LEARNING QUANTUM STATES WITH GENERATIVE MODELS

Juan Carrasquilla
@carrasqu

September 17th, 2018

Toronto Deep Learning Series

VECTOR INSTITUTE | INSTITUT VECTEUR

UNIVERSITY OF WATERLOO

VECTOR
INSTITUTE

INSTITUT
VECTEUR

About

- The Vector Institute is an independent, not-for-profit entity dedicated to research in artificial intelligence with a focus on excellence in machine learning and deep learning including applications in natural sciences.
- Vector launched in 2017 with support from the Government of Ontario, Government of Canada, and the private sector.

Researchers


Geoffrey Hinton
Chief Scientific Advisor


Richard Zemel
Research Director


Alán Aspuru-Guzik


Jimmy Ba


Juan Carrasquilla


David Duvenaud


Murat Erdogdu


Amir-massoud Farahmand


Sanja Fidler


David Fleet


Brendan Frey


Marzyeh Ghassemi


Anna Goldenberg

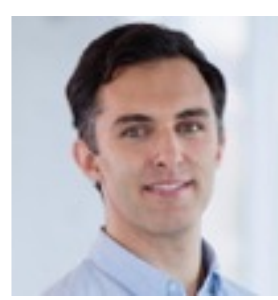

Roger Grosse


Alireza Makhzani

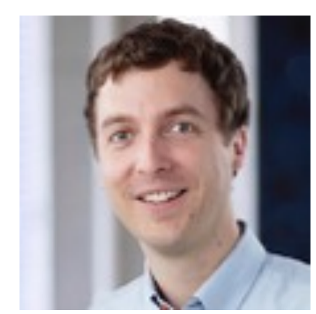

Quaid Morris

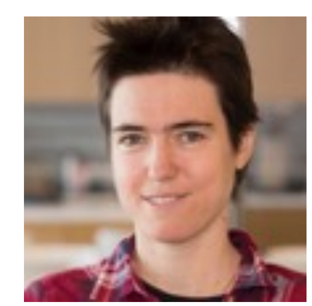

Sageev Oore


Pascal Poupart


Daniel Roy


Frank Rudzicz


Graham Taylor


Raquel Urtasun

Over the next few years it is expected that Vector will grow to include several hundred researchers, students, engineers and others.

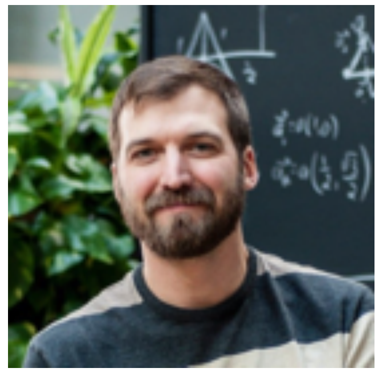

vectorinstitute.ai | @VectorInst

2 PhD positions and short-term visits

- If you know students interested in working at the intersection between machine learning, condensed matter theory, numerical simulations of quantum many-body systems, and quantum computers
- Possibility to interact and collaborate with ML and quantum computing experts
- Possibility to interact with industry if you want
- Get in touch with me carrasqu@vectorinstitute.ai
- https://vectorinstitute.ai/team/juan-felipe-carrasquilla/

MACHINE LEARNING/MANY-BODY PHYSICS FRIENDS

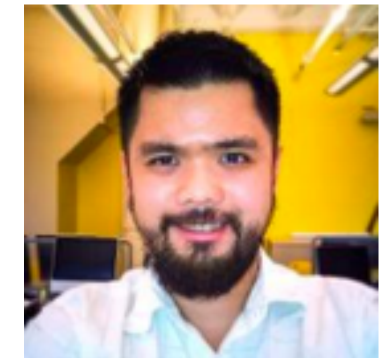
Leandro Aolita Universidade Federal do Rio de Janeiro



Roger Melko (U. Waterloo and Perimeter)

Giacomo Torlai (U. Waterloo and Perimeter)

Ehsan Khatami (San Jose State U)



Kelvin Chng (San Jose State U)

Giuseppe Carleo (Flatiron Institute)

Matthias Troyer (Microsoft, ETH)

Peter Broecker (U. Cologne)

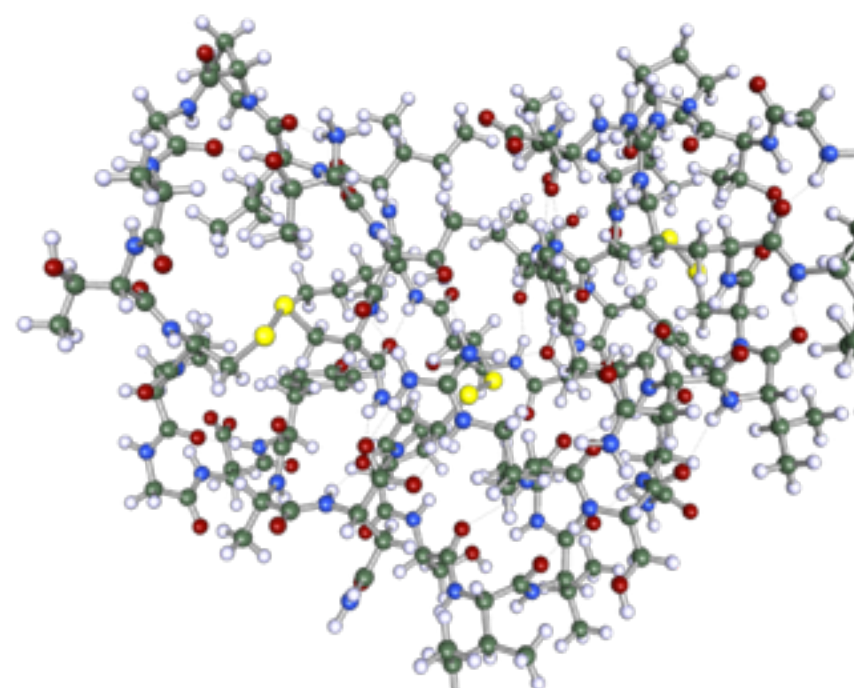
Simon Trebst (U. Cologne)

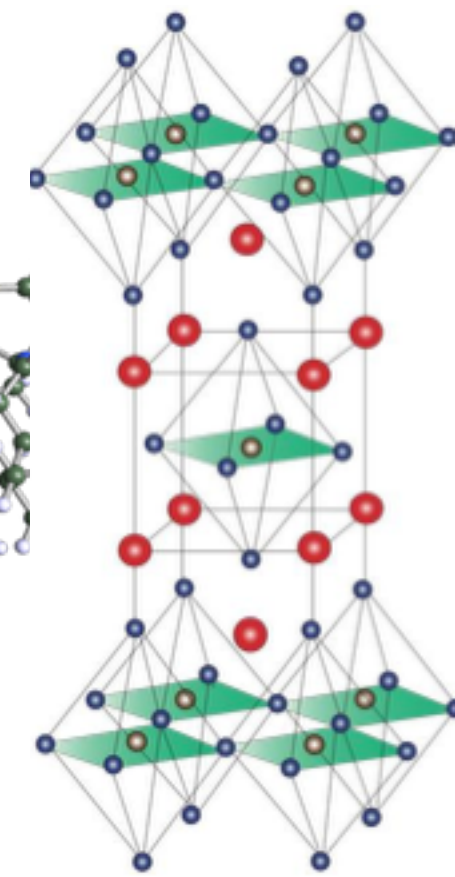
Guglielmo mazzola (ETH)

THE MANY-BODY PROBLEM IN QUANTUM MECHANICS

- Generic specification of a quantum state requires resources exponentially large in the number of degrees of freedom N
- Today's best supercomputers can solve the wave equation **exactly** for systems with a maximum of ~45 particles.
- Storing the state of a 273 spin system requires a computer with more bits than there are atoms in the universe
- Yet, technologically relevant problems in chemistry, condensed matter physics, and quantum computing are much larger than 273.
- Quantum computing

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



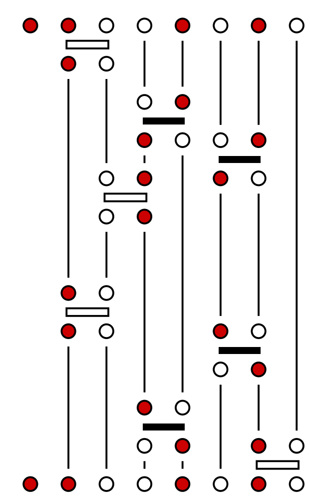


THERE IS STILL HOPE FOR CLASSICAL ALGORITHMS

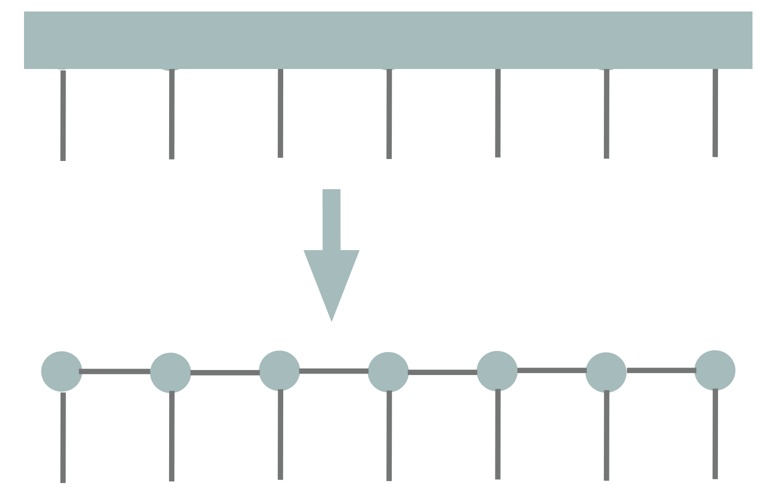
- Nature is sometimes compassionate: many-body systems can be typically characterized by an amount of information smaller than the maximum capacity of the corresponding state space.
- Quantum Monte Carlo and other numerical methods based on Tensor Networks exploit this fact and are able to accurately study large quantum system in practice with limited amount of resources.
- Machine learning community deals with equally high dimensional problems ($N \sim 256^2$) and battle the curse of dimensionality successfully with impressive results in a wide spectrum of scientific and technologically relevant areas of research.

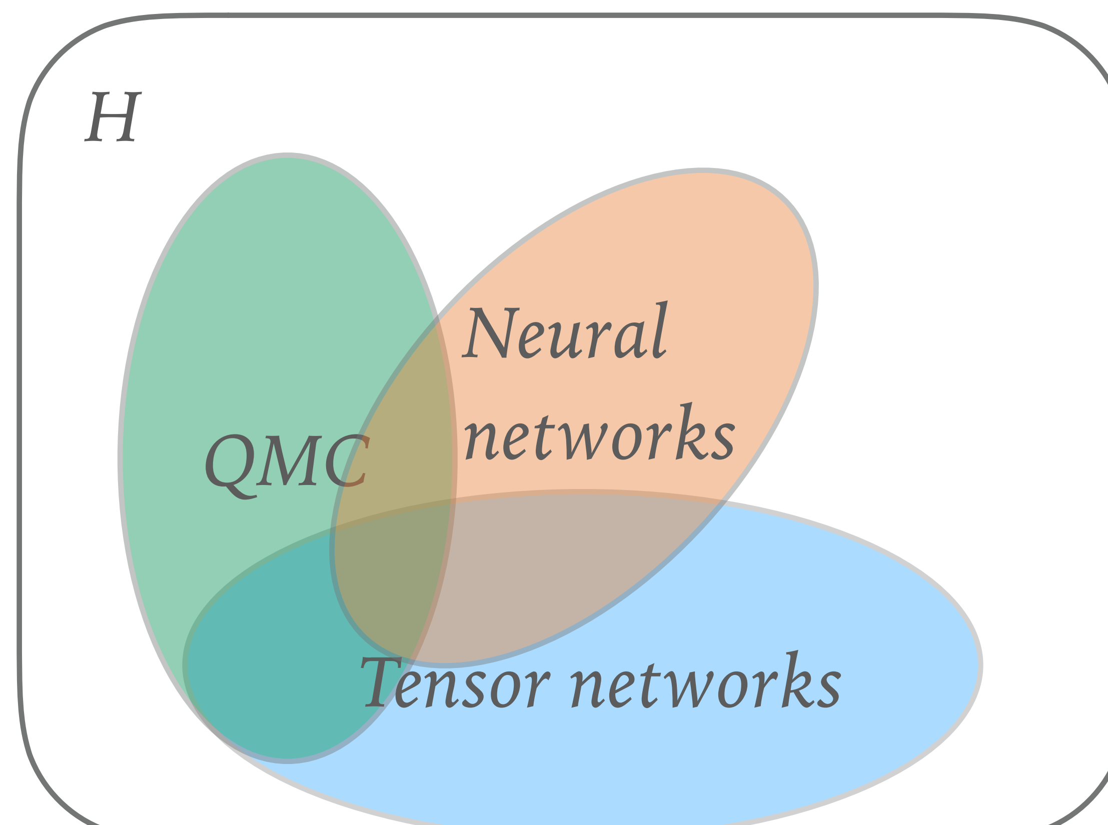
$|\Psi\rangle$ *vector with* 2^N

QMC



Low entanglement MPS and other TN



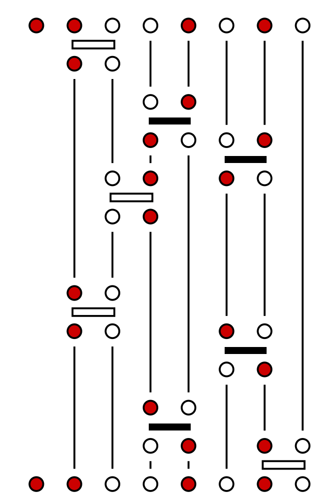


THERE IS STILL HOPE FOR CLASSICAL ALGORITHMS

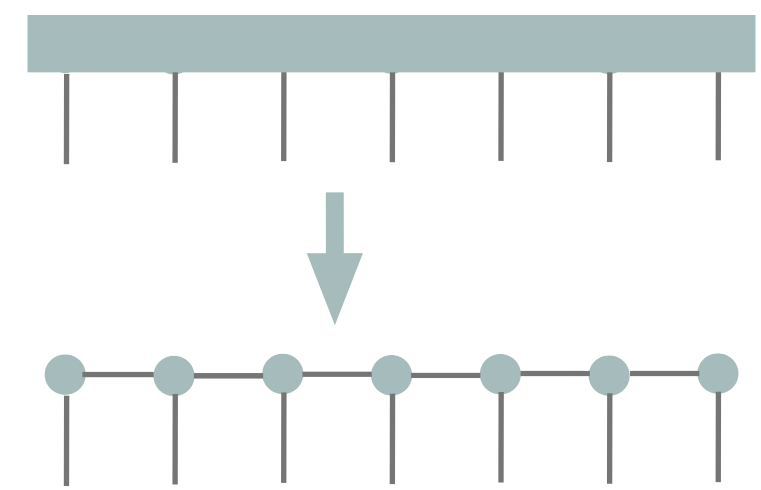
- Nature is sometimes compassionate: many-body systems can be typically characterized by an amount of information smaller than the maximum capacity of the corresponding state space.
- Quantum Monte Carlo and other numerical methods based on Tensor Networks exploit this fact and are able to accurately study large quantum system in practice with limited amount of resources.
- Machine learning community deals with equally high dimensional problems and battle the curse of dimensionality successfully with impressive results in a wide spectrum of scientific and technologically relevant areas of research.

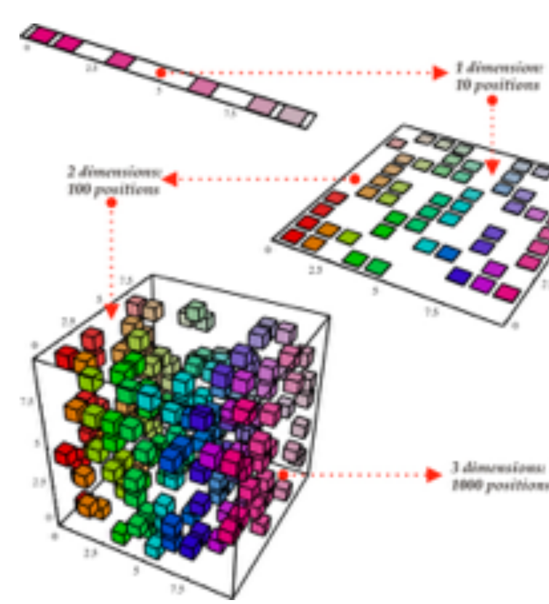
$|\Psi\rangle$ *vector with* 2^N

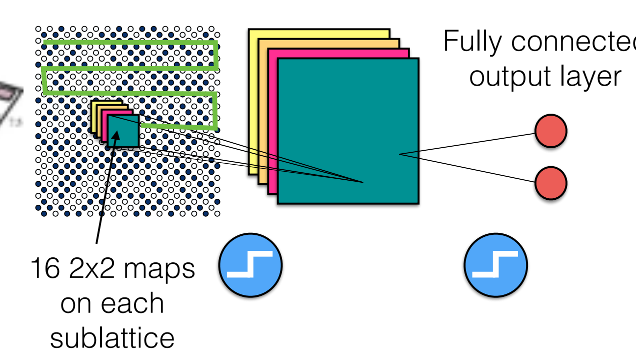
QMC



Low entanglement MPS and other TN







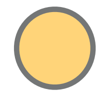
QUANTUM AND CLASSICAL MANY-BODY PHYSICS HAS NOT BEEN THE EXCEPTION

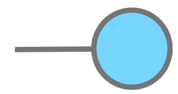
- ML phases of matter/phase transitions (Carrasquilla, Melko 1605.01735, Wang 1606.00318, Zhang, Kim, 1611.01518)
- New ML inspired ansatz for quantum many-body systems (Carleo, Troyer 1606.02318, Deng, Li, Das Sarma, 1701.04844, Deng, Li, Das Sarma 1609.09060, Carrasquilla, Melko 1605.01735)
- Accelerated Monte Carlo simulations (Huang, Wang 1610.02746)
- Quantum state preparation guided by ML (Bukov, Day, Sels, Weinberg, Polkovnikov and Mehta 1705.00565)
- Renormalization group analyses, RBMs, PCA (Bradde, Bialek 1610.09733, Koch-Janusz, Ringel 1704.06279,Mehta, Schwab,1410.3831)
- Quantum state tomography based on RBMs (Torlai, Mazzola, Carrasquilla, Troyer, Melko, Carleo, 1703.05334)
- ML based decoders for topological codes (Torlai, Melko 1610.04238, Varsamopoulos, Criger, Bertels, 1705.00857)
- Supervised Learning with Quantum-Inspired Tensor Networks (Stoudenmire, Schwab 1605.05775, Novikov, Trofimov, Oseledets, 1605.03795)
- Quantum Boltzmann machines (Amin, Andriyash, Rolfe, Kulchytskyy, Melko, 1601.02036, Kieferova, Wiebe, 1612.05204,)
- Quantum machine learning algorithms to accelerate learning (Biamonte, Wittek, Pancotti, Rebentrost, Wiebe, Lloyd, 1611.09347)

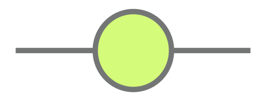
And many more

QUANTUM STATE TOMOGRAPHY

NOTATION SLIDE

S  *Scalar*

V_i  *Vector*

$W_{i,j}$  *Matrix*

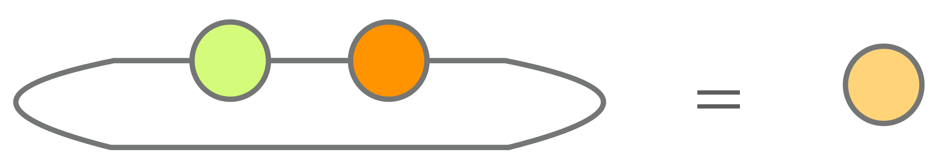
$$C_k = \sum_j W_{k,j} V_j$$

Matrix vector Multiplication

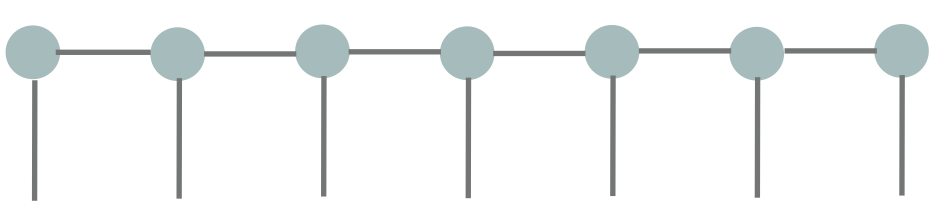
$$|C|^2 = \sum_k C_k C_k^*$$

norm

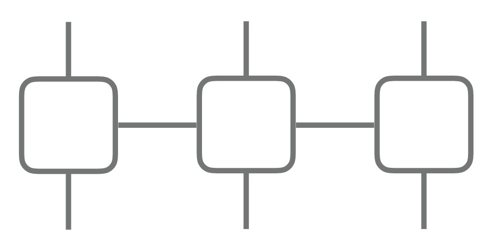
$$\mathrm{Tr}\; W M = \sum_{k,j} W_{k,j} M_{j,k}$$



Matrix Product States (MPS)



Matrix Product Operators (MPO)



QUANTUM STATE TOMOGRAPHY

Quantum state tomography is the process of reconstructing the quantum state by **measurements** on the system. It "is the gold standard for verification and benchmarking of quantum devices"*

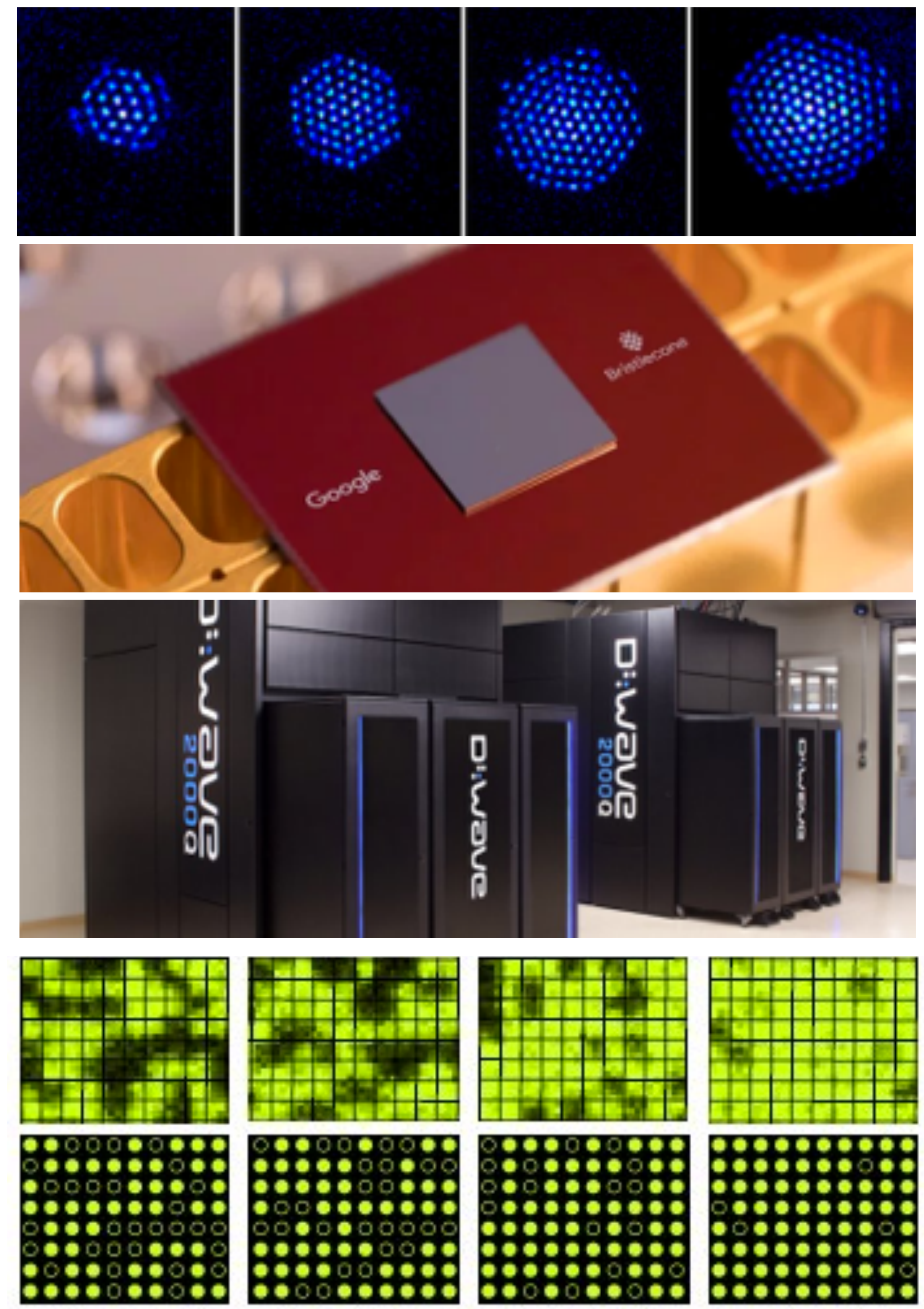
Useful for:

- Characterizing optical signals
- Diagnosing and detecting errors in state preparation, e.g. states produced by quantum computers reliably.
- Entanglement verification

* Efficient quantum state tomography. Marcus Cramer, Martin B. Plenio, Steven T. Flammia, Rolando Somma, David Gross, Stephen D. Bartlett, Olivier Landon-Cardinal, David Poulin & Yi-Kai Liu. Nature Communications volume **1**, Article number: 149

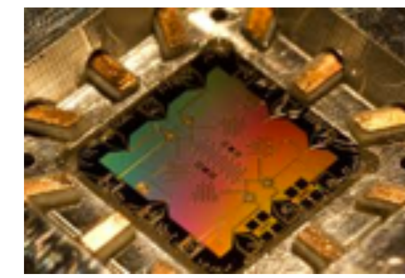
QUANTUM STATE TOMOGRAPHY

- 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.



QUANTUM STATE TOMOGRAPHY

Required ingredients



- A quantum system that can be prepared repeatedly
- Set of **measurements**.
- A training procedure and a model (full density matrix, MPS, MPO, neural networks)
- Certification

A typical state tomography protocol prepares many copies of ρ which are measured in diverse ways, and finally the outcomes of those measurements (data) are analyzed to produce an estimate ρ*.

* Efficient quantum state tomography. Marcus Cramer, Martin B. Plenio, Steven T. Flammia, Rolando Somma, David Gross, Stephen D. Bartlett, Olivier Landon-Cardinal, David Poulin & Yi-Kai Liu. Nature Communications volume 1, Article number: 149

TRADITIONALLY QST REQUIRES EXPONENTIAL RESOURCES

Examples

- Linear inversion method requires inverting a large matrix and results in an explicit representation of the density matrix

$P(\mathbf{a}) = \mathrm{Tr}\,\rho M^{\mathbf{a}}$ experimental histogram $\rightarrow$ $p_{\text{hist}}(a)$

$$\hat{\rho} = \sum_{a,a'} T^{-1}_{a,a'} p_{\text{hist}}(a') M^{(a)}$$

- Issues: Exponential scaling both in the representation and inversion problem
- Potentially unphysical density matrices $\hat{\rho}$

On the Measurement of Qubits. Daniel F. V. James, Paul G. Kwiat, William J. Munro, Andrew G. White. Physical Review A 64, 052312 (2001)

TRADITIONALLY QST REQUIRES EXPONENTIAL RESOURCES

Examples

- Maximum likelihood estimation. Requires an explicit "physical" density matrix → representation scales poorly

$$L(\hat{\rho}) = \prod_a P(a)^{f_a}$$

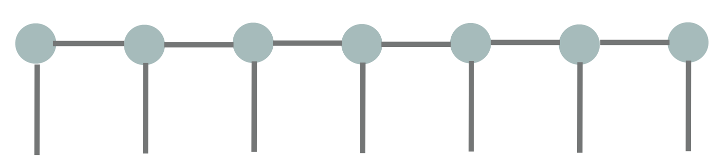
Maximize probability of observed data with respect to a parametrization of $\hat{\rho}$

- Issues: Exponential scaling in the parametrization
- Estimation of errors due to finite statistics in the measurements is difficult

On the Measurement of Qubits. Daniel F. V. James, Paul G. Kwiat, William J. Munro, Andrew G. White. Physical Review A 64, 052312 (2001)

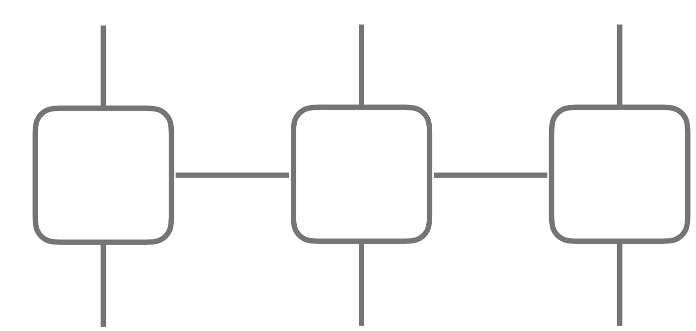
HOW TO MAKE QST EFFICIENT?

- Introduce a parametrization of the quantum state with good scaling if non-trivial structural information on the quantum systems under consideration is utilized: MPS[1] and MPO[2] tomography



[1] Efficient quantum state tomography. Marcus Cramer, Martin B. Plenio, Steven T. Flammia, Rolando Somma, David Gross, Stephen D. Bartlett, Olivier Landon-Cardinal, David Poulin & Yi-Kai Liu. Nature Communications volume **1**, Article number: 149

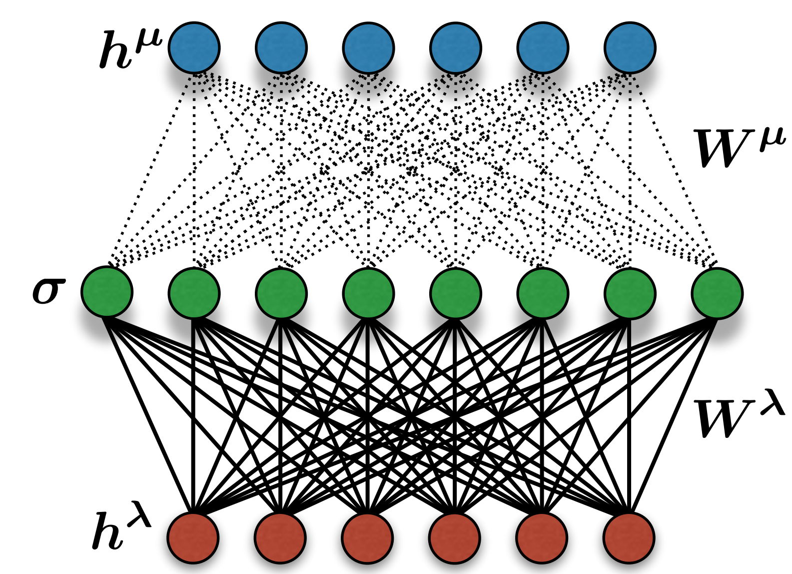
[2] A scalable maximum likelihood method for quantum state tomography T Baumgratz1, A Nüßeler, M Cramer and M B Plenio. New Journal of Physics, Volume 15, December 2013



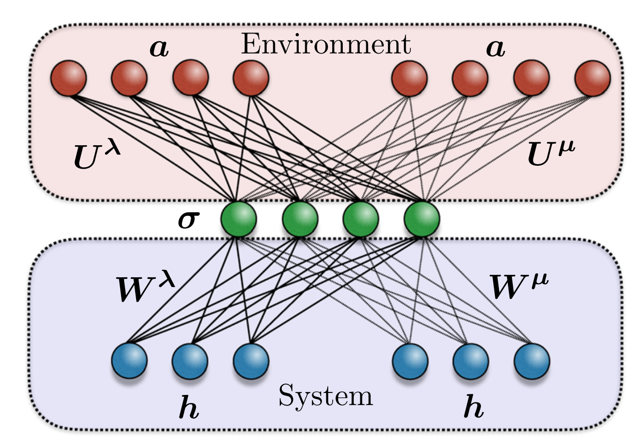
HOW TO MAKE QST EFFICIENT?

- Introduce a parametrization of the quantum state with good scaling if non-trivial structural information on the quantum systems under consideration is utilized: Restricted Boltzmann machines both for pure[3] states and mixed states[4]

[3] Neural-network quantum state tomography. G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Nat. Phys. 14, 447 (2018).



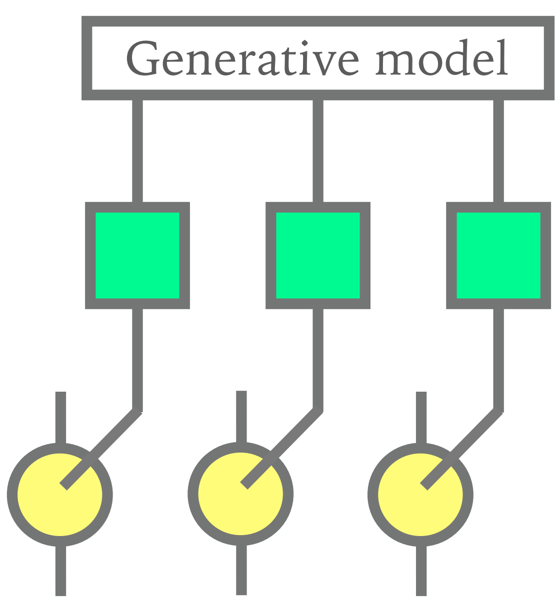
[4] Latent Space Purification via Neural Density Operators. Giacomo Torlai and Roger G. Melko. Phys. Rev. Lett. 120, 240503 (2018)



IN THIS TALK

- ~~Parametrize the quantum state~~ We parametrize the Born's rule in terms of generative models. Measurements: informationally complete positive operator valued measures (POVM)
- Use this idea to learn states from synthetic measurements mimicking experimental data

$$\rho_{\mathrm{model}} = (\mathbf{T}^{-1} P_{\mathrm{model}})^T \mathbf{M}$$



MEASUREMENTS: POSITIVE OPERATOR VALUED MEASURES (POVM)

Measurements:

$$\mathbf{M} = \{M^{(a)} \mid a \in \{1, ..., m\}\}$$

Informationally complete POVMs

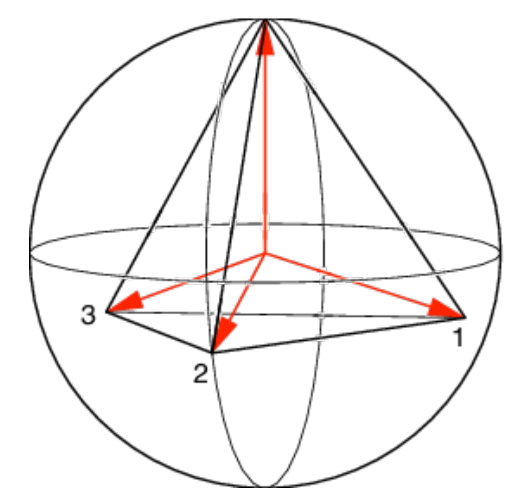
$$\sum_i M^{(a)} = \mathbb{1}$$

m is at least $D^2 = 2^{2N}$ and M should span the entire Hilbert space

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

Born rule. Defines a distribution over the generalized measurements

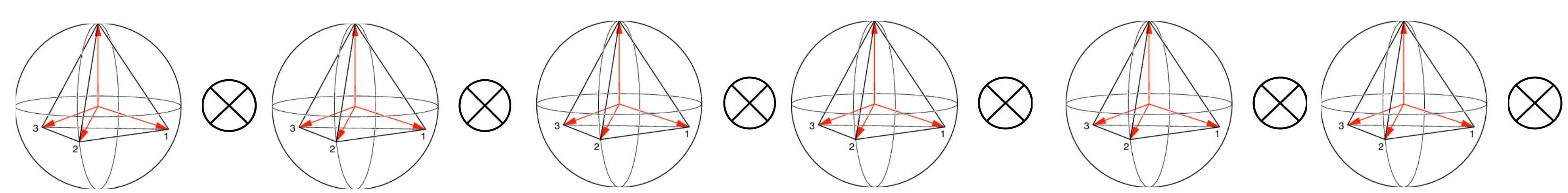
Tetrahedral measurement for one qubit



$$M^{(i)} = \frac{1}{4}\left(\mathbb{1} + \boldsymbol{s}^{(i)} \cdot \boldsymbol{\sigma}\right), \quad i = 1, ..., 4$$

$$\mathbf{s}^{(1)} = (0, 0, 1), \ \mathbf{s}^{(2)} = \left(\frac{2\sqrt{2}}{3}, 0, -\frac{1}{3}\right), \ \mathbf{s}^{(3)} = \left(-\frac{\sqrt{2}}{3}, \sqrt{\frac{2}{3}}, -\frac{1}{3}\right), \ \mathbf{s}^{(4)} = \left(-\frac{\sqrt{2}}{3}, -\sqrt{\frac{2}{3}}, -\frac{1}{3}\right)$$

For multiqubit systems



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

Experimental schemes see: Phys. Rev. A 86, 062107 (2012), Phys. Rev. A 83, 051801(R) (2011), Phys. Rev. A 91, 042101 (2015)

MEASUREMENTS: POSITIVE OPERATOR VALUED MEASURES (POVM)

Measurements: $$\mathbf{M} = \{M^{(a)} \mid a \in \{1, ..., m\}\}$$

Informationally complete POVMs

$$\sum_i M^{(a)} = \mathbb{1}$$

m is at least $D^2 = 2^{2N}$ and M should span the entire Hilbert space

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

Born rule. Defines a distribution over the generalized measurements

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 $$\boldsymbol{M}_{\text{Pauli}} \otimes \boldsymbol{M}_{\text{Pauli}} \otimes \boldsymbol{M}_{\text{Pauli}} \otimes \boldsymbol{M}_{\text{Pauli}} \otimes \boldsymbol{M}_{\text{Pauli}} \otimes \boldsymbol{M}_{\text{Pauli}} \otimes$$

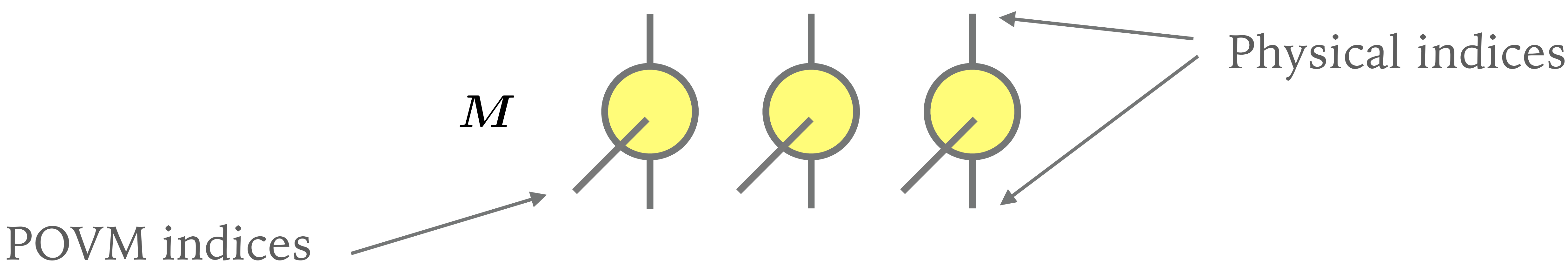
$$\boldsymbol{M} = \left\{ M^{(a_1)} \otimes M^{(a_2)} \otimes \ldots M^{(a_N)} \right\}_{a_1, \ldots a_N}$$

Experimental schemes see: Phys. Rev. A 86, 062107 (2012), Phys. Rev. A 83, 051801(R) (2011), Phys. Rev. A 91, 042101 (2015)

MEASUREMENTS: POSITIVE OPERATOR VALUED MEASURES (POVM)

Important: both types of measurements can be implemented in experiments and synthetically with numerics for MPS states

Tensor network representation of the POVM



Experimental schemes see: Phys. Rev. A 86, 062107 (2012), Phys. Rev. A 83, 051801(R) (2011), Phys. Rev. A 91, 042101 (2015)

LEARNING MIXED STATES USING POSITIVE OPERATOR VALUED MEASURE

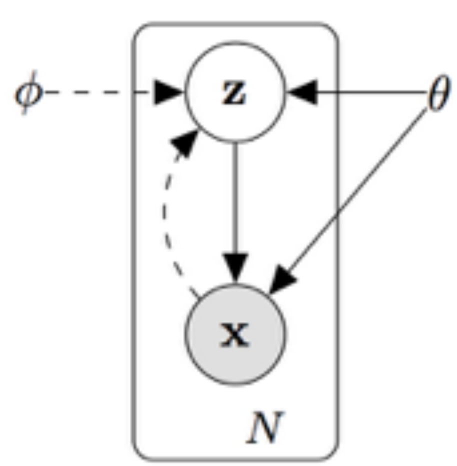
Measurements: $$\mathbf{M} = \{M^{(a)} \mid a \in \{1, ..., m\}\}$$

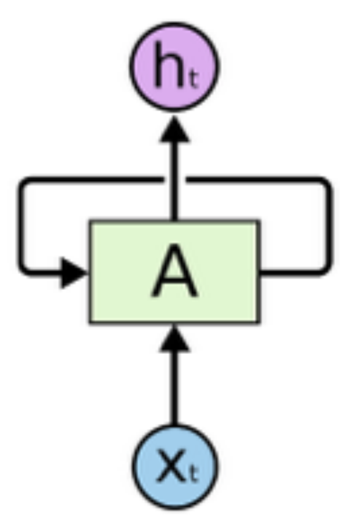
Informationally complete POVMs $$\sum_i M^{(a)} = \mathbb{1}$$

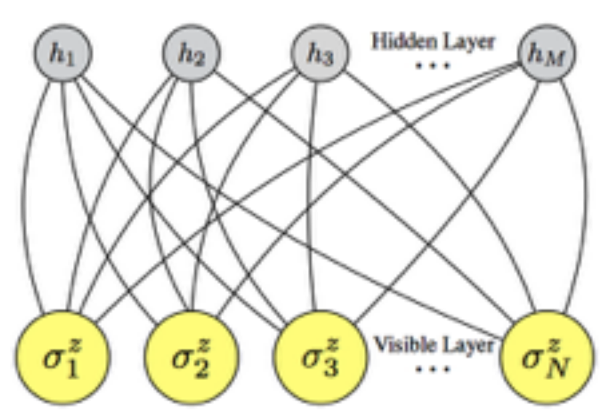
m is at least $D^2 = 2^{2N}$ and M should span the entire Hilbert space

For multiqubit systems $$P(\mathbf{a}) = \mathrm{Tr}\, \rho M^{\mathbf{a}}$$ **=> Unsupervised learning of P(a)**

$$P_{\mathrm{model}}(\mathbf{a}) \longrightarrow$$ **VAE, RBM, GAN, autoregressive models, any model**







Goodfellow et al. Deep Learning (2016)

INFORMATIONAL COMPLETENESS

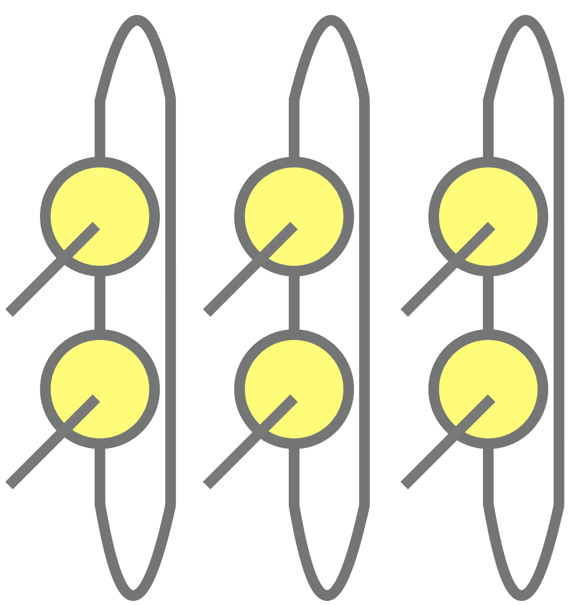
If the POVM is informationally complete then

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

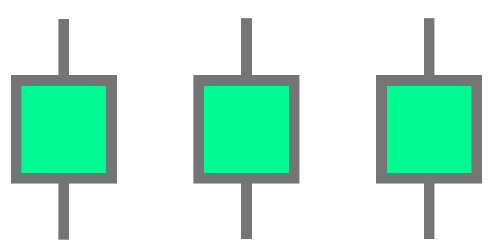
Insert this relation into Born's rule $\quad P(a) = \sum_{a'} O_\rho(a') Tr[M^{(a)} M^{(a')}] = \sum_{a'} O_\rho(a') T_{a'a}$

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

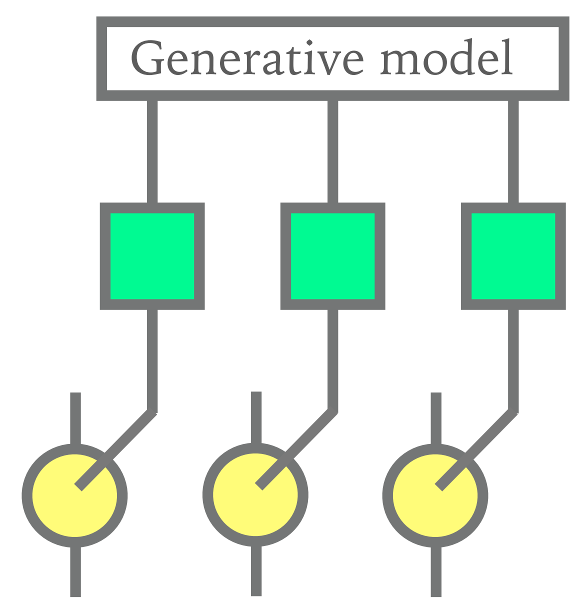
$$T_{\alpha,\beta} = \mathrm{Tr}\ M^\alpha M^\beta$$



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

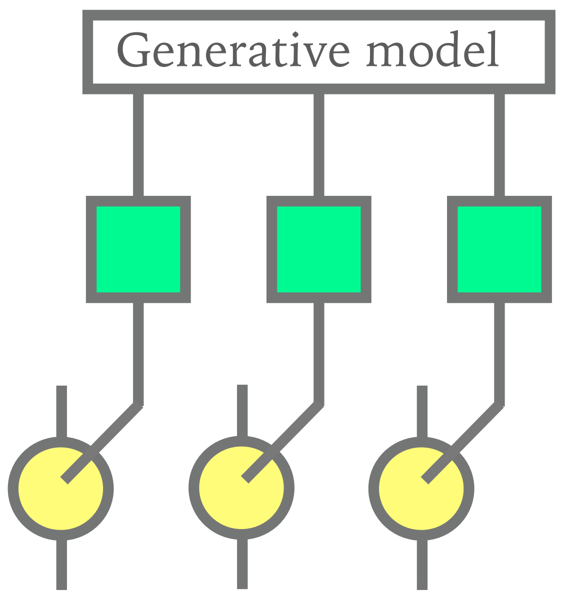


$$\rho_{\text{model}} = (\mathbf{T}^{-1} P_{\text{model}})^T \mathbf{M}$$



MODEL FOR THE DENSITY MATRIX

$$\rho_{\text{model}} = (\mathbf{T}^{-1} P_{\text{model}})^T \mathbf{M}$$



- Factorization of the state in terms of a probability distribution and a set of tensors
- All the entanglement comes from the structure of the P(a)
- P(a) is approximated by powerful neural network probabilistic models

CERTIFICATION: FIDELITY, CLASSICAL FIDELITY, CORRELATION FUNCTIONS

- Fidelity
- Classical fidelity in POVM space
- Kullback–Leibler divergence in POVM space
- Correlation functions

$$F\left(\rho,\sigma\right)=\mathrm{Tr}\left[\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right]^2$$

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

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

LEARNING MIXED STATES USING POVM: PROCEDURE

- Prepare a desired quantum state repeatedly
- Perform generalized measurements
- Reconstruct P(a) using unsupervised learning
- "Invert" $\rho = \sum_{a,a'} T^{-1}_{a,a'} P(a') M^{(a)}$
- Perform some sort of certification via fidelity or classical fidelity or measure correlation functions etc.

Combines the idea of MLE with linear inversion in one method (observables/certification)

RESULTS

GHZ + NOISE, GROUND STATES OF QUANTUM SYSTEMS APPROXIMATED BY MPS STATES

RESULTS ON SYNTHETIC DATASETS FOR GHZ STATES

Pure GHZ

$$|\Psi_0\rangle \equiv \alpha\,|0\rangle^{\otimes N} + \beta\,|1\rangle^{\otimes N}$$

$$\begin{aligned}\varrho_0 &\coloneqq |\Psi_0\rangle\langle\Psi_0| \\ &= |\alpha|^2\,|0\rangle\langle 0|^{\otimes N} + |\beta|^2\,|1\rangle\langle 1|^{\otimes N} + \left(\alpha\beta^*\,|0\rangle\langle 1|^{\otimes N} + \text{ h.c.}\right)\end{aligned}$$

GHZ with

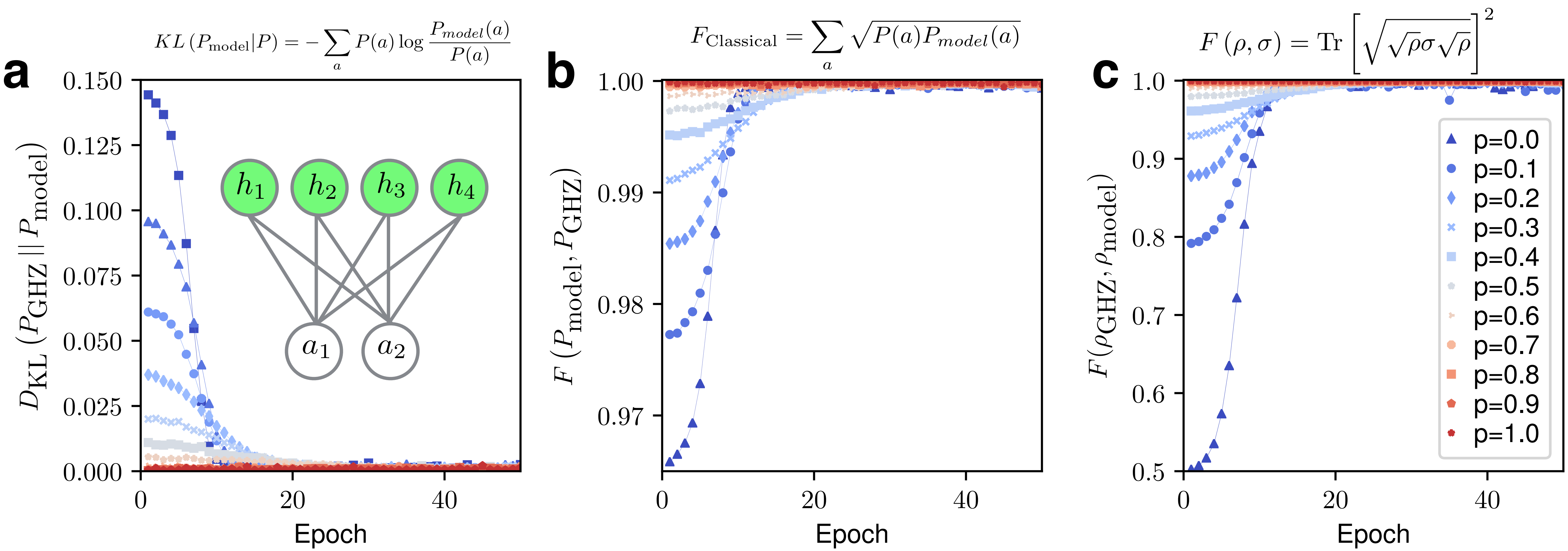
$$\mathcal{E}_i\varrho_0 = (1-p)\,\varrho_0 + \frac{p}{3}\left(\sigma_i^{(1)}\,\varrho_0\,\sigma_i^{(1)} + \sigma_i^{(2)}\,\varrho_0\,\sigma_i^{(2)} + \sigma_i^{(3)}\,\varrho_0\,\sigma_i^{(3)}\right)$$

Locally depolarized

generalized GHZ

states : a model of a decohering qubit where with probability 1 − p the qubit remains intact, while with probability p an “error” occurs.

LEARNING 2 QUBIT LOCALLY DEPOLARIZED GHZ



This result is obtained by parametrizing the P(a) with an RBM with multinomial visible units (4 states per qubit for the corresponding to the outcomes of the POVMs)

$$v_i = 0, 1, 2, 3, 4$$

$$h_j = 0, 1$$

Restricted Boltzmann Machines for Collaborative Filtering. Salakhutdinov, Mnih, Hinton. 2007

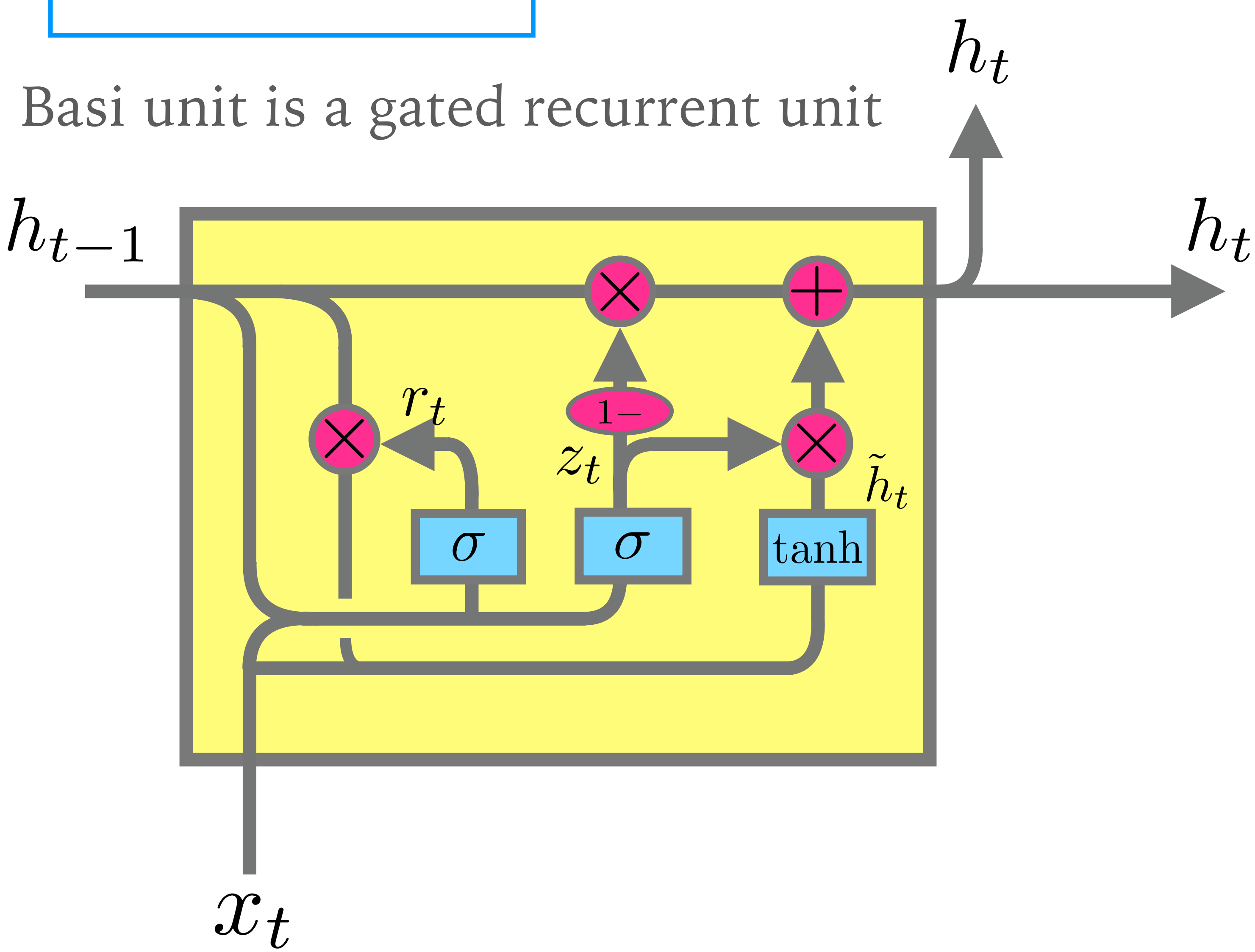
RECURRENT NEURAL NETWORK MODEL AND RESULTS

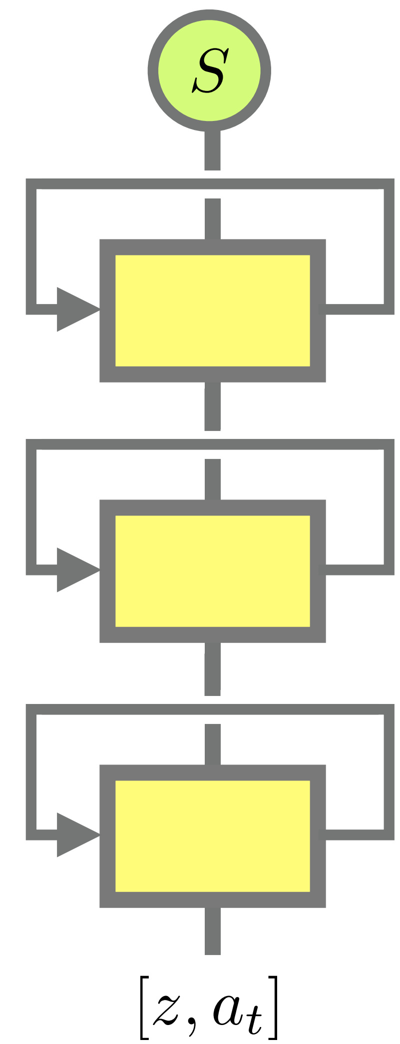
GHZ results

40 qubit p=0
Fc = 0.9992(4)
80 qubits p =0.01
Fc = 0.9988(1)
KL = 0.0050(2)

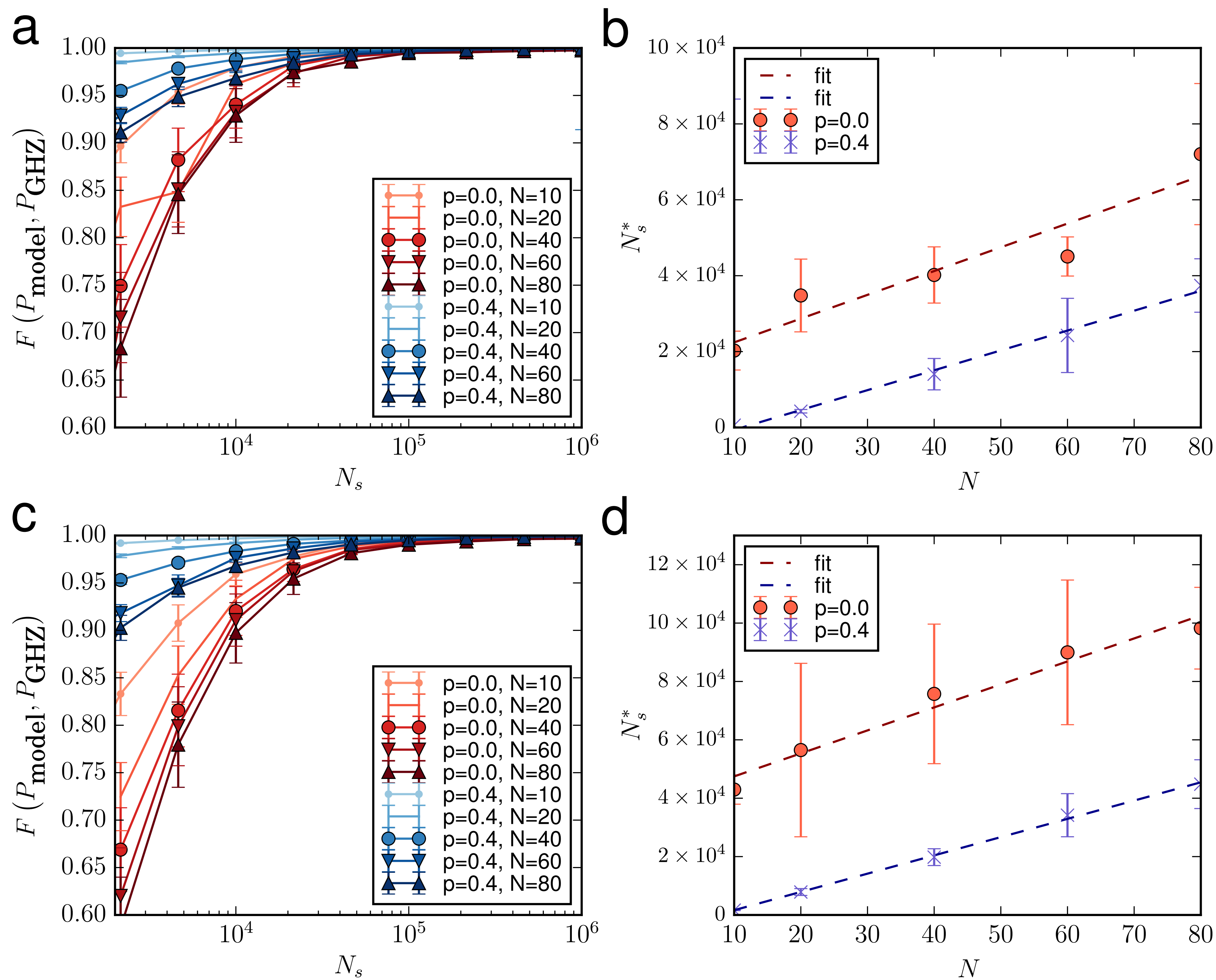
Full model stacks three of these units and adds a softmax dense layer at each time step

Basi unit is a gated recurrent unit



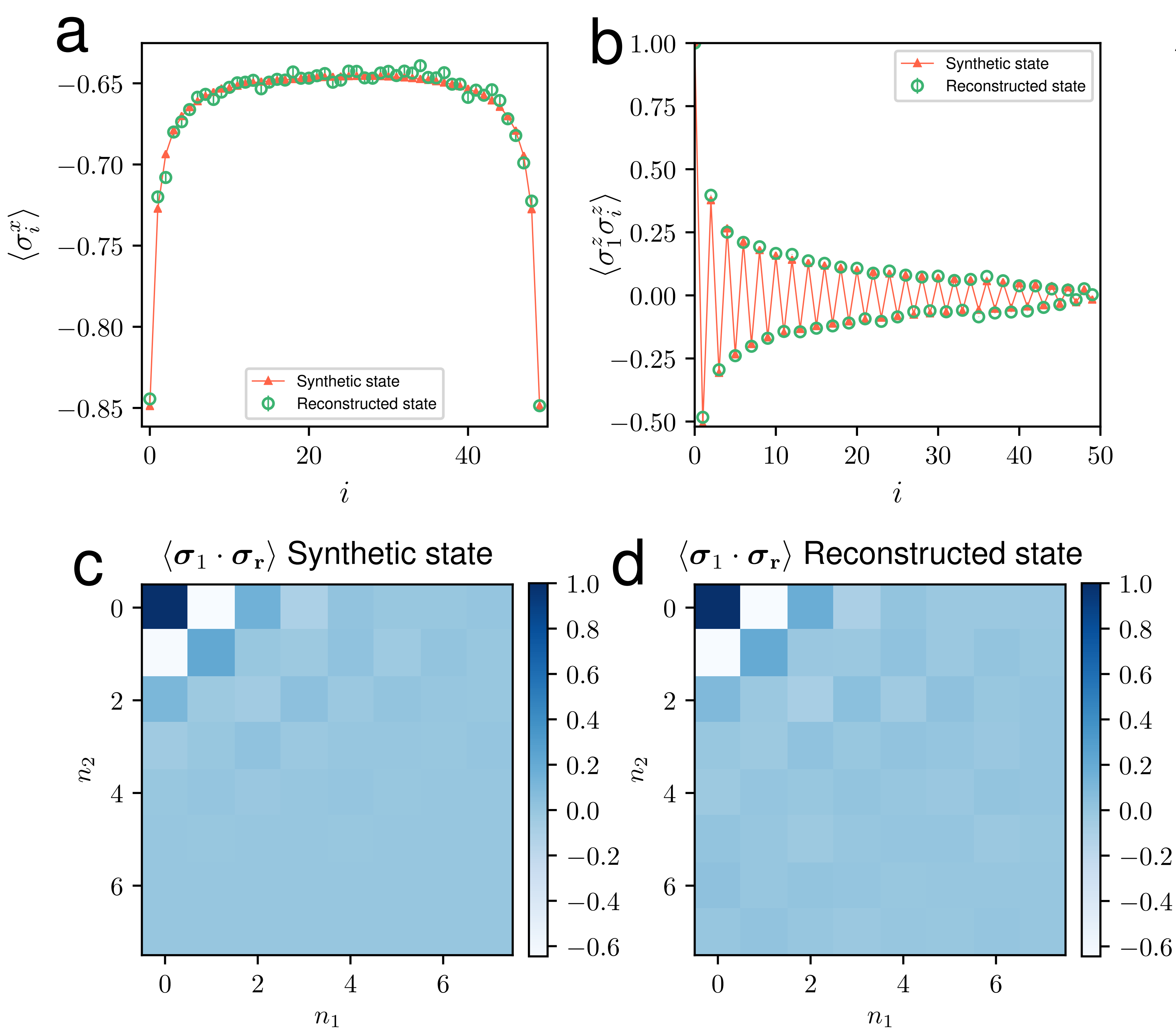


NUMERICAL INVESTIGATION OF THE SAMPLE COMPLEXITY OF LEARNING



Favorable scaling!

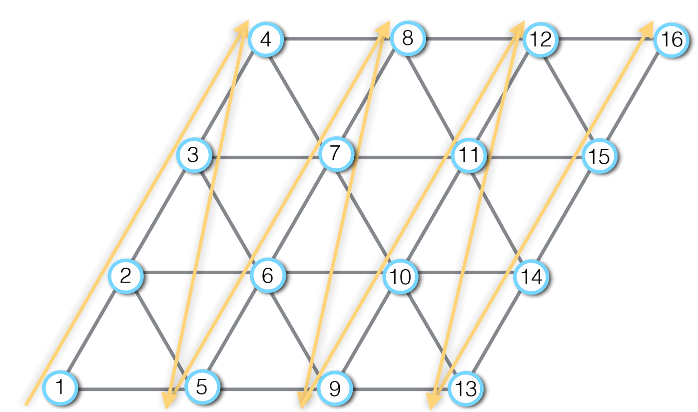
GROUND STATES OF LOCAL HAMILTONIANS



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

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

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



DISCUSSION

- How to develop intuition for which generative model performs best given a certain physical situation: spatial dimensionality? Entanglement?
- Why RNN is still so effective in 2 dimension? Quantum states are easy?
- Regularization issues with our ansatz (see next slide)

REGULARIZATION ISSUES

Requisite to specifying a quantum state through P(a), one must have an understanding of the allowed probabilities since not every choice of P(a) will give rise to positive semi-definite density matrix.

Requires regularization that is tricky to perform for large systems

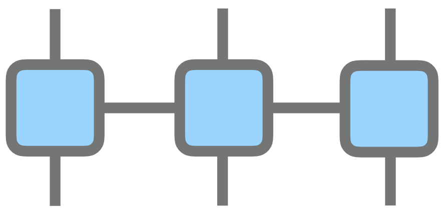
For this tomography problem in particular we just use a lot of data

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

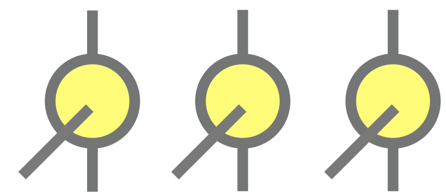
FOR LARGER GHZ STATES USE MPS/MPO

GENERATING **SYNTHETIC** DATA FOR BIG SYSTEMS EFFICIENTLY USING TENSOR NETWORKS

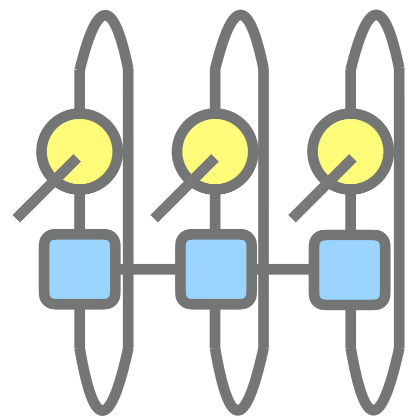
ρ



$\mathbf{M}$

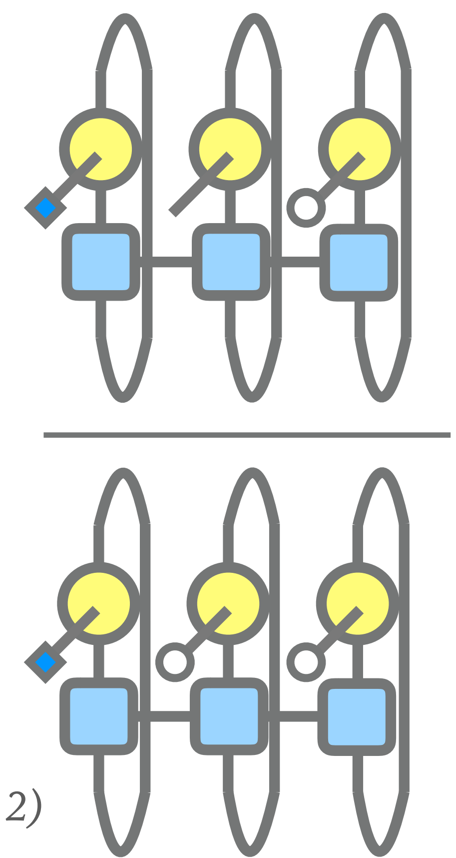


$$P\left(\mathbf{a}\right) = \mathrm{Tr}\, \rho M^{\mathbf{a}}$$



Can be sampled in polynomial time exactly (zipper algorithm)

$$P(a_i|a_{<i}) = \frac{\sum_{a_{j>i}} P(\mathbf{a})}{\sum_{a_{j\geq i}} P(\mathbf{a})}$$

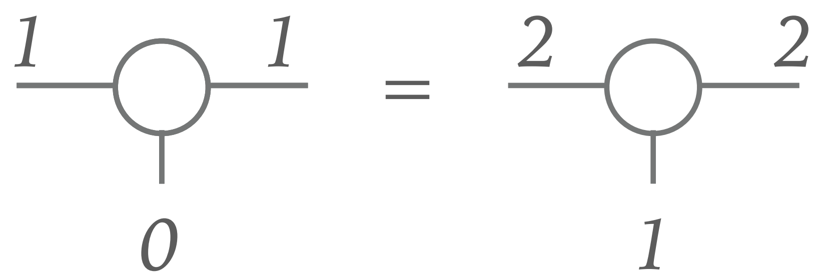


$$P(a_2|a_1)$$

AJ Ferris and G. Vidal. PRB 85, 165146 (2012)

FOR LARGER GHZ STATES USE MPS/MPO

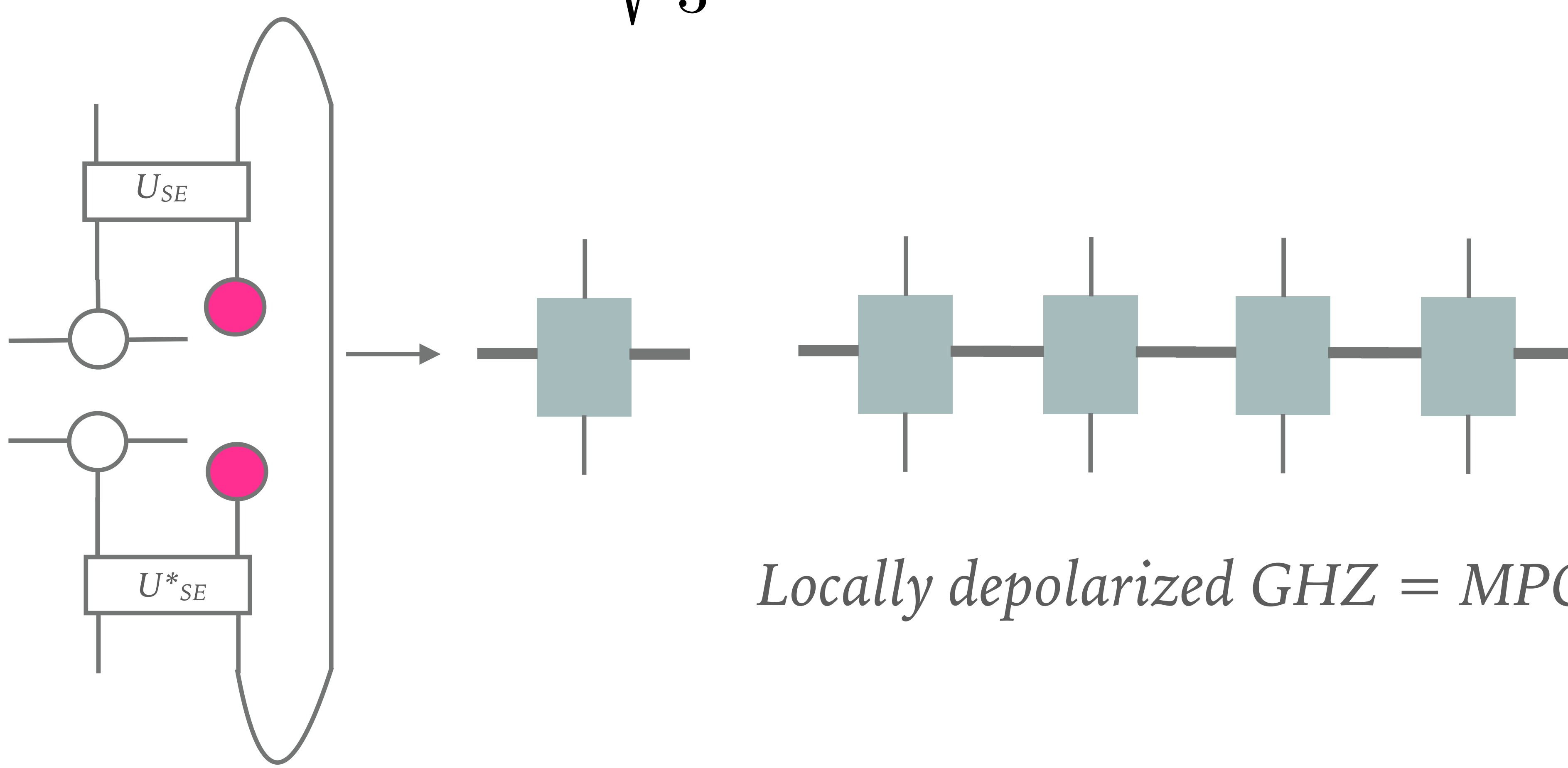
Pure GHZ = MPS



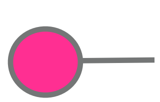
$= 2^{-1/(2N)}$

Locally depolarized GHZ

$$U_{SE} = \mathbb{1}_S \otimes |0\rangle\langle 0|_E + \sqrt{\frac{p}{3}}\,(\sigma_S^x \otimes |1\rangle\langle 0|_E + \sigma_S^y \otimes |2\rangle\langle 0|_E + \sigma_S^z \otimes |3\rangle\langle 0|_E)$$



Locally depolarized GHZ = MPO



$= (1\ 0\ 0\ 0)$

Preskill's notes http://www.theory.caltech.edu/people/preskill/ph219/

DEPOLARIZING CHANNEL

3.4.1 Depolarizing channel

The *depolarizing channel* is a model of a decohering qubit that has particularly nice symmetry properties. We can describe it by saying that, with probability $1-p$ the qubit remains intact, while with probability p an "error" occurs. The error can be of any one of three types, where each type of error is equally likely. If $\{|0\rangle, |1\rangle\}$ is an orthonormal basis for the qubit, the three types of errors can be characterized as:

1. Bit flip error: $\begin{smallmatrix}|0\rangle\mapsto|1\rangle \\ |1\rangle\mapsto|0\rangle\end{smallmatrix}$ or $|\psi\rangle \mapsto \boldsymbol{\sigma}_1|\psi\rangle, \boldsymbol{\sigma}_1 = \left(\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right)$,
2. Phase flip error: $\begin{smallmatrix}|0\rangle\mapsto|0\rangle \\ |1\rangle\mapsto-|1\rangle\end{smallmatrix}$ or $|\psi\rangle \mapsto \boldsymbol{\sigma}_3|\psi\rangle, \boldsymbol{\sigma}_3 = \left(\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right)$,
3. Both: $\begin{smallmatrix}|0\rangle\mapsto+i|1\rangle \\ |1\rangle\mapsto-i|0\rangle\end{smallmatrix}$ or $|\psi\rangle \mapsto \boldsymbol{\sigma}_2|\psi\rangle, \boldsymbol{\sigma}_2 = \left(\begin{smallmatrix}0 & -i\\ i & 0\end{smallmatrix}\right)$.

If an error occurs, then $|\psi\rangle$ evolves to an ensemble of the three states $\boldsymbol{\sigma}_1|\psi\rangle, \boldsymbol{\sigma}_2|\psi\rangle, \boldsymbol{\sigma}_3|\psi\rangle$, all occurring with equal likelihood.

Unitary representation. The depolarizing channel mapping qubit A to A can be realized by an isometry mapping A to AE, where E is a four-dimensional environment, acting as

$$\begin{aligned}\boldsymbol{U}_{A\to AE} : |\psi\rangle_A \mapsto & \sqrt{1-p}\ |\psi\rangle_A \otimes |0\rangle_E \\ & + \sqrt{\frac{p}{3}}\left(\boldsymbol{\sigma}_1|\psi\rangle_A \otimes |1\rangle_E + \boldsymbol{\sigma}_2|\psi\rangle_A \otimes |2\rangle_E + \boldsymbol{\sigma}_3|\psi\rangle_A \otimes |3\rangle_E\right).\end{aligned} \tag{3.83}$$

Preskill's notes http://www.theory.caltech.edu/people/preskill/ph219/

NEUMARK'S DILATION THEOREM

How one can implement a general quantum measurement by performing a unitary on the system of interest and an ancilla, followed by a von Neumann measurement of the ancilla.

$$P(a) = \mathrm{Tr}\rho M^{(a)} = \mathrm{Tr}\left[\mathbf{I}_s \otimes |a\rangle\langle a|_p U_{sp}\rho \otimes |0\rangle\langle 0|_p U_{sp}^{\dagger}\right]$$

- Requires projective measurements in only one basis set

Experimental schemes see: Phys. Rev. A 86, 062107 (2012), Phys. Rev. A 83, 051801(R) (2011), Phys. Rev. A 91, 042101 (2015)