SIMULATING QUANTUM DYNAMICS WITH WITH NEURAL MACHINE TRANSLATION

Juan Carrasquilla

Vector Institute

Canada CIFAR AI chair

IPAM From Passive to Active: Generative and Reinforcement Learning with Physics workshop

Los Angeles, CA Sept. 27th 2019

VECTOR INSTITUTE | INSTITUT VECTEUR CIFAR NSERC CRSNG UNIVERSITY OF WATERLOO

COLLABORATORS

Di Luo (U. Illinois-Urbana-Champaign)

Felipe Pérez (Layer6 AI)

Maksims Volkovs (Layer6 AI)

Leandro Aolita (Universidade Federal do Rio de Janeiro, ICTP-SAIFR)

FEYNMAN 1981:

Simulating Physics with Computers

Richard P. Feynman

Department of Physics, California Institute of Technology, Pasadena, California 91107

Received May 7, 1981

1. INTRODUCTION

On the program it says this is a keynote speech—and I don't know what a keynote speech is. I do not intend in any way to suggest what should be in this meeting as a keynote of the subjects or anything like that. I have my own things to say and to talk about and there's no implication that anybody needs to talk about the same thing or anything like it. So what I want to talk about is what Mike Dertouzos suggested that nobody would talk about. I want to talk about the problem of simulating physics with computers and I mean that in a specific way which I am going to explain. The reason for doing this is something that I learned about from Ed Fredkin, and my entire interest in the subject has been inspired by him. It has to do with learning something about the possibilities of computers, and also something about possibilities in physics. If we suppose that we know all the physical laws perfectly, of course we don't have to pay any attention to computers. It's interesting anyway to entertain oneself with the idea that we've got something to learn about physical laws; and if I take a relaxed view here (after all I'm here and not at home) I'll admit that we don't understand everything.

The first question is, What kind of computer are we going to use to simulate physics? Computer theory has been developed to a point where it realizes that it doesn't make any difference; when you get to a *universal computer*, it doesn't matter how it's manufactured, how it's actually made. Therefore my question is, Can physics be simulated by a universal computer? I would like to have the elements of this computer *locally interconnected*, and therefore sort of think about cellular automata as an example (but I don't want to force it). But I do want something involved with the

Motivated and shaped the field of quantum computing which hopes to revolutionize computation through exploitation of quantum mechanical effects

FEYNMAN 1981:

Simulating Physics with Computers

Richard P. Feynman

Motivated the field of quantum computing

5. CAN QUANTUM SYSTEMS BE PROBABILISTICALLY SIMULATED BY A CLASSICAL COMPUTER?

Now the next question that I would like to bring up is, of course, the interesting one, i.e., Can a quantum system be probabilistically simulated by a classical (probabilistic, I'd assume) universal computer? In other words, a computer which will give the same probabilities as the quantum system does. If you take the computer to be the classical kind I've described so far, (not the quantum kind described in the last section) and there're no changes in any laws, and there's no hocus-pocus, the answer is certainly, No! This is called the hidden-variable problem: it is impossible to represent the results of quantum mechanics with a classical universal device. To learn a little bit about it, I say let us try to put the quantum equations in a form as close as

This is all still true today and is **fundamentally** linked to the notion of quantum speed-up in quantum computing.

Feynman concludes:

the great difficulty. The only difference between a probabilistic classical world and the equations of the quantum world is that somehow or other it appears as if the probabilities would have to go negative, and that we do not know, as far as I know, how to simulate. Okay, that's the fundamental problem. I don't know the answer to it, but I wanted to explain that if I try my best to make the equations look as near as possible to what would be imitable by a classical probabilistic computer, I get into trouble.

WE CAN'T SIMULATE QUANTUM STATES AND DYNAMICS EXACTLY USING PROBABILITY, BUT CAN WE DO IT APPROXIMATELY?

TEXT GENERATION USING NEURAL LANGUAGE MODELS

SYSTEM PROMPT (HUMAN-WRITTEN)

In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.

MODEL COMPLETION (MACHINE-WRITTEN, 10 TRIES)

The scientist named the population, after their distinctive horn, Ovid's Unicorn. These four-horned, silver-white unicorns were previously unknown to science.

Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved.

Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow.

Pérez and the others then ventured further into the valley. "By the time we reached the top of one peak, the water looked blue, with some crystals on top," said Pérez.

Pérez and his friends were astonished to see the unicorn herd. These creatures could be seen from the air without having to move too much to see them – they were so close they could touch their horns.

While examining these bizarre creatures the scientists discovered that the creatures also spoke some fairly regular English. Pérez stated, "We can see, for example, that they have a common 'language,' something like a dialect or dialectic."

Dr. Pérez believes that the unicorns may have originated in Argentina, where the animals were believed to be descendants of a lost race of people who lived

Prompt

Output

Pretty coherent over a page of text

https://openai.com/blog/better-language-models/

TEXT GENERATION USING NEURAL LANGUAGE MODELS

https://talktotransformer.com/

Prompt

quantum supremacy has been seen in a series of experiments carried out at the Joint Quantum Institute in Cambridge and at the University of California, Santa Barbara. There, for instance, the team has shown that quantum entanglement, a phenomenon that allows information to be transmitted without the need for a remote link, is a robust, useful feature of the system, and that experiments with multiple quantum processors could generate high-resolution measurements of the quantum states of a single quantum processor.

Quantum effects are fundamental to modern science and technology, including quantum teleportation, which allows quantum states of atoms to be transmitted without the need for inter-atomic distance.

But, as scientists at the National Physical Laboratory in the UK and the Joint Quantum Institute reported this week, it is possible to exploit that effect from different quantum processors, which could improve the reliability of the measurement and the accuracy of the results of the experiments. "Quantum entanglement does not happen very often, but it has great implications if you understand and manipulate it correctly," says Professor Mark Gassaway, head of the physics department at the NI

Output

generating samples from a variety of inputs —>close to human quality and long-range coherence over a **page or more of text**

Reproducing correctly strong correlations present in language (power law)

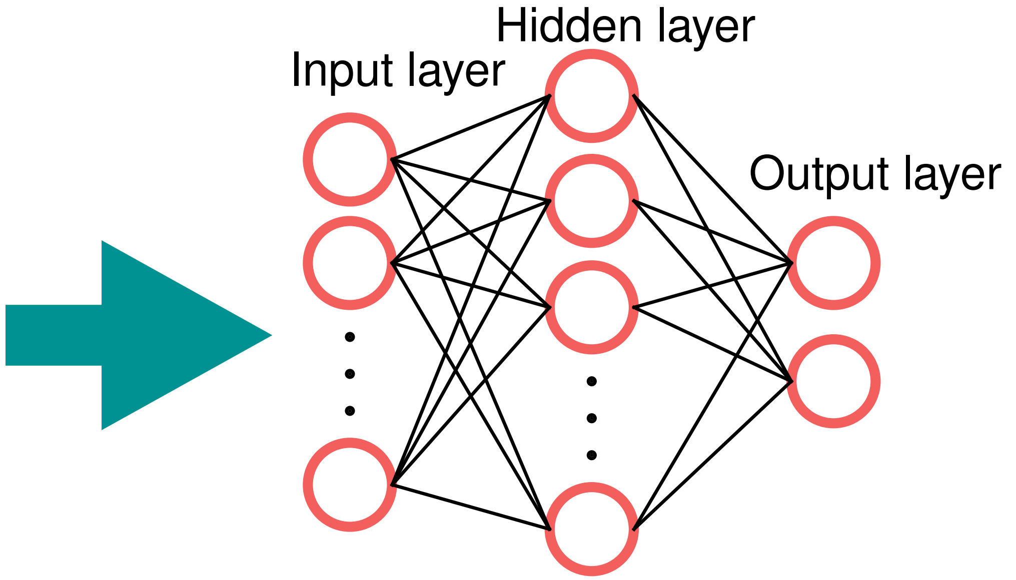
These techniques are based on powerful neural probabilistic language models

NEURAL LANGUAGE MODELS

- A neural network language model is a language model based on neural networks
- Neural networks are powerful universal function approximators and can in principle compute any function
- It is natural to ask whether these models represent complex quantum states

$$P(\text{output text}|\text{input text})$$

$$P(\text{English}|\text{Spanish})$$



DIMENSIONALITY OF QUANTUM SYSTEMS VS NEURAL MACHINE TRANSLATION

$|\Psi\rangle$ *vector with* 2^N

- Today's best supercomputers can solve the wave equation **exactly** for systems with a maximum of ~45 spins.

$$2^N \sim 3.5 \times 10^{13}$$

- Language translation models live in very high dimensional spaces too (example from "Attention is all you need")

$$\text{Vocab. Size}^{\text{Max length of sentence}}$$

$$8000^{100} \sim 2.03 \times 10^{390}$$

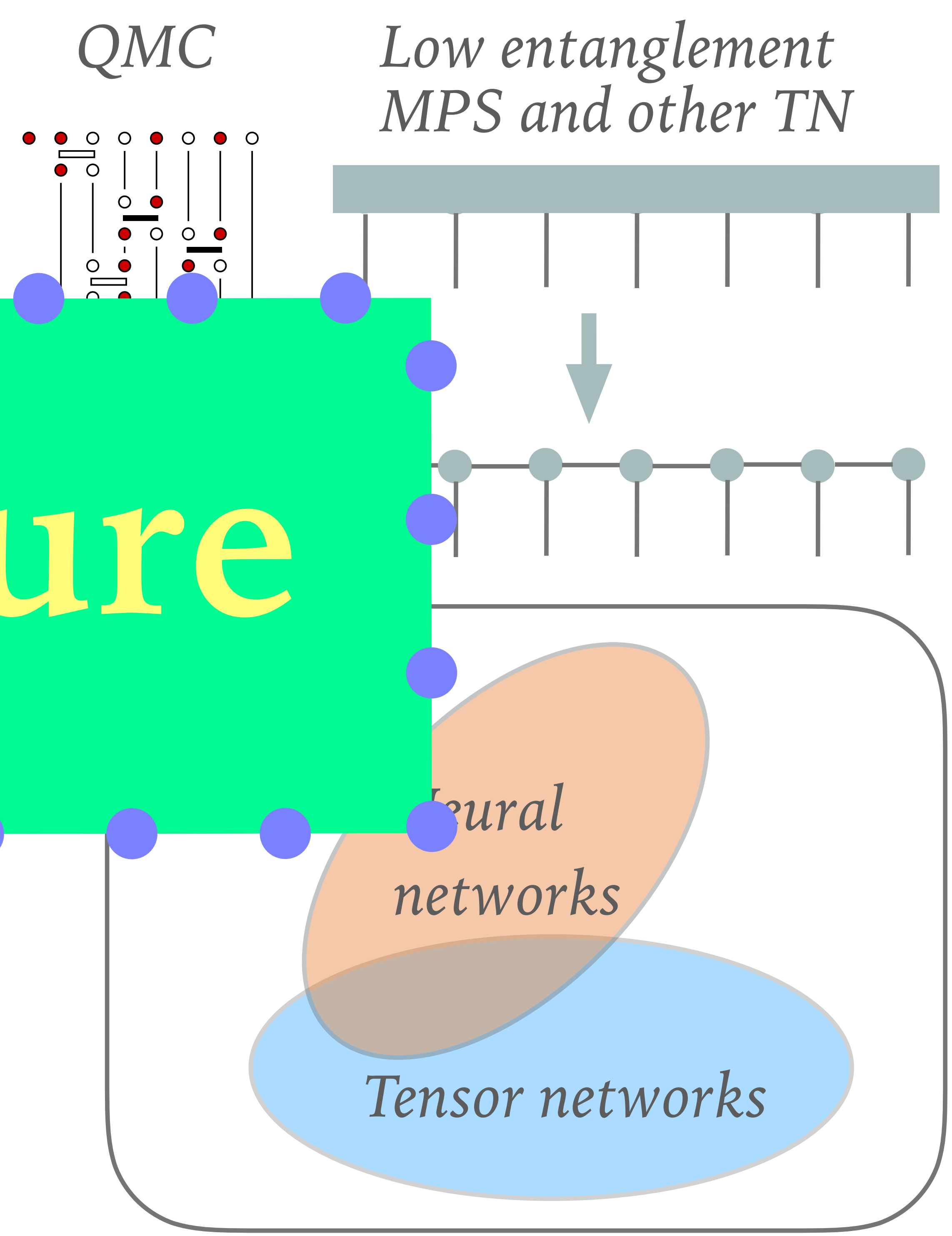
Storage of these distributions requires a computer with a memory which exceeds in size a number of universes bigger than there are atoms in the known universe.

QUANTUM STATES, NATURAL IMAGES, NATURAL LANGUAGES ARE "PHYSICAL"

- The amount of informati[...] for quantum states, lan[...]age modelling, computer vis[...]s smaller than the max[...]um capacity[...]
- Quantu[...] numeri[...] Tensor N[...]
- Both qua[...] [...] have a lot of (**shared**) structure and symmetry that we can exploit

Structure

$|\Psi\rangle$ *vector with* 2^N

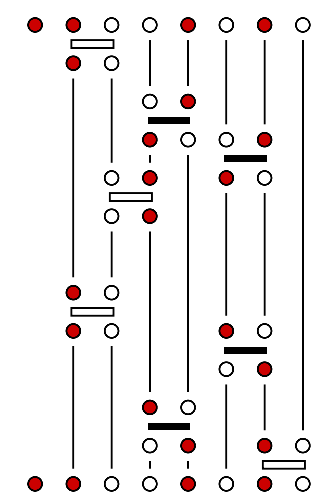


QUANTUM STATES, NATURAL IMAGES, NATURAL LANGUAGES ARE "PHYSICAL"

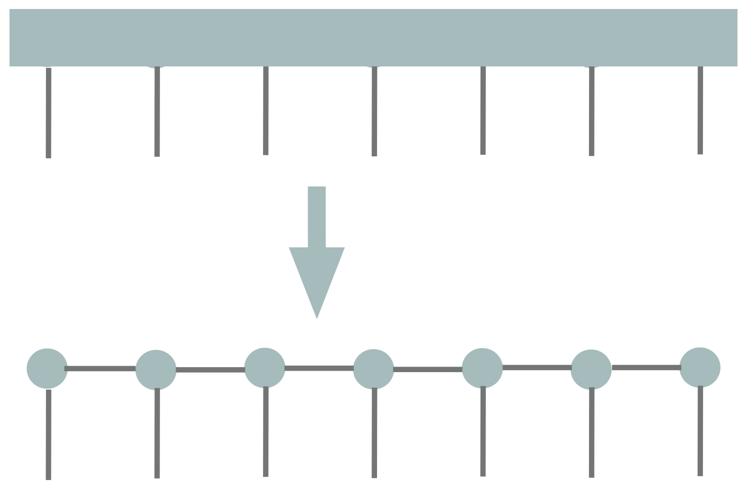
- The amount of information for quantum states, language modelling, computer vision, is smaller than the maximum capacity
- Quantum Monte Carlo and other numerical methods based on Tensor Networks exploit this fact.
- Both quantum and ML problems have a lot of (**shared**) structure and symmetry that we can exploit

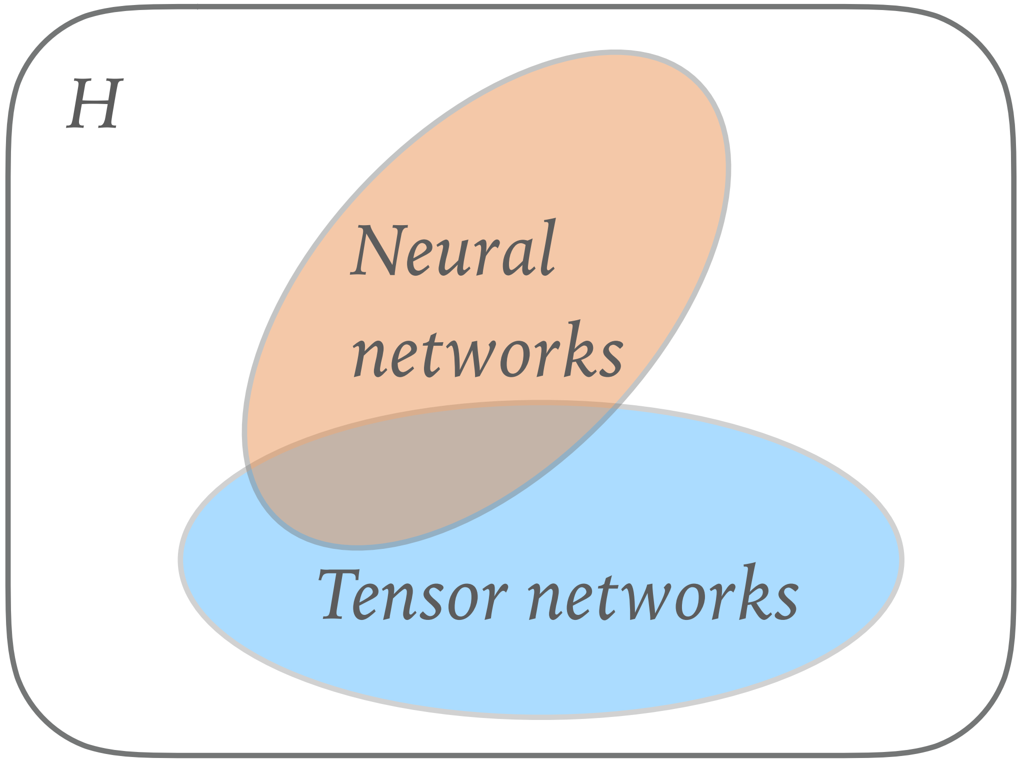
$|\Psi\rangle$ *vector with* 2^N

QMC



Low entanglement MPS and other TN





CAN WE USE THE POWER OF GENERATIVE MODELS FOR THE SIMULATION OF QUANTUM SYSTEMS?

CAN WE BRING QUANTUM THEORY CLOSER TO MACHINE LEARNING?

IN THIS TALK

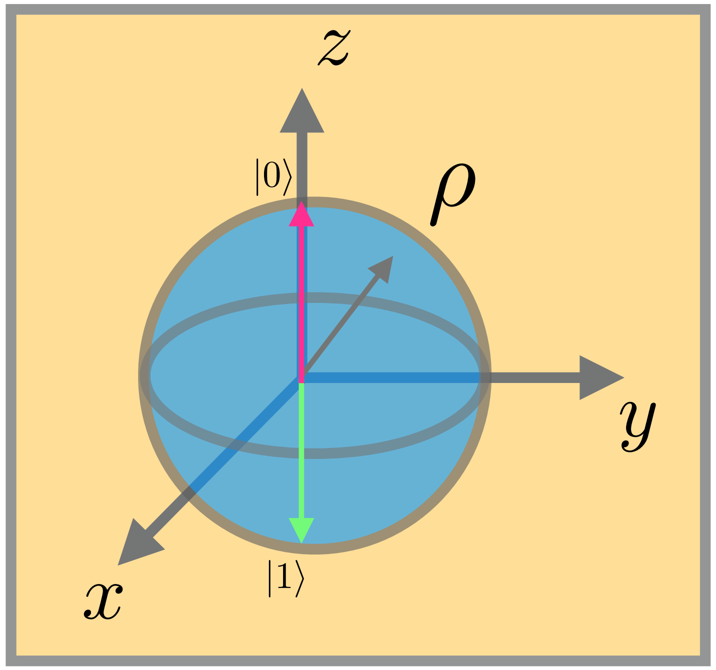
- I will introduce a formulation of quantum theory that brings quantum mechanics and machine learning close to each other
- We use generative models, in particular RNNs and transformers to parametrize quantum states.
- I will show an example to motivate why this may be a good idea in the context of quantum state reconstruction
- I will show you a heuristic to simulate a quantum circuits

QUANTUM STATES, MEASUREMENTS, AND PROBABILITY DISTRIBUTIONS

HOW IS A QUANTUM STATE TRADITIONALLY DESCRIBED?

- A **density matrix** describes the statistical state of a system in quantum mechanics. Everything we can possibly know about a quantum system is encoded in the density matrix.
- A quantum state is a positive semidefinite, Hermitian operator of trace 1 acting on the state space.
- The family of quantum states forms a convex set. For one qubit: Bloch sphere.

$$\rho$$



HOW TO REPRESENT A QUANTUM STATE WITH ONLY PROBABILITY?

MEASUREMENTS: POSITIVE OPERATOR-VALUED MEASURE (POVM)

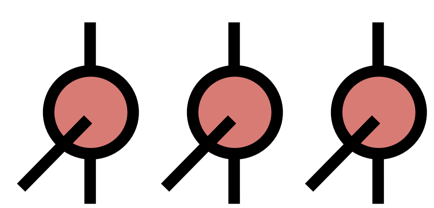
- POVM elements $\mathbf{M} = \{M^{(a)} \mid a \in \{1, ..., m\}\}$
- Positive semidefinite operators $\sum_i M^{(a)} = \mathbb{1}$
- Born Rule $P(\mathbf{a}) = \mathrm{Tr}\, \rho M^{\mathbf{a}}$

INFORMATIONALLY COMPLETE POVM

- The measurement statistics $P(\mathbf{a})$ contains all of the information about the state.
- Relation between ρ and distribution $P(\mathbf{a})$ can be inverted.

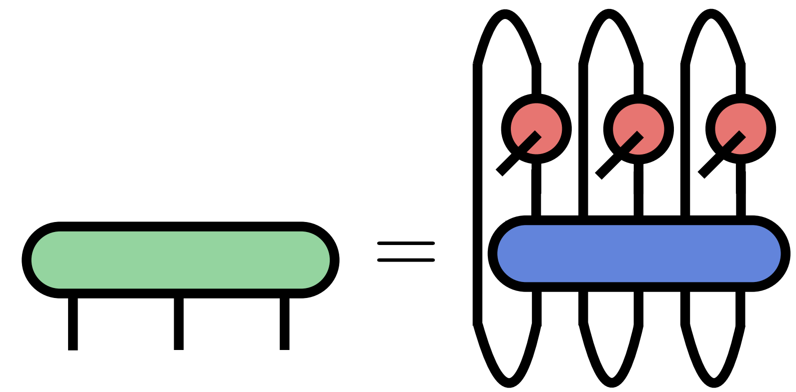
TAKE A SINGLE QUBIT POVM AND MAKE A TENSOR PRODUCT

$$\boldsymbol{M} = \left\{M^{(a_1)} \otimes M^{(a_2)} \otimes \ldots M^{(a_N)}\right\}_{a_1,\ldots a_N}$$



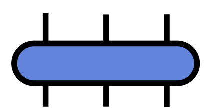
GRAPHICAL NOTATION AND INVERSE

Born rule $P(\mathbf{a}) = \mathrm{Tr}\, \rho M^{\mathbf{a}}$



If the POVM is informationally complete then

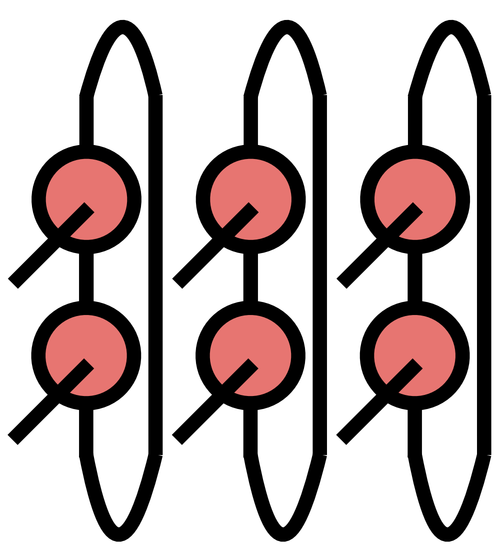
$$\rho = \sum_a O_\rho(a) M^{(a)}$$



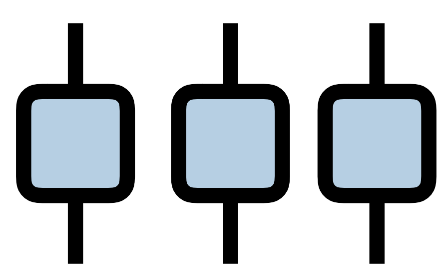
Insert this relation into Born's rule $P(a) = \sum_{a'} O_\rho(a') Tr[M^{(a)} M^{(a')}] = \sum_{a'} O_\rho(a') T_{a'a}$

$$\rho = \sum_{a,a'} T^{-1}_{a,a'} P(a') M^{(a)}$$

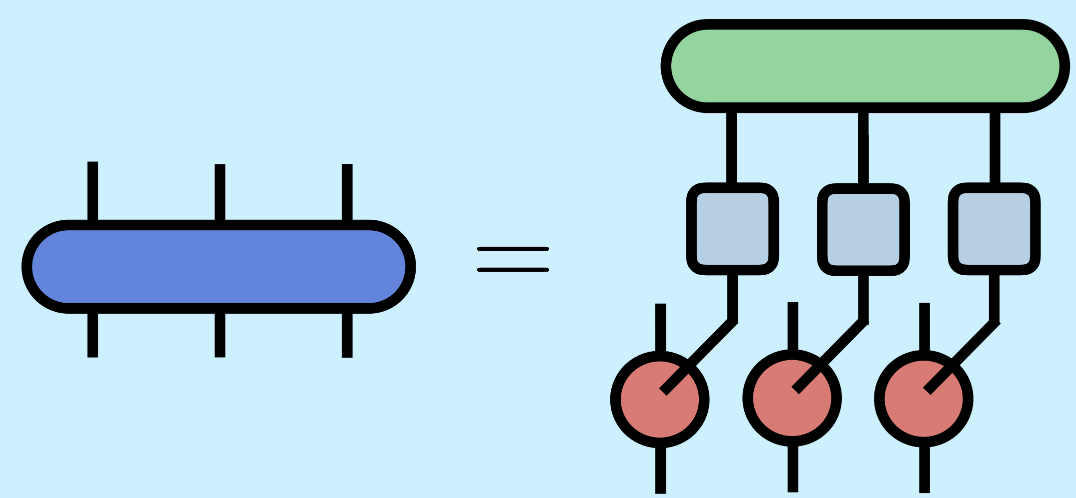
$$T_{\alpha,\beta} = \mathrm{Tr}\ M^\alpha M^\beta$$



$\mathbf{T}^{-1}$

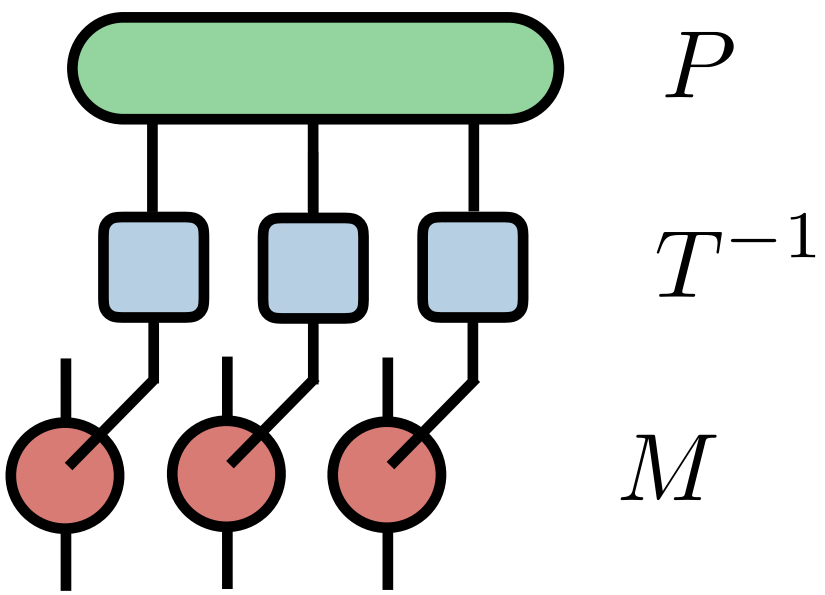


$$\rho = \sum_{a,a'} T^{-1}_{a,a'} P(a') M^{(a)}$$



REPRESENTATION THE QUANTUM STATE

$$\rho = (T^{-1}P)^T M$$



- Factorization of the state in terms of a probability distribution and a set of tiny tensors —> Wavefunction positivization
- All the **entanglement** and potential complexity of the state comes from the structure of the P(a)
- Very efficient to handle numerically for some tasks
- Sign structure of the state is in the tiny tensors

Carrasquilla, Torlai, Melko, Aolita. Nature Machine Intelligence 1, 200 (2019)

INSIGHT: PARAMETRIZE STATISTICS OF MEASUREMENTS AND INVERT

$$P(\mathbf{a}) = \mathrm{Tr}\, \rho M^{\mathbf{a}}$$

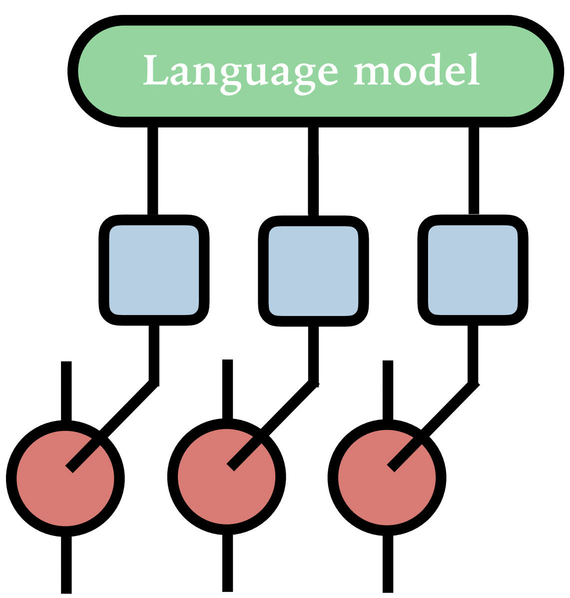
=> Create a neural model of P(a)

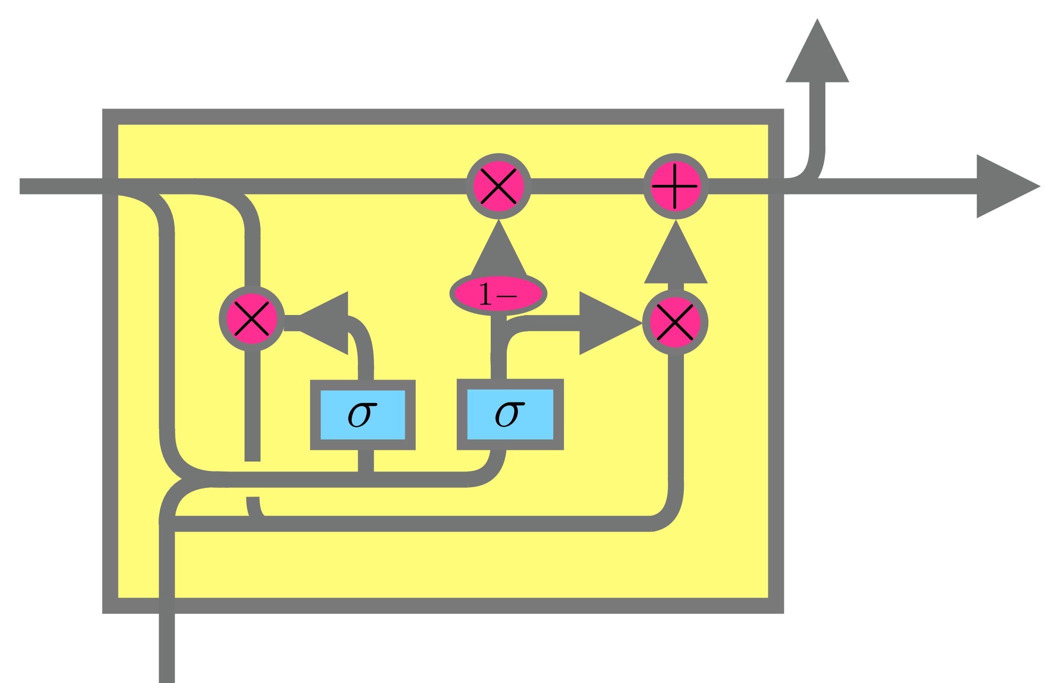
Autoregressive models (RNNs and transformer)

1. Allow for exact sampling
2. Tractable density $P_{\mathrm{model}}(\mathbf{a})$
3. Traditionally used in neural machine translation

$P_{\mathrm{model}}(\mathbf{a})$ →

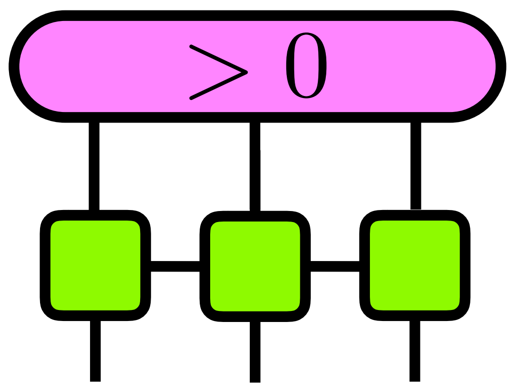
$\rho_{\mathrm{model}} =$





REMARK: WAVEFUNCTION POSITIVZATION WITH AD

$$|\Psi\rangle \quad = $$



Wavefunction positivization via automatic differentiation

Giacomo Torlai, Juan Carrasquilla, Matthew T. Fishman, Roger G. Melko, Matthew P. A. Fisher

(Submitted on 11 Jun 2019)

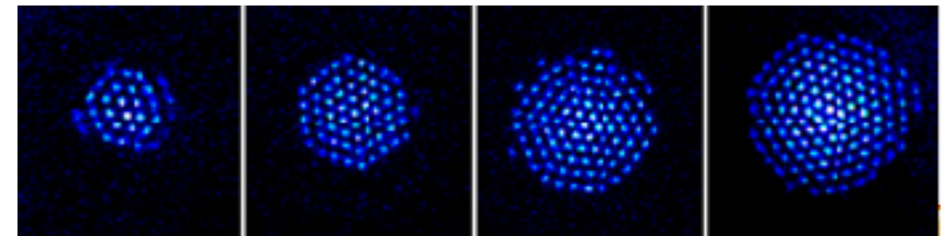
We introduce a procedure to systematically search for a local unitary transformation that maps a wavefunction with a non-trivial sign structure into a positive-real form. The transformation is parametrized as a quantum circuit compiled into a set of one and two qubit gates. We design a cost function that maximizes the average sign of the output state and removes its complex phases. The optimization of the gates is performed through automatic differentiation algorithms, widely used in the machine learning community. We provide numerical evidence for significant improvements in the average sign, for a two-leg triangular Heisenberg ladder with next-to-nearest neighbour and ring-exchange interactions. This model exhibits phases where the sign structure can be removed by simple local one-qubit unitaries, but also an exotic Bose-metal phase whose sign structure induces "Bose surfaces" wit
a fermionic character and a higher entanglement that requires two-qubit gates.

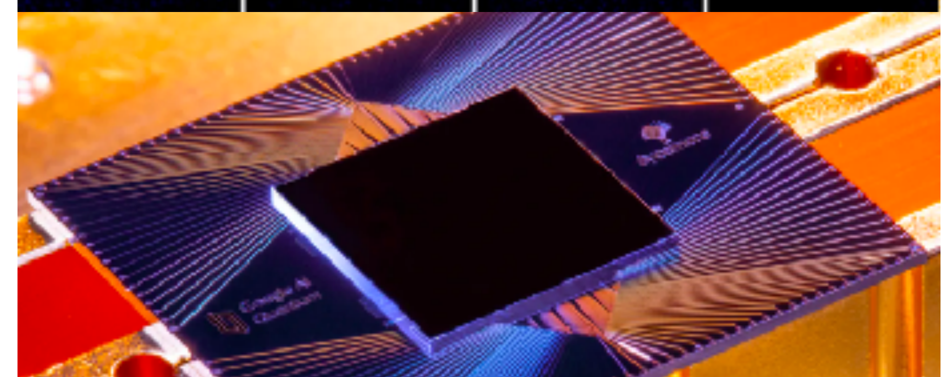
Comments: 9 pages, 5 figures
Subjects: **Quantum Physics (quant-ph)**; Strongly Correlated Electrons (cond-mat.str-el)
Cite as: arXiv:1906.04654 [quant-ph]
(or arXiv:1906.04654v1 [quant-ph] for this version)

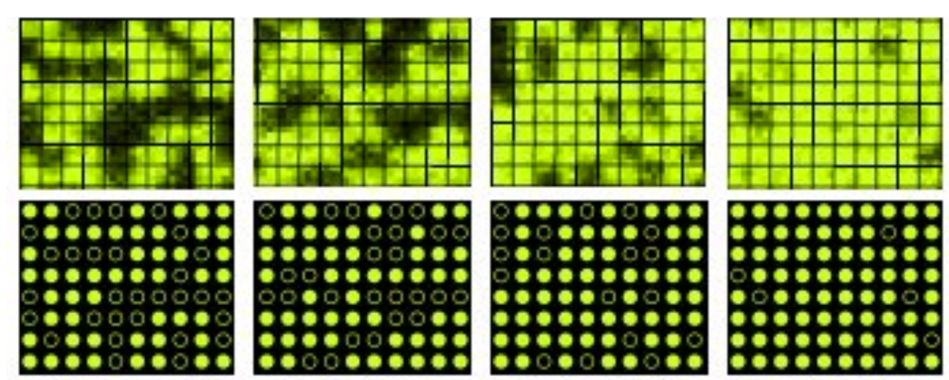
EXAMPLE: LEARN A QUANTUM STATE FROM MEASUREMENTS

NEED TO GO BEYOND STANDARD QUANTUM STATE TOMOGRAPHY

- Progress in controlling large quantum systems.
- Availability of arbitrary measurements performed with great accuracy.
- The bottleneck limiting progress in the estimation of states: curse of dimensionality.







SYNTHETIC QUANTUM DEVICES ARE GROWING FAST

nature.com > nature > articles > article

nature
International journal of science

Article | Published: 29 November 2017

Probing many-body dynamics on a 51-atom quantum simulator

Hannes Bernien, Sylvain Schwartz, Alexander Keesling, Harry Levine, Ahmed Omran, Hannes Pichler, Soonwon Choi, Alexander S. Zibrov, Manuel Endres, Markus Greiner, Vladan Vuletić & Mikhail D. Lukin

Nature **551**, 579–584 (30 November 2017) | Download Citation

nature.com > nature > letters > article

nature
International journal of science

Letter | Published: 29 November 2017

Observation of a many-body dynamical phase transition with a 53-qubit quantum simulator

J. Zhang, G. Pagano, P. W. Hess, A. Kyprianidis, P. Becker, H. Kaplan, A. V. Gorshkov, Z.-X. Gong & C. Monroe

Nature **551**, 601–604 (30 November 2017) | Download Citation

nature.com > nature > letters > article

nature
International journal of science

Letter | Published: 22 August 2018

Observation of topological phenomena in a programmable lattice of 1,800 qubits

Andrew D. King, Juan Carrasquilla, […] Mohammad H. Amin

Nature **560**, 456–460 (2018) | Download Citation

PHYSICAL REVIEW X

Highlights Recent Subjects Accepted Collections Authors Referees Search Press About Staff

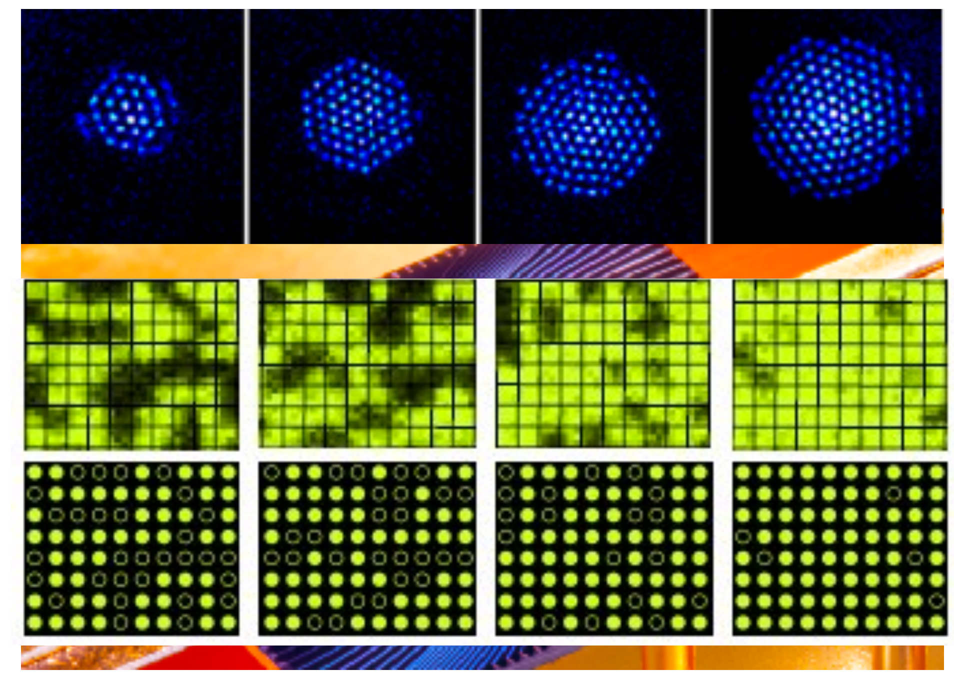
Open Access

Quantum Chemistry Calculations on a Trapped-Ion Quantum Simulator

Cornelius Hempel, Christine Maier, Jonathan Romero, Jarrod McClean, Thomas Monz, Heng Shen, Petar Jurcevic, Ben P. Lanyon, Peter Love, Ryan Babbush, Alán Aspuru-Guzik, Rainer Blatt, and Christian F. Roos

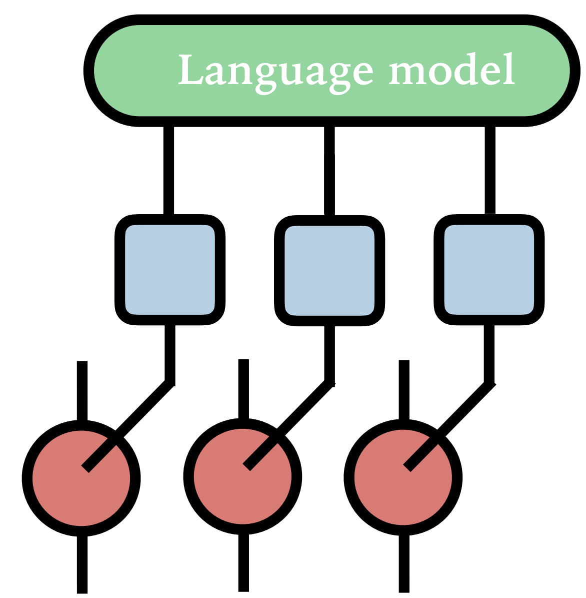
Phys. Rev. X **8**, 031022 – Published 24 July 2018

NEED TO GO BEYOND STANDARD QUANTUM STATE TOMOGRAPHY

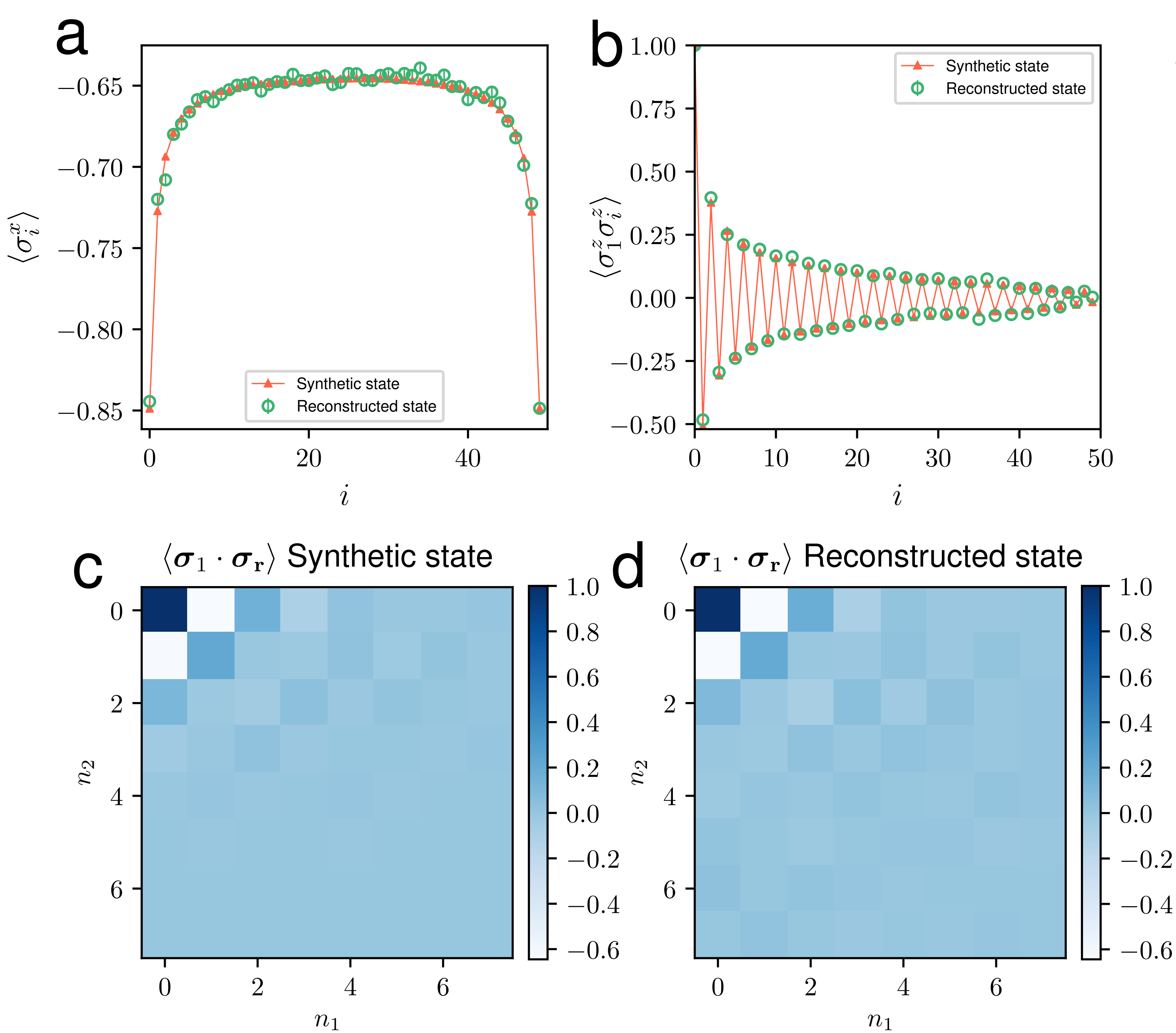


- Prepare an unknown quantum state
- Apply a measurement that probes enough information about the quantum state
- Repeat and collect the statistics of the measurement
- Infer a reconstruction of the state consistent with the measurement outcomes

$\rho^* =$



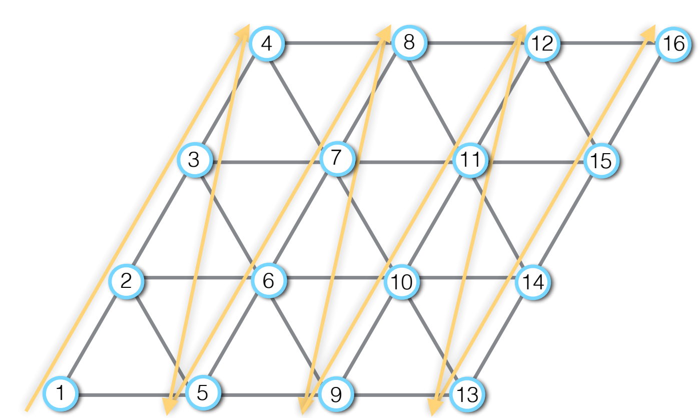
LEARNING GROUND STATES OF LOCAL HAMILTONIANS FROM DATA



$$\mathcal{H} = J\sum_{ij}\sigma_i^z\sigma_j^z + h\sum_i \sigma_i^x$$

N=50 spins. P(a) is a deep (3 layer GRU) recurrent neural network language model.

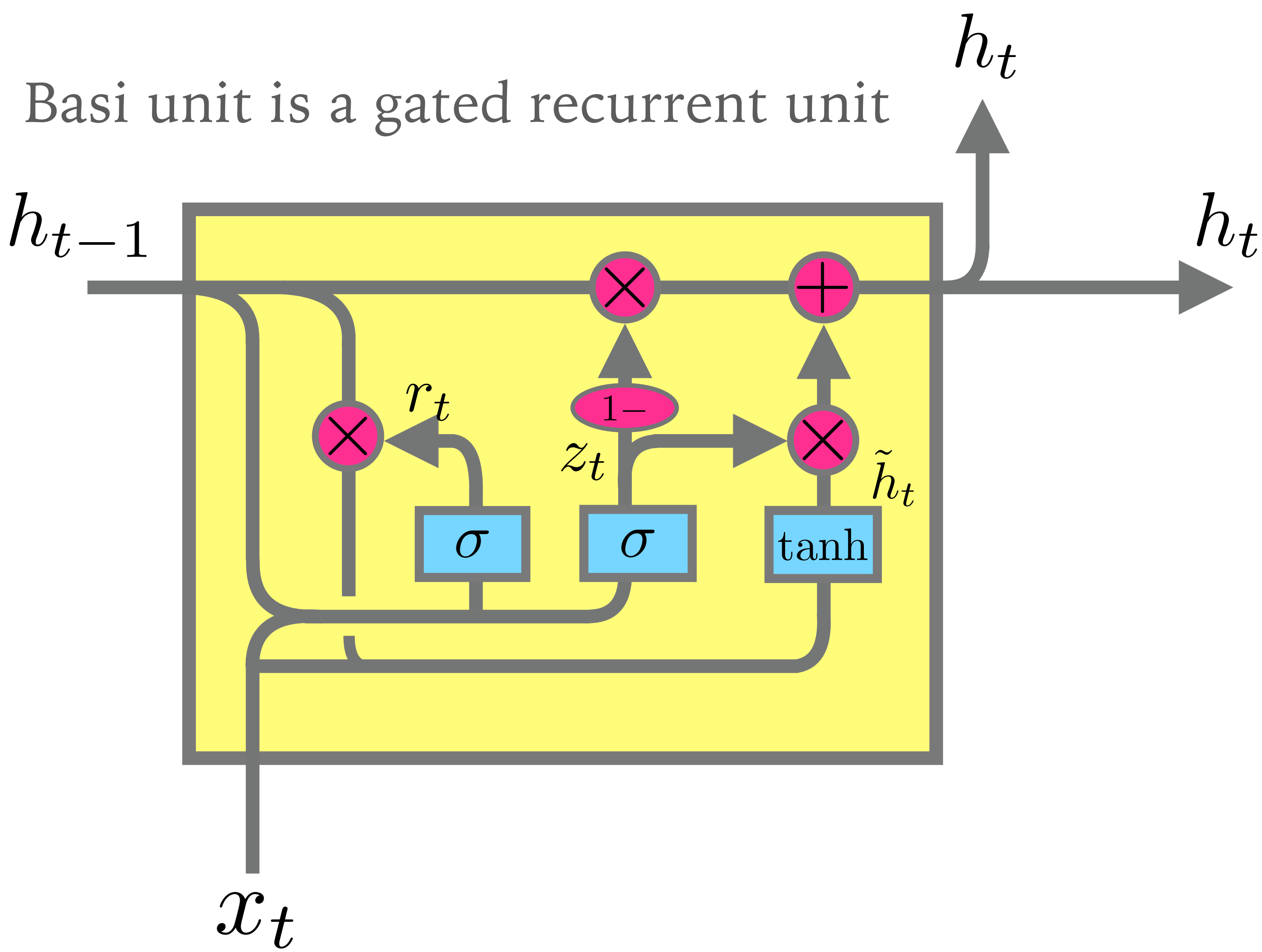
$$H = J\sum_{i,j}\boldsymbol{\sigma}_i.\boldsymbol{\sigma}_j$$



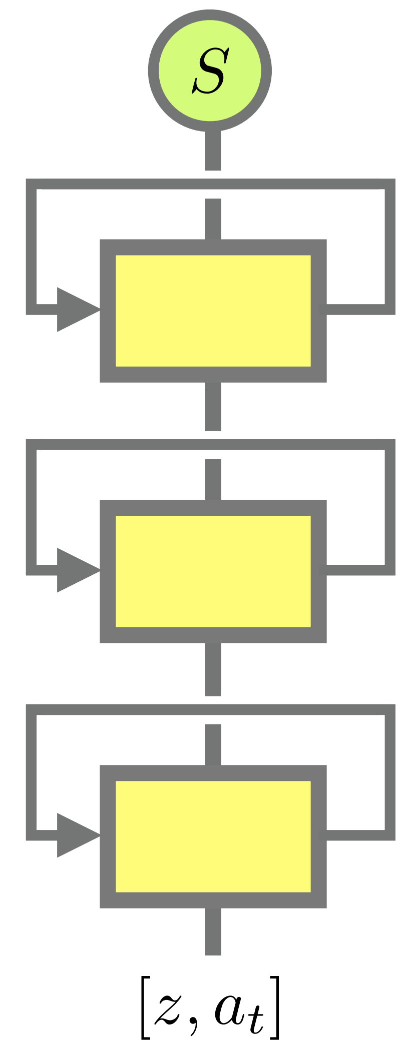
Carrasquilla, Torlai, Melko, Aolita. Nature Machine Intelligence 1, 200 (2019)

RECURRENT NEURAL NETWORK MODEL

Basi unit is a gated recurrent unit



Full model stacks three of these units and adds a softmax dense layer at each “time” step



BUT QUANTUM THEORY GOES BEYOND REPRESENTATION. DYNAMICS (E.G. SCHRODINGER EQUATION)? MEASUREMENTS?

UNITARY DYNAMICS AND QUANTUM CHANNELS

$$\varrho_U = U\varrho U^\dagger$$

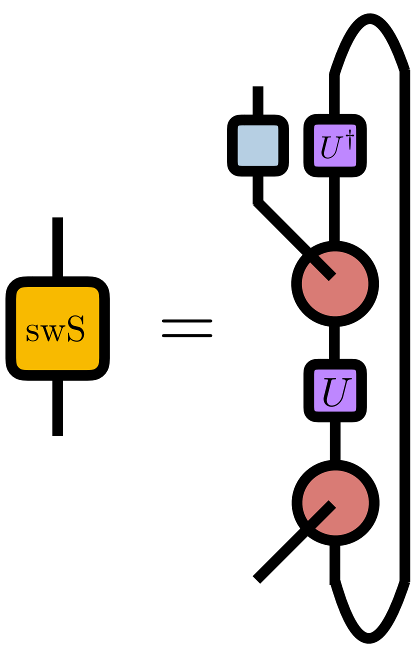
$$P_U(\boldsymbol{a}'') = \mathrm{Tr}\left[U\varrho U^\dagger M^{(\boldsymbol{a}'')}\right] = \sum_{\boldsymbol{a}'} O_{\boldsymbol{a}''\boldsymbol{a}'} P(\boldsymbol{a}')$$

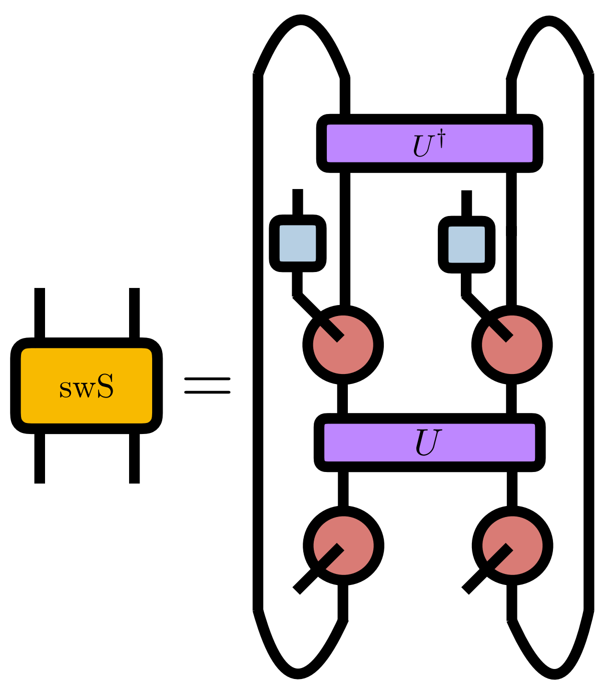
$$O_{\boldsymbol{a}''\boldsymbol{a}'} = \sum_{\boldsymbol{a}} \mathrm{Tr}\left[U M^{(\boldsymbol{a})} U^\dagger M^{(\boldsymbol{a}'')}\right] T^{-1}_{\boldsymbol{a},\boldsymbol{a}'}$$

Probabilistic gates: **Somewhat** (or quasi-) stochastic matrices

$$P_U(\boldsymbol{a}'') = \sum_{\boldsymbol{a}'} O_{\boldsymbol{a}''\boldsymbol{a}'} P(\boldsymbol{a}')$$

Evolution of probability is **somewhat** classical :)





If the starting unitaries are k-local, the swS matrices are also k-local

Somewhat stochastic matrices

Branko Ćurgus, Robert I. Jewett

(Submitted on 3 Sep 2007)

The standard theorem for regular stochastic matrices is generalized to matrices with no sign restriction on the entries. The condition that column sums be equal to 1 is kept, but the regularity condition is replaced by a condition on the ℓ_1-distances between columns.

UNITARY DYNAMICS AND QUANTUM CHANNELS

$$\varrho_U = U \varrho U^\dagger$$

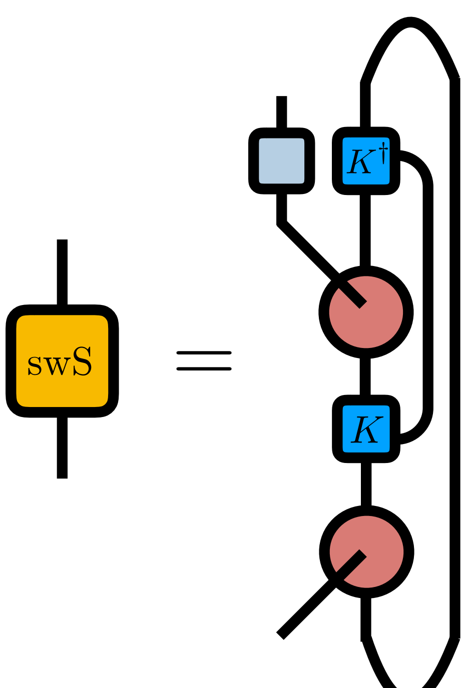
$$P_U\left(\boldsymbol{a}''\right) = \mathrm{Tr}\left[U \varrho U^\dagger M^{(\boldsymbol{a}'')}\right] = \sum_{\boldsymbol{a}'} O_{\boldsymbol{a}''\boldsymbol{a}'} P(\boldsymbol{a}')$$

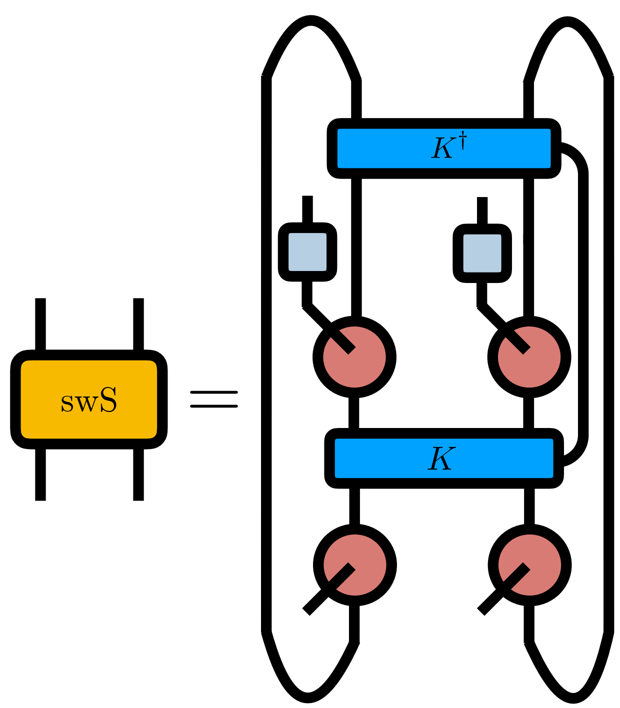
$$O_{\boldsymbol{a}''\boldsymbol{a}'} = \sum_{\boldsymbol{a}} \mathrm{Tr}\left[U M^{(\boldsymbol{a})} U^\dagger M^{(\boldsymbol{a}'')}\right] T^{-1}_{\boldsymbol{a},\boldsymbol{a}'}$$

Probabilistic gates: **Somewhat** (or quasi-) stochastic matrices

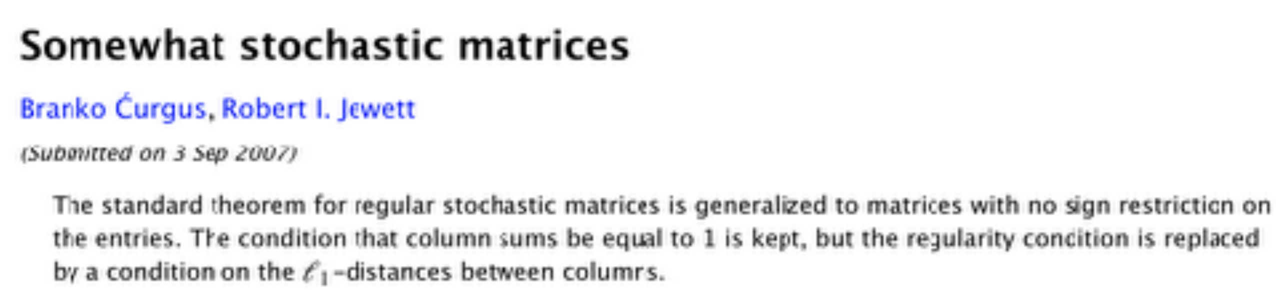
$$P_U\left(\boldsymbol{a}''\right) = \sum_{\boldsymbol{a}'} O_{\boldsymbol{a}''\boldsymbol{a}'} P(\boldsymbol{a}')$$

Evolution of probability is **somewhat** classical :)





If the starting unitaries are k-local, the swS matrices are also k-local



Somewhat stochastic matrices

Branko Ćurgus, Robert I. Jewett

(Submitted on 3 Sep 2007)

The standard theorem for regular stochastic matrices is generalized to matrices with no sign restriction on the entries. The condition that column sums be equal to 1 is kept, but the regularity condition is replaced by a condition on the ℓ_1-distances between columns.

QUANTUM DYNAMICS

Schrödinger equation

$$i\frac{\partial P\left(\boldsymbol{a}'',t\right)}{\partial t}=\sum_{\boldsymbol{a},\boldsymbol{a}'}\mathrm{Tr}\left(\left[\mathcal{H},M^{(\boldsymbol{a})}\right]M^{(\boldsymbol{a}'')}\right)T^{-1}_{\boldsymbol{a},\boldsymbol{a}'}P\left(\boldsymbol{a}',t\right)$$

BORN RULE

$$i\frac{\partial\rho}{\partial t}=[\mathcal{H},\rho]$$

"Solution"

$$\boldsymbol{P}(t)=e^{-iAt}\boldsymbol{P}(0)\qquad A_{\boldsymbol{a}''\boldsymbol{a}'}=\sum\nolimits_{\boldsymbol{a}}T^{-1}_{\boldsymbol{a},\boldsymbol{a}'}\left[\mathrm{Tr}\left(\left[\mathcal{H},M^{(\boldsymbol{a})}\right]M^{(\boldsymbol{a}'')}\right)\right]$$

QUANTUM DYNAMICS OF OPEN QUANTUM SYSTEMS

Linblad equation

$$i\frac{\partial P\left(\boldsymbol{a},t\right)}{\partial t} = \sum_{\boldsymbol{a}} A_{\boldsymbol{a}'',\boldsymbol{a}'} P\left(\boldsymbol{a},t\right)$$

$$\boldsymbol{P}(t) = e^{-iAt}\boldsymbol{P}(0)$$

$$\begin{aligned} A_{\boldsymbol{a}''\boldsymbol{a}'} = \sum_{\boldsymbol{a}} T^{-1}_{\boldsymbol{a},\boldsymbol{a}'} \Bigg(\mathrm{Tr}\left(\left[\mathcal{H}, M^{(\boldsymbol{a})}\right] M^{(\boldsymbol{a}'')}\right) \\ + \sum_k \left[-\frac{i}{2}\mathrm{Tr}\left(\{L_k^\dagger L_k, M^{(\boldsymbol{a})}\} M^{(\boldsymbol{a}'')}\right)\right. \\ \left. + \; i\,\mathrm{Tr}\left(L_k M^{(\boldsymbol{a})} L_k^\dagger M^{(\boldsymbol{a}'')}\right)\right]\Bigg) \end{aligned}$$

MEASUREMENTS

- Suppose we want to measure the quantum state. The measurement is described by some other POVM $\Pi^{(\boldsymbol{b})}$

$$P(\boldsymbol{b}) = \sum_{\boldsymbol{a},\boldsymbol{a}'} P(\boldsymbol{a}')\, T^{-1}_{\boldsymbol{a},\boldsymbol{a}'}\, \mathrm{Tr}\left[M^{(\boldsymbol{a})}\Pi^{(\boldsymbol{b})}\right] = \sum_{\boldsymbol{a}'} q(\boldsymbol{b}|\boldsymbol{a}')P(\boldsymbol{a}')$$

$$q(\boldsymbol{b}|\boldsymbol{a}') = \sum_{\boldsymbol{a}} T^{-1}_{\boldsymbol{a},\boldsymbol{a}'}\, \mathrm{Tr}\left[M^{(\boldsymbol{a})}\Pi^{(\boldsymbol{b})}\right]$$

- can be characterized as a **somewhat** conditional probability since its entries can either be positive or negative but its trace over b is the identity.
- evocative resemblance with the law of total probability—> **quantum law of total probability** in quantum Bayesianism.

CIRCUITS AND TENSOR NETWORKS IN OUR LANGUAGE

Quantum circuits and quantum computing

$$P_U(\mathbf{a}'') = \sum_{\mathbf{a}'} P(\mathbf{a}')O_{\mathbf{a}',\mathbf{a}''}$$

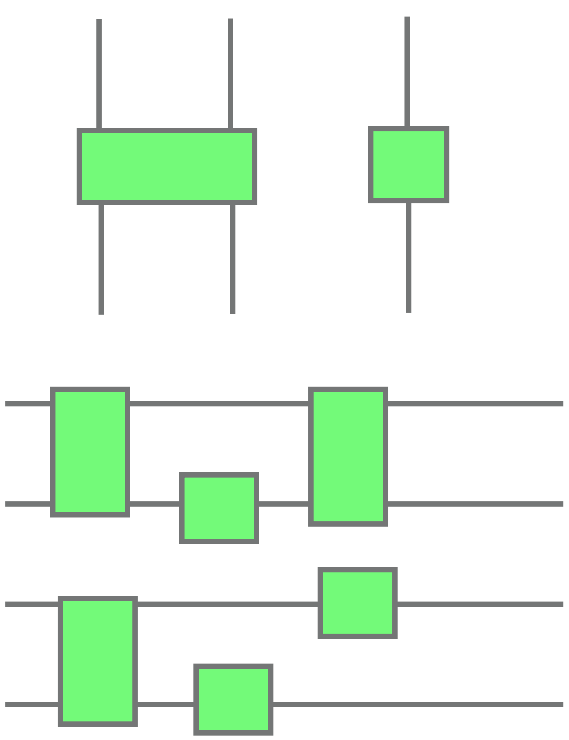
$$O_{\mathbf{a}'\mathbf{a}''} = \sum_{\mathbf{a}} \mathrm{Tr}(UM^{(\mathbf{a})}U^\dagger M^{(\mathbf{a}'')})T^{-1}_{\mathbf{a},\mathbf{a}'}$$

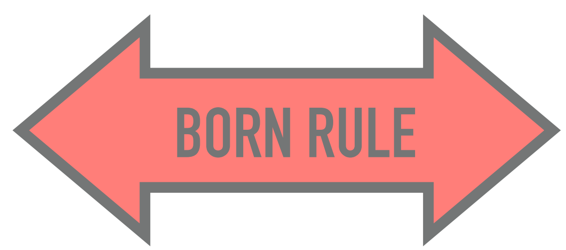
$$\rho_U = U\rho U^\dagger$$

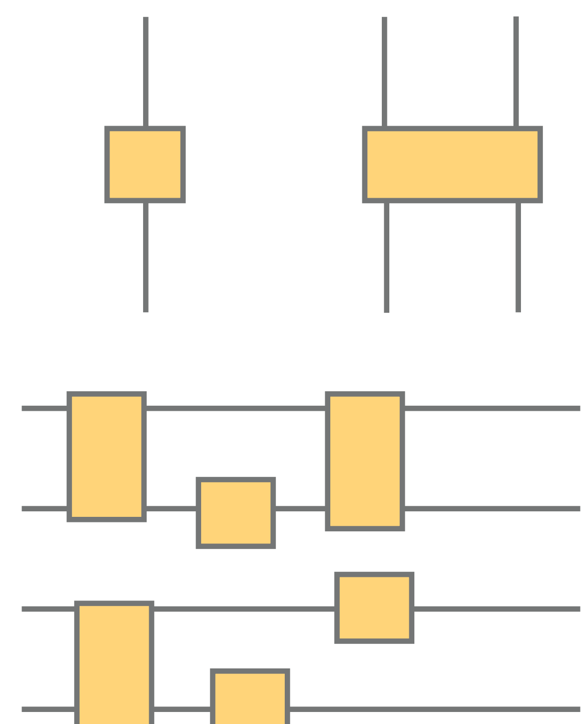
Unitary matrices U

Completely positive (CP) trace preserving map

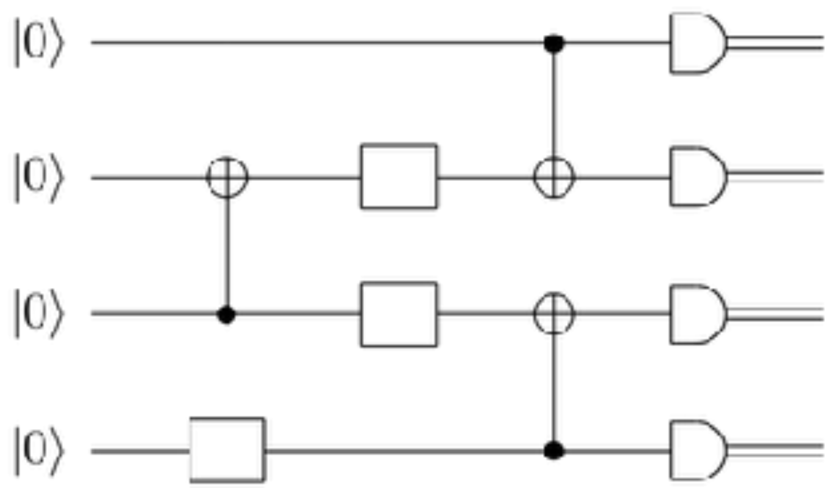
Tensor networks and quantum circuits







QUANTUM CIRCUIT



- Start the quantum device in a simple product state
- Apply a sequence of simple unitary matrices acting on the initial state
- To obtain the result, usually measure in the computational basis.

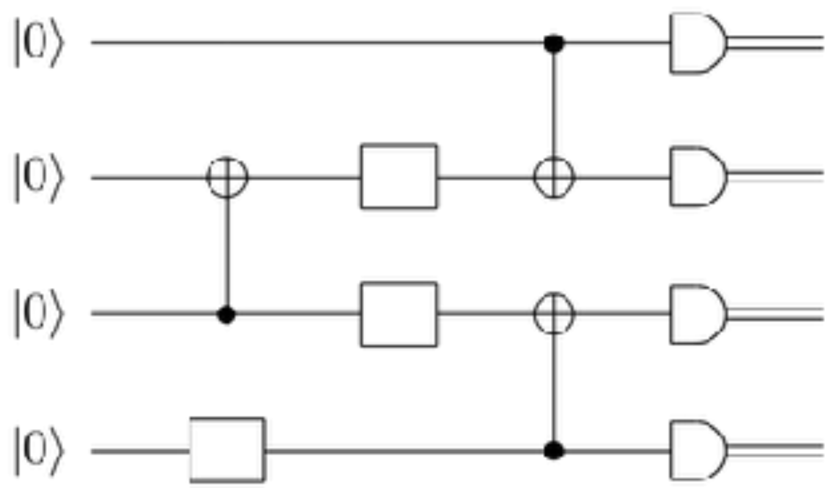
That what a general quantum computation is. The quantum algorithms (set of unitaries) are designed so that measuring the evolved quantum state results in the solution of a computational problem

$$\rho_U = U_N \ldots U_1 \rho_0 U_1^\dagger \ldots U_N^\dagger$$

BORN RULE

$$\mathbf{P}_U = \mathbf{O}_N \ldots \mathbf{O}_2 \mathbf{O}_1 \mathbf{P}_0$$

QUANTUM CIRCUIT



- Start the quantum device in a simple product state
- Apply a sequence of simple unitary matrices acting on the initial state
- To obtain the result, usually measure in the computational basis.

Looks similar to Green's function Monte Carlo, but it has sign problem

Because O are somewhat stochastic

$$\rho_U = U_N \ldots U_1 \rho_0 U_1^\dagger \ldots U_N^\dagger$$

BORN RULE

$$\mathbf{P}_U = \mathbf{O}_N \ldots \mathbf{O}_2 \mathbf{O}_1 \mathbf{P}_0$$

SIMULATING QUANTUM CIRCUITS WITH NEURAL MACHINE TRANSLATION

APPLY ONE GATE

$$P_U(\mathbf{a}'') = \sum_{\mathbf{a}'} P(\mathbf{a}') O_{\mathbf{a}',\mathbf{a}''}$$

Take an initial distribution

Multiply it by a somewhat stochastic matrix

Results in an evolved distribution Pu(a)

APPLY ONE GATE

$$P_U(\mathbf{a}'') = \sum_{\mathbf{a}'} P(\mathbf{a}') O_{\mathbf{a}',\mathbf{a}''}$$

Introduce a model $P_\theta(\mathbf{a})$

Compute "distance" between model and evolved $P_U(\mathbf{a})$ through sampling

Minimize distance

$$D_{\mathrm{KL}}\left(P_U || P_\theta\right) = -\sum_{\mathbf{a}} P_U\left(\mathbf{a}\right) \ln \frac{P_\theta\left(\mathbf{a}\right)}{P_U\left(\mathbf{a}\right)}$$

$$D_{\mathrm{KL}}\left(P_U || P_\theta\right) = H(P_U, P_\theta) - H(P_U)$$

$$H(P_U, P_\theta) = -\sum_{\mathbf{a}} P_U\left(\mathbf{a}\right) \ln P_\theta\left(\mathbf{a}\right) = -\sum_{\mathbf{a},\mathbf{a}'} P\left(\mathbf{a}'\right) O_{\mathbf{a}\mathbf{a}'} \ln P_\theta\left(\mathbf{a}\right)$$

SIMULATING QUANTUM CIRCUITS WITH RNN

$$H(P_U, P_\theta) = -\sum_{\mathbf{a}} P_U(\mathbf{a}) \ln P_\theta(\mathbf{a}) = -\sum_{\mathbf{a},\mathbf{a}'} P(\mathbf{a}') O_{\mathbf{a}\mathbf{a}'} \ln P_\theta(\mathbf{a})$$

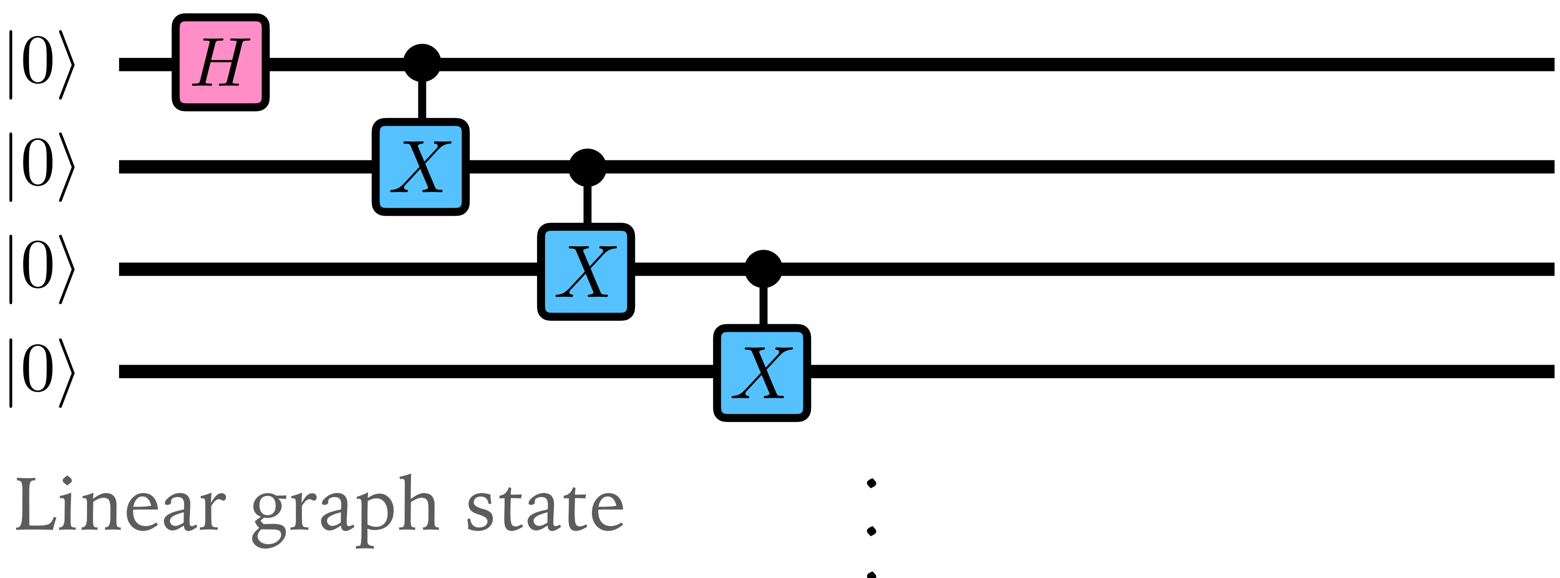
$$H(P_U, P_\theta) \approx -\frac{1}{N_s} \sum_{\mathbf{a}' \sim P} \sum_{\mathbf{a}} O_{\mathbf{a},\mathbf{a}'} \ln P_\theta(\mathbf{a})$$

Minimize cross entropy to search for an approximation to Pu from samples drawn from P

RESULTS: STATE PREPARATION FOR SIMPLE STATES

GHZ state

a

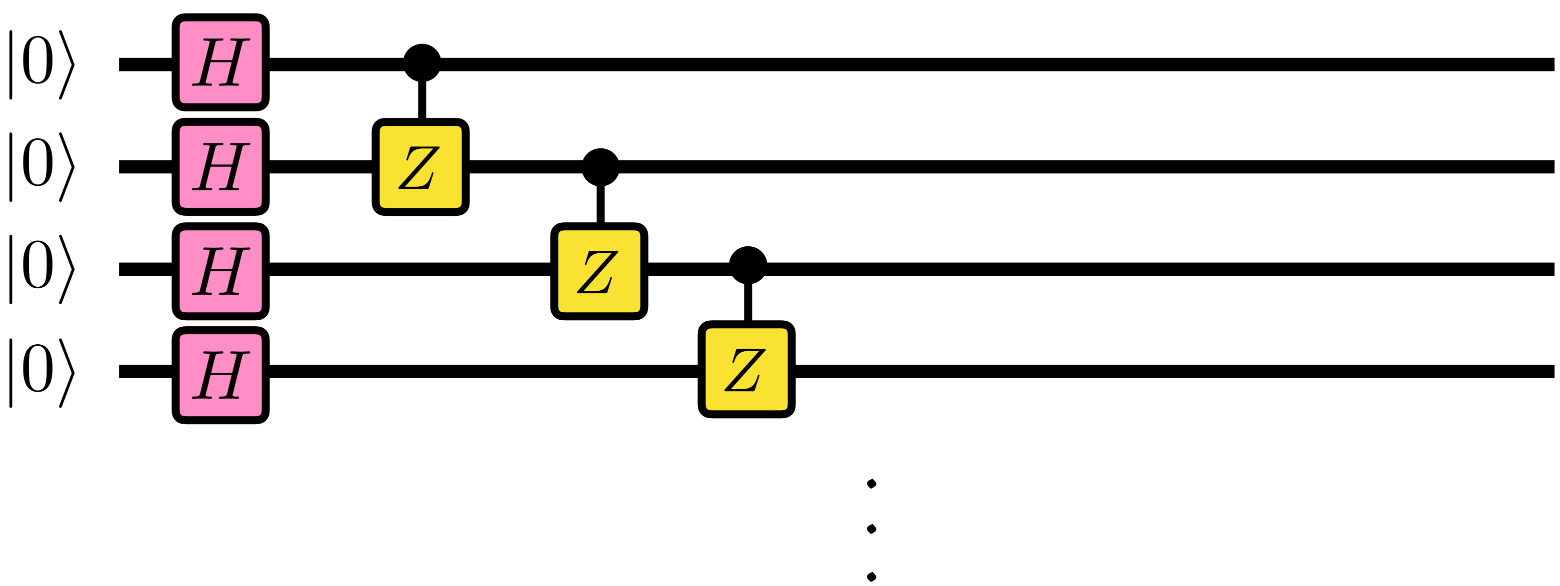


$$\mathrm{CNOT} = cX = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$H_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

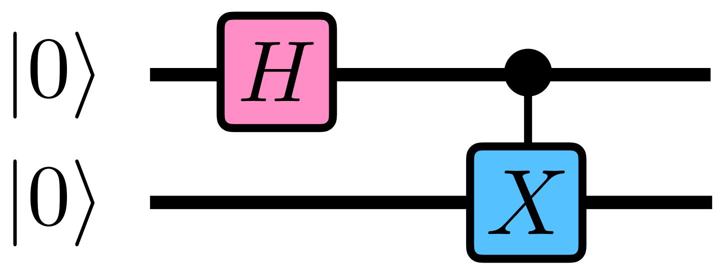
Linear graph state

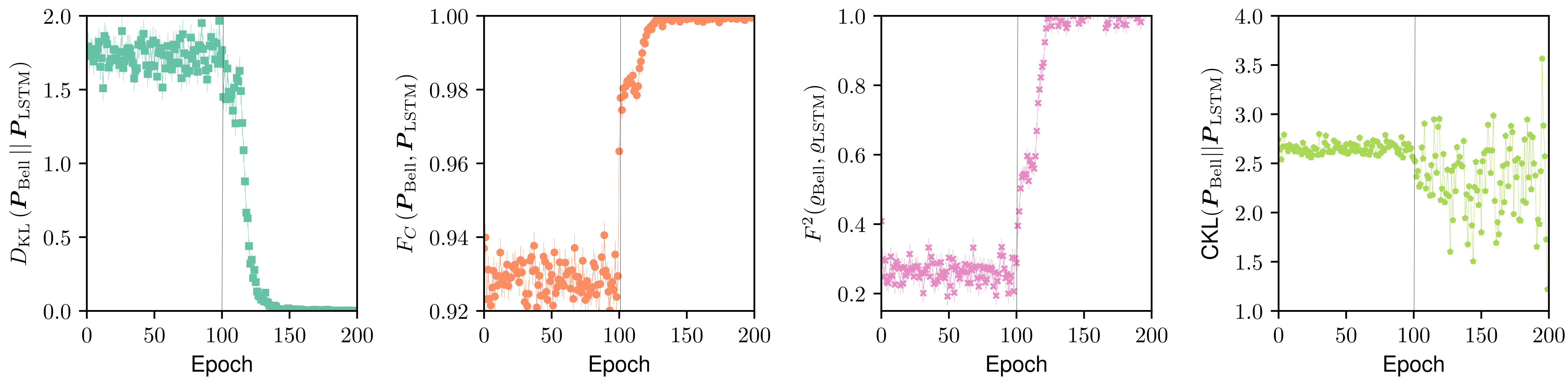
b



$$cZ = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

TRAINING DYNAMICS OF THE BELL STATE PREPARATION





Ns=10000

Batch size=100

LSTM model with two stacked layers with hidden states d=10 followed by a softmax

$$M_{4P}^{(0)} = \frac{1}{3}\,|0\rangle\langle 0|$$

$$M_{4P}^{(1)} = \frac{1}{3}\,|+\rangle\langle +|$$

$$M_{4P}^{(2)} = \frac{1}{3}\,|r\rangle\langle r|$$

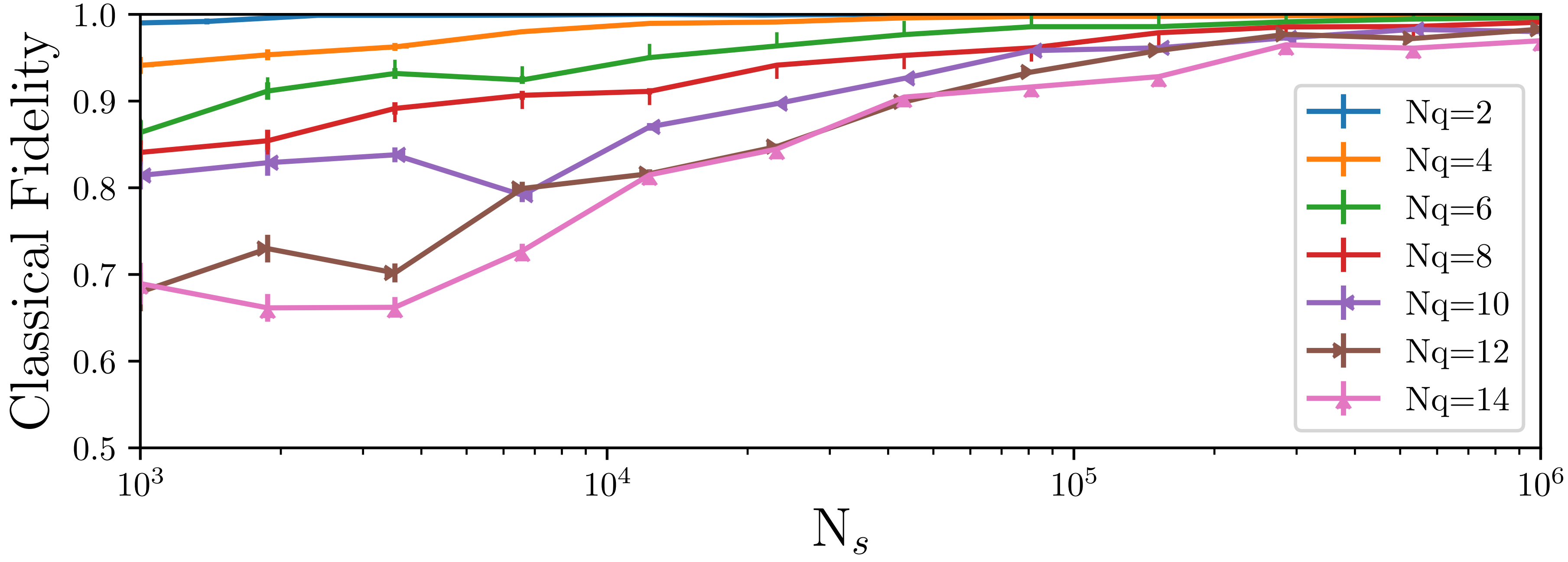
$$M_{4P}^{(3)} = \mathbf{1} - M^{(0)} - M^{(1)} - M^{(2)}$$

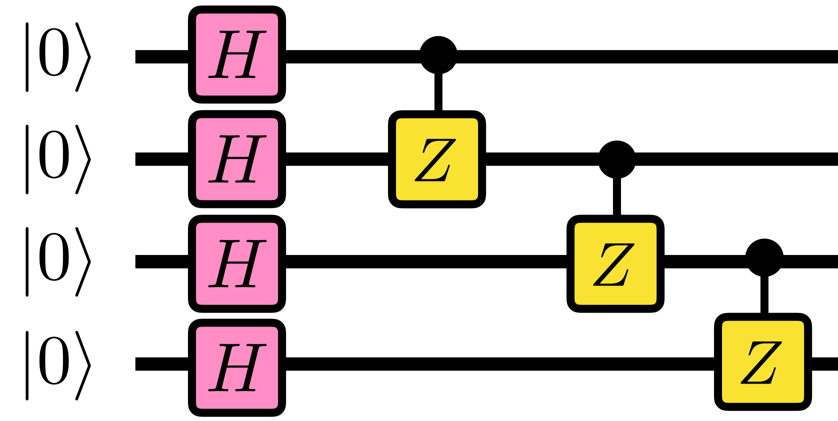
$$KL\left(P_{\mathrm{model}}|P\right) = -\sum_a P(a)\log\frac{P_{model}(a)}{P(a)}$$

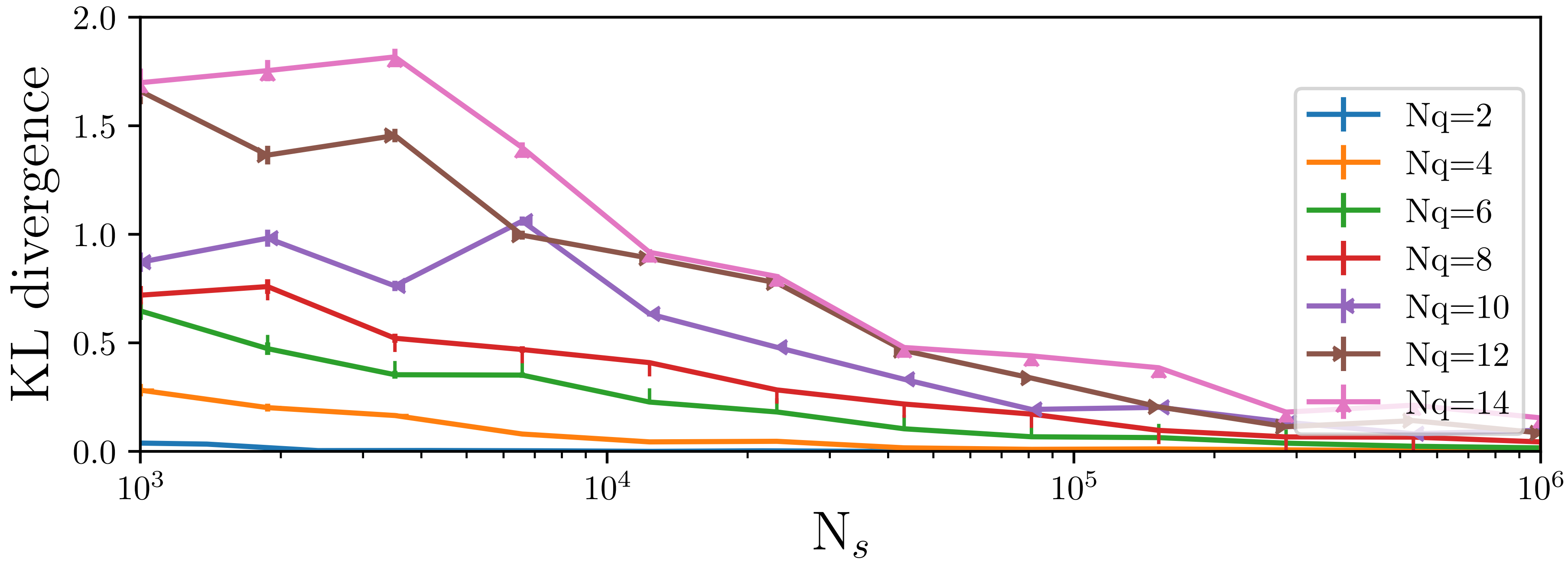
$$F_{\mathrm{Classical}} = \sum_a \sqrt{P(a)P_{model}(a)}$$

$$F\left(\rho,\sigma\right) = \mathrm{Tr}\left[\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right]$$

SAMPLE COMPLEXITY ANALYSIS OF THE LEARNING PROBLEM: GRAPH STATE







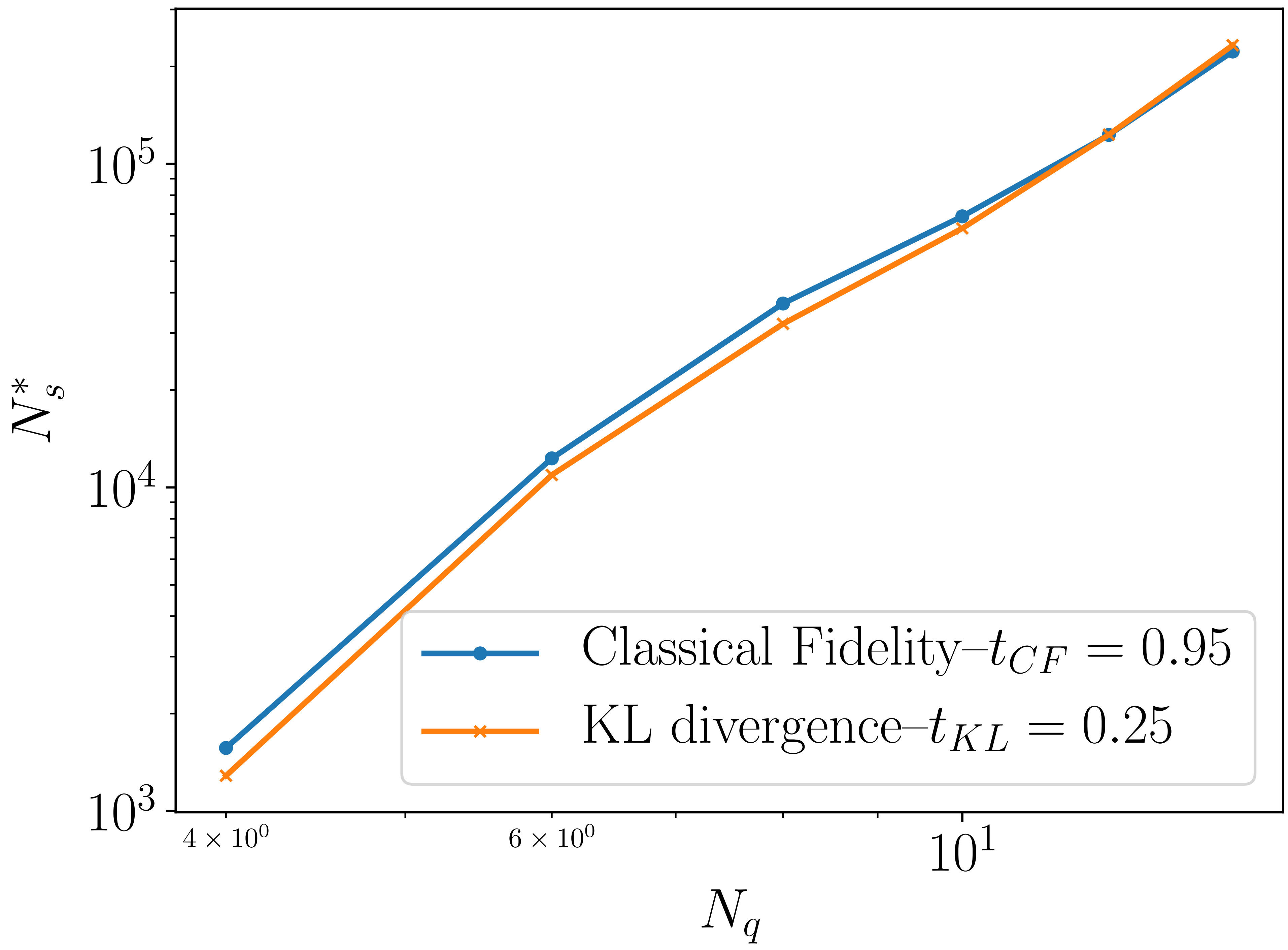
20 qubits with Transformer

CF= 0.9767 +/- 0.0001

Ns=10x10^6

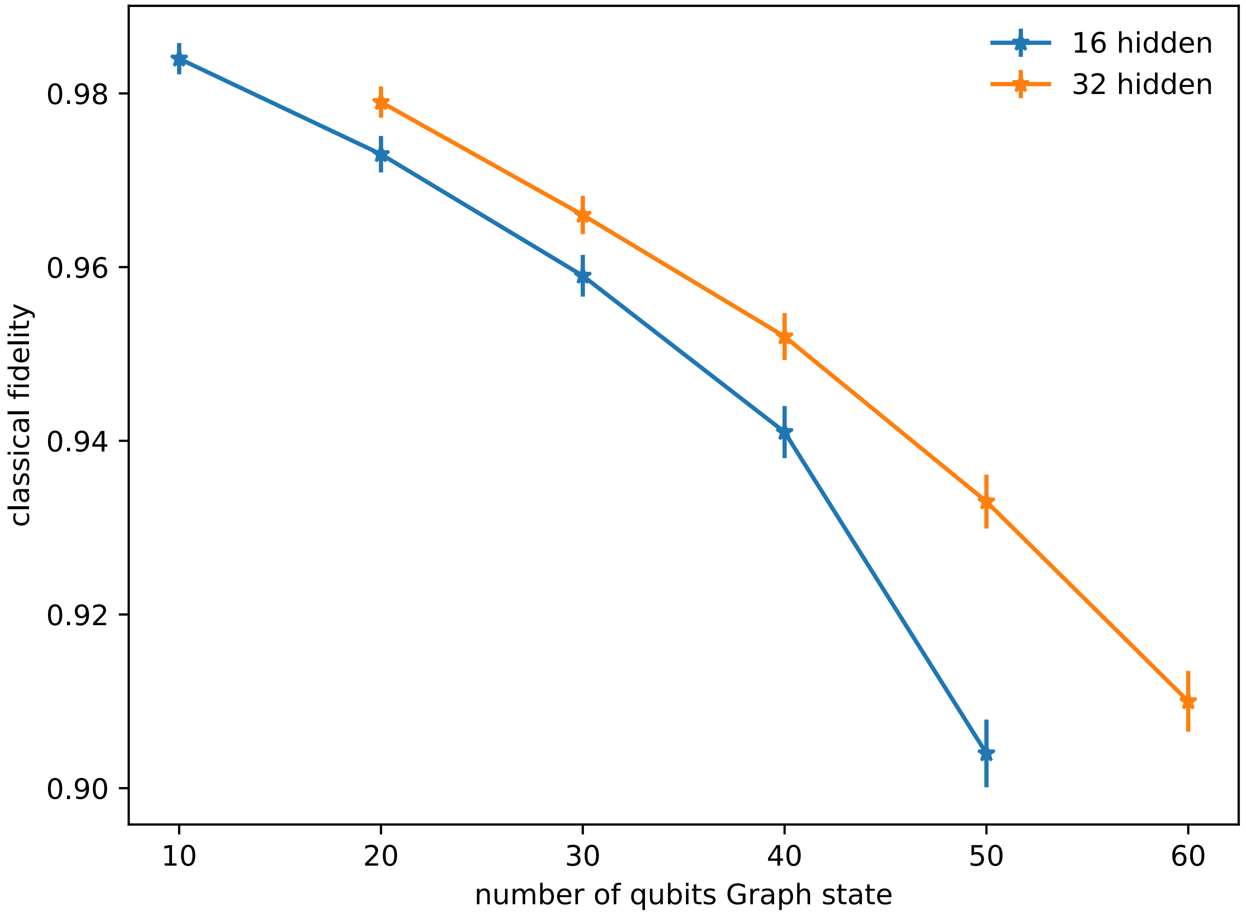
Ns is number of samples per gate application

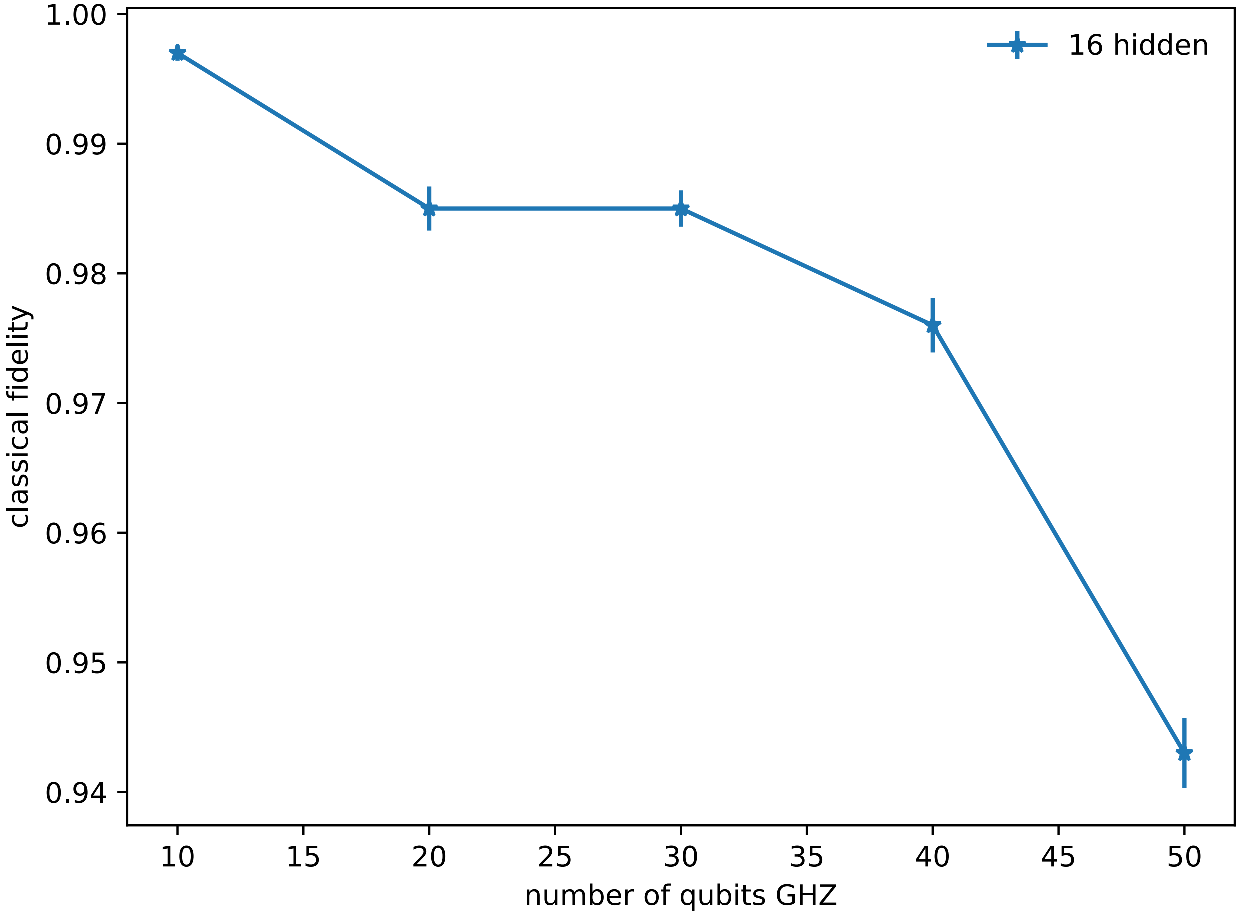
SAMPLE COMPLEXITY ANALYSIS OF THE LEARNING PROBLEM



Linear fit gives a slope of ~ 4.0, high complexity but still looks polynomial.

TRANSFORMER IS MUCH MORE SCALABLE

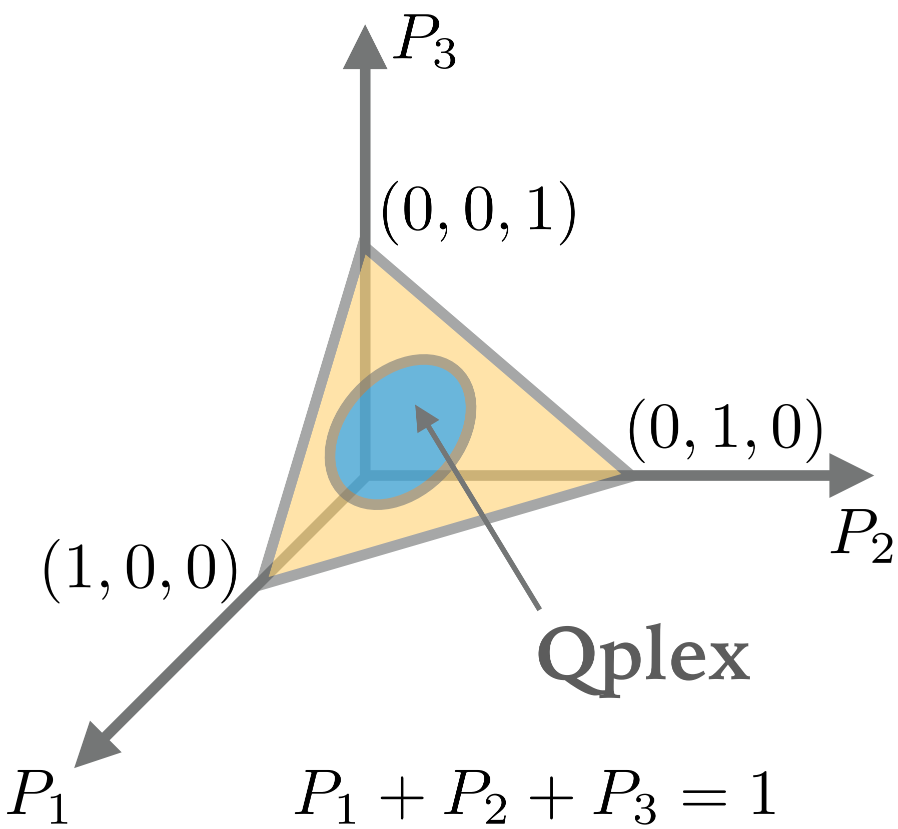




KNOWN BUGS

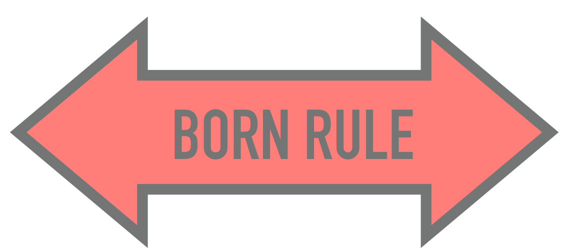
THE STANDARD SIMPLEX AND QUANTUM STATES

In probability, the points of the standard *n*-simplex in (*n* + 1)-space are the space of possible parameters (probabilities) of the categorical distribution on *n* + 1 possible outcomes.

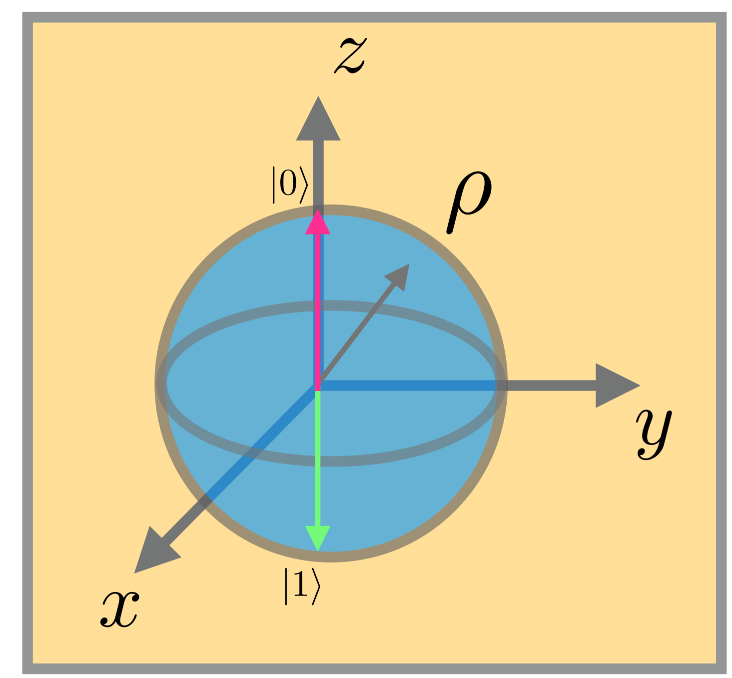


$$P_1 + P_2 + P_3 = 1$$

$$P(\mathbf{a}) = \mathrm{Tr}\,\rho M^{\mathbf{a}}$$

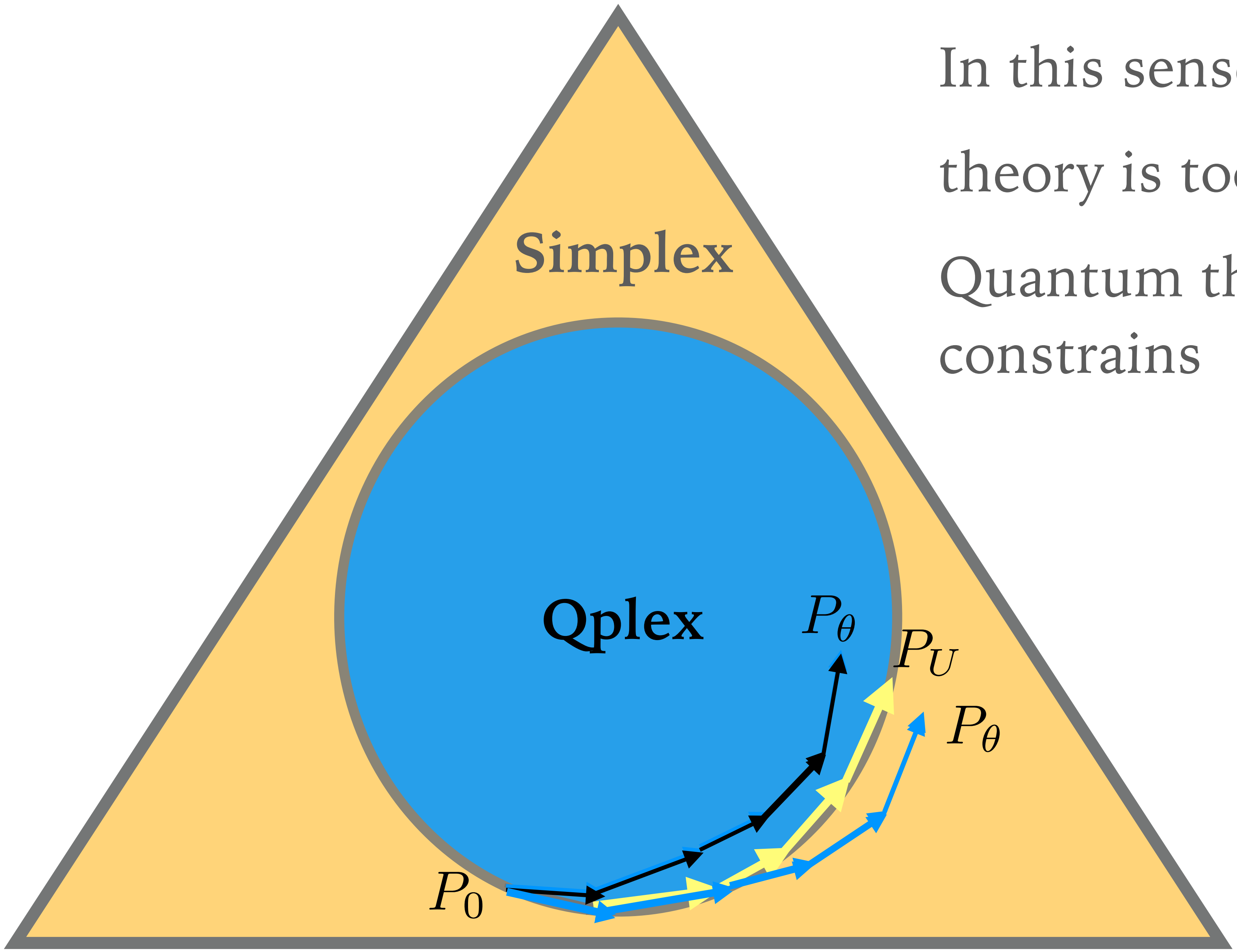


$$\rho = \sum_{a,a'} T^{-1}_{a,a'} P(a') M^{(a)}$$



$\{x \in \mathbb{R}^{k+1} : x_0 + x_1 + \ldots + x_k = 1, x_i \geq 0, i = 0, 1, \ldots k\}$ "All convex bodies behave a bit like Euclidean balls." Keith Ball

POSITIVITY AND VISUALIZING THE TRAINING IN THE Q-PLEX



In this sense probability theory is too general and Quantum theory needs constrains

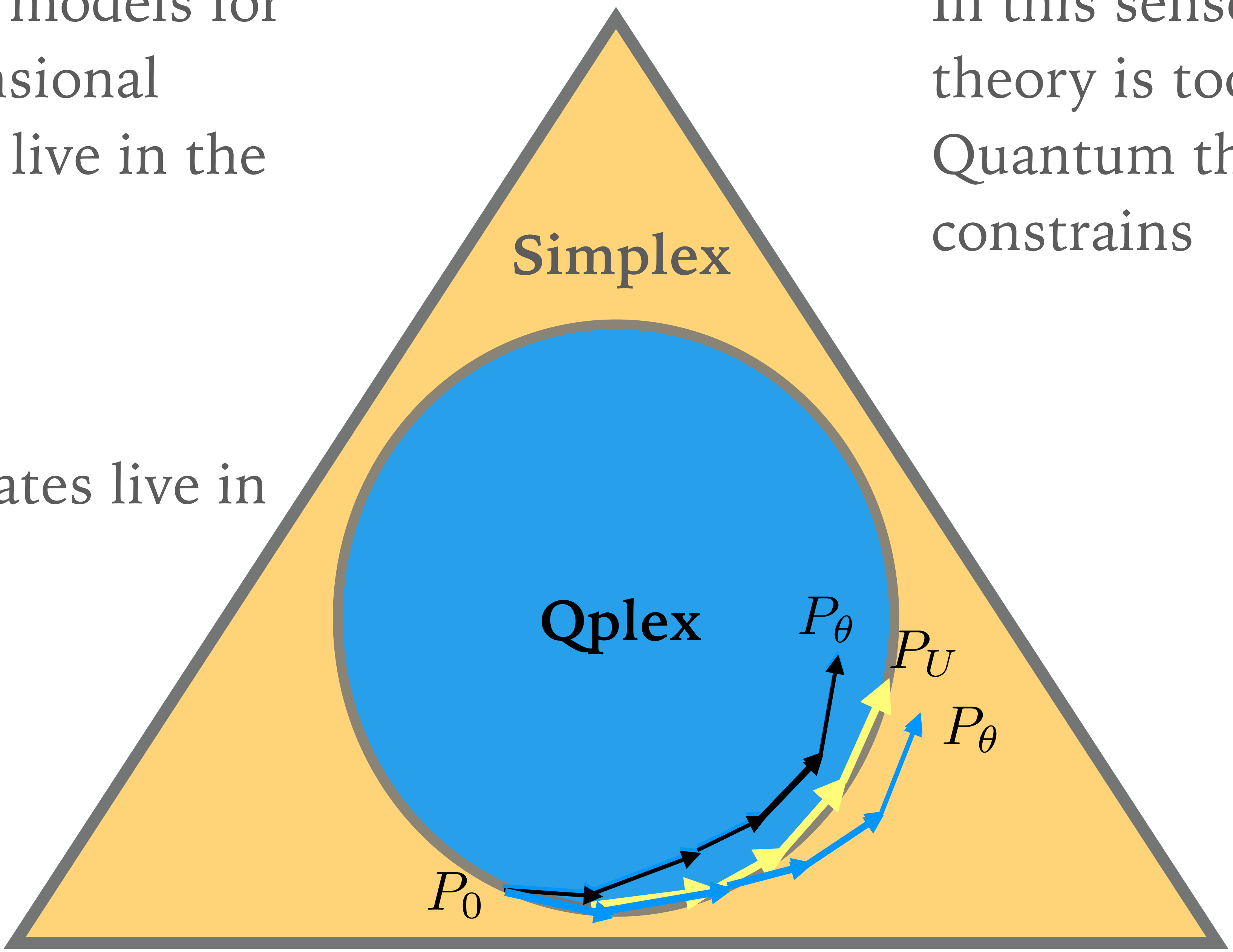
Introducing the Qplex: A Novel Arena for Quantum Theory. Marcus Appleby, Christopher A. Fuchs, Blake C. Stacey, Huangjun Zhu. arXiv:1612.03234 [quant-ph]

POSITIVITY AND VISUALIZING THE TRAINING IN THE QPLEX

Generative models for high dimensional probability live in the simplex

In this sense probability theory is too general and Quantum theory needs constrains

Quantum states live in the qplex



Introducing the Qplex: A Novel Arena for Quantum Theory. Marcus Appleby, Christopher A. Fuchs, Blake C. Stacey, Huangjun Zhu. arXiv:1612.03234 [quant-ph]

CLASSICAL FIDELITY IS NOT A STRONG METRIC OF FIDELITY

- $C_F \leq F$ Classical fidelity is overconfident
- Not a bug but a feature (?) since these probabilistic measurements of distance guarantee only that "most" measurements are correct.
- Perhaps a better measurement of fidelity for current quantum devices or approximate quantum computing

The Learnability of Quantum States

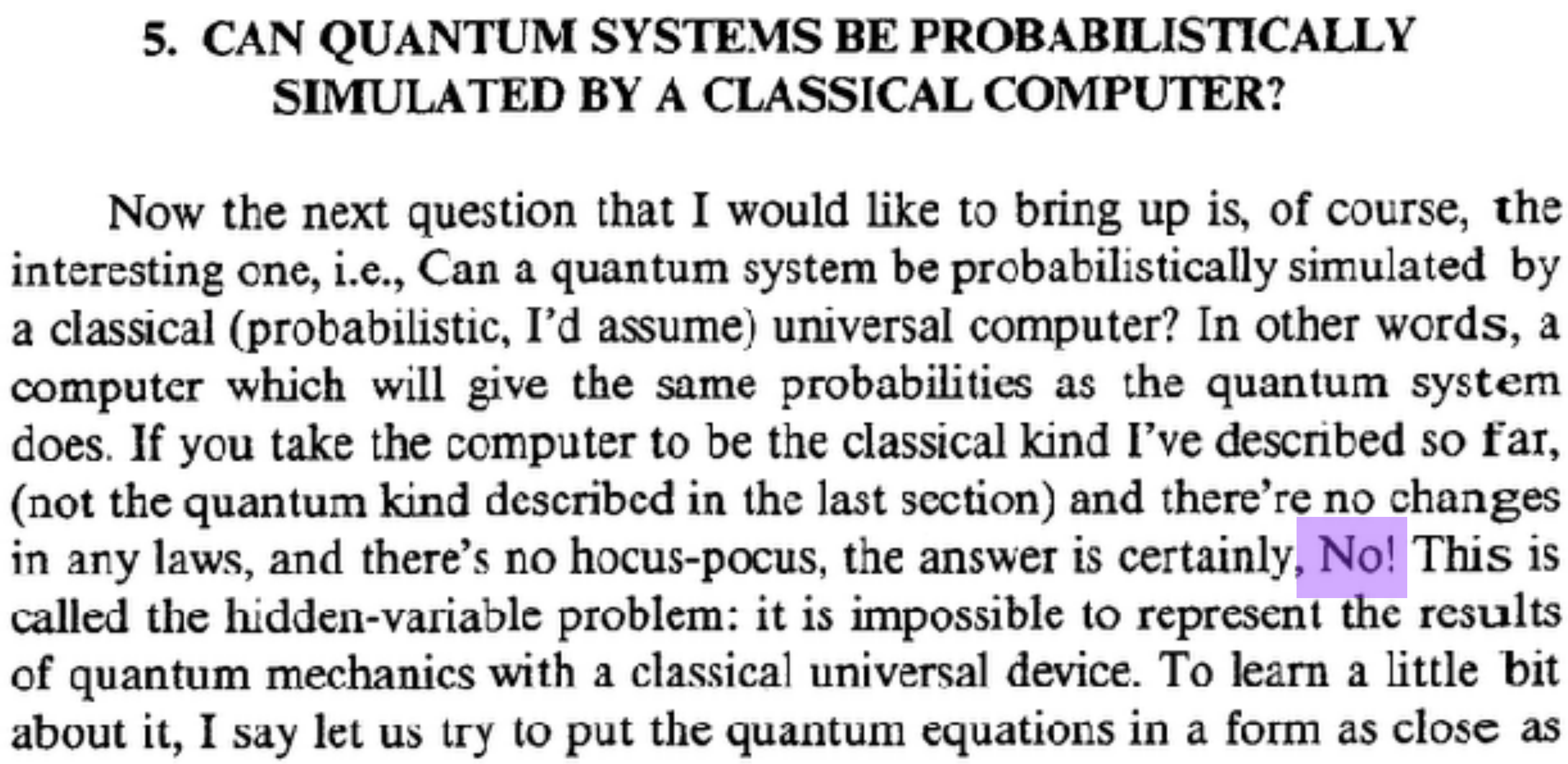
Scott Aaronson

(Submitted on 18 Aug 2006 (v1), last revised 4 Mar 2007 (this version, v3))

Traditional quantum state tomography requires a number of measurements that grows exponentially with the number of qubits n. But using ideas from computational learning theory, we show that "for most practical purposes" one can learn a state using a number of measurements that grows only linearly with n. Besides possible implications for experimental physics, our learning theorem has two applications to quantum computing: first, a new simulation of quantum one-way communication protocols, and second, the use of trusted classical advice to verify untrusted quantum advice.

Comments: 30 pages; added discussion of adaptive measurements, moved proofs to appendix, and corrected various minor errors
Subjects: Quantum Physics (quant-ph)
DOI: 10.1098/rspa.2007.0113
Cite as: arXiv:quant-ph/0608142
(or arXiv:quant-ph/0608142v3 for this version)

CAN WE SIMULATE QUANTUM SYSTEMS WITH PROBABILITY?



5. CAN QUANTUM SYSTEMS BE PROBABILISTICALLY SIMULATED BY A CLASSICAL COMPUTER?

Now the next question that I would like to bring up is, of course, the interesting one, i.e., Can a quantum system be probabilistically simulated by a classical (probabilistic, I'd assume) universal computer? In other words, a computer which will give the same probabilities as the quantum system does. If you take the computer to be the classical kind I've described so far, (not the quantum kind described in the last section) and there're no changes in any laws, and there's no hocus-pocus, the answer is certainly, No! This is called the hidden-variable problem: it is impossible to represent the results of quantum mechanics with a classical universal device. To learn a little bit about it, I say let us try to put the quantum equations in a form as close as

Answer is still NO (duh), since evolving these distributions remains a challenge (eg circuits evolution has a sign problem). However we have introduced a **heuristic** to do it using RNNs

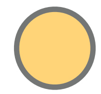
CONCLUSIONS

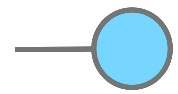
- Can quantum systems be probabilistically simulated in a classical computer? No
- But our reformulation of quantum theory + ML can help. A similar formulation is used in quantum Bayesian theory.
- Heuristic using language translation models to simulate quantum circuits.
- Good: Optimization problem is relatively easy
- But: requires a lot of samples/tricks to improve variance in the gradients
- Parallelize over GPUs
- Would like to find generative models directly living in the qplex

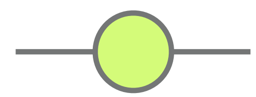
OUTLOOK

- Run other simple quantum algorithms on the transformer
- Real-time dynamics, Linblad equation, etc

NOTATION SLIDE: PENROSE NOTATION

S  *Scalar*

V_i  *Vector*

$W_{i,j}$  *Matrix*

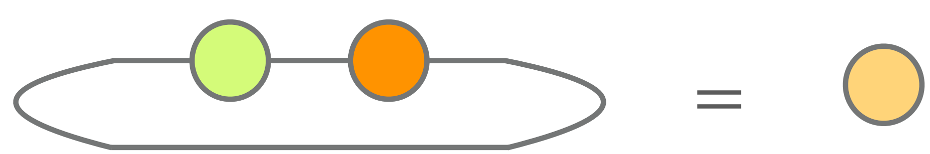
$$C_k = \sum_j W_{k,j} V_j$$

Matrix vector Multiplication

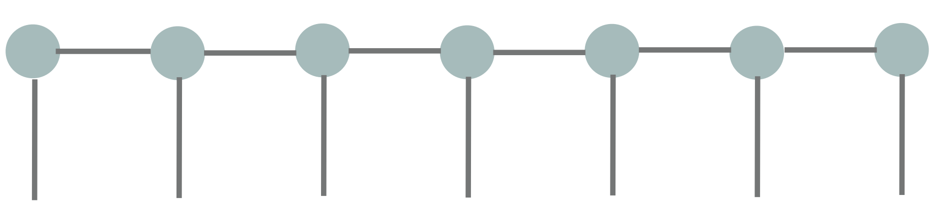
$$|C|^2 = \sum_k C_k C_k^*$$

norm

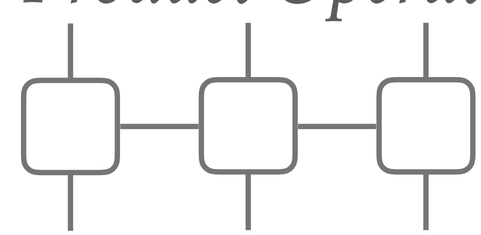
$$\text{Tr } W M = \sum_{k,j} W_{k,j} M_{j,k}$$



Matrix Product States (MPS)



Matrix Product Operators (MPO)

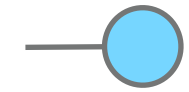


https://www.math3ma.com/blog/matrices-as-tensor-network-diagrams

NOTATION SLIDE: DIRAC NOTATION

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad |\Psi\rangle = \alpha|0\rangle + \beta|1\rangle$$

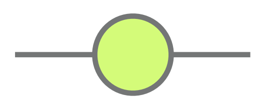
Dirac representation of Vectors



$$\langle\Psi| = [\alpha^* \;\; \beta^*]$$

Conjugate transpose

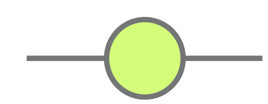
$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$



matrices

$$|0\rangle\langle 0| = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \qquad |1\rangle\langle 1| = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Outer product



Quantum computing for the very curious https://quantum.country/qcvc

MEASUREMENTS: POSITIVE OPERATOR VALUED MEASURES (POVM)

CONSTRUCTING POVMS: TAKE A SINGLE QUBIT POVM AND MAKE A TENSOR PRODUCT OF MANY

Pauli measurement for one qubit

$$\boldsymbol{M}_{\text{Pauli}} := \{M^{(0)} := p(3) \times |0\rangle\langle 0|\,, M^{(1)} := p(3) \times |1\rangle\langle 1|\,,$$
$$M^{(+)} := p(1) \times |+\rangle\langle +|\,, M^{(-)} := p(1) \times |-\rangle\langle -|\,,$$
$$M^{(r)} := p(2) \times |r\rangle\langle r|\,, M^{(l)} := p(2) \times |l\rangle\langle l|\,\},$$

Experimental realization:

- Choose a random direction (x,y,z) with probability 1/3
- Measure along the chosen direction
- 6 possible outcomes: If x -> + or -; if y -> r or l; if z 0 or 1

MEASUREMENTS: POSITIVE OPERATOR VALUED MEASURES (POVM)

CONSTRUCTING POVMS: TAKE A SINGLE QUBIT POVM AND MAKE A TENSOR PRODUCT OF MANY

Pauli measurement for one qubit

$$\begin{aligned}\boldsymbol{M}_{\text{Pauli}} := \{ & M^{(0)} := p(3) \times |0\rangle\langle 0| \, , M^{(1)} := p(3) \times |1\rangle\langle 1| \, , \\ & M^{(+)} := p(1) \times |+\rangle\langle +| \, , M^{(-)} := p(1) \times |-\rangle\langle -| \, , \\ & M^{(r)} := p(2) \times |r\rangle\langle r| \, , M^{(l)} := p(2) \times |l\rangle\langle l| \, \},\end{aligned}$$

For multiqubit systems $\boldsymbol{M}_{\text{Pauli}} \otimes \boldsymbol{M}_{\text{Pauli}} \otimes \boldsymbol{M}_{\text{Pauli}} \otimes \boldsymbol{M}_{\text{Pauli}} \otimes \boldsymbol{M}_{\text{Pauli}} \otimes \boldsymbol{M}_{\text{Pauli}} \otimes$

we will consider

$$\boldsymbol{M} = \left\{ M^{(a_1)} \otimes M^{(a_2)} \otimes \ldots M^{(a_N)} \right\}_{a_1, \ldots a_N}$$

Experimental realization: pick a random direction with probability 1/3, then measure in that direction on each qubit independently

Easy to implement in gate-based QC (Qiskit, Cirq, Rigetti, etc.)

HOW IS THIS USEFUL FOR QUANTUM TECHNOLOGY?

MEASUREMENTS

- Suppose we want to measure the quantum state. The measurement is described by some other POVM $\Pi^{(\boldsymbol{b})}$

$$P(\boldsymbol{b}) = \sum_{\boldsymbol{a},\boldsymbol{a}'} P(\boldsymbol{a}')\, T^{-1}_{\boldsymbol{a},\boldsymbol{a}'}\, \mathrm{Tr}\left[M^{(\boldsymbol{a})}\Pi^{(\boldsymbol{b})}\right] = \sum_{\boldsymbol{a}'} q(\boldsymbol{b}|\boldsymbol{a}')P(\boldsymbol{a}')$$

$$q(\boldsymbol{b}|\boldsymbol{a}') = \sum_{\boldsymbol{a}} T^{-1}_{\boldsymbol{a},\boldsymbol{a}'}\, \mathrm{Tr}\left[M^{(\boldsymbol{a})}\Pi^{(\boldsymbol{b})}\right]$$

- can be characterized as a **somewhat** conditional probability since its entries can either be positive or negative but its trace over b is the identity.
- evocative resemblance with the law of total probability—> quantum law of total probability in quantum Bayesianism.

QUASIPROBABILITY

PAPER • OPEN ACCESS

Negative quasi-probability as a resource for quantum computation

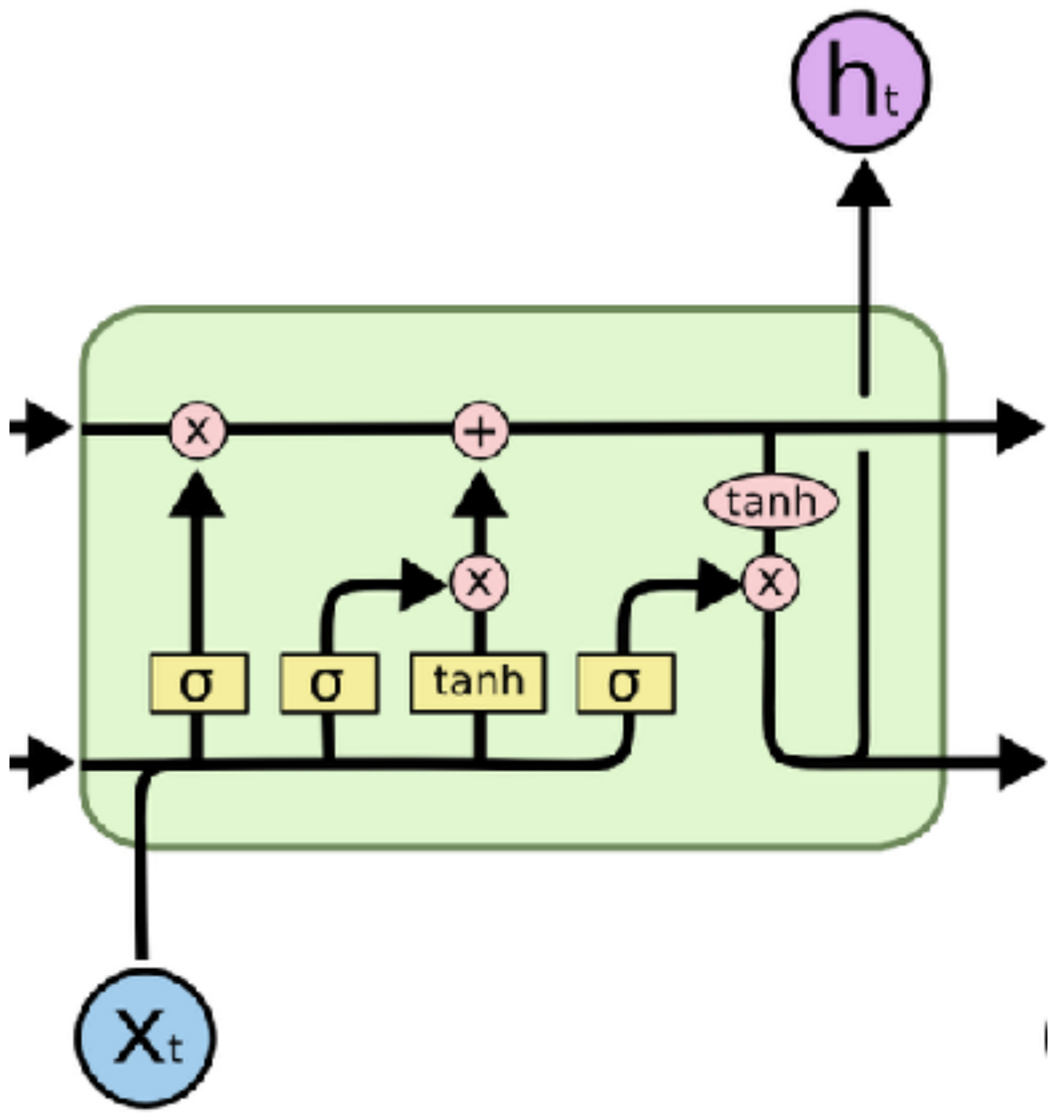
Victor Veitch[1], Christopher Ferrie[1], David Gross[2] and Joseph Emerson[1]

Published 8 November 2012 • IOP Publishing and Deutsche Physikalische Gesellschaft
New Journal of Physics, Volume 14, November 2012

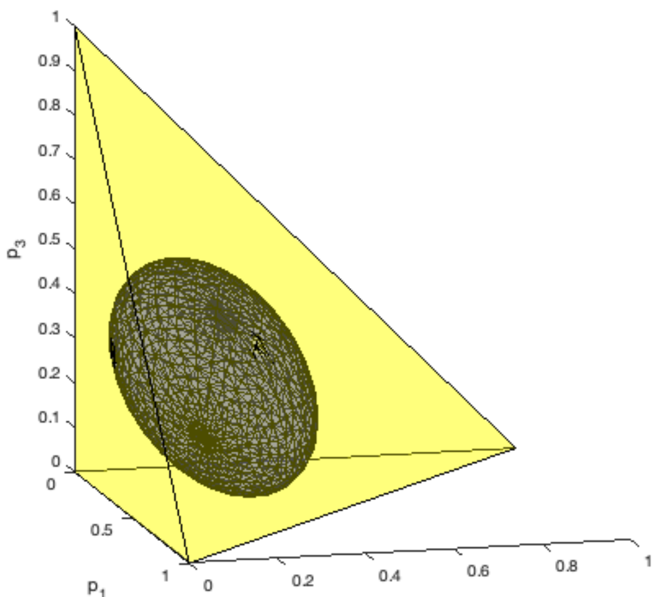
Abstract

A central problem in quantum information is to determine the minimal physical resources that are required for quantum computational speed-up and, in particular, for fault-tolerant quantum computation. We establish a remarkable connection between the potential for quantum speed-up and the onset of negative values in a distinguished quasi-probability representation, a discrete analogue of the Wigner function for quantum systems of odd dimension. This connection allows us to resolve an open question on the existence of bound states for magic state distillation: we prove that there exist mixed states outside the convex hull of stabilizer states that cannot be distilled to non-stabilizer target states using stabilizer operations. We also provide an efficient simulation protocol for Clifford circuits that extends to a large class of mixed states, including bound universal states.

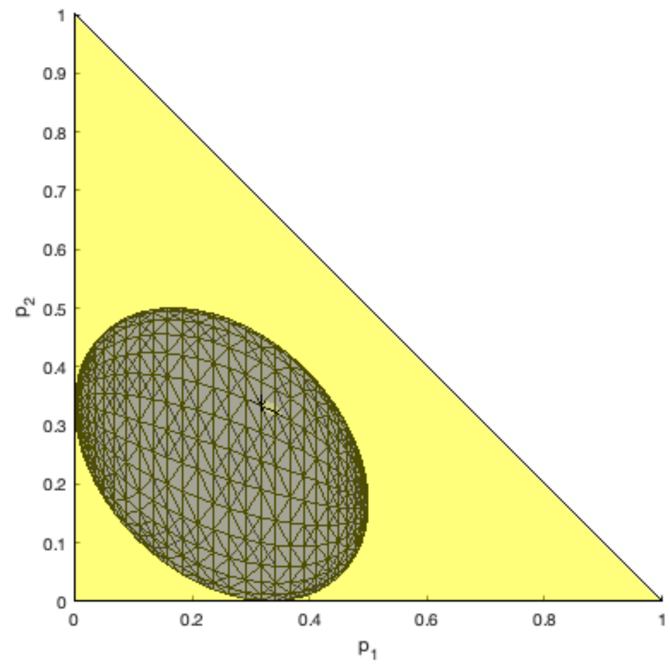
LSTM CELL



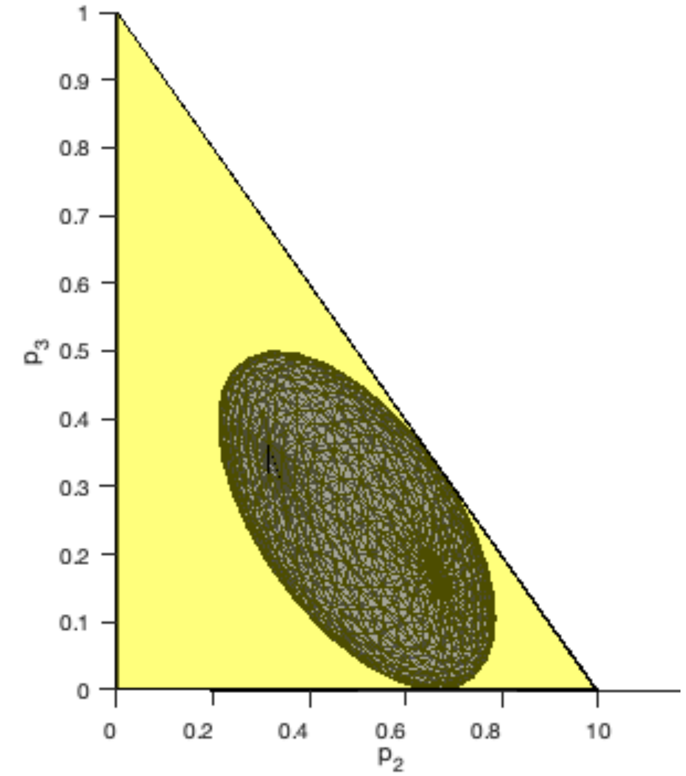
QPLEXES



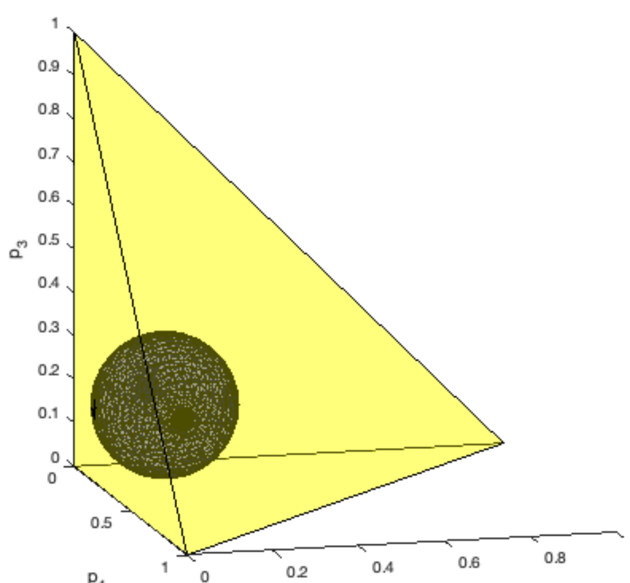
(a) view 1



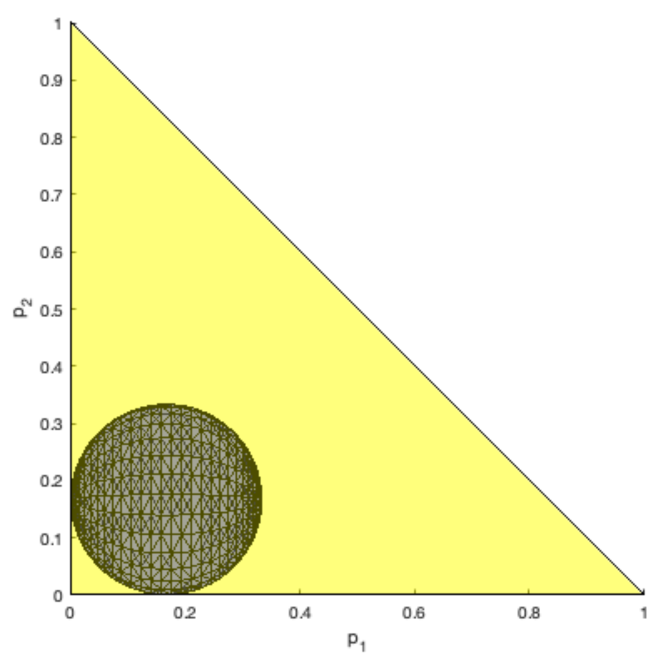
(b) view 2



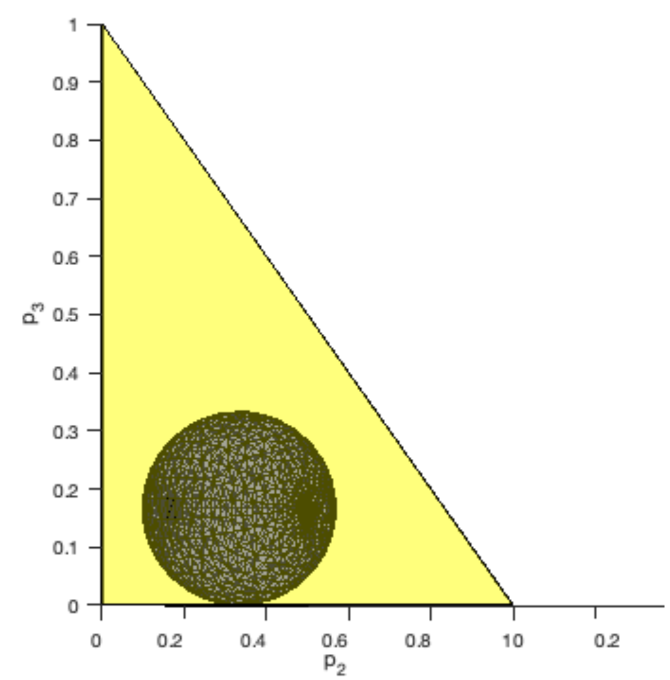
(c) view 3



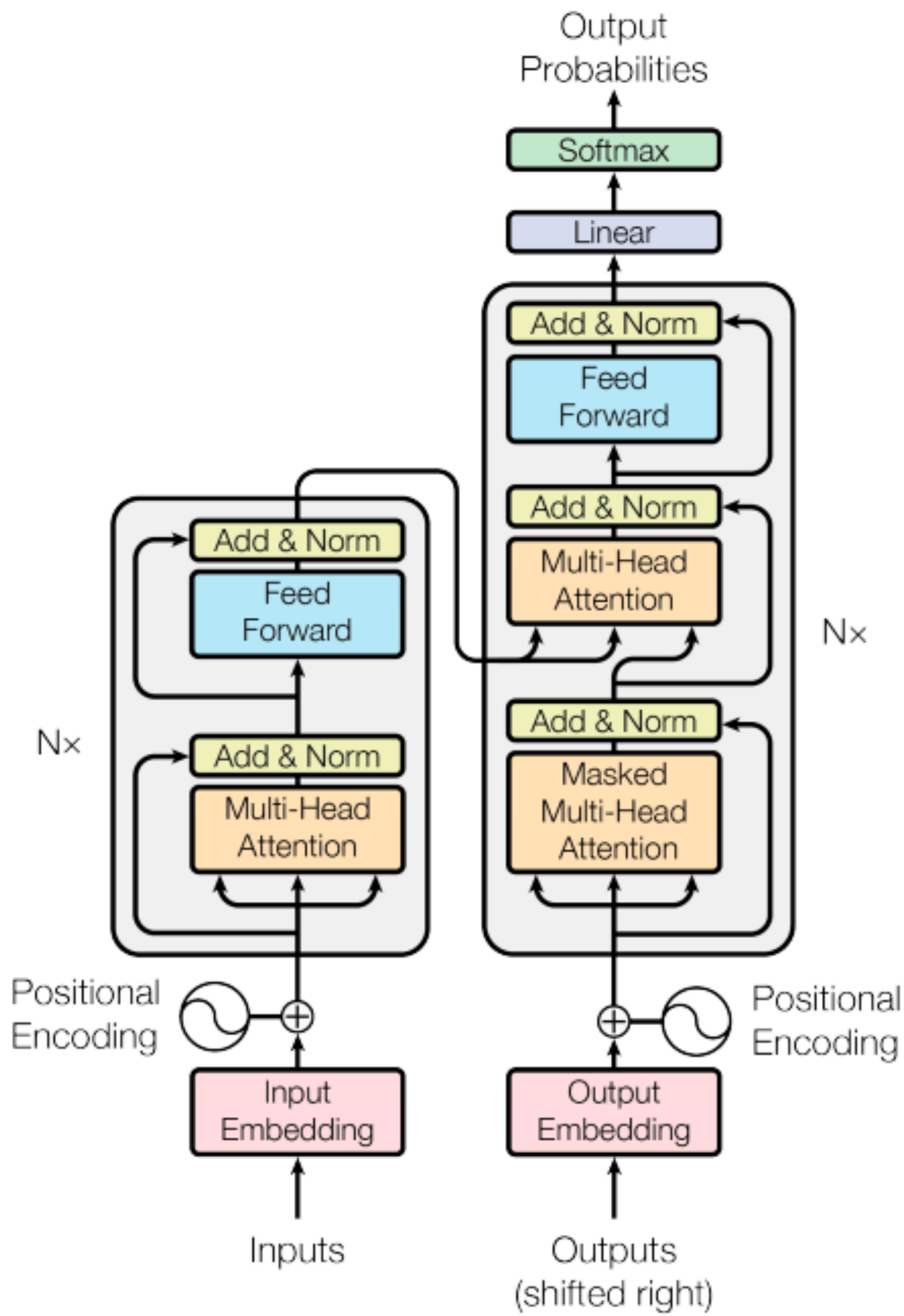
(a) view 1



(b) view 2



(c) view 3


Output Probabilities
Softmax
Linear
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Add & Norm
Masked Multi-Head Attention
Nx
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Nx
Positional Encoding
Positional Encoding
Input Embedding
Output Embedding
Inputs
Outputs (shifted right)