A PCP Characterization of AM

Andrew Drucker [*]

MIT

February 18, 2010

Abstract

We introduce a 2-round stochastic constraint-satisfaction problem, and show that its approximation version is complete for (the promise version of) the complexity class AM. This gives a 'PCP characterization' of AM analogous to the PCP Theorem for NP. Similar characterizations have been given for higher levels of the Polynomial Hierarchy, and for PSPACE; however, we suggest that the result for AM might be of particular significance for attempts to derandomize this class.

To test this notion, we pose some 'Randomized Optimization Hypotheses' related to our stochastic CSPs that (in light of our result) would imply collapse results for AM. Unfortunately, the hypotheses appear over-strong, and we present evidence against them. In the process we show that, if some language in NP is hard-on-average against circuits of size $2^{\Omega(n)}$, then there exist hard-on-average optimization problems of a particularly elegant form.

All our proofs use a powerful form of PCPs known as Probabilistically Checkable Proofs of Proximity, and demonstrate their versatility. We also use known results on randomness-efficient soundness- and hardness-amplification. In particular, we make essential use of the Impagliazzo-Wigderson generator; our analysis relies on a recent Chernoff-type theorem for expander walks.

1 Introduction

1.1 Background: PCPs and complexity classes

A *Constraint Satisfaction Problem (CSP)* is a collection $\psi(x)$ of Boolean-valued constraints over variables on a bounded-size alphabet Σ. A CSP in which each constraint depends on at most k variables is called a *k-CSP*. A natural computational task is to determine the maximum fraction of constraints that can be satisfied by any assignment. Cook's Theorem [Pap94] states that this problem is NP-complete, and the landmark PCP Theorem of Arora et al. [ALM+98] implies that, for a sufficiently small constant $\varepsilon > 0$, it is NP-hard even to output an estimate that is within ε of this maximum fraction (where in both results we may take $k = 3, \Sigma = \{0, 1\}$).

Given the importance of the PCP Theorem for complexity theory, researchers have looked for analogues of the result for complexity classes other than NP. The PCP Theorem can be seen as stating that it is NP-hard to determine within ε the value of a 1-player 'solitaire' game defined by a 3-CSP. It is equally possible to study games played on a k-CSP in which 2 players alternate in setting values to designated blocks of variables, with one player trying to maximize the fraction

---

of satisfied clauses and the other trying to minimize this fraction. These games were explored in several works. Ko and Lin [KL94] showed that approximating the value of such a game is hard for the j-th level of the Polynomial Hierarchy, if the game lasts for j moves. In more recent work of Haviv et al. [HRTS07] this result was shown to hold even if each variable is allowed to appear in at most a constant number of constraints.

If the game is allowed to last polynomially many rounds, the approximation problem becomes PSPACE-hard as shown by Condon et al. [CFLS95]. The same authors showed the approximation problem for poly(n) rounds is also PSPACE-hard if a maximizing player plays against a *random* player [CFLS97] (where the game's value now is the *expected* number of satisfied clauses under optimum play by the maximizer). Moreover, all of the hardness-of-approximation results mentioned so far are in fact completeness results for the corresponding promise classes, so they can be viewed as giving 'PCP characterizations' of NP, PSPACE, and the Polynomial Hierarchy.

One class that did not receive a PCP characterization based on CSP games was the Arthur-Merlin class AM. In fact, there are few known natural complete problems for AM (technically, for its promise version, prAM; we don't know if AM, a semantic class, has *any* complete problems. See Sec. 2.3 for the definition of prAM.). To this author's knowledge there is only one *approximation* problem previously known to be prAM-complete: Mossel and Umans [MU02] give a prAM-completeness result for approximating the VC dimension of set systems. This striking result does not fall within the framework of CSP games given above.

1.2 Our results

In this paper we present a PCP characterization of prAM. We consider 'stochastic' 2-CSPs $\psi(r, z)$, where r is a collection of Boolean variables and z a collection of variables over an alphabet Σ. Let $\mathrm{Val}_\psi(r, z)$ be the fraction of constraints of ψ satisfied by (r, z). In Section 4 we prove:

Theorem 1. *There is a finite alphabet Σ and a constant $\varepsilon > 0$, such that it is* prAM-*complete to distinguish between the following two sets of 2-CSPs:*

$$\Pi_{YES} = \{\psi : \text{ for all } r \text{ there exists } z \text{ such that } \mathrm{Val}_\psi(r, z) = 1\};$$

$$\Pi_{NO} = \{\psi : \text{ with probability } 1 - \exp(-\Omega(|r|)) \text{ over random } r, \text{ Max}_z[\mathrm{Val}_\psi(r, z)] < 1 - \varepsilon\}.$$

In particular, this implies that $\varepsilon/2$-approximating the value of the 2-round game associated with $\psi(r, z)$ (where the first player plays randomly) is AM-hard.

AM is a class for which we feel such a PCP characterization might be especially important. There is compelling evidence that AM = NP, or at least that significant derandomization of AM is possible (see [SU07] for an overview of this line of research). One approach to to try and derandomize AM is to directly attack the 'easiest' AM-hard problems, and a problem like the one provided by Theorem 1 seems like a plausible candidate.

How might such an attack proceed? We make a concrete suggestion in the form of two 'Randomized Oracle Hypotheses'. In what follows $\psi(r, z)$ is a 2-CSP over ℓ Boolean variables (r) and m variables (z) over a finite alphabet Σ.

Hypothesis A (Randomized Optimization Hypothesis for P/poly). *Fix any $\delta > 0$. For every 2-CSP $\psi(r, z)$, there exists a circuit $C_\psi(r) : \{0, 1\}^\ell \to \Sigma^m$ of size $O(\mathrm{poly}(|\psi|))$, such that with probability at least $1/\mathrm{poly}(\ell)$ over a random $r \in \{0, 1\}^\ell$, we have*

$$\mathrm{Val}_\psi(r, C_\psi(r)) \geq \mathrm{Max}_z[\mathrm{Val}_\psi(r, z)] - \delta.$$

In a nutshell, this hypothesis suggests that it is easy to approximately-optimize over z for a random choice of r, if we allow our optimizer to depend nonuniformly on the 2-CSP ψ. (Such nonuniformity is clearly necessary, in light of the PCP Theorem for NP.) This hypothesis, if true, would yield a collapse result for AM. In Section 5 we prove the following claim by a straightforward application of Theorem 1:

Claim 2. *Hypothesis A implies* AM = MA.

A strengthened hypothesis could have the stronger implication that AM = NP. Consider the following:

Hypothesis B (Randomized Optimization Hypothesis for NC^0). *For any $\delta > 0$, there is an integer $t = t(\delta) > 0$ such that the following holds. For every 2-CSP $\psi(r, z)$, there exists a function $F_\psi(r) : \{0,1\}^\ell \to \Sigma^m$, where each output coordinate of F_ψ depends on at most t bits of r, and such that with probability at least $1 - \delta$ over r,*

$$\mathrm{Val}_\psi(r, F_\psi(r)) \geq \mathrm{Max}_z[\mathrm{Val}_\psi(r, z)] - \delta.$$

In Section 5 we prove:

Claim 3. *Hypothesis B implies* AM = NP.

Given the potential consequences of these hypotheses, what chance do they have of being true? Unfortunately, it seems that each is unlikely. In Section 6 we prove two results, each to the effect that, if NP decision problems are hard on average for exponential-size circuits, then both hypotheses fail in a strong way. We state these results next. A language L is called $p(n)$-*hard for size* $s(n)$ if for every circuit C of size $s(n)$, $\mathrm{Pr}_{x \in \{0,1\}^n}[C(x) = L(x)] \leq p(n)$.

Theorem 4. *Suppose there exists a $\gamma_1 > 0$ and an $L \in \mathsf{NP} \cap \mathsf{coNP}$ that is $(1 - 1/\mathrm{poly}(n))$-hard for size $2^{\gamma_1 n}$. Then there exists $c, \gamma_2, \theta > 0$ and a polynomial-time constructible family $\{\psi_n(r, w)\}_{n>0}$ of 2-CSPs (with $|r| = cn, |w| = d(n) = O(\mathrm{poly}(n))$), such that:*
  (4.i) for all r, there exists a w such that $\mathrm{Val}_{\psi_n}(r, w) = 1$;
  (4.ii) for all n, if $C : \{0,1\}^{cn} \to \{0,1\}^{d(n)}$ is a circuit of size at most $2^{\gamma_2 n}$, then

$$\Pr_r[\mathrm{Val}_{\psi_n}(r, C(r)) > 1 - \theta] \leq \exp\{-\Omega(n)\}.$$

Theorem 5. *There is an $\varepsilon_0 > 0$ such that the following holds. Suppose there exists a $\gamma_1 > 0$ and an $L \in \mathsf{NP}$ that is $(1/2 + \varepsilon_0)$-hard for size $2^{\gamma_1 n}$. Then there exists a $c > 0$, a polynomial-time constructible family $\{\psi_n(r, w)\}_{n>0}$ of 2-CSPs (with $|r| = cn, |w| = d(n) = O(\mathrm{poly}(n))$), and $\gamma_2, \theta > 0$, such that:*
  (5.i) With probability $\geq 1 - \exp\{-\Omega(n)\}$ over r, there exists w with $\mathrm{Val}_{\psi_n}(r, w) = 1$;
  (5.ii) If $C : \{0,1\}^{c(n)} \to \{0,1\}^{d(n)}$ is any circuit of size at most $2^{\gamma_2 n}$, then

$$\Pr_r[\mathrm{Val}_{\psi_n}(r, C(r)) > 1 - \theta] \leq \exp\{-\Omega(n)\}.$$

We note that the hypothesis in Theorem 5 is implied by the hypothesis that there exists a *balanced* function $L \in \mathsf{NP}$ that is $(1 - 1/\operatorname{poly}(n))$-hard for some size $s(n) = 2^{\Omega(n)}$; this follows from a result of O'Donnell [O'D02] (see also Healy et al. [HVV06], where the needed form of O'Donnell's result is made explicit and proved in a stronger form).

Theorems 4 and 5 both say that if NP (or $\mathsf{NP} \cap \mathsf{coNP}$) has sufficiently hard problems, then this hardness can be 'concentrated' into a kind of 'inapproximability-on-average' result for an optimization problem associated with a single, uniform family of (stochastic) CSPs. Note that the two results offer a tradeoff: Theorem 5 gives a slightly weaker conclusion from a presumably likelier hardness assumption. The assumptions in the above results are strong but, we feel, plausible. But at the very least, these results suggest that the approach we suggested to showing new upper-bounds on the power of AM must be modified to have a reasonable chance of succeeding.

We feel, however, that the Random Optimization Hypotheses are worthy of study in their own right, even if they turn out to be false; we pose some concrete questions about them at the end of the paper. We also feel that Theorems 4 and 5 are interesting for the study of average-case hardness in NP, and that the CSP families they produce might have further applications in complexity theory.

1.3 Our methods

All of our three main results–Theorems 1, 4, and 5–are essentially hardness results for computational tasks associated with 2-CSPs. In each case the reduction with which we prove our result uses a powerful type of PCP known as Probabilistically Checkable Proofs of Proximity (PCPPs). PCPPs were introduced independently by Ben-Sasson et al. [BSGH+06] and by Dinur and Reingold [DR06], and the PCPPs we use were developed by Dinur [Din07] (in [DR06, Din07] PCPPs are referred to as 'Assignment Testers'). In Section 3, we derive a variant form of PCPPs (Lemma 9) that is more useful for our purposes.

Lemma 9 gives a general reduction (similar to past uses of PCPPs, e.g., in [Din07]) in which we start with a two-argument circuit $Q(r, w)$ and efficiently produce a 2-CSP $\psi(r, z)$. The basic hope for our reduction is as follows: first, for any r, if the restricted circuit $Q(r, \cdot)$ is satisfiable (i.e., there exists w such that $Q(r, w) = 1$), then the restricted 2-CSP $\psi(r, \cdot)$ should be satisfiable as well. Second, if $Q(r, \cdot)$ is unsatisfiable, then any assignment to $\psi(r, \cdot)$ should violate an $\Omega(1)$-fraction of the constraints in ψ. Unfortunately, this second requirement is too strong and cannot be met. What we *can* guarantee is that if r is 'far' in Hamming distance from any r' for which $Q(r', \cdot)$ is satisfiable, then for any z, (r, z) violates an $\Omega(1)$-fraction of constraints of ψ.

How does this reduction help us prove the prAM-hardness result in Theorem 1? Any instance x of a promise problem $\Pi = (\Pi_{YES}, \Pi_{NO})$ defines a predicate $Q(r, w)$ computed by a poly-size circuit. If $x \in \Pi_{YES}$ then for all r, $Q(r, \cdot)$ is satisfiable; while if $x \in \Pi_{NO}$ then for a 2/3 fraction of r, $Q(r, \cdot)$ is unsatisfiable. In order to apply our reduction, we need a stronger condition in the second case: a random choice of r should be far from any r' for which $Q(r', \cdot)$ is satisfiable. In other words, we need an extremely low error probability in our underlying Arthur-Merlin protocol. This cannot be achieved by straightforward parallel repetition, but it is provided by a theorem of Bellare et al. [BGG93] which gives a randomness-efficient soundness-amplification for AM. Interestingly, Mossel and Umans [MU02] also used such amplification for their AM-hardness-of-approximation result on VC dimension, but for rather different reasons (unrelated to PCPs).

Our prAM-hardness proof is, we feel, more straightforward than the existing proofs of the analogous results for PSPACE and the Polynomial Hierarchy, modulo our use of sophisticated tools (PCPPs and efficient soundness-amplification) which we apply in a 'black-box' fashion. Of course

we do not rule out that our result could be also proved more directly by adapting ideas from the earlier papers (which use some of the same property-testing ideas that have gone into constructions of PCPPs). But we feel that PCPPs in particular, which have already found applications in PCP construction, coding theory, and property testing (see [BSHLM08] for an overview), are a versatile tool which could be of more widespread use in complexity theory. In a very recent example of their utility, Williams [Wil] applied PCPPs to the study of circuit lower bounds.

Next we discuss our methods in Theorems 4 and 5. Our transformation in Lemma 9 from the circuit Q to the 2-CSP ψ has a further useful property: we can reduce the problem of *finding* satisfying assignments to $Q(r, \cdot)$, to the problem of finding nearly-optimal assignments to $\psi(r, \cdot)$. Roughly speaking, we show the following. Suppose there is an algorithm $P(r)$ producing an assignment z, such that with some probability p over r, $(r, P(r))$ satisfies 'almost all' of the constraints of ψ; then there is a second algorithm $\tilde{P}(r)$ such that $Q(r, \tilde{P}(r)) = 1$ with probability $p' \geq 2^{-\varepsilon|r|}p$ (where $\varepsilon > 0$ can be chosen arbitrarily small). This property of the reduction is somewhat more novel, although the techniques we use (involving error-correcting codes) still follow previous works.

To apply our reduction, we use the hardness assumptions in Theorems 4 and 5 to produce predicates $Q(r, w)$ such that $Q(r, \cdot)$ is satisfiable with high probability, while any 'small' witness-producing circuit C fails to solve the search problem associated with Q: that is, $Q(r, C(r)) = 0$ with high probability. Because of the exponential loss factor $2^{-\varepsilon|r|}$ in our reduction, we need the search problem associated with Q to be *extremely* hard: we need every 'small' circuit C to succeed with probability at most $\exp\{-\Omega(|r|)\}$ over r in achieving $Q(r, C(r)) = 1$.

To produce such extremely hard search problems from a more 'mild' hardness assumption, we use existing hardness-amplification techniques. In particular, we use the well-known Impagliazzo-Wigderson generator [IW97]. This generator, on input parameter n, takes a seed r of length $O(n)$, and produces n 'pseudorandom' outputs $g_1, \ldots, g_n$ each of length n. The generator has the property that if language L is mildly hard for sufficiently small (but exponential-size) circuits, then any sufficiently smaller circuit has success probability $\leq \exp\{-\Omega(n)\}$ in correctly guessing the n-bit string $(L(g_1), \ldots, L(g_n))$. Then, if our hard language L is in $\mathsf{NP} \cap \mathsf{coNP}$ (as in Theorem 4), defining our predicate Q is straightforward: we let $Q(r, w) = 1$ iff w contains 'proofs' for the n values $(L(g_1), \ldots, L(g_n))$.

If our hard language is merely in NP (as in Theorem 5), we need to work harder. In this case, we let $Q(r, w) = 1$ iff w contains proofs that $L(g_i) = 1$, for a 'sufficient number' of the strings g_i. The idea is that if a small circuit $C(r)$ could with some noticeable probability guess such proofs for 'almost all' the indices i for which $L(g_i) = 1$, then C could be modified to correctly guess $(L(g_1), \ldots, L(g_n))$ with noticeable probability, contrary to the properties of the generator. Making this idea work involves showing that the set size $|\{i \in [n] : L(g_i) = 1\}|$ is highly concentrated around its expectation. For this we rely on a recently proved concentration result called the Strong Chernoff Bound for Expander Walks [WX05, WX08, Hea08]. This result is perfectly suited to analyze the Impagliazzo-Wigderson generator (which is partly defined in terms of walks on expander graphs).

The precise form of our assumptions in Theorems 4 and 5 are dictated by the hardness-amplification tools currently available. In particular, sufficiently strong hardness-amplification is only available if we make a hardness assumption against nonuniform, exponential-sized circuits. We believe versions of Theorems 4 and 5 should be possible for a *uniform* hardness assumption; recently Impagliazzo et al. [IJKW08] made partial progress towards the hardness-amplification tools needed.

2 Preliminaries

2.1 Basic definitions

We presume familiarity with basic notions in complexity theory, in particular familiarity with the classes P, NP, and AM. We define promise classes and the promise class prAM in Section 2.3.

For a language $L \subseteq \{0,1\}^*$, we use $L(x)$ to denote the characteristic function of L. We use $|x|$ to denote the length of a string x over some (possibly non-Boolean) alphabet Σ. $d(x,y)$ denotes the Hamming distance between strings $x, y \in \Sigma^n$, and $d(x,S)$ is the generalized Hamming distance between $x \in \Sigma^n$ and a set $S \subseteq \Sigma^n$. If $d(x,S) \leq c$ we say x is c-*close to* S, otherwise x is c-*far from* S. Similarly, for $\alpha \in [0,1]$, if $d(x,S) \leq \alpha n$ we say x is α-*close in relative distance to* S, otherwise x is α-*far in relative distance from* S.

$H(t) : [0,1] \to [0,1]$ denotes the binary entropy function, $H(t) = -t \log t - (1-t) \log(1-t)$ for $t \in (0,1)$ and $H(0) = H(1) = 0$. We let $V_{n,k}$ denote the discrete volume of the Hamming sphere of radius k in $\{0,1\}^n$; that is,

$$V_{n,k} := \sum_{0 \leq i \leq k} \binom{n}{i},$$

and we use the known bound $V_{n,\alpha n} \leq 2^{H(\alpha)n}$ (valid for $\alpha \in [0,1/2]$).

When we speak of circuits, unless otherwise mentioned we mean deterministic Boolean circuits of fanin-two, and we measure circuit size (denoted $|C|$ for circuit C) as the number of gates. For functions $p(n) \in [0,1]$, $s(n) \geq 0$ we say that a language L is $p(n)$-*hard for size* $s(n)$ if for every Boolean circuit C of size $\leq s(n)$, $\Pr_{x \in \{0,1\}^n}[C(x) = L(x)] \leq p(n)$. We extend this definition to general functions: we say that a function $F : \{0,1\}^n \to \{0,1\}^m$ is $p(n)$-*hard for size* $s(n)$ if for every m-output Boolean circuit C of size $\leq s(n)$, $\Pr_{x \in \{0,1\}^n}[C(x) = F(x)] \leq p(n)$.

2.2 CSPs, PCPPs, and codes

Fix an integer $k \geq 1$. A k-*local Constraint Satisfaction Problem*, or k-*CSP*, over finite alphabet Σ is a collection $\psi(x) = \psi_1(x), \ldots \psi_m(x)$ of Boolean-valued functions on the input $x = (x_1, \ldots x_n) \in \Sigma^n$, where each ψ_j depends only on some k variables of x and is specified by a k-tuple $I_j \subseteq [n]$ and a truth-table on these k variables. Define $\mathrm{Val}_\psi(x)$, the *value of ψ on x*, as the fraction of constraints ψ_j that are satisfied by x (i.e. such that $\psi_j(x) = 1$).

Next we define PCPPs. Fix a circuit $C(x)$ on n Boolean input variables, a finite alphabet Σ, and a parameter $\beta > 0$. We say that a k-CSP ψ is a *PCPP for C over Σ with security β* if:

1. ψ is defined on variable set (x,z), where x are the Boolean input variables to C and z are auxiliary 'proof' variables taking values in Σ;

2. For any $x \in \{0,1\}^n$, if $C(x) = 1$ then there exists a setting of z such that $Val_\psi(x,z) = 1$;

3. For all $x \in \{0,1\}^n$ and z, $\mathrm{Val}_\psi(x,z) \leq 1 - \beta \cdot \frac{d(x,C^{-1}(1))}{n}$.

The *proof size* of ψ is the number of variables in z.

The following positive result on PCPPs is due to Dinur.

Theorem 6. *[Din07, Cor. 9.3] There is a constant-size alphabet Σ_0, a constant $\beta > 0$, and a polynomial-time algorithm that, given a circuit $Q(x)$ of size t, produces a 2-CSP $\psi_Q(x,z)$ that is a PCPP for Q over Σ_0 with security β. Moreover, the proof size of ψ is $O(\mathrm{poly}(t))$.*

Following the techniques of earlier papers working with PCPPs, we will use PCPPs in conjunction with efficient error-correcting codes. A (binary) *code* is an injective map $E : \{0,1\}^N \to \{0,1\}^{N'}$ where $N' \geq N$. We also use E to denote the image of the map, i.e., we consider $E \subseteq \{0,1\}^{N'}$. The *minimum distance* of the code is the minimum over distinct $u, v \in E$ of $d(u, v)$. An algorithm D *decodes E from an η fraction of errors* if, given any string u at relative distance at most η from some $u' \in E$, $A(u)$ outputs u'. Note that for such decoding to be possible, the minimum distance must be greater than $2\eta N'$.

We will use the following well-known fact:

Theorem 7. *There is a polynomial-time computable code E for all input lengths N with output length $N' = O(N)$, and an $\eta > 0$, such that E can be polynomial-time decoded from an η fraction of errors.*

Many such constructions are known; recently Goldwasser et al. [GGH⁺08] gave a construction in which the decoder algorithm can be implemented in AC^0, i.e., with constant-depth, polynomial-size Boolean circuits.

2.3 Promise problems and prAM

A *promise problem* is a pair $\Pi = (\Pi_{YES}, \Pi_{NO})$ of disjoint subsets of $\{0,1\}^*$ (the 'yes' and 'no' instances, respectively). For a function $s(n) \in [0,1]$, we say that $(\Pi_{YES}, \Pi_{NO}) \in \mathsf{prAM}_{1,s(n)}$ if there exists a polynomial-time randomized algorithm $M(x, r, w)$, with $|r|, |w| = O(\text{poly}(n))$ such that:

1. (Completeness) If $x \in \Pi_{YES}$, then with probability 1 over the random string r, there exists a $w = w(r)$ such that $M(x, r, w) = 1$;

2. (Soundness) If $x \in \Pi_{NO}$ and $|x| = n$, then the probability over the random string r that there exists a w such that $M(x, r, w) = 1$, is at most $s(n)$.

The algorithm M defines an 'Arthur-Merlin protocol': we consider that a polynomially bounded verifier Arthur chooses a random 'challenge' r for the computationally unbounded Merlin, who sees r and gives a response w which Arthur accepts or rejects..

We define $\mathsf{prAM} = \mathsf{prAM}_{1,1/3}$. A promise problem $\Pi_1 = (\Pi_{YES}, \Pi_{NO})$ is prAM-*hard* if for all $\Pi' = (\Pi'_{YES}, \Pi'_{NO})$ in prAM, there exists a polynomial-time computable reduction $R(x)$, such that, if $x \in \Pi'_{YES}$, then $R(x) \in \Pi_{YES}$, while if $x \in \Pi'_{NO}$, then $R(x) \in \Pi_{NO}$. We say that Π is prAM-*complete* if Π is in prAM and is prAM-hard.

It is not hard to see that for any $\Pi \in \mathsf{prAM}$, the soundness parameter $1/3$ in the protocol can be made exponentially small in n, by parallel repetition of the original protocol. However, we require soundness-amplification that is more efficient in its use of randomness. This is provided by a result of Bellare et al. [BGG93]. They state their theorem for AM, not for prAM, but the proof carries over without changes to the promise setting and we state it for this setting.

Theorem 8. *[BGG93] Let $\Pi = (\Pi_{YES}, \Pi_{NO}) \in \mathsf{prAM}$, where $M(x, r, w)$ is a polynomial-time predicate defining an Arthur-Merlin protocol for Π. Let $n = |x|$, and fix a polynomial $m(n)$. Then there exists an Arthur-Merlin protocol for Π defined by a polynomial-time predicate $M'(x, r', w')$, with $|w'| \leq O(\text{poly}(n)), |r| \leq |r'| \leq O(|r| + m(n))$, and with soundness $2^{-m(n)}$.*

The randomness-efficiency in the above result has been improved in more recent work (see [MU02] for a discussion), but we do not need or use these improvements.

2.4 AM-k-CSPs

By an *AM-k-CSP* we mean a k-CSP $\psi(r, z)$, where r are Boolean and z may be non-Boolean. We call r the 'Arthur-variables' and z the 'Merlin-variables'. Informally speaking, we are interested in the game in which the r are first set uniformly by Arthur, and then Merlin sets z to try to maximize the fraction of constraints of ψ satisfied by (r, z).

For any fixed $k \geq 1$, soundness parameter $s = s(|r|) \in [0, 1]$, alphabet Σ, and fixed $\varepsilon \in (0, 1]$, we define the promise problem $\mathsf{Gap - AM - k - CSP}_{1, 1-\varepsilon, s(|r|)} = (\Pi_{AM-CSP, YES}, \Pi_{AM-CSP, NO})$ as follows. Both 'yes' and 'no' instances are AM-k-CSPs over Σ. If $\psi(r, z) \in \Pi_{AM-CSP, YES}$, we are promised that for all choices of r, there exists a z such that $\mathrm{Val}_\psi(r, z) = 1$. If $\psi(r, z) \in \Pi_{AM-CSP, NO}$, we are promised that only for at most an $s(|r|)$ fraction of strings r does there exist a z with $\mathrm{Val}_\psi(r, z) > 1 - \varepsilon$.

3 An Augmented PCPP

As a tool for proving Theorems 1, 4, and 5, we prove the following 'augmented' version of the PCPP Theorem (Theorem 6), which we derive from Theorem 6. We remark that the proof of our prAM-completeness result (Theorem 1) uses only condition (9.i) of the Lemma below; this first part of the Lemma is quite similar to previous uses of PCPPs. Also, our use of error-correcting codes will only be important for establishing condition (9.ii).

Lemma 9. *There is a finite alphabet Σ_0 such that the following holds. For any $\varepsilon > 0$ there is a $\nu > 0$ and a polynomial-time algorithm A that takes as input a Boolean circuit $C = C(r, w)$. A outputs a 2-CSP $\psi(r, z)$, where $|z| = O(\mathrm{poly}(|C|))$ and the variables of z are over Σ_0. Letting $\ell = |r|$, ψ has the following properties:*

(9.i) For all r, if there is a w such that $C(r, w) = 1$, then there is a z such that $\mathrm{Val}_\psi(r, z) = 1$. On the other hand, if r is $\alpha\ell$-far from any r' for which $C(r', \cdot)$ is satisfiable, then for all z, $\mathrm{Val}_\psi(r, z) < 1 - \Omega(\alpha)$.

(9.ii) Suppose $P(r)$ is any (possibly randomized) procedure such that with probability at least $p = p(\ell)$ over a uniform $r \in \{0, 1\}^\ell$ and any randomness in P, $P(r)$ outputs a z such that $\mathrm{Val}_\psi(r, z) > 1 - \nu$.

Then there exists a deterministic procedure $\tilde{P}(r)$, such that with probability at least $p(\ell) \cdot 2^{-\varepsilon\ell}$ over uniform r, $\tilde{P}(r)$ outputs a w such that $C(r, w) = 1$. Moreover, $\tilde{P}(r)$ is computable by a nonuniform, $\mathrm{poly}(|C|)$-sized circuit that makes a single oracle call to P on the same input length (with P's randomness fixed nonuniformly).

Proof. Let $N := |w|$. We can assume, by padding w if necessary, that $N \geq \ell$. Let $E : \{0, 1\}^N \to \{0, 1\}^{N'}$ be the error-correcting code given by Theorem 7 (with $N' = O(N)$), applied to inputs of length N.

Let $b = b(\ell) := \lceil \frac{N'}{\ell} \rceil$. Define a predicate $Q(r_1, r_2, \ldots, r_b, u)$, with $|r_i| = |r| = \ell$ for $i \leq b$ and $|u| = N'$, by the following rule: $Q(r_1, r_2, \ldots r_b, u) = 1$ iff $r_1 = r_2 = \ldots, = r_b$, $u = E(w)$ for some w, and $C(r_1, w) = 1$. Clearly we can efficiently construct a circuit of size $O(\mathrm{poly}(|C|))$ computing Q. Note that by our setting of b, there are more variables in the blocks r_j than in u.

Let $\psi_0 = \psi_Q((r_1, \ldots, r_b, u), Z)$ be the PCPP 2-CSP for Q over alphabet Σ_0 given by Theorem 6, efficiently constructible and of size $O(\mathrm{poly}(|C|))$. We take ψ_0 and make two changes. First, we

8

substitute the variables of r for the corresponding variables of each vector r_i. Second, we allow the variables of u to range over all of Σ_0 (we may assume $\{0,1\} \subseteq \Sigma_0$), and modify each constraint to reject in the case where one or more of its u-variables are set to a non-Boolean value.

We denote the resulting 2-CSP by $\psi(r, u, Z)$. We claim that this efficiently constructible 2-CSP satisfies the conditions of Lemma 9's statement, with $z := (u, Z)$ and Σ_0 as in Theorem 6. Our setting of $\nu > 0$ will be determined later.

First, we show that condition (9.i) is satisfied. Consider any $r \in \{0,1\}^\ell$. Suppose that there exists $w \in \{0,1\}^N$ such that $C(r, w) = 1$. Then $Q(r, r, \ldots, r, E(w)) = 1$. Using the completeness property of PCPPs, there exists a Z such that $\mathrm{Val}_\psi((r, E(w)), Z) = 1$. On the other hand, say r is $\alpha\ell$-far from any r' for which $C(r', \cdot)$ is satisfiable. Given any $u \in \Sigma_0^{N'}$, let us choose some string $u' \in \{0,1\}^{N'}$ which agrees with u on any variable where u is Boolean. We observe that $(r, r, \ldots, r, u')$ is $\alpha/2$-far in relative distance from $Q^{-1}(1)$. By the soundness property of PCPPs, for any choice of Z, $\mathrm{Val}_\psi((r, u'), Z) < 1 - \frac{\alpha\beta}{2}$, where $\beta > 0$ is the constant from Theorem 6. Also, by the way we defined ψ, $\mathrm{Val}_\psi((r, u), Z) \leq \mathrm{Val}_\psi((r, u'), Z)$. We have verified condition (9.i).

Now we turn to condition (9.ii). Let $P(r)$ be as described in (9.ii). Note that by averaging, we may fix (nonuniformly) some value of the randomness used by P while preserving the lower-bound $p(\ell)$ on its success probability over the choice of r; we do so and consider P a deterministic algorithm from now on. We set $\nu := \eta\beta\gamma/4$, where η is the constant in Theorem 7, β is the constant in Theorem 6, and $\gamma \in (0, 1)$ is a small constant to be announced.

Let $P'(r)$ be the procedure that, on input r, computes $z = P(r) = (u, Z)$ and runs the polynomial-time decoder for E on u, yielding a string $w \in \{0,1\}^N$. Let P' output w.

We analyze the behavior of P'. Let $z = (u, Z)$ be any output of $P(r)$ such that $\mathrm{Val}_\psi(r, z) > 1 - \nu$. By the soundness property of PCPPs, the string $(r, r, \ldots, r, u)$ must be $\frac{\nu}{\beta} = \frac{\eta\gamma}{4}$-close in relative distance to some string $(r_1, \ldots, r_b, u')$ for which $Q(r_1, \ldots, r_b, u') = 1$ (and thus $r_1 = \ldots = r_b$ and $u' \in E$). Since $|u'| = N' \leq b(\ell) \cdot |r| \leq 2N'$, we find that $d(r, r_1) < \gamma\ell$ and $d(u, u') < \eta N'$. The latter inequality implies that when P' applies the polynomial-time decoder to u, it correctly recovers $w = E^{-1}(u')$. Since $Q(r_1, \ldots, r_b, u') = 1$, we have $C(r_1, w) = 1$.

To analyze $\tilde{P}$, say that a string $r \in \{0,1\}^n$ is *good* if $P'(r)$ outputs a w such that there exists an r' at distance at most $\gamma\ell$ from r, such that $C(r', w) = 1$. Our analysis of P', combined with our original assumption about the success probability of P, shows that at least a $p(\ell)$ fraction of strings r are good.

Now we define the procedure $\tilde{P}(r)$: $\tilde{P}(r)$ first chooses a vector $v \in \{0,1\}^l$ uniformly from the set of all strings of Hamming weight at most γl, then outputs $P'(r + v)$. Note that, if r is selected uniformly, $r + v$ is also uniform and, after conditioning on its value, r is uniformly distributed over all strings at distance at most $\gamma\ell$ from $r + v$. Thus, conditioning on $r + v$ being good, we have at least a $1/V_{\ell, \gamma\ell} \geq 2^{-H(\gamma)\ell}$ chance that $C(r, P'(r + v)) = 1$. So the overall success probability of $\tilde{P}(r)$ is at least $p(\ell) \cdot 2^{-H(\gamma)\ell}$. Since $H(\gamma) \to 0$ as $\gamma \to 0$, we may choose $\gamma > 0$ so that the success probability is at least $p(\ell) \cdot 2^{-\varepsilon\ell}$.

P' is clearly a polynomial-time algorithm making one call to P, while $\tilde{P}$ simply makes one call to P' after its random sampling and bitwise addition mod 2. The choice of v can be nonuniformly fixed in a way that does not decrease the success probability, so $\tilde{P}$ can be implemented with the resources claimed. Thus we have verified condition (9.ii), completing the proof of the Lemma. $\square$

4 PCP Characterization of prAM

In this section we prove Theorem 1, which we restate in the terminology of Section 2.4:

Theorem 1 (restated). *There is a finite alphabet Σ and a constant $\varepsilon > 0$, such that* $\mathsf{Gap - AM - 2 - CSP}_{1,1-\varepsilon,\exp\{-\Omega(|r|)\}}$ *is* prAM-*complete.*

Proof. First, we claim that for any $s(|r|) = o(1)$ and $\varepsilon > 0$, $\mathsf{Gap - AM - 2 - CSP}_{1,1-\varepsilon,s(|r|)} = (\Pi_{AM-CSP,YES}, \Pi_{AM-CSP,NO})$ is in prAM. The protocol is as follows: given a 2-CSP $\psi(r, z)$, Arthur picks r uniformly and Merlin responds with a setting of z. Arthur accepts iff $\mathrm{Val}_\psi(r, z) = 1$. If $\psi \in \Pi_{AM-CSP,YES}$, then clearly Arthur accepts with probability 1 when Merlin responds optimally. If $\psi \in \Pi_{AM-CSP,NO}$, then Arthur accepts with probability at most $s(|r|)$, which is greater than $2/3$ for large enough $|r|$. (For instances with $|r|$ below this threshold, Arthur can simply request certificates $z(r)$ for every setting of r and verify that each satisfies $\mathrm{Val}_\psi(r, z(r)) = 1$.)

Thus our main task is to show that the promise problem is prAM-hard, for appropriate choice of parameters. Let $\Pi = (\Pi_{YES}, \Pi_{NO}) \in$ prAM, and let $M_1(x, r_1, w_1)$ be a polynomial-time-computable predicate defining an Arthur-Merlin protocol for Π. We use parameters $n = |x|$, $\ell_1(n) = |r|$; by definition of prAM we have $\ell_1(n) = O(\mathrm{poly}(n))$ and $|w_1| = O(\mathrm{poly}(n))$. By padding r_1 if necessary we may assume $\ell_1(n) \geq n$. Apply Theorem 8 to M_1, with the setting $m(n) := \ell_1(n)$. Thus we get a new Arthur-Merlin protocol $M_2(x, r_2, w_2)$ for Π, with $|r_2| = \ell_2(n) \in [n, \ldots, D \cdot \ell_1(n)]$ (for some fixed $D > 0$), $|w_2| = O(\mathrm{poly}(n))$, and with soundness $2^{-\ell_1(n)}$.

Given an input $x \in \Pi_{YES} \cup \Pi_{NO}$, we construct a $\mathrm{poly}(n)$-sized circuit $C(r_2, w_2) = C_x(r_2, w_2)$ that accepts iff $M_2(x, r_2, w_2) = 1$. To this circuit we apply the algorithm A of Lemma 9 (with a setting of $\varepsilon > 0$ to be announced), yielding a 2-CSP $\psi = \psi(r_2, z)$ which we make the output of our reduction.

We show the correctness of the reduction. First, suppose that $x \in \Pi_{YES}$. Then for each choice of r_2, there exists a w_2 such that $M_2(x, r_2, w_2) = 1$. By condition (9.i) of Lemma 9, there exists z such that $\mathrm{Val}_\psi(r_2, z) = 1$. Thus $\psi \in \Pi_{AM-CSP,YES}$.

Now suppose that $x \in \Pi_{NO}$. Then by the soundness property of M_2, the number of strings r_2 for which $M_2(x, r_2, \cdot)$ is satisfiable is at most $2^{-\ell_1(n)} \cdot 2^{\ell_2(n)} \leq 2^{(1-\frac{1}{D})\ell_2(n)}$. Thus the number of r_2 for which there exists an r' at distance $\leq \alpha \ell_2(n)$ from r_2, such that $M_2(x, r', \cdot)$ is satisfiable, is at most

$$V_{\ell_2(n),\alpha \ell_2(n)} \cdot 2^{(1-\frac{1}{D})\ell_2(n)} \leq 2^{(H(\alpha)+1-\frac{1}{D})\ell_2(n)}.$$

Choosing $\alpha > 0$ such that $H(\alpha) < \frac{1}{D}$, we find that with probability $\geq 1 - \exp\{-\Omega(\ell_2(n))\}$ over a uniform choice of r_2, r_2 is $\alpha \ell_2(n)$-far from any r' such that $C(r', \cdot)$ is satisfiable. For such r_2 and for any z, condition (9.i) of Lemma 9 tells us that $\mathrm{Val}_\psi(r_2, z) < 1 - \Omega(\alpha)$.

Thus if we fix $\varepsilon > 0$ as an appropriately small constant and choose an appropriate $s(|r_2|) = \exp\{-\Omega(|r_2|)\}$, we have $\psi \in \Pi_{AM-CSP,NO}$. This completes the proof of correctness for our reduction. $\qquad \square$

5 Randomized Optimization Hypotheses Imply Collapse of AM

What significance might Theorem 1, our 'PCP characterization of AM', have for the project of trying to prove new upper bounds on the power of this class? In the Introduction we gave two hypotheses inspired by Theorem 1. Each of these hypotheses, if true, would have major implications

for the study of AM; this is the content of Claims 2 and 3 from the Introduction, which we prove next.

Proof of Claim 2. Let $L \in$ AM; then $(L, \overline{L}) \in$ prAM. Given an instance x, let Arthur run the reduction in Theorem 1 on input x, producing a 2-CSP $\psi(r, z)$. Let Merlin send Arthur a polynomial-sized circuit $C : \{0,1\}^{\ell} \to \{0,1\}^m$, with $\delta := \varepsilon$ (here ε is from Theorem 1). Then Arthur runs C on a sufficiently large $(O(\mathrm{poly}(n)))$ number of random choices of r, accepting only if he finds an r such that $\mathrm{Val}_{\psi}(r, C(r)) \geq 1 - \varepsilon$.

First suppose $x \in L$; then by the completeness property of our reduction, for all r there exists a z for which $\mathrm{Val}_{\psi}(r, z) = 1$. If Merlin sends the circuit C_{ψ} assumed to exist by Hypothesis A, then with at least $1/\mathrm{poly}(|\ell|)$ probability over r, $\mathrm{Val}_{\psi}(r, C(r)) \geq 1 - \varepsilon$. So if Arthur samples a sufficiently large (polynomial) number of strings r, Arthur will accept with probability $> 2/3$.

Next suppose $x \notin L$; then our reduction guarantees that for all but an exponentially small fraction of strings r, for all z $\mathrm{Val}_{\psi}(r, z) < 1 - \varepsilon$. So Arthur's acceptance probability is negligible no matter what circuit Merlin sends. Thus we have an MA protocol for L. $\qquad\square$

Proof of Claim 3. We apply Hypothesis B with $\delta := \varepsilon/3$, yielding a value $t = t(\delta)$. Let $L \in$ AM be given, and let Arthur run the reduction from Theorem 1 on input x, yielding an instance $\psi(r, z)$. Let Merlin send a description of a function $F(r)$, where each output of F depends on at most t bits of r (note F can be described in polynomial size). Arthur performs explicit variable-substitutions $z = F(r)$ in ψ and uses linearity of expectation to exactly compute $\mathbb{E}_r[\mathrm{Val}_{\psi}(r, F(r))]$.

If $x \in L$ and Merlin sends F_{ψ} as given by Hypothesis B, this expectation is at least $(1 - \delta)^2 > 1 - 2\varepsilon/3$. On the other hand, if $x \notin L$ then, regardless of the function sent, this expectation is at most $(1 - \varepsilon) + \exp\{-\Omega(|r|)\}$. Thus for $|r|$ large enough we can distinguish the two cases. (If $|r|$ is below a fixed threshold, Arthur can instead request that Merlin send optimal values $z(r)$ for each r.) Arthur's computations are deterministic and polynomial-time, so the above defines an NP protocol for L. $\qquad\square$

Note that Claim 3 would hold even if we weakened Hypothesis B, allowing each coordinate of $F(r)$ to depend on $t(\delta, n) = O_{\delta}(\log n)$ coordinates. We state Hypothesis B in a stronger form because, although we believe it is false, we don't know how to disprove it unconditionally even in the form given.

6   Evidence Against the Randomized Optimization Hypotheses

Next we use Lemma 9, in conjunction with known results about amplification of hardness, to prove Theorems 4 and 5. That is, under various complexity-theoretic assumptions, we exhibit families of 2-CSPs $\psi(r, z)$ for which it is hard on average to approximately optimize over z, for randomly chosen r. As mentioned earlier, the conclusions of both Theorems are easily seen to falsify both of our Randomized Optimization Hypotheses, and we consider this evidence that these hypotheses are probably false.

First, amplification of hardness in NP $\cap$ coNP from $(1 - 1/\mathrm{poly}(n))$-hardness to $2/3$-hardness is made possible by the following result of Impagliazzo [Imp95, essentially Thm. 2]:

Theorem 10. *[Imp95] Suppose that there exists a language L, a function $s(n)$, and a $c > 0$, such that L is $(1 - \frac{1}{n^c})$-hard for size $s(n)$. Then for any $c' > 0$, there exists another language L' such*

that L' is $(\frac{1}{2} + O(\frac{1}{n^{c'}}))$-hard for size $\frac{s(n)}{n^{O(1)}}$. Moreover, L' is polynomial-time truth-table reducible to L.

Lemma 11. *Suppose that there exists a language $L \in \mathsf{NP} \cap \mathsf{coNP}$ and $\gamma, c > 0$ such that L is $(1 - \frac{1}{n^c})$-hard for size $2^{\gamma n}$. Then there exists another language $L' \in \mathsf{NP} \cap \mathsf{coNP}$ and a $\gamma' > 0$ such that L' is $2/3$-hard for size $2^{\gamma' n}$ (for sufficiently large n).*

Proof. Apply Theorem 10, with $s(n) := 2^{\gamma n}$ and with any $c' > 0$ and $\gamma' \in (0, \gamma)$, and use the fact that $\mathsf{NP} \cap \mathsf{coNP}$ is closed under polynomial-time reducibilities, i.e., $\mathsf{P}^{\mathsf{NP} \cap \mathsf{coNP}} = \mathsf{NP} \cap \mathsf{coNP}$. $\qquad\square$

Next, the Impagliazzo-Wigderson pseudorandom generator [IW97] allows us to amplify 'moderate' hardness of the type produced by Lemma 11 into 'extreme' hardness, albeit of a function problem rather than a decision problem (in [IW97] additional techniques are used to produce extremely hard decision problems, but we do not follow this path). The next definition follows [IW97] (and previous works). Given a language L, an integer $c \geq 1$, a parameter $k = k(n)$, and a function $G(r) : \{0,1\}^{cn} \to \{0,1\}^{k \times n}$ (called a 'generator' function), define $L^k \circ G : \{0,1\}^{cn} \to \{0,1\}^k$ by

$$(L^k \circ G)(r) := (L(G_1(r)), L(G_2(r)), \ldots, L(G_k(r))),$$

where the string $G(r)$ is divided into k blocks $G_1(r), \ldots, G_k(r)$, each of length n. The basic idea is that if $G(r)$ is appropriately 'pseudorandom', then the collection $G_1(r), \ldots, G_k(r)$ should behave in important respects like a truly independent collection of random strings. In particular, if it is somewhat hard to compute $L(x)$ for a random x, it should be very hard to compute $(L^k \circ G)(r)$ correctly when k is large.

The following result (a restatement of [IW97, Thm. 2.12]) gives the main hardness-amplification property of the generator defined in that paper, which we denote G_{IW}.

Theorem 12. *[IW97] For any $\gamma > 0$, there are $\gamma', c > 0$, and a polynomial-time computable $G_{IW} : \{0,1\}^{cn} \to \{0,1\}^{n \times n}$, such that: if L is $2/3$-hard for size $2^{\gamma n}$, then $(L^n \circ G_{IW})(r)$ is $2^{-\gamma' n}$-hard for size $2^{\gamma' n}$.*

(Recall our definition of average-case hardness for general functions from Section 2.1.)

Proof of Theorem 4. We begin by applying Lemma 11 to our language $L \in \mathsf{NP} \cap \mathsf{coNP}$, yielding a language $L' \in \mathsf{NP} \cap \mathsf{coNP}$ that is $2/3$-hard for circuits of size $2^{\gamma_0 n}$ for some $\gamma_0 > 0$. Then we apply Theorem 12 to L'; we derive a $\gamma' > 0$, such that $((L')^n \circ G_{IW})(r)$ is $2^{-\gamma' n}$-hard for size $2^{\gamma' n}$.

Since $L' \in \mathsf{NP} \cap \mathsf{coNP}$, there exists a polynomial-time witness predicate $M(x, w)$, producing outputs from $\{0, 1, ?\}$, satisfying:

1. for all (x, w), $M(x, w) \in \{L'(x), ?\}$;

2. for all x, there exists a w such that $M(x, w) = L'(x)$;

3. $|w| = O(\text{poly}(n))$.

Let $t(n) = |w|$. We reformat M if necessary to ensure that the first bit of w consists of a 'claim' bit, call it w_{cl}, such that for any (x, w) with $M(x, w) = L'(x)$, we have $w_{cl} = L'(x)$. Next we define $M'(x, w)$, which outputs 1 if $M(x, w) \in \{0, 1\}$, 0 otherwise. M' is also polynomial-time computable.

Define a predicate $Q(r, w_1, \ldots, w_n) : \{0,1\}^{cn} \times \{0,1\}^{n \times t(n)} \to \{0,1\}$ as follows: $Q(r, w_1, \ldots, w_n) = 1$ iff for all $i \in [n]$, $M'(G_{IW,i}(r), w_i) = 1$. Q is polynomial-time computable since G_{IW} and M' are, so let Q_n be a $O(\text{poly}(n))$-sized circuit for Q on input parameter n. Clearly Q_n is efficiently constructible.

We claim that Q defines a hard-on-average search problem. To see this, suppose $C(r) : \{0,1\}^{cn} \to \{0,1\}^{n \times t(n)}$ is any circuit of size at most $2^{\gamma' n}$ which has some $p(n)$ probability over r of outputting a collection $w_1, \ldots, w_n$ for which $Q(r, w_1, \ldots, w_n) = 1$. Then we may construct a circuit $C'(r)\{0,1\}^{cn} \to \{0,1\}^n$ that simply restricts the output of $C(r)$ to the 'claim' bits of the strings $w_1, \ldots, w_n$ that C produces. Observe that $C'(r)$ has a $p(n)$ chance (over r) of correctly outputting $((L')^n \circ G_{IW})(r)$. Moreover, $C'(r)$ also has size bounded by $2^{\gamma' n}$. We conclude $p(n) \leq 2^{-\gamma' n}$.

We invoke Lemma 9 with $\varepsilon := \gamma'/(2c)$, yielding a poly-time algorithm A (and an associated $\nu > 0$). We apply this A to Q_n, yielding a 2-CSP $\psi_n(r, z)$ (here $|z| = d(n) = O(\text{poly}(n))$). We claim that the 2-CSP family $\{\psi_n(r, z)\}_{n>0}$ satisfies the conditions of Theorem 4.

To see this, first note that for all r, $M'(r, \cdot)$ is satisfiable; so, there exists $w_1, \ldots, w_n$ such that $Q_n(r, w_1, \ldots, w_n) = 1$. Thus by condition (9.i) of Lemma 9, there exists z such that $\text{Val}_{\psi_n}(r, z) = 1$. So condition (4.i) is satisfied.

To establish condition (4.ii), let $\gamma_2 := \varepsilon$ and $\theta := \nu$. Suppose $C(r') : \{0,1\}^{cn} \to \{0,1\}^{|w'|}$ is a circuit of size at most $2^{\gamma_2 n}$, such that with some probability $q(n)$, $\text{Val}_{\psi_n}(r', C(r')) > 1 - \theta$. By condition (9.ii) of Lemma 9, there exists a circuit $\tilde{C}(r) : \{0,1\}^{cn} \to \{0,1\}^{n \times t(n)}$, such that with probability at least $q(n) \cdot 2^{-\varepsilon(cn)}$ over r, $Q_n(r, \tilde{C}(r)) = 1$. Moreover, $\tilde{C}$ is of size at most $|C| + O(\text{poly}(n))$, which for large enough n is less than $2^{\gamma' n}$. By our previous analysis we find that $q(n) \cdot 2^{-\varepsilon(cn)} \leq 2^{-\gamma' n}$, i.e., $q(n) \leq 2^{(\varepsilon c - \gamma')n} = 2^{-\gamma' n/2} = \exp\{-\Omega(n)\}$. We have proved condition (4.ii), for our settings of γ_2, θ. $\qquad\square$

Next we turn to prove Theorem 5. For this, we need a more detailed analysis of the generator G_{IW}. Besides the hardness-amplification property summarized in Theorem 12, G_{IW} has another useful property: with very high probability over r, the fraction of the strings $G_{IW,1}(r), \ldots, G_{IW,n}(r)$ which are in L is close to $|L \cap \{0,1\}^n|/2^n$, that is, close to the fraction we'd expect if these strings were drawn independently and uniformly. To prove this fact (not proved or used in [IW97]), we first describe the generator in more detail.

The input r to G_{IW} consists of two parts, $r = (r_a, r_b)$. $G_{IW}(r_a, r_b)$ is defined blockwise for $i \in [n]$ as

$$G_{IW,i}(r_a, r_b) = K_i(r_a) + K'_i(r_b),$$

with $K_i : \{0,1\}^{|r_a|} \to \{0,1\}^n$, $K'_i : \{0,1\}^{|r_b|} \to \{0,1\}^n$, and with $+$ denoting bitwise addition mod 2. The definition of K'_i is not important to us; let us describe the functions K_i. The string r_a defines a random walk of length n (counting the starting vertex) on an explicit expander graph $\mathcal{G}_n$ with vertex set $\{0,1\}^n$. $\mathcal{G}_n$ is 16-regular with normalized second eigenvalue λ_n at most some fixed $\lambda < 1$. $v_i = K_i(r_a)$ represents the i-th vertex visited in this walk. v_1 is a uniform element, and each subsequent step v_{i+1} is a uniform choice from among the neighbors of v_i. (Note that this can be achieved with $|r_a| = O(n)$ random bits as claimed.)

We will use the following powerful result, called the Strong Chernoff Bound for Expander Walks, proved by Healy [Hea08].

Theorem 13. *[Hea08] Let $G = (V, E)$ be a d-regular graph with second eigenvalue λ, let $m > 0$, and let $f_1, \ldots, f_m : V \to [0, 1]$ have expectations $\mu_1, \ldots, \mu_m$ (over a uniform choice of $v \in V$).*

13

Taking a random walk $v_1, \ldots, v_m$ on G with uniform starting-point, we have for all $\varepsilon > 0$,

$$\Pr\left[\left|\sum_{i \leq m} f_i(v_i) - \sum_{i \leq m} \mu_i\right| \geq \varepsilon m\right] \leq 2e^{-\frac{\varepsilon^2(1-\lambda)m}{4}}.$$

A more general claim was made earlier by Wigderson and Xiao [WX05], but the proof contained an error, as pointed out in [WX08]. (A valid proof of the Theorem above, with different constants, can still be extracted from [WX05].)

For any $r \in \{0,1\}^{cn}$, let $\sharp(r) := |\{i \in [n] : L(G_{IW,i}(r)) = 1\}|$. Theorem 13 implies the following concentration bound for the generator G_{IW}:

Lemma 14. *Let L be an arbitrary language. Let $c_n = |L \cap \{0,1\}^n|/2^n$. Then for any fixed $\delta > 0$,*

$$\Pr_r[|\sharp(r) - c_n \cdot n| \geq \delta n] \leq \exp\{-\Omega(n)\}.$$

Proof. Recall that $r = (r_a, r_b)$. We show that the above inequality is true after conditioning on any value of r_b; this will prove the Lemma. Let $(v'_1, \ldots, v'_n) = (K'_1(r_b), \ldots, K'_n(r_b))$. Then for $i \in [n]$, $G_{IW,i}(r) \in L$ iff $K_i(r_a) + v'_i \in L$, or equivalently $K_i(r_a) \in L_n + v'_i$ (where $L_n := L \cap \{0,1\}$ and $L_n + v'_i = \{x + v_i : x \in L_n\}$).

Define $f_i : \{0,1\}^n \to \{0,1\}$ to be the characteristic function of $L_n + v'_i$. Clearly $\mu_i = c_n$ for all i. The result now follows by a direct application of Theorem 13, using the fact that $\mathcal{G}_n$ has second eigenvalue bounded away from 1. $\qquad\square$

Proof of Theorem 5. Our choice of ε_0, determined later, will be no larger than $1/6$, so by Theorem 12, our hypothesis implies there is a $\gamma' > 0$, a $c > 0$, and a polynomial-time $G_{IW} : \{0,1\}^{cn} \to \{0,1\}^{n \times n}$, such that: for any circuit $C : \{0,1\}^n \to \{0,1\}^n$ of size at most $2^{\gamma'n}$,

$$\Pr_r[C(r) = (L^n \circ G_{IW})(r)] \leq 2^{-\gamma'n}.$$

Let $M(x,w)$ be a polynomial-time verifier for L: $x \in L$ iff there exists w such that $M(x,w) = 1$. Let $t(n) = |w| = O(\text{poly}(n))$.

Let $\sharp(r)$ be as defined after Theorem 13. If $C : \{0,1\}^{cn} \to \{0,1\}^{t(n) \times n}$ is a circuit producing n strings $w_1, \ldots, w_n$, each of length $t(n)$, define

$$\sharp_C(r) := |\{i \in [n] : M(G_{IW,i}(r), w_i) = 1\}|.$$

Claim 15. *There exists $\gamma'' > 0$ such that the following holds. If $C(r) : \{0,1\}^{cn} \to \{0,1\}^{t(n) \times n}$ is a circuit of size at most $2^{\gamma''n}$, then*

$$\Pr_r[\sharp(r) < \sharp_C(r) + \gamma''n] < 2^{-\gamma''n}.$$

Proof (of Claim 15). Let $\alpha > 0$. Say $C(r) : \{0,1\}^{cn} \to \{0,1\}^{t(n) \times n}$ is a circuit of size at most $2^{\alpha n}$, such that $\Pr_r[\sharp(r) < \sharp_C(r) + \alpha n] \geq 2^{-\alpha n}$. Consider the following randomized procedure that attempts to compute $(L^n \circ G_{IW})(r)$:

- Let $C(r) = (w_1, \ldots, w_n) \in \{0,1\}^{t(n) \times n}$. Let $I \subseteq [n]$ be the indices i for which $M(G_{IW,i}(r), w_i) = 1$. Pick a random subset J of $[n]$, uniformly from the set of all subsets of size less than αn (including the empty set). Output the characteristic vector of $I \cup J$.

We analyze this procedure. Suppose r is any input for which $\sharp(r) < \sharp_C(r) + \alpha n$. Note that by definition of M, we always have $I \subseteq \{i \in [n] : L(G_{IW,i}(r)) = 1\}$. Then there exists a $J \subseteq [n] \setminus S$, of size less than αn, such that $I \cup J = \{i \in [n] : L(G_{IW,i}(r)) = 1\}$. Thus conditioned on this event, the procedure succeeds with probability at least $\frac{1}{V_{n,\alpha n}} \geq 2^{-H(\alpha)n}$. So the overall success probability is at least $2^{-\alpha n} \cdot 2^{-H(\alpha)n} = 2^{-(\alpha + H(\alpha))n}$.

Let us nonuniformly fix a setting J that maximizes the procedure's success probability, and use this choice to run the procedure. The result is a (nonuniform) circuit of size $2^{\alpha n} + O(\mathrm{poly}(n))$, with success probability $\geq 2^{-(\alpha + H(\alpha))n}$. For α sufficiently small this contradicts the hardness of $(L^n \circ G_{IW})$, proving the claim. $\qquad\square$

Now we set $\varepsilon_0 := \min(1/6, \gamma''/4)$. Fix any circuit $C : \{0,1\}^{cn} \to \{0,1\}^{t(n) \times n}$ of size at most $2^{\gamma'' n}$. We use Lemma 14 applied to $\delta := \varepsilon_0$, and the previous Claim, to find that, with probability $\geq 1 - \exp\{-\Omega(n)\}$ over r, we have the simultaneous inequalities $(c_n + \varepsilon_0)n > \sharp(r) > (c_n - \varepsilon_0)n$ and $\sharp(r) \geq \sharp_C(r) + \gamma'' n$. Call a string r with this property C-typical.

What is c_n? We claim it must lie in $[1/2 - \varepsilon_0, 1/2 + \varepsilon_0]$. For otherwise, a size-1 circuit could guess $L(x)$ with probability greater than $1/2 + \varepsilon_0$ by guessing the majority value on length n, contrary to our hardness assumption about L. Thus for a C-typical r, $\sharp_C(r) \leq (1/2 + \varepsilon_0)n - \gamma'' n < (1/2 - 3\gamma''/4)n$ and also $\sharp(r) \geq (1/2 - \varepsilon_0)n - \varepsilon_0 n \geq (1/2 - \gamma''/2)n$.

Defining η as some rational number in the interval $(1/2 - 3\gamma''/4, 1/2 - \gamma''/2)$, define a predicate $Q(r, w_1, \ldots, w_n) : \{0,1\}^{cn} \times \{0,1\}^{n \times t(n)} \to \{0,1\}$ as follows: $Q(r, w_1, \ldots, w_n) = 1$ iff for at least an η fraction of indices i we have $M(G_{IW,i}(r), w_i) = 1$. Q is itself polynomial-time computable, computed by some uniform family $\{Q_n\}_{n>0}$ of poly-size circuits. We have the key property that for a C-typical r, there exist $w_1, \ldots, w_n$ such that $Q(r, w_1, \ldots, w_n) = 1$, yet $Q(r, C(r)) = 0$.

Invoke Lemma 9 with $\varepsilon := \gamma''/(2c)$, yielding an algorithm A (and an associated $\nu > 0$). Then we claim $\{A(Q_n)\}_{n>0} = \{\psi_n(r,z)\}_{n>0}$ is the desired family of 2-CSPs (here $|r| = cn, |z| = d(n) = O(\mathrm{poly}(n))$). First we verify condition (5.i). Consider any r for which there exists a $w_1, \ldots, w_n$ such that $Q_n(r, w_1, \ldots, w_n) = 1$. By condition (9.i) of Lemma 9, we find that in this case there exists z such that $\mathrm{Val}_{\psi_n}(r, z) = 1$. Since all but an $\exp\{-\Omega(n)\}$ fraction of r have this property, condition (5.i) is satisfied.

To establish condition (5.ii), fix γ_2 as any value in $(0, \gamma'')$ and let $\theta := \nu$. Suppose $C(r) : \{0,1\}^{cn} \to \{0,1\}^{d(n)}$ is a circuit of size at most $2^{\gamma_2 n}$, such that with some probability $q(n)$, $\mathrm{Val}_{\psi_n}(r, C(r)) > 1 - \theta$. By condition (9.ii) of Lemma 9, there exists a circuit $\tilde{C}(r) : \{0,1\}^{cn} \to \{0,1\}^{n \times t(n)}$, such that with probability at least $q(n) \cdot 2^{-\varepsilon(cn)}$ over r, $Q_n(r, \tilde{C}(r)) = 1$. Note that such an r fails to be $\tilde{C}$-typical. Moreover, $\tilde{C}$ is of size at most $|C| + O(\mathrm{poly}(n))$, which for large enough n is less than $2^{\gamma'' n}$.

So, by our previous analysis we find that $q(n) \cdot 2^{-\varepsilon cn} \leq 2^{-\gamma'' n}$, i.e., $q(n) \leq 2^{(\varepsilon c - \gamma'')n} = 2^{-\gamma'' n/2} = \exp\{-\Omega(n)\}$. We have proved condition (5.ii). This completes the proof of Theorem 5. $\qquad\square$

Finally, we note that versions of Theorems 4 and 5 can be proved, in which both the hypotheses and conclusions apply, not to general circuits, but to the class of TC^0 circuits (i.e., constant-depth Boolean circuits with majority gates), or any circuit class containing TC^0. This is because all the reductions involved can be carried out in TC^0. (For a discussion of why the Impagliazzo and Impagliazzo-Wigderson constructions amplify hardness in TC^0, see Agrawal [Agr01]; the difficulties in amplifying hardness in lower classes like AC^0 were explored by Shaltiel and Viola [SV08].)

7   Questions for Further Research

- Does our approximation problem remain prAM-complete if each variable in the CSP $\psi(r, z)$ is restricted to appear in only a constant number of constraints? (The 'expander-replacement' technique [PY91, Pap94] allows us to restrict the occurrences of z-variables in our prAM-completeness proof; it is the 'stochastic' r-variables which pose a challenge.) Alternatively, can one perhaps show that under this restriction the problem lies in NP?

- Can we unconditionally disprove Hypothesis B? Given the sharp limitations of NC^0 circuits this might be possible.

- Can PCP ideas be used to give new upper bounds on the class AM?

- Find more applications of PCPPs in complexity theory.

8   Acknowledgements

References

[Agr01]   Manindra Agrawal. Hard sets and pseudo-random generators for constant depth circuits. In *FSTTCS*, pages 58–69, 2001.

[ALM⁺98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, 1998.

[BGG93]   Mihir Bellare, Oded Goldreich, and Shafi Goldwasser. Randomness in interactive proofs. *Computational Complexity*, 3:319–354, 1993.

[BSGH⁺06] Eli Ben-Sasson, Oded Goldreich, Prahladh Harsha, Madhu Sudan, and Salil P. Vadhan. Robust PCPs of proximity, shorter PCPs, and applications to coding. *SIAM J. Comput.*, 36(4):889–974, 2006.

[BSHLM08] Eli Ben-Sasson, Prahladh Harsha, Oded Lachish, and Arie Matsliah. Sound 3-query PCPPs are long. In *ICALP (1)*, pages 686–697, 2008.

[CFLS95]   Anne Condon, Joan Feigenbaum, Carsten Lund, and Peter W. Shor. Probabilistically checkable debate systems and nonapproximability of pspace-hard functions. *Chicago J. Theor. Comput. Sci.*, 1995, 1995.

[CFLS97]   Anne Condon, Joan Feigenbaum, Carsten Lund, and Peter W. Shor. Random debaters and the hardness of approximating stochastic functions. *SIAM J. Comput.*, 26(2):369–400, 1997.

[Din07]   Irit Dinur. The PCP theorem by gap amplification. *J. ACM*, 54(3):12, 2007.

[DR06]   Irit Dinur and Omer Reingold. Assignment testers: Towards a combinatorial proof of the PCP theorem. *SIAM J. Comput.*, 36(4):975–1024, 2006.

[GGH+08]   Shafi Goldwasser, Dan Gutfreund, Alexander Healy, Tali Kaufman, and Guy N. Rothblum. A (de)constructive approach to program checking. In *STOC*, pages 143–152, 2008.

[Hea08]   Alexander Healy. Randomness-efficient sampling within NC^1. *Computational Complexity*, 17(1):3–37, 2008.

[HRTS07]   Ishay Haviv, Oded Regev, and Amnon Ta-Shma. On the hardness of satisfiability with bounded occurrences in the polynomial-time hierarchy. *Theory of Computing*, 3(1):45–60, 2007.

[HVV06]   Alexander Healy, Salil P. Vadhan, and Emanuele Viola. Using nondeterminism to amplify hardness. *SIAM J. Comput.*, 35(4):903–931, 2006.

[IJKW08]   Russell Impagliazzo, Ragesh Jaiswal, Valentine Kabanets, and Avi Wigderson. Uniform direct product theorems: simplified, optimized, and derandomized. In *STOC*, pages 579–588, 2008.

[Imp95]   Russell Impagliazzo. Hard-core distributions for somewhat hard problems. In *FOCS*, pages 538–545, 1995.

[IW97]   Russell Impagliazzo and Avi Wigderson. *P = BPP* if *E* requires exponential circuits: Derandomizing the XOR lemma. In *STOC*, pages 220–229, 1997.

[KL94]   Ker-I Ko and Chih-Long Lin. Non-approximability in the polynomial-time hierarchy. Technical Report 94-2, Dept. of Computer Science, SUNY at Stony Brook, 1994.

[MU02]   Elchanan Mossel and Christopher Umans. On the complexity of approximating the VC dimension. *J. Comput. Syst. Sci.*, 65(4):660–671, 2002.

[O'D02]   Ryan O'Donnell. Hardness amplification within NP. In *STOC*, pages 751–760, 2002.

[Pap94]   Christos H. Papadimitriou. *Computational Complexity*. Addison Wesley, 1994.

[PY91]   Christos H. Papadimitriou and Mihalis Yannakakis. Optimization, approximation, and complexity classes. *J. Comput. Syst. Sci.*, 43(3):425–440, 1991.

[SU07]   Ronen Shaltiel and Christopher Umans. Low-end uniform hardness vs. randomness tradeoffs for AM. In *STOC*, pages 430–439, 2007.

[SV08]   Ronen Shaltiel and Emanuele Viola. Hardness amplification proofs require majority. In *STOC*, pages 589–598, 2008.

[Wil]   Ryan Williams. Improving exhaustive search implies superpolynomial lower bounds. To appear in STOC 2010.

[WX05]   Avi Wigderson and David Xiao. A randomness-efficient sampler for matrix-valued functions and applications. In *FOCS*, pages 397–406, 2005.

[WX08]   Avi Wigderson and David Xiao. Derandomizing the Ahlswede-Winter matrix-valued Chernoff bound using pessimistic estimators, and applications. *Theory of Computing*, 4(1):53–76, 2008.