Electronic Colloquium on Computational Complexity, Report No. 127 (2022)

Kolmogorov Complexity Characterizes Statistical Zero Knowledge

Eric Allender

Rutgers University, NJ, USA

Shuichi Hirahara

National Institute of Informatics, Japan

Harsha Tirumala

Rutgers University, NJ, USA

Abstract

We show that a decidable promise problem has a non-interactive statistical zero-knowledge proof system if and only if it is randomly reducible to a promise problem for Kolmogorov-random strings, with a superlogarithmic additive approximation term. This extends recent work by Saks and Santhanam (CCC 2022). We build on this to give new characterizations of Statistical Zero Knowledge SZK, as well as the related classes $\mathsf{NISZK_L}$ and $\mathsf{SZK_L}$.

2012 ACM Subject Classification Complexity Classes; Problems, reductions and completeness; Circuit complexity

Keywords and phrases Kolmogorov Complexity, Interactive Proofs

Funding *Eric Allender*: Supported in part by NSF Grants CCF-1909216 and CCF-1909683.
Shuichi Hirahara: Supported in part by JST, PRESTO Grant Number JPMJPR2024, Japan
Harsha Tirumala: Supported in part by NSF Grants CCF-1909216 and CCF-1909683.

1 Introduction

In this paper, we give the first non-trivial characterization of a computational complexity class in terms of reducibility to the Kolmogorov random strings.

Some readers may be surprised that this is possible. After all, the set of Kolmogorov random strings is undecidable, and undecidable sets typically do not figure prominently in complexity-theoretic investigations.[1] But what does it mean to be reducible to the Kolmogorov-random strings? Let us consider the prefix-free Kolmogorov complexity K (which is one of the most-studied types of Kolmogorov complexity), and recall that different universal Turing machines U give a slightly different Kolmogorov measure K_U. Then if we say "A is reducible to the K-random strings" we probably mean that A is reducible to the K_U random strings, no matter which universal machine U we are using. But it turns out that the class of languages that can be solved in polynomial time with an oracle that returns $K_U(q)$ for any query q—*regardless* of which universal machine U is used—is a complexity class that contains NEXP and lies in EXPSPACE [23, 13, 29].[2] There has been substantial interest in obtaining a precise understanding of which problems can be reduced in this way to the Kolmogorov complexity function under different notions of reducibility [2, 3, 9, 7, 8, 12, 13, 14, 20, 23, 30, 29, 32, 33, 46], but until now, no previously studied complexity class has been characterized in this way, with the exception of P [8, 46]. (The

[1] We do wish to highlight the recent work of Ilango, Ren, and Santhanam [37], who related the existence of one-way functions to the *average case* complexity of computing Kolmogorov complexity.

[2] More specifically, it is shown in [13] that all decidable sets with this property lie in EXPSPACE, and it is shown in [23] that there are no undecidable sets with this property. Hirahara shows in [30] that every set in $\mathsf{EXP}^{\mathsf{NP}}$ (and hence in NEXP) has this property.

ISSN 1433-8092

characterizations of P obtained in this way can be viewed as showing that certain limited polynomial-time reductions are useless when using the Kolmogorov complexity function as an oracle.)

Faced with this lack of success, it was proposed in [3, Open Question 4.8] that a more successful approach might be to consider reductions to *approximations* to the Kolmogorov complexity function. Saks and Santhanam [46] took the first significant step in this direction, by showing the following results:

▶ **Theorem 1** (Saks & Santhanam [46]). **1.** *Although (by the work of Hirahara [30]) every language in* $\mathsf{EXP}^{\mathsf{NP}}$ *is reducible in deterministic polynomial time to any function that differs from* K *by at most an additive* $O(\log n)$ *term, no decidable language outside of* P *is reducible to all approximations to* K *that differ by an error margin* $e(n) = \omega(\log n)$ *via an "honest" deterministic polynomial-time nonadaptive reduction.*

2. *Although (by the work of Hirahara [29]) every language in* NEXP *is reducible via randomized nonadaptive reductions to any function that differs from* K *by at most an additive* $O(\log n)$ *term, no decidable language outside of* $\mathsf{AM} \cap \mathsf{coAM}$ *is reducible to all approximations to* K *that differ by an error margin* $e(n) = \omega(\log n)$ *via an "honest" probabilistic polynomial-time nonadaptive reduction.*

3. *No decidable language outside of* SZK *is randomly m-reducible to each* $\omega(\log n)$ *approximation to the* K*-random strings.*

This is not the first time that the complexity class SZK (for *Statistical Zero Knowledge* has arisen in the context of investigations relating to Kolmogorov complexity. In particular, SZK and its "non-interactive" subclass NISZK have been studied in connection with a version of time-bounded Kolmogorov complexity, which in turn is studied because of its connection with the Minimum Circuit Size Problem (MCSP) [11, 14]. These problems lie at the heart of what has come to be called *meta-complexity*: the study of the computational difficulty of answering questions about complexity.

Allender [2] proposed an intriguing research program towards the $\mathsf{P} = \mathsf{BPP}$ conjecture. The class P can be characterized by the class of languages reducible to the set of Kolmogorov-random strings under polynomial-time disjunctive truth-table reductions [8]. Similarly, he conjectured that BPP can also be characterized by polynomial-time truth-table reductions to the set of Kolmogorov-random strings, and envisioned that such a completely new characterization of complexity classes would give us new insights into BPP, especially from the perspective of computability theory. Unfortunately, his conjecture was refuted by Hirahara [30] under a plausible complexity-theoretic assumption.

In this paper, we show that SZK, NISZK and their logspace variants SZK_L and $\mathsf{NISZK}_\mathsf{L}$ can be characterized by reductions to approximations to the Kolmogorov complexity function. We envision that our new characterization of these complexity classes would improve our understanding of zero knowledge interactive proof systems in future. Zero knowledge interactive proof systems have many applications in cryptographic protocols, and they have been studied very widely. We refer the reader to the excellent survey by Vadhan for more background [47]. For our purposes, the complexity classes of interest to us (SZK, NISZK, SZK_L, and $\mathsf{NISZK}_\mathsf{L}$) can be defined in terms of their complete problems. But first, we need to define some basic notions and provide some background.

2 Preliminaries

We assume familiarity with basic complexity classes such as P, L, and AC^0; we view these as classes of *functions*, as well as of *languages*. We also will refer to the class of functions computed in NC^0, where each output bit depends on at most $O(1)$ input bits. For circuit complexity classes such as NC^0, and AC^0, by default we assume that the circuit families are "First-Order-uniform" as discussed in [5, 18, 38]. This coincides with Dlogtime-uniform AC^0, and what one might call "Dlogtime-uniform AC^0-uniform" NC^0. (We refer the reader to [49] for more background on circuit uniformity.) When we need to refer to *nonuniform* circuit complexity, we will be explicit.

All of these classes give rise to restrictions of Karp reducibility $\leq_{\mathrm{m}}^{\mathsf{P}}$, such as $\leq_{\mathrm{m}}^{\mathsf{L}}, \leq_{\mathrm{m}}^{\mathsf{AC}^0}$, and $\leq_{\mathrm{m}}^{\mathsf{NC}^0}$. We will also discuss *projections* ($\leq_{\mathrm{m}}^{\mathsf{proj}}$), which are $\leq_{\mathrm{m}}^{\mathsf{NC}^0}$ reductions in which each output bit depends on at most one input bit. Thus projections are computed by circuits consisting of constants, wires, and NOT gates.

A *promise problem* A is a pair of disjoint sets (Y_A, N_A) of YES instances and NO instances, respectively. A *solution* to a promise problem is any set B such that $Y_A \subseteq B$ and $N_A \subseteq \overline{B}$. A *don't-care instance* of A is any string that is not in $Y_A \cup N_A$. A *language* can be viewed as a promise problem that has no don't-care instances.

We say that a promise problem $A = (Y, N)$ is *decidable* if Y and N are decidable sets. Observe that if $B = (Y', N')$ with $Y' \subseteq Y$ and $N' \subseteq N$, then any solution to A is also a solution to B. Such subproblems of decidable promise problems are intuitively "decidable", but are not necessarily decidable according to our definition. Since there are uncountably many subsets of Y and N for any nontrivial promise problem, clearly not every intuitively "decidable" promise problem can be decidable.

When defining reductions between two promise problems A and B, there are two options. Either

- for every solution S to B there is a reduction from A to S, or
- there is a reduction that correctly decides A when given any solution S for B.

As it turns out, these two notions are equivalent [28, 43]. Thus we shall always use the second approach, when defining notions of reducibility between promise problems.

We assume that the reader is familiar with Kolmogorov complexity; more background on this topic can be found in references such as [41, 25]. Briefly, $K_U(x|y) = \min\{|d| : U(d, y) = x\}$, and $K_U(x) = K(x|\lambda)$ where λ denotes the empty string.[3] Although this definition depends on the choice of the Turing machine U, we pick some "universal" machine U' and define $K(x|y)$ to be $K_{U'}(x|y)$; for every machine U, there is a constant c such that $K(x|y) \leq K_U(x|y) + c$. One important non-trivial fact regarding Kolmogorov complexity is known as *symmetry of information*:

▶ **Theorem 2.** *(Symmetry of Information)*

$$K(x, y) = K(x) + K(x|y) \pm O(\log(K(x, y))).$$

Let $\widetilde{R}_K$ be the promise problem $(Y_{\widetilde{R}_K}, N_{\widetilde{R}_K})$ where $Y_{\widetilde{R}_K}$ contains all strings y such that $K(y) \geq |y|/2$ and the NO instances $N_{\widetilde{R}_K}$ consists of those strings y where $K(y) \leq |y|/2 - e(|y|)$ for some approximation error term $e(n)$, where $e(n) = \omega(\log n)$ and $e(n) = n^{o(1)}$. All of our theorems hold for any $e(n)$ in this range. We will sometimes assume that $e(n)$ is computable in AC^0, which is true for most approximation terms of interest.

---

[3] This is actually the definition of so-called "plain" Kolmogorov complexity, although the letter K is traditionally used for the "prefix-free" Kolmogorov complexity. These two measures differ by at most a logarithmic term, and our theorems hold for either measure. For simplicity, we have presented the simpler definition.

Since the approximation error $e(n)$ is superlogarithmic, it is worth noting that $\widetilde{R}_K$ can be defined equivalently either in terms of prefix-free or plain Kolmogorov complexity (because these two measures are within an additive logarithmic term of each other).

Any *language* that is reducible to $\widetilde{R}_K$ via any of the reducibilities that we consider is decidable, by a theorem of [23]. However, it is not known whether this carries over in any meaningful way to promise problems.

The reader may wonder about the justification for the threshold $K(y) \geq |y|/2$ in the definition of $\widetilde{R}_K$. The following proposition indicates that, for large error bounds $e(n)$, using a larger threshold reduces to $\widetilde{R}_K$. Later, we show a related result for smaller thresholds.

▶ **Proposition 3.** *Let $A = (Y, N)$ be the promise problem where $Y = \{y : K(y) \geq t(|y|)\}$ for some AC^0-computable threshold $t(n) \geq \frac{n}{2}$, and where $N = \{y : K(y) \leq t(|y|) - |y|^\epsilon\}$ for some $1 > \epsilon > 0$. Then $A \leq_{\mathrm{m}}^{\mathsf{NC}^0} \widetilde{R}_K$.*

Proof. Let $\delta = \frac{\epsilon}{2}$. Given an instance y of length n (for all large n), in AC^0 we can find the least integer $i < n$ such that $2t(n) - n + 5\log n + (2(2n)^\delta - n^\epsilon) \leq i \leq 2t(n) - n - 3\log n$.

Let $z = y0^i$. Then $K(z) \leq K(y) + 2\log i + O(1)$. Similarly, $K(y) \leq K(z) + 2\log i + O(1)$, and hence $K(z) \geq K(y) - 2\log i - O(1)$.

Thus if $y \in Y$, then $K(z) \geq t(n) - 2\log i - O(1) > (t(n) - \frac{n}{2}) + \frac{n}{2} - 3\log n \geq \frac{n+i}{2} = \frac{|z|}{2}$. And if $y \in N$, then $K(z) \leq t(n) - n^\epsilon + 2\log i + O(1) < (t(n) - \frac{n}{2}) + \frac{n}{2} - n^\epsilon + 2\log i + O(1) \leq \frac{n+i}{2} - (n+i)^\delta = \frac{|z|}{2} - |z|^\delta < \frac{|z|}{2} - e(|z|)$.

Thus $y \in Y$ implies $z \in Y_{\widetilde{R}_K}$ and $y \in N$ implies $z \in N_{\widetilde{R}_K}$. ◀

Randomized reductions play a central role in the results that we will be presenting. Here is the basic definition:

▶ **Definition 4.** *A promise problem $A = (Y, N)$ is $\leq_{\mathrm{m}}^{\mathsf{RP}}$-reducible to $B = (Y', N')$ with threshold θ if there is a polynomial p and a deterministic Turing machine M running in time p such that*

- *$x \in Y$ implies $\Pr_{r \in \{0,1\}^{p(|x|)}}[M(x, r) \in Y'] \geq \theta$.*
- *$x \in N$ implies $\Pr_{r \in \{0,1\}^{p(|x|)}}[M(x, r) \in N'] = 1$.*

Randomized reductions were introduced by Adleman and Manders, as a probabilistic generalization of $\leq_{\mathrm{m}}^{\mathsf{P}}$ reducibility[4] [1]. They used the threshold $\theta = \frac{1}{2}$. One of the most important applications of randomized reductions is the theorem of Valiant and Vazirani [48], where they showed that SAT reduces to Unique Satisfiability (USAT) via a randomized reduction, with threshold $\theta = \frac{1}{4n}$.[5] The reader may expect that—as is so often the case with probabilistic notions in computational complexity theory—the choice of threshold is arbitrary, and can be changed with no meaningful consequences. However, this does not appear to be true; we refer the reader to the work of Chang, Kadin, and Rohatgi [24] for a discussion of this point. As they point out, different thresholds are appropriate in different situations. If $A \leq_{\mathrm{m}}^{\mathsf{RP}} B$ with threshold $\frac{1}{4n}$ (for instance), where the set $\mathrm{OR}_B = \{(x_1, \ldots, x_k) : \exists i, x_i \in B\} \leq_{\mathrm{m}}^{\mathsf{P}} B$, then it is indeed true that $A \leq_{\mathrm{m}}^{\mathsf{RP}} B$ with threshold $1 - \frac{1}{2^n}$ [24]. But Chang, Kadin, and Rohatgi point out that it is far from clear that USAT has this property. We are concerned here with problems that are $\leq_{\mathrm{m}}^{\mathsf{RP}}$-reducible to $\widetilde{R}_K$; just as in the case with randomized reductions to USAT, we must be careful about which threshold θ we choose. For the remainder of

[4] We assume that the reader is familiar with Karp reducibility $\leq_{\mathrm{m}}^{\mathsf{P}}$.
[5] Recently, there have also been several papers showing that certain meta-complexity-theoretic problems are NP-complete under randomized reductions, including [10, 31, 34, 35, 36, 42, 44].

this paper, we will use the threshold $\theta = 1 - \frac{1}{n^{\omega(1)}}$. (For a discussion of why we select this threshold, see Remark 12.)

The following proposition is the counterpart to Proposition 3, for thresholds smaller than $\frac{n}{2}$.

▶ **Proposition 5.** *Let $A = (Y, N)$ be the promise problem where $Y = \{y : K(y) \geq t(|y|)\}$ for some polynomial-time computable threshold $t(n) \leq \frac{n}{2}$, and where $N = \{y : K(y) \leq t(|y|) - |y|^\epsilon\}$ for some $1 > \epsilon > 0$. Then $A \leq_m^{\mathsf{RP}} \widetilde{R}_K$.*

Proof. Given an instance y of length n (for all large n), in polynomial time we can find the least integer $i < n$ such that $2t(n) - 2n^\epsilon + 2e(3n) + 4\log n \leq i \leq 2t(n) - e(n) - 2c\log n$ (for a constant c that will be picked later).

Pick a random string r of length n. Let $z = yr0^i$. Then $K(z) \leq K(y) + 2\log i + |r|$. Also, by symmetry of information, $K(z) \geq K(yr0^i|y0^i) + K(y0^i) - c' \log n$ (for some fixed constant c', and hence with probability at least $1 - \frac{1}{n^{\omega(1)}}$, $K(z) \geq (n - \frac{e(n)}{2}) + K(y) - c\log n$ (for some fixed c, which is the constant c that we use above in defining i).

Thus if $y \in Y$, then with high probability $K(z) \geq t(n) + (n - \frac{e(n)}{2}) - c\log n > n + \frac{i}{2} = \frac{|z|}{2}$. And if $y \in N$, then $K(z) \leq (t(n) - n^\epsilon) + 2\log i + |r| \leq n + \frac{i}{2} - e(3n) \leq \frac{|z|}{2} - e(|z|)$.

Thus $y \in Y$ implies $z \in Y_{\widetilde{R}_K}$ (with probability $\geq 1 - \frac{1}{n^{\omega(1)}}$), and $y \in N$ implies $z \in N_{\widetilde{R}_K}$. ◀

We will also need a "two-sided error" version of random reducibility, analogous to the relationship between RP and BPP.

▶ **Definition 6.** *A promise problem $A = (Y, N)$ is $\leq_m^{\mathsf{BPP}}$-reducible to $B = (Y', N')$ with threshold $\theta > \frac{1}{2}$ if there is a polynomial p and a deterministic Turing machine M running in time p such that*

- *$x \in Y$ implies $\Pr_{r \in \{0,1\}^{p(|x|)}}[M(x, r) \in Y'] \geq \theta$.*
- *$x \in N$ implies $\Pr_{r \in \{0,1\}^{p(|x|)}}[M(x, r) \in N'] \geq \theta$.*

The complexity classes SZK (Statistical Zero Knowledge) and NISZK (Non-Interactive Statistical Zero Knowledge) are defined in terms of interactive proof protocols (with a *Prover* interacting with a probabilistic polynomial-time *Verifier*, together with a *Simulator* that can produce a distribution on transcripts that is statistically close to the distribution on messages that would be exchanged by the prover and the verifier on YES instances. But for our purposes, it will suffice (and be simpler) to present alternative definitions of these classes, in terms of their standard complete problems.

▶ **Definition 7** (Promise-EA)**.** *Let a circuit $C : \{0,1\}^m \to \{0,1\}^n$ represent a probability distribution X on $\{0,1\}^n$ induced by the uniform distribution on $\{0,1\}^m$. We define Promise-EA to be the promise problem*

$$Y_{\mathsf{EA}} = \{(C, k) \mid H(X) > k + 1\}$$
$$N_{\mathsf{EA}} = \{(C, k) \mid H(X) < k - 1\}$$

where $H(X)$ denotes the entropy of X.

▶ **Theorem 8** ([27])**.** EA *is complete for* NISZK *under $\leq_m^{\mathsf{P}}$ reductions.*

We will actually take this as a definition; we say that (Y, N) is in NISZK if and only if $(Y, N) \leq_m^{\mathsf{P}} \mathsf{EA}$.

▶ **Definition 9** (Promise-SD). SD *(Statistical Difference) is the promise problem*

$$Y_{\mathsf{SD}} = \left\{(C, D) \;\middle|\; \Delta(C, D) > \frac{2}{3}\right\},$$
$$N_{\mathsf{SD}} = \left\{(C, D) \;\middle|\; \Delta(C, D) < \frac{1}{3}\right\}.$$

where $\Delta(C, D)$ denotes the statistical distance between the distributions represented by the circuits C and D.

▶ **Theorem 10** ([45]). SD *is complete for* SZK *under* $\leq^{\mathsf{P}}_{\mathrm{m}}$ *reductions.*

Thus we will define SZK to be the class of promise problems (Y, N) such that $(Y, N) \leq^{\mathsf{P}}_{\mathrm{m}} \mathsf{SD}$.

3 A New Characterization of NISZK

We are now ready to present the characterization of NISZK by reductions to the set of Kolmogorov-random strings.

▶ **Theorem 11.** *The following are equivalent, for any decidable promise problem A:*

1. $A \in \mathsf{NISZK}$.
2. $A \leq^{\mathsf{RP}}_{\mathrm{m}} \widetilde{R}_K$.
3. $A \leq^{\mathsf{BPP}}_{\mathrm{m}} \widetilde{R}_K$.

Proof. In order to show that $A \in \mathsf{NISZK}$ implies $A \leq^{\mathsf{RP}}_{\mathrm{m}} \widetilde{R}_K$, it suffices to reduce the NISZK-complete problem EA to $\widetilde{R}_K$. This follows easily from the proof given in [14, Corollary 18], combined with [27, Lemma 3.2]. Specifically, Lemma 3.2 in [27] shows that the following promise problem is complete for NISZK: All instances are of the form $(C, 1^s)$, where C is a circuit with m inputs and n outputs, representing a distribution (also denoted C) on $\{0,1\}^n$. $(C, 1^s)$ is a YES instance if C has statistical distance at most 2^{-s} from the uniform distribution on $\{0,1\}^n$. $(C, 1^s)$ is in the set of NO instances if the support of C has size at most 2^{n-s}. Furthermore, the reduction g from EA to A has the property that the parameter s is at least n^{ϵ} for some constant $\epsilon > 0$. Also, it is observed in Lemma 4.1 of [27] that the mapping $(C, 1^s) \mapsto (C, n-3)$ (i.e., the mapping that leaves the circuit C unchanged) is a reduction from A to EA. To summarize: these results from [27] show that the following subproblem of EA is also hard for NISZK under $\leq^{\mathsf{P}}_{\mathrm{m}}$ reductions: The set Y of YES instances consists of pairs $(C, n-3)$ where the entropy of C is greater than $n-2$, and the set N of NO instances consists of pairs $(C, n-3)$ where the support of C has size at most $2^{n-n^{\epsilon}}$.

Corollary 18 of [14] states that every promise problem in NISZK reduces to the problem of computing the time-bounded Kolmogorov complexity KT via a probabilistic reduction that makes at most one query along any computation path. But here we observe that the same approach can be used to obtain a $\leq^{\mathsf{RP}}_{\mathrm{m}}$ reduction to $\widetilde{R}_K$. Corollary 18 of [14] relies on the proof of Theorem 17 in the same paper (which in turn relies on the techniques of [16]), which presents a probabilistic algorithm M that takes an instance $(C, n-3)$ of EA (as described above), and constructs a string y that is the concatenation of t random samples from C (i.e., $y = C(r_1)C(r_2)\ldots C(r_t)$ for uniformly chosen random strings $r_1, \ldots, r_t$, for some polynomially-large t). Lemma 16 of [14] shows that, with probability exponentially close to 1, if $(C, n-3)$ is a YES instance of EA, then the time-bounded Kolmogorov complexity $\mathsf{KT}(y)$ is greater than a threshold θ of the form $\theta = t(n-2) - t^{1-\alpha}$ for some constant $\alpha > 0$. In the argument of [14, Theorem 17], t can be chosen to be an arbitrarily large polynomial

n^k. Thus we have $\theta > n^k(n-3)$ for all large n, and hence for all large YES instances we have $\mathsf{KT}(y) > n^k(n-3) = \ell - \ell^\delta$ for some $\delta < 1$, where $|y| = tn = \ell$. The focus of [14] was on the measure KT, but (as was previously observed in [4, Theorem 1]) the analysis in Lemma 16 carries over unchanged to the setting of non-resource-bounded Kolmogorov complexity K. Thus, with high probability, the probabilistic routine, when given a YES instance of EA, produces a string y where $K(y) \geq |y| - |y|^\delta$.

On the other hand, if $(C, n-3)$ is a NO instance, then the support of C has size at most 2^{n-n^ϵ}, and thus any string z in the support of C has $K(z|C) \leq n - n^\epsilon + O(1)$. Thus any string y that is produced by M in this case has $K(y) \leq t(n-n^\epsilon) + |C| + O(1) = n^k(n-n^\epsilon)) + |C| + O(1)$. Since $t = n^k$ was chosen to be large (with respect to the length of the input instance $(C, n-3)$), we may assume $|C| < n^{k+\epsilon} - 4n^k$. Thus if $(C, n-3)$ is any large NO instance, we have $K(y) < n^k(n-4) = \ell - \ell^{\delta'}$ for some $\delta' > \delta$. To summarize, with probability 1, the probabilistic routine, when given a NO instance of EA, produces a string y where $K(y) \geq |y| - |y|^{\delta'} \geq (|y| - |y|^\delta) - |y|^\epsilon$ for some $\epsilon > 0$. We can now conclude that $\mathsf{EA} \leq_{\mathrm{m}}^{\mathsf{RP}} \widetilde{R}_K$ by appealing to Proposition 3.

To complete the proof of the theorem, we need to show that if A is any decidable promise problem that has a randomized poly-time m-reduction ($\leq_{\mathrm{m}}^{\mathsf{BPP}}$) with error $1/n^{\omega(1)}$ to the promise problem $\widetilde{R}_K$ then $A \in \mathsf{NISZK}$. This was essentially shown by Saks and Santhanam [46, Theorem 39], but we present a complete argument here. Let M be the probabilistic machine that computes this $\leq_{\mathrm{m}}^{\mathsf{BPP}}$ reduction.

Let $y = f(x, r) \in \{0,1\}^m$ denote the output that M produces, where x is an instance of A and r denotes the randomness used in the reduction. (As in the proof of [46, Theorem 39], we may assume that, for each x, all outputs of the form $f(x, r)$ have the same length.) Given an $x \in \{0,1\}^n$, observe that there is a polynomial-sized circuit C_x such that $C_x(r) = f(x, r)$. According to the correctness of the reduction, we have

$$x \in Y_A \Rightarrow \Pr_r[M(x,r) \in Y_{\widetilde{R}_K}] \geq 1 - 1/n^{\omega(1)} \text{ and}$$

$$x \in N_A \Rightarrow \Pr_r[M(x,r) \in N_{\widetilde{R}_K}] \geq 1 - 1/n^{\omega(1)}.$$

In other words, if x is a YES instance, then $K(y) \geq |y|/2$ with probability at least $1 - 1/n^{\omega(1)}$ and if x is a NO instance, then $K(y) \leq |y|/2 - e(|y|)$ with probability at least $1 - 1/n^{\omega(1)}$. (Recall that $e(n)$ is the error term in the approximation $\widetilde{R}_K$.) We will now show that there is an entropy threshold that separates these two distributions, which will provide an NISZK upper bound on resolving A.

Claim: If x is a YES instance, then the entropy of the distribution $C_x(r)$ is at least $m/2 - e(m)/2 + 1$ and if x is a NO instance, then the entropy of $C_x(r)$ is at most $m/2 - e(m)/2 - 1$.

We first show that if the claim holds, then $A \in \mathsf{NISZK}$. Let $k = m/2 - e(m)/2$. The reduction given above reduces membership in A to the Entropy Approximation (EA) problem on the circuit description C_x with threshold k. Given x, we can compute the map $x \mapsto C_x$ in time $n^{O(1)}$. Recall that EA is compete for NISZK. Since NISZK is closed under $\leq_{\mathrm{m}}^{\mathsf{P}}$ reductions, we can conclude that $A \in \mathsf{NISZK}$.

Proof of claim:

Assume not and let x be the lexicographically first string that violates the above claim (for some length n). Since the reduction is a computable function, and since A is a decidable promise problem, $K(x) = O(\log n)$. We have the following two cases to consider:

Case 1 - x is a YES instance: From the correctness of the reduction we have that with probability $1 - 1/n^{\omega(1)}$ the output y is a string with Kolmogorov complexity at least $|m|/2$.

Since x is a violator, we have $H(C_x(r)) < k + 1 = m/2 - e(m)/2 + 1$.
On one hand, the distribution $C_x(r)$ has large enough probability mass on the high-complexity strings. On the other hand, we have that since x is a low-complexity string itself, the elements of $C_x(r)$ with highest mass can be identified by short descriptions. This leads to a contradiction of simultaneously having large enough mass on the low and the high K-complexity strings.
Let t be the entropy of the distribution $C_x(r)$. Let $Y = \{y_1 \dots y_{2^{t+\log m}}\}$ be the heaviest elements (in terms of probability mass) of $C_x(r)$ in decreasing order. Conditioned on x, the K complexity of any of these strings y_i is at most $t + O(\log m)$. Since $K(x) = O(\log n) = O(\log m)$, we have $K(y_i) \leq t + O(\log m) < m/2$. Next, we will show that there is at least mass $\frac{1}{m}$ on these strings within $C_x(r)$. This will contradict the correctness of the reduction for $x \in L$ since it cannot output strings with K complexity at most $|m|/2$ with probability $1/n^{\Omega(1)}$.
Assume not, i.e., the mass on elements of Y is at most $\frac{1}{m}$. Observe that elements of $Sup(C_x(r)) - Y$ have mass no more than $2^{-(t+\log m)}$ each. Then, the contribution to entropy by these elements is at least $(1 - 1/m)(t + \log m) > t$ (which is a contradiction).

Case 2 - x is a NO instance: From the correctness of the reduction we have that with probability at least $1 - 1/n^{\omega(1)}$ the output $f(x, r)$ is a string with K complexity at most $m/2 - e(m)$. Since x is a violator, we also have $H(C_x(r)) > k - 1 = m/2 - e(m)/2 - 1$.
We claim that the following holds:

$$\Pr_{y \sim f(x,r)}[K(y) > m/2 - e(m)] \geq 1/m.$$

Assume not. Then, the entropy of $f(x, r)$ is at most $(1/m)(m) + (1 - 1/m)(m/2 - e(m)) \leq m/2 - e(m) + 1 < m/2 - e(m)/2 - 1$, which contradicts the lower bound on the entropy of $f(x, r)$ above.
Since the claim holds, with probability at least $1/m$ the output of the reduction is not an element of the set $N_{\tilde{R}_K}$. Thus, the reduction fails with probability $1/n^{\Omega(1)}$.

◀

▶ Remark 12. The proof of the preceding theorem illustrates why we define the error threshold in our randomized reductions to be $\frac{1}{n^{\omega(1)}}$. If we assumed that A were $\leq_{\mathrm{m}}^{\mathsf{BPP}}$-reducible to $\widetilde{R}_K$ with an inverse polynomial threshold (say $q(n)^{-1}$), then (as in the proof of [46, Theorem 39] we may modify the reduction so that the length of each output produced has length $Q(n) = \omega(q(n))$ (by padding with some uniformly-random bits). For strings x that are NO instances of A, when the reduction to $\widetilde{R}_K$ fails with probability $1/q(n)$, our calculation of the entropy of C_x will involve a term of $\frac{1}{q(n)}Q(n)$ (because the queries made in this case can have nearly $Q(n)$ bits of entropy). This is more than the entropy gap between the distributions corresponding to the YES and NO outputs.

▶ Remark 13. Although our focus in this paper is in $\widetilde{R}_K$, we note that one can also define an analogous problem $\widetilde{R}_{\mathsf{KT}}$ in terms of the time-bounded measure KT. The approach used in Theorem 11 also shows that every problem in NISZK is $\leq_{\mathrm{m}}^{\mathsf{BPP}}$ reducible to $\widetilde{R}_{\mathsf{KT}}$, although we do not know how to show hardness under $\leq_{\mathrm{m}}^{\mathsf{RP}}$ reductions. (A random sample from the low-entropy distribution is guaranteed to always have low K-complexity, but the tools of [14, 16] only guarantee that the output has low KT-complexity with high probability.)

4 More Powerful Reductions

Just as $\leq_m^{\mathsf{RP}}$ and $\leq_m^{\mathsf{BPP}}$ reducibilities generalize the familiar $\leq_m^{\mathsf{P}}$ (Karp) reducibility to the setting of probabilistic computation, so also are there probabilistic generalizations of deterministic non-adaptive reductions (also known as truth-table reductions). Before presenting these probabilistic generalizations, let us review the previously-studied deterministic non-adaptive reducibilities that are relevant for this investigation. Some of them may be unfamiliar to the reader.

Ladner, Lynch, and Selman [40] considered several possible ways to define polynomial-time versions of the truth-table reducibility that had been studied in computability theory, before settling on the definition of $\leq_{tt}^{\mathsf{P}}$ reducibility below. They considered only reductions between *languages*; the corresponding generalization to *promise problems* is due to [45]. In order to state this generalization formally, let us define the characteristic function χ_A of a promise problem $A = (Y, N)$ to take on the following values in three-valued logic:

- If $x \in Y$, then $\chi_A(x) = 1$.
- If $x \in N$, then $\chi_A(x) = 0$.
- If $x \notin (Y \cup N)$, then $\chi_A(x) = *$.

A Boolean circuit with n variables, when given an assignment in $\{0, 1, *\}^n$, can be evaluated using the usual rules of three-valued logic. (See, e.g., [45, Definition 4.6].)

▶ **Definition 14.** *Let $A = (Y, N)$ and $B = (Y', N')$ be promise problems. We say $A \leq_{tt}^{\mathsf{P}} B$ if there is a function f computable in polynomial time, such that, for all x, $f(x)$ is of the form $(C, z_1, z_2, \ldots, z_k)$ where C is a Boolean circuit with k input variables, and $(z_1, \ldots, z_k)$ is a list of queries, with the property that*

- *If $x \in Y$, then $C(\chi_B(z_1), \ldots, \chi_B(z_k)) = 1$.*
- *If $x \in N$, then $C(\chi_B(z_1), \ldots, \chi_B(z_k)) = 0$.*

This definition ensures that the circuit C, viewed as an ordinary circuit in 2-valued logic, correctly decides membership for all $x \in (Y \cup N)$ when given any solution S for B as an oracle.

If C is a Boolean formula, instead of a circuit, then one obtains the so-called "Boolean formula reducibility" (denoted by $A \leq_{bf}^{\mathsf{P}} B$), which was discussed in [40] and studied further in [39, 22]. (See also [21, 6].)

▶ **Theorem 15.** $\mathsf{SZK} = \{A : A \leq_{bf}^{\mathsf{P}} \mathsf{EA}\}$.

Proof. $\mathsf{EA} \in \mathsf{NISZK} \subseteq \mathsf{SZK}$. Sahai and Vadhan [45, Corollary 4.14] showed that SZK is closed under NC^1-truth-table reductions, but the proof carries over immediately to $\leq_{bf}^{\mathsf{P}}$ reductions. Thus $\{A : A \leq_{bf}^{\mathsf{P}} \mathsf{EA}\} \subseteq \mathsf{SZK}$. The other inclusion was shown in [27, Proposition 5.4]. ◀

Notably, it is still an open question if SZK is closed under $\leq_{tt}^{\mathsf{P}}$ reducibility.

Our characterization of SZK in terms of reductions to $\widetilde{R}_K$ relies on the following probabilistic generalization of $\leq_{bf}^{\mathsf{P}}$:

▶ **Definition 16.** *Let $A = (Y, N)$ and $B = (Y', N')$ be promise problems. We say $A \leq_{bf}^{\mathsf{BPP}} B$ with threshold $\theta > \frac{1}{2}$ if there are functions f and g computable in* **deterministic** *polynomial time, and a polynomial p, such that, for all x, $f(x)$ is a Boolean formula C (with $k = |x|^{O(1)}$ variables), with the property that*

- *If* $x \in Y$, *then* $C(\chi_{g,B}(x,1),\ldots,\chi_{g,B}(x,k)) = 1$,
- *If* $x \in N$, *then* $C(\chi_{g,B}(x,1),\ldots,\chi_{g,B}(x,k)) = 0$,

where

- $\chi_{g,B}(x,i) = 1$ *if* $\Pr_{r\in\{0,1\}^{p(|x|)}}[g(x,i,r) \in Y'] \geq \theta$
- $\chi_{g,B}(x,i) = 0$ *if* $\Pr_{r\in\{0,1\}^{p(|x|)}}[g(x,i,r) \in N'] \geq \theta$
- $\chi_{g,B}(x,i) = *$ *otherwise.*

Intuitively, $\leq_{\text{bf}}^{\mathsf{BPP}}$ reductions generalize $\leq_{\text{bf}}^{\mathsf{P}}$ reductions, in that the queries are now generated probabilistically, and the probability that any query returns a definite YES or NO answer is bounded away from $\frac{1}{2}$.

The following proposition is immediate from the definitions.

▶ **Proposition 17.** *If* $A \leq_{\text{bf}}^{\mathsf{P}} B$ *and* $B \leq_{\text{m}}^{\mathsf{BPP}} C$ *with threshold* θ, *then* $A \leq_{\text{bf}}^{\mathsf{BPP}} C$ *with threshold* θ.

▶ **Corollary 18.** $\mathsf{SZK} \subseteq \{A : A \leq_{\text{bf}}^{\mathsf{BPP}} \widetilde{R}_K\}$ *with threshold* $1 - \frac{1}{n^{\omega(1)}}$.

Proof. Immediate from Theorem 15 and Theorem 11. ◀

There are (at least) three other variants of probabilistic nonadaptive reducibility that we should mention. The first of these is the notion that goes by the name "nonadaptive BPP reducibility" or "randomized nonadaptive reductions" in work such as [46, 14, 19] and elsewhere.

▶ **Definition 19.** *Let* $A = (Y, N)$ *and* $B = (Y', N')$ *be promise problems. We say* $A \leq_{\text{tt}}^{\mathsf{BPP}} B$ *if there are a function* f *computable in polynomial time and a polynomial* p *such that, for all* x *and all* r *of length* $p(|x|)$, $f(x,r)$ *is of the form* $(C, z_1, z_2, \ldots, z_k)$ *where* C *is a Boolean circuit with* k *input variables, and* $(z_1,\ldots,z_k)$ *is a list of queries, with the property that*

- *If* $x \in Y$, *then* $\Pr_r[C(\chi_B(z_1),\ldots,\chi_B(z_k)) = 1] \geq \frac{2}{3}$.
- *If* $x \in N$, *then* $\Pr_r[C(\chi_B(z_1),\ldots,\chi_B(z_k)) = 0 \geq \frac{2}{3}$.

(The threshold $\frac{2}{3}$ *can be replaced by any threshold between* n^{-k} *and* 2^{-n^k}, *by the usual method of taking the majority vote of several independent trials.)*

Saks and Santhanam showed that if $A \leq_{\text{tt}}^{\mathsf{BPP}} \widetilde{R}_K$ via a reduction that satisfies an additional "honesty" condition, then $A \in \mathsf{AM} \cap \mathsf{coAM}$ [46]. The most important ways in which $\leq_{\text{bf}}^{\mathsf{BPP}}$ and $\leq_{\text{tt}}^{\mathsf{BPP}}$ reducibility differ from each other, are (1) in $\leq_{\text{bf}}^{\mathsf{BPP}}$ reducibility, the query evaluation is performed by a Boolean formula, instead of a circuit, and (2) in $\leq_{\text{tt}}^{\mathsf{BPP}}$ reducibility, the circuit that is chosen, to do the evaluation, depends on the choice of random bits, whereas in $\leq_{\text{bf}}^{\mathsf{BPP}}$ reducibility, the formula is chosen deterministically. Making different choices in these two dimensions gives rise to two other notions:

▶ **Definition 20.** *Let* $A = (Y, N)$ *and* $B = (Y', N')$ *be promise problems. We say* $A \leq_{\text{rbf}}^{\mathsf{BPP}} B$ *if there are a function* f *computable in polynomial time and a polynomial* p *such that, for all* x *and all* r *of length* $p(|x|)$, $f(x,r)$ *is of the form* $(C, z_1, z_2, \ldots, z_k)$ *where* C *is a Boolean formula with* k *input variables, and* $(z_1,\ldots,z_k)$ *is a list of queries, with the property that*

- *If* $x \in Y$, *then* $\Pr_r[C(\chi_B(z_1),\ldots,\chi_B(z_k)) = 1] \geq \frac{2}{3}$.
- *If* $x \in N$, *then* $\Pr_r[C(\chi_B(z_1),\ldots,\chi_B(z_k)) = 0] \geq \frac{2}{3}$.

(The threshold $\frac{2}{3}$ *can be replaced by any threshold between* n^{-k} *and* 2^{-n^k}, *simply by incorporating a Boolean formula that takes the majority vote of several independent trials.).*

The notation $\leq_{\mathrm{rbf}}^{\mathsf{BPP}}$ is intended to suggest "random Boolean formula", since the Boolean formula is chosen randomly.

▶ **Definition 21.** *Let $A = (Y, N)$ and $B = (Y', N')$ be promise problems. We say $A \leq_{\mathrm{circ}}^{\mathsf{BPP}} B$ with threshold $\theta > \frac{1}{2}$ if there are functions f and g computable in* **deterministic** *polynomial time, and a polynomial p, such that, for all x, $f(x)$ is a Boolean circuit (with $k = |x|^{O(1)}$ variables), with the property that*

- *If $x \in Y$, then $C(\chi_{g,B}(x, 1), \ldots, \chi_{g,B}(x, k)) = 1$,*
- *If $x \in N$, then $C(\chi_{g,B}(x, 1), \ldots, \chi_{g,B}(x, k)) = 0$,*

where

- *$\chi_{g,B}(x, i) = 1$ if $\Pr_{r \in \{0,1\}^{p(|x|)}}[g(x, i, r) \in Y'] \geq \theta$*
- *$\chi_{g,B}(x, i) = 0$ if $\Pr_{r \in \{0,1\}^{p(|x|)}}[g(x, i, r) \in N'] \geq \theta$*
- *$\chi_{g,B}(x, i) = *$ otherwise.*

We show in this paper that SZK is the class of problems $\leq_{\mathrm{bf}}^{\mathsf{BPP}}$ reducible to $\widetilde{R}_K$. We are not able to show that the class of problems $\leq_{\mathrm{rbf}}^{\mathsf{BPP}}$ reducible to $\widetilde{R}_K$ is contained in SZK, although we do observe that SZK is closed under this type of reducibility.

▶ **Theorem 22.** $\mathsf{SZK} = \{A : A \leq_{\mathrm{rbf}}^{\mathsf{BPP}} \mathsf{EA}\}$.

Proof. The inclusion of SZK in $\{A : A \leq_{\mathrm{rbf}}^{\mathsf{BPP}} \mathsf{EA}\}$ is immediate from Theorem 15. For the other direction, let $A \leq_{\mathrm{rbf}}^{\mathsf{BPP}} \mathsf{EA}$. Thus there are a function f computable in polynomial time, and a polynomial p such that, for all x and all r of length $p(|x|)$, $f(x, r)$ is of the form $(C, z_1, z_2, \ldots, z_k)$, where evaluating the Boolean formula $C(\chi_B(z_1), \ldots, \chi_B(z_k))$ gives a correct answer for all $x \in Y \cup N$ with error at most 2^{-n^2}. Here is a zero-knowledge interactive protocol for A. The verifier sends a random string r to the prover. The prover and the verifier can each compute $f(x, r) = (C, z_1, z_2, \ldots, z_k)$, and then (as in [45, Corollary 4.14], compute an instance (D, E) of SD such that (D, E) is a YES instance of SD if $C(\chi_B(z_1), \ldots, \chi_B(z_k)) = 1$, and (D, E) is a NO instance of SD if $C(\chi_B(z_1), \ldots, \chi_B(z_k)) = 0$. The prover and the verifier can then run the SZK protocol for the SD instance (D, E). The verifier clearly accepts each YES instance with high probability, and cannot be convinced to accept any NO instance with more than negligible probability. The simulator, given input x, will generate the string r uniformly at random, and then compute $f(x, r)$ and compute the instance (D, E) as above, and then produce the transcript that is produced by the SD simulator on input (D, E). It is straightforward to observe that, if $x \in Y$, then this distribution is very close to the distribution induced by the honest prover and verifier. ◀

5 A New Characterization of SZK

▶ **Theorem 23.** *The following are equivalent, for any decidable promise problem A:*

1. *$A \in \mathsf{SZK}$.*
2. *$A \leq_{\mathrm{bf}}^{\mathsf{BPP}} \widetilde{R}_K$ with threshold $1 - \frac{1}{n^{\omega(1)}}$.*

Proof. Corollary 18 states that all problems in SZK $\leq_{\mathrm{bf}}^{\mathsf{BPP}}$-reduce to $\widetilde{R}_K$. Thus we need only show the converse containment. Let $A \leq_{\mathrm{bf}}^{\mathsf{BPP}} \widetilde{R}_K$. As in the proof of Theorem 11, we will build circuits $C_{x,i}(r)$ that model the computation that produces the i^{th} query that is asked on input x, when using random bits r. As in the proof of Theorem 11, we claim that if a $1 - \frac{1}{n^{\omega(1)}}$ fraction of the strings of the form $C_{x,i}(r)$ are in $Y_{\widetilde{R}_K}$, then $C_{x,i}$ represents a distribution with entropy at least $m/2 - e(m)/2 + 1$, and if a $1 - \frac{1}{n^{\omega(1)}}$ fraction of the strings of the form $C_{x,i}(r)$ are in $N_{\widetilde{R}_K}$, then $C_{x,i}$ represents a distribution with entropy at most $m/2 - e(m)/2 - 1$. Indeed, the proof is essentially identical. Assume that there are infinitely many x that are not don't care instances, where replacing the $\widetilde{R}_K$ oracle with the EA oracle does not yield the correct answer. Given n, we can find the lexicographically-least string x of length n for which the reduction fails. Since the reduction fails, there must be some i such that the i^{th} query in the formula yields the wrong answer. Thus, given (n, i), we can find x and build the circuit $C_{x,i}$ of Kolmogorov complexity $O(\log n)$ that yields a correct answer when given $\widetilde{R}_K$ as an oracle, but fails when queries are made to EA instead. The analysis is identical to the argument in the proof of Theorem 11. ◀

We have nothing to say, regarding the problems that are reducible to $\widetilde{R}_K$ via $\leq_{\text{tt}}^{\mathsf{BPP}}$ or $\leq_{\text{rbf}}^{\mathsf{BPP}}$ reductions, other than to refer to the $\mathsf{AM} \cap \mathsf{coAM}$ upper bound provided by Saks and Santhanam [46]. We do have a somewhat better bound to report, regarding $\leq_{\text{circ}}^{\mathsf{BPP}}$ reducibility.

▶ **Theorem 24.** *The following are equivalent, for any decidable promise problem A:*

1. $A \leq_{\text{circ}}^{\mathsf{BPP}} \widetilde{R}_K$ *with threshold* $1 - \frac{1}{n^{\omega(1)}}$.
2. $A \leq_{\text{tt}}^{\mathsf{P}} \mathsf{EA}$.
3. $A \leq_{\text{tt}}^{\mathsf{P}} B$ *for some* $B \in \mathsf{SZK}$.

Proof. Items 2 and 3 are equivalent, by Theorem 15. Similarly, if $A \leq_{\text{tt}}^{\mathsf{P}} B$ for some $B \in \mathsf{SZK}$, then we know that $A \leq_{\text{tt}}^{\mathsf{P}} B \leq_{\text{bf}}^{\mathsf{BPP}} \widetilde{R}_K$. The composition of a $\leq_{\text{tt}}^{\mathsf{P}}$ reduction with a $\leq_{\text{bf}}^{\mathsf{BPP}}$ reduction is clearly a $\leq_{\text{circ}}^{\mathsf{BPP}}$ reduction. Finally, the proof of the remaining implication follows along the same lines as the proof of Theorem 23. ◀

6 Less Powerful Reductions

The standard complete problems EA and SD remain complete for NISZK and SZK, respectively, even under more restrictive reductions such as $\leq_{\text{m}}^{\mathsf{L}}$ and $\leq_{\text{m}}^{\mathsf{NC}^0}$. In this section, we show that it is worthwhile considering probabilistic versions of $\leq_{\text{m}}^{\mathsf{L}}, \leq_{\text{m}}^{\mathsf{AC}^0}$ and $\leq_{\text{m}}^{\mathsf{NC}^0}$ reducibility to $\widetilde{R}_K$.

▶ **Definition 25.** *For a class $\mathcal{C}$, a promise problem $A = (Y, N)$ is $\leq_{\text{m}}^{\mathsf{R}\mathcal{C}}$-reducible to $B = (Y', N')$ with threshold θ if there are a function $f \in \mathcal{C}$ and a polynomial p such that*

- $x \in Y$ *implies* $\Pr_{r \in \{0,1\}^{p(|x|)}}[f(x,r) \in Y'] \geq \theta$.
- $x \in N$ *implies* $\Pr_{r \in \{0,1\}^{p(|x|)}}[f(x,r) \in N'] = 1$.

A is $\leq_{\text{m}}^{\mathsf{BP}\mathcal{C}}$-reducible to B with threshold θ if there are a function $f \in \mathcal{C}$ and a polynomial p such that

- $x \in Y$ *implies* $\Pr_{r \in \{0,1\}^{p(|x|)}}[f(x,r) \in Y'] \geq \theta$.
- $x \in N$ *implies* $\Pr_{r \in \{0,1\}^{p(|x|)}}[f(x,r) \in N'] \geq \theta$.

We are particularly interested in the cases $\mathcal{C} = \mathsf{L}, \mathcal{C} = \mathsf{AC}^0$, and $\mathcal{C} = \mathsf{NC}^0$. Note especially that, in the definitions of $\leq_{\text{m}}^{\mathsf{RL}}$ and $\leq_{\text{m}}^{\mathsf{BPL}}$, the logspace computation has full (two-way) access to the random bits r. This is consistent with the way that probabilistic logspace computation is used in the context of the "verifier" and "simulator" in the complexity classes $\mathsf{SZK}_{\mathsf{L}}$ and $\mathsf{NISZK}_{\mathsf{L}}$ [26, 14].

$\mathsf{SZK}_{\mathsf{L}}$, the "logspace version" of SZK, was introduced in [26], primarily as a tool to discuss the complexity of problems involving distributions realized by extremely limited circuits (such as NC^0 circuits). It is shown in [26] that $\mathsf{SZK}_{\mathsf{L}}$ contains many of the problems of cryptographic significance that lie in SZK. $\mathsf{NISZK_L}$ was introduced in [14] as the "non-interactive" counterpart to $\mathsf{SZK_L}$, by analogy with NISZK, primarily as a tool to investigate the complexity of computing time-bounded Kolmogorov complexity. It was subsequently studied in [15], where it was shown to be robust to several changes to the definition. It is shown in [26, 14] that complete problems for $\mathsf{SZK_L}$ and $\mathsf{NISZK_L}$ arise by considering restrictions of the standard complete problems for SZK and NISZK where the distributions under consideration are represented either by branching programs (in $\mathsf{EA_{BP}}$), or by NC^0 circuits where each output bit depends on at most 4 input bits (in $\mathsf{SD_{NC^0}}$ and $\mathsf{EA_{NC^0}}$).

Following the pattern we established in Section 2, we now define $\mathsf{SZK_L}$ and $\mathsf{NISZK_L}$ in terms of their complete problems, rather than presenting the definitions in terms of interactive proofs:

▶ **Definition 26.** $\mathsf{SZK_L} = \{A : A \leq_{\mathrm{m}}^{\mathsf{proj}} \mathsf{SD_{NC^0}}\} = \{A : A \leq_{\mathrm{m}}^{\mathsf{L}} \mathsf{SD_{BP}}\}$
$\mathsf{NISZK_L} = \{A : A \leq_{\mathrm{m}}^{\mathsf{proj}} \mathsf{EA_{NC^0}}\} = \{A : A \leq_{\mathrm{m}}^{\mathsf{L}} \mathsf{EA_{BP}}\}$.

▶ **Theorem 27.** *The following are equivalent, for any decidable promise problem A:*

- $A \in \mathsf{NISZK_L}$
- $A \leq_{\mathrm{m}}^{\mathsf{RNC}^0} \widetilde{R}_K$
- $A \leq_{\mathrm{m}}^{\mathsf{BPNC}^0} \widetilde{R}_K$
- $A \leq_{\mathrm{m}}^{\mathsf{RAC}^0} \widetilde{R}_K$
- $A \leq_{\mathrm{m}}^{\mathsf{BPAC}^0} \widetilde{R}_K$
- $A \leq_{\mathrm{m}}^{\mathsf{RL}} \widetilde{R}_K$
- $A \leq_{\mathrm{m}}^{\mathsf{BPL}} \widetilde{R}_K$

Proof. The proof that $A \in \mathsf{NISZK}$ implies $A \leq_{\mathrm{m}}^{\mathsf{RNC}^0} \widetilde{R}_K$ proceeds as in the proof of Theorem 11, except that we appeal to [14, Corollary 43] (presenting a nonuniform $\leq_{\mathrm{m}}^{\mathsf{proj}}$ reduction from $\mathsf{EA_{NC^0}}$ to $\widetilde{R}_K$), instead of Corollary 18 in that paper. In more detail: as in the proof of Theorem 11, given x, the reduction constructs a sequence of independent copies of EA, but now each distribution is represented by an NC^0 circuit. The proof of Corollary 43 in [14] shows that these NC^0 circuits can be constructed via uniform *projections*, and thus each output bit is computed by a gadget that is connected to $O(1)$ random bits (i.e., the bits that are fed into the circuit computing the distribution), along with at most one bit from the input x (determining the circuitry internal to the gadget). The rest of the analysis is similar to that in the proof of Theorem 11.

If A is decidable and $A \leq_{\mathrm{m}}^{\mathsf{BPL}} \widetilde{R}_K$, then, as in the proof of Theorem 11, we build a device $C_x(r)$ that simulates the computation that produces queries to $\widetilde{R}_K$ on input x. However, now C_x is a branching program, and thus we replace queries to $\widetilde{R}_K$ by queries to $\mathsf{EA_{BP}}$. Again, the analysis is similar to that in the proof of Theorem 11. ◀

We end this section, with an analogous characterization of $\mathsf{SZK_L}$.

▶ **Definition 28.** *Let $A = (Y, N)$ and $B = (Y', N')$ be promise problems. We say $A \leq_{\mathrm{bf}}^{\mathsf{L}} B$ if there is a function f computable in logspace such that, for all x, $f(x)$ is of the form $(C, z_1, z_2, \ldots, z_k)$ where C is a Boolean formula with k input variables, and $(z_1, \ldots, z_k)$ is a list of queries, with the property that*

- *If $x \in Y$, then $C(\chi_B(z_1), \ldots, \chi_B(z_k)) = 1$.*
- *If $x \in N$, then $C(\chi_B(z_1), \ldots, \chi_B(z_k)) = 0$.*

Earlier work that studied $\leq_{\rm bf}^{\sf L}$ reducibility can be found in [21, 6].

We say $A \leq_{\rm bf}^{\sf BPL} B$ with threshold $\theta > \frac{1}{2}$ if there are functions f and g computable in **deterministic** *logspace, and a polynomial p, such that, for all x, $f(x)$ is a Boolean formula (with $k = |x|^{O(1)}$ variables), with the property that*

- *If $x \in Y$, then $C(\chi_{g,B}(x,1),\ldots,\chi_{g,B}(x,k)) = 1$,*
- *If $x \in N$, then $C(\chi_{g,B}(x,1),\ldots,\chi_{g,B}(x,k)) = 0$,*

where

- *$\chi_{g,B}(x,i) = 1$ if $\Pr_{r\in\{0,1\}^{p(|x|)}}[g(x,i,r) \in Y'] \geq \theta$*
- *$\chi_{g,B}(x,i) = 0$ if $\Pr_{r\in\{0,1\}^{p(|x|)}}[g(x,i,r) \in N'] \geq \theta$*
- *$\chi_{g,B}(x,i) = *$ otherwise.*

(Similarly, one can define AC^0 versions of $\leq_{\rm bf}^{\sf L}$, although, since an AC^0 circuit cannot evaluate a Boolean formula, we do not pursue that direction here.)

▶ **Theorem 29.** *The following are equivalent, for any decidable promise problem A:*

- *$A \in \mathsf{SZK_L}$.*
- *$A \leq_{\rm bf}^{\sf L} \mathsf{EA}_{\mathsf{NC}^0}$.*
- *$A \leq_{\rm bf}^{\sf BPL} \widetilde{R}_K$ with threshold $1 - \frac{1}{n^{\omega(1)}}$.*

Proof. The first two items are equivalent, because (a) $\mathsf{SZK_L}$ is closed under $\leq_{\rm bf}^{\sf L}$ reducibility [15], and (b) the argument in [27], showing that SZK $\leq_{\rm bf}^{\sf L}$-reduces to NISZK carries over directly to $\mathsf{SZK_L}$ and $\mathsf{NISZK_L}$.

Since $\mathsf{EA}_{\mathsf{NC}^0}$ is complete for $\mathsf{NISZK_L}$, Theorem 27 implies that every $A \in \mathsf{NISZK_L}$ is $\leq_{\rm bf}^{\sf BPL}$-reducible to $\widetilde{R}_K$. The argument that every decidable A that $\leq_{\rm bf}^{\sf BPL}$-reduces to $\widetilde{R}_K$ lies in $\mathsf{SZK_L}$ is similar to the argument in Theorem 23. ◀

7 Discussion

There are not many examples of natural computational problems that are known or conjectured to lie outside of P, such that the class of problems reducible to them via $\leq_{\rm m}^{\sf P}$ and $\leq_{\rm m}^{\sf L}$ (or $\leq_{\rm m}^{\mathsf{AC}^0}$) reductions differ (or are conjectured to differ). Is it the case that the problems reducible to $\widetilde{R}_K$ via $\leq_{\rm m}^{\sf RP}$ and $\leq_{\rm m}^{\sf RL}$ (or $\leq_{\rm m}^{\mathsf{RAC}^0}$) reductions differ? Or should this be taken as evidence that NISZK and $\mathsf{NISZK_L}$ coincide?

Similarly, there are not many examples of natural computational problems such that the classes of problems reducible to them via $\leq_{\rm tt}^{\sf P}$ and $\leq_{\rm bf}^{\sf P}$ reductions differ (or are conjectured to differ). For example, these reducibilities coincide for SAT [22]. Is it the case that $\leq_{\rm bf}^{\sf BPP}$ and $\leq_{\rm circ}^{\sf BPP}$ reducibilities differ for $\widetilde{R}_K$? Or should this be taken as evidence that SZK is closed under $\leq_{\rm tt}^{\sf P}$ reducibility?

Perhaps our new characterizations of statistical zero knowledge classes will be useful in answering these questions.

It is known that every promise problem in $\mathsf{NISZK_L}$ reduces to $\widetilde{R}_K$ via *nonuniform projections* [14, 4]. The following quote from [4] is worth paraphrasing here:

> …no complexity class larger than $\mathsf{NISZK_L}$ is known to be (non-uniformly) $\leq_{\rm m}^{\mathsf{AC}^0}$ reducible to the Kolmogorov-random strings [14]. It seems unlikely that this is optimal.

The discussion in [4] was referring to reductions to an oracle for the *exact* Kolmogorov-complexity function. Our results show that, for reductions to an *approximation* to the Kolmogorov-complexity function, $\mathsf{NISZK_L}$ *is* essentially "optimal".

Finally, let us observe that our new characterizations of $\mathsf{NISZK_L}$ may open new avenues of attack on questions such as whether $\mathsf{NP} = \mathsf{NL}$. MKTP, the problem of computing KT complexity, lies in NP and is hard for $\mathsf{co\text{-}NISZK_L}$ under nonuniform projections [14]. If $\mathsf{MKTP} \in \mathsf{NISZK_L}$, then there must be a nonuniform projection f that takes strings of low KT-complexity (and hence low K-complexity) to strings of high K complexity, and simultaneously maps strings of high KT complexity to strings of low K-complexity. It is plausible that one could show unconditionally that no such projection can exist. Among other things, this would show that $\mathsf{NP} \neq \mathsf{DET}$ (where DET is the complexity class, containing NL, of problems that reduce to the determinant) since $\mathsf{DET} \subseteq \mathsf{NISZK_L}$ [14]. In this vein, let us also remark that Kolmogorov complexity has already proved useful in developing nonrelativizing proof techniques [31], and also that the machinery of perfect randomized encodings (which were developed in [17] and which are essential to the results of [14]) also does not seem to relativize in any obvious way.

Acknowledgments

EA and HT were supported in part by NSF Grants CCF-1909216 and CCF-1909683. SH was supported in part by JST, PRESTO Grant Number JPMJPR2024, Japan. We thank Sam Buss, Johannes Köbler, and Uwe Schöning for discussions concerning Boolean formula reducibility.

References

1 Leonard M. Adleman and Kenneth L. Manders. Reducibility, randomness, and intractability (abstract). In *Proceedings of the 9th Annual ACM Symposium on Theory of Computing (STOC)*, pages 151–163. ACM, 1977. doi:10.1145/800105.803405.

2 Eric Allender. Curiouser and curiouser: The link between incompressibility and complexity. In *Proc. Computability in Europe (CiE)*, volume 7318 of *Lecture Notes in Computer Science*, pages 11–16. Springer, 2012. doi:10.1007/978-3-642-30870-3_2.

3 Eric Allender. The complexity of complexity. In *Computability and Complexity: Essays Dedicated to Rodney G. Downey on the Occasion of his 60th Birthday*, volume 10010 of *Lecture Notes in Computer Science*, pages 79–94. Springer, 2017. doi:10.1007/978-3-319-50062-1_6.

4 Eric Allender. Vaughan Jones, Kolmogorov complexity, and the new complexity landscape around circuit minimization. *New Zealand journal of mathematics*, 52, 2021. doi:10.53733/148.

5 Eric Allender, José L. Balcázar, and Neil Immerman. A first-order isomorphism theorem. *SIAM J. Comput.*, 26(2):557–567, 1997. doi:10.1137/S0097539794270236.

6 Eric Allender, David A. Mix Barrington, Tanmoy Chakraborty, Samir Datta, and Sambuddha Roy. Planar and grid graph reachability problems. *Theory of Computing Systems*, 45(4):675–723, 2009. doi:10.1007/s00224-009-9172-z.

7 Eric Allender, Harry Buhrman, Luke Friedman, and Bruno Loff. Reductions to the set of random strings: The resource-bounded case. *Logical Methods in Computer Science*, 10(3), 2014. doi:10.2168/LMCS-10(3:5)2014.

8 Eric Allender, Harry Buhrman, and Michal Koucký. What can be efficiently reduced to the Kolmogorov-random strings? *Annals of Pure and Applied Logic*, 138:2–19, 2006.

9 Eric Allender, Harry Buhrman, Michal Kouckỳ, Dieter Van Melkebeek, and Detlef Ronneburger. Power from random strings. *SIAM Journal on Computing*, 35(6):1467–1493, 2006. doi:10.1007/978-3-662-03927-4.

10 Eric Allender, Mahdi Cheraghchi, Dimitrios Myrisiotis, Harsha Tirumala, and Ilya Volkovich. One-way functions and a conditional variant of MKTP. In *41st IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS)*, volume 213 of *LIPIcs*, pages 7:1–7:19. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021. doi:10.4230/LIPIcs.FSTTCS.2021.7.

11 Eric Allender and Bireswar Das. Zero knowledge and circuit minimization. *Information and Computation*, 256:2–8, 2017. Special issue for MFCS '14. doi:10.1016/j.ic.2017.04.004.

12 Eric Allender, George Davie, Luke Friedman, Samuel B. Hopkins, and Iddo Tzameret. Kolmogorov complexity, circuits, and the strength of formal theories of arithmetic. *Chicago Journal of Theoretical Computer Science*, 2013(5), April 2013. doi:10.4086/cjtcs.2013.005.

13 Eric Allender, Luke Friedman, and William Gasarch. Limits on the computational power of random strings. *Information and Computation*, 222:80–92, 2013. ICALP 2011 Special Issue. doi:10.1016/j.ic.2011.09.008.

14 Eric Allender, John Gouwar, Shuichi Hirahara, and Caleb Robelle. Cryptographic hardness under projections for time-bounded Kolmogorov complexity. In *32nd International Symposium on Algorithms and Computation (ISAAC)*, volume 212 of *LIPIcs*, pages 54:1–54:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021. doi:10.4230/LIPIcs.ISAAC.2021.54.

15 Eric Allender, Jacob Gray, Saachi Mutreja, Harsha Tirumala, and Pengxiang Wang. Robustness for space-bounded statistical zero knowledge. Technical report, 2023. In preparation.

16 Eric Allender, Joshua A Grochow, Dieter Van Melkebeek, Cristopher Moore, and Andrew Morgan. Minimum circuit size, graph isomorphism, and related problems. *SIAM Journal on Computing*, 47(4):1339–1372, 2018. doi:10.1137/17M1157970.

17 Benny Applebaum, Yuval Ishai, and Eyal Kushilevitz. Cryptography in NC^0. *SIAM Journal on Computing*, 36(4):845–888, 2006. doi:10.1137/S0097539705446950.

18 David A. Mix Barrington, Neil Immerman, and Howard Straubing. On uniformity within NC^1. *Journal of Computer and System Sciences*, 41(3):274–306, 1990. doi:10.1016/0022-0000(90)90022-D.

19 Andrej Bogdanov and Luca Trevisan. On worst-case to average-case reductions for NP problems. *SIAM J. Comput.*, 36(4):1119–1159, 2006. doi:10.1137/S0097539705446974.

20 Harry Buhrman, Lance Fortnow, Michal Koucký, and Bruno Loff. Derandomizing from random strings. In *25th IEEE Conference on Computational Complexity (CCC)*, pages 58–63. IEEE, 2010. doi:10.1109/CCC.2010.15.

21 Harry Buhrman, Edith Spaan, and Leen Torenvliet. The relative power of logspace and polynomial time reductions. *Computational Complexity*, 3:231–244, 1993. doi:10.1007/BF01271369.

22 Samuel R. Buss and Louise Hay. On truth-table reducibility to SAT. *Information and Computation*, 91(1):86–102, 1991. doi:10.1016/0890-5401(91)90075-D.

23 Mingzhong Cai, Rodney Downey, Rachel Epstein, Steffen Lempp, and Joseph Miller. Random strings and tt-degrees of Turing complete c.e. sets. *Logical Methods in Computer Science*, 10(3):1–24, 2014. doi:10.2168/LMCS-10(3:15)2014.

24 Richard Chang, Jim Kadin, and Pankaj Rohatgi. On unique satisfiability and the threshold behavior of randomized reductions. *Journal of Computer and System Sciences*, 50(3):359–373, 1995. doi:10.1006/jcss.1995.1028.

25 R. Downey and D. Hirschfeldt. *Algorithmic Randomness and Complexity*. Springer, 2010.

26 Zeev Dvir, Dan Gutfreund, Guy N Rothblum, and Salil P Vadhan. On approximating the entropy of polynomial mappings. In *Second Symposium on Innovations in Computer Science*, 2011.

27 Oded Goldreich, Amit Sahai, and Salil Vadhan. Can statistical zero knowledge be made non-interactive? or On the relationship of SZK and NISZK. In *Annual International Cryptology Conference*, pages 467–484. Springer, 1999. doi:10.1007/3-540-48405-1_30.

28 Joachim Grollmann and Alan L. Selman. Complexity measures for public-key cryptosystems. *SIAM J. Comput.*, 17(2):309–335, 1988. doi:10.1137/0217018.

29 Shuichi Hirahara. Unexpected hardness results for Kolmogorov complexity under uniform reductions. In *Proccedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 1038–1051. ACM, 2020. doi:10.1145/3357713.3384251.

30 Shuichi Hirahara. Unexpected power of random strings. In *11th Innovations in Theoretical Computer Science Conference, ITCS*, volume 151 of *LIPIcs*, pages 41:1–41:13. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2020. doi:10.4230/LIPIcs.ITCS.2020.41.

31 Shuichi Hirahara. NP-hardness of learning programs and partial MCSP. In *63rd IEEE Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2022. To appear.

32 Shuichi Hirahara and Akitoshi Kawamura. On characterizations of randomized computation using plain Kolmogorov complexity. *Computability*, 7(1):45–56, 2018. doi:10.3233/COM-170075.

33 Shuichi Hirahara and Osamu Watanabe. On nonadaptive reductions to the set of random strings and its dense subsets. In Ding-Zhu Du and Jie Wang, editors, *Complexity and Approximation - In Memory of Ker-I Ko*, volume 12000 of *Lecture Notes in Computer Science*, pages 67–79. Springer, 2020. doi:10.1007/978-3-030-41672-0_6.

34 Rahul Ilango. Approaching MCSP from above and below: Hardness for a conditional variant and $AC^0[p]$. In *11th Innovations in Theoretical Computer Science Conference (ITCS)*, volume 151 of *LIPIcs*, pages 34:1–34:26. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020. doi:10.4230/LIPIcs.ITCS.2020.34.

35 Rahul Ilango. Constant depth formula and partial function versions of MCSP are hard. In *61st IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pages 424–433. IEEE, 2020. doi:10.1109/FOCS46700.2020.00047.

36 Rahul Ilango, Bruno Loff, and Igor Carboni Oliveira. NP-hardness of circuit minimization for multi-output functions. In *35th Computational Complexity Conference (CCC)*, volume 169 of *LIPIcs*, pages 22:1–22:36. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020. doi:10.4230/LIPIcs.CCC.2020.22.

37 Rahul Ilango, Hanlin Ren, and Rahul Santhanam. Robustness of average-case meta-complexity via pseudorandomness. In *54th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 1575–1583. ACM, 2022. doi:10.1145/3519935.3520051.

38 Neil Immerman. *Descriptive complexity*. Graduate texts in computer science. Springer, 1999. doi:10.1007/978-1-4612-0539-5.

39 Johannes Köbler, Uwe Schöning, and Klaus W. Wagner. The difference and truth-table hierarchies for NP. *RAIRO Theor. Informatics Appl.*, 21(4):419–435, 1987. doi:10.1051/ita/1987210404191.

40 Richard E. Ladner, Nancy A. Lynch, and Alan L. Selman. A comparison of polynomial time reducibilities. *Theoretical Computer Science*, 1(2):103–123, 1975. doi:10.1016/0304-3975(75)90016-X.

41 Ming Li and Paul M. B. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications, 4th Edition*. Texts in Computer Science. Springer, 2019. doi:10.1007/978-3-030-11298-1.

42 Yanyi Liu and Rafael Pass. On one-way functions from NP-complete problems. In *37th Computational Complexity Conference (CCC)*, volume 234 of *LIPIcs*, pages 36:1–36:24. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022. doi:10.4230/LIPIcs.CCC.2022.36.

43 Kenneth W. Regan. A uniform reduction theorem - extending a result of J. Grollmann and A. Selman. In *Proc. International Conference on Automata, Languages, and Programming (ICALP)*, volume 226 of *Lecture Notes in Computer Science*, pages 324–333. Springer, 1986. doi:10.1007/3-540-16761-7_82.

44 Hanlin Ren and Rahul Santhanam. Hardness of KT characterizes parallel cryptography. In *36th Computational Complexity Conference (CCC)*, volume 200 of *LIPIcs*, pages 35:1–35:58. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021. doi:10.4230/LIPIcs.CCC.2021.35.

45 Amit Sahai and Salil P. Vadhan. A complete problem for statistical zero knowledge. *J. ACM*, 50(2):196–249, 2003. doi:10.1145/636865.636868.

46 Michael Saks and Rahul Santhanam. On randomized reductions to the random strings. In *37th Computational Complexity Conference (CCC)*, volume 234 of *LIPIcs*, pages 29:1–29:30. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022. doi:10.4230/LIPIcs.CCC.2022.29.

47 Salil Vadhan. *A Study of Statistical Zero-Knowledge Proofs*. Springer, 2023. To appear.

48 Leslie G. Valiant and Vijay V. Vazirani. NP is as easy as detecting unique solutions. *Theoretical Computer Science*, 47(3):85–93, 1986. doi:10.1016/0304-3975(86)90135-0.

49 Heribert Vollmer. *Introduction to circuit complexity: a uniform approach*. Springer Science & Business Media, 1999. doi:10.1007/978-3-662-03927-4.

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il