Electronic Colloquium on Computational Complexity, Report No. 155 (2019)

Pseudorandomness and the Minimum Circuit Size Problem

Rahul Santhanam*
University of Oxford

November 5, 2019

Abstract

We explore the possibility of basing one-way functions on the average-case hardness of the fundamental Minimum Circuit Size Problem ($\mathsf{MCSP}[s]$), which asks whether a Boolean function on n bits specified by its truth table has circuits of size $s(n)$.

1. (Pseudorandomness from Zero-Error Average-Case Hardness) We show that for a given size function s, the following are equivalent: Pseudorandom distributions supported on strings describable by $s(O(n))$-size circuits exist; Hitting sets supported on strings describable by $s(O(n))$-size circuits exist; $\mathsf{MCSP}[s(O(n))]$ is zero-error average-case hard. Using similar techniques, we show that Feige's hypothesis for random k-CNFs implies that there is a pseudorandom distribution (with constant error) supported entirely on satisfiable formulas. Underlying our results is a general notion of *semantic sampling*, which might be of independent interest.
2. (A New Conjecture) In analogy to a known universal construction of succinct hitting sets against arbitrary polynomial-size adversaries, we propose the Universality Conjecture: there is a universal construction of succinct pseudorandom distributions against arbitrary polynomial-size adversaries. We show that under the Universality Conjecture, the following are equivalent: One-way functions exist; Natural proofs useful against sub-exponential size circuits do not exist; Learning polynomial-size circuits with membership queries over the uniform distribution is hard; $\mathsf{MCSP}[2^{\epsilon n}]$ is zero-error hard on average for some $\epsilon > 0$; Cryptographic succinct hitting set generators exist.
3. (Non-Black-Box Results) We show that for weak circuit classes $\mathfrak{C}$ against which there are natural proofs [RR97], pseudorandom functions secure against poly-size circuits in $\mathfrak{C}$ imply superpolynomial lower bounds in P against poly-size circuits in $\mathfrak{C}$. We also show that for a certain natural variant of MCSP, there is a polynomial-time reduction from approximating the problem well in the worst case to solving it on average. These results are shown using non-black-box techniques, and in the first case we show that there is no black-box proof of the result under standard crypto assumptions.

*E-mail: rahul.santhanam@cs.ox.ac.uk

ISSN 1433-8092

1 Introduction

We investigate the relationship between the complexity of the Minimum Circuit Size Problem (MCSP) [All17] and the existence of various kinds of pseudorandom objects, such as hitting sets and pseudorandom sets, succinctly describable or not, in the cryptographic regime. There are two broad regimes of pseudorandom constructions: the complexity-theoretic regime, where the generator has more resources than the adversary, and the cryptographic regime, where the adversary can have more resources than the generator. In the complexity-theoretic regime, there is a beautiful theory [NW94, IW97] giving equivalences between the existence of hard problems in E (linear exponential time), and explicit constructions of hitting set generators and pseudorandom generators.

In the cryptographic regime, the celebrated result of [HILL99] gives an equivalence between pseudorandom generators and one-way functions, which are in many ways the fundamental cryptographic primitive. However, as we point out, there are several gaps in our understanding of the relationships between hardness and pseudorandomness in the cryptographic regime. To some extent, these gaps reflect an imperfect knowledge of average-case hardness for NP in general. In this paper, we make the argument that the average-case complexity of MCSP in particular is deeply relevant to pseudorandomness in the cryptographic regime.

We first discuss the questions that motivate us.

1.1 Our Questions

1.1.1 One-way Functions

From a cryptographer's point of view, one-way functions are an extremely robust and useful primitive, forming the basis for a range of important crypo constructions [KL14]. However, one-way functions do not fit very neatly into the complexity landscape. The hardness assumptions required to do cryptography seem stronger than the assumptions traditionally studied in complexity theory, and bridging this gap is an important open problem.

Question 1 Is there a natural well-studied problem in NP whose average-case hardness with respect to some natural distribution is equivalent to the existence of one-way functions?

An attempt to base one-way functions on a standard complexity assumption was made by Ostrovsky and Wigderson [OW93], who showed that a weak variant of one-way functions called auxiliary-input one-way functions follow from the assumption that ZK (computational zero-knowledge) is not contained in BPP. It is unknown whether the existence of one-way functions is equivalent to the existence of auxiliary-input one-way functions. An equivalence result would imply that one-way functions could be based on worst-case hardness of ZK, which "would have a major impact on cryptography" [ABX08].

Question 2 Are one-way functions equivalent to auxiliary-input one-way functions?

One-way functions are also relevant to proof complexity. The main result of [RR97] is that if one-way functions exist, "natural proofs" of lower bounds against superpolynomial-size circuits do not. Here a "natural proof" is a property of Boolean functions that is satisfied with significant probability by a random Boolean function, and is moreover easy to check. Most standard algebraic

and combinatorial lower bound techniques yield properties of this form.

Thus one-way functions rule out natural proofs of lower bounds, but could it be the case that neither one-way functions exist nor natural proofs? Both objects are useful, the first for cryptographic applications and the second for proof theory, so it would be nice to have at least one of them available, even if it is impossible to have both.

Question 3 Can we base one-way functions on the non-existence of natural proofs?

1.1.2 Pseudorandomness

In many contexts where pseudorandomness is relevant, there is a distinction between a one-sided "hitting" notion where the goal is to find an explicit set of points that hits every "easy" dense set and a two-sided "pseudorandom" notion, where the question is to find an explicit distribution on points that approximates every easy set. The support of a pseudorandom distribution is a hitting set, but it is not clear in general how to get pseudorandom distributions from hitting sets. Sometimes the two notions coincide, and this makes for a cleaner theory. This is the case for example with the complexity-theoretic regime, where the results of [NW94, IW97] imply that hitting set generators (HSGs) and pseudorandom generators (PRGs) are equivalent.

How about the cryptographic regime? What is the power of HSGs in this setting? Surprisingly, this question doesn't seem to have received much attention, though it seems a very natural one.

Part of the reason might be that, in contrast to PRGs, HSGs are not obviously robust. For instance, while we can increase the "stretch" of a PRG simply by iterating it, this does not seem to work with HSGs.

Question 4 Can HSGs be stretched in the cryptographic regime? For example, does a HSG with seed length $N^{0.5}$ (where N is the output length) imply one with seed length $N^{0.25}$?

Question 4 might seem purely curiosity-driven, but in fact it has relevance to proof complexity and cryptography against non-deterministic adversaries, as suggested by Rudich [Rud97]. The notion of cryptographic pseudorandomness does not make sense against non-deterministic adversaries, as a non-deterministic adversary could break the PRG by simply guessing the seed and checking that it maps to his/her input. However, a crucial observation is that the notion of HSG still does make sense against non-deterministic adversaries. Rudich defined the notion of a 'demi-bit' to capture hitting sets in this setting, and asked questions about whether they can be stretched. Indeed, Question 4 is a version of his question for deterministic adversaries.

Cryptographic PRGs are equivalent to the existence of one-way functions, as shown by [HILL99]. Is there a comparable connection for HSGs?

Question 5 In the cryptographic regime, is the existence of HSGs equivalent to some notion of average-case hardness for a decision or search problem?

Finally, an ideal situation in terms of the cleanness of the resulting theory would be if HSGs are simply equivalent to PRGs.

Question 6 Are HSGs equivalent to PRGs in the cryptographic regime?

1.1.3 The Minimum Circuit Size Problem and Learning

The Minimum Circuit Size Problem (MCSP), which asks if a given string is the truth table of a function with small circuits, is both fundamental and elusive. It asks a very natural question about the complexity of a string, and the naturalness and importance of the problem have been clear since work by Soviet researchers in the 50s and 60s [Tra84]. However, the problem has also been elusive in terms of classifying its complexity. It is said that Levin delayed publication of his seminal paper on NP-completness because he was hoping to show MCSP is NP-complete [AKRR11]. Nearly 50 years later, we still lack even a clear belief about whether MCSP is NP-complete or not.

Natural proofs are essentially zero-error algorithms for MCSP on average over the uniform distribution [RR97, HS17], so it is no surprise that the "natural proofs" paper revived work on MCSP and its place in the complexity of landscape. There has been a long line of recent works on the problem [KC00, AHM+08, AD14, AGM15, AHK15, MW15, Wil16, AH17, HW16, HP15, HS17, OS17]. In fact, the work by Williams [Wil16] implicitly studies connections between MCSP and derandomization, but he is interested in connections to complexity-theoretic derandomization, while we are interested in the cryptographic regime, which raises some very different issues.

Despite recent work, major questions still remain about how the problem relates to other problems. Indeed, a number of the cited works deal with the difficulties of constructing reductions to MCSP.

Question 7 Is MCSP or one of its variants equivalent in complexity to some other complexity class or notion of independent interest?

MCSP comes supplied with a parameter - the size s for which we wish to know whether the input truth table has circuits of that size. Often complexity results about MCSP are fairly robust to this parameter, but there is no formal justification known for this. It was suggested in [HS17] that the problem might be easier to show robust to its parameter in terms of average-case complexity.

Question 8 Does the average-case easiness of MCSP with parameter $2^{n/4}$ imply the average-case easiness of MCSP with parameter $2^{n/2}$?

The MCSP problem is closely connected to Valiant's PAC learning model [Val84]. Learning algorithms can be thought of as methods to solve the *search* version of MCSP - they are given access to the truth table of a Boolean function with small circuits, and need to efficiently find a small circuit that approximates the truth table well. Valiant observed in his original paper that polynomial-size circuits are not learnable in his model under cryptographic assumptions, and it was observed in [PW90] that non-learnability follows from the assumption that one-way functions exist. However the precise complexity of learning in various models is still not known despite more than three decades of work [Val84, KV89, BFKL93, ABX08, DLS14, Vad17]. In particular, one can ask if there is a converse to the hardness result of [Val84, PW90] for some natural learning model.

Question 9 For some natural learning model, does non-learnability of polynomial-size circuits imply the existence of one-way functions?

MCSP has stubbornly resisted attempts to show that it is NP-complete. It is natural to wonder if it has structural features that distinguish it from other NP-complete problems, such as for

example random self-reducibility or a worst-case to average-case reduction. This might give some evidence that MCSP is not NP-complete - it is known under standard complexity assumptions that NP-complete problems do not have black-box worst-case to average-case reductions [FF93, BT06].

Question 10 Is there a worst-case to average-case reduction for MCSP? Or to ask a more relaxed question, does average-case easiness of MCSP imply non-trivial approximations for the problem?

1.2 Results

1.2.1 Pseudorandomness from Zero-Error Average-Case Hardness

We begin by showing connections between zero-error average-case hardness of MCSP, and the existence of *succinct* hitting sets and pseudorandom distributions. As mentioned before, we are inspired by the connections between these notions in the complexity-theoretic setting [NW94, IW97], and are looking for analogous results in the cryptographic setting.

Let us first explain what we mean by zero-error average-case hardness of MCSP. Given any size function $s : \mathbb{N} \to \mathbb{N}$, $\mathsf{MCSP}[s]$ has a very natural distribution on inputs associated with it, namely the uniform distribution. $\mathsf{MCSP}[s]$ is heavily biased over the uniform distribution since the overwhelming majority of truth tables correspond to hard Boolean functions. So it does not make sense to study bounded-error notions of average-case hardness over the uniform distribution - the trivial algorithm that always says 'no' will do very well. Instead, we consider zero-error algorithms, i.e., algorithms that always output the correct answer or '?', and output the correct answer with noticeable probability over the uniform distribution.

It turns out that this notion of average-case hardness is fairly robust. It was shown in [HS17] that average-case hardness is essentially equivalent to the non-existence of Razborov-Rudich natural proofs. Thus the main result of Razborov and Rudich [RR97] showing that natural proofs don't exist if one-way functions exist can also be interpreted as showing zero-error average-case hardness of $\mathsf{MCSP}[s]$ for any $s = 2^{\Omega(n)}$ under the assumption that one-way functions exist. A major motivation for this paper is the question of whether the *converse* holds, i.e., whether one-way functions can be based on zero-error average-case hardness of $\mathsf{MCSP}[s]$ for reasonable size functions s. Indeed, by the connection we just mentioned with natural proofs, this is equivalent to Question 3.

It is known in the cryptographic setting that the existence of pseudorandom functions of circuit complexity $2^{\epsilon n}$ for any $\epsilon > 0$ is equivalent to the existence of one-way functions [HILL99, GGM86]. In order to make progress toward basing one-way functions on zero-error average-case hardness of MCSP, we first show how to derive weaker versions of pseudorandom functions. Succinct hitting sets and succinct pseudorandom distributions are examples of such weaker objects, as we explain below.

A succinct set (resp. distribution) is a set of (resp. distribution over) strings, each of which, when interpreted as the truth table of a Boolean function, has circuits that are not too large. Succinct hitting sets are simply succinct sets that are hitting sets against arbitrary poly-size adversaries. A succinct pseudorandom distribution is a succinct distribution that is pseudorandom against arbitrary poly-size adversaries.

To compare these notions with pseudorandom functions, we note that pseudorandom functions are essentially equivalent to succinct PRGs, i.e., succinct pseudorandom distributions that are efficiently samplable. Similarly, a succinct HSG is an efficiently computable function whose range is a collection of succinct hitting sets.

It is fairly straightforward to show that zero-error average-case hardness of $\mathsf{MCSP}[\tilde{O}(s)]$, $\tilde{O}(s(n))$-succinct hitting sets and $\tilde{O}(s(n))$-succinct HSGs are all equivalent. We give this argument in Subsection 3.1 of Section 3. The equivalence between the second and third notions follows from the existence of *universal succinct HSGs*, i.e., a fixed polynomial-time construction that is guaranteed to be a succinct HSG if succinct hitting sets exist. This is essentially folklore - the idea is to use the mapping from circuits to the truth tables of functions they compute. Indeed, this has been a very fruitful technique in complexity theory - the "easy witness" method of [Kab00, IKW01].

What seems less straightforward is to get a connection between succinct hitting sets and succinct pseudorandom sets. This is what we manage to show.

Theorem 1. *(Informal Statement) The following are equivalent:*

1. *For all $\epsilon > 0$, there are succinct hitting sets supported on truth tables of circuit complexity $2^{\epsilon n}$.*
2. *For all $\epsilon > 0$, MCSP with parameter $2^{\epsilon n}$ is zero-error hard on average.*
3. *For all $\epsilon > 0$, there are succinct pseudorandom distributions supported on truth tables of circuit complexity $2^{\epsilon n}$.*

It is perhaps surprising that based on a *zero-error* average-case notion of hardness for MCSP, we are able to get pseudorandom distributions that are indistinguishable from random with respect to *bounded two-sided error*.

The proof of Theorem 1 proceeds via a certain Sampler-Distinguisher game we define here, inspired by the PRF-Distinguisher game in [OS17]. The analysis of the game uses the approximate Min-Max theorem of [Alt94, LY94] and is fairly general. We note that the approximate Min-Max theorem has been used in many different contexts in complexity theory and pseudorandomness (see [VZ13]), but as far as we are aware, our use of it here to derive pseudorandomness from zero-error average-case hardness is novel.

Indeed our techniques can be used to establish an analogous result for Satisfiability based on Feige's hypothesis [Fei02] that random k-CNFs of linear density are hard to refute. Feige's hypothesis is essentially a hypothesis about zero-error average-case hardness of k-SAT under a natural distribution on k-CNFs, just as the non-existence of natural proofs is a hypothesis about zero-error average-case hardness of MCSP. One way of stating the hypothesis is that errorless polynomial-time algorithms, which always output the correct answer or '?' and output the correct answer with high probability over randomly chosen k-CNFs of linear density, do not exist.

In this case again, we get a consequence for pseudorandom distributions. Just as the pseudorandom distributions obtained in Theorem 1 are succinct, i.e., supported on YES instances of $\mathsf{MCSP}[s]$, here the pseudorandom distributions obtained are supported on satisfiable instances.

Theorem 2. *(Informal Statement) For any fixed integer k, if random k-CNFs with any linear number of clauses are hard to refute for non-uniform poly-time algorithms, then for each $\epsilon > 0$ there is a pseudorandom distribution over satisfiable formulas with error ϵ.*

We abstract out the main idea behind the proofs of Theorems 1 and 2 to show a more general connection between zero-error average-case hardness for a language $Q \subseteq \{0,1\}^*$ and the existence of pseudorandom distributions supported on a language $Q' \subseteq \{0,1\}^*$. We show that such a connection exists whenever there are samplers satisfying a certain *semantic* condition. Recall that a sampler

f from n bits to m bits with seed length ℓ is a polynomial-time computable function such that for any bounded function g on m bits, for *most* inputs x of length n, the expectation of $g(f(x, U_\ell))$ is close to the expectation of $g(U_m)$. Samplers are closely related to randomness extractors [Vad12]. We consider samplers with an additional condition: whenever $x \in Q$, for every y of length ℓ, $f(x, y) \in Q'$. We call such samplers (Q, Q')-semantic samplers, and show that the existence of semantic samplers with good parameters implies a strong connection between zero-error average-case hardness and pseudorandomness.

We note that this connection between samplers and pseudorandomness is *different* to the well-known connection of Trevisan [Tre01] between hardness-based pseudorandom generators and extractors. Indeed, Trevisan's connection does not involve any semantic property of the extractor.

1.2.2 A New Conjecture and Its Consequences

Theorem 1 partly addresses Questions 5 and 6 in Subsection 1.1, but it is not clear what it says about the others. Motivated by a belief that the questions in Subsection 1.1 are all connected and part of a unified picture, we propose a natural conjecture about universal succinct PRGs that results in such a picture.

A universal succinct PRG is a fixed polynomial-time computable function f such that if succinct pseudorandom distributions exist, then f induces such distributions on uniformly chosen seeds.

Universality Conjecture (Informal Statement) There exist universal succinct PRGs.

We briefly discuss the context for the Conjecture. Universality is a phenomenon that is widely observed in complexity theory and the foundations of cryptography. For example, there are universal one-way functions [Lev84], and via the equivalence between one-way functions and PRGs [HILL99], there are universal PRGs in the cryptographic setting. In the complexity-theoretic setting, universal HSGs and PRGs follow from the equivalence of HSGs, PRGs and circuit lower bounds for $\mathsf{E} \stackrel{\text{def}}{=} \mathsf{DTIME}(2^{O(n)})$, together with the existence of complete problems for $\mathbb{E}$. Given the variety of such universal examples, and the variety of ways for establishing that they exist, it is perhaps not unreasonable to expect them in the context of succinct PRGs.

An even closer analogy is to the case of succinct HSGs, which are the one-sided error version of succinct PRGs. As described in the previous subsection, universal succinct HSGs can be shown to exist by interpreting circuits succinctly describing hitting sets as seeds to a hitting set generator. This is a folklore argument that we formalize in Section 3 as Proposition 2. The Universality Conjecture is simply the analogue of Proposition 2 for pseudorandom distributions.

Next we turn to the consequences of our Conjecture for the questions raised in the introduction. We make a few clarifications about notation in the informal statement of our main result below. N always refer to the output size of some generator and is a power of 2, and $n \stackrel{\text{def}}{=} \log(N)$. By default, we take the succinctness parameter (i.e., the size of circuits representing each element of the range) to be the same as the seed length. To clarify any further notational issues, please refer to Section 2.

Theorem 3. *(Informal Statement) If the Univerality Conjecture is true, then the following hold:*

1. *One-way functions exist iff there is an $\epsilon < 1$ such that MCSP with parameter $2^{\epsilon n}$ is zero-error hard on average.*

2. *One-way functions exist iff auxiliary one-way functions exist.*

3. *One-way functions exist iff natural proofs against* $\mathsf{SIZE}(\text{poly})$ *do not exist.*

4. *For any* $0 < \epsilon < \delta < 1$*, a succinct HSG with seed length* N^δ *implies a succinct HSG with seed length* N^ϵ*.*

5. *For any* $\epsilon < 1$*, a succinct HSG with seed length* N^ϵ *exists iff* MCSP *with parameter* $2^{\epsilon n}$ *is zero-error hard on average.*

6. *For any* $\epsilon < 1$*, a succinct HSG with seed length* N^ϵ *exists iff a PRG with seed length* N^ϵ *exists.*

7. *For any* $0 < \epsilon < \delta$*,* MCSP *with parameter* $2^{\epsilon n}$ *is zero-error hard on average iff* MCSP *with parameter* $2^{\delta n}$ *is zero-error hard on average.*

8. *One-way functions exist iff polynomial-size circuits cannot be learned with membership queries under the uniform distribution in polynomial time.*

Items 1 to 3 of Theorem 3 answer Questions 1 to 3. Items 4 to 6 answer Questions 4 to 6, but for *succinct* HSGs rather than HSGs. Question 7 is also answered by Item 1. Questions 8 and 9 are answered by Items 7 and 8.

Of the items in Theorem 3, it is perhaps Item 1 which is of most interest. Most candidate one-way functions are based on problems in $\mathsf{NP} \cap \mathsf{coNP}$. MCSP, however, is believed by many to be NP-complete, and therefore unlikely to be in coNP. Indeed, a conjecture of Rudich [Rud97] even asserts that MCSP is hard on average for coNP.

Item 1 also has implications for Impagliazzo's "Five Worlds" [Imp95]. In particular, if the Conjecture were true and we could base one-way functions on average-case hardness of MCSP, we would get evidence against the existence of Pessiland: a world where there are problems that are hard on average, but one-way functions do not exist.

We briefly explain how the Conjecture helps to show these results. For every item in Theorem 3, one of the two directions of the equivalence was known before, and it is the Conjecture, together with Theorem 1, that enables us to show the reverse direction. The crucial aspect of the Conjecture is that it allows us to derive succinct PRGs from succinct pseudorandom distributions. Once we have a PRG, we are able to stretch the PRG using standard techniques, and this enables us to close the chain of implications between average-case hardness of MCSP, succinct HSGs and succinct PRGs.

We discuss some reasons why it might be useful to consider the Universality Conjecture.

First note that common beliefs in the crypto and complexity communities support the Conjecture. Indeed, most complexity theorists and cryptographers believe that one-way functions exist. If one-way functions exist, by [HILL99] and [GGM86], pseudorandom function generators exist, and any pseudorandom function generator is trivially a universal succinct PRG.

Of course, the issue is not simply *truth* but also *provability*. Is it likely the Universality Conjecture will be proved in the near future? We do not have a strong belief about this, but there is at least some reason to hope that more sophisticated versions of the proof technique of Theorem 1 might help. More precisely, understanding uniformity and succinctness of approximate strategies for Min-Max games [Alt94, LY94] is a possible direction.

Regardless of whether the Conjecture will be settled in the near future, it could function as an organizing principle connecting various fundamental phenomena that we still don't understand well, including the complexity of MCSP, the hardness of learning, the relationship between uniform and non-uniform versions of one-way functions, the structure of zero knowledge, reducibilities between average-case problems over the uniform distribution, and the role of pseudorandomness in proof complexity, among others.

1.2.3 Non-Black-Box Results

Finally, we show a couple of *non-black-box* results about MCSP. As a meta-complexity problem, i.e., a problem whose instances themselves encode computations, MCSP seems particularly amenable to non-black-box techniques. Hopefully this amenability will come in useful in establishing strong unconditional connections between the hardness of MCSP and the existence of one-way functions. Basing one-way functions on NP-hardness in a black-box fashion has unlikely consequences [AGGM06], so if MCSP did turn out to be NP-complete and we wished to base one-way functions on its *worst-case* hardness, we would need to use non-black-box techniques.

Our first result is about circuit classes that are weak in the sense that there are natural proofs useful against them. For such circuit classes (which include AC^0, $\mathsf{AC}^0[p]$ for prime p, etc.), we establish a surprising implication from lower bounds for MCSP to lower bounds for P.

Theorem 4. *(Informal Statement) Let $\mathfrak{D}$ be any circuit class closed under projections for which there are natural proofs against $\mathfrak{D}$. If there is a constant k such that* MCSP *with parameter n^k is zero-error hard on average against $\mathfrak{D}$, then* P *is not contained in $\mathfrak{D}$.*

The proof of Theorem 4 is non black-box, i.e., it does not work when the circuit class $\mathfrak{D}$ against which we are arguing is given access to an oracle. Indeed, we can prove that under standard crypto assumptions, there is no black-box implication of the sort we show.

The reason we find the implication in Theorem 4 somewhat surprising is that the hypothesis is about the hardness of a problem in NP and unlikely to be in P, indeed even believed by many to be NP-complete. Yet the conclusion is about super-polynomial lower bounds within P!

Our next result addresses the relaxed version of Question 10. We observe that a recent search-to-decision reduction of [CIKK17] for a slight variant of MCSP called AveMCSP (where the question is whether there is a small circuit that computes the input truth table correctly on a 0.9 fraction of inputs) actually gives an approximation to average-case reduction. We note that Shuichi Hirahara [Hir18] independently observed an analogous approximation to average-case reduction for the standard version of MCSP, based on [CIKK16], however the approximation factor there is much larger.

Theorem 5. *(Informal Statement) There is an approximation to average-case reduction for* AveMCSP.

Theorem 5 has implications for the NP-hardness of problems such as MCSP and AveMCSP. For several NP-hard problems, the theory of probabilistically checkable proofs establishes strong inapproximability results under the $\mathsf{NP} \neq \mathsf{P}$ assumption. Thus an approximation to average-case reduction can be considered morally similar to a worst-case to average-case reduction. It is known under standard complexity assumptions that NP-complete problems do not have black-box worst-case to average-case reductions [FF93, BT06].

Does this suggest that AveMCSP is unlikely to be NP-complete? Not quite! The proof of Theorem 5 is *also* non black-box, even though the ideas are quite different from the proof of Theorem 4.

1.3 Related Work

The first paper to connect the complexity of MCSP with the existence of one-way functions was the "natural proofs" paper of Razborov and Rudich [RR97]. Razborov and Rudich do not explicitly consider MCSP - indeed, this problem was only defined in subsequent work of Kabanets and Cai [KC00]. The main result of [RR97] states that exponentially-hard one-way functions imply the non-existence of natural proofs with poly-size constructibility that are useful against polynomial-size circuits. This can be re-interpreted [HS17] as saying that if exponentially-hard one-way functions exist, then $\mathsf{MCSP}[\mathrm{poly}(n)]$ is zero-error hard on average. Thus an average-case algorithm for MCSP can be used to break *any* one-way function. The main question we consider in this paper is whether the converse holds.

Various works have attempted to connect the existence of one-way functions and their variants with other complexity notions. Ostrovsky and Wigderson [OW93] showed that if $\mathsf{ZK} \neq \mathsf{BPP}$, then auxiliary-input one-way functions exist. They also showed that if ZK is hard for BPP on average (in the bounded-error sense), then one-way functions exist. [IL90] and [BFKL93] (see also [ABX08]) show equivalences between the existence of one-way functions and certain average-case hardness assumptions for learning.

More recently, interest in the complexity of MCSP and its connections with learning and pseudorandomness has been re-awakened by the result of [CIKK16] which shows that natural proofs useful against a circuit class $\mathfrak{C}$ imply efficient learning algorithms for $\mathfrak{C}$, under certain reasonable conditions on $\mathfrak{C}$. Building on this, [OS17] show equivalences between various forms of learning, as well as a dichotomy between learning and pseudorandomness in a certain parameter regime.

A very recent work of Hirahara [Hir18] (who obtained his results independently from ours) builds on [CIKK16] to show that solving MCSP on average in the zero-error sense implies efficient non-trivial approximability of the minimum circuit size. Thus, if MCSP were NP-hard even to approximate non-trivially, the existence of hard-on-average problems in NP would follow. This suggests the possibility of excluding Impagliazzo's Heuristica world [Imp95] where NP is hard in the worst case but not hard on average by studying MCSP and showing that it is NP-hard to approximate.

In contrast to [Hir18], our focus in this paper is on finding evidence against the existence of Pessiland: another world of Impagliazzo's [Imp95] where there are hard-on-average problems but one-way functions do not exist.

2 Preliminaries

2.1 Notation

For convenience, we use $f\colon K \to N$ to denote a function mapping K bits to N bits, i.e., $f\colon \{0,1\}^K \to \{0,1\}^N$. We let $\mathcal{F}_{K\to N}$ denote the family of all functions $f\colon K \to N$. We sometimes view a boolean function $f\colon N \to 1$ as a subset of $\{0,1\}^N$.

Throughout this paper, we use capital letters to denote the input length and output length of a function when we are interested in interpreting the output as the truth table of a Boolean function. In such a case, we will use a lowercase letter to denote the logarithm of the number represented by the corresponding uppercase letter.

We let $\mathcal{U}_L$ denote the uniform distribution over $\{0,1\}^L$. We occasionally abuse notation and identify a set S with the uniform distribution over the set. We say that a function $f\colon K \to N$

ε-fools a family of functions $\mathcal{D} \subseteq \mathcal{F}_{N\to 1}$ if $|\Pr[g(f(\mathcal{U}_K)) = 1] - \Pr[g(\mathcal{U}_N) = 1]| \leq \varepsilon$ for every $g \in \mathcal{D}$.

A function $g\colon N \to 1$ is γ-dense if $\Pr[g(\mathcal{U}_N) = 1] \geq \gamma$. We say that $f \in \mathcal{F}_{K\to N}$ *hits* g is $f(\{0,1\}^K) \cap g^{-1}(1) \neq \emptyset$.

Given a Boolean function $f \in \mathcal{F}_{n\to 1}$, $\mathtt{tt}(f)$ is the 2^n-bit string which represents the truth table of f in the standard way, and conversely, given a string $y \in \{0,1\}^{2^n}$, $\mathtt{fn}(y)$ is the Boolean function in $\mathcal{F}_{n\to 1}$ whose truth table is represented by y.

Given a circuit class $\mathfrak{C}$, we use $\mathfrak{C}[s(n)]$ to refer to the class of functions computed by $\mathfrak{C}$-circuits of size $s(n)$. Given a language $L \subseteq \{0,1\}^*$, co-L denotes the complement of L.

We say a circuit class $\mathfrak{C}$ is standard if there is a quasi-linear time Turing machine which, given a representation $\tilde{C}$ of a circuit C from the class and an input x, computes $C(x)$. All commonly studied circuit classes contained in the class of general Boolean circuits are standard.

For a size bound $s\colon \mathbb{N} \to \mathbb{N}$, we use quasi-$s(n)$ to denote a function of growth rate $s(n) \cdot \mathsf{poly}(\log s(n))$.

The density of a set $A \subseteq \{0,1\}^N$ is $|A|/2^N$. Given a language L and an integer n, $L_n = L \cap \{0,1\}^n$ denotes the slice of L at length n.

2.2 Pseudorandomness and Average-Case Hardness

Let $K\colon \mathbb{N} \to \mathbb{N}$ be a function such that $K(N) < N$ for all N, and $\epsilon : \mathbb{N} \to [0,1]$ be a function. Moreover, let $\mathfrak{C}$ be a complexity class and $\mathfrak{D}$ be a class of functions. A $\mathfrak{C}$-PRG against $\mathfrak{D}$ with seed length K and error ϵ is a sequence $f = \{f_N\}_{N\in\mathbb{N}}$ of functions such that (*i*) $f \in \mathcal{F}_{K(N)\to N}$; (*ii*) $f \in \mathfrak{C}$; and (*iii*) the output of f_N $\epsilon(N)$-fools every function in $\mathcal{F}_{N\to 1} \cap \mathfrak{D}$, whenever N is sufficiently large. If ϵ is left unspecified, it will be taken to be $1/N^{\omega(1)}$ by default. Similarly $\mathfrak{C}$ will be taken to be $\mathsf{DTIME}(\mathrm{poly}(N))$ by default, and $\mathfrak{D}$ will be taken to be $\mathsf{SIZE}(\mathrm{poly}(N))$ by default. Note that in this default situation, the generator is computable in a *fixed* polynomial time in its output length, but must be secure against arbitrary poly-size distinguishers.

We say $f = \{f_N\}_{N\in\mathbb{N}}$ is a $\mathfrak{C}$-HSG against $\mathfrak{D}$ if instead of condition (*iii*) above the output of f_N hits every set $A_N \subseteq \{0,1\}^N$ of density $\epsilon(N)$ in $\mathfrak{D}$. Again the error parameter is taken to be $1/N^{\omega(1)}$ by default.

It is instructive to consider the following examples: standard cryptographic PRGs are default PRGs with seed length $K(N) = N^{\Omega(1)}$, while complexity-theoretic (Nisan-Wigderson) PRGs are E-PRGs against $\mathsf{SIZE}(N)$ with error $1/N$, where the seed length can be anything between $\log N$ and $N - 1$.

We say the seed length K is non-trivial if $K(N) < N$.

We will be interested in *succinct* pseudorandom distributions and hitting sets. Let $\mathfrak{C}$ be a circuit class and $\epsilon : \mathbb{N} \to [0,1]$ be an error bound. We say that there are $\mathfrak{C}$-succinct pseudorandom distributions against $\mathfrak{D}$ with error ϵ if there is a set $S \subseteq \{0,1\}^*$ and a probability distribution μ_S supported on S such that for each N a power of 2, (*i*) For each $y \in S_N$, $\mathtt{fn}(y) \in \mathfrak{C}$; and (*ii*) $|\Pr_{y\leftarrow\mu_S}[g(y) = 1] - \Pr_{y\leftarrow\mathcal{U}_N}[g(y) = 1]| \leq \epsilon(N)$ for every $g \in \mathcal{D}$. By default, we will take $\epsilon(N)$ to be $1/N^{\omega(1)}$ as usual.

Similarly, we say that there are $\mathfrak{C}$-succinct hitting sets against $\mathfrak{D}$ with error ϵ if there is a set $S \subseteq \{0,1\}^*$ such that for each N a power of 2, (*i*) For each $y \in S_N$, $\mathtt{fn}(y) \in \mathfrak{C}$; and (*ii*) For each set $A_N \subseteq \{0,1\}^N$ of density $\epsilon(N)$, S_N has non-empty intersection with A_N.

A $\mathfrak{C}$-succinct PRG is simply a PRG that induces a $\mathfrak{C}$-succinct pseudorandom distributions when a seed of any given length is chosen uniformly at random, and a $\mathfrak{C}$-succinct HSG is a HSG whose

range is a collection of $\mathfrak{C}$-succinct hitting sets.

Definition 1 (Universal Succinct Generators)**.** *Let $\mathfrak{C}$ be a circuit class and $K : \mathbb{N} \to \mathbb{N}$ be a function. A universal $\mathfrak{C}$-succinct HSG with seed length K is a poly-time computable sequence $f = \{f_N\}_{N\in\mathbb{N}}$ of functions from $K(N)$ bits to N bits such that if there are $\mathfrak{C}$-succinct hitting sets against $\mathsf{SIZE}(\text{poly})$ with error $\epsilon(N)$, then f is a $\mathfrak{C}$-succinct HSG with error $\epsilon(N)$. A universal $\mathfrak{C}$-succinct PRG with seed length K is a poly-time computable family f of functions from $K(N)$ bits to N bits such that if there are $\mathfrak{C}$-succinct pseudorandom distributions against $\mathsf{SIZE}(\text{poly})$ with error $\epsilon(N)$, then f is a $\mathfrak{C}$-succinct PRG with error $\epsilon(N)$. A non-uniform universal $\mathfrak{C}$-succinct PRG with seed length K is a poly-size computable family f of functions from $K(N)$ bits to N bits such that if there are $\mathfrak{C}$-succinct pseudorandom distributions against $\mathsf{SIZE}(\text{poly})$ with error $\epsilon(N)$, then f is a $\mathfrak{C}$-succinct $\mathsf{SIZE}(\text{poly})$-PRG with error $\epsilon(N)$.*

Given $\epsilon : \mathbb{N} \to \mathbb{R}$, a one-way function with security $\epsilon(N)$ is a polynomial-time computable sequence of functions $f = \{f_N : K(N) \to N\}$ for some $K(N) = N^{\Omega(1)}$, such that for any sequence $\{C_N\}$ of polynomial-size circuits, for large enough N, $\Pr_{x\sim U_N}[f_N(C_N(1^N, f(x))) = f_N(x)] \leq \epsilon(N)$. A one-way function is called *weak* if it has security $1 - 1/\operatorname{poly}(N)$, and *strong* if it has security $1/N^{\omega(1)}$. One-way functions are often defined against uniform adversaries; in this work, we only consider security against non-uniform adversaries.

Informally, an auxiliary-input one-way function is a poly-time computable sequence of functions f with an 'auxiliary' input z such that for any poly-size adversary, there is an infinite set of auxiliary inputs z for which f is hard to invert. For a formal definition and a good discussion of the significance of auxiliary-input one-way functions, see [Vad06].

We require a notion of zero-error average-case hardness for languages. Given a parameter $\epsilon : \mathbb{N} \to [0, 1]$, a language $Q \subseteq \{0, 1\}^*$, and a circuit class $\mathfrak{D}$, we say that Q is zero-error average-case feasible for $\mathfrak{D}$ with success probability ϵ if there is a sequence of circuits $D_N \in \mathfrak{D}$ such that for each N, D_N always outputs 0,1 or '?', never outputs the wrong answer for any input to Q, and outputs '?' with probability at most $1 - \epsilon(N)$. We say Q is zero-error average-case easy if it is average-case feasible for $\mathsf{SIZE}[\text{poly}]$ with success probability $1/N^{O(1)}$.

We say that Q is zero-error average-case infeasible for $\mathfrak{D}$ with success probability ϵ if for every sequence of circuits $D_N \in \mathfrak{D}$ such that D_N always outputs 0,1 or '?', for large enough N, D_N either outputs the wrong answer for some input to Q or outputs '?' with probability greater than $1 - \epsilon(N)$. We say Q is zero-error average-case hard if it is average-case infeasible for $\mathsf{SIZE}[\text{poly}]$ with success probability $1/N^{O(1)}$.

Note that hardness is not just defined as the complement of easiness - hardness and easiness are both required to hold on almost all input lengths.

2.3 MCSP, Natural Proofs and Learning

$\mathsf{MCSP}[s(n)]$ denotes the Minimum Circuit Size Problem for a size parameter $s(n)$, where the truth table given as input to the problem has size 2^n. In other words, an input truth table is a positive instance if and only if it is computed by a circuit of size at most $s(n)$. This notation is adopted for convenience. Thus the most interesting functions $s(n)$ would satisfy $n \leq s(n) \leq 2^n$. Given a class $\mathfrak{C}$, $\mathfrak{C}$-$\mathsf{MCSP}[s(n)]$ denotes MCSP for $\mathfrak{C}$-circuits. In case the input size N is not a power of two, we consider the input truth table to be the first $2^{\lfloor \log(N) \rfloor}$ bits of the input.

We also require some variants of MCSP. Given an approximation parameter $\delta : \mathbb{N} \to [0, 1]$, and size functions $c, s : \mathbb{N} \to \mathbb{N}$ with $n \leq c(n) \leq s(n) \leq 2^n$ for all n, δ-$\mathsf{AveMCSP}[c, s]$ is the

following promise problem: YES instances are truth tables of Boolean functions that can be $\delta(n)$-approximated by circuits of size at most $c(n)$, and NO instances are truth tables of Boolean functions that cannot be δ-approximated by circuits of size at most $s(n)$. When δ is left unspecified, we take it to be 0.9 for convenience.

We say that $\mathfrak{R} = \{\mathcal{R}_n\}_{n\in\mathbb{N}}$ is a *combinatorial property* (of boolean functions) if $\mathcal{R}_n \subseteq \mathcal{F}_n$ for all n. We use $L_{\mathfrak{R}}$ to denote the language of truth-tables of functions in $\mathfrak{R}$. Formally, $L_{\mathfrak{R}} = \{y \mid y = \mathtt{tt}(f)$ for some $f \in \mathcal{R}_n$ and $n \in \mathbb{N}\}$.

Definition 2 (**Natural Properties** [RR97])**.** *Let $\mathfrak{R} = \{\mathcal{R}_n\}$ be a combinatorial property, $\mathfrak{C}$ a circuit class, and $\mathfrak{D}$ a (uniform or non-uniform) complexity class. We say that $\mathfrak{R}$ is a $\mathfrak{D}$-natural property useful against $\mathfrak{C}[s(n)]$ if there is $n_0 \in \mathbb{N}$ such that the following holds*:

(*i*) Constructivity. $L_{\mathfrak{R}} \in \mathfrak{D}$.

(*ii*) Density. *For every $n \geq n_0$,* $\Pr_{f\sim\mathcal{F}_n}[f \in \mathcal{R}_n] \geq 1/2$.

(*iii*) Usefulness. *For every $n \geq n_0$, we have $\mathcal{R}_n \cap \mathcal{C}_n[s(n)] = \emptyset$.*

By default, $\mathfrak{D}$ is taken to be $\mathsf{SIZE}[\mathrm{poly}]$.

It has been observed that the density parameter in the definition above can be amplified using a straightforward reduction (cf. [CIKK16]).

Proposition 1. *[HS17] Let $s : \mathbb{N} \to \mathbb{N}$ be a size function.* $\mathsf{MCSP}[s(n)]$ *is zero-error easy on average iff there are* $\mathsf{SIZE}(\mathrm{poly})$*-natural properties useful against* $\mathsf{SIZE}[s(n)]$.

We give only a brief description of learning in Valiant's PAC model. We will be concerned here only with learning using membership queries under the uniform distribution. A learning algorithm in this context is a probabilistic oracle algorithm making queries to an oracle for some function f. Given a circuit class $\mathfrak{C}$ and a function $T : \mathbb{N} \to \mathbb{N}$, a time T learner for $\mathfrak{C}$ is an oracle algorithm which given input 1^n and oracle access to a function f from $\mathcal{C}$ halts in time $T(n)$ and outputs with high probability a circuit computing f correctly on a 0.9 fraction of inputs of length n.

3 Hitting Sets, Pseudorandom Distributions and the Hardness of MCSP

3.1 Succinct HSGs and MCSP

We first observe that there is a universal succinct HSG.

Proposition 2. *Let $\mathfrak{C}$ be any standard circuit class, and $s : \mathbb{N} \to \mathbb{N}$ be any function such that $n \leq s(n) \leq 2^n$ for all n. There is a universal $\mathfrak{C}[s(n)]$-succinct HSG U with seed length* quasi-$s(n)$.

Proof. The idea of the proof is simple. We define a HSG U with seed length quasi-$s(n)$ that interprets its seed as the representation of a $\mathfrak{C}$-circuit of size $s(n)$, and outputs the truth table of the function computed by this circuit. U can be computed in time $2^n\mathsf{quasi}$-$s(n)$, which is quasi-quadratic in $N = 2^n$.

We need to show that U is a universal $\mathfrak{C}[s(n)]$-succinct HSG. The $\mathfrak{C}[s(n)]$-succinctness is immediate from the definition of the HSG - every output of the HSG is the truth table of a Boolean function in $\mathfrak{C}[s(n)]$. For the universality, suppose that there is some collection H of $\mathfrak{C}[s(n)]$-succinct

hitting sets. Every string in this collection is the truth table of a Boolean function in $\mathfrak{C}[s(n)]$. Hence H is contained in the range of U, from which it follows that the range of U is a collection of hitting sits. □

Next we observe that the existence of succinct hitting set generators is equivalent to hardness for MCSP. This observation is closely related to similar observations in the theory of Boolean or arithmetic circuit complexity and meta-mathematics of lower bounds [Wil16, GKSS17, FSV17].

Proposition 3. *Let $\mathfrak{C}$ be a standard circuit class and $\mathfrak{D}$ be a family of functions. Moreover, let $s(n) \geq n$. For every large n, the following are equivalent:*

1. *There are $\mathfrak{C}$[quasi-$s(n)$]-succinct hitting sets against $\mathfrak{D}$*
2. *There is a $\mathfrak{C}$[quasi-$s(n)$]-succinct HSG with seed length quasi-$s(n)$ against $\mathfrak{D}$*
3. *$\mathfrak{C}$-MCSP[quasi-$s(n)$] is zero-error average-case hard against $\mathfrak{D}$.*

Proof. The second item follows from the first by Proposition 2.

Next we show that the third item follows from the second. Let H be a $\mathfrak{C}$[quasi-$s(n)$]-succinct HSG against $\mathfrak{D}$. Assume for the sake of contradiction that $\mathfrak{C}$-MCSP[quasi-$s(n)$] is zero-error average-case solvable with success probability $1/N^{O(1)}$ in $\mathfrak{D}$, where $N = 2^n$. This implies that there is a subset A of co-$\mathfrak{C}$-MCSP[quasi-$s(n)$] with density $1/N^{O(1)}$. For each large N, A_N has no strings of $\mathfrak{C}$-circuit complexity $\leq$ quasi-$s(n)$, hence A_N does not intersect the range of H_N for such N. Yet the density of A_n is at least $1/N^{O(1)}$, which contradicts the assumption that H is a HSG against $\mathfrak{D}$.

Finally we show that the first item follows from the third. Suppose $\mathfrak{C}$-MCSP[quasi-$s(n)$] is zero-error average-case hard against $\mathfrak{D}$. Consider the set of truth tables of functions with $\mathfrak{C}$-complexity at most quasi-$s(n)$. We show that this is a collection of $\mathfrak{C}$[quasi-$s(n)$]-succinct hitting sets against $\mathfrak{D}$. The succinctness condition follows from the definition of the set. To show that this is a collection of hitting sets, suppose to the contrary that there is a set $A \in \mathfrak{D}$ such that A_N has density $1/N^{O(1)}$ for each N a power of 2, and A does not intersect the collection. Then we can define a zero-error average-case algorithm which outputs 0 for $x \in A$ and '?' otherwise. This algorithm only outputs 0 on truth tables that do not have quasi-$s(n)$ size $\mathfrak{C}$-circuits, hence it always outputs the correct answer when it does not output '?'. By the density condition on A, the algorithm has success probability $1/N^{O(1)}$.

□

3.2 From Zero-Error Average-Case Hardness to Succinct Pseudorandomness

In this section, we show that zero-error average-case hardness for MCSP in fact yields succinct *pseudorandom* distributions, where the complexity parameter of the succinctness is slightly worse than that of the MCSP problem we assume to be hard.

In fact, we show something more general for languages $Q, Q' \subseteq \{0,1\}^*$: assuming the existence of certain 'semantic' samplers, zero-error average-case hardness over the uniform distribution for a problem Q implies the existence of pseudorandom distributions supported on YES instances of Q'. The main idea for showing the implication to pseudorandomness is to analyze a family of *Sampler-Distinguisher* zero-sum games, which we introduce in this work. This is inspired by, but different from, the PRF-Distinguisher game analyzed in [OS17]. The strategies of the row player

in the Sampler-Distinguisher game are YES instances of a given length for the problem Q we wish to solve. The strategies of the column player are circuits from some circuit class $\mathfrak{D}$, corresponding to the class against which we are analyzing average-case hardness.

The payoff corresponding to a row (instance) x and column (circuit) D is defined as the average of $D(x_j), j = 1 \dots k$ minus the expectation of D on inputs of length $|x_j|$. Here the strings x_j are generated from x to have the following properties. First, when x is a YES instance of Q, the x_j's are YES instances of Q'. Second, when y is random, the set $\{y_i\}$ is a good sampler with high probability, meaning that it can be used to estimate the expectation of any bounded function to within reasonable error.

We are able to argue that when the Row player wins the game, there are succinct pseudorandom distributions against $\mathfrak{D}$, and when the column player wins the game, there is a zero-error average case algorithm for Q. The argument in the second case relies on the approximately optimal succinct strategies for zero-sum games given by [Alt94, LY94].

Lemma 1. *[Alt94, LY94] Let M be a $r \times c$ matrix with entries in $[-1, 1]$ representing the payoffs of a zero-sum game. Let $v(M)$ denote the value of the game. Let $\delta < 1$ be a parameter. Then there is a strategy for the row (Min) player supported uniformly on at most $10\log(c)/\delta^2$ pure strategies that guarantees her a payoff at most $v(M)+\delta$ and a strategy for the column (Max) player supported uniformly on at most $10\log(r)/\delta^2$ pure strategies that guarantees here a payoff at least $v(M) - \delta$.*

We will need a special case of the standard Hoeffding inequalities.

Proposition 4. *[Hoe63] Let $X_1 \dots X_n$ be independent random variables taking values in $[-1, 1]$. Let $\bar{X} \stackrel{\text{def}}{=} \sum_i X_i/n$ denote the empirical mean of these variables. Then, for any $t > 0$, $\Pr(|\bar{X} - E(\bar{X})| \geq t) \leq 2e^{-t^2 n/2}$.*

We now define the notion of semantic sampler we require.

Definition 3. *Let $Q, Q' \subseteq \{0,1\}^*$ be languages, $\ell, m : \mathbb{N} \to \mathbb{N}$ and $\epsilon, \delta : \mathbb{N} \to [0, 1]$ be functions. A poly-time computable sequence of functions $f = \{f_N : \{0,1\}^N \times \{0,1\}^{\ell(N)} \to \{0,1\}^{m(N)}\}$ is a (Q, Q')-semantic sampler f with seed length ℓ, output length m, accuracy ϵ and error δ if:*

1. *(Semantic condition) For large enough $N \in \mathbb{N}$, for all $x \in \{0,1\}^N$ and $y \in \{0,1\}^{\ell(N)}$, $x \in Q$ implies $f_N(x, y) \in Q'$.*
2. *(Sampling condition) For large enough $N \in \mathbb{N}$, for every function $g : \{0,1\}^{m(N)} \to [-1, 1]$, for all but a $\delta(N)$ fraction of N-bit strings x, $|E_{z \in U_{m(N)}} g(z) - E_{y \in U_{\ell(N)}} g(f_N(x, y))| \leq \epsilon(N)$.*

We are ready to prove our general connection between zero-error average-case hardness and pseudorandomness.

Theorem 6. *Let $Q, Q' \subseteq \{0,1\}^*$ be languages, $m : \mathbb{N} \to \mathbb{N}$ be a surjective function such that $m(N) \leq N$ for all $N \in \mathbb{N}$, and $\epsilon, \delta : \mathbb{N} \to [0, 1]$ be functions. Suppose there is a (Q, Q')-semantic sampler f with output length m, accuracy ϵ and error δ. If Q is zero-error average-case infeasible for $\mathsf{SIZE}[\text{poly}]$ with success probability $1 - \delta(N)$, then there are pseudorandom distributions with error $30\epsilon(N)$ against $\mathsf{SIZE}[\text{poly}]$ that are supported on Q'.*

Proof. Let $t : \mathbb{N} \to \mathbb{N}$ be a size function, and f be the (Q, Q')-semantic sampler in the statement of the theorem. We define the (f, t) Sampler-Distinguisher family of zero-sum games at level N by their payoff functions as follows. For each positive integer N, the set A_N corresponding to strategies of the Row (Min) player is defined to be the set of N-bit strings x such that $x \in Q$. The set B_N corresponding to strategies of the Column (Max) player is defined to be the set of Boolean circuits with input length $m(N)$ and size at most $t(N)$.

Now we define the Column player's payoff function $P_N : A_N \times B_N \to [-1, 1]$ as follows. Given a string $x \in A_N$ and a circuit $D \in B_N$, $P_N(x, D) = E_{y \in U_{\ell(N)}} D(f_N(x, y)) - E_{z \in U_{m(N)}} D(z)$. Since the class of Boolean circuits is closed under complement, we can assume wlog that the value of the game is non-negative for each N.

Let $\gamma : \mathbb{N} \to [0, 1]$ be a monotonically decreasing error parameter to be specified later. There are two cases: either for every polynomially bounded t, for almost all N, the value of the (f, t) Sampler-Distinguisher game at level N is at most $\gamma(N)$, or there exists some polynomially bounded t such that for infinitely many N, the value of the game at level N is greater than $\gamma(N)$.

In the first case, using Lemma 1, for each $t(N)$, for N large enough, there is a $\gamma(N)/2$-approximately optimal strategy of the Row player for the game at level N that is a uniform distribution on a multiset R of strings x of length N in Q.

Now the induced distribution on strings $z = f_N(x, y)$ obtained by picking x uniformly from R and y uniformly from strings of length $\ell(N)$ is a pseudorandom distribution with error at most $3\gamma(N)/2$ against circuits with input length $m(N)$ and size at most $t(N)$, just by linearity of expectation. Then we have that there are pseudorandom distributions with error at most $3\gamma(N)/2$ against $\mathsf{SIZE}(\text{poly})$ that are supported on Q', using the fact that $R \subseteq Q$ and that $f(x, y) \in Q'$ whenever $x \in Q$ and $y \in \{0, 1\}^{\ell(N)}$. We will eventually choose $\gamma(N) = 20\epsilon(N)$, which yields pseudorandom distributions with error at most $30\epsilon(N)$.

In the second case, let t be polynomially bounded in N and I' be an infinite set of input lengths such that the value of the (f, t) Sampler-Distinguisher game is at least $\gamma(N')$ on each $N' \in I'$. Define the set I of input lengths as follows: $N \in I$ iff $m(N) \in I'$. Since I' is infinite and m is surjective, I is also infinite. We show that there is a sequence of circuits $\{C_N\}$ of polynomial size such that for each input length $N \in I$, C_N solves Q well on average, contradicting the assumption that Q is average-case hard against $\mathsf{SIZE}[\text{poly}]$.

Since the value of the game is at least $\gamma(N')$ on each $N' \in I'$, we have that for each $N' \in I'$, there is an approximately optimal strategy for the Column player supported uniformly on at most $O(N'/\gamma(N')^2)$ pure strategies that guarantees her a payoff at least $\gamma(N')/2$. Note that each pure strategy is simply a circuit of size $O(t)$. For each $N' \in I'$, define the probabilistic circuit $C'_{N'}$ on input z of length N' as follows. $C'_{N'}$ samples uniformly a circuit D from the set of circuits $S_{N'}$ in the support of the approximate strategy for the Column player, and outputs $D(z)$. $C'_{N'}$ can be implemented straightforwardly to have size at most $O(tN'/\gamma(N')^2)$. Let v denote the value $E_{z \leftarrow U_{N'}, D \leftarrow S_{N'}} D(z)$.

Next we define the circuit C_N. We first construct a probabilistic circuit and then show how to fix the randomness. On input x of length N, the circuit samples y uniformly at random from $\{0, 1\}^{\ell(N)}$ and computes $f_N(x, y)$. C_N runs the probabilistic circuit C' independently $100N/(\gamma(N))^2$ times on $f_N(x, y)$ for uniformly and independently chosen $y \in \{0, 1\}^{\ell(N)}$. It computes the average v' of the outputs obtained by C' over all these runs, and checks if $v' - v \geq \epsilon/4$. In order to implement this check, the value v (which does not depend on x) is hardcoded into C. If the majority of checks succeed, C_N outputs '?', else it outputs 0.

Let $N \in I$ and $N' = m(N)$. We show that if x of length N is a YES instance of Q, C_N always outputs '?' with probability $> 1 - 2^{-N}$, and for a $1 - o(1)$ fraction of inputs of length N, C_N outputs 0 with probability $> 1 - 2^{-N}$. By fixing the randomness of C_N using Adleman's trick, we get deterministic circuits which *always* output '?' on YES inputs, and output 0 on almost all NO instances. Moreover, they output either '0' or '?' on every NO instance. This implies that for each $N \in I$, C_N is a circuit that solves Q well on average.

We first establish our claim for YES instances. Suppose x of length N is in Q. This implies that for all $y \in \{0,1\}^{\ell(N)}$, $f_N(x,y) \in Q'$ by the semantic condition on the (Q,Q')-semantic sampler f. Hence, for each $y \in \{0,1\}^{\ell(N)}$, $f_N(x,y)$ is a pure strategy for the Row player in the (s'',t) Sampler-Distinguisher game at level N'. Since $N' \in I'$, the game at level N' has value greater than $\gamma(N')$. In particular, this means that the Row player's strategy of playing $f(x, U_{\ell N})$ yields payoff at least $\gamma(N')/2$ in expectation to the Column player when the Column player plays the approximately optimal strategy corresponding to the probabilistic circuit C'. Thus, over the randomness of C' and random choice of $y \in \{0,1\}^{\ell(N)}$, $E[C'(f_N(x,y))] > v + \gamma(N')/2$. Since γ is a monotonically decreasing function, and $N' \leq N$, we have that $E[C'(f_N(x,y))] > v + \gamma(N)/2$. Applying Proposition 4, when C' is simulated $100N/(\gamma(N))^2$ times independently on uniformly chosen $y \in \{0,1\}^{\ell(N)}$ to compute the empirical average v', $v' > v + \gamma(N)/4$ with probability $> 1 - 2^{-N}$. Hence C_N outputs '?' with at least this probability, as claimed.

Next we argue that when $\gamma(N)$ is chosen to be $20\epsilon(N)$, for all but approximately $\delta(N)$ fraction of inputs of length N, C_N outputs 0 with probability $> 1 - 2^{-N}$. Intuitively, this follows from the fact that a large enough randomly chosen set of strings is a good sampler for a function with range $[-1, 1]$.

More formally, consider the quantity v defined above as the expectation of $C'(z)$ over uniformly chosen N'-bit z and randomness of the circuit C'. We upper bound the probability, for a uniformly chosen x of length N, that the empirical average v' of the outputs obtained by $100N/(\gamma(N))^2$ independent runs of $C'(f_N(x,y))$ for uniformly chosen y, is greater than $v + \gamma(N)/4$.

By the sampling condition for the (Q,Q')-semantic sampler f, with probability at least $1 - \delta(N)$ over uniformly chosen x of length N, $|E_{y \in U_{\ell(N)}} C'(f_N(x,y)) - v| \leq \epsilon(N)$. Call a string x "good" if this inequality is satisfied. Applying Proposition 4 again, for any x, with probability $> 1 - 2^{-N}$ over the internal randomness of C, $|v' - E_{y \in U_{\ell(N)}} C'(f_N(x,y))| \leq \gamma(N)/5$. In particular, for good x, by the triangle inequality, we have that with probability $> 1 - 2^{-N}$, $|v' - v| < \gamma/5 + \epsilon(N) \leq \gamma(N)/5 + \gamma(N)/20 = \gamma(N)/4$. Thus, for all but a $\delta(N)$ fraction of strings x of length N, C_N outputs 0 with probability $> 1 - 2^{-N}$, as claimed. □

We now show how to use Theorem 6 to derive pseudorandom distributions from zero-error average-case hardness for MCSP and SAT, by showing the existence of appropriate semantic samplers. Indeed, in both cases, we will use the simple sampler defined below.

Definition 4. *Given a function $q : \mathbb{N} \to \mathbb{N}$ such that $q(N) \leq N$ is a power of two for all N, we define the $q(N)$-projection sampler to be the function sequence $Proj = \{Proj_N : \{0,1\}^N \times \{0,1\}^{\log(q(N))} \to \{0,1\}^{\lfloor N/q(N) \rfloor}$, where for any string $x = x_1 x_2 \dots x_N$ of length N and a string y of length $\log(q(N))$ interpreted in the standard way as an integer in $[q(N)]$, $Proj_N(x,y) = x_{y\lfloor N/q(N) \rfloor} \dots x_{(y+1)\lfloor N/q(N) \rfloor - 1}$. In other words, $Proj_N(x,y)$ is the contiguous substring of x of length $\lfloor N/q(N) \rfloor$ beginning at index $y\lfloor N/q(N) \rfloor$.*

We first show how to apply Theorem 6 in the case of MCSP.

Theorem 7. *Let $\mathfrak{C}$ be any circuit class closed under projections. Let $s : \mathbb{N} \to \mathbb{N}$ be a size function, and $q : \mathbb{N} \to \mathbb{N}$ be any super-constant poly-time computable function such that $q(N) \leq N$ is a power of two for all N. Let $\epsilon : \mathbb{N} \to \mathbb{R}$ be any poly-time computable monotonically decreasing error function such that $\epsilon(N) = \omega(1/\sqrt{q(N)})$. If $\mathfrak{C}$-MCSP$[s(n)]$ is zero-error average-case infeasible for* SIZE[poly] *with success probability $1 - o(1)$, then there are $\mathfrak{C}[s']$-succinct pseudorandom distributions with error $O(\epsilon(N))$ against* SIZE[poly]*, where $s' : \mathbb{N} \to \mathbb{N}$ is any size function such that $s'(n - \log(q(2^n))) \geq s(n)$.*

Proof. Let Q be $\mathfrak{C}$-MCSP$[s(n)]$, and Q' be $\mathfrak{C}$-MCSP$[s'(n)]$, where s' is any size function such that $s'(n - 2\log(q(2^n))) \geq s(n)$, as in the statement of the theorem. We show that the $q(N)$-projection sampler $Proj$ is in fact a (Q, Q')-semantic sampler with error $o(1)$ and accuracy $\epsilon(N)$, and then apply Theorem 6.

First we show the semantic condition. Let x of length N be the truth table of a Boolean function with $\mathfrak{C}$-circuits of size at most $s(n)$, where $n = \log(N)$. We show that for any $y \in \{0,1\}^{\log(q(n)}$, $Proj_N(x, y)$ is the truth table of a Boolean function with $\mathfrak{C}$-circuits of size at most $s'(n)$. Indeed, $Proj_N(x, y)$ is the truth table of a function on $n - \log(q(N))$ bits, obtained from x by fixing the first $\log(q(n))$ input bits. Let C_N be a $\mathfrak{C}$-circuit of size $s(n)$ computing $\mathtt{fn}(x)$. By fixing $\log(q(n))$ input bits of C_N to constants, and using the fact that $\mathfrak{C}$ is closed under projections, we obtain a $\mathfrak{C}$-circuit of size at most $s(n)$ computing $\mathtt{fn}(Proj(x, y))$, which is a function with $n - \log(q(N))$ input bits. Since $s'(n - \log(q(N)) \geq s(n)$, we have that $\mathtt{fn}(Proj(x, y))$ is in $\mathfrak{C}$-MCSP$[s'(n)]$.

Next we establish the sampling condition. Note that for x uniformly chosen in $\{0,1\}^N$, the $q(N)$ strings $Proj_N(x, y), y \in [q(N)]$ are uniformly and independently distributed in $\{0,1\}^{\lfloor N/q(N) \rfloor}$. Applying Proposition 4, we have that with for any function $g : \{0,1\}^{\lfloor N/q(N) \rfloor} \to [-1, 1]$, for at least a $1 - o(1)$ fraction of x, $|E_{z \in U_{\lfloor N/q(N) \rfloor}} g(z) - E_{y \in U_{\log(q(N))}} g(Proj_N(x, y))| \leq \epsilon(N)$, where $\epsilon(N)$ is as in the statement of the theorem.

Now applying Theorem 6, the statement of the theorem follows immediately. □

Theorem 7 gives a connection from zero-error average-case hardness of MCSP to succinct pseudorandom distributions. It is perhaps surprising that zero-error average-case hardness in this context implies the existence of pseudorandom distributions that approximate any poly-size circuit well with respect to *two-sided error*. This is reminiscent of the situation with complexity-theoretic pseudorandom generators, where the generator is allowed to run in time exponential in the seed length. Known equivalences between worst-case hardness of exponential time and existence of complexity-theoretic PRGs [NW94, IW97] also imply an equivalence between complexity-theoretic PRGs and complexity-theoretic hitting set generators. Such an equivalence is unknown in the cryptographic setting, and is a major motivation for our work. Theorem 7 provides partial unconditional evidence for such an equivalence extending to the cryptographic setting. Indeed, combining Theorem 7 with Proposition 3 for appropriately chosen parameters, we get the following equivalence between succinct hitting sets and succinct pseudorandom distributions in the medium-error regime.

Corollary 1. *Let $\mathfrak{C}$ be a standard circuit class closed under projections. Moreover, let $s(n) \geq n$. For every large n, the following are equivalent:*

1. *For every $\delta < 1/2$, there are $\mathfrak{C}[s(O(n))]$-succinct hitting sets with error $N^{-\delta}$ against* SIZE[poly]
2. *For every $\delta < 1/2$, $\mathfrak{C}$-MCSP$[s(O(n))]$ is zero-error average-case infeasible for* SIZE[poly] *with success probability $N^{-\delta}$.

3. *For every $\delta < 1/2$, there are $\mathfrak{C}[s(O(n))]$-*succinct *pseudorandom distributions with error $N^{-\delta}$ against* SIZE[poly]

Proof. The second item follows from the first using the proof idea of Proposition 3. By using a direct implication from the first item to the third item of Proposition 3, we avoid the quasi-linear blow-up in the parameter, and preserve the error to within a neglibible additive term.

The third item follows from the second by using Theorem 7 with $\epsilon(N) = N^{\delta}$ and $k(N) = N^{2\delta+\gamma}$ for some γ such that $2\delta + \gamma < 1$. Note that for this parameter choice of k, $s'(N) = s(O(N))$, as desired.

The first item follows trivially from the third, since pseudorandom sets are also hitting sets. □

Remark. For MCSP, the equivalence above can be extended to the setting of negligible error by using samplers more sophisticated than the projection sampler used in the proof of Theorem 7, such as samplers corresponding to the Nisan-Wigderson generator. However, this extension comes at the cost of generality - it does not work for $\mathfrak{C}$-MCSP for arbitrary $\mathfrak{C}$ closed under projections.

Corollary 1 implies Theorem 1 by choosing s appropriately.

We next show how to apply Theorem 6 to SAT, by showing an analogous construction of pseudorandom satisfiable formulas based on Feige's hypothesis [Fei02] that random k-CNF formulas with a linear number of clauses are hard to refute. In our terminology, Feige's hypothesis states that random k-SAT is zero-error average-case hard. Here, as elsewhere in this paper, we consider non-uniform algorithms rather than uniform ones.

We would like to construct a pseudorandom distribution supported on satisfiable formulae based on Feige's hypothesis, similar to the construction of succinct pseudorandom distribution in Theorem 7 based on the zero-error average-case hardness of MCSP. However, we need to be careful in how we encode our formulas. With an arbitrary encoding, it might be the case that satisfiable formulas are easily distinguishable from random strings. This would the case, for example, if any valid encoding of any formula began with a 1. To solve this issue, we specify a natural information-theoretically efficient encoding of formulas as follows.

We consider k-CNF formulas on N variables, where k is a constant and N is a power of 2. We encode a formula ϕ with cN clauses by using $(n+1)kcN$ bits of information. Each clause is encoded by $(n+1)k$ bits, n bits to encode each variable, and 1 bit to encode whether the variable is positive or negative. These blocks of bits are simply concatenated together. Thus every string of $(n+1)kcN$ represents a unique k-CNF in a valid way.

Assuming the encoding of inputs to SAT above, we can show the following result.

Theorem 8. *Fix a positive integer k. If for every $c > 0$, k-*SAT *on cN clauses is zero-error infeasible for* SIZE[poly] *with success probability $\Omega(1)$, then for every $\epsilon > 0$ there are pseudorandom distributions with error $O(\epsilon)$ supported entirely on satisfiable formulas.*

Proof. We provide a sketch, since the proof is quite similar to that of Theorem 7.

Let $\epsilon > 0$ be any constant, and q be a power of two to be chosen large enough as a function of ϵ. Let Q be k-SAT with cN clauses for some c a large enough power of two, encoded as described above, and Q' be k-SAT with cN/q clauses. We show that that the q-projection sampler is a (Q, Q')-semantic sampler with error ϵ and accuracy ϵ.

For the semantic condition, we simply note that any sub-formula of a satisfiable formula is itself satisfiable, and therefore the q-projection sampler maps satisfiable formulas to satisfiable formulas

in our encoding, whenever q is a power of two. The sampling condition is easy to check by again using Proposition 4.

Now the theorem follows by applying Theorem 6.

□

Theorem 8 is a formal statement of Theorem 2.

We use the simple projection sampler in the proof of Theorem 8. We might be able to reduce the error for the pseudorandom sets by using more sophisticated samplers, however it is tricky to ensure that satisfiable formulas remain satisfiable when the sampler is applied.

4 On Universal Succinct PRGs

We first state our main conjecture formally.

Conjecture 1 (Universality Conjecture)**.** *For every $\epsilon < 1$, there is a universal $\mathsf{SIZE}(2^{\epsilon n})$-succinct PRG with non-trivial seed length.*

We need the standard fact that a PRG can be stretched by iteration.

Lemma 2. *For each $0 < \epsilon < 1$, if there is a PRG with non-trivial seed length, there is a PRG with seed length N^{ϵ}.*

We also need the construction of succinct PRGs from PRGs due to [GGM86].

Lemma 3. *[GGM86] For each $\epsilon > \delta < 1$, if there is a PRG with seed length N^{δ}, there is a $\mathsf{SIZE}[2^{\epsilon n}]$-succinct PRG*

Theorem 9. *Under Conjecture 1, the following are equivalent:*

1. *There is a one-way function secure against $\mathsf{SIZE}(\mathrm{poly})$.*
2. *There is an non-uniform one-way function secure against $\mathsf{SIZE}(\mathrm{poly})$.*
3. *There is a $\mathsf{SIZE}[2^{\delta n}]$-succinct HSGs secure against $\mathsf{SIZE}[\mathsf{poly}(N)]$, for some $0 < \delta < 1$.*
4. *There are $\mathsf{SIZE}[2^{\varepsilon n}]$-succinct HSGs secure against $\mathsf{SIZE}[\mathsf{poly}(N)]$, for any $0 < \varepsilon < 1$.*
5. *There is a PRG secure against $\mathsf{SIZE}[\mathsf{poly}(N)]$ with non-trivial seed length.*
6. *$\mathsf{MCSP}[2^{\delta n}]$ is hard on average against $\mathsf{SIZE}[\mathsf{poly}(N)]$ for some $0 < \delta < 1$.*
7. *$\mathsf{MCSP}[2^{\varepsilon n}]$ is hard on average against $\mathsf{SIZE}[\mathsf{poly}(N)]$ for every $0 < \varepsilon < 1$.*
8. *There are no $\mathsf{SIZE}(\mathrm{poly})$-natural proofs against $\mathsf{SIZE}(2^{\epsilon n})$ for any $\epsilon > 0$.*
9. *Polynomial-size circuits cannot be PAC-learned with membership queries over the uniform distribution in polynomial time.*

Proof. We establish the equivalence through a series of implications.

(3) implies (6): Follows from Proposition 3.

(6) implies (5): By setting $k(N)$ to be a sufficiently small power of N in Theorem 7, we get that there are $\mathsf{SIZE}[2^{\delta' n}]$-succinct pseudorandom distributions with error $1/N^{\gamma}$ for some $\gamma > 0$. Using Conjecture 1 we get that there is a PRG G with non-trivial seed length and error $1/N^{\gamma}$. The PRG G yields a weak one-way function, and by using the standard conversion from weak to strong one-way functions [Yao82, Gol01], and then applying the HILL construction [HILL99], we get a PRG with non-trivial seed length and negligible error.

(5) implies (4): By using Lemma 2 and Lemma 3, we get $\mathsf{SIZE}[2^{\varepsilon n}]$-succinct PRGs for any $\varepsilon > 0$. This trivially implies $\mathsf{SIZE}[2^{\varepsilon n}]$-succinct HSGs for any $\varepsilon > 0$.

(4) implies (7): Again follows from Proposition 3.

(7) implies (3): Trivial.

(7) equivalent to (8): Shown in [HS17].

(1) equivalent to (5): Shown in [HILL99].

(1) implies (2): Trivial.

(2) implies (8): Given any infinite set I of auxiliary inputs, by applying the constructions of [HILL99] and [GGM86] to auxiliary-input one-way functions, for any $\varepsilon > 0$, we get a $\mathsf{SIZE}[2^{\varepsilon n}]$-succinct distribution which can be distinguished from uniform by $\mathsf{SIZE}(\text{poly})$-natural proofs against $\mathsf{SIZE}(2^{\epsilon n})$. This implies that the auxiliartinput one-way function can be inverted on I, in contradiction to the assumption that for each poly-size adversary, there is some infinite set of inputs on which the function is hard to invert.

(6) equivalent to (8): Shown in [CIKK16].

□

Theorem 9 can easily be seen to imply all the items in Theorem 3.

The equivalences above give a fairly clean picture of connections between various fundamental notions, modulo Conjecture 1. We now discuss some of the more interesting individual connections in more detail.

The connection between auxiliary-input one-way functions and one-way functions has been an important question in cryptography since the former notion was introduced in [OW93]. The notion of auxiliary-input one-way functions has played an important role in the study of zero-knowledge [OW93, Vad06] and learning [ABX08]. In particular, it follows from the main result of [OW93] that under Conjecture 1, if there is a language with zero-knowledge proofs that is not in polynomial size, then one-way functions exist[1].

The notion of HSGs has not been much studied in cryptography, and this is perhaps because it is not obvious how to use HSGs in crypto applications. One of the issues is that it is not clear how to stretch a HSG, i.e., increase the gap between seed length and output length. As a consequence of the Conjecture, succinct HSGs are stretchable. It remains unclear whether the same is true for standard HSGs under plausible assumptions.

One of the main questions about MCSP is how robust its complexity is with respect to the size parameter s. Known results about the complexity of the problem are not very sensitive to the size parameter, but there are no known equivalences between the complexity of $\mathsf{MCSP}[s]$ and the complexity of $\mathsf{MCSP}[s']$ for s and s' that are different. As a consequence of the Conjecture, we

[1]Note that we are using security against non-uniform adversaries throughout our work.

get such an equivalence in the average-case setting. It would be interesting to try to establish this equivalence unconditionally.

The equivalence we get between hardness of learning and one-way functions (modulo the Conjecture) is the first such equivalence of which we are aware for a natural worst-case notion of learning. It is shown in [BFKL93] that hardness of PAC-learning *on average* implies the existence of one-way functions. The question of whether the hardness of PAC-learning (over any distribution, and without membership queries) implies the existence of one-way functions is posed in [ABX08]. It would be nice if we could use the Conjecture to resolve this question.

Perhaps the most interesting connection is the equivalence between the average-case hardness of MCSP over the uniform distribution and the existence of one-way functions. This has potential applications for the construction of a natural universal one-way function [Lev03]. One also wonders if under the Conjecture, there are other natural problems and distributions such that the average-case hardness of the problem under the distibution is equivalent to the existence of one-way functions. One would like a richer theory of reducibility between average-case problems, and equivalences of this sort might help.

5 MCSP and Circuit Lower Bounds against Weak Classes

Pseudorandomness is intimately connected to circuit lower bounds. Complexity-theoretic PRGs are equivalent to circuit lower bounds for E, and cryptographic PRGs are equivalent to one-way functions and hence imply circuit lower bounds for NP. In this section, we consider "weak" circuit classes, i.e., circuit classes $\mathfrak{C}$ for which there are natural proofs useful against $\mathfrak{C}[\mathrm{poly}]$. We show that for weak circuit classes $\mathfrak{C}$, succinct hitting sets imply lower bounds for P against $\mathfrak{C}[\mathrm{poly}]$. Using the connection between succinct hitting sets and average-case hardness for MCSP, we show that zero-error average-case lower bounds for $\mathsf{MCSP}[\mathrm{poly}]$ against $\mathfrak{C}[\mathrm{poly}]$ imply lower bounds for P against $\mathfrak{C}[\mathrm{poly}]$. This is surprising in that we establish a hardness consequence for P based on a hardness assumption about a problem not believed to be in P. Indeed, we show that our result is inherently non-black box if one-way functions exist.

Lemma 4. *Let $\mathfrak{C}$ be a weak circuit class closed under projections. If there is a constant k such that there are $\mathsf{SIZE}(n^k)$-succinct hitting sets against $\mathfrak{C}[\mathrm{poly}]$, then $\mathsf{P} \not\subseteq \mathfrak{C}[\mathrm{poly}]$*

Proof. Let $\mathfrak{C}$ be a weak circuit class closed under projections. Suppose there is a constant k such that there are $\mathsf{SIZE}(n^k)$-succinct hitting sets against $\mathfrak{C}[\mathrm{poly}]$. Using Proposition 2, we have that there is a $\mathsf{SIZE}(n^k)$-succinct HSG U with seed length $\mathsf{quasi}\text{-}n^k$ against $\mathfrak{C}[\mathrm{poly}]$. Now consider the following function $f(x,i)$, where i is of length n and x is of length $\mathsf{quasi}\text{-}n^k$. $f(x,i)$ is defined to be 1 iff the i'th bit of $G(x)$=1. Now, since Boolean circuits are a standard class, given seed x to G and index i of $G(x)$, the i'th bit of $G(x)$ is computable in time $\mathsf{quasi}\text{-}s(n)$, which is $\mathsf{quasi}\text{-}n^k$. Thus $f \in \mathsf{P}$.

Now we use a win-win analysis. If $\mathsf{P} \not\subseteq \mathfrak{C}[\mathrm{poly}]$, we are done. Hence we can assume that $\mathsf{P} \subseteq \mathfrak{C}[\mathrm{poly}]$. This implies that $f \in \mathfrak{C}[n^{k'}]$ for some constant k'. Since $\mathfrak{C}$ is closed under projections, it follows that every string in the range of U is the truth table of a function with $\mathfrak{C}$-circuits of size at most $n^{k'}$.

Since $\mathfrak{C}$ is weak, there are natural proofs useful against $\mathfrak{C}[\mathrm{poly}]$. This implies that there is a set $A \subseteq \mathsf{SIZE}(\mathrm{poly})$ of density $1/2$ such that no string y for which $\mathtt{fn}(y)$ is in $\mathfrak{C}[n^{k'}]$ belongs to A .

Now again using the assumption that $\mathsf{P} \subseteq \mathsf{C}[\text{poly}]$, we have that $\mathsf{SIZEpoly} \subseteq \mathfrak{C}[\text{poly}]$, and hence $A \in \mathfrak{C}[\text{poly}]$. But now A is a set in $\mathfrak{C}[\text{poly}]$ of density $1/2$ which does not intersect the range of U non-trivially, contradicting the assumption that U is a hitting set generator against $\mathfrak{C}[\text{poly}]$. □

The smallest complexity class within which weakness of $\mathfrak{C}$ is known to imply a super-polynomial lower bound against $\mathfrak{C}$ is ZPEXP [OS17]. Lemma 4 shows that if additionally there are succinct hitting sets against $\mathfrak{C}[\text{poly}]$, the function for which we get a lower bound is much more explicit - it is in P.

Next we combine Lemma 4 with Proposition 3 to get a surprising implication.

Theorem 10. *Let $\mathfrak{C}$ be a weak circuit class closed under projections. If there is a constant k such that $\mathsf{MCSP}[n^k]$ is zero-error average-case hard against $\mathfrak{C}[\text{poly}]$, then P is not contained in $\mathfrak{C}[\text{poly}]$.*

Proof. if there is a constant k such that $\mathsf{MCSP}[n^k]$ is zero-error average-case hard against $\mathfrak{C}[\text{poly}]$, then by Proposition 3, we get that there are $\mathsf{SIZE}(n^k)$-succinct hitting sets against $\mathfrak{C}[\text{poly}]$. Now applying Lemma 4, we get the desired consequence. □

Theorem 10 is essentially a restatement of Theorem 4.

Theorem 10 is a partial converse to the following corollary to Theorem 2 from [OS17].

Theorem 11. *Let $\mathfrak{C}[\text{poly}]$ be any circuit class closed under composition with poly-size AC^0. If there is a language in P that cannot be approximated on $1/2 + 1/\operatorname{poly}(n)$ fraction of inputs by $\mathfrak{C}$-circuits of polynomial size, then $\mathsf{MCSP}[2^{n/2}]$ is zero-error average-case hard against $\mathfrak{C}[\text{poly}]$.*

Note that there are examples of weak circuit classes such as AC^0 and $\mathsf{AC}^0[p]$ which satisfy the condition in Theorem 11.

Theorem 10 is a rare example of a non-black-box reduction between two problems - the reduction does not work when the circuit class against which we are arguing is given access to an oracle. Indeed, under standard crypto assumptions, there is no black-box reduction from $\mathsf{MCSP}[n^k]$ to P, for k chosen large enough.

Theorem 12. *Let $\mathfrak{C}$ be any Boolean circuit class which contains the projection functions. If there are one-way functions of exponential hardness, there is a constant k for which there is no black-box reduction from zero-error average-case hardness of $\mathsf{MCSP}[n^k]$ against $\mathfrak{C}[\text{poly}]$ to $\mathsf{P} \not\subseteq \mathfrak{C}[\text{poly}]$.*

Proof. If there were such a black-box reduction, then for each oracle A, $\mathsf{P} \subseteq \mathfrak{C}^A[\text{poly}]$ would imply $\mathsf{MCSP}[n^k]$ is zero-error easy on average for $\mathfrak{C}^A[\text{poly}]$. But now consider an oracle A that is complete for P under projections. The antecedent trivially holds for such an oracle, but if the consequent held, we would have $\mathsf{MCSP}[n^k]$ is zero-error easy on average for $\mathsf{SIZE}[\text{poly}]$ for any k, which by the "natural proofs" argument of Razborov and Rudich [RR97] would invert any one-way function in sub-exponential time. □

6 An approximation to average-case reduction for AveMCSP

Finally, we observe that a certain search-to-decision reduction for a variant of MCSP given in recent work [CIKK17] actually yields a non black-box approximation to average-case reduction. We will use the following variant of the Nisan-Wigderson generator [CIKK17], for which the output of the generator has small average-case circuit complexity for most seeds when the function on which the generator is based has small average-case circuit complexity.

Theorem 13. *[NW94, CIKK17] There is a fixed constant $d > 1$ such that for any constant $c > d$ there is a sequence of functions $\{G_m : \{0,1\}^{m^c} \times \{0,1\}^{O(\log(m))} \to \{0,1\}^m\}$ computable in polynomial time such that*

1. *Given $y \in \{0,1\}^{m^c}$ and for any $t = \log(m)^{\omega(1)}$, if $\mathtt{fn}(y)$ can be computed correctly on 0.9 fraction of inputs by circuits of size t, then with probability $1 - o(1)$ over choice of the seed $z \in \{0,1\}^{O(\log(m))}$, $\mathtt{fn}(G_m(y,z))$ can be computed correctly on 0.9 fraction of inputs by circuits of size t^2.*

2. *Given $y \in \{0,1\}^{m^c}$, if $\mathtt{fn}(y)$ cannot be computed correctly on 0.9 fraction of inputs by circuits of size m^d, then $G_m(y,\cdot)$ is a PRG with error $1/m$ against $\mathsf{SIZE}(m)$.*

Theorem 14. *For any $\delta > 0$ and $k > 0$, there is $\epsilon > 0$ such that if $\mathsf{AveMCSP}[2^{n/2}]$ is zero-error easy on average for circuits of size N^k, then $\mathsf{AveMCSP}[2^{\epsilon n}, 2^{\delta n}]$ has polynomial-size circuits.*

Proof. Suppose there is $k > 0$ such that $\mathsf{AveMCSP}[2^{n/2}]$ is zero-error easy on average for circuits of size N^k. Without loss of generality, this means that there exists a set $A \subseteq \{0,1\}^*$ with circuits of size N^k such that A has density at least 0.01 and A does not contain any strings y for which $\mathtt{fn}(y)$ has circuits of size at most $2^{n/2}$.

We show how to solve $\mathsf{AveMCSP}[2^{\epsilon n}, 2^{\delta n}]$ in polynomial size, for some ϵ fixed during the argument which depends on δ and k. Given an input z of length N, our non-uniform polynomial time algorithm runs as follows. First it finds m such that $m^c = N$, where d is the fixed constant given by Theorem 13, and c is chosen to be $2kd/\delta$ in the construction of Theorem 13. It then computes $G_{m^k}(z,x)$ for each $x \in \{0,1\}^{O(\log(m))}$ in polynomial time, truncating each output string to its first m bits. It calculates the fraction η of these strings that belong to A. If $\eta \geq 0.005$, it rejects, else it accepts. This algorithm can clearly be implemented by circuits of polynomial size.

In the following argument, we use 'average-case circuit complexity' to mean the size of the smallest circuit computing the function correctly on a 0.9 fraction of inputs.

We argue that when $\mathtt{fn}(z)$ has average-case circuit complexity at least N^δ, the algorithm rejects, and that when $\mathtt{fn}(z)$ has average-case circuit complexity at most N^ϵ for appropriately chosen ϵ, the algorithm accepts. For the first item, note that when $\mathtt{fn}(z)$ has average-case circuit complexity at least N^δ, by the second part of Theorem 13, $G_{m^k}(z,\cdot)$ is a $1/m^k$-error PRG against circuits of size m^k. Since A has circuits of size m^k and also has density at least 0.01, this means that $G_{m^k}(z,\cdot)$ fools A, and hence the fraction η calculated by the algorithm in this case is at least $0.01 - o(1)$, which is at least 0.005 for large enough N.

Now if $\mathtt{fn}(z)$ has average-case circuit complexity at most N^ϵ, where we are yet to fixe ϵ, we have from the first part of Theorem of Theorem 13, for at least a $1 - o(1)$ fraction of seeds x, $G_{m^k}(z,x)$ has average-case circuit complexity at most $N^{2\epsilon}$. Note that we are truncating the output of the PRG to the first m bits, hence the function represented by the output has length $m = N^{\delta/2kd}$. If ϵ is chosen so that $\epsilon < \delta/8kd$, we have that with probability $1 - o(1)$ over choice of seed x, the truncated output of the PRG has average-case circuit complexity at most $2^{n/2}$. Since A does not contain any strings y for which $\mathtt{fn}(y)$ has circuits of size at most $2^{n/2}$, we have that the fraction η is calculated to be $o(1) < 0.005$ for large enough N. □

Theorem 14 is a formal version of Theorem 5.

Note that the approximation to average-case reduction in Theorem 14 has a very unusual feature - the gap in the approximation version depends on the *complexity* of the average-case algorithm.

In particular, this reduction is not black-box, meaning that the reduction does not extend to the case where the average-case algorithm uses an oracle.

7 Future Work

The Universality Conjecture opens up several directions for future work. The first is to derive further interesting implications from the Conjecture. For example, is the Learning is Hard assumption of [ABX08] equivalent to the non-existence of natural proofs under the Conjecture?

The second is to establish the connections and equivalences we seek unconditionally, or under weaker forms of the Conjecture. Our unconditional results such as Theorem 1 are a step in this direction.

The third is to develop approaches to proving the Conjecture. A useful step here would be to come up with an explicit candidate for universal succinct PRGs.

A fourth direction is to explore consequences of the Conjecture being false. As mentioned before, the Conjecture holds if one-way functions exist, and therefore the failure of the Conjecture would imply the non-existence of one-way functions. But perhaps even stronger consequences follow from the failure of the Conjecture, lending further support to it?

More generally, there are other pathways to connecting average-case hardness of MCSP to the existence of one-way functions. In this work, we have explored average-case hardness in the zero-error sense. But perhaps it would be easier to use average-case hardness of MCSP in the bounded-error sense to construct one-way functions. One question here is to come up with a natural distribution under which MCSP is hard on average in the bounded-error sense - clearly the uniform distribution is not a valid candidate.

8 Acknowledgments

Discussions with Shuichi Hirahara and Igor Carboni Oliveira were very helpful at an early stage of this research. Thanks to Igor for his detailed comments on an early draft of this work. Thanks to Shuichi for telling me about his independent work [Hir18] and for alerting me to the relevance of auxiliary-input one-way functions. Thanks also to Andrej Bogdanov and Hoeteck Wee for e-mail correspondence about cryptographic hitting set generators.

This work was supported in part by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2014)/ERC Grant Agreement No. 615075.

References

[ABX08] Benny Applebaum, Boaz Barak, and David Xiao. On basing lower-bounds for learning on worst-case assumptions. In *Proceedings of 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 211–220, 2008.

[AD14] Eric Allender and Bireswar Das. Zero knowledge and circuit minimization. In *Symposium on Mathematical Foundations of Computer Science* (MFCS), pages 25–32, 2014.

[AGGM06] Adi Akavia, Oded Goldreich, Shafi Goldwasser, and Dana Moshkovitz. On basing one-way functions on np-hardness. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing, Seattle, WA, USA, May 21-23, 2006*, pages 701–710, 2006.

[AGM15] Eric Allender, Joshua A. Grochow, and Cristopher Moore. Graph isomorphism and circuit size. *CoRR*, abs/1511.08189, 2015.

[AH17] Eric Allender and Shuichi Hirahara. New insights on the (non-)hardness of circuit minimization and related problems. In *International Symposium on Mathematical Foundations of Computer Science* (MFCS), pages 54:1–54:14, 2017.

[AHK15] Eric Allender, Dhiraj Holden, and Valentine Kabanets. The minimum oracle circuit size problem. In *International Symposium on Theoretical Aspects of Computer Science* (STACS), pages 21–33, 2015.

[AHM+08] Eric Allender, Lisa Hellerstein, Paul McCabe, Toniann Pitassi, and Michael E. Saks. Minimizing disjunctive normal form formulas and AC0 circuits given a truth table. *SIAM J. Comput.*, 38(1):63–84, 2008.

[AKRR11] Eric Allender, Michal Koucký, Detlef Ronneburger, and Sambuddha Roy. The pervasive reach of resource-bounded Kolmogorov complexity in computational complexity theory. *J. Comput. Syst. Sci.*, 77(1):14–40, 2011.

[All17] Eric Allender. The complexity of complexity. In *Computability and Complexity - Essays Dedicated to Rodney G. Downey on the Occasion of His 60th Birthday*, pages 79–94, 2017.

[Alt94] Ingo Althofer. On sparse approximations to randomized strategies and convex combinations. *Linear Algebra and its Applications*, 199:339–355, 1994.

[BFKL93] Avrim Blum, Merrick Furst, Michael Kearns, and Richard Lipton. Cryptographic primitives based on hard learning problems. In *Proceedings of 13th Annual International Cryptology Conference*, pages 278–291, 1993.

[BT06] Andrej Bogdanov and Luca Trevisan. On worst-case to average-case reductions for NP problems. *SIAM J. Comput.*, 36(4):1119–1159, 2006.

[CIKK16] Marco L. Carmosino, Russell Impagliazzo, Valentine Kabanets, and Antonina Kolokolova. Learning algorithms from natural proofs. In *Conference on Computational Complexity* (CCC), pages 10:1–10:24, 2016.

[CIKK17] Marco L. Carmosino, Russell Impagliazzo, Valentine Kabanets, and Antonina Kolokolova. Agnostic learning from tolerant natural proofs. In *Approximation, Randomization, and Combinatorial Optimization*, pages 35:1–35:19, 2017.

[DLS14] Amit Daniely, Nati Linial, and Shai Shalev-Shwartz. From average case complexity to improper learning complexity. In *Proceedings of 46th Annual Symposium on Theory of Computing*, pages 441–448, 2014.

[Fei02] Uriel Feige. Relations between average case complexity and approximation complexity. In *Proceedings on 34th Annual ACM Symposium on Theory of Computing*, pages 534–543, 2002.

[FF93] Joan Feigenbaum and Lance Fortnow. Random-self-reducibility of complete sets. *SIAM J. Comput.*, 22(5):994–1005, 1993.

[FSV17] Michael A. Forbes, Amir Shpilka, and Ben Lee Volk. Succinct hitting sets and barriers to proving algebraic circuits lower bounds. In *Proceedings of the 49th Annual ACM Symposium on Theory of Computing*, pages 653–664, 2017.

[GGM86] Oded Goldreich, Shafi Goldwasser, and Silvio Micali. How to construct random functions. *J. ACM*, 33(4):792–807, 1986.

[GKSS17] Joshua A. Grochow, Mrinal Kumar, Michael E. Saks, and Shubhangi Saraf. Towards an algebraic natural proofs barrier via polynomial identity testing. *Electronic Colloquium on Computational Complexity (ECCC)*, 24:9, 2017.

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

[HILL99] Johan Håstad, Russell Impagliazzo, Leonid 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 *59th IEEE Annual Symposium on Foundations of Computer Science*, pages 247–258, 2018.

[Hoe63] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[HP15] John M. Hitchcock and Aduri Pavan. On the NP-completeness of the minimum circuit size problem. In *Conference on Foundation of Software Technology and Theoretical Computer Science* (FSTTCS), pages 236–245, 2015.

[HS17] Shuichi Hirahara and Rahul Santhanam. On the average-case complexity of MCSP and its variants. In *Computational Complexity Conference* (CCC), pages 7:1–7:20, 2017.

[HW16] Shuichi Hirahara and Osamu Watanabe. Limits of minimum circuit size problem as oracle. In *Conference on Computational Complexity* (CCC), pages 18:1–18:20, 2016.

[IKW01] Russell Impagliazzo, Valentine Kabanets, and Avi Wigderson. In search of an easy witness: Exponential time vs. probabilistic polynomial time. In *Proceedings of the 16th Annual IEEE Conference on Computational Complexity*, pages 2–12, 2001.

[IL90] Russell Impagliazzo and Leonid A. Levin. No better ways to generate hard NP instances than picking uniformly at random. In *31st Annual Symposium on Foundations of Computer Science*, 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.

[IW97] Russell Impagliazzo and Avi Wigderson. P = BPP if E requires exponential circuits: Derandomizing the XOR lemma. In *Proceedings of the 29th Annual ACM Symposium on the Theory of Computing (STOC)*, pages 220–229, 1997.

[Kab00] Valentine Kabanets. Easiness assumptions and hardness tests: Trading time for zero error. In *Proceedings of the 15th Annual IEEE Conference on Computational Complexity*, pages 150–157, 2000.

[KC00] Valentine Kabanets and Jin-yi Cai. Circuit minimization problem. In *Symposium on Theory of Computing* (STOC), pages 73–79, 2000.

[KL14] Jonathan Katz and Yehuda Lindell. *Introduction to Modern Cryptography, Second Edition*. CRC Press, 2014.

[KV89] Michael Kearns and Leslie Valiant. Cryptographic limitations on learning boolean formulae and finite automata. In *Proceedings of the 21st Annual ACM Symposium on Theory of Computing*, pages 433–444, 1989.

[Lev84] Leonid Levin. Randomness conservation inequalities; information and independence in mathematical theories. *Information and Control*, 61(1):15–37, 1984.

[Lev03] Leonid Levin. The tale of one-way functions. *Problems of Information Transmission*, 39(1):92–103, 2003.

[LY94] Richard J. Lipton and Neal E. Young. Simple strategies for large zero-sum games with applications to complexity theory. In *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing, 23-25 May 1994, Montréal, Québec, Canada*, pages 734–740, 1994.

[MW15] Cody Murray and Ryan Williams. On the (non) NP-hardness of computing circuit complexity. In *Conference on Computational Complexity* (CCC), pages 365–380, 2015.

[NW94] Noam Nisan and Avi Wigderson. Hardness vs randomness. *J. Comput. Syst. Sci.*, 49(2):149–167, 1994.

[OS17] Igor Carboni Oliveira and Rahul Santhanam. Conspiracies between learning algorithms, circuit lower bounds, and pseudorandomness. In *Computational Complexity Conference* (CCC), pages 18:1–18:49, 2017.

[OW93] Rafail Ostrovsky and Avi Wigderson. One-way fuctions are essential for non-trivial zero-knowledge. In *Proceedings of Second Israel Symposium on Theory of Computing Systems*, pages 3–17, 1993.

[PW90] Leonard Pitt and Manfred Warmuth. Prediction-preserving reducibility. *Journal of Computer and System Sciences*, 41(3):430–467, 1990.

[RR97] Alexander A. Razborov and Steven Rudich. Natural proofs. *J. Comput. Syst. Sci.*, 55(1):24–35, 1997.

[Rud97] Steven Rudich. Super-bits, demi-bits, and np/qpoly-natural proofs. In *Randomization and Approximation Techniques in Computer Science*, pages 85–93, 1997.

[Tra84] Boris A. Trakhtenbrot. A survey of Russian approaches to perebor (brute-force searches) algorithms. *IEEE Annals of the History of Computing*, 6(4):384–400, 1984.

[Tre01] Luca Trevisan. Extractors and pseudorandom generators. *J. ACM*, 48(4):860–879, 2001.

[Vad06] Salil Vadhan. An unconditional study of computational zero knowledge. *SIAM Journal on Computing*, 36(4):1160–1214, 2006.

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

[Vad17] Salil Vadhan. On learning vs. refutation. In *Proceedings of the 30th Conference on Learning Theory*, pages 1835–1848, 2017.

[Val84] Leslie Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

[VZ13] Salil Vadhan and Colin Zheng. A uniform min-max theorem with applications in cryptography. In *33rd Annual Cryptology Conference,*, pages 93–110, 2013.

[Wil16] R. Ryan Williams. Natural proofs versus derandomization. *SIAM J. Comput.*, 45(2):497–529, 2016.

[Yao82] Andrew Yao. Theory and applications of trapdoor functions (extended abstract). In *Proceedings of 23rd Annual Symposium on Foundations of Computer Science*, pages 80–91, 1982.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il