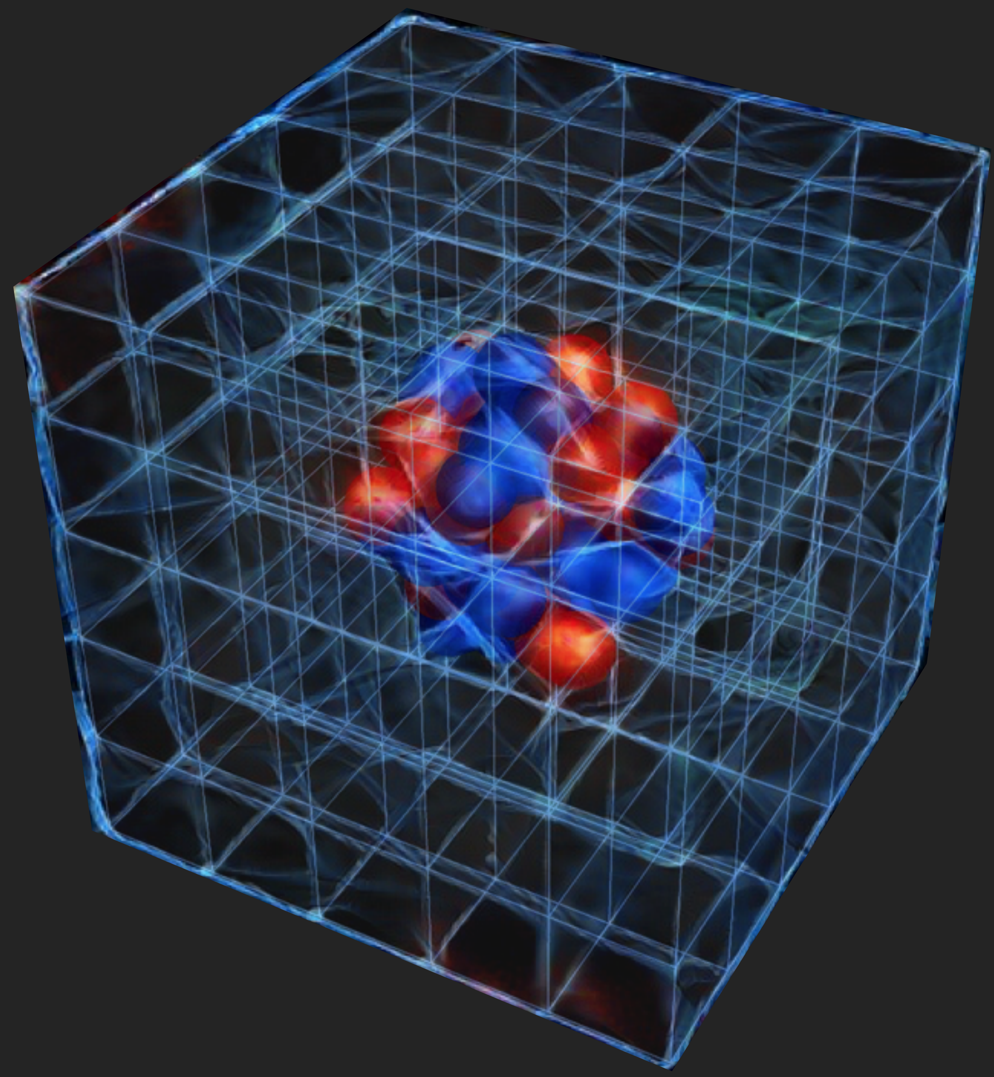
# ML for multi-scale algorithms in lattice field theory



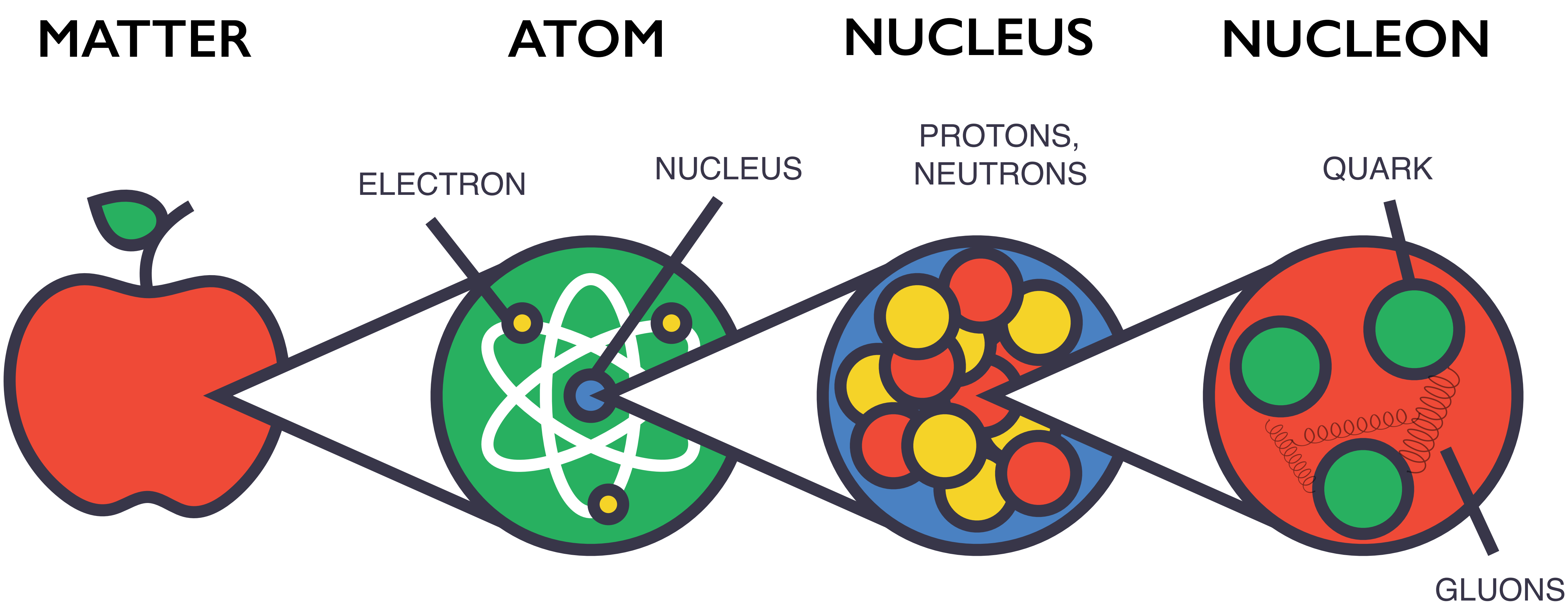
Phiala Shanahan

Massachusetts Institute of Technology

# The structure of matter

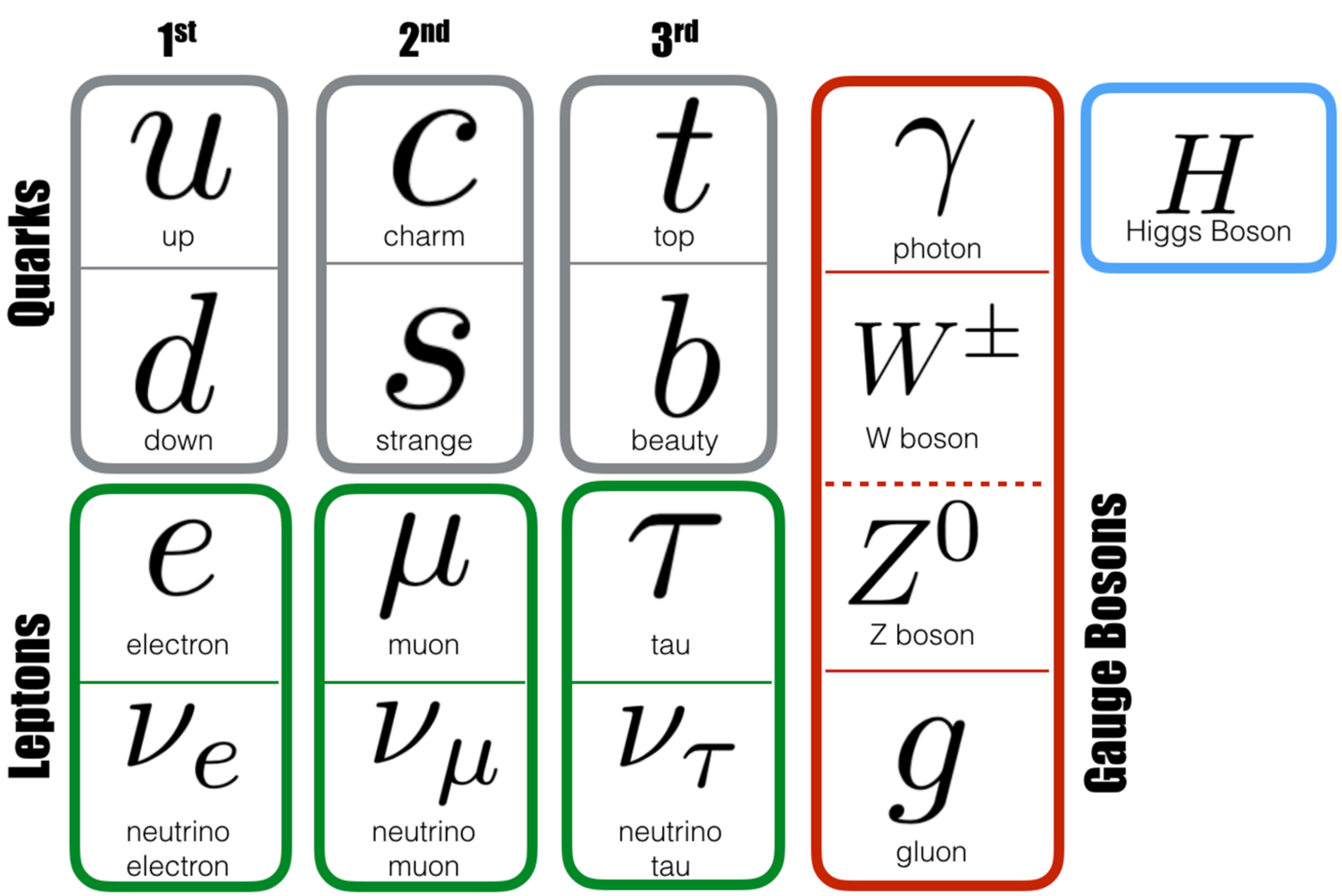
What is everything made of?

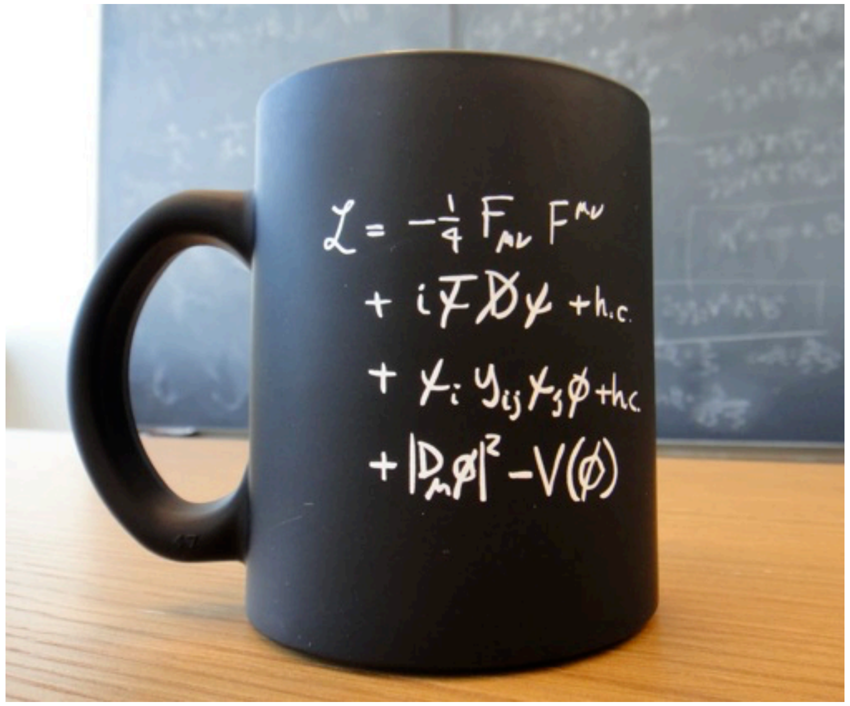
What laws describe the properties of matter?



# The structure of matter

## The Standard Model of nuclear and particle physics



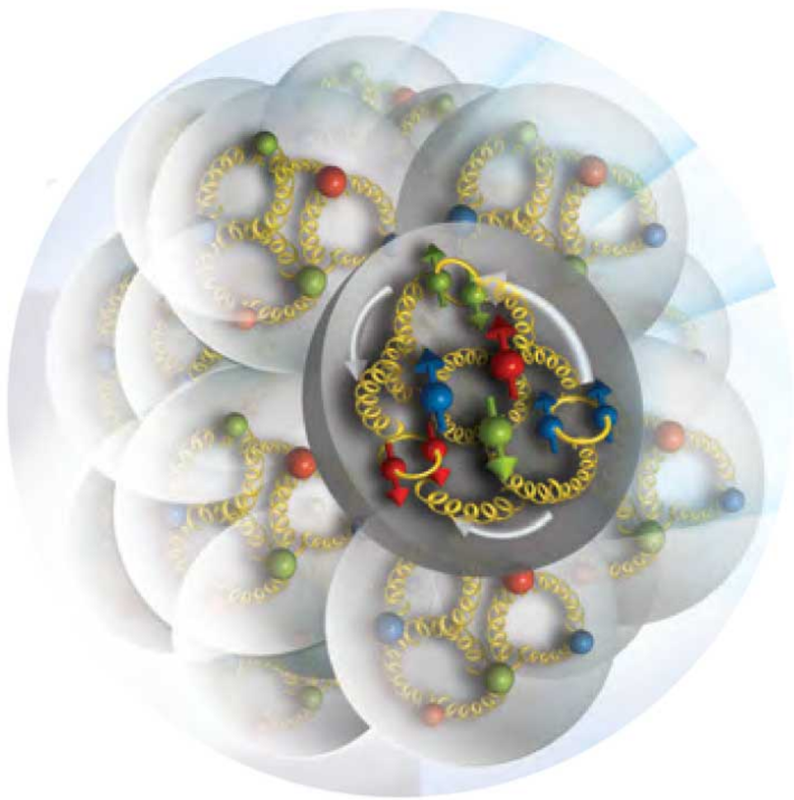



# The structure of matter

**Nuclear structure from the Standard Model**

Emergence of complex structure in nature

Backgrounds and benchmarks for searches for new physics

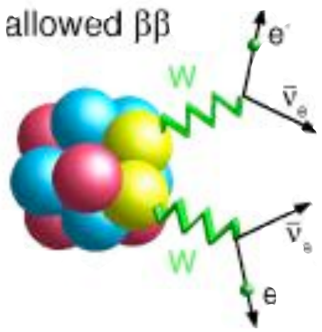


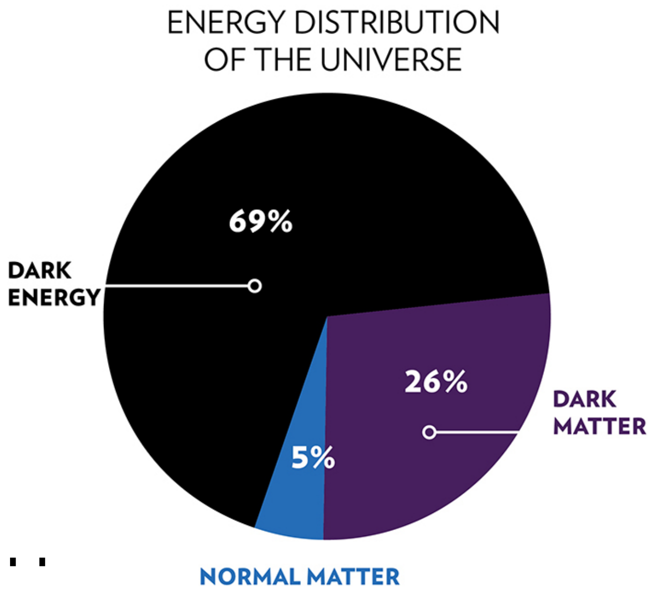
# The search for new physics

**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…



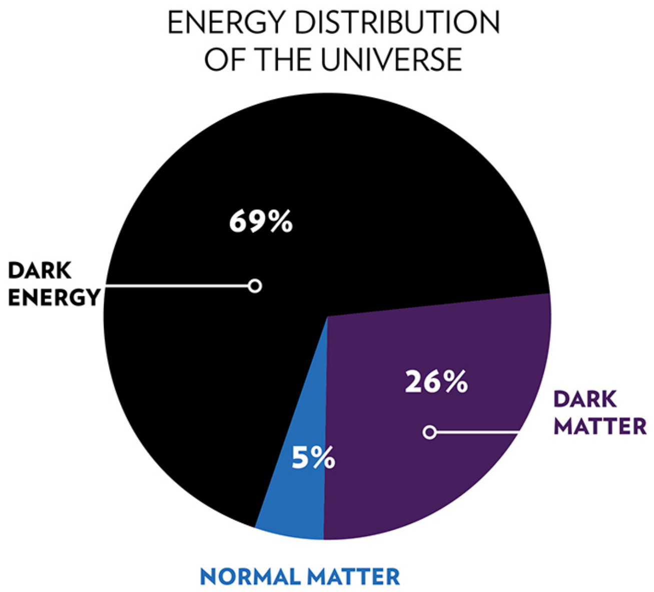
# The search for new physics

**Precise experiments seek new physics at the "Intensity Frontier"**

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

EXPERIMENTS USE NUCLEAR TARGETS

NEED TO UNDERSTAND STANDARD MODEL PHYSICS OF NUCLEI

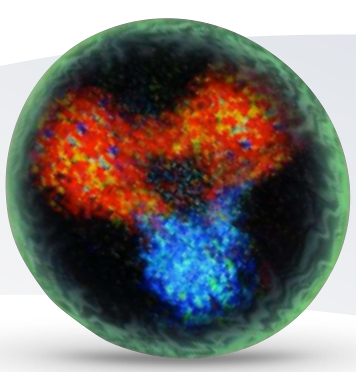


# Strong interactions

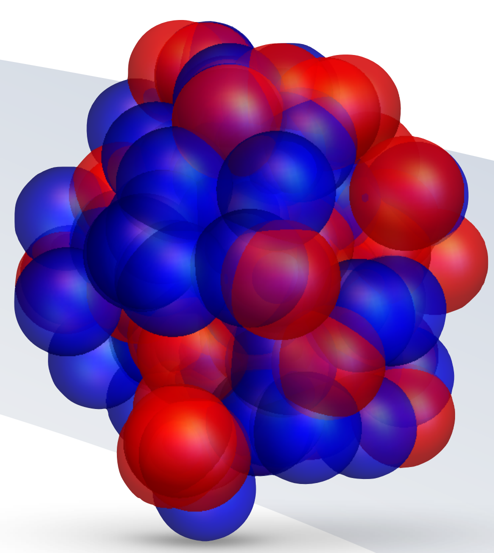
Study nuclear structure from the strong interactions

**Quantum Chromodynamics (QCD)**

Strongest of the four forces in nature

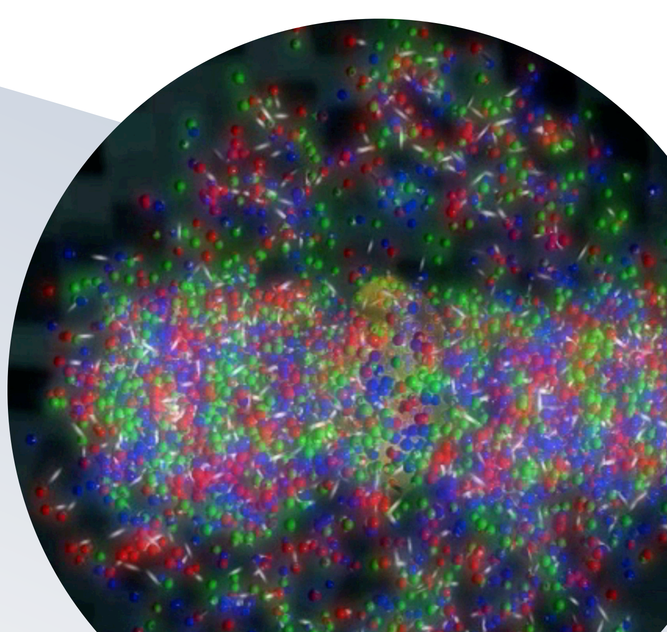


Binds quarks and gluons into protons, neutrons, pions etc.



Binds protons and neutrons into nuclei

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



# Strong interactions

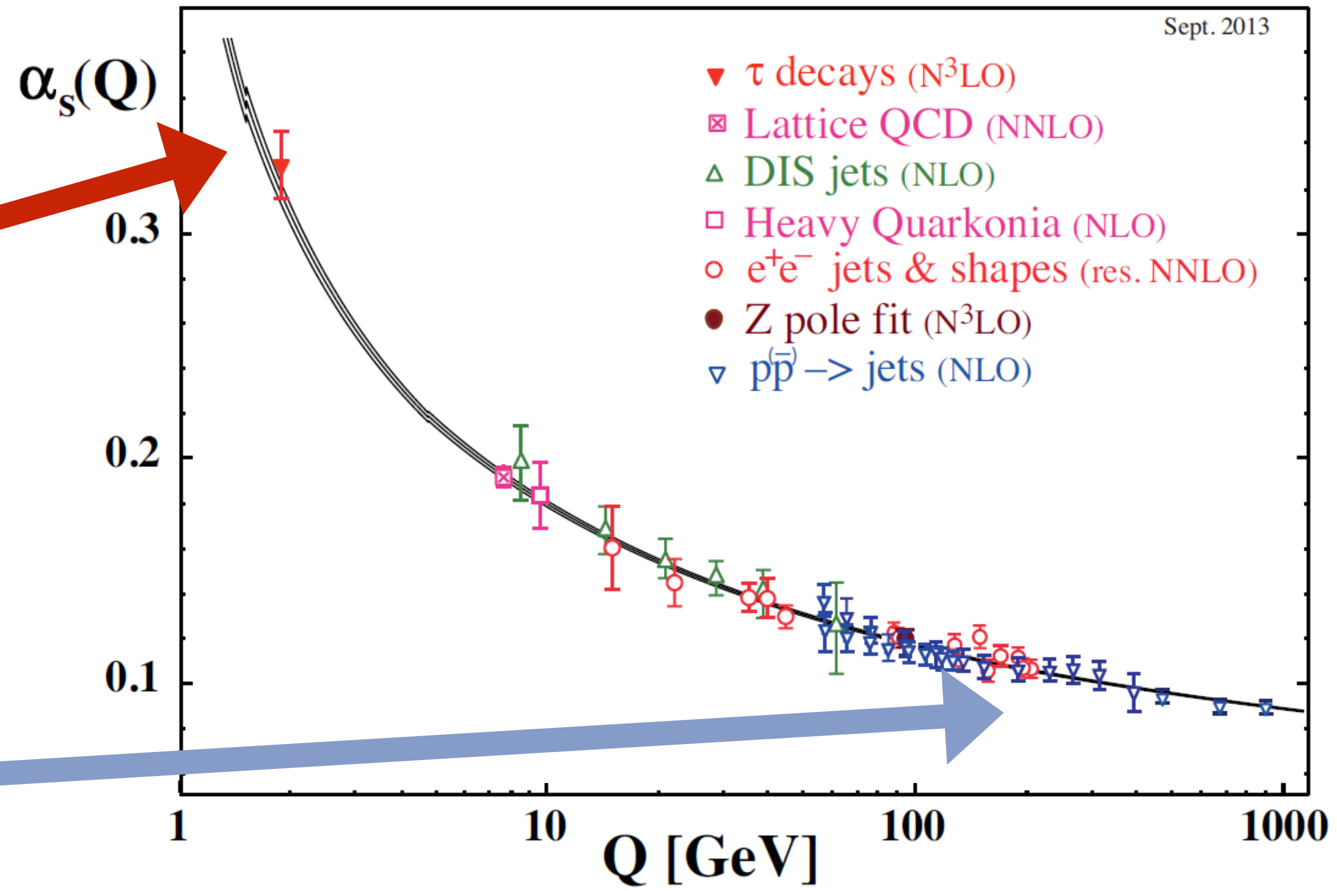
Interaction strength depends on energy
[Gross, Politzer, Wilczek, Nobel 2004]

**Low-energy QCD is non-perturbative**

Perturbation theory at high energies

$$\mathcal{O}_{\text{exact}} = \mathcal{O}_0 + \mathcal{O}_1 \alpha_s + \mathcal{O}_2 \alpha_s^2 + \dots$$

Strong coupling

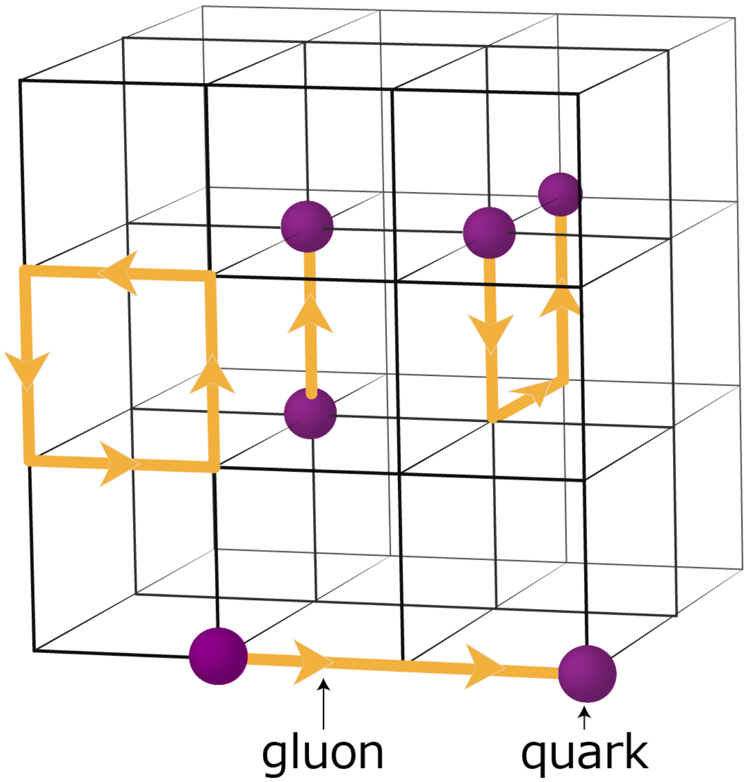



Energy

# Lattice QCD

Numerical first-principles approach to non-perturbative QCD

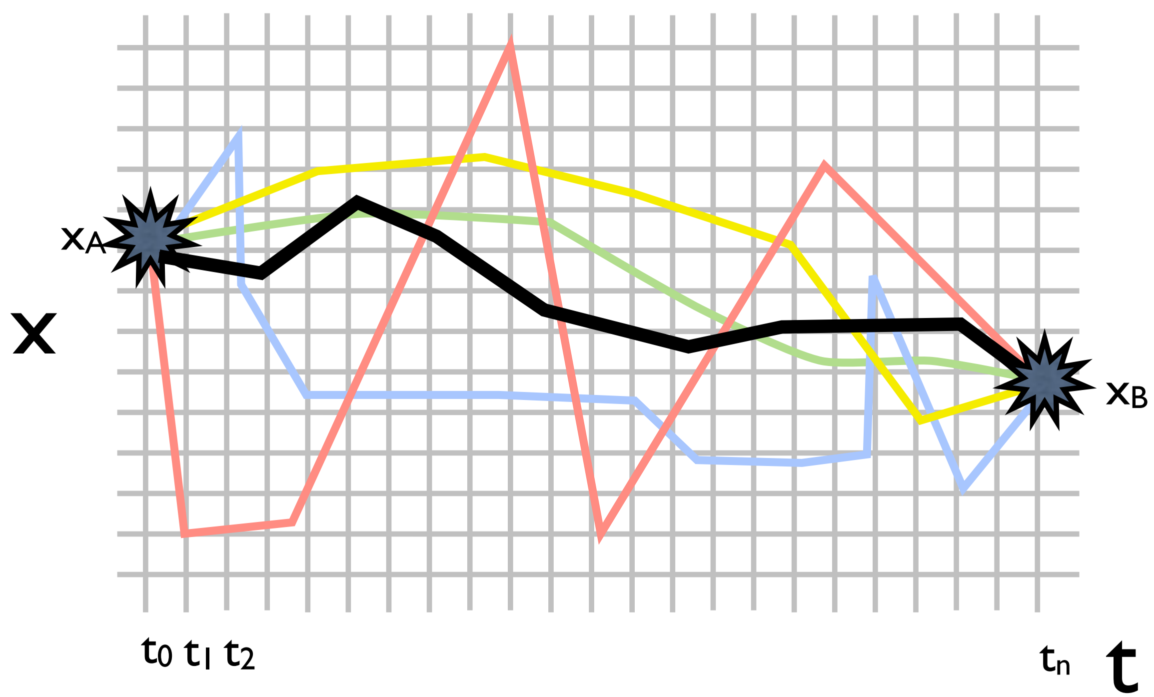
- Discretise QCD onto 4D space-time lattice
- Calculate physical quantities
- Run on supercomputers and dedicated clusters
- Take limit of vanishing discretisation, infinite volume, physical quark masses



# Lattice QCD

Numerical first-principles approach to
non-perturbative QCD

- 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)

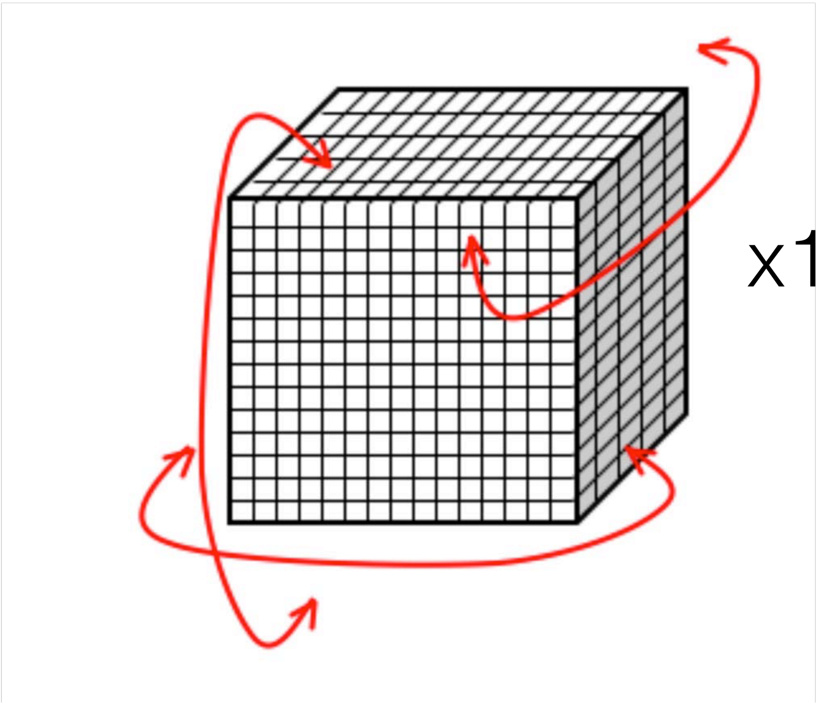



- Evaluate by importance sampling
- Paths near classical action dominate
- Calculate physics on a set (ensemble) of samples of the quark and gluon fields

# Lattice QCD

Numerical first-principles approach to
non-perturbative QCD

- Euclidean space-time
  - Finite lattice spacing $a$
  - Volume $L^3 \times T = 64^3 \times 128$
  - Boundary conditions




Approximate the QCD path integral by **Monte Carlo**

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int \mathcal{D}A \mathcal{D}\overline{\psi} \mathcal{D}\psi \mathcal{O}[A, \overline{\psi}\psi] e^{-S[A,\overline{\psi}\psi]} \rightarrow \langle \mathcal{O} \rangle \simeq \frac{1}{N_{\text{conf}}} \sum_i^{N_{\text{conf}}} \mathcal{O}([U^i])$$

with field configurations $U^i$ distributed according to $e^{-S[U]}$

# Doing lattice QCD

- Correlation decays exponentially with distance in time:

$$C_2(t) = \sum_n Z_n \exp(-E_n t)$$

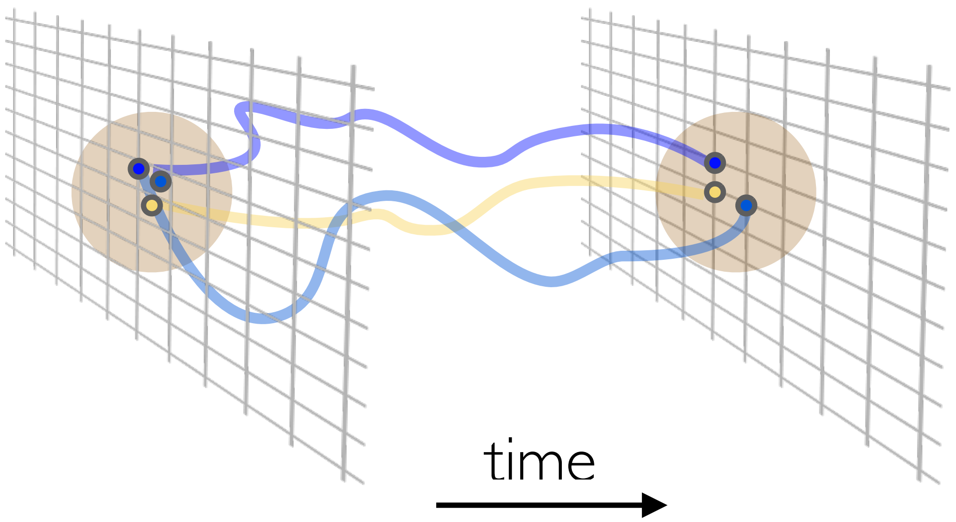
$n$ ← all eigenstates with q#'s of proton

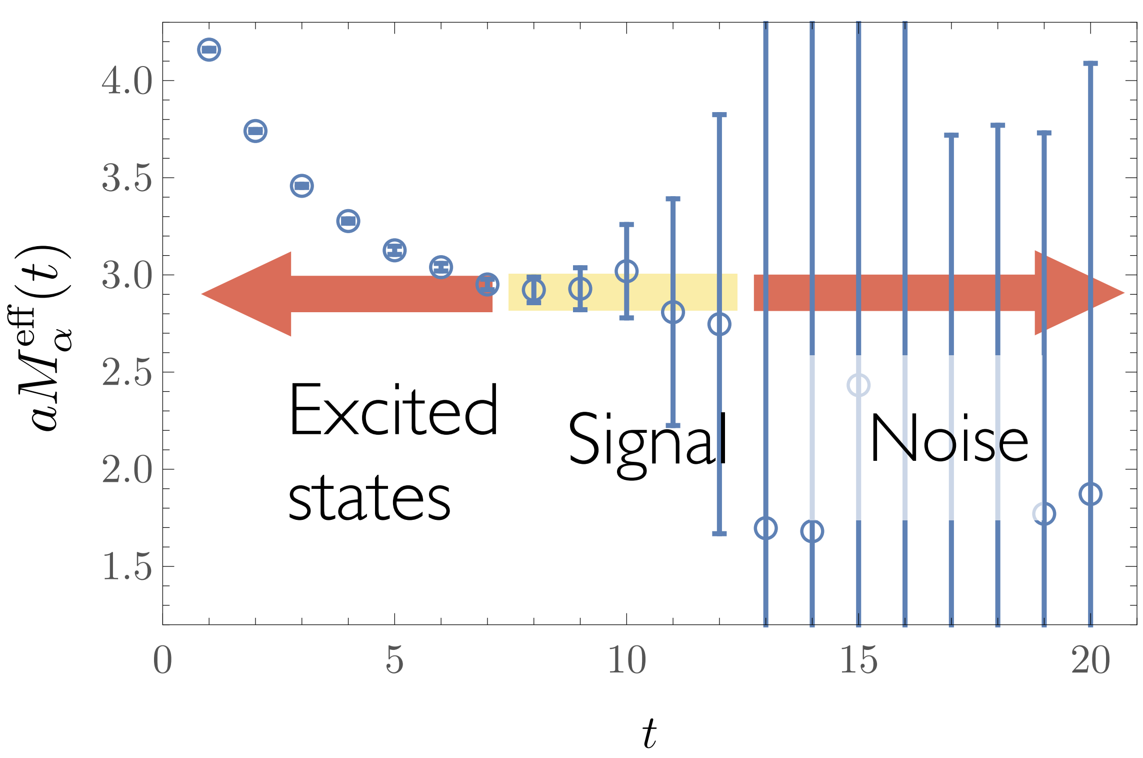
At late times:

$$\to Z_0 \exp(-E_0 t)$$

- Ground state mass revealed through "effective mass plot"

$$M(t) = \ln\left[\frac{C_2(t)}{C_2(t+1)}\right] \xrightarrow{t\to\infty} E_0$$



# Lattice QCD

## Workflow of a lattice QCD calculation

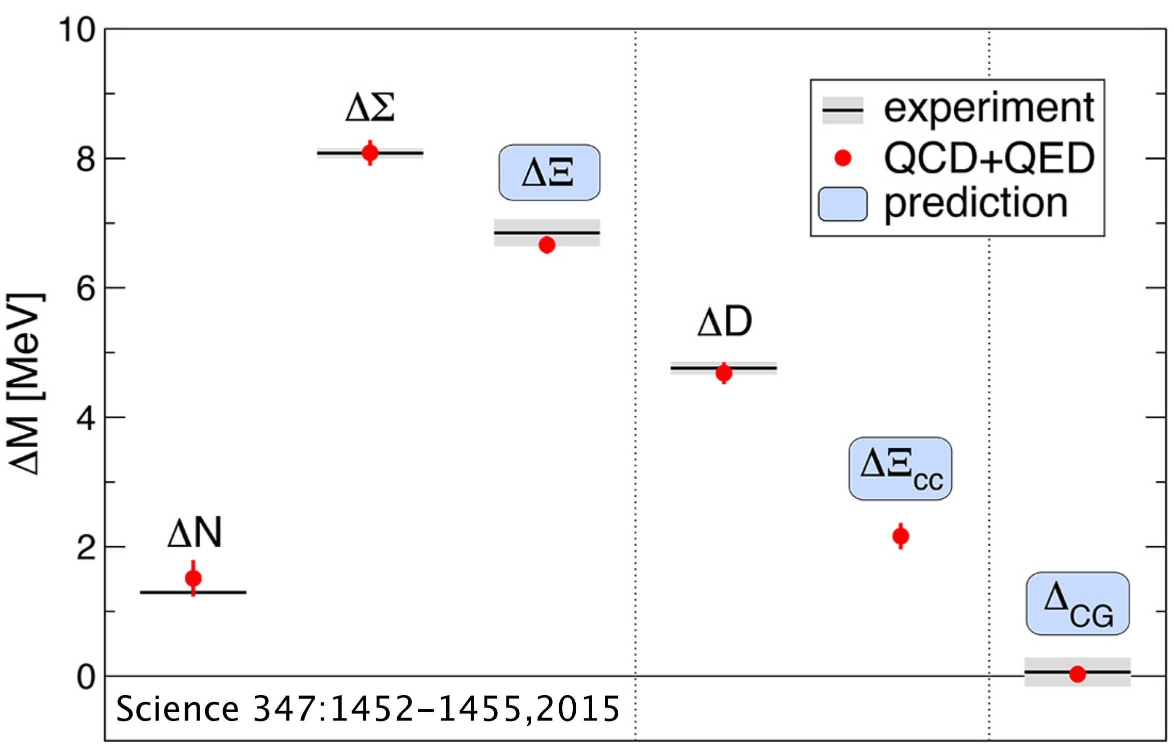
1. Generate configurations via Hybrid Monte Carlo
   - Leadership-class computing
   - ~100K cores or 1000GPUs, 10's of TF-years
   - O(100-1000) configurations, each ~10-100GB


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

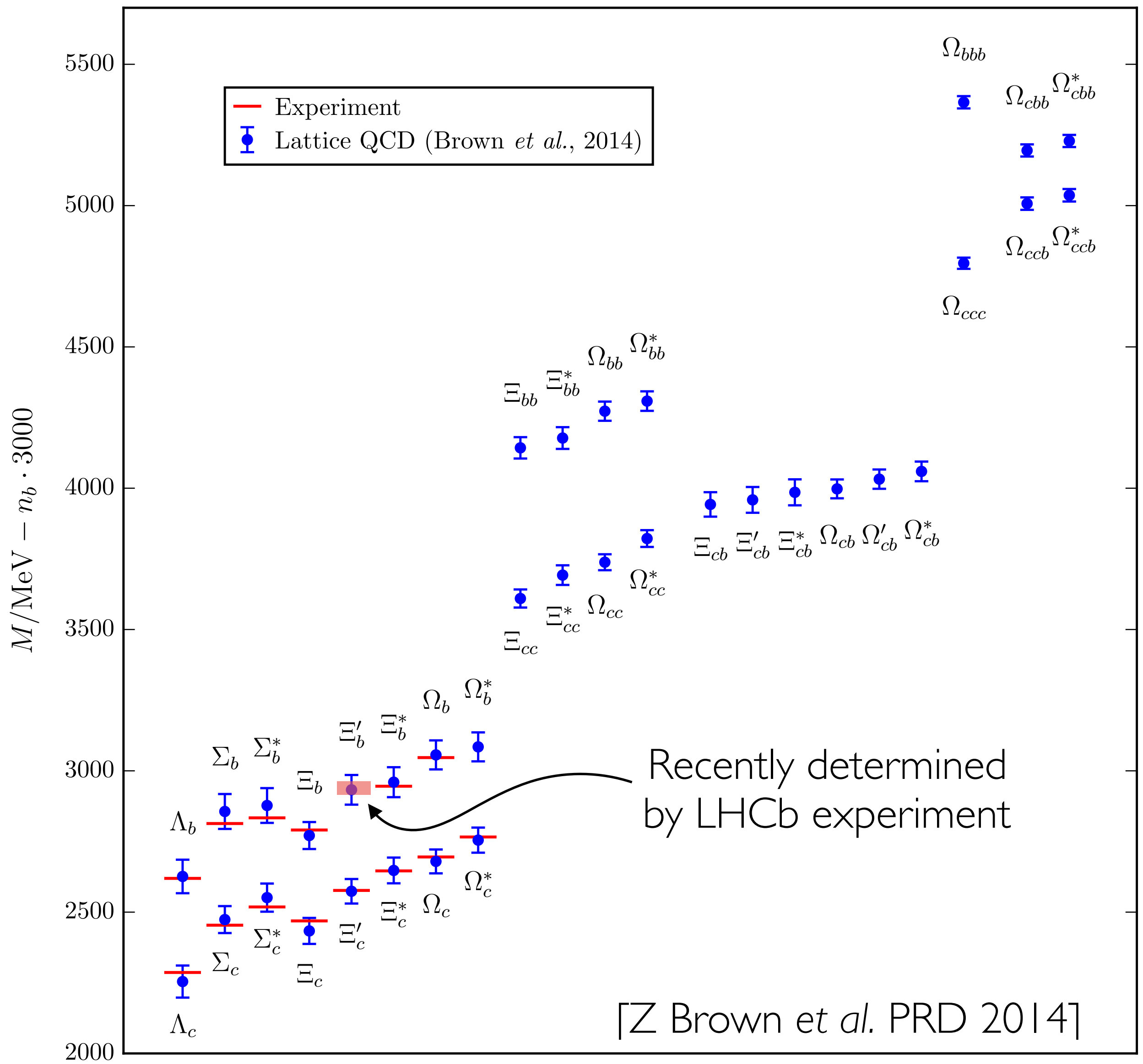
3. Contract into correlation functions
   - ~few GPUs
   - O(100k-1M) copies

# Lattice QCD works

- Ground state hadron spectrum reproduced
- p-n mass splitting reproduced
- …



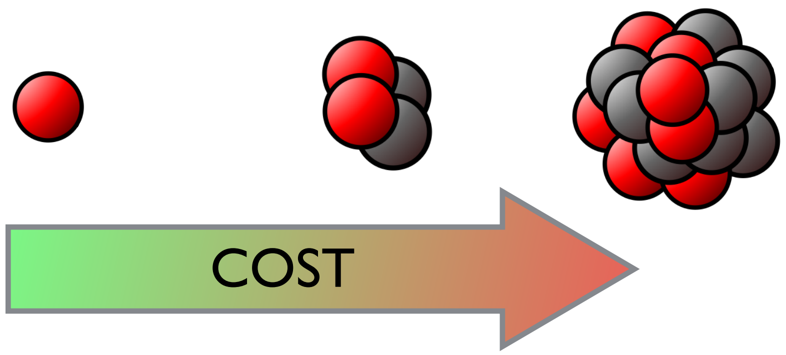

- Predictions for new states with controlled uncertainties



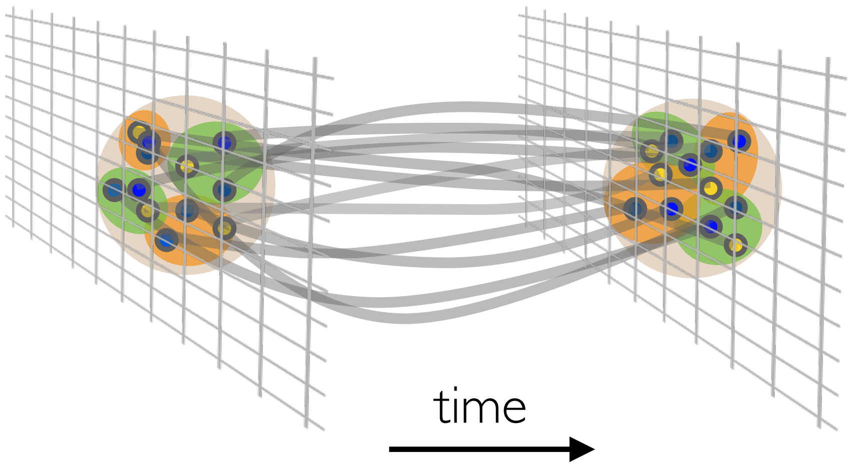
# Nuclear physics from LQCD

## Nuclei on the lattice: HARD

- **Noise:**
  Statistical uncertainty grows exponentially with number of nucleons




- **Complexity:**
  Number of contractions grows factorially




**Calculations possible for A<5** (unphysically heavy quark masses)

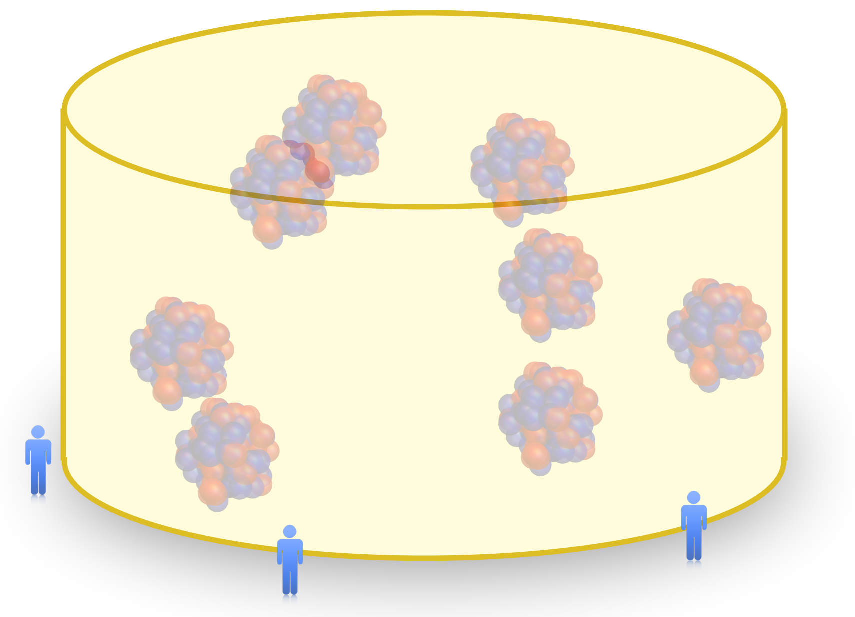
# Motivation: ML for LQCD

## First-principles nuclear physics beyond A=4

**How finely tuned is the emergence of nuclear structure in nature?**

**Interpretation of intensity-frontier experiments**

- Scalar matrix elements in A=131
  XENON1T dark matter direct detection search
- Axial form factors of Argon A=40
  DUNE long-baseline neutrino expt.
- Double-beta decay rates of Calcium A=48



Exponentially harder problems → Need exponentially improved algorithms

# Machine learning for LQCD

## APPROACH

Machine learning as ancillary tool for lattice QCD

- Accelerate gauge-field generation
- Optimise extraction of physics from gauge field ensemble

} Will need to accelerate all stages of lattice QCD workflow to achieve physics goals

- **ONLY apply where quantum field theory can be rigorously preserved**

# Lattice QCD

## Workflow of a lattice QCD calculation


1. Generate 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 gauge field in size, many per config

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

# Lattice QCD

## Workflow of a lattice QCD calculation

1. Generate 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 gauge field in size, many per config

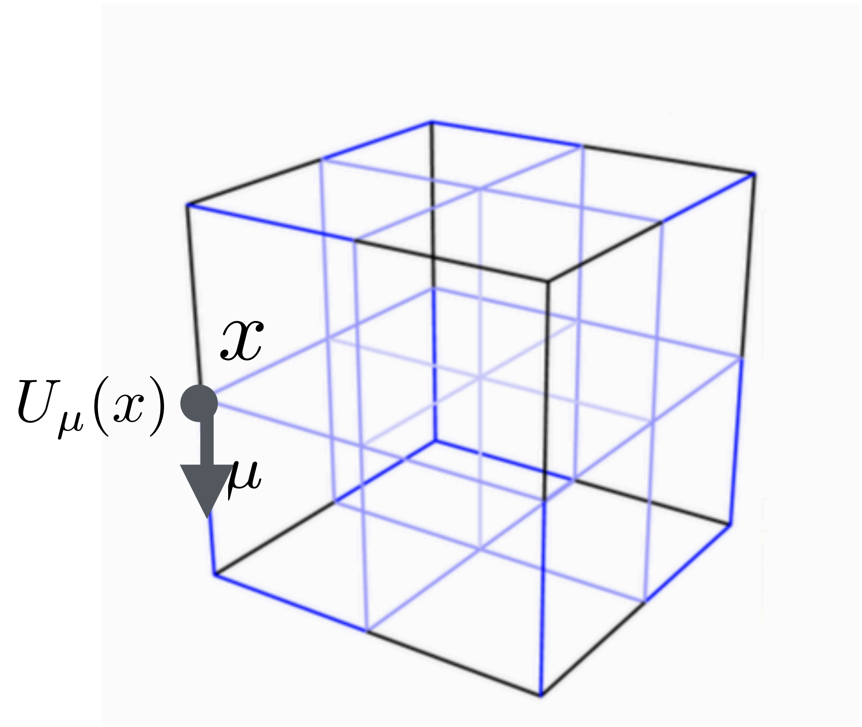
3. Contract into correlation functions
   - ~few GPUs
   - O(100k-1M) copies

# Lattice QCD

## Workflow of a lattice QCD calculation

1 Generate configurations via Hybrid Monte Carlo

- Gauge field configurations represented by ~$10^7$ links $U_\mu(x)$ encoded as SU(3) matrices (3x3 complex matrix $M$ with $det[M] = 1$, $M^{-1} = M^\dagger$) i.e., ~$10^9$ double precision numbers




- Configurations sample probability distribution corresponding to LQCD action (function that defines the quark and gluon dynamics)

  ➡ Weighted averages over configurations determine physical observables of interest

- Calculations use ~$10^3$ 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

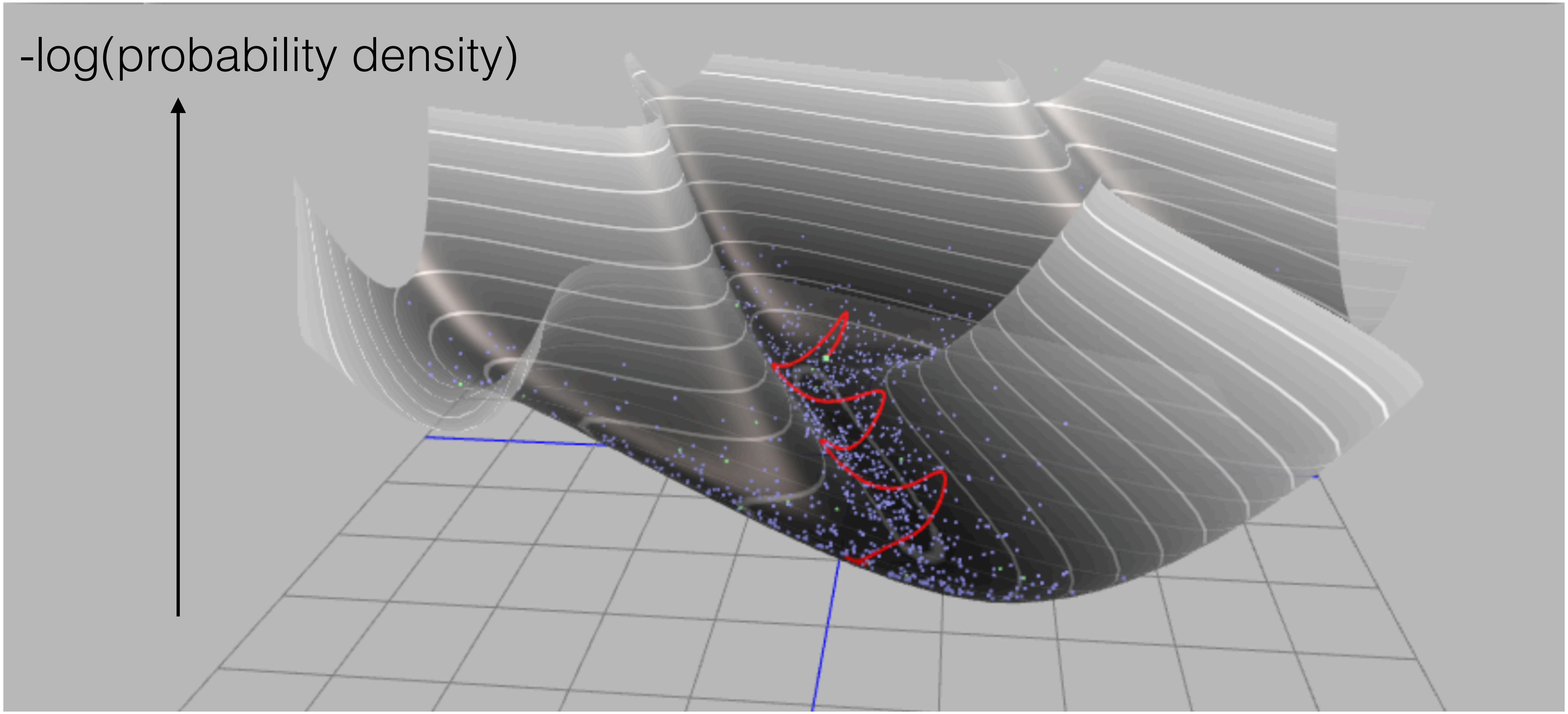
**BUT**

- Short trajectories for high acceptance

# Generate QCD gauge fields

**QCD gauge field configurations sampled via**

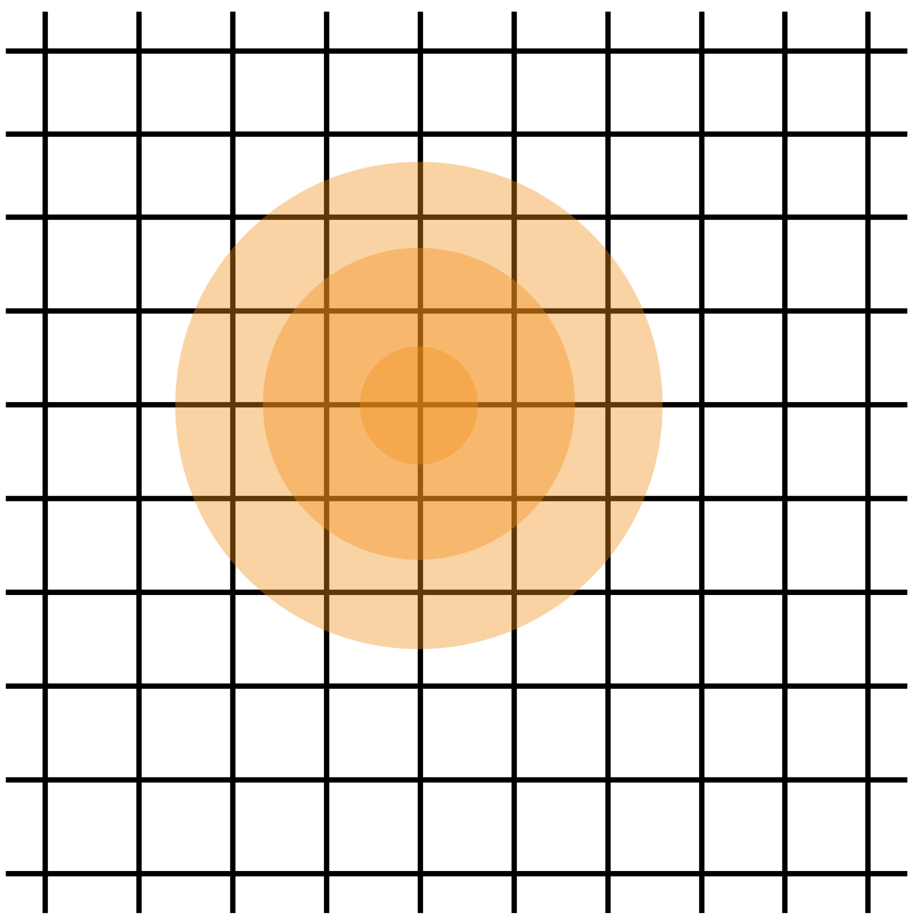
Hamiltonian dynamics + Markov Chain Monte Carlo



# Generate QCD gauge fields

**QCD gauge field configurations sampled via**

Hamiltonian dynamics + Markov Chain Monte Carlo



**Updates diffusive**

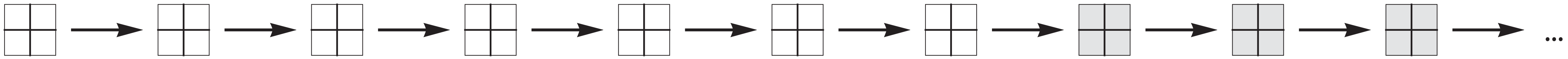
Lattice spacing → 0

Number of updates to change fixed physical length scale → ∞

**“Critical slowing-down”**
**of generation of uncorrelated samples**

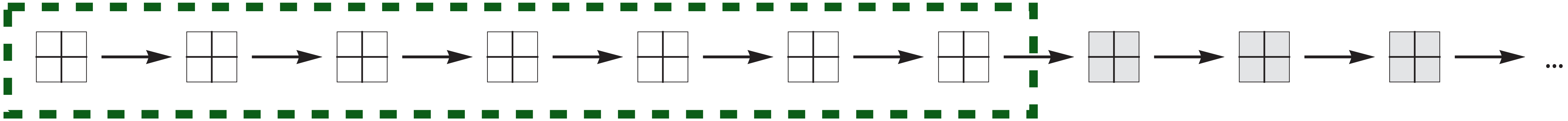
# Multi-scale HMC updates

Usual HMC algorithm



# Multi-scale HMC updates

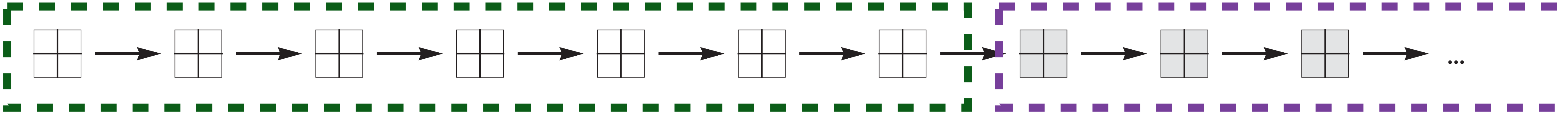
Usual HMC algorithm



Long thermalisation time

# Multi-scale HMC updates

Usual HMC algorithm

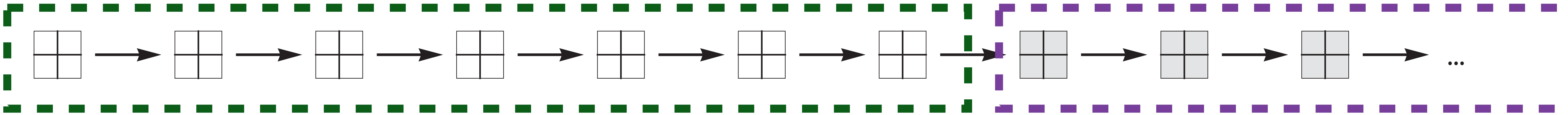


Long thermalisation time

Many steps to decorrelate (sample every $N^{th}$) decorrelation time grows exponentially with inverse lattice spacing

# Multi-scale HMC updates

Usual HMC algorithm



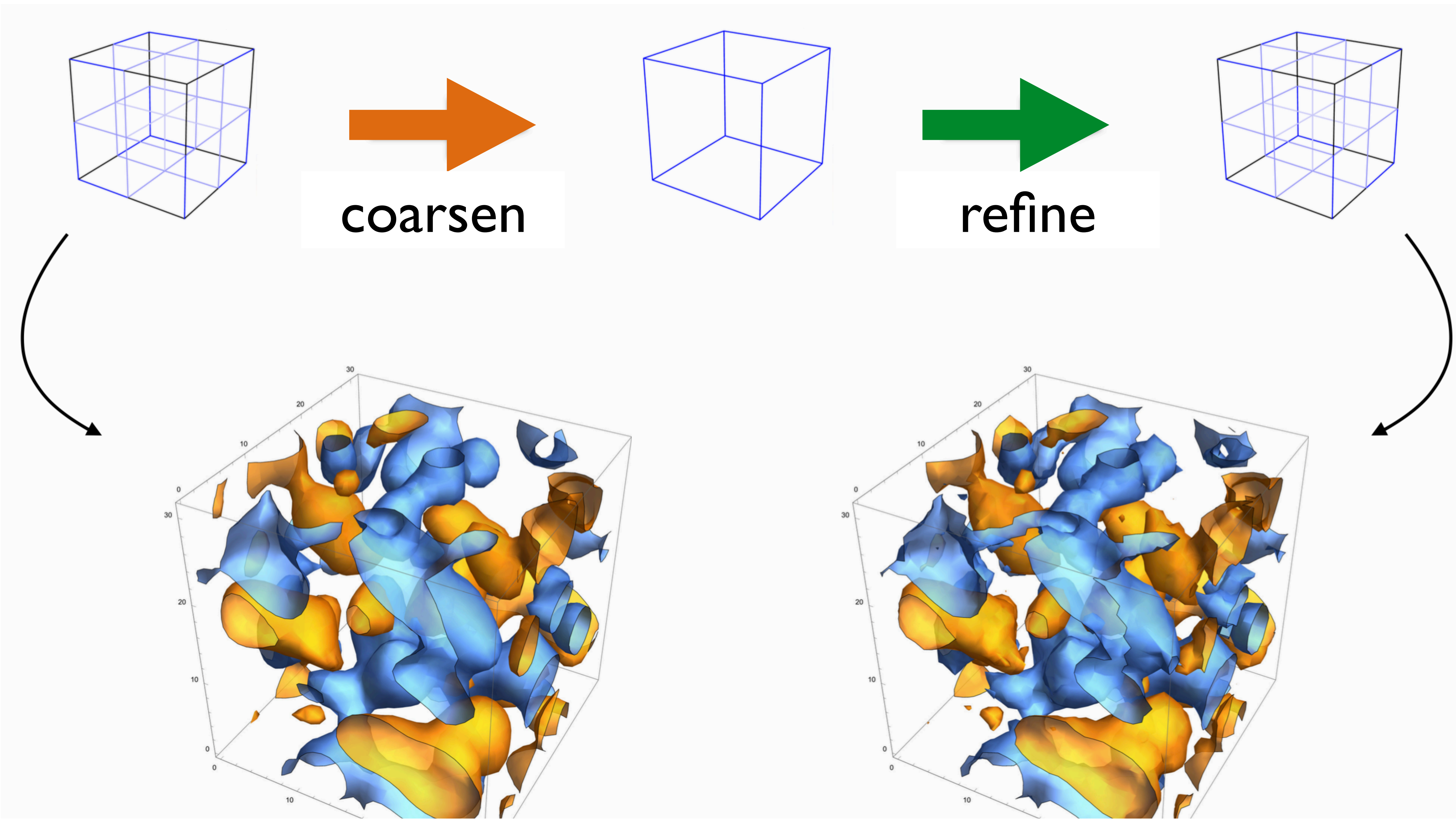

Long thermalisation time

Many steps to decorrelate (sample every $N^{th}$) decorrelation time grows exponentially with inverse lattice spacing

Accelerate gauge field generation with new algorithms?

# Multi-scale HMC updates

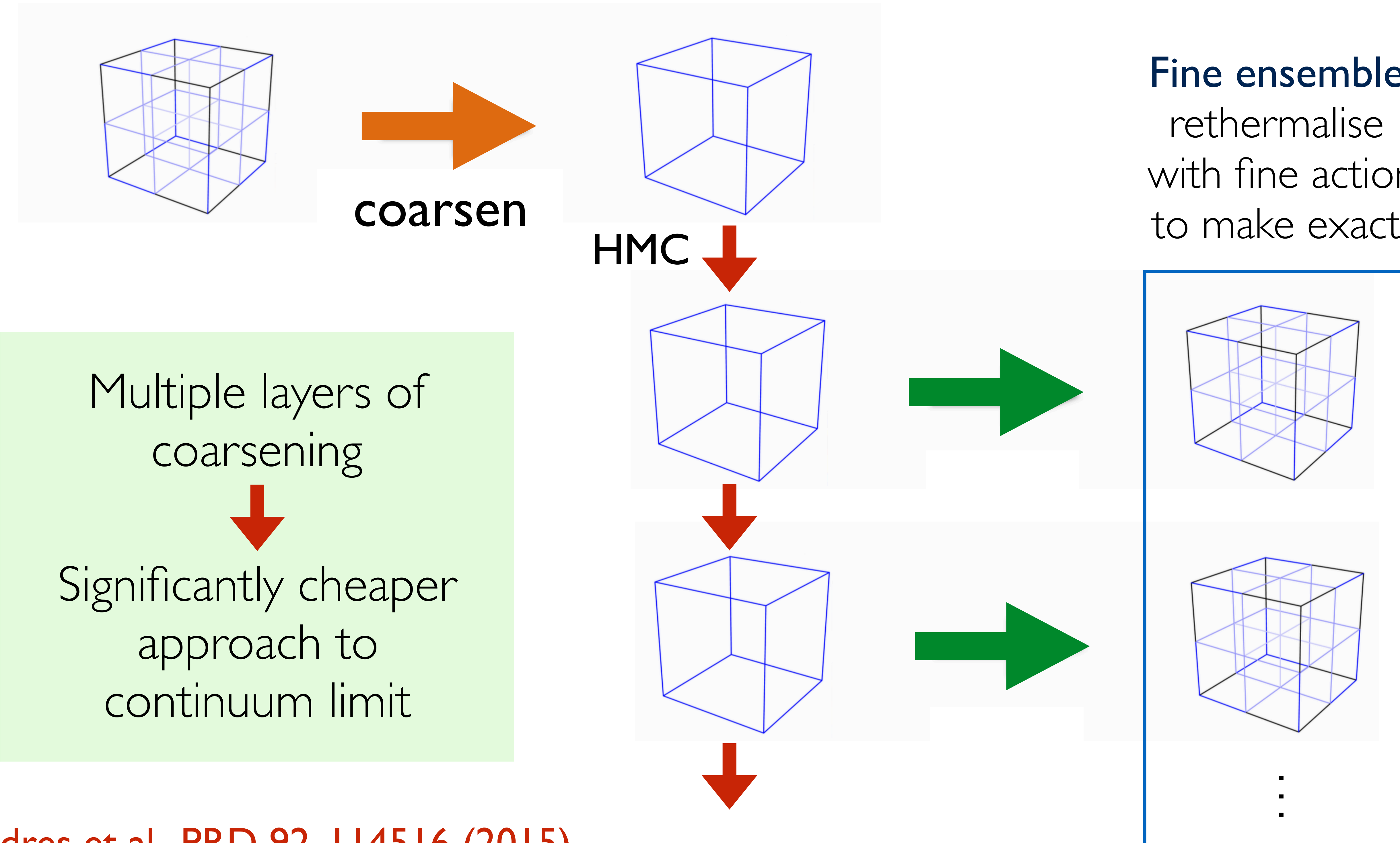
Given coarsening and refinement procedures…




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

# Multi-scale HMC updates

## Perform HMC updates at coarse level




Multiple layers of coarsening

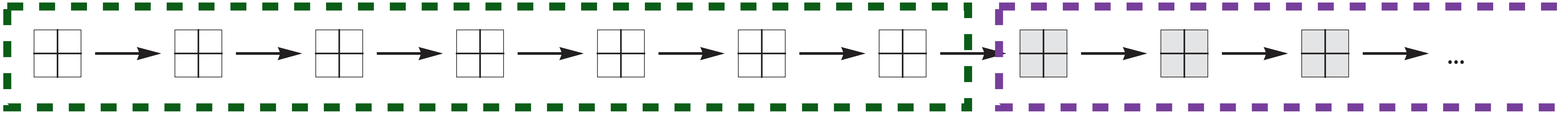
↓

Significantly cheaper approach to continuum limit

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

# Multi-scale HMC updates

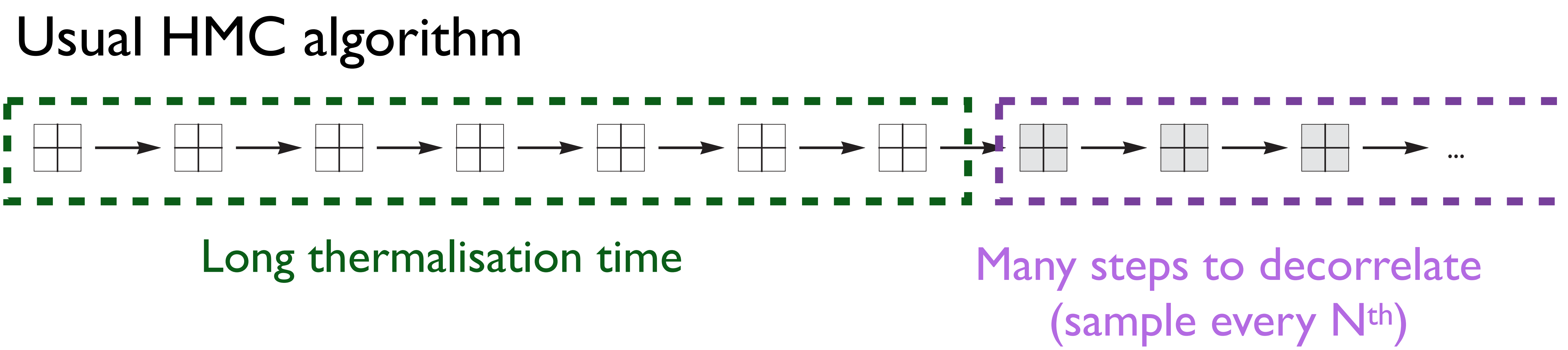
Usual HMC algorithm

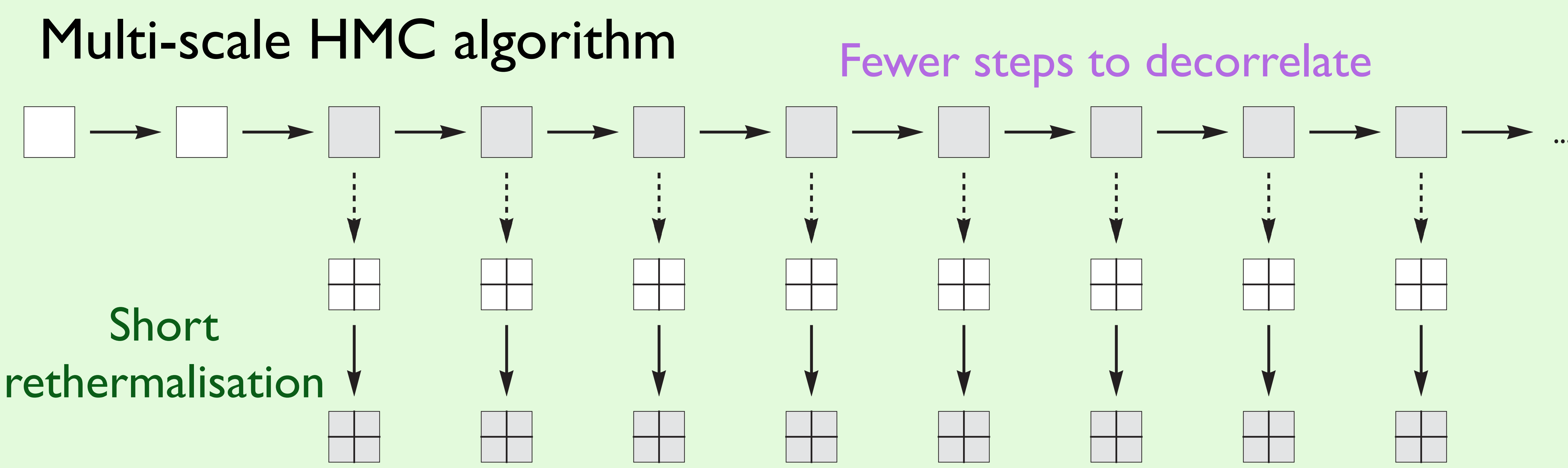



Long thermalisation time

Many steps to decorrelate
(sample every $N^{th}$)
decorrelation time grows exponentially with inverse lattice spacing

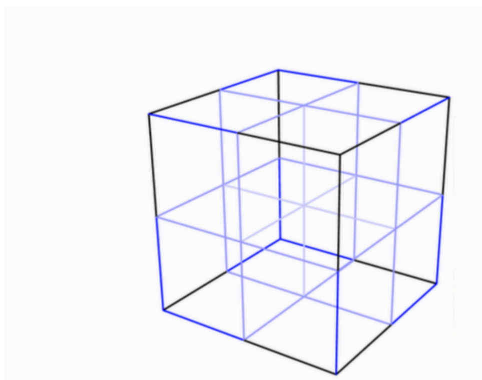
# Multi-scale HMC updates



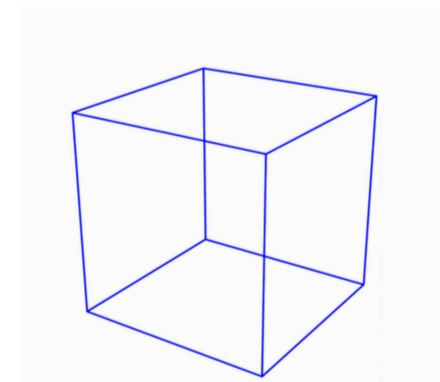



# Multi-scale HMC updates

## 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**

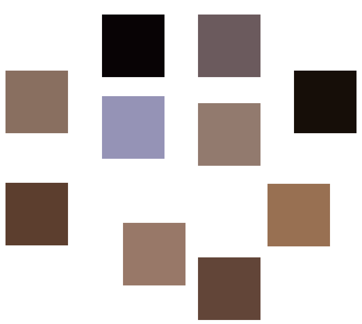
Solve regression problem directly: "Given a coarse ensemble, what parameters generated it?"

# Parameter matching via NN

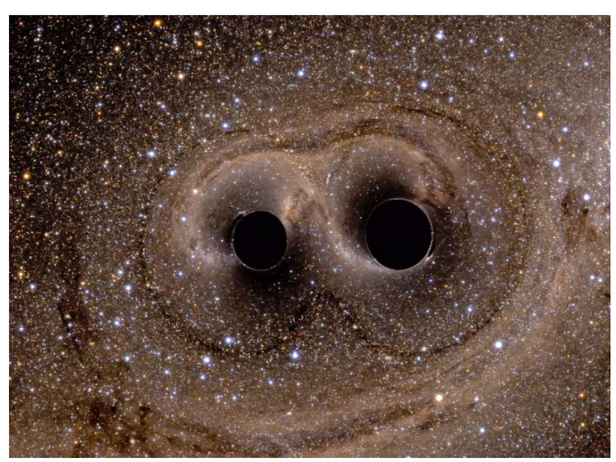
Parameter regression is a problem suited to neural networks

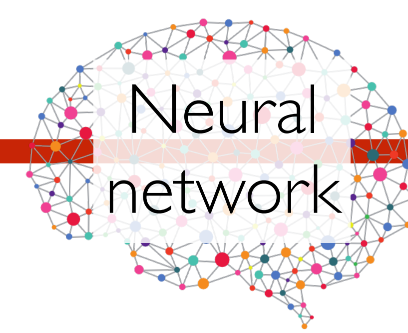
## Image recognition

Pixel



Image






Label

“Colliding black holes”

# Parameter matching via NN

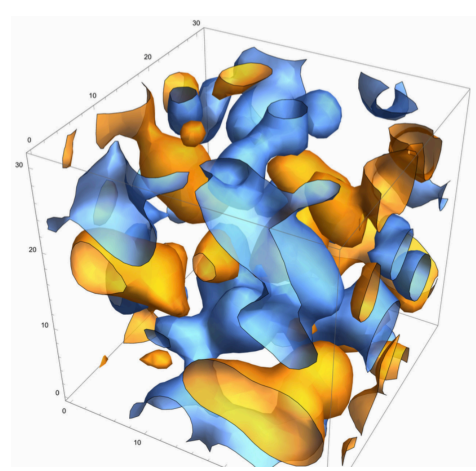
Parameter regression is a problem suited to neural networks

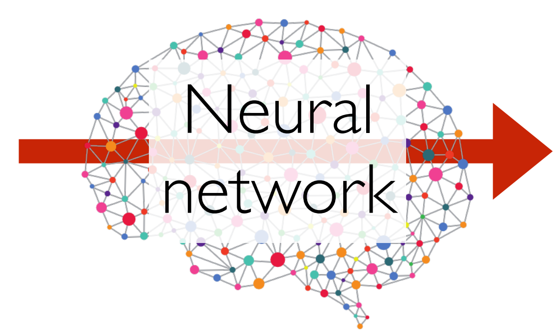
## Parameter identification

Element of a colour matrix at one discrete space-time point

0 6 7 3 5 2 8 4 6 1

Ensemble of lattice QCD gauge field configurations

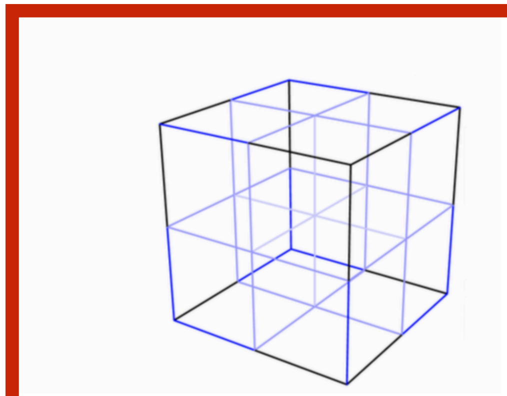




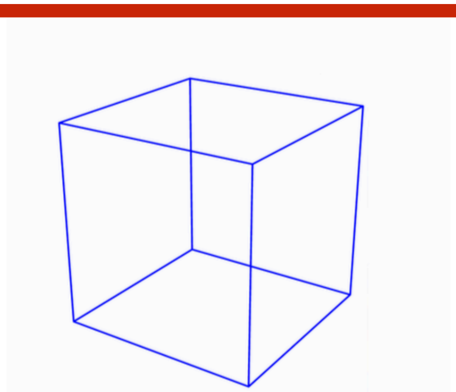

Label

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

**CIFAR benchmark image set for machine learning**

- $32 \times 32$ pixels $\times$ 3 cols $\simeq 3000$ numbers
- 60000 samples
- Each image has meaning
- Local structures are important
- Translation-invariance within frame

**Ensemble of lattice QCD gauge fields**

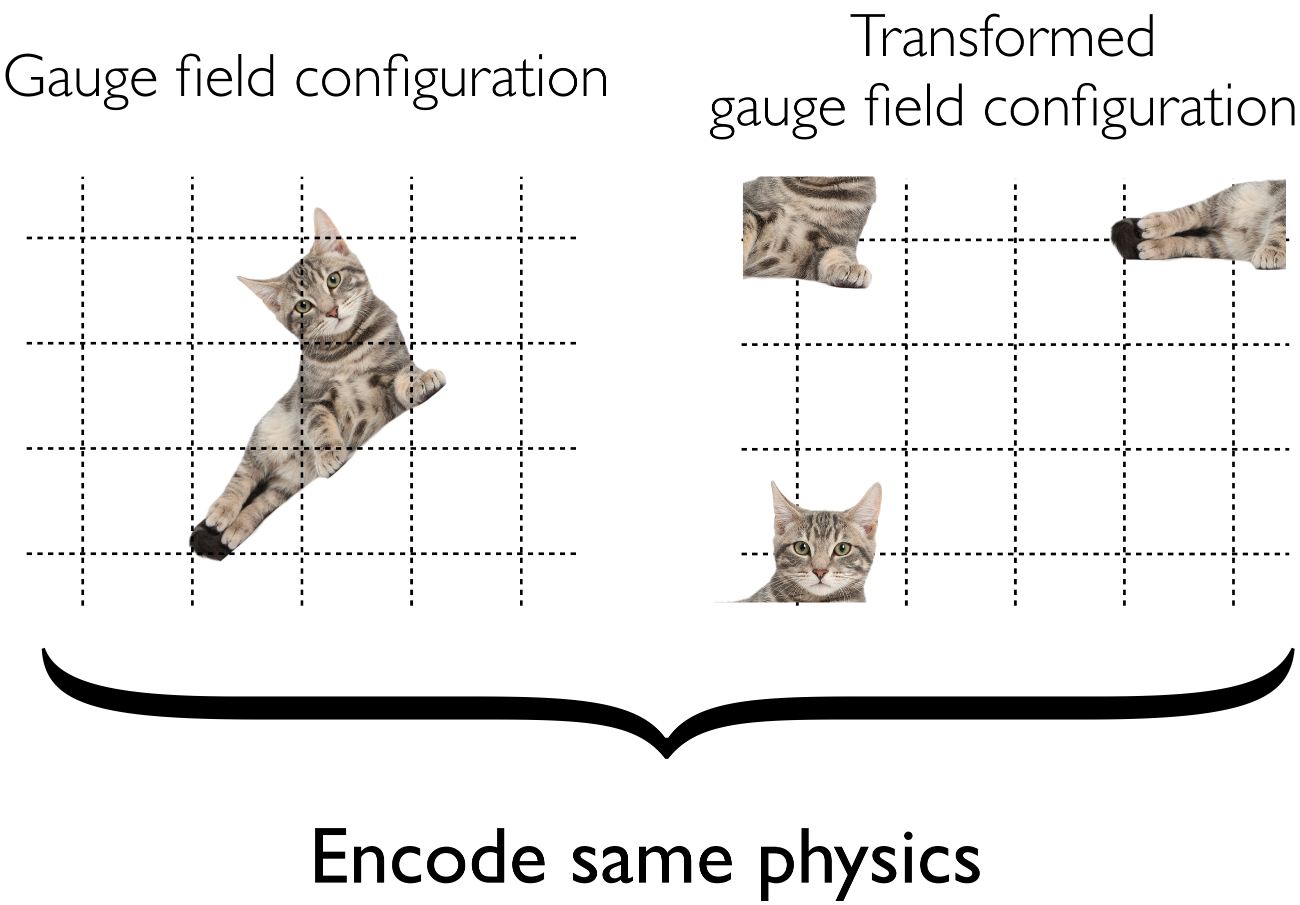
- $64^3 \times 128 \times 4 \times N_c^2 \times 2 \simeq 10^9$ numbers
- ~1000 samples
- Ensemble of gauge fields has meaning
- Long-distance correlations are important
- Gauge and translation-invariant with periodic boundaries

# Symmetries of LQCD gauge fields

**Physics** encoded by lattice QCD gauge fields
is invariant under specific field transformations

- **Rotation (4D)**
- **Translation**

with periodic
boundary conditions

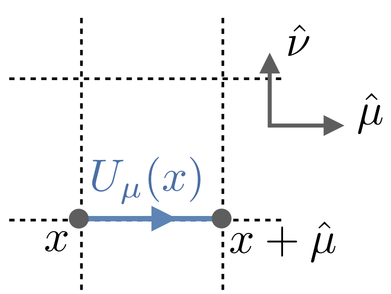


# Symmetries of LQCD gauge fields

**Physics** encoded by lattice QCD gauge fields
is invariant under specific field transformations

- **Gauge transformation**

  Separate group transformation of each link matrix $U_\mu(x)$

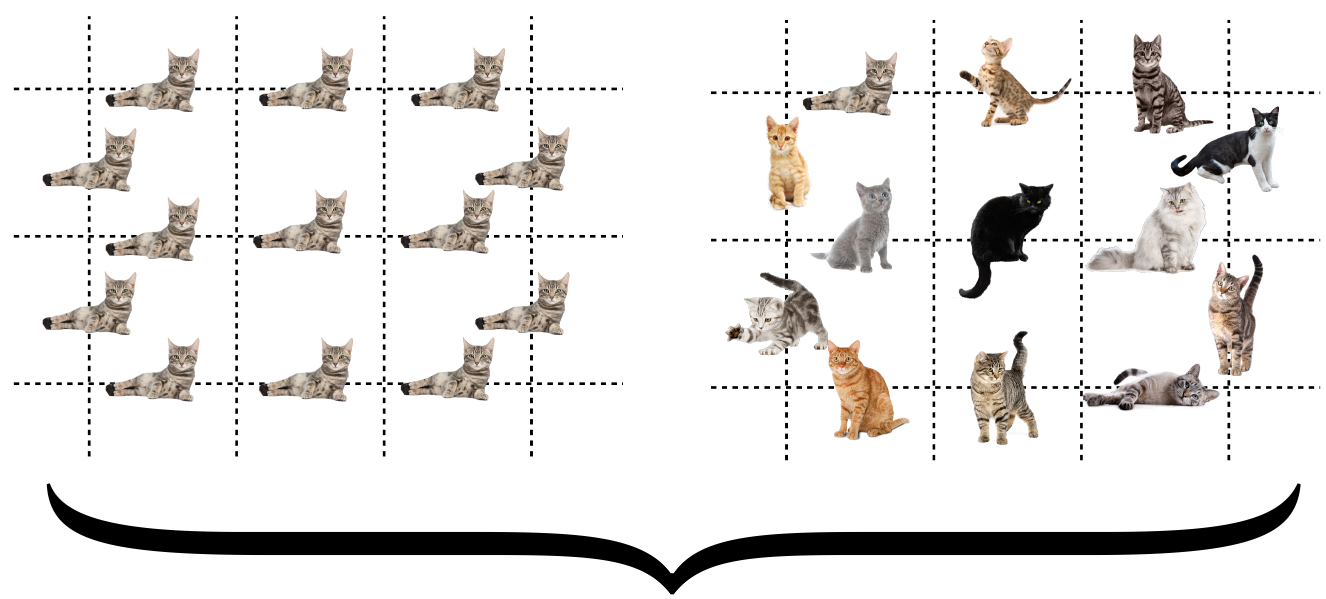



$$U_\mu(x) \to U'_\mu(x) = \Omega(x) U_\mu(x) \Omega^\dagger(x+\hat{\mu})$$

$$\text{for all } \Omega(x) \in \mathrm{SU}(3)$$

Gauge field configuration

Transformed gauge field configuration



**Encode same physics**

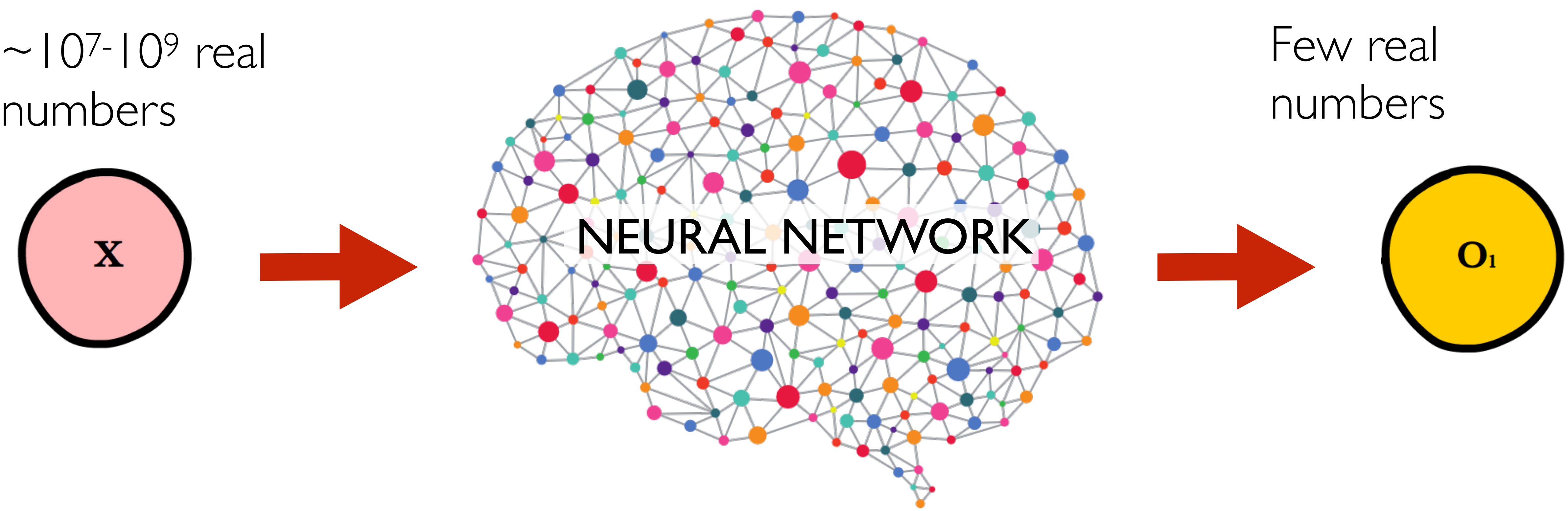
# Regression by neural network

Lattice QCD gauge field

~$10^7$-$10^9$ real numbers

Parameters of lattice action

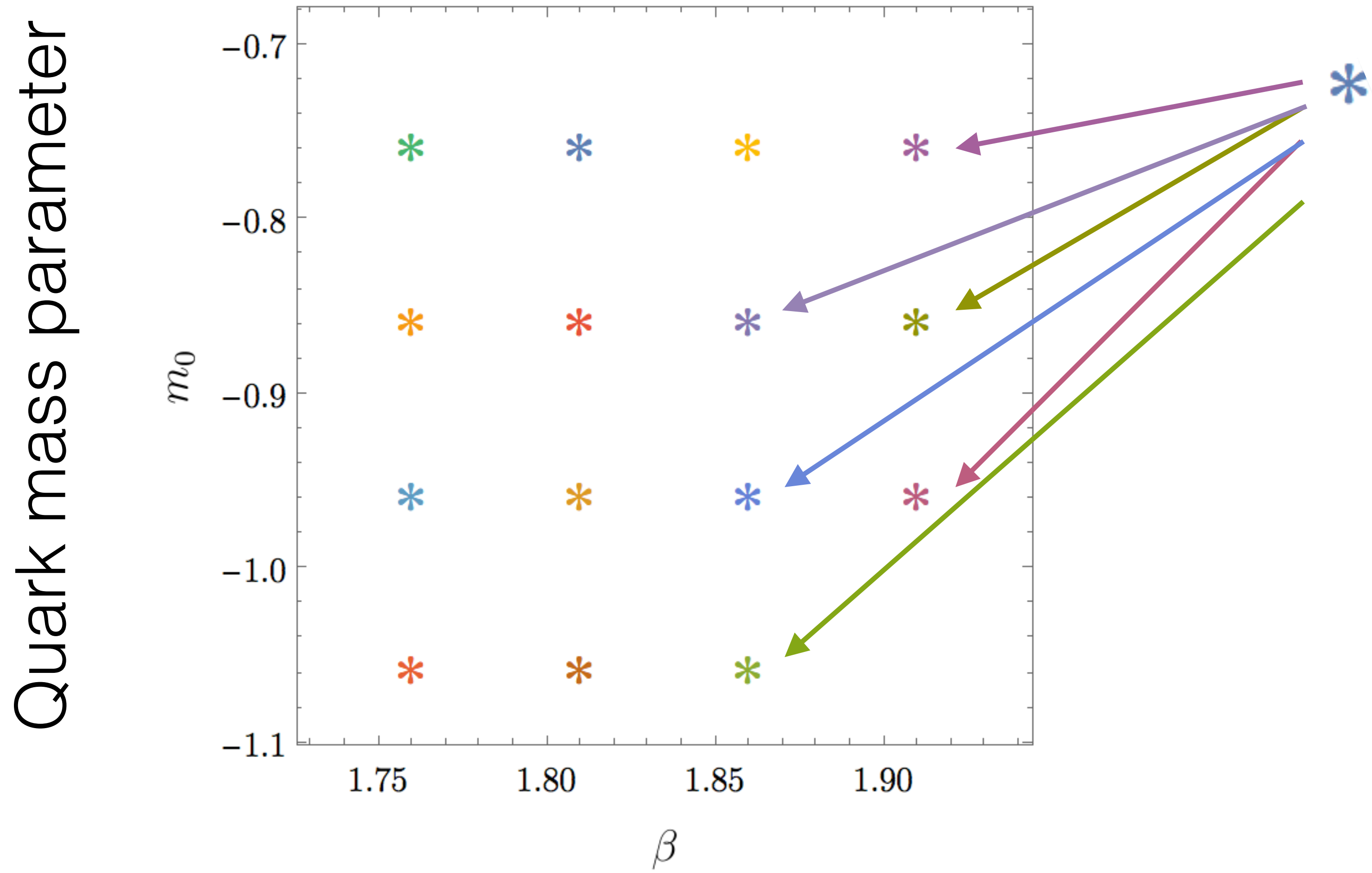
Few real numbers




- **Complete:** not restricted to affordable subset of physics parameters
- **Fast:** once trained over a parameter range

# Training data

## Training and validation datasets




Parameter related to lattice spacing

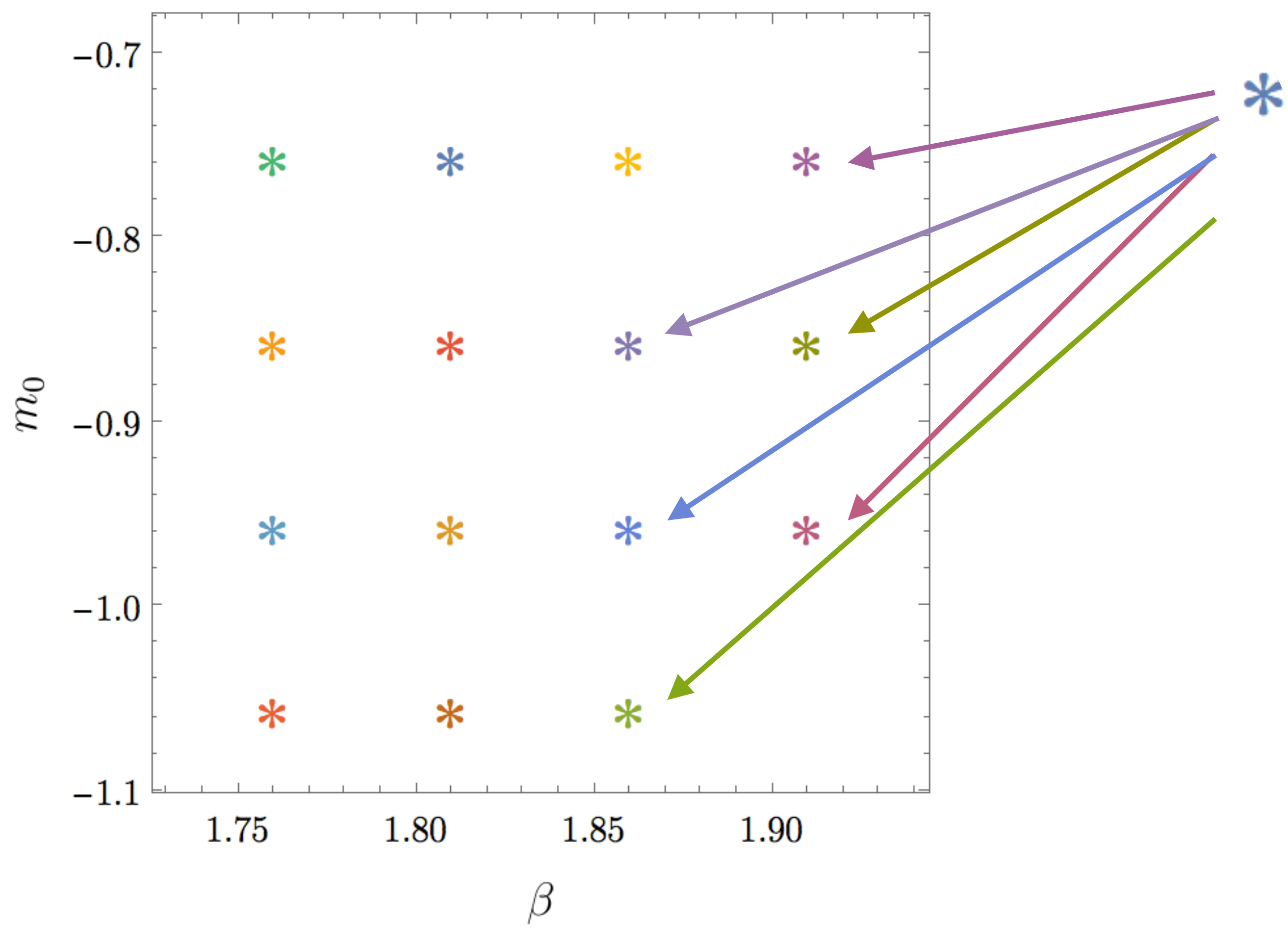
Parameters of training and validation datasets

**Cheaper to produce large sets at a few sets of parameters than well-distributed samples**

# Training data

## Training and validation datasets

Quark mass parameter




Parameter related to lattice spacing

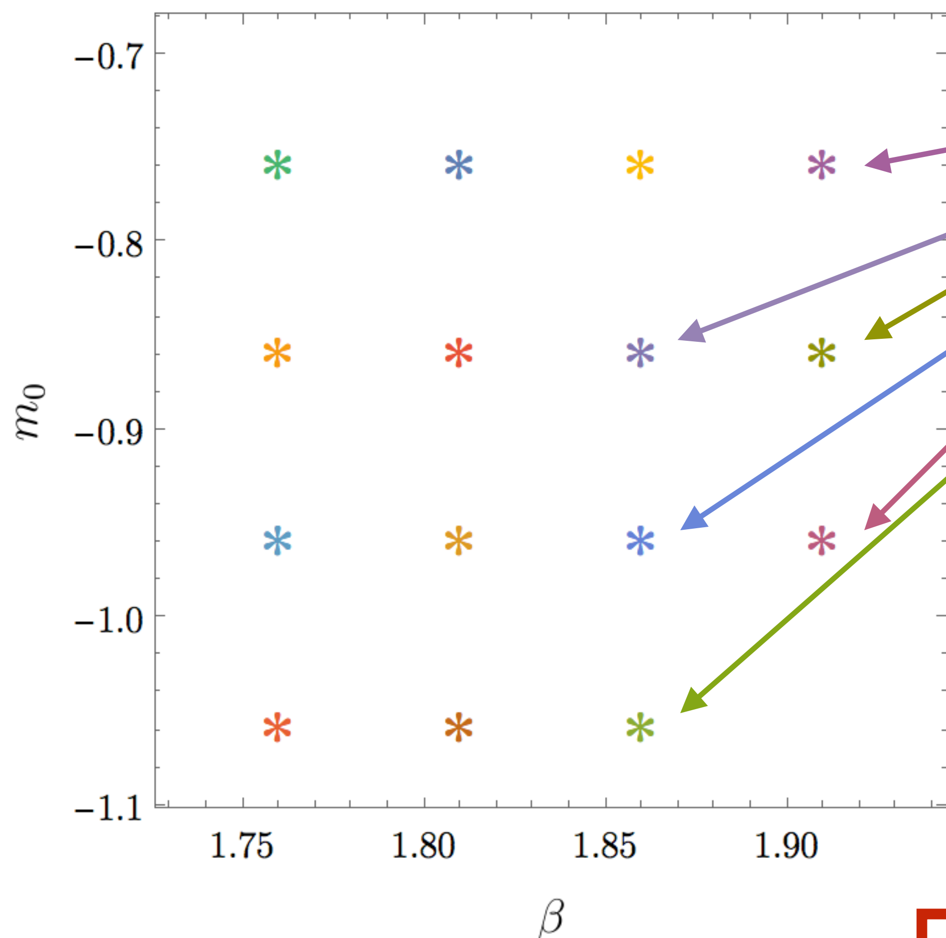
Parameters of training and validation datasets

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

# Training data

## Training and validation datasets

Quark mass parameter

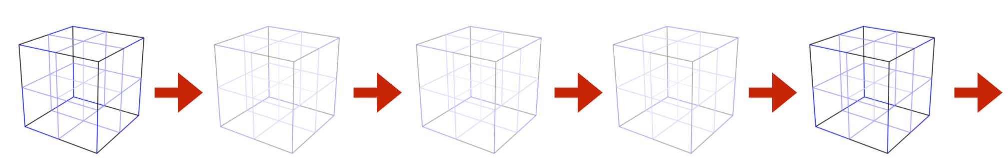



Parameter related to lattice spacing

* 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



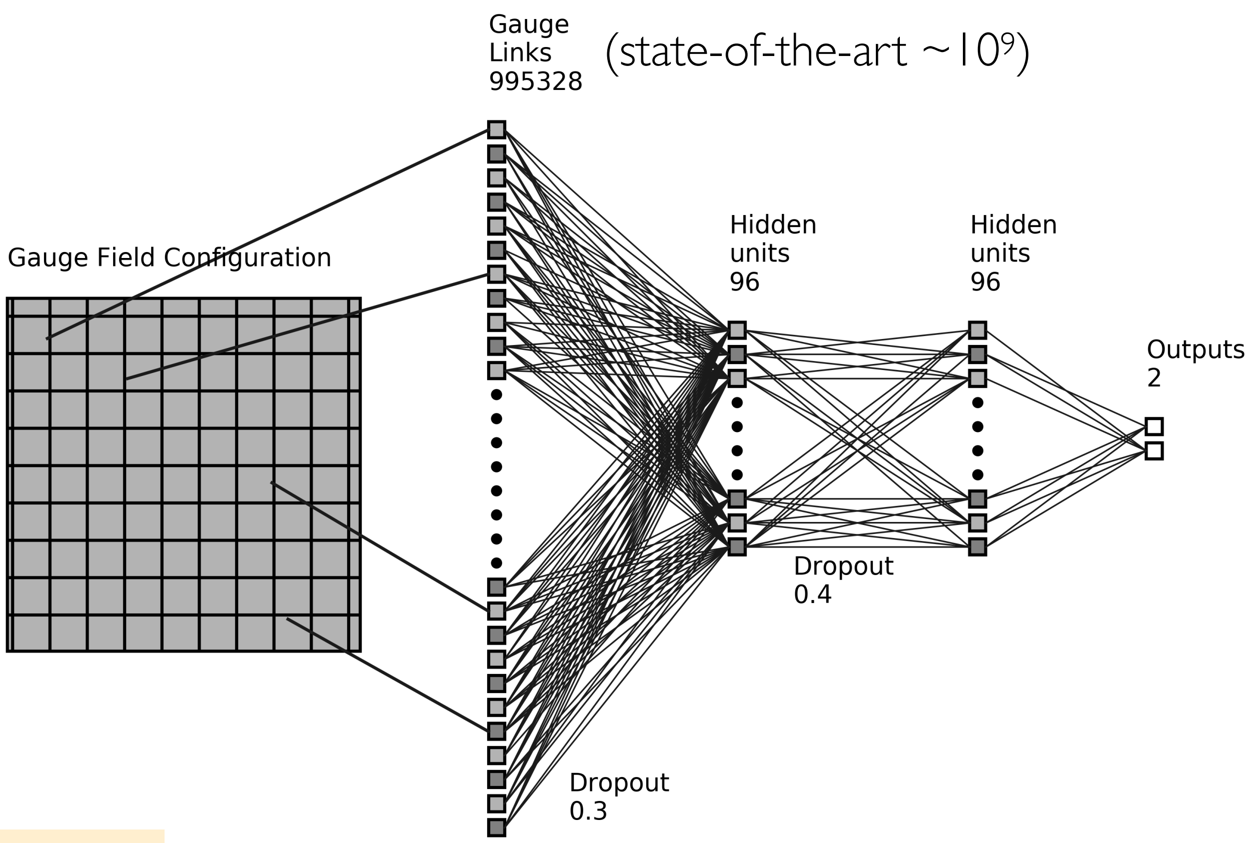
# Naive neural network

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

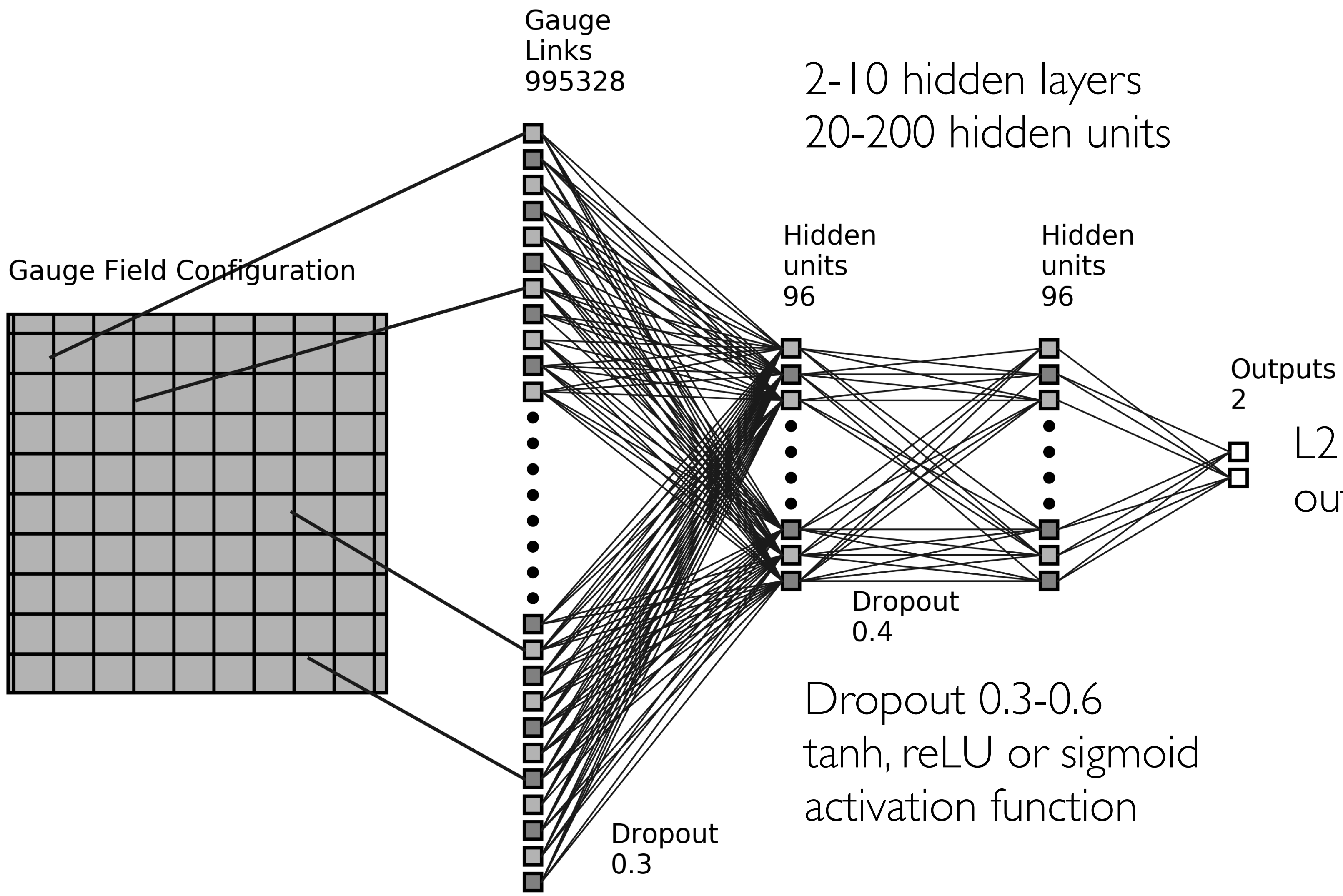
"Inverted data hierarchy" → Recipe for overfitting!



# Naive neural network

Simplest approach ⟷ Ignore physics symmetries

Flatten gauge links into a vector (state-of-the-art ~$10^9$ double-precision real numbers)




2-10 hidden layers
20-200 hidden units

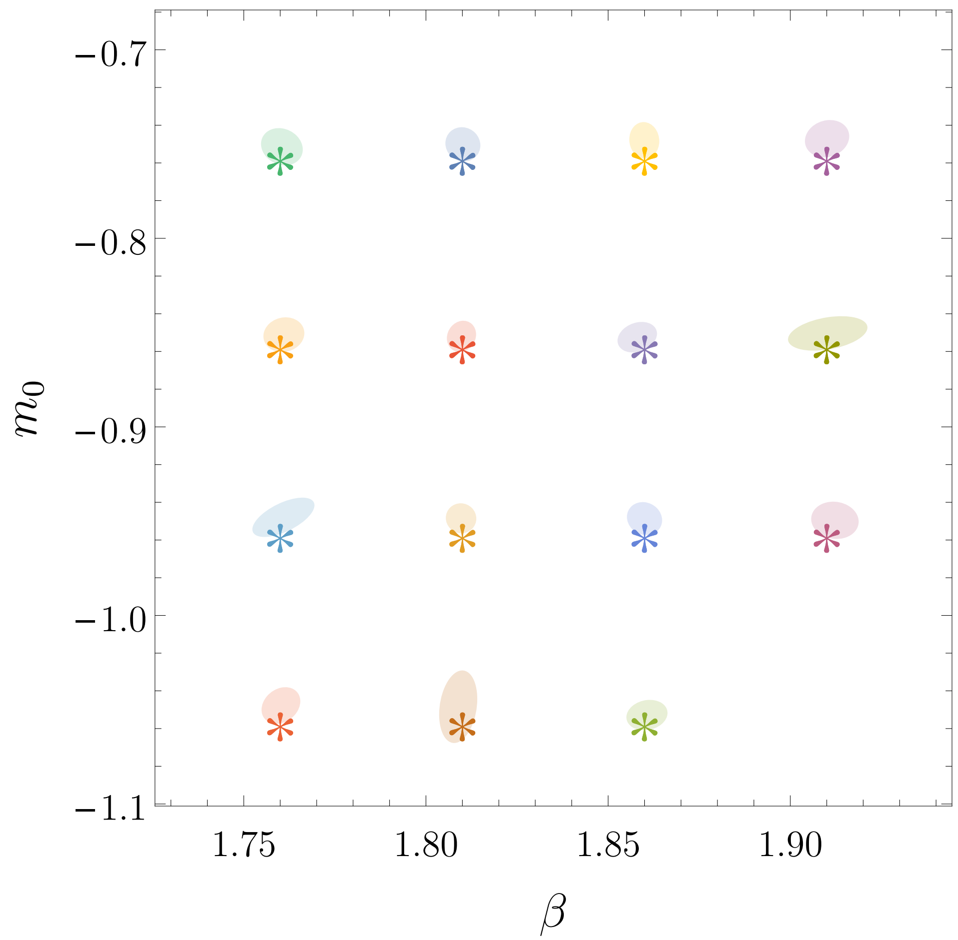
L2 loss in 2D output space

Dropout 0.3-0.6
tanh, reLU or sigmoid activation function

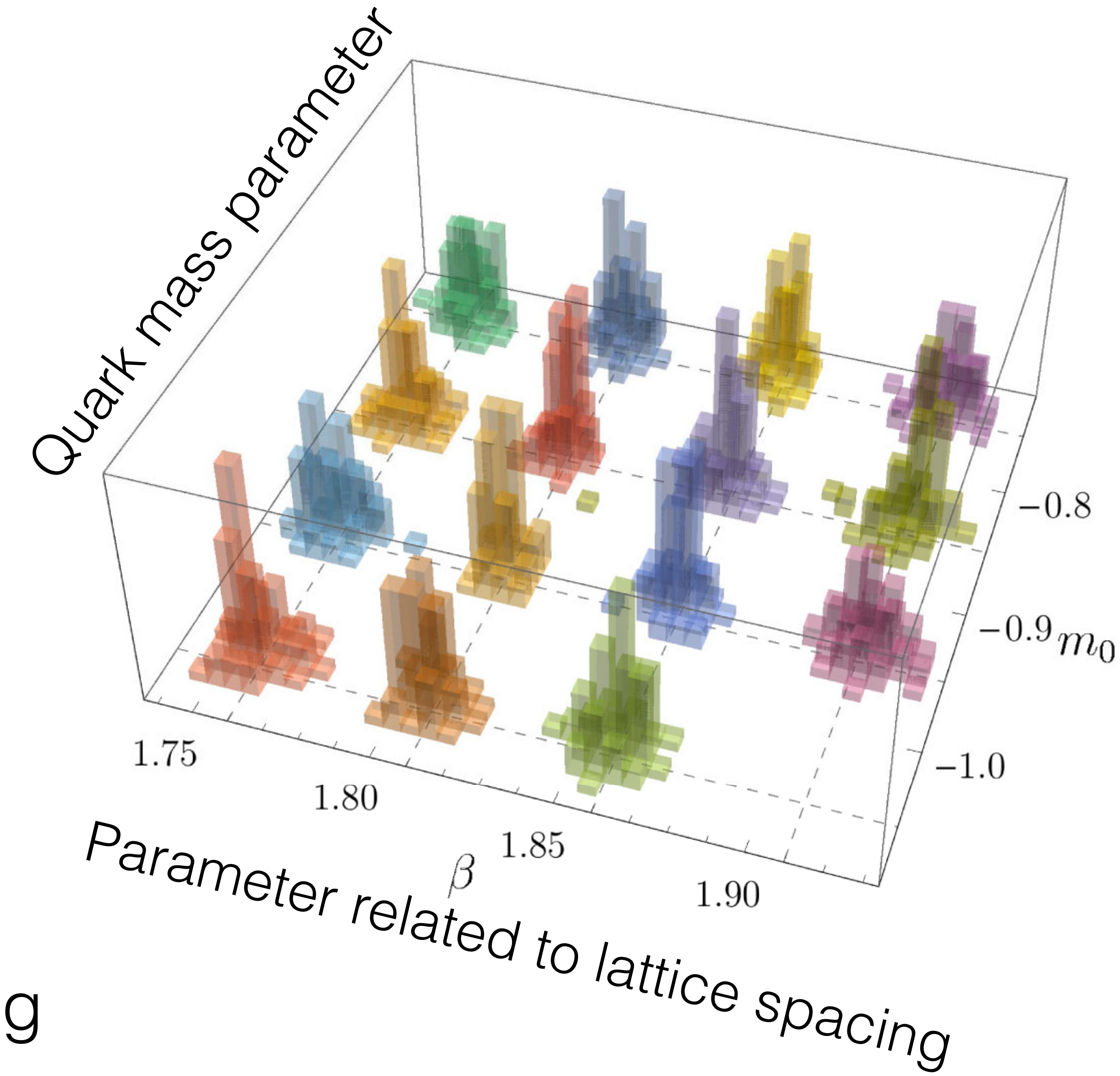
# Naive neural network

## Neural net predictions on validation data sets

Quark mass parameter




Parameter related to lattice spacing

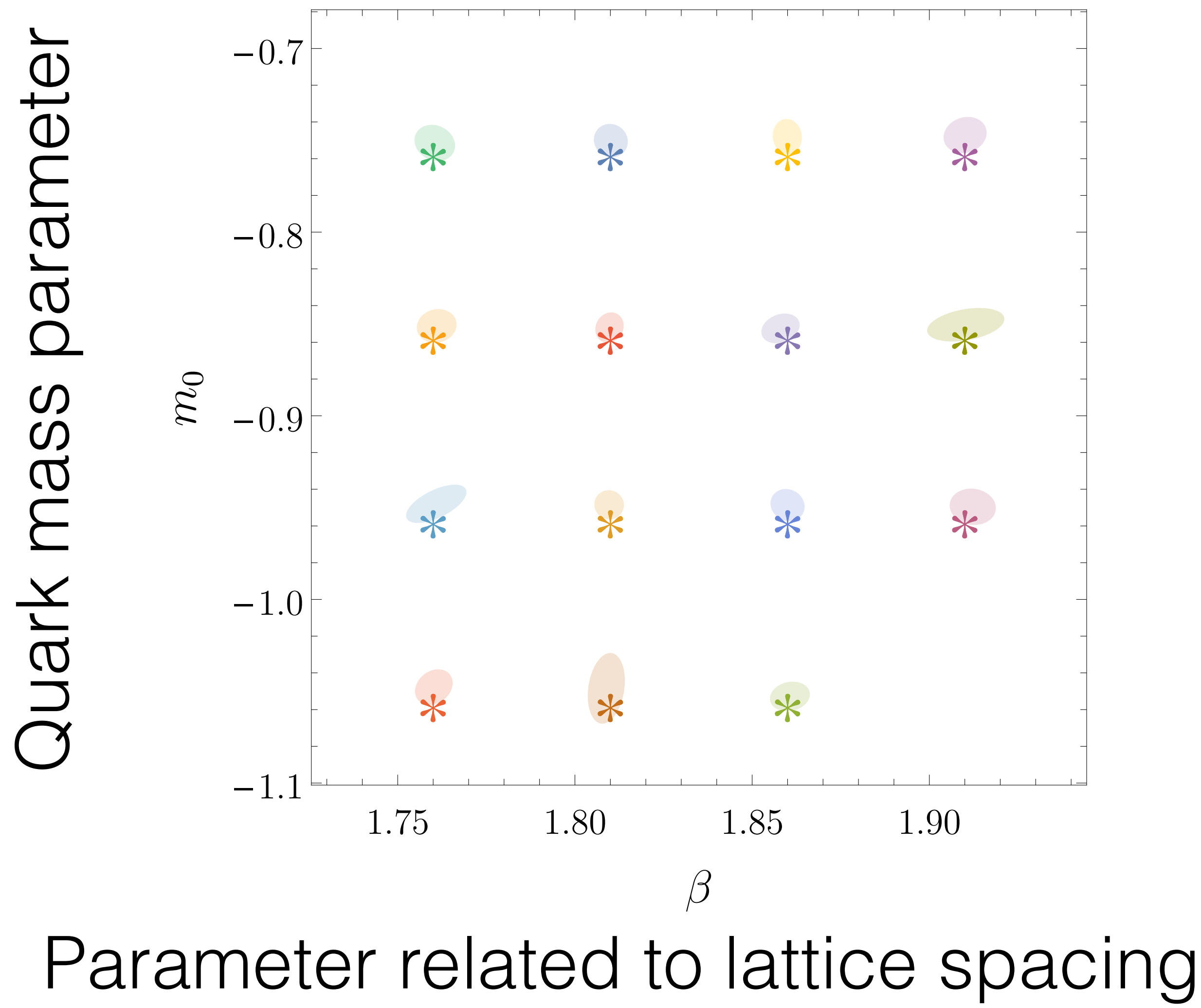



* True parameter values

Confidence interval from ensemble of gauge fields

# Naive neural network

Neural net predictions
on validation data sets




* True parameter values

Confidence interval from ensemble of gauge fields

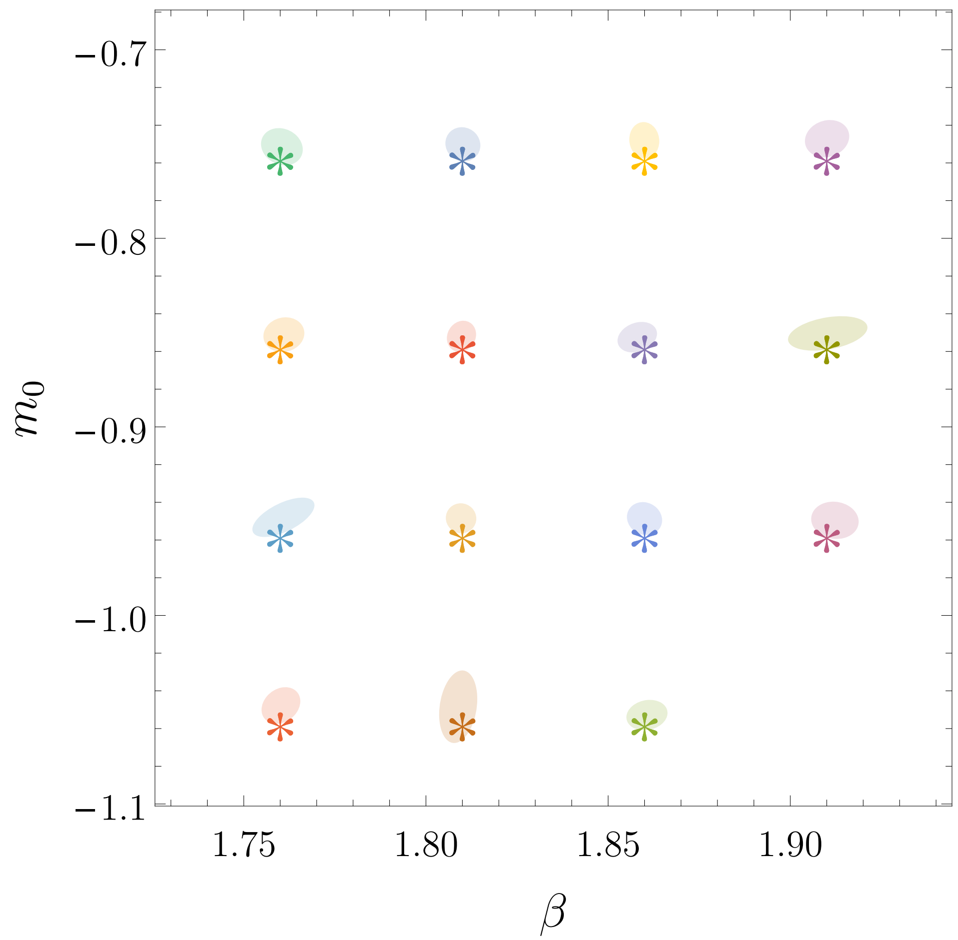
**SUCCESS?**

No sign of overfitting

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

# Naive neural network

Neural net predictions on validation data sets




Quark mass parameter

Parameter related to lattice spacing

* 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

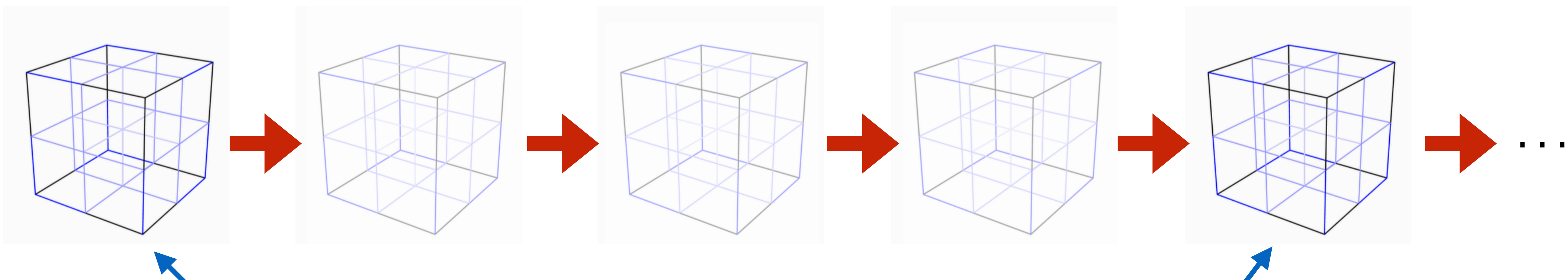
**BUT fails to generalise to**

- 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**

# Autocorrelation times

**Autocorrelation time:** measure of correlation length in HMC trajectory (how many steps separate independent samples)

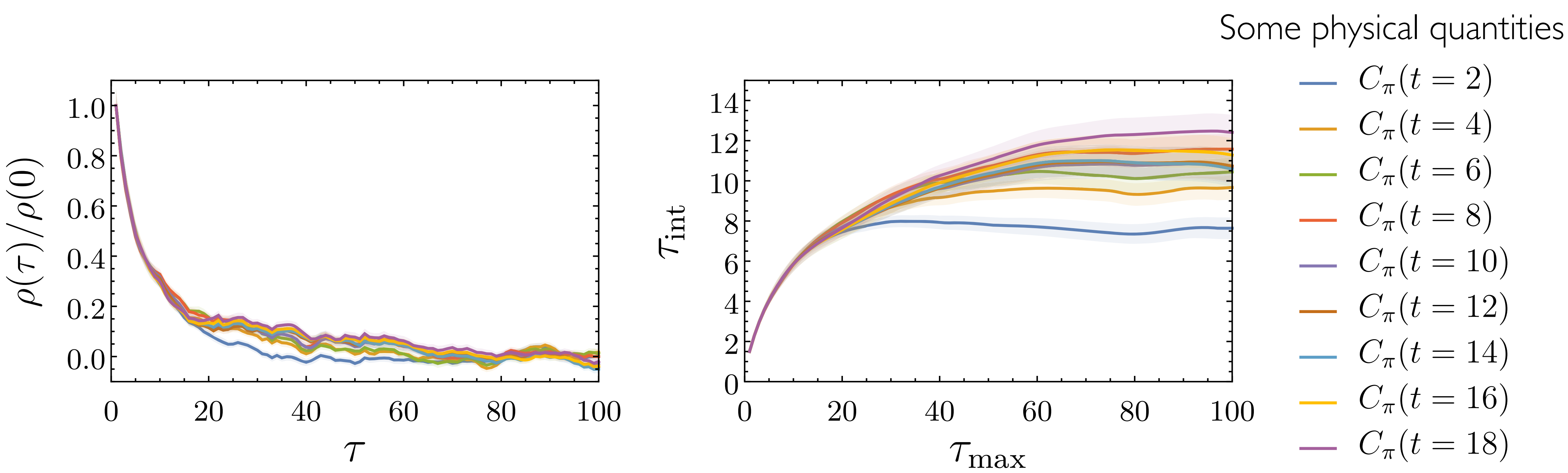
Autocorrelation $$\rho(\tau) = \sum_{\tau'} \langle (\mathcal{O}(\tau') - \langle \mathcal{O} \rangle)(\mathcal{O}(\tau' + \tau) - \langle \mathcal{O} \rangle) \rangle$$

$\langle \ldots \rangle$ : average over confs

$\mathcal{O}$: some physical quantity

$\tau$: HMC trajectory time

Autocorrelation time: $$\tau_{\mathrm{int}} = \frac{1}{2} + \lim_{\tau_{\mathrm{max}} \to \infty} \frac{1}{\rho(0)} \sum_{\tau=0}^{\tau_{\mathrm{max}}} \rho(\tau),$$



# NN autocorrelation times

Define an autocorrelation time for the neural network

**Train network to classify configurations from pairs of HMC streams generated with the same parameters**

- Same network structure as for regression task
- Softmax in final layer: probability interpretation of network output
- Categorical cross-entropy loss
- **Training data:** consecutive confs in HMC stream
  **Test data:** after the end of the sequence used for training

Trajectory from stream $\gamma$
$\tau$ steps in MC time after the end of sequence used as training data

$$\rho(\tau) = 2[P_\gamma(c^\gamma(\tau))] - 1$$

Probability (from classifier output) that $c$ is in stream $\gamma$

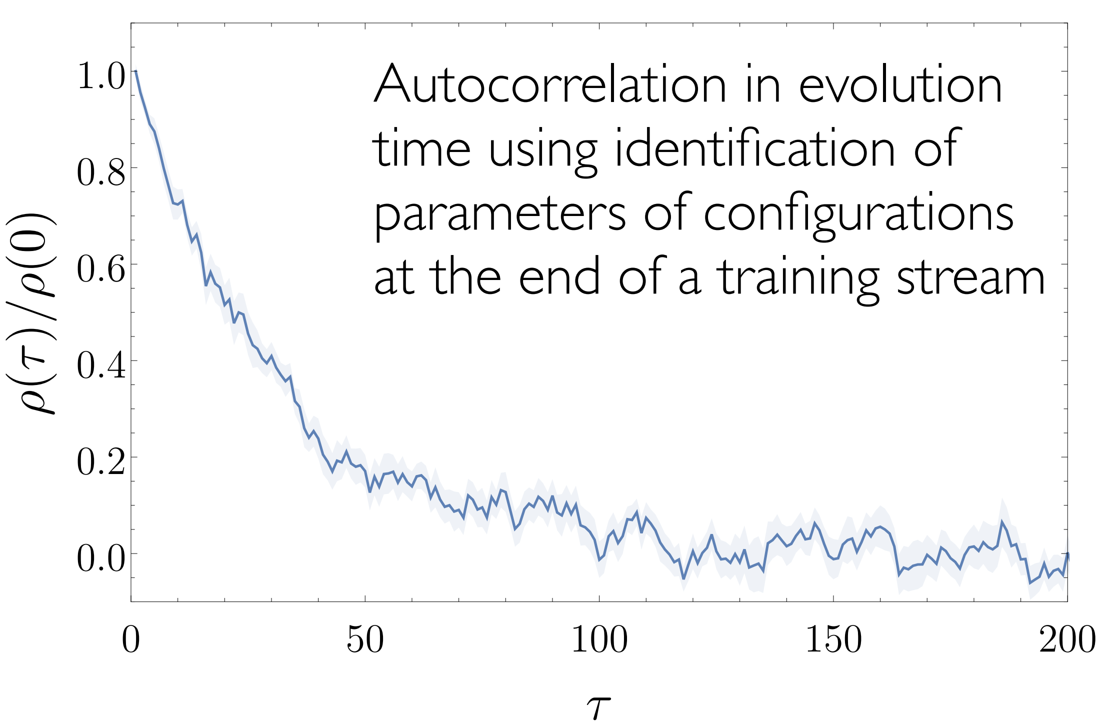
# Naive neural network

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

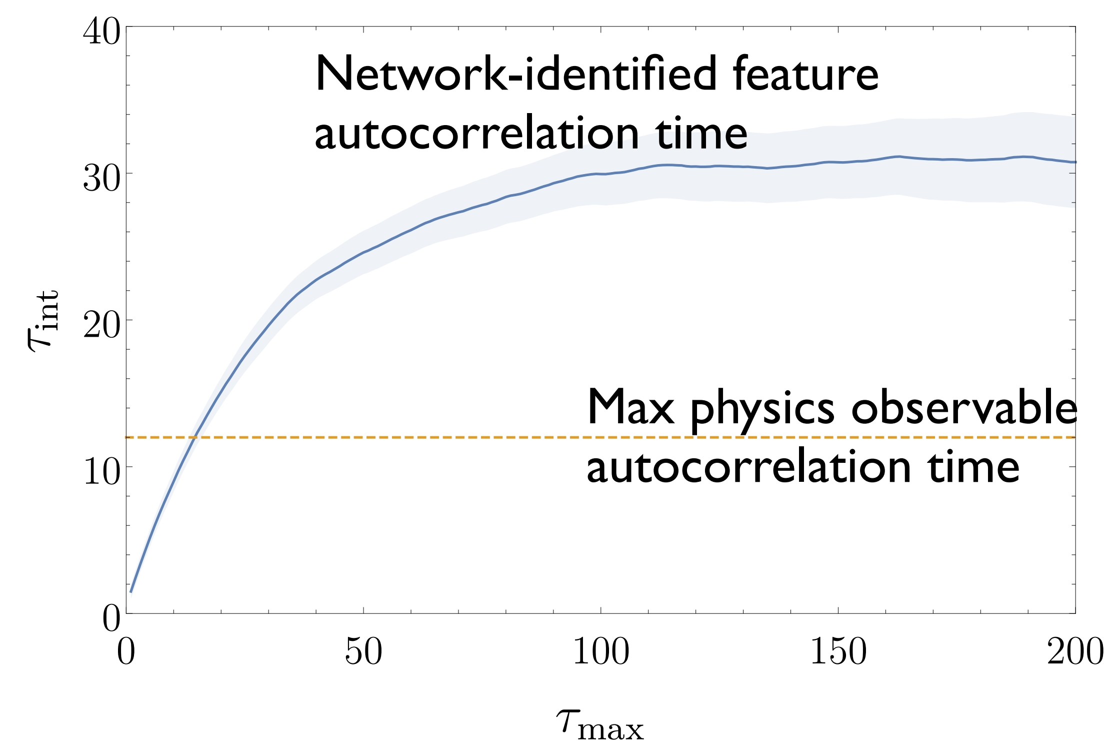
**BUT**

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

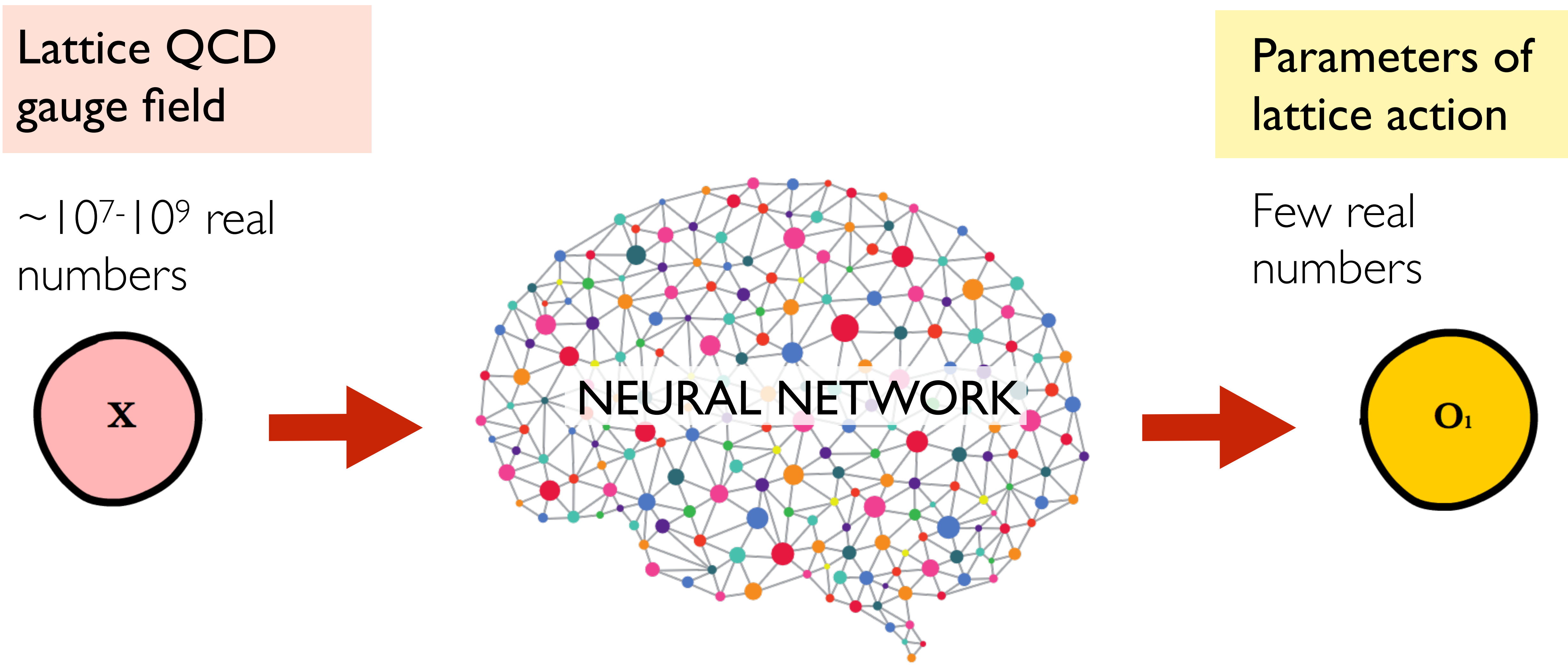
**Network feature autocorrelation**




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



# Regression by neural network

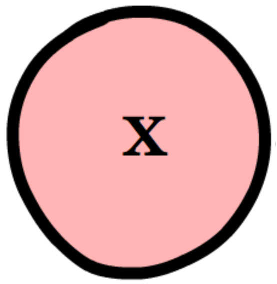



- **Complete:** not restricted to affordable subset of physics parameters
- **Fast:** once trained over a parameter range

# Regression by neural network

**Lattice QCD gauge field**

~$10^7$-$10^9$ real numbers

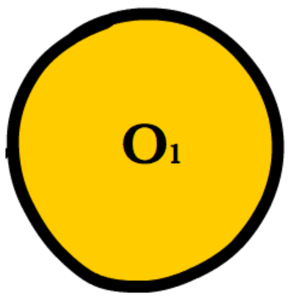



**Custom network structures (or data preprocessing)**

- Respects gauge-invariance, translation-invariance, boundary conditions
- Emphasises QCD-scale physics
- Range of neural network structures find same minimum

**Parameters of lattice action**

Few real numbers

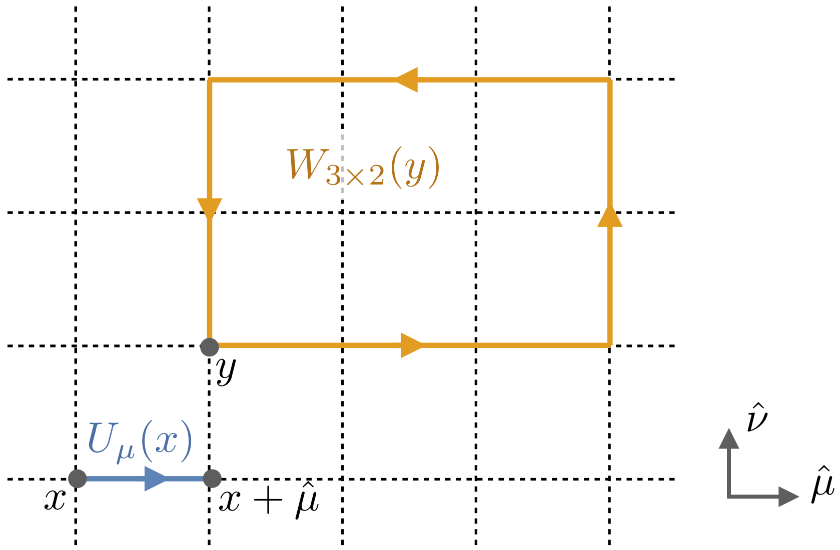



- **Complete:** not restricted to affordable subset of physics parameters
- **Fast:** once trained over a parameter range

# Symmetry-preserving network

Network based on symmetry-invariant features

Closed Wilson loops
(gauge-invariant)

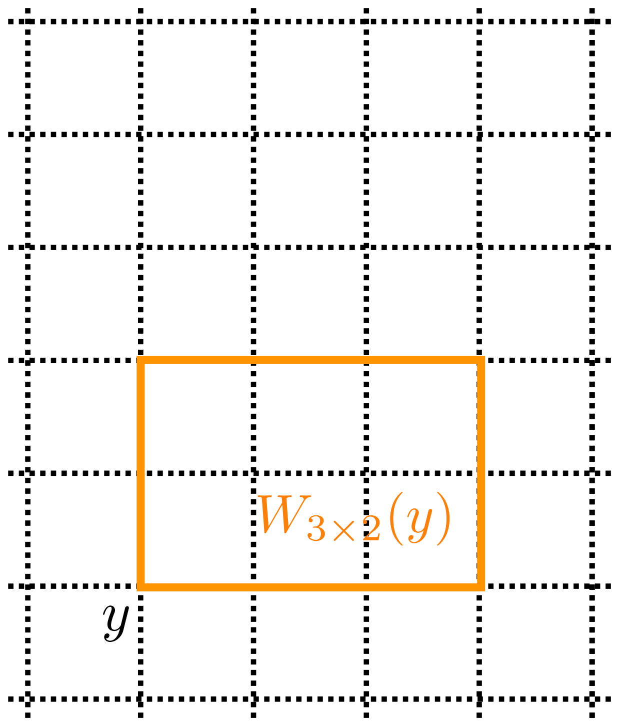



- Loops
- Correlated products of loops at various length scales
- Volume-averaged and rotation-averaged

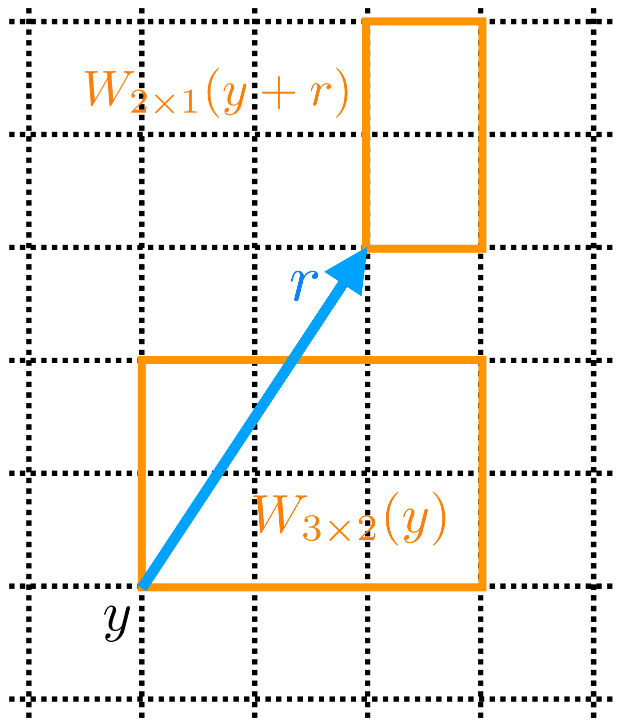
# Symmetry-preserving network

Network based on symmetry-invariant features

Loops

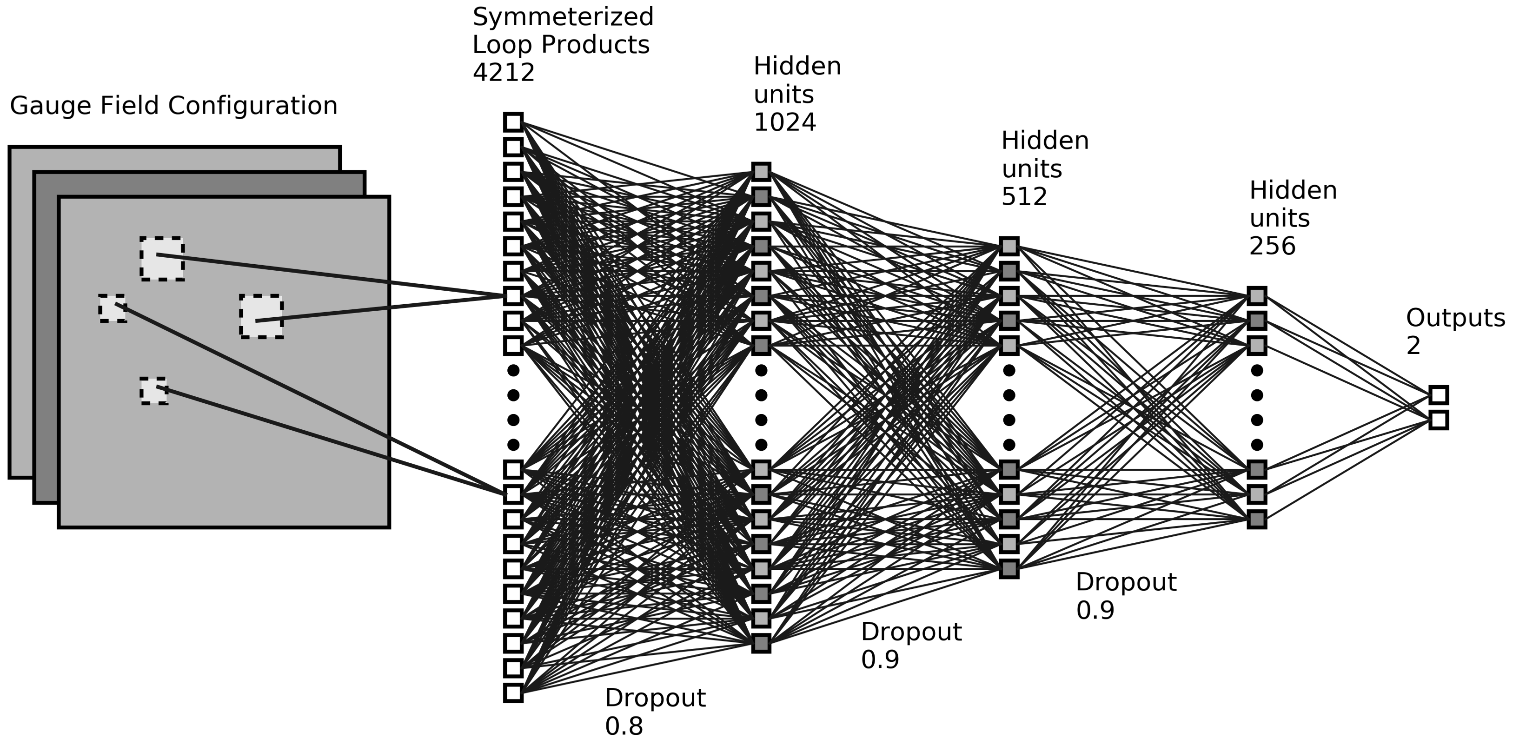



Correlated products of loops at various length scales



# Symmetry-preserving network

## Network based on symmetry-invariant features



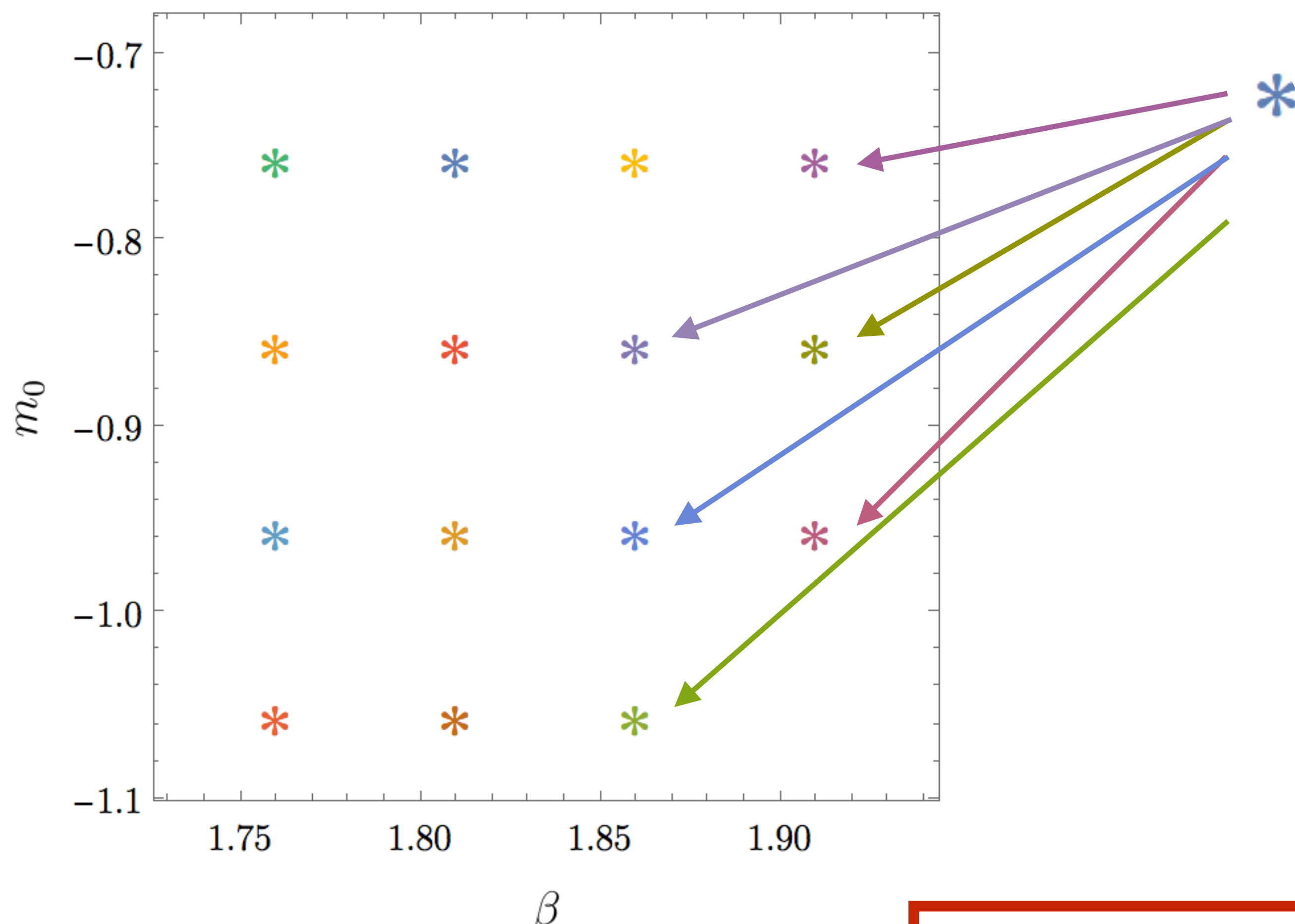

- Fully-connected network structure
- First layer samples from set of possible symmetry-invariant features

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

# Training data

## Training and validation datasets

Quark mass parameter

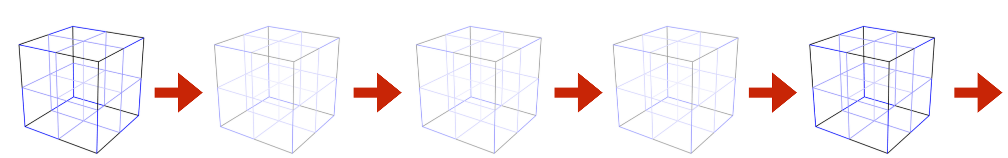



Parameter related to lattice spacing

* Parameters of training and validation datasets

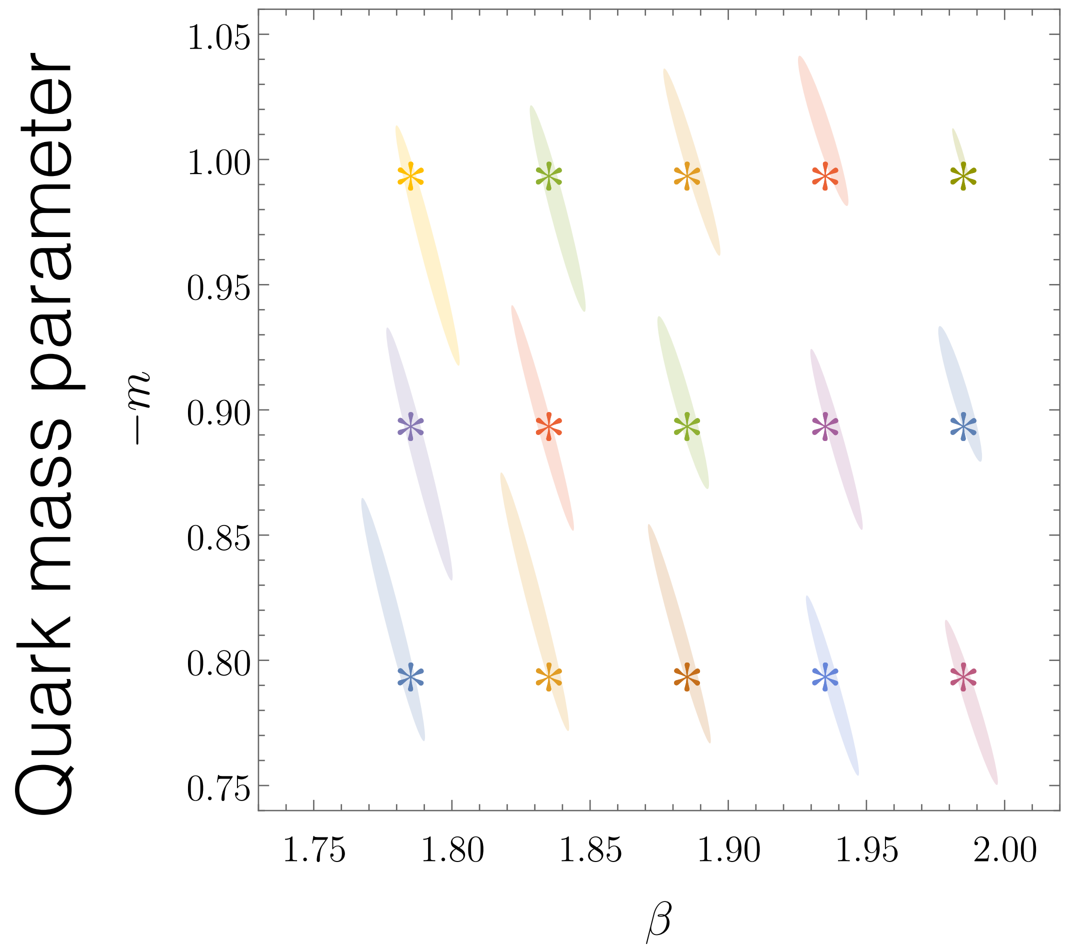
- O(10,000) independent configurations generated at each point
- Validation configurations randomly selected from generated streams

Spacing in evolution stream >> correlation time of physics observables



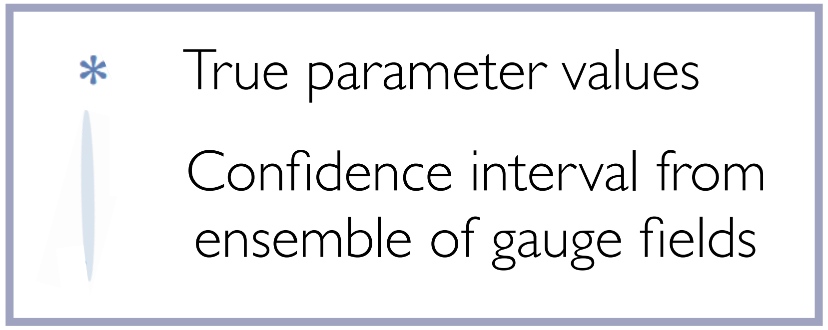
# Gauge field parameter regression

Neural net predictions
on validation data sets



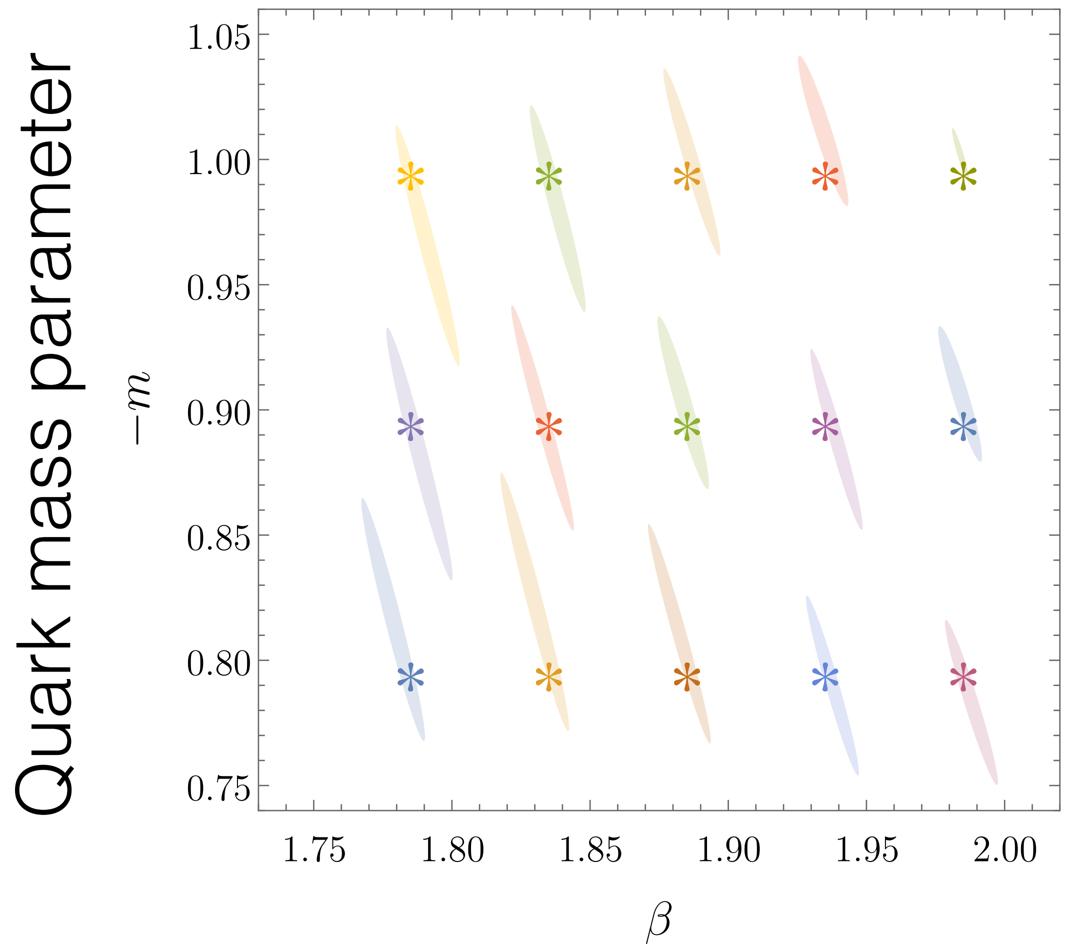

Parameter related
to lattice spacing

- Uncertainty in direction of constant 1x1 plaquette (simplest gauge-invariant object)
- 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)



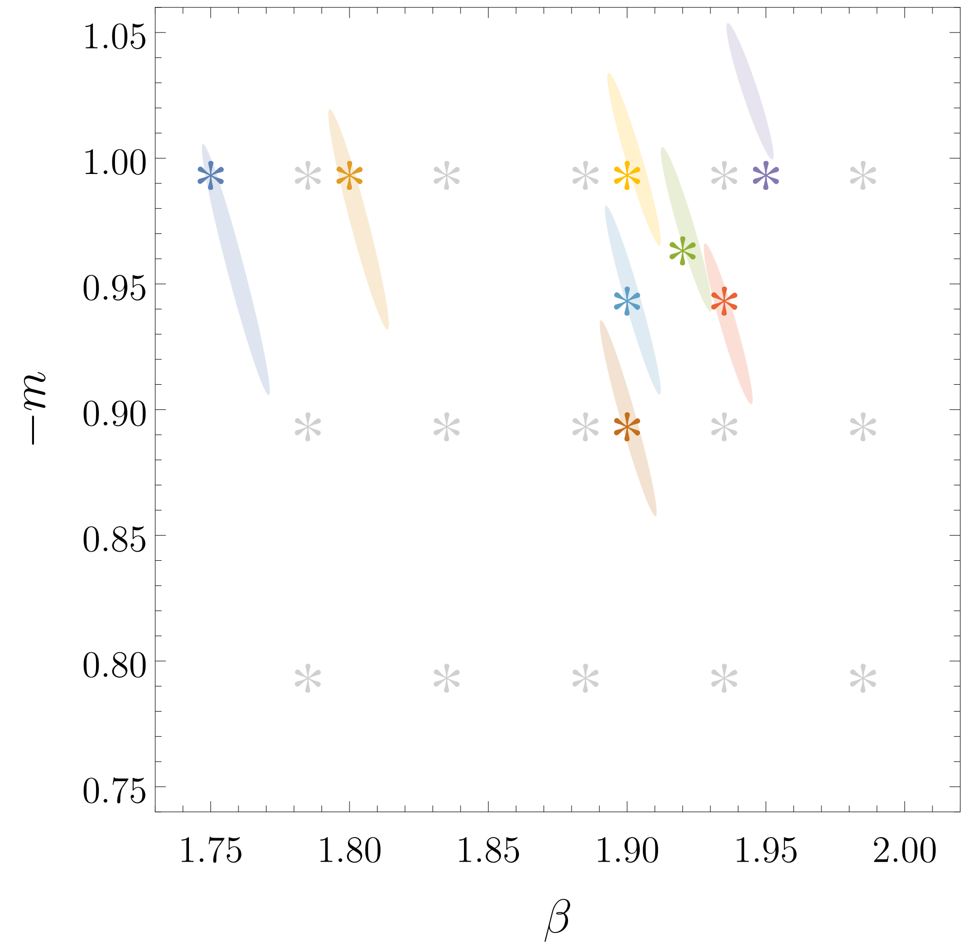
# Gauge field parameter regression

## Neural net predictions on validation data sets




Parameter related to lattice spacing

## Predictions on new datasets



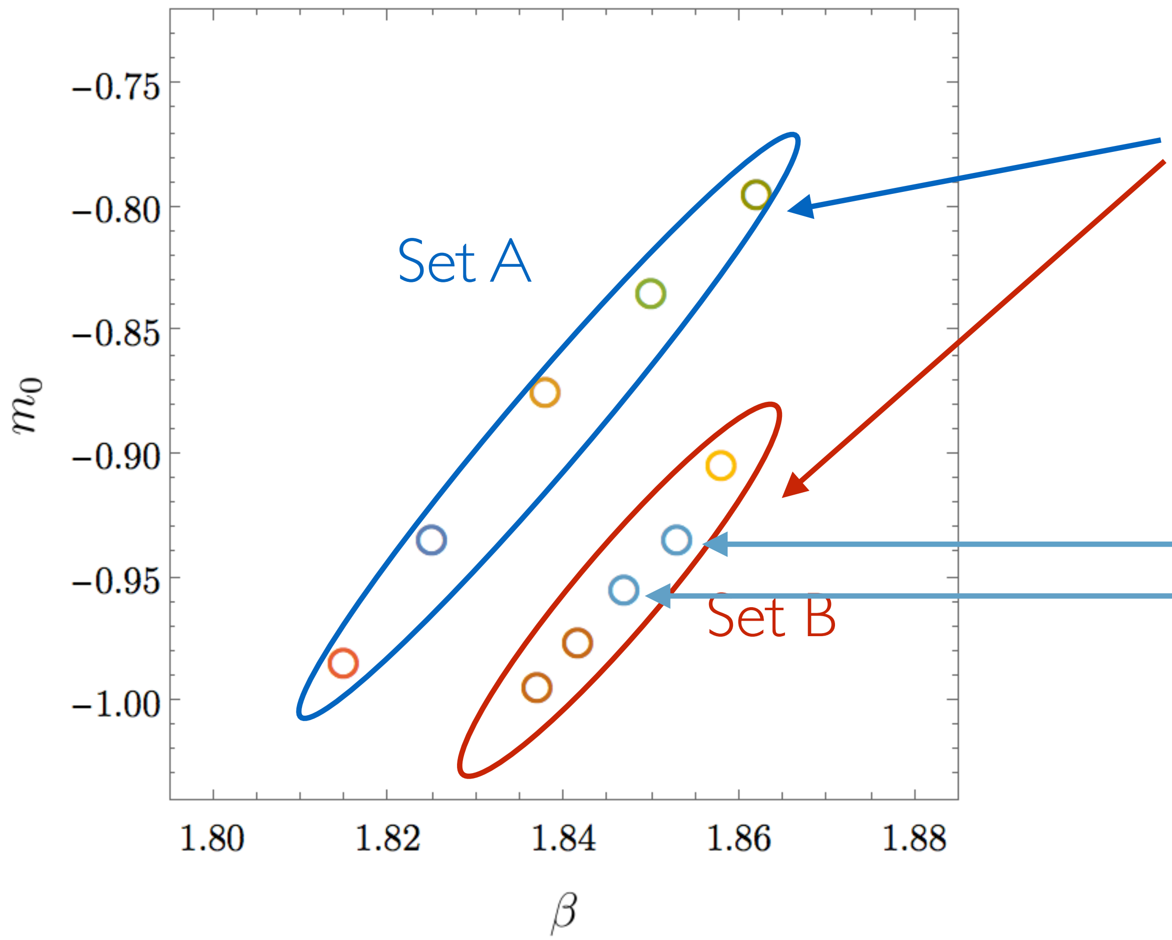

* True parameter values

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



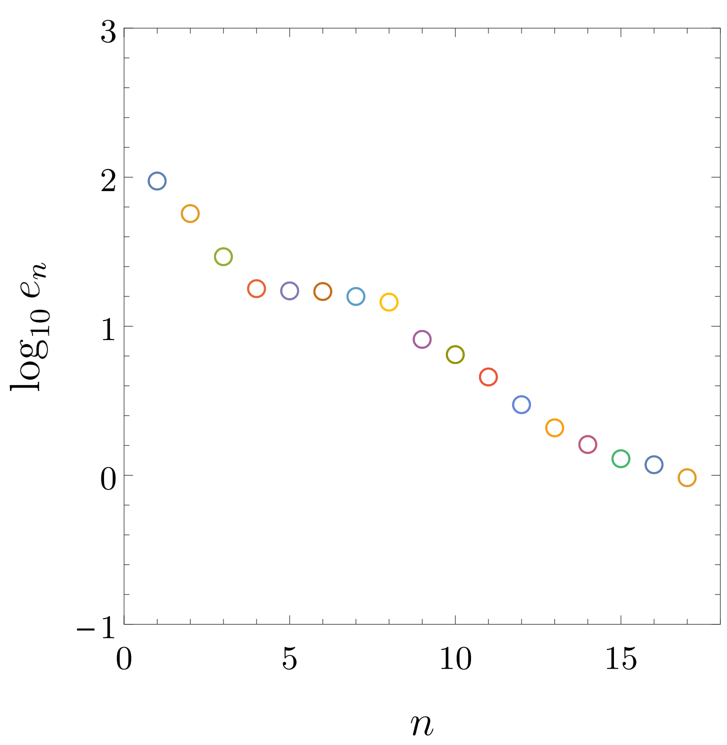

Sets along lines of **constant 1x1 Wilson loop** (most precise feature allowed by network)

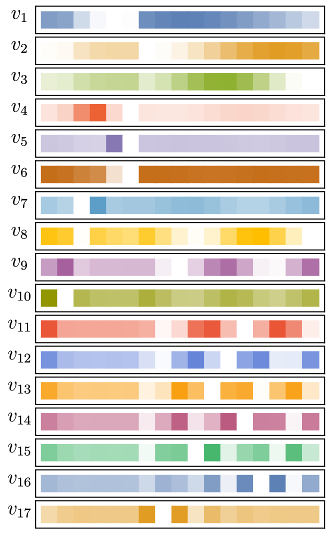
Much closer spacing than separation of training ensembles

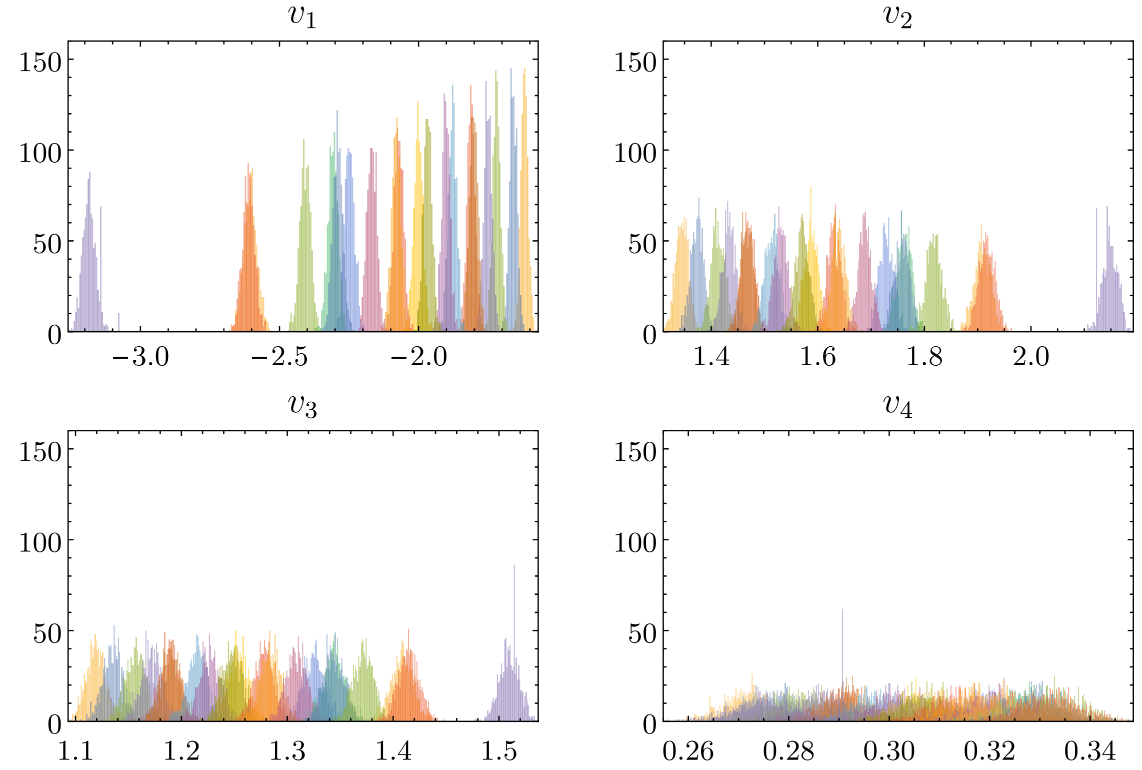
# How does neural network regression perform compared with other approaches?

Consider very closely-spaced validation ensembles at new parameters

**Principal component analysis:** find eigenvalues/vectors of matrix of cross-correlations of Wilson loops on training dataset



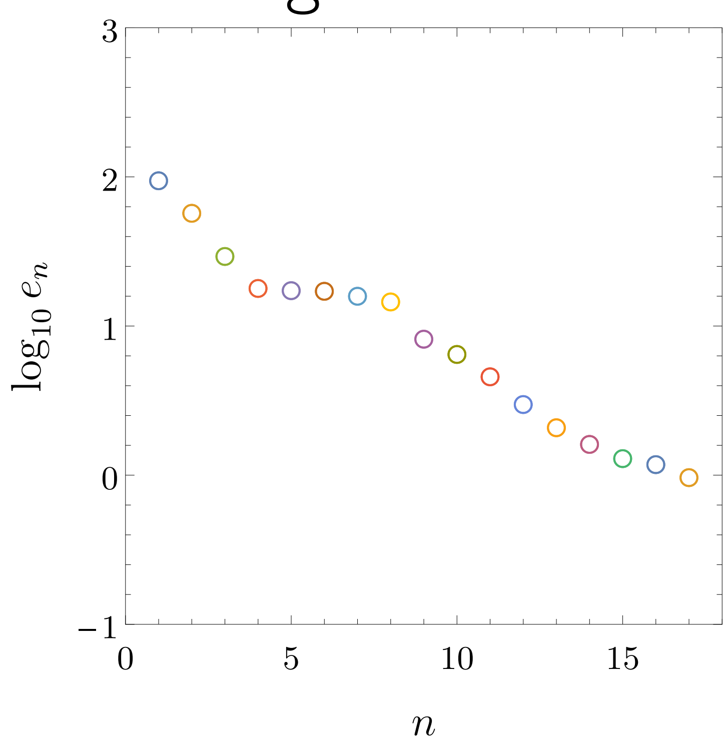



# Tests of network success

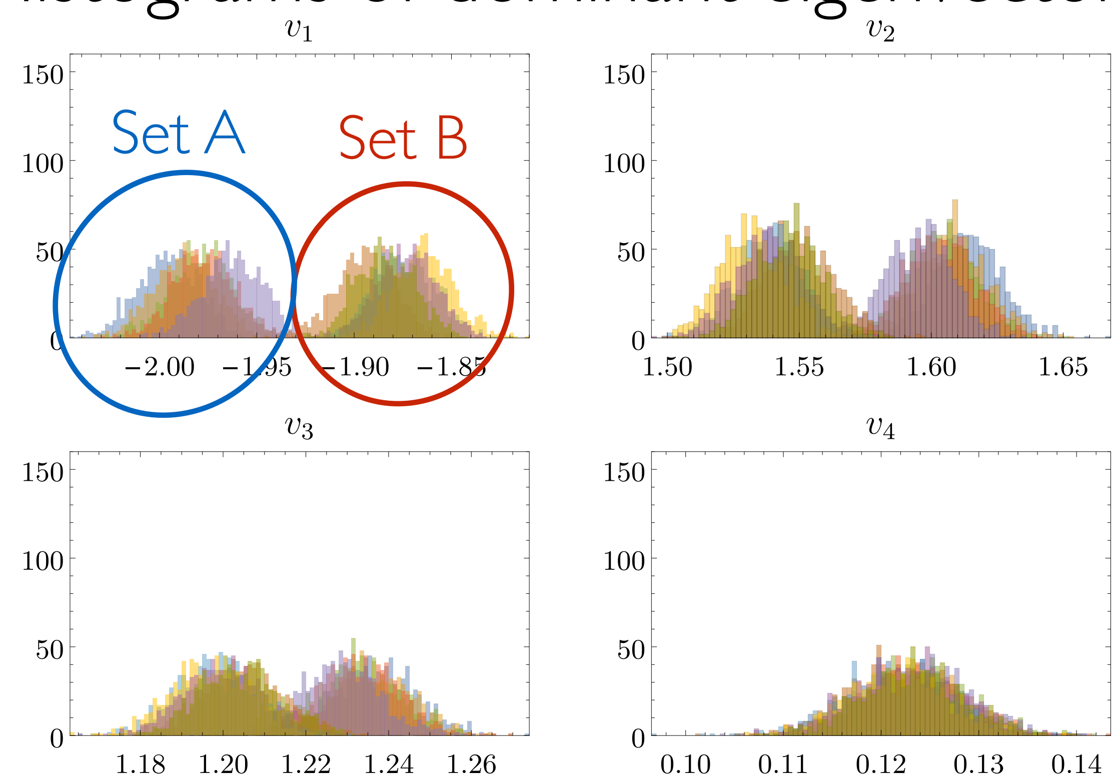
**How does neural network regression perform compared with other approaches?**

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

Eigenvalues



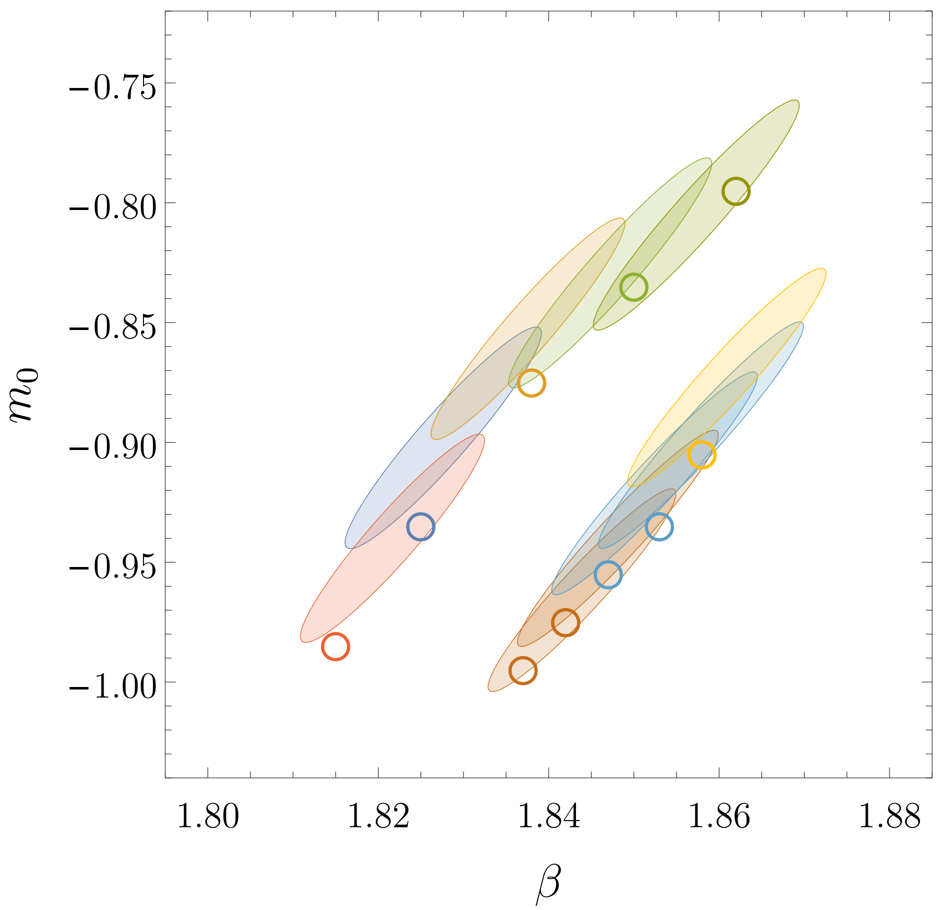

Histograms of dominant eigenvectors



# Tests of network success

**How does neural network regression perform compared with other approaches?**

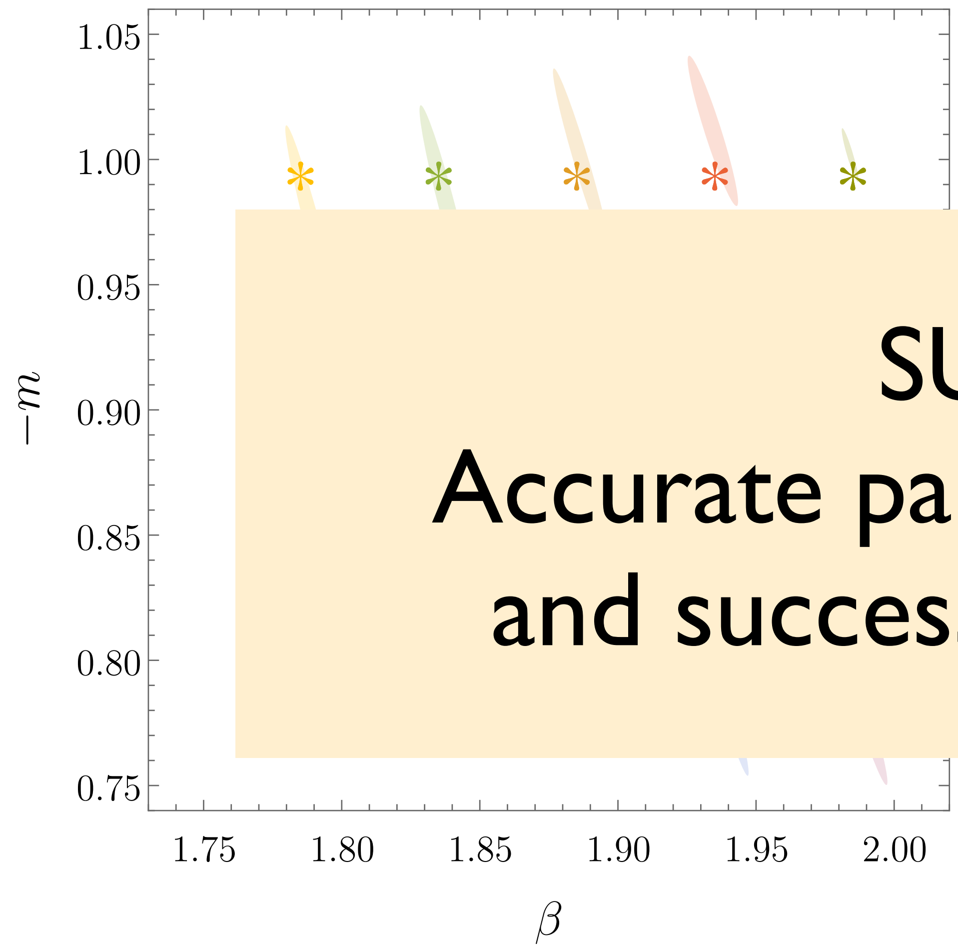
Consider very closely-spaced validation ensembles at new parameters: **distinguishable to trained neural network**




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

# Gauge field parameter regression

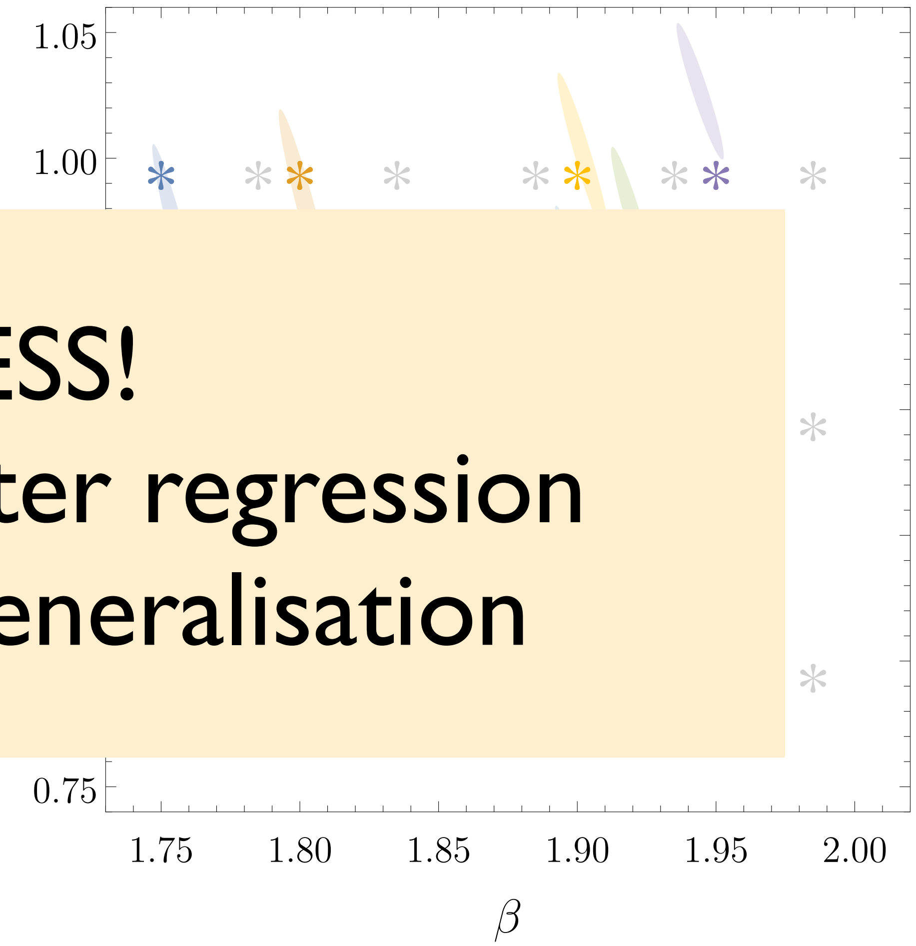
Neural net predictions on validation data sets




Quark mass parameter

Parameter related to lattice spacing

Predictions on new datasets




**SUCCESS!**

**Accurate parameter regression and successful generalisation**

* True parameter values

Confidence interval from ensemble of gauge fields

# 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

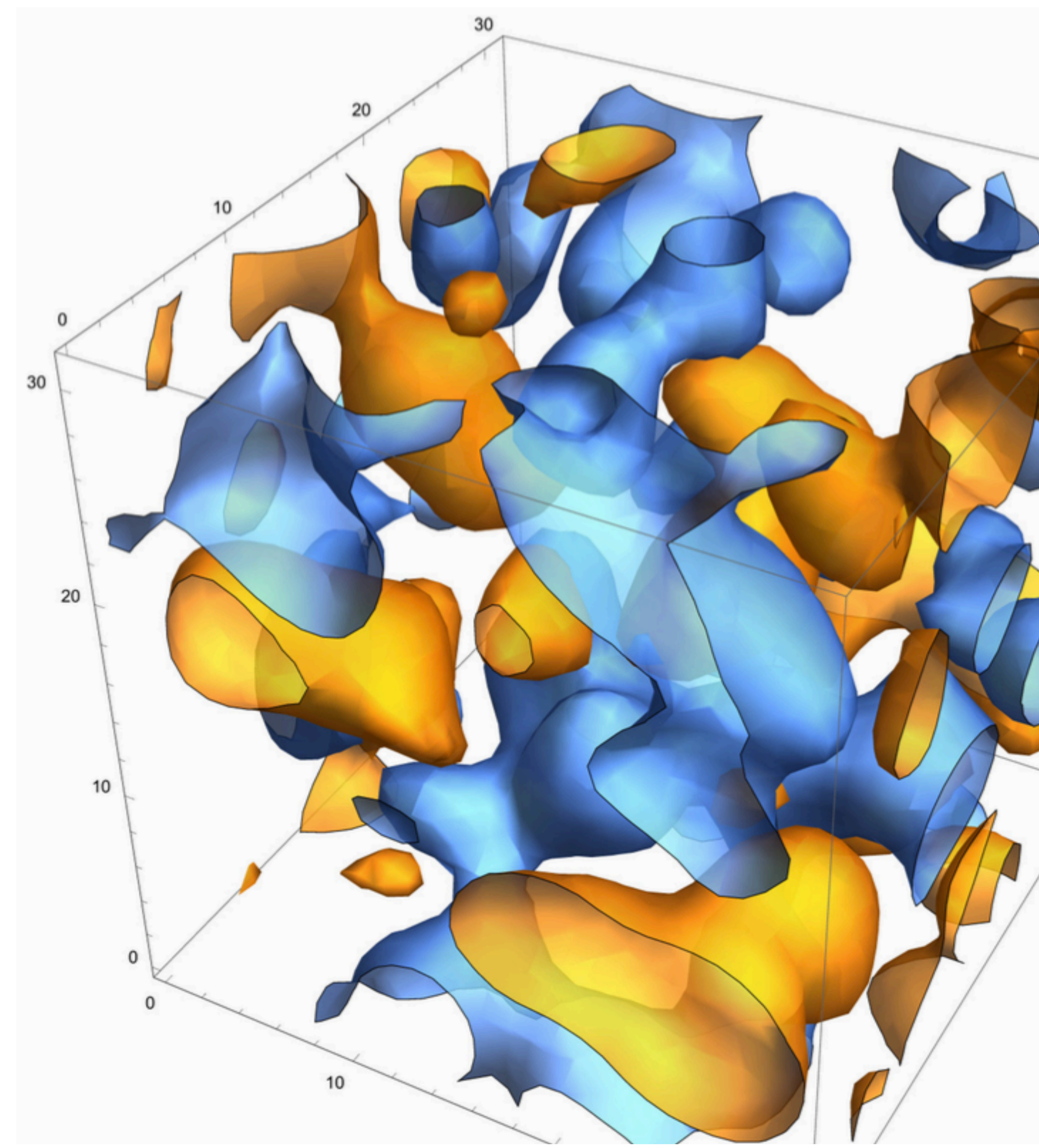
## Accelerate gauge-field generation

- Multi-scale matching PROOF OF PRINCIPLE ✓
- Generative models to replace expensive HMC IN PROGRESS

**Gurtej Kanwar**

**Michael Albergo**

- Learn parameters of a complicated pure-gauge action (cheap) to reproduce action with dynamical fermions (expensive)

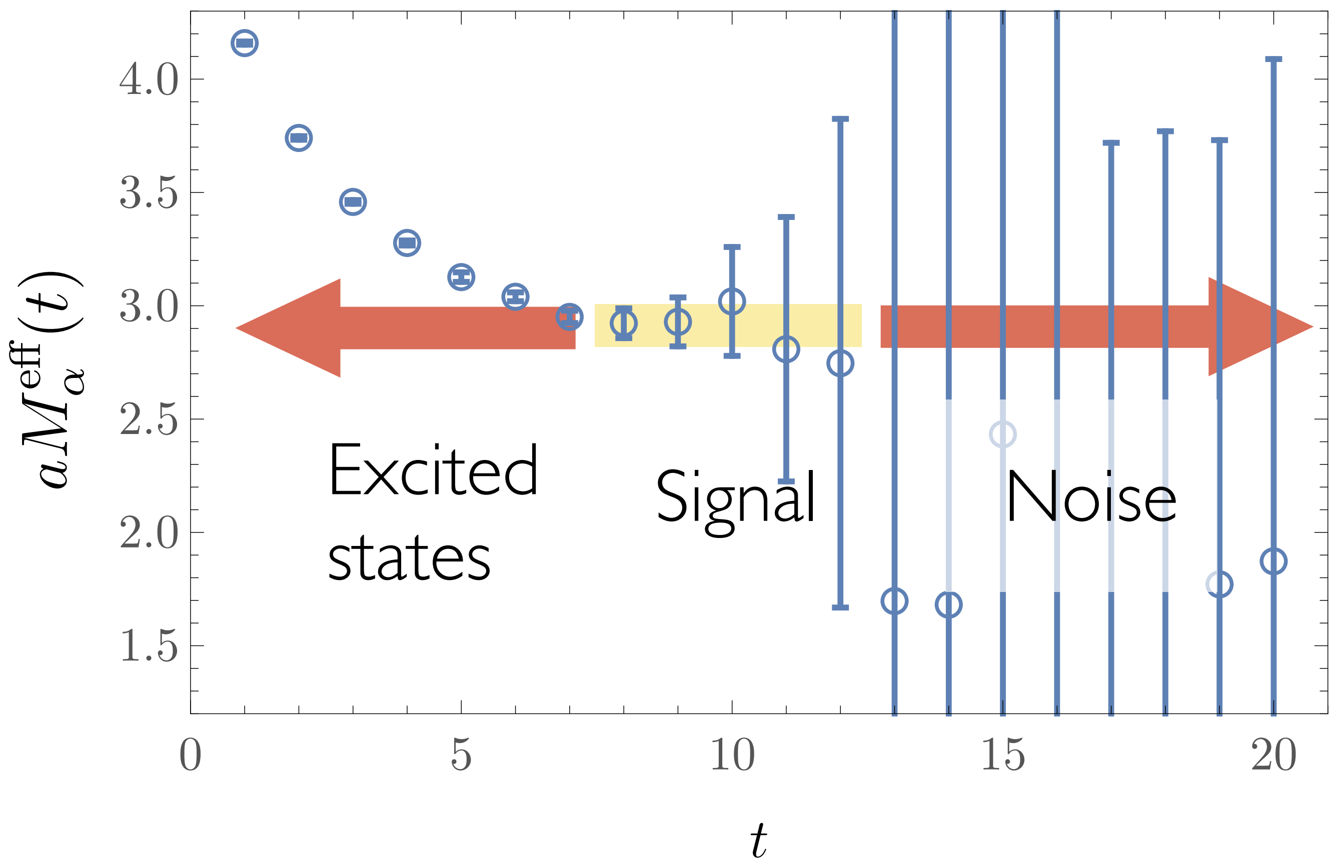


# Machine learning QCD

## 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

**EXPONENTIAL IMPROVEMENTS**




Huge potential to enable first-principles nuclear physics studies