Opportunities of AI

Gert Aarts

AIMLAC
UKRI CENTRE FOR DOCTORAL TRAINING IN ARTIFICIAL INTELLIGENCE, MACHINE LEARNING AND ADVANCED COMPUTING

Swansea University
Prifysgol Abertawe

FONDAZIONE BRUNO KESSLER

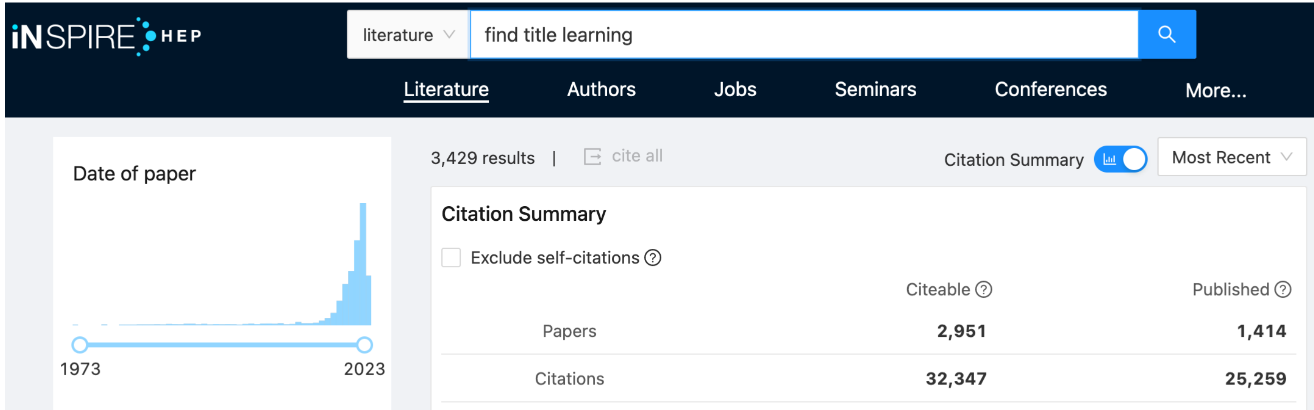
ECT*
EUROPEAN CENTRE FOR THEORETICAL STUDIES IN NUCLEAR PHYSICS AND RELATED AREAS

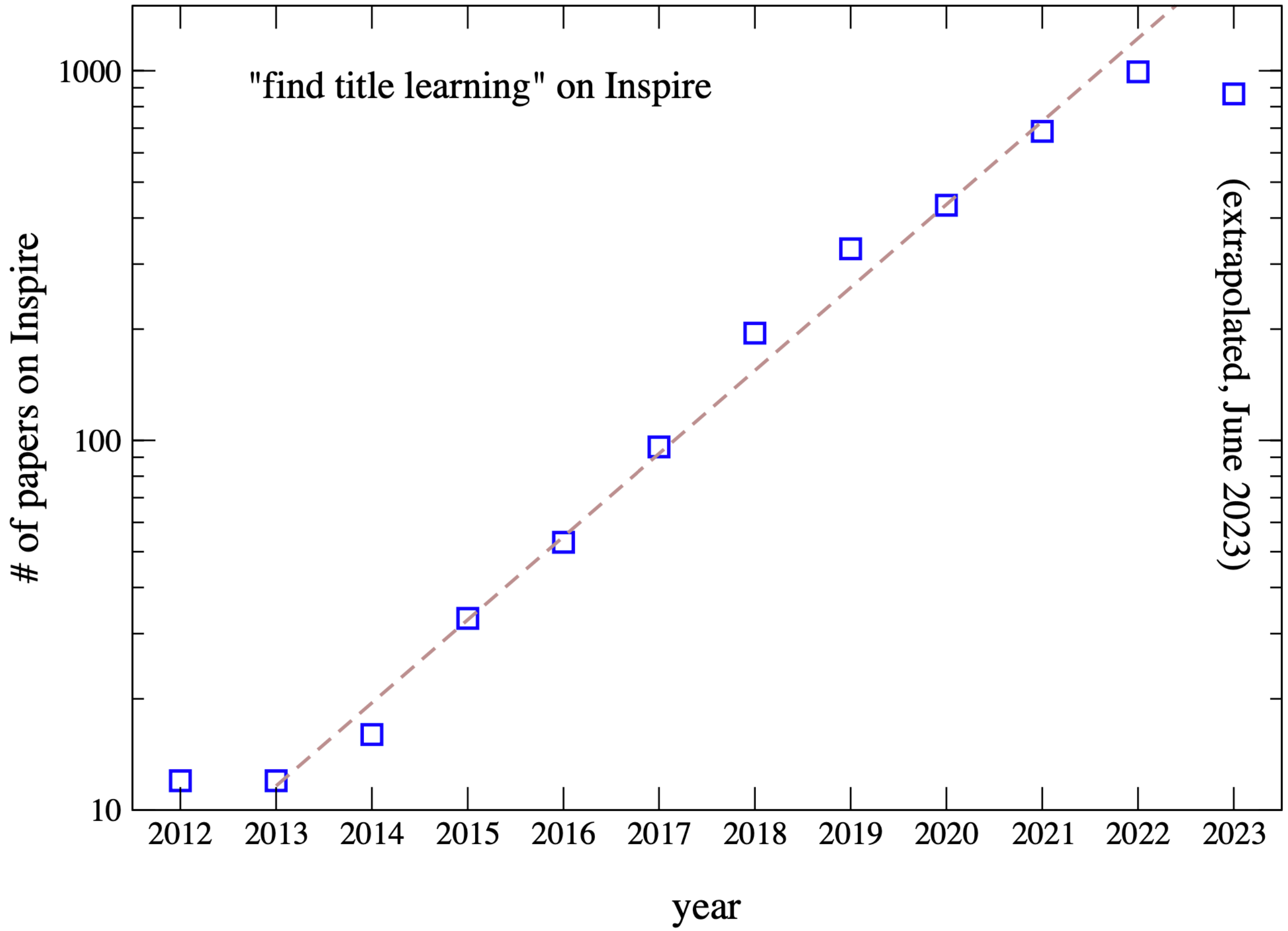
ExaTEPP, Swansea, June 2023

Introduction

- past six years or so has seen a rapid rise of applications of machine learning (ML) in fundamental science, particle physics, theoretical physics
- of course ML has been around for quite some time, especially in experimental particle physics
- nevertheless, there is an **exponential** increase in activity

Introduction





- find title **learning** on the iNSPIRE data base (high-energy physics)
- exponential growth!

Opportunities in ~~AI~~ machine learning

- opportunities in experimental particle physics, astronomy, gravitational waves, …
  (not discussed here)

- theoretical physics, lattice field theory

Applications of Machine Learning to Lattice Quantum Field Theory
Boyda, Aarts, Lucini et al, contribution to Snowmass 2022
arXiv:2202.05838 [hep-lat]

Outline: ML in lattice field theory

- configurations – generating ensembles, tuning algorithms
- observables – correlators, thermodynamics, …
- analysis – fitting, phase classification, ill-posed inverse problems, …
- more generally: which method to use, why does it (not) work, understand ML

Applications of Machine Learning to Lattice Quantum Field Theory, Boyda, Aarts, Lucini et al, arXiv:2202.05838 [hep-lat]

Outline: ML in lattice field theory

- configurations – generating ensembles, tuning algorithms

  normalising flow, gauge equivariance

- observables – correlators, thermodynamics, …

- analysis – fitting, phase classification, ill-posed inverse problems, …

  neural network outputs as physical observables

- more generally: which method to use, why does it (not) work

  field theory approaches to understanding ML

Generating configurations

- well-known problems in MCMC: critical slowing down, topological freezing
- generate configurations starting from “simple” distribution
- perform change of variables to reach desired distribution: invertible map
- simple example

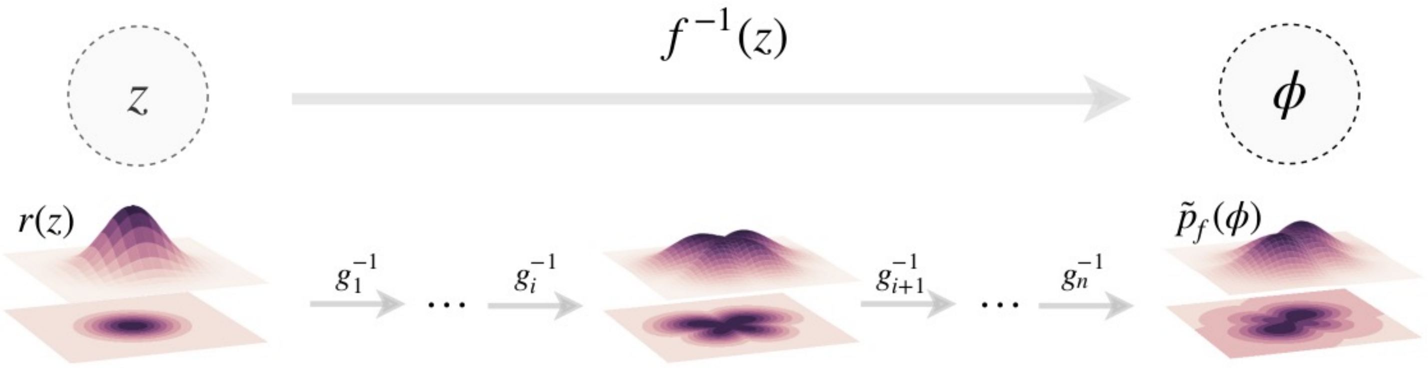
  Box-Mueller transformation: from uniform distribution to Gaussian distribution

normalising flow, trivialising map

- many applications in e.g. image generation in ML literature
- applications to lattice field theory (since 2019)

Flow-based generative models for Markov chain Monte Carlo in lattice field theory
Albergo, Kanwar, Shanahan, Phys. Rev. D 100 (2019) 3, 034515 [1904.12072 [hep-lat]]

Generating configurations: normalising flow



- from Gaussian distribution $r(z)$ to desired distribution $p(\phi)$
- generated by neural network, sequence of invertible (matrix+shift) transformations
- trained by minimising distance between learned and target distribution
- due to checkerboard structure: Jacobian of learned transformation is trivial
- “provably exact”: insert Metropolis-Hastings step at the end

Introduction to Normalizing Flows for Lattice Field Theory, Albergo et al, arXiv:2101.08176 [hep-lat]
Aspects of scaling and scalability for flow-based sampling of lattice QCD, Abbott et al, 2211.07541 [hep-lat]

Normalising flow: applications to QCD

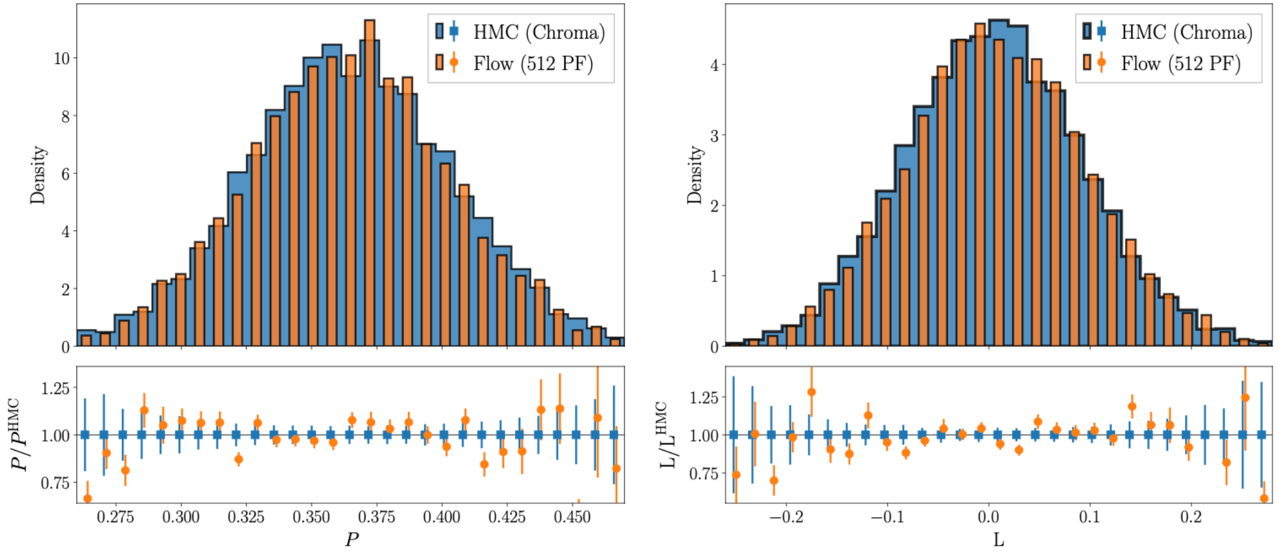
challenges:

- higher dimensions: not 2d (images), but 3d and 4d spacetime
- gauge symmetry: large internal symmetry, do not want to sample redundant dof
- construct gauge equivariant coupling layers (commute with gauge transformations)
- gauge invariant input distribution → gauge invariant output distribution

first application in 4d QCD

with $N_f = 2$ on a 4^4 lattice

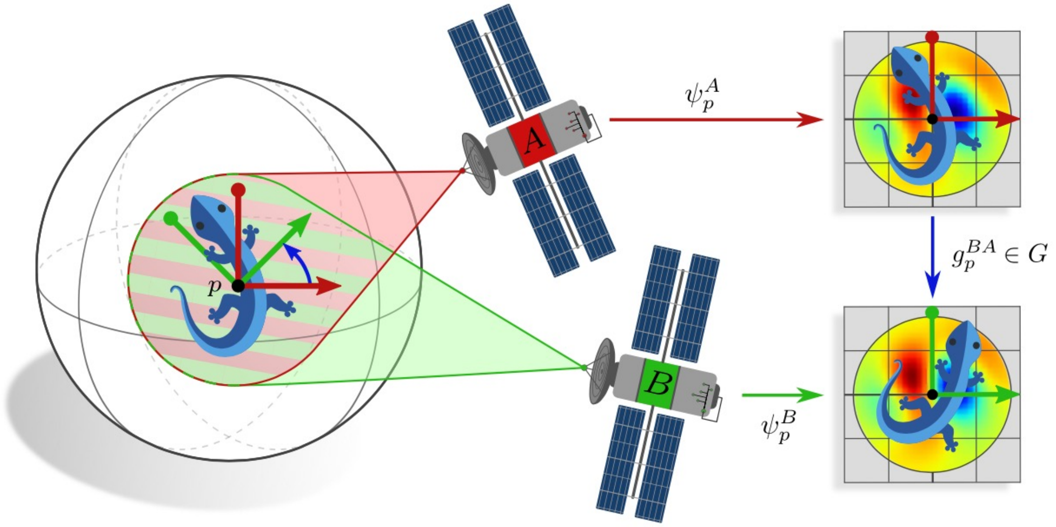
- scalability?



(a) Plaquette (b) Polyakov loop

Sampling QCD field configurations with gauge-equivariant flow models, Abbott et al, PoS LATTICE2022 (2023) 036 [2208.03832 [hep-lat]]

Gauge equivariance



deep connections to recent developments in ML

- coordinate independence
- local reference frame
- Convolutional Neural Nets on Riemannian manifolds

applications in

- vision
- medical imaging
- climate patterns

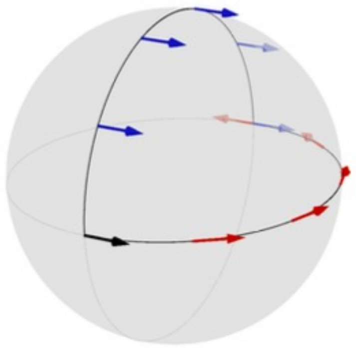


Figure 2. On curved spaces, parallel transport is path dependent. The black vector is transported to the same point via two different curves, yielding different results. The same phenomenon occurs for other geometric objects, including filters.

Coordinate Independent Convolutional Networks--Isometry and Gauge Equivariant Convolutions on Riemannian Manifolds, Weiler, Forré, Verlinde, Welling, arXiv preprint arXiv:2106.06020 [cs.LG]
Gauge equivariant convolutional networks and the icosahedral CNN, Cohen, Weiler, Kicanaoglu, Welling International conference on Machine learning, 1321-1330 [arXiv:1902.04615v3 [cs.LG]

More LFT applications of gauge equivariance

- lattice gauge equivariant convolutional neural networks,

  Favoni, Ipp, Mueller, Schuh, Phys. Rev. Lett. 128 (2022) 3, arXiv:2012.12901 [hep-lat]

- gauge-equivariant pooling layers for preconditioners in lattice QCD

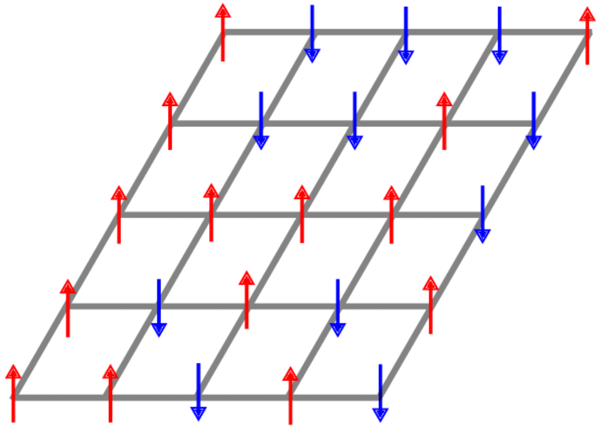
  Lehner and Wettig, arXiv:2304.10438 [hep-lat]

✓ active field, interesting cross-talk with other ML applications

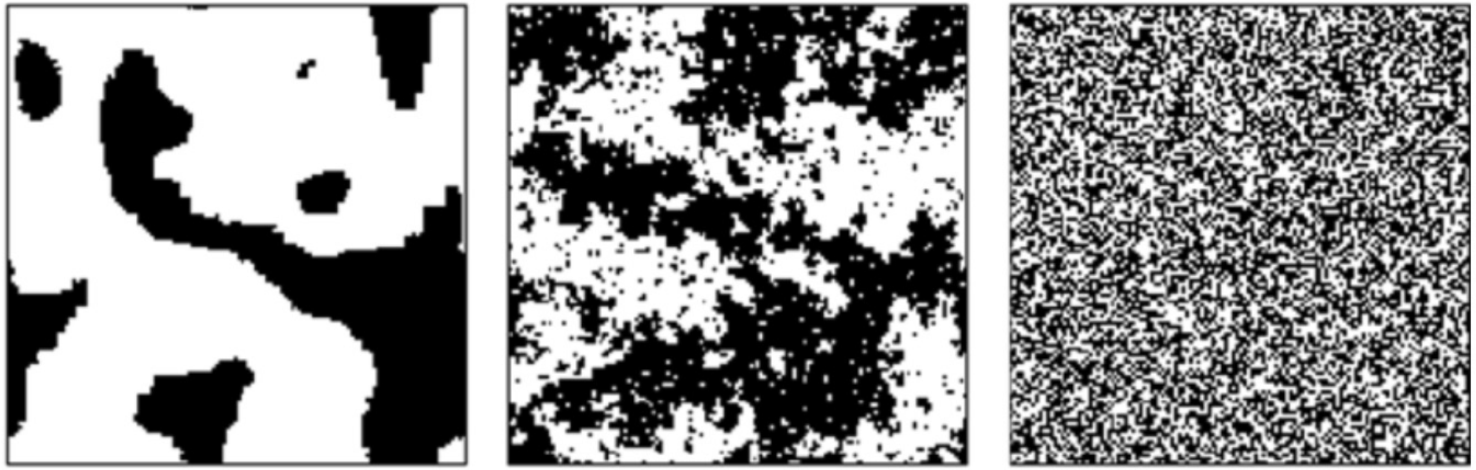
✓ requires “theoretical physicists/QFT experts” to master formalism

opportunity for LFT community to contribute to ML world

Classification of phases of matter



- matter can exist in different phases
- prototype: 2d Ising model -> ordered/disordered or cold/hot phases
- task: determine phase a system is in, determine critical coupling or temperature



Ordered -- ? -- Disordered

Published: 13 February 2017

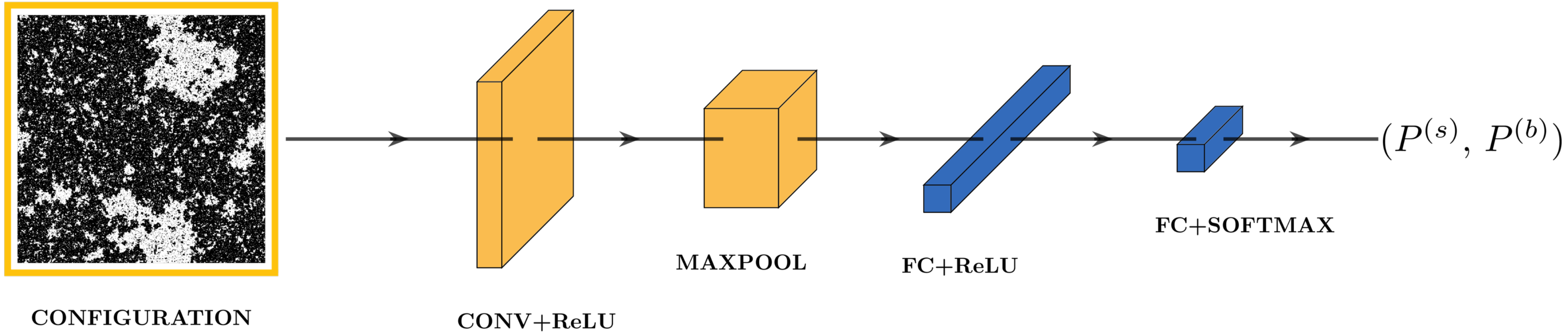
Machine learning phases of matter

Juan Carrasquilla & Roger G. Melko

Nature Physics **13**, 431–434(2017) | Cite this article

Phase classification: (by now) standard procedure

- use your favourite architecture, e.g. Convolutional Neural Network

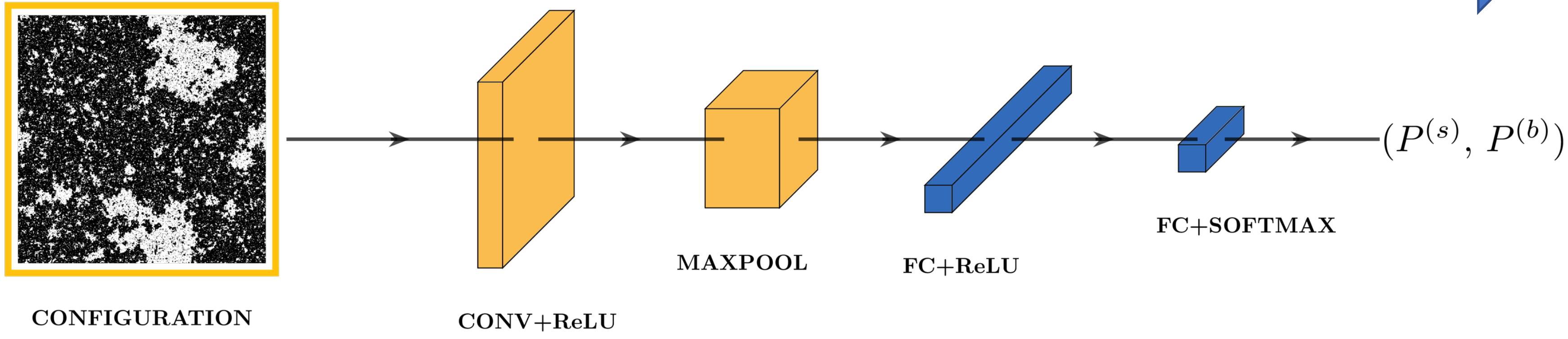


- input: train on sets of configurations away from the transition
- output: assign probability to be in ordered or disordered phase
- standard supervised classification problem
- apply to unseen configurations and predict

What can we add?

- ○ give a physical interpretation to neural network (NN) prediction
- ○ interpret output from a NN as an observable in a statistical system
- input: configurations, distributed according to Boltzmann weight
- output: observable, “order parameter” in statistical system

$$< P > = \frac{1}{Z} \sum_i P_i \, e^{-\beta E_i}$$



Extending machine learning classification capabilities with histogram reweighting,
Bachtis, Aarts, Lucini, Phys. Rev. E 102 (2020) 033303 [2004.14341 [cond-mat.stat-mech]]

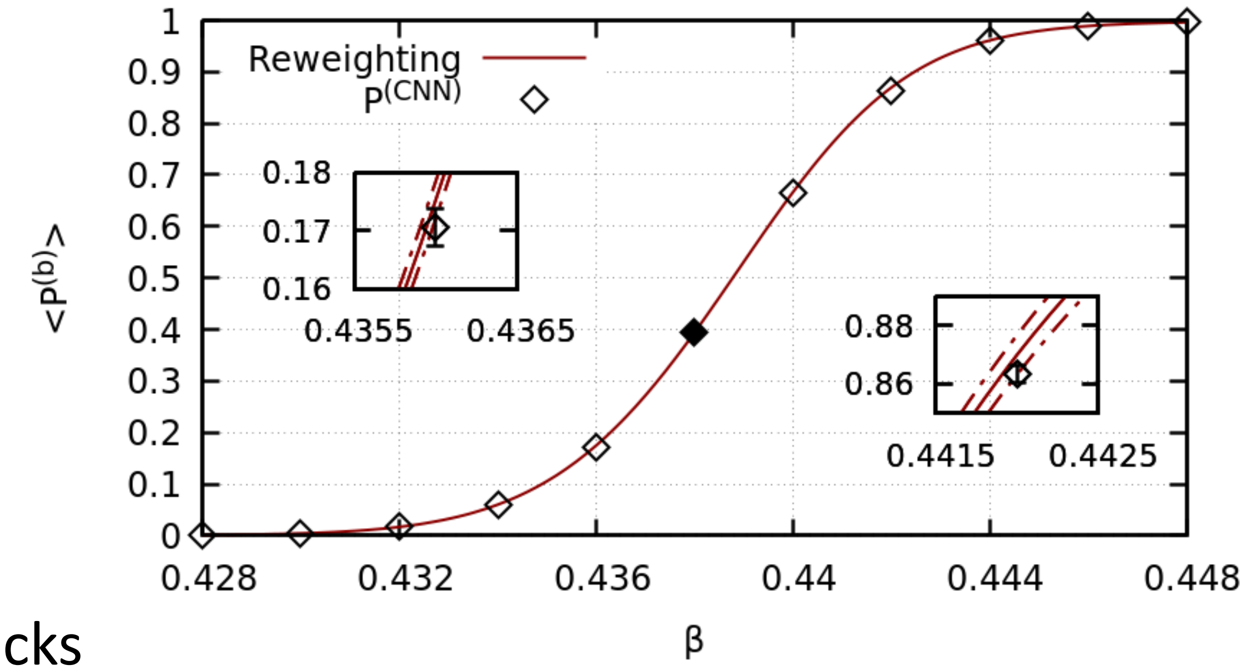
Output of NN as physical observable

- once you accept this: opens up possibility to use “standard” numerical/statistical methods

→ histogram reweighting: extrapolation to other parameter values

- starting from computation at given β_0: extrapolate to other β values

$$< P > (\beta) = \frac{\sum P_i e^{-(\beta-\beta_0)E_i}}{\sum e^{-(\beta-\beta_0)E_i}}$$



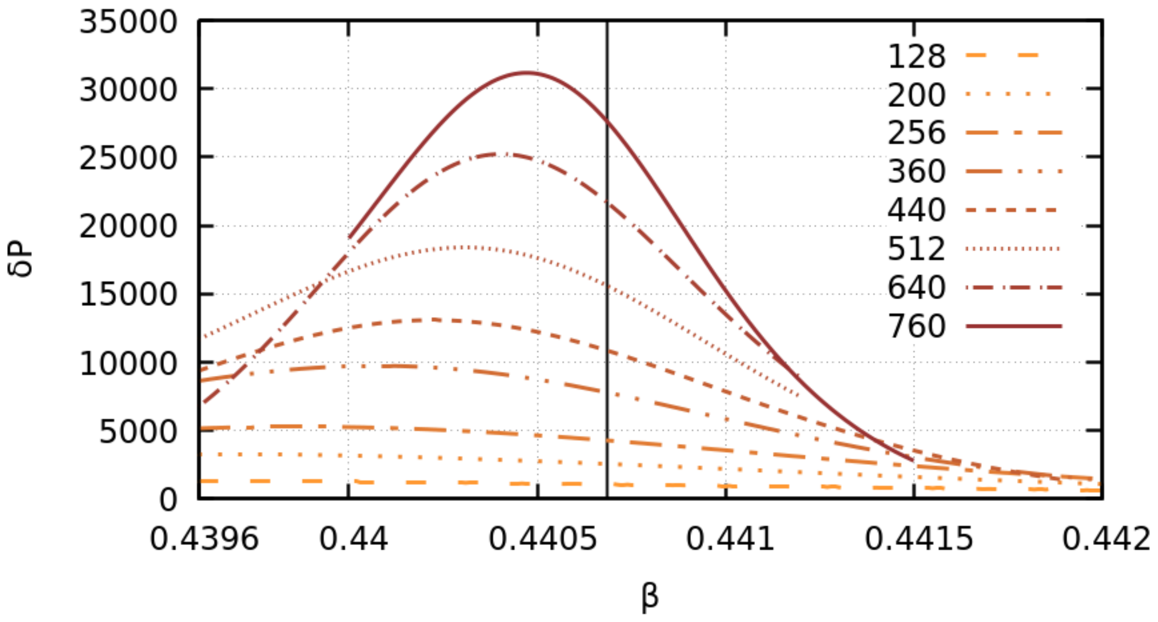
- ✓ filled diamond at β_0
- ✓ line obtained by reweighting in β
- ✓ open diamonds are independent cross checks

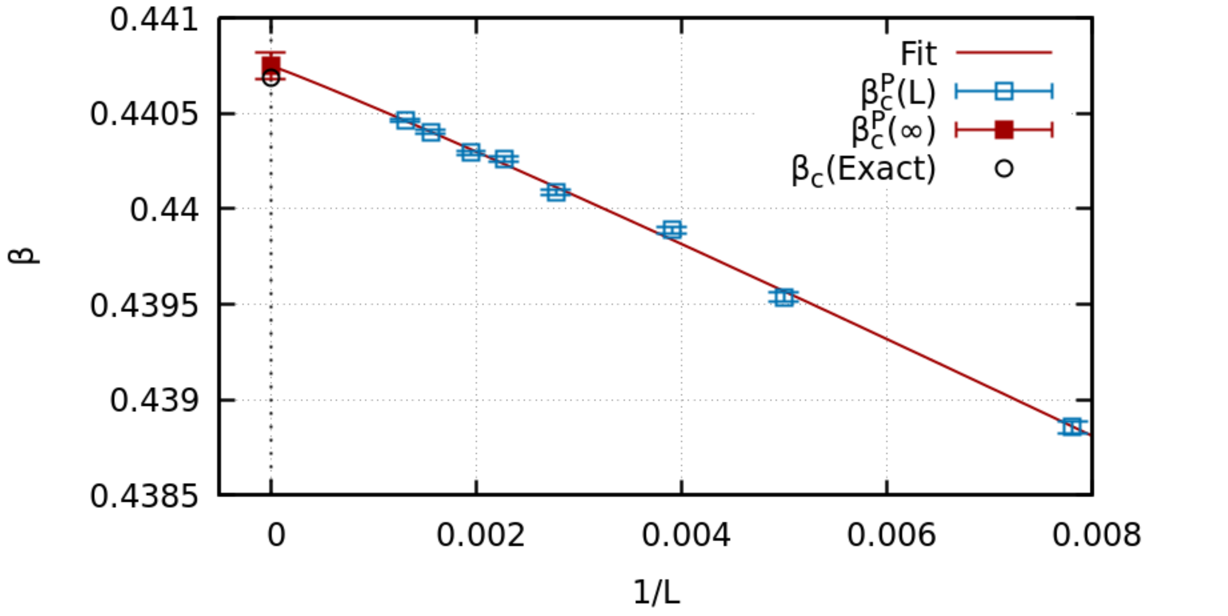
2d Ising model

Critical behaviour from NN observables

- determine L dependent susceptibility δP and its maximum at $\beta_c(L)$

2d Ising model

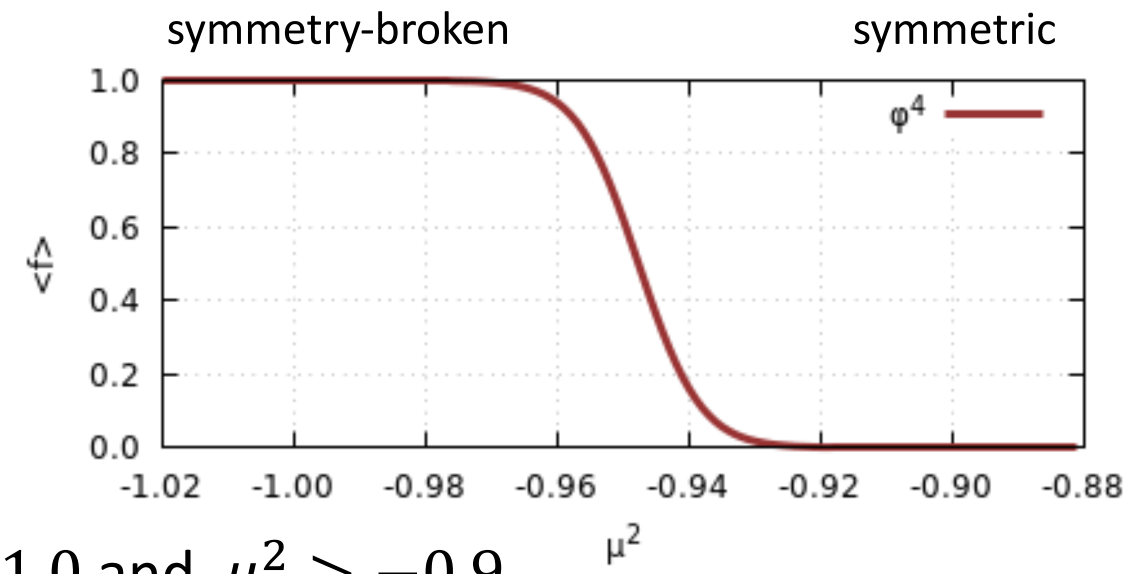




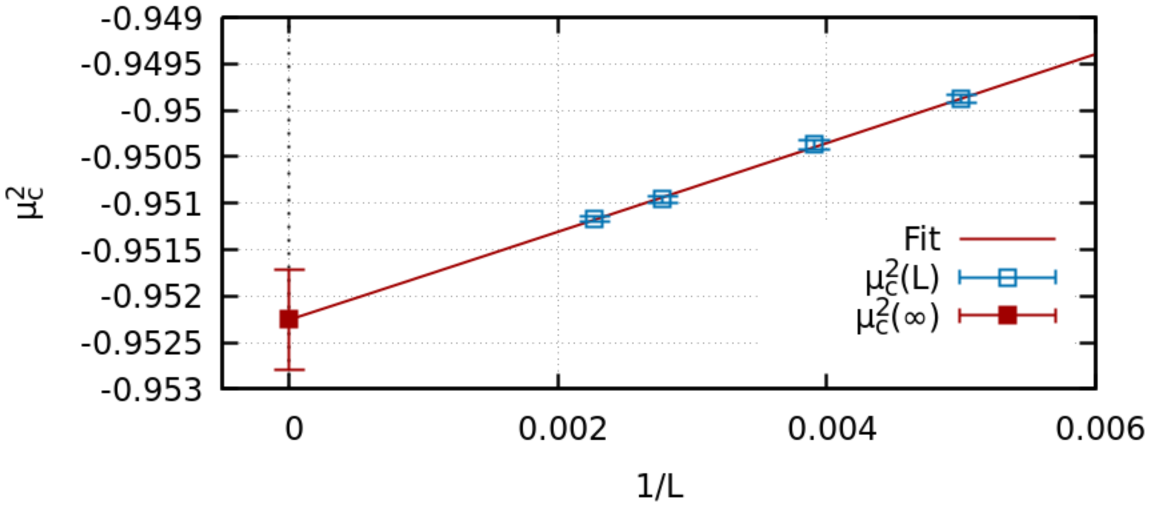
extract critical properties from NN observables only

	β_c	ν	γ/ν
CNN+Reweighting	0.440749(68)	0.95(9)	1.78(4)
Exact	$\ln(1+\sqrt{2})/2$ ≈ 0.440687	1	$7/4$ $=1.75$

φ^4 scalar field theory



- reweight in mass parameter, μ^2
- identify regions where phase is clear
- transfer learning: retrain NN using $\mu^2 < -1.0$ and $\mu^2 > -0.9$
- repeat finite-size scaling analysis as in 2d Ising model



	μ_c^2	ν	γ/ν
CNN+Reweighting	-0.95225(54)	0.99(34)	1.78(7)

- same universality class as 2d Ising model
- critical mass in agreement with results obtained with standard methods (Binder cumulant, susceptibility)

Mapping distinct phase transitions to a neural network, Bachtis, Aarts, Lucini
Phys. Rev. E 102 (2020) 053306 [2007.00355 [cond-mat.stat-mech]]

Quantum field-theoretical machine learning

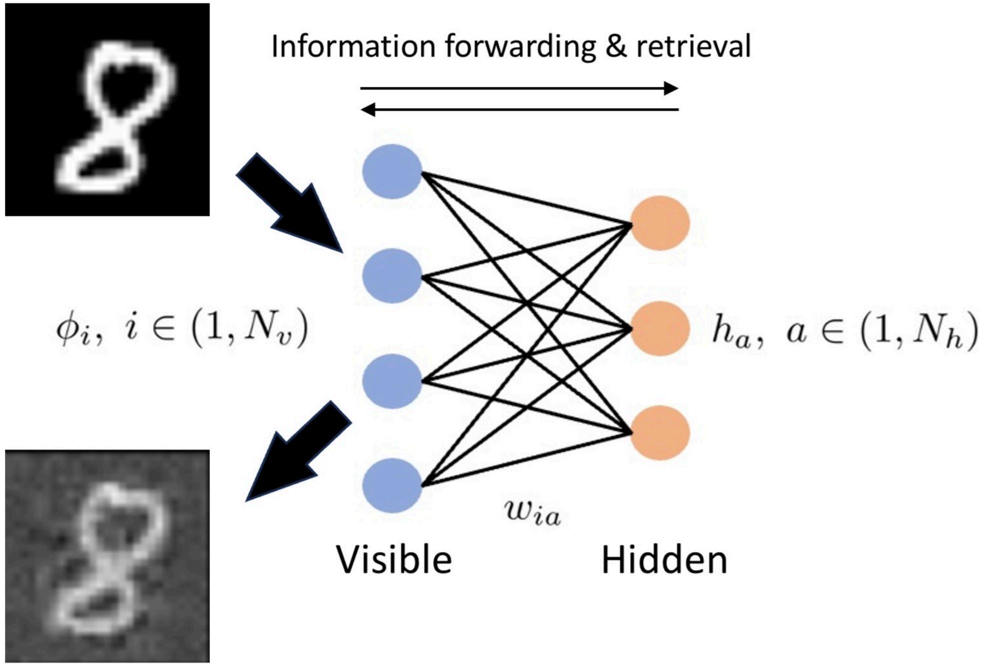
- improve understanding of ML using QFT techniques

- propose new formulations using QFT intuition

Quantum field-theoretic machine learning,
Bachtis, Aarts, Lucini, Phys. Rev. D 103 (2021) 074510
[2102.09449 [hep-lat]]
Aarts, Lucini, Park (in preparation)

Example: Restricted Boltzmann Machine

two-layer generative network



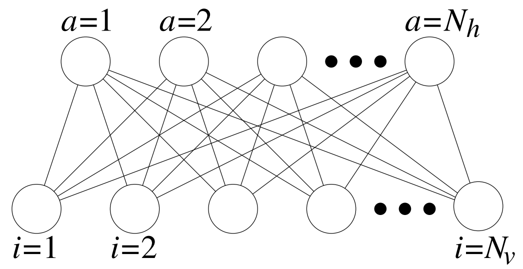
- visible layer: to encode probability distribution
- hidden layer: to encode correlations

- restricted: no connections within a layer

- degrees of freedom on two layers can be spins, say ± 1, or continuous, or mixed

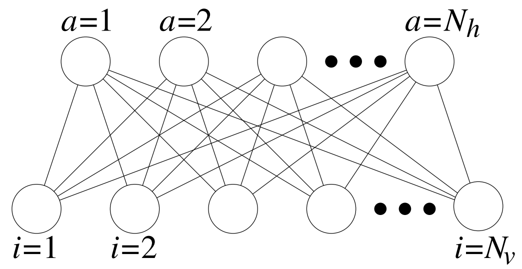
- energy-based method: $p(\phi, h) = \frac{1}{Z} e^{-S(\phi,h)}$ $\qquad Z = \int D\phi Dh\, e^{-S(\phi,h)}$

Scalar field RBM



- use field theory approach: start with "free fields": Gaussian-Gaussian RBM
- distribution: $$p(\phi, h) = \frac{1}{Z} e^{-S(\phi,h)} \qquad Z = \int D\phi Dh\, e^{-S(\phi,h)}$$
- energy (or action): $$S(\phi, h) = \sum_i \frac{1}{2}\mu_i^2 \phi_i^2 + \sum_a \frac{1}{2\sigma^2}(h_a - \eta_a)^2 - \sum_{i,a} \phi_i w_{ia} h_a$$
- no interaction between nodes in a layer, bilinear coupling between layers
- start with quadratic terms, add interactions later, e.g. ϕ^4 terms

Gaussian scalar field RBM



- two Gaussian fields with bilinear coupling

$$S(\phi, h) = \sum_i \frac{1}{2}\mu_i^2\phi_i^2 + \sum_a \frac{1}{2\sigma^2}(h_a - \eta_a)^2 - \sum_{i,a} \phi_i w_{ia} h_a$$

- 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 $J_i = \sum_a w_{ia}\eta_a$

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

Some results for $N_h = N_v$

explicit representations for the weight matrix W

1. $\mathcal{K}$ is symmetric, positive-definite: use Cholesky decomposition $\mathcal{K} = LL^T$ with L real lower triangular matrix with positive diagonal entries hidden node a connects to visible nodes with $i \leq a$ only

   $$W = L$$

2. $\mathcal{K}$ is symmetric, positive-definite: diagonalise using orthogonal transformation

   $$\mathcal{K} = ODO^T = O\sqrt{D}O^T O\sqrt{D}O^T \quad \rightarrow \quad W = W^T = O\sqrt{D}O^T$$

3. non-uniqueness: internal symmetry $W \to WO_R$ with $O_R O_R^T = \mathbb{1}$ leaves WW^T invariant, reshuffles hidden nodes $\phi^T W h \to \phi^T W O_R h = \phi^T W h'$

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

Some results for $N_h = N_v$

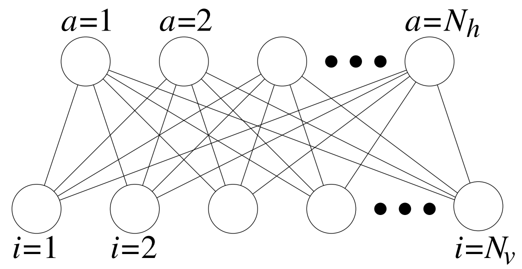
(infinitely) many solutions for weight matrix:

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 $W \to WO_R$ → $\phi^T W h \to \phi^T W O_R h = \phi^T W h'$

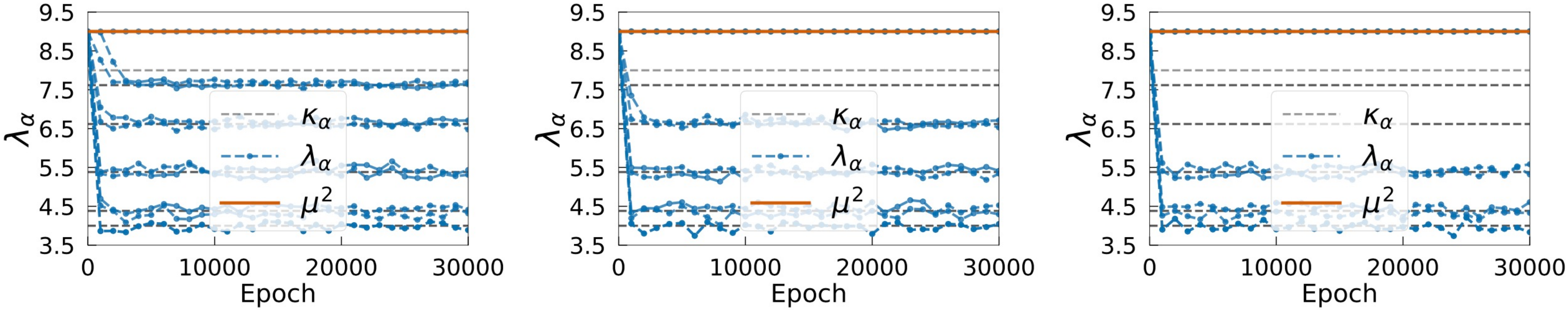
in practice

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

Result for $N_h < N_v$



- number of hidden nodes act as an ultraviolet regulator
- spectrum of induced quadratic operator on visible layer $\sum_{ij} \phi_i K_{ij} \phi_j$



(a) $N_h = 9$ **(b)** $N_h = 7$ **(c)** $N_h = 5$

- exact spectrum (κ) reproduced by RBM (λ) from smallest eigenvalue upwards

Scalar field Restricted Boltzmann Machine as an ultraviolet regulator, Aarts, Lucini, Park, in preparation

Gaussian RBM for MNIST data with $N_h \leq N_h$

- applicable to typical ML data sets, such as MNIST
- analyse RBM using language and intuition from LFT
- obtain some straightforward but rigorous results

(a) $N_h = 784$

(b) $N_h = 225$

(c) $N_h = 64$

(d) $N_h = 36$

(e) $N_h = 16$

(f) $N_h = 4$

- include interactions using statistical field theory methods, e.g. perturbation theory

Scalar field Restricted Boltzmann Machine as an ultraviolet regulator, Aarts, Lucini, Park, in preparation

Outlook

- inspiring connection between problems in lattice field theory and machine learning
- new solutions to old problems/old solutions to new problems
- insights work both ways: plenty of opportunities for impact in LFT and ML