PAC-Learning and Reconstruction of Quantum States

Scott Aaronson (University of Texas at Austin)

KITP, October 12, 2017

The Problem

A state of n entangled qubits requires **2^n** complex numbers to specify, even approximately:

$$|\psi\rangle = \sum_{x\in\{0,1\}^n} \alpha_x |x\rangle$$

Why is this a problem?

-Quantum state tomography

-Foundations of quantum mechanics (cf. Dorit's talk)

Can we tame the exponential beast?

Idea: "Shrink quantum states down to reasonable size" by viewing them operationally

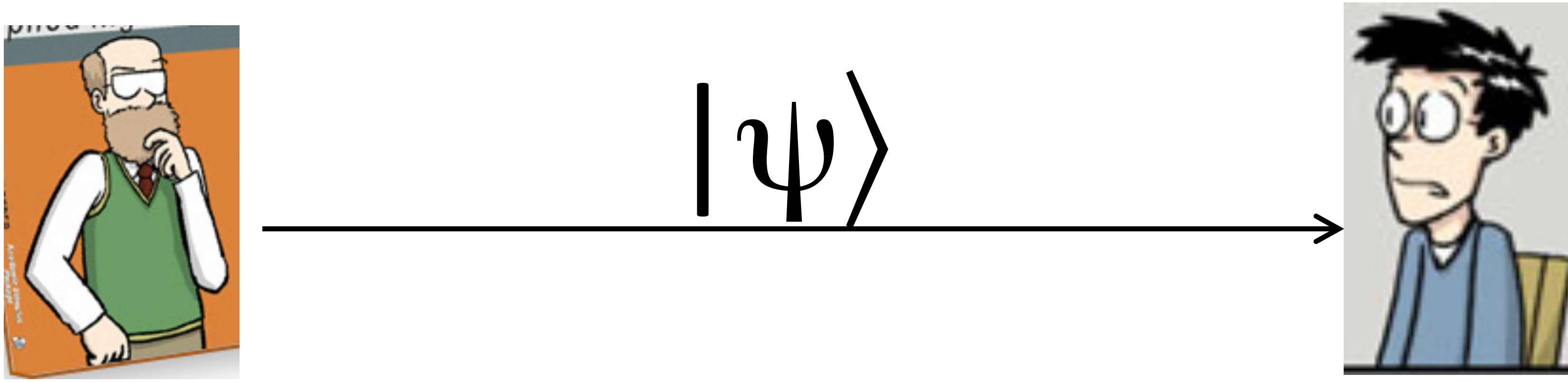
Analogy: A probability distribution over n-bit strings *also* takes ~2^n bits to specify. But seeing a sample only provides n bits

In this talk, I'll survey 14 years of results showing how tools from computational learning theory can be used to upper-bound the "effective size" of quantum states

[A. 2004], [A. 2006], [A.-Drucker 2010], [A. 2017], [A. et al. 2017]

Lesson: "The linearity of QM helps tame the exponentiality of QM"

The Absent-Minded Advisor Problem



Can you hand all your grad students the same $n^{O(1)}$-qubit quantum state $|\psi\rangle$, so that by measuring their copy of $|\psi\rangle$ in a suitable basis, each student can learn the {0,1} answer to their n-bit thesis question?

NO [Nayak 1999, Ambainis et al. 1999]

Indeed, quantum communication is no better than classical for this task as $n \to \infty$

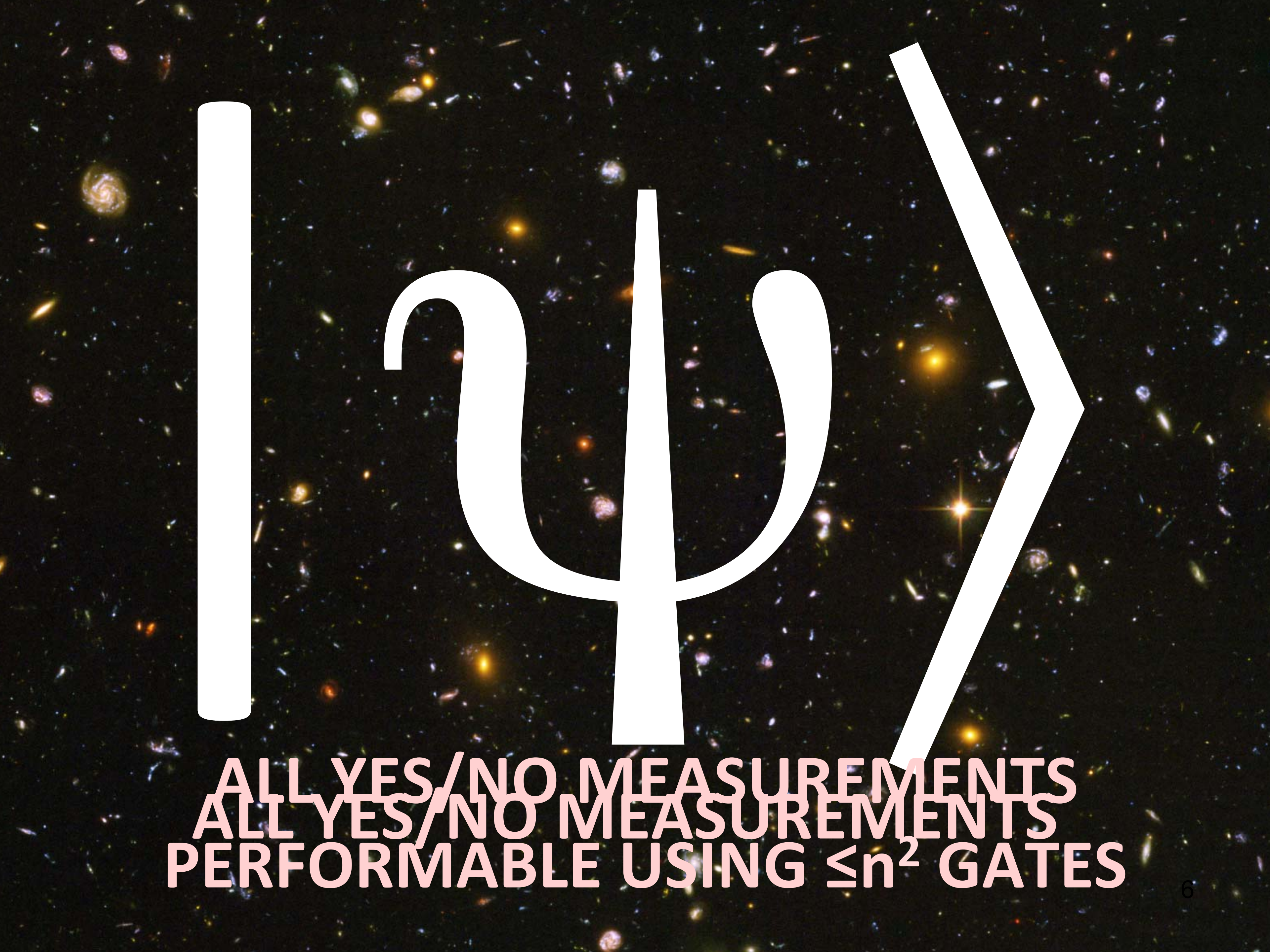
On the Bright Side…

Suppose Alice wants to describe an n-qubit quantum state $|\psi\rangle$ to Bob, well enough that, for any 2-outcome measurement E in some finite set S, Bob can estimate Pr[E($|\psi\rangle$) accepts] to constant additive error

Theorem (A. 2004): In that case, it suffices for Alice to send Bob only

$$\sim n \log n \cdot \log|S|$$

classical bits (trivial bounds: **exp(n), |S|**)



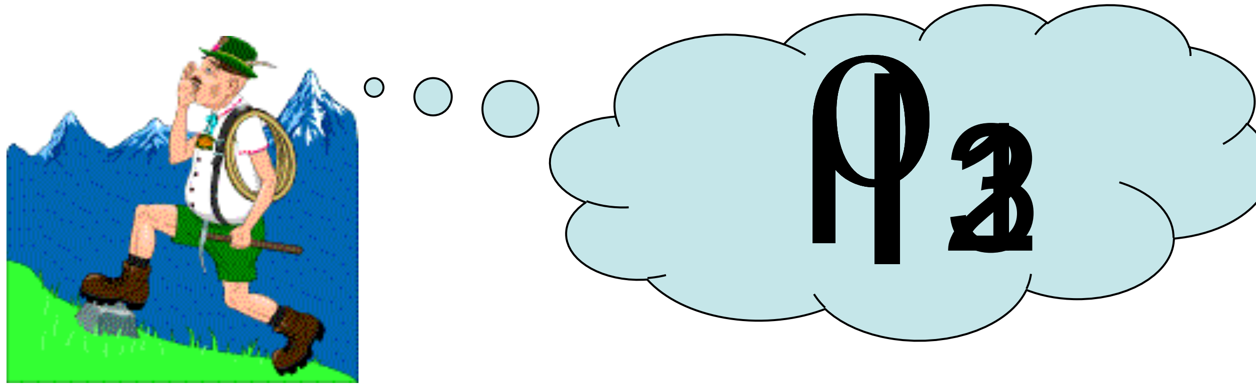
$$|\psi\rangle$$

ALL YES/NO MEASUREMENTS

ALL YES/NO MEASUREMENTS

PERFORMABLE USING $\leq n^2$ GATES

How does the theorem work?



Alice is trying to describe the n-qubit state $\rho=|\psi\rangle\langle\psi|$ to Bob

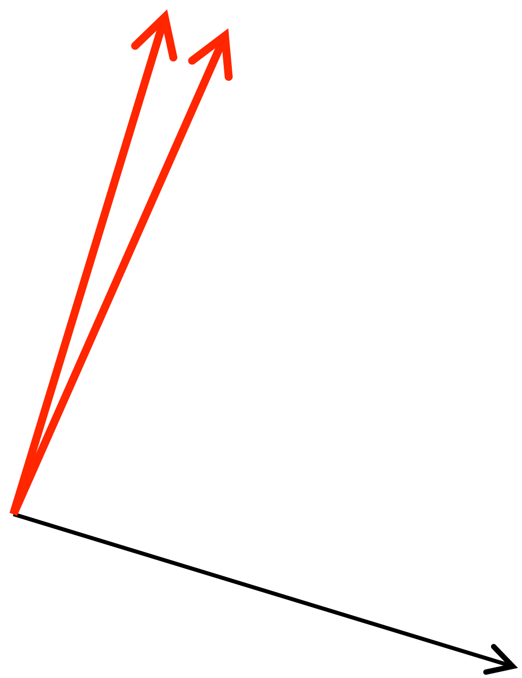
In the beginning, Bob knows nothing about ρ, so he guesses it's the **maximally mixed state** ρ_0=I (actually $I/2^n = I/D$)

Then Alice helps Bob **improve**, by repeatedly telling him a measurement $E_t \in S$ on which his current guess ρ_{t-1} badly fails

Bob lets ρ_t be the state obtained by starting from ρ_{t-1}, then performing E_t and **postselecting** on getting the right outcome

Question: Why can Bob keep reusing the same state?

To ensure that, we actually need to **"amplify"** $|\psi\rangle$ to $|\psi\rangle^{\otimes \log(n)}$, slightly increasing the number of qubits from n to n log n



Gentle Measurement / Almost As Good As New Lemma:

Suppose a measurement of a mixed state ρ yields a certain outcome with probability $\geq 1-\varepsilon$

Then after the measurement, we still have a state ρ' that's $\sqrt{\varepsilon}$-close to ρ in trace distance

Crucial Claim: Bob's iterative learning procedure will "converge" on a state ρ_T that behaves like $|\psi\rangle^{\otimes \log(n)}$ on all measurements in the set S, after at most T=O(n log(n)) iterations

Proof: Let p_t = Pr[first t postselections all succeed]. Then

$$I = \frac{1}{2^{n\log n}} \sum |\psi_i\rangle\langle\psi_i|, \quad |\psi_1\rangle = |\psi\rangle^{\otimes \log n}$$

Solving, we find that t = O(n log(n))

So it's enough for Alice to tell Bob about T=O(n log(n)) measurements $E_1,\ldots,E_T$, using log|S| bits per measurement

Complexity theory consequence:

BQP/qpoly ⊆ PostBQP/poly

(Open whether **BQP/qpoly**=**BQP/poly**)

Quantum Occam's Razor Theorem

Let $|\psi\rangle$ be an unknown entangled state of n particles

Suppose you just want to be able to estimate the acceptance probabilities of **most** measurements E drawn from some probability distribution μ

Then it suf… =O(n):

"Quantum states are PAC-learnable"

1.Choose m …

2.Go into your lab and estimate acceptance probabilities of all of them on $|\psi\rangle$

3.Find **any** "hypothesis state" approximately consistent with all measurement outcomes

How do we find the hypothesis state?

Here's one way: let $b_1,\ldots,b_m$ be the outcomes of measurements $E_1,\ldots,E_m$

Then choose a hypothesis state σ to minimize

$$\sum_{i=1}^{m}\left(\mathrm{Tr}\left(E_i\sigma\right)-b_i\right)^2$$

This is a convex programming problem, which can be solved in time polynomial in $D=2^n$ (good enough in practice for $n\leq 15$ or so)

Optimized, **linear-time** iterative method for this problem: [Hazan 2008]

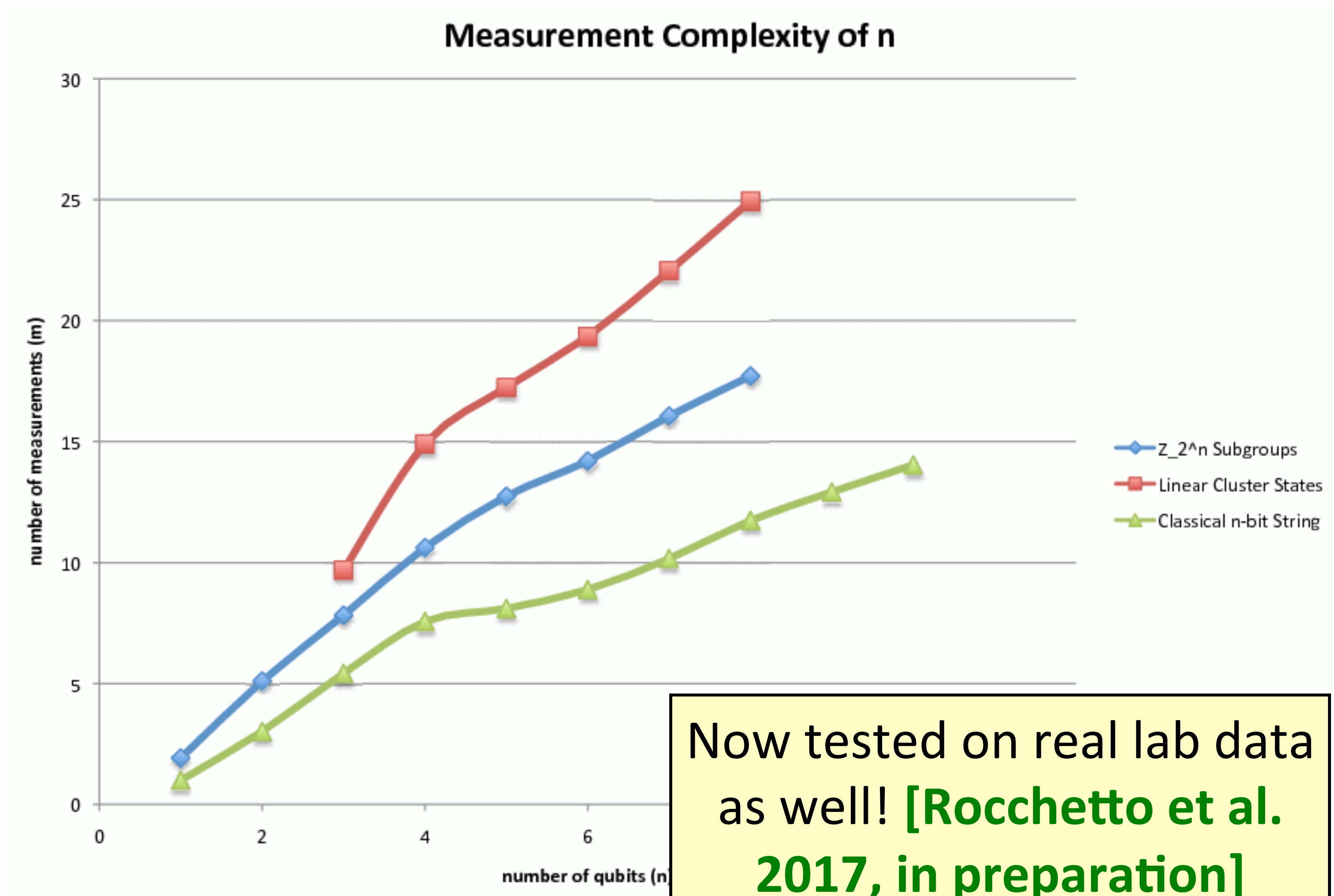
Numerical Simulation

[A.-Dechter 2008]

We implemented Hazan's algorithm in MATLAB, and tested it on simulated quantum state data

We studied how the number of sample measurements m needed for accurate predictions scales with the number of qubits n, for n≤10

Result of experiment: My theorem appears to be true…



Now tested on real lab data as well! **[Rocchetto et al. 2017, in preparation]**

Combining My Postselection and Quantum Learning Results

[A.-Drucker 2010]: Given an n-qubit state ρ and complexity bound T, there exists an efficient measurement V on poly(n,T) qubits, such that any state that "passes" V can be used to efficiently estimate ρ's response to any yes-or-no measurement E that's implementable by a circuit of size T

Application: Trusted quantum advice is equivalent to trusted *classical* advice + *untrusted* quantum advice

In complexity terms: **BQP/qpoly ⊆ QMA/poly**

Proof uses boosting-like techniques, together with results on ε-covers and fat-shattering dimension

New Result [A. 2017, in preparation]: "Shadow Tomography"

Theorem: Let ρ be an unknown D-dimensional state, and let $E_1,\ldots,E_M$ be known 2-outcome measurements. Suppose we want to know $\mathrm{Tr}(E_i\rho)$ to within additive error $\pm\varepsilon$, for **all** $i \in [M]$. We can achieve this, with high probability, given only k copies of ρ, where

$$k = \widetilde{O}\left(\frac{\log^4 M \cdot \log D}{\varepsilon^4}\right)$$

Open Problem: Dependence on D removable?

Challenge: How to measure $E_1,\ldots,E_M$ without destroying our few copies of ρ in the process!

Proof Idea

Theorem [A. 2006, fixed by Harrow-Montanaro-Lin 2016]:

Let ρ be an unknown D-dimensional state, and let $E_1,\ldots,E_M$ be known 2-outcome measurements. Suppose we're promised that either there exists an i such that $Tr(E_i\rho) \geq c$, or else $Tr(E_i\rho) \leq c-\varepsilon$ for all $i \in [M]$. We can decide which, with high probability, given only k copies of ρ, where

$$k = O\left(\frac{\log M}{\varepsilon^2}\right)$$

Indeed, can **find** an i with $Tr(E_i\rho) \geq c-\varepsilon$, if given $O(\log^4 M / \varepsilon^2)$ copies of ρ

Now run my postselected learning algorithm [A. 2004], but using the above to find the E_i's to postselect on!

Online Learning of Quantum States

A., Chen, Hazan, Nayak 2017

Corollary of my Postselected Learning Algorithm: Let ρ be an ur… 2- outc… follov… ne, we're… t

We ca…

Can use convex optimization tools to improve the $1/\varepsilon^3$ to $1/\varepsilon^2$, and also to prove a **regret bound**: Even if the data you get isn't consistent with any actual quantum state, you can output a sequence of hypotheses that, after T iterations, has made only O(√(Tn)) more mistakes than the best hypothesis you could find after having seen all the data

$$k = O\left(\frac{\cdots}{\varepsilon^3} \log \log \frac{\cdots}{\varepsilon} \right)$$

Summary

The tools of learning theory let us show that often, the "exponentiality" of quantum states is more bark than bite

I.e. there's a short classical string that specifies how the quantum state behaves, on any 2-outcome measurement you could actually perform

Applications from complexity theory to experimental quantum state tomography…

Alas, these results don't generalize to many-outcome measurements, or to learning quantum **processes**

Biggest future challenge: Find subclasses of states and measurements for which these learning procedures are *computationally* efficient (Stabilizer states: Rocchetto 2017)