Title: Machine Learning Lecture

Speakers: Roger Melko

Collection: Machine Learning 2023/24

Date: April 25, 2024 - 11:30 AM

URL: https://pirsa.org/24040053

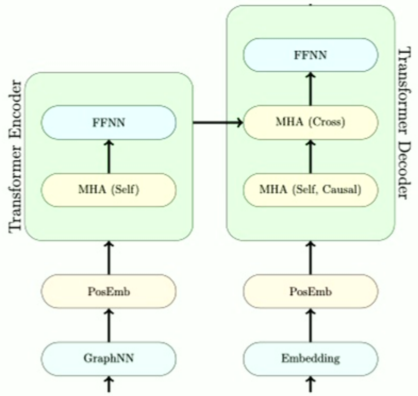
Machine Learning for Many-Body Physics 2023/2024

Roger Melko

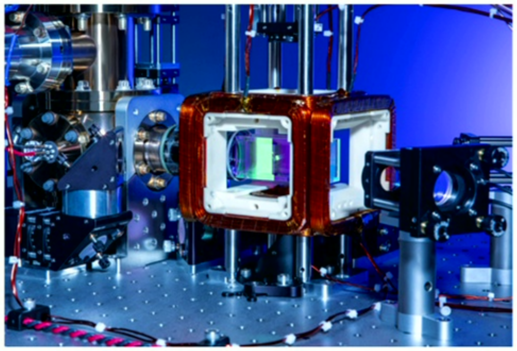
UNIVERSITY OF WATERLOO

QUANTUM INTELLIGENCE LAB

Outline

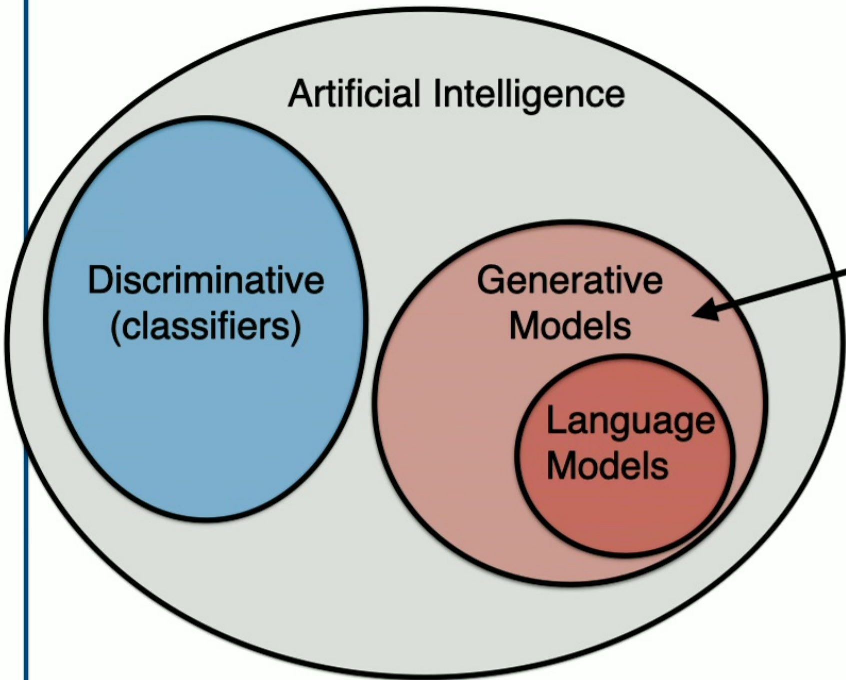


Generative models and Language models



Applications to quantum simulation & beyond

The age of generative models in AI



`/imagine a stormtrooper family in the style of Norman Rockwell`

Midjourney:

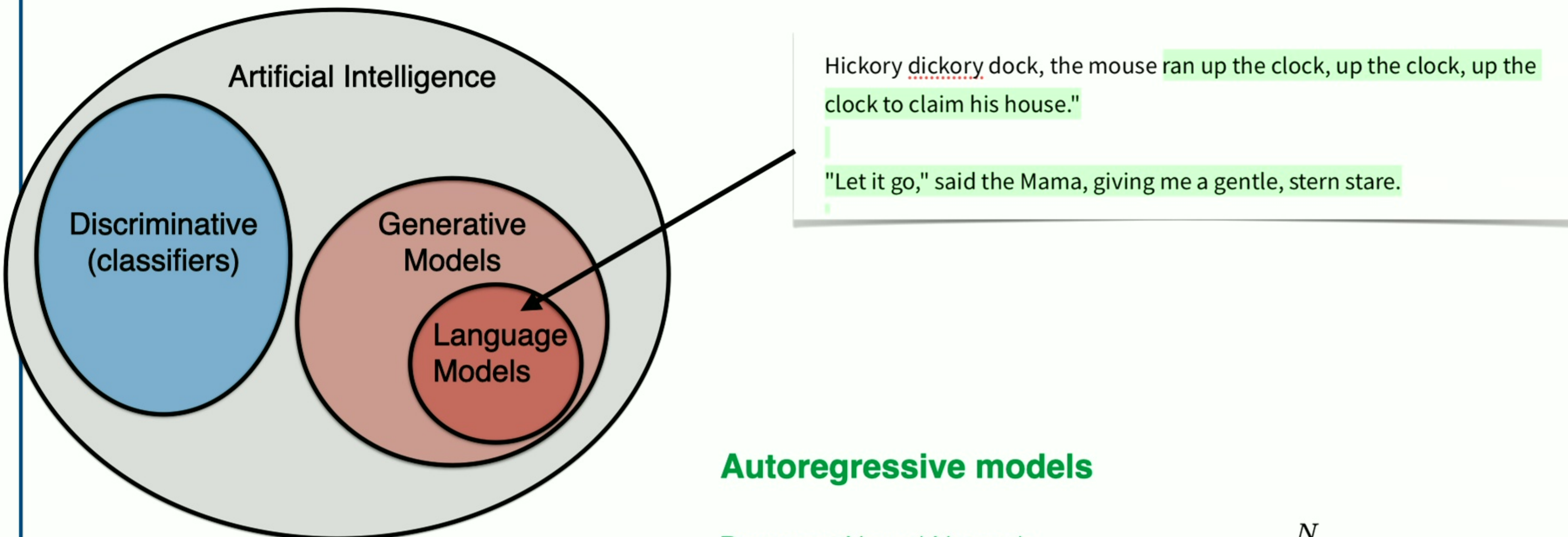
Diffusion models
Stable diffusion

Energy-based models

$$p(\mathbf{x}) = \frac{1}{Z} e^{-E(\mathbf{x})}$$

Hopfield networks
Restricted Boltzmann machines

The age of generative models in AI



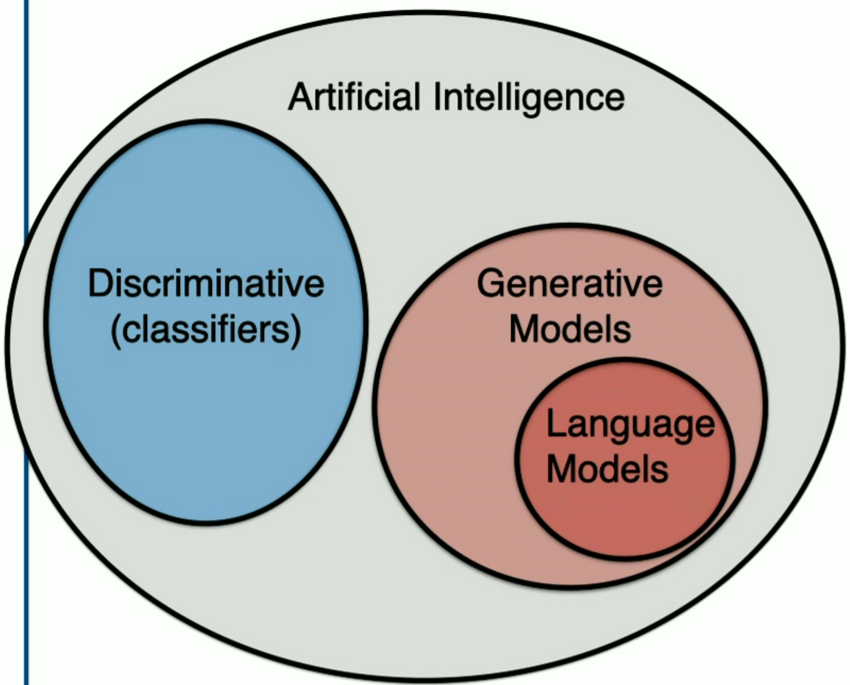
Autoregressive models

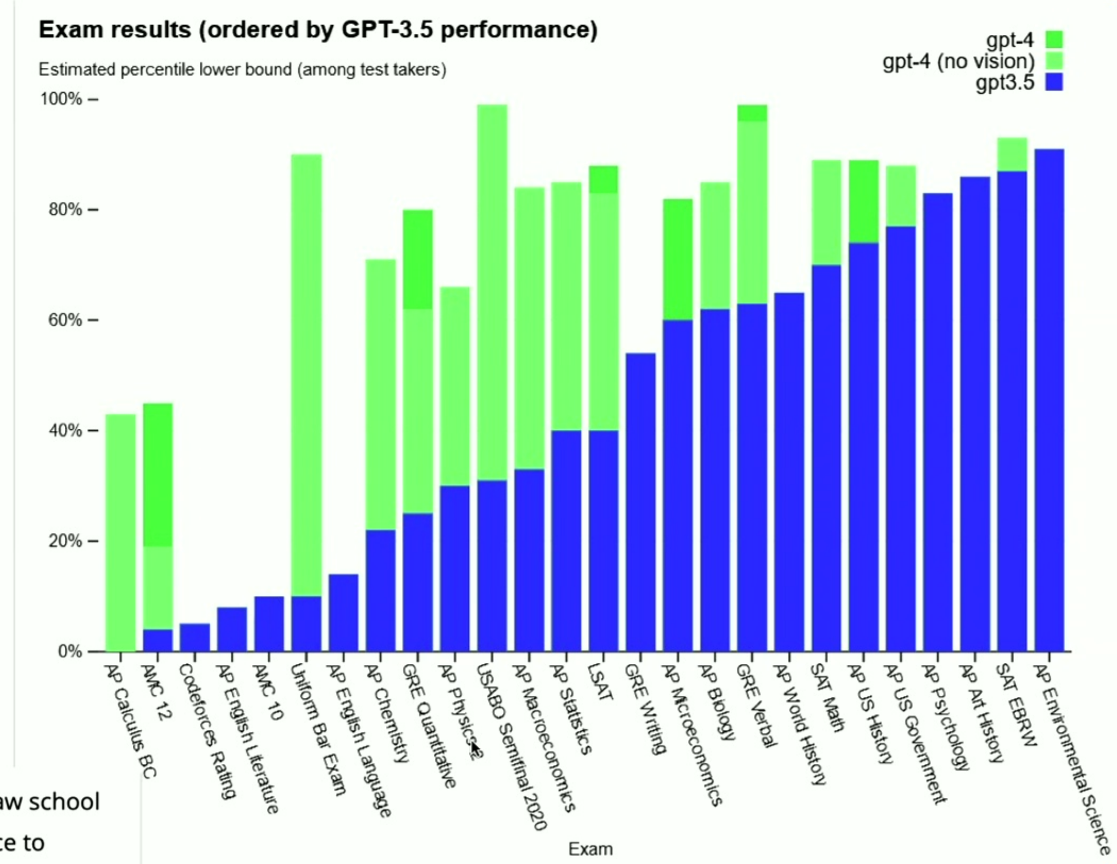
Recurrent Neural Network,
LSTM, GRU
Transformer
GPT

$$p(\mathbf{x}) = \prod_{i=1}^{N} p(x_i | x_{<i})$$

The age of generative models in AI

https://cdn.openai.com/papers/gpt-4.pdf

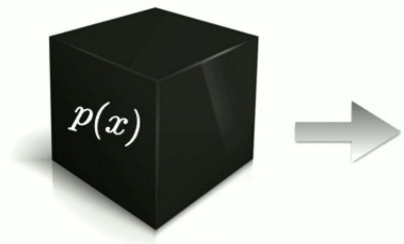




GPT4's score would put it in a good position to be admitted into a top 20 law school and is only a few marks short of the reported scores needed for acceptance to prestigious schools such as Harvard, Stanford, Princeton or Yale.

Generative models: density estimation

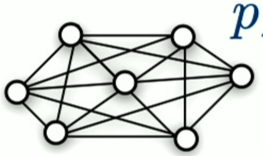
In the typical setting, generative models *learn* to approximate a target probability distribution that underlies a dataset:



$$\mathbf{x}_1 = (1,0,0,1,1,1,0,0,0,0,\cdots,1)$$
$$\mathbf{x}_2 = (1,1,1,0,1,1,0,1,1,1,\cdots,1)$$
$$\mathbf{x}_3 = (0,1,1,0,0,1,0,1,0,1,\cdots,0)$$
$$\vdots$$

Use this data to *train* the optimal parameters in some representation of the unknown target distribution/state

a model:



p_λ

goal:

$$p_\lambda(\mathbf{x}) \approx p(\mathbf{x})$$

loss function:

$$\mathcal{L} = \langle \log p_\lambda(\mathbf{x}) \rangle_p$$

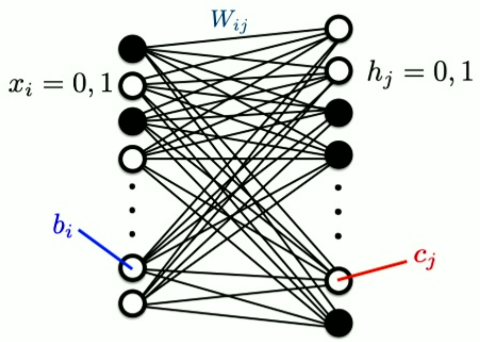
$$\boxed{\mathcal{L} = F_\lambda = \langle H_{\mathrm{target}} \rangle_\lambda - TS(p_\lambda)}$$

training:

$$\lambda' = \lambda - \eta \nabla \mathcal{L}$$

Computational physics perspectives

Energy-based (RBM)



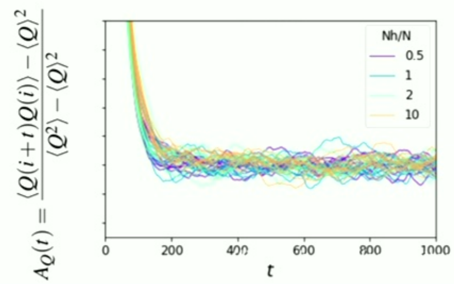
$$p_\lambda = \frac{1}{Z_\lambda} e^{-E_\lambda(\mathbf{x},\mathbf{h})}$$

$$p_\lambda(\mathbf{x}) = \sum_{\mathbf{h}} p_\lambda(\mathbf{x},\mathbf{h})$$

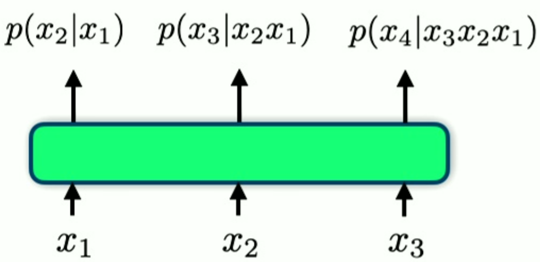
$$E_\lambda(\mathbf{x},\mathbf{h}) = -\sum_{ij} W_{ij} x_i h_j - \sum_i b_i x_i - \sum_j c_j h_j$$

$\lambda = \{W, b, c\}$ model parameters

- Requires MCMC sampling
- Generated samples are autocorrelated



Autoregressive

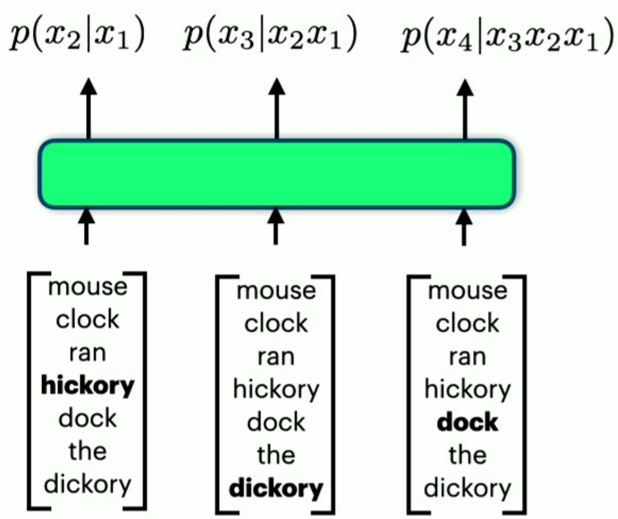


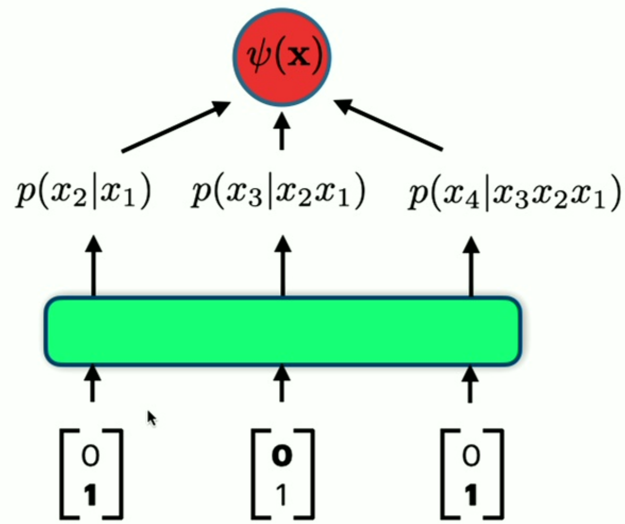
$$p_\lambda(x_1, \cdots x_N) = \prod_{i=1}^{N} p(x_i | x_1, \ldots, x_{i-1})$$

- Joint distribution is normalized
- Generated samples are uncorrelated
- Requires the definition of a sequence

Sequences: languages and qubits

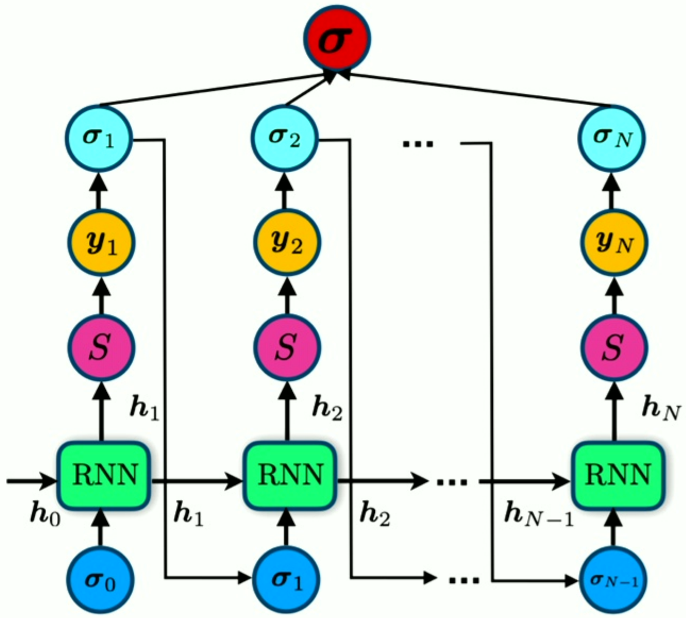
Autoregressive models produce the conditional probability of each *token* (word or qubit) in a sequence





Recurrent Neural Networks

Lipton, Berkowitz, Elkan, arXiv:1506.00019
Hibat-Allah, Ganahl, Hayward, RGM, Carrasquilla, arXiv:2002.02973



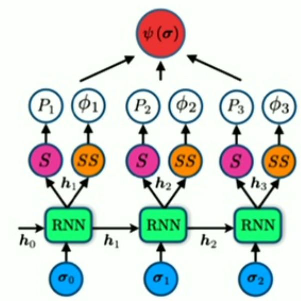
$$\boldsymbol{y}_n = \left[p(\sigma_n = 0|\boldsymbol{\sigma}_{<n}), p(\sigma_n = 1|\boldsymbol{\sigma}_{<n})\right]$$

$$\boldsymbol{y}_n \equiv S\left(U\boldsymbol{h}_n + \mathbf{c}\right) \qquad S(v_j) = \frac{\exp(v_j)}{\sum_i \exp(v_i)}$$

$$\boldsymbol{h}_n = f\left(W[\boldsymbol{h}_{n-1}; \boldsymbol{\sigma}_{n-1}] + \boldsymbol{b}\right)$$

This architecture can be adapted to learn pure & mixed quantum states

$$|\psi_\lambda\rangle = \sum_{\boldsymbol{\sigma}} e^{i\phi(\boldsymbol{\sigma})}\sqrt{p(\boldsymbol{\sigma})}|\boldsymbol{\sigma}\rangle$$



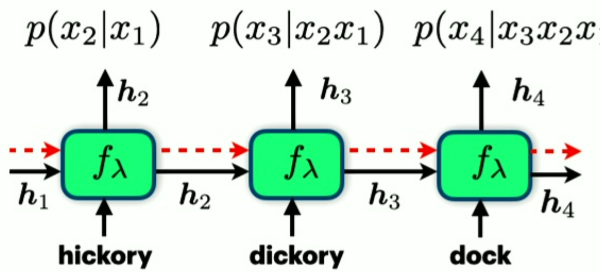
Carrasquilla, Torlai, RGM, Aolita, arXiv:1810.10584

Recurrent Neural Networks

Lipton, Berkowitz, Elkan, arXiv:1506.00019
Hibat-Allah, Ganahl, Hayward, RGM, Carrasquilla, arXiv:2002.02973

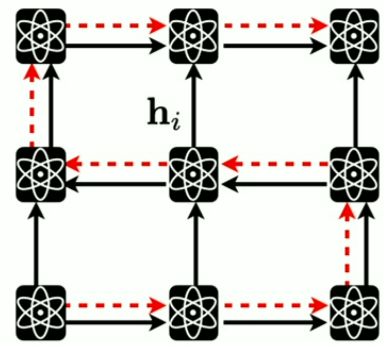
- Long-range correlations (context) passed through a hidden state vector

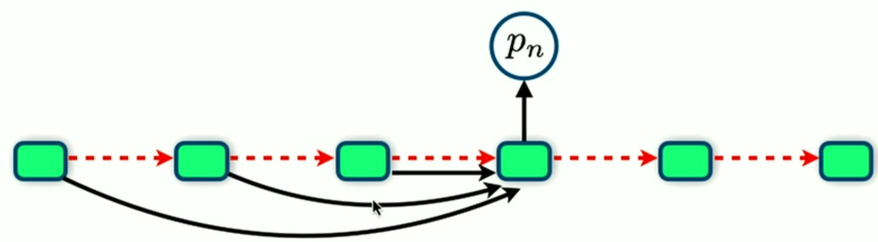
language



$$\mathbf{h}_i = f_\lambda(\mathbf{h}_{i-1}, x_{i-1})$$

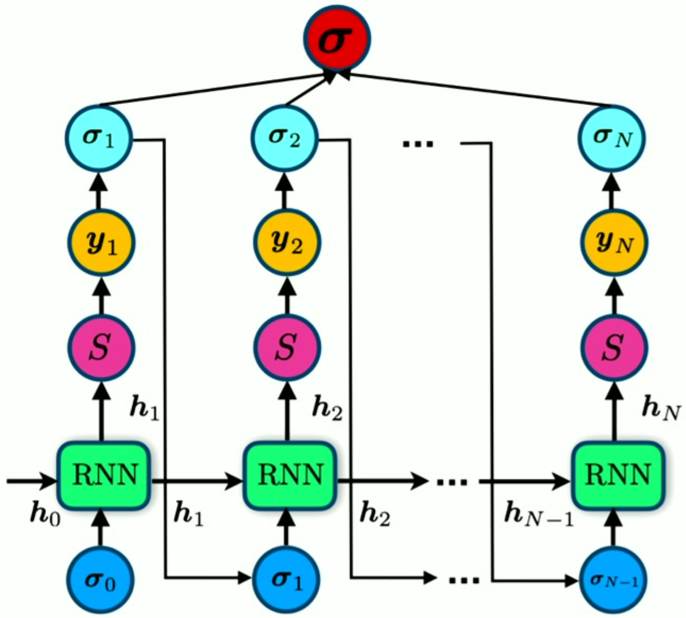
qbit arrays





Recurrent Neural Networks

Lipton, Berkowitz, Elkan, arXiv:1506.00019
Hibat-Allah, Ganahl, Hayward, RGM, Carrasquilla, arXiv:2002.02973



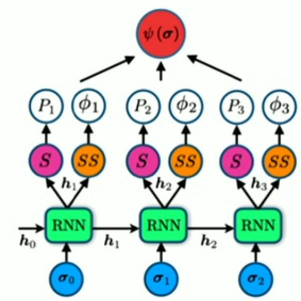
$$\boldsymbol{y}_n = \left[p(\sigma_n = 0|\boldsymbol{\sigma}_{<n}), p(\sigma_n = 1|\boldsymbol{\sigma}_{<n})\right]$$

$$\boldsymbol{y}_n \equiv S\left(U\boldsymbol{h}_n + \mathbf{c}\right) \qquad S(v_j) = \frac{\exp(v_j)}{\sum_i \exp(v_i)}$$

$$\boldsymbol{h}_n = f\left(W[\boldsymbol{h}_{n-1}; \boldsymbol{\sigma}_{n-1}] + \boldsymbol{b}\right)$$

This architecture can be adapted to learn pure & mixed quantum states

$$|\psi_\lambda\rangle = \sum_{\boldsymbol{\sigma}} e^{i\phi(\boldsymbol{\sigma})}\sqrt{p(\boldsymbol{\sigma})}|\boldsymbol{\sigma}\rangle$$



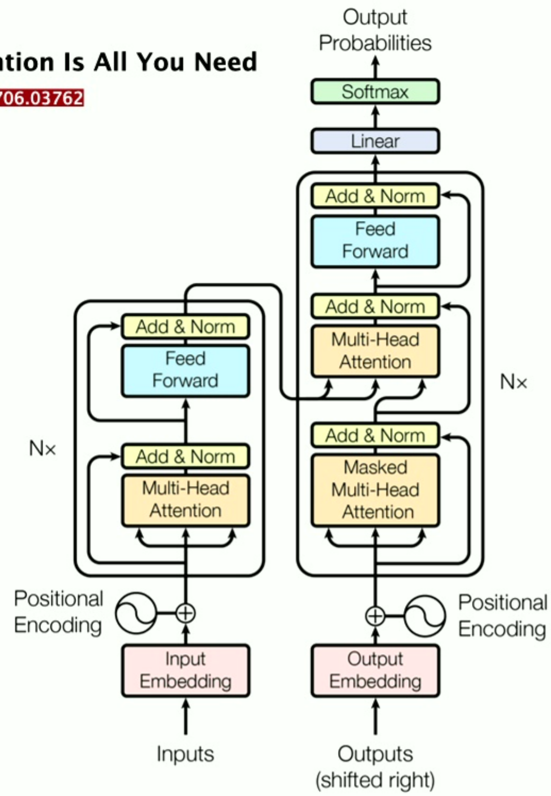
Carrasquilla, Torlai, RGM, Aolita, arXiv:1810.10584

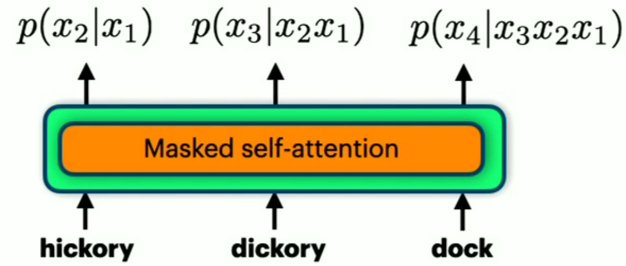
The Transformer

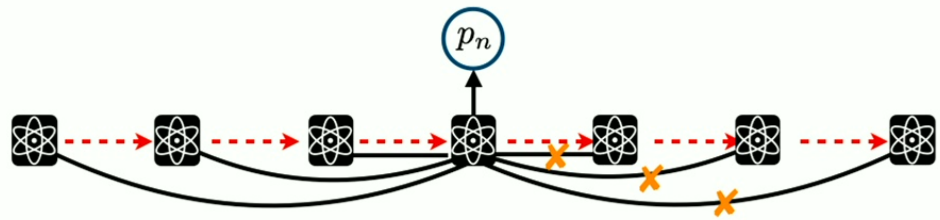
- The basis of modern Large Language Models (LLMs)

Attention Is All You Need

arXiv:1706.03762





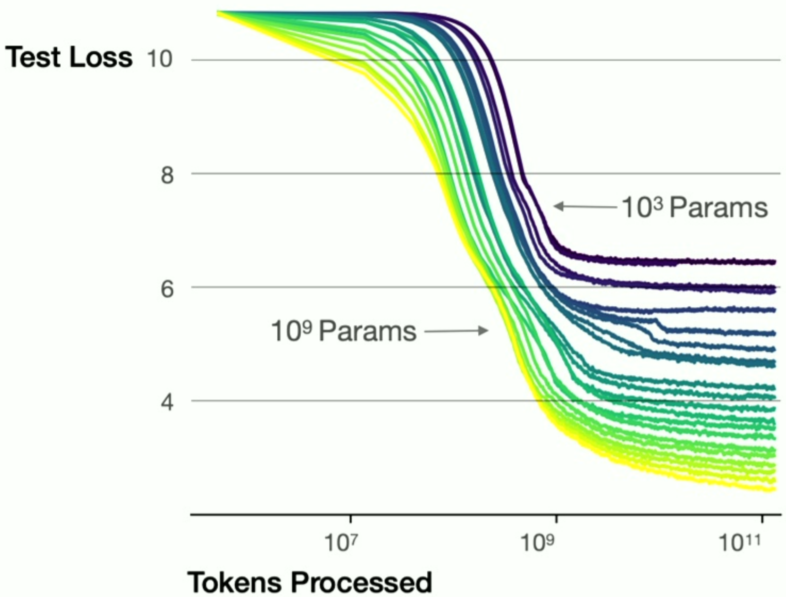


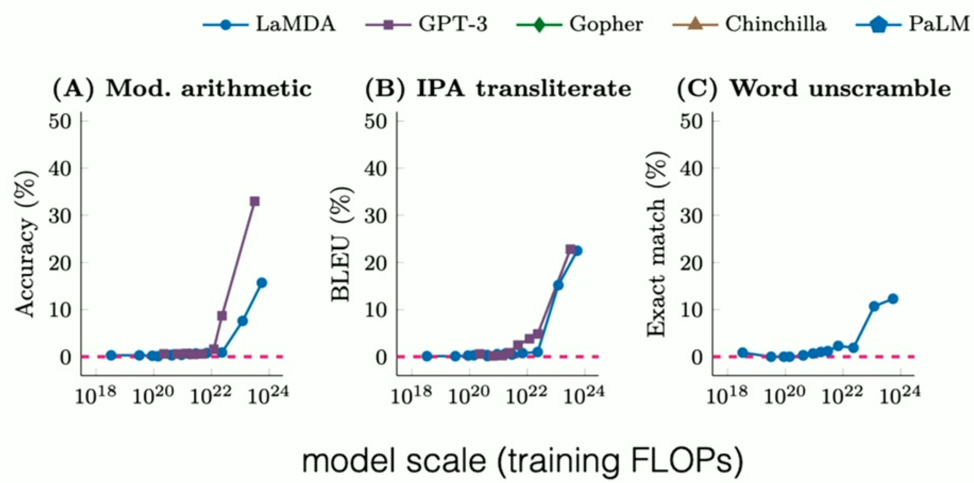
The self-attention and its correlation matrix are useful to introduce correlations between qubits separated at any distance in the quantum system. This is analogous to a two-body Jastrow factor which induces pairwise long-distance correlations between the bare degrees of freedom (i.e. spins, qubits, electrons) in a wavefunction.

Probabilistic Simulation of Quantum Circuits with the Transformer,
J. Carrasquilla, Di Luo, F. Pérez, A. Milsted, B. Clark, M. Volkovs, L. Aolita, Phys. Rev. A 104, 032610 (2021)

Transformer LLMs: scaling and emergence

Scaling Laws for Neural Language Models, arxiv:2001.08361
Emergent Abilities of Large Language Models, arXiv:2206.07682

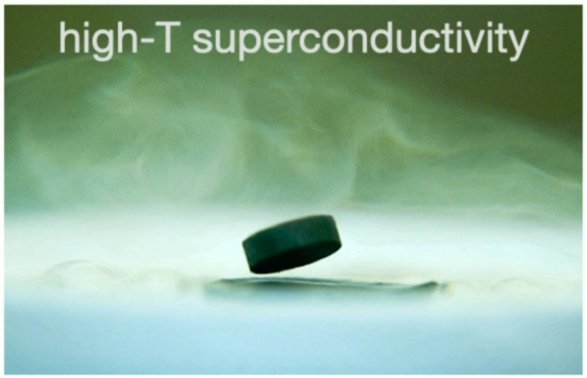


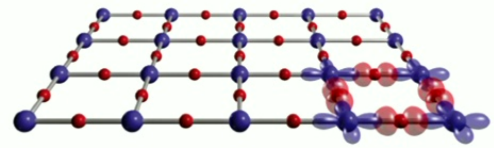


"Emergence is when quantitative changes in a system result in qualitative changes in behavior."

Steinhardt (2022) , Anderson, (1972)

The Explainability crisis in Condensed matter physics





$$H = -t \sum_{\langle ij \rangle, \sigma} (c^\dagger_{i\sigma} c_{j\sigma} + c_{i\sigma} c^\dagger_{j\sigma}) + U \sum_i n_{i\uparrow} n_{i\downarrow}$$

In the cuprates, we don't know how nature achieves the superconducting state.

Even our best calculations of the microscopic interactions are unable to recreate the emergence of the macroscopic phenomena…

Condensed matter physics and (machine) learning

Philip W. Anderson

Nobel Prize 1977
(Semiconductors, superconductivity, magnetism)

Harvard, Bell Labs, Cambridge, Princeton

The ability to reduce everything to simple fundamental laws does not imply the ability to start from those laws and reconstruct the universe

The constructionist hypothesis breaks down when confronted with the twin difficulties of scale and complexity.

Discussion: language models

Breakthroughs in scale arose from large **parameterization**, **data**, and **compute**

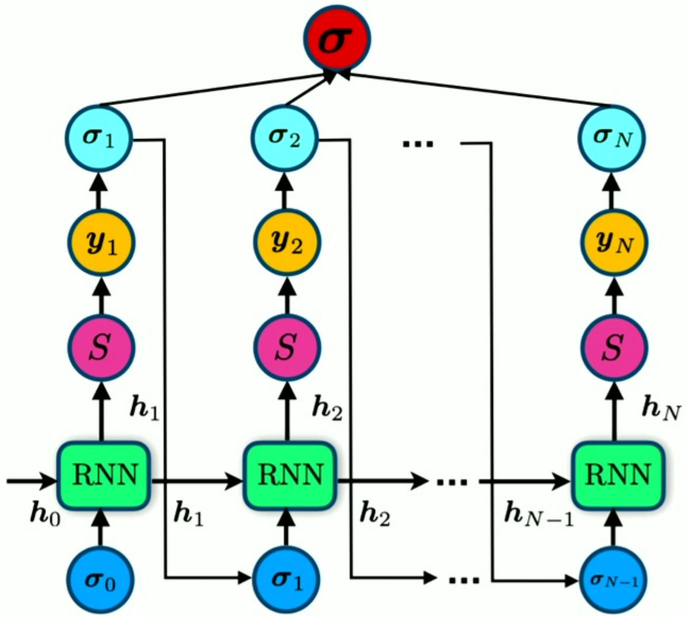
- GPT4 contains over 1 trillion parameters
- training involves an Avogadro's number of floating point operations
- costs $100M to train or more
- today, essentially the entire available internet is used for training data

Are language models poised to make breakthroughs in **quantum physics**?

- Scaling is the central challenge in building quantum computers
- A goal of quantum simulators its to demonstrate emergence
- Do we lose interpretability/explainability with scale?
- Scaling is as much an economic problem as a physics/engineering one

Recurrent Neural Networks

Lipton, Berkowitz, Elkan, arXiv:1506.00019
Hibat-Allah, Ganahl, Hayward, RGM, Carrasquilla, arXiv:2002.02973

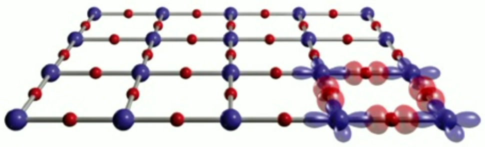


$$\boldsymbol{y}_n = \left[p(\sigma_n = 0|\boldsymbol{\sigma}_{<n}), p(\sigma_n = 1|\boldsymbol{\sigma}_{<n})\right]$$

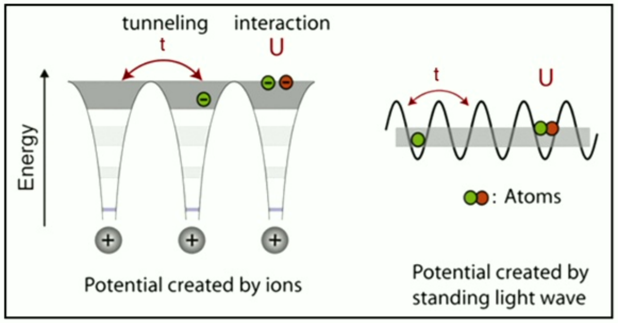
$$\boldsymbol{y}_n \equiv S\left(U\boldsymbol{h}_n + \mathbf{c}\right) \qquad S(v_j) = \frac{\exp(v_j)}{\sum_i \exp(v_i)}$$

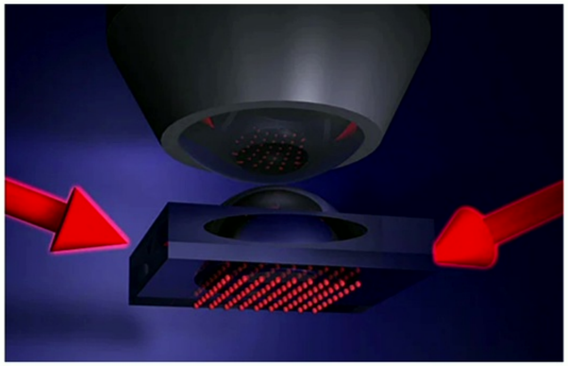
$$\boldsymbol{h}_n = f\left(W[\boldsymbol{h}_{n-1}; \boldsymbol{\sigma}_{n-1}] + \boldsymbol{b}\right)$$

Quantum "simulation" (or emulation)



$$H = -t \sum_{\langle ij \rangle, \sigma} (c^{\dagger}_{i\sigma} c_{j\sigma} + c_{i\sigma} c^{\dagger}_{j\sigma}) + U \sum_{i} n_{i\uparrow} n_{i\downarrow}$$

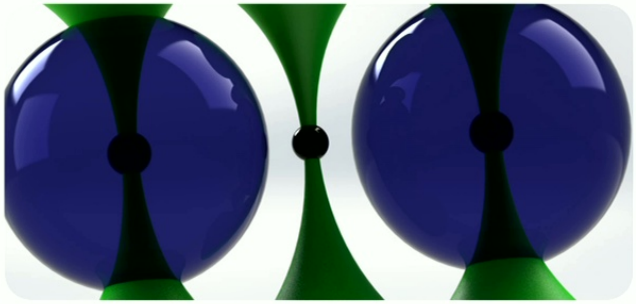




Can the emergence of unexplained macroscopic phenomena be understood by building quantum computers?

Data obtained on challenging models could be used to drastically improve variational calculations…

Rydberg atom arrays



- Neutral atoms (Rb, Sr) are loaded into a lattice formed by an array of optical tweezers

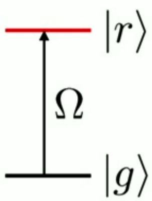
- Atoms can be in their ground state, or an excited state with a large principle quantum number (a Rydberg state). They form a strongly-interacting system.

- Single-atom resolved fluorescent imaging provides projective measurements

- Arrays of atoms are currently used for simulation (groundstates, critical phenomena), dynamics, solving combinatorial optimization problems, etc.

Rydberg Blockade Hamiltonian

Jaksch, Cirac, Zoller, Rolston, Cote, Lukin, Phys. Rev. Lett. 85, 2208 (2000)
Lukin, Fleischhauer, Cote, Duan, Jaksch, Cirac, Zoller, Phys. Rev. Lett. 87, 037901 (2001)
Fendley, Sengupta, Sachdev, Phys. Rev. B 69, 075106 (2004)



$$H = \Omega \sum_i \sigma_i^x - \Delta \sum_i n_i + \sum_{i<j} V_{ij} n_i n_j$$

$$V(R) = \frac{\Omega}{(R/R_b)^6}$$

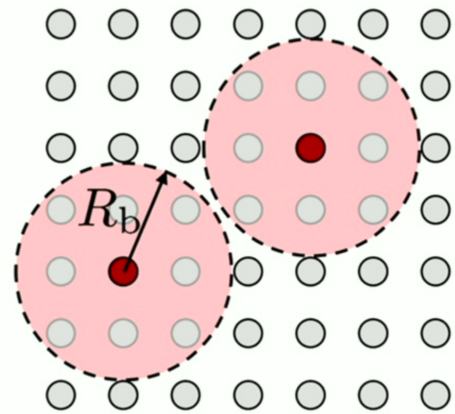
$$\sigma^x = |g\rangle\langle r| + |r\rangle\langle g| \qquad n = |r\rangle\langle r|$$

This Hamiltonian is stoquastic:

- Perron-Frobenius implies the groundstate is real positive

$$\psi_\lambda(z) = \sqrt{p_\lambda(z)}$$

- Allows for QMC with no sign problem

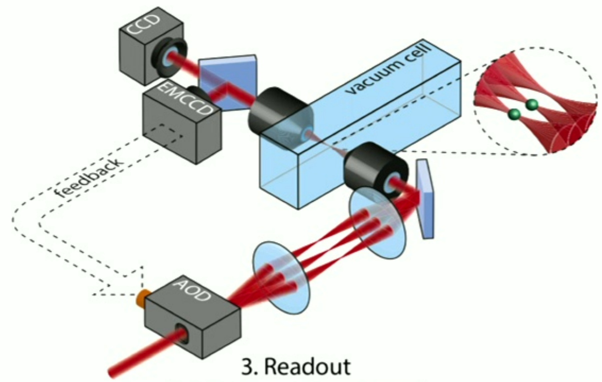


Browaeys, Lahaye, Nature Physics 16, 132 (2020)

Experimental Rydberg Arrays

Endres et. al. Science 354, 1024 (2016)

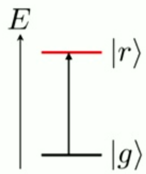
Ebadi et. al. arXiv:2012.12281
Nature 595, 227 (2021)

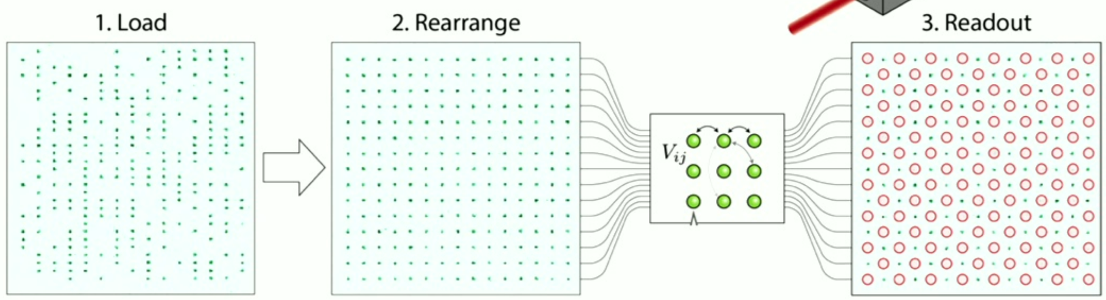


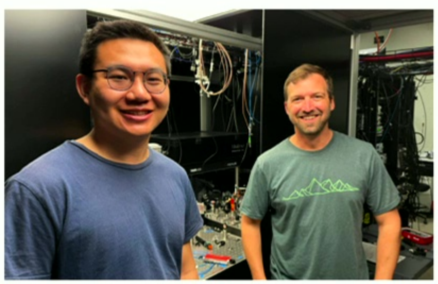
HARVARD UNIVERSITY

|QuEra>

$^{87}\mathrm{Rb}$



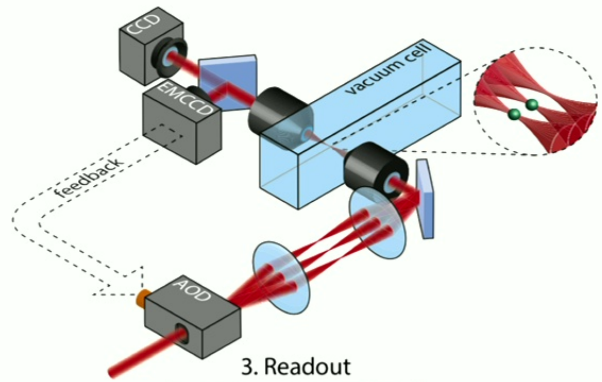




Experimental Rydberg Arrays

Endres et. al. Science 354, 1024 (2016)

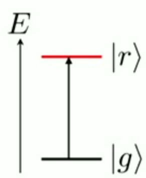
Ebadi et. al. arXiv:2012.12281
Nature 595, 227 (2021)

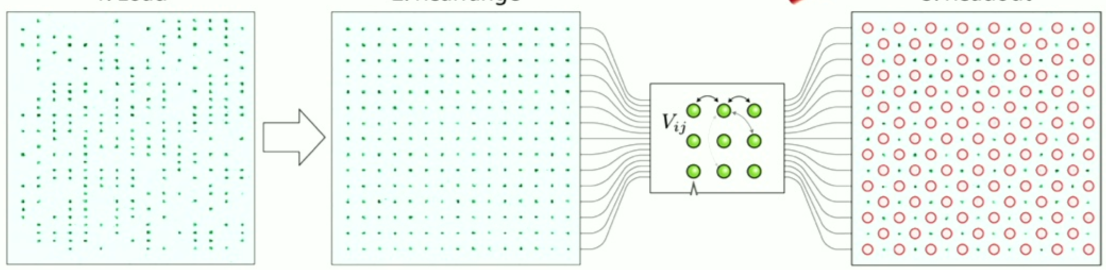


HARVARD UNIVERSITY

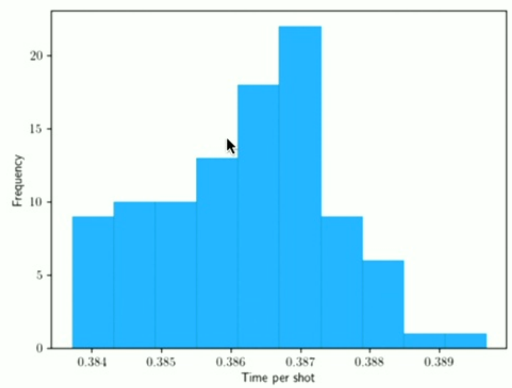
|QuEra⟩

$^{87}\mathrm{Rb}$





Jonathon Kambulow
PSI START 2023



Single-atom resolved fluorescent imaging provides projective measurements of $|g\rangle$

Measurements are destructive

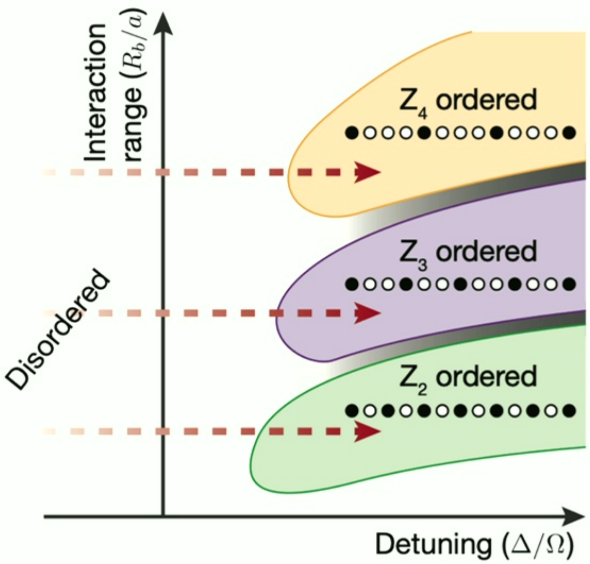
1D atom array

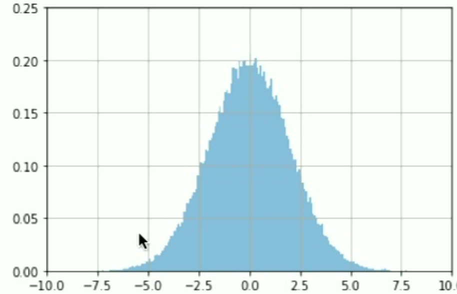
nature

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

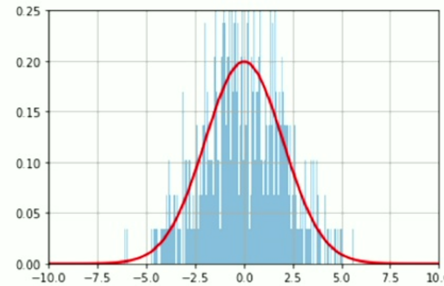
Hannes Bernien, Sylvain Schwartz, Alexander Keesling, Harry Levine, Ahmed Omran, Hannes Pichler, Soonwon Choi, Alexander S. Zibrov, Manuel Endres, Markus Greiner, Vladan Vuletić & Mikhail D. Lukin

Nature **551**, 579–584 (2017)

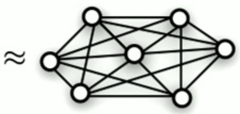




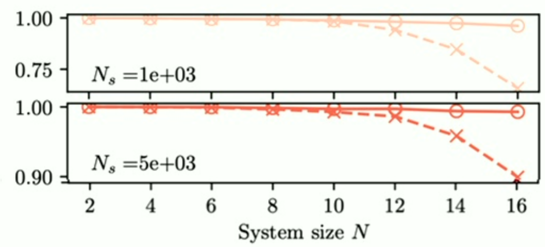
$$p(x) \approx \frac{1}{||\mathcal{D}||} \sum_{\mathbf{x}_k \in \mathcal{D}} \delta_{\mathbf{x},\mathbf{x}_k}$$



$$p_\lambda(x) \approx$$



Fidelity improvements for RBMs vs. frequency-distribution reconstructions, for a selection of dataset sizes Ns.



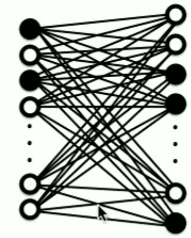
G. Torlai

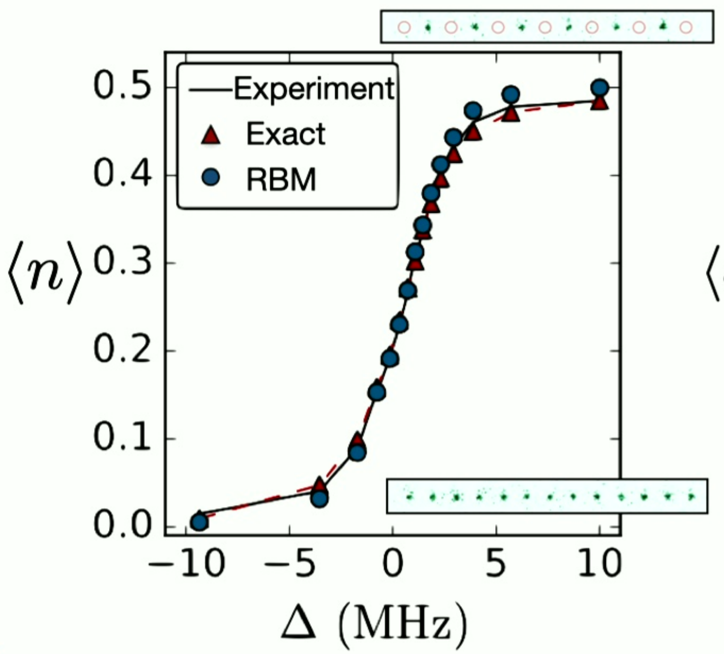
B. Timar

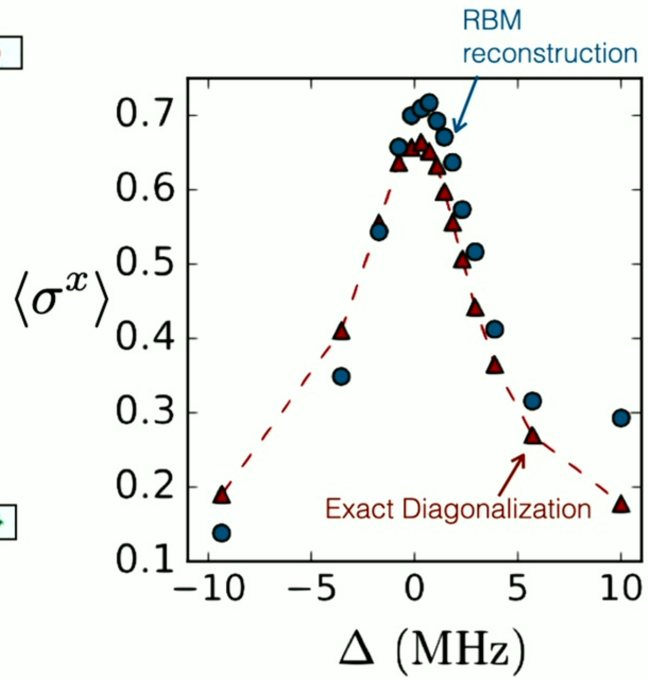
Data-driven state reconstruction

Torlai, Timar, van Nieuwenburg, Levine, Omran, Keesling, Bernien, Greiner, Vuletić, Lukin, RGM, Endres, Phys. Rev. Lett. 123, 230504 (2019)

- 8 atoms, 3000 shots
- one-dimensional AFM chain
- 3000 projective measurements per detuning parameter
- RBM trained a energy-based model, produced estimators







Stoquastic Hamiltonian:

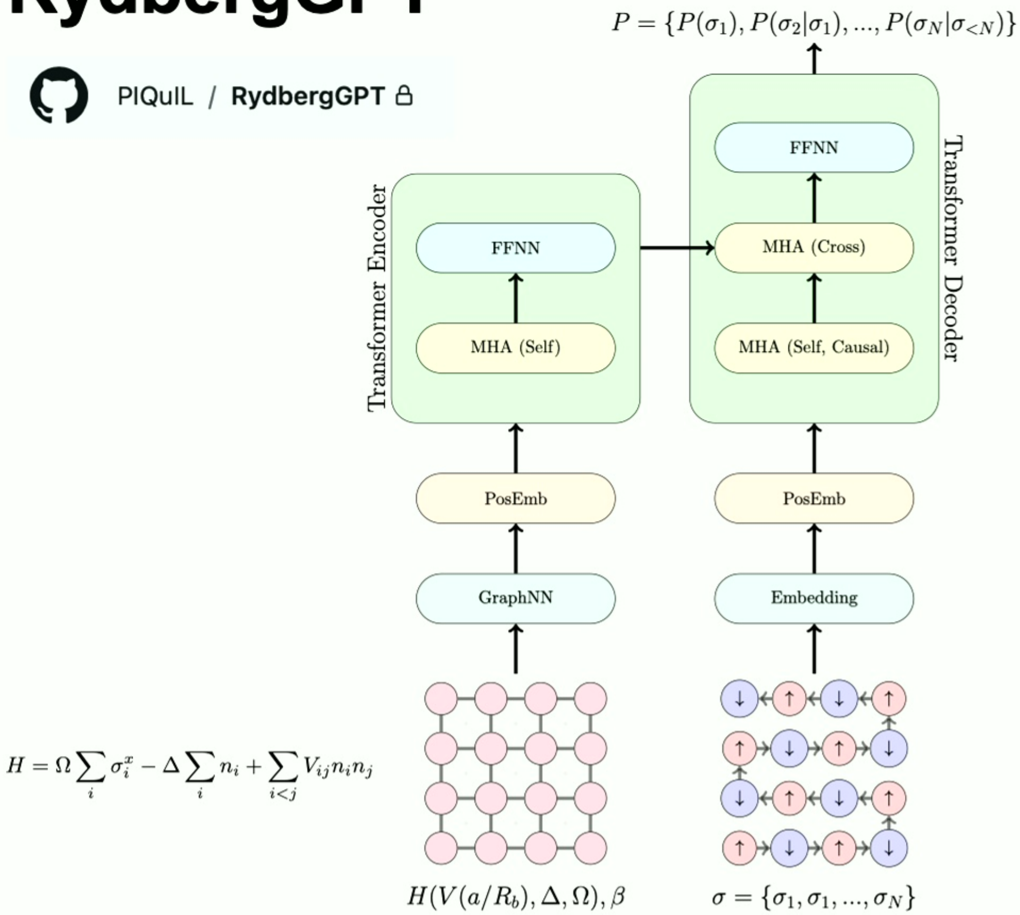
$$\psi_\lambda(z) = \sqrt{p_\lambda(z)}$$

$$\langle \mathcal{O} \rangle = \sum_{zz'} \psi_\lambda(z)\psi_\lambda(z')\mathcal{O}_{zz'}$$

$$= \sum_{z} \psi^2_\lambda(z) \underbrace{\boxed{\sum_{z'} \frac{\psi_\lambda(z')}{\psi_\lambda(z)} \mathcal{O}_{zz'}}}_{\text{“local” estimator}}$$

RydbergGPT

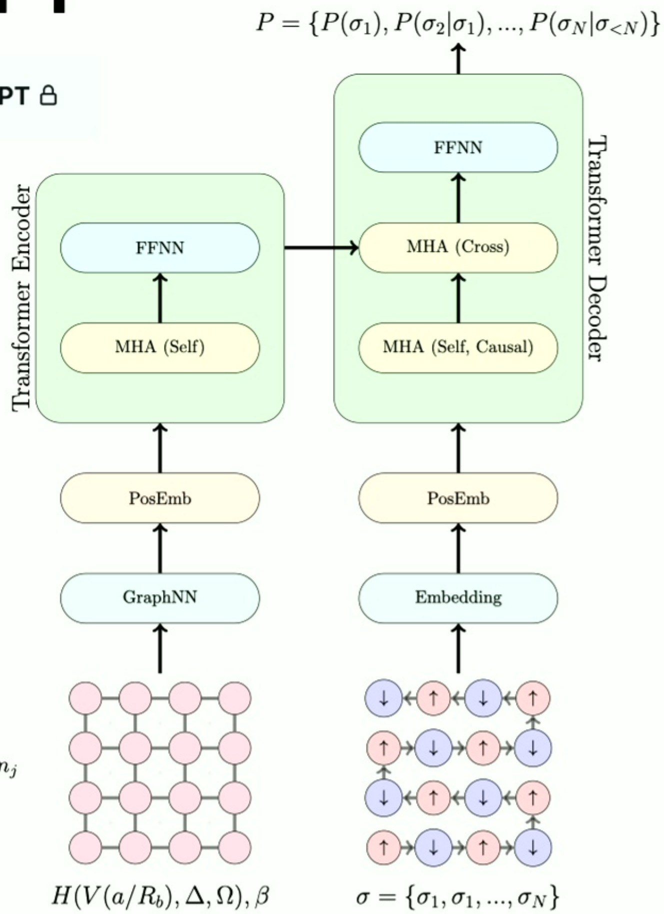
PIQuIL / **RydbergGPT**



RydbergGPT

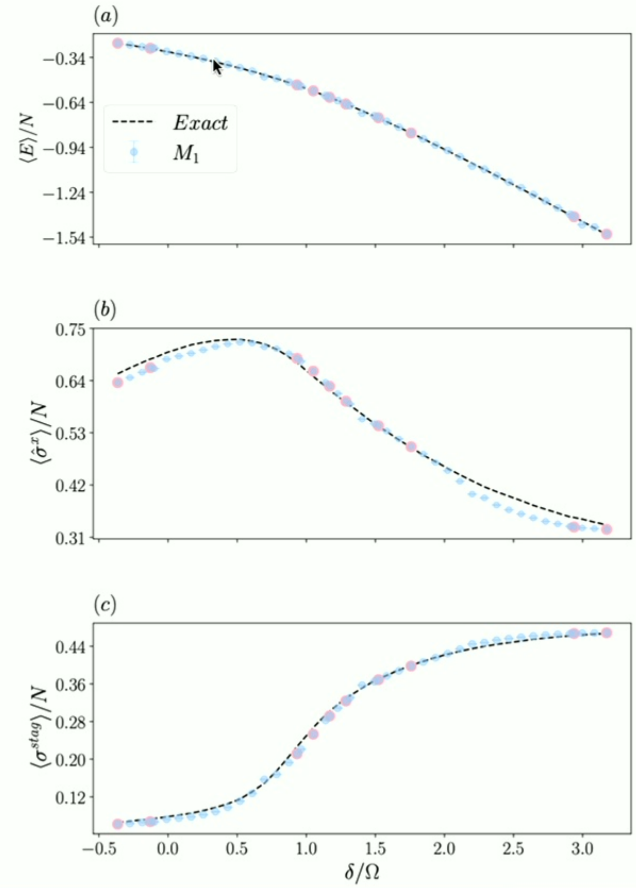
PIQuIL / **RydbergGPT**

$$H = \Omega \sum_i \sigma_i^x - \Delta \sum_i n_i + \sum_{i<j} V_{ij} n_i n_j$$



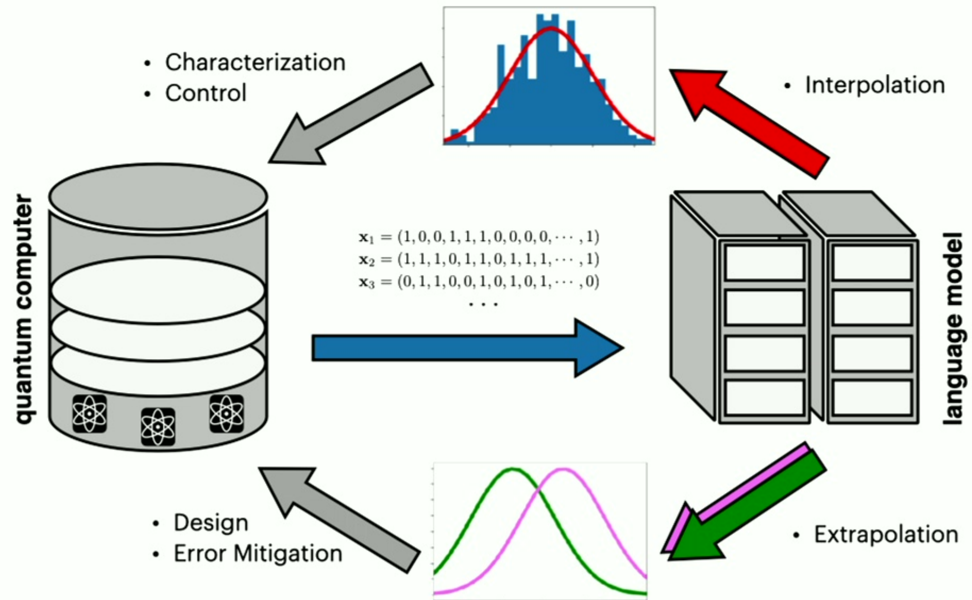
RydbergGPT

David Fitzek,[1] Yi Hong Teoh,[2] Hin Pok Fung,[2] Gebremedhin A. Dagnew,[2] Ejaaz Merali,[2] M. Schuyler Moss,[2] Benjamin MacLellan,[2] and Roger Melko[2,3]



Discussion: language models for quantum simulation

- LLMs are increasingly demonstrating *scaling* & *emergence*
- Quantum measurement data is well suited to train generative models
- Large models will be a flexible tool to aid quantum computer design & control
- It would be interesting to *scale* LLMs using quantum data… but can we afford to?



RGM & J. Carrasquilla,
Nature Computational Science 4, 11–18 (2024)