Title: [VIRTUAL] Emergent Classicality from Information Bottleneck

Speakers:

Collection: Machine Learning for Quantum Many-Body Systems

Date: June 15, 2023 - 10:00 AM

URL: https://pirsa.org/23060044

Abstract: Our universe is quantum, but our everyday experience is classical. Where is the boundary between quantum and classical worlds? How does classical reality emerge in quantum many-body systems? Does the collapse of the quantum states involve intelligence? These are fundamental questions that have puzzled physicists and philosophers for centuries. The recent development of quantum information science and artificial intelligence offers new opportunities to investigate these old problems. In this talk, we present our preliminary research on using a transformer-based language model to process randomized measurement data collected from Schrödinger's cat quantum state. We show that the classical reality emerges in the language model due to the information bottleneck: although our training data contains the full quantum information of Schrödinger's cat, a weak language model can only learn the classical reality of the cat from the data. Our study opens up a new avenue for using the big data generated on noisy intermediate-scale quantum (NISQ) devices to train generative models for representation learning of quantum operators, which might be a step toward our ultimate goal of creating an artificial intelligence quantum physicist.

Emergent Clasicallity from Information Bottleneck

Yi-Zhuang You
(UCSD, Physics)
May 2023

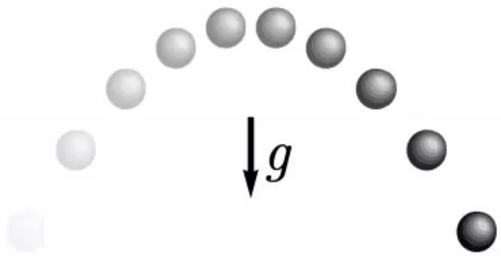
Art by Midjourney AI

What is Classicality?

- In physics, the word "classical" is used in contrast to "quantum": classical physics refers to physics before quantum mechanics.

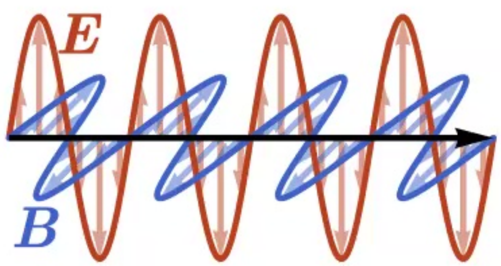
Issac Newton

$$F = ma$$



James Maxwell

$$\mathrm{d} \star F = J$$



- Classical physics is deterministic.
- It works pretty well in the macroscopic world.

How is Quantum Differed from Classical?

- In the early 20th century, it was realized that classical physics does not quite apply to the microscopic world.

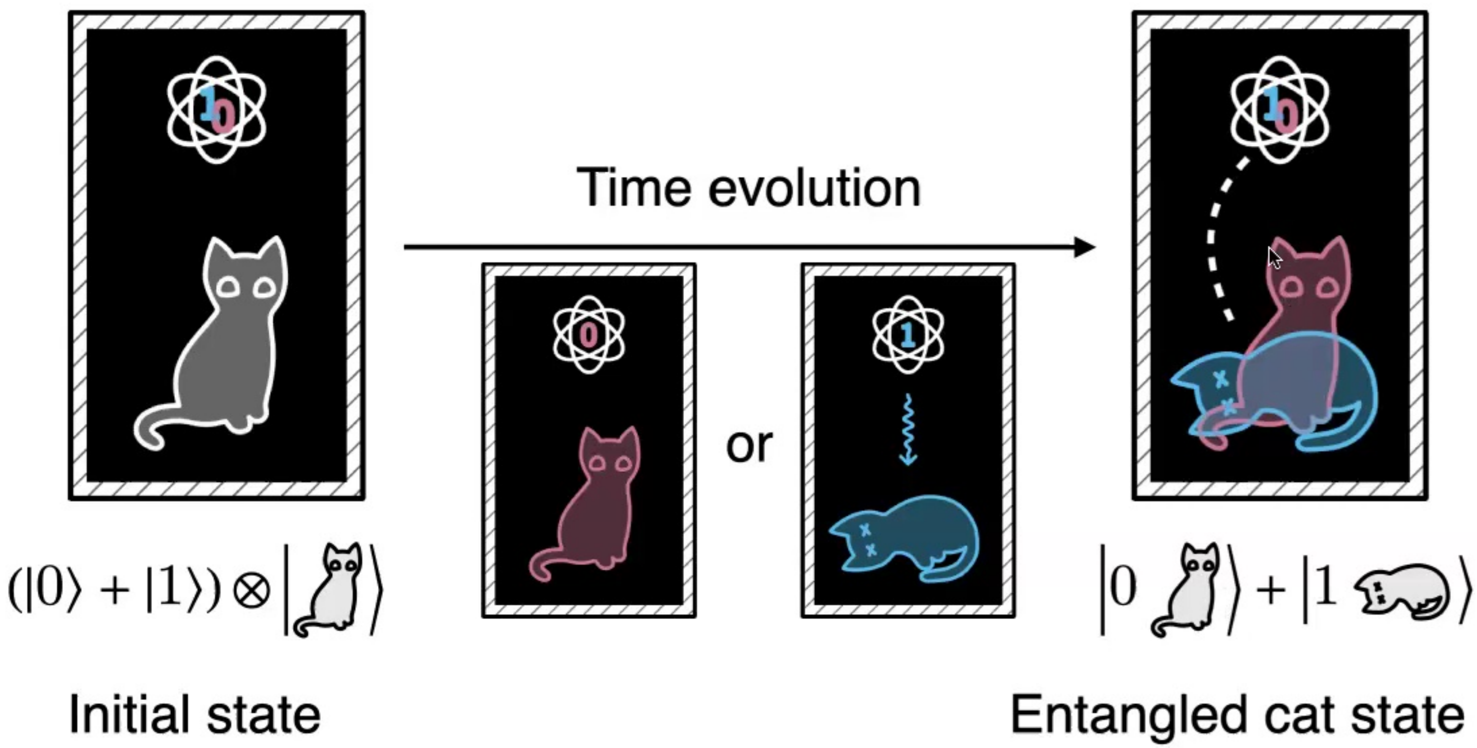
Max Planck Albert Eistein Niels Bohr E. Schrödinger W. Heisenberg

- A new branch of physics — quantum mechanics — was established. It is intrinsically probabilistic.
- Quantum mechanics is more exotic: it describes the square root of probability — called probability amplitude.

$$\psi(x) \sim \pm\sqrt{p(x)}$$

Schrödinger's Cat

- Quantum superposition can become weirder when it comes to states of multiple qubits — a famous example is Schrödinger's cat.

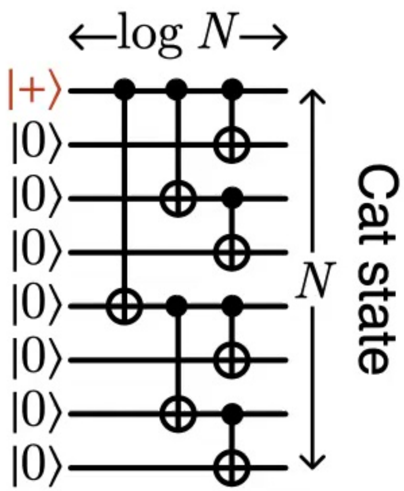


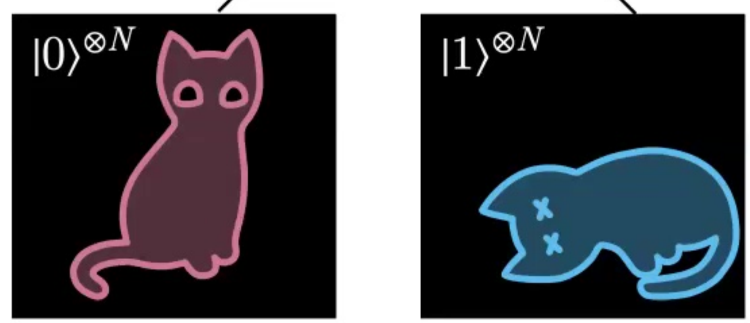
Schrödinger's Cat

- The cat state can be modeled by a multi-qubit GHZ state,

$$|+\rangle := \frac{|0\rangle + |1\rangle}{\sqrt{2}} \xrightarrow[\text{(Quantum circuit)}]{\text{Time evolution}} \frac{|0\rangle^{\otimes N} + |1\rangle^{\otimes N}}{\sqrt{2}}$$

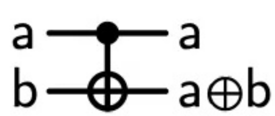
which can be prepared by a quantum circuit in log N depth (time).





Greenberger, Horne, Zeilinger 1989

CNOT (controlled-NOT) gate

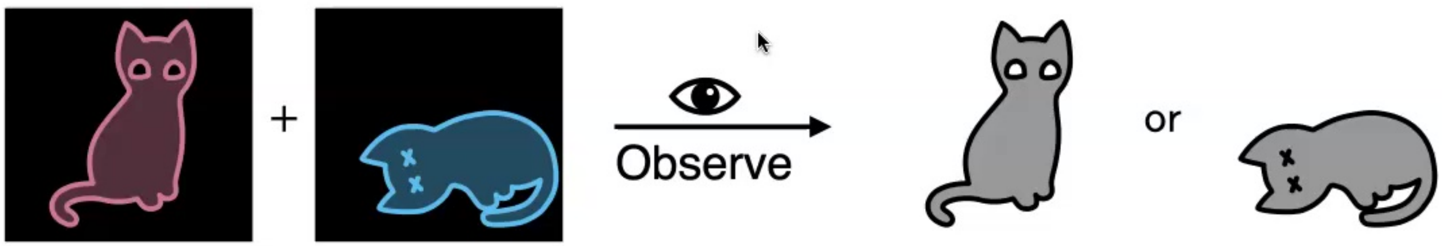


$a, b \in \{0, 1\}$

$|a\rangle \otimes |b\rangle \to |a\rangle \otimes |a \oplus b\rangle$

Quantum State Collapse

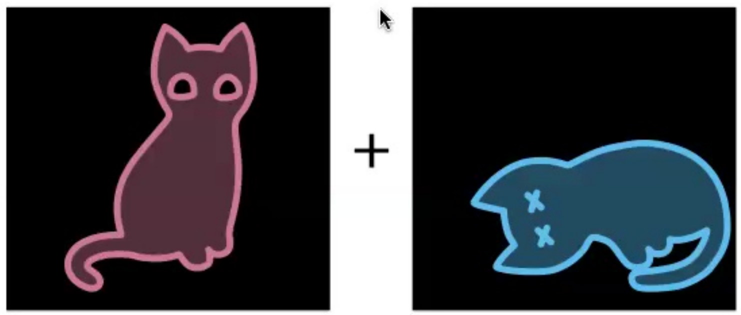
- But we never see a superposition cat in reality. Why?
- **Copenhagen Interpretation**: Observing the cat would cause the superposition to collapse into one of the two classical realities: cat alive or cat dead.



- What happens during the observation?
- Who qualifies as an observer?
- Should the observer be conscious/intelligent? …

Quantum State Collapse

- **Modern understanding**: randomized measurement + classical data processing.
  - Measurement: the system interacts with the environment.

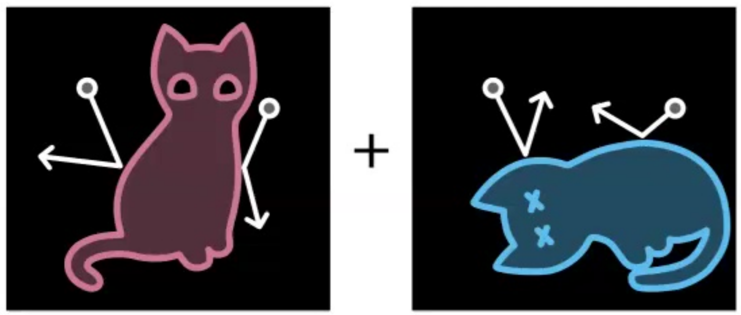


  - Interaction → entanglement (information sharing).
  - Information loss = entropy increase:
    pure cat state → mixed state ensemble of alive and dead.
  - This process is called quantum decoherence. No intelligence is required at this step.

Joos, Zeh 1985; Ghirard, Rimini, Weber 1986, Zurek 2003.

Quantum State Collapse

- **Modern understanding**: randomized measurement + classical data processing.
  - Measurement: the system interacts with the environment.

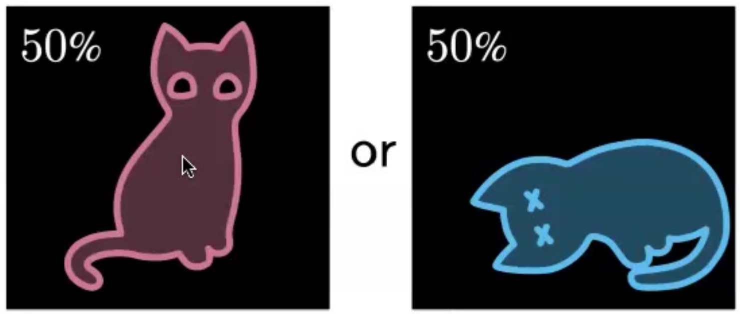


  - Interaction → entanglement (information sharing).
  - Information loss = entropy increase:
    pure cat state → mixed state ensemble of alive and dead.
  - This process is called quantum decoherence. No intelligence is required at this step.

Joos, Zeh 1985; Ghirard, Rimini, Weber 1986, Zurek 2003.

Quantum State Collapse

- **Modern understanding**: randomized measurement + classical data processing.
  - Measurement: the system interacts with the environment.

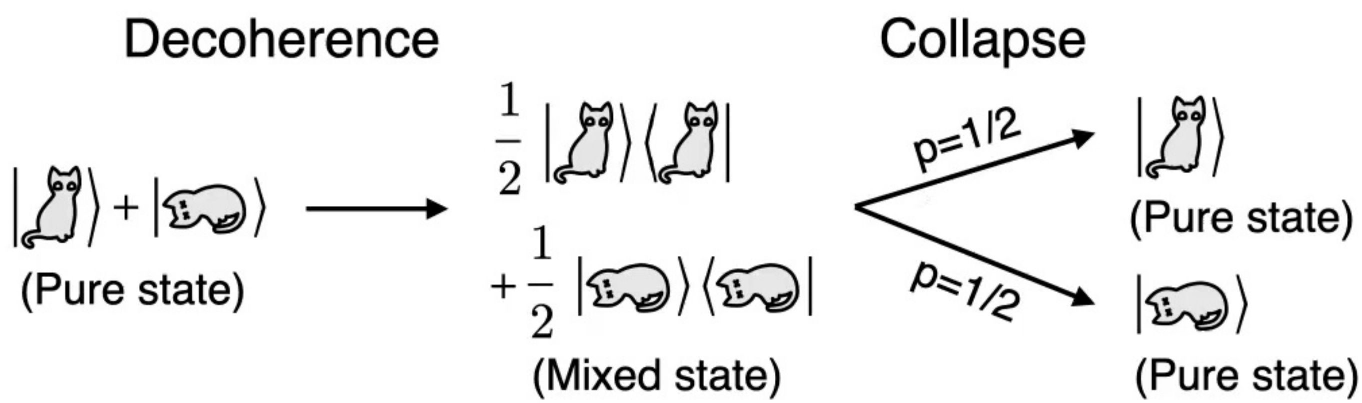


  - Interaction → entanglement (information sharing).
  - Information loss = entropy increase: pure cat state → mixed state ensemble of alive and dead.
  - This process is called quantum decoherence. No intelligence is required at this step.

Joos, Zeh 1985; Ghirard, Rimini, Weber 1986, Zurek 2003.

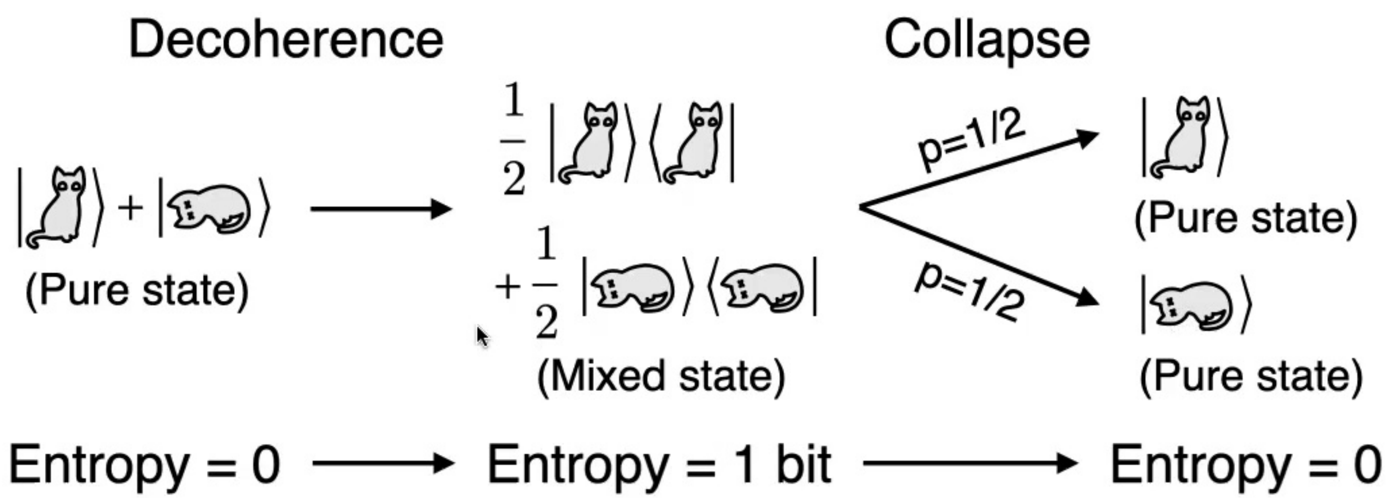
Quantum State Collapse

- **Modern understanding**: randomized measurement + classical data processing.
  - Emergent classical reality: how to collapse from the mixed state back to one of the alive/dead pure states



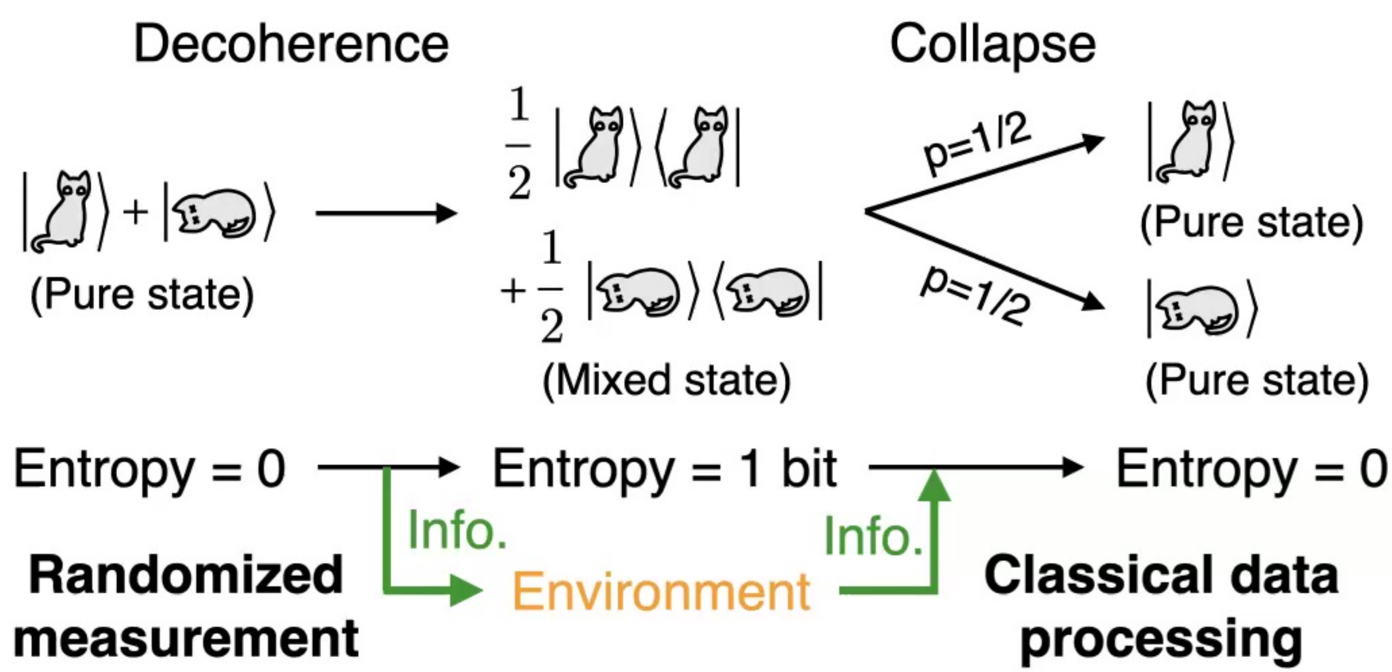
Quantum State Collapse

- **Modern understanding**: randomized measurement + classical data processing.
  - Emergent classical reality: how to collapse from the mixed state back to one of the alive/dead pure states



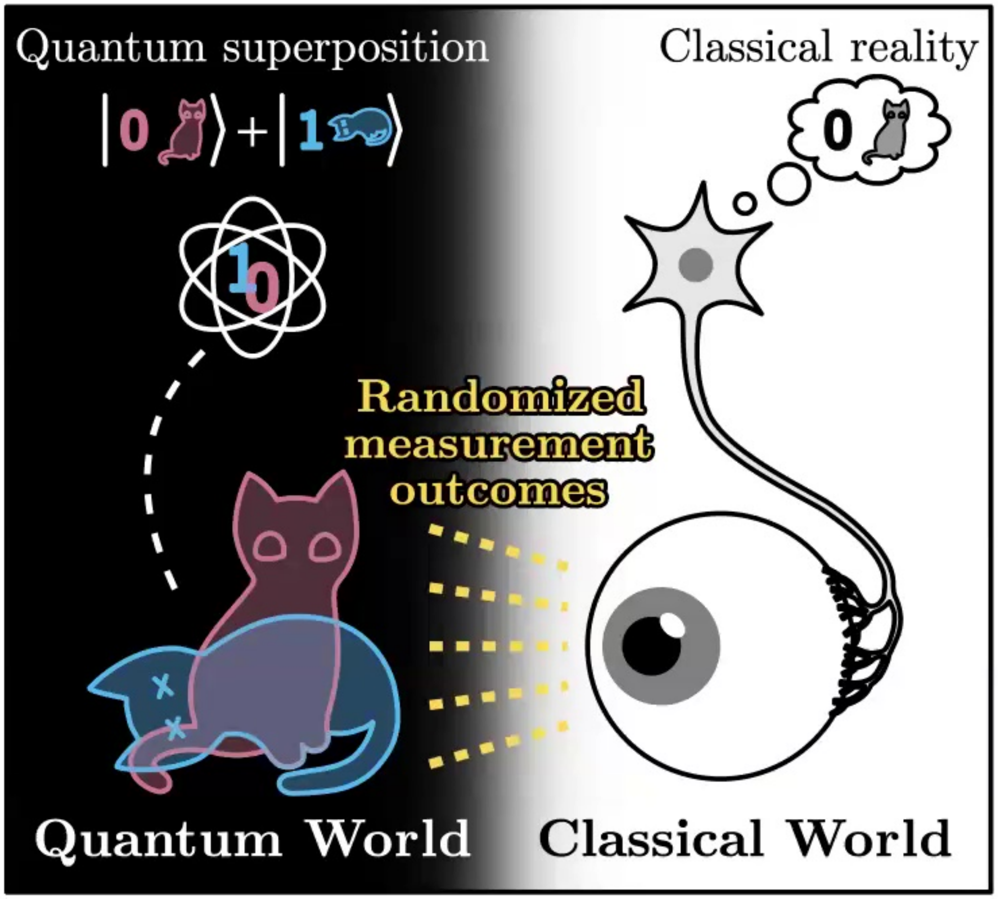
Quantum State Collapse

- **Modern understanding**: randomized measurement + classical data processing.
  - Emergent classical reality: how to collapse from the mixed state back to one of the alive/dead pure states



General Idea

- **Idea**: use AI to process randomized measurement data.



Randomize Measurement

- Randomized measurement — estimate properties of an unknown quantum state by measuring random observables.
  - Philosophy: measure first, ask questions later.
- **Measurement scheme**:
  - Prepare an N-qubit GHZ state $|\Psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes N} + |1\rangle^{\otimes N})$
  - Perform random & local measurements:
    - Draw a sequence of Pauli observables uniformly
      $$\boldsymbol{x} = (x_1, x_2, \cdots, x_N), \quad x_i \in \{X, Y, Z\}$$
    - Independently measure each qubit i by its corresponding observable x_i
  - Collect measurement outcomes as a sequence
    $$\boldsymbol{y} = (y_1, y_2, \cdots, y_N), \quad y_i \in \{\pm 1\}$$
  - Repeat …

Elben et al. 2022

Randomize Measurement

- Randomized measurements collect a large amount of data.
- **Data structure**: a pair of sequences

$$(\boldsymbol{x}, \boldsymbol{y}) \quad \boldsymbol{x} \in \{X, Y, Z\}^{\times N}, \boldsymbol{y} \in \{\pm 1\}^{\times N}$$

- **Data distribution**: $p(\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})$

$$p(\boldsymbol{x}) = 3^{-N} \quad \text{(Uniform, trivial)}$$

$$p(\boldsymbol{y}|\boldsymbol{x}) = \langle\Psi| \bigotimes_i \frac{1 + y_i x_i}{2} |\Psi\rangle$$

- **Classical post-processing**: $(\boldsymbol{x}, \boldsymbol{y})$ are also called classical shadows, from which the quantum state can be recovered.

$$\rho := |\Psi\rangle\langle\Psi| = \mathop{\mathbb{E}}_{(\boldsymbol{x},\boldsymbol{y})} \bigotimes_i \frac{1 + 3y_i x_i}{2}$$

Huang, Kueng, Preskill 2020

Randomize Measurement

- Randomized measurements collect a large amount of data.

$$(\boldsymbol{x}, \boldsymbol{y}) \quad \boldsymbol{x} \in \{X, Y, Z\}^{\times N}, \boldsymbol{y} \in \{\pm 1\}^{\times N}$$

$$(\boldsymbol{x}, \boldsymbol{y}) \sim p(\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})$$

- Examples ($N = 4$): classical shadows of Schrödinger's cat

x: XYYX	x: XYYZ	x: XZYX	x: XZXX	x: XYXY	x: XYZZ
y: −+−−	y: −+−−	y: −−+−	y: −+−+	y: +−++	y: +−++
x: XZZZ	x: XZXY	x: YZZZ	x: YZXX	x: YXXY	x: YXZX
y: +−−−	y: ++−−	y: −−−−	y: −++−	y: ++−+	y: ++++
x: YYXZ	x: YYYX	x: ZYXX	x: ZZYX	x: ZZZY	x: ZZZX
y: ++−+	y: +++−	y: −+−−	y: −−−+	y: −−−+	y: ++++

Generative Modeling of Classical Shadows

- **Objective**: to model the conditional distribution $p(\boldsymbol{y}|\boldsymbol{x})$ of measurement outcomes given local observables.

  `x: ZZXY` Observables (question)

  `y: ---+` Outcomes (answer)

  - This maps to a chat completion task in natural language processing. — We can train a transformer-based generative language model to perform the task.

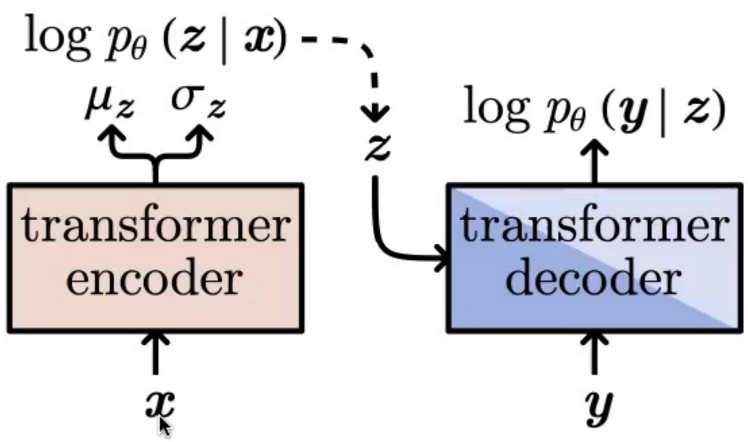
    Vaswani et al. 2017; Devlin et al. 2019

  - After training, the model can replace the quantum experiment to answer questions about the underlying quantum state (the cat state). — It can "speak" the quantum language.

Generative Modeling of Classical Shadows

- **Objective**: to model the conditional distribution $p(\boldsymbol{y}|\boldsymbol{x})$ of measurement outcomes given local observables.
- **Architecture**: transformer-based β-VAE

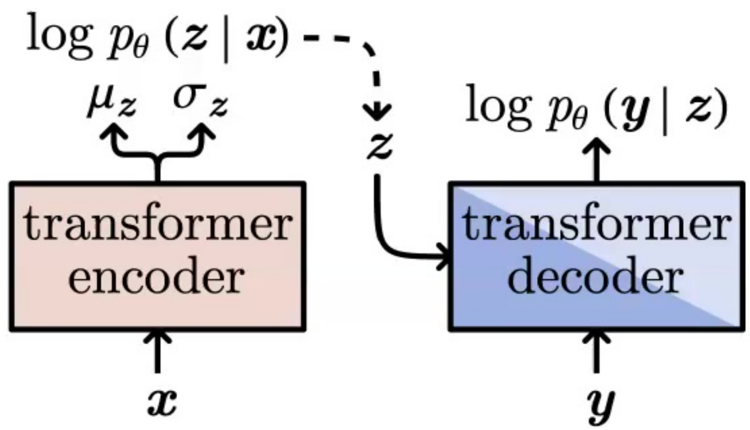
$$p_\theta(\boldsymbol{y}|\boldsymbol{x}) = \int_{\boldsymbol{z}} p_\theta(\boldsymbol{y}|\boldsymbol{z}) p_\theta(\boldsymbol{z}|\boldsymbol{x})$$



Kingma, Welling 2014; Henderson, Fehr 2022

Generative Modeling of Classical Shadows

- **Loss function** (ELBO): $\mathcal{L} = \underset{(\boldsymbol{x},\boldsymbol{y})\sim p_{\mathrm{dat}}}{\mathbb{E}} \mathcal{L}(\boldsymbol{x},\boldsymbol{y})$



$$\mathcal{L}(\boldsymbol{x},\boldsymbol{y}) = -\underset{\boldsymbol{z}\sim p_\theta(\boldsymbol{z}|\boldsymbol{x})}{\mathbb{E}} \log p_\theta(\boldsymbol{y}|\boldsymbol{z})$$ Negative log-likelihood

$$+ \beta \mathrm{KL}[p_\theta(\boldsymbol{z}|\boldsymbol{x}) \| p_{\mathcal{N}}(\boldsymbol{z})]$$ KL regularization

- Hyperparameter β enables us to impose a variational information bottleneck on the transformer.

Kingma, Welling 2014; Henderson, Fehr 2022

Model Evaluation

- **Evaluation metric**: fidelity — a measure of the closeness between quantum states.
  - Original state ($|\Psi\rangle$ - the GHZ state):

$$|\Psi\rangle\langle\Psi| = \rho = \mathop{\mathbb{E}}_{(\boldsymbol{x},\boldsymbol{y})\sim p_{\mathrm{dat}}} \bigotimes_i \frac{1+3y_i x_i}{2}$$

  - Reconstructed state:

$$\tilde{\rho} = \mathop{\mathbb{E}}_{(\boldsymbol{x},\tilde{\boldsymbol{y}})\sim p_{\mathrm{mdl}}} \bigotimes_i \frac{1+3\tilde{y}_i x_i}{2}$$

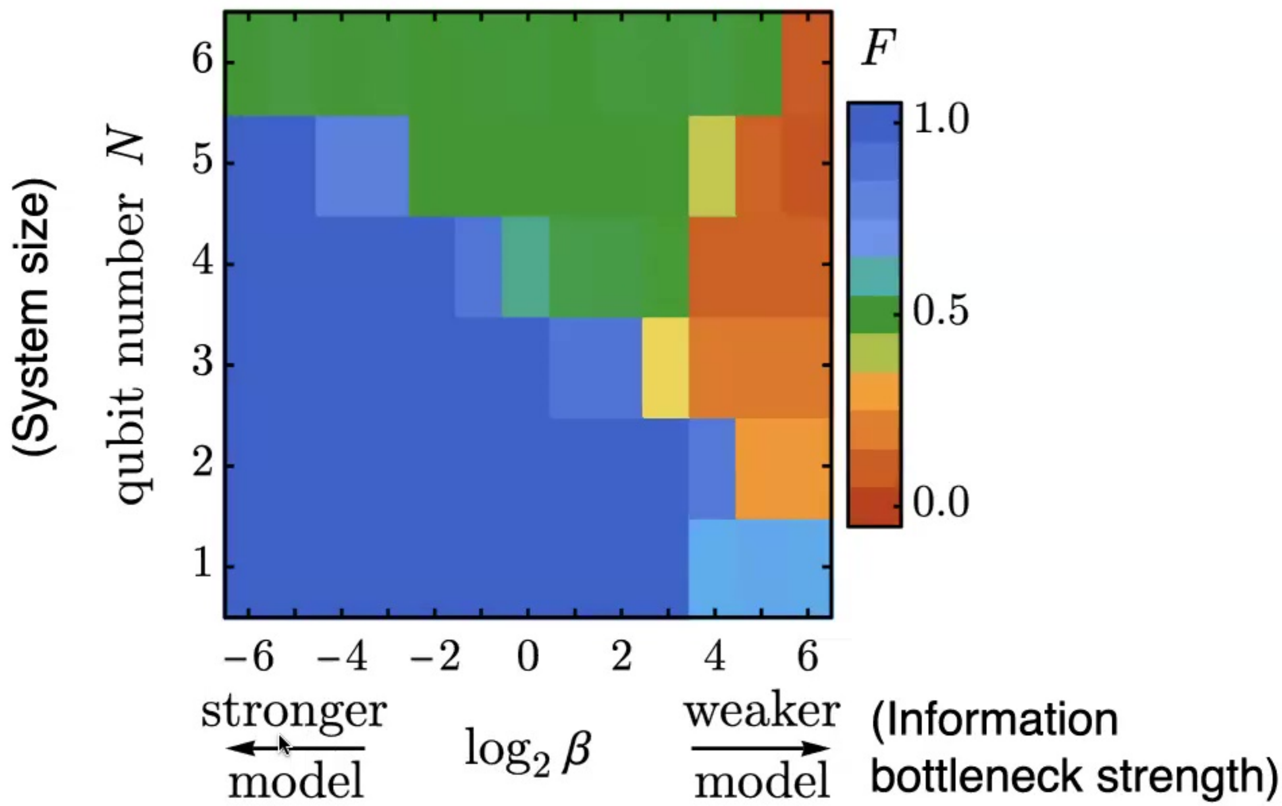
  - Fidelity (the probability of observing $\tilde{\rho}$ given $|\Psi\rangle$)

$$F(\rho,\tilde{\rho}) := \left(\mathrm{Tr}\sqrt{\sqrt{\rho}\tilde{\rho}\sqrt{\rho}}\right)^2 = \langle\Psi|\tilde{\rho}|\Psi\rangle$$

  In general, $0 \leq F(\rho,\tilde{\rho}) \leq 1$ (the larger the better).

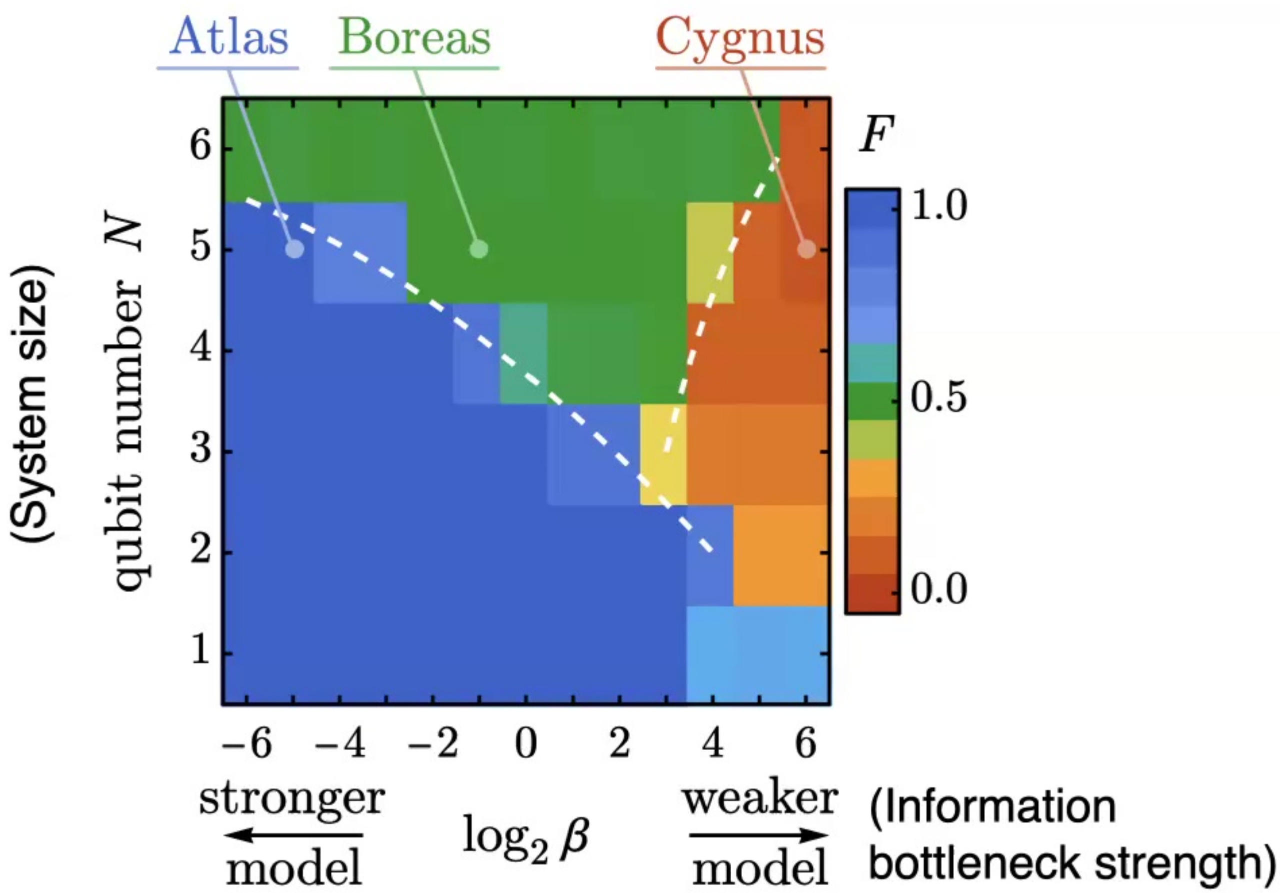
Model Evaluation

- Fidelity of the model reconstructed quantum state



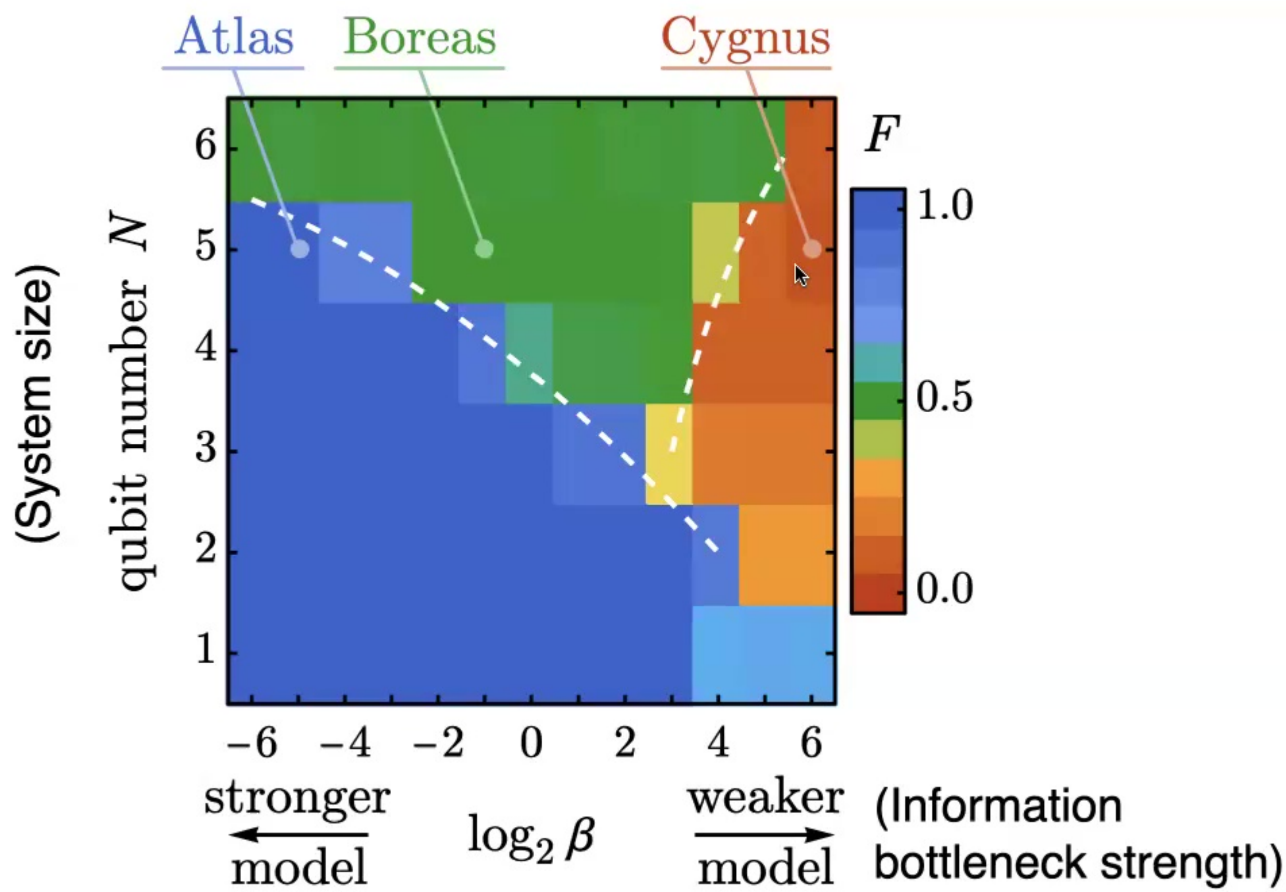
Model Evaluation

- Fidelity of the model reconstructed quantum state



Model Evaluation

- Fidelity of the model reconstructed quantum state



One-Shot Cat Classification

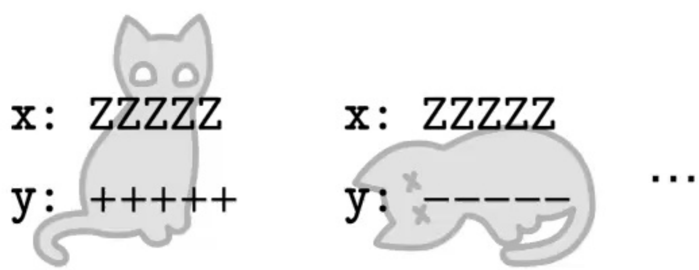
- To understand the difference between Atlas, Boreas and Cygnus, let us chat with them!
  - We can ask them for the "one-shot cat classification".

Task: given a one-shot observation of the cat, determine if it is alive or dead.

↓

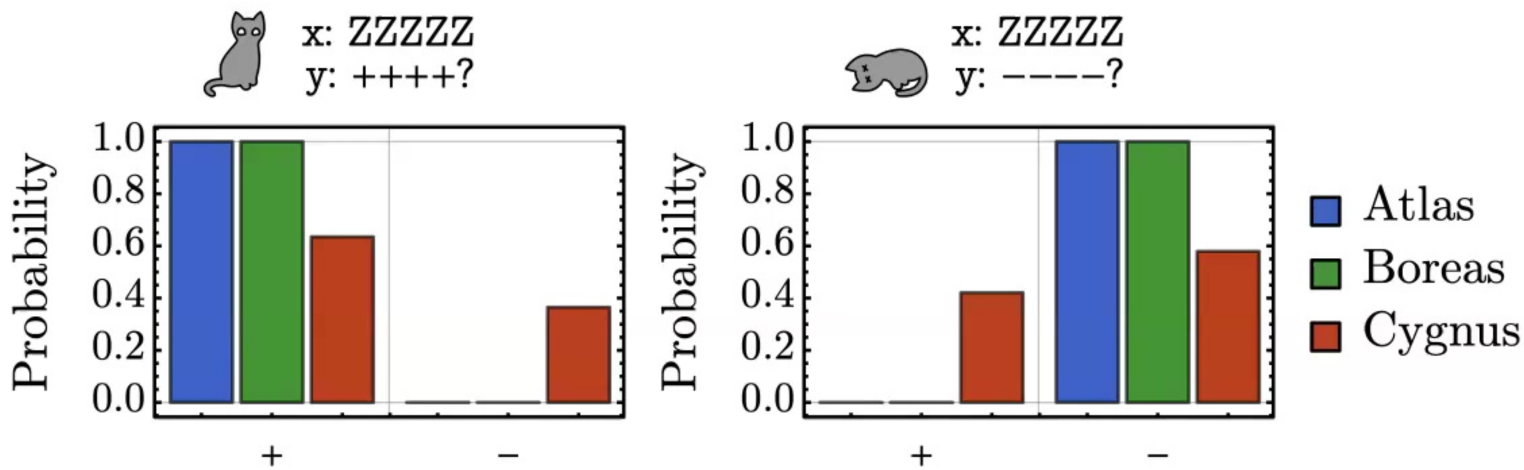
Prompt: x: ZZZZZ y: ++++? x: ZZZZZ y: ----? ...

Expectation:



One-Shot Cat Classification

- In-distribution classification task



- Atlas and Boreas can perfectly determine the life and death of the cat.
- However, Cygnus is a weaker model and cannot make a clear judgment about the classical reality.

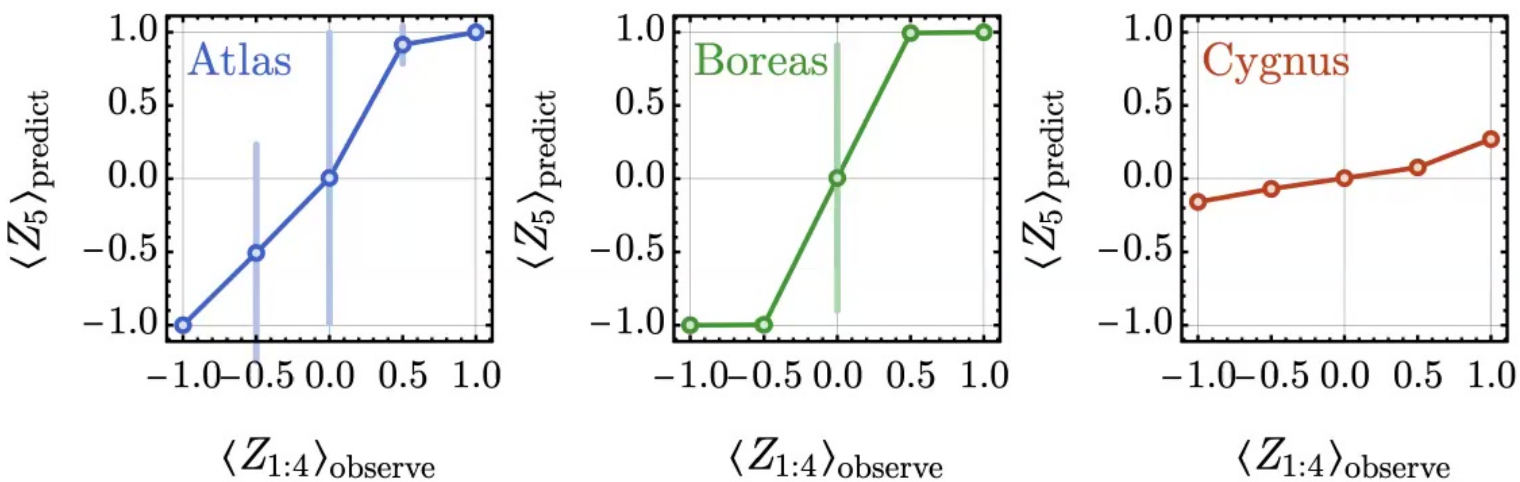
One-Shot Cat Classification

- Out-of-distribution classification task
  - What about the following prompt?

x: ZZZZZ

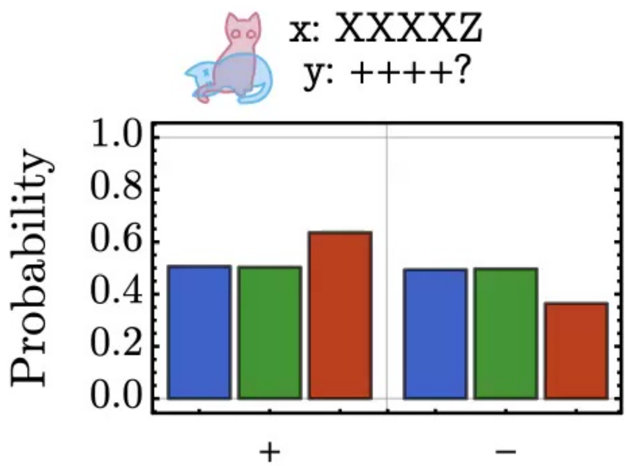
y: $\underbrace{-++-}_{Z_{1:4}}?$ ← Z_5

(This never appears in the classical shadow data of the GHZ state.)



One-Shot Cat Classification

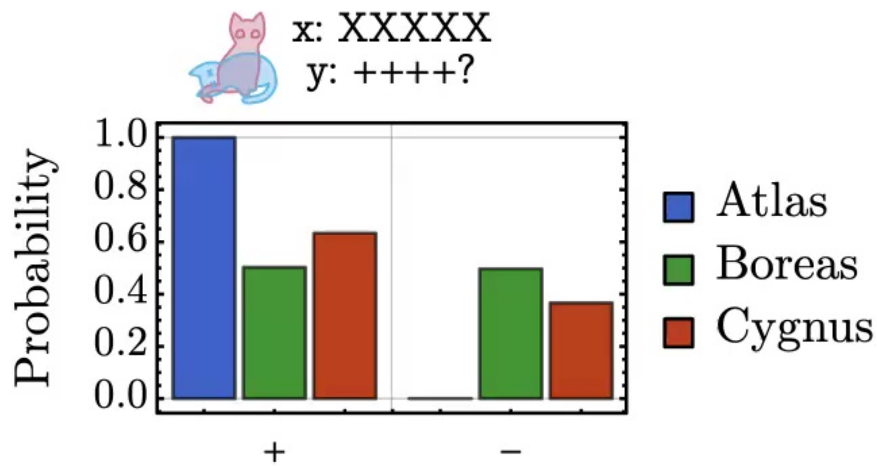
- Local Z-measurements destroy the quantum coherence of the cat state. Can we preserve the coherence?
- Consider local X-measurements:



Q: Is the Schrödinger cat alive or dead?

(+) Alive.

(−) Dead.

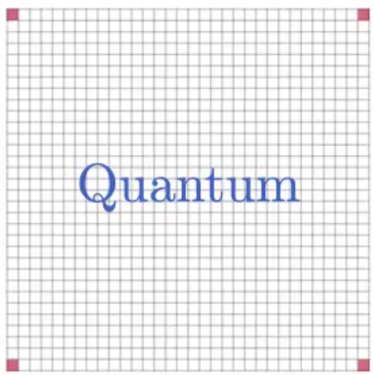
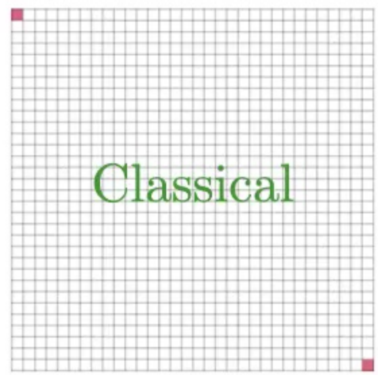
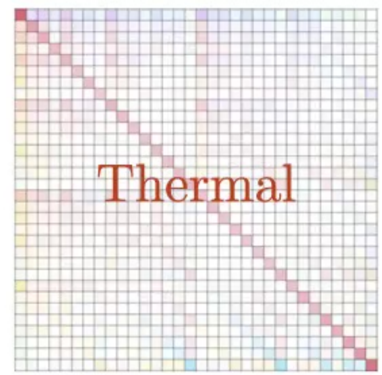


Q: What is the sign of quantum coherence?

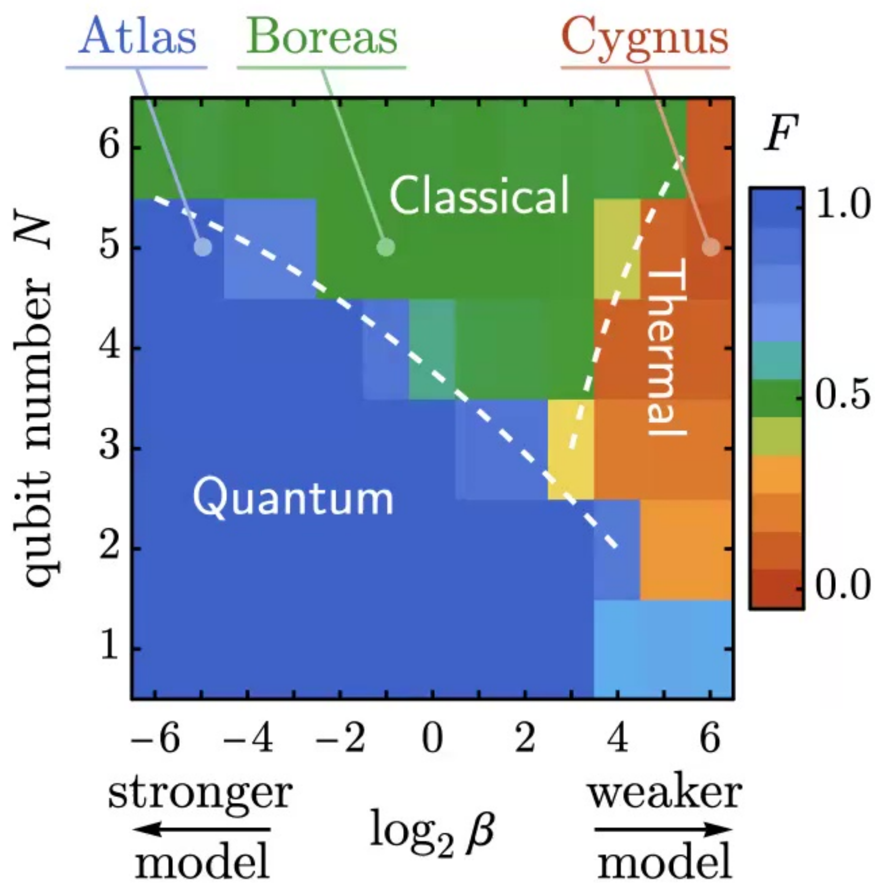
(+) Positive.

(−) Negative.

Characterize Representative Models

Model	Atlas	Boreas	Cygnus
$Z_{1:4} \to Z_5$ accuracy ($\uparrow$)	1.000	1.000	0.607
$X_{1:4} \to X_5$ accuracy ($\uparrow$)	1.000	0.503	0.634
$\tilde{\rho}$			
$F(\rho, \tilde{\rho})$ ($\uparrow$)	1.000	0.500	0.063
$S(\tilde{\rho})$ [bit] ($\downarrow$)	0.206	1.190	4.410

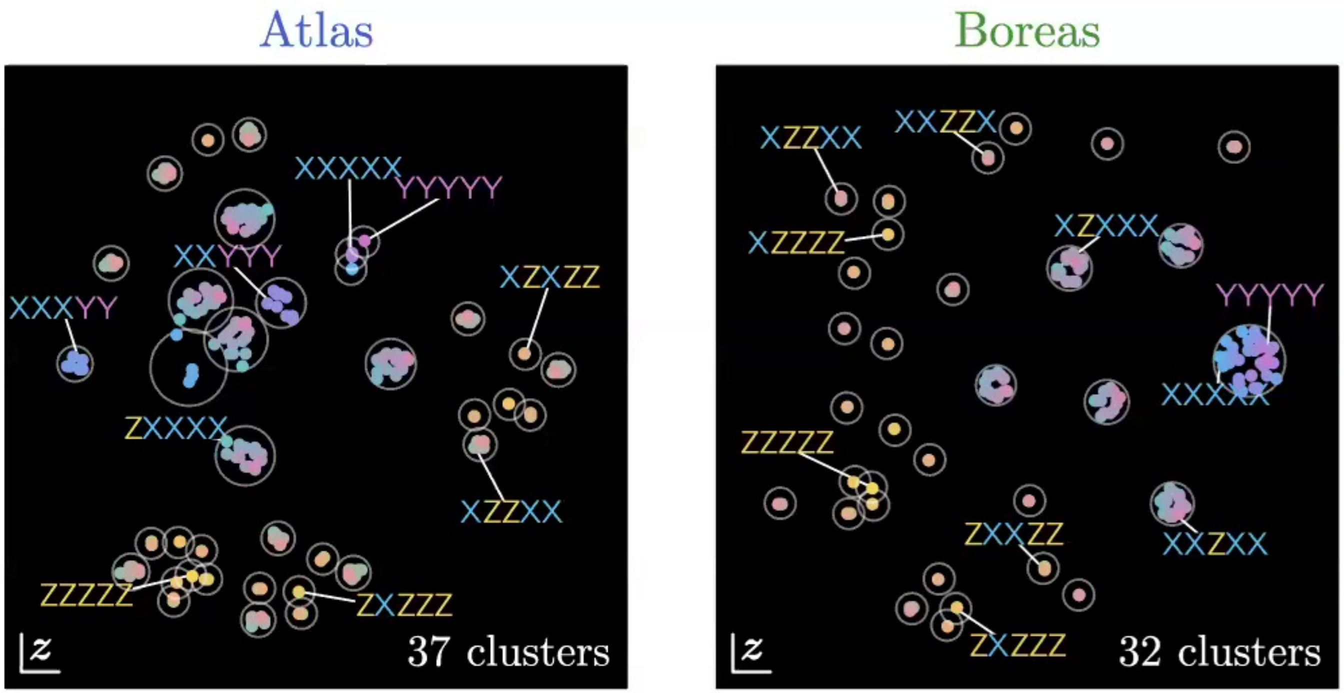
Emergent Classicality



- Classicality emerges with increasing —
  - Qubit number (system size),
  - Information bottleneck strength.
- Our world appears classical because —
  - It involves too many qubits.
  - We do not have enough classical data processing capability.

What Does the Latent Space Look Like?

- t-SNE visualization of operator embeddings.



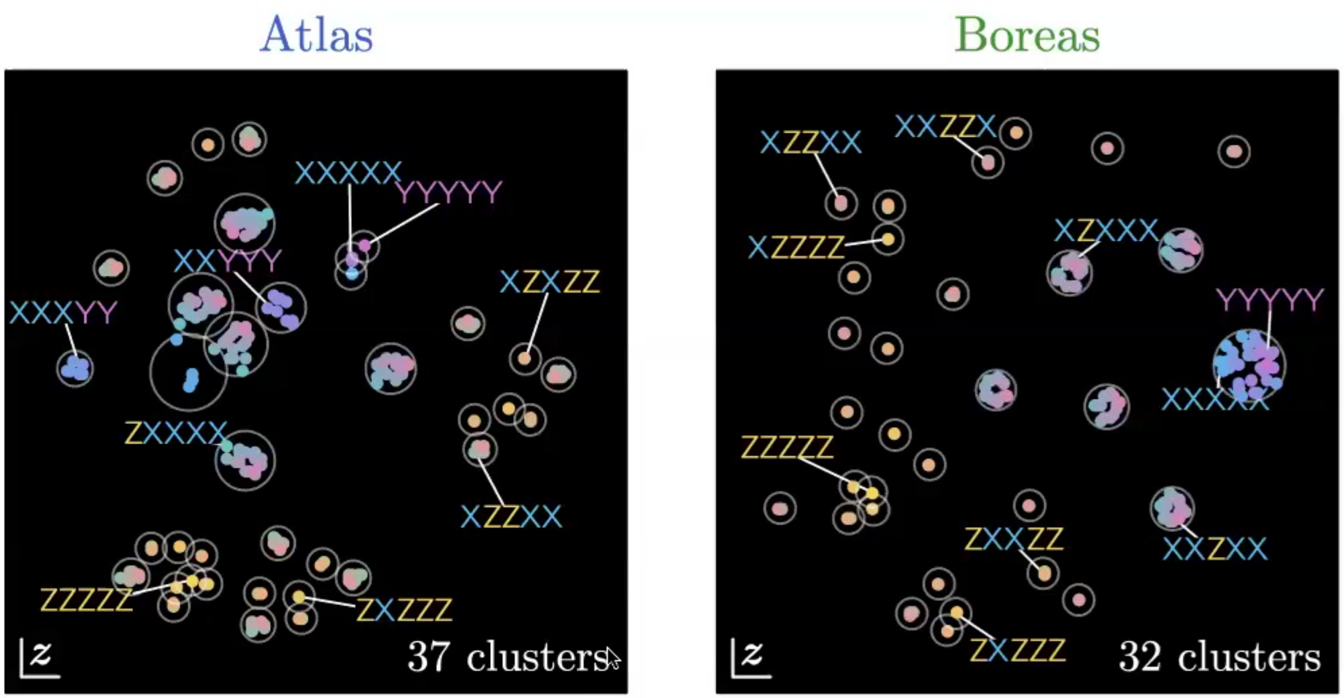
- Each dot represents a sequence of observables.

$$\boldsymbol{x} \in \{X, Y, Z\}^N \xrightarrow[\text{Encoder}]{\text{Transformer}} \boldsymbol{z}$$

van der Maaten, Hinton 2008

What Does the Latent Space Look Like?

- t-SNE visualization of operator embeddings.



- Each dot represents a sequence of observables.

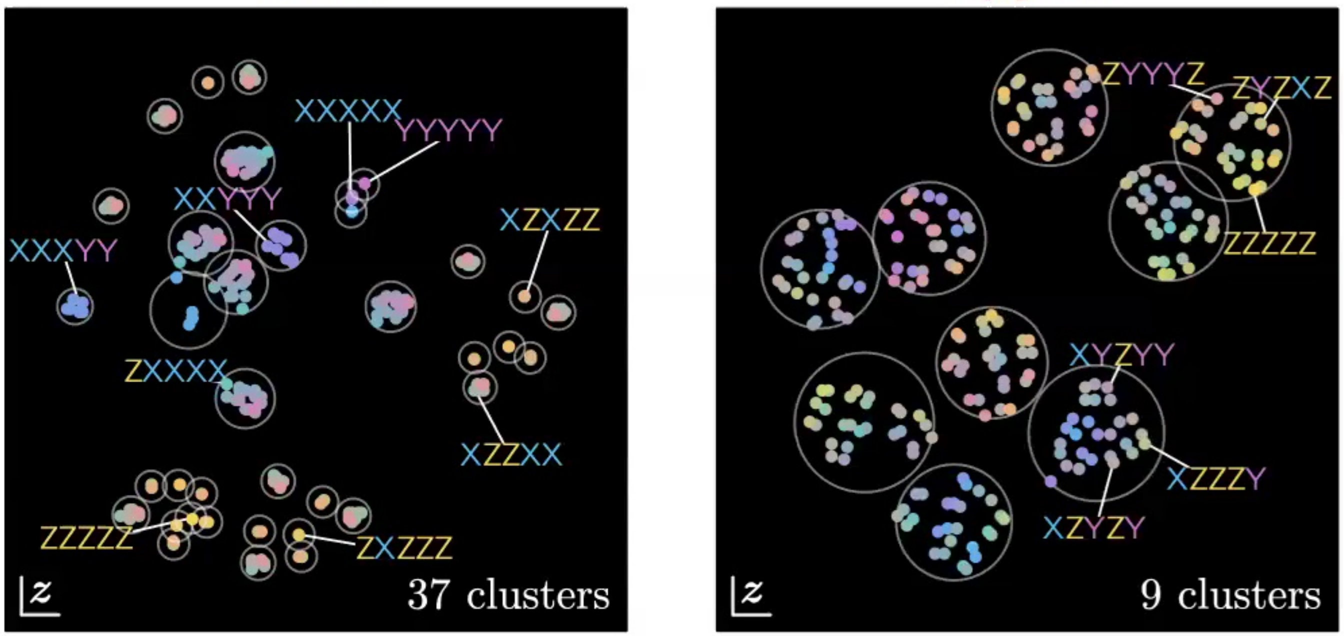
$$\boldsymbol{x} \in \{X, Y, Z\}^N \xrightarrow[\text{Encoder}]{\text{Transformer}} \boldsymbol{z}$$

van der Maaten, Hinton 2008

What Does the Latent Space Look Like?

- t-SNE visualization of operator embeddings.

Atlas / Cygnus



- Each dot represents a sequence of observables.

$$\boldsymbol{x} \in \{X, Y, Z\}^N \xrightarrow[\text{Encoder}]{\text{Transformer}} \boldsymbol{z}$$

van der Maaten, Hinton 2008

Summary

- We use a transformer-based language model to process randomized measurement data collected from Schrödinger's cat quantum state.
  - Classical reality emerges in the language model due to the information bottleneck.
  - Implying a fundamental limitation on our ability to understand the full quantum nature of the universe.
  - A new avenue for using unlabeled classical shadow data to train generative models for representation learning of quantum operators
    — a step toward realizing AI quantum physicists.