Separating QMA from QCMA with a classical oracle

Mark Zhandry (Stanford University *)

Based on join work with John Bostanci, Jonas Haferkamp, Chinmay Nirkhe

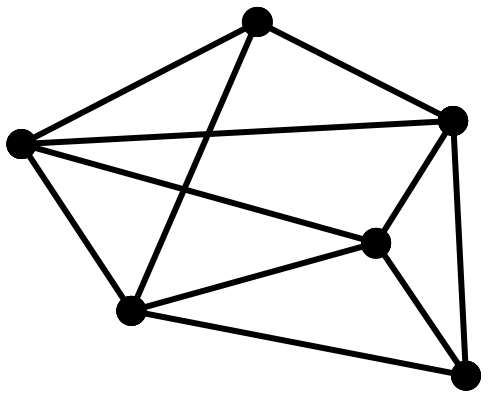
* Project started while I was still at NTT Research

P vs NP

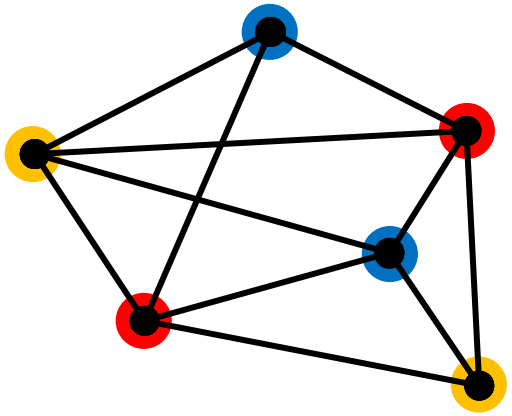
P: problems that can be solved efficiently on a classical computer

NP: decision problems that can be solved efficiently on a classical computer, given a verifiable ***witness***

Is this graph 3-colorable?



Witness:



Why **P** vs **NP**?

(Extended) Church-Turing Thesis: *Anything* computable by nature is captured by (efficient) classical computation

Under ECT:

P are problems we can actually hope to solve at scale

NP are problems whose solutions have a real-world impact at scale

Enter Quantum

(Extended) Church-Turing Thesis: *Anything* computable by nature is captured by (efficient) classical computation

(Extended) Quantum Church-Turing Thesis: *Anything* computable by nature is captured by (efficient) **quantum** computation

What do our favorite complexity classes look like in a quantum world?

The quantum analog of **P** is clear:

BQP: problems that can be solved efficiently on a quantum computer

That is, problems we can actually hope to solve at scale

What is the right quantum analog of **NP**?

Two axes of quantumness:

The verifier: **Real-world impact = efficient quantum verifier**

The witness: Do we care about classical or quantum states?

What is the right quantum analog of **NP**?

QMA: decision problems that can be solved efficiently on a *quantum* computer, given a verifiable *quantum* witness

Does this quantum system have a low-energy ground state?

Witness: the ground state

QCMA: decision problems that can be solved efficiently on a *quantum* computer, given a verifiable *classical* witness

Does this quantum system have a low-energy state that can be prepared by a small quantum circuit?

Witness: the quantum circuit

What is the right quantum analog of **NP**?

QMA: decision problems that can be solved efficiently on a quantum computer, given a verifiable *quantum* witness

Is there a drug that cures cancer

Witness: the drug

QCMA: decision problems that can be solved efficiently on a quantum computer, given a verifiable *classical* witness

Is there a drug that cures cancer, *and* can be efficiently prepared

Witness: the method of preparation

Advantage of QMA: fully quantum, captures almost all problems of interest

Advantage of QCMA: may be what we *really* want in our solutions to problems

Wouldn’t it be great if there was no debate at all?

Does QCMA = QMA?

First proposed by [Aharonov-Naveh'02]

Like other major complexity class questions (P vs NP, BQP vs QMA, etc), pretty much everyone believes they are different, but proving it seems far beyond our capabilities

So how do we nevertheless justify our intuition that **QMA ≠ QCMA**?

Oracle Separations

Provide an oracle **O** relative to which we can formally prove $\mathbf{QMA^O \neq QCMA^O}$

Advantage: separation becomes a query complexity question, which we *may* be able solve without overcoming long-standing barriers

Disadvantage: no longer directly connected to the real world

Despite being disconnected from the real world, oracle separations are extremely popular in both complexity theory and cryptography, because they at least allow us to say *something*

A slight cheat: throughout this talk, I'm going to think of the oracle **O** as the "input", and the quantum algorithm has to learn something about **O** by making quantum queries

By standard complexity-theory arguments, a separation for oracle "inputs" implies a standard oracle separation

A quantum oracle separating QMA from QCMA

[Aaronson-Kuperberg'07]

Let $|\psi\rangle$ be a Haar random state

Flips sign of $|\psi\rangle$, leaves everything orthogonal untouched

Let $U = \mathbf{I} - 2|\psi\rangle\langle\psi|$ or identity

Problem: decide which is the case, given oracle access to U

Clearly in oracle-QMA: witness $|\psi\rangle$

Not in oracle-QCMA:

Distinguisher must query on $|\psi\rangle$ to distinguish two cases

➡ Somehow Haar random $|\psi\rangle$ described by poly-sized classical string ✗

[Aaronson-Kuperberg'07] shows that any lower-bound technique that "relativizes" to unitary oracles cannot separate QMA from QCMA

However, it is often viewed as less-than-satisfactory evidence for a separation

- At the time, classical oracle separations were considered "standard" while quantum oracle separations "non-standard"
- Quantum oracle is an inherently quantum "input" – is it telling us anything about real-world problems where the input is classical?
- Separation arises due to information theory (exponential description size of Haar random states) – is it really telling us anything about computation?

A major goal for >15 years: a **classical** oracle separation

(classical function that can be queried in superposition)

Why?

- Long been considered more “standard”
- “Input” is now a classical object – maybe it’s more reflective of real-world problems
- Individual accepting inputs have low description complexity → QCMA verifier can make non-trivial queries, so any separation must rely on inherently (query) complexity-theoretic aspects
- Classical oracles capture more techniques (e.g. QCMA = perfect QCMA)

[Jordan-Kobayashi-Nagaj-Nishimura’11]

Numerous partial results

[Lutomirski ’11] Candidate, no formal analysis

$$\sum_x \alpha_x |x\rangle \mapsto \sum_x \alpha_x |\Pi(x)\rangle$$

[Fefferman-Kimmel’15]: Relative to “in-place” permutation oracle. If given inverse access, problem becomes in **BQP**

[Natarajan-Nirkhe’22]: Requires witness independent of part of oracle

[Li-Liu-Pelecanos-Yamakawa’23, Ben-David-Kundu’24]: Assuming QCMA verifier makes classical queries or has bounded “adaptivity”

[Z’25]: Under a (refuted) query complexity conjecture

[Liu-Mutreja-Yuen’25]: Assuming AA-like conjecture

Our Work

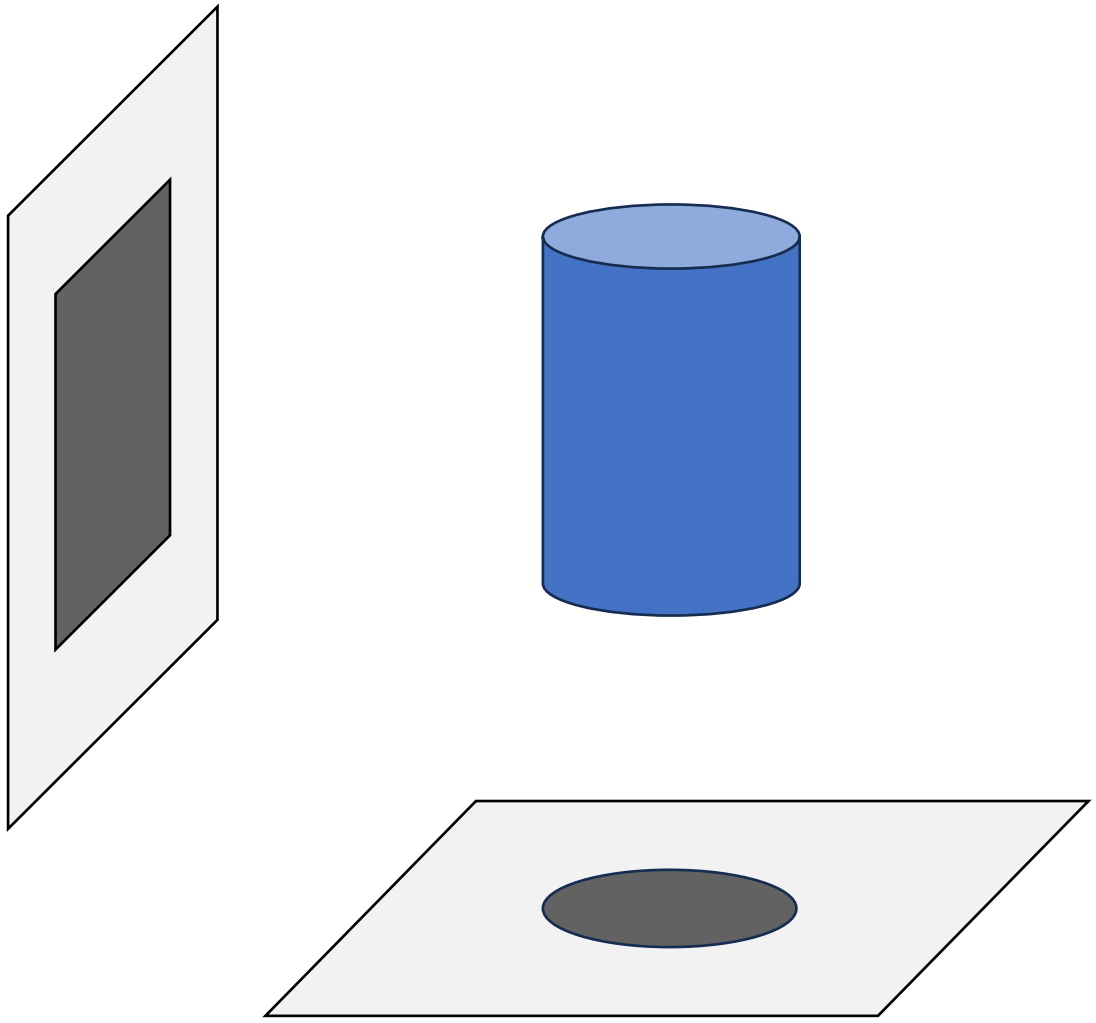
Thm: There exists a classical oracle relative to which

$$\mathbf{QCMA^O \neq QMA^O}$$

Theorem proved in [Bostanci-Haferkamp-Nirkhe-**Z**'25], building on an approach initially developed in [**Z**'24]

Shadows

Hadamard basis

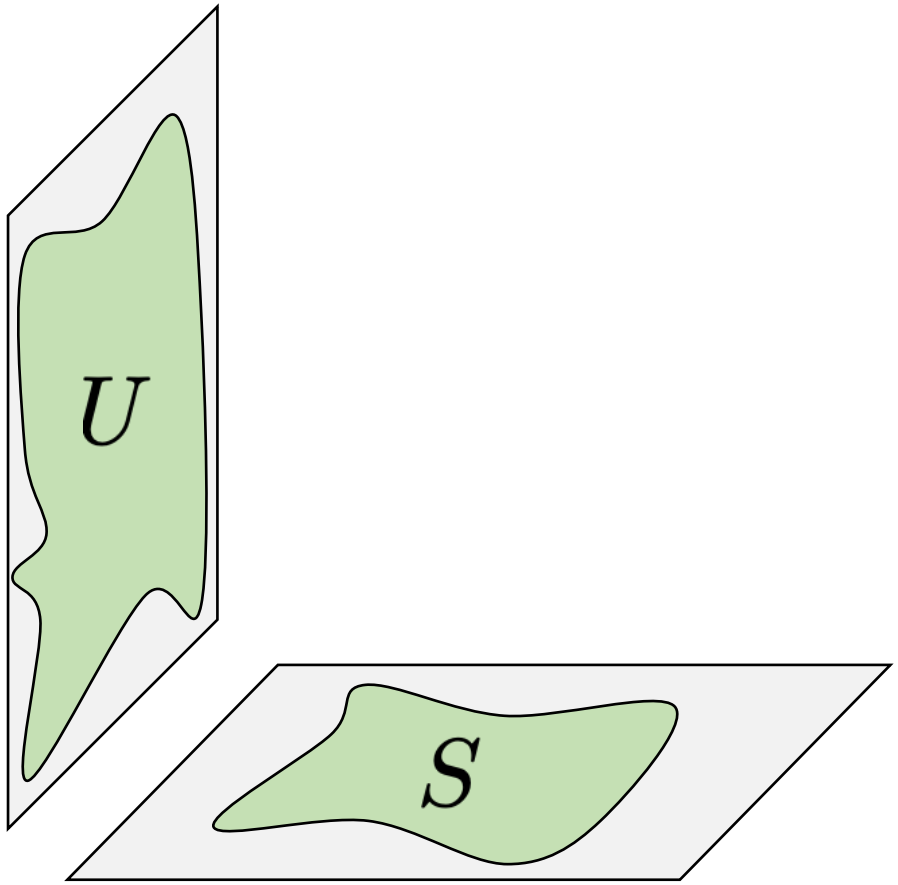


Computational basis

The shadow of a state $|\psi\rangle$ in basis B is the set of "heavy" points when measuring $|\psi\rangle$ in B

Spectral Forrelation

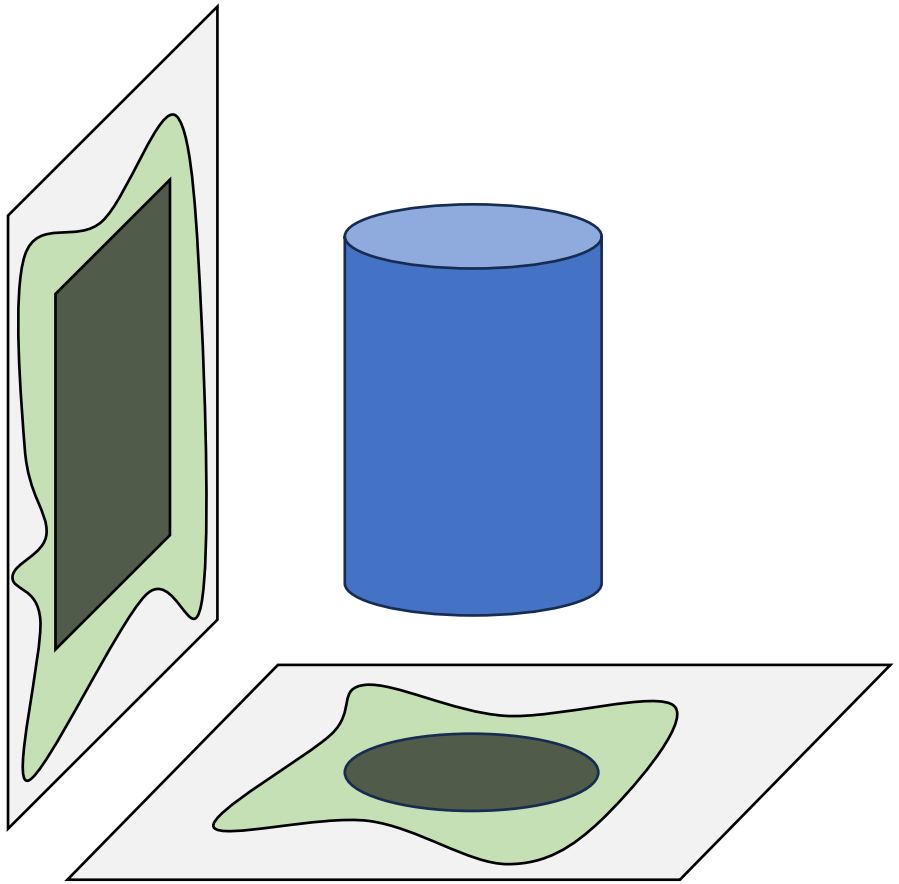
Given two sets S, U (as membership oracles), decide if there is a state whose shadows lie within those sets



Spectral Forrelation

Given two sets S, U (as membership oracles), decide if there is a state whose shadows lie within those sets

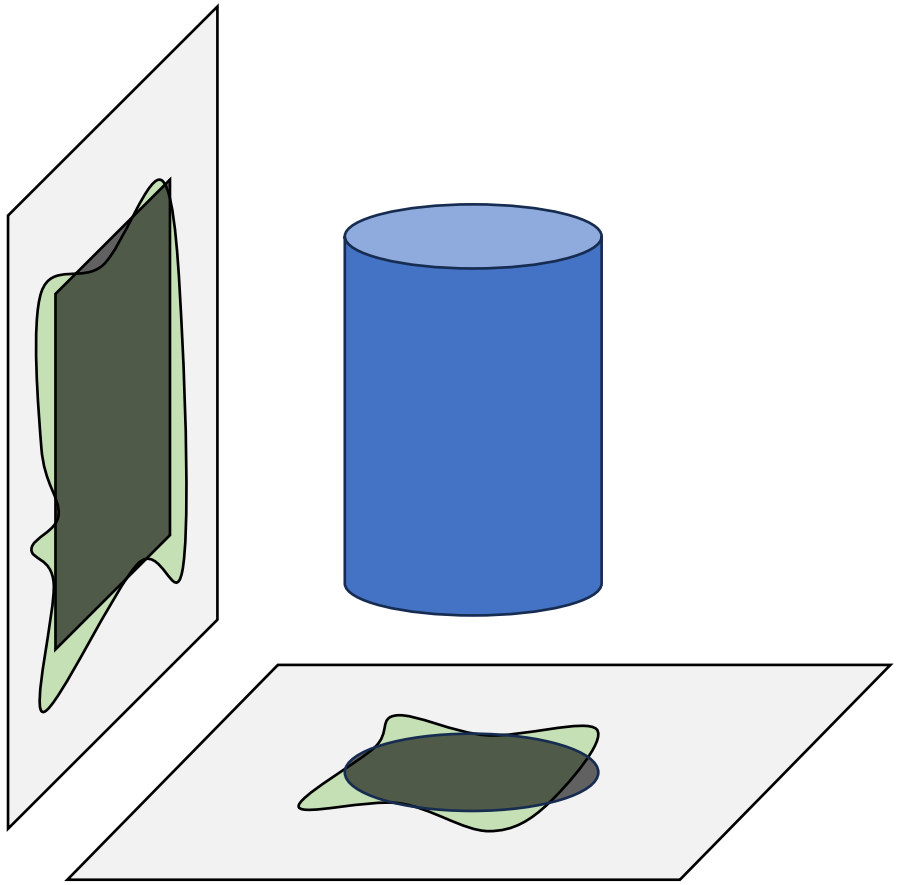
Yes instance:



Spectral Forrelation

Given two sets S, U (as membership oracles), decide if there is a state whose shadows lie within those sets

Yes instance:

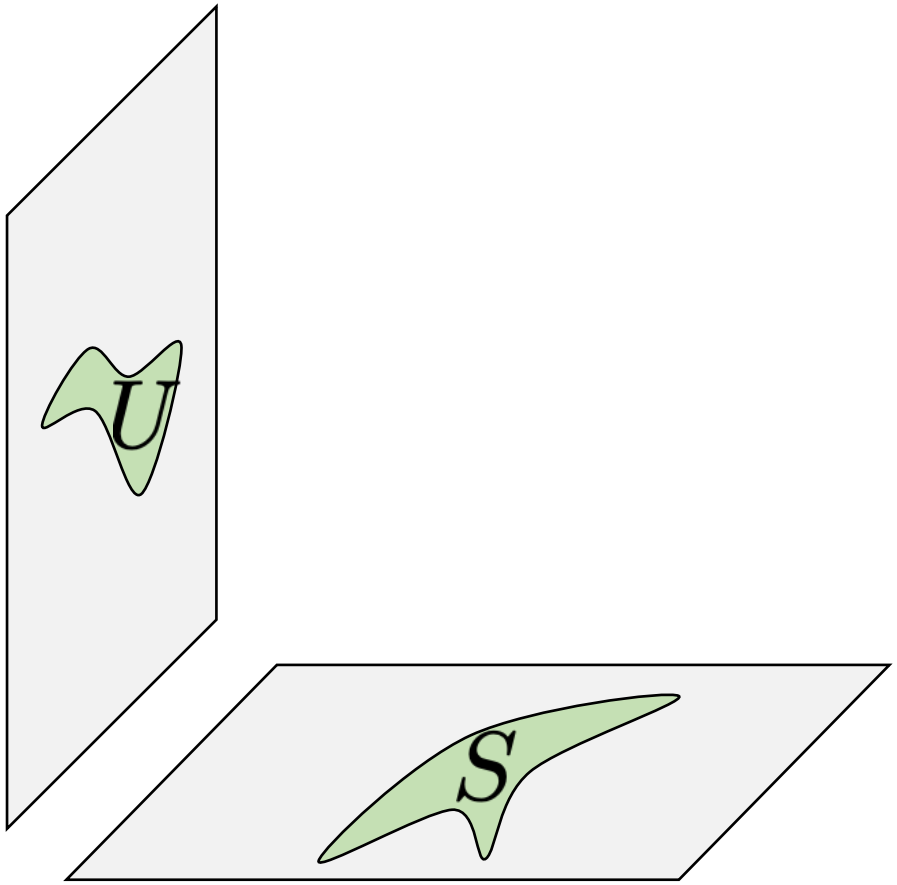


We allow some amount of error

Spectral Forrelation

Given two sets S, U (as membership oracles), decide if there is a state whose shadows lie within those sets

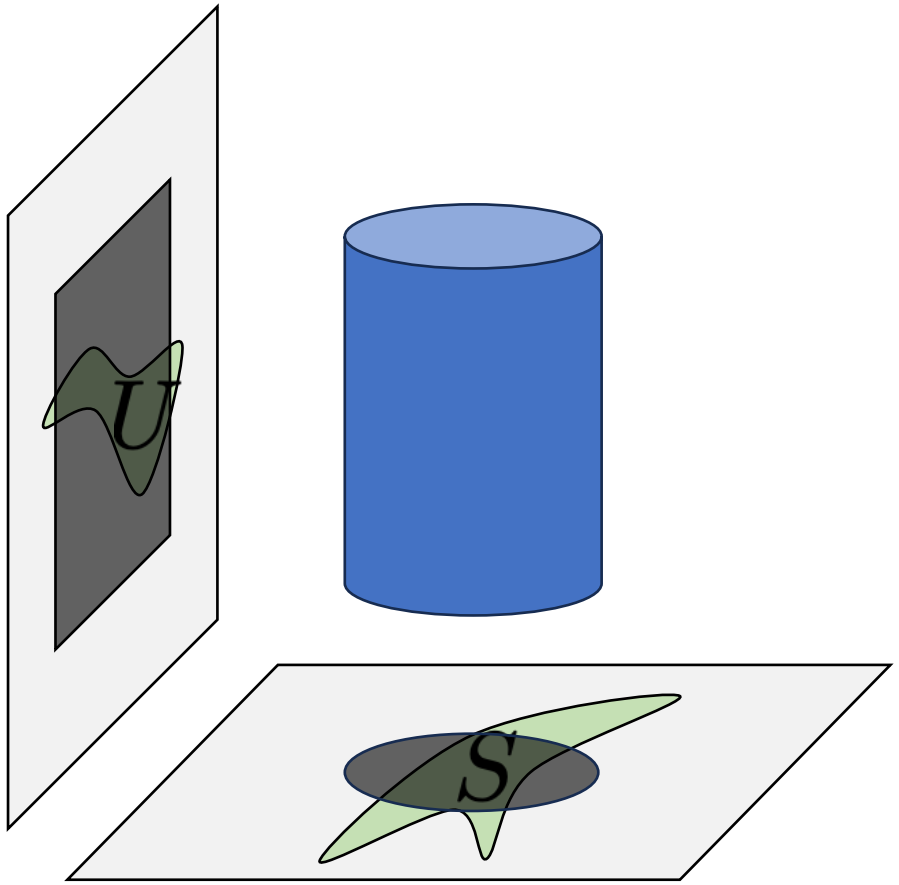
No instance:



Spectral Forrelation

Given two sets S, U (as membership oracles), decide if there is a state whose shadows lie within those sets

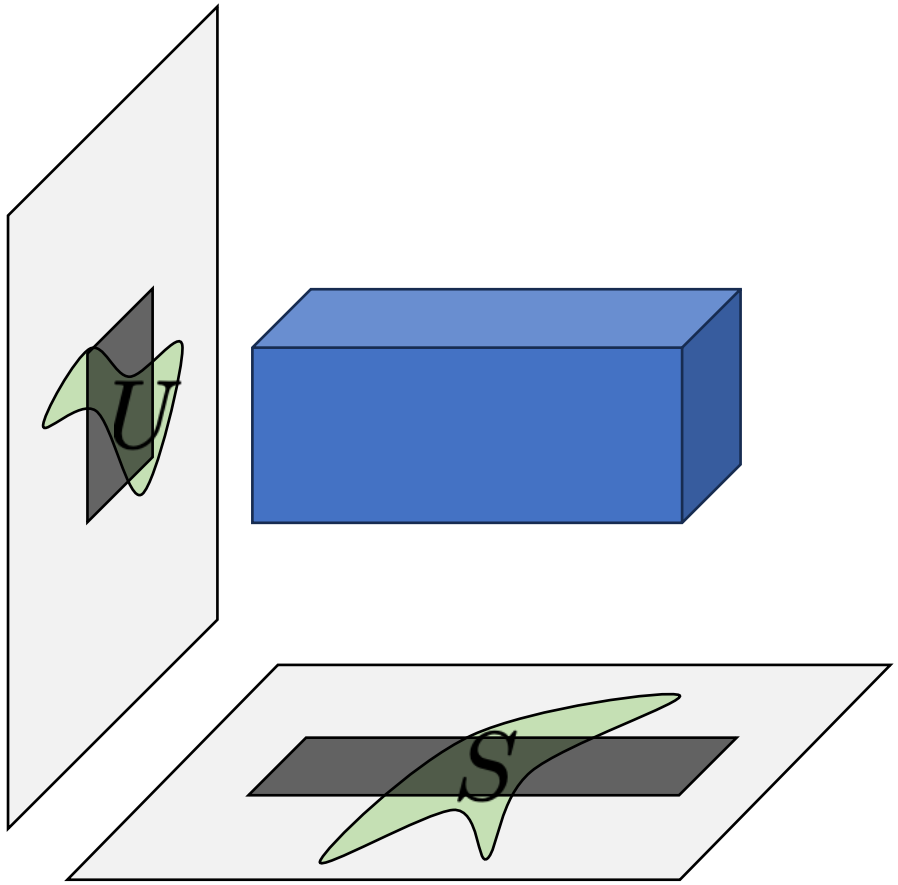
No instance:



Spectral Forrelation

Given two sets S, U (as membership oracles), decide if there is a state whose shadows lie within those sets

No instance:



Not Hard Thm: Spectral Forrelation $\in$ Oracle-QMA

Witness = state with given shadows

Verifier = query to check that support consistent with given shadow

Much Harder Theorem:
Spectral Forrelation $\notin$ Oracle-QCMA

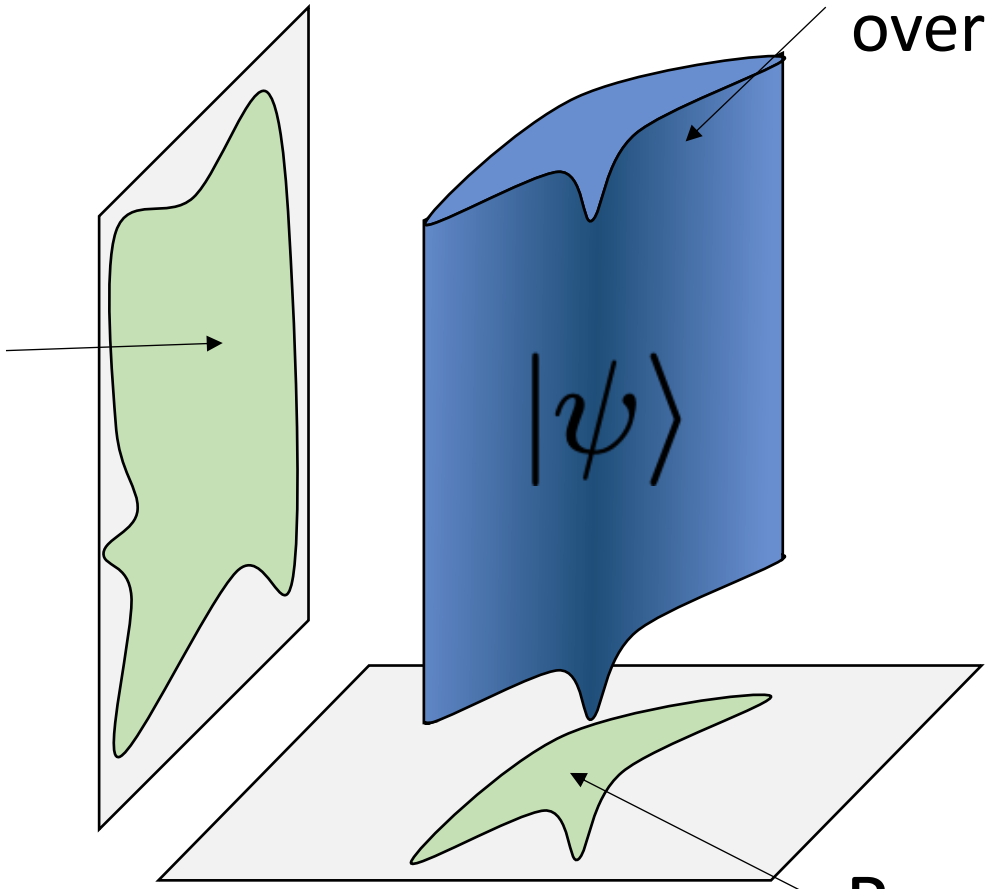
Key Challenge: We have no idea how a QCMA verifier may work, and it may decide Spectral Forrelation without ever constructing the witness state

Need some approach to rule out general QCMA verifiers that doesn't simultaneously rule out QMA verifiers

Idea 0: Strong Yes Instances

Uniform superposition over S

U = Approx. Hadamard shadow of $|\psi\rangle$

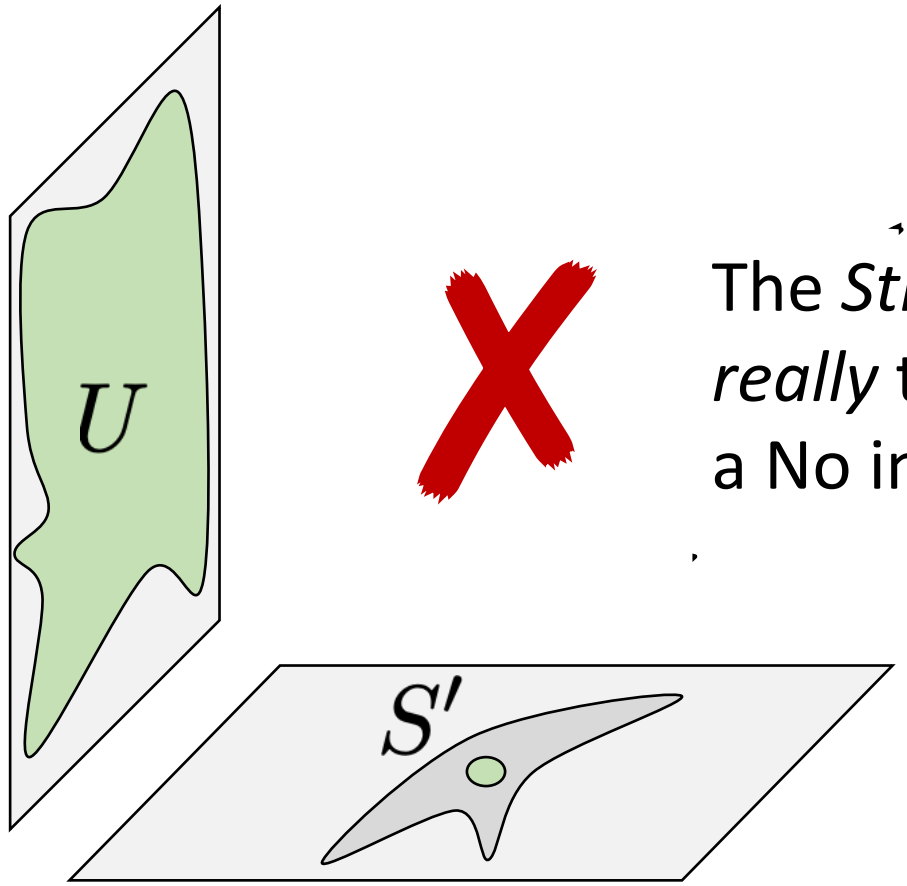


✓ by construction

Random sparse S

Say, $|S| = 2^{n/10}$

Idea 0: Strong Yes Instances



The *Strong Yes Property*: For any *really* tiny $S' \subseteq S, (S', U)$ is a No instance

Intuition

If we ignore U, finding points in S is hard

Therefore, any superposition that passes verification intuitively must be derived from witness

Classical witness may contain poly-many points in S, but by *Strong Yes Property*, any superposition over such points cannot pass verification

Idea 1: QCMA → Repeated Sampling

Thm: Any QCMA verifier can be turned into an efficient non-trivial sampler for S

If V is a QCMA verifier for Spectral Forrelation,

$\exists w$ such that $V(w)$ distinguishes (S, U) from $(\{\}, U)$

➡ $V^{\{\},U}(w)$ must query on accepting input to S

➡ Measure random query, obtain $x_1 \in S$ with inverse poly probability $n^{-O(1)}$

Idea 1: QCMA → Repeated Sampling

Thm: Any QCMA verifier can be turned into an efficient non-trivial sampler for S

But we can actually do the same with a QMA verifier!!!

What can we do with a QCMA verifier that we can't do with QMA?

Idea 1: QCMA → Repeated Sampling

Thm: Any QCMA verifier can be turned into an efficient non-trivial **repeated** sampler for S

Repeat! $V(w)$ distinguishes (S, U) from $(\{x_1\}, U)$

➡ $V^{\{x_1\},U}(w)$ must query on accepting input to $S \setminus \{x_1\}$

➡ Measure random query, obtain $x_2 \in S \setminus \{x_1\}$ with inverse poly probability $n^{-O(1)}$

Can't do with QMA, since obtaining x_1 may have destroyed witness

Idea 1: QCMA → Repeated Sampling

Thm: Any QCMA verifier can be turned into an efficient non-trivial **repeated** sampler for S

Repeat! v trials ➡ v distinct points in S with prob $n^{-O(v)}$

Compare to random guessing: $\left(2^{-9n/10}\right)^v = 2^{-O(nv)}$

QCMA verifier gives an exponential improvement over random guessing

Idea 1: QCMA → Repeated Sampling

Moving to repeated sampling accomplishes two things:

- Exponential advantage over trivial
- Allows us to remove witness

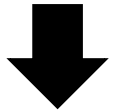
If witness length q, guess witness, incurring 2^{-q} loss

v distinct points in S with prob $2^{-q}n^{-O(v)}$

For $v \gg q$, still exponential advantage

Recap so far

QCMA verifier



Sampler $\mathsf{Samp}^U()$ yielding v distinct points in S
with probability $n^{-O(v)}$ for large enough v

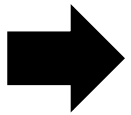
Such a sampler trivially impossible without access to U

Even impossible with access to S but not U [Hamoudi-Magniez'20]

But maybe U reveals too much info about S ?

At a conceptual level, [Z'24] got up to this point, but got stuck

Basically hypothesized that U indistinguishable from random oracle



Simulate U to get oracle-free sampler

Substantiated hypothesis based on conjectured relationship between relative on marginals and overall absolute error

But conjecture turns out to be false!

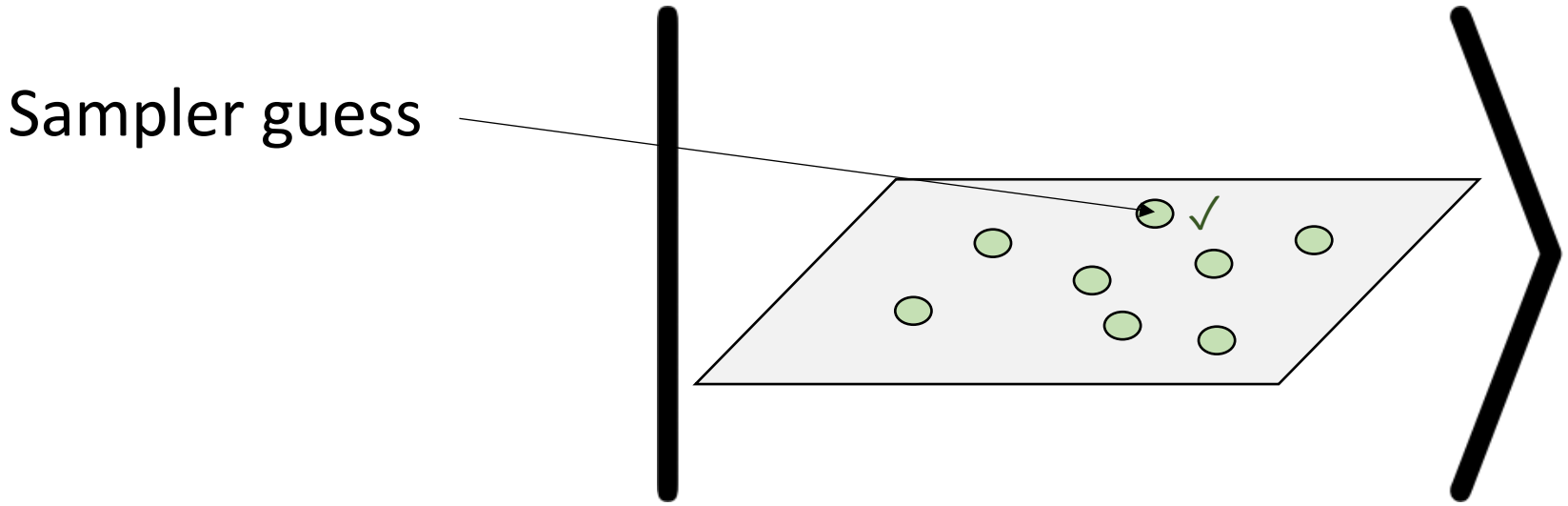
Idea 2: Purification → Bosons

Instead of choosing a random set S, purify
to obtain a superposition over all S

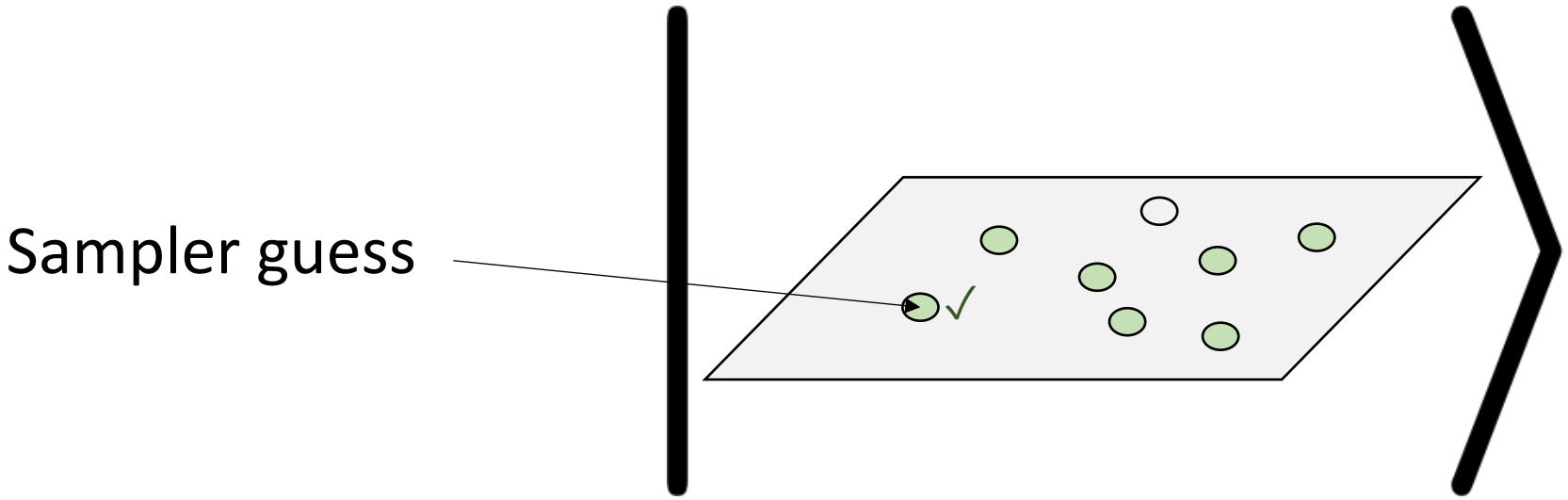
We will actually take S to be a multi-set,
where each point is placed uniformly,
without requiring distinctness

By sparseness, allowing repeats negligibly changes distribution

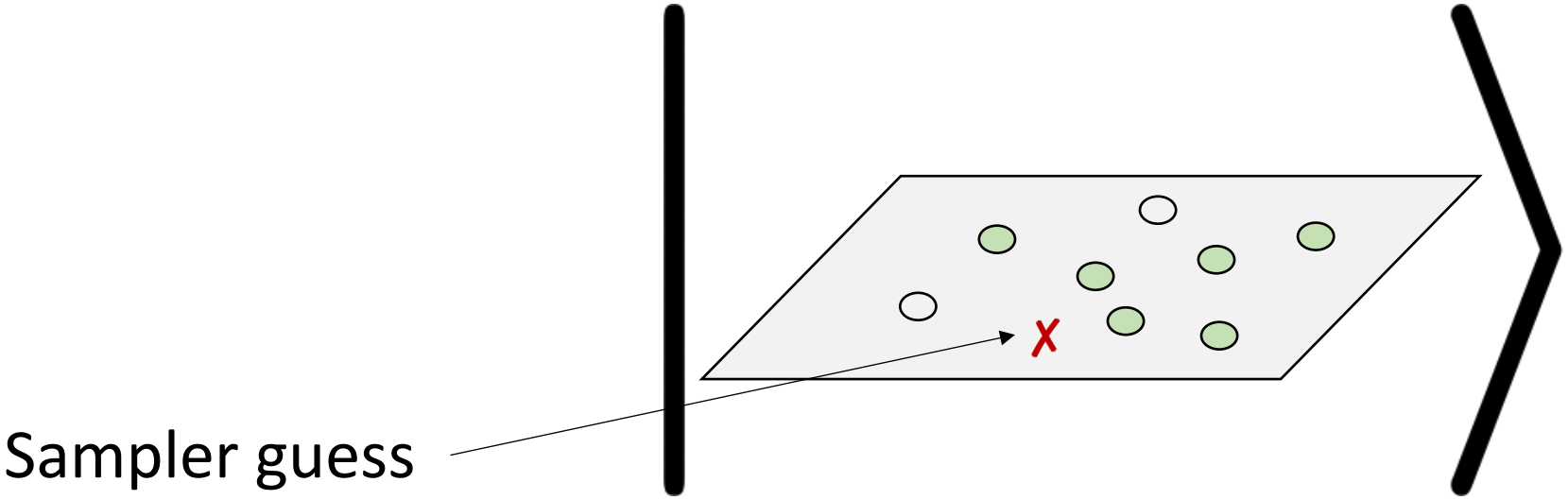
Each output of sampler is a position, and we check if there is a boson at that position, and if so, remove it



Each output of sampler is a position, and we check if there is a boson at that position, and if so, remove it



Each output of sampler is a position, and we check if there is a boson at that position, and if so, remove it



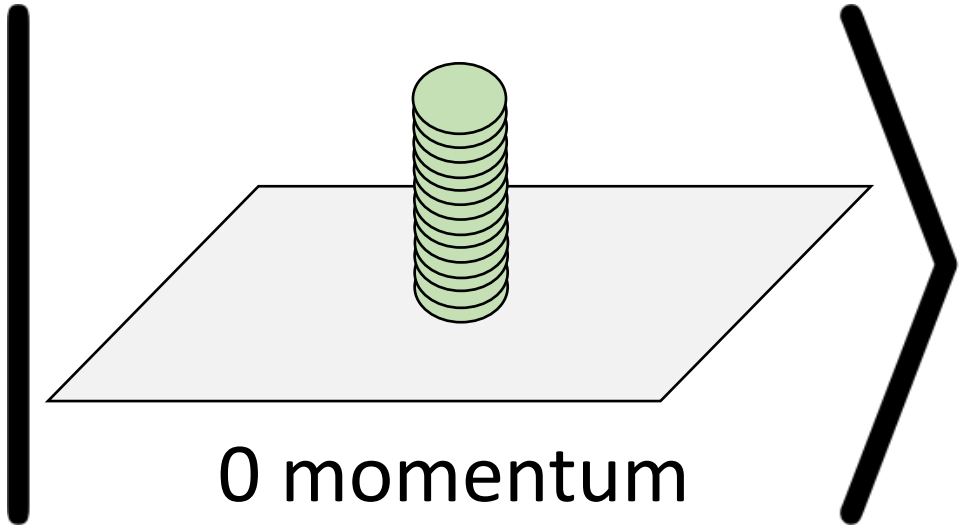
Sampler succeeds if each position has a boson

Idea 2: Purification → Bosons

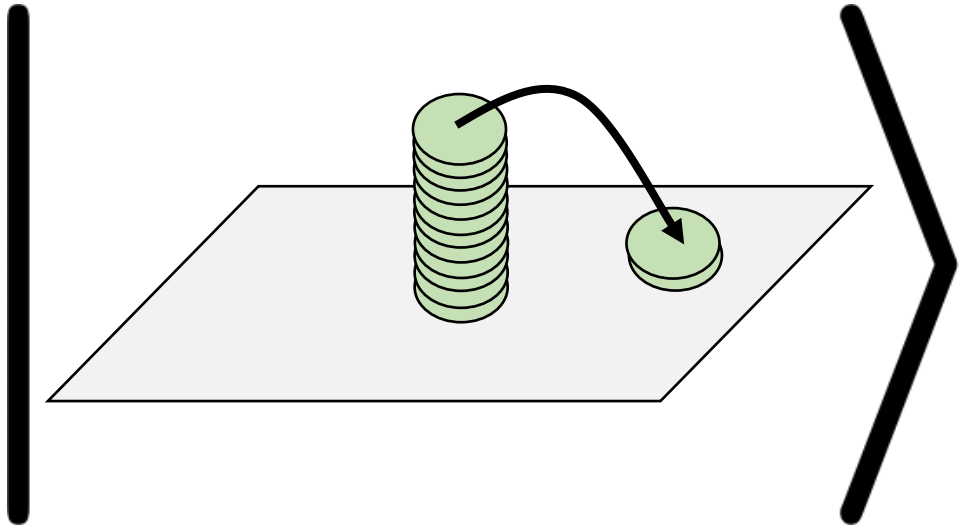
If sampler makes no queries to U, easy to show necessary bound

Now we need to understand how queries to U change the system

Look at momentum (that is, Hadamard) basis



Query U on y → add momentum y to **two** random bosons



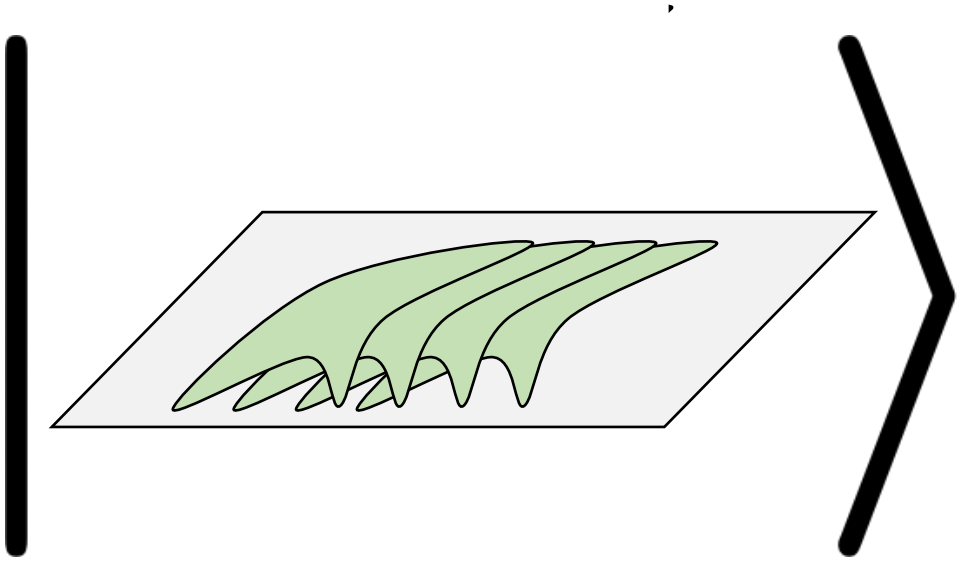
Two bosons due to us putting points in U based on norm^2 of amplitudes in $|\psi\rangle$

Technically some chance of moving 4 bosons, 6, bosons, etc, but we'll ignore

Key observations:

(1) Regardless of queries to U, total momentum stays 0 (mod 2)

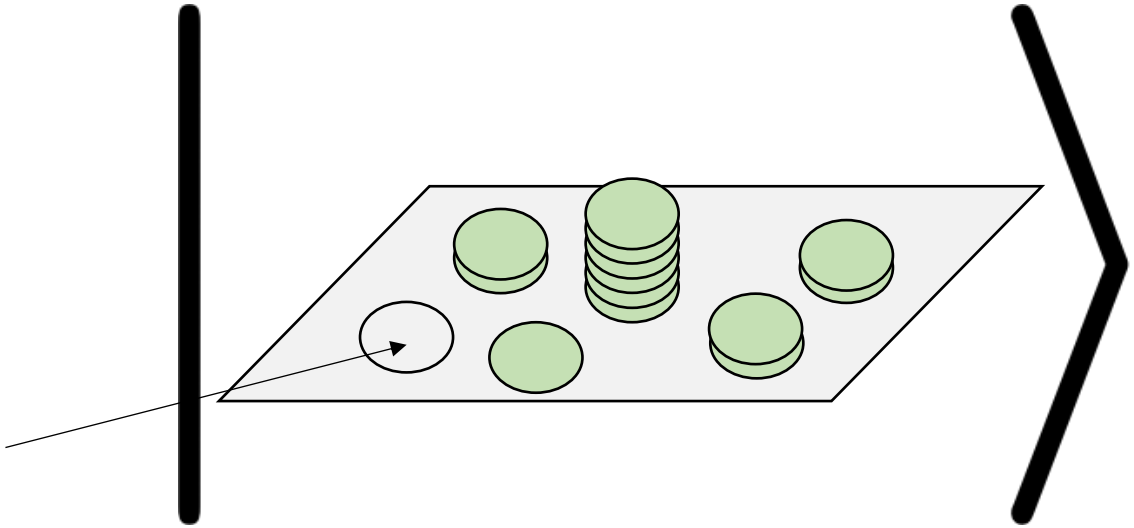
➡ By uncertainty principle, shift invariant in position basis



➡ Probability of finding first point = $|S|/2^n$

Still not enough! Once first point is found, finding more points may be easy since points are correlated, preventing us from obtaining a sufficiently small bound

In boson terms: post-selecting on the first successful guess, remaining bosons may have non-zero momentum



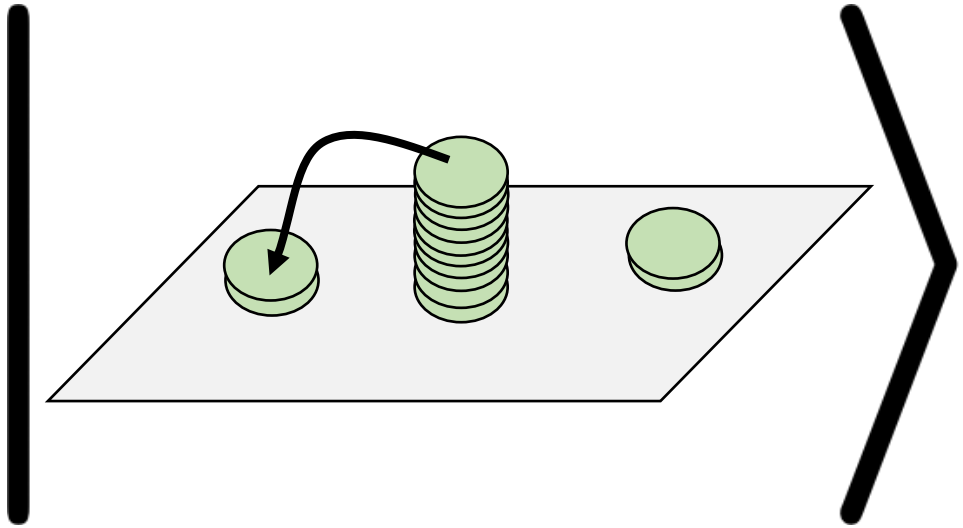
→ no position shift invariance

→ no guarantees on subsequent success probability

Key observations:

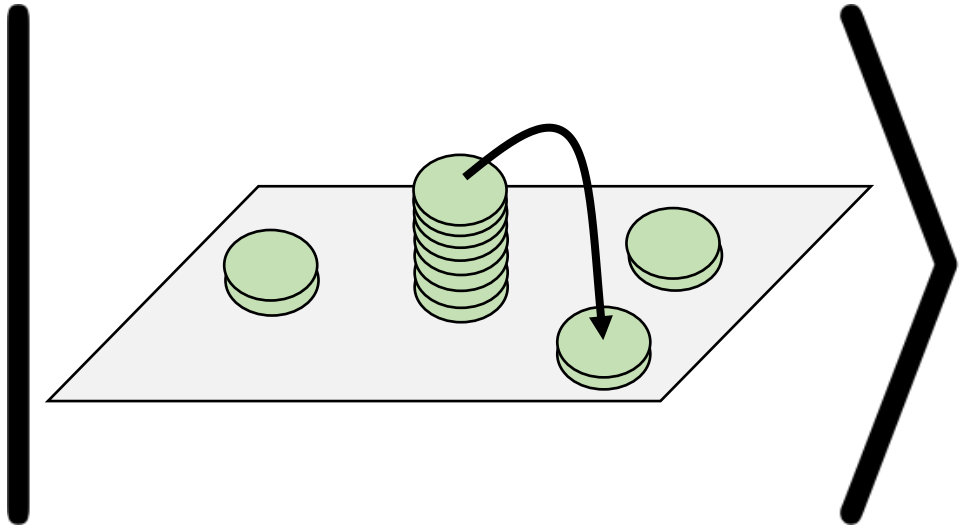
$|S| \gg \#(\text{queries})$

(2) Since most bosons have 0 momentum, each query to U will most likely kick off 0-momentum bosons



Key observations:

(2) Since most bosons have 0 momentum, each query to U will most likely kick off 0-momentum bosons



➡ Most of superposition will have non-zero bosons paired off

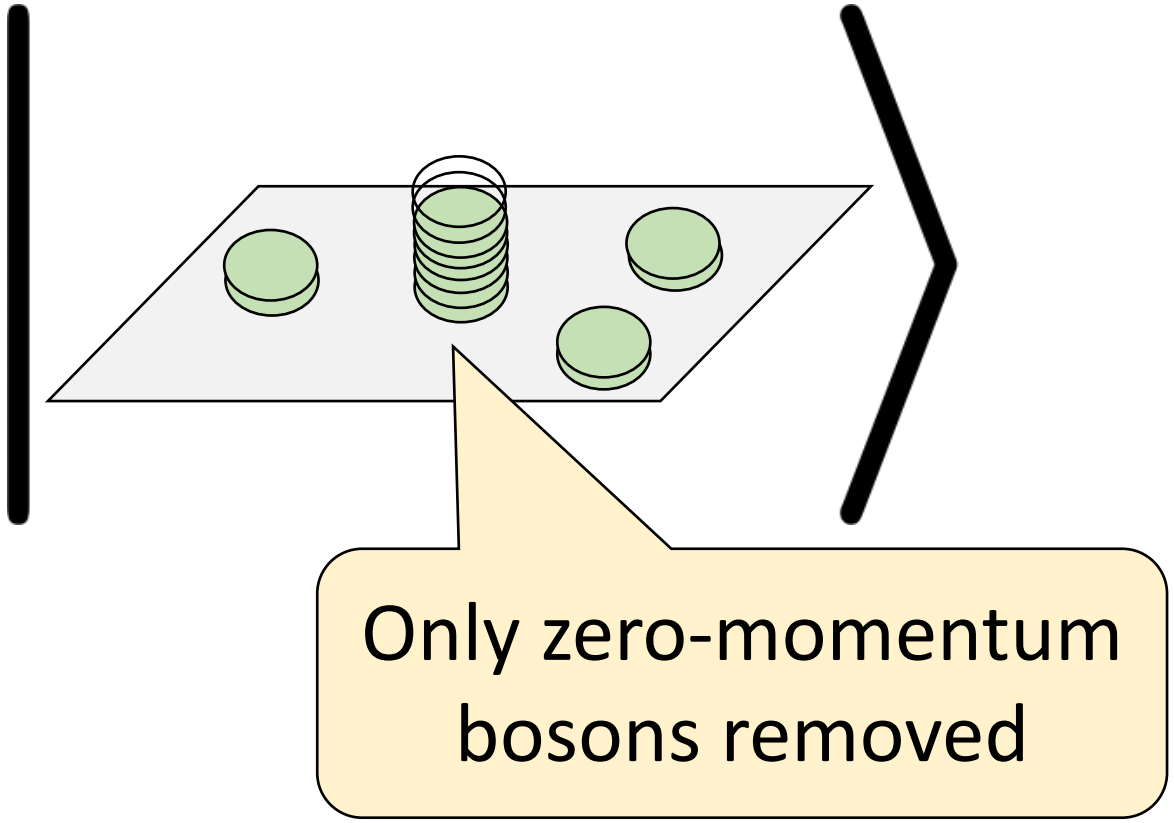
Key observations:

(3) Each pair of non-zero-momentum bosons still has zero total momentum for that pair

➡ Probability of finding these bosons = $2 \times \#(\text{queries})/2^n$

➡ Any found boson will almost certainly have zero momentum

Thus, post-selected state after several successful guesses will still have zero momentum on remaining bosons



Result: finding each subsequent boson has exponentially small probability → overall success probability sufficiently small

Actually proving this is very delicate: post-selection involves multiplying by exponential re-normalization constant → errors potentially blow up

Proof recap

Supposed QCMA verifier → efficient repeated sampler with success probability $n^{-O(v)}$

But any efficient sampler has success probability $2^{-O(nv)}$

Contradiction! QCMA verifier does not exist

Subsequent work

[Bostanci-Huang-Vaikuntanathan'26] a new, dramatically simpler, separation based on [Yamakawa-**Z**'22]

Also yields a classical oracle separation between **BQP/qpoly** and **BQP/poly**

Speaking at QIQC on May 5th!

Thanks!