Electronic Colloquium on Computational Complexity, Report No. 62 (2019)

Quantum Lower Bounds for Approximate Counting via Laurent Polynomials*

Scott Aaronson† Robin Kothari‡ William Kretschmer§ Justin Thaler¶

Abstract

This paper proves new limitations on the power of quantum computers to solve *approximate counting*—that is, multiplicatively estimating the size of a nonempty set $S \subseteq [N]$.

Given only a membership oracle for S, it is well known that approximate counting takes $\Theta\left(\sqrt{N/|S|}\right)$ quantum queries. But what if a quantum algorithm is also given "QSamples"—i.e., copies of the state $|S\rangle = \sum_{i \in S} |i\rangle$—or even the ability to apply reflections about $|S\rangle$? Our first main result is that, even then, the algorithm needs either $\Theta\left(\sqrt{N/|S|}\right)$ queries or else $\Theta\left(\min\left\{|S|^{1/3}, \sqrt{N/|S|}\right\}\right)$ reflections or samples. We also give matching upper bounds.

We prove the lower bound using a novel generalization of the polynomial method of Beals et al. to *Laurent polynomials*, which can have negative exponents. We lower-bound Laurent polynomial degree using two methods: a new "explosion argument" that pits the positive- and negative-degree parts of the polynomial against each other, and a new formulation of the dual polynomials method.

Our second main result rules out the possibility of a black-box Quantum Merlin-Arthur (or QMA) protocol for proving that a set is large. More precisely, we show that, even if Arthur can make T quantum queries to the set $S \subseteq [N]$, and also receives an m-qubit quantum witness from Merlin in support of S being large, we have $Tm = \Omega\left(\min\left\{|S|, \sqrt{N/|S|}\right\}\right)$. This resolves the open problem of giving an oracle separation between SBP, the complexity class that captures approximate counting, and QMA.

Note that QMA is "stronger" than the queries+QSamples model in that Merlin's witness can be anything, rather than just the specific state $|S\rangle$, but also "weaker" in that Merlin's witness cannot be trusted. Intriguingly, Laurent polynomials *also* play a crucial role in our QMA lower bound, but in a completely different manner than in the queries+QSamples lower bound. This suggests that the "Laurent polynomial method" might be broadly useful in complexity theory.

*This paper subsumes preprints arXiv:1808.02420 and arXiv:1902.02398 by the first and third authors, respectively.

†University of Texas at Austin. Email: `aaronson@cs.utexas.edu`. Supported by a Vannevar Bush Fellowship from the US Department of Defense, a Simons Investigator Award, and the Simons "It from Qubit" collaboration.

‡Quantum Architectures and Computation group (QuArC), Microsoft Research, Redmond, WA, USA. Email: `robin.kothari@microsoft.com`.

§University of Texas at Austin. Email: `kretsch@cs.utexas.edu`. Supported by a Simons Investigator award.

¶Georgetown University. Email: `justin.thaler@georgetown.edu`.

ISSN 1433-8092

Contents

1 Introduction

The *quantum query complexity of approximate counting* was one of the first topics studied in quantum algorithms. Given a nonempty finite set $S \subseteq [N]$ (here and throughout, $[N] = \{1, \ldots, N\}$), suppose we want to estimate its cardinality, $|S|$, to within some multiplicative accuracy ε. Approximate counting is a fundamental task with a rich history in classical computer science, including the work of Stockmeyer [Sto85], who showed that approximate counting is in the polynomial hierarchy, and Sinclair and Jerrum [SJ89], who showed the equivalence between approximate counting and approximate sampling that enabled the development of new algorithms using Markov chains.

In the query model (see [BdW02]), we assume we are given a membership oracle for S: one that, for any $i \in [N]$, returns whether $i \in S$. How many queries must we make, as a function of both N and $|S|$, to solve approximate counting with high probability?

For classical randomized algorithms, one can show that $\Theta(N/|S|)$ membership queries are necessary and sufficient, for approximate counting to within some constant accuracy $\varepsilon > 0$. Moreover, *any* accuracy ε is achievable at the cost of a $O(1/\varepsilon^2)$ multiplicative overhead. Intuitively, in the worst case, we might need $\Theta(N/|S|)$ queries just to find *any* elements from S, but once we do, estimating their frequency is just a standard statistics problem. Furthermore, for the $O(N/|S|)$ estimation strategy to work, we don't need to suppose (circularly) that $|S|$ is approximately known in advance, but can decide when to halt dynamically, depending on when the first element in S is found.

In the quantum setting, we can query the membership oracle on superpositions of inputs. Here Brassard et al. [BHT98a, BHMT02] gave an algorithm for approximate counting that makes only $O\big(\sqrt{N/|S|}\big)$ queries, for any constant $\varepsilon > 0$. Moreover, they showed how to achieve any accuracy ε with $O(1/\varepsilon)$ multiplicative overhead [BHMT02, Theorem 15]. To do so, one uses amplitude amplification, the basic primitive of Grover's search algorithm [Gro96]. The original algorithm of Brassard et al. also used quantum phase estimation, in effect *combining* Grover's algorithm with Shor's period-finding algorithm. However, it's a folklore fact that one can remove the phase estimation, and adapt Grover search with an unknown number of marked items, to get an approximate count of the number of marked items as well.

On the lower bound side, it follows immediately from the optimality of Grover's algorithm (i.e., the BBBV Theorem [BBBV97]) that even with a quantum computer, $\Omega\big(\sqrt{N/|S|}\big)$ queries are *needed* for approximate counting to any constant accuracy. Hence the classical and quantum complexity of approximate counting with membership queries is completely understood.

In this paper we study approximate counting in models of computation that go beyond membership queries. Before delving into the models, let us define a simple decision version of approximate counting that will be convenient to use throughout this paper.

Problem 1 (Approximate Counting)**.** *In the* $\mathsf{ApxCount}_{N,w}$ *problem, our goal is to decide whether a nonempty set* $S \subseteq [N]$ *satisfies* $|S| \geq 2w$ *(YES) or* $|S| \leq w$ *(NO), promised that one of these is the case.*

1.1 Do quantum samples and reflections help?

Quantum samples. In practice, when trying to estimate the size of a set $S \subseteq [N]$, often we can do more than make membership queries to S. At the least, often we can efficiently generate nearly uniform *samples* from S, for instance by using Markov Chain Monte Carlo techniques. To give two examples, if S is the set of perfect matchings in a bipartite graph, or the set of grid points in a high-dimensional convex body, then we can efficiently sample S using the seminal algorithms of Jerrum, Sinclair, and Vigoda [JSV04] or of Dyer, Frieze, and Kannan [DFK91], respectively.

Sometimes we can even "QSample" S—a term coined in 2003 by Aharonov and Ta-Shma [ATS03], and which simply means that we can approximately prepare the uniform superposition

$$|S\rangle := \frac{1}{\sqrt{|S|}} \sum_{i \in S} |i\rangle \tag{1}$$

via a polynomial-time quantum algorithm (where "polynomial" here means $\text{poly}(\log N)$). Because we need to uncompute any garbage, the ability to prepare $|S\rangle$ as a coherent superposition is a more stringent requirement than the ability to sample S. Indeed, as Aharonov and Ta-Shma [ATS03] pointed out, the quantum lower bound for finding collisions [Aar02, AS04] has the corollary that, in the black-box setting, there are classes of sets S that can be efficiently sampled but *not* efficiently QSampled.

On the other hand, Aharonov and Ta-Shma [ATS03], and Grover and Rudolph [GR02], observed that many interesting sets S can be QSampled as well. In particular, this holds for all sets S such that we can approximately count not only S itself, but also the restrictions of S obtained by fixing bits of its elements. Or, what's known to be equivalent [SJ89], it holds for all sets S such that we can efficiently sample not only the uniform distribution over S elements, but also the conditional distributions obtained by fixing bits. So in particular, the set of perfect matchings in a bipartite graph, and the set of grid points in a convex body, can both be efficiently QSampled. There are other sets that can be QSampled but not because of this reduction. A simple example would be a set S such that $|S| \geq \frac{N}{\text{polylog} N}$: in that case we can efficiently prepare $|S\rangle$ using postselection, but approximately counting S's restrictions might be hard.

Quantum reflections. We can further generalize the setting above to allow not only QSamples, but also reflections about $|S\rangle$: that is, applications of the unitary transformation

$$\mathcal{R}_S := \mathbb{1} - 2|S\rangle\langle S|, \tag{2}$$

which has eigenvalue -1 for $|S\rangle$ and eigenvalue $+1$ for all states orthogonal to $|S\rangle$. The ability to perform the unitary $\mathcal{R}_S$ follows in a completely black-box way from the ability to prepare the state $|S\rangle$ unitarily. More concretely, let U be the unitary that performs the map $U|0\rangle = |S\rangle$, for some canonical starting state $|0\rangle$. Since we know the circuit U, we can also implement $U^\dagger$, by reversing the order of all the gates and replacing all the gates with their adjoints. Then $\mathcal{R}_S$ is simply

$$\mathcal{R}_S = \mathbb{1} - 2|S\rangle\langle S| = U \left(\mathbb{1} - 2|0\rangle\langle 0|\right) U^\dagger. \tag{3}$$

Note that *a priori*, QSamples and reflections about $|S\rangle$ could be incomparable resources; it is not obvious how to simulate either one using the other. On the other hand, it is known how to apply a quantum channel that is ε-close to $\mathcal{R}_S$ (in the diamond norm) using $\Theta(1/\varepsilon)$ copies of $|S\rangle$ [LMR14, KLL+17].

Results. It is now natural to ask[1] whether one could solve approximate counting efficiently, using *any* combination of $\text{poly}(\log N)$ queries, samples, and reflections.

In this work, we show emphatically that the answer is no:

[1] We thank Paul Burchard (personal communication) for bringing this question to our attention.

Theorem 2. *Let Q be a quantum algorithm that makes T queries to the membership oracle for S, and uses a total of R copies of $|S\rangle$ and reflections about $|S\rangle$. If Q decides whether $|S| = w$ or $|S| = 2w$ with high probability, promised that one of those is the case, then either*

$$T = \Omega\left(\sqrt{\frac{N}{w}}\right) \qquad \text{or} \qquad R = \Omega\left(\min\left\{w^{1/3}, \sqrt{\frac{N}{w}}\right\}\right). \tag{4}$$

So if (for example) we set $w := N^{3/5}$, then any quantum algorithm must either query S, or use the state $|S\rangle$ or reflections about $|S\rangle$, at least $\Omega(N^{1/5})$ times. This means that there's at most a quadratic speedup compared to classical approximate counting.

We also prove that the lower bounds in Theorem 2 are optimal. As mentioned before, Brassard et al. [BHT98a] gave a quantum algorithm to solve the problem using $T = O(\sqrt{N/w})$ queries alone, which proves the optimality of the lower bound on the number of queries.

On the other hand, it's easy to solve the problem using $O(\sqrt{w})$ copies of $|S\rangle$ alone, by simply measuring each copy of $|S\rangle$ in the computational basis and then searching for birthday collisions. Alternately, we can solve the problem using $O\left(\frac{N}{w}\right)$ copies of $|S\rangle$ alone, by projecting onto the state $|\psi\rangle = \frac{1}{\sqrt{N}}(|1\rangle + \cdots + |N\rangle)$ or its orthogonal complement. This measurement succeeds with probability $|\langle S|\psi\rangle|^2 = \frac{|S|}{N}$, so we can approximate $|S|$ by simply counting how many measurements succeed.

In Section 3.2 we improve on these algorithms by using samples *and* reflections, and thereby establish that Theorem 2 is tight.

Theorem 3. *There is a quantum algorithm that solves $\mathsf{ApxCount}_{N,w}$ with high probability using R copies of $|S\rangle$ and reflections about $|S\rangle$, where $R = O\left(\min\left\{w^{1/3}, \sqrt{\frac{N}{w}}\right\}\right)$.*

The Laurent polynomial method. In our view, at least as interesting as Theorem 2 is the technique used to achieve it. In 1998, Beals et al. [BBC+01] famously observed that, if a quantum algorithm Q makes T queries to an input X, then Q's acceptance probability can be written as a real multilinear polynomial in the bits of X, of degree at most $2T$. And thus, crucially, if we want to *rule out* a fast quantum algorithm to compute some function $f(X)$, then it suffices to show that any real polynomial p that approximates f pointwise must have high degree. This general transformation, from questions about quantum algorithms to questions about polynomials, has been used to prove many results that were not known otherwise at the time, including the quantum lower bound for the collision problem [Aar02, AS04] and the first direct product theorems for quantum search [Aar05a, KŠdW07].

In our case, even in the simpler model with only queries and samples (and no reflections), the difficulty is that the quantum algorithm starts with many copies of the state $|S\rangle$. As a consequence of this—and specifically, of the $1/\sqrt{|S|}$ normalizing factor in $|S\rangle$—when we write the average acceptance probability of our algorithm as a function of $|S|$, we find that we get a *Laurent polynomial*: a polynomial that can contain both positive and negative integer powers of $|S|$. The degree of this polynomial (the highest power of $|S|$) encodes the sum of the number of queries, the number of copies of $|S\rangle$, and the number of uses of $\mathcal{R}_S$, while the "anti-degree" (the highest power of $|S|^{-1}$) encodes the sum of the number of copies of $|S\rangle$ and uses of $\mathcal{R}_S$. We're thus faced with the task of lower-bounding the degree and the anti-degree of a Laurent polynomial that's bounded in $[0, 1]$ at integer points and that encodes the approximate counting problem.

We then lower bound the degree of Laurent polynomials that approximate $\mathsf{ApxCount}_{N,w}$, showing that degree $\Theta\left(\min\left\{w^{1/3}, \sqrt{N/w}\right\}\right)$ is necessary. We show this using two very different arguments. The first approach, which we call the "explosion argument," is shorter but yields suboptimal lower bounds, whereas the second approach using "dual polynomials" yields the optimal lower bound.

Before describing these techniques at a high level, observe that there are *rational* functions[2] of degree $O(\log(N/w))$ that approximate $\mathsf{ApxCount}_{N,w}$. This follows, for example, from Aaronson's $\mathsf{PostBQP} = \mathsf{PP}$ theorem [Aar05b], or alternately from the classical result of Newman [N+64] that shows for any $k > 0$, there is a rational polynomial of degree $O(k)$ that pointwise approximates the sign function on domain $[-n, -1] \cup [1, n]$ to error $1 - n^{-1/k}$. Thus, our proof relies on the fact that Laurent polynomials are an extremely special kind of rational function.

Overview of the explosion argument. Our first proof uses an "explosion argument" that, as far as we know, is new in quantum query complexity. We separate out the purely positive degree[3] and purely negative degree parts of our Laurent polynomial as $q(|S|) = u(|S|) + v(1/|S|)$, where u and v are ordinary polynomials. We then show that, if u and v both have low enough degree, namely $\deg(u) = o\left(\sqrt{N/w}\right)$ and $\deg(v) = o\left(w^{1/4}\right)$, then we get "unbounded growth" in their values. That is: for approximation theory reasons, either u or v must attain large values, far outside of $[0, 1]$, at some integer values of $|S|$. But that means that, for q itself to be bounded in $[0, 1]$ (and thus represent a probability), the other polynomial must *also* attain large values. And that, in turn, will force the first polynomial to attain even larger values, and so on forever—thereby proving that these polynomials could not have existed.

Overview of the method of dual polynomials. Our second argument obtains the (optimal) lower bound stated in Theorem 2, via a novel adaptation of the so-called *method of dual polynomials*.

With this method, to lower-bound the approximate degree of a Boolean function f, one exhibits an explicit *dual polynomial* ψ for f, which is a dual solution to a certain linear program. Roughly speaking, a dual polynomial ψ is a function mapping the domain of f to $\mathbb{R}$ that is (a) uncorrelated with any polynomial of degree at most d, and (b) well-correlated with f.

Approximating a univariate function g via low-degree Laurent polynomials is also captured by a linear program, but the linear program is more complicated because Laurent polynomials can have negative-degree terms. We analyze the value of this linear program in two steps.

In Step 1, we transform the linear program so that it refers only to ordinary polynomials rather than Laurent polynomials. Although simple, this transformation is crucial, as it lets us bring techniques developed for ordinary polynomials to bear on our goal of proving Laurent polynomial degree lower bounds.

In Step 2, we explicitly construct an optimal dual witness to the transformed linear program from Step 1. We do so by first identifying two weaker dual witnesses: ψ_1, which witnesses that *ordinary* (i.e., purely positive degree) polynomials encoding approximate counting require degree at least $\Omega\left(\sqrt{N/w}\right)$, and ψ_2, which witnesses that purely negative degree polynomials encoding approximate counting require degree $\Omega(w^{1/3})$. The first witness is derived from prior work of Bun and Thaler [BT13], while the second builds on a non-constructive argument of Zhandry [Zha12].

[2] A rational function of degree d is of the form $\frac{p(x)}{q(x)}$, where p and q are both real polynomials of degree at most d.

[3] Throughout this paper we allow any "purely positive degree" Laurent polynomial and any "purely negative degree" Laurent polynomial to include a constant (degree zero) term.

Finally, we show how to "glue together" ψ_1 and ψ_2, to get a dual witness ψ showing that any general Laurent polynomial that encodes approximate counting must have either positive degree $\Omega\big(\sqrt{N/w}\big)$ or negative degree $\Omega(w^{1/3})$.

Overview of the upper bound. To recap, Theorem 2 shows that any quantum algorithm for $\mathsf{ApxCount}_{N,w}$ needs either $\Theta(\sqrt{N/w})$ queries or $\Theta\big(\min\{w^{1/3}, \sqrt{N/w}\}\big)$ samples and reflections. Since we know from the work of Brassard, Høyer, Tapp [BHT98a] that the problem can be solved with $O(\sqrt{N/w})$ queries alone, it remains only to show the matching upper bound using samples and reflections.

First we describe a simple algorithm that uses $O(\sqrt{N/w})$ samples and reflections. If we take one copy of $|S\rangle$, and perform a projective measurement onto $|\psi\rangle = \frac{1}{\sqrt{N}}(|1\rangle + \cdots + |N\rangle)$ or its orthogonal complement, the measurement will succeed with probability $|\langle S|\psi\rangle|^2 = |S|/N$. We can now use amplitude amplification [BHMT02] to distinguish the probabilities w/N and $2w/N$, and this will cost $O(\sqrt{N/w})$ repetitions. Note that amplitude amplification requires reflecting about the initial state, $|S\rangle$, so we use $O(\sqrt{N/w})$ reflections about $|S\rangle$ and one copy of $|S\rangle$.

Our second algorithm solves the problem with $O(w^{1/3})$ reflections and samples and is based on the quantum collision-finding algorithm [BHT98b]. We first use $O(w^{1/3})$ copies of $|S\rangle$ to learn $w^{1/3}$ distinct elements in S. We now know a fraction of elements in S, and this fraction is either $w^{-2/3}$ or $\frac{1}{2}w^{-2/3}$. We then use amplitude amplification (or quantum counting) to distinguish these two cases, which costs $O(w^{1/3})$ repetitions, where each repetition uses a reflection about $|S\rangle$.

1.2 Do quantum witnesses help?

We have shown that copies of $|S\rangle$ and reflections about $|S\rangle$ do not help solve approximate counting efficiently. This raises a question: what about other quantum states, besides $|S\rangle$?

In the commonly-studied *Merlin-Arthur* setting, a skeptical verifier (Arthur) receives a quantum witness state $|\psi\rangle$ from an all-powerful but untrustworthy prover (Merlin), in support of some predetermined conclusion such as S being large. Arthur then needs to verify $|\psi\rangle$, via some algorithm that satisfies the twin properties of *completeness* and *soundness*. That is, if the answer to the $\mathsf{ApxCount}_{N,w}$ instance is YES, then there must exist some $|\psi\rangle$ that causes Arthur to accept with high probability, while if the answer is NO, then every $|\psi\rangle$ must cause Arthur to reject with high probability. We call such a protocol a QMA (Quantum Merlin-Arthur) protocol.

There are two resources to consider: the length of the quantum witness and the number of queries Arthur makes to the membership oracle for S. A QMA protocol for $\mathsf{ApxCount}_{N,w}$ is efficient if both are $\mathrm{poly}(\log N)$. Does such an efficient protocol exist? Our second main result answers this question in the negative:

Theorem 4. *Consider a* QMA *protocol that solves* $\mathsf{ApxCount}_{N,w}$. *If the protocol receives a quantum witness of length m, and makes T queries to the membership oracle for S, then*

$$m \cdot T = \Omega\left(\min\left\{w, \sqrt{N/w}\right\}\right). \tag{5}$$

QMA lower bounds. In certain special cases, it is trivial to lower-bound QMA query complexity: for example, the standard BBBV Theorem [BBBV97] immediately implies the existence of an oracle relative to which $\mathsf{coNP} \not\subset \mathsf{QMA}$, and directly related to that, it is easy to show that the *complement* of $\mathsf{ApxCount}_{N,w}$ is not in QMA (i.e., that there are no short QMA witnesses proving that a set S

is *small*). Outside of those cases, though, lower-bounding QMA query complexity is challenging. Essentially all QMA lower bounds in the literature have exploited the containment $\mathsf{QMA} \subseteq \mathsf{SBQP}$, where SBQP is a complexity class that models quantum algorithms with tiny acceptance and rejection probabilities.

We say that a function f has SBQP query complexity at most k if there exists a k-query quantum algorithm that

- outputs 1 with probability $\geq \alpha$ when $f(x) = 1$, and
- outputs 1 with probability $\leq \alpha/2$ when $f(x) = 0$,

for any α that the algorithm chooses in advance. Note that when $\alpha = 2/3$, we recover standard quantum query complexity. But α could be also be exponentially small, which makes SBQP algorithms very powerful.

Nevertheless, one can establish significant limitations on SBQP algorithms, by using a variation of the polynomial method of Beals et al. [BBC+01]. If a function f can be evaluated by an SBQP algorithm with k queries, then there exists a real polynomial p of degree $2k$ such that $p(X) \in [0, 1]$ whenever $f(X) = 0$ and $p(X) \geq 2$ whenever $f(X) = 1$. The minimum degree of such a polynomial is sometimes called "one-sided approximate degree" [BT15, Tha16].

The relationship between SBQP and QMA protocols is very simple: if f has a QMA protocol that receives an m-qubit witness and makes T queries, then it also has an SBQP algorithm that makes $O(mT)$ queries. This was essentially observed by Marriott and Watrous [MW05, Remark 3.9] and used by Aaronson [Aar12] to show an oracle relative to which $\mathsf{SZK} \not\subset \mathsf{QMA}$; we also state it for completeness as Lemma 16.

Thus, an obvious route to ruling out a QMA protocol for $\mathsf{ApxCount}_{N,w}$ would be to rule out an SBQP algorithm. Unfortunately, this cannot work, since $\mathsf{ApxCount}_{N,w}$ *does* have an SBQP algorithm: namely, the algorithm that simply picks an $i \in [N]$ uniformly at random, and accepts if and only if $i \in S$. So clearly we need a different strategy.

Proof overview. To get around the issue of $\mathsf{ApxCount}_{N,w}$ being in SBQP, we use a clever strategy that was previously used by Göös et al. [GLM+16], and that was suggested to us by Thomas Watson (personal communication).

Our strategy exploits the fact that QMA is closed under intersection, but (at least relative to oracles, and as we'll show) SBQP is not.

More precisely, given a function f, let $\mathsf{AND}_2 \circ f$ be the AND of two copies of f on separate inputs. Then if f has small QMA query complexity, it's not hard to see that $\mathsf{AND}_2 \circ f$ does as well: Merlin simply sends witnesses corresponding to both inputs; then Arthur checks both of them independently. While it's not completely obvious, one can verify that a dishonest Merlin would gain nothing by entangling the two witness states. Hence if $\mathsf{ApxCount}_{N,w}$ had an efficient QMA protocol, then so would $\mathsf{AND}_2 \circ \mathsf{ApxCount}_{N,w}$, with the witness size and query complexity increasing by only a constant factor.

By contrast, even though $\mathsf{ApxCount}_{N,w}$ *does* have an efficient SBQP algorithm, we will show that $\mathsf{AND}_2 \circ \mathsf{ApxCount}_{N,w}$ does not. This is the technical core of our proof.

In more detail, suppose by way of contradiction that $\mathsf{ApxCount}_{N,w}$ had an efficient QMA protocol. Then as we said above, $\mathsf{AND}_2 \circ \mathsf{ApxCount}_{N,w}$ would have an efficient QMA protocol as well—which means that the latter function would *also* have an efficient SBQP algorithm. But by using the

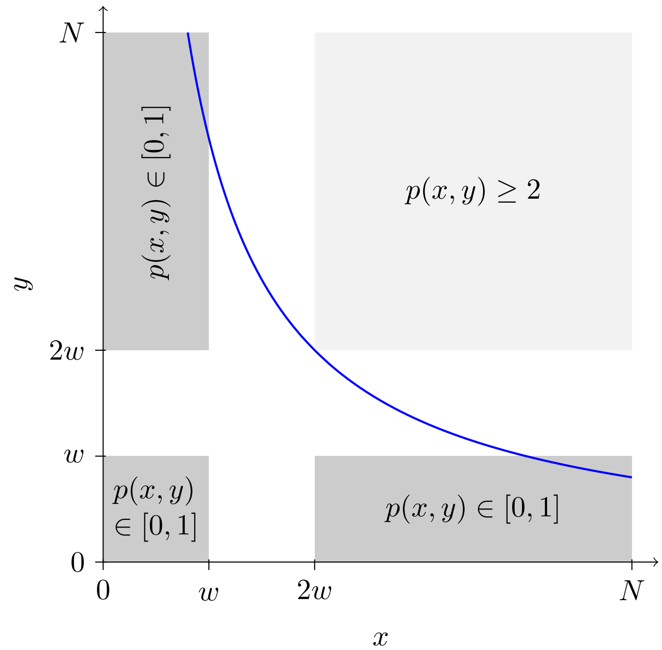

Figure 1: The behavior of the (Minsky–Papert symmetrized) bivariate polynomial $p(x,y)$ at integer points (x,y) in the proof of Theorem 5. The polynomial q obtained by erase-all-subscripts symmetrization is not depicted. We later restrict q to a hyperbola similar to the one drawn in blue.

polynomial method, we will show that $\mathsf{AND}_2 \circ \mathsf{ApxCount}_{N,w}$ does *not* have a query-efficient SBQP algorithm, thereby yielding the desired contradiction.

Theorem 5. *Consider an* SBQP *algorithm for* $\mathsf{AND}_2 \circ \mathsf{ApxCount}_{N,w}$ *that makes* T *queries to membership oracles for the two instances of* $\mathsf{ApxCount}_{N,w}$. *Then* $T = \Omega\left(\min\{w, \sqrt{N/w}\}\right)$.

Note that Theorem 5 is quantitatively optimal, as we'll exhibit matching upper bounds as well. Combined with Lemma 16 (the connection between QMA and SBQP), Theorem 5 immediately implies Theorem 4.

At a high level, the proof of Theorem 5 assumes that there's an efficient SBQP algorithm for $\mathsf{AND}_2 \circ \mathsf{ApxCount}_{N,w}$. This assumption yields a low-degree one-sided approximating polynomial for the problem in $2N$ Boolean variables, where N variables come from each $\mathsf{ApxCount}_{N,w}$ instance. We then symmetrize the polynomial (using the standard Minsky–Papert symmetrization argument [MP88]) to obtain a bivariate polynomial in two variables x and y that represent the Hamming weight of the original instances. This yields a polynomial $p(x,y)$ that for *integer pairs* x, y (also called lattice points) satisfies $p(x,y) \in [0,1]$ when either $x \in \{0,\ldots,w\}$ and $y \in \{0,\ldots,w\} \cup \{2w,\ldots,N\}$, or (symmetrically) $y \in \{0,\ldots,w\}$ and $x \in \{0,\ldots,w\} \cup \{2w,\ldots,N\}$. If both $x \in \{2w,\ldots,N\}$ and $y \in \{2w,\ldots,N\}$, then $p(x,y) \geq 2$. This polynomial p is depicted in Figure 1.

One difficulty is that we have a guarantee on the behavior of p at lattice points only, whereas the rest of our proof requires better control over the polynomial, even at non-integer points. To obtain better control over our polynomial at non-integer points, we use a newer symmetrization argument due to Servedio, Tan, and Thaler [STT12] that we call "erase all subscripts" symmetrization (see

Lemma 12).[4] This symmetrization yields a polynomial q of the same degree as p. Both types of symmetrization play an important role in our analysis, as we use p to bound q when the polynomials have degree $o(w)$, using tools from approximation theory and Chernoff bounds.

The second difficulty is that we want to lower-bound the degree of a bivariate polynomial, but almost all known lower bound techniques apply only to univariate polynomials. To reduce the number of variables (from 2 to 1) in a degree-preserving way, we pass a *hyperbola* through the xy plane (see Figure 1) and consider the polynomial q restricted to the hyperbola. Doing so gives us a new univariate *Laurent* polynomial $\ell(t)$, whose positive and negative degree is at most $\deg(q)$. This Laurent polynomial has an additional symmetry, which stems from the fact that $\mathsf{AND}_2 \circ \mathsf{ApxCount}_{N,w}$ is the AND of two identical problems (namely, $\mathsf{ApxCount}_{N,w}$). We leverage this symmetry to view $\ell(t)$, a Laurent polynomial in t, as an ordinary univariate polynomial in $t + 1/t$ of degree $\deg(q)$. Finally, we appeal to classical results in approximation theory to argue that this univariate polynomial must have degree $\Omega\left(\sqrt{N/w}\right)$.

There are two aspects of this that we find surprising: first, that Laurent polynomials appear at all, and second, that they seem to appear in a completely different way than they did in the proof of Theorem 2, despite the close connection between the two statements. For Theorem 2, Laurent polynomials were fundamentally needed just to describe the quantum algorithm's acceptance probability, whereas for Theorem 5, ordinary polynomials sufficed; Laurent polynomials appeared only when we restricted a bivariate polynomial to a hyperbola in the plane. In any case, our results suggest that the "Laurent polynomial method" might be useful for other problems as well.

Complexity-theoretic perspective. The complexity class SBP, which is sandwiched between MA (Merlin-Arthur) and AM (Arthur-Merlin), is captured by the following complete problem: given a polynomial-time computable Boolean function f and an integer w, decide whether the number of inputs that f accepts is less than w or greater than $2w$, promised that one of these is the case. The class SBQP (discussed above, following Theorem 4), first defined by Kuperberg [Kup15], is a quantum analogue of SBP that contains both SBP and QMA.

By the usual connection between oracle separations and query complexity lower bounds, Theorem 4 implies the first oracle separation between SBP and QMA—i.e., there exists an oracle A such that $\mathsf{SBP}^A \not\subset \mathsf{QMA}^A$. Prior to our work, it was known that there exist oracles A, B such that $\mathsf{SBP}^A \not\subset \mathsf{MA}^A$ [BGM06] and $\mathsf{AM}^B \not\subset \mathsf{QMA}^B$ [Ver92], but the relation between SBP and QMA remained elusive.[5] Figure 2 shows the known inclusion relations among these classes (all of which hold relative to all oracles).

Previous oracle separation techniques failed because they either used lower-bound techniques specific to MA, or they established lower bounds against PP, which contains QMA (see Figure 2) .

Now the reason that $\mathsf{SBP}^A \not\subset \mathsf{QMA}^A$ (our new result) is much harder to establish than $\mathsf{SBP}^A \not\subset \mathsf{MA}^A$ (from [BGM06]) is the following: We know lower bound methods for MA that do not also apply to SBP. In particular, MA is contained in Σ_2, which is not known to contain SBP, and Böhler et al. exploit this relationship to prove the oracle separation $\mathsf{SBP}^A \not\subset \mathsf{MA}^A$. However, the primary method for proving QMA lower bounds is exploit the fact that $\mathsf{QMA} \subseteq \mathsf{SBQP}$ and to show lower bounds on SBQP. Unfortunately, SBQP also contains SBP, and hence can solve approximate counting efficiently.

[4] It is possible to obtain a suboptimal lower bound without using erase-all-subscripts symmetrization. See arXiv:1902.02398 for details.

[5] It is interesting to note that in the non-relativized world, under plausible derandomization assumptions [MV99], we have $\mathsf{NP} = \mathsf{MA} = \mathsf{SBP} = \mathsf{AM}$. In this scenario, all these classes are equal, and all are contained in QMA.

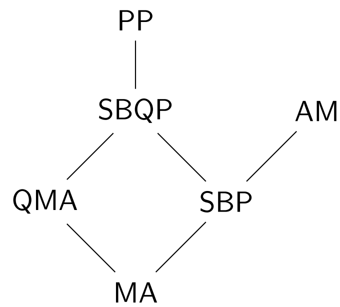

Figure 2: Relationships between complexity classes. An upward line indicates that a complexity class is contained in the one above it relative to all oracles.

As described in the proof overview above, we resolve this by using a difference in structural properties of the involved complexity classes. While QMA is easily seen to be closed under intersection, SBQP is not obviously closed under intersection, and in fact our results also show the existence of an oracle relative to which SBQP is not closed under intersection. Hence we separate SBP and QMA by taking a complete problem for SBP and considering the AND of two copies of the problem, which we then show is not in SBQP in the black-box setting.

2 Preliminaries

In this section we introduce some definitions and known facts about polynomials and complexity classes.

2.1 Approximation theory

We will use several results from approximation theory, each of which has previously been used (in some form) in other applications of the polynomial method to quantum lower bounds. We start with the basic inequality of A.A. Markov [Mar90].

Lemma 6 (Markov)**.** *Let p be a real polynomial, and suppose that*

$$\max_{x,y\in[a,b]} |p(x)-p(y)| \leq H. \tag{6}$$

Then for all $x\in[a,b]$, we have

$$\left|p'(x)\right| \leq \frac{H}{b-a}\deg(p)^2. \tag{7}$$

We'll also need a bound that was explicitly stated by Paturi [Pat92], and which amounts to the folklore fact that, among all degree-d polynomials that are bounded within a given range, the Chebyshev polynomials have the fastest growth outside that range.

Lemma 7 (Paturi)**.** *Let p be a real polynomial, and suppose that $|p(x)|\leq 1$ for all $|x|\leq 1$. Then for all $x\leq 1+\mu$, we have*

$$|p(x)| \leq \exp\left(2\deg(p)\sqrt{2\mu+\mu^2}\right). \tag{8}$$

We now state a useful corollary of Lemma 7, which says (in effect) that slightly shrinking the domain of a low-degree real polynomial can only modestly shrink its range.

Corollary 8. *Let p be a real polynomial of degree d, and suppose that*

$$\max_{x,y\in[a,b]} |p(x)-p(y)| \geq H. \tag{9}$$

Let $\varepsilon \leq \frac{1}{100d^2}$ and $a' := a+\varepsilon(b-a)$. Then

$$\max_{x,y\in[a',b]} |p(x)-p(y)| \geq \frac{H}{2}. \tag{10}$$

Proof. Suppose by contradiction that

$$|p(x)-p(y)| < \frac{H}{2} \tag{11}$$

for all $x,y \in [a',b]$. By affine shifts, we can assume without loss of generality that $|p(x)| < \frac{H}{4}$ for all $x \in [a',b]$. Then by Lemma 7, for all $x \in [a,b]$ we have

$$|p(x)| < \frac{H}{4} \cdot \exp\left(2d\sqrt{2\left(\frac{1}{1-\varepsilon}-1\right)+\left(\frac{1}{1-\varepsilon}-1\right)^2}\right) \leq \frac{H}{2}. \tag{12}$$

But this violates the hypothesis. ■

We will also need a bound that relates the range of a low-degree polynomial on a discrete set of points to its range on a continuous interval. The following lemma generalizes a result due to Ehlich and Zeller [EZ64] and Rivlin and Cheney [RC66], who were interested only in the case where the discrete points are evenly spaced.

Lemma 9. *Let p be a real polynomial of degree at most $\sqrt{k}$, and let $0 = z_1 < \cdots < z_M = k$ be a list of points such that $z_{i+1} - z_i \leq 1$ for all i (the simplest example being the integers $0,\ldots,k$). Suppose that*

$$\max_{x,y\in[0,k]} |p(x)-p(y)| \geq H. \tag{13}$$

Then

$$\max_{i,j} |p(z_i)-p(z_j)| \geq \frac{H}{2}. \tag{14}$$

Proof. Suppose by contradiction that

$$|p(z_i)-p(z_j)| < \frac{H}{2} \tag{15}$$

for all i,j. By affine shifts, we can assume without loss of generality that $|p(z_i)| < \frac{H}{4}$ for all i. Let

$$c := \max_{x\in[0,k]} \frac{|p(x)|}{H/4}. \tag{16}$$

If $c \leq 1$, then the hypothesis clearly fails, so assume $c > 1$. Suppose that the maximum, $|p(x)| = \frac{cH}{4}$, is achieved between z_i and z_{i+1}. Then by basic calculus, there exists an $x^* \in [z_i, z_{i+1}]$ such that

$$\left|p'(x^*)\right| > \frac{2(c-1)}{z_{i+1} - z_i} \cdot \frac{H}{4} \geq \frac{(c-1)H}{2}. \tag{17}$$

So by Lemma 6,

$$\frac{(c-1)H}{2} < \frac{cH/4}{k} \deg(p)^2. \tag{18}$$

Solving for c, we find

$$c < \frac{2k}{2k - \deg(p)^2} \leq 2. \tag{19}$$

But if $c < 2$, then $\max_{x \in [0,k]} |p(x)| < \frac{H}{2}$, which violates the hypothesis. ■

We also use a related inequality due to Coppersmith and Rivlin [CR92] that bounds a polynomial on a continuous interval in terms of a bound on a discrete set of points, but now with the weaker assumption that the degree is at most k, rather than $\sqrt{k}$. This gives a substantially weaker bound.

Lemma 10 (Coppersmith and Rivlin)**.** *Let p be a real polynomial of degree at most k, and suppose that $|p(x)| \leq 1$ for all integers $x \in \{0, 1, \ldots, k\}$. Then there exist universal constants a, b such that for all $x \in [0, k]$, we have*

$$|p(x)| \leq a \cdot \exp\left(b \deg(p)^2 / k\right). \tag{20}$$

2.2 Symmetric polynomials

Univariate symmetrizations. Our starting point is the well-known *symmetrization lemma* of Minsky and Papert [MP88] (see also Beals et al. [BBC+01] for its application to quantum query complexity), by which we can often reduce questions about multivariate polynomials to questions about univariate ones.

Lemma 11 (Minsky–Papert symmetrization)**.** *Let $p : \{0,1\}^N \to \mathbb{R}$ be a real multilinear polynomial of degree d, and let $q : \{0, 1, \ldots, N\} \to \mathbb{R}$ be defined as*

$$q(k) := \mathbb{E}_{|X|=k}[p(X)]. \tag{21}$$

Then q can be written as a real polynomial in k of degree at most d.

We now introduce a different, lesser known notion of symmetrization, due to Servedio, Tan, and Thaler [STT12], which we call the *erase-all-subscripts* symmetrization for reasons to be explained shortly.

Lemma 12 (Erase-all-subscripts symmetrization)**.** *Let $p : \{0,1\}^N \to \mathbb{R}$ be a real multilinear polynomial of degree d, and for any real number $k \in [0,1]$, let M_k denote the distribution over $\{0,1\}^N$, wherein each coordinate is selected independently to be 1 with probability k. Let $q : [0,1] \to \mathbb{R}$ be defined as*

$$q(k) := \mathbb{E}_{M_k}[p(X)]. \tag{22}$$

Then q can be written as a real polynomial in k of degree at most d.

Proof. (Appears in [STT12, Proof of Theorem 3]). Given the multivariate polynomial expansion of p, we can obtain q easily just by "erasing all the subscripts in each variable". For example, if $p(x_1, x_2, x_3) = 2x_1x_2 + x_2x_3 + x_2$, we replace every x_i with k to obtain $q(k) = 2k\cdot k + k\cdot k + k = 3k^2 + k$. This follows from linearity of expectation along with the fact that M_k is defined to be the product distribution wherein each coordinate has expected value k. ■

We highlight the following key difference between Minsky–Papert symmetrization and the erase-all-subscripts symmetrization. Let $p : \{0,1\}^N \to [0,1]$ be a real multivariate polynomial whose evaluations at Boolean inputs are in $[0,1]$, i.e., for all $x \in \{0,1\}^n$, we have $p(x) \in [0,1]$. If q is the the erase-all-subscripts symmetrization of p, then q takes values in $[0,1]$ at all *real-valued* inputs in $[0,1]$: $q(k) \in [0,1]$ for all $k \in [0,1]$. If q is the Minsky–Papert symmetrization of p, then it is only guaranteed to take values in $[0,1]$ at *integer-valued* inputs in $[0,N]$, i.e., $q(k) \in [0,1]$ is only guaranteed to hold at $k \in \{0,1,\ldots,N\}$. This is the main reason we use erase-all-subscripts symmetrization in this work.

Bivariate symmetrizations. In this paper, it will be convenient to consider bivariate versions of both Minsky–Papert and erase-all-subscripts symmetrization, and their applications to oracle separations. To this end, define $X \in \{0,1\}^N$, the "characteristic string" of the set $S \subseteq [N]$, by $x_i = 1$ if $i \in S$ and $x_i = 0$ otherwise. Let $\mathcal{O}_S$ denote the unitary that performs a membership query to S, defined as

$$\mathcal{O}_S \left|i\right\rangle \left|b\right\rangle = (1 - 2bx_i) \left|i\right\rangle \left|b\right\rangle \tag{23}$$

for any index $i \in [N]$ and bit $b \in \{0,1\}$.

Because we study oracle intersection problems, it is often convenient to think of an algorithm as having access to *two* oracles, wherein the first bit in the oracle register selects the choice of oracle. As a consequence, we need a slight generalization of a now well-established fact in quantum complexity: that the acceptance probability of a quantum algorithm with an oracle can be expressed as a polynomial in the bits of the oracle string.

Lemma 13 (Symmetrization with two oracles). *Let $Q^{\mathcal{O}_{S_0}, \mathcal{O}_{S_1}}$ be a quantum algorithm that makes T queries to a pair of membership oracles for sets $S_0, S_1 \subseteq [N]$. Let D_μ denote the distribution over subsets of $[N]$ wherein each element is selected independently with probability $\frac{\mu}{N}$. Then there exist bivariate real polynomials $q(s,t)$ and $p(x,y)$ of degree at most $2T$ satisfying:*

$$\text{for all real numbers } s,t \in [0,N], \quad q(s,t) = \mathbb{E}_{\substack{S_0 \sim D_s,\\ S_1 \sim D_t}} \left[\Pr[Q^{\mathcal{O}_{S_0}, \mathcal{O}_{S_1}} \text{ accepts}]\right], \text{ and}$$

$$\text{for all integers } x,y \in \{0,1,\ldots,N\}, \quad p(x,y) = \mathbb{E}_{\substack{|S_0|=x,\\ |S_1|=y}} \left[\Pr[Q^{\mathcal{O}_{S_0}, \mathcal{O}_{S_1}} \text{ accepts}]\right].$$

Proof. Take $X = X_0|X_1$ to be the concatenation of the characteristic strings of the two oracles, and let $S \subseteq [2N]$ be such that X is the characteristic string of S. Then, Lemma 4.2 of Beals et al. [BBC+01] tells us that there is a real multilinear polynomial $r(X)$ of degree at most $2T$ in the bits of X such that $r(X) = \Pr[Q^{\mathcal{O}_S} \text{ accepts}]$.

Observe that r has a meaningful probabilistic interpretation over arbitrary inputs in $[0,1]$. A vector $X \in [0,1]^{2N}$ of probabilities corresponds to a distribution over $\{0,1\}^{2N}$ wherein each bit is chosen from a Bernoulli distribution with the corresponding probability. Because r is multilinear, r

in fact computes the expectation of the acceptance probability over this distribution. In particular, the polynomial

$$q(s,t) = r\Bigg(\underbrace{\frac{s}{N},\ldots,\frac{s}{N}}_{N \text{ times}},\underbrace{\frac{t}{N},\ldots,\frac{t}{N}}_{N \text{ times}}\Bigg) = \mathbb{E}_{\substack{S_0 \sim D_s,\\ S_1 \sim D_t}}\left[\Pr[Q^{\mathcal{O}_{S_0},\mathcal{O}_{S_1}} \text{ accepts}]\right] \tag{24}$$

corresponds to selecting $S_0 \sim D_s$ and $S_1 \sim D_t$. The total degree of q is obviously at most the degree of r, by the same reasoning as in the proof of Lemma 12.

To construct p, we apply the symmetrization lemma of Minsky and Papert [MP88] to symmetrize r, first with respect to X_0, then with respect to X_1:

$$p_0(x, X_1) = \mathbb{E}_{|S_0|=x}\, r(X_0, X_1) = \mathbb{E}_{|S_0|=x}\left[\Pr[Q^{\mathcal{O}_{S_0},\mathcal{O}_{S_1}} \text{ accepts}]\right] \tag{25}$$

$$p(x,y) = \mathbb{E}_{|S_1|=y}\, p_0(x, X_1) = \mathbb{E}_{\substack{|S_0|=x,\\ |S_1|=y}}\left[\Pr[Q^{\mathcal{O}_{S_0},\mathcal{O}_{S_1}} \text{ accepts}]\right] \tag{26}$$

The degree of p is at most the degree of r, due to Lemma 11. ∎

We remark that, as a consequence of their definitions in Lemma 13, p and q satisfy:

$$q(s,t) = \mathbb{E}\left[p(X,Y)\right], \tag{27}$$

where X and Y are drawn from N-trial binomial distributions with means s and t, respectively.

Symmetric Laurent polynomials. Finally, we state a useful fact about Laurent polynomials:

Lemma 14 (Symmetric Laurent polynomials)**.** *Let $\ell(x)$ be a real Laurent polynomial of positive and negative degree d that satisfies $\ell(x) = \ell(1/x)$. Then there exists a (ordinary) real polynomial q of degree d such that $\ell(x) = q(x + 1/x)$.*

Proof. $\ell(x) = \ell(1/x)$ implies that the coefficients of the x^i and x^{-i} terms are equal for all i, as otherwise $\ell(x) - \ell(1/x)$ would not equal the zero polynomial. Thus, we may write $\ell(x) = \sum_{i=0}^{d} a_i \cdot (x^i + x^{-i})$ for some coefficients a_i. So, it suffices to show that $x^i + x^{-i}$ can be expressed as a polynomial in $x + 1/x$ for all $0 \le i \le d$.

We prove by induction on i. The case $i = 0$ corresponds to constant polynomials. For $i > 0$, by the binomial theorem, observe that $(x + 1/x)^i = x^i + x^{-i} + r(x)$ where r is a degree $i - 1$ real Laurent polynomial satisfying $r(x) = r(1/x)$. By the induction assumption, r can be expressed as a polynomial in $x + 1/x$, so we have $x^i + x^{-i} = (x + 1/x)^i - r(x)$ is expressed as a polynomial in $x + 1/x$. ∎

2.3 Complexity classes

Though SBP and SBQP can be defined in terms of counting complexity functions, for our purposes it is easier to work with the following equivalent definitions (see Böhler et al. [BGM06]):

Definition 15. *The complexity class SBP consists of the languages L for which there exists a probabilistic polynomial time algorithm M and a polynomial σ with the following properties:*

1. *If $x \in L$, then $\Pr[M(x)$ accepts$] \ge 2^{-\sigma(|x|)}$.*

2. If $x \notin L$, then $\Pr[M(x) \text{ accepts}] \leq 2^{-\sigma(|x|)}/2$.

The complexity class SBQP *is defined analogously, wherein the classical algorithm is replaced with a quantum algorithm.*

A classical (respectively, quantum) algorithm that satisfies the above promise for a particular language will be referred to as an SBP (respectively, SBQP) algorithm throughout. Using this definition, a query complexity relation between QMA protocols and SBQP algorithms follows from the procedure of Marriott and Watrous [MW05], which shows that one can exponentially improve the soundness and completeness errors of a QMA protocol without increasing the witness size. See Aaronson [Aar12, Lemma 5] for a proof of the following lemma:

Lemma 16 (Guessing lemma)**.** *Suppose there is a* QMA *protocol for some problem that makes T queries and receives an m-qubit witness. Then there is an* SBQP *algorithm for the same problem that makes $O(mT)$ queries.*

3 Approximate counting with quantum samples and reflections

3.1 The Laurent polynomial method

By using Minsky–Papert symmetrization (Lemma 11), we now prove the key fact that relates quantum algorithms, of the type we're considering, to real Laurent polynomials in one variable. The following lemma generalizes the connection between quantum algorithms and real polynomials established by Beals et al. [BBC+01].

Lemma 17. *Let Q be a quantum algorithm that makes T queries to $\mathcal{O}_S$, uses R_1 copies of $|S\rangle$, and makes R_2 uses of the unitary $\mathcal{R}_S$. Let $R := R_1 + 2R_2$. For $k \in \{1, \ldots, N\}$, let*

$$q(k) := \mathbb{E}_{|S|=k}\left[\Pr\left[Q^{\mathcal{O}_S, \mathcal{R}_S}\left(|S\rangle^{\otimes R_1}\right) \text{ accepts}\right]\right]. \tag{28}$$

Then q can be written a univariate Laurent polynomial, with maximum exponent at most $2T + R$ and minimum exponent at least $-R$.

Proof. Let $|\psi_{\text{initial}}\rangle$ denote the initial state of the algorithm, which we can write as

$$\begin{aligned}|\psi_{\text{initial}}\rangle = |S\rangle^{\otimes R_1} &= \left(\frac{1}{\sqrt{|S|}}\sum_{i\in S}|i\rangle\right)^{R_1} = \left(\frac{1}{\sqrt{|S|}}\sum_{i\in[N]}x_i|i\rangle\right)^{R_1} \\ &= \frac{1}{|S|^{R_1/2}}\sum_{i_1,\ldots,i_{R_1}\in[N]} x_{i_1}\cdots x_{i_{R_1}}\,|i_1,\ldots,i_{R_1}\rangle\,.\end{aligned}$$

Thus, each amplitude is a complex multilinear polynomial in $X = (x_1, \ldots, x_N)$ of degree R_1, divided by $|S|^{R_1/2}$.

Throughout the algorithm, each amplitude will remain a complex multilinear polynomial in X divided by some power of $|S|$. Since $x_i^2 = x_i$ for all i, we can always maintain multilinearity without loss of generality.

Like Beals et al. [BBC+01], we now consider how the polynomial degree of each amplitude and the power of $|S|$ in the denominator change as the algorithm progresses. We have to handle 3 different kinds of unitaries that the quantum circuit may use: the membership query oracle $\mathcal{O}_S$, unitaries independent of the input, and the reflection unitary $\mathcal{R}_S$.

The first two cases are handled as in Beals et al. Since $\mathcal{O}_S$ is a unitary whose entries are degree-1 polynomials in X, each use of this unitary increases a particular amplitude's degree as a polynomial by 1 and does not change the power of $|S|$ in the denominator. Second, input-independent unitary transformations only take linear combinations of existing polynomials and hence do not increase the degree of the amplitudes or the power of $|S|$ in the denominator. Finally, we consider the reflection unitary $\mathcal{R}_S = \mathbb{1} - 2|S\rangle\langle S|$. The $(i,j)^{\text{th}}$ entry of this operator is $1 - \frac{2x_i x_j}{|S|} = \frac{|S| - 2x_i x_j}{|S|}$. Since $|S| = \sum_i x_i$, this is a degree-2 polynomial divided by $|S|$. Hence applying this unitary will increase the degree of the amplitudes by 2 and increase the power of $|S|$ in the denominator by 1.

In conclusion, we start with each amplitude being a polynomial of degree R_1 divided by $|S|^{R_1/2}$. T queries to the membership oracle will increase the degree of each amplitude by at most T and leave the power of $|S|$ in the denominator unchanged. R_2 uses of the reflection unitary will increase the degree by at most $2R_2$ and the power of $|S|$ in the denominator by R_2. It follows that Q's final state has the form

$$|\psi_{\text{final}}\rangle = \sum_z \alpha_z(X)\,|z\rangle\,, \tag{29}$$

where each $\alpha_z(X)$ is a complex multilinear polynomial in X of degree at most $R_1 + 2R_2 + T = R + T$, divided by $|S|^{R_1/2 + R_2} = |S|^{R/2}$. Since X itself is real-valued, it follows that the real and imaginary parts of $\alpha_z(X)$, considered individually, are real multilinear polynomials in X of degree at most $R + T$ divided by $|S|^{R/2}$.

Hence, if we let

$$p(X) := \Pr\left[Q^{\mathcal{O}_S, \mathcal{R}_S}\left(|S\rangle^{\otimes R_1}\right) \text{ accepts}\right], \tag{30}$$

then

$$p(X) = \sum_{\text{accepting } z} |\alpha_z(X)|^2 = \sum_{\text{accepting } z} \left(\operatorname{Re}^2 \alpha_z(X) + \operatorname{Im}^2 \alpha_z(X)\right) \tag{31}$$

is a real multilinear polynomial in X of degree at most $2(R+T)$, divided through (in every monomial) by $|S|^R = |X|^R$.

Now consider

$$q(k) := \mathbb{E}_{|X|=k}\left[p(X)\right]. \tag{32}$$

By Lemma 11, this is a real univariate polynomial in $|X|$ of degree at most $2(R+T)$, divided through (in every monomial) by $|S|^R = |X|^R$. Or said another way, it's a real Laurent polynomial in $|X|$, with maximum exponent at most $R + 2T$ and minimum exponent at least $-R$. ∎

3.2 Upper bounds

Before proving our lower bounds on the degree of Laurent polynomials approximating $\mathsf{ApxCount}_{N,w}$, we establish some simpler *upper bounds*. We show upper bounds on Laurent polynomial degree and in the queries, samples, and reflections model.

Laurent polynomial degree of approximate counting. We now describe a *purely negative* degree Laurent polynomial of degree $O(w^{1/3})$ for approximate counting. This upper bound will serve as an important source of intuition when we prove the (matching) lower bound of Theorem 2 (see Section 3.4.3). We are thankful to user "fedja" on MathOverflow for describing this construction.[6]

Lemma 18 (fedja). *For all w, there is a real polynomial p such that $|p(1/k)| \leq 1$ for all $k \in [2w]$, and $p\left(\frac{1}{w}\right) \leq \frac{1}{3}$ while $p\left(\frac{1}{2w}\right) \geq \frac{2}{3}$, and $\deg(p) = O\left(w^{1/3}\right)$.*

Proof. Assuming for simplicity that w is a perfect cube, consider

$$u(x) := (1-x)(1-2x)\cdots\left(1-w^{1/3}x\right). \tag{33}$$

Notice that $\deg(u) = w^{1/3}$ and $u\left(\frac{1}{k}\right) = 0$ for all $k \in \left[w^{1/3}\right]$. Furthermore, we have $|u(x)| \leq 1$ for all $x \in \left[0, \frac{1}{w^{1/3}}\right]$, and also $u(x) \in \left[1 - O\left(\frac{1}{w^{1/3}}\right), 1\right]$ for all $x \in \left[0, \frac{1}{w}\right]$. Now, let v be the Chebyshev polynomial of degree $w^{1/3}$, affinely adjusted so that $|v(x)| \leq 1$ for all $x \in \left[0, \frac{1}{w^{1/3}}\right]$ rather than all $|x| \leq 1$, and with a large jump between $\frac{1}{2w}$ and $\frac{1}{w}$. Then the product, $p(x) := u(x)v(x)$, has degree $2w^{1/3}$ and satisfies all the requirements. ■

Interestingly, if we restrict our attention to purely negative degree Laurent polynomials, then a matching lower bound is not too hard to show. In the same MathOverflow post, user fedja also proves the following, which can also be shown using earlier work of Zhandry [Zha12, Proof of Theorem 7.3]):

Lemma 19. *Let p be a real polynomial, and suppose that $|p(1/k)| \leq 1$ for all $k \in [2w]$, and that $p\left(\frac{1}{w}\right) \leq \frac{1}{3}$ while $p\left(\frac{1}{2w}\right) \geq \frac{2}{3}$. Then $\deg(p) = \Omega\left(w^{1/3}\right)$.*

Section 3.3 and Section 3.4 below take the considerable step of extending Lemma 19 from the setting of purely negative degree Laurent polynomials to general Laurent polynomials.

Upper bounds in the queries, samples, and reflections model. Although we showed that there is a purely negative degree Laurent polynomial of degree $O(w^{1/3})$ for $\mathsf{ApxCount}_{N,w}$, this does not imply the existence of a quantum algorithm in the queries, samples, and reflections model with similar complexity.

We now show that our lower bounds in the queries, samples, and reflections model (in Theorem 2) are tight (up to constants). This is Theorem 3 in the introduction, restated here for convenience:

Theorem 3. *There is a quantum algorithm that solves $\mathsf{ApxCount}_{N,w}$ with high probability using R copies of $|S\rangle$ and reflections about $|S\rangle$, where $R = O\left(\min\left\{w^{1/3}, \sqrt{\frac{N}{w}}\right\}\right)$.*

Proof. We describe two quantum algorithms for this problem with the two stated complexities.

The first algorithm uses $O(w^{1/3})$ samples and reflections. This algorithm is reminiscent of the original collision finding algorithm of Brassard, Høyer, and Tapp [BHT98b]. We first use $O(w^{1/3})$ copies of $|S\rangle$ to learn a set $M \subset S$ of size $w^{1/3}$ by simply measuring copies of $|S\rangle$ in the computational basis. Now we know that the ratio $|S|/|M|$ is either $w^{2/3}$ or $2w^{2/3}$. Now consider running Grover's algorithm on the set S where the elements in M are considered the "marked" elements. Grover's algorithm alternates reflections about the uniform superposition over the set being searched, S, with an operator that reflects about the marked elements in M. The first reflection is simply $\mathcal{R}_S$, which

[6] See https://mathoverflow.net/questions/302113/real-polynomial-bounded-at-inverse-integer-points

we have access to. The second unitary can be constructed since we have an explicit description of the set M. Now Grover's algorithm can be used to distinguish whether the fraction of marked elements is $1/w^{2/3}$ or half of that, and the cost will be $O(w^{1/3})$.

The second algorithm uses $O\left(\sqrt{\frac{N}{w}}\right)$ samples and reflections. We start with the state $|S\rangle$ and perform the two-outcome measurement that projects the state onto the uniform superposition over all N elements, $\sum_{i\in[N]}|i\rangle$. The probability of obtaining this measurement outcome is exactly the square of the inner product between this state and $|S\rangle$, which is $|S|/N$. This probability is either w/N or $2w/N$. We can distinguish these two probabilities using amplitude amplification, and the number of steps of amplitude amplification required is $O(\sqrt{N/w})$. Note that amplitude amplification will require reflections about the starting state $|S\rangle$, which we have available. ■

Note that both the algorithms presented above generalize to the situation where we want to distinguish $|S| = w$ from $|S| = (1+\varepsilon)w$. For the first algorithm, we now pick a subset M of size $w^{1/3}/\varepsilon^{2/3}$. Now we want to $(1+\varepsilon)$-approximate the fraction of marked elements, which is either $1/(w\varepsilon)^{2/3}$ or $(1+\varepsilon)^{-1}$ times that. This can be done with approximate counting [BHMT02, Theorem 15], and the cost will be $O\left(\frac{1}{\varepsilon}(w\varepsilon)^{1/3}\right) = O\left(\frac{w^{1/3}}{\varepsilon^{2/3}}\right)$.

The second algorithm is simpler to generalize, as we only need to distinguish the two probabilities more accurately. This will incur a $O(1/\varepsilon)$ overhead, since we want to approximate the probability to $(1+\varepsilon)$ relative error. This will yield an algorithm that uses $O\left(\frac{1}{\varepsilon}\sqrt{\frac{N}{w}}\right)$ copies and reflections about $|S\rangle$. Another way to derive this bound is to consider running the usual approximate counting algorithm that uses queries to S. The membership oracle in that algorithm is only used to reflect about the subspace of all elements in S. However, we can simply replace that reflection with a reflection about $|S\rangle$. This is because the analysis of the algorithm reveals that the state of the algorithm always remains in a two-dimensional subspace, and within that subspace the reflection about elements in S and reflection about $|S\rangle$ are identical operators.

3.3 Lower bound using the explosion argument

We now show a weaker version of Theorem 2 using the explosion argument described in the introduction. The difference between the following theorem and Theorem 2 is the exponent of w in the lower bound.

Theorem 20. *Let Q be a quantum algorithm that makes T queries to the membership oracle for S, and uses a total of R copies of $|S\rangle$ and reflections about $|S\rangle$. If Q decides whether $|S| = w$ or $|S| = 2w$ with success probability at least $2/3$, promised that one of those is the case, then either*

$$T = \Omega\left(\sqrt{\frac{N}{w}}\right) \qquad \text{or} \qquad R = \Omega\left(\min\left\{w^{1/4}, \sqrt{\frac{N}{w}}\right\}\right). \tag{34}$$

Proof. Since we neglect multiplicative constants in our lower bounds, let us allow the algorithm to use up to R copies of $|S\rangle$ and R uses of $\mathcal{R}_S$. Let

$$q\left(k\right) := \mathbb{E}_{|S|=k}\left[\Pr\left[Q^{\mathcal{O}_S,\mathcal{R}_S}\left(|S\rangle^{\otimes R}\right) \text{ accepts}\right]\right]. \tag{35}$$

Then by Lemma 17, we can write q as a Laurent polynomial, like so:

$$q\left(k\right) = u\left(k\right) + v\left(1/k\right), \tag{36}$$

where u is a real polynomial in k with $\deg(u) = O(T+R)$, and v is a real polynomial in $1/k$ with $\deg(v) = O(R)$. So to prove the theorem, it suffices to show that either $\deg(u) = \Omega\left(\sqrt{\frac{N}{w}}\right)$, or else $\deg(v) = \Omega\left(w^{1/4}\right)$. To do so, we'll assume that $\deg(u) = o\left(\sqrt{\frac{N}{w}}\right)$ and $\deg(v) = o\left(w^{1/4}\right)$, and derive a contradiction.

Our high-level strategy is as follows: we'll observe that, if approximate counting is successfully being solved, then either u or v must attain a large first derivative somewhere in its domain. By the approximation theory lemmas that we proved in Section 2.1, this will force that polynomial to have a large range—even on a subset of integer (or inverse-integer) points. But the sum, $u(k) + v(1/k)$, is bounded in $[0,1]$ for all $k \in [N]$. So if one polynomial has a large range, then the other does too. But this forces the *other* polynomial to have a large derivative somewhere in its domain, and therefore (by approximation theory) to have an even larger range, forcing the first polynomial to have an even larger range to compensate, and so on. As long as $\deg(u)$ and $\deg(v)$ are both small enough, this endless switching will force both u and v to attain *unboundedly* large values—with the fact that one polynomial is in k, and the other is in $1/k$, crucial to achieving the desired "explosion." Since u and v are polynomials on compact sets, such unbounded growth is an obvious absurdity, and this will give us the desired contradiction.

In more detail, we will study the following quantities.

$$
\begin{array}{ll}
G_u := \max_{x,y\in\left[\sqrt{w},2w\right]} \left|u(x)-u(y)\right| & G_v := \max_{x,y\in\left[\frac{1}{N},\frac{1}{w}\right]} \left|v(x)-v(y)\right| \\
\Delta_u := \max_{x\in\left[\sqrt{w},2w\right]} \left|u'(x)\right| & \Delta_v := \max_{x\in\left[\frac{1}{N},\frac{1}{w}\right]} \left|v'(x)\right| \\
H_u := \max_{x,y\in\left[\sqrt{w},N\right]} \left|u(x)-u(y)\right| & H_v := \max_{x,y\in\left[\frac{1}{N},\frac{1}{\sqrt{w}}\right]} \left|v(x)-v(y)\right| \\
I_u := \max_{x,y\in[w,N]} \left|u(x)-u(y)\right| & I_v := \max_{x,y\in\left[\frac{1}{2w},\frac{1}{\sqrt{w}}\right]} \left|v(x)-v(y)\right| \\
L_u := \max_{x,y\in\{w,\ldots,N\}} \left|u(x)-u(y)\right| & L_v := \max_{x,y\in\left\{\sqrt{w},\ldots,2w\right\}} \left|v\left(\frac{1}{x}\right)-v\left(\frac{1}{y}\right)\right|
\end{array}
\tag{37}
$$

We have $0 \leq q(k) \leq 1$ for all $k \in [N]$, since in those cases $q(k)$ represents a probability. Since Q solves approximate counting, we also have $q(w) \leq \frac{1}{3}$ and $q(2w) \geq \frac{2}{3}$. This means in particular that either

(i) $u(2w) - u(w) \geq \frac{1}{6}$, and hence $G_u \geq \frac{1}{6}$, or else

(ii) $v\left(\frac{1}{2w}\right) - v\left(\frac{1}{w}\right) \geq \frac{1}{6}$, and hence $G_v \geq \frac{1}{6}$.

We will show that either case leads to a contradiction.

We have the following inequalities regarding u:

$$
\begin{array}{ll}
G_u \geq L_v - 1 & \text{by the boundedness of } q \\
\Delta_u \geq \frac{G_u}{2w} & \text{by basic calculus} \\
H_u \geq \frac{\Delta_u\left(N-\sqrt{w}\right)}{\deg(u)^2} & \text{by Lemma 6} \\
I_u \geq \frac{H_u}{2} & \text{by Corollary 8} \\
L_u \geq \frac{I_u}{2} & \text{by Lemma 9}
\end{array}
\tag{38}
$$

Here the fourth inequality uses the fact that, setting $\varepsilon := \frac{\sqrt{w}}{N}$, we have $\deg(u) = o\left(\frac{1}{\sqrt{\varepsilon}}\right)$ (thereby satisfying the hypothesis of Corollary 8), while the fifth inequality uses the fact that $\deg(u) = o\left(\sqrt{N}\right)$.

Meanwhile, we have the following inequalities regarding v:

$$\begin{array}{ll} G_v \geq L_u - 1 & \text{by the boundedness of } q \\ \Delta_v \geq G_v w & \text{by basic calculus} \\ H_v \geq \frac{\Delta_v\left(\frac{1}{\sqrt{w}} - \frac{1}{N}\right)}{\deg(v)^2} & \text{by Lemma 6} \\ I_v \geq \frac{H_v}{2} & \text{by Corollary 8} \\ L_v \geq \frac{I_v}{2} & \text{by Lemma 9} \end{array} \tag{39}$$

Here the fourth inequality uses the fact that, setting $\varepsilon := \frac{1/2w}{1/\sqrt{w}} = \frac{1}{2\sqrt{w}}$, we have $\deg(v) = o\left(\frac{1}{\sqrt{\varepsilon}}\right)$ (thereby satisfying the hypothesis of Corollary 8). The fifth inequality uses the fact that, if we set $V(x) := v(x/w)$, then the situation satisfies the hypothesis of Lemma 9: we are interested in the range of V on the interval $\left[\frac{1}{2}, \sqrt{w}\right]$, compared to its range on discrete points $\frac{w}{\sqrt{w}}, \frac{w}{\sqrt{w}+1}, \ldots, \frac{w}{2w}$ that are spaced at most 1 apart from each other; and we also have $\deg(V) = \deg(v) = o\left(w^{1/4}\right)$.

All that remains is to show that, if we insert either $G_u \geq \frac{1}{6}$ or $G_v \geq \frac{1}{6}$ into the coupled system of inequalities above, then we get unbounded growth and the inequalities have no solution. Let us collapse the two sets of inequalities to

$$\begin{aligned} L_u &\geq \frac{1}{4} \frac{N - \sqrt{w}}{\deg(u)^2} \frac{G_u}{2w} = \Omega\left(\frac{N}{w\deg(u)^2} G_u\right), \\ L_v &\geq \frac{1}{4} \frac{\frac{1}{\sqrt{w}} - \frac{1}{N}}{\deg(v)^2} G_v w = \Omega\left(\frac{\sqrt{w}}{\deg(v)^2} G_v\right). \end{aligned}$$

Hence

$$\begin{aligned} G_u &\geq L_v - 1 = \Omega\left(\frac{\sqrt{w}}{\deg(v)^2} G_v\right) - 1, \\ G_v &\geq L_u - 1 = \Omega\left(\frac{N}{w\deg(u)^2} G_u\right) - 1. \end{aligned}$$

By the assumption that $\deg(v) = o\left(w^{1/4}\right)$ and $\deg(u) = o\left(\sqrt{\frac{N}{w}}\right)$, we have $\frac{\sqrt{w}}{\deg(v)^2} \gg 1$ and $\frac{N}{w\deg(u)^2} \gg 1$. Plugging in $G_u \geq \frac{1}{6}$ or $G_v \geq \frac{1}{6}$, this is enough to give us unbounded growth. ■

3.4 Lower bound using dual polynomials

In this section we use the method of dual polynomials to establish our main result, Theorem 2, restated for convenience:

Theorem 2. *Let Q be a quantum algorithm that makes T queries to the membership oracle for S, and uses a total of R copies of $|S\rangle$ and reflections about $|S\rangle$. If Q decides whether $|S| = w$ or $|S| = 2w$ with high probability, promised that one of those is the case, then either*

$$T = \Omega\left(\sqrt{\frac{N}{w}}\right) \qquad \text{or} \qquad R = \Omega\left(\min\left\{w^{1/3}, \sqrt{\frac{N}{w}}\right\}\right). \tag{4}$$

Let $p(r)$ be a univariate Laurent polynomial of negative degree D_1 and positive degree D_2. That is, let $p(r)$ be of the form

$$p(r) = a_0/r^{D_1} + a_1/r^{D_1-1} + \cdots + a_{D_1-1}/r + a_{D_1} + a_{D_1+1} \cdot r + \cdots + a_{D_2+D_1} \cdot r^{D_2}. \tag{40}$$

Theorem 2 follows by combining the Laurent polynomial method (Lemma 17) and the following theorem.

Theorem 21. *Let $\varepsilon < 1$. Suppose that p has negative degree D_1 and positive degree D_2 and satisfies the following properties.*

- $|p(w) - 1| \leq \varepsilon$
- $|p(2w) + 1| \leq \varepsilon$
- $|p(\ell)| \leq 1 + \varepsilon$ *for all* $\ell \in \{1, 2, \ldots, n\}$

Then either $D_1 \geq \Omega\left(w^{1/3}\right)$ *or* $D_2 \geq \Omega\left(\sqrt{N/w}\right)$.

In fact, our proof of Theorem 21 will show that the lower bound holds even if $|p(\ell)| \leq 1 + \varepsilon$ only for $\ell \in \{w^{1/3}, w^{1/3} + 1, \ldots, w\} \cup \{2w, 2w + 1, \ldots, N\}$. We refer to a Laurent polynomial p satisfying the three properties of Theorem 21 as an *approximation for approximate counting*.

Proof of Theorem 21. Let p be any Laurent polynomial satisfying the hypothesis of Theorem 21. We begin by transforming p into a (standard) polynomial q in a straightforward manner. This transformation is captured in the following lemma, whose proof is so simple that we omit it.

Lemma 22. *If p satisfies the properties of Theorem 21, then the polynomial $q(r) = p(r) \cdot r^{D_1} = a_0 + a_1 r + \cdots + a_{D_1+D_2} r^{D_1+D_2}$ is a (standard) polynomial of degree at most $D_1 + D_2$, and q satisfies the following three properties.*

- $\left|q(w) - w^{D_1}\right| \leq \varepsilon \cdot w^{D_1}$
- $\left|q(2w) + (2w)^{D_1}\right| \leq \varepsilon \cdot (2w)^{D_1}$
- $|q(\ell)| \leq (1 + \varepsilon)\, \ell^{D_1}$ *for all* $\ell \in \{1, 2, \ldots, N\}$

We now turn to showing that, for any constant $\varepsilon < 1$, no polynomial q can satisfy the conditions of Lemma 22 unless $D_1 \geq \Omega(w^{1/3})$ or $D_2 \geq \Omega\left(\sqrt{N/w}\right)$.

Consider the following linear program. The variables of the linear program are ε, and the $D_2 + D_1 + 1$ coefficients of q.

$$\boxed{\begin{array}{ll} \text{minimize} & \varepsilon \\ \text{such that} & \\ & |q(w) - w^{D_1}| \leq \varepsilon \cdot w^{D_1} \\ & |q(2w) + (2w)^{D_1}| \leq \varepsilon \cdot (2w)^{D_1} \\ & |q(\ell)| \leq (1 + \varepsilon) \cdot \ell^{D_1} \text{ for all } \ell \in \{1, 2, \ldots, N\} \\ & \varepsilon \geq 0 \end{array}} \tag{41}$$

Standard manipulations reveal the dual.

$$\boxed{\begin{array}{l} \text{maximize } \phi(w)\cdot w^{D_1} - \phi(2w)\cdot(2w)^{D_1} - \sum_{\ell\in\{1,\ldots,N\},\ell\notin\{w,2w\}} |\phi(\ell)|\cdot\ell^{D_1} \\ \text{such that} \\ \sum_{\ell=1}^{N}\phi(\ell)\cdot\ell^j = 0 \text{ for } j=0,1,2,\ldots,D_1+D_2 \\ \sum_{\ell=1}^{N}|\phi(\ell)|\cdot\ell^{D_1} = 1 \\ \phi\colon \mathbb{R}\to\mathbb{R} \end{array}} \tag{42}$$

Theorem 21 will follow if we can exhibit a solution ϕ to the dual linear program achieving value $\varepsilon > 0$, for some setting of $D_1 \geq \Omega(w^{1/3})$ and $D_2 \geq \Omega\left(\sqrt{N/w}\right)$.[7] We now turn to this task.

3.4.1 Constructing the dual solution

For a set $T \subseteq \{0,1,\ldots,N\}$, define

$$Q_T(t) = \prod_{i=0,1,\ldots,N, i\notin T}(t-i). \tag{43}$$

Let $c > 2$ be a constant that we will choose later (the bigger we choose c to be, the better the objective value achieved by our final dual witness. But choosing a bigger c will also lower the degrees D_1, D_2 of Laurent polynomials against which our lower bound will hold).

We now define two sets T_1 and T_2. The size of T_1 will

$$d_1 := \lfloor (w/c)^{1/3} \rfloor = \Theta\left(w^{1/3}\right) \tag{44}$$

and the size of T_2 will be d_2 for

$$d_2 := \lfloor \sqrt{N/(cw)} \rfloor = \Theta\left(\sqrt{N/w}\right). \tag{45}$$

Let

$$T_1 = \left\{\lfloor w/(ci^2)\rfloor \colon i = 1,2,\ldots,d_1\right\} \tag{46}$$

and

$$T_2 = \left\{c\cdot i^2\cdot w \colon i=1,2,\ldots,d_2 := \sqrt{N/(cw)}\right\}. \tag{47}$$

Finally, define

$$T = \{w, 2w\} \cup T_1 \cup T_2. \tag{48}$$

At last, define $\Phi\colon \{0,1,\ldots,N\} \to \mathbb{R}$ via

$$\Phi(t) = (-1)^t \cdot \binom{N}{t} \cdot Q_T(t). \tag{49}$$

Our final dual solution ϕ will be a scaled version of Φ. Specifically, Φ itself does not satisfy the second constraint of the dual linear program, that $\sum_{\ell=1}^{N}|\Phi(\ell)|\cdot\ell^{D_1} = 1$. So letting

$$C = \sum_{\ell=1}^{N}|\Phi(\ell)|\cdot\ell^{D_1}, \tag{50}$$

our final dual witness ϕ will be Φ/C.

[7] We will alternatively refer to such dual solutions ϕ as *dual witnesses*, since they act as a witness to the fact that any low-degree Laurent polynomial p approximating the approximate counting problem must have large error.

The sizes of T_1 and T_2. Clearly, under the above definition of T_2, $|T_2| = d_2$ as claimed above. It is not as immediately evident that $|T_1| = d_1$: to establish this, we must show that for distinct $i, j \in \{1, 2, \ldots, d_1\}$, $\lfloor w/(ci^2) \rfloor \neq \lfloor w/(cj^2) \rfloor$. This is handled in the following easy lemma.

Lemma 23. *Let $i \neq j$ be distinct numbers in $\{1, \ldots, d_1\}$ and $c > 2$ be a constant. Then as long as $d_1 < (w/c)^{1/3}$, it holds that $\lfloor w/(ci^2) \rfloor \neq \lfloor w/(cj^2) \rfloor$.*

Proof. Assume without loss of generality that $i > j$. Then $w/(cj^2) - w/(ci^2)$ is clearly minimized when $i = d_1$ and $j = i - 1$. For the remainder of the proof, fix $i = d_1$. In this case,

$$\begin{aligned} w/(cj^2) - w/(ci^2) \geq w/\left(c(i-1)^2\right) - w/\left(ci^2\right) &= \frac{wi^2 - w\,(i-1)^2}{c \cdot i^2 \cdot (i-1)^2} \\ &= \frac{w}{c} \cdot \frac{2i-1}{i^2\,(i-1)^2} \geq \frac{w}{c} \cdot \frac{2i-1}{i^4} \geq \frac{w}{ci^3} \geq 1. \end{aligned} \tag{51}$$

Here, the final inequality holds because $i^3 = d_1^3 \leq w/c$.

Equation (51) implies the lemma, as two numbers whose difference is at least 1 cannot have the same integer floor. ■

Lemma 23 is false for $d_1 = \omega(w^{1/3})$, highlighting on a technical level why one cannot choose d_1 larger than $\Theta(w^{1/3})$ without the entire construction and analysis of Φ breaking down.

3.4.2 Intuition: "gluing together" two simpler dual solutions

Before analyzing the dual witnesses Φ and ϕ constructed in Equation (49) and Equation (50), in this subsection and the next, we provide detailed intuition for why the definitions of Φ and ϕ are natural, and briefly overview their analysis.

A dual witness for purely positive degree (i.e., approximate degree). Suppose we were merely interested in showing an approximate degree lower bound of $\Omega(\sqrt{N/w})$ for approximate counting (i.e., a lower bound on the degree of traditional polynomials that distinguish input w from $2w$, and are bounded at all other integer inputs in $1, \ldots, N$). This is equivalent to exhibiting a solution to the dual linear program with $D_1 = 0$. A valid dual witness ϕ_1 for this simpler case is to also use Equation (49), but to set

$$T = \{w, 2w\} \cup T_2, \tag{52}$$

rather than $T = \{w, 2w\} \cup T_1 \cup T_2$.

We will explain intuition for why Equation (52) is a valid dual solution for the approximate degree of approximate counting in the next subsection. For now, we wish to explain how this construction relates to prior work. In [BT13], for any constant $\delta > 0$, a dual witness is given for the fact that the $(1-\delta)$-approximate degree of OR is $\Omega(\sqrt{N})$. This dual witness *nearly* corresponds to the above, with $w = 1$. Specifically, Bun and Thaler [BT13] use the set $T = \{0, 1\} \cup \{ci^2 \colon i = 1, 2, \ldots, \sqrt{N/c}\}$, and they show that almost all of the "mass" of this dual witness is located on the inputs 0 and 1, i.e.,

$$|\Phi(0)| + |\Phi(1)| \geq (1-\delta) \cdot \sum_{i=2}^{N} |\Phi(i)|\,. \tag{53}$$

Here, the bigger c is chosen to be, the smaller the value of δ for which Equation (53) holds.

In the case of $w = 1$, our dual witness for approximate counting differs from this only in that $\{0, 1\}$ is replaced with $\{1, 2\}$. This is because, in order to show a lower bound for distinguishing input $w = 1$ from input $2w = 2$, we want almost all of the mass to be on inputs $\{1, 2\}$ rather than $\{0, 1\}$ (this is what will ensure that the objective function of the dual linear program is large).

For general w, we want most of the mass of ψ to be concentrated on inputs w and $2w$. Accordingly, relative to the $w = 1$ case, we effectively multiply *all* points in T by w, and one can show that this does not affect the calculation regarding concentration of mass.

A dual witness for purely negative degree. Now, suppose we were merely interested in showing that Laurent polynomials of *purely negative* degree require degree $\Omega(w^{1/3})$ to approximate the approximate counting problem. This is equivalent to exhibiting a solution to the dual linear program with $D_2 = 0$. Then a valid dual witness ϕ_2 for this simpler case is to also use Equation (49), but to set

$$T = \{w, 2w\} \cup T_1. \tag{54}$$

Again, we will give intuition for why this is a valid dual solution in the next subsection (Section 3.4.3). For now, we wish to explain how this construction relates to prior work. Essentially, the $\Omega\left(w^{1/3}\right)$-degree lower bound for Laurent polynomials with *only negative* powers was proved by Zhandry [Zha12, Theorem 7.3]. Translating Zhandry's theorem into our setting is not entirely trivial, and he did not explicitly construct a solution to our dual linear program. However (albeit with significant effort), one can translate his argument to our setting to show that Equation (54) gives a valid dual solution to prove a lower bound against Laurent polynomials with only negative powers.

Gluing them together. The above discussion explains that the key ideas for constructing dual solutions ϕ_1, ϕ_2 witnessing degree lower bounds for Laurent polynomials of *only negative* or *only positive* powers were essentially already known, or at least can be extracted from prior work with enough effort. In this work, we are interested in proving lower bounds for Laurent polynomials with both positive and negative powers. Our dual solution Φ essentially just "glues together" the dual solutions that can be derived from prior work. By this, we mean that the set T of integer points on which our Φ is nonzero is the *union* of the corresponding sets for ϕ_1 and ϕ_2 individually. Moreover, this union is nearly disjoint, as the only points in the intersection of the two sets being unioned are w and $2w$.

Overview of the analysis. To show that we have constructed a valid solution to the dual linear program (Equation (42)), we must establish that (a) Φ is uncorrelated with every polynomial of degree at most $D_1 + D_2$ and (b) Φ is well-correlated with any function g that evaluates to $+1$ on input w, to -1 on input $2w$, and is bounded in $[-1, 1]$ elsewhere. In (b), the correlation is taken with respect to an appropriate weighting of the inputs, that on input $\ell \in [N]$ places mass proportional to ℓ^{D_1}.

The definition of Φ as a "gluing together" of ϕ_1 and ϕ_2 turns out, in a straightforward manner, to ensure that Φ is uncorrelated with polynomials of degree at $D_1 + D_2$. All that remains is to show that Φ is well-correlated with g under the appropriate weighting of inputs. This turns out to be technically demanding, but ultimately can be understood as stemming from the fact that ϕ_1 and ϕ_2

are individually well-correlated with g (albeit, in the case of ϕ_2, under a *different* weighting of the inputs than the weighting that is relevant for Φ).

3.4.3 Intuition via complementary slackness

We now attempt to lend some insight into why the dual witnesses ϕ_1 and ϕ_2 for the purely positive degree and purely negative degree take the form that they do. This section is deliberately slightly imprecise in places, and builds on intuition that has been put forth in prior works proving approximate degree lower bounds via dual witnesses [BT13, Tha16, BKT18].

Notice that ϕ_1 is precisely defined so that $\phi_1(i) = 0$ for any $i \notin \{w, 2w\} \cup T_2$, and similarly $\phi_2(i) = 0$ for any $i \notin \{w, 2w\} \cup T_1$. The intuition for why this is reasonable comes from complementary slackness, which states that an optimal dual witness should equal 0 except on inputs that correspond to primal constraints that are *made tight by an optimal primal solution*. By "constraints made tight by an optimal primal solution", we mean constraints that, for the optimal primal solution, hold with equality rather than (strict) inequality.

Unpacking that statement, this means the following. Suppose that q is an optimal solution to the primal linear program of Section 3.4, meaning it minimizes the error ε amongst all polynomials of the same same degree. The constraints made tight by q are precisely those inputs ℓ at which q hits its "maximum error" (e.g., an input ℓ such that $|q(\ell)| = (1+\varepsilon) \cdot \ell^{D_1}$). We call these inputs *maximum-error* inputs for q. Complementary slackness says that there is an optimal solution to the dual linear program (Equation (42)) that equals 0 at all inputs that are not maximum-error inputs for q.

In both the purely positive degree case, and the purely negative degree case, we know roughly what primal optimal solutions q look like, and moreover we know what roughly their maximum-error points look like. In the first case, the maximum-error points are well-approximated by the points in T_2, and in the purely negative degree case, the maximum error points are well-approximated by the points in T_1. Let us explain.

Purely positive degree case. Let T_d be the degree d Chebyshev polynomial of the first kind. It can be seen that $P(\ell) = T_{\sqrt{N}}(1 + 2/N - \ell/N)$ satisfies $P(1) \geq 2$, while $|P(\ell)| \leq 1$ for $\ell = 2, 3, \ldots, N$. That is, up to scaling, P approximates the approximate counting problem for $w = 1$, and its known that its degree is within a constant factor of optimal.

It is known that the extreme points of T_d are of the following form, for $k = 1, \ldots, d$:

$$\cos\left(\frac{(2k-1)}{2d}\pi\right) \approx 1 - k^2/(2d^2), \tag{55}$$

where the approximation uses the Taylor expansion of the cosine function around 0. Equation (55) means that the extreme points of P are roughly those inputs ℓ such that $1+2/N-\ell/N \approx 1-k^2/(2d^2)$, where $d = \sqrt{N}$. Such ℓ are roughly of the form $\ell \approx c \cdot i^2$ for some constant c, as i ranges from 1 up to $\Theta(N^{1/2})$.

More generally, when $w \geq 1$, an asymptotically optimal approximation for distinguishing input w from $2w$ is $P(\ell) = T_{\sqrt{N/w}}(1 + 2w/N - \ell/(wN))$. The extreme points of P are roughly of the form $\ell \approx c \cdot i^2 \cdot w$ for some constant c, as i ranges from 1 up to $\Theta(\sqrt{N/w})$, which is exactly the form of the points in our set T_2.

Purely negative degree case. In Lemma 18, we exhibited a simple, purely negative degree Laurent polynomial p (i.e., $p(\ell)$ is a standard polynomial in $1/\ell$) with degree $D_1 = w^{1/3}$ that solves the approximate counting problem (the construction is due to MathOverflow user "fedja"). Roughly speaking, p can be written as a product $p(\ell) = u(\ell) \cdot v(\ell)$, where $u(\ell)$ has the roots $\ell = 1, 2, \ldots, w^{1/3}$, and $v(\ell)$ is (an affine transformation) of a Chebyshev polynomial of degree $w^{1/3}$, applied to $1/\ell$. One can easily look at this construction and see that $p(\ell)$ outputs *exactly* the correct value on inputs $\{1, 2, \ldots, w^{1/3}\}$, so these are not maximum error points for p. Moreover, the analysis of the maximum error points for Chebyshev polynomials above can be applied to show that the maximum error points of p are roughly of the form ℓ such that $1/\ell = c \cdot i^2/w$ for some constant c, with i ranging from 1 up to $\Theta(w^{1/3})$. This means that the extreme points are roughly of the form $\ell \approx \frac{w}{ci^2}$, which is why our set T_1 consists of points of the form $\lfloor \frac{w}{ci^2} \rfloor$ (the floors are required because we are proving lower bounds against polynomials whose behavior is only constrained at integer inputs).

3.4.4 Analysis of the dual solution Φ

Lemma 24. *Let $d_1 = |T_1|$ and $d_2 = |T_2|$. Then for any $j = 0, 1, \ldots, d_1 + d_2$, it holds that*

$$\sum_{\ell=1}^{N} \Phi(\ell) \cdot \ell^j = 0.$$

Proof. A basic combinatorial fact is that for any polynomial Q of degree at most $N-1$, the following identity holds:

$$\sum_{\ell=0}^{N} \binom{N}{\ell} (-1)^\ell Q(\ell) = 0. \tag{56}$$

Observe that for any $j \leq d_1 + d_2 + 1$,

$$Q_T(\ell) \cdot \ell^j \text{ is a polynomial in } \ell \text{ of degree at most } N-1. \tag{57}$$

Furthermore, $\Phi(0) = 0$, because $0 \notin T$. Hence

$$\sum_{\ell=0}^{N} \binom{N}{\ell} (-1)^\ell Q_T(\ell) \cdot \ell^j = \sum_{\ell=1}^{N} \binom{N}{\ell} (-1)^\ell Q_T(\ell) \cdot \ell^j. \tag{58}$$

Thus, we can calculate:

$$\begin{aligned} \sum_{\ell=1}^{N} \Phi(\ell) \cdot \ell^j &= \sum_{\ell=1}^{N} (-1)^\ell \cdot \binom{N}{\ell} \cdot Q_T(\ell) \cdot \ell^j \\ &= \sum_{\ell=0}^{N} (-1)^\ell \cdot \binom{N}{\ell} \cdot Q_T(\ell) \cdot \ell^j = 0. \end{aligned}$$

Here, the second equality follows from Equation (58), while the third follows from Equations (56) and (57). ∎

Let us turn to analyzing Φ's value on various inputs. Clearly the following condition holds:

$$\Phi(\ell) = 0 \text{ for all } \ell \notin T. \tag{59}$$

Next, observe that for any $r \in T$,

$$|\Phi(r)| = N! \cdot \frac{1}{\prod_{j \in T, j \neq r} |r-j|}.$$

Consider any quantity $c \cdot i^2 \cdot w \in T_2$. Then

$$\begin{aligned}
\left|\Phi(c \cdot w \cdot i^2)\right| / |\Phi(w)| &= \frac{\prod_{j \in T, j \neq w} |w-j|}{\prod_{j \in T, j \neq c \cdot i^2 \cdot w} |w \cdot c \cdot i^2 - j|} \\
&= \frac{|w-2w| \cdot \left(\prod_{j=1}^{d_2} |w - c \cdot j^2 \cdot w|\right) \cdot \left(\prod_{j=1}^{d_1} \left(w - \left\lfloor \frac{w}{cj^2} \right\rfloor\right)\right)}{|c \cdot i^2 \cdot w - w| \cdot |c \cdot i^2 \cdot w - 2w| \cdot \left(\prod_{j=1, j \neq i}^{d_2} |w \cdot c \cdot i^2 - w \cdot c \cdot j^2|\right) \cdot \left(\prod_{j=1}^{d_1} \left(w \cdot c \cdot i^2 - \left\lfloor \frac{w}{c \cdot j^2} \right\rfloor\right)\right)} \\
&= \frac{c^{d_2} \cdot \left(\prod_{j=1}^{d_2} \left(j^2 - \frac{1}{c}\right)\right) \cdot \prod_{j=1}^{d_1} \left(w - \left\lfloor \frac{w}{c \cdot j^2} \right\rfloor\right)}{(ci^2 - 1) \cdot (ci^2 - 2) \cdot c^{d_2 - 1} \cdot \left(\prod_{j=1, j \neq i}^{d_2} |i^2 - j^2|\right) \cdot \left(\prod_{j=1}^{d_1} \left(w \cdot c \cdot i^2 - \left\lfloor \frac{w}{c \cdot j^2} \right\rfloor\right)\right)} \\
&\leq \frac{c \cdot \left(\prod_{j=1}^{d_2} \left(j^2 - \frac{1}{c}\right)\right) \cdot \prod_{j=1}^{d_1} \left(w - \left\lfloor \frac{w}{c \cdot j^2} \right\rfloor\right)}{(ci^2 - 1) \cdot (ci^2 - 2) \cdot \left(\prod_{j=1, j \neq i}^{d_2} |i^2 - j^2|\right) \cdot \left(\prod_{j=1}^{d_1} \left(w \cdot c \cdot i^2 - \frac{w}{c \cdot j^2}\right)\right)}
\end{aligned} \tag{60}$$

Now, observe that

$$\begin{aligned}
\prod_{j=1}^{d_1} \left(w - \left\lfloor \frac{w}{c \cdot j^2} \right\rfloor\right) &\leq \prod_{j=1}^{d_1} \left(w - \frac{w}{cj^2} + 1\right) = \prod_{j=1}^{d_1} w \cdot \left(1 - \frac{1}{cj^2}\right) \cdot \left(1 + \frac{1}{w \cdot \left(1 - \frac{1}{cj^2}\right)}\right) \\
&\leq \prod_{j=1}^{d_1} w \cdot \left(1 - \frac{1}{cj^2}\right) \left(1 + \frac{1}{(1 - 1/c) \cdot w}\right) \leq \left(\prod_{j=1}^{d_1} w \cdot \left(1 - \frac{1}{cj^2}\right)\right) \cdot (1 + o(1)).
\end{aligned} \tag{61}$$

Hence, we see that Expression (60) is bounded by

$$\begin{aligned}
&\frac{c \cdot \left(\prod_{j=1}^{d_2} \left(j^2 - \frac{1}{c}\right)\right) \cdot \left(\prod_{j=1}^{d_1} \left(1 - \frac{1}{c \cdot j^2}\right)\right) \cdot (1 + o(1))}{(ci^2 - 1) \cdot (ci^2 - 2) \cdot \left(\prod_{j=1, j \neq i}^{d_2} |i^2 - j^2|\right) \cdot \left(\prod_{j=1}^{d_1} \left(c \cdot i^2 - \frac{1}{c \cdot j^2}\right)\right)} \\
&\leq \frac{c \cdot (d_2!)^2 \cdot \left(\prod_{j=1}^{d_1} \left(1 - \frac{1}{c \cdot j^2}\right)\right) \cdot (1 + o(1))}{(ci^2 - 1) \cdot (ci^2 - 2) \cdot \left(\prod_{j=1, j \neq i}^{d_2} |i - j| \, |i + j|\right) \cdot (c \cdot i^2)^{d_1} \cdot \left(\prod_{j=1}^{d_1} \left(1 - \frac{1}{c^2 \cdot i^2 \cdot j^2}\right)\right)} \\
&= \frac{c \cdot (d_2!)^2 \cdot 2i^2 \cdot \left(\prod_{j=1}^{d_1} \left(1 - \frac{1}{c \cdot j^2}\right)\right) \cdot (1 + o(1))}{(ci^2 - 1) \cdot (ci^2 - 2) \cdot (d_2 + i)! \, (d_2 - i)! \cdot (c \cdot i^2)^{d_1} \cdot \left(\prod_{j=1}^{d_1} \left(1 - \frac{1}{c^2 \cdot i^2 \cdot j^2}\right)\right)} \\
&\leq \frac{c \cdot 2i^2 \cdot (d_2!)^2 \cdot (1 + o(1))}{(ci^2 - 1) \, (ci^2 - 2) \cdot (d_2 + i)! \, (d_2 - i)! \cdot (c \cdot i^2)^{d_1}} \leq \frac{2 \, (1 + o(1))}{\left(1 - \frac{1}{c \cdot i^2}\right) \cdot (c \cdot i^2 - 2) \cdot (c \cdot i^2)^{d_1}}.
\end{aligned} \tag{62}$$

In the penultimate inequality, we used the fact that $\frac{(d_2!)^2}{(d_2+i)!(d_2-i)!} = \frac{\binom{2d_2}{d_2+i}}{\binom{2d_2}{d_2}} \leq 1$.

Next, consider any quantity $\left\lfloor \frac{w}{c \cdot i^2} \right\rfloor \in T_1$. Then

$$
\begin{aligned}
&\left| \Phi \left(\left\lfloor \frac{w}{c \cdot i^2} \right\rfloor \right) \right| / \left| \Phi(w) \right| \\
&= \frac{|w - 2w| \left(\prod_{j=1}^{d_2} |w - cj^2 w| \right) \left(\prod_{j=1}^{d_1} \left(w - \left\lfloor \frac{w}{cj^2} \right\rfloor \right) \right)}{\left(w - \left\lfloor \frac{w}{c \cdot i^2} \right\rfloor\right) \cdot \left(2w - \left\lfloor \frac{w}{c \cdot i^2} \right\rfloor\right) \left(\prod_{j=1}^{d_2} \left(w \cdot c \cdot j^2 - \left\lfloor \frac{w}{c \cdot i^2} \right\rfloor \right) \right) \prod_{j=1, j \neq i}^{d_1} \left| \left\lfloor \frac{w}{c \cdot i^2} \right\rfloor - \left\lfloor \frac{w}{c \cdot j^2} \right\rfloor \right|} \\
&\leq \frac{|w - 2w| \left(\prod_{j=1}^{d_2} |w - cj^2 w| \right) \left(\prod_{j=1}^{d_1} \left(w - \left\lfloor \frac{w}{cj^2} \right\rfloor \right) \right)}{\left(w - \frac{w}{c \cdot i^2}\right) \cdot \left(2w - \frac{w}{c \cdot i^2}\right) \left(\prod_{j=1}^{d_2} \left(w \cdot c \cdot j^2 - \frac{w}{c \cdot i^2} \right) \right) \prod_{j=1, j \neq i}^{d_1} \left| \left\lfloor \frac{w}{c \cdot i^2} \right\rfloor - \left\lfloor \frac{w}{c \cdot j^2} \right\rfloor \right|} \\
&\leq \frac{|w - 2w| \left(\prod_{j=1}^{d_2} |w - cj^2 w| \right) \left(\prod_{j=1}^{d_1} \left(w - \frac{w}{cj^2} \right) \right) \cdot (1 + o(1))}{\left(w - \frac{w}{c \cdot i^2}\right) \cdot \left(2w - \frac{w}{c \cdot i^2}\right) \left(\prod_{j=1}^{d_2} \left(w \cdot c \cdot j^2 - \frac{w}{c \cdot i^2} \right) \right) \prod_{j=1, j \neq i}^{d_1} \left| \left\lfloor \frac{w}{c \cdot i^2} \right\rfloor - \left\lfloor \frac{w}{c \cdot j^2} \right\rfloor \right|} \qquad (63)
\end{aligned}
$$

Here, the final inequality used Equation (61). Let us consider the expression $\prod_{j=1, j \neq i}^{d_1} \left| \left\lfloor \frac{w}{c \cdot i^2} \right\rfloor - \left\lfloor \frac{w}{c \cdot j^2} \right\rfloor \right|$. This quantity is *at least*

$$
\begin{aligned}
\prod_{j=1, j \neq i}^{d_1} \left(\left| \frac{w}{c \cdot i^2} - \frac{w}{c \cdot j^2} \right| - 1 \right) &= w^{d_1 - 1} \cdot \prod_{j=1, j \neq i}^{d_1} \frac{\left| j^2 - i^2 \right| - \frac{ci^2 j^2}{w}}{ci^2 j^2} \\
&= w^{d_1 - 1} \cdot \prod_{j=1, j \neq i}^{d_1} \frac{|j - i| \cdot |j + i| - \frac{ci^2 j^2}{w}}{ci^2 j^2} \\
&= \left(\frac{w}{ci^2} \right)^{d_1 - 1} \cdot \prod_{j=1, j \neq i}^{d_1} \frac{|j - i| \cdot |j + i| - \frac{ci^2 j^2}{w}}{j^2} \qquad (64)
\end{aligned}
$$

We claim that Expression (64) is at least

$$
\left(\frac{w}{ci^2} \right)^{d_1 - 1} \cdot \frac{1}{2}. \qquad (65)
$$

In the case that $c = 2$ and d_1 is (at most) $w^{1/3}$, this is precisely [Zha12, Claim 4]. We will ultimately take c to be a constant strictly greater than 2 and hence $d_1 = \lfloor (w/c)^{1/3} \rfloor$ is a constant factor smaller than $w^{1/3}$. The proof of [Zha12, Claim 4] works with cosmetic changes in this case. For completeness, we present a derivation of the claim in Appendix A.

Equation (65) implies that Expression (63) is at most:

$$
\begin{aligned}
&\frac{|w - 2w| \left(\prod_{j=1}^{d_2} |w - cj^2 w| \right) \left(\prod_{j=1}^{d_1} \left(w - \frac{w}{cj^2} \right) \right) \cdot (1 + o(1))}{\left(w - \frac{w}{c \cdot i^2}\right) \cdot \left(2w - \frac{w}{c \cdot i^2}\right) \left(\prod_{j=1}^{d_2} \left(w \cdot c \cdot j^2 - \frac{w}{c \cdot i^2} \right) \right) \left(\frac{w}{ci^2} \right)^{d_1 - 1} \cdot \frac{1}{2}} \\
&\quad = \frac{2 \left(\prod_{j=1}^{d_2} |1 - cj^2| \right) \left(\prod_{j=1}^{d_1} \left(1 - \frac{1}{cj^2} \right) \right) \cdot (1 + o(1))}{\left(1 - \frac{1}{c \cdot i^2}\right) \cdot \left(2 - \frac{1}{c \cdot i^2}\right) \left(\prod_{j=1}^{d_2} \left(c \cdot j^2 - \frac{1}{c \cdot i^2} \right) \right) \left(\frac{1}{ci^2} \right)^{d_1 - 1}} \\
&\quad = \frac{2 \left(\prod_{j=1}^{d_2} (j^2 - 1/c) \right) \left(\prod_{j=1}^{d_1} \left(1 - \frac{1}{cj^2} \right) \right) \cdot (1 + o(1))}{\left(1 - \frac{1}{c \cdot i^2}\right) \cdot \left(2 - \frac{1}{c \cdot i^2}\right) \left(\prod_{j=1}^{d_2} \left(j^2 - \frac{1}{c^2 \cdot i^2} \right) \right) \left(\frac{1}{ci^2} \right)^{d_1 - 1}}
\end{aligned}
$$

$$\leq \frac{2\,(1+o(1))}{\left(1-\frac{1}{c\cdot i^2}\right)\cdot\left(2-\frac{1}{c\cdot i^2}\right)\left(\frac{1}{ci^2}\right)^{d_1-1}} \leq 4\cdot\left(ci^2\right)^{d_1-1}. \tag{66}$$

Summarizing Equations (62) and (66), we have shown that: for any quantity $c\cdot i^2\cdot w\in T_2$,

$$\left|\Phi(c\cdot w\cdot i^2)\right| / |\Phi(w)| \leq \frac{2\,(1+o(1))}{\left(1-\frac{1}{c\cdot i^2}\right)\cdot(c\cdot i^2-2)\cdot(c\cdot i^2)^{d_1}} \tag{67}$$

and for any quantity $\left\lfloor\frac{w}{c\cdot i^2}\right\rfloor\in T_1$,

$$\left|\Phi\left(\left\lfloor\frac{w}{c\cdot i^2}\right\rfloor\right)\right| / |\Phi(w)| \leq 4\cdot\left(ci^2\right)^{d_1-1}. \tag{68}$$

Let $\phi=\Phi/C$, where C is as in Equation (50). Let $D_1=d_1$ and $D_2=d_2$. Lemma 24 implies that ϕ is a feasible solution for the dual linear program of Section 3.4.1. We now show that, for any constant $\delta>0$, by choosing c to be a sufficiently large constant (that depends on δ), we can ensure that ϕ achieves objective value $1-2\delta$.

Let

$$\begin{aligned} A &= |\Phi(w)|\cdot w^{D_1}, \\ B &= |\Phi(2w)|\cdot(2w)^{D_1}, \end{aligned}$$

and

$$E=\sum_{i=1}^{d_1}|\Phi(\lfloor w/ci^2\rfloor)|\cdot\left(\lfloor w/ci^2\rfloor\right)^{D_1}+\sum_{i=1}^{d_2}|\Phi(\lfloor w\cdot ci^2\rfloor)|\cdot\left(w\cdot c\cdot i^2\right)^{D_1}.$$

By Equation (59), $C=A+B+E$.

Moreover, observe that $\operatorname{sgn}(\Phi(w))=-\operatorname{sgn}(\Phi(2w))$, so without loss of generality we may assume $\Phi(w)\geq 0$ and $\Phi(2w)\leq 0$ (if not, then replace Φ with $-\Phi$ throughout).

We now claim that by choosing c to be a sufficiently large constant, we can ensure that $E\leq\delta\cdot A$. To see this, observe that Equations (67) and (68), along with the fact that $D_1=d_1$ and $D_2=d_2$ implies that

$$\begin{aligned} E/A &\leq \frac{1}{w^{D_1}}\left[\left(\sum_{i=1}^{d_1}\left(\lfloor w/ci^2\rfloor\right)^{D_1}\cdot 4\cdot\left(ci^2\right)^{d_1-1}\right)+\left(\sum_{i=1}^{d_2}\left(w\cdot c\cdot i^2\right)^{D_1}\frac{2\left(1-\frac{1}{c\cdot i^2}\right)(1+o(1))}{(c\cdot i^2-2)\cdot(c\cdot i^2)^{d_1}}\right)\right] \\ &\leq \frac{1}{w^{D_1}}\left[\left(\sum_{i=1}^{d_1}\left(w/ci^2\right)^{D_1}\cdot 4\cdot\left(ci^2\right)^{d_1-1}\right)+\left(\sum_{i=1}^{d_2}\left(w\cdot c\cdot i^2\right)^{D_1}\frac{2\left(1-\frac{1}{c\cdot i^2}\right)(1+o(1))}{(c\cdot i^2-2)\cdot(c\cdot i^2)^{d_1}}\right)\right] \\ &\leq 4\left(\sum_{i=1}^{d_1}\frac{1}{c\cdot i^2}\right)+\left(\sum_{i=1}^{d_2}\frac{2\,(1+o(1))}{\left(1-\frac{1}{c\cdot i^2}\right)(c\cdot i^2-2)}\right) \end{aligned}$$

Since $\sum_{i=1}^{\infty}1/(ci^2)\leq\frac{\pi^2}{6c}$, we see that choosing c to be a sufficiently large constant depending on δ ensures that $E/A\leq\delta$ as desired.

Hence, ϕ achieves objective value at least

$$\begin{aligned} &\phi(w)\cdot w^{D_1}-\phi(2w)\cdot(2w)^{D_1}-\sum_{\ell\in\{1,\ldots,N\},\ell\notin\{w,2w\}}|\phi(\ell)|\cdot\ell^{D_1} \\ &\qquad\geq\frac{A+B-E}{A+B+E}\geq\frac{(1-\delta)A+B}{(1+\delta)A+B}\geq 1-2\delta. \end{aligned}$$

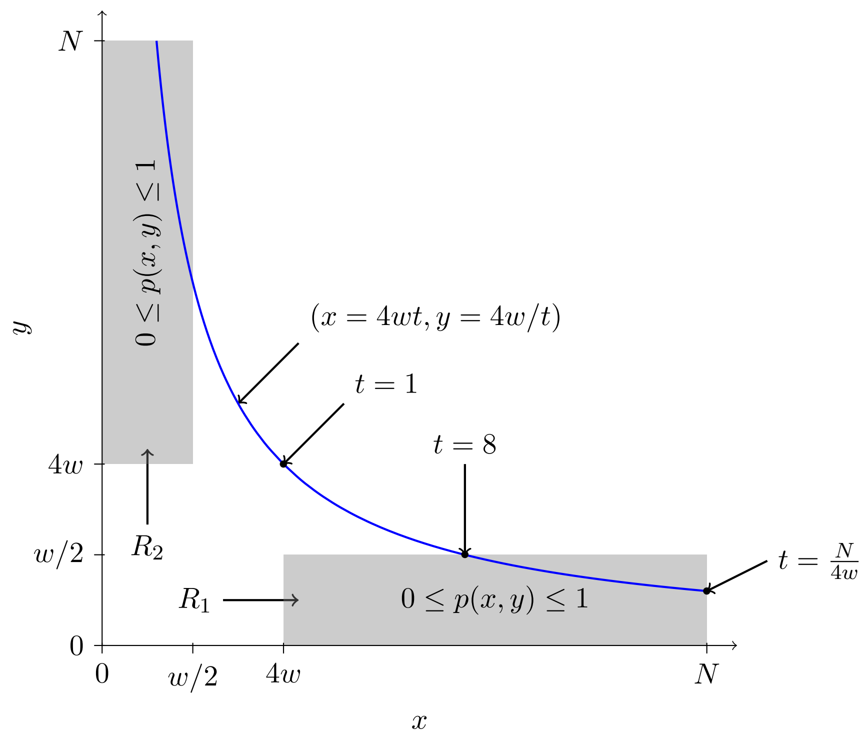

Figure 3: Diagram of Theorem 25 (not drawn to scale).

4 Approximate counting with a quantum witness

As described in the introduction, in this section we first lower bound the SBQP complexity of the $\mathsf{AND}_2 \circ \mathsf{ApxCount}_{N,w}$ problem (Theorem 5), and then use Lemma 16 to prove a lower bound on the QMA complexity of $\mathsf{ApxCount}_{N,w}$ (Theorem 4).

4.1 Lower bound for SBQP algorithms

Our lower bound on the SBQP complexity of $\mathsf{AND}_2 \circ \mathsf{ApxCount}_{N,w}$ hinges on the following theorem. The theorem uses Laurent polynomials to prove a degree lower bound for bivariate polynomials that satisfy an upper bound on an "L"-shaped pair of rectangles and a lower bound at a nearby point:

Theorem 25. *Let $0 < w < 32w < N$. Let $R_1 = [4w, N] \times [0, w/2]$ and $R_2 = [0, w/2] \times [4w, N]$ be disjoint rectangles in the plane, and let $L = R_1 \cup R_2$. Let $p(x, y)$ be a real polynomial of degree d with the following properties:*

1. *$p(4w, 4w) \geq 1.5$.*
2. *$0 \leq p(x, y) \leq 1$ for all $(x, y) \in L$.*

Then $d = \Omega(\sqrt{N/w})$.

Proof. Observe that if $p(x, y)$ satisfies the statement of the theorem, then so does $p(y, x)$. This is because the constraints in the statement of the theorem are symmetric in x and y (in particular, because R_1 and R_2 are mirror images of one another along the line $x = y$; see Figure 3). As a result, we may assume without loss of generality that p is symmetric, i.e., $p(x, y) = p(y, x)$. Else, we may replace p by $\frac{p(x,y)+p(y,x)}{2}$ because the set of polynomials that satisfy the inequalities in the statement of the theorem are closed under convex combinations.

Consider the hyperbolic parametric curve $(x = 4wt, y = 4w/t)$ as it passes through R_1 (see Figure 3). We can view the restriction of $p(x, y)$ to this curve as a Laurent polynomial $\ell(t) = p(4wt, 4w/t)$ of positive and negative degree d. The bound of $p(x, y)$ on all of R_1 implies that $|\ell(t)| \leq 1$ when $t \in [8, \frac{N}{4w}]$ and that $\ell(1) \geq 1.5$ (see Figure 3). Moreover, the condition that $p(x, y)$ is symmetric implies that $\ell(t) = \ell(1/t)$.

By Lemma 14 for symmetric Laurent polynomials, $\ell(t)$ can be viewed as a degree d polynomial $q(t + 1/t)$. Under the transformation $s = t + 1/t$, q satisfies $|q(s)| \leq 1$ for $s \in [8 + 1/8, \frac{N}{4w} + \frac{4w}{N}]$ and $q(2) \geq 1.5$. Note that the length of the interval $[8 + 1/8, \frac{N}{4w} + \frac{4w}{N}]$ is $\Theta(N/w)$ because $w < N$. By an appropriate affine transformation of q, we can conclude from Lemma 7 with $\mu = \Theta(w/N)$ that $d = \Omega(\sqrt{N/w})$. ∎

Why is Theorem 25 useful? One may be tempted to apply this theorem directly to the polynomial $p(x, y)$ obtained in Lemma 13 to conclude a degree lower bound (and thus a query complexity lower bound), as the "L"-shaped pair of rectangles $L = R_1 \cup R_2$ correspond to "no" instances of $\mathsf{AND}_2 \circ \mathsf{ApxCount}_{N,w}$, while $(4w, 4w)$ corresponds to a "yes" instance. However, even though $p(x, y)$ is bounded at lattice points in L, it need not be bounded along the entirety of L.[8]

To obtain a lower bound, we instead use the connection between the polynomials $p(x, y)$ and $q(s, t)$ from Lemma 13, and establish Theorem 5 from the introduction, restated for convenience:

Theorem 5. *Consider an* SBQP *algorithm for* $\mathsf{AND}_2 \circ \mathsf{ApxCount}_{N,w}$ *that makes* T *queries to membership oracles for the two instances of* $\mathsf{ApxCount}_{N,w}$*. Then* $T = \Omega\left(\min\{w, \sqrt{N/w}\}\right)$.

Proof. Let $N > 32w$ (otherwise the theorem holds trivially). Since Q is an SBQP algorithm, we may suppose that Q accepts with probability at least 2α on a "yes" instance and with probability at most α on a "no" instance (note that α may be exponentially small in N). Take $p(x, y)$ and $q(s, t)$ to be the symmetrized bivariate polynomials of degree at most $2T$ defined in Lemma 13. Define $L' = ([0, w] \times [0, w]) \cup ([0, w] \times [2w, N]) \cup ([2w, N] \times [0, w])$. The conditions on the acceptance probability of Q for all S_0, S_1 that satisfy the $\mathsf{ApxCount}_{N,w}$ promise imply that $p(x, y)$ satisfies these corresponding conditions:

1. $1 \geq p(x, y) \geq 2\alpha$ for all $(x, y) \in ([2w, N] \times [2w, N]) \cap \mathbb{Z}^2$.
2. $0 \leq p(x, y) \leq \alpha$ for all $(x, y) \in L' \cap \mathbb{Z}^2$.

Our strategy is to show that if $T = o(w)$, then these conditions on p imply that the polynomial $q(s, t) \cdot \frac{0.9}{\alpha}$ satisfies the statement of Theorem 25 for all sufficiently large w. This in turn implies $T = \Omega(\sqrt{N/w})$. This allows us conclude that either $T = \Omega(w)$ or $T = \Omega(\sqrt{N/w})$, which proves the theorem.

[8] One can nevertheless use this intuition to obtain a nontrivial (though suboptimal) lower bound by inspecting p alone. Using the Markov brothers' inequality (Lemma 6), if $\deg(p) = o(\sqrt{w})$, then the bounds on $p(x, y)$ at lattice points in L imply that $|p(x, y)| \leq 1 + o_w(1)$ for all $(x, y) \in L$. Thus, Theorem 25 applies if $\deg(p) = o(\sqrt{w})$, so overall we get a lower bound of $\Omega\left(\min\left\{\sqrt{w}, \sqrt{N/w}\right\}\right)$ for the SBQP query complexity of $\mathsf{AND}_2 \circ \mathsf{ApxCount}_{N,w}$. See arXiv:1902.02398 for details.

Suppose $T = o(w)$, so that $p(x, y)$ and $q(s, t)$ both have degree $o(w)$. We begin by bounding $p(x, y)$ at the lattice points (x, y) outside of L'. We claim the following:

(a) $|p(x, y)| \le \alpha \cdot a \cdot \exp(bd^2/w) \le \alpha \cdot a \cdot \exp(bd)$ whenever $(x, y) \in L'$ and either x or y is an integer, where a and b are the constants from Lemma 10. This follows from Lemma 10 by fixing either x or y to be an integer and viewing the resulting restriction of $p(x, y)$ as a univariate polynomial in the other variable.

(b) $|p(x, y)| \le \alpha \cdot a \cdot \exp(bd) \cdot \exp(2\sqrt{3}d) = \alpha \cdot a \cdot \exp((b + 2\sqrt{3})d)$ whenever $x \in [w, 2w]$, $y \in [0, w]$, and y is an integer. This follows Lemma 7: consider the univariate polynomial $p(\cdot, y)$ on the intervals $[0, w]$ and $[2w, 3w]$, where it is bounded by (a).

(c) $|p(x, y)| \le \alpha \cdot a \cdot \exp((b+2\sqrt{3})d) \cdot a \cdot \exp(bd^2/w) \le \alpha \cdot a^2 \cdot \exp((2b+2\sqrt{3})d)$ whenever $x \in [w, 2w]$ and $y \in [0, w]$. This follows from Lemma 10: consider the univariate polynomial $p(x, \cdot)$ on the interval $[0, w]$, where it is bounded at integer points by (b).

(d) $|p(x, y)| \le \alpha \cdot a^2 \cdot \exp((2b + 2\sqrt{3})d) \cdot \exp(4dy/w) = \alpha \cdot a^2 \cdot \exp((2b + 2\sqrt{3} + 4y/w)d)$ whenever $x \in [0, N]$, $y \in [w + 1, N]$, and x is an integer. This follows from Lemma 7: consider the univariate polynomial $p(x, \cdot)$ on the interval $[0, w]$, where it is bounded by (a) when $x \in [0, w]$ or $x \in [2w, N]$, or bounded by (c) when $x \in [w, 2w]$. By an affine shift, this corresponds to applying Lemma 7 with $\mu = 2y/w - 2$, with the observation that $\sqrt{2\mu + \mu^2} < \mu + 2$.

We now use this to upper bound $q(s, t)$ when $s \in [4w, N]$ and $t \in [0, w/2]$. Let X and Y be drawn from N-trial binomial distributions with means s and t, respectively, so that $q(s, t) = \mathbb{E}[p(X, Y)]$. Using the above bounds and basic probability, we have

$$\begin{aligned}
\mathbb{E}[p(X, Y)] \le \alpha \cdot \Bigg(&\Pr[X \ge 2w, Y \le w] + \Pr[X \le 2w, Y \le w] \cdot a \cdot \exp\left(\left(b + 2\sqrt{3}\right) d\right) + \\
&+ \sum_{y=w+1}^{N} \Pr[Y = y] \cdot a^2 \cdot \exp\left(\left(2b + 2\sqrt{3} + 4y/w\right) d\right) \Bigg) \qquad (69) \\
\le \alpha \cdot \Bigg(&1 + \Pr[X \le 2w] \cdot a \cdot \exp\left(\left(b + 2\sqrt{3}\right) d\right) + \\
&+ \sum_{y=w+1}^{N} \Pr[Y \ge y] \cdot a^2 \cdot \exp\left(\left(2b + 2\sqrt{3} + 4y/w\right) d\right) \Bigg). \qquad (70)
\end{aligned}$$

The probabilities above are easily bounded with a Chernoff bound:

$$\begin{aligned}
\mathbb{E}[p(X, Y)] \le \alpha \cdot \Bigg(1 + a \cdot \exp\left(\left(b + 2\sqrt{3}\right) d - w/2\right) + \\
+ \sum_{y=w+1}^{N} a^2 \cdot \exp\left(\left(2b + 2\sqrt{3} + 4y/w\right) d - y/6\right) \Bigg). \qquad (71)
\end{aligned}$$

Because a and b are universal constants from Lemma 10, when $d = o(w)$, the first exponential term becomes arbitrarily small for all sufficiently large w. Moreover, for all sufficiently large w, the remaining sum becomes bounded by a geometric sum. For some constant c, we have

$$\sum_{y=w+1}^{N} a^2 \cdot \exp\left(\left(2b + 2\sqrt{3} + 4y/w\right) d - y/6\right) \le \sum_{y=w+1}^{\infty} c \cdot \exp\left(-y/12\right)$$

$$\leq \frac{c}{1-\exp(-1/12)} \cdot \exp(-w/12)$$
$$= o_w(1).$$

Thus we conclude that $0 \leq q(s,t) \leq \alpha \cdot (1+o_w(1))$ when $s \in [4w, N]$ and $t \in [0, w/2]$ (i.e., $(s,t) \in R_1$ in the statement of Theorem 25). By symmetry, we can conclude the same bound when $s \in [0, w/2]$ and $t \in [4w, N]$ (i.e., $(s,t) \in R_2$ in the statement of Theorem 25).

Now, we lower bound $q(4w, 4w)$. Let X and Y be drawn from independent N-trial binomial distributions with mean $4w$, so that $q(4w, 4w) = \mathbb{E}[p(X,Y)]$. Then we have

$$\begin{aligned} \mathbb{E}[p(X,Y)] &\geq 2\alpha \cdot \Pr[X \geq 2w, Y \geq 2w] \\ &\geq 2\alpha \cdot (1 - \Pr[X \leq 2w] - \Pr[Y \leq 2w]) \\ &\geq 2\alpha \cdot (1 - 2\exp(-w/2)) \\ &\geq 2\alpha \cdot (1 - o_w(1)) \end{aligned}$$

We conclude that $q(s,t) \cdot \frac{0.9}{\alpha}$ satisfies the statement of Theorem 25 for all sufficiently large w. ■

We remark that this lower bound is tight, i.e., there exists an SBQP algorithm that makes $O\left(\min\left\{w, \sqrt{N/w}\right\}\right)$ queries. The $O(\sqrt{N/w})$ upper bound follows from the BQP algorithm of Brassard, Høyer, and Tapp [BHT98a]. The $O(w)$ upper bound is in fact an SBP upper bound with the following algorithmic interpretation: first, guess $w+1$ items randomly from each of S_0 and S_1. Then, verify using the membership oracle that the first $w+1$ items all belong to S_0 and that the latter $w+1$ items all belong to S_1, accepting if and only if this is the case. Clearly, this accepts with nonzero probability if and only if $|S_0| \geq w+1$ and $|S_1| \geq w+1$.

4.2 Lower bound for QMA

We now show some implications for QMA that follow from the SBQP lower bound of Theorem 5. Using the guessing lemma (Lemma 16), we can quantitatively lower bound the QMA complexity of $\mathsf{ApxCount}_{N,w}$ as stated in the introduction:

Theorem 4. *Consider a QMA protocol that solves $\mathsf{ApxCount}_{N,w}$. If the protocol receives a quantum witness of length m, and makes T queries to the membership oracle for S, then*

$$m \cdot T = \Omega\left(\min\left\{w, \sqrt{N/w}\right\}\right). \tag{5}$$

Proof. Running the verifier, V, a constant number of times with fresh witnesses to reduce the soundness and completeness errors, one obtains a verifier with soundness and completeness errors $1/6$ that receives an $O(m)$-length witness and makes $O(T)$ queries. Repeating twice with two oracles and computing the AND, one obtains a QMA verifier $V'^{\mathcal{O}_{S_0}, \mathcal{O}_{S_1}}$ for $\mathsf{AND}_2 \circ \mathsf{ApxCount}_{N,w}$ with soundness and completeness errors $1/3$ that receives an $O(m)$-length witness and makes $O(T)$ queries. Applying the guessing lemma (Lemma 16) to V', there exists an SBQP algorithm $Q^{\mathcal{O}_{S_0}, \mathcal{O}_{S_1}}$ for $\mathsf{AND}_2 \circ \mathsf{ApxCount}_{N,w}$ that makes $O(m \cdot T)$ queries. Theorem 5 tells us that $m \cdot T = \Omega\left(\min\left\{w, \sqrt{N/w}\right\}\right)$. ■

Theorem 5 also implies several oracle separations:

Corollary 26. *There exists an oracle A and a pair of languages L_0, L_1 such that:*

1. $L_0, L_1 \in \mathsf{SBP}^A$
2. $L_0 \cap L_1 \notin \mathsf{SBQP}^A$.
3. $\mathsf{SBP}^A \not\subset \mathsf{QMA}^A$.

Proof. For an arbitrary function $A : \{0,1\}^* \to \{0,1\}$ and $i \in \{0,1\}$, define $A_i^n = \{x \in \{0,1\}^n : A(i,x) = 1\}$. Define the unary language $L_i^A = \{1^n : |A_i^n| \geq 2^{n/2}\}$. Observe that as long as A satisfies the promise $|A_i^n| \geq 2^{n/2}$ or $|A_i^n| \leq 2^{n/2-1}$ for all $n \in \mathbb{N}$, then $L_i^A \in \mathsf{SBP}^A$. Intuitively, the oracles A that satisfy this promise encode a pair of $\mathsf{ApxCount}_{N,w}$ instances $|A_0^n|$ and $|A_1^n|$ for every $n \in \mathbb{N}$ where $N = 2^n$ and $w = 2^{n/2-1}$.

Theorem 5 tells us that an SBQP algorithm Q that makes $o(2^{n/4})$ queries fails to solve $\mathsf{AND}_2 \circ \mathsf{ApxCount}_{N,w}$ on *some* pair (S_0, S_1) that satisfies the promise. Thus, one can construct an A such that $L_0, L_1 \in \mathsf{SBP}^A$ and $L_0 \cap L_1 \notin \mathsf{SBQP}^A$, by choosing (A_0^n, A_1^n) so as to diagonalize against all SBQP algorithms.

Because QMA^A is closed under intersection for any oracle A, and because $\mathsf{QMA}^A \subseteq \mathsf{SBQP}^A$ for any oracle A, it must be the case that either $L_0 \notin \mathsf{QMA}^A$ or $L_1 \notin \mathsf{QMA}^A$. ■

We remark that this gives an alternative construction of an oracle relative to which SBP is not closed under intersection. To our knowledge, this is the first that uses the polynomial method directly.

5 Discussion and open problems

5.1 Approximate counting with samples and queries only

If we consider the model where we only have membership queries and samples (but no reflections), then the best upper bound we can show is $O\left(\min\left\{\sqrt{w}, \sqrt{N/w}\right\}\right)$, using the sampling algorithm that looks for birthday collisions, and the quantum counting algorithm. It would be interesting to improve the lower bound further in this case, but it is clear that the Laurent polynomial approach cannot do so, since it hits a limit at $w^{1/3}$. Hence a new approach is needed to tackle the model without reflections.

We now give what we think is a viable path to solve this problem. Specifically, we observe that our problem—of lower-bounding the number of copies of $|S\rangle$ *and* the number of queries to $\mathcal{O}_S$ needed for approximate counting of S—can be reduced to a pure problem of lower-bounding the number of copies of $|S\rangle$. To do so, we use a hybrid argument, closely analogous to an argument recently given by Zhandry [Zha17] in the context of quantum money.

Given a subset $S \subseteq [L]$, let $|S\rangle$ be a uniform superposition over S elements. Then let

$$\rho_{L,w,k} := \mathbb{E}_{S \subseteq [L] \,:\, |S| = w}\left[(|S\rangle\langle S|)^{\otimes k}\right] \tag{72}$$

be the mixed state obtained by first choosing S uniformly at random subject to $|S| = w$, then taking k copies of $|S\rangle$. Given two mixed states ρ and σ, recall also that the *trace distance*, $\|\rho - \sigma\|_{\mathrm{tr}}$, is the maximum bias with which ρ can be distinguished from σ by a single-shot measurement.

Theorem 27. *Let $2w \leq L \leq N$. Suppose $\|\rho_{L,w,k} - \rho_{L,2w,k}\|_{\mathrm{tr}} \leq \frac{1}{10}$. Then any quantum algorithm Q requires either $\Omega\left(\sqrt{\frac{N}{L}}\right)$ queries to $\mathcal{O}_S$ or else $\Omega(k)$ copies of $|S\rangle$ to decide whether $|S| = w$ or $|S| = 2w$ with success probability at least $2/3$, promised that one of those is the case.*

Proof. Choose a subset $S \subseteq [N]$ uniformly at random, subject to $|S| = w$ or $|S| = 2w$, and consider S to be fixed. Then suppose we choose $U \subseteq [N]$ uniformly at random, subject to both $|U| = L$ and $S \subseteq U$. Consider the hybrid in which Q is still given R copies of the state $|S\rangle$, but now gets oracle access to $\mathcal{O}_U$ rather than $\mathcal{O}_S$. Then so long as Q makes $o\left(\sqrt{\frac{N}{L}}\right)$ queries to its oracle, we claim that Q cannot distinguish this hybrid from the "true" situation (i.e., the one where Q queries $\mathcal{O}_S$) with $\Omega(1)$ bias. This claim follows almost immediately from the BBBV Theorem [BBBV97]. In effect, Q is searching the set $[N] \setminus S$ for any elements of $U \setminus S$ (the "marked items," in this context), of which there are $L - |S|$ scattered uniformly at random. In such a case, we know that $\Omega\left(\sqrt{\frac{N-|S|}{L-|S|}}\right) = \Omega\left(\sqrt{\frac{N}{L}}\right)$ quantum queries are needed to detect the marked items with constant bias.

Next suppose we first choose $U \subseteq [N]$ uniformly at random, subject to $|U| = L$, and consider U to be fixed. We then choose $S \subseteq U$ uniformly at random, subject to $|S| = w$ or $|S| = 2w$. Note that this produces a distribution over (S, U) pairs identical to the distribution that we had above. In this case, however, since U is fixed, queries to $\mathcal{O}_U$ are no longer relevant. The only way to decide whether $|S| = w$ or $|S| = 2w$ is by using our copies of $|S\rangle$—of which, by assumption, we need $\Omega(k)$ to succeed with constant bias, even after having fixed U. ∎

One might think that Theorem 27 would lead to immediate improvements to our lower bound for the queries and samples model. In practice, however, the best lower bounds that we currently have, even purely on the number of copies of $|S\rangle$, come from the Laurent polynomial method (Theorem 2)! Having said that, we are optimistic that one could obtain a lower bound that beats Theorem 2 at least when w is small, by combining Theorem 27 with a brute-force computation of trace distance.

5.2 Approximate counting to multiplicative factor $1+\varepsilon$

Throughout, we considered the task of approximating $|S|$ to within a multiplicative factor of 2. But suppose our task was to distinguish the case $|S| \leq w$ from the case $|S| \geq (1+\varepsilon) w$; then what is the optimal dependence on ε? As discussed in Section 1, with only membership queries the quantum query complexity of this problem is $O\left(\frac{1}{\varepsilon}\sqrt{\frac{N}{w}}\right)$. One can also show without too much difficulty that in the queries+QSamples model, the problem can be solved with

$$O\left(\min\left\{\frac{\sqrt{w}}{\varepsilon^2}, \frac{1}{\varepsilon}\sqrt{\frac{N}{w}}\right\}\right) \tag{73}$$

queries and copies of $|S\rangle$.

As observed after Theorem 3, the problem can also be solved with

$$O\left(\min\left\{\frac{w^{1/3}}{\varepsilon^{2/3}}, \frac{1}{\varepsilon}\sqrt{\frac{N}{w}}\right\}\right) \tag{74}$$

samples and reflections. On the lower bound side, what generalizations of Theorem 2 can we prove that incorporate ε? We note that the explosion argument doesn't automatically generalize; one would need to modify something to continue getting growth in the polynomials u and v after the first iteration. The lower bound using dual polynomials should generalize, but back-of-the-envelope calculations show that the lower bound does not match the upper bound.

5.3 Other questions

Non-oracular example of our result. Is there any interesting real-world example of a class of sets for which QSampling and membership testing are both efficient, but approximate counting is not? (I.e., is there an interesting non-black-box setting that appears to exhibit the behavior that this paper showed can occur in the black-box setting?)

Improved QMA lower bounds. The QMA lower bound trade-off between m and T for $\mathsf{ApxCount}_{N,w}$ is not optimal in general: when $w = O(1)$, there is no QMA protocol for $\mathsf{ApxCount}_{N,w}$ that receives a constant size witness and makes a constant number of queries for large N. Since our SBQP lower bound is optimal, proving better QMA lower bounds will require new techniques.

The Laurent polynomial connection. At a deeper level, is there is any meaningful connection between our two uses of Laurent polynomials? And what other applications can be found for the Laurent polynomial method?

Acknowledgments

We are grateful to many people: Paul Burchard, for suggesting the problem of approximate counting with queries and QSamples; MathOverflow user "fedja" for letting us include Lemmas 18 and 19; Ashwin Nayak, for extremely helpful discussions, and for suggesting the transformation of linear programs used in our extension of the method of dual polynomials to the Laurent polynomial setting; Thomas Watson, for suggesting the intersection approach to proving an SBP vs. QMA oracle separation; and Patrick Rall, for helpful feedback on writing.

A Establishing Equation 65

A.1 A clean calculation establishing a loose version of equation 65

For clarity of exposition, we begin by presenting a relatively clean calculation that establishes a slightly loose version of Equation (65). Using just this looser bound, we would be able to establish that Equation (65) holds (with the constant $1/2$ replaced by a slightly smaller constant) so long as we set d_1 to be $\Theta\left(w^{1/3}/\log w\right)$. A slightly more involved calculation (cf. Appendix A.2) is required to establish Equation (65) for our desired value of $d_1 = \lfloor (w/c)^{1/3} \rfloor$.

Expression (64) equals

$$\left(\frac{w}{ci^2}\right)^{d_1-1} \cdot \frac{i^2}{((d_1)!)^2} \cdot \prod_{j=1, j\neq i}^{d_1} \left(|j-i| \cdot |j+i| - \frac{ci^2 j^2}{w}\right)$$

$$= \left(\frac{w}{ci^2}\right)^{d_1-1} \cdot \frac{i^2}{((d_1)!)^2} \cdot \prod_{j=1,j\neq i}^{d_1} (|j-i| \cdot |j+i|) \cdot \left(1 - \frac{ci^2 j^2}{w \cdot |j-i||j+i|}\right)$$

$$= \left(\frac{w}{ci^2}\right)^{d_1-1} \cdot \frac{(d_1+i)!(d_1-i)!}{2\left((d_1)!\right)^2} \cdot \prod_{j=1,j\neq i}^{d_1} \left(1 - \frac{ci^2 j^2}{w \cdot |j-i||j+i|}\right) \tag{75}$$

$$\geq \left(\frac{w}{ci^2}\right)^{d_1-1} \cdot \frac{1}{2} \cdot \prod_{j=1,j\neq i}^{d_1} \left(1 - \frac{ci^2 j^2}{w \cdot |j-i||j+i|}\right)$$

$$\geq \left(\frac{w}{ci^2}\right)^{d_1-1} \cdot \frac{1}{2} \cdot \left(1 - \sum_{j=1,j\neq i}^{d_1} \frac{ci^2 j^2}{w \cdot |j-i||j+i|}\right)$$

$$\geq \left(\frac{w}{ci^2}\right)^{d_1-1} \cdot \frac{1}{2} \cdot \left(1 - \frac{ci^2}{w} \sum_{j=1,j\neq i}^{d_1} \frac{j^2}{|j-i||j+i|}\right). \tag{76}$$

Let us consider the expression $\sum_{j=1,j\neq i}^{d_1} \frac{j^2}{|j-i||j+i|}$. If $i^2 \notin [j^2/2, 3j^2/2]$, then the j'th term in this sum is at most 2. Hence, letting H_i denote the ith Harmonic number and using the fact that $H_i \leq \ln(i+1)$,

$$\begin{aligned}
&\sum_{j=1,j\neq i}^{d_1} \frac{j^2}{|j-i||j+i|} \\
&\leq 2 \cdot d_1 + \sum_{j=\lfloor\sqrt{2/3}i\rfloor}^{\lfloor\sqrt{2}i\rfloor} \frac{j^2}{|j-i||j+i|} \\
&\leq 2 \cdot d_1 + \sum_{j=\lfloor\sqrt{2/3}\cdot i\rfloor}^{\lceil\sqrt{2}\cdot i\rceil} \frac{j}{|j-i|} \\
&\leq 2d_1 + \sqrt{2} \cdot i \cdot \sum_{j=1}^{(\sqrt{2}-1)\cdot i} 2/j \\
&\leq 2d_1 + 2\sqrt{2} \cdot i \cdot H_i \leq 2d_1 + 2\sqrt{2} i \ln(i+1).
\end{aligned} \tag{77}$$

We conclude that if d_1 were set to a value less than $w^{1/3}/(100 \cdot c^2 \cdot \ln(w))$ (rather than to $\lfloor (w/c)^{1/3} \rfloor$), then Expression (76) is at least

$$\left(\frac{w}{ci^2}\right)^{d_1-1} \cdot \frac{1-1/c}{2}. \tag{78}$$

A.2 The tight bound

To obtain the tight bound, we need a tighter sequence of inequalities following Expression (75). Specifically, Expression (75) is bounded below by:

$$\geq \left(\frac{w}{ci^2}\right)^{d_1-1} \cdot \frac{1}{2}\left(1+\frac{i}{2d_1}\right)^i \cdot \prod_{j=1, j\neq i}^{d_1}\left(1-\frac{ci^2j^2}{w\cdot|j-i||j+i|}\right)$$

$$\geq \left(\frac{w}{ci^2}\right)^{d_1-1} \cdot \frac{1}{2} \cdot e^{i^2/(2d_1)} \cdot \prod_{j=1, j\neq i}^{d_1}\left(1-\frac{ci^2j^2}{w\cdot|j-i||j+i|}\right)$$

$$\geq \left(\frac{w}{ci^2}\right)^{d_1-1} \cdot \frac{1}{2} \cdot e^{i^2/(2d_1)} \cdot \prod_{j=1, j\neq i}^{d_1}\left(1-\frac{ci^2j^2}{w\cdot|j-i||j+i|}\right) \tag{79}$$

The rough idea of how to proceed is as follows. Equation (77) implies that for $i \ll w^{1/3}/\ln w$, the factor

$$F_1 := \prod_{j=1, j\neq i}^{d_1}\left(1-\frac{ci^2j^2}{w\cdot|j-i||j+i|}\right)$$

is at some a positive constant, and hence Expression (79) is bounded below by the desired quantity. If $i \gtrsim w^{1/3}/\ln w$, then Equation (77) does not yield a good bound on this factor, leaving open the possibility that this factor is subconstant. But in this case, the factor $F_2 := e^{i^2/(2d_1)} \geq e^{\tilde{\Omega}(d_1)}$, and the largeness of F_2 dominates the smallness of F_1.

In more detail, let $x_{i,j} = \frac{ci^2j^2}{w\cdot|i-j|j+i||}$. Then for all $i \neq j$ such that $i, j \leq d_1$,

$$x_{i,j} \leq \frac{c\cdot d_1^2(d_1-1)^2}{(2d_1-1)\cdot w} \leq \frac{c\cdot d_1^3}{2w} \leq 1/2, \tag{80}$$

where in the final inequality we used the fact that $d_1 \leq (w/c)^{1/3}$.

Using the fact that $1-x \geq e^{-x-x^2}$ for all $x \in [0, 1/2]$, we can write

$$F_1 \geq \prod_{j=1, j\neq i}^{d_1} e^{-x_{i,j}-x_{i,j}^2}.$$

Hence,

$$F_1 \cdot F_2 \geq \exp\left(i^2/(2d_1) - \sum_{j=1, j\neq i}^{d_1} -x_{i,j} - x_{i,j}^2\right).$$

From Equations (77) and (80), we know that

$$\sum_{j=1, j\neq i}^{d_1} x_{i,j} + x_{i,j}^2 \leq \frac{ci^2}{w}\cdot\left(3d_1 + 3\sqrt{2}i\ln(i+1)\right) \leq \frac{ci^2}{w}\cdot(4d_1\ln(d_1)).$$

Hence,

$$F_1 \cdot F_2 \geq \exp\left(i^2/(2d_1) - \frac{ci^2}{w}\cdot 4c\ln(d_1)\right)$$

$$
\begin{aligned}
&= \exp\left(i^2\left(\frac{1}{2d_1} - \frac{4c^2\ln(d_1)}{w}\right)\right) \\
&\geq \exp\left(i^2 \cdot \frac{1}{2d_1} \cdot (1 - o(1))\right) \\
&\geq 1.
\end{aligned}
$$

Equation (65) follows.

References

[Aar02] Scott Aaronson. Quantum lower bound for the collision problem. In *Proceedings of the thirty-fourth annual ACM symposium on Theory of computing - STOC 2002*, pages 635–642, 2002. quant-ph/0111102. doi:10.1145/509907.509999. [pp. 4, 5]

[Aar05a] Scott Aaronson. Limitations of quantum advice and one-way communication. *Theory of Computing*, 1(1):1–28, 2005. Earlier version in CCC'2004. quant-ph/0402095. doi:10.4086/toc.2005.v001a001. [p. 5]

[Aar05b] Scott Aaronson. Quantum computing, postselection, and probabilistic polynomial-time. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 461(2063):3473–3482, 2005. quant-ph/0412187. doi:10.1098/rspa.2005.1546. [p. 6]

[Aar12] Scott Aaronson. Impossibility of succinct quantum proofs for collision-freeness. *Quantum Info. Comput.*, 12(1-2):21–28, January 2012. URL: http://dl.acm.org/citation.cfm?id=2231036.2231039. [pp. 8, 16]

[AS04] Scott Aaronson and Yaoyun Shi. Quantum lower bounds for the collision and the element distinctness problems. *Journal of the ACM*, 51(4):595–605, 2004. doi:10.1145/1008731.1008735. [pp. 4, 5]

[ATS03] Dorit Aharonov and Amnon Ta-Shma. Adiabatic quantum state generation and statistical zero knowledge. In *Proceedings of the thirty-fifth ACM symposium on Theory of computing - STOC 2003*, pages 20–29, 2003. quant-ph/0301023. doi:10.1145/780542.780546. [p. 4]

[BBBV97] Charles H. Bennett, Ethan Bernstein, Gilles Brassard, and Umesh Vazirani. Strengths and weaknesses of quantum computing. *SIAM Journal on Computing*, 26(5):1510–1523, 1997. quant-ph/9701001. doi:10.1137/S0097539796300933. [pp. 3, 7, 36]

[BBC+01] Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald de Wolf. Quantum lower bounds by polynomials. *Journal of the ACM*, 48(4):778–797, 2001. Earlier version in FOCS'1998, pp. 352-361. quant-ph/9802049. doi:10.1145/502090.502097. [pp. 5, 8, 13, 14, 16]

[BdW02] Harry Buhrman and Ronald de Wolf. Complexity measures and decision tree complexity: a survey. *Theoretical Computer Science*, 288:21–43, 2002. doi:10.1016/s0304-3975(01)00144-x. [p. 3]

[BGM06] Elmar Böhler, Christian Glaßer, and Daniel Meister. Error-bounded probabilistic computations between MA and AM. *Journal of Computer and System Sciences*, 72(6):1043 – 1076, 2006. doi:10.1016/j.jcss.2006.05.001. [pp. 10, 15]

[BHMT02] Gilles Brassard, Peter Høyer, Michele Mosca, and Alain Tapp. *Quantum amplitude amplification and estimation*, volume 305, pages 53–74. American Mathematical Society, 2002. doi:10.1090/conm/305/05215. [pp. 3, 7, 19]

[BHT98a] Gilles Brassard, Peter Høyer, and Alain Tapp. Quantum counting. In *Automata, Languages and Programming*, pages 820–831, 1998. arXiv:quant-ph/9805082. doi:10.1007/bfb0055105. [pp. 3, 5, 7, 34]

[BHT98b] Gilles Brassard, Peter Høyer, and Alain Tapp. Quantum cryptanalysis of hash and claw-free functions. In *LATIN'98: Theoretical Informatics*, pages 163–169, 1998. doi:10.1007/BFb0054319. [pp. 7, 18]

[BKT18] Mark Bun, Robin Kothari, and Justin Thaler. The polynomial method strikes back: tight quantum query bounds via dual polynomials. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 297–310, 2018. doi:10.1145/3188745.3188784. [p. 26]

[BT13] Mark Bun and Justin Thaler. Dual lower bounds for approximate degree and Markov-Bernstein inequalities. In *Automata, Languages, and Programming*, pages 303–314, 2013. doi:10.1007/978-3-642-39206-1_26. [pp. 6, 24, 26]

[BT15] Mark Bun and Justin Thaler. Hardness amplification and the approximate degree of constant-depth circuits. In *Automata, Languages, and Programming*, pages 268–280, 2015. doi:10.1007/978-3-662-47672-7_22. [p. 8]

[CR92] Don Coppersmith and T. J. Rivlin. The growth of polynomials bounded at equally spaced points. *SIAM J. Math. Anal.*, 23(4):970–983, July 1992. doi:10.1137/0523054. [p. 13]

[DFK91] Martin Dyer, Alan Frieze, and Ravi Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *Journal of the ACM*, 38(1):1–17, jan 1991. Earlier version in STOC'1989. doi:10.1145/102782.102783. [p. 4]

[EZ64] H. Ehlich and K. Zeller. Schwankung von Polynomen zwischen Gitterpunkten. *Mathematische Zeitschrift*, 86:41–44, 1964. doi:10.1007/BF01111276. [p. 12]

[GLM+16] Mika Göös, Shachar Lovett, Raghu Meka, Thomas Watson, and David Zuckerman. Rectangles are nonnegative juntas. *SIAM Journal on Computing*, 45(5):1835–1869, 2016. doi:10.1137/15M103145X. [p. 8]

[GR02] Lov Grover and Terry Rudolph. Creating superpositions that correspond to efficiently integrable probability distributions. arXiv:quant-ph/0208112, 2002. [p. 4]

[Gro96] Lov K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing - STOC 1996*, pages 212–219, 1996. quant-ph/9605043. doi:10.1145/237814.237866. [p. 3]

[JSV04] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries. *Journal of the ACM*, 51(4):671–697, 2004. Earlier version in STOC'2001. doi:10.1145/1008731.1008738. [p. 4]

[KLL+17] Shelby Kimmel, Cedric Yen-Yu Lin, Guang Hao Low, Maris Ozols, and Theodore J. Yoder. Hamiltonian simulation with optimal sample complexity. *npj Quantum Information*, 3(1), 2017. doi:10.1038/s41534-017-0013-7. [p. 4]

[KŠdW07] Hartmut Klauck, Robert Špalek, and Ronald de Wolf. Quantum and classical strong direct product theorems and optimal time-space tradeoffs. *SIAM Journal on Computing*, 36(5):1472–1493, 2007. Earlier version in FOCS'2004. quant-ph/0402123. doi:10.1137/05063235X. [p. 5]

[Kup15] Greg Kuperberg. How hard is it to approximate the Jones polynomial? *Theory of Computing*, 11(6):183–219, 2015. doi:10.4086/toc.2015.v011a006. [p. 10]

[LMR14] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost. Quantum principal component analysis. *Nature Physics*, 10(9):631–633, 2014. arXiv:1307.0401. doi:10.1038/nphys3029. [p. 4]

[Mar90] Andrei Andreyevich Markov. On a question by di mendeleev. *Zapiski Imperatorskoi Akademii Nauk*, 62:1–24, 1890. [p. 11]

[MP88] Marvin Minsky and Seymour A. Papert. *Perceptrons (2nd edition)*. MIT Press, 1988. First appeared in 1968. [pp. 9, 13, 15]

[MV99] Peter Bro Miltersen and N. V. Vinodchandran. Derandomizing Arthur-Merlin games using hitting sets. In *40th Annual Symposium on Foundations of Computer Science*, pages 71–80, October 1999. doi:10.1109/SFFCS.1999.814579. [p. 10]

[MW05] Chris Marriott and John Watrous. Quantum Arthur–Merlin games. *computational complexity*, 14(2):122–152, June 2005. doi:10.1007/s00037-005-0194-x. [pp. 8, 16]

[N+64] Donald J Newman et al. Rational approximation to $|x|$. *The Michigan Mathematical Journal*, 11(1):11–14, 1964. doi:10.1307/mmj/1028999029. [p. 6]

[Pat92] Ramamohan Paturi. On the degree of polynomials that approximate symmetric Boolean functions. In *Proceedings of the twenty-fourth annual ACM symposium on Theory of computing - STOC 1992*, pages 468–474, 1992. doi:10.1145/129712.129758. [p. 11]

[RC66] T. J. Rivlin and E. W. Cheney. A comparison of uniform approximations on an interval and a finite subset thereof. *SIAM Journal on Numerical Analysis*, 3(2):311–320, 1966. doi:10.1137/0703024. [p. 12]

[SJ89] Alistair Sinclair and Mark Jerrum. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation*, 82(1):93–133, 1989. doi:10.1016/0890-5401(89)90067-9. [pp. 3, 4]

[Sto85] Larry Stockmeyer. On approximation algorithms for #P. *SIAM Journal on Computing*, 14(4):849–861, 1985. doi:10.1137/0214060. [p. 3]

[STT12] Rocco Servedio, Li-Yang Tan, and Justin Thaler. Attribute-efficient learning andweight-degree tradeoffs for polynomial threshold functions. In *Proceedings of the 25th Annual Conference on Learning Theory*, volume 23 of *Proceedings of Machine Learning Research*, pages 14.1–14.19, 2012. URL: http://proceedings.mlr.press/v23/servedio12.html. [pp. 9, 13, 14]

[Tha16] Justin Thaler. Lower Bounds for the Approximate Degree of Block-Composed Functions. In *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*, volume 55 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 17:1–17:15. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016. doi:10.4230/LIPIcs.ICALP.2016.17. [pp. 8, 26]

[Ver92] Nikolai K. Vereshchagin. On the power of PP. In *Proceedings of the Seventh Annual Structure in Complexity Theory Conference*, pages 138–143, 1992. doi:10.1109/SCT.1992.215389. [p. 10]

[Zha12] Mark Zhandry. How to construct quantum random functions. In *Proceedings of the 2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, FOCS '12, pages 679–687. IEEE, 2012. doi:10.1109/FOCS.2012.37. [pp. 6, 18, 25, 29]

[Zha17] Mark Zhandry. Quantum lightning never strikes the same state twice. arXiv:1711.02276, 2017. [p. 35]

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il