Title: Pretty-Good Tomography

Date: Aug 26, 2008 09:30 AM

URL: http://pirsa.org/08080036

Abstract: I\'ll survey recent results from quantum computing theory showing that,if one just wishes to learn enough about a quantum state to predictthe outcomes of most measurements that will actually be made, then itoften suffices to perform exponentially fewer measurements than wouldbe needed in quantum state tomography. I\'ll then describe the resultsof a numerical simulation of the new quantum state learning approach.The latter is joint work with Eyal Dechter.

Pretty-Good Tomography

Scott Aaronson

MIT

There's a problem…

There's a problem…

To do tomography on an entangled state of n qubits, we need exp(n) measurements

There’s a problem…

To do tomography on an entangled state of n qubits, we need exp(n) measurements

Does this mean that a generic state of (say) 10,000 particles can never be “learned” within the lifetime of the universe?

There’s a problem…

To do tomography on an entangled state of n qubits, we need exp(n) measurements

Does this mean that a generic state of (say) 10,000 particles can never be “learned” within the lifetime of the universe?

If so, this is certainly a practical problem—but to me, it’s a **conceptual** problem as well

There's a problem…

To do tomography on an entangled state of n qubits, we need exp(n) measurements

Does this mean that a generic state of (say) 10,000 particles can never be "learned" within the lifetime of the universe?

If so, this is certainly a practical problem—but to me, it's a **conceptual** problem as well

What **is** a quantum state?

What **is** a quantum state?

A “state of the world”? A “state of knowledge”?

What **is** a quantum state?

A "state of the world"? A "state of knowledge"?

Whatever else it is, should at least be a **useful hypothesis** that encapsulates previous observations and lets us predict future ones

What **is** a quantum state?

A "state of the world"? A "state of knowledge"?

Whatever else it is, should at least be a **useful hypothesis** that encapsulates previous observations and lets us predict future ones

How "useful" is a hypothesis that takes 10^{5000} bits even to write down?

What **is** a quantum state?

A "state of the world"? A "state of knowledge"?

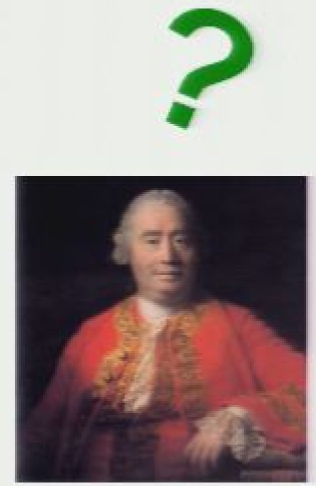
Whatever else it is, should at least be a **useful hypothesis** that encapsulates previous observations and lets us predict future ones

How "useful" is a hypothesis that takes 10^{5000} bits even to write down?

Seems to bolster the arguments of quantum computing skeptics who think quantum mechanics will break down in the "large N limit"

Really we're talking about Hume's Problem of Induction…

You see 500 ravens. Every one is black. Why does that give you any grounds whatsoever for expecting the next raven to be black?



The answer, according to computational learning theory: In practice, we always restrict attention to some class of hypotheses vastly smaller than the class of all logically conceivable hypotheses

Probably Approximately Correct (PAC) Learning

Set S called the **sample space**

Probability distribution D over S

Class C of **hypotheses**: functions from S to {0,1}

Unknown function $f \in C$

Goal: Given $x_1, \ldots, x_m$ drawn independently from D, together with $f(x_1), \ldots, f(x_m)$, output a hypothesis $h \in C$ such that

$$\Pr_{x \in D}\left[h(x) = f(x)\right] \geq 1 - \varepsilon,$$

with probability at least $1-\delta$ over $x_1, \ldots, x_m$

Occam's Razor Theorem

Valiant 1984: If the hypothesis class C is finite, then any hypothesis consistent with

$$m = O\left(\frac{1}{\varepsilon}\log\frac{|C|}{\delta}\right)$$

random samples will also be consistent with a 1-ε fraction of future data, with probability at least 1-δ over the choice of samples

Occam's Razor Theorem

Valiant 1984: If the hypothesis class C is finite, then any hypothesis consistent with

$$m = O\left(\frac{1}{\varepsilon}\log\frac{|C|}{\delta}\right)$$

random samples will also be consistent with a 1-ε fraction of future data, with probability at least 1-δ over the choice of samples

"Compression implies prediction"

Occam's Razor Theorem

Va … thesis class C is finite, then any
b

But the number of quantum states is infinite!

$$O\left(\frac{1}{\varepsilon}\log\frac{|C|}{\delta}\right)$$

random samples will also be consistent with a 1-ε fraction of future data, with probability at least 1-δ over the choice of samples

"Compression implies prediction"

Occam's Razor Theorem

Va... thesis class C is finite, then any h...

$$O\left(\frac{1}{\varepsilon}\log\ldots|C|\right)$$

random samples will als... fraction of future data, w... the choice of samples

But the number of quantum states is infinite!

And even if we discretize, it's still **doubly** exponential in the number of qubits!

"Compression implies prediction"

A Hint of What's Possible…

Theorem [A. 2004]: Any n-qubit quantum state can be "simulated" using O(n log n log m) classical bits, where m is the number of (binary) measurements whose outcomes we care about.

A Hint of What's Possible…

Theorem [A. 2004]: Any n-qubit quantum state can be "simulated" using O(n log n log m) classical bits, where m is the number of (binary) measurements whose outcomes we care about.

Let $E=(E_1,\ldots,E_m)$ be two-outcome POVMs on an n-qubit state ρ. Then given (classical descriptions of) E and ρ, we can produce a classical string of

$$\tilde{O}\left(\frac{n\log n}{\varepsilon^2}\cdot\log m\right)$$

bits, from which $\mathrm{Tr}(E_i\rho)$ can be estimated to within additive error ε given any E_i (without knowing ρ).

A Hint of What's Possible…

Theorem [A. 2004]: Any n-qubit quantum state can be "simulated" using O(n log n log m) classical bits, where m is the number of (binary) measure… …bout.

> **Proof idea:** Start with the maximally mixed state as your hypothesis, then find a "Darwinian training set" of measurements within $\{E_1,\ldots,E_m\}$ such that postselecting on their outcomes improves the hypothesis

Let E=(E… …on an n-qubit stat… …tions of) E and ρ, … …of

bits, from which $Tr(E_i\rho)$ can be estimated to within additive error ε given any E_i (without knowing ρ).

Quantum Occam's Razor Theorem
[A. 2006]

Let ρ be an n-qubit state, and let D be a distribution over two-outcome measurements.

Suppose we draw measurements $E_1,\ldots,E_m$ independently from D, and then find a hypothesis state σ that minimizes

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

Then $\Pr_{E\in D}\left[\left|\mathrm{Tr}\left(E\sigma\right)-\mathrm{Tr}\left(E\rho\right)\right|\leq\gamma\right]\geq 1-\varepsilon$

with probability at least 1-δ over $E_1,\ldots,E_m$, provided

$$m\geq\frac{C}{\gamma^4\varepsilon^2}\left(\frac{n}{\gamma^4\varepsilon^2}\log^2\frac{1}{\gamma\varepsilon}+\log\frac{1}{\delta}\right) \qquad (C \text{ a constant})$$

A Hint of What's Possible…

Theorem [A. 2004]: Any n-qubit quantum state can be "simulated" using O(n log n log m) classical bits, where m is the number of (binary) measure… …bout.

Let E=(E… …on an n-qubit stat… …tions of) E and ρ, … …of

Proof idea: Start with the maximally mixed state as your hypothesis, then find a "Darwinian training set" of measurements within $\{E_1,\ldots,E_m\}$ such that postselecting on their outcomes improves the hypothesis

bits, from which $Tr(E_i\rho)$ can be estimated to within additive error ε given any E_i (without knowing ρ).

Quantum Occam's Razor Theorem
[A. 2006]

Let ρ be an n-qubit state, and let D be a distribution over two-outcome measurements.

Suppose we draw measurements $E_1,\ldots,E_m$ independently from D, and then find a hypothesis state σ that minimizes

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

(b_i = outcome of E_i)

Then $$\Pr_{E\in D}\left[\left|\mathrm{Tr}\left(E\sigma\right)-\mathrm{Tr}\left(E\rho\right)\right|\leq\gamma\right]\geq 1-\varepsilon$$

with probability at least 1-δ over $E_1,\ldots,E_m$, provided

$$m\geq\frac{C}{\gamma^4\varepsilon^2}\left(\frac{n}{\gamma^4\varepsilon^2}\log^2\frac{1}{\gamma\varepsilon}+\log\frac{1}{\delta}\right)$$

(C a constant)

Quantum Occam's Razor Theorem

[A. 2006]

Let ρ be an n-qubit state, and let D be a distribution over two-outcome measurements.

Suppose we draw measurements $E_1,\ldots,E_m$ independently from D, an Proof builds on results on *quantum* minimizes

random access codes due to

Ambainis et al. and Nayak, and on me of E_i)

learning of real-valued concept

Then $\Pr_{E \in D}$ classes due to Alon et al. and

Bartlett and Long

with probability at least 1-δ over $E_1,\ldots,E_m$, provided

$$m \geq \frac{C}{\gamma^4 \varepsilon^2}\left(\frac{n}{\gamma^4 \varepsilon^2}\log^2\frac{1}{\gamma\varepsilon}+\log\frac{1}{\delta}\right)$$ (C a constant)

Beyond the Bayesian and Max-Lik creeds: a third way?

Beyond the Bayesian and Max-Lik creeds: a third way?

We're not assuming any prior over states

Removes a lot of problems!

Beyond the Bayesian and Max-Lik creeds: a third way?

Beyond the Bayesian and Max-Lik creeds: a third way?

We're not assuming any prior over states

Removes a lot of problems!

Beyond the Bayesian and Max-Lik creeds: a third way?

We're not assuming any prior over states

Removes a lot of problems!

Beyond the Bayesian and Max-Lik creeds: a third way?

We're not assuming any prior over states

Removes a lot of problems!

Instead we assume a distribution over **measurements**

Beyond the Bayesian and Max-Lik creeds: a third way?

We're not assuming any prior over states

Removes a lot of problems!

Instead we assume a distribution over **measurements**

Why might that be preferable for some applications?

We can control which measurements to apply, but not what the state is

Generalization to process tomography?

Beyond the Bayesian and Max-Lik creeds: a third way?

We're not assuming any prior over states

Removes a lot of problems!

Instead we assume a distribution over **measurements**

Why might that be preferable for some applications?

We can control which measurements to apply, but not what the state is

Generalization to process tomography?

Generalization to process tomography?

No!

Suppose $U|x\rangle=(-1)^{f(x)}|x\rangle$, for some random Boolean function $f:\{0,1\}^n \rightarrow \{0,1\}$

Then the values of f(x) constitute 2^n **independently accessible** bits to be learned about

Yet each measurement provides at most n of the bits, by Holevo's Theorem

Hence, no analogue of my learning theorem is going to be true

How do we actually **find** σ?

Extension to k-outcome measurements?

Extension to k-outcome measurements?

Sure, if we increase the number of sample measurements m by a poly(k) factor

Note that there's no hope of learning to simulate 2^n-outcome measurements (i.e. measurements on all n qubits) after poly(n) sample measurements

How do we actually find σ?

Let $b_1,\ldots,b_m$ be the binary 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 the Hilbert space dimension $N=2^n$

In general, we can't hope for better than this—for basic computational complexity reasons

How do we actually find σ?

Let $b_1,\ldots,b_m$ be the binary outcomes of measurements $E_1,\ldots,E_m$

Then choose a hypothesis state σ to minimize

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

How do we actually find σ?

Let $b_1, \ldots, b_m$ be the binary 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 the Hilbert space dimension $N=2^n$

In general, we can't hope for better than this—for basic computational complexity reasons

Extension to k-outcome measurements?

Custom Convex Programming Method

[E. Hazan, 2008]

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

Set $S_0 := I/N$

For $t:=0$ to ∞

Compute smallest eigenvector v_t of $\nabla f(S_t)$

Compute step size α_t that minimizes $f(S_t+\alpha_t(v_t v_t^* - S_t))$

Set $S_{t+1} := S_t + \alpha_t(v_t v_t^* - S_t)$

Custom Convex Programming Method

[E. Hazan, 2008]

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

Set S_0 := I/N

For t:=0 to ∞

Compute smallest eigenvector v_t of $\nabla f(S_t)$

Compute step size α_t that minimizes $f(S_t+\alpha_t(v_t v_t^* - S_t))$

Set $S_{t+1} := S_t + \alpha_t(v_t v_t^* - S_t)$

Theorem (Hazan): This algorithm returns an ε-optimal solution after only $\log(m)/\varepsilon^2$ iterations.

Implementation
[A. & Dechter 2008]

We implemented Hazan's algorithm in MATLAB

Code available on request

Using MIT's computing cluster, we then did numerical simulations to check experimentally that the learning theorem is true

Custom Convex Programming Method

[E. Hazan, 2008]

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

Set $S_0 := I/N$

For $t:=0$ to ∞

Compute smallest eigenvector v_t of $\nabla f(S_t)$

Compute step size α_t that minimizes $f(S_t+\alpha_t(v_t v_t^* - S_t))$

Set $S_{t+1} := S_t + \alpha_t(v_t v_t^* - S_t)$

Theorem (Hazan): This algorithm returns an ε-optimal solution after only $\log(m)/\varepsilon^2$ iterations.

Generalization to process tomography?

Beyond the Bayesian and Max-Lik creeds: a third way?

We're not assuming any prior over states

Removes a lot of problems!

Extension to k-outcome measurements?

Sure, if we increase the number of sample measurements m by a poly(k) factor

Note that there's no hope of learning to simulate 2^n-outcome measurements (i.e. measurements on all n qubits) after poly(n) sample measurements

Custom Convex Programming Method
[E. Hazan, 2008]

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

Set S_0 := I/N

For t:=0 to ∞

Compute smallest eigenvector v_t of $\nabla f(S_t)$

Compute step size α_t that minimizes $f(S_t+\alpha_t(v_t v_t^*-S_t))$

Set $S_{t+1} := S_t + \alpha_t(v_t v_t^*-S_t)$

Theorem (Hazan): This algorithm returns an ε-optimal solution after only $\log(m)/\varepsilon^2$ iterations.

Implementation
[A. & Dechter 2008]

We implemented Hazan's algorithm in MATLAB

Code available on request

Using MIT's computing cluster, we then did numerical simulations to check experimentally that the learning theorem is true

Experiments We Ran

Implementation
[A. & Dechter 2008]

We implemented Hazan's algorithm in MATLAB

Code available on request

Using MIT's computing cluster, we then did numerical simulations to check experimentally that the learning theorem is true

Experiments We Ran

Experiments We Ran

1. **Classical States (sanity check).** States have form $\rho = |x\rangle\langle x|$, measurements check if i^{th} bit is 1 or 0, distribution over measurements is uniform.

Experiments We Ran

1. **Classical States (sanity check).** States have form $\rho=|x\rangle\langle x|$, measurements check if i^{th} bit is 1 or 0, distribution over measurements is uniform.

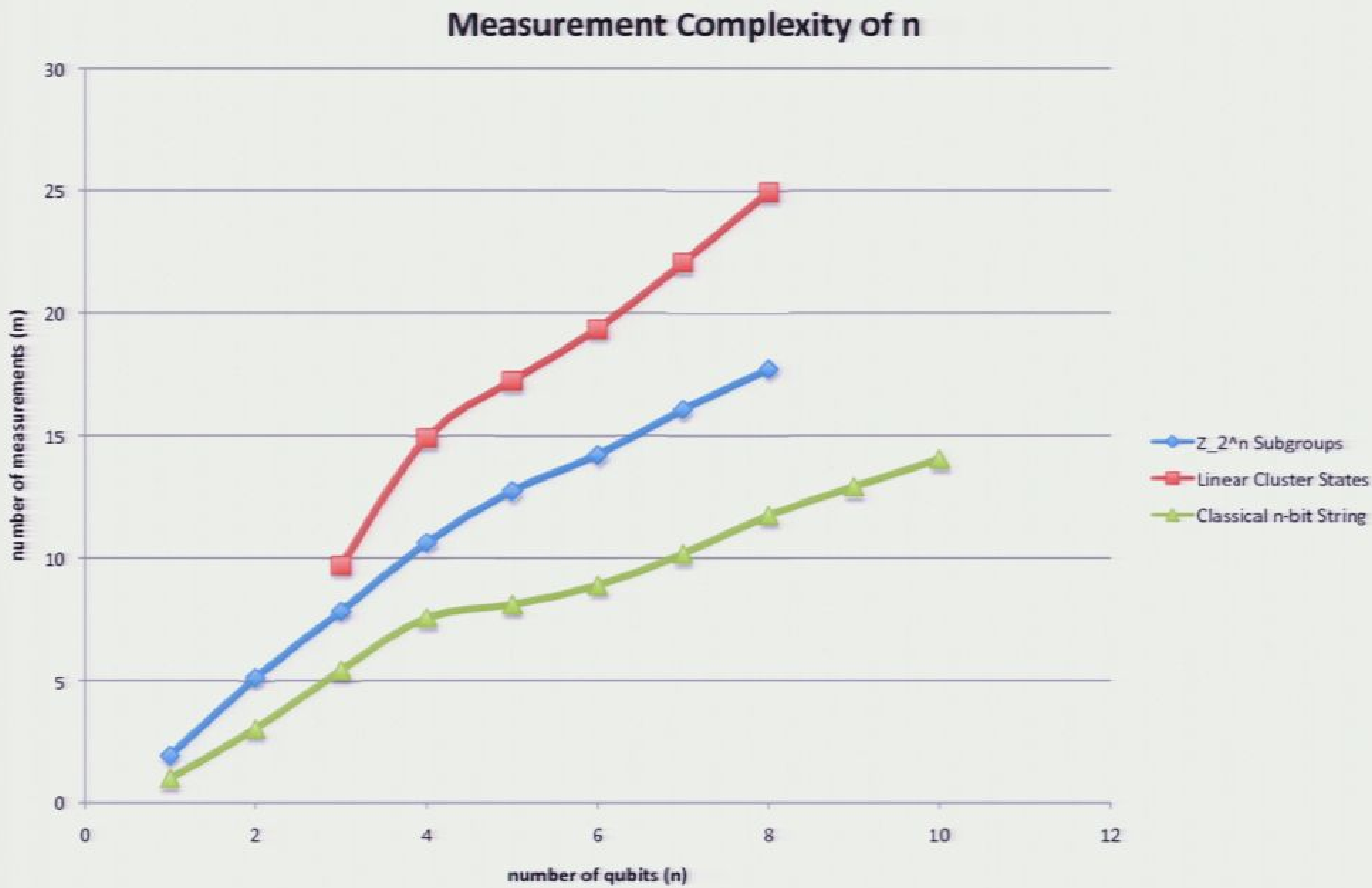
2. **Linear Cluster States.** States are n qubits, prepared by starting with $|+\rangle^{\otimes n}$ and then applying conditional phase ($P|xy\rangle=(-1)^{xy}|xy\rangle$) to each neighboring pair. Measurements check three randomly-chosen neighboring qubits, in a basis like $\{|0\rangle|+\rangle|0\rangle,|1\rangle|+\rangle|1\rangle,|0\rangle|-\rangle|1\rangle\}$. Acceptance probability is always ¾.

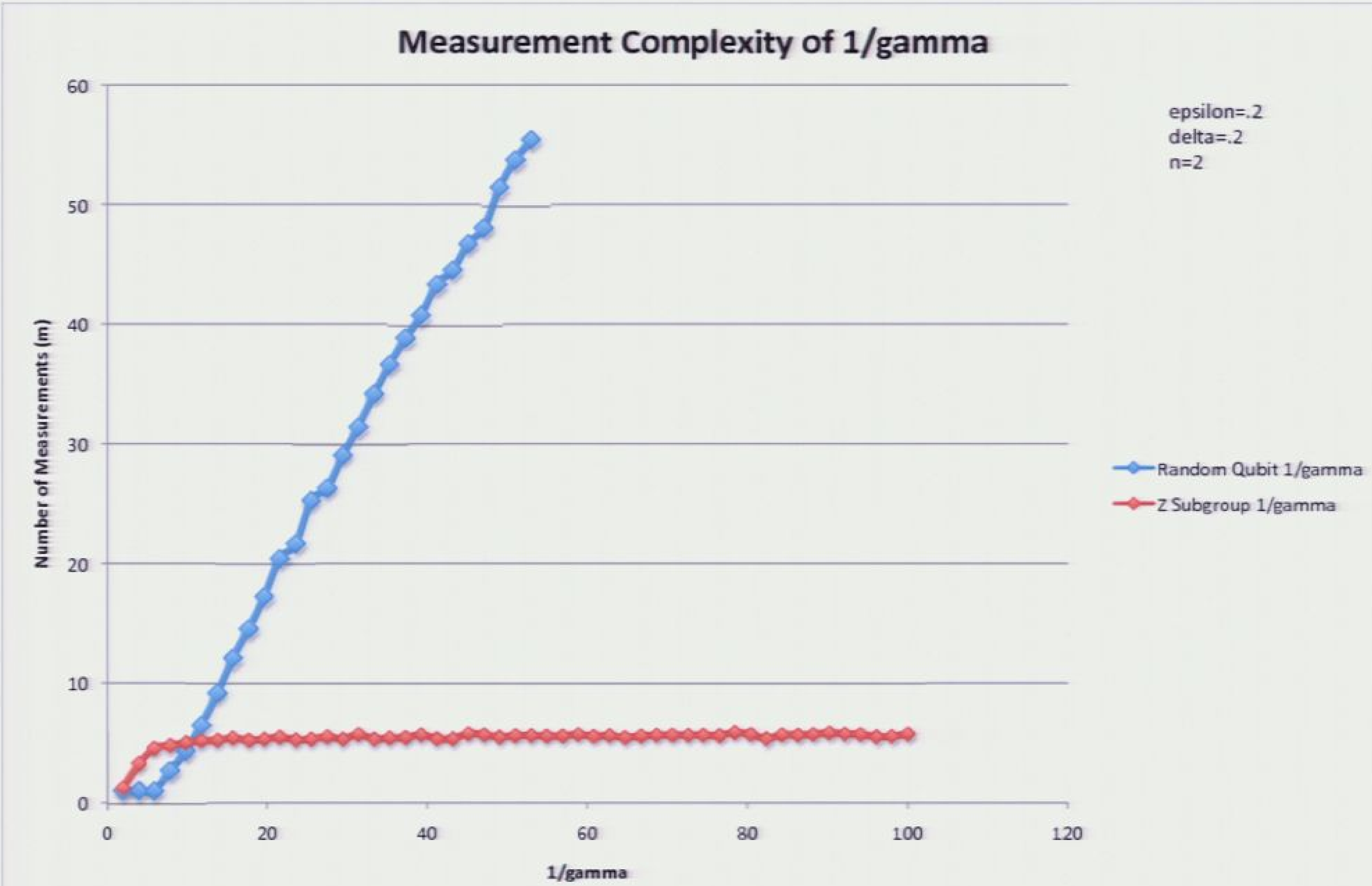
3. Z_2^n Subgroup States. Let H be a subgroup of $G=Z_2^n$ of order 2^{n-1}. States $\rho=|H\rangle\langle H|$ are equal superpositions over H. There's a measurement E_g for each element $g\in G$, which checks whether $g\in H$:

$$E_g = \frac{1}{2} I_n + \frac{1}{4} U_g + \frac{1}{4} U_g^*$$

where $U_g|h\rangle=|gh\rangle$ for all $h\in G$. E_g accepts with probability 1 if $g\in H$, or ½ if $g\notin H$.

Inspired by [Watrous 2000]; meant to showcase pretty-good tomography with non-commuting measurements.


Measurement Complexity of n
number of measurements (m)
number of qubits (n)
0
5
10
15
20
25
30
0
2
4
6
8
10
12
Z_2^n Subgroups
Linear Cluster States
Classical n-bit String


Measurement Complexity of 1/gamma
epsilon=.2
delta=.2
n=2
Number of Measurements (m)
1/gamma
Random Qubit 1/gamma
Z Subgroup 1/gamma
0
10
20
30
40
50
60
0
20
40
60
80
100
120

Open Problems

Open Problems

Find more convincing applications of our learning theorem

Open Problems

Find more convincing applications of our learning theorem

Find special classes of states for which learning can be done using **computation** time polynomial in the number of qubits

Open Problems

Find more convincing applications of our learning theorem

Find special classes of states for which learning can be done using **computation** time polynomial in the number of qubits

Improve the parameters of the learning theorem

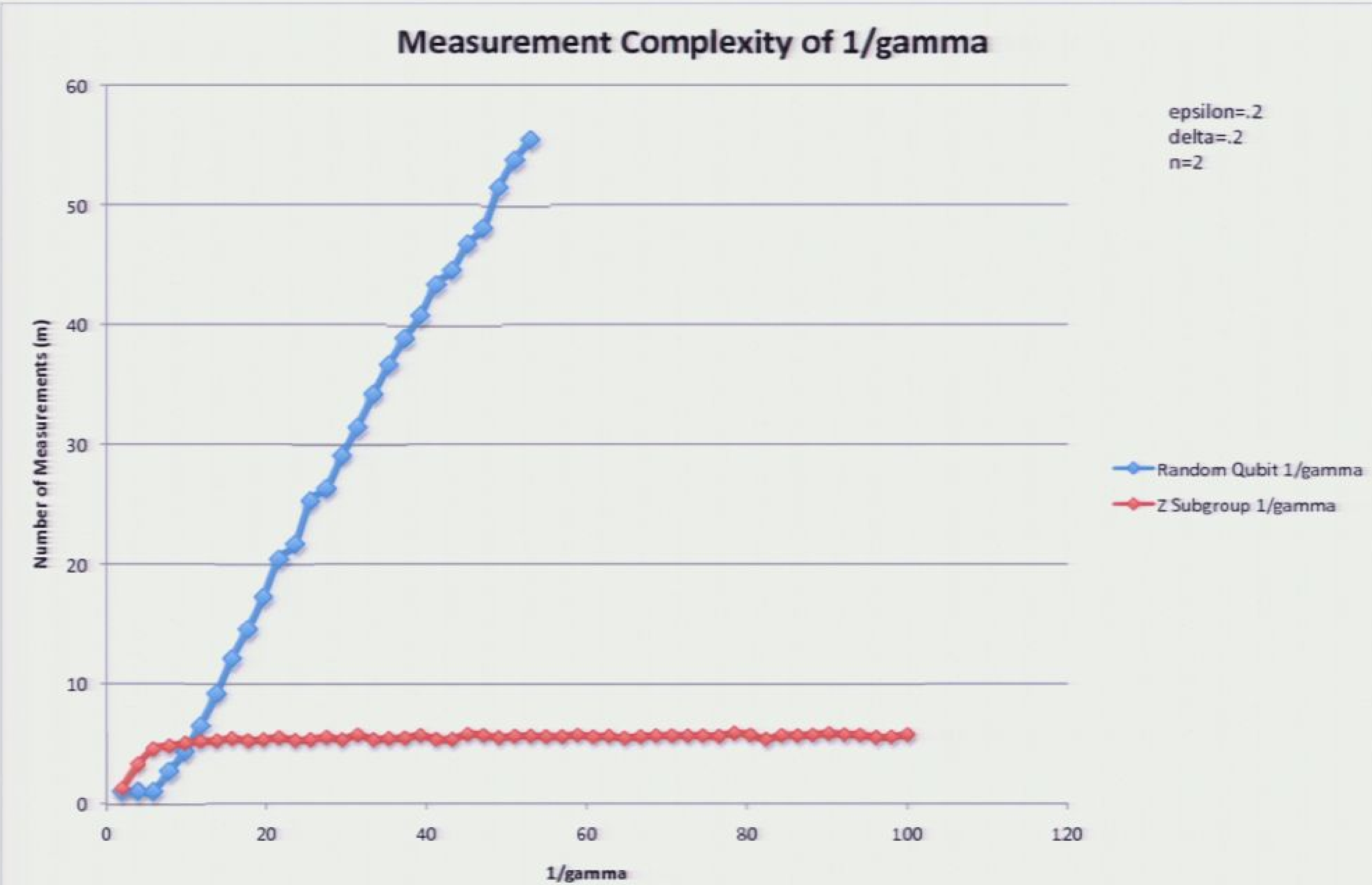
Open Problems

Find more convincing applications of our learning theorem

Find special classes of states for which learning can be done using **computation** time polynomial in the number of qubits

Improve the parameters of the learning theorem

Experimental demonstration!


Measurement Complexity of 1/gamma
Number of Measurements (m)
1/gamma
0
10
20
30
40
50
60
0
20
40
60
80
100
120
epsilon=.2
delta=.2
n=2
Random Qubit 1/gamma
Z Subgroup 1/gamma