Machine Learning and the Renormalization Group

May 27 – 31, 2024
ECT*

Supporting institutions

STRUCTURES CLUSTER OF EXCELLENCE

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

FONDAZIONE BRUNO KESSLER

ECT* EUROPEAN CENTRE FOR THEORETICAL STUDIES IN NUCLEAR PHYSICS AND RELATED AREAS

RIKEN iTHEMS 理化学研究所 数理創造プログラム RIKEN Interdisciplinary Theoretical and Mathematical Sciences Program

Building Interactions with Deep Autoregressive Networks

Lingxiao Wang(王 凌霄) (RIKEN)

Collaborators: Lianyi He(THU), Yin Jiang and Tian Xu(Beihang Uni.), Kai Zhou(CUHK-Shenzhen/FIAS),…

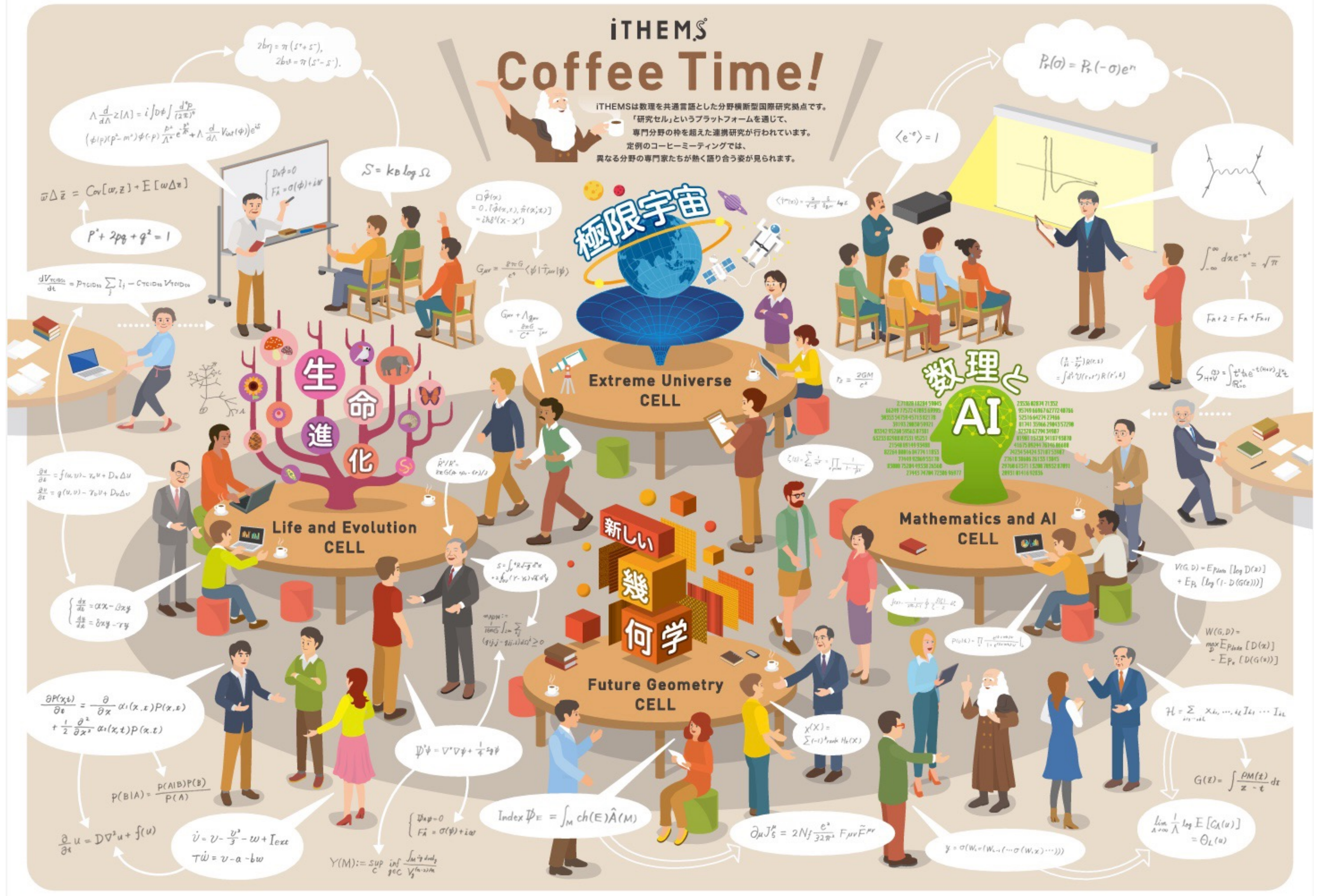
ArXiv: 2007.01037, 2405.10493, 24xx.xxxxx

May 31, 2024, “Machine Learning and Renormalization Group” Workshop, ECT*

RIKEN-iTHEMS

RIKEN iTHEMS
Interdisciplinary Theoretical and
Mathematical Sciences Program
https://ithems.riken.jp/

March 1, 2024



Theorists in mathematics, biology, physics, chemistry, comp/info sciences under one-roof

since Nov.1, 2016

iTHEMS 理化学研究所 数理創造プログラム
RIKEN Interdisciplinary Theoretical and Mathematical Sciences Program

About iTHEMS

HOME / About iTHEMS / Working Groups / DEEP-IN Working Group

DEEP-IN Working Group

"DEEP learning for INverse problems (DEEP-IN) in Sciences" working group (April 1st, 2024 -)

DEEP-IN Working Group Website

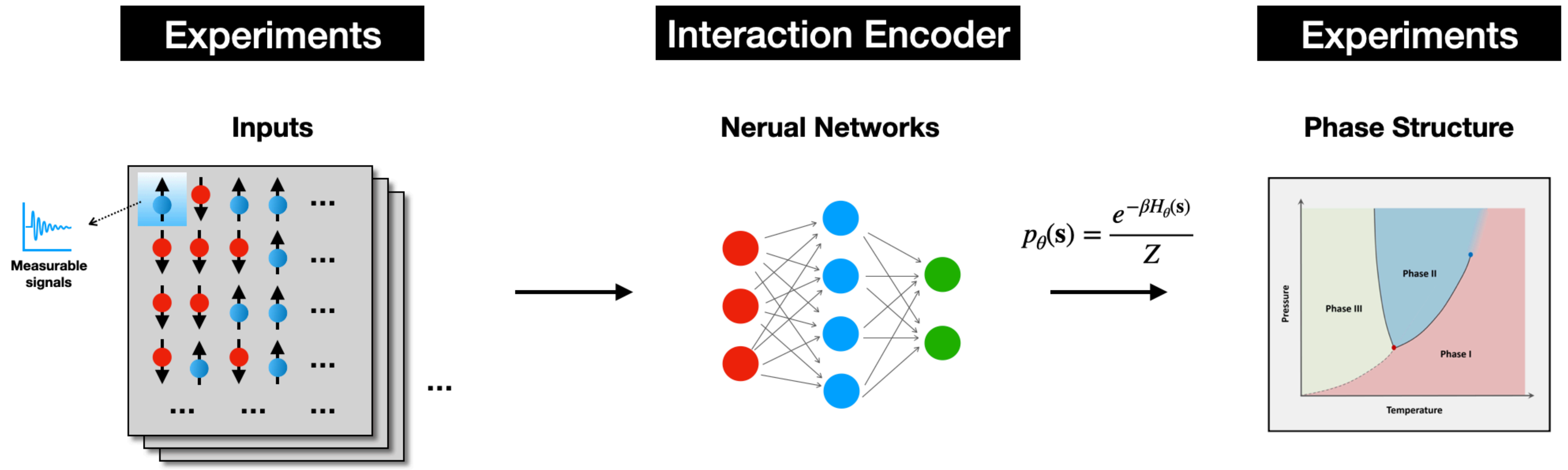
Objectives

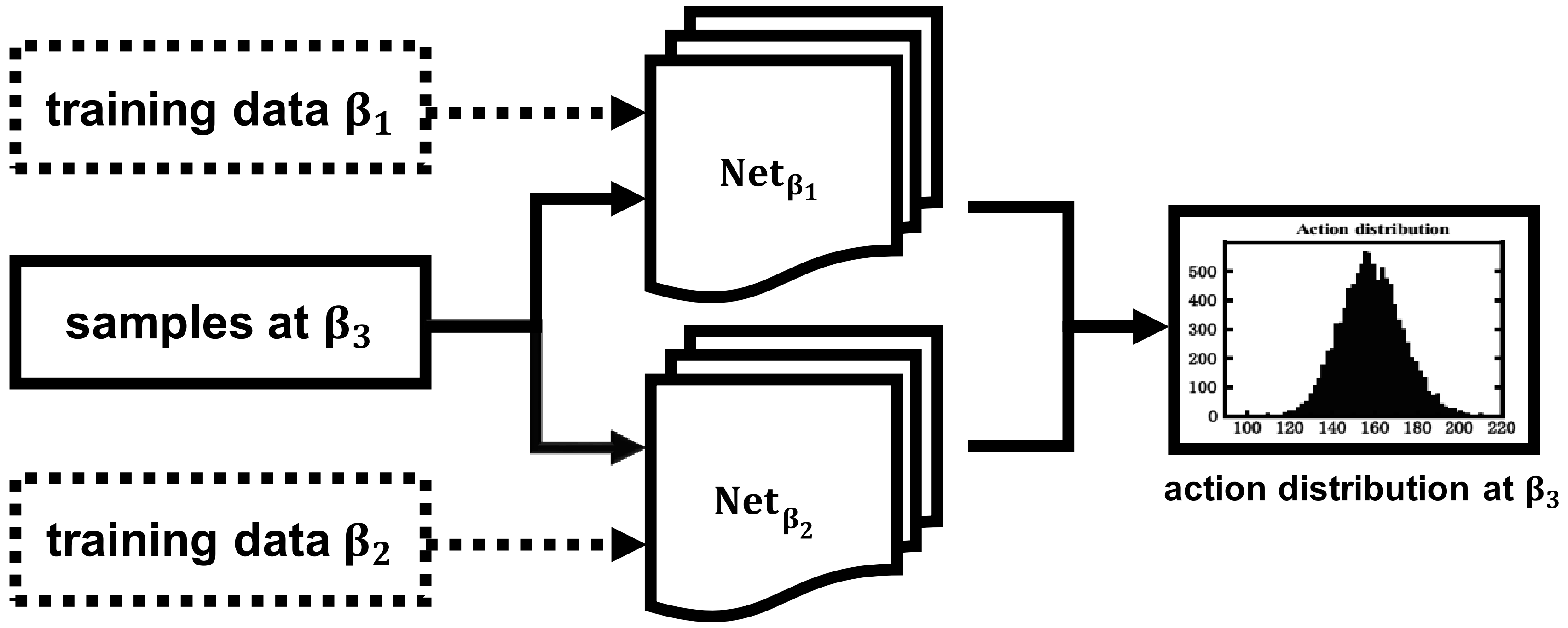
The essence of discovery in sciences has always been rooted in the reverse engineering of natural phenomena and observational data. This paradigm of deducing the underlying laws of nature from observable outcomes forms the cornerstone of our scientific inquiry. The DEEP-IN working group is established with the recognition that the elucidation of such complex phenomena demands a fusion of physics insights and advanced deep learning methodologies. Historically, the exploration of sceinces has relied heavily on intuition and empirical exploration, with methods like inference playing a significant role in our understanding.

Facilitators:
Lingxiao Wang (RIKEN iTHEMS) *Contact at lingxiao.wang@riken.jp
Catherine Beauchemin (RIKEN iTHEMS)
Akinori Tanaka (RIKEN AIP/RIKEN iTHEMS)
Enrico Rinaldi (Quantinuum K.K./RIKEN iTHEMS)
Akira Harada (NITIC/RIKEN iTHEMS)

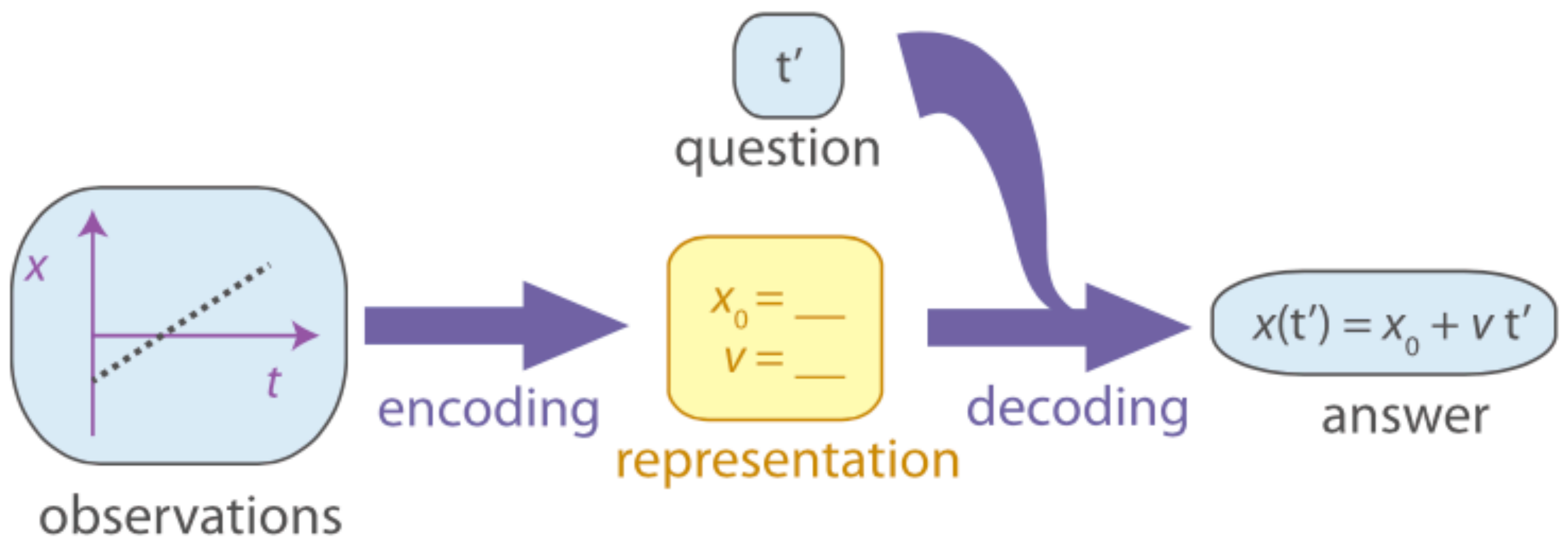
Outline

- **Neural Network as Universal Emulator**
- **Learn from Observations**
  - Building Microscopic Interactions
  - Autoregressive Networks
- **Classical Systems**
  - Ferromagnetic Phase Transition
  - Topological Phase Transition
- **Quantum Systems**
- **Outlooks**

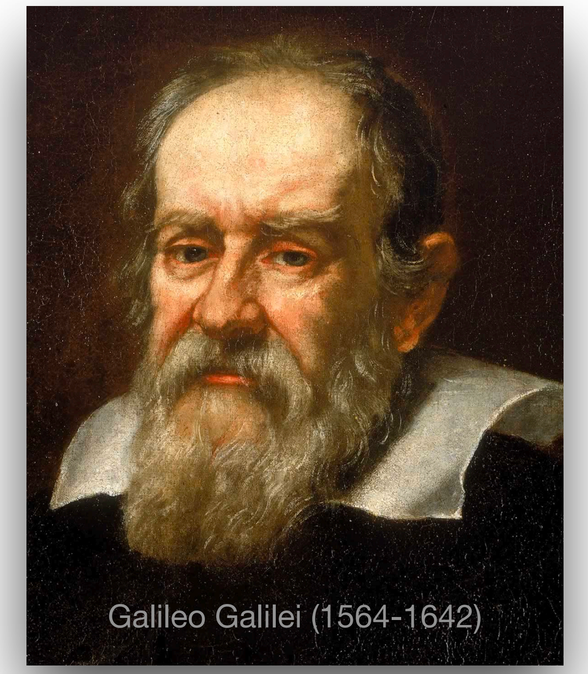




Machine Learning and Physics

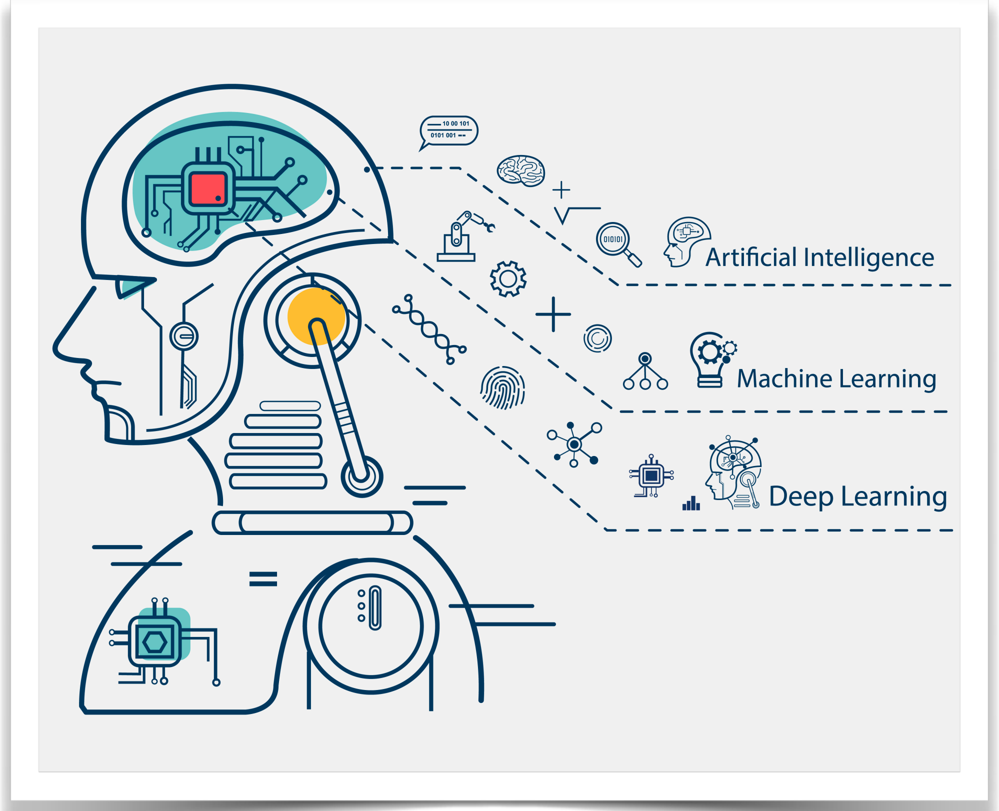


Phys.Rev. Lett. **124**, 010508 (2020)



An inverse problem in science is the process of inferring from a set of observations the causal factors that produced them.

Data, X



Machine, $\{\theta\}$

Prediction

Estimation

Machine Learning and Inference

Maximum Likelihood Estimation(MLE)

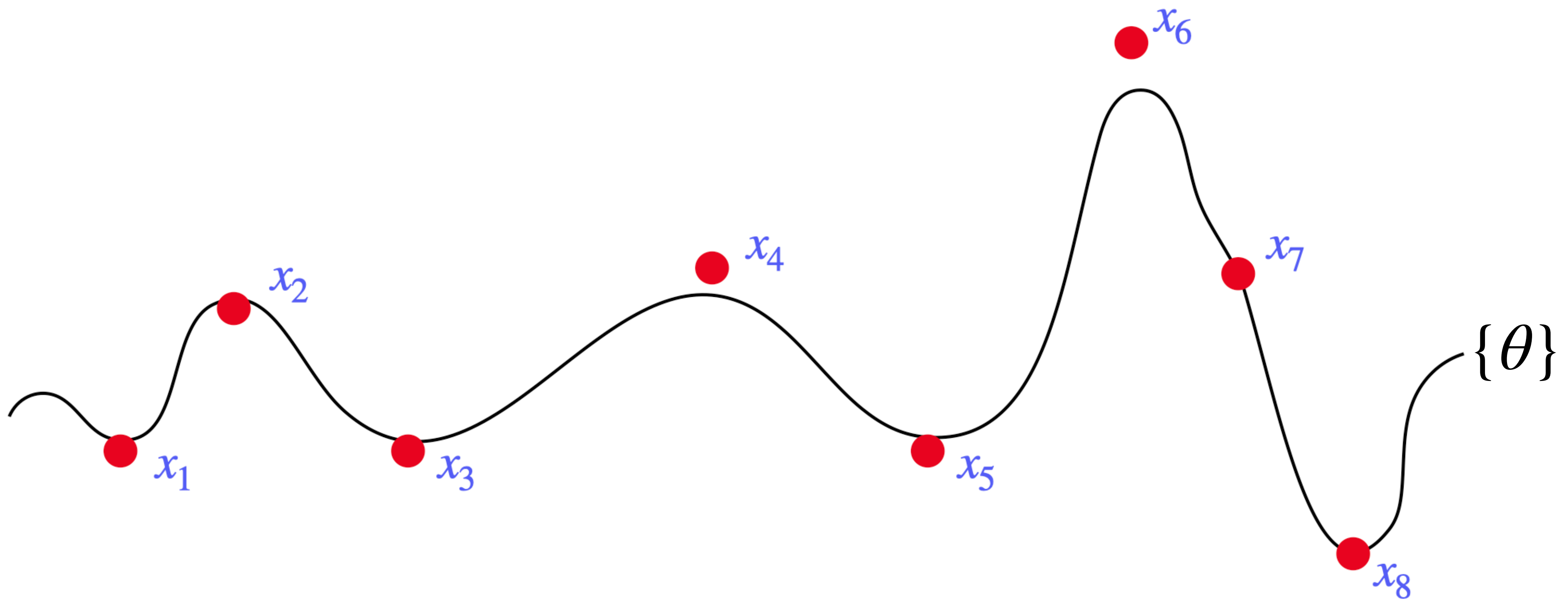
$$\max_{\theta} \prod_{i=1}^{N} p(\mathbf{x}_i \mid \theta)$$

Bayesian Inference

Maximum A Posterior(MAP)

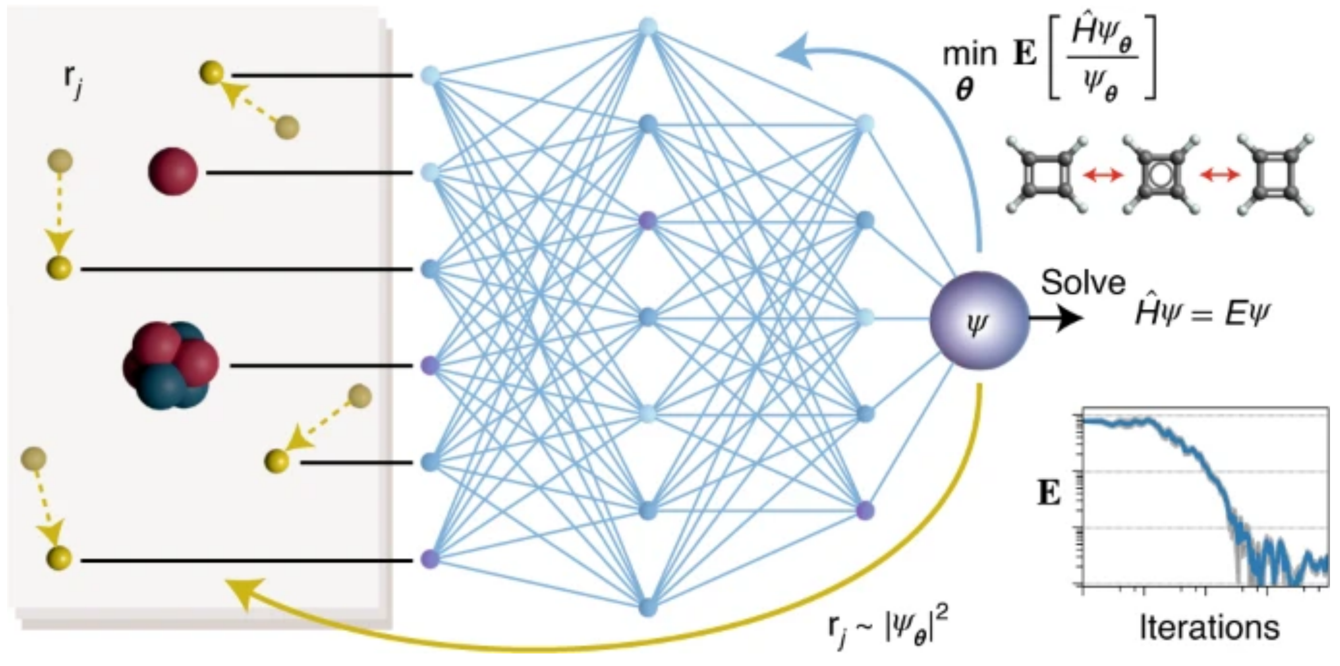
$$p(\theta \mid X) = \frac{p(X \mid \theta)\pi(\theta)}{p(X)}$$

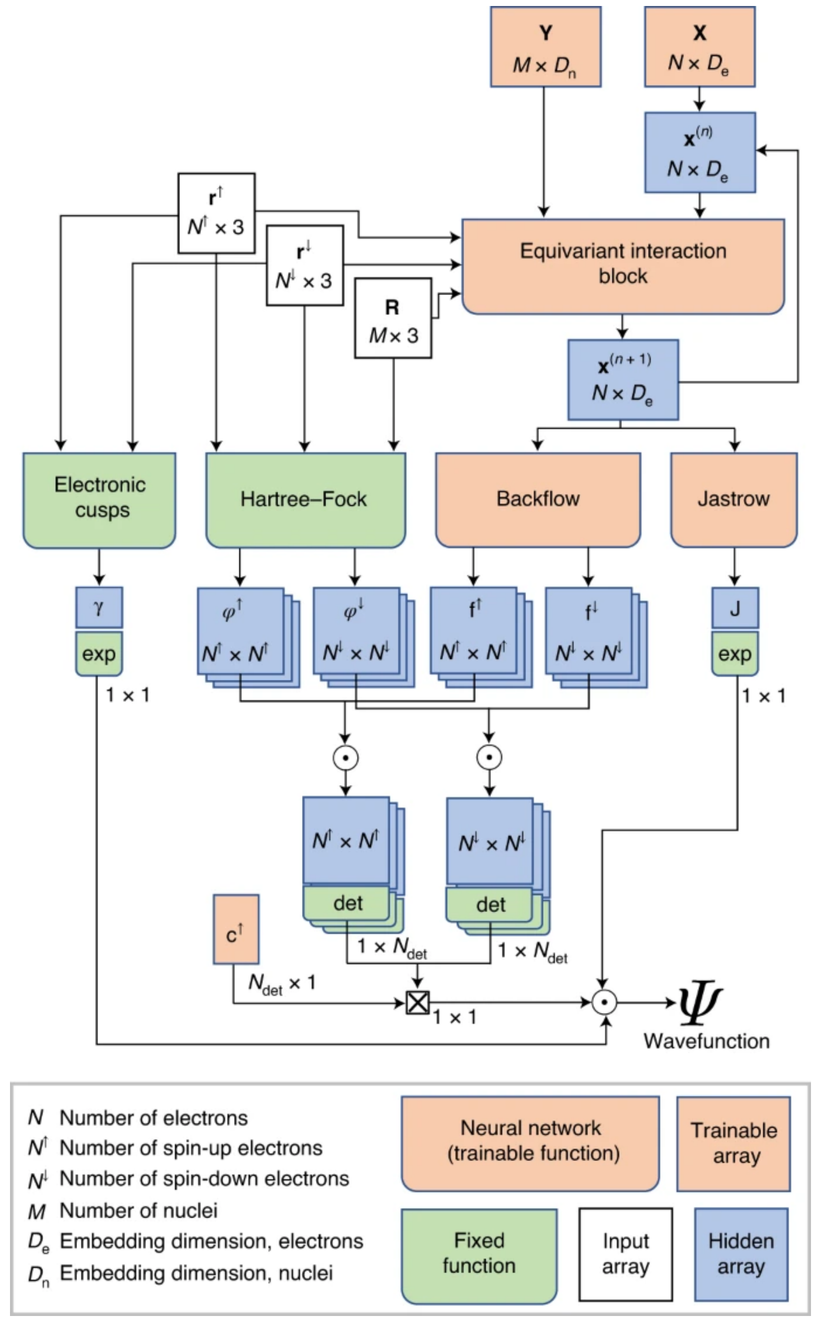
Posterior $p(\theta \mid X)$, **Prior** $\pi(\theta)$, **Evidence** $p(X)$



Neural Network as Universal Emulator

Many-Body Wave Functions





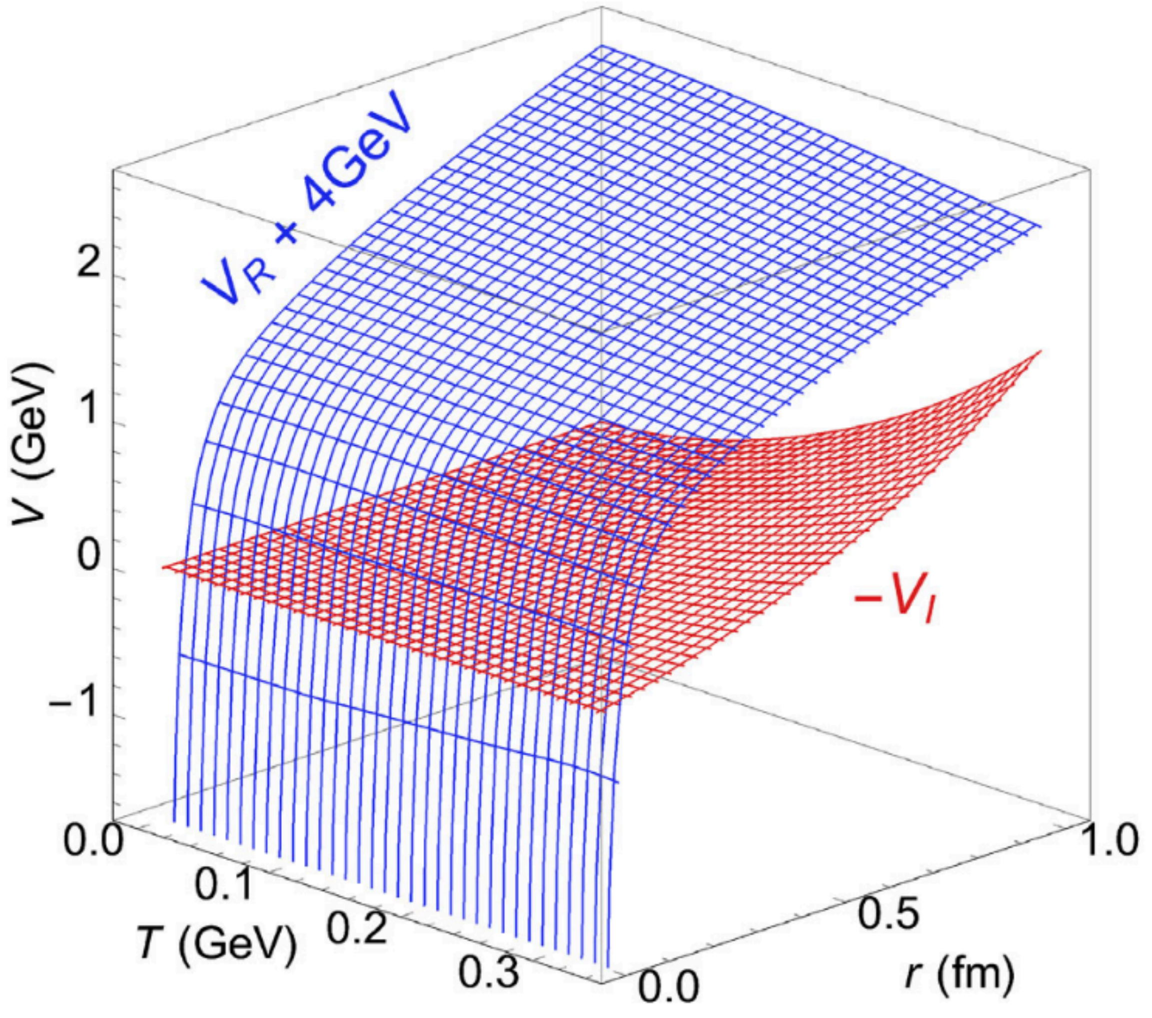
Architecture of the newly developed PauliNet wavefunction ansatz.

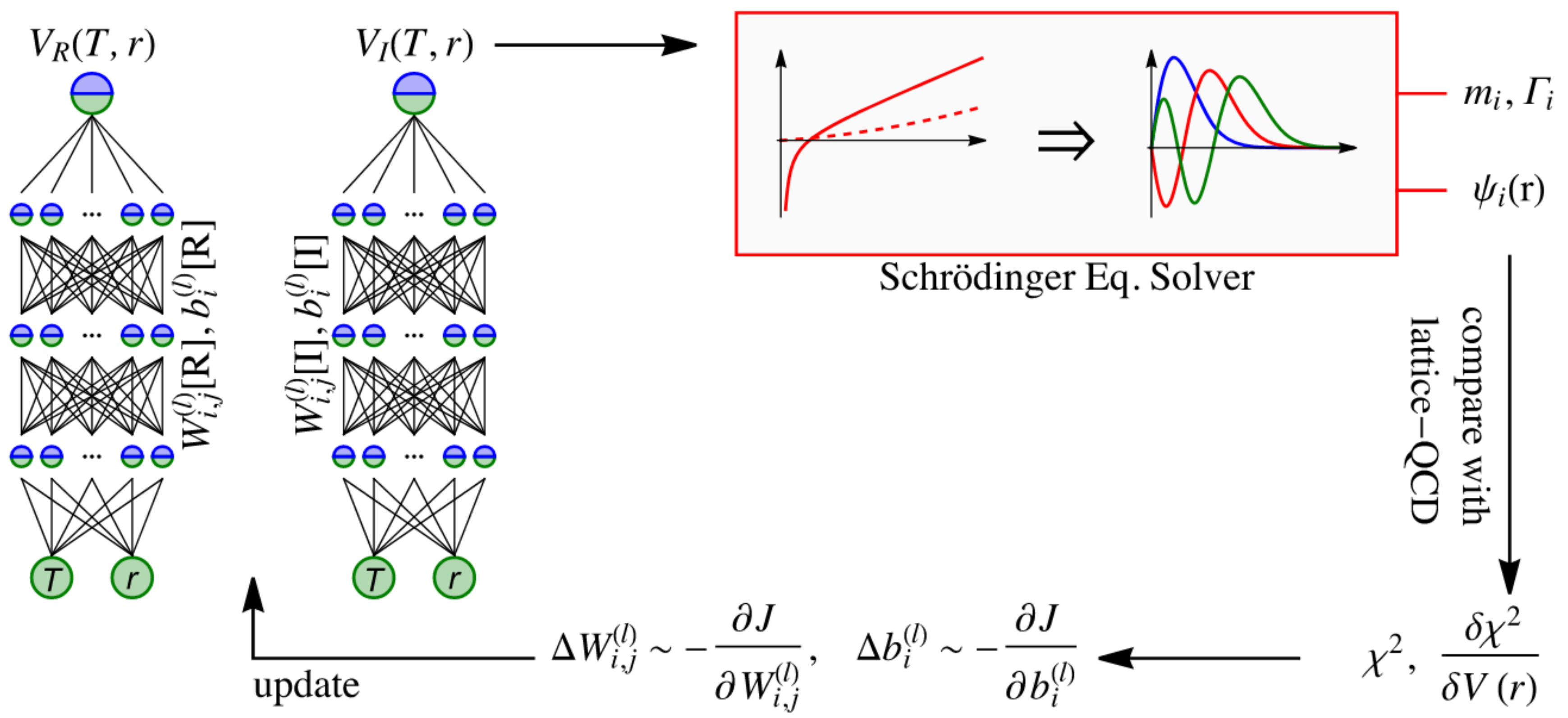
Hermann, J., Schätzle, Z. & **Noé, F.** , Nat. Chem. 12, 891–897 (2020).

Neural Network as Universal Emulator

Many-Body Wave Functions

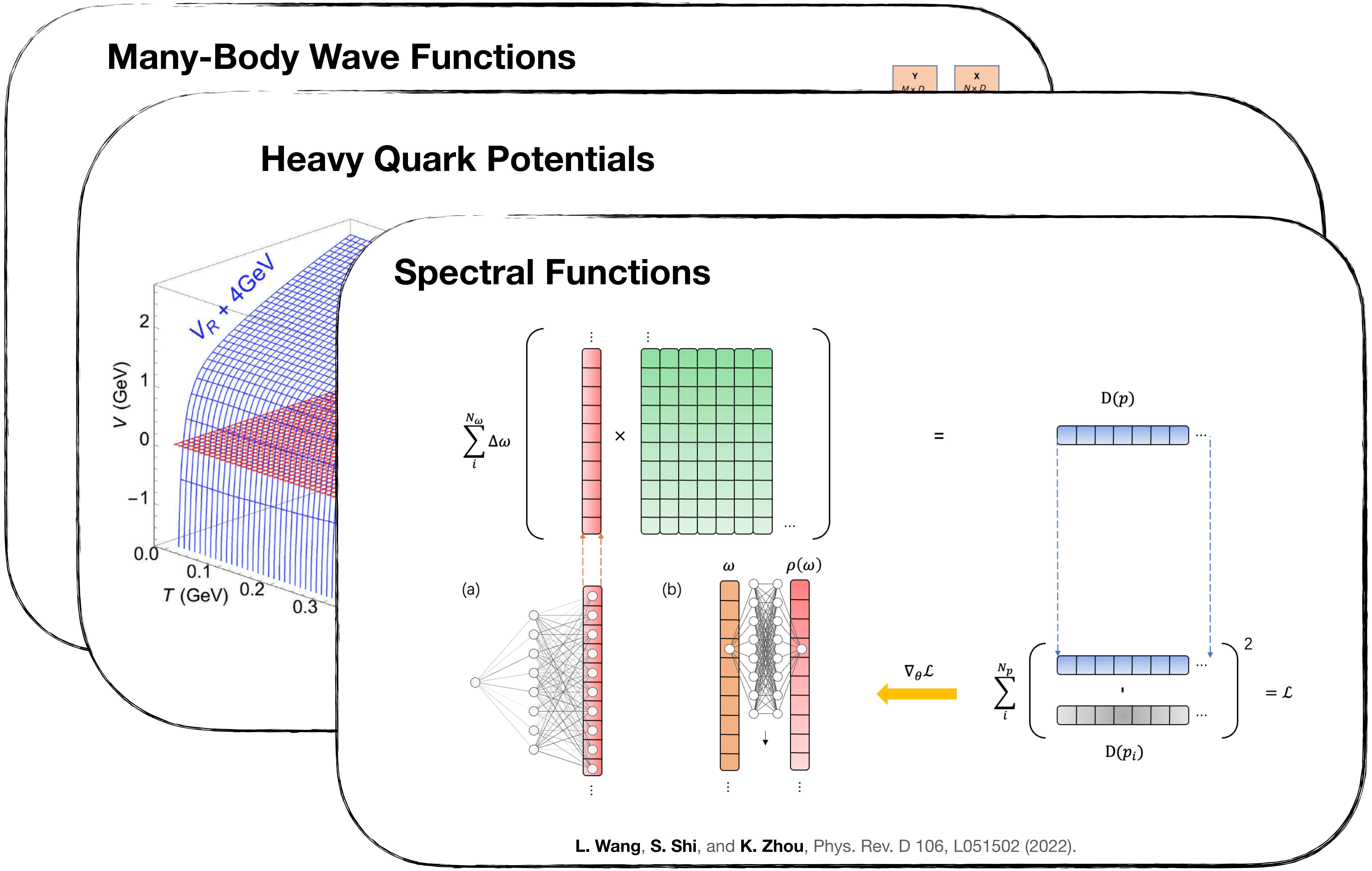
Heavy Quark Potentials





S. Shi, K. Zhou, J. Zhao, S. Mukherjee, and P. Zhuang, Phys. Rev. D **105**, 014017 (2022).

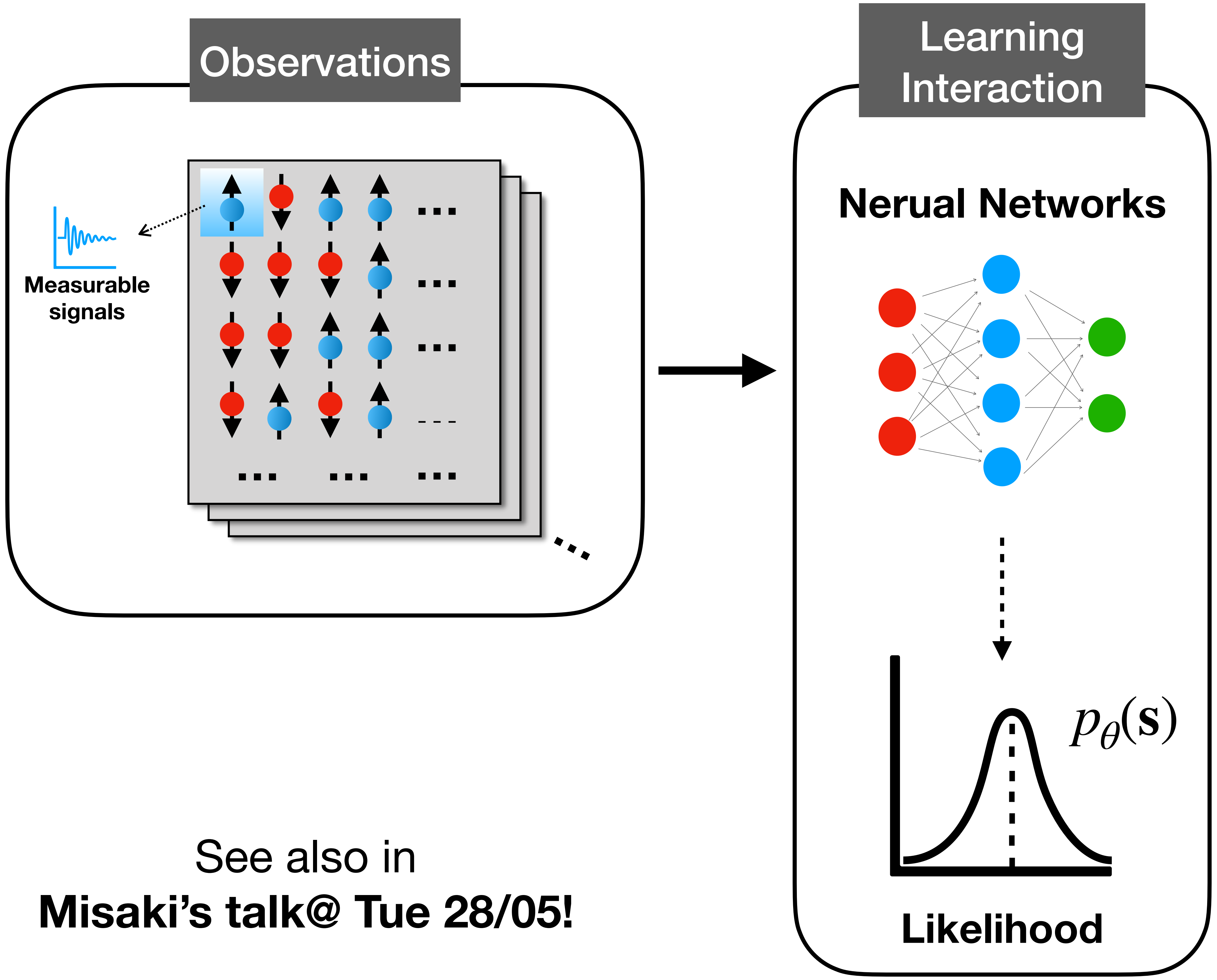
Neural Network as Universal Emulator



- **K. Zhou, L. Wang,** L.-G. Pang, and S. Shi, Prog. Part. Nucl. Phys. 135, 104084 (2023).
- K. Cranmer, **G. Kanwar,** S. Racanière, D. J. Rezende, and P. E. Shanahan, Nat Rev Phys 1 (2023).
- G. Carleo, etc., Rev. Mod. Phys. 91, 045002 (2019).
- V. Dunjko and H. J. Briegel, Rep. Prog. Phys. 81, 074001 (2018).

Can We Learn Microscopic Interactions From Observations Directly?

Learn Microscopic Interactions



See also in
Misaki's talk@ Tue 28/05!

$$\max_\theta \prod_{i=1}^{N} p_\theta(\mathbf{s}_i)$$

$$p(\mathbf{s}) \leftarrow p_\theta(\mathbf{s}) \equiv \frac{e^{-\frac{H_\theta(\mathbf{s})}{T}}}{Z}$$

$$H_\theta(\mathbf{s}, T) = -T \ln p_\theta(\mathbf{s}) - T \ln Z$$

$$\frac{\Delta H_\theta(\mathbf{s}, T)}{T'} \equiv -\frac{T}{T'}(\ln p_\theta(\mathbf{s} + \delta \mathbf{s}) - \ln p_\theta(\mathbf{s}))$$

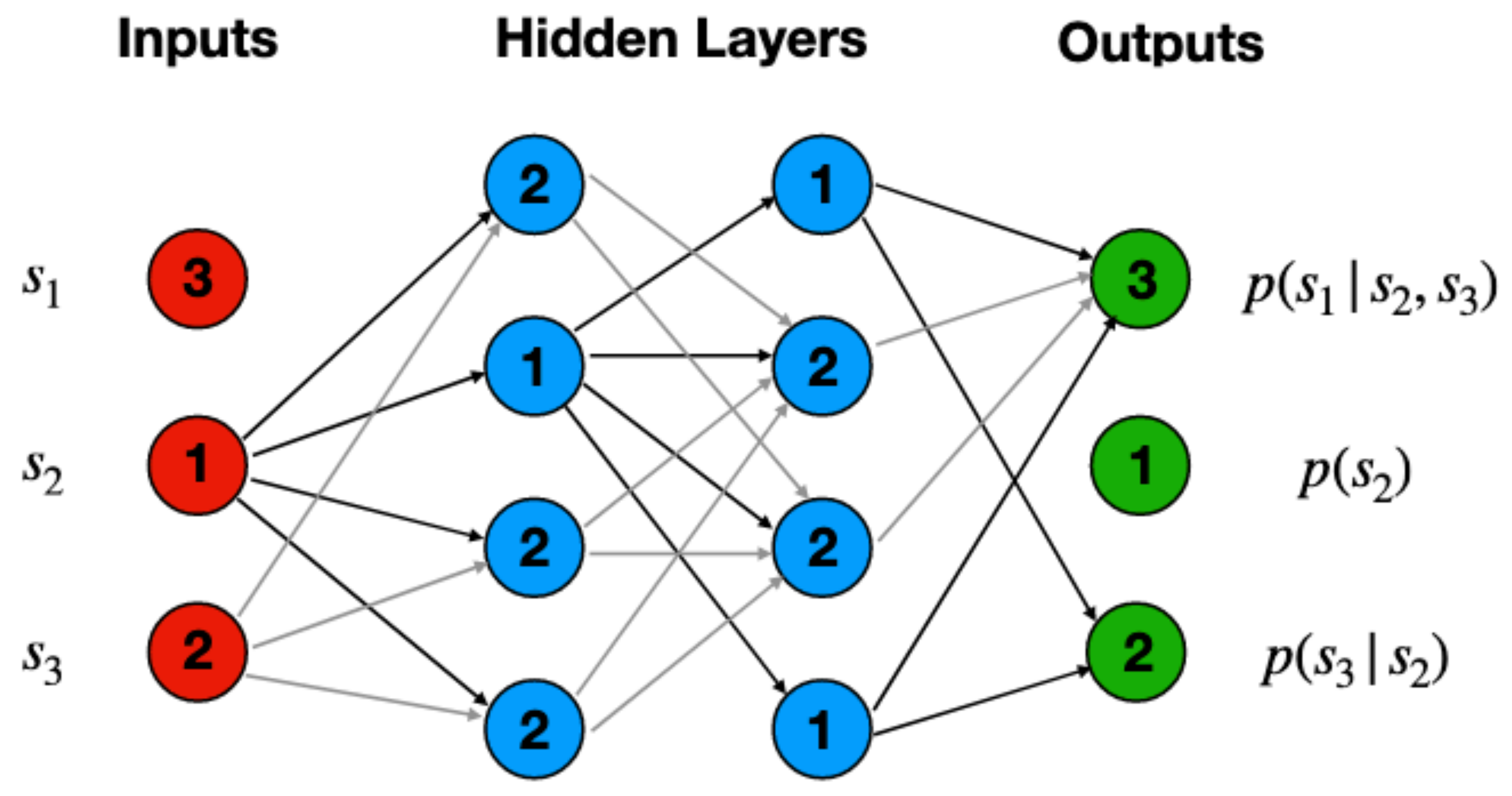
Autoregressive Networks

$$p_\theta(s) = \prod_{i=1}^{N} p(s_i \,|\, s_1, \ldots, s_{i-1})$$

- Example (L=3) for 1D spin model

$$p_\theta(s) = p(s_3 \,|\, s_2, s_1)p(s_2 \,|\, s_1)p(s_1)$$

- $p(s_i \,|\, s_{<i})$ can be any naive distribution



Bernoulli distribution for Ising model

$$p(s_i \,|\, s_{<i}) = q_i \delta_{s_i,+1} + (1 - q_i)\delta_{s_i,-1}$$

$q_1 = f(s_1 = +1), q_2 = f(s_2 = +1 \,|\, s_1), q_3 = f(s_3 = +1 \,|\, s_2, s_1)$

discrete *d.o.f.s*

Chinese Phys. Lett. 39, 120502 (2022)

Beta distribution for continuous *d.o.f.*, $X \sim Beta(a, b)$

$$p_\theta(s_i \,|\, s_1, \cdots, s_{i-1}) = \frac{\Gamma(a_i + b_i)}{\Gamma(a_i)\Gamma(b_i)} s_i^{a_i - 1}(1 - s_i)^{b_i - 1}$$

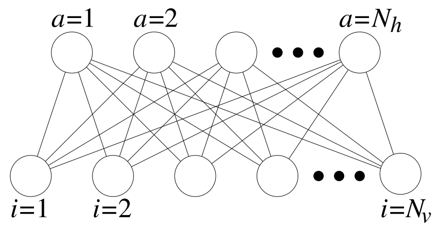
$\Gamma(a)$ is gamma function, $s_i = \theta_i/2\pi \in [0,1), (a_i, b_i) > 0$

continuous *d.o.f.s*

Autoregressive Networks

$$p_\theta(s) = \prod_{i=1}^{N} p(s_i \mid s_1, \ldots, s_{i-1})$$

Network is parametrized by a triangular matrix L, which ensures that s_i is independent with s_j when $j \geq i$. This is named as autoregressive property in machine learning.



Gaussian scalar field RBM

- induced distribution on visible layer

$$p(\phi) = \int Dh\, p(\phi, h) = \frac{1}{Z} \exp\left(-\frac{1}{2}\sum_{i,j} \phi_i K_{ij} \phi_j + \sum_i J_i \phi_i\right)$$

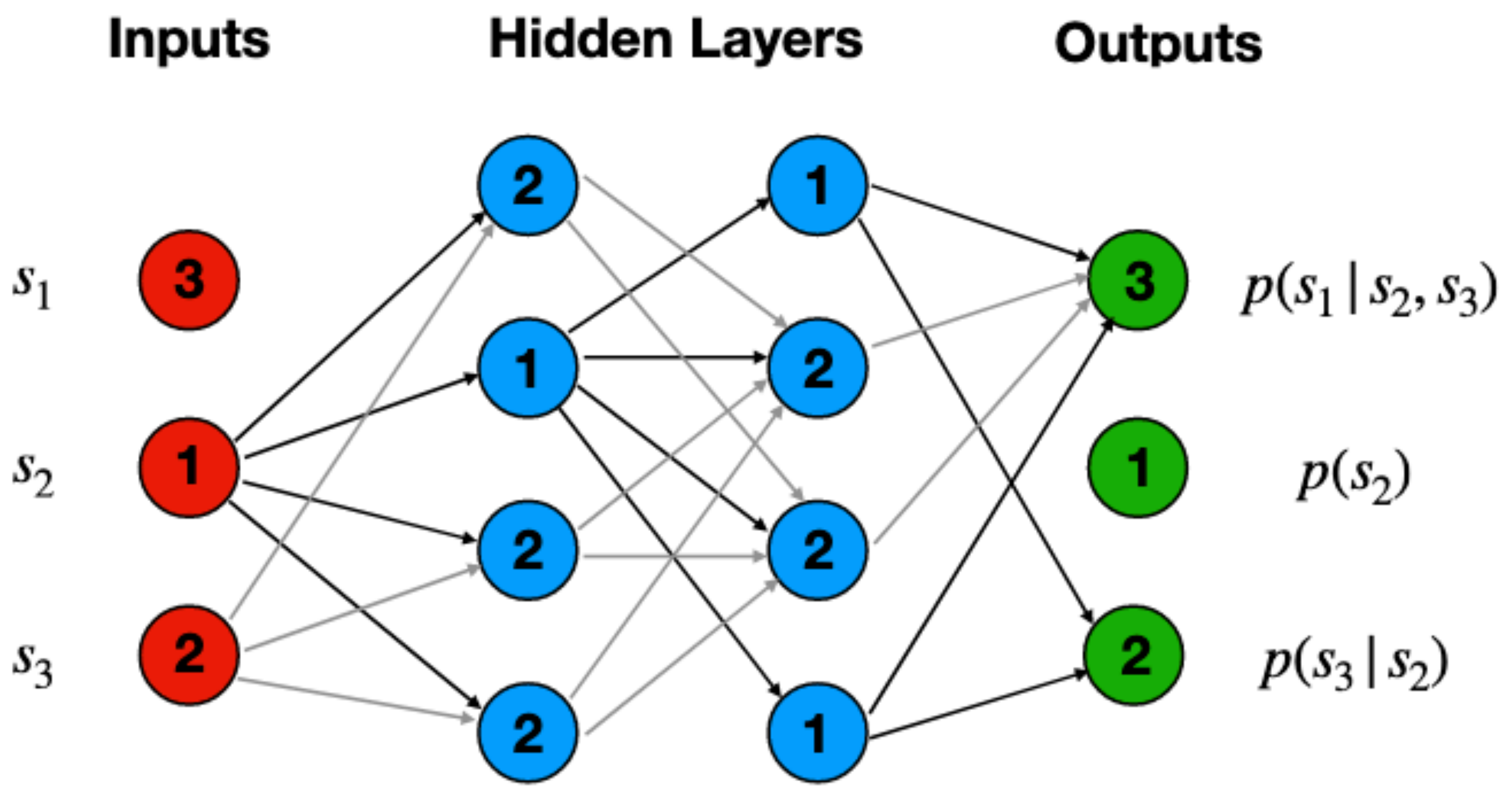
- scalar field with kinetic (all-to-all) term $K_{ij} = \mu_i^2 \delta_{ij} - \sigma^2 \sum_a w_{ia} w_{aj}^T$

  and source $J_i = \sum_a w_{ia} \eta_a$

- unusual Gaussian LFT: what is the weight matrix W and bias η?

Gert's slides@XQCD2023

arXiv:2007.01037



$$WW^T = \frac{1}{\sigma^2}(\mu^2 \mathbb{1} - K^\phi) \equiv \mathcal{K}$$

Exact results for $N_h = N_v$

(infinitely) many solutions for weight matrix: $\mathcal{K}$ is symmetric and positive-definite

1. Cholesky decomposition $\mathcal{K} = LL^T$: $W = L$ triangular
2. diagonalisation $\mathcal{K} = ODO^T = O\sqrt{D}O^T O\sqrt{D}O^T$: $W = W^T = O\sqrt{D}O^T$
3. non-uniqueness: internal symmetry $W \to WO_R$ → $\phi^T W h \to \phi^T W O_R h = \phi^T W h'$

in practice

- all equally valid, realisation depends on initialisation
- non-observable degeneracy due to internal symmetry on hidden layer

Autoregressive Networks

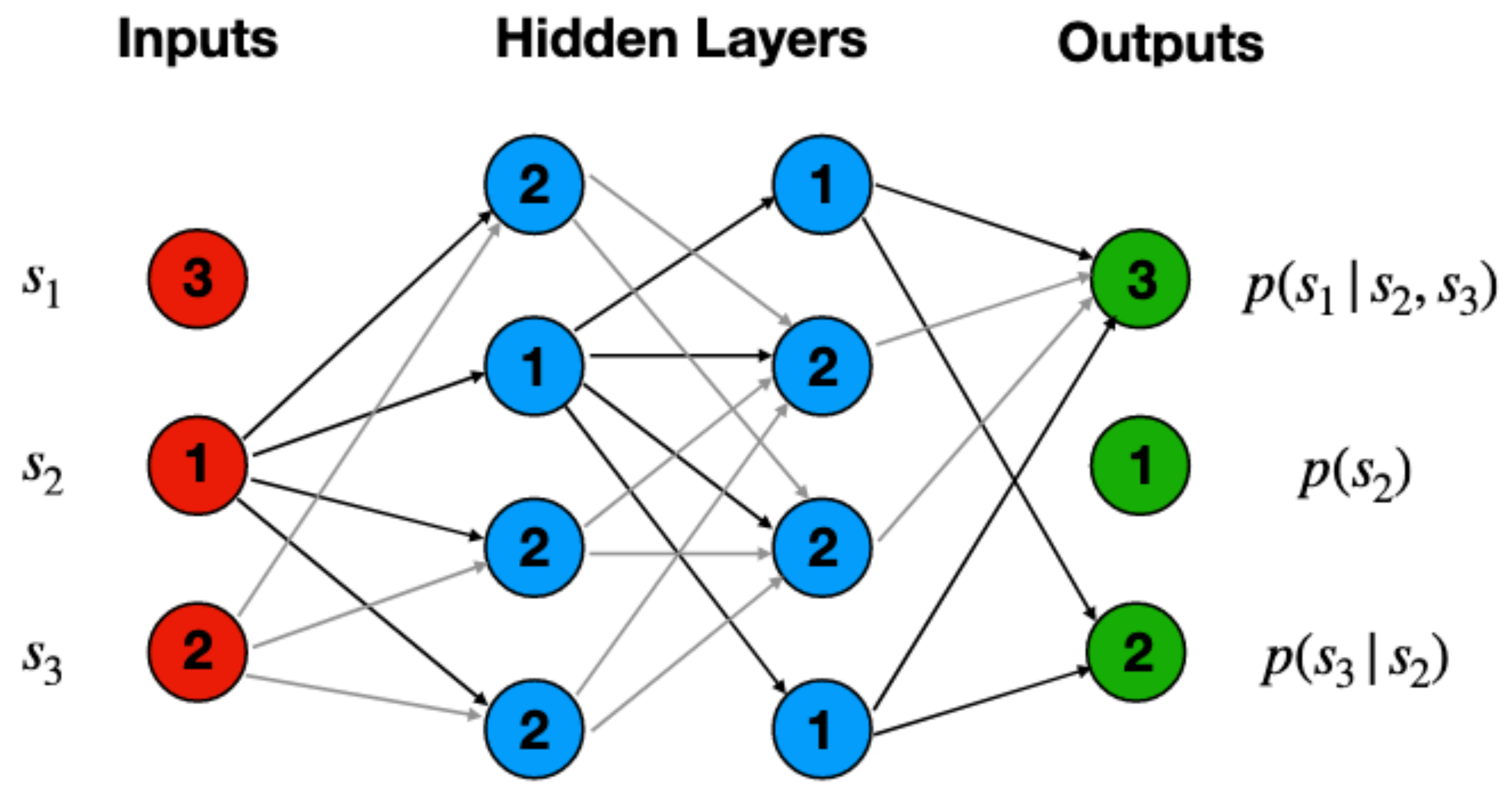
$$p_\theta(s) = \prod_{i=1}^{N} p(s_i \mid s_1, \ldots, s_{i-1})$$

- Example (L=3) for 1D spin model

$$p_\theta(s) = p(s_3 \mid s_2, s_1) p(s_2 \mid s_1) p(s_1)$$

- $p(s_i \mid s_{<i})$ can be any naive distribution

arXiv:2007.01037



Autoregressive networks are variants of RBM with $N_h = N_v$ and triangular weight matrix!

no "UV cut-off"

G. Aarts, B. Lucini, and C. Park, Scalar Field Restricted Boltzmann Machine as an Ultraviolet Regulator**,** *Phys. Rev. D 109, 034521 (2024).*

Learning Interactions

$$\max_{\theta} \prod_{i=1}^{N} p_\theta(\mathbf{s}_i)$$

arXiv:2007.01037

on specific degrees of freedom(*d.o.f.s*)

1. Prepare data-set from observations

$$\mathbf{s} \sim q_{\text{data}}$$

2. Put them into the deep autoregressive network(DAN)

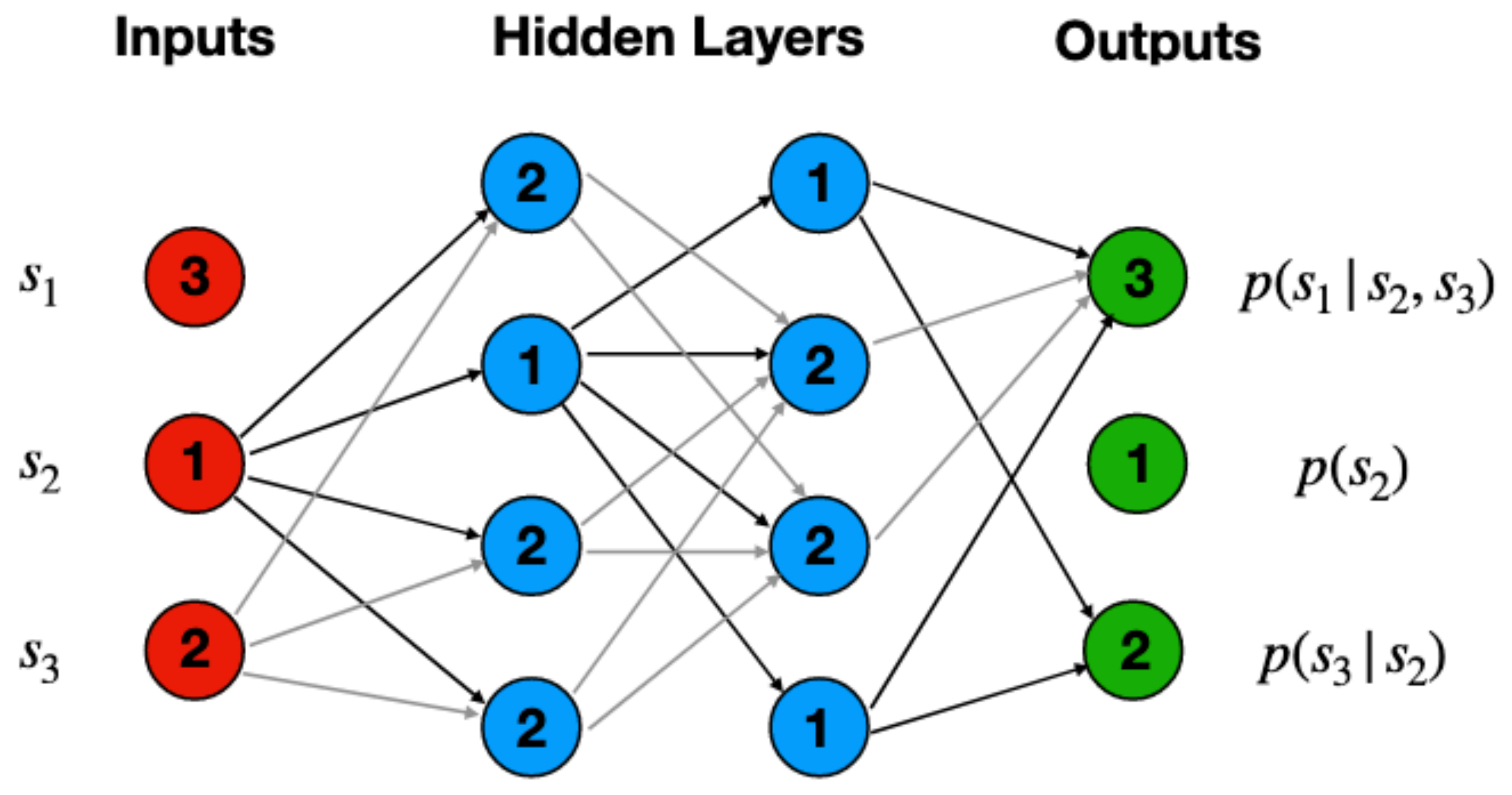
$$p_\theta(\mathbf{s}) = \prod_{i=1}^{N} p(s_i \mid s_1, \ldots, s_{i-1}, \theta)$$

3. Minimize the negative log-liklihood(NLL)

$$\mathscr{L} = -\sum_{\mathbf{s}\sim q_{data}} \sum_{d=1}^{N} \log(p(s_d \mid \mathbf{s}_{<d}, \theta))$$

4. Get your DAN represented Hamiltonian

$$H_\theta(\mathbf{s}, T) = -T \ln p_\theta(\mathbf{s})$$

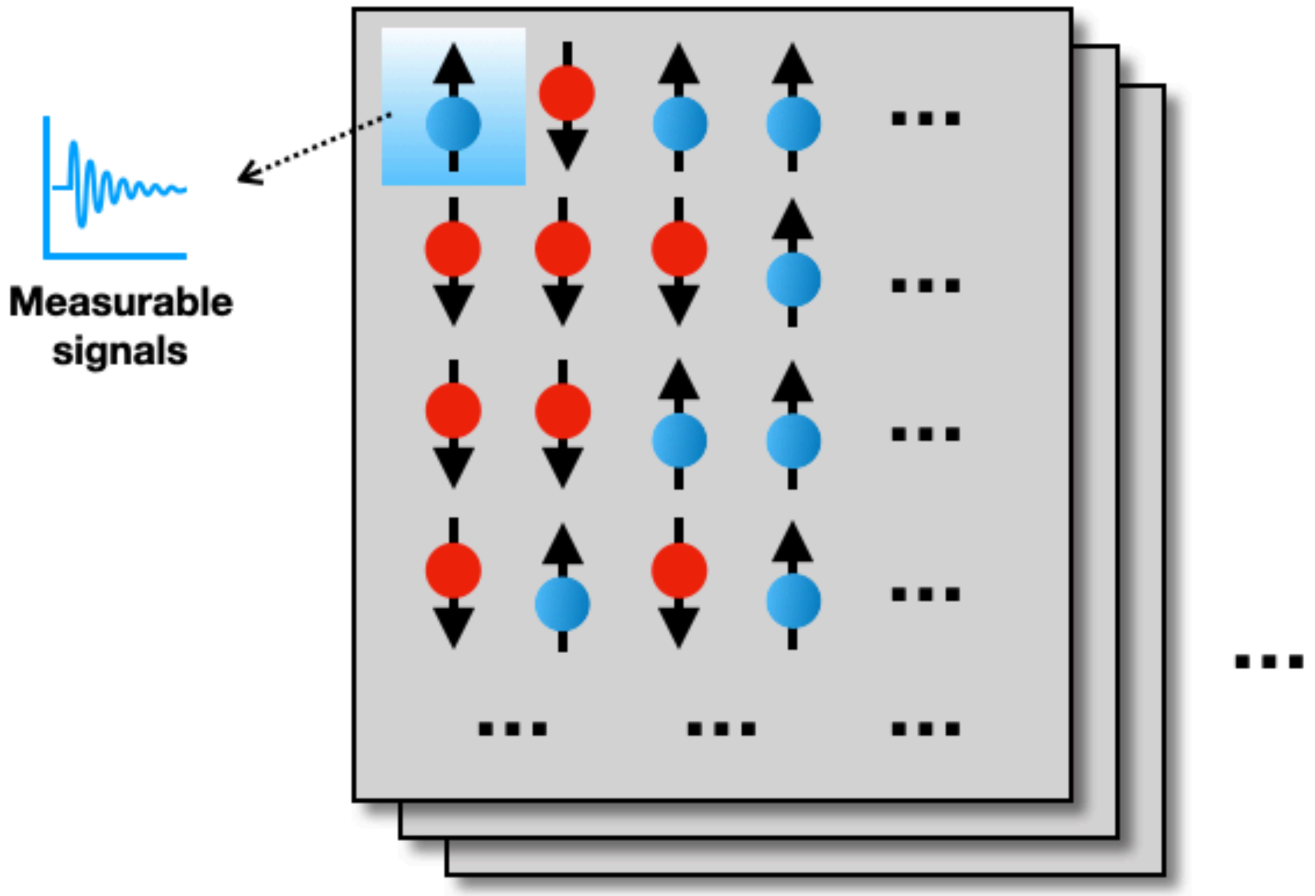


What Can We Do With Neural Network Interactions?

Detect Phase Transitions

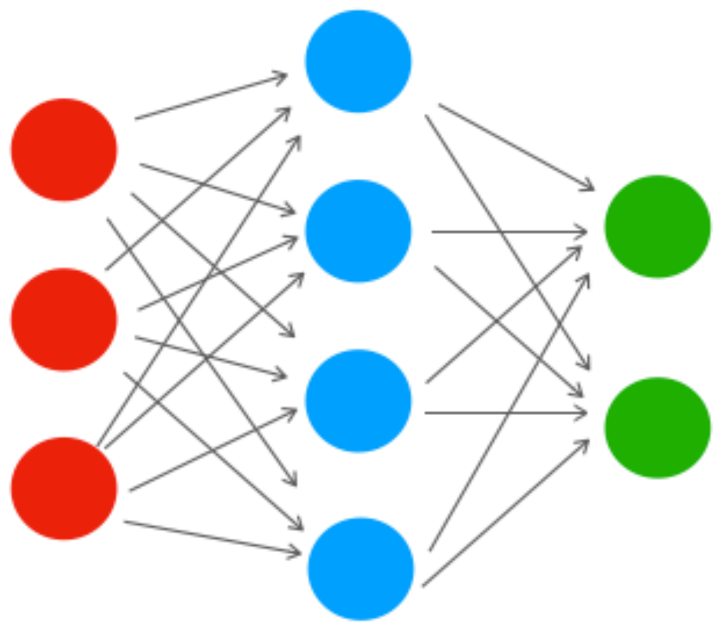
Experiments

Inputs



Interaction Encoder

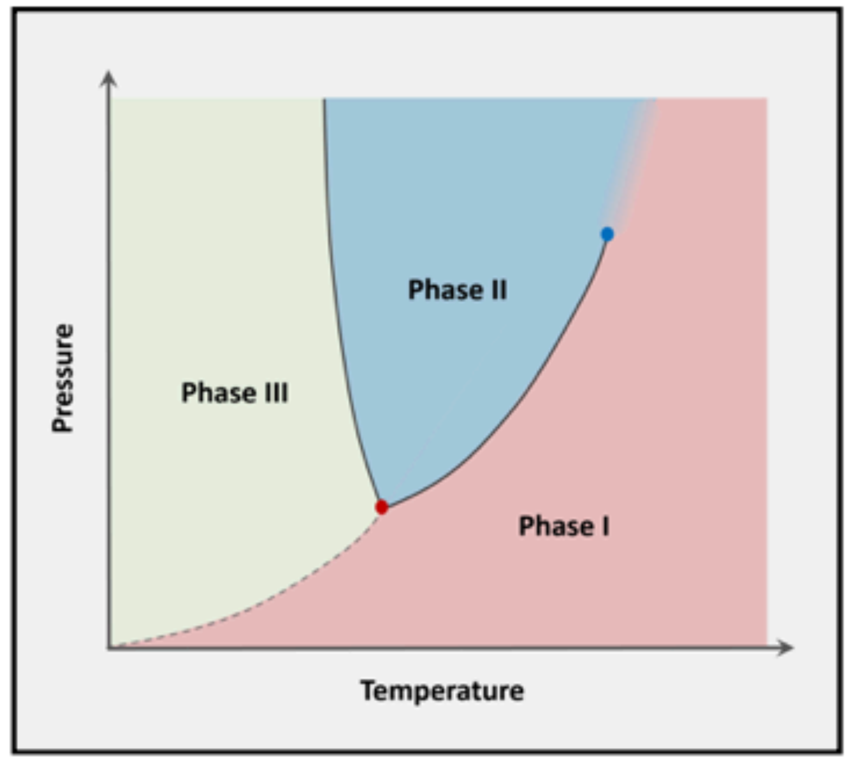
Nerual Networks



$$p_\theta(\mathbf{s}) = \frac{e^{-\beta H_\theta(\mathbf{s})}}{Z}$$

Experiments

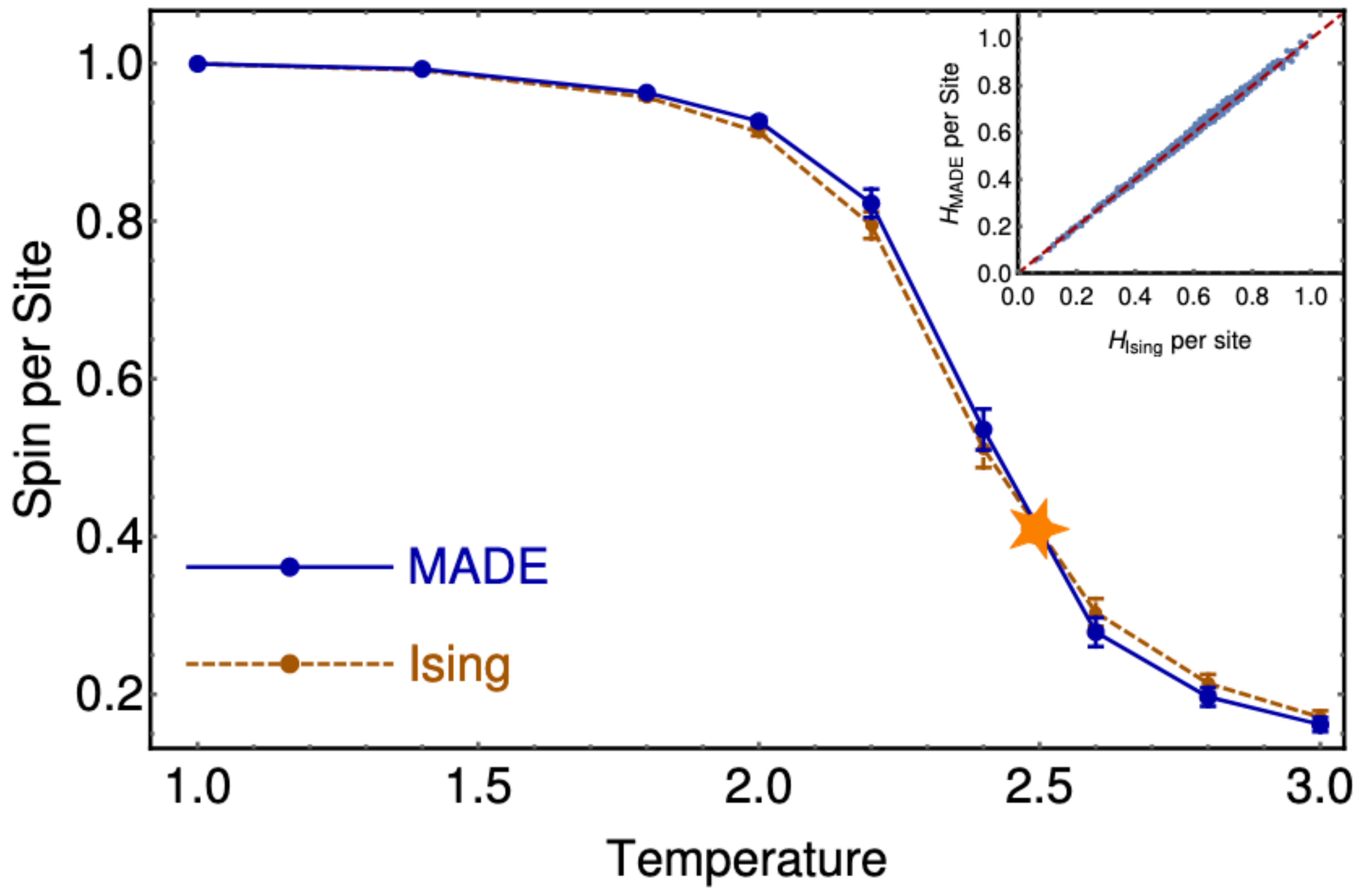
Phase Structure



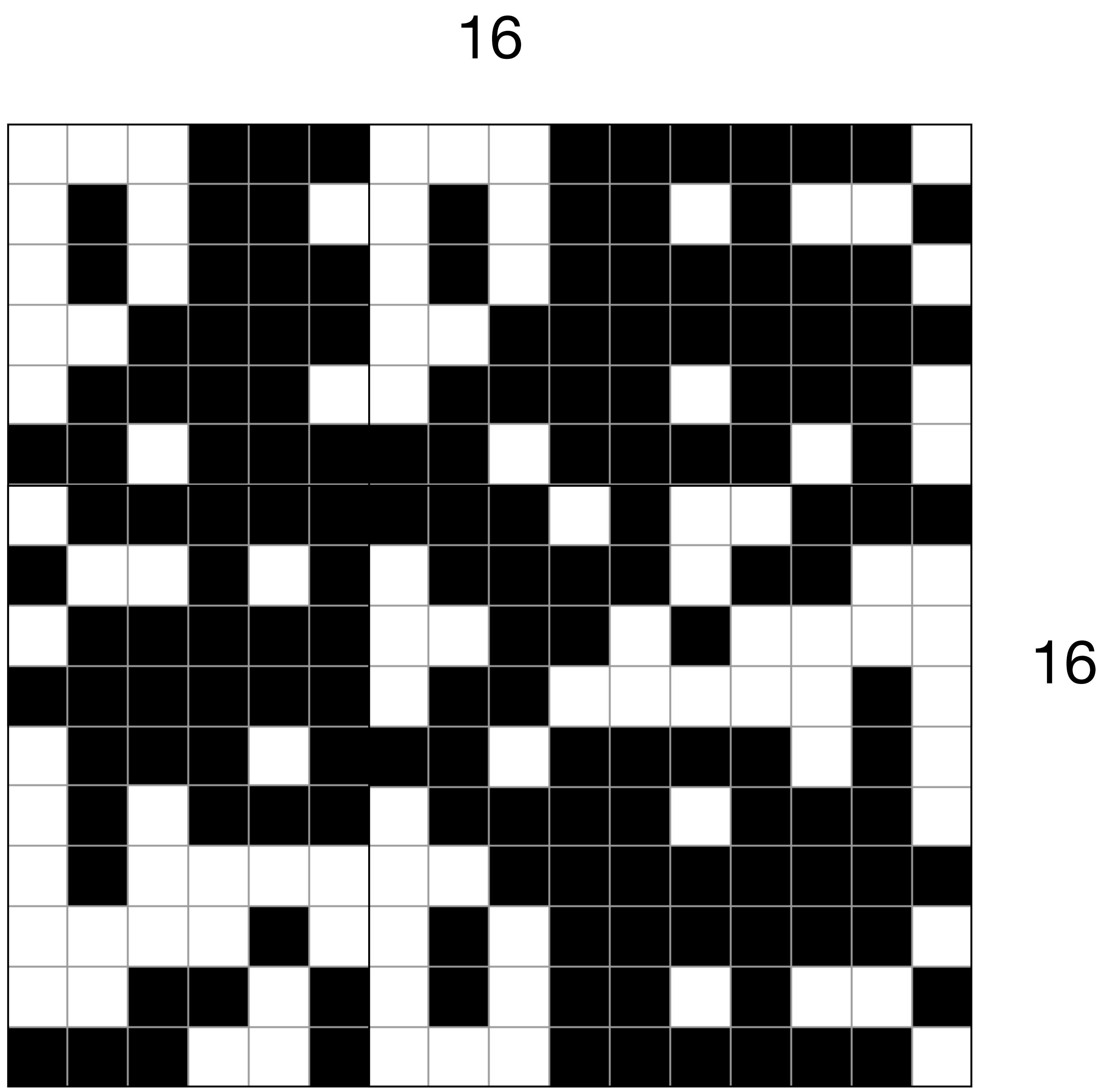
Ferromagnetic Phase Transitions

1. 2D Ising Model

arXiv:2007.01037



Masked Autoencoder for Distribution Estimation **(MADE)**

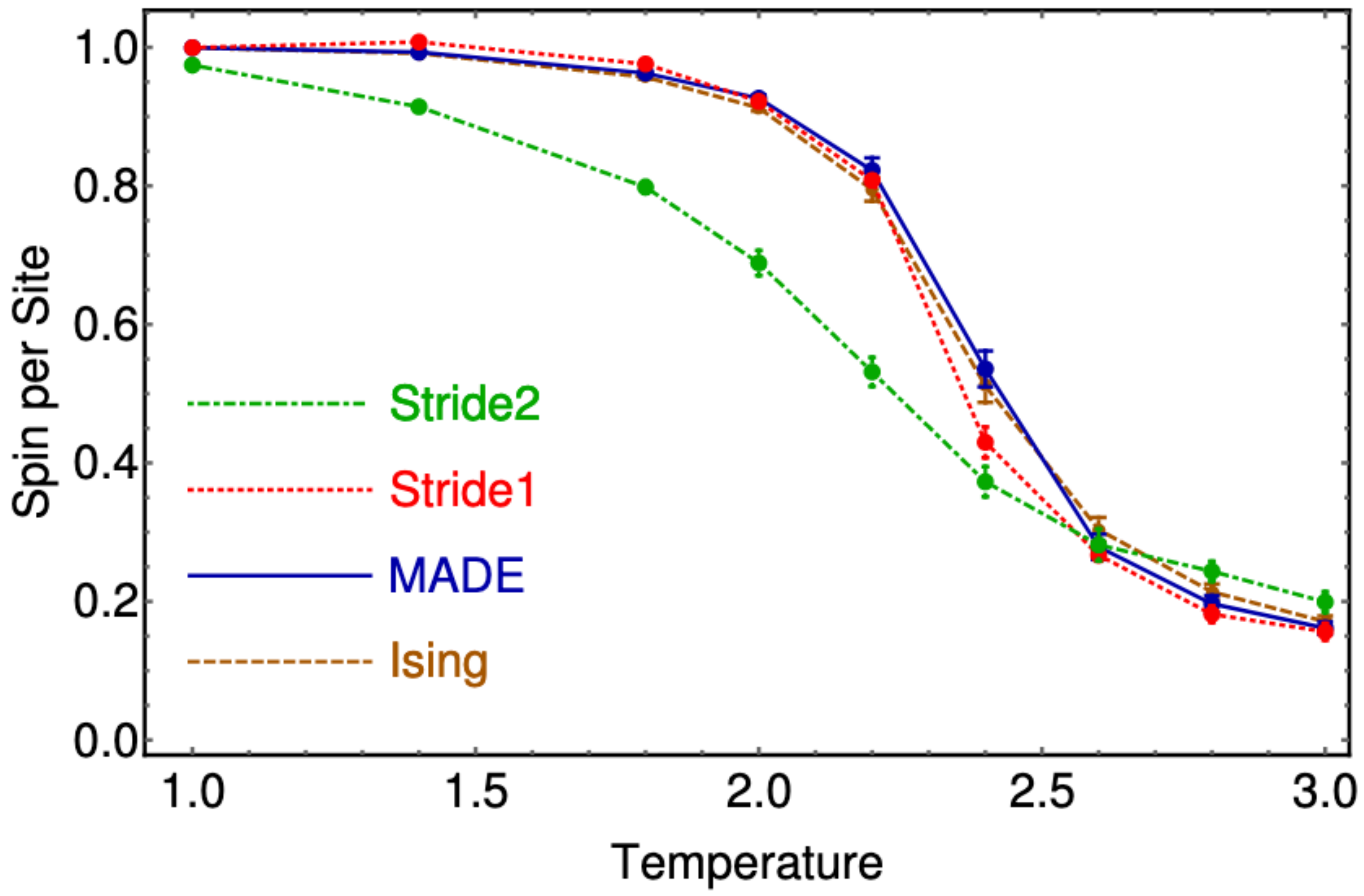


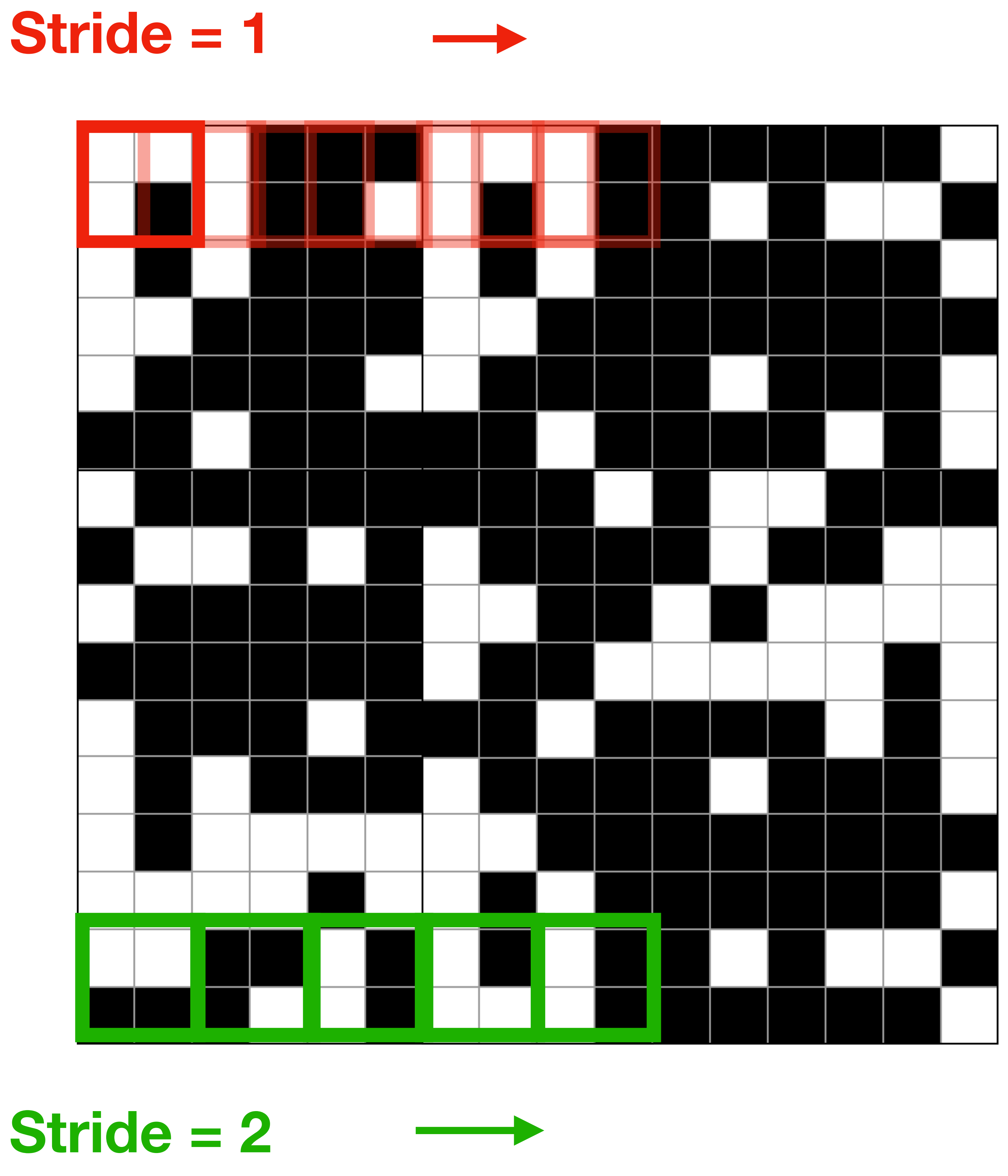
$$H(\mathbf{s}) = -\sum_{<i,j>} s_i s_j$$

Ferromagnetic Phase Transitions

1. 2D Ising Model

arXiv:2007.01037



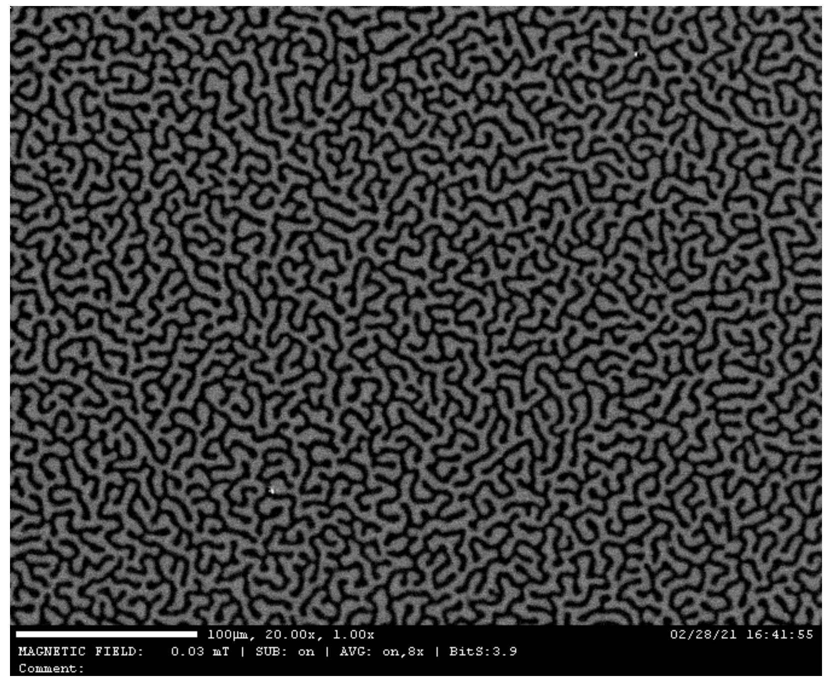


2×2 **spin-block**

Ferromagnetic Phase Transitions

2. Ferromagnetic Materials

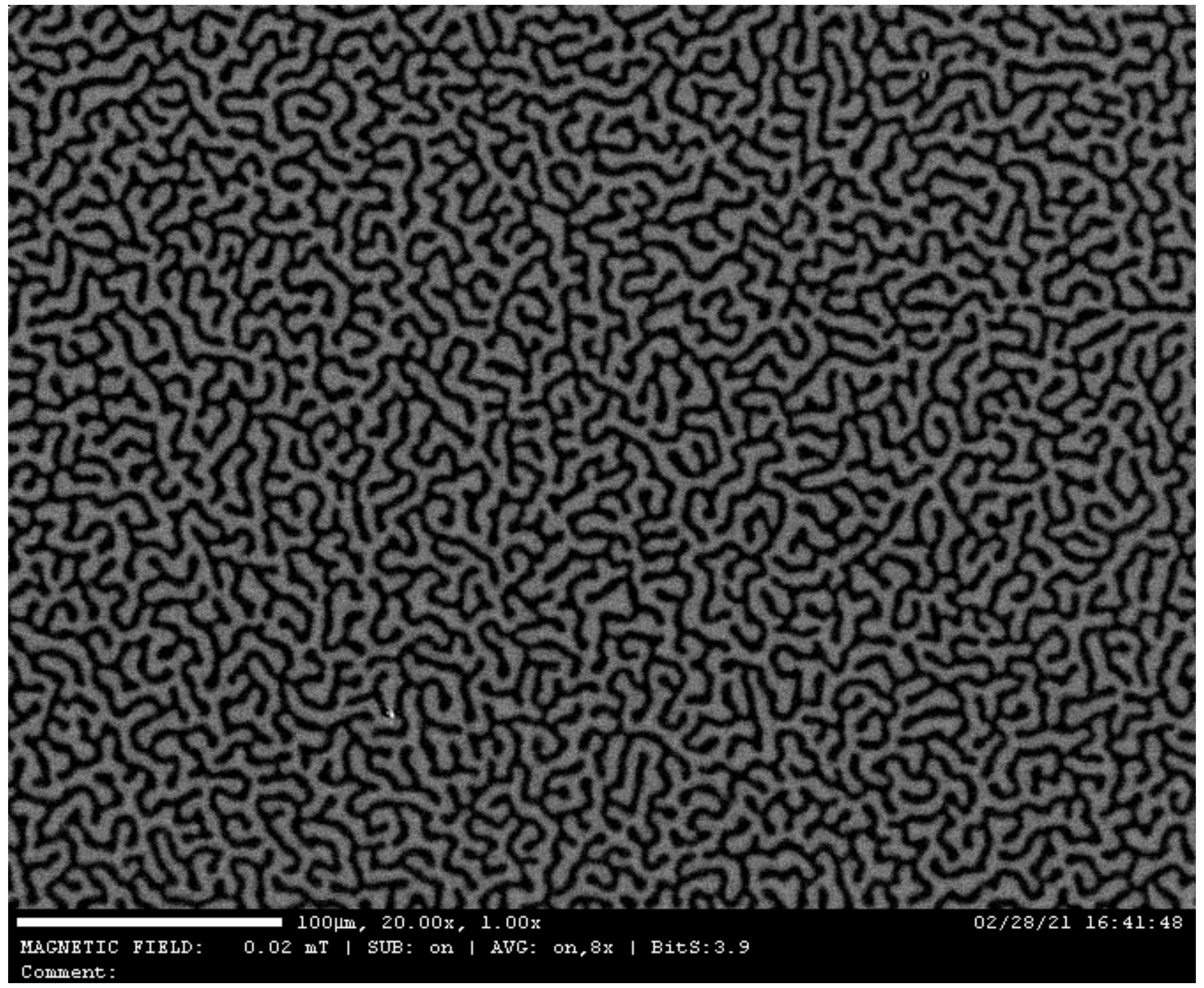
In preparation

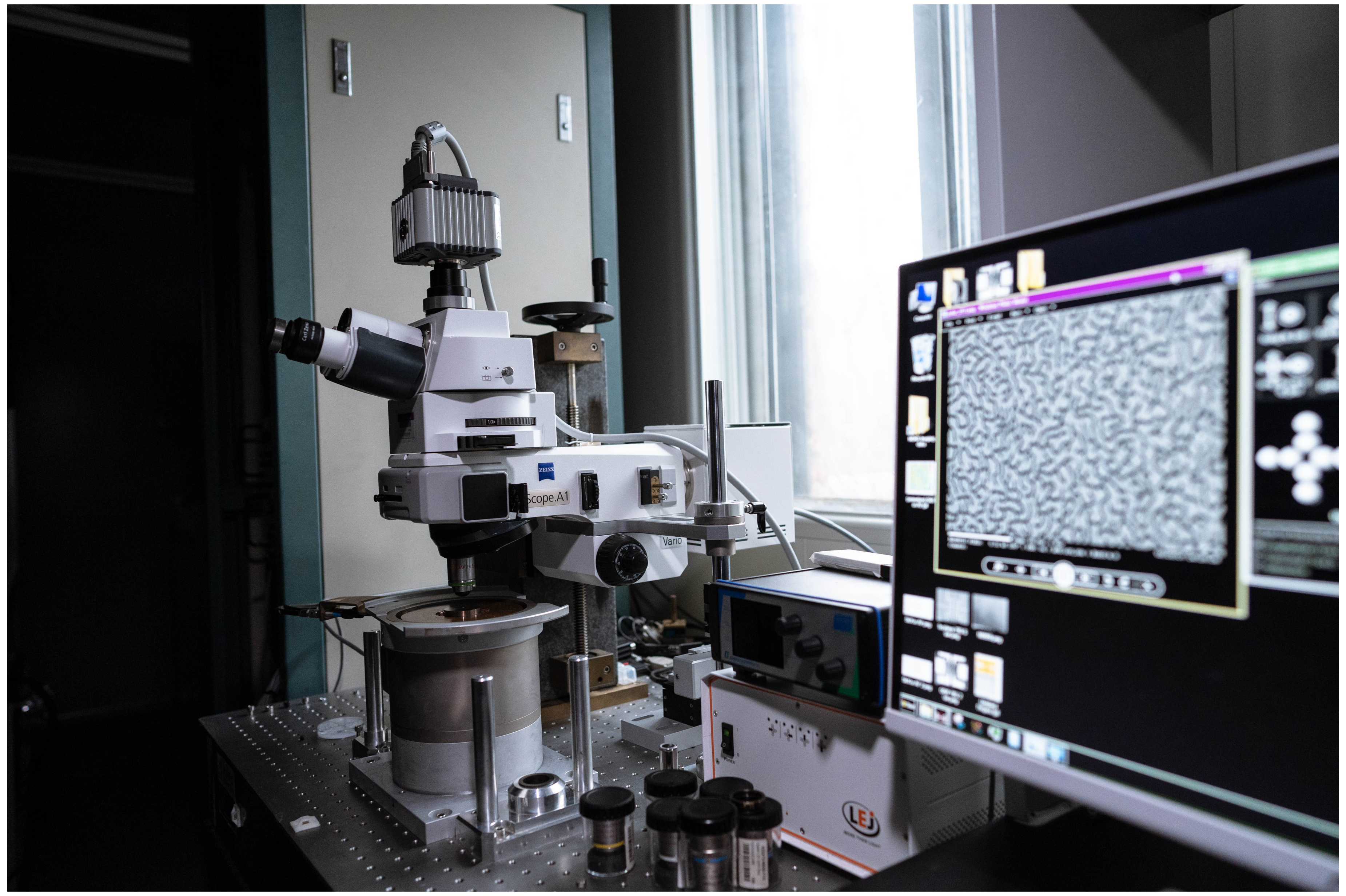


...

3000 images

Resolution: 0.5 μm





MOKE microscope@THU

PhysRevLett.125.027206

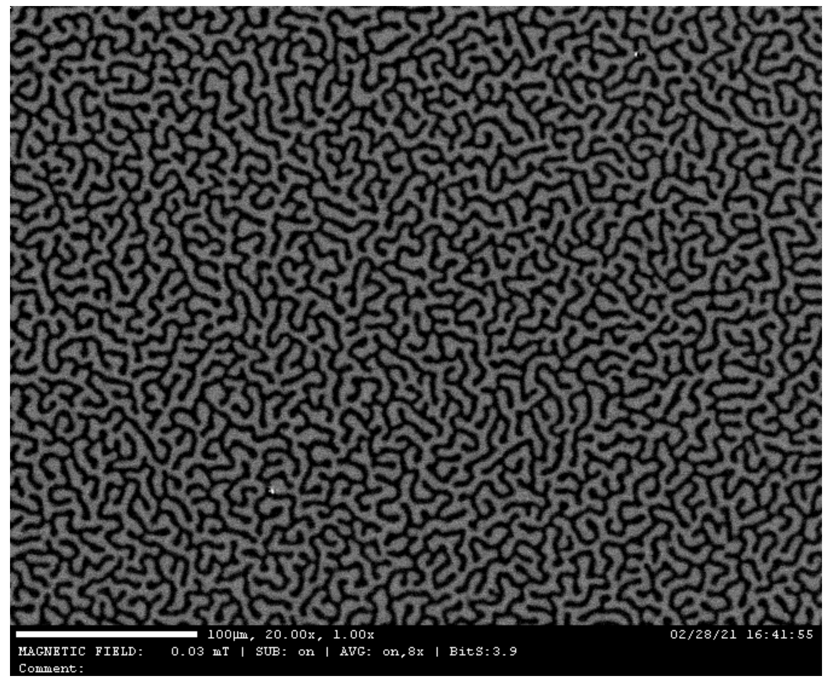
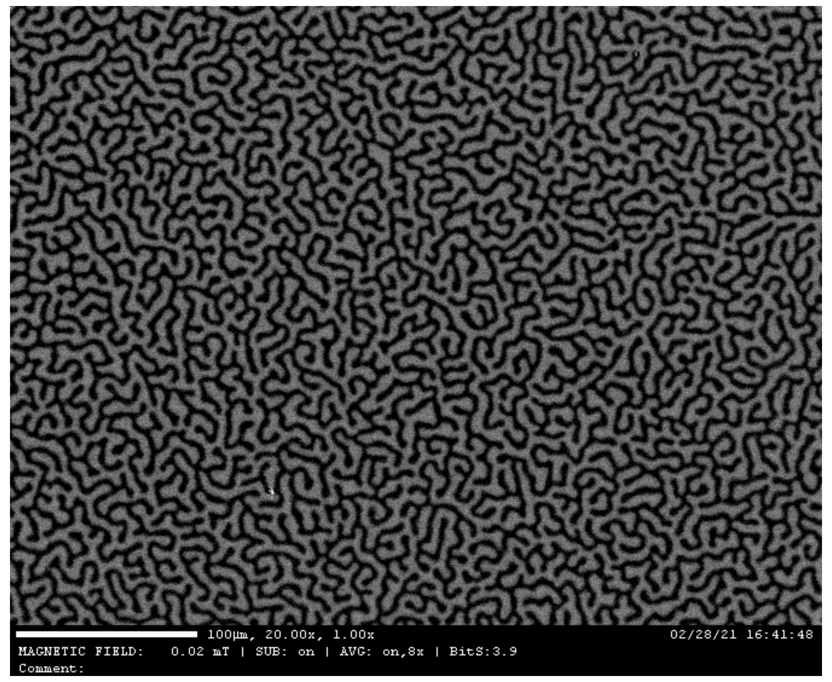
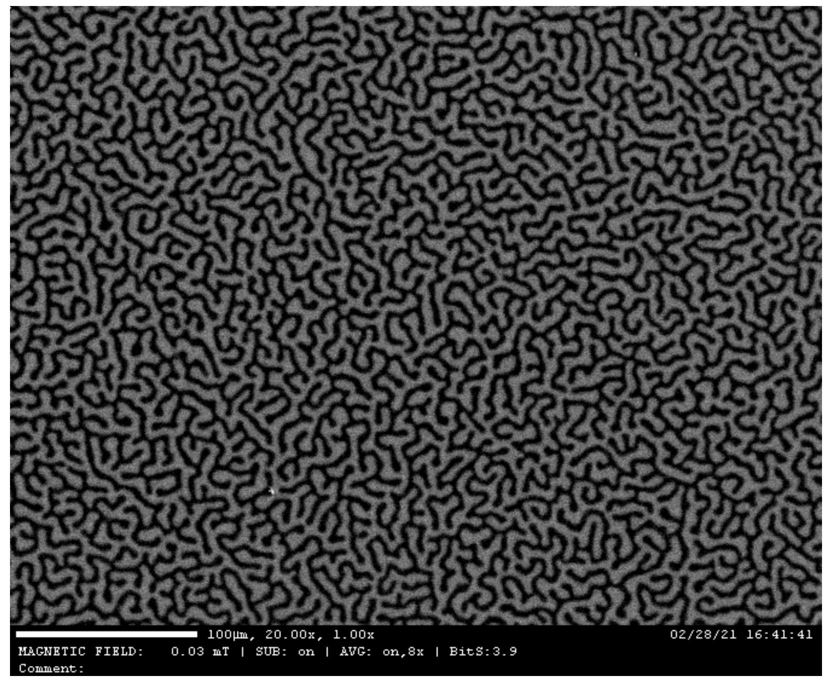
with Le Zhao and Wan-Jun Jiang

Magneto-optic Kerr effect (MOKE) microscope to capture images
for the **magnetic domains** appearing inside a **Ta/CoFeB/TaO_x thin film**
at **room temperature T = 296 K**

Ferromagnetic Phase Transitions

2. Ferromagnetic Materials

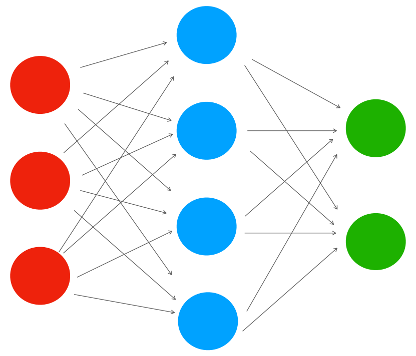
In preparation

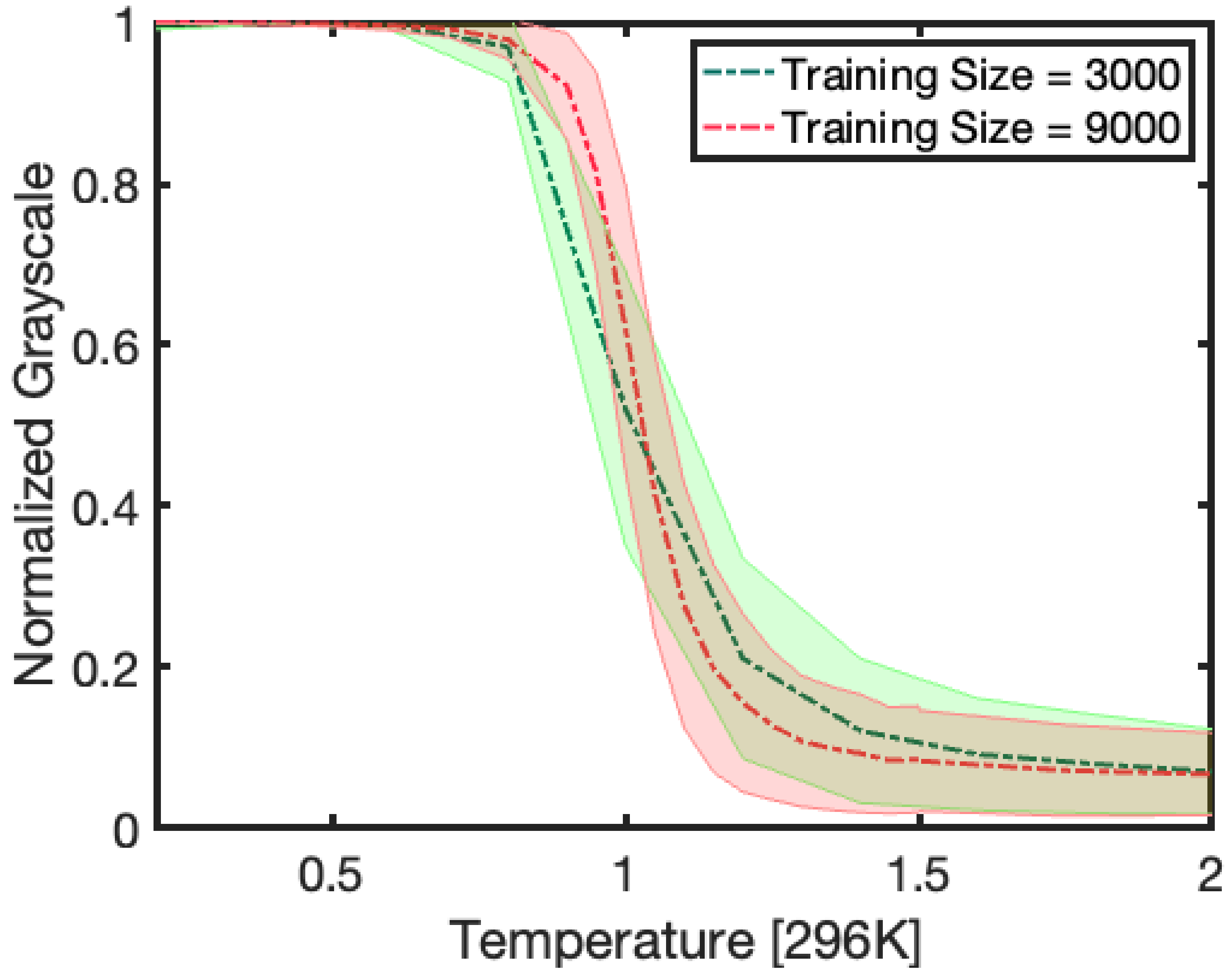
PhysRevLett.125.027206

...

3000 images



$$p_\theta(\mathbf{s}) = \frac{e^{-\beta H_\theta(\mathbf{s})}}{Z}$$



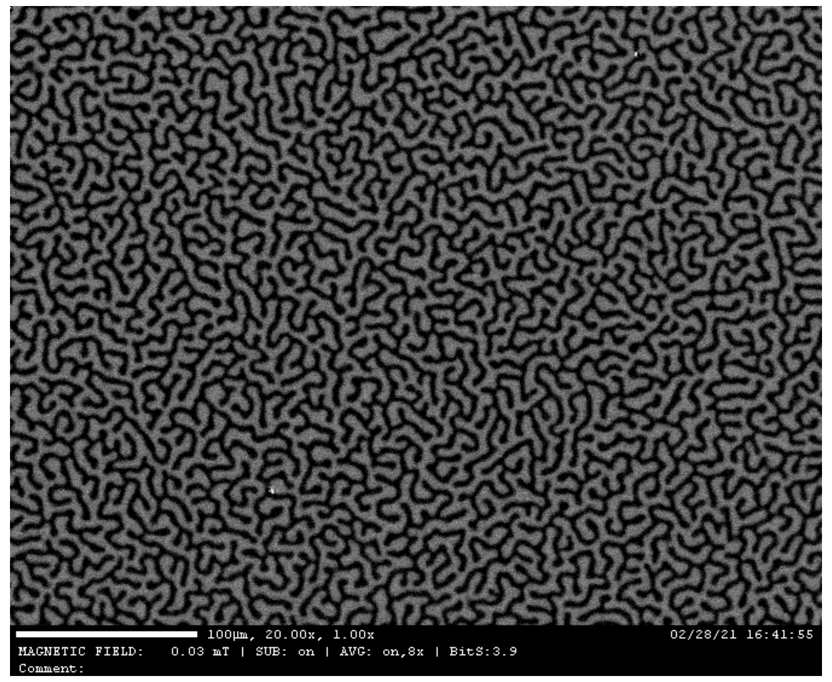
Magneto-optic Kerr effect (MOKE) microscope to capture images for the **magnetic domains** appearing inside a **Ta/CoFeB/TaO_x thin film** at **room temperature T = 296 K**

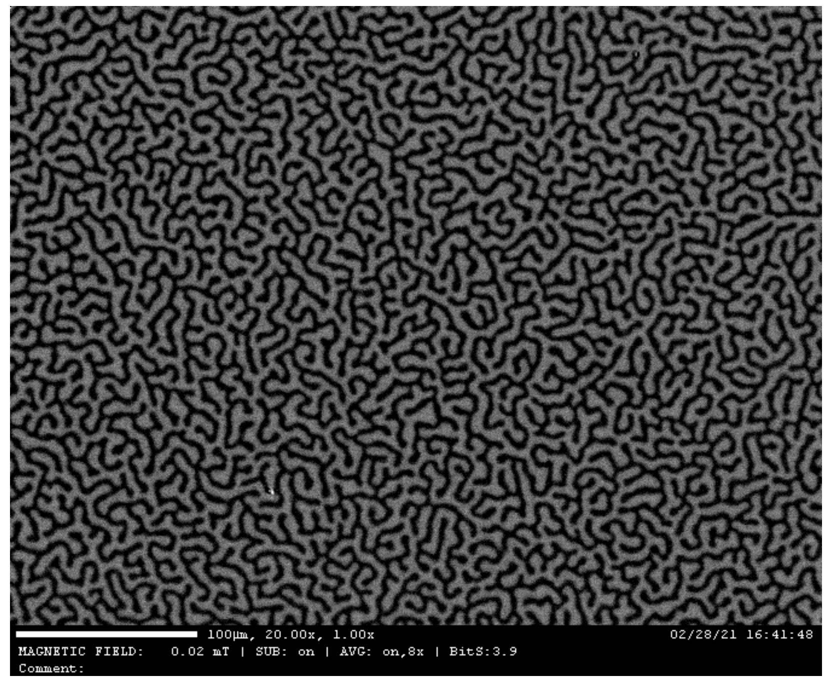
with Le Zhao and Wan-Jun Jiang

Ferromagnetic Phase Transitions

2. Ferromagnetic Materials

In preparation





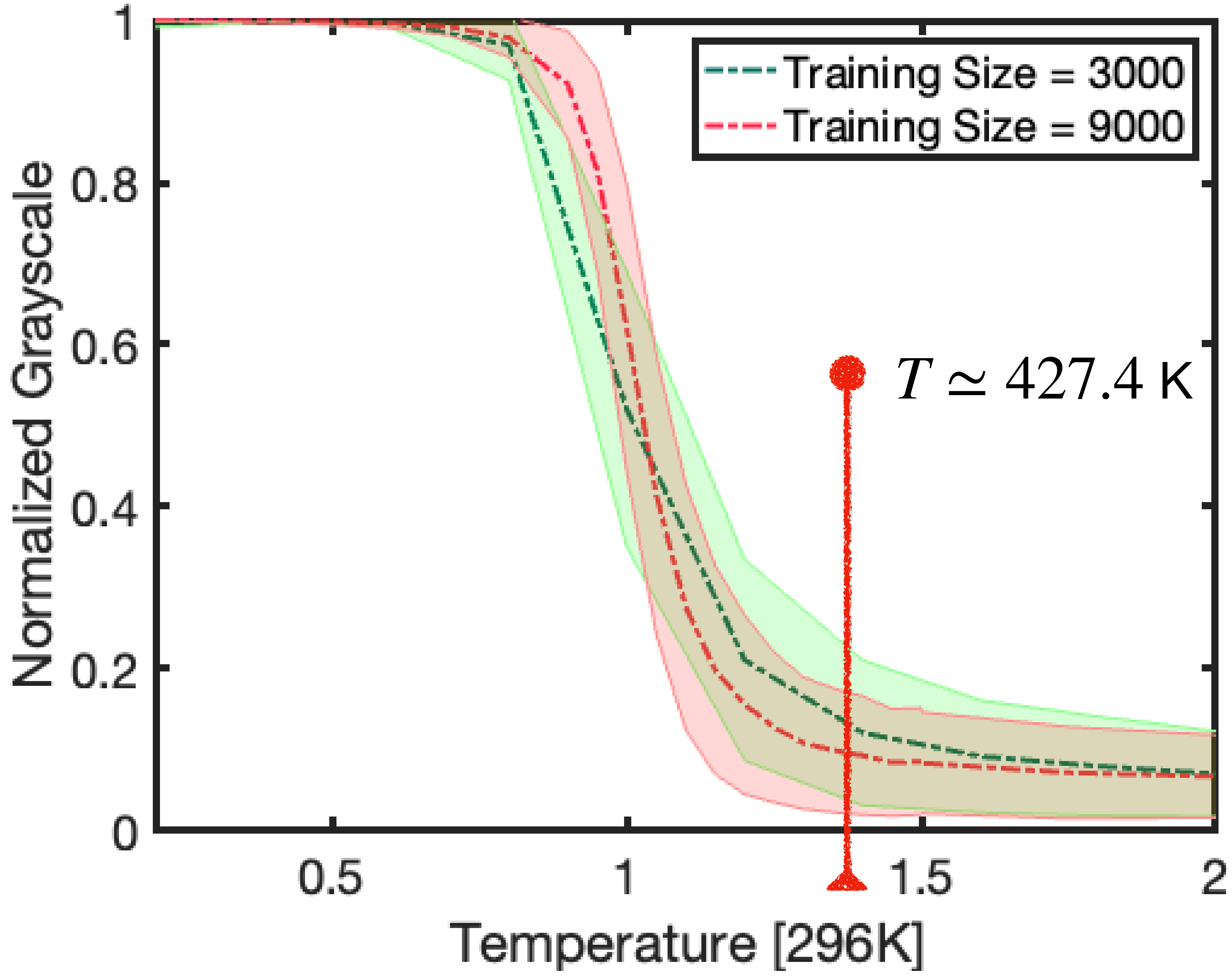
PhysRevLett.125.027206

3000 images

T [K]	50	100	150	200	250	300
M_s [emu/cc]	961.16	925.25	877.73	818.17	753.31	673.427

$$M_s(T) = M_s(0)(1 - T/T_C)^{1/3}$$

$$T_C = 427.4 \pm 2.9 \text{ K}$$

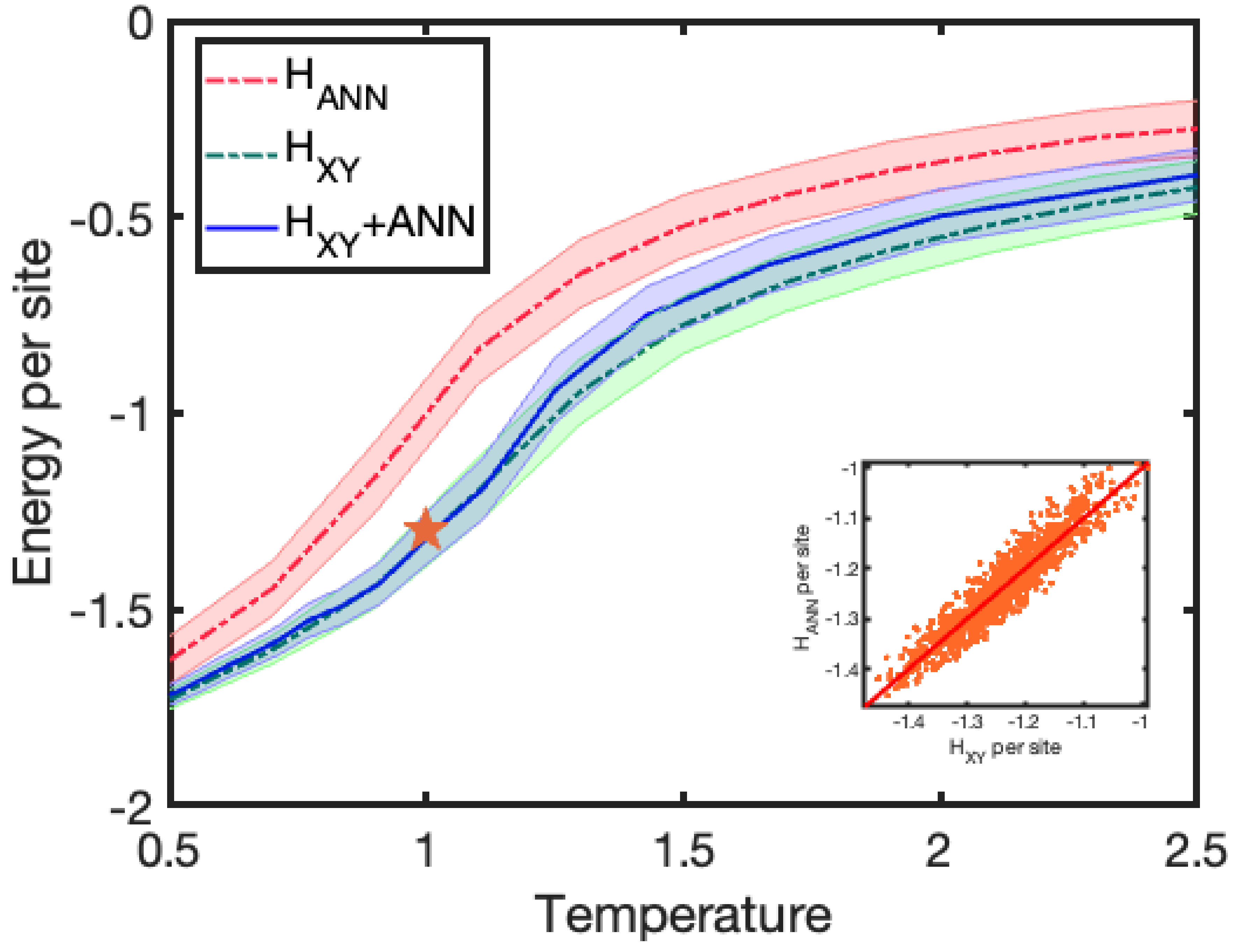


Topological Phase Transitions

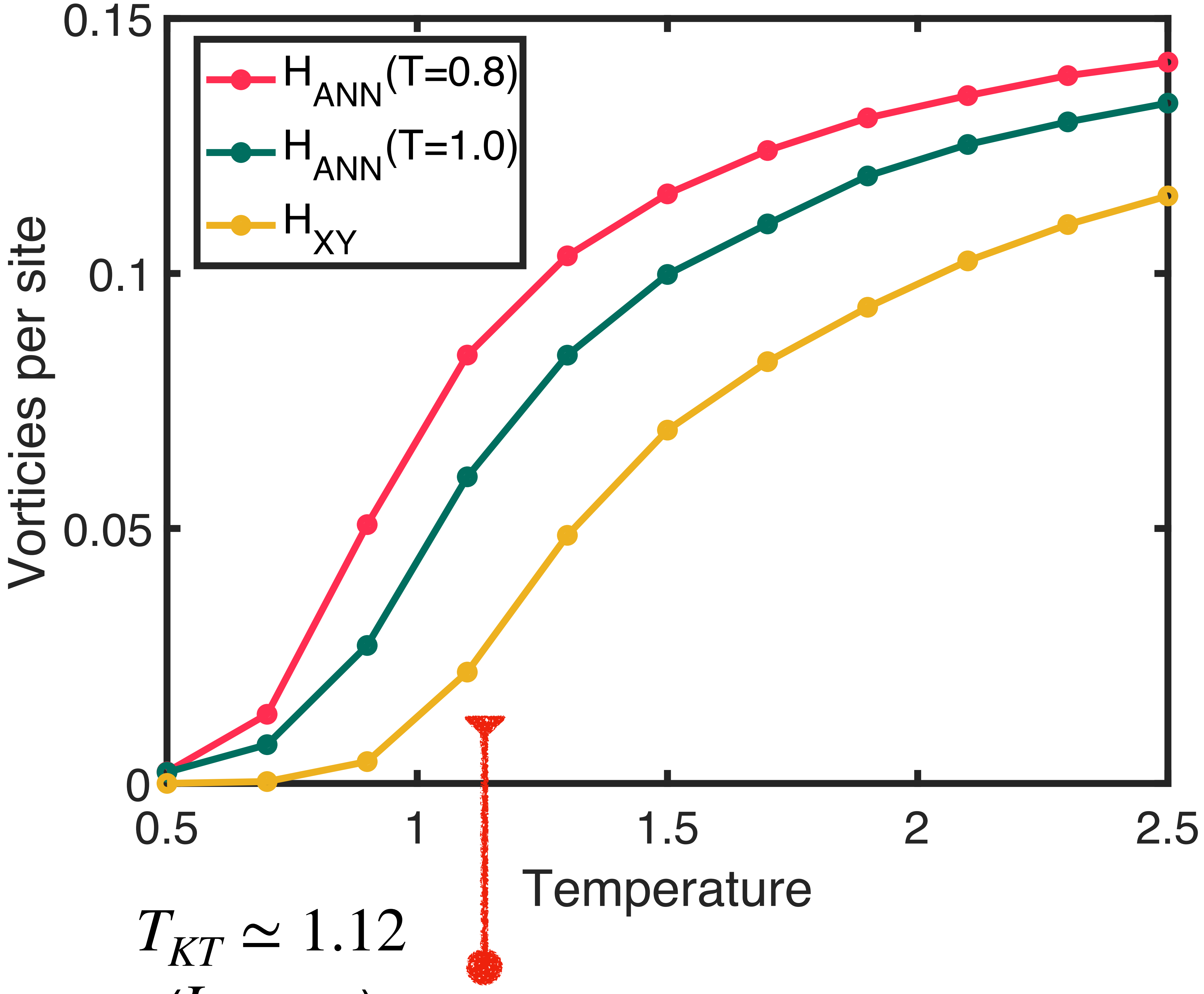
3. 2D XY Model

In preparation

$$H_{\rm XY} = -\sum_{<i,j>} s_i s_j = -\sum_{<i,j>} \cos(\phi_i - \phi_j)$$



9000 training configurations with L = 16

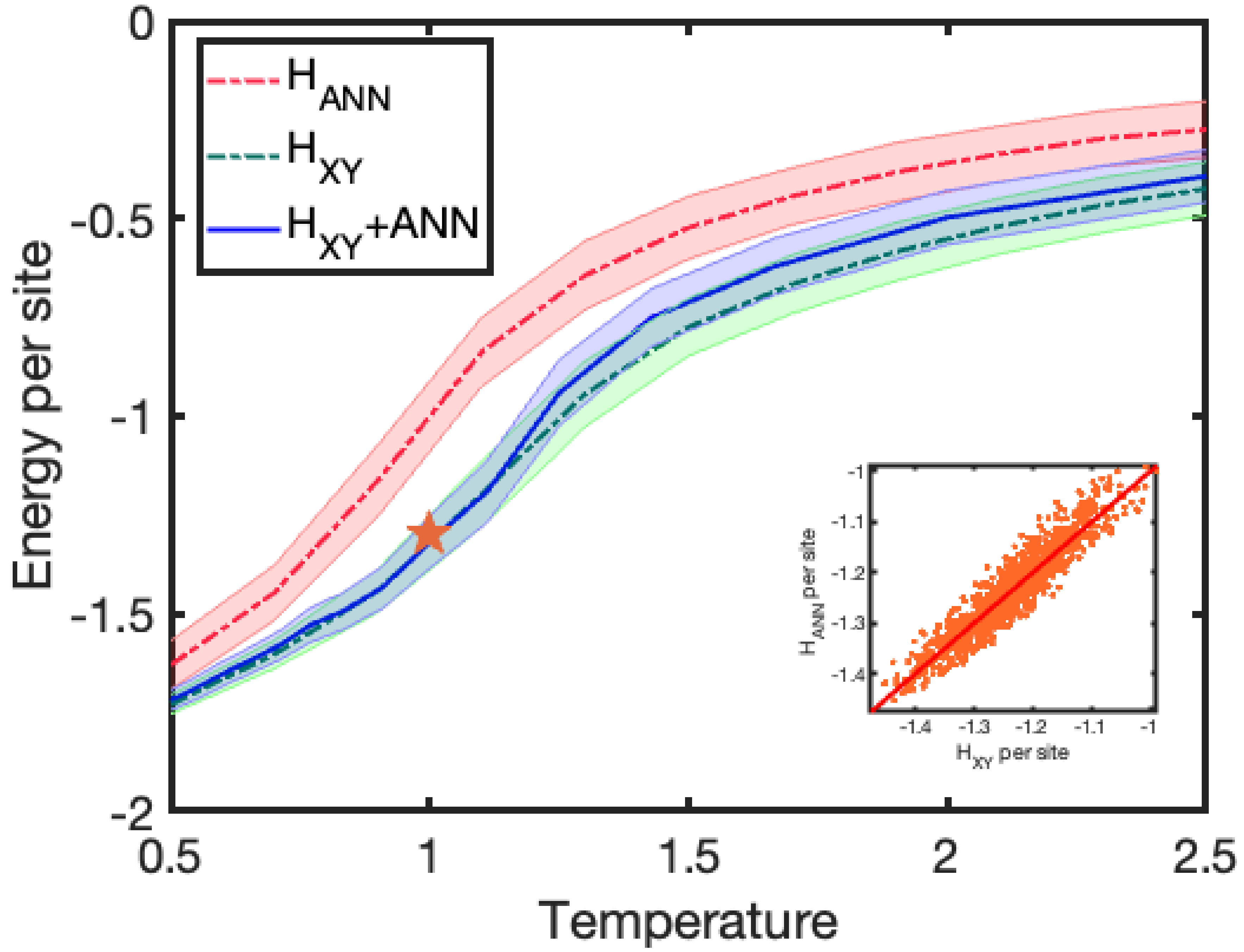


$T_{KT} \simeq 1.12$
$(L \to \infty)$

Topological Phase Transitions

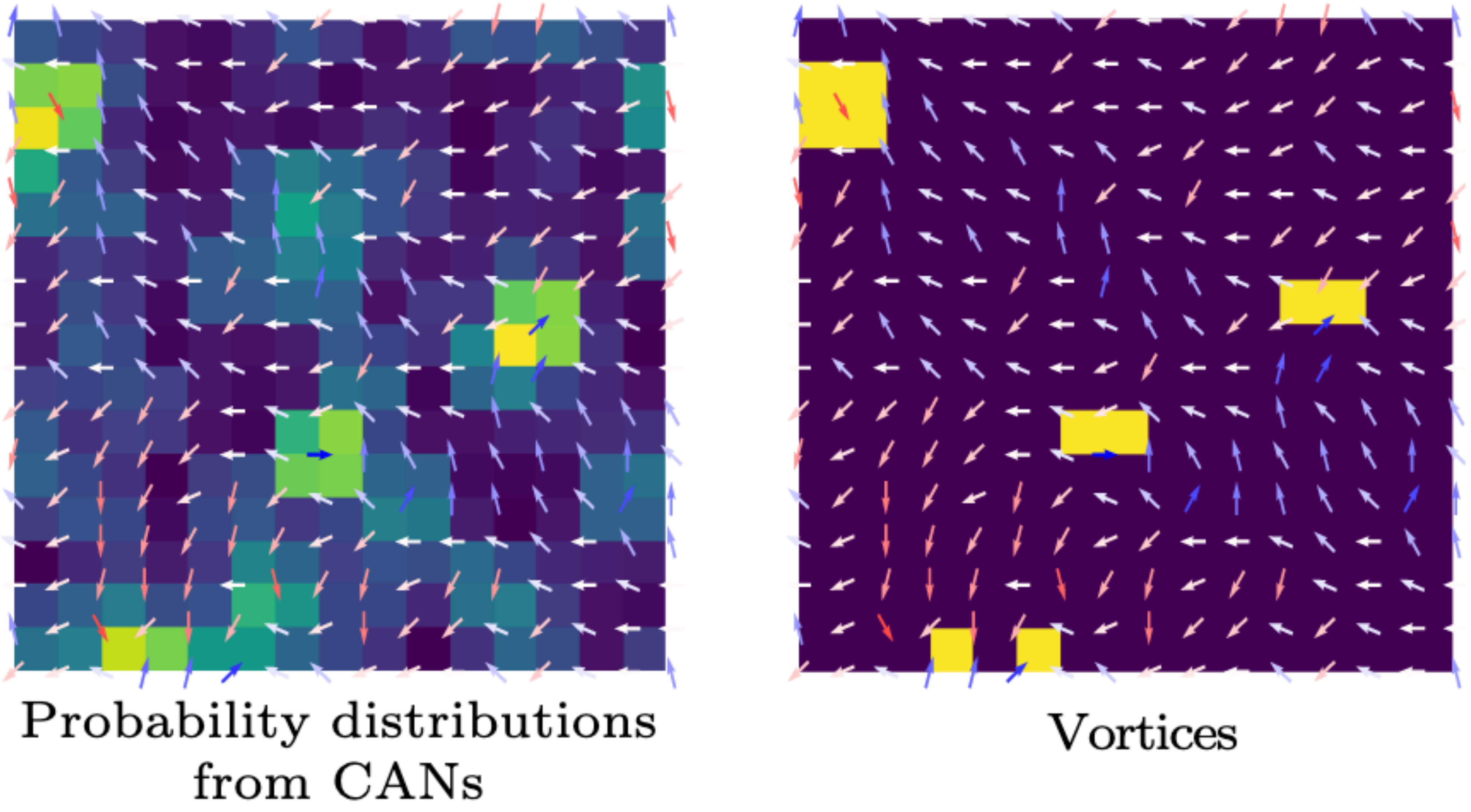
3. 2D XY Model

$$H_{\rm XY} = -\sum_{<i,j>} s_i s_j = -\sum_{<i,j>} \cos(\phi_i - \phi_j)$$



9000 training configurations with L = 16

Chinese Phys. Lett. 39, 120502 (2022)



$$\delta q_{[i,j]} \equiv \sum_{[i,j]} q_\theta(s_{ij})$$

$$v = (1/2\pi) \oint_C \nabla\phi(\boldsymbol{r}) \cdot d\boldsymbol{r}$$

$q_\theta(s_{ij})$ **conditional probability differences in the same given direction**

Can It Still Work for Quantum Systems?

Finite-Temperature Fields

$$Z = \int D\Phi \exp(-S[\Phi])$$

$$S[\Phi] = \sum \Delta\tau(\Delta x)^3[(\frac{\Delta\Phi}{\Delta\tau})^2 + (\nabla\Phi)^2 + V(\Phi)$$

$$= \sum (\Delta x)^3 \left[\frac{(\Delta\Phi)^2}{\Delta\tau} + \Delta\tau((\nabla\Phi)^2 + V(\Phi)) \right]$$

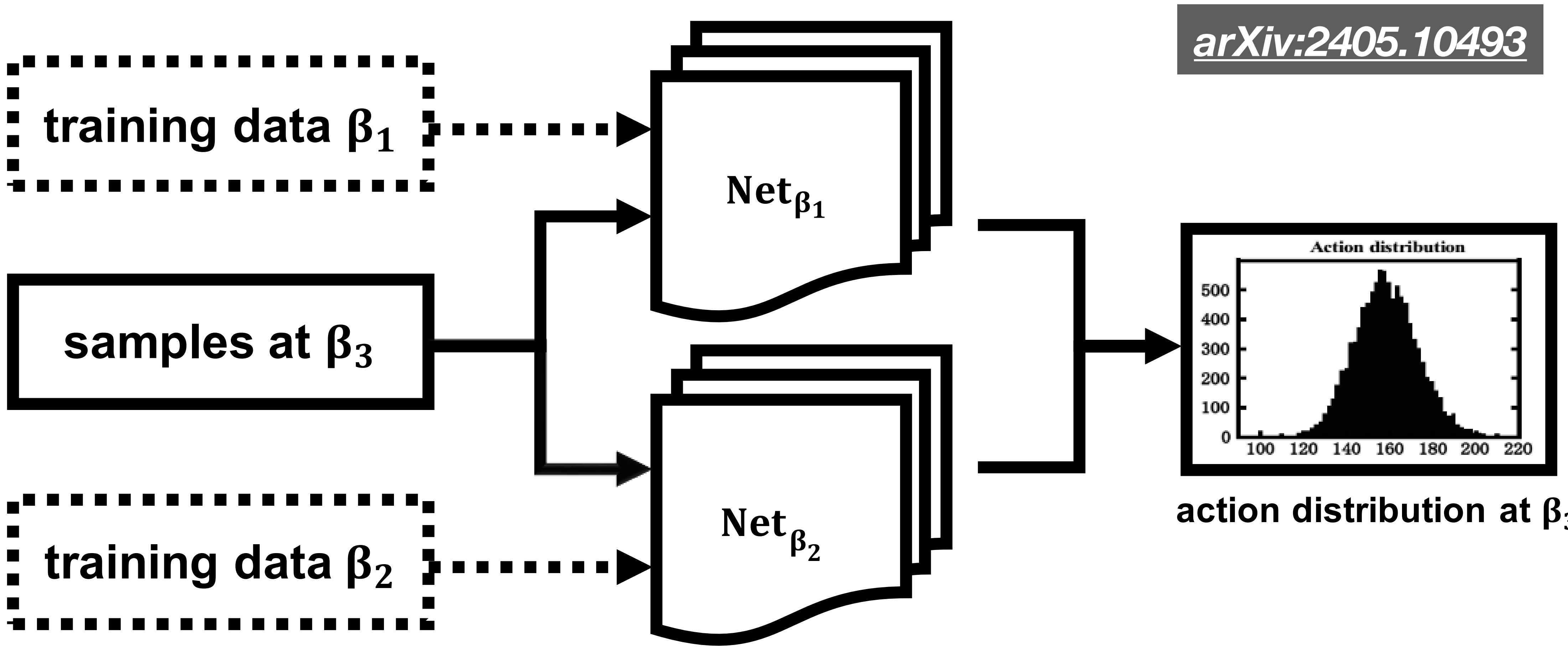
$$= \beta^{-1}K + \beta V$$

$$\Delta\tau = \beta/N_\tau$$

$$K \equiv N_\tau \sum (\Delta x)^3(\Delta\Phi)^2$$

$$V \equiv N_\tau^{-1} \sum (\Delta x)^3[(\nabla\Phi)^2 + V(\Phi)]$$

arXiv:2405.10493



action distribution at β_3

$$S_1[\Phi] = \beta_1^{-1}K[\Phi] + \beta_1 V[\Phi] + C_1$$

$$S_2[\Phi] = \beta_2^{-1}K[\Phi] + \beta_2 V[\Phi] + C_2,$$

$$S_3[\Phi] = \frac{\beta_1(\beta_3^2 - \beta_2^2)}{\beta_3(\beta_1^2 - \beta_2^2)}S_1 + \frac{\beta_2(\beta_1^2 - \beta_3^2)}{\beta_3(\beta_1^2 - \beta_2^2)}S_2 + C_3$$

$$C_3 = \frac{\beta_1(\beta_2^2 - \beta_3^2)}{\beta_3(\beta_1^2 - \beta_2^2)}C_1 + \frac{\beta_2(\beta_3^2 - \beta_1^2)}{\beta_3(\beta_1^2 - \beta_2^2)}C_2$$

Action Estimation

4. 0+1 D Quantum Field

arXiv:2405.10493

$$\mathscr{L} = \frac{1}{2}\left(\frac{dx}{d\tau}\right)^2 + V_k(x) \qquad V_k(x) = \frac{\lambda_k}{4}\left(x^2 - \frac{\mu_k^2}{2k}\right)^2$$

$$Z = \int_{x(\beta)=x(0)} Dx\ e^{-S_E[x(\tau)]}$$

$$= \int \prod_{j=-N+1}^{N+1} \frac{dx_j}{\sqrt{2\pi a}} \times \exp\left\{-\sum_{i=-N+1}^{N+1}\left[\frac{(x_{i+1}-x_i)^2}{2a} + aV_k(x_i)\right]\right\}$$

$$S_E[x(\tau)] = \int_0^\beta d\tau\ \mathscr{L}_E[x(\tau)] = \int_0^\beta d\tau\ \left[\frac{1}{2}\left(\frac{dx}{d\tau}\right)^2 + V_k(x)\right]$$

$$x(\tau) = \pm\frac{\mu_k}{\sqrt{\lambda_k}}\tanh\left[\frac{\mu_k}{\sqrt{2}}(\tau-\tau_0)\right]$$

Kink/Anti-Kink solutions reach $\pm\mu_k/\sqrt{\lambda_k}$ at $\tau = \pm\infty$

Numerical Simulations

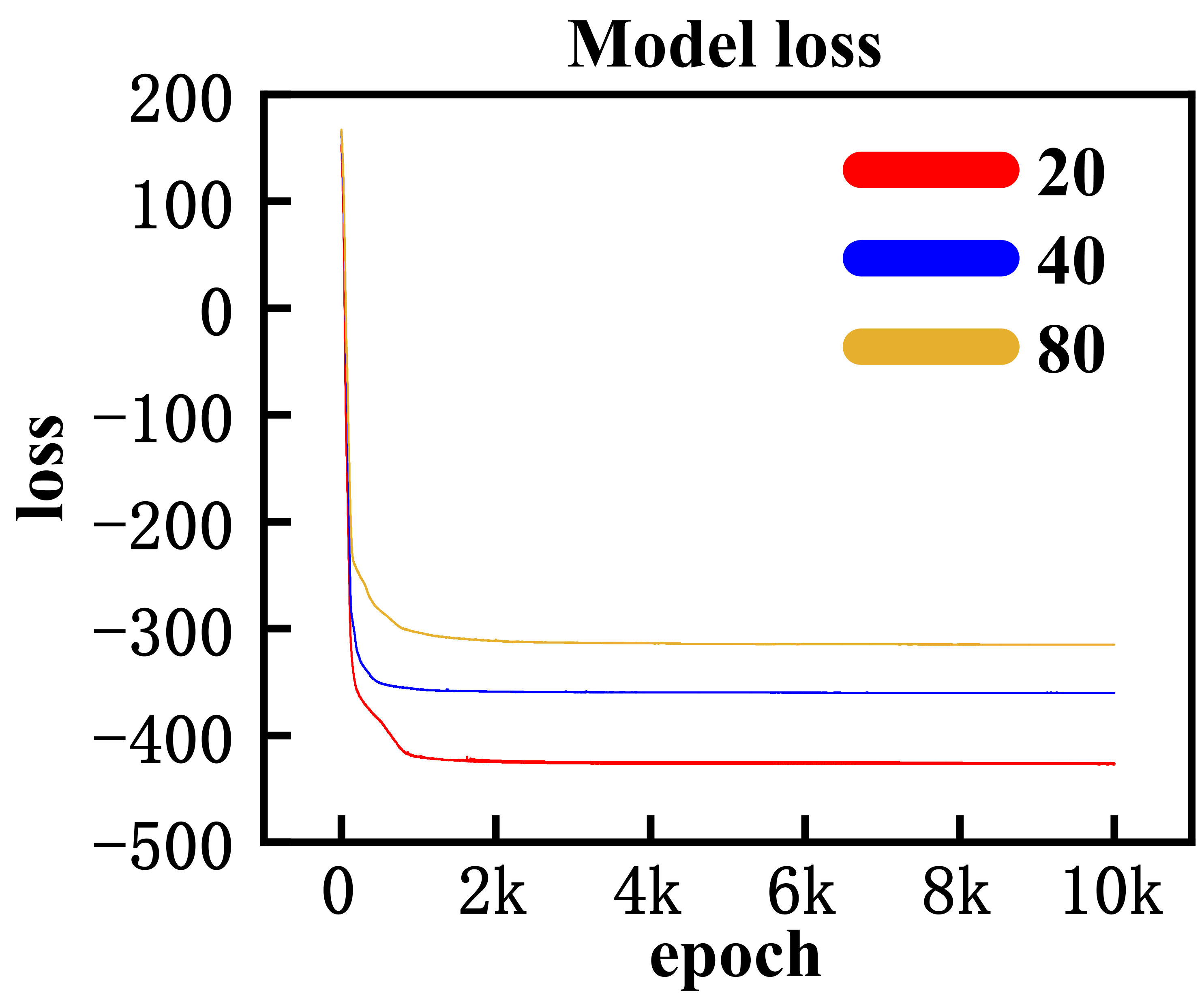
$$\lambda_k = 4 \qquad \mu_k/\sqrt{\lambda_k} = 1.4$$

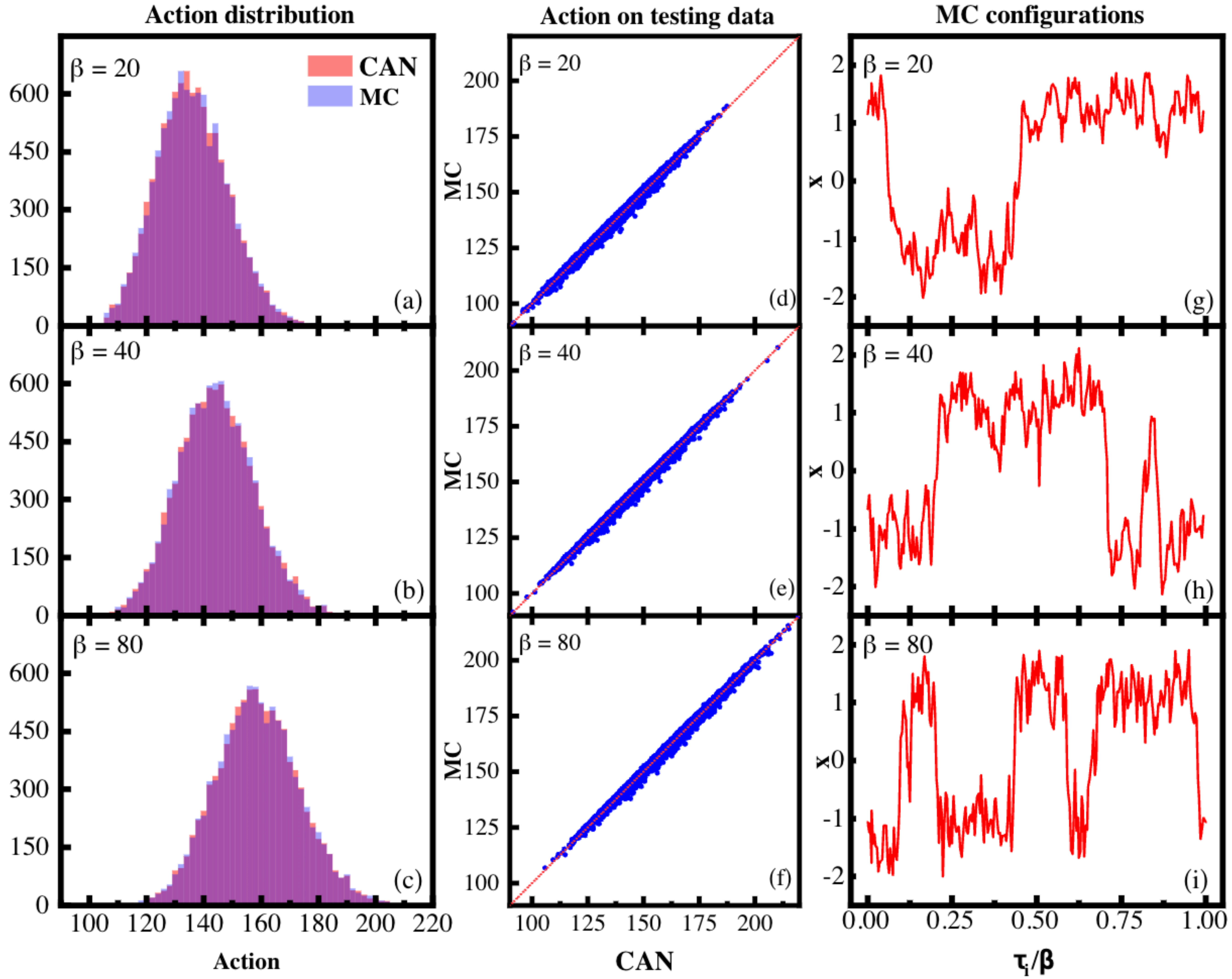
$$\beta = T^{-1} = 80,40,20 \qquad N_{MC} = 5\times 10^6$$

Action Estimation

4. 0+1 D Quantum Field

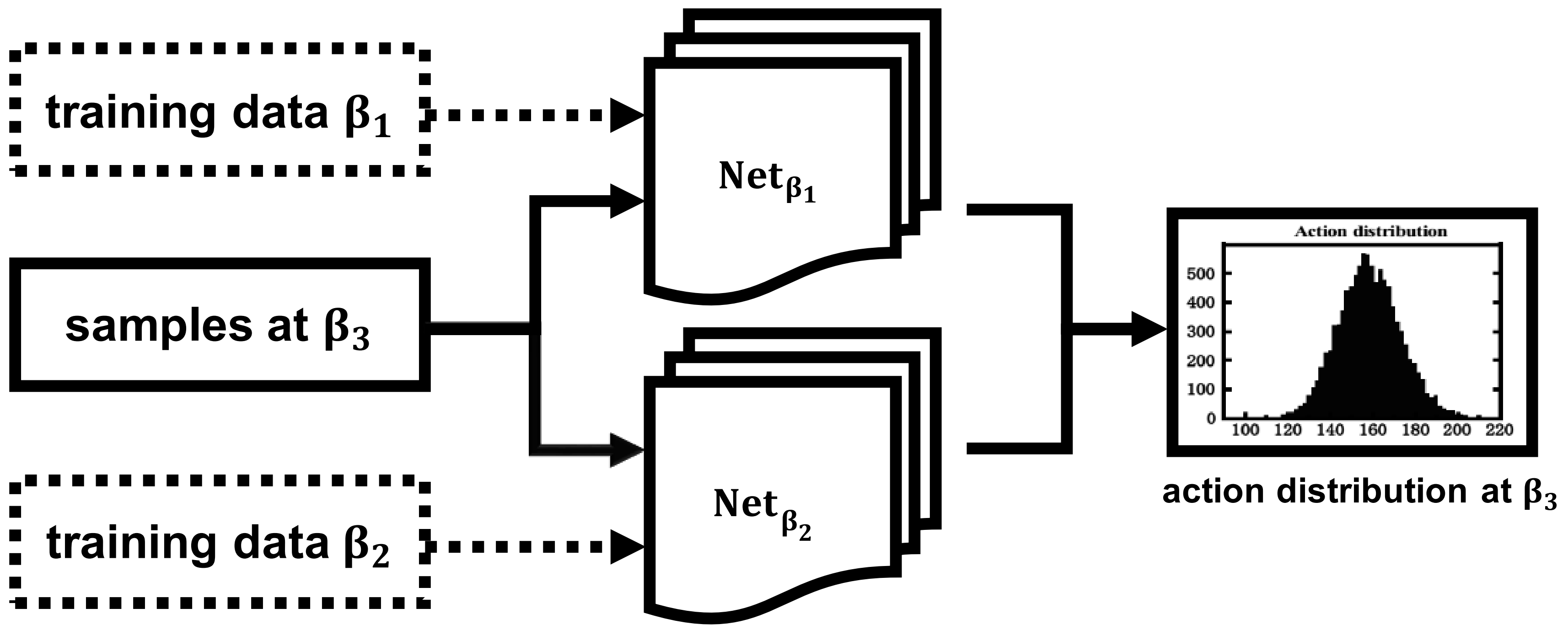
arXiv:2405.10493

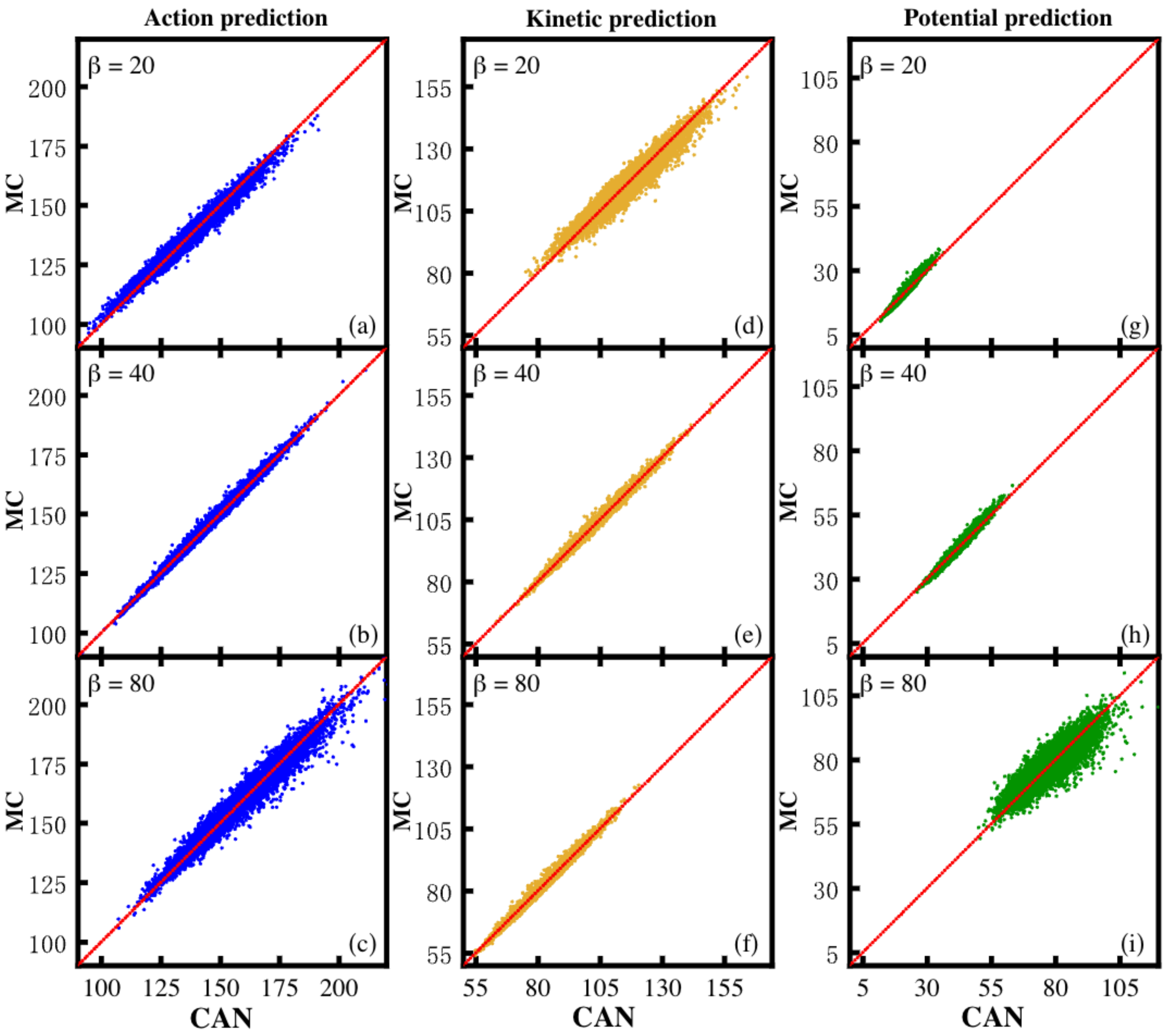




Action Prediction

4. 0+1 D Quantum Field

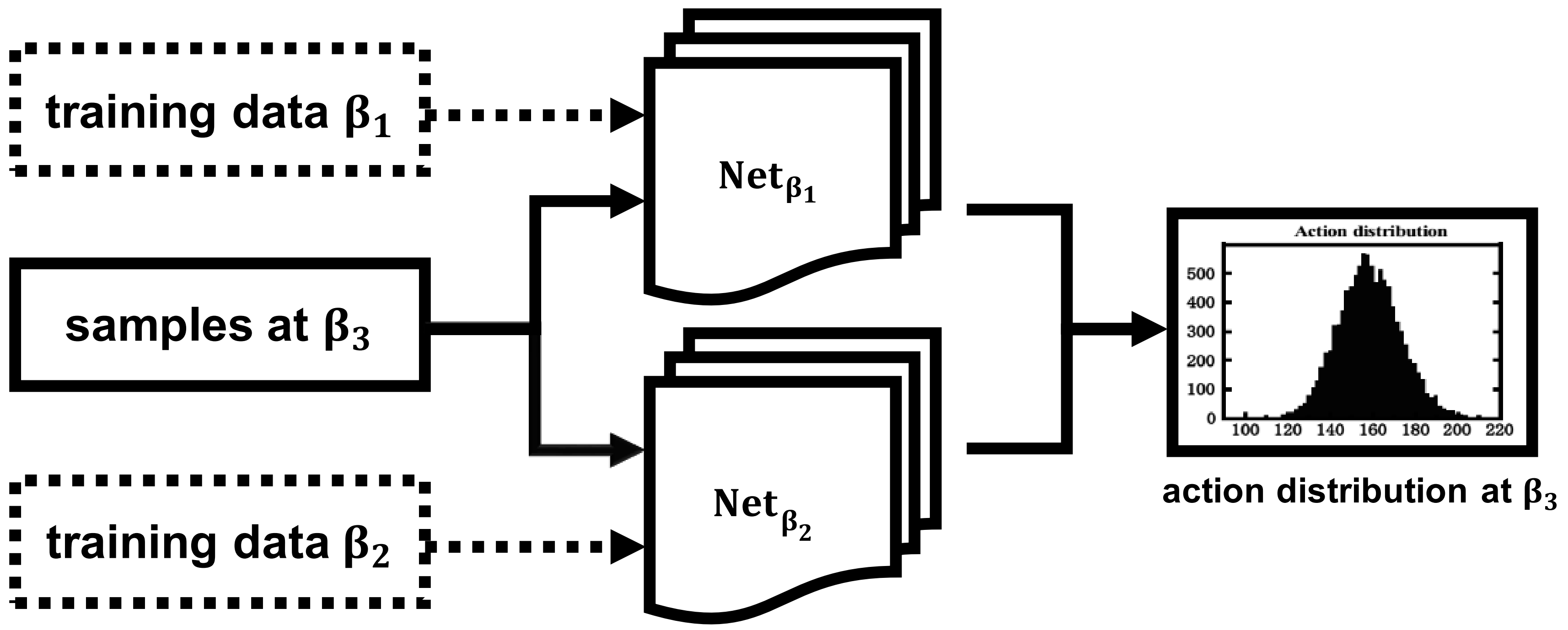




arXiv:2405.10493

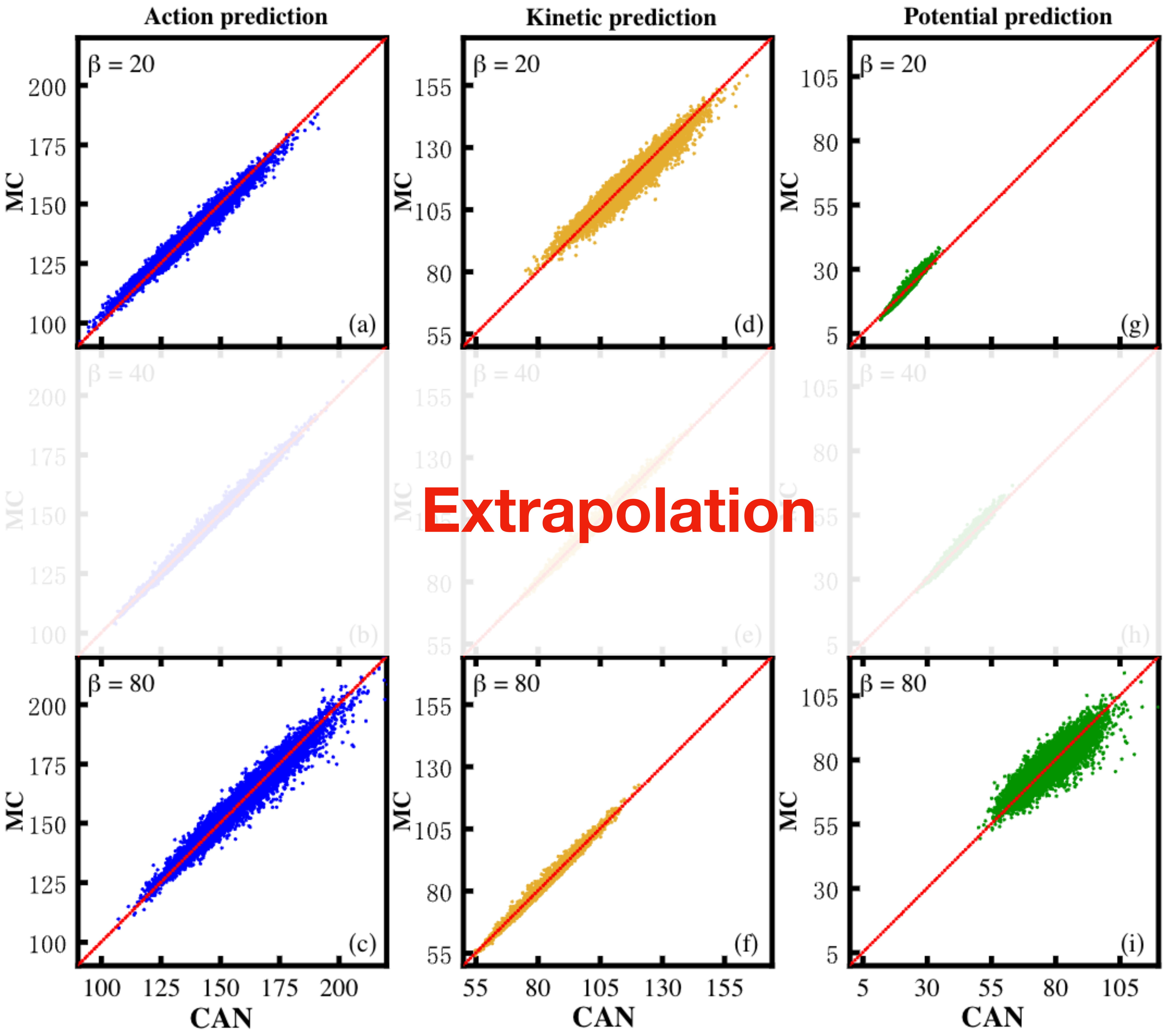
Action Prediction

4. 0+1 D Quantum Field



$\beta_1, \beta_2 = 40{,}80$

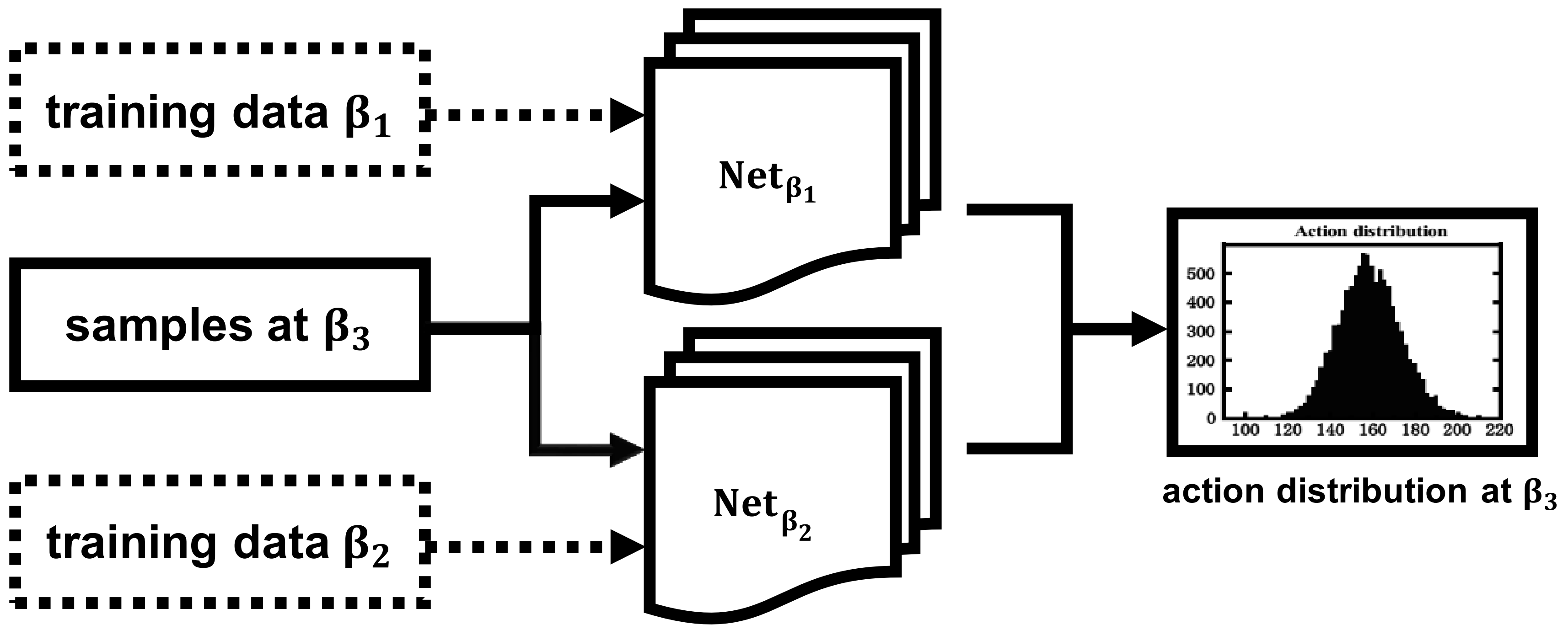
$\beta_1, \beta_2 = 20{,}40$

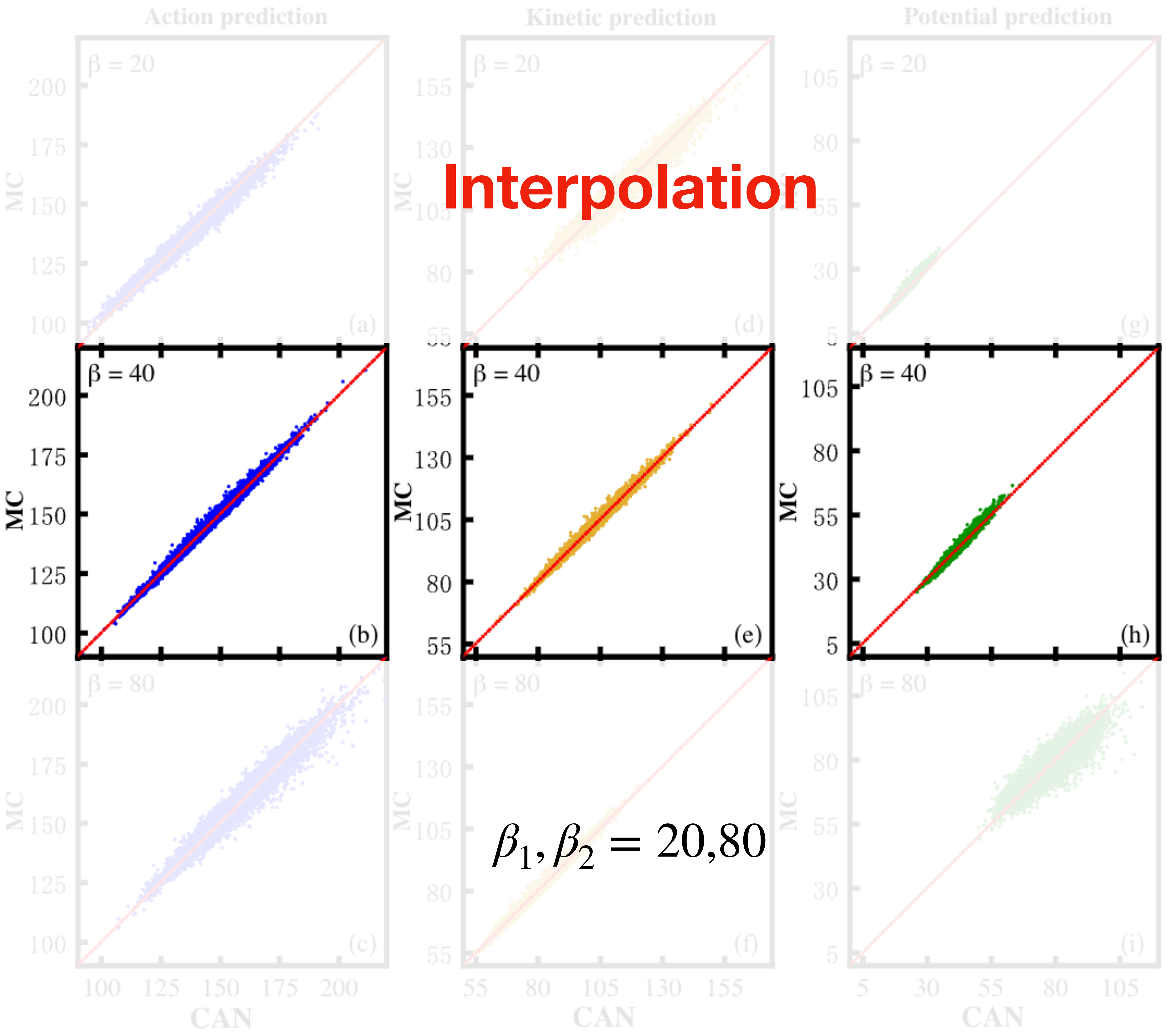


arXiv:2405.10493

Action Prediction

4. 0+1 D Quantum Field

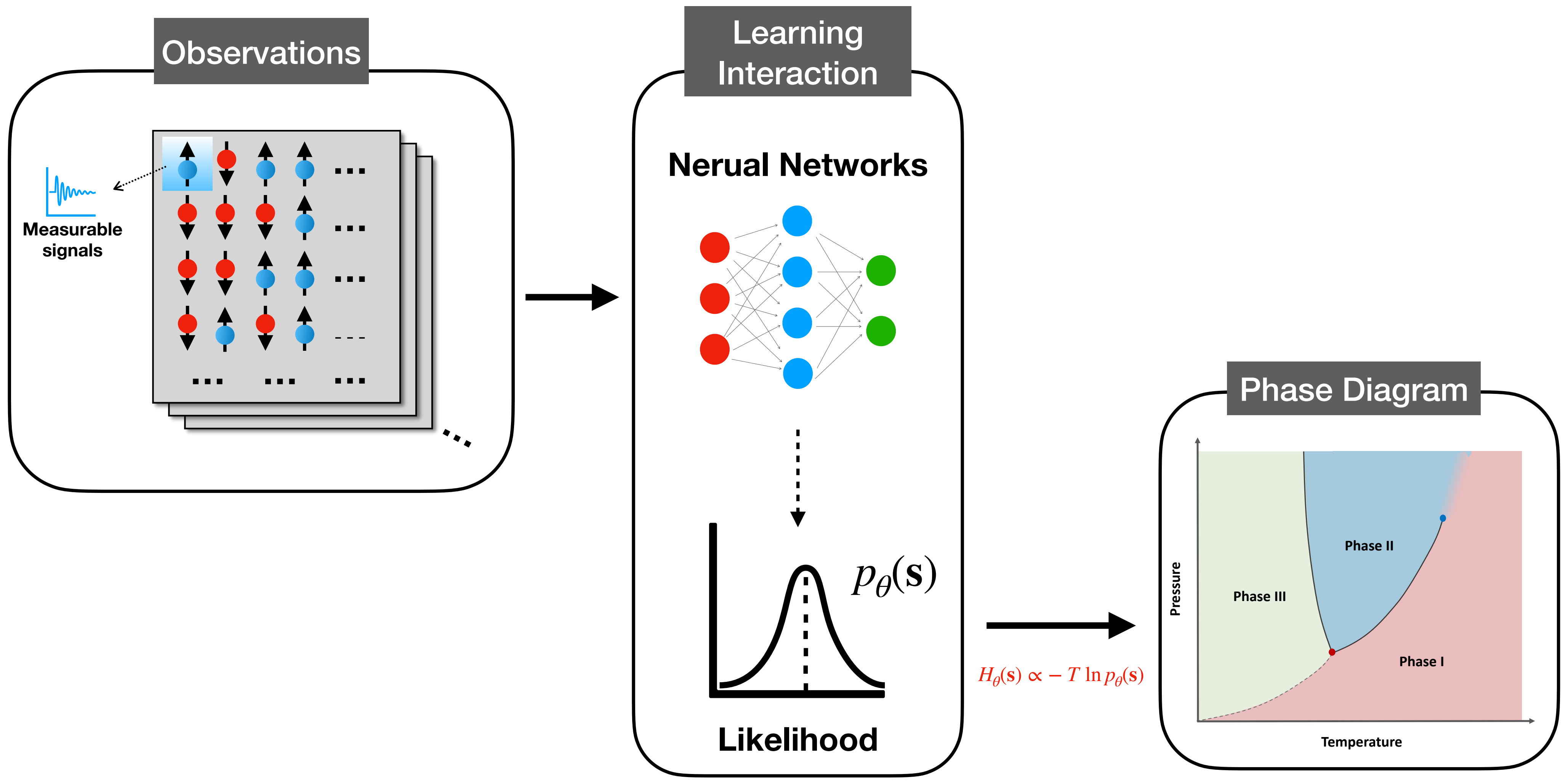


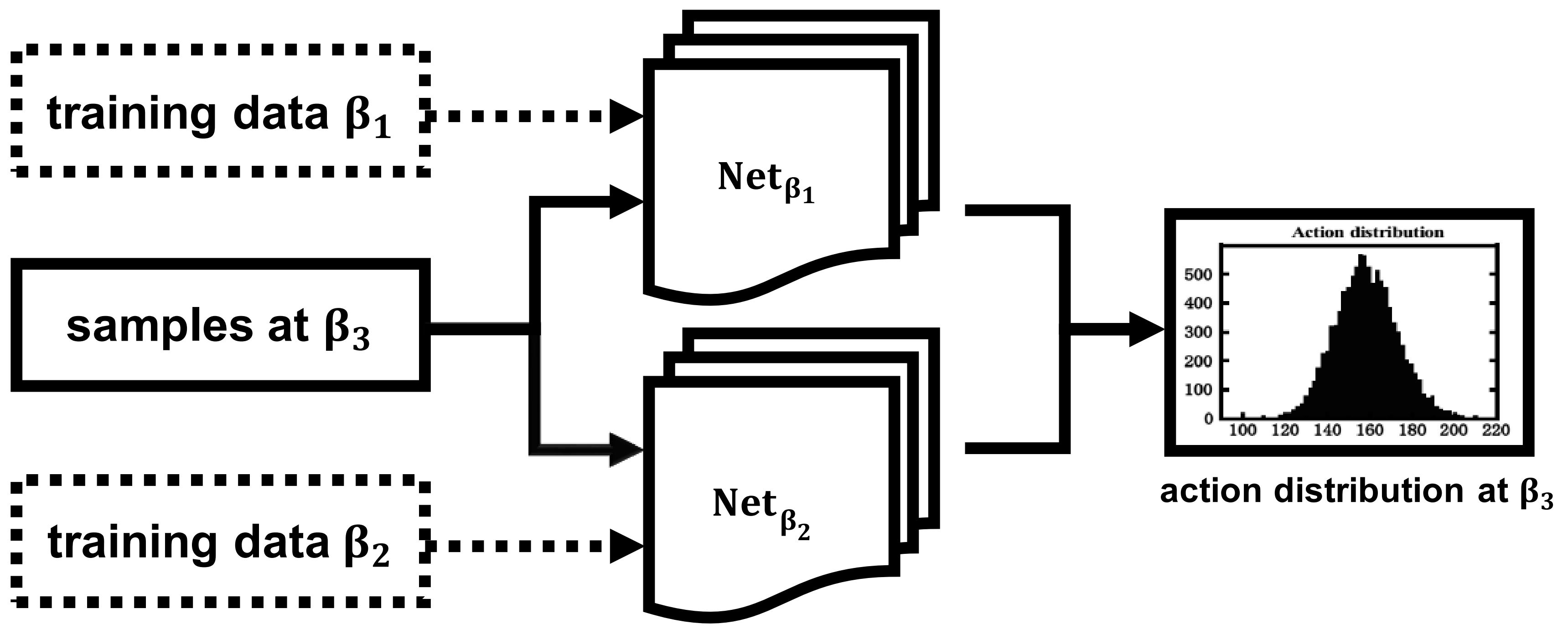


arXiv:2405.10493

Summary

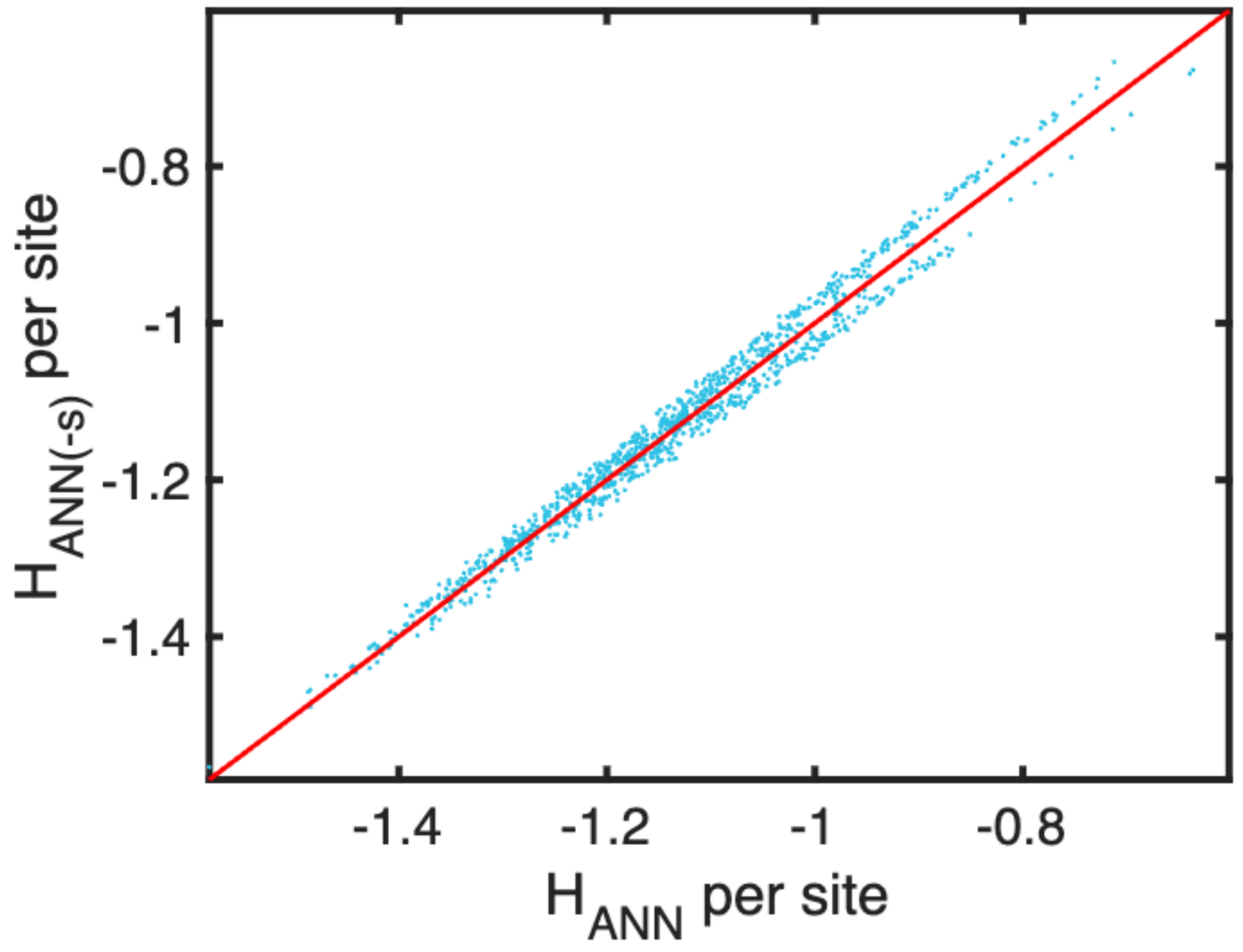
- **Building Interactions with DNNs**
  - Boltzmann Factor as Prior
  - Density Estimation
  - Autoregressive Networks
- **Applications**
  - Detecting Phase Transitions
    - Ferromagnetic PT
    - Topological PT
  - Estimating and Predicting Actions
    - 0+1 dimensional FTQT
- **Future works**
  - Finite-density quantum fields
  - Building effective interactions from macroscopic observations -> Renormalization Group ?!





Thank You !

Detecting Symmetries



Z2 flip for 2D Ising model

$$H_{ANN}(\hat{O}\mathbf{s}) \overset{?}{=} H_{ANN}(\mathbf{s})$$

θ [2π]	0	0.2	0.4	0.6	0.8	1
$H_{\mathrm{ANN}}(\theta\|\mathbf{s})$	-1.2397	-1.2394	-1.2406	-1.2413	-1.2417	-1.2397

O(2) rotation for 2D XY model

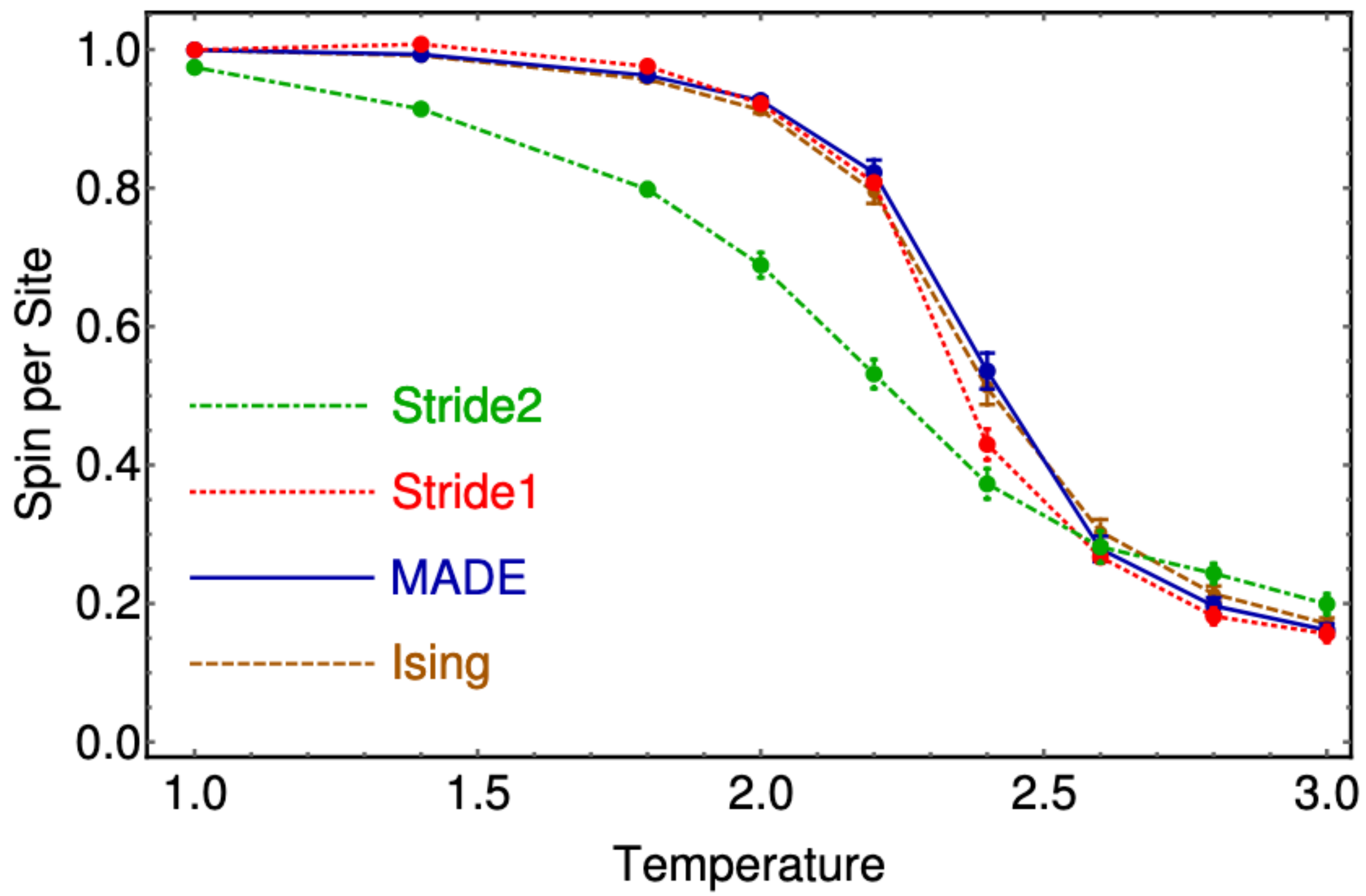
Virtual operations on observations to test learned neural network interactions

Backups - Ising Models

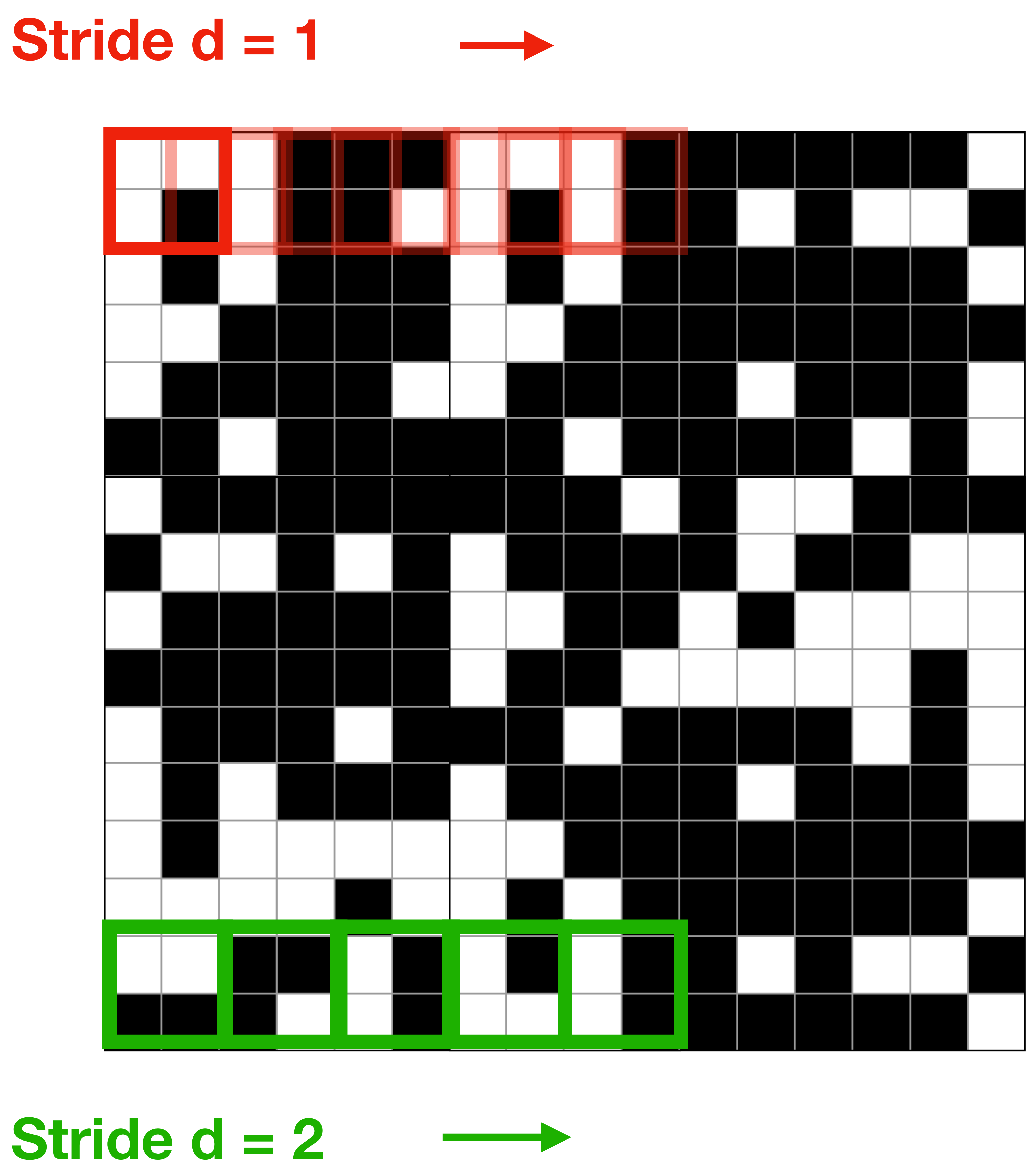
$$H(\mathbf{s}) = -\sum_{<i,j>} s_i s_j$$

$$S_{I,J} = s_{i,j} + s_{i+1,j} + s_{i,j+1} + s_{i+1,j+1}$$

$$i = 1 + d(I-1), j = 1 + d(J-1)$$



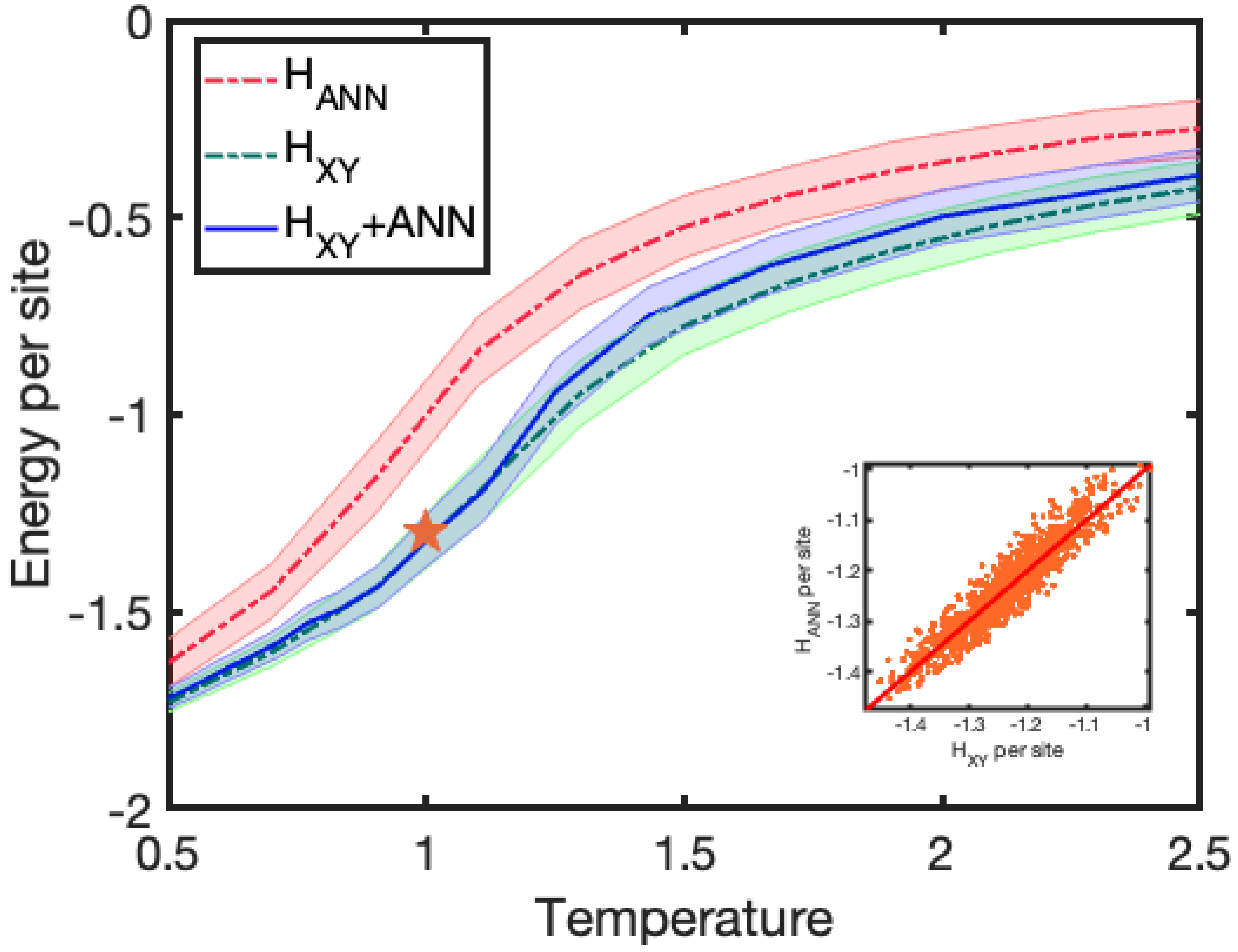
arXiv:2007.01037



60000 configurations

Backups - Important Sampling

$$H_{\rm XY} = -\sum_{<i,j>} s_i s_j = -\sum_{<i,j>} \cos(\phi_i - \phi_j)$$



9000 training configurations with L = 16

Chinese Phys. Lett. 39, 120502 (2022)

$$\langle O \rangle = \int O(\mathbf{s})(p(\mathbf{s})/q_\theta(\mathbf{s}))q_\theta(\mathbf{s})d\mathbf{s}$$

$$\langle O \rangle \simeq \sum_{\mathbf{s}\sim q_\theta(\mathbf{s})} O(\mathbf{s})w(\mathbf{s})$$

$$w(\mathbf{s}) = (p(\mathbf{s})/q_\theta(\mathbf{s}))/A$$

$$A = \sum_{\mathbf{s}\sim q_\theta(\mathbf{s})} p(\mathbf{s})/q_\theta(\mathbf{s})$$

Backups - CANs

Continious Spin Configurations

Chinese Phys. Lett. 39, 120502 (2022)

Autoregressive Networks model Likelihood $q_\theta(s)$ explicitly

$$q_\theta(s) \rightarrow p(s) = \frac{e^{-E(s)}}{Z}$$

Optimization **Minimizing loss function**

Variational free energy

$$F_q = \sum_s q_\theta(s)(E(s) + (ln q_\theta(s))/\beta)$$

Kullback-Leibler (KL) divergence

$$D_{KL}(q_\theta||p) = \sum_s q_\theta(s) ln(\frac{q_\theta(s)}{p(s)}) = \beta(F_q - F) \geq 0$$