Electronic Colloquium on Computational Complexity, Report No. 35 (2023)

A Duality Between One-Way Functions and Average-Case Symmetry of Information

Shuichi Hirahara* Rahul Ilango† Zhenjian Lu‡ Mikito Nanashima§

Igor C. Oliveira¶

March 22, 2023

Abstract

Symmetry of Information (SoI) is a fundamental property of Kolmogorov complexity that relates the complexity of a pair of strings and their conditional complexities. Understanding if this property holds in the *time-bounded* setting is a longstanding open problem. In the nineties, Longpré and Mocas [LM93] and Longpré and Watanabe [LW95] established that if SoI holds for time-bounded Kolmogorov complexity then cryptographic one-way functions do not exist, and asked if a converse holds.

We show that one-way functions exist *if and only if* (probabilistic) time-bounded SoI fails on average, i.e., if there is a samplable distribution of pairs (x, y) of strings such that SoI for pK^t complexity fails for many of these pairs. Our techniques rely on recent perspectives offered by probabilistic Kolmogorov complexity and meta-complexity, and reveal further equivalences between inverting one-way functions and the validity of key properties of Kolmogorov complexity in the time-bounded setting: (average-case) language compression and (average-case) conditional coding.

Motivated by these results, we investigate correspondences of this form for the worst-case hardness of NP (i.e., $\mathsf{NP} \not\subseteq \mathsf{BPP}$) and for the average-case hardness of NP (i.e., $\mathsf{DistNP} \not\subseteq \mathsf{HeurBPP}$), respectively. Our results establish the existence of similar *dualities* between these computational assumptions and the failure of results from Kolmogorov complexity in the time-bounded setting. In particular, these characterizations offer a novel way to investigate the main hardness conjectures of complexity theory (and the relationships among them) through the lens of Kolmogorov complexity and its properties.

*National Institute of Informatics, Japan. Email: s_hirahara@nii.ac.jp
†Massachusetts Institute of Technology, US. Email: rilango@mit.edu
‡University of Oxford, UK. Email: zhenjian.lu@cs.ox.ac.uk
§Tokyo Institute of Technology, Japan. Email: nanashima.m.aa@is.c.titech.ac.jp
¶University of Warwick, UK. Email: igor.oliveira@warwick.ac.uk

ISSN 1433-8092

Contents

1 Introduction

1.1 Context and Motivation

A basic and fundamental property in Shannon's information theory is *Symmetry of Information* (SoI). Informally, SoI says that for any two random variables X and Y the amount of information that X reveals about Y is the same as the amount of information that Y reveals about X. Formally, it says that

$$I(X;Y) \;=\; H(Y) - H(Y \mid X) \;=\; H(X) - H(X \mid Y)\;,$$

where H denotes the entropy function. Equivalently, symmetry of information is often written as:

$$H(X,Y) \;=\; H(Y) + H(X \mid Y) \;=\; H(X) + H(Y \mid X)\;,$$

where $H(X,Y)$ denotes the entropy of the jointly distributed random variable (X,Y).

Symmetry of information is also a basic and fundamental property of Kolmogorov complexity, which can be viewed as an algorithmic analogue of information theory. The Kolmogorov complexity $\mathsf{K}(x)$ of a string x is the length of the smallest program p that outputs x (when p is fed to an a priori fixed universal Turing machine). The conditional Kolmogorov complexity of x given y, written $\mathsf{K}(x \mid y)$, is defined similarly, with the difference that y is provided as input to the universal machine. Zvonkin and Levin [ZL70] (actually, [ZL70] credits Levin and Kolmogorov independently) show that the SoI property from information theory also holds for Kolmogorov complexity, with an additive logarithmic loss. Formally, for any strings x and y,

$$\mathsf{K}(x,y) \;\approx\; \mathsf{K}(y) + \mathsf{K}(x \mid y) \;\approx\; \mathsf{K}(x) + \mathsf{K}(y \mid x)\;,$$

up to an additive factor of order $\pm O(\log(|x|+|y|))$ in each equation. Written this way, symmetry of information roughly says that: (i) to describe both x and y it suffices to first describe y optimally without considering x and then describe x optimally assuming access to a description of y; and (ii) there is no significantly better way to describe a pair of strings x, y. Note that (i) is easily seen to hold, while (ii) is non-trivial and states that

$$\mathsf{K}(x,y) \;\geq\; \mathsf{K}(x \mid y) + \mathsf{K}(y) - O(\log(|x|+|y|))\;.$$

Symmetry of information has found numerous applications in a variety of areas (see the textbooks [SUV17, LV19] for a comprehensive introduction) and is widely regarded as one of the main pillars of the theory of Kolmogorov complexity (see, e.g., [Lee06]).

Time-Bounded Kolmogorov Complexity. A disadvantage of Kolmogorov complexity is that it does not take into account the complexity of generating the string x. This issue is particularly significant in applications to algorithms and complexity theory, where the running time is a crucial parameter. Remarkably, in his seminal paper [Kol65, Section 4], Kolmogorov also proposed the study of *t-time-bounded Kolmogorov complexity*, denoted by $\mathsf{K}^t(x)$, which is the shortest size of a program that prints x in time at most t. Similarly to Kolmogorov complexity, the theory of time-bounded Kolmogorov complexity has been widely investigated and has led to several influential results and applications (see, e.g., [Sip83, Ko91, ABK+06, AF09, Hir18, OPS19, LP20, Hir21]).

Motivated by the prominent role of symmetry of information in Kolmogorov complexity, the existence of a *time-bounded* symmetry of information principle has been considered since the early

years of algorithmic information theory. According to Levin (see [LR05]), already in the sixties Kolmogorov suggested time-bounded versions of symmetry of information as an interesting research question [Kol68]. Unfortunately, the classical proof that SoI holds for (time-unbounded) Kolmogorov complexity requires an exhaustive search, and as such, the same argument is not applicable in the time-bounded setting.

One-Way Functions and Time-Bounded Symmetry of Information. In the nineties, Longpré and Mocas [LM93] and Longpré and Watanabe [LW95] established a connection between time-bounded SoI and the existence of cryptographic one-way functions. More precisely, they proved that if SoI holds for time-bounded Kolmogorov complexity then one-way functions do not exist. Since one-way functions are both necessary and sufficient for the existence of a variety of fundamental cryptographic primitives, such as private-key encryption [GM84], pseudorandom generators [HILL99], digital signatures [Rom90], and commitment schemes [Nao91], their result further highlights the significance of understanding the validity of SoI in the time-bounded setting.

Longpré and Mocas [LM93] asked if a converse result holds, i.e., if time-bounded symmetry of information characterizes the non-existence of one-way functions. Similarly, Longpré and Watanabe [LW95] mentioned that their ultimate goal would be to prove some if and only if statement regarding symmetry of information. In the same paper, they showed that time-bounded SoI holds if $\mathsf{P} = \mathsf{NP}$, which is stronger than the assumption that one-way functions do not exist. Recent papers (Hirahara [Hir22b], Goldberg and Kabanets [GK22], and Goldberg, Kabanets, Lu, and Oliveira [GKLO22]) improved this by deriving time-bounded SoI from weaker assumptions on average-case complexity of NP. However, establishing a *characterization* of the existence of one-way functions (or of any other computational assumption) through SoI has remained elusive.

1.2 Results

We confirm the existence of a tight relationship between symmetry of information and cryptography, by establishing the first *characterization* of one-way functions using SoI. More precisely, our results show that one-way functions exist if and only if time-bounded SoI fails on average, i.e., if there is a samplable distribution of pairs (x, y) of strings such that time-bounded SoI fails for many of these pairs.

In order to state unconditional characterizations, we work in the setting of probabilistic Kolmogorov complexity, i.e., our results are stated for the measures rK^t and pK^t. These measures naturally extend the theory of time-bounded Kolmogorov complexity to the realm of probabilistic algorithms. Intuitively, this is a more suitable perspective in our context, given that the security of a one-way function refers to probabilistic polynomial-time adversaries. Nevertheless, under standard derandomization assumptions our results can also be stated for the classical notion of K^t complexity employed in early papers in the area.

Before formally stating our results, we briefly review the necessary notions from probabilistic Kolmogorov complexity. We discuss additional related work in Section 1.5.

Probabilistic Kolmogorov Complexity. Thanks to the ubiquitous role of randomness in algorithms and complexity, probabilistic Kolmogorov complexity has found a number of applications in recent years (e.g., [Oli19, LO21, LOS21, GKLO22, LOZ22, Hir22b]). As alluded to above, we consider two notions that extend (deterministic) time-bounded Kolmogorov complexity K^t to the setting of randomized algorithms: rK^t complexity and pK^t complexity. We briefly review these

notions below, referring to Section 2 for their formal definitions.

For a string x, we let $\mathsf{rK}^t(x)$ denote the shortest size of a randomized program that prints x with probability at least $2/3$ when running for at most t steps. We refer to $\mathsf{rK}^t(x)$ as the *randomized t-time bounded Kolmogorov complexity of x*. Intuitively, there is a short and efficient randomized program that prints x with high probability. The code of this program serves as a description of x.

On the other hand, we let $\mathsf{pK}^t(x)$ denote the smallest integer k such that, with probability at least $2/3$ over the choice of a random string $w \sim \{0,1\}^t$, there is a (deterministic) program that when given w prints x within at most t steps. In other words, for a typical random string w, the string x has a t-time bounded description of length at most k given w. We refer to $\mathsf{pK}^t(x)$ as the *probabilistic t-time bounded Kolmogorov complexity of x*. Informally, this notion can be interpreted as K^t in the presence of a random string shared by all parties involved in a computation.

Under a standard derandomization assumption, Goldberg, Kabanets, Lu, and Oliveira [GKLO22] proved that, for every string x, $\mathsf{K}^t(x)$, $\mathsf{rK}^t(x)$, and $\mathsf{pK}^t(x)$ are the same up to an additive factor of $O(\log |x|)$ and at most a polynomial overhead in the running time t. (Roughly speaking, this is similar in nature to the conjectured collapse $\mathsf{NP} = \mathsf{MA} = \mathsf{AM}$.) For this reason, the results presented next also hold for K^t complexity, under a standard derandomization assumption. However, as in previous works, probabilistic Kolmogorov complexity allows us to obtain unconditional statements. We refer to the survey [LO22] for more information about rK^t and pK^t and their applications in algorithms and complexity theory.

1.2.1 One-Way Functions, Symmetry of Information, and Kolmogorov Complexity

As alluded to above, our main results establish a *duality* between the (non-)existence of one-way functions and the validity of the symmetry of information principle in the time-bounded setting for most pairs of strings. More generally, we establish that the preservation in the time-bounded setting of average-case versions of other key principles from Kolmogorov complexity is completely captured by one-way functions.

Theorem 1 (Duality Between One-way Functions and Properties of pK)**.** *The following are equivalent.*

1. *Infinitely-often one-way functions do not exist.*
2. **(Average-Case Symmetry of Information)** *For every polynomial-time samplable distribution family $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and for every polynomial q, there exists a polynomial p such that for every computable time bound $t\colon \mathbb{N} \to \mathbb{N}$ with $t(n) \geq p(n)$ and for all large enough n,*

$$\Pr_{(x,y)\sim\mathcal{D}_n}\left[\mathsf{pK}^{t(n)}(x,y) \geq \mathsf{pK}^{t(n)}(x \mid y) + \mathsf{pK}^{t(n)}(y) - \log t(n)\right] \geq 1 - \frac{1}{q(n)}.$$

3. **(Average-Case Conditional Coding)** *For every polynomial-time samplable distribution family $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and for every polynomial q, there exists a polynomial p such that for all large enough n,*

$$\Pr_{(x,y)\sim\mathcal{D}_n}\left[\mathsf{pK}^{p(n)}(x \mid y) \leq \log \frac{1}{\mathcal{D}_n(x \mid y)} + \log p(n)\right] \geq 1 - \frac{1}{q(n)}.$$

4. **(Average-Case Language Compression)** *For every recursively enumerable set* $L \subseteq \{\{0,1\}^n \times \{0,1\}^n\}_{n\in\mathbb{N}}$*, every polynomial-time samplable distribution family* $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$*, where each* $\mathcal{D}_n$ *is over* $\{0,1\}^n \times \{0,1\}^n$*, and every polynomial* q*, there exists a polynomial* p *such that for all large enough* n*,*

$$\Pr_{(x,y)\sim\mathcal{D}_n}\left[x \in L_y \implies \mathsf{pK}^{p(n)}(x \mid y) \leq \log|L_y| + \log p(n)\right] \geq 1 - \frac{1}{q(n)},$$

where $L_y = \{x \in \{0,1\}^n \mid (x,y) \in L\}$.

Observe that a coding theorem for pK^t with optimal parameters is known to hold unconditionally [LOZ22]. In contrast, Theorem 1 Item 3 considers *conditional* coding, where we relate conditional pK^t complexity and conditional probability. This exhibits a contrast between (time-unbounded) Kolmogorov complexity, where coding and conditional coding can be established, and time-bounded Kolmogorov complexity, where coding holds but conditional coding does not hold under a cryptographic assumption.

Note that in Theorem 1 (Items 2-4) we stated high probability versions of each property of Kolmogorov complexity. As a consequence of our proof, we can show that the low probability (i.e., non-negligible) and high probability versions of these statements are all equivalent (see Section 3). The result is also robust with respect to almost-everywhere versus infinitely-often statements, i.e., it is possible to prove that one-way functions do not exist if and only if the average-case Kolmogorov complexity properties hold infinitely often (see Appendix A).

An interesting aspect of the average-case setting is that we can employ a statement of symmetry of information where the same time bound $t(n)$ appears on both sides of the inequality (Theorem 1 Item 2). In recent papers that establish worst-case SoI under an easiness assumption (e.g., [Hir22b, GK22, GKLO22]), there is a loss of parameters and the inequalities are of the form $\mathsf{pK}^{t(n)}(x,y) \geq \mathsf{pK}^{p(t(n))}(x \mid y) + \mathsf{pK}^{p(t(n))}(y) - \log p(t(n))$, for some polynomial $p(\cdot)$.

Relevance to the Foundations of Cryptography. Our result reveals a deep relationship between the existence of secure cryptography and the failure of the symmetry of information principle for efficient computations. We discuss this in more detail now.

Consider the output of a polynomial-time computable function $y = f(x)$, and assume for simplicity that f is an injective function. Since given x we can efficiently recover y, x contains all the necessary information about y. On the other hand, intuitively, we can break a candidate one-way function f if $y = f(x)$ contains sufficient information for us to efficiently recover x from it. Longpré and Watanabe [LW95] made this intuition formal in the context of an arbitrary polynomial-time computable function f, i.e., they proved that if SoI holds in the time-bounded setting then secure one-way functions do not exist. In other words, the existence of one-way functions necessarily breaks the symmetry of information between y and x in the time-bounded setting. Indeed, this must hold for a non-negligible fraction of such pairs of strings.

Our result completes the picture by showing that a failure of symmetry of information for a non-negligible fraction of pairs (x,y) of strings produced by a samplable distribution is all we need to construct key cryptographic primitives such as one-way functions, private-key encryption, digital signatures, commitment schemes, etc. In other words, the existence of secure cryptography can be formally characterized by a break of the computational/information symmetry between the input and output of an efficiently computable function with respect to

polynomial-time computations.

In our next result, we consider two natural questions posed to us by Watanabe [Wat22]:

- Is it possible to get an unconditional equivalence between the non-existence of a one-way function and symmetry of information for rK?
- The usual notions of time-bounded Kolmogorov complexity do not refer to the complexity of producing a succinct encoding. Can we efficiently compute a short program that "witnesses" the symmetry of information inequality $\mathsf{rK}^t(x \mid y) \lesssim \mathsf{rK}^t(x,y) - \mathsf{rK}^t(y)$?

We are able to answer these questions in the quasi-polynomial-time regime. In the statement below, a quasi-polynomial is a function of the form $\exp(\log^c n)$, for some constant $c \in \mathbb{N}$.

Theorem 2 (Duality Between One-way Functions and Properties of rK)**.** *The following are equivalent.*

1. *Infinitely-often polynomial-time-computable one-way functions secure against quasi-polynomial-time randomized algorithms do not exist.*
2. **(Average-Case Symmetry of Information)** *For every polynomial-time samplable distribution family $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and for every quasi-polynomial q, there exists a quasi-polynomial p such that for every computable time bound $t\colon \mathbb{N} \to \mathbb{N}$ with $t(n) \geq p(n)$ and for all large enough n,*

$$\Pr_{(x,y)\sim\mathcal{D}_n}\left[\mathsf{rK}^{t(n)}(x,y) \geq \mathsf{rK}^{t(n)}(x \mid y) + \mathsf{rK}^{t(n)}(y) - \log t(n)\right] \geq 1 - \frac{1}{q(n)}.$$

3. **(Average-Case Symmetry of Information with an Efficient Encoder)** *In addition to Item 2, there exists a quasi-polynomial-time randomized algorithm M that takes $(x,y) \sim \mathcal{D}_n$ and, with probability $\geq 1 - 1/q(n)$ over the choice of (x,y) and the internal randomness of M, outputs the description of a randomized program of length $\leq \mathsf{rK}^{t(n)}(x,y) - \mathsf{rK}^{t(n)}(y) + \log t(n)$ that takes y as input and outputs x with high probability in time $p(n)$.*
4. **(Average-Case Approximation of K by $\mathsf{rK}^{\mathsf{quasipoly}}$)** *For every polynomial-time samplable distribution family $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and for every quasi-polynomial q, there exists a quasi-polynomial p such that for all large enough n,*

$$\Pr_{(x,y)\sim\mathcal{D}_n}\left[\mathsf{rK}^{p(n)}(x \mid y) \leq \mathsf{K}(x \mid y) + \log p(n)\right] \geq 1 - \frac{1}{q(n)}.$$

5. **(Average-Case Approximation of K by $\mathsf{rK}^{\mathsf{quasipoly}}$ with an Efficient Encoder)** *In addition to Item 4, there exists a quasi-polynomial-time randomized algorithm M that takes $(x,y) \sim \mathcal{D}_n$ as input and, with probability $\geq 1 - 1/q(n)$ over the choice of (x,y) and the internal randomness of M, outputs the description of a randomized program of length $\leq \mathsf{K}(x \mid y) + \log p(n)$ that takes y as input and outputs x with high probability in time $p(n)$.*

In our opinion, Theorem 2 Item 4 is particularly striking. It shows that, under the ability to invert one-way functions, time-bounded Kolmogorov complexity and Kolmogorov complexity essentially coincide for most strings generated by a samplable distribution. In a sense, this explains why all key properties of Kolmogorov complexity survive on average if one-way functions do not exist.

An unexpected consequence of the equivalences in Theorem 2 is that if time-bounded SoI holds on average then time-bounded SoI holds on average with an efficient encoder.

For succinctness, we emphasized different aspects of Kolmogorov complexity in the equivalences appearing in Theorem 1 and Theorem 2. We stress that it is not difficult to adapt our techniques so that in both statements we obtain the same set of results (e.g., an average-case conditional coding statement in Theorem 2 or the approximation of K by pK in Theorem 1). The only fundamental difference between these characterizations is that for rK our techniques can only be applied in the quasi-polynomial regime.

1.2.2 Complexity Theory Through the Lens of Kolmogorov Complexity

Inspired by these results, we begin investigating whether other central questions in complexity theory could also be captured through structural properties of time-bounded Kolmogorov complexity. First, we consider the average-case complexity of NP. Since the assumption that NP is easy on average is stronger than the assumption that one-way functions do not exist, it is natural to suspect that a stronger form of the aforementioned average-case principles might hold in this case.

While in Theorem 1 Items 2-4 we sampled a pair (x, y) of strings from $\mathcal{D}$, our next result will consider a more general way of sampling (x, y). In more detail, we consider a distribution $\mathcal{C}$ supported over $\{0,1\}^n$ and a distribution $\mathcal{D}$ supported over $\{0,1\}^n \times \{0,1\}^n$. In order to sample a pair (x, y), we first sample $y \sim \mathcal{C}$, then sample $x \sim \mathcal{D}(\cdot \mid y)$. It turns out that such a change completely captures the difference between inverting one-way functions and solving problems in NP on average.

Theorem 3 (Duality Between DistNP vs $\mathsf{HeurBPP}$ and Kolmogorov Complexity)**.** *The following are equivalent.*

1. $\mathsf{DistNP} \subseteq \mathsf{HeurBPP}$.
2. **(Independent Average-Case Conditional Coding)** *For every polynomial-time samplable distribution families $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ and $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and each $\mathcal{C}_n$ is over the support of the second half of $\mathcal{D}_n$, and for every polynomial q, there exists a polynomial p such that for all large enough n,*
$$\mathop{\mathbf{Pr}}_{y\sim\mathcal{C}_n, x\sim\mathcal{D}_n(\cdot|y)}\left[\mathsf{pK}^{p(n)}(x \mid y) \leq \log \frac{1}{\mathcal{D}_n(x \mid y)} + \log p(n)\right] \geq 1 - \frac{1}{q(n)}.$$
3. **(Independent Average-Case Language Compression)** *For every recursively enumerable set $L \subseteq \{\{0,1\}^n \times \{0,1\}^n\}_{n\in\mathbb{N}}$, for every polynomial-time samplable distribution families $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ and $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and each $\mathcal{C}_n$ is over the support of the second half of $\mathcal{D}_n$, and for every polynomial q, there exists a polynomial p such that for all large enough n,*
$$\mathop{\mathbf{Pr}}_{y\sim\mathcal{C}_n, x\sim\mathcal{D}_n(\cdot|y)}\left[x \in L_y \implies \mathsf{pK}^{p(n)}(x \mid y) \leq \log |L_y| + \log p(n)\right] \geq 1 - \frac{1}{q(n)}.$$

We also show that the assumption that $\mathsf{DistNP} \subseteq \mathsf{HeurBPP}$ implies a stronger form of Average-Case Symmetry of Information called Independent Average-Case Symmetry of Information (see Theorem 32 in Section 5). However, this result is not yet an equivalence. We discuss this in more detail in Section 1.4.

Finally, we consider the worst-case complexity of NP, and the possibility of capturing this setting through Kolmogorov complexity.

Theorem 4 (Duality Between NP vs BPP and Kolmogorov Complexity)**.** *The following are equivalent.*

1. $\mathsf{NP} \subseteq \mathsf{BPP}$.
2. **(Worst-Case Conditional Coding)** *For every polynomial-time samplable distribution family $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, there exists a polynomial p such that for all large enough n and $(x,y) \in \mathsf{Support}(\mathcal{D}_n)$,*
$$\mathsf{pK}^{p(n)}(x \mid y) \leq \log \frac{1}{\mathcal{D}_n(x \mid y)} + \log p(n).$$
3. **(Worst-Case Language Compression)** *For every polynomial-time computable set $L \subseteq \{\{0,1\}^n \times \{0,1\}^n\}_{n\in\mathbb{N}}$, there exists a polynomial p such that for all large enough n and all $x, y \in \{0,1\}^n$,*
$$x \in L_y \implies \mathsf{pK}^{p(n)}(x \mid y) \leq \log |L_y| + \log p(n).$$

Moreover, the above equivalence continues to hold if we replace pK with rK in Item 2 and Item 3.

Similarly to Theorem 3, the role of symmetry of information in the setting of Theorem 4 remains unclear. We refer to Section 1.4 for a discussion on this.

These duality results uncover a far-reaching correspondence between computational assumptions and key aspects of time-bounded Kolmogorov complexity. In particular, these characterizations offer a novel way to investigate the main hardness conjectures of complexity theory (and the relationships among them) through the lens of Kolmogorov complexity and its properties.

1.3 Techniques

In this section, we explain the main ideas behind our proofs. We focus on Theorems 1, 2, and 3.

Theorem 1: OWFs vs Average-Case SoI for pK^t via Conditional Coding. The equivalence between the non-existence of one-way functions and average-case symmetry of information is proved via average-case conditional coding. That is, we first show that one-way functions do not exist if and only if average-case conditional coding holds. We then argue that symmetry of information and conditional coding are equivalent in the average-case setting.

Part 1: OWFs vs Conditional Coding. First, we explain how the non-existence of one-way functions implies average-case conditional coding. Consider an arbitrary samplable distribution $\{\mathcal{D}_n\}_n$ over $\{0,1\}^n \times \{0,1\}^n$ and assume one-way functions do not exist. Our goal is to, on average when

(x, y) is sampled from $\mathcal{D}_n$, give a short (length roughly $\log \frac{1}{\mathcal{D}_n(x|y)}$) and efficient (time at most $p(n)$) description of x given access to y.

To gain some intuition, let us ignore the efficient part for now and recall how to prove the conditional coding theorem in the time-unbounded setting. First, we can assume without loss of generality that $D_n(x \mid y) \geq 2^{-n}$, since otherwise the desired bound on the complexity of x given y is trivial. One description of x would be to first describe the probability $p = \mathcal{D}_n(x \mid y)$ that x is sampled from $\mathcal{D}_n(\cdot \mid y)$ and then describe the index of x in the set $\{x' : \mathcal{D}_n(x' \mid y) \geq p\}$, which contains at most $1/p = 1/\mathcal{D}_n(x \mid y)$ elements. In general, this gives an (inefficient) description for x of length at least n bits (to describe the value $\mathcal{D}_n(x \mid y)$) plus $\log \frac{1}{\mathcal{D}_n(x|y)}$ bits (to describe the index). Thus, this description's length is worse than the trivial bound of $n + O(1)$ for x! To improve this, one uses a standard trick in Kolmogorov complexity: instead of describing $\mathcal{D}_n(x \mid y)$, one describes the largest power of two less than or equal to $\mathcal{D}_n(x \mid y)$. Let α be this value. Then, given α, one can also specify the index of x in the set $\{x' : \mathcal{D}_n(x' \mid y) \geq \alpha\}$. This gives a description for x given y of length at most $\log n$ (to describe α) plus at most $\log \frac{1}{\alpha} = O(1) + \log \frac{1}{\mathcal{D}_n(x|y)}$, as desired. However, this description is not an efficient one. Indeed, even assuming one-way functions do not exist, it is unclear how to easily compute the i-th element of the set $\{x' : \mathcal{D}_n(x' \mid y) \geq \alpha\}$ on average.

Instead, we take a different approach that relies on hashing. Our description of x given y will still include α, but instead of specifying the index of x in the set $\{x' : \mathcal{D}_n(x' \mid y) \geq \alpha\}$, we will specify the hash value $v = H(x)$ of x where H is a randomly chosen pairwise independent hash function. Setting parameters appropriately, we can guarantee that with high probability that v is of length $\log \frac{1}{\mathcal{D}_n(x|y)} + O(1)$ and that $v \neq H(x')$ for all $x' \in \{x' : \mathcal{D}_n(x' \mid y) \geq \alpha\}$. Thus, the value v uniquely specifies x in the set $\{x' : \mathcal{D}_n(x' \mid y) \geq \alpha\}$ and so this gives a description of x given y of length $(\log \frac{1}{\mathcal{D}_n(x|y)} + O(1)) + \log n + \log n$ (the first term comes from specifying v, the second from specifying α, and the third from specifying H).

We show that this description is also efficient on average assuming one-way functions do not exist. A first attempt might be to consider the candidate one-way function that takes as input randomness r, a parameter α, and a random hash function H, and outputs $(H(x), y, H, \alpha)$ where (x, y) is sampled from $\mathcal{D}_n$ using the randomness r. Since one-way functions do not exist, this function can be inverted on average. Thus, to try to go from the description (v, α, H) back to x one could try to run the inverter to find an x' such that $H(x') = v$. The difficulty is that x' is not guaranteed to be in the set $\{x' : \mathcal{D}_n(x' \mid y) \geq \alpha\}$ and so x' might not equal x, even though they both hash to v!

To get around this, we show that, assuming one-way functions do not exist, there exists an efficient algorithm B such that on average $B(x, y)$ outputs a constant factor approximation of $\mathcal{D}_n(x \mid y)$. Crucially, however $B(x, y)$ *never overestimates* $\mathcal{D}_n(x \mid y)$ (even in the worst-case). To prove the existence of B we build on [IL89, IL90]. Assuming we have B, we can then consider the candidate one-way function that takes as input randomness r, a parameter α, and a random hash function H, then samples (x, y) from $\mathcal{D}_n$ using the randomness r and outputs $\perp$ if $B(x, y) \leq \alpha$ and otherwise outputs $(H(x), y, H, \alpha)$. This means that the one-way function always outputs $\perp$ on any x' where $\mathcal{D}_n(x' \mid y) < \alpha$. Using this new candidate one-way function, we can fix the aforementioned difficulty in the previous paragraph and efficiently go from the description (v, α, H) back to x, as desired.

We note that pK complexity is particularly useful when implementing the plan described above, since we need random bits (e.g., to obtain H) and additional information that depends on the

choice of these random bits (e.g., the hash value $v = H(x)$ once we have H).

To show that average-case conditional coding theorem allows us to break any candidate one-way function $f\colon \{0,1\}^n \to \{0,1\}^n$, we consider a (polynomial-time samplable) distribution $\mathcal{D}$ which samples $(x, f(x))$, where x is uniformly random. Note that this distribution can be equivalently viewed as first sampling $y := f(z)$ for a uniformly random z and then sampling $x \sim \mathcal{D}(\cdot \mid y)$, where, crucially, $\mathcal{D}(\cdot \mid y)$ is uniformly distributed on $f^{-1}(y)$. Now assuming that average-case conditional coding holds, we have for most pairs (x, y) sampled from $\mathcal{D}$,

$$\mathsf{pK}^{\mathsf{poly}(n)}(x \mid y) \lesssim \log \frac{1}{\mathcal{D}(x \mid y)}.$$

Then, by an averaging argument, for most $y := f(z)$ (where z is uniformly random) the above condition holds with high probability over x sampled from $\mathcal{D}(\cdot \mid y)$. Since $\mathcal{D}(\cdot \mid y)$ is uniformly distributed on $f^{-1}(y)$, this means that for most (say at least half) $x \in f^{-1}(y)$, it is the case that

$$\mathsf{pK}^{\mathsf{poly}(n)}(x \mid y) \lesssim \log \frac{1}{\mathcal{D}(x \mid y)} = \log |f^{-1}(y)|. \tag{1}$$

A useful property of probabilistic Kolmogorov complexity is that if $\mathsf{pK}^{\mathsf{poly}(n)}(a \mid b)$ is at most k, where $a, b \in \{0,1\}^n$, then there is a universal randomized algorithm USamp that, given b as input, runs in $\mathsf{poly}(n)$ time and outputs a with probability at least $1/O\big(n \cdot 2^k\big)$ (see Definition 21 and Proposition 22). Then combining this fact with Equation (1), we get that for at least half of the $x \in f^{-1}(y)$, $\mathsf{USamp}(y)$ outputs x with probability at least $1/O\big(n \cdot |f^{-1}(y)|\big)$, which implies that it outputs *some* $x' \in f^{-1}(y)$ with probability at least $1/O(n)$. This gives an efficient algorithm that finds a pre-image of y with high probability, for most y.

Part 2: Conditional Coding vs Symmetry of Information. It remains to show the equivalence between average-case conditional coding and average-case symmetry of information. We describe how to get the latter from the former. Roughly speaking, if average-case conditional coding holds, then we have for every polynomial-time samplable distribution $\mathcal{D}$ over $\{0,1\}^n \times \{0,1\}^n$ and for almost all pairs (x, y) sampled from $\mathcal{D}$,

$$\mathsf{pK}^{\mathsf{poly}(n)}(x \mid y) \lesssim \log \frac{1}{\mathcal{D}(x \mid y)}.$$

By the fact that $\mathcal{D}(x \mid y) = \mathcal{D}(x, y)/\mathcal{D}'(y)$, where $\mathcal{D}'$ is the marginal distribution of $\mathcal{D}$ on the second half, we can rewrite the above as

$$\mathsf{pK}^{\mathsf{poly}(n)}(x \mid y) \lesssim \log \frac{1}{\mathcal{D}(x, y)} - \log \frac{1}{\mathcal{D}'(y)}. \tag{2}$$

Remember that for symmetry of information, we aim to show

$$\mathsf{pK}^{\mathsf{poly}(n)}(x \mid y) \lesssim \mathsf{pK}^{\mathsf{poly}(n)}(x, y) - \mathsf{pK}^{\mathsf{poly}(n)}(y). \tag{3}$$

Therefore, it suffices to show that $\mathsf{pK}^{\mathsf{poly}(n)}(x, y) \gtrsim \log \frac{1}{\mathcal{D}(x,y)}$ and that $\mathsf{pK}^{\mathsf{poly}(n)}(y) \lesssim \log \frac{1}{\mathcal{D}'(y)}$. Note that the second inequality is exactly the (ordinary) coding theorem for $\mathsf{pK}^{\mathsf{poly}}$, which holds unconditionally thanks to [LOZ22]. For the first inequality, we use an "incompressibility" property of Kolmogorov complexity, which says that for every distribution $\mathcal{E}$, almost all elements z sampled

from $\mathcal{E}$ have (resource-unbounded) Kolmogorov complexity at least $\log \frac{1}{\mathcal{E}(z)}$ minus some small additive term (see Lemma 9). Since it can be shown that $\mathsf{pK}^{\mathsf{poly}(n)}(x,y)$ is lower bounded by $\mathsf{K}(x,y)$ (modulo some additive logarithmic term), we get that the first inequality holds for almost all (x,y) sampled from $\mathcal{D}$.

Similarly, to show that average-case symmetry of information implies average-case conditional coding, we can "reverse" the above argument and show Equation (2) from Equation (3), in which case we need to show $\mathsf{pK}^{\mathsf{poly}(n)}(x,y) \lesssim \log \frac{1}{\mathcal{D}(x,y)}$ and $\mathsf{pK}^{\mathsf{poly}(n)}(y) \gtrsim \log \frac{1}{\mathcal{D}'(y)}$. These again follow from the coding theorem and the "incompressibility" property for $\mathsf{pK}^{\mathsf{poly}}$.

Theorem 2: OWFs vs Average-Case SoI for rK^t via Meta-Complexity. To show the equivalence between average-case SoI for rK^t and the non-existence of quasi-polynomial-time variants of one-way functions, we employ a general approach of showing symmetry of information from meta-complexity [Hir22b]. It was shown in [Hir22b] that the existence of an efficient algorithm that approximates resource-bounded Kolmogorov complexity implies a corresponding version of SoI. We apply a similar proof technique to the average-case setting, and show that Item 1 implies Item 4 in Theorem 2, i.e., $\mathsf{rK}^{\mathsf{poly}}$ is approximated by K if a one-way function does not exist. For simplicity, we consider polynomial-time bounds in this proof overview.

The key technical ingredient is a *pseudorandom generator construction* $G_k \colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^k$ [TV07]. A pseudorandom generator construction takes a "hard" string $x \in \{0,1\}^n$ and a seed $z \in \{0,1\}^d$ and outputs a pseudorandom sequence $G_k(x;z)$ with the following reconstruction property. If a function D distinguishes the output distribution of $G_k(x;\text{-})$ from the uniform distribution, then the D-oracle Kolmogorov complexity of x is small. In other words, if the D-oracle Kolmogorov complexity of x is large, then the output distribution $G_k(x;\text{-})$ looks pseudorandom to D. Following [Hir22b], we use the specific pseudorandom generator construction of [RRV02], which satisfies the reconstruction property that

$$\mathsf{rK}^{\mathsf{poly}(n),D}(x) \leq k + O(\log^3 n) \tag{4}$$

and the seed length is $d = O(\log^3 n)$. Moreover, a pseudorandom generator construction has an "advice function", which outputs the witness for Equation (4), namely, the description of a D-oracle randomized program of length $k + O(\log^3 n)$ that prints x with high probability. This property enables us to show SoI with an efficient encoder.

We instantiate the approach of [Hir22b] using the approximation algorithm of [IRS21]. Under the assumption that there is no one-way function, it was shown in [IRS21] that for every polynomial-time samplable distribution $\mathcal{D}$, there exists an efficient average-case algorithm A that approximates the *resource-unbounded* Kolmogorov complexity $\mathsf{K}(x)$ of a string $x \sim \mathcal{D}$. We observe that this algorithm enables us to approximate the conditional Kolmogorov complexity $\mathsf{K}(x \mid y)$ as well because

$$\mathsf{K}(x \mid y) \approx \mathsf{K}(x,y) - \mathsf{K}(y)$$

by the symmetry of information for K. Let A be an average-case algorithm that approximates $\mathsf{K}(x \mid y)$ on input $(x,y) \sim \mathcal{D}$.

Using the algorithm A, let us explain how to prove $\mathsf{rK}^{\mathsf{poly}}(x \mid y) \approx \mathsf{K}(x \mid y)$ for most $(x,y) \sim \mathcal{D}$. The idea is to try to distinguish the pseudorandom generator construction $G_k(x;\text{-})$ from the uniform distribution by using the approximation algorithm A. On the one hand, observe that

$$A(G_k(x;z),y) \approx \mathsf{K}(G_k(x;z) \mid y) \lesssim \mathsf{K}(x \mid y) + |z|$$

because $G_k(x;z)$ is computable given x, z and k as input. Note that $|z| = O(\log^3 n)$, which is negligible. On the other hand, by a standard counting argument, we have

$$A(w,y) \approx \mathsf{K}(w \mid y) \gtrsim k$$

for a random $w \sim \{0,1\}^k$. These two inequalities show that when $k \gtrsim \mathsf{K}(x \mid y)$, the algorithm $A(\text{-}, y)$ can distinguish $G_k(x;\text{-})$ from the uniform distribution. By the reconstruction property from Equation (4), we obtain $\mathsf{rK}^{\mathsf{poly}(n)}(x \mid y) \leq k \approx \mathsf{K}(x \mid y)$, which completes the proof.

In fact, there are important technical details hidden in the outline above. We need to choose $k \approx \mathsf{K}(x \mid y)$ depending on (x,y), which is chosen randomly from a distribution $\mathcal{D}$. Thus, the distribution $(G_k(x;z),y)$ may not be polynomial-time samplable in general. This is problematic for us because A is guaranteed to work correctly only with respect to polynomial-time samplable distributions. The issue is not present in the worst-case setting [Hir22b]. Fortunately, it turns out that there is a simple way to circumvent this issue. We consider a distribution that randomly chooses $k \sim [2n]$ as the input distribution of A. Using that A works correctly with high probability, we can use A to approximate $\mathsf{K}(G_k(x;z) \mid y)$ for *every* $k \in [2n]$ for a randomly chosen $(x,y) \sim \mathcal{D}$.

Once we obtain the approximation of $\mathsf{rK}^{\mathsf{poly}}$ by K, SoI for $\mathsf{rK}^{\mathsf{poly}}$ easily follows from SoI for K.

To prove that SoI for $\mathsf{rK}^{\mathsf{poly}}$ implies the non-existence of one-way functions, a natural idea is to try to follow our approach for Theorem 1. However, an unconditional (ordinary) coding theorem for rK is currently unknown, and this was an important ingredient in that argument. Instead, we adapt the proof ideas of [LW95] to the average-case setting. In general, they proved that if SoI for $\mathsf{K}^{\mathsf{poly}}$ holds with an additive error of $e(n)$, then any one-way function can be inverted in time $n^{O(1)} \cdot 2^{O(e(n))}$. In our case, the reconstruction property of Equation (4) incurs an additive error of $O(\log^3 n)$, which makes the running time of the inverter quasi-polynomial. More generally, this is why the equivalence is proved in the quasi-polynomial time regime.

This completes our sketch of the proof of Theorem 2.

The proof overviews presented above highlight two perspectives that can be leveraged to obtain a characterization of the existence one-way functions via average-case symmetry of information: (*i*) employing conditional coding as a bridge between the two statements, and (*ii*) a meta-computational approach through meta-complexity. The two approaches come with different benefits, and shed light on distinct aspects of the duality between the statements. In terms of generality, we note that the meta-computational approach can also be implemented in the setting of pK^t, which provides a different proof of Theorem 1. On the other hand, the same method does not seem to work in a worst-case complexity setting, since in the worst-case the time-unbounded and time-bounded Kolmogorov complexities of a string can be quite far from each other. In contrast, the conditional coding perspective can be employed to prove Theorem 4 (see Section 6).

Theorem 3: Characterizing DistNP vs HeurBPP via Conditional Extrapolation. The average-case easiness of NP is derived from the *independent* version of conditional coding and language compression in a way that is similar to the other characterization results for one-way functions and the worst-case complexity of NP. Remember that the *independent* version of the statements comes with two samplable distributions $\mathcal{C}$ and $\mathcal{D}$, where $\mathcal{D}$ is over $\{0,1\}^n \times \{0,1\}^n$, and $\mathcal{C}$ is over conditionings of the second half element of $\mathcal{D}$. To solve an NP problem L on average under a samplable distribution $\mathcal{E}$, we set $\mathcal{D}$ to the distribution of $(x, y \circ b)$, where $x \sim \{0,1\}^n$,

$y \sim \mathcal{E}$, and $b = 1$ if and only if x is a witness for $y \in L$ (otherwise, $b = 0$).[1] By letting $\mathcal{C}$ be the distribution of $y \circ 1$ for $y \sim \mathcal{E}$, a pair (x, y) selected as $y \circ 1 \sim \mathcal{C}$ and $x \sim \mathcal{D}(\cdot \mid y \circ 1)$ is distributed over the witness-instance pairs for L. Crucially, the marginal distribution of y corresponds to $\mathcal{E}$, and in case y is a positive instance, the marginal distribution of x is uniform over the witnesses for y. Using this, the average-case easiness of L follows from the observation that the efficient search of witness from the language compression holds even in the average-case settings with respect to witness-instance pairs.

To show the opposite direction, we introduce a new concept of *conditional extrapolation*, which may be of independent interest. The conditional extrapolation for a joint distribution $\mathcal{D}$ over $\{0,1\}^* \times \{0,1\}^*$ is a probabilistic algorithm $\mathsf{CondExt}$ that is given a string y in the support of the second half of $\mathcal{D}$ and selects a sample x according to $\mathcal{D}(\cdot \mid y)$ with a small statistical error (note that $\mathcal{D}(\cdot \mid y)$ is not efficiently samplable in general even if $\mathcal{D}$ is samplable). We are interested in achieving this when y is sampled from a *different* distribution $\mathcal{C}$. In more detail, we consider the following statement involving distributions $\mathcal{C}$ and $\mathcal{D}$:

Conditional Extrapolation. There exists a probabilistic polynomial-time algorithm $\mathsf{CondExt}$ such that for all $\varepsilon^{-1}, \delta^{-1} \in \mathbb{N}$,

$$\Pr_{y \sim \mathcal{C}}\left[\mathsf{L}_1\Big(\mathsf{CondExt}(y; 1^{\varepsilon^{-1}}, 1^{\delta^{-1}}), \mathcal{D}(\cdot \mid y)\Big) \leq \varepsilon\right] \geq 1 - \delta,$$

where we use the notation L_1 to refer to the total variation distance between two distributions.

As a key lemma in our proof of equivalence, we show that $\mathsf{DistNP} \subseteq \mathsf{HeurBPP}$ holds if and only if conditional extrapolation is feasible for every samplable joint distribution $\mathcal{D}$ *on average* over the choice of the conditional string y, where y is selected according to an arbitrary samplable distribution $\mathcal{C}$. The proof of the lemma is based on an adaptation of the proof of the well-known equivalence result between one-way functions and distributional one-way functions [IL89], where we apply a heuristic scheme for the search version of the circuit SAT problem instead of an inverting algorithm.

Conditional extrapolation yields the implication from the average-case easiness of NP to the *independent* version of conditional coding, language compression, and symmetry of information. More specifically, we apply the conditional extrapolation to join two independent samplable distributions $\mathcal{C}$ and $\mathcal{D}$ and make a *samplable* distribution of (x, y) for $y \sim \mathcal{C}$ and $x \sim \mathcal{D}(\cdot \mid y)$ by regarding the sampling process as $y \sim \mathcal{C}$ and $x \sim \mathsf{CondExt}(y)$. Namely, by conditional extrapolation, we can derive the *independent* version of the statements from the corresponding average-case statements for joint samplable distributions. Since the latter is shown under the non-existence of one-way functions (a consequence of $\mathsf{DistNP} \subseteq \mathsf{HeurBPP}$), we obtain the *independent* version of the statements from $\mathsf{DistNP} \subseteq \mathsf{HeurBPP}$.

1.4 Open Problems

A summary of our results appears in Appendix B. Note that in the context of one-way functions we have obtained a precise characterization through average-case symmetry of information, average-case conditional coding, and average-case language compression. On the other hand, the exact role of symmetry of information remains mysterious in the correspondences for the worst-case easiness

[1] The actual construction is a little more involved in order to meet the conditions in the statement; e.g., $\mathcal{C}$ is over the support of the second half of $\mathcal{D}$.

of NP and for the average-case easiness of NP. In light of Theorem 3 and Theorem 32, we ask the following question.

Problem 5. *Is Independent Average-Case Symmetry of Information* equivalent *to* DistNP $\subseteq$ HeurBPP*?*

Next, we consider *worst-case* (time-bounded) symmetry of information. Hirahara [Hir22b], Goldberg and Kabanets [GK22], and Goldberg, Kabanets, Lu, and Oliveira [GKLO22] showed that the *errorless* average-case easiness of DistNP implies worst-case symmetry of information. For instance, [GKLO22] proved that worst-case symmetry of information holds for pK^t under the assumption that DistNP $\subseteq$ AvgBPP. While we believe that some of our results can be adapted to show correspondences between DistNP $\subseteq$ AvgBPP and certain "certified" average-case versions of key principles from Kolmogorov complexity, a proof that worst-case symmetry of information implies a corresponding easiness assumption for NP remains elusive.

Problem 6. *Is there a natural computational assumption that is equivalent to Worst-Case Symmetry of Information?*

It would also be interesting to obtain an unconditional analogue of Theorem 2 in the polynomial time regime. This is connected to the advice complexity of the reconstruction procedure from [RRV02], which incurs a poly-logarithmic additive factor in our bounds on Kolmogorov complexity.

1.5 Related Work

In this section we discuss the broader context surrounding our results, providing pointers to related research directions and the most relevant recent developments.

As mentioned above, Longpré and Watanabe [LW95] showed that if P = NP then worst-case time-bounded SoI holds. This result has been improved by [GK22, Hir22b] (see also the subsequent paper [GKLO22]), where it was shown that the same conclusion holds under the weaker assumption that NP admits errorless heuristic schemes. In contrast, our results establish the first equivalence between a natural computational assumption (inverting one-way functions) and average-case time-bounded SoI.

Longpré [Lon86] proved that symmetry of information holds for a space-bounded notion of Kolmogorov complexity, while Ronneburger [Ron04] established that it fails for Levin's Kt complexity. Lee and Romashchenko [LR05] investigate symmetry of information for variants of distinguishing complexity (CD^t), including non-deterministic distinguishing complexity and non-deterministic distinguishing complexity with randomness. On the other hand, here we are concerned with variants of time-bounded Kolmogorov complexity that are equivalent to K^t under standard hardness assumptions. Liu and Pass [LP22] proved that a general form of time-bounded symmetry of information for strings x and y of different lengths fails when K^t complexity is defined with respect to RAM-machines (as opposed to Turing machines).

Liu and Pass [LP20] (see also [LP21, RS21, IRS21]) showed an equivalence between inverting one-way functions and the error-prone average-case easiness of computing K^t complexity. Our results (Theorem 1 and Theorem 2) are incomparable to theirs, as we do not consider an equivalence between two computational assumptions (inverting one-way functions and easiness of computing K^t), i.e., we relate instead one-way functions and the validity of key principles from Kolmogorov complexity in the time-bounded setting. It is worth noting that several additional characterizations of one-way functions are known, e.g., [IL89, IL90, HILL99, Gol90].

Several papers have considered coding [AF09, LO21, LOZ22] and language compression [BLM00, BFL01, BLvM05, Hir21] in the time-bounded setting. While an optimal coding theorem for pK^t is known unconditionally [LOZ22], the existence of a result of this form for rK^t and K^t is currently only known under a derandomization assumption (see, e.g., [AF09]). Weak forms of language compression hold for variants of distinguishing complexity (see [BLvM05]). In general, our results and the literature on this topic indicate that language compression most likely does not hold for time-bounded Kolmogorov complexity measures.

One of the proofs discussed in Section 1.3 relies on techniques from meta-complexity, a rapidly developing area which investigates the complexity of computational problems and tasks that are themselves about computations and their complexity. We refer to the surveys [All21, Hir22a] for an overview of recent results in this area.

Acknowledgements. We thank the anonymous STOC reviewers for their comments and suggestions. We would like to thank Hanlin Ren for asking a question about conditional coding that led to part of these investigations and for initial discussions. We are grateful to Osamu Watanabe for conversations that motivated Theorem 2. We also thank Rahul Santhanam for related discussions. This work received support from the Royal Society University Research Fellowship URF\R1\191059, the EPSRC New Horizons Grant EP/V048201/1, and the Centre for Discrete Mathematics and its Applications (DIMAP) at the University of Warwick.

2 Preliminaries

2.1 Basic Definitions and Notation

For a string $w \in \{0,1\}^*$, we use $|w| \in \mathbb{N}$ to denote its length. The empty string is denoted by ϵ. For any $w \in \{0,1\}^*$ and any $i \leq |w|$, we let $w_{[i]}$ denote the i-bit prefix of w.

Probability Distributions. Due to our investigation of conditional coding, we will be mostly interested in distributions supported over pairs of strings. Unless stated otherwise, we use $\mathcal{D} = \{\mathcal{D}_n\}_{n\in\mathbb{N}}$ to denote an ensemble of polynomial-time samplable distributions, where each $\mathcal{D}_n$ is supported over $\{0,1\}^{\ell_1(n)} \times \{0,1\}^{\ell_2(n)}$, and ℓ_1 and ℓ_2 are polynomials satisfying $\ell_1(n), \ell_2(n) \geq n$.[2] We let PSamp be the collection of ensembles of distributions that can be sampled in polynomial time. When n is clear from context, we might simply write $\mathcal{D}$ instead of $\mathcal{D}_n$. We use $\mathcal{D}^{(2)}$ to refer to the marginal distribution of the second half element of $\mathcal{D}$.

We use $\mathcal{D}_n(x,y)$ to denote the probability that the pair (x,y) is sampled from $\mathcal{D}_n$. Similarly, $\mathcal{D}_n(x \mid y)$ denotes the probability x is sampled from $\mathcal{D}_n$ given y is sampled.

One-Way Functions. We will be concerned with one-way functions that are secure against uniform probabilistic polynomial-time algorithms (PPTs). As usual, we say that an efficiently computable collection $f = \{f_n\}_{n\geq 1}$ satisfying $f_n \colon \{0,1\}^n \to \{0,1\}^{\mathsf{poly}(n)}$ is a *one-way function* (OWF) if for every PPT algorithm A and constant $c \geq 1$ and for every sufficiently large n, we have

$$\Pr_{x\sim\{0,1\}^n}[f(A(1^n, f(x))) = f(x)] \leq n^{-c}. \tag{5}$$

[2] Recall that $\mathcal{D}$ can be sampled in polynomial time if there is a polynomial-time algorithm Samp such that $\mathsf{Samp}(1^n, r)$ is distributed according to $\mathcal{D}_n$ when r is a uniformly random string of length $\mathsf{poly}(n)$.

It is well known that the existence of one-way functions does not crucially depend on the success probability in this definition (i.e., weak and strong one-way functions are equivalent) and on the output length of each f_n (we can assume output length $m = n$ without loss of generality). We refer to [Gol01] for more details.

Similarly, we say that f is an *infinitely-often one-way function* (i.o. OWF) if for every PPT A and constant $c \geq 1$ as above, there are infinitely many values of n such that Equation (5) holds.

Time-Bounded Kolmogorov Complexity. Let U be a Turing machine. Given a positive integer t and a string $x \in \{0,1\}^*$, we let

$$\mathsf{K}^t_U(x) = \min_{p \in \{0,1\}^*} \Big\{ |p| \;\mid\; U(p, \epsilon) \text{ outputs } x \text{ in at most } t \text{ steps} \Big\}.$$

We say that $\mathsf{K}^t_U(x)$ is the *t-time-bounded Kolmogorov complexity of x* (with respect to U). As usual, we fix U to be a time-optimal machine [LV19], and drop the index U when referring to time-bounded Kolmogorov complexity measures. In addition, we use $\mathsf{K}(x)$ to denote the (time-unbounded) Kolmogorov complexity of x.

It will be useful to consider a randomized variant of K^t where instead of having a deterministic machine that prints x, we consider a randomized machine that generates x with high probability. Given a probability parameter $\delta \in [0,1]$ and a positive integer t, we let

$$\mathsf{rK}^t_\delta(x) = \min_{p \in \{0,1\}^*} \Big\{ |p| \;\mid\; \Pr_{r \sim \{0,1\}^t}[U(p,r) \text{ outputs } x \text{ in at most } t \text{ steps}] \geq \delta \Big\}.$$

denote the *t-time-bounded randomized Kolmogorov complexity of x*. For simplicity, we omit δ when $\delta = 2/3$, i.e., when x is printed with high probability.

We also make use of another probabilistic variant of $\mathsf{K}^t(x)$ introduced by Goldberg, Kabanets, Lu, and Oliveira [GKLO22], which we define next. For a string x, the *probabilistic t-time-bounded Kolmogorov complexity of x* is defined as

$$\mathsf{pK}^t(x) = \min \left\{ k \in \mathbb{N} \;\middle|\; \Pr_{r \sim \{0,1\}^{t(|x|)}} \Big[\exists\, p \in \{0,1\}^k \text{ s.t. } U(p,r) \text{ outputs } x \text{ within } t(|x|) \text{ steps} \Big] \geq \frac{2}{3} \right\}. \tag{6}$$

It is known that the deterministic and probabilistic time-bounded Kolmogorov complexities of a string essentially coincide, under a plausible circuit lower bound assumption: for every string x and time bound $t(n) \geq n$, $\mathsf{pK}^t(x) \leq \mathsf{K}^t(x)$ and $\mathsf{K}^{\mathsf{poly}(t)}(x) \leq \mathsf{pK}^t(x) + O(\log |x|)$ [GKLO22]. We refer to [LO22] for more background on probabilistic time-bounded Kolmogorov complexity and its applications.

These definitions can be extended to *conditional* Kolmogorov complexity in the natural way. For instance, in $\mathsf{pK}^t(x \mid y)$ the machine U is also given access to the input string y in Equation (6) above. For concreteness, we assume that y is given in a separate input tape.

Average-case Complexity. Recall that a pair $(L, \mathcal{D})$ is a *distributional problem* if $L \subseteq \{0,1\}^*$ and $\mathcal{D} = \{\mathcal{D}_n\}_{n \in \mathbb{N}}$ is a distribution family, where each $\mathcal{D}_n$ is over $\{0,1\}^*$.

We let DistNP denote the set of distributional problems $(L, \mathcal{D})$ with $L \in \mathsf{NP}$ and $\mathcal{D} \in \mathsf{PSamp}$.

A distributional problem $(L, \mathcal{D})$ is said to admit a (error-prone) *heuristic scheme* if there exists a probabilistic polynomial-time algorithm A such that for every $n, k \in \mathbb{N}$,

$$\Pr_{x \sim \mathcal{D}_n, A} \Big[A(x; 1^n, 1^k) \neq L(x) \Big] \leq 1/k.$$

We let $\mathsf{HeurBPP}$ denote the set of distribution problems that admit a heuristic scheme. For more information about average-case complexity, we refer to [BT06].

2.2 Technical Lemmas

Kolmogorov Complexity and Coding Results. We will need the following results.

Theorem 7 (Efficient Coding Theorem [LOZ22])**.** *For every polynomial-time samplable distribution family $\{\mathcal{D}_n\}_n$, there exists a polynomial p such that for every $x \in \mathsf{Support}(\mathcal{D}_n)$*

$$\mathsf{pK}^{p(n)}(x) \leq \log \frac{1}{\mathcal{D}_n(x)} + \log p(n).$$

Lemma 8 ([GKLO22])**.** *There is a universal constant $c > 0$ such that the following holds. For every time bound $t \in \mathbb{N}$ and $x \in \{0,1\}^n$,*

$$\mathsf{K}(x \mid t) \leq \mathsf{pK}^t(x) + c \log n.$$

Lemma 9. *The following two hold.*

1. *For any distribution family $\{\mathcal{D}_n\}_n$ over $\{0,1\}^n$,*

$$\Pr_{x \sim \mathcal{D}_n}\left[\mathsf{K}(x) < \log \frac{1}{\mathcal{D}_n(x)} - \alpha\right] < \frac{1}{2^\alpha}.$$

2. *For any distribution family $\{\mathcal{D}_n\}_n$ over $\{0,1\}^n \times \{0,1\}^n$ and any $y \in \mathsf{support}(\mathcal{D}_n^{(2)})$, where $\mathcal{D}_n^{(2)}$ is the marginal distribution of $\mathcal{D}_n$ on the second half,*

$$\Pr_{x \sim \mathcal{D}_n(\cdot|y)}\left[\mathsf{K}(x \mid y) < \log \frac{1}{\mathcal{D}_n(x \mid y)} - \alpha\right] < \frac{1}{2^\alpha}.$$

Proof. We show the second item. The first item can be shown in a similar way.

Fix any $y \in \mathsf{support}(\mathcal{D}_n^{(2)})$, we have

$$\mathop{\mathbf{E}}_{x \sim \mathcal{D}_n(\cdot|y)}\left[\frac{2^{\mathsf{K}(x|y)}}{\mathcal{D}_n(x \mid y)}\right] = \sum_{x \in \mathsf{Support}(\mathcal{D}_n(x|y))} \mathcal{D}_n(x \mid y) \cdot \frac{2^{\mathsf{K}(x|y)}}{\mathcal{D}_n(x \mid y)} = \sum_{x \in \mathsf{Support}(\mathcal{D}_n)} 2^{\mathsf{K}(x|y)} \leq 1,$$

Where the last inequality follows from Kraft's inequality.

By Markov's inequality, we obtain that

$$\Pr_{x \sim \mathcal{D}_n(\cdot|y)}\left[\frac{2^{\mathsf{K}(x|y)}}{\mathcal{D}_n(x \mid y)} > 2^\alpha\right] < \frac{\mathbf{E}_{x \sim \mathcal{D}_n(\cdot|y)}\left[\frac{2^{\mathsf{K}(x|y)}}{\mathcal{D}_n(x|y)}\right]}{2^\alpha} < \frac{1}{2^\alpha},$$

which implies

$$\Pr_{x \sim \mathcal{D}_n(\cdot|y)}\left[\mathsf{K}(x \mid y) < \log \frac{1}{\mathcal{D}_n(x \mid y)} - \alpha\right] < \frac{1}{2^\alpha},$$

as desired. □

Pairwise Independent Hash Family. We review below a standard result result about hash functions.

Definition 10 (Pairwise Independent Hash Family)**.** For $m, n \in \mathbb{N}$, a family of hash functions $\mathcal{H} := \{h \colon \{0,1\}^n \to \{0,1\}^m\}$ is called *pairwise independent* if, for any distinct $x, x' \in \{0,1\}^n$, and any $y, y' \in \{0,1\}^m$, we have

$$\Pr_{h \sim \mathcal{H}}\left[h(x) = y \wedge h(x') = y'\right] = \frac{1}{2^{2m}}.$$

Theorem 11 (See e.g., [Vad12, Problem 3.3])**.** *Let $m, n \in \mathbb{N}$. There is a family of pairwise independent hash functions $H_{n,m} := \{h_w \colon \{0,1\}^n \to \{0,1\}^m\}_w$, where each h_w is indexed by a $w \in \{0,1\}^{n+m}$. Moreover, given n, m, w, and x, $h_w(x)$ can be computed in time $\mathsf{poly}(n, m)$.*

Proposition 12. *Let $n, m, \alpha \in \mathbb{N}$ and let $\mathcal{H} := \{h \colon \{0,1\}^n \to \{0,1\}^{m+\alpha}\}$ be a pairwise independent hash family. Then for every $S \subseteq \{0,1\}^n$ with $|S| \leq 2^m$, with probability at least $1 - 1/2^\alpha$, we have*

$$H(x) \neq H(x') \text{ for all } x' \in S \setminus \{x\}.$$

Proof. Fix some $x' \in S \setminus \{x\}$. Because the hash family is pairwise independent, the probability that $H(x) = H(x')$ is at most $1/2^{m+\alpha}$. Union-bounding over all $x' \neq x$ in S (there are at most $2^m - 1$ of them), we get that the probability that $H(x) \neq H(x')$ for all $x' \in S \setminus \{x\}$ is at least $1 - 1/2^\alpha$. □

3 One-Way Functions, Average-Case Conditional Coding, Language Compression and Symmetry of Information

Theorem 13. *The following are equivalent.*

1. *Infinitely-often one-way functions do not exist.*
2. **(Strong Average-Case Conditional Coding)** *For every polynomial-time samplable distribution family $\{\mathcal{D}_n\}_n$, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and for every polynomial q, there exists a polynomial p such that for all n,*

$$\Pr_{(x,y)\sim\mathcal{D}_n}\left[\mathsf{pK}^{p(n)}(x \mid y) \leq \log \frac{1}{\mathcal{D}_n(x \mid y)} + \log p(n)\right] \geq 1 - \frac{1}{q(n)}.$$

3. **(Weak Average-Case Conditional Coding)** *For every polynomial-time samplable distribution family $\{\mathcal{D}_n\}_n$, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, there exists polynomials p and q such that for all n,*

$$\Pr_{(x,y)\sim\mathcal{D}_n}\left[\mathsf{pK}^{p(n)}(x \mid y) \leq \log \frac{1}{\mathcal{D}_n(x \mid y)} + \log p(n)\right] \geq \frac{1}{q(n)}.$$

4. **(Strong Average-Case Language Compression)** *For every recursive enumerable set $L \subseteq \{\{0,1\}^n \times \{0,1\}^n\}_n$, every polynomial-time samplable distribution family $\{\mathcal{D}_n\}_n$, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and every polynomial q, there exists a polynomial p such that for all n,*

$$\Pr_{(x,y)\sim\mathcal{D}_n}\left[x \in L_y \implies \mathsf{pK}^{p(n)}(x \mid y) \leq \log |L_y| + \log p(n)\right] \geq 1 - \frac{1}{q(n)}.$$

5. **(Weak Average-Case Language Compression)** *For every polynomial-time recursive enumerable set $L \subseteq \{\{0,1\}^n \times \{0,1\}^n\}_n$ and every polynomial-time samplable distribution family $\{\mathcal{D}_n\}_n$, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, there exist polynomials p and q such that for all n,*

$$\Pr_{(x,y)\sim\mathcal{D}_n}\left[x \in L_y \implies \mathsf{pK}^{p(n)}(x \mid y) \leq \log|L_y| + \log p(n)\right] \geq \frac{1}{q(n)}.$$

6. **(Strong Average-Case Symmetry of Information)** *For every polynomial-time samplable distribution family $\{\mathcal{D}_n\}_n$, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and for every polynomial q, there exists a polynomial p such that for all computable time bound t with $t(n) \geq p(n)$ and for all n,*

$$\Pr_{(x,y)\sim\mathcal{D}_n}\left[\mathsf{pK}^{t(n)}(x,y) \geq \mathsf{pK}^{t(n)}(x \mid y) + \mathsf{pK}^{t(n)}(y) - \log t(n)\right] \geq 1 - \frac{1}{q(n)}.$$

7. **(Weak Average-Case Symmetry of Information)** *For every polynomial-time samplable distribution family $\{\mathcal{D}_n\}_n$, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, there exist polynomials p and q such that for all computable time bound t with $t(n) \geq p(n)$ and for all n,*

$$\Pr_{(x,y)\sim\mathcal{D}_n}\left[\mathsf{pK}^{t(n)}(x,y) \geq \mathsf{pK}^{t(n)}(x \mid y) + \mathsf{pK}^{t(n)}(y) - \log t(n)\right] \geq \frac{1}{q(n)}.$$

Proof. We first show that the first 5 items (non-existence of infinitely-often OWFs, average-case conditional coding, and average-case language compression) are equivalent, as follows. Item 1 implies Item 2 (via Lemma 14). Item 2 implies Item 3, and Item 4 implies Item 5 trivially. Item 2 implies Item 4, and Item 3 implies Item 5 (via Lemma 18 and Lemma 19 respectively). Item 5 implies Item 1 (via Lemma 20).

We then establish equivalence between average-case conditional coding and average-case symmetry of information, by showing that Item 2 implies Item 6 (via Lemma 23), and that Item 7 implies Item 3 (via Lemma 24). □

3.1 Strong Average-Case Conditional Coding from Inverting OWFs

Lemma 14 (Item 1 ⇒ Item 2 in Theorem 13)**.** *If infinitely-often one-way functions do not exist, then strong average-case conditional coding holds.*

We need the following technical theorem.

Theorem 15 ([IL90, IL89]; see also [IRS21, Theorem 20])**.** *Assume infinitely-often one-way functions do not exist. Let $\{\mathcal{D}_m\}_m$ be a family of polynomial-time samplable distributions, where each $\mathcal{D}_m$ is over $\{0,1\}^m$, and let $\alpha_0 \geq 1$ be any constant. There exist a constant $c \geq 1$ and a randomized polynomial-time algorithm A such that for all m,*

$$\Pr_{z\sim\mathcal{D}_m, A}\left[A(z) \geq \frac{\mathcal{D}_m(z)}{c}\right] \geq 1 - \frac{1}{m^{\alpha_0}},$$

and for all $z \in \{0,1\}^m$

$$\Pr_A\left[A(z) \leq \mathcal{D}_m(z)\right] \geq 1 - \frac{1}{m^{\alpha_0}}.$$

Lemma 16. *Assume infinitely-often one-way functions do not exist. Let $\{\mathcal{D}_n\}_n$ be a family of polynomial-time samplable distributions, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and let $\alpha \geq 1$ be any constant. There exist a constant $c \geq 1$ and a randomized polynomial-time algorithm B such that for all n,*

$$\Pr_{(x,y)\sim\mathcal{D}_n,B}\left[B(x,y) \geq \frac{\mathcal{D}_n(x \mid y)}{c} \text{ and } B(x',y) \leq c \cdot \mathcal{D}_n(x' \mid y) \text{ for all } x' \in \mathsf{Support}(\mathcal{D}_n(\cdot \mid y))\right] \geq 1 - \frac{1}{n^\alpha}.$$

Proof. Let $\mathcal{D}'_n$ be the marginal distribution of $\mathcal{D}_n$ on the second half of the output, i.e., $\mathcal{D}'_n$ samples $(x,y) \sim \mathcal{D}_n$ and outputs y.

Let $\alpha_0 := 6\alpha$. Let $A_\mathcal{D}$ be the algorithm that, given $(x,y) \in \{0,1\}^n \times \{0,1\}^n$, runs for $\mathsf{poly}(n)$ times the algorithm from Theorem 15 with respect to $\mathcal{D}_n$ and takes the median. Let $A_{\mathcal{D}'}$ be the algorithm that, given $y \in \{0,1\}^n$, runs the algorithm from Theorem 15 with respect to $\mathcal{D}'_n$. Finally, let B be the algorithm that, given input $(x,y) \in \{0,1\}^n \times \{0,1\}^n$, outputs $\tilde{A}_\mathcal{D}(x,y)/\tilde{A}_{\mathcal{D}'}(y)$, where $\tilde{A}_\mathcal{D}(x,y) = \max\{\tilde{A}_\mathcal{D}(x,y), 1/2^{t(n)}\}$, $\tilde{A}_{\mathcal{D}'}(y)\max\{\tilde{A}_\mathcal{D}(x,y), 1/2^{t(n)}\}$, and t corresponds to amount of randomness used by the machine that samples $\{\mathcal{D}_n\}$. Also, B can use the same randomness for running both $A_\mathcal{D}$ and $A_{\mathcal{D}'}$.

We argue the correctness of B. First consider an algorithm A from Theorem 15 with respect to $\mathcal{D}_n$. By averaging, we have

$$\Pr_{(x,y)\sim\mathcal{D}_n}\left[\Pr_A\left[A(x,y) \geq \frac{\mathcal{D}_n(x,y)}{c}\right] \geq 1 - \frac{1}{n^{\alpha_0/2}}\right] \geq 1 - \frac{1}{n^{\alpha_0/2}}. \tag{7}$$

Also, for all $(a,b) \in \{0,1\}^n \times \{0,1\}^n$

$$\Pr_A\left[A(a,b) \leq \mathcal{D}_n(a,b)\right] \geq 1 - \frac{1}{n^{\alpha_0}}. \tag{8}$$

By standard concentration bound, we note that the randomized algorithm $A_\mathcal{D}$ (which repeats A for $\mathsf{poly}(n)$ times and takes the median) makes the "success" probability of A in both Equation (7) and Equation (8) become $1 - 1/\exp(n)$. In particular, we have

$$\Pr_{(x,y)\sim\mathcal{D}_n}\left[\Pr_{A_\mathcal{D}}\left[A_\mathcal{D}(x,y) \geq \frac{\mathcal{D}_n(x,y)}{c}\right] \geq 1 - 1/\exp(n)\right] \geq 1 - \frac{1}{n^{\alpha_0/2}},$$

which implies

$$\Pr_{(x,y)\sim\mathcal{D}_n,A_\mathcal{D}}\left[A_\mathcal{D}(x,y) \geq \frac{\mathcal{D}_n(x,y)}{c}\right] \geq 1 - \frac{1}{n^{\alpha_0/3}}, \tag{9}$$

Also, by applying a union bound over all (a,b) on the new Equation (8) but with exponentially small failure probability, we get

$$\Pr_{A_\mathcal{D}}\left[A_\mathcal{D}(a,b) \leq \mathcal{D}_n(a,b) \text{ for all } (a,b)\right] \geq 1 - \frac{1}{n^{\alpha/2}}. \tag{10}$$

By combining Equation (9) and Equation (10), we get

$$\begin{aligned}
&\Pr_{(x,y)\sim\mathcal{D}_n,A_\mathcal{D}}\left[A_\mathcal{D}(x,y) \geq \frac{\mathcal{D}_n(x,y)}{c} \quad \text{and} \quad A_\mathcal{D}(a,y) \leq \mathcal{D}_n(a,y) \text{ for all } a \in \{0,1\}^n\right] \\
&\geq \Pr_{(x,y)\sim\mathcal{D}_n,A_\mathcal{D}}\left[A_\mathcal{D}(x,y) \geq \frac{\mathcal{D}_n(x,y)}{c} \quad \text{and} \quad A_\mathcal{D}(a,b) \leq \mathcal{D}_n(a,b) \text{ for all } (a,b)\right] \\
&\geq 1 - \frac{1}{2n^\alpha}.
\end{aligned} \tag{11}$$

Note that since every element in the support of $\mathcal{D}_n$ has probability mass at least $1/2^{t(n)}$, for $\tilde{A}_{\mathcal{D}}$ (which outputs the larger of $A_{\mathcal{D}}$ and $1/2^{t(n)}$), Equation (11) yields

$$\begin{aligned}&\mathop{\mathbf{Pr}}_{(x,y)\sim\mathcal{D}_n,\tilde{A}_{\mathcal{D}}}\left[\tilde{A}_{\mathcal{D}}(x,y) \geq \frac{\mathcal{D}_n(x,y)}{c} \quad \text{and} \quad \tilde{A}_{\mathcal{D}}(a,y) \leq \mathcal{D}_n(a,y) \text{ for all } a \in \mathsf{Support}(\mathcal{D}_n(\cdot \mid y))\right] \\ &\geq 1 - \frac{1}{2n^{\alpha}}. \end{aligned} \tag{12}$$

On the other hand, for the algorithm $A_{\mathcal{D}'}$, we have

$$\mathop{\mathbf{Pr}}_{(x,y)\sim\mathcal{D}_n,A_{\mathcal{D}'}}\left[\frac{\mathcal{D}'_n(y)}{c} \leq A_{\mathcal{D}'}(y) \leq \mathcal{D}'_n(y)\right] = \mathop{\mathbf{Pr}}_{y\sim\mathcal{D}'_n,A_{\mathcal{D}'}}\left[\frac{\mathcal{D}'_n(y)}{c} \leq A_{\mathcal{D}'}(y) \leq \mathcal{D}'_n(y)\right] \geq 1 - \frac{1}{n^{\alpha_0}}. \tag{13}$$

Again, since every element in the support of $\mathcal{D}'_n$ has probability mass at least $1/2^{t(n)}$, Equation (13) still holds if we replace $A_{\mathcal{D}'}$ with $\tilde{A}_{\mathcal{D}'}$ (which outputs the larger of $A_{\mathcal{D}'}$ and $1/2^{t(n)}$).

Fix any (x,y) in the support of $\mathcal{D}_n$ and any randomness for running B. Note that when both

$$\tilde{A}_{\mathcal{D}}(x,y) \geq \frac{\mathcal{D}_n(x,y)}{c} \quad \text{and} \quad \tilde{A}_{\mathcal{D}}(a,y) \leq \mathcal{D}_n(a,y) \text{ for all } a \in \mathsf{Support}(\mathcal{D}_n(\cdot \mid y))$$

and

$$\frac{\mathcal{D}'_n(y)}{c} \leq \tilde{A}_{\mathcal{D}'}(y) \leq \mathcal{D}'_n(y)$$

are true, we have for the algorithm B,

$$\frac{\mathcal{D}_n(x \mid y)}{c} := \frac{\mathcal{D}_n(x,y)}{c} \cdot \frac{1}{\mathcal{D}'_n(y)} \leq B(x,y) \leq \mathcal{D}_n(x,y) \cdot \frac{c}{\mathcal{D}'_n(y)} =: c \cdot \mathcal{D}_n(x \mid y),$$

and also

$$B(a,y) \leq \mathcal{D}_n(a,y) \cdot \frac{c}{\mathcal{D}'_n(y)} := c \cdot \mathcal{D}_n(a \mid y) \text{ for all } a \in \mathsf{Support}(\mathcal{D}_n(\cdot \mid y)).$$

By a union bound, the above happens with probability at least $1 - 1/n^{\alpha}$. □

We are now ready to show Lemma 14.

Proof of Lemma 14. Let $\{\mathcal{D}_n\}$ be any polynomial-time samplable distribution family, and let M be the machine such that $M(1^n, \mathcal{U}_{n^{c_{\mathcal{D}}}})$ is distributed according to $\mathcal{D}_n$.

Let α be an arbitrary constant, and let B be the algorithm from Lemma 16 with respect to $\{\mathcal{D}_n\}_n$. Also, we view B as a deterministic algorithm that takes n^{c_B} bits of randomness, where c_B is some constant.

Let us define the following polynomial-time computable function f.

> On input $(r_{\mathcal{D}}, r_B, w) \in \{0,1\}^{n^{c_{\mathcal{D}}}} \times \{0,1\}^{3n^{c_{\mathcal{D}}}} \times \{0,1\}^{n^{c_B}}$, we first run $M(1^n, r_{\mathcal{D}})$ to obtain some (x,y). We then run $B(x,y;r_B)$ to obtain $p_0 \in [0,1]$ and let p be the greatest power of two less than p_0/c, where c is the constant from Lemma 16. Finally, we interpret w as the encoding of a hash function H from a pairwise independent hash family, mapping n bits to $\log(1/p) + \alpha \log n$ bits[3] and output $(H(x), y, p, r_B, w)$.

[3]This can done by using the construction of pairwise independent hash family from Theorem 11. Note that we can obtain a hash function H from the family using only the first $n + \log(1/p) + \alpha$ bits of w.

For an output $z := (H(x), y, p, r_B, w)$ of f, we say that z is *good* if

1. $p \geq \mathcal{D}_n(x \mid y)/(2c^2)$, and
2. for every $(r_{\mathcal{D}}, r_B, w) \in f^{-1}(z)$, we have $M(1^n, r_{\mathcal{D}}) = (x, y)$. In other words, there does not exist (x', y) such that $x' \neq x$, but $M(1^n, r'_{\mathcal{D}}) = (x', y)$ and $f(r'_{\mathcal{D}}, r_B, w) = (H(x), y, p, r_B, w)$.

Claim 17. *We have*

$$\Pr_{r_{\mathcal{D}}, r_B, w}[f(r_{\mathcal{D}}, r_B, w) \text{ is good }] \geq 1 - \frac{2}{n^\alpha}.$$

Proof of Claim 17. Consider Lemma 16, we have

$$\Pr_{\substack{r_{\mathcal{D}}, r_B \\ (x,y):=M(1^n, r_{\mathcal{D}})}}\left[B(x, y; r_B) \geq \frac{\mathcal{D}_n(x \mid y)}{c} \text{ and } B(a, y; r_B) \leq c \cdot \mathcal{D}_n(a \mid y) \text{ for all } a \in \mathsf{Support}(\mathcal{D}_n(\cdot \mid y))\right] \geq 1 - \frac{1}{n^\alpha}.$$

Whenever this event happens, we have in the output of f,

$$\frac{\mathcal{D}_n(x \mid y)}{2c^2} \leq \frac{B(x, y; r_B)}{2c} \leq p < \frac{B(x, y; r_B)}{c} \leq \mathcal{D}_n(x \mid y),$$

and for all (x', y, r_B) that produces the same p, it must be the case that $\mathcal{D}_n(x' \mid y) > p$. We call such a pair (r_D, r_B) *successful* if the above holds.

We have

$$\begin{aligned}
&\Pr_{r_{\mathcal{D}}, r_B, w}[f(r_{\mathcal{D}}, r_B, w) \text{ is not good }] \\
&\leq \Pr_{r_{\mathcal{D}}, r_B, w}[f(r_{\mathcal{D}}, r_B, w) \text{ is not good } \mid (r_{\mathcal{D}}, r_B) \text{ is successful}] + \Pr_{r_{\mathcal{D}}, r_B}[(r_{\mathcal{D}}, r_B) \text{ is not successful}] \\
&\leq \Pr_{r_{\mathcal{D}}, r_B, w}[f(r_{\mathcal{D}}, r_B, w) \text{ is not good } \mid (r_{\mathcal{D}}, r_B) \text{ is successful}] + \frac{1}{n^\alpha}. \qquad (14)
\end{aligned}$$

Fix any (r_D, r_B) that is successful. Consider a random w, Let

$$(H(x), y, p, r_B, w) := f(r_{\mathcal{D}}, r_B, w),$$

where $(x, y) := M(1^n, r_{\mathcal{D}})$ and H is obtained from w (viewed as the encoding of a hash function from a pairwise independent hash family), mapping n bits to $\log(1/p) + \alpha \log n$ bits. Note that by Proposition 12, with probability at least $1 - 1/n^\alpha$, H is a "good" hash function that isolates x from the other elements in the set

$$S_{y,p} := \{a : \mathcal{D}_n(a \mid y) \geq p\}.$$

Now suppose $(H(x), y, p, r_B, w)$ is *not good*, this means there exists (x', y), which is obtained from $M(1^n, r'_{\mathcal{D}})$ for some $r'_{\mathcal{D}}$, such that together with r_B and w, produces $(H(x'), y, p, r_B, w)$ with $H(x') = H(x)$. Note that since (r_D, r_B) is successful, this implies that $\mathcal{D}(x' \mid y) > p$, which means H fails to isolates x within $S_{y,p}$. Therefore, we conclude that

$$\Pr_{r_{\mathcal{D}}, r_B, w}[f(r_{\mathcal{D}}, r_B, w) \text{ is not good } \mid (r_{\mathcal{D}}, r_B) \text{ is successful}] \leq 1/n^\alpha,$$

which, combined with Equation (14), completes the proof of the claim. □

Since infinitely-often one-way functions do not exist (hence infinitely-often weak one-way functions do not exist), for all α, there exists a PPT algorithm Invert that breaks f for all n with probability at least $1 - 1/n^\alpha$ (over a random input $(r_{\mathcal{D}}, r_B, w)$ to f and the internal randomness of Invert). Consider the PPT algorithm Rec that runs Invert, obtains some $(r_{\mathcal{D}}, r_B, w)$ and output x from $(x, y) := M(1^n, r_{\mathcal{D}})$. Note that if $(H(x), y, p, r_B, w)$ is good, then on a successful inversion, Rec will output x. Therefore, we have

$$\Pr_{\substack{(x,y)\sim\mathcal{D}_n \\ r_B,w,\mathsf{Rec}}}\left[\mathsf{Rec}(H(x), y, p, r_B, w) = x \text{ and } p \geq \frac{\mathcal{D}_n(x \mid y)}{2c^2}\right] \geq 1 - \frac{3}{n^\alpha},$$

where we take a union bound over the probability that the output is not good (which is less than $2/n^\alpha$) and that Invert fails (which is less than $1/n^\alpha$). By an averaging argument, with probability at least $1 - \sqrt{3/n^\alpha}$ over $(x, y) \sim \mathcal{D}_n$, we have

$$\Pr_{r_B,w,\mathsf{Rec}}\left[\mathsf{Rec}(H(x), y, p, r_B, w) = x \text{ and } p \geq \frac{\mathcal{D}_n(x \mid y)}{2c^2}\right] \geq 1 - \sqrt{\frac{3}{n^\alpha}} \geq \frac{2}{3}. \tag{15}$$

Note that Equation (15) implies an upper bound for $\mathsf{pK}^t(x \mid y)$, for some polynomial t. Indeed, for each such (x, y), if we sample r_B, w and the internal randomness Rec uniformly at random, then with high probability there exist $v := H(x)$ and p, where

$$|v| = \log p + \alpha \log n \leq \log \frac{1}{\mathcal{D}_n(x \mid y)} + \log 2c^2 + \alpha \log n,$$

and p can be encoded using $O(\log n)$ bits, such that, given y, one can print x, by simulating $\mathsf{Rec}(v, y, p, r_B, w)$. Finally, note that α was chosen to be arbitrary. Therefore, the above gives a strong average-case coding theorem as desired. □

3.2 Average-Case Conditional Coding implies Average-Case Language Compression

Lemma 18 (Item 2 ⇒ Item 4 in Theorem 13)**.** *Strong average-case conditional coding theorem implies strong average-case language compression.*

Proof. Assuming strong average-case conditional coding, we have that for some polynomial p_0,

$$\Pr_{(x,y)\sim\mathcal{D}_n}\left[\mathsf{pK}^{p_0(n)}(x \mid y) \leq \log \frac{1}{\mathcal{D}_n(x \mid y)} + \log p_0(n)\right] \geq 1 - \frac{1}{2q(n)}.$$

Also, by Lemma 9, there exists a polynomial p_1 such that

$$\Pr_{(x,y)\sim\mathcal{D}_n}\left[\mathsf{K}(x \mid y) \geq \log \frac{1}{\mathcal{D}_n(x \mid y)} - \log p_1(n)\right] \geq 1 - \frac{1}{2q(n)}.$$

Therefore, with probability at least $1 - 1/q(n)$ over $(x, y) \sim \mathcal{D}_n$,

$$\mathsf{pK}^{p_0(n)}(x \mid y) \leq \mathsf{K}(x \mid y) + \log(p_0(n) \cdot p_1(n)). \tag{16}$$

Also, it is easy to see that for every $x, y \in \{0,1\}^n$,

$$x \in L_y \implies \mathsf{K}(x \mid y) \leq \log |L_y| + O(1).$$

Then whenever Equation (16) is true, we have

$$x \in L_y \implies \mathsf{pK}^{p_0(n)}(x \mid y) \leq \log |L_y| + \log(p_0(n) \cdot p_1(n)) + O(1),$$

which implies that there exists a polynomial p such that

$$x \in L_y \implies \mathsf{pK}^{p(n)}(x \mid y) \leq \log |L_y| + \log p(n),$$

as desired. □

Lemma 19 (Item 3 ⇒ Item 5 in Theorem 13)**.** *Weak average-case conditional coding theorem implies weak average-case language compression.*

Proof Sketch. The proof can be adapted from that of Lemma 18 in a straightforward manner. We omit the details here. □

3.3 Inverting OWFs from Weak Average-Case Language Compression

Lemma 20 (Item 5 ⇒ Item 1 in Theorem 13)**.** *If weak average-case language compression holds, then infinitely-often one-way functions do not exist.*

We need the following notion of a universal sampler.

Definition 21 (Universal Time-Bounded Sampler)**.** Let $n, t \in \mathbb{N}$ and $y \in \{0,1\}^*$. The universal sampler $\mathsf{USamp}(1^n, 1^t, y)$ does the follow.

1. Pick a uniformly random $k \sim [O(n)]$,
2. Pick a uniformly random $r \sim \{0,1\}^t$,
3. Pick a uniformly random $d \sim \{0,1\}^k$,
4. Outputs x which is the output of a universal oracle Turing machine (fixed in advance) U, on input d with an oracle to the bits of y and r (i.e. $U^{y,r}(d)$), running for t steps.

Note that USamp runs in polynomial time. The following proposition follows easily from the definitions of pK^t and USamp.

Proposition 22. *For every $n, t, \ell \in \mathbb{N}$, $x \in \{0,1\}^n$ and $y \in \{0,1\}^*$, if $\mathsf{pK}^t(x \mid y) \leq k$, then $\mathsf{USamp}(1^n, 1^t, y)$ outputs x with probability $\Omega\big(1/(n \cdot 2^k)\big)$, where USamp is the universal sampler defined in Definition 21.*

We now show Lemma 20.

Proof of Lemma 20. First of all, consider any polynomial-time computable set $L \subseteq \{\{0,1\}^n \times \{0,1\}^n\}_n$ and any polynomial-time samplable distribution $\mathcal{D}_n$ whose support is L. By weak average-case language compression, there exist a polynomial p and a constant $c > 0$ such that for all n, with probability at least $1/n^c$ over $(x, y) \sim \mathcal{D}_n$,

$$\mathsf{pK}^{p(n)}(x \mid y) \leq \log |L_y| + \log p(n), \tag{17}$$

Say y is *good* if with probability at least $1/(2n^c)$ when we sample x from $\mathcal{D}_n(\cdot \mid y)$, we have that Equation (17) holds for (x, y). By a simple counting argument, with probability at least $1/(2n^c)$ when (x, y) is sampled according to $\mathcal{D}_n$, y is good.

We work by contradiction. Suppose $f\colon \{0,1\}^n \to \{0,1\}^n$ is an infinitely-often one-way function. Let $\mathcal{D}_n$ be the distribution given as follows: sample $x \sim \{0,1\}^n$ uniformly at random and output $(x, f(x))$. Also, let L be the polynomial-time computable set $L := \{(x, f(x))\}_{x\in\{0,1\}^*}$.

Note that by construction, for $y \in \{0,1\}^n$ that is in the image of f, we have $L_y := f^{-1}(y)$. Also, $\mathcal{D}_n(\cdot \mid y)$ is uniformly distributed on L_y. By the above discussion, with probability at least $1/(2n^c)$ over $x \sim \{0,1\}^n$, $y := f(x)$ is good in the sense that for at least $1/(2n^c)$ of the $x' \in L_y$, it holds that

$$\mathsf{pK}^{p(n)}(x' \mid y) \le \log |L_y| + \log p(n), \tag{18}$$

Let $S_y \subseteq L_y$ be the set of x' such that Equation (18) holds. Note that if y is good,

$$|S_y| \ge \frac{|L_y|}{2n^c}.$$

Fix any good y. By Proposition 22 and Equation (18), $\mathsf{USamp}(1^n, 1^{p(n)}, y)$ outputs each $x' \in S_y$ with probability at least

$$\frac{1}{O(n \cdot p(n) \cdot |L_y|)}.$$

Hence the probability that $\mathsf{USamp}(1^n, 1^{p(n)}, y)$ outputs some $x' \in S_y$ is at least

$$|S_y| \cdot \frac{1}{O(n \cdot p(n) \cdot |L_y|)} \ge \frac{1}{O(p(n) \cdot n^c)}.$$

In other words, with probability at least $1/(2n^c)$ over $x \sim \{0,1\}^n$ (in which case $f(x)$ is good), $\mathsf{USamp}(1^n, 1^{p(n)}, f(x))$ outputs some pre-image of $f(x)$ with probability at least $1/O(p(n) \cdot n^c)$, which breaks the security of f. □

3.4 Average-Case Conditional Coding Implies Average-Case Symmetry of Information

Lemma 23 (Item 2 ⇒ Item 6 in Theorem 13)**.** *Strong average-case conditional coding implies strong average-case symmetry of information.*

Proof. Let $\{\mathcal{D}_n\}$ be any polynomial-time samplable distribution family, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$ and let t be any computable time bound such that $t(n) \ge p(n)$, where p is some sufficiently large polynomial specified later. We want to show that strong average-case symmetry of information holds for $\{\mathcal{D}_n\}$. Let $\mathcal{D}'_n$ be the marginal distribution of $\mathcal{D}_n$ on the second half of the output. Assuming strong average-case coding theorem, for every constant $c > 0$, there exists a polynomial p_0, such that for all n, with probability at least $1 - 1/(2n^c)$ over $(x, y) \sim \mathcal{D}_n$, we have

$$\begin{aligned}
\mathsf{pK}^{p_0(n)}(x \mid y) &\le \log \frac{1}{\mathcal{D}_n(x \mid y)} + \log p_0(n) \\
&= \log \frac{\mathcal{D}'_n(y)}{\mathcal{D}_n(x, y)} + \log p_0(n) \\
&= \log \frac{1}{\mathcal{D}_n(x, y)} - \log \frac{1}{\mathcal{D}'_n(y)} + \log p_0(n).
\end{aligned} \tag{19}$$

Now let $b > 0$ be a sufficiently large constant, we have

$$\begin{aligned}
&\mathop{\mathbf{Pr}}_{(x,y)\sim\mathcal{D}_n}\left[\mathsf{pK}^{t(n)}(x,y) < \log\frac{1}{\mathcal{D}_n(x,y)} - b\log n\right] \\
&\leq \mathop{\mathbf{Pr}}_{(x,y)\sim\mathcal{D}_n}\left[\mathsf{K}(x,y) < \left(\log\frac{1}{\mathcal{D}_n(x,y)} - b\log n\right) + O(\log n)\right] && \text{(by Lemma 8)} \\
&< \frac{1}{n^c}. && \text{(by Item 1 of Lemma 9)}
\end{aligned}$$

In other words, with probability least $1 - 1/n^c$ over $(x,y) \sim \mathcal{D}_n$, we have

$$\log\frac{1}{\mathcal{D}_n(x,y)} \leq \mathsf{pK}^{t(n)}(x,y) + b\log n. \tag{20}$$

Also note that since $\mathcal{D}_n$ is polynomial-time samplable, so is $\mathcal{D}'_n$. By the coding theorem (Theorem 7), there exists a polynomial p_1 such that

$$\mathsf{pK}^{p_1(n)}(y) \leq \log\frac{1}{\mathcal{D}'_n(y)} + \log p_1(n). \tag{21}$$

Substituting Equation (20) and Equation (21) into Equation (19), we get that with probability at least $1 - 1/n^c$ over $(x,y) \sim \mathcal{D}_n$, we have

$$\mathsf{pK}^{p_0(n)}(x \mid y) \leq \left(\mathsf{pK}^{t(n)}(x,y) + b\log n\right) - \left(\mathsf{pK}^{p_1(n)}(y) - \log p_1(n)\right) + \log p_0(n),$$

which implies

$$\mathsf{pK}^{t(n)}(x \mid y) \leq \mathsf{pK}^{t(n)}(x,y) - \mathsf{pK}^{t(n)}(y) + \log t(n),$$

as long as p is a sufficiently large polynomial. □

3.5 Average-Case Symmetry of Information Implies Average-Case Conditional Coding

Lemma 24 (Item 7 ⇒ Item 3 in Theorem 13)**.** *Weak average-case symmetry of information implies weak average-case conditional coding.*

Proof. The proof is similar to that of Lemma 23 but works "backwards". We present the details for completeness.

Let $\{\mathcal{D}_n\}$ be any polynomial-time samplable distribution family, where each $\mathcal{D}_n$ is supported over $\{0,1\}^n \times \{0,1\}^n$. We want to show that weak average-case conditional coding holds for $\{\mathcal{D}_n\}$. Let $\mathcal{D}'_n$ be the marginal distribution of $\mathcal{D}_n$ on the second half of the output, i.e., $\mathcal{D}'_n$ samples $(x,y) \sim \mathcal{D}_n$ and outputs y.

Assuming weak average-case symmetry of information, there exist a polynomial p_0 and a constant $c > 0$, such that for all polynomial t that is greater than p_0 and for all n, with probability at least $1/n^c$ over $(x,y) \sim \mathcal{D}_n$, we have

$$\mathsf{pK}^{t(n)}(x \mid y) \leq \mathsf{pK}^{t(n)}(x,y) - \mathsf{pK}^{t(n)}(y) + \log t(n) \tag{22}$$

On the one hand, by applying the coding theorem (Theorem 7) to $\mathcal{D}_n$, given that t is a sufficiently large polynomial,

$$\mathsf{pK}^{t(n)}(x,y) \le \log \frac{1}{\mathcal{D}_n(x,y)} + \log t(n). \tag{23}$$

On the other hand, let $b > 0$ be a sufficiently large constant, we have

$$\begin{aligned}
&\Pr_{(x,y)\sim\mathcal{D}_n}\left[\mathsf{pK}^{t(n)}(y) < \log \frac{1}{\mathcal{D}'_n(y)} - b\log n\right] \\
&\le \Pr_{y\sim\mathcal{D}'_n}\left[\mathsf{K}(y) < \left(\log \frac{1}{\mathcal{D}'_n(y)} - b\log n\right) + O(\log n\right] && \text{(by Lemma 8)} \\
&< \frac{1}{2n^c}. && \text{(by Item 1 of Lemma 9)}
\end{aligned}$$

In other words, with probability least $1 - 1/(2n^c)$ over $(x,y) \sim \mathcal{D}_n$, we have

$$\mathsf{pK}^{t(n)}(y) \ge \log \frac{1}{\mathcal{D}'_n(y)} - b\log n. \tag{24}$$

Substituting Equation (24) and Equation (23) into Equation (22), we get that with probability at least $1/(2n^c)$ over $(x,y) \sim \mathcal{D}_n$, we have

$$\begin{aligned}
\mathsf{pK}^{t(n)}(x \mid y) &\le \left(\log \frac{1}{\mathcal{D}_n(x,y)} + \log t(n)\right) - \left(\log \frac{1}{\mathcal{D}'_n(y)} - b\log n\right) + \log t(n) \\
&= \log \frac{\mathcal{D}'_n(y)}{\mathcal{D}_n(x,y)} + \log t(n) + b\log n + \log t(n) \\
&= \log \frac{1}{\mathcal{D}_n(x \mid y)} + (b+2)\cdot \log t(n),
\end{aligned}$$

which implies that there exists a polynomial p such that

$$\mathsf{pK}^{p(n)}(x \mid y) \le \log \frac{1}{\mathcal{D}_n(x \mid y)} + \log p(n),$$

as desired. □

4 One-Way Functions and Average-Case Symmetry of Information for $\mathsf{rK}^{\mathsf{quasipoly}}$

We present the characterization of the existence of a one-way function secure against quasi-polynomial-time adversaries by average-case symmetry of information for $\mathsf{rK}^{\mathsf{poly}}$.

Theorem 25. *The following are equivalent.*

1. *Infinitely-often polynomial-time-computable one-way functions secure against quasi-polynomial-time randomized algorithms do not exist.*
2. **(Approximation of K by $\mathsf{rK}^{\mathsf{quasipoly}}$)** *For every polynomial-time samplable distribution family $\{\mathcal{D}_n\}_n$, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and for every quasi-polynomial q, there exists a quasi-polynomial p such that for all n,*

$$\Pr_{(x,y)\sim\mathcal{D}_n}\left[\mathsf{rK}^{p(n)}(x \mid y) \le \mathsf{K}(x \mid y) + \log p(n)\right] \ge 1 - \frac{1}{q(n)}.$$

3. **(Approximation of K by $\mathsf{rK}^{\mathsf{quasipoly}}$ with an Efficient Encoder)** *In addition to Item 2, there exists a quasi-polynomial-time randomized algorithm M that takes $(x,y) \sim \mathcal{D}_n$ as input and, with probability $1 - 1/q(n)$ over the choice of (x,y) and the internal randomness of M, outputs the description of a randomized program of length $\mathsf{K}(x \mid y) + \log p(n)$ that takes y as input and outputs x in time $p(n)$.*

4. **(Average-Case Symmetry of Information)** *For every polynomial-time samplable distribution family $\{\mathcal{D}_n\}_n$, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and for every quasi-polynomial q, there exists a quasi-polynomial p such that for all computable time bounds t with $t(n) \geq p(n)$ and for all n,*

$$\Pr_{(x,y)\sim\mathcal{D}_n}\left[\mathsf{rK}^{t(n)}(x,y) \geq \mathsf{rK}^{t(n)}(x \mid y) + \mathsf{rK}^{t(n)}(y) - \log t(n)\right] \geq 1 - \frac{1}{q(n)}.$$

5. **(Average-Case Symmetry of Information with an Efficient Encoder)** *In addition to Item 4, there exists a quasi-polynomial-time randomized algorithm M that takes $(x,y) \sim \mathcal{D}_n$ and, with probability $1 - 1/q(n)$ over the choice of (x,y) and the internal randomness of M, outputs the description of a randomized program of length $\mathsf{rK}^{t(n)}(x,y) - \mathsf{rK}^{t(n)}(y) + \log t(n)$ that takes y as input and outputs x in time $p(n)$.*

Proof. Item 1 $\Rightarrow$ Item 3 is due to Lemma 28.

Item 3 $\Rightarrow$ Item 2 and Item 5 $\Rightarrow$ Item 4 are obvious.

Item 2 $\Rightarrow$ Item 4 and Item 3 $\Rightarrow$ Item 5 can be proved as follows. Using the assumption, with high probability over $(x,y) \sim \mathcal{D}_n$, we have

$$\mathsf{rK}^{p(n)}(x \mid y) \leq \mathsf{K}(x \mid y) + \log p(n).$$

and

$$\mathsf{rK}^{p(n)}(y) \leq \mathsf{K}(y) + \log p(n).$$

By summing the two inequalities, we obtain

$$\begin{aligned}\mathsf{rK}^{p(n)}(x \mid y) + \mathsf{rK}^{p(n)}(y) &\leq \mathsf{K}(x \mid y) + \mathsf{K}(y) + 2\log p(n)\\ &\leq \mathsf{K}(x,y) + O(\log p(n))\\ &\leq \mathsf{rK}^{t(n)}(x,y) + O(\log p(n)),\end{aligned}$$

where the second inequality holds by the symmetry of information for Kolmogorov complexity. Noting that $\mathsf{rK}^{p(n)}(x \mid y) + \mathsf{rK}^{p(n)}(y) \geq \mathsf{rK}^{t(n)}(x \mid y) + \mathsf{rK}^{t(n)}(y)$, the claim follows.

Item 4 $\Rightarrow$ Item 1 is due to Lemma 29. □

4.1 Approximating K by $\mathsf{rK}^{\mathsf{quasipoly}}$ from Inverting Quasi-Polynomial OWFs

We use the following pseudorandom generator construction.

Lemma 26. *There exists a polynomial p such that, for all sufficiently large $n, m, t \in \mathbb{N}$ such that $m \leq 2n$ and $t \geq n$, there exists a "pseudorandom generator construction"*

$$G_m \colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$$

such that for every $x \in \{0,1\}^n$ *and any function* $D\colon \{0,1\}^m \times \{0,1\}^t \to \{0,1\}$*, if*

$$\left| \Pr_{\substack{z\sim\{0,1\}^d \\ w'\sim\{0,1\}^t}} \left[D(G_m(x;z);w')=1\right] - \Pr_{\substack{w\sim\{0,1\}^m \\ w'\sim\{0,1\}^t}} \left[D(w;w')=1\right] \right| \geq \frac{1}{m},$$

then

$$\mathsf{rK}^{p(t),D}(x) \leq m + O(\log^3 n).$$

Here, $d = O(\log^3 n)$ *and* G_m *can be computed in time* $\mathsf{poly}(n)$*. Moreover, there exists a* D*-oracle randomized polynomial-time algorithm* α_m *that takes* x *as input and outputs the description of a* D*-oracle randomized program of length* $m + O(\log^3 n)$ *that prints* x *in time* $p(t)$*.*

Proof Sketch. As in [Hir22b], we use the pseudorandom generator construction of [RRV02] (where a list-decodable error-correcting code is used in place of an error-correcting code). Then we obtain a triple (G_m, A_m, R) such that

$$\begin{aligned} G_m &\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m, \\ A_m &\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m, \\ R &\colon \{0,1\}^m \times \{0,1\}^d \times \{0,1\}^r \to \{0,1\}^n \end{aligned}$$

with the following properties: For every $x \in \{0,1\}^n$ and any $D\colon \{0,1\}^m \times \{0,1\}^t \to \{0,1\}$ such that

$$\left| \Pr_{\substack{z\sim\{0,1\}^d \\ w'\sim\{0,1\}^t}} \left[D(G(x;z);w')=1\right] - \Pr_{\substack{w\sim\{0,1\}^m \\ w'\sim\{0,1\}^t}} \left[D(w;w')=1\right] \right| \geq \frac{1}{m},$$

it holds that

$$\Pr_{\substack{w\sim\{0,1\}^r \\ z\sim\{0,1\}^d}} \left[x = R^D(A_m(x,z),z,w)\right] \geq \frac{1}{\mathsf{poly}(m)} =: \epsilon.$$

Here, we have $d = O(\log^3 n)$ and $r = O(t)$. Moreover, G_m and A_m can be computed in time $\mathsf{poly}(n)$ and R^D can be computed in time $\mathsf{poly}(n)$ with oracle access to D. By an averaging argument, with probability at least $\epsilon/2$ over a random choice of $z \sim \{0,1\}^d$, it holds that

$$\Pr_{w\sim\{0,1\}^r} \left[x = R^D(A_m(x,z),z,w)\right] \geq \frac{\epsilon}{2}. \tag{25}$$

For any z that satisfies this, we may construct a randomized D-oracle program M^D that prints x with probability $\epsilon/2$ as follows: M^D takes $A_m(x,z)$ and z as hard-wired input and simulates $R^D(A_m(x,z),z,w)$ for a random choice of $w \sim \{0,1\}^r$. By Equation (25), this algorithm M^D witnesses

$$\mathsf{rK}^{\mathsf{poly}(t),D}_{\epsilon/2}(x) \leq m + O(\log^3 n)$$

for some polynomial poly. By using the randomness-efficient coding of [LO21], the success probability $\frac{\epsilon}{2}$ can be amplified to $\frac{2}{3}$ with additional $O(\log(1/\epsilon))$ bits of information; thus, we get

$$\mathsf{rK}^{p(t),D}(x) \leq m + O(\log^3 n)$$

for some polynomial p.

To see the "moreover" part, observe that whether z satisfies Equation (25) or not can be approximately checked in polynomial time with oracle access to D by randomly sampling w. For such z, the algorithm α_m outputs the description of M^D whose success probability is amplified by [LO21]. □

We also use a conditional version of [IRS21]. For simplicity, we state it for polynomial-time algorithms, but the analogous result holds for quasi-polynomial-time algorithms.

Lemma 27. *If there exists no infinitely-often one-way function, then for every polynomial-time samplable distribution family $\{\mathcal{D}_n\}_n$, there exist a randomized polynomial-time algorithm $A_{\mathcal{D}}$ and a polynomial p such that for all $n \in \mathbb{N}$,*

$$\Pr_{(x,y)\sim\mathcal{D}_n}[\mathsf{K}(x \mid y) \le A(x,y) \le \mathsf{K}(x \mid y) + \log p(n)] \ge 1 - \frac{1}{p(n)}.$$

Proof. By [IRS21], for every polynomial-time samplable distribution family $\{\mathcal{E}_n\}_n$, there exist a randomized polynomial-time algorithm $A_{\mathcal{E}}$ and a polynomial p such that for all $n \in \mathbb{N}$,

$$\Pr_{x\sim\mathcal{E}_n}[\mathsf{K}(x) \le A(x) \le \mathsf{K}(x) + \log p(n)] \ge 1 - \frac{1}{p(n)}.$$

For a given distribution $\mathcal{D}$, let $\mathcal{E}$ be the uniform mixture of the two distributions, (x,y) and y for $(x,y) \sim \mathcal{E}$. Define a new algorithm A' to be one that takes (x,y) as input and outputs $A_{\mathcal{E}}(x,y) - A_{\mathcal{E}}(y)$.

We prove that A' approximates $\mathsf{K}(x \mid y)$ on average. Fix $n \in \mathbb{N}$ and let $\epsilon := 1/p(n)$. By the definition of $\mathcal{E}$, with probability at least $1 - 2\epsilon$ over a random choice of $(x,y) \sim \mathcal{D}$, it holds that

$$\mathsf{K}(x,y) \le A_{\mathcal{E}}(x,y) \le \mathsf{K}(x,y) + O(\log n)$$

and

$$\mathsf{K}(y) \le A_{\mathcal{E}}(y) \le \mathsf{K}(y) + O(\log n)$$

Under this event, we also have

$$\mathsf{K}(x,y) - \mathsf{K}(y) - O(\log n) \le A'(x,y) = A_{\mathcal{E}}(x,y) - A_{\mathcal{E}}(y) \le \mathsf{K}(x,y) - \mathsf{K}(y) + O(\log n).$$

By the symmetry of information for Kolmogorov complexity, we have

$$\mathsf{K}(x \mid y) - O(\log n) \le \mathsf{K}(x,y) - \mathsf{K}(y) \le \mathsf{K}(x \mid y) + O(\log n).$$

The claim follows from these two inequalities. □

Lemma 28 (Item 1 ⇒ Item 3 in Theorem 25)**.** *If infinitely-often one-way functions secure against quasi-polynomial randomized algorithms do not exist, then for every polynomial-time samplable distribution family $\{\mathcal{D}_n\}_n$, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and for every quasi-polynomial q, there exists a quasi-polynomial p such that for all n,*

$$\Pr_{(x,y)\sim\mathcal{D}_n}\left[\mathsf{rK}^{p(n)}(x \mid y) \le \mathsf{K}(x \mid y) + \log p(n)\right] \ge 1 - \frac{1}{q(n)}.$$

Moreover, there exists a quasi-polynomial-time randomized algorithm M that takes $(x,y) \sim \mathcal{D}_n$ as input and, with probability $1 - 1/q(n)$ over the choice of (x,y) and the internal randomness of M, outputs the description of a randomized program of length $\mathsf{K}(x \mid y) + \log p(n)$ that takes y as input and outputs x in time $p(n)$.

Proof. By Lemma 27, for every polynomial-samplable distribution $\mathcal{E}$ and every quasi-polynomial q, there exist a quasi-polynomial p and a randomized quasi-polynomial-time algorithm $A_{\mathcal{E}}$ such that for all $n \in \mathbb{N}$,

$$\mathop{\mathbf{Pr}}_{(x,y)\sim\mathcal{E}}[\mathsf{K}(x \mid y) \le A_{\mathcal{E}}(x,y) \le \mathsf{K}(x \mid y) + \log p(n)] \ge 1 - \frac{1}{q(n)}.$$

Consider the distribution $\mathcal{F}$ of (x, y, k, z, w), where $(x, y) \sim \mathcal{D}$, $k \sim [2n]$, and $z \sim \{0,1\}^d$, and $w \sim \{0,1\}^k$. Here, $d = O(\log^3 n)$ is (an upper bound of) the seed length of G_k from Lemma 26. Let $\mathcal{E}_1$ be the distribution of $(G_k(x; z), y)$ for $(x, y, k, z, w) \sim \mathcal{F}$. Similarly, let $\mathcal{E}_2$ be the distribution of (w, y) for $(x, y, k, z, w) \sim \mathcal{F}$. We define $\mathcal{E}$ to be the uniform mixture of $\mathcal{E}_1$ and $\mathcal{E}_2$.

Since the pseudorandom generator construction G_k of Lemma 26 is computable, we have

$$\mathsf{K}(G_k(x; z) \mid y) \le \mathsf{K}(x \mid y) + |z| + O(\log n) \le \mathsf{K}(x \mid y) + \log p(n),$$

where the last inequality holds by choosing a large enough quasi-polynomial $p \ge q$. Therefore, we obtain

$$\mathop{\mathbf{Pr}}_{(x,y,k,z,w)\sim\mathcal{F}}[A_{\mathcal{E}}(G_k(x; z), y) \le \mathsf{K}(x \mid y) + 2\log p(n)] \ge 1 - \frac{2}{q(n)},$$

where the factor 2 in $\frac{2}{q(n)}$ comes from the fact that $\mathcal{E}_1$ is identical to $\mathcal{E}$ with probability $\frac{1}{2}$. By a union bound over all $k^* \in [2n]$, we obtain

$$\mathop{\mathbf{Pr}}_{(x,y,k,z,w)\sim\mathcal{F}}[\forall k^* \in [2n],\ A_{\mathcal{E}}(G_{k^*}(x; z), y) \le \mathsf{K}(x \mid y) + 2\log p(n)] \ge 1 - \frac{4n}{q(n)}.$$

In particular, for a given (x, y), let $k^* = k^*(x, y) := \mathsf{K}(x \mid y) + 3\log p(n)$. Then, we have

$$\mathop{\mathbf{Pr}}_{(x,y,k,z,w)\sim\mathcal{F}}[A_{\mathcal{E}}(G_{k^*}(x; z), y) \le k^* - \log p(n)] \ge 1 - \frac{4n}{q(n)}. \tag{26}$$

By a simple counting argument, we have $\mathsf{K}(w \mid y) \ge |w| - \log q(n) > |w| - \log p(n)$ with probability at least $1 - \frac{1}{q(n)}$ over a choice of $w \sim \{0,1\}^k$. Thus, we obtain

$$\mathop{\mathbf{Pr}}_{(x,y,k,z,w)\sim\mathcal{F}}[A_{\mathcal{E}}(w, y) > k^* - \log p(n)] \ge 1 - \frac{4n}{q(n)}. \tag{27}$$

Let $D_{k,y} \colon \{0,1\}^k \times \{0,1\}^t \to \{0,1\}$ be the function such that $D_{k,y}(w; r) := 1$ if and only if $A_{\mathcal{E}}(w, y; r) \le k - \log p(n)$, where r denotes the internal randomness of $A_{\mathcal{E}}$. It follows from Equations (26) and (27) that

$$\mathbf{Pr}\left[D_{k^*,y}(G_{k^*}(x; z); r) = 1\right] - \mathbf{Pr}\left[D_{k^*,y}(w; r) = 1\right] \ge 1 - \frac{8n}{q(n)},$$

where the probabilities are over $(x, y, k, z, w) \sim \mathcal{F}$ and $r \sim \{0,1\}^t$. By an averaging argument, with probability at least $1 - \frac{16}{q(n)}$ over a random choice of $(x, y) \sim \mathcal{D}$, it holds that

$$\mathbf{Pr}\left[D_{k^*,y}(G_{k^*}(x; z); r) = 1\right] - \mathbf{Pr}\left[D_{k^*,y}(w; r) = 1\right] \ge \frac{1}{2}.$$

Under this event, applying Lemma 26 to $D_{k^*,y}$, we obtain

$$\mathsf{rK}^{p'(n)}(x \mid y) \le k^* + O(\log^3 n)$$

for some quasi-polynomial p'. Since $k^* = \mathsf{K}(x,y) + 3\log p(n)$, it follows that

$$\Pr_{(x,y)\sim\mathcal{D}}\left[\mathsf{rK}^{p'(n)}(x \mid y) \leq \mathsf{K}(x \mid y) + O(\log p(n))\right] \geq 1 - \frac{16n}{q(n)}.$$

The "moreover" part follows from the "moreover" part of Lemma 26. □

4.2 Inverting Quasi-Polynomial OWFs from Average-Case Symmetry of Information for $\mathsf{rK}^{\mathsf{quasipoly}}$

Lemma 29 (Item 4 ⇒ Item 1 in Theorem 25)**.** *If Average-Case Symmetry of Information for $\mathsf{rK}^{\mathsf{quasipoly}}$ (Item 4 in Theorem 25) holds, then infinitely-often polynomial-time-computable one-way functions secure against quasi-polynomial-time randomized algorithms do not exist.*

Proof. Let f be an arbitrarily polynomial-time computable length-preserving function. We show that if Item 4 holds, then there exists a quasi-polynomial-time adversary A that inverts f with probability at least $1 - 1/n$ over the choice of input x for f and randomness for A.

We consider a polynomial-time samplable distribution $\mathcal{D} = \{\mathcal{D}_n\}$, where for each $n \in \mathbb{N}$, $\mathcal{D}_n$ is a distribution of $(x, f(x))$ for $x \sim \{0,1\}^n$. Under the assumption that Item 4 holds, it holds that, for any sufficiently large quasi-polynomial p and for any $n \in \mathbb{N}$,

$$\Pr_{(x,f(x))\sim\mathcal{D}_n}\left[\mathsf{rK}^{p(n)}(x, f(x)) \geq \mathsf{rK}^{p(n)}(x \mid f(x)) + \mathsf{rK}^{p(n)}(f(x)) - \log p(n)\right] \geq 1 - \frac{1}{8n}.$$

Since f is polynomial-time computable, we can select a sufficiently large quasi-polynomial p such that $\mathsf{rK}^{p(n)}(x, f(x)) \leq \mathsf{rK}^{p(n)/2}(x)$ for every $n \in \mathbb{N}$ and every $x \in \{0,1\}^n$. Therefore,

$$\Pr_{(x,f(x))\sim\mathcal{D}_n}\left[\mathsf{rK}^{p(n)/2}(x) \geq \mathsf{rK}^{p(n)}(x \mid f(x)) + \mathsf{rK}^{p(n)}(f(x)) - \log p(n)\right] \geq 1 - \frac{1}{8n}.$$

For every $t \in \mathbb{N}$ and every $x, y \in \{0,1\}^*$, let $\mathrm{rQ}^t(x \mid y)$ be a probability that $U(\pi, r \mid y)$ outputs x in t steps for $i \sim [t]$, $\pi \sim \{0,1\}^i$, and $r \sim \{0,1\}^t$. By the definition of rK^t, we have that $\mathrm{rQ}^t(x \mid y) \geq (2/3) \cdot t^{-1} 2^{-\mathsf{rK}^t(x|y)}$ for every $t \in \mathbb{N}$ and every $x, y \in \{0,1\}^*$.

Therefore, we have

$$\Pr_{(x,f(x))\sim\mathcal{D}_n}\left[\mathrm{rQ}^{p(n)}(x \mid f(x)) \geq \frac{2^{\mathsf{rK}^{p(n)}(f(x))}}{(3/2)\cdot p(n)^2 \cdot 2^{\mathsf{rK}^{p(n)/2}(x)}}\right] \geq 1 - \frac{1}{8n}. \tag{28}$$

We first show Item 1 by assuming the following claim.

Claim 30. *There exists a quasi-polynomial q such that for every $n \in \mathbb{N}$*

$$\Pr_{x\sim\{0,1\}^n}\left[\frac{2^{\mathsf{rK}^{p(n)}(f(x))}}{2^{\mathsf{rK}^{p(n)/2}(x)}} \geq \frac{1}{p(n)q(n)\cdot|f^{-1}(f(x))|}\right] \geq 1 - \frac{1}{8n}.$$

Let p' be a quasi-polynomial defined as $p'(n) = (3/2)p(n)^3 q(n)$. By inequality (28), Claim 30, and the union bound, we have that, for every $n \in \mathbb{N}$,

$$\Pr_{x\sim\{0,1\}^n}\left[\mathrm{rQ}^{p(n)}(x \mid f(x)) \geq \frac{1}{p'(n)\cdot|f^{-1}(f(x))|}\right] \geq 1 - \frac{1}{4n}.$$

Let $G \subseteq \{0,1\}^n$ be the subset of $x \in \{0,1\}^n$ that satisfies the event in the expression above, i.e., $x \in G$ iff $\mathrm{rQ}^{p(n)}(x \mid f(x)) \geq 1/(p'(n)|f^{-1}(f(x))|)$ holds. Let $B := \{0,1\}^n \setminus G$. Then, it trivially holds that $\mathbf{Pr}_{x' \sim \{0,1\}^n}[x' \in B] \leq 1/(4n)$.

We can assume that a random choice $x' \sim \{0,1\}^n$ is performed as the following two-step procedure: (i) $x \sim \{0,1\}^n$ and (ii) $x' \sim |f^{-1}(f(x))|$. Then, by Markov's inequality,

$$\mathop{\mathbf{Pr}}_{x \sim \{0,1\}^n}\left[\mathop{\mathbf{Pr}}_{x' \sim |f^{-1}(f(x))|}\left[x' \in B\right] > \frac{1}{2}\right] \leq \frac{1}{2n}.$$

Let $G' \subseteq \{0,1\}^n$ be the subset of $x \in \{0,1\}^n$ that does not satisfy the event in the expression above, i.e., $x \in G'$ iff $\mathbf{Pr}_{x' \sim |f^{-1}(f(x))|}[x' \in B] \leq 1/2$ iff $\mathbf{Pr}_{x' \sim |f^{-1}(f(x))|}[x' \in G] \geq 1/2$. Then, it holds that $\mathbf{Pr}_{x \sim \{0,1\}^n}[x \in G'] \geq 1 - 1/(2n)$.

Furthermore, for any $x \in G'$, we have

$$\begin{aligned}
\sum_{x' \in f^{-1}(f(x))} \mathrm{rQ}^{p(n)}(x' \mid f(x)) &\geq \sum_{x' \in G \cap f^{-1}(f(x))} \mathrm{rQ}^{p(n)}(x' \mid f(x)) \\
&\geq \sum_{x' \in G \cap f^{-1}(f(x))} \frac{1}{p'(n) \cdot |f^{-1}(f(x))|} \\
&= \frac{|G \cap f^{-1}(f(x))|}{p'(n) \cdot |f^{-1}(f(x))|} \\
&= \frac{1}{p'(n)} \mathop{\mathbf{Pr}}_{x' \sim |f^{-1}(f(x))|}\left[x' \in G\right] \\
&\geq \frac{1}{2p'(n)},
\end{aligned}$$

where the last inequality holds since $x \in G'$. Thus,

$$\mathop{\mathbf{Pr}}_{x \sim \{0,1\}^n}\left[\sum_{x' \in f^{-1}(f(x))} \mathrm{rQ}^{p(n)}(x' \mid f(x)) \geq \frac{1}{2p'(n)}\right] \geq \mathop{\mathbf{Pr}}_{x \sim \{0,1\}^n}[x \in G'] \geq 1 - \frac{1}{2n}. \tag{29}$$

Based on the inequality above, we construct the quasi-polynomial-time adversary A that inverts f as follows: On input 1^n and y, where $y = f(x)$ for $x \sim \{0,1\}^n$, the adversary A executes the universal Turing machine $U(\pi, r \mid y)$ in $p(n)$ steps for $i \sim [t]$, $\pi \sim \{0,1\}^i$, and $r \sim \{0,1\}^t$, and obtains $x' \sim U(\pi, r \mid y)$ repeatedly. For each obtained sample x', the adversary A checks whether $f(x') = y$, and if so, A outputs the inverse element x'. If A cannot obtain any inverse element after $\mathsf{poly}(n, p'(n))$ sampling processes, then A outputs $\perp$.

We discuss the correctness of A. Suppose that a hidden random seed $x \in \{0,1\}^n$ satisfies the condition in inequality (29) (this event occurs with probability at least $1 - 1/(2n)$). Then, by inequality (29), for each sampling process according to $U(x, r \mid f(x))$, the probability that the sample x' satisfies that $f(x') = f(x)$ is

$$\sum_{x' \in f^{-1}(f(x))} \mathrm{rQ}^{p(n)}(x' \mid f(x)) \geq \frac{1}{2p'(n)}.$$

Therefore, by repeating the sampling process $\mathsf{poly}(n, p'(n))$ times, we can amplify the success probability that some inverse element is sampled to $1 - 1/(2n)$. By the union bound, the success

probability of $A(1^n, y)$ is at least $1 - 1/n$ over the choice of $y = f(x)$ and the randomness of A (for sampling). It is easily verified that A halts in polynomial time in $n, p(n)$, and $p'(n)$, which is a quasi-polynomial in n.

The remaining is the proof of Claim 30.

Proof of Claim 30. Fix $n \in \mathbb{N}$ arbitrarily. For every $x \in \{0,1\}^n$, by considering the trivial program that outputs the embedded string x, it holds that

$$\mathsf{rK}^{p(n)/2}(x) \leq n + O(1).$$

Therefore,

$$2^{\mathsf{rK}^{p(n)/2}(x)}/O(1) \leq 2^n. \tag{30}$$

Now, we evaluate $2^{\mathsf{rK}^{p(n)}(x)}$. Let $\mathcal{D}'_n$ be the distribution of $f(x)$ for $x \sim \{0,1\}^n$, i.e., the marginal distribution of the second half element of $\mathcal{D}_n$. It is easy to verify that for every $y \in \mathrm{Im} f$,

$$\mathcal{D}'_n(y) = \frac{|f^{-1}(y)|}{2^n}.$$

By simple calculations,

$$\begin{aligned}
\mathop{\mathbf{E}}_{y\sim\mathcal{D}'_n}\left[\frac{2^{-\mathsf{rK}^{p(n)}(y)}}{\mathcal{D}'_n(y)}\right] &= \sum_{y\in\mathsf{Support}(\mathcal{D}'_n)} 2^{-\mathsf{rK}^{p(n)}(y)} \\
&\leq (2/3)^{-1}p(n)\cdot \sum_{y\in\mathsf{Support}(\mathcal{D}'_n)} \mathrm{rQ}^{p(n)}(y) \\
&\leq 2p(n).
\end{aligned}$$

By Markov's inequality,

$$\mathop{\mathbf{Pr}}_{x\sim\{0,1\}^n}\left[\frac{2^{-\mathsf{rK}^{p(n)}(f(x))}}{\mathcal{D}'_n(f(x))} \leq 16p(n)n\right] = \mathop{\mathbf{Pr}}_{y\sim\mathcal{D}'_n}\left[\frac{2^{-\mathsf{rK}^{p(n)}(y)}}{\mathcal{D}'_n(y)} \leq 16p(n)n\right] \geq 1 - \frac{1}{8n}. \tag{31}$$

For every $x \in \{0,1\}^n$ satisfying inequality (31), we have

$$\begin{aligned}
2^{\mathsf{rK}^{p(n)}(f(x))} &\geq \frac{1}{16p(n)n\cdot\mathcal{D}'_n(f(x))} \\
&= \frac{2^n}{16p(n)n\cdot|f^{-1}(f(x))|} \\
&\geq \frac{2^{\mathsf{rK}^{p(n)/2}(x)}}{O(1)\cdot 16p(n)n\cdot|f^{-1}(f(x))|},
\end{aligned}$$

where the last inequality follows from inequality (30).

By rearranging the above, we conclude that

$$\mathop{\mathbf{Pr}}_{x\sim\{0,1\}^n}\left[\frac{2^{\mathsf{rK}^{p(n)}(f(x))}}{2^{\mathsf{rK}^{p(n)/2}(x)}} \geq \frac{1}{O(1)\cdot 16p(n)n\cdot|f^{-1}(f(x))|}\right] \geq 1 - \frac{1}{8n},$$

as desired. □

This completes the proof of Lemma 29. □

5 DistNP vs HeurBPP, Independent Average-Case Conditional Coding and Language Compression

In this section, we show the following relationships between the average-case easiness of NP and the *independent* variants of average-case conditional coding, language compression, and symmetry of information.

Theorem 31. *The following are equivalent.*

1. $\mathsf{DistNP} \subseteq \mathsf{HeurBPP}$.
2. **(Independent Average-Case Conditional Coding)** *For every samplable distribution families $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ and $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and each $\mathcal{C}_n$ is over $\mathsf{Support}(\mathcal{D}_n^{(2)})$, and for every polynomial q, there exists a polynomial p such that for all large enough $n \in \mathbb{N}$,*
$$\Pr_{y\sim\mathcal{C}_n, x\sim\mathcal{D}_n(\cdot|y)}\left[\mathsf{pK}^{p(n)}(x \mid y) \le \log\frac{1}{\mathcal{D}_n(x \mid y)} + \log p(n)\right] \ge 1 - \frac{1}{q(n)}.$$
3. **(Independent Average-Case Language Compression)** *For every computable set $L \subseteq \{\{0,1\}^n \times \{0,1\}^n\}_n$, for every samplable distribution families $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ and $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and each $\mathcal{C}_n$ is over $\mathsf{Support}(\mathcal{D}_n^{(2)})$, and for every polynomial q, there exists a polynomial p such that for all n,*
$$\Pr_{y\sim\mathcal{C}_n, x\sim\mathcal{D}_n(\cdot|y)}\left[x \in L_y \implies \mathsf{pK}^{p(n)}(x \mid y) \le \log|L_y| + \log p(n)\right] \ge 1 - \frac{1}{q(n)}.$$
4. **(Conditional Extrapolation)** *For every samplable distribution families $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ and $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and each $\mathcal{C}_n$ is over $\mathsf{Support}(\mathcal{D}_n^{(2)})$, there exists a probabilistic polynomial-time algorithm $\mathsf{CondExt}$ such that for all $n, \varepsilon^{-1}, \delta^{-1} \in \mathbb{N}$,*
$$\Pr_{y\sim\mathcal{C}_n}\left[\mathsf{L}_1\left(\mathsf{CondExt}(y; 1^{\varepsilon^{-1}}, 1^{\delta^{-1}}), \mathcal{D}_n(\cdot \mid y)\right) \le \varepsilon\right] \ge 1 - \delta.$$

Theorem 32. *If $\mathsf{DistNP} \subseteq \mathsf{HeurBPP}$ holds, then the following holds:*

5. **(Independent Average-Case Symmetry of Information)** *For every samplable distribution families $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ and $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and each $\mathcal{C}_n$ is over $\mathsf{Support}(\mathcal{D}_n^{(2)})$, and for every polynomial q, there exists a polynomial p such that for all enough large $n \in \mathbb{N}$ and for all computable time bound $t\colon \mathbb{N} \to \mathbb{N}$ with $t(n) \ge p(n)$ and for all n,*
$$\Pr_{y\sim\mathcal{C}_n, x\sim\mathcal{D}_n(\cdot|y)}\left[\mathsf{pK}^{t(n)}(x, y) \ge \mathsf{pK}^{t(n)}(x \mid y) + \mathsf{pK}^{t(n)}(y) - \log t(n)\right] \ge 1 - \frac{1}{q(n)}.$$

The proofs of Theorems 31 and 32 are presented in the subsequent sections. In Section 5.1, we show Item 1 ⇒ Item 4. In Section 5.2, we show Item 4 ⇒ Items 2, 3, and 5. In Section 5.3, we show Item 2 ⇒ Item 1 and Item 3 ⇒ Item 1.

5.1 Conditional Extrapolation from Average-Case Easiness of NP

In this section, we extend the well-known reduction from inverting functions to *distributionally* inverting functions [IL89] to the case of average-case easiness of NP and show the following lemma.

Lemma 33 (Item 1 $\Rightarrow$ Item 4 in Theorem 31)**.** *If* $\mathsf{DistNP} \subseteq \mathsf{HeurBPP}$ *holds, then for every samplable distribution families* $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ *and* $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$*, where each* $\mathcal{D}_n$ *is over* $\{0,1\}^n \times \{0,1\}^n$*, and each* $\mathcal{C}_n$ *is over* $\mathsf{Support}(\mathcal{D}_n^{(2)})$*, there exists a probabilistic polynomial-time algorithm* $\mathsf{CondExt}$ *such that for all* $n, \varepsilon^{-1}, \delta^{-1} \in \mathbb{N}$*,*

$$\Pr_{y\sim\mathcal{C}_n}\left[\mathsf{L}_1\left(\mathsf{CondExt}(y; 1^{\varepsilon^{-1}}, 1^{\delta^{-1}}), \mathcal{D}_n(\cdot \mid y)\right) \le \varepsilon\right] \ge 1-\delta.$$

To show Lemma 33, we use the following search-to-decision reduction.

Theorem 34 (Search-to-Decision Reduction [BCGL92])**.** *If* $\mathsf{DistNP} \subseteq \mathsf{HeurBPP}$*, then for every* $(L, \mathcal{D}) \in \mathsf{DistNP}$*, where* L *is determined by an* NP *relation* $R_L \subseteq \{0,1\}^* \times \{0,1\}^*$*, there exists a probabilistic polynomial-time algorithm* M *such that for every* $n, \varepsilon^{-1} \in \mathbb{N}$*,*

$$\Pr_{x\sim\mathcal{D}_n}\left[x \notin L \vee \Pr_M[R_L(x, M(x, 1^n, 1^{\varepsilon^{-1}}))] \ge 1-2^{-n}\right] \ge 1-\varepsilon.$$

Proof of Lemma 33. Let $\mathcal{C} = \{\mathcal{C}_n\}_{n\in\mathbb{N}}$ and $\mathcal{D} = \{\mathcal{D}_n\}_{n\in\mathbb{N}}$ be arbitrary samplable distributions satisfying the conditions in Lemma 33. Let $\ell(n)$ be a polynomial that represents the seed length required for sampling according to $\mathcal{D}_n$. Based on the standard way to transform Turing machines into uniformly computable circuits, we obtain a uniformly computable polynomial-size circuit family $D = \{D_n\}_{n\in\mathbb{N}}$, where $D_n\colon \{0,1\}^{\ell(n)} \to \{0,1\}^n \times \{0,1\}^n$ for each $n \in \mathbb{N}$, such that the distribution of $D_n(r)$ for $r \sim \{0,1\}^{\ell(n)}$ is statistically equivalent to $\mathcal{D}_n$. We use the notation $D_n^{(1)}$ (resp. $D_n^{(2)}$) to refer to the subcircuit of D that outputs the first (resp. second) half of the element, i.e., $D_n(r) = (D_n^{(1)}(r), D_n^{(2)}(r))$ for every $r \in \{0,1\}^{\ell(n)}$.

We define an NP language $\mathsf{InvCirc}$ as follows: For every $x \in \{0,1\}^*$ and every (binary representation of) circuit C,

$$\langle x, C\rangle \in \mathsf{InvCirc} \iff \exists r \in \{0,1\}^* \text{ such that } C(r) = x.$$

We define a samplable distribution $\mathcal{E} = \{\mathcal{E}_n\}_{n\in\mathbb{N}}$ as follows. For each $n, m \in \mathbb{N}$, let $\mathcal{H}_{n,m}$ be the pairwise independent hash family that maps n bits to m bits. For each $n \in \mathbb{N}$, we define $\mathcal{E}_n$ as a distribution of $\langle h, i, y, v \circ 0^{\ell(n)+\log^2 n - i}\rangle$, where $h \sim \mathcal{H}_{\ell(n),\ell(n)+\log^2 n}$, $i \sim [\ell(n)+\log^2 n]$, $y \sim \mathcal{C}_n$, and $v \sim \{0,1\}^i$. For readability, we identify $v \circ 0^{\ell(n)+\log^2 n - i}$ with v when n and i are clear in context. For each $n \in \mathbb{N}$, we let E_n denote a circuit that is given $r \in \{0,1\}^{\ell(n)}, h' \in \mathcal{H}_{\ell(n),\ell(n)+\log^2 n}, i' \in [\ell(n)+\log^2 n]$ and outputs $\langle h', i', D_n^{(2)}(r), h'(r)_{[i']} \circ 0^{\ell(n)+\log^2 n - i'}\rangle$. Since D_n is uniformly computable in polynomial time, $D_n^{(2)}$ and E_n are also uniformly computable in polynomial time. Thus, the distribution family $\{\langle \mathcal{E}_n, E_n\rangle\}_{n\in\mathbb{N}}$ is samplable, and $(\mathsf{InvCirc}, \{\langle \mathcal{E}_n, E_n\rangle\}_n) \in \mathsf{DistNP}$.

By the assumption that $\mathsf{DistNP} \subseteq \mathsf{HeurBPP}$ and Theorem 34, there exists a probabilistic polynomial-time algorithm M such that for every $n, \varepsilon^{-1} \in \mathbb{N}$,

$$\Pr_{z=\langle h,i,y,v\rangle\sim\mathcal{E}_n}\left[\langle\langle h, i, y, v\rangle, E_n\rangle \notin \mathsf{InvCirc} \vee \Pr_M[E_n(M(z, 1^n, 1^{\varepsilon^{-1}})) = \langle h, i, y, v\rangle] \ge 1-2^{-n}\right] \ge 1-\varepsilon$$

For simplicity, we assume that M does not fail, i.e., we assume that

$$\mathop{\mathbf{Pr}}_{z=\langle h,i,y,v\rangle\sim\mathcal{E}_n}\left[\langle h,i,y,v\rangle \notin \mathrm{Im}E_n \ \vee \ E_n(M(z,1^n,1^{\varepsilon^{-1}})) = \langle h,i,y,v\rangle\right] \geq 1-\varepsilon.$$

This assumption is valid in the following sense: The algorithm $\mathsf{CondExt}$ (specified later) executes M only polynomially many times. Thus, the probability that M does not satisfy the above is negligible. Therefore, this affects the result negligibly, and we can manage this by selecting slightly better parameters.

The outline of the proof is the following: First, we construct an algorithm M' that approximates the number of inverses of a given $y \sim \mathcal{C}_n$ with respect to $D_n^{(2)}$ based on M. Then, we construct the extrapolation algorithm $\mathsf{CondExt}$ based on M and M'.

We construct the approximation algorithm M' as follows: On input $y \sim \mathcal{C}_n$ and $1^{\delta^{-1}}$, where $\delta^{-1} \in \mathbb{N}$ is an additional error parameter, M' executes $M(\langle\langle h_{i,j}, i, y, v_{i,j}\rangle, E_n\rangle, 1^n, 1^{(\ell(n)+\log^2 n)\gamma^2})$ for each $i \in [\ell(n)+\log^2 n]$ and $j \in [\ell(n)]$, where $h_{i,j} \sim \mathcal{H}_{\ell(n),\ell(n)+\log^2 n}$, $v_{i,j} \sim \{0,1\}^i$, and $\gamma := \max\{\delta^{-1}, 16\}$. Then, M' outputs the maximum value of i (denoted by i^*) satisfying that M succeeds in finding an inverse with respect to E_n for some $j \in [\ell(n)]$, i.e., i^* is the maximum value f i satisfying that there exists $j \in [\ell(n)]$ such that

$$E_n(M(\langle h_{i,j}, i, y, v_{i,j}\rangle, 1^n, 1^{(\ell(n)+\log^2 n)\gamma^2})) = \langle h_{i,j}, i, y, v_{i,j}\rangle.$$

If there is no such i^*, then let $i^* = 0$.

It is easy to verify that M' halts in polynomial time. Let $R_y = \{r \in \{0,1\}^{\ell(n)} : D^{(2)}(r) = y\}$, $N_y := |R_y|$, and $n_y := \lfloor \log N_y \rfloor$. Then, M' satisfies the following claim.

Claim 35. *For every parameter $\delta^{-1} \in \mathbb{N}$, it holds that*

$$\mathop{\mathbf{Pr}}_{y\sim\mathcal{C}_n, M'}\left[n_y + 2 \leq i^* \leq n_y + \log\delta^{-1}\right] \geq 1 - 2(\ell(n)^2 + \ell(n)\log^2 n + 1)\cdot\delta$$

Proof of Claim 35. First, we show the upper bound. Fix y and each $h_{i,j}$ arbitrarily. For each i and j, the number of hash values of R_y, i.e., $|\{h_{i,j}(r)_{[i]} : r \in R_y\}|$ is at most $|R_y| = N_y$. Notice that M can find an inverse of $\langle h_{i,j}, i, y, v_{i,j}\rangle$ only if $v_{i,j}$ corresponds to one of the at most N_y hash values. Thus, M is successful with probability at most $N_y \cdot 2^{-i}$ over the choice of $v_{i,j} \sim \{0,1\}^i$. Particularly, for every $i > n_y + \log\delta^{-1}$ and every $j \in [\ell(n)]$, the success probability of M is at most

$$N_y \cdot 2^{-i} < 2^{n_y+1} \cdot 2^{-(n_y+\log\delta^{-1})} = 2\delta.$$

By the union bound, with probability at least $1 - 2\ell(n)(\ell(n)+\log^2 n)\delta$, the algorithm M fails to find an inverse for all $i \in [\ell(n)+\log^2 n]$ with $i > n_y + \log\delta^{-1}$ and for all $j \in [\ell(n)]$. In this case, it holds that $i^* \leq n_y + \log\delta^{-1}$.

Next, we show the lower bound. Let $i = n_y + 2$. For readability, we omit parameters 1^n and $1^{(\ell(n)+\log^2 n)\gamma^2}$ for M below. Remember that

$$\mathop{\mathbf{Pr}}_{y,h,i',v}\left[\langle h,i',y,v\rangle \notin \mathrm{Im}E_n \vee E_n(M(\langle h,i',y,v\rangle)) = \langle h,i',y,v\rangle\right] \geq 1 - (\ell(n)+\log^2 n)^{-1}\gamma^{-2},$$

where $y \sim \mathcal{C}_n$, $h \sim \mathcal{H}_{\ell(n),\ell(n)+\log^2 n}$, $i' \sim [\ell(n)+\log^2 n]$, and $v \sim \{0,1\}^i$.

By Markov's inequality, with probability at least $1 - \gamma^{-1} \geq 1 - \delta$ over the choice of $y \sim \mathcal{C}_n$,

$$\Pr_{h,i',v}\left[\langle h, i', y, v\rangle \notin \mathrm{Im}E_n \vee E_n(M(\langle h, i', y, v\rangle)) = \langle h, i', y, v\rangle\right] \geq 1 - (\ell(n) + \log^2 n)^{-1}\gamma^{-1}.$$

Since i' corresponds to $i = n_y + 2$ with probability $(\ell(n) + \log^2 n)^{-1}$, we have

$$\Pr_{h,v}\left[\langle h, i, y, v\rangle \notin \mathrm{Im}E_n \vee E_n(M(\langle h, i, y, v\rangle)) = \langle h, i, y, v\rangle\right] \geq 1 - \gamma^{-1}.$$

We fix any $y \in \{0,1\}^n$ satisfying the above inequality. Remember that $R_y = \{r \in \{0,1\}^{\ell(n)} : D^{(2)}(r) = y\}$. For every $r, r' \in R_y$ with $r \neq r'$, the collision probability that $h(r)_{[i]} = h(r')_{[i]}$ is 2^{-i} over the choice of h. By the union bound, for every $r \in R_y$, the probability that there exists another $r' \in R_y \setminus \{r\}$ satisfying that $h(r)_{[i]} = h(r')_{[i]}$ is at most $|R_y| \cdot 2^{-i} = |R_y| \cdot 2^{-n_y-2} \leq 2^{-1}$. For each $r \in R_y$ and $h \in \mathcal{H}_{\ell(n),\ell(n)+\log^2 n}$, let $A_{r,h}$ be a random variable that takes 1 iff there exists no $r' \in R_y \setminus \{r\}$ such that $h(r)_{[i]} = h(r')_{[i]}$ (otherwise, $A_{r,h} = 0$). Then, it holds that

$$\mathop{\mathbf{E}}_{h}\left[\sum_{y \in R_y} A_{y,h}\right] = \sum_{y \in R_y} \mathop{\mathbf{E}}_{h}[A_{y,h}] \geq \frac{N_y}{2}.$$

For each h, the number of hash values of R_y, i.e., $|\{h(r)_{[i]} : r \in R_y\}|$ is at least $\sum_{y \in R_y} A_{y,h}$. Thus, the random value $v \sim \{0,1\}^i$ corresponds to one of the hash values with probability at least $2^{-i} \cdot \sum_{y \in R_y} A_{y,h}$. Therefore, we have

$$\Pr_{h,v}\left[\langle h, i, y, v\rangle \in \mathrm{Im}E_n\right] \geq \mathop{\mathbf{E}}_{h}\left[2^{-i} \cdot \sum_{y \in R_y} A_{y,h}\right] \geq \frac{N_y}{2 \cdot 2^i} \geq \frac{2^{n_y}}{2 \cdot 2^{n_y+2}} = \frac{1}{8}$$

Since $\gamma \geq 16$, we obtain that

$$\begin{aligned}
&\Pr_{h,v}\left[E_n(M(\langle h, i, y, v\rangle)) \neq \langle h, i, y, v\rangle | \langle h, i, y, v\rangle \in \mathrm{Im}E_n\right] \\
&\quad = \Pr_{h,v}\left[\langle h, i, y, v\rangle \in \mathrm{Im}E_n \wedge E_n(M(\langle h, i, y, v\rangle)) \neq \langle h, i, y, v\rangle | \langle h, i, y, v\rangle \in \mathrm{Im}E_n\right] \leq 8\gamma^{-1} \leq 2^{-1},
\end{aligned}$$

and

$$\begin{aligned}
&\Pr_{h,v}\left[E_n(M(\langle h, i, y, v\rangle)) = \langle h, i, y, v\rangle\right] \\
&\quad \geq \Pr_{h,v}\left[\langle h, i, y, v\rangle \in \mathrm{Im}E_n\right] \cdot \Pr_{h,v}\left[E_n(M(\langle h, i, y, v\rangle)) = \langle h, i, y, v\rangle | \langle h, i, y, v\rangle \in \mathrm{Im}E_n\right] \\
&\quad \geq \frac{1}{8} \cdot \frac{1}{2} \geq \frac{1}{16}.
\end{aligned}$$

Since $(h_{i,j}, v_{i,j})$ is selected at independently random for each $j \in [\ell(n)]$, there is no $(h_{i,j}, v_{i,j})$ for which M is successful with probability at most $(1 - 1/16)^{\ell(n)} \leq 2^{-\Omega(\ell(n))}$. Notice that if this event does not occur, then $i^* \geq i = n_y + 2$. Thus, we conclude that

$$\Pr_{y \sim \mathcal{C}_n, M'}[i^* \geq n_y + 2] \geq (1 - \delta)(1 - 2^{-\Omega(\ell(n))}) \geq 1 - 2\delta,$$

where we assume that $\delta \geq 2^{-\Omega(\ell(n))}$; otherwise, we can directly compute N_y and n_y by trying all $r \in \{0,1\}^{\ell(n)}$ in time $2^{\ell(n)} \leq \mathsf{poly}(\delta^{-1})$.

By the union bound, we have

$$\begin{aligned}\Pr_{y\sim\mathcal{C}_n,M'}\left[n_y+2\leq i^*\leq n_y+\log\delta^{-1}\right] &\geq 1-2\ell(n)(\ell(n)+\log^2 n)\delta-2\delta\\ &= 1-2(\ell(n)^2+\ell(n)\log^2 n+1)\delta.\end{aligned}$$

This completes the proof of Claim 35. □

We construct the extrapolation algorithm $\mathsf{CondExt}$ based on M and M': On input $y\sim\mathcal{C}_n, 1^{\varepsilon^{-1}}, 1^{\delta^{-1}}$, (i) $\mathsf{CondExt}$ executes $i^*\leftarrow M'(y,1^{2\delta^{-1}\gamma_0\gamma_1})$ and sets $\tilde{i}:=i^*+2\log\gamma_2-1$, where $\gamma_0=2(\ell(n)^2+\ell(n)\log^2 n+1)$, $\gamma_1=4\varepsilon^{-1}$, and $\gamma_2=\max\{8\varepsilon^{-1},2\}$, (if M' outputs 0, then $\mathsf{CondExt}$ outputs an empty symbol ϵ and halts); (ii) $\mathsf{CondExt}$ picks $h\sim\mathcal{H}_{\ell(n),\ell(n)+\log^2 n}$ and repeats the following execution $m:=4\gamma_0\gamma_1\gamma_2^2\delta^{-1}\ell(n)$ times: execute

$$M(\langle\langle h,\tilde{i},y,v\rangle,E_n\rangle,1^n,1^{2\delta^{-1}(\ell(n)+\log^2 n)\gamma_3})$$

for $r\sim\{0,1\}^{\tilde{i}}$ (where r is selected independently for each trial) and $\gamma_3=4\gamma_0\gamma_1\gamma_2^4\delta^{-1}$. If M outputs a valid inverse $(r',h,\tilde{i})$ of $\langle h,\tilde{i},y,v\rangle$ (the validity can be easily verified), then $\mathsf{CondExt}$ outputs $C^{(1)}(r')$ and halts; and (iii) if no inverse was found within the m trials, then $\mathsf{CondExt}$ outputs an error symbol $\perp$ or an arbitrarily string and halts. It is not hard to verify that $\mathsf{CondExt}$ halts in polynomial time.

We verify the correctness of $\mathsf{CondExt}$. For readability, we omit the unary parameters for $\mathsf{CondExt}$, M', and M.

By Claim 35,

$$\Pr_{y\sim\mathcal{C}_n,M'}\left[n_y+2\leq i^*\leq n_y+\log\delta^{-1}\gamma_0\gamma_1+1\right]\geq 1-\frac{\delta}{2}\cdot 2(\ell(n)^2+\ell(n)\log^2 n+1)\gamma_0^{-1}\gamma_1^{-1}=1-\frac{\delta}{2}\gamma_1^{-1}$$

Thus, by Markov's inequality, the following holds with probability at least $1-\delta/2$ over the choice of $y\sim\mathcal{C}_n$:

$$\Pr_{M'}\left[n_y+2\leq i^*\leq n_y+\log\delta^{-1}\gamma_0\gamma_1+1\right]\geq 1-\gamma_1^{-1} \tag{32}$$

Furthermore, on step (ii), we have

$$\Pr_{y,h,i,v}\left[\langle h,i,y,v\rangle\notin \mathrm{Im}E_n\vee E_n(M(\langle h,i,y,v\rangle))=\langle h,i,y,v\rangle\right]\geq 1-\frac{\delta}{2(\ell(n)+\log^2 n)}\cdot\gamma_3^{-1},$$

where $y\sim\mathcal{C}_n, h\sim\mathcal{H}_{\ell(n),\ell(n)+\log^2 n}$, $i\sim[\ell(n)+\log^2 n]$, and $v\sim\{0,1\}^i$. By Markov's inequality, the following holds with probability at least $1-\delta/2$ over the choice of $y\sim\mathcal{C}_n$:

$$\Pr_{h,i,v}\left[\langle h,i,y,v\rangle\notin \mathrm{Im}E_n\vee E_n(M(\langle h,i,y,v\rangle))=\langle h,i,y,v\rangle\right]\geq 1-\frac{1}{(\ell(n)+\log^2 n)}\cdot\gamma_3^{-1}.$$

In this case, it holds that

$$\forall i\in[\ell(n)+\log^2 n]\quad \Pr_{h,v}\left[\langle h,i,y,v\rangle\notin \mathrm{Im}E_n\vee E_n(M(\langle h,i,y,v\rangle))=\langle h,i,y,v\rangle\right]\geq 1-\gamma_3^{-1}, \tag{33}$$

because each i is selected with probability $(\ell(n) + \log^2 n)^{-1}$.

By the union bound, with probability $1 - \delta$ over the choice of $y \sim \mathcal{C}_n$, inequalities (32) and (33) are satisfied. For the theorem, it suffices to show that, under this condition,

$$\mathsf{L}_1\left(\mathsf{CondExt}(y), \mathcal{D}_n(\cdot \mid y)\right) \leq \varepsilon.$$

Thus, we fix y that satisfies (32) and (33) arbitrarily and show the above.

On the execution of $\mathsf{CondExt}$, let B an event that M' does not output i^* satisfying that

$$n_y + 2 \leq i^* \leq n_y + \log \delta^{-1}\gamma_0\gamma_1 + 1$$

Then, we show the following claims:

Claim 36. $\mathbf{Pr}[B] \leq \varepsilon/4$.

Claim 37. *Under the condition that $\neg B$, it holds that* $\mathsf{L}_1\left(\mathsf{CondExt}(y), \mathcal{D}_n(\cdot \mid y)\right) \leq 3\varepsilon/4$

These claims imply Lemma 33 as

$$\mathsf{L}_1\left(\mathsf{CondExt}(y), \mathcal{D}_n(\cdot \mid y)\right) \leq 1 \cdot \mathbf{Pr}[B] + 3\varepsilon/4 \cdot \mathbf{Pr}[\neg B] \leq \varepsilon.$$

Thus, the remaining of the proof is to show Claims 36 and 37.

Proof of Claim 36. The claim immediately follows from inequality (32) and $\gamma_1 = 4\varepsilon^{-1}$. □

Proof of Claim 37. Under the condition $\neg B$, we have

$$n_y + 2\log\gamma_2 + 1 \leq \tilde{i} \leq n_y + 2\log\gamma_2 + \log\delta^{-1}\gamma_0\gamma_1.$$

For each $r \in R_y$ and $h \in \mathcal{H}_{\ell(n),\ell(n)+\log^2 n}$, we use the notation $A_{r,h}$ again to refer to the binary random variable that takes 1 iff there exists no $r' \in R_y \setminus \{r\}$ such that $h(r)_{[\tilde{i}]} = h(r')_{[\tilde{i}]}$. Then, for each $r \in R_y$, it holds that $\mathbf{Pr}_h[A_{r,h} = 0] \leq N_y \cdot 2^{-\tilde{i}}$ by the union bound; thus,

$$\mathop{\mathbf{E}}_h\left[\sum_{r\in R_y} A_{r,h}\right] = \sum_{r\in R_y}\mathop{\mathbf{E}}_h\left[A_{r,h}\right] \geq N_y \cdot (1 - N_y 2^{-\tilde{i}}) \geq N_y \cdot (1 - 2^{n_y}2^{-(n_y+2\log\gamma_2+1)}) \geq N_y \cdot (1 - \frac{1}{2\gamma_2^2}).$$

By Markov's inequality,

$$\mathop{\mathbf{Pr}}_h\left[\sum_{r\in R_y} A_{r,h} \geq (1 - \frac{1}{\gamma_2}) \cdot N_y\right] \geq 1 - \frac{1}{2\gamma_2}.$$

Furthermore, by inequality (33),

$$\mathop{\mathbf{Pr}}_{h,v}\left[\langle h, \tilde{i}, y, v\rangle \notin \mathrm{Im}E_n \vee E_n(M(\langle h, \tilde{i}, y, v\rangle)) = \langle h, \tilde{i}, y, v\rangle\right] \geq 1 - \gamma_3^{-1}.$$

By Markov's inequality, with probability at least $1 - 1/(2\gamma_2)$ over the choice of h,

$$\mathop{\mathbf{Pr}}_{v}\left[\langle h, \tilde{i}, y, v\rangle \notin \mathrm{Im}E_n \vee E_n(M(\langle h, \tilde{i}, y, v\rangle)) = \langle h, \tilde{i}, y, v\rangle\right] \geq 1 - 2\gamma_2\gamma_3^{-1}. \tag{34}$$

Let $S_h = \{r \in R_y : A_{r,h} = 1\}$ and $T_h = \{h(r)_{[\tilde{i}]} : r \in S_h\}$. Notice that $|S_h| = |T_h| = \sum_{r \in R_y} A_{r,h}$ holds. We call a hash function h satisfying inequality (34) and $|T_h| \geq (1 - \gamma_2^{-1}) \cdot N_y$ a good hash function. By the union bound, h is good with probability at least $1 - \gamma_2^{-1}$ over the choice of $h \sim \mathcal{H}_{\ell(n),\ell(n)+\log^2 n}$.

We fix a good hash function h arbitrarily. Then,

$$\Pr_{v \sim \{0,1\}^{\tilde{i}}}[v \in T_h] = |T_h| \cdot 2^{-\tilde{i}} \geq (1 - \gamma_2^{-1}) \cdot 2^{n_y} \cdot 2^{-(n_y + 2\log\gamma_2 + \log\delta^{-1}\gamma_0\gamma_1)} = (2\gamma_0\gamma_1\gamma_2^2)^{-1}\delta,$$

where we use the fact that $\gamma_2 \geq 2$.

If $v \in T_h$, then $\langle h, \tilde{i}, y, v\rangle \in \mathrm{Im}E_n$. Therefore, we have

$$\begin{aligned}
&\Pr_v\left[E_n(M(\langle h, \tilde{i}, y, v\rangle)) \neq \langle h, \tilde{i}, y, v\rangle | v \in T_h\right] \\
&= \Pr_v\left[\langle h, \tilde{i}, y, v\rangle \in \mathrm{Im}E_n \wedge E_n(M(\langle h, \tilde{i}, y, v\rangle)) \neq \langle h, \tilde{i}, y, v\rangle | v \in T_h\right] \leq 2\gamma_2\gamma_3^{-1} \cdot 2\gamma_0\gamma_1\gamma_2^2\delta^{-1} = \gamma_2^{-1}.
\end{aligned}$$

Let $S'_h = \{r \in S_h : M(\langle h, \tilde{i}, y, h(r)_{[\tilde{i}]}\rangle)) = \langle h, \tilde{i}, y, h(r)_{[\tilde{i}]}\rangle\}$ and $T'_h = \{h(r)_{[\tilde{i}]} : r \in S'_h\}$. Then, we have $|S'_h| = |T'_h|$ and $|S'_h| \geq (1 - \gamma_2^{-1})|S_h| \geq (1 - \gamma_2^{-1})^2 N_y \geq (1 - 2\gamma_2^{-1})N_y$.

Under the condition that $v \in T'_h$, $\mathsf{CondExt}$ obtains $r \in S'_h$ from M at uniformly random over S'_h. The total variation distance between the uniform distribution over S'_h and the uniform distribution over R_y is at most $2\gamma_2^{-1}$. Therefore, under the condition I_h that M finds a valid inverse $(r', h, \tilde{i})$ of $\langle h, \tilde{i}, y, v\rangle$ (i.e., $D^{(2)}(r') = y$), the total variation distance between the conditional probability $\tilde{D}$ of r' and the uniform distribution over R_y (denoted by U_{R_y}) is bounded as follows:

$$\mathsf{L}_1(\tilde{D}, R_y | I_h) \leq 2\gamma_2^{-1} + 1 \cdot \Pr_v[v \notin T'_h | v \in \{h(r)_{[\tilde{i}]} : r \in R_y\}] \leq 2\gamma_2^{-1} + 2\gamma_2^{-1}N_y \cdot N_y^{-1} = 4\gamma_2^{-1} \leq \varepsilon/2.$$

In this case,

$$\mathsf{L}_1(\mathsf{CondExt}(y), \mathcal{D}(\cdot \mid y) | I_h) = \mathsf{L}_1(D^{(1)}(\tilde{D}), D^{(1)}(R_y) | I_h) \leq \mathsf{L}_1(\tilde{D}, R_y | I_h) \leq \varepsilon/2.$$

Next, we show that, for every good h, the algorithm M finds a valid inverse (i.e., the condition I_y is satisfied) at some trial with probability at least $1 - 2^{-\Omega(\ell(n))}$ over the choices of $v \sim \{0,1\}^{\tilde{i}}$. This implies the claim as

$$\begin{aligned}
\mathsf{L}_1(\mathsf{CondExt}(y), \mathcal{D}^{(1)}(\cdot \mid y)) &\leq 1 \cdot \Pr_h[h \text{ is not good}] + \mathsf{L}_1(\mathsf{CondExt}(y), \mathcal{D}^{(1)}(\cdot \mid y) | h \text{ is good}, I_h) \\
&\quad + 1 \cdot \Pr[M \text{ finds no valid inverse for all } m \text{ trials} | h \text{ is good}] \\
&\leq \gamma_2^{-1} + \varepsilon/2 + 2^{-\Omega(\ell(n))} \\
&\leq \varepsilon/8 + \varepsilon/2 + \varepsilon/8 = 3\varepsilon/4,
\end{aligned}$$

where we assume that $\varepsilon/8 \geq 2^{-\Omega(\ell(n))}$; otherwise, we can try all $r \in \{0,1\}^{\ell(n)}$ and perfectly simulate $\mathcal{D}^{(1)}(\cdot \mid y)$ in time $2^{\ell(n)} \leq \mathsf{poly}(\varepsilon^{-1})$.

Remember that

$$\Pr_v[v \in T_h] \geq (2\gamma_0\gamma_1\gamma_2^2)^{-1}\delta \text{ and } \Pr_v\left[E_n(M(\langle h, \tilde{i}, y, v\rangle)) = \langle h, \tilde{i}, y, v\rangle | v \in T_h\right] \geq 1 - \gamma_2^{-1} \geq 2^{-1}.$$

Therefore,

$$\begin{aligned}
\Pr_v\left[E_n(M(\langle h, \tilde{i}, y, v\rangle)) = \langle h, \tilde{i}, y, v\rangle\right] &\geq \Pr_v\left[E_n(M(\langle h, \tilde{i}, y, v\rangle)) = \langle h, \tilde{i}, y, v\rangle | v \in T_h\right] \cdot \Pr_v[v \in T_h] \\
&\geq (4\gamma_0\gamma_1\gamma_2^2)^{-1} \cdot \delta.
\end{aligned}$$

Therefore, M fails to find some inverse with probability at most $1-(4\gamma_0\gamma_1\gamma_2^2)^{-1}\cdot\delta$. Thus, the probability that M finds no inverse in all m trials is at most

$$(1-(4\gamma_0\gamma_1\gamma_2^2)^{-1}\cdot\delta)^m \le (1-(4\gamma_0\gamma_1\gamma_2^2)^{-1}\cdot\delta)^{4\gamma_0\gamma_1\gamma_2^2\delta^{-1}\ell(n)} \le 2^{-\Omega(\ell(n))},$$

as desired. □

This completes the proof of Lemma 33. □

5.2 Consequences of Conditional Extrapolation

We show that conditional extrapolation (Item 4) implies Items 2, 3 and 5 in Theorems 31 and 32. We will use the following proposition, which was observed in [IL90].

Proposition 38. *If Conditional Extrapolation (Item 4 in Theorem 31) holds, then there exists no infinitely-often one-way functions.*

Proof. Let f be an arbitrarily polynomial-time computable length-preserving function. Let $\mathcal{C}=\{\mathcal{C}_n\}_{n\in\mathbb{N}}$ (resp. $\mathcal{D}=\{\mathcal{D}_n\}_{n\in\mathbb{N}}$) be a samplable distribution, where each $\mathcal{C}_n$ (resp. each $\mathcal{D}_n$) is a distribution of $f(x)$ for $x\sim\{0,1\}^n$ (resp. $(x,f(x))$ for $x\sim\{0,1\}^n$).

By the assumption, there exists a probabilistic polynomial-time algorithm $\mathsf{CondExt}$ such that for all $n,\varepsilon^{-1},\delta^{-1}\in\mathbb{N}$,

$$\Pr_{y\sim\mathcal{C}_n}\left[\mathsf{L}_1\left(\mathsf{CondExt}(y;1^{\varepsilon^{-1}},1^{\delta^{-1}}),\mathcal{D}_n(\cdot\mid y)\right)\le\varepsilon\right]\ge 1-\delta.$$

For every $y\in\mathsf{Support}(\mathcal{C}_n)=\mathrm{Im}f_n$, the conditional distribution $\mathcal{D}_n(\cdot\mid y)$ is a uniform distribution over $f^{-1}(y)$. Therefore, for a given $y=f_n(x)$ (where $x\sim\{0,1\}^n$), the algorithm $\mathsf{CondExt}(y;1^{2n},1^{2n})$ outputs an inverse element with probability at least $1-(1/(2n)+1/(2n))=1-1/n$. □

Lemma 39 (Item 4 ⇒ Items 2 in Theorem 31)**.** *If Conditional Extrapolation (Item 4 in Theorem 31) holds, then for every samplable distribution families $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ and $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$, where each $\mathcal{D}_n$ is over $\{0,1\}^n\times\{0,1\}^n$, and each $\mathcal{C}_n$ is over $\mathsf{Support}(\mathcal{D}_n^{(2)})$, and for every polynomial q, there exists a polynomial p such that for all large enough $n\in\mathbb{N}$,*

$$\Pr_{y\sim\mathcal{C}_n,x\sim\mathcal{D}_n(\cdot|y)}\left[\mathsf{pK}^{p(n)}(x\mid y)\le\log\frac{1}{\mathcal{D}_n(x\mid y)}+\log p(n)\right]\ge 1-\frac{1}{q(n)}.$$

Proof. Let $\mathcal{C}=\{\mathcal{C}_n\}_{n\in\mathbb{N}}$ and $\mathcal{D}=\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ be arbitrary samplable distributions satisfying the conditions in Lemma 39. Let q be an arbitrary polynomial.

By Conditional Extrapolation, there exists a probabilistic polynomial-time algorithm $\mathsf{CondExt}$ such that for all $n\in\mathbb{N}$,

$$\Pr_{y\sim\mathcal{C}_n}\left[\mathsf{L}_1\left(\mathsf{CondExt}(y;1^{8q(n)},1^{8q(n)}),\mathcal{D}_n(\cdot\mid y)\right)\le\frac{1}{8q(n)}\right]\ge 1-\frac{1}{8q(n)}. \tag{35}$$

We define a distribution family $\mathcal{E}=\{\mathcal{E}_n\}_{n\in\mathbb{N}}$ as follows: each $\mathcal{E}_n$ is a distribution of (x,y) for $y\sim\mathcal{C}_n$ and $x\sim\mathsf{CondExt}(y;1^{8q(n)},1^{8q(n)})$. Since $\mathsf{CondExt}$ is a probabilistic polynomial-time algorithm, the distribution family $\mathcal{E}$ is samplable. Let $\mathcal{E}^*=\{\mathcal{E}_n^*\}_{n\in\mathbb{N}}$ be a distribution family, where each $\mathcal{E}_n^*$ is a

distribution of (x, y) for $y \sim \mathcal{C}_n$ and $x \sim \mathcal{D}(\cdot \mid y)$. Then, by inequality (35), we have that, for each $n \in \mathbb{N}$,

$$\mathsf{L}_1(\mathcal{E}_n, \mathcal{E}_n^*) \leq \frac{1}{8q(n)} + \frac{1}{8q(n)} = \frac{1}{4q(n)}.$$

By Conditional Extrapolation and Proposition 38, there is no infinitely-often one-way function. Thus, by Theorem 13 (Item 1 $\Rightarrow$ Item 2) for the samplable distribution $\mathcal{E}$, there exists a polynomial p such that for all large enough n,

$$\mathop{\mathbf{Pr}}_{(x,y)\sim\mathcal{E}_n}\left[\mathsf{pK}^{p(n)}(x \mid y) \leq \log \frac{1}{\mathcal{E}_n(x \mid y)} + \log p(n)\right] \geq 1 - \frac{1}{4q(n)}.$$

Since $\mathsf{L}_1(\mathcal{E}_n, \mathcal{E}_n^*) \leq 1/(4q(n))$, we have

$$\begin{aligned}\mathop{\mathbf{Pr}}_{y\sim\mathcal{C}_n, x\sim\mathcal{D}_n(\cdot|y)}\left[2^{\mathsf{pK}^{p(n)}(x|y)}\mathcal{E}_n(x \mid y) \leq p(n)\right] &\geq \mathop{\mathbf{Pr}}_{(x,y)\sim\mathcal{E}_n}\left[2^{\mathsf{pK}^{p(n)}(x|y)}\mathcal{E}_n(x \mid y) \leq p(n)\right] - \frac{1}{4q(n)} \\ &\geq 1 - \frac{1}{2q(n)}.\end{aligned}$$

For Lemma 39, it suffices to show that

$$\mathop{\mathbf{Pr}}_{y\sim\mathcal{C}_n, x\sim\mathcal{D}_n(\cdot|y)}\left[\mathcal{D}_n(x \mid y) \leq 8q(n) \cdot \mathcal{E}_n(x \mid y)\right] \geq 1 - \frac{1}{2q(n)}, \tag{36}$$

because it implies that, by the union bound,

$$\mathop{\mathbf{Pr}}_{y\sim\mathcal{C}_n, x\sim\mathcal{D}_n(\cdot|y)}\left[2^{\mathsf{pK}^{p(n)}(x|y)}\mathcal{D}_n(x \mid y) \leq 8p(n)q(n)\right] \geq 1 - \frac{1}{q(n)}$$

and

$$\mathop{\mathbf{Pr}}_{y\sim\mathcal{C}_n, x\sim\mathcal{D}_n(\cdot|y)}\left[\mathsf{pK}^{p(n)}(x \mid y) \leq \log \frac{1}{\mathcal{D}_n(x \mid y)} + \log 8p(n)q(n)\right] \geq 1 - \frac{1}{q(n)}.$$

We show inequality (36). Notice that $\mathcal{E}_n(x \mid y) = \mathbf{Pr}[\mathsf{CondExt}(y; 1^{8q(n)}, 1^{8q(n)}) = x]$. Therefore, by inequality (35),

$$\mathop{\mathbf{Pr}}_{y\sim\mathcal{C}_n}\left[\mathsf{L}_1\left(\mathcal{E}_n(\cdot \mid y), \mathcal{D}_n(\cdot \mid y)\right) \leq \frac{1}{8q(n)}\right] \geq 1 - \frac{1}{8q(n)}.$$

We fix y that satisfies the event above arbitrarily, i.e., $\mathsf{L}_1\left(\mathcal{E}_n(\cdot \mid y), \mathcal{D}_n(\cdot \mid y)\right) \leq 1/(8q(n))$ holds.

We have

$$\mathop{\mathbf{E}}_{x\sim\mathcal{E}_n(\cdot|y)}\left[\frac{\mathcal{D}_n(x \mid y)}{\mathcal{E}_n(x \mid y)}\right] \leq \sum_x \mathcal{D}_n(x \mid y) \leq 1.$$

Thus, by Markov's inequality,

$$\mathop{\mathbf{Pr}}_{x\sim\mathcal{E}_n(\cdot|y)}\left[\mathcal{D}_n(x \mid y) \leq 8q(n) \cdot \mathcal{E}_n(x \mid y)\right] = \mathop{\mathbf{Pr}}_{x\sim\mathcal{E}_n(\cdot|y)}\left[\frac{\mathcal{D}_n(x \mid y)}{\mathcal{E}_n(x \mid y)} \leq 8q(n)\right] \geq 1 - \frac{1}{8q(n)}.$$

Since $\mathsf{L}_1\left(\mathcal{E}_n(\cdot \mid y), \mathcal{D}_n(\cdot \mid y)\right) \leq 1/(8q(n))$,

$$\begin{aligned}\mathop{\mathbf{Pr}}_{x\sim\mathcal{D}_n(\cdot|y)}\left[\mathcal{D}_n(x \mid y) \leq 8q(n) \cdot \mathcal{E}_n(x \mid y)\right] &\geq \mathop{\mathbf{Pr}}_{x\sim\mathcal{E}_n(\cdot|y)}\left[\mathcal{D}_n(x \mid y) \leq 8q(n) \cdot \mathcal{E}_n(x \mid y)\right] - \frac{1}{8q(n)} \\ &\geq 1 - \frac{1}{4q(n)}.\end{aligned}$$

By the union bound, we conclude that

$$\Pr_{y\sim\mathcal{C}_n, x\sim\mathcal{D}_n(\cdot|y)}\left[\mathcal{D}_n(x \mid y) \le 8q(n)\cdot\mathcal{E}_n(x \mid y)\right] \ge 1 - \frac{3}{8q(n)} \ge 1 - \frac{1}{2q(n)},$$

as desired. □

Lemma 40 (Item 4 ⇒ Items 3 in Theorem 31)**.** *If Conditional Extrapolation (Item 4 in Theorem 31) holds, then for every computable set $L \subseteq \{\{0,1\}^n \times \{0,1\}^n\}_n$, for every samplable distribution families $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ and $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and each $\mathcal{C}_n$ is over $\mathsf{Support}(\mathcal{D}_n^{(2)})$, and for every polynomial q, there exists a polynomial p such that for all n,*

$$\Pr_{y\sim\mathcal{C}_n, x\sim\mathcal{D}_n(\cdot|y)}\left[x \in L_y \implies \mathsf{pK}^{p(n)}(x \mid y) \le \log|L_y| + \log p(n)\right] \ge 1 - \frac{1}{q(n)}.$$

Proof. Let $L \subseteq \{\{0,1\}^n \times \{0,1\}^n\}_n$ be an arbitrary computable set. Let $\mathcal{C} = \{\mathcal{C}_n\}_{n\in\mathbb{N}}$ and $\mathcal{D} = \{\mathcal{D}_n\}_{n\in\mathbb{N}}$ be arbitrary samplable distributions satisfying the conditions in Lemma 40. Let q be an arbitrary polynomial.

By Conditional Extrapolation, there exists a probabilistic polynomial-time algorithm $\mathsf{CondExt}$ such that for all $n \in \mathbb{N}$,

$$\Pr_{y\sim\mathcal{C}_n}\left[\mathsf{L}_1\left(\mathsf{CondExt}(y; 1^{4q(n)}, 1^{4q(n)}), \mathcal{D}_n(\cdot \mid y)\right) \le \frac{1}{4q(n)}\right] \ge 1 - \frac{1}{4q(n)}. \tag{37}$$

We define a distribution family $\mathcal{E} = \{\mathcal{E}_n\}_{n\in\mathbb{N}}$ as follows: each $\mathcal{E}_n$ is a distribution of (x, y) for $y \sim \mathcal{C}_n$ and $x \sim \mathsf{CondExt}(y; 1^{4q(n)}, 1^{4q(n)})$. Since $\mathsf{CondExt}$ is a probabilistic polynomial-time algorithm, the distribution family $\mathcal{E}$ is samplable. Let $\mathcal{E}^* = \{\mathcal{E}_n^*\}_{n\in\mathbb{N}}$ be a distribution family, where each $\mathcal{E}_n^*$ is a distribution of (x, y) for $y \sim \mathcal{C}_n$ and $x \sim \mathcal{D}(\cdot \mid y)$. Then, by inequality (37), we have that, for each $n \in \mathbb{N}$,

$$\mathsf{L}_1(\mathcal{E}_n, \mathcal{E}_n^*) \le \frac{1}{4q(n)} + \frac{1}{4q(n)} = \frac{1}{2q(n)}.$$

By Conditional Extrapolation and Proposition 38, there is no infinitely-often one-way function. Thus, by Theorem 13 (Item 1 ⇒ Item 4) for the samplable distribution $\mathcal{E}$, there exists a polynomial p such that for all n,

$$\Pr_{(x,y)\sim\mathcal{E}_n}\left[x \in L_y \implies \mathsf{pK}^{p(n)}(x \mid y) \le \log|L_y| + \log p(n)\right] \ge 1 - \frac{1}{2q(n)}.$$

Since $\mathsf{L}_1(\mathcal{E}_n, \mathcal{E}_n^*) \le 1/(2q(n))$, we conclude that

$$\Pr_{y\sim\mathcal{C}_n, x\sim\mathcal{D}_n(\cdot|y)}\left[x \in L_y \implies \mathsf{pK}^{p(n)}(x \mid y) \le \log|L_y| + \log p(n)\right] \ge 1 - \frac{1}{q(n)},$$

as desired. □

We can also show Theorem 32 in the same way as Lemma 40, where the only difference is that we use Item 1 ⇒ Item 6 in Theorem 13 instead of Item 1 ⇒ Item 4.

5.3 Average-Case Easiness of NP from Conditional Coding and Language Compression

We verify that the *independent* variants of average-case conditional coding and language compression are sufficient for the average-case easiness of NP.

To derive DistNP $\subseteq$ HeurBPP, we use the following useful lemma.

Lemma 41 ([Imp95, Proposition 3])**.** *If every distributional* NP *problem has a probabilistic polynomial-time heuristic algorithm of failure probability at most n^{-2} over the choice of instances (where n is an instance size*[4]*), then* DistNP $\subseteq$ HeurBPP*.*

First, we show Item 3 $\Rightarrow$ Item 1 in Theorem 31.

Lemma 42 (Item 3 $\Rightarrow$ Item 1 in Theorem 31)**.** *Suppose that for every computable set $L \subseteq \{\{0,1\}^n \times \{0,1\}^n\}_n$, for every samplable distribution families $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ and $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and each $\mathcal{C}_n$ is over $\mathsf{Support}(\mathcal{D}_n^{(2)})$, and for every polynomial q, there exists a polynomial p such that for all n,*

$$\Pr_{y\sim\mathcal{C}_n, x\sim\mathcal{D}_n(\cdot|y)}\left[x \in L_y \implies \mathsf{pK}^{p(n)}(x \mid y) \leq \log|L_y| + \log p(n)\right] \geq 1 - \frac{1}{q(n)}.$$

Then, DistNP $\subseteq$ HeurBPP *holds.*

Proof. Let $(L, \{\mathcal{C}_n\}_{n\in\mathbb{N}})$ be an arbitrary distributional NP problem such that each $\mathcal{C}_n$ is over $\{0,1\}^n$. By Lemma 41, it suffices to construct a probabilistic polynomial-time A such that for every $n \in \mathbb{N}$,

$$\Pr_{y\sim\mathcal{C}_n}\left[\Pr_A[A(y) = L(y)] \geq 3/4\right] \geq 1 - n^{-2}.$$

Let V_L be the polynomial-time computable verifier for L. Without loss of generality (by padding), we assume that the length of every yes instance $y \in L$ and the length of every witness x for $y \in L$ are the same, i.e., $|x| = |y|$ (with respect to V_L). For every $y \in L$, let L_y be a set of V_L-witness x for y.

First, we assume the following claim and construct the algorithm A for $(L, \{\mathcal{C}_n\}_{n\in\mathbb{N}})$.

Claim 43. *Under the assumption (i.e., independent average-case language compression), there exists a polynomial $p(n)$ such that for every $n \in \mathbb{N}$,*

$$\Pr_{y\sim\mathcal{C}_n}\left[y \in L \implies \Pr_{x\sim L_y}\left[\mathsf{pK}^{p(n)}(x \mid y) \leq \log|L_y| + \log p(n)\right] \geq 3/4\right] \geq 1 - \frac{1}{n^2}. \tag{38}$$

We consider the following algorithm A: On a given instance $y \in \{0,1\}^n$, the algorithm A samples $x \sim \mathsf{Usamp}(1^n, 1^{p(n)}, y)$ repeatedly $O(np(n))$ times, where Usamp is the universal sampler in Definition 21, and outputs 1 iff some sample x satisfies that $V_L(y,x) = 1$ (otherwise, A outputs 0). A is trivially a probabilistic polynomial-time algorithm.

To verify the correctness of A. Fix $n \in \mathbb{N}$ arbitrarily. For the correctness, it suffices to show that for every y satisfying the event in inequality (38), the probability that $A(y) = L(y)$ is at least 3/4.

[4] Although the original statement is for errorless and deterministic algorithms and for samplable distributions over instances whose size varies, it is not hard to verify that the same proof works for HeurBPP and distributions over instances of fixed size by a simple padding argument.

Notice that $A(y)$ always output 0 for every $y \notin L$ since there is no x satisfying that $V_L(y,x) = 1$. Therefore, we only consider an arbitrary $y \in L$ satisfying the event in inequality (38), i.e.,

$$\Pr_{x \sim L_y}\left[\mathsf{pK}^{p(n)}(x \mid y) \leq \log |L_y| + \log p(n)\right] \geq 3/4.$$

Let $G_y \subseteq L_y$ be a subset of $x \in L_y$ satisfying the event above. By Proposition 22, the probability that some witness $x \in L_y$ is sampled from $\mathsf{Usamp}(1^n, 1^{p(n)}, y)$ is at least

$$\begin{aligned}
\Omega\left(\frac{1}{n}\right) \cdot \sum_{x \in L_y} \frac{1}{2^{\mathsf{pK}^{p(n)}(x|y)}} &\geq \Omega\left(\frac{1}{n}\right) \cdot \sum_{x \in G_y} \frac{1}{2^{\mathsf{pK}^{p(n)}(x|y)}} \\
&\geq \Omega\left(\frac{1}{n}\right) \cdot \sum_{x \in G_y} \frac{1}{p(n)|L_y|} \\
&\geq \Omega\left(\frac{1}{np(n)}\right) \cdot \frac{|G_y|}{|L_y|} \\
&\geq \Omega\left(\frac{1}{np(n)}\right) \cdot \frac{3}{4}.
\end{aligned}$$

Since A repeats sampling from $\mathsf{Usamp}(1^n, 1^{p(n)}, y)$ $O(np(n))$ times, the failure probability that A fails to find a witness for y is reduced to $1/4$.

Thus, the remaining is the proof of Claim 43.

Proof of Claim 43. Consider a polynomial-time computable set $L' := \{(x1, y1) : V_L(y,x) = 1\}$. Note that for any input $y \in L \cap \{0,1\}^{n-1}$, $L'_{y1} \subseteq \{0,1\}^n$ is the set of $x1$ for $x \in L_y$. Let $\mathcal{C}' = \{\mathcal{C}'_n\}_{n\in\mathbb{N}}$ be a samplable distribution such that each $\mathcal{C}'_n$ is a distribution of $y1$ for $y \sim \mathcal{C}_{n-1}$. Let $\mathcal{D} = \{\mathcal{D}_n\}_{n\in\mathbb{N}}$ be a samplable distribution such that each $\mathcal{D}_n$ is a distribution of (x,y) determined by $x', y' \sim \{0,1\}^{n-1}$ as

$$(x,y) = \begin{cases} (x'1, y'1) & \text{if } V_L(y', x') = 1 \\ (0^n, y'1) & \text{if } x' = 0^{n-1} \text{ and } V_L(y', x') = 0 \\ (x'0, y'0) & \text{otherwise.} \end{cases}$$

For each $n \in \mathbb{N}$, the distribution $\mathcal{C}'_n$ is over $\{0,1\}^n$, the distribution $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and

$$\mathsf{Support}(\mathcal{C}'_n) \subseteq \{y1 : y \in \{0,1\}^{n-1}\} \subseteq \mathsf{Support}(\mathcal{D}^{(2)}_n).$$

Therefore, by the assumption (i.e., independent average-case language compression), there exists a polynomial p such that for all $n \in \mathbb{N}$,

$$\Pr_{y \sim \mathcal{C}'_n, x \sim \mathcal{D}_n(\cdot|y)}\left[x \in L'_y \implies \mathsf{pK}^{p(n)}(x \mid y) \leq \log |L'_y| + \log p(n)\right] \geq 1 - \frac{1}{8n^2}.$$

Fix $n \in \mathbb{N}$ arbitrarily. It is not hard to verify that for every $y \in \mathsf{Support}(\mathcal{C}_n)$, the conditional distribution $\mathcal{D}_n(\cdot \mid y1)$ is statistically equivalent to the uniform distribution over $L'_{y1} \cup \{0^{n+1}\}$. Therefore,

$$\Pr_{y \sim \mathcal{C}_n, x \sim L'_{y1} \cup \{0^{n+1}\}}\left[x \in L'_{y1} \implies \mathsf{pK}^{p(n+1)}(x \mid y) \leq \log |L'_{y1}| + \log p(n+1)\right] \geq 1 - \frac{1}{8(n+1)^2}.$$

Let $p'(n) = p(n+1)$. By Markov's inequality and the fact that $|L'_{y1}| = |L_y|$,

$$\Pr_{y\sim\mathcal{C}_n}\left[\Pr_{x\sim L'_{y1}\cup\{0^{n+1}\}}\left[x \in L'_{y1} \wedge \mathsf{pK}^{p'(n)}(x \mid y) > \log|L_y| + \log p'(n)\right] \le 1/8\right] \ge 1 - \frac{1}{(n+1)^2} \ge 1 - \frac{1}{n^2}.$$

Fix an arbitrary $y \in \mathsf{Support}(\mathcal{C}_n)$ that satisfies the event above. If $y \in L$, then it holds that $L'_{y1} \neq \emptyset$ and $\Pr_{x\sim L'_{y1}\cup\{0^{n+1}\}}[x \in L'_{y1}] \ge 1/2$. Therefore, if $y \in L$, then

$$\begin{aligned}\Pr_{x\sim L_y}\left[\mathsf{pK}^{p'(n)}(x \mid y) > \log|L_y| + \log p'(n)\right] &= \Pr_{x\sim L'_{y1}}\left[x \in L'_{y1} \wedge \mathsf{pK}^{p'(n)}(x \mid y) > \log|L_y| + \log p'(n)\right] \\ &\le 1/8 \cdot (1/2)^{-1} = 1/4.\end{aligned}$$

Hence, we conclude that

$$\Pr_{y\sim\mathcal{C}_n}\left[y \in L \implies \Pr_{x\sim L_y}\left[\mathsf{pK}^{p'(n)}(x \mid y) \le \log|L_y| + \log p'(n)\right] \ge 3/4\right] \ge 1 - \frac{1}{n^2},$$

as desired. □

This completes the proof of Lemma 42. □

We also show Item 2 ⇒ Item 1 in Theorem 31 in a similar way.

Lemma 44 (Item 2 ⇒ Item 1 in Theorem 31)**.** *Suppose that for every samplable distribution families $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ and $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and each $\mathcal{C}_n$ is over $\mathsf{Support}(\mathcal{D}_n^{(2)})$, and for every polynomial q, there exists a polynomial p such that for all large enough $n \in \mathbb{N}$,*

$$\Pr_{y\sim\mathcal{C}_n, x\sim\mathcal{D}_n(\cdot|y)}\left[\mathsf{pK}^{p(n)}(x \mid y) \le \log\frac{1}{\mathcal{D}_n(x \mid y)} + \log p(n)\right] \ge 1 - \frac{1}{q(n)}.$$

Then, $\mathsf{DistNP} \subseteq \mathsf{HeurBPP}$ *holds.*

Proof. Let $(L, \{\mathcal{C}_n\}_{n\in\mathbb{N}})$ be an arbitrary distributional NP problem such that each $\mathcal{C}_n$ is over $\{0,1\}^n$. Let V_L be the polynomial-time computable verifier for L. Without loss of generality (by padding), we assume that the length of every yes instance $y \in L$ and the length of every witness x for $y \in L$ are the same, i.e., $|x| = |y|$ (with respect to V_L). For every $y \in L$, let L_y be a set of V_L-witness x for y.

It suffices to show the following claim by the same argument as in the proof of Lemma 42.

Claim 45. *Under the assumption (i.e., independent average-case conditional coding), there exists a polynomial $p(n)$ such that for every $n \in \mathbb{N}$,*

$$\Pr_{y\sim\mathcal{C}_n}\left[y \in L \implies \Pr_{x\sim L_y}\left[\mathsf{pK}^{p(n)}(x \mid y) \le \log|L_y| + \log p(n)\right] \ge 3/4\right] \ge 1 - \frac{1}{n^2}.$$

Proof of Claim 45. Let $\mathcal{C}' = \{\mathcal{C}'_n\}_{n\in\mathbb{N}}$ be a samplable distribution such that each $\mathcal{C}'_n$ is a distribution of $y1$ for $y \sim \mathcal{C}_{n-1}$. Let $\mathcal{D} = \{\mathcal{D}_n\}_{n\in\mathbb{N}}$ be a samplable distribution such that each $\mathcal{D}_n$ is a distribution of (x, y) determined by $x', y' \sim \{0,1\}^{n-1}$ as

$$(x, y) = \begin{cases} (x'1, y'1) & \text{if } V_L(y', x') = 1 \\ (0^n, y'1) & \text{if } x' = 0^{n-1} \text{ and } V_L(y', x') = 0 \\ (x'0, y'0) & \text{otherwise.} \end{cases}$$

For each $n \in \mathbb{N}$, the distribution $\mathcal{C}'_n$ is over $\{0,1\}^n$, the distribution $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and

$$\mathsf{Support}(\mathcal{C}'_n) \subseteq \{y1 : y \in \{0,1\}^{n-1}\} \subseteq \mathsf{Support}(\mathcal{D}^{(2)}_n).$$

Therefore, by the assumption (i.e., independent average-case conditional coding), there exists a polynomial p such that for all $n \in \mathbb{N}$,

$$\Pr_{y\sim\mathcal{C}'_n, x\sim\mathcal{D}_n(\cdot|y)}\left[\mathsf{pK}^{p(n)}(x \mid y) \leq \log\frac{1}{\mathcal{D}_n(x \mid y)} + \log p(n)\right] \geq 1 - \frac{1}{8n^2}.$$

Fix $n \in \mathbb{N}$ arbitrarily. For every $y \in L$, let $L'_y = \{x1 : x \in L_y\}$. It is not hard to verify that for every $y \in \mathsf{Support}(\mathcal{C}_n)$, the conditional distribution $\mathcal{D}_n(\cdot \mid y1)$ is statistically equivalent to the uniform distribution over $L'_y \cup \{0^{n+1}\}$. Therefore,

$$\Pr_{y\sim\mathcal{C}_n, x\sim L'_y\cup\{0^{n+1}\}}\left[\mathsf{pK}^{p(n+1)}(x \mid y) \leq \log|L'_y \cup \{0^{n+1}\}| + \log p(n+1)\right] \geq 1 - \frac{1}{8(n+1)^2}.$$

Let p' be a polynomial defined as $p'(n) = p(n+1)$. By Markov's inequality and the fact that $|L'_y| = |L_y|$,

$$\Pr_{y\sim\mathcal{C}_n}\left[\Pr_{x\sim L'_y\cup\{0^{n+1}\}}\left[\mathsf{pK}^{p'(n)}(x \mid y) > \log(|L_y|+1) + \log p'(n)\right] \leq 1/8\right] \geq 1 - \frac{1}{(n+1)^2} \geq 1 - \frac{1}{n^2}.$$

Fix an arbitrary $y \in \mathsf{Support}(\mathcal{C}_n)$ that satisfies the event above. If $y \in L$, then it holds that $L'_y \neq \emptyset$, $\Pr_{x\sim L'_y\cup\{0^{n+1}\}}[x \in L'_y] \geq 1/2$, and $|L_y| + 1 \leq 2|L_y|$. Therefore, if $y \in L$, then

$$\begin{aligned}
\Pr_{x\sim L_y}\left[\mathsf{pK}^{p'(n)}(x \mid y) > \log 2|L_y| + \log p'(n)\right] &\leq \Pr_{x\sim L_y}\left[\mathsf{pK}^{p'(n)}(x \mid y) > \log(|L_y|+1) + \log p'(n)\right] \\
&= \Pr_{x\sim L'_y}\left[\mathsf{pK}^{p'(n)}(x \mid y) > \log(|L_y|+1) + \log p'(n)\right] \\
&\leq 1/8 \cdot (1/2)^{-1} = 1/4.
\end{aligned}$$

Hence, we conclude that

$$\Pr_{y\sim\mathcal{C}_n}\left[y \in L \implies \Pr_{x\sim L_y}\left[\mathsf{pK}^{p'(n)}(x \mid y) \leq \log|L_y| + \log p'(n) + 1\right] \geq 3/4\right] \geq 1 - \frac{1}{n^2},$$

as desired. □

This completes the proof of Lemma 44. □

6 NP vs BPP, Worst-Case Conditional Coding and Language Compression

Theorem 46. *The following are equivalent.*

1. $\mathsf{NP} \subseteq \mathsf{BPP}$.

2. **(Worst-Case Conditional Coding)** *For every polynomial-time samplable distribution family $\{\mathcal{D}_n\}_n$, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, there exists a polynomial p such that for all large enough n, and $(x,y) \in \mathsf{Support}(\mathcal{D}_n)$*

$$\mathsf{pK}^{p(n)}(x \mid y) \leq \log \frac{1}{\mathcal{D}_n(x \mid y)} + \log p(n).$$

3. **(Worst-Case Language Compression)** *For every polynomial-time computable set $L \subseteq \{\{0,1\}^n \times \{0,1\}^n\}_n$, there exists a polynomial p such that for all n and all $x, y \in \{0,1\}^n$,*

$$x \in L_y \implies \mathsf{pK}^{p(n)}(x \mid y) \leq \log |L_y| + \log p(n).$$

Moreover, the above holds if we replace pK *with* rK.

Proof. We show below that Item 1 implies Item 2 (via Lemma 47), that Item 2 implies Item 3 (via Lemma 51) and that Item 3 implies Item 1 (via Lemma 52).

To see the "moreover" part, first note that we have $\mathsf{pK}^t(x) \leq \mathsf{rK}^t(x)$ for every $x \in \{0,1\}^*$ and $t \in \mathbb{N}$. Then it suffices to show that assuming $\mathsf{NP} \subseteq \mathsf{BPP}$, there exists a polynomial τ such that $\mathsf{rK}^{\tau(t)}(x) \leq \mathsf{pK}^t(x) + \log \tau(t)$. Indeed, if $\mathsf{NP} \subseteq \mathsf{BPP}$, then $\mathsf{PH} \subseteq \mathsf{BPP}$, which implies $\mathsf{EH} \subseteq \mathsf{BPE}$. Also, it is easy to see that $\mathsf{E}^{\Sigma_3^{\mathsf{P}}}$ contains a language that requires NP-oracle circuits of size at least $2^{\Omega(n)}$, for almost all input lengths n. Therefore, we have $\mathsf{BPE} \not\subseteq \mathsf{i.o.NSIZE}\left[2^{\Omega(n)}\right]$. By [GKLO22, Proposition 66 (Item 3)], this implies that there is a polynomial τ such that $\mathsf{rK}^{\tau(t)}(x) \leq \mathsf{pK}^t(x) + \log \tau(t)$, for every x and t. □

6.1 Conditional Coding from Easiness of NP

Lemma 47 (Item 1 ⇒ Item 2 in Theorem 46)**.** *If $\mathsf{NP} \subseteq \mathsf{BPP}$, then worst-case conditional coding holds.*

We need the following theorem for estimating probabilities of an efficiently samplable distribution.

Theorem 48 ([Sto85])**.** *For every polynomial-time samplable distribution family $\{\mathcal{V}_n\}_n$, where each $\mathcal{V}_n$ is over $\{0,1\}^{\mathsf{poly}(n)}$, there exists a polynomial-time deterministic algorithm A with access to a Σ_2^{P}-oracle such that for every input $z \in \{0,1\}^{\mathsf{poly}(n)}$,*

$$\mathcal{V}_n(z)/1.01 \;\leq\; A(z) \;\leq\; 1.01 \cdot \mathcal{V}_n(z).$$

Lemma 49. *For every polynomial-time samplable distribution family $\{\mathcal{D}_n\}_n$, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, there exists a polynomial-time deterministic algorithm B with access to a Σ_2^{P}-oracle such that for input $(x,y) \in \mathsf{Support}(\mathcal{D}_n)$,*

$$\mathcal{D}_n(x \mid y)/2 \;\leq\; B(x,y) \;\leq\; 2 \cdot \mathcal{D}_n(x \mid y).$$

Proof. Since $D_n(x \mid y) = \mathcal{D}_n(x,y)/\mathcal{D}_n(y)$ for every $(x,y) \in \mathsf{Support}(\mathcal{D}_n)$, where $\mathcal{D}_n(y)$ refers to the marginal distribution of $\mathcal{D}_n$ over the second half of the input, to obtain the desired approximation it is enough to estimate probabilities of two efficiently samplable distributions. The result then follows from Theorem 48. □

We are now ready to prove Lemma 47.

Proof of Lemma 47. Let $\{\mathcal{D}_n\}_n$ be a polynomial-time samplable distribution family. Fix any (x, y) in the support of $\mathcal{D}_n$. Let p be greatest power of two less than $\mathcal{D}_n(x \mid y)$. Note that

$$\frac{\mathcal{D}_n(x \mid y)}{2} \leq p < \mathcal{D}_n(x \mid y). \tag{39}$$

Let

$$S := \{z \colon \mathcal{D}_n(z \mid y) \geq p/4\}.$$

Observe that $|S| \leq 4/p \leq 8/\mathcal{D}_n(x \mid y)$.

Let $m := \lceil \log |S| \rceil$, and let $\mathcal{H}$ be a pairwise independent hash family with $O(n)$-bit seed length mapping n bits to $m + \log n$ bits (see Theorem 11). By Proposition 12, a random hash function in the family isolates x from the other elements in S with probability at least $1 - 1/n$

We now show that

$$\mathsf{pK}^{p(n)}(x \mid y) \leq \log \frac{1}{\mathcal{D}_n(x \mid y)} + \log p(n),$$

where p is a polynomial (that depends on $\{\mathcal{D}_n\}_n$).

Consider a random hash function $H \in \mathcal{H}$, mapping n-bits to $m + \log n$ bits. (A description of m will be hard-coded to the program that reconstructs x. The integer n can be recovered from $|y|$ or hard-coded as well.) If H is good in the sense that it isolates x from the other elements in S, which happens with high probability, then we show that we can use $v := H(x)$ and p to reconstruct x given y. Before arguing this, note that v can be described using

$$m + \log n = \lceil \log |S| \rceil + \log n \leq \log \frac{1}{\mathcal{D}_n(x \mid y)} + O(\log n)$$

bits, while the value p can be described using $\log \log(1/p) \leq \log n^{O(1)} = O(\log n)$ bits because we chose p to be a power of two and $\mathcal{D}_n$ is polynomial-time samplable. (In total, as explained below, this leads to a description of length $\log(1/\mathcal{D}_n(x \mid y)) + O(\log n)$ bits as desired.)

Let B be the polynomial-time algorithm equipped with a Σ_2^{P}-oracle in Lemma 49 that, given $(a, b) \in \mathsf{Support}(\mathcal{D}_n)$, estimates $\mathcal{D}_n(a \mid b)$ up to a (multiplicative) factor of 2.

Claim 50. *If H isolates x from the other elements in S, then x is the unique string in the support of $\mathcal{D}_n(\cdot \mid y)$ that satisfies both $B(x, y) \geq p/2$ and $H(x) = v$.*

Proof of Claim 50. On the one hand, by the correctness of B (recall Lemma 49) and Equation (39), we have

$$B(x, y) \geq \frac{\mathcal{D}_n(x \mid y)}{2} > \frac{p}{2}.$$

On the other hand, suppose there exists some x' in the support of $\mathcal{D}_n(\cdot \mid y)$ such that $x' \neq x$ but $B(x', y) \geq p/2$ and $H(x') = H(x)$. By the correctness of B, this implies that

$$\mathcal{D}_n(x' \mid y) \geq \frac{B(x', y)}{2} \geq \frac{p}{4},$$

which means $x' \in S$. However, since H isolates x from the other elements in S, we cannot have $H(x') = H(x)$. This completes the proof of Claim 50. □

Now we show how to compute x using the description $(v := H(x), p, H, m, y)$, assuming that H isolates x within S. We define the following language L:

$$(i, v, p, H, m, y) \in L \iff \exists\, x \in \mathsf{Support}(\mathcal{D}_n(\cdot \mid y)) \text{ s.t. } B(x, y) \geq p/2,\ H(x) = v \text{ and } x_i = 1.$$

It is easy to see that $L \in \Sigma_3^{\mathsf{P}}$. By our assumption that $\mathsf{NP} \subseteq \mathsf{BPP}$, we get that $\mathsf{PH} \subseteq \mathsf{BPP}$ and consequently $L \in \mathsf{BPP}$. Then we can recover the i-th bit of x by checking whether $(i, v, p, H, m, y) \in L$.

The correctness follows easily from Claim 50. On the other hand, since H is random and does not contribute to the description length, given the discussion above it is easy to see that the desired $\mathsf{pK}^{\mathsf{poly}}(x \mid y)$ upper bound also holds. □

6.2 Conditional Coding implies Language Compression

Lemma 51 (Item 2 ⇒ Item 3 in Theorem 46)**.** *Worst-case conditional coding implies worst-case language compression.*

Proof. Consider the distribution family $\{\mathcal{D}_n\}_n$, where $\mathcal{D}_n$ is given by the following sampling algorithm:

1. Sample a uniformly random $x \sim \{0,1\}^{n-1}$,
2. Sample a uniformly random $y \sim \{0,1\}^{n-1}$,
3. Output $\begin{cases} (x0, y0), & \text{if } x \in L_y \\ (1^n, 1^n), & \text{otherwise.} \end{cases}$

Note that for every $y \in \{0,1\}^n$, $\mathcal{D}_n(\cdot \mid y0)$ is uniformly distributed on $\{x0 : x \in L_y\}$. By worst-case conditional coding, there is a polynomial p_0 such that for every $x \in L_y$

$$\mathsf{pK}^{p_0(n)}(x0 \mid y0) \leq \log \frac{1}{\mathcal{D}_n(x0 \mid y0)} + \log p_0(n) = \log |L_y| + \log p_0(n),$$

which implies

$$\mathsf{pK}^{p(n)}(x \mid y) \leq \log |L_y| + \log p_0(n)$$

for some polynomial p. □

6.3 Easiness of NP from Language Compression

Lemma 52 (Item 3 ⇒ Item 1 in Theorem 46)**.** *If worst-case language compression holds, then* $\mathsf{NP} \subseteq \mathsf{BPP}$.

Proof. Without loss of generality (by padding), we show how to solve every language $L' \in \mathsf{NP}$ where a yes instance $y \in L' \cap \{0,1\}^n$ admits some witness x of length n (with respect to a fixed verifier $V_{L'}$).

Consider the polynomial-time computable set $L := \{(x, y) : V_{L'}(y, x) = 1\}$. Fix any input $y \in L \cap \{0,1\}^n$. Note that $L_y \subseteq \{0,1\}^n$ is the set of $V_{L'}$-witnesses for y. By worst-case language compression, there exist a polynomial p such that for every $x \in L_y$,

$$\mathsf{pK}^{p(n)}(x \mid y) \leq \log |L_y| + \log p(n).$$

Then by Proposition 22, for *each* $x \in L_y$, $\mathsf{USamp}(1^n, 1^{p(n)}, y)$ outputs x with probability at least

$$\frac{1}{O(n \cdot p(n) \cdot |L_y|)}.$$

Hence the probability that $\mathsf{USamp}(1^n, 1^{p(n)}, y)$ outputs *some* $x \in L_y$ is at least

$$|L_y| \cdot \frac{1}{O(n \cdot p(n) \cdot |L_y|)} \geq \frac{1}{O(n \cdot p(n))}.$$

In other words, $\mathsf{USamp}(1^n, 1^{p(n)}, y)$ outputs a witness of y with probability at least $1/\mathsf{poly}(n)$. By standard amplification, this yields an efficient randomized algorithm for solving L' with high probability. □

References

[ABK+06] Eric Allender, Harry Buhrman, Michal Koucký, Dieter van Melkebeek, and Detlef Ronneburger. Power from random strings. *SIAM J. Comput.*, 35(6):1467–1493, 2006.

[AF09] Luis Filipe Coelho Antunes and Lance Fortnow. Worst-case running times for average-case algorithms. In *Conference on Computational Complexity* (CCC), pages 298–303, 2009.

[All21] Eric Allender. Vaughan Jones, Kolmogorov Complexity, and the new complexity landscape around circuit minimization. *New Zealand journal of mathematics*, 52:585–604, 2021.

[BCGL92] Shai Ben-David, Benny Chor, Oded Goldreich, and Michael Luby. On the theory of average case complexity. *J. Comput. Syst. Sci.*, 44(2):193–219, 1992.

[BFL01] Harry Buhrman, Lance Fortnow, and Sophie Laplante. Resource-bounded Kolmogorov complexity revisited. *SIAM J. Comput.*, 31(3):887–905, 2001.

[BLM00] Harry Buhrman, Sophie Laplante, and Peter B. Miltersen. New bounds for the language compression problem. In *Conference on Computational Complexity* (CCC), pages 126–130, 2000.

[BLvM05] Harry Buhrman, Troy Lee, and Dieter van Melkebeek. Language compression and pseudorandom generators. *Comput. Complex.*, 14(3):228–255, 2005.

[BT06] Andrej Bogdanov and Luca Trevisan. Average-case complexity. *Found. Trends Theor. Comput. Sci.*, 2(1), 2006.

[GK22] Halley Goldberg and Valentine Kabanets. A simpler proof of the worst-case to average-case reduction for polynomial hierarchy via symmetry of information. *Electron. Colloquium Comput. Complex.*, 7:1–14, 2022.

[GKLO22] Halley Goldberg, Valentine Kabanets, Zhenjian Lu, and Igor C. Oliveira. Probabilistic kolmogorov complexity with applications to average-case complexity. In *Computational Complexity Conference* (CCC), pages 16:1–16:60, 2022.

[GM84] Shafi Goldwasser and Silvio Micali. Probabilistic encryption. *J. Comput. Syst. Sci.*, 28(2):270–299, 1984.

[Gol90] Oded Goldreich. A note on computational indistinguishability. *Inf. Process. Lett.*, 34(6):277–281, 1990.

[Gol01] Oded Goldreich. *The Foundations of Cryptography - Volume 1: Basic Techniques*. Cambridge University Press, 2001.

[HILL99] Johan Håstad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby. A pseudorandom generator from any one-way function. *SIAM J. Comput.*, 28(4):1364–1396, 1999.

[Hir18] Shuichi Hirahara. Non-black-box worst-case to average-case reductions within NP. In *Symposium on Foundations of Computer Science* (FOCS), pages 247–258, 2018.

[Hir21] Shuichi Hirahara. Average-case hardness of NP from exponential worst-case hardness assumptions. In *Symposium on Theory of Computing* (STOC), pages 292–302, 2021.

[Hir22a] Shuichi Hirahara. Meta-computational average-case complexity: A new paradigm toward excluding heuristica. *Bull. EATCS*, 136, 2022.

[Hir22b] Shuichi Hirahara. Symmetry of information from meta-complexity. In *Computational Complexity Conference* (CCC), pages 26:1–26:41, 2022.

[IL89] Russell Impagliazzo and Michael Luby. One-way functions are essential for complexity based cryptography (extended abstract). In *Symposium on Theory of Computing* (STOC), pages 230–235, 1989.

[IL90] Russell Impagliazzo and Leonid A. Levin. No better ways to generate hard NP instances than picking uniformly at random. In *Symposium on Theory of Computing* (STOC), pages 812–821, 1990.

[Imp95] Russell Impagliazzo. A personal view of average-case complexity. In *Proceedings of the Tenth Annual Structure in Complexity Theory Conference*, pages 134–147, 1995.

[IRS21] Rahul Ilango, Hanlin Ren, and Rahul Santhanam. Hardness on any samplable distribution suffices: New characterizations of one-way functions by meta-complexity. *Electron. Colloquium Comput. Complex.*, page 82, 2021.

[Ko91] Ker-I Ko. On the complexity of learning minimum time-bounded Turing machines. *SIAM J. Comput.*, 20(5):962–986, 1991.

[Kol65] Andrey N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems of information transmission*, 1(1):1–7, 1965.

[Kol68] Andrey N. Kolmogorov. Several theorems about algorithmic entropy and algorithmic amount of information (a talk at a Moscow Math. Soc. meeting on 10/31/67). In Usp. Mat. Nauk, volume 23, page 201, 1968.

[Lee06] Troy Lee. *Kolmogorov complexity and formula lower bounds*. PhD thesis, University of Amsterdam, 2006.

[LM93] Luc Longpré and Sarah Mocas. Symmetry of information and one-way functions. *Inf. Process. Lett.*, 46(2):95–100, 1993.

[LO21] Zhenjian Lu and Igor C. Oliveira. An efficient coding theorem via probabilistic representations and its applications. In *International Colloquium on Automata, Languages, and Programming* (ICALP), pages 94:1–94:20, 2021.

[LO22] Zhenjian Lu and Igor C. Oliveira. Theory and applications of probabilistic Kolmogorov complexity. *Bull. EATCS*, 137, 2022.

[Lon86] Luc Longpré. *Resource bounded Kolmogorov complexity, a link between computational complexity and information theory.* PhD thesis, Cornell University, 1986.

[LOS21] Zhenjian Lu, Igor C. Oliveira, and Rahul Santhanam. Pseudodeterministic algorithms and the structure of probabilistic time. In *Symposium on Theory of Computing* (STOC), pages 303–316, 2021.

[LOZ22] Zhenjian Lu, Igor Carboni Oliveira, and Marius Zimand. Optimal coding theorems in time-bounded Kolmogorov complexity. In *International Colloquium on Automata, Languages, and Programming* (ICALP), pages 92:1–92:14, 2022.

[LP20] Yanyi Liu and Rafael Pass. On one-way functions and Kolmogorov complexity. In *Symposium on Foundations of Computer Science* (FOCS), pages 1243–1254, 2020.

[LP21] Yanyi Liu and Rafael Pass. On the possibility of basing cryptography on EXP≠BPP. In *International Cryptology Conference* (CRYPTO), pages 11–40, 2021.

[LP22] Yanyi Liu and Rafael Pass. On one-way functions from NP-complete problems. In *Computational Complexity Conference* (CCC), pages 36:1–36:24, 2022.

[LR05] Troy Lee and Andrei E. Romashchenko. Resource bounded symmetry of information revisited. *Theor. Comput. Sci.*, 345(2-3):386–405, 2005.

[LV19] Ming Li and Paul M. B. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications, 4th Edition.* Texts in Computer Science. Springer, 2019.

[LW95] Luc Longpré and Osamu Watanabe. On symmetry of information and polynomial time invertibility. *Inf. Comput.*, 121(1):14–22, 1995.

[Nao91] Moni Naor. Bit commitment using pseudorandomness. *J. Cryptol.*, 4(2):151–158, 1991.

[Oli19] Igor C. Oliveira. Randomness and intractability in Kolmogorov complexity. In *International Colloquium on Automata, Languages, and Programming* (ICALP), pages 32:1–32:14, 2019.

[OPS19] Igor C. Oliveira, Ján Pich, and Rahul Santhanam. Hardness magnification near state-of-the-art lower bounds. In *Computational Complexity Conference* (CCC), pages 27:1–27:29, 2019.

[Rom90] John Rompel. One-way functions are necessary and sufficient for secure signatures. In *Symposium on Theory of Computing* (STOC), pages 387–394, 1990.

[Ron04] Detlef Ronneburger. *Kolmogorov Complexity and Derandomization*. PhD thesis, Rutgers University, 2004.

[RRV02] Ran Raz, Omer Reingold, and Salil P. Vadhan. Extracting all the Randomness and Reducing the Error in Trevisan's Extractors. *J. Comput. Syst. Sci.*, 65(1):97–128, 2002.

[RS21] Hanlin Ren and Rahul Santhanam. Hardness of KT characterizes parallel cryptography. In *Computational Complexity Conference* (CCC), pages 35:1–35:58, 2021.

[Sip83] Michael Sipser. A complexity theoretic approach to randomness. In *Symposium on Theory of Computing* (STOC), pages 330–335, 1983.

[Sto85] Larry J. Stockmeyer. On approximation algorithms for #P. *SIAM J. Comput.*, 14(4):849–861, 1985.

[SUV17] Alexander Shen, Vladimir A. Uspensky, and Nikolay Vereshchagin. *Kolmogorov complexity and algorithmic randomness*. American Mathematical Society, 2017.

[TV07] Luca Trevisan and Salil P. Vadhan. Pseudorandomness and average-case complexity via uniform reductions. *Computational Complexity*, 16(4):331–364, 2007.

[Vad12] Salil P. Vadhan. Pseudorandomness. *Foundations and Trends in Theoretical Computer Science*, 7(1-3):1–336, 2012.

[Wat22] Osamu Watanabe. Personal Communication, 2022.

[ZL70] Alexander K. Zvonkin and Leonid A. Levin. The complexity of finite objects and the algorithmic concepts of randomness and information. *UMN (Russian Math. Surveys)*, 25(6):83–124, 1970.

A Infinitely-Often Characterization via Extrapolation

In this section, we consider infinitely-often versions of the properties of Kolmogorov complexity stated in Theorem 13. Instead of saying that the property holds for all n, the infinitely-often version only requires that it holds for infinitely many n.

Theorem 53. *The following are equivalent.*

1. *One-way functions do not exist.*
2. *Infinitely-often strong average-case conditional coding holds.*
3. *Infinitely-often weak average-case conditional coding holds.*
4. *Infinitely-often strong average-case language compression holds.*
5. *Infinitely-often weak average-case language compression holds.*
6. *Infinitely-often strong average-case symmetry information holds.*
7. *Infinitely-often weak average-case symmetry information holds.*

Proof. The proof is analogous to that of Theorem 13. In fact, it is easy to see that all the proofs in Section 3 work even in the infinitely-often setting, except for the one in Section 3.1 that shows strong average-case conditional coding holds assuming infinitely-often secure one-way function do not exist. In Lemma 56, we show that if (almost-everywhere secure) one-way function do not exist, then infinitely-often strong average-case conditional coding holds. □

We need the following theorem regarding *extrapolation*.

Theorem 54 (Implicit in [IL90, IL89])**.** *If almost everywhere (resp. infinitely-often) secure one-way functions do not exist, then for every samplable distribution family $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, there exists a probabilistic polynomial-time algorithm Ext such that for all $\varepsilon^{-1}, \delta^{-1} \in \mathbb{N}$ and for infinitely many (resp. all) n,*

$$\Pr_{y\sim\mathcal{D}_n^{(2)}}\left[\mathsf{L}_1\left(\mathsf{Ext}(y; 1^{\varepsilon^{-1}}, 1^{\delta^{-1}}), \mathcal{D}_n(\cdot \mid y)\right) \le \varepsilon\right] \ge 1-\delta,$$

where $\mathcal{D}_n^{(2)}$ denotes the marginal distribution of the second element of $\mathcal{D}_n$.

We also need the following variant of the coding theorem for pK.

Theorem 55 ([LOZ22])**.** *Let $\mathcal{D}$ be a distribution over $\{0,1\}^n$. Suppose $\mathcal{D}$ can be sampled using a randomized program $M_{\mathcal{D}}$ that runs in $\mathsf{poly}(n)$ time. Then there exists a polynomial p such that for every $x \in \{0,1\}^n$,*

$$\mathsf{pK}^{p(n)}(x \mid M_{\mathcal{D}}) \le \log\frac{1}{\mathcal{D}(x)} + \log p(n).$$

Lemma 56 (Item 1 ⇒ Item 2 in Theorem 53)**.** *If one-way functions do not exist, then infinitely-often strong average-case conditional coding holds. That is, for every polynomial-time samplable distribution family $\{\mathcal{D}_n\}_n$, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and for every polynomial q, there exists a polynomial p such that for infinitely many n,*

$$\Pr_{(x,y)\sim\mathcal{D}_n}\left[\mathsf{pK}^{p(n)}(x \mid y) \le \log\frac{1}{\mathcal{D}_n(x \mid y)} + \log p(n)\right] \ge 1 - \frac{1}{q(n)}.$$

Proof. Let $\mathcal{D}_n^{(2)}$ denote the marginal distribution of the second element of $\mathcal{D}_n$.

Since there is no one-way function, by Theorem 54, there exists a polynomial-time randomized algorithm Ext such that for infinitely many n,

$$\Pr_{y\sim\mathcal{D}_n^{(2)}}\left[\mathsf{L}_1\left(\mathsf{Ext}\left(y; 1^{4q(n)}, 1^{2q(n)}\right), \mathcal{D}_n(\cdot \mid y)\right) \le \frac{1}{4q(n)}\right] \ge 1 - \frac{1}{2q(n)}.$$

Fix any n for which Ext satisfies the above condition, and let $\mathcal{D}'_y$ be the distribution given by $\mathsf{Ext}\left(y; 1^{4q(n)}, 1^{2q(n)}\right)$. Then we have with probability at least $1 - 1/(2q(n))$ over $y \sim \mathcal{D}_n^{(2)}$,

$$\mathsf{L}_1\left(\mathcal{D}'_y, \mathcal{D}_n(\cdot \mid y)\right) \le \frac{1}{4q(n)}.$$

Fix any y such that the above holds, we claim the following.

Claim 57. *We have*

$$\mathop{\mathbf{Pr}}_{x\sim\mathcal{D}_n(\cdot|y)}\left[\mathcal{D}'_y(x) \geq \frac{\mathcal{D}_n(x \mid y)}{4q(n)}\right] \geq 1 - \frac{1}{2q(n)}$$

Proof. Since $\mathsf{L}_1\left(\mathcal{D}'_y, \mathcal{D}_n(\cdot \mid y)\right) \leq 1/(4q(n))$ hods for y, we have

$$\mathop{\mathbf{Pr}}_{x\sim\mathcal{D}(\cdot|y)}\left[\frac{\mathcal{D}_n(x \mid y)}{\mathcal{D}'_y(x)} > 4q(n)\right] \leq \frac{1}{4q(n)} + \mathop{\mathbf{Pr}}_{x\sim\mathcal{D}'_y}\left[\frac{\mathcal{D}_n(x \mid y)}{\mathcal{D}'_y(x)} > 4q(n)\right]. \tag{40}$$

Also,

$$\mathop{\mathbf{E}}_{x\sim\mathcal{D}'_y}\left[\frac{\mathcal{D}_n(x \mid y)}{\mathcal{D}'_y(x)}\right] = \sum_{x\in\mathsf{Support}(\mathcal{D}'_y)} \mathcal{D}'_y(x) \cdot \frac{\mathcal{D}_n(x \mid y)}{\mathcal{D}'_y(x)} = \sum_{x\in\mathsf{Support}(\mathcal{D}'_y)} \mathcal{D}_n(x \mid y) \leq 1.$$

Then By Markov's inequality, we obtain that

$$\mathop{\mathbf{Pr}}_{x\sim\mathcal{D}'_y}\left[\frac{\mathcal{D}_n(x \mid y)}{\mathcal{D}'_y(x)} > 4q(n)\right] < \frac{\mathbf{E}_{x\sim\mathcal{D}'_y}\left[\frac{\mathcal{D}_n(x|y)}{\mathcal{D}'_y(x)}\right]}{4q(n)} \leq \frac{1}{4q(n)}. \tag{41}$$

Combining Equation (40) and Equation (41) completes the proof of the claim. □

Note that D'_y is $\mathsf{poly}(n)$-time samplable given y, since Ext runs in polynomial time. By the coding theorem for pK (Theorem 55), there exists some polynomial p_0 such that for every $x \in \{0,1\}^n$,

$$\mathsf{pK}^{p_0(n)}(x \mid y) \leq \log \frac{1}{\mathcal{D}'_y(x)} + \log p_0(n).$$

Combining this with Claim 57, we get that with probability at least $1-1/(2q(n))$ over $x \sim \mathcal{D}_n(\cdot \mid y)$,

$$\mathsf{pK}^{p_0(n)}(x \mid y) \leq \log \frac{1}{\mathcal{D}'_y(x)} + \log p_0(n) \leq \log \frac{1}{\mathcal{D}(x \mid y)} + \log(4q(n)) + \log p_0(n).$$

Therefore, by a union bound, with probability at least $1 - 1/q(n)$ over $(x, y) \sim \mathcal{D}_n$, we have

$$\mathsf{pK}^{p(n)}(x \mid y) \leq \log \frac{1}{\mathcal{D}(x \mid y)} + \log p(n),$$

where p is some polynomial. □

B Summary of the Dualities Between Complexity Theory and Kolmogorov Complexity

In this section, we summarize the dualities between complexity theory and the preservation of key properties of Kolmogorov complexity in the time-bounded setting. For simplicity, we focus on our results for pK^t complexity in the polynomial-time regime.

	$\iff$	$\implies$
$\mathsf{NP} \subseteq \mathsf{BPP}$	Worst-Case Conditional Coding Worst-Case Language Compression	Worst-Case SoI [5]
$\mathsf{DistNP} \subseteq \mathsf{HeurBPP}$	Independent Average-Case Conditional Coding Independent Average-Case Language Compression	Independent Average-Case SoI
$\not\exists$ i.o.OWFs	Average-Case Conditional Coding Average-Case Language Compression Average-Case SoI	

Conditional Coding:

1. **(Worst-Case Conditional Coding)** There exists a polynomial p such that for all n, and $(x, y) \in \mathsf{Support}(\mathcal{D}_n)$
$$\mathsf{pK}^{p(n)}(x \mid y) \leq \log \frac{1}{\mathcal{D}_n(x \mid y)} + \log p(n).$$

2. **(Independent Average-Case Conditional Coding)** Let $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ and $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$ be samplable distribution families, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and each $\mathcal{C}_n$ is over the support of the second half of $\mathcal{D}_n$. For every polynomial q, there exists a polynomial p such that for all n,
$$\Pr_{y\sim\mathcal{C}_n, x\sim\mathcal{D}_n(\cdot|y)}\left[\mathsf{pK}^{p(n)}(x \mid y) \leq \log \frac{1}{\mathcal{D}_n(x \mid y)} + \log p(n)\right] \geq 1 - \frac{1}{q(n)}.$$

3. **(Average-Case Conditional Coding)** Let $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ be samplable distribution family, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$. For every polynomial q, there exists a polynomial p such that for all n,
$$\Pr_{(x,y)\sim\mathcal{D}_n}\left[\mathsf{pK}^{p(n)}(y \mid x) \leq \log \frac{1}{\mathcal{D}_n(x \mid y)} + \log p(n)\right] \geq 1 - \frac{1}{q(n)}.$$

Language Compression:

1. **(Worst-Case Language Compression)** Let $L \subseteq \{\{0,1\}^n \times \{0,1\}^n\}_{n\in\mathbb{N}}$ be a polynomial-time computable set. There exists a polynomial p such that for all n, and $y \in \{0,1\}^n$,
$$y \in L_y \implies \mathsf{pK}^{p(n)}(x \mid y) \leq \log |L_y| + \log p(n).$$

[6] In fact, in order to establish worst-case SoI it suffices to assume $\mathsf{DistNP} \subseteq \mathsf{AvgBPP}$. See [Hir22b, GK22, GKLO22].

2. **(Independent Average-Case Language Compression)** Let $L \subseteq \{\{0,1\}^n \times \{0,1\}^n\}_{n\in\mathbb{N}}$ be a recursively enumerable set. Let $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ and $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$ be samplable distribution families, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and each $\mathcal{C}_n$ is over the support of the second half of $\mathcal{D}_n$. For every polynomial q, there exists a polynomial p such that for all n,

$$\mathop{\mathbf{Pr}}_{y\sim\mathcal{C}_n, x\sim\mathcal{D}_n(\cdot|y)}\left[y \in L_y \implies \mathsf{pK}^{p(n)}(x \mid y) \leq \log|L_y| + \log p(n)\right] \geq 1 - \frac{1}{q(n)}.$$

3. **(Average-Case Language Compression)** Let $L \subseteq \{\{0,1\}^n \times \{0,1\}^n\}_{n\in\mathbb{N}}$ be a recursively enumerable set. Let $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ be samplable distribution family, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$. For every polynomial q, there exists a polynomial p such that for all n,

$$\mathop{\mathbf{Pr}}_{(x,y)\sim\mathcal{D}_n}\left[x \in L_y \implies \mathsf{pK}^{p(n)}(x \mid y) \leq \log|L_y| + \log p(n)\right] \geq 1 - \frac{1}{q(n)}.$$

Symmetry of Information:

1. **(Worst-Case Symmetry of Information)** There exists a polynomial p such that for all $t \geq 2n$ and for all n and all $x, y \in \{0,1\}^n$,

$$\mathsf{pK}^t(x,y) \geq \mathsf{pK}^{p(t)}(x \mid y) + \mathsf{pK}^{p(t)}(y) - \log p(t).$$

2. **(Independent Average-Case Symmetry of Information)** Let $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ and $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$ be samplable distribution families, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$, and each $\mathcal{C}_n$ is over the support of the second half of $\mathcal{D}_n$. For every polynomial q, there exists a polynomial p such that for every computable time bound $t\colon \mathbb{N} \to \mathbb{N}$ with $t(n) \geq p(n)$ and for all n,

$$\mathop{\mathbf{Pr}}_{y\sim\mathcal{C}_n, x\sim\mathcal{D}_n(\cdot|y)}\left[\mathsf{pK}^{t(n)}(x,y) \geq \mathsf{pK}^{t(n)}(x \mid y) + \mathsf{pK}^{t(n)}(y) - \log t(n)\right] \geq 1 - \frac{1}{q(n)}.$$

3. **(Average-Case Symmetry of Information)** Let $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ be samplable distribution family, where each $\mathcal{D}_n$ is over $\{0,1\}^n \times \{0,1\}^n$. For every polynomial q, there exists a polynomial p such that for every computable time bound $t\colon \mathbb{N} \to \mathbb{N}$ with $t(n) \geq p(n)$ and for all n,

$$\mathop{\mathbf{Pr}}_{(x,y)\sim\mathcal{D}_n}\left[\mathsf{pK}^{t(n)}(x,y) \geq \mathsf{pK}^{t(n)}(x \mid y) + \mathsf{pK}^{t(n)}(y) - \log t(n)\right] \geq 1 - \frac{1}{q(n)}.$$

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il