Free Bits, PCPs and Non-Approximability—Towards Tight Results[*]

(Version 5: July 1997)

Mihir Bellare[†] Oded Goldreich[‡] Madhu Sudan[§]

Abstract

This paper continues the investigation of the connection between probabilistically checkable proofs (PCPs) the approximability of NP-optimization problems. The emphasis is on proving *tight* non-approximability results via consideration of measures like the "free bit complexity" and the "amortized free bit complexity" of proof systems.

The first part of the paper presents a collection of new proof systems based on a new error-correcting code called the long code. We provide a proof system which has amortized free bit complexity of $2+\epsilon$, implying that approximating Max Clique within $N^{\frac{1}{3}-\epsilon}$, and approximating the Chromatic Number within $N^{\frac{1}{5}-\epsilon}$, are hard assuming NP $\neq$ coRP, for any $\epsilon > 0$. We also derive the first explicit and reasonable constant hardness factors for Min Vertex Cover, Max2SAT, and Max Cut, and improve the hardness factor for Max3SAT. We note that our non-approximability factors for MaxSNP problems are appreciably close to the values known to be achievable by polynomial time algorithms. Finally we note a general approach to the derivation of strong non-approximability results under which the problem reduces to the construction of certain "gadgets."

The increasing strength of non-approximability results obtained via the PCP connection motivates us to ask how far this can go, and whether PCPs are inherent in any way. The second part of the paper addresses this. The main result is a "reversal" of the FGLSS connection: where the latter had shown how to translate proof systems for NP into NP-hardness of approximation results for Max Clique, we show how any NP-hardness of approximation result for Max Clique yields a proof system for NP. Roughly our result says that for any constant f if Max Clique is NP-hard to approximate within $N^{1/(1+f)}$ then NP $\subseteq \overline{\mathrm{FPCP}}[\log, f]$, the latter being the class of languages possessing proofs of logarithmic randomness and amortized free bit complexity f. This suggests that PCPs are inherent to obtaining non-approximability results. Furthermore the tight relation suggests that reducing the amortized free bit complexity is necessary for improving the non-approximability results for Max Clique.

The third part of our paper initiates a systematic investigation of the properties of PCP and FPCP as a function of the various parameters: randomness, query complexity, free bit complexity, amortized free bit complexity, proof size, etc. We are particularly interested in "triviality" results, which indicate which classes are *not* powerful enough to capture NP. We also distill the role of randomized reductions in this area, and provide a variety of useful transformations between proof checking complexity classes.

[*] In honor of Shimon Even's 60th birthday. To appear in *SIAM Journal on Computing*. Extended abstract appeared in *Proceedings of the* 0th *Symposium on Foundations of Computer Science*, IEEE, 1995.

[†] Department of Computer Science and Engineering, University of California at San Diego, La Jolla, CA 92093, USA. E-mail: `mihir@cs.ucsd.edu`. Supported in part by NSF CAREER Award CCR-9624439 and a 1996 Packard Foundation Fellowship in Science and Engineering. Some of this work was done when the author was at IBM.

[‡] Department of Computer Science and Applied Mathematics, Weizmann Institute of Sciences, Rehovot, Israel. E-mail: `oded@wisdom.weizmann.ac.il`. Supported in part by grant No. 92-00226 from the US–Israel Binational Science Foundation (BSF), Jerusalem, Israel.

[§] Laboratory for Computer Science, MIT, 545 Technology Sq., Cambridge, MA 02139, USA. E-mail: `madhu@theory.lcs.mit.edu`. Some of this work was done when the author was at IBM.

Contents

1 Introduction

In the Max Clique problem we are given a graph G and must find the value of MaxClique$(G) = \max\{\, |S| \,:\, S \text{ is a clique in } G\,\}$. This is an example of an NP-optimization problem, of which others are to find the chromatic number of a graph; to find the size of the smallest vertex cover; etc. These problems arise in many settings, and efficient solutions are much desired. Unfortunately, many important NP-optimization problems (those mentioned above in particular) are NP-hard to solve. So algorithm designers seek efficient (polynomial time) approximation algorithms.

An approximation algorithm delivers a number that is supposed to be close to optimal. The quality of the algorithm is measured in terms of how close this number is to optimal. For example, if $\mu(N) \geq 1$ is a function of the number N of vertices in a graph G, then we say an algorithm A approximates Max Clique within μ, or is factor μ approximation algorithm, if MaxClique$(G)/\mu(N) \leq A(G) \leq$ MaxClique(G) for every graph G. (For a minimization problem like Chromatic Number, we require instead that ChromNum$(G) \leq A(G) \leq \mu(N) \cdot$ ChromNum(G) where ChromNum(G) is the chromatic number of G.)

The search for efficient approximation algorithms achieving good factors has met with varied success. For some problems, good approximation algorithms were found. For some important problems, including Max Clique and Chromatic Number, the best approximation algorithms found achieved factors only marginally better than the trivial factor of N. For others, like Minimum Vertex Cover, simple algorithms achieving reasonable factors were discovered quite quickly, but it was unclear whether one could do better. Algorithm designers want to know whether this is due to some inherent intractability, or only to the lack of cleverness in algorithm design.

Some early non-approximability results were able to indicate (at least for some problems) that very good approximation (ie. achieving factors very close to optimal) can be NP-hard. But the real breakthrough came more recently, when a strong hardness of approximation result for Max Clique was shown by establishing a connection between Max Clique and the existence of probabilistically checkable proof (PCP) systems for NP. Since then, similar connections have been found to other optimization problems. Meanwhile with the construction of more efficient proof systems, the factors within which approximation is shown hard continue to increase. Indeed, in some cases, even tight results seem in sight.

This paper continues the development of the connection between PCPs and hardness of approximation with the goal of getting tight results. On the one hand, we continue past work by building new proof systems and obtaining improved non-approximability results; on the other hand we open some new directions with an exploration of the limits of the PCP connection.

In what follows we provide a little background and then a high level overview of our results. The rich history of the ideas in this area is overviewed in Section 1.3, and more detailed histories are provided in the body of the paper.

1.1 Some background and definitions

We will be informal and as brief as possible; formal definitions can be found in Section 2.

Proof systems and parameters. A probabilistic proof system is described by a probabilistic, polynomial time *verifier* V. It takes an input x of length n and tosses coins R. It has oracle access to a poly(n) length string σ describing the proof: to access a bit it writes a $O(\log n)$ bit address and is returned the corresponding bit of the proof. Following its computation it will either accept or reject its input x. The accepting probability, denoted $\texttt{ACC}\,[\,V(x)\,]$, is the maximum, over all σ, of the probability (over R) that V accepts x on coins R and proof string σ. While the task is typically language recognition (namely to recognize whether x is in some fixed language L) we will, more generally, consider promise problems (A, B) consisting of a set A of "positive" instances and a set B of "negative" instances [ESY]. A languages L is identified with the promise problem $(L, \overline{L})$.

Of interest in the applications are various parameters of the system. The completeness probability $c = c(n)$ and the soundness probability $s = s(n)$ are defined in the usual ways. In case $c = 1$ we say that the system has *perfect completeness.* The gap is $g = c/s$. The query complexity is the maximum (over all coin tosses and proof strings) of the number of bits of the proof that are examined by the verifier. The free-bit complexity, roughly speaking, is the logarithm of number of possible accepting configurations of V on coins R and input x. (For example a verifier which makes 3 queries and accepts iff the parity of the answers is odd has 4 accepting configuration and thus free-bit complexity 2.)

Either the query or the free-bit complexity may be considered in amortized form: e.g. the amortized free-bit complexity is the free-bit complexity (of a proof system with perfect completeness) divided by the logarithm of the gap. (That is, the number of free-bits needed per factor of 2 increase in the gap.) Also, either the query or free-bit complexity may be considered on the average, the average being over the random string of the verifier.

Denote by $\mathrm{PCP}_{c,s}[r, q]$ the class of promise problems recognized by verifiers tossing r coins, having query complexity q, and achieving completeness probability c and soundness probability s. $\mathrm{FPCP}_{c,s}[r, f]$ is defined analogously with f being the free-bit complexity. $\overline{\mathrm{PCP}}[r, q]$ is defined analogously with q being the amortized query complexity, and $\overline{\mathrm{FPCP}}[r, f]$ is defined analogously with f the amortized free-bit complexity.

Max Clique approximation. Although we look at many optimization problems there is a particular focus on Max Clique. Recall the best known polynomial time approximation algorithm for Max Clique achieves a factor of only $N^{1-o(1)}$ [BoHa], scarcely better than the trivial factor of N. (Throughout the paper, when discussing the Max Clique problem, or any other problem about graphs, N denotes the number of vertices in the graph.) Does there exist an $N^{1-\epsilon}$ factor approximation algorithm for Max Clique for some $\epsilon < 1$? An additional motivation for searching for such "weak" approximation algorithms was suggested by Blum [Bl]. He showed that a polynomial-time $N^{1-\epsilon}$-factor approximation algorithm for Max Clique implies a polynomial time algorithm to color a three colorable graph with $O(\log N)$ colors [Bl], which is much better than currently known [KMS]. But perhaps $N^{1-o(1)}$ is the best possible. Resolving the approximation complexity of this basic problem seems, in any case, to be worth some effort.

Gaps in clique size. Hardness of approximation (say of Max Clique) is typically shown via the construction of promise problems with gaps in max clique size. Specifically, let $\mathsf{Gap}\text{-}\mathrm{MaxClique}_{c,s}$ be the promise problem (A, B) defined as follows: A is the set of all graphs G with $\mathrm{MaxClique}(G)/N \geq c(N)$, and B is the set of all graphs G with $\mathrm{MaxClique}(G)/N < s(N)$. The gap is defined as c/s. Now, a hardness result will typically specify a value of the gap $g(N) = c(N)/s(N)$ for which $\mathsf{Gap}\text{-}\mathrm{MaxClique}_{c,s}$ is NP-hard under a (randomized) Karp reduction. This means that there is no polynomial time algorithm to approximate the Max Clique size of an N node graph within $g(N)$ unless NP has randomized polynomial time algorithms. Gap problems can be similarly defined for all the other optimization problems we consider. From now on, we discuss approximation in terms of these gap problems.

The connection: Making gaps from proofs. The FGLSS-reduction [FGLSS] is a reduction of a promise problem (A, B) to $\mathsf{Gap}\text{-}\mathrm{MaxClique}_{c,s}$ for some appropriate c, s defined by the reduction. It works by using a verifier V of a pcp system for (A, B) to map any instance $x \in A \cup B$ to a graph G_x so that $\mathrm{MaxClique}(G_x)$ reflects $\mathtt{ACC}\,[\,V(x)\,]$. For the best results one typically uses a randomized form of this reduction due to [BeSc, Zuc] and it is this that we will assume henceforth.

A NP-hard gap problem is obtained roughly as follows. First, one exhibits an appropriate proof system for NP. Then one applies the FGLSS reduction. The factor indicated hard depends on the proof system parameters. A key element in getting better results has been the distilling of appropriate pcp-parameters. The sequence of works [FGLSS, ArSa, ALMSS, BGLR, FeKi1, BeSu] lead us through a sequence of parameters: query complexity, free-bit complexity and, finally, for the best known results, amortized free-bit complexity. The connection in terms of amortized free-bits can be stated as follows: if NP reduces to $\overline{\mathrm{FPCP}}[\log, f]$ then NP also reduces to $\mathsf{Gap}\text{-MaxClique}_{c,s}$, with gap $c(N)/s(N) = N^{1/(1+f)}$. (In both cases the reduction is via randomized Karp reductions, and terms of $\epsilon > 0$ which can be arbitrarily small are ignored.) In particular if NP $\subseteq \overline{\mathrm{FPCP}}[\log, f]$ then approximating the max clique size of an N vertex graph within $N^{1/(1+f)}$ in polynomial time is not possible unless NP has efficient randomized polynomial time algorithms.

1.2 Overview of our results

1.2.1 New proof systems and non-approximability results

This section describes the new proof systems that we construct and the non-approximability results that we derive from them.

New proof systems. We present several new ways of capturing NP via probabilistic proof systems, summarized below and in Figure 1:

(1) For every $\epsilon > 0$ it is the case that NP $\subseteq \overline{\mathrm{FPCP}}[\log, 2+\epsilon]$.

(2) NP $\subseteq \mathrm{PCP}_{1,1/2}[\log, 11]$.

(3) NP $\subseteq \mathrm{FPCP}_{1,s}[\log, 2]$ for $s = 0.794$.

(4) NP $\subseteq \mathrm{PCP}_{1,s}[\log, 3]$ for any $s > 0.85$.

Some of these results are motivated by applications, others purely as interesting items in proof theory.

The search for proof systems of low amortized free-bit complexity is motivated of course by the FGLSS reduction. Bellare and Sudan [BeSu] have shown that NP $\subseteq \overline{\mathrm{FPCP}}[\log, 3+\epsilon]$ for every $\epsilon > 0$. The first result above improves upon this, presenting a new proof system with amortized free-bit complexity $2+\epsilon$.

The question of how low one can get the (worst-case and average) query complexity required to attain soundness error 1/2 was investigated a lot in earlier works because they were applying the result to obtain Max Clique hardness results. We now know we can do better with amortized free-bit complexity. Nevertheless, the original question is still one to which we are curious to know the answer.

Minimizing the soundness error obtainable using only two (non-amortized!) free-bits is important for a more pragmatic reason. It enables us to get the first explicit and reasonably strong constant non-approximability result for the Min Vertex Cover problem. This application is discussed below.

Finally, the soundness achievable using only three query bits is natural to consider given the results on the Max 3SAT gap problem. Indeed, if there is an NP-hard Max 3SAT gap problem with certain

focus	error	queries	free-bits	previous related result
3 queries	0.85	3	2	error $\frac{72}{73}$ via MaxSAT [BeSu]
2 free-bits	0.794	$O(1)$	2	
error 1/2	$\frac{1}{2}$	11	7	32 queries (24 on average) [FeKi1]
amortized free-bits	$O(2^{-m})$	2^{3m}	$2m$	$3m$ free-bits [BeSu]

Figure 1: *New PCP Systems for NP, all with logarithmic randomness.*

Problem	**Approx**		**Non-Approx**		
	Factor	Due to	Our Factor	Previous Factor	Assumption
Max3SAT	1.258	[Yan, GoWi2, TSSW]	1.038	$1+\frac{1}{72}$ [BeSu]	$P \neq NP$
MaxE3SAT	$1+\frac{1}{7}$	folklore	$1+\frac{1}{26}$	unspecified [ALMSS]	$P \neq NP$
Max2SAT	1.075	[GoWi2, FeGo]	1.013	$1+\frac{1}{504}$ (implied [BeSu])	$P \neq NP$
Max$\oplus$SAT	2	folklore	$1+\frac{1}{7}$		$P \neq NP$
MaxCUT	1.139	[GoWi2]	1.014	unspecified [ALMSS]	$P \neq NP$
MinVC	$2-o(1)$	[BaEv2, MoSp]	$1+\frac{1}{15}$	unspecified [ALMSS]	$P \neq NP$
Max-Clique	$N^{1-o(1)}$	[BoHa]		$N^{\frac{1}{4}}$ [BeSu]	$NP \not\subseteq co\tilde{RP}$
			$N^{\frac{1}{3}}$	$N^{\frac{1}{5}}$	$coRP \neq NP$
			$N^{\frac{1}{4}}$	$N^{\frac{1}{6}}$ [BeSu]	$P \neq NP$
Chromatic Number	$N^{1-o(1)}$	[BoHa]		$N^{\frac{1}{10}}$ [BeSu]	$NP \not\subseteq co\tilde{RP}$
			$N^{\frac{1}{5}}$	$N^{\frac{1}{7}}$ [Fu]	$coRP \neq NP$
			$N^{\frac{1}{7}}$	$N^{\frac{1}{14}}$ [BeSu]	$P \neq NP$

Figure 2: *Approximation factors attainable by polynomial-time algorithms (Approx) versus factors we show are hard to achieve (Non-Approx). MaxE3SAT (resp., Max⊕SAT) denote the maximization problem for CNF formulae having exactly 3 different literals in each clause (resp., a conjunction of parity clauses).*

gap then one can easily get a three query proof system with the same gap. But in fact one can do better as indicated above.

New non-approximability results. The results are summarized in Figure 2. (In the last items, we ignore terms of N^ϵ where $\epsilon > 0$ is an arbitrarily small positive constant.) Refer to Section 2.4.2 for the definitions of the problems.

The conclusion for Max Clique follows, of course, from the FGLSS-reduction and the first proof system listed above. The conclusion for the Chromatic Number follows from a recent reduction of Fürer [Fu], which in turn builds on reductions in [LuYa, KLS, BeSu]. (Fürer's work and ours are contemporaneous and thus we view the $N^{1/5}$ hardness result as jointly due to both papers.)

The improvements for the MaxSNP problems are perhaps more significant than the improvements for the Max Clique problem: We see hardness results for MaxSNP problems that are comparable to the factors achieved by known polynomial time approximation algorithms.

We are obtaining the first explicit and reasonable non-approximability factor for Max2SAT, MaxCUT and minimum Vertex Cover. Recall that the latter is approximable within $2 - o(1)$ [BaEv2, MoSp]. Our results for MaxCUT and Max2SAT show that it is infeasible to find a solution with value which is only a factor of 1.01 from optimal. This may be contrasted with the recent results of [GoWi2, FeGo] which shows that solutions which are within 1.14 and 1.075, respectively, of the optimum are obtainable in polynomial time. Thus, even though we do not know if the "pcp approach" allows to get the best possible non-approximability results for these problems, we feel that the current results are not ridiculously far from the known upper bounds.

General framework. We emphasize a general framework for the derivation of strong non-approximability results for MaxSNP problems which results from our tests and proof systems. We use direct reductions from verifiers to the problems of interest. (This follows and extends [BGLR], prior to which results had used "generic" reductions, which did not take advantage of the nature of the tests performed by the verifier.) In particular, in our case it turns out that the verifier only performs two kinds of tests — (1) verify that $a+b+c=\sigma \pmod 2$; and (2) verify that $a+b_c=\sigma_c$, where a,b,b_0,b_1,c are answer bits obtained from the oracle and the σ's are fixed bits. By constructing local gadgets (i.e., one gadget per random coin toss sequence) to verify each of the verifier's tests, we achieve better non-approximability results than using more general reductions. In particular our work seems to suggest that optimizing for gadgets which "check" the two conditions listed above will lead to reasonably good lower bounds for many MaxSNP problems. In this way, obtaining a non-approximability result for a particular problem is reduced to the construction of appropriate "gadgets" to "represent" two simple functions.

Techniques. The main technical contribution is a new error-correcting code which we have called the "long code. This code encodes an n-bit string as a 2^{2^n} bit string which consists of the value of every boolean function on the n-bit string. It is easy to see such codes have large Hamming distance. We show that this code is also easily "testable" and "correctable", and derive the new proof systems based on this.

As in all recent constructions of efficient pcp's our construction also relies on the use of recursive construction of verifiers, introduced by Arora and Safra [ArSa]. We have the advantage of being able to use, at the outer level, the recent verifier of Raz [Raz], which was not available to previous authors. The inner level verifier relies on the use of a "good" encoding scheme. Beginning with [ALMSS], constructions of this verifier have used the Hadamard Code; in this paper we use instead the long code.

1.2.2 Proofs and approximation: Potential and limits

As the above indicates, non-approximability results are getting steadily stronger, especially for Max Clique. How far can they go? And, in minimizing amortized free-bits, are we on the right track? Are there other ways? The next set of results provides answers to these kinds of questions.

Reversing the connection: Making proofs from gaps. The FGLSS Reduction Lemma indicates that one route to good non-approximability results for Max Clique is to show NP $\subseteq \overline{\mathrm{FPCP}}[\log, f]$ for values of f which are as small as possible. We present a "reverse connection" which says that, in a sense, this is the *only* way to proceed. Namely, we "invert" the above FGLSS-reduction. Roughly, we show that, for any constant f, the following statements are equivalent:

(1) NP reduces to Gap-MaxClique$_{c,s}$ with gap $c(N)/s(N) = N^{1/(1+f)}$.

(2) NP reduces to $\overline{\mathrm{FPCP}}[\log, f]$.

The (2)⇒(1) direction is the FGLSS-reduction; The (1)⇒(2) direction is our reversed connection. (The statement ignores terms of $\epsilon > 0$ which can be arbitrarily small. The proof and a more precise statement are in Section 8.) In both cases the reduction is randomized. Furthermore the statement holds both for Karp and for Cook reductions. Also, if (1) holds with a deterministic Karp reduction then NP $\subseteq \overline{\mathrm{FPCP}}'[\log, f]$, where FPCP$'$ is defined as being the amortized free-bit complexity of proof systems with almost-perfect completeness (i.e., $c = 1 - o(1)$).

In other words *any* method of proving NP-hardness of Max Clique approximation to a factor of $N^{1/(1+f)}$ implies that NP has proof systems of amortized free-bit complexity f.

We stress both the "qualitative" and the "quantitative" aspects of this result. Qualitatively, it provides an answer to the following kind of a question: "What do proofs have to do with approximating clique size, and can we not prove non-approximability results without using proof checking?" The result indicates that proofs are inherent, and explains, perhaps, why hardness results avoiding the proof connection have not appeared.

However, at this stage it is the quantitative aspect that interests us more. It says that to get tighter results on Max Clique hardness, we must construct proof systems to minimize the amortized free-bit complexity. So our current efforts (recall that we have the amortized free-bit complexity down to two, yielding a $N^{1/3}$ hardness for Max Clique) are in the right direction. To prove that, say Max Clique is hard to approximate within $\sqrt{N}$, our reverse connection says we must construct proof systems with amortized free-bit complexity one.

Yet the reverse connection does more than guide our choice of parameters. It is also a useful conceptual tool since it allows us to go from graphs to proof systems and vice versa, in the process perhaps gaining some property. As an example we show how all known hardness results for chromatic number can be viewed (with almost no loss in efficiency) as reductions from Max Clique — even though these were essentially hardness results based on proof checking. Other examples demonstrating the usefulness of the equivalence may be found in Section 8.4. We believe that exploring and exploiting further this duality is a fruitful avenue to pursue.

A LOWER BOUND ON AMORTIZED FREE-BITS. Having shown that the minimization of amortized free-bits is unavoidable, we asked ourselves how low we can take them. Our approach here was to look at current techniques and assess their limitations. We stress that this approach makes various assumptions about methods, and is intended to show that significantly novel techniques are required to go further. But it does NOT suggest an *inherent* limitation.

We show that, under the framework used within this and previous papers on this subject, amortized free-bit complexity of 2 seems to be a natural barrier: any proof system in this framework must use $2-\epsilon$ amortized free-bits, where $\epsilon > 0$ as usual can be arbitrarily small. The result, including a definition of what we mean by the "framework," is in Section 9. Loosely speaking, it considers proof systems which, among other things, probe two oracles in order to check that one oracle is "close" to a codeword (i.e., a codeword test) and the second oracle encodes a projection of the information encoded in the first oracle (i.e., a projection test).

In retrospect, our lower bounds justify Håstad's two deviations from these techniques; specifically, his relaxation of the codeword test [H1] and his relaxation of the projection test [H2]. Specifically, Håstad [H1, H2] has constructed a pcp system (for NP) of amortized free-bit complexity ϵ, $\forall \epsilon > 0$. This was done in two stages/papers. In his first paper [H1], Håstad builds on the framework presented in the current work but introduces a relaxed codeword test which is conducted within amortized free-bit complexity ϵ. In his second paper [H2], Håstad abandons the current framework and utilizes a relaxed projection test which is conducted within amortized free-bit complexity ϵ. Our lower bounds justify Håstad's deviations from the intuitive but more stringent forms of the codeword and projection tests.

1.2.3 Properties and transforms of PCP and FPCP

Probabilistic proofs involve a vast arena of complexity parameters: query complexity, free-bit complexity, amortized free-bit complexity, randomness, and proof sizes to name a few. Some might, at first glance, seem less "natural" than others; yet all are important in applications. A better understanding of the basic properties and relations between these parameters would help move us forward.

We initiate, therefore, a systematic investigation of the properties of pcp complexity classes as a function of the parameter values. Besides providing new results we take the opportunity to state and prove a few folklore ones.

A contribution of this work is to distill and formalize the role of randomized reductions. These transforms provide an elegant and concise way to state connections between PCPs and approximability, or just between different kinds of proof systems, and make it easier to manipulate the many connections that exist to derive new results.

We begin with "triviality results," namely results which say that certain parameter combinations yield classes probably not capable of capturing NP.

For simplicity we restrict attention in this part to classes of languages, not classes of promise problems.

Triviality results. Perhaps the first thing to ask is whether, instead of amortized free-bit complexity, we could work with any of the simpler measures. After all $\overline{\mathrm{FPCP}}[\log, f]$ contains each of the following classes: (1) $\mathrm{PCP}_{1,1/2}[\log, f]$; (2) $\overline{\mathrm{PCP}}[\log, f]$; (3) $\mathrm{FPCP}_{1,1/2}[\log, f]$. Thus it would suffice to minimize the query complexity to get error 1/2; or the amortized query complexity; or the free-bit complexity to get error 1/2. However it turns out these complexities will not enable us to reach our target (of reducing the complexity to almost zero and thus proving that clique is hard to approximate to within a $N^{1-\epsilon}$ factor, for every $\epsilon > 0$). This is because the following classes are all contained in P:

(1) $\mathrm{PCP}_{1,1/2}[\,\log, 2\,]$

(2) $\overline{\mathrm{PCP}}[\,\log,\ 1\,]$

(3) $\mathrm{FPCP}_{1,1/2}[\log, 1]$.

Thus, we cannot expect to construct pcp systems for NP with either query complexity 2 (this is actually folklore predating our work); or amortized query complexity 1; or free-bit complexity 1. However it is a feature of amortized free-bit complexity that so far it seems entirely possible that NP reduces to $\overline{\mathrm{FPCP}}[\log, f]$ with f an arbitrarily small constant. Indeed, if we believe (conjecture) that Max Clique is hard to approximate with $N^{1-\epsilon}$ for any $\epsilon > 0$ then such proof systems must exist, by virtue of the equivalence stated above. In fact, even if we do not believe that Max Clique is hard to approximate with $N^{1-\epsilon}$ for any $\epsilon > 0$, it turns out that the amortized query bit parameter will be too weak to capture the hardness of the clique function. In fact, if Max Clique is hard to approximate to within N^{α}, then the best hardness result obtainable from the amortized query bit parameter would be of the form $N^{\frac{\alpha}{2-\alpha}}$. This is shown by invoking Corollary 10.11 which shows that the amortized query complexity parameter is always one unit larger than the amortized free-bit parameter (and we know that the amortized free bit parameter captures the hardness of Max Clique tightly).

Other results. We have already mentioned above that strict limitations on various query parameters make PCP very weak. Actually, for every $s < 1$, $\mathrm{PCP}_{1,s}[\,\log, 2\,]$ and $\mathrm{FPCP}_{1,s}[\log, 1]$ collapse to P. This means that pcp systems with perfect completeness are very weak when restricted to either *two queries* or to *free-bit complexity one*. However, pcp systems with completeness error and the very same query (resp., free-bit) bounds are not so weak. In particular, it is well known that $\mathrm{NP} = \mathrm{PCP}_{c,s}[\log, 2]$ for some $0 < s < c < 1$ (e.g., by using the NP-hardness of approximating Max2SAT). We show that $\mathrm{NP} = \mathrm{FPCP}_{c,s}[\log, 1]$ for some $0 < s < c < 1$ (specifically, $c = \frac{1}{2}$ and $s = 0.8 \cdot c$). Furthermore, for some smaller $0 < s < c < 1$, the following holds

$$\mathrm{NP} = \mathrm{FPCP}_{c,s}[\log, 0] \tag{1}$$

(specifically, with $c = \frac{1}{4}$ and $s = \frac{1}{5}$). We find the last assertion quite intriguing. It seems to indicate that one needs to be very careful when making conjectures regarding free-bit complexity. Furthermore, one has to be very careful also when making conjectures regarding amortized free-bit complexity; for example, the result $\mathrm{P} = \overline{\mathrm{PCP}}[\,\log,\ 1\,]$ holds also when one allows non-perfect completeness (in the definition of $\overline{\mathrm{PCP}}[\,\cdot,\ \cdot\,]$) as long as the gap is greater than 2^q per q queries, but an analogous result cannot hold for *two-sided error* amortized free-bit complexity (i.e., $\overline{\mathrm{FPCP}}[\,\cdot,\ \cdot\,]$).

Trying to understand the power of pcp systems with low free-bit complexity, we have waived the bound on the randomness complexity. Recall that in this case pcp systems are able to recognize non-deterministic exponential time (i.e., NEXPT = $\mathrm{PCP}_{1,1/2}[\mathrm{poly},\mathrm{poly}]$) [BFL]. Thus, it may be of interest to indicate that for every $s < 1$,

$$\mathrm{FPCP}_{1,s}[\mathrm{poly}, 0] \subseteq \mathrm{coNP} \quad (2)$$

$$\mathrm{FPCP}_{1,s}[\mathrm{poly}, 1] \subseteq \mathrm{PSPACE} \quad (3)$$

It seems that $\mathrm{FPCP}_{1,1/2}[\mathrm{poly}, 0]$ is not contained in BPP, since Quadratic Non-Residuosity and Graph Non-Isomorphism belong to the former class. (Specifically, the interactive proofs of [GMR] and [GMW] can be viewed as a pcp system with polynomial randomness, query complexity 1 and free-bit complexity 0.) Thus, it seems that the obvious observation $\mathrm{PCP}_{1,s}[\mathrm{poly}, 1] \subseteq \mathrm{AM}$ (for every $s < 1$, where AM stands for one round Arthur-Merlin games), would also be hard to improve upon.

Transformations between proof systems. We provide various useful transformation of pcp systems. These transformations are analogous to transformations that can be applied to graphs with respect to the Max Clique problem. In view of the relation (mentioned above), between FPCP and the gap-clique promise problem, this analogy is hardly surprising.

One type of transformation amplifies the gap (i.e., the ratio between completeness and soundness bounds) of the proof system while preserving its amortized free-bit complexity and incurring a relatively small additional cost in the randomness complexity. Specifically, using a randomized reduction we can transform $\mathrm{FPCP}_{1,\frac{1}{2}}[\log, f]$ into $\mathrm{FPCP}_{1,2^{-k}}[\log +k, k \cdot f]$ (ignoring multiplicative factors of $1+\epsilon$ for arbitrarily small $\epsilon > 0$). This transformation is analogous to the well-known transformation of Berman and Schnitger [BeSc]. Alternatively, using a known deterministic amplification method based on [AKS, LPS] one can transform $\mathrm{FPCP}_{1,\frac{1}{2}}[\log, f]$ into $\mathrm{FPCP}_{1,2^{-k}}[\log +2k, k \cdot f]$. (To the best of our knowledge this transformation has never appeared with a full proof.) Both alternatives are important ingredients in transforming pcp results into clique in-approximability results via the FGLSS method.

A second type of transformation moves the location of the gap (or, equivalently, the completeness parameter). The gap itself is preserved by the transformation but moving it is related to changing the free-bit complexity (and thus the amortized free-bit complexity is not preserved). Moving the gap 'up' requires increasing the free-bit complexity, whereas moving the gap 'down' allows to decrease the free-bit complexity. For example, we randomly reduce $\mathrm{FPCP}_{c,s}[\log, f]$ to $\mathrm{FPCP}_{1,s\cdot\log}[\log, f + \log(1/c) + \log\log]$. On the other hand, for every $k \leq f$, we (deterministically) reduce $\mathrm{FPCP}_{c,s}[\log, f]$ into $\mathrm{FPCP}_{\frac{c}{2^k},\frac{s}{2^k}}[\log, f-k]$, provided that the original system has at least 2^k accepting configurations per each possible sequence of coin-tosses. (This condition is satisfied in many natural pcp systems, even for $k = f$.)

1.3 History

Early work in non-approximability includes that of Garey and Johnson [GJ1] showing that it is NP-hard to approximate the chromatic factor within a factor less than two. The indication of higher factors, and results for other problems, had to wait for the interactive proof approach.

Interactive proofs were introduced by Goldwasser, Micali and Rackoff [GMR] and Babai [Bab]. Ben-Or, Goldwasser, Kilian and Wigderson [BGKW] extended these ideas to define a notion of multi-prover interactive proofs. Fortnow, Rompel and Sipser [FRS] showed that the class, MIP, of languages possessing multi-prover interactive proofs equals the class of languages which have (using todays terms) probabilistically checkable proofs (of unrestricted, and thus polynomial, randomness and query complexity).

First indication to the power of interactive proof systems was given in [GMW], where it was shown that interactive proofs exist for Graph Non-Isomorphism (whereas this language is not known to be in NP). However, the real breakthrough came with the result of Lund, Fortnow, Karloff and

Nisan [LFKN] who used algebraic methods to show that all coNP languages (and actually, all languages in $\mathrm{P}^{\#P}$) have interactive proof systems. These techniques were used by Shamir [Sha] to show that IP = PSPACE.

A central result that enabled the connection to hardness of approximation is that of Babai, Fortnow and Lund [BFL]. They showed that the class MIP equals the class NEXP (i.e., languages recognizable in non-deterministic exponential time). The latter result has been "scaled-down" to the NP-level by two independent groups of researchers. Babai, Fortnow, Lund and Szegedy [BFLS] showed that if the input is encoded using a special error-correcting code (for which encoding and decoding can be performed in polynomial-time) then NP has transparent proof systems (i.e., it is possible to verify the correctness of the proof in poly-logarithmic time). Feige, Goldwasser, Lovász, Safra and Szegedy [FGLSS] showed that NP has probabilistically checkable proofs of poly-logarithmic randomness and query complexity; namely, $\mathrm{NP} \subseteq \mathrm{PCP}_{1,1/2}[r, q]$, where $r(n) = q(n) = O(\log n \cdot \log\log n)$.

A hardness of approximation result based on interactive proofs was first proved by Condon [Con]. The breakthrough PCP connection to approximation was made by Feige, Goldwasser, Lovász, Safra and Szegedy [FGLSS]. They showed that $\mathrm{NP} \subseteq \mathrm{PCP}_{1,s}[r, q]$ implies that approximating the maximum clique in a $2^{r(n)+q(n)}$-vertices graph to within a $1/s(n)$ factor is infeasible (i.e., not doable in polynomial-time), provided that NP is not in $\mathrm{Dtime}(2^{O(r+q)})$. (Here n is the length of the input x to the pcp verifier.) Combined with the above-mentioned results, they obtained the first in a sequence of strong non-approximability results for Max Clique: a non-approximability factor of $2^{\log^{1-\epsilon} N}$, $\forall \epsilon > 0$, assuming NP does not have quasi-polynomial time algorithms.

After the work of [FGLSS] the field took off in two major directions. One was to extend the interactive proof approach to prove the non-approximability of other optimization problems. Direct reductions from proofs were used to show the hardness of quadratic programming [BeRo, FeLo], Max3SAT [ALMSS], set cover [LuYa], and other problems [Be]. The earlier work of Papadimitriou and Yannakakis introducing the class MaxSNP [PaYa] now came into play; by reduction from Max3SAT it implied hardness of approximation for any MaxSNP-hard problem. Also, reductions from Max Clique lead to hardness results for the chromatic number [LuYa] and other problems [Zuc].

The other direction was to increase factors, and reduce assumptions, for existing hardness of approximation results. This involves improving the efficiency of the underlying proof systems and/or the efficiency of the reductions.

The first stage of this enterprise started with the work of Arora and Safra [ArSa]. They showed that $\mathrm{NP} \subseteq \mathrm{PCP}_{1,1/2}[\log, o(\log)]$. This provided the first strong NP-hardness result for Max Clique (specifically, a hardness factor of $2^{\sqrt{\log N}}$). This work introduced the idea of recursive proof checking, which turned out to play a fundamental role in all subsequent developments. Interestingly, the idea of encoding inputs in an error-correcting form (as suggested in [BFLS]) is essential to make "recursion" work. Arora, Lund, Motwani, Sudan and Szegedy [ALMSS] reduced the query complexity of pcp systems for NP to a constant, while preserving the logarithmic randomness complexity; namely, they showed that $\mathrm{NP} = \mathrm{PCP}_{1,1/2}[\log, O(1)]$. This immediately implied the NP-hardness of approximating Max Clique within N^{ϵ}, for some $\epsilon > 0$. Furthermore, it also implied that Max-3-Sat is NP-hard to approximate to within some constant factor [ALMSS] and so is any MaxSNP-hard problem [PaYa].

The second stage of this enterprise started with the work of Bellare, Goldwasser, Lund and Russell [BGLR]. The goal was to improve (increase) the constant ϵ in the exponent of the hardness of approximation factor for Max Clique, and also to improve the constant values of the hardness factors in the MaxSNP hardness results. They presented new proof systems minimizing query complexity and exploited a slightly improved version of the FGLSS-reduction due to [BeSc, Zuc] to get a $N^{1/30}$ hardness of approximation factor for Max Clique. Feige and Kilian [FeKi1], however, observed that one should work with free-bits, and noted that the free-bit complexity of the system of [BGLR] was 14, yielding a $N^{1/15}$ hardness factor. Bellare and Sudan then suggested the notion of amortized free-bits.

Problem	**EASY to Approx. Factor**		**HARD to Approx. Factor**			**Tight?**
	Factor	Due to	Our Factor	NEW Factor	Assumption	
Max3SAT	$1+\frac{1}{7}+\epsilon$	[KaZw] (new)	$1+\frac{1}{26}$	$1+\frac{1}{7}-\epsilon$ [H3]	$\mathrm{P} \neq \mathrm{NP}$	Yes
MaxE3SAT	$1+\frac{1}{7}$	folklore	$1+\frac{1}{26}$	$1+\frac{1}{7}-\epsilon$ [H3]	$\mathrm{P} \neq \mathrm{NP}$	Yes
Max2SAT	1.075	[GoWi2, FeGo]	1.013	1.047 [H3]	$\mathrm{P} \neq \mathrm{NP}$	No
Max$\oplus$SAT	2	folklore	$1+\frac{1}{7}-\epsilon$	$2-\epsilon$ [H3]	$\mathrm{P} \neq \mathrm{NP}$	Yes
MaxCUT	1.139	[GoWi2]	1.014	1.062 [H3]	$\mathrm{P} \neq \mathrm{NP}$	No
MinVC	$2-o(1)$	[BaEv2, MoSp]	$1+\frac{1}{15}$	$1+\frac{1}{6}-\epsilon$ [H3]	$\mathrm{P} \neq \mathrm{NP}$	No
Max-Clique	$N^{1-o(1)}$	[BoHa]	$N^{\frac{1}{3}-\epsilon}$	$N^{1-\epsilon}$ [H2]	$\mathrm{coRP} \neq \mathrm{NP}$	Yes
			$N^{\frac{1}{4}-\epsilon}$	$N^{\frac{1}{2}-\epsilon}$ [H2]	$\mathrm{P} \neq \mathrm{NP}$	No
Chromatic No.	$N^{1-o(1)}$	[BoHa]	$N^{\frac{1}{5}-\epsilon}$	$N^{1-\epsilon}$ [FeKi2]	$\mathrm{coRP} \neq \mathrm{NP}$	Yes

Figure 3: *State of the art regarding easy and hard approximation factors (updated July 1997). Here* $\epsilon > 0$ *is an arbitrarily small constant.*

They constructed proof systems achieving amortized free-bit complexity three, and in thus obtained a $N^{1/4}$ hardness for Max Clique assuming $\mathrm{NP} \not\subseteq \mathrm{co}\widetilde{\mathrm{RP}}$.

Detailed histories for specific topics are given in Sections 2.2.3 and 2.4.3.

1.4 Related work

Following the presentation of our results, Arora has also investigated the limitations of proof checking techniques in proving non-approximability results [Ar]. Like in our free-bit lower bound result, he tries to assess the limitations of current techniques by making some assumptions about these techniques and then showing a lower bound. His focus is on the reductions, which he assumes are "code like." In this setting he can show that one should not expect to prove non-approximability of Max Clique within $N^{1/2}$. (The assumptions made by us and by Arora do not seem to be comparable: neither implies the other. In retrospect, both sets of assumptions could be by-passed – as done by Håstad [H1, H2].)

1.5 Subsequent work

Our prophecy by which the PCP approach is leading to tight non-approximability results is in the process of materializing (see Figure 3). By now, tight results are known for central problems such as Min-Set-Cover (cf., [LuYa, BGLR, Fe2]), Max-Clique (cf., [H1, H2]), Min-Coloring ([FeKi2]), and Max-3SAT (cf., [H3]). The latter results were obtained subsequently to the current work, and while building on it.

Amortized free-bits and Max-Clique. The most intriguing problem left open by our work has been resolved by Håstad [H1, H2]. He proved our conjecture (cf., [BGS2]) by which, for every $\epsilon > 0$, it is the case that $\mathrm{NP} \subseteq \overline{\mathrm{FPCP}}[\log, \epsilon]$. The Long-Code, introduced in this work, plays a pivotal role in Håstad's work. He also uses the idea of folding. Applying the FGLSS-reduction to the new proof system, Håstad obtains a tight result for Max-Clique by showing that for every $\epsilon > 0$, assuming

NP $\neq$ coRP, there is no polynomial time algorithm to approximate Max-Clique within a factor of $N^{1-\epsilon}$.

Improved 3-query proofs and Max-SNP. Another challenge, one we even did not dare state, was achieved as well: Håstad [H3] has recently obtained optimal non-approximability results to MaxSNP problems such as Max-E3-SAT. Furthermore, he has improved over all our non-approximability results for MaxSNP problems, obtaining non-approximability factors of 22/21 and 17/16 for Max-2-SAT and Max-CUT, respectively. Underlying these results is a new proof system for NP which yields NP $\subseteq \mathrm{PCP}_{1-\epsilon,0.5}[\log, 3]$, for any $\epsilon > 0$. In addition, Håstad [H3] shows that NP is contained in $\mathrm{PCP}_{1,0.75+\epsilon}[\log, 3]$ (and it follows that NP $\subseteq \mathrm{PCP}_{1,0.5}[\log, 9]$). The Long-Code plays a pivotal role in all these proof systems.

Improved 2-free-bits proofs and Min-VC. The above-mentioned proof system of Håstad [H3] uses two (non-amortized) free-bits, and so NP $\subseteq \mathrm{FPCP}_{1-\epsilon,0.5}[\log, 2]$, for every $\epsilon > 0$. This sets the non-approximability bound for Min Vertex-Cover at $\frac{7}{6} - \epsilon$.

Chromatic Number. Feige and Kilian [FeKi2] have introduced a new approach to showing hardness of approximability of ChromNum, based on a new measure of proof checking complexity called the covering complexity. By modifying our proof systems so as to preserve the amortized free-bit complexity and achieve low covering complexity, they proved a that approximating ChromNum within $N^{1/3}$ is hard unless NP = coRP. They were able to similarly modify Håstad's proof systems [H1, H2] and thereby improve the hard factor to $N^{1-\epsilon}$, for any $\epsilon > 0$.

Gadgets. Another research direction, suggested in early versions of this work [BGS2], was taken on by Trevisan et. al. [TSSW] who initiated a systematic study of the construction of gadgets. In particular, they showed that the gadgets we have used in our reductions to the MaxSAT problems were optimal, and constructed better (and optimal) gadgets for reduction to MaxCUT.

Weights. An important issue neglected in our treatment of MaxSNP problems is the issue of weights. For example, in our MaxSAT problems we have allowed the same clause to appear many times in the formula, which can be considered as allowing "small" weights. Certainly, one may want non-approximability results for the unweighted case (where one does not allow multiple occurrences of the same clause). This issue is treated in a subsequent paper by Crescenzi et. al. [CST]. Essentially, they show that the unweighted cases of all problems considered in our paper are as hard as the weighted cases.

1.6 Directions for further research

Although the most intriguing open problems suggested in previous versions of this work [BGS2] have been resolved, some very interesting problems remain. We mention a few.

2-free-bits proofs and Min-VC. As we show, NP $\subseteq \mathrm{FPCP}_{c,s}[\log, f]$ implies that approximating Min Vertex-Cover up to a $\frac{2^f - s}{2^f - c}$ factor is NP-hard. This motivates us to ask whether the following, increasingly stronger, conjectures hold.

(1) NP $\subseteq \mathrm{FPCP}_{1-\epsilon,\epsilon}[\log, 2]$ (or even NP $\subseteq \mathrm{FPCP}_{1,\epsilon}[\log, 2]$) for every $\epsilon > 0$. This would imply a hardness factor of $\frac{4}{3} - \epsilon$ for MinVC.

(2) For $f \stackrel{\mathrm{def}}{=} \log_2 3$, NP $\subseteq \mathrm{FPCP}_{1-\epsilon,\epsilon}[\log, f]$ (or even NP $\subseteq \mathrm{FPCP}_{1,\epsilon}[\log, f]$) for every $\epsilon > 0$. This would imply a hardness factor of $\frac{3}{2} - \epsilon$.

(3) NP $\subseteq \mathrm{FPCP}_{1-\epsilon,\epsilon}[\log, 1]$ for every $\epsilon > 0$. This would imply a hardness factor of $2 - \epsilon$.

Recall that $\mathrm{FPCP}_{1,s}[\log, 1] \subseteq \mathrm{P}$, for every $s < 1$, whereas $\mathrm{NP} \subseteq \mathrm{FPCP}_{1-\epsilon,0.5}[\log, 2]$ [H3]. It will be interesting (though of no application for MinVC) to know whether $\mathrm{NP} \subseteq \mathrm{FPCP}_{1,0.5+\epsilon}[\log, 2]$.

Perfect versus imperfect completeness. Håstad's work [H3] is indeed the trigger for the last question and similarly we wonder whether $\mathrm{NP} \subseteq \mathrm{PCP}_{1,0.5+\epsilon}[\log, 3]$. Non-perfect completeness seems to be useful in [H3], but it is to be seen if this is inherent. Similar issues arise with respect to some results in the current work (e.g., see our transformations for increasing acceptance probability of proof systems).

De-randomization. We know that $\overline{\mathrm{FPCP}}[\log, f]$ is *randomly* reducible to $\mathrm{FPCP}_{1,2^{-k}}[\log + (1+\epsilon)k, (1+\epsilon)k \cdot f]$. On the other hand, the former class is contained in (i.e., is *deterministically* reduced to) the class $\mathrm{FPCP}_{1,2^{-k}}[\log + (2+\epsilon)k, (1+\epsilon)k \cdot f]$, for arbitrarily small $\epsilon > 0$. Can one obtain the best of both worlds; namely, a deterministic reduction of $\overline{\mathrm{FPCP}}[\log, f]$ to, say, $\mathrm{FPCP}_{1,2^{-k}}[\log + (1+\epsilon)k, (1+\epsilon)k \cdot f]$, for arbitrarily small $\epsilon > 0$. An affirmative answer will allow us to infer from $\mathrm{NP} \subseteq \overline{\mathrm{FPCP}}[\log, f]$ that approximating Max Clique to within an $N^{\frac{1}{1+f+\epsilon}}$ factor is NP-hard (rather than NP-hard under randomized reductions).

One ingredient of our method for reversing the FGLSS-reduction is the randomized reduction of the class $\mathrm{FPCP}_{c,s}[\log, f]$ to the class $\mathrm{FPCP}_{1,\frac{\log}{c}\cdot s}[\log, f + \log(1/c) + \log\log]$. (This statement is proved using the ideas in Section 11. An alternative exposition, making use of a randomized graph-layering process, is given in Section 8.) Anyhow, randomness plays an essential role in obtaining a pcp system with perfect completeness.[1] The question is whether the class $\mathrm{FPCP}_{c,s}[\log, f]$ is contained in the class $\mathrm{FPCP}_{1,\frac{\log}{c}\cdot s}[\log, f + \log(1/c) + \log\log]$ (rather than being randomly reducible to it).

1.7 Previous versions of this paper

An extended abstract of this work appeared in the proceedings of the FOCS 95 conference [BGS1]. It was backed up by the first versions of this large manuscript [BGS2], posted on ECCC. The paper went through several revisions due to improvements and corrections in the results. These were regularly posted on ECCC (as revisions to [BGS2]). This is the fifth version of the work.

1.8 Organization

This introduction is followed by a section that contains definitions as well as detailed histories. The main content of the paper is divided into three parts:

(1) *Part 1: New proof systems and non-approximability results*, consisting of Sections 3 to 7, contains the materiel discussed in Section 1.2.1. See overview in Section 3.1.

(2) *Part 2: Proofs and approximation: Potential and limitations*, consisting of Sections 8 and 9, contains the materiel discussed in Section 1.2.2. Specifically, Section 8 contain the "reverse reduction" of Clique hardness to PCP, and Section 9 contains lower bounds on the free-bit complexity of certain tasks.

(3) *Part 3: PCP: Properties and Transformations*, consisting of Sections 10 and 11, contains the materiel discussed in Section 1.2.3. Specifically, Section 10 studies the expressive power of PCP systems with certain parameters, and Section 11 contains transformations among PCP classes.

[1]This makes our results more elegant, but actually – as indicated in Section 8, we could have settled for "almost perfect" completeness which suffices for presenting an inverse of the "FGLSS-reduction".

1.9 Acknowledgments

We thank Uri Feige, Johan Håstad, Viggo Kann, Marcos Kiwi and Luca Trevisan for carefully reading the previous versions of our work and pointing out several flaws and improvements. We thank two (anonymous) referees for their careful reading and many comments towards improving the presentation.

2 Definitions and histories

2.1 General notation and definitions

For integer n let $[n] = \{1, \ldots, n\}$. A graph always means an undirected graph with no self-loops, unless otherwise indicated. We let $\|G\|$ denote the number of vertices in graph $G = (V, E)$.

A probabilistic machine K has one or more inputs $x_1, x_2, \ldots$. It tosses coins to create a random string R, usually of some length $r(\cdot)$ which is a function of the (lengths of the) inputs. We let $K(x_1, x_2, \ldots; R)$ denote the output of K when it uses the random string R. Typically we are interested in the probability space associated to a random choice of R.

A function is *admissible* if it is polynomially bounded and polynomial-time computable. We will ask that all functions measuring complexity (e.g. the query complexity $q = q(n)$) be admissible.

In defining complexity classes we will consider promise problems rather than languages. (This convention is adopted since approximation problems are easily cast as promise problems.) Following Even et. al. [ESY], a promise problem is a pair of disjoint sets (A, B), the first being the set of "positive" instances and the second the set of "negative" instances. A language L is identified with $(L, \overline{L})$.

2.2 Proof systems

2.2.1 Basic Setting

A verifier is a probabilistic machine V taking one or more inputs and also allowed access to one or more oracles. Let x denote the sequence of all inputs to V and let n denote its length. During the course of its computation on coins R and input x, the verifier makes queries of its oracles. Its final decision to accept or reject is a function $\mathsf{DEC}_V(x, a; R)$ of x, R and the sequence a of all the bits obtained from the oracle in the computation. Contrary to standard terminology, acceptance in this paper will correspond to outputting 0 and rejection to outputting 1.

Oracles are formally functions, with the context specifying for each the domain and range. Sometimes, however, an algorithm will be given a string s as an oracle. (Giving a verifier s as an oracle models a "written proof" model in which someone has "written" s somewhere and the verifier wants to check it.) This is to be interpreted in the natural way; namely the oracle takes an index i and returns the i-th bit of s. Let π denote the sequence (tuple) of all proof oracles supplied to the verifier V. Now for verifier V examining the proofs π and having input x, we let

$$\mathtt{ACC}\,[\,V^{\pi}(x)\,] \;=\; \Pr_R\,[V^{\pi}(x; R) = 0]$$

denote the probability that V accepts when its random string is R. We then let

$$\mathtt{ACC}\,[\,V(x)\,] \;=\; \max_{\pi}\; \mathtt{ACC}\,[\,V^{\pi}(x)\,]\;.$$

This is the maximum accepting probability, over all possible choices of proof sequences π. (The domain from which the proofs are chosen depends, as mentioned above, on the context.)

Let $\mathsf{pattern}_V(x; R)$ be the set of all sequences a such that $\mathsf{DEC}_V(x, a; R) = 0$. (That is, all sequences of oracle answers leading to acceptance). A *generator* for V is a poly(n)-time computable function G

such that $\mathsf{pattern}_V(x;R) = G(x,R)$ for all x, R. (That is, it can efficiently generate the set of accepted patterns.)

2.2.2 Parameters

We are interested in a host of parameters that capture various complexity measures of the proof checking process. They are all functions of the length n of the input x given to the verifier V. In the following σ denotes the concatenation of all the proof strings given to the verifier. Also recall we are interested in proof systems for promise problems (A, B) rather than just for languages.

$\mathsf{coins} =$	Number of coins tossed by verifier. Typically denoted r.
$\mathsf{pflen} =$	Length of the proof provided to the verifier. Typically denoted l.
$c =$	Completeness probability. Namely $\min\{\, \mathtt{ACC}\,[\,V(x)\,] \,:\, x \in A \text{ and } \lvert x\rvert = n \,\}$.
$s =$	Soundness probability. Namely $\max\{\, \mathtt{ACC}\,[\,V(x)\,] \,:\, x \in B \text{ and } \lvert x\rvert = n \,\}$.
$g =$	Gap. Namely c/s.

Now we move to various measures of the "information" conveyed by the oracle to the verifier. For simplicity we consider here only oracles that return a single bit on each query; that is, they correspond to strings, or "written proofs."

$\mathsf{query} =$	The query complexity on input x is the maximum, over all possible random strings R of V, of the number of bits of σ accessed by V on input x. The query complexity of the system $q = q(n)$ is the maximum of this over all inputs $x \in A \cup B$ of length n.
$\mathsf{query}_{\rm av} =$	The average query bit complexity on input x is the average, over R, of the number of bits of the proof σ accessed by V on input x and coins R. The average query complexity of the system is the maximum of this over all $x \in A \cup B$ of length n. Typically denoted $q_{\rm av}$.
$\overline{\mathsf{query}} =$	V is said to have amortized query bit complexity $\bar{q}$ if $q/\lg(g) \le \bar{q}$ where q is the query bit complexity and g is the gap, and, furthermore, q is at most logarithmic in n.
$\mathsf{free} =$	The free bit complexity of V is f if there is a generator G such that $\lvert G(x,R)\rvert \le 2^f$ for all R and all $x \in A \cup B$ of length n.
$\mathsf{free}_{\rm av} =$	The average free bit complexity of V is $f_{\rm av}$ if there is a generator G such that $\mathbf{E}_R\,[\lvert G(x,R)\rvert] \le 2^{f_{\rm av}}$ for all $x \in A \cup B$ of length n.
$\overline{\mathsf{free}} =$	V is said to have amortized free bit complexity $\bar{f}$ if $f/\lg(g) \le \bar{f}$ where f is the free bit complexity and g is the gap.

Notice that amortized query complexity is restricted to be at most logarithmic. We don't need to explicitly make this restriction for the amortized free bit complexity since it is a consequence of the efficient generation condition.

In case the completeness parameter equals 1 (i.e., $c = 1$), we say that the system is of *perfect completeness.* In case the completeness parameter, c, satisfies $c(n) = 1 - o(1)$, we say that the system is of *almost-perfect completeness.*

The consideration of combinations of all these parameters give rise to a potentially vast number of different complexity classes. We will use a generic notation in which the parameter values are specified by name, except that, optionally, the completeness and soundness can, if they appear, do so as subscripts. Thus for example we have things like:

$$\mathrm{PCP}_{c,s}[\,\mathsf{coins} = r\,;\ \mathsf{query} = q\,;\ \mathsf{pflen} = 2^r\,;\ \mathsf{free} = f \ldots]\,.$$

However most often we'll work with the following abbreviations:

$$\begin{aligned}
\mathrm{PCP}_{c,s}[r,q] &\stackrel{\mathrm{def}}{=} \mathrm{PCP}_{c,s}[\,\mathsf{coins} = r\,;\ \mathsf{query} = q\,] \\
\overline{\mathrm{PCP}}_c[\,r,\ q\,] &\stackrel{\mathrm{def}}{=} \mathrm{PCP}_{c,\cdot}[\,\mathsf{coins} = r\,;\ \overline{\mathsf{query}} = q\,] \\
\mathrm{FPCP}_{c,s}[r,f] &\stackrel{\mathrm{def}}{=} \mathrm{PCP}_{c,s}[\,\mathsf{coins} = r\,;\ \mathsf{free} = f\,] \\
\mathrm{FPCP}_{c,s}[r,f,l] &\stackrel{\mathrm{def}}{=} \mathrm{PCP}_{c,s}[\,\mathsf{coins} = r\,;\ \mathsf{free} = f\,;\ \mathsf{pflen} = l\,] \\
\overline{\mathrm{FPCP}}_c[\,r,\ f\,] &\stackrel{\mathrm{def}}{=} \mathrm{PCP}_{c,\cdot}[\,\mathsf{coins} = r\,;\ \overline{\mathsf{free}} = f\,]\,.
\end{aligned}$$

We stress that in the definitions of the amortized classes, $\overline{\mathrm{PCP}}_c[\,r,\ q\,]$ and $\overline{\mathrm{FPCP}}_c[\,r,\ f\,]$, we refer to the completeness parameter c (but not to the soundness parameter). In case $c = 1$, we may omit this parameter and shorthand the amortized classes of perfect completeness by $\overline{\mathrm{PCP}}[\,r,\ q\,]$ and $\overline{\mathrm{FPCP}}[\,r,\ f\,]$, respectively. Namely,

$$\begin{aligned}
\overline{\mathrm{PCP}}[\,r,\ q\,] &\stackrel{\mathrm{def}}{=} \overline{\mathrm{PCP}}_1[\,r,\ q\,] \\
\overline{\mathrm{FPCP}}[\,r,\ f\,] &\stackrel{\mathrm{def}}{=} \overline{\mathrm{FPCP}}_1[\,r,\ f\,]
\end{aligned}$$

2.2.3 History of proof systems

Models and parameters. The model underlying what are now known as "probabilistically checkable proofs" is the "oracle model" of Fortnow, Rompel and Sipser [FRS], introduced as an equivalent version (with respect to language recognition power) of the multi-prover model of Ben-Or, Goldwasser, Kilian and Wigderson [BGKW]. Interestingly, as shown by [BFLS, FGLSS], this framework can be applied in a meaningful manner also to languages in NP. These works provide the verifier V with a "written" proof, modeled as an oracle to which V provides the "address" of a bit position in the proof string and is returned the corresponding bit of the proof. Babai et. al. [BFLS] suggested a model in which the inputs are encoded in a special (polynomial-time computable and decodable) error-correcting code and the verifier works in poly-logarithmic time. Here we follow the model of Feige et. al. [FGLSS] where the verifier is probabilistic polynomial-time (as usual) and one considers finer complexity measures such as the query and randomness complexity. The FGLSS-reduction (cf., [FGLSS]), stated in terms of the query complexity (number of binary queries), randomness complexity and error probability of the proof system, has focused attention on the above model and these parameters. The class $\mathrm{PCP}_{1,1/2}[r,q]$ was made explicit by [ArSa].

The parameterization was expanded by [BGLR] to explicitly consider the answer size (the oracle was allowed to return more than one bit at a time) and query size– their notation included five parameters: randomness, number of queries, size of each query, size of each answer, and error probability. They also similarly parameterized (single round) multi-prover proofs, drawing attention to the analogue with pcp. This served to focus attention on the roles of various parameters, both in reductions and in constructions. Also they introduced the consideration of average query complexity, the first in a sequence of parameter changes towards doing better for clique.

Free bits are implicit in [FeKi1] and formalized in [BeSu]. Amortized free bits are introduced in [BeSu] but formalized a little better here.

Due to	q	$q_{\rm av}$
[ALMSS]	some constant	some constant
[BGLR]	36	29
[FeKi1]	32	24
This paper	11	10.9

Figure 4: *Worst case (q) and average ($q_{\rm av}$) number of queries needed to get $1/2$ soundness with logarithmic randomness; that is, results of the form of Eq. (4).*

Proof sizes were considered in [BFLS, PoSp]. We consider them here for a different reason – they play an important role in that the randomized FGLSS reduction [BeSc, Zuc] depends actually on this parameter (rather than on the randomness complexity).

The discussion of previous proof systems is coupled with the discussion of Max Clique in Section 2.4.3. We conclude the current section, by discussing two somewhat related topics: query minimization and constant-prover proof systems.

Query complexity minimization. One seeks results of the form

$$\text{NP} \;=\; \text{PCP}_{1,1/2}[\,\mathsf{coins} = \log\,;\;\mathsf{query} = q\,;\;\mathsf{query}_{\rm av} = q_{\rm av}\,]\;. \qquad (4)$$

This was originally done for deriving NP-hardness results for the approximation of MaxClique, but subsequent work has indicated that other parameters actually govern this application. Still, the query complexity of a proof system remains a most natural measure, and it is an intriguing question as to how many bits of a proof you need to look at to detect an error with a given probability. Specifically, we consider the question of determining the values of $q, q_{\rm av}$ for which Eq. (4) holds.

The fundamental result of [ALMSS] states that $q, q_{\rm av}$ may be constants (independent of the input length). Reductions in the values of these constants were obtained since then and are depicted in Figure 4. See Section 6 for our results.

Role of constant-prover proofs in PCP – perspective. Constant-prover proofs have been instrumental in the derivation of non-approximability results in several ways. One of these is that they are a good starting point for reductions— examples of such are reductions of two-prover proofs to quadratic programming [BeRo, FeLo] and set cover [LuYa]. However, it is a different aspect of constant prover proofs that is of more direct concern to us. This aspect is the use of constant-prover proof systems as the penultimate step of the recursion, and begins with [ALMSS]. It is instrumental in getting PCP systems with only a constant number of queries. Their construction requires that these proof systems have low complexity: error which is any constant, and randomness and answer sizes that are preferably logarithmic. The number of provers and the randomness and query complexity determine the quality of many non-approximability results (e.g., poly-logarithmic rather than logarithmic complexities translate into non-approximability results using assumptions about quasi-polynomial time classes rather than polynomial time ones). The available constant-prover proof systems appear in Figure 5 and are discussed below. Throughout this discussion we consider proof systems obtaining an *arbitrary small constant* error probability.

The two-prover proofs of Lapidot-Shamir and Feige-Lovász [LaSh, FeLo] had poly-logarithmic randomness and answer sizes, so [ALMSS] used a modification of these, in the process increasing the

Due to	**Provers**	**Coins**	**Answer size**	**Canonical?**	**Can be made canonical?**
[LaSh, FeLo]	2	polylog	polylog	No	Yes [BeSu]
[ALMSS]	poly(ϵ^{-1})	log	polylog	No	?
[BGLR]	4	log	polyloglog	No	?
[Tar]	3	log	$O(1)$	No	?
[FeKi1]	2	log	$O(1)$	No	At cost of one more prover [BeSu]
[Raz]	2	log	$O(1)$	Yes	(NA)

Figure 5: *Constant prover PCPs achieving error which is a fixed, but arbitrarily small, constant ϵ.*

number of provers to a constant much larger than two. The later constructions of few-prover proofs of [BGLR, Tar, FeKi1] lead to better non-approximability results.

Bellare and Sudan [BeSu] identified some extra features of constant prover proofs whose presence they showed could be exploited to further increase the non-approximability factors. These features are captured in their definition of canonical verifiers (cf. Section 3.4). But the proof systems of [FeKi1] that had worked above no longer sufficed— they are not canonical. So instead [BeSu] used (a slight modification of) the proofs of [LaSh, FeLo], thereby incurring poly-logarithmic randomness and answer sizes, so that the assumptions in their non-approximability results pertain to quasi-polynomial time classes. (Alternatively they modify the [FeKi1] system to a canonical three-prover one, but then incur a decrease in the non-approximability factors due to having more provers).

A breakthrough result in this area is Raz's Parallel Repetition Theorem which implies the existence of a two-provers proof system with logarithmic randomness and constant answer size [Raz]. Furthermore, this proof system is canonical.

2.3 Reductions between problems and classes

We will consider reductions between promise problems. A deterministic Karp reduction from (A_1, B_1) to (A_2, B_2) is a polynomial time function T which for all x satisfies: if $x \in A_1$ then $T(x) \in A_2$ and if $x \in B_1$ then $T(x) \in B_2$. A randomized Karp reduction from (A_1, B_1) to (A_2, B_2) is a probabilistic, polynomial time function T which takes two arguments: an input x and a security parameter k, the latter written in unary. The transformation is required to have the property that

$$\begin{aligned} x \in A_1 \quad &\Longrightarrow \quad \Pr\left[\, T(x, 1^k) \in A_2 \,\right] \stackrel{\text{def}}{=} p_1(x,k) \geq 1 - 2^{-k} \\ x \in B_1 \quad &\Longrightarrow \quad \Pr\left[\, T(x, 1^k) \in B_2 \,\right] \stackrel{\text{def}}{=} p_2(x,k) \geq 1 - 2^{-k} \ . \end{aligned}$$

The probability is over the coin tosses of T. We say the reduction has perfect completeness if $p_1 = 1$ and perfect soundness if $p_2 = 1$. Notice a deterministic reduction corresponds to a randomized one in which $p_1 = p_2 = 1$. We write $(A_1, B_1) \leq_R^K (A_2, B_2)$ if there is a randomized Karp reduction from (A_1, B_1) to (A_2, B_2). If the reduction is deterministic we omit the subscript of "R," or, sometimes, for emphasis, replace it by a subscript of "D."

An example is the randomized FGLSS transformation [FGLSS, BeSc, Zuc]. Here (A_1, B_1) is typically an NP-complete language L, and (A_2, B_2) is Gap-MaxClique$_{c,s}$ for some c, s which are determined by the transformation. (See Section 2.4 for definition of latter.) This transformation has perfect soundness, while, on the other hand, it is possible to get $p_1 = 1 - 2^{-\text{poly}(n)}$.

Similarly one can define (randomized) Cook reductions. The notation for these reductions is $\leq_R^C$.

Let C be a complexity class (e.g. NP). We say that C reduces to (A_2, B_2) if for every (A_1, B_1) in C it is the case that (A_1, B_1) reduces to (A_2, B_2). An example is to say that NP reduces to $\mathsf{Gap}\text{-MaxClique}_{c,s}$. We say that C_1 reduces to C_2, where C_1 and C_2 are complexity classes, if for every (A_1, B_1) in C_1 there is an (A_2, B_2) in C_2 such that (A_1, B_1) reduces to (A_2, B_2). An example is to say that NP reduces to $\overline{\mathrm{FPCP}}[\log, f\,]$. The notation of $\leq_R^K$ or $\leq_R^C$ extends to these cases as well.

Notice that our definition of reducibility ensures that this relation is transitive.

For simplicity we sometimes view a randomized reduction T as a function only of x, and write $T(x)$. In such a case it is to be understood that the security parameter has been set to some convenient value, such as $k = 2$.

Historical Note. We've followed the common tradition regarding the names of polynomial-time reductions: many-to-one reductions are called Karp-reductions whereas (polynomial-time) Turing reductions are called Cook-reductions. This terminology is somewhat unfair towards Levin whose work on NP-completeness [Lev] was independent of those of Cook [Coo] and Karp [Kar]. Actually, the reductions considered by Levin are more restricted as they also efficiently transform the corresponding NP-witnesses (this is an artifact of Levin's desire to treat search problems rather than decision problems). In fact, such reductions (not surprisingly termed Levin-reductions) are essential for results such as Corollary 8.15. (Yet, this is the only example in the current paper.)

2.4 Approximation problems and quality

We discuss optimization problems, approximation algorithms for them, and how hardness is shown via the production of "hard gaps." We then list all the problems considered in this paper.

2.4.1 Optimization problems, approximation and gaps

Optimization problems. An *optimization problem* $\Phi = (S, g, \|\cdot\|, \|\cdot\|^*, \mathrm{opt})$ is specified by:

- A function S associating to any *instance* w a *solution set* $S(w) \neq \emptyset$.
- An *objective function* g associating to any instance w and solution $y \in S(w)$ a non-negative real number $g(w, y)$. This number is sometimes called the *value* of solution y.
- Two *norm* functions $\|\cdot\|, \|\cdot\|^*$, the first admissible, the second polynomial time computable, each associating to any instance w a non-negative real number; their roles will be explained later.
- An indication $\mathrm{opt} \in \{\min, \max\}$ of the *type* of optimization, whether maximization or minimization.

The task, given w, is to either maximize (this if opt = max) or minimize (this if opt = min), the value $g(w, y)$, over all $y \in S(w)$.

Definition 2.1 Let $\Phi = (S, g, \|\cdot\|, \|\cdot\|^*, \mathrm{opt})$ be an optimization problem. The *optimum* value for instance w is denoted $\Phi(w)$ and defined by

$$\Phi(w) \;=\; \begin{cases} \max_{y \in S(w)} g(w, y) & \text{if } \Phi \text{ is a maximization problem} \\ \min_{y \in S(w)} g(w, y) & \text{if } \Phi \text{ is a minimization problem.} \end{cases}$$

We sometimes consider the *normalized optimum,* defined by $\overline{\Phi}(w) = \Phi(w)/\|w\|^*$.

The above definition illustrates the role of the second norm: it is to normalize the optimum. Thus $\|\cdot\|^*$ will usually be chosen to make $0 \leq \overline{\Phi}(w) \leq 1$, depending on the problem.

Approximation. An *approximation algorithm* for $\Phi = (S, g, \|\cdot\|, \|\cdot\|^*, \mathrm{opt})$ is an algorithm A which on input w tries to output a number as close as possible to $\Phi(w)$. Unless otherwise indicated, an approximation algorithm runs in time polynomial in the length of w.

While the complexity of the algorithm is measured as a function of the length of the input, the approximation quality is often measured as a function of some other measure associated to the input. This is what we have called the first norm of w and denoted $\|w\|$. For example, for graph problems the first norm is typically the number of vertices in the graph.

The notion of an approximation algorithm achieving a certain approximation factor is different depending on whether it is a maximization problem or a minimization problem.

Definition 2.2 An approximation algorithm A for optimization problem $\Phi = (S, g, \|\cdot\|, \|\cdot\|^*, \mathrm{opt})$ is said to achieve a factor $\mu(\cdot) \geq 1$ if for all instances w its output $A(w)$ satisfies

- $\dfrac{\Phi(w)}{\mu(\|w\|)} \leq A(w) \leq \Phi(w)$ if Φ is a maximization problem, or
- $\Phi(w) \leq A(w) \leq \mu(\|w\|) \cdot \Phi(w)$ if Φ is a minimization problem.

Note that as per this definition, our convention is that an approximation factor is always a number at least one. In some other places, the approximation factor, at least in the case of minimization problems, is a number less than one: they set it to the reciprocal of what we set it.

Gap problems. We are interested in instances of an optimization problem for which the optimum is promised to be either "very high" or "very low." We capture this by associating to any optimization problem a promise problem, depending on a pair of "thresholds" c, s, both admissible functions of the first norm and satisfying $0 < s(\cdot) < c(\cdot)$. It is convenient to make the definition in terms of the normalized optimum rather than the optimum. We consider maximization and minimization problems separately.

Definition 2.3 Let $\Phi = (S, g, \|\cdot\|, \|\cdot\|^*, \max)$ be a maximization problem, and let $0 < s(\cdot) < c(\cdot)$ be admissible functions of the first norm. Define

$$\begin{aligned} Y &= \{\, w : \overline{\Phi}(w) \geq c(\|w\|) \,\} \\ N &= \{\, w : \overline{\Phi}(w) < s(\|w\|) \,\} \\ \textsf{Gap-}\Phi_{c,s} &= (Y, N) \ . \end{aligned}$$

The *gap* of the promise problem is defined to be c/s.

It is important that the inequality in the definitions of Y, N is strict in one case and not in the other. The same is true below although the order is reversed.

Definition 2.4 Let $\Phi = (S, g, \|\cdot\|, \|\cdot\|^*, \min)$ be a minimization problem, and let $0 < s(\cdot) < c(\cdot)$ be admissible functions of the first norm. Define

$$\begin{aligned} Y &= \{\, w : \overline{\Phi}(w) \leq s(\|w\|) \,\} \\ N &= \{\, w : \overline{\Phi}(w) > c(\|w\|) \,\} \\ \textsf{Gap-}\Phi_{c,s} &= (Y, N) \ . \end{aligned}$$

The *gap* of the promise problem is defined to be c/s.

In this way, each of the many optimization problems we consider will give rise to a gap problem.

SHOWING NON-APPROXIMABILITY VIA GAPS. Hardness of approximation of some optimization problem is shown by reducing NP to $\mathsf{Gap}\text{-}\Phi_{c,s}$ via a (possibly randomized) Karp reduction. (This is called producing "hard gaps.") The following proposition says that we can show Φ is hard to approximate within a factor equal to the gap by showing NP $\leq^K$ $\mathsf{Gap}\text{-}\Phi_{c,s}$, and the assumption under which the non-approximability result holds depends on the type of reduction. It is this proposition that motivates the consideration of gap problems.

Proposition 2.5 Optimization problem Φ has no factor c/s approximation algorithm

Under this assumption:	If this is true:
P $\neq$ NP	NP $\leq^K_D$ $\mathsf{Gap}\text{-}\Phi_{c,s}$
NP $\neq$ coRP	NP $\leq^K_R$ $\mathsf{Gap}\text{-}\Phi_{c,s}$ via a reduction with perfect completeness
NP $\not\subseteq$ BPP	NP $\leq^K_R$ $\mathsf{Gap}\text{-}\Phi_{c,s}$

Proof: Let us illustrate by proving the first of the three claims under the assumption that the problem is one of maximization. We proceed by contradiction. Given a (polynomial time) algorithm A that achieves an approximation factor of $\mu = c/s$ for Φ, we present a polynomial time algorithm B to decide L, where L is any language in NP. Let T be a (deterministic, Karp) reduction of L to $\mathsf{Gap}\text{-}\Phi_{c,s}$. On input x our algorithm B computes $w = T(x)$. Next it computes $\alpha = s(\|w\|) \cdot \|w\|^*$. Finally, B invokes A on w, outputs 1 (indicating $x \in L$) if $A(w) \geq \alpha$ and 0 otherwise (indicating $x \notin L$).

Since A runs in polynomail-time and the functions $s, \|\cdot\|, \|\cdot\|^*$ are polynomial time computable (by assumption), the algorithm B runs in polynomial time. We claim that B is always right. To see this, first let Y, N be the two parts of the promise problem $\mathsf{Gap}\text{-}\Phi_{c,s}$ as per Definition 2.3. Now consider two cases.

First suppose $x \in L$. Then $w = T(x) \in Y$ because T is a correct reduction of L to $\mathsf{Gap}\text{-}\Phi_{c,s}$. So $\overline{\Phi}(w) \geq c(\|w\|)$ by Definition 2.3. But then, starting from Definition 2.2 and simplifying, we have

$$A(w) \;\geq\; \frac{\Phi(w)}{c(\|w\|)/s(\|w\|)} \;=\; \frac{\overline{\Phi}(w)\cdot\|w\|^*}{c(\|w\|)/s(\|w\|)} \geq \frac{c(\|w\|)\cdot\|w\|^*}{c(\|w\|)/s(\|w\|)} \;=\; s(\|w\|)\cdot\|w\|^* \;=\; \alpha\,.$$

Thus, B will output 1, as desired.

Now suppose $x \notin L$. Then $w = T(x) \in N$. So $\overline{\Phi}(w) < s(\|w\|)$ by Definition 2.3. Starting from Definition 2.2 and simplifying, we have

$$A(w) \;\leq\; \Phi(w) \;=\; \overline{\Phi}(w)\cdot\|w\|^* \;<\; s(\|w\|)\cdot\|w\|^* \;=\; \alpha\,.$$

Thus B will output 0 as desired.

The proofs for the other cases are similar (and thus omitted). ▌

2.4.2 Some optimization problems we consider

A *formula* is a set of clauses (i.e., or-clauses) over some set of literals. We consider various classes of formulae. In particular, 3-SAT formulae (at most three literals in each clause), E3-SAT formulae (exactly three different literals in each clause) and 2-SAT formulae (at most two literals in each clause). We use the generic notation X-SAT to stand for some specified class of formulae; thus the above correspond to X $\in \{3, \mathrm{E3}, 2\}$. To each value of X we associate an optimization problem:

Problem: MaxXSAT
Instance: X-SAT formula φ
Solutions: An assignment v which associates to any variable x of φ a boolean value $v(x) \in \{0, 1\}$. (Not necessarily a satisfying assignment!)
Objective Function: The value of an assignment v is the number of clauses in φ that v makes true
Norm: The norm $\|\varphi\|$ of formula φ is the number of clauses in it, and $\|\varphi\|^*$ is the same
Type: Maximization

In particular we have optimization problems Max2SAT, Max3SAT, MaxE3SAT, and their corresponding gap problems.

Problem: MaxLinEq
Instance: A set of linear equations over GF(2)
Solutions: An assignment v which associates to any variable x in the set of equations a value $v(x) \in$ GF(2)
Objective Function: The value of an assignment v is the number of equations that v satisfies
Norm: Both norms are set to the number of equations in the instance
Type: Maximization

Problem: MaxCUT
Instance: G, w, where $G = (V, E)$ is a graph and $w \colon E \to \mathcal{R}^+$ is a weight function
Solutions: A *cut* $S, \overline{S}$ in G, meaning a partition $V = S \cup \overline{S}$ of V into disjoint sets
Objective Function: The value of a cut is its weight $w(S, \overline{S})$, the sum of the weights of the edges with one endpoint in S and the other in $\overline{S}$.
Norm: $\|G, w\| = |V|$ and $\|G, w\|^* = \sum_{e \in E} w(e)$
Type: Maximization

Problem: MinVC
Instance: Graph $G = (V, E)$
Solutions: A *vertex cover* in G, meaning a set $V' \subseteq V$ such that $V' \cap \{u, v\} \neq \emptyset$ for every $\{u, v\} \in E$.
Objective Function: The value of vertex cover V' is it size, meaning the number of vertices in it
Norm: $\|G\| = \|G\|^* = |V|$ is the number of vertices in the graph
Type: Minimization

Similarly for any integer B we can define MinVC-B, the version of MinVC in which the instance is a graph of degree B.

Problem: MaxClique
Instance: Graph $G = (V, E)$
Solutions: A *clique* in G, meaning a set $C \subseteq V$ such that $\{u, v\} \in E$ for every pair u, v of distinct vertices in C
Objective Function: The value of clique C is its size, meaning the number of vertices in it
Norm: $\|G\| = \|G\|^* = |V|$ is the number of vertices in the graph
Type: Maximization

Problem: ChromNum
Instance: Graph $G = (V, E)$
Solutions: A *coloring* of G, meaning a map $c \colon V \to \{1, \ldots, k\}$, for some k, such that $c(u) \neq c(v)$ for any $\{u, v\} \in E$
Objective Function: The value of c is the number k of colors it uses

Norm: $\|G\| = \|G\|^* = |V|$ is the number of vertices in the graph
Type: Minimization

As per our general notation, $\Phi(w)$ is the optimum for instance w of optimization problem Φ. Given the above, this means, for example, that MaxClique(G) is the maximum clique size in graph G; ChromNum(G) is the chromatic number of G; MinVC(G) is the minimum vertex cover size of G; etc. The corresponding normalized versions get bars overhead. We will use these notations in what follows.

By the above, MaxXSAT(φ) is the maximum number of simultaneously satisfiable clauses in φ. We abuse notation slightly by dropping the "X", writing just MaxSAT(φ). Similarly for the normalized version.

2.4.3 History of approximability results for these problems

Satisfiability problems. Max3SAT is the canonical Max-SNP complete problem [PaYa]. A polynomial-time algorithm due to Yannakakis [Yan] approximates it to within a factor of $4/3 < 1.334$ (see Goemans and Williamson [GoWi1] for an alternate algorithm). Currently the best known polynomial-time algorithm for Max3SAT achieves a factor of 1.258 (and is due to Trevisan et. al. [TSSW] which in turn build on Goemans and Williamson [GoWi2]). For MaxE3SAT, which is also Max-SNP complete, a very simple algorithm achieves an approximation of $8/7 \leq 1.143$ (where $7/8$ is the expected fraction of clauses satisfied by a uniformly chosen assignment).

Max2SAT is also Max-SNP complete [GJS, PaYa]. This problem is particularly interesting because it has been the focus of recent improvements in the approximation factor attainable in polynomial-time. Specifically, Goemans and Williamson [GoWi2] exhibited a polynomial time algorithm achieving an approximation factor of $\frac{1}{0.878} \approx 1.139$, and subsequently Feige and Goemans [FeGo] exhibited an algorithm achieving $\frac{1}{0.931} \approx 1.074$.

Non-approximability results for Max-SNP problems begin with [ALMSS] who proved that there exists a constant $\epsilon > 0$ such that $\mathsf{Gap\text{-}3SAT}_{1,1-\epsilon}$ is NP-hard. They did this by providing a reduction from a given NP language L to the promise problem in question, constructed by encoding as a 3-SAT instance the computation of a $\mathrm{PCP}_{1,1/2}[\,\log, O(1)\,]$ verifier for an NP-complete language, the variables in the instance corresponding to bits in the proof string. The basic paradigm of their reduction has been maintained in later improvements.

Figure 6 depicts the progress. Improvements (in the constant value of the non-approximability factor) begin with [BGLR]. They used Hadamard code based inner verifiers following [ALMSS]. They also introduced a framework for better analysis, and improved some previous analyses; we exploit in particular their better analyses of linearity testing (cf. Section 3.5) and of Freivalds's matrix multiplication test (cf. Lemma 3.16). The improvement of Feige and Kilian [FeKi1] was obtained via new proof systems; that of [BeSu] by use of the canonicity property of constant prover proofs and some optimizations. (See Section 2.2.3 for a discussion of the role of constant-prover proofs in this context).

Garey, Johnson and Stockmeyer [GJS] had provided, as early as 1976, a reduction of Max3SAT to Max2SAT which showed that if the former is non-approximable within $(k+1)/k$ then the latter is non-approximable within $(7k+1)/(7k)$. With the best previous non-approximability factor for Max3SAT (namely 66/65) we would only get a 456/455 factor non-approximability for Max2SAT. In fact, even using our new Max3SAT result we would get a hardness factor of only 185/184. See Section 4.2 for our results.

Linear equations. The MaxLinEq problem is known to be Max-SNP complete (see [BrNa] or [Pet]).

We remark that the problem of *maximizing* the number of satisfiable equations should not be confused with the "complementary" problem of *minimizing* the number of violated constraints, inves-

Due to	**Assuming**	**Factor**	**Technique**
[ALMSS]	P $\neq$ NP	some constant	NP $\subseteq$ PCP$_{1,1/2}$[log, $O(1)$]; Reduction of this to Max3SAT.
[BGLR]	$\widetilde{\text{P}} \neq \widetilde{\text{NP}}$	94/93	Framework; better analyses; uses proof systems of [LaSh, FeLo].
[BGLR]	P $\neq$ NP	113/112	New four-prover proof systems.
[FeKi1]	P $\neq$ NP	94/93	New two-prover proof systems.
[BeSu]	$\widetilde{\text{P}} \neq \widetilde{\text{NP}}$	66/65	Canonicity and some optimizations.
[BeSu]	P $\neq$ NP	73/72	Canonicity and some optimizations.
This paper	P $\neq$ NP	27/26	Long code and new proof systems.

Figure 6: *Non-approximability results for Max3SAT indicating the factor shown hard and the assumption under which this was done.*

tigated by Arora et. al. [ABSS]. Also the case of maximum satisfiable linear constraints over larger fields (of size q) has been considered by Amaldi and Kann [AmKa], who show that this problem is hard to approximate to within a factor of q^ϵ for some universal $\epsilon > 0$. See Section 4.2.2 for our results.

Max Cut. In 1976, Sahni and Gonzales [SaGo] gave a simple 2-approximation algorithm for this problem. Recently, in a breakthrough result, Goemans and Williamson [GoWi2] gave a new algorithm which achieves a ratio of $\frac{1}{0.878} = 1.139$ for this problem. On the other hand, [PaYa] give an approximation preserving reduction from Max3SAT to MaxCUT. Combined with [ALMSS] this shows that there exists a constant $\alpha > 1$ such that approximating MaxCUT within a factor of α is NP-hard. No explicit bounds were given since and even using the best known hardness results for MAX 3SAT, one suspects that the bound for MaxCUT would not be very large, since the reduction uses constructions of constant degree expanders etc. See Section 4.3 for our results.

Vertex cover. There is a simple polynomial time algorithm to approximate MinVC in unweighted graphs within a factor of 2. The algorithm, due to F. Gavril (cf. [GJ2]), consists of taking all vertices which appear in a maximal matching of the graph. For weighted graphs, Bar-Yehuda and Even [BaEv1] and Hochbaum [Hoc], gave algorithms achieving the same approximation factor. The best known algorithm today achieves a factor only slightly better, namely $2 - (\log\log|V|)/(2\log|V|)$ [BaEv2, MoSp].

Evidence to the hardness of approximating MinVC was given by Bar-Yehuda and Moran who showed that, for every $k \geq 2$ and $\epsilon > 0$, a $1 + \frac{1}{k} - \epsilon$ approximator for (finding) a minimum vertex cover would yield an algorithm for coloring $(k+1)$-colorable graphs using only logarithmically many colors [BaMo]. The version of MinVC in which one restricts attention to graphs of degree bounded by a constant B, is Max-SNP complete for suitably large B [PaYa]. In particular they provide a reduction from Max3SAT. Combined with [ALMSS] this implies the existence of a constant $\delta > 0$ such that approximating MinVC within a factor of $1 + \delta$ is hard unless P = NP. No explicit value of δ has been stated until now. Indeed, the value that could be derived, even using the best existing non-approximability results for Max3SAT, will be very small, because of the cost of the reduction of [PaYa], which first reduces Max3SAT to its bounded version using expanders, and then reduces this to MinVC-B. See Section 5.2 for our results.

Max Clique. The best known polynomial time approximation algorithm for Max Clique achieves a factor of only $N^{1-o(1)}$ [BoHa], scarcely better than the trivial factor of N. There is not even a heuristic algorithm that is conjectured to do better. (The Lovász Theta function had been conjectured to approximate the Max Clique size within $\sqrt{N}$ but this conjecture was disproved by Feige [Fe1].)

Prior to 1991, no non-approximability results on Max Clique were known. In 1991 the connection to proofs was made by Feige et. al. [FGLSS]. The FGLSS reduction says that $\mathrm{PCP}_{1,e}[\,\mathsf{coins}=r;\,\mathsf{query}=q\,]$ Karp reduces to $\mathsf{Gap}\text{-}\mathrm{MaxClique}_{c,s}$ via a reduction running in time $\mathrm{poly}(2^{r+q})$, and with the gap c/s being a function of (r, q and) the error e. In applying it one works with PCP classes containing NP. One obtains a result saying Max Clique has no polynomial time approximation algorithm achieving a certain factor, under an assumption about the deterministic time complexity of NP (the time complexity depends on r, q and the factor on these, but, most importantly, on the error e). In particular, these authors were able to "scale-down" the proof system of [BFL] to indicate strong non-approximability factors of $2^{\log^{1-\epsilon} N}$ for all $\epsilon > 0$, assuming NP is not in quasi-polynomial deterministic time. They also initiated work on improving the factors and assumptions via better proof systems. The best result in their paper is indicated in Figure 7.

Arora and Safra [ArSa] reduced the randomness complexity of a PCP verifier for NP to logarithmic — they showed $\mathrm{NP} = \mathrm{PCP}_{1,1/2}[\,\mathsf{coins}=\log;\,\mathsf{query}=\sqrt{\log N}\,]$. On the other hand, it is easy to see that that random bits can be recycled for error-reduction via the standard techniques of [AKS, CW, ImZu]. The consequence was the first NP-hardness result for Max Clique approximation. The corresponding factor was $2^{\sqrt{\log N}}$.

Arora et. al. [ALMSS] showed that $\mathrm{NP} = \mathrm{PCP}_{1,1/2}[\,\mathsf{coins}=\log;\,\mathsf{query}=O(1)\,]$, which implied that there exists an $\epsilon > 0$ for which approximating Max Clique within N^{ϵ} was NP-complete. The number of queries was unspecified, but indicated to be $\approx 10^4$, so $\epsilon \approx 10^{-4}$. Later work has focused on reducing the constant value of ϵ in the exponent.

In later work a slightly tighter form of the FGLSS reduction due to [BeSc, Zuc] has been used. It says that $\mathrm{PCP}_{1,1/2}[\,\mathsf{coins}=r\,;\,\mathsf{query}_{\mathrm{av}}=q_{\mathrm{av}}\,]$ reduces, via a randomized Karp reduction, to $\mathsf{Gap}\text{-}\mathrm{MaxClique}_{c,s}$ for some c, s satisfying $c(N)/s(N) = N^{1/(1+q_{\mathrm{av}})}$, and with the running time of the reduction being $\mathrm{poly}(2^r)$. (We assume $q_{\mathrm{av}} = O(1)$ for simplicity.) (We omit factors of N^{ϵ} where $\epsilon > 0$ can be arbitrarily small, here and in the following.) Thus the hardness factor was tied to the (average) number of queries required to get soundness error 1/2. Meanwhile the assumption involved the probabilistic, rather than deterministic time complexity of NP– it would be $\mathrm{NP} \not\subseteq \mathrm{co}\widetilde{\mathrm{RP}}$ if $r = \mathrm{polylog}(n)$ and $\mathrm{NP} \neq \mathrm{coRP}$ if $r = \log(n)$.

New proof systems of [BGLR] were able to obtain significantly smaller query complexity: they showed $\mathrm{NP} \subseteq \mathrm{PCP}_{1,1/2}[\,\mathsf{coins}=\mathrm{polylog}\,;\,\mathsf{query}=24\,]$ and $\mathrm{NP} \subseteq \mathrm{PCP}_{1,1/2}[\,\mathsf{coins}=\log\,;\,\mathsf{query}=29\,]$. This leads to their hardness results shown in Figure 7. However, significantly reducing the (average) number of bits queried seemed hard.

However, as observed by Feige and Kilian, the performance of the FGLSS reduction actually depends on the free-bit complexity which may be significantly smaller than the query complexity [FeKi1]. Namely, the factor in the above mentioned reduction is $N^{1/(1+f)}$ where f is the free-bit complexity. They observed that the proof system of [BGLR] has free-bit complexity 14, yielding a $N^{1/15}$ hardness of approximation factor.

The notion of amortized free-bits was introduced in [BeSu]. They observed that the performance of the reduction depended in fact on this quantity, and that the factor was $N^{1/(1+\bar{f})}$ where $\bar{f}$ is the amortized free bit complexity. They then showed that $\mathrm{NP} \subseteq \overline{\mathrm{FPCP}}[\mathrm{polylog}, 3]$. This lead to a $N^{1/4}$ hardness factor assuming $\mathrm{NP} \neq \mathrm{co}\widetilde{\mathrm{RP}}$. See Section 7 for our results.

Chromatic Number. The first hardness result for the chromatic number is due to Garey and Johnson [GJ1]. They showed that if $\mathrm{P} \neq \mathrm{NP}$ then there is no polynomial time algorithm that can achieve a

Due to	Factor	Assumption
[FGLSS]	$2^{\log^{1-\epsilon} N}$ for any $\epsilon > 0$	NP $\not\subseteq \widetilde{\mathrm{P}}$
[ArSa]	$2^{\sqrt{\log N}}$	P $\neq$ NP
[ALMSS]	N^ϵ for some $\epsilon > 0$	P $\neq$ NP
[BGLR]	$N^{1/30}$	NP $\neq$ coRP
[BGLR]	$N^{1/25}$	NP $\not\subseteq$ co$\widetilde{\mathrm{RP}}$
[FeKi1]	$N^{1/15}$	NP $\neq$ coRP
[BeSu]	$N^{1/6}$	P $\neq$ NP
[BeSu]	$N^{1/4}$	NP $\not\subseteq$ co$\widetilde{\mathrm{RP}}$
This paper	$N^{1/4}$	P $\neq$ NP
This paper	$N^{1/3}$	NP $\neq$ coRP

Figure 7: *Some Milestones in the project of proving non-approximability of the Clique number.*

factor less than 2. This remained the best result until the connection to proof systems, and the above mentioned results, emerged.

Hardness results for the chromatic number were obtained via reduction from Max Clique. A N^ϵ factor hardness for Max Clique translates into a N^δ factor hardness for the Chromatic number[2], with $\delta = \delta(\epsilon)$ a function of ϵ. In all reductions $\delta(\epsilon) = \min\{h(\epsilon), h(0.5)\}$, for some function h. The bigger h, the better the reduction.

The first reduction, namely that of Lund and Yannakakis [LuYa], obtained $h(\epsilon) = \epsilon/(5 - 4\epsilon)$. Via the Max Clique hardness results of [ArSa, ALMSS] this implies the chromatic number is hard to approximate within N^δ for some $\delta > 0$. But, again, δ is very small. Improvements to δ were derived both by improvements to ϵ and improvements to the function h used by the reduction.

A subsequent reduction of Khanna, Linial and Safra [KLS] is simpler but in fact slightly less efficient, having $h(\epsilon) = \epsilon/(5 + \epsilon)$. A more efficient reduction is given by [BeSu] – they present a reduction obtaining $h(\epsilon) = \epsilon/(3 - 2\epsilon)$. Our $N^{1/3}$ hardness for Clique would yield, via this, a $N^{1/7}$ hardness for the chromatic number. But more recently an even more efficient reduction has become available, namely that of Fürer [Fu]. This reduction achieves $h(\epsilon) = \epsilon/(2 - \epsilon)$, and thereby we get our $N^{1/5}$ hardness.

Following the appearance of our results, Feige and Kilian [FeKi2] have introduced a new approach to showing hardness of approximability of ChromNum. See discussion in Section 1.5.

Randomized and de-randomized error reduction. As mentioned above, randomized and de-randomized error reduction techniques play an important role in obtaining the best Clique hardness results via the FGLSS method. Typically, one first reduces the error so that its logarithm relates to the query (or free-bit) complexity and so that the initial randomness cost can be ignored (as long as it were logarithmic). (Otherwise, one would have needed to construct proof systems which minimize also this parameter; i.e., the constant factor in the logarithmic randomness complexity.)

[2]Actually all the reductions presented here, make assumptions regarding the structure of the graph and hence do not directly yield the hardness results stated here. However, as a consequence of some results from this paper, we are able to remove the assumptions made by the earlier papers and hence present those results in a simpler form. See Section 8.4 for details.

The randomized error reduction method originates in the work of Berman and Schnitger [BeSc] were it is applied to the Clique Gap promise problem. An alternative description is given by Zuckerman [Zuc]. Another alternative description, carried out in the proof system, is presented in Section 11.

The de-randomized error reduction method consists of applying general, de-randomized, error-reduction techniques to the proof system setting.[3] The best method knows as the "Expander Walk" technique is due to Ajtai, Komlos and Szemeredi [AKS] (see also [CW, ImZu]). It is easy to see that this applies in the pcp context. (The usage of these methods in the pcp context begins with [ArSa].) It turns out that the (constant) parameters of the expander, specifically the ratio $\rho \stackrel{\rm def}{=} \frac{\log_2 d}{\log_2 \lambda}$, where d is the degree of the expander and λ is the second eigenvalue (of its adjacency matrix), play an important role here. In particular, $\rho - 1$ determines how much we lose with respect to the randomized error reduction (e.g., $\mathrm{NP} \in \overline{\mathrm{FPCP}}[\log, f\,]$ translates to a hardness factor of $N^{\frac{1}{1+f}}$ under $\mathrm{NP} \not\subseteq \mathrm{BPP}$ and to a hardness factor of $N^{\frac{1}{\rho+f}}$ under $\mathrm{NP} \neq \mathrm{P}$). Thus the Ramanujan Expander of Lubotzky, Phillips and Sarnak [LPS] play an important role yielding $\rho \approx 2$ (cf. Proposition 11.4), which is the best possible.

[3] An alternative approach, applicable to the Gap-Clique problem and presented in [AFWZ], is to "de-randomize" the graph product construction of [BeSc].

Part I
New proof systems and non-approximability results

3 The Long Code and its machinery

3.1 New PCPs and Hardness Results – Overview and guidemap

The starting point for all our proof systems is a two-prover proof system achieving arbitrarily small but fixed constant error with logarithmic randomness and constant answer size, as provided by Raz [Raz]. This proof system has the property that the answer of the second prover is supposed to be a predetermined function of the answer of the first prover. Thus, verification in it amounts to checking that the first answer satisfies some predicate and that the second answer equals the value obtained from the first answer. Following the "proof composition" paradigm of Arora and Safra [ArSa], the proof string provided to the PCP verifier will consist of "encodings" of the answers of the two provers under a suitable code. The PCP verifier will then check these encodings. As usual, we will check both that these encodings are valid and that they correspond to answers which would have been accepted by the original verifier.

Our main technical contribution is a new code, called the *long code*, and means to check it. The long code of an n-bit information word a is the sequence of 2^{2^n} bits consisting of the values of all possible boolean functions at a. The long code is certainly a disaster in terms of coding efficiency, but it has big advantages in the context of proof verification, arising from the fact that it carries enormous amounts of data about a. The difficulty will be to check that a prover claiming to write the long code of some string a is really doing so.

The long code is described in Section 3.3. In Section 3.5 we provide what we call the "atomic" tests for this code. These tests and their analysis are instrumental to all that follows. Section 3.4 is also instrumental to all that follows. This section sets up the framework for recursive proof checking which is used in all the later proof systems.

The atomic tests are exploited in Section 4.1, to construct a verifier that queries the proof at 3 locations and performs one of two simple checks on the answers obtained. These simple checks are implemented by gadgets of the MaxSNP problem at hand, yielding the non-approximability results. Section 4.2 presents gadgets which are CNF formulae of the corresponding type and Section 4.3 presents Max-CUT gadgets. The non-approximability results for Max3SAT, MaxE3SAT, Max2SAT and MaxCUT follow. The verifier of Section 4.1 benefits from another novel idea which is referred to as *folding* (see Section 3.3). Folding contributes to the improved results for Max3SAT, MaxE3SAT, Max2SAT and Max-CUT, but not to the results regarding Max Clique (and Chromatic Number).

A reasonable non-approximability result for MinVC (Minimum Vertex Cover) can be obtained by the above methodology, but a better result is obtained by constructing a different verifier, for NP languages, that uses exactly two free-bits. This verifier is then used to create a reduction of NP to MinVC via the FGLSS reduction and the standard Karp reduction. This approach is presented in Section 5 where we try to minimizing the soundness error attainable using exactly two free-bits.

In Section 6 we minimize the number of bits queried in a PCP to attain soundness error 1/2 – the result is not of direct applicability, but it is intriguing to know how low this number can go.

We then turn to Max Clique (and Chromatic Number). In Section 7.1 we provide the "iterated" tests. (Here the atomic tests are invoked (sequentially) many times. These invocations are not independent of each other.) This leads to a proof system in which the number of *amortized free-bits* used is two. We then draw the implications for Max Clique (and Chromatic Number). A reader interested only in the (amortized) free-bit and Max Clique results can proceed directly from Section 3.5 to and Section 7.

The improvement in the complexities of the proof systems is the main source of our improved non-approximability results. In addition we also use (for the Max-SAT and Max-CUT problems) a recent improvement in the analysis of linearity testing [BCHKS], and introduce (problem specific) gadgets which represent the various tests of the resulting PCP system.

3.2 Preliminaries to the Long Code

Here $\Sigma = \{0,1\}$ will be identified with the finite field of two elements, the field operations being addition and multiplication modulo two. If X and Y are sets, then $\mathsf{Map}(X,Y)$ denotes the set of all maps of X to Y. For any m we regard Σ^m as a vector space over Σ, so that strings and vectors are identified.

If $a \in \Sigma^m$ then $a^{(i)}$ denotes its i-th bit. Similarly, if f is any function with range Σ^m then $f^{(i)}$ denotes the i-th bit of its output.

Linearity. Let G, H be groups. A map $f\colon G \to H$ is *linear* if $f(x+y) = f(x) + f(y)$ for all $x, y \in G$. Let $\textsc{Lin}(G,H)$ denote the set of all linear maps of G to H.

When $G = \Sigma^n$ and $H = \Sigma$, a function $f\colon G \to H$ is linear if and only if there exists $a \in \Sigma^n$ such that $f(x) = \sum_{i=1}^n a^{(i)} x^{(i)}$ for all $x \in \Sigma^n$.

Distance. The *distance* between functions f_1, f_2 defined over a common finite domain D is

$$\mathrm{Dist}(f_1,f_2) \;=\; \Pr_{x \overset{R}{\leftarrow} D}\left[f_1(x) \neq f_2(x)\right] .$$

Functions f_1, f_2 are *ϵ-close* if $\mathrm{Dist}(f_1,f_2) < \epsilon$. If f maps a group G to a group H we denote by $\mathrm{Dist}(f,\textsc{Lin})$ the minimum, over all $g \in \textsc{Lin}(G,H)$, of $\mathrm{Dist}(f,g)$. (Note the notation does not specify G, H which will be evident from the context). We are mostly concerned with the case where G is a vector space V over Σ and H is Σ. Notice that in this case we have $\mathrm{Dist}(f,\textsc{Lin}) \leq 1/2$ for all $f\colon V \to \Sigma$.

Boolean Functions. Let l be an integer. We let $\mathcal{F}_l \overset{\text{def}}{=} \mathsf{Map}(\Sigma^l,\Sigma)$ be the set of all maps of Σ^l to Σ. We regard $\mathcal{F}_l$ as a vector space (of dimension 2^l) over Σ. Addition and multiplication of functions are defined pointwise.

We let $\mathcal{L}_m \subseteq \mathcal{F}_m$ be the set $\textsc{Lin}(\Sigma^m,\Sigma)$ of linear functions of Σ^m to Σ, and let $\mathcal{L}_m^* = \mathcal{L}_m - \{0\}$ be the non-zero linear functions.

Let $g \in \mathcal{F}_m$ and $\vec{f} = (f_1,\ldots,f_m) \in \mathcal{F}_l^m$. Then $g \circ \vec{f}$ denotes the function in $\mathcal{F}_l$ that assigns the value $g(f_1(x),\ldots,f_m(x))$ to $x \in \Sigma^l$.

The Monomial Basis. For each $S \subseteq [l]$ we let $\chi_S \in \mathcal{F}_l$ be the monomial corresponding to S, defined for $x \in \Sigma^l$ by

$$\chi_S(x) \;=\; \prod_{i \in S} x^{(i)} .$$

The empty monomial, namely $\chi_\emptyset$, is defined to be the constant-one function (i.e., $\chi_\emptyset(x) \;=\; \bar{1}$, for all $x \in \Sigma^l$). The functions $\{\chi_S\}_{S \subseteq [l]}$ form a basis for the vector space $\mathcal{F}_l$ which we call the *monomial basis*. This means that for each $f \in \mathcal{F}_l$, there exists a unique vector $\mathcal{C}(f) = (\,C_f(S)\,)_{S \subseteq [l]} \in \Sigma^{2^l}$ such that

$$f \;=\; \sum_{S \subseteq [l]} C_f(S) \cdot \chi_S . \tag{5}$$

The expression on the right hand side of Equation (5) is called the *monomial series* for f, and the members of $\mathcal{C}(f)$ are called the *coefficients of f* with respect to the monomial basis. We note that

$\mathcal{C}$: $\mathcal{F}_l \to \Sigma^{2^l}$ is a bijection. (The Monomial Basis is reminisent of the Fourier Basis, but the two are actually different.)

3.3 Evaluation operators, the Long Code, and Folding

Evaluation operators. Let $a \in \Sigma^l$. We define the map E_a: $\mathcal{F}_l \to \Sigma$ by $E_a(f) = f(a)$ for all $f \in \mathcal{F}_l$. We say that a map A: $\mathcal{F}_l \to \Sigma$ is an *evaluation operator* if there exists some $a \in \Sigma^l$ such that $A = E_a$. We now provide a useful characterization of evaluation operators. First we need a definition.

Definition 3.1 (Respecting the monomial basis): A map A: $\mathcal{F}_l \to \Sigma$ is said to *respect the monomial basis* if

(1) $A(\chi_\emptyset) = 1$ and

(2) $\forall\, S, T \subseteq [l] \;\; : \;\; A(\chi_S) \cdot A(\chi_T) = A(\chi_{S \cup T})$.

Proposition 3.2 (Characterization of the evaluation operator): A map $\tilde{A}$: $\mathcal{F}_l \to \Sigma$ is an evaluation operator if and only if it is linear and respects the monomial basis.

Proof: Let $a \in \Sigma^l$. It is easy to see that E_a is linear: $E_a(f+g) = (f+g)(a) = f(a)+g(a) = E_a(f) + E_a(g)$. It is also easy to see E_a respects the monomial basis. Firstly we have $E_a(\chi_\emptyset) = \chi_\emptyset(a) = 1$. Next, for every $S, T \subseteq [l]$,

$$E_a(\chi_S) \cdot E_a(\chi_T) \;=\; \chi_S(a) \cdot \chi_T(a) \;=\; \prod_{i \in S} a^{(i)} \cdot \prod_{i \in T} a^{(i)} \,.$$

However $x^2 = x$ for any $x \in \Sigma$ so

$$\prod_{i \in S} a^{(i)} \cdot \prod_{i \in T} a^{(i)} \;=\; \prod_{i \in S \cup T} a^{(i)} \;=\; \chi_{S \cup T}(a) \;=\; E_a(\chi_{S \cup T})$$

Now we turn to the converse. Let $\tilde{A}$: $\mathcal{F}_l \to \Sigma$ be linear and respect the monomial basis. For $i = 1, \ldots, l$, let $a_i \stackrel{\text{def}}{=} \tilde{A}(\chi_{\{i\}})$, and let $a \stackrel{\text{def}}{=} a_1 \ldots a_l$. We claim that $\tilde{A} = E_a$. The proof is as follows. We first claim that

$$\forall\, S \subseteq [l] \;\; : \;\; \tilde{A}(\chi_S) = \chi_S(a) \,. \tag{6}$$

Since $\tilde{A}$ respects the monomial basis we have $\tilde{A}(\chi_\emptyset) = 1$ which in turn equals $\chi_\emptyset(a)$, proving Eq. (6) for $S = \emptyset$. To establish Eq. (6) for $S = \{i_1, \ldots, i_t\} \neq \emptyset$, we write

$$\tilde{A}(\chi_S) \;=\; \tilde{A}\left(\chi_{\{i_1\} \cup \cdots \cup \{i_t\}}\right) \;=\; \prod_{j=1}^{t} \tilde{A}(\chi_{\{i_j\}}) \;=\; \prod_{j=1}^{t} a_{i_j} \;=\; \chi_S(a) \,.$$

where the second equality is due to the fact that $\tilde{A}$ respects the monomial basis. This establishes Eq. (6). Now for any $f \in \mathcal{F}_l$ we can use the linearity of $\tilde{A}$ to see that

$$\tilde{A}(f) \;=\; \tilde{A}\left(\sum_S C_f(S) \cdot \chi_S)\right) \;=\; \sum_S C_f(S) \cdot \tilde{A}(\chi_S) \;=\; \sum_S C_f(S) \cdot \chi_S(a) \;=\; f(a) \;=\; E_a(f) \,.$$

Thus $\tilde{A} = E_a$. ▮

The Long Code. Intuitively, the encoding of $a \in \{0,1\}^l$ (via the long code) is the 2^{2^l} bit string which in position $f \in \mathcal{F}_l$ stores the bit $f(a)$. The Long Code is thus an extremely "redundant" code, encoding an l-bit string by the values, at a, of *all* functions in $\mathcal{F}_l$.

Definition 3.3 (Long Code:) The *long code* $E\colon \Sigma^l \to \mathsf{Map}(\mathcal{F}_l, \Sigma)$ is defined for any $a \in \Sigma^l$ by $E(a) = E_a$.

In some natural sense E is the longest possible code: E is the longest code which is not repetitive (i.e., does not have two positions which are identical in all codewords).

We let $\mathrm{Dist}(A, \textsc{Eval}) = \min_{a\in\Sigma^l} \mathrm{Dist}(A, E_a)$ be the distance from A to a closest codeword of E. It is convenient to define $E^{-1}(A) \in \Sigma^l$ as the lexicographically least $a \in \Sigma^l$ such that $\mathrm{Dist}(A, E_a) = \mathrm{Dist}(A, \textsc{Eval})$. Notice that if $\mathrm{Dist}(A, \textsc{Eval}) < 1/4$ then there is exactly one $a \in \Sigma^l$ such that $\mathrm{Dist}(A, E_a) = \mathrm{Dist}(A, \textsc{Eval})$, and so $E^{-1}(A)$ is this a.

The long code is certainly a disaster in terms of coding efficiency, but it has a big advantage in the context of proof verification. Consider, for example, the so-called "circuit test" (i.e., testing that the answer of the first prover satisfies some predetermined predicate/circuit). In this context one needs to check that the codeword encodes a string which satisfies a predetermined predicate (i.e., the codeword encodes some $w \in \{0,1\}^n$ which satisfies $h(w) = 0$, for some predetermined predicate h). The point is that the value of this predicate appears explicitly in the codeword itself, and furthermore it can be easily "self-corrected" by probing the codeword for the values of the functions f and $f + h$, for a uniformly selected function $f : \{0,1\}^n \to \{0,1\}$ (as all these values appear explicitly in the codeword). Actually, the process of verifying, via self-correction, that the value under h is zero can be incorporated into the task of checking the validity of the codeword; this is done by the notion of "$(h, 0)$-folding" (see below). The fact that we can avoid testing whether the codeword encodes a string which satisfies a given function (or that this testing does not cost us anything) is the key to the complexity improvements in our proof systems (over previous proof systems in which a "circuit test" was taking place).

Folding. The intuition behind the notion we will now define is like this. When A is the long code of some string x for which it is known that $h(x) = b$ for some function h and bit b, then half the bits of A become redundant because they can be computed from the other half, via $A(f) = A(f + h) - b$. This phenomenon enables us to reduce the proof checking complexity. To capture it we now define the notion of folding.

Fix $\prec$ to be some canonical, polynomial-time computable total order (reflexive, antisymmetric, transitive) on the set $\mathcal{F}_l$. Given functions $A\colon \mathcal{F}_l \to \Sigma$ and $h \in \mathcal{F}_l \setminus \{\bar{0}\}$ (i.e., h is not the constant function $\bar{0}$) and bit $b \in \Sigma$, the (h, b)*-folding* of A is the function $A_{(h,b)}\colon \mathcal{F}_l \to \Sigma$ given by

$$A_{(h,b)}(f) = \begin{cases} A(f) & \text{if } f \prec h + f \\ A(f + h) - b & \text{otherwise.} \end{cases}$$

(Notice that the above is well-defined for any $h \neq \bar{0}$.) For sake of technical simplicity (see Definition 3.9), we define the $(\bar{0}, 0)$-folding of A to be A itself; namely, $A_{(\bar{0},0)}(f) = A(f)$, for every $f \in \mathcal{F}_l$. As shown below, the (h, b)-folding of a function A is forced to satisfy $A_{(h,b)}(f + h) = A_{(h,b)}(f) + b$, for every $f \in \mathcal{F}_l$ (whereas A itself may not necessarily satisfy these equalities). Before proving this, let us generalize the notion of folding to folding over several, specifically two, functions $h_1, h_2 \in \mathcal{F}_l$ (and bits $b_1, b_2 \in \Sigma$).

Definition 3.4 (Folding): Let $f, h_1, h_2 \in \mathcal{F}_l$. The (h_1, h_2)*-span of* f, denoted $\textsc{span}_{h_1,h_2}(f)$, is defined as the set $\{\, f + \sigma_1 h_1 + \sigma_2 h_2 : \sigma_1, \sigma_2 \in \Sigma \,\}$. Let $\textsc{mincoef}_{h_1,h_2}(f)$ be the pair (σ_1, σ_2) of elements of Σ for which $f + \sigma_1 h_1 + \sigma_2 h_2$ is the smallest function (according to $\prec$) in $\textsc{span}_{h_1,h_2}(f)$. Let $A\colon \mathcal{F}_l \to \Sigma$. Assume h_1, h_2 are distinct and non-zero. Let $b_1, b_2 \in \Sigma$. The *folding* of A *over* (h_1, b_1) *and* (h_2, b_2), denoted $A_{(h_1,b_1),(h_2,b_2)}$, is defined for every $f \in \mathcal{F}_l$ by

$$A_{(h_1,b_1),(h_2,b_2)}(f) = A(f + \sigma_1 h_1 + \sigma_2 h_2) - \sigma_1 b_1 - \sigma_2 b_2\ ,$$

where $(\sigma_1, \sigma_2) = \textsc{mincoef}_{h_1,h_2}(f)$.

The definition extends naturally to the following two cases. In case $(h_1, b_1) = (h_2, b_2)$, folding over the two (identical) pairs is defined as folding over one pair. In case $h_1 \equiv \bar{0}$ and $b_1 = 0$, folding over both (h_1, b_1) and (h_2, b_2) is defined as folding over (h_2, b_2). Note that folding over two pairs is invariant under the order between the pairs; namely, $A_{(h_1,b_1),(h_2,b_2)} \equiv A_{(h_2,b_2),(h_1,b_1)}$. Finally, observe that a function $A\colon \mathcal{F}_l \to \Sigma$ that is folded over two functions (i.e., over both (h_1, b_1) and (h_2, b_2)) is folded over each of them (i.e., over each (h_i, b_i)).

Proposition 3.5 (Folding forces equalities): Let $A\colon \mathcal{F}_l \to \Sigma$, $h_1, h_2 \in \mathcal{F}_l$ and $b_1, b_2 \in \Sigma$ (with $b_i = 0$ in case $h_i \equiv \bar{0}$). Then, for every $f \in \mathcal{F}_l$,

$$A_{(h_1,b_1),(h_2,b_2)}(f + h_1) = A_{(h_1,b_1),(h_2,b_2)}(f) + b_1$$

Proof: By definition, $A_{(h_1,b_1),(h_2,b_2)}(f) = A(f + \sigma_1 h_1 + \sigma_2 h_2) - \sigma_1 b_1 - \sigma_2 b_2$, where the function $f + \sigma_1 h_1 + \sigma_2 h_2$ is the smallest function in $\textsc{span}_{h_1,h_2}(f)$. Since $\textsc{span}_{h_1,h_2}(f + h_1) \equiv \textsc{span}_{h_1,h_2}(f)$, we have $A_{(h_1,b_1),(h_2,b_2)}(f + h_1) = A(f + \sigma_1 h_1 + \sigma_2 h_2) - (\sigma_1 - 1)b_1 - \sigma_2 b_2$. The claim follows. ▮

As a corollary to the above (combined with the self-correcting paradigm [BLR]), we get

Proposition 3.6 (folding and the evaluation operator): Let $A\colon \mathcal{F}_l \to \Sigma$, $h \in \mathcal{F}_l$, $b \in \Sigma$ and $a \in \Sigma^l$. Suppose that for any $f \in \mathcal{F}_l$ it is the case that $A(f + h) = A(f) + b$. Then $\text{Dist}(A, E_a) < 1/2$ implies $h(a) = b$. Consequently, if $\text{Dist}(A_{(h,b),(h',b')}, E_a) < 1/2$ then $h(a) = b$, provided $b = 0$ if $h \equiv \bar{0}$.

Proof: By the hypothesis, we have $A(h+f) = A(f)+b$, for every $f \in \mathcal{F}_l$. Suppose that $\text{Dist}(A, E_a) < 1/2$. Then, noting that E_a is linear and applying a self-correction process (cf., Corollary 3.14 below), we get $E_a(h) = b$. Using the definition of the Evaluator operator (i.e., $E_a(h) = h(a)$) we have $h(a) = b$. The consequence for $A_{(h,b),(h',b')}$ follows since by Proposition 3.5 we have $A_{(h,b),(h',b')}(f+h) = A_{(h,b),(h',b')}(f) + b$ for any $f \in \mathcal{F}_l$. ▮

The verifiers constructed below make virtual access to “folded” functions rather than to the functions themselves. Virtual access to a folding of A is implemented by actual accessing A itself according to the definition of folding (e.g., say one wants to access $A_{(h,0)}$ at f then one determines whether $f \prec h + f$ or not and accesses either $A(f)$ or $A(f + h)$, accordingly). One benefit of folding in our context is illustrated by Proposition 3.6; in case a (h, b)-folded function is close to a codeword (in the long code), we infer that the codeword encodes a string a satisfying $h(a) = b$. We will see that folding (the long code) over $(h, 0)$ allows us to get rid of a standard ingredient in proof verification; the so-called “circuit test”.

In the sequel, we will use folding over the pairs $(h, 0)$ and $(\bar{1}, 1)$, where $h \in \mathcal{F}_l$ is an arbitrary function (typically not identically zero) and $\bar{1}$ is the constant-one function. Folding over $(\bar{1}, 1)$ allows us to simplify the “codeword” test (w.r.t. the long-code).

3.4 Recursive verification of proofs

This section specifies the basic structure of proof construction, and in particular provides the definitions of the notions of inner and outer verifiers which will be used throughout. It is useful to understand these things before proceeding to the tests.

Overview. The constructions of efficient proofs that follow will exploit the notion of recursive verifier construction due to Arora and Safra [ArSa]. We will use just one level of recursion. We first define a notion of a *canonical outer verifier* whose intent is to capture two-prover one-round proof systems [BGKW] having certain special properties; these verifiers will be our starting point. We then define a canonical inner verifier. Recursion is captured by an appropriate definition of a composed verifier whose attributes we relate to those of the original verifiers in Theorem 3.12.

The specific outer verifier we will use is one obtained by a recent work of Raz [Raz]. We will construct various inner verifiers based on the long code and the tests in Section 3.5 and Section 7.1. Theorem 3.12 will be used ubiquitously to combine the two.

Comparison with previous work. For a history and a better understanding of the role of constant-prover proof systems in this context, see Section 2.4.3. In comparison, our definition of outer verifiers (below) asks for almost the same canonicity properties as in [BeSu]. (The only difference is that they have required σ to be a projection function, whereas we can deal with an arbitrary function. But we don't take advantage of this fact.) In addition we need answer sizes of $\log\log n$ as opposed to the $O(\log n)$ of previous methods, for reasons explained below. This means that even the (modified) [LaSh, FeLo] type proofs won't suffice for us. We could use the three-prover modification of [FeKi1] but the cost would wipe out our gain. Luckily this discussion is moot since we can use the recent result of Raz [Raz] to provide us with a canonical two-prover proof having logarithmic randomness, constant answer size, and any constant error. This makes an ideal starting point. To simplify the definitions below we insisted on constant answer size and two provers from the start.

The inner verifiers used in all previous works are based on the use of the Hadamard code constructions of [ALMSS]. (The improvements mentioned above are obtained by checking this same code in more efficient ways). We instead use a new code, namely the long code, as the basis of our inner verifiers. Note the codewords (in the long code) have length double exponential in the message, explaining our need for $\log\log n$ answer sizes in the outer verifier. We also incorporate into the definitions the new idea of folding which we will see means we don't need a circuit test (a hint towards this fact is already present in the definition of a good inner verifier).

3.4.1 Outer verifiers

As mentioned above, outer verifiers will model certain special kinds of two-prover, one-round proof systems. We think of the verifier as provided with a pair of proof oracles π, π_1, and allowed one query to each. The desired properties concern the complexity of the system and a certain behavior in the checking of the proof, as we now describe.

Let $r_1, s, s_1 \colon \mathcal{Z}^+ \to \mathcal{Z}^+$ and let l and l_1 be positive integers. An (l, l_1)-*canonical outer verifier* V_{outer} takes as input $x \in \Sigma^n$, and has oracle access to a pair of proofs $\bar{\pi} \colon [s(n)] \to \Sigma^l$ and $\bar{\pi}_1 \colon [s_1(n)] \to \Sigma^{l_1}$. It does the following.

- Picks a random string R_1 of length $r_1(n)$.
- Computes, as a function of x and R_1, queries $q \in [s(n)]$ and $q_1 \in [s_1(n)]$, and a (circuit computing a) function $\sigma \colon \Sigma^l \to \Sigma^{l_1}$ (which is determined by x and R_1). Determines, based on x and q, a function $h \colon \Sigma^l \to \Sigma$ (and computes an appropriate representation of it).
  (We stress that h does not depend on R_1, only on q and x).
- Lets $a = \bar{\pi}(q)$ and $a_1 = \bar{\pi}_1(q_1)$.
- If $h(a) \neq 0$ then rejects.
- If $\sigma(a) \neq a_1$ then rejects.
- Otherwise accepts.

We call s, s_1 the *proof sizes* for V_{outer} and r_1 the *randomness* of V_{outer}.

Recall that by the conventions in Section 2, $\texttt{ACC}\,[\, V_{\text{outer}}^{\bar{\pi},\bar{\pi}_1}(x) \,]$ denotes the probability, over the choice of R_1, that V_{outer} accepts, and $\texttt{ACC}\,[\, V_{\text{outer}}(x) \,]$ denotes the maximum of $\texttt{ACC}\,[\, V_{\text{outer}}^{\bar{\pi},\bar{\pi}_1}(x) \,]$ over all possible proofs $\bar{\pi}, \bar{\pi}_1$.

Definition 3.7 (goodness of outer verifier): Outer verifier V_{outer} is ϵ-*good* for language L if for all x it is the case that
(1) $x \in L$ implies $\texttt{ACC}\,[\,V_{\text{outer}}(x)\,] = 1$.
(2) $x \notin L$ implies $\texttt{ACC}\,[\,V_{\text{outer}}(x)\,] \le \epsilon$.

Employing the FRS-method [FRS] to any PCP(log,O(1))-system for NP (e.g., [ALMSS]) one gets a canonical verifier which is δ-good for some $\delta < 1$. (Roughly, the method is to take the given pcp system, send all queries to one oracle, and, as a check, a random one of them to the other oracle.) Using the Parallel Repetition Theorem of Raz, we obtain our starting point –

Lemma 3.8 (Construction of outer verifiers): Let $L \in$ NP. Then for every $\epsilon > 0$ there exist positive integers l, l_1 and c such that there exists an (l, l_1)-canonical outer verifier which is ϵ-good for L and uses randomness $r(n) = c \log_2 n$.

Actually, Raz's Theorem [Raz] enables one to assert that l, l_1 and c are $O(\log \epsilon^{-1})$; but we will not need this fact. Also, the function σ determined by this verifier is always a projection, but we don't use this fact either.

3.4.2 Inner verifiers

Inner verifiers are designed to efficiently verify that the encoding of answers, which a (canonical) outer verifier expects to see, indeed satisfy the checks which this outer verifier performs. Typically, the inner verifier performs a combination of a codeword test (i.e., tests that each oracle is indeed a proper encoding relative to a fixed code – in our case the Long Code), a projection test (i.e., that the decoding of the second answer corresponds to the value of σ applied to the decoding of the first), and a "circuit test" (i.e., that the decoding of the first answer evaluates to 0 under the function h).

Let $r_2, l, l_1 \in \mathcal{Z}^+$. A (l, l_1)-*canonical inner verifier* V_{inner} takes as inputs functions $\sigma \colon \Sigma^l \to \Sigma^{l_1}$ and $h \in \mathcal{F}_l$. (It may also take additional inputs, depending on the context). It has oracle access to a pair of functions $A \colon \mathcal{F}_l \to \Sigma$ and $A_1 \colon \mathcal{F}_{l_1} \to \Sigma$, and uses r_2 random bits. The parameters $\delta_1, \delta_2 > 0$ in the following should be thought as extremely small: in our constructions, they are essentially 0 (see comment below).

Definition 3.9 (goodness of inner verifier): An inner verifier V_{inner} is $(\rho, \delta_1, \delta_2)$-*good* if for all σ, h as above–
(1) Suppose $a \in \Sigma^l$ is such that $h(a) = 0$. Let $a_1 = \sigma(a) \in \Sigma^{l_1}$. Then $\texttt{ACC}\,[\,V_{\text{inner}}^{E_a, E_{a_1}}(\sigma, h)\,] = 1$.
(2) Suppose A, A_1 are such that $\texttt{ACC}\,[\,V_{\text{inner}}^{A, A_1}(\sigma, h)\,] \ge \rho$. Then there exists $a \in \Sigma^l$ such that:
(2.1) $\text{Dist}(A_{(h,0),(\bar{1},1)}, E_a) < 1/2 - \delta_1$.
(2.2) $\text{Dist}(A_1, E_{\sigma(a)}) < 1/2 - \delta_2$.

We stress that although the inner verifier has access to the oracle A (and the hypothesis in condition (2) of Definition 3.9 refers to its computations with oracle A), the conclusion in condition (2.1) refers to A folded over both $(h, 0)$ and $(\bar{1}, 1)$, where $\bar{1}$ is the constant-one function. (Typically, but not necessarily, the verifier satisfying Definition 3.9 accesses the virtual oracle $A_{(h,0),(\bar{1},1)}$ by actual access to A according to the definition of folding.) Furthermore, by Proposition 3.6, condition (2.1) implies that $h(a) = 0$. (Thus, there is no need to explicitly require $h(a) = 0$ in order to make Theorem 3.12 work.) We comment that the upper bounds in conditions (2.1) and (2.2) are chosen to be the largest ones which still allow us to prove Theorem 3.12 (below). Clearly, the complexity of the inner verifier decreases as these bounds increase. This is the reason for setting δ_1 and δ_2 to be extremely small.

We stress that this optimization is important for the MaxSNP results but not for the Max Clique result. In the latter case, we can use δ_i's greater than $\frac{1}{4}$ which simplifies a little the analysis of the composition of verifiers (below).

Remark 3.10 (a tedious one): *The above definition allows h to be identically zero (although this case never occurs in our constructions nor in any other reasonable application). This is the reason that we had to define folding over (0,0) as well. An alternative approach would have been to require $h \not\equiv 0$ and assert that this is the case with respect to the outer verifier of Lemma 3.8.*

3.4.3 Composition of verifiers

We now describe the *canonical* composition of a (canonical) outer verifier with a (canonical) inner verifier. Let V_{outer} be a (l, l_1)-canonical outer verifier with randomness r_1 and proof sizes s, s_1. Let V_{inner} be a (l, l_1)-canonical inner verifier with randomness r_2. Their composed verifier $\langle V_{\text{outer}}, V_{\text{inner}}\rangle$ takes as input $x \in \Sigma^n$ and has oracle access to proofs $\pi\colon [s(n)] \times \mathcal{F}_l \to \Sigma$ and $\pi_1\colon [s_1(n)] \times \mathcal{F}_{l_1} \to \Sigma$. We ask that it does the following –

- Picks random strings for both V_{outer} and V_{inner}; namely, picks a random string R_1 of length $r_1(n)$ and a random string R_2 of length $r_2(n)$.
- Computes queries q and q_1 and functions σ and h as V_{outer} would compute them given x, R_1
- Outputs $V_{\text{inner}}^{A,A_1}(\sigma, h; R_2)$ where $A(\cdot) = \pi(q, \cdot)$ and $A_1(\cdot) = \pi_1(q_1, \cdot)$.

The randomness complexity of the composed verifier is $r_1 + r_2$ whereas its query and free-bit complexities equal those of V_{inner}.

We show how the composed verifier $\langle V_{\text{outer}}, V_{\text{inner}}\rangle$ inherits the goodness of the V_{outer} and V_{inner}. To do so we need the following Lemma. It is the counterpart of a claim in [BGLR, Lemma 3.5] and will be used in the same way. The lemma is derived from a coding theory bound which is slight extension of bounds in [McSl, Ch. 7] (see Appendix).

Lemma 3.11 Suppose $0 \le \delta \le 1/2$ and $A\colon \mathcal{F}_l \to \Sigma$. Then there are at most $1/(4\delta^2)$ codewords that have distance less than $1/2 - \delta$ from A. That is,

$$\left|\{\, a \in \Sigma^l \;:\; \mathrm{Dist}(A, E_a) \le 1/2 - \delta \,\}\right| \;\le\; \frac{1}{4\delta^2}\,.$$

Furthermore, for $\delta > 1/4$ the above set contains at most one string.

Proof: We know that E_a is linear for any a (cf. Proposition 3.2). So it suffices to upper bound the size of the set

$$\mathcal{A} \;=\; \{\, X \in \textsc{Lin}(\mathcal{F}_l, \Sigma) \;:\; \mathrm{Dist}(A, X) \le 1/2 - \delta \,\}\,.$$

This set has the same size as

$$\mathcal{B} \;=\; \{\, X - A \;:\; X \in \textsc{Lin}(\mathcal{F}_l, \Sigma) \text{ and } \mathrm{Dist}(A, X) \le 1/2 - \delta \,\}\,.$$

Let $n = 2^{2^l}$ and identify $\mathsf{Map}(\mathcal{F}_l, \Sigma)$ with Σ^n in the natural way. Let $w(\cdot)$ denote the Hamming weight. Now note that $Z = X - A \in \mathcal{B}$ implies $w(Z)/n = \mathrm{Dist}(X, A) \le 1/2 - \delta$. Furthermore if $Z_1 = X_1 - A$ and $Z_2 = X_2 - A$ are in $\mathcal{B}$ then $\mathrm{Dist}(Z_1, Z_2) = \mathrm{Dist}(X_1, X_2)$ and the latter is $1/2$ if $X_1 \ne X_2$, since X_1, X_2 are linear. Thus, $\mathcal{B}$ is a set of binary vectors of length n, each of weight at most $(0.5 - \delta)n$, and any two of distance at least $0.5n$ apart. Invoking Lemma A.1 (with $\alpha = \delta$ and $\beta = 0$), we upper bound the size of $\mathcal{B}$ as desired. Finally, when $\delta > 1/4$ the triangle inequality implies that we cannot have $a_1 \ne a_2$ so that $\mathrm{Dist}(A, E_{a_i}) \le 1/2 - \delta < 1/4$ for both $i = 1, 2$. ▌

In some applications of the following theorem, $\delta_1, \delta_2 > 0$ will first be chosen to be so small that they may effectively be thought of as 0. (This is done in order to lower the complexities of the inner verifiers.) Once the δ_i's are fixed, ϵ will be chosen to be so much smaller (than the δ_i's) that $\epsilon/(16\delta_1^2\delta_2^2)$ may be thought of as effectively 0. The latter explains why we are interested in outer verifiers which achieve a constant, but arbitrarily small, error ϵ. For completeness we provide a proof, following the ideas of [ArSa, ALMSS, BGLR].

Theorem 3.12 (the composition theorem): Let V_{outer} be a (l, l_1)-canonical outer verifier. Suppose it is ϵ-good for L. Let V_{inner} be an (l, l_1)-canonical inner verifier that is $(\rho, \delta_1, \delta_2)$-good. Let $V = \langle V_{\text{outer}}, V_{\text{inner}} \rangle$ be the composed verifier, and let $x \in \Sigma^*$. Then —
(1) If $x \in L$ then $\texttt{ACC}\,[\,V(x)\,] = 1$
(2) If $x \notin L$ then $\texttt{ACC}\,[\,V(x)\,] \le \rho + \frac{\epsilon}{16\delta_1^2\delta_2^2}$.
For $\delta_1, \delta_2 > 1/4$ the upper bound in (2) can be improved to $\rho + \epsilon$.

(The latter case (i.e., $\delta_1, \delta_2 > 1/4$) suffices for the Max Clique results.)

Proof: Let $n = |x|$, and let s, s_1 denote the proof sizes of V_{outer}.

Suppose $x \in L$. By Definition 3.7 there exist proofs $\bar{\pi}$: $[s(n)] \to \Sigma^l$ and $\bar{\pi}_1$: $[s_1(n)] \to \Sigma^{l_1}$ such that $\texttt{ACC}\,[\,V_{\text{outer}}^{\bar{\pi},\bar{\pi}_1}(x)\,] = 1$. Let π: $[s(n)] \times \mathcal{F}_l \to \Sigma$ be defined by $\pi(q,f) = E_{\bar{\pi}(q)}(f)$. (In other words, replace the l bit string $\bar{\pi}(q)$ with its 2^{2^l} bit encoding under the long code, and let the new proof provide access to the bits in this encoding). Similarly let π_1: $[s_1(n)] \times \mathcal{F}_{l_1} \to \Sigma$ be defined by $\pi_1(q_1, f_1) = E_{\bar{\pi}_1(q_1)}(f_1)$. Now one can check that the item (1) properties in Definitions 3.7 and 3.9 (of the outer and inner verifier, respectively) imply that $\texttt{ACC}\,[\,V^{\pi,\pi_1}(x)\,] = 1$.

Now suppose $x \notin L$. Let π: $[s(n)] \times \mathcal{F}_l \to \Sigma$ and π_1: $[s_1(n)] \times \mathcal{F}_{l_1} \to \Sigma$ be proof strings for V. We will show that $\texttt{ACC}\,[\,V^{\pi,\pi_1}(x)\,] \le \rho + \epsilon/(16\delta_1^2\delta_2^2)$. Since π, π_1 were arbitrary, this will complete the proof.

We set $N_1 = \lfloor 1/(4\delta_1^2) \rfloor$ and $N_2 = \lfloor 1/(4\delta_2^2) \rfloor$ (with $N_1 = 1$ if $\delta_1 > 1/4$ and $N_2 = 1$ if $\delta_2 > 1/4$). The idea to show $\texttt{ACC}\,[\,V^{\pi,\pi_1}(x)\,] \le \rho + N_1 N_2 \cdot \epsilon$ is as follows. We will first define a collection of N_1 proofs $\bar{\pi}^1, \ldots, \bar{\pi}^{N_1}$ and a collection of N_2 proofs $\bar{\pi}_1^1, \ldots, \bar{\pi}_1^{N_2}$ so that each pair $(\bar{\pi}^i, \bar{\pi}_1^j)$ is a pair of oracles for the *outer* verifier. Next we will partition the random strings R_1 of the *outer* verifier into two categories, depending on the performance of the *inner* verifier on the inputs (i.e., the functions σ, h and the oracles A, A_1) induced by R_1. On the "bad" random strings of the *outer* verifier, the *inner* verifier will accept with probability at most ρ; on the "good" ones, we will use the soundness of the *inner* verifier to infer that that the *outer* verifier accepts under some oracle pair $(\bar{\pi}^i, \bar{\pi}_1^j)$, for $i \in [N_1]$ and $j \in [N_2]$. The soundness of the *outer* verifier will be used to bound the probability of such acceptances.

We now turn to the actual analysis. We define N_1 proofs $\bar{\pi}^1, \ldots, \bar{\pi}^{N_1}$: $[s(n)] \to \Sigma^l$ as follows. Fix $q \in [s(n)]$ and let $A = \pi(q, \cdot)$. Let $B_q = \{\, a \in \Sigma^l \;:\; \text{Dist}(A_{(h,0),(\bar{1},1)}, E_a) \le 1/2 - \delta_1 \,\}$. (Notice that for this set to be well-defined we use the fact that h is well-defined given q.) Note that $|B_q| \le N_1$ by Lemma 3.11. Order the elements of B_q in some canonical way, adding dummy elements to bring the number to exactly N_1, so that they can be written as $a^1(q), \ldots, a^{N_1}(q)$. Now set $\bar{\pi}^i(q) = a^i(q)$ for $i = 1, \ldots, N_1$. In a similar fashion we define $\bar{\pi}_1^j(q_1) = a_1^j(q_1)$ for $j = 1, \ldots, N_2$, where each $a_1^j = a_1^j(q_1)$ satisfies $\text{Dist}(\pi_1(q_1, \cdot), E_{a_1^j}) \le 1/2 - \delta_2$.

Let R_1 be a random string of V_{outer}. We say that R_1 is *good* if

$$\texttt{ACC}\,[\,V_{\text{inner}}^{\pi(q,\cdot),\pi_1(q_1,\cdot)}(\sigma, h)\,] \;\ge\; \rho \;,$$

where q, q_1, σ, h are the queries and functions specified by R_1. If R_1 is not good we say it is *bad*. The claim that follows says that if R_1 is good then there is some choice of the above defined proofs which leads the outer verifier to accept on coins R_1.

Claim. Suppose R_1 is good. Then there is an $i \in [N_1]$ and a $j \in [N_2]$ such that $V_{\text{outer}}^{\bar{\pi}^i,\bar{\pi}_1^j}(x;R_1) = 0$.

Proof. Let q, q_1, σ, h be the queries and functions specified by R_1. Let $A = \pi(q,\cdot)$ and $A_1 = \pi_1(q_1,\cdot)$ (be the oracles accessed by the inner verifier). Since R_1 is good we have $\texttt{ACC}\,[\,V_{\text{inner}}^{A,A_1}(\sigma,h)\,] \geq \rho$. So by Item (2) of Definition 3.9 there exists $a \in \Sigma^l$ such that $\text{Dist}(A_{(h,0),(\bar{1},1)}, E_a) < 1/2 - \delta_1$ and $\text{Dist}(A_1, E_{\sigma(a)}) < 1/2 - \delta_2$. Let $a_1 = \sigma(a)$. Since $\text{Dist}(A_{(h,0),(\bar{1},1)}, E_a) \leq 1/2 - \delta_1$ it must be the case that $a \in B_q$, and hence there exists $i \in [N_1]$ such that $a = \bar{\pi}^i(q)$. Similarly $\text{Dist}(A_1, E_{\sigma(a)}) < 1/2 - \delta$ implies that there is some $j \in [N]$ such that $a_1 = \bar{\pi}_1^j(q_1)$. By Proposition 3.6 we have $h(a) = 0$, and we have $\sigma(a) = a_1$ by (the above) definition. Now, by definition of the (execution of the) canonical outer verifier, $V_{\text{outer}}^{\bar{\pi}^i,\bar{\pi}_1^j}(x;R_1) = 0$ holds. □

By conditioning we have $\texttt{ACC}\,[\,V^{\pi,\pi_1}(x)\,] \leq \alpha + \beta$ where

$$\begin{aligned} \alpha &= \Pr_{R_1}[R_1 \text{ is good}] \\ \beta &= \Pr_{R_1,R_2}[\,V^{\pi,\pi_1}(x;R_1R_2) = 0 \mid R_1 \text{ is bad}\,] \,. \end{aligned}$$

The definition of badness implies $\beta \leq \rho$. On the other hand we can use the Claim to see that

$$\begin{aligned} \alpha &\leq \Pr_{R_1}\left[\exists i \in [N_1]\,,\, j \in [N_2] \,:\, V_{\text{outer}}^{\bar{\pi}^i,\bar{\pi}_1^j}(x;R_1) = 0\right] \\ &\leq \textstyle\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} \Pr_{R_1}\left[V_{\text{outer}}^{\bar{\pi}^i,\bar{\pi}_1^j}(x;R_1) = 0\right] \\ &\leq N_1N_2 \cdot \epsilon \,, \end{aligned}$$

the last by the soundness of V_{outer} (i.e., Item (2) of Definition 3.7). Using the bound on N_1 and N_2, the proof is concluded. ▌

3.5 The atomic tests

Motivation. Our constructions of proofs systems will use the outer verifier of Lemma 3.8, composed via Theorem 3.12 with inner verifiers to be constructed. The brunt of our constructions is the construction of appropriate inner verifiers. The inner verifier will have oracle access to a function $A \colon \mathcal{F}_l \to \Sigma$ and a function $A_1 \colon \mathcal{F}_{l_1} \to \Sigma$. In all our applications, A is supposed to be a folding of an encoding of the answer a of the first prover (in a two-prover proof system) and A_1 is supposed to be the encoding of the answer a_1 of the second prover. The verifier will perform various tests to determine whether these claims are true. The design of these tests is the subject of this subsection.

The atomic tests we provide here will be used directly in the proof systems for showing non-approximability of Max3SAT, Max2SAT and MaxCUT. Furthermore, they are also the basis of iterated tests which will lead to proof systems of amortized free-bit complexity ≈ 2, which in turn are used for the Max Clique and Chromatic Number results. We remark that for the applications to the above-mentioned MaxSNP problems it is important to have the best possible analysis of our atomic tests, and what follows strives to this end. We stress that the exposition and analysis of these tests, in this subsection, is independent of the usage of the codes in our proof systems.

Testing for a codeword. The first task that concerns us is to design a test which, with high probability, passes if and only if A is close to an evaluation operator (i.e., a valid codeword). The

The Atomic Tests. Here $A\colon \mathcal{F}_l \to \Sigma$ and $A_1\colon \mathcal{F}_{l_1} \to \Sigma$ are the objects being tested. The tests also take additional inputs or parameters: below $f, f_1, f_2, f_3 \in \mathcal{F}_l$; $g \in \mathcal{F}_{l_1}^m$; and $\sigma\colon \Sigma^l \to \Sigma^{l_1}$.

$\mathbf{LinTest}(A; f_1, f_2)$ (Linearity Test)
If $A(f_1) + A(f_2) = A(f_1 + f_2)$ then output 0 else output 1.

$\mathbf{MBTest}(A; f_1, f_2, f_3)$ (Respecting-Monomial-Basis Test)
If $A(f_1) = 0$ then check if $A(f_1 \cdot f_2 + f_3) = A(f_3)$
Otherwise (i.e. $A(f_1) = 1$) then check if $A(f_1 \cdot f_2 + f_2 + f_3) = A(f_3)$
Output 0 if the relevant check succeeded, else output 1.

$\mathbf{ProjTest}_\sigma(A, A_1; f, g)$ (Projection Test)
If $A_1(g) = A(g \circ \sigma + f) - A(f)$ then output 0, else output 1.

The Passing Probabilities. These are the probabilities we are interested in:

$$\begin{aligned}
\textsc{LinPass}(A) &= \Pr_{f_1,f_2 \overset{R}{\leftarrow} \mathcal{F}_l} [\,\mathbf{LinTest}(A; f_1, f_2) = 0\,] \\
\textsc{MBPass}(A) &= \Pr_{f_1,f_2,f_3 \overset{R}{\leftarrow} \mathcal{F}_l} [\,\mathbf{MBTest}(A; f_1, f_2, f_3) = 0\,] \\
\textsc{ProjPass}_\sigma(A, A_1) &= \Pr_{f \overset{R}{\leftarrow} \mathcal{F}_l\, ;\, g \overset{R}{\leftarrow} \mathcal{F}_{l_1}} [\,\mathbf{ProjTest}_\sigma(A, A_1; f, g) = 0\,]
\end{aligned}$$

Figure 8: *The atomic tests and their passing probabilities.*

idea is to exploit the characterization of Proposition 3.2. Thus we will perform (on A) a linearity test, and then a “Respect of Monomial Basis” test. Linearity testing is well understood, and we will use the test of [BLR], with the analyses of [BLR, BGLR, BCHKS]. The main novelty is the Respect of Monomial Basis Test.

Circuit and projection. Having established that A is close to some evaluation operator E_a, we now want to test two things. The first is that $h(a) = 0$ for some predetermined function h. This test which would normally be implemented by “self-correction” (i.e., evaluating $h(a)$ by uniformly selecting $f \in \mathcal{F}_l$ and computing $A(f + h) - A(f)$) is not needed here, since in our applications we will use an $(h, 0)$-folding of A instead of A. Thus, it is left to test that the two oracles are consistent in the sense that A_1 is not too far from an evaluation operator which corresponds to $\sigma(a)$ for some predetermined function σ.

Self-correction. The following self-correction lemma is due to [BLR] and will be used throughout.

Lemma 3.13 (Self Correction Lemma [BLR]): Let $A, \tilde{A}\colon \mathcal{F}_l \to \Sigma$ with $\tilde{A}$ linear, and let $x = \mathrm{Dist}(A, \tilde{A})$. Then for every $g \in \mathcal{F}_l$:

$$\Pr_{f \overset{R}{\leftarrow} \mathcal{F}_l} \left[A(f + g) - A(f) = \tilde{A}(g)\right] \;\geq\; 1 - 2x\ .$$

Proof:

$$\begin{aligned}
&\Pr_{f \overset{R}{\leftarrow} \mathcal{F}_l}\left[A(f+g) - A(f) = \tilde{A}(g)\right] \\
&\quad\geq\ \Pr_{f \overset{R}{\leftarrow} \mathcal{F}_l}\left[A(f+g) = \tilde{A}(f+g) \text{ and } A(f) = \tilde{A}(f)\right] \\
&\quad\geq\ 1 - \left(\Pr_{f \overset{R}{\leftarrow} \mathcal{F}_l}\left[A(f+g) \neq \tilde{A}(f+g)\right] + \Pr_{f \overset{R}{\leftarrow} \mathcal{F}_l}\left[A(f) \neq \tilde{A}(f)\right]\right) .
\end{aligned}$$

However each of the probabilities in the last expression is bounded above by x. ▮

Corollary 3.14 Let $A, \tilde{A}\colon \mathcal{F}_l \to \Sigma$ with $\tilde{A}$ linear, and suppose $x \overset{\text{def}}{=} \text{Dist}(A, \tilde{A}) < 1/2$. Suppose also that $A(f+h) = A(f) + \sigma$, for some fixed $h \in \mathcal{F}_l$, $\sigma \in \Sigma$ and every $f \in \mathcal{F}_l$. Then $\tilde{A}(h) = \sigma$.

Proof: By the hypothesis, we have $A(f+h) - A(f) = \sigma$ for all functions f. Thus, we can write

$$\Pr_{f \overset{R}{\leftarrow} \mathcal{F}_l}\left[A(f+h) - A(f) = \tilde{A}(h)\right] \ = \ \Pr_{f \overset{R}{\leftarrow} \mathcal{F}_l}\left[\sigma = \tilde{A}(h)\right] .$$

But the right hand side (and hence the left) is either 0 or 1 (as both h and σ are fixed). However, by Lemma 3.13 the left hand side is bounded below by $1 - 2x > 0$ and so the corollary follows. ▮

Convention. All our tests output a bit, with 0 standing for accept and 1 for reject.

3.5.1 Atomic linearity test

The *atomic linearity test* shown in Figure 8 is the one of Blum, Luby and Rubinfeld [BLR]. We want to lower bound the probability $1 - \text{LinPass}(A)$ that the test rejects when its inputs f_1, f_2 are chosen at random, as a function of $x = \text{Dist}(A, \text{Lin})$. The following lemma, due to Bellare et. al. [BCHKS], gives the best known lower bound today.

Lemma 3.15 [BCHKS] Let $A\colon \mathcal{F}_l \to \Sigma$ and let $x = \text{Dist}(A, \text{Lin})$. Then $1 - \text{LinPass}(A) \geq \Gamma_{\text{lin}}(x)$, where the function $\Gamma_{\text{lin}}\colon [0, 1/2] \to [0, 1]$ is defined as follows:

$$\Gamma_{\text{lin}}(x) \overset{\text{def}}{=} \begin{cases} 3x - 6x^2 & 0 \leq x \leq 5/16 \\ 45/128 & 5/16 \leq x \leq 45/128 \\ x & 45/128 \leq x \leq 1/2. \end{cases}$$

The above lower bound is composed of three different bounds with "phase transitions" at $x = \frac{5}{16}$ and $x = \frac{45}{128}$. It was shown in [BCHKS] (see below) that this combined lower bound is close to the best one possible.

Perspective. The general problem of linearity testing as introduced and studied by Blum et. al. [BLR] is stated as follows: Given a function $A\colon G \to H$, where G, H are groups, obtain a lower bound on r_A as a function of x_A, where

$$\begin{aligned}
r_A &= \Pr_{a,b \overset{R}{\leftarrow} G}\left[A(a) + A(b) \neq A(a+b)\right] \\
x_A &= \text{Dist}(A, \text{Lin}) .
\end{aligned}$$

Blum et. al. showed that $r_A \geq \frac{2}{9}x_A$, for every A. Their analysis was used in the proof system and Max3SAT non-approximability result of [ALMSS]. Interest in the tightness of the analysis began with [BGLR], with the motivation of improving the Max3SAT non-approximability results. They showed that $r_A \geq 3x_A - 6x_A^2$, for every A. This establishes the first segment of the lower bound quoted above (i.e., of the function $\Gamma_{\rm lin}$). Also, it is possible to use [BLR] to show that $r_A \geq 2/9$ when $x_A \geq 1/4$. Putting these together implies a two segment lower bound with phase transition at the largest root of the equation $3x - 6x^2 = \frac{2}{9}$ (i.e., at $\frac{1}{4} + \frac{\sqrt{33}}{36}$). This lower bound was used in the Max3SAT analyses of [BGLR] and [BeSu].

However, for our applications (i.e., linearity testing over $\mathcal{F}_l$ as in Lemma 3.15), the case of interest is when the underlying groups are $G = {\rm GF}(2)^n$ and $H = {\rm GF}(2)$ (since $\mathcal{F}_l$ may be identified with ${\rm GF}(2)^n$ for $n = 2^l$). The work of [BCHKS] focused on this case and improved the bound on r_A for the case $x_A \geq \frac{1}{4}$ where $A\colon {\rm GF}(2)^n \to {\rm GF}(2)$. Specifically, they showed that $r_A \geq 45/128$ for $x_A \geq \frac{1}{4}$ which establishes the second segment of $\Gamma_{\rm lin}$. They also showed that $r_A \geq x_A$, for every $A\colon {\rm GF}(2)^n \to {\rm GF}(2)$. Combining the three lower bounds, they have derived the three-segment lower bound stated in Lemma 3.15.

The optimality of the above analysis has been demonstrated as well in [BCHKS]. Essentially[4], for every $x \leq 5/16$ there are functions $A\colon {\rm GF}(2)^n \to {\rm GF}(2)$ witnessing $r_A = \Gamma_{\rm lin}(x_A)$ with $x_A = x$. For the interval $(\frac{5}{16}, \frac{1}{2}]$, no tight results are known. Instead, [BCHKS] reports of computer constructed examples of functions $A\colon {\rm GF}(2)^n \to {\rm GF}(2)$ with x_A in every interval $[\frac{k}{100}, \frac{k+1}{100}]$, for $k = 32, 33, ..., 49$, and $r_A < \Gamma_{\rm lin}(x_A) + \frac{1}{20}$. Furthermore, they showed that there exist such functions with both x_A and r_A arbitrarily close to $\frac{1}{2}$.

3.5.2 Monomial basis test

Having determined that A is close to linear, the *atomic respect of monomial basis* test makes sure that the linear function close to A respects the monomial basis. Let us denote the latter function (i.e., the linear function closest to A) by $\tilde{A}$. Recalling Definition 3.1 we need to establish two things: namely, that $\tilde{A}(\chi_\emptyset) = 1$ and that $\tilde{A}(\chi_S) \cdot \tilde{A}(\chi_T) = \tilde{A}(\chi_{S\cup T})$, for every $S, T \subseteq [l]$. Recall that we do not have access to $\tilde{A}$ but rather to A; still, the Self-Correction Lemma provides an obvious avenue to bypass the difficulty provided ${\rm Dist}(A, \tilde{A}) < 1/4$. This would have yielded a solution but quite a wasteful one (though sufficient for the Max Clique and Chromatic Number results). Instead, we adopt the following more efficient procedure.

Firstly, by considering only oracles folded over $(\bar{1}, 1)$, we need not check that $\tilde{A}(\chi_\emptyset) = 1$. (This follows by combining Corollary 3.14 and the fact that the $(\bar{1}, 1)$-folded oracle A satisfies $A(f + \bar{1}) = A(f) + 1$, for all $f \in \mathcal{F}_l$.) Secondly, we test that $\tilde{A}(\chi_S) \cdot \tilde{A}(\chi_T) = \tilde{A}(\chi_{S\cup T})$, for every $S, T \subseteq [l]$, by taking random linear combinations of the S's and T's to be tested. Such linear combinations are nothing but uniformly selected functions in $\mathcal{F}_l$. Namely, we wish to test $\tilde{A}(f) \cdot \tilde{A}(g) = \tilde{A}(f \cdot g)$, where f and g are uniformly selected in $\mathcal{F}_l$. Since A is close to $\tilde{A}$, we can inspect $A(f)$ (resp., $A(g)$) rather than $\tilde{A}(f)$ (resp., $\tilde{A}(g)$) with little harm. However, $f \cdot g$ is not uniformly distributed (when f and g are uniformly selected in $\mathcal{F}_l$) and thus Self-Correction will be applied here. The resulting test is

$$A(f_1) \cdot A(f_2) = A(f_1 \cdot f_2 + f_3) - A(f_3) \qquad (7)$$

This test was analyzed in a previous version of this work [BGS2]; specifically, this test was shown to reject a folded oracle A, with $\tilde{A}$ (the linear function closest to A) which does not respect the monomial basis, with probability at least $(1 - 2x) \cdot (\frac{3}{8} - x + \frac{x^2}{2}) = \frac{3}{8} - \frac{7}{4}x + \frac{5}{2}x^2 - x^3$, where $x = {\rm Dist}(A, \tilde{A})$. Here we present an adaptive version of the above test, which performs even better. We observe that if

[4]Actually, the statement holds only for x's which are integral multiples of 2^{-n}

$A(f_1)=0$ then there is no need to fetch $A(f_2)$ (since the l.h.s. of Eq. (7) is zero regardless of $A(f_2)$). Thus, we merely test whether $A(f_1\cdot f_2+f_3)-A(f_3) \;=\; 0$. But what should be done if $A(f_1)=1$? In this case we may replace f_1 by $f_1+\bar{1}$ (yielding $A(f_1+\bar{1}) = A(f_1)+1 = 0$) and test whether $A((f_1+\bar{1})\cdot f_2+f_3)-A(f_3) \;=\; 0$. The resulting test is depicted in Figure 8.

A TECHNICAL LEMMA. First we recall the following lemma of [BGLR] which provides an improved analysis of Freivalds's matrix multiplication test in the special case when the matrices are symmetric with common diagonal.

Lemma 3.16 (symmetric matrix multiplication test [BGLR]): Let M_1, M_2 be N-by-N symmetric matrices over Σ which agree on their diagonals. Suppose that $M_1 \neq M_2$. Then

$$\Pr_{x,y \xleftarrow{R} \Sigma^N}[xM_1y \neq xM_2y] \;\geq\; \frac{3}{8}\;.$$

Furthermore, $\Pr_{x \xleftarrow{R} \Sigma^N}[xM_1 \neq xM_2] \;\geq\; 3/4$.

Proof: Let $M \stackrel{\rm def}{=} M_1 - M_2$. The probability that a uniformly selected combination of the rows of M yields an all-zero vector is 2^{-r}, where r is the rank of M. Since M is symmetric, not identically zero and has a zero diagonal, it must have rank at least 2. Thus, $\Pr_{x \xleftarrow{R} \Sigma^N}\left[xM \neq 0^N\right] \;\geq\; 3/4$ and the lemma follows. ▌

RMB DETECTORS. Suppose that A is actually linear. In that case, the following lemma provides a condition under which A respects the monomial basis. We start with a definition.

Definition 3.17 (RMB detector): Let $A\colon \mathcal{F}_l \to \Sigma$ and $f \in \mathcal{F}_l$. We say that f is a *detector for A* if

$$\Pr_{g \xleftarrow{R} \mathcal{F}_l}[A(f'\cdot g)\neq 0] \;\geq\; 1/2\;.$$

where $f' = f$ if $A(f)=0$ and $f' = f+\bar{1}$ otherwise.

The number of detectors is clearly related to the rejection probability of the RMB test. Suppose that A (or rather $\tilde{A}$) is linear. Clearly, if A respects the monomial basis then it has no detectors. On the other hand, the following lemma asserts that if A does not respect the monomial basis then it has many detectors.

Lemma 3.18 (RMB test for linear functions): Suppose $\tilde{A}\colon \mathcal{F}_l \to \Sigma$ is linear, $\tilde{A}(\chi_\emptyset)=1$ and $\tilde{A}$ does not respect the monomial basis. Then at least a 3/4 fraction of the functions in $\mathcal{F}_l$ are detectors for $\tilde{A}$.

Proof: Let $N=2^l$. We define a pair of N-by-N matrices whose rows and columns are indexed by the subsets of $[l]$. Specifically, for $S,T \subseteq [l]$, we set

$$\begin{aligned} M_1[S,T] &= \tilde{A}(\chi_S)\cdot\tilde{A}(\chi_T) \\ M_2[S,T] &= \tilde{A}(\chi_{S\cup T})\;. \end{aligned}$$

Clearly, both M_1 and M_2 are symmetric, and they agree on the diagonal. Using $\tilde{A}(\chi_\emptyset)=1$ we have, for every $T\subseteq[l]$,

$$M_1[\emptyset,T] = \tilde{A}(\chi_\emptyset)\cdot\tilde{A}(\chi_T) = 1\cdot\tilde{A}(\chi_T) = M_2[\emptyset,T] \tag{8}$$

By the hypothesis that $\tilde{A}$ does not respects the monomial basis it follows that $M_1 \neq M_2$. Our aim is to relate the inequality of the above matrices to the existence of detectors for $\tilde{A}$. We first express the condition $\tilde{A}(fg)=\tilde{A}(f)\cdot\tilde{A}(g)$ in terms of these matrices.

Recall that $\mathcal{C}\colon \mathcal{F}_l \to \Sigma^{2^l}$ is the transformation which to any $f \in \mathcal{F}_l$ associates the vector $(\, C_f(S)\,)_{S\subseteq[l]}$ whose entries are the coefficients of f in its monomial series. Using the linearity of $\tilde{A}$ we note that

$$\begin{aligned}
\tilde{A}(f)\cdot\tilde{A}(g) &= \tilde{A}\left(\textstyle\sum_S C_f(S)\cdot\chi_S\right)\cdot\tilde{A}\left(\textstyle\sum_T C_g(T)\cdot\chi_T\right)\\
&= \left[\textstyle\sum_S C_f(S)\cdot\tilde{A}(\chi_S)\right]\cdot\left[\textstyle\sum_T C_g(T)\cdot\tilde{A}(\chi_T)\right]\\
&= \textstyle\sum_{S,T} C_f(S)\cdot\tilde{A}(\chi_S)\cdot\tilde{A}(\chi_T)\cdot C_g(T)\\
&= \mathcal{C}(f)M_1\mathcal{C}(g)\,.
\end{aligned}$$

For the next step we first need the following.

Fact. Let $f, g \in \mathcal{F}_l$ and $U \subseteq [l]$. Then $C_{fg}(U) = \sum_{S\cup T=U} C_f(S)\cdot C_g(T)$.

Using this fact (and the linearity of $\tilde{A}$) we have:

$$\begin{aligned}
\tilde{A}(fg) &= \tilde{A}\left(\textstyle\sum_U C_{fg}(U)\cdot\chi_U\right)\\
&= \textstyle\sum_U C_{fg}(U)\cdot\tilde{A}(\chi_U)\\
&= \textstyle\sum_U\sum_{S\cup T=U} C_f(S)\cdot C_g(T)\cdot\tilde{A}(\chi_U)\\
&= \textstyle\sum_{S,T} C_f(S)\cdot C_g(T)\cdot\tilde{A}(\chi_{S\cup T})\\
&= \mathcal{C}(f)M_2\mathcal{C}(g)\,.
\end{aligned}$$

Since $\tilde{A}$ is linear and $\tilde{A}(\bar{1}) = 1$ (as $\bar{1} = \chi_\emptyset$), we can rephrase the condition $A(f'\cdot g) \neq 0$, where $f' = f$ if $\tilde{A}(f) = 0$ and $f' = f + \bar{1}$ otherwise, as $A(f'\cdot g) \neq A(f')\cdot A(g)$. Thus, for every f (setting f' as above), we conclude that

$$A(f'\cdot g) \neq A(f')\cdot A(g) \;\text{ if and only if }\; \mathcal{C}(f')M_2\mathcal{C}(g) \neq \mathcal{C}(f')M_1\mathcal{C}(g)\,.$$

A key observation is that $\mathcal{C}(f)$ and $\mathcal{C}(f')$ are identical in all entries except, possibly, for the entry corresponding to $\emptyset$ (i.e., $C_f(S) = C_{f'}(S)$ for all $S \neq \emptyset$). On the other hand, by Eq. (8), we have $M_1[\emptyset,\cdot] = M_2[\emptyset,\cdot]$. Thus,

$$A(f'\cdot g) \neq A(f')\cdot A(g) \;\text{ if and only if }\; \mathcal{C}(f)M_2\mathcal{C}(g) \neq \mathcal{C}(f)M_1\mathcal{C}(g)\,.$$

Now we note that $\mathcal{C}$ is a bijection, so that if h is uniformly distributed in $\mathcal{F}_l$ then $\mathcal{C}(h)$ is uniformly distributed in Σ^{2^l}. Fixing any $f \in \mathcal{F}_l$ and setting f' as above, we have, for $x = \mathcal{C}(f)$,

$$\begin{aligned}
\Pr_{g\xleftarrow{R}\mathcal{F}_l}\left[\tilde{A}(f')\cdot\tilde{A}(g) \;=\; \tilde{A}(f'g)\right] &= \Pr_{g\xleftarrow{R}\mathcal{F}_l}\left[\mathcal{C}(f)M_1\mathcal{C}(g) \;=\; \mathcal{C}(f)M_2\mathcal{C}(g)\right]\\
&= \Pr_{y\xleftarrow{R}\Sigma^{2^l}}\left[xM_1y \;=\; xM_2y\right].
\end{aligned}$$

The latter probability is 1/2 if $xM_1 \neq xM_2$ and zero otherwise. Invoking Lemma 3.16 we conclude that the first case, which coincides with f being a detector for $\tilde{A}$, holds for at least 3/4 fraction of the $f \in \mathcal{F}_l$. The lemma follows. ▌

Lemma 3.18 suggests that if we knew A was linear we could test that it respects the monomial basis by picking f, g at random and testing whether $A(f'g) = 0$, where $f' = f$ if $A(f) = 0$ and $f' = f + \bar{1}$ otherwise. The lemma asserts that in case A is linear and does not respect the monomial basis we will have

$$\Pr_{f,g\xleftarrow{R}\mathcal{F}_l}\left[A(f'g) \neq 0\right] \geq \frac{3}{4}\cdot\frac{1}{2}$$

where 3/4 is a lower bound on the probability that f is a detector for A and $\Pr_{g \overset{R}{\leftarrow} \mathcal{F}_l}[A(f'g) \neq 0] \geq \frac{1}{2}$ for any detector f (by definition). However, we only know that A is close to linear. Still we can perform an approximation of the above test via self-correction of the value $A(f'g)$. This, indeed, is our test as indicated in Figure 8.

The RMB test. We are interested in lower bounding the probability $1 - \text{MBPass}(A)$ that the test rejects when f_1, f_2, f_3 are chosen at random, as a function of the distance of A to a linear function $\tilde{A}$, given that $\tilde{A}$ does not respect the monomial basis. We assume that A satisfies $A(f + \bar{1}) = A(f) + 1$ (for all $f \in \mathcal{F}_l$), as is the case in all our applications (since we use verifiers which access a $(\bar{1}, 1)$-folded function). The first item of the following lemma is in spirit of previous analyses of analogous tests. The second item is somewhat unusual and will be used only in our construction of verifiers of free-bit complexity 2 (cf., Section 5).

Lemma 3.19 (RMB test — final analysis): Let $A, \tilde{A}\colon \mathcal{F}_l \to \Sigma$ be functions such that $\tilde{A}$ linear but does not respect the monomial basis. Let $x = \text{Dist}(A, \tilde{A})$. Suppose that the function A satisfies $A(f + \bar{1}) = A(f) + 1$, for all $f \in \mathcal{F}_l$. Then

1. $1 - \text{MBPass}(A) \;\geq\; \Gamma_{\text{RMB}}(x) \overset{\text{def}}{=} \frac{3}{8} \cdot (1 - 2x)$.
2. $\Pr_{f_1, f_3 \overset{R}{\leftarrow} \mathcal{F}_l}[\exists f_2 \in \mathcal{F}_l \text{ s.t. } \mathbf{MBTest}(A; f_1, f_2, f_3) = 1] \;\geq\; 2 \cdot \Gamma_{\text{RMB}}(x)$.

In particular, the lemma holds for $A_{(h,0),(\bar{1},1)}$, where $A\colon \mathcal{F}_l \to \Sigma$ is arbitrary and $h \in \mathcal{F}_l$. We will consider the linear function closest to $A_{(h,0),(\bar{1},1)}$, denoted $\tilde{A}$, and the case in which $\tilde{A}$ does not respect the monomial basis. (In this case $\text{Dist}(A_{(h,0),(\bar{1},1)}, \tilde{A})) = \text{Dist}(A_{(h,0),(\bar{1},1)}, \text{Lin}) \leq 1/2$.)

Proof: As a preparation to using Lemma 3.18, we first show that $\tilde{A}(\bar{1}) = 1$. For $x < 1/2$ this is justified by Corollary 3.14 (using the hypothesis $A(f + \bar{1}) = A(f) + 1$, $\forall f \in \mathcal{F}_l$). Otherwise (i.e., in case $x \geq 1/2$) the claimed lower bound (i.e., $\frac{3}{8} \cdot (1 - 2x) \leq 0$) holds vacuously.

Using Lemma 3.18 and Lemma 3.13 we lower bound the rejection probability of the test as follows:

$$\begin{aligned}
1 - \text{MBPass}(A) \;&\geq\; \Pr_{f_1 \overset{R}{\leftarrow} \mathcal{F}_l}\left[f_1 \text{ is a detector for } \tilde{A}\right] \\
&\quad \cdot \min_{f \text{ is a } \tilde{A}\text{-detector}} \left\{\Pr_{f_2, f_3 \overset{R}{\leftarrow} \mathcal{F}_l}[\mathbf{MBTest}(A; f, f_2, f_3) = 1]\right\} \\
&\geq\; \frac{3}{4} \cdot \min_{f \text{ is a } \tilde{A}\text{-detector}} \left\{\Pr_{f_2, f_3 \overset{R}{\leftarrow} \mathcal{F}_l}[A(f'f_2 + f_3) \neq A(f_3)]\right\} \\
&\geq\; \frac{3}{4} \cdot \min_{f \text{ is a } \tilde{A}\text{-detector}} \left\{\Pr_{f_2, f_3 \overset{R}{\leftarrow} \mathcal{F}_l}\left[0 \neq \tilde{A}(f'f_2) \;=\; A(f'f_2 + f_3) - A(f_3)\right]\right\} \\
&\geq\; \frac{3}{4} \cdot \frac{1}{2} \cdot \min_{f',\, g \text{ s.t. } \tilde{A}(f'g) \neq 0} \left\{\Pr_{f_3 \overset{R}{\leftarrow} \mathcal{F}_l}\left[\tilde{A}(f' \cdot g) \;=\; A(f' \cdot g + f_3) - A(f_3)\right]\right\} \\
&\geq\; \frac{3}{8} \cdot (1 - 2x)
\end{aligned}$$

where the second inequality uses Lemma 3.18, the fourth inequality follows by the definition of a detector for $\tilde{A}$ (by which $\Pr_{g \overset{R}{\leftarrow} \mathcal{F}_l}\left[\tilde{A}(f'g) \neq 0\right] \geq 1/2$), and the last inequality follows by Lemma 3.13. This concludes the proof of Part (1). Part (2) is proven analogously with the exception that we don't lose a factor of two in the fourth inequality (since here f_2 is not selected at random but rather set existentially). ▌

Remark 3.20 *An RMB test for arbitrary A's (rather than ones satisfying $A(f+\bar{1}) = A(f)+1$, $\forall f \in \mathcal{F}_l$) can be derived by augmenting the above test with a test of $A(f+\bar{1}) = A(f)+1$ for uniformly chosen $f \in \mathcal{F}_l$. The analysis of the augmented part is as in the circuit test (below).*

3.5.3 Atomic projection test

The final test checks that the second function A_1 is not too far from the evaluation operator E_{a_1} where $a_1 = \sigma(a)$ is a function of the string a whose evaluation operator is close to A. Here, unlike previous works (for instance [BeSu]), σ may be an arbitrary mapping from Σ^l to Σ^{l_1} rather than being a projection (i.e., satisfying $\sigma(x) = x^{(i_1)} \dots x^{(i_{l_1})}$ for some sequence $1 \le i_1 < \cdots < i_{l_1} \le l$ and all $x \in \Sigma^l$). Thus, the term "projection test" is adopted for merely historical reasons.

Lemma 3.21 Let $A\colon \mathcal{F}_l \to \Sigma$ and let $\sigma\colon \Sigma^l \to \Sigma^{l_1}$ be a function. Let $a \in \Sigma^l$ and let $x = \mathrm{Dist}(A, E_a)$. Let $a_1 = \sigma(a) \in \Sigma^{l_1}$. Then $1 - \textsc{ProjPass}_\sigma(A, A_1) \ge \mathrm{Dist}(A_1, E_{a_1}) \cdot (1-2x)$.

Proof: We lower bound the rejection probability as follows:

$$\begin{aligned}
&\Pr_{f \xleftarrow{R} \mathcal{F}_l\,;\, g \xleftarrow{R} \mathcal{F}_{l_1}} [A_1(g) \ne A(g\circ\sigma + f) - A(f)] \\
&\quad\ge\; \Pr_{f \xleftarrow{R} \mathcal{F}_l\,;\, g \xleftarrow{R} \mathcal{F}_{l_1}} [A_1(g) \ne E_a(g\circ\sigma) \text{ and } A(g\circ\sigma+f) - A(f) = E_a(g\circ\sigma)] \\
&\quad\ge\; \Pr_{g \xleftarrow{R} \mathcal{F}_{l_1}} [A_1(g) \ne E_a(g\circ\sigma)] \cdot (1-2x)\,.
\end{aligned}$$

Here we used Lemma 3.13 in the last step. Now we note that $E_a(g\circ\sigma) = (g\circ\sigma)(a) = g(\sigma(a)) = E_{a_1}(g)$. Hence the first term in the above product is just

$$\Pr_{g \xleftarrow{R} \mathcal{F}_{l_1}} [A_1(g) \ne E_{a_1}(g)] \;=\; \mathrm{Dist}(A_1, E_{a_1})\,.$$

This concludes the proof. ∎

3.5.4 Atomic circuit test

For sake of elegancy, we present also an atomic Circuit Test, denoted $\mathbf{CircTest}_h(A; f)$. The test consists of checking whether $A(h+f) = A(f)$ and it outputs 0 if equality holds and 1 otherwise. Assuming that A is close to some evaluation operator E_a, the atomic circuit test uses self-correction [BLR] to test that a given function h has value 0 at a. As explained above, this test is not needed since all our proof systems will use a $(h, 0)$-folding (of A) and thus will impose $h(a) = 0$. The analysis lower bounds the rejection probability, as a function of the distance of A from linear, given that $h(a) = 1$.

Lemma 3.22 Let $A\colon \mathcal{F}_l \to \Sigma$ and let $a \in \Sigma^l$. Let $h \in \mathcal{F}_l$ and $x = \mathrm{Dist}(A, E_a)$. If $h(a) = 1$ then $1 - \textsc{CircPass}_h(A) \ge 1 - 2x$, where

$$\textsc{CircPass}_h(A) \stackrel{\text{def}}{=} \Pr_{f \xleftarrow{R} \mathcal{F}_l} [\, \mathbf{CircTest}_h(A; f) = 0 \,]$$

4 A new 3-query PCP and improved MaxSNP hardness results

4.1 The MAX SNP verifier

In this section we present a simple verifier which performs one of two simple checks, each depending on only three queries. This verifier will be the basis for the non-approximability results for several MaxSNP problems, in particular Max3SAT, Max2SAT and MaxCUT, whence the name.

4.1.1 The inner verifier

Figure 9 describes an inner verifier. Our verifier is adaptive; that is, some of its queries are determine as a function of answers to previous queries. (The adaptivity is not obvious from Figure 9; it is rather 'hidden' in the RMB Test — see Section 3.5.2). Thus, adaptivity is used to improve the performance of our verifier and to strengthen the non-approximability results which follow (cf., previous versions of this paper [BGS2]).

The inner verifier, $V_{\rm SNPinner}$, takes the usual length parameters l, l_1 as well as additional (probability) parameters p_1, p_2 and p_3 such that $p_1 + p_2 + p_3 = 1$. It performs just one test: with probability p_1 the linearity test; with probability p_2 the respect of monomial basis test; and with probability p_3 the projection test. Formally, this is achieved by picking p at random and making cases based on its value.[5] To improve the results, we perform the tests on a folding of A over both $(h, 0)$ and $(\bar{1}, 1)$ (i.e., on $A_{(h,0),(\bar{1},1)}$). We stress that $A_{(h,0),(\bar{1},1)}$ is a virtual oracle which is implemented by the verifier which accesses the actual oracle A (on points determined by the definition of folding). We now examine the goodness of $V_{\rm SNPinner}$. Recall the definitions of $\Gamma_{\rm lin}(x)$ (specifically, note that $\Gamma_{\rm lin}(x) \geq x$) and $\Gamma_{\rm RMB}(x) = \frac{3}{8}(1 - 2x)$, for all x.

Informally, the following lemma considers all the possible strategies of a "dishonest" prover and indicates the probability (denoted $1 - \rho$) with which the verifier detects an error (when run against

[5] For simplicity p is depicted as being chosen as a random real number between 0 and 1. Of course we cannot quite do this. But we will see later that the values of p_1, p_2, p_3 in our final verifiers are appropriate constants. So in fact an appropriate choice of p can be made using $O(1)$ randomness, which is what we will implicitly assume.

The Max-SNP inner verifier. Given functions $h \in \mathcal{F}_l$ and $\sigma \colon \Sigma^l \to \Sigma^{l_1}$, the verifier has access to oracles for $A \colon \mathcal{F}_l \to \Sigma$ and $A_1 \colon \mathcal{F}_{l_1} \to \Sigma$. In addition it takes three $[0,1]$ valued parameters p_1, p_2 and p_3 such that $p_1 + p_2 + p_3 = 1$.

```
    Pick $p \stackrel{R}{\leftarrow} [0,1]$.

    Case: $p \leq p_1$ :
        Pick $f_1, f_2 \stackrel{R}{\leftarrow} \mathcal{F}_l$.
        $\mathbf{LinTest}(A_{(h,0),(\bar{1},1)}; f_1, f_2)$.

    Case: $p_1 < p \leq p_1 + p_2$ :
        Pick $f_1, f_2, f_3 \stackrel{R}{\leftarrow} \mathcal{F}_l$.
        $\mathbf{MBTest}(A_{(h,0),(\bar{1},1)}; f_1, f_2, f_3)$.

    Case: $p_1 + p_2 < p$ :
        Pick $f \stackrel{R}{\leftarrow} \mathcal{F}_l$ and $g \stackrel{R}{\leftarrow} \mathcal{F}_{l_1}$.
        $\mathbf{ProjTest}_\sigma(A_{(h,0),(\bar{1},1)}, A_1; f, g)$.
```

Remark: access to $A_{(h,0),(\bar{1},1)}(f)$ is implemented by accessing either $A(f)$, $A(f+h)$, $A(f+\bar{1})$ or $A(f + h + \bar{1})$.

Figure 9: *The Max-SNP inner verifier* $V_{\rm SNPinner}$

such strategies). The three cases correspond to the events that

(1) the function $A_{(h,0),(\bar{1},1)}$ may be very far from being linear;

(2) the function $A_{(h,0),(\bar{1},1)}$ is x-close to linear, for some $x < \frac{1}{2} - \delta_1$, but is not x-close to a valid codeword (i.e., to a linear function which respects the monomial basis); and

(3) the function $A_{(h,0),(\bar{1},1)}$ is x-close to linear but the encoding of $\sigma(E^{-1}(A_{(h,0),(\bar{1},1)}))$ is very far from the function A_1.

Lemma 4.1 (soundness of V_{SNPinner}): Suppose $\delta_1, \delta_2 > 0$ and $l, l_1 \in \mathcal{Z}^+$. Suppose $p_1, p_2, p_3 \in [0,1]$ satisfy $p_1 + p_2 + p_3 = 1$. Then the (l, l_1)-canonical inner verifier V_{SNPinner} is $(\rho, \delta_1, \delta_2)$-good, where $1 - \rho = \min(T_1, T_2, T_3)$ and

(1) $T_1 \stackrel{\text{def}}{=} p_1 \cdot (\frac{1}{2} - \delta_1)$

(2) $T_2 \stackrel{\text{def}}{=} \min_{x \le 1/2 - \delta_1} \left[\, p_1 \cdot \Gamma_{\text{lin}}(x) + p_2 \cdot \Gamma_{\text{RMB}}(x) \,\right]$

(3) $T_3 \stackrel{\text{def}}{=} \min_{x \le 1/2 - \delta_1} \left[\, p_1 \cdot \Gamma_{\text{lin}}(x) + p_3 \cdot (\frac{1}{2} - \delta_2)(1 - 2x) \,\right]$.

Proof: We consider an arbitrary pair of oracles, (A, A_1), and the behavior of V_{SNPinner} when given access to this pair of oracles. Our analysis is broken up into cases depending on (A, A_1); specifically, the first case-partition depends on the distance of $A_{(h,0),(\bar{1},1)}$ (i.e., the folding of A) from linear functions. We show that, in each case, either the verifier rejects with probability bounded below by one of the three quantities (above) or the oracle pair is such that rejection is not required.

Let $x = \text{Dist}(A_{(h,0),(\bar{1},1)}, \text{LIN})$.

Case 1: $x \ge \frac{1}{2} - \delta_1$. Lemma 3.15 implies that $1 - \text{LINPASS}(A_{(h,0),(\bar{1},1)}) \ge \Gamma_{\text{lin}}(x) \ge x \ge \frac{1}{2} - \delta_1$. (The second inequality follows from the fact that $\Gamma_{\text{lin}}(x) \ge x$ for all x.) Since V_{SNPinner} performs the atomic linearity test with probability p_1 we have

$$1 - \texttt{ACC}\,[\,V^{A,A_1}_{\text{SNPinner}}(\sigma, h)\,] \;\ge\; p_1 \cdot (\frac{1}{2} - \delta_1) \;\ge\; 1 - \rho \tag{9}$$

Case 2: $x \le \frac{1}{2} - \delta_1$. Lemma 3.15 implies that $1 - \text{LINPASS}(A_{(h,0),(\bar{1},1)}) \ge \Gamma_{\text{lin}}(x)$ and so the probability that V_{SNPinner} performs the linearity test and rejects is at least $p_1 \cdot \Gamma_{\text{lin}}(x)$. Now let $\tilde{A}$ be a linear function such that $\text{Dist}(A_{(h,0),(\bar{1},1)}, \tilde{A}) = x$. We consider the following sub-cases.

Case 2.1: $\tilde{A}$ does not respect the monomial basis. In this case Part (1) of Lemma 3.19 implies that $1 - \text{MBPASS}(A_{(h,0),(\bar{1},1)}) \ge \Gamma_{\text{RMB}}(x)$. So the probability that V_{SNPinner} performs the atomic respect of monomial basis test and rejects is at least $p_2 \cdot \Gamma_{\text{RMB}}(x)$. Since the event that the verifier performs a linearity test and the event that it performs a respect of monomial basis test are mutually exclusive, we can add the probabilities of rejection and thus get

$$1 - \texttt{ACC}\,[\,V^{A,A_1}_{\text{SNPinner}}(\sigma, h)\,] \;\ge\; p_1 \cdot \Gamma_{\text{lin}}(x) + p_2 \cdot \Gamma_{\text{RMB}}(x) \;\ge\; 1 - \rho \tag{10}$$

Case 2.2: $\tilde{A}$ respects the monomial basis. By Proposition 3.2, $\tilde{A}$ is an evaluation operator. So there exists $a \in \Sigma^l$ such that $\tilde{A} = E_a$. So $\text{Dist}(A_{(h,0),(\bar{1},1)}, E_a) = x$. Let $a_1 = \sigma(a)$. The proof splits into two further sub-cases.

Case 2.2.1: $d \stackrel{\text{def}}{=} \text{Dist}(A_1, E_{a_1}) \ge \frac{1}{2} - \delta_2$. By Lemma 3.21 we have $1 - \text{PROJPASS}_\sigma(A_{(h,0),(\bar{1},1)}, A_1) \ge d \cdot (1 - 2x) \ge (1/2 - \delta_2) \cdot (1 - 2x)$. So the probability that V_{SNPinner} performs the projection test and rejects is at least $p_3 \cdot (1/2 - \delta_2)(1 - 2x)$. Thus, adding probabilities as in Case (2.1), we get

$$1 - \texttt{ACC}\,[\,V^{A,A_1}_{\text{SNPinner}}(\sigma, h)\,] \;\ge\; p_1 \cdot \Gamma_{\text{lin}}(x) + p_3 \cdot (1/2 - \delta_2)(1 - 2x) \;\ge\; 1 - \rho \tag{11}$$

Case 2.2.2: Else, we have $x = \mathrm{Dist}(A_{(h,0),(\bar{1},1)}, E_a) \le 1/2 - \delta_1$ and $\mathrm{Dist}(A_1, E_{a_1}) < 1/2 - \delta_2$. Thus the functions $A_{(h,0),(\bar{1},1)}$ and A_1 satisfy conditions (2.1) and (2.2) in Definition 3.9.

Observe that the only case which does not yield $1 - \mathtt{ACC}\,[V^{A,A_1}_{\mathrm{PCPinner}}(\sigma,h)] \ge 1-\rho$ is Case (2.2.2). However, Case (2.2.2) satisfies conditions (2.1) and (2.2) of Definition 3.9. Thus, V_{PCPinner} satisfies condition (2) of Definition 3.9. Clearly, V_{PCPinner} also satisfies condition (1) of Definition 3.9, and thus the lemma follows. ▌

The upper bound on the soundness error of V_{SNPinner}, provided by Lemma 4.1, is somewhat complicated to grasp. Fortunately, using $\Gamma_{\mathrm{RMB}}(x) = \frac{3}{8}(1-2x)$ and $\Gamma_{\mathrm{lin}}(x) \ge x$, for all $x \le 1/2$, we can simplify the expression as follows.

Claim 4.2 Let T_1, T_2 and T_3 be as in Lemma 4.1, $\delta = \max(\delta_1,\delta_2) > 0$ and $p_1,p_2,p_3 \in [0,1]$ satisfy $p_1+p_2+p_3 = 1$. Then, $T_2 \ge \min\{\frac{1}{2}p_1, \frac{3}{8}p_2\}$, $T_3 \ge \min\{\frac{1}{2}p_1, \frac{1}{2}p_3\} - \delta$, and

$$\min\{T_1,T_2,T_3\} \ge \min\left\{\frac{1}{2}p_1, \frac{3}{8}p_2, \frac{1}{2}p_3\right\} - \delta\ .$$

Interestingly, this lower bound is tight.

Proof: Clearly, $T_1 = (\frac{1}{2} - \delta_1)p_1 \ge \frac{1}{2}p_1 - \delta$. To analyze T_2, let $h(x) \stackrel{\mathrm{def}}{=} p_1 \cdot \Gamma_{\mathrm{lin}}(x) + p_2 \cdot \Gamma_{\mathrm{RMB}}(x)$.

Fact 1: $\min_{x \le 1/2}\{h(x)\} = \min\{\frac{3}{8}p_2, \frac{1}{2}p_1\} = \min\{h(0), h(1/2)\}$.

proof: by considering two cases and using $\Gamma_{\mathrm{lin}}(x) \ge x$ and $\Gamma_{\mathrm{RMB}}(x) = \frac{3}{8} - \frac{3}{4}x$.

case 1: $p_1 \ge \frac{3}{4}p_2$

$$h(x) \ge p_1 x + \frac{3}{8}p_2 - \frac{3}{4}p_2 x = \frac{3}{8}p_2 + (p_1 - \frac{3}{4}p_2)\cdot x \ge \frac{3}{8}p_2$$

case 2: $p_1 \le \frac{3}{4}p_2$

$$h(x) \ge p_1 x + \frac{3}{8}p_2 - \frac{3}{4}p_2 x = \frac{1}{2}p_1 + (\frac{3}{4}p_2 - p_1)\cdot(\frac{1}{2} - x) \ge \frac{1}{2}p_1$$

The fact follows by observing that $h(0) = \frac{3}{8}p_2$ and $h(1/2) = \frac{1}{2}p_1$. □

Thus, we have $T_2 = \min_{x \le 1/2 - \delta_1}[h(x)] \ge \min\{\frac{1}{2}p_1, \frac{3}{8}p_2\}$. The term T_3 is analyzed similarly, by defining $g(x) \stackrel{\mathrm{def}}{=} p_1 \cdot \Gamma_{\mathrm{lin}}(x) + p_3 \cdot (1-2x)/2$, and using the following fact.

Fact 2: $\min_{x \le 1/2}\{g(x)\} = \min\{\frac{1}{2}p_3, \frac{1}{2}p_1\} = \min\{g(0), g(1/2)\}$.

proof: by considering two cases and using $\Gamma_{\mathrm{lin}}(x) \ge x$.

case 1: $p_1 \ge p_3$

$$g(x) \ge p_1 x + \frac{1}{2}p_3 - p_3 x = \frac{1}{2}p_3 + (p_1 - p_3)\cdot x \ge \frac{1}{2}p_3$$

case 2: $p_1 \le p_3$

$$g(x) \ge p_1 x + \frac{1}{2}p_3 - p_3 x = \frac{1}{2}p_1 + (p_3 - p_1)\cdot(\frac{1}{2} - x) \ge \frac{1}{2}p_1$$

The fact follows by observing that $g(0) = \frac{1}{2}p_3$ and $g(1/2) = \frac{1}{2}p_1$. □

Thus, we have $T_3 \ge \min_{x \le 1/2 - \delta_2}[g(x)] - \delta \ge \min\{\frac{1}{2}p_1, \frac{1}{2}p_3\} - \delta$. The claim follows. ▌

4.1.2 Main application: the MaxSNP verifier

We are now ready to state the main result of this section. It is a simple verifier for NP which achieves soundness error approaching 85% while performing one of two very simple tests.

Proposition 4.3 (The MaxSNP Verifier): For any $\gamma > 0$ and for any language $L \in \mathrm{NP}$, there exists a verifier V_{SNP} for L such that

- V_{SNP} uses logarithmic randomness and is perfectly complete;
- V_{SNP} has soundness error $\frac{17}{20} + \gamma$; and
- on access to an oracle π (and according to the outcome of the verifier's coin tosses), the verifier V_{SNP} performs one of the following actions:
  1. Parity check: V_{SNP} determines a bit b, makes three queries q_1, q_2 and q_3, and rejects if $\pi(q_1) \oplus \pi(q_2) \oplus \pi(q_3) \neq b$.
  2. RMB check: V_{SNP} determines two bits b_0, b_1, makes three out of four predetermined queries, q_1, q_2, q_3 and q_4, and rejects if either $(\pi(q_1) = 0) \wedge (\pi(q_2) \oplus \pi(q_4) \neq b_0)$ or $(\pi(q_1) = 1) \wedge (\pi(q_3) \oplus \pi(q_4) \neq b_1)$.

     That is, the verifier inspects $\pi(q_1)$ and consequently checks either $\pi(q_2) \oplus \pi(q_4) \overset{?}{=} b_0$ or $\pi(q_3) \oplus \pi(q_4) \overset{?}{=} b_1$.

  Furthermore, the probability (over its coin tosses) that V_{SNP} performs a parity check is $\frac{3}{5}$ (and the probability that V_{SNP} performs a RMB check is $\frac{2}{5}$).

Proof: Set $\delta_1 = \delta_2 = \gamma/2$ and $\epsilon = \frac{\gamma}{2} \cdot (16\delta_1^2\delta_2^2) = \frac{\gamma^5}{2} > 0$. Now, let l and l_1 be integers such that the outer verifier, V_{outer}, guaranteed by Lemma 3.8 is (l, l_1)-canonical and ϵ-good for L. Consider the (l, l_1)-canonical inner verifier V_{SNPinner}, working with the parameters p_1, p_2 and p_3 set to minimize its error. Obviously this calls for setting $\frac{1}{2}p_1 = \frac{3}{8}p_2 = \frac{1}{2}p_3$, which yields

$$p_1 = \frac{3}{10} \;\; ; \;\; p_2 = \frac{4}{10} \;\; ; \;\; p_3 = \frac{3}{10} \tag{12}$$

Let V_{SNP} be the verifier obtained by composing V_{outer} with V_{SNPinner}.

We start by analyzing the soundness error of V_{SNP}. By Lemma 4.1 and Claim 4.2, we know that the inner verifier V_{SNPinner}, with p_i's as in Eq. (12), is $(\rho, \delta_1, \delta_2)$-good, for

$$\begin{aligned} \rho &\leq 1 - \frac{1}{2} \cdot p_3 + \delta_1 \\ &= 1 - \frac{3}{20} + \frac{1}{2} \cdot \gamma \end{aligned}$$

Invoking Theorem 3.12, we upper bound the soundness error of V_{SNP} by $1 - \frac{3}{20} + \frac{1}{2} \cdot \gamma + \frac{\epsilon}{16\delta_1^2\delta_2^2}$ which by the setting of ϵ yields the claimed bound (of $0.85 + \gamma$).

Clearly, V_{SNP} uses logarithmic randomness, has perfect completeness, and its computation on the answers of the oracles are determined by V_{SNPinner}. It is left to observe that each of the three tests (i.e., Linearity, Monomial-Basis, and Projection), performed by V_{SNPinner}, is either a Parity Check or an RMB Check and that the latter occurs with probability 0.4. First observe that with probability p_1, V_{SNPinner} performs $\mathbf{LinTest}(A_{(h,0),(\bar{1},1)}; f_1, f_2)$, where $g_1, g_2 \in \mathcal{F}_l$. Recall that query f to $A_{(h,0),(\bar{1},1)}$ translates to a query in the set $\{f, f+h, f+\bar{1}, f+h+\bar{1}\}$ answered by A and that the answer is possibly complemented (by adding 1 mod 2). Thus, the above linearity test translates to checking the exclusive-or of three values of A against a predetermined bit b (i.e., this bit is determined by the number of times which have shifted a potential query by $\bar{1}$). Similarly, $\mathbf{MBTest}(A_{(h,0),(\bar{1},1)}; f_1, f_2, f_3)$ translates to an RMB Check with b_0, b_1 and the ordering of the second/third function is determined by the folding over $\bar{1}$. Finally, we observe that the projection test, performed by V_{SNPinner}, also amounts to a Parity Check; this time, on answers taken from two different oracles (which can actually be viewed as one oracle). ▌

Remark 4.4 (A tedious one): *The probability that verifier V_{SNP}, of the above proposition, makes two identical queries is negligible. Specifically, it can be made smaller than γ (mentioned in the proposition). Thus, we can ignore this case*[6] *in the next two sections and assume, without loss of generality, that all queries are distinct.*

Implementing the MaxSNP verifier via Gadgets. In the following sections we use the verifier of Proposition 4.3 to obtain hardness results for various variants of MaxSAT as well as for MaxCUT. The hardness results are obtained by constructing an instance of the problem at hand so that the instance represent the verifier's computation on input x. The primary aspect of the reduction is the construction of gadgets which reflect the result of the verifier's computation (i.e., accept/reject) after performing one of the two types of checks, i.e., parity check or RMB check. We define a performance measure for a gadget and then relate the hardness result achieved to the performance measure obtained by the gadgets in use.[7]

Sources of our improvements. The explicit statement of a generic verifier for deriving Max SNP hardness results is a novelty of our paper. Thus, a quantitative comparison to previous works is not readily available. Certainly, we improve over these works thanks to the use of the new LongCode-based inner-verifier, the atomic tests and their analysis in Section 3.5, the new idea of folding, and the improved analysis of linearity testing due to [BCHKS].

4.1.3 Another application: minimizing soundness error in 3-query pcp

As a direct corollary to Proposition 4.3, we obtain

Theorem 4.5 For any $s > 0.85$, $\text{NP} \subseteq \text{PCP}_{1,s}[\,\mathsf{coins} = \log\,;\ \mathsf{query} = 3\,;\ \mathsf{free} = 2\,]$.

4.2 Satisfiability problems

In this section we mainly deal with CNF formulae. However the last subsection deals with formulae consisting of a conjunction of PARITY (rather than OR) clauses. Refer to Section 2.4 for definitions, in particular for what is the problem MaxXSAT and the promise problem Gap-XSAT. Recall that $\text{MaxSAT}(()\varphi)$ is the maximum number of simultaneously satisfiable clauses in formula φ and $\overline{\text{MaxSAT}}(\varphi) = \text{MaxSAT}(\varphi)/\|\varphi\|$ be the normalized version, where $\|\varphi\|$ is the number of clauses in formula φ. See Section 2.4.3 for description of previous work.

[6] Formally, suppose that when it occurs the verifier performs some standard check on fixed different queries. This modification increases the soundness error by at most γ which tends to zero anyhow.

[7] Given that the performance of the various gadgets might be different for the different checks, one might suspect that it might have been a better idea to first construct the gadgets and then to optimize the soundness of V_{SNP} keeping in mind the relative performance measures of the two kinds of gadgets being employed. Surprisingly enough it turns out (cf., [BGS2]) that the optimization is not a function of the performance of the gadgets and indeed the choice of parameters p_1, p_2 and p_3 as in Equation (12) is optimal for the following reductions.

A consequence of the following theorem (apply Proposition 2.5) is that, assuming P $\neq$ NP there is no polynomial time algorithm to approximate: (1) Max3SAT within a factor of 1.038; (2) MaxE3SAT within a factor of 1.038; (3) Max2SAT within a factor of 1.013.

Theorem 4.6 (MaxSAT non-approximability results): The following problems are NP-hard–

(1) $\mathsf{Gap}\text{-}3\mathrm{SAT}_{c,s}$ with $c = 1$ and $s = 26/27$.

(2) $\mathsf{Gap}\text{-}\mathrm{E3SAT}_{c,s}$ with $c = 1$ and $s = 26/27$.

(3) $\mathsf{Gap}\text{-}2\mathrm{SAT}_{c,s}$ for some $0 < s < c < 1$ satisfying $c > 0.9$ and $c/s = 74/73$.

Actually, Items (1) and (2) hold for any $s > 1 - \frac{3}{80}$ whereas Item (3) holds as long as $\frac{c}{s} < 1 + \frac{3}{217}$. Item (1) is implied by Item (2) so we will prove only the latter.

4.2.1 The Hardness of MaxE3SAT and Max2SAT

Gadgets. In the context of MaxSAT problems, we may easily replace a condition of the form $a+b+c = 1$ by $\bar{a}+b+c = 0$, where $\bar{a}$ is the negation of the variable a. Thus, the task of designing gadgets is simplified, and we need to implement two (simplified) types of checks: the Parity Check (checking that $a+b = c$ for a, b and c obtained from the oracle) and the RMB-Check for a, b_0, b_1 and c obtained from the oracle). Accordingly a Parity Check (PC) gadget, $\mathrm{PC}(a, b, c, x_1, x_2, \ldots, x_n)$, is a set of clauses over three *distinguished* variables a, b, c and n auxiliary variables $x_1, \ldots, x_n$. It is an (α, β)-*PC gadget* if the following is true: If $a+b = c$ then $\mathrm{MaxSAT}(\mathrm{PC}(a, b, c, x_1, x_2, \ldots, x_n)) = \alpha$; else it is at most $\alpha - \beta$. Similarly a Respect-Monomial-Basis Check (RMBC) gadget, $\mathrm{RMBC}(a, b_0, b_1, c, x_1, \ldots, x_n)$, is a set of clauses over four *distinguished* variables a, b_0, b_1, c and n auxiliary variables $x_1, \ldots, x_n$. It is an (α, β)-*RMBC gadget* if the following is true: If $b_a = c$ then $\mathrm{MaxSAT}(\mathrm{RMBC}(a, b_0, b_1, c, x_1, x_2, \ldots, x_n)) = \alpha$; else it is at most $\alpha - \beta$. We stress that in both cases the maximum number of clauses which are simultaneously satisfied is at most α. A gadget is said to be a X-SAT gadget if, as a formula, it is a X-SAT formula.

The following lemma describes how gadgets of the above form can be used to obtain the hardness of MaxSAT.

Lemma 4.7 (MaxSAT implementation of a verifier): Let V be a verifier for L of logarithmic randomness, with perfect completeness and soundness s, such that V performs either a single Parity Check (with probability q) or a single RMB check (with probability $1-q$). Furthermore, suppose that in either case, the verifier never makes two identical queries. If there exists an (α_1, β)-Parity-Check X-SAT gadget containing m_1 clauses and an (α_2, β)-RMBC X-SAT gadget containing m_2 clauses then L reduces to $\mathsf{Gap}\text{-}\mathrm{XSAT}_{c',s'}$ for

$$\begin{aligned} c' &= \frac{\alpha_1 q + \alpha_2(1-q)}{m_1 q + m_2(1-q)} \\ s' &= \frac{\alpha_1 q + \alpha_2(1-q) - (1-s)\beta}{m_1 q + m_2(1-q)} \end{aligned}$$

In particular $\frac{c'}{s'} \geq 1 + \frac{(1-s)\beta}{\alpha_1 q + \alpha_2(1-q) - (1-s)\beta}$.

Remark 4.8 *In the above lemma, we have assumed that both the PC and RMBC gadgets have the same second parameter β. This assumption is not really a restriction since we can transform a pair of a (α_1, β_1)-PC gadget and (α_2, β_2)-RMBC gadget into a pair of a $(\alpha_1\beta_2, \beta_1\beta_2)$-PC gadget and a $(\alpha_2\beta_1, \beta_1\beta_2)$-RMBC gadget, thereby achieving this feature. (Actually, what really matters are the fractions α_i/β.)*

Proof: Let $\mathrm{PC}(a, b, c, x_1, \ldots, x_{n_1})$ be the Parity Check gadget and let $\mathrm{RMBC}(a, b, c, d, x_1, \ldots, x_{n_2})$ be the RMBC gadget. We encode V's computation on input x by a CNF formula φ_x. Corresponding to every bit $\pi[q]$ of the proof (oracle) accessed by the verifier V we create a variable $y[q]$. In addition we create some auxiliary variables $y_{\mathrm{Aux}}[R, i]$ for each random string R used by the verifier V and i going from 1 to $\max(n_1, n_2)$. For each such R we will construct a formula φ_R which encodes the computation of the verifier when its coins are R. The union of all these formulae will be our φ_x.

On random string R if the verifier performs a parity check on bits $\pi[q_1], \pi[q_2]$ and $\pi[q_3]$, then φ_R consists of the clauses $PC(y[q_1], y[q_2], y[q_3], y_{\mathrm{Aux}}[R, 1], \ldots, y_{\mathrm{Aux}}[R, n_1])$. On the other hand if the verifier performs a RMB check on bits $\pi[q_1], \pi[q_2], \pi[q_3], \pi[q_4]$, then φ_R consists of the clauses $RMBC(y[q_1], y[q_2], y[q_3], y[q_4], y_{\mathrm{Aux}}[R, 1], \ldots, y_{\mathrm{Aux}}[R, n_2])$.

Let N denote the number of possible random strings used by V. Observe that the number of clauses in φ_x equals $m_1 \cdot qN + m_2 \cdot (1-q)N$. We now analyze the value of $\mathrm{MaxSAT}(\varphi_x)$.

If $x \in L$ then there exists an oracle π such that $V^\pi(x)$ always accepts. Consider the assignment $y[q] = \pi[q]$ (i.e., $y[q]$ is true iff $\pi[q] = 1$). Then for every R, there exists an assignment to the variables $y_{\mathrm{Aux}}[R, i]$ such that the number of clauses of φ_R that are satisfied by this assignment is α_1 if R corresponds to a Parity Check and α_2 if R corresponds to a RMB-check. Since qN of the gadgets are PC-gadgets and $(1-q)N$ of the gadgets are RMBC-gadgets, we have $\mathrm{MaxSAT}(\varphi_x) \geq qN\alpha_1 + (1-q)N\alpha_2$, and the expression for c' follows.

Now consider the case when $x \notin L$. We prove below that if there exists an assignment which satisfies $qN\alpha_1 + (1-q)N\alpha_2 - (1-s)N\beta$ clauses of φ_x, then there exists an oracle π such that $V^\pi(x)$ accepts with probability at least s. Since we know this can not happen we conclude that $\mathrm{MaxSAT}(\varphi_x) < qN\alpha_1 + (1-q)N\alpha_2 - (1-s)N\beta = s'|\varphi_x|$.

To prove the above claim, we convert any assignment to the variables y into an oracle π in the natural way, i.e., $\pi[q] = 1$ iff $y[q]$ is true. Now by the property of the gadgets if a PC gadget $PC(y[q_1], y[q_2], y[q_3], y_{\mathrm{Aux}}[R, 1], \ldots)$ has more than $\alpha_1 - \beta$ clauses satisfied then $\pi[q_1] \oplus \pi[q_2] = \pi[q_3]$. In turn this implies that the verifier V accepts π on random string R. A similar argument can be made about the random strings R which correspond to RMB checks. We also use the property that a PC (resp., RMB) gadget cannot have more than α_1 (resp., α_2) satisfied clauses, even if the claim it checks does hold. Thus, if an assignment satisfies $qN \cdot (\alpha_1 - \beta) + (1-q)N \cdot (\alpha_2 - \beta) + sN\beta$ clauses, then there must exist sN random strings R on which V accepts. This proves the claim and the lemma follows. ▮

Figure 10 describes gadgets which will be used for our MaxE3SAT construction: notice they are *exact*-3-SAT gadgets. We have a $(4, 1)$-PC gadget, PC_3, and a $(4, 1)$-RMB gadget, RMBC_3, each consisting of 4 clauses in which all the clauses have exactly three variables. Both gadgets have no auxiliary variables. The $\mathrm{PC}_3(a, b, c)$ gadget is merely the canonical 3CNF of the expression $a + b + c = 0$. The first two clauses in the $\mathrm{RMBC}_3(a, b, b', c)$ gadget are the canonical 3CNF of the expression $(a = 0) \Rightarrow (b = c)$, whereas the latter two clauses are the canonical 3CNF of the expression $(a = 1) \Rightarrow (b' = c)$. Figure 11 similarly describes 2-SAT gadgets for our Max2SAT construction. We have a $(11, 1)$-PC gadget, PC_2, and a $(11, 1)$-RMB gadget, RMBC_2, each consisting of 12 clauses. Each gadget has four auxiliary variables. The auxiliary variable $x_{\tau\sigma}$ in the PC_2 gadget is supposed to be the indicator of the event $((a = \sigma) \wedge (b = \tau))$. Thus, $a + b = c$ allows to satisfy 11 clauses by appropriately setting the indicator variables (e.g., if $a = b = c = 0$ then setting $x_{00} = 1$ and the other $x_{\tau\sigma}$'s to 0 satisfies all clauses except the last one). The RMBC_2 gadget is composed of two parts; the first six clauses handle the expression $(a = 0) \Rightarrow (b = c)$, whereas the latter six clauses are for the expression $(a = 1) \Rightarrow (b' = c)$.

Lemma 4.9 (SAT gadgets): The following gadgets exist

- E3-SAT gadgets: a $(4,1)$-PC gadget of 4 clauses and a $(4,1)$-RMB gadget of 4 clauses.
- 2-SAT gadgets: a $(11,1)$-PC gadget of 12 clauses and a $(11,1)$-RMB gadget of 12 clauses.

Remark 4.10 *In previous versions of this work [BGS2], it was observed that a ratio of 4 between the number of clauses and the second parameter (i.e., β) is minimal for both E3-SAT gadgets. Several questions regarding the α/β ratios achievable by 3-SAT and 2-SAT gadgets were posed. Answers were subsequently provided in [TSSW], which undertakes a general study of the construction of optimal gadgets.*

Proof of Lemma 4.9: We use the gadgets presented in Figure 10 and Figure 11. The claim regarding E3-SAT follows from the motivating discussion above (i.e., by which these gadgets are merely the canonical 3CNF expressions for the corresponding conditions). Thus, the E3-SAT gadgets are satisfiable if and only if the corresponding condition (i.e., parity or RMB) holds, and the first part of the lemma follows.

We now turn to the 2-SAT gadgets in Figure 11, starting with the PC-gadget $\mathrm{PC}_2(a,b,c,x_{00},x_{01},x_{10},x_{11})$.

- We first claim that if $a+b=c$ then we can satisfy 11 clauses. This is done by setting each $x_{\tau\sigma}$ to 1 if and only if both $a=\sigma$ and $b=\tau$. Clearly, this assignment satisfies the three clauses in which the variable x_{ab} appears (the first two by a and b and the last by x_{ba}). Out of the other 9 clauses, 6 (i.e., those in which an auxiliary variable appears negated) are satisfied by the 0-assignment to the other 3 auxiliary variables, and 2 (i.e., of the 3 in which an auxiliary variable appears unnegated) are satisfied by the variable c.
- We next claim that no assignment for which $a+b=c$ can satisfy all 12 clauses. Let $a=\sigma$, $b=\tau$ and $c=\sigma+\tau$ be an arbitrary partial assignment and consider the three clauses in which the variable $x_{\overline{\tau}\,\overline{\sigma}}$ appears. To satisfy any of the first two clauses we must have $x_{\overline{\tau}\,\overline{\sigma}}=0$ but this cannot satisfy the third clause unless $c\neq\sigma+\tau$, in contradiction to our hypothesis.
- Finally, we show that no assignment for which $a+c\neq c$ can satisfy more than 10 clauses. Let $a=\sigma$, $b=\tau$ and $c=1+\sigma+\tau$ be an arbitrary partial assignment and consider the three clauses in which the variable $x_{\tau\overline{\sigma}}$ appears. To satisfy the first clause we must have $x_{\tau\overline{\sigma}}=0$ but this cannot satisfy the third clause unless $c=\sigma+\tau$, in contradiction to our hypothesis. Applying the same analysis to the clauses in which the variable $x_{\overline{\tau}\sigma}$ appears, the claim follows.

Finally, we consider the RMB-gadget $\mathrm{RMB}_2(a,b,b',c,x_{00},x_{11},y_{00},y_{11})$. This gadget is the conjunction of two analogous 2CNF formulae, each consisting of six clauses. We first consider the first six clauses and the expression $(a=0)\Rightarrow(b=c)$.

The Max-E3-SAT Gadgets.

$$\mathrm{PC}_3(a,b,c)=\{(a\vee b\vee\overline{c}),\ (a\vee\overline{b}\vee c),\ (\overline{a}\vee b\vee c),\ (\overline{a}\vee\overline{b}\vee\overline{c})\}$$

$$\mathrm{RMBC}_3(a,b,b',c)=\{(a\vee b\vee\overline{c}),\ (a\vee\overline{b}\vee c),\ (\overline{a}\vee b'\vee\overline{c}),\ (\overline{a}\vee\overline{b'}\vee c),\ \}$$

Figure 10: *The MaxE3SAT Gadgets*

- Suppose $a = 1$. Then, regardless of the values of b and c, we can satisfy all six clauses by setting $x_{00} = x_{11} = 0$.
- Suppose $a = 0$ and $b = c = \sigma$, for $\sigma \in \{0, 1\}$. Then, we can satisfy five out of the six clauses by setting $x_{\sigma\sigma} = 1$ and $x_{\overline{\sigma}\,\overline{\sigma}} = 0$. On the other hand, it is not possible to satisfy all six clauses, since this requires setting $x_{00} = x_{11} = 1$ (to satisfy the 3rd and 6th clauses), which in turn requires setting both b and $\overline{b}$ to 1.
- Suppose $a = 0$ and $b \neq c$. In this case we claim that no truth assignment can satisfy more than 4 clauses. The claim is proven by contradiction. We already know that no truth assignment can satisfy all clauses. Suppose that some truth assignment satisfies five (or more) clauses. Then, for some $\sigma \in \{0, 1\}$, we must satisfy all clauses in which the variable $x_{\sigma\sigma}$ appears. This requires setting $x_{\sigma\sigma} = 1$ (to satisfy the 3rd or/and the 6th clause), which in turn forces us to set b and c to σ (to satisfy the other two clauses), in contradiction to the case hypothesis.

The last six clauses are analyzed analogously. We conclude that if the RMB condition holds then we can satisfy $6 + 5 = 11$ clauses and that we can satisfy at most 11 clauses. Furthermore, in case the RMB condition does not hold we can satisfy at most $4 + 6 = 10$ clauses. The lemma follows. ▮

Proof of Theorem 4.6: The theorem follows by applying Lemma 4.7 to the verifier of Proposition 4.3 and the gadgets of Lemma 4.9. Details follow.

Recall that by Remark 4.4, we may assume that the verifier does not make two identical queries. Applying Lemma 4.7 to the verifier of Proposition 4.3 we obtain a reduction of any language in NP to $\mathsf{Gap}\text{-XSAT}_{c',s'}$ for values of c' and s' determined as a function of the gadget parameters, the probability parameter q and the soundness s of the verifier of Proposition 4.3. Specifically, we observe that for E3-SAT we have $c' = 1$ (since $\alpha_i = m_i$ for $i = 1, 2$), whereas for 2-SAT we have $0.9 < c' < 1$ (since

The MAX 2SAT Gadgets.

$$\begin{aligned}
&\mathrm{PC}_2(a, b, c, x_{00}, x_{01}, x_{10}, x_{11}) = \\
&\quad \{(\overline{x_{00}} \vee \overline{a}), (\overline{x_{00}} \vee \overline{b}), (x_{00} \vee c), \\
&\qquad (\overline{x_{01}} \vee a), (\overline{x_{01}} \vee \overline{b}), (x_{01} \vee \overline{c}), \\
&\qquad (\overline{x_{10}} \vee \overline{a}), (\overline{x_{10}} \vee b), (x_{10} \vee \overline{c}), \\
&\qquad (\overline{x_{11}} \vee a), (\overline{x_{11}} \vee b), (x_{11} \vee c)\}
\end{aligned}$$

$$\begin{aligned}
&\mathrm{RMBC}_2(a, b, b', c, x_{00}, x_{11}, y_{00}, y_{11}) = \\
&\quad \{(\overline{x_{00}} \vee \overline{b}), (\overline{x_{00}} \vee \overline{c}), (a \vee x_{00}), \\
&\qquad (\overline{x_{11}} \vee b), (\overline{x_{11}} \vee c), (a \vee x_{11}), \\
&\qquad (\overline{y_{00}} \vee \overline{b'}), (\overline{y_{00}} \vee \overline{c}), (\overline{a} \vee y_{00}), \\
&\qquad (\overline{y_{11}} \vee b'), (\overline{y_{11}} \vee c), (\overline{a} \vee y_{11})\}.
\end{aligned}$$

Figure 11: *The Max2SAT Gadgets*

$\frac{\alpha_i}{m_i} = \frac{11}{12}$ for $i = 1, 2$). In both cases, $\beta = 1$ and the expression for c'/s' is given by

$$1 + \frac{1-s}{q\alpha_1 + (1-q)\alpha_2 - (1-s)} \tag{13}$$

where s and q are determined by Proposition 4.3; that is (for every $\gamma > 0$)

$$s = 1 - \frac{3}{20} + \gamma \tag{14}$$

$$q = \frac{3}{5} \tag{15}$$

Substituting Eq. (14) and (15) in Eq. (13), and letting $\gamma \to 0$, we get

$$\frac{c'}{s'} \to 1 + \frac{3}{12\alpha_1 + 8\alpha_2 - 3}.$$

The bounds for E3-SAT and 2-SAT now follow by using the α_i's values of Lemma 4.9. In particular, for E3-SAT we get $s' \to 77/80$ and for 2-SAT we get $\frac{c'}{s'} \to 1 + \frac{3}{217}$. ▌

We conclude this subsection by presenting a variant of Lemma 4.7. This variant refers only to 3SAT formulae, but makes no restrictions on the verifier in the PCP system.

Lemma 4.11 (Max3SAT implementation of a generic verifier): Let $L \in \mathrm{PCP}_{1,1-\delta}[\log, 3]$, for some $0 < \delta < 1$. Then, L reduces to $\mathsf{Gap}\text{-}3\mathrm{SAT}_{1,1-\frac{\delta}{4}}$.

Proof: Let V be a verifier as guaranteed by the hypothesis. Building on Lemma 4.7, it suffices to show that the computation of V on any possible random-tape can be captured by a 3CNF formula with at most 4 clauses. We consider the depth-3 branching program which describes the acceptance of V on a specific random-tape. (The variables in this program correspond to queries that the verifier may make on this fixed random-tape. Since the verifier may be adaptive, different variables may appear on different paths.) In case this tree has at most 4 rejecting leaves (i.e., marked FALSE) writing corresponding 3CNF clauses (which state that these paths are not followed) we are done. Otherwise, we consider the 4 depth-1 subtrees. For each such subtree we do the following. In case both leaves are marked FALSE we write a 2CNF clause (which states that this subtree is not reached at all). In case a single leaf is marked FALSE we write one 3CNF clause (as above), and if no leaf is marked FALSE we write nothing. ▌

Remark 4.12 *The above argument can be easily extended to show that, for any* $0 \le \epsilon, \delta < 1$,

$$\mathrm{PCP}_{1-\epsilon,1-\delta}[\log, 3] \le_D^K \mathsf{Gap}\text{-}\mathit{3SAT}_{1-\epsilon,1-\frac{\delta}{4}}$$

4.2.2 Maximum Satisfiable Linear Constraints (Parity Clauses)

Analogously to the MaxSAT problems considered above, we consider parity/linear clauses rather than disjunctive clauses. In other words, we are given a system of linear equations over GF(2), and need to determine the maximum number of equations which may be simultaneously satisfied. The problem in question is MaxLinEq (cf. Section 2.4.2). See Section 2.4.3 for status and discussion of previous work. Here we provide an explicit hardness factor via a direct reduction from the MaxSNP verifier.

Theorem 4.13 $\mathsf{Gap}\text{-}\mathrm{MaxLinEq}_{c,s}$ is NP-hard for $c = 6/7$ and any $\frac{c}{s} < 8/7$.

Proof: The theorem follows by constructing appropriate gadgets. A PC-gadget is straightforward here and so we have a $(1,1)$-PC gadget. We present a $(3,2)$-RMB gadget consisting of 4 equations. Specifically, for $\mathrm{RMB}(a, b_0, b_1, c)$ we present the equations $b_0 + c = 0$, $a + b_0 + c = 0$, $b_1 + c = 0$ and $a + b_1 + c = 1$. Observe that we can think of the RMB gadget as a $(1.5, 1)$-gadget with 2 clauses (or, equivalently, think of the parity gadget as a $(2,2)$-gadget with 2 clauses).

We claim that the above 4 equations are indeed a $(3,2)$-gadget for $b_a = c$. First observe that if $a = 0$ and $b_0 = c$ (resp., if $a = 1$ and $b_1 = c$) then the first (resp., last) two equations hold. On the other hand, if $a = 0$ and $b_0 \neq c$ (resp., if $a = 1$ and $b_1 \neq c$) then the first (resp., last) two equations are both violated. Finally, if $a = 0$ (resp., if $a = 1$) then, regardless of the values of b_0, b_1, c, exactly one of the last (resp., first) two equations hold. Thus, the claim holds.

Proceeding as in the proof of Theorem 4.6, we obtain a hardness for $\mathsf{Gap}\text{-}\mathrm{MaxLinEq}_{c',s'}$, where

$$\frac{c'}{s'} \to 1 + \frac{3}{12\alpha_1 + 8\alpha_2 - 3} = 1 + \frac{3}{12 + 12 - 3} = \frac{8}{7}$$

and $c' = \frac{3\alpha_1 + 2\alpha_2}{3m_1 + 2m_2} = \frac{6}{7}$. ▮

4.3 MaxCUT

Refer to Section 2.4 for the definition of the MaxCUT problem and the associated gap problem $\mathsf{Gap}\text{-}\mathrm{MaxCUT}_{c,s}$. See Section 2.4.3 for discussion of status and previous work. The following theorem (combined with Proposition 2.5) shows that MaxCUT is NP-hard to approximate to within a factor of 1.014. We note that the result of the following theorem holds also when the weights of the graph are presented in unary (or, equivalently, when considering unweighted graphs with parallel edges).

Theorem 4.14 (MaxCUT non-approximability result): $\mathsf{Gap}\text{-}\mathrm{MaxCUT}_{c,s}$ is NP-hard for some c, s satisfying $c > 0.6$ and $c/s > 1.014$ (anything below 72/71).

A weaker result can be obtained for simple graphs without weights or parallel edges. In particular, one may reduce the MaxCUT problem for graphs with parallel edges to MaxCUT for simple graphs, by replacing every edge by a path of 3 edges. This causes a loss of a factor of 3 in the hardness factor; that is, we would get a hardness factor of 214/213 for the MaxCUT problem restricted to simple graphs. A better reduction which preserves the non-approximation ratio has been recently suggested by Crescenzi et. al. [CST].

Gadgets. Unlike with MaxSAT problem, here we cannot negate variables at zero cost. Still, we first define simplified gadgets for Parity and RMB checking and make the necessary adaptations inside Lemma 4.15.

Gadgets will be used to express the verifier's computation in terms of cuts in graphs. A parity check gadget $\mathrm{PC\text{-}CUT}(a, b, c, T; x_1, \ldots, x_n)$ is a weighted graph on $n+4$ vertices. Of these three vertices a, b, c correspond to oracle queries made by the verifier. The vertex T will be a special vertex mapping cuts to truth values so that a vertex corresponding to an oracle query is considered set to 1 if it resides in the T-side of the cut (i.e., a is considered set to 1 by a cut $(S, \overline{S})$ iff either $a, T \in S$ or $a, T \in \overline{S}$). The gadget is an (α, β)-PC gadget if $\mathrm{MaxCUT}(\mathrm{PC\text{-}CUT}(a, b, c, T; x_1, \ldots, x_n))$ is exactly α when restricted to cuts which induce $a + b = c$ (i.e., either 0 or 2 of the vertices $\{a, b, c\}$ lie on the same side of the cut as T), and is at most $\alpha - \beta$ when restricted to cuts for which $a + b \neq c$. A cut gadget to check if $a + b \neq c$ can be defined similarly. Similarly a weighted graph $\mathrm{RMBC\text{-}CUT}(a, b_0, b_1, c, T; x_1, \ldots, x_n)$ is an (α, β)-RMBC gadget if it satisfies the property that $\mathrm{MaxCUT}(\mathrm{RMBC\text{-}CUT}(a, b_0, b_1, c, T; x_1, \ldots, x_n))$ is exactly α

when restricted to cuts satisfying $b_a = c$ and is at most $\alpha - \beta$ otherwise. Cut gadgets for the other generalized RMB checks (checking if $b_a \neq c$, or $b_a = c + a$ or $b_a \neq c + a$) can be defined similarly. The following lemma (similar to Lemma 4.7) shows how to use the above forms of gadgets to derive a reduction from NP to Gap-MaxCUT.

Lemma 4.15 (MaxCUT implementation of a verifier): Let V be a verifier for L of logarithmic randomness, with perfect completeness and soundness s, such that V performs either a single Parity Check (with probability q) or a single RMB check (with probability $1 - q$). Here, we refer to the generalized checks as defined in Proposition 4.3. Furthermore, suppose that in either case, the verifier never makes two identical queries. If there exists an $(\alpha_1 - \beta, \beta)$-PC gadget consisting of edges of total weight w_1 and an $(\alpha_2 - \beta, \beta)$-RMBC gadget consisting of edges of total weight w_2 then L reduces to Gap-MaxCUT$_{c',s'}$ for $c' = \frac{\alpha_1 q + \alpha_2(1-q)}{w_1 q + w_2(1-q)}$ and $s' = \frac{\alpha_1 q + \alpha_2(1-q) - (1-s)\beta}{w_1 q + w_2(1-q)}$. In particular $c'/s' \geq 1 + \frac{(1-s)\beta}{\alpha_1 q + \alpha_2(1-q) - (1-s)\beta}$.

Remark 4.16 *Actually, the conclusion of the lemma holds provided all the* generalized *parity check (resp., RMB-check) functions have (α_1, β)-gadgets (resp., (α_2, β)-gadgets).*

Proof: Let PC-CUT$(a, b, c, T, x_1, \ldots, x_{n_1})$ denote the Parity Check gadget and RMBC-CUT$(a, b_0, b_1, c, T, x_1, \ldots, x_{n_2})$ denote the RMBC gadget. These are simplified gadgets as defined above. Increasing the α value by β, we can easily obtain the general gadgets as defined in Proposition 4.3. For example, to check that $a + b + c = 1$ we introduce a gadget which in addition to the variables $a, b, c, T, x_1, \ldots, x_{n_1})$, has an auxiliary vertex, denoted $\bar{a}$. The new gadget consists of the edge $(a, \bar{a})$ having weight β together with the weighted graph PC-CUT$(\bar{a}, b, c, T, x_1, \ldots, x_{n_1})$. Clearly, the result is an (α_1, β)-gadget for $a + b + c = 1$. Likewise we can check the condition $b_a + c = \tau_a$, where τ_0, τ_1 are any fixed bits as follows. In case $\tau_0 = \tau_1 = 1$ we introduce an auxiliary vertex $\bar{c}$, connect it to c by an edge of weight β and use the graph RMBC-CUT$(a, b_0, b_1, \bar{c}, T, x_1, \ldots, x_{n_2})$. In case $\tau_0 = 0$ and $\tau_1 = 1$ we introduce an auxiliary vertex $\bar{b}_1$, connect it to b_1 by an edge of weight β and use the graph RMBC-CUT$(a, b_0, \bar{b}_1, c, T, x_1, \ldots, x_{n_2})$. The case $\tau_0 = 1$ and $\tau_1 = 0$ is analogous, whereas $\tau_0 = \tau_1 = 0$ is obtained by the simplified gadget itself. Thus, we have (α_2, β)-gadgets for all cases of the RMB Check. Throughout the rest of the proof, PC-CUT and RMBC-CUT denote the generalized gadgets.

We create a graph G_x and weight function w_x which encodes the actions of the verifier V on input x. The vertices of G_x are as follows:

(1) For every bit $\pi[q]$ of the proof queried by the verifier V, the graph G_x has a vertex $v_{\pi[q]}$.
(2) For every random string R tossed by the verifier V, we create vertices $v_{R,i}$, for i going from 1 to $\max\{n_1, n_2\}$.
(3) There will be one special vertex T.

The edges of G_x are defined by the various gadgets. We stress that the same edge may appear in different gadgets (and its weight in these gadgets may be different). The graph G_x is defined by taking all these edges and thus it is a graph (or multi-graph) with parallel edges and weights. The natural conversion of G_x into a graph with no parallel edges replaces the parallel edges between two vertices with a single edge whose weight is the sum of the weights of the original edges. Alternatively, since the weights are constants which do not depend on x, we can transform G_x into a unweighted graph with parallel edges.

Suppose that on random string R the verifier V queries the oracle for bits $\pi[q_1]$, $\pi[q_2]$ and $\pi[q_3]$, and then does a parity check on these three bits. Then corresponding to this random string we add the weighted edges of the graph G_R to the graph G_x where $G_R =$ PC-CUT$(v_{\pi[q_1]}, v_{\pi[q_2]}, v_{\pi[q_3]}, T; v_{R,1}, \ldots, v_{R,n_1})$. Alternatively, if the verifier V performs a respect of monomial basis test on the bits

$\pi[q_1]$, $\pi[q_2]$, $\pi[q_3]$ and $\pi[q_4]$, then we add the weighted edges of the graph $G_R = \text{RMBC-CUT}(v_{\pi[q_1]}, v_{\pi[q_2]}, v_{\pi[q_3]}, v_{\pi[q_4]}, T; v_{R,1}, \ldots, v_{R,n_2})$.

Let N denote the number of possible random strings used by V. Observe that the total weight of the edges of G_x is $w_1 qN + w_2(1-q)N$. We now analyze the value of $\text{MaxCUT}(G_x)$.

If $x \in L$ then there exists an oracle π such that $V^\pi(x)$ always accepts. We define a cut $(S, \bar{S})$ in G_x in the following way: We place $T \in S$ and for every query q we place $v_{\pi[q]} \in S$ iff $\pi[q] = 1$. Then for each R, there exists an placement of the vertices $v_{R,i}$ so that the size of the cut induced in G_R is α_1 if R corresponds to V performing a Parity Check and α_2 if R corresponds to V performing an RMB check. The weight of the so obtained cut is $\alpha_1 qN + \alpha_2(1-q)N$.

Now consider $x \notin L$. We claim that if there exists a cut $(S, \bar{S})$ such that the weight of the cut is greater than $qN\alpha_1 + (1-q)N\alpha_2 - (1-s)N\beta$, then there exists an oracle π, such that $V^\pi(x)$ accepts with probability at least s. Since we know this can not happen we conclude that $\text{MaxCUT}(G_x) < qN\alpha_1 + (1-q)N\alpha_2 - (1-s)N\beta$. To prove the claim, we convert any cut in G_x into an oracle π where $\pi[q] = 1$ iff T and $v_{\pi[q]}$ lie on the same side of the cut. Now by the property of the gadgets if a graph $G_R = \text{PC-CUT}(y[q_1], y[q_2], y[q_3], T; x_1, \ldots, x_{n_1})$ contributes more than a weight of $\alpha_1 - \beta$ to the cut, then V accepts π on random string R. (Similarly if the graph G_R is an RMBC-gadget and contributes more than $\alpha_2 - \beta$ to the cut then V accepts π on random string R.) Recall that no gadget can contribute more than the corresponding α to any cut. Thus if the total weight of the cut is more than $(\alpha_1 - \beta)qN + (\alpha_2 - \beta)(1-q)N + sN \cdot \beta$, then V accepts on at least sN random strings. This proves the claim and the lemma follows. ▌

We now turn to the construction of cut-gadgets. Our first gadget, denoted $\text{PC-CUT}(a, b, c, T; \text{AUX})$, is a complete graph defined on five vertices $\{a, b, c, T, \text{AUX}\}$. The weight function, w, assign to the edge $\{u, v\}$ weight $w_u \cdot w_v$, where $w_a = w_b = w_c = w_T = 1$ and $w_{\text{AUX}} = 2$. The following claim shows how $\text{PC-CUT}(a, b, c, T; \text{AUX})$ functions as a parity check gadget.

Claim 4.17 (MaxCUT PC-gadget): $\text{PC-CUT}(a, b, c, T; \text{AUX})$ is a $(9, 1)$-parity check gadget consisting of edges of total weight 14.

Proof: Recall that the edges in the graph are of two types: (1) edges to AUX having weight 2; and (2) other edges having weight 1. Thus, the total weight of the edges is $4 \cdot 2 + 6 \cdot 1 = 14$. The weight function is decomposed as a product of vertices "weights" and so we can express the weight of a cut $(S, \overline{S})$ by the corresponding product $(\sum_{u \in S} w_u) \cdot (\sum_{v \in \overline{S}} w_v)$. It turns out that the weight of a cut is maximized when the weight of the vertices on both sides are equal and specifically equal $\frac{6}{2} = 3$. Thus, the maximum cut has weight $3^2 = 9$. Furthermore, a max-cut must have AUX and exactly one of the other vertices on one side. On the other hand, all other cuts (i.e., in which the vertex weights are not split evenly) have weight at most 8. Using the above characterization of a max-cut we conclude that the max-cut may have one of the two forms:

(1) AUX resides in the same side with T: since a, b and c are on the other side, the induced assignment is $a = b = c = 0$ which satisfies the parity condition.

(2) AUX resides in the same side with $x \in \{a, b, c\}$: this induces $x = 0$ and an assignment of 1 to the other two variables and thus the parity condition is satisfied again.

Thus a max-cut corresponds to an assignment which satisfies the parity condition and each such assignment (can be extended to) corresponds to a max-cut. The claim follows. ▌

The second gadget, denoted $\text{RMBC-CUT}(a, b_0, b_1, c, T; \text{AUX}_1, \text{AUX}_2, \text{AUX}_3, a')$, is composed of two graphs denoted G_1 and G_2, respectively. To motivate the construction we first observe that the condition $b_a = c$ (i.e., $(a = 0) \Rightarrow (b_0 = c)$ and $(a = 1) \Rightarrow (b_1 = c)$) is equivalent to the conjunction of

$(b_0 = b_1) \Rightarrow (b_0 = c)$ and $(b_0 \neq b_1) \Rightarrow (a+b_0+c = 0)$. The graph $G_1(b_0, b_1, c; \textsc{Aux}_1)$ will take care of the first implication. It consists of the vertex set $\{b_0, b_1, c, \textsc{Aux}_1\}$, the unit-weight edges $\{b_0, \textsc{Aux}_1\}$ and $\{b_1, \textsc{Aux}_1\}$, and a weight 2 edge $\{c, \textsc{Aux}_1\}$. The graph $G_2(a, b_0, b_1, c, T; \textsc{Aux}_2, \textsc{Aux}_3, a')$, taking care of the second implication, consists of two subgraphs PC-CUT$(a, b_0, c, T; \textsc{Aux}_2)$ and $\overline{\text{PC-CUT}}(a, b_1, c, T; \textsc{Aux}_3, a')$, where the latter is supposed to "check" $a + b_1 + c = 1$. Specifically, $\overline{\text{PC-CUT}}(a, b, c, T; \textsc{Aux}, a')$ consists of the graph PC-CUT$(a', b, c, T; \textsc{Aux})$ and a unit-weight edge $\{a, a'\}$. The following claim shows exactly how good this gadget is in "verifying" that $b_a = c$.

Claim 4.18 (MaxCUT RMB-gadget): RMBC-CUT$(a, b_0, b_1, c, T; \textsc{Aux}_1, \textsc{Aux}_2, \textsc{Aux}_3, a')$ is a $(22, 2)$-RMBC gadget consisting of edges of total weight 33.

Proof: Clearly, the total edge weight is $4 + 14 + (14 + 1) = 33$. We analyze the performance of each of the two sub-gadgets, G_1 and G_2, considering three cases. Recall that each of two sub-gadgets corresponds to a condition of the form $E \Rightarrow E'$, where both E and E' are linear conditions on two/three variables. The first case corresponds to both E and E' being satisfied (i.e., "good case"), the second case (called "neutral") corresponds to E not being satisfied, whereas the third case (called "bad") corresponds to E being satisfied and E' being violated. We start with G_1.

Fact 1: Consider the set of all cuts in $G_1(b_0, b_1, c; \textsc{Aux})$.

(1) *Good Case* ($b_0 = b_1 = c$): If b_0, b_1 and c are all in same side of the cut then we can place Aux so that the cut has weight 4. On the other hand, there is no cut with weight more than 4.

(2) *Neutral Case* ($b_0 \neq b_1$): If b_0 and b_1 are on opposite sides of the cut, we can always place Aux so that the weight of the cut is 3. On the other hand, 3 is the maximum cut-weight for such cuts.

(3) *Bad Case* ($b_0 = b_1 \neq c$): If b_0 and b_1 are the same side of the cut and c is on the opposite side then, no matter where Aux is placed, the cut-weight is 2.

proof: The lower bounds for Items (1) and (2) are proven by placing Aux on the opposite side to c. The upper bounds for Items (1) and (2) are obvious (since 4 is the total edge weight in G_1 and since placing b_0 and b_1 on opposite sides does not allow placing Aux opposite to both of them). Item (3) is obvious as in each of the two cases we get a cut of weight 2. $\square$

Fact 2: Consider the set of all cuts in $G_2(a, b_0, b_1, c, T; \textsc{Aux}, \textsc{Aux}', a')$.

(1) *Good Case* ($b_0 \neq b_1$ and $a + b_0 + c = 0$): If b_0 and b_1 are on opposite sides and $a + b_0 + c = 0$ then we can place Aux, Aux$'$ so that the cut has weight 19. On the other hand, there is no cut with weight more than 19.

(2) *Neutral Case* ($b_0 = b_1$): If b_0 and b_1 are on the same side of the cut, we can place Aux, Aux$'$ so that the weight of the cut is 18. On the other hand, 18 is the maximum cut-weight for such cuts.

(3) *Bad Case* ($b_0 \neq b_1$ and $a + b_0 + c \neq 0$): If b_0 and b_1 are on opposite sides and $a + b_0 + c \neq 0$ then 17 is the maximum cut-weight for such cuts.

proof: Recall that $G_2(a, b_0, b_1, c, T; \textsc{Aux}, \textsc{Aux}', a')$ consists of the subgraphs PC-CUT$(a, b_0, c, T; \textsc{Aux})$ and PC-CUT$(a', b_1, c, T; \textsc{Aux}')$, and the edge $\{a, a'\}$.

- Item (1) follows by Claim 4.17 (where for the lower bound we place a' opposite to a and use $a' + b_1 = a + b_0$).
- The upper bound of Item (2) follows from Claim 4.17 by first observing that if both $a + b_0 + c = 0$ and $a' + b_1 + c = 0$ then $a = a'$ (since in this case $b_0 = b_1$). Thus, either we obtain maximum weight of 9 in both PC gadgets (and lose the edge $\{a, a'\}$) or we do not obtain the weight 9 in both PC gadgets – either way the bound follows.

- The lower bound of Item (2) follows by first observing that when we place a' opposite to a, either $a + b_0 + c = 0$ or $a' + b_1 + c = 0$ holds. Extending the argument of Claim 4.17, we next observe that if the parity condition is not satisfied we can still place the auxiliary vertex to obtain a cut of weight 8. Thus, we obtain a cut of weight $1 + 8 + 9 = 18$ as claimed.
- Item (3) follows from Claim 4.17 by first observing that if $b_0 \neq b_1$ and $a + b_0 + c \neq 0$ then $a + b_1 + c = 0$. Thus, PC-CUT$(a, b_0, c, T; \textsc{Aux})$ contributes at most weight 8 to the cut (use Claim 4.17) whereas either $a = a'$ or PC-CUT$(a', b_1, c, T; \textsc{Aux}')$ also contributes at most 8. As above, in the former case (i.e., $a = a'$) the edge $\{a, a'\}$ is not in the cut. Thus in either cases the maximum cut weight is 17 (obtained by either $2 \cdot 8 + 1$ or $8 + 9$).

This concludes the proof of Fact 2. □

The Claim now follows by combining the two facts. First recall that the RMB condition is equivalent to the conjunction of $(b_0 = b_1) \Rightarrow (b_0 = c)$ and $(b_0 \neq b_1) \Rightarrow (a + b_0 + c = 0)$. If the RMB condition holds then we obtain the Good Case weight from one sub-gadget, say G_i, and the Neutral Case weight from the other (i.e., G_{3-i}). (The value of $i \in \{1, 2\}$ depends on whether $b_0 = b_1$ or not.) Thus, the total weight equals 22 (obtained by either $4 + 18$ or $3 + 19$). If the RMB condition does not hold then we obtain the Bad Case weight from one sub-gadget and the Neutral Case weight from the other. Thus, the total weight equals 20 (obtained by either $2 + 18$ or $3 + 17$). The claim follows. ▮

Proof of Theorem 4.14: The theorem follows by combining Proposition 4.3, Lemma 4.15, Claim 4.17 and Claim 4.18 (when regarding the RMB gadget as a $(11, 1)$-gadget rather than a $(22, 2)$-gadget). Details follows.

As in the proof of Theorem 4.6, when applying Lemma 4.15 to the verifier in Proposition 4.3, we obtain the same expression for the gap, c'/s', for which NP $\leq_D^K$ Gap-MaxCUT$_{c',s'}$; namely,

$$\begin{aligned}\frac{c'}{s'} &\rightarrow 1 + \frac{(1-s)\beta}{q \cdot \alpha_1 + (1-q) \cdot \alpha_2 - (1-s)\beta} \\ &= 1 + \frac{3}{12\alpha_1 + 8\alpha_2 - 3}.\end{aligned}$$

Recall that here $\alpha_1 - 1 = 9$ and $\alpha_2 - 1 = 11$ (rather than $\alpha_1 = 9$ and $\alpha_2 = 11$ – see Lemma 4.15). The above simplifies to $1 + \frac{3}{213} = \frac{72}{71}$ and the bound on $\frac{c'}{s'}$ follows. As for c', it equals $\frac{3\alpha_1 + 2\alpha_2}{3m_1 + 2m_2} > 0.6$. ▮

5 Free bits and vertex cover

It is known that approximating the minimum vertex cover of a graph to within a $1 + \epsilon$ factor is hard, for some $\epsilon > 0$ [PaYa, ALMSS]. However, we do not know of any previous attempt to provide a lower bound for ϵ. An initial attempt may use VC-gadgets that implement the various tests in V_{SNPinner}, analogously to the way it was done in the previous sections for the Max SAT versions and Max Cut. This yields a lower bound of $\epsilon > \frac{1}{43} > 0.023$ (see details in previous versions of this work [BGS2]). However, a stronger result is obtained via free-bit complexity:[8] We apply the FGLSS-reduction to a proof system (for NP) of low free-bit complexity; specifically to a proof system which uses 2 free-bits and has soundness error below 0.8. Consequently, the clique size, in case the original input is in the language, is at least one fourth (1/4) of the size of the graph which means that translating clique-approximation factors to VC-approximation factors yields only a loss of a factor of

[8] Furthermore, there seems to be little hope that the former approach can ever yield an improvement over the better bounds subsequently obtained by Håstad [H3].

The Enhanced RMB Test. Again, $A\colon \mathcal{F}_l \to \Sigma$ is the object being tested, and the test take additional inputs or parameters $f_1, f_2 \in \mathcal{F}_l$.

EMBTest$(A; f_1, f_2)$ (Enhanced Monomial-Basis Test)
For every $f \in \mathcal{F}_l$, invoke **MBTest**$(A; f_1, f, f_2)$.
Output 0 if all invocations answered with 0, else output 1.

The Passing Probability:

$$\mathrm{EMBP{\small ASS}}(A) \;=\; \Pr_{f_1,f_2 \overset{R}{\leftarrow} \mathcal{F}_l}[\,\mathbf{EMBTest}(A; f_1, f_2) = 0\,]$$

Figure 12: *The Enhanced RMB test and its passing probability.*

3. Since the FGLSS-transformation translates the completeness/soundness ratio to the gap-factor for approximating clique, our first goal is to construct for NP a proof system which uses two free-bits and has soundness error as low as possible. We remark that the proof system of Section 6 uses 7 free-bits and achieves soundness error less than 1/2. The reader may observe that, following the above approach, it is not worthwhile to use the proof system of Section 6 or any proof systems which achieves a soundness error of 1/2 at the cost of 5 free-bits or more. On the other hand, in light of the results of Section 10, we cannot hope for a proof system of free-bit complexity 1 for NP.

5.1 Minimizing the error achievable with two free bits

The pcp system of Proposition 4.3 had free-bit complexity 2 (and query-complexity 3). However, a smaller soundness error can be achieved if we make more queries. Our starting point is Part (2) of Lemma 3.19 which suggests an RMB-test with a detection probability that is twice as big, still using 2 free-bits (alas $2^l + 2$ rather than 3 queries). Specifically, we consider an *enhanced* RMB test which on input $f_1, f_2 \in \mathcal{F}_l$, goes over all $f \in \mathcal{F}_l$ invoking the Atomic RMB test with input functions f_1, f, f_2. The enhanced RMB Test, denoted **EMBtest**, is depicted in Figure 12. Further improvement is obtained by "packing" together the Linearity Test and the Enhanced RMB Test (in contrast to V_{SNPinner} in which these tests were performed exclusively). Both tests make three queries of which two are common, and the answers to these queries determine the answer to the third query (which is different in the two tests). The resulting inner verifier, denoted V_{2inner}, is depicted in Figure 13. As V_{SNPinner}, the verifier V_{2inner} works with functions/oracles A that are folded twice — once across $(h, 0)$ and once across $(\bar{1}, 1)$.

The following corollary is immediate from Part (2) of Lemma 3.19.

Corollary 5.1 (analysis of the Enhanced Monomial-Basis Test): Let $A, \tilde{A}\colon \mathcal{F}_l \to \Sigma$ with A satisfying $A(f + \bar{1}) = A(f) + 1$ for all f and $\tilde{A}$ linear but not respecting the monomial basis. Let $x = \mathrm{Dist}(A, \tilde{A})$. Then

$$1 - \mathrm{EMBP{\small ASS}}(A) \;\geq\; \frac{3}{4} \cdot (1 - 2x)$$

The following lemma is analogous to Lemma 4.1. Loosely speaking, it considers three possible strategies of a "dishonest" prover and indicates the probability with which the verifier detects an error.

Lemma 5.2 (soundness of V_{2inner}): Let $\delta_1, \delta_2 > 0$, $0 \le p \le 1$ and $l, l_1 \in \mathcal{Z}^+$. Then the (l, l_1)-canonical inner verifier V_{2inner} (with parameter p) is $(\rho, \delta_1, \delta_2)$-good, where $1 - \rho = \min(T_1, T_2, T_3)$ and

(1) $T_1 \overset{\text{def}}{=} (\frac{1}{2} - \delta_1) \cdot p$

(2) $T_2 \overset{\text{def}}{=} p \cdot \min_{x \le 1/2 - \delta_1} [\, \max(\Gamma_{\text{lin}}(x)\, ,\, \frac{3}{4} \cdot (1 - 2x))\,]$

(3) $T_3 \overset{\text{def}}{=} \min_{x \le 1/2 - \delta_1} [\, p \cdot \Gamma_{\text{lin}}(x) + (1 - p) \cdot (\frac{1}{2} - \delta_2)(1 - 2x)\,]$.

Proof: The analysis is broken up into several cases as in the proof of Lemma 4.1. Let $x = \text{Dist}(A_{(h,0),(\bar{1},1)}, \textsc{Lin})$.

Case 1: $x \ge 1/2 - \delta_1$. Lemma 3.15 implies that $1 - \textsc{LinPass}(A_{(h,0),(\bar{1},1)}) \ge \Gamma_{\text{lin}}(x) \ge x \ge 1/2 - \delta_1$. Since V_{2inner} performs the atomic linearity test with probability p, we have in this case

$$1 - \texttt{ACC}\,[\, V_{\text{2inner}}^{A,A_1}(\sigma, h)\,] \ge p \cdot (1/2 - \delta_1)$$

Case 2: $x < 1/2 - \delta_1$. Again, Lemma 3.15 implies that $1 - \textsc{LinPass}(A_{(h,0),(\bar{1},1)}) \ge \Gamma_{\text{lin}}(x)$ and

$$1 - \texttt{ACC}\,[\, V_{\text{2inner}}^{A,A_1}(\sigma, h)\,] \;\ge\; p \cdot \Gamma_{\text{lin}}(x)$$

follows. Now let $\tilde{A}$ be a linear function such that $\text{Dist}(A_{(h,0),(\bar{1},1)}, \tilde{A}) = x$. We consider the following sub-cases.

Case 2.1: $\tilde{A}$ does not respect the monomial basis. In this case Corollary 5.1 implies that $1 - \textsc{EMBPass}(A_{(h,0),(\bar{1},1)}) \ge \frac{3}{4}(1 - 2x)$. So the probability that V_{2inner} rejects is at least $p \cdot \frac{3}{4}(1 - 2x)$.

The two free-bit inner verifier. Given functions $h \in \mathcal{F}_l$ and $\sigma \colon \Sigma^l \to \Sigma^{l_1}$, the verifier has access to oracles for $A \colon \mathcal{F}_l \to \Sigma$ and $A_1 \colon \mathcal{F}_{l_1} \to \Sigma$. In addition it takes a parameter $p \in [0, 1]$.

```
Pick $q \overset{R}{\leftarrow} [0,1]$.

Case: $q \le p$ :
    Pick $f_1, f_2 \overset{R}{\leftarrow} \mathcal{F}_l$.
    **LinTest**$(A_{(h,0),(\bar{1},1)}; f_1, f_2)$.
    **EMBTest**$(A_{(h,0),(\bar{1},1)}; f_1, f_2)$.

Case: $q > p$ :
    Pick $f \overset{R}{\leftarrow} \mathcal{F}_l$ and $g \overset{R}{\leftarrow} \mathcal{F}_{l_1}$.
    **ProjTest**$_\sigma(A_{(h,0),(\bar{1},1)}, A_1; f, g)$.
```

Remark: access to $A_{(h,0),(\bar{1},1)}(f)$ is implemented by accessing either $A(f)$ or $A(f + h)$ or $A(f + \bar{1})$ or $A(f + h + \bar{1})$.

Figure 13: *The two free-bit inner verifier* V_{2inner}

Combining the two lower bounds on $1 - \texttt{ACC}\,[\,V_{\text{2inner}}^{A,A_1}(\sigma,h)\,]$, we get

$$1 - \texttt{ACC}\,[\,V_{\text{2inner}}^{A,A_1}(\sigma,h)\,] \;\geq\; p \cdot \max(\Gamma_{\text{lin}}(x), \frac{3}{4}(1-2x))$$

Case 2.2: $\tilde{A}$ respects the monomial basis. By Proposition 3.2, $\tilde{A}$ is an evaluation operator. So there exists $a \in \Sigma^l$ such that $\tilde{A} = E_a$. So $\text{Dist}(A_{(h,0),(\bar{1},1)}, E_a) = x$. Let $a_1 = \sigma(a)$. The proof splits into two further sub-cases.

Case 2.2.1: $d \stackrel{\text{def}}{=} \text{Dist}(A_1, E_{a_1}) \geq 1/2 - \delta_2$. By Lemma 3.21 we have $1 - \textsc{ProjPass}_\sigma(A_{(h,0),(\bar{1},1)}, A_1) \geq d \cdot (1-2x) \geq (1/2 - \delta_2)\cdot(1-2x)$. So the probability that V_{2inner} performs the projection test and rejects is at least $(1-p)\cdot(1/2-\delta_2)(1-2x)$. To this we add the probability of the exclusively disjoint event in which the verifier performs the Linearity Test and rejects, obtaining

$$1 - \texttt{ACC}\,[\,V_{\text{2inner}}^{A,A_1}(\sigma,h)\,] \;\geq\; p \cdot \Gamma_{\text{lin}}(x) + (1-p)\cdot(1/2-\delta_2)(1-2x)$$

Case 2.2.2: Else, we have $x = \text{Dist}(A_{(h,0),(\bar{1},1)}, E_a) < 1/2 - \delta_1$ and $\text{Dist}(A_1, E_{a_1}) < 1/2 - \delta_2$. Thus the functions $A_{(h,0),(\bar{1},1)}$ and A_1 satisfy conditions (2.1) and (2.2) in Definition 3.9.

Similarly to the proof of Lemma 4.1, we infer that the lower bound on $1-\rho$ is as claimed and the lemma follows. ▌

We now simplify the soundness bound of the lemma. The proof of the first item uses the fact that $\Gamma_{\text{lin}}(x) \geq 45/128$ for all $x \geq 1/4$. The second item uses the fact that $\Gamma_{\text{lin}}(x) \geq x$ for all $x \leq 1/2$.

Claim 5.3 :
(1) $\min_{x \leq 1/2-\delta_1} [\max(\Gamma_{\text{lin}}(x)\,,\, \frac{3}{4}(1-2x))] \geq \frac{45}{128}$.
(2) $\min_{x \leq 1/2-\delta_1} [\, p\cdot\Gamma_{\text{lin}}(x) + (1-p)\cdot(\frac{1}{2}-x)\,] \geq \frac{1}{2}\cdot\min(p\,,\,1-p)$.
(3) Let T_1, T_2 and T_3 be as in Lemma 5.2. Then

$$\min(T_1,T_2,T_3) \;\geq\; \min\left\{\frac{45}{128}\cdot p\;,\;\frac{1}{2}\cdot(1-p)\right\} - \max(\delta_1,\delta_2)$$

Interestingly, the lower bound provided by Item (3) is tight. Optimization calls for setting $\frac{45}{128}\cdot p = \frac{1}{2}\cdot(1-p)$, which yields $p = \frac{64}{109}$ and a soundness bound of $1 - \frac{45}{128}p + \max(\delta_1,\delta_2) = 1 - \frac{45}{218} + \max(\delta_1,\delta_2)$.

Proof: Towards proving Part (1) we consider two cases.

Case 1.1: $x \geq 1/4$. In this case, by definition of Γ_{lin}, we have

$$\max(\Gamma_{\text{lin}}(x)\,,\,\frac{3}{4}(1-2x)) \;\geq\; \Gamma_{\text{lin}}(x) \;\geq\; \frac{45}{128}$$

Case 1.2: $x \leq 1/4$. In this case we have

$$\max(\Gamma_{\text{lin}}(x)\,,\,\frac{3}{4}(1-2x)) \;\geq\; \frac{3}{4}(1-2x) \;\geq\; \frac{3}{8} \;>\; \frac{45}{128}$$

This establishes Part (1). Towards proving Part (2) we consider two different cases.

Case 2.1: $p \leq (1-p)$. In this case

$$p\cdot\Gamma_{\text{lin}}(x) + (1-p)\cdot(\frac{1}{2}-x) \;\geq\; p\cdot x + p\cdot(\frac{1}{2}-x) \;=\; \frac{p}{2}$$

Case 2.2: $p \geq (1-p)$. In this case

$$p \cdot \Gamma_{\text{lin}}(x) + (1-p) \cdot (\frac{1}{2} - x) \;\geq\; (1-p) \cdot x + (1-p) \cdot (\frac{1}{2} - x) \;=\; \frac{1-p}{2}$$

This establishes Part (2). To prove Part (3) use Parts (1) and (2) to lower bound T_2 and T_3, respectively, and get

$$\begin{aligned} \min(T_1, T_2, T_3) &\geq \min\left\{(\frac{1}{2} - \delta_1) \cdot p \,,\, \frac{45}{128} \cdot p \,,\, \frac{1}{2} \cdot \min(p\,,\, 1-p) - \delta_2\right\} \\ &\geq \min\left\{\frac{45}{128} \cdot p \,,\, \frac{1}{2} \cdot (1-p)\right\} - \max(\delta_1, \delta_2) \end{aligned}$$

The claim follows. ∎

Composing the above inner verifier with an adequate outer verifier, we get

Theorem 5.4 $\text{NP} \subseteq \text{FPCP}_{1,s}[\log, 2]$, and furthermore, there is a constant q such that $\text{NP} \subseteq \text{PCP}_{1,s}[\,\mathsf{coins} = \log\,;\, \mathsf{free} = 2\,;\, \mathsf{query} = q\,]$, for any $s > \frac{173}{218} \approx 0.79357798$.

Proof: Let $\delta = s - \frac{173}{218}$, $\delta_1 = \delta_2 = \delta/3$ and $\epsilon = \frac{\delta}{3} \cdot 16\delta_1^2\delta_2^2 = \frac{16\delta^5}{243}$. Now, let l and l_1 be integers such that the outer verifier, V_{outer}, guaranteed by Lemma 3.8, is (l, l_1)-canonical and ϵ-good for $L \in \text{NP}$. Consider the (l, l_1)-canonical inner verifier $V_{2\text{inner}}$ working with parameter $p = 64/109$. Using Lemma 5.2 and Claim5.3, we conclude that $V_{2\text{inner}}$ is (ρ, δ, δ)-good for $\rho = 1 - \frac{45}{218} + \max(\delta_1, \delta_2)$.

Composing V_{outer} and $V_{2\text{inner}}$ we obtain a verifier, $V_{2\text{free}}$, which by Theorem 3.12 has soundness error bounded above by $\frac{173}{218} + \max(\delta_1, \delta_2) + \frac{\epsilon}{16\delta_1^2\delta_2^2} = s$, as required. Furthermore, $V_{2\text{free}}$ uses logarithmically many coins. We claim that $V_{2\text{free}}$ has query complexity $2^l + 2$ and free-bit complexity 2. The claim is obvious in case $V_{2\text{inner}}$ performs the Projection test. Otherwise, $V_{2\text{inner}}$ performs a Linearity Test with parameters f_1 and f_2 and an enhanced RMB Tests with the same parameters. Clearly, the answers on f_1 and f_2 determine the acceptable (by Linearity Test) answer on $f_1 + f_2$. The key observation is that the former two answers also determine all 2^l acceptable answers in the enhanced RMB test (i.e., for every $f \in \mathcal{F}_l$, the answer on $f_1' \cdot f + f_2$ should equal the answer on f_2, where $f_1' = f_1$ if the answer on f_1 is zero and $f_1' = f_1 + \bar{1}$ otherwise). ∎

By repeating the above proof system three times, we obtain

Corollary 5.5 $\text{NP} \subseteq \text{FPCP}_{1,1/2}[\log, 6]$.
Furthermore, there is a constant q such that $\text{NP} \subseteq \text{PCP}_{1,1/2}[\,\mathsf{coins} = \log\,;\, \mathsf{free} = 6\,;\, \mathsf{query} = q\,]$.

Proof: There exists $\epsilon > 0$ such that $\left(\frac{173}{218} + \epsilon\right)^3 \leq \frac{1}{2}$. ∎

5.2 Hardness of vertex cover

Refer to Section 2.4 for the definition of the MinVC problem and the associated gap problem $\mathsf{Gap}\text{-MinVC}_{c,s}$, and to Section 2.4.3 for status and previous work.

Going from Free bits to VC. Instead of reducing from Max3SAT, we first use Theorem 5.4 to get gaps in Clique size, and then apply the standard reduction.

Proposition 5.6 $\mathrm{FPCP}_{c,s}[\log, f] \leq_D^K \textsf{Gap-}\mathrm{MinVC}_{c',s'}$ for $s' = 1 - 2^{-f}c$ and $\frac{c'}{s'} = 1 + \frac{c-s}{2^f - c}$.

Proof: The FGLSS reduction says that $\mathrm{FPCP}_{c,s}[\log, f] \leq_D^K \textsf{Gap-}\mathrm{MaxClique}_{c'',s''}$ where $c'' = 2^{-f} \cdot c$ and $s'' = 2^{-f} \cdot s$. (See Section 2.4 for definition of $\textsf{Gap-}\mathrm{MaxClique}$.) Now we apply the standard Karp reduction (of MaxClique to MinVC) which maps a graph G to its complement $\overline{G}$, noting that $\overline{\mathrm{MinVC}}(\overline{G}) = 1 - \overline{\mathrm{MaxClique}}(G)$. Thus $\textsf{Gap-}\mathrm{MaxClique}_{c'',s''} \leq_D^K \textsf{Gap-}\mathrm{MinVC}_{1-s'',1-c''}$. Now set $c' = 1 - s''$ and $s' = 1 - c''$ and note

$$\frac{c'}{s'} = \frac{1-s''}{1-c''} = \frac{1-s2^{-f}}{1-c2^{-f}} = 1 + \frac{c-s}{2^f - c}.$$

This completes the proof. ▌

Our results. We obtain the first explicit and reasonable constant factor non-approximability result for MinVC. A consequence of the following theorem is that, assuming P $\neq$ NP there is no polynomial time algorithm to approximate MinVC within a factor of 1.0688.

Theorem 5.7 $\textsf{Gap-}\mathrm{MinVC}_{c,s}$ is NP-complete for some c, s satisfying $c/s \geq 1.0688 > 16/15$. Moreover $s = 3/4$.

Proof: Follows immediately from Proposition 5.6 and Theorem 5.4. Namely, for any $s' > 173/218$, $\mathrm{NP} \subseteq \mathrm{FPCP}_{1,s'}[\log, 2] \leq_D^K \textsf{Gap-}\mathrm{MinVC}_{c,s}$ for $s = 1 - 2^{-2} = \frac{3}{4}$ and $\frac{c}{s} = 1 + \frac{1-s'}{2^2-1} = 1 + \frac{1-s'}{3}$. Thus, $\frac{c}{s} = 1 + \frac{15}{218} - \frac{\epsilon}{3} > 1.068807 - \frac{\epsilon}{3}$, where $\epsilon \overset{\mathrm{def}}{=} s' - \frac{173}{218}$. ▌

We remark that a special case of Proposition 5.6 in which the statement is restricted to $f = 0$ would have sufficed for proving the above theorem. The reason being that we could have applied Proposition 11.8 to Theorem 5.4 and obtained $\mathrm{NP} \subseteq \mathrm{FPCP}_{1/4,s/4}[\log, 0]$, for $s = 0.7936$, which by the special case of Proposition 5.6 is reducible to $\textsf{Gap-}\mathrm{MinVC}_{c',s'}$ with $s' = 1 - \frac{1}{4} = \frac{3}{4}$ and $\frac{c'}{s'} = 1 + \frac{(1/4)-(s/4)}{1-(1/4)} = 1 + \frac{1-s}{3}$ (as above). Interestingly, the special case of Proposition 5.6 can be "reversed": namely, $\textsf{Gap-}\mathrm{MinVC}_{c',s'}$ is reducible to $\mathrm{FPCP}_{c,s}[\log, 0]$ with $s = 1 - c'$, $c = 1 - s'$ and $\frac{c}{s} = \frac{1-s'}{1-c'}$ (which reverses $\frac{c'}{s'} = \frac{1-s}{1-c} = 1 + \frac{c-s}{1-c}$). The key fact in proving this "reverse reduction" is Corollary 8.5 which asserts that $\textsf{Gap-}\mathrm{MaxClique}_{c,s} \leq_D^K \mathrm{FPCP}_{c,s}[\log, 0]$. However, we do not know if it is possible to "reverse" the other step in the alternative proof; namely, whether $\mathrm{FPCP}_{c,s}[\log, 0]$ is reducible to $\mathrm{FPCP}_{4c,4s}[\log, 2]$ (our reverse transformation is weaker – see Proposition 11.6).

6 Minimizing the number of queries for soundness 0.5

The problem we consider here is to minimize the values of q (and q_{av}) for which we can construct PCPs for NP using q queries in the worst case (and q_{av} on the average) to achieve a soundness error of 1/2. We allow only logarithmic randomness. See Section 2.2.3 for description of past records.

Sources of our improvements. The principal part of our improvement comes from the use of the new long code based inner verifier, the atomic tests and their analysis in Section 3.5, and the new idea of folding. By repeating the proof system of Theorem 4.5 five times, we obtain that Eq. (4) holds for $q = 15$. (Note that $5 = \min\{i \in \mathsf{N} : 0.85^i < 0.5\}$.) A straightforward implementation of the recycling technique of [BGLR] yields that Eq. (4) holds for $q = 12$ and $q_{\mathrm{av}} = 11.74$. Using, a more careful implementation of this technique, we reduce the query complexity by an additional bit.

6.1 The PCP inner verifier

Our result is based on the construction of the (l, l_1)-canonical inner verifier V_{PCPinner} depicted in Figure 14. In addition to its standard inputs h, σ it takes parameters $p_1, p_2, p_3 \geq 0$ so that $p_1+p_2+p_3 = 1$. The inner verifier V_{PCPinner} combines the atomic tests in three different ways.

(1) Some tests are performed independently (i.e., the main steps in Figure 14);

(2) Some tests are performed while re-using some queries (i.e., the tests in Step (2) re-use f_3);

(3) Some tests are performed in a mutual exclusive manner (i.e., the tests in Step (3));

As in previous sections, the tests are executed on the function $A_{(h,0),(\bar{1},1)}$ to which the verifier has an effective oracle access given his access to A. By inspection it is clear that the total number of accesses to the oracles for A and A_1 is $3+5+3 = 11$ (whereas the free-bit complexity is $2+3+2 = 7$). We now examine the goodness of V_{PCPinner}. Recall the definitions of the functions $\Gamma_{\text{lin}}(x)$ (from Lemma 3.15) and $\Gamma_{\text{RMB}}(x) = \frac{3}{8}(1 - 2x)$ (from Lemma 3.19).

Lemma 6.1 (soundness of V_{PCPinner}): For any $0 < \delta_1, \delta_2 < 0.1$ and any l, l_1, p_1, p_2 and p_3, satisfying $p_1 + p_2 + p_3 = 1$ and $5p_1 = 2p_2$, the (l, l_1)-canonical inner verifier V_{PCPinner} is $(\rho, \delta_1, \delta_2)$-good, where $1 - \rho$ is the minimum of the following three quantities

(1) $\frac{1}{2} + \frac{p_1}{10} - \delta_1$;

The PCP inner verifier. This (l, l_1)-canonical inner verifier is given functions $h \in \mathcal{F}_l$ and σ: $\Sigma^l \to \Sigma^{l_1}$, and has access to oracles for A: $\mathcal{F}_l \to \Sigma$ and A_1: $\mathcal{F}_{l_1} \to \Sigma$. In addition it takes three non-negative parameters p_1, p_2 and p_3 which sum-up to 1.

```
Pick functions $f_1, \ldots, f_8 \overset{R}{\leftarrow} \mathcal{F}_l$ and $g_1, g_2 \overset{R}{\leftarrow} \mathcal{F}_{l_1}$.

Step 1: Linearity Test
    $\mathbf{LinTest}(A_{(h,0),(\bar{1},1)}; f_1, f_2)$.

Step 2: Combined RMB and Projection Test
    $\mathbf{MBTest}(A_{(h,0),(\bar{1},1)}; f_3, f_4, f_5)$.
    $\mathbf{ProjTest}_\sigma(A_{(h,0),(\bar{1},1)}, A_1; f_3, g_1)$.

Step 3: Invoking $V_{\text{SNPinner}}$ with parameters $p_1, p_2, p_3$.
    Pick $p \overset{R}{\leftarrow} [0, 1]$.
    Case $p \leq p_1$ :                 $\mathbf{LinTest}(A_{(h,0),(\bar{1},1)}; f_6, f_7)$.
    Case $p_1 < p \leq p_1 + p_2$ :     $\mathbf{MBTest}(A_{(h,0),(\bar{1},1)}; f_6, f_7, f_8)$.
    Case $p_1 + p_2 < p$ :              $\mathbf{ProjTest}_\sigma(A_{(h,0),(\bar{1},1)}, A_1; f_6, g_2)$.

Accept iff all the above tests accept.
```

Remark: access to $A_{(h,0),(\bar{1},1)}(f)$ is implemented by accessing either $A(f)$, $A(f+h)$, $A(f+\bar{1})$ or $A(f + h + \bar{1})$.

Figure 14: *The PCP inner verifier* V_{PCPinner}

(2) $1-(11/14)^3 - \frac{p_3}{1-p_3} > 0.51494168 - \frac{p_3}{1-p_3}$;
(3) $\min\{\frac{1}{2}+\frac{p_3}{20}-\delta_2\,,\, 1-(0.55218507+\delta_2)\cdot(1-\frac{45}{128}p_1)\}$
Furthermore, if $p_1 > 10\delta_1$, $p_3 > 20\delta_2$ and $p_3 \le 0.01$ then $1-\rho > \frac{1}{2}$.

Proof: We split the analysis into several cases based on the value of $x = \mathrm{Dist}(A_{(h,0),(\bar{1},1)}, \textsc{Lin})$.

Case 1: $x \ge \frac{1}{2}-\delta_1$. Lemma 3.15 implies that $\textsc{LinPass}(A_{(h,0),(\bar{1},1)}) \le 1-\Gamma_{\mathrm{lin}}(x) \le 1-x \le \frac{1}{2}+\delta_1$. Thus, in this case

$$\mathtt{ACC}\,[\,V^{A,A_1}_{\mathrm{PCPinner}}(\sigma,h)\,] \;\le\; \rho_1 \stackrel{\mathrm{def}}{=} (1-p_1)\cdot\left(\frac{1}{2}+\delta_1\right)+p_1\cdot\left(\frac{1}{2}+\delta_1\right)^2 \;<\; \frac{1}{2}+\delta_1-\frac{p_1}{10}$$

(The last inequality is due to $\delta_1 < 0.1$.) Using $p_1 > 10\delta_1$ we get $\rho_1 < 1/2$.

Case 2: $x < \frac{1}{2}-\delta_1$. Let $\tilde{A}\colon \mathcal{F}_l \to \Sigma$ be a linear function such that $\mathrm{Dist}(A_{(h,0),(\bar{1},1)}, \tilde{A}) = x$. The proof splits into two subcases.

Case 2.1: $\tilde{A}$ does not respect the monomial basis. In this case, by Lemmas 3.15 and 3.19 we have

$$\begin{aligned}
\mathtt{ACC}\,[\,V^{A,A_1}_{\mathrm{PCPinner}}(\sigma,h)\,] \;&\le\; (1-\Gamma_{\mathrm{lin}}(x))\cdot(1-\Gamma_{\mathrm{RMB}}(x))\cdot(1-p_1\Gamma_{\mathrm{lin}}(x)-p_2\Gamma_{\mathrm{RMB}}(x)) \\
&<\; (1-\Gamma_{\mathrm{lin}}(x))\cdot(1-\Gamma_{\mathrm{RMB}}(x)) \\
&\quad\cdot\left(1-\frac{p_1}{p_1+p_2}\cdot\Gamma_{\mathrm{lin}}(x)-\frac{p_2}{p_1+p_2}\cdot\Gamma_{\mathrm{RMB}}(x)+\frac{p_3}{1-p_3}\right) \\
&<\; \alpha\cdot\beta\cdot[q\alpha+(1-q)\beta]+\frac{p_3}{1-p_3}
\end{aligned}$$

where $q \stackrel{\mathrm{def}}{=} \frac{p_1}{p_1+p_2}$, $\alpha = 1-\Gamma_{\mathrm{lin}}(x)$ and $\beta = 1-\Gamma_{\mathrm{RMB}}(x)$. Using $p\cdot x+(1-p)\cdot y \ge x^p\cdot y^{1-p}$, we show that $\alpha\cdot\beta\cdot[q\alpha+(1-q)\beta] \le [\frac{1+q}{3}\alpha+\frac{2-q}{3}\beta]^3$. Specifically,

$$\begin{aligned}
\left[\frac{1+q}{3}\alpha+\frac{2-q}{3}\beta\right]^3 \;&=\; \left[\frac{2}{3}\cdot(\frac{1}{2}\alpha+\frac{1}{2}\beta)+\frac{1}{3}\cdot(q\alpha+(1-q)\beta)\right]^3 \\
&\ge\; \left(\frac{1}{2}\alpha+\frac{1}{2}\beta\right)^{\frac{2}{3}\cdot 3}\cdot(q\alpha+(1-q)\beta)^{\frac{1}{3}\cdot 3} \\
&=\; \left[\frac{1}{2}\cdot\alpha+\frac{1}{2}\cdot\beta\right]^2\cdot(q\alpha+(1-q)\beta) \\
&\ge\; \alpha\cdot\beta\cdot(q\alpha+(1-q)\beta)
\end{aligned}$$

Combining the above with Claim 4.2 (i.e., the lower bound on T_2), we obtain (for every $x < 1/2$)

$$\begin{aligned}
\mathtt{ACC}\,[\,V^{A,A_1}_{\mathrm{PCPinner}}(\sigma,h)\,] \;&<\; \left[1-\frac{1+q}{3}\cdot\Gamma_{\mathrm{lin}}(x)-\frac{2-q}{3}\cdot\Gamma_{\mathrm{RMB}}(x)\right]^3+\frac{p_3}{1-p_3} \\
&\le\; \left[1-\min\left(\frac{1+q}{6},\frac{2-q}{8}\right)\right]^3+\frac{p_3}{1-p_3}
\end{aligned}$$

Observe that $\min(\frac{1+q}{6},\frac{2-q}{8})$ is maximized at $q=2/7$ where its value is $3/14$. Indeed this value of q is consistent with $p_1 = \frac{2}{7}\cdot(p_1+p_2)$ and so, in this case, we get

$$\mathtt{ACC}\,[\,V^{A,A_1}_{\mathrm{PCPinner}}(\sigma,h)\,] \;\le\; \rho_2 \stackrel{\mathrm{def}}{=} \left[\frac{11}{14}\right]^3+\frac{p_3}{1-p_3} \;<\; 0.48505832+\frac{p_3}{1-p_3}$$

Using $p_3 \le 0.01$ we get $\rho_2 < 1/2$.

Case 2.2: $\tilde{A}$ respects the monomial basis. By Proposition 3.2, $\tilde{A}$ is an evaluation operator. So there exists $a \in \Sigma^l$ such that $\tilde{A} = E_a$. So $\mathrm{Dist}(A_{(h,0),(\bar{1},1)}, E_a) = x$. Let $a_1 = \sigma(a)$. The proof splits into two further sub-cases.

Case 2.2.1: $d \overset{\text{def}}{=} \mathrm{Dist}(A_1, E_{a_1}) \ge 1/2 - \delta_2$. By Lemma 3.21 we have $\textsc{ProjPass}_\sigma(A_{(h,0),(\bar{1},1)}, A_1) \le 1 - d \cdot (1-2x) < \frac{1}{2} + x + \delta_2$. Letting $\Gamma_{\mathrm{PRJ}}(x) \overset{\text{def}}{=} \frac{1}{2} - x - \delta_2$, we get in this case

$$\mathtt{ACC}\,[\,V^{A,A_1}_{\mathrm{PCPinner}}(\sigma,h)\,] \;\le\; \rho_3 \overset{\text{def}}{=} (1-\Gamma_{\mathrm{lin}}(x)) \cdot (1-\Gamma_{\mathrm{PRJ}}(x)) \cdot (1 - p_1\Gamma_{\mathrm{lin}}(x) - p_3\Gamma_{\mathrm{PRJ}}(x))$$

We upper bound ρ_3 by considering three sub-cases (corresponding to the segments of Γ_{lin}).

Case 2.2.1.1: $x \le 1/4$. In this case we use $\Gamma_{\mathrm{lin}}(x) \ge 3x(1-2x)$ and obtain

$$\begin{aligned}
\rho_3 &< (1-\Gamma_{\mathrm{lin}}(x)) \cdot (1-\Gamma_{\mathrm{PRJ}}(x)) \cdot (1 - p_3\Gamma_{\mathrm{PRJ}}(x)) \\
&< (1-3x(1-2x)) \cdot (\frac{1}{2} + x + \delta_2) \cdot (1 - \frac{p_3}{10}) \\
&< \frac{1}{2} \cdot \left[1 - x + 12x^3\right] \cdot \left[1 - \frac{p_3}{10}\right] + \delta_2 \\
&\le \frac{1}{2} \cdot \left[1 - \frac{p_3}{10}\right] + \delta_2
\end{aligned}$$

where the last inequality uses the fact that the function $x - 12x^3$ is non-negative in the interval $[0, 1/4]$. Using $p_3 > 20\delta_2$ we obtain $\rho_3 < 1/2$.

Case 2.2.1.2: $x \ge 1/4$ and $x \le 45/125$. In this case we use $\Gamma_{\mathrm{lin}}(x) \ge 45/128 = \Gamma_{\mathrm{lin}}(45/128)$ and $\Gamma_{\mathrm{PRJ}}(x) \ge \Gamma_{\mathrm{PRJ}}(45/128)$ and obtain

$$\begin{aligned}
\rho_3 &< (1-\Gamma_{\mathrm{lin}}(x)) \cdot (1-\Gamma_{\mathrm{PRJ}}(x)) \cdot (1 - p_1\Gamma_{\mathrm{lin}}(x)) \\
&\le (1-\Gamma_{\mathrm{lin}}(45/128)) \cdot (1-\Gamma_{\mathrm{PRJ}}(45/128)) \cdot (1 - p_1\Gamma_{\mathrm{lin}}(45/128)) \\
&< \frac{83}{128} \cdot \left(\frac{109}{128} + \delta_2\right) \cdot \left(1 - p_1\frac{45}{128}\right) \\
&< (0.55218507 + \delta_2) \cdot \left(1 - p_1\frac{45}{128}\right)
\end{aligned}$$

Using $\delta_2 < \frac{p_3}{10} < 0.001$ and $p_1 \ge \frac{2}{7} \cdot 0.99 > 0.28$, we obtain $\rho_3 < 0.5532 \cdot 0.902 < 0.499$.

Case 2.2.1.3: $x \ge 45/128$. In this case we use $\Gamma_{\mathrm{lin}}(x) \ge x \ge 45/128$ and obtain

$$\begin{aligned}
\rho_3 &< (1-\Gamma_{\mathrm{lin}}(x)) \cdot (1-\Gamma_{\mathrm{PRJ}}(x)) \cdot (1 - p_1\Gamma_{\mathrm{lin}}(x)) \\
&< (1-x) \cdot (\frac{1}{2} + x + \delta_2) \cdot (1 - p_1\frac{45}{128})
\end{aligned}$$

The latter expression decreases in the interval $[\frac{45}{128}, \frac{1}{2}]$ and is hence maximized at $x = 45/128$. Thus we obtain the same expression as in Case 2.2.1.2, and the bound on ρ_3 follows identically.

We conclude that in Case (2.2.1) we have

$$\rho_3 \;<\; \max\left[\frac{1}{2} - \frac{p_3}{20} + \delta_2 \;,\; (0.55218507 + \delta_2) \cdot (1 - p_1\frac{45}{128})\right]$$

and under the hypothesis regarding p_1, p_3 and δ_2, we always have $\rho_3 < 0.5$.

Case 2.2.2: Else, we have $x = \mathrm{Dist}(A_{(h,0),(\bar{1},1)}, E_a) \leq 1/2 - \delta_1$ and $\mathrm{Dist}(A_1, E_{a_1}) < 1/2 - \delta_2$. Thus the functions $A_{(h,0),(\bar{1},1)}$ and A_1 satisfy the properties required in conditions (2.1) and (2.2) of Definition 3.9.

Let $\rho \stackrel{\mathrm{def}}{=} \max\{\rho_1, \rho_2, \rho_3\}$. We conclude that the only case which allows $\mathtt{ACC}\,[\,V_{\mathrm{PCPinner}}^{A,A_1}(\sigma, h)\,] > \rho$ is Case (2.2.2) which also satisfies conditions (2.1) and (2.2) of Definition 3.9. Thus, V_{PCPinner} satisfies condition (2) of Definition 3.9. Clearly, V_{PCPinner} also satisfies condition (1) of Definition 3.9, and thus the lemma follows. ▌

6.2 The new proof system

Combining the above inner verifier with an adequate outer verifier, we obtain a pcp system for NP with query complexity 11.

Theorem 6.2 $\mathrm{NP} = \mathrm{PCP}_{1,1/2}[\,\mathsf{coins} = \log\,;\ \mathsf{query} = 11\,;\ \mathsf{query}_{\mathrm{av}} = 10.89\,;\ \mathsf{free} = 7\,]$.

Proof: We consider a canonical (l, l_1)-inner verifier V_{PCPinner} with parameters $p_3 = 0.001$, $p_1 = \frac{2}{7} \cdot 0.999$ and $p_2 = \frac{5}{7} \cdot 0.999$. By Lemma 6.1, V_{PCPinner} is $(\rho, \delta_1, \delta_2)$-good for $\delta_1 = \delta_2 = 0.00001$ and $\rho = 0.49999$. We now choose an appropriate outer verifier. Let $\epsilon = 16 \cdot (0.5 - \rho)\delta_1^2\delta_2^2$. Lemma 3.8 provides us with l and l_1 such that an ϵ-good (l, l_1)-canonical outer verifier V_{outer} with randomness $O(\log n)$ exists. Let $V = \langle V_{\mathrm{outer}}, V_{\mathrm{PCPinner}} \rangle$ be the composition of V_{outer} and V_{PCPinner} according to the definitions in Section 3.4. This verifier has randomness $O(\log n)$. Apply Theorem 3.12 to see that V has completeness parameter 1 and soundness parameter $\rho + \epsilon/(16\delta_1^2\delta_2^2) = 1/2$. The query (and free-bit) complexity of V is the same as that of V_{PCPinner} above (i.e., 11 and 7, respectively).

To obtain the bound on the average query complexity, we observe that we can afford not to perform the RMB test with some small probability. Specifically, Case (2.1) in the proof of Lemma 6.1, which is the only case where the RMB test is used, yields error of $0.48505832 + \frac{p_3}{1-p_3}$. Thus, if we modify V_{PCPinner} so that, whenever the RMB test is invoked it is performed only with probability 0.973, we get that Case (2.1) detects violation with probability at least $(1 - 0.48505832 - 0.0010011) \cdot 0.973 > 0.50006$. Consequently, the modified inner verifier errs with probability bounded away from 1/2 and so does the composed verifier. The modification decreases the average query complexity by $(1-0.973)\cdot(2+p_2\cdot 3) > 0.027\cdot 4.12 > 0.11$. (The reduction is both from Step (2) and the second case in Step (3).) The theorem follows. ▌

7 Amortized free-bits and Max Clique hardness

7.1 The iterated tests

The "iterated tests" will be used in the next section to derive a proof system for NP having amortized free-bit complexity ≈ 2. Intuitively, we will be running each of the atomic tests many times, but, to keep the free-bit count low, these will NOT be independent repetitions. Rather, following [BeSu], we will run about $2^{O(m)}$ copies of each test in a way which is pairwise, or "almost" pairwise independent, to lower the error probability to $O(2^{-m})$. This will be done using $2m$ free-bits. Specifically, we will select uniformly m functions in $\mathcal{F}_l$ (and m functions in $\mathcal{F}_{l_1}$) and invoke the atomic tests with functions resulting from all possible linear combinations of the selected functions.

7.1.1 Linearity and randomness

We begin with some observations relating stochastic and linear independence. Note that $\mathcal{L}_m$ is a sub-vector-space of $\mathcal{F}_m$, and in particular a vector space over Σ in its own right. So we can discuss the linear independence of functions in $\mathcal{L}_m$. We say that $\vec{L} = (L_1, \ldots, L_k) \in \mathcal{L}_m^k$ is linearly independent if $L_1, \ldots, L_k$ are linearly independent. Furthermore we say that $\vec{L}_1 = (L_{1,1}, \ldots, L_{1,k})$ and $\vec{L}_2 = (L_{2,1}, \ldots, L_{2,k})$ are *mutually linearly independent* if the $2k$ functions $L_{1,1}, L_{2,1}, \ldots, L_{1,k}, L_{2,k}$ are linearly independent.

Lemma 7.1 For $\vec{L} = (L_1, \ldots, L_k) \in \mathcal{L}_m^k$ let $J_{\vec{L}}\colon \mathcal{F}_l^m \to \mathcal{F}_l^k$ be defined by $J_{\vec{L}}(\vec{f}) = (L_1 \circ \vec{f}, \ldots, L_k \circ \vec{f})$, for $\vec{f} = (f_1, ..., f_m)$. Fix $\vec{L}$ and consider the probability space defined by having $f_1, ..., f_m$ be uniformly and independently distributed over $\mathcal{F}_l$. Regard the $J_{\vec{L}}$'s as random variables over the above probability space. Then

(1) If $\vec{L}$ is linearly independent then $J_{\vec{L}}$ is uniformly distributed in $\mathcal{F}_l^k$.

(2) If $\vec{L}_1, \vec{L}_2$ are mutually linearly independent then $J_{\vec{L}_1}$ and $J_{\vec{L}_2}$ are independently distributed.

The proof of this lemma is quite standard and thus omitted: It amounts to saying that linearly independent combinations of stochastically independent random variables result in stochastically independent random variables.

The analysis of the Iterated Projection test (see Figure 15) can be done in a relatively straightforward way, given the above, because the invoked projection test uses a single linear combination of each sequence of random functions, rather than several such combinations (as in the other iterated tests). Thus we begin with the iterated projection tests. The analysis of the other iterated tests, where the atomic tests are invoked on two/three linear combinations of the same sequence of random function, require slightly more care. The corresponding lemmas could have been proven using the notion of "weak pairwise independence" introduced in [BeSu]. However, we present here an alternative approach.

7.1.2 Iterated projection test

The *iterated projection test* described in Figure 15 takes as input vectors $\vec{f}, \vec{g} \in \mathcal{F}_l^m$ and also a linear function $L \in \mathcal{L}_m$. Note that $f = L \circ \vec{f}$ is in $\mathcal{F}_l$, and $g = L \circ \vec{g}$ is in $\mathcal{F}_{l_1}$. The test is just the atomic projection test on f and g. The following lemma says that if the passing probability $\textsc{ProjPass}_A^m()$, representing 2^m invocations of the atomic projection test, is even slightly significant and if A is close to E_a, then A_1 is close to the encoding of the projection of a.

Lemma 7.2 There is a constant c_3 such that the following is true. Let $\sigma\colon \Sigma^l \to \Sigma^{l_1}$ be a function. Let $a \in \Sigma^l$ be such that $\text{Dist}(E_a, A) \leq 1/4$, and let $a_1 = \sigma(a) \in \Sigma^{l_1}$. If $\textsc{ProjPass}_\sigma^m(A, A_1) \geq c_3 \cdot 2^{-m}$ then $\text{Dist}(E_{a_1}, A_1) \leq 0.1$.

Proof: The proof is similar to that of [BeSu, Lemma 3.5]. Let $\epsilon_1 = \text{Dist}(A_1, E_{a_1})$ and assume it is at least 0.1. We show that there is a constant c_3 such that $\textsc{ProjPass}_h^m(A) < c_3 \cdot 2^{-m}$.

Let $N = |\mathcal{L}_m^*| = 2^m - 1$. For $L \in \mathcal{L}_m^*$ let $X_L\colon \mathcal{F}_l^m \times \mathcal{F}_{l_1}^m \to \Sigma$ be defined by

$$X_L(\vec{f}, \vec{g}) \stackrel{\text{def}}{=} \mathbf{ProjTest}_\sigma^m(A, A_1; \vec{f}, \vec{g}, L) \;=\; \mathbf{ProjTest}_\sigma(A, A_1; L \circ \vec{f},\, L \circ \vec{g})\,.$$

Regard it as a random variable over the uniform distribution on $\mathcal{F}_l^m \times \mathcal{F}_{l_1}^m$. Let $X = \sum_{L \in \mathcal{L}_m^*} X_L$. It suffices to show that $\Pr[\, X = 0\,] \leq O(1/N)$.

Lemma 7.1 implies that $\{X_L\}_{L \in \mathcal{L}_m^*}$ are pairwise independent, identically distributed random variables. Let $L \in \mathcal{L}_m^*$ and let $p = \mathbf{E}[X_L]$. Again using Lemma 7.1 we have

$$\begin{aligned} p &= \Pr_{\vec{f} \xleftarrow{R} \mathcal{F}_l^m \,;\, \vec{g} \xleftarrow{R} \mathcal{F}_{l_1}^m} \left[\mathbf{ProjTest}_\sigma(A, A_1; L \circ \vec{f},\, L \circ \vec{g}) = 1 \right] \\ &= \Pr_{f \xleftarrow{R} \mathcal{F}_l \,;\, g \xleftarrow{R} \mathcal{F}_{l_1}} \left[\mathbf{ProjTest}_\sigma(A, A_1; f, g) = 1 \right] . \end{aligned}$$

But by Lemma 3.21, p is at least $\epsilon_1(1 - 2\epsilon) \geq 0.05$, since $\epsilon \stackrel{\text{def}}{=} \text{Dist}(E_a, A) \leq 1/4$. We can conclude by applying Chebyshev's inequality. Namely,

$$\Pr[\, X = 0 \,] \;\leq\; \Pr[\, |X - Np| \geq Np \,] \;\leq\; \frac{Np}{(Np)^2} \;\leq\; \frac{20}{N}$$

as desired. ▌

7.1.3 Technical claim

For analyzing the other two tests we will use the following simple claim.

Claim 7.3 Let $k \geq 1$ and $N = 2^m$. Then $\mathcal{L}_m^k$ contains a subset S of cardinality $\frac{N}{2^{2k}}$ such that every $\vec{L}_1 \neq \vec{L}_2 \in S$ are mutually linearly independent.

Proof: Let $\vec{L} \in \mathcal{L}_m^k$ be linearly independent. Then, the probability that L chosen uniformly in $\mathcal{L}_m$ is linearly independent of $\vec{L}$ is $1 - \frac{2^k}{N}$. Thus, the probability that a uniformly chosen $\vec{L}' \in \mathcal{L}_m^k$ is mutually linearly independent of $\vec{L}$ is greater than $1 - \sum_{i=1}^{k} \frac{2^{k+i-1}}{N} > 1 - \frac{2^{2k}}{N}$. Now, consider a graph with vertex set $\mathcal{L}_m^k$ and edges connecting pairs of mutually linearly independent sequences (i.e., $\vec{L}_1$ and $\vec{L}_2$ are connected if and only they are mutually linearly independent). This graph has N^k vertices and every vertex which is linearly independent has degree greater than $(1 - \frac{2^{2k}}{N}) \cdot N^k$. Clearly this graph has a clique of size $\frac{N}{2^{2k}}$ (e.g., consider a greedy algorithm which picks a vertex of maximal degree among all vertices connected to the previously selected vertices). Noting that a clique corresponds to a set of mutually linear independent sequences, we are done. ▌

7.1.4 Iterated linearity test

The *iterated linearity test* described in Figure 15 takes as input a vector $\vec{f} \in \mathcal{F}_l^m$ and also linear functions $L_1, L_2 \in \mathcal{L}_m$. Note that $f_1 = L_1 \circ \vec{f}$ and $f_2 = L_2 \circ \vec{f}$ are in $\mathcal{F}_l$. The test is just the atomic linearity test on these inputs. The following lemma says that if the passing probability is even slightly significant, then A is almost linear.

Lemma 7.4 There is a constant c_1 such that if $\text{LinPass}^m(A) \geq c_1 \cdot 2^{-m}$ then $\text{Dist}(A, \text{Lin}) \leq 0.1$.

Proof: Assume that $\epsilon \stackrel{\text{def}}{=} \text{Dist}(A, \text{Lin}) \geq 0.1$. We show that there is a constant c_1 such that $\text{LinPass}^m(A) < c_1 \cdot 2^{-m}$. For $\vec{L} = (L_1, L_2) \in \mathcal{L}_m^2$ let $X_{\vec{L}} \colon \mathcal{F}_l^m \to \Sigma$ be defined by

$$X_{\vec{L}}(\vec{f}) \stackrel{\text{def}}{=} \mathbf{LinTest}^m(A; \vec{f}, L_1, L_2) \;=\; \mathbf{LinTest}(A; L_1 \circ \vec{f},\, L_2 \circ \vec{f}) \,.$$

Regard it as a random variable over the uniform distribution on $\mathcal{F}_l^m$. Let $S \subset \mathcal{L}_m^2$ be a set as guaranteed by Claim 7.3 and $X = \sum_{\vec{L} \in S} X_{\vec{L}}$. It suffices to show that $\Pr[\, X = 0 \,] \leq O(2^{-m})$. (Thus

The Iterated Tests. Here $A\colon \mathcal{F}_l \to \Sigma$ and $A_1\colon \mathcal{F}_{l_1} \to \Sigma$ are the objects being tested. The tests also take additional inputs or parameters: below $\vec{f} \in \mathcal{F}_l^m$; $\vec{g} \in \mathcal{F}_{l_1}^m$; $L, L_1, L_2, L_3 \in \mathcal{L}_m$; and $\sigma\colon \Sigma^l \to \Sigma^{l_1}$. The tests are specified in terms of the atomic tests of Figure 8.

$$\mathbf{LinTest}^m(A; \vec{f}, L_1, L_2) = \mathbf{LinTest}(A; L_1 \circ \vec{f},\ L_2 \circ \vec{f}).$$

$$\mathbf{MBTest}^m(A; \vec{f}, L_1, L_2, L_3) = \mathbf{MBTest}(A; L_1 \circ \vec{f},\ L_2 \circ \vec{f},\ L_3 \circ \vec{f}).$$

$$\mathbf{ProjTest}_\sigma^m(A, A_1; \vec{f}, \vec{g}, L) = \mathbf{ProjTest}_\sigma(A, A_1; L \circ \vec{f},\ L \circ \vec{g}).$$

The Passing Probabilities. These are the probabilities we are interested in:

$$\begin{aligned}
\textsc{LinPass}^m(A) &= \Pr_{\vec{f} \overset{R}{\leftarrow} \mathcal{F}_l^m}\left[\forall\, L_1, L_2 \in \mathcal{L}_m \ :\ \mathbf{LinTest}^m(A; \vec{f}, L_1, L_2) = 0\right] \\
\textsc{MBPass}^m(A) &= \Pr_{\vec{f} \overset{R}{\leftarrow} \mathcal{F}_l^m}\left[\forall\, L_1, L_2, L_3 \in \mathcal{L}_m \ :\ \mathbf{MBTest}^m(A; \vec{f}, L_1, L_2, L_3) = 0\right] \\
\textsc{ProjPass}_\sigma^m(A, A_1) &= \Pr_{\vec{f} \overset{R}{\leftarrow} \mathcal{F}_l^m;\, \vec{g} \overset{R}{\leftarrow} \mathcal{F}_{l_1}^m}\left[\forall\, L \in \mathcal{L}_m \ :\ \mathbf{ProjTest}_\sigma^m(A, A_1; \vec{f}, \vec{g}, L) = 0\right]
\end{aligned}$$

Figure 15: *The iterated tests and their passing probabilities.*

our analysis of $\textsc{LinPass}^m(A)$ is based only on a small fraction of all possible invocations of the iterated linear test; yet, this small fraction corresponds to a sufficiently large number of invocations.)

Using Lemma 7.1, it follows that the random variables $\{X_{\vec{L}}\}_{\vec{L} \in S}$ are pairwise independent and that for every $\vec{L} \in S$

$$p \stackrel{\text{def}}{=} \Pr_{\vec{f} \overset{R}{\leftarrow} \mathcal{F}_l^m}\left[X_{\vec{L}}(\vec{f}) = 1\right] \;=\; \Pr_{f_1, f_2 \overset{R}{\leftarrow} \mathcal{F}_l}\left[\mathbf{LinTest}(A; f_1, f_2) = 1\right].$$

By Lemma 3.15, $p \geq \Gamma_{\text{lin}}(\epsilon)$ and so $p \geq 3\epsilon - 6\epsilon^2$ if $\epsilon \leq 1/4$ and $p \geq 45/128$ otherwise. In either case, for $\epsilon \geq 0.1$, we get $p > 0.2$. Now by Chebyshev's inequality we have

$$\Pr\left[\, X = 0 \,\right] \leq \Pr\left[\, |X - N'p| \geq N'p \,\right] \leq \frac{1}{N'p} < \frac{5}{N'}$$

where $N' \stackrel{\text{def}}{=} |S| = 2^m / 2^{2 \cdot 2} = 2^m/16$. The lemma follows. ∎

7.1.5 Iterated RMB test

The *iterated respect of monomial basis test* in Figure 15 takes an input $\vec{f}$ and also three linear functions $L_1, L_2, L_3 \in \mathcal{L}_m$. For simplicity of exposition, we assume that A is folded over $(\bar{1}, 1)$. (This assumption is justified by our usage of the test – see next subsection.) If the probability $\textsc{MBPass}^m(A)$ is significant, we can conclude that the linear function close to A respects the monomial basis.

Lemma 7.5 There is a constant c_2 such that the following is true. Let $A\colon \mathcal{F}_l \to \Sigma$ so that $A(f+\bar{1}) = A(f)+1$, for every $f \in \mathcal{F}_l$. Let $\epsilon \le 0.1$ so that A is ϵ-close to a linear function $\tilde{A}$ and suppose that $\text{MBP}\textsc{ass}^m(A) \ge c_2 \cdot 2^{-m}$. Then $\tilde{A}$ respects the monomial basis.

Proof: Assume that $\tilde{A}$ is linear but does not respect the monomial basis. We will show that there is a constant c_2 such that $\text{MBP}\textsc{ass}^m(A) < c_2 \cdot 2^{-m}$.

For $\vec{L} = (L_1, L_2, L_3) \in \mathcal{L}_m^3$ let $X_{\vec{L}}\colon \mathcal{F}_l^m \to \Sigma$ be defined by

$$X_{\vec{L}}(\vec{f}) \stackrel{\text{def}}{=} \mathbf{MBTest}^m(A; \vec{f}, L_1, L_2, L_3) \;=\; \mathbf{MBTest}(A; L_1 \circ \vec{f},\, L_2 \circ \vec{f},\, L_3 \circ \vec{f})\;.$$

Regard it as a random variable over the uniform distribution on $\mathcal{F}_l^m$. Again, let $S \subset \mathcal{L}_m^3$ be a set as guaranteed by Claim 7.3 (in this case $|S| = 2^m/2^{2\cdot 3}$), and $X = \sum_{\vec{L}\in S} X_{\vec{L}}$. It suffices to show that $\Pr[\, X = 0\,] \le O(2^{-m})$.

Using Lemma 7.1, it follows that the random variables $\{X_{\vec{L}}\}_{\vec{L}\in S}$ are pairwise independent and that for every $\vec{L} \in S$

$$p \stackrel{\text{def}}{=} \Pr_{\vec{f} \stackrel{R}{\leftarrow} \mathcal{F}_l^m}\left[X_{\vec{L}}(\vec{f}) = 1\right] \;=\; \Pr_{f_1,f_2,f_3 \stackrel{R}{\leftarrow} \mathcal{F}_l}\left[\mathbf{MBTest}(A; f_1, f_2, f_3) = 1\right]\;.$$

By Lemma 3.19, $p \ge 3/8 - 7\epsilon/4 + 5\epsilon^2/2 - \epsilon^3$. Using $\epsilon \le 0.1$, it follows that $p > 0.2$. Using Chebyshev's inequality, as in the previous proof, we are done. ▌

Remark 7.6 *For general A's (which are not folded over $(\bar{1}, 1)$), a similar result can be proven by augmenting the iterated RMB test so that on input A, $\vec{f}$ and $\vec{L} = (L_1, L_2, L_3)$ it also checks if $A((L_1 \circ \vec{f}) + \bar{1}) = A(L_1 \circ \vec{f}) + 1$.*

7.1.6 Putting some things together

The last two lemmas above allow us to conclude that if $A_{(h,0),(\bar{1},1)}$ passes the first two tests with any significant probability then $A_{(h,0),(\bar{1},1)}$ is close to some evaluation operator E_a so that $h(a) = 0$. Thus, again, there is no need for a "circuit test".

Corollary 7.7 There is a constant c such that the following is true. Let $A\colon \mathcal{F}_l \to \Sigma$, and suppose $\text{L}\textsc{in}\text{P}\textsc{ass}^m(A_{(h,0),(\bar{1},1)}) \ge c \cdot 2^{-m}$ and $\text{MBP}\textsc{ass}^m(A_{(h,0),(\bar{1},1)}) \ge c \cdot 2^{-m}$. Then there is a string $a \in \Sigma^l$ such that $\text{Dist}(E_a, A_{(h,0),(\bar{1},1)}) \le 0.1$ and $h(a) = 0$.

Proof: Let c be the larger of the constants from Lemmas 7.4 and 7.5. By the first lemma there is a linear $\tilde{A}$ such that $\text{Dist}(A_{(h,0),(\bar{1},1)}, \tilde{A}) < 0.1$. Now the second lemma implies that $\tilde{A}$ respects the monomial basis (using the fact that $A_{(h,0),(\bar{1},1)}(f+\bar{1}) = A_{(h,0),(\bar{1},1)}(f)+1$ for all f's). So Proposition 3.2 says that $\tilde{A}$ is an evaluation function. Finally, by Proposition 3.6, we have $h(a) = 0$. ▌

7.2 NP in amortized free-bit complexity 2

Sources of our improvements We adopt the basic framework of the construction of proof systems with low free-bit complexity as presented in [BeSu]. Our improvement comes from the use of the new long code instead of the Hadamard code as a basis for the construction of inner verifiers. This allows us to save one bit in the amortized free-bit complexity. The reason being that the long code contains explicitly all functions of the encoded string whereas the Hadamard code contains only linear combinations of the bits of the string. Typically, we need to check that the verifier accepts a string and this condition is unlikely to be expressed by a linear combination of the bits of the string. Thus, one needs to keep also the linear combinations of all two-bit products and using these extra combinations (via self-correcting) increases the amortized free-bit by one. Instead, as seen above, the long code allows us to directly handle any function. The fact that we take linear combinations of these functions should not confuse the reader; these are linear combinations of random functions rather than being linear combinations of random LINEAR functions (as in [BeSu]).

Our construction of a proof systems with amortized free-bit complexity of two bits is obtained by composing the (l, l_1)-canonical outer verifier of Lemma 3.8 with a (l, l_1)-canonical inner verifier, denoted $V_{\text{free-in}}$, which is depicted in Figure 16. The inner verifier $V_{\text{free-in}}$ consists of invoking the three iterated tests of Figure 15. In addition, $V_{\text{free-in}}$ also applies the linearity test to the oracle A_1. This is not done in order to improve the rejection probability of $V_{\text{free-in}}$ (in case the oracles A and A_1 are far from being fine), but rather in order to decrease the number of accepting configurations (and consequently the free-bit complexity). We also remark that $V_{\text{free-in}}$ invokes the iterated tests while providing them with access to a double folding of A (i.e., $A_{(h,0),(\bar{1},1)}$) rather than to A itself. This eliminates the need for checking that A encodes a string which evaluates to zero under h and simplifies the iterated RMB test (see remark at the end of subsection 7.1.5). However, unlike in previous subsections, these simplifications do not buy us anything significant (here), since the additional testing could have been done without any additional cost in free-bits.

Lemma 7.8 There exists a constant c such that the following is true. Let l, l_1, m be integers. Then the (l, l_1)-canonical inner verifier $V_{\text{free-in}}$ with parameter m is $(\rho, \delta_1, \delta_2)$-good, where $\rho = c \cdot 2^{-m}$ and $\delta_i = 0.4$, for $i = 1, 2$.

Proof: Here the analysis can be less careful than in analogous statements such as in Lemmas 4.1 and 5.2. Using Corollary 7.7, with respect to the oracle $A_{(h,0),(\bar{1},1)}$, we conclude that if $A_{(h,0),(\bar{1},1)}$ passed both the iterated Linearity and RMB Tests with probability at least $c \cdot 2^{-m}$ then there exists a string $a \in \Sigma^l$ such that $\text{Dist}(E_a, A_{(h,0),(\bar{1},1)}) \le 0.1 = \frac{1}{2} - \delta_1 < 1/4$ and $h(a) = 0$. Using Lemma 7.2, we conclude that if $(A_{(h,0),(\bar{1},1)}, A_1)$ passed the iterated Projection Test, with probability at least $c_3 \cdot 2^{-m}$,

The free inner verifier. Given functions $h \in \mathcal{F}_l$ and $\sigma\colon \Sigma^l \to \Sigma^{l_1}$, the verifier has access to oracles for $A\colon \mathcal{F}_l \to \Sigma$ and $A_1\colon \mathcal{F}_{l_1} \to \Sigma$. It also takes an integer parameter m.

Random choices:	$\vec{f} \overset{R}{\leftarrow} \mathcal{F}_l^m$; $\vec{g} \overset{R}{\leftarrow} \mathcal{F}_{l_1}^m$
$\forall\, L_1, L_2 \in \mathcal{L}_m$:	$\textbf{LinTest}^m(A_{(h,0),(\bar{1},1)}; \vec{f}, L_1, L_2)$
$\forall\, L_1, L_2, L_3 \in \mathcal{L}_m$:	$\textbf{MBTest}^m(A_{(h,0),(\bar{1},1)}; \vec{f}, L_1, L_2, L_3)$
$\forall L \in \mathcal{L}_m$:	$\textbf{ProjTest}_\sigma^m(A_{(h,0),(\bar{1},1)}, A_1; \vec{f}, \vec{g}, L)$
$\forall\, L_1, L_2 \in \mathcal{L}_m$:	$\textbf{LinTest}^m(A_1; \vec{g}, L_1, L_2)$

Remark: access to $A_{(h,0),(\bar{1},1)}(f)$ is implemented by accessing either $A(f)$, $A(f+h)$, $A(f+\bar{1})$ or $A(f + h + \bar{1})$.

Figure 16: *The free inner verifier* $V_{\text{free-in}}$

then $\mathrm{Dist}(E_{\sigma(a)}, A_1) < 0.1 = \frac{1}{2} - \delta_2$. Setting $\rho = c' \cdot 2^{-m}$, where $c' = \max\{c, c_3\}$, we conclude that $V_{\text{free-in}}$ satisfies condition (2) of Definition 3.9. Clearly, $V_{\text{free-in}}$ also satisfies condition (1) and the lemma follows. ▌

Proposition 7.9 Let l, l_1, m be integers. Then the (l, l_1)-canonical inner verifier $V_{\text{free-in}}$ with parameter m uses $2m$ free-bits.

Proof: We consider only accepting computations of $V_{\text{free-in}}$. We start by observing that all oracle values obtained from A, during the iterated Linearity Test (on $A_{(h,0),(\bar{1},1)}$), are determined by the values of A in locations $f'_1, f'_2, ..., f'_m$, where each f'_i is one of the four functions $f_i, f_i + h, f_i + \bar{1}$ and $f_i + h + \bar{1}$. Likewise, all oracle values obtained from A, during the iterated RMB Test, are determined by the values of A in these locations $f'_1, f'_2, ..., f'_m$. Finally, all oracle values obtained from A, during the iterated Projection Test, are determined by the values of A_1 in locations $L \circ \vec{g}$ (for all L's) and the values of A in the locations $f'_1, f'_2, ..., f'_m$.

Now we use the fact that $V_{\text{free-in}}$ applies an iterated Linearity Test to the oracle A_1. It follows that all oracle values obtained from A_1, in accepting computations of $V_{\text{free-in}}$, are determined by the values of A_1 in locations $g_1, g_2, ..., g_m$.

We conclude that, in accepting computations of $V_{\text{free-in}}$, all values obtained from the oracles are determined by $2m$ bits (i.e., $A(f'_1), ..., A(f'_m)$ and $A_1(g_1), ..., A_1(g_m)$). ▌

Composing the canonical outer verifier of Lemma 3.8 and the canonical inner verifier $V_{\text{free-in}}$, we get the following

Theorem 7.10 For any $\epsilon > 0$ it is the case that $\mathrm{NP} \subseteq \overline{\mathrm{FPCP}}[\log, 2+\epsilon]$.

Proof: Given an NP language L and an integer m (see below), we use Lemma 3.8 to construct a 2^{-m}-good outer verifier, denoted V_{outer}, for L. Recall that this outer verifier uses logarithmic randomness (actually the randomness depends linearly on m which is a constant). Next, compose V_{outer} with the $(c \cdot 2^{-m}, 0.4, 0.4)$-good inner verifier, $V_{\text{free-in}}$, guaranteed by Lemma 7.8, where $V_{\text{free-in}}$ uses m as its integer parameter. The composed verifier has free-bit complexity $2m$ (as inherited from $V_{\text{free-in}}$ by Proposition 7.9). By Theorem 3.12 the soundness error of the composed verifier is at most $c \cdot 2^{-m} + 2^{-m}$. Selecting m to be sufficiently large (i.e., $m = \frac{2+\epsilon}{\epsilon} \cdot \log_2(c+1)$), the theorem follows. ▌

7.3 Hardness of MaxClique

Refer to Section 2.4 for definitions of the MaxClique and ChromNum problems and their associated gap problems, and to Section 2.4.3 for a description of previous work. Using the FGLSS-transformation, we get

Theorem 7.11 For any $\epsilon > 0$

(1) $\mathrm{NP} \leq_R^K$ Gap-MaxClique$_{c,s}$ for $s(N) = N^\epsilon/N$ and $c(N) = N^{1/3}/N$.

(2) $\mathrm{NP} \leq_D^K$ Gap-MaxClique$_{c,s}$ for $s(N) = N^\epsilon/N$ and $c(N) = N^{1/4}/N$.

Proof: For Part (1) we use Corollary 11.3 (below), with $r = O(\log n)$ and $k = \frac{r}{\epsilon}$. We get that NP is randomly reducible to a pcp system with randomness $r + k + O(1)$, free-bit complexity $(2+\epsilon)k$ and error probability 2^{-k}. The FGLSS-graph corresponding to the resulting pcp system has size $N = 2^{(r+k+O(1))+(2+\epsilon)k}$ and a gap in clique size of factor 2^k, which can be rewritten as $N^{1/(1+2+2\epsilon)}$. The clique size in case of input not in the language is 2^r which can be rewritten as N^ϵ. Substituting

ϵ for $\epsilon/2$, the claim of Part (1) follows. For Part (2) we use Corollary 11.5, and get a pcp system for NP with randomness $r + (2+\epsilon)k$, free-bit complexity $(2+\epsilon)k$ and error probability 2^{-k}. Using the FGLSS-construction on this system, the claim of Part (2) follows. ▌

Combining the above with a recent reduction of Fürer [Fu], we get

Theorem 7.12 For any $\epsilon > 0$
(1) NP $\leq_R^K$ Gap-ChromNum$_{c,s}$ for $c(N)/s(N) = N^{\frac{1}{5}-\epsilon}$.
(2) NP $\leq_D^K$ Gap-ChromNum$_{c,s}$ for $c(N)/s(N) = N^{\frac{1}{7}-\epsilon}$.

Part II
Proofs and Approximation: Potential and Limitations

8 The reverse connection and its consequences

Feige et al. [FGLSS] describe a procedure which takes a verifier V, and an input x and constructs a graph, which we denote $\mathcal{G}_V(x)$, whose vertices correspond to possible accepting transcripts in V's computation and edges corresponding to consistent/non-conflicting computations. They then show the following connection between the maximum (over all possible oracles) acceptance probability of the verifier and the clique size in the graph. Recall that $\texttt{ACC}\,[\,V(x)\,] = \max_\pi \Pr_R\,[V^\pi(x;R)=0]$ is the maximum accepting probability. Also recall that MaxClique(G) is the maximum clique size.

Theorem 8.1 ([FGLSS]) If, on input x, a verifier V tosses r coins then the following relationship holds:

$$\texttt{ACC}\,[\,V(x)\,] = \frac{\text{MaxClique}(\mathcal{G}_V(x))}{2^r}\;.$$

In this section we essentially show an inverse of their construction.

8.1 The Clique-Gap Verifier

We stress that by the term *graph* we mean an undirected simple graph (i.e., no self-loops or parallel edges).

Theorem 8.2 (Clique verifier of ordinary graphs): There exists a verifier, denoted W, of logarithmic randomness-complexity, logarithmic query-length and zero free-bit complexity, that, on input an N-node graph G, satisfies

$$\texttt{ACC}\,[\,W(G)\,] = \frac{\text{MaxClique}(G)}{N}\;.$$

Furthermore, $\mathcal{G}_W(G)$ is isomorphic to G where the isomorphism is easily computable. Lastly, given a proof/oracle π we can construct in polynomial-time a clique of size pN in G, where p is the probability that W accepts G with oracle access to π.

Proof: On input a graph G on N nodes, the verifier W works with proofs of length $\binom{N}{2} - |E(G)|$. The proof π is indexed by the edges in $\overline{G}$ (i.e., non-edges in G). For clarity we assume that the binary value $\pi(\{u,v\})$ is either u or v. This is merely a matter of encoding (i.e., consider a 1-1 mapping of the standard set of binary values, $\{0,1\}$, to the set $\{u,v\}$). On input G and access to oracle π, the verifier W acts as follows:

- Picks uniformly a vertex u in the vertex set of G.
- For every $\{u,v\} \in E(\overline{G})$, the verifier W queries the oracle at $\{u,v\}$ and rejects if $\pi(\{u,v\}) \neq u$.
- If the verifier did not reject by now (i.e., all queries were answered by u), it accepts.

Properties of W. Clearly, W tosses $\log_2 N$ coins. Also, once W picks a vertex u, the only pattern it may accepts is $(u,u,\ldots,u)$. Thus the free-bit complexity of W is 0. To analyze the probability that W accepts the input G, when given the best oracle access, we first prove the following:

Claim. The graphs $\mathcal{G}_W(G)$ and G are isomorphic.

Proof. The proof is straightforward. One needs first to choose an encoding of accepting transcripts of the computation of W on input G. We choose to use the "full transcript" in which the random coins as well as the entire sequence of queries and answers is specified. Thus, a generic accepting transcript has the form

$$T_u \stackrel{\text{def}}{=} (u, (\{u, v_1\}, u), ..., (\{u, v_d\}, u))$$

where u is the random vertex selected by the verifier and $\{v_1, ..., v_d\}$ the set of non-neighbors of u. We stress that T_u is the only accepting transcript in which the verifier has selected the vertex u. Also, for each vertex u, the transcript T_u is accepting. Thus, we may consider the 1-1 mapping, ϕ, that maps T_u to u. We claim that ϕ is an isomorphism between $\mathcal{G}_W(G)$ and G.

Suppose that T_u and T_v are adjacent in $\mathcal{G}_W(G)$. Then, by definition of the FGLSS graph, these transcripts are consistent. It follows that the same query can not appear in both (accepting) transcripts (otherwise it would have been given conflicting answers). By definition of W we conclude that (u, v) is NOT a non-edge; namely, $(\phi(T_u), \phi(T_v)) = (u, v) \in E(G)$. Suppose, on the other hand, that $(u, v) \in E(G)$. It follows that the query $\{u, v\}$ does not appear in either T_u or T_v. Since no other query may appear in both transcripts, we conclude that the transcripts are consistent and thus T_u and T_v are adjacent in $\mathcal{G}_G(W)$. $\square$

By Theorem 8.1 it now follows that the probability that W accepts on input G, given the best oracle, is MaxClique$(\mathcal{G}_W(G))/N$ which by the above equals MaxClique$(G)/N$. Furthermore, given a proof π which makes W accept G with probability p, the accepting random strings of W constitute a clique of size pN in $\mathcal{G}_W(G)$. These accepting random strings can be found in polynomial-time and they encode vertices of G (which form a clique in G). ▌

We now generalize the above construction to get verifiers which indicate the existence of large cliques in layered graphs. An *(L, M, N)-layered graph* is an N-vertex graph in which the vertices are arranged in L layers so that there are no edges between vertices in the same layer and there are at most M vertices in each layer. We use a convention by which, whenever a layered graph is given to some algorithm, a partition into layers is given along with it (i.e., is implicit in the encoding of the graph).

Theorem 8.3 (Clique verifier for layered graphs): There exists a verifier, denoted W, of logarithmic randomness-complexity and logarithmic query-length that, on input an (L, M, N)-layered graph G has free-bit complexity $\log_2 M$, average free-bit complexity $\log_2(N/L)$ and satisfies

$$\texttt{ACC}\,[\,W(G)\,] = \text{MaxClique}(G)/L\ .$$

Furthermore, $\mathcal{G}_W(G)$ is isomorphic to G where the isomorphism is easily computable. Lastly, given a proof/oracle π we can construct in polynomial-time a clique of size pL in G, where p is the probability that W accepts G with oracle access to π.

Proof: On input a (L, M, N)-layered graph G, the verifier W works with proofs consisting of two parts. The first part assigns every layer (i.e., every integer $i \in [L]$) a vertex in the layer (i.e., again we use a redundant encoding by which the answers are vertex names rather then an index between 1 and the number of vertices in the layer). The second part assigns pairs of non-adjacent (in G) vertices, a binary value, which again is represented as one of the two vertices. On input G and access to oracle π, the verifier W acts as follows:

- Picks uniformly a layer i in $\{1, ..., L\}$.
- Queries π at i and obtains as answer a vertex u. If u is not in the i^{th} layer of G then the verifier rejects. (Otherwise, it continues as follows.)

- For every $\{u,v\} \in E(\overline{G})$, the verifier W queries the oracle at $\{u,v\}$ and rejects if $\pi(\{u,v\}) \neq u$. (Actually, it is not needed to query the oracle on pairs of vertices belonging to the same layer.)
- If the verifier did not reject by now (i.e., all queries were answered by u), it accepts.

Properties of W. Here W tosses $\log_2 L$ coins. Once the first query of W is answered, specifying a vertex u, the only pattern it may accept in the remaining queries is $(u,u,\ldots,u)$. Thus, the free-bit complexity of W is $\log_2 M$, accounting for the first query which may be answered arbitrarily in $\{1,...,m\}$, where $m \leq M$ is the number of vertices in the chosen layer. The average free-bit complexity is $\log_2(N/L)$ (as N/L is the average number of vertices in a layer of the graph G). Again, we can prove that $\mathcal{G}_W(G) = G$ and the theorem follows.

Proof. Here, the accepting transcripts of W, on input G, correspond to a choice of a layer, i, and a vertex in the i^{th} layer (since once a vertex is specified by the first answer there is only one accepting way to answer the other queries). Thus, a generic accepting transcript has the form

$$T_u \stackrel{\text{def}}{=} (i,(i,u),(\{u,v_1\},u),...,(\{u,v_d\},u))$$

where i is the layer selected by the verifier, u is a vertex in the i^{th} layer of G and $\{v_1,...,v_d\}$ the set of non-neighbors of u. Again, T_u is the only accepting transcript in which the verifier has selected the vertex u, and for each vertex u, the transcript T_u is accepting. Again, we consider the 1-1 mapping, ϕ, that maps T_u to u, and show that it is an isomorphism between $\mathcal{G}_W(G)$ and G.

Suppose that T_u and T_v are adjacent in $\mathcal{G}_G(W)$. Then, by definition of the FGLSS graph, these transcripts are consistent. We first note that u and v cannot appear in the same layer of G (otherwise the first query in the transcript would yield conflicting answers). Again, the same two-vertex query can not appear in both (accepting) transcripts, and we conclude that $(\phi(T_u),\phi(T_v)) = (u,v) \in E(G)$. Suppose, on the other hand, that $(u,v) \in E(G)$. Clearly, u and v belong to different layers and as before the query (u,v) does not appear in either T_u or T_v. Since no other two-vertex query may appear in both transcripts, we conclude that the transcripts are consistent and thus T_u and T_v are adjacent in $\mathcal{G}_G(W)$. $\square$

The theorem follows as before. ▌

Remark 8.4 *The clique verifier W is adaptive: the answer to its first query determines (all) the other queries. We wonder if it is possible to construct a non-adaptive clique verifier with properties as claimed in Theorem 8.3.*

8.2 Reversing the FGLSS reduction

We are interested in problems exhibiting a gap in Max-Clique size between positive and negative instances. Recall that $\overline{\text{MaxClique}}(G) = \text{MaxClique}(G)/N$ is the fraction of nodes in a maximum clique of N-node graph G. Also recall from Section 2.4 that the Gap-$\text{MaxClique}_{c,s}$ promise problem is (A,B) where A is the set of all graphs G with $\overline{\text{MaxClique}}(G) \geq c(N)$, and B is the set of all graphs G with $\overline{\text{MaxClique}}(G) < s(N)$. The *gap* of this problem is defined to be c/s. As a direct consequence of Theorem 8.2, we get

Corollary 8.5 For all functions $c,s\colon \mathcal{Z}^+ \to [0,1]$ we have Gap-$\text{MaxClique}_{c,s} \in \text{FPCP}_{c,s}[\log,0,\text{poly}]$.

The above corollary transforms the gap in the promise problem into a gap in a pcp system. However, the accepting probabilities in this pcp system are very low (also on yes-instances). Below, we use Theorem 8.3 to obtain pcp systems with perfect (resp., almost-perfect) completeness for this promise problem. We start by presenting two randomized reductions of the promise problem to a layer version. Alternative methods are presented in Section 11 (cf., Proposition 11.6).

Proposition 8.6 (Layering the clique promise problem):

(1) (Obtaining a perfect layering): There exists a polynomial-time randomized transformation, T, of graphs into layered graphs so that, on input a graph G, integers C and L, the transformation outputs a subgraph $H = T(G, C, L)$ of G in L layers such that if $\text{MaxClique}(G) \geq C$ then

$$\Pr[\,\text{MaxClique}(H) < L\,] < L \cdot 2^{-\frac{C}{2L}}$$

Furthermore, with probability $1 - L \cdot 2^{-N/3L}$, no layer of H contains more than $2 \cdot \frac{N}{L}$ nodes.

(2) (Using logarithmic randomness): There exists a polynomial-time randomized transformation, T, of graphs into layered graphs so that, on input a graph G, integers C and L, the transformation outputs a subgraph $H = T(G, C, L)$ of G in L layers such that if $\text{MaxClique}(G) \geq C$ then

$$\Pr[\,\text{MaxClique}(H) \leq (1-\epsilon)\cdot L\,] < \frac{L}{\epsilon C}$$

for every $\epsilon \in [0, 1]$. Furthermore, the transformation uses logarithmically many coins. Also, with probability $1 - \frac{L}{\epsilon N}$, at most ϵL layers of H contains more than $2 \cdot \frac{N}{L}$ nodes.

Proof: The *first transformation* consists of assigning to each vertex of G a randomly chosen layer of H. Namely, we construct the graph H which is a subgraph of G by uniformly selecting for each vertex v a layer $l(v) \in [L]$ and copying only the edges of G which connect vertices placed in different layers (of H). The construction can be carried out in random polynomial-time and we show that if the original graph has a clique of size C then with high probability the resulting graph has a clique of size L, provided $L \ll \frac{C}{2}\log_2 L$.

Claim 1. Suppose that G has a clique of size C, denoted S. Then, the probability that all vertices in S were placed in less than L layers is at most $L \cdot 2^{-\frac{C}{2L}}$.

Proof. We start by bounding, for each i, the probability that no vertex of S is placed in the i^{th} layer. For each $v \in S$, we introduce the 0-1 random variable ζ_v so that $\zeta_v = 1$ if v is placed in the i^{th} layer (i.e., $l(v) = i$) and $\zeta_v = 0$ otherwise. Let $t \stackrel{\text{def}}{=} C/L$. Then, $\mathbf{E}[\sum_{v\in S}\zeta_v] = t$. Using a multiplicative Chernoff bound [MoRa], we get

$$\begin{aligned}\Pr[\,\forall v \in S : l(v) \neq i\,] &= \Pr\left[\sum_{v\in S}\zeta_v = 0\right]\\ &< 2^{-\frac{t}{2}}\end{aligned}$$

Call the i^{th} layer *bad* if no vertex of S is placed in it. By the above, the probability that there exists a bad layer is smaller than $L \cdot 2^{-t/2}$, and the claim follows. □

It is left to bound the probability that a particular layer contains more than twice the expected number of vertices. Using again a multiplicative Chernoff bound, this probability is at most $2^{-N/3L}$ and the first part of the proposition follows.

The *second transformation* consists of selecting randomly a Universal$_2$ Hashing function (a.k.a., pair-wise independent hash function) mapping the vertices of the graph G into the layer-set $[L]$. Namely, suppose that the function h was chosen, then we construct the graph H which is a subgraph of G by placing a vertex v (of G) in layer $h(v)$ of H, and copying only the edges of G which connect vertices placed in different layers (of H). The construction can be carried out in polynomial-time using only logarithmic randomness (for the selection of the hashing function). We show that if the original graph has a clique of size C then with high probability the resulting graph has a clique of size almost L, provided $L \ll C$.

Claim 2. Suppose that G has a clique of size C, denoted S. Then, the probability that all vertices in S were placed in less than $(1-\epsilon) \cdot L$ layers is at most $\frac{L}{\epsilon C}$.

Proof. Again, we bound, for each i, the probability that no vertex of S is placed in the i^{th} layer. For each $v \in S$, we introduce the 0-1 random variable ζ_v so that $\zeta_v = 1$ if $h(v) = i$ and $\zeta_v = 0$ otherwise. Let $t \stackrel{\text{def}}{=} C/L$ and $\zeta \stackrel{\text{def}}{=} \sum_{v\in S} \zeta_v$. Then, $\mathbf{E}[\zeta] = t$ (which is greater than 1, otherwise the claim holds vacuously). Using the pairwise independence of h and Chebyshev's inequality, we get

$$\begin{aligned} \Pr[\forall v \in S : h(v) \neq i] &= \Pr[\zeta = 0] \\ &\leq \frac{\mathbf{Var}[\sum_{v\in S} \zeta_v]}{t^2} \\ &< \frac{C/L}{t^2} = \frac{1}{t} \end{aligned}$$

Call the i^{th} layer *bad* if no vertex of S is placed in it. By the above, the expected number of bad layers is smaller than $L \cdot \frac{1}{t}$, so by Markov inequality the probability that more than ϵL layers are bad is at most $1/\epsilon t$. The claim follows. □

Again, it is left to bound the probability that a particular layer contains more than $M \stackrel{\text{def}}{=} 2N/L$. Using Chebyshev's inequality again, this probability is at most L/N. Thus, the expected number of layers having more than M vertices is at most L^2/N and it follows that the probability that ϵL layers contain more than M vertices each is at most $\frac{L^2/N}{\epsilon L} = \frac{L}{\epsilon N}$. The second part of the proposition follows. ∎

Combining Theorem 8.3 and Proposition 8.6, we obtain the following. (Refer to Section 2.3 for what it means for a promise problem to reduce to a complexity class.)

Proposition 8.7 (Reversing the FGLSS-reduction, general form:) For any polynomial-time computable, positive functions $c, s, \epsilon \colon \mathcal{Z}^+ \to [0,1]$ we have

(1) (Randomized reduction to a pcp with perfect completeness):

$$\mathsf{Gap}\text{-MaxClique}_{c,s} \leq_R^K \text{FPCP}_{1,s'}[\log, f']$$

where $f'(N) \stackrel{\text{def}}{=} \log_2(1/c(N)) + \log_2 \log_2 N + 2$ and $s'(N) \stackrel{\text{def}}{=} 2\log_2 N \cdot \frac{s(N)}{c(N)}$.

(2) (A pcp with almost-perfect completeness):

$$\mathsf{Gap}\text{-MaxClique}_{c,s} \in \text{FPCP}_{1-4\epsilon,s'}[\log, f']$$

where $f'(N) \stackrel{\text{def}}{=} 1 + \log_2(1/c(N)) + 2\log_2(1/\epsilon(N))$ and $s'(N) \stackrel{\text{def}}{=} \frac{1}{\epsilon(N)^2} \cdot \frac{s(N)}{c(N)}$.

Proof: For the *second part*, we construct a verifier for the promise problem as follows. On input an N-vertex graph G, the verifier computes $C \stackrel{\text{def}}{=} N \cdot c(N)$, $\epsilon \stackrel{\text{def}}{=} \epsilon(N)$ and $L \stackrel{\text{def}}{=} \epsilon^2 C$. It invokes the second transformation of Proposition 8.6, obtaining a (L, N, N)-layered graph $H = T(G, C, L)$. (We stress that this transformation requires only logarithmically many coin tosses.) Next, the verifier modifies H into H' by omitting (the minimum number of) vertices so that no layer of H' has more than $2N/L$ vertices. Finally, the verifier invokes the clique-verifier W of Theorem 8.3 on input H'.

The free-bit complexity of the verifier constructed above is $\log_2(2N/L) = 1+\log_2(1/c(N))+2\log_2(1/\epsilon(N))$. Suppose that G is a no-instance of the promise problem. Using $\text{MaxClique}(H') \leq \text{MaxClique}(G)$ and Theorem 8.3, it follows that the constructed verifier accepts G with probability at most $\frac{\text{MaxClique}(H')}{L} \leq \frac{s(N)}{\epsilon^2(N)\cdot c(N)}$. Suppose, on the other hand, that G is a yes-instance of the promise problem. Then, with probability at least $1 - \frac{L}{\epsilon C} = 1-\epsilon$ we have $\text{MaxClique}(H) \geq (1-\epsilon)\cdot L$, and with probability at least $1 - \frac{L}{\epsilon N} > 1-\epsilon$ we have $\text{MaxClique}(H') \geq \text{MaxClique}(H) - \epsilon L$. Thus, with probability at least $1-2\epsilon$, we have $\text{MaxClique}(H') \geq (1-2\epsilon)\cdot L$. It follows that the constructed verifier, when given oracle access to an appropriate proof, accepts G with probability at least $1-4\epsilon$.

For the *first part*, we define a promise problem which refers to gaps in cliques of layered graphs. Specifically,

Definition. For any function $\ell : \mathcal{Z}^+ \to \mathcal{Z}^+$ and $s : \mathcal{Z}^+ \to [0,1]$, we define the promise problem $\mathsf{Gap}\text{-}\text{LG}_{\ell,s}$ be the pair (A, B), where–

(1) A is the set of all $(\ell(N), \frac{2N}{\ell(N)}, N)$-layered graphs G with $\text{MaxClique}(G) = \ell(N)$, and

(2) B is the set of all $(\ell(N), \frac{2N}{\ell(N)}, N)$-layered graphs G with $\text{MaxClique}(G) < s(N)\cdot\ell(N)$.

The *gap* of this problem is defined to be $1/s$.

Using the first transformation of Proposition 8.6, we obtain $\mathsf{Gap}\text{-MaxClique}_{c,s} \leq_R^K \mathsf{Gap}\text{-}\text{LG}_{\ell,s'}$, where $\ell(N) = \frac{c(N)\cdot N}{2\log_2 N}$ and $s'(N) = \frac{s(N)\cdot N}{\ell(N)} = 2\log_2 N \cdot \frac{s(N)}{c(n)}$. On the other hand, Theorem 8.3 asserts that $\mathsf{Gap}\text{-}\text{LG}_{\ell,s'} \in \text{FPCP}_{1,s'}[\log, f']$, where $f'(N) \overset{\text{def}}{=} \log_2(2N/\ell(N))$. Observing that $f'(N) = 1 + \log_2 \frac{2\log_2 N}{c(N)}$ (which equals $\log_2(1/c(N)) + \log_2\log_2 N + 2$), the proposition follows. ▌

Each of the two parts of Proposition 8.7 shows that the well-known method of obtaining clique-approximation results from efficient pcp systems (cf., [FGLSS, BeSc, Zuc, FeKi1, BeSu]) is "complete" in the sense that if clique-approximation can be shown NP-hard then this can be done via this method. The precise statement is given in Theorems 8.10 and 8.11 (below). As a preparatory step, we first provide an easier-to-use form of the above proposition. The restriction that f be a constant is only for notational simplicity (as otherwise, given f as a function of $N = \|G\|$, one needs to repharse it as a function of $n = |x|$).

Proposition 8.8 (Reversing the FGLSS-reduction, easy to use form:) Let $f > 0$ be a constant and c, s: $\mathcal{Z}^+ \to [0,1]$ be polynomial-time computable so that

$$\frac{c(N)}{s(N)} \geq N^{\frac{1}{1+f}}$$

Then, for every $\epsilon > 0$,

(1) (Randomized reduction to a pcp with perfect completeness):

$$\mathsf{Gap}\text{-MaxClique}_{c,s} \leq_R^K \overline{\text{FPCP}}[\log, f+\epsilon]$$

(2) (A pcp with almost-perfect completeness):

$$\mathsf{Gap}\text{-MaxClique}_{c,s} \in \overline{\text{FPCP}}_{1-o(1)}[\log, f+\epsilon]$$

Proof: We merely invoke Proposition 8.7, and calculate the amortized free-bit complexity of the resulting verifier. We may assume that $s(N) \geq 1/N$. Thus (using $c(N)/s(N) \geq N^{\frac{1}{1+f}}$), we have $c(N) \geq N^{\frac{1}{1+f}}/N = N^{\frac{-f}{1+f}}$ and $1/c(N) \leq N^{\frac{f}{1+f}}$.

For Part 1, we let $\alpha(N) \stackrel{\rm def}{=} 2\log_2 N$, and set $f'(N) \stackrel{\rm def}{=} \log_2(1/c(N)) + \log_2 \alpha(N)$ and $s'(N) \stackrel{\rm def}{=} \alpha(N) \cdot \frac{s(N)}{c(N)}$. By invoking Proposition 8.7 (Part 1) we find that $\textsf{Gap}\text{-MaxClique}_{c,s} \leq_R^K \text{FPCP}_{1,s'}[\log, f']$ and $\textsf{Gap}\text{-MaxClique}_{c,s} \leq_R^K \overline{\text{FPCP}}[\log, \overline{f'}]$, for $\overline{f'} = \frac{f'}{\log(1/s')}$, follows. It now remains to argue that for any $\epsilon > 0$, $\overline{f'} \leq f + \epsilon$.

Using the lower bounds on $c(N)$ and $c(N)/s(N)$, we obtain $f'(N) \leq \frac{f}{1+f} \log_2 N + \log_2 \alpha(N)$ and $\log(1/s'(N)) \geq \frac{1}{1+f} \cdot \log_2 N - \log_2 \alpha(N)$. Selecting a sufficiently small $\delta > 0$ and using $\log_2 \alpha(N) < \delta \cdot \log_2 N$, we get

$$\begin{aligned} \overline{f'} &\leq \frac{\frac{f}{1+f}\log_2 N + \log_2 \alpha(N)}{\frac{1}{1+f}\log_2 N - \log_2 \alpha(N)} \\ &< \frac{\frac{f}{1+f} + \delta}{\frac{1}{1+f} - \delta} \end{aligned}$$

and so Part 1 follows. For Part 2, we let α be a slowly decreasing function s.t. $\alpha(N) = o(1)$ but $\log_2(1/\alpha(N)) = o(\log N)$. We set $f'(N) \stackrel{\rm def}{=} \log_2(1/c(N)) + 2\log_2(1/\alpha(N))$ and $s'(N) \stackrel{\rm def}{=} \frac{1}{\alpha(N)^2} \cdot \frac{s(N)}{c(N)}$. By invoking Proposition 8.7 (Part 2) we get $\textsf{Gap}\text{-MaxClique}_{c,s} \in \text{FPCP}_{1-\alpha,s'}[\log, f']$. Since $\alpha(N) = o(1)$, we conclude that $\textsf{Gap}\text{-MaxClique}_{c,s} \in \overline{\text{FPCP}}_{1-o(1)}[\log, \overline{f'}]$ for $\overline{f'} = \frac{f'}{\log_2(1/s')}$. Again, it remains to argue that for any $\epsilon > 0$, $\overline{f'} \leq f + \epsilon$. Using the lower bound on $c(N)$ and $c(N)/s(N)$, we obtain $f'(N) \leq \frac{f}{1+f} \log_2 N - 2\log_2 \alpha(N)$ and $\log_2(1/s'(N)) = 2\log_2 \alpha(N) + \frac{1}{1+f} \log_2 N$. Selecting a sufficiently small $\delta > 0$ and using $\log_2(1/\alpha(N)) < \delta \cdot \log_2 N$, we get

$$\begin{aligned} \overline{f'} &\leq \frac{\frac{f}{1+f}\log_2 N + 2\log_2(1/\alpha(N))}{\frac{1}{1+f}\log_2 N - 2\log_2(1/\alpha(N))} \\ &< \frac{\frac{f}{1+f} + \delta}{\frac{1}{1+f} - \delta} \end{aligned}$$

and Part 2 follows. ▌

8.3 Main Consequences

Let us first state the FGLSS-reduction.

Theorem 8.9 (The FGLSS-reduction, revisited:) Let $f > 0$ be a constant and $c, s \colon \mathcal{Z}^+ \to [0,1]$. Then, for every $\epsilon > 0$,

$$\overline{\text{FPCP}}_{1-o(1)}[\log, f] \leq_R^K \textsf{Gap}\text{-MaxClique}_{c,s}$$

where $c(N)/s(N) \geq N^{1/(1+f+\epsilon)}$. Furthermore, in case the proof system is of perfect completeness, we have $c(N) = N^{-f/(1+f+\epsilon)}$ and $s(N) = N^{-(1+f)/(1+f+\epsilon)}$.

Proof: We first amplify the gap of the pcp-verifier (cf., Corollary 11.3) and then by apply the bare FGLSS-reduction (see Theorem 8.1 and [FGLSS]) to the amplified verifier. Specifically, for any problem

Π in $\overline{\mathrm{FPCP}}[\log, f]$, we first obtain $\Pi \leq_R^K \mathrm{FPCP}_{1,2^{-t}}[(1+\epsilon)\cdot t, f\cdot t]$, where $t(n) = \gamma \log_2 n$ (with the constant γ determined by the constant $\epsilon > 0$). The FGLSS-reduction now yields a graph of size $N \stackrel{\mathrm{def}}{=} 2^{(1+\epsilon+f)\cdot t(n)}$ with gap $2^{t(n)}$ (which can be written as $N^{\frac{1}{1+\epsilon+f}}$). Specifically, the clique size for a YES-instance (resp., NO-instance) is at least $2^{(1+\epsilon)\cdot t(n)} = N^{\frac{1+\epsilon}{1+\epsilon+f}}$ (resp., at most $2^{\epsilon\cdot t(n)} = N^{\frac{\epsilon}{1+\epsilon+f}}$).

A similar procedure may be applied for any Π in $\overline{\mathrm{FPCP}}_{1-o(1)}[\log, f]$. Specifically, by definition, for some function m, $\Pi \in \mathrm{FPCP}_{c,2^{-m}\cdot c}[\log, m\cdot f]$, for $c(n) = 1-o(1)$ (but we are not going to use the bound on c). Using Proposition 11.1 and Proposition 11.2 (Part 2), we first obtain $\Pi \leq_R^K \mathrm{FPCP}_{c',2^{-t}\cdot c'}[(1+\epsilon)\cdot t, f\cdot t]$, where $c'(n) = c(n)^{t(n)/m(n)}$ and $t(n) = \gamma \log_2 n$ (with the constant γ determined by the constant $\epsilon > 0$). The FGLSS-reduction now yields a graph of size $N \stackrel{\mathrm{def}}{=} 2^{(1+\epsilon+f)\cdot t(n)}$ with gap $2^{t(n)}$ as above. ▌

Interestingly, the gap (for MaxClique) created by FGLSS-reduction is independent of the location of the gap in the pcp system. The main result of this section is –

Theorem 8.10 Let f be a constant. Then the following statements are equivalent:

(1) For all $\epsilon > 0$ it is the case that NP reduces to Gap-MaxClique$_{c,s}$ with gap $c(N)/s(N) = N^{1/(1+f+\epsilon)}$.

(2) For all $\epsilon > 0$ it is the case that NP reduces to $\overline{\mathrm{FPCP}}[\log, f+\epsilon]$.

In both items the reduction is randomized. Furthermore the equivalence holds both for Karp and for Cook reductions.

Proof: The direction (2) $\Rightarrow$ (1) follows again by Theorem 8.9. The reverse direction follows by Part 1 of Proposition 8.8. ▌

An alternative statement is provided by the following theorem. Here the second item (existence of pcp systems with certain parameters) is weaker than in the previous theorem, but this allows the (1) $\Rightarrow$ (2) direction to be proven via a deterministic reduction (instead of the randomized reduction used in the analogous proof above). Recall that $\overline{\mathrm{FPCP}}_{1-o(1)}[\,\cdot,\, f\,]$ is the class of problems having a proof system with almost-perfect completeness (i.e., $c = 1 - o(1)$) and amortized free-bit complexity f.

Theorem 8.11 Let f be a constant. Then the following statements are equivalent:

(1) For all $\epsilon > 0$ it is the case that NP reduces to Gap-MaxClique$_{c,s}$ with gap $c(N)/s(N) = N^{1/(1+f+\epsilon)}$.

(2) For all $\epsilon > 0$ it is the case that NP reduces to $\overline{\mathrm{FPCP}}_{1-o(1)}[\,\log,\, f+\epsilon\,]$.

In both items the reduction is randomized and the equivalence holds both for Karp and for Cook reductions. Furthermore, if Item (1) holds with respect to deterministic reductions so does Item (2). Thus, if Item (1) holds with a deterministic Karp reduction then $\mathrm{NP} \subseteq \overline{\mathrm{FPCP}}_{1-o(1)}[\,\log,\, f+\epsilon\,]$.

Proof: The direction (2) $\Rightarrow$ (1) follows by applying Theorem 8.9. The reverse direction follows by Part 2 of Proposition 8.8. ▌

8.4 More Consequences

The equivalence between clique and FPCP described above turns out be a useful tool in the study of the hardness of the clique and chromatic number problems. Here we describe some applications. The first application is merely a rephrasing of the known reductions from the Max Clique problem to the Chromatic number problem in a simpler and more convenient way. The remaining applications use the fact that the equivalence between FPCP and Max Clique allows us to easily shift gaps, in the Max Clique problem, from one place to another. Loosely speaking, these applications use the fact that the complexity of the promise problem Gap-MaxClique$_{c,s}$ remains unchanged when changing the parameters c and s so the $\frac{\log_2 c(N)}{\log_2 s(N)}$ remains invariant. We stress that the ratio $\frac{c(N)}{s(N)}$ does not remain invariant.

Rephrasing reductions from Max Clique to Chromatic Number. Starting with the work of Lund and Yannakakis [LuYa], there have been several works on showing the hardness of approximating the Chromatic number, which reduce the Max Clique problem to the Chromatic number problem: see Section 2.4.3 for a description. Yet none of these results could be stated cleanly in terms of a reduction from Max Clique to Chromatic Number without loss of efficiency - i.e., the theorems could not be stated as saying "If approximating Max Clique to within a factor of N^α is NP-hard, then approximating Chromatic Number to within a factor of $N^{h(\alpha)}$ is NP-hard." The reason for the lack of such a statement is that these reductions use the structure of the graph produced by applying an FGLSS-reduction to a FPCP result, and are hence really reductions from FPCP to Chromatic Number rather than reductions from Max Clique to Chromatic Number. However now we know that FPCP and Max Clique are equivalent, so we can go back and rephrase the old statements. Thus results of [LuYa, KLS, BeSu, Fu] can be summarized as:

For every $\alpha, \epsilon, \gamma > 0$, $\text{Gap-MaxClique}_{N^{\alpha-1},N^{\epsilon-1}} \leq_R^K \text{Gap-ChromNum}_{N^{-(\epsilon+\gamma)},N^{-h(\alpha)}}$, where

(1) $h(\alpha) = \min\{\frac{1}{6}, \frac{\alpha}{5-4\alpha}\}$ [LuYa].

(2) $h(\alpha) = \min\{\frac{1}{11}, \frac{\alpha}{5+\alpha}\}$ [KLS].

(3) $h(\alpha) = \min\{\frac{1}{4}, \frac{\alpha}{3-2\alpha}\}$ [BeSu].

(4) $h(\alpha) = \min\{\frac{1}{3}, \frac{\alpha}{2-\alpha}\}$ [Fu].

We note that it is an open problem whether one can get a reduction in which $h(\alpha) \to 1$ as $\alpha \to 1$. We also note that Fürer's reduction is randomized while the rest are deterministic.

Reductions among Max Clique Problems. Next we present an invariance of the Gap Clique problem with respect to shifting of the gaps. The following result has also been independently observed by Feige [Fe1], where he uses a randomized graph product to show the result. Our description uses the properties of fpcp and its equivalence to clique approximation.

Theorem 8.12 Let $k, \epsilon_1, \epsilon_2$ be real numbers such that $k \geq 1$ and $0 \leq \epsilon_1 < \epsilon_2 \leq 1$. Then the following hold:

(1) $\text{Gap-MaxClique}_{N^{-\epsilon_2},N^{-k\epsilon_2}} \leq_D^K \text{Gap-MaxClique}_{N^{-\epsilon_1},N^{-k\epsilon_1}}$. (Deterministic reduction.)

(2) $\text{Gap-MaxClique}_{N^{-\epsilon_1},N^{-k\epsilon_1}} \leq_R^K \text{Gap-MaxClique}_{\frac{1}{2}\cdot N^{-\epsilon_2},2\cdot N^{-k\epsilon_2}}$.

Proof: Part (1) is proved via a well-known graph theoretic trick. Let G be an instance of $\text{Gap-MaxClique}_{N^{-\epsilon_2},N^{-k\epsilon_2}}$ with N nodes. We take the graph-product of G with a complete graph on m nodes, to get a graph H on $M = mN$ nodes. (By a graph-product of two graphs $G_1(V_1, E_1)$ and $G_2(V_2, E_2)$ we mean a graph with vertex set $V_1 \times V_2$ where vertices (u_1, u_2) and (v_1, v_2) are connected iff $(u_i, v_i) \in E_i$ for both $i = 1, 2$.) We choose m so that if G has a clique of size $N^{1-\epsilon_2}$, then H has a clique of size $M^{1-\epsilon_1}$. Specifically, setting $m = N^{\frac{\epsilon_2-\epsilon_1}{\epsilon_1}}$, the requirement is satisfied (as a clique of size $N^{1-\epsilon_2}$ in G yields a clique of size $m \cdot N^{1-\epsilon_2} = N^{\frac{\epsilon_2-\epsilon_1}{\epsilon_1}+1-\epsilon_2} = M^{\frac{\epsilon_1}{\epsilon_2}\cdot\frac{\epsilon_2(1-\epsilon_1)}{\epsilon_1}}$ in H.) Under this choice of m we will show that if G has no cliques of size $N^{1-k\epsilon_2}$ then H has no cliques of size $M^{1-k\epsilon_1}$. This will complete the proof of part (1).

Suppose H has a clique of size $M^{1-\epsilon_1}$. Then, by construction, G must have a clique of size

$$\begin{aligned} \frac{M^{1-\epsilon_1}}{m} &= \frac{N^{1-\epsilon_1}}{m^{\epsilon_1}} \\ &= N^{1-\epsilon_1-\frac{\epsilon_2-\epsilon_1}{\epsilon_1}\cdot\epsilon_1} \end{aligned}$$

and the claim follows.

For part (2) we use the equivalence between FPCP and gaps in MaxClique and apply amplification properties of FPCP. Let $c(N) = N^{-\epsilon_1}$ and $s(N) = N^{-k\epsilon_1}$. Then, using Corollary 8.5 (for line 1 below), Proposition 11.1 (for line 2) and Part (2) of Proposition 11.2 (for line 3), we get

$$\begin{aligned}
\textsf{Gap}\text{-MaxClique}_{N^{-\epsilon_1},N^{-k\epsilon_1}} &\in \text{FPCP}_{c,s}[\log_2 N, 0, N^2] \\
&\subseteq \text{FPCP}_{c^t,s^t}[t\cdot\log_2 N, 0, N^2] \quad \text{(for any integer constant } t \geq 1.) \\
&\leq_R^K \text{FPCP}_{\frac{1}{2}\cdot c^t, 2\cdot s^t}[\log_2(N^2/s^t), 0, N^2]
\end{aligned}$$

The choice of the integer t will be determined later.

Now, we go back to the clique-gap promised problem. Applying the FGLSS-reduction to the pcp class $\text{FPCP}_{\frac{1}{2}\cdot c^t, 2\cdot s^t}[\log_2(N^2/s^t), 0, N^2]$ we obtain an instance of $\textsf{Gap}\text{-MaxClique}_{\frac{1}{2}N^{-\epsilon_1 t}, 2N^{-k\epsilon_1 t}}$ on an M-vertex graph, where $M = \frac{N^2}{s^t} = N^{2+k\epsilon_1 t}$. To clarify the last assertion and the rest of the proof, we introduce the notation $\textsf{Gap}\text{-MaxClique}_{\alpha(N),\beta(N)}(N)$ which makes explicit the size parameter to which the promise problem refers. Thus, letting $\gamma \stackrel{\text{def}}{=} \frac{t}{2+tk\epsilon_1}$, we have obtained

$$\textsf{Gap}\text{-MaxClique}_{N^{-\epsilon_1},N^{-k\epsilon_1}}(N) \leq_R^K \textsf{Gap}\text{-MaxClique}_{\frac{1}{2}M^{-\gamma\epsilon_1}, 2M^{-k\cdot\gamma\epsilon_1}}(M)$$

(with M polynomial in N). Now, part (2) follows by setting t so that $\gamma = \frac{t}{2+tk\epsilon_1} \geq \frac{\epsilon_2}{\epsilon_1}$ and $t = \lceil \frac{2\epsilon_2}{(1-k\epsilon_2)\epsilon_1} \rceil$ will do. (Actually, we get $\textsf{Gap}\text{-MaxClique}_{N^{-\epsilon_1},N^{-k\epsilon_1}}(N) \leq_R^K \textsf{Gap}\text{-MaxClique}_{\frac{1}{2}M^{-\epsilon_2'}, 2M^{-k\epsilon_2'}}(M)$, for $\epsilon_2' \geq \epsilon_2$, but this can be corrected by invoking item (1).) ▌

The following theorem was first shown by Blum [Bl], using the technique of randomized graph products. It essentially uses the gap-shifting idea to show that a seemingly very weak approximator to the clique (say, $N^{1-\epsilon}$-approximation algorithm for some $\epsilon > 0$), can be used to obtain a very good approximator to the clique number in graphs which are guaranteed to have very large cliques. In particular, using such an algorithm, if a graph has a clique of size $\frac{N}{k}$, then a clique of size $\frac{N}{k^{\frac{1}{\epsilon}}}$ can be found in such a graph in polynomial time. As observed by Blum, this can be translated into significantly better algorithms for approximate coloring of a three colorable graph than known currently (see Item (1) in Corollary 8.15 below). Here we derive the theorem using FPCP and the gap-shifting techniques. The parameters are generalized so as to be able to conclude, say, that even if we have a $\frac{N}{2^{\sqrt{\log_2 N}}}$-approximation (for Max Clique), then we can obtain non-trivially good algorithms for 3-coloring (see Item (2) in Corollary 8.15).

Theorem 8.13 Let $\alpha \in [0,1]$, $\beta \in [0,1/2)$ and $k > 1$. Define $\epsilon : \mathcal{Z}^+ \to \mathcal{R}^+$, $c \in \mathcal{R}^+$ and $g : \mathcal{Z}^+ \to \mathcal{R}^+$ so that

$$\begin{aligned}
\epsilon(N) &= \frac{\alpha}{\log_2^\beta N} \\
c &= \frac{2}{\log_2 k} \\
\text{and } \log_2 g(N) &= \left(\frac{c^\beta \log_2 k}{\alpha}\right)^{1/(1-\beta)} \log_2^{\beta/(1-\beta)} N.
\end{aligned}$$

Then there is a randomized poly$(N^{2+c\log_2 g(N)})$-time reduction of instances of $\textsf{Gap}\text{-MaxClique}_{1/k,1/g}$ to M-vertex instances of $\textsf{Gap}\text{-MaxClique}_{\frac{1}{2}M^{-\epsilon(M)}, 2M^{-1+\epsilon(M)}}$.

Remark 8.14 *Observe that $g(N) = N^{o(1)}$. Also, for $\beta = 0$ we have $\epsilon(N) = \alpha$ and $g(N) = k^{\frac{1}{\alpha}}$. Thus, the theorem states that given a $\frac{1}{4}M^{1-2\alpha}$ approximator for clique one can solve* Gap-*MaxClique*$_{1/k,1/k'}$ *in polynomial-time, where $k' = k^{1/\alpha}$.*

Proof: As usual we first reduce Gap-MaxClique to FPCP and then amplify.

$$\begin{array}{rcl}
\textsf{Gap}\text{-MaxClique}_{1/k,1/g} & \in & \text{FPCP}_{1/k,1/g}[\log_2 N, 0, N^2] \\
& \subseteq & \text{FPCP}_{(1/k)^t,(1/g)^t}[t\log_2 N, 0, N^2] \text{ (for any function } t : \mathcal{Z}^+ \to \mathcal{Z}^+.\text{)} \\
& \leq_R^K & \text{FPCP}_{\frac{1}{2}(1/k)^t,2(1/g)^t}[\log_2 N^2 g^t, 0, N^2]
\end{array}$$

We now show that by setting $t = c\log_2 N$ and using the FGLSS-reduction, the above reduces in poly(M)-time to Gap-MaxClique$_{\frac{1}{2}M^{-\epsilon},2M^{-\epsilon+1}}$ in an M vertex graph, where $M = N^2 g(N)^t$.

In case the graph is a no-instance the size of the clique is most $2(1/g(N))^t \cdot M = 2N^2$. In the case the graph is a yes-instance then the clique size is at least $\frac{1}{2}(1/k)^t \cdot M$. Thus it suffices to show that $2N^2 \leq 2M^{\epsilon(M)}$ and $2k^t \leq 2M^{\epsilon(M)}$, respectively. Taking logs in both cases it suffices to show that

$$2\log_2 N \leq \epsilon(M)\log_2 M \quad (16)$$
$$t\log_2 k \leq \epsilon(M)\log_2 M \quad (17)$$

We first lower bound the right hand side of both equations.

$$\begin{array}{rcl}
\epsilon(M)\log_2 M & = & \alpha\log_2^{1-\beta} M \\
& \geq & \alpha\log_2^{1-\beta}(g(N)^t) \\
& \geq & \alpha t^{1-\beta}\log_2^{1-\beta} g(N) \\
& = & \alpha\cdot(c\log_2 N)^{1-\beta}\cdot\left(c^\beta\frac{\log_2 k}{\alpha}\log_2^\beta N\right) \\
& = & c\log_2 N\log_2 k
\end{array}$$

Inequality (16) now follows from the fact that $c\log_2 k = 2$. Inequality (17) follows from the fact that $t = c\log_2 N$. ▮

The following result was derived as a corollary by Blum [Bl] and shows the application of the above theorem to coloring graphs with low-chromatic number with relatively small number of colors. We warn the reader that the corollary does not follow directly from the above theorem; this is because it uses a Levin-reduction[9] from the search version of chromatic number to the search version of the clique problem. However, it is possible to define search versions of all the gap problems above appropriately and verify that all the reductions work for the search problems as well (i.e., they are in fact Levin-reductions). Thus the following can be derived as a corollary to the above.

Corollary 8.15 Let $k < \infty$.

(1) For $\epsilon > 0$, given an $N^{1-\epsilon}$ approximator to the clique, one can color any k-colorable graph on M nodes with $O(k^{1/\epsilon}\log M)$ colors in polynomial time.

(2) For $\epsilon(N) = \omega((\log N)^{-1/2})$, given an $N^{1-\epsilon(N)}$ approximator to the clique, one can color any k-colorable graph on M nodes with $M^{o(1)}$-colors in time $M^{O(\log M)}$.

[9]A Levin-reduction is a polynomial-time many-to-one reduction which is augmented by corresponding polynomial-time witness transformations.

9 On the Limitations of Some Common Approaches

In this section we provide lower bounds on the free-bit complexity of two tasks which are central to all existing ("low-complexity") probabilistically checkable proofs. Specifically, we consider the task of checking that a string (given by oracle access) is "close" to a valid codeword and the task of checking that one oracle is an encoding of a projection of a string encoded by a second oracle. Here a string is considered *close* to the code if its distance from some codeword is less than half the distance of the code. Loosely speaking, we show that each of these tasks has amortized free-bit complexity of at least *one* (and this is tight by the codes and tests presented in Section 7). Furthermore, we show that the amortized free-bit complexity of performing both tasks (with respect to the same given oracles) is at least *two* (and also this is tight by Section 7).

Our original motivation in proving these lower bounds was to indicate that a paradigm shift is required in order to improve over our PCP systems of amortized free-bit complexity 2 (for NP). In retrospect, the paradigm shifts have amounted to the relaxation of the codeword test in [H1] and to the relaxation of the projection test in [H2]. Thus, our lower bounds may be considered as a justification for these (somewhat unnatural) relaxations.

In particular, the lower bound on the complexity of the codeword test relies on the particular interpretation of 'closeness' used above (i.e., being at distance less than *half* the distance of the code). This requirement is not essential as can be seen in Section 3.4, where we show that relaxed codeword tests, in which closeness means approximately the distance of the code, also suffice. Håstad's relaxation of the codeword test is different, yet it also suffices for the purpose of constructing PCP systems of amortized free-bit complexity 1 (for NP) [H1]. The lower bound on the complexity of the projection test seems more robust. Yet, as shown by Håstad in [H2], the projection requirements can be by-passed as well, yielding pcp systems of amortized free-bit complexity tending to 0.

9.1 The tasks

Our definitions of the various tasks/tests are quite minimal and do not necessarily suffice for PCP applications. However, as we are proving lower bounds this only makes our results stronger.

Loosely speaking, the first task consists of testing that an oracle encodes a valid codeword, or is "close" to a valid codeword, with respect to an error-correcting code of non-trivial distance (i.e., distance greater than 1). The condition regarding the distance of the code is essential since the task is easy with respect to the identity map (which is a code of distance 1). We remark that testing "closeness" to codewords with respect to codes of large distance is essential in all known pcp constructions [BFLS, FGLSS, ArSa, ALMSS, BGLR, FeKi1, BeSu].

The *absolute distance* between two words $w, u \in \{0,1\}^n$, denoted $\Delta(w,u)$, is the number of bits on which w and u disagree. We say that the code $E : \{0,1\}^* \mapsto \{0,1\}^*$ has *absolute distance* d if for every m and every $x \neq y \in \{0,1\}^m$ the absolute distance between $E(x)$ and $E(y)$ is at least $d(m)$. The absolute distance between a word w and a code E, denoted $\Delta_E(w)$, is defined as the minimum absolute distance between w and a codeword of E.

Definition 9.1 (Codeword test): Let $E : \{0,1\}^m \to \{0,1\}^n$ be a code of absolute distance $d > 1$. A *codeword test* (with respect to E) is an oracle machine, T, such that $T^{E(a)}(R)$ accepts for all a, R. The error probability of T is defined as the maximum accepting probability of T over oracles A of absolute distance at least $\lfloor d/2 \rfloor$ from the code E; namely,

$$\max_{A \in \{0,1\}^n \text{ s.t. } \Delta_E(A) \geq \lfloor d/2 \rfloor} \left\{ \Pr_R \left[T^A(R) \text{ accepts} \right] \right\}$$

(Nothing is required with respect to non-codewords which are "close" to the code.)

The second task is defined with respect to a "projection function" π and a pair of codes, E_1 and E_2. Loosely speaking, the task consists of checking if the string E_1-encoded by the first oracle is mapped by π to the string that is E_2-encoded by the second oracle.

Definition 9.2 (projection test): Let E_1: $\{0,1\}^m \to \{0,1\}^n$ and E_2: $\{0,1\}^k \to \{0,1\}^{n'}$ be two codes and let $\pi : \{0,1\}^m \to \{0,1\}^k$ be a function. A *projection test* (with respect to the above) is a two-oracle machine, T, such that $T^{E_1(a),E_2(\pi(a))}(R)$ accepts for all a, R. The error probability of T is defined as the maximum accepting probability of T over oracles pairs $(E_1(a), E_2(b))$ where $b \neq \pi(a)$; namely,

$$\max_{a,b \text{ s.t. } \pi(a)\neq b} \left\{ \Pr_R \left[T^{E_1(a),E_2(b)}(R) \text{ accepts} \right] \right\}$$

(Nothing is required with respect to non-codewords.)

Finally, we consider a test T which combines the two tests above; namely, T takes two oracles A and B and performs a codeword test on A and a projection test on the pair (A, B).

Definition 9.3 (combined test): Let E_1: $\{0,1\}^m \to \{0,1\}^n$ be a code of absolute distance $d > 1$ and E_2: $\{0,1\}^k \to \{0,1\}^{n'}$ be two codes and let $\pi : \{0,1\}^m \to \{0,1\}^k$ be a function. A *combined test* for (E_1, E_2, π) is a two-oracle machine T such that $T^{E_1(a),E_2(\pi(a))}(R)$ accepts on all a, R. The error probability of T is defined as the maximum accepting probability of T over oracles pairs (A, B) where either $\Delta_{E_1}(A) \geq \lfloor d/2 \rfloor$ or $A = E_1(a)$, $B = E_2(b)$ but $\pi(a) \neq b$; namely,

$$\max_{(A,B)\in S} \left\{ \Pr_R \left[T^{A,B}(R) \text{ accepts} \right] \right\}.$$

where $S \stackrel{\text{def}}{=} \{(A, B) : (\Delta_{E_1}(A) \geq \lfloor d/2 \rfloor) \text{ or } (\exists a, b \text{ s.t. } A = E_1(a) \text{ and } B = E_2(b) \text{ and } \pi(a) \neq b)\}$.

(Nothing is required with respect to non-codeword pairs, (A, B), which are "close" to some pair $(E_1(a), E_2(b))$ with $\pi(a) \neq b$.)

Conventions and Notations

The *pattern* of test T on access to oracle A (resp., oracles A and B) when using coin-sequence R consists of (R and) the sequence of queries and answers made by T. Namely, this pattern, denoted $\mathsf{pattern}_T(A; R)$ (resp., $\mathsf{pattern}_T(A, B; R)$), is defined as the sequence $(R, q_1, a_1, ..., q_t, a_t)$ where q_i is the i^{th} query made by T on coin-sequence R and after receiving the answers $a_1, ..., a_{i-1}$. We include the queries in the pattern for sake of clarity (but they can be easily reconstructed from the coin-sequence and the answers). In case T uses two oracles, we may assume that the queries specify to which oracle they are addressed. For simplicity, we assume in the rest of this subsection that the test has access to one oracle, denoted A.

The set $\mathsf{Acc}_T(R)$ is defined to be the set of accepting patterns of T on coin-sequence R. Clearly,

$$\mathsf{Acc}_T(R) = \{\mathsf{pattern}_T(A; R) : T^A(R) \text{ accepts}\}$$

Recall that T is said to have free-bit complexity f if for each possible coin-sequence R it holds that $|\mathsf{Acc}_T(R)| \leq 2^f$. We say that T has *average free-bit complexity* f_{av} if $\mathbf{E}_R [|\mathsf{Acc}_T(R)|] \leq 2^{f_{\text{av}}}$, when the expectation is taken uniformly over all possible coin-sequences. The amortized free-bit complexity of a test is defined as $\frac{f_{\text{av}}}{\log_2(1/\epsilon)}$, where f_{av} is the average free-bit complexity of the test and ϵ is its error probability.

9.2 Lower Bound for the Codeword Test

Proposition 9.4 For any code of absolute distance greater than 1, the Codeword Test has amortized free-bit complexity of at least $1 - o(1)$.

The amortization in the above proposition is to be understood as taking place on a fixed number of free-bits whereas the length of the oracle grows. Actually, we can allow both the oracle-length and the free-bit count to grow, provided that the logarithm of the number of codewords grows faster than the free-bit complexity. Alternatively, we can consider a fixed oracle length and a fix bound on the number of free-bits. Actually, this is done in the following technical lemma from which the above proposition follows.

Lemma 9.5 Let $E : \{0,1\}^m \mapsto \{0,1\}^n$ be a code of absolute distance $d > 1$, and let T be a codeword test with respect to E having average free-bit complexity f_{av}. Then, T has error probability at least $\max(2 - 2^{f_{\text{av}}}, \frac{1}{F} - \frac{1}{M})$, where $F = 2^{f_{\text{av}}}$ and $M = 2^m$.

In particular, if $f_{\text{av}} = 0$ then the error is 1, and for $f_{\text{av}} \geq 1$ the error is at least $\frac{1}{F} - \frac{1}{M}$.

Proof: Fix an arbitrary coin-sequence R, and let F_R denote the cardinality of the set $\mathsf{Acc}_T(R)$.

Let a_1, a_2 be selected independently and uniformly in $\{0,1\}^m$, and consider the codewords $E(a_1)$ and $E(a_2)$. With probability $\frac{1}{M}$ we have $a_1 = a_2$ and otherwise $\Delta(E(a_1), E(a_2)) \geq d$. From a_1 and a_2, we construct an oracle $A(a_1, a_2)$ as follows: If $a_1 = a_2$, then $A = E(a_1)$. Otherwise, we construct $A(a_1, a_2)$ so that it agrees with the value of the bits of both $E(a_i)$'s whenever they are the same and is at distance $\lceil d/2 \rceil$ from $E(a_1)$. This can be done as follows: let S be the set of positions on which $E(a_1)$ and $E_2(a_2)$ disagree and let S' be a subset of S of cardinality $\lceil d/2 \rceil$. Then $A(a_1, a_2)$ equals $E(a_1)$ on all positions not in S' (and equals $E(a_2)$ on the positions in S').

We claim that, when $a_1 \neq a_2$, the oracle $A \stackrel{\text{def}}{=} A(a_1, a_2)$ is at distance at least $\lfloor d/2 \rfloor$ from the code (i.e., $\Delta_E(A) \geq \lfloor d/2 \rfloor$). This can be proved as follows: Consider any $a \in \{0,1\}^m$ and observe that by the triangle inequality

$$\Delta(A, E(a)) \geq \Delta(E(a_1), E(a)) - \Delta(E(a_1), A) \geq d - \lceil d/2 \rceil = \lfloor d/2 \rfloor$$

We now claim that

$$\Pr_{a_1,a_2}\left[T^{A(a_1,a_2)}(R) \text{ accepts}\right] \geq \frac{1}{F_R}$$

where the probability is taken uniformly over all possible choices of $a_1, a_2 \in \{0,1\}^m$. The key observation is that if $\mathsf{pattern}_T(E(a_1); R)$ equals $\mathsf{pattern}_T(E(a_2); R)$, then $\mathsf{pattern}_T(A(a_1, a_2); R)$ will be equal to $\mathsf{pattern}_T(E(a_1); R)$ (since no query of $T(R)$ falls in the set S – defined above). Thus, since $T^{E(a_1)}(R)$ accepts, $T^{A(a_1,a_2)}(R)$ must accept too. This suggests to lower bound the probability that $T^{A(a_1,a_2)}(R)$ accepts by the probability that $\mathsf{pattern}_T(E(a_1); R) = \mathsf{pattern}_T(E(a_2); R)$. Consider an enumeration, $\alpha_1, ..., \alpha_{F_R}$, of the patterns in $\mathsf{Acc}_T(R)$ and denote by p_i the probability that $\mathsf{pattern}_T(E(a); R)$ equals the i^{th} pattern in this enumeration, when a is uniformly selected in $\{0,1\}^m$ (i.e., $p_i \stackrel{\text{def}}{=} \Pr_a[\mathsf{pattern}_T(E(a); R) = \alpha_i]$). Thus, when a_1 and a_2 are picked at random, the probability that $\mathsf{pattern}_T(E(a_1); R) = \mathsf{pattern}_T(E(a_2); R)$ is $\sum_{i=1}^{F_R} p_i^2$. Subject to the condition $\sum_i p_i = 1$, the quantity $\sum_{i=1}^{F_R} p_i^2$ is lower bounded by $\frac{1}{F_R}$ (with an equality occurring when the p_i's are equal).

The following observations now bound the error of T:

$$\begin{aligned}
\Pr_{a_1,a_2}\left[T^{A(a_1,a_2)}(R) \text{ accepts and } a_1 \neq a_2\right] &\geq \Pr_{a_1,a_2}\left[T^{A(a_1,a_2)}(R) \text{ accepts}\right] - \Pr_{a_1,a_2}[a_1 = a_2] \\
&\geq \frac{1}{F_R} - \frac{1}{M}
\end{aligned}$$

All the above holds for any coin-sequence R. Now, we let R be uniformly chosen and get

$$\begin{aligned}\Pr_{R,a_1,a_2}\left[T^{A(a_1,a_2)}(R) \text{ accepts and } a_1 \neq a_2\right] &\geq \mathbf{E}_R\left[\frac{1}{F_R}\right] - \frac{1}{M} \\ &\geq \frac{1}{F} - \frac{1}{M}\end{aligned}$$

(The last inequality follows by Jensen's inequality.) Thus there must exist oracles a_1 and a_2 with $a_1 \neq a_2$ such that

$$\Pr_R\left[T^{A(a_1,a_2)}(R) \text{ accepts}\right] \geq \frac{1}{F} - \frac{1}{M}$$

But the oracle $A(a_1, a_2)$ above satisfies $\Delta_E(A(a_1, a_2)) \geq \lfloor d/2 \rfloor$ implying that the error of T is at least $\frac{1}{F} - \frac{1}{M}$.

To prove that the error is at least $2 - 2^{f_{\mathrm{av}}}$, we observe that if $F_R = 1$ for some coin-sequence R then $\mathsf{pattern}_T(E(a_1); R) = \mathsf{pattern}_T(E(a_2); R)$, for every two $a_1, a_2 \in \{0,1\}^m$. It follows that, for every $a_1 \neq a_2$, given access to the oracle $A(a_1, a_2)$ and using coin-sequence R, the test T accepts (and is wrong in doing so). Thus, for every $a_1 \neq a_2$,

$$\Pr_R\left[T^{A(a_1,a_2)}(R) \text{ accepts}\right] \;\geq\; \Pr_R\left[F_R = 1\right] \;=\; 1 - \Pr_R\left[F_R > 1\right]$$

and the error bound follows by using $\Pr_R\left[F_R - 1 > 0\right] \leq \mathbf{E}_R\left[F_R - 1\right] = F - 1$. ▌

Proof of Proposition 9.4: Let T be a test for the code $E : \{0,1\}^* \to \{0,1\}^*$ so that E maps m-bit strings into $n(m)$-bit strings. Suppose that T has average free-bit complexity $f(m)$ and error $\epsilon(m)$, as a function of m (the length of strings encoded by the oracle). We first assume that $f(m) \geq 1$. Using Lemma 9.5 (and letting $\rho(m) \stackrel{\text{def}}{=} 2^{f(m)-m}$), we lower bound the amortized free-bit complexity of T as follows

$$\begin{aligned}\frac{f(m)}{\log_2(1/\epsilon(m))} &\geq \frac{f(m)}{-\log_2(\frac{1}{2^{f(m)}} - \frac{1}{2^m})} \\ &= \frac{f(m)}{f(m) - \log_2(1-\rho(m))} \\ &> \frac{f(m)}{f(m) + \rho(m)} \\ &> 1 - \rho(m)\end{aligned}$$

(For the last inequality, we have assumed $f(m) \geq 1$.) Thus, for this case, the proposition follows by our convention that the number of codewords (denoted 2^m) grows faster than exponential in the free-bit complexity $f(m)$ (i.e., $\rho(m) = \frac{2^{f(m)}}{2^m} \to 0$ with $n \to \infty$). Finally, we need to address the case in which $f(m) \geq 1$ does not hold. We consider two sub-cases. In the first sub-case, we assume that $f(m) \to 0$ for some subsequence of the m's. For these m's, we use Lemma 9.4's assertion that $\epsilon(m) \geq 2 - 2^{f(m)}$. Setting $g(m) \stackrel{\text{def}}{=} 2^{f(m)} - 1$, we lower bound the amortized free-bit complexity by

$$\begin{aligned}\frac{f(m)}{\log_2(1/\epsilon(m))} &\geq \frac{\log_2(1+g(m))}{-\log_2(1-g(m))} \\ &\to \frac{g(m)}{g(m)}\end{aligned}$$

For the other sub-case, we have $f(m) \geq t$, for some constant $t > 0$. Applying T for t times we get a test T' with average free-bit complexity $t \cdot f(m) \geq 1$ and error $\epsilon'(m) = \epsilon(m)^t$, which maintains the amortized free-bit complexity of T (since $\frac{f(m)}{-\log_2 \epsilon(m)} = \frac{t \cdot f(m)}{-\log_2 \epsilon'(m)}$). Applying the above analysis to T', the proposition follows. ▌

9.3 Lower Bound for the Projection Test

A *projection function* is a function $\pi : \{0,1\}^* \mapsto \{0,1\}^*$ having the property that for every m there exists a k so that π maps $\{0,1\}^m$ onto $\{0,1\}^k$.

Proposition 9.6 For any pair of codes used in the two oracles and any projection function, the Projection Test has amortized free-bit complexity of at least $1 - o(1)$.

Again, the proposition is proved by a technical lemma. Actually, the lemma refers to any function $\pi : \{0,1\}^m \mapsto \{0,1\}^k$ and its conclusion depends on the cardinality of the range of π (which in case of a projection function equals 2^k). Abusing notations we let $\pi(S) \stackrel{\text{def}}{=} \{\pi(a) : a \in S\}$.

Lemma 9.7 Let $E_1 : \{0,1\}^m \mapsto \{0,1\}^n$, $E_2 : \{0,1\}^k \mapsto \{0,1\}^{n'}$ and $\pi : \{0,1\}^m \mapsto \{0,1\}^k$ be as in Definition 9.2, and T be a projection test with respect to them having average free-bit complexity f_{av}. Then, T has error probability at least $\frac{1}{F} - \frac{1}{K}$, where $K = |\pi(\{0,1\}^m)|$ and $F = 2^{f_{\text{av}}}$. Furthermore, if $K > 1$ then T has error probability at least $2 - 2^{f_{\text{av}}}$.

Proof: Fixing an arbitrary coin-sequence R, let $F_R \stackrel{\text{def}}{=} |\{\mathsf{Acc}_T(R)\}|$. We consider the behavior of the test T when given oracle access to a pair of randomly and independently selected codewords. Specifically, let $S \subset \{0,1\}^m$ be a set of K strings such that for every $b \in \pi(\{0,1\}^m)$ there exists an $a \in S$ satisfying $\pi(a) = b$. We consider the behavior of T when given access to the oracles $E_1(a)$ and $E_2(\pi(a'))$, where a and a' are independently and uniformly selected in S. With probability $\frac{1}{K}$, we have $\pi(a) = \pi(a')$. On the other hand we claim that, given access to such pair of random oracles, T accepts with probability at least $\frac{1}{F_R}$. Once the claim is proven, the lemma follows (as in the proof of the previous lemma).

Consider the set of all F_R possible accepting patterns of T on access to oracles, $E_1(a)$ and $E_2(\pi(a))$, where $a \in S$. Each such pattern consists of a pair (α, β), where α (resp., β) denotes the transcript of the test's interaction with $E_1(a)$ (resp., $E_2(\pi(a))$). Enumerating all possible F_R patterns, we denote by p_i the probability that the i^{th} pattern occurs, when T is given access to the oracle-pair $(E_1(a), E_2(\pi(a))$ where a is uniformly selected in S. Namely,

$$p_i \stackrel{\text{def}}{=} \Pr_{a \in S}\left[\mathsf{pattern}_T(E_1(a), E_2(\pi(a)); R) = (\alpha_i, \beta_i)\right]$$

where (α_i, β_i) is the i^{th} accepting pattern for $T(R)$. Clearly,

$$\Pr_{a,a' \in S}\left[\mathsf{pattern}_T(E_1(a), E_2(\pi(a)); R) = \mathsf{pattern}_T(E_1(a'), E_2(\pi(a')); R) = (\alpha_i, \beta_i)\right] = p_i^2 \tag{18}$$

We now claim that the probability that a pair of independently chosen random oracles (i.e., $(E_1(a), E_2(b))$ selected by uniformly selecting $a, a' \in S$ and setting $b = \pi(a')$) leads to the i^{th} pattern is at least p_i^2; namely,

$$\Pr_{a,a' \in S}\left[\mathsf{pattern}_T(E_1(a), E_2(\pi(a')); R) = (\alpha_i, \beta_i)\right] \geq p_i^2 \tag{19}$$

Eq. (19) is proven by a cut-and-paste argument: Suppose $\mathsf{p} \stackrel{\text{def}}{=} \mathsf{pattern}_T(E_1(a), E_2(\pi(a)); R)$ equals $\mathsf{p}' \stackrel{\text{def}}{=} \mathsf{pattern}_T(E_1(a'), E_2(\pi(a')); R)$ and consider a computation of $T^{E_1(a), E_2(\pi(a'))}(R)$. Proceeding by induction, and assuming that the first t queries are answered as in p, we conclude that the $t+1^{\text{st}}$ query (in our "mixed" computation) is identical to the $t+1^{\text{st}}$ query in $\mathsf{p} = \mathsf{p}'$. If this query is directed to the fist oracle then it is answered by $E_1(a)$ (as in p) and otherwise it is answered by $E_2(\pi(a'))$ (as in p'). In both cases the answer matches the $t+1^{\text{st}}$ answer in $\mathsf{p} = \mathsf{p}'$. We conclude that whenever $\mathsf{p} = \mathsf{p}'$, the computation of $T^{E_1(a),E_2(\pi(a'))}(R)$ encounters the same pattern (p). Thus, the probability that the computation of $T^{E_1(a),E_2(\pi(a'))}(R)$ encounters the i^{th} pattern is lower bounded by the expression in Eq. (18), and Eq. (19) follows. (We remark that for non-adaptive tests, the probability that the i^{th} pattern is encountered equals $\sum_{i=1}^{F_R} p_i' p_i''$, where p_i' (resp., p_i'') is the sum of all p_j's satisfying $\alpha_j = \alpha_i$ (resp., $\beta_j = \beta_i$). Actually, the same holds for any test which selects its queries for each oracle independently of answers obtained from the other oracle.)

Using Eq. (19), we get

$$\begin{aligned}
\Pr_{a,a' \in S}\left[\mathsf{pattern}_T(E_1(a), E_2(\pi(a')); R) \in \mathsf{Acc}_T(R)\right] \;&\geq\; \sum_{i=1}^{F_R} p_i^2 \\
&\geq\; \frac{1}{F_R}
\end{aligned}$$

and the main part of the lemma follows. Again, the furthermore part follows by observing for $F_R = 1$, $\mathsf{pattern}_T(E_1(a), E_2(\pi(a)); R) = \mathsf{pattern}_T(E_1(a'), E_2(\pi(a')); R)$, for every two $a, a' \in \{0,1\}^m$. Again, this implies that, for every $a_1 \neq a_2$, given access to the oracle-pair $(E_1(a), E_2(\pi(a')))$ and using coin-sequence R, the test T (wrongly) accepts. ▌

9.4 Lower Bound for the Combined Test

Proposition 9.8 For any pair of codes used in the two oracles, so that the first code has absolute distance greater than 1, and for any projection function, the Combined Test has amortized free-bit complexity of at least $2 - o(1)$.

Again, the proposition is proved by a technical lemma. Loosely speaking, the lemma asserts that a combined test of free-bit complexity $2f$ must have error probability at least $\frac{1}{8} \cdot 2^{-f}$. The lower bound extends to the case where $2f$ is a bound on the average free-bit complexity; the error probability in this case can be lower bounded by $\frac{3}{64} \cdot 2^{-f}$ – see details below. It follows that the amortized free-bit complexity of such a test must be at least $\frac{2f}{f+5} \approx 2$ (for large f's). The restriction to large f's does not really weaken the result. Suppose on the contrary that there exists a test with amortized free-bit complexity f_{am}. Then, for any sufficient large t, we can obtain a test with free-bit complexity $2f \stackrel{\text{def}}{=} t \cdot f_{\text{am}}$ and error 2^{-t}. By the above $\frac{t \cdot f_{\text{am}}}{t} \geq \frac{2f}{f+5} \approx 2$ (as f is now large).

Lemma 9.9 Let $E_1 : \{0,1\}^m \mapsto \{0,1\}^n$ be a code of absolute distance greater than 1, $E_2 : \{0,1\}^k \mapsto \{0,1\}^{n'}$, and $\pi : \{0,1\}^m \mapsto \{0,1\}^k$ be a projection function. Suppose that T is a combined codeword and projection test with respect to the above having free-bit complexity $2f$. Then, T has error probability at least $\frac{1}{8F} - \frac{1}{2K} - \frac{1}{4M}$, where $K = 2^k$, $F = 2^f$, and M is the minimum, over all $b \in \{0,1\}^k$, of the number of $a \in \{0,1\}^m$ projected by π to b (i.e., $M \stackrel{\text{def}}{=} \min_{b \in \{0,1\}^k}\{|\{a : \pi(a) = b\}|\}$). Furthermore, if $2f < 1$ and $\max\{M, K\} > 1$ then T has error probability 1.

Proof: The "furthermore" part follows immediately by any of the furthermore parts of Lemma 9.5 or Lemma 9.7 (as 2^{2f} must be an integer and so $2f < 1$ implies $f = 0$). The proof of the main part of the lemma uses both strategies employed in the proofs of Lemmas 9.5 and 9.7. We consider two cases. The first case is that for some $E_2(b)$, half of the possible (coin-sequences) R's have at most F accepting patterns with respect to the coin-sequence R and second oracle $B = E_2(b)$. In this case we employ the strategy used in the proof of Lemma 9.5, restricted to oracles constructed by combining two uniformly selected codewords $E_1(a_i)$'s satisfying $\pi(a_i) = b$. The second case is that for every $b \in \{0,1\}^k$, for half of the possible (coin-sequences) R's, the number of accepting patterns with respect to the coin-sequence R and second oracle $B = E_2(b)$ is at least F. In this case we show that many possible B's must fit into fewer than $\frac{F^2}{F}$ accepting patterns and we may employ the strategy used in the proof of Lemma 9.7. Details follow.

In the sequel $\delta \in [0,1]$ is a constant to be determined later. (In the above motivating discussion we have used $\delta = \frac{1}{2}$ but a better bound follows by letting δ be larger.)

Case 1: there exists $b \in \{0,1\}^k$ so that for at least $(1-\delta)$ fraction of the possible (coin-sequences) R's, hereafter called *good*, the number of accepting patterns with respect to the coin-sequence R and second oracle (fixed to) $B = E_2(b)$ is at most F.

Fixing this b, we consider M possible a's satisfying $\pi(a) = b$. Employing the argument of Lemma 9.5, we get that for each of these *good* R's, a random oracle A (constructed using two uniformly chosen a's as above) is wrongly accepted with probability at least $\frac{1}{F} - \frac{1}{M}$. By an averaging argument, it follows that there exists a pair of oracles (A, B) on which T errs with probability at least

$$(1-\delta) \cdot \left(\frac{1}{F} - \frac{1}{M}\right) \tag{20}$$

Case 2: for every $b \in \{0,1\}^k$, for at least a δ fraction of the possible (coin-sequences) R's, the number of accepting patterns with respect to the coin-sequence R and second oracle $B = E_2(b)$ is at least F.

Let $\gamma < \delta$ be a parameter to be determined later. By a counting argument, for at least a $\frac{\delta-\gamma}{1-\gamma}$ fraction of the possible R's, hereafter called *good*, there exists a set, denoted Π_R, of at least $\gamma \cdot 2^k$ possible $b \in \{0,1\}^k$ so that there are at least F accepting patterns which are consistent with coin-sequence R and second oracle fixed to $B = E_2(b)$. (Namely, let g denote the fraction of good R's. Then $g + (1-g)\cdot\gamma \geq \delta$ and $g \geq \frac{\delta-\gamma}{1-\gamma}$ follows.)

Let $S \subset \{0,1\}^m$ be a set of 2^k strings, defined as in the proof of Lemma 9.7, so that π maps S onto $\{0,1\}^k$. Fixing a good coin-sequence R, we adapt the strategy used in the proof of Lemma 9.7 as follows. We consider a set $S_R \subseteq S$ of $|\Pi_R|$ strings so that π maps S_R onto Π_R, and enumerate the accepting patterns which occur when the test, using coins R, is given access to a oracle-pair $(E_1(a), E_2(\pi(a)))$, where a is uniformly chosen in S_R. We first claim that there are at most F such patterns. Namely,

Claim: For any good R, $|\{\mathsf{pattern}_T(E_1(a), E_2(\pi(a)); R) : a \in S_R\}| \leq F$.

Proof: By definition of Π_R, for each $b \in \Pi_R$, there are at least F accepting patterns consistent with the coin-sequence R and the second oracle $E_2(b)$ (and out of them only one fits the first oracle $E_1(a)$ where $a \in S_R$ and $\pi(a) = b$). By a cut-and-paste argument, if (R, α, β) and (R, α', β) are accepting patterns for second-oracle $E_2(b)$ and if (R, α, β) is an accepting pattern for second-oracle $E_2(b')$ then (R, α', β) is also an accepting pattern for second-oracle $E_2(b')$. It follows that the accepting patterns of two $E_2(b)$'s either collide or do not intersect. Thus, the number of accepting patterns for the various $(E_1(a), E_2(\pi(a)))$'s, where $a \in S_R$, is at most $\frac{F^2}{F} = F$ and the claim follows. $\Box$

Now we consider what happens if one selects independently and uniformly $a, a' \in S$. Following the proof of Lemma 9.7, with probability $\frac{1}{K}$, we have $\pi(a) = \pi(a')$ (and otherwise $\pi(a) \neq \pi(a')$). On the other hand, given access to such pair of random oracles, the test accepts with probability at least $\gamma^2 \cdot \frac{1}{F}$. (The γ^2 factor is due to the probability that $a, a' \in S_R$, whereas the $\frac{1}{F}$ factor corresponds to the analysis which supposes that a and a' are uniformly selected in S_R).

The above analysis holds for any good coin-sequence R. Using the lower bound on the fraction of good R's, it follows that for a $\frac{\delta-\gamma}{1-\gamma}$ fraction of the R's, the probability that the test errs, on coin-sequence R when given access to a random pair of oracles (selected as above), is at least $\frac{\gamma^2}{F} - \frac{1}{K}$. By an averaging argument, there exists a pair of oracles for which the test errs with probability

$$\frac{\delta-\gamma}{1-\gamma} \cdot \left(\frac{\gamma^2}{F} - \frac{1}{K}\right) \tag{21}$$

Setting $\delta = \frac{3}{4}$ and $\gamma = \frac{1}{2}$ we lower bound the expressions in Eq. (20) and (21) by $\frac{1}{4F} - \frac{1}{4M}$ and $\frac{1}{8F} - \frac{1}{2K}$, respectively, and the lemma follows. ▌

To prove a bound for the case of average free-bit complexity $2f$, we first apply Markov's Inequality and conclude that all but an ϵ fraction of the coin-sequences have at most $G^2 \stackrel{\text{def}}{=} \frac{F^2}{\epsilon}$ accepting patterns (in which this fixed coin-sequence appears). (We can use any $0<\epsilon<1$.) We then consider only those coin sequences (and apply the same argument as above to each of them). The averaging argument at the end of the above proof then yields that there exists an oracle-pair on which T errs on at least a $\frac{1}{8G} - \frac{1}{2K} - \frac{1}{4M}$ fraction of these coin-sequences. It follows that this oracle makes T err with probability at least $(1-\epsilon) \cdot (\frac{1}{8G} - \frac{1}{2K} - \frac{1}{4M})$ (which equals $(1-\epsilon) \cdot (\frac{\sqrt{\epsilon}}{8F} - \frac{1}{2K} - \frac{1}{4M})$). Using $\epsilon = \frac{1}{4}$, we get a lower bound of $\frac{3}{64F} - \frac{3}{8K} - \frac{3}{16M}$.

Part III
PCP: Properties and Transformations

10 The Complexity of PCP and FPCP

In this section we present several results regarding the complexity of languages acceptable by probabilistically checkable proofs having, respectively, small query complexity, small amortized-query complexity and small free-bit complexity. Thus, in the current section, notations such as $\mathrm{PCP}_{c,s}[r,q]$ stand for classes of languages. The results can be extended to classes of promise problems having such probabilistically checkable proofs.

In this section, $\mathrm{MIP}_{c,s}[r,p]$ denotes the class of languages accepted by a (one-round) p-prover interactive proof system in which r is the randomness complexity, c is a lower bound on the probability of accepting yes-instances and s is an upper bound on the probability of accepting no-instances. The corresponding class for probabilistically checkable proofs is $\mathrm{PCP}_{c,s}[r,q]$, where q denotes the number of queries. In both classes only binary queries are allowed (indeed this is less standard for MIP).

10.1 MIP versus PCP

The first part of the following lemma is folklore and is stated here for sake of completeness.

Lemma 10.1 For all admissible functions c, s, r, p.
(1) $\mathrm{MIP}_{c,s}[r,p] \subseteq \mathrm{PCP}_{c,s}[r,p]$.
(2) $\mathrm{MIP}_{c,s}[r,p] \subseteq \mathrm{MIP}_{c,2s}[r,p-1]$.

Proof: Part (1) follows from the definition of PCP and MIP. Part (2) is shown as follows. Let V be an (r,p)-restricted MIP verifier. We define V' – an $(r,p-1)$-restricted verifier who on input x behaves as follows:

- V' tosses coins $\bar{c}$ for V.
- V' refers the first $p-1$ queries of V to the corresponding $p-1$ provers obtaining answers (bits) $a_1,\ldots,a_{p-1}$, respectively.
- V' accepts if and only if there exists $a_p \in \{0,1\}$ such that V would accept answers $a_1,\ldots,a_p$ on input x and random string $\bar{c}$.

Suppose that provers $P_1,\ldots,P_p$ convince V to accept x with probability δ. Then, the provers $P_1,\ldots,P_{p-1}$ convince V' to accept x with probability at least δ (because if $V(x)$ accepts the transcript $(\bar{c},a_1,...,a_p)$ then $V'(x)$ will accept the transcript $(\bar{c},a_1,...,a_{p-1})$). This justifies the bound on the completeness probability of V'. Suppose, on the other hand, that provers $P_1,\ldots,P_{p-1}$ cause V' to accept x with probability δ. Consider a uniformly selected strategy for another prover, denoted P_p (i.e., choose a random response for every question). Then, the probability that provers $P_1,\ldots,P_p$ cause V to accept input x is at least $\frac{1}{2}\cdot\delta$ (because if $V'(x)$ accepts the transcript $(\bar{c},a_1,...,a_{p-1})$, then there exists a value $a_p \in \{0,1\}$ so that $V(x)$ will accept the transcript $(\bar{c},a_1,...,a_p)$, and with probability one half P_p's answer equals this a_p). This justifies the bound on the soundness probability of V'. ▮

Containments of PCP systems in MIP systems are more problematic. The reader is referred to a paper by Ta-Shma [Ta-S]. That paper also contains a proof of the following result due to Bellare, Goldreich and Safra:

$$\mathrm{PCP}_{c,s}(\log,q) \subseteq \mathrm{MIP}_{c,q^q\cdot s}(\log,q)$$

Here we only consider the non-adaptive case, and obtain a different bound on the soundness parameter:

Proposition 10.2 Suppose $L \in \mathrm{PCP}_{c,s}(r,q)$ with a non-adaptive verifier. Then $L \in \mathrm{MIP}_{c',s'}(r + O(\log q), q)$, where $c' = c + p \cdot (1-c)$, $s' = s + p \cdot (1-s)$ for any $p \geq \lfloor q/2 \rfloor / (1 + \lfloor q/2 \rfloor)$.

For $q = 3$, we may set $p = 0.5$ and obtain $c' = (c+1)/2$ and $s' = (s+1)/2$.

Proof: We start with a non-adaptive PCP verifier of q queries and construct a q-prover system as follows. First we uniformly select coin tosses for the PCP verifier, which defines q queries (here is where we use non-adaptivity). Next,

- With probability p we select a query uniformly among these q queries, and forward it to all q provers. We accept iff all provers answer in the same manner.
- With probability $1-p$ we simulate the PCP system as follows. We uniformly select $i \in [q]$ and refer the j^{th} query of the verifier to the $(i+j)^{\text{th}}$ prover. We accept iff the PCP verifier would have accepted

Clearly, by setting all MIP-provers to equal the good oracle (of the PCP system), inputs in the language are accepted with probability at least $p \cdot 1 + (1-p) \cdot c = c + p \cdot (1-c)$.

We now bound the acceptance probability of the MIP system for an input not in the language. Fix an arbitrary sequence of MIP-provers. Let δ denote the probability, taken over the queries selected by the PCP-verifier as above, that the MIP-provers differ on a random query. Define an oracle so that on each query it equals the majority of the prover's answers (ties, in case of even q are broken arbitrarily). Then, the probability that the MIP system accepts is bounded above by

$$p \cdot (1-\delta) + (1-p) \cdot (s + \lfloor q/2 \rfloor \cdot \delta) \tag{22}$$

To justify the second term consider the simulation of the PCP system (which takes place with probability $1-p$). In case the answers given by all MIP-provers equal the corresponding answers of the PCP-oracle (defined above), we bound the acceptance probability by soundness of the PCP system. Otherwise, there must be a query on which the relevant MIP-prover differs from the PCP-oracle. For each query this happens with probability at most $\frac{\lfloor q/2 \rfloor}{q}$ (as, by definition, only a minority of provers differ from the oracle). Using the Union Bound, Eq. (22) follows. Using the definition of p, we have

$$\begin{aligned} p \cdot (1-\delta) + (1-p) \cdot (s + \lfloor q/2 \rfloor \cdot \delta) &= p + (1-p) \cdot s - \delta \cdot (p - (1-p) \cdot \lfloor q/2 \rfloor) \\ &\leq s + p \cdot (1-s) \end{aligned}$$

and the proposition follows. ▮

10.2 Query complexity and amortized query complexity

The following proposition explores the limitations of probabilistically checkable proof systems which use logarithmic randomness and upto three queries. Some of the qualitative assertions are well-known; for example, when considering perfect completeness, 3 queries are the minimum needed (and sufficient [ALMSS]) to get above P.

Proposition 10.3 (PCP systems with logarithmic randomness and at most 3 queries):

(1) (PCP with 1 query is weak): For all admissible functions $s, c : \mathcal{Z}^+ \to [0,1]$, so that s is strictly smaller than c, $\mathrm{PCP}_{c,s}[\log, 1] = \mathrm{P}$.

(2) (One-sided error pcp with 2 queries is weak): For all admissible functions $s : \mathcal{Z}^+ \to [0,1]$ strictly less than 1, $\mathrm{PCP}_{1,s}[\log, 2] = \mathrm{P}$.

(3) (Two-sided error pcp with 2 queries is not weak): There exists $0 < s < c < 1$ so that $\mathrm{PCP}_{c,s}[\log, 2] = \mathrm{NP}$. Furthermore, this holds for some $c > 0.9$ and $s < \frac{73}{74}c$.

(4) (One-sided error pcp with 3 queries is not weak): $\mathrm{PCP}_{1,0.85+\epsilon}[\log, 3] = \mathrm{NP}$, $\forall \epsilon > 0$.

(5) (One-sided error pcp with 3 queries is not very strong): $\forall s < 0.18$, $\mathrm{PCP}_{1,s}[\log, 3] = \mathrm{P}$. Furthermore, $\forall s \leq 0.299$, $\mathbf{na}\mathrm{PCP}_{1,s}[\log, 3] = \mathrm{P}$, where $\mathbf{na}\mathrm{PCP}$ is a restriction of PCP in which the verifier is required to be non-adaptive.

We remark that $\mathrm{PCP}_{1,0.8999}[\log, 3] = \mathrm{NP}$ with a non-adaptive verifier was presented in an earlier version of this paper [BGS2]. Using Proposition 10.2, we have $\mathrm{MIP}_{1,0.95}[\log, 3] = \mathrm{NP}$.

Proof of Proposition 10.3, Part (1): An oracle π maximizing the acceptance probability can be constructed by scanning all possible random pads (random strings) and setting $\pi(q)$ so that it "satisfies" the majority of random-pads for which the verifier makes query q. ▌

Proof of Proposition 10.3, Part (2): The folklore proof commonly deals only with the non-adaptive case. In general, the verifier V, demonstrating that $L \in \mathrm{PCP}_{1,s}[\log, 2]$, may be adaptive. We assume, without loss of generality, that V always makes at least one query. Thus, after making the first query, V decides whether to accept, reject or make an additional query and accept only a specific answer for it. Thus, the computation of V on input x, random pad $\bar{c}$ and access to a generic oracle can be captured by two Horn clauses, each corresponding to a different answer-value for the first query. Specifically, suppose that V queries the oracle at location i and upon receiving value σ accepts iff location j have value τ. Then, we write the Horn clause $\pi_i^\sigma \rightarrow \pi_j^\tau$. (In case V always accepts (resp., rejects) after obtaining value σ from oracle location i, we write the clause $\pi_i^\sigma \rightarrow \mathbf{T}$ (resp., $\pi_i^\sigma \rightarrow \mathbf{F}$).) In addition, for every i, we write the Horn clauses $\pi_i^0 \rightarrow (\neg\pi_i^1)$ and $(\neg\pi_i^0) \rightarrow \pi_i^1$. Thus, the computation of V on input x and access to a generic oracle can be captured by a Horn formula, denoted ϕ_x, in which Horn clauses correspond to the various (polynomially many) possible (random-pad,first-answer) pairs. Furthermore, ϕ_x can be constructed in polynomial-time given x (and V). Using a (polynomial-time) decision procedure for satisfiability of Horn Formulae, we are done. (Alternatively, we can use the linear-time decision procedure for 2-SAT due to Even et. al. [EIS].) ▌

Proof of Proposition 10.3, Part (4): To see that $\mathrm{PCP}_{1,s}[\log, \mathrm{poly}] \subseteq \mathrm{NP}$, for every $s < 1$, consider a non-deterministic machine which tries to guess an oracle which makes the verifier (of the above system) always accept. The other direction (of Part (4)) is shown in Theorem 4.5. ▌

Proof of Proposition 10.3, Part (3): To see that $\mathrm{PCP}_{c,s}[\log, \mathrm{poly}] \subseteq \mathrm{NP}$, for every $s < c$, consider a non-deterministic machine which tries to guess an oracle which makes the verifier accept with probability at least c. The $\mathrm{NP} \subseteq \mathrm{PCP}_{c,s}[\log, 2]$ result follows from the hardness of approximating Max2SAT. Specifically, suppose that $L \leq_D^K \mathsf{Gap}\text{-}2\mathrm{SAT}_{c,s}$. Then we can present a $\mathrm{PCP}_{c,s}[\log, 2]$ system for L as follows. On input x, the verifier in this system performs the reduction (of L to the promise problem) obtaining a 2CNF formula ϕ_x. Next it uniformly selects a clause of ϕ_x and queries the oracle for the values of the variables in this clause (accepting accordingly). Using Theorem 4.6 (Part 3), $\mathrm{NP} \leq_D^K \mathsf{Gap}\text{-}2\mathrm{SAT}_{c,s}$ for some $c > 0.9$ and $s < \frac{73}{74} \cdot c$, and $\mathrm{NP} \subseteq \mathrm{PCP}_{c,s}[\log, 2]$ follows. ▌

Remark 10.4 *The ratio c/s has been subsequently increased to $(10/9) - \epsilon$, for any $\epsilon > 0$ (cf., [TSSW, H3]).*

Proof of Proposition 10.3, Part (5): The result for general verifiers follows from Lemma 4.11 and the fact that MaxSAT can be approximated to within a $0.795 = 0.75 + \frac{0.18}{4}$ factor in polynomial-time (cf., [TSSW]). The (tedious) proof of the non-adaptive case can be found in earlier versions of this paper [BGS2]. The paper of Trevisan et. al. [TSSW] contains a stronger result which holds for all verifiers; that is, $\mathrm{PCP}_{1,0.367}[\log, 3] = \mathrm{P}$. ▌

The latter result (i.e., $\mathrm{PCP}_{1,0.367}[\log, 3] = \mathrm{P}$) is weaker than what can be proven for MIP proof systems (see next corollary). This contrast may provide a testing ground to separate PCP from MIP, a question raised by [BGLR].

Corollary 10.5 For $s < 1/2$, $\mathrm{MIP}_{1,s}[\mathsf{coins} = \log, \mathsf{provers} = 3] = \mathrm{P}$.

Proof: Combining (the two parts of) Lemma 10.1 and (Part 2 of) Proposition 10.3, we have $\mathrm{MIP}_{1,s}[\log, 3] \subseteq \mathrm{MIP}_{1,2s}[\log, 2] \subseteq \mathrm{PCP}_{1,2s}[\log, 2] \subseteq \mathrm{P}$. ▌

A general result which relates the query complexity of a probabilistically checkable proof system and the ratio between the acceptance probabilities of yes-instances and no-instances, follows –

Lemma 10.6 For all admissible functions c, s, q, r, l such that $\frac{c}{s} > 2^q$,

$$\mathrm{PCP}_{c,s}[r, q] \subseteq \mathrm{RTIME}\left(\mathrm{poly}\left(\frac{n}{c - 2^q s}\right)\right)$$

Furthermore, $\mathrm{PCP}_{c,s}[r, q] \subseteq \mathrm{PSPACE}$, and if r and q are both logarithmically bounded then $\mathrm{PCP}_{c,s}[r, q] = \mathrm{P}$.

Proof: Let $L \in \mathrm{PCP}_{c,s}[r, q]$ and V be a verifier demonstrating this fact. Observe that for $x \in L$, the probability that V accepts x, given access to a random oracle, is at least $\frac{c}{2^q}$. On the other hand, for $x \notin L$, the probability that V accepts x, given access to any oracle, is at most $s < \frac{c}{2^q}$. Thus, we can decide if x is in L by simulating the execution of V with access to a random oracle and estimating the acceptance probability, over V's random choices and all possible oracles. In particular, we can estimate this probability upto an $\epsilon \overset{\text{def}}{=} \frac{1}{2} \cdot (s - \frac{c}{2^q})$ additive term, with very high probability, by taking $\mathrm{poly}(1/\epsilon)$ samples. Alternatively, we can compute this probability in polynomial-space. Finally, in case r and q are both logarithmically bounded, we can (exactly) compute the probability that V accepts x, given access to a random oracle. To this end we loop through all possible random-pads for V and for each pad consider all possibilities of setting the oracle bits examined by V. Thus, for $s < \frac{c}{2^q}$, we get a deterministic polynomial-time decision procedure. ▌

The last assertion in the above lemma (i.e., $\mathrm{PCP}_{c,s}[\log, q] = \mathrm{P}$ for $\frac{c}{s} > 2^q$) cannot be strengthen by omitting the (logarithmic) bound on q since $\mathrm{NP} = \mathrm{PCP}_{1,0}[0, \mathrm{poly}]$. On the other hand, recalling the definition of $\overline{\mathrm{PCP}}$ we immediately get

Corollary 10.7 Let $\epsilon : \mathcal{Z}^+ \to [0, 1]$ be an admissible function strictly greater than 0. Then, for every admissible function $c : \mathcal{Z}^+ \to [0, 1]$,

$$\overline{\mathrm{PCP}}_c[\log,\ 1 - \epsilon] = \mathrm{P}$$

In particular, this holds for $c = 1$.

Proof: $L \in \overline{\mathrm{PCP}}_c[\log,\ 1 - \epsilon]$ implies that for some logarithmically bounded function m, we have $L \in \mathrm{PCP}_{c,2^{-m}\cdot c}[\log, (1 - \epsilon) \cdot m]$ and the corollary follows. ▌

PCP WITH SUPER-LOGARITHMIC RANDOMNESS. The above results are focused on pcp systems with logarithmic randomness. Proof systems with unrestricted randomness (as considered in the next proposition) may also provide some indication to the effect of very low query complexity. The results we obtain are somewhat analogous to those of Proposition 10.3. Recall that $\mathrm{PCP}_{1,\frac{1}{2}}[\mathrm{poly},\mathrm{poly}]$ equals NEXPT (Non-deterministic exponential time) [BFL]. Thus, the power of pcp systems with polynomial randomness has to be compared against NEXPT.

Proposition 10.8 (general PCP systems with at most 3 queries):

(1) (PCP with 1 query is relatively very weak): For all admissible functions $s, c : \mathcal{Z}^+ \to [0,1]$, so that $c(n) - s(n)$ is non-negligible[10]

$$\mathrm{PCP}_{c,s}[\mathrm{poly}, 1] \subseteq \mathrm{AM}$$

where AM is the class of languages having one-round Arthur-Merlin proof systems (cf., [Bab]).

(2) (One-sided error pcp with 2 queries is relatively weak): For all admissible functions $s : \mathcal{Z}^+ \to [0,1]$ strictly less than 1, $\mathrm{PCP}_{1,s}[\mathrm{poly}, 2] \subseteq \mathrm{PSPACE}$.

(3) (Two-sided error pcp with 2 queries is not weak): On the other hand, there exists $0 < s < c < 1$ so that $\mathrm{PCP}_{c,s}[\mathrm{poly}, 2] = \mathrm{NEXPT}$.

(4) (One-sided error pcp with 3 queries is not weak): $\mathrm{PCP}_{1,0.85+\epsilon}[\mathrm{poly}, 3] = \mathrm{NEXPT}$, $\forall \epsilon > 0$.

(5) (One-sided error pcp with 3 queries is not very strong): $\forall s < \frac{1}{8}$, $\mathrm{PCP}_{1,s}[\mathrm{poly}, 3] = \mathrm{PSPACE}$. Furthermore, $\forall s \leq 0.299$, $\mathbf{na}\mathrm{PCP}_{1,s}[\mathrm{poly}, 3] = \mathrm{PSPACE}$.

The first part of the proposition may be hard to improve since, as indicated in Proposition 10.9 Part (6), Graph Non-Isomorphism is in $\mathrm{PCP}_{1,\frac{1}{2}}[\mathrm{poly}, 1]$.

Proof of Proposition 10.8, Part (1): We first observe that a 1-query pcp system is actually a one-round interactive proof system (cf., [GMR]). (The completeness and soundness bounds are as in the pcp system.) Using well-known transformations we obtain the claimed result. Specifically, we first reduce the error of the interactive proof by parallel repetition, next transform it into an Arthur-Merlin interactive proof [GS], and finally transform it into an Arthur-Merlin interactive proof of perfect completeness [FGMSZ]. We stress that all the transformations maintain the number of rounds upto a constant and that the constant-round Arthur-Merlin hierarchy collapses to one-round [Bab]. ▌

Proof of Proposition 10.8, Parts (3) and (4): For these parts we observe that the proof systems used in the corresponding parts of the proof of Proposition 10.3, do "scale-up". Specifically, it is easy to see that the outer verifier used for all proof systems in this paper does scale-up, yielding a canonical outer verifier of randomness complexity $O(\log(T(n))$ for any language in $\mathrm{Ntime}(T(n))$, provided $n < T(n) < 2^{\mathrm{poly}(n)}$. Furthermore, all inner-verifiers used in the paper operate on constant sized oracles and so the composed verifier maintains the time and randomness complexities of the outer verifier. In particular, the verifier used for establishing Theorem 4.5 can be scaled-up to yield Part (4). The same holds for the verifier used for establishing Part (3) of Proposition 10.3. (Note that although the exposition of the proof in Proposition 10.3 is in terms of reducing NP to Max2SAT, what actually happens is that the verifier used to establish the NP-hardness of Max2SAT (cf., Section 4.2) is implemented by a verifier which makes only two queries (out of a constant number of possibilities).) ▌

Proof of Proposition 10.8, Part (2): Following the strategy of the proof of the analogous part in Proposition 10.3, we obtain a polynomial-space reduction of $L \in \mathrm{PCP}_{1,s}[\mathrm{poly}, 2]$ to the set of satisfiable 2-Horn formulae (i.e., Horn formulae in which each clause has at most 2 literals). Namely, on input x, the reduction uses space $\mathrm{poly}(|x|)$ and produces a Horn formula ϕ_x (of size exponential in $|x|$) so

[10] A function $f : \mathcal{Z}^+ \to \mathcal{Z}^+$ is called non-negligible if there exists a positive polynomial p so that $\forall n :\ f(n) > \frac{1}{p(n)}$.

that $x \in L$ iff ϕ_x is satisfiable. Using a poly-logarithmic decision procedure for satisfiability of 2-Horn formulae[11], we can decide if ϕ_x is satisfiable using poly($|x|$)-space. ▮

Proof of Proposition 10.8, Part (5): The result for non-adaptive verifiers follows from Part (2) by using the same strategy as in the analogous proof in Proposition 10.3. The result for general verifiers follows by the Furthermore-part of Lemma 10.6 (i.e., $\mathrm{PCP}_{c,s}[\mathrm{poly}, q] = \mathrm{PSPACE}$ for $\frac{c}{s} > 2^q$). ▮

10.3 Free-bit complexity

The class $\mathrm{FPCP}_{c,s}[r, f]$ is defined analogously to the class $\mathrm{PCP}_{c,s}[r, q]$, except that we consider the free-bit complexity (denoted f) instead of the query complexity (denoted q). The following proposition demonstrates the limitations of probabilistically checkable proof systems with free-bit complexity bounded by 1. We do not believe that similar limitations hold for amortized free-bit complexity.[12]

The first three items refer to proof systems with logarithmic randomness. The second item shows that such systems with perfect completenss and free-bit complexity 1 only exists for P (and are hence weak). In contrast, the first item shows the crucial role of perfect completeness in the former negative result: Specifically, proof systems with two-sided error (non-perfect completeness) having free-bit complexity *zero* suffice for $\mathcal{NP}$. The third item asserts that the second item cannot be strengthened with respect to increasing the free-bit complexity. Proof systems with unrestricted randomness (as considered in the last 3 items) may also provide some indication to the effect of very low free-bit complexity. The last item can be viewed as (weak) evidence that the result in the fourth item cannot be "drastically improved" (e.g., to yield $\mathrm{FPCP}_{1,s}[\mathrm{poly}, 0] \subseteq \mathrm{BPP}$).

We make essential use of the ability to efficiently generate accepting computations, and the results may not hold otherwise.[13]

Proposition 10.9 (PCP systems with low free-bit complexity): Let $s : \mathcal{Z}^+ \to [0, 1]$ be an admissible function strictly smaller than 1. Then,

(1) (PCP with logarithmic randomness and 0 free-bit):
- There exists $s < 0.794$ so that $\mathrm{NP} \subseteq \mathrm{FPCP}_{\frac{1}{4},\frac{s}{4}}[\log, 0]$. Thus, $\mathrm{NP} = \overline{\mathrm{FPCP}}_{\frac{1}{4}}[\log, 0]$.
- For every $\epsilon > 0$, $\mathrm{NP} \subseteq \mathrm{FPCP}_{1-\epsilon, 1-\frac{16}{15}\cdot\epsilon}[\log, 0]$.
- For every $\epsilon > 0$, $\mathrm{FPCP}_{1-\epsilon, 1-2\cdot\epsilon}[\log, 0] \subseteq \mathrm{P}$.

(2) (Limitations of PCP with logarithmic randomness and 1 free-bit):
$\mathrm{FPCP}_{1,s}[\log, 1] = \mathrm{P}$. Also, $\mathrm{FPCP}_{1,1-(1/\mathrm{poly})}[\mathsf{coins} = \mathrm{poly}\ ;\ \mathsf{free} = 1\ ;\ \mathsf{pflen} = \mathrm{poly}] \subseteq \mathrm{BPP}$.

(3) ("Tightness" of Item 2): There exists $s < 0.794$ so that
- $\mathrm{NP} \subseteq \mathrm{FPCP}_{1,s}[\log, 2]$;
- $\mathrm{NP} \subseteq \mathrm{FPCP}_{1,\frac{1+s}{2}}[\log, f]$ where $f = \log_2 3$ (i.e., $2^f = 3$);

[11]For example, consider the following procedure. Given a 2-Horn formula, we construct a directed graph in which the vertices are the literals of the formula and there is an directed edge from literal x to literal y if the formula contains the clause $x \to y$. One can easily verify that the formula is not satisfied iff there exists a variable for which every truth assignment yields a contradiction (i.e., "forcing paths" to contradicting values – cf., [EIS]). Thus, a non-deterministic logspace machine can guess this variable and check that both possible truth assignments (to it) yield contradictions. The latter checking reduces to guessing the variable for which a conflicting assignment is implied and verifying the conflict via s-t directed connectivity. Since the latter task is in $\mathcal{NL}$, we are done. (Actually, 2SAT is complete for co$\mathcal{NL}$; see [JLL].)

[12] The conjecture was stated for systems with perfect completeness, and has been subsequently proven by Håstad [H2] (who proved that $\mathrm{NP} = \overline{\mathrm{FPCP}}_1[\log, \epsilon]$, for every $\epsilon > 0$). For systems with two-sided error probability, we knew that they can recognize $\mathcal{NP}$ languages using zero free-bits – see below.

[13] We note, however, that the more relaxed notion of free-bits may be less relevant to proving hardness of approximation results.

- $\mathrm{NP} \subseteq \mathrm{FPCP}_{\frac{1}{2},\frac{s}{2}}[\log, 1]$.

(4) (General pcp with 0 free-bit): $\mathrm{FPCP}_{1,s}[\mathrm{poly}, 0] \subseteq \mathrm{coNP}$.

(5) (general pcp with 1 free-bit): $\mathrm{FPCP}_{1,s}[\mathrm{poly}, 1] \subseteq \mathrm{PSPACE}$.

(6) (Examples for pcp with 0 free-bit): Graph Non-Isomorphism, GNI, has a PCP system with perfect completeness and soundness bound $\frac{1}{2}$, in which the verifier makes a single query and this query is free. Namely,

$$\mathtt{GNI} \in \mathrm{FPCP}_{1,\frac{1}{2}}[\mathsf{coins} = \mathrm{poly}\ ;\ \mathsf{free} = 0\ ;\ \mathsf{query} = 1]$$

The same holds for QNR ("Quadratic Non-Residuosity" (cf., [GMR])) the set of integer pairs (x, N) so that x is a quadratic non-residue modulo N.

Proof of Proposition 10.9, Part (3): The first claim of Part 3 is justified by Theorem 5.4. Applying Proposition 11.9 to this verifier (which indeed satisfies the condition of this proposition), yields the second claim of Part 3. Applying Proposition 11.8 to the same verifier (with $k = 1 < f = 2$), the third claim of Part 3 follows. ▌

Proof of Proposition 10.9, Part (1): Applying Proposition 11.8 (with $k = f = 2$) to the the verifier of Theorem 5.4, the first claim of Part 1 follows. To prove the second claim, we apply Proposition 11.10 to the first claim and obtain $\mathrm{NP} \subseteq \mathrm{FPCP}_{1-\delta\cdot(1-0.25),1-\delta\cdot(1-0.2)}[\log, 0]$ (which holds for any δ). Substituting $\delta = \frac{4}{3}\epsilon$, the second claim follows.

The last claim follow by the relationship between the Minimum Vertex Cover problem and the class $\mathrm{FPCP}_{c,s}[\log, 0]$ – see proof of Proposition 5.6. Specifically, consider the FGLSS reduction/graph of a proof system witnessing $L \in \mathrm{FPCP}_{1-\epsilon,1-2\epsilon}[\log, 0]$ (actually consider the complement graph where one asks about the size of the independent set). Then, for each $x \in L$ this graph has a vertex cover of density at most ϵ, whereas for $x \notin L$ this graph has no vertex cover of density 2ϵ. Using Gavril's approximation algorithm (cf. [GJ2]), these two cases are distinguishable in polynomial-time and so the third claim follows. ▌

Proof of Proposition 10.9, Part (4): Let $L \in \mathrm{PCP}_{1,s}[\mathrm{poly}, 0]$ and V be a verifier demonstrating this fact. By definition, for every possible sequence of coin tosses for V, there exists at most one accepting configuration (of oracle answers to the queries made by V). Furthermore, by definition, this accepting configuration (if it exists) can be generated in polynomial time, from the coin-sequence. Following is a non-deterministic procedure that accepts $\overline{L}$. It starts by guessing two sequences of coin tosses for V, generating the corresponding accepting configurations and checking whether they are consistent. (The input is accepted by this non-deterministic procedure iff the two coin-sequences guessed yield conflicting configurations.) Clearly, if $x \in L$ then, for all possible pairs of coin-sequences, accepting configurations exist and are consistent (since an oracle which *always* makes V accept x does exist). Thus, $x \in L$ is never accepted by the non-deterministic procedure. On the other hand, if all pairs of coin-sequences yield accepting and mutually consistent configurations then an oracle which *always* makes V accept x emerges. Thus, for every $x \notin L$ there exists a guess which makes the non-deterministic procedure accept x. ▌

Proof of Proposition 10.9, Parts (2) and (5): Here we consider proofs with free-bit complexity 1. Thus, for each possible sequence of coin tosses, there exist at most two accepting configurations (which again can be efficiently found given the coin-sequence). We refer to these two possible accepting configuration as to the 1-configuration and the 2-configuration of the coin-sequence. In case a specific coin-sequence has less than two accepting configurations, we introduce dummy configurations so that now each coin-sequence has two associated configurations. Given an input x to such a pcp system, we consider the following 2CNF formula representing all possible computations of the verifier with a generic oracle. For each possible sequence of coin tosses, $\overline{c}$, we introduce a pair of Boolean variables, $\pi^1_{\overline{c}}$ and $\pi^2_{\overline{c}}$, representing which of the two associated configurations is encountered (e.g., $\pi^1_{\overline{c}} = T$ means that the 1-configuration is encountered). To enforce that a single accepting configuration is encountered we introduce the clauses $(\pi^1_{\overline{c}} \vee \pi^2_{\overline{c}})$ and $((\neg\pi^1_{\overline{c}}) \vee (\neg\pi^2_{\overline{c}}))$. In addition, in case the σ-configuration of $\overline{c}$ is not accepting (but rather a dummy configuration) we introduce the clause $(\neg\pi^\sigma_{\overline{c}})$ thus "disallowing" a computation in which it is encountered. Finally, for each pair of coin-sequences we introduce clauses disallowing inconsistencies. Namely, suppose that the σ-configuration of $\overline{c}$ is inconsistent with the τ-configuration of $\overline{c}'$, then we introduce the clause $((\neg\pi^\sigma_{\overline{c}}) \vee (\neg\pi^\tau_{\overline{c}'}))$, which is logically equivalent to $\neg(\pi^\sigma_{\overline{c}} \wedge \pi^\tau_{\overline{c}'})$. The resulting 2CNF formula, ϕ_x, is satisfiable if and only if there exists an oracle which causes V to accept x with probability 1. Thus, given x, we need to test if ϕ_x is satisfiable. We consider two cases.

(1) In case V uses logarithmically many coins, the 2CNF formula ϕ_x can be generated from x in polynomial-time. Using a polynomial-time decision procedure for satisfiability of 2CNF formulae, we conclude that $\mathrm{FPCP}_{1,s}[\log, 1] = \mathrm{P}$. Furthermore, using Proposition 11.2, we can randomly reduce $\mathrm{FPCP}_{1,1-(1/\mathrm{poly})}[\mathrm{poly}, \mathsf{free} = 1, \mathsf{pflen} = \mathrm{poly}]$ to $\mathrm{FPCP}_{1,1-(1/\mathrm{poly})}[\log, \mathsf{free} = 1]$, and $\mathrm{FPCP}_{1,1-(1/\mathrm{poly})}[\mathrm{poly}, \mathsf{free} = 1, \mathsf{pflen} = \mathrm{poly}] \subseteq \mathrm{BPP}$ follows. This establishes Part (2).

(2) In general (V may make polynomially many coin tosses), the 2CNF formula ϕ_x may have exponential (in $|x|$) length. Yet, it can be generated from x in polynomial-space. Using a polylogarithmic-space decision procedure for satisfiability of 2CNF formulae[14], we can decide if ϕ_x is satisfiable using poly($|x|$)-space. Part (5) (i.e., $\mathrm{FPCP}_{1,s}[\mathrm{poly}, 1] \subseteq \mathrm{PSPACE}$) follows.

▌

Proof of Proposition 10.9, Part (6): We merely note that the interactive proof presented in [GMW] for Graph Non-Isomorphism[15] constitute a 1-query pcp system with perfect completeness and soundness bound $\frac{1}{2}$. Furthermore, the query made by the verifier has a unique acceptable answer and thus the free-bit complexity of this system is zero. The same holds for the interactive proof presented in [GMR] for Quadratic Non-Residuosity `QNR`, which is actually the inspiration to the proof in [GMW]. ▌

10.4 Query complexity versus free-bit complexity

The following proposition quantifies the intuition that not all queries are "undetermined" (i.e., that the free-bit complexity is lower than the query complexity). Furthermore, as a corollary we obtain that the amortized (average) free-bit complexity is at least 1 unit less than the amortized query complexity.

Proposition 10.10 For admissible functions c, s, r, q such that $r(n), q(n) = O(\log n)$.

$$\mathrm{PCP}_{c,s}[r, q] \subseteq \mathrm{PCP}_{c,s}[\,\mathsf{coins} = r\,;\, \mathsf{free}_{\mathrm{av}} = q - \log_2(1/s)\,] \tag{23}$$

Furthermore, for every admissible function t, $\mathrm{PCP}_{c,s}[r, q] \subseteq \mathrm{FPCP}_{c,(2^t+1)\cdot s}[r, q - t]$.

Proof: Let $L \in \mathrm{PCP}_{c,s}[r, q]$ and let V be the verifier demonstrating this. Fix an input $x \in \Sigma^n$, and let $r = r(n), q = q(n), s = s(n)$. For a random string $R \in \{0,1\}^r$, let F^x_R denote the number of

[14]For example, note that 2CNF formulae can be written in Horn form and use the procedure described in the proof of Proposition 10.8 Part (2).

[15]On input a pair of graphs, G_0 and G_1, the verifier uniformly selects $i \in \{0, 1\}$ and generates a random isomorphic copy of G_i, denoted H. This graph H is the single query made by the verifier, which accepts if and only if the answer equals i.

accepting patterns of V, i.e., $F_R^x = |\mathsf{pattern}_V(x;R)|$. We first claim that if $\mathbf{E}_R\left[F_R^x\right] > 2^q \cdot s$, then $x \in L$. This is true since a random oracle π is accepted with probability at least $\mathbf{E}_R\left[F_R^x \cdot 2^{-q}\right]$, and so if the claim were not to hold we would have reached contradiction to the soundness condition (i.e., $x \notin L$ is accepted with probability strictly larger than s).

We now construct a verifier, denoted V', witnessing $L \in \mathrm{FPCP}^{\mathrm{av}}_{c,s}[r, q - \log_2(1/s)]$. On input x, the verifier first computes $\mathbf{E}_R\left[F_R^x\right]$ (by scanning all possible R's and generating all accepting patterns for each of them). If $\mathbf{E}_R\left[F_R^x\right]) > 2^q \cdot s$, then V' accepts x (without querying the oracle). Otherwise (i.e., if $\mathbf{E}_R\left[F_R^x\right]) \leq 2^q \cdot s$), then V' simulates V and accepts if V accepts. It follows that the average free-bit complexity of V' on input x equals the corresponding quantify for V, provided the latter is at most $q - \log_2(1/s)$, and equals zero otherwise. The first part of the proposition follows.

To establish the second part, for some $t = t(n)$, we construct a verifier V'' which, on input x, proceeds as follows. First, V'' computes $q \stackrel{\mathrm{def}}{=} \mathbf{E}_R\left[F_R^x\right]$ and accepts if $q > s2^q$ (just as V'). In case $q \leq s2^q$, the new verifier proceeds differently: It randomly selects R as V does and computes F_R^x. If $F_R^x > 2^{q-t}$ then V'' accepts and otherwise it invokes V on input x and coins R. Clearly, this guarantees that the free-bit complexity of V'' is at most $q - t$. To analyze the soundness of V'', note that when $\mathbf{E}_R\left[F_R^x\right] \leq s2^q$, it follows that $\Pr_R\left[F_R^x > 2^{q-t}\right] \leq 2^t \cdot s$ (Markov Inequality). Thus, the soundness error of V'' is at most $s + 2^t s$ and the second part follows. ▌

By computing the amortized average free-bit complexity of the class of languages in the right hand side of Eq. (23) above, we obtain the following consequence.

Corollary 10.11 For admissible functions c, r, q with $r(n), q(n) = O(\log n)$,

$$\overline{\mathrm{PCP}}_c[\,r,\,q\,] \subseteq \overline{\mathrm{FPCP}}^{\mathrm{av}}_c[\,r,\,q-1\,].$$

where $\overline{\mathrm{FPCP}}^{\mathrm{av}}_{\cdot}[\,\cdot,\,f\,]$ denotes a class analogous to $\overline{\mathrm{FPCP}}_{\cdot}[\,\cdot,\,f\,]$ in which average free-bit complexity is measured instead of (worst-case) free-bit complexity.

Proof: For some function m, we have

$$\overline{\mathrm{PCP}}_c[\,r,\,q\,] \subseteq \mathrm{PCP}_{c,c\cdot 2^{-m}}[r, qm] \subseteq \mathrm{FPCP}^{\mathrm{av}}_{c,c\cdot 2^{-m}}[r, qm-m] \subseteq \overline{\mathrm{FPCP}}^{\mathrm{av}}_c[\,r,\,q-1\,].$$

where the second inclusion is due to Eq. (23). ▌

The above corollary clinches the argument that the amortized query complexity is incapable of capturing the approximability of the clique function. Previously we had argued thus based on the assumption that the clique number may be hard to approximate to within $N^{\frac{1}{2}}$ (i.e., establishing such a clique NP-hardness would require showing that $\mathrm{NP} \subseteq \overline{\mathrm{PCP}}[\,\log,\,1-\epsilon\,]$, for every $\epsilon > 0$, which is impossible[16] as we've shown that $\overline{\mathrm{PCP}}[\,\log,\,1-\epsilon\,] \subseteq \mathrm{P}$). Now, we can remove this assumption also.[17] Suppose that, for some g (e.g., $g = \frac{3}{2}$), MaxClique is NP-hard to approximate to within a $N^{1/(1+g)}$ factor, but it can be approximated to within a $N^{1/(1+g-\delta)}$ factor in polynomial-time, for every $\delta > 0$ (actually, it suffices to postulate that MaxClique can be approximated to within a $N^{1/g}$ factor in polynomial-time). Furthermore, supposed that the hardness result is demonstrated by showing that $\mathrm{NP} \subseteq \overline{\mathrm{PCP}}[\,\log,\,g-\epsilon\,]$, for every $\epsilon > 0$. Then, using the above corollary, we get $\mathrm{NP} \subseteq \overline{\mathrm{FPCP}}^{\mathrm{av}}[\,\log,\,g-1-\epsilon\,]$, for every $\epsilon > 0$, and an NP-hardness result of clique approximation[18] upto a $N^{1/(1+(g-1-\epsilon)+\epsilon)} = N^{1/g}$ follows,

[16]The entire discussion assumes $\mathrm{P} \neq \mathrm{NP}$. The discussion is anyhow moot otherwise.

[17]In retrospect, there is no reason to remove this assumption as it has been proven to hold in [H2]. However, this was not known at the time the current work was done.

[18] Here we use the observation that the FGLSS-reduction works also for amortized *average* free-bit complexity.

in contradiction to our hypothesis that such approximations could be achieved in polynomial time. To summarize, attempts to establish the factor $N^{1/(g+1)}$ for which it is NP-hard to approximate Max-Clique via amortized query complexity will always fall at least one unit away from the truth; whereas amortized free-bit complexity will yield the right answer.

11 Transformations of FPCP Systems

We present several useful transformations which can be applied to pcp systems. These fall into two main categories:

(1) Transformations which amplify the (completeness versus soundness) gap of the proof system, while preserving (or almost preserving) its amortized free-bit complexity.

(2) Transformations which move the gap location (or, equivalently, the completeness parameter). The gap itself is almost preserved but moving it changes the free-bit complexity (and thus the amortized free bit complexity is not preserved). Specifically, moving the gap 'up' requires increasing the free-bit complexity, whereas moving the gap 'down' allows to decrease the free-bit complexity.

Most of these transformations are analogous to transformations which can be applied to graphs with respect to the Max-Clique approximation problem. In view of the relation between FPCP and the clique promise problem (shown in Section 8), this analogy is hardly surprising.

In this section, we use a more extensive FPCP notation which refers to promise problems (rather than to languages) and introduce an additional parameter – the proof length. Specifically, $\mathrm{FPCP}_{c,s}[r,f,l]$ refers to randomness complexity r, free-bit complexity f and proof-length l.

11.1 Gap amplifications maintaining amortized free-bit complexity

We start by stating the simple fact that the ratio between the completeness and soundness bounds (also referred to as gap) is amplified (i.e., raise to the power k) when one repeats the pcp system (k times). Note, however, that if the original system is not perfectly complete then the completeness bound in the resulting system gets decreased.

Proposition 11.1 (simple gap amplification): For all $c,s : \mathcal{Z}^+ \to [0,1]$ and $k : \mathcal{Z}^+ \to \mathcal{Z}^+$,

$$\mathrm{FPCP}_{c,s}[r,f,l] \subseteq \mathrm{FPCP}_{c^k,s^k}[kr,kf,l].$$

Proof: Let $(Y,N) \in \mathrm{FPCP}_{c,s}[r,f,l]$ and let V be a verifier witnessing this with query complexity $q : \mathcal{Z}^+ \to \mathcal{Z}^+$. Given $k : \mathcal{Z}^+ \to \mathcal{Z}^+$, we define a verifier $V^{(k)}$ as follows: On input $x \in \{0,1\}^n$, let $r = r(n), k = k(n), f = f(n), l = l(n)$ and $q = q(n)$.

- $V^{(k)}$ picks k random strings $\overline{c}^{(1)},\ldots,\overline{c}^{(k)}$ uniformly and independently in $\{0,1\}^r$.
- For $i = 1$ to k, verifier $V^{(k)}$ simulates the actions of V on input x and random string $\overline{c}^{(i)}$. Verifier $V^{(k)}$ accepts if V accepts on each of these k instances.

Clearly, $V^{(k)}$ tosses kr coins and examines the l-bit long oracle in at most kq bits, where at most kf of these are free. For every x, if the probability that V accepts x, given access to oracle π, is p then the probability that $V^{(k)}$ accepts x, given access to π is exactly p^k. Thus, $(Y,N) \in \mathrm{FPCP}_{c^k,s^k}[kr,kf,l]$, and oracles can be transformed (by identity) from one pcp system to the other. ▌

Next, we show that in some sense the randomness-complexity of a proof system need not be higher than logarithmic in the length of the proofs/oracles employed. Specifically, we show how to randomly reduce languages proven by the first kind of systems into languages proven by the second kind. Thus, whenever one is interested in the computational complexity of languages proven via pcp systems, one may assume that the system is of the second type. Recall that $\leq_R^K$ denotes a randomized Karp reduction.

Proposition 11.2 (reducing randomness): There exists a constant $\gamma > 0$ so that

(1) (for perfect completeness): For every two admissible functions $s, \epsilon : \mathcal{Z}^+ \to [0,1]$,

$$\mathrm{FPCP}_{1,s}[r,f,l] \leq_R^K \mathrm{FPCP}_{1,s'}[r',f,l]$$

where $s' = (1+\epsilon)\cdot s$ and $r' = \gamma + \log_2(l/\epsilon^2 s)$.

(2) (for two-sided error): For every four admissible functions $c, s, \epsilon_1, \epsilon_2 : \mathcal{Z}^+ \to [0,1]$,

$$\mathrm{FPCP}_{c,s}[r,f,l] \leq_R^K \mathrm{FPCP}_{c',s'}[r',f,l]$$

where $c' = 1-(1+\epsilon_1)\cdot(1-c) \geq c-\epsilon_1$, $s' = (1+\epsilon_2)\cdot s$
and $r' = \gamma + \max\{-\log_2(\epsilon_1^2(1-c))\ ,\ \log_2(l) - \log_2(\epsilon_2^2 s)\}$.

Proof: The proof is reminiscent of Adleman's proof that $\mathcal{RP} \subseteq \mathrm{P/poly}$ [Ad]. Suppose we are given a pcp system for which we want to reduce the randomness complexity. The idea is that it suffices to choose the random pad for the verifier out of a relatively small set of possibilities (instead than from all 2^r possibilities). Furthermore, most small sets (i.e., sets of size linear in l) are good for this purpose. This suggest randomly mapping an input x for the original pcp system into an input (x,R) for the new system, where R is a random set of $m = O(l)$ possible random-pads for the original system. The new verifier will select a random-pad uniformly in R, thus using only $\log_2|R|$ random coins, and run the original verifier using this random-pad. Details follow.

We start with the simpler case stated in Part (1). Let $(Y,N) \in \mathrm{FPCP}_{1,s}[r,f,l]$ and V be a verifier demonstrating this fact. The random reduction maps $x \in \{0,1\}^n$ to (x,R), where R is a uniformly chosen m-multi-subset of $\{0,1\}^r$ for $l \stackrel{\mathrm{def}}{=} l(n)$, $r \stackrel{\mathrm{def}}{=} r(n)$, $s \stackrel{\mathrm{def}}{=} s(n)$, $\epsilon \stackrel{\mathrm{def}}{=} \epsilon(n)$ and $m \stackrel{\mathrm{def}}{=} \frac{\gamma l}{\epsilon^2 s}$. (The constant γ is chosen to make the Chernoff bound, used below, hold.) On input (x,R), the new verifier V' uniformly selects $\overline{c} \in R$ and invokes V with input x and random-pad $\overline{c}$. Clearly, the complexities of V' are as claimed above. Also, assuming that V always accepts x, when given access to an oracle π then, for every possible pair (x,R) to which x is mapped, V' always accepts (x,R) when given access to the oracle π. It remains to upper bound, for each $x \notin L$ and most R's, the probability that V' accepts (x,R) when given access to an arbitrary oracle.

Fixing any $x \notin L$ and any oracle π, we bound the probability that V', give access to π, accepts (x,R) for most R's. A set R is called *bad* for x with respect to π if for more than a s' fraction of the $\overline{c} \in R$ the verifier V accepts x when given access to π and random-pad $\overline{c}$. Let $R = (\overline{r}^{(1)}, ..., \overline{r}^{(m)})$ be a uniformly selected multi-set. For every $i \in [m]$ (a possible random choice of V'), we define a 0-1 random variable ζ_i so that it is 1 iff V on random-pad $\overline{r}^{(i)}$ and access to oracle π accepts the input x. Clearly, the ζ_i's are mutually independent and each equals 1 with probability $\delta \leq s$. Using a multiplicative Chernoff Bound (cf. [MoRa, Theorem 4.3]), the probability that a random R is bad (for x w.r.t. π) is bounded by

$$\Pr\left[\sum_{i=1}^{m} \zeta_i \geq (1+\epsilon)\cdot ms\right] < 2^{-\Omega(\epsilon^2 \cdot ms)}$$

Thus, by the choice of m, the probability that a random R is bad for x, with respect to any fixed oracle, is smaller than $\frac{1}{4}\cdot 2^{-l}$. Since they are only 2^l relevant oracles, the first part of the proposition follows.

For the second part of the proposition, we repeat the same argument, except that now we need to take care of the completeness bound in the resulting pcp system. This is done similarly to the way we dealt with the soundness bound, except that we do not need to consider all possible oracles – it suffices to consider the best oracle for any $x \in Y$. When applying the multiplicative Chernoff bound it is important to note that, since we are interested in the rejection-event, the relevant expectation is $m \cdot (1-c)$ (and not $m \cdot c$). Thus, as long as $m \geq \frac{2\gamma}{\epsilon_1^2(1-c)}$, at least $\frac{3}{4}$ of the possible sets R cause V' to accept $x \in Y$ with probability at least $1-(1+\epsilon_1)\cdot(1-c) = c-(1-c)\epsilon_1$. The second part of the proposition follows. ▌

Combining Propositions 11.1 and 11.2, we obtain a randomized reduction of pcp systems which yields the effect of Proposition 11.1 at much lower (and in fact minimal) cost in the randomness complexity of the resulting pcp system. This reduction is analogous to the well-known transformation of Berman and Schnitger [BeSc]. The reduction (in either forms), plays a central role in deriving clique approximation results via the FGLSS method: applying the FGLSS-reduction to proof systems obtained via the second item (below), one derives graphs of size $N \stackrel{\text{def}}{=} 2^{(1+\epsilon+f)\cdot t}$ with clique-gap 2^t (which can be rewritten as $N^{1/(1+f+\epsilon)}$). For sake of simplicity, we only state the case of perfect completeness.

Corollary 11.3 (probabilistic gap amplification at minimal randomness cost):

(1) (Combining the two propositions): For every admissible $k : \mathcal{Z}^+ \to \mathcal{Z}^+$,

$$\mathrm{PCP}_{1,1/2}[\mathsf{coins}=r\,;\,\mathsf{query}=q\,;\,\mathsf{free}=f\,;\,\mathsf{pflen}=l\,] \leq_R^K \mathrm{FPCP}_{1,2^{-k+1}}[r+\log_2 q+O(1)+k, kf, l]\,.$$

(2) (using amortized free-bit complexity): For every $\epsilon>0$, there exists a constant c so that

$$\overline{\mathrm{FPCP}}[\log,\, f, l\,] \leq_R^K \mathrm{FPCP}_{1,2^{-t}}[(1+\epsilon)\cdot t, f\cdot t, l]$$

where $t(n) = c\log_2 n$.

Proof: Suppose that $(Y,N) \in \mathrm{FPCP}_{1,1/2}[r,f,l]$. Clearly, $l \leq 2^r \cdot q$, where $q(n) = \mathrm{poly}(n)$ is the query complexity of the verifier. Then, applying Proposition 11.1, we get $(Y,N) \in \mathrm{FPCP}_{1,1/2^k}[kr, kf, 2^r\cdot q]$. Applying Part (1) of Proposition 11.2, we obtain $(Y,N) \leq_R^K \mathrm{FPCP}_{1,\frac{1}{2^{k-1}}}[r', kf]$, where $r' = O(1)+\log_2(2^r q/2^{-k}) = O(1)+r+k+\log_2 q$. The first part of the corollary follows.

Suppose now that a language L has a proof system as in the hypothesis of the second part. Then, there exists a logarithmically bounded function m so that $L \in \mathrm{FPCP}_{1,1/2^m}[r, mf, l]$, where $r(n) \leq \alpha\cdot\log_2 n$ and $l(n) \leq n^\beta$ for some constants α and β. Invoking a similar argument (to the above), we get $L \leq_R^K \mathrm{FPCP}_{1,\frac{1}{2^{km-1}}}[r', k\cdot mf]$, where $r'(n) = O(1)+km+(\alpha+\beta)\cdot\log_2 n$. Now, setting $k(n)$ so that $k(n)\cdot m(n) \geq \frac{\alpha+\beta}{\epsilon}\cdot\log_2 n$, the corollary follows. ▌

An alternative gap amplification procedure which does not employ randomized reductions is presented below. This transformation increases the randomness complexity of the pcp system more than the randomized reduction presented above (i.e., $r' \approx O(r)+2k$ rather than $r' \approx r+k$ as in Item (1) of Corollary 11.3). The transformation is used to obtain in-approximability results under the assumption $\mathrm{P} \neq \mathrm{NP}$ (rather than under $\mathrm{NP} \not\subseteq \mathrm{BPP}$). Again, we only state the case of perfect completeness.

Proposition 11.4 (deterministic gap amplification at low randomness cost): For every $\epsilon, s > 0$ and every admissible function $k : \mathcal{Z}^+ \to \mathcal{Z}^+$

$$\mathrm{FPCP}_{1,s}[r,f,l] \subseteq \mathrm{FPCP}_{1,s^k}[O(r)+(2+\epsilon)\cdot k\cdot\log(1/s), (1+\epsilon)\cdot kf, l].$$

Actually, the constant in the O-notation is $\min\{1, \frac{2+(4/\epsilon)}{\log_2(1/s)}\}$.

We use random walks on expander graphs for error reduction (cf., [AKS, CW]). The value of the constant multiplier of $k\log(1/s)$ in the randomness complexity of the resulting pcp system, depends on the expander graph used. Specifically, using a degree d expander graph with second eigenvalue λ yields a factor of $\frac{\log_2 d}{1+\log_2 \lambda}$. Thus, it is essential to use Ramanujan graphs [LPS] in order to obtain the claimed constant of $2+\epsilon$.

Proof of Proposition 11.4: For simplicity assume $s = 1/2$. The idea is to use a "pseudorandom" sequence generated by a random walk on an expander graph in order to get error reduction at moderate randomness cost. Specifically, we will use a Ramanujan expander graph of constant degree d and second eigenvalue $\lambda \approx 2\sqrt{d}$ (cf., [LPS]). The constant d will be determined so that $d > 2^{4+\frac{8}{\epsilon}}$ (and $d < 2^{6+\frac{8}{\epsilon}}$). It is well-known that a random walk of length t in an expander avoids a set of density ρ with probability at most $(\rho + \frac{\lambda}{d})^t$ (cf., [AKS, Kah]). Thus, as a preparation step, we reduce the error probability of the pcp system to

$$p \stackrel{\text{def}}{=} \frac{\lambda}{d} = \frac{2}{\sqrt{d}} \tag{24}$$

This is done using the trivial reduction of Proposition 11.1. We derive a proof system with error probability p, randomness complexity

$$r' \stackrel{\text{def}}{=} r \cdot \log_2(1/p) = r \cdot \log_2(\sqrt{d}/2) = O(r) \tag{25}$$

and free-bit complexity

$$f' \stackrel{\text{def}}{=} f \cdot \log_2(1/p) = f \cdot \log_2(\sqrt{d}/2) \tag{26}$$

(In case we start with soundness error s, where $s > p$, the multiplier will be $\log_{1/s}(1/p)$ instead of $\log_2(1/p)$.) Now we are ready to apply the expander walk technique. Using an expander walk of length t, we transform the proof system into one in which the randomness complexity is $r' + (t-1)\cdot \log_2 d$, the free-bit complexity is $tf' = tf \cdot \log_2(\sqrt{d}/2)$ and the error probability is at most $(2p)^t = (4/\sqrt{d})^t = 2^{-k}$, where $k \stackrel{\text{def}}{=} t \cdot \log_2(\sqrt{d}/4)$. Using $\log_2 d > \frac{8}{\epsilon} + 4$, we can bound the randomness complexity by

$$\begin{aligned} r' + t\log_2 d &= r' + \frac{\log_2 d}{\frac{1}{2}\cdot(\log_2 d) - 2} \cdot k \\ &< r' + (2+\epsilon)\cdot k \end{aligned}$$

and the free-bit complexity by

$$\begin{aligned} tf \cdot \log_2(\sqrt{d}/2) &= \frac{\frac{1}{2}\cdot(\log_2 d) - 1}{\frac{1}{2}\cdot(\log_2 d) - 2} \cdot kf \\ &< (1+\epsilon)\cdot kf \end{aligned}$$

The proposition follows. ▌

Using Proposition 11.4, we obtain the following corollary which is used in deriving clique inapproximability results under the P $\neq$ NP assumption, via the FGLSS method: applying the FGLSS-reduction to proof systems obtained via this corollary, one derives graphs of size $N \stackrel{\text{def}}{=} 2^{(2+\epsilon+f)\cdot t}$ with clique-gap 2^t (which can be rewritten as $N^{1/(2+f+\epsilon)}$).

Corollary 11.5 For every $\epsilon > 0$ there exists a constant c so that

$$\overline{\text{FPCP}}[\log,\ f, l\,] \subseteq \text{FPCP}_{1,2^{-t}}[(2+\epsilon)\cdot t, (1+\epsilon)f\cdot t, l]$$

where $t(n) = c\log_2 n$.

11.2 Trading-off gap location and free-bit complexity

The following transformation is analogous to the randomized layering procedure for the clique promise problem (i.e., Proposition 8.6). The transformation increases the acceptance probability bounds at the expense of increasing the free-bit complexity.

Proposition 11.6 (increasing acceptance probabilities and free-bit complexity):

(1) (using a randomized reduction which preserves the randomness of the proof system): For all admissible functions $c, s : \mathcal{Z}^+ \to [0,1]$, and $r, f, m : \mathcal{Z}^+ \to \mathcal{Z}^+$,

$$\mathrm{FPCP}_{c,s}[r,f] \leq_R^K \mathrm{FPCP}_{c',s'}[r, f + \log_2 m]$$

where $c' = 1 - 4(1-c)^m$ and $s' = m \cdot s$. In case $c' > 1 - 2^{-r}$, we have then $c' = 1$.

(2) (inclusion which moderately increases the randomness of the proof system): For all admissible functions $c, s : \mathcal{Z}^+ \to [0,1]$, and $r, f, m : \mathcal{Z}^+ \to \mathcal{Z}^+$,

$$\mathrm{FPCP}_{c,s}[r,f] \subseteq \mathrm{FPCP}_{c',s'}[r', f + \log_2 m]$$

- where if $m \leq 1/c$ then $r' = 2 \cdot \max\{r, \log m\}$, $c' = \frac{m}{2} \cdot c$ and $s' = m \cdot s$;
- and otherwise (i.e., for $m > 1/c$), $r' = O(\max\{r, \log m\} + mc)$, $c' = 1 - 2^{-\Theta(mc)}$ and $s' = m \cdot s$.

Proof: Suppose we are given a pcp system for which we want to increase the acceptance probability bound in the completeness condition. The idea is to allow the new verifier to select m random-pads for the original verifier and query the oracle as to which pad to use. A straightforward implementation of this idea will increase the randomness complexity of the verifier by a factor of m. Instead, we use two alternative implementations, which yield the two parts of the proposition. In both implementations the free-bit complexity increases by $\log_2 m$ and the soundness bound increases by a factor of m.

The first implementation employs a technique introduced by Lautemann (in the context of $\mathcal{BPP}$) [Lau]. Using a randomized reduction, we supply the new verifier with a sequence of m possible "shifts" that it may effect. The new verifier selects one random-pad for the original verifier and generates m shifts of this pad. Now, the new verifier queries the oracle as to which of these shifts it should use as a random-pad for the original verifier. Details follow.

We first present a random reduction mapping $x \in \{0,1\}^n$ to (x, S), where S is a uniformly chosen m-multi-subset of $\{0,1\}^r$, for $r \stackrel{\mathrm{def}}{=} r(n)$. On input (x, S), the new verifier V' uniformly selects $\overline{c} \in \{0,1\}^r$ and queries the oracle on $(x, \overline{c})$ receiving an answer $i \in [m]$. Intuitively, V' asks which shift of the random-pad to use. Finally, V' invokes V with input x and random-pad $\overline{c} \oplus \overline{s}_i$, where $\overline{s}_i$ is the i^{th} string in S. Clearly, the complexities of V' are as claimed above. Also, assuming that V accepts x with probability δ, we get that, for every S, verifier V' accepts (x, S) with probability at most $m \cdot \delta$. On the other hand suppose that, when given access to oracle π, verifier V accepts x with probability δ. It follows that there exists a set R of $\delta 2^r$ random-pads for V so that if V uses any $\overline{c} \in R$ (and queries oracle π) then it accepts x. Fixing any $\overline{c} \in \{0,1\}^r$, we ask what is the probability, for a uniformly chosen $S = \{\overline{s}_i : i \leq m\}$, that there exists an $i \in [m]$ so that $\overline{c} \oplus \overline{s}_i \in R$. Clearly, the answer is $1 - (1-\delta)^m$. Thus, for uniformly chosen $S \in (\{0,1\}^r)^m$ and $\overline{c} \in \{0,1\}^r$,

$$\Pr\left[\, \exists i \in [m] \ \text{ s.t. } \ \overline{c} \oplus \overline{s}_i \in R \,\right] = 1 - (1-\delta)^m$$

By Markov Inequality, with probability at least $\frac{3}{4}$, a uniformly chosen $S = \{\overline{s}_i\}$ has the property that for at least $1 - 4 \cdot (1-\delta)^m$ of the $\overline{c}$'s (in $\{0,1\}^r$) there exists an $i \in [m]$ so that $\overline{c} \oplus \overline{s}_i \in R$. Part (1) of the proposition follows.

To prove Part (2) of the proposition, we use an alternative implementation of the above idea, which consists of letting the new verifier V' generate a "pseudorandom" sequence of possible random-pads by itself. V' will then query the oracle as to which random-pad to use, in the simulation of V, and complete its computation by invoking V with the specified random-pad. To generate the "pseudorandom" sequence we use the sampling procedure of [BGG]. Specifically, for $m \leq 1/c$ this merely amounts to generating a pairwise independent sequence of uniformly distributed strings in $\{0,1\}^r$, which can be done using randomness $\max\{2r, 2\log_2 m\}$. Otherwise (i.e., for $m > 1/c$) the construction of [BGG] amounts to generating $\Theta(cm)$ such related sequences, where the sequences are related via a random walk on a constant degree expander. Part (2) follows. ▮

The following corollary exemplifies the usage of the above proposition. In case $c(n) = n^{-\alpha}$ and $r(n) = O(\log n)$, the gap is preserved (upto a logarithmic factor) and the free-bit complexity increases by a $\log_2 1/c$ additive term. Thus, the corollary provides an alternative way of deriving the reverse-FGLSS transformation (say, Proposition 8.7) from the simple clique verifier of Theorem 8.2. Specifically, one may apply the following corollary to the simple clique verifier of Theorem 8.2, instead of combining the layered-graph verifier[19] (of Theorem 8.3), and the graph-layering process of Proposition 8.6.

Corollary 11.7 For all admissible $r, f : \mathcal{Z}^+ \to \mathcal{Z}^+$, so that $\forall n : \; r(n) \geq 2$,

$$\mathrm{FPCP}_{c,s}[r,f] \; \leq_R^K \; \mathrm{FPCP}_{1, r\cdot\frac{s}{c}}[r, f + \log_2 r + \log_2(1/c)]$$

(Compare to Item (1) of Proposition 8.7.)

We conclude this subsection with another transformation which is reminiscent to an assertion made in Section 8. The following transformation has an opposite effect than the previous one, reducing the free-bit complexity at the expense of lowering the bounds on acceptance probability.

Proposition 11.8 (decreasing acceptance probabilities and free-bit complexity): For all admissible functions $c, s : \mathcal{Z}^+ \to [0,1]$, and $r, f, k : \mathcal{Z}^+ \to \mathcal{Z}^+$ so that $k \leq f$, if $L \in \mathrm{FPCP}_{c,s}[r,f]$ then $L \in \mathrm{FPCP}_{\frac{c}{2^k}, \frac{s}{2^k}}[r+k, f-k]$. Furthermore, in case each random-pad in the original pcp system has at least 2^k accepting configurations, the average free-bit complexity of the resulting system is $f_{\mathrm{av}} - k$, where f_{av} is the average free-bit complexity of the original system.

Proof: Let V be a verifier satisfying the condition of the proposition. We construct a new verifier V' that on input $x \in \{0,1\}^n$, setting $r = r(n)$, $k = k(n)$ and $f = f(n)$, acts as follows. Verifier V' uniformly selects a random-pad $\bar{c} \in \{0,1\}^r$ for V, and generates all possible accepting configurations with respect to $V(x)$ and random-pad $\bar{c}$. In case there are less than 2^k accepting configurations we add *dummy configurations* to reach the 2^k count. We now partition the set of resulting configurations (which are accepting and possibly also dummy) into 2^k parts of about the same size (i.e., some parts may have one configuration more than others). Actually, if we only care about average free-bit complexity then any partition of the accepting configurations into 2^k non-empty parts will do. The new verifier, V', uniformly selects $i \in [2^k]$ thus specifying one of these parts, denoted A_i. Next, V' invokes V with random-pad $\bar{c}$ and accepts if and only if the oracle's answers form an accepting configuration which is in A_i (i.e., resides in the selected portion of the accepting configurations). (We stress that in case $\bar{c}$ has less than 2^k accepting configurations and the selected A_i does not contain any accepting configuration then V' rejects on coins $(i, \bar{c})$.) Clearly, the randomness complexity of the new verifier is $r + k$.

[19]which generalizes the simple clique verifier

To analyze the other parameters of V', we fix any $x \in \{0,1\}^n$. For sake of simplicity, we first assume that the number of accepting configurations of V for any random-pad is a power of 2. Then the number of accepting configurations of V' for any random-pad $(\bar{c}, i) \in \{0,1\}^r \times [2^k]$ is 2^{m-k}, where 2^m is the number of accepting configurations of V on random-pad $\bar{c}$. Thus, the free-bit complexity of V' is $f - k$. Finally, we relate the acceptance probability of V' to that of V. This is done by reformulating the execution of V' with oracle π as consisting of two steps. First V' invokes V with access to π. If V reaches a rejecting configuration then V' rejects as well; otherwise (i.e., when V reaches an accepting configuration), V accepts with probability 2^{-k} (corresponding to uniformly selecting $i \in [2^k]$). It follows that on input x and access to oracle π, the verifier V' accepts with probability $\frac{\delta}{2^k}$, where δ denotes the probability that V accepts input x when given access to oracle π.

In general, our simplifying assumption that the number of accepting configurations of V is a power of 2, may not hold and the analysis becomes slightly more cumbersome. Firstly, the number of accepting configurations of V' for a random-pad $(\bar{c}, i)$ is either $\lceil M/2^k \rceil$ or $\lfloor M/2^k \rfloor$, where M is the number of accepting configurations of V on random-pad $\bar{c}$. Thus, in the worse-case the number of accepting configurations for V' (on random-pad $(\bar{c}, i)$) is $\lceil M/2^k \rceil$ and it follows that the free-bit complexity of V' is $\log_2 \lceil 2^f/2^k \rceil = f - k$. Furthermore, the expected number of accepting configurations (for a fixed $\bar{c}$ and uniformly chosen $i \in [2^k]$) is exactly $M/2^k$ (even if $M < 2^k$). Thus, if the extra condition holds then the free-bit complexity of V' equals $f_{\text{av}} - k$. Finally, observe that the argument regarding the acceptance probabilities remains unchanged (and actually it does not depend on the partition of the accepting configurations into 2^k non-empty parts). The proposition follows. ▌

11.3 Other effects on acceptance probabilities and free-bit complexity

Following is an alternative transformation which reduces the free-bit complexity. However, unlike Proposition 11.8, the following does not decrease the acceptance parameters. Furthermore, the transformation increases the soundness parameter and so does *not* preserve the gap (between the completeness and soundness parameters).

Proposition 11.9 (decreasing free-bit complexity without decreasing acceptance probabilities): Let $c, s : \mathcal{Z}^+ \to [0,1]$ be admissible functions and $r, f, k : \mathcal{Z}^+ \to \mathcal{Z}^+$. Suppose $L \in \text{FPCP}_{c,s}[r, f]$ with a verifier for which the first k oracle-answers for each random-pad allow at most 2^{f-k} accepting configurations. Then $L \in \text{FPCP}_{c',s'}[r + k, f']$, where $c' = 1 - \frac{1-c}{2^k}$, $s' = 1 - \frac{1-s}{2^k}$, and $f' = \log_2(2^{f-k} + 2^k - 1)$.

The above can be further generalized; yet the current paper only utilizes the special case in which $c = 1$ (specifically, in the proof of Part 3 in Proposition 10.9, we use $f = 2$ and $k = 1$ obtaining $f' = \log_2 3$, $c' = 1$ and $s' = \frac{1+s}{2}$).

Proof: The proof is similar to the proof of Proposition 11.8. Again, we consider a verifier V as guaranteed by the hypothesis and let A_i be the set of (at most 2^{f-k}) accepting configurations which are consistent with the i^{th} possibility of k oracle-answers to the first k queries. Denote the i^{th} possibility by α_i (i.e., all configurations in A_i start with α_i). We construct a new verifier, V', which uniformly selects a random-pad $\bar{c}$ for V and $i \in [2^k]$ (specifying a part A_i as above). The verifier V' makes the first k queries of V and if the answers differ from α_i then V' halts and accepts.[20] Otherwise, V' continues the emulation of V and accepts iff V accepts.

Clearly, V' uses $r + k$ coin-tosses. The accepting configurations of V' on random-pad $(\bar{c}, i)$ are those in A_i as well as the "truncated V configurations" α_j, for $j \neq i$. Thus, there are at most $2^{f-k} + 2^k - 1$

[20] In contrast, the verifier constructed in the proof of Proposition 11.8, rejects in case of such a mismatch.

accepting configurations. Suppose $V^\pi(x)$ accepts with probability p, then V' accepts input x with oracle access to π with probability $(1-2^{-k}) + 2^{-k} \cdot p = 1 - \frac{1-p}{2^k}$. The proposition follows. ▌

Finally, we present a simplified version of the above transformation. Here the acceptance probabilities are increased without affecting the free-bit complexity (either way).

Proposition 11.10 (increasing acceptance probabilities while preserving free-bit complexity): Let $c, s, \delta : \mathcal{Z}^+ \to [0,1]$ be admissible functions and $r, f : \mathcal{Z}^+ \to \mathcal{Z}^+$. Then

$$\mathrm{FPCP}_{c,s}[r,f] \subseteq \mathrm{FPCP}_{c',s'}[r + \log(1/\delta), f]$$

where $c' = 1 - \delta \cdot (1-c)$ and $s' = 1 - \delta \cdot (1-s)$.

Proof: Let V be a verifier for $L \in \mathrm{FPCP}_{c,s}[r,f]$. We construct a new verifier, V', which with probability δ invokes V and otherwise accepts regardless of the input. The proposition follows. ▌

References

[Ad] L. Adleman. Two theorems on random polynomial time. *Proceedings of the* 19th *Symposium on Foundations of Computer Science*, IEEE, 1978, pp. 75–83.

[AKS] M. Ajtai, J. Komlos and E. Szemeredi. Deterministic Simulation in Logspace. *Proceedings of the* 19th *Annual Symposium on the Theory of Computing*, ACM, 1987, pp. 132–140.

[AFWZ] N. Alon, U. Feige, A. Wigderson, D. Zuckerman. Derandomized Graph Products. *Computational Complexity*, Vol. 5, No. 1, 1995, pp. 60–75.

[ASE] N. Alon, J. Spencer and P. Erdos. *The Probabilistic Method.* John Wiley and Sons, 1992.

[AmKa] E. Amaldi and V. Kann. The complexity and approximability of finding maximum feasible subsystems of linear relations. *Theoretical Computer Science*, Vol. 147, 1995, pp 181–210.

[Ar] S. Arora. Reductions, Codes, PCPs and Inapproximability. *Proceedings of the* 36th *Symposium on Foundations of Computer Science*, IEEE, 1995, pp. 404–413.

[ABSS] S. Arora, L. Babai, J. Stern and Z. Sweedyk. The hardness of approximate optima in lattices, codes and systems of linear equations. *Journal of Computer and System Sciences*, Vol. 54, No. 2, 1997, pp. 317–331.

[ALMSS] S. Arora, C. Lund, R. Motwani, M. Sudan and M. Szegedy. Proof verification and intractability of approximation problems. *Proceedings of the* 33rd *Symposium on Foundations of Computer Science*, IEEE, 1992, pp. 14–23.

[ArSa] S. Arora and S. Safra. Probabilistic checking of proofs: a new characterization of NP. *Proceedings of the* 33rd *Symposium on Foundations of Computer Science*, IEEE, 1992, pp. 2–13.

[Bab] L. Babai. Trading Group Theory for Randomness. *Proceedings of the* 17th *Annual Symposium on the Theory of Computing*, ACM, 1985, pp. 421–429.

[BFL] L. Babai, L. Fortnow and C. Lund. Non-deterministic Exponential time has two-prover interactive protocols. *Computational Complexity*, Vol. 1, 1991, pp. 3–40. (See also addendum in Vol. 2, 1992, pp. 374.)

[BFLS] L. Babai, L. Fortnow, L. Levin, and M. Szegedy. Checking computations in polylogarithmic time. *Proceedings of the* 23rd *Annual Symposium on the Theory of Computing*, ACM, 1991, pp. 21–31.

[BaEv1] R. Bar-Yehuda and S. Even. A linear time approximation algorithm for the weighted vertex cover problem. *Journal of Algorithms*, Vol. 2, 1981, pp. 198–201.

[BaEv2] R. Bar-Yehuda and S. Even. A local ratio theorem for approximating the weighted vertex cover problem. In *Analysis and Design of Algorithms for Combinatorial Problems*, Vol. 25 of Annals of Discrete Math, Elsevier, 1985.

[BaMo] R. Bar-Yehuda and S. Moran. On approximation problems related to the independent set and vertex cover problems. *Discrete Applied Mathematics*, Vol. 9, 1984, pp. 1–10.

[Be] M. Bellare. Interactive proofs and approximation: reductions from two provers in one round. *Proceedings of the Second Israel Symposium on Theory and Computing Systems*, IEEE, 1993, pp. 266–274.

[BCHKS] M. Bellare, D. Coppersmith, J. Håstad, M. Kiwi and M. Sudan. Linearity testing in characteristic two. *IEEE Transactions on Information Theory* Vol. 42, No. 6, November 1996, pp. 1781–1795.

[BGG] M. Bellare, O. Goldreich and S. Goldwasser. Randomness in interactive proofs. *Computational Complexity,* Vol. 3, No. 4, 1993, pp. 319–354.

[BGS1] M. Bellare, O. Goldreich and M. Sudan. Free Bits, PCPs and Non-Approximability — Towards Tight Results. Extended abstract of this paper, *Proceedings of the* 36th *Symposium on Foundations of Computer Science*, IEEE, 1995, pp. 422–431.

[BGS2] M. Bellare, O. Goldreich and M. Sudan. Free Free Bits, PCPs and Non-Approximability — Towards Tight Results. Preliminary versions of this paper. TR95-024 of ECCC, the *Electronic Colloquium on Computational Complexity.* May 1995 (revised Sept. 1995, Jan. 1996, Dec. 96). See http://www.eccc.uni-trier.de/eccc/.

[BGLR] M. Bellare, S. Goldwasser, C. Lund and A. Russell. Efficient probabilistically checkable proofs and applications to approximation. *Proceedings of the* 25th *Annual Symposium on the Theory of Computing*, ACM, 1993, pp. 294–304. (See also Errata sheet in *Proceedings of the* 26th *Annual Symposium on the Theory of Computing*, ACM, 1994, pp. 820–820).

[BeRo] M. Bellare and P. Rogaway. The complexity of approximating a quadratic program. *Journal of Mathematical Programming B*, Vol. 69, No. 3, September 1995, pp. 429–441. Also in *Complexity of Numerical Optimization*, Ed. P. M. Pardalos, World Scientific, 1993.

[BeSu] M. Bellare and M. Sudan. Improved non-approximability results. *Proceedings of the* 26th *Annual Symposium on the Theory of Computing*, ACM, 1994, pp. 184–193.

[BGKW] M. Ben-Or, S. Goldwasser, J. Kilian and A. Wigderson. Multi-Prover interactive proofs: How to remove intractability assumptions. *Proceedings of the* 20th *Annual Symposium on the Theory of Computing*, ACM, 1988, pp. 113–131.

[BeSc] P. Berman and G. Schnitger. On the complexity of approximating the independent set problem. *Information and Computation*, Vol. 96, 1992, pp. 77–94.

[Bl] A. Blum. Algorithms for approximate graph coloring. Ph. D Thesis, Dept. of Computer Science, MIT, 1991.

[BLR] M. Blum, M. Luby and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences*, Vol. 47, 1993, pp. 549–595.

[BoHa] R. Boppana and M. Haldórsson. Approximating maximum independent sets by excluding subgraphs. *BIT*, Vol. 32, No. 2, 1992.

[BrNa] J. Bruck and M. Naor. The hardness of decoding with preprocessing. *IEEE Transactions on Information Theory*, Vol. 36, No. 2, 1990, pp. 381–385.

[CW] A. Cohen and A. Wigderson. Dispersers, deterministic amplification, and weak random sources. *Proceedings of the* 30th *Symposium on Foundations of Computer Science*, IEEE, 1989, pp. 14–19.

[Con] A. Condon. The complexity of the max word problem and the power of one-way interactive proof systems. *Computational Complexity*, Vol. 3, 1993, pp. 292–305.

[Coo] S. Cook. The complexity of theorem-proving procedures. *Proceedings of the* 3rd *Annual Symposium on the Theory of Computing*, ACM, 1971, pp. 151–158.

[CrKa] P. Crescenzi and V. Kann. A compendium of NP optimization problems. Technical Report, Dipartimento di Scienze dell'Informazione, Università di Roma "La Sapienza", SI/RR-95/02, 1995. The list is updated continuously. The latest version is available via `http://www.nada.kth.se/~viggo/problemlist/compendium.html`.

[CST] P. Crescenzi, R. Silvestri and L. Trevisan. To weight or not to weight: where is the question? *Proceedings of the Fourth Israel Symposium on Theory and Computing Systems*, IEEE, 1996.

[EIS] S. Even, A. Itai and A. Shamir. On the complexity of timetable and multicommodity flow problems. *SIAM J. on Computing*, Vol. 5, 1976, pp. 691–703.

[ESY] S. Even, A. Selman and Y. Yacobi. The complexity of promise problems with applications to public-key cryptography. *Information and Control*, Vol. 2, 1984, 159–173.

[Fe1] U. Feige. Randomized graph products, chromatic numbers, and the Lovász theta function. *Proceedings of the* 27th *Annual Symposium on the Theory of Computing*, ACM, 1995, pp. 635–640.

[Fe2] U. Feige. Set Cover. A threshold of ln n for approximating set cover. In Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing, pages 314–318, 1996.

[FeGo] U. Feige and M. Goemans. Approximating the value of two prover proof systems, with application to Max-2SAT and Max-DICUT. *Proceedings of the Third Israel Symposium on Theory and Computing Systems*, IEEE, 1995, pp. 182–189.

[FGLSS] U. Feige, S. Goldwasser, L. Lovász, S. Safra, and M. Szegedy. Interactive proofs and the hardness of approximating cliques. *Journal of the ACM*, Vol. 43, No. 2, 1996, pp. 268–292.

[FeKi1] U. Feige and J. Kilian. Two prover protocols – Low error at affordable rates. *Proceedings of the* 26th *Annual Symposium on the Theory of Computing*, ACM, 1994, pp. 172–183.

[FeKi2] U. Feige and J. Kilian. Zero-knowledge and the chromatic number. *Proceedings of the* 11th *Annual Conference on Computational Complexity*, IEEE, 1996.

[FeLo] U. Feige and L. Lovász. Two-prover one round proof systems: Their power and their problems. *Proceedings of the* 24th *Annual Symposium on the Theory of Computing*, ACM, 1992, pp. 733-744.

[FRS] L. Fortnow, J. Rompel and M. Sipser. On the power of multiprover interactive protocols. *Theoretical Computer Science*, Vol. 134, No. 2, 1994, pp. 545–557.

[Fu] M. Fürer. Improved hardness results for approximating the chromatic number. *Proceedings of the* 36th *Symposium on Foundations of Computer Science*, IEEE, 1995, pp. 414–421.

[FGMSZ] M. Fürer, O. Goldreich, Y. Mansour, M. Sipser, and S. Zachos. On completeness and soundness in interactive proof systems. In *Advances in Computing Research: a research annual*, Vol. 5 (Randomness and Computation, S. Micali, ed.), 1989, pp. 429–442.

[GJ1] M. Garey and D. Johnson. The complexity of near optimal graph coloring. *Journal of the ACM*, Vol. 23, No. 1, 1976, pp. 43–49.

[GJ2] M. Garey and D. Johnson. *Computers and Intractability: A guide to the theory of NP-completeness*. W. H. Freeman and Company, 1979.

[GJS] M. Garey, D. Johnson and L. Stockmeyer. Some simplified NP-complete graph problems. *Theoretical Computer Science*, Vol. 1, 1976, pp. 237–267.

[GoWi1] M. Goemans and D. Williamson. New 3/4-approximation algorithms for the maximum satisfiablity problem. *SIAM Journal on Discrete Mathematics*, Vol. 7, No. 4, 1994, pp. 656–666.

[GoWi2] M. Goemans and D. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, Vol. 42, No. 6, 1995, pp. 1115–1145.

[GMW] O. Goldreich, S. Micali, and A. Wigderson. Proofs that yield nothing but their validity, or all languages in NP have zero-knowledge proof systems. *Journal of the ACM*, Vol. 38, No. 1, July 1991, pp. 691–729.

[GMR] S. Goldwasser, S. Micali, and C. Rackoff. The knowledge complexity of interactive proofs. *SIAM J. Computing*, Vol 18, No. 1, 1989, pp. 186–208.

[GS] S. Goldwasser and M. Sipser. Private coins versus public coins in interactive proof systems. *Proceedings of the* 18th *Annual Symposium on the Theory of Computing*, ACM, 1986, pp. 59–68.

[H1] J. Håstad. Testing of the long code and hardness for clique. *Proceedings of the* 28th *Annual Symposium on the Theory of Computing*, ACM, 1996, pp. 11–19.

[H2] J. Håstad. Clique is hard to approximate within $n^{1-\epsilon}$. *Proceedings of the* 37th *Symposium on Foundations of Computer Science*, IEEE, 1996, pp. 627–636.

[H3] J. Håstad. Getting optimal in-approximability results. *Proceedings of the* 29th *Annual Symposium on the Theory of Computing*, ACM, 1997, pp. 1–10.

[Hoc] D. Hochbaum. Efficient algorithms for the stable set, vertex cover and set packing problems. *Discrete Applied Mathematics*, Vol 6, 1983, pp. 243–254.

[ImZu] R. Impagliazzo and D. Zuckerman. How to recycle random bits. *Proceedings of the* 30th *Symposium on Foundations of Computer Science*, IEEE, 1989, pp. 248–253.

[JLL] N. Jones, Y. Lien and W. Laaser. New problems complete for non-deterministic log space. *Math. Systems Theory*, Vol. 10, 1976, pp. 1–17.

[Kah] N. Kahale. Eigenvalues and expansion of regular graphs. *Journal of the ACM*, Vol. 42, No. 5, 1995, pp. 1091–1106.

[KKLP] V. Kann, S. Khanna, J. Lagergren and A. Panconesi. On the hardness of approximating MAX k-CUT and its dual. Technical Report of the Department of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm, TRITA-NA-P9505, 1995.

[KMS] D. Karger, R. Motwani and M. Sudan. Approximate graph coloring by semidefinite programming. *Proceedings of the* 35th *Symposium on Foundations of Computer Science*, IEEE, 1994, pp. 2–13.

[KaZw] H. Karloff and U. Zwick. A 7/8-eps approximation algorithm for MAX 3SAT? To appear in *Proceedings of the* 38th *Symposium on Foundations of Computer Science*, IEEE, 1997.

[Kar] R. Karp. Reducibility among combinatorial problems. *Complexity of Computer Computations,* Miller and Thatcher (eds.), Plenum Press, New York, 1972.

[KLS] S. Khanna, N. Linial and S. Safra. On the hardness of approximating the chromatic number. *Proceedings of the Second Israel Symposium on Theory and Computing Systems*, IEEE, 1993, pp. 250–260.

[LaSh] D. Lapidot and A. Shamir. Fully parallelized multi-prover protocols for NEXP-time. *Proceedings of the* 32nd *Symposium on Foundations of Computer Science*, IEEE, 1991, pp. 13–18.

[Lau] C. Lautemann. BPP and the polynomial hierarchy. *Information Processing Letters*, Vol. 17, No. 4, 1983, pp. 215–217.

[Lev] L. Levin. Universal'nyĭe perebornyĭe zadachi (universal search problems : in russian). *Problemy Peredachi Informatsii*, Vol. 9, No. 3, 1973, pp. 265–266.

[LPS] A. Lubotzky, R. Phillips and P. Sarnak. Explicit Expanders and the Ramanujan Conjectures. *Proceedings of the* 18th *Annual Symposium on the Theory of Computing*, ACM, 1986, pp. 240–246.

[LuYa] C. Lund and M. Yannakakis. On the hardness of approximating minimization problems. *Journal of the ACM*, Vol. 41, No.5, 1994, pp. 960–981.

[LFKN] C. Lund, L. Fortnow, H. Karloff, and N. Nisan. Algebraic Methods for Interactive Proof Systems. *Journal of the ACM*, Vol. 39, No. 4, 1992, pp 859–868.

[McSl] F. MacWilliams and N. Sloane. *The theory of error-correcting codes.* North-Holland, 1981.

[MoSp] B. Monien and E. Speckenmeyer. Ramsey numbers and an approximation algorithm for the vertex cover problem. *Acta Informatica*, Vol. 22, No. 1, 1985, pp. 115–123.

[MoRa] R. Motwani and P. Raghavan. *Randomized algorithms.* Cambridge University Press, 1995.

[PaYa] C. Papadimitriou and M. Yannakakis. Optimization, approximation, and complexity classes. *Journal of Computer and System Sciences*, Vol. 43, 1991, pp. 425–440.

[Pet] E. Petrank. The Hardness of Approximations: Gap Location. TR–754, Department of Computer Science, Technion – Israel Institute of Technology, 1992.

[PoSp] A. Polishchuk and D. Spielman. Nearly-linear size holographic proofs. *Proceedings of the* 26th *Annual Symposium on the Theory of Computing*, ACM, 1994, pp. 194–203.

[Raz] R. Raz. A parallel repetition theorem. *Proceedings of the* 27th *Annual Symposium on the Theory of Computing*, ACM, 1995, pp. 447–456.

[SaGo] S. Sahni and T. Gonzales. P-complete approximation problems. *Journal of the ACM*, Vol. 23, 1976, pp. 555–565.

[Sha] A. Shamir. IP=PSPACE. *Journal of the ACM*, Vol. 39, No. 4, 1992, pp. 869–877.

[Tar] G. Tardos. Multi-prover encoding schemes and three prover proof systems. *Journal of Computer and System Sciences*, Vol. 53, No. 2, October 1996, pp. 251–260.

[Ta-S] A. Ta-Shma. A Note on PCP vs. MIP. *Information Processing Letters*, Vol. 58, No. 3, 1996, pp. 135–140.

[TSSW] L. Trevisan, G. Sorkin, M. Sudan and D. Williamson. Gadgets, approximation and linear programming. *Proceedings of the* 37th *Symposium on Foundations of Computer Science*, IEEE, 1996, pp. 617–626.

[Yan] M. Yannakakis, On the approximation of maximum satisfiability. *Journal of Algorithms*, Vol. 17, 1994, pp. 475–502.

[Zuc] D. Zuckerman. On unapproximable versions of NP-complete problems. *SIAM J. on Computing*, Vol. 25, No. 6, 1996, pp. 1293–1304.

A The coding theory bound

We provide here the coding theory bound used in the proof of Lemma 3.11. It is a slight extension of bounds in [McSl, Ch. 17] which consider only vectors of weight exactly w rather than at most w. For sake of completeness, we include a proof of this bound. In discussing binary vectors, the weight is the number of ones in the vector and the distance between two vectors is the number of places in which they disagree.

Lemma A.1 Let $B = B(n,d,w)$ be the maximum number of binary vectors of length n, each with weight at most w, and any two being distance at least d apart. Then $B \le (1-2\beta)/(4\alpha^2 - 2\beta)$, provided $\alpha^2 > \beta/2$, where $\alpha = (1/2) - (w/n)$ and $\beta = (1/2) - (d/n)$.

Proof: Consider an arbitrary sequence, $v_1, ..., v_M$, of n-vectors which are at mutual distance at least $n/2$. Let us denote by $v_{i,j}$ the j^{th} entry in the i^{th} vector, by w_i the weight of the i^{th} vector, and by $\overline{w}$ the average value of the w_i's. Define

$$S \stackrel{\text{def}}{=} \sum_{i=1}^{M} \sum_{j=1}^{M} \sum_{k=1}^{n} v_{i,k} v_{j,k}$$

Then, on one hand

$$\begin{aligned} S &= \sum_{i=1}^{M} \sum_{k=1}^{n} v_{i,k}^2 + \sum_{1 \le i \ne j \le M} \sum_{k=1}^{n} v_{i,k} v_{j,k} \\ &\le \sum_i w_i + \sum_{1 \le i \ne j \le M} \frac{w_i + w_j - d}{2} \\ &= M\overline{w} + M(M-1) \cdot (\overline{w} - (d/2)) \end{aligned}$$

where the inequality follows from observing that, for $i \ne j$,

$$\begin{aligned} w_i + w_j &= 2|\{k : v_{i,k} = v_{j,k} = 1\}| + |\{k : v_{i,k} \ne v_{j,k}\}| \\ &\ge 2 \sum_{k=1}^{n} v_{i,k} v_{j,k} + d \end{aligned}$$

On the other hand $S = \sum_{k=1}^{n} |\{i : v_{i,k} = 1\}|^2$. This allows to lower bound S by the minimum of $\sum_k x_k^2$ subject to $\sum_k x_k = M\overline{w}$. The minimum is obtained when all x_k's are equal and yields

$$S \ge n \cdot \left(\frac{M\overline{w}}{n} \right)^2$$

Confronting the two bounds, we get

$$\frac{M \cdot \overline{w}^2}{n} \le M \cdot \overline{w} - (M-1) \cdot (d/2)$$

which yields $(\frac{\overline{w}^2}{n} - \overline{w} + \frac{d}{2})M \le \frac{d}{2}$. Letting $\overline{\alpha} = (1/2) - (\overline{w}/n)$ and using $\overline{\alpha}^2 \ge \alpha^2 > \beta/2$, we get

$$M \le \frac{1 - 2\beta}{4\overline{\alpha}^2 - 2\beta}$$

and the lemma follows by observing that the bound maximizes when $\alpha = \overline{\alpha}$. ▌