From lattice field theory to machine learning and back

Gert Aarts

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

Swansea University
Prifysgol Abertawe

EuCAIFCon Apr 2024

Swansea ML-LFT group

Chanju Park, Diaa Habibi, Shiyang Chen, GA, Biagio Lucini, Matteo Favoni cdt-aimlac.org/

AIMLAC

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

- 5 cohorts of O(11) PhD students each
- first cohort has graduated
- final cohort of 17 students started in Oct 2023

- 60% particle physics/astronomy
- 25% computer science/maths
- 15% health

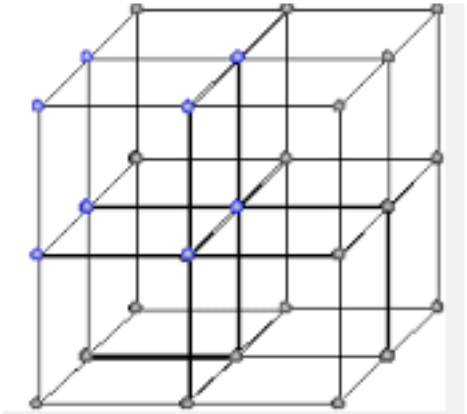
Prifysgol Aberystwyth University · Prifysgol Bangor University · University of Bristol · Cardiff University Prifysgol Caerdydd · Swansea University Prifysgol Abertawe · Supercomputing Wales Uwchgyfrifiadura Cymru

Lattice field theory (LFT) ←→ ML

- many fascinating connections between ML and statistical physics/field theory
- what can we learn that is relevant for → or ← direction

lattice field theory: very loosely speaking

- many fluctuating degrees of freedom on a lattice/graph <-> Markov random fields
- thermalisation and non-equilibrium dynamics <-> learning
- stochastic quantisation and Langevin dynamics <-> diffusion models
- renormalisation group
- …



Conceptual and practical questions

can experience in quantum field theory help in understanding ML and vice versa?

- quantum field-theoretic machine learning

  Dimitrios Bachtis, GA, Biagio Lucini, Phys. Rev. D 103 (2021) 074510 [2102.09449 [hep-lat]]

- scalar field restricted Boltzmann machines as an ultraviolet regulator

  Chanju Park, Biagio Lucini, GA, Phys. Rev. D 109 (2024) 034521 [2309.15002 [hep-lat]]

- stochastic quantisation and diffusion models

  Lingxiao Wang, GA, Kai Zhou, JHEP (to appear) [2309.17082 [hep-lat]]

  NeurIPS 2023 2311.03578 [hep-lat]

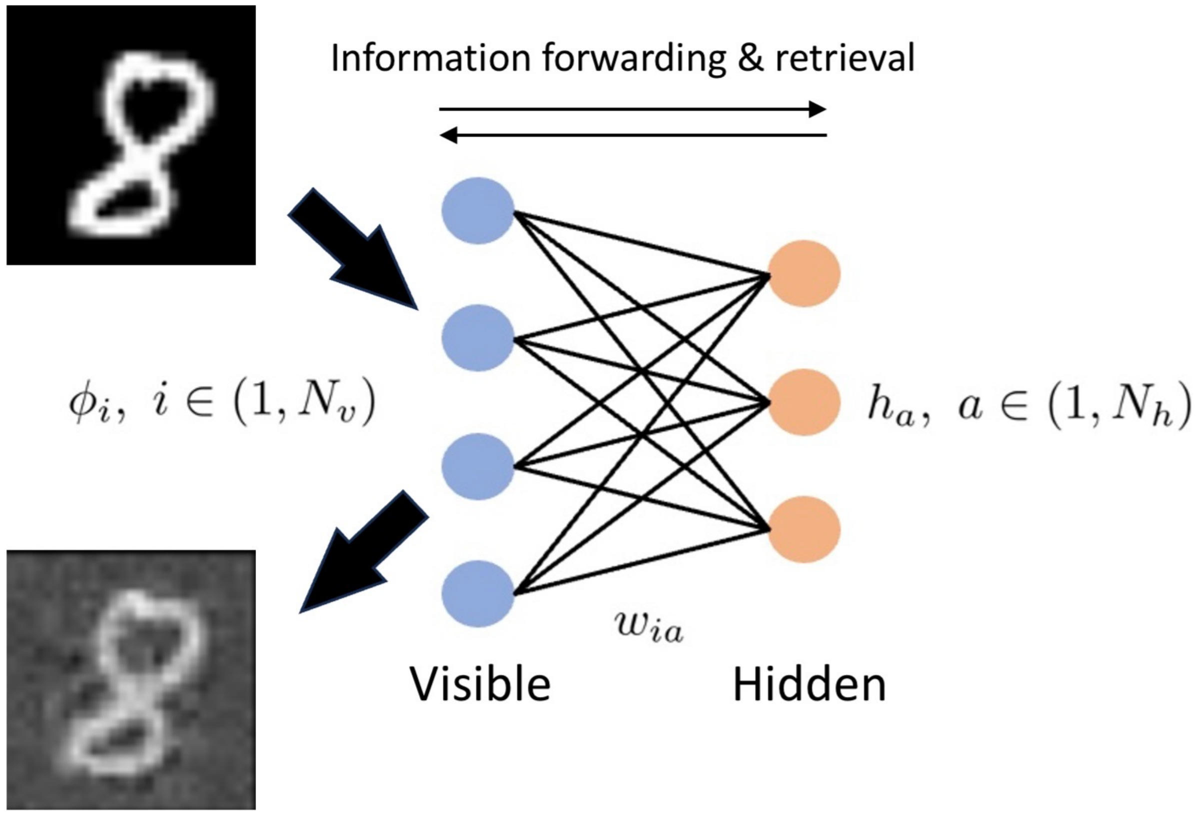
Topic 1

scalar field restricted Boltzmann machines as an ultraviolet regulator

Chanju Park, Biagio Lucini, GA

Phys. Rev. D 109 (2024) 034521 [2309.15002 [hep-lat]]

Restricted Boltzmann Machine: generative network

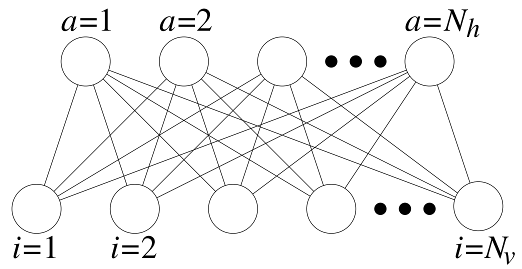


- energy-based method
- probability distribution
- binary or continuous d.o.f.

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

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

Scalar field RBM



- treat RBM as a lattice field theory with action

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

- only quadratic terms, add interactions later, e.g ϕ^4 terms
- learn weight matrix w_{ia} and bias η_a (put to 0 below)
- induced distribution on visible layer

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

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?
- learn from data or directly from known distribution
- simplest case: target theory = LFT of free scalar field in 1 or 2d, $K^{\phi} \approx p^2 + m^2$

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

Infinite # of solutions ($N_h = N_v$)

infinite # solutions for weight matrix, use that $\mathcal{K}$ 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

- realisation depends on initialisation
- non-observable degeneracy due to internal symmetry on hidden layer

$$K_{ij} = \mu_i^2 \delta_{ij} - \sigma^2 \sum_a w_{ia} w_{aj}^T$$

Dependence on N_h and μ^2

- what if $N_h < N_v$? role of hyperparameter μ^2?

- use LFT insights: target distribution is scalar field theory
- method: SVD of weight matrix $W = U\,\Xi\,V^T$

- quadratic kernel $K = \mu^2 \mathbb{1} - \sigma_h^2 U \Xi \Xi^T U^T = U\left[\mu^2 \mathbb{1} - \sigma_h^2 \Xi \Xi^T\right] U^T \equiv U D_K U^T$

- eigenvalues of quadratic kernel

$$D_K = \mathrm{diag}\big(\underbrace{\mu^2 - \sigma_h^2 \xi_1^2, \mu^2 - \sigma_h^2 \xi_2^2, \ldots, \mu^2 - \sigma_h^2 \xi_{N_h}^2}_{N_h}, \underbrace{\mu^2, \ldots, \mu^2}_{N_v - N_h}\big)$$

- both N_h and μ^2 act as **ultraviolet regulators** train RBM with persistent contrastive divergence

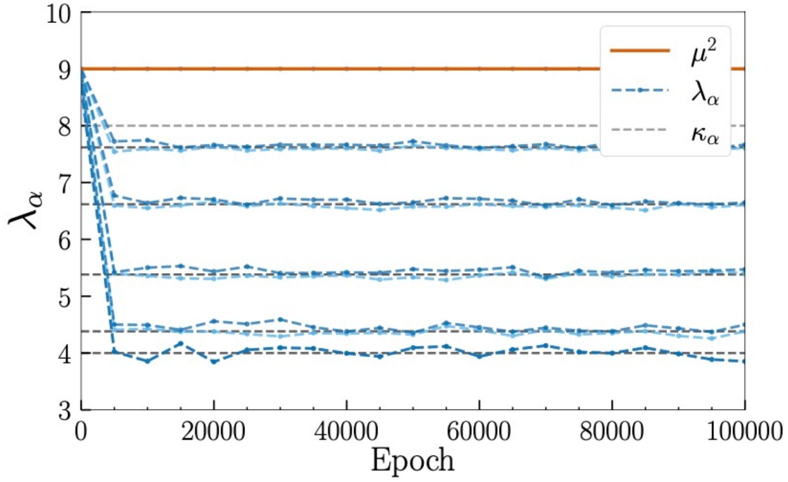
What if $N_h < N_v$?

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

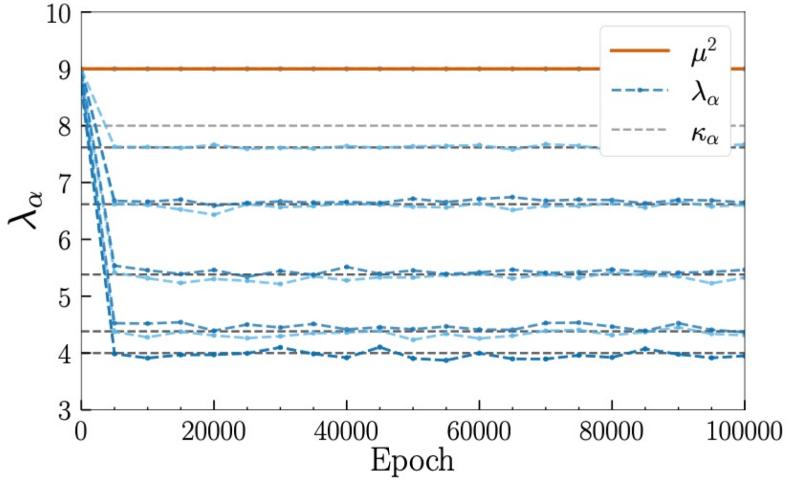
- exact spectrum:
  $$\kappa \sim p^2 + m^2$$

- reproduced by RBM (λ) from smallest eigenvalue upwards

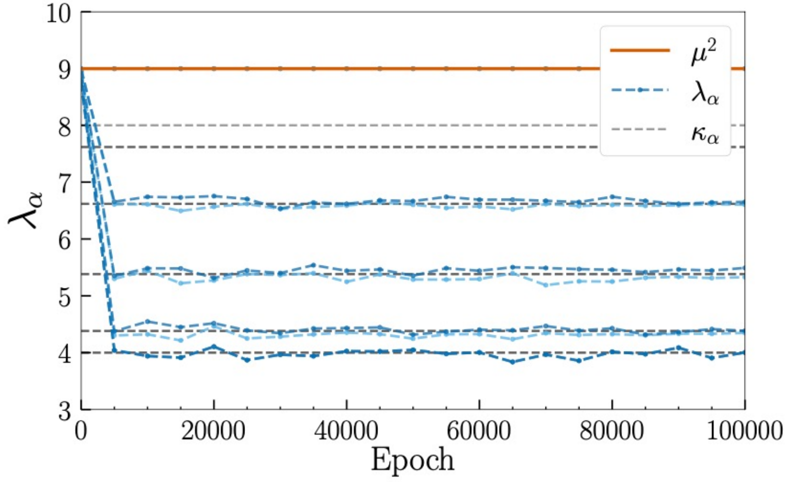
- higher modes are moved to cut-off scale (μ^2)



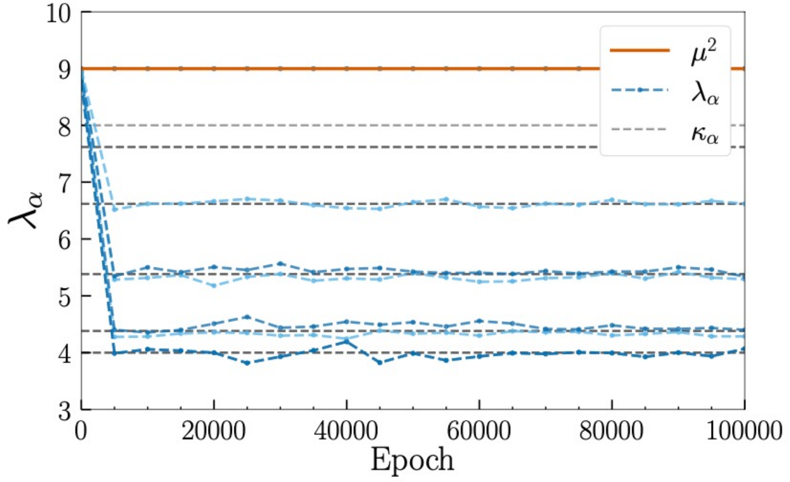
(a) $N_h = 9$



(b) $N_h = 8$



(c) $N_h = 7$



(d) $N_h = 6$

What if RBM mass $\mu^2 < \lambda_{max}$?

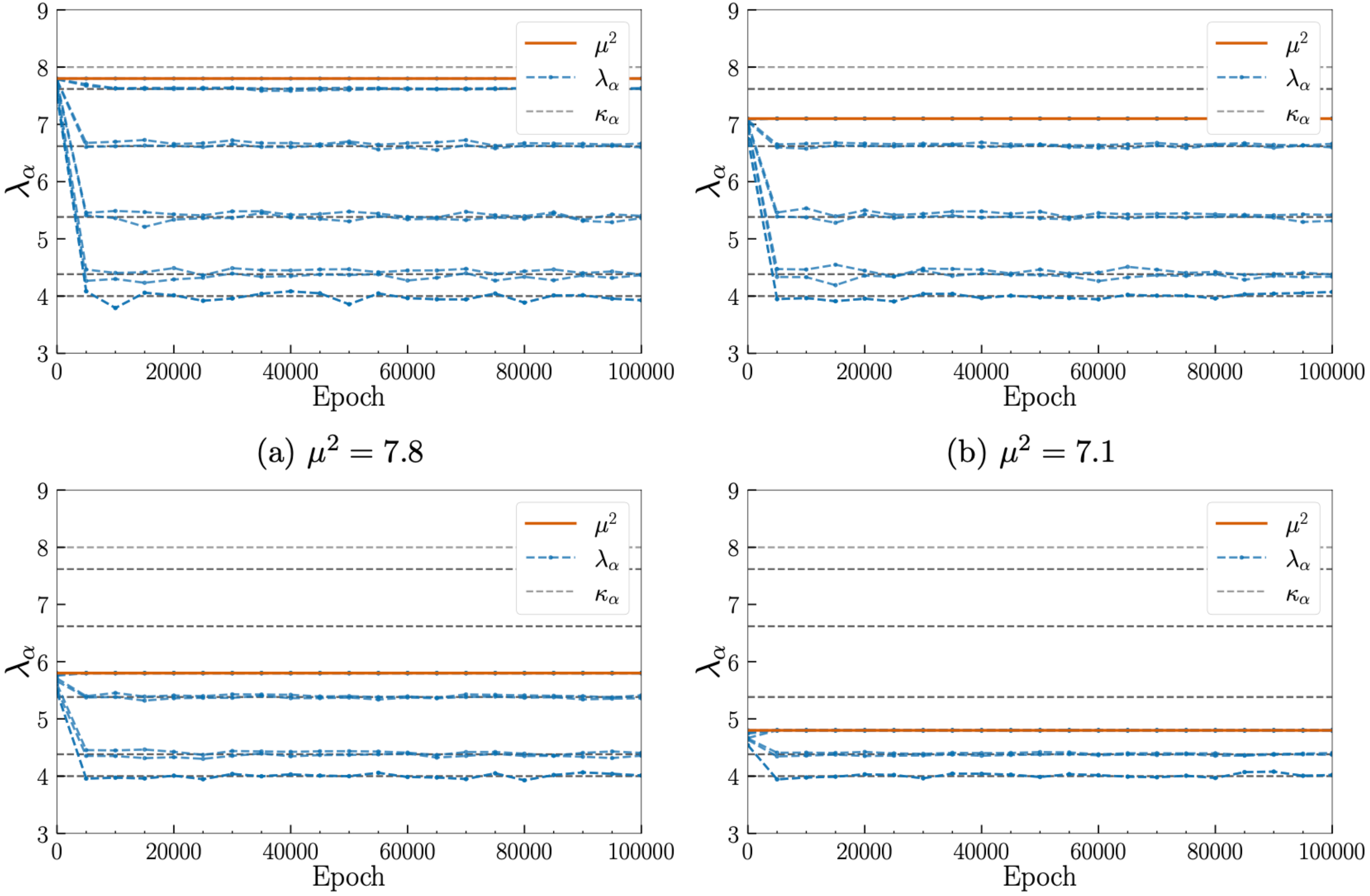
- example: scalar LFT
  with $N_v = 10$ nodes

- exact spectrum:

$$\kappa \sim p^2 + m^2$$

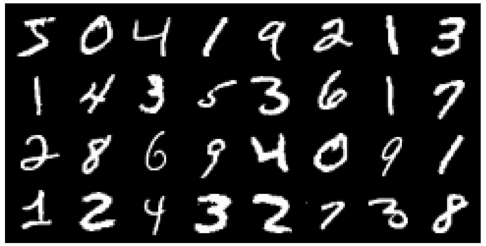
- reproduced by RBM (λ) from smallest eigenvalue upwards

- higher modes are suppressed at cut-off scale (μ^2)

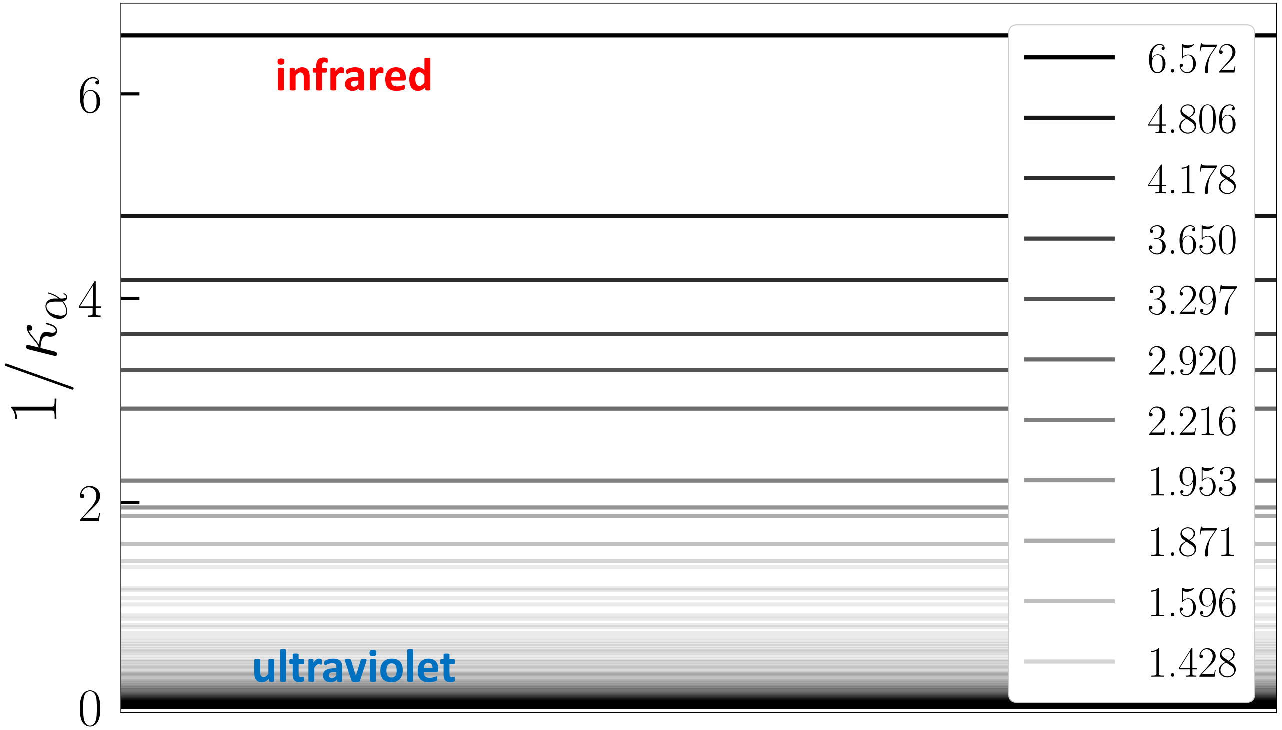


(a) $\mu^2 = 7.8$ (b) $\mu^2 = 7.1$ (c) $\mu^2 = 5.8$ (d) $\mu^2 = 4.8$

RBM as ultraviolet regulator

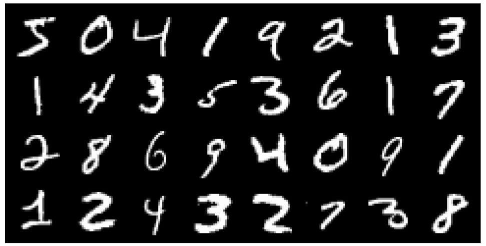


- relevant for “real” data sets? MNIST: 28x28 images of digits
- compute spectrum of two-point correlator $K_{ij}^{-1} = < \phi_i \phi_j >_{data}$
- inverse spectrum $1/\kappa$
- infrared safe
- ultraviolet divergent

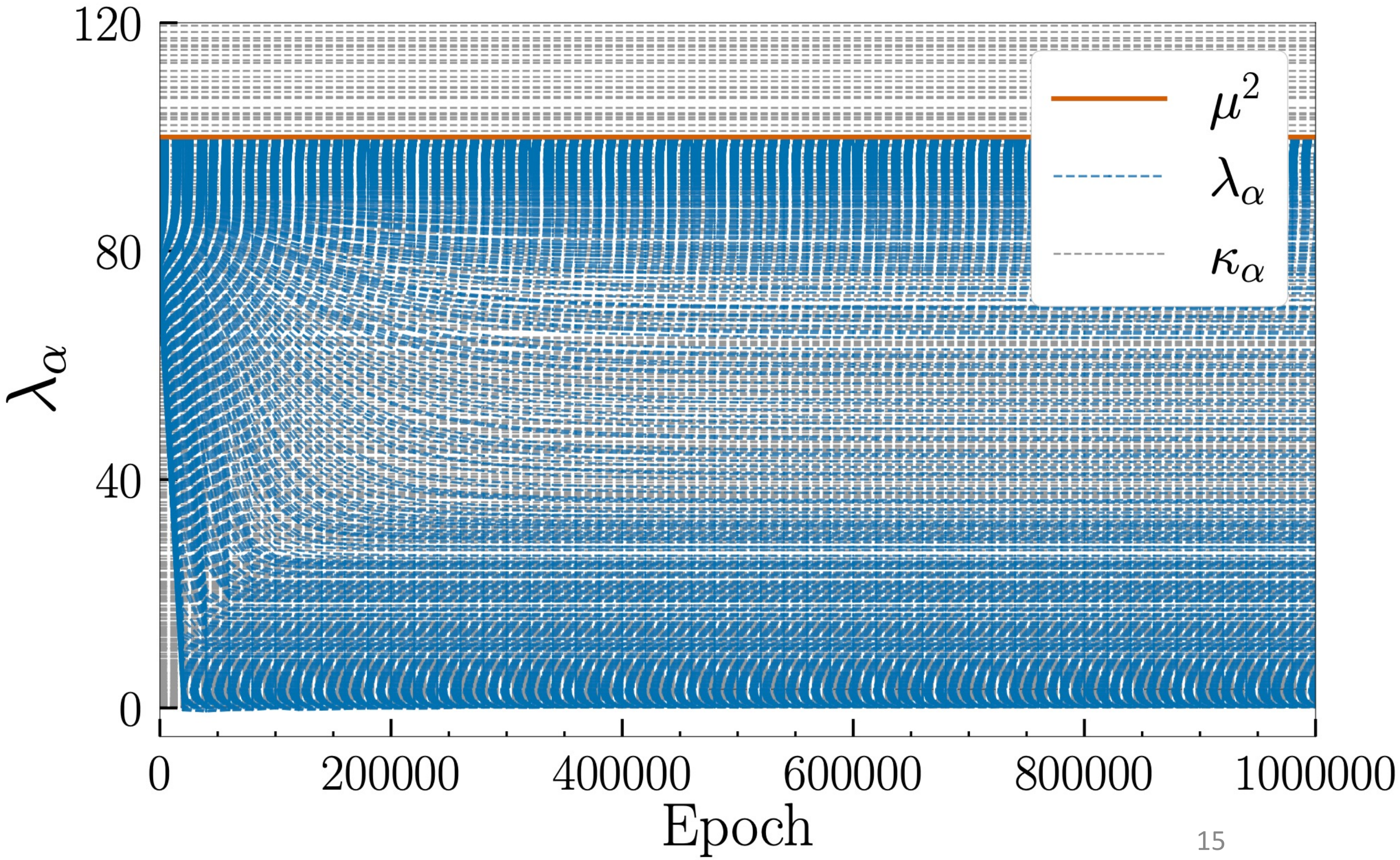


784 eigenvalues

MNIST with fixed RBM mass



- $N_v = N_h = 784$
- fixed RBM mass $\mu^2 = 100$
- spectrum regulated
- infrared modes learned correctly



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) $N_h = 4$

Topic 2

stochastic quantisation and diffusion models

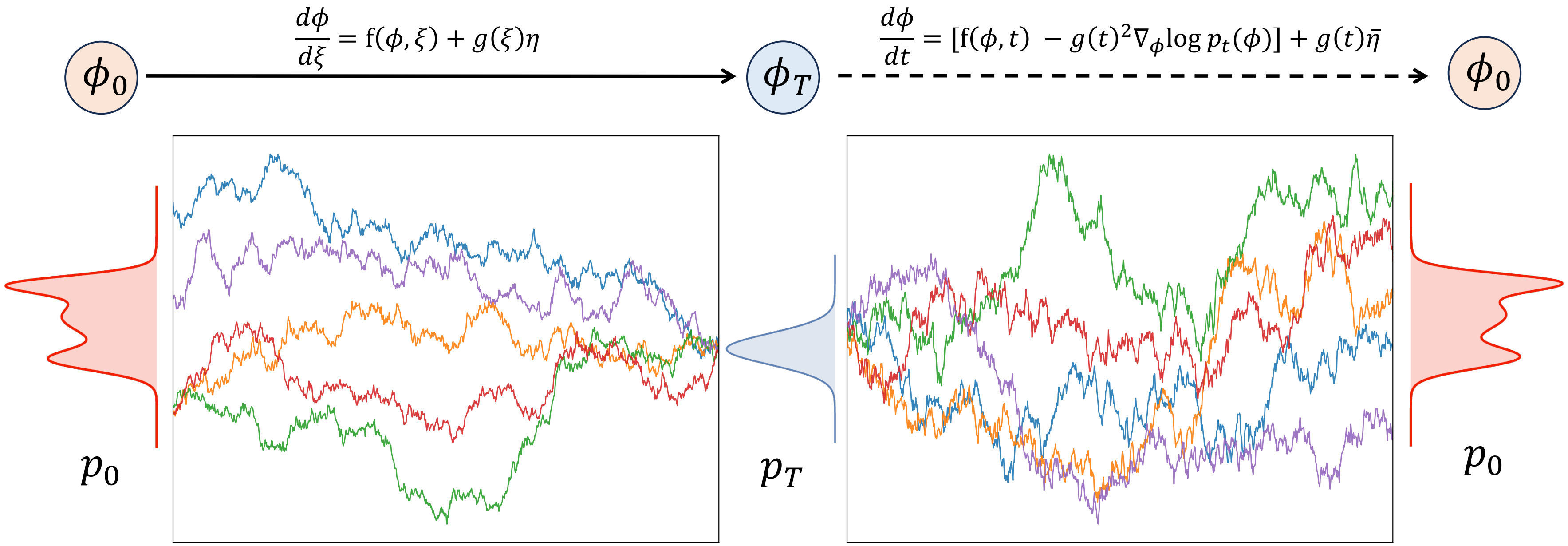
Lingxiao Wang, GA, Kai Zhou

JHEP (to appear) [2309.17082 [hep-lat]]

NeurIPS 2023 2311.03578 [hep-lat]

Diffusion models

- solve stochastic process with a particular drift/force/score
- drift is learnt during forward diffusion process, starting from data
- new configurations are generated via backward process using learnt drift



Stochastic quantisation

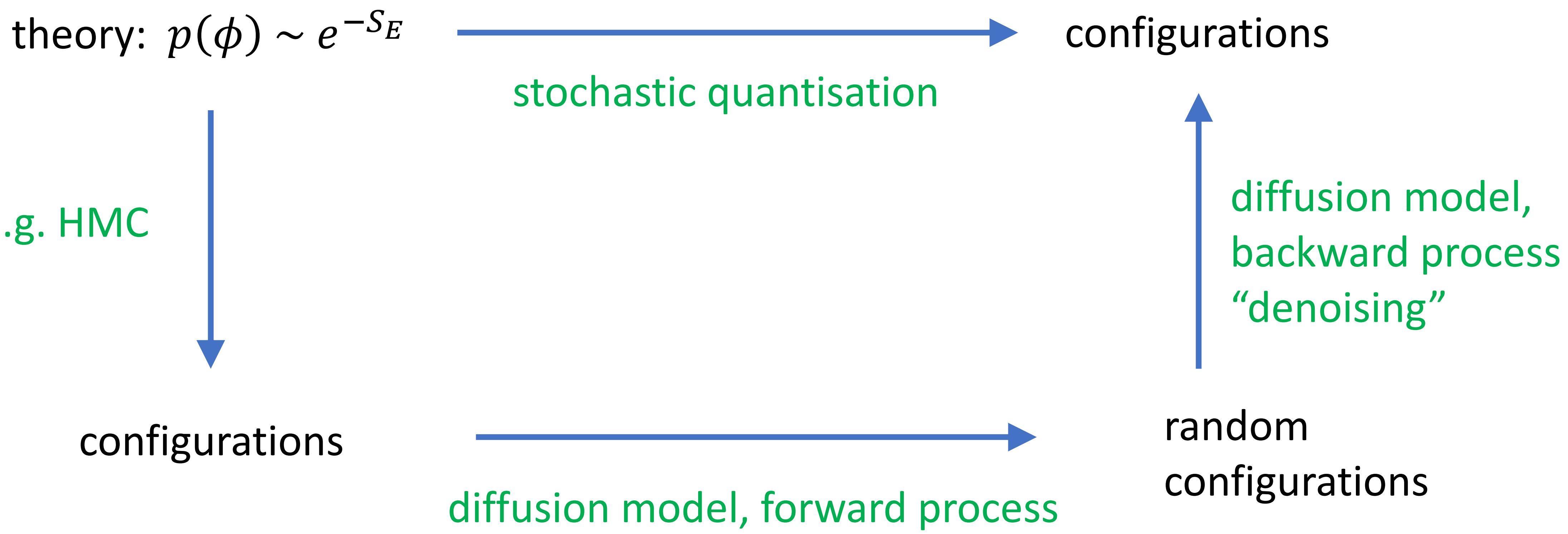
- ideas well-known in quantum field theory: stochastic quantisation (Parisi & Wu 1980)
- path integral quantisation via a stochastic process in fictitious time

$$\frac{\partial\phi(x,\tau)}{\partial\tau} = -\frac{\delta S_E[\phi]}{\delta\phi(x,\tau)} + \eta(x,\tau) \qquad \langle\eta(x,\tau)\eta(x',\tau')\rangle = 2\alpha\delta(x-x')\delta(\tau-\tau')$$

- equilibrium solution ($\tau \to \infty$): distribution $p(\phi) \sim e^{-S_E}$
- convergence guaranteed for real actions due to properties of Fokker-Planck equation
- create samples from Euclidean path integral
- applied to non-abelian gauge theories and QCD in 1980s, but superseded by other methods such as Hybrid Monte Carlo (HMC) [stepsize dependence, efficiency]

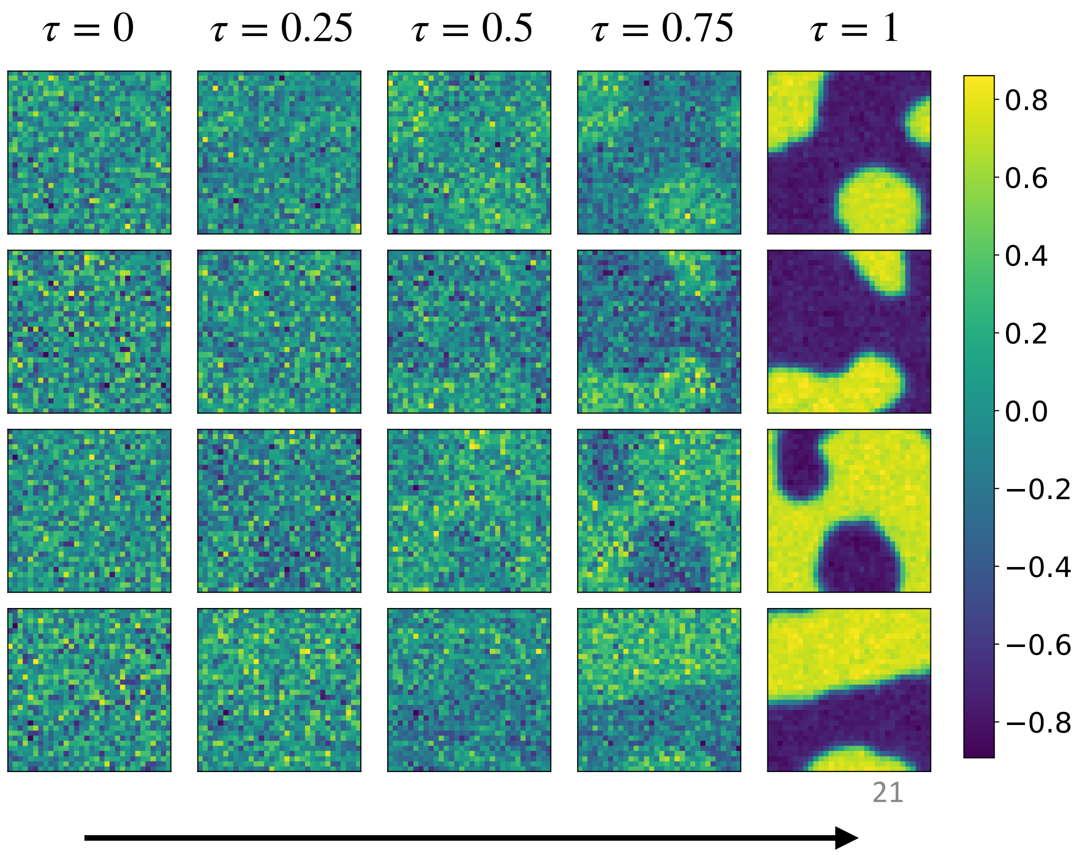
Stochastic quantisation and diffusion models

- diffusion models as an alternative approach to stochastic quantisation



Diffusion model for 2d ϕ^4 scalar theory

- 32^2 lattice, choice of action parameters in symmetric and broken phase
- training data set generated using Hybrid Monte Carlo (HMC)
- variance expanding DM trained using U-Net architecture



generating configurations:

- broken phase
- “denoising” (backward process)
- large-scale clusters emerge, as expected

Diffusion model for 2d ϕ^4 scalar theory

generating configurations in symmetric phase

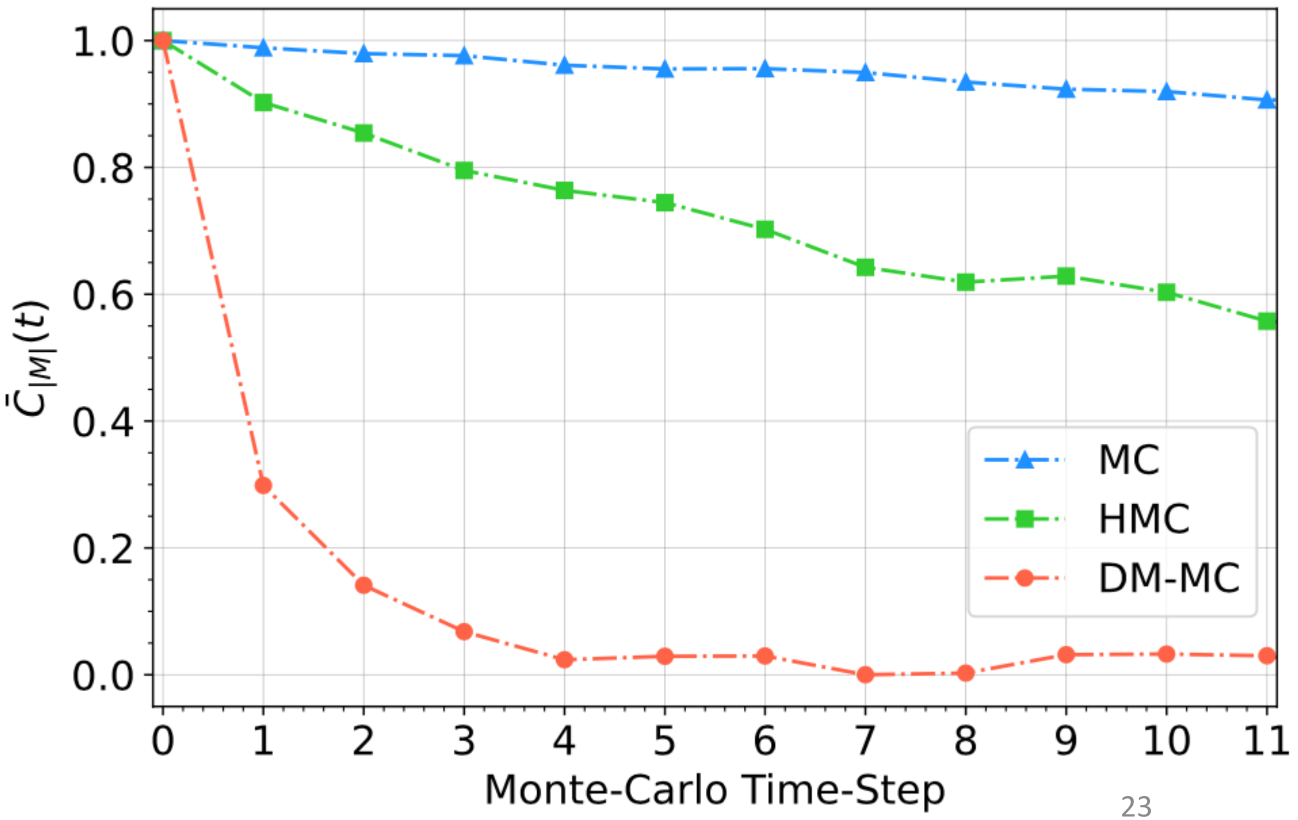
- compute magnetisation $< M >$, susceptibility χ_2, Binder cumulant U_L
- compare with test HMC data set (with same statistics)

data-set	$\langle M \rangle$	χ_2	U_L
Training (HMC)	0.0012± 0.0007	2.5160 ± 0.0457	0.1042 ± 0.0367
Testing (HMC)	0.0018 ± 0.0015	2.4463 ± 0.1099	-0.0198 ± 0.1035
Generated (DM)	0.0017± 0.0015	2.4227 ± 0.1035	0.0484 ± 0.0959

- good agreement is observed

Diffusion model for 2d ϕ^4 scalar theory

- auto-correlation time (first comparison)
- normalised auto-correlation function



overall:

- proof of principle
- expected results obtained
- need to do detailed comparison of precision, speed and scalability

Summary and outlook

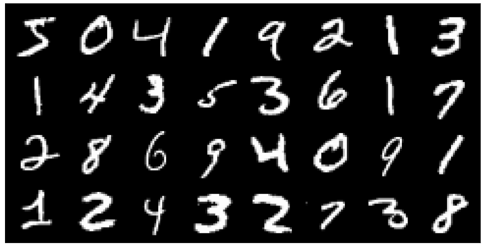
- can experience in quantum field theory help in understanding ML and vice versa?

- two examples of interplay between lattice field theory and ML
  - scalar field RBM as ultraviolet regulator
  - stochastic quantisation and diffusion models

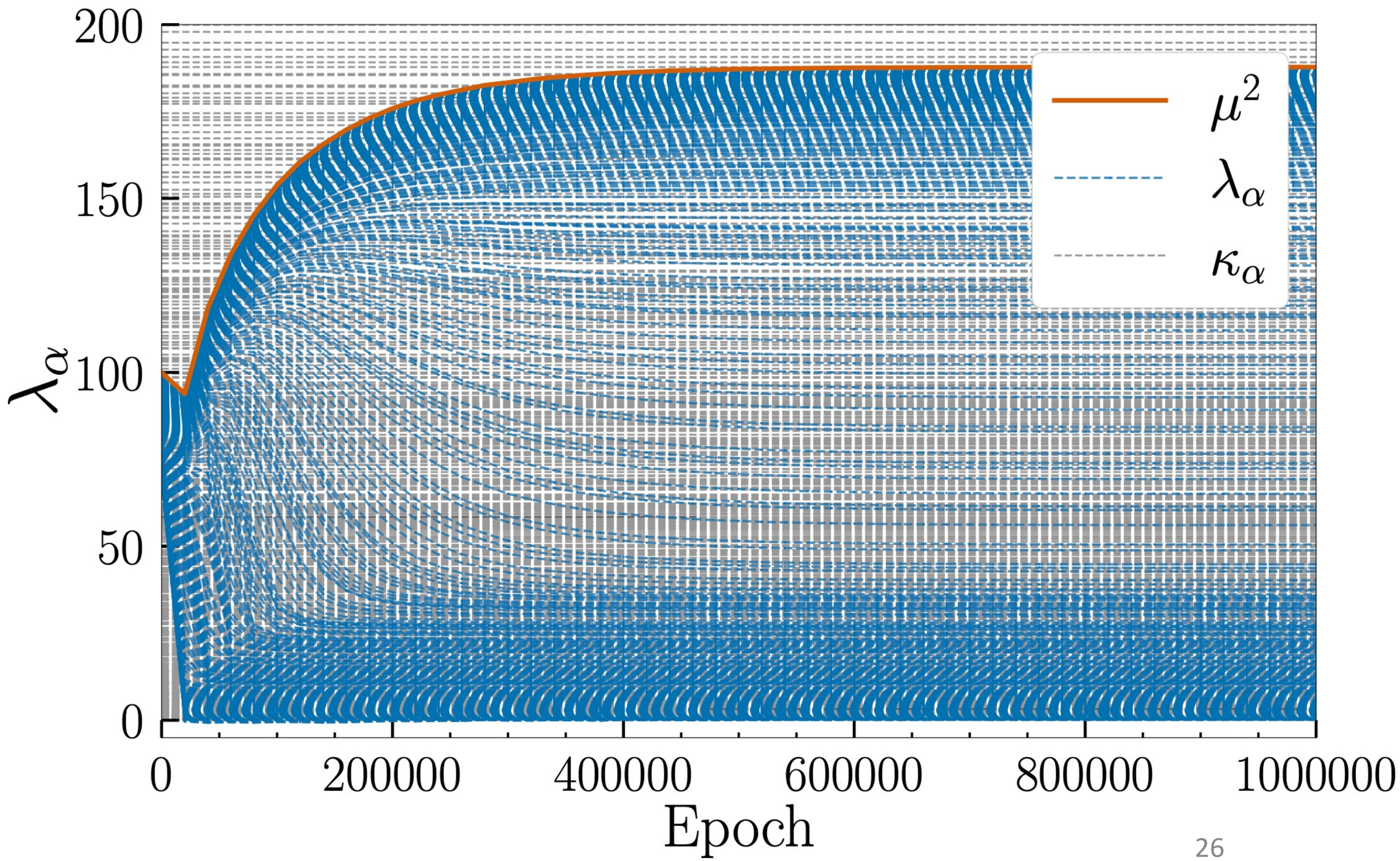
- interplay between statistical/lattice field theory and ML
- many directions to explore

Back-up

MNIST with dynamic RBM mass



- $N_v = N_h = 784$
- dynamical RBM mass μ^2 is learned as well
- spectrum regulated
- ultraviolet cut-off μ^2 increases to include more modes



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 standard field theories, analysis in preparation

Stochastic quantisation: complex actions

- approach not limited to real-valued distributions/actions
- extend Langevin process to complex manifold: complex Langevin dynamics (Parisi 1981)
- complexify d.o.f.: real scalar → complex scalar, $U \in SU(N) \to U \in SL(N, \mathbb{C})$

- convergence not guaranteed, no general solution of Fokker-Planck equation
- a posteriori justification (GA, Seiler, Stamatescu 2009, Nagata, Nishimura, Shimasaki 2016)

- applied to problems at finite (baryon) chemical potential (GA & Stamatescu 2008, Aarts 2009)
- success in some theories, QCD remains difficult (e.g. Sexty 2019)
- introductory lectures: GA, J.Phys.Conf.Ser. 706 (2016) 2, 022004 [1512.05145 [hep-lat]]