Machine Learning Physics
From Quantum Mechanics to Quantum Field Theory

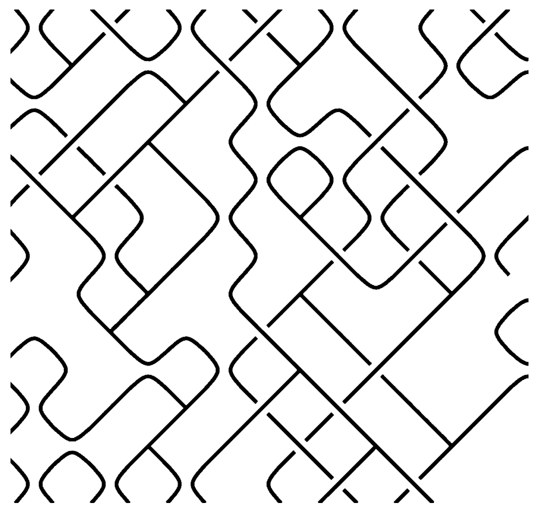
Yi-Zhuang You
University of California, San Diego

July, 2020

Machine Learning Physics

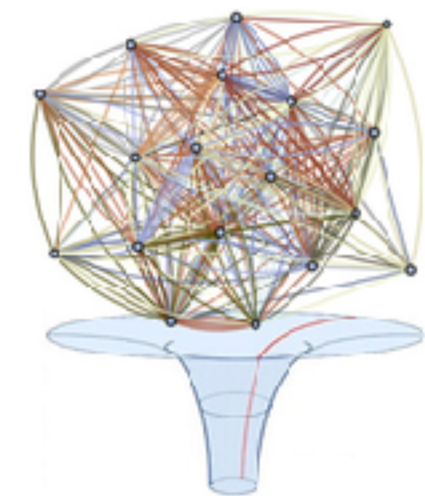
- Emergent phenomenon — a central theme of condensed matter physics.

Weyl semimetal
(emergent particle)



String net condensation
(emergent force)

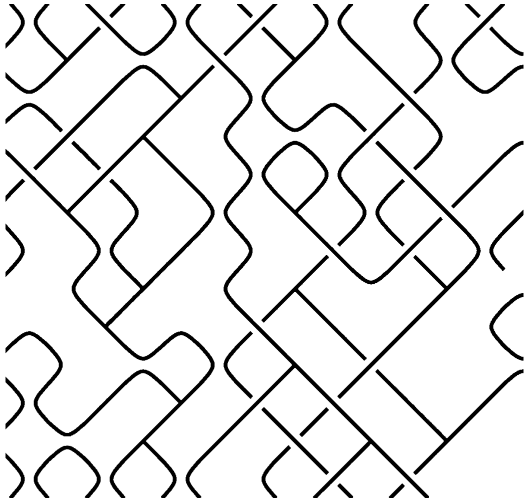
ER = EPR
(emergent spacetime)

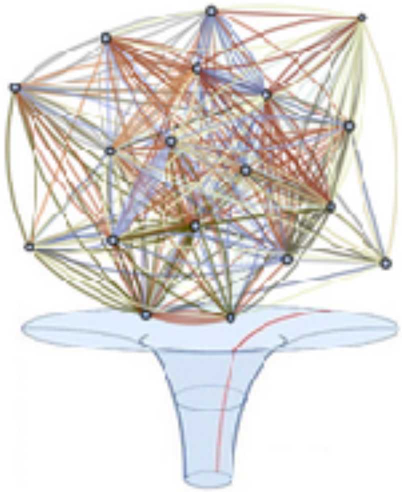


SYK model
(emergent gravity)

Machine Learning Physics

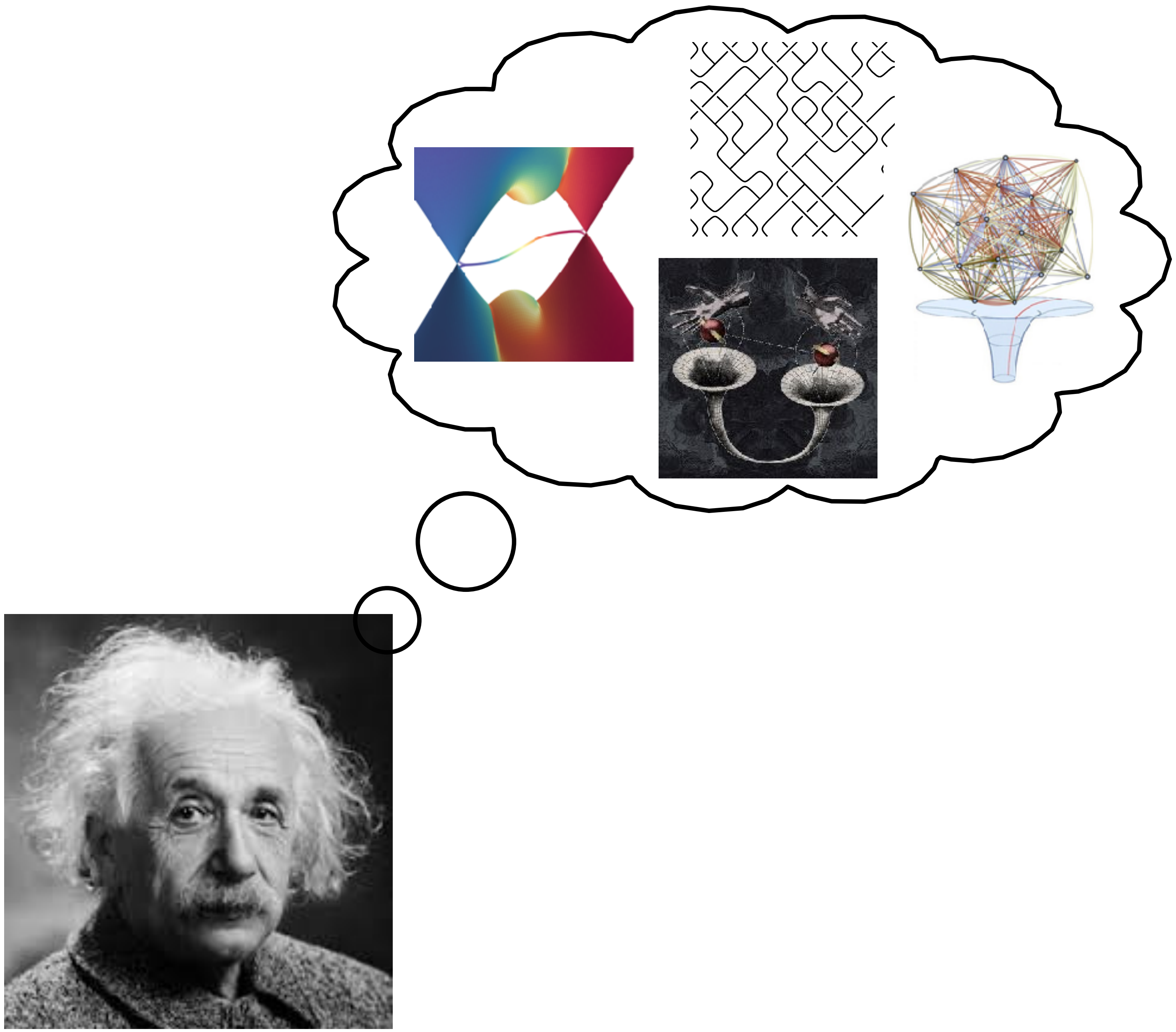
- Aren't all these physics theories themselves also emergent phenomena?



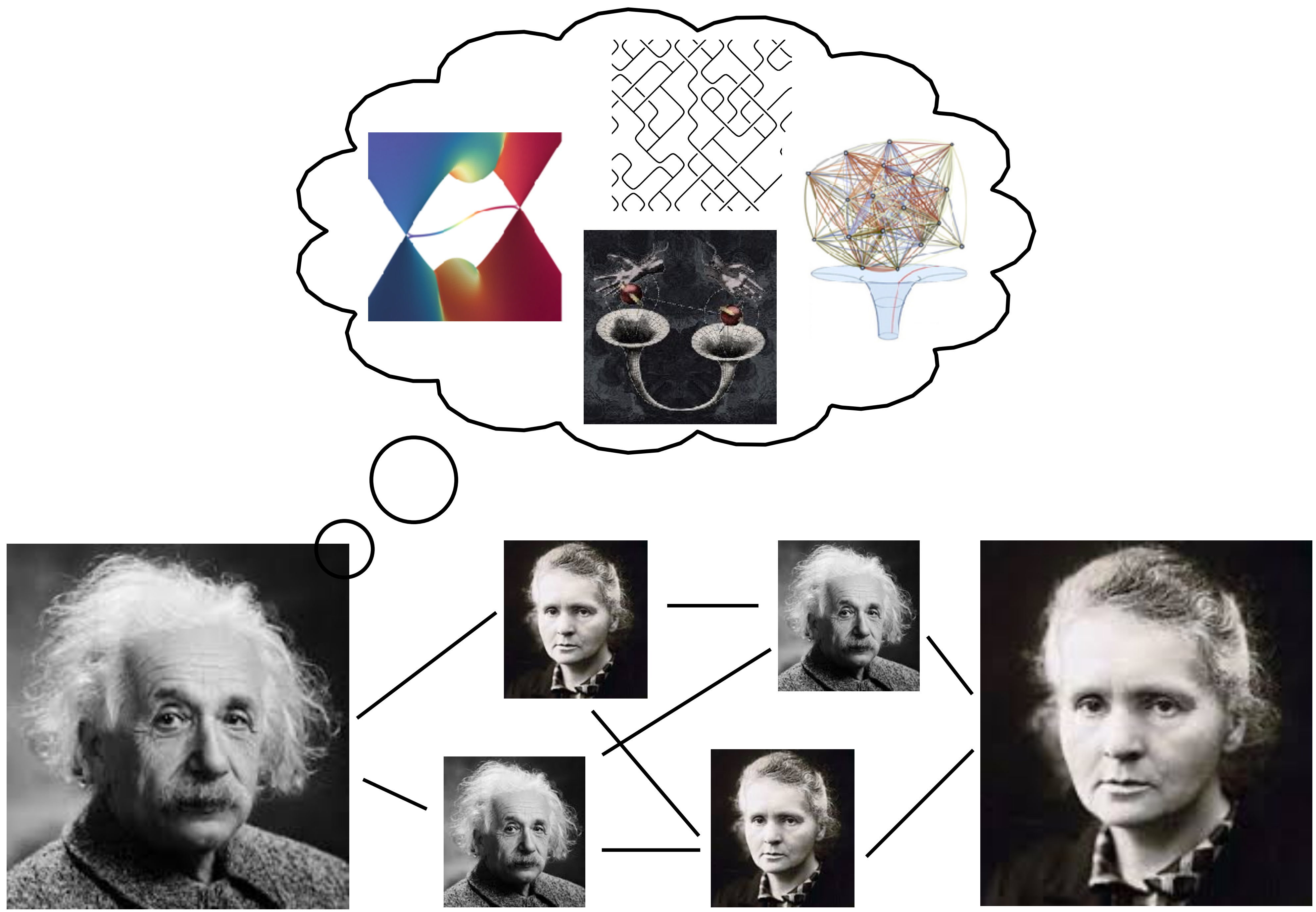
Machine Learning Physics

- Aren't all these physics theories themselves also emergent phenomena?



Machine Learning Physics

- Aren't all these physics theories themselves also emergent phenomena?

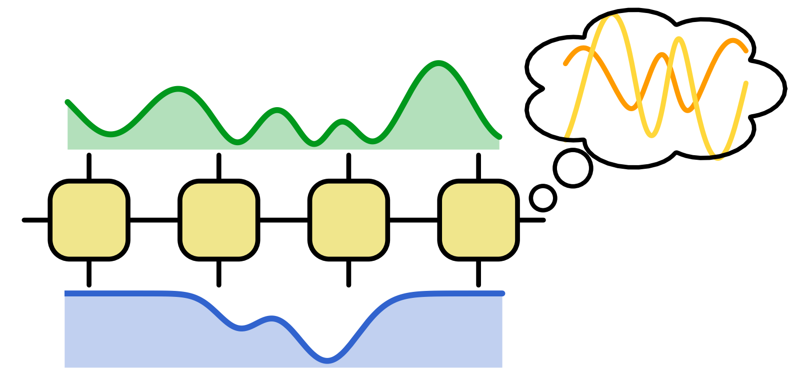


Machine Learning Physics

- Goal: investigate whether artificial neural networks can be used to discover physical concepts and laws from observation data.

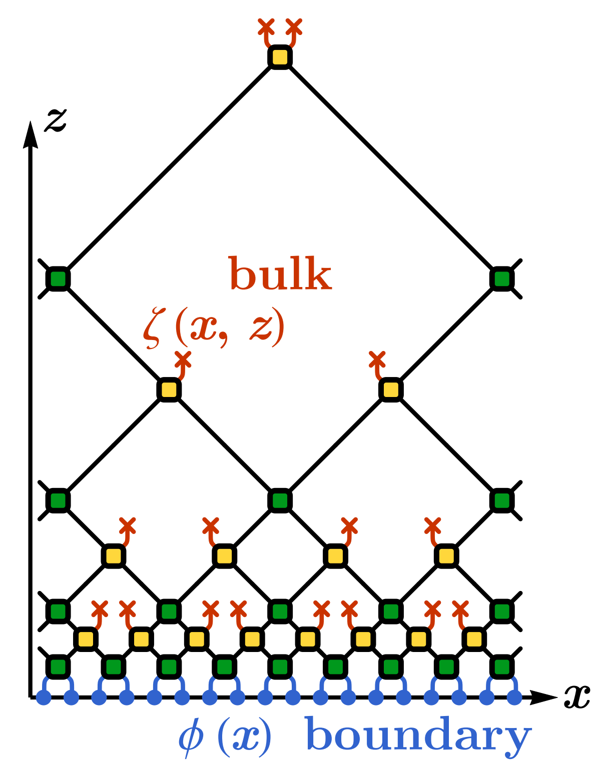
- Examples
  - Machine learning quantum mechanics — recurrent autoencoder

    C Wang, H Zhai, Y-Z You. arXiv: 1901.11103



  - Machine learning renormalization group and holographic duality — flow-based deep generative model

    H Hu, S-H Li, L Wang, Y-Z You. arXiv: 1903.00804



Machine Learning Physics

- Goal: investigate whether artificial neural networks can be used to discover physical concepts and laws from observation data.

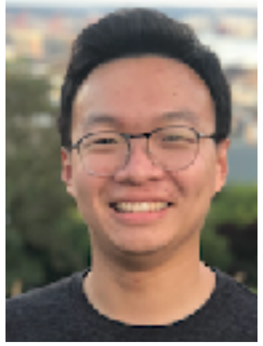
- Examples
  - Machine learning quantum mechanics — recurrent autoencoder

    C Wang, H Zhai, Y-Z You. arXiv: 1901.11103

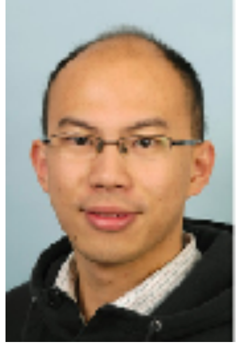
  - Machine learning renormalization group and holographic duality — flow-based deep generative model

    H Hu, S-H Li, L Wang, Y-Z You. arXiv: 1903.00804


Ce Wang Hui Zhai
(Tsinghua University)



Hong-Ye Hu
(UCSD)

Shuo-Hui Li Lei Wang
(IOP, CAS)

Machine Learning
Quantum Mechanics

Potential and Density Data

- Suppose quantum mechanics has not been formulated so far
- Yet, amazingly, we know how to perform cold atom experiments of Bose-Einstein condensate (BEC)

Potential profile
(optical speckles,
optical tweezers …)

BEC Density profile
(in-situ measurement)

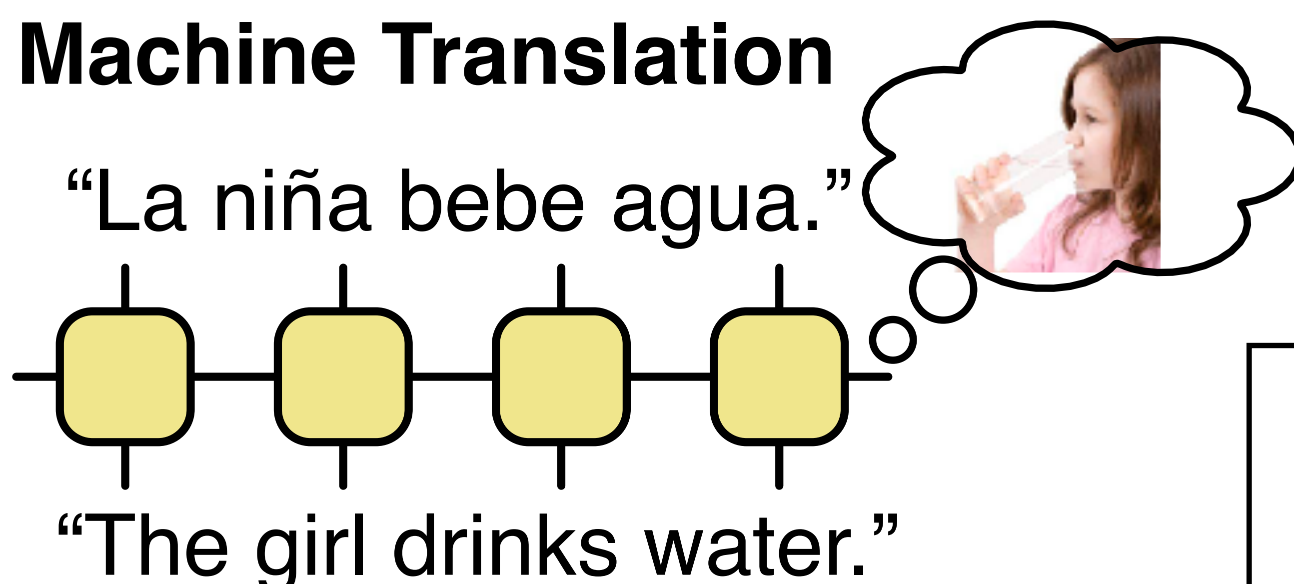
Billy et. al., Nature (2008),
Henderson et.al., NJP (2009) …

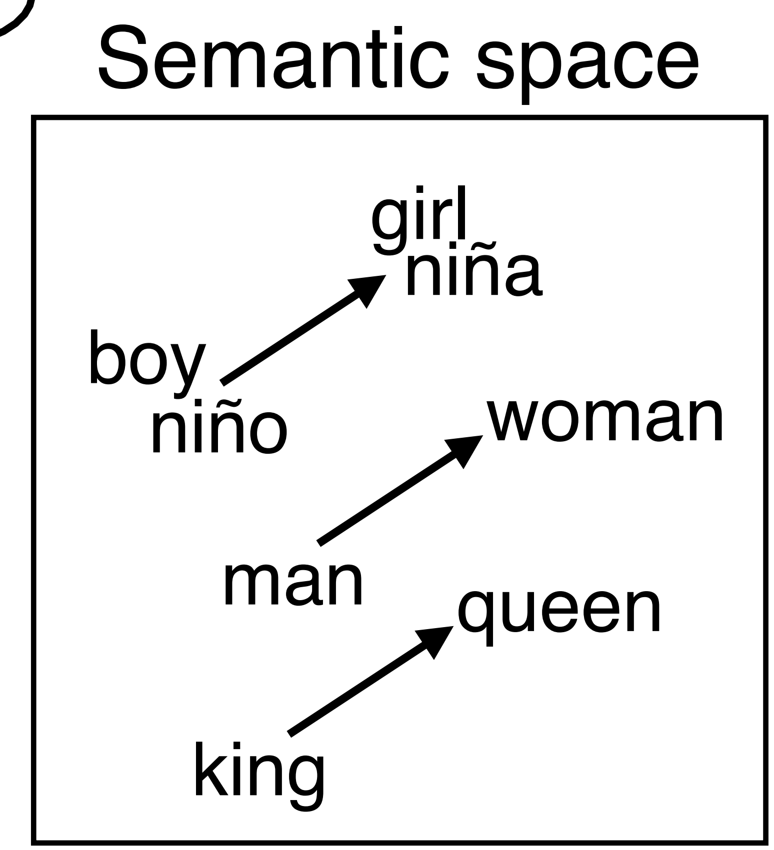
Issac Tamblyn's talk

- Questions
  - Can quantum mechanics (QM) be discovered as the most natural theory to explain the experiment?
  - Will the machine develop alternative form of QM?

Inspiration from Machine Translation

- Motivation: developments in machine translation
  - Sequence-to-sequence mapping (RNN, LSTM …)





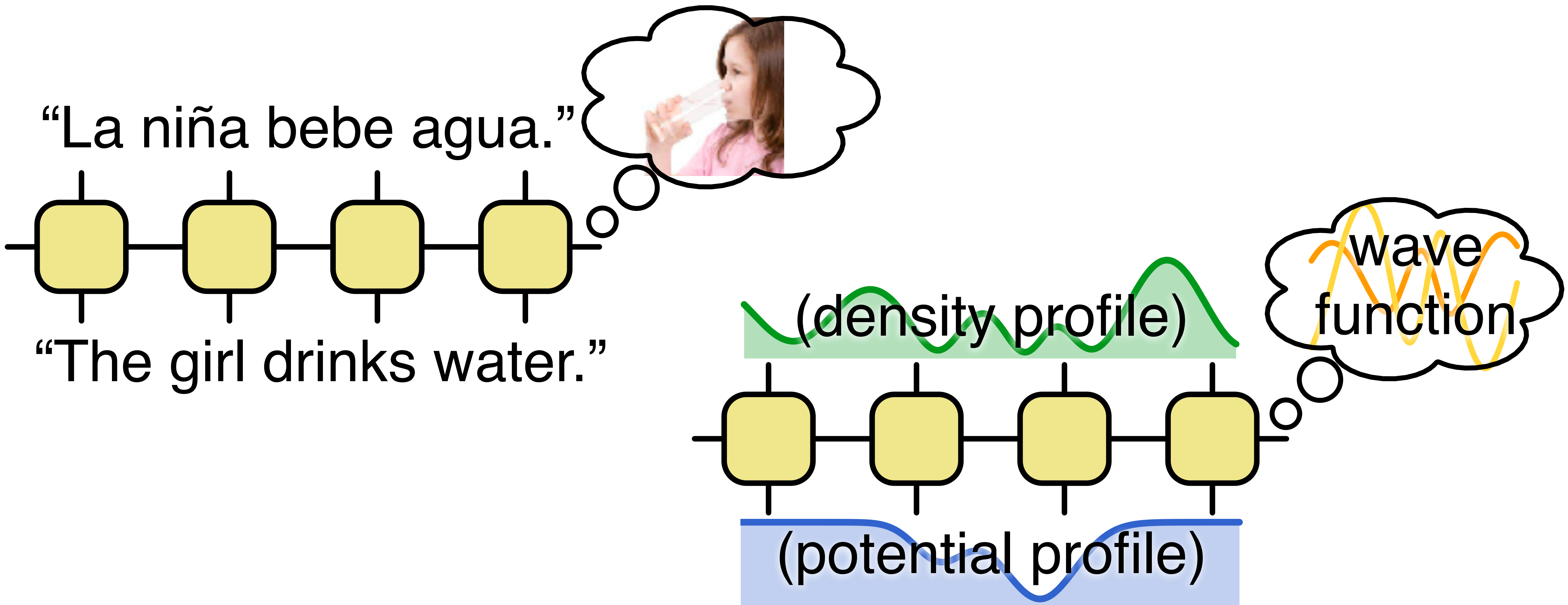
  - Representation learning (word2vec …)

king – man + woman ≈ queen

T Mikolov, SW Yih, G Zweig. NAACL-HLT-2013
also Paul’s talk

Inspiration from Machine Translation

- Motivation: developments in machine translation
  - Train the neural network model to perform a task
  - Discover concepts and relations in representation space

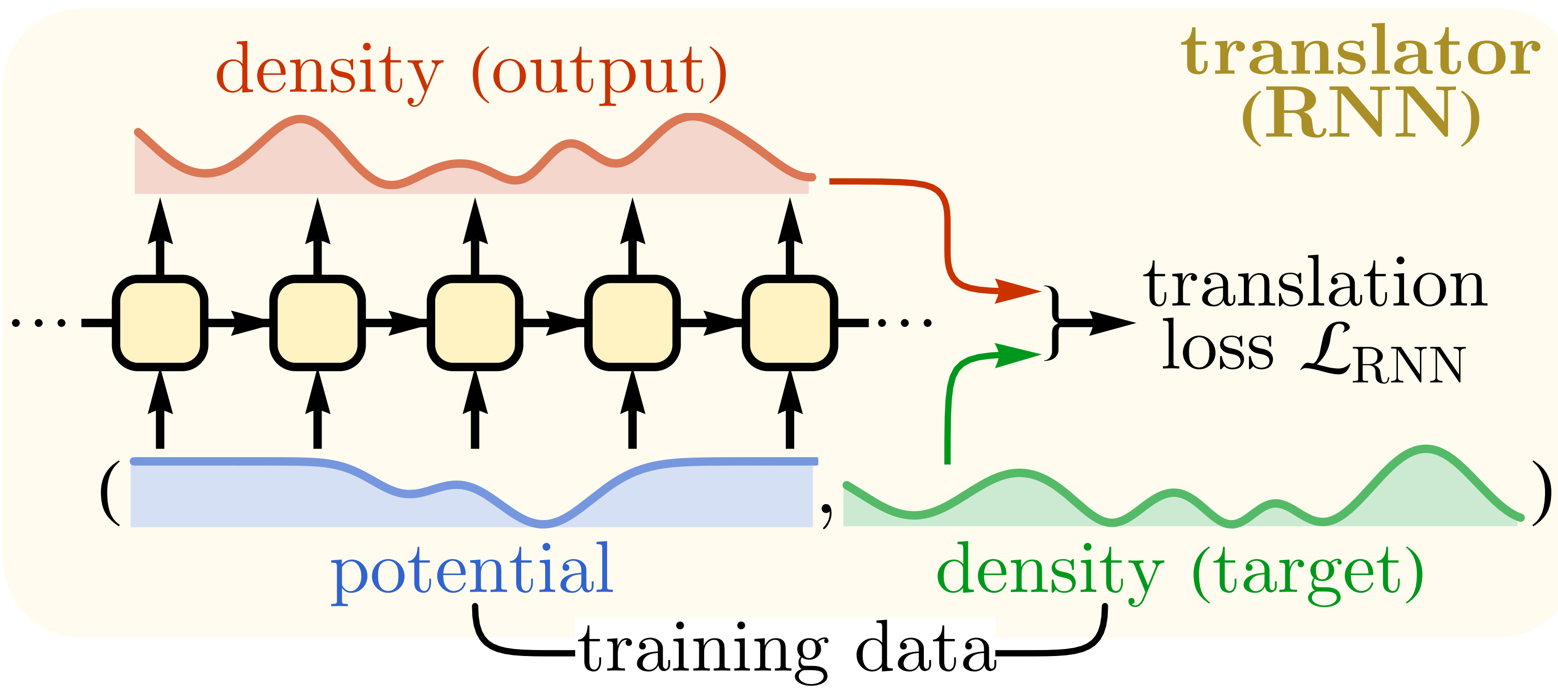


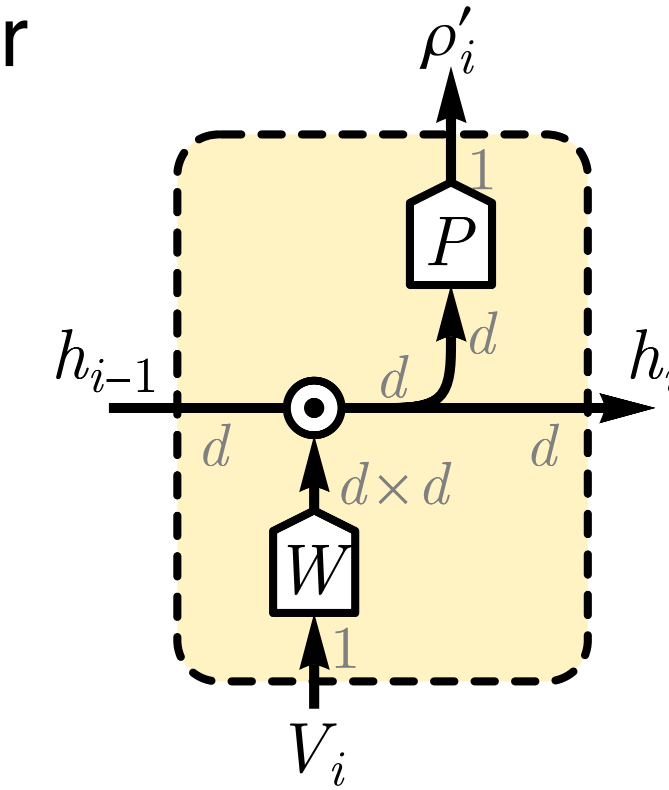
Similar setup but different task:
S Pilati, P Pieri, Scientific
Reports (2019)

- **Task**: potential-to-density mapping
- **Latent variables**: wave function?

Potential-to-Density Translator

- Recurrent neural network (RNN) translator

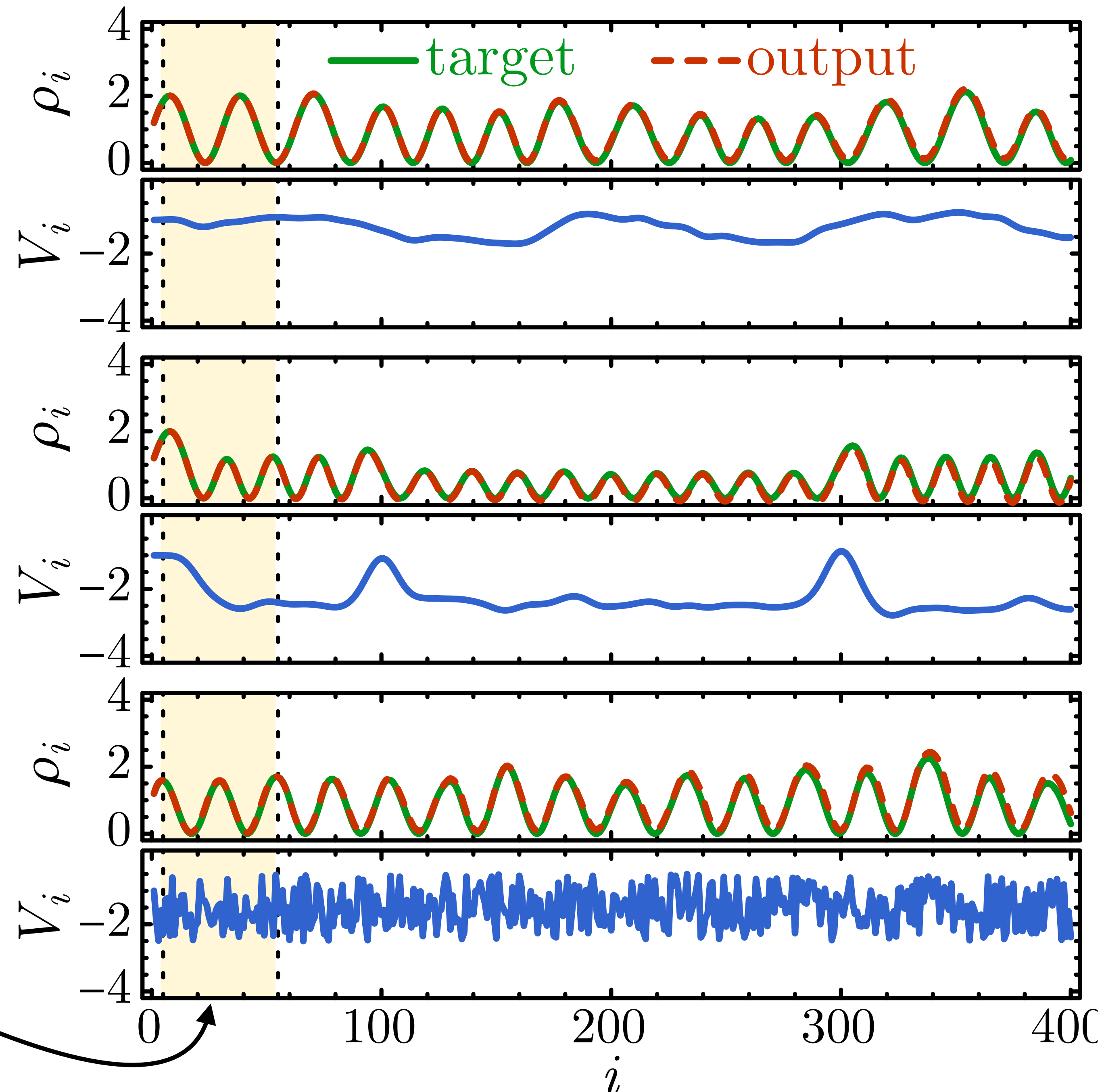




- Discretize the 1D space, collect training data by simulation
- Input: potential sequence V_i
- Update: hidden state $h_i = W(V_i) \cdot h_{i-1}$
- Output: density sequence $\rho'_i = P(h_i)$
- Minimize translation loss $\mathcal{L}_{\mathrm{RNN}} = \sum_{i \in \mathrm{window}} (\rho'_i - \rho_i)^2$

Performance of the Translator

- Performance of the RNN translator



smooth & shallow

smooth & deep

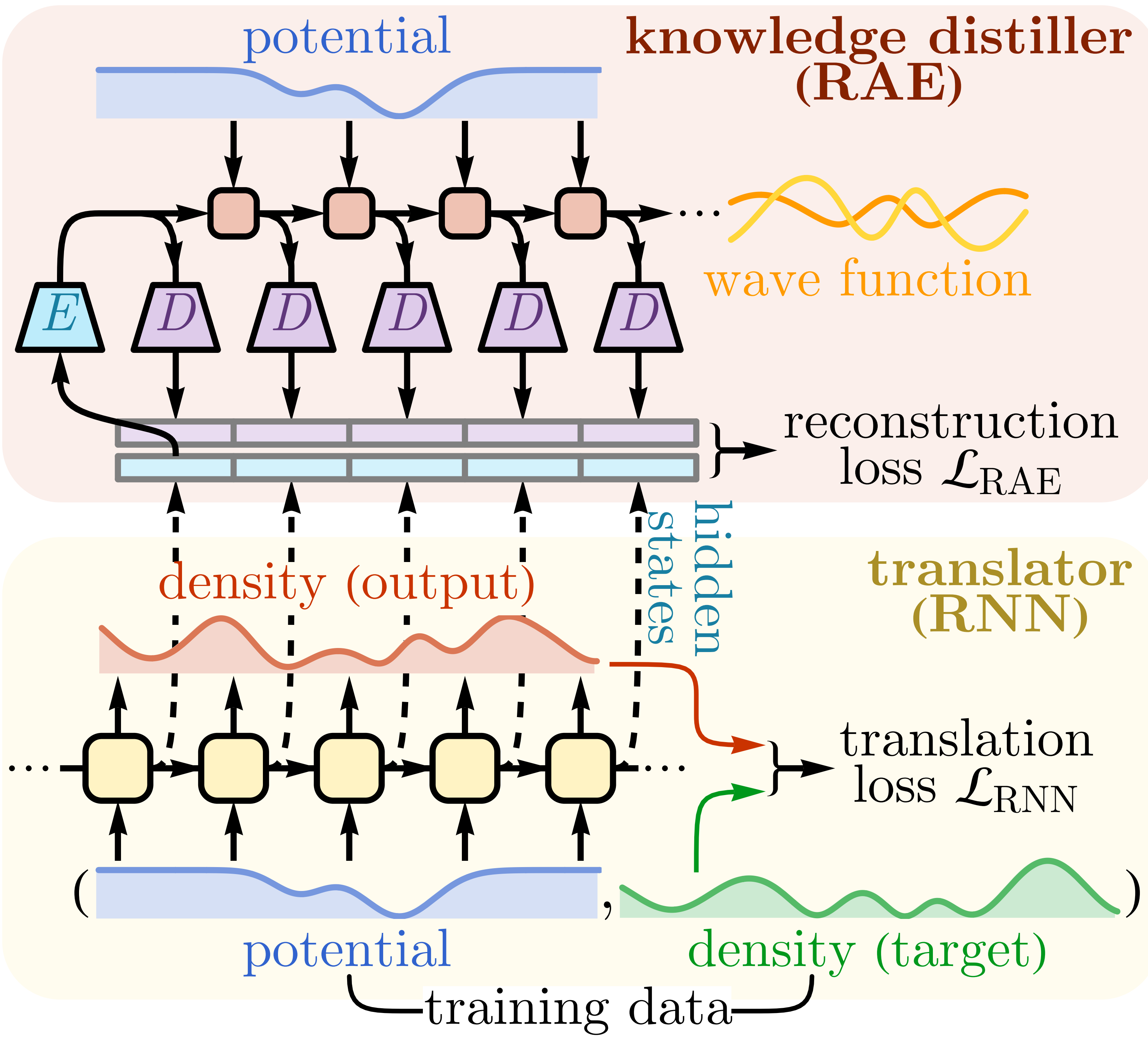
rough

Trained over window of 50

Automatic scalable to larger systems

Introspective Learning

- Introspective Learning

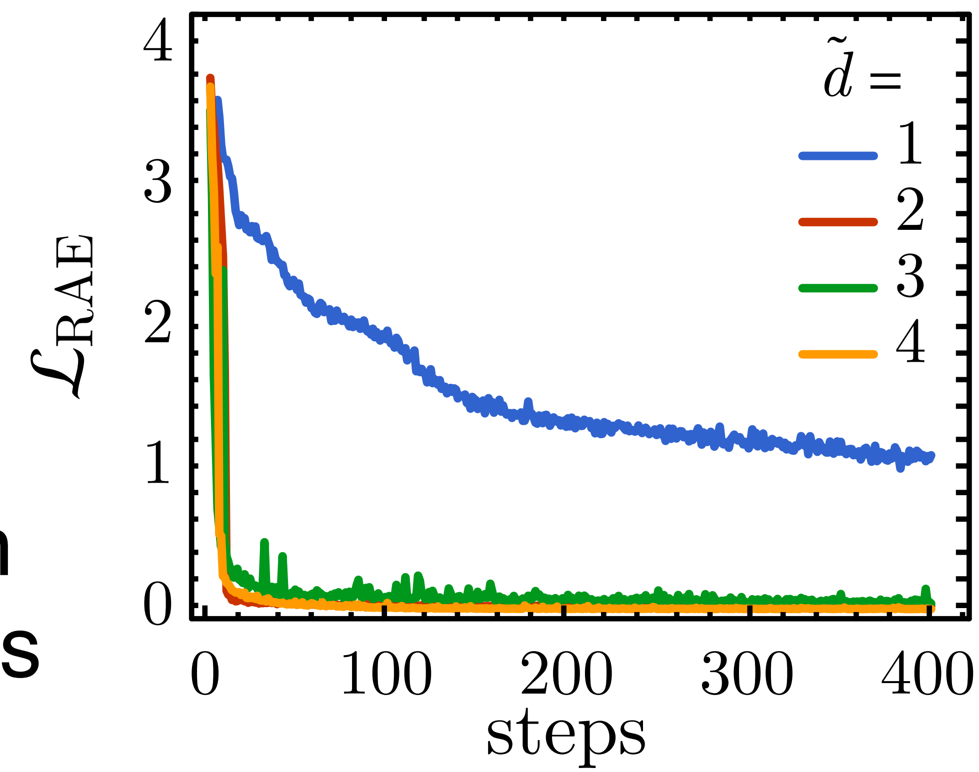


High-level machine only interface with the neural activation of the low-level machine

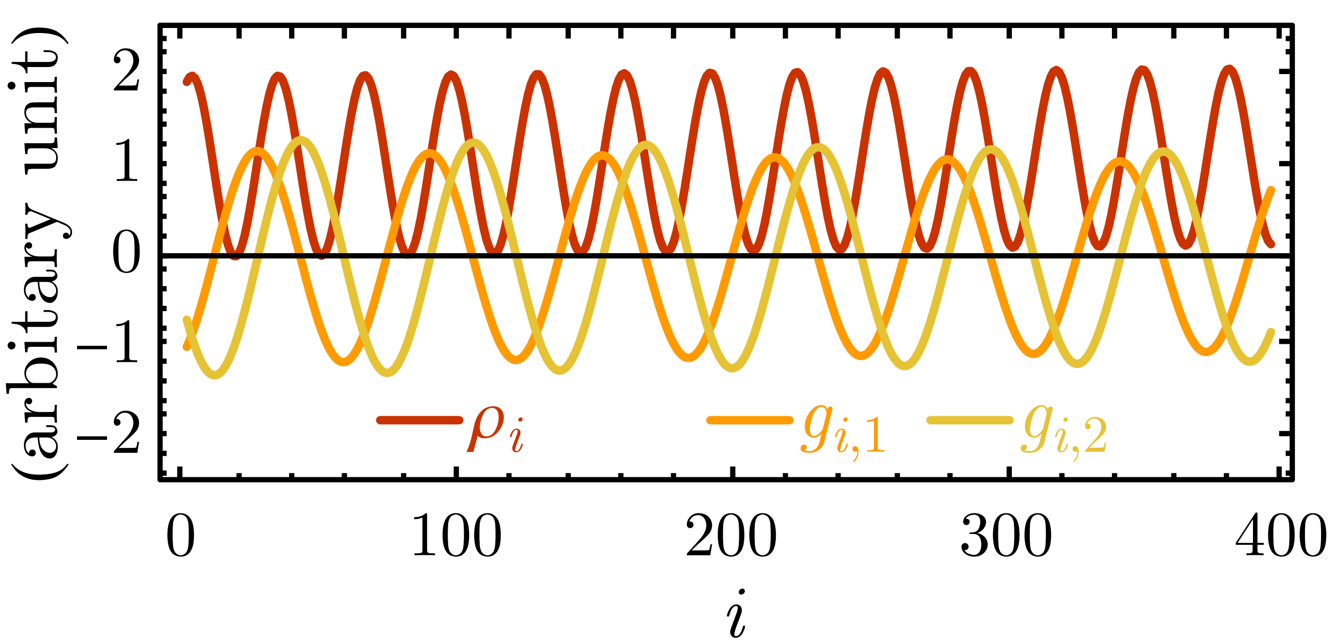
Low-level machine deal with training / experimental data

Emergent Quantum Mechanics

- Imposing information bottleneck
  - Squeezing the latent space dim
  - Monitor the reconstruction loss of the knowledge distiller
  - Abrupt increase of loss only when latent dim < 2 ⇒ two real variables



- Quantum wave function and its 1st order derivative



Update rules

$$\begin{bmatrix} g_{i+1,1} \\ g_{i+1,2} \end{bmatrix} = \begin{bmatrix} 1 & a \\ aV_i & 1 \end{bmatrix} \begin{bmatrix} g_{i,1} \\ g_{i,2} \end{bmatrix}$$

matching Schrödinger Eq.

$$\partial_x^2 \psi(x) = V(x)\psi(x)$$

Alternative Forms of Quantum Mechanics

- If we relax the information bottle neck → alternative forms of quantum machines can also emerge, e.g.

$$\partial_x \begin{bmatrix} \rho(x) \\ \rho'(x) \\ \rho''(x) \end{bmatrix} = \begin{bmatrix} 0 & 2 & 0 \\ V(x) & 0 & 1 \\ 0 & 2V(x) & 0 \end{bmatrix} \begin{bmatrix} \rho(x) \\ \rho'(x) \\ \rho''(x) \end{bmatrix}$$

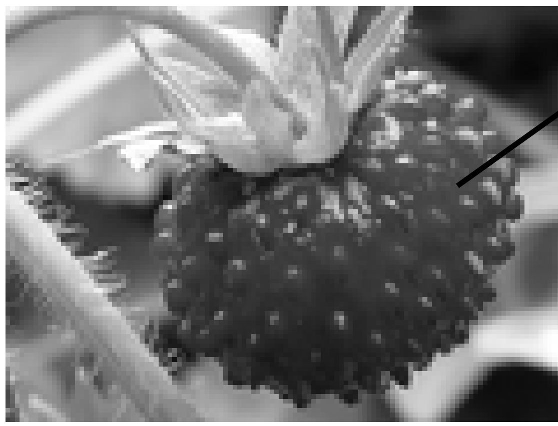
  - Hidden variables: density $\rho(x) = |\psi(x)|^2$ and derivatives
  - But requires at least three real variables

- Wave function + Schrödinger equation formulation of QM is indeed the most parsimonious theory that have emerged in our neural network.

Machine Learning Renormalization Group

Quantum Field Theory as Image Dataset

- A field: a mapping from spacetime to some target manifold

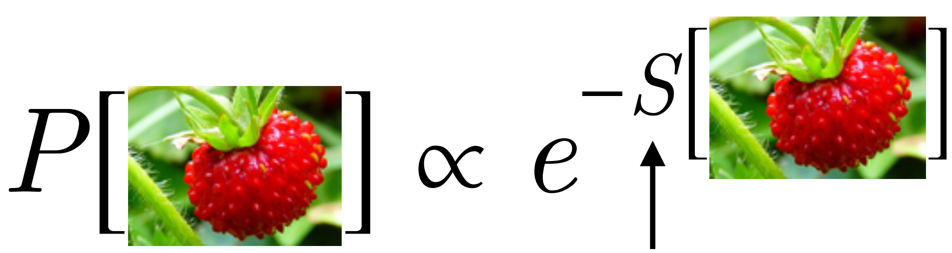


Scalar fields

Vector fields

...

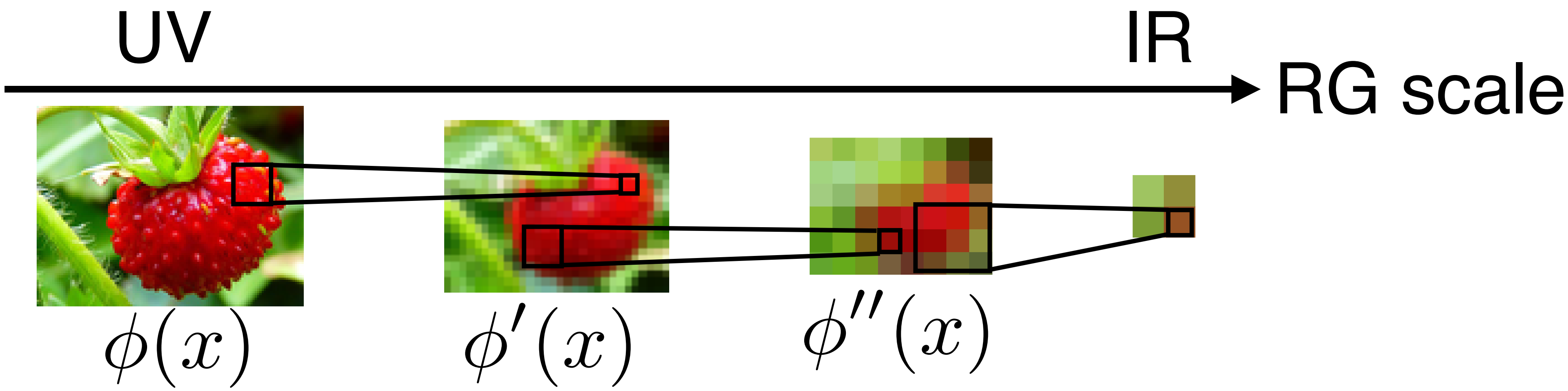
- A quantum field theory (QFT): a model that assigns an action (= negative log likelihood) to every field configuration.



- Can we build a generative model to represent a QFT?

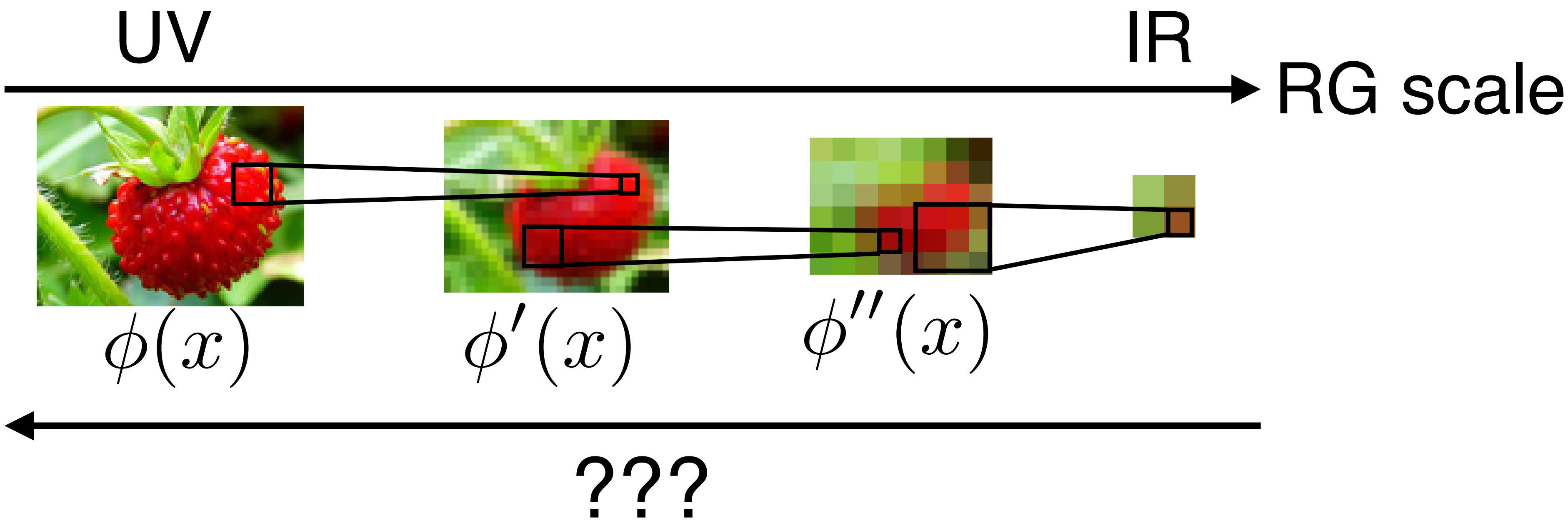
Renormalization Group as Generative Model

- Renormalization "group" (RG): progressively coarse-graining the field (like a convolutional neural network)



Renormalization Group as Generative Model

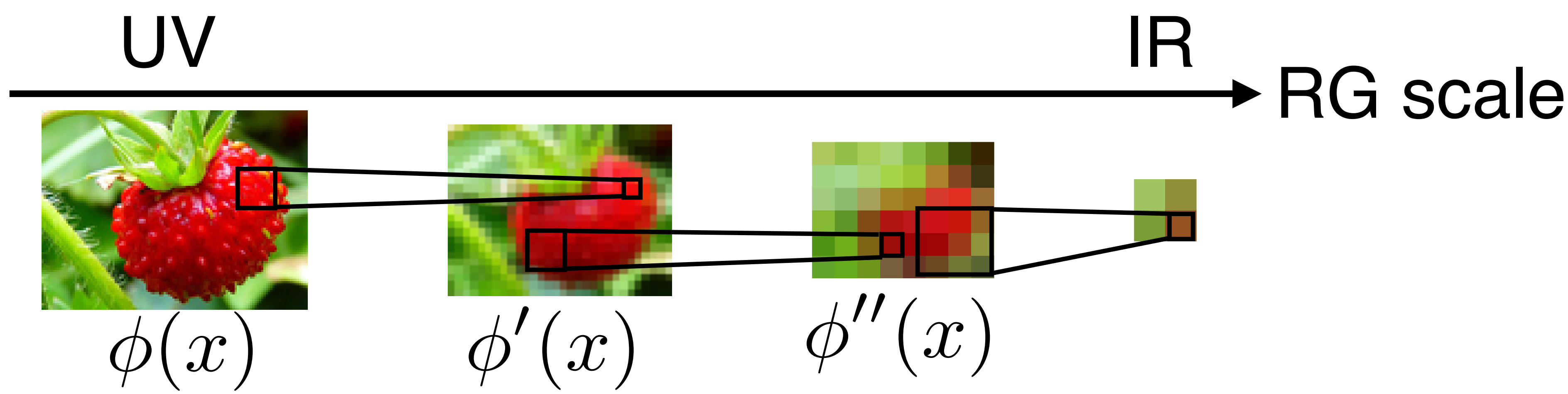
- Renormalization "group" (RG): progressively coarse-graining the field (like a convolutional neural network)



Traditional RG is not invertible…

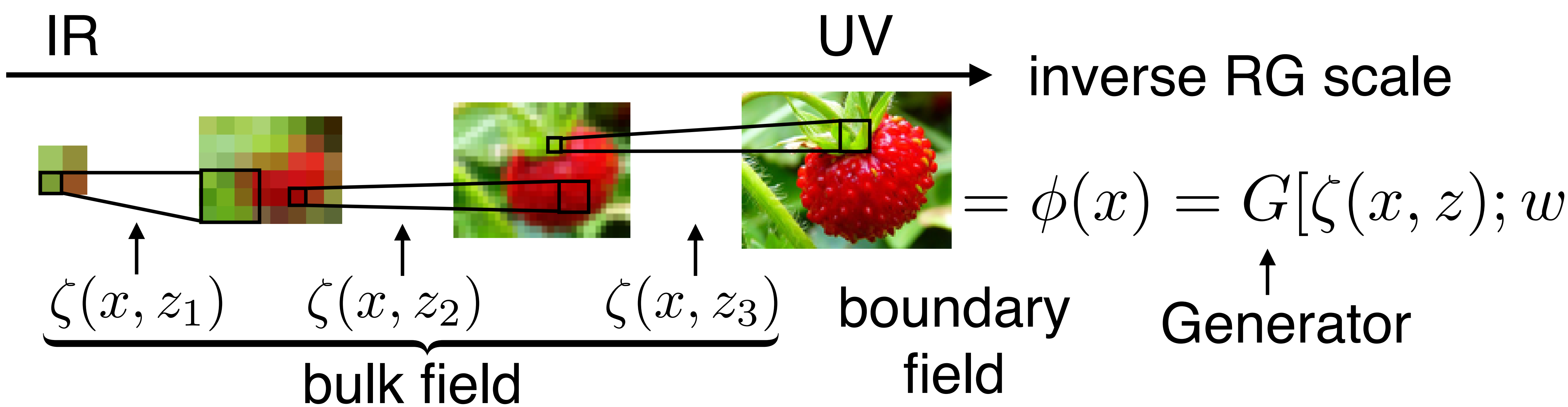
Renormalization Group as Generative Model

- Renormalization "group" (RG): progressively coarse-graining the field, like a convolutional neural network



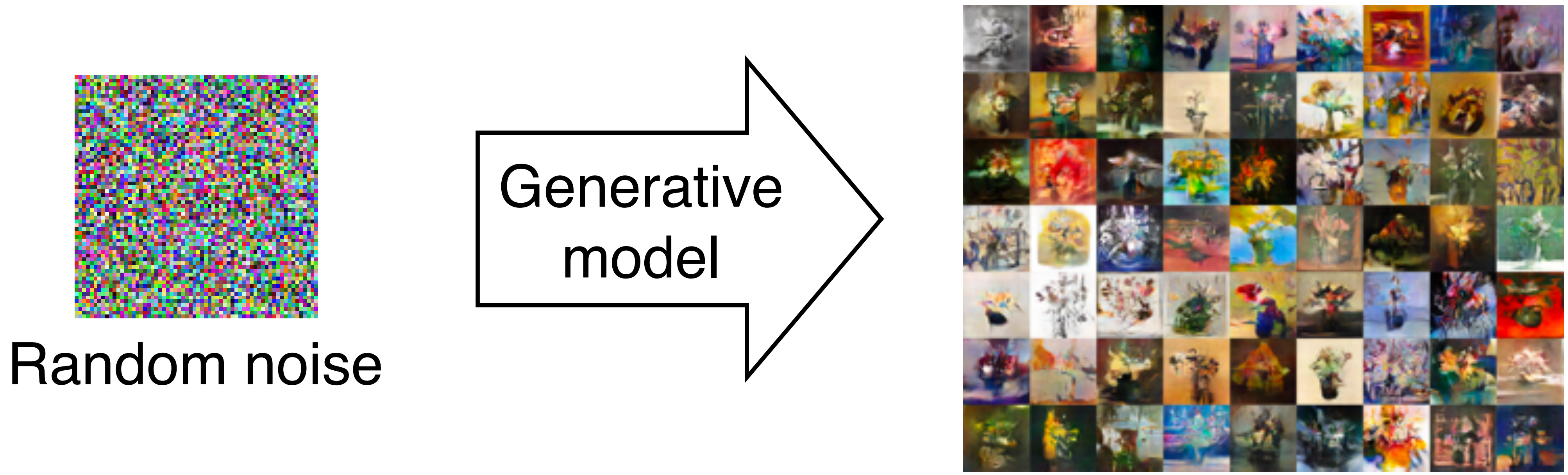
- Inverse RG: a hierarchical generative model

Cédric Bény, NJP (2013)



Flow-Based Generative Models

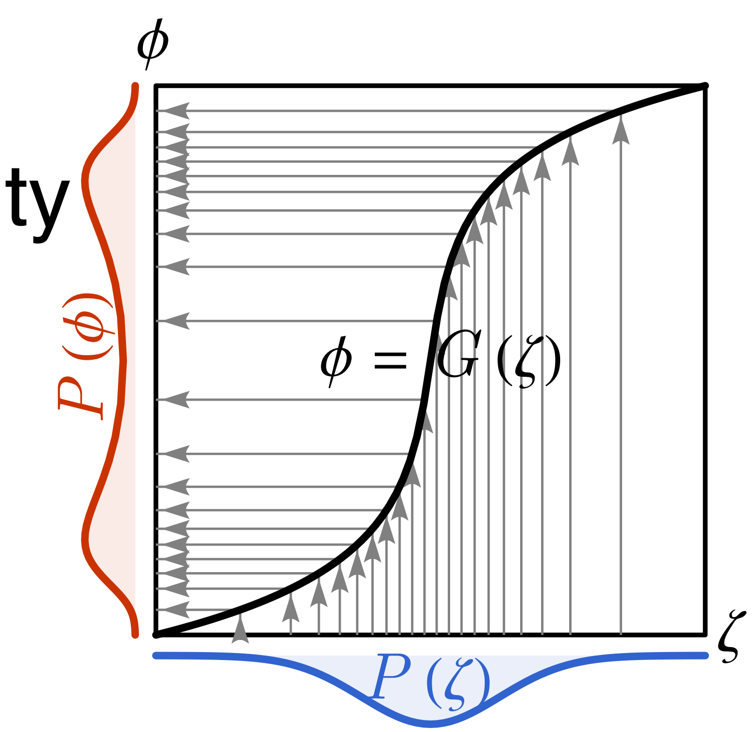
- Flow-based generative model: generate images from noise (latent variables) by an invertible non-linear transformation



- Toy model: a random number generator
  - Generative model deforms the probability distribution, sample ζ to generate ϕ

$$\phi = G(\zeta)$$

$$P(\phi) = P(\zeta)\left(\frac{\partial G(\zeta)}{\partial \zeta}\right)^{-1}$$



Flow-Based Generative Models

- What are the advantages of flow-based models compared to energy-based models (e.g. Boltzmann machines)?
  - Differentiable log likelihood allows gradient to propagate through probability to train the model.

$$\mathcal{L} = \mathrm{KL}(P_{\text{post}} \| P_{\text{target}}) \quad P_{\text{post}}(\phi) = P_{\text{prior}}(\zeta) \left(\frac{\partial G(\zeta)}{\partial \zeta} \right)^{-1}$$

  - Direct sampling generates new samples efficiently

$$\phi = G(\zeta)$$

  - Bijectivity allows inference of latent encoding

$$\zeta = G^{-1}(\phi)$$

- Generative models with tractable likelihood
  - Flow-based: Zhang, E, Wang (2018)
  - Autoregressive: Wu, Wang, Zhang, PRL(2019), Sharir et.al. (2019)
  - Tensor networks: Han et.al. PRX(2018)

Neural Network Renormalization Group

- Generative model deforms noise to QFT

Li, Wang, PRL (2018)
Hu, Li, Wang, You (2019)

Posterior (boundary) ↓ Prior (bulk) ↓ Generator ↓

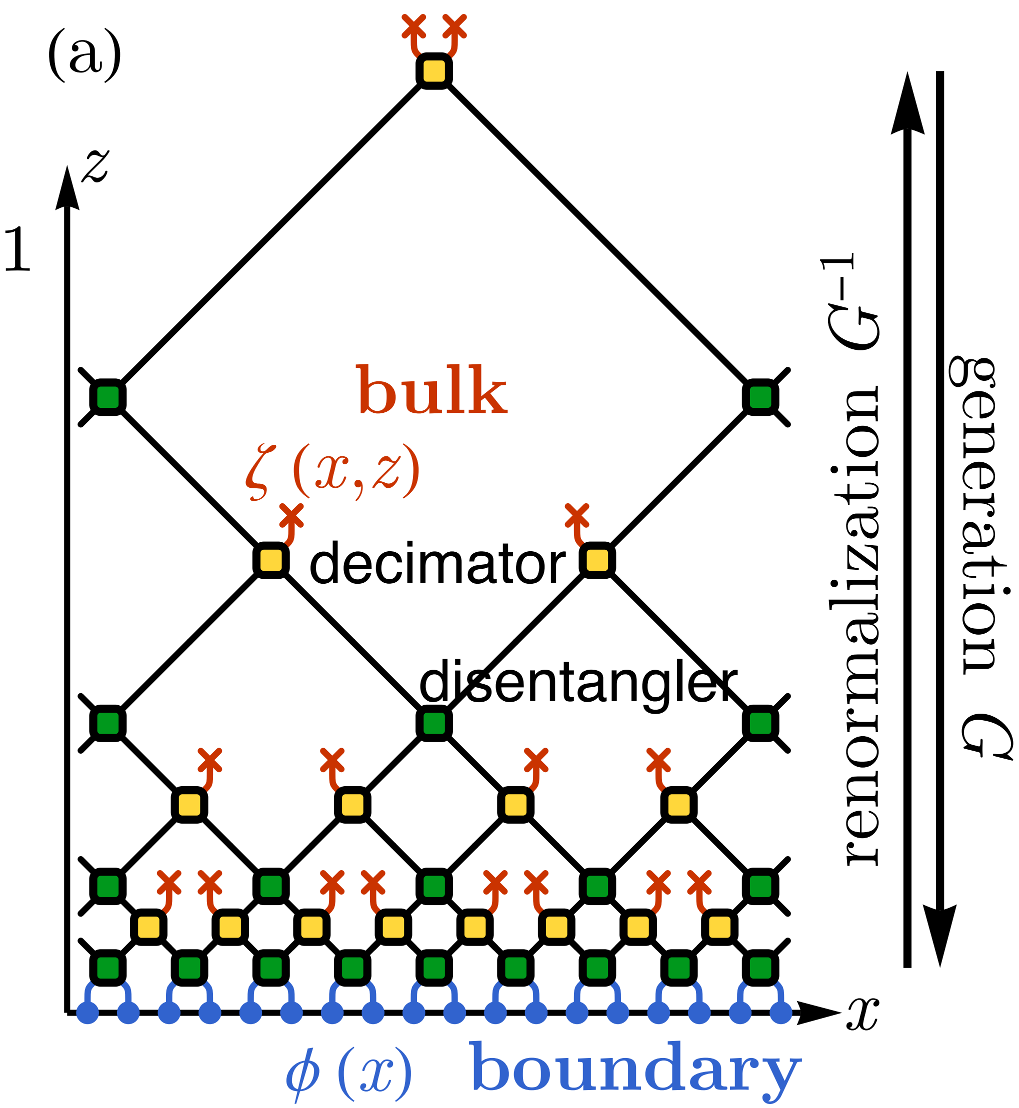
$$P_{\text{post}}[\phi] = P_{\text{prior}}[\zeta] \det\left(\frac{\delta G[\zeta]}{\delta \zeta}\right)^{-1}$$

Model distribution

↕ Minimize KL divergence

Target distribution (QFT)

$$P_{\text{target}}[\phi] = e^{-S_{\text{QFT}}[\phi]}/Z_{\text{QFT}}$$

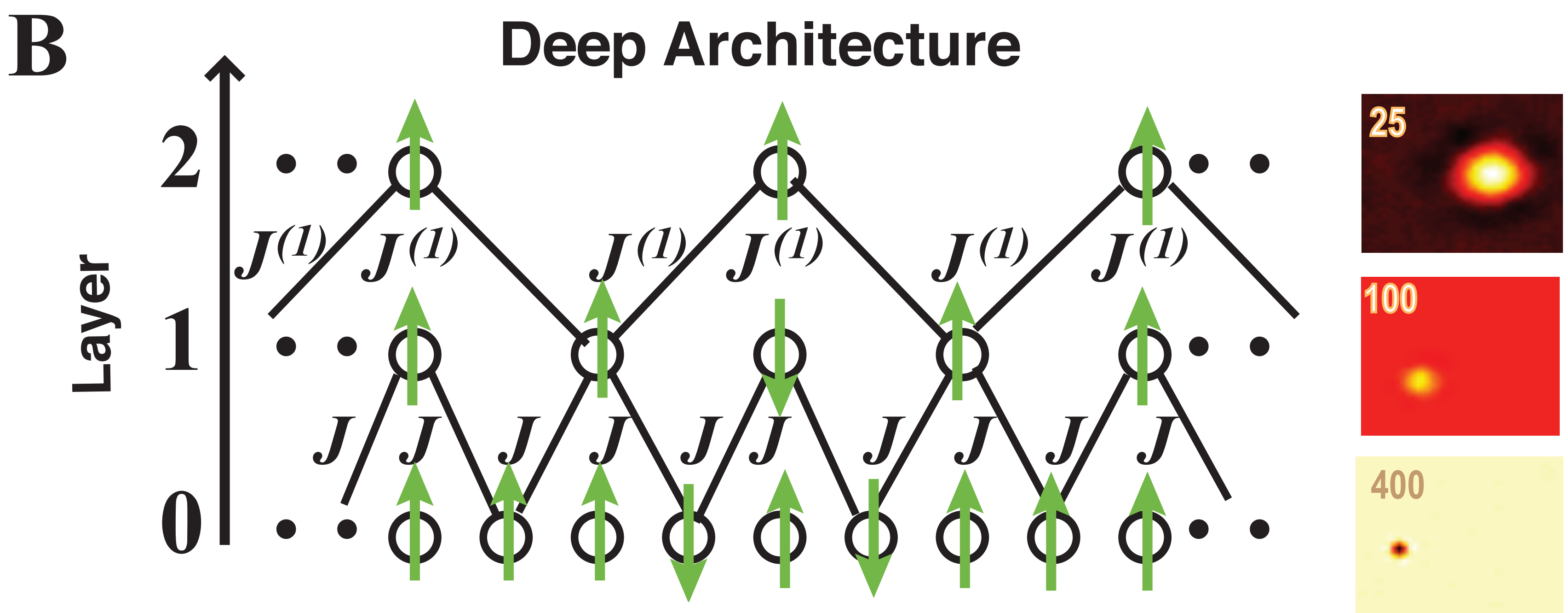


MERA network - Vidal (2006)

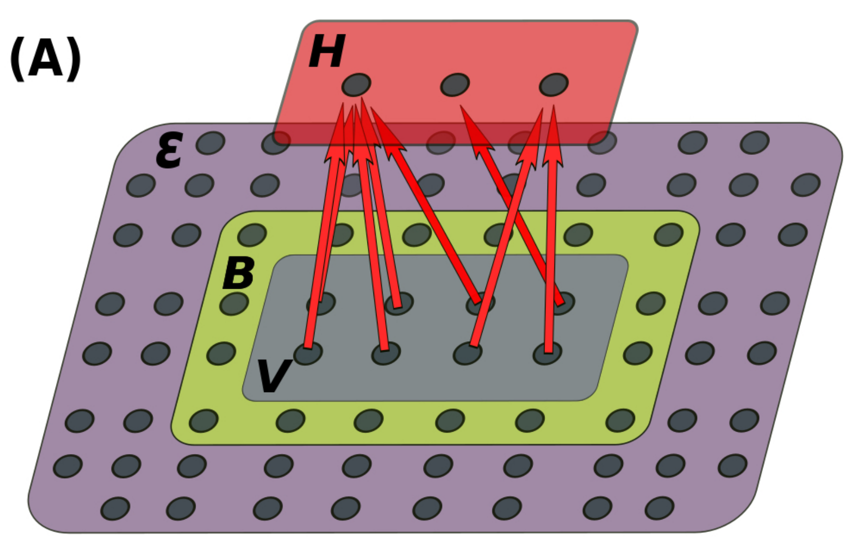
- How to choose the prior distribution?
Our choice: independent Gaussian $P_{\text{prior}}[\zeta] \propto e^{-\|\zeta\|^2}$

Information Theoretic Goal of RG

- Renormalization Group = Deep Learning? Mehta, Schwab (2014)



- Maximal Real-Space Mutual Information (maxRMI) principle



Koch-Janusz, Ringel (2017)

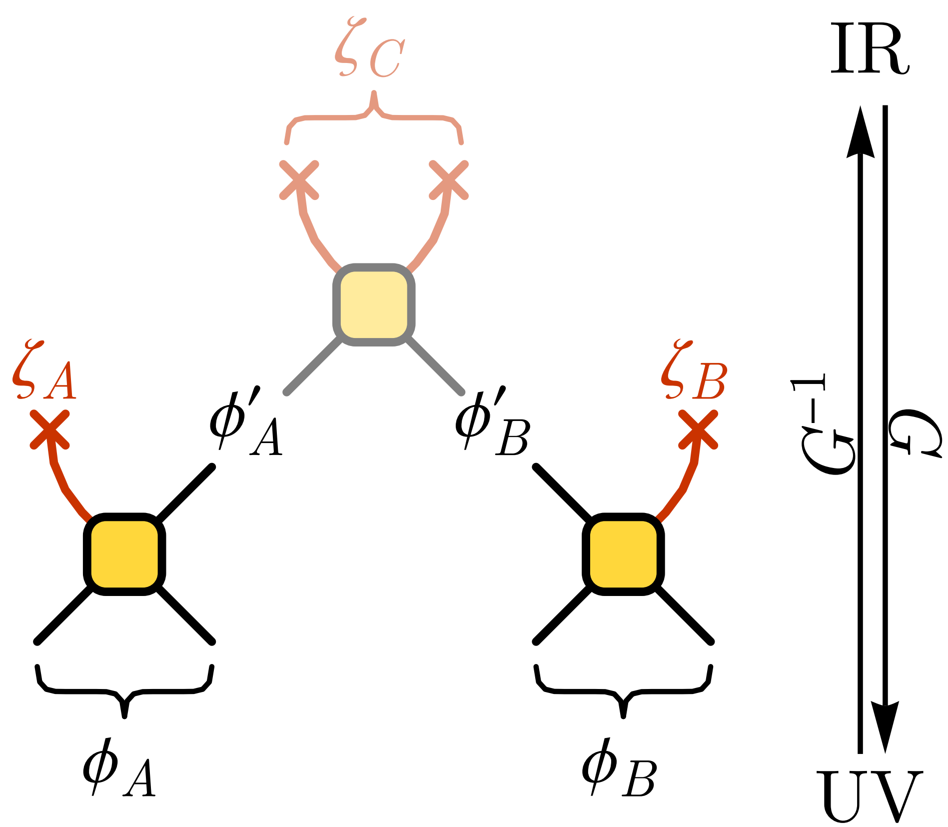
$$\max I(H, \mathcal{E})$$

Coarse-grained freedom → H

Environment → $\mathcal{E}$

Information Theoretic Goal of RG

- Minimal Bulk Mutual Information (minBMI) principle



- maxRMI: $\max I(\phi'_A : \phi_B)$.
- minBMI: $\min I(\zeta_A : \zeta_B)$

Two objectives are related

$$I(\phi'_A : \phi_B) + I(\zeta_A : \zeta_B)$$
$$= I(\phi_A, \phi_B) = \text{const.}$$

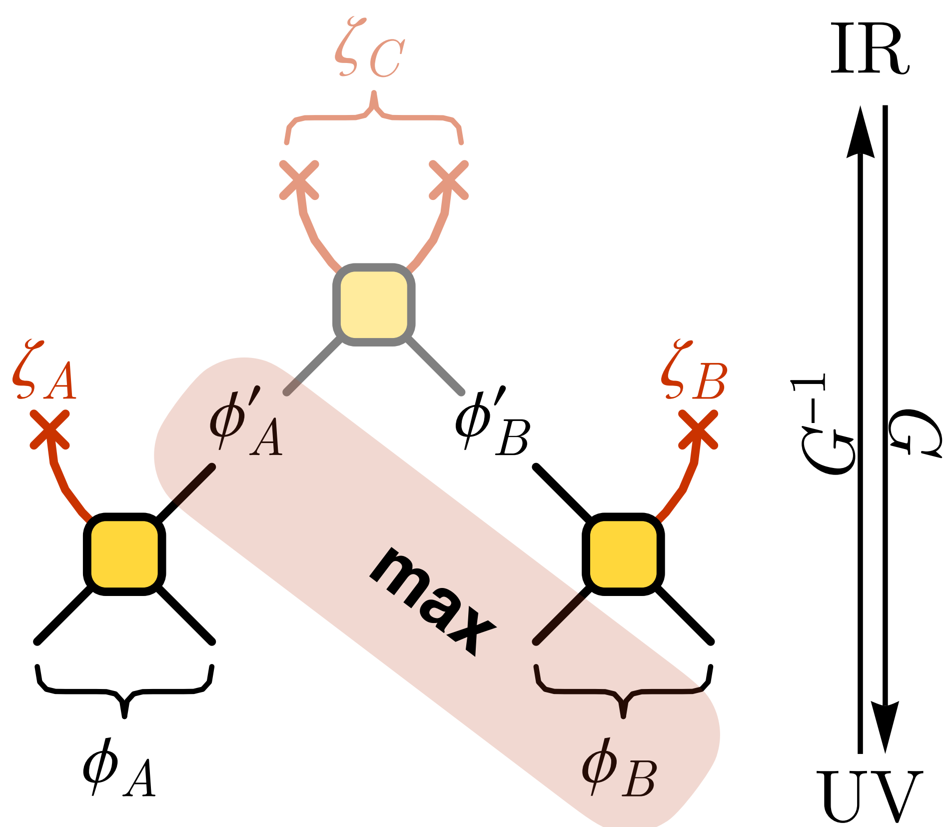
Hu, Li, Wang, You (2109)

- The objectives are two-folded
  - RG: disentangle the QFT in the bulk $P_{\text{prior}}[\zeta] \propto e^{-\|\zeta\|^2}$
  - Inverse RG: reproduce the QFT on the boundary

$$\min \mathsf{KL}(P_{\text{post}}[\phi] || e^{-S_{\text{QFT}}[\phi]})$$

Information Theoretic Goal of RG

- Minimal Bulk Mutual Information (minBMI) principle



- maxRMI: $\max I(\phi'_A : \phi_B)$.
- minBMI: $\min I(\zeta_A : \zeta_B)$

Two objectives are related

$$I(\phi'_A : \phi_B) + I(\zeta_A : \zeta_B)$$
$$= I(\phi_A, \phi_B) = \text{const.}$$

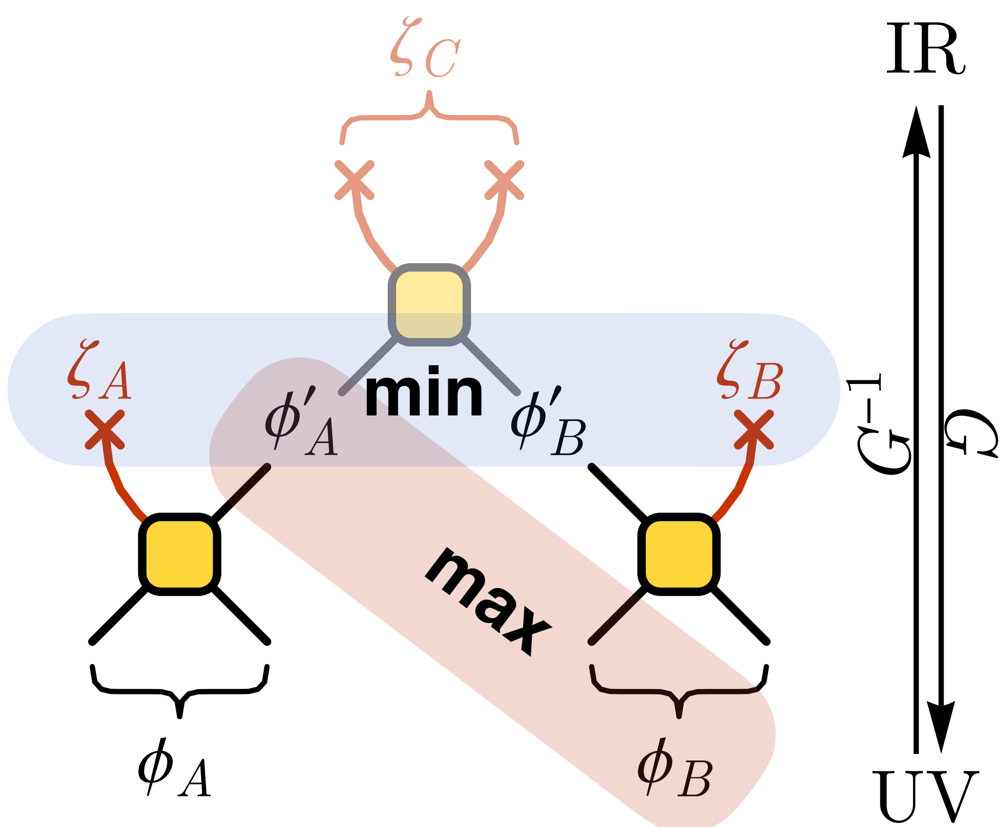
Hu, Li, Wang, You (2109)

- The objectives are two-folded
  - RG: disentangle the QFT in the bulk $P_{\text{prior}}[\zeta] \propto e^{-\|\zeta\|^2}$
  - Inverse RG: reproduce the QFT on the boundary

$$\min \mathsf{KL}(P_{\text{post}}[\phi] || e^{-S_{\text{QFT}}[\phi]})$$

Information Theoretic Goal of RG

- Minimal Bulk Mutual Information (minBMI) principle



- maxRMI: $\max I(\phi'_A : \phi_B)$.
- minBMI: $\min I(\zeta_A : \zeta_B)$

Two objectives are related

$$I(\phi'_A : \phi_B) + I(\zeta_A : \zeta_B)$$
$$= I(\phi_A, \phi_B) = \text{const.}$$

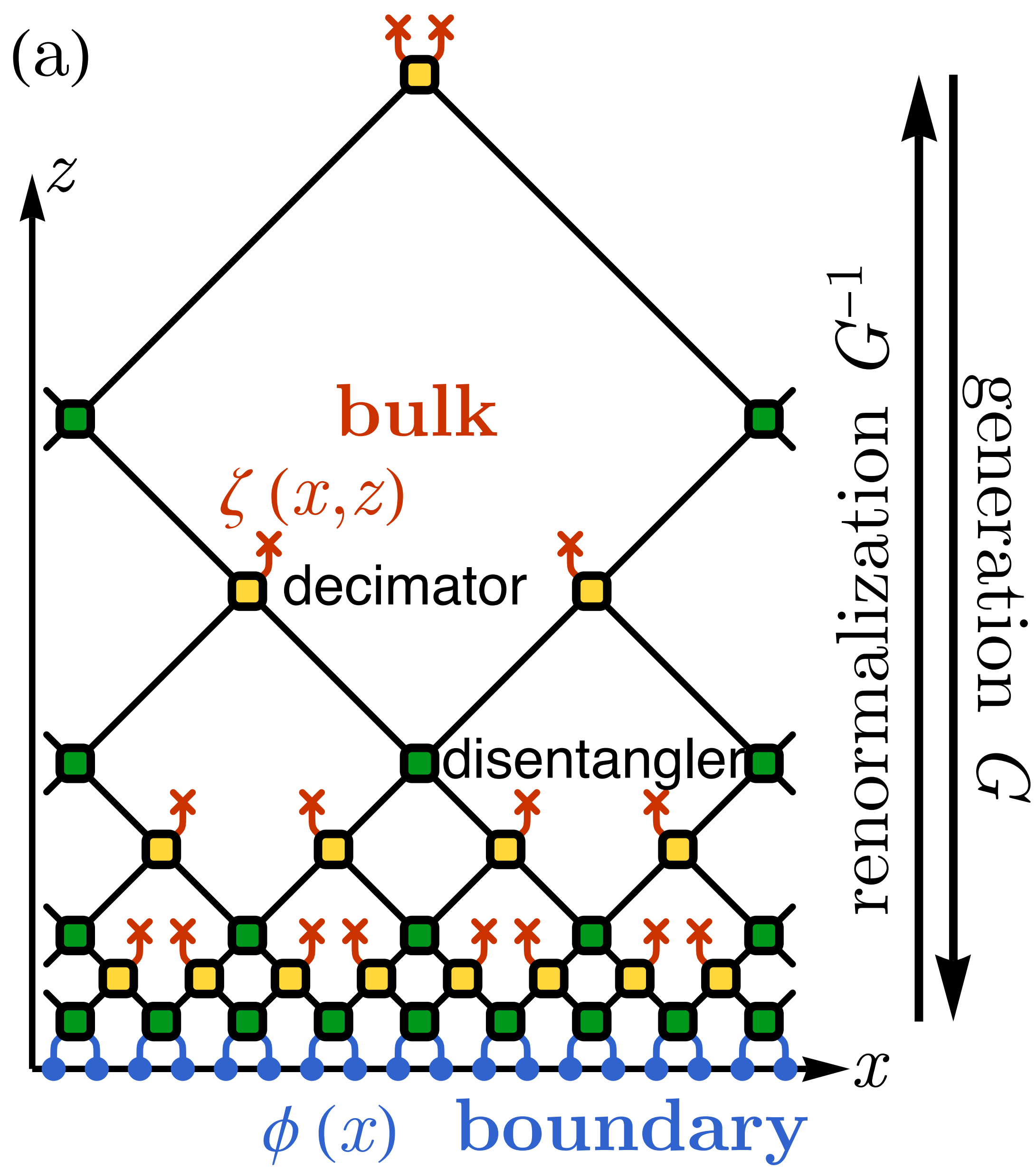
Hu, Li, Wang, You (2109)

- The objectives are two-folded
  - RG: disentangle the QFT in the bulk $P_{\text{prior}}[\zeta] \propto e^{-\|\zeta\|^2}$
  - Inverse RG: reproduce the QFT on the boundary

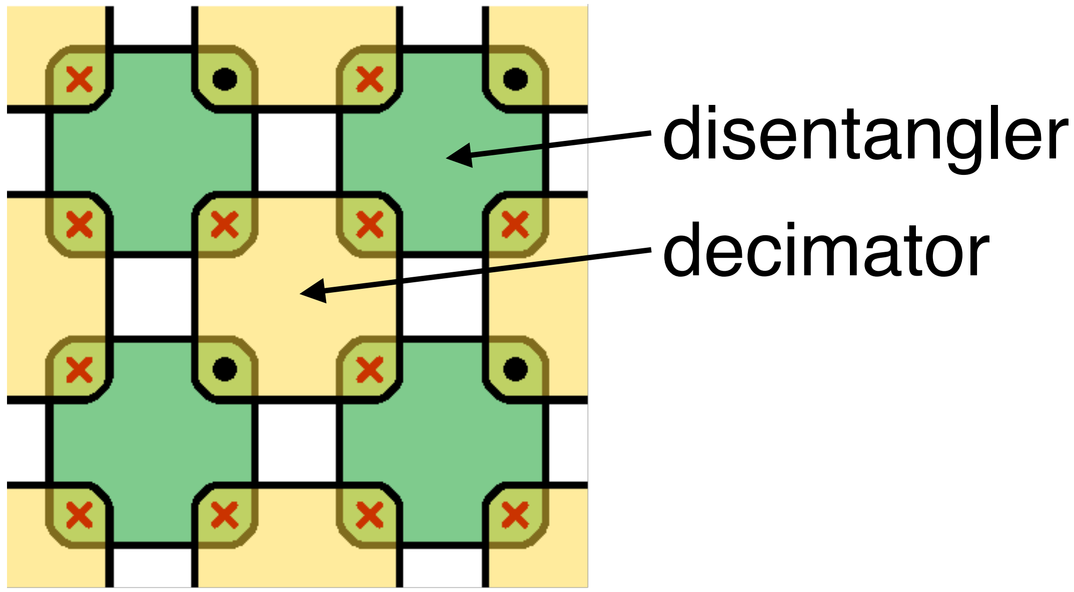
$$\min \mathsf{KL}(P_{\text{post}}[\phi] || e^{-S_{\text{QFT}}[\phi]})$$

Neural Network Architecture

- Flow-based hierarchical generative model



(side view of the network)

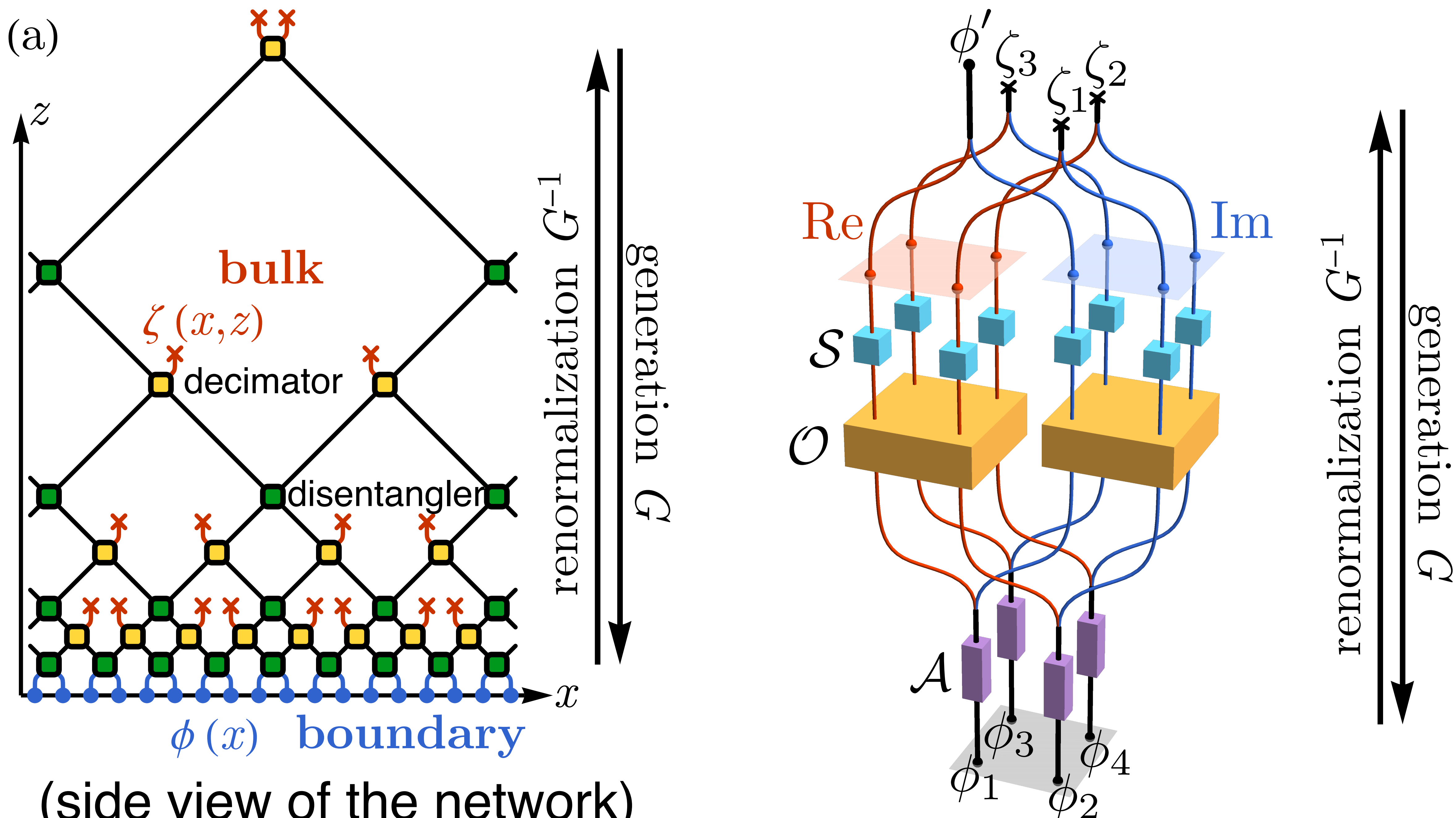


(top view of one layer)

- Both disentangler and decimator are realized as a $\mathbb{C}^4 \leftrightarrow \mathbb{C}^4$ bijector

Neural Network Architecture

- Flow-based hierarchical generative model



(side view of the network)

Training Scheme

- Objective function

$$\mathcal{L} = \mathsf{KL}(P_{\text{post}}[\phi] \| P_{\text{target}}[\phi])$$

$$= \mathop{\mathbb{E}}_{\zeta \sim P_{\text{prior}}} S_{\text{QFT}}[G[\zeta]] + \ln P_{\text{prior}}[\zeta] - \ln \det \left(\frac{\delta G[\zeta]}{\delta \zeta} \right)$$

  - Sample ζ from bulk, push to the boundary $\phi = G[\zeta]$
  - Forward: evaluate loss function
  - Backward: propagate gradient to train bijectors

$$\frac{\mathrm{d}G}{\mathrm{d}t} = -\frac{\delta \mathcal{L}}{\delta G}$$

- Machine learns the optimal RG (optimal EHM) G from the QFT action S_{QFT}

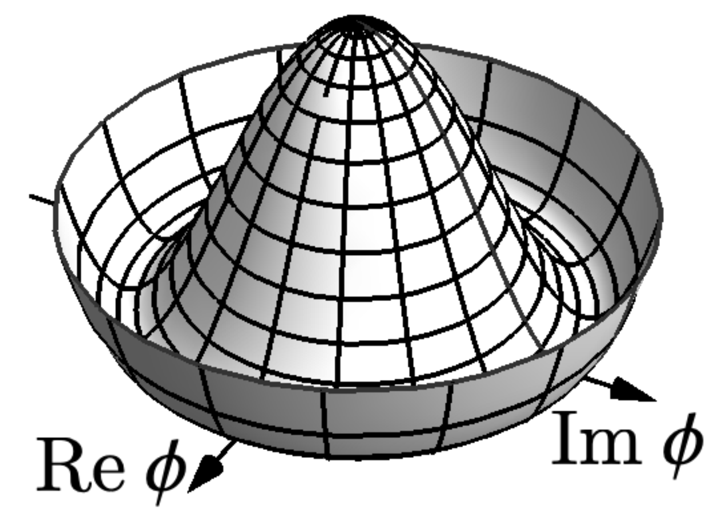
Complex ϕ^4 Model in 2D

- Lattice field theory on square lattice

$$S[\phi] = -t \sum_{\langle ij \rangle} \phi_i^* \phi_j + \sum_i \left(\mu |\phi_i|^2 + \lambda |\phi_i|^4 \right)$$

- Effectively 2D XY model $\phi_i = \sqrt{\rho} e^{\mathrm{i}\theta_i}$

$$S[\theta] = -\frac{1}{T} \sum_{\langle ij \rangle} \cos(\theta_i - \theta_j)$$



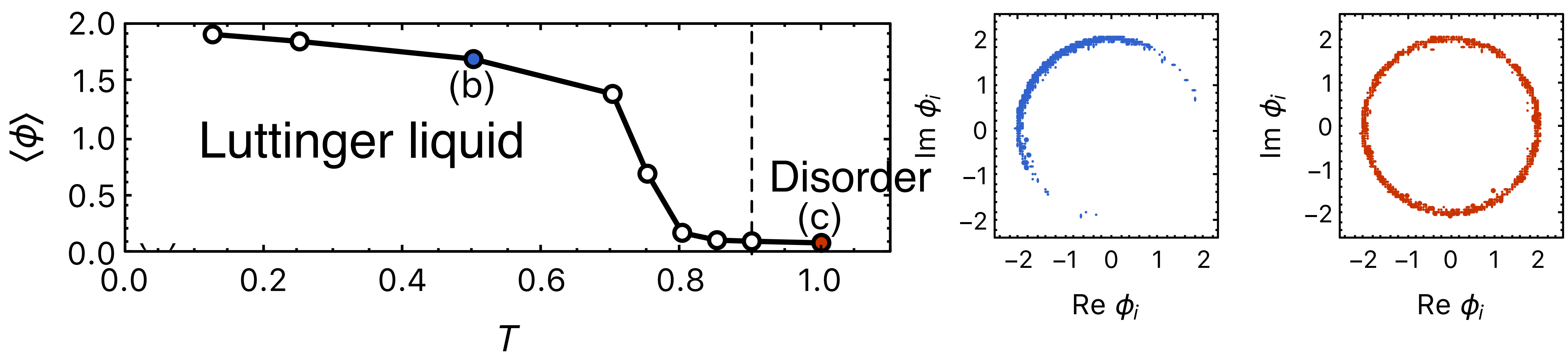
$\langle \phi_i^* \phi_j \rangle \sim r_{ij}^{\alpha}$ $\langle \phi_i^* \phi_j \rangle \sim e^{-r_{ij}/\xi}$

T (effectively tuned by t)

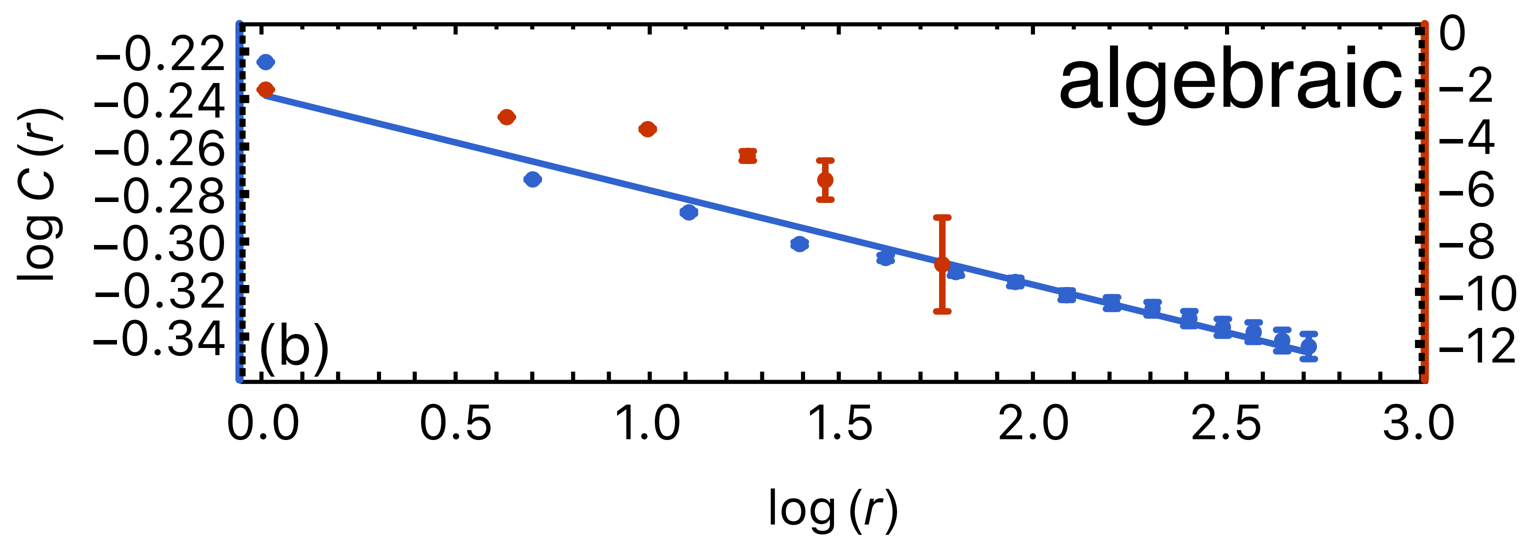
Luttinger liquid (CFT) — T_{KT} — Disordered

Performance of the Generative Model

- Let us first make sure that the machine learns the correct physics from the given action.
  - Phase diagram (32x32 finite size lattice)

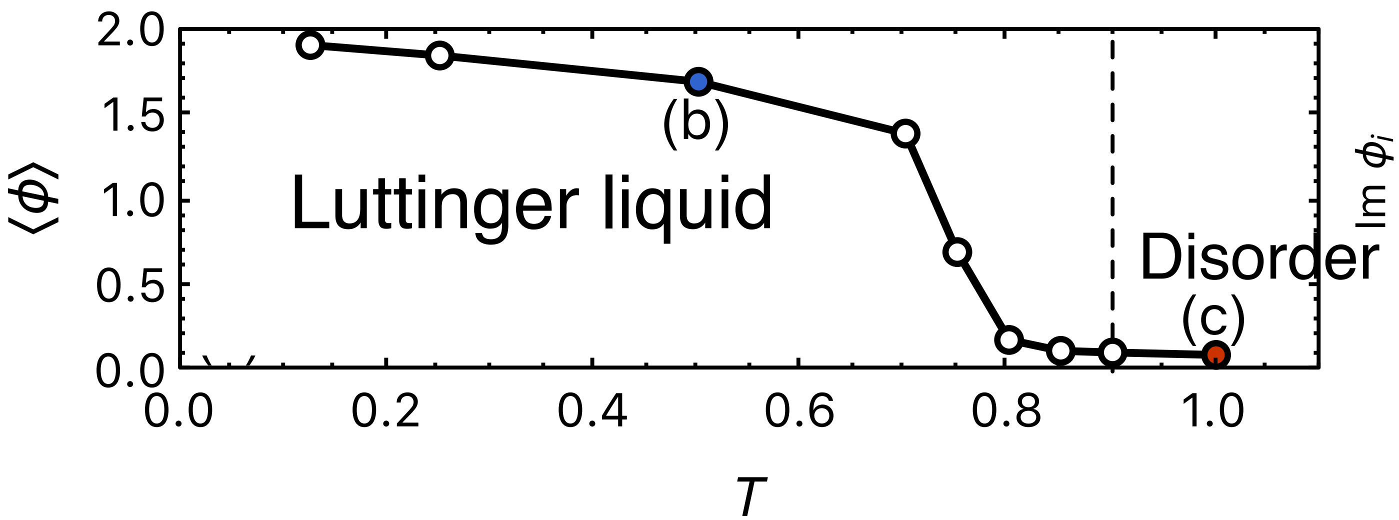


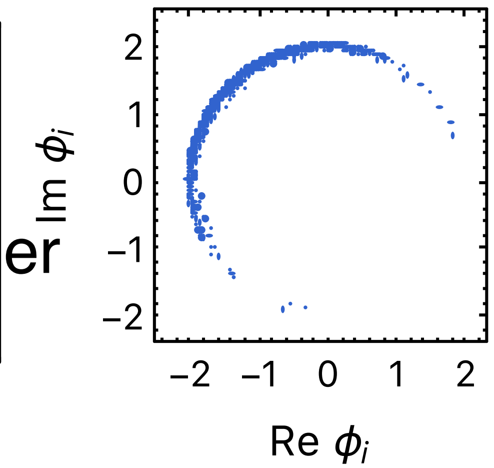
  - Correlation function

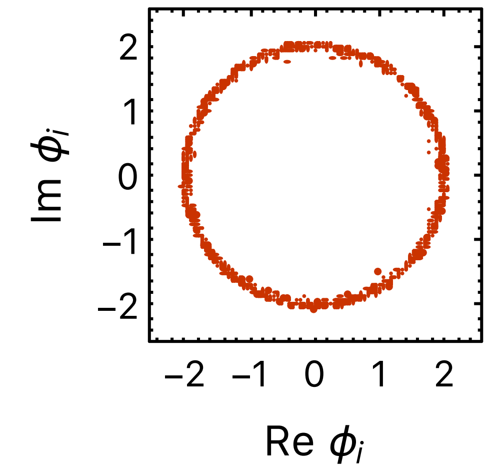


Performance of the Generative Model

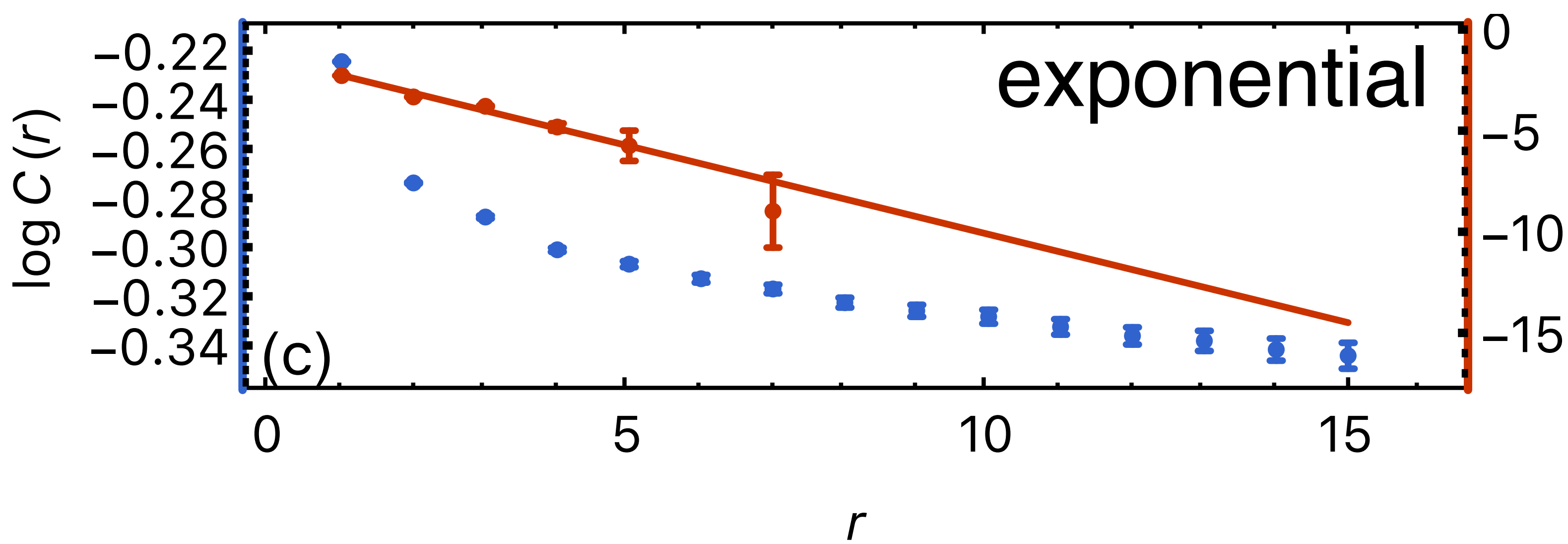
- Let us first make sure that the machine learns the correct physics from the given action.
  - Phase diagram (32x32 finite size lattice)







  - Correlation function

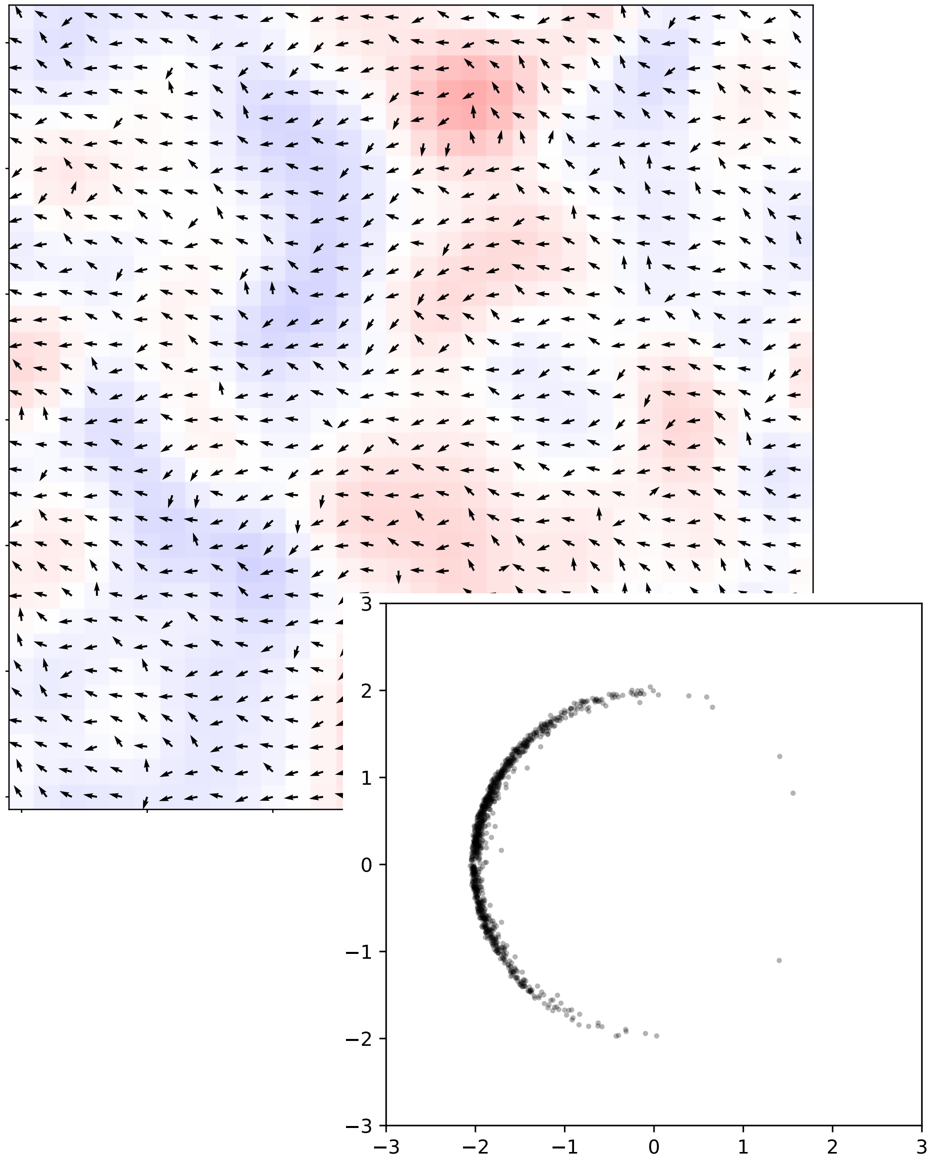


Efficient Sampling from the Bulk

- Sampling: holographic mapping from bulk to boundary
  - Massive field in the bulk → Critical field on the boundary
  - Local update in the bulk → Global update on the boundary

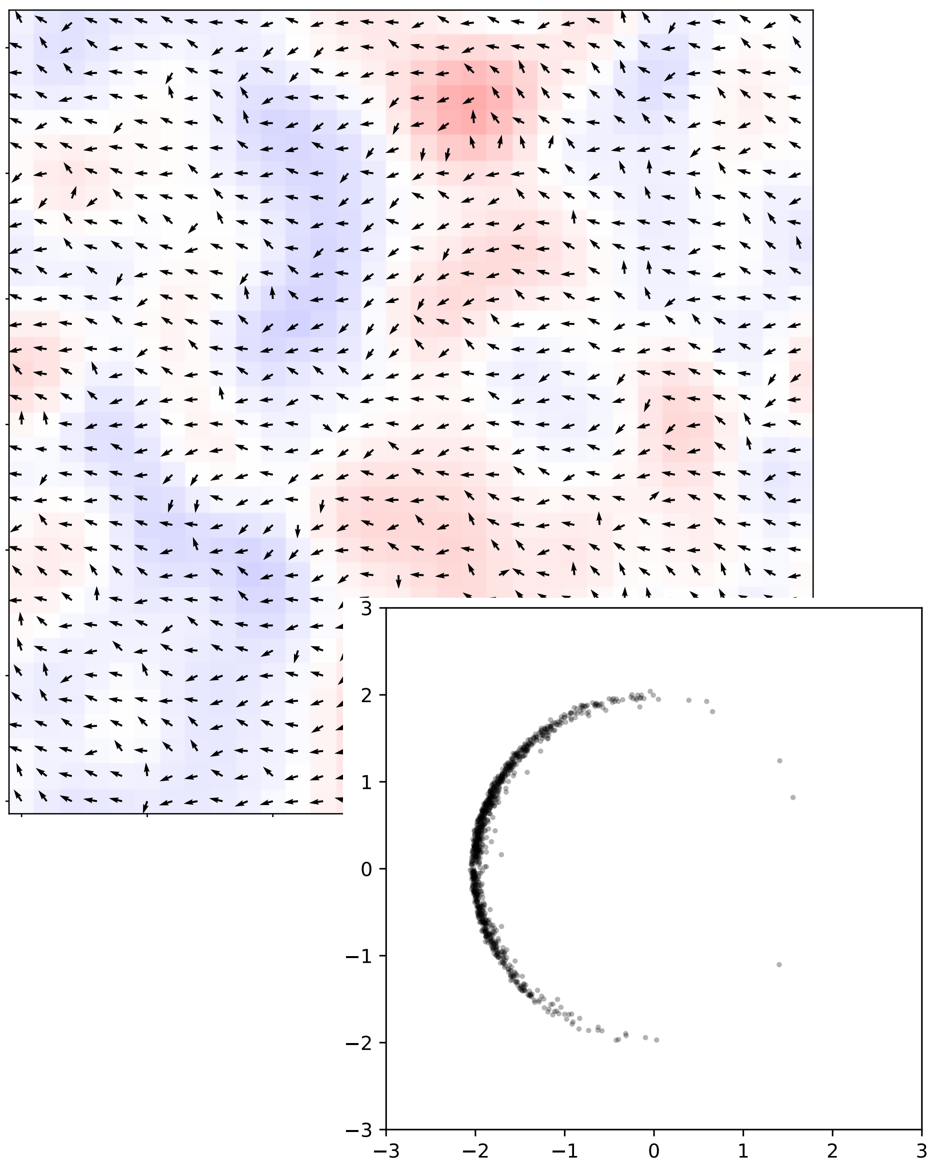
Efficient Sampling from the Bulk

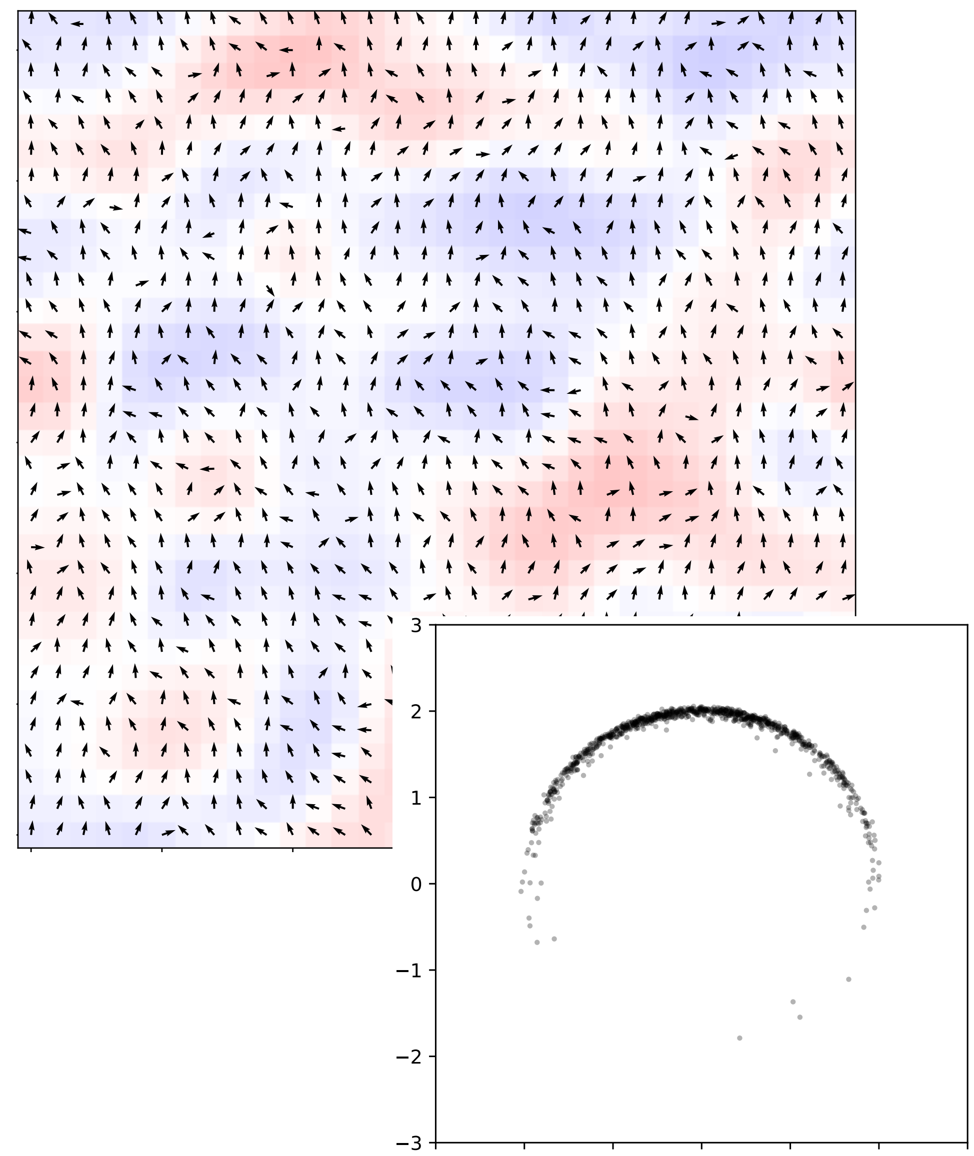
- Sampling: holographic mapping from bulk to boundary
  - Massive field in the bulk → Critical field on the boundary
  - Local update in the bulk → Global update on the boundary



Efficient Sampling from the Bulk

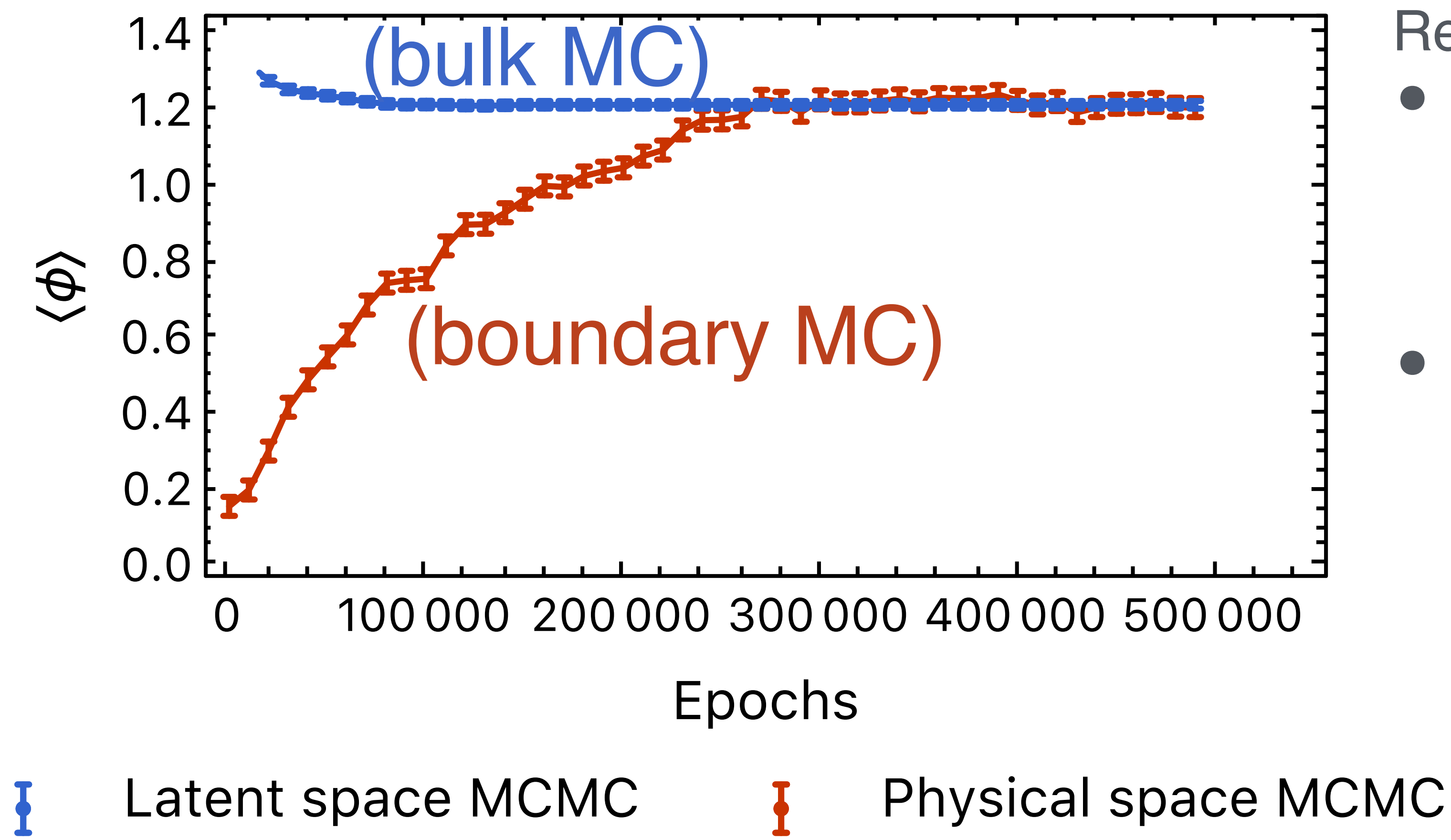
- Sampling: holographic mapping from bulk to boundary
  - Massive field in the bulk → Critical field on the boundary
  - Local update in the bulk → Global update on the boundary





Efficient Sampling from the Bulk

- Sampling: holographic mapping from bulk to boundary
  - Massive field in the bulk → Critical field on the boundary
  - Local update in the bulk → Global update on the boundary
- Order parameters converges faster using bulk MCMC.



Related topics:

- Self-learning MC
  Huang, Wang, PRB (2017)
  Liu, Qi, Meng, Fu, PRB(2017)
  …
- Super-resolution sampling
  Efthymiou, Beach, Melko (2019)

Summary

- We demonstrated several examples of machine learning physics. The common theme:
  - Train the machine on a task (but we don't use it!)
  - Open up the neural network for emergent physics

	Task	Emergent physics
ML Quantum Mechanics arXiv: 1901.11103	Potential-density mapping	Wave function + Schrödinger eq.
ML Holographic Mapping arXiv:1903.00804	Quantum field generation	RG scheme, bulk effective theory