ML in fundamental physics

Gert Aarts

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

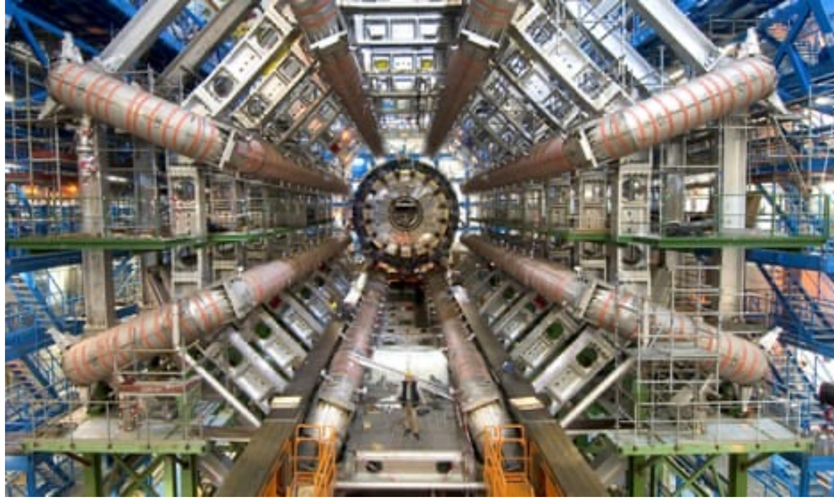
Swansea University
Prifysgol Abertawe

FONDAZIONE BRUNO KESSLER

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

Distributed Algorithms CDT Liverpool, Nov 29, 2023

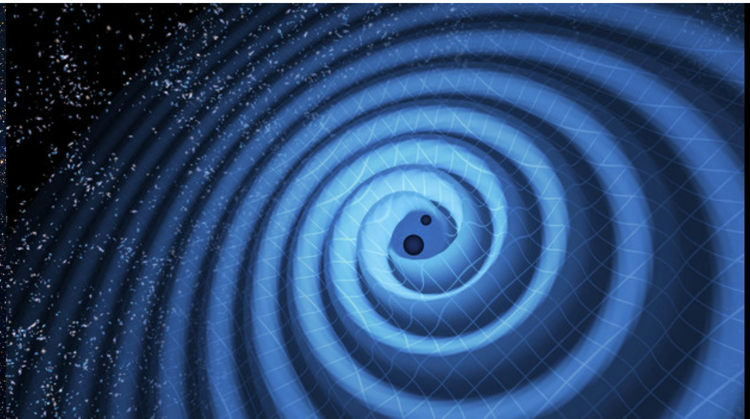
Fundamental physics

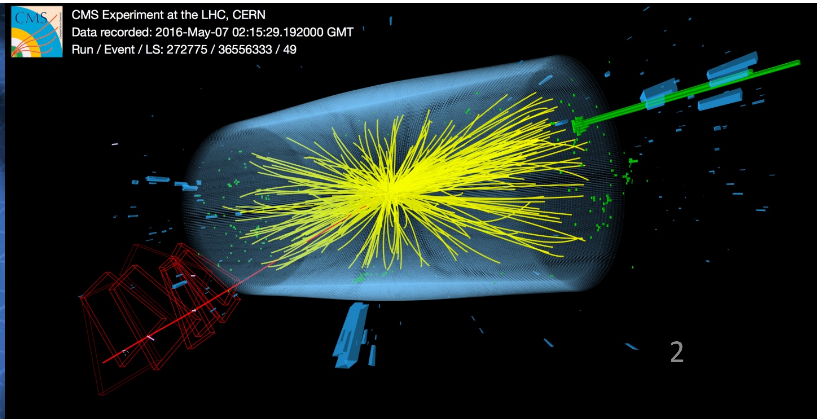


- understood to mean **particle physics** and **astronomy**

- big questions about the very small: elementary particles, quarks and gluons and the very large: galaxies, black holes, gravitational waves, the universe

- no immediate application in society, but long-term interest (and impact!) in understanding nature



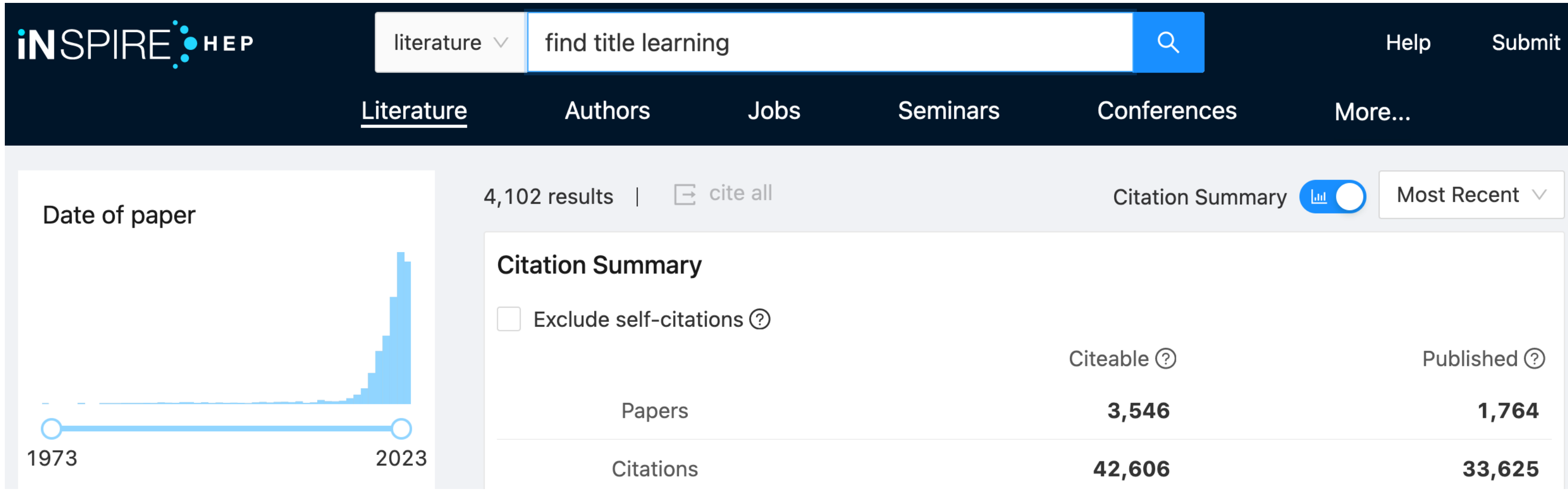


High-energy physics (HEP)

- past seven years or so have seen a rapid rise of applications of ML

- of course, ML has been around for quite some time, especially in experimental particle physics

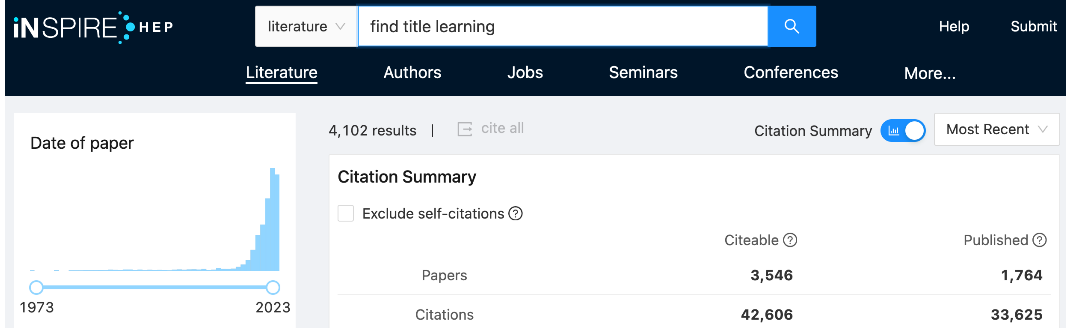
- nevertheless, there is an **exponential** increase in activity

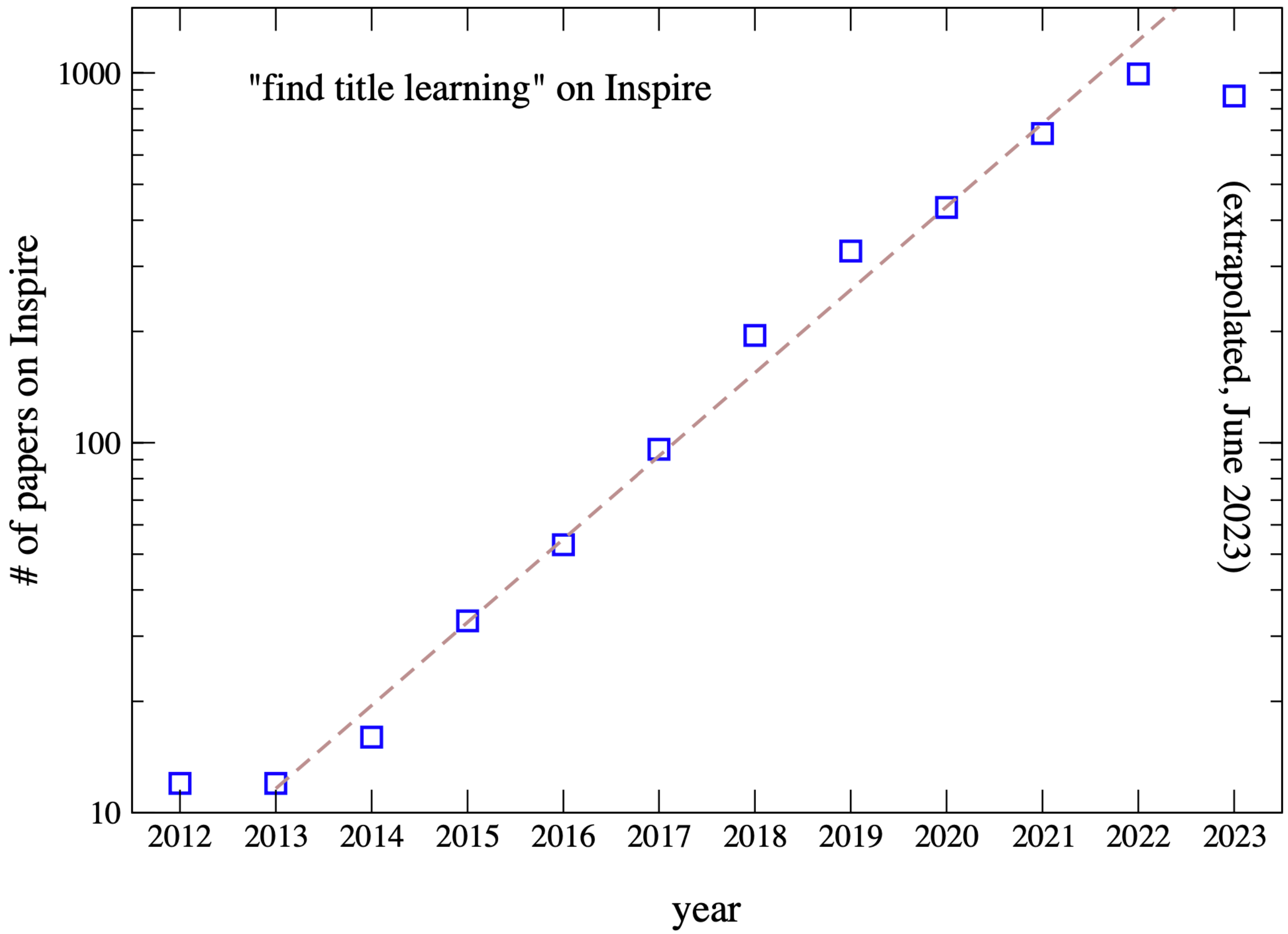


"find title **learning**" on the iNSPIRE-HEP data base

dedicated HEP database, collecting all papers in the field

High-energy physics (HEP)





find title **learning** on the iNSPIRE-HEP data base

exponential increase

cdt-aimlac.org/

AIMLAC

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

- 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

AIMLAC: some project titles

- coevolution of galaxies and supermassive black holes in the Euclid era
- searching for New Physics with the CMS experiment at the Large Hadron Collider
- a cold and dusty universe: understanding the cosmic dust and cold gas in nearby galaxies
- tests of the dark sector with gravitational waves
- exploiting GAIA data and understanding the galaxies’ past histories with ML
- simulation-based inference of gravitational waves signals from black holes and neutron stars
- searches for Beyond-Standard-Model signatures with jets + missing energy
- monsters in the dark: gas, dust and star formation around supermassive black holes
- AI techniques for extracting source information from Square Kilometre Array (SKA) datasets
- deep learning for real-time gravitational wave detection
- …

ML in fundamental physics

- lots of applications of ML in particle physics, astronomy, gravitational waves, …

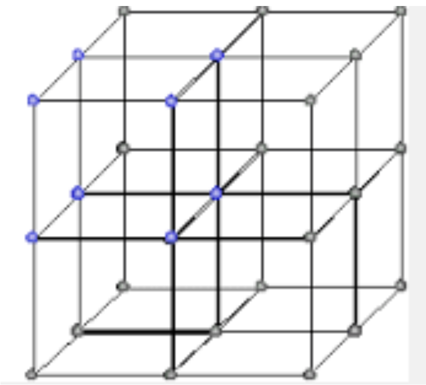
AI/ML for physics

- considerable overlap between concepts in theoretical physics and in ML

- interesting cross-talk to explore algorithms and improve understanding

physics for AI/ML

My interests and background

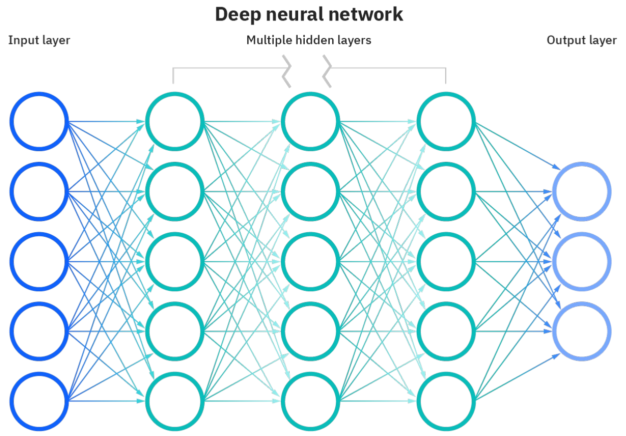


theoretical physics → particle physics → quantum field theory → lattice field theory

- understand the strong nuclear force (quarks & gluons) from first principles
- systems of many fluctuating degrees of freedom (quantum fields) on a spacetime lattice
- interactions via fluctuating links, determined by physical theory

ML models/neural networks

- train ML model to complete some task
- many degrees of freedom ("neurons") on a multi-layered network
- nodes interact via weights, determined by loss function



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, 2309.15002 [hep-lat]

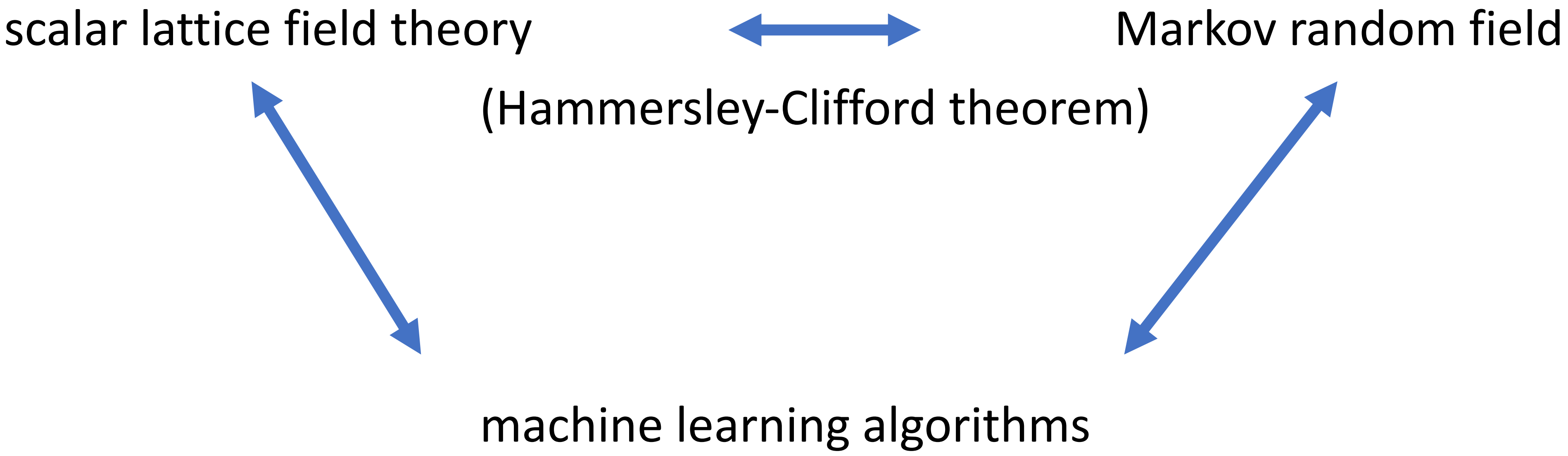
- stochastic quantisation and diffusion models

  Lingxiao Wang, GA, Kai Zhou, 2309.17082 [hep-lat]

  NeurIPS 2023 2311.03578 [hep-lat]

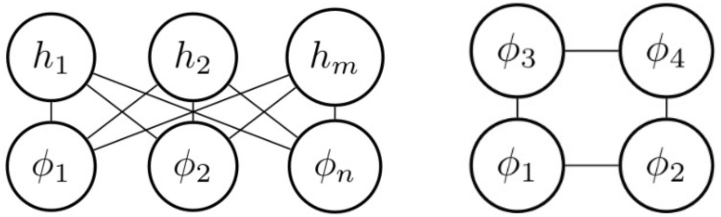
lattice field theory and Markov random fields

derive machine learning algorithms from discretized Euclidean field theories



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

Probability distribution $\mathrm{p}(\varphi)$



probability distribution $\mathrm{p}(\varphi)$ defined as product of nonnegative functions over maximal cliques in graph (or lattice):

$$p(\phi) = \frac{1}{Z} \prod_{c \in C} \psi_c(\phi)$$

then $\mathrm{p}(\varphi)$ satisfies local Markov property

and set of random variables φ define a Markov random field

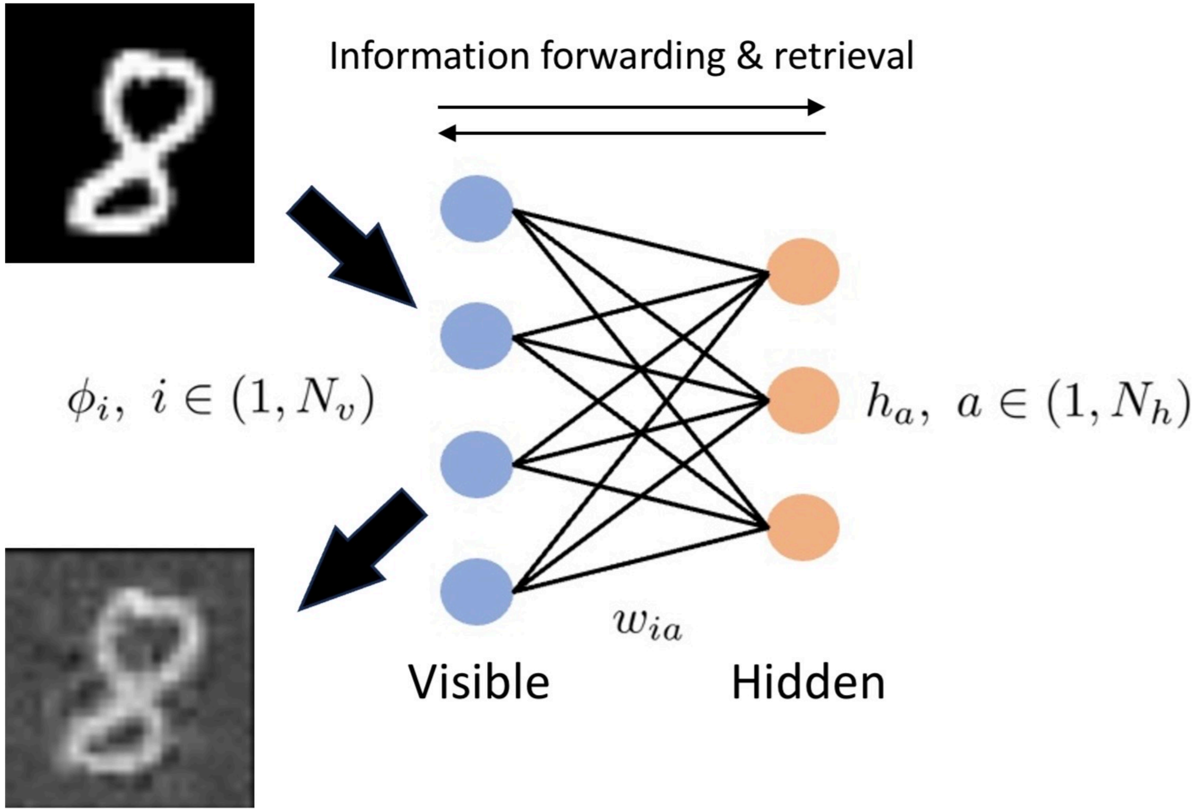
$\mathrm{p}(\varphi)$ defines lattice field theory

Theorem 1 (Hammersley-Clifford.) *A strictly positive distribution p satisfies the local Markov property of an undirected graph $\mathcal{G}$, if and only if p can be represented as a product of nonnegative potential functions ψ_c over $\mathcal{G}$, one per maximal clique $c \in C$, i.e.,*

$$p(\phi) = \frac{1}{Z} \prod_{c \in C} \psi_c(\phi), \qquad (2)$$

where $Z = \int_{\boldsymbol{\phi}} \prod_{c \in C} \psi_c(\boldsymbol{\phi}) d\boldsymbol{\phi}$ is the partition function and $\boldsymbol{\phi}$ are all possible states of the system.

Restricted Boltzmann Machine: generative network

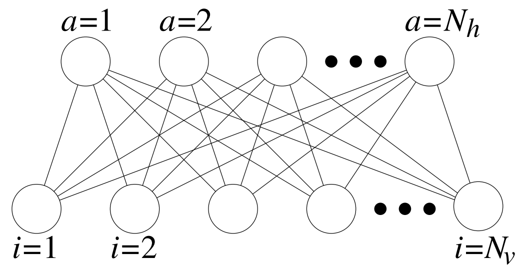


- energy-based method
- probability distribution
- binary or continuous d.o.f.

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

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

Scalar field RBM



- treat RBM as a **lattice field theory** with action

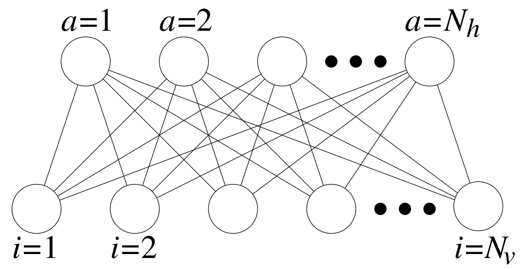
$$S(\phi,h)=\sum_i \frac{1}{2}\mu_i^2\phi_i^2+\sum_a \frac{1}{2\sigma^2}(h_a-\eta_a)^2-\sum_{i,a}\phi_i w_{ia} h_a$$

- only quadratic terms: learn weight matrix w_{ia} and bias η_a
- induced distribution on visible layer

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

- kinetic (all-to-all) term $K_{ij}=\mu_i^2\delta_{ij}-\sigma^2\sum_a w_{ia}w_{aj}^T$ and source $J_i=\sum_a w_{ia}\eta_a$

Chanju Park, Biagio Lucini, GA, 2309.15002 [hep-lat]



Gaussian scalar field RBM

- kinetic (all-to-all) term $K_{ij} = \mu_i^2 \delta_{ij} - \sigma^2 \sum_a w_{ia} w_{aj}^T$ and source $J_i = \sum_a w_{ia} \eta_a$

- kinetic term and source should match target distribution/data

- for instance: target is scalar field theory with $K^\phi \approx p^2 + m^2$

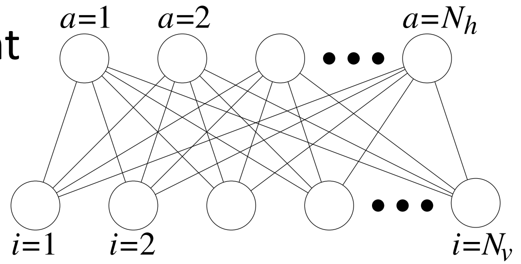
- ❖ dependence on N_h: what if $N_h < N_v$? role of hyperparameter μ^2?

- ❖ both N_h and μ^2 act as **ultraviolet regulators**

- regulate the spectrum of the quadratic operator K^ϕ of the target distribution/data

What if $N_h < N_v$?

train RBM with persistent contrastive divergence

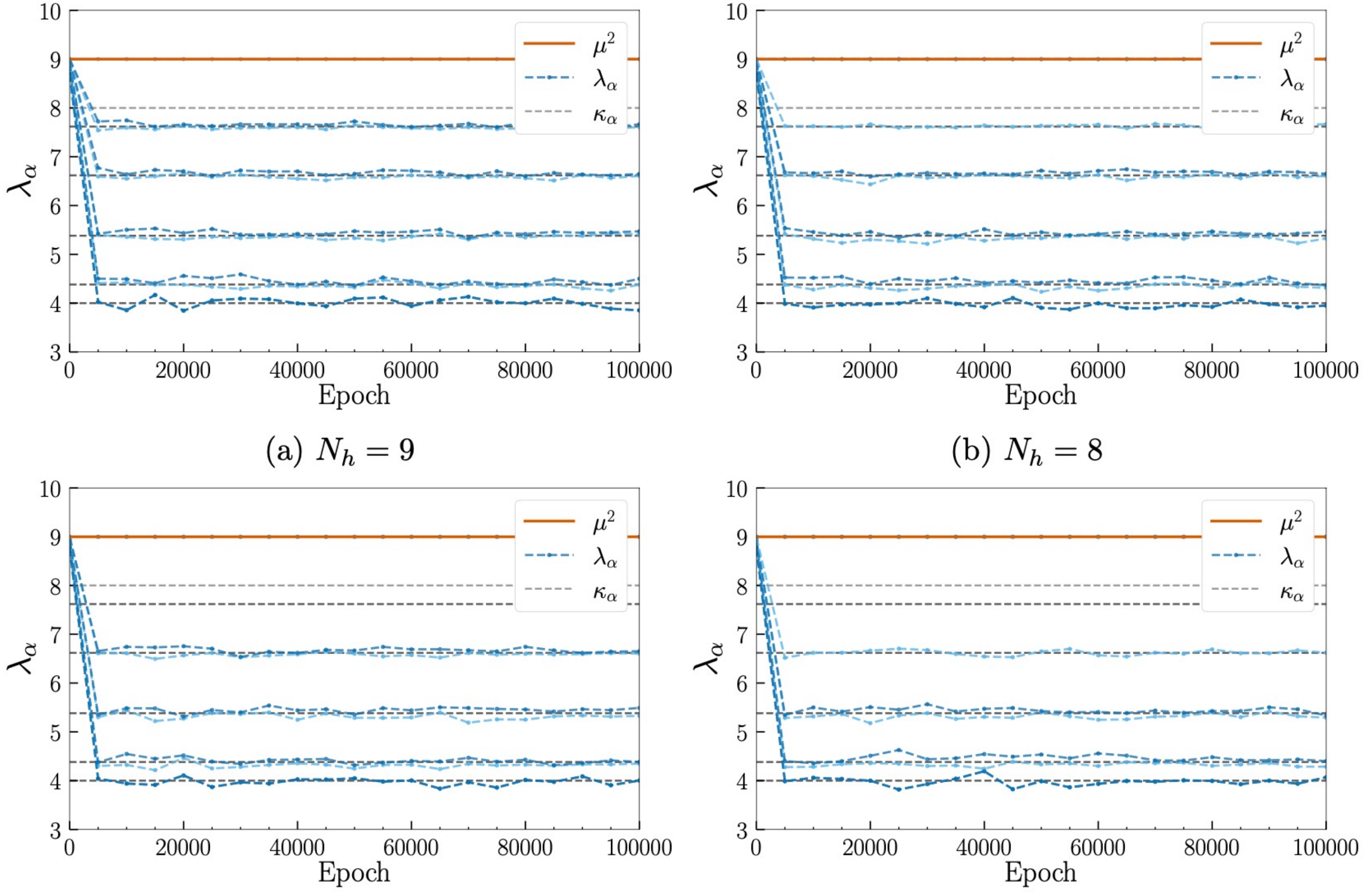


- example: scalar LFT
  with $N_v = 10$ nodes

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

- reproduced by RBM (λ) from smallest eigenvalue upwards

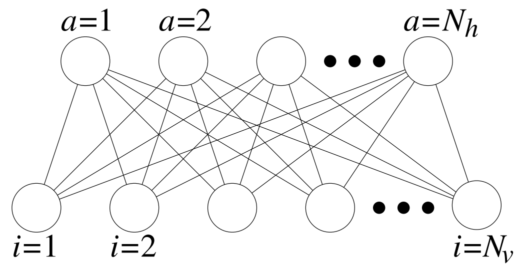
- higher modes are moved to cut-off scale (μ^2)



(a) $N_h = 9$ (b) $N_h = 8$

(c) $N_h = 7$ (d) $N_h = 6$

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



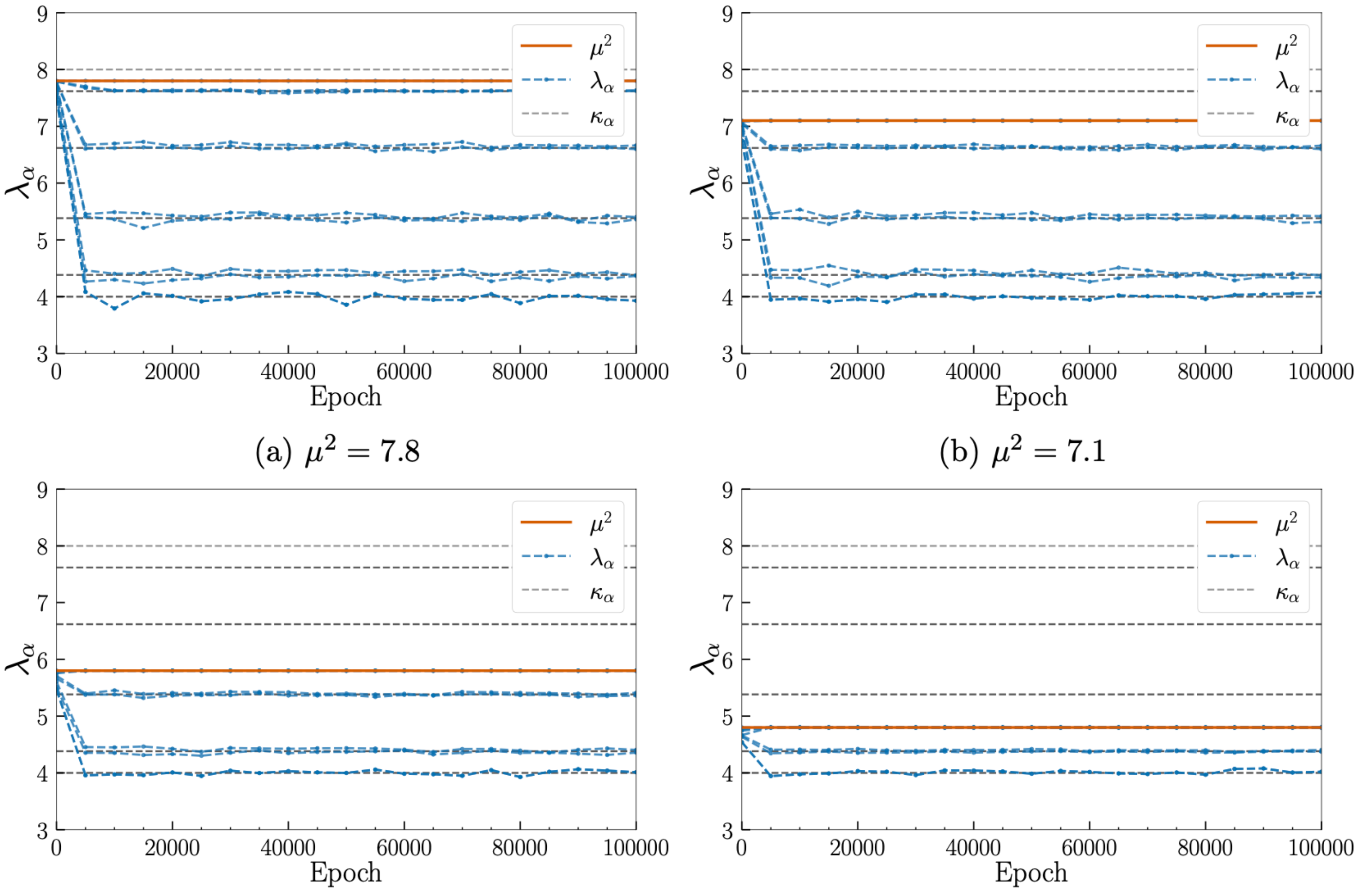
- example: scalar LFT
  with $N_v = 10$ nodes

- exact spectrum:

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

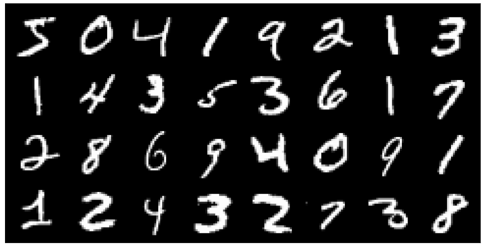
- reproduced by RBM (λ) from smallest eigenvalue upwards

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

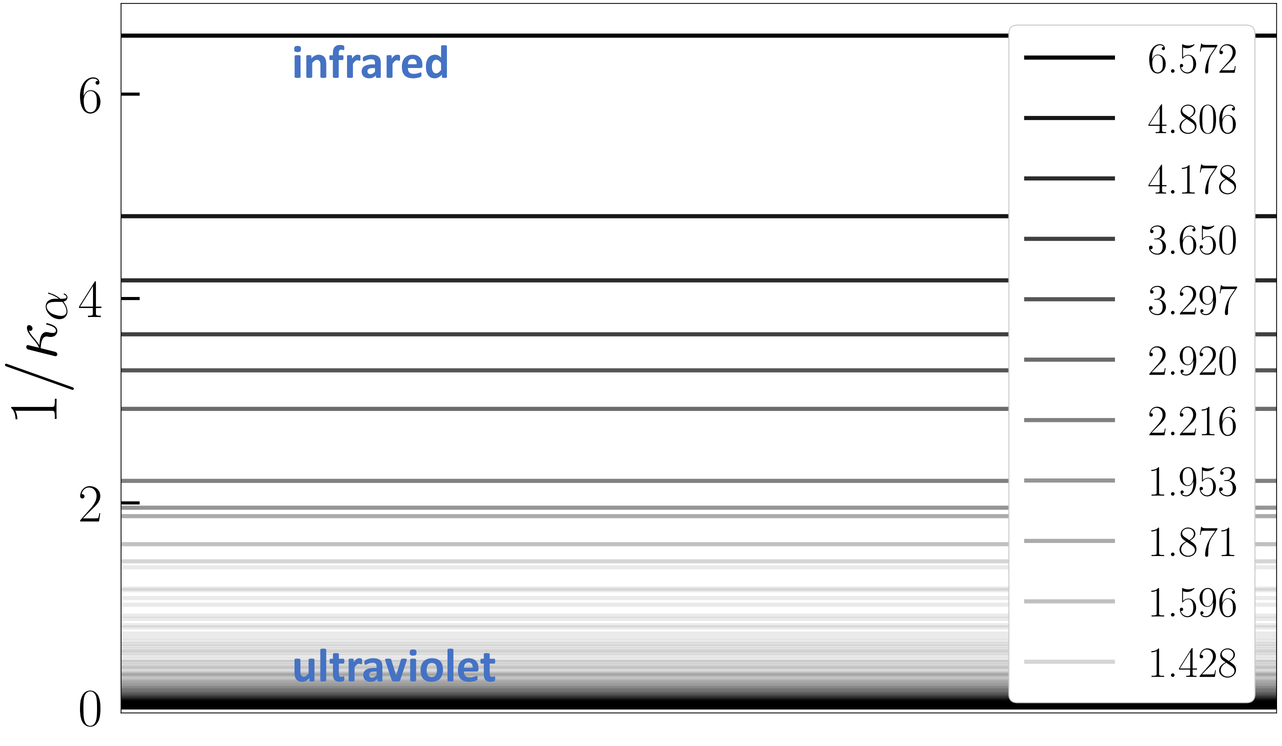


(a) $\mu^2 = 7.8$ (b) $\mu^2 = 7.1$ (c) $\mu^2 = 5.8$ (d) $\mu^2 = 4.8$

RBM as ultraviolet regulator

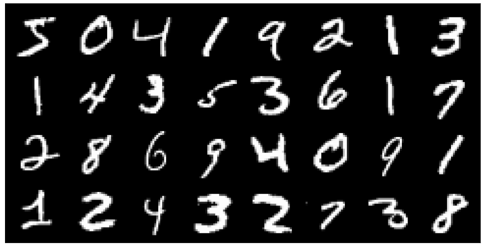


- relevant for “real” data sets? MNIST
- compute spectrum of two-point correlator $K_{ij}^{-1} = < \phi_i \phi_j >_{data}$
- inverse spectrum $1/\kappa$
- infrared safe
- ultraviolet divergent

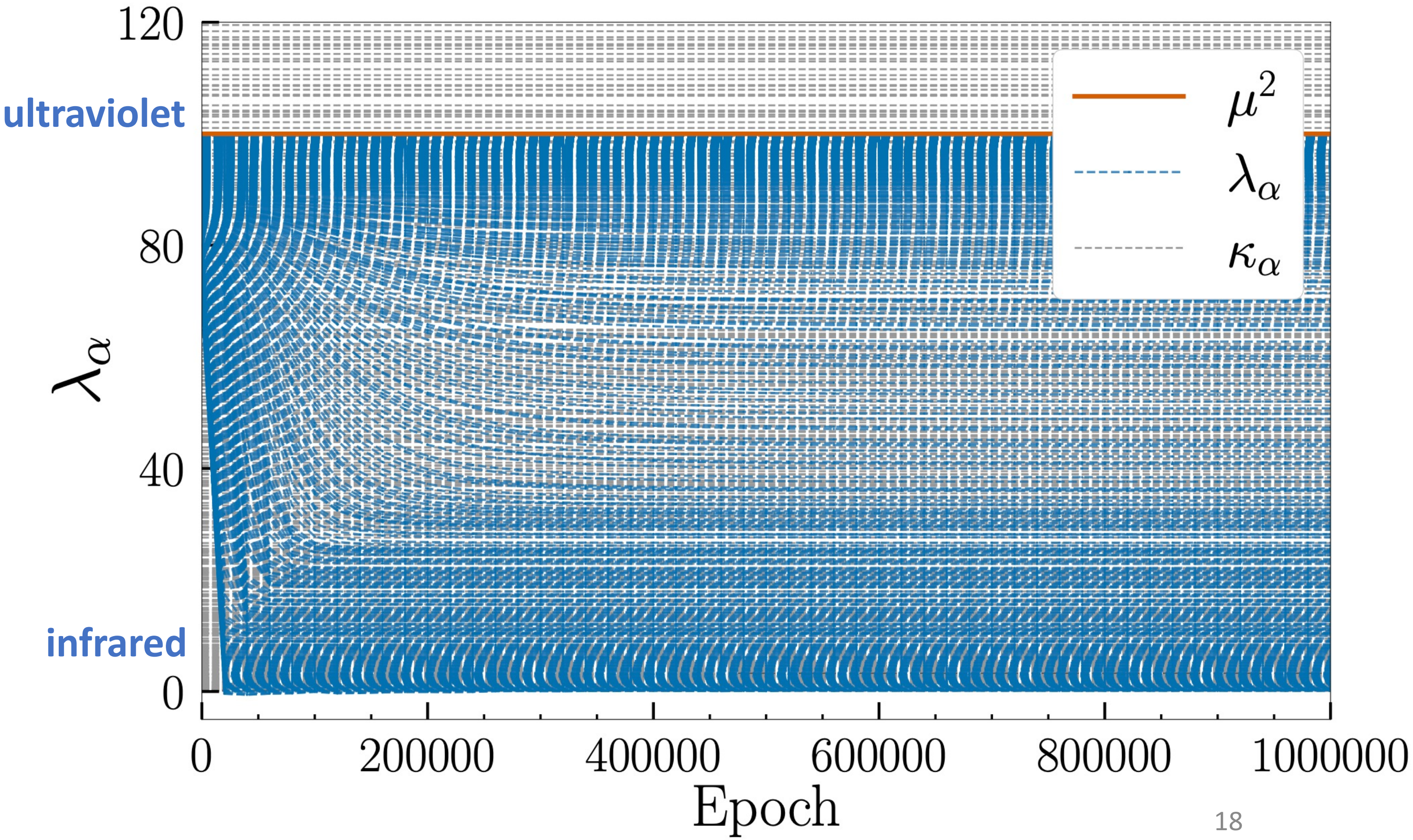


784 eigenvalues

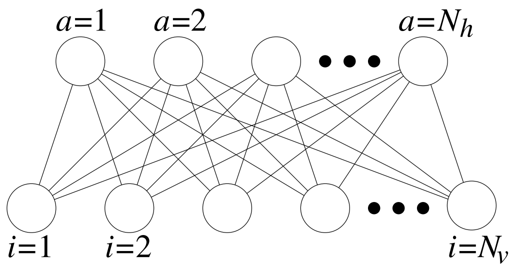
MNIST with fixed RBM mass



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



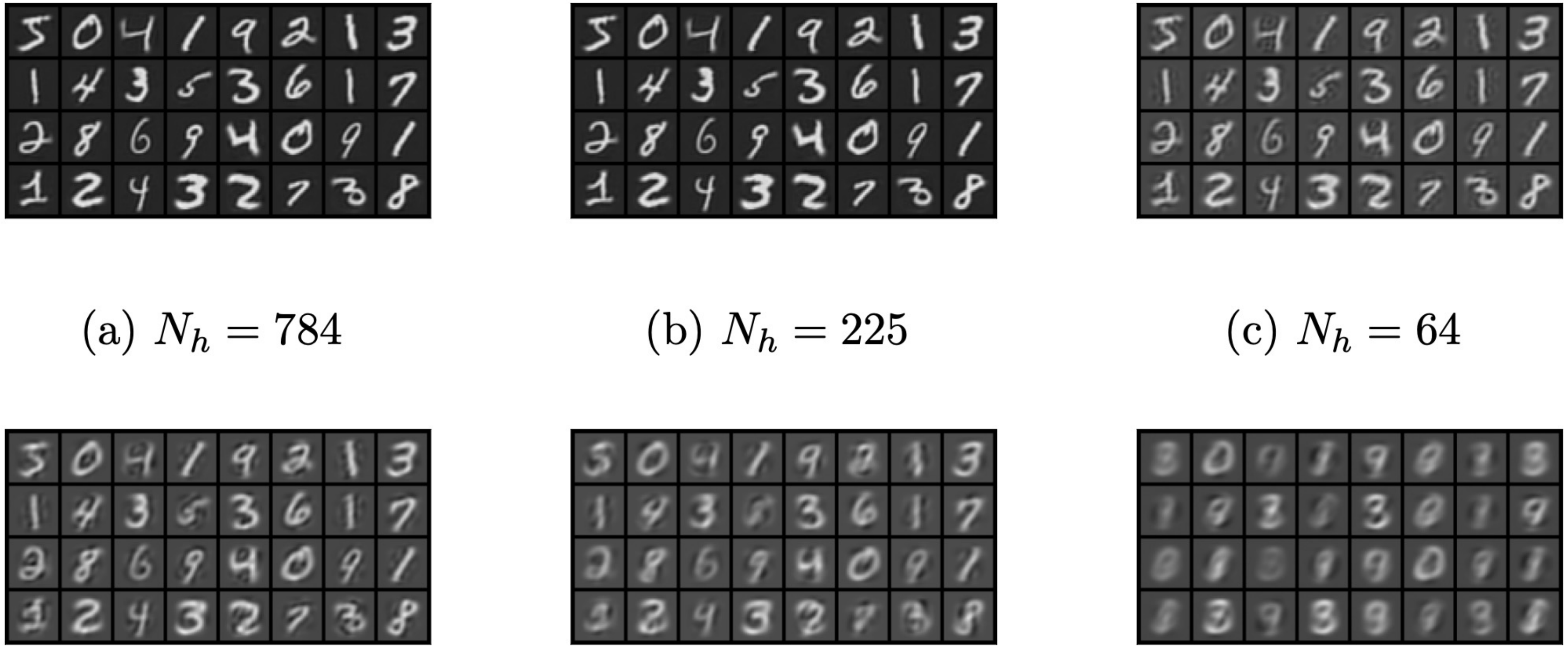
MNIST with $N_h \leq N_v$



what is the effect of including incomplete spectrum?

removal of ultraviolet modes affects generative power

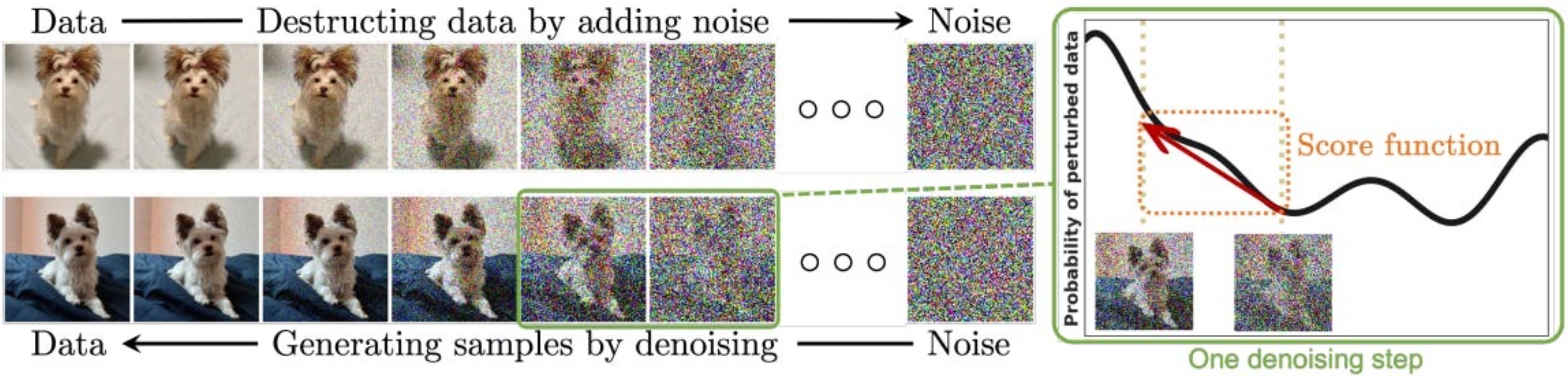
language of quantum fields helps in understanding (at least for me!)



(a) $N_h = 784$ (b) $N_h = 225$ (c) $N_h = 64$

(d) $N_h = 36$ (e) $N_h = 16$ (f) $N_h = 4$

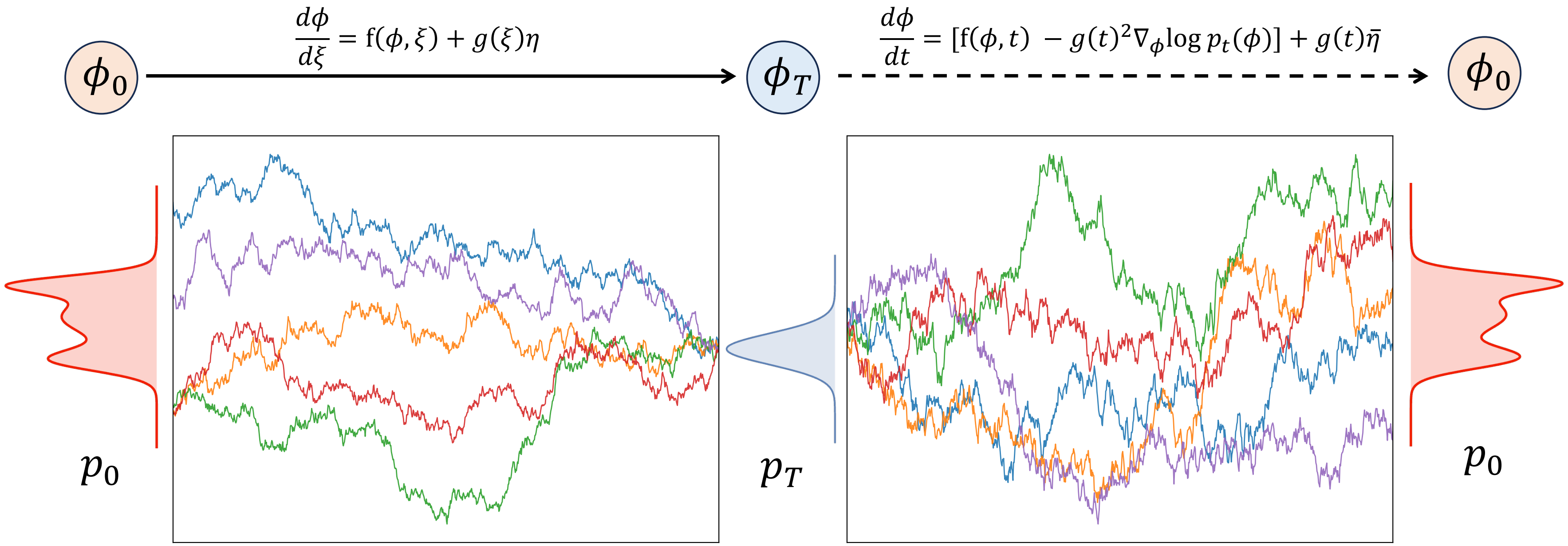
Diffusion models: generative ML



https://encord.com/blog/diffusion-models/

Diffusion models: generative ML

- 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

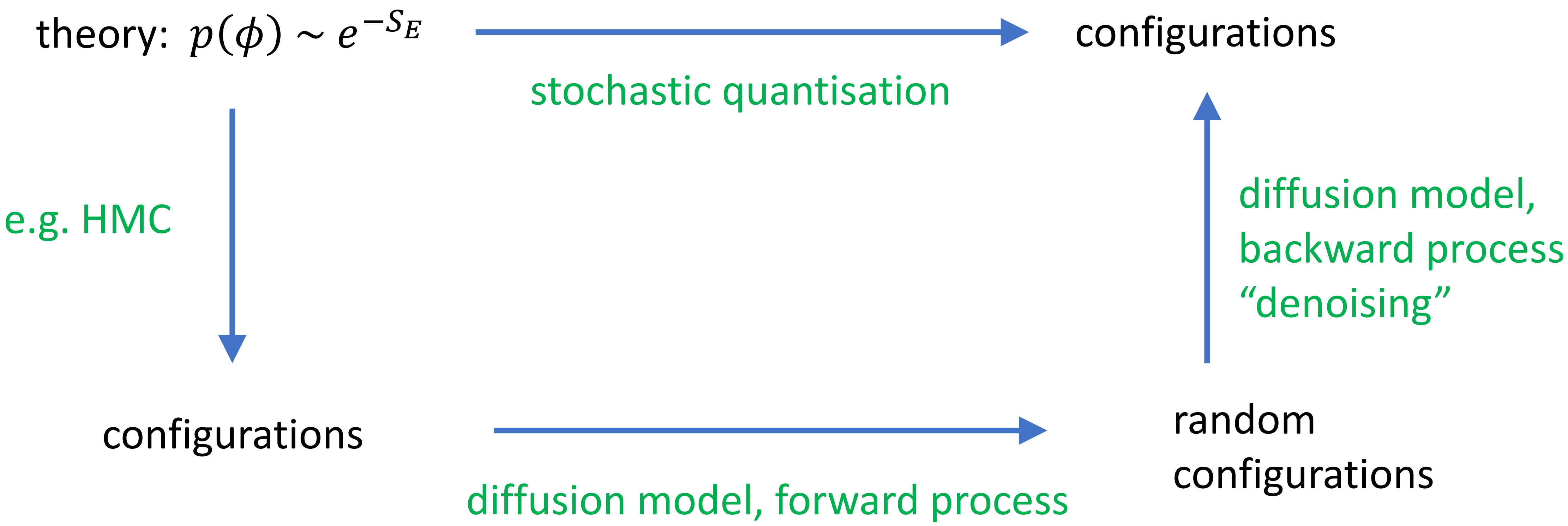
- ideas well-known in quantum field theory: stochastic quantisation (Parisi & Wu 1980)
- path integral quantisation via a stochastic process in fictitious time

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

- 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



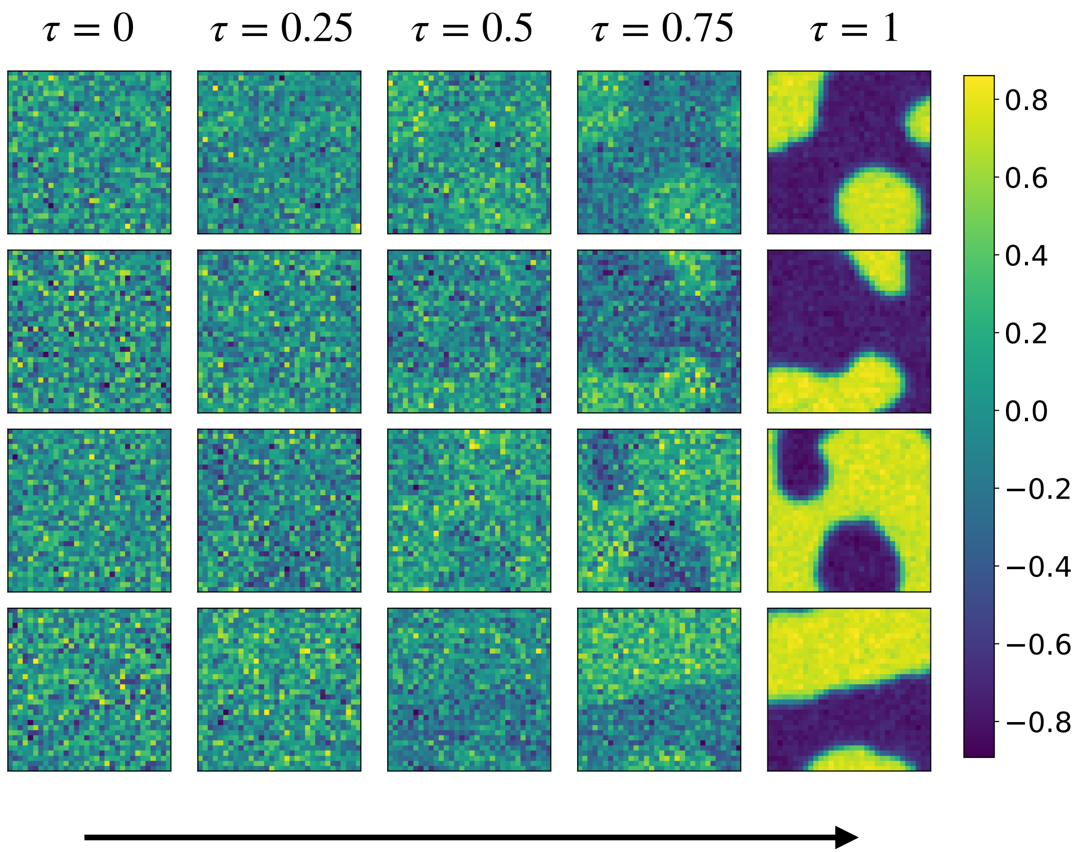
Diffusion model for 2d ϕ^4 scalar theory

- 32^2 lattice, 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

use diffusion models to generate configurations in field theory



Lingxiao Wang, GA, Kai Zhou, 2309.17082 [hep-lat]

Summary: ML in fundamental physics

- lots of applications of ML in particle physics, astronomy, gravitational waves, …

AI/ML for physics

- considerable overlap between concepts in theoretical physics and in ML
- interesting cross-talk to explore algorithms and improve understanding

physics for AI/ML