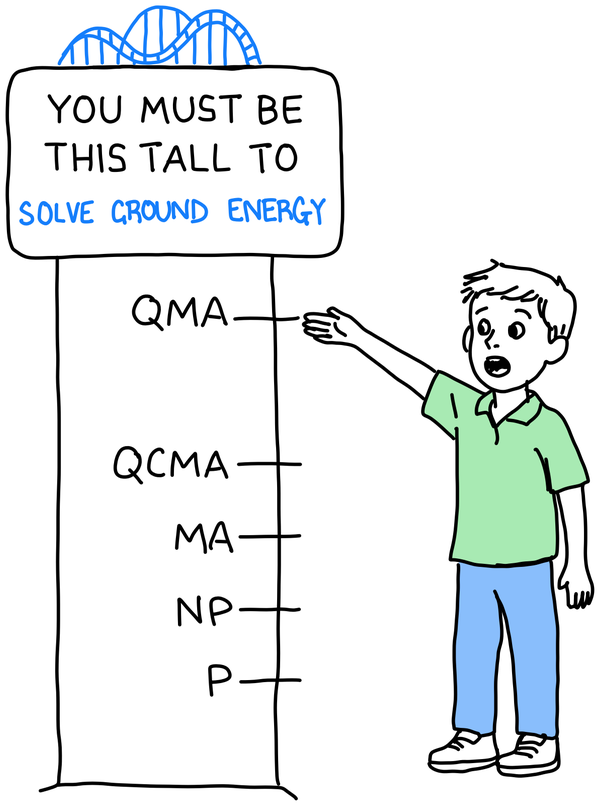
How to prove query lower bounds for QMA or QCMA

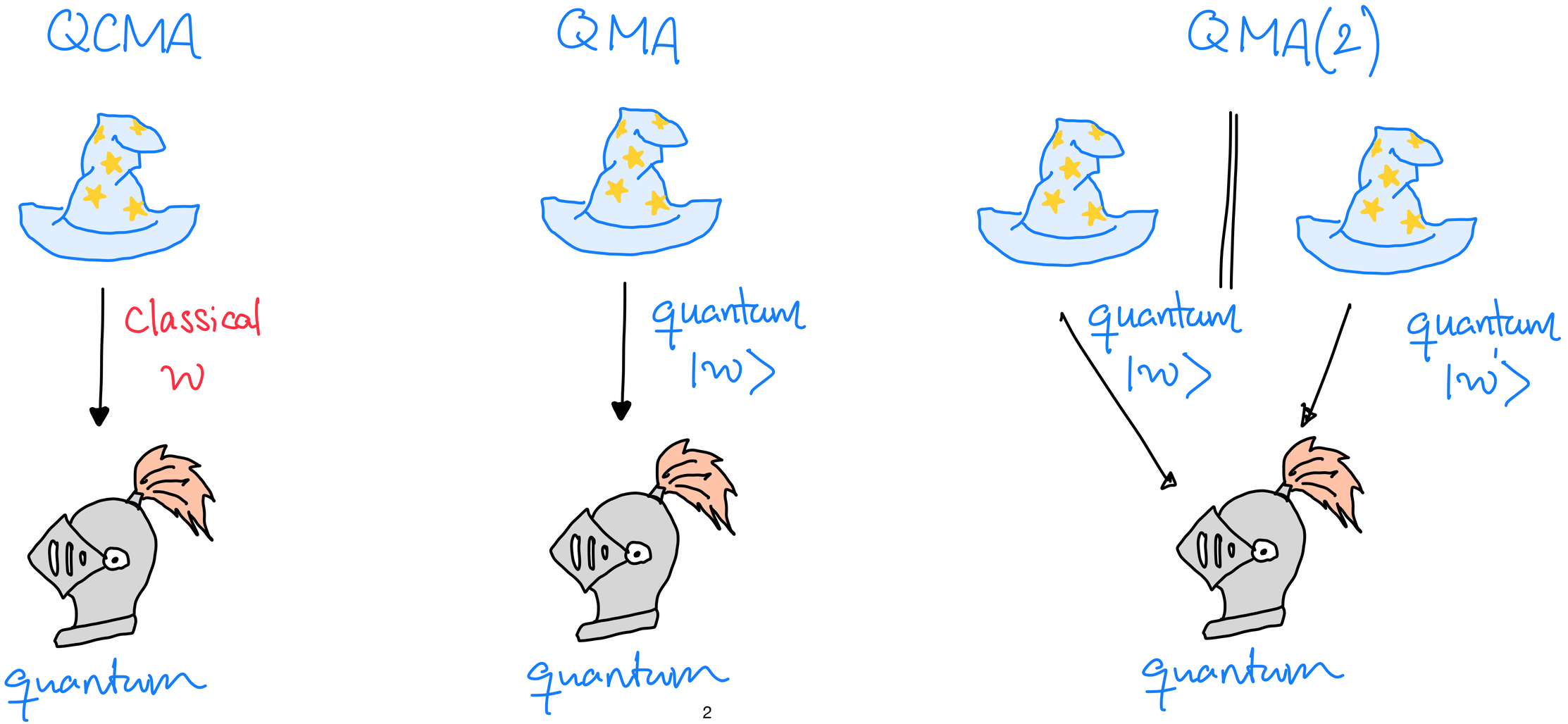
Based on results in

Bostanci, Haferkamp, Nirkhe, Zhandry (2025),
Bostanci, Grewal, Haferkamp, Huang, Hwang, Natarajan, Nirkhe (2026).



Chinmay Nirkhe | University of Washington

The multiple faces of quantum NP



Separating the many faces of quantum NP

- One of the goals of quantum complexity theory is to prove query complexity aka oracle separations between the various faces of quantum **NP**
- Containment by simulation is often easy but it has been relatively unclear whether the containment was strict or not
- Oracle complexity separations give first evidence for the pure complexity classes in a model where unconditional results are easier to come by
- Recently, we proved a *classical* oracle separation between **QMA** and **QCMA**
- Recently, we prove a *quantum* oracle separation between **QMA(2)** and **QMA**

How do we prove a problem is not in QCMA or QMA?

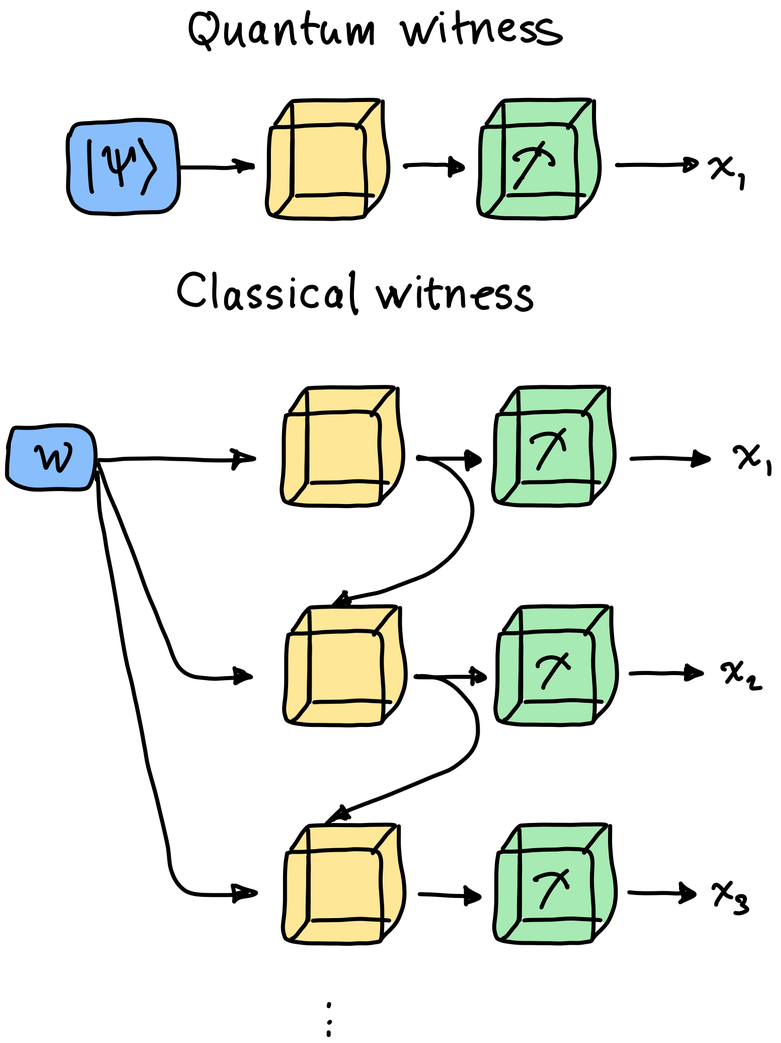
- The goal of today is not to give you proofs of the latest results which are oracle separations between complexity classes.
- In fact, I'll only briefly say what the separating oracle problems are!
- Instead, I'll try to teach you about how we got started thinking about proving query complexity lower bounds for **QCMA** and **QMA**
  - I'll introduce a meta-strategy which I think is powerful and warrants further exploration
  - My hope is that this will lead to new or simpler proofs
  - Many open questions exists in query complexity — particularly cryptography!

Separating complexity classes $\mathscr{X}$ and $\mathscr{Y}$

- Suppose we know of two complexity classes with $\mathscr{X} \subseteq \mathscr{Y}$
- Identify a “differentiating” property of $\mathscr{X}$ that strictly does not hold for $\mathscr{Y}$
  - Ex. For **QCMA** $\subseteq$ **QMA**, we can clone the (classical) witnesses for **QCMA** but we can’t (generically) clone the witnesses for **QMA**
  - Ex. For **QMA** $\subseteq$ **QMA(2)**, both the “guess-the-witness” trick and in-place amplification are valid **QMA** transforms that don’t hold for **QMA(2)**
- Assume that a $\mathscr{X}$ protocol exists; exploit the differentiating property to prove a contradiction. The same contradiction argument won’t hold for $\mathscr{Y}$.

Proving a QCMA query complexity lower bound

“Differentiating” property for QCMA and QMA



- Let V describe the behavior of the verifier in either a **QCMA** or **QMA** protocol
- For **QMA**, given a witness $|\psi\rangle$, you can run $V|\psi\rangle$ and measure to get one classical answer x_1. But this “destroys” $|\psi\rangle$ and you can’t recover it.
- For **QCMA**, since the witness w is classical, preemptively make many copies allowing you to produce many samples.

Today's warmup example: Set size estimation

- Oracle problem: given access to an oracle $\mathcal{O} : \{0,1\}^n \to \{0,1\}$ decide if the number of 1's in the function (i.e. Hamming weight) is
  - (Yes) greater than $2^{n/2}$
  - (No) less than $2^{n/3}$.
- **Theorem:** Any (query) **QCMA** algorithm for Set Size Estimation with t queries and m length witnesses requires either

  (a) $t \geq \Omega(2^{n/4})$ or (b) $m \geq 2^{n/3} \cdot (n - 4\log t)$

A word on oracle queries

- Quantum algorithms are allowed to access queryable functions in *superposition*
- In this case, let $S \subseteq \{0,1\}^n$ be the subset where the oracle is 1.
- For example, query access to $\mathcal{O}$ can be described as the linear extension of

  $$|x\rangle|y\rangle \mapsto |x\rangle|y\rangle \oplus \mathcal{O}_S(x)\rangle$$

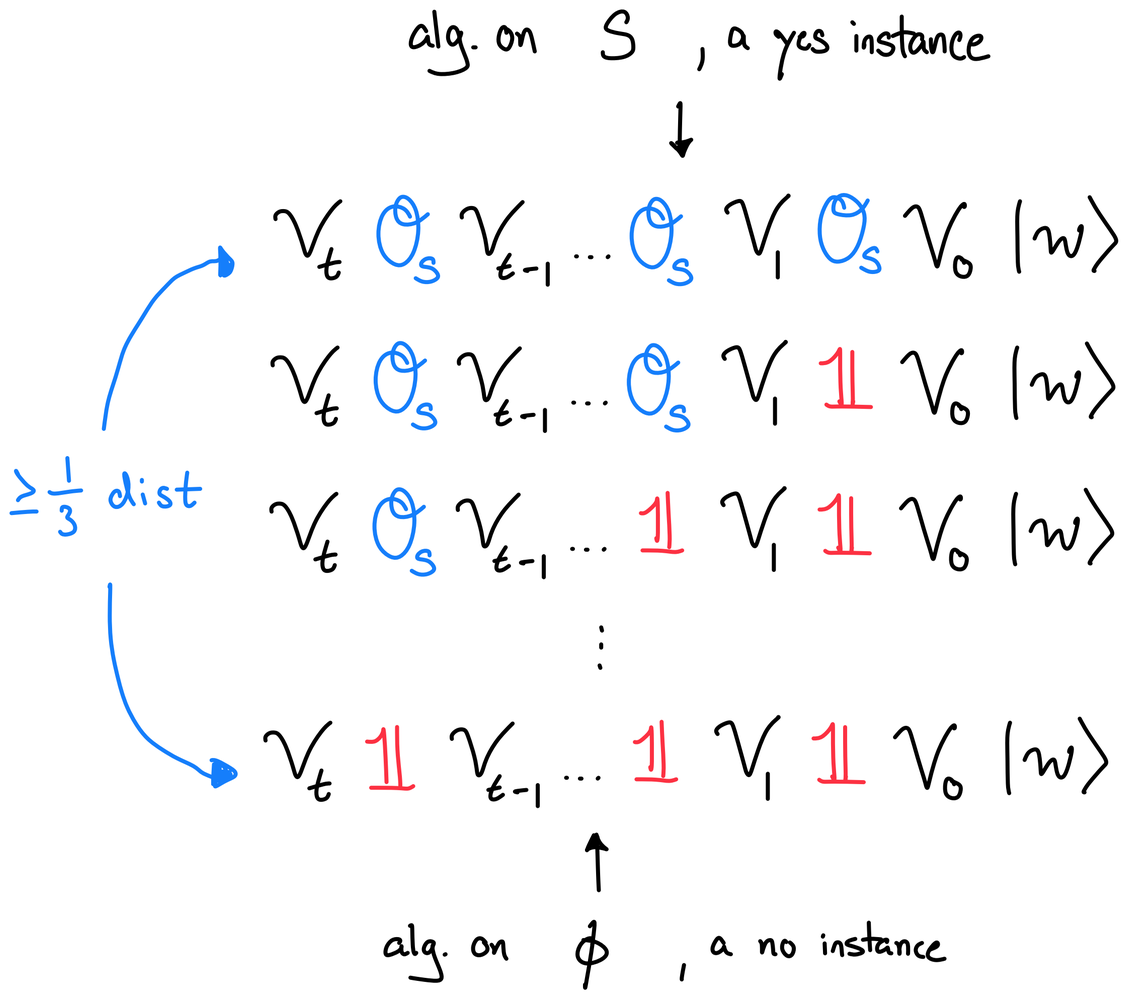
- This is equivalent (up to Hadamard transform) to $|x\rangle|b\rangle \mapsto (-1)^{b\cdot\mathcal{O}(x)}|x\rangle|b\rangle$
  - For simplicity, I'll ignore the control bit b and just write $|x\rangle \mapsto (-1)^{\mathcal{O}(x)}|x\rangle$.
  - In which case unitary $\mathcal{O}_S = \mathbf{1} - 2 \cdot [\text{Projector onto } S]$.

Samplers from QCMA + yes instance

- Assume there exists a **QCMA** algorithm with length m witness and makes t queries to the oracle S.
- Assume $|S| \geq 2^{n/2}$ (yes instance) and then for all $\Delta \subset S$, $|\Delta| \leq 2^{n/3}$ is a no instance.
- Write the algorithm as alternation

  $$V_t \mathcal{O}_S V_{t-1} \mathcal{O}_S \dots V_1 \mathcal{O}_S V_0 |w\rangle$$

  where w is optimal witness for S.



Samplers from QCMA + yes instance

- Assume there exists a **QCMA** algorithm with length m witness and makes t queries to the oracle S.
- Assume $|S| \geq 2^{n/2}$ (yes instance) and then for all $\Delta \subset S$, $|\Delta| \leq 2^{n/3}$ is a no instance.
- Write the algorithm as alternation

  $$V_t \mathcal{O}_S V_{t-1} \mathcal{O}_S \cdots V_1 \mathcal{O}_S V_0 |w\rangle$$

  where w is optimal witness for S.

For some $j \in \{1, \ldots, t\}$, the distance between

$$V_t \mathcal{O}_S V_{t-1} \cdots \mathcal{O}_S V_j \mathbb{1} V_{j-1} \cdots V_1 \mathbb{1} V_0 |w\rangle$$

and $$V_t \mathcal{O}_S V_{t-1} \cdots \mathbb{1} V_j \mathbb{1} V_{j-1} \cdots V_1 \mathbb{1} V_0 |w\rangle$$

is at least $\frac{1}{3t}$.

$$\Rightarrow \left\| (\mathcal{O}_S - \mathbb{1}) V_j \mathbb{1} V_{j-1} \cdots V_1 \mathbb{1} V_0 |w\rangle \right\| \geq \frac{1}{3t}$$

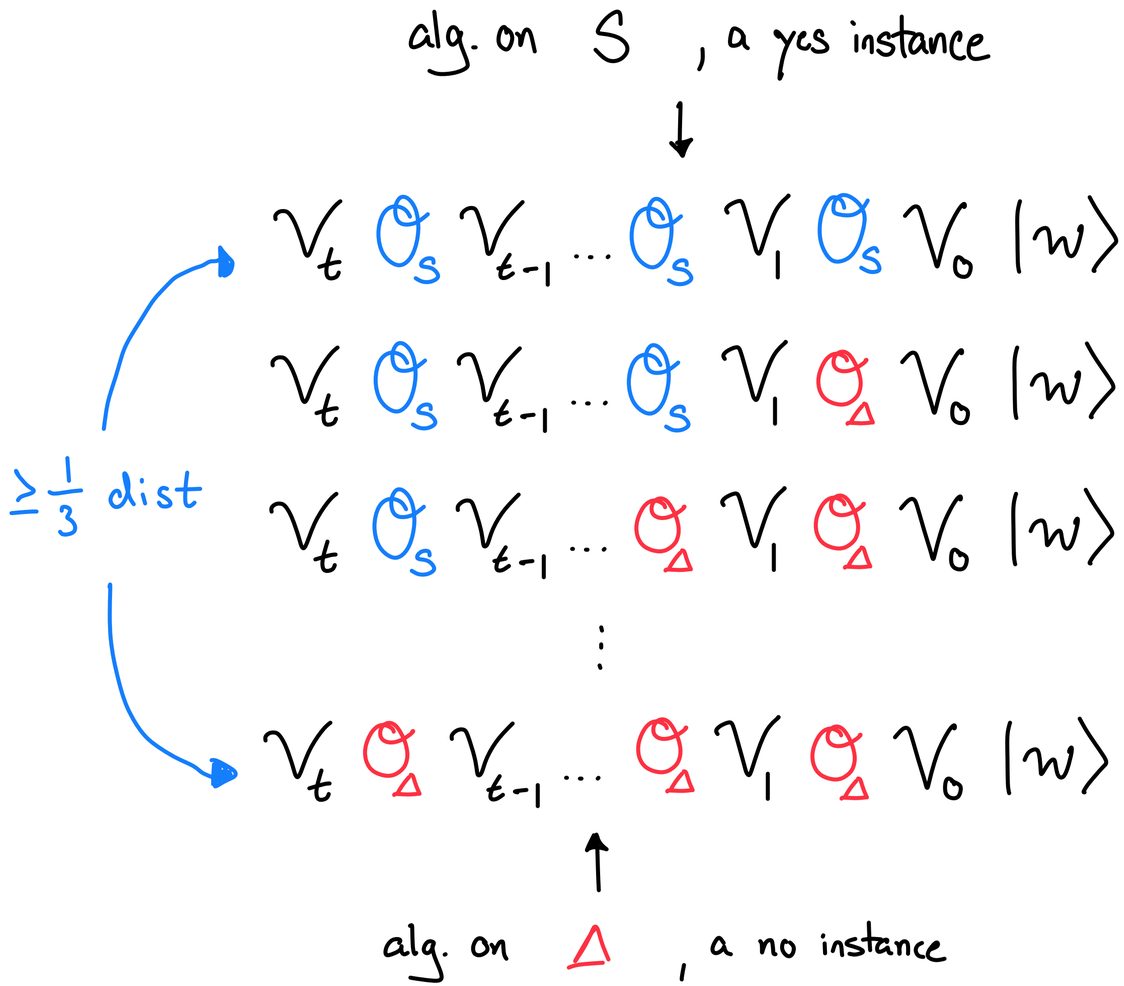
$$\left\| \Pi_S V_j \mathbb{1} V_{j-1} \cdots V_1 \mathbb{1} V_0 |w\rangle \right\| \geq \frac{1}{6t}.$$

Samplers from QCMA + yes instance

- This yields a sampling algorithm which makes no queries and produces a sample from S with probability $\geq 1/(36t^2)$:
- **Basic Sampler**(w)**:**
  - Pick a random $j \in \{1,\ldots,t\}$.
  - Run $V_j V_{j-1} \ldots V_1 V_0 \,|\, w\rangle$ and measure query register.
- This constructs a point $x_1 \in S$. How do we find a point $x_2 \in S \backslash \{x_1\}$?

Samplers from QCMA + yes instance

- Assume there exists a **QCMA** algorithm with length m witness and makes t queries to the oracle S.
- Assume $|S| \geq 2^{n/2}$ (yes instance) and then for all $\Delta \subset S$, $|\Delta| \leq 2^{n/3}$ is a no instance.



Here, $O_\Delta = \sum_{b,x} (-1)^{b \cdot (x \in \Delta)} |b,x\rangle\langle b,x|$.

This is controlled-membership access to Δ.

For example, $\Delta = \{x_1\}$.

Samplers from QCMA + yes instance

- For any $\Delta \subset S, |\Delta| < 2^{n/3}$, this yields a sampling algorithm which makes no queries and produces a sample from $S \setminus \Delta$ with probability $\geq 1/(36t^2)$.
- Let $\mathcal{O}_\Delta = \mathrm{Id} - 2\sum_{x \in \Delta} |x\rangle\langle x|$.
- **Basic Sampler**(Δ, w)**:**
  - Pick a random $j \in \{1, \ldots, t\}$.
  - Run $V_j \mathcal{O}_\Delta V_{j-1} \mathcal{O}_\Delta \ldots \mathcal{O}_\Delta V_1 \mathcal{O}_\Delta V_0 |w\rangle$ and measure query register.

Samplers from QCMA + yes instance

- We built a sampler which generates multiple samples from S with decent probability when S is a yes instance.
- **Cumulative Sampler**(w)**:**
  - Start with $\Delta = \emptyset$.
  - For $i = 1,\ldots, v$
    - Run **Basic Sampler**(Δ, w) and produce output x_i.
    - If $x_i \in \Delta$, abort! Else, update $\Delta \leftarrow \Delta \cup \{x_i\}$.
  - Output $\Delta = \{x_1, \ldots, x_v\}$

Each sample x_i is $\in S$ with probability $\geq 1/(36t^2)$ conditioned on $x_1, \ldots, x_{i-1} \in S$.

Probability that all v samples are $\in S$ is at least $\geq (1/36t^2)^v$.

Samplers from QCMA + yes instance

- We built a sampler which generates multiple samples from S with decent probability (without any queries!) when S is a yes instance.
- **Cumulative Sampler:**
  - Uniformly randomly sample $w \in \{0,1\}^m$.
  - Start with $\Delta = \emptyset$.
  - For $i = 1, \ldots, v$
    - Run **Basic Sampler**(Δ, w) and produce output x_i.
    - If $x_i \in \Delta$, abort! Else, update $\Delta \leftarrow \Delta \cup \{x_i\}$.
  - Output $\Delta = \{x_1, \ldots, x_v\}$

Each sample x_i is $\in S$ with probability $\geq 1/(36t^2)$ conditioned on $x_1, \ldots, x_{i-1} \in S$ and w is a good witness for S.

Probability that all v samples are $\in S$ is at least $\geq 2^{-m}(1/36t^2)^v$.

Trivial upper bounds

- As long as $v \leq 2^{n/3}$, we have a sampler that makes no queries and yet produces v samples from S with probability $\geq 2^{-m}(1/36t^2)^v$.
- Simultaneously, the max guessing probability is $\leq (2^{-n/2})^v$ as **no queries**!
- Solving this inequality:

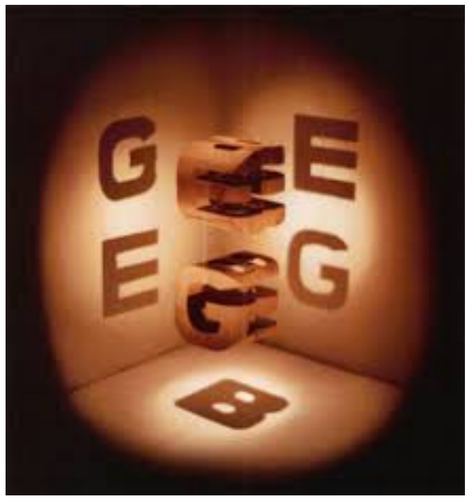
  Either $t \geq \frac{1}{6}2^{n/4}$ OR $m \geq \frac{1}{2}2^{n/3} \cdot (n - 4\log 6t)$

What have we learned?

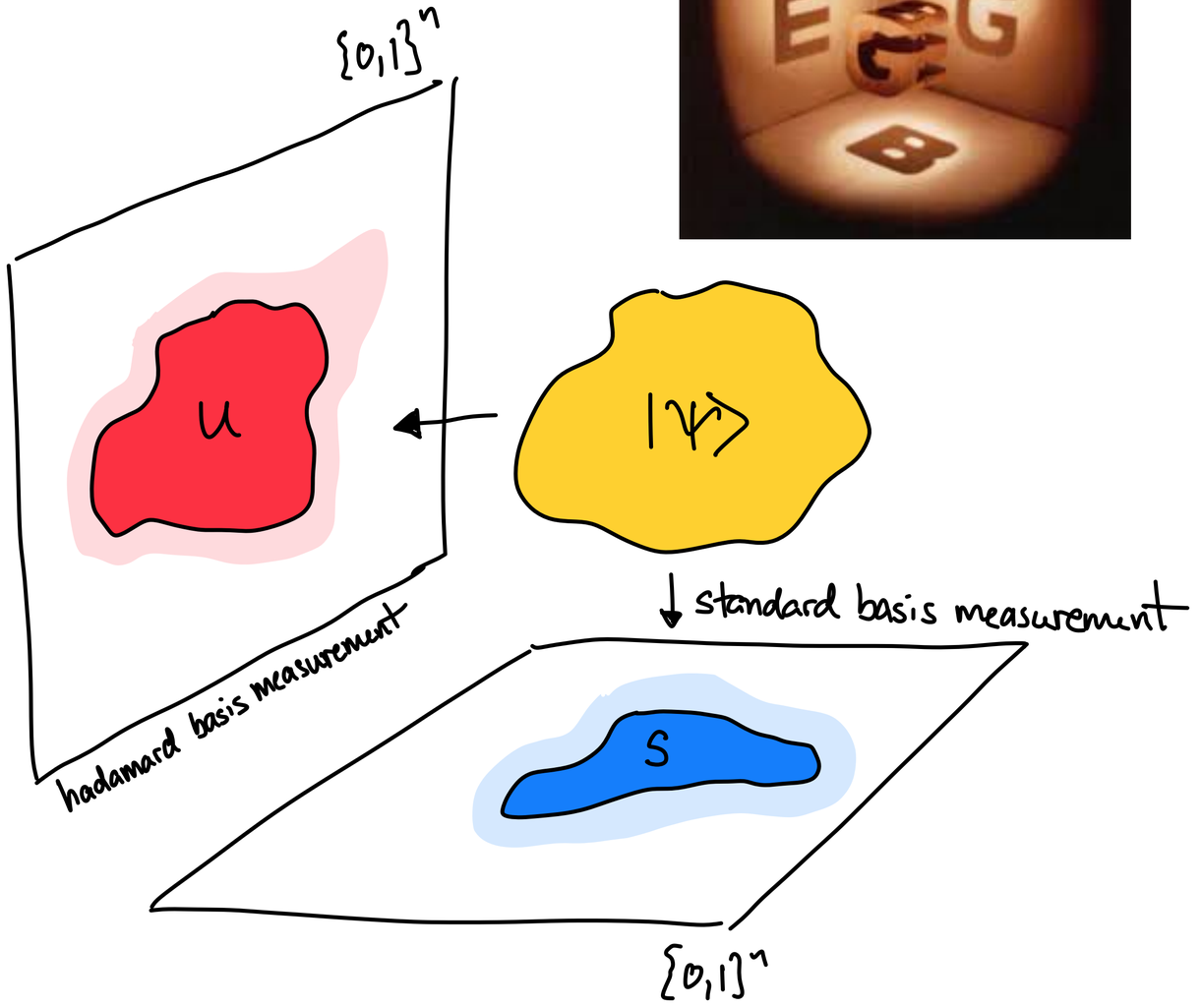
- Identify a property (Cloning) ⟶ The property must be too good to be true
- Technically, we proved that Set Size Estimation is outside **ClonableQMA**
- This idea of generating a sampler was the basis for the **QMA** vs **QCMA** separation
  - More is needed to construct a problem that is in **QMA** (as Set Size Estimation isn't in **QMA**, either)
  - This in turn makes proving the sampling upper bound harder (but doable!)

[Skip link]

Separating problem: Spectral Forrelation

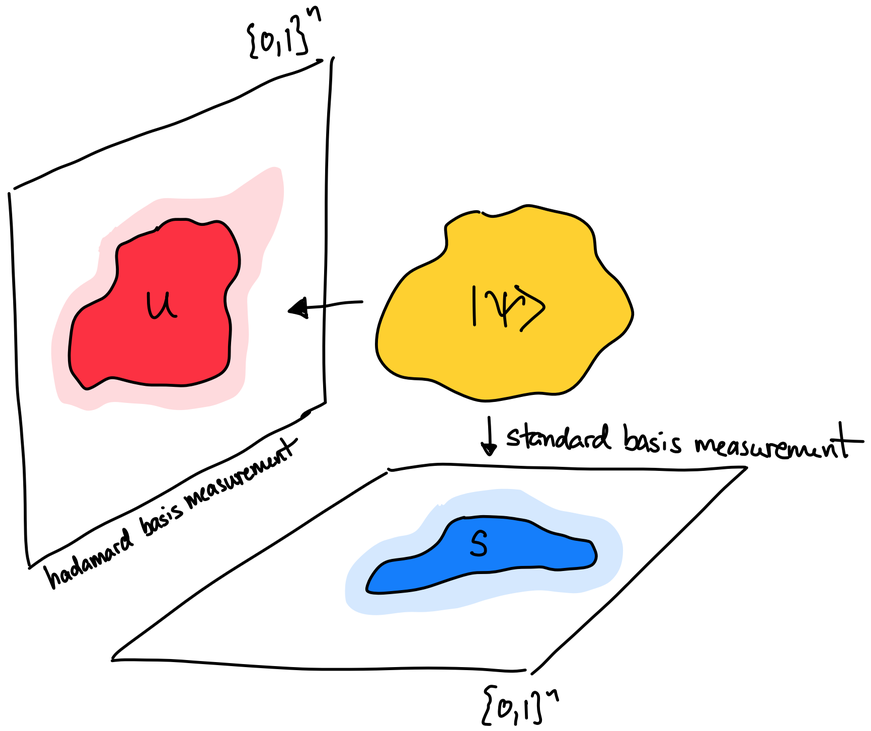


- Subsets $S, U \subseteq \{0,1\}^n$ are α-spectrally Forrelated iff there exists a state $|\psi\rangle$ such that $\|\Pi_U H^{\otimes n} \Pi_S |\psi\rangle\|^2 = \alpha$.
- Morally, if $|\psi\rangle$ is
  - Measured in the standard basis, the output is in S w.h.p. and
  - Measured in the Hadamard basis, the output is in U w.h.p.



Spectral Forrelation is in QMA

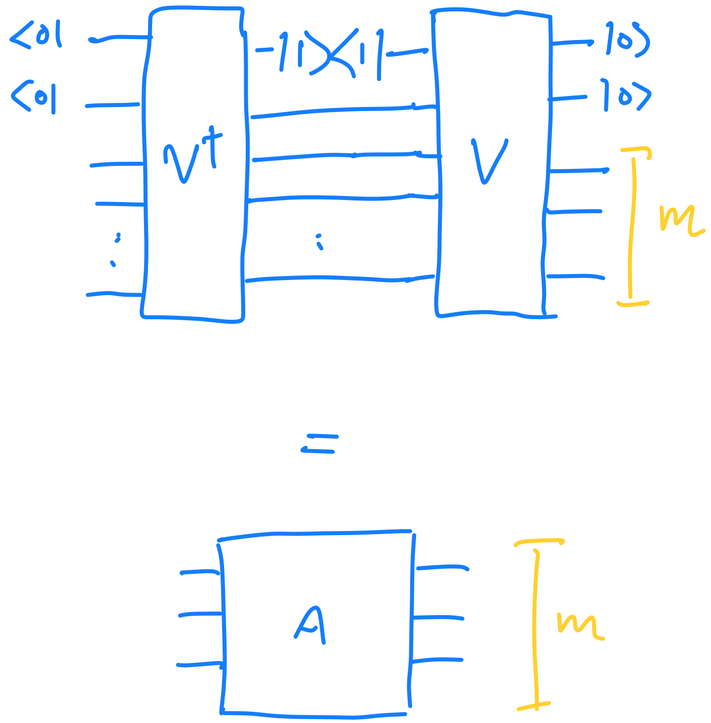
- **Computational task** (α, β)**:**
  - Given oracle access to $S, U \subseteq \{0,1\}^n$
  - Decide if the pair are (yes instance) $\geq \alpha$-spectrally Forrelated
  - Or decide if the pair are (no instance) $\leq \beta$-spectrally Forrelated
- **This computation task is in QMA:**
  - Merlin sends over quantum state $|\psi\rangle$
  - Arthur uses S oracle to measures if state is supported on S. Rejects if not.
  - Else, applies $H^{\otimes n}$. Arthur uses U oracle to measures if state is supported on U. Rejects if not.
  - Accepts otherwise. Acceptance probability = $\|\Pi_U H^{\otimes n} \Pi_S |\psi\rangle\|^2$.



Proving a QMA query complexity lower bound

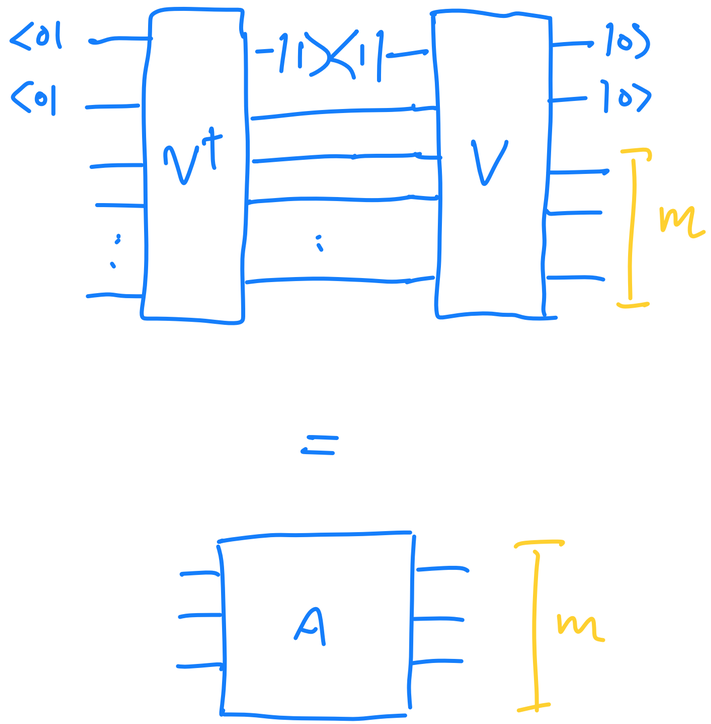
“Differentiating” property for QMA and QMA(2)

- Let V be the description of a verifier with the input hardcoded (completeness/soundness of $2/3, 1/3$)
- Then, the max acceptance probability is $\|A\|_\infty$
- **QMA** in-place amplification exploits this expression to create a new verifier V' with completeness/soundness $1 - 2^{-2m}, 2^{-2m}$
- New verifier simulates V a total of $O(m)$ times
- In-place amplification is not known for **QMA(2)**



“Differentiating” property for QMA and QMA(2)

- V' with completeness/soundness$1 - 2^{-2m}, 2^{-2m}$
- Run V' on a random witness —i.e., $2^{-m} \cdot I$
- This protocol accepts
  - with prob. $\geq 2^{-m+1}$ if V is a yes instance
  - with prob. $\leq 2^{-2m}$ if V is a no instance
- Exponential multiplicative gap in success prob.!

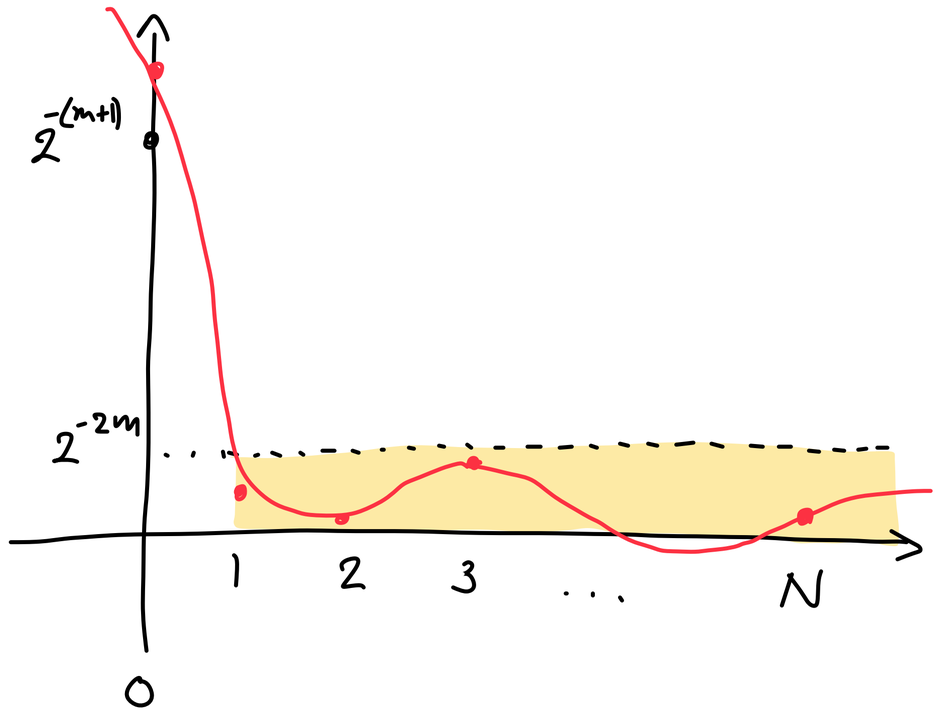


Query complexity problem not in QMA

- The IS-ZERO problem is whether a function $\mathcal{O} : \{0,1\}^n \to \{0,1\}$ is
  - if $\mathcal{O} \equiv 0$ everywhere (yes instance)
  - if $\mathcal{O} \not\equiv 0$ (no instance) or
- It is the negation of the OR problem (which is in **QMA**)
- **Theorem:** Any (query) **QMA** algorithm with t queries and m-qubit witness for IS-ZERO requires $t^2 m \geq \Omega(2^n)$.

The only polynomial degree bound you need to know

- Let $g : \mathbb{R} \to \mathbb{R}$ be a polynomial such that
  - $g(0) \geq 2^{-(m+1)}$
  - $g(x) \in [0, 2^{-2m}]$ for all $x = 1, \ldots, N$.
- Then, $\deg g \geq \Omega\left(\sqrt{Nm}\right)$.



Putting the ideas side-by-side

- If V solves IS-ZERO with t queries and m-qubit witness then there exists a $O(tm)$ query protocol W with NO witness such that
  - If $\mathcal{O} \equiv 0$, then W accepts with prob. $\geq 2^{-(m+1)}$
  - If $\mathcal{O} \not\equiv 0$, then W accepts with prob. $\leq 2^{-2m}$
- **Fact** (Beals et. al.) the new protocol can be expressed as a degree $O(tm)$ polynomial in the entries of $\mathcal{O}$ (when viewed as a truth table)
- Observe that the IS-ZERO problem is symmetric! In fact, whether $\mathcal{O}$ is non-zero only depends on the Hamming weight of $\mathcal{O}$'s truth table
- Let $g(x) := \mathop{\mathbb{E}}_{\mathcal{O}:|\mathcal{O}|=x} W^{\mathcal{O}}$. Then observe that g is also a degree $O(tm)$ polynomial but in one variable!

Putting the ideas side-by-side

- If V solves IS-ZERO with t queries and m-qubit witness then there exists a $O(tm)$ query protocol W with NO witness such that
  - If $\mathcal{O} \equiv 0$, then W accepts with prob. $\geq 2^{-(m+1)}$
  - If $\mathcal{O} \not\equiv 0$, then W accepts with prob. $\leq 2^{-2m}$
- Let $g(x) := \mathop{\mathbb{E}}_{\mathcal{O}:|\mathcal{O}|=x} W^{\mathcal{O}}$. Then observe that g is also a degree $O(tm)$ polynomial but in one variable!
  - If $x = 0$, then $g(x) \geq 2^{-(m+1)}$.
  - If $x \in \{1,\ldots,2^n\}$, then $g(x) \in [0, 2^{-2n}]$.
- $\implies \Omega\left(\sqrt{2^n m}\right) \leq \deg g \leq O(tm) \implies \Omega(2^n) \leq t^2 m.$

polynomial deg

query assumption

What have we learned?

- Identify a property (amplification) ⟶ The property must be too good to be true
- Technically, we proved that IS-ZERO is outside **SBQP**, a variation of **BQP** where completeness and soundness are multiplicatively separated
- This idea of using the polynomial method was the basis for the **QMA(2)** vs **QMA** separation
  - We don't believe IS-ZERO is in **QMA(2)**. For the actual separating quantum oracle,
  - More work is needed to construct a univariate polynomial
  - The original polynomial from the query algorithm is in the entries of the unitary. This makes symmetrizing the polynomial more complicated (but doable!)

[Skip link]

Separating problem: Entangled subspace problem

- Canonical problem: Let Π be a projector acting on Hilbert space $\mathbb{C}^{2^n} \otimes \mathbb{C}^{2^n}$.
  - Yes instance: $\exists\ |\psi_1\rangle, |\psi_2\rangle$ s.t. $\left\| \ |\Pi|\psi_1\rangle \otimes |\psi_2\rangle \ \right\| = 1,$
  - No instance: $\forall\ |\psi_1\rangle, |\psi_2\rangle$, $\left\| \ |\Pi|\psi_1\rangle \otimes |\psi_2\rangle \ \right\| \leq 1/3.$
- For any Π that is classically describable, this problem is in QMA(2).
- If $\Pi = \dfrac{I+U}{2}$ for a reflection unitary U, we can define a (quantum query) **QMA(2)** problem where complexity is measured in terms of uses of U.

Separating problem: Entangled subspace problem

- Symmetric subspace — spanned by vectors of the form $|\psi\rangle \otimes |\psi\rangle$
- Antisymmetric subspace — the orthogonal subspace to symmetric
- $\Pi_{\pm} = \dfrac{I \pm \mathrm{SWAP}}{2}$ defines the projector onto sym. and anti-sym. subspaces

$$\mathbb{1}_{2^n - d} \otimes \Pi_{\pm}^{(d)} \otimes \mathbb{1}_{2^n - d}$$

embed d-dim subspace into $2n$ qubits.

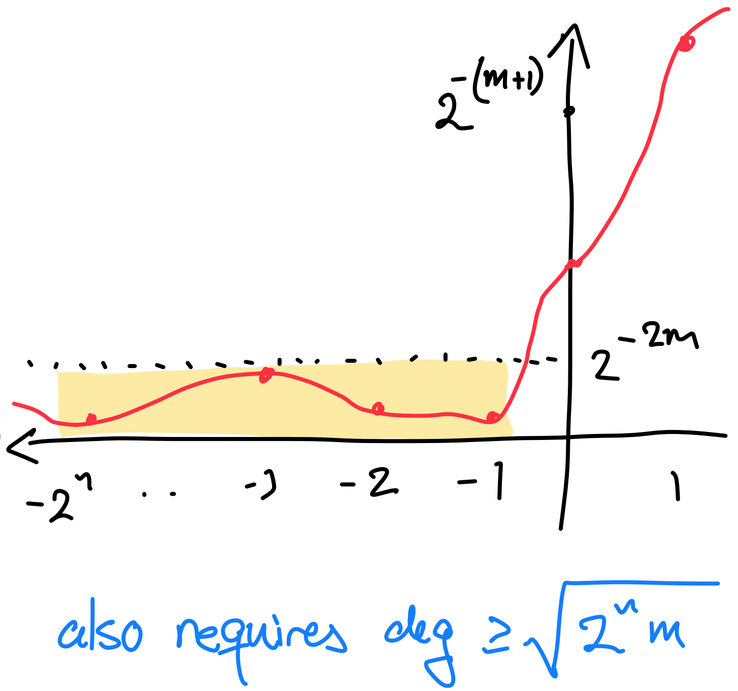
For $\Pi_+^{(d)}$, this is a yes instance

For $\Pi_-^{(d)}$, this is a no instance.

Fact: If Π is a yes instance $\iff$ $(V \otimes W)\Pi(V \otimes W^\dagger)$ is a yes instance.

Separating problem: Entangled subspace problem

- For any $d \geq 1$, and *any* unitaries V, W,
  $$\left(V^\dagger \otimes W^\dagger\right)\left(I_{2^n-d} \otimes \Pi_{\pm}^{(d)} \otimes I_{2^n-d}\right)(V \otimes W)$$
  is + = yes, - = no instance
- For any $d \geq 1$, this gives us a large distribution over yes/no instances by sampling the unitaries V, W Haar randomly
- This let's us symmetrize into a polynomial that only depends on d! Moreover, its the **same** polynomial $g(d)$ for both yes and no instances!
  - For symmetric projectors, the average for dim. d is given by $g(d)$
  - For anti-symm. projectors, the average for dim. d is given by $g(-d)$



also requires deg $\geq \sqrt{2^n m}$

Wrapping up a.k.a. meta strategies

Meta strategies for lower bounds

- Query complexity bounds are easiest to prove for quantum algorithms with no witnesses and bounded number of queries
- Figure out a way to get rid of the witness! Guessing the witness is a good strategy.
- Adversary method, polynomial method, compressed oracle are great ways to prove lower bounds for quantum algorithms with no witnesses.
- Show that your original algorithm implies a “too good to be true” algorithm which has an easy query lower bound.
- Don’t be afraid if your “too good to be true” algorithm isn’t physical or has exponentially small success probabilities! Even though they are small — they could still be much bigger than you should expect!

Lower bounds with corresponding upper bounds

- Both problems we looked at today, Set Size Estimation and IS-ZERO, are “hard” query complexity problems for their respective classes
- I showed their lower bound proofs (in the entirety) because they are easy to explain and prove bounds for
- For the actual query/oracle separations of **QMA** v **QCMA** or **QMA(2)** v **QMA**, we needed to construct problems that were actually in the larger class
  - This requires providing “more structure” which puts the problem in the larger class
  - This new structure gives the smaller class more to work with — but in both cases, we were still able to prove lower bounds
- Hopefully more query complexity results will be proven through this meta-technique!

