Machine learning for lattice field theory and back

Gert Aarts

XQCD, July 2023

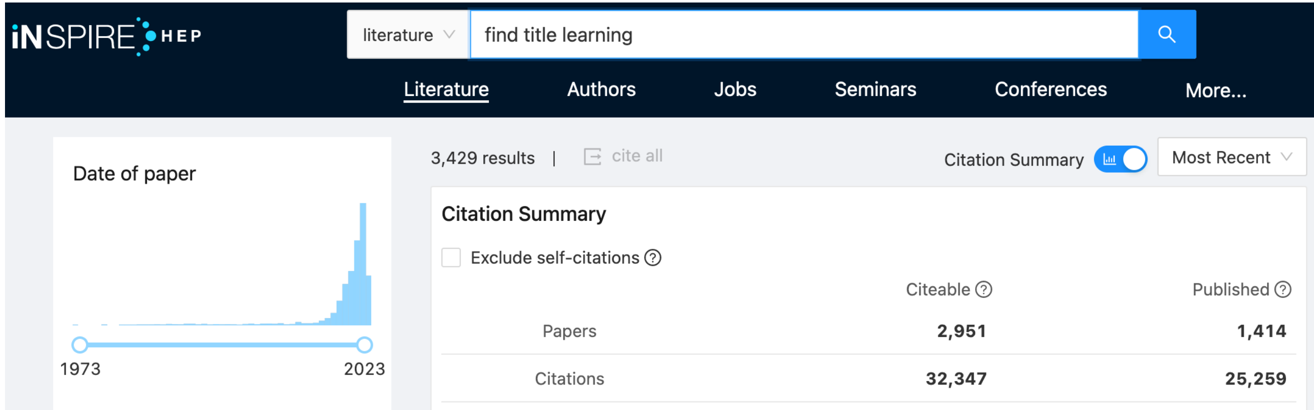
Introduction

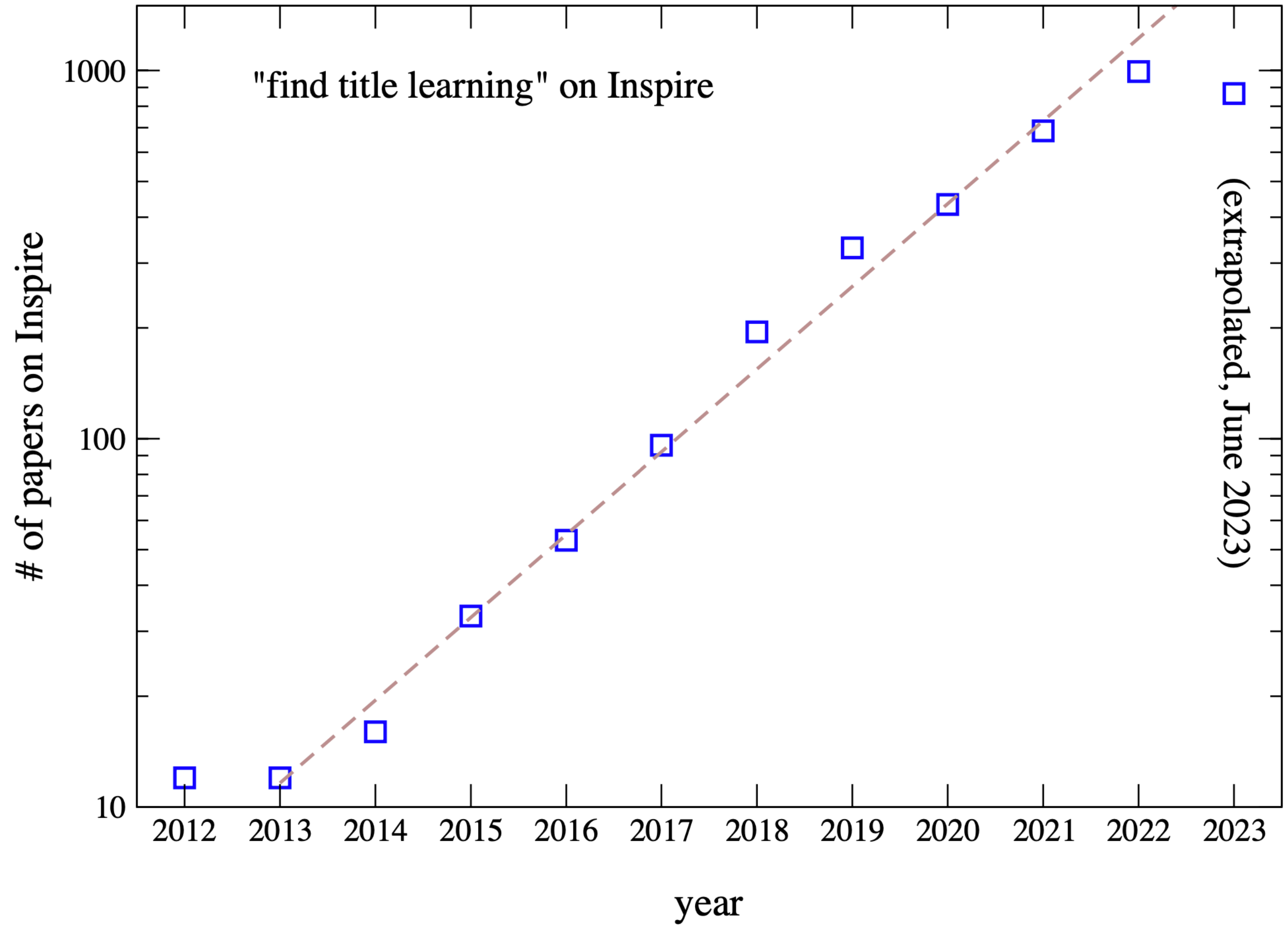
- past six years or so have seen a rapid rise of applications of machine learning (ML) in fundamental science, particle physics, theoretical physics

- 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
- exponential growth!

ML in lattice field theory

explored in all aspects of LFT:

- 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
SNOWMASS paper, Boyda, Aarts, Lucini et al, arXiv:2202.05838 [hep-lat]

Outline

- two-page introduction to supervised ML
  - classification: order-disorder transition (by now classic application)
  - generating ensembles: normalising flow (popular application)
  - quantum field-theoretical machine learning (new conceptual ideas to explore)
- biased towards own work and interests in lattice field theory

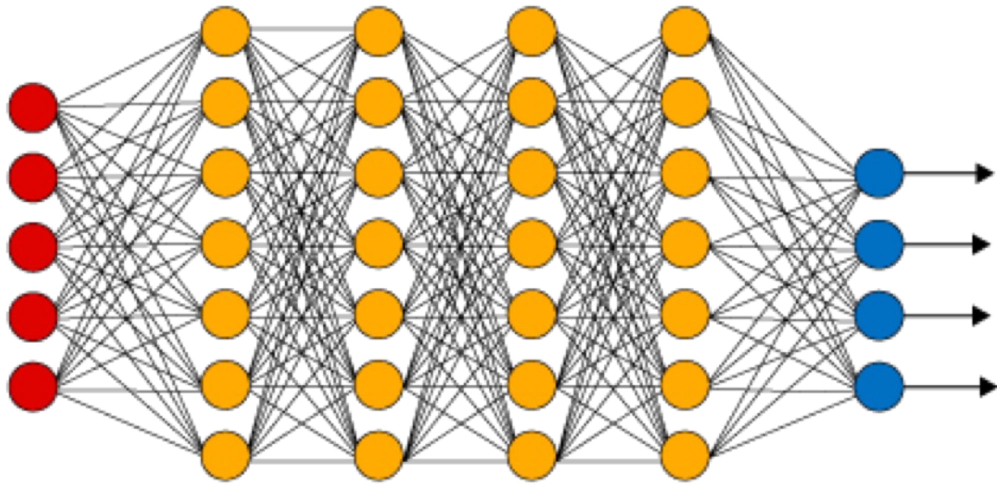
One-slide introduction to supervised ML

- attempt to fit a function or probability distribution to describe lots of data
- can be an actual function (regression) or a classification boundary (dog vs cat)
- functional form is not known: use a “universal approximator”, such as neural network
- linear combinations with weights and biases + nonlinear “activation” functions
- many, many, many internal parameters, determine these using training data
- generalise, make predictions for unseen cases, generate new instances, …

Neural networks, deep learning

- input layer – hidden layers – output layer

- many degrees of freedom ("neurons'')
  associated with sites of network



https://dvl.in.tum.de/teaching/i2dl-ws18/

- weights w_{ij} (connections) and biases b_i (on-site) are tunable parameters

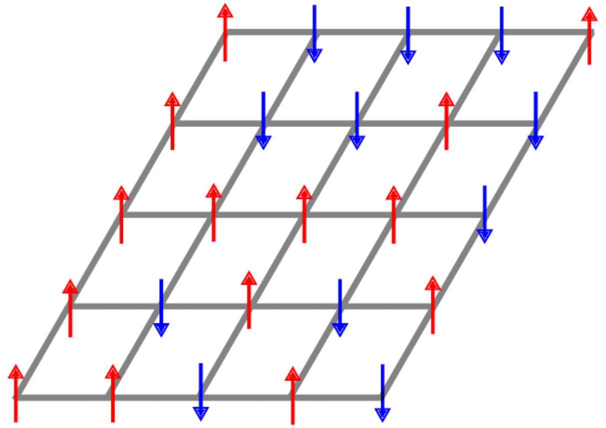
- learning: parameters are adjusted by minimising some cost or loss function

- NN should then encode some probability distribution and generate/classify/generalise

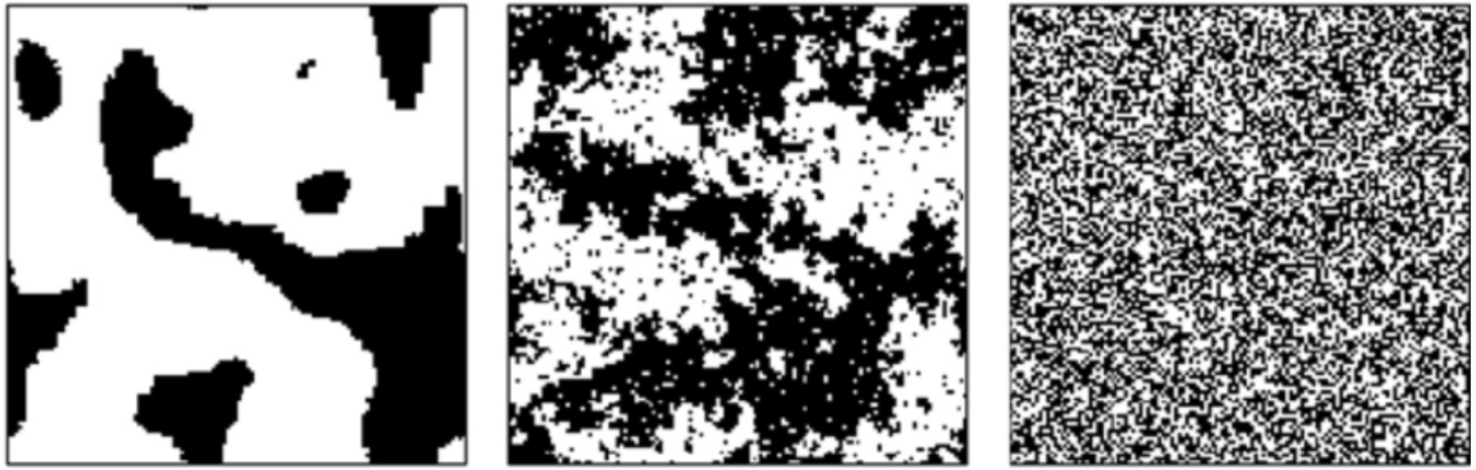
Outline

- **classification of phases of matter: order-disorder transition**
- generating ensembles: normalising flow
- quantum field-theoretical machine learning

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

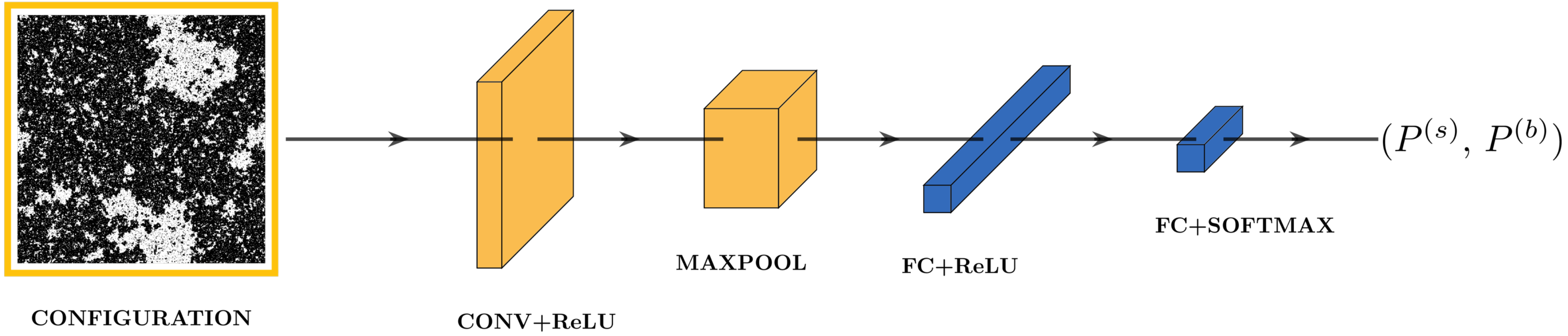


1350 cites

Ordered -- ? -- Disordered

Phase classification: (by now) standard procedure

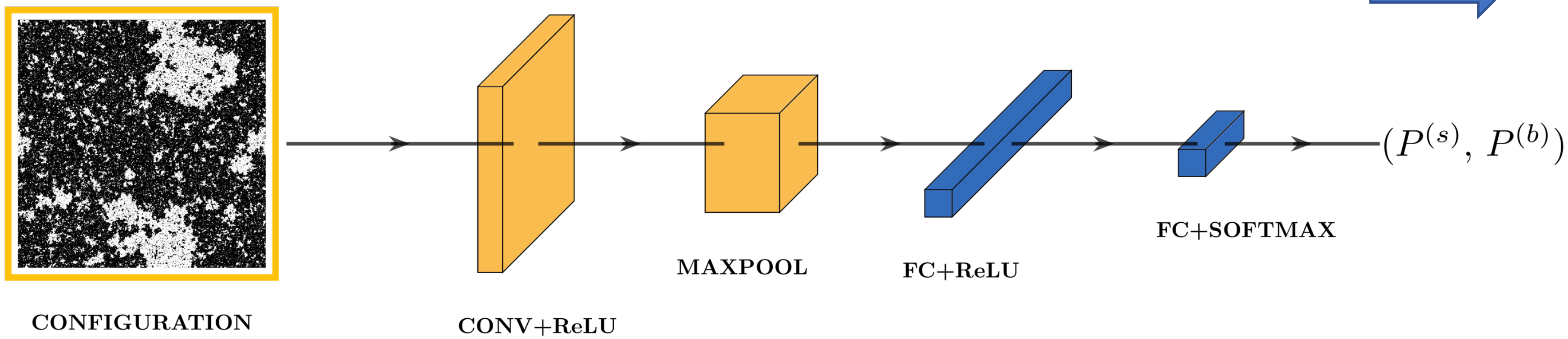
- use your favourite architecture, e.g. Convolutional Neural Network



- 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]]

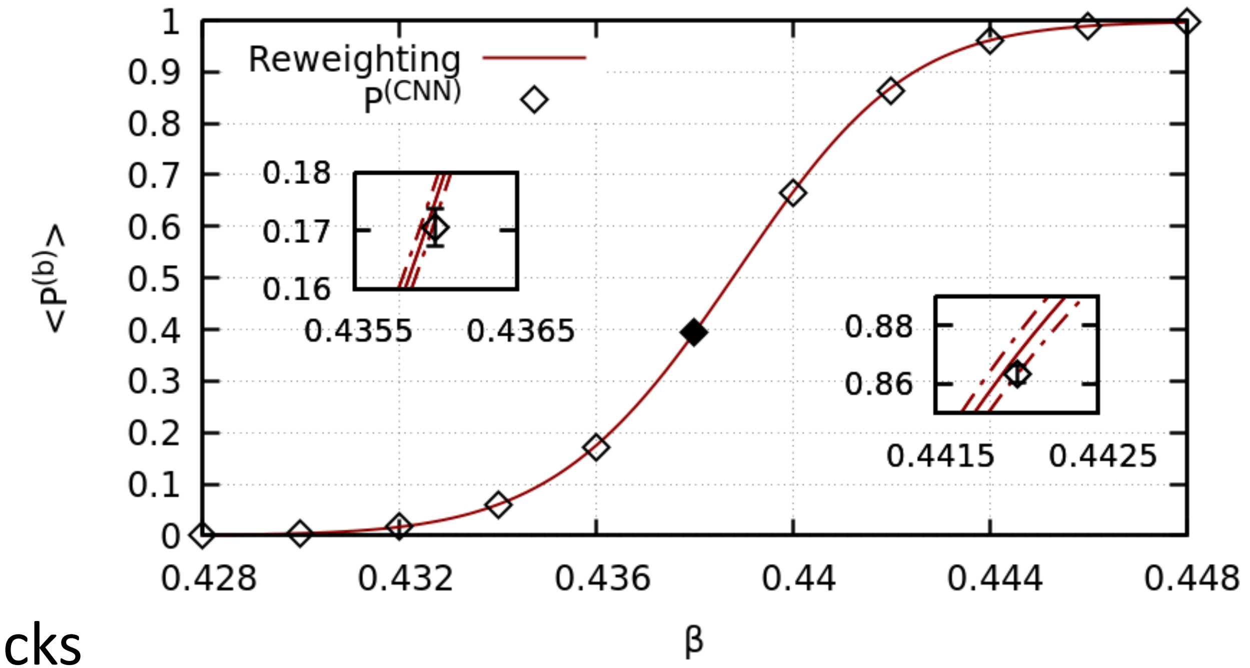
Output of NN as physical observable

- 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

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



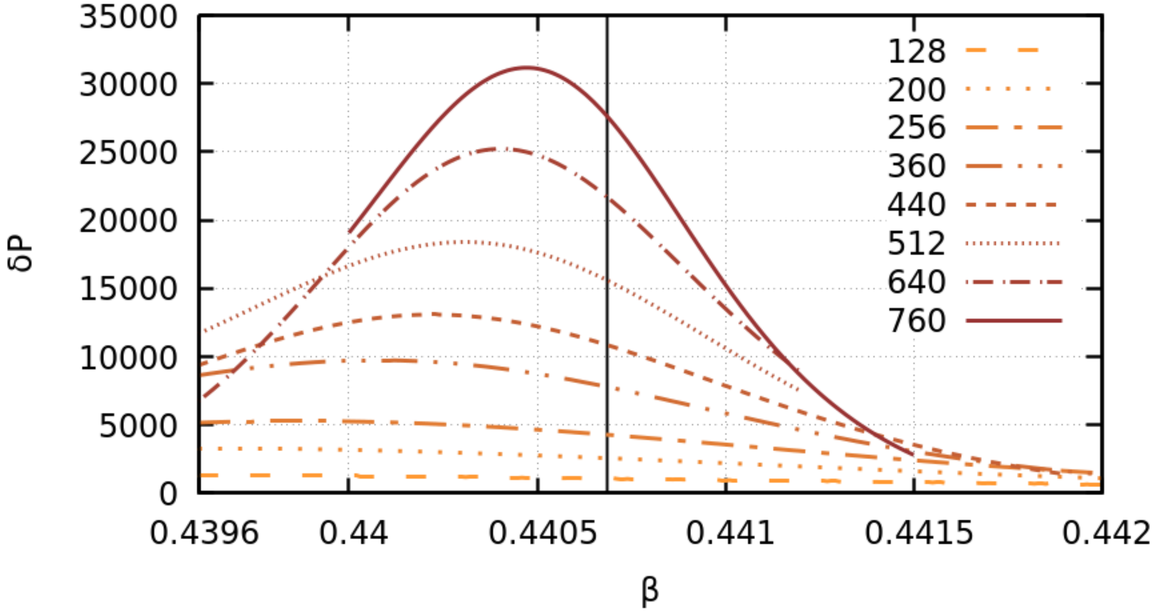
✓ filled diamond at β_0
✓ line obtained by reweighting in β
✓ open diamonds are independent cross checks

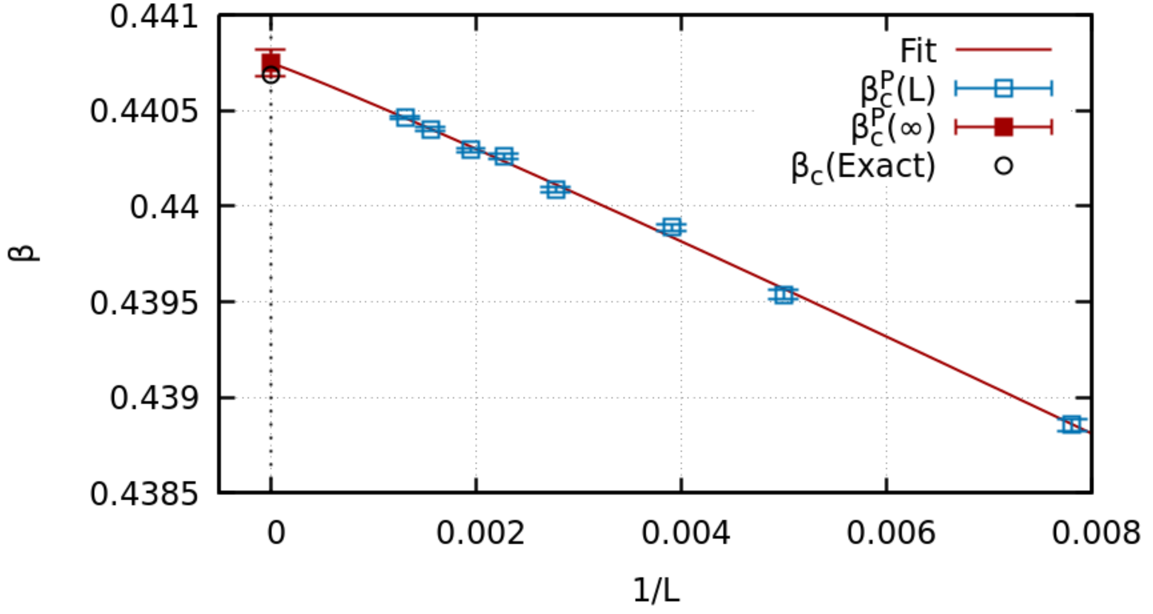
2d Ising model

Critical behaviour from NN observables

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

2d Ising model

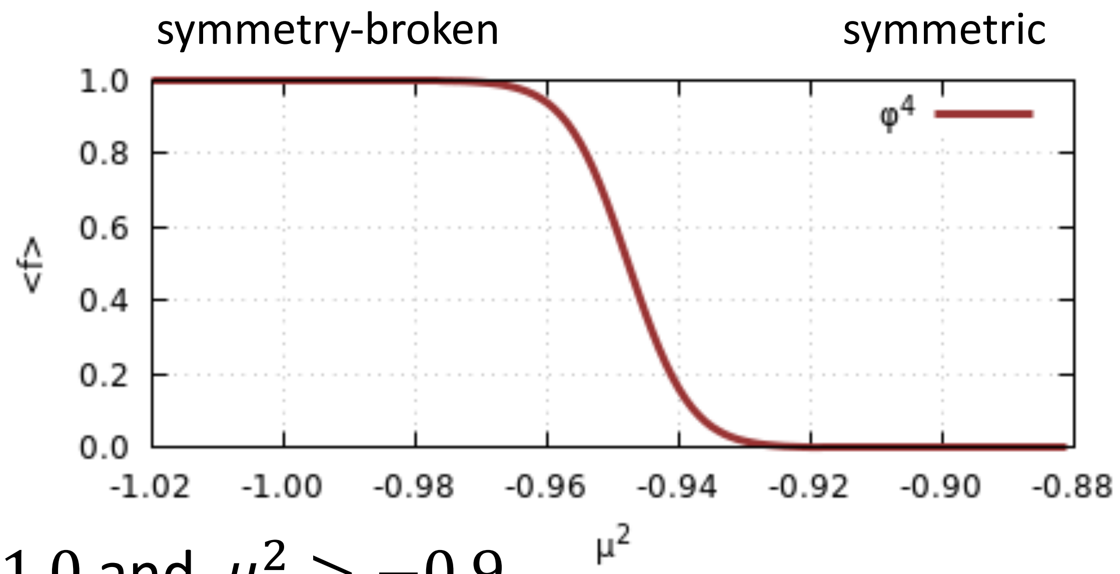




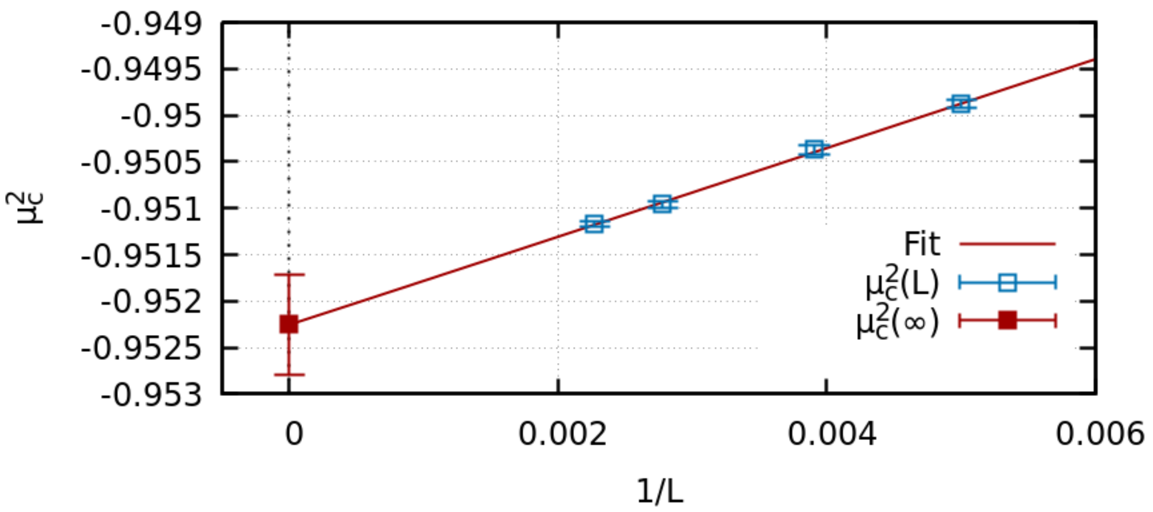
extract critical properties from NN observables only

	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]]

Outline

- classification of phases of matter: order-disorder transition
- **generating ensembles: normalising flow**
- quantum field-theoretical machine learning

Generating configurations in LFT

- 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

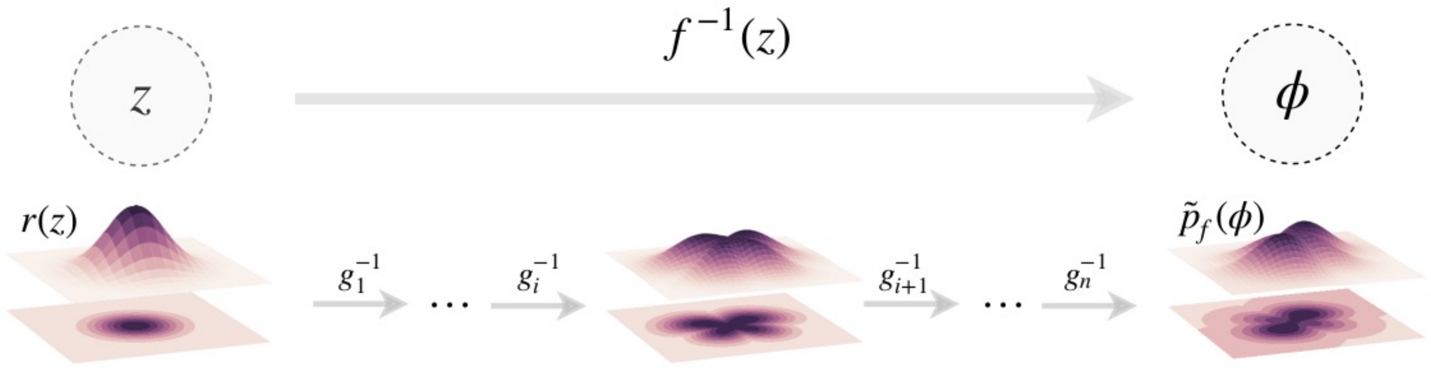
  Box-Mueller transformation: from uniform distribution to Gaussian distribution

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]

Generating configurations: normalising flow

- target distribution: $p_{\text{target}}(\phi)$, learned distribution: $p_{\text{learned}}(\phi)$

- compare distributions e.g. with Kullback-Leibler divergence

$$D(p_{\text{learned}} \,||\, p_{\text{target}}) = \int D\phi \;\, p_{\text{learned}}(\phi) \ln \frac{p_{\text{learned}}(\phi)}{p_{\text{target}}(\phi)} \geq 0$$

- philosophy: much easier to sample from learned distribution via trained network

Normalising flow: applications to QCD

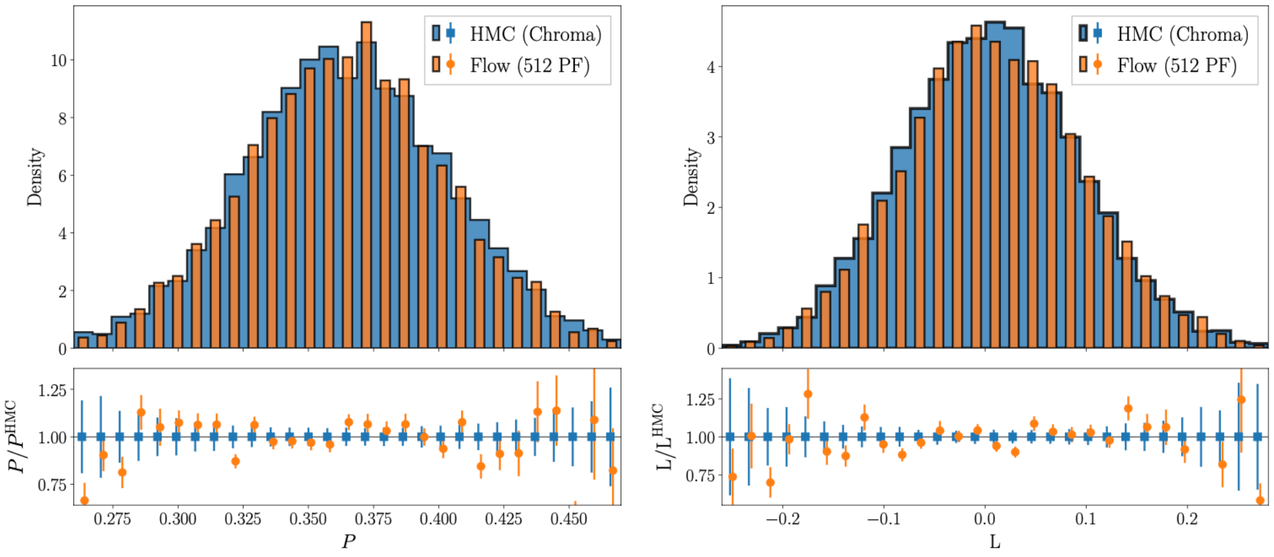
challenges:

- 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

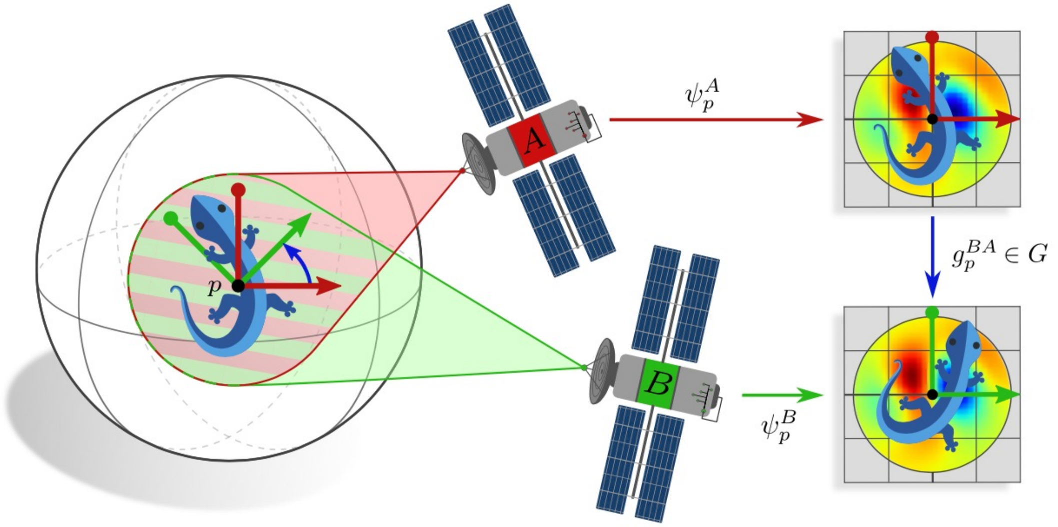
- scalability?



(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

see posters by
Daniel Schuh and
Matteo Favoni

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]

Normalising flow

- many variants being developed
- see talks at recent ECT* workshop
- https://indico.ectstar.eu/event/171/

If you are interested in organising a workshop at ECT* in 2024: please click this link
deadline September 20, 2023

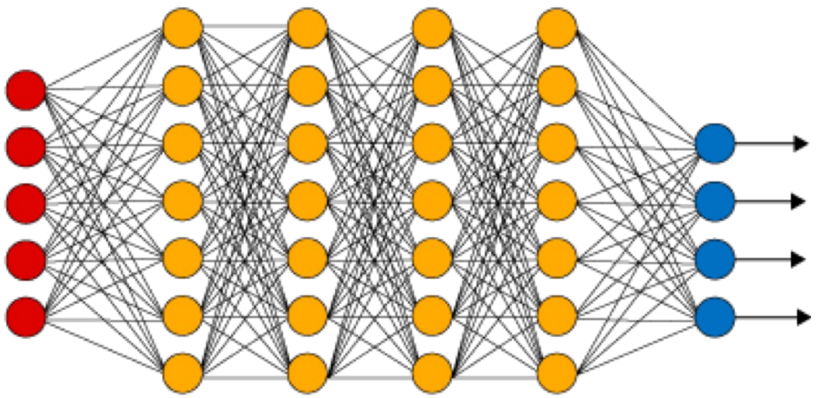
Outline

- classification of phases of matter: order-disorder transition
- generating ensembles: normalising flow
- **quantum field-theoretical machine learning**

Can we understand ML using QFT methods?

neural network:

- system with many fluctuating degrees of freedom
- connected via links
- represents probability distribution



- can we use stat mech/QFT methods?
- yes, already studied since 1980s

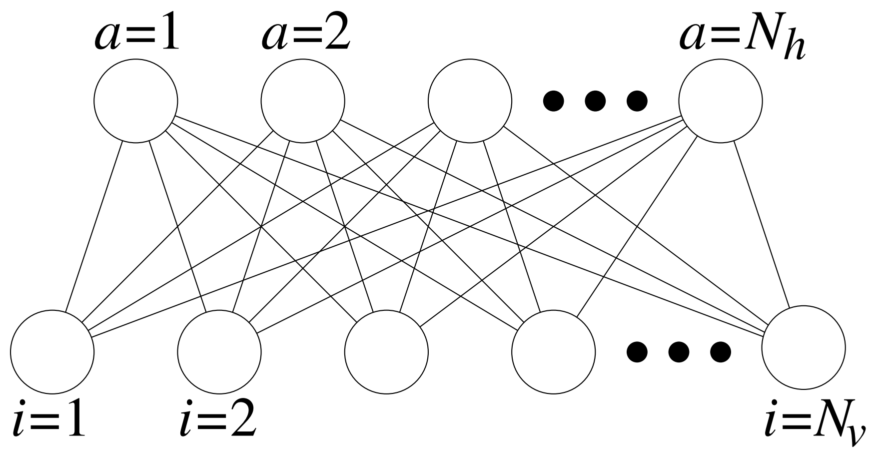
take a fresh look using our favourite tools

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

Restricted Boltzmann Machine

simplest example of a NN: two-layer generative network

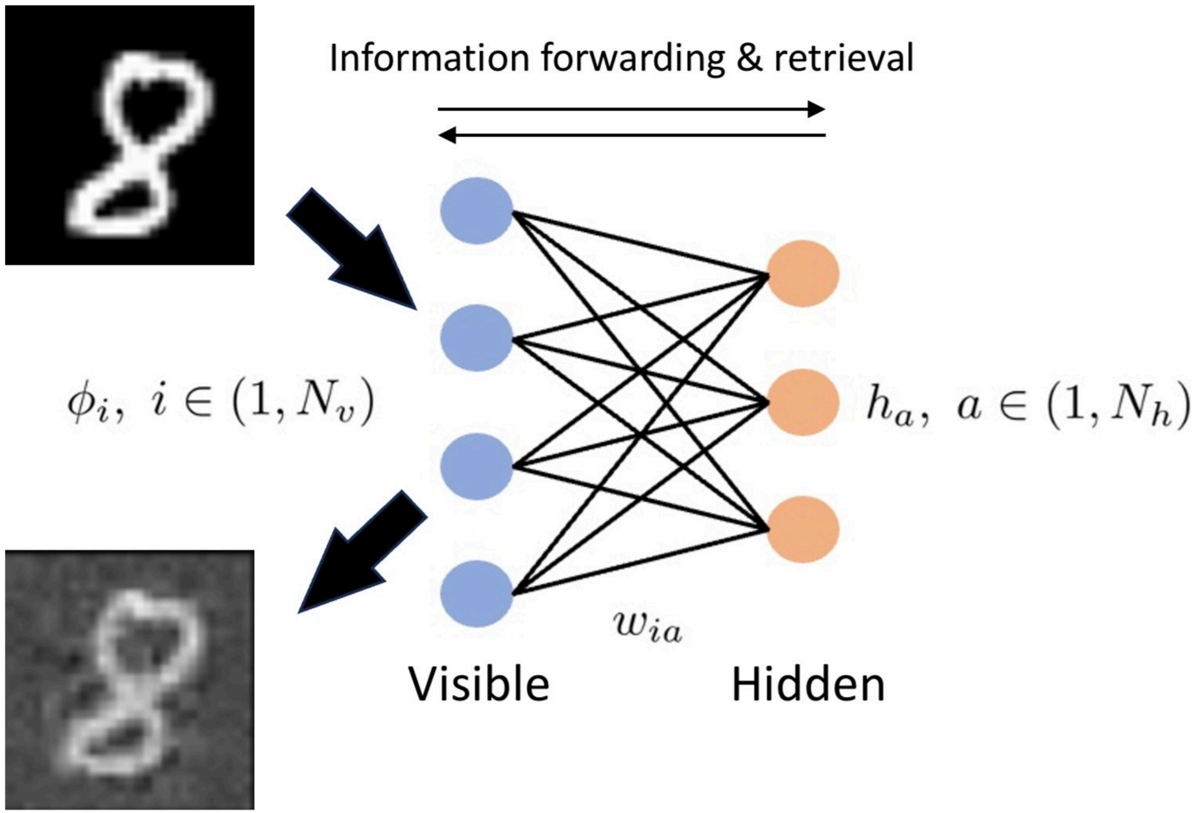
- visible layer: to encode probability distribution
- hidden layer: to encode correlations
- restricted: no connections within a layer



- degrees of freedom on two layers can be spins, say ±1, or continuous, or mixed
- weights connect the nodes, biases on the nodes

Aarts, Lucini, Park, in preparation

RBM: generative network

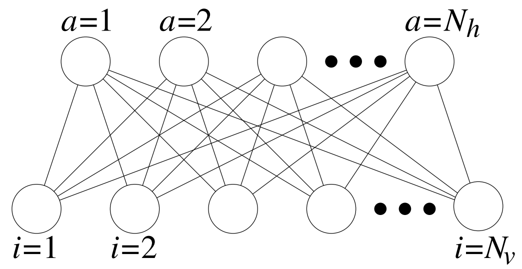


- energy-based method
- probability distribution

$$p(\phi, h) = \frac{1}{Z} e^{-S(\phi, h)}$$

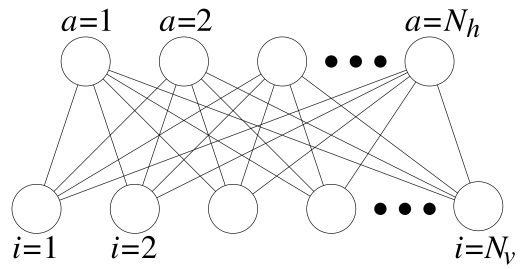
$$Z = \int D\phi Dh \, e^{-S(\phi, h)}$$

Scalar field RBM



- use field theory approach: start with "free fields": Gaussian-Gaussian RBM
- distribution: $p(\phi, h) = \frac{1}{Z} e^{-S(\phi,h)}$ $\quad Z = \int D\phi Dh\, e^{-S(\phi,h)}$
- 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
- only quadratic terms, add interactions later, e.g. ϕ^4 terms

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?

RBM probability function

- RBM should reproduce target distribution, determine $K_{ij} = \mu_i^2 \delta_{ij} - \sigma^2 \sum_a w_{ia} w_{aj}^T$
- learn from data or directly from known distribution
- simplest case: target theory = LFT of free scalar field in 1 or 2d
- kinetic matrix is LFT inverse propagator, $K^{\phi} \approx p^2 + m^2$
- can solve for weight matrix $WW^T = \frac{1}{\sigma^2}\left(\mu^2 \mathbb{1} - K^{\phi}\right) \equiv \mathcal{K}$

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

Exact results for $N_h = N_v$

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

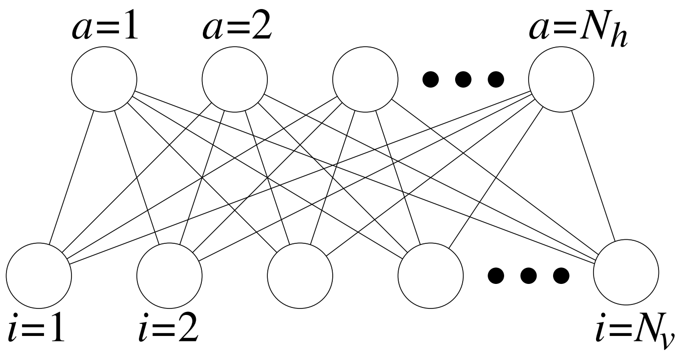
1. Cholesky decomposition $\mathcal{K} = LL^T$: $W = L$ triangular

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



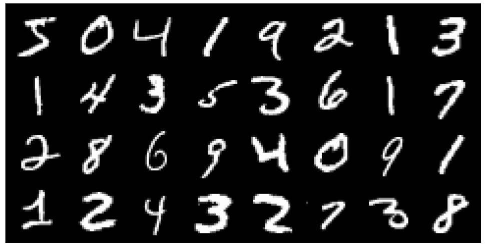
RBM training/architecture

- analytical solution usually not available
- train RBM numerically by maximising log-likelihood function

questions we would like to understand, for NNs in general and here:

- what determines the optimal architecture? number of hidden nodes?
- what is the role of internal parameters, e.g. RBM mass parameter μ^2?
- what if these parameters are chosen incorrectly?
- can one make generic statements?

use LFT insights, apply to MNIST data



What if $N_h < N_v$?

- analytical result: number of hidden nodes acts as an ultraviolet regulator
- consider spectrum of induced quadratic operator on visible layer $\sum_{ij} \phi_i K_{ij} \phi_j$
- exact spectrum of target distribution is reproduced from infrared scale upwards
- RBM is an ultraviolet regulator

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

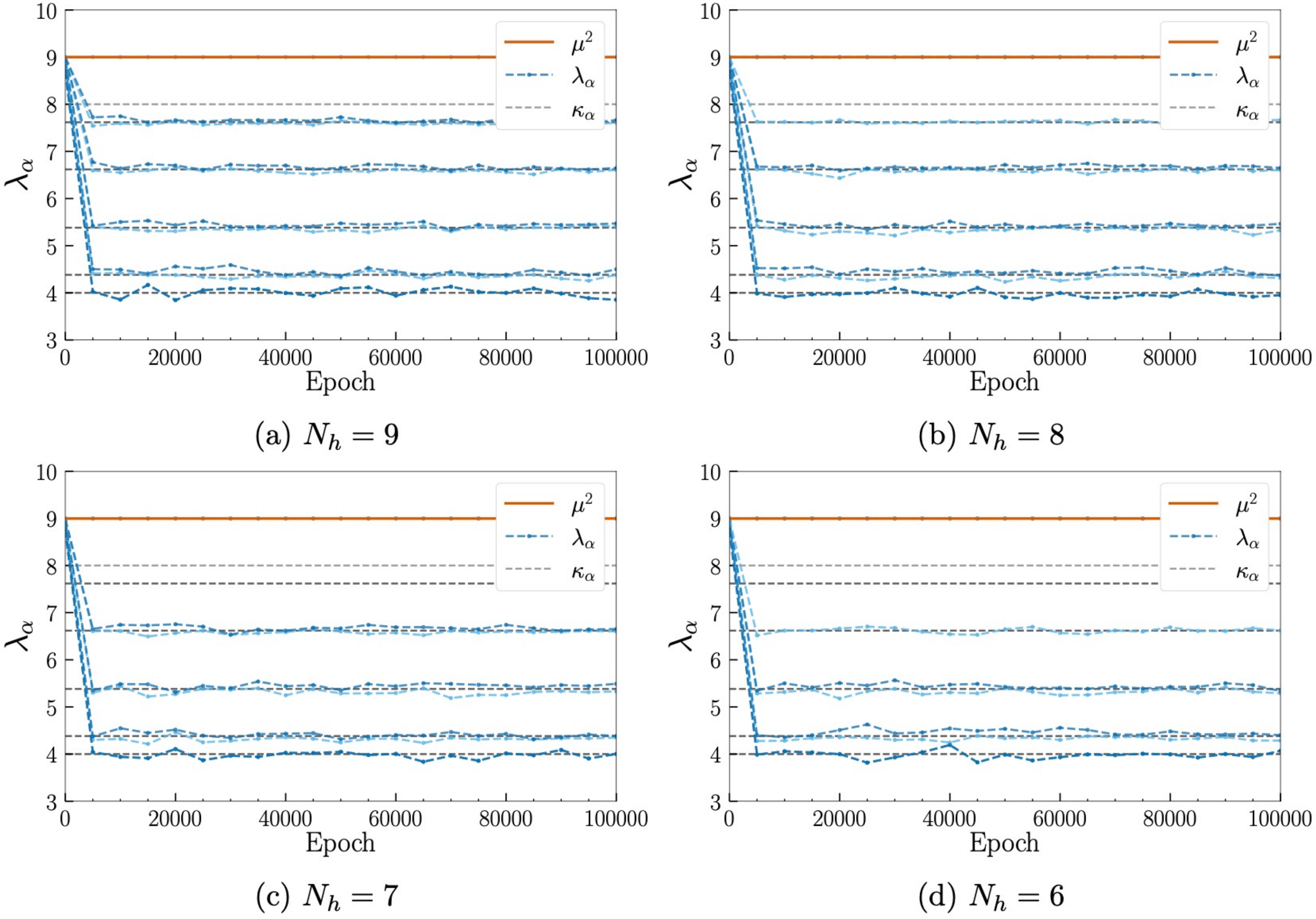
What if $N_h < N_v$?

- example: 1D scalar LFT
  with $N_v = 10$ nodes

- exact spectrum (κ)

- reproduced by RBM (λ) from smallest eigenvalue upwards

- higher modes are moved to cut-off scale (μ^2)



(a) $N_h = 9$ (b) $N_h = 8$ (c) $N_h = 7$ (d) $N_h = 6$

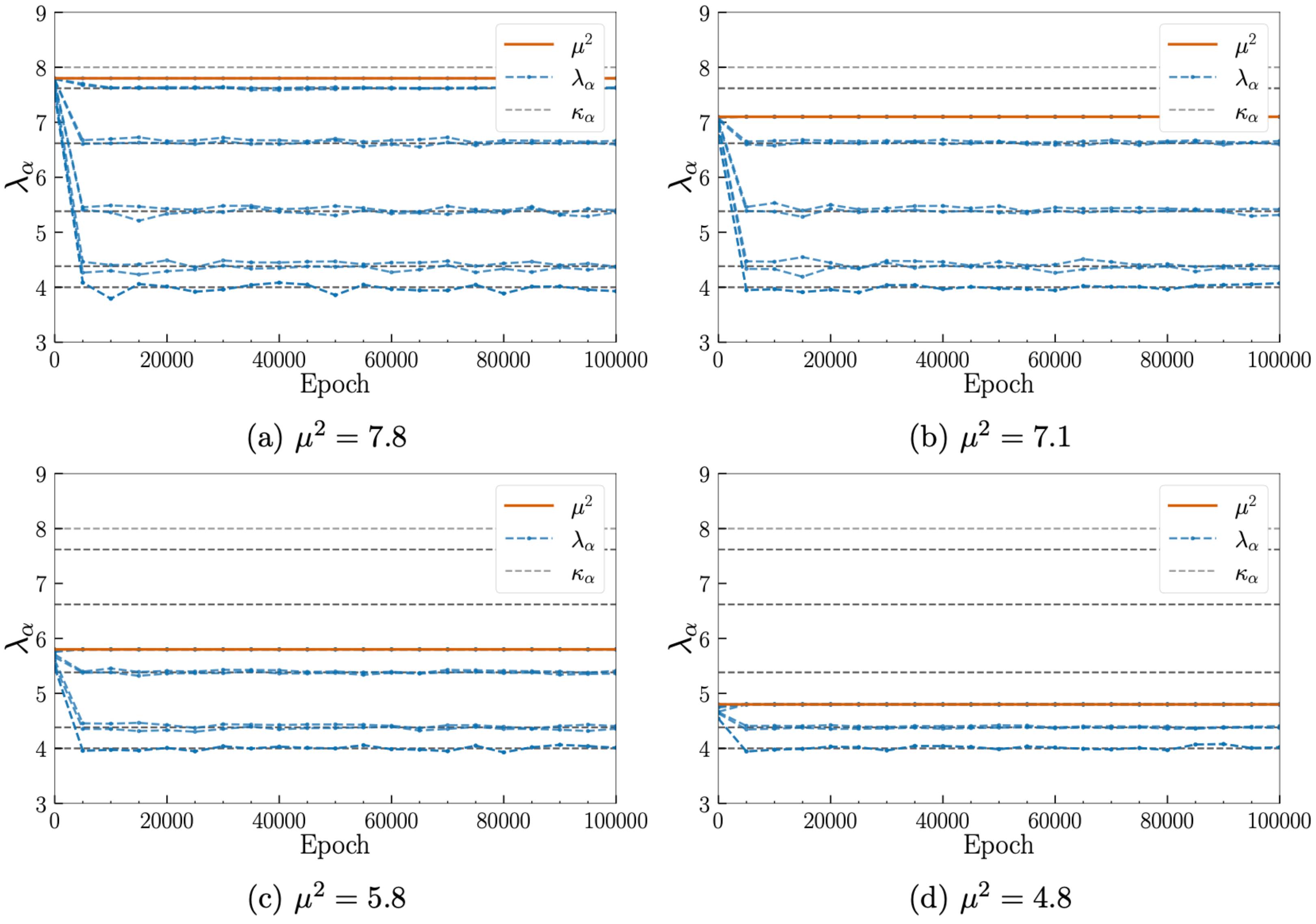
What if RBM mass μ^2 is wrongly chosen?

- analytical result: μ^2 acts as an ultraviolet regulator
- consider spectrum of induced quadratic operator on visible layer $\sum_{ij} \phi_i K_{ij} \phi_j$
- exact spectrum of target distribution is reproduced from infrared scale upwards
- RBM is an ultraviolet regulator

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

What if RBM mass μ^2 is wrongly chosen?

- example: 1D scalar LFT
  with $N_v = 10$ nodes
- exact spectrum (κ)
- reproduced by RBM (λ) from smallest eigenvalue upwards
- higher modes are suppressed at cut-off scale (μ^2)



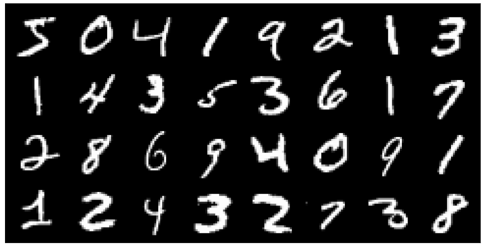
(a) $\mu^2 = 7.8$

(b) $\mu^2 = 7.1$

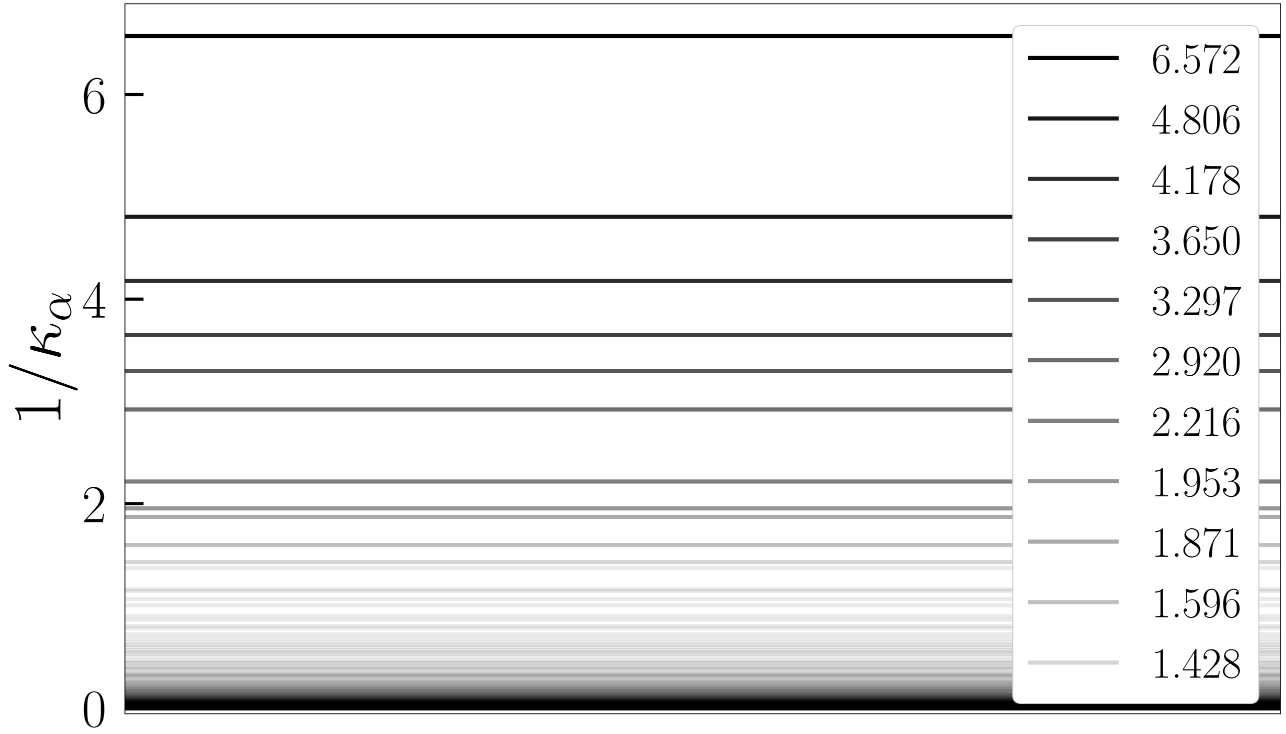
(c) $\mu^2 = 5.8$

(d) $\mu^2 = 4.8$

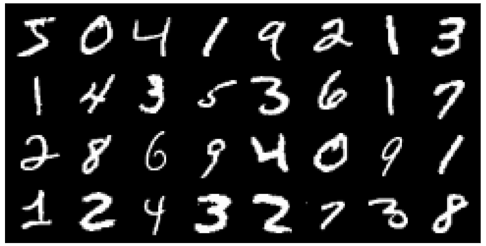
Application to MNIST data



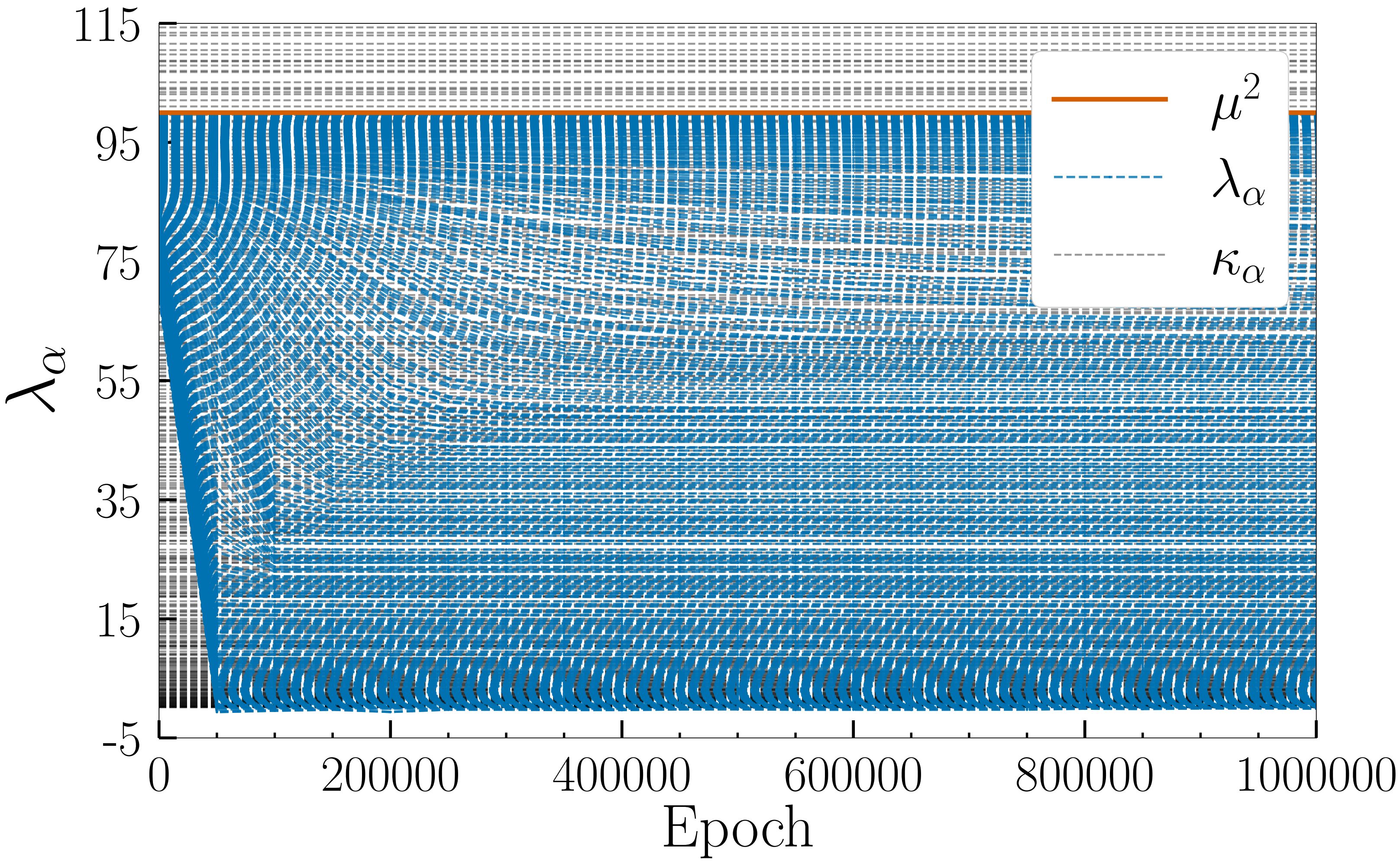
- standard data set to test ML methods
- 28x28 images of digits
- 748x784 correlation matrix
- inverse spectrum
- infrared safe
- ultraviolet divergent



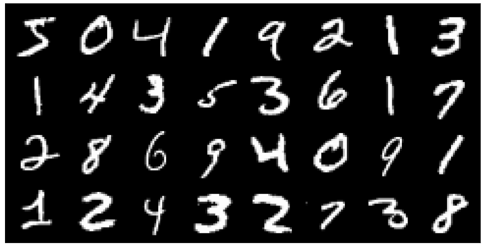
MNIST with fixed RBM mass



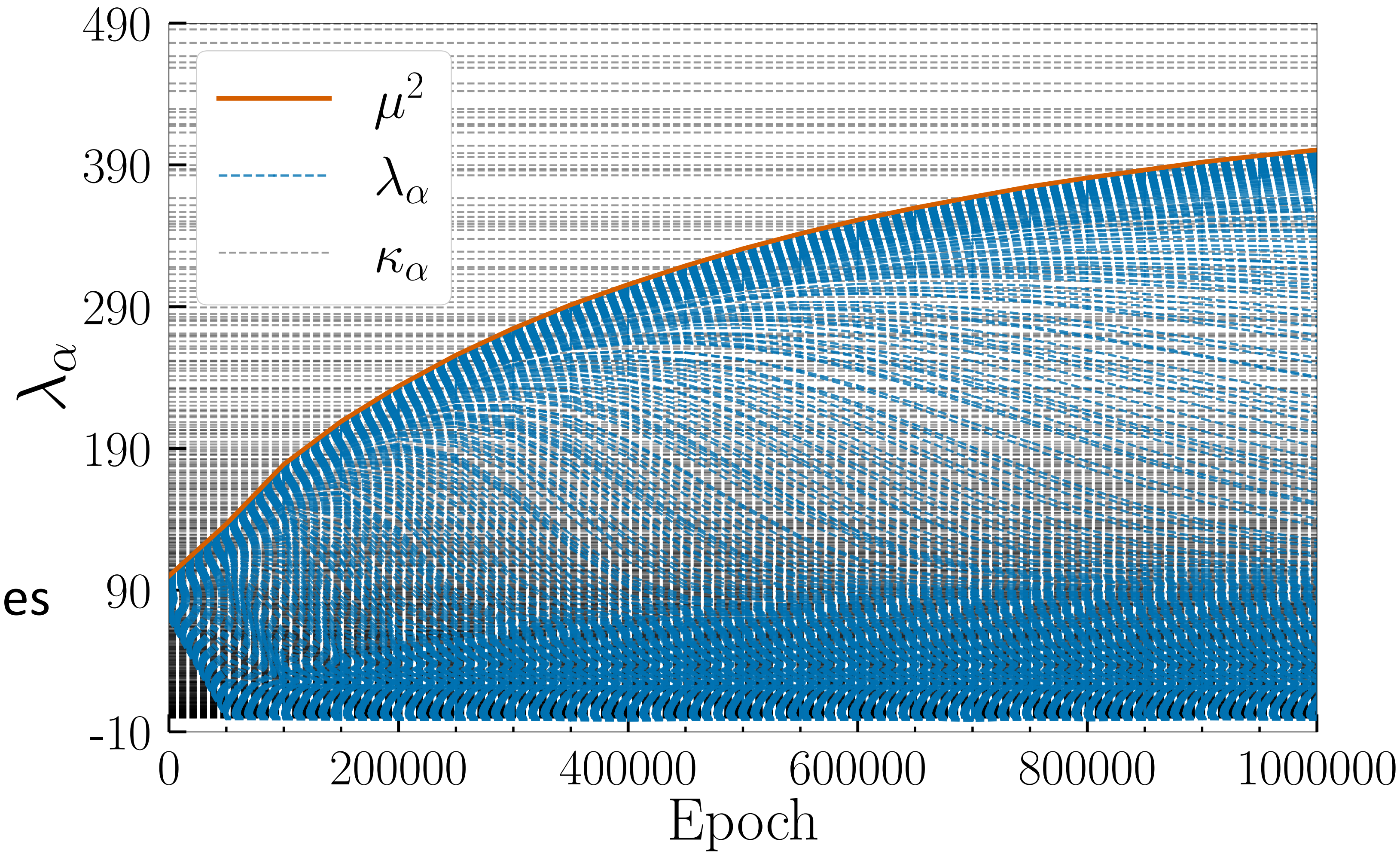
- $N_v = N_h = 748$
- fixed RBM mass $\mu^2 = 100$
- spectrum regulated
- infrared modes learned correctly



MNIST with dynamic RBM mass



- $N_v = N_h = 748$
- dynamical RBM mass μ^2 is learned as well
- spectrum regulated
- ultraviolet cut-off μ^2 increases to include more modes



MNIST with $N_h \leq N_v$

what is the effect of including incomplete spectrum?

removal of ultraviolet modes affects generative power

(a) $N_h = 784$

(b) $N_h = 225$

(c) $N_h = 64$

(d) $N_h = 36$

(e) $N_h = 16$

(f) $N_h = 4$

Interacting scalar field RBM

- Gaussian RBMs can learn Gaussian distributions
- in LFT language: need to include interactions
- various ways to do so, depending on properties of target distribution

- QFT-ML approach: add local potential terms on nodes, e.g. ϕ^4 terms

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

- standard RBM approach: use binary hidden layer $h_a = \pm 1$

Scalar-Bernoulli RBM: hidden binary nodes

- induced distribution $p(\phi) = \frac{1}{Z} \exp\left(-S_\phi(\phi) + \sum_a \sum_{n=1}^{\infty} c_n \psi_a^{2n} \right)$ with $\psi_a = \sum_i \phi_i w_{ia} - \eta_a$
- generates all-to-all interactions of all powers of ϕ
- at leading order in W same kinetic term as in Gaussian case
- example of quartic term (taking $\eta_a = 0$ for simplicity) $\sum_a \sum_{i,j,k,l} (\phi_i w_{ia}) (\phi_j w_{ja}) (\phi_k w_{ka}) (\phi_l w_{la})$
- highly non-local, very different from local field theories, analysis in preparation

Summary

- applications of ML to many problems in fundamental physics

- three examples:
  - phase classification and physics interpretation
  - ensemble generation with normalising flow
  - Restricted Boltzmann Machines as toy models to understand ML

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

Outline

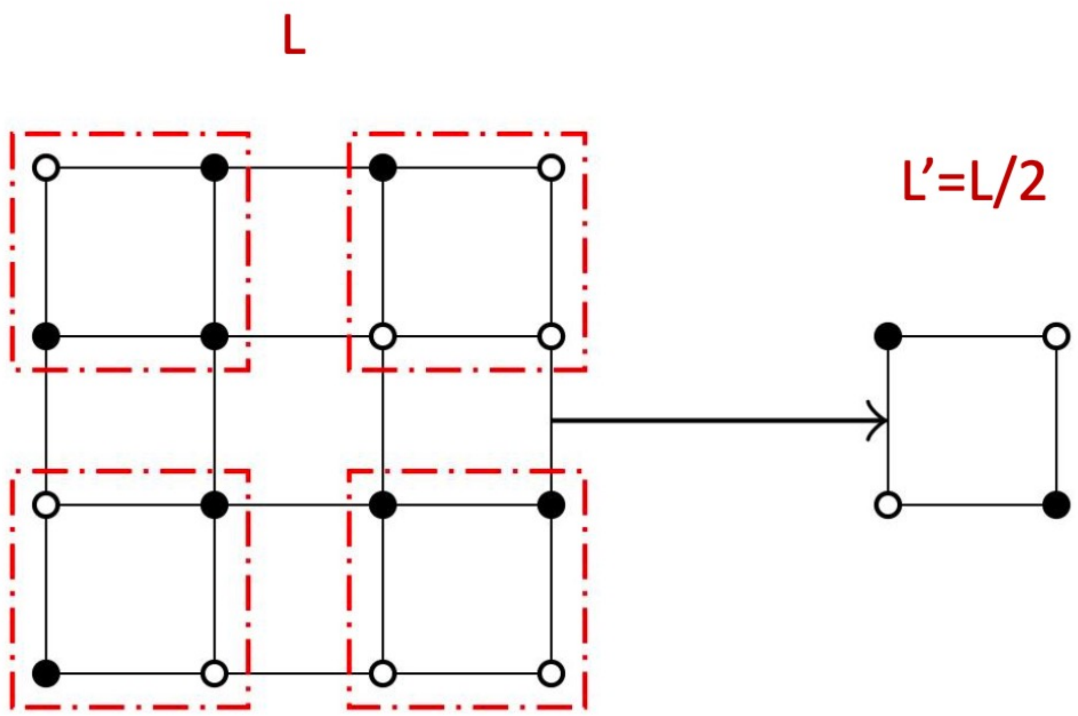
- classification of phases of matter: order-disorder transition
- generating ensembles: normalising flow
- quantum field-theoretical machine learning
- **inverse renormalisation group**

Renormalisation Group (RG)

- standard renormalisation group: coarse-graining,
  blocking transformation, integrating out degrees of freedom, …

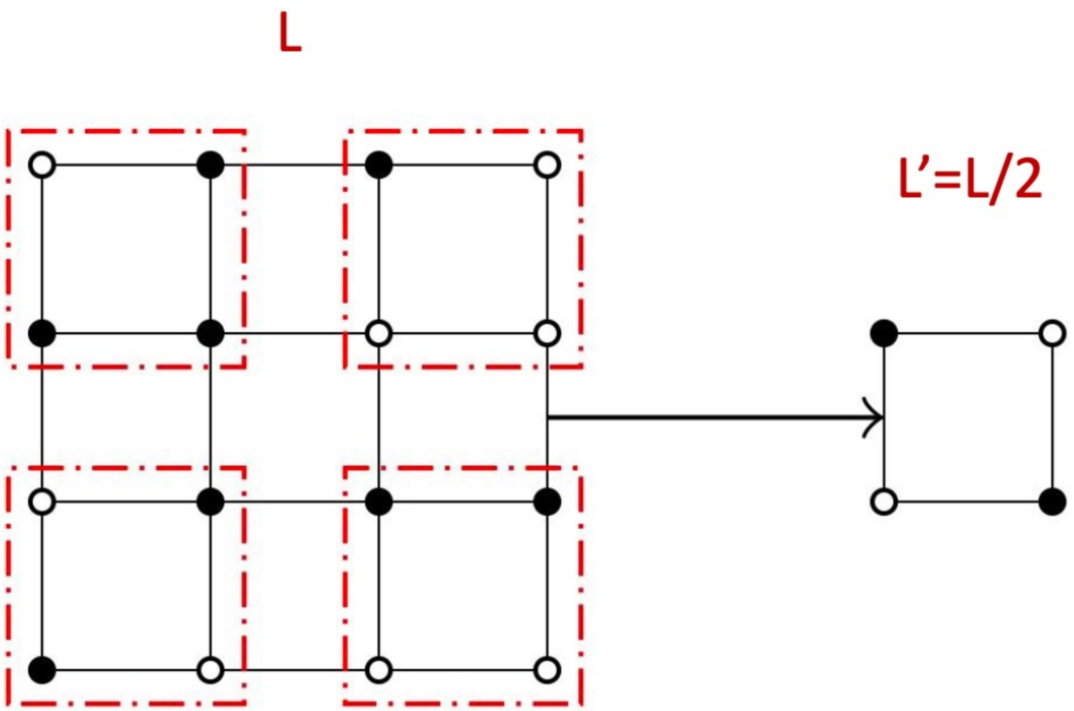
- Ising model: Kadanoff block spin
- majority rule
- reduction of degrees of freedom
- study critical scaling

- not invertible: semi-group



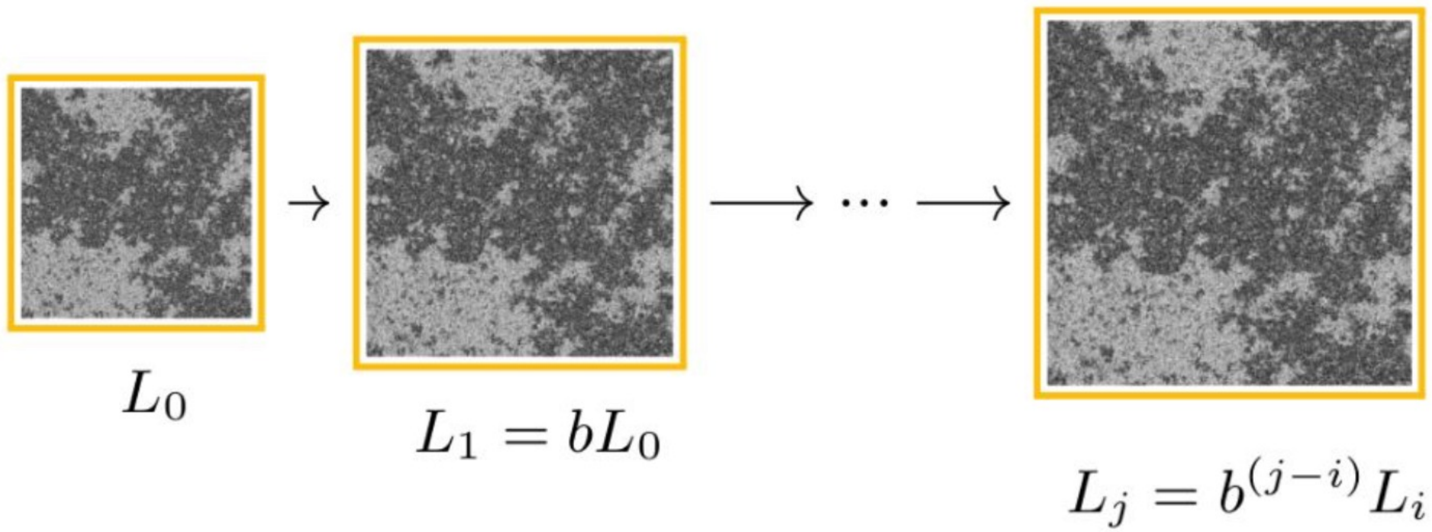
Renormalisation group

- generates flow in parameter space
- due to repeated blocking: run out of degrees of freedom
- need to start with large system to apply RG step multiple times
- large systems, close to a transition,
  suffer from critical slowing down



Inverse renormalisation group

- what if we could invert the RG?
- add degrees of freedom, fill in the 'details'
- inverse flow in parameter space
- can be applied arbitrary number of steps
- evade critical slowing down

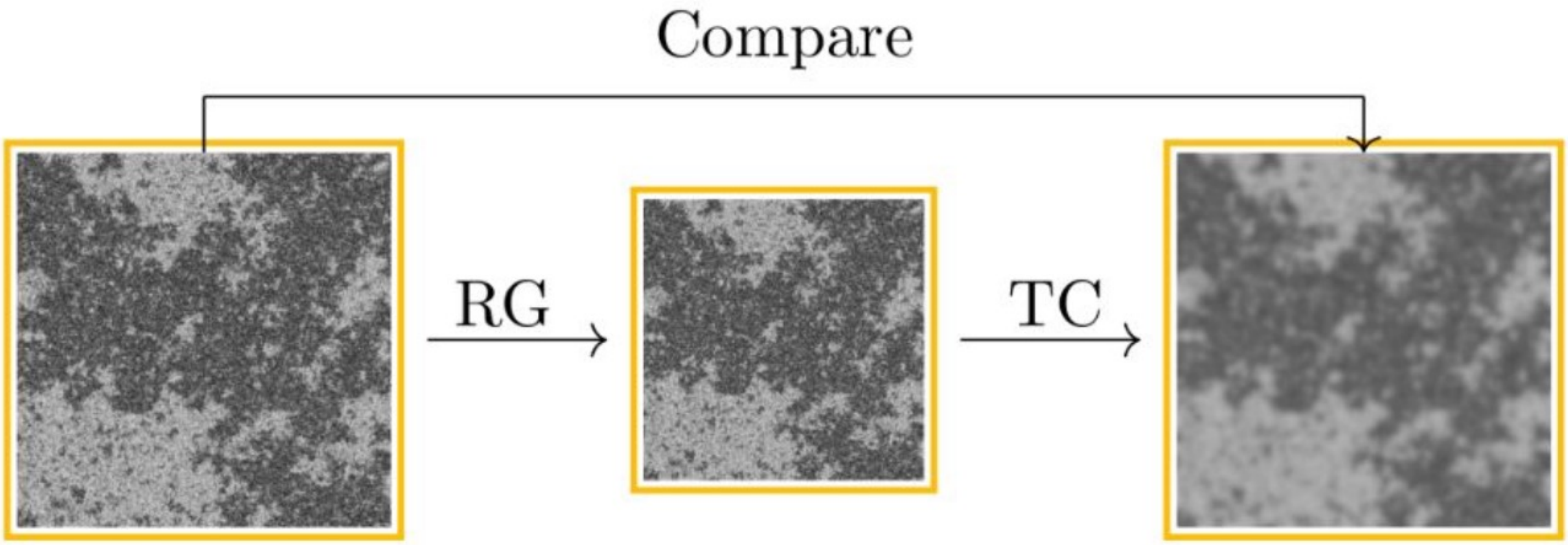


for Ising model: Inverse Monte Carlo Renormalization Group Transformations for Critical Phenomena, D. Ron, R. Swendsen, A. Brandt, Phys. Rev. Lett. 89, 275701 (2002)

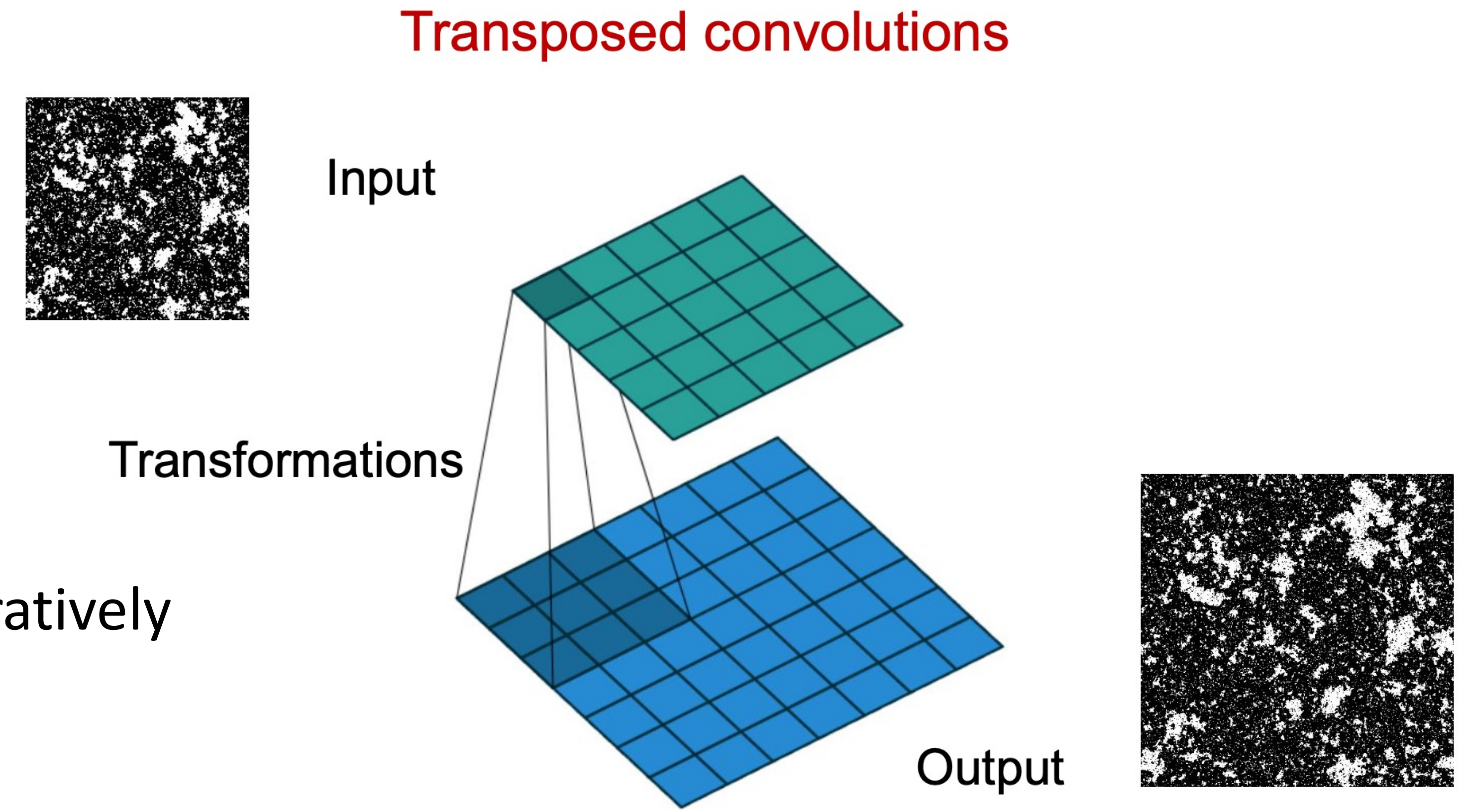
Inverse renormalisation group in quantum field theory, Bachtis, Aarts and Lucini
Phys. Rev. Lett. 128 (2022) 081603 [2107.00466 [hep-lat]]

How to devise an inverse transformation?

- new degrees of freedom should be introduced
- learn a set of transformations (*transposed convolutions*) to invert a standard RG step
- minimise difference between original and constructed configuration



Inverse renormalisation group



- local transformation
- apply inverse transformations iteratively
- evade critical slowing down
- generate flow in parameter space
- invariance at critical point

Application to φ^4 scalar field theory

- repeated steps
- locking in on critical point

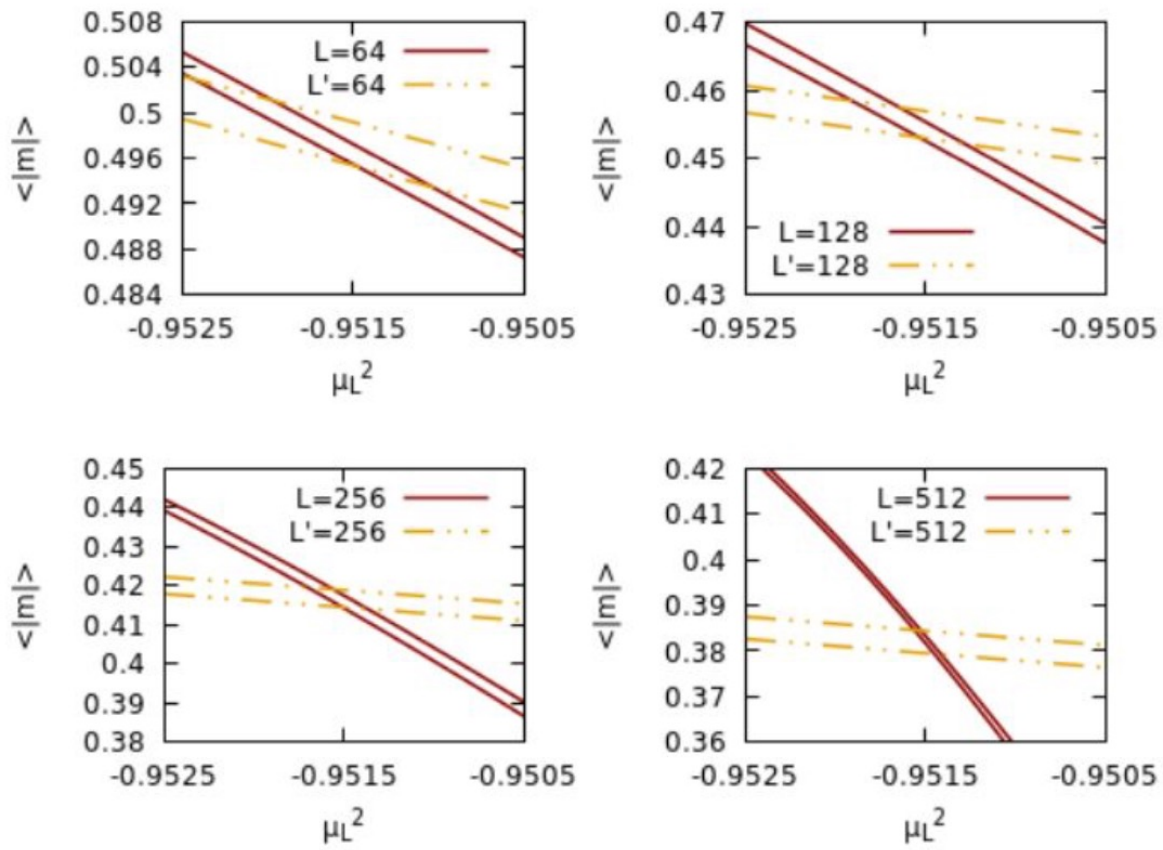


TABLE I. Values of the critical exponents γ/ν and β/ν. The original system has lattice size $L = 32$ in each dimension and its action has coupling constants $\mu_L^2 = -0.9515$, $\lambda_L = 0.7$, and $\kappa_L = 1$. The rescaled systems are obtained through inverse renormalization group transformations.

L_i/L_j	32/64	32/128	32/256	32/512	64/128	64/256	64/512	128/256	128/512	256/512
γ/ν	1.735(5)	1.738(5)	1.741(5)	1.742(5)	1.742(5)	1.744(5)	1.744(5)	1.745(5)	1.745(5)	1.746(5)
β/ν	0.132(2)	0.130(2)	0.128(2)	0.128(2)	0.128(2)	0.127(2)	0.127(2)	0.126(2)	0.126(2)	0.126(2)

Application to φ^4 scalar field theory

- start with lattice of size 32^2 and apply IRG steps repeatedly
- $32^2 \to 64^2 \to 128^2 \to 256^2 \to 512^2$
- IRG flow towards critical point
- extract critical exponents
  γ/υ and β/υ from comparison
  between two volumes
- constructed a large (512^2) lattice
  very close to criticality
  without critical slowing down