Universiteit Leiden

Felix Frohnert & Evert van Nieuwenburg

Explainable Representation Learning of Quantum States

Utilizing Machine Learning to Understand Small Quantum Systems

22/11/2023

Audience Question

You are presented with a large set of random 2-qubit states and are tasked with selecting a new numerical representation of a singular variable. You are aiming to achieve optimal fidelity in reconstructing each state from your new representation. Which variable would you choose and what would be its physical meaning?

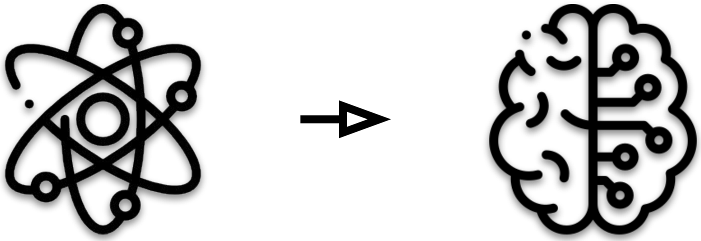
Computer-Assisted Scientific Understanding

A broad overview and motivation for this project

Computer-Assisted Scientific Understanding

Quantum Science is Difficult

… but sometimes machine learning can help: Tuning - Design - Representation



Computer-Assisted Scientific Understanding

Quantum Science is Difficult

... but sometimes machine learning can help: Tuning - Design - Representation

Interpretable Machine Learning

We want to gain scientific understanding from the machine-learned solutions!

Computer-Assisted Scientific Understanding

Quantum Science is Difficult

... but sometimes machine learning can help: Tuning - Design - Representation

Interpretable Machine Learning

We want to gain scientific understanding from the machine-learned solutions!

Three Dimensions of Computer-Assisted Scientific Understanding

1. Computational Microscope
2. Resource of Inspiration
3. Agent of Understanding

Krenn et al. Nat. Rev. Phys. 4 (2022)

Context: Machine Learning as a Computational Microscope

Neural Quantum States

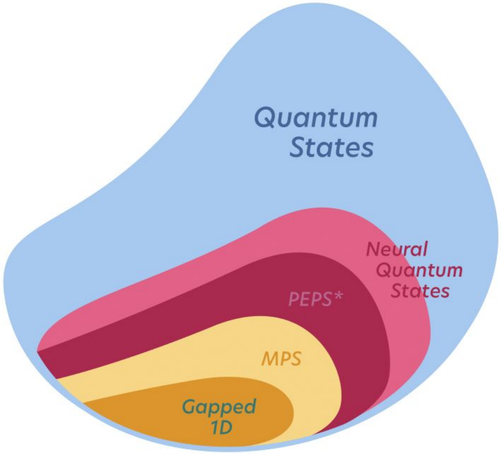
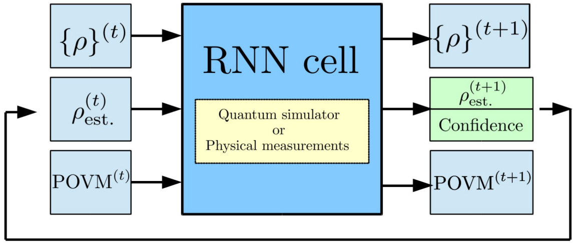


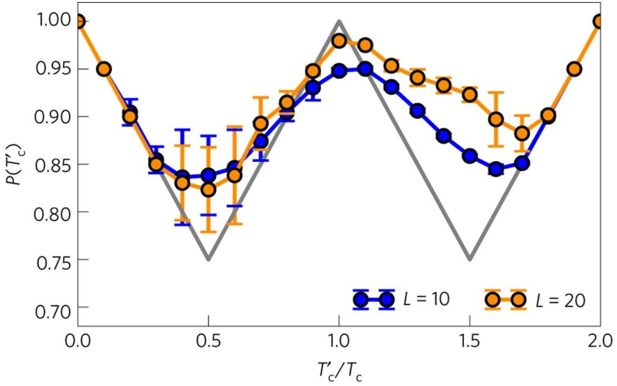
Image from Dawid et al. (2022)

Quantum Tomography



Quek et al. npj Quantum Inf. 7 (2021)

Phase Classification



van Nieuwenburg et al. Nat. Phys. 13 (2017)

Context: Machine Learning as a Computational Microscope

Neural Quantum States

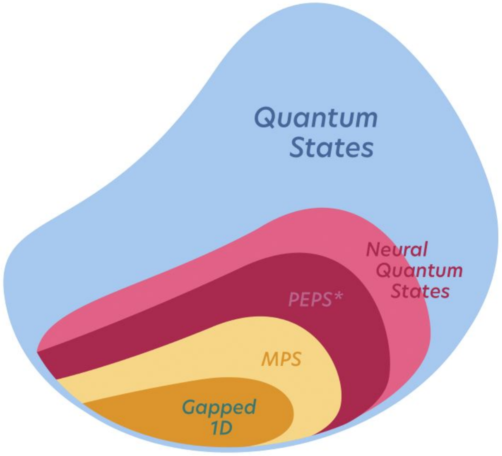
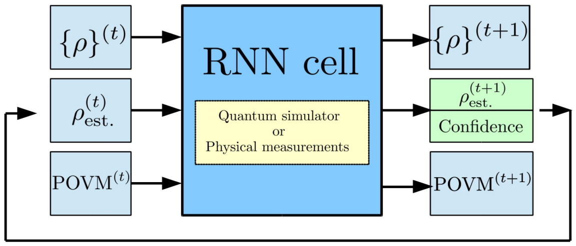


Image from Dawid et al. (2022)

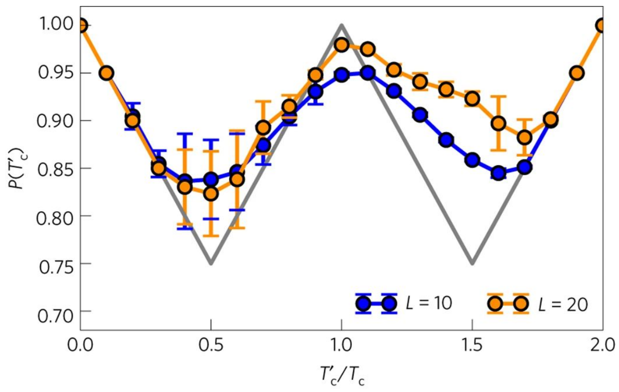
Quantum Tomography



(Not this talk!)

Quek et al. npj Quantum Inf. 7 (2021)

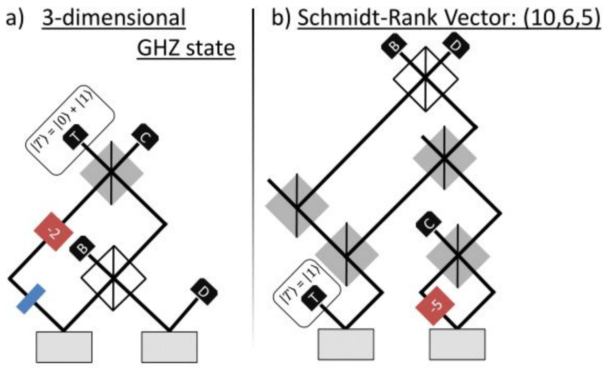
Phase Classification



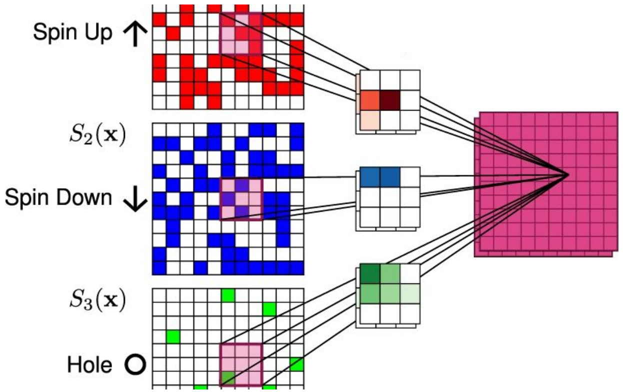
van Nieuwenburg et al. Nat. Phys. 13 (2017)

Context: Machine Learning as a Resource of Inspiration

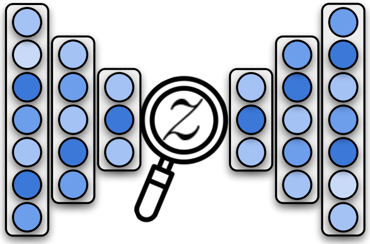
Identifying Surprises | **Interpretable Results** | **Interpretable Internal States**



Krenn et al. PRL 116 (2016)



Miles et al. Nat. Commun. 12 (2021)



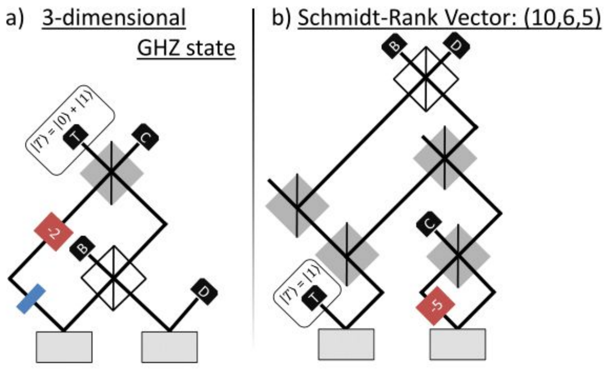
Frohnert et al. (2023)

Context: Machine Learning as a Resource of Inspiration

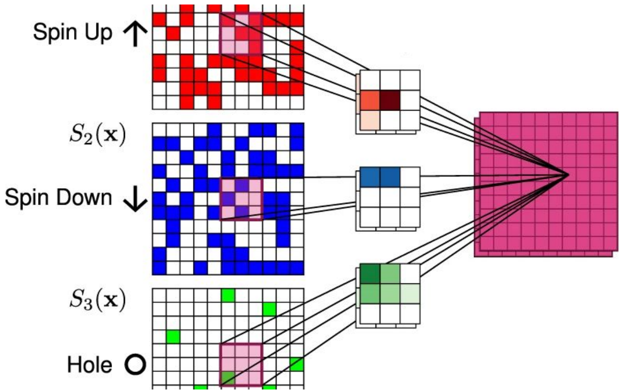
Identifying Surprises

Interpretable Results

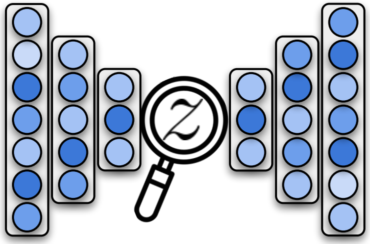
Interpretable Internal States



Krenn et al. PRL 116 (2016)



Miles et al. Nat. Commun. 12 (2021)

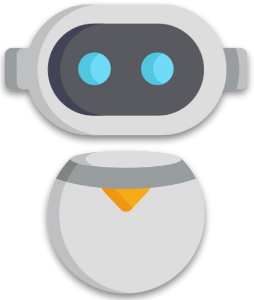


(This talk!)

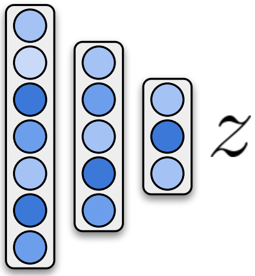
Frohnert et al. (2023)

Simple Case-Study: How do machines represent quantum states?

$$|\psi(\alpha)\rangle \longrightarrow$$



$$\longrightarrow$$



Quantum State

Machine Learning Agent

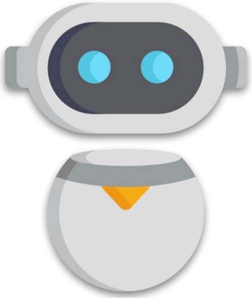
Compression Task

Simple Case-Study: How do machines represent quantum states?

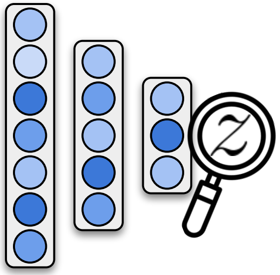
Investigate the Learned Internal Representation

$$|\psi(\alpha)\rangle$$

Quantum State



Machine Learning Agent



Compression Task

Simple Case-Study: How do machines represent quantum states?

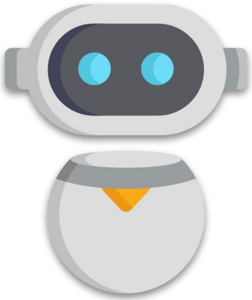
Investigate the Learned Internal Representation

1

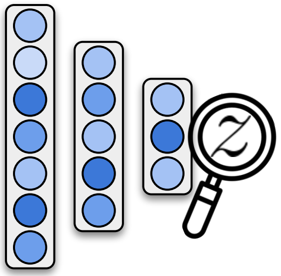
How does the agent learn to perform the compression?

$$|\psi(\alpha)\rangle$$

Quantum State



Machine Learning Agent



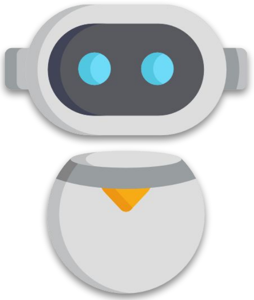
Compression Task

Simple Case-Study: How do machines represent quantum states?

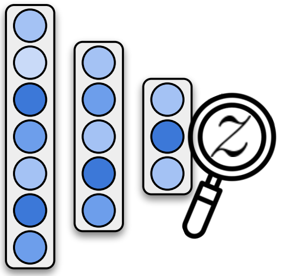
Investigate the Learned Internal Representation

1. How does the agent learn to perform the compression?
2. What (quantum) features are relevant for the agent?

$$|\psi(\alpha)\rangle \longrightarrow$$



$\longrightarrow$



Quantum State — Machine Learning Agent — Compression Task

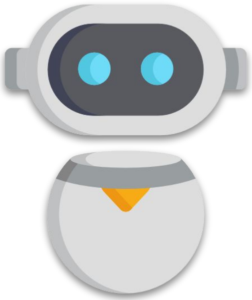
Simple Case-Study: How do machines represent quantum states?

Investigate the Learned Internal Representation

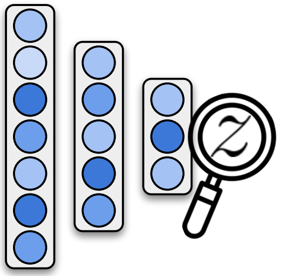
1. How does the agent learn to perform the compression?
2. What (quantum) features are relevant for the agent?
3. Is the learned representation human interpretable?

$$|\psi(\alpha)\rangle$$

Quantum State



Machine Learning Agent

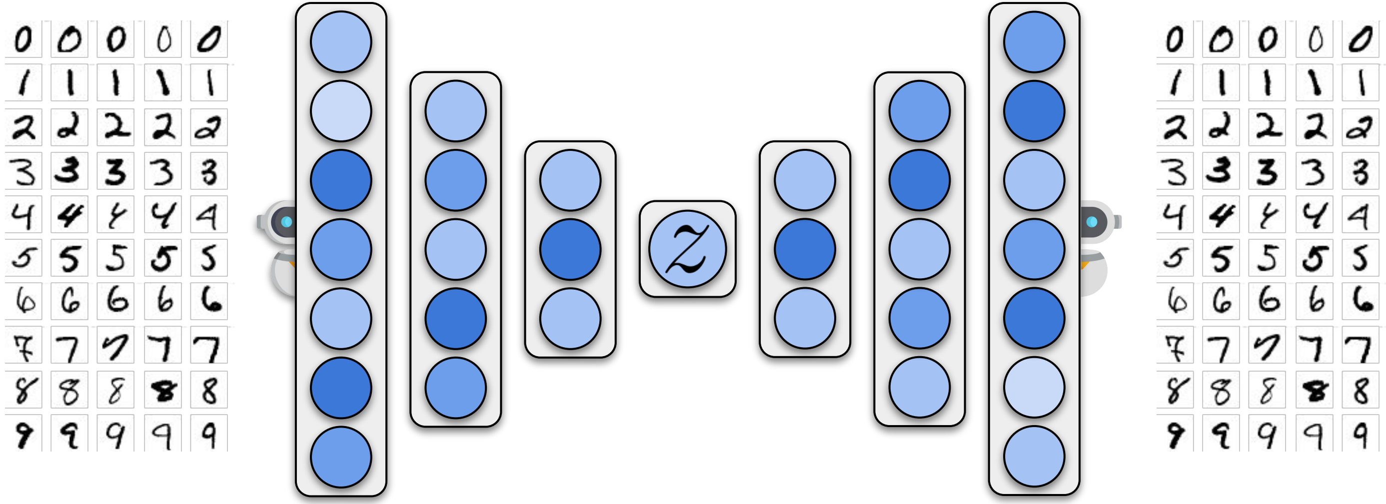


Compression Task

Methodology

Using Variational Autoencoders to extract interpretable features

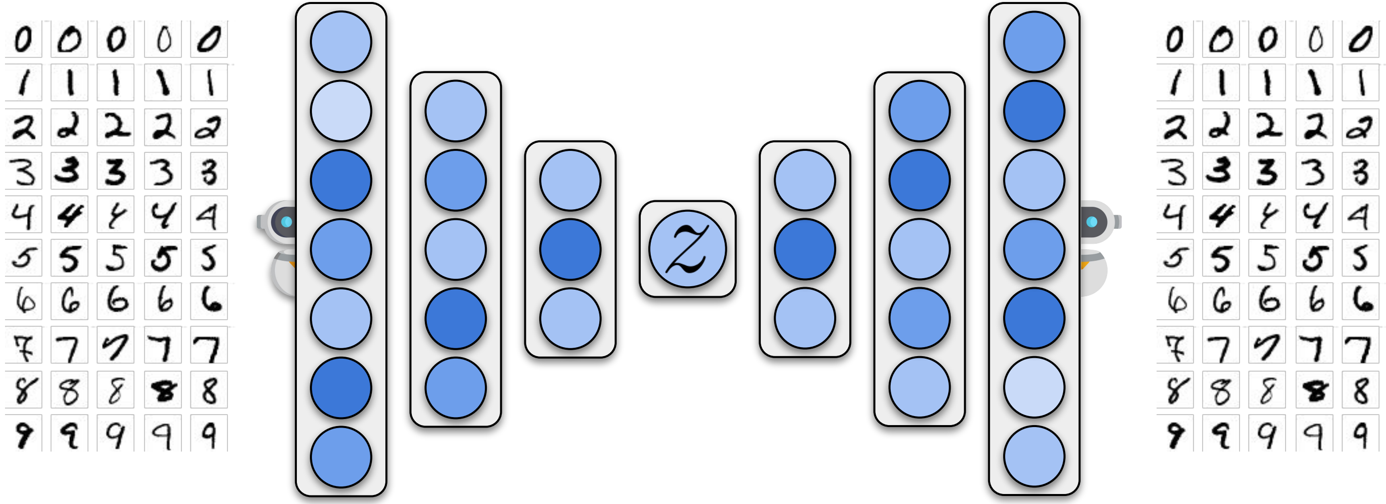
Neural-Network Autoencoders



Unsupervised learning of efficient compressed representation via information bottleneck

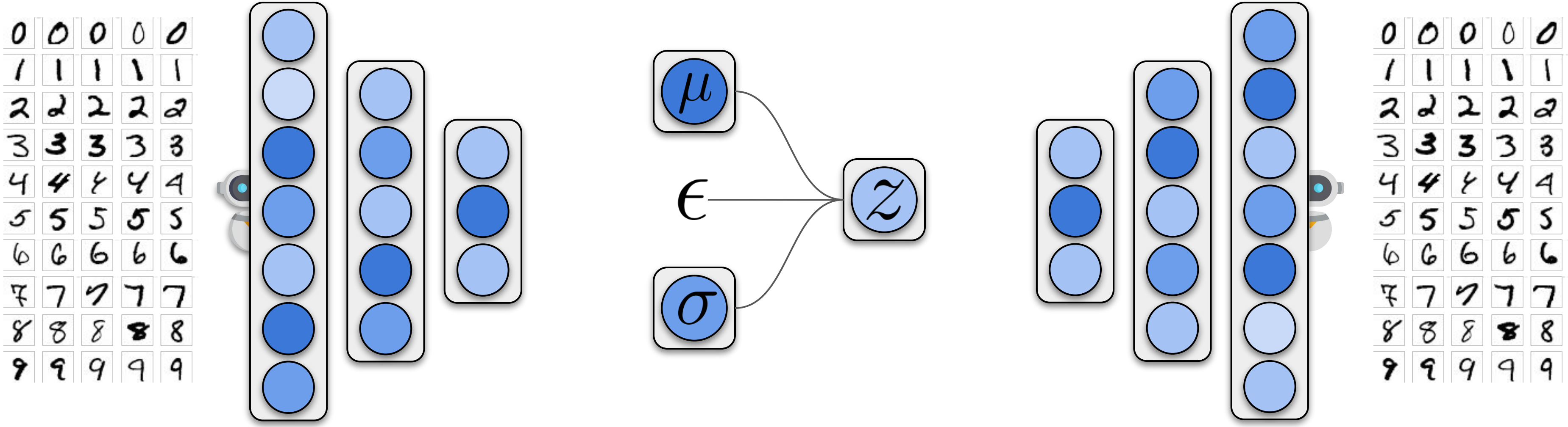
Kramer A. (1991)

Neural-Network Autoencoders



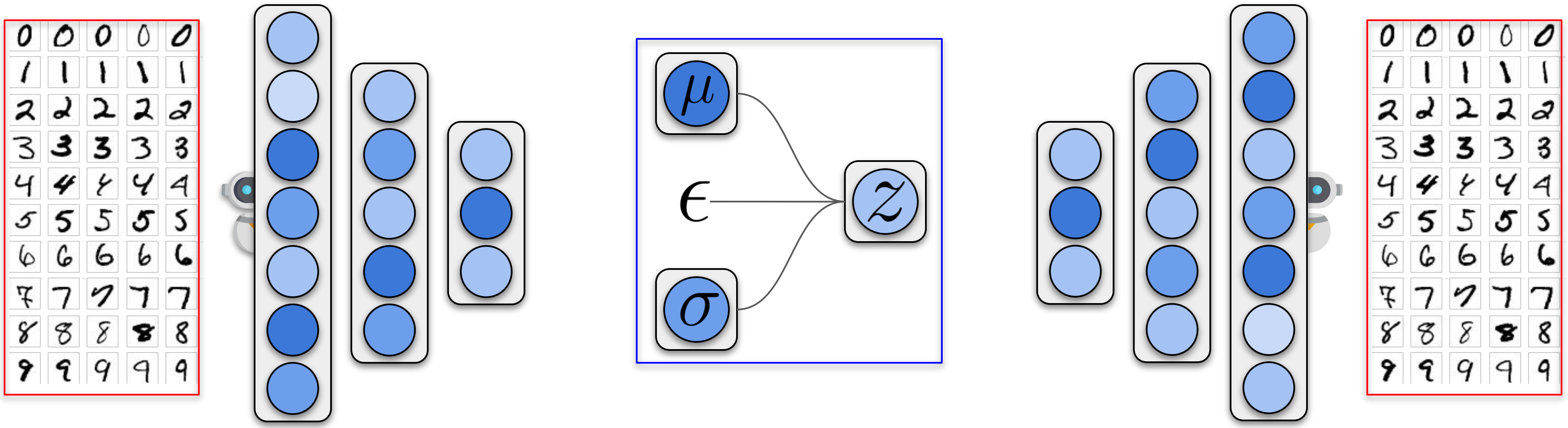
Latent space Z acts as feature extractor

Neural-Network Variational Autoencoders



Probabilistic encoding and decoding to generate a latent space with continuous and structured representations

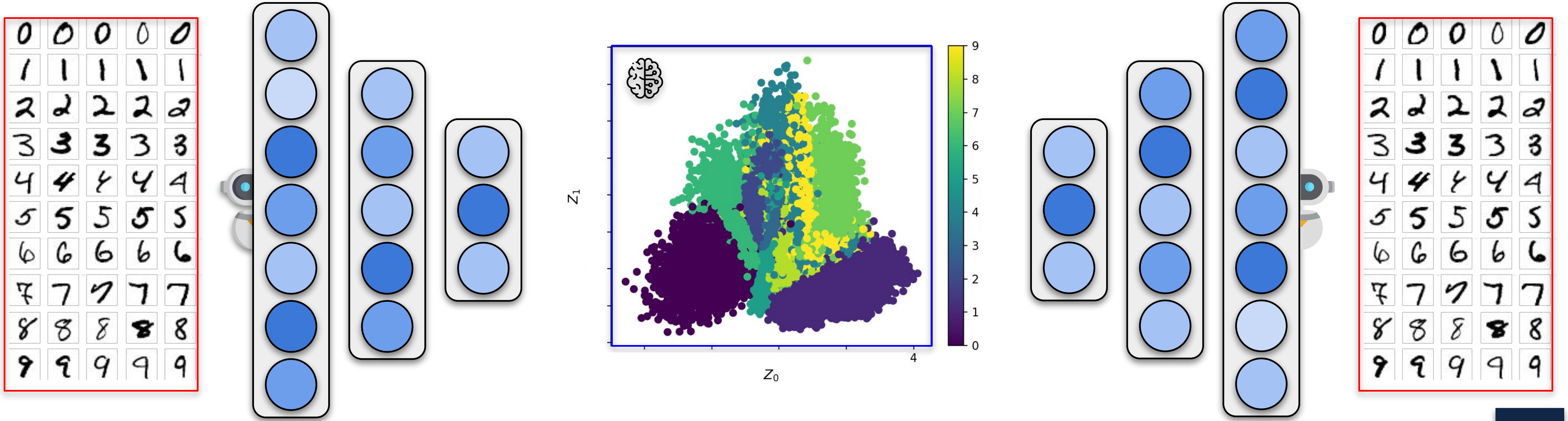
Neural-Network Variational Autoencoders



$$\mathcal{L}(\theta, \phi; \mathbf{x}) = \boxed{E_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x} \mid \mathbf{z})\right]} - \boxed{\beta} \cdot \boxed{KL(q_\phi(\mathbf{z} \mid \mathbf{x}) || p_\theta(\mathbf{z}))}$$

Neural-Network Variational Autoencoders

Why is this learned representation interesting?

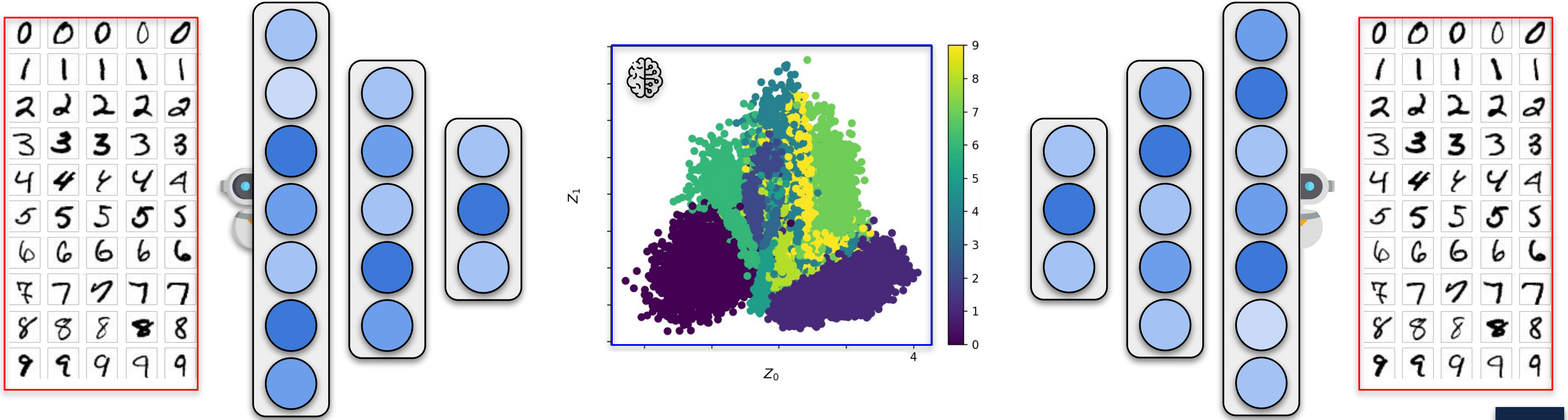


Neural-Network Variational Autoencoders

Why is this learned representation interesting?

1

Captures generative factors

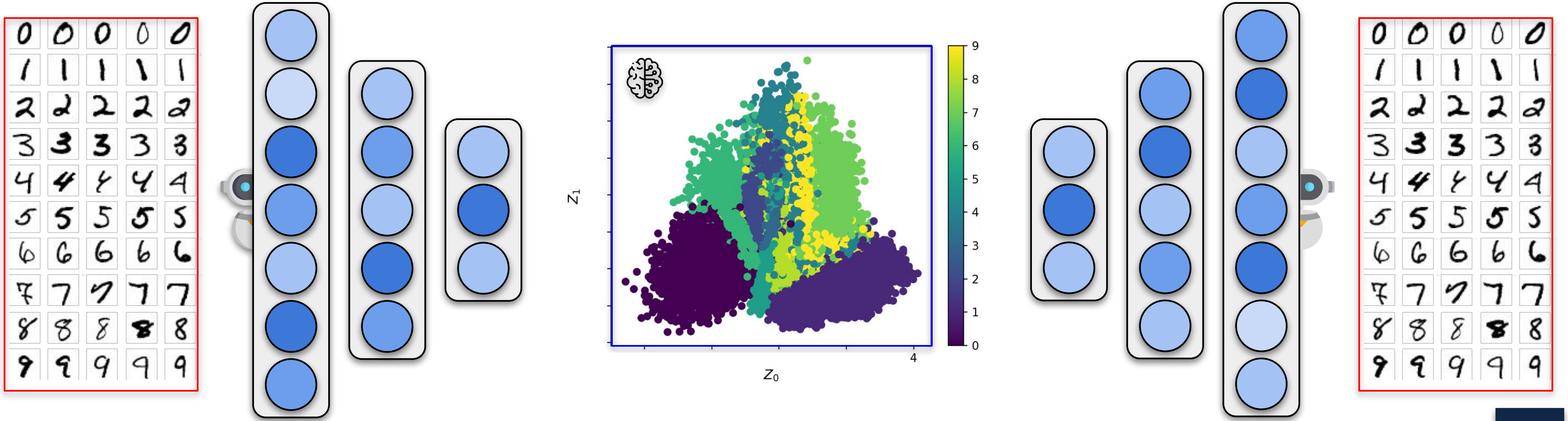


Neural-Network Variational Autoencoders

Why is this learned representation interesting?

1 Captures generative factors

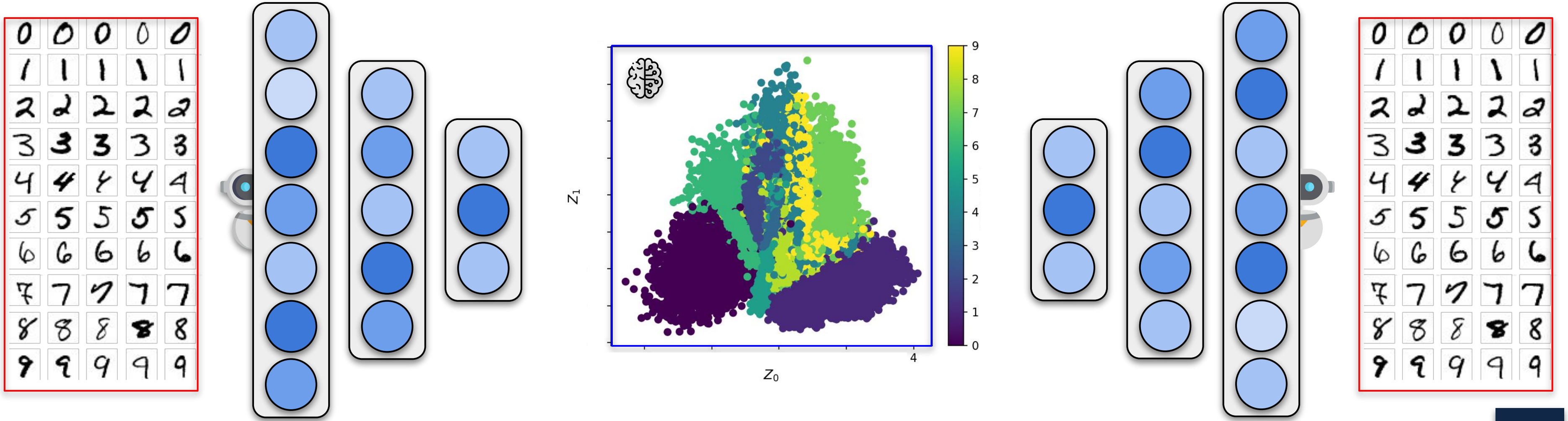
2 Interpretable/Meaningful



Neural-Network Variational Autoencoders

Why is this learned representation interesting?

1. Captures generative factors
2. Interpretable/Meaningful
3. Tunable regularization



Applications

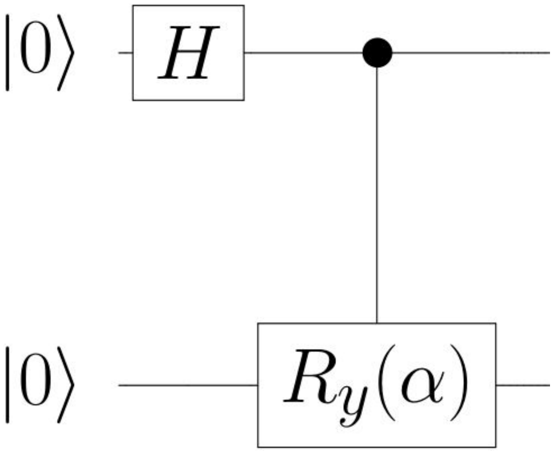
Proof-of-principle quantum system

Proof-of-Principle Quantum System with Non-Trivial Properties

Proof-of-Principle Quantum System with Non-Trivial Properties

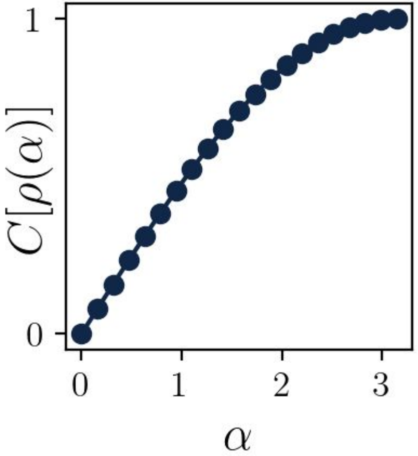
2-Qubit Quantum Circuit

Density Matrices Parametrized by α



Entanglement Properties

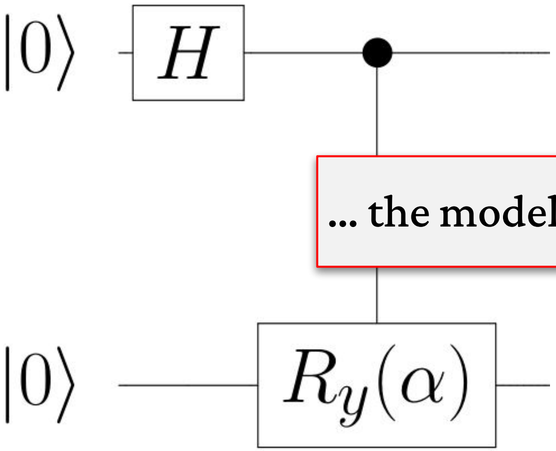
From separable to maximally entangled



Proof-of-Principle Quantum System with Non-Trivial Properties

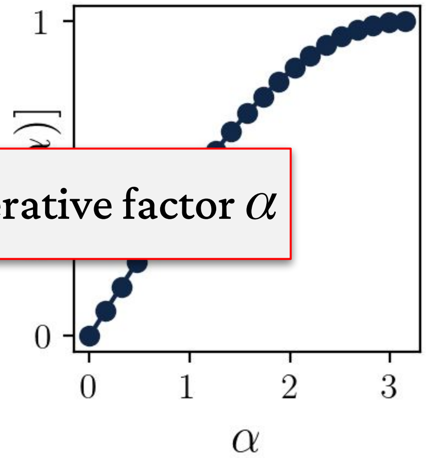
2-Qubit Quantum Circuit

Density Matrices Parametrized by α



Entanglement Properties

From separable to maximally entangled



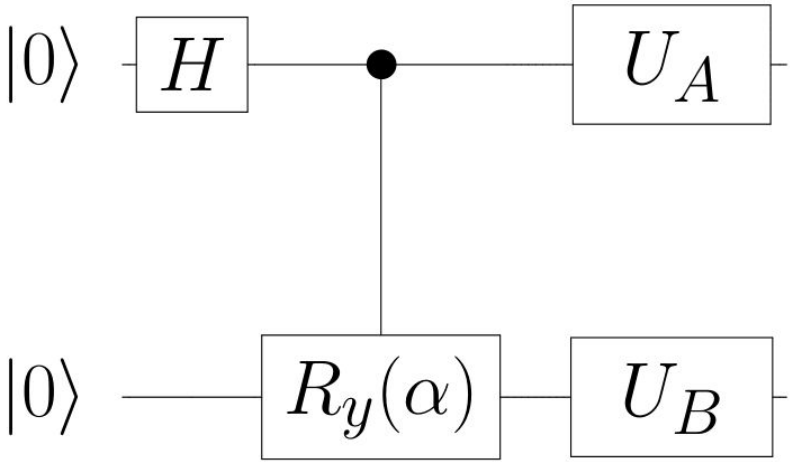
... the model would simply learn the generative factor α

Proof-of-Principle Quantum System with Non-Trivial Properties

2-Qubit Quantum Circuit

Density Matrices Parametrized by α

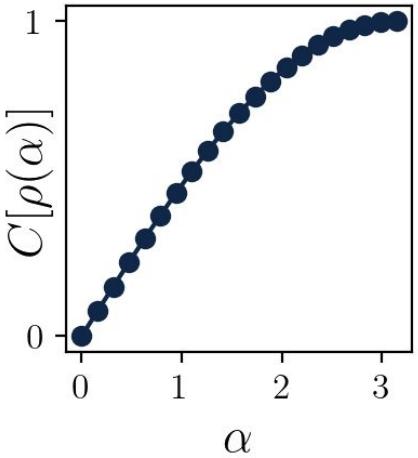
Unitaries randomize local information



Entanglement Properties

From separable to maximally entangled

Non-local features remain invariant

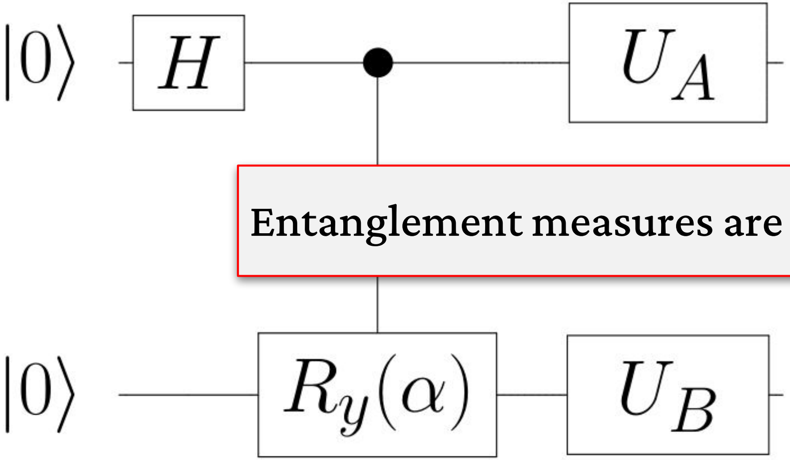


Proof-of-Principle Quantum System with Non-Trivial Properties

2-Qubit Quantum Circuit

Density Matrices Parametrized by α

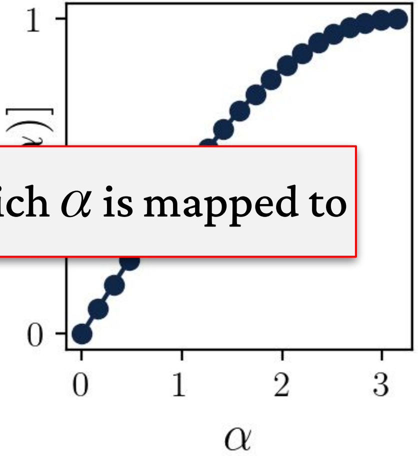
Unitaries randomize local information



Entanglement Properties

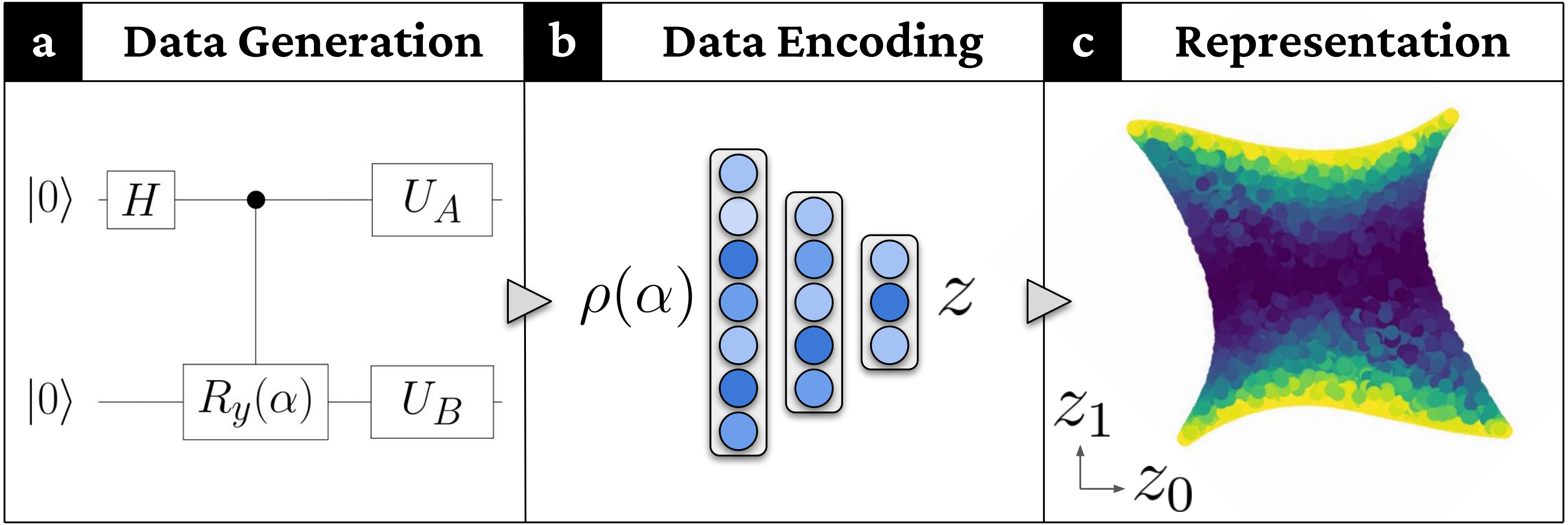
From separable to maximally entangled

Non-local features remain invariant



Entanglement measures are a quantity to which α is mapped to

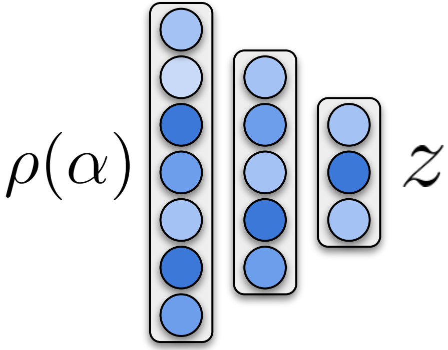
Intermediate Summary



Does the model learn to recognize entanglement properties, if so, in which representation?

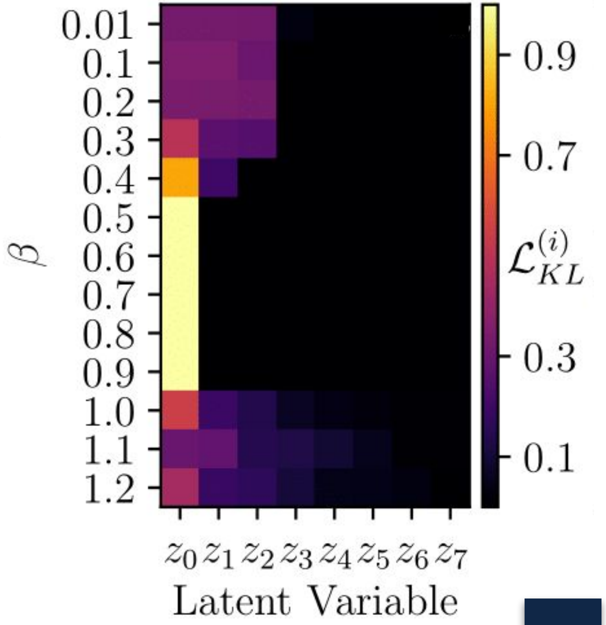
Tuning Regularization β for an Interpretable Representation

Encoding Quantum States · **Factorized Representations** · **Activation of Latent Space**



$$\mathcal{L}_{KL}^{(i)} = \beta \cdot KL(q_\phi(\mathbf{z} \mid \mathbf{x}) || p_\theta(\mathbf{z}))$$

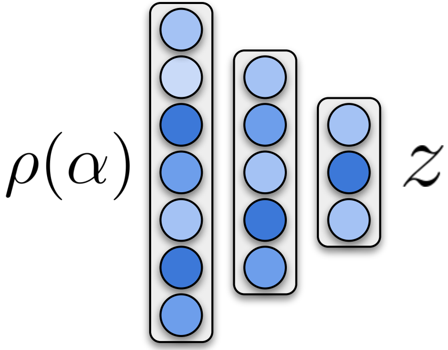
Tune β by measuring individual latent variable activity



Burgess et al. (2018)

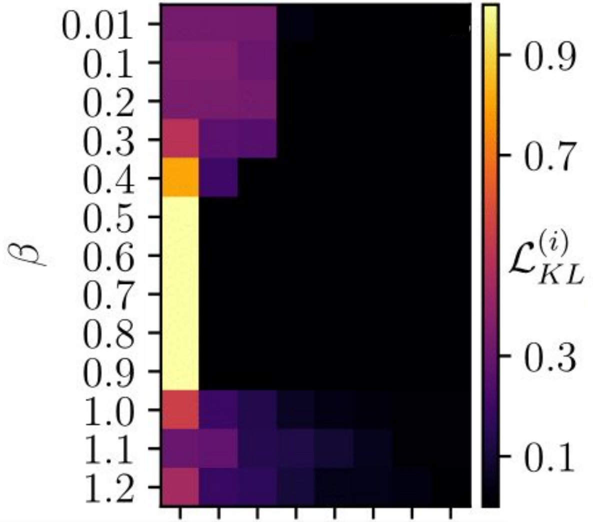
Tuning Regularization β for an Interpretable Representation

Encoding Quantum States | **Factorized Representations** | **Activation of Latent Space**



$$\mathcal{L}_{KL}^{(i)} = \beta \cdot KL(q_\phi(\mathbf{z} \mid \mathbf{x}) || p_\theta(\mathbf{z}))$$

Tune β by measuring individual latent variable activity



Goal: Find representation where #active variables equals #generative factors in dataset

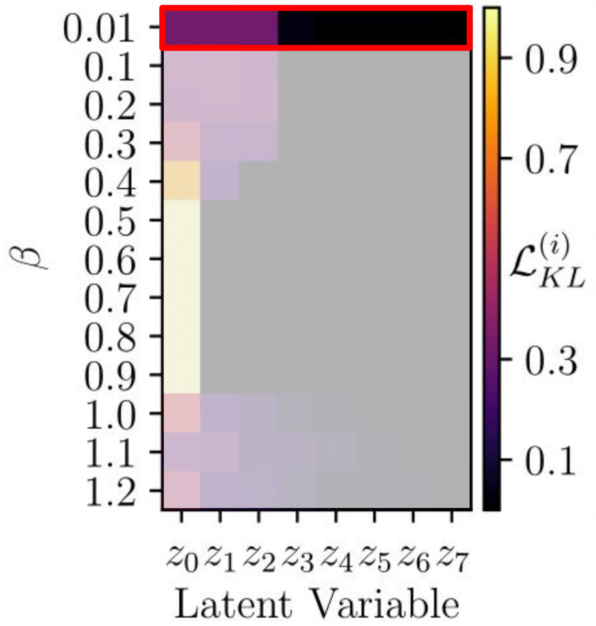
Burgess et al. (2018)

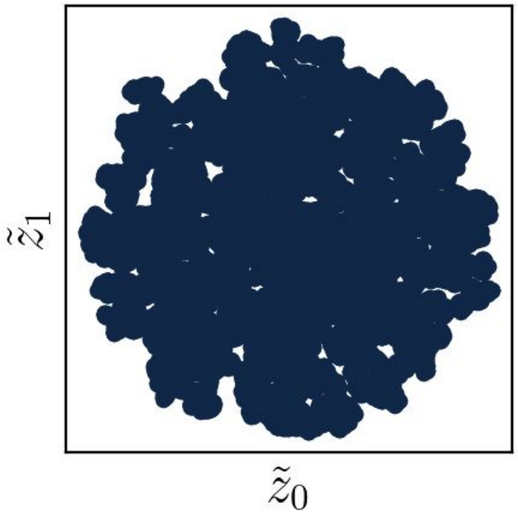
Investigating the Representation at Low β

Activation of Latent Space

t-SNE of Latent Space

Interpretation of Result





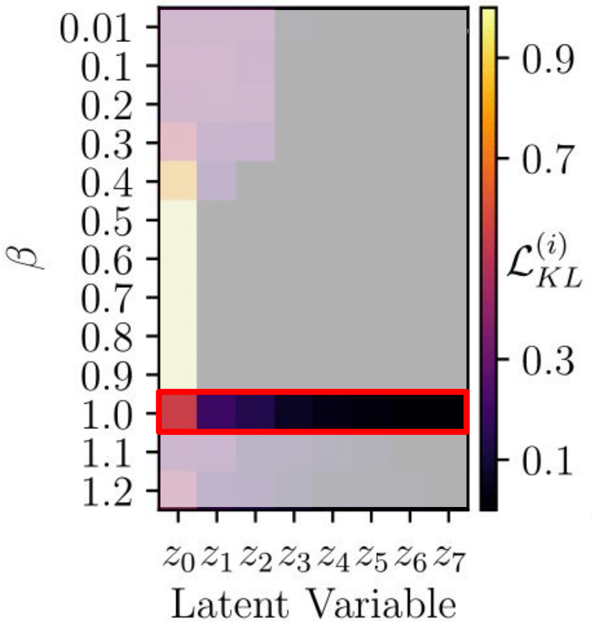
Efficient but with no easily interpretable structure

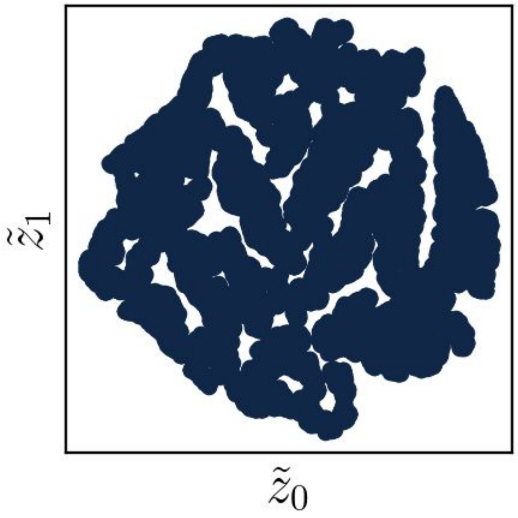
Investigating the Representation at High β

Activation of Latent Space

t-SNE of Latent Space

Interpretation of Result





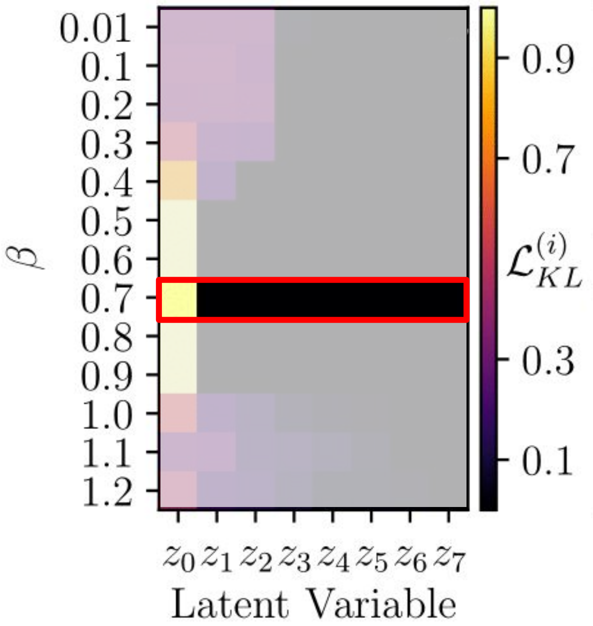
Posterior collapse with no easily interpretable structure

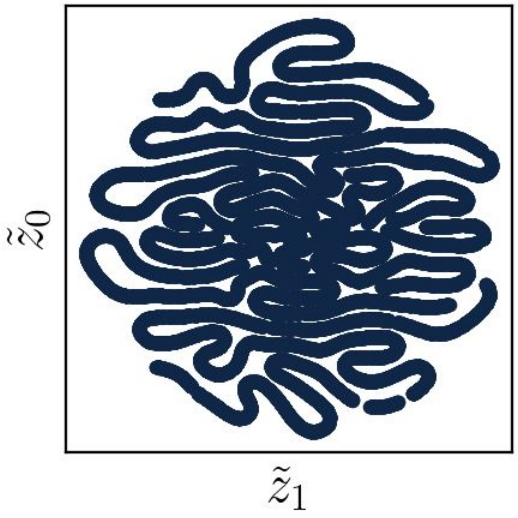
Investigating the Representation at Tuned β

Activation of Latent Space

t-SNE of Latent Space

Interpretation of Result





Something

Interesting

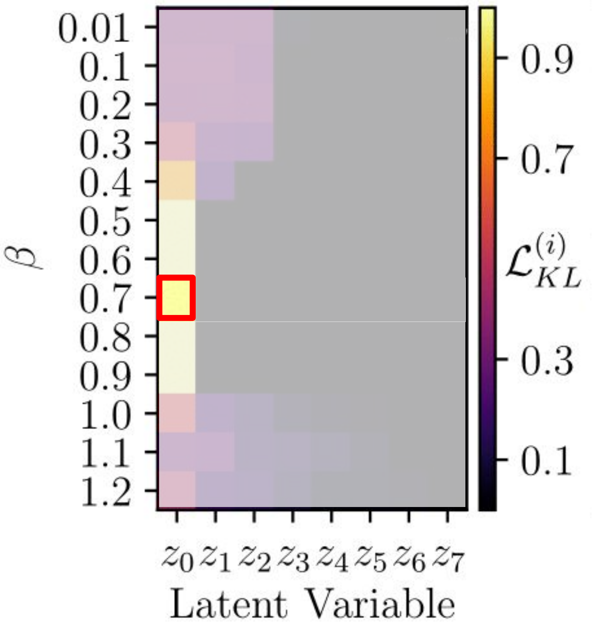
Happened!

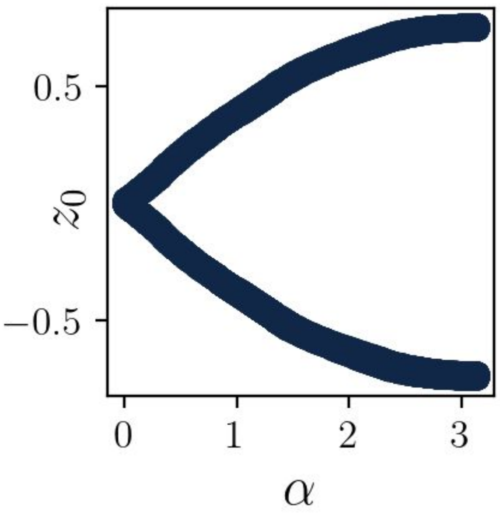
Investigating the Representation at Tuned β

Activation of Latent Space

Single Latent Variable

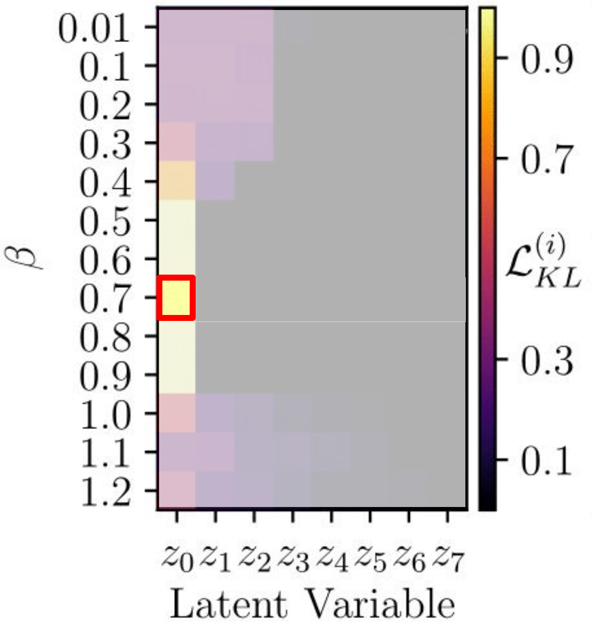
Interpretation of Result



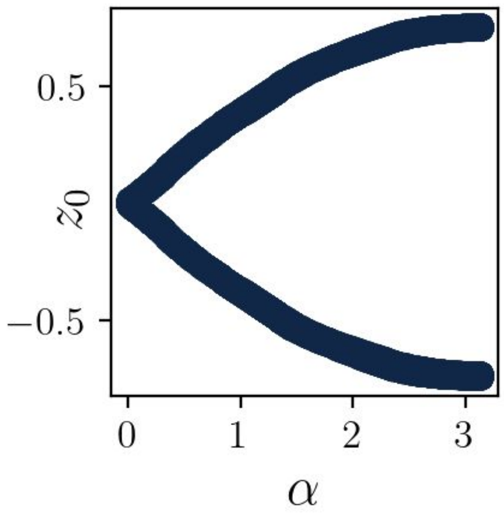


Investigating the Representation at Tuned β

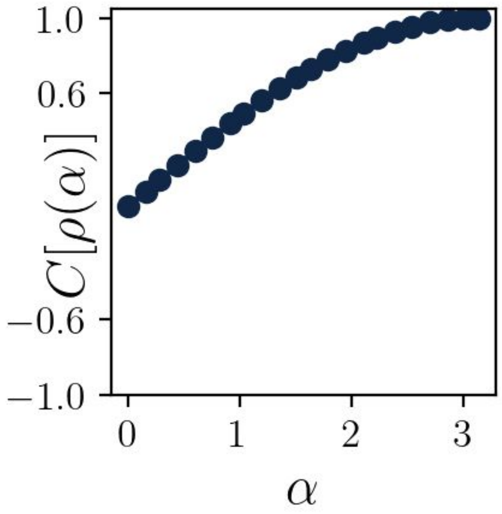
Activation of Latent Space



Single Latent Variable

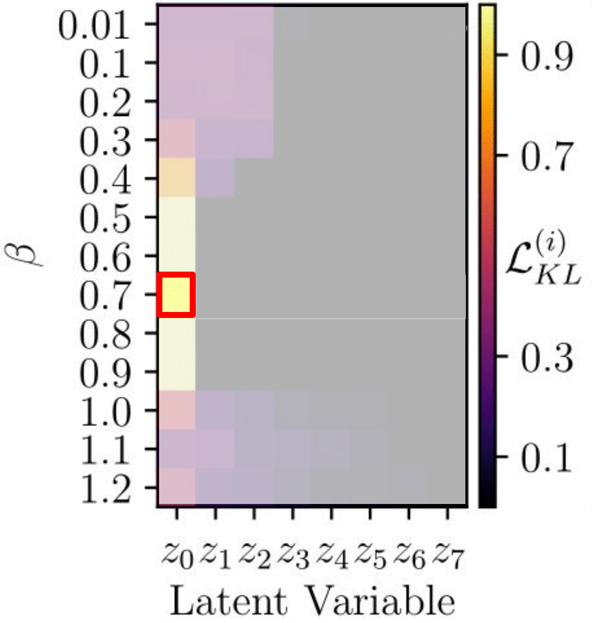


Interpretation of Result

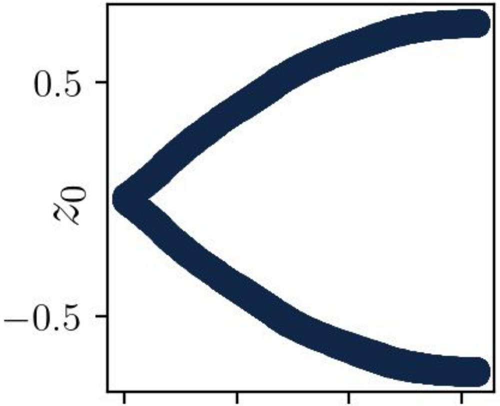


Investigating the Representation at Tuned β

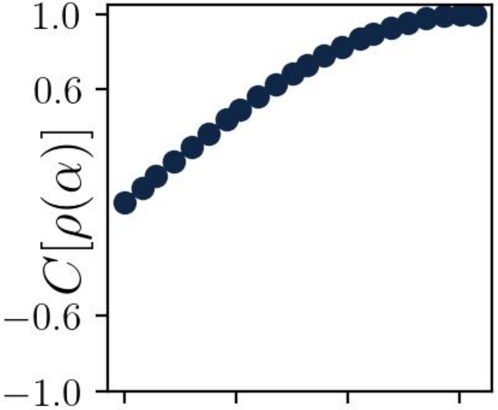
Activation of Latent Space



Single Latent Variable



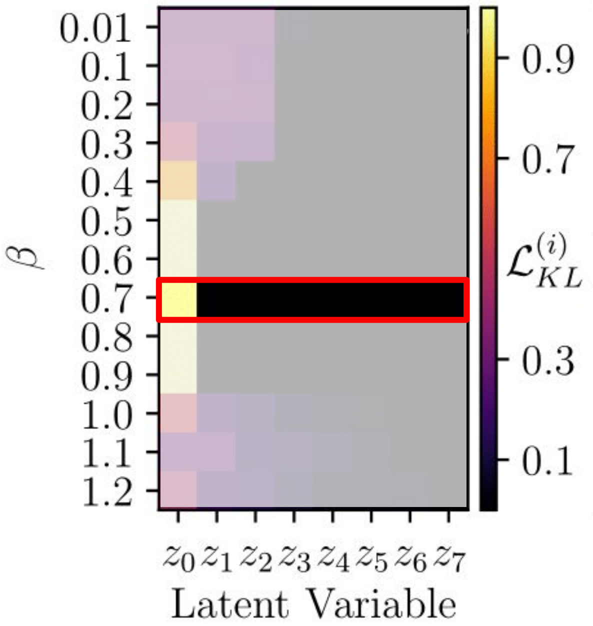
Interpretation of Result



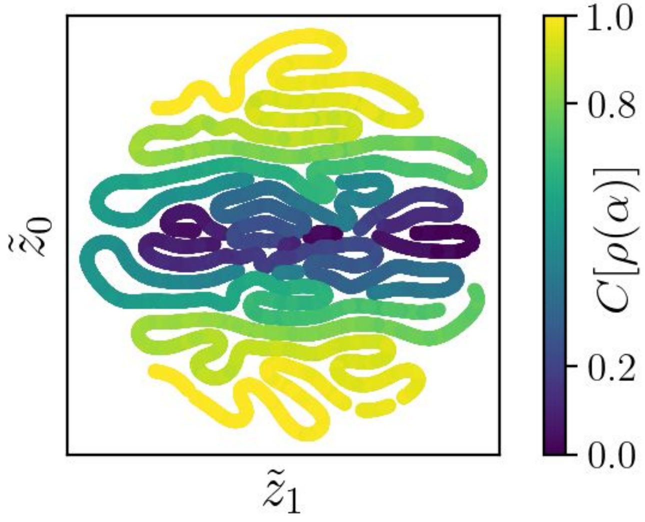
Symmetry around 0 due to regularization against normal prior

Investigating the Representation at Tuned β

Activation of Latent Space



t-SNE of Latent Space



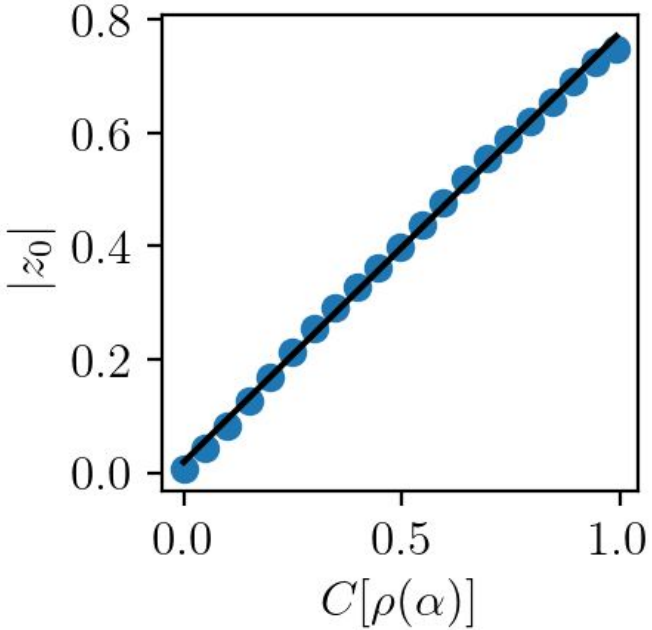
Interpretation of Result

Representation is structured with entanglement information

Using the Representation as Resource of Inspiration

Using the Representation as Resource of Inspiration

Relation to Concurrence



Machine-Designed Entanglement Measure

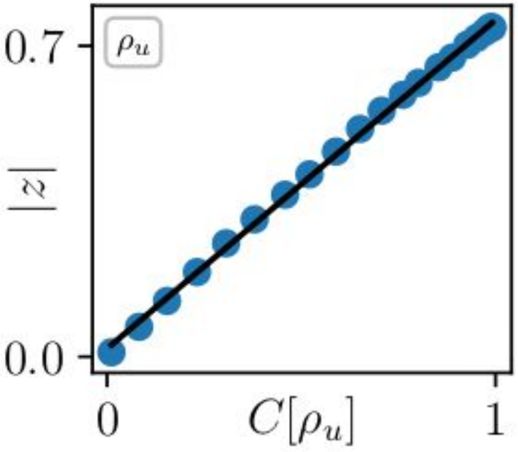
1. The agent finds that entanglement is important for this quantum system.

2. Use the representation to approximate an entanglement monotone with equal information content as concurrence (negativity).

Using the Representation as Resource of Inspiration

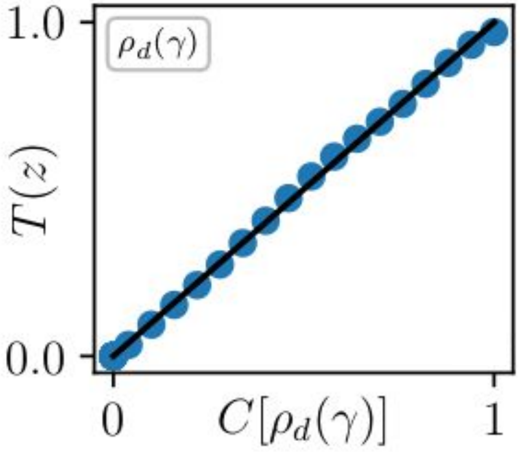
Generalizability to Random Pure States

$$\rho_u = U_{AB}|00\rangle\langle 00|U_{AB}^\dagger$$



Generalizability to Mixed States

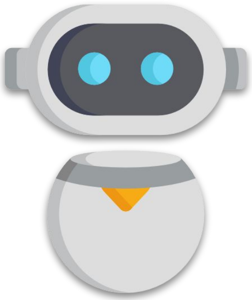
$$\rho_d(\gamma) = (1-\gamma)\rho(\pi) + \frac{\gamma}{2}\mathbb{1}$$



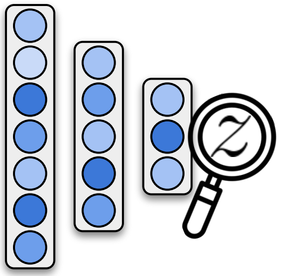
The agent constructs its own understanding of entanglement

Let’s summarize what we have learned

$$|\psi(\alpha)\rangle \longrightarrow$$



$\longrightarrow$



Quantum State | Machine Learning Agent | Compression Task

The agent constructs its own understanding of entanglement

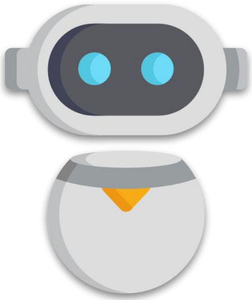
Let's summarize what we have learned

1

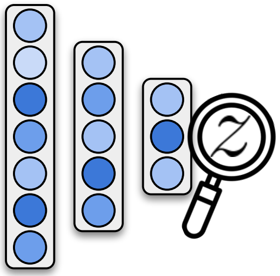
The agent learned to extract entanglement information

$$|\psi(\alpha)\rangle$$

Quantum State



Machine Learning Agent



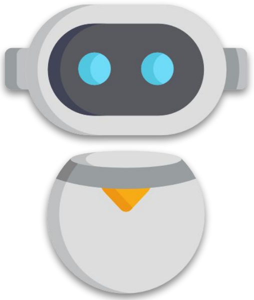
Compression Task

The agent constructs its own understanding of entanglement

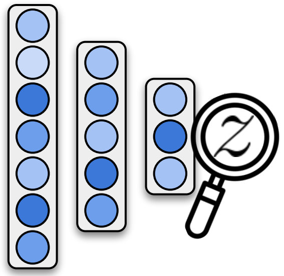
Let's summarize what we have learned

1. The agent learned to extract entanglement information
2. The representation of entanglement can be compacted into a single number

$$|\psi(\alpha)\rangle \longrightarrow$$



$\longrightarrow$



Quantum State

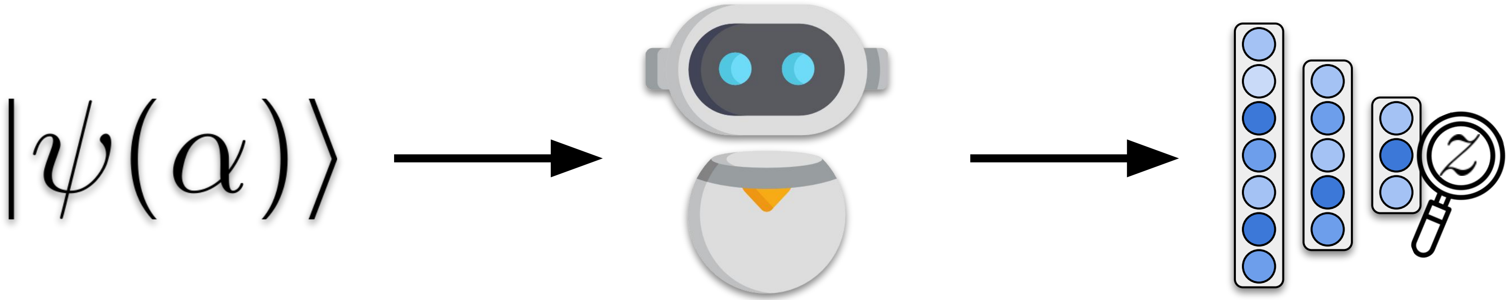
Machine Learning Agent

Compression Task

The agent constructs its own understanding of entanglement

Let's summarize what we have learned

1. The agent learned to extract entanglement information
2. The representation of entanglement can be compacted into a single number
3. The discovered entanglement monotone matches concurrence



Quantum State — Machine Learning Agent — Compression Task

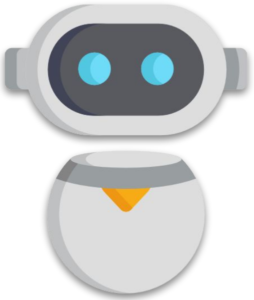
The agent constructs its own understanding of entanglement

Let's summarize what we have learned

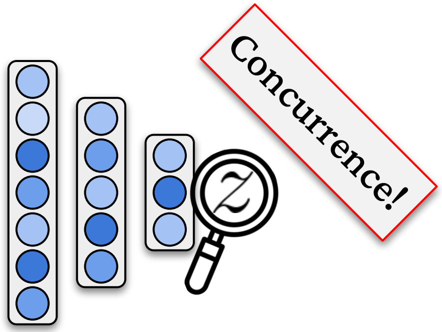
1. The agent learned to extract entanglement information
2. The representation of entanglement can be compacted into a single number
3. The discovered entanglement monotone matches concurrence

$$|\psi(\alpha)\rangle$$

Quantum State



Machine Learning Agent



Audience Question

Outlook

Extending the methodology to larger system sizes

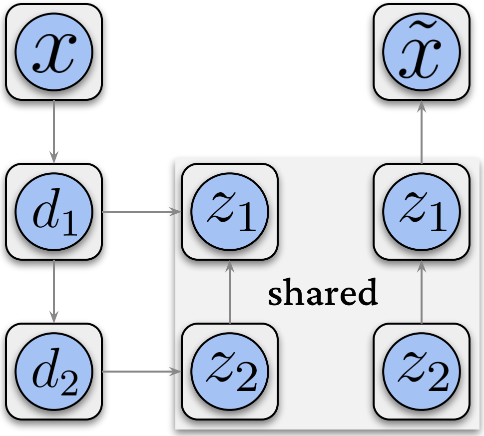
Outlook

Possible Extension

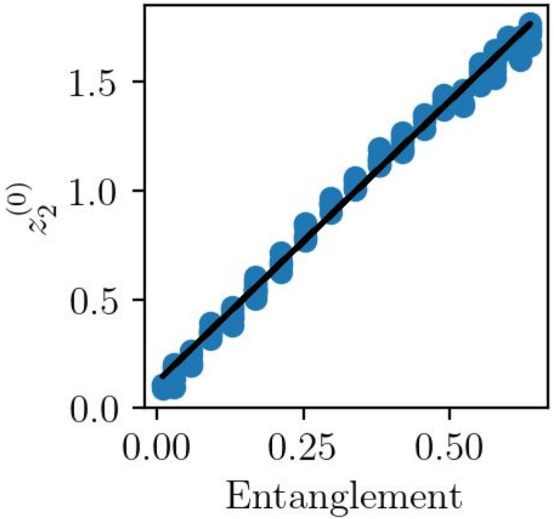
Ladder-VAE

W-like Entangled States

1. 3-Qubit Entanglement ✓
2. Multiple parameters
3. N-Qubit Entanglement
4. Changing the Input Data



Hierarchical Latent Representation



Contact

Felix Frohnert

f.frohnert@
liacs.leidenuniv.nl

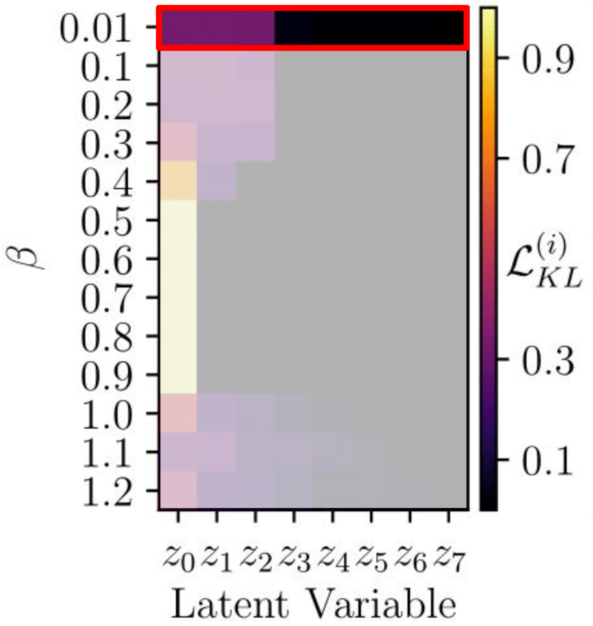
arXiv:2306.05694

Evert van Nieuwenburg

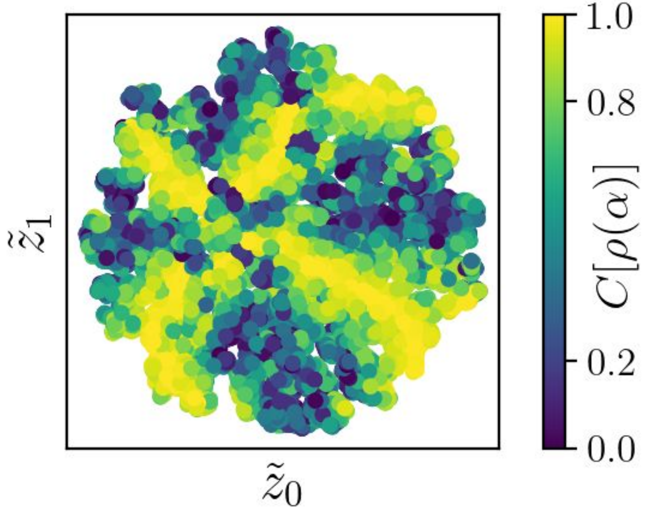
CAI Group

Investigating the representation at low β

Activation of Latent Space



t-SNE of Latent Space

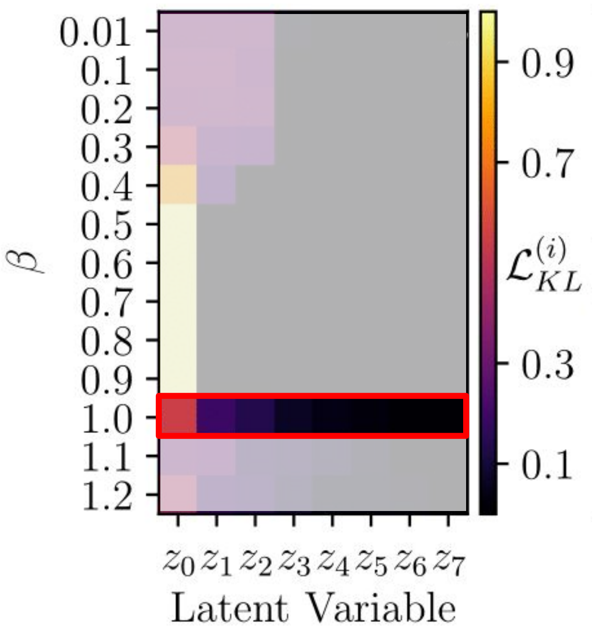


Interpretation of Result

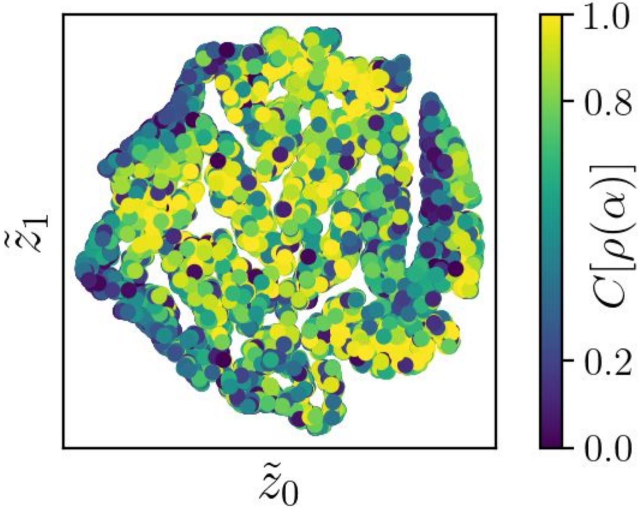
Efficient but no clear interpretable structure

Investigating the representation at high β

Activation of Latent Space



t-SNE of Latent Space



Interpretation of Result

Posterior Collapse