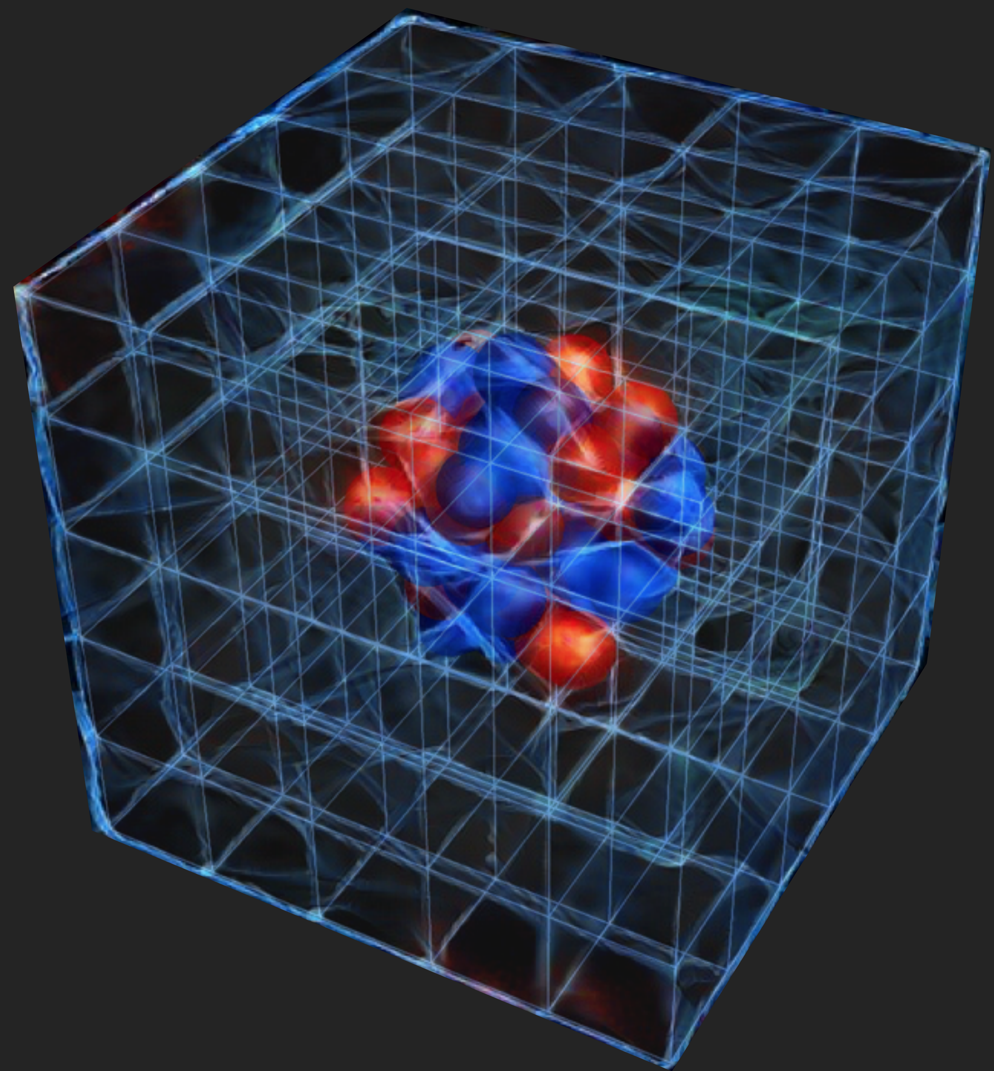
Machine Learning for Lattice Field Theory

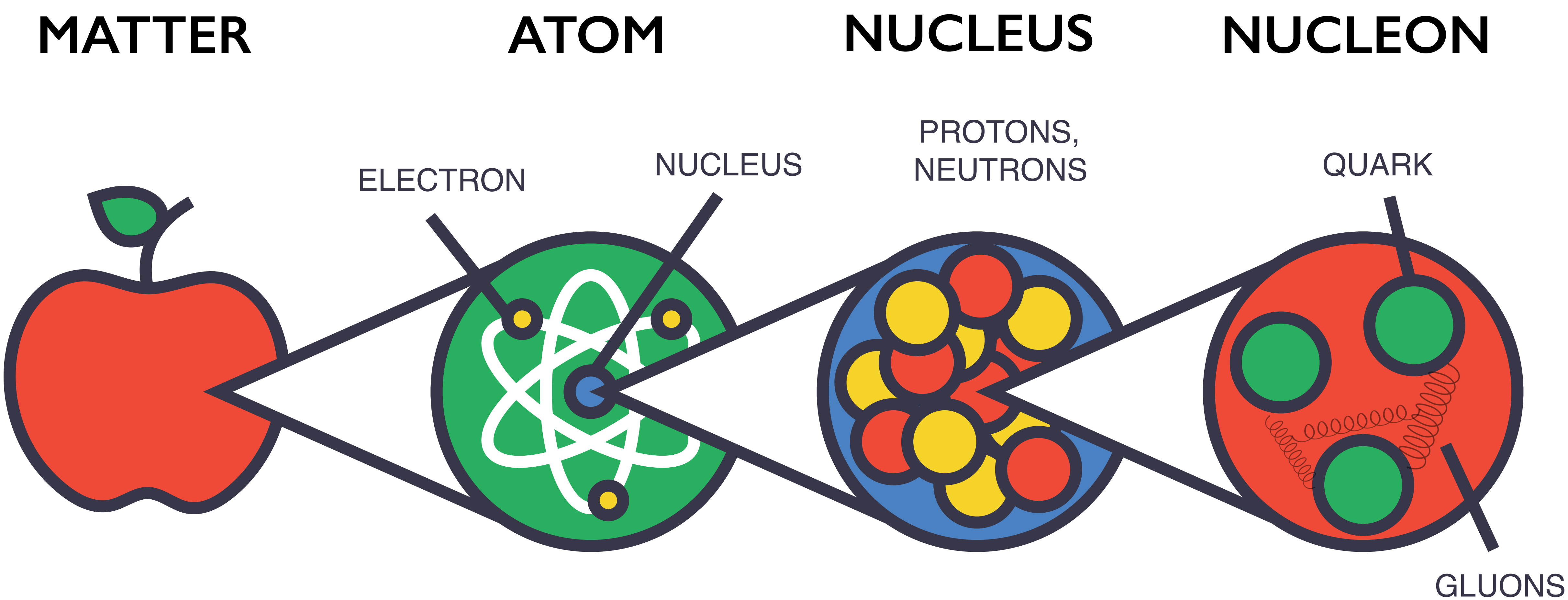


Phiala Shanahan

Massachusetts Institute of Technology

The structure of matter

Nuclear physics from the
Standard Model of fundamental particles

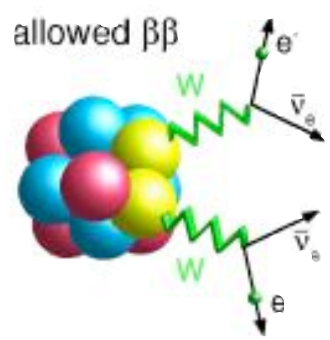


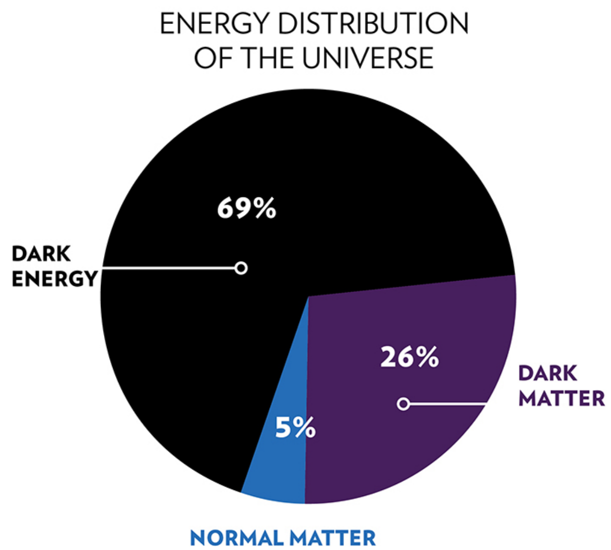
The search for new physics

Precise experiments seek new physics at the "Intensity Frontier"

- Sensitivity to probe the rarest Standard Model interactions
- Search for beyond—Standard-Model effects

- Dark matter direct detection
- Neutrino physics
- Charged lepton flavour violation, **ββ**-decay, proton decay, neutron-antineutron oscillations…





CHALLENGE: understand the physics of nuclei used as targets

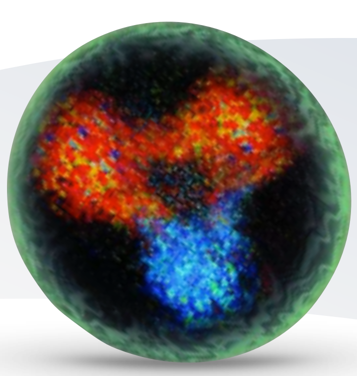
Strong interactions

Study nuclear structure from the strong interactions

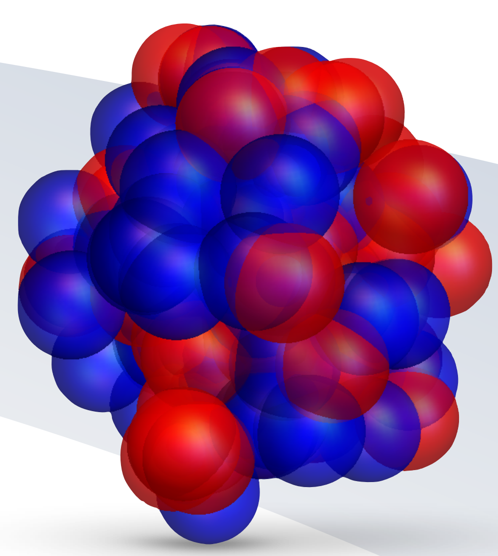
Quantum Chromodynamics (QCD)

Strongest of the four forces in nature

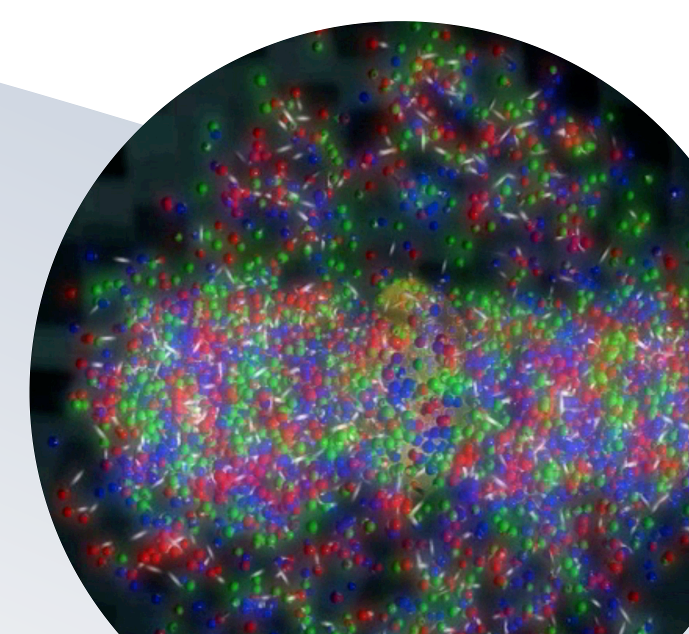
Forms other types of exotic matter e.g., quark-gluon plasma



Binds quarks and gluons into protons, neutrons, pions etc.



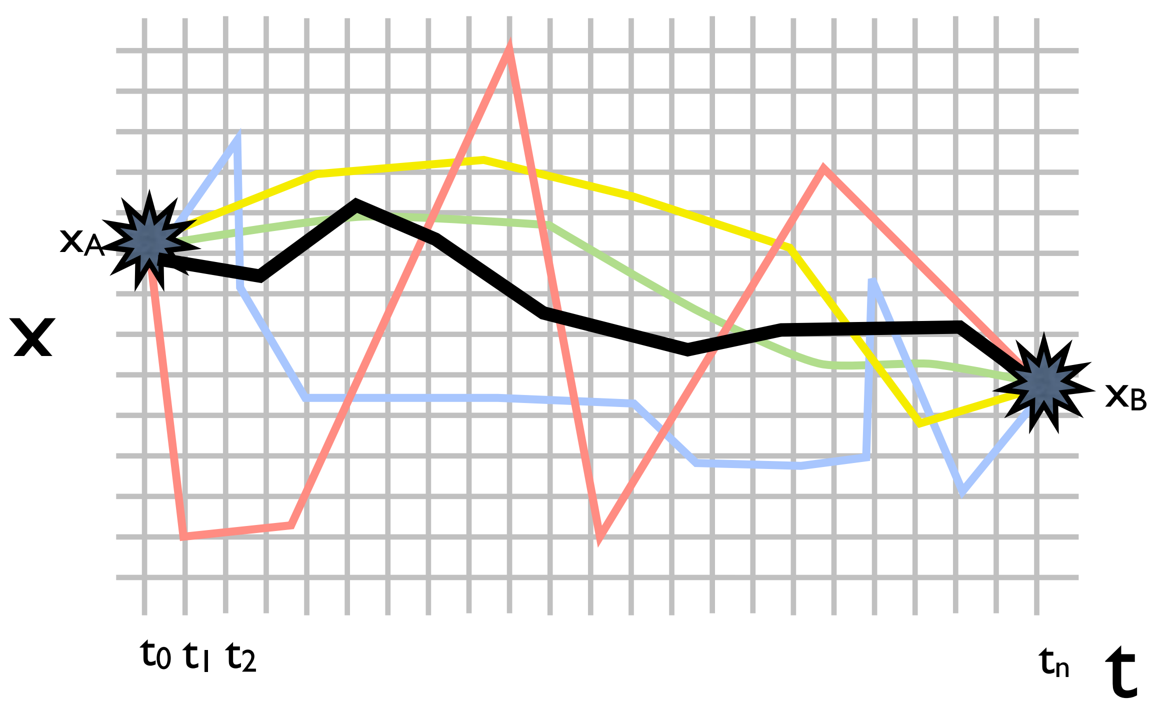
Binds protons and neutrons into nuclei



Lattice QCD

Numerical first-principles approach to
non-perturbative QCD

- Discretise QCD onto 4D space-time lattice
- QCD equations ⟷ integrals over the values of quark and gluon fields on each site/link (QCD path integral)
- ~10^{12} variables (for state-of-the-art)



- Evaluate by importance sampling
- Paths near classical action dominate
- Calculate physics on a set (ensemble) of samples of the quark and gluon fields

Lattice QCD

Workflow of a lattice QCD calculation

1 Generate field configurations via Hybrid Monte Carlo

- Leadership-class computing
- ~100K cores or 1000GPUs, 10's of TF-years
- O(100-1000) configurations, each ~10-100GB

2 Compute propagators

- Large sparse matrix inversion
- ~few 100s GPUs
- 10x field config in size, many per config

3 Contract into correlation functions

- ~few GPUs
- O(100k-1M) copies

Computational cost grows exponentially with size of nuclear system

Lattice QCD

Workflow of a lattice QCD calculation

1. Generate field configurations via Hybrid Monte Carlo
   - Leadership-class computing
   - ~100K cores or 1000GPUs, 10's of TF-years
   - O(100-1000) configurations, each ~10-100GB

2. Compute propagators
   - Large sparse matrix inversion
   - ~few 100s GPUs
   - 10x field config in size, many per config

3. Contract into correlation functions
   - ~few GPUs
   - O(100k-1M) copies

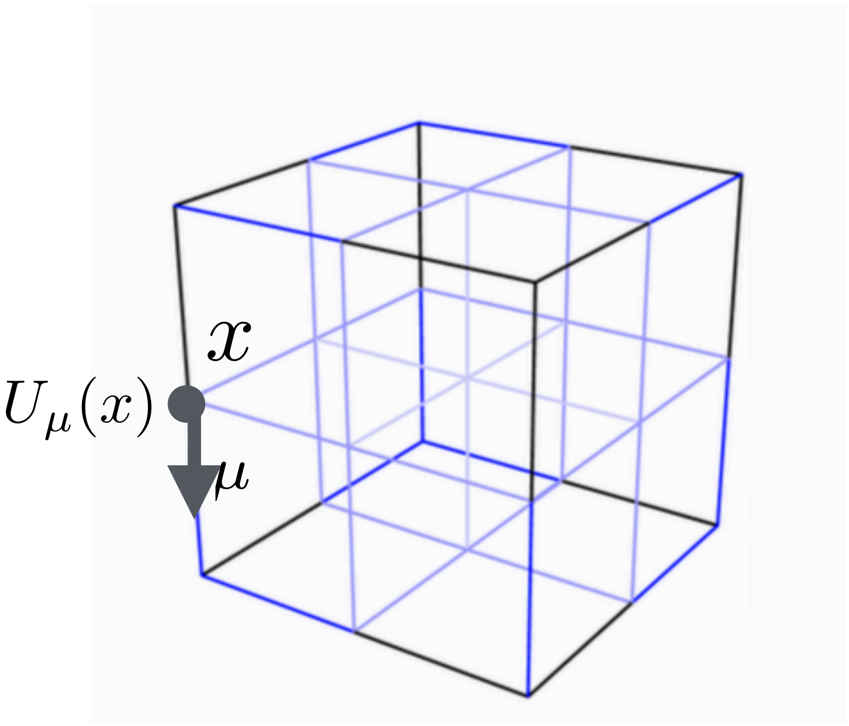
Computational cost grows exponentially with size of nuclear system

Lattice QCD

Generate field configurations $\phi(x)$ with probability

$$P[\phi(x)] \sim e^{-S[\phi(x)]}$$

- Gauge field configurations represented by $\sim 10^{10}$ links $U_\mu(x)$ encoded as SU(3) matrices (3x3 complex matrix M with $det[M] = 1$, $M^{-1} = M^\dagger$) i.e., $\sim 10^{12}$ double precision numbers



- Configurations sample probability distribution corresponding to LQCD action $S[\phi]$ (function that defines the quark and gluon dynamics)

  ➡ Weighted averages over configurations determine physical observables of interest

- Calculations use $\sim 10^3$ configurations

Generate QCD gauge fields

Generate field configurations $\phi(x)$ with probability

$$P[\phi(x)] \sim e^{-S[\phi(x)]}$$

Molecular dynamics

Classical motion with

conjugate

$$H = \sum_x \frac{\pi^2(x)}{2} + S[\phi(x)]$$

- Reversible
- Volume-preserving

BUT

- Energy non-conservation for numerical integrators

Markov Chain Monte Carlo

Propose update using integrated molecular dynamics trajectory

Accept/ reject with probability

$$\alpha = \min(1, e^{(-S[\phi'(x)]+S[\phi(x)])})$$

- Numerical error corrected by accept/reject

BUT

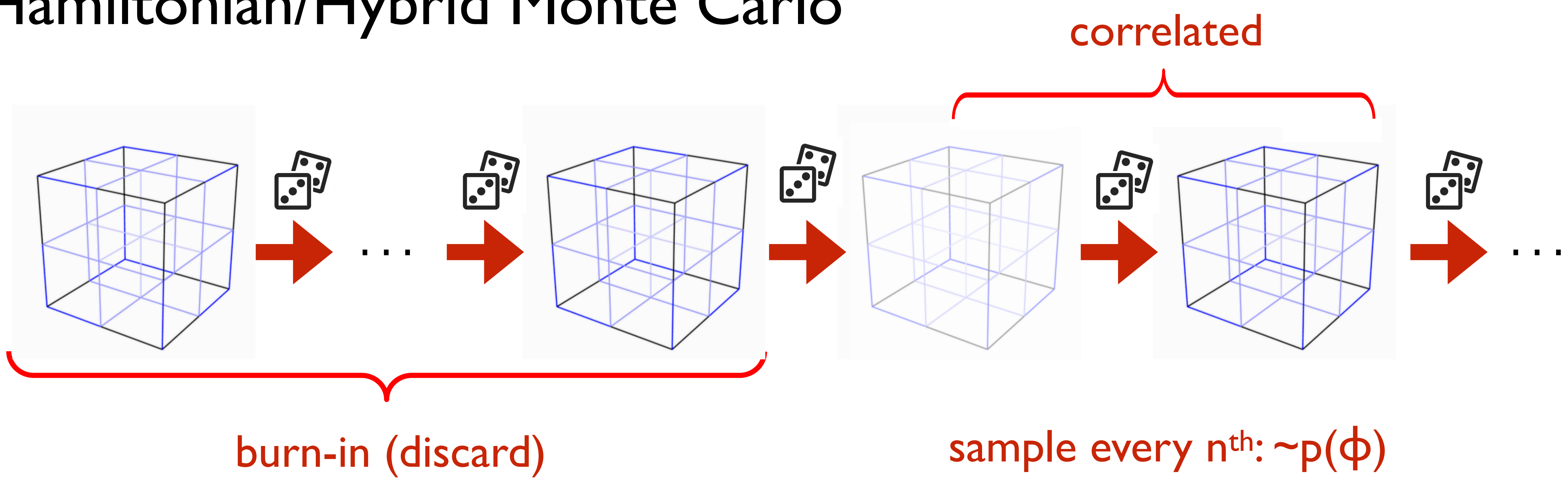
- Short trajectories for high acceptance

Generate QCD gauge fields

Generate field configurations $\phi(x)$ with probability

$$P[\phi(x)] \sim e^{-S[\phi(x)]}$$

Hamiltonian/Hybrid Monte Carlo

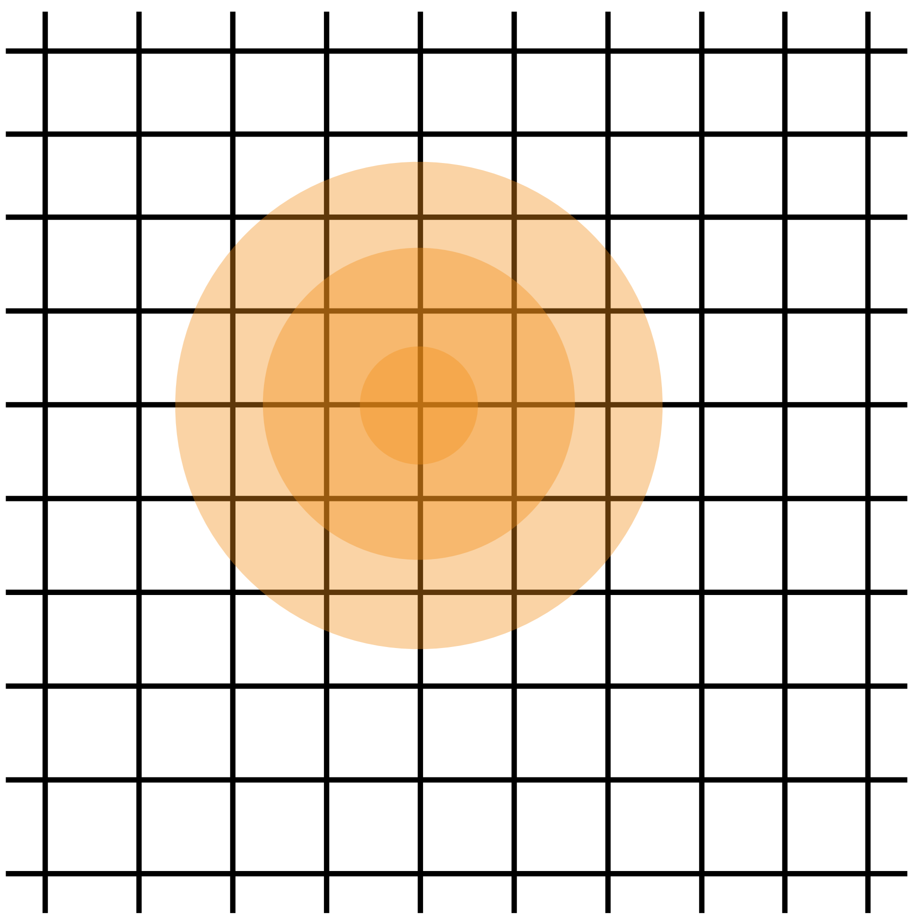


Burn-in time and correlation length dictated by Markov chain **'autocorrelation time'**: shorter autocorrelation time implies less computational cost

Accelerating Lattice QCD

QCD gauge field configurations sampled via

Hamiltonian dynamics + Markov Chain Monte Carlo



Updates diffusive

Lattice spacing → 0

Number of updates to change fixed physical length scale → ∞

"Critical slowing-down"
of generation of uncorrelated samples

Accelerating Lattice QCD

QCD gauge field configurations sampled via

Hamiltonian dynamics + Markov Chain Monte Carlo

"Critical slowing-down"
of generation of uncorrelated samples

Autocorrelation measure

$$\tau_{\mathcal{O}}^{\text{int}} = \frac{1}{2} + \lim_{\tau_{\max}\to\infty} \sum_{\tau=1}^{\tau_{\max}} \frac{\rho_{\mathcal{O}}(\tau)}{\rho_{\mathcal{O}}(0)}$$

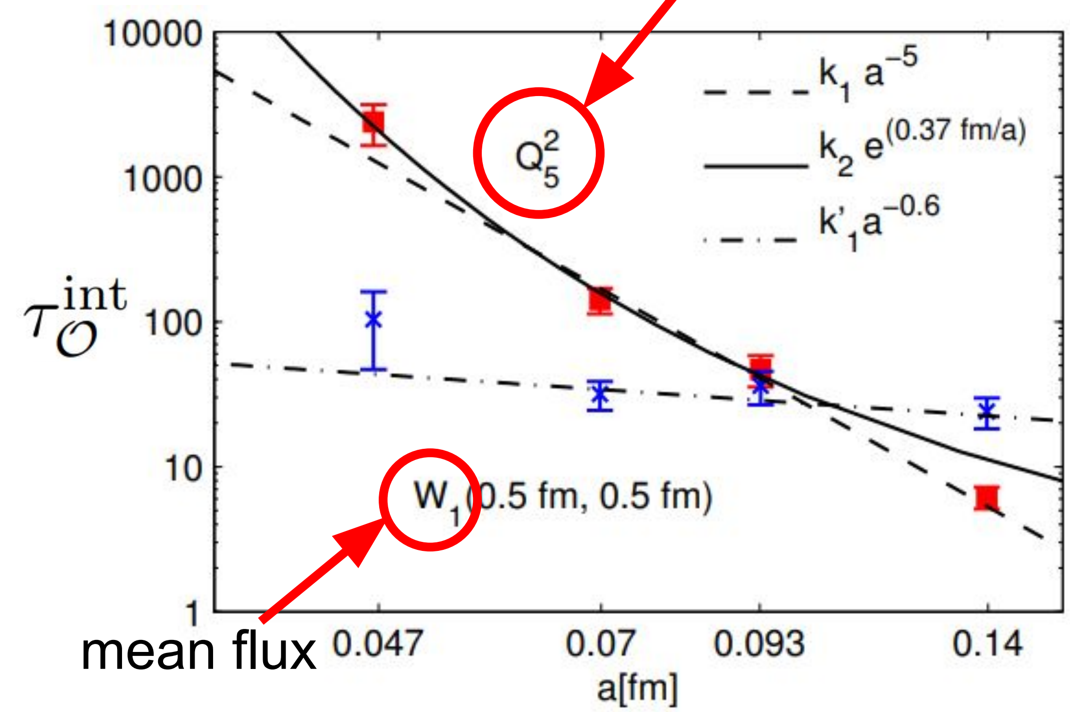
$$\tau_{\mathcal{O}}^{\text{int}} = \alpha_{\mathcal{O}} L^{z_{\mathcal{O}}}$$

Critical exponent

Correlation of observable $\mathcal{O}$ on configurations separated by τ Markov Chain steps

topological charge

[Schaefer et al. / ALPHA collaboration 1009.5228]



mean flux

critical limit

Machine learning QCD

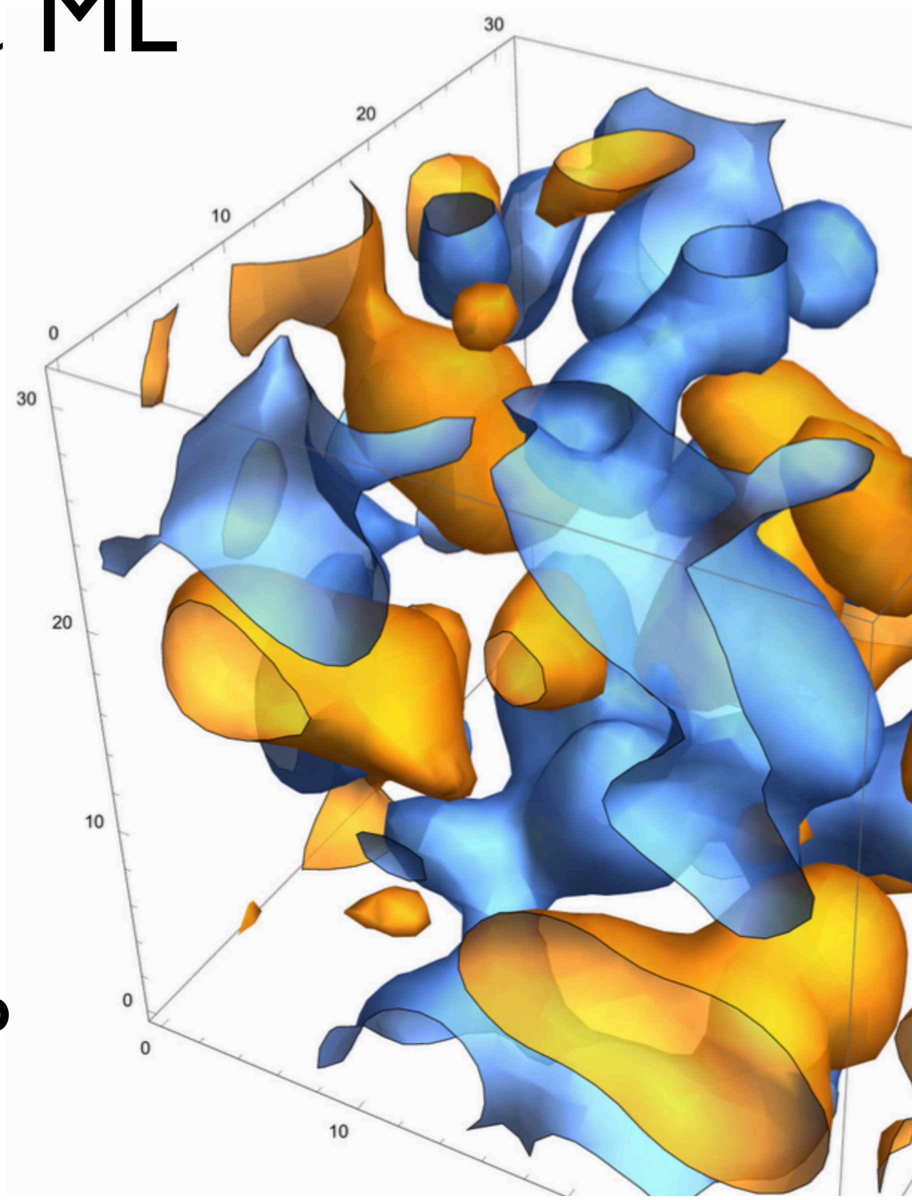
Accelerate gauge-field generation via ML

1. Multi-scale algorithms:
   parallels with image recognition
   Shanahan et al., PRD 97, 094506 (2018)

2. Generative models to replace Hybrid Monte-Carlo
   parallels with image generation
   Albergo et al., arXiv:1904.12072 (2019)

Gurtej Kanwar (MIT)

Michael Albergo (NYU)



Consider only approaches which rigorously preserve quantum field theory in applicable limits

Machine learning QCD

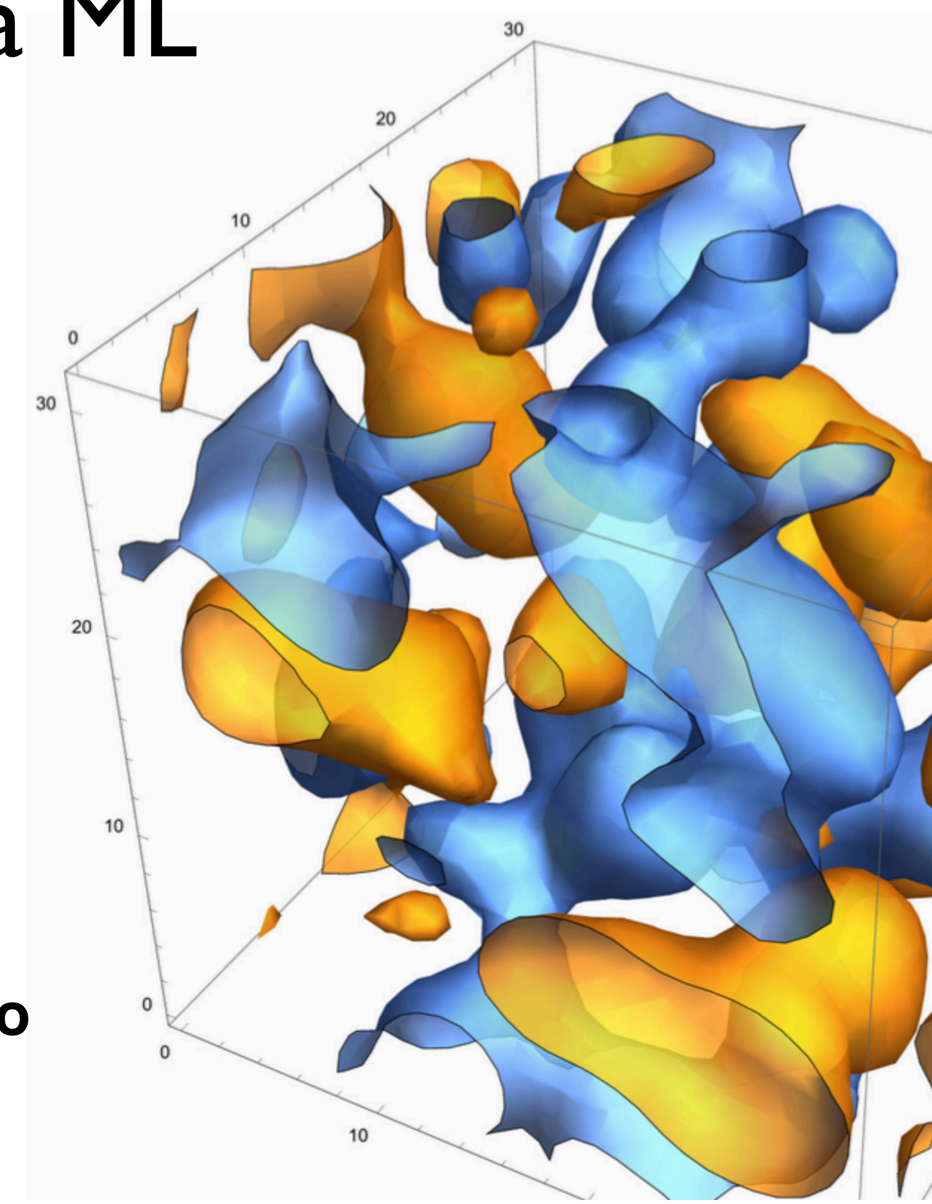
Accelerate gauge-field generation via ML

1. Multi-scale algorithms:
   parallels with image recognition
   Shanahan et al., PRD 97, 094506 (2018)

2. Generative models to replace Hybrid Monte-Carlo
   parallels with image generation
   Albergo et al., arXiv:1904.12072 (2019)

Gurtej Kanwar (MIT)

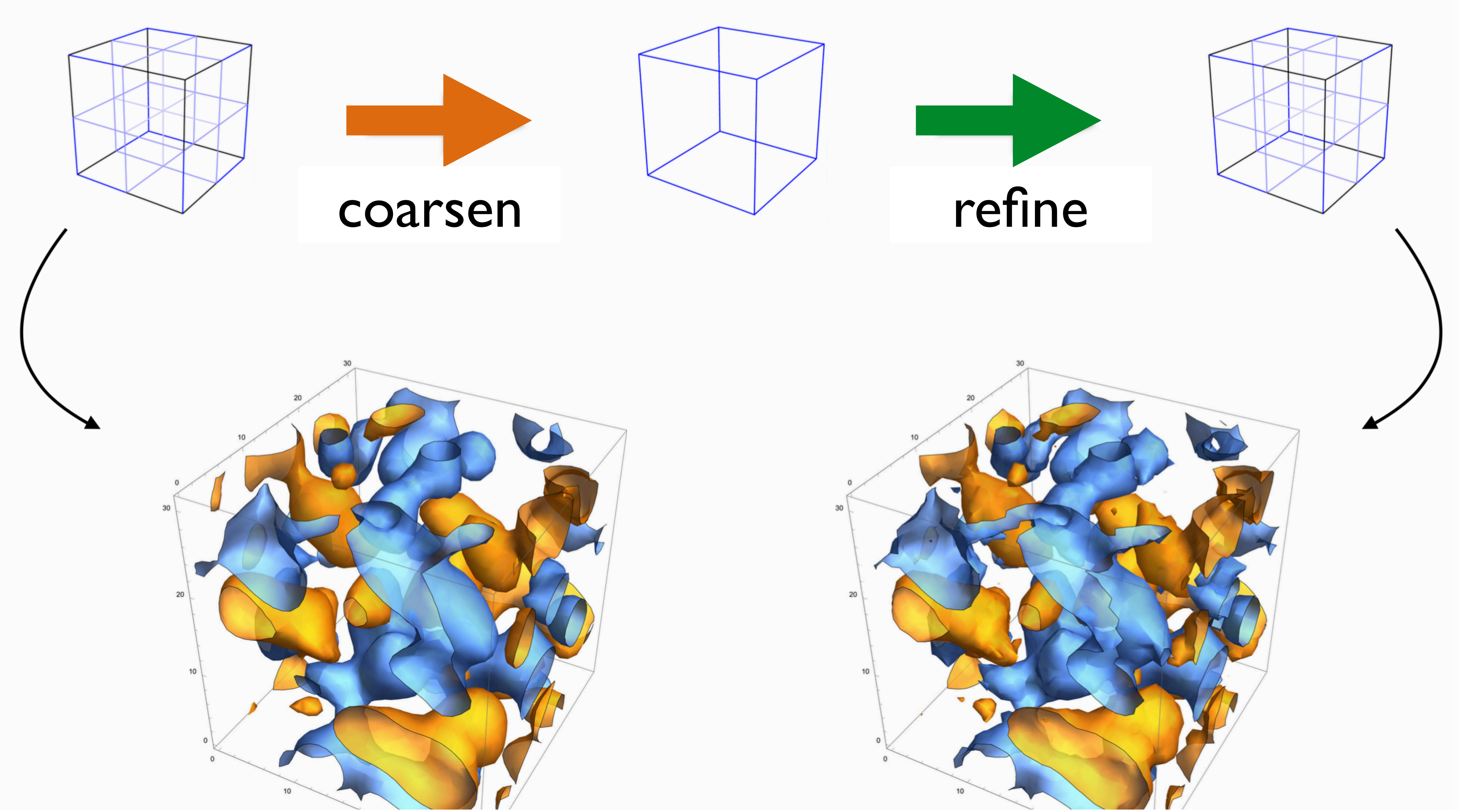
Michael Albergo (NYU)



Consider only approaches which rigorously preserve quantum field theory in applicable limits

Multi-scale HMC updates

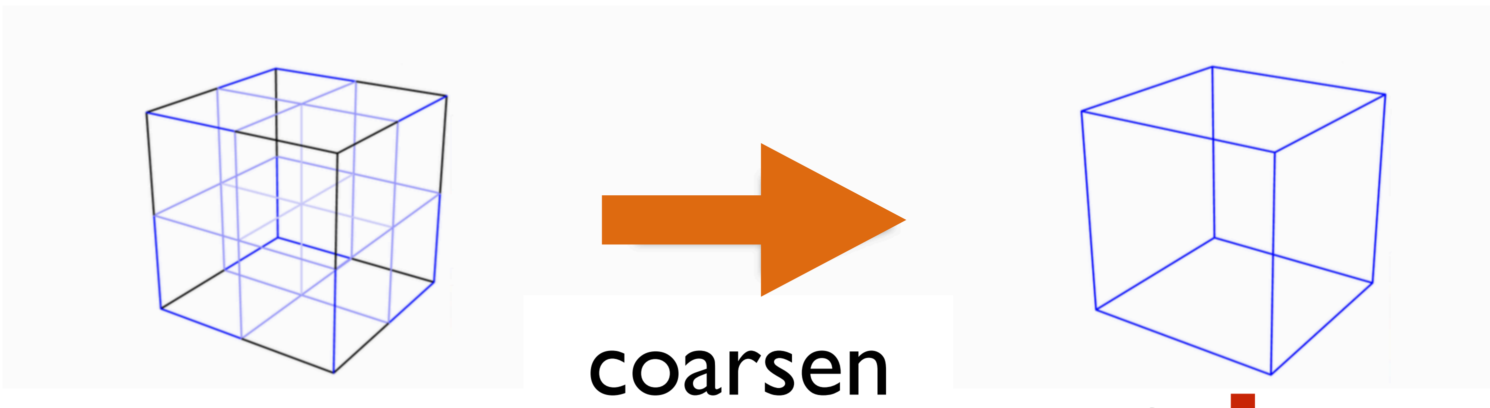
Given coarsening and refinement procedures…



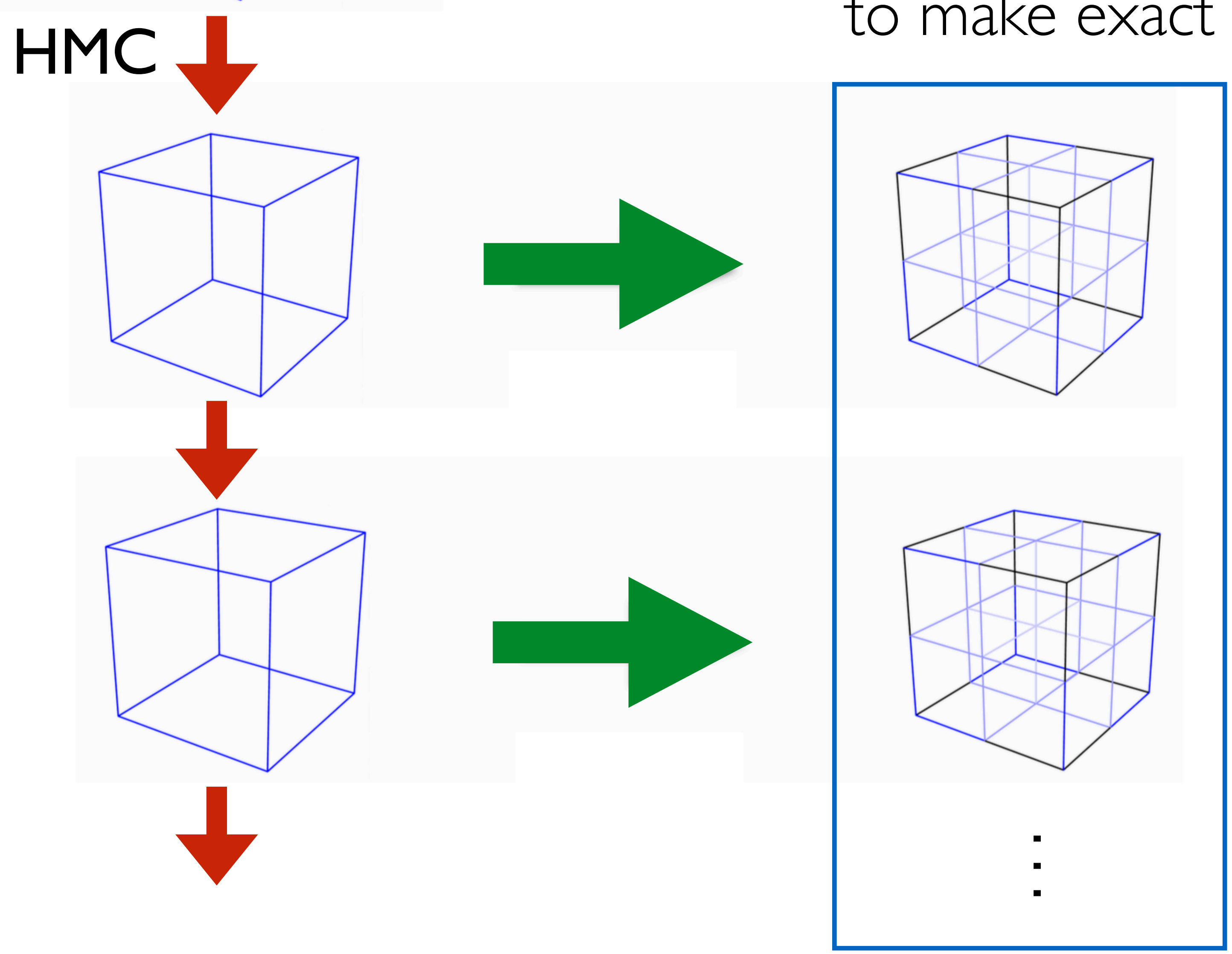
Endres et al., PRD 92, 114516 (2015)

Multi-scale HMC updates

Perform HMC updates at coarse level



Fine ensemble
rethermalise
with fine action
to make exact



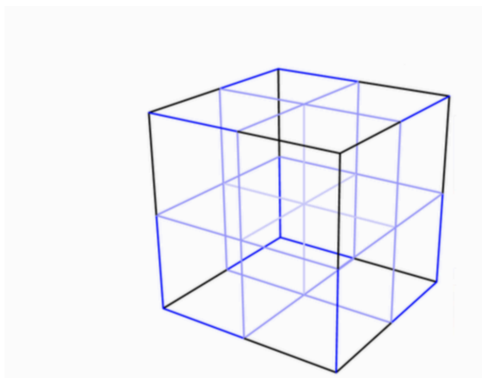
Multiple layers of coarsening

Significantly cheaper approach to continuum limit

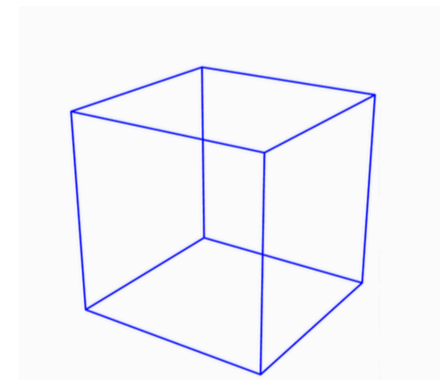
Endres et al., PRD 92, 114516 (2015)

Multi-scale HMC updates

Perform HMC updates at coarse level



encode same
long-distance
physics



MUST KNOW
parameters of coarse QCD action that reproduce ALL physics parameters of fine simulation

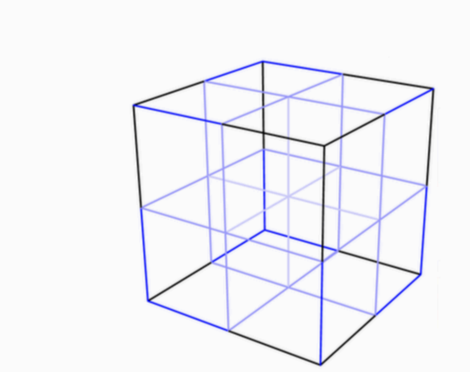
Map a subset of physics parameters in the coarse space and match to coarsened ensemble

OR

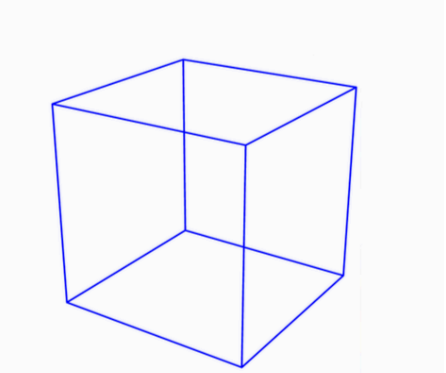
Solve regression problem directly: "Given a coarse ensemble, what parameters generated it?"

Multi-scale HMC updates

Perform HMC updates at coarse level

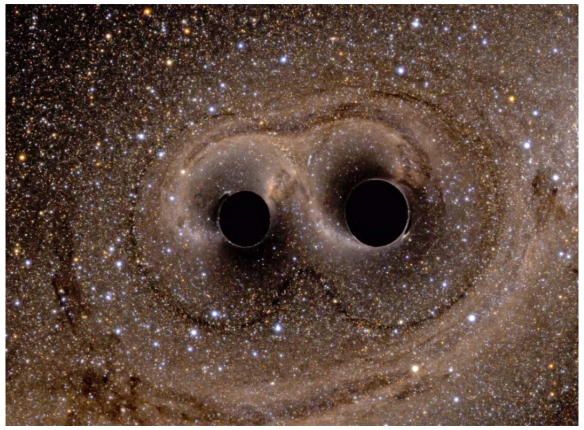


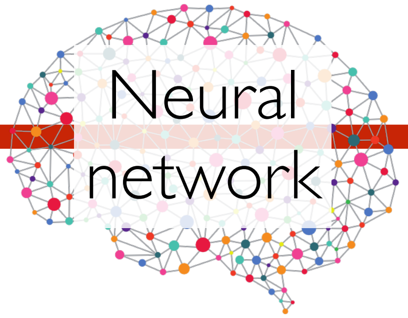
Match action parameters at different scales



Parameter regression is a problem suited to neural networks

Image





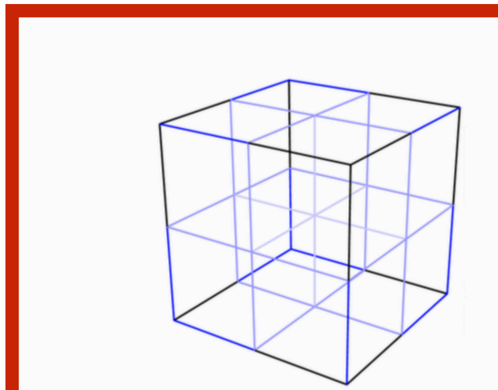
Neural network

Label

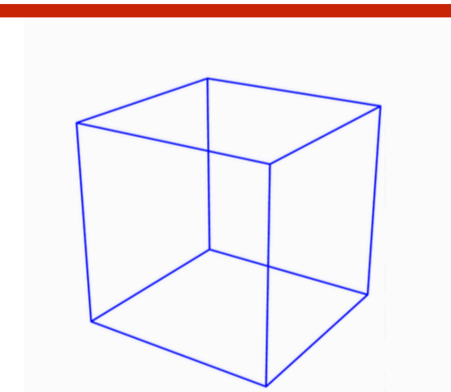
"Colliding black holes"

Multi-scale HMC updates

Perform HMC updates at coarse level

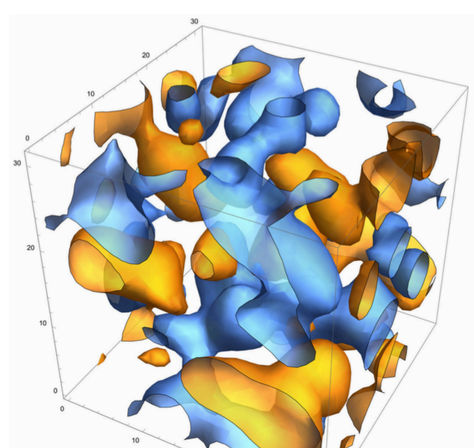


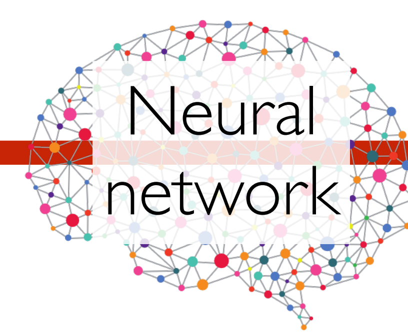
Match action parameters at different scales



Parameter regression is a problem suited to neural networks

Coarsened ensemble of lattice QCD gauge field configurations

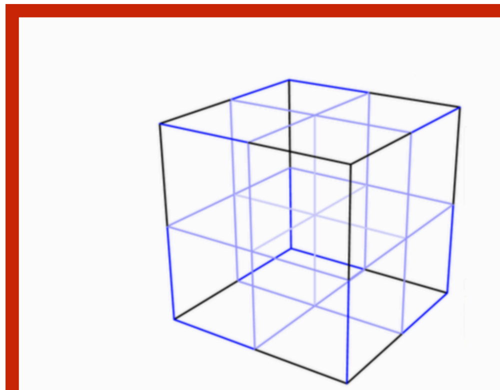




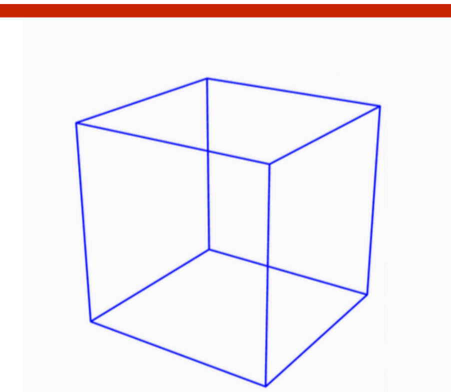
Label

Parameters of action

Parameter matching via NN



Match action parameters at different scales



- Parameter matching can be done on smaller (cheaper) ensembles than state-of-the-art target ensembles
- Regression does not need to be exact — corrected by rethermalisation (also, cannot be exact; a given gauge field could, with some probability, have been generated from a different action)
- All of the work is at the coarse scale

TASK: given a coarse configuration, find the corresponding action parameters

Training (and training data) needed in coarse space only

Machine learning LQCD

Ensemble of lattice QCD gauge fields

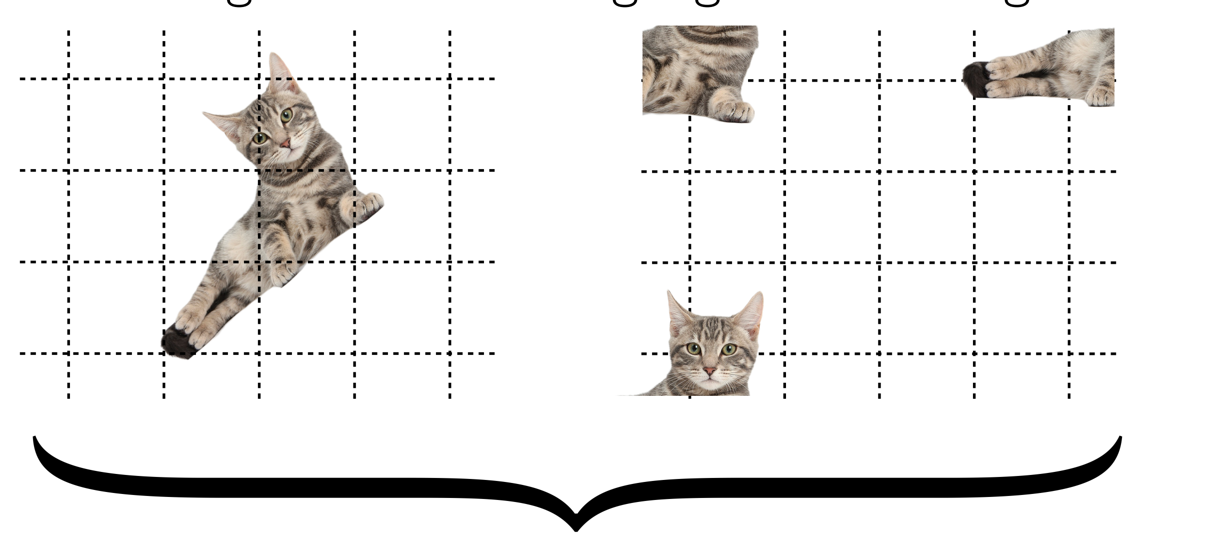
- $64^3 \times 128 \times 4 \times N_c^2 \times 2 \simeq 10^9$ numbers
- ~1000 samples
- Ensemble of gauge fields has meaning
- Long-distance correlations are important
- Gauge and translation-invariant with periodic boundaries

Physics is invariant under specific field transformations

- **Rotation, translation (4D), with boundary conditions**

Gauge field configuration

Transformed gauge field configuration



Encode same physics

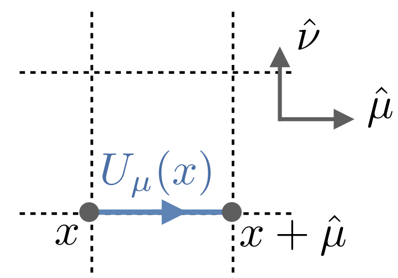
Machine learning LQCD

Ensemble of lattice QCD gauge fields

- $64^3 \times 128 \times 4 \times N_c^2 \times 2 \simeq 10^9$ numbers
- ~ 1000 samples
- Ensemble of gauge fields has meaning
- Long-distance correlations are important
- Gauge and translation-invariant with periodic boundaries

Physics is invariant under specific field transformations

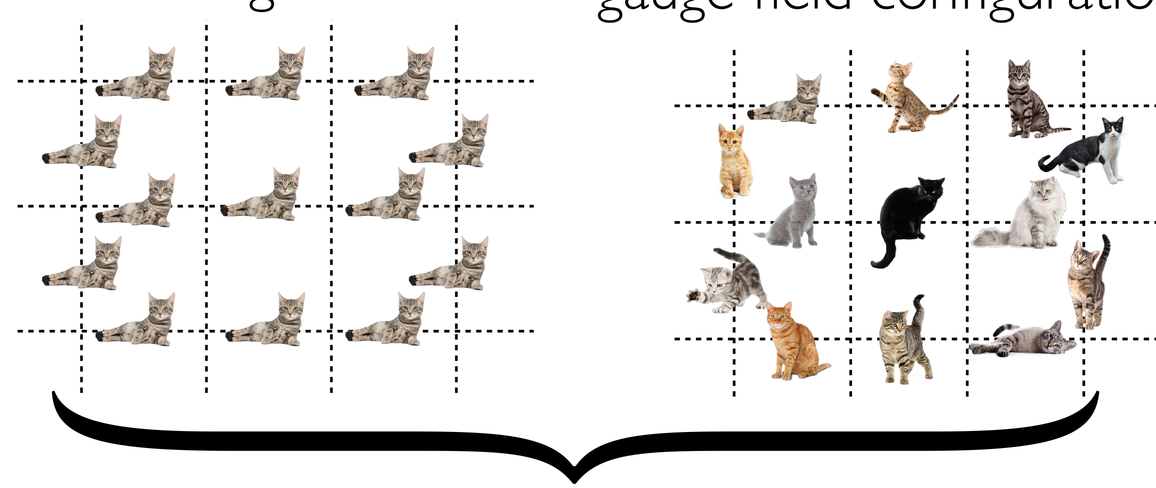
- **Gauge transformation**



$$U_\mu(x) \to \Omega(x) U_\mu(x) \Omega^\dagger(x+\hat{\mu})$$
$$\text{for all } \Omega(x) \in \mathrm{SU}(3)$$

Gauge field configuration

Transformed gauge field configuration



Encode same physics

Naive neural network

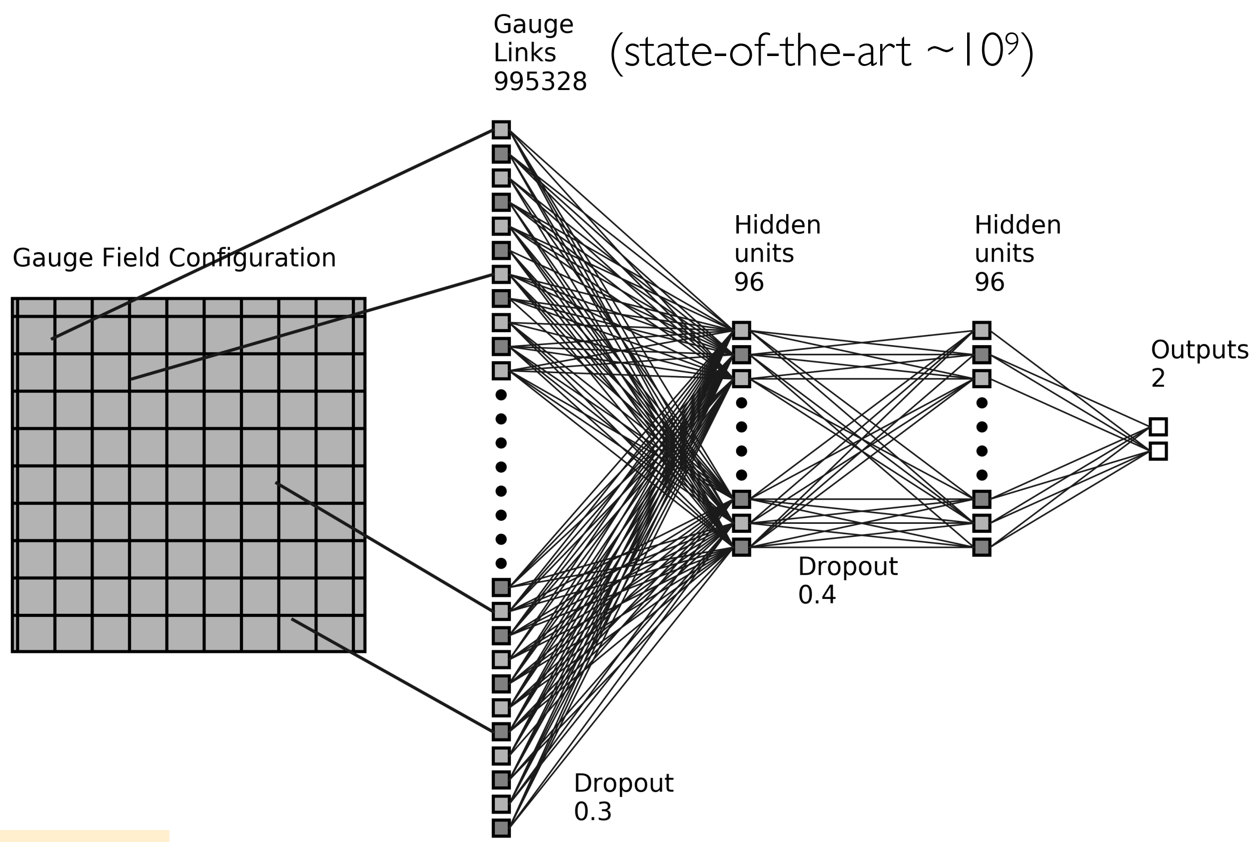
Simplest approach ⟺ Ignore physics symmetries

Train simple neural network on regression task

- Fully-connected structure
- Far more degrees of freedom than number of training samples available

"Inverted data hierarchy" → Recipe for overfitting!

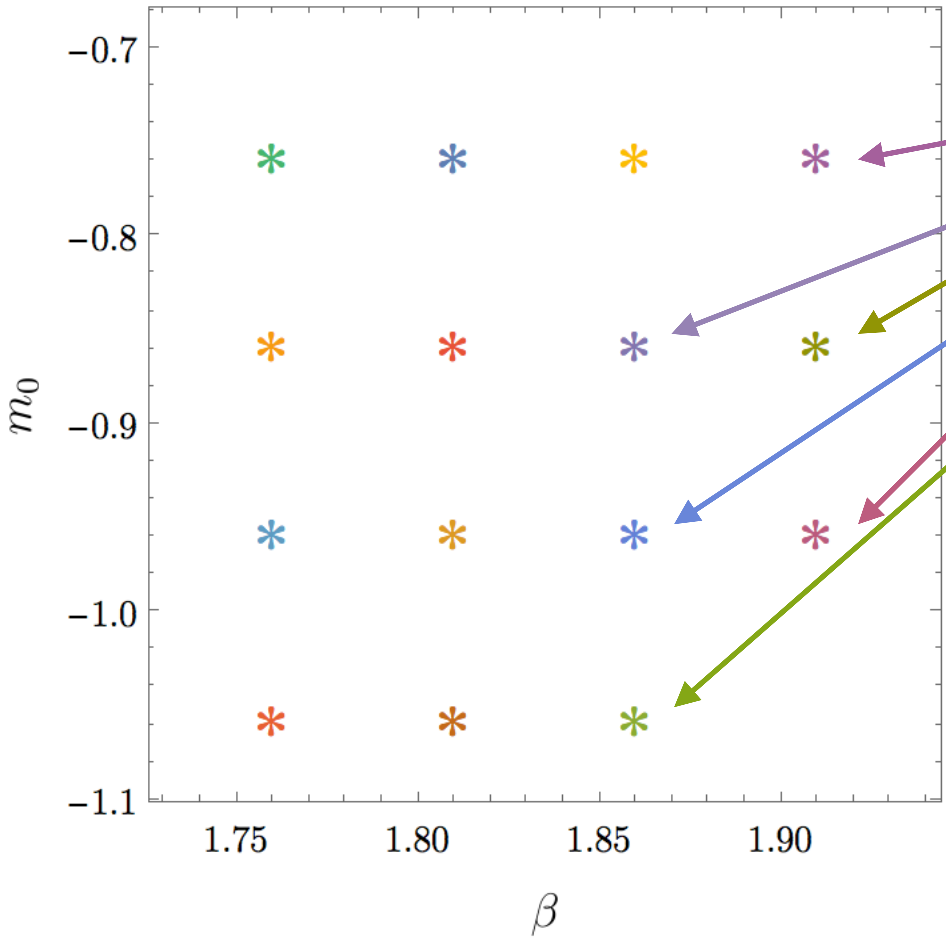
(state-of-the-art ~10^9)



Naive neural network

Training and validation datasets

Quark mass parameter



Parameter related to lattice spacing

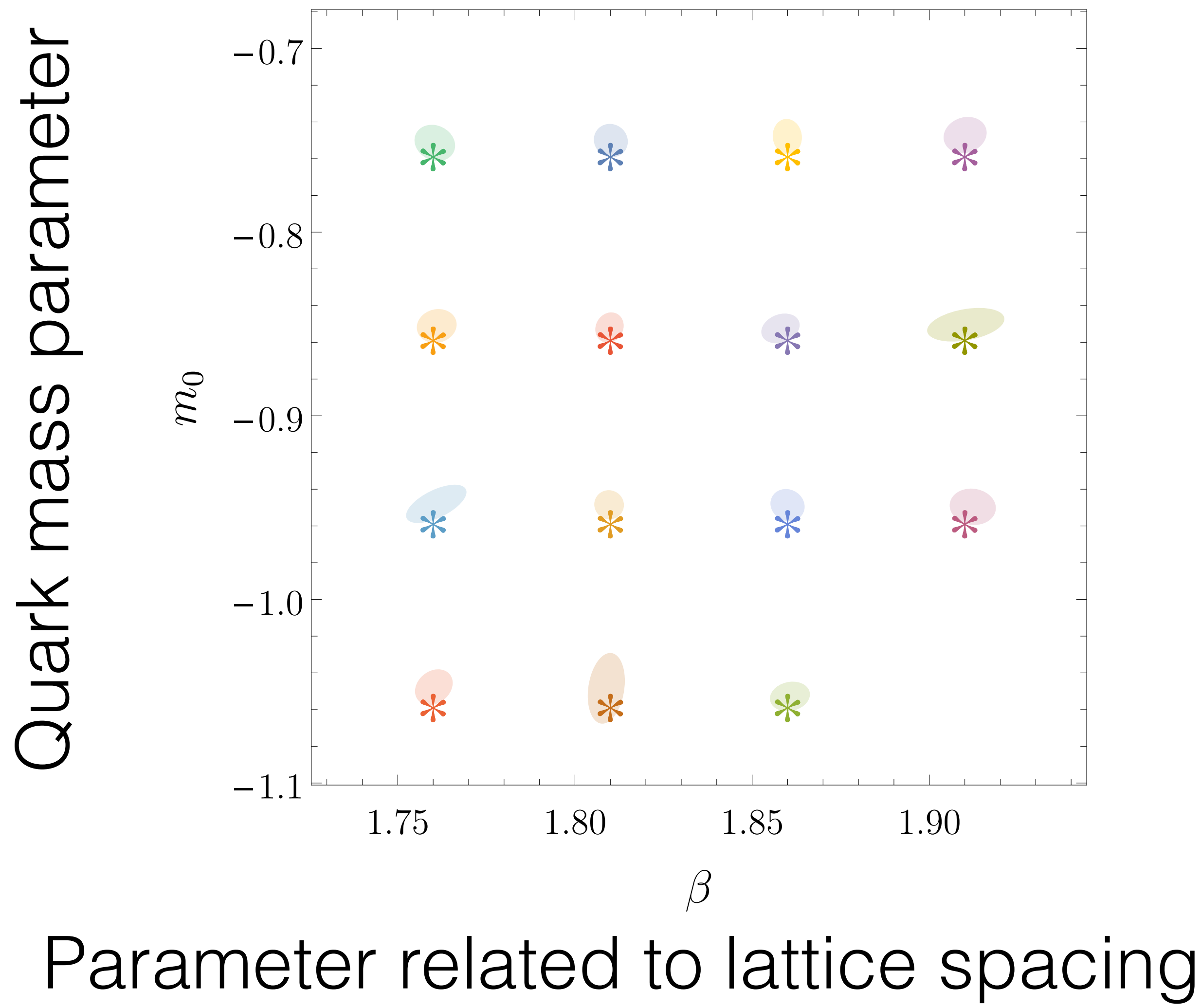
* Parameters of training and validation datasets

- O(10,000) independent configurations generated at each point
- Validation configurations randomly selected from generated streams

Spacing in evolution stream >> correlation time of physics observables

Naive neural network

Neural net predictions on validation data sets



* True parameter values

Confidence interval from ensemble of gauge fields

SUCCESS?

No sign of overfitting

- Training and validation loss equal
- Accurate predictions for validation data

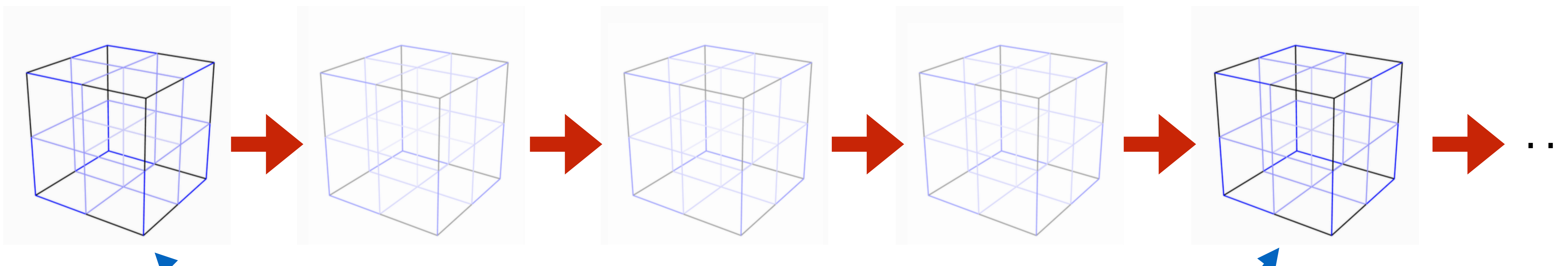
BUT fails to generalise to

- Ensembles at other parameters
- New streams at same parameters

NOT POSSIBLE IF CONFIGS ARE UNCORRELATED

Naive neural network

Stream of generated gauge fields at given parameters



Training/validation data selected from configurations spaced to be decorrelated (by physics observables)

- Network succeeds for validation configs from same stream as training configs
- Network fails for configs from new stream at same parameters

Network has identified feature with a longer correlation length than any known physics observable

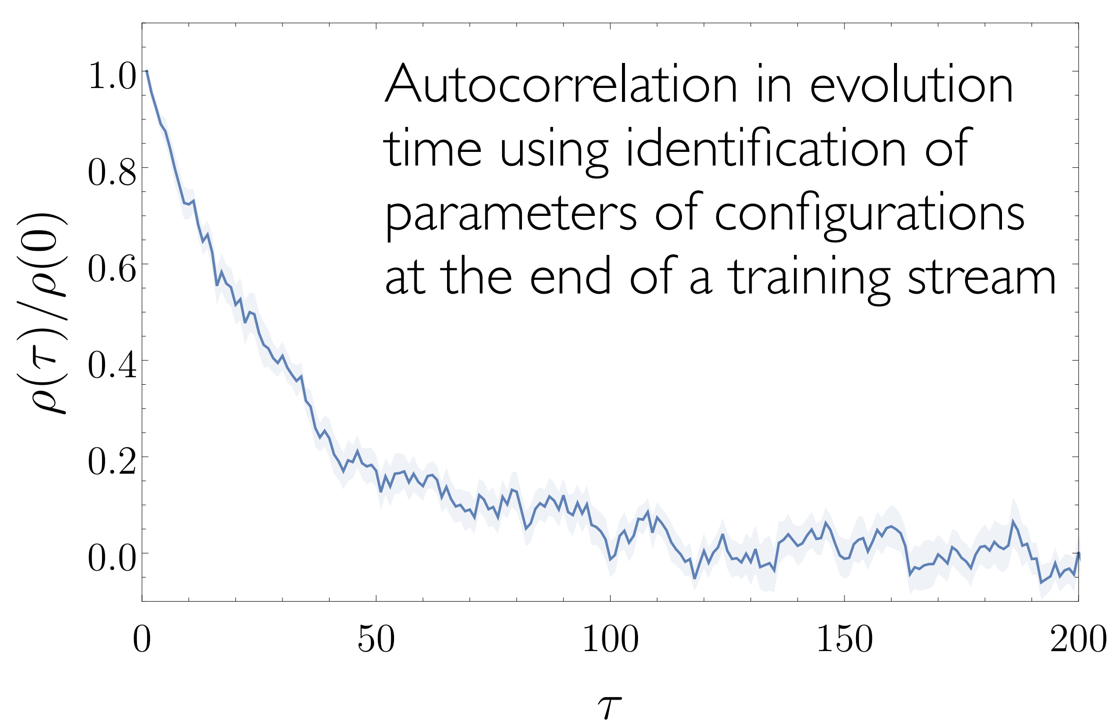
Naive neural network

- Naive neural network that does not respect symmetries fails at parameter regression task

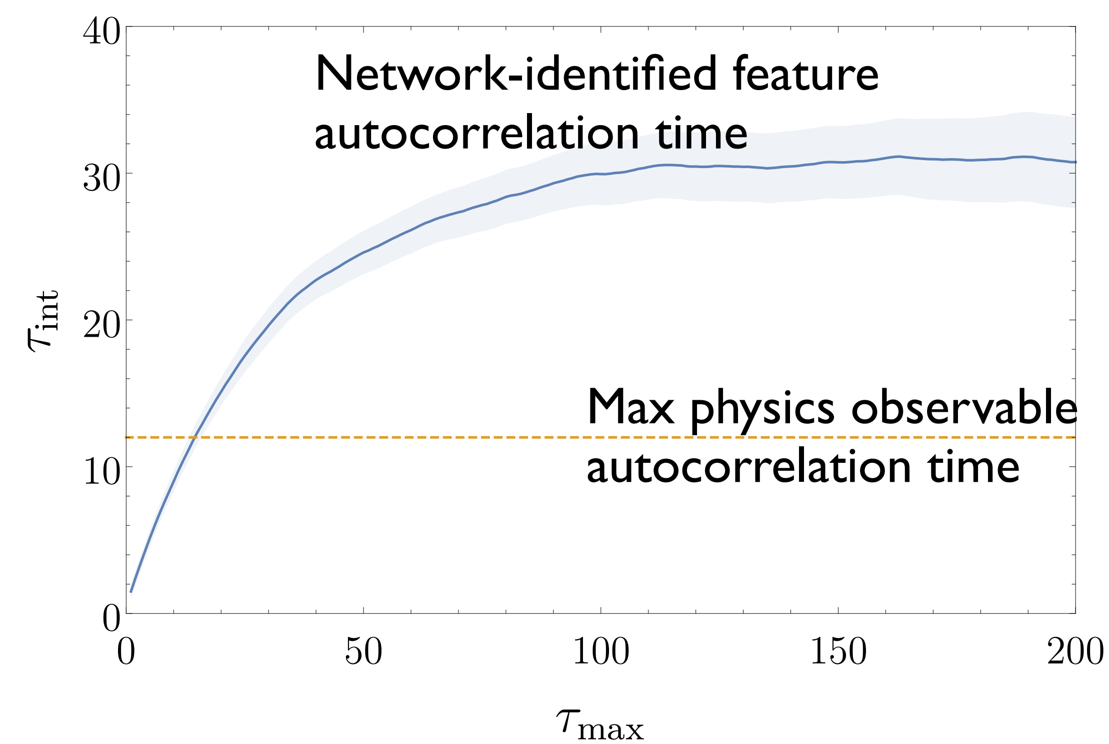
BUT

- Identifies unknown feature of gauge fields with a longer correlation length than any known physics observable

Network feature autocorrelation



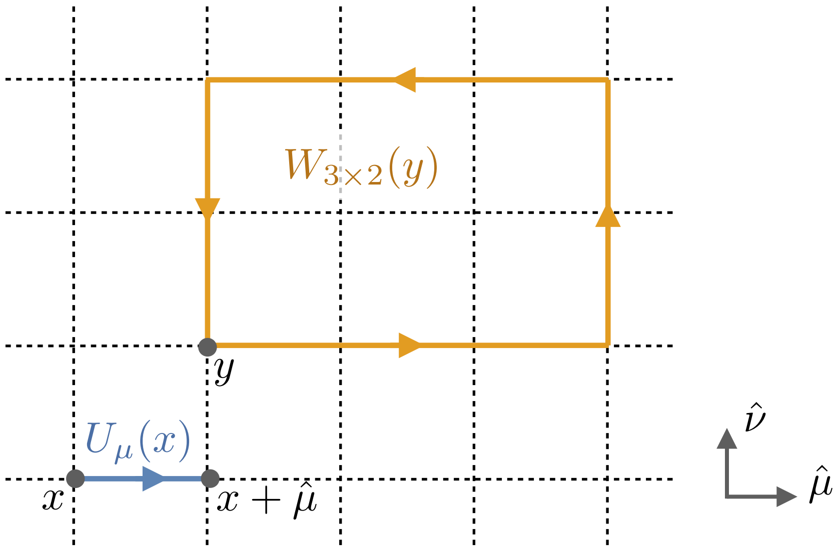
$$\tau_{\rm int} = \frac{1}{2} + \lim_{\tau_{\rm max}\to\infty} \frac{1}{\rho(0)} \sum_{\tau=0}^{\tau_{\rm max}} \rho(\tau)$$



Symmetry-preserving network

Network based on symmetry-invariant features

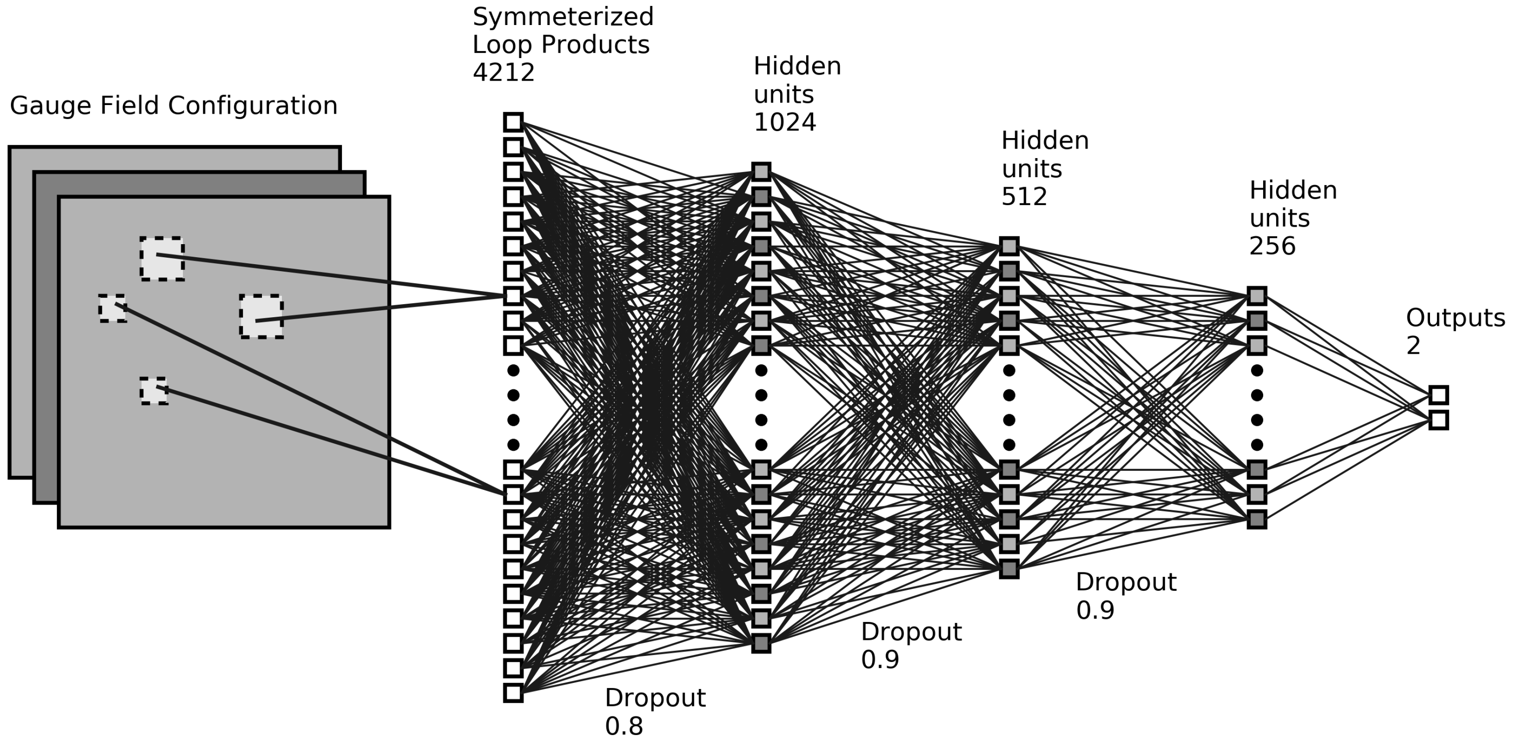
Closed Wilson loops
(gauge-invariant)



- Loops
- Correlated products of loops at various length scales
- Volume-averaged and rotation-averaged

Symmetry-preserving network

Network based on symmetry-invariant features



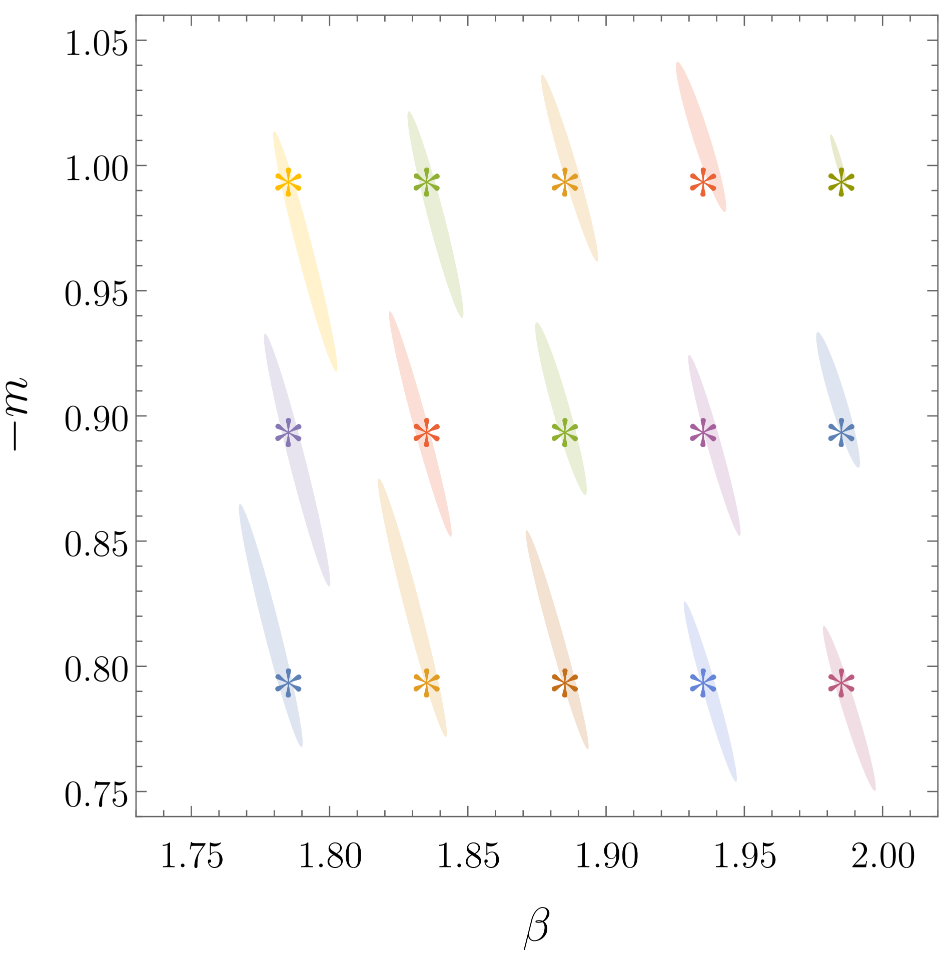
- Fully-connected network structure
- First layer samples from set of possible symmetry-invariant features

Number of degrees of freedom of network comparable to size of training dataset

Gauge field parameter regression

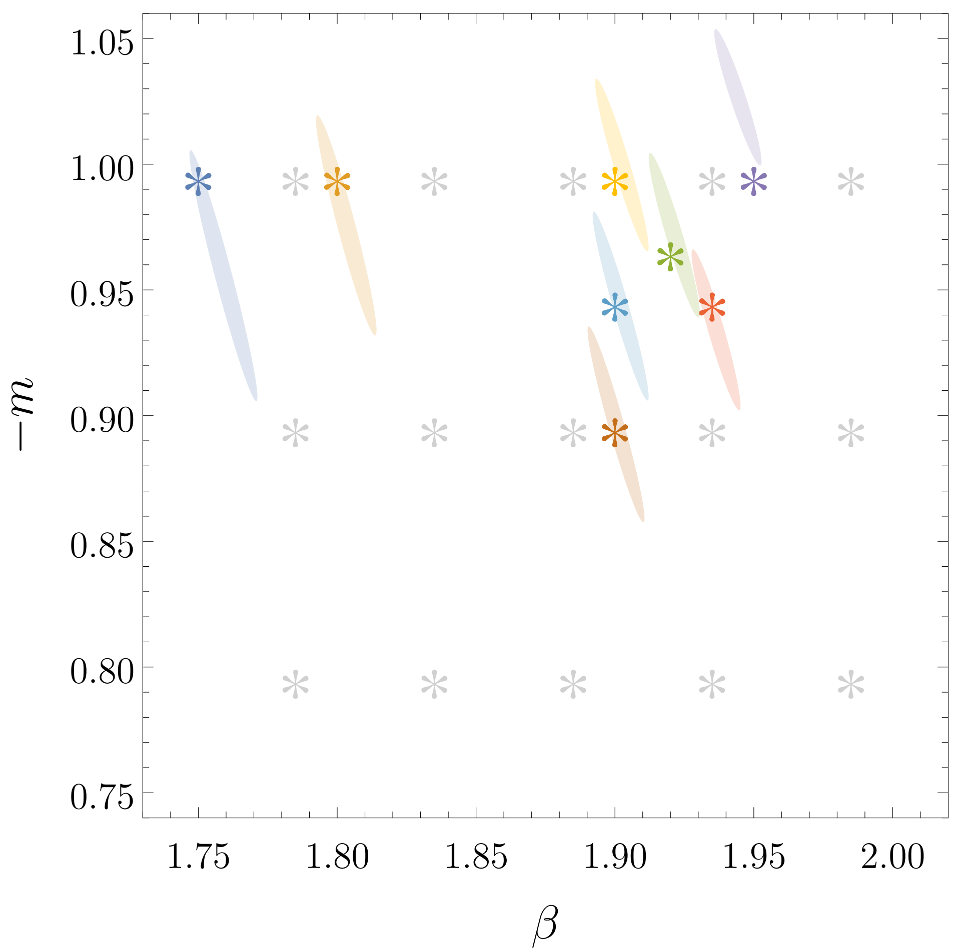
Neural net predictions on validation data sets

Quark mass parameter



Parameter related to lattice spacing

Predictions on new datasets



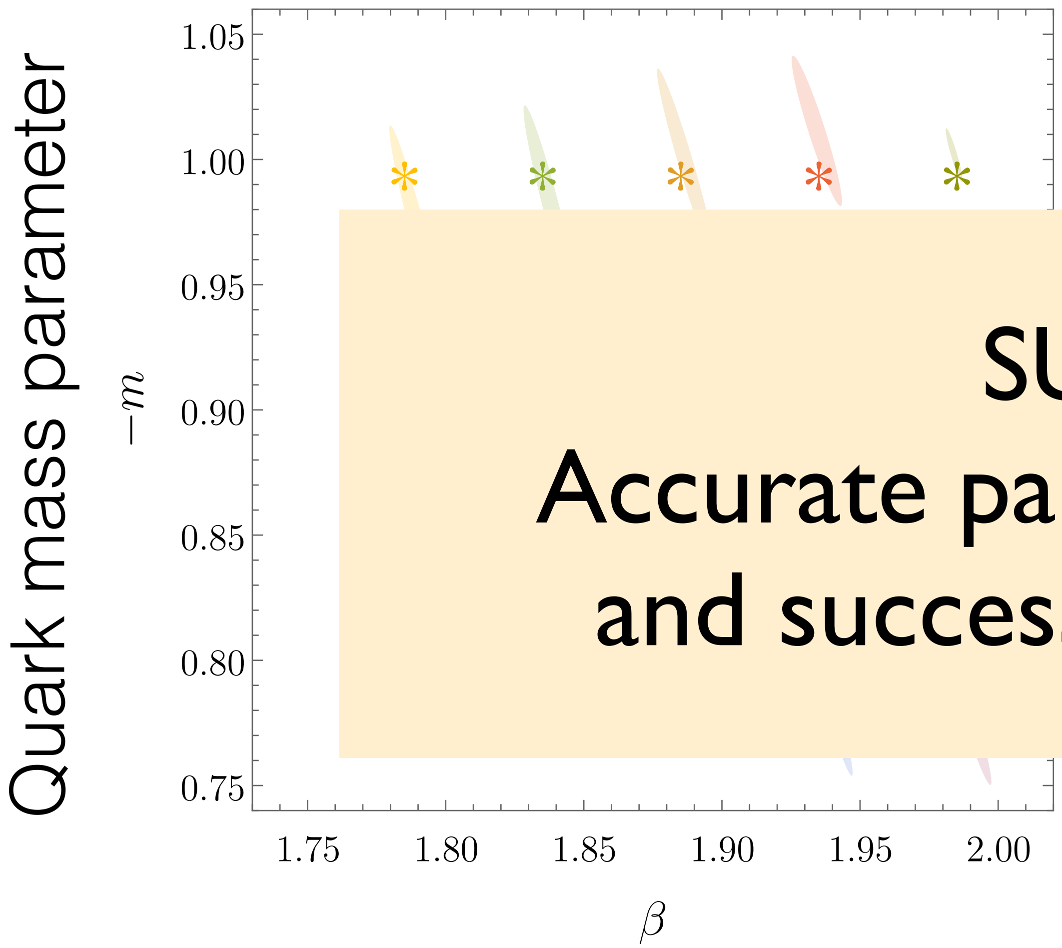
* True parameter values

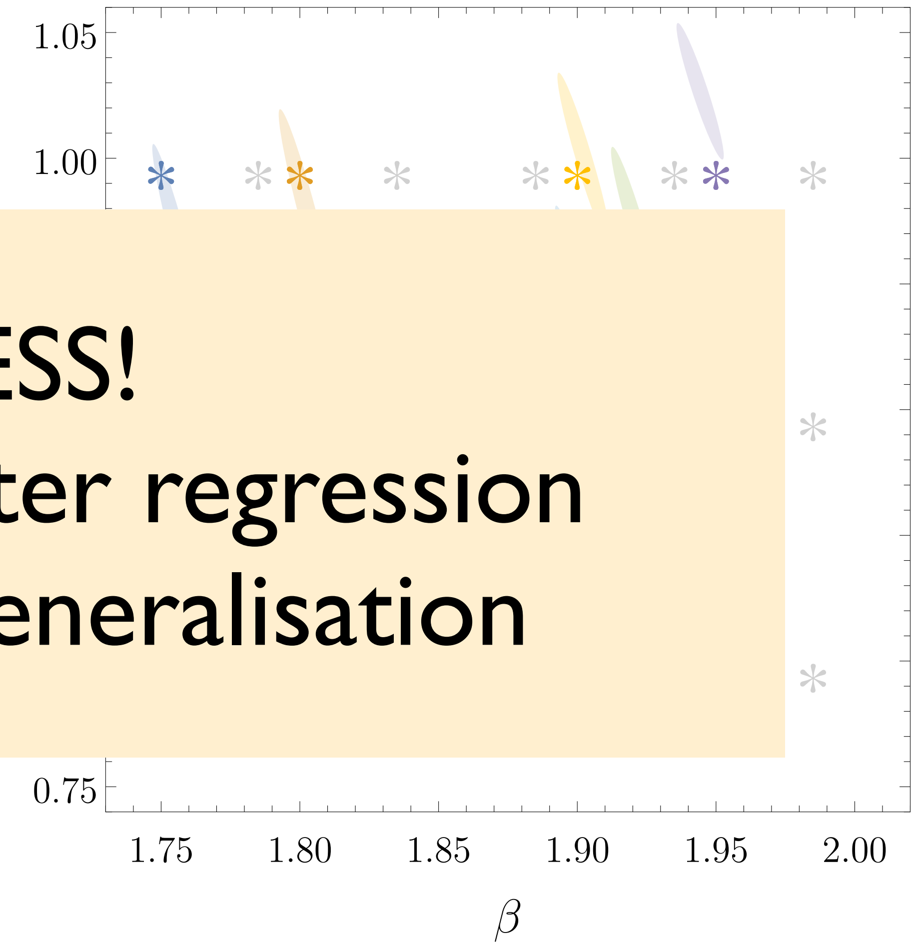
Confidence interval from ensemble of gauge fields

Gauge field parameter regression

Neural net predictions on validation data sets

Predictions on new datasets





SUCCESS!
Accurate parameter regression and successful generalisation

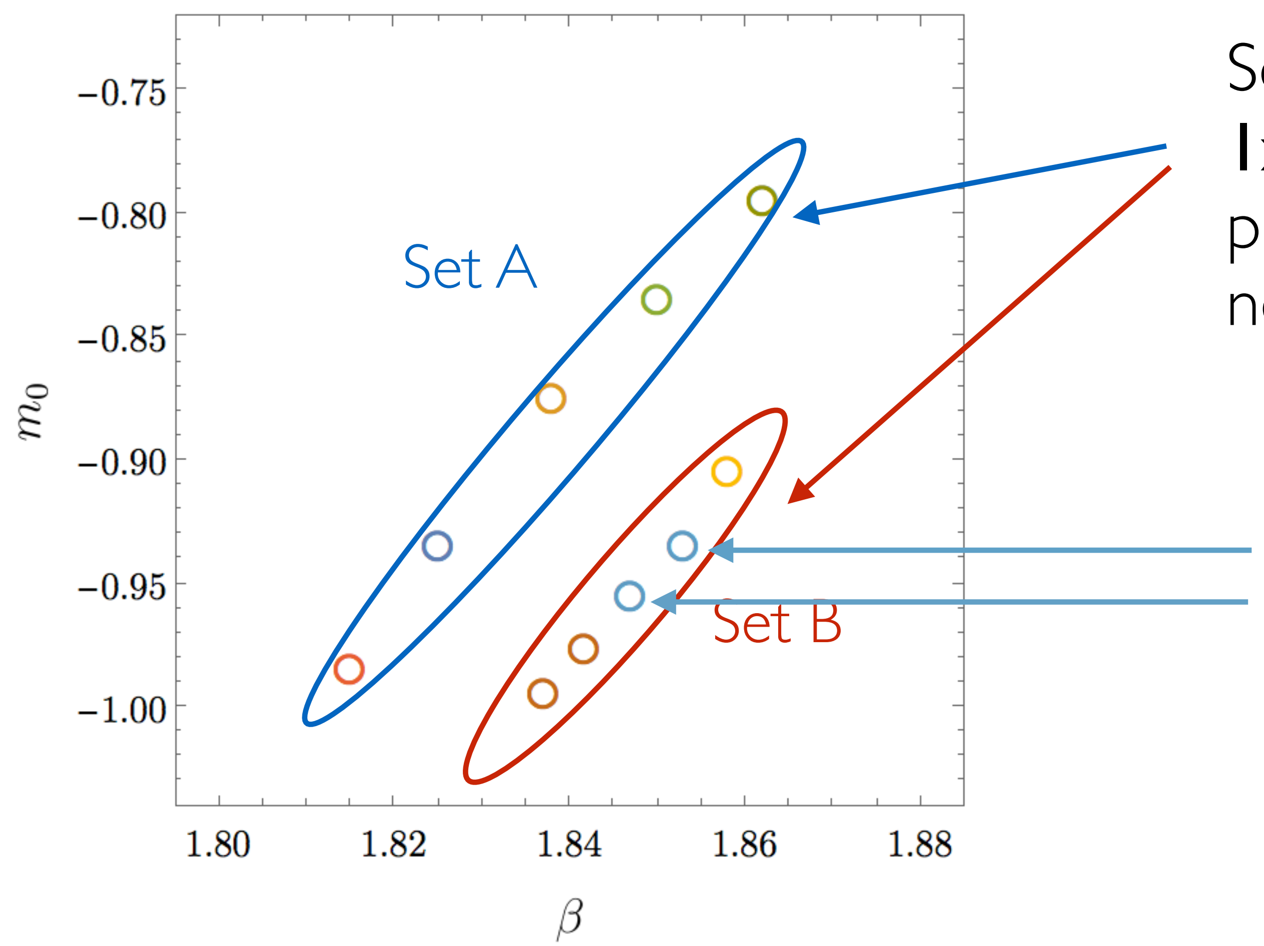
Parameter related to lattice spacing

* True parameter values
Confidence interval from ensemble of gauge fields

Tests of network success

How does neural network regression perform compared with other approaches?

Consider very closely-spaced validation ensembles at new parameters



Sets along lines of **constant 1x1 Wilson loop** (most precise feature allowed by network)

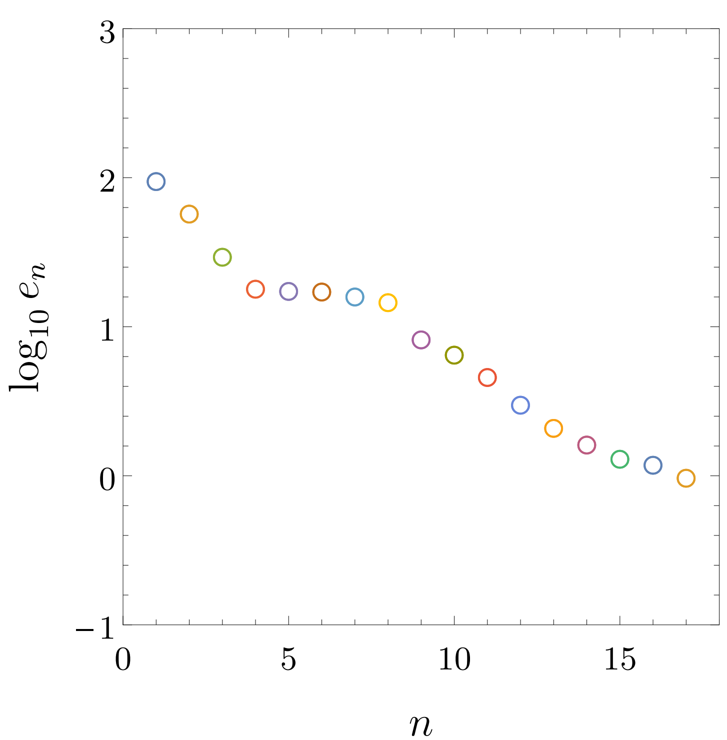
Much closer spacing than separation of training ensembles

Tests of network success

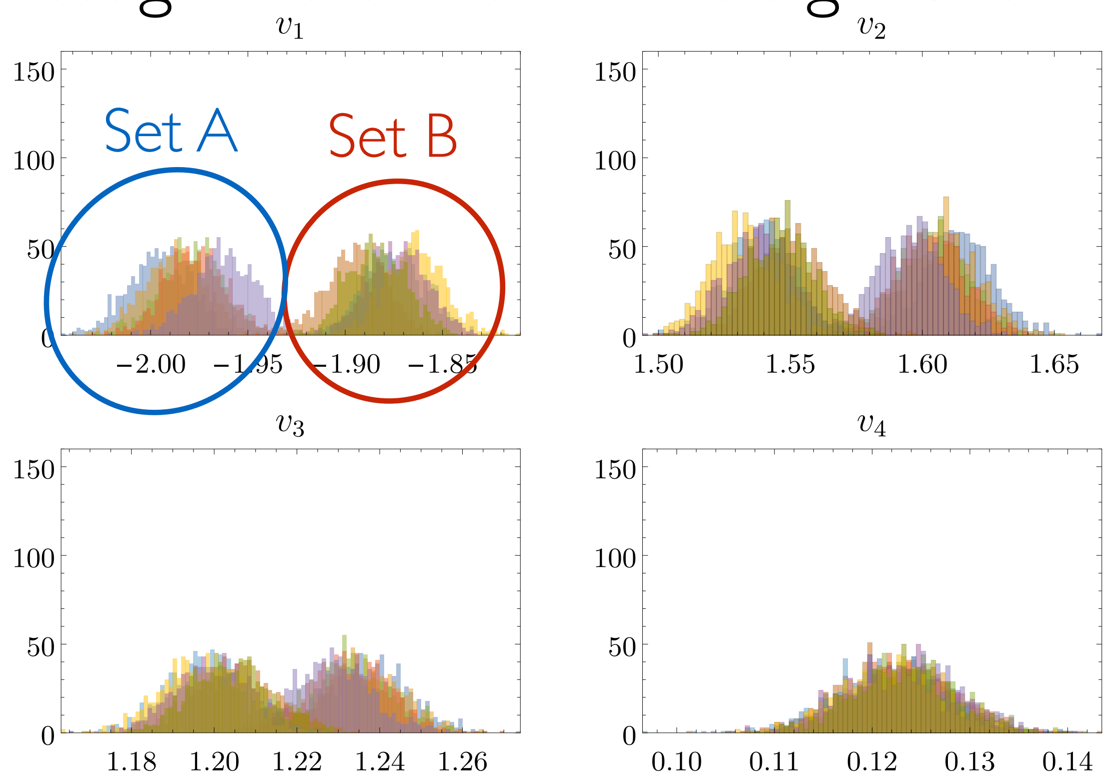
How does neural network regression perform compared with other approaches?

Consider very closely-spaced validation ensembles at new parameters: **not distinguishable to principal component analysis in loop space**

Eigenvalues



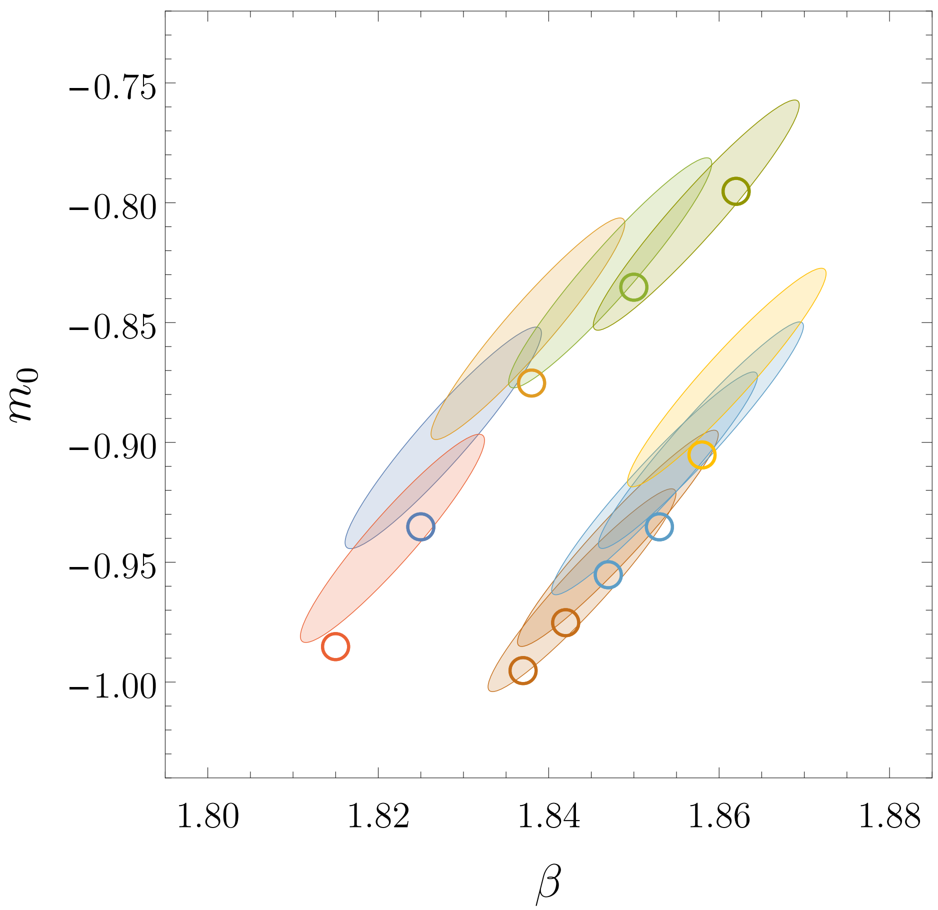
Histograms of dominant eigenvectors



Tests of network success

How does neural network regression perform compared with other approaches?

Consider very closely-spaced validation ensembles at new parameters: **distinguishable to trained neural network**



- Correct ordering of central values
- Accurate regression differences even at very fine resolution

Gauge field parameter regression

PROOF OF PRINCIPLE

Step towards fine lattice generation at reduced cost

1. Generate one fine configuration
2. Find matching coarse action
3. HMC updates in coarse space
4. Refine and rethermalise

Accurate matching minimises cost of updates in fine space

Guarantees correctness

Shanahan, Trewartha, Detmold, PRD (2018) [1801.05784]

Machine learning QCD

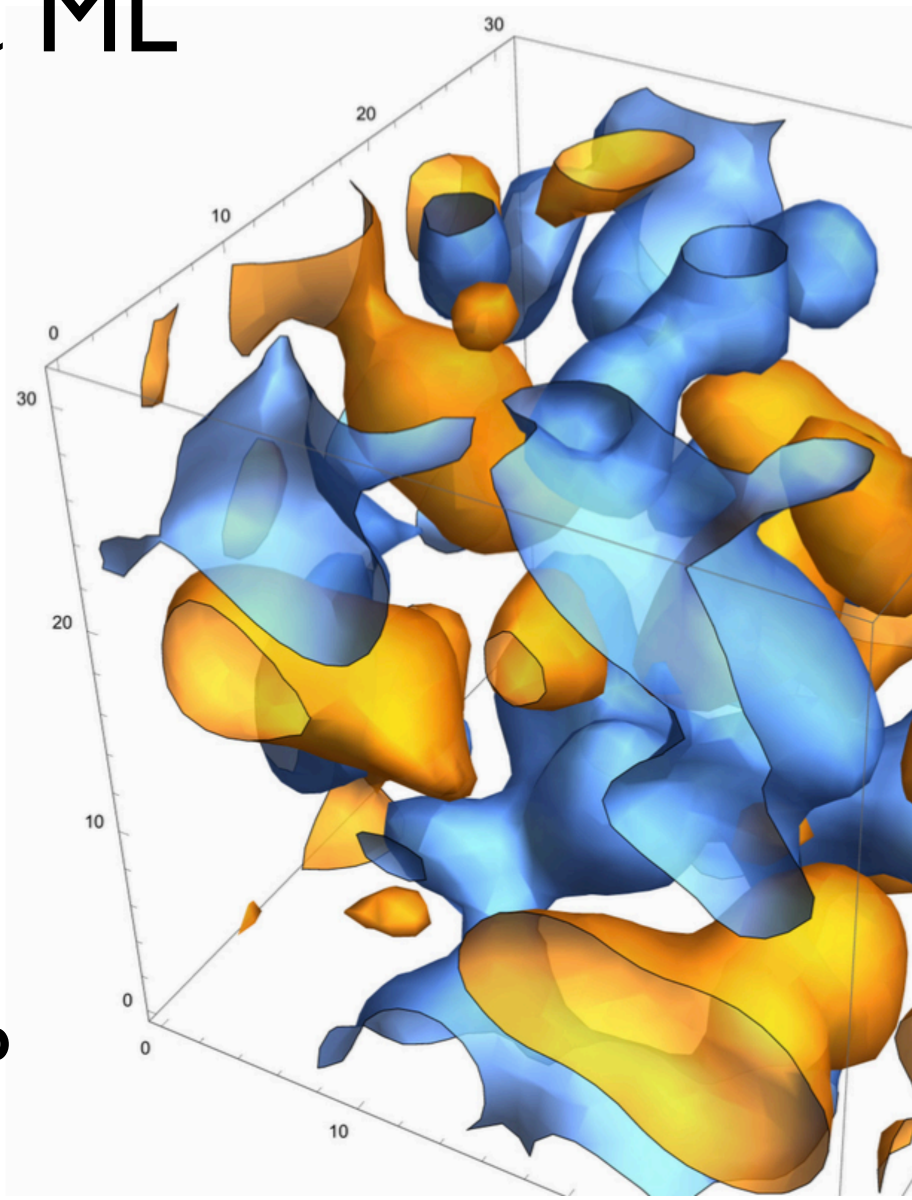
Accelerate gauge-field generation via ML

1. Multi-scale algorithms:
   parallels with image recognition
   Shanahan et al., PRD 97, 094506 (2018)

2. Generative models to replace Hybrid Monte-Carlo
   parallels with image generation
   Albergo et al., arXiv:1904.12072 (2019)

Gurtej Kanwar (MIT)

Michael Albergo (NYU)

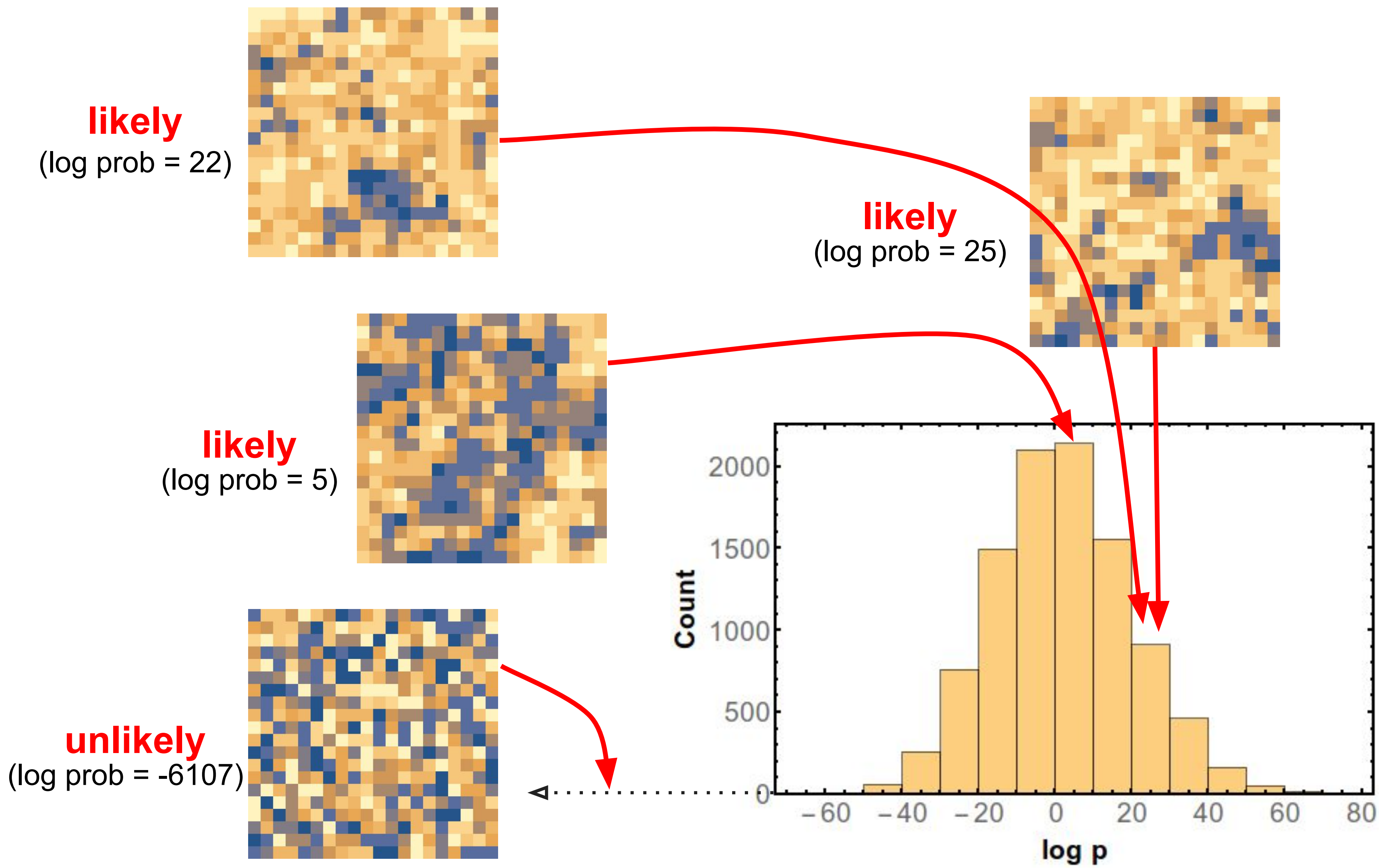


Consider only approaches which rigorously preserve quantum field theory in applicable limits

Sampling gauge field configs

Generate field configurations $\phi(x)$ with probability

$$P[\phi(x)] \sim e^{-S[\phi(x)]}$$



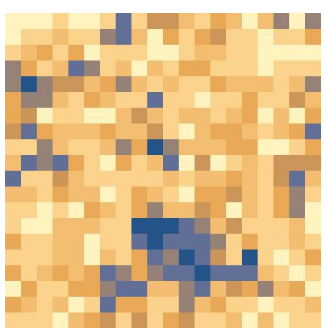
Sampling gauge field configs

Generate field configurations $\phi(x)$ with probability

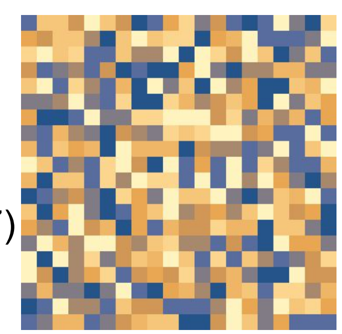
$$P[\phi(x)] \sim e^{-S[\phi(x)]}$$

Parallels with image generation problem

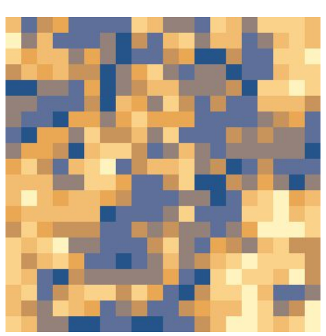
likely
(log prob = 22)



unlikely
(log prob = -6107)



likely
(log prob = 5)

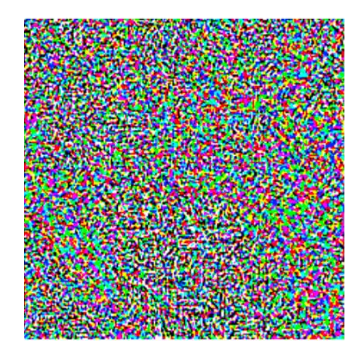


likely

[Karras, Lane, Aila / NVIDIA 1812.04948]

likely

unlikely



Sampling gauge field configs

- **Probability density can be computed for a given sample** (up to normalization)

$$p(..) = e^{-S(...)}/Z$$

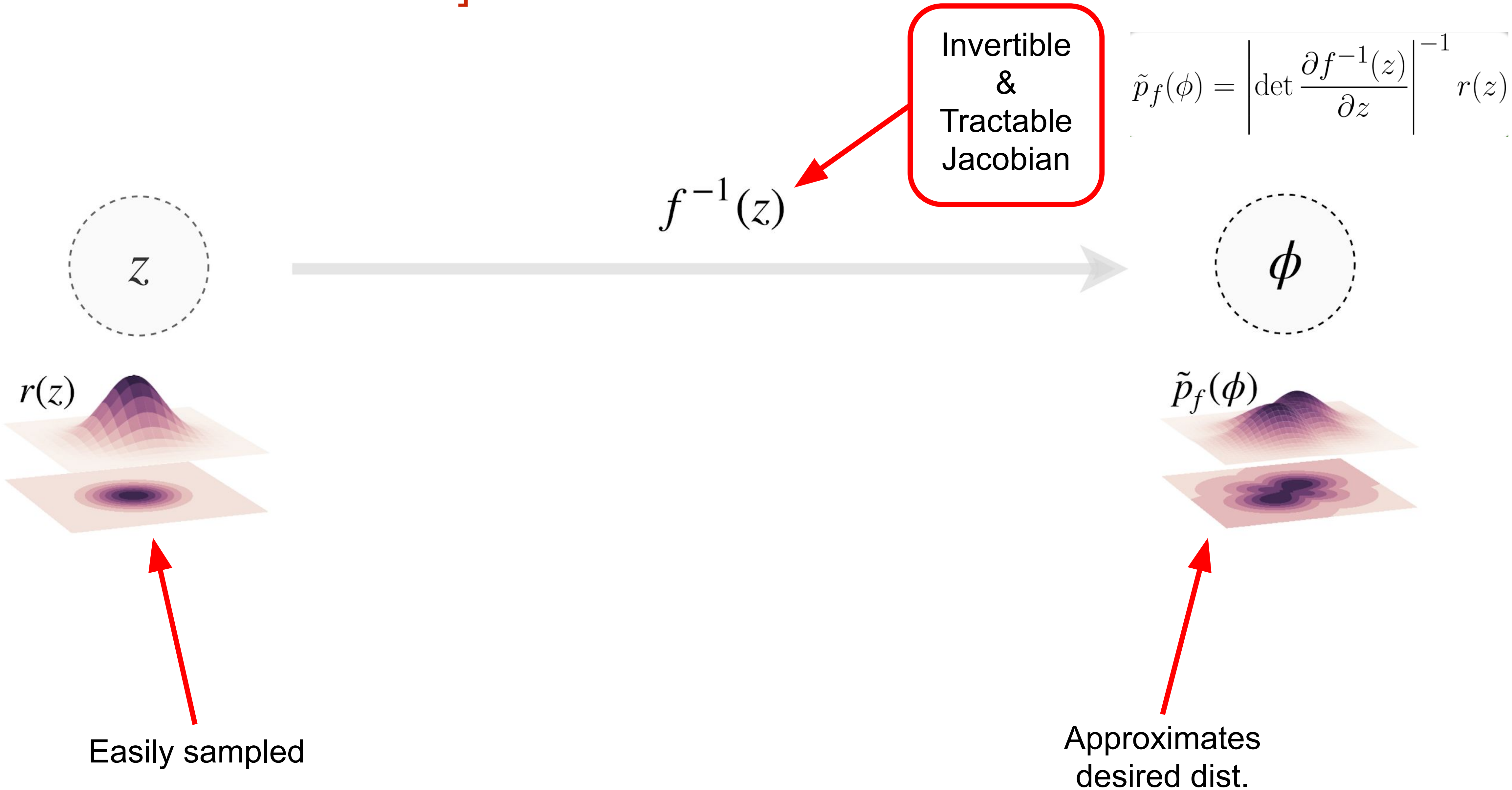
- **Physics distributions have precise symmetries**
  - Lattice symmetries (translation, rotation, reflection)
  - Internal symmetries (gauge symmetries mixing field components)
- **Data hierarchies are challenging**
  - 10^9 to 10^{12} variables per configuration
  - O(1000), samples available (fewer than # degrees of freedom per config)

→ **Hard to use training paradigms that rely on existing samples from distribution**

Generative flow models

Flow-based models learn a change-of-variables that transforms a known distribution to the desired distribution

[Rezende & Mohamed 1505.05770]

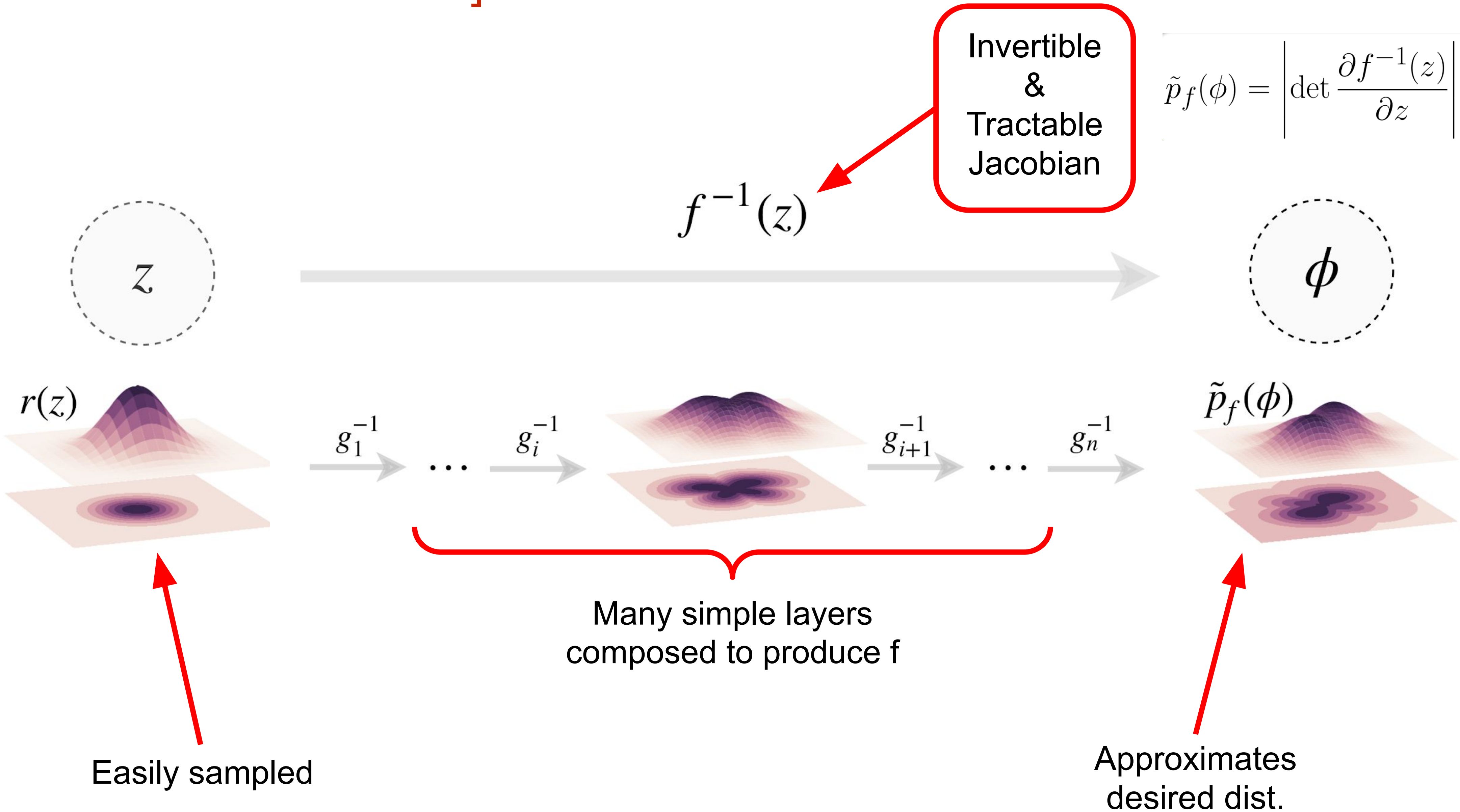


Generative flow models

Flow-based models learn a change-of-variables that transforms a known distribution to the desired distribution

[Rezende & Mohamed 1505.05770]

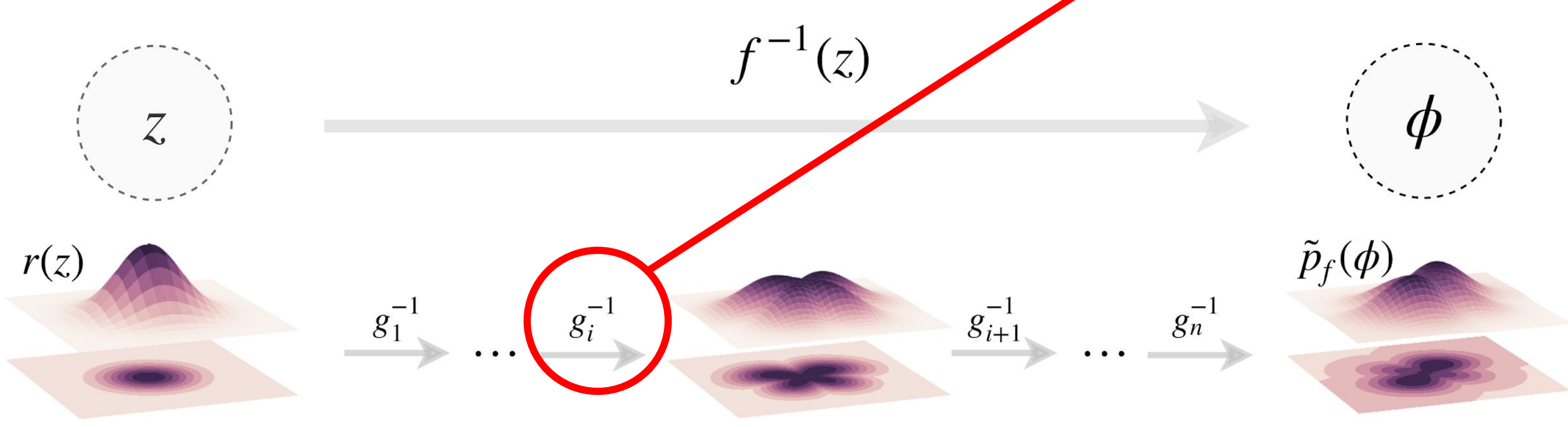
$$\tilde{p}_f(\phi) = \left| \det \frac{\partial f^{-1}(z)}{\partial z} \right|^{-1} r(z)$$

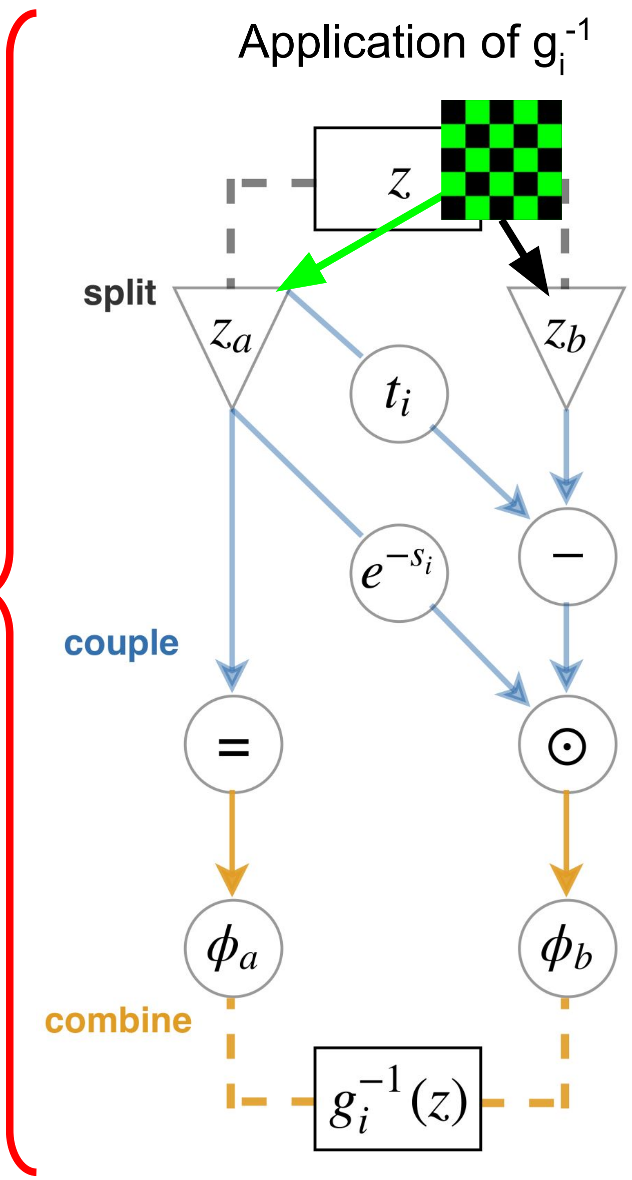


Generative flow models

Choose real non-volume preserving flows:
[Dinh et al. 1605.08803]

- Affine transformation of half of the variables:
  - scaling by exp(s)
  - translation by t
  - s and t arbitrary neural networks depending on untransformed variables only
- Simple inverse and Jacobian



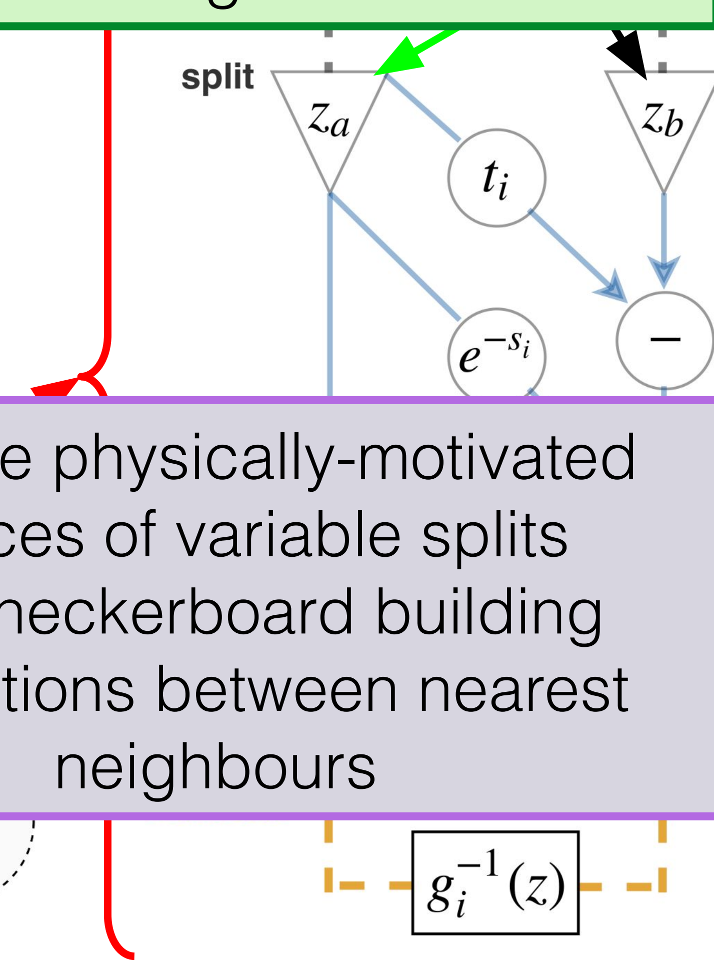


Generative flow models

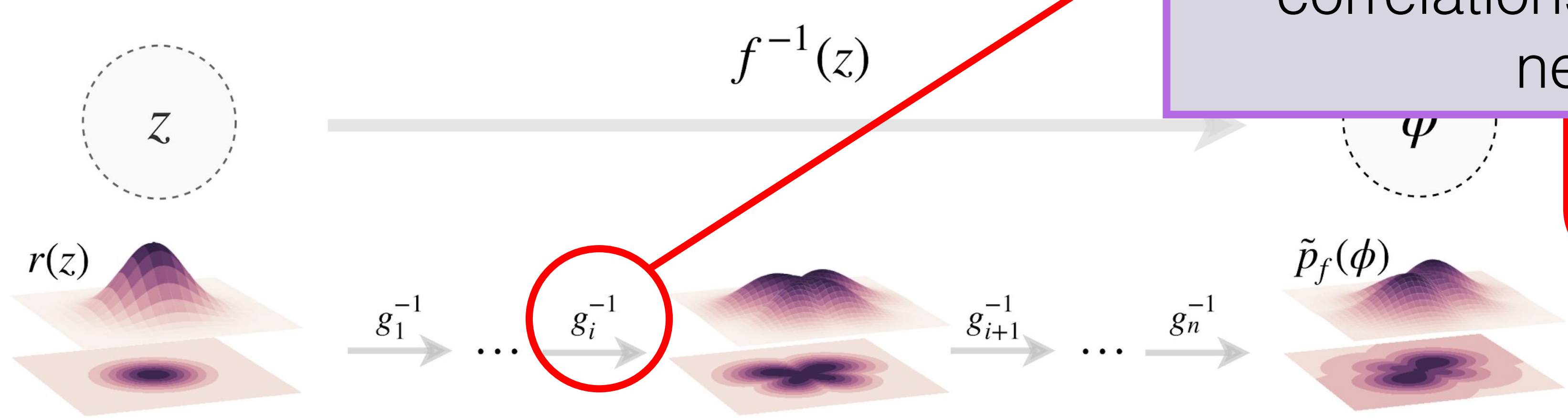
Choose real **non-volume preserving** flows:
[Dinh et al. 1605.08803]

- Affine transformation of **half of the variables**:
  - scaling by exp(s)
  - translation by t
  - s and t arbitrary neural networks depending on untransformed variables only
- Simple inverse and Jacobian

Density can be squished/stretched by change-of-variables



Can use physically-motivated choices of variable splits e.g. checkerboard building correlations between nearest neighbours



Training the model

Target distribution is known up to normalisation

$$p(\phi) = e^{-S(\phi)}/Z$$

Train to minimise shifted KL divergence: [Zhang, E, Wang 1809.10188]

shift removes unknown normalisation Z

$$\begin{aligned} L(\tilde{p}_f) &:= D_{KL}(\tilde{p}_f || p) - \log Z \\ &= \underbrace{\int \prod_j d\phi_j \, \tilde{p}_f(\phi)} \left(\log \tilde{p}_f(\phi) + S(\phi) \right) \end{aligned}$$

allows **self-training**: sampling with respect to model distribution $\tilde{p}_f(\phi)$ to estimate loss

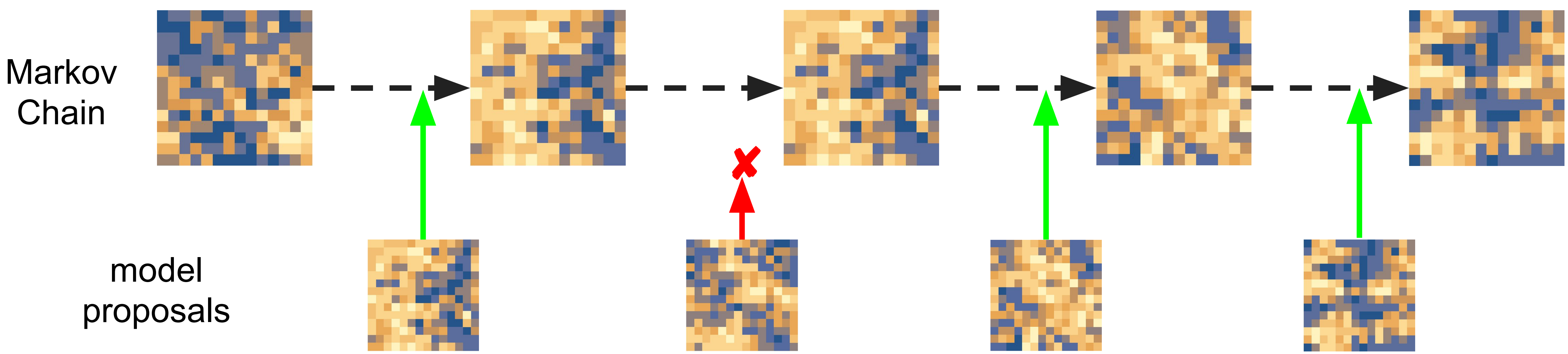
Exactness via Markov chain

Guarantee exactness of generated distribution by forming a Markov chain: accept/reject with Metropolis-Hastings step

Acceptance probability

$$A(\phi^{(i-1)}, \phi') = \min\left(1, \frac{\tilde{p}(\phi^{(i-1)})}{p(\phi^{(i-1)})} \frac{p(\phi')}{\tilde{p}(\phi')}\right)$$

proposal independent of previous sample



Exactness via Markov chain

Guarantee exactness of generated distribution by forming a Markov chain: accept/reject with Metropolis-Hastings step

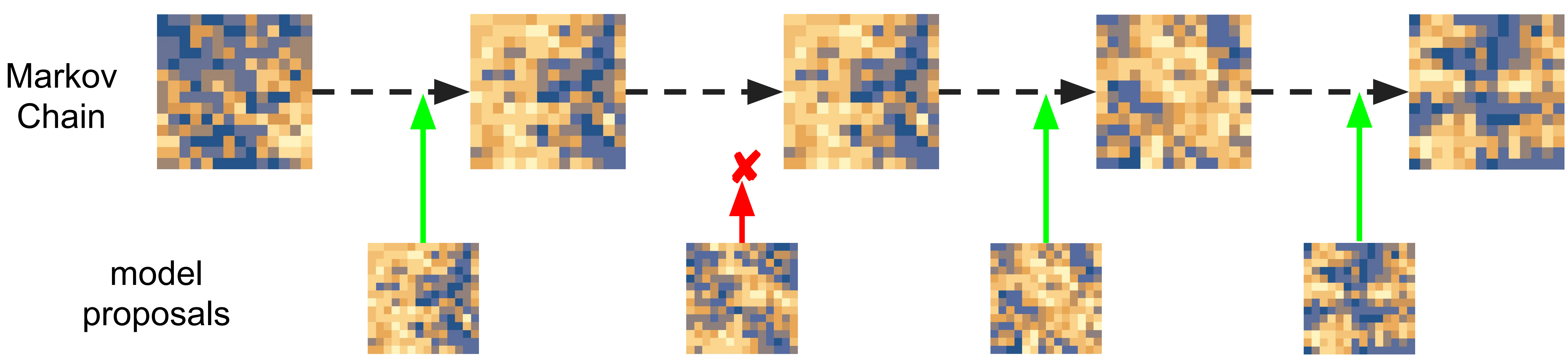
Acceptance probability

$$A(\phi^{(i-1)}, \phi') = \min\left(1, \frac{\boxed{\tilde{p}(\phi^{(i-1)})}}{\boxed{p(\phi^{(i-1)})}} \frac{\boxed{p(\phi')}}{\boxed{\tilde{p}(\phi')}}\right)$$

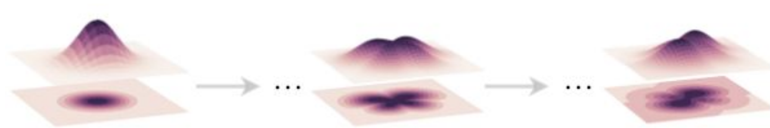
True dist

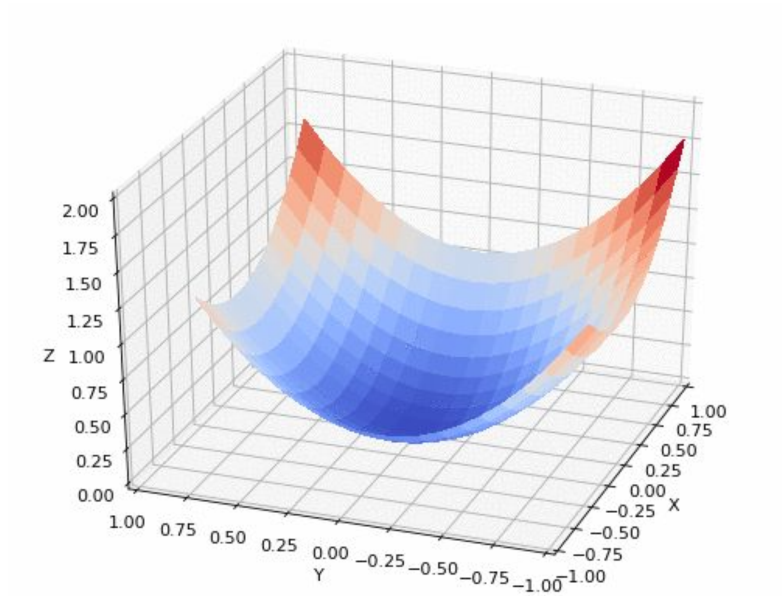
Model dist

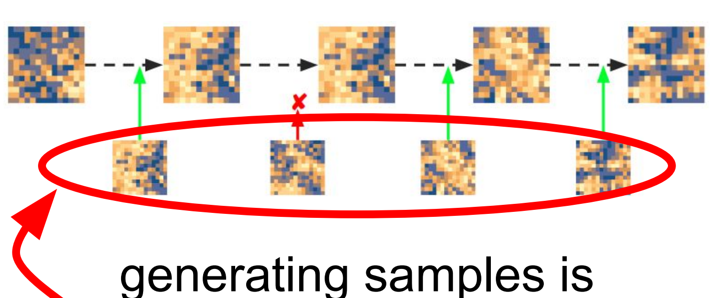
proposal independent of previous sample



Fields via flow models







generating samples is
"embarrassingly parallel"

Parameterize flow using Real NVP coupling layers

Each layer contains arbitrary neural nets **s** and **t**

Training step

Draw samples from model

Compute loss function

Gradient descent

Desired accuracy?

Save trained model

Markov chain using samples from model

Summary chart: Tej Kanwar

Application: scalar field theory

First application: scalar lattice field theory

- One real number $\phi(x) \in (-\infty, \infty)$ per lattice site x (2D lattice)
- Action: kinetic terms and quartic coupling

$$S(\phi) = \sum_x \left(\sum_y \frac{1}{2}\phi(x)\Box(x,y)\phi(y) + \frac{1}{2}m^2\phi(x)^2 + \lambda\phi(x)^4 \right)$$

5 lattice sizes: $L^2 = \{6^2, 8^2, 10^2, 12^2, 14^2\}$ with parameters tuned for analysis of critical slowing down

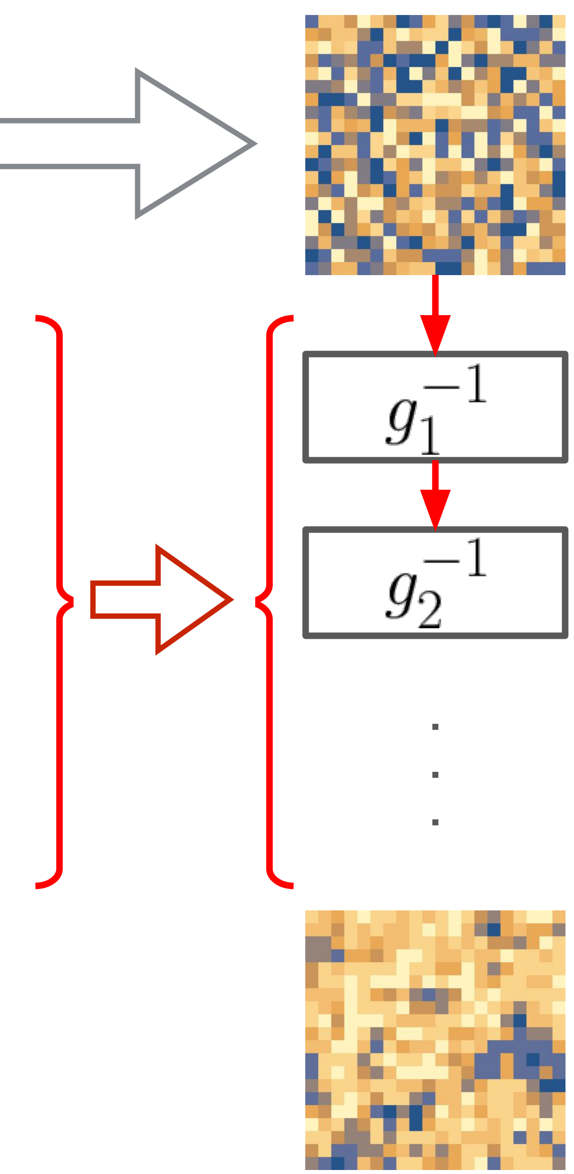
	E1	E2	E3	E4	E5
L	6	8	10	12	14
m^2	-4	-4	-4	-4	-4
λ	6.975	6.008	5.550	5.276	5.113
$m_p L$	3.96(3)	3.97(5)	4.00(4)	3.96(5)	4.03(6)

Application: scalar field theory

First application: scalar lattice field theory

- Prior distribution chosen to be uncorrelated Gaussian:

$$\phi(x) \sim \mathcal{N}(0,1)$$



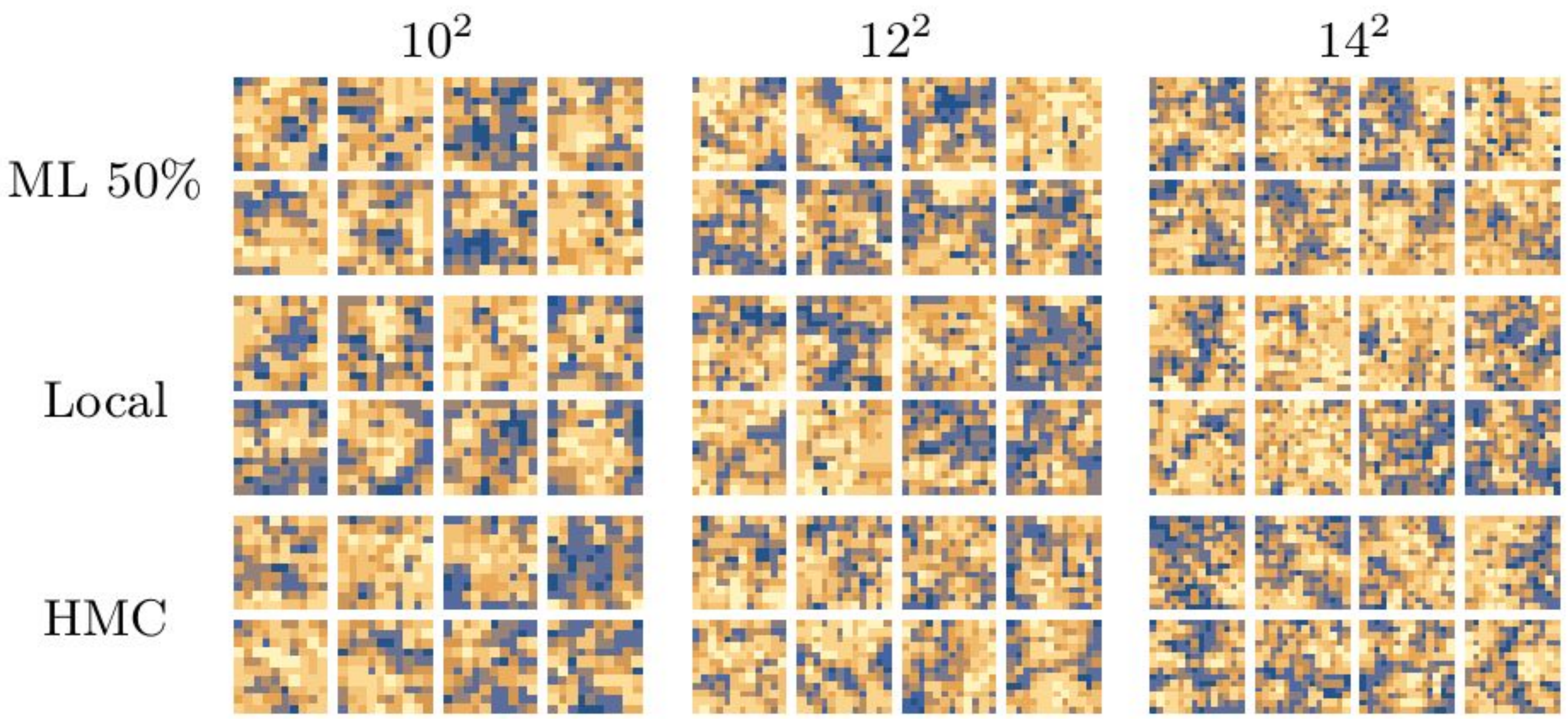
- Real non-volume-preserving (NVP) couplings
  - 8-12 Real NVP coupling layers
  - Alternating checkerboard pattern for variable split
  - NNs with 2-6 fully connected layers with 100-1024 hidden units

- Train using shifted KL loss with Adam optimizer
  - Stopping criterion: fixed acceptance rate in Metropolis-Hastings MCMC

Application: scalar field theory

First application: scalar lattice field theory

Compare with standard updating algorithms: 'local', 'HMC'

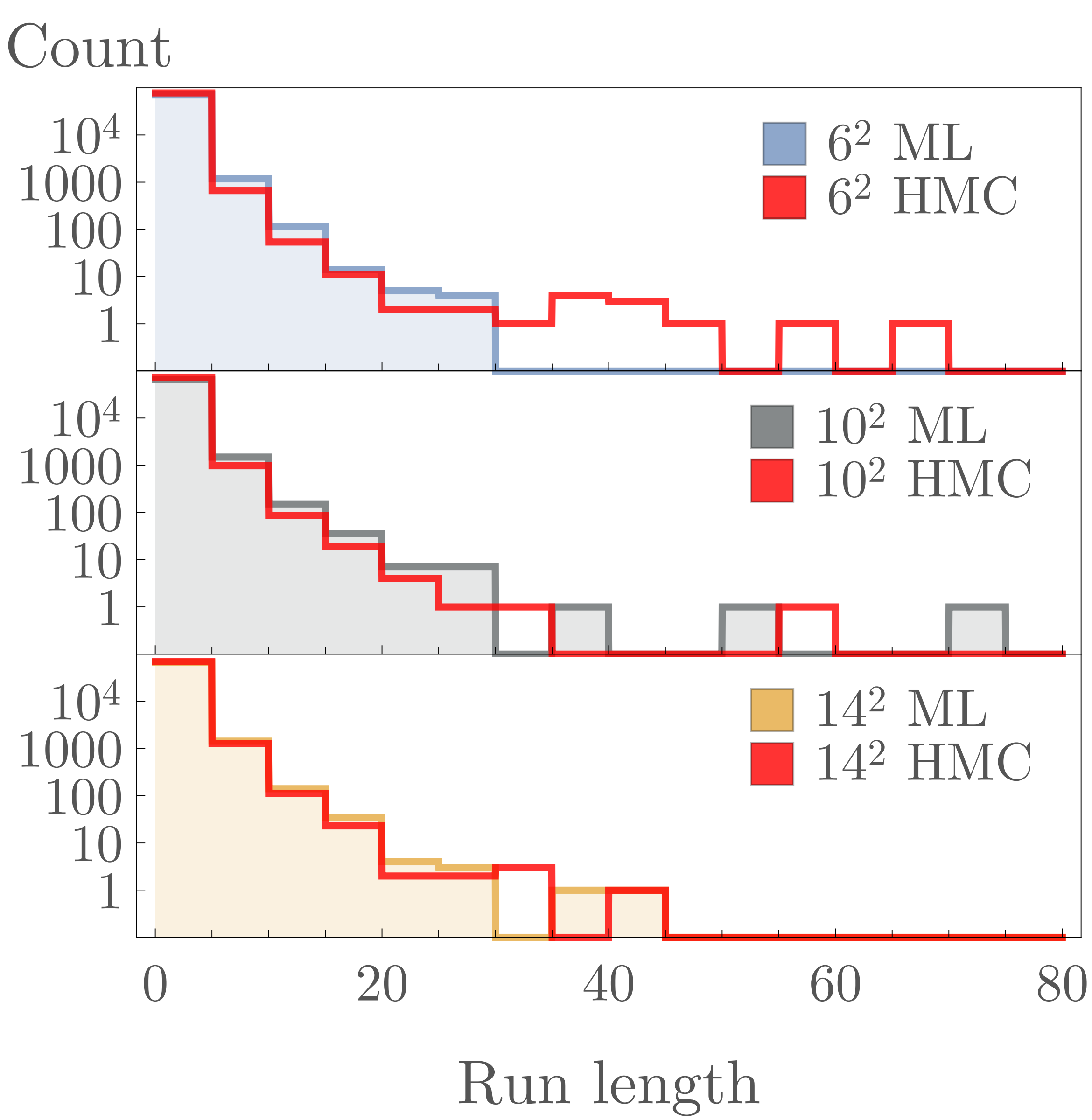


ML model produces **varied samples** and **correlations at the right scale**

Application: scalar field theory

First application: scalar lattice field theory

Compare with standard updating algorithms: 'local', 'HMC'

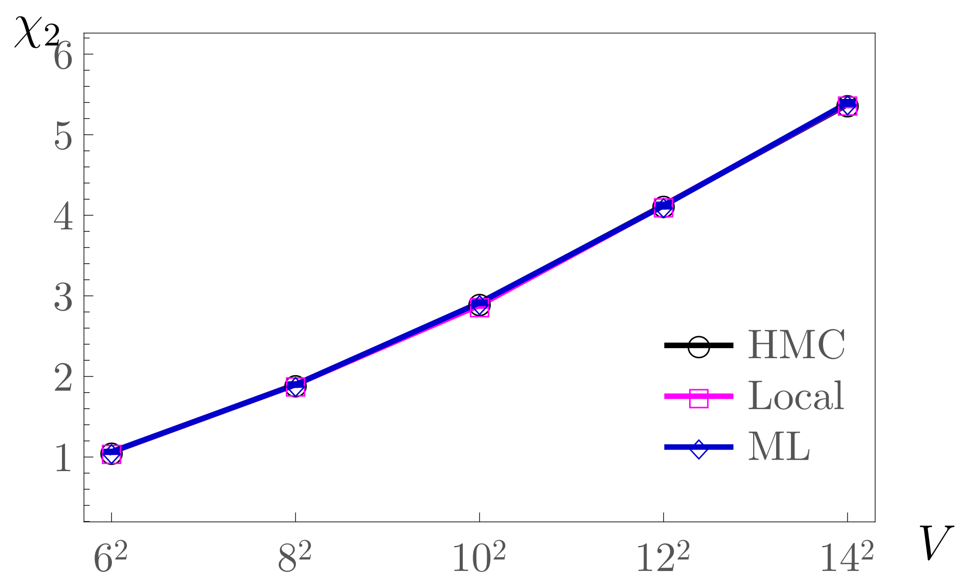


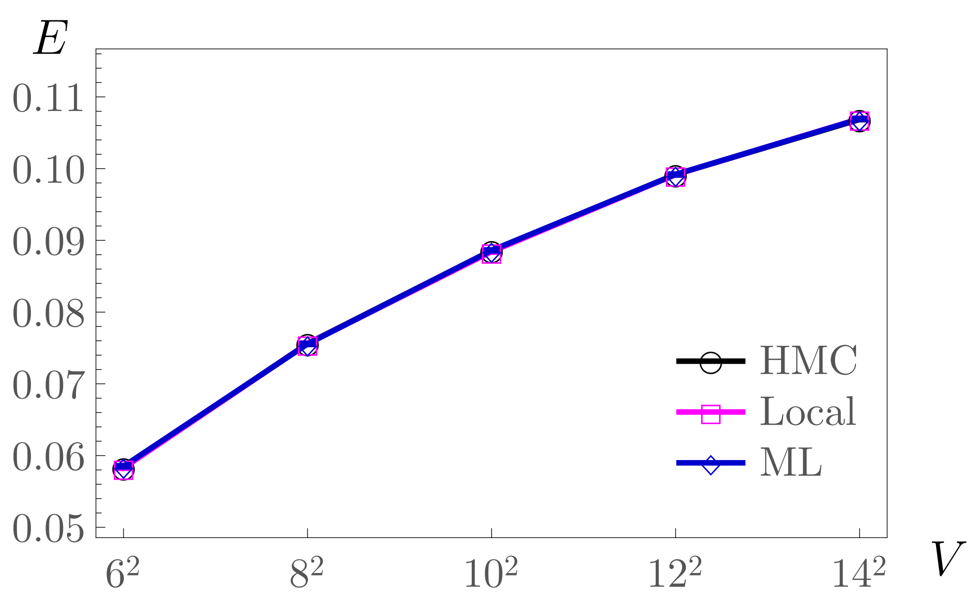
Rejection runs in the Metropolis-Hastings accept/reject step are **comparable to those in Hamiltonian Monte-Carlo** tuned to same acceptance

Application: scalar field theory

First application: scalar lattice field theory

Compare with standard updating algorithms: 'local', 'HMC'





Physical observables match computed on ensembles generated from ML model and from standard methods

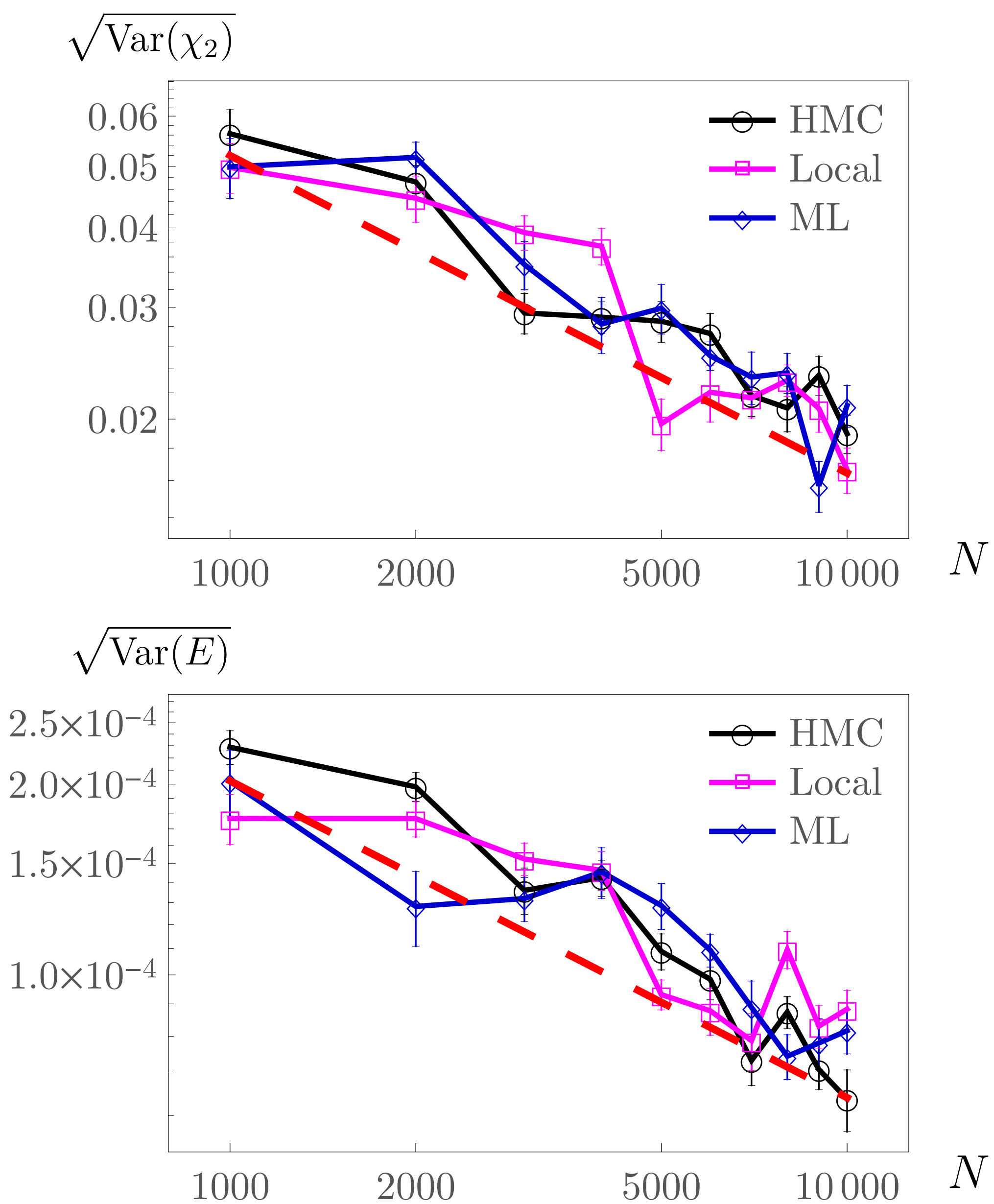
Two-point susceptibility $\chi_2 = \sum_x G_c(x)$

Ising limit energy $E = \frac{1}{d} \sum_{1 \le \mu \le d} G_c(\hat{\mu})$

Application: scalar field theory

First application: scalar lattice field theory

Compare with standard updating algorithms: 'local', 'HMC'



Uncertainties in physical observables follow statistical scaling as the number of samples is increased

red dashed curve:

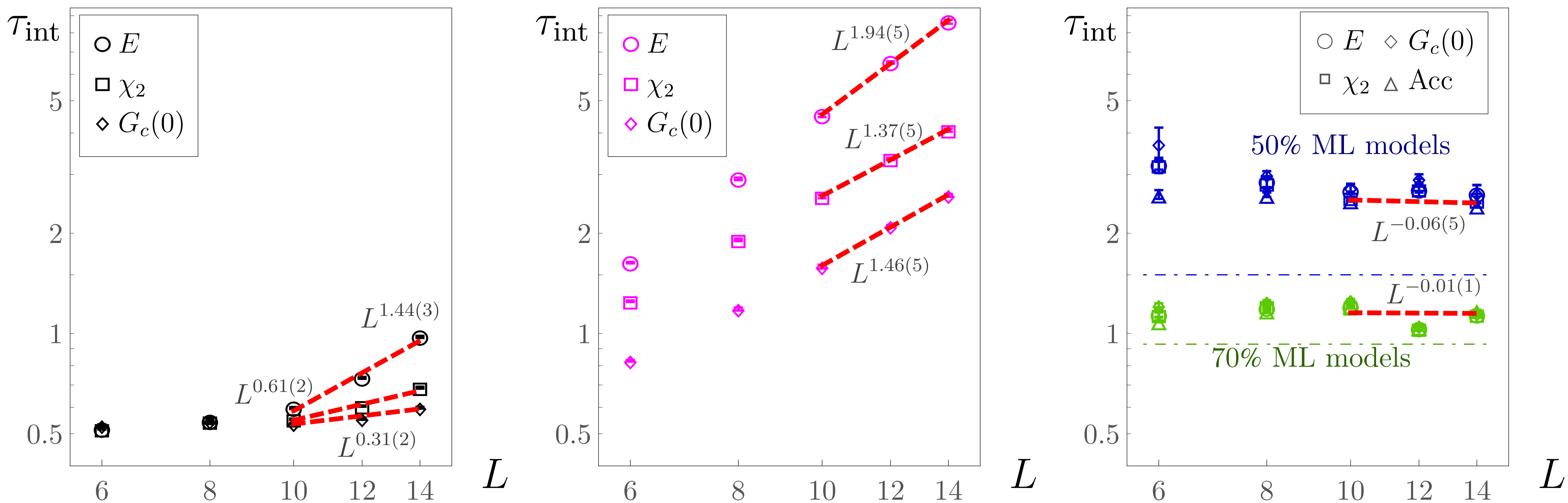
$$\propto 1/\sqrt{N}$$

Application: scalar field theory

First application: scalar lattice field theory

Success: Critical slowing down is eliminated

Cost: Up-front training of the model



(a) HMC ensembles

(b) Local Metropolis ensembles

(c) Flow-based MCMC ensembles

Dynamical critical exponents consistent with zero

Next steps

Target application: lattice quantum chromodynamics for particle and nuclear physics

- 4 dimensions, typical lattice size $48^3 \times 96$
- Gauge theory i.e., fields take matrix values at each site, with physics invariant under symmetry transformations

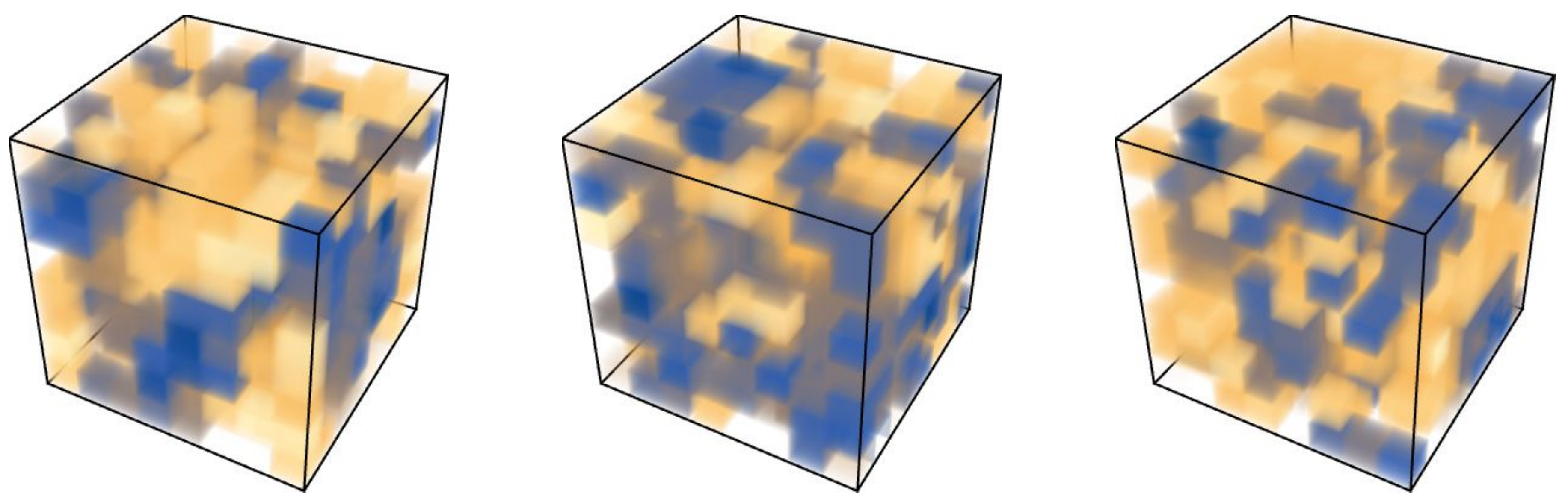
1. Scale number of dimensions
2. Scale number of degrees of freedom
3. Methods for gauge theories

Next steps

1. Scale number of dimensions

Costs scale up, but no theoretical obstacle
Successful tests for scalar field theory

30% acc, no hyperparameter tuning required



Samples generated for ϕ^4 theory with $V=8^3$, $m^2=-6.0$, $\lambda=14.590$
$mL \sim 4$, matching CSD investigation of [Vierhaus, Thesis, doi:10.18452/14138]

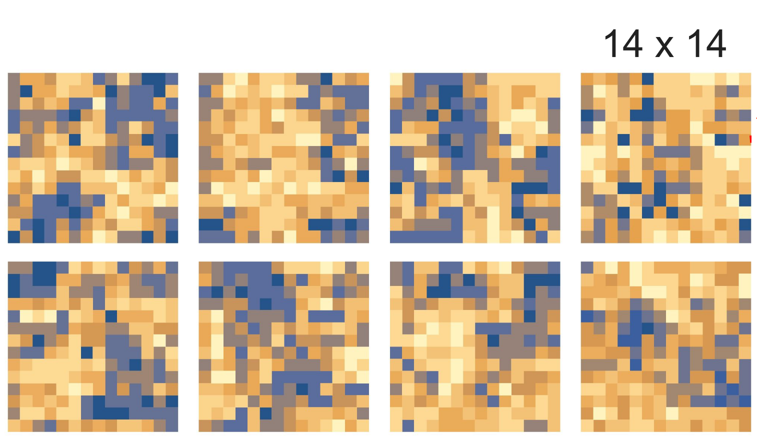
Next steps

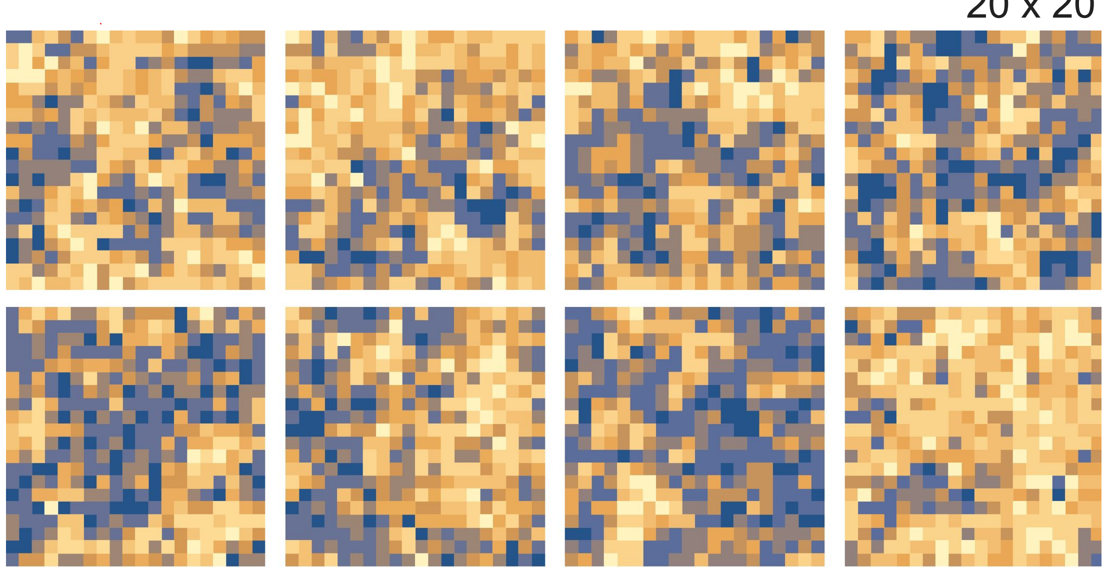
2. Scale number of degrees of freedom

Instead of fully-connected neural networks, consider convolutions and hierarchical structure

- Explicitly preserves translational invariance
- Makes scaling physical volume easy

Successful tests for scalar field theory

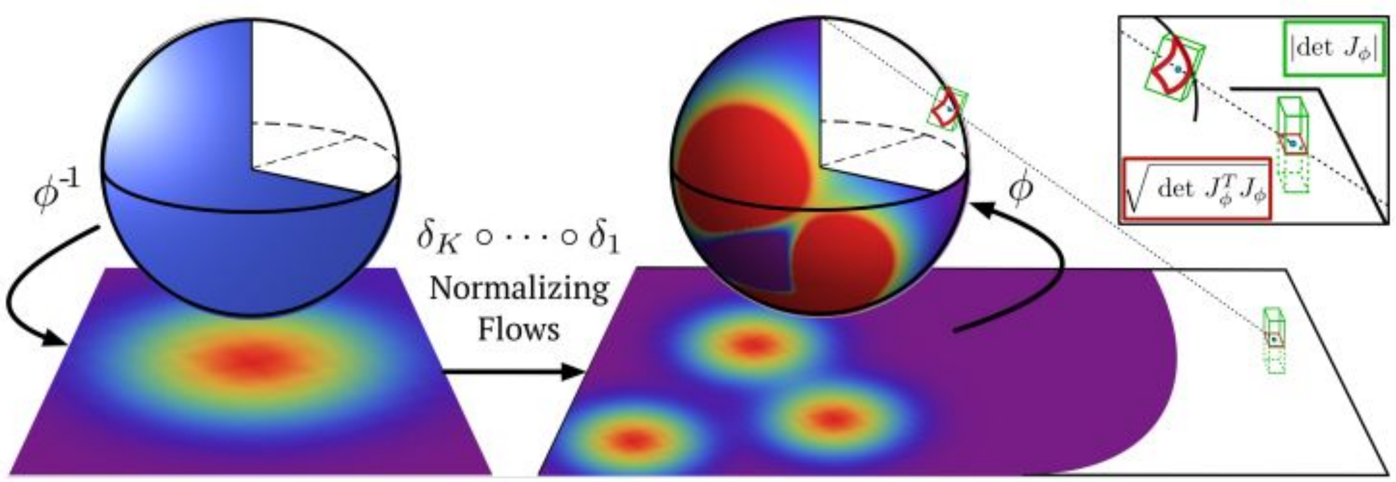




Next steps

3. Methods for gauge theories

- Stereographic projection coupled with standard methods [Gemici, Rezende, Mohammed 1611:02304]

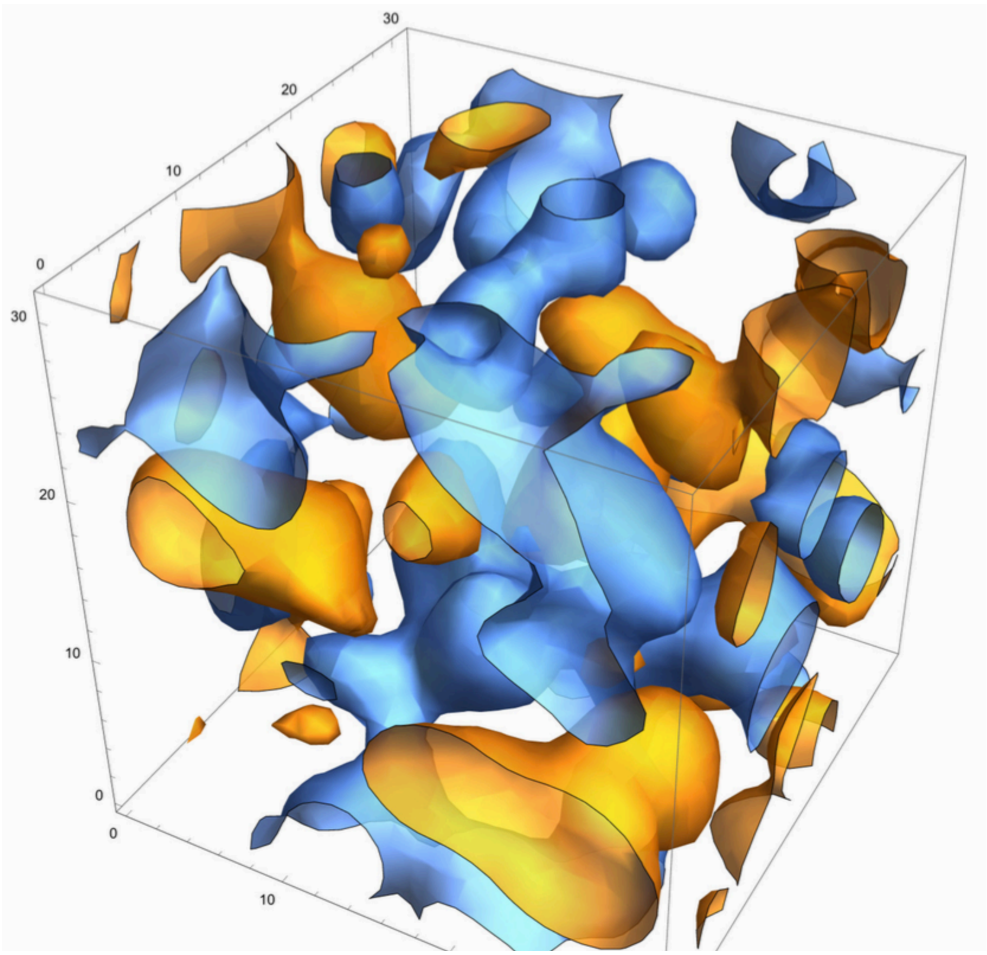


- Other approaches?

Machine learning for LQFT

Accelerate gauge-field generation

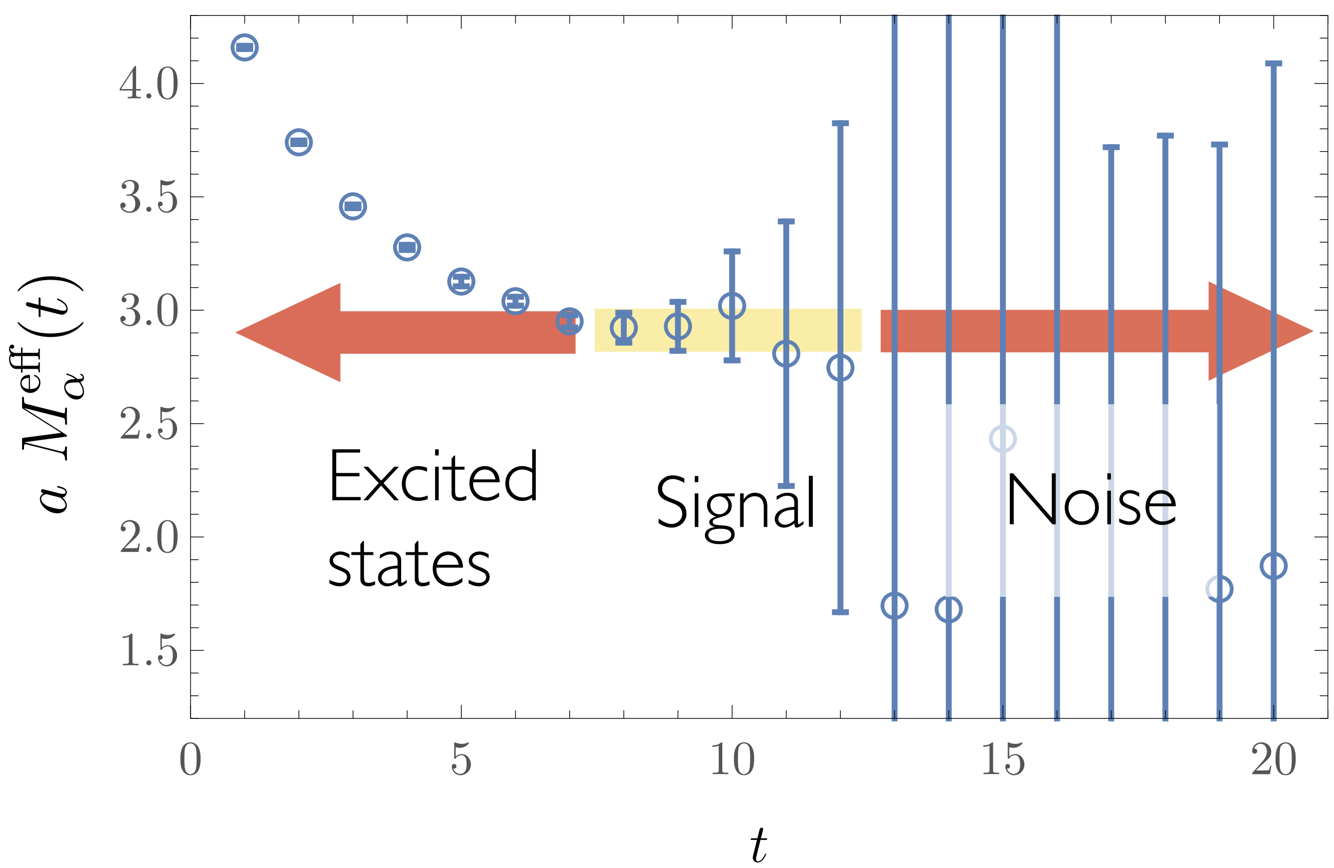
- Multi-scale matching PROOF OF PRINCIPLE ✓
- Generative models to replace expensive HMC PROOF OF PRINCIPLE ✓
- Learn parameters of a complicated pure-gauge action (cheap) to reproduce action with dynamical fermions (expensive)



Machine learning for LQFT

Optimise extraction of physics from gauge fields

- Optimise source operator construction
  - ➡ beat down excited states
- New analysis approaches to maximise signal-to-noise
  - ➡ beat down noise



Accelerated algorithms have huge potential to enable first-principles nuclear physics studies