Electronic Colloquium on Computational Complexity, Report No. 169 (2023)

Extracting Randomness from Samplable Distributions, Revisited

Marshall Ball*, Dana Dachman-Soled**, Eli Goldin***, and Saachi Mutreja†

Abstract. Randomness extractors provide a generic way of converting sources of randomness that are merely unpredictable into almost uniformly random bits. While in general, deterministic randomness extraction is impossible, it is possible if the source has some structural constraints.

While much of the literature on deterministic extraction has focused on sources with strong independence properties, a natural class where deterministic extraction is possible is sources that can sampled by a polynomial size circuit, Levin [SIAM J Comp'86]. Trevisan and Vadhan [FOCS'00] explicitly constructed deterministic randomness extractors for this class of sources, assuming very strong circuit lower bounds.

We suggest that there is perhaps an even more reasonable model of natural sources of randomness than Levin's: sources sampled by polynomial size quantum circuits. Under a suitable circuit lower bound, we show that Trevisan and Vadhan's extractor indeed works for this class.

Along the way, we substantially improve their analysis in the classical case, showing that a circuit lower bound against NP-circuits suffice in the classical case (as opposed to a lower bounds on Σ_5-circuits, as shown by Trevisan and Vadhan). Moreover, we show that under this assumption, it is possible to handle sources sampled by postselecting circuits (a variant of nondeterministic circuits). We show that this model is sufficient to capture randomness extraction in the presence of efficiently computable leakage.

1 Introduction

Randomness is an essential resource in computing. It is necessary for nearly all cryptographic tasks, such as achieving semantically secure symmetric key encryption. Similarly, many fundamental tasks in the domain of distributed computing, such as byzantine agreement or testing equality of two strings with low communication, are impossible to achieve deterministically. In a similar vein, certain tasks in differential privacy are also impossible without randomness.

Yet, where do these random bits come from? When constructing randomized protocols or procedures, the protocol/procedure designer almost always assumes access to independent, unbiased random bits. However, natural sources of randomness available to our machines are almost invariably far from such idealized sources of randomness, and moreover the particulars of their distributions are unknown to us. The question then becomes what algorithmic tasks can be accomplished with access to weakly random sources. Randomness extractors provide a generic means of deterministically converting a weakly random source into (almost) uniformly random independent bits, so that we may use constructions in our idealized models. This motivates the following general question:

> Can we deterministically extract uniformly random bits from naturally occurring weakly random sources?

It is well known that deterministic extraction from arbitrary weakly random sources is impossible, but is possible if the sources have some structure. While one rich line of work on deterministic (or seedless) randomness extractors has studied sources with strong independence properties, it is

* New York University. marshall.ball@cs.nyu.edu

** University of Maryland. danadach@umd.edu. Dana Dachman-Soled is supported in part by NSF grants CNS-2154705 and CNS-1933033.

*** New York University. eli.goldin@nyu.edu

† Columbia University. saachi@berkeley.edu

ISSN 1433-8092

unclear if naturally occurring sources can be assumed to be independent.[1] The extended Church-Turing thesis motivates another model for naturally occurring entropic sources: *sources sampled by polynomial size circuits*. [16] Indeed, Trevisan and Vadhan constructed efficient extractors for such classes. [26]

The starting point of the present work is the simple observation that the universe, and hence natural sources, are generated by quantum phenomena. And even if quantum computing never materializes in practice, it is quite plausible that local natural physical phenomena cannot be efficiently simulated by classical circuits, but can be efficiently simulated by *theoretical* quantum circuits.

1.1 Our Results

We demonstrate that it is possible to deterministically (classically) extract random bits from weakly random sources sampled by polynomial size quantum circuits, assuming lower bounds on quantum circuits with postselection, a nondeterministic analog of quantum circuits. In fact, we show it is possible to deterministically extract random bits from a significantly larger, nondeterministic class of sources. Importantly, this implies the ability to extract randomness after observing arbitrary efficiently computed leakage on the source (provided some entropy remains after seeing the leakage).

Additionally, we improve what is known in the classical case. We show that it is possible to extract randomness from sources that are samplable by polynomial size circuits assuming lower bounds on nondeterministic circuits (as opposed to circuits with gates computing Σ_5-complete problems, as is the case in [26]).[2] Again, we show that indeed it is possible to extract from a larger class that includes sources uniform over a set recognized by a polynomial size circuit (also known as recognizable sources [21]), from the same assumption. Prior to our work, similar results were only known assuming lower bounds on Σ_3-circuits. [3][3]

In all cases, we can extract almost all the randomness from sources with linear min-entropy (there exists $\gamma > 0$ such that for all x, $\Pr[X = x] \leq 2^{-(1-\gamma)n}$, where n is the length of the source). Unfortunately, like all prior work, the output of our extractors is only inverse-polynomially close to uniform.[4]

Classical and quantum postselecting samplers. We consider a notion of nondeterministic samplers that generalizes samplable sources. We say that a source X (supported on $\{0,1\}^n$) is *sampled by a postselecting circuit* C (in a class $\mathcal{C}$), if C outputs $n+1$ bits, $C \to (x, b)$ such that X is identically distributed to the distribution sampled by C, conditioned on $b = 1$, namely $\Pr[X = x'] = \Pr_{C \to (x,b)}[x = x' | b = 1]$ for all x'. In particular, we are concerned with the case that the class $\mathcal{C}$ is either (randomized) polynomial size classical circuits, in which case we say the source is samplable by postselecting circuits, or $\mathcal{C}$ is polynomial size quantum circuits (with sufficient min-entropy), in which case we say the source is samplable by postselecting quantum circuits.

Note that sources sampled by postselecting classical polynomial size circuits correspond to sources sampled by polynomial size circuits whose random bits may themselves be drawn uniformly

[1] Moreover, even very limited quantitative relaxations of independence quickly render extraction impossible. [8,4]

[2] If the samplable source has very high min entropy, $n - O(\log n)$, then it was known how to extract from hardness against non-deterministic circuits. [26]

[3] Again, if the recognizable source was known to have very high min-entropy, $n - O(\log n)$ it is known how to extract from lower bounds on *deterministic* circuits. [17].

[4] Applebaum et al. showed that this inherent in all black-box nondeterministic reductions. [3]

from a set recognized by a polynomial size circuit. Clearly, this class generalizes both samplable and recognizable sources.[5]

A motivation for considering such classes of sources is that they capture samplable sources induced by external observation or side-channel leakage. For example, it is unlikely that a physical source exists in a vacuum and is only observed by the extractor itself. If the extractor works for nondetermnistic samplable sources, then so long as the source has enough conditional min-entropy, then the output of the extractor will be independent of the leakage (and safe to use in a sensitive task).

Nondeterministic circuit models and hardness. Before stating our results, we must briefly describe the circuit classes we assume hardness against.

A classical nondeterministic circuit, C can be thought of as a deterministic circuit C' that takes input, x and a witness, w: for any input x, $C(x) = 1$ if and only there exists w such that $C'(x, w) = 1$.

In the quantum regime, we are concerned with a fairly strong analog of nondeterminism: quantum circuits with postselection [1]. These are quantum circuits (with classical description) that can *condition on a measurement being 1* before the output is measured. We say that such a circuit decides a language if such a circuit (conditioned on the first measurement outcome being 1) disagrees with the language on any input x with probability at most $1/3$.

Both of these circuit classes are quite strong. In particular, uniform polytime quantum computation with postselection, PostBQP is known to be equivalent to PP [1]. However, it is nonetheless reasonable to conjecture that there are classical deterministic computations which do not admit superpolynomial speedups even if the computation is both non-uniform and nondeterministic or non-uniform, quantum, and postselecting.

The former classical assumption has been considered before in the context of derandomizing AM [18]. We are not aware of a situation where the latter assumption has been made, but the connection with PP gives a classical interpretation: a set admits a postselecting quantum circuit family if and only if there is a family of randomized classical circuits that accept every string in the language with probability strictly greater than $1/2$, and reject every string not in the set with probability at least $1/2$.

Main Theorems. Now we can (informally) state our results. Our main classical result is the following:

Informal Theorem 1 (Extractors for Classical Sources (Theorem 2)) *If there is a problem in $E = DTIME(2^{O(n)})$ with nondeterministic circuit complexity $2^{\Omega(n)}$, then for any constant c, there is an explicit deterministic extractor for sources samplableable by size n^c postselecting circuits with linear min-entropy (whose output is $1/\mathrm{poly}(n)$-close to uniform).*

Our main quantum theorem is the following:

Informal Theorem 2 (Extractors for Quantum Sources (Theorem 3)) *If there is a problem in $E = DTIME(2^{O(n)})$ with postselecting quantum circuit complexity $2^{\Omega(n)}$, then for any constant c, there is an explicit deterministic extractor for sources samplable by size n^c postselecting quantum circuits with linear min-entropy (whose output is $1/\mathrm{poly}(n)$-close to uniform).*

[5] Guo et al. [14] consider a similar analog in the algebraic setting: sources sampled by polynomials evaluated on varieties (generalizing polynomial sources [12] and variety sources [11].

We remark that regardless of whether this strong hardness assumption is true, explicit hard functions for postselecting quantum circuits are *required* to extract from this source class.

In both cases, our extractor is essentially the same extractor as that of Trevisan and Vadhan. [26] If f is an E-complete problem, $\widetilde{f}$ is its low degree extension, and $2Ext$ is a sufficiently good two-source extractor, our extractor will simply be $Ext(x,i) = 2Ext(\widetilde{f}(x), i)$.

Where our result differs from Trevisan and Vadhan's is that we give a novel analysis of the extractor. At the core of our analysis are new nondeterministic algorithms for an optimal parameter agnostic learning problem we call *gap probability maximization*. We refer the reader to the detailed technical overview below for details.

2 Detailed Technical Overview

In this section, we explain in detail our approach to lifting Trevisan and Vadhan's proof that a hard function for Σ_5-circuits gives a good extractor for samplable sources [26] to the quantum realm. Through this explanation, it will become clear how we extend this result to nondeterministically samplable sources, as well as how we reduce the Σ_5 hardness requirement all the way to Σ_1.

As a warmup, we will describe how to lift Trevisan and Vadhan's proof that a boolean function f hard *on average* for NP-circuits is itself a good extractor [26]. The classical argument for this goes as follows: Let $\mathcal{S}$ be a flat (i.e. all outputs in the support have equal probability) source biasing f to 1. Then the following NP-circuit can compute $f(x)$:

On input x, nondeterministically check if x is in the range of $\mathcal{S}$, and if so output 1. Otherwise, output a random bit.

This approach can be augmented to non-flat $\mathcal{S}$ as long as we can solve the probability estimation problem, which asks that given a randomized circuit C and an output x, compute $\Pr_r[C(r) \to x]$ up to $(1 \pm \epsilon)$ multiplicative error. However, it is known that NP-circuits can solve the probability estimation problem in size polynomial in $size(C)$. This means that if f is hard for NP-circuits of size s, then it is an extractor for sources samplable by $s - O(n)$-size circuits for some concrete polynomial.

To extend this argument to the quantum world, all that is necessary is that we be able to do quantum probability estimation. That is, we need some model that can solve the following problem: given a quantum circuit C and an output x, compute $\Pr_r[C(r) \to x]$ up to $(1 \pm \epsilon)$ multiplicative error.

It turns out that to solve this problem for quantum circuits, we require quantum circuits with *postselection* (for more details on this equivalence see Section B). Postselection refers to the ability for algorithms to conditionally sample. In the quantum setting, this refers to the ability for quantum algorithms to produce the residual state resulting from measuring in the standard basis and receiving result 1. Thus, a postselecting circuit is a quantum circuit with the additional ability to postselect.

Quantum circuits with postselection are considered in depth by Aaronson in [1]. It is not known how to implement postselection with a quantum computer, but it does not directly contradict the laws of quantum mechanics. In particular, Aaronson shows that PostBQP, the class of uniform postselecting circuits, is equivalent to PP.

Solving probability estimation using postselecting circuits implies that if a function is hard on average for quantum circuits with postselection, then it is an extractor for quantum samplable sources. However, it would be better to be able to show extraction from a worst-case assumption. In fact, Trevisan and Vadhan were able to extend their average case classical result to a worst-case hardness assumption, resulting in the following-theorem:

Theorem 1. *If there is a problem in $E = DTIME(2^{O(n)})$ with Σ_5-circuit complexity $2^{\Omega(n)}$ for all n, then there is a constant $\delta > 0$ such that for all n, there is a $((1-\delta)n, 1/n)$-deterministic extractor $\mathrm{Ext}_{n,s} : \{0,1\}^n \to \{0,1\}^{(1-O(\delta))n}$ against circuit-size n^c such that $\mathrm{Ext}_{n,s}$ is computable in time $\mathrm{poly}(n^c)$ (with exponent depending on δ).*

As postselecting hardness was enough to lift the average case variation of this theorem to the quantum setting, we postulate (and will later prove) the following quantumization:

Proposition 1. *If there is a problem in $E = DTIME(2^{O(n)})$ with postselecting quantum circuit complexity $2^{\Omega(n)}$, then there is a constant $\delta > 0$ such that for all n, there is a $((1-\delta)n, 1/n)$-deterministic extractor $\mathrm{Ext}_{n,s} : \{0,1\}^n \to \{0,1\}^{(1-O(\delta))n}$ against circuit-size n^c such that $\mathrm{Ext}_{n,s}$ is computable in time $\mathrm{poly}(n^c)$ (with exponent depending on δ).*

One would hope that the same technique as in the average case hardness argument would apply when quantumizing this result. However, the proof for Theorem 1 relies strongly on the fact that Σ_i circuits can do probability estimation for Σ_{i-1} circuits. In fact, the proof involves several instances of such "ladder-climbing", accomplishing some task on Σ_{i-1} circuits using Σ_i circuits.

It is not clear how one would do "ladder-climbing" for postselecting quantum circuits. One may hope that postselecting quantum circuits themselves can accomplish tasks like probability estimation for postselecting quantum circuits. Unfortunately, this seems unlikely to be true. We note that since $PostBQP = PP$, $PH \subseteq P^{PostBQP} = P^{\#P}$. This doesn't say anything definitive, but it is not clear how to reduce adaptive counting queries to a single threshold query. To provide more concrete evidence, we show in Sections 4.2 and 4.3 a concrete problem on quantum circuits, solvable by postselecting quantum circuits, for which the natural approach will not extend to a solution for postselecting quantum circuits.

One may wonder whether "ladder-climbing" is necessary to construct extractors from worst-case hardness assumptions. We show that it is not necessary, giving us the following improvement to Trevisan and Vadhan's classical result.

Proposition 2. *If there is a problem in $E = DTIME(2^{O(n)})$ with $NP_{||}$-circuit complexity $2^{\Omega(n)}$ for all n, then there is a constant $\delta > 0$ such that for all n, there is a $((1-\delta)n, 1/n)$-deterministic extractor $\mathrm{Ext}_{n,s} : \{0,1\}^n \to \{0,1\}^{(1-O(\delta))n}$ against circuit-size n^c such that $\mathrm{Ext}_{n,s}$ is computable in time $\mathrm{poly}(n^c)$ (with exponent depending on δ).*

Our proof for this improvement will indeed lift easily to the quantum setting, allowing us to prove Proposition 1. In fact, in our main body we will prove both theorems simultaneously.

Note that Theorem 1 has a nondeterminism gap. That is, we require hardness against Σ_5-circuits to obtain extractors for deterministic sources.

In fact, a small modification to our proof technique improves the result so that it provides extractors for *postselecting* (classically) samplable sources, and this improvement trivially lifts to the quantum setting. Informally, we call a source $\mathcal{S}$ a postselecting samplable source if there exists

a circuit C outputting x, b such that $\mathcal{S}$ is the distribution on x conditioned on $b = 1$. That is, if $b = f(r)$ for some efficient f, we give the circuit the ability to uniformly sample from $f^{-1}(1)$. In general, this task can be implemented by an NP-circuit [15,5], and so this class of sources is slightly weaker than those samplable by NP-circuits.[6]

We observe that both samplable and recongnizable sources are samplable by postselecting circuits. Any sampling circuit C gives a postselecting sampling circuit C' by setting $C'(r) = (C(r), 1)$. Any recognizing circuit C gives a postselecting sampling circuit C' by setting $C'(r) = (r, C(r))$.

Formally, our main results are captured by the following theorems:

Theorem 2. *If there is a problem in $E = DTIME(2^{O(n)})$ with $NP_{||}$-circuit complexity $2^{\Omega(n)}$ for all n, then there is a constant $\delta > 0$ such that for all n, there is a $((1-\delta)n, 1/n)$-deterministic extractor $\mathrm{Ext}_{n,s} : \{0,1\}^n \to \{0,1\}^{(1-O(\delta))n}$ against postselecting circuit-size n^c such that $\mathrm{Ext}_{n,s}$ is computable in time $\mathrm{poly}(n^c)$ (with exponent depending on δ).*

Moreover, because our reduction makes nonadaptive queries (and hence we only need to assume that E is hard for exponential size circuits that make nonadaptive NP queries), we need only assume hardness against nondeterministic circuits by using a collapse theorem due to Shaltiel and Umans [23].

Corollary 1. *If there is a problem in $E = DTIME(2^{O(n)})$ with nondeterministic-circuit complexity $2^{\Omega(n)}$ for all n, then there is a constant $\delta > 0$ such that for all n, there is a $((1-\delta)n, 1/n)$-deterministic extractor $\mathrm{Ext}_{n,s} : \{0,1\}^n \to \{0,1\}^{(1-O(\delta))n}$ against postselecting circuit-size n^c such that $\mathrm{Ext}_{n,s}$ is computable in time $\mathrm{poly}(n^c)$ (with exponent depending on δ).*

Theorem 3. *If there is a problem in $E = DTIME(2^{O(n)})$ with postselecting quantum circuit complexity $2^{\Omega(n)}$ for all n, then there is a constant $\delta > 0$ such that for all n, there is a $((1-\delta)n, 1/n)$-deterministic extractor $\mathrm{Ext}_{n,s} : \{0,1\}^n \to \{0,1\}^{(1-O(\delta))n}$ against postselecting quantum circuit-size n^c such that $\mathrm{Ext}_{n,s}$ is computable in time $\mathrm{poly}(n^c)$ (with exponent depending on δ).*

[1] shows that a PP oracle can simulate postselecting quantum computation. Thus, we get the following corollary

Corollary 2. *If there is a problem in $E = DTIME(2^{O(n)})$ with PP-circuit complexity $2^{\Omega(n)}$ for all n, then there is a constant $\delta > 0$ such that for all n, there is a $((1-\delta)n, 1/n)$-deterministic extractor $\mathrm{Ext}_{n,s} : \{0,1\}^n \to \{0,1\}^{(1-O(\delta))n}$ against postselecting quantum circuit-size n^c such that $\mathrm{Ext}_{n,s}$ is computable in time $\mathrm{poly}(n^c)$ (with exponent depending on δ).*

2.1 Classical extractors from ladder-climbing

For the beginning of this overview, we only discuss extractors with one bit output. The extension to multi-bit output is discussed in Section 2.5

To generate an extractor from worst-case hardness, Trevisan and Vadhan rely on two ladder climbing techniques:

[6] In fact, it turns out that our notion of postselecting samplable sources is more powerful than sources samplable by "single-valued nondeterministic circuits," [19,22] a generalization of NP $\cap$ coNP to (a) computing functions, and (b) the non-uniform setting (which we won't formally define here). A simple rejection sampling argument implies that any extractor for the class of sources samplable by single-valued nondeterministic sources must be a hard to compute function for this class. It follows that hardness for this computational is indeed necessary in order to extract from postselecting samplable sources. Moreover, E being hard for exponential size nondeterministic circuits is in fact equivalent to E being hard for exponential size single-valued nondeterministic circuits. [22,2]

1. Probability estimation: Given a Σ_{i-1}-circuit C, there is a Σ_i-circuit estimating $\Pr_r[C(r) \to x]$ to a $(1 \pm \epsilon)$ multiplicative factor.
2. Uniform sampling: Given a boolean Σ_{i-1}-circuit C, there is a Σ_i-circuit sampling uniformly from $C^{-1}(1)$.

These ladder climbing techniques can prove the following two claims:

1. There is a very good worst-to-average case reduction for polynomial evaluation stepping up the ladder.
2. For functions $E(\cdot,\cdot)$ satisfying a "combinatorial list-decoding" property, there is a very efficient way to find points biasing $E(\cdot, \mathcal{S})$ for samplable $\mathcal{S}$

Stated more formally,

1. For a degree d polynomial $p : \mathbb{F}^t \to \mathbb{F}$, given a size-$s$ Σ_i which computes p correctly on a $c\sqrt{d/|\mathbb{F}|}$ fraction of points, there is a size-$poly(s)$ Σ_{i+2} circuit which computes p everywhere.
2. There exists a probabilistic Σ_{i+2}-circuit $DECODE$ of polynomial size such that the following holds: Let $\mathcal{S}$ be a source of density δ samplable by size-s Σ_i-circuits. Let $E(\cdot,\cdot)$ be a boolean function computable by size-$poly(n)$ circuits satisfying combinatorial list-decoding. If $E(w, C(r))$ is ϵ-biased to 1, then $C(\mathcal{S}, \epsilon) = w$ with probability $\Omega(\delta\epsilon^2)$

Trevisan and Vadhan use these claims to show that if you have a samplable distribution (X, I) of density δ which biases $E(p(x), i)$, then there is a Σ_5-circuit computing $p(x)$ everywhere. The approach here is simple: if I_x is the distribution I conditioned on $X = x$, then $DECODE(I_x)$ is a Σ_3-circuit computing $p(x)$ with some small probability. Then, the very efficient worst-to-average case reduction gives a Σ_5-circuit computing $p(x)$ everywhere. This immediately gives that $EXT(x, i) := E(p(x), i)$ is a good 1-bit extractor. Some additional care (but no further levels of nondeterminism) are required to extend this proof to multi-bit outputs.

For the proof of both of these claims, the two ladder climbing techniques described at the beginning of this section are used in sequence. For expository purposes, we will sketch the proof of the worst to average reduction for polynomial decoding.

2.2 Strong worst to average case reductions from ladder climbing

The proof they use for this claim relies on the following lemma from [25].

Lemma 1. *Let C be any function and let $L_{z,x} := (1-t)z + x$ denote the line between z and x. Then there exists a $z \in \mathbb{F}$, $\gamma > 0$, such that for $15/16$ of the values of x,*
- $\Pr_{\mathbb{F} \to u}[p(L_{z,x}(u)) = C(L_{z,x}(u))] \geq \gamma$
-For all univariate degree d $h : \mathbb{F} \to \mathbb{F}$ such that $h \neq p \circ L_{z,x}$, either $h(0) \neq z$ or $\Pr_{\mathbb{F} \to u}[h(u) = C(L_{z,x}(u))] \leq \frac{\gamma}{2}$.

Let C be a Σ_i circuit evaluating $p(x)$ on a $c\sqrt{d/|\mathbb{F}|}$ fraction of points.

We define C' to be the circuit which takes in a univariate h, chooses a random u from $\mathbb{F}$, and outputs 1 if $h(0) = z$ and $h(u) = C(L_{z,x}(u))$.

We further define C'' to be the Σ_{i+1}-circuit which takes an input x, gets an estimate $\widetilde{\gamma}$ for $\Pr[C'(x) = 1]$, and outputs 1 if $\widetilde{\gamma} \geq \frac{3}{4}\gamma$. The key lemma immediately shows that the only input accepted by C'' is $p \circ L_{z,x}$

It is then clear that running uniform sampling on C'' will find $p \circ L_{z,x}$ with high probability, and so outputting $(p \circ L_{z,x})(1)$ will find $p(x)$ with high probability.

2.3 Our techniques

Our key observation comes from the fact that the purpose of running uniform sampling and probability estimation in sequence is to solve a task we call the gap probability maximization problem. We define this problem as follows:

Say we are given a boolean randomized algorithm $\widetilde{C}$ and a constant γ with the following promise:

1. There exists some x^* such that $\Pr[\widetilde{C}(x^*) \to 1] \geq \gamma$
2. For all $x \neq x^*$, $\Pr[\widetilde{C}(x) \to 1] \leq \frac{\gamma}{2}$

The GPM problem asks us to find x^*.

We show in Section 4 that the gap maximization problem can be solved for an input circuit $\widetilde{C}$ by an NP-circuit C. Moreover, this circuit C only needs non-adaptive calls to the NP gates. This means that one only needs to step up the hierarchy once in order to achieve highly efficient worst to average case reductions for polynomial decoding, as well as bias finding for codes satisfying combinatorial list decoding. Thus, this observation immediately reduces our hardness assumption to hardness for Σ_3-circuits.

To get us all the way down to nondeterministic circuits, we note that gap probability maximization can be used to directly compute $p(x)$ using a source which biases our extractor. Once we have trimmed the layers of nondeterminism produced by stacking uniform sampling and probability estimation, the extra layers of nondeterminism are purely an artifact of modularity. This gets us to NP-circuits that use their NP gates non-adaptively. From there, we can apply a result of Shaltiel and Umans [24] that implies that if E is hard for exponential size nondeterministic circuits, then E is hard for exponential size non-adaptive NP-circuits.

The essential ingredient for combining the two claims used by Trevisan and Vadhan is the following improved key lemma

Lemma 2. *Let $\mathbb{F} = GF(2^{n'})$ be a field of size $q = 2^{n'}$. Let $p : \mathbb{F}^t \to \mathbb{F}$ be any degree d polynomial, let $E : \mathbb{F} \times \{0,1\}^{n'} \to \{0,1\}^m$ satisfy $(2^{-4m}, 2^{-0.1n'}, 2^{0.2n'})$-combinatorial list decoding, and let $\mathcal{S}$ be any distribution of density δ such that*

$$\left| \Pr_{\mathcal{S} \to (u,i)}[E(p(u), i) = 1] - \frac{1}{2^m} \right| \geq \frac{\epsilon}{2^m}.$$

For any ϵ, δ satisfying

$$\frac{d}{q} \leq \frac{\epsilon^2 \delta^2}{64 \cdot 2^{0.4n'+2m}}$$
$$\epsilon\delta \geq 128 \cdot 2^{m-0.1n'}$$

the following holds: There exists a z such that for $\frac{15}{16}$ values of x,
- $\left| \Pr_{S \to (u,i)}[E(p(u), i) = 1 | u \in L_{z,x}] - \frac{1}{2^m} \right| \geq \frac{\epsilon}{3 \cdot 2^m}$
- *For all $h : \mathbb{F} \to \mathbb{F}$ such that $h \neq p \circ L_{z,x}$, either* $\left| \Pr_{S \to (u,i)}[E(h(L_{z,x}^{-1}(u)), i) = 1 | u \in L_{z,x}] - \frac{1}{2^m} \right| \leq \frac{\epsilon}{6 \cdot 2^m}$ *or $h(0) \neq p(z)$.*

Once this lemma has been proved, a gap probability maximization solver immediately gets us the result. Let $\mathcal{S}$ be some distribution of density δ biasing $EXT(x, i) = E(p(x), i)$. Our algorithm for evaluating $p(x)$ operates as follows:

Algorithm 1 GPM evaluator for $p(x)$

We will define $C'(h)$ as follows:
  If $h(0) \neq p(z)$, output 0.
  Sample $\mathcal{S} \to (u, i)$.
  Say the test passes if u lies on the line and $E(h(L_{z,x}^{-1}(u)), i) = 1$:
  Output 1 if and only if the test passes $O(1/\epsilon)$ times.
    ▷ The key lemma tells us that we can run gap probability maximization on C' to compute $p \circ L_{z,x}$.
Output $(p \circ L_{z,x})(1) = p(x)$.

We remark that the proof of this key lemma is highly non-trivial. Full details are included in Section 5.1. Roughly, the first half of this lemma comes from a double application of Chebyshev's inequality. The second half of the lemma comes from proving a list decoding property, which implies that the number of univariate polynomials which bias $E(h(L(u)), i)$ on a random line L is small. We give a more thorough intuition for the proof of the second half of this lemma in Section 5.3.

To lift this result, all we need to do is show that gap probability maximization can be solved for quantum circuits using postselecting quantum circuits. We show this in Section 4.2.

2.4 Improvement to postselecting samplers

We remark that it is easy to extend our result to postselecting samplers. We simply replace the line "sample $\mathcal{S} \to (u, i)$" with "sample $\mathcal{S} \to (u, i, b)$ and fail if $b = 0$". The idea here is that nondeterminism allows us to condition our distribution for free, and so it costs nothing to add in the additional condition stemming from using a postselecting sampler.

2.5 Multi-bit output

To extend our result to multi-bit output, we use the same argument as Trevisan and Vadhan. In particular, it is not hard to extend from 1-bit output to logarithmic length output, and our

formal proof will go straight to logarithmic output. Once we have a deterministic extractor with logarithmic length output, we can use this extractor to create a logarithmic length seed for a seeded extractor (defined formally in Section 3.5). That is, our final multi-bit extractor will be as follows

$$\mathrm{Ext}(a,(x,i)) = \mathrm{Ext}_1(a, E(p(x),i))$$

where Ext_1 is a seeded extractor with appropriate parameters.

2.6 Leakage Resilience

Finally, we consider a notion of leakage-resilient extractors against samplable sources. Informally, leakage resilience requires that the output of the extractor remain close to uniform even when some side information about the underlying source is revealed. Note that arbitrary leakage resilience is impossible, as the leakage could be the output of the extractor itself. However, as our goal is to model physical extractors, it is natural to consider leakage which is itself samplable.

It is natural to consider leakage any samplable function of the randomness source. However, we choose to consider a stronger notion, where the leakage is provided by the sampling circuit itself. This way, the leakage can depend on the randomness used to generate the source (including in the quantum setting). Note that providing resilience against arbitrary length leakage, even in this restricted model, is impossible as the leakage may simply be the underlying randomness used to generate the source. Thus, we additionally require that the source have high min-entropy conditioned on its leakage.

We show that given a deterministic extractor against nondeterministic samplable sources, either classical or quantum, then we get a leakage-resilient deterministic extractor for free. To show this, we rely on the fact that our extractor works against the source defined by conditioning the original source on the leakage being any particular value.

Theorem 4. *Let* Ext *be a* (k,ϵ)*-deterministic extractor against nondeterministic sources samplable by size-*s *(quantum) circuits. Then, for all* $c > 0$*,* Ext *is a leakage-resilient* $(k+c, \epsilon + 2^{-c})$*-deterministic extractor against sources samplable by size-*$O(s)$ *postselecting (quantum) circuits.*

We remark that in our notion of leakage-resilience, we only consider *classical* leakage. In the quantum setting, there is also a notion of min-entropy, which was defined originally in [20]. This definition has been previously been used to capture randomness extraction [7,9,6] in the presence of *quantum* side-information.

We do not make any claims about quantum leakage-resilience. If quantum computers do not exist, the power of quantum computing must come only from the real world. Therefore, any adversary wishing to use quantum side information to distinguish the output of an extractor from random must first make some efficient measurement, which equivalently could be made by the source directly. Nevertheless, constructing deterministic extractors against quantum samplable sources secure even in the presence of quantum side information is an interesting open question.

3 Preliminaries

3.1 Notation

Throughout this paper, we often consider distributions $\mathcal{S}$ over product spaces $\mathcal{A} \times \mathcal{B}$. For some subset $X \subseteq \mathcal{A}$, we define $\mathcal{S}|_X$ to be $\mathcal{S}$ conditioned on the first output being in X. Formally, for all

$(a,b) \in \mathcal{A} \times \mathcal{B}$,

$$\Pr_{\mathcal{S}|_X \to (a',b')}[(a',b') = (a,b)] = \Pr_{\mathcal{S} \to (a,b)}[(a',b') = (a,b) | a' \in X]$$

3.2 Types of Nondeterministic Circuits

Let $\mathcal{P}$ be any complexity class and fix some $\mathcal{P}$-complete problem $\pi_{\mathcal{P}}$. A $\mathcal{P}$-circuit is a circuit with access to oracle gates for $\pi_{\mathcal{P}}$. We will primarily be concerned with Σ_i circuits. We will refer to Σ_1-circuits primarily as NP-circuits. The class of circuits that has all its Σ_1 gates in the same layer, i.e. circuits making SAT queries non-adaptively, is referred to as $NP_{||}$-circuits.

We rely on a collapse theorem for E due to Shaltiel and Umans [24] of which the following is a special case:

Theorem 5 (Corollary of [24, Theorem 3.2]). *If every language in E has $NP_{||}$-circuits of size $s(n)$, then every language in E has non-deterministic circuits of size $s(n)^{O(1)}$.*

A quantum circuit of size s is a sequence of unitaries $U_1, \ldots, U_s$ where each U_i is taken from some universal gate set. A quantum circuit has an input register of length m, $\ell \leq s$ ancilla qubits, and an output register of length n. We use $C(x)$ to refer to the distribution on the output register of $(U_s \ldots U_1)(|x\rangle \otimes |0\rangle^{\otimes \ell} \otimes |0\rangle^{\otimes n})$ after measuring in the standard basis.

We will use the following formulation of postselecting quantum circuits. A postselecting quantum circuit C has the same format as a quantum circuit, except it has an additional postselection register. We use $C(x)$ to refer to the distribution on the output register of $(U_s \ldots U_1)(|x\rangle \otimes |0\rangle^{\otimes \ell} \otimes |0\rangle^{\otimes n})$ after measuring in the standard basis, conditioned on the measurement of the postselection register being 1. Note that here the size of a postselecting quantum circuit is the size of the corresponding quantum circuit.

Aaronson proved in [1] that this model is equivalent to the model of quantum circuits with the ability to perform arbitrary postselections. Note that we must be somewhat careful here, as it is necessary that postselecting quantum circuits not be allowed to intersperse measurement and postselection.

Definition 1. *The Σ_i-circuit complexity of a boolean function f is the size of the smallest Σ_i-circuit C such that $C(x) = f(x)$ for all x.*

Definition 2. *The postselecting quantum circuit complexity of a boolean function f is the size of the smallest postselecting quantum circuit C such that for all x,*

$$f(x) = 1 \Rightarrow \Pr[C(x) = 1] \geq \frac{2}{3}$$

$$f(x) = 0 \Rightarrow \Pr[C(x) = 1] \leq \frac{1}{3}$$

Definition 3. *Let $\mathfrak{C}$ be a circuit model of computation and let L be a language. We define $f_n^L : \{0,1\}^n \to \{0,1\}$ by $f_n^L(x) = 1 \iff x \in L$. The $\mathfrak{C}$-complexity of L is the function $s(n) :=$ the $\mathfrak{C}$-complexity of f_n^L.*

3.3 Min-entropy and density

Definition 4. *Let X, Y be random variables. We define the min-entropy of X conditioned on Y:*

$$H_\infty(X|Y) := -\log \mathbb{E}_{Y \to y}[\max_x \Pr[X = x|Y = y]]$$

The unconditional min-entropy of X is the min-entropy of X conditioned on a constant. That is,

$$H_\infty(X) := -\log(\max_x \Pr[X = x])$$

Oftentimes, it is more convenient for us to reframe min-entropy from the framework of density. Formally,

Definition 5. *Let X be a random variable over some space $\mathcal{X}$. We say X has density δ if*

$$\max_x \Pr[X = x] \leq \frac{1}{\delta\,|\mathcal{X}|}$$

Note that a random variable X over $\{0,1\}^n$ has density δ if and only if the min-entropy of X is $\geq n - \log\frac{1}{\delta}$. Density also satisfies the following useful property.

Proposition 3. *Let (X, Y) be a random variable over some product space $\mathcal{X} \times \mathcal{Y}$. If (X, Y) has density δ, then X and Y both have density δ.*

To see this, observe that

$$\Pr[X \to x] = \sum_y \Pr[(X, Y) \to (x, y)] \leq |Y|\frac{1}{\delta|X||Y|} = \frac{1}{\delta|X|}$$

3.4 Classes of sources

A randomness source $\mathcal{S}$ is a distribution over some space $\mathcal{X}$. A class of sources $\mathfrak{S}$ is a set of randomness sources. We define several relevant classes of sources.

Definition 6. *We say that a distribution $\mathcal{S}$ is samplable by size-s circuits if there exists a circuit C of size s such that*

$$\Pr_r[C(r) = x] = \Pr[\mathcal{S} \to x]$$

for all x.

Definition 7. *We say that a distribution $\mathcal{S}$ is samplable by size-s quantum circuits if there exists a quantum circuit C of size s such that*

$$\Pr[C \to x] = \Pr[\mathcal{S} \to x]$$

for all x.

Note that as quantum circuits can sample their own randomness, we no longer need to quantify the probability that C outputs x by some randomness space.

We also define a notion of postselecting samplable sources. A source is samplable by postselecting circuits if we allow the circuit to condition on one of its outputs being 1.

Definition 8. *Let $\mathfrak{S}$ be a class of sources over $\mathcal{X} \times \{0,1\}$. We define a new class of sources $\mathfrak{S}_{nd}$ over $\mathcal{X}$, which we call postselecting $\mathfrak{S}$. We say that $\mathcal{S}' \in \mathfrak{S}_{nd}$ if there exists a source $(\mathcal{S}, b) \in \mathfrak{S}$ such that for all x',*

$$\Pr_{\mathcal{S}' \to x}[x = x'] = \Pr_{\mathcal{S} \to (x,b)}[x = x' | b = 1].$$

Observe that when $\mathfrak{S}$ is the class of sources samplable by size-s quantum circuits, we see that $\mathfrak{S}_{nd}$ is the class of sources samplable by size-s postselecting quantum circuits.

3.5 Statistical Distance and Extractors

Definition 9. *For two distributions X, Y, the statistical distance between X and Y is*

$$SD(X,Y) := \frac{1}{2} \sum_x |\Pr[X = x] - \Pr[Y = x]|$$

We say that X and Y are ϵ-close if $SD(X,Y) \leq \epsilon$.

Notation 1 *We say that U_n is the uniform distribution over $\{0,1\}^n$.*

Definition 10. *We say that a function $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ is a (k, ϵ) seeded extractor if for every distribution $\mathcal{S}$ over $\{0,1\}^n$ such that $H_\infty(\mathcal{S}) \geq k$, $\mathrm{Ext}(\mathcal{S}, U_t)$ is ϵ-close to U_m.*

Definition 11. *Let $\mathfrak{S}$ be some class of sources. We say that a function $EXT : \{0,1\}^n \to \{0,1\}^m$ is a (k, ϵ) (deterministic) extractor against $\mathfrak{S}$ if for every distribution $\mathcal{S} \in \mathfrak{S}$ such that $H_\infty(\mathcal{S}) \geq k$, $EXT(\mathcal{S})$ is ϵ-close to U_m.*

Definition 12. *Let $\mathfrak{S}$ be some class of sources. We say that a function $EXT : \{0,1\}^n \to \{0,1\}^m$ is a (k, ϵ) leakage-resilient extractor against $\mathfrak{S}$ if for every distribution $(\mathcal{S}, L) \in \mathfrak{S}$ such that $H_\infty(\mathcal{S}|L) \geq k$, $EXT(\mathcal{S})$ is ϵ-close to U_m.*

Definition 13. *We say that a function $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ is a (k_X, k_Y, ϵ) two source extractor if for every pair of distributions X, Y over $\{0,1\}^n$ with $H_\infty(X) \geq k_X$ and $H_\infty(Y) \geq k_Y$, $\mathrm{Ext}(X,Y)$ is ϵ-close to U_m.*

3.6 Combinatorial list decoding

Definition 14. *We say that a function $E : \{0,1\}^{n'} \times \{0,1\}^{n'} \to \{0,1\}^m$ satisfies (δ, t, ϵ)-combinatorial list decoding if, for all $a \in \{0,1\}^m$, the following holds:*

Let $\mathcal{S}$ be any distribution over $\{0,1\}^{n'}$ of density δ. Then,

$$\left| \left\{ w : \left| \Pr_{\mathcal{S} \to i}[E(w,i) = a] - \frac{1}{2^m} \right| \geq \epsilon \right\} \right| \leq t$$

Lemma 3 (Adapted from [3]). *If $E : \{0,1\}^{n'} \times \{0,1\}^{n'} \to \{0,1\}^m$ is a $\left(k_X, k_Y, \frac{\epsilon}{2^m}\right)$-two source extractor, then E satisfies $(2^{n'-k_Y}, 2^{k_X+1}, \epsilon)$ combinatorial-list decoding.*

Proof. Fix any $\mathcal{S}$ of density $2^{n'-b_Y}$. Then $\mathcal{S}$ has min-entropy b_Y. Assume towards contradiction that there exists some a such that

$$\left|\left\{w : \left|\Pr_{\mathcal{S}\to i}[E(w,i)=a] - \frac{1}{2^m}\right| \geq \epsilon\right\}\right| > 2^{b_X+1}$$

Without loss of generality, we can assume that

$$\left|\left\{w : \Pr_{\mathcal{S}\to i}[E(w,i)=a] \geq \frac{1}{2^m} + \epsilon\right\}\right| \geq 2^{b_X}$$

Define X to be the uniform distribution over such strings w. $H_{\min}(X) \geq k_X$. So

$$\Pr_{X\to w,\mathcal{S}\to i}[E(w,i)=a] \geq \frac{1}{2^m} + \epsilon$$

which contradicts E being a $\frac{\epsilon}{2^m}$ two source extractor.

Theorem 6. *[10] There exists a* $\left(0.2n', 0.9n', 2^{-0.2n'}\right)$ *two source extractor computable in time* $\text{poly}(n')$.

Corollary 3. *For every* $m \leq 0.1n'$*, there exists a function* $E : \{0,1\}^{n'} \to \{0,1\}^{n'} \to \{0,1\}^m$ *computable in time poly*(n') *satisfying* $(2^{-0.1n'}, 2^{0.2n'}, 2^{-0.1n'})$*-combinatorial list decoding.*

4 The Gap Probability Maximization Problem

We repeat our definition of the gap maximization problem (GPM) from the technical overview:

Say we are given a boolean randomized algorithm $\widetilde{C}$ and a constant γ with the following promise:
- There exists some x^* such that $\Pr[\widetilde{C}(x^*) \to 1] \geq \gamma$
- For all $x \neq x^*$, $\Pr[\widetilde{C}(x) \to 1] \leq \frac{\gamma}{2}$

The GPM problem asks us to find x^*.

In this section, we will show that the GPM problem can be efficiently solved for classical circuits using $NP_{||}$-circuits. Formally,

Theorem 1 *Let* $\gamma, s > 0$*, and let* $\mathcal{C}$ *be the class of boolean valued circuits of size* s *with input space* X *such that there exists an* $x^* \in X$ *satisfying*
- $\Pr[\widetilde{C}(x^*) \to 1] \geq \gamma$*,*
- *For all* $x \neq x^*$*,* $\Pr[\widetilde{C}(x) \to 1] \leq \frac{\gamma}{2}$*.*

Then $C(\widetilde{C}) = x^*$*. There exists a poly*(s)*-size* $\text{NP}_{||}$*-circuit* C *such that for all* $\widetilde{C} \in \mathcal{C}$*,* $C(\widetilde{C}) = x^*$*.*

In fact, the size of C will be polynomial in γ and s. But since $\gamma \leq 1$, we can thus upper bound the size of C by a polynomial in s. Note that this means the problem is easier for small values of γ. Intuitively, this is because the difficulty lies in reducing the number of witnesses for "bad" $x \neq x^*$.

Note that nondeterminism is indeed necessary to solve this problem, as such a C can be used to solve SAT. In particular, if $\widetilde{C}_\phi(x;r)$ is a circuit evaluating ϕ on input x, then for all ϕ with exactly one witness x^*, $C_1(\widetilde{C}_\phi) = x^*$.

We also show a quantum analogue of this theorem.

Theorem 7. *Let $\gamma, s > 0$, and let $\mathcal{C}$ be the class of boolean valued quantum circuits of size s with input space X and ℓ ancilla qubits such that there exists an $x^* \in X$ satisfying*
- $\Pr[\widetilde{C}(x^*) \to 1] \geq \gamma$,
- For all $x \neq x^*$, $\Pr[\widetilde{C}(x) \to 1] \leq \frac{\gamma}{2}$.
Then there exists a poly(s)-size PostBQP-circuit C such that for all $\widetilde{C} \in \mathcal{C}$, $\Pr[C(\widetilde{C}) \to x^*] \geq \frac{2}{3}$.

We also remark that it is not necessary that these algorithms be given γ, since they can try every $\gamma = \frac{1}{2^i}$ for each $1 \leq i \leq s$ in time $poly(s)$.

4.1 GPM solving for classical circuits

In this section, we prove Theorem 1 by relying on the following lemma.

Lemma 4. *Let $S = \{a_1, \ldots, a_{|S|}\}$. Choose $h : \{0,1\}^s \to [N]$ at random from a family of pairwise independent hash functions. Let S_h be the random variable defined by $S_h := S \cap \{r : h(r) = 0\}$. Then $\mathbb{E}[|S_h|] = \frac{|S|}{N}$ and $Var(|S_h|) \leq \frac{|S|}{N}$.*

Proof. Note that $|S_h| = \sum \mathbb{1}_{h(a_i)=0}$. We also have $\mathbb{E}_h[\mathbb{1}^2_{h(a_i)=0}] = \mathbb{E}_h[\mathbb{1}_{h(a_i)=0}] = \frac{1}{N}$. The lemma then follows by linearity of expectation and pairwise independence.

We now proceed to the proof of Theorem 1.

Proof. We will first define a randomized (non-adaptive) NP-circuit $Test$ solving GPM for classical circuits with some positive probability. That is, for all $\widetilde{C} \in \mathcal{C}$, $Test$ will satisfy

$$\Pr[Test(\widetilde{C}) = x^*] \geq 1 - 2^{-O(s \log s)}$$

As there are at most $2^{O(s \log s)}$ circuits of size s, by the union bound we have that there exists some randomness r such that for all $\widetilde{C}$ satisfying the property,

$$Test(\widetilde{C}; r) = x^*$$

Thus, nonuniformly fixing r gives us the C we want for the theorem statement.

To help us frame the argument, we will instead think of $\widetilde{C}$ as a circuit taking an input in X and a witness in $\{0,1\}^s$. We say that r is a witness for x if $C(x; r) = 1$. The property of C that we require is that there are at least $2^s\gamma$ witnesses for x, but for any $x \neq x^*$ there are at most $2^{r-1}\gamma$ witnesses.

$Test$ will use $\widetilde{C}$ to define a new circuit $\widetilde{C}'$ such that with high probability x^* has at least one witness under $\widetilde{C}'$, but any $x \neq x^*$ has NO witnesses. $Test$ will then operate by nondeterministically finding a $(\widetilde{x}, \widetilde{r})$ such that $\widetilde{C}'(\widetilde{x}; \widetilde{r}) = 1$, and outputting $\widetilde{x}$.

Given a sequence of hash functions $h_1, \ldots, h_k$ and a circuit C, $\widetilde{C}'(x)$ will first sample t distinct preimages of 0 for each h_i: $r_{i,1}, \ldots, r_{i,t}$. $\widetilde{C}'$ will then set $test_i^x$ to be true if $\widetilde{C}(x; r_{i,j}) = 1$ for all j. If the number of $test_i^x$'s satisfied is beyond the threshold t', $\widetilde{C}'(x)$ will output 1.

To argue why this works, let us first think of the case where $k = 1$. There will exist some randomness such that $test_i$ is true if and only if t witnesses for x under $\widetilde{C}$ fall into the subspace defined by $h_1^{-1}(0)$. For an appropriately chosen output length of h, with constant probability there

will be some randomness such that $test_i$ is satisfied for x^*. For any $x \neq x^*$, the probability that $test_i$ is satisfied for x will be less than some smaller constant, and thus these two events will be distinguishable.

Amplification of this process inside $\widetilde{C}'$ will then guarantee that with high probability there is a witness for x^* but no witness for x under $\widetilde{C}'$. Finally, because x^* is unique, we can find each bit of it with parallel calls to a SAT oracle.

We formally define $Test(\widetilde{C})$ in Algorithm 4.1. Here, N, k, t, and t' are all variables to be set later, and $\mathcal{H} \subseteq \{h : \{0,1\}^s \to [N]\}$ is a family of pairwise independent hash functions.

Algorithm 2 $Test(\widetilde{C})$

Sample $h_1, \ldots, h_k \xleftarrow{\$} \mathcal{H}$.
Construct a circuit $\widetilde{C}'$ as follows:

On input $(x, \{r_{i,j}\}_{[k]\times[t]})$ with $x \in X$, $r_{i,j} \in \{0,1\}^s$

If there exists i, j, j' such that $r_{i,j} = r_{i,j'}$, output 0.
If there exists i, j, such that $h_i(r_{i,j}) \neq 0$, output 0.
Let $test_i^x$ be 1 if and only if $\widetilde{C}(x; r_{i,j}) = 1$ for all $j \in [t]$.

Output 1 if and only if $\sum_{i\in[k]} test_i \geq t'$.

For $i = 1, \ldots, |x|$, define $\widetilde{C}'_i$ to the be the circuit: $\widetilde{C}'_i(x, r) := \widetilde{C}'(x, r) \wedge (x_i = 1)$
For $i = 1, \ldots, |x|$, set $\hat{x}_i = 1$ if and only if $\widetilde{C}'_i$ is satisfiable (via parrallel SAT calls).
Output $\hat{x}$.

Let A be the event that there exists a witness for x^* under $\widetilde{C}'$. Let B be the event that there does not exist a witness for any $x \neq x^*$. As long as A and B both hold, $Test(\widetilde{C})$ will succeed.

Formally, call $\{r_{i,j}\}$ valid if the first two tests in $\widetilde{C}'$ pass for $r_{i,j}$. That is, for each i, $r_{i,1}, \ldots, r_{i,t}$ are distinct, and $h(r_{i,j}) = 0$ for all i, j.

We define $R_{i,x}$ to be an indicator that there exists a witness r such that $test_i^x$ is 1. Then, we can view A as the event that $\sum_i R_{i,x^*} \geq t'$ and B as the event that $\sum_i R_{i,x} < t'$ for all $x \neq x^*$.

Define $S^x := \{r | C(x; r) \to 1\}$. It is clear that $\mathbb{E}_{h_i}[R_{i,x}] = \Pr_{h_i}[|S_h^x| \geq t]$.

We will begin by bounding $\mathbb{E}[R_{i,x^*}]$. Let $S' \subseteq S^{x^*}$ be some subset of size $\gamma 2^s$. We know that $|S^{x^*}| \geq \gamma 2^s$. Using the key lemma and applying Chebyshev's inequality gives us

$$\begin{aligned}
\mathbb{E}_{h_i}[R_{i,x^*}] &= \Pr_{h_i}[|S_h^{x^*}| \geq t] \\
&\geq \Pr_{h_i}[|S'_h| \geq t] \\
&\geq 1 - \Pr_{h_i}\left[\left|\frac{|S'|}{N} - |S'_h|\right| > \frac{|S'|}{N} - t\right] \\
&\geq 1 - \frac{\gamma 2^s}{N} \frac{1}{\left(\frac{\gamma 2^s}{N} - t\right)^2}
\end{aligned}$$

Similarly, for $x \neq x^*$, we have $|S^x| \leq \gamma 2^{s-1}$, and so we can use similar techniques to get

$$\begin{aligned}\mathbb{E}_{h_i}[R_{i,x}] &= \Pr_{h_i}[|S_h^x| \geq t] \\ &\leq \Pr[|S_h^x| - \frac{|S^x|}{N} \geq t - \frac{|S^x}{N}] \\ &\leq \frac{2^s}{2N} \frac{1}{\left(t - \frac{\gamma 2^s}{2N}\right)^2}\end{aligned}$$

Setting $t = 48$, $N = \dfrac{\gamma 2^s}{64}$ gives us

$$\mathbb{E}_{h_i}[R_{i,x^*}] \geq \frac{3}{4}$$

$$\mathbb{E}_{h_i}[R_{i,x\neq x^*}] \leq \frac{1}{8}$$

Thus, if we set $t' = \dfrac{k}{2}$, we should have $\Pr[\sum_{i\in[k]} R_{i,x^*}]$ is large and for $x \neq x^*$, $\Pr[\sum_{i\in[k]} R_{i,x}]$ is small.

Thus, the Chernoff bound gives us that

$$\begin{aligned}\Pr_{h_1,\ldots,h_k}[\overline{A}] &= \Pr_{h_1,\ldots,h_k}[\sum_{i\in[k]} R_{i,x^*} < t'] \\ &\leq \Pr_{h_1,\ldots,h_k}[\sum_{i\in[k]} R_{i,x^*} \leq \left(1 - \frac{1}{3}\right)\frac{3}{4}k] \leq e^{-\frac{k}{24}}\end{aligned}$$

and (by using the union bound)

$$\begin{aligned}\Pr_{h_1,\ldots,h_k}[\overline{B}] &= \Pr[\exists\ x \neq x^* : \sum_{i\in[k]} R_{i,x} \geq \frac{k}{2}] \\ &\leq \sum_{x\neq x^*} \Pr[\sum_{i\in[k]} R_{i,x} \geq (1+3)\frac{k}{8}] \\ &\leq 2^s \cdot e^{-\frac{9k}{40}}\end{aligned}$$

Together, this means

$$\Pr[A \text{ and } B] \geq 1 - e^{-\frac{k}{24}} - 2^s e^{-\frac{9k}{40}}$$

and so for $k = O(s\log(s))$, we get that $Test$ succeeds with all but $2^{-O(s\log s)}$ probability.

Note that the size of $Test$ is $poly(s, k, t, t', \log N)$. But $k = O(s\log(s))$, $t = 48 = O(1)$, $t' = k/2 = O(s)$, and $\log N = \log \dfrac{\gamma 2^s}{64} = O(s)$, and so $Test$ runs in time $poly(s)$.

4.2 GPM solving for quantum circuits

In this section we prove Theorem 7

Proof. Using the principle of deferred measurement, we will assume that $\widetilde{C}(x)$ produces a pure state $|\phi_x\rangle$ and outputs a measurement of the first register of $|\phi_x\rangle$ under the standard basis. We will write $|\phi_x\rangle = \alpha_x |0\rangle |\phi_x^0\rangle + \beta_x |1\rangle |\phi_x^1\rangle$.

We define $C(\widetilde{C})$ as follows. First, run $\widetilde{C}$ on $\frac{1}{\sqrt{|X|}} \sum_{x \in X} |x\rangle$ to generate

$$|\psi_1\rangle \propto \sum_{x \in X} |\phi_x\rangle |x\rangle .$$

Then, postselect on the first register of $|\psi_1\rangle$ being 1 to get

$$|\psi_2\rangle \propto \sum_{x \in X} \beta_x |\phi_x^1\rangle |x\rangle$$

Repeating this process k times gives the state

$$|\psi_2\rangle^{\otimes k} \propto \sum_{x_1,\ldots,x_k} \beta_{x_1} \cdots \beta_{x_k} |\phi_{x_1}^1\rangle \cdots |\phi_{x_k}^1\rangle |x_1\rangle \cdots |x_k\rangle$$

Postselecting on the predicate $x_1 = x_2 = \cdots = x_k$ gives us

$$|\psi_3\rangle \propto \sum_x \beta_x^k |\phi_x^1\rangle^{\otimes k} |x\rangle^{\otimes k}$$

$C(\widetilde{C})$ will produce $|\psi_3\rangle$, and output the result of a measurement on the last register in the standard basis.

Then,

$$\begin{aligned} \Pr[C(\widetilde{C}) \to x^*] &= \frac{\beta_{x^*}^{2k}}{\beta_{x^*}^{2k} + \sum_{x \neq x^*} \beta_x^{2k}} \\ &\geq \frac{\gamma^k}{\gamma^k + (|X|-1)\frac{\gamma^k}{2^k}} \end{aligned}$$

If we set $k \geq 2 \log |X|$, then

$$\Pr[C(\widetilde{C}) \to x^*] > \frac{\gamma^k}{\frac{3}{2}\gamma^k} = \frac{2}{3}$$

Note that $C(\widetilde{C})$ just simulates $\widetilde{C}$, $\log |X|$ times. Since $\log |X| \leq size(\widetilde{C})$, this means that C is of size $poly(size(\widetilde{C}))$.

4.3 On GPM for postselecting quantum circuits

One may wish to extend this result to get a GPM solver for postselecting quantum circuits. The naive approach to this would be to define $|\phi_x\rangle$ to be the state produced by $\widetilde{C}(x)$ after postselection but before measurement. However, this approach will not work. The subtlety here is that there is

no obvious way to produce a superposition over these states, as postselecting on $\widetilde{C}\left(\sum_{x\in X}|x\rangle\right)$ will not give the desired result.

The counterexample for this is the following circuit $\widetilde{C}(x)$ defined as follows:
First, it produces the state

$$|\phi\rangle = \sqrt{1-\epsilon}\,|0\rangle\,|x\rangle + \sqrt{\epsilon}\,|1\rangle\,|1\rangle$$

Then, it postselects on the second register measuring to 1, and outputs the result of a measurement on the first register.

Let $|\phi_x\rangle$ be the state resultant from $\widetilde{C}(x)$ before postslection and let $|\phi'_x\rangle$ be the state resultant after postselection. We have

$$|\phi_0\rangle = \sqrt{1-\epsilon}\,|0\rangle\,|0\rangle + \sqrt{\epsilon}\,|1\rangle\,|1\rangle\,; |\phi'_0\rangle = |1\rangle$$

and

$$|\phi_1\rangle = \sqrt{1-\epsilon}\,|0\rangle\,|1\rangle + \sqrt{\epsilon}\,|1\rangle\,|1\rangle\,; |\phi'_1\rangle = \sqrt{1-\epsilon}\,|0\rangle + \sqrt{\epsilon}\,|1\rangle$$

It is clear that the probability that $\widetilde{C}$ accepts 0 is more than twice the probability it accepts 1, and so our GPM algorithm should 0 on input $\widetilde{C}$. However, let us consider what happens when we try to produce $|\phi_0\rangle + |\phi_1\rangle$ by postselecting on $\widetilde{C}$ applied to the uniform superposition. We start with

$$(\sqrt{1-\epsilon}\,|00\rangle + \sqrt{\epsilon}\,|11\rangle)\,|0\rangle + (\sqrt{1-\epsilon}\,|01\rangle + \sqrt{\epsilon}\,|11\rangle)\,|1\rangle$$

and so postselecting on the second register being 1 leaves us with the residual state proportional to

$$\sqrt{\epsilon}\,|11\rangle\,|0\rangle + \sqrt{1-\epsilon}\,|01\rangle\,|1\rangle + \sqrt{\epsilon}\,|11\rangle\,|1\rangle$$

But this is very different from $|\phi'_0\rangle + |\phi'_1\rangle$. In particular, measuring the first register will produce 1 with very high probability. Thus, continuing to run our GPM procedure using this state will output 1.

This argument shows that a naive extension of the GPM protocol described in the previous section will not be able to solve GPM for postselecting circuits. However, this does not mean that no postselecting algorithm can solve GPM for postselecting circuits. Although our techniques bypass this, we leave this challenge as an open question.

5 Extractors from hardness assumptions

In this section, we focus on proving the following theorem.

Theorem 8. *Let $p : \mathbb{F}^t \to \mathbb{F}$ be any polynomial of degree d with $|\mathbb{F}| = q = 2^{n'}$. Let $E : \mathbb{F}\times\{0,1\}^{n'} \to \{0,1\}^m$ satisfy $(2^{-0.2n'}, 2^{-0.1n'}, 2^{0.2n'})$-combinatorial list decoding.*

For any ϵ, δ satisfying

$$\frac{d}{q} \leq \frac{\epsilon^2\delta^2}{64\cdot 2^{0.4n'+2m}}$$
$$\epsilon\delta \geq 128\cdot 2^{m-0.1n'}$$

the following holds:

If there exists a size-s postselecting (quantum/classical) samplable source S of density δ with output space $\mathbb{F}^t \times \{0,1\}^{n'}$ such that

$$\left| \Pr_{S \to (x,i)}[E(p(x), i) = b] - \frac{1}{2^m} \right| \geq \frac{\epsilon}{2^m}$$

then there exists a (postselecting/$NP_{||}$)-circuit C of size poly $\left(s, 2^m, \frac{1}{\epsilon}\right)$ computing p everywhere.

First we instantiate Theorem 8 with the E from Corollary 3. Theorems 2 and 3 then follow by applying a similar argument to the one used by [26] in proving Theorem 5.8 from Theorem 5.3. The full proofs of these theorems are deferred to Appendix C. Theorem 5 then gives us Corollary 1.

To prove Theorem 8, we will rely on the following (restated) key lemma as well as the existence of an efficient implementation of a circuit solving the gap probability maximization problem:

Lemma 5. *KEY LEMMA:*
Let $\mathbb{F} = GF(2^{n'})$ be a field of size $q = 2^{n'}$. Let $p : \mathbb{F}^t \to \mathbb{F}$ be any degree d polynomial, let $E : \mathbb{F} \times \{0,1\}^{n'} \to \{0,1\}^m$ satisfy $(2^{-4m}, 2^{-0.1n'}, 2^{0.2n'})$-combinatorial list decoding, and let $\mathcal{S}$ be any distribution of density δ such that

$$\left| \Pr_{\mathcal{S} \to (u,i)}[E(p(u), i) = 1] - \frac{1}{2^m} \right| \geq \frac{\epsilon}{2^m}.$$

For any ϵ, δ satisfying

$$\frac{d}{q} \leq \frac{\epsilon^2 \delta^2}{64 \cdot 2^{0.4n'+2m}}$$
$$\epsilon\delta \geq 128 \cdot 2^{m-0.1n'}$$

the following holds: There exists a z such that for $\frac{15}{16}$ values of x,
- $\left| \Pr_{S \to (u,i)}[E(p(u), i) = 1 | u \in L_{z,x}] - \frac{1}{2^m} \right| \geq \frac{\epsilon}{3 \cdot 2^m}$
- For all $h : \mathbb{F} \to \mathbb{F}$ such that $h \neq p \circ L_{z,x}$, either $\left| \Pr_{S \to (u,i)}[E(h(L_{z,x}^{-1}(u)), i) = 1 | u \in L_{z,x}] - \frac{1}{2^m} \right| \leq \frac{\epsilon}{6 \cdot 2^m}$ or $h(0) \neq p(z)$.

The proof of this lemma will be presented in Section 5.1. We now prove Theorem 8 from Lemma 5.

Proof. Towards contradiction, we will assume that there is some samplable distribution $\mathcal{S}$ which biases $E(p(x), i)$ for some output. Without loss of generality, assume that

$$\left| \Pr_{S \to (x,i)}[E(p(x), i) = 1] - \frac{1}{2^m} \right| \geq \frac{\epsilon}{2^m}.$$

Algorithm 3 $\widetilde{C}_x(h)$

If $h(0) \neq p(z)$, output 0.

Sample $(u_1, i_1, b_1), \ldots, (u_k, i_k, b_k) \xleftarrow{\$} SAMP$.

Output 1 if and only if for all j:

- $b_j = 1$
- $u_j \in L_{z,x}(\mathbb{F})$ or
- $E(h(L_{z,x}^{-1}(u_j)), i_j) \neq 1$.

If $\Pr_{S \to (x,i)}[E(p(x), i) = 1] \leq \frac{1}{2^m} - \frac{\epsilon}{2^m}$, we define $\tilde{C}_x(h)$ as follows:

Let $PASS$ be the event that all but the last test succeeds for all j. If $PASS$ occurs, then (u_j, i_j) is distributed according to $\mathcal{S}$ conditioned on the event that the first output is on the line. Thus, for any h such that $h(0) = p(z)$,

$$\Pr[\widetilde{C}_x(h) \to 1] = \Pr[PASS] \Pr_{\mathcal{S} \to (u,i)}[E(p(u), i) \neq 1 | u \in L_{z,x}(\mathbb{F})]^k$$

If we nonuniformly fix z to be the same as from the key lemma, we then have that

$$\Pr[\widetilde{C}_x(p \circ L_{z,x}) \to 1] \geq \Pr[PASS] \left(1 - \frac{1}{2^m} + \frac{\epsilon}{3 \cdot 2^m}\right)^k$$

and

$$\Pr[\widetilde{C}_x(h') \to 1] \leq \Pr[PASS] \left(1 - \frac{1}{2^m} + \frac{\epsilon}{6 \cdot 2^m}\right)^k$$

for all $h' \neq p \circ L_{z,x}$. Then, for $k = O\left(\frac{2^m}{\epsilon}\right)$, we can say that

$$\left(1 - \frac{1}{2^m} + \frac{\epsilon}{3 \cdot 2^m}\right)^k \geq 2 \left(1 - \frac{1}{2^m} + \frac{\epsilon}{6 \cdot 2^m}\right)^k$$

So nonuniformly fixing $\gamma = \Pr[PASS] \left(1 - \frac{1}{2^m} + \frac{\epsilon}{3 \cdot 2^m}\right)^k$ in Theorem 1 (or Theorem 7 for the quantum case) gives us a ($NP_{||}$/postselecting quantum)-circuit computing p on 15/16 of its inputs. Polynomial reconstruction then gives a ($NP_{||}$/postselecting quantum)-circuit computing p everywhere.

If $\Pr_{S \to (x,i)}[E(p(x), i) = 1] \geq \frac{1}{2^m} + \frac{\epsilon}{2^m}$, we define $\tilde{C}_x(h)$ as follows:

If we nonuniformly fix z to be the same as from the key lemma, we then have that

$$\Pr[\widetilde{C}_x(p \circ L_{z,x}) \to 1] \geq \Pr[PASS] \left(\frac{1}{2^m} + \frac{\epsilon}{3 \cdot 2^m}\right)^k$$

and

$$\Pr[\widetilde{C}_x(h') \to 1] \leq \Pr[PASS] \left(\frac{1}{2^m} + \frac{\epsilon}{6 \cdot 2^m}\right)^k$$

Algorithm 4 $\widetilde{C}_x(h)$

If $h(0) \neq p(z)$, output 0.

Sample $(u_1, i_1, b_1), \ldots, (u_k, i_k, b_k) \xleftarrow{\$} SAMP$.

Output 1 if and only if for all j:

- $b_j = 1$
- $u_j \in L_{z,x}(\mathbb{F})$ or
- $E(h(L_{z,x}^{-1}(u_j)), i_j) = 1$.

for all $h' \neq p \circ L_{z,x}$. Then, for some $k = O\left(\frac{2^m}{\epsilon}\right)$, we can say that

$$\left(\frac{1}{2^m} + \frac{\epsilon}{3 \cdot 2^m}\right)^k \geq 2\left(\frac{1}{2^m} + \frac{\epsilon}{6 \cdot 2^m}\right)^k$$

So nonuniformly fixing $\gamma = \Pr[PASS]\left(\frac{1}{2^m} + \frac{\epsilon}{3}\right)^k$ in Theorem 1 (or Theorem 7 for the quantum case) gives us the result for the same reason as in the previous case.

Observe that $\widetilde{C}_x$ and thus also the reconstruction algorithm is of size $poly\left(s, 2^m, \frac{1}{\epsilon}\right)$, and so we are done. Note that in order to run the GPM solver, we need C to be a $NP_{||}$-circuit or a PostBQP-circuit, depending on whether we are operating in the classical or quantum world.

5.1 Proof of Key Lemma

We will prove our key lemma by relying on two sublemmas. The proofs of these lemmas will be delegated to sections 5.2 and 5.3 respectively. The proofs of these claims are adapted from the proofs of Claims 37 and 38 in the full version of [25], although the proof of Lemma 7 is significantly more involved. We give a full intuition for the proof of Lemma 7 at the beginning of its relevant subsection.

Lemma 6. *Let $\sigma = \frac{1}{\delta^2 q} + \frac{2^{2m}}{\epsilon^2 \delta^2 q} + \frac{1}{q^t}$. Choose $a, b \in \mathbb{F}^t$ uniformly at random. We define $L : \mathbb{F} \to \mathbb{F}^t$ to be the line between a and b. Formally, $L(t) := at + b(1-t)$. Then, with probability $1 - \sigma$ over the choice of a, b, $a \neq b$,*

$$\Pr_{\mathcal{S} \to (x,i)}[x \in L(\mathbb{F})] \in \left(\frac{1}{2}\frac{q}{q^t}, \frac{3}{2}\frac{q}{q^t}\right)$$

and

$$\left| \Pr_{\mathcal{S} \to (x,i)}[E(p(x), i) = 1 | x \in L(\mathbb{F})] - \frac{1}{2^m} \right| \geq \frac{\epsilon}{3 \cdot 2^m}$$

Lemma 7. *Let L be a non-trivial line such that*

$$\Pr_{\mathcal{S} \to (x,i)}[x \in L(\mathbb{F})] \geq \frac{1}{2}\frac{q}{q^t}.$$

Define $\overline{\mathcal{S}} = \mathcal{S}|_L$. We call a polynomial $h : \mathbb{F} \to \mathbb{F}$ "bad" if

$$\left| \Pr_{\overline{S} \to (x,i)}[E(h(L^{-1}(x)), i) = 1] - \frac{1}{2^m} \right| \geq \frac{\epsilon}{6 \cdot 2^m}.$$

If

$$\sqrt{\frac{d}{q}} \leq \frac{5}{48}\frac{\epsilon\delta}{2^{0.2n'+m}}$$

$$\epsilon\delta \geq 128 \cdot 2^{m-0.1n'}$$

then

$$|\{h : h \text{ } is \text{ } \text{"}bad\text{"}\}| \leq \frac{16 \cdot 2^{0.4n'+m}}{\epsilon}$$

To prove our key lemma, we first consider the experiment where the two points z, x are chosen uniformly at random from $\mathbb{F}^t$, and we consider the line $L_{z,x}$ between them. The probability over the choice of z, x that Lemma 6 fails to hold is $\leq \sigma$, and as long as Lemma 6 holds, so will Lemma 7. To evaluate the probability of both conditions of the key theorem holding, we define E to be the event over uniform random variables z and x that

$$\left| \text{ bias of } \mathcal{S}|_L \text{ on } E(p(L^{-1}(x)), i) - \frac{1}{2^m} \right| < \frac{\epsilon}{3 \cdot 2^m}$$

$$\text{or } \exists h \neq p \circ L_{z,x} \text{ s.t. } h(0) = p(z) \text{ and } \left| \text{ bias of } \mathcal{S}|_L \text{ on } E(h(L^{-1}(x)), i) - \frac{1}{2^m} \right| < \frac{\epsilon}{3 \cdot 2^m}$$

We upper bound the probability of event E.

$$\Pr_{z,x}[E] \leq \Pr_{z,x}\left[E \cup |\text{bad h }| > \frac{16 \cdot 2^{0.4n'+m}}{\epsilon}\right]$$

from the union bound.

$$\begin{aligned}
\Pr_{z,x}[E] &\leq \Pr_{z,x}\left[E \cup |\text{bad h }| > \frac{16 \cdot 2^{0.4n'+m}}{\epsilon}\right] \\
&\leq \Pr_{z,x}\left[\left| \text{ bias of } S|_L \text{ on } E(p(L^{-1}(x)), i) - \frac{1}{2^m} \right| > \frac{\epsilon}{6 \cdot 2^m} \cup |\text{bad h }| > \frac{16 \cdot 2^{0.4n'+m}}{\epsilon}\right] \\
&+ \Pr_{z \leftarrow L}\left[|\text{bad h }| \leq \frac{16 \cdot 2^{0.4n'+m}}{\epsilon} \cap \exists \text{bad } h \text{ s.t. } h(0) = p(z)\right] \\
&\leq \sigma + \frac{16 \cdot 2^{0.4n'+m}}{\epsilon}\frac{d}{q}
\end{aligned}$$

Thus,

$$1 - \Pr_{z,x}[E] \geq 1 - \sigma - \frac{16 \cdot 2^{0.4n'+m}}{\epsilon}\frac{d}{q}.$$

Therefore, as long as

$$\sigma + \frac{16 \cdot 2^{0.4n'+m}}{\epsilon}\frac{d}{q} \leq \frac{1}{32}$$

and the conditions from Lemma 7 hold, an averaging argument gives us that there exists some z such that for 15/16 values of x, the properties we care about in the lemma statement hold. But note that $\frac{d}{q} \leq \frac{\epsilon^2\delta^2}{64 \cdot 2^{0.4n'+2m}}$ implies that

$$\sigma + \frac{16 \cdot 2^{0.4n'+m}}{\epsilon}\frac{d}{q} \leq \frac{1}{32}$$

and so we are done.

5.2 Proof of Lemma 6

Proof. Observe that the error is of the form $\frac{1 \pm \epsilon}{2^m}$.
First, let's assume that $\Pr_{\mathcal{S} \to (x,i)}[E(p(x), i) = 1] \leq \frac{1-\epsilon}{2^m}$
Let A be the event that

$$\Pr_{\mathcal{S} \to (x,i)}[x \in L(\mathbb{F})] \in \left(\frac{1}{2}\frac{q}{q^t}, \frac{3}{2}\frac{q}{q^t}\right)$$

and let B be the event that

$$\Pr_{\mathcal{S} \to (x,i)}[E(p(x), i) = 1 | x \in L(\mathbb{F})] \leq \frac{1}{2^m} - \frac{\epsilon}{3 \cdot 2^m}.$$

Note that A and B are defined as events over the random choice of $a, b \in \mathbb{F}^t$ defining L. To show that A and B hold with high probability, we define a new event B' such that A and B' together imply B. Thus,

$$\Pr[A \text{ and } B] \geq \Pr[A \text{ and } B'] \geq 1 - \Pr[\overline{A}] - \Pr[\overline{B'}].$$

Therefore, it will only remain to bound $\Pr[\overline{A}]$ and $\Pr[\overline{B'}]$. To set up, we will define a few variables:
- $p_x := \Pr[\mathcal{S} \to (x, \cdot)]$ will be the probability that $\mathcal{S}$ outputs x as its first output.
- $p_L := \Pr_{\mathcal{S} \to (x,i)}[x \in L(\mathbb{F})]$ will be the probability that $\mathcal{S}$ outputs a point on L. Note that A is the event that $p_L \in \left(\frac{1}{2}\frac{q}{q^t}, \frac{3}{2}\frac{q}{q^t}\right)$.
- $w_x := \Pr_{\mathcal{S} \to (x',i')}[E(p(x'), i') = 1 | x' = x]$ will be the bias of $E(p(x'), i')$ conditioned on x. This can be thought of as the "weight" of x on the bias of our extractor.
- $\widetilde{w_x} := \frac{1}{2^m} - w_x$ will be the weight of x shifted by $\frac{1}{2^m}$.

Note that we can write the event A and B in terms of these variables. In particular, A is the event that $p_L \in \left(\frac{1}{2}\frac{q}{q^t}, \frac{3}{2}\frac{q}{q^t}\right)$. We also have

$$\begin{aligned}
\Pr_{\mathcal{S} \to (x,i)}[E(p(x), i) = 1 | x \in L(\mathbb{F})] &= \sum_{x \in L(\mathbb{F})} \Pr_{\mathcal{S} \to (x',i')}[E(p(x'), i') | x' = x] \Pr_{\mathcal{S} \to (x',i')}[x' = x | x' \in L] \\
&= \sum_{x \in L(\mathbb{F})} w_x \frac{\Pr[\mathcal{S} \to x]}{\Pr_{\mathcal{S} \to (x',i')}[x' \in L]} \\
&= \frac{1}{2^m} - \frac{1}{p_L} \sum_{x \in L(\mathbb{F})} \widetilde{w_x} p_x
\end{aligned}$$

Observe that if $p_L \leq \frac{3}{2}\frac{q}{q^t}$ and $\sum_{x\in\mathbb{F}^t} \widetilde{w_x} p_x \geq \frac{\epsilon}{2}\frac{q}{q^t}$, then

$$\Pr_{\mathcal{S}\to(x,i)}[E(p(x),i)=1 | x \in L(\mathbb{F})] \leq \frac{1}{2^m} - \frac{\epsilon}{3} \leq \frac{1}{2^m} - \frac{\epsilon}{3\cdot 2^m}$$

. and so B holds. Thus, we define B' to be the event that L is non-trivial and

$$\sum_{a\in\mathbb{F}} \widetilde{w_{L(a)}} p_{L(a)} \geq \frac{\epsilon}{2}\frac{q}{q^t}$$

It is clear that A and B' together imply B. We now proceed to bounding the probabilities of A and B'. Both of these arguments will boil down to a simple application of Chebyshev's inequality.
A) Since L is chosen by selecting two points $a, b \in \mathbb{F}^t$ uniformly at random and setting $L(0) = a$, $L(1) = b$, we have that $L(0)$ and $L(1)$ are independent, uniformly random variables over $\mathbb{F}^t$. But we note that for any pair of indices i, j, it would be equivalent to choose L by setting $L(i) = a$ and $L(j) = b$. Thus, $\{L(a)\}_{a\in\mathbb{F}}$ is a collection of pairwise independent uniform random variables. Define $R_a := p_{L(a)}$. $\{R_a\}_{a\in\mathbb{F}}$ is a set of pairwise independent random variables. Note that $\mathbb{E}_L[R_a] = \mathbb{E}_{x\xleftarrow{\$}\mathbb{F}^t}[p_x]$. But $\sum_{x\in\mathbb{F}^t} p_x = 1$, and so $\mathbb{E}_L[R_a] = \frac{1}{q^t}$. Also, $R_a \in \left[0, \frac{1}{\delta q^t}\right]$ by the density requirement on $\mathcal{S}$. Thus, we get $Var(R_a) \leq \frac{1}{4\delta^2 q^{2t}}$ by Popoviciu's inequality on variances. Observe that $p_L = \sum_{a\in\mathbb{F}} R_a$. Thus, $\mathbb{E}[p_L] = \frac{q}{q^t}$ and $Var(p_L) \leq \frac{q}{4\delta^2 q^{2t}}$ by pairwise independence. Chebyshev gives us

$$\Pr[\overline{A}] = \Pr\left[\left|p_L - \frac{q}{q^t}\right| \geq \frac{1}{2}\frac{q}{q^t}\right] \leq Var(R)\cdot\frac{4q^{2t}}{q} = \frac{1}{\delta^2 q}$$

B') We will use the same approach as for A, but with a different random variable. Define $\widetilde{R_a} := \widetilde{w_{L(a)}} p_{L(a)}$. Note that

$$\begin{aligned}
\sum_{x\in\mathbb{F}^t} \widetilde{w_x} p_x &= \frac{1}{2^m} - \left(\sum_{x\in\mathbb{F}^t} \Pr_{\mathcal{S}\to(x',i')}[E(p(x'),i')=1 | x'=x]\Pr[\mathcal{S}\to x]\right) \\
&= \frac{1}{2^m} - \Pr_{\mathcal{S}\to(x,i)}[E(p(x),i)=1] \\
&= \frac{\epsilon}{2^m}
\end{aligned}$$

Thus, $\mathbb{E}_L[\widetilde{R_a}] = \mathbb{E}_{x\xleftarrow{\$}\mathbb{F}^t}[\widetilde{w_x} p_x] = \frac{\epsilon}{q^t 2^m}$. Also, $\widetilde{R_a} \in \left(-\frac{1}{2}\frac{1}{\delta q^t}, \frac{1}{2}\frac{1}{\delta q^t}\right)$ and so

$$Var(\widetilde{R_a}) \leq \frac{1}{4\delta^2 q^{2t}}$$

Define $\widetilde{R} = \sum_{a \in \mathbb{F}} \widetilde{R_a}$. Note that as long as the line is non-trivial, $\widetilde{R} = \sum_{x \in L(\mathbb{F})} \widetilde{w_x} p_x$. By linearity of expectation and linearity of variance for pairwise independent random variables, we have

$$\mathbb{E}[\widetilde{R}] = \frac{q\epsilon}{q^t 2^m}$$

$$Var(\widetilde{R}) \leq \frac{1}{\delta^2 q^{2t}}$$

$$\Pr[\overline{B'}] \leq \Pr\left[L \text{ is trivial or } \left|\widetilde{R} - \frac{q\epsilon}{q^t 2^m}\right| \geq \frac{1}{2}\frac{q\epsilon}{q^t 2^m}\right] \leq \frac{1}{q^t} + \frac{2^{2m}}{\epsilon^2 \delta^2 q}$$

And so setting $\sigma = \frac{1}{q^t} + \frac{1}{\delta^2 q} + \frac{2^{2m}}{\epsilon^2 \delta^2 q}$ gives us our claim.

Now, let's assume that $\Pr_{\mathcal{S} \to (x,i)}[E(p(x), i) = 1] \geq \frac{1+\epsilon}{2^m}$

Let A be the event that

$$\Pr_{\mathcal{S} \to (x,i)}[x \in L(\mathbb{F})] \in \left(\frac{1}{2}\frac{q}{q^t}, \frac{3}{2}\frac{q}{q^t}\right)$$

and let B be the event that

$$\Pr_{\mathcal{S} \to (x,i)}[E(p(x), i) = 1 | x \in L(\mathbb{F})] \geq \frac{1}{2^m} + \frac{\epsilon}{3 \cdot 2^m}.$$

Note that A and B are defined as events over the random choice of $a, b \in \mathbb{F}^t$ defining L. To show that A and B hold with high probability, we define a new event B' such that A and B' together imply B. Thus,

$$\Pr[A \text{ and } B] \geq \Pr[A \text{ and } B'] \geq 1 - \Pr[\overline{A}] - \Pr[\overline{B'}].$$

Therefore, it will only remain to bound $\Pr[\overline{A}]$ and $\Pr[\overline{B'}]$. To set up, we will define a few variables:

- $p_x := \Pr[\mathcal{S} \to (x, \cdot)]$ will be the probability that $\mathcal{S}$ outputs x as its first output.
- $p_L := \Pr_{\mathcal{S} \to (x,i)}[x \in L(\mathbb{F})]$ will be the probability that $\mathcal{S}$ outputs a point on L. Note that A is the event that $p_L \in \left(\frac{1}{2}\frac{q}{q^t}, \frac{3}{2}\frac{q}{q^t}\right)$.
- $w_x := \Pr_{\mathcal{S} \to (x',i')}[E(p(x'), i') = 1 | x' = x]$ will be the bias of $E(p(x'), i')$ conditioned on x. This can be thought of as the "weight" of x on the bias of our extractor.
- $\widetilde{w_x} := w_x - \frac{1}{2^m}$ will be the weight of x shifted by $\frac{1}{2^m}$.

Note that we can write the event A and B in terms of these variables. In particular, A is the event that $p_L \in \left(\frac{1}{2}\frac{q}{q^t}, \frac{3}{2}\frac{q}{q^t}\right)$. We also have

$$\begin{aligned}
\Pr_{\mathcal{S} \to (x,i)}[E(p(x), i) = 1 | x \in L(\mathbb{F})] &= \sum_{x \in L(\mathbb{F})} \Pr_{\mathcal{S} \to (x',i')}[E(p(x'), i') | x' = x] \Pr_{\mathcal{S} \to (x',i')}[x' = x | x' \in L] \\
&= \sum_{x \in L(\mathbb{F})} w_x \frac{\Pr[\mathcal{S} \to x]}{\Pr_{\mathcal{S} \to (x',i')}[x' \in L]} \\
&= \frac{1}{p_L} \sum_{x \in L(\mathbb{F})} \widetilde{w_x} p_x + \frac{1}{2^m}
\end{aligned}$$

Observe that if $p_L \leq \frac{3}{2}\frac{q}{q^t}$ and $\sum_{x \in \mathbb{F}^t} \widetilde{w_x} p_x \geq \frac{\epsilon}{2}\frac{q}{q^t}$, then

$$\Pr_{\mathcal{S} \to (x,i)}[E(p(x), i) = 1 | x \in L(\mathbb{F})] \geq \frac{1}{2^m} + \frac{\epsilon}{3} \geq \frac{1}{2^m} + \frac{\epsilon}{3 \cdot 2^m}$$

. and so B holds.
We can now proceed identically to the previous case.

5.3 Proof of Lemma 7

The proof of Lemma 4.1 in [26] relies on a weaker version of this lemma, which is originally stated in [13]. We repeat that statement here as a proposition:

Proposition 4. *Theorem 4.4 from [13].*
Consider $\mathcal{C}$ a $[N, D]_q$ code. That is, $\mathcal{C} \subseteq [q]^N$ and has a minimum distance between codewords of D. Let $R \in [q]^N$. Define $\gamma := 1 - \frac{D}{N}$. Let $C_1, \ldots, C_m \in \mathcal{C}$ such that for all j,

$$\Pr_{U \to x}[(C_j)_{(x)} = R_{(x)}] \geq \epsilon.$$

Then if $\epsilon \geq \sqrt{2\gamma}$, we have $m \leq 2/\epsilon$.

If R is a circuit estimating p on average, then setting $\mathcal{C}$ to be a Reed-Muller polynomial code immediately gives a bound on the number of univariate degree d polynomials h which can agree with R on any fixed line.

In order to prove this Lemma, we will use a similar approach. In particular, we will show that if we have a set of $h_1, \ldots, h_m$ such that

$$\left| \Pr_{\overline{S} \to (x,i)}[E(h_i(L^{-1}(x)), i) = 1] - \frac{1}{2^m} \right| \geq \frac{\epsilon}{2^m}$$

then this will imply some statement we can plug into Theorem 4.

Note that the first issue we run into is that Theorem 4 requires indices to be drawn from the uniform distribution. However, it turns out that this theorem easily lifts to the case where x is sampled from an arbitrary high density distribution. In particular, we have

Proposition 5. *Consider $\mathcal{C}$ a $[N, D]_q$ code. That is, $\mathcal{C} \subseteq [q]^N$ and has a minimum distance between codewords of D.*
Let $R \in [q]^N$, $\mathcal{S}$ be a distribution of density δ. Define $\gamma := 1 - \frac{D}{N}$. Let $C_1, \ldots, C_m \in \mathcal{C}$ such that for all j,

$$\Pr_{\mathcal{S} \to x}[(C_j)_{(x)} = R_{(x)}] \geq \epsilon.$$

Then if $\epsilon \geq \sqrt{\frac{2\gamma}{\delta}}$, we have $m \leq 2/\epsilon$.

We prove this modified proposition in Section A.

In order to begin our reasoning, we define

$$B(x) := \{y : E(y, \mathcal{S}_x) \text{ is biased}\}$$

We can argue that for each h_i, we have

$$\Pr_{\overline{\mathcal{S}} \to x}[(h \circ L^{-1})(x) \in B(x)] \geq \epsilon'$$

for some ϵ' depending on ϵ. Intuitively, this is because we know that $E((h \circ L^{-1})(x), i)$ is biased, and so $(h \circ L^{-1})(x)$ should bias $\mathcal{S}_x$ with some reasonable probability over $\overline{\mathcal{S}} \to x$.z

Furthermore, we know by the definition of combinatorial list-decoding that if $\mathcal{S}_x$ has high density, then $B(x)$ is small. Thus, if we had that $\mathcal{S}_x$ has high density for every x, then an averaging argument would tell us that there is some subset $h_{i_1}, \ldots, h_{i_{m/|B(x)|}}$ and some index $i \in \{1, 2, \ldots, |B(x)|\}$ such that

$$\Pr_{\overline{\mathcal{S}} \to x}[h_{i_j}(x) = B(x)_{(i)}] \geq \frac{\epsilon'}{|B(x)|}$$

Unfortunately, it is not the case that $\mathcal{S}_x$ has high density for every x. However, it turns out that it is enough for us to show that $\mathcal{S}_x$ has high density with high probability over the choice of $\overline{\mathcal{S}} \to x$. This can be shown to be true via simple probability arguments.

That is, the outline of our proof of this claim goes as follows:
We define a set A of points on the line such that for all $a \in A$, $\mathcal{S}_a$ has high density and $\Pr_{\overline{\mathcal{S}} \to x}[x \in A]$ is high.

We then follow the reasoning above to show that there is a large subset of $h_1, \ldots, h_m$ such that each polynomial in the subset is biased to land in the the i^{th} element of $B(a)$ for all $a \in A$.

Finally, we apply Proposition 5 to the code of polynomials restricted to A.

Formal proof of Lemma 7:

For every non-trivial line L such that $\Pr_{\mathcal{S} \to (x,i)}[x \in L(\mathbb{F})] \geq \frac{1}{2}\frac{q}{q^t}$, $\overline{\mathcal{S}}$ has density $\overline{\delta} := \frac{\delta}{2}$

To see this, consider any $(x', i') \in L(\mathbb{F}) \times \{0,1\}^{n'}$.

$$\Pr_{\mathcal{S} \to (x,i)}[(x,i) = (x',i') | x \in L] = \frac{\Pr_{\mathcal{S} \to (x,i)}[(x,i) = (x',i')]}{\Pr_{\mathcal{S} \to (x,i)}[x \in L(\mathbb{F})]} \leq \frac{\frac{1}{\delta q^t 2^{n'}}}{\frac{q}{2q^t}} = \frac{1}{\overline{\delta} q 2^{n'}}$$

We define the α-heavy-set $A_\alpha \subseteq L(\mathbb{F})$ of $\overline{\mathcal{S}}$ as

$$A_\alpha := \{a \in L(\mathbb{F}) | \Pr[\overline{\mathcal{S}} \to (a, \cdot)] \geq \frac{\alpha}{\overline{\delta} q}\}$$

These are the points which have a particularly high probability of occurring.

We prove the following properties about A_α.

1. $\Pr_{\overline{\mathcal{S}} \to (x,i)}[x \in A_\alpha] \geq 1 - \gamma$, where $\gamma := \frac{\alpha}{1-\alpha}\frac{1-\overline{\delta}}{\overline{\delta}}$.
2. Let $\mathcal{S}_a = \overline{\mathcal{S}_a} := \overline{\mathcal{S}}|_{\{a\}}$. Then, $\mathcal{S}_a$ has density α.

3. $\mathcal{S}|_{A_\alpha} = \overline{\mathcal{S}}|_{A_\alpha}$ has density $\overline{\delta}(1-\gamma)$.
4. Define

$$B_\beta(a) := \left\{ b \in \mathbb{F} : \left| \Pr_{\mathcal{S}_a \to i}[E(b,i) = 1] - \frac{1}{2^m} \right| \geq \beta \right\}.$$

Then, if $\alpha \geq 2^{-0.1n'}, \beta \geq 2^{-0.2n'}$, for all $a \in A_\alpha$, $|B_\beta(a)| \leq 2^{0.2n'}$.

Proof of Property 1:

Proof. Observe that $1 \leq |A_\alpha| \cdot \frac{1}{\overline{\delta}q} + (q - |A_\alpha|) \cdot \frac{\alpha}{\overline{\delta}q}$, which implies that

$$1 - \frac{\alpha}{\overline{\delta}} \leq |A_\alpha| \cdot (\frac{1}{\overline{\delta}q} - \frac{\alpha}{\overline{\delta}q})$$

$$\begin{aligned} |A_\alpha| &\geq \frac{1 - \frac{\alpha}{\overline{\delta}}}{\frac{1}{\overline{\delta}q} - \frac{\alpha}{\overline{\delta}q}} \\ &\geq \frac{\frac{\overline{\delta}-\alpha}{\overline{\delta}}}{\frac{1-\alpha}{\overline{\delta}q}} \\ &\geq q\frac{\overline{\delta}-\alpha}{1-\alpha} \end{aligned}$$

Now, we can bound the probability of $x \notin A_\alpha$ by property 1, and its weight. Elements not in A_α have weight at most $\frac{\alpha}{\overline{\delta}q}$. Thus,

$$\begin{aligned} \Pr_{\overline{\mathcal{S}} \to (x,i)}[x \notin A_\alpha] &\leq (q - |A_\alpha|) \cdot \frac{\alpha}{\overline{\delta}q} \\ &\leq (q - q\frac{\overline{\delta}-\alpha}{1-\alpha}) \cdot \frac{\alpha}{\overline{\delta}q} \\ &= q(1 - \frac{\overline{\delta}-\alpha}{1-\alpha}) \cdot \frac{\alpha}{\overline{\delta}q} \\ &= \frac{1-\overline{\delta}}{1-\alpha} \cdot \frac{\alpha}{\overline{\delta}} \end{aligned}$$

Thus, $\Pr_{\overline{\mathcal{S}} \to (x,i)}[x \in A_\alpha] \geq 1 - \frac{1-\overline{\delta}}{1-\alpha} \cdot \frac{\alpha}{\overline{\delta}} \geq 1 - \gamma$, where $\gamma = \frac{1-\overline{\delta}}{1-\alpha} \cdot \frac{\alpha}{\overline{\delta}}$.

Proof of Property 2:

Proof. We know that $\overline{\mathcal{S}}_a = \overline{\mathcal{S}}|_{\{a\}}$. Thus,

$$\begin{aligned}
\Pr[\overline{\mathcal{S}}_a \to i'] &= \Pr[\overline{\mathcal{S}} \to (x,i) | x = a] \\
&= \frac{\Pr_{\overline{\mathcal{S}} \to (x,i)}[x = a \cap i = i']}{\Pr_{\overline{\mathcal{S}} \to (x,i)}[x = a]} \\
&= \frac{\frac{\overline{\delta}}{\overline{\delta} q |I|}}{\frac{\alpha}{\overline{\delta} q}} \\
&= \frac{1}{\alpha |I|}
\end{aligned}$$

Proof of Property 3:

Proof.

$$\begin{aligned}
\Pr_{\overline{S_{A_\alpha}} \to i}[i = i'] &= \Pr_{\overline{S} \to (x,i)}[(x,i) = (x',i') | x \in A] \\
&= \frac{\Pr_{\overline{S} \to (x,i)}[(x,i) = (x',i') \cap x \in A]}{\Pr[x \in A]} \\
&\leq \frac{\Pr_{\overline{S} \to (x,i)}[(x,i) = (x',i')]}{\Pr[x \in A]} \\
&\leq \frac{\frac{\overline{\delta}}{\overline{\delta} q |I|}}{1 - \gamma} \\
&\leq \frac{1}{\overline{\delta} q |I| (1 - \gamma)} \\
&\leq \overline{\delta}(1 - \gamma)
\end{aligned}$$

Proof of Property 4:

Proof. Property 4 follows immediately from the definition of combinatorial list-decoding as well as the fact that $\overline{\mathcal{S}}_a$ has density α.

To prove Lemma 6, we introduce the following $[r, r-d]_q$ code where $r = |A_\alpha| \geq q\frac{\overline{\delta} - \alpha}{1 - \alpha}$(from property 1):
For a polynomial $h : \mathbb{F} \to \mathbb{F}$ of degree $d+1$, let

$$C_h = h(L^{-1}(a_1)), h(L^{-1}(a_2)), \ldots h(L^{-1}(a_r))$$

This code is in $[q]^r$. Every codeword of this code is defined by a different polynomial $h : \mathbb{F} \to \mathbb{F}$ of degree $d+1$. Since 2 polynomials of degree $d+1$ can agree on at most d points, the distace between any 2 codewords is atleast $r - d$.

First, we will prove this lemma for the case when there exist m polynomials $\{f_1, \ldots, f_m\} : \mathbb{F} \to \mathbb{F}$ such that,

$$\forall j \in \{1, \ldots, m\}, \Pr_{\overline{S} \to (x,i)}[\langle f_j(L^{-1}(x)), i \rangle = 1] \geq \frac{1}{2^m} + \frac{\epsilon}{6 \cdot 2^m}$$

Then, $\Pr_{\overline{S}\to(x,i)}[\langle f_j(L^{-1}(x)), i\rangle = 1]$

$$= \Pr_{\overline{S}|_A\to(a,i)}[\langle f_j(L^{-1}(a)), i\rangle = 1] \cdot \Pr \overline{S} \to x[x \in A] + \Pr_{\overline{S}_{|\overline{A^c}}\to(x,i)}[\langle f_j(L^{-1}(x)), i\rangle = 1] \cdot \Pr_{\overline{S}\to x}[x \notin A]$$
$$\leq \Pr_{\overline{S}|_A\to(a,i)}[\langle f_j(L^{-1}(a)), i\rangle = 1] + \Pr_{\overline{S}\to x}[x \notin A]$$

Rearranging the equations,

$$\Pr_{\overline{S}|_A\to(a,i)}[\langle f_j(L^{-1}(a)), i\rangle = 1] \geq \Pr_{\overline{S}\to(x,i)}[\langle f_j(L^{-1}(x)), i\rangle = 1] - \Pr_{\overline{S}\to x}[x \notin A] \geq \frac{1}{2^m} + \frac{\epsilon}{6\cdot 2^m} - \Pr_{\overline{S}\to x}[x \notin A]$$

Now, we will lower bound the $\Pr_{\overline{S}|_A\to(a,i)}[f_j(L^{-1}(a)) \in B_{\alpha,\beta}(a)]$.

We have,

$$\begin{aligned}\Pr_{\overline{S}|_A\to(a,i)}[\langle f_j(L^{-1}(a)), i\rangle = 1] \leq & \Pr_{\overline{S}|_A\to(a,i)}[f_j(L^{-1}(a)) \in B_{\alpha,\beta}(a)] \\ & + \Pr_{\overline{S}|_A\to(a,i)}[\langle f_j(L^{-1}(a)), i\rangle = 1 | f_j(L^{-1}(a)) \notin B_{\alpha,\beta}(a)] \\ \leq & \Pr_{\overline{S}|_A\to(a,i)}[f_j(L^{-1}(a)) \in B_{\alpha,\beta}(a)] + \frac{1}{2^m} + \beta,\end{aligned}$$

where the last inequality follows from property 4.
Thus,

$$\begin{aligned}\Pr_{\overline{S}|_A\to(a,i)}[f_j(L^{-1}(a)) \in B_{\alpha,\beta}(a)] &\geq \Pr_{\overline{S}|_A\to(a,i)}[\langle f_j(L^{-1}(a)), i\rangle = 1] - \frac{1}{2^m} - \beta \\ &\geq \frac{1}{2^m} + \frac{\epsilon}{6\cdot 2^m} - \Pr_{\overline{S}\to x}[x \notin A] - \frac{1}{2^m} - \beta \\ &\geq \frac{\epsilon}{6\cdot 2^m} - \beta - \gamma \\ &=: \epsilon^*,\end{aligned}$$

where the last inequality follows from property 1.
Intuitively this implies that the probability that the set of functions $\{f_1, \ldots, f_m\}$ is in the set of biased functions B_β is at least ϵ^*.
Now, we define $B_\beta(a) = \{b_1(a), b_2(a), \ldots b_s(a)\}$. Since $\forall j \in \{1, 2, \ldots, m\}, \Pr_{\overline{S}|_A\to(a,i)}[f_j(L^{-1}(a)) \in B_\beta(a)] \geq \epsilon^*$, it implies that for each j, there must exist an index z_j such that $\Pr_{\overline{S}|_A\to(a,i)}[f_j(L^{-1}(a)) = b_{z_j}(a)] \geq \frac{\epsilon^*}{s}$.
An averaging argument gives us that there must exist an index z such that at least m/s of the functions in the set $\{f_1, \ldots, f_m\}$ (call this set $\{g_1, g_2, \ldots, g_{m/s}\}$) are such that

$$\Pr_{\overline{S}|_A}[g_j(L^{-1}(a)) = b_z(a)] \geq \frac{\epsilon^*}{s}$$

If this is not true, since z comes from a set of size s then the total number of functions is $< m$ which is a contradiction.

From the definition of the code we defined, $\{b_z(a)\}_{a\in A_\alpha} \subseteq [q]^s$. To apply Proposition 6, we can let $R_a = b_z(a)$. In our case, $\epsilon = \dfrac{\epsilon^*}{s}$ and $\gamma = \dfrac{d}{r}$. We also know that $\delta = \overline{\delta}(1-\gamma)$ from property 3. Therefore, if $\dfrac{\epsilon^*}{s} \geq \dfrac{\sqrt{2d}}{\sqrt{\overline{\delta}(1-\gamma)\cdot r}}$, then $\dfrac{m}{s} \leq \dfrac{2s}{\epsilon^*}$, which implies that $m \leq \dfrac{2s^2}{\epsilon^*}$.

Now, we will set the parameters α and β so that the condition $\dfrac{\epsilon^*}{s} \geq \dfrac{\sqrt{2d}}{\sqrt{\overline{\delta}(1-\gamma)\cdot r}}$ is satisfiable and such that property 4 holds.

Let $\alpha = \dfrac{\epsilon\delta}{128\cdot 2^m}, \beta = \dfrac{\epsilon}{16\cdot 2^m}$. If $\dfrac{\epsilon\delta}{128\cdot 2^m} \geq 2^{-0.1n'}$ then property 4 holds. Recall that $\epsilon^* = \dfrac{\epsilon}{6\cdot 2^m} - \beta - \gamma$. Let $\dfrac{\epsilon}{6\cdot 2^m} = \epsilon'$.
Rearranging this, our requirement becomes:

$$\begin{aligned}
\epsilon' &= \epsilon^* + \beta + \gamma \\
&\geq \frac{s\sqrt{2d}}{\sqrt{\overline{\delta}(1-\gamma)\cdot r}} + \beta + \gamma \\
&= \frac{s\sqrt{4d}}{\sqrt{\delta(1-\gamma)\cdot r}} + \beta + \gamma \\
&= \frac{s\sqrt{4d}}{\sqrt{\delta(1-\frac{1-\overline{\delta}}{1-\alpha}\cdot\frac{\alpha}{\overline{\delta}})\cdot q^{\frac{\overline{\delta}-\alpha}{1-\alpha}}}} + \beta + \frac{1-\overline{\delta}}{1-\alpha}\cdot\frac{\alpha}{\overline{\delta}} \\
&= \frac{s\sqrt{4d}}{\sqrt{\delta(1-\frac{2-\delta}{1-\alpha}\cdot\frac{\alpha}{\overline{\delta}})\cdot q^{\frac{\frac{\delta}{2}-\alpha}{1-\alpha}}}} + \beta + \frac{2-\delta}{1-\alpha}\cdot\frac{\alpha}{\overline{\delta}} \\
&= 2^{0.2n'}\frac{\sqrt{4d}}{\sqrt{\delta(1-\frac{2-\delta}{1-\alpha}\cdot\frac{\alpha}{\overline{\delta}})\cdot q^{\frac{\frac{\delta}{2}-\alpha}{1-\alpha}}}} + \beta + \frac{2-\delta}{1-\alpha}\cdot\frac{\alpha}{\overline{\delta}}
\end{aligned}$$

Observe that the following hold,

$$\frac{2-\delta}{1-\alpha}\cdot\frac{\alpha}{\overline{\delta}} \leq \frac{\epsilon}{16\cdot 2^m}$$

$$\beta \leq \frac{\epsilon}{16\cdot 2^m}$$

and

$$\frac{\frac{\delta}{2}-\alpha}{1-\alpha} \leq \frac{\delta}{2}$$

Substituting these values, in our expression for ϵ', we get the new requirement:

$$\epsilon' \geq 2^{0.2n'}\sqrt{\frac{d}{q}\frac{4}{\delta}} + \frac{\epsilon}{16\cdot 2^m} + \frac{\epsilon}{16\cdot 2^m}$$

If

$$\frac{\sqrt{d}}{\sqrt{q}} \leq \frac{\delta}{8} \frac{\frac{\epsilon}{6\cdot 2^m} - \frac{\epsilon}{16\cdot 2^m} - \frac{\epsilon}{16\cdot 2^m}}{2^{0.2n'}} = \frac{5\epsilon\delta}{48\cdot 2^{0.2n'+m}},$$

then our requirement is satisfied. Then, since $\epsilon^* = \epsilon' - \beta - \frac{1-\overline{\delta}}{1-\alpha}\cdot\frac{\alpha}{\overline{\delta}}$, we get:

$$\epsilon^* \geq \frac{\epsilon}{6\cdot 2^m} - \frac{\epsilon}{8\cdot 2^m} - \frac{\epsilon}{8\cdot 2^m} \geq \frac{5\epsilon}{48\cdot 2^m} \geq \frac{\epsilon}{8\cdot 2^m}$$

This implies that,

$$m \leq \frac{2s^2}{\epsilon^*} \leq \frac{16\cdot 2^{0.4n'+m}}{\epsilon}.$$

and so we are done.

For the other case, we now assume that there exist m polynomials $\{f_1,\ldots,f_m\} : \mathbb{F}\to\mathbb{F}$ such that,

$$\forall j\in\{1,\ldots,m\}, \Pr_{\overline{S}\to(x,i)}[\langle f_j(L^{-1}(x)), i\rangle = 1] \leq \frac{1}{2^m} - \frac{\epsilon}{6\cdot 2^m}$$

This implies that $\forall j\in\{1,\ldots,m\}, \Pr_{\overline{S}\to(x,i)}\left[\langle f_j(L^{-1}(x)), i\rangle \neq 1\right] \geq 1 - \frac{1}{2^m} + \frac{\epsilon}{6\cdot 2^m}$.

Then, $\Pr_{\overline{S}\to(x,i)}[\langle f_j(L^{-1}(x)), i\rangle \neq 1]$

$$= \Pr_{\overline{S}|_A\to(a,i)}[\langle f_j(L^{-1}(a)), i\rangle \neq 1]\cdot \Pr\overline{S}\to x[x\in A] + \Pr_{\overline{S}_{|A^c}\to(x,i)}[\langle f_j(L^{-1}(x)), i\rangle \neq 1]\cdot \Pr_{\overline{S}\to x}[x\notin A]$$
$$\leq \Pr_{\overline{S}|_A\to(a,i)}[\langle f_j(L^{-1}(a)), i\rangle \neq 1] + \Pr_{\overline{S}\to x}[x\notin A]$$

Rearranging the equations,

$$\Pr_{\overline{S}|_A\to(a,i)}[\langle f_j(L^{-1}(a)), i\rangle \neq 1] \geq \Pr_{\overline{S}\to(x,i)}[\langle f_j(L^{-1}(x)), i\rangle \neq 1] - \Pr_{\overline{S}\to x}[x\notin A] \geq 1 - \frac{1}{2^m} + \frac{\epsilon}{6\cdot 2^m} - \Pr_{\overline{S}\to x}[x\notin A]$$

Now, we will lower bound $\Pr_{\overline{S}|_A\to(a,i)}[f_j(L^{-1}(a))\in B_{\alpha,\beta}(a)]$.

We can re-write $\Pr_{\overline{S}|_A\to(a,i)}[\langle f_j(L^{-1}(a)), i\rangle \neq 1]$ as

$$\Pr_{\overline{S}|_A\to(a,i)}[\langle f_j(L^{-1}(a)), i\rangle \neq 1 | f_j(L^{-1}(a))\in B_{\alpha,\beta}(a)]\cdot \Pr_{\overline{S}|_A\to(a,i)}[f_j(L^{-1})\in B_{\alpha,\beta}(a)]$$
$$+ \Pr_{\overline{S}|_A\to(a,i)}[\langle f_j(L^{-1}(a)), i\rangle \neq 1 | f_j(L^{-1}(a))\notin B_{\alpha,\beta}(a)]\cdot \Pr_{\overline{S}|_A\to(a,i)}[f_j(L^{-1}(a))\notin B_{\alpha,\beta}(a)]$$

Therefore,

$$\Pr_{\overline{S}|_A\to(a,i)}[\langle f_j(L^{-1}(a)), i\rangle \neq 1] \leq \Pr_{\overline{S}|_A\to(a,i)}[f_j(L^{-1}(a))\in B_{\alpha,\beta}(a)]$$
$$+ \Pr_{\overline{S}|_A\to(a,i)}[\langle f_j(L^{-1}(a)), i\rangle \neq 1 | f_j(L^{-1}(a))\notin B_{\alpha,\beta}(a)]$$

Observe that $\Pr_{\overline{S}|_A \to (a,i)}[\langle f_j(L^{-1}(a)), i\rangle \neq 1 | f_j(L^{-1}(a)) \notin B_{\alpha,\beta}(a)]$

$$\begin{aligned}
&= 1 - \Pr_{\overline{S}|_A \to (a,i)}[\langle f_j(L^{-1}(a)), i\rangle = 1 | f_j(L^{-1}(a)) \notin B_{\alpha,\beta}(a)] \\
&\leq 1 - (\frac{1}{2^m} - \beta) \\
&\leq 1 - \frac{1}{2^m} + \beta
\end{aligned}$$

where the first inequality follows from property 4. Thus,

$$\begin{aligned}
\Pr_{\overline{S}|_A \to (a,i)}[f_j(L^{-1}(a)) \in B_{\alpha,\beta}(a)] &\geq \Pr_{\overline{S}|_A \to (a,i)}[\langle f_j(L^{-1}(a)), i\rangle \neq 1] - 1 + \frac{1}{2^m} - \beta \\
&\geq 1 - \frac{1}{2^m} + \frac{\epsilon}{6 \cdot 2^m} - \Pr_{\overline{S} \to x}[x \notin A] - 1 + \frac{1}{2^m} - \beta \\
&\geq \frac{\epsilon}{6 \cdot 2^m} - \beta - \gamma \\
&= \epsilon^*,
\end{aligned}$$

where the last inequality follows from property 1.
Intuitively this implies that the probability that the set of functions $\{f_1, \ldots, f_m\}$ is in the set of biased functions $B_{\alpha,\beta}$ is at least ϵ^*.
Now, we define $B_{\alpha,\beta}(a) = \{b_1(a), b_2(a), \ldots b_s(a)\}$. Since $\forall j \in \{1, 2, \ldots, m\}, \Pr_{\overline{S}|_A \to (a,i)}[f_j(L^{-1}(a)) \in B_{\alpha,\beta}(a)] \geq \epsilon^*$, it implies that for each j, there must exist an index z_j such that $\Pr_{\overline{S}|_A \to (a,i)}[f_j(L^{-1}(a)) = b_{z_j}(a)] \geq \frac{\epsilon^*}{s}$.
An averaging argument gives us that there must exist an index z such that at least m/s of the functions in the set $\{f_1, \ldots, f_m\}$ (call this set $\{g_1, g_2, \ldots, g_{m/s}\}$) are such that

$$\Pr_{\overline{S}|_A}[g_j(L^{-1}(a)) = b_z(a)] \geq \frac{\epsilon^*}{s}$$

If this is not true, since z comes from a set of size s then the total number of functions is $< m$ which is a contradiction. From the definition of the code we defined, $\{b_z(a)\}_{a \in A_\alpha} \subseteq [q]^s$. To apply Proposition 6, we can let $R_a = b_z(a)$. In our case, $\epsilon = \frac{\epsilon^*}{s}$ and $\gamma = \frac{d}{r}$. We also know that $\delta = \overline{\delta}(1-\gamma)$ from property 3.
The rest of the proof is identical to the previous case.

6 Leakage

Theorem 9. *Let EXT be a (k, ϵ)-deterministic extractor against nondeterministic sources samplable by size-s (quantum) circuits. Then, for all $c > 0$, EXT is a leakage-resilient $(k + c, \epsilon + 2^{-c})$-deterministic extractor against nondeterministic sources samplable by size-$O(s)$ (quantum) circuits.*

Lemma 8. *Let X, L be any random variables such that $H_\infty(X|L) \geq k$. Then, for any constant $c > 0$, $\Pr_{L \to \ell}[H_\infty(X|L = \ell) \geq k - c] \geq 1 - \frac{1}{2^c}$.*

Proof. We immediately have that

$$\mathbb{E}_{L\to\ell}[\max_x \Pr[X = x|L = \ell]] \leq 2^{-k}$$

Markov's inequality then gives us that

$$\Pr_{L\to\ell}[\max_x \Pr[X = x|L = \ell] \geq 2^{-k+c}] \leq \frac{1}{2^c}$$

which is exactly what we want.

Proof. Fix any source $\mathcal{S} \to (X, L)$ of min-entropy $k+c$. For each possible leakage ℓ, we define a new nondeterministic source $X|_\ell$ defined by running $\mathcal{S} \to (X, L)$, conditioning on $L = \ell$, and outputting X. Note that the size of this source is linear in s.

We call an ℓ bad if $H_\infty(X_\ell) < k$. The lemma tells us that $\Pr_{L\to\ell}[\ell \text{ is bad}] \leq 2^{-c}$.

We can then calculate $\Delta((EXT(X), L), (U, L))$ as follows

$$\begin{aligned}
\Delta((EXT(X, L), L), (U, L)) &= \frac{1}{2}\sum_{x,\ell} |\Pr[EXT(X) = x \text{ and } L = \ell] - \Pr[U = x \text{ and } L = \ell]| \\
&= \frac{1}{2}\sum_{x,\ell} |\Pr[EXT(X) = x|L = \ell]\Pr[L = \ell] - \Pr[U = x]\Pr[L = \ell]| \\
&\leq \Pr[L \text{ "bad"}] + \frac{1}{2}\sum_{\ell \text{ "good"}} \left(\Pr[L = \ell]\sum_x |\Pr[EXT(X_\ell) = x] - \Pr[U = x]|\right) \\
&\leq 2^{-c} + \sum_{\ell \text{ "good"}} \Pr[L = \ell]\Delta((Ext(X_\ell), L), (U, L))
\end{aligned}$$

But since $H_\infty(X_\ell) \geq k$ for ℓ "good", we have that $\Delta((Ext(X_\ell), L), (U, L)) \leq \epsilon$ and so

$$\Delta((EXT(X, L), L), (U, L)) \leq 2^{-c} + \epsilon$$

References

1. Scott Aaronson. Quantum computing, postselection, and probabilistic polynomial-time, 2004.
2. Eric Allender, Michal Koucký, Detlef Ronneburger, and Sambuddha Roy. Derandomization and distinguishing complexity. In *18th Annual IEEE Conference on Computational Complexity (Complexity 2003), 7-10 July 2003, Aarhus, Denmark*, pages 209–220. IEEE Computer Society, 2003.
3. Benny Applebaum, Sergei Artemenko, Ronen Shaltiel, and Guang Yang. Incompressible functions, relative-error extractors, and the power of nondeterministic reductions. *computational complexity*, 25(2):349–418, Jun 2016.
4. Marshall Ball, Oded Goldreich, and Tal Malkin. Randomness extraction from somewhat dependent sources. In Mark Braverman, editor, *13th Innovations in Theoretical Computer Science Conference, ITCS 2022, January 31 - February 3, 2022, Berkeley, CA, USA*, volume 215 of *LIPIcs*, pages 12:1–12:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022.
5. Mihir Bellare, Oded Goldreich, and Erez Petrank. Uniform generation of np-witnesses using an np-oracle. *Information and Computation*, 163(2):510–526, 2000.
6. Mario Berta, Omar Fawzi, Volkher Scholz, and Oleg Szehr. Variations on classical and quantum extractors. In *2014 IEEE International Symposium on Information Theory*. IEEE, jun 2014.
7. Mario Berta, Omar Fawzi, and Stephanie Wehner. Quantum to classical randomness extractors. In Reihaneh Safavi-Naini and Ran Canetti, editors, *CRYPTO 2012*, volume 7417 of *LNCS*, pages 776–793. Springer, Heidelberg, August 2012.

8. Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity (extended abstract). In *26th Annual Symposium on Foundations of Computer Science, Portland, Oregon, USA, 21-23 October 1985*, pages 429–442. IEEE Computer Society, 1985.
9. Anindya De, Christopher Portmann, Thomas Vidick, and Renato Renner. Trevisan's extractor in the presence of quantum side information. *SIAM Journal on Computing*, 41(4):915–940, jan 2012.
10. Yevgeniy Dodis, Ariel Elbaz, Roberto Oliveira, and Ran Raz. Improved randomness extraction from two independent sources. In Klaus Jansen, Sanjeev Khanna, José D. P. Rolim, and Dana Ron, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 334–344, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.
11. Zeev Dvir. Extractors for varieties. In *Proceedings of the 24th Annual IEEE Conference on Computational Complexity, CCC 2009, Paris, France, 15-18 July 2009*, pages 102–113. IEEE Computer Society, 2009.
12. Zeev Dvir, Ariel Gabizon, and Avi Wigderson. Extractors and rank extractors for polynomial sources. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2007), October 20-23, 2007, Providence, RI, USA, Proceedings*, pages 52–62. IEEE Computer Society, 2007.
13. Oded Goldreich, Ronitt Rubinfeld, and Madhu Sudan. Learning polynomials with queries: The highly noisy case. In *36th FOCS*, pages 294–303. IEEE Computer Society Press, October 1995.
14. Zeyu Guo, Ben Lee Volk, Akhil Jalan, and David Zuckerman. Extractors for images of varieties. In Barna Saha and Rocco A. Servedio, editors, *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023, Orlando, FL, USA, June 20-23, 2023*, pages 46–59. ACM, 2023.
15. Mark R. Jerrum, Leslie G. Valiant, and Vijay V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science*, 43:169–188, 1986.
16. Leonid A. Levin. Average case complete problems. *SIAM J. Comput.*, 15(1):285–286, 1986.
17. Fu Li and David Zuckerman. Improved extractors for recognizable and algebraic sources. In Dimitris Achlioptas and László A. Végh, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2019, September 20-22, 2019, Massachusetts Institute of Technology, Cambridge, MA, USA*, volume 145 of *LIPIcs*, pages 72:1–72:22. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019.
18. Peter Bro Miltersen and N. V. Vinodchandran. Derandomizing arthur-merlin games using hitting sets. In *40th Annual Symposium on Foundations of Computer Science, FOCS '99, 17-18 October, 1999, New York, NY, USA*, pages 71–80. IEEE Computer Society, 1999.
19. Peter Bro Miltersen and N. V. Vinodchandran. Derandomizing arthur-merlin games using hitting sets. In *40th Annual Symposium on Foundations of Computer Science, FOCS '99, 17-18 October, 1999, New York, NY, USA*, pages 71–80. IEEE Computer Society, 1999.
20. Renato Renner. Security of quantum key distribution, 2006.
21. Ronen Shaltiel. Weak derandomization of weak algorithms: Explicit versions of yao's lemma. In *Proceedings of the 24th Annual IEEE Conference on Computational Complexity, CCC 2009, Paris, France, 15-18 July 2009*, pages 114–125. IEEE Computer Society, 2009.
22. Ronen Shaltiel and Christopher Umans. Simple extractors for all min-entropies and a new pseudo-random generator. In *42nd Annual Symposium on Foundations of Computer Science, FOCS 2001, 14-17 October 2001, Las Vegas, Nevada, USA*, pages 648–657. IEEE Computer Society, 2001.
23. Ronen Shaltiel and Christopher Umans. Pseudorandomness for approximate counting and sampling. In *20th Annual IEEE Conference on Computational Complexity (CCC 2005), 11-15 June 2005, San Jose, CA, USA*, pages 212–226. IEEE Computer Society, 2005.
24. Ronen Shaltiel and Christopher Umans. Pseudorandomness for approximate counting and sampling. *Comput. Complex.*, 15(4):298–341, 2006.
25. Madhu Sudan, Luca Trevisan, and Salil P. Vadhan. Pseudorandom generators without the XOR lemma (extended abstract). In *31st ACM STOC*, pages 537–546. ACM Press, May 1999.
26. Luca Trevisan and Salil P. Vadhan. Extracting randomness from samplable distributions. In *41st FOCS*, pages 32–42. IEEE Computer Society Press, November 2000.
27. David Zuckerman. Randomness-optimal oblivious sampling. *Random Structures & Algorithms*, 11(4):345–367, 1997.

A Modified list-decoding lemma

Proposition 6. *Modified from [13].*
Consider $\mathcal{C}$ a $[N, D]_q$ code. That is, $\mathcal{C} \subseteq [q]^N$ and has a minimum distance between codewords of D.

Let $R \in [q]^N$, $\mathcal{S}$ be a distribution of density δ over $[N]$. Define $\gamma := 1 - \frac{D}{N}$. Let $C_1, \ldots, C_m \in \mathcal{C}$ such that for all j,

$$\Pr_{\mathcal{S} \to x}[(C_j)_{(x)} = R_{(x)}] \geq \epsilon.$$

Then if $\epsilon \geq \sqrt{\frac{2\gamma}{\delta}}$, we have $m \leq 2/\epsilon$.

Proof. Let $\chi_j(x)$ be an indicator for the event that $(C_j)_{(x)} = R_{(x)}$. Inclusion-exclusion gives us that, for any $m' \leq m$,

$$\begin{aligned}
1 &\geq \Pr_{\mathcal{S} \to x}[\exists\ j \leq m' : \chi_j(x)] \\
&= \sum_{j \leq m'} \Pr_{\mathcal{S} \to x}[\chi_j(x)] - \sum_{j_1 \neq j_2 \leq m'} \Pr_{\mathcal{S} \to x}[\chi_{j_1}(x) \cdot \chi_{j_2}(x)] \\
&\geq \sum_{j \leq m'} \Pr_{\mathcal{S} \to x}[\chi_j(x)] - \sum_{j_1 \neq j_2 \leq m'} \Pr_{\mathcal{S} \to x}[(C_{j_1})_{(x)} = (C_{j_2})_{(x)}] \\
&\geq m'\epsilon - \sum_{j_1 \neq j_2 \leq m'} \Pr_{\mathcal{S} \to x}[(C_{j_1})_{(x)} = (C_{j_2})_{(x)}]
\end{aligned}$$

And so, if we get a bound on $\Pr_{\mathcal{S} \to x}[(C_{j_1})_{(x)} = (C_{j_2})_{(x)}]$, we can get an equation bounding m'. But note that

$$\begin{aligned}
\Pr_{\mathcal{S} \to x}[(C_{j_1})_{(x)} = (C_{j_2})_{(x)}] &= \sum_x \Pr[\mathcal{S} \to x]\, \mathbb{1}_{(C_{j_1})_{(x)} = (C_{j_2})_{(x)}} \\
&\leq \sum_x \frac{1}{\delta N}\, \mathbb{1}_{(C_{j_1})_{(x)} = (C_{j_2})_{(x)}} \\
&= \frac{1}{\delta} \Pr_{U \to x}[(C_{j_1})_{(x)} = (C_{j_2})_{(x)}] \\
&\leq \frac{\gamma}{\delta}
\end{aligned}$$

where the last inequality comes from the fact that $\Pr_{U \to x}[(C_{j_1})_{(x)} = (C_{j_2})_{(x)}] \leq \frac{D - N}{N} = \gamma$.

This then gives us the equation

$$1 \geq m'\epsilon - \frac{m'(m'-1)}{2}\gamma$$

Define $\beta = \frac{\gamma}{2\delta}$ and $g(y) := \beta y^2 - (\epsilon + \beta)y + 1$. The equation above tells us that $g(y) \geq 0$ for all $m' \leq m$.

Looking closely at the proof of Theorem 4.2 from [13], we see that they show the following properties of the roots α_1, α_2 of g, conditioned on $(\beta + \epsilon)^2 - 4\beta > \beta^2$:

1. α_1, α_2 are both non-negative reals
2. $|\alpha_1 - \alpha_2| > 1$
3. $\min(\alpha_1, \alpha_2) < \frac{2}{\epsilon + \beta}$

But note that as long as $\epsilon \geq \sqrt{\frac{2\gamma}{\delta}}$, we have $(\beta + \epsilon)^2 - 4\beta > \beta^2$. Thus, as for all $m' < m$, $g(m') > 0$, we must have that $m \leq min(\alpha_1, \alpha_2) < \frac{2}{\epsilon + \beta} \leq \frac{2}{\epsilon}$.

B Quantum probability estimation requires postselection

Let f be a boolean function decided by a postselecting circuit C. We observe that if we can solve probability estimation for C when viewed as a quantum circuit, then we can evaluate f. This means that to solve probability estimation generally, we require a model of computation stronger than postselecting circuits. Formally,

Proposition 7. *Let $f : \{0,1\}^n \to \{0,1\}$ be a boolean function. Let C be a quantum circuit of size s such that for all x*

$$\Pr_{C(x)\to(y,b)}[y = f(x)|b = 1] \geq \frac{2}{3}$$

Then, if there exists a quantum circuit C' of size $poly(s)$ such that for some sufficiently small constant ϵ, and for all $x \in \{0,1\}^n, y \in \{0,1\}, b \in \{0,1\}$

$$\Pr_{C'(x,y,b)\to\widetilde{p}}[\widetilde{p} \in (1 \pm \epsilon) \Pr_{C(x)\to(y',b')}[(y',b') = (y,b)]] \geq \frac{2}{3}$$

then there exists a quantum circuit C'' of size $poly(s)$ such that

$$\Pr_{C''(x)\to y}[C''(x) = f(x)] \geq \frac{2}{3}$$

Proof. First, we will amplify C' so that it succeeds with probability $\frac{99}{100}$. Define $C''(x)$ as follows:
-Run $C'(x,1,1) \to \widetilde{p}_1$ and $C'(x,0,1) \to \widetilde{p}_0$
-If $\widetilde{p}_0 \geq \widetilde{p}_1$, output 1. Otherwise, output 0.

Let $\gamma = \Pr_{C(x)\to(y,b)}[b = 1]$. Note that $\Pr_{C(x)\to(y,b)}[y = f(x)|b = 1] \geq \frac{99}{100}$ and so we have

$$\Pr_{C(x)\to(y,b)}[y = f(x) \text{ and } b = 1] \geq \frac{2}{3}\gamma.$$

Similarly,

$$\Pr_{C(x)\to(y,b)}[y = f(x) \text{ and } b = 1] \leq \frac{1}{3}\gamma.$$

Thus, with probability $\geq \left(\frac{99}{100}\right)^2$, for sufficiently small ϵ we will have that $\widetilde{p}_{f(x)} \geq (1-\epsilon)\frac{2}{3}\gamma \geq \frac{1}{2}\gamma$ and $\widetilde{p}_{1-f(x)} \leq (1+\epsilon)\frac{1}{3}\gamma < \frac{1}{2}\gamma$. Therefore, C'' evaluates f correctly with probability $\left(\frac{99}{100}\right)^2 \geq \frac{2}{3}$.

We also show how to achieve probability estimation using postselection. Formally,

Proposition 8. *For all s, $\epsilon > 0$, there exists a circuit C such that, if $\mathcal{S}$ is a random source samplable by a size-s circuit $\widetilde{C}$, then*

$$\Pr_{C(\widetilde{C},x)\to\widetilde{p}}[\widetilde{p} \in \left(1 \pm \frac{1}{2}\right) \Pr[\mathcal{S} \to x]] \geq \frac{2}{3}$$

Proof. Note that estimating the probability that $\widetilde{C}$ outputs x is the same as estimating the probability that the circuit defined by running $\widetilde{C}$ and checking if the output is x outputs 1. Thus, without loss of generality we will assume that $\widetilde{C}$ is boolean.

Let $|\psi_0\rangle = a\,|0\rangle\,|\phi_0\rangle + b\,|1\rangle\,|\phi_1\rangle$ be the state output by $\widetilde{C}$ before measuring the output register.

For any constants c, d, we can produce the state

$$\begin{aligned} |\psi_1\rangle &\propto c\,|0\rangle\,(X \otimes I)\,|\psi_1\rangle + d\,|1\rangle\,|\psi_1\rangle \\ &= bc\,|00\rangle\,|\phi_1\rangle + ac\,|01\rangle\,|\phi_0\rangle + ad\,|10\rangle\,|\phi_0\rangle + bd\,|11\rangle\,|\phi_1\rangle \end{aligned}$$

Postselecting on the second qubit being 1 gives

$$|\psi_2\rangle \propto ac\,|01\rangle\,|\phi_0\rangle + bd\,|11\rangle\,|\phi_1\rangle$$

Measuring the first qubit then outputs 1 with probability $\sigma := \dfrac{b^2d^2}{a^2c^2 + b^2d^2}$. In particular, we will consider the case where $c^2 = 1$, $d^2 = 2^i$.

If $\Pr[\widetilde{C} \to 1] = b^2 \geq \dfrac{1}{2^i}$, then

$$\sigma = \frac{b^2d^2}{a^2c^2 + b^2d^2} \geq \frac{1}{a^2c^2 + 1} \geq \frac{1}{2}$$

If $\dfrac{1}{2^{i+2}} \leq b^2 \leq \dfrac{1}{2^{i+1}}$, then $a^2 \geq 1 - \dfrac{1}{2^{i+1}}$ and so

$$a^2c^2 + b^2d^2 \geq \frac{1}{4} + 1 - \frac{1}{2^{i+1}} \geq \frac{9}{8}.$$

This then means that

$$\sigma \leq \frac{\frac{1}{2}}{\frac{9}{8}} \leq \frac{4}{9}$$

If $b^2 \leq \dfrac{1}{2^{i+2}}$, then

$$\sigma \leq \frac{1}{4} \leq \frac{4}{9}$$

Since there is a constant gap between $\dfrac{1}{2}$ and $\dfrac{4}{9}$, this technique will allow us to detect using a polynomial size postselecting circuit whether $\Pr[\widetilde{C} \to 1] \leq \dfrac{1}{2^i}$. Trying every i from 1 to s provides an algorithm for probability estimation.

C Formal proofs of main theorems from Theorem 8

Theorem 10. *If there is a problem in $E = DTIME(2^{O(n)})$ with $NP_{||}$-circuit complexity $2^{\Omega(n)}$, there exists a constant $\delta > 0$ such that for every constant $c > 0$, for every sufficiently large n and for every $m \leq \omega \log n$, there is a $\left((1-\delta)n, \dfrac{1}{n^c}\right)$ deterministic extractor* $\mathrm{Ext} : \{0,1\}^n \to \{0,1\}^m$ *against sources samplable by size n^c circuits. Furthermore,* Ext *is computable in time* $\mathrm{poly}(n^c)$.

Proof. We know that there exists constants γ, γ' such that for all ℓ, there is a function $f : \{0,1\}^\ell \to \{0,1\}$ which requires circuits of size $2^{\gamma\ell}$ and is computable in time $2^{\gamma'\ell}$. Without loss of generality, we will assume $\gamma \leq 1$.

Let $\alpha \in (0,1)$ be a constant to be set later. Set $\ell = \left\lceil \frac{c}{\gamma\alpha} \log n \right\rceil$ and set $s \coloneqq n^{\frac{c}{\alpha}}$, $s' \coloneqq n^c$. Observe that

$$s = 2^{\frac{c}{\alpha}\log n} \leq 2^{\gamma\ell}$$

and so $f : \{0,1\}^\ell \to \{0,1\}$ is hard for circuits of size s.

Let $t \coloneqq \lceil \ell / \log s' \rceil, n' \coloneqq \lceil n/(t+1) \rceil, \mathbb{F} :- GF(q \coloneqq 2^{n'})$. Let $\tau : \{0,1\}^\ell \to [s']^t$ be any natural injective map. There exists a polynomial $p : \mathbb{F}^t \to \mathbb{F}$ of degree at most s' in each variable such that for all $x \in \{0,1\}^\ell$, $f(x) = p(\tau(x))$. p can be computed in time $\text{poly}(n, 2^\ell)$ and has total degree $\leq d \coloneqq s' \cdot t$.

Let $E : \{0,1\}^{n'} \times \{0,1\}^{n'} \to \{0,1\}^m$ be the function from Claim 3 satisfying $(2^{-0.1n'}, 2^{0.2n'}, 2^{-0.1n'})$-combinatorial list decoding computable in time $\text{poly}(n')$. We define $\text{Ext} : \{0,1\}^n \to \{0,1\}^m$ by

$$\text{Ext}(x, i) = E(p(x), i \circ 0^{(t+1)n'-n}).$$

Note that this exists because $m \leq \omega \log n \leq 0.1n'$ for all sufficiently large n.

Set $\epsilon \coloneqq \frac{1}{s'}$. Suppose that there is a distribution X on $\{0,1\}^n$ with min-entropy $n \cdot (1 - (\alpha \log s')/\ell)$ samplable by size s' postselecting circuits such that for some x, $\left| \Pr[\text{Ext}^f_{n,\ell,s}(X) = x] - \frac{1}{2^m} \right| \geq \frac{\epsilon}{2^m}$. Define $X' = X \circ 0^{(t+1)n'-n}$. Then X' has density at least

$$\delta = \frac{2^{n \cdot (1 - (\alpha \log s')/\ell)}}{2^{(t+1)n'}}$$

And so, as long as

$$\begin{aligned} m &\leq c_E n' \\ \frac{d}{q} &\leq \frac{\epsilon^2 \delta^2}{64 \cdot 2^{(0.4n' + 2m)}} \\ \epsilon\delta &\geq 128 \cdot 2^{-(0.1n' - m)} \end{aligned}$$

applying Theorem 8 gives us that if Ext is not a $\left(\frac{1}{s'}, (1 - \alpha \log s'/\ell)n\right)$-extractor, then we have a circuit of size $\text{poly}(s', 2^m)$ computing f everywhere. Then as long as $\text{poly}(s', 2^m) \leq s$ and the conditions hold, then Ext is a $\left(\frac{1}{s'}, (1 - \alpha \log s'/\ell)n\right)$-extractor against $NP_{||}$-circuits of size s'.

To begin to show these, we observe the following properties:

$$\begin{aligned} t = \lceil \ell / \log s' \rceil &= \left\lceil \frac{1}{\gamma\alpha} \right\rceil \\ \frac{\log s'}{\ell} &\leq \frac{\log s'}{\frac{1}{\gamma\alpha} \log s'} = \gamma\alpha \\ n' = \left\lceil \frac{n}{t+1} \right\rceil &\geq \frac{n}{t+1} \geq \frac{n}{2t} \geq \frac{\gamma\alpha}{4} n \\ \delta = 2^{-(t+1)n' + n - (\alpha n \log s')/\ell} &\geq 2^{-(t+1) - \alpha n \log s'/\ell} \geq 2^{-\frac{2}{\gamma\alpha} - \gamma\alpha^2 n} \end{aligned}$$

Note that the condition $\frac{d}{q} \leq \frac{\epsilon^2\delta^2}{64 \cdot 2^{(0.4n'+2m}}$ holds as long as

$$\frac{s't}{2^{n'}} \leq \frac{\epsilon^2\delta^2}{64 \cdot 2^{0.4n'+2m}}$$

so it suffices to show that

$$\frac{64s't}{\epsilon^2\delta^2} \leq 2^{0.6n'-2m}$$

But we have

$$\begin{aligned} &\frac{64s't}{\epsilon^2\delta^2} \\ &\leq 64n^{3c}\left\lceil\frac{1}{\gamma\alpha}\right\rceil 2^{\frac{2}{\gamma\alpha}+\gamma\alpha^2 n} \\ &\leq 2^{2\gamma\alpha^2 n} \end{aligned}$$

where the last inequality only holds for sufficiently large n. Similarly,

$$\begin{aligned} &2^{0.6n'-2m} \\ &\geq 2^{0.4n'} \\ &\geq 2^{0.1\gamma\alpha n} \end{aligned}$$

But it is clear that for sufficiently small α, $2\gamma\alpha^2 n \leq 0.1\gamma\alpha n$.

Similarly, the condition $\epsilon\delta \geq 128 \cdot 2^{m-0.1n'}$ holds since

$$\begin{aligned} &\frac{1}{\epsilon\delta} \\ &\leq n^{2c}2^{\frac{2}{\gamma\alpha}+\gamma\alpha^2 n} \\ &\leq 2^{2\gamma\alpha^2 n} \end{aligned}$$

for sufficiently large n and

$$2^{0.1n'-m} \geq 2^{0.04n'} \geq 2^{0.01\gamma\alpha n}$$

and for sufficiently small α, $2^{2\gamma\alpha^2 n} \leq 2^{0.01\gamma\alpha n}$.

Thus, it remains to be seen that $\text{poly}(s', 2^m) \leq s$. But as $m \leq \omega \log n$, $\text{poly}(s', 2^m) = \text{poly}(s')$ and so it is clear that there is some sufficiently small α such that $\text{poly}(s') \leq s$.

Thus, setting $\delta = \gamma\alpha^2 \geq \alpha \log s'/\ell$ gives us the result.

Lemma 9 (Lemma 5.6 from [26]). *There is a constant $\alpha > 0$ such that the following holds. Let X be a distribution of min-entropy $n_1 + n_2 - \Delta$ ranging over $\{0,1\}^{n_1+n_2}$ and let us view X as a pair (X_1, X_2) where X_1 ranges over $\{0,1\}^{n_1}$ and X_2 ranges over $\{0,1\}^{n_2}$. Let X be samplable by a postselecting circuit of size s. Let $\text{Ext}_1 : \{0,1\}^{n_1} \times \{0,1\}^t \to \{0,1\}^{m_1}$ be a $(n_1 - \Delta, \epsilon)$ two-source extractor, and let $\text{Ext}_2 : \{0,1\}^{n_2} \to \{0,1\}^{m_2}$ be a $(n_2 - \Delta - \log(1/\epsilon), \epsilon)$ deterministic extractor against postselecting circuits of size s^α. Then $\text{Ext}(X_1, X_2) = \text{Ext}_1(X_1, \text{Ext}(X_2))$ is 3ϵ-close to uniform.*

Theorem 11 ([27]). *For every $\gamma > 0$, there is a constant c_γ and an explicit construction of a $\left((1-2\gamma)n, \frac{1}{6n}\right)$-extractor $\text{Ext} : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ where $t = c_\gamma \log n$ and $m = (1-3\gamma)n$.*

Combining Theorem 10, Lemma 9, and Theorem 11 gives us Theorem 2. Following the same argument, but replacing $NP_{||}$-circuits with postselecting quantum circuits gives us Theorem 3.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il