper-bound the formulas in Equation (7), we need the following important properties

(a) *Boundedness of correction terms*, that is ideally $|\theta^i(x.z)| = O(\mathrm{poly}(|\mathcal{Z}|)\cdot\gamma)$.
(b) *Acute angle between the correction and the error*, that is $\theta_x^i \cdot (g_x - h_x^i) \geqslant 0$.

Below we present an outline of the proof, discussing more technical parts in the appendix.

Proof outline. Indeed, with these assumptions we prove an upper bound on the energy function, namely

$$E_1 + E_2 \leqslant O\left(\mathrm{poly}(|\mathcal{Z}|)\cdot\left(t\gamma + \gamma^{-1}\right)\right), \tag{8}$$

which follows from the properties (a) and (b) above (they are proved in Claim 4 and Claim 3 below, and the inequality on $E_1 + E_2$ is derived in Claim 5). Note that, except a factor $\mathrm{poly}(|\mathcal{Z}|)$, our formula (not the proof, though) is identical to the bound used in [TTV09] (see Claim 3.4 in the eprint version). Indeed, our theorem is, to some extent, an extension to the main result in [TTV09] to cover the conditional case, where $|\mathcal{X}| > 1$. The main difference is that we show how to simulate a short leakage $|Z|$ given X, whereas [TTV09] shows how to simulate

[11] The RAM model

Z alone, under the assumption that the distribution of Z is dense in the uniform distribution (the min-entropy gap being small)[12].

Since the bound above is valid for any step t, and since on the other hand we have $t\epsilon \leqslant \Delta^t$ after t steps of the algorithm, we achieve a contradiction (to the number of steps) setting $\gamma = \epsilon/\mathrm{poly}(|\mathcal{Z}|)$. Indeed, suppose that $t\epsilon \leqslant A|\mathcal{Z}|^B(\gamma^{-1} + t\gamma)$ for some positive constants A, B. Since the step size γ can be chosen arbitrarily, we can set $\gamma = \frac{\epsilon}{2A|\mathcal{Z}|^B}$ which yields $\frac{t\epsilon}{2} \leqslant \frac{2A^2|\mathcal{Z}|^B}{\epsilon}$ or $t \leqslant 4A^2|\mathcal{Z}|^B\epsilon^{-2}$, which means that the algorithm terminates after at most $T = \mathrm{poly}(|\mathcal{Z}|)\epsilon^{-2}$ steps. Our proof goes exactly this way, except some extra optimization do obtain better exponent A.

We stress that it outputs only a *signed measure*, not a probability distribution yet. However, because of property (a) the negative mass is only of order $\mathrm{poly}(|\mathcal{Z}|)\epsilon$ and the function we end with can be simply rescaled (we replace negative masses by 0 and normalize the function dividing by a factor $1 - m$ where m is the total negative mass). With this transformation, we keep the expected advantage $O(\epsilon)$ and lose an extra factor $O(|\mathcal{Z}|)$ in the complexity. We can then. Finally, we need to remember that we construct only a probability distribution function, not a sampler. Transforming it into a sampler yields an overhead of $O(\mathcal{Z})$. This discussion shows that it is possible to build a sampler of complexity $\mathrm{poly}(|\mathcal{Z}|)\epsilon^{-2}$ with respect to $\mathcal{D}$. A more careful inspection of the proof shows that we can actually achieve the claimed bound $|\mathcal{Z}|^5\epsilon^{-2}$ (see Remark 4 at the end of the proof).

Technical Discussion We note that condition (b) somehow means that mass cuts should go in the right direction, as it is much simpler to prove that Algorithm 1 terminates when there are no correction terms θ^t; thus we don't want to go in a wrong direction and ruin the energy gain. Concrete bounds on properties (a) and (b) are given in Claims 3 and 4.

In Algorithm 1 in every round we shift only one negative point mass (see Line 12). However, since this point mass is chosen to be as big as possible and since h^{t+1} and h^t differ only by a small term $\gamma \cdot \overline{\mathrm{D}}^{t+1}$ except the mass shift θ^{t+1}, one can expect that we have the negative mass under control. Indeed, this is stated precisely in Claim 3 below.

Claim 3 (The total negative mass is small). Let

$$\mathsf{NegativeMass}(h^t(x,\cdot)) = -\sum_z \min(h^t(x,z),0)$$

be the total negative mass in $h^t(x,z)$ as the function of z. Then we have

$$\mathsf{NegativeMass}(h^t(x,\cdot) < |\mathcal{Z}|^3\gamma. \tag{9}$$

[12] It's not possible to extend the result from [TTV09] directly, the issue is that the constraint on the marginal distribution are not preserved. That's why [JP14] and this paper require much more extra work.

for every x and every t. In fact, for all x, z and t we have the following stronger bound

$$\max_z \left|\min(h^t(x,z),0)\right| < |\mathcal{Z}|\gamma.$$

The proof is based on a recurrence relation that links $\mathsf{NegativeMass}(h^{t+1}(x,\cdot)$ with $\mathsf{NegativeMass}(h^t(x,\cdot)$, and appears in Appendix C.

Claim 4 (The angle formed by the correction and the difference vector is acute). For every x, t we have $\mathsf{Angle}\left(\theta_x^{t+1}, g_x - h_x^{t+1}\right) \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$.

The proof appears in Appendix D.

Having established Claims 3 and 4 we are now in position to prove a concrete bound in Equation (8). To this end, we give upper bounds on E_1 and E_2, defined in Equation (7), separately.

Claim 5 (Algorithm 1 terminates after a small number of steps.). The energy function in Claim 2 can be bounded as follows

$$E_1 \leqslant \gamma^{-1}\left(1 + 2|\mathcal{Z}|^2\gamma + |\mathcal{Z}|t\gamma^2 + |\mathcal{Z}|^3 t\gamma^2\right), \quad E_2 \leqslant 2|\mathcal{Z}|^2 t\gamma.$$

In particular, we conclude that with $\gamma = \frac{\epsilon}{8|\mathcal{Z}|^4}$ the algorithm terminates after at most $t = O(|\mathcal{Z}|^3)\epsilon^{-2}$ steps.

First, note that by Claim 4 we have $-\sum_{i=0}^{t-1} \theta_x^{i+1} \cdot \left(g_x - h_x^{i+1}\right) \leqslant 0$. Second, by definition of the sequence $(h^i)_i$ we have $-\sum_{i=0}^{t-1} \theta_x^{i+1} \cdot \left(h_x^{i+1} - h_x^i\right) = -\sum_{i=0}^{t-1} \theta_x^{i+1} \cdot \theta_x^{i+1} - \sum_{i=0}^{t-1} \gamma\theta_x^{i+1} \cdot \overline{\mathrm{D}}_x^{i+1}$ which is at most $2|\mathcal{Z}|^3 t\gamma^2$, because of Equation (9) (the sum of absolute correction terms $\sum_z |\theta^{i+1}(x,z)|$ is, by definition, twice the total negative mass, and $|\overline{\mathrm{D}}^{i+1}(x,z)| \leqslant 1$). This proves that

$$E_2 \leqslant \frac{1}{\gamma} \cdot 2|\mathcal{Z}|^3 t\gamma^2 = 2|\mathcal{Z}|^3 t\gamma.$$

To bound E_1, note that we have to bounds two non-negative terms, namely $\frac{1}{2}\sum_i \left(h_x^{i+1} - h_x^i\right)^2$ and $\left(h_x^t - h_x^0\right) \cdot g_x$. As for the first one, we have

$$\left(h_x^{i+1} - h_x^i\right)^2 = \left(\gamma\overline{\mathrm{D}}_x^{i+1} + \theta_x^{i+1}\right)^2 \leqslant 2(\gamma\overline{\mathrm{D}}_x^{i+1})^2 + 2\left(\theta_x^{i+1}\right)^2,$$

where the inequality follows by the Cauchy-Schwarz inequality[13]. We trivially have $\left(\overline{\mathrm{D}}_x^{i+1}\right)^2 \leqslant |\mathcal{Z}|$ (because of $|\overline{\mathrm{D}}(x,z)| \leqslant 1$). By the definition of correction terms in Equation (4) we have $\left(\theta_x^{i+1}\right)^2 = \sum_z (\theta^{i+1}(x,z))^2 < 2(\theta^{i+1}(x,z_0))^2$, where $\theta^{i+1}(x,z_0)$ is the smallest negative mass , which is at most $(2|\mathcal{Z}|^3\gamma)^2$ by Equation (9) . Thus, we have $\left(h_x^{i+1} - h_x^i\right)^2 \leqslant 2|\mathcal{Z}|\gamma^2 + 8|\mathcal{Z}|^6\gamma^2$. To bund $\left(h_x^t - h_x^0\right) \cdot g_x$ note that $-h_x^0 \cdot g_x \leqslant 0$ and that $h_x^t \cdot g_x \leqslant \max_z |h^t(x,z)|$ (because

[13] Or cam be concluded from the parallelogram identity $(x+y)^2 + (x-y)^2 = x^2 + y^2$

$g(x,z) \geqslant 0$ and $\sum_x g(x,z) = 1$) which means $h^t_x \cdot g_x \leqslant 1 + 2\mathsf{NegativeMass}(h^t_x)$ (as $\sum_z \max(h^t(x,z),0) = 1 - \sum_z \min(h^t(x,z),0) = 1 + \mathsf{NegativeMass}(h^t_x)$ and $-\sum_z \min(h^t(x,z),0) = \mathsf{NegativeMass}(h^t_x)$ by $\sum_z \max(h^t(x,z) = 1$ and the definition of the total negative mass). This allows us to estimate E_1 as follows

$$E_1 \leqslant \gamma^{-1}\left(1 + 2|\mathcal{Z}|^3\gamma + |\mathcal{Z}|t\gamma^2 + 4|\mathcal{Z}|^6 t\gamma^2\right)$$

After t steps, the energy is at least $t\epsilon$. On the other hand, it at most $E_1 + E_2$. Since $|\mathcal{Z}|, |\mathcal{Z}|^3 \leqslant |\mathcal{Z}|^6$, we obtain

$$t\epsilon < \gamma^{-1} + 2|\mathcal{Z}|^3 + 7|\mathcal{Z}|^6 t\gamma$$

Since this is true for any positive γ, we choose $\gamma = \frac{\epsilon}{14|\mathcal{Z}|^6}$, which gives us (slightly weaker than claimed)

$$t < 32|\mathcal{Z}|^6\epsilon^{-2}.$$

Remark 4 (Optimized bounds). By the second part of Claim 3 we have $|\theta^i(x,z)| < |\mathcal{Z}|\gamma$ for every x, z and i. An inspection of the discussion above shows that this allows us to improve the bounds on E_1, E_2

$$E_1 \leqslant \gamma^{-1}\left(1 + 2|\mathcal{Z}|^2\gamma + |\mathcal{Z}|t\gamma^2 + |\mathcal{Z}|^2 t\gamma^2\right), \quad E_2 \leqslant 2|\mathcal{Z}|^2 t\gamma$$

Setting $\gamma = \frac{\epsilon}{8|\mathcal{Z}|^2}$ we get $E_1 + E_2 \leqslant 20|\mathcal{Z}|^2\epsilon^{-1}$ and $t \leqslant 20|\mathcal{Z}|^2\epsilon^{-2}$.

This finishes the proof of the claim.

From Claim 5 we conclude that after $t = O\left(|\mathcal{Z}|^2\epsilon^{-2}\right)$ steps we end up with a function $h = h^t$ that is (s,ϵ)-indistinguishable from g, because the algorithm terminated (and, clearly, has the complexity at most $O\left(|\mathcal{Z}|^3\epsilon^{-2}\right)$ relative to circuits of size s (including an overhead of $O(|\mathcal{Z}|)$ to compute $\overline{\mathrm{D}}$ from D). To finish the proof, we need to solve two issues

Claim 6 (From the signed measure to the probability measure). Let h^t be the output of the algorithm. Define the probability distribution

$$h(x,z) = \frac{\max(h^t(x,z),0)}{\sum_{z'}\max(h^t(x,z'),0)}$$

for every x, z. Then $h^t(x,\cdot)$ and $h(x,\cdot)$ are $O(\epsilon)$-statistically close for every x.

To prove the claim, we note that $\sum_{z'}\max(h^t(x,z'),0)$ equals $1+\beta$ where $\beta = \mathsf{NegativeMass}(h^t(x,\cdot)$. Thus we have $|h(x,z) - h^t(x,z)| \leqslant |h^t(x,z)| \cdot \frac{\beta}{1+\beta}$. Since $\sum_{z'}|h^t(x,z')| = \sum_{z'}\max(h^t(x,z'),0) - \sum_{z'}\min(h^t(x,z'),0) = 1 + 2\beta$, we get $\sum|h(x,z) - h^t(x,z)| = O(\beta)$ which is $O(\epsilon)$ by Claim 3 for γ defined as in Claim 5.

Recall that we have constructed an approximating probability measure h for the probability mass function g, which is not a sampler yet. However, we can fix it by rejection sampling, as shown below.

Claim 7 (From the pmf to the sampler). There exists a (probabilistic) function $h_{\mathsf{sim}} : \mathcal{X} \to \mathcal{Z}$ which calls $h(x,z)$ (defined as above) at most $O(|\mathcal{Z}|\log(1/\epsilon))$ times and for every x the distribution of its output is ϵ-close to $h(x,\cdot)$.

The proof goes by a simple rejection sampling argument: we sample a point $z \leftarrow \mathcal{Z}$ at random and reject with probability $h(x,z)$. The rejection probability in one turn is $\frac{1}{|\mathcal{Z}|}$. If we repeat the experiment $|\mathcal{Z}|\log(1/\epsilon)|$ then the probability of rejection in every round is only ϵ. On the other hand, conditioned on the opposite event, we get the distribution identical to $h(x,\cdot)$. So the distance is at most ϵ as claimed. note that

The last two claims prove that the distribution of $h_{sim}(x)$ is $(s, O(\epsilon))$-close to $h^t_x = h^t(x,\cdot)$, for every x. Since h^t, as a function of x, z is (s,ϵ)-close to g, and g is the conditional distribution of $Z|X$, we obtain

$$X, h_{sim}(X) \approx^{s,O(\epsilon)} X, Z$$

and the complexity of the final sampler $h_{sim}(X)$ is $O(|\mathcal{Z}|^5\epsilon^{-2})$

4 Time-success ratio under algebraic transformations

In Lemma 1 below we provide a quantitative analysis of how the time-success ratio changes under concrete formulas in security reductions.

Lemma 1 (Time-success ratio for algebraic transformations). *Let a, b, c and A, B, C be positive constants. Suppose that P' is secure against adversaries (s', ϵ'), whenever P is secure against adversaries (s, ϵ), where*

$$\begin{aligned} s' &= s \cdot c\epsilon^C - b\epsilon^{-B} \\ \epsilon' &= a\epsilon^A. \end{aligned} \tag{10}$$

In addition, suppose that the following condition is satisfied

$$A \leqslant C + 1. \tag{11}$$

Then the following is true: if P is 2^k-secure, then P' is $2^{k'}$-secure (in the sense of Definition 4) where

$$k' = \begin{cases} \frac{A}{B+C+1}k + \frac{A}{B+C+1}(\log c - \log b) - \log a, & b \geqslant 1 \\ \frac{A}{C+1}k + \frac{A}{C+1}\log c - \log a, & b = 0 \end{cases} \tag{12}$$

The proof is elementary though not immediate. It can be found in [Skó15].

Remark 5 (On the technical condition (11)). This condition is satisfied in almost all applications, at in the reduction proof typically ϵ' cannot be better (meaning higher exponent) than ϵ. Thus, quite often we have $A \leqslant 1$.

References

BL13. Ahto Buldas and Risto Laanoja, *Security proofs for hash tree time-stamping using hash functions with small output size*, Information Security and Privacy (Colin Boyd and Leonie Simpson, eds.), Lecture Notes in Computer Science, vol. 7959, Springer Berlin Heidelberg, 2013, pp. 235–250 (English).

CLP15. Kai-Min Chung, Edward Lui, and Rafael Pass, *From weak to strong zero-knowledge and applications*, Theory of Cryptography - 12th Theory of Cryptography Conference, TCC 2015, Warsaw, Poland, March 23-25, 2015, Proceedings, Part I, 2015, pp. 66–92.

DP08. Stefan Dziembowski and Krzysztof Pietrzak, *Leakage-resilient cryptography*, Proceedings of the 2008 49th Annual IEEE Symposium on Foundations of Computer Science (Washington, DC, USA), FOCS '08, IEEE Computer Society, 2008, pp. 293–302.

DP10. Yevgeniy Dodis and Krzysztof Pietrzak, *Leakage-resilient pseudorandom functions and side-channel attacks on feistel networks*, Advances in Cryptology - CRYPTO 2010, 30th Annual Cryptology Conference, Santa Barbara, CA, USA, August 15-19, 2010. Proceedings, 2010, pp. 21–40.

DTT09. Anindya De, Luca Trevisan, and Madhur Tulsiani, *Non-uniform attacks against one-way functions and prgs*, Electronic Colloquium on Computational Complexity (ECCC) **16** (2009), 113.

FK99. Alan M. Frieze and Ravi Kannan, *Quick approximation to matrices and applications.*, Combinatorica **19** (1999), no. 2, 175–220.

FPS12. Sebastian Faust, Krzysztof Pietrzak, and Joachim Schipper, *Practical leakage-resilient symmetric cryptography*, Cryptographic Hardware and Embedded Systems - CHES 2012 - 14th International Workshop, Leuven, Belgium, September 9-12, 2012. Proceedings, 2012, pp. 213–232.

FR12. Benjamin Fuller and Leonid Reyzin, *Computational entropy and information leakage*, Cryptology ePrint Archive, Report 2012/466, 2012, http://eprint.iacr.org/.

GW11. Craig Gentry and Daniel Wichs, *Separating succinct non-interactive arguments from all falsifiable assumptions.*, STOC (Lance Fortnow and Salil P. Vadhan, eds.), ACM, 2011, pp. 99–108.

Imp95. Russell Impagliazzo, *Hard-core distributions for somewhat hard problems*, In 36th Annual Symposium on Foundations of Computer Science, IEEE, 1995, pp. 538–545.

JP14. Dimitar Jetchev and Krzysztof Pietrzak, *How to fake auxiliary input*, Theory of Cryptography TCC 2014 (Yehuda Lindell, ed.), Lecture Notes in Computer Science, vol. 8349, Springer, 2014, pp. 566–590.

LM94. Michael George Luby and Luby Michael, *Pseudorandomness and cryptographic applications*, Princeton University Press, Princeton, NJ, USA, 1994.

Pie09. Krzysztof Pietrzak, *A leakage-resilient mode of operation*, Advances in Cryptology - EUROCRYPT 2009 (Antoine Joux, ed.), Lecture Notes in Computer Science, vol. 5479, Springer Berlin Heidelberg, 2009, pp. 462–482 (English).

Pie15. ———, *private communication, may*, 2015.

RTTV08. Omer Reingold, Luca Trevisan, Madhur Tulsiani, and Salil Vadhan, *Dense subsets of pseudorandom sets*, Proceedings of the 2008 49th Annual IEEE Symposium on Foundations of Computer Science (Washington, DC, USA), FOCS '08, IEEE Computer Society, 2008, pp. 76–85.

Skó15. Maciej Skórski, *Time-advantage ratios under simple transformations: Applications in cryptography*, Cryptography and Information Security in the Balkans - Second International Conference, BalkanCryptSec 2015, Koper, Slovenia, September 3-4, 2015, Revised Selected Papers, 2015, pp. 79–91.

TTV09. Luca Trevisan, Madhur Tulsiani, and Salil Vadhan, *Regularity, boosting, and efficiently simulating every high-entropy distribution*, Proceedings of the 2009 24th Annual IEEE Conference on Computational Complexity (Washington, DC, USA), CCC '09, IEEE Computer Society, 2009, pp. 126–136.

VZ13. Salil Vadhan and ColinJia Zheng, *A uniform min-max theorem with applications in cryptography*, Advances in Cryptology CRYPTO 2013 (Ran Canetti and JuanA. Garay, eds.), Lecture Notes in Computer Science, vol. 8042, Springer Berlin Heidelberg, 2013, pp. 93–110 (English).

YS13. Yu Yu and François-Xavier Standaert, *Practical leakage-resilient pseudorandom objects with minimum public randomness*, Proceedings of the 13th International Conference on Topics in Cryptology (Berlin, Heidelberg), CT-RSA'13, Springer-Verlag, 2013, pp. 223–238.

A More on the flaw in [JP14]

In the original setting we have $\mathcal{Z} = \{0,1\}^\lambda$. In the proof of the claimed better bound $O\left(s \cdot 2^{3\lambda}\epsilon^{-2}\right)$ there is a mistake on page 18 (eprint version), when the authors enforce a signed measure to be a probability measure by a mass shifting argument. The number M defined there is in fact a function of x and is hard to compute, whereas the original proof amuses that this is a constant independent of x. During iterations of the boosting loop, this number is used to modify distinguishers class step by step, which drastically blows up the complexity (exponentially in the number of steps, which is already polynomial in ϵ). In the min-max based proof giving the bound $O\left(s \cdot 2^{3\lambda}\epsilon^{-4}\right)$ a fixable flaw is a missing factor of 2^λ in the complexity (page 16 in the eprint version), which is because what is constructed in the proof is only a probability mass function, not yet a sampler [Pie15].

B Proof of Claim 2

We can rewrite Equation (6) as

$$\begin{aligned}\Delta^t &= \frac{1}{\gamma} \mathop{\mathbb{E}}_{x\sim X}\left[\sum_{i=0}^{t-1}\left(\left(h_x^{i+1}-h_x^i\right)-\theta_x^{i+1}\right)\cdot\left(g_x-h_x^i\right)\right] \\ &= \frac{1}{\gamma} \mathop{\mathbb{E}}_{x\sim X}\left[\sum_{i=0}^{t-1}\left(h_x^{i+1}-h_x^i\right)\cdot\left(g_x-h_x^i\right)-\sum_{i=0}^{t-1}\theta_x^{i+1}\cdot\left(g_x-h_x^i\right)\right] \qquad (13)\end{aligned}$$

First, note that

$$\begin{aligned}\sum_{i=0}^{t-1}\left(h_x^{i+1}-h_x^i\right)\cdot\left(g_x-h_x^i\right)=&\\ &=\left(h_x^t-h_x^0\right)\cdot g_x-\sum_{i=0}^{t-1}h_x^i\cdot\left(h_x^{i+1}-h_x^i\right)\\ &=\left(h_x^t-h_x^0\right)\cdot g_x+\frac{1}{2}\sum_{i=0}^{t-1}\left(h_x^{i+1}-h_x^i\right)\cdot\left(h_x^{i+1}-h_x^i\right)+\\ &\quad-\frac{1}{2}\sum_{i=0}^{t-1}\left(h_x^{i+1}+h_x^i\right)\cdot\left(h_x^{i+1}-h_x^i\right)\\ &=\left(h_x^t-h_x^0\right)\cdot g_x+\frac{1}{2}\sum_{i=0}^{t-1}\left(h_x^{i+1}-h_x^i\right)^2-\frac{1}{2}\left(\left(h_x^t\right)^2-\left(h_x^0\right)^2\right)\end{aligned}\tag{14}$$

As to the second term in Equation (13), we observe that

$$-\sum_{i=0}^{t-1}\theta_x^{i+1}\cdot\left(g_x-h_x^i\right)=-\sum_{i=0}^{t-1}\theta_x^{i+1}\cdot\left(g_x-h_x^{i+1}\right)-\sum_{i=0}^{t-1}\theta_x^{i+1}\cdot\left(h_x^{i+1}-h_x^i\right)\tag{15}$$

C Proof of Claim 3

Proof (Proof of Claim 3). We start by comparing the total negative mass in the functions $h^{t+1}=h^t+\overline{\mathrm{D}}^{t+1}+\theta^{t+1}$ and h^t. Suppose first that $\tilde{h}^t(x,z_0)<0$ where $z_0=z_{\min}^t(x)$. Since $\sum_{z\neq z_0}\tilde{h}^{t+1}=1-\tilde{h}^{t+1}(x,z_0)$, there exists z_1 such that $\tilde{h}^{t+1}(x,z_1)\geqslant\frac{1-\tilde{h}^{t+1}(x,z_0)}{|\mathcal{Z}|-1}>0$. Combining this with Equation (4) we obtain

$$h^{t+1}(x,z_1)=\tilde{h}^{t+1}(x,z_1)+\frac{\tilde{h}^{t+1}(x,z_0)}{|\mathcal{Z}|-1}\geqslant\frac{1}{|\mathcal{Z}|-1}\tag{16}$$

These observations together with Equation (3) give us

$$\begin{aligned}
\sum_{z\in\mathcal{Z}} \min\left(h^{t+1}(x,z),0\right) &= \sum_{z\in\mathcal{Z}} \min\left(\tilde{h}^{t+1}(x,z)+\theta^{t+1}(x,z),0\right)\\
&= \sum_{z\in\mathcal{Z}\setminus\{z_0,z_1\}} \min\left(\tilde{h}^{t+1}(x,z)+\frac{\tilde{h}^{t+1}(x,z_0)}{|\mathcal{Z}|-1},0\right)\\
&\geqslant \sum_{z\in\mathcal{Z}\setminus\{z_0,z_1\}} \min\left(\tilde{h}^{t+1}(x,z),0\right)+(|\mathcal{Z}|-2)\cdot\frac{\tilde{h}^{t+1}(x,z_0)}{|\mathcal{Z}|-1}\\
&= \sum_{z\in\mathcal{Z}} \min\left(\tilde{h}^{t+1}(x,z),0\right)+(|\mathcal{Z}|-2)\cdot\frac{\tilde{h}^{t+1}(x,z_0)}{|\mathcal{Z}|-1}-\tilde{h}^{t+1}(x,z_1)\\
&= \sum_{z\in\mathcal{Z}} \min\left(\tilde{h}^{t+1}(x,z),0\right)+\min\left(\frac{\tilde{h}^{t+1}(x,z_0)}{|\mathcal{Z}|-1},0\right)
\end{aligned} \tag{17}$$

where the inequality line follows from $\tilde{h}^{t+1}(x,z_0)<0$ and Equation (16). But by the definition of $z_0=z^t_{\min}(x)$ we have $\tilde{h}^{t+1}(x,z_0)=\min_z \tilde{h}^{t+1}(x,z)$. Since this value is negative, we get

$$\tilde{h}^{t+1}(x,z_0)\leqslant\frac{1}{|\mathcal{Z}|-1}\cdot\sum_{z\in\mathcal{Z}}\min\left(\tilde{h}^{t+1}(x,z),0\right) \tag{18}$$

Combining Equation (17) and Equation (18) we obtain

$$-\sum_{z\in\mathcal{Z}}\min\left(h^{t+1}(x,z),0\right)\leqslant-\left(1-\frac{1}{(|\mathcal{Z}|-1)^2}\right)\sum_{z\in\mathcal{Z}}\min\left(\tilde{h}^{t+1}(x,z),0\right). \tag{19}$$

Since $|h^{t+1}(x,z)-\tilde{h}^t(x,z)|\leqslant\gamma$ by Equation (3), we get the following recursion

$$-\sum_{z\in\mathcal{Z}}\min\left(h^{t+1}(x,z),0\right)\leqslant-\left(1-\frac{1}{(|\mathcal{Z}|-1)^2}\right)\sum_{z\in\mathcal{Z}}\min\left(h^{t}(x,z),0\right)+|\mathcal{Z}|\gamma \tag{20}$$

which can be rewritten as

$$\mathsf{NegativeMass}\left(h^{t+1}(x,\cdot)\right)<\left(1-\frac{1}{|\mathcal{Z}|^2}\right)\mathsf{NegativeMass}\left(h^{t}(x,\cdot)\right)+|\mathcal{Z}|\gamma. \tag{21}$$

which is in addition trivially true if $\tilde{h}^{t+1}(x,z)\geqslant 0$ for all z. Since we have $\mathsf{NegativeMass}\left(h^0(x,\cdot)\right)=0$, expanding this recursion till $t=0$ gives an upper bound $|\mathcal{Z}|\gamma\cdot\sum_{j\leqslant t+1}\left(1-|\mathcal{Z}|^{-2}\right)^j$ which is smaller than by $|\mathcal{Z}|^3\gamma$ by the convergence of the geometric series. This finishes the proof of the first part.

To prove the second part, recall that by the definition of z_0 we have $\tilde{h}^{t+1}(x, z_0) = \min_z \tilde{h}^{t+1}(x, z)$. Suppose that $\tilde{h}^{t+1}(x, z_0) < 0$ (that is, there is a negative mass in $\widetilde{h}^{t+1}(x, \cdot)$). Now, by the definition of h^{t+1}, we get

$$\begin{aligned}\max_z \left|\min(h^{t+1}(x, z), 0)\right| &= \max_{z \neq z_0} \left|\min(h^{t+1}(x, z), 0)\right| \\ &= \max_{z \neq z_0} \left|\min\left(\widetilde{h}^{t+1}(x, z) + \frac{|\widetilde{h}^{t+1}(x, z_0)|}{|\mathcal{Z}| - 1}, 0\right)\right|.\end{aligned}$$

Suppose that $\widetilde{h}^{t+1}(x, z) + \frac{|\widetilde{h}^{t+1}(x, z_0)|}{|\mathcal{Z}|-1} \leqslant 0$ for some z. Then, by the definition of z_0, we also have

$$\begin{aligned}0 &\geqslant \widetilde{h}^{t+1}(x, z) + \frac{|\widetilde{h}^{t+1}(x, z_0)|}{|\mathcal{Z}| - 1} \\ &\geqslant \widetilde{h}^{t+1}(x, z_0) + \frac{|\widetilde{h}^{t+1}(x, z_0)|}{|\mathcal{Z}| - 1} \\ &= -\left(1 - \frac{1}{|\mathcal{Z}| - 1}\right)\left|\widetilde{h}^{t+1}(x, z_0)\right|.\end{aligned}$$

From this we conclude that for *any* z we have

$$\min\left(\widetilde{h}^{t+1}(x, z) + \frac{|\widetilde{h}^{t+1}(x, z_0)|}{|\mathcal{Z}| - 1}, 0\right) \geqslant -\left(1 - \frac{1}{|\mathcal{Z}| - 1}\right)\left|\widetilde{h}^{t+1}(x, z_0)\right|.$$

and thus

$$\max_{z \neq z_0}\left|\min\left(\widetilde{h}^{t+1}(x, z) + \frac{|\widetilde{h}^{t+1}(x, z_0)|}{|\mathcal{Z}| - 1}, 0\right)\right| \leqslant \left(1 - \frac{1}{|\mathcal{Z}| - 1}\right)\left|\widetilde{h}^{t+1}(x, z_0)\right|$$

which means that (still assuming that $\widetilde{h}^{t+1}(x, z_0) < 0$)

$$\max_z \left|\min(h^{t+1}(x, z), 0)\right| \leqslant \left(1 - \frac{1}{|\mathcal{Z}| - 1}\right) \max_z \left|\min\left(\widetilde{h}^{t+1}(x, z), 0\right)\right|.$$

Note that $0 \geqslant \min\left(\widetilde{h}^{t+1}(x, z), 0\right) \geqslant \min\left(h^t(x, z), 0\right) - \gamma$ by the definition of h^{t+1} and $\widetilde{h}^{t+1}$. Then

$$\max_z \left|\min(h^{t+1}(x, z), 0)\right| \leqslant \left(1 - \frac{1}{|\mathcal{Z}| - 1}\right) \max_z \left|\min(h^t(x, z), 0)\right| + \gamma.$$

Note that this inequality is true even if $\widetilde{h}^{t+1}(x, z_0) = 0$, that is $\widetilde{h}^{t+1}(x, z) \geqslant 0$ for all z as then $h^{t+1}(x, z) \geqslant 0$ for all z. By expanding this recursion, and noticing that $\min(h^0(x, z), 0) = 0$ for all x, z by definition, we get

$$\max_z \left|\min(h^{t+1}(x, z), 0)\right| \leqslant \gamma \sum_{j=0}^{t} \left(1 - \frac{1}{|\mathcal{Z}| - 1}\right)^j < |\mathcal{Z}|\gamma.$$

D Proof of Claim 4

Proof. If $\theta^{t+1}(x,z) = 0$ then there is nothing to prove. Suppose that $\theta^{t+1}(x,z) < 0$. Let $z_0 = z^t_{\min}(x)$. According to Equation (4) we have $\theta^{t+1}(x,z_0) = -\tilde{h}^{t+1}(x,z_0)$ and $\theta^{t+1}(x,z) = \frac{\tilde{h}^{t+1}(x,z_0)}{\#\mathcal{Z}-1}$ for $z \neq z_0$. Therefore

$$\begin{aligned}
\theta^{t+1}_x \cdot \left(g_x - \tilde{h}^{t+1}_x\right) &= -\tilde{h}^{t+1}(x,z_0)\left(g(x,z_0) - \tilde{h}^{t+1}(x,z_0)\right) + \\
&\quad + \sum_{z \neq z_0} \frac{\tilde{h}^{t+1}(x,z_0)}{|\mathcal{Z}|-1} \cdot \left(g(x,z) - \tilde{h}^{t+1}(x,z)\right) \\
&= -\tilde{h}^{t+1}(x,z_0)\left(g(x,z_0) - \tilde{h}^{t+1}(x,z_0)\right) \\
&\quad - \frac{\tilde{h}^{t+1}(x,z_0)}{|\mathcal{Z}|-1}\left(g(x,z_0) - \tilde{h}^{t+1}(x,z_0)\right) \qquad (22)
\end{aligned}$$

and

$$-\theta^{t+1}_x \cdot \theta^{t+1}_x = -\tilde{h}^{t+1}(x,z_0) \cdot \tilde{h}^{t+1}(x,z_0)\left(1 + \frac{1}{|\mathcal{Z}-1|}\right). \qquad (23)$$

Putting Equations (22) and (23) together we obtain

$$\begin{aligned}
\theta^{t+1}_x \cdot \left(g_x - h^{t+1}_x\right) &= \theta^{t+1}_x \cdot \left(g_x - \tilde{h}^{t+1}_x\right) - \theta^{t+1}_x \cdot \theta^{t+1}_x \\
&= -\left(1 + \frac{1}{|\mathcal{Z}|-1}\right)\tilde{h}^{t+1}(x,z_0) \cdot g(x,z_0)
\end{aligned}$$

which is positive because $\tilde{h}^{t,r}(x,z_0) < 0$ and $g(x,z_0) \geqslant 0$. This proves Claim 4.

Chapter 5

Best Generic Attacks Against Pseudoentropy

Non-Uniform Attacks Against Pseudoentropy*

Krzysztof Pietrzak**, Maciej Skorski***

IST Austria

Abstract De, Trevisan and Tulsiani [CRYPTO 2010] show that every distribution over n-bit strings which has constant statistical distance to uniform (e.g., the output of a pseudorandom generator mapping $n-1$ to n bit strings), can be distinguished from the uniform distribution with advantage ϵ by a circuit of size $O(2^n\epsilon^2)$.
We generalize this result, showing that a distribution which has less than k bits of min-entropy, can be distinguished from any distribution with k bits of δ-smooth min-entropy with advantage ϵ by a circuit of size $O(2^k\epsilon^2/\delta^2)$. As a special case, this implies that any distribution with support at most 2^k (e.g., the output of a pseudoentropy generator mapping k to n bit strings) can be distinguished from any given distribution with min-entropy $k+1$ with advantage ϵ by a circuit of size $O(2^k\epsilon^2)$.
Our result thus shows that pseudoentropy distributions face basically the same non-uniform attacks as pseudorandom distributions.

Keywords: pseudoentropy, non-uniform attacks

1 Introduction

De, Trevisan and Tulsiani [DTT10] show a non-uniform attack against any pseudorandom generator (PRG) which maps $\{0,1\}^{n-1} \to \{0,1\}^n$. For any $\epsilon \geq 2^{-n/2}$, their attack achieves distinguishing advantage ϵ and can be realized by a circuit of size $O\left(2^n\epsilon^2\right)$. Their attack doesn't even need the PRG to be efficiently computable.

In this work we consider a more general question, where we ask for attacks distinguishing a distribution from any distribution with slightly higher min-entropy. We generalize [DTT10], showing a non-uniform attack which, for any $\epsilon, \delta > 0$, distinguishes any distribution with $< k$ bits of min-entropy from any distribution with k bits of δ-smooth min-entropy with advantage ϵ, and where the distinguisher is of size $O(2^k\epsilon^2/\delta^2)$. As a corollary we recover the [DTT10] result, showing that the output of any pseudoentropy generator $\{0,1\}^k \to \{0,1\}^n$ can be distinguished from any variable with min-entropy $k+1$ with advantage ϵ by circuits of size $O(2^k\epsilon^2)$.

* The full version is available at https://arxiv.org/abs/1704.08678

** Supported by the European Research Council, ERC consolidator grant (682815 - TOCNeT).

*** Supported by the European Research Council, ERC consolidator grant (682815 - TOCNeT).

– From a theoretical perspective, we prove where the separation between pseudoentropy and smooth min-entropy lies, by classifying how powerful computationally bounded adversaries can be so they can still be fooled to "see" more entropy than there really is.
– From a more practical perspective, our result shows that using pseudoentropy instead of pseudorandomness (which for many applications is sufficient and allows for saving in entropy *quantity* [DPW14]), will not give improvements in terms of *quality* (i.e., the size and advantage of distinguishers considered), at least not against generic non-uniform attacks.

1.1 Notation and Basic Definitions

Two variables X and Y are (s,ϵ) indistinguishable, denoted $X \sim_{s,\epsilon} Y$, if for all boolean circuits D of size $|D| \leq s$ we have $|\Pr[D(X)=1] - \Pr[D(Y)=1]| \leq \epsilon$. The statistical distance of X and Y is $d_1(X;Y) \stackrel{\mathsf{def}}{=} \sum_x |P_X(x) - P_Y(x)|$ (where $P_X(x) \stackrel{\mathsf{def}}{=} \Pr[X=x]$), the Euclidean distance of X and Y is $d_2(P_X;P_Y) \stackrel{\mathsf{def}}{=} \sqrt{\sum_x (P_X(x)-P_Y(x))^2}$. A variable X has min-entropy k if it doesn't take any particular outcome with probability greater 2^{-k}, it has δ-smooth min-entropy k [RW05], if it's δ close to some distribution with min-entropy k. X has k bits of HILL pseudoentoentry of quality (s,ϵ) if there exists a Y with min-entropy k that is (s,ϵ) indistinguishable from X, we use the following standard notation for these notions

min-entropy: $\mathsf{H}_\infty(X) \stackrel{\mathsf{def}}{=} -\log \max_x (\Pr[X=x])$.
smooth min-entropy: $\mathsf{H}^\delta_\infty(X) \stackrel{\mathsf{def}}{=} \max_{Y, d_1(X;Y)\leq\delta} \mathsf{H}_\infty(Y)$.
HILL pseudoentropy: $\mathsf{H}^{\mathsf{HILL}}_{s,\epsilon}(X) \stackrel{\mathsf{def}}{=} \max_{Y, Y\sim_{(s,\epsilon)}X} \mathsf{H}_\infty(Y)$.

1.2 Our Contribution

In this work give generic non-uniform attacks on pseudoentropy distributions. A seemingly natural goal is to consider a distribution X with $\mathsf{H}_\infty(X) \leq k$ bits of min-entropy, strictly larger $\mathsf{H}^{\mathsf{HILL}}_{s,\epsilon}(X) \geq k+1$ bits of HILL entropy, and then give an upper bound on s in terms of ϵ. This does not work as there are X where $\mathsf{H}_\infty(X) \ll \mathsf{H}^\delta_\infty(X)$,[1] and as by definition $\mathsf{H}^\delta_\infty(X) = \mathsf{H}^{\mathsf{HILL}}_{\infty,\delta}(X)$, we can have a large entropy gap $\mathsf{H}^{\mathsf{HILL}}_{\infty,\delta}(X) - \mathsf{H}_\infty(X)$ even when considering unbounded adversaries against HILL entropy. For this reason, in our main technical result Lemma 1 below, we must consider distributions with bounded *smooth* min-entropy. This makes the statement of the lemma somewhat technical. In practice, the distributions considered often have bounded support, for example because they were generated from a short seed by a deterministic process (like a pseudorandom generator). In this case we can drop the smoothness requirement as stated in Theorem 1 below.

[1] Consider an X which is basically uniform over $\{0,1\}^n$, but has mass δ on one particular point, then $\log \delta^{-1} = \mathsf{H}_\infty(X) \ll \mathsf{H}^\delta_\infty(X) = n$.

Lemma 1 (Nonuniform attacks against pseudoentropy). *Suppose that $X \in \{0,1\}^n$ does not have k bits of δ-smooth min-entropy, i.e., $\mathsf{H}^{\delta}_{\infty}(X) < k$, then for any ϵ we have*

$$\mathsf{H}^{\mathsf{HILL}}_{\tilde{O}(2^k\epsilon^2\delta^{-2}),\epsilon}(X) < k$$

where $\tilde{O}(\cdot)$ hides a factor linear in n.

Theorem 1. *Let $f : \{0,1\}^k \to \{0,1\}^n$ be a deterministic (not necessarily efficient) function. Then we have*

$$\mathsf{H}^{\mathsf{HILL}}_{\tilde{O}(2^k\epsilon^2),\epsilon}(f(U_k)) \leq k+1.$$

more generally, for any X over $\{0,1\}^n$ with support of size $\leq 2^k$

$$\mathsf{H}^{\mathsf{HILL}}_{\tilde{O}(2^k\epsilon^2),\epsilon}(X) \leq k+1.$$

Remark 1 (Concluding best attacks against PRGs). For the special case $n = k+1$ we recover the bound for pseudo*random* generators from [DTT10].

Proof (Proof of Theorem 1). The theorem follows from Lemma 1 when $\delta = 1/2$; consider any X with support of size $\leq 2^k$, then $\mathsf{H}^{\delta}_{\infty}(X) \leq k+1$, as no matter how we cut probability mass of $1-\delta = 1/2$ over 2^k elements, one element will have the weight at least 2^{-k-1}.

1.3 Proof Outline

A Weaker Result as a Ball-Bins Problem We outline the proof of a somewhat weakened version of Theorem 1 in the language of balls and bins. For every Y of min-entropy $k' = k + \Omega(1)$ we want to distinguish Y from $X = f(U_k)$. Suppose for simplicity that Y is flat and f is injective, so that X is also flat. Our strategy will be to hash the points randomly into two bins and take advantage of the fact that the *average maximum load* is closer to $\frac{1}{2}$ when we sample from Y than when drawing from X. The reason is that Y has more balls, so by the law of large numbers, we expect the load to be "more concentrated" around the mean.

Think of throwing balls (inputs x) into two bins (labeled by -1 and 1). If the balls come from the support of X, the expected maximum load (over two bins) equals $\approx 2^{k-1} + \sqrt{2/\pi} \cdot 2^{k/2}$. Similarly, if the balls come from the support of Y, then maximum load is $2^{k'-1} + \sqrt{2/\pi} \cdot 2^{k'/2}$. In terms of the average load (the load normalized by the total number of balls)

$$\begin{aligned} &\mathsf{AverageMaxLoad}(X) \approx 0.5 + \sqrt{2/\pi} \cdot 2^{-k/2} && \text{w.h.p. when drawing from } X \\ &\mathsf{AverageMaxLoad}(Y) \approx 0.5 + \sqrt{2/\pi} \cdot 2^{-k'/2} && \text{w.h.p. when drawing from } Y \end{aligned}$$

As $k' = k + \Omega(1)$ we obtain (with good probability)

$$\mathsf{AverageMaxLoad}(X) - \mathsf{AverageMaxLoad}(Y) = \Omega(2^{-k/2}).$$

Letting D be one of these bins assignments we obtain a distinguisher with advantage $\epsilon = \Omega(2^{-k/2})$. To generate the assignments efficiently we relax the assumption about choosing bins and assume only that the choices of bins are independent for any group of $\ell = 4$ balls. The fourth moment method allows us to keep sufficiently good probabilistic guarantees on the maximum load.

The General Case by Random Walk Techniques

A high-level outline and comparison to [DTT10] Below in Figure 1 we sketch the flow of our argument.

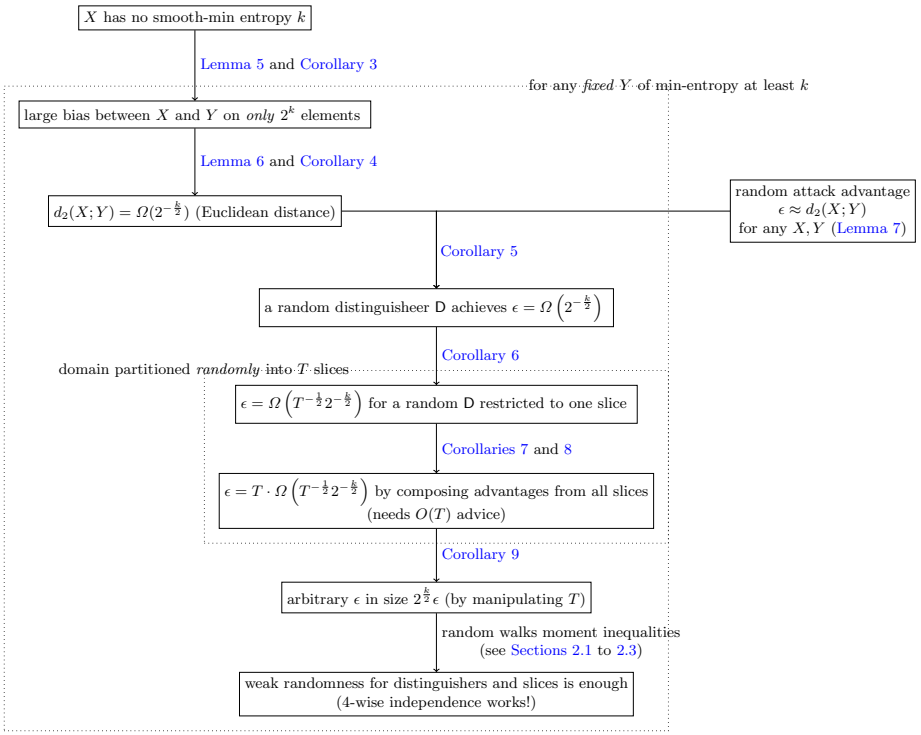


Figure 1: The map of our proof.

Our starting point is the proof from [DTT10]. They use the fact that a random mapping $\mathsf{D} : \{0,1\}^n \to \{-1,1\}$ likely distinguishes any two distributions X and Y over $\{0,1\}^n$ with advantage being the Euclidean distance $d_2(X;Y) \stackrel{\mathsf{def}}{=} \sqrt{\sum_x (P_X(x) - P_Y(x))^2}$.

For any X and Y with constant statistical distance $\sum_x |P_X(x) - P_Y(x)| = \Theta(1)$ (which is the case for the PRG setting where $Y = U_n$ and $X = PRG(U_{n-1})$) this yields a bound $\Omega\left(2^{-\frac{n}{2}}\right)$. This bound can be then amplified, at the cost of

extra advice, by partitioning the domain $\{0,1\}^n$ and combining corresponding advantages (advice basically encodes if there is a need for flipping the output). Finally one can show that 4-wise independence provides enough randomness for this argument, which makes sampling D efficient. Our argument deviates from this approach in two important aspects.

The first difference is that in the pseudoentropy case we can improve the advantage from $\Omega\left(2^{-\frac{n}{2}}\right)$, where n is the logarithm of the support of the variables considered, to $\Omega\left(2^{-\frac{k}{2}}\right)$, where k is the min-entropy of the variable we want to distinguish from. The reason is that being statistically far from any k-bit min-entropy distributions implies a *large bias on already* 2^k *elements*. This fact (see Lemma 5 and Corollary 3, and also Figure 3) is a new characterization of smooth min-entropy of independent interest.

The second subtlety arises when it comes to amplify the advantage over the partition slices. For the pseudorandomness case it is enough to split the domain in a deterministic way, for example by fixing prefixes of n-bit strings, in our case this is not sufficient. For us a "good" partition must shatter the 2^k-element high-biased set, which can be arbitrary. Our solution is to use *random partitions*, in fact, we show that using 4-universal hashing is sufficient. Generating base distinguishers and partitions at the same time makes probability calculations more involved.

Technical calculations are based on the fourth moment method, similarly as in [DTT10]. The basic idea is that for settings where the second and fourth moment are easy to compute (e.g. sums of independent symmetric random variables) we can obtain good upper and lower bounds on the first moment. In the context of algorithmic applications these techniques are usually credited to [Ber97]. Interestingly, exploiting natural relations to *random walks*, we show that calculations immediately follow by adopting classical (almost one century old) tools and results [MZ38,Khi24]. Our technical novelty is an application of moment inequalities due to Marcinkiewicz-Zygmund and Paley-Zygmund, which allow us to prove slightly more than just the existence of an attack. Namely we generate it with constant success probability.

Advantage $\Omega(2^{-k/2})$ Consider any X with δ-smooth min-entropy smaller than k. This requirement can be seen as a statement about the "shape" of the distribution. Namely, the mass of X that is above the threshold 2^{-k} equals at least δ, that is

$$\sum_x \max(P_X(x) - 2^{-k}, 0) \geqslant \delta.$$

For an illustration see Figure 2.
We construct our attack based on this observation. Define the advantage of a function D for distributions X and Y as

$$\mathsf{Adv}^{\mathsf{D}}(X;Y) = \left|\sum_x \mathsf{D}(x)(P_X(x) - P_Y(x))\right|$$

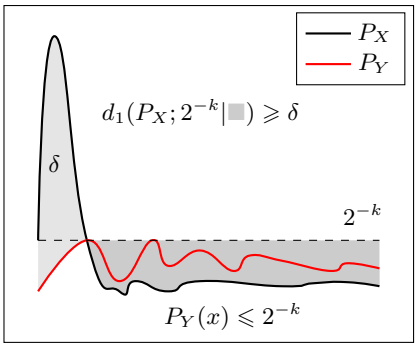


Figure 2: An intuition behind the attack. Random ± 1-weights make the bias equal to the ℓ_2-distance of P_X and P_Y. This distance can be bounded in terms of the ℓ_1 distance, which concentrates mass difference δ on less than 2^k elements (the region in gray).

(writing also $\mathsf{Adv}^{\mathsf{D}}_S$ when the summation is restricted to a subset S). Consider a random distinguisher $\mathsf{D} : \{0,1\}^n \to \{-1,1\}$. Random variables $\mathsf{D}(x)$ for different x are independent, have zero-mean and second moment equal to 1. Therefore the expected square of of the advantage, over the choice of D, equals

$$\mathbb{E}\left[\left(\mathsf{Adv}^{\mathsf{D}}(X;Y)\right)^2\right] = \mathbb{E}\left|\sum_x \mathsf{D}(x)(P_X(x) - P_Y(x))\right|^2 = \sum_x (P_X(x) - P_Y(x))^2$$

Let S be the set of x such that $P_X(x) > 2^{-k}$. For any Y of min-entropy at least k we obtain

$$\begin{aligned}\sum_{x\in S}(P_X(x) - P_Y(x))^2 &\geqslant \sum_{x\in S}(P_X(x) - 2^{-k})^2 \\ &\geqslant |S|^{-1}\left(\sum_{x\in S}\left(P_X(x) - 2^{-k}\right)\right)^2 \geqslant 2^{-k}\delta^2\end{aligned}$$

where the first inequality follows because $P_Y(x) \leqslant 2^{-k} < P_X(x)$ for $x \in S$, the second inequality is by the standard inequality between the first and second norm, and the third inequality follows because we showed that $\Pr[X \in S] \geqslant |S| \cdot 2^{-k} + \delta$ (illustrated in Figure 2) which also implies $|S|^{-1} \geqslant 2^{-k}$.

By the previous formula on the expected squared advantage this means that

$$\mathbb{E}\left[\left(\mathsf{Adv}^{\mathsf{D}}(X;Y)\right)^2\right] \geqslant 2^{-k}\delta^2$$

for at least one choice of D. This implies

$$\mathsf{Adv}^{\mathsf{D}}(X;Y) \geqslant 2^{-\frac{k}{2}}\delta.$$

A random D as defined would be of size exponential in n, but since we used only the second moment in calculations, it suffices to generate $\mathsf{D}(x)$ as pairwise independent random variables. By assuming 4-wise independence – which can be computed by $O(n^2)$ size circuits – we can prove slightly more, namely that a constant fraction of generated D's are good distinguishers. This property will be important for the next step, where we amplify the advantage assuming larger distinguishers.

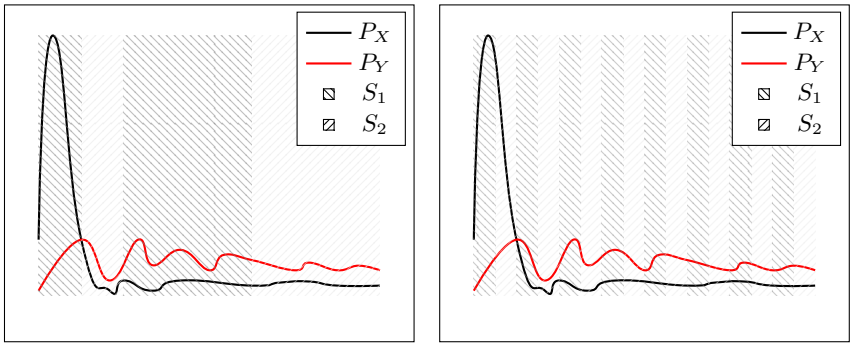
Amplifying the advantage by slicing the domain Consider a random and equitable partition $\{S_i\}_{i=1}^{T}$ of the set $\{0,1\}^n$. From the previous analysis we know that a random distinguisher achieves advantage $\epsilon = d_2(P_X; P_Y)$ over the whole domain. Note that (for any, not necessarily random partition $\{S_i\}_i$) we have

$$(d_2(P_X; P_Y))^2 = \sum_{i=1}^{T} (d_2(P_X; P_Y|S_i))^2$$

where $d_2(P_X; P_Y|S_i)$ is the restriction of the distance to the set S_i (by restricting the summation to S_i). From a random partition we expect the mass difference between P_X and P_Y to be *distributed evenly* among the partition slices (see Figure 3(b)). Based on the last equation, we expect

$$d_2(P_X; P_Y|S_i) \approx \frac{d_2(P_X; P_Y)}{\sqrt{T}}$$

to hold with high probability over $\{S_i\}_i$. In fact, if the mass difference is not well



(a) An example of a "bad" partition. Almost all advantage is captured by one partition slice S_1.

(b) An example of a "good" partition. The advantage is evenly distributed among slices S_1, S_2.

Figure 3: Illustration of good and bad partitions.

balanced amongst the slices (in the extreme case, concentrated on one slice) our argument will not offer any gain over the previous construction (see Figure 3(a)).

By applying the previous argument to individual slices, for every i we can obtain an advantage $\mathsf{Adv}^{\mathsf{D}}_{S_i}(X;Y) = \Omega\left((T^{-\frac{1}{2}}2^{-\frac{k}{2}})\delta\right)$ when restricted to the set S_i (with high probability over the choice of D and $\{S_i\}_i$). Now if the sets S_i are *efficiently recognizable*, we can combine them into a better distinguisher. Namely for every i we chose a value $\beta_i \in \{-1, 1\}$ such that D's advantage (before taking the absolute value) restricted to S_i has sign β_i, and set

$$\hat{\mathsf{D}}(x) = \beta_i \mathsf{D}(x), \text{ where } i \text{ is such that } x \in S_i,$$

then the advantage equals (with high probability over D and the S_i's)

$$\mathsf{Adv}^{\hat{\mathsf{D}}}(X;Y) = \sum_{i=1}^{T} \mathsf{Adv}^{\mathsf{D}}_{S_i}(X;Y) = \Omega\left(T^{\frac{1}{2}}2^{-\frac{k}{2}}\delta\right)$$

We need to specify a 4-wise independent hash for D, another 4-wise independent hash for deciding in which of the T slices an element lies, and T bits to encode the β_i's. Thus for a given T the size of $\hat{\mathsf{D}}$ will be $T + \tilde{O}(n)$. Using the above equation, we then get a smooth tradeoff $s = O(2^k\epsilon^2\delta^{-2})$ between the advantage ϵ and the circuit size s. This discussion shows that to complete the argument we need the following two properties of the partition (a) the mass difference between P_X and P_Y is (roughly) equidistributed among slices and (b) the membership in partition slices can be efficiently decided.

Efficient slicing using 4-wise independence To complete the argument, we assume that T is a power of 2, and generate the slicing by using a 4-universal hash function $h : \{0,1\}^n \to \{0,1\}^{\log T}$. The i-th slice S_i is defined as $\{x \in \{0,1\}^n : h(x) = i\}$. These assumptions are enough to prove that

$$\mathbb{E}\,\mathsf{Adv}^{\hat{\mathsf{D}}}_{S_i}(X;Y) = \Omega\left(T^{-\frac{1}{2}}d_2(P_X;P_Y)\right) = \Omega\left(T^{-\frac{1}{2}}2^{-\frac{k}{2}}\delta\right).$$

Interestingly, the expected advantage (left-hand side) cannot be computed directly. The trick here is to bound it in terms of the second and fourth moment. The above inequality, coupled with bounds on second moments of the advantage $\mathsf{Adv}^{\hat{\mathsf{D}}}_{S_i}$ (obtained directly), allows us to prove that

$$\Pr\left[\sum_{i=1}^{T} \mathsf{Adv}^{\hat{\mathsf{D}}}_{S_i} \geqslant \Omega(1) \cdot T^{\frac{1}{2}}2^{-\frac{k}{2}}\delta\right] > \Omega(1).$$

This shows that there exists the claimed distinguisher $\hat{\mathsf{D}}$. In fact, a *constant fraction* of generated (over the choice of D and $\{\S_i\}_i$) distinguishers $\hat{\mathsf{D}}$'s works.

Random walks From a technical point of view, our method involves computing higher moments of the advantages to obtain concentration and anti-concentration results. The key observation is that the advantage written down as

$$\mathsf{Adv}^{\mathsf{D}}_{S_i}(X;Y) = \left|\sum_x (P_X(x) - P_Y(x))\mathbf{1}_{S_i}(x)\mathsf{D}(x)\right|$$

which can be then studied as a *random walk*

$$\mathsf{Adv}^{\mathsf{D}}_{S_i}(X;Y) = \left|\sum_x \xi_{i,x}\right|$$

with zero-mean increments $\xi_{i,x} = (P_X(x) - P_Y(x))\mathbf{1}_{S_i}(x)\mathsf{D}(x)$. The difference with respect to classical model is that the increments are only ℓ-wise independent (for $\ell = 4$). However, the classical moment bounds still apply (see Sections 2.2 and 2.3 for more details).

2 Preliminaries

2.1 Interpolation Inequalities

Interpolation inequalities show how to bound the p-th moment of a random variable if we know bounds on one smaller and one higher moment. The following result is known also as *log-convexity of L_p norms*, and can be proved by the Hölder Inequality

Lemma 2 (Moments interpolation). *For any $p_1 < p < p_2$ and any bounded random variable Z we have*

$$\|Z\|_p \leqslant \left(\|Z\|_{p_1}\right)^{\theta} \left(\|Z\|_{p_2}\right)^{1-\theta}$$

where θ is such that $\frac{\theta}{p_1} + \frac{1-\theta}{p_2} = \frac{1}{p}$, and for any r we define $\|Z\|_r = \left(\mathbb{E}\,|Z|^r\right)^{\frac{1}{r}}$.

Alternatively, we can *lower bound* a moment given *two higher moments*. This is very useful when higher moments are easier to compute. In this work will bound first moments from below when we know the second and the fourth moment (which are easier to compute as they are even-order moments)

Corollary 1. *For any bounded Z we have $\mathbb{E}\,|Z| \geqslant \frac{\left(\mathbb{E}\,|Z|^2\right)^{\frac{3}{2}}}{\left(\mathbb{E}\,|Z|^4\right)^{\frac{1}{2}}}$.*

2.2 Moments of random walks

For a random walk $\sum_x \xi(x)$, where $\xi(x)$ are independent with zero-mean, we have good control over the moments, namely $\mathbb{E}\left|\sum_x \xi(x)\right|^p = \Theta(1) \cdot \left(\sum_x \mathrm{Var}(\xi(x))\right)^{\frac{p}{2}}$ where constants depend on p. This result is due to Marcinkiewicz and Zygmund [MZ38] who extended the former result of Khintchine [Khi24]. Below we notice that for *small moments* p it suffices to assume only p-wise independence (most often used versions assume fully independence)

Lemma 3 (Strengthening of Marcinkiewicz-Zygmund's Inequality for $p = 4$). *Suppose that $\{\xi(x)\}_{x\in\mathcal{X}}$ are 4-wise independent, with zero mean. Then*

we have

$$\frac{1}{\sqrt{3}}\left(\sum_{x\in\mathcal{X}}\mathrm{Var}(\xi(x))\right)^{\frac{1}{2}} \leqslant \mathbb{E}\left|\sum_{x\in\mathcal{X}}\xi(x)\right| \leqslant \left(\sum_{x\in\mathcal{X}}\mathrm{Var}(\xi(x))\right)^{\frac{1}{2}}$$

$$\mathbb{E}\left|\sum_{x\in\mathcal{X}}\xi(x)\right|^2 = \sum_{x\in\mathcal{X}}\mathrm{Var}(\xi(x))$$

$$\left(\sum_{x\in\mathcal{X}}\mathrm{Var}(\xi(x))\right)^2 \leqslant \mathbb{E}\left|\sum_{x\in\mathcal{X}}\xi(x)\right|^4 \leqslant 3\left(\sum_{x\in\mathcal{X}}\mathrm{Var}(\xi(x))\right)^2$$

The proof appears in Section 4.1.

2.3 Anticontentration bounds

Lemma 4 (Paley-Zygmund Inequality). *For any positive random variable Z and a parameter $\theta \in (0,1)$ we have*

$$\Pr\left[Z > \theta\,\mathbb{E}\,Z\right] \geqslant (1-\theta)^2\frac{(\mathbb{E}\,Z)^2}{\mathbb{E}\,Z^2}.$$

By applying Lemma 4 to the setting of Lemma 3, and choosing $\theta = \frac{1}{\sqrt{3}}$ we obtain

Corollary 2 (Anticoncentration for walks with 4-wise independent increments). *Suppose that $\{\xi(x)\}_{x\in\mathcal{X}}$ are 4-wise independent with zero-mean, then we have*

$$\Pr\left[\left|\sum\xi(x)\right| > \frac{1}{3}\left(\sum\mathrm{Var}(\xi(x))\right)^{\frac{1}{2}}\right] > \frac{1}{17}.$$

where the summation is over $x\in\mathcal{X}$.

3 Proof of Lemma 1

Lemma 5 (Characterizing smooth min-entropy). *For any random variable X with values in a finite set $\mathcal{X}$, any δ and k we have the following equivalence*

$$H_\infty^\delta(X) \geqslant k \iff \sum_{x\in\mathcal{X}}\max\left(P_X(x) - 2^{-k}, 0\right) \leqslant \delta.$$

The proof appears in Section 4.2. We will work with the following equivalent statement

Corollary 3 (No smooth min-entropy k implies bias w.r.t. distributions of min-entropy k over at most 2^k elements). *We have $\mathsf{H}^{\delta}_{\infty}(X) < k$ if and only if there exists a set S of at most 2^k elements such that*

$$\sum_{x \in S} |P_X(x) - P_Y(x)| > \delta$$

for all Y of min-entropy at least k.

Proof (Proof of Corollary 3). The direction $\Longleftarrow$ trivially follows by the definition of smooth min-entropy. Now assume $\mathsf{H}^{\delta}_{\infty}(X) < k$. Let S be the set of all x such that $P_X(x) > 2^{-k}$, then $|S| < 2^k$, and moreover by Lemma 5 we have $\sum_{x \in S} \left(P_X(x) - 2^{-k}\right) > \delta$. In particular for any Y of min-entropy k (i.e., $P_Y(x) \leqslant 2^{-k}$ for all x)

$$\sum_{x \in S} \left(P_X(x) - P_Y(x)\right) > \delta$$

Lemma 6 (Bias implies Euclidean distance). *For any distributions P_X, P_Y on $\mathcal{X}$ and any subset S of $\mathcal{X}$ we have*

$$\left(\sum_{x \in S} \left(P_X(x) - P_Y(x)\right)^2\right)^{\frac{1}{2}} > |S|^{-1/2} \sum_{x \in S} |P_X(x) - P_Y(x)| .$$

Proof. By the Jensen Inequality we have

$$|S|^{-1} \left(\sum_{x \in S} \left(P_X(x) - P_Y(x)\right)^2\right) > \left(|S|^{-1} \sum_{x \in S} |P_X(x) - P_Y(x)|\right)^2$$

which is equivalent to the statement.

Corollary 4 (No smooth min-entropy implies Euclidean distance to min-entropy distributions). *Suppose that $\mathsf{H}^{\delta}_{\infty}(X) < k$. Then for any Y of min-entropy at least k we have $\left(\sum_x |P_X(x) - P_Y(x)|^2\right)^{\frac{1}{2}} > 2^{-\frac{k}{2}} \delta$.*

Proof (Proof of Corollary 4). It suffices to combine Lemma 6 and Corollary 3.

By Corollary 2 we conclude that the advantage of a random distinguisher for any two measures (in our case P_X and P_Y) equals the Euclidean distance.

Lemma 7 (The advantage of a random distinguisher equals the Euclidean distance). *Let $\{\mathsf{D}(x)\}_{x \in \{0,1\}^n}$ be 4-wise independent as indexed by x and such that $\mathsf{D}(x)$ outputs a random element from $\{-1, 1\}$. Then for any set S we have*

$$\left|\sum_{x \in S} \mathsf{D}(x)(P_X(x) - P_Y(x))\right| > \frac{1}{3} \cdot d_2(P_X; P_Y|S)$$

with probability $\frac{1}{17}$ over the choice of D (the result actually holds for any measures in place of P_X, P_Y).

For our case, that is the setting in Lemma 6, we obtain

Corollary 5 (A random attack achieves $\Omega\left(2^{-k}\delta\right)$ with significant probability). *For X, Y as in Corollary 4, and D as in Lemma 7 we have $\mathsf{Adv}^{\mathsf{D}}(X;Y) \geq \frac{1}{3}\cdot 2^{-\frac{k}{2}}\delta$ w.p. $\frac{1}{17}$ over D.*

3.1 Partitioning the domain into T slices

Let $h : \{0,1\}^n \to [1 \dots 2^t]$, where $t = \lceil \log T \rceil$, be a 4-universal hash function. Define $S_i = \{x : h(x) = i\}$, $\Delta(x) = P_X(x) - P_Y(x)$ and consider advantages on slices S_i

$$\mathsf{Adv}^{\mathsf{D}}_{S_i}(X;Y) = \left|\sum_x \Delta(x)\mathsf{D}(x)\mathbf{1}_{S_i}(x)\right|$$

The following corollary shows that on each of our T slices, we get the advantage $T^{-\frac{1}{2}}2^{-\frac{k}{2}}\delta$. The proof appears in Section 4.3.

Corollary 6 ((Mixed) moments of slice advantages). *For D, $\{S_u\}_u$ as above and every i, j*

$$\mathbb{E}_{\mathsf{D},\{S_u\}_u}\mathsf{Adv}^{\mathsf{D}}_{S_i}(X;Y) \geqslant 3^{-\frac{1}{2}}T^{-\frac{1}{2}}\cdot d_2\left(P_X;P_Y\right)$$
$$\mathbb{E}_{\mathsf{D},\{S_u\}_u}\left(\mathsf{Adv}^{\mathsf{D}}_{S_i}(X;Y)\,\mathsf{Adv}^{\mathsf{D}}_{S_j}(X;Y)\right) \leqslant T^{-1}\cdot d_2(P_X;P_Y)^2$$

(the statement is valid for arbitrary measures in place of P_X, P_Y).

Denote $Z = \sum_i \mathsf{Adv}^{\mathsf{D}}_{S_i}(X;Y)$. Using Lemma 4 with $\theta = \frac{1}{\sqrt{3}}$ where we compute $\mathbb{E}\,Z^2$ and $\mathbb{E}\,Z$ according to Corollary 6 we obtain $\Pr\left[|Z| > \frac{1}{\sqrt{3}}\cdot\mathbb{E}\,|Z|\right] \geqslant \frac{1}{17}$. Bounding once again $\mathbb{E}\,|Z|$ as in Corollary 6 we get

Corollary 7 (Total advantage on all parition slices). *For X, Y as in Corollary 4, D and S_i defined above we have*

$$\Pr_{\mathsf{D},\{S_u\}_u}\left[\sum_{i=1}^{T}\mathsf{Adv}^{\mathsf{D}}_{S_i}(X;Y) \geqslant \frac{1}{3}\cdot T^{\frac{1}{2}}2^{-\frac{k}{2}}\delta\right] \geqslant \frac{1}{17}.$$

(for general X, Y the lower bound is $\Omega(1)\cdot T^{\frac{1}{2}}\cdot d_2(P_X;P_Y)$).

The corollary shows that the *total absolute advantage* over all partition slices, is as expected. Since $\{S_i\}_i$ is a partition we have

$$\sum_{i=1}^{T}\mathsf{Adv}^{\mathsf{D}}_{S_i}(X;Y) = \sum_{i=1}^{T}\left|\sum_{x\in S_i}\left(P_X(x)-P_Y(x)\right)\mathsf{D}(x)\right| = \sum_x\left(P_X(x)-P_Y(x)\right)\mathsf{D}(x)\beta(x)$$

where for $\beta_i \stackrel{def}{=} \operatorname{sgn}\left(\sum_{x\in S_i}\left(P_X(x')-P_Y(x)\right)\mathsf{D}(x)\right)$ (the sign of the advantage on the i-th slice) we define $\beta(x) = \beta_i$ where S_i contains x. This shows that by

"flipping" the distinguisher output on the slices we achieve the sum of individual advantages. Since the bit $\beta(x)$ can be computed with $O(T) + \tilde{O}(n)$ advice (the complexity of the function $i \to \beta_i$ plus the complexity of finding i for a given x) we obtain

Corollary 8 (Computing total advantage by one distinguisher). *For X, Y as in Corollary 4, D and $\{S_i\}_i$ defined above there exists a modification to D which in time $\tilde{O}(n)$ and advice $O(T)$ achieves advantage $\frac{1}{3} \cdot T^{\frac{1}{2}} 2^{-\frac{k}{2}} \delta$ with probability $\frac{1}{17}$.*

Finally by setting $\epsilon = T^{\frac{1}{2}} 2^{-\frac{k}{2}} \delta$ and manipulating T we arrive at

Corollary 9 (Continue tradeoff). *For any ϵ there exists T such that the distinguisher in Corollary 8 has advantage ϵ and circuit complexity $s = O\left(2^k \epsilon^2 \delta^{-2}\right)$.*

4 Omitted Proofs

4.1 Proof of Lemma 3 (Strengthening of Marcinkiewicz-Zygmund's Inequality for $p = 4$)

Let $Z = \sum_x \xi(x)$. Since $\xi(x)$ are (in particular) 2-wise independent with zero mean, we get

$$\mathbb{E}\left(\sum_x \xi(x)\right)^2 = \sum_{x,y} \mathbb{E}\left(\xi(x)\xi(y)\right) = \sum_{x=y} \mathbb{E}\left(\xi(x)\xi(y)\right) = \sum_x \mathrm{Var}(\xi(x)).$$

(the summation taken over $x, y \in \mathcal{X}$). The fourth moment is somewhat more complicated

$$\begin{aligned}
\mathbb{E}\left(\sum_x \xi(x)\right)^4 &= \sum_{x_1,x_2,x_3,x_4} \mathbb{E}\left(\xi(x_1)\xi(x_2)\xi(x_3)\xi(x_4)\right) \\
&= \sum_{x_1=x_2=x_3=x_4} \mathbb{E}\left(\xi(x_1)\xi(x_2)\xi(x_3)\xi(x_4)\right) + \\
&\quad + 3 \sum_{x_1=x_2\neq x_3=x_4} \mathbb{E}\left(\xi(x_1)\xi(x_2)\xi(x_3)\xi(x_4)\right) \\
&= \sum_x \mathbb{E}\,\xi(x)^4 + 3 \sum_{x\neq y} \mathbb{E}\,\xi(x)^2\, \mathbb{E}\,\xi(y)^2 \\
&= 3\left(\sum_x \mathbb{E}\,\xi(x)^2\right)^2 - 2\sum_x \mathbb{E}\,\xi(x)^4
\end{aligned}$$

The second equality follows because whenever $\xi(x)$ occurs in an odd power, for example $x = x_1 \neq x_2 = x_3 = x_4$, the expectation is zero (this way one can simplify and bound also higher moments, see [SSS93]). It remains to estimate

the first moment. By Corollary 1 and bounds on the second and fourth moment we have just computed we obtain

$$\frac{1}{\sqrt{3}} \cdot \left(\sum_{x \in \mathcal{X}} \mathrm{Var}(\xi(x)) \right)^{\frac{1}{2}} \leqslant \mathbb{E} \left| \sum_{x \in \mathcal{X}} \xi(x) \right|$$

and the upper bound follows by Jensen's Inequality (with constant 1).

4.2 Proof of Lemma 5 (Characterizing smooth min-entropy)

Suppose that $\mathsf{H}_\infty^\delta(X) \geqslant k$. then, by definition, there is Y such that $\mathsf{H}_\infty(Y) \geqslant k$ and $\sum_{x:P_X(x)>P_Y(x)} P_X(x) - P_Y(x) \leqslant \delta$. Since all the summands are positive and since $P_Y(x) \leqslant 2^{-k}$, ignoring those x for which $P_Y(x) < 2^{-k}$ yields

$$\sum_{x:P_X(x)>2^{-k}} P_X(x) - P_Y(x) \leqslant \delta.$$

Again, since $P_Y(x) \leqslant 2^{-k}$ we obtain

$$\sum_{x:P_X(x)>2^{-k}} P_X(x) - 2^{-k} \leqslant \delta,$$

which finishes the proof of the "$\Longrightarrow$" part.

Assume now that $\delta' = \sum_{x \in \mathcal{X}} \max\left(P_X(x) - 2^{-k}, 0\right) \leqslant \delta$. Note that

$$\sum_{x \in \mathcal{X}} \max\left(P_X(x) - \frac{1}{2^k}, 0\right) + \sum_{x \in \mathcal{X}} \max\left(\frac{1}{2^k} - P_X(x), 0\right) =$$
$$= 2 \sum_{x \in \mathcal{X}} \left| P_X(x) - \frac{1}{2^k} \right| \geqslant 2 \sum_{x \in \mathcal{X}} \max\left(P_X(x) - \frac{1}{2^k}, 0\right)$$

and therefore we have $\sum_{x \in \mathcal{X}} \max\left(2^{-k} - P_X(x), 0\right) \geqslant \delta'$. By this observation we can construct a distribution Y by shifting δ' of the mass of P_X from the set $S^- = \{x : P_X(x) > 2^{-k}\}$ to the set $\{x : 2^{-k} \geqslant P_X(x)\}$ in such a way that we have $P_Y(x) \leqslant 2^{-k}$ for all x. Thus $\mathsf{H}_\infty(Y) \geqslant k$ and since a δ' fraction of the mass is shifted and redistributed we have $d_1(X;Y) \leqslant \delta'$. This finishes the proof of the "$\Longleftarrow$" part.

4.3 Proof of Corollary 6 ((Mixed) moments of slice advantages)

For shortness denote $\Delta(x) = P_X(x) - P_Y(x)$ and $\mathsf{Adv}_{S_i}^{\mathsf{D}} = \mathsf{Adv}_{S_i}^{\mathsf{D}}(X;Y)$.

For any fixed i we apply Lemma 3 to the family $f_x = \Delta(x)\mathsf{D}(x)\mathbf{1}_{S_i}(x)$ which is 4-wise independent, zero-mean, and with the second moment $T^{-1}\sum_x \Delta(x)^2$. We obtain

$$\mathbb{E}\,\mathsf{Adv}_{S_i}^{\mathsf{D}} \geqslant 3^{-\frac{1}{2}} \left(T^{-1} \sum_x \Delta(x)^2 \right)^{\frac{1}{2}}$$

which is the first inequality claimed in the corollary. In turn, again by Lemma 3

$$\mathbb{E}\left(\mathsf{Adv}^{\mathsf{D}}_{S_i}\right)^2 = T^{-1} \cdot \sum_x \Delta(x)^2.$$

Since this holds for any i, by Cauchy-Schwarz we get for any i, j

$$\mathbb{E}\,\mathsf{Adv}^{\mathsf{D}}_{S_i}\mathsf{Adv}^{\mathsf{D}}_{S_j} \leqslant \sqrt{\mathbb{E}\left(\mathsf{Adv}^{\mathsf{D}}_{S_i}\right)^2 \cdot \mathbb{E}\left(\mathsf{Adv}^{\mathsf{D}}_{S_j}\right)^2} \leqslant T^{-1} \cdot \sum_x \Delta(x)^2.$$

which proves the second inequality in the corollary.

References

Ber97. Bonnie Berger, *The fourth moment method*, SIAM J. Comput. **26** (1997), no. 4, 1188–1207.

DPW14. Yevgeniy Dodis, Krzysztof Pietrzak, and Daniel Wichs, *Key derivation without entropy waste*, Advances in Cryptology – EUROCRYPT 2014 (PhongQ. Nguyen and Elisabeth Oswald, eds.), Lecture Notes in Computer Science, vol. 8441, Springer Berlin Heidelberg, 2014, pp. 93–110 (English).

DTT10. Anindya De, Luca Trevisan, and Madhur Tulsiani, *Time space tradeoffs for attacks against one-way functions and prgs*, Advances in Cryptology - CRYPTO 2010, 30th Annual Cryptology Conference, Santa Barbara, CA, USA, August 15-19, 2010. Proceedings, 2010, pp. 649–665.

Khi24. Aleksandr Khintchine, *Über einen satz der wahrscheinlichkeitsrechnung*, Fundamenta Mathematicae **6** (1924), no. 1, 9–20 (ger).

MZ38. J. Marcinkiewicz and A. Zygmund, *Quelques théorèmes sur les fonctions indépendantes*, Studia Mathematica **7** (1938), no. 1, 104–120 (fre).

RW05. Renato Renner and Stefan Wolf, *Simple and tight bounds for information reconciliation and privacy amplification*, Advances in Cryptology - ASIACRYPT 2005, 11th International Conference on the Theory and Application of Cryptology and Information Security, Chennai, India, December 4-8, 2005, Proceedings, 2005, pp. 199–216.

SSS93. Jeanette P. Schmidt, Alan Siegel, and Aravind Srinivasan, *Chernoff-hoeffding bounds for applications with limited independence*, Proceedings of the Fourth Annual ACM/SIGACT-SIAM Symposium on Discrete Algorithms, 25-27 January 1993, Austin, Texas., 1993, pp. 331–340.

Chapter 6

Geometric Characterization

Metric Pseudoentropy: Characterizations and Applications

Maciej Skorski
maciej.skorski@gmail.com

Cryptology and Data Security Group, University of Warsaw

Abstract. Metric entropy is a computational variant of entropy, often used as a convenient substitute of HILL Entropy, slightly stronger and standard notion for entropy in cryptographic applications. In this paper we develop a general method to characterize metric-type computational variants of entropy, in a way depending only on properties of a chosen class of test functions (adversaries). As a consequence, we obtain a nice and elegant geometric interpretation of metric entropy. We apply these characterization to simplify and modularize proofs of some important results, in particular: (a) computational dense model theorem, (b) derivation of the improved version of Leftover Hash Lemma and (c) equivalence between unpredictability entropy and HILL entropy for short strings.

1 Introduction

1.1 Computational Entropy

Entropy. Entropy, as a measure of uncertainty or randomness, is a fundamental notion in information-theory. The most known metric of entropy is Shannon Entropy [Sha48]. For cryptographic applications such as *extracting randomness*, it is more convenient to work with so called min-entropy, which gives an upper bound on the probability that computationally unbounded adversary can guess a value sampled according to a given distribution. A slightly weaker but also very useful, especially in the context of *hashing*, is the notion of collision entropy which upperbounds the probability that two independent samples of a given distribution collide.

Defining computational variants of entropy. Computational analogues of entropy can be defined in different ways. In any case, we need to formalize that a distribution has, from a computational point of view, the same of almost the same properties like a distribution having "true" information-theoretic entropy. This might be based on hardness of compressing-decompressing, hardness of prediction or hardness of distinguishing. In this paper we follow the last approach, which is most widely used. A good survey of different entropy notions and their properties can be found in [BSW03] and [Rey11]. We stress that, contrarily to the information-theoretic case, for computational entropy it's not only the *amount* of entropy that matters but also its *quality* is important.

Computational Indistinguishability. Indistinguishability is a fundamental concept in computational complexity and cryptography. For two distributions X, Y taking values in the same space, a class $\mathcal{D}$ of $[0,1]$-valued functions (refereed to as the "attackers class") and a parameter ϵ (refereed to as the "distinguishing advantage"), we say that X and Y are $(\mathcal{D}, \epsilon)$-indistinguishable if for all $\mathrm{D} \in \mathcal{D}$ we have $|\mathbf{E}\,\mathrm{D}(X) - \mathbf{E}\,\mathrm{D}(Y)| \leqslant \epsilon$. An attacker D can distinguish X and Y if $\mathbf{E}\,\mathrm{D}(X) - \mathbf{E}\,\mathrm{D}(Y) > 0$ or $\mathbf{E}\,\mathrm{D}(X) - \mathbf{E}\,\mathrm{D}(Y) < 0$, and the far from 0 this difference is, the better "advantage" he achieves. Sometimes we want to define indistinguishability between two sets $\mathbb{X}$ and $\mathbb{Y}$ of probability distributions. We can formalize this by saying that no single adversary D can achieve bigger than 0 advantage for *every* pair (X, Y) where X comes from $\mathbb{X}$ and Y comes from $\mathbb{Y}$. Since the expectation $\mathbf{E}\,\mathrm{D}(X)$ can be thought as the scalar product of vectors representing D and the distribution of X, the concept of indistinguishability is *exactly* the same concept as the idea of *separating hyperplanes*.

Computational Entropy. Having formalized the concept of "computational closeness", one can define the "computational" entropy, called also pseudoentropy, of a distribution X by one of the following ways:

(a) (stronger) X is computationally indistinguishable from a *single* distribution having required amount of information-theoretic entropy (min-entropy, Shannon Entropy etc.)
(b) (weaker) is computationally indistinguishable from a *set* of all distributions having required amount of information-theoretic entropy.

Both approaches turn out to be useful. Setting the underlying information-theoretic entropy measure to be the min-entropy, for case (a) we obtain the notion of HILL entropy [HILL99] which directly generalizes the notion of pseudorandomness, whereas for case (b) we get the notion of the so called Metric Entropy [BSW03]. Roughly speaking, with HILL entropy one generalizes most of information-theoretic facts about entropy, into the computational setting. Metric entropy is commonly thought as a less intuitive and understood notion than HILL entropy. Quite surprisingly it has been proven to be technically more convenient in many problems. The typical approach is to work with metric entropy and to convert it to HILL entropy (which is possible with some loss in quality [BSW03]). For example, the use of metric entropy simplifies and improves the proof of the computational variant of the dense model theorem [FOR12], applicable in leakage-resilient cryptography [DP08]. Notions of pseudoentropy have found also important applications in general complexity theory, for example in [VZ12] a HILL-like variant of Shannon entropy is used to simplify the construction of a PRG from a one-way function. These two examples show also that the notion of pseudoentropy is a key ingredient of important or even breakthrough results and as such is worth of studying.

Worst Case Distributions. In problems which involve computational indistinguishability it is often convenient to know the distributions which makes the attacker's advantage maximal. This distribution is typically subjected to some entropy restrictions. In particular, one might ask the following question

> Given D and X, what is the best (minimal) attacker advantage $|\Delta^D| = |\mathbf{E}\,\mathrm{D}(X) - \mathbf{E}\,\mathrm{D}(Y)|$ over all distributions Y of entropy as least k?

An answer to this question yields a bound on how (computationaly) close is X to the set of all distributions of entropy k. Such problems arises naturally where one uses HILL and Metric entropy, see for instance [BSW03, CKLR11, VZ12].

1.2 Our Results

Summary of our contribution. As mentioned, the concept of characterizing the "worst case" distribution which optimizes the attacker advantage is very common, thought not always explicitly stated [BSW03, CKLR11, FOR12, RTTV08]. In this paper we give a uniform treatment of this idea and use to obtain characterizations for pseudoentropy and other interesting corollaries.

Characterizing Metric Pseudoentropy via Optimizing Attacker's Advantage. Using standard constrained optimization techniques, we develop a general method to characterize metric-type pseudoentropy. A characterization is based on *explicitly* calculating the distribution which minimizes the attacker's advantage, subject to entropy constraints. These characterizations could be used in studying properties of variants of pseudoentropy based on entropy different than min-entropy. In particular, they could be applied in studying the problem of comparing the amount of metric pseudoentropy against *deterministic* and *randomized* adversaries, or verifying the so called "chain rule". We also unify the definitions of metric and HILL entropy in a nice geometric way.

Applications: the power of pseudoentropy characterizations. Our technique leads to interesting corollaries besides the basic properties of pseudoentropy. From the characterization of metric pseudo-entropy we immediately obtain the computational Dense Model Theorem [RTTV08, DP08, FOR12] Extending our characterization into the conditional case when side information is available to the attacker, we reprove equivalence between unpredictability and indistinguishability based definition of pseudoentropy for short strings [VZ12]. Finally, from the characterization of collision-pseudoentropy we derive the improved Leftover Hash Lemma [BDK+11]. Our results show that metric entropy is a powerful tool which deserves the systematic study.

2 Preliminaries

Entropy notions. The min-entropy of a distribution X equals $\mathbf{H}_\infty(X) = -\log(\max_x \Pr[X = x])$. The collision entropy of X is $\mathbf{H}_2(X) = -\log(\sum_x \Pr[X = x]^2)$. If there is side information Z, we define the average conditional min-entropy [DORS08] of X given Z by $\widetilde{\mathbf{H}}_\infty(X|Z) = -\log(\mathbf{E}_{z\leftarrow Z} \max_x \Pr[X = x|Z = z])$.

Computational advantage. The advantage of an attacker D in distinugishing random variables X and Y, which take values in the same space, is defined to be $\Delta^{\mathrm{D}}(X;Y) = \mathbf{E}\,\mathrm{D}(X) - \mathbf{E}\,\mathrm{D}(Y)$.

Computational Entropy. There is many ways to define computational analogues of entropy. We follow the most popular approach, which is based on the concept of computational indistinguishability.

Definition 1 (HILL Pseudoentropy [HILL99]). *Let X be a distribution with the following property: there exists Y of min-entropy at least k such that for all circuits* D *of size at most s we have $|\Delta^{\mathrm{D}}(X;Y)| \leqslant \epsilon$. Then we say that X has k bits of HILL min-entropy of quality (s,ϵ) and denote by $\mathbf{H}_{\infty}^{\mathrm{HILL},(s,\epsilon)}(X) \geqslant k$.*

Remark 1 (HILL entropy against different circuits classes). It is known that for HILL entropy all kind of circuits: deterministic boolean, deterministic real valued and randomized boolean, are equivalent (for the same size s). That's why we can abbreviate the notation and omit declaring circuits type in Definition 1.

Definition 2 (Metric Pseudoentropy [BSW03]). *Let X be a distribution with the following property: for every deterministic boolean (respectively: deterministic real valued or boolean randomized) circuit* D *of size at most s there exists Y of min-entropy at least k such that $|\Delta^{\mathrm{D}}(X;Y)| \leqslant \epsilon$. Then we say that X has k bits of deterministic (respectively: deterministic real valued or boolean randomized) metric min-entropy of quality (s,ϵ) and denote by $\mathbf{H}_{\infty}^{\mathrm{M},\mathrm{det}\{0,1\},(s,\epsilon)}(X)$ (respectively: $\mathbf{H}_{\infty}^{\mathrm{M},\mathrm{det}[0,1],(s,\epsilon)}(X)$ and $\mathbf{H}_{\infty}^{\mathrm{M},\mathrm{rand}\{0,1\},(s,\epsilon)}(X)$).*

Definitions of HILL and metric entropy for entropy notions different than min-entropy, for instance collision entropy can be obtained by replacing min-entropy with collision entropy in Definition 1 and Definition 2.

Remark 2 (Metric Entropy against different circuits class). For metric min-entropy, it does not matter if the deterministic circuits are boolean or real valued (see [Rey11] and the errata of [BSW03]). However, this is not true for the conditional case and does not extend to other entropy notions.

Computational Entropy - Side information. Sometimes we assume that information Z correlated to X might be available to an adversary.

Definition 3 (Conditional HILL Pseudoentropy [HLR07]). *Let X, Z be a joint distribution with the following property: there exists Y of average conditional min-entropy at least k given Z such that for all circuits* D *of size at most s we have $|\Delta^{\mathrm{D}}(X,Z;Y,Z)| \leqslant \epsilon$. Then we say that X given Z has k bits of HILL min-entropy of quality (s,ϵ) and denote by $\mathbf{H}_{\infty}^{\mathrm{HILL},(s,\epsilon)}(X|Z) \geqslant k$.*

Remark 3 (HILL entropy against different circuits classes). Similarly to Remark 2, here all kinds of circuits: deterministic boolean, deterministic real valued and randomized boolean, are equivalent (for the same size s).

Definition 4 (Conditional Metric Pseudoentropy [FOR12]). *Let X, Z be a joint distribution with the following property: for every deterministic boolean (respectively: deterministic real valued or boolean randomized) circuit* D *of size at most s there exists Y of average conditional min entropy at least k given Z*

such that $|\Delta^{\mathrm{D}}(X,Z;Y,Z)| \leqslant \epsilon$. *Then we say that* X *given* Z *has* k *bits of deterministic (respectively: deterministic real valued or boolean randomized) metric min-entropy of quality* (s,ϵ) *and denote by* $\mathbf{H}_{\infty}^{\mathrm{M},\mathrm{det}\{0,1\},(s,\epsilon)}(X|Z)$ *(respectively:* $\mathbf{H}_{\infty}^{\mathrm{M},\mathrm{det}[0,1],(s,\epsilon)}(X|Z)$ *and* $\mathbf{H}_{\infty}^{\mathrm{M},\mathrm{rand}\{0,1\},(s,\epsilon)}(X|Z)$*).*

There is a variant of conditional pseudoentropy where (X,Z) is required to be computationally close to (Y,Z') but Z' is not necessarily the same as Z. This notion is called the "relaxed" HILL entropy [Rey11] and denoted by $\mathbf{H}^{\mathrm{HILL-rlx},(s,\epsilon)}(X)$ (for metric variants $\mathbf{H}^{\mathrm{M-rlx},\mathrm{det}\{0,1\},(s,\epsilon)}(X)$ and $\mathbf{H}^{\mathrm{M-rlx},\mathrm{det}[0,1],(s,\epsilon)}(X)$). Typically we want Z to be the same as Z'[1] but this relaxed notion is also useful [GW11, Rey11]. It satisfies the so called chain rule, a property desired in leakage-resilient cryptography, which doesn't hold for HILL entropy [KPW13].

Relations between HILL and Metric Pseudoentropy. For any "reasonable" notion of (information-theoretic) entropy, metric and HILL variants are equivalent up to some loss in quality parameters s,ϵ.

Lemma 1 (HILL vs Metric Pseudoentropy, [BSW03]). *Let* $\mathbf{H}$ *be an entropy notion which is concave*[2]*. Then for any* n*-bit random variable* X *we have*

$$\mathbf{H}^{\mathrm{HILL},(s',\epsilon')}(X) \geqslant \mathbf{H}^{\mathrm{M},\mathrm{det}[0,1],(s,\epsilon)}(X)$$

where $\delta \in (0,1)$ *is arbitrary,* $s' = \mathcal{O}\left(s \cdot \delta^2/n\right)$ *and* $\epsilon' = \epsilon + \delta$*. The same is true for conditional pseudoentropy and relaxed pseudoentropy, with* $s' = \mathcal{O}\left(s \cdot \frac{\delta^2}{n+m}\right)$ *where* m *is the length of* Z*.*

3 Characterizing Metric Pseudoentropy

In what follows we assume that $\mathbf{H}$ is a concave entropy notion (like min-entropy or collision entropy), and that all distributions and distinguishers are over $\{0,1\}^n$.

3.1 Connections to separating hyperplanes

We start with the following simple observation, which gives a nice geometrical formulation of the definition of pseudo-entropy. We say that the sets $\mathbb{X}$ and $\mathbb{Y}$ of probability distributions are $(\mathcal{D},\epsilon)$-indistinguishable if there exists *no* adversary D such that $|\mathbf{E}\,\mathrm{D}(X) - \mathbf{E}\,\mathrm{D}(Y)| \geqslant \epsilon$ for all $X \in \mathbb{X}$ and all $Y \in \mathbb{Y}$. It is easy to see that if $\mathbb{X}$ and $\mathbb{Y}$ are convex and if $\mathcal{D}$ is closed under complements (that is $\mathrm{D} \in \mathcal{D}$ implies $\mathbf{1} - \mathrm{D} \in \mathcal{D}$) then this is equivalent to

There is no $\mathrm{D} \in \mathcal{D}$ such that: $\mathbf{E}\,\mathrm{D}(X) - \mathbf{E}\,\mathrm{D}(Y) \geqslant \epsilon$ for all $X \in \mathbb{X}, Y \in \mathbb{Y}$.

[1] For instance, when Z represents information that adversary might have learned

[2] That is, a convex combination of distributions with entropy at least k is a distribution with entropy at least k. This assumption is fulfilled for most notions, for example for all Renyi entropies which include min-entropy and collision entropy

We can interpret the expectation $\mathbf{E}\,\mathrm{D}(X)$ as the scalar product $\langle \mathrm{D}, \mathbf{P}_X \rangle$ by identifying D and distributions of X with the vectors in $\mathbb{R}^{2^n}$. Hence we can write the above condition as

$$\text{There is no } \mathrm{D} \in \mathcal{D} \text{ such that: } \langle \mathrm{D}, \mathbf{P}_X - \mathbf{P}_Y \rangle \geqslant \epsilon \text{ for all } X \in \mathbb{X}, Y \in \mathbb{Y},$$

which means that the distinguisher D is precisely a *separating hyperplane*. If $\mathcal{D}$ is a circuit class, $\mathbb{X} = \{X\}$ and $\mathbb{Y} = \{Y : \mathbf{H}(Y) \geqslant k\}$ we obtain[3]

Corollary 1 (Alternative definitions of metric and HILL entropy). *Let X be an n-bit random variable and let $\mathbf{H}$ be a concave entropy notion. Then*

(a) $\mathbf{H}^{\mathrm{HILL},(s,\epsilon)}(X) \geqslant k$ *iff X is $(\mathcal{D}, \epsilon)$-indistinguishable from* some *Y of entropy $\mathbf{H}$ at least k, where $\mathcal{D}$ is the class of boolean circuits*[4] *of size s with n-inputs.*
(b) $\mathbf{H}^{\mathrm{M},\det\{0,1\},(s,\epsilon)}(X) \geqslant k$ *iff X is $(\mathcal{D}, \epsilon)$-indistinguishable from* the set *of* all *Y of entropy $\mathbf{H}$ at least k,*

where $\mathcal{D}$ is the class of all deterministic boolean circuits of size s with n-inputs (analogously for randomized and deterministic real valued circuits).

3.2 Reduction to constrained optimization

By the "geometric" view on pseudoentropy, given in Corollary 1, we obtain the following characterization of pseudoentropy.

Lemma 2 (Characterization of metric pseudoentropy). *Let X and $\mathbf{H}$ be as in Corollary 1. Then $\mathbf{H}^{\mathrm{M},\det\{0,1\},(s,\epsilon)}(X) \geqslant k$, respectively $\mathbf{H}^{\mathrm{M},\det[0,1],(s,\epsilon)}(X) \geqslant k$ if and only if for every boolean (respectively real valued) deterministic circuit* D *of size at most s we have*

$$\mathbf{E}\,\mathrm{D}(X) \leqslant \mathbf{E}\,\mathrm{D}(Y^*) + \epsilon,$$

where Y^* is optimal to the following optimization problem

$$\begin{array}{ll} \underset{Y}{\text{maximize}} & \mathbf{E}\,\mathrm{D}(Y) \\ \text{s.t.} & \mathbf{H}(Y) \geqslant k \end{array}. \tag{1}$$

This results is useful if we can solve the optimization problem in Equation (1). In the next subsections we explain how to solve it in general and discuss the two concrete and simple cases: min-entropy and collision entropy.

[3] We can assume that the class circuits of size at most s is closed under complements because every complement is of size at most $s + 1$. Formally we need to start with size $s' = s + 1$ but we omit this negligible difference
[4] Randomized or deterministic- it makes no difference

3.3 Maximizing expectations under convex constraints

We can characterize optimal solutions of (1) in terms of Lagrange multipliers. Due to convexity, the characterization is both: necessary and sufficient.

Lemma 3 (Maximizing expectation under convex constraints). *Let* D *be a real-valued vector in* $\mathbb{R}^d$ *and* f *be a differentiable convex real-valued function on* $\mathbb{R}^d$*. Assume that* a *is a number such that* $\min_p f(p) < a$ *where the minimum is over all probability vectors, and consider the following optimization program*

$$\begin{aligned} \underset{(p_i)_i}{\text{maximize}} \quad & \sum_i \mathrm{D}_i p_i \\ \text{s.t.} \quad & \begin{cases} f(p) \leqslant a \\ -p_i \leqslant 0 \\ \sum_i p_i = 1 \end{cases}. \end{aligned} \tag{2}$$

Then a feasible point $p = p^*$ *is optimal to* (2) *if and only if there exist* $\lambda_1 \geqslant 0, \lambda_2 \geqslant 0$ *and* $\lambda_{3i} \in \mathbb{R}$ *for* $i = 1, \ldots, m$ *such that the following relations hold*

$$\mathrm{D}_i = \lambda_1 (\nabla f(p^*))_i - \lambda_{3i} + \lambda_2 \qquad \text{for } i = 1, \ldots, m \tag{3}$$

and the following complementary conditions are satisfied:

$$\begin{aligned} p_i \cdot \lambda_{3i} &= 0 \\ (f(p) - a) \cdot \lambda_1 &= 0 \end{aligned} \tag{4}$$

Proof. The Slater Constraint Qualification holds, by the assumption on a, and we have strong duality. In other words, the first order Karush-Kuhn-Tucker condition is sufficient and necessary [BV04]. The numbers $\lambda_1, \lambda_2, \lambda_{3i}$ are exactly KKT multipliers for the convex program in Equation (2), and Equation (3) states that the gradient of the objective function is a combination of gradients of constraints. The condition in Equation (4) means that we take only active constraints into account. Finally, to the inequality constraints we assign non-negative multipliers which explains the requirement $\lambda_1 \geqslant 0$ and $\lambda_{3i} \geqslant 0$. □

Remark 4. If f is not differentiable, we replace the gradient of f in optimality conditions by the *subdifferential* of f, which always exists for a convex function.

3.4 Characterization of metric min entropy

For $\mathbf{H} = \mathbf{H}_\infty$ we obtain from Lemma 3 the following simple characterization of pseudoentropy based on min-entropy (see [BSW03] for a restricted variant). The proof appears in Appendix A.

Theorem 1 (Characterization of metric min-entropy). *Let* X *be an* n*-bit r.v.. Then* $\mathbf{H}_\infty^{\mathrm{M},\mathrm{det}\{0,1\},(s,\epsilon)}(X) \geqslant k$*, respectively* $\mathbf{H}_\infty^{\mathrm{M},\mathrm{det}[0,1],(s,\epsilon)}(X) \geqslant k$ *if and*

only if for every boolean (respectively real valued) deterministic circuit D *of size at most* s *with* n *inputs we have*

$$\mathbf{E}\,\mathrm{D}(X) \leqslant \mathbf{E}\,\mathrm{D}(Y^*) + \epsilon,$$

where Y^* *is uniform over the set of* 2^k *values of* x *which correspond to the biggest values of* $\mathrm{D}(x)$.

Extending Lemma 3 by additional constraints to cover the case of side information, we obtain the characterization of conditional metric entropy, commeting in Appendix B on its proof.

Theorem 2 (Characterization of conditional metric min-entropy). *Let* X *and* Z *be, respectively,* n *and* m*-bit random variables. Then* $\mathbf{H}_\infty^{\mathrm{M},\mathrm{det}\{0,1\},(s,\epsilon)}(X) \geqslant k$ *(respectively* $\mathbf{H}_\infty^{\mathrm{M},\mathrm{det}[0,1],(s,\epsilon)}(X) \geqslant k$*) iff for every boolean (respectively real valued) deterministic circuit* D *of size at most* s *on* $\{0,1\}^{n+m}$ *we have*

$$\mathbf{E}\,\mathrm{D}(X,Z) \leqslant \mathbf{E}\,\mathrm{D}(Y^*,Z) + \epsilon,$$

for Y^* *such that* $Y^*|Z=z$ *is uniform over the set* $\{\mathrm{D}(x,z) \geqslant t(z)\}$ *for every* z*, where the thresholds* $t(z)$ *satisfy the following two conditions*

$$\mathop{\mathbf{E}}_{x\leftarrow U_n} \mathbf{E}\max(\mathrm{D}(x,z) - t(z)) = \mathrm{const} \quad \text{for all } z$$
$$\mathop{\mathbf{E}}_{z\leftarrow Z}\left[1/\#\left\{x : \mathrm{D}(x,z) \geqslant t(z)\right\}\right] \leqslant 2^{-k} \leqslant \mathbf{E}\left[1/\#\left\{x : \mathrm{D}(x,z) > t(z)\right\}\right].$$

3.5 Characterization of metric collision entropy

The characterization of the worst-case collision entropy distribution is slightly different. It is *proportional* to a distinguisher, after taking a threshold. The proof appears in Appendix C.

Theorem 3 (Characterization of metric collision entropy).
Let X *be an* n*-bit r.v. and* $k < n$ *be integer. Then* $\mathbf{H}_2^{\mathrm{M},\mathrm{det}\{0,1\},(s,\epsilon)}(X) \geqslant k$*, respectively* $\mathbf{H}_2^{\mathrm{M},\mathrm{det}[0,1],(s,\epsilon)}(X) \geqslant k$ *if and only if for every boolean (respectively real valued) deterministic circuit* D *of size at most* s *with* n *inputs, we have*

$$\mathbf{E}\,\mathrm{D}(X) \leqslant \mathbf{E}\,\mathrm{D}(Y^*) + \epsilon,$$

where Y^* *is of colision entropy* k *such that* $\lambda \cdot \mathbf{P}_{Y^*}(x) = \max(\mathrm{D}(x) - t, 0)$ *for some* $t \in \mathbb{R}$ *and* $\lambda > 0$.

Remark 5. Note that t is a solution of $\mathbf{E}\,\mathrm{D}'(U)^2 = 2^{n-k}\left(\mathbf{E}\,\mathrm{D}'(U)\right)^2$ where $\mathrm{D}'(x) = \max(\mathrm{D}(x) - t, 0)$ and $\lambda = 2^n\,\mathbf{E}\,\mathrm{D}'(U)$. It follows that $\mathbf{E}\,\mathrm{D}'(Y^*) = 2^{n-k}\,\mathbf{E}\,\mathrm{D}'(U) = \mathbf{E}\,\mathrm{D}'(U) + \sqrt{\mathrm{Var}\mathrm{D}'(U)} \cdot \sqrt{2^{n-k}-1}$.

4 Applications

4.1 Computational Dense Model Theorem

We say that a distribution A is γ-dense in B if we have $\Pr[A = x] \leqslant \Pr[B = x]/\gamma$. The Dense Model Theorem is the statement of the following form: if X is (s, ϵ)-indistinguishable from the uniform distribution R and X' is γ-dense in X, then there exists a distribution R' which is γ-dense in R and is (s', ϵ')-indistinguishable from X', where s' and ϵ' depends as explicit functions on s and ϵ. In this sense, R is a dense "model" for X'. The dense model theorem was proved first by Tao and Ziegler [TZ08]. It's efficient versions[5] have found important applications in complexity theory and cryptography [RTTV08, DP08, FOR12], see also [TTV09]. Below we recall a version with improved parameters, stated in language of pseudoentropy and called the "leakage lemma":

Theorem 4 (Leakage Lemma [DP08, FOR12]). *Let X be an n-bit random variable such that $\mathbf{H}_\infty^{\mathrm{HILL},(s,\epsilon)}(X) \geqslant k$ and let Z be correlated with X. Then we have $\mathbf{H}_\infty^{\mathrm{HILL},(s',\epsilon')}(X|_{Z=z}) \geqslant k'$ where $k' = k - \log(1/\Pr[Z = z])$, $s' = \mathcal{O}\left(s \cdot \delta^2/n\right)$ and $\epsilon' = \epsilon/\Pr[Z = z] + \delta$, for any $\delta \in (0, 1)$.*

The lemma states that the amount of pseudoentropy due to leakage of t bits of information decreases roughly by t, hence its name. The original proof was simplified by the use of metric entropy [FOR12]. We show how it can be simplified even further: just few lines using the basic facts about metric entropy!

Proof. If we can prove that

$$\mathbf{H}_\infty^{\mathrm{M},\mathrm{det}\{0,1\},(s,\epsilon/\Pr[Z=z]))}(X|_{Z=z}) \geqslant \mathbf{H}_\infty^{\mathrm{M},\mathrm{det}\{0,1\},(s,\epsilon)}(X) - \log(1/\Pr[Z = z])$$

then the result will follow by Lemma 1 and Remark 2. Note that by Theorem 1 for any X we have $\mathbf{H}_\infty^{\mathrm{M},\mathrm{det}\{0,1\},(s,\epsilon)}(X) \geqslant k$ if and only if $\mathbf{E}\,\mathrm{D}(X) \leqslant \frac{|\mathrm{D}|}{2^k} + \epsilon$ for all boolean D of size at most s. From this we get

$$\mathbf{E}\,\mathrm{D}(X|_{Z=z}) \leqslant \mathbf{E}\,\mathrm{D}(X)/\Pr[Z = z] \leqslant |\mathrm{D}|/2^k \Pr[Z = z] + \epsilon/\Pr[Z = z]$$

for any D. Since the characterization is also sufficient, the results follows. □

4.2 Equivalence of HILL Entropy and Unpredictability Entropy for short strings

Unpredictability entropy. The notion of unpredictability entropy is based on the (assumed) hardness of guessing X given auxiliary information Z. More formally, we have $\mathbf{H}^{\mathrm{Unp},s}(X|Z) \geqslant k$ if and only if no adversary of size at most s can predict X given Z better than with probability 2^{-k}. For Z independent of X or of the relatively short length, this reduces to the min-entropy of X[6].

[5] With the loss at most $\mathrm{poly}(1/\delta)$ in s and ϵ. In the original proof the loss is $\exp(1/\delta)$

[6] Provided that $s > 2^m n$ so that the adversary can hardcore his best guess.

Seperation from HILL entropy. If f is a one-way function, U is the uniform distribution and $X = U, Z = f(U)$ then we see that $X|Z$ has large amount of unpredictability. It is also easy to see that $X|Z$ has almost no HILL entropy.

Equivalence for short strings. On the positive side, using metric entropy and the characterization in Theorem 2, we reprove the following result of Vadhan and Zheng who established the equivalence when X is short[7]

Theorem 5 ([VZ12]). *Suppose that X and Z are, respectively, n and m-bit random variables. Then $\mathbf{H}_{\infty}^{\mathrm{HILL},(s',\epsilon)}(X|Z) \gtrsim \mathbf{H}^{\mathrm{Unp},s}(X|Z)$ with $s' = \frac{s}{\mathrm{poly}(2^n,1/\epsilon)}$.*

The original proof is based on a result similar to Theorem 2 proved in a much more complicated way. We note that this part is a trivial consequence of KKT optimality conditions and also simplify the rest of the proof.

Proof (Sketch). We prove that $\mathbf{H}_{\infty}^{\mathrm{M},\mathrm{det}[0,1],(s',\epsilon)}(X|Z) < k$ implies $\mathbf{H}^{\mathrm{Unp},s}(X|Z) < k$. Suppose not, then we have $\mathbf{E}\,\mathrm{D}(X,Z) - \mathbf{E}\,\mathrm{D}(Y,Z) \geqslant \epsilon$ for all Y such that $\widetilde{\mathbf{H}}_{\infty}(X|Z) \geqslant k$. Let Y^* be the distribution which minimizes this expression, that is which maximizes $\mathbf{E}\,\mathrm{D}(Y,Z)$. Let $t(z)$ be as in Theorem 2 and denote $\mathrm{D}'(x,z) = \max(\mathrm{D}(x,z) - t(z), 0)$ and let $\lambda = \sum_x \mathrm{D}'(x,z)$ (according to Theorem 2 this sum does not depend on z). Consider the following predictor A:

$$\text{On input } z \text{ sample } x \text{ according to the probability } \Pr[A(z) = x] = \mathrm{D}'(x,z)/\lambda$$

Note that $Y^*|_{Z=z}$ is uniform over the set $\{x : \mathrm{D}'(x,z) > 0\}$. By Theorem 2 (the sufficiency part) it follows that Y^* is also maximal for D. For every z we have $\mathbf{E}\,\mathrm{D}'(Y^*|_{Z=z},z) = \mathbf{E}\,\mathrm{D}(Y^*|Z=z,z) - t(z)$. We have also $\mathbf{E}\,\mathrm{D}'(X|_{Z=z},z) \geqslant \mathbf{E}\,\mathrm{D}(X|_{Z=z},z) - t(z)$ by the definition of D'. This proves

$$\mathbf{E}\,\mathrm{D}'(X,Z) - \mathbf{E}\,\mathrm{D}'(Y,Z) \geqslant \epsilon \text{ for all } Y \text{ such that } \widetilde{\mathbf{H}}_{\infty}(X|Z) \geqslant k.$$

It is easy to observe that

$$\Pr_{z\leftarrow Z}[A(Z) = X] = \frac{\mathbf{E}\,\mathrm{D}'(X,Z)}{\lambda} > \mathop{\mathbf{E}}_{z\leftarrow}\left[\frac{\mathbf{E}\,\mathrm{D}'(Y|_{Z=z},z)}{\sum_x \mathrm{D}'(x,z)}\right] \geqslant \mathop{\mathbf{E}}_{z\leftarrow Z} 2^{-\mathbf{H}_{\infty}(Y^*|_{Z=z})}$$

which is at least 2^{-k}. The circuit $\mathrm{D}'(x,z)$ is of complexity $2^m \cdot \mathrm{size}(\mathrm{D})$, which is too big. However, if the domain of x is small, we can approximate the numbers $t(z)$ given λ from relations in Theorem 2 (and even λ, from the second relation, for the uniform setting). Indeed, knowing that $\mathbf{E}\max(\mathrm{D}(U,z) - t(z)) = \lambda$, we estimate $\mathbf{E}\max(\mathrm{D}(U,z) - t)$ for fixed t and then find a "right" value $t = t(z)$ by the binary search. This way for every z we can approximate $\mathrm{D}'(\cdot,z)$, and hence the distribution $\Pr[A(z) = x]$, up to a maximal error $\delta \ll 2^{-k}$ and with overwhelming probability $1 - \exp(-\mathrm{poly}(1/\delta))$, using $\mathrm{poly}(1/\delta)$ samples of D. On average over z we predict X with probability $2^{-k} - \delta \approx 2^{-k}$. □

[7] Logarithmically in the security parameter

4.3 Improved Leftover Hash Lemma for square-secure applications

In the key derivation problem we want to derive a secure m-bit key for some application P from an *imperfect* source of randomness X. The generic approach is to use a randomness extractor. However, as implied by the RT-bounds [RTS00], the min-entropy in X needs to be at least $m + 2\log(1/\epsilon)$ if we want the derived key to be ϵ-secure. Fortunately, as shown by Barak et. al [BDK+11], for many cryptographic applications, one can reduce this loss by half, that is to $L = \log(1/\epsilon)$. To this end, they introduce the class of *square-secure* applications, where the squared advantage, over the uniform choice of keys, of every bounded attacker is small[8]. This class contains for example all unpredictability applications, stateless chosen plaintext attack secure encryption and weak pseudo-random functions. The reduction of entropy loss follows by combining universal hashing with the following lemma

Lemma 4 ([BDK+11]). *For a function* $\mathrm{D} : \{0,1\}^\ell \to [-1,1]$ *and* $X \in \{0,1\}^\ell$ *of collision entropy* k *we have*

$$\mathbf{E}\,\mathrm{D}(X) \leqslant \mathbf{E}\,\mathrm{D}(U_\ell) + \sqrt{\mathrm{Var}\mathrm{D}(U_\ell)} \cdot \sqrt{2^{\ell-k} - 1}.$$

To see this, let $\mathrm{Win}_{\mathsf{A}}(r,h)$, for arbitrary attacker $\mathsf{A} \in \mathcal{A}$, be the probability that A breaks the key r given in addition[9] h and let $\mathrm{D}_{\mathcal{A}}(r,h) = \mathrm{Win}_{\mathsf{A}}(r,h) - \frac{1}{2}$ be its advantage. Let X be any n-bit random variable of min-entropy $m+\log(1/\epsilon)$. We apply a randomly chosen universal hash function[10] H from n to m bits. It is easy to see that $H(X), H$ is a distribution with collision entropy $m+\log|\mathcal{H}|-\log(1+\epsilon)$. From the lemma it follows now that

$$\mathbf{E}\,\mathrm{D}_{\mathsf{A}}(H(X),H) \leqslant \mathbf{E}\,\mathrm{D}_{\mathsf{A}}(U,H) + \sqrt{\mathrm{Var}\mathrm{D}_{\mathsf{A}}(U,H)} \cdot \sqrt{\epsilon}$$

If we assume that $\max_h \mathbf{E}\,\mathrm{D}_{\mathsf{A}}(U,h) \leqslant \epsilon$ (which means ϵ-security against $\mathcal{A}$ with the uniform key) and that $\max_h \mathbf{E}\,\mathrm{D}_{\mathsf{A}}(U,h)^2 \leqslant \sigma$ with $\sigma = \mathcal{O}(\epsilon)$ (which means σ-square-security against $\mathcal{A}$ with the uniform key) then we achieve $\mathcal{O}(\epsilon)$ security for the *extracted* key, with entropy loss only $\log(1/\epsilon)$.

An alternative proof. We show that Theorem 3 implies Lemma 4. Indeed, set $k = \ell$ and $\epsilon = 0$ in Theorem 3. Let Y^* be the distribution of collision entropy at least $k = \ell$ which maximizes $\mathbf{E}\,\mathrm{D}(Y)$, and let t, λ and D' be as in the characterization. Denote $S = \{x : \mathrm{D}(x) \geqslant t\}$ and let $\mathrm{D}|_S$ be the restriction of D to the set S. Note that $Y^*|S \stackrel{d}{=} Y^*$ maximizes $\mathrm{D}|_S$ and $\mathrm{D}|_S(x) = \mathrm{D}'|_S(X) + t$ for every $x \in S$. By Remark 5 we get

$$\mathbf{E}\,\mathrm{D}(X) \leqslant \mathbf{E}\,\mathrm{D}(Y^*) = \mathbf{E}\,\mathrm{D}|_S(Y^*|S) = \mathbf{E}\,\mathrm{D}|_S(U_S) + \sqrt{\mathrm{Var}\mathrm{D}_S(U_S)} \cdot \sqrt{|S|2^{-k} - 1}.$$

[8] Which essentially means that the probability that an attacker break the key is concentrated over keys

[9] For the uniformly chosen key this doesn't help the adversary, at least in the nonuniform model

[10] A family $\mathcal{H}$ functions from n to m bits is universal if $\Pr_{h \leftarrow \mathcal{H}}[h(x) = h(x')] = 2^{-m}$ for $x \neq x'$

We show that one can replace S by the $\{0,1\}^\ell$ on the right hand side. This will follow by the following general lemma

Lemma 5. *Let X be a random variable, $c > 1$ be a constant and S be an event of probability $\mathbf{P}(S) > c^{-1}$. Then*

$$\mathbf{E}[X|S] + \sqrt{\mathrm{Var}[X|S]} \cdot \sqrt{c\mathbf{P}(S) - 1} \leqslant \mathbf{E}[X] + \sqrt{\mathrm{Var}[X]} \cdot \sqrt{c-1} \qquad (5)$$

The proof follows by a few algebraic manipulations and is omitted.

4.4 Some further applications

Lower bounds on square security. Using the characterization from Theorem 3 one can derive some non-trivial lower bounds on square-security needed for key derivation. We discuss this problem in a separate paper.

References

BDK+11. Boaz Barak, Yevgeniy Dodis, Hugo Krawczyk, Olivier Pereira, Krzysztof Pietrzak, François-Xavier Standaert, and Yu Yu, *Leftover hash lemma, revisited*, CRYPTO'11, Springer-Verlag, 2011, pp. 1–20.

BSW03. Boaz Barak, Ronen Shaltiel, and Avi Wigderson, *Computational analogues of entropy.*, RANDOM-APPROX, vol. 2764, Springer, 2003, pp. 200–215.

BV04. Stephen Boyd and Lieven Vandenberghe, *Convex optimization*, Cambridge University Press, New York, NY, USA, 2004.

CKLR11. Kai-Min Chung, Yael Tauman Kalai, Feng-Hao Liu, and Ran Raz, *Memory delegation*, Proceedings of the 31st Annual Conference on Advances in Cryptology, CRYPTO'11, Springer-Verlag, 2011, pp. 151–165.

DORS08. Yevgeniy Dodis, Rafail Ostrovsky, Leonid Reyzin, and Adam Smith, *Fuzzy extractors: How to generate strong keys from biometrics and other noisy data*, SIAM J. Comput. **38** (2008), no. 1, 97–139.

DP08. Stefan Dziembowski and Krzysztof Pietrzak, *Leakage-resilient cryptography*, FOCS '08, IEEE Computer Society, 2008, pp. 293–302.

FOR12. Benjamin Fuller, Adam O'Neill, and Leonid Reyzin, *A unified approach to deterministic encryption: New constructions and a connection to computational entropy*, TCC'12, Springer-Verlag, 2012, pp. 582–599.

GW11. Craig Gentry and Daniel Wichs, *Separating succinct non-interactive arguments from all falsifiable assumptions*, STOC '11, ACM, 2011, pp. 99–108.

HILL99. Johan Hastad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby, *A pseudorandom generator from any one-way function*, SIAM J. Comput. **28** (1999), no. 4, 1364–1396.

HLR07. Chun-Yuan Hsiao, Chi-Jen Lu, and Leonid Reyzin, *Conditional computational entropy, or toward separating pseudoentropy from compressibility*, EUROCRYPT '07, Springer-Verlag, 2007, pp. 169–186.

KPW13. S. Krenn, K. Pietrzak, and A. Wadia, *A counterexample to the chain rule for conditional hill entropy*, TCC'13, Springer-Verlag, 2013, pp. 23–39.

Rey11. Leonid Reyzin, *Some notions of entropy for cryptography*, ICITS'11, Springer-Verlag, 2011, pp. 138–142.

RTS00. Jaikumar Radhakrishnan and Amnon Ta-Shma, *Bounds for dispersers, extractors, and depth-two superconcentrators*, SIAM JOURNAL ON DISCRETE MATHEMATICS **13** (2000), 2000.

RTTV08. Omer Reingold, Luca Trevisan, Madhur Tulsiani, and Salil P. Vadhan, *Dense subsets of pseudorandom sets*, 49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA, 2008, pp. 76–85.

SGP15. Maciej Skorski, Alexander Golovnev, and Krzysztof Pietrzak, *Condensed unpredictability*, Automata, Languages, and Programming - 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part I, 2015, pp. 1046–1057.

Sha48. C. E. Shannon, *A mathematical theory of communication*, Bell system technical journal **27** (1948).

TTV09. Luca Trevisan, Madhur Tulsiani, and Salil Vadhan, *Regularity, boosting, and efficiently simulating every high-entropy distribution*, CCC '09, IEEE Computer Society, 2009, pp. 126–136.

TZ08. Terence Tao and Tamar Ziegler, *The primes contain arbitrarily long polynomial progressions*, Acta Mathematica **201** (2008), no. 2, 213–305 (English).

VZ12. S. Vadhan and C. J. Zheng, *Characterizing pseudoentropy and simplifying pseudorandom generator constructions*, STOC '12, ACM, 2012, pp. 817–836.

Zhe14. Jia Zheng, *A uniform min-max theorem and characterizations of computational randomness*, Ph.D. thesis, University of Harvard, 2014.

A Proof of Theorem 1

Consider the maximization program in Lemma 2 for a given D and under the min-entropy entropy. We identify functions and measures over n-bit strings with vectors in $\mathbb{R}^d$ where $d = 2^n$.

For the min-entropy case the objective is as in Equation (2) and the constraints are $f(p) = \max_i p_i \leqslant 2^{-k}$, $\sum_i p_i = 1$ where p represents a probability vector. Since p is non-differentiable but convex, we should use subgradients in place of gradients in Lemma 3. But instead of using the convex optimization machinery we give an elementary argument. It is well-known that a min-entropy distribution is a convex combination of flat sources, and since the expectation is linear we conclude that it is maximized on a flat source. Now clearly it makes sense to associate this flat source with biggest values of D. Note that for non-integer values of k this is till true except that "flat" source is actually uniform on $\lfloor 2^k \rfloor$ values and has one extra point with the "remainder" positive mass of $2^{-k}\lfloor 2^k \rfloor$.

This proves that the expectation of a given distinguisher D, under the min entropy constraint, is maximized by a distribution Y^* of the form as in Theorem 1. The characterization claimed in the theorem follows now by Lemma 2.

B Proof of Theorem 2

We proceed as before, considering the maximization program in Lemma 2 but under the *average conditional min-entropy* constraints. The characterization of

the maximizing distribution Y^* appears already, along with a detailed proof, in the paper [SGP15] (Lemma 2 in the appendix[11]). It also essentially reproves a result implict in [Zhe14]. By this characterization (also based on KKT multipliers) the optimal $Y^*|Z$ is such that for thresholds $t(z)$ as in Theorem 2 we put $Y^*|Z = z$ to have zero mass when $\mathrm{D}(x,z) < t(z)$, make it uniform when $\mathrm{D}(x,z) > t(z)$ and put the smaller (pointwise smaller than $\max_x \Pr[Y^* = x]$) mass arbitrarily over the set $\mathrm{D}(x,z) = t(z)$. To guarantee we can have a distributon of conditional entropy k we need to satisfy

$$\mathop{\mathbf{E}}_{z \leftarrow Z} \left[1/\#\left\{x : \mathrm{D}(x,z) \geqslant t(z)\right\}\right] \leqslant 2^{-k} \leqslant \mathbf{E}\left[1/\#\left\{x : \mathrm{D}(x,z) > t(z)\right\}\right].$$

where the right-hand side reflects the freedom in choosing the distribution on $\{x : \mathrm{D}(x,z) = t(z)\}$.

This proves that the expectation of a given distinguisher D, under the conditional min-entropy constraint, is maximized by a distribution Y^* of the form as in Theorem 2. The characterization claimed in the theorem follows now by Lemma 2.

C Proof of Theorem 3

Consider the maximization program in Lemma 2 for a given D and under the collision entropy. We identify functions and measures over n-bit strings with vectors in $\mathbb{R}^d$ where $d = 2^n$. For collision entropy the constraint f reads as $f(p) = \sum_i p_i^2$, and $a = 2^{-k}$. The gradient of f at p equals $2p$, and thus by Lemma 3 the point p is optimal if and only if for all coordinates i we have

$$\mathrm{D}_i = 2\lambda_1 \cdot p_i - \lambda_{3i} + \lambda_2$$

where λ_3 is non-negative and $\lambda_{3i} p_i = 0$. In case there are many maximas p we can choose such that $f(p) = a$, that is the entropy is exactly k. Now we argue that we can assume $\lambda_1 \neq 0$. Indeed, otherwise we have $\mathrm{D}_i = \lambda_2$ whenever $p_i > 0$, so that D is constant on the support of p. The support S of p has at least 2^k elements because Jensen's inequality $|S|^{-1} = |S|^{-1} \sum_{i \in S} p_i^2 \geqslant \left(|S|^{-1} \sum_{i \in S} p_i\right)^2$ implies $|S| \geqslant 2^k$ (this is true for each concave entropy notion). If this is the case we can take p' which is uniform over 2^k elements and it is also the optimum (the objective is still the value of D). We then have $\mathrm{D}_i = 2\lambda_1 \cdot p'_i + \lambda_2$ for i in the support of p' and some $\lambda_1 > 0$,$\lambda_2 \in \mathbb{R}$ (we can take $\lambda_1 = 1$ and choose λ_2 accordingly). For the remaining values of i (outside the support of p') the value of D_i cannot be bigger (as otherwise p' and p wouldn't be optimal) so that finally we can write $\mathrm{D}_i = 2\lambda_1 \cdot p'_i + \lambda_2 - \lambda_{3i}$ where λ_{3i} equals zero when $p'_i > 0$ and is non-negative otherwise. We conclude that we can assume $\lambda_1 \neq 0$.

We now rewrite the first-order condition as

$$2\lambda_1 \cdot p_i = \mathrm{D}_i - \lambda_2 + \lambda_{3i}$$

[11] The paper is also a chapter of this thesis.

If $p_i \neq 0$ then we have $\lambda_{3i} = 0$ so that $2\lambda_1 \cdot p_i = \mathrm{D}_i - \lambda_2$. If $p_i = 0$ then $0 = 2\lambda_1 \cdot p_i = \mathrm{D}_i - \lambda_2 + \lambda_{3i} \geqslant \mathrm{D}_i - \lambda_2$. Summing up, we can write

$$2\lambda_1 \cdot p_i = \max\left(\mathrm{D}_i - \lambda_2, 0\right)$$

This proves that the expectation of a given distinguisher D, under the colision entropy constraint, is maximized by a distribution Y^* of the form as in Theorem 3 (use p_i as Y^*, set $\lambda := 2\lambda_1 > 0$ and $t = \lambda_2$). The characterization claimed in the theorem follows now by Lemma 2.