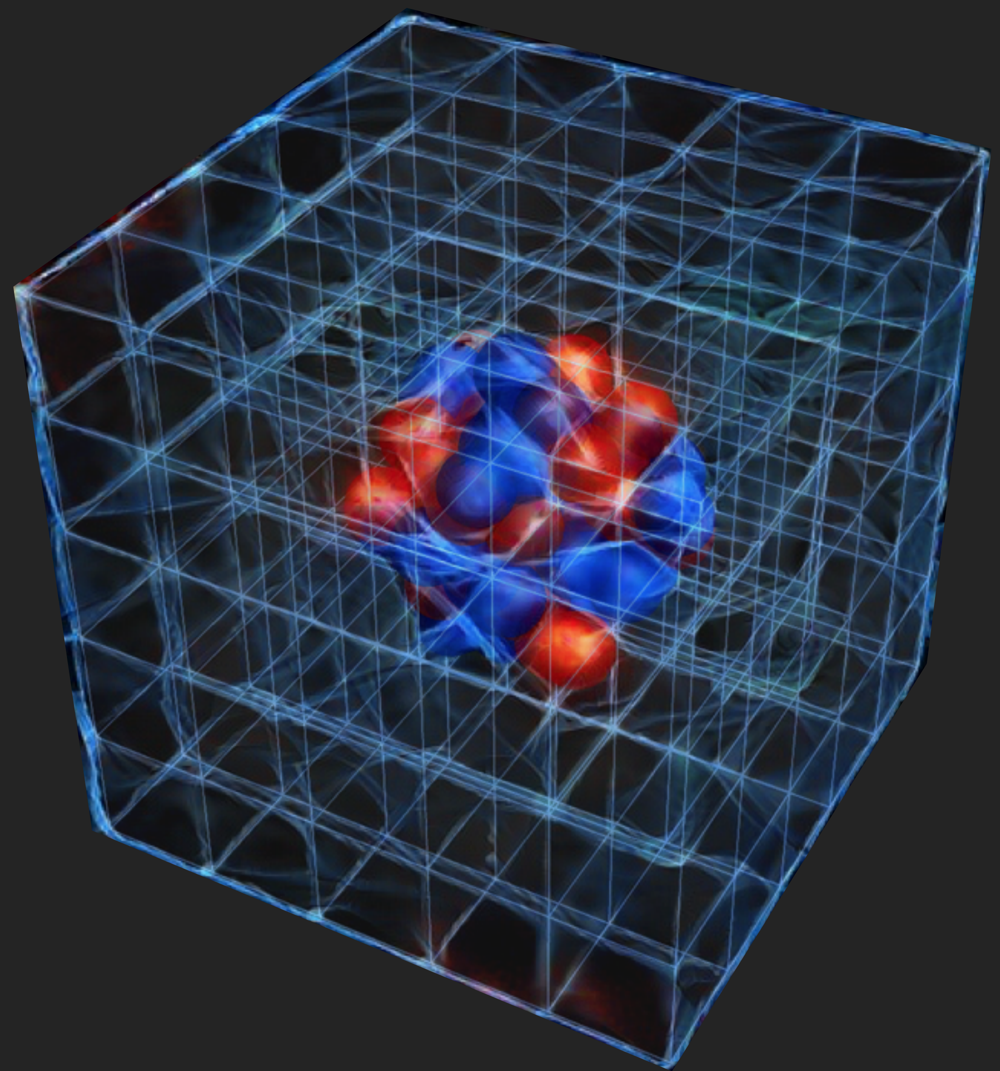
Machine learning for lattice field theory



Phiala Shanahan

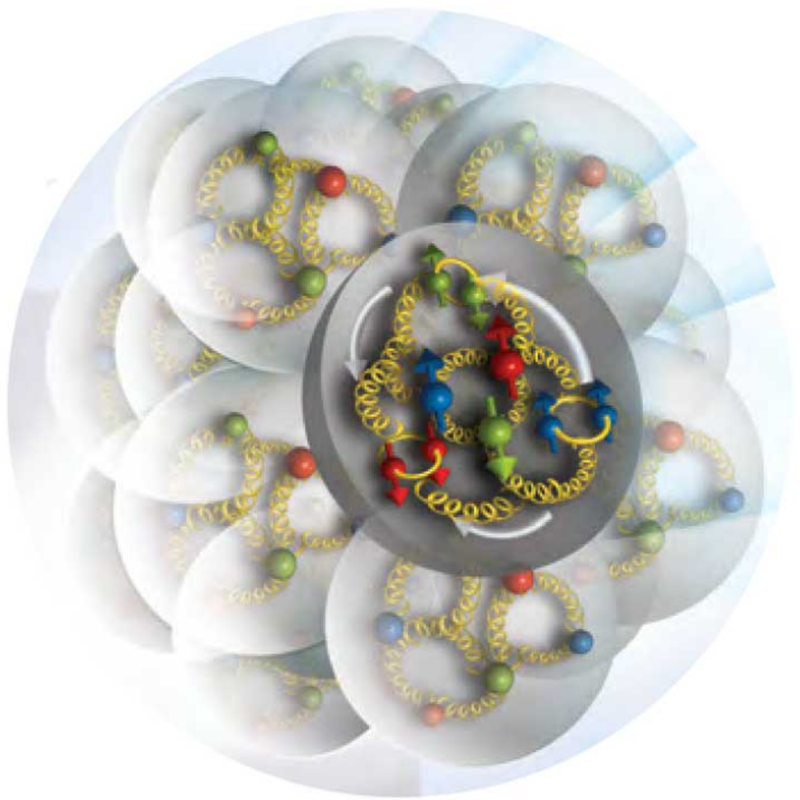
Massachusetts Institute of Technology

Matter from QCD

Hadron and nuclear physics from the Standard Model

Emergence of complex structure in nature

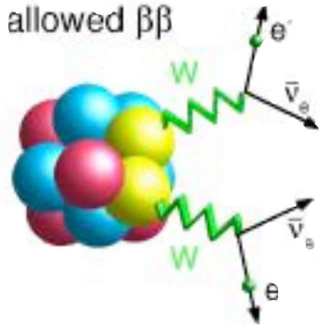
Backgrounds and benchmarks for searches for new physics



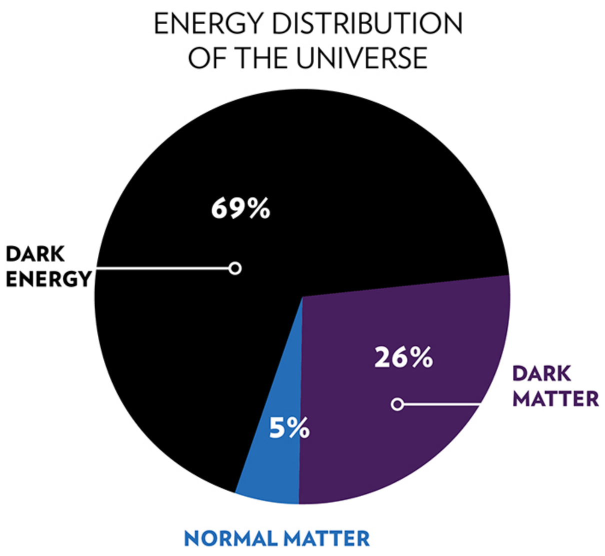
The search for new physics

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…

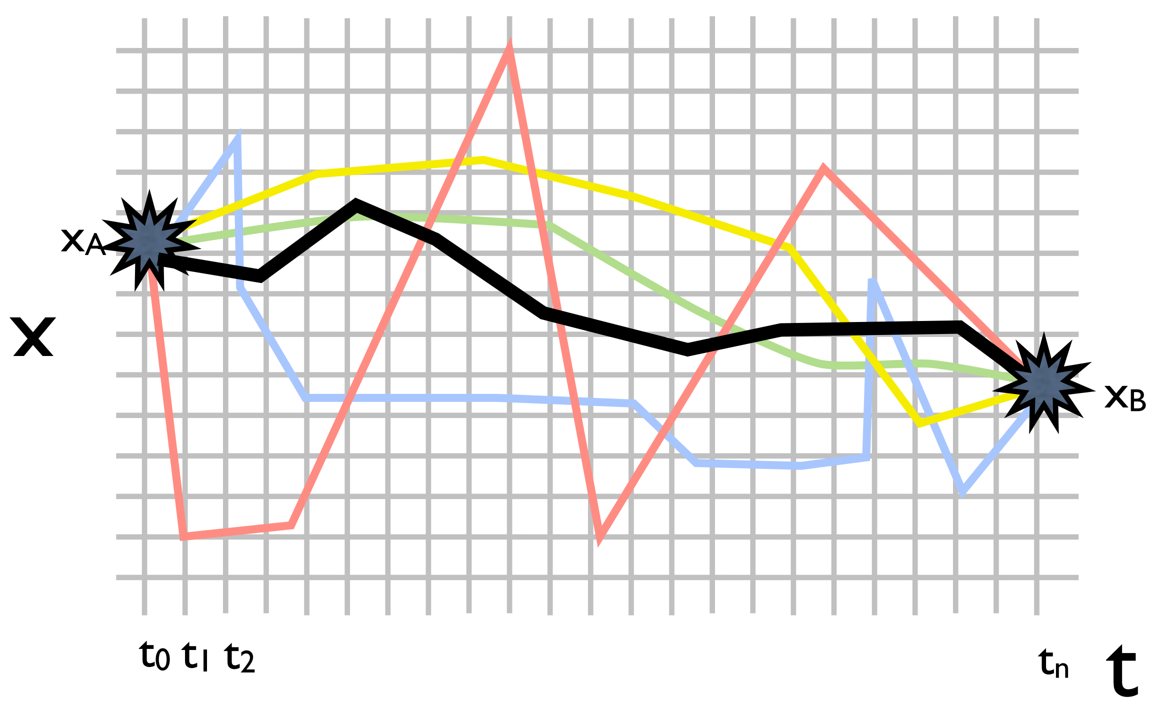


CHALLENGE: understand the physics of nuclei used as targets

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)

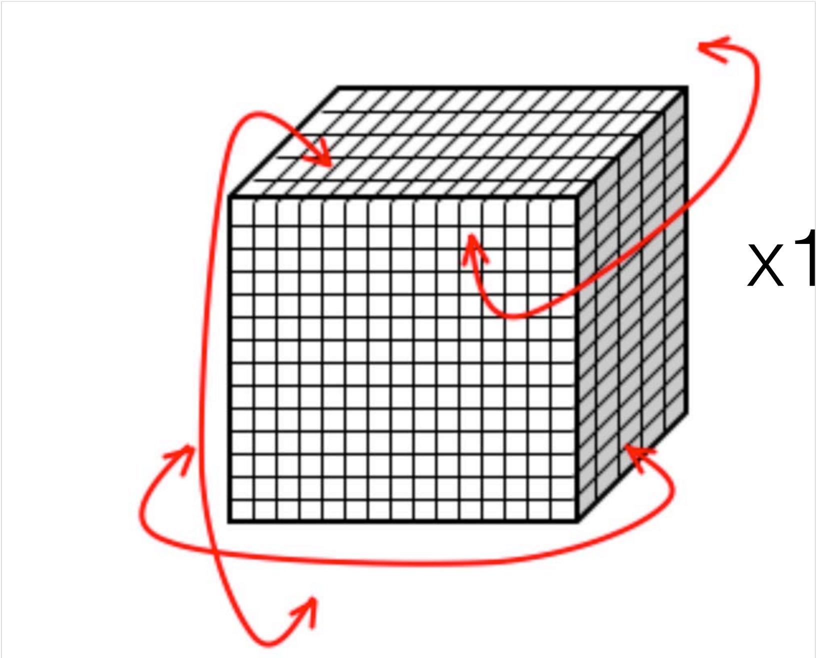


- Evaluate by importance sampling
- Paths near classical action dominate
- Calculate physics on a set (ensemble) of samples of the quark and gluon fields

Lattice QCD

Numerical first-principles approach to
non-perturbative QCD

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



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

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

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

Lattice QCD

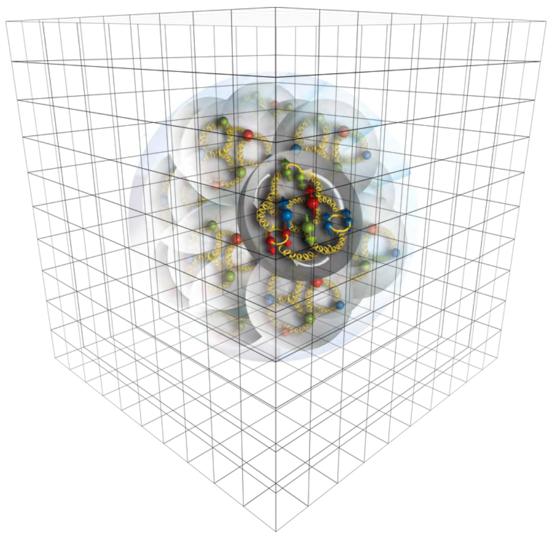
Numerical first-principles approach to
non-perturbative QCD

INPUT

- Lattice QCD action has same free parameters as QCD: quark masses, α_S
- Fix quark masses by matching to measured hadron masses, e.g., π, K, D_s, B_s for u, d, s, c, b
- One experimental input to fix lattice spacing in GeV (and also α_S), e.g., $2S$-$1S$ splitting in Υ, or f_π or Ω mass

OUTPUT

Calculations of all other quantities are QCD predictions



Machine learning for LQCD

MACHINE LEARNING IS

A class of tools for optimising the parameters of **complex models** to describe data

In the context of LQCD, must rigorously account/correct for the effects of modelling in provably exact/unbiased ways

MACHINE LEARNING IS NOT

A black box or model-independent solution to e.g., inverse problems

Applications without **formal quantification and propagation of the effects of modelling, correlations, and systematics**, compromise the rigour of LQCD

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

Machine learning for LQCD

Existing efforts to apply ML tools to many aspects of the lattice QCD workflow

Field configuration generation by e.g.,

- Multi-scale approaches
- Accelerated HMC
- Direct sampling methods
- …

Shanahan et al., Phys.Rev.D 97 (2018)
Albergo et al., Phys.Rev.D 100 (2019)
Rezende et al., 2002.02428 (2020)
Kanwar et al., Phys.Rev.Lett. 125 (2020)
Boyda et al., 2008.05456 (2020)

Tanaka and Tomiya, 1712.03893 (2017)
Zhou et al., Phys.Rev.D 100 (2019)
Li et al., PRX 10 (2020)
Pawlowski and Urban 1811.03533 (2020)
Nagai, Tanaka, Tomiya 2010.11900 (2020)

Efficient computations of correlation functions/observables

Yoon, Bhattacharya, Gupta, Phys. Rev. D 100, 014504 (2019)
Zhang et al, Phys. Rev. D 101, 034516 (2020)
Nicoli et al., 2007.07115 (2020)

Sign-problem avoidance via contour deformation of path integrals

Alexandruet al., Phys. Rev. Lett. 121 (2020),
Detmold et al., 2003.05914 (2020)

Analysis, order parameters, insights

Tanaka and Tomiya, Journal of the Physical Society of Japan, 86 (2017)
Wetzel and Scherzer, Phys. Rev. B 96 (2017)
S. Blùcher et al., Phys. Rev. D 101 (2020)
Boyda et al., 2009.10971 (2020)

*Early developmental stage — many of these papers use toy theories instead of QCD
*Much more related work in e.g., condensed matter context

Generate QCD gauge fields

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

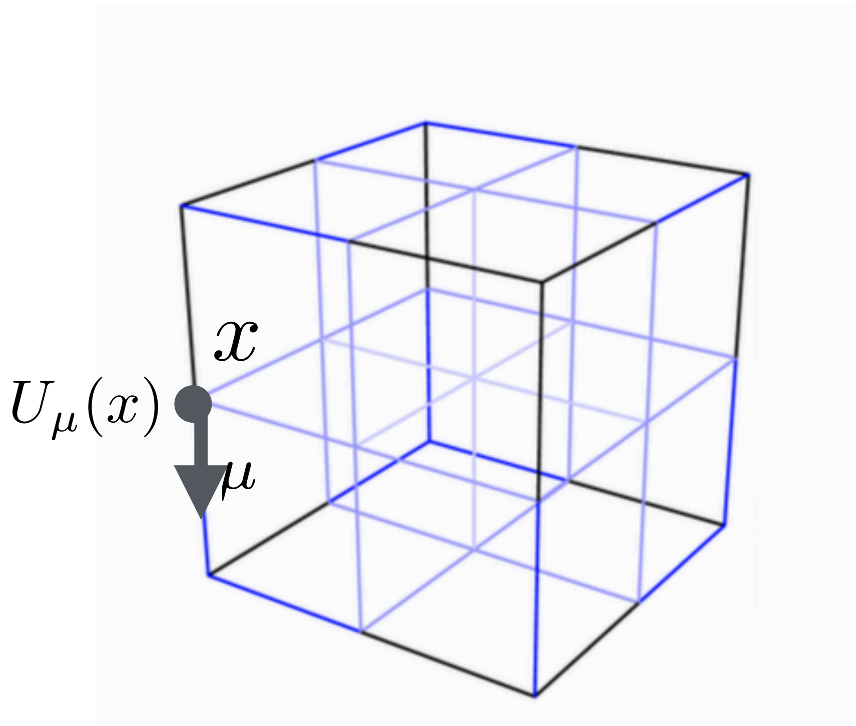
$$P[\phi(x)] \sim e^{-S[\phi(x)]}$$

- Gauge field configurations represented by ~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., ~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 ~10^3 configurations

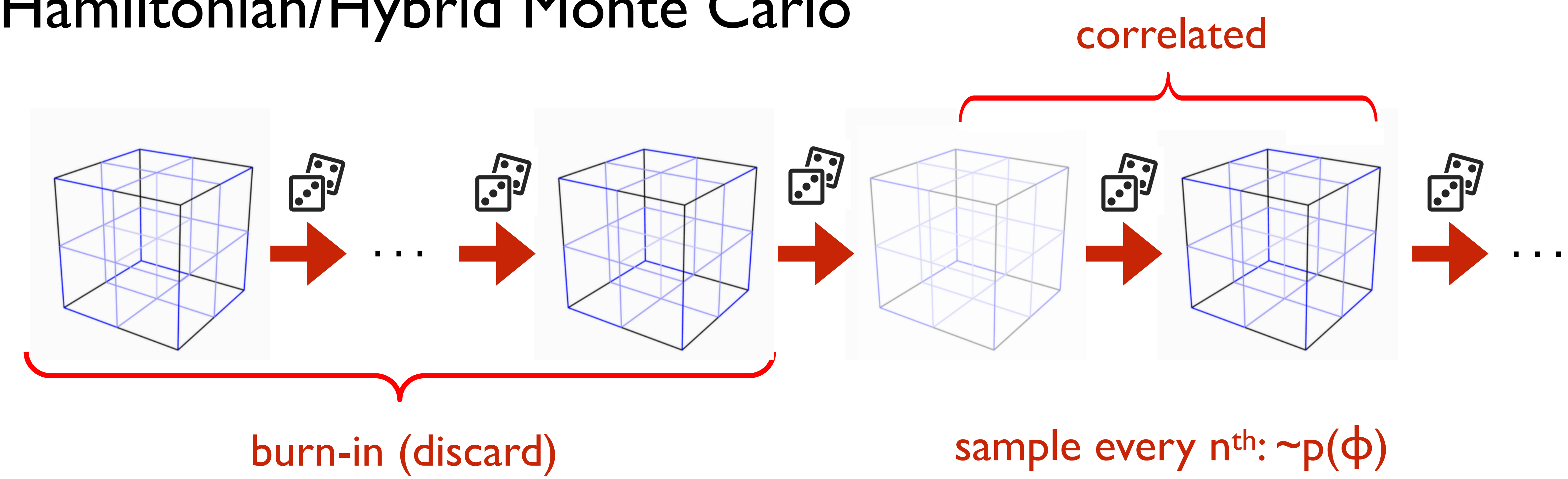


Generate QCD gauge fields

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

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

Hamiltonian/Hybrid Monte Carlo

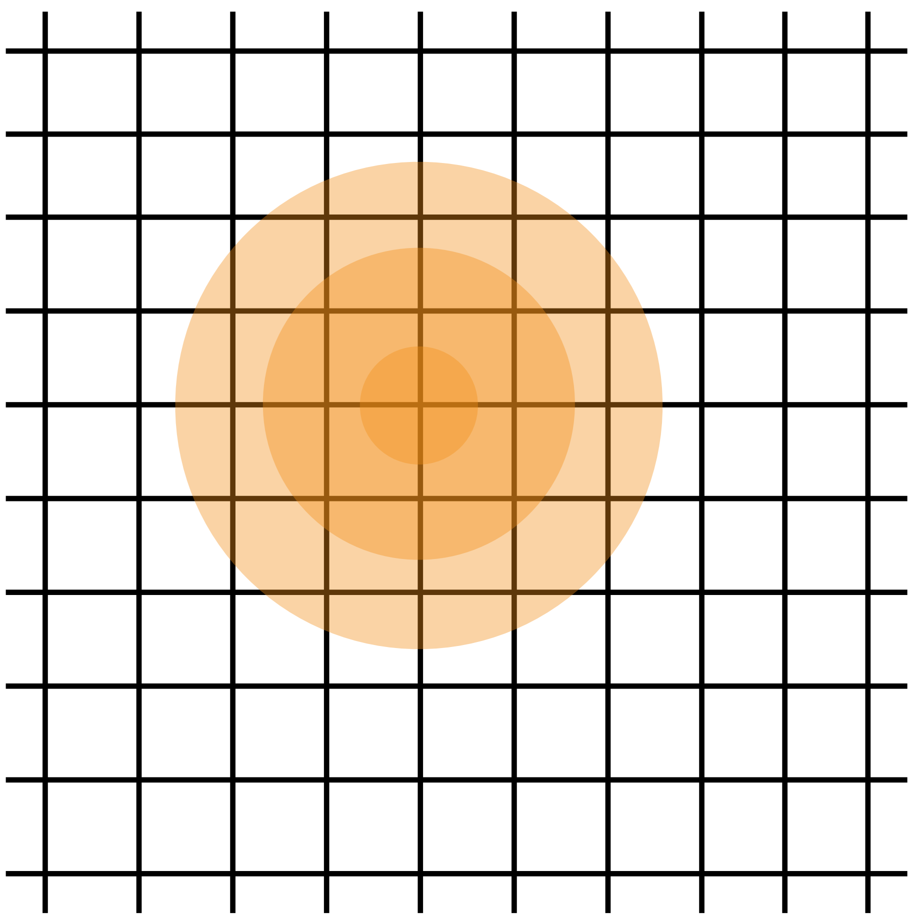


Burn-in time and correlation length dictated by Markov chain **'autocorrelation time'**: shorter autocorrelation time implies less computational cost

Generate QCD gauge fields

QCD gauge field configurations sampled via

Hamiltonian dynamics + Markov Chain Monte Carlo



Updates diffusive

Lattice spacing → 0

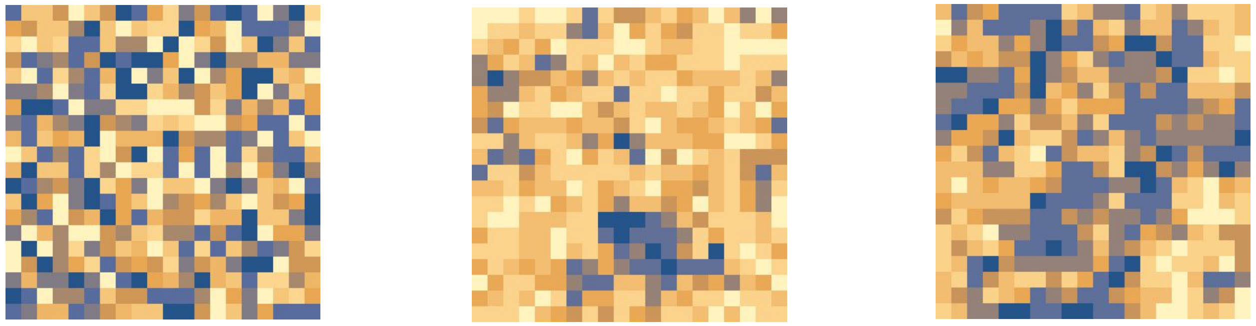
Number of updates to change fixed physical length scale → ∞

“Critical slowing-down” of generation of uncorrelated samples

Scalar lattice field theory

Test case: scalar lattice field theory

- One real number $\phi(x) \in (-\infty, \infty)$ per lattice site x (2D lattice)



- Action: kinetic terms and quartic coupling

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

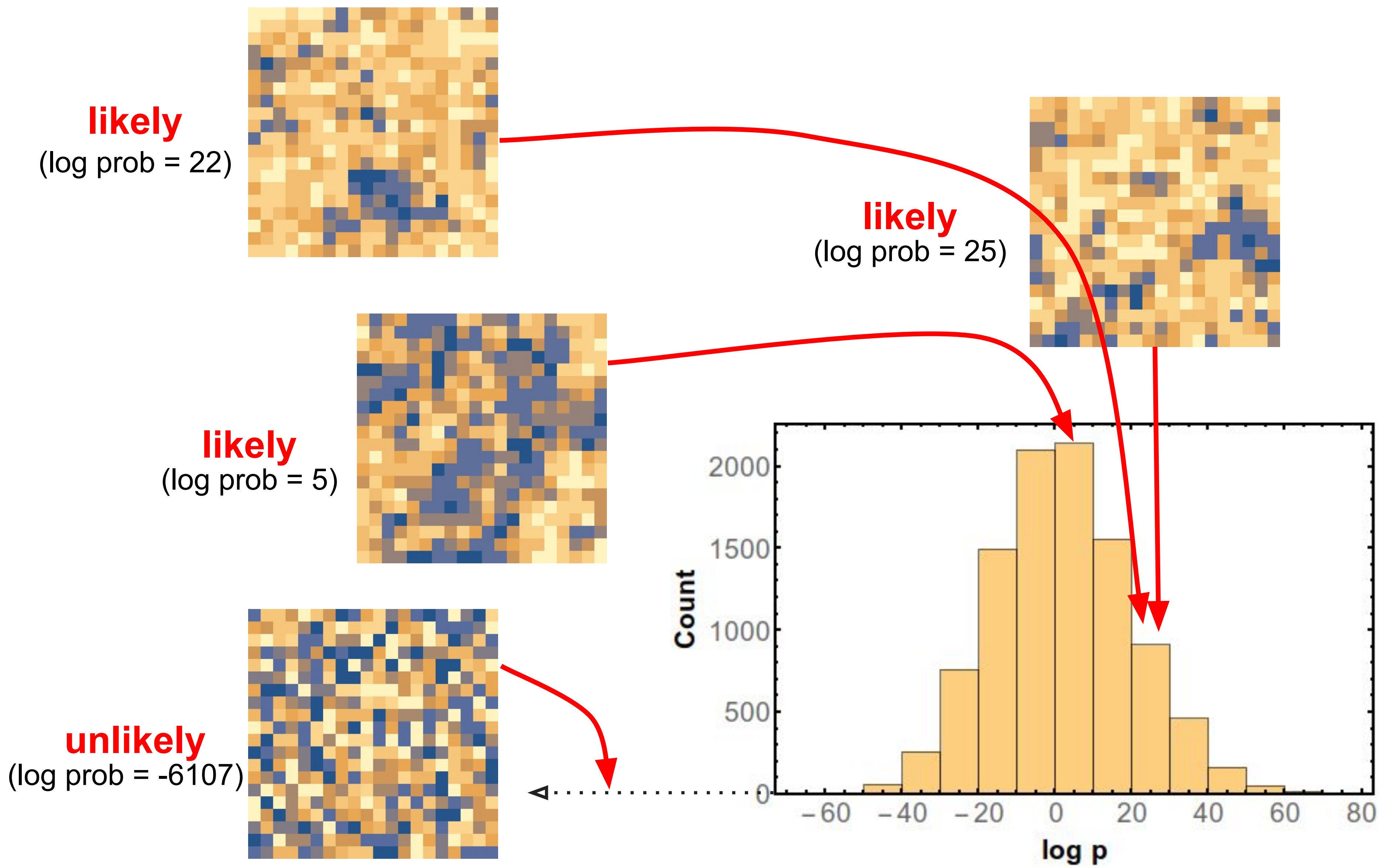
Generate field configurations $\phi(x)$ with probability

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

Sampling gauge field configs

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

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



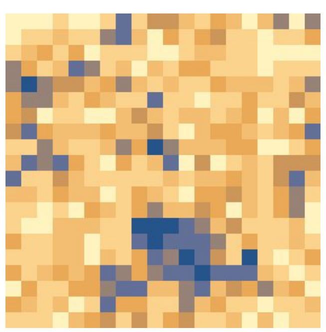
Sampling gauge field configs

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

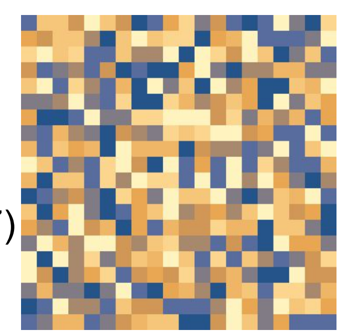
$$P[\phi(x)] \sim e^{-S[\phi(x)]}$$

Parallels with image generation problem

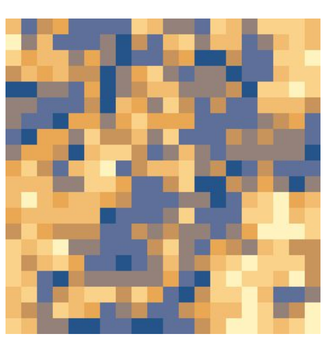
likely
(log prob = 22)



unlikely
(log prob = -6107)



likely
(log prob = 5)



likely

[Karras, Lane, Aila / NVIDIA 1812.04948]

likely

unlikely

Machine learning QCD

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

CIFAR benchmark image set for machine learning

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

Machine learning QCD

Ensemble of lattice QCD gauge fields

- meaning
- Long-distance correlations are important
- Gauge and translation-invariant with periodic boundaries

CIFAR benchmark image set for machine learning

- 32 x 32 pixels x 3 cols ≃3000 numbers
- Translation-invariant within frame

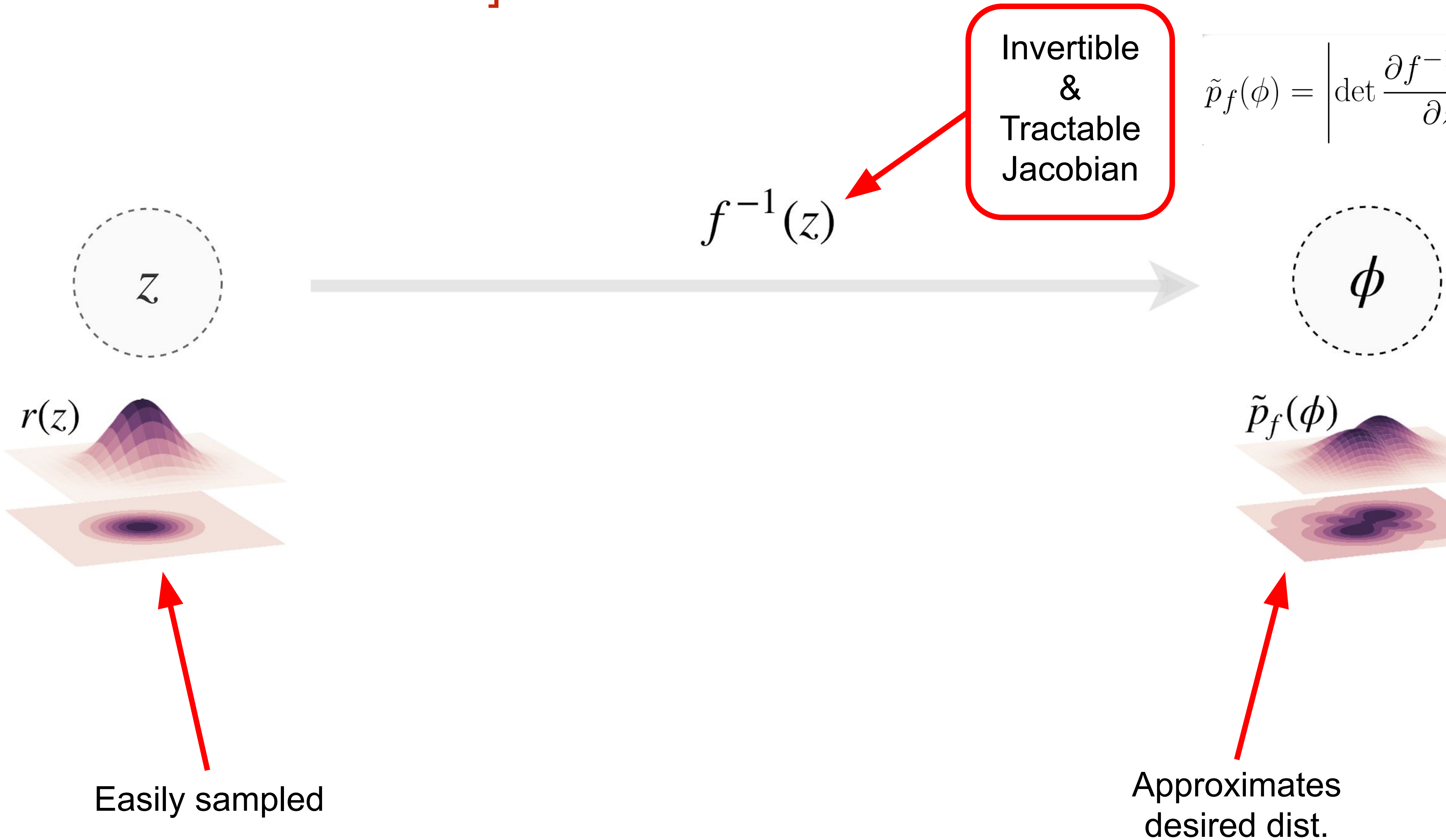
Out-of-the-box ML tools are not appropriate
Need custom ML for physics from the ground up

Generative flow models

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

[Rezende & Mohamed 1505.05770]

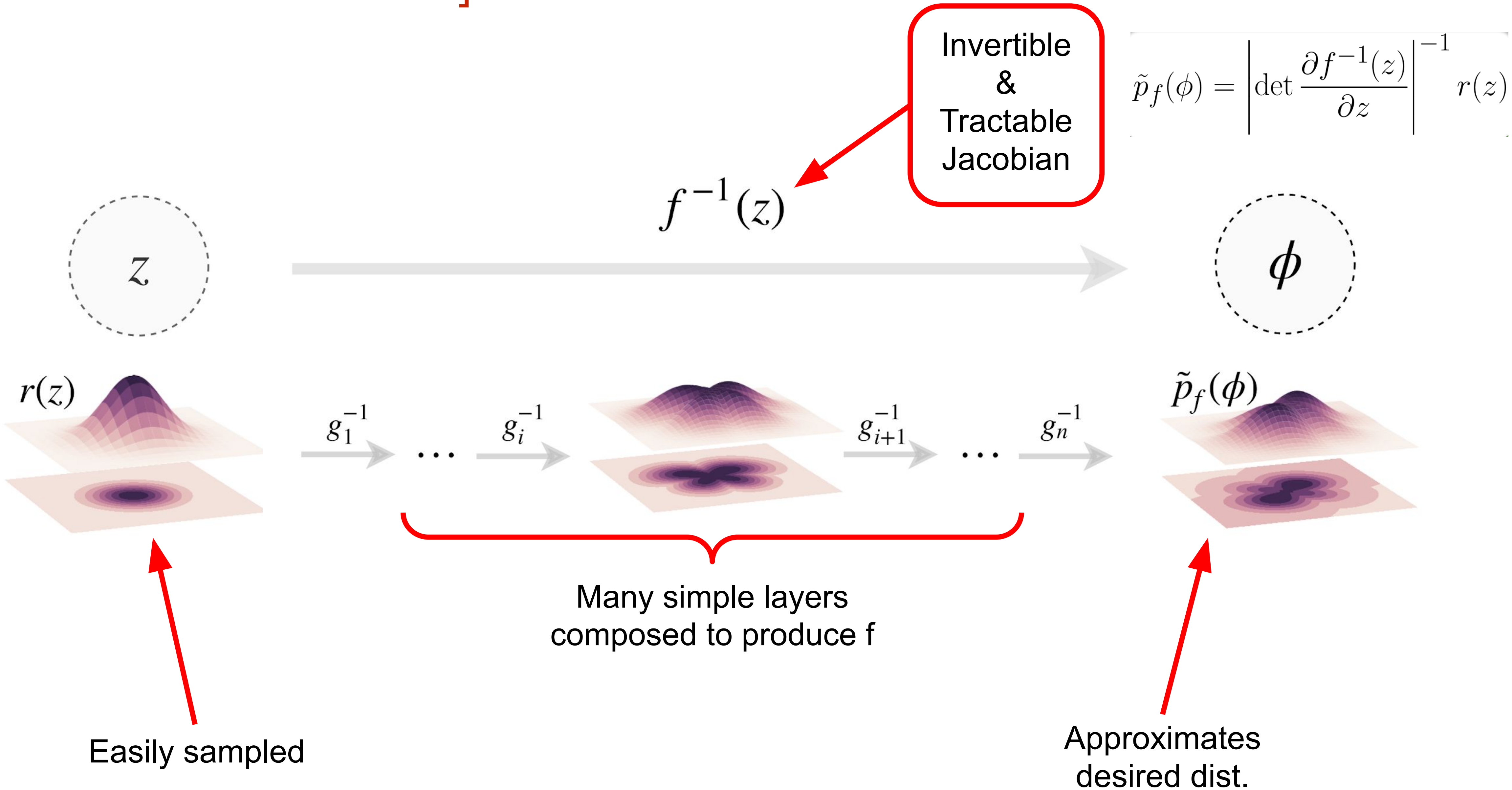
$$\tilde{p}_f(\phi) = \left|\det \frac{\partial f^{-1}(z)}{\partial z}\right|^{-1} r(z)$$



Generative flow models

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

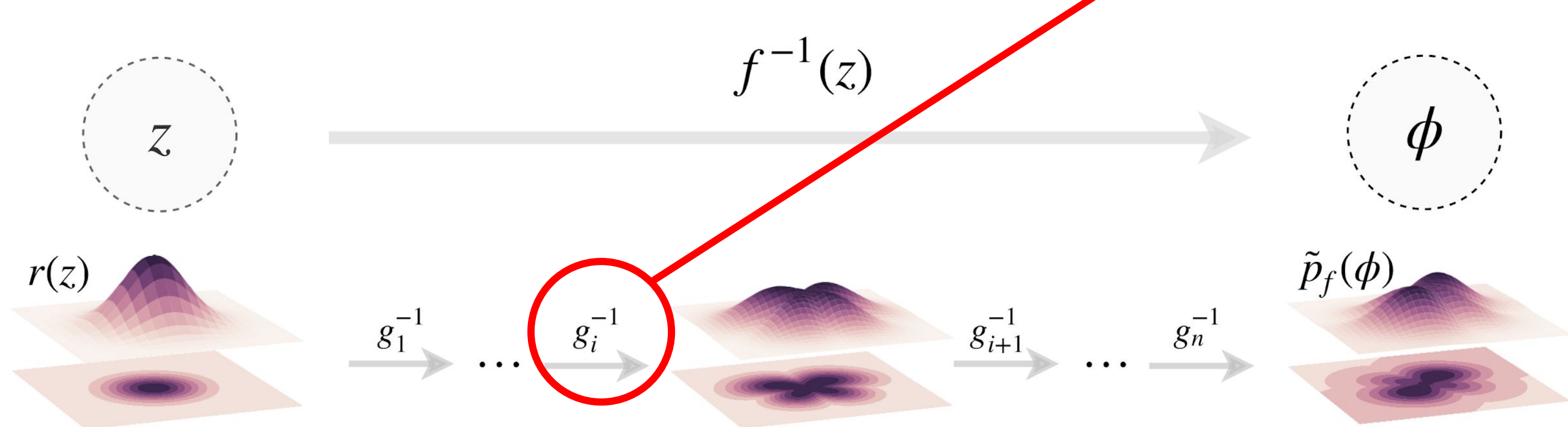
[Rezende & Mohamed 1505.05770]

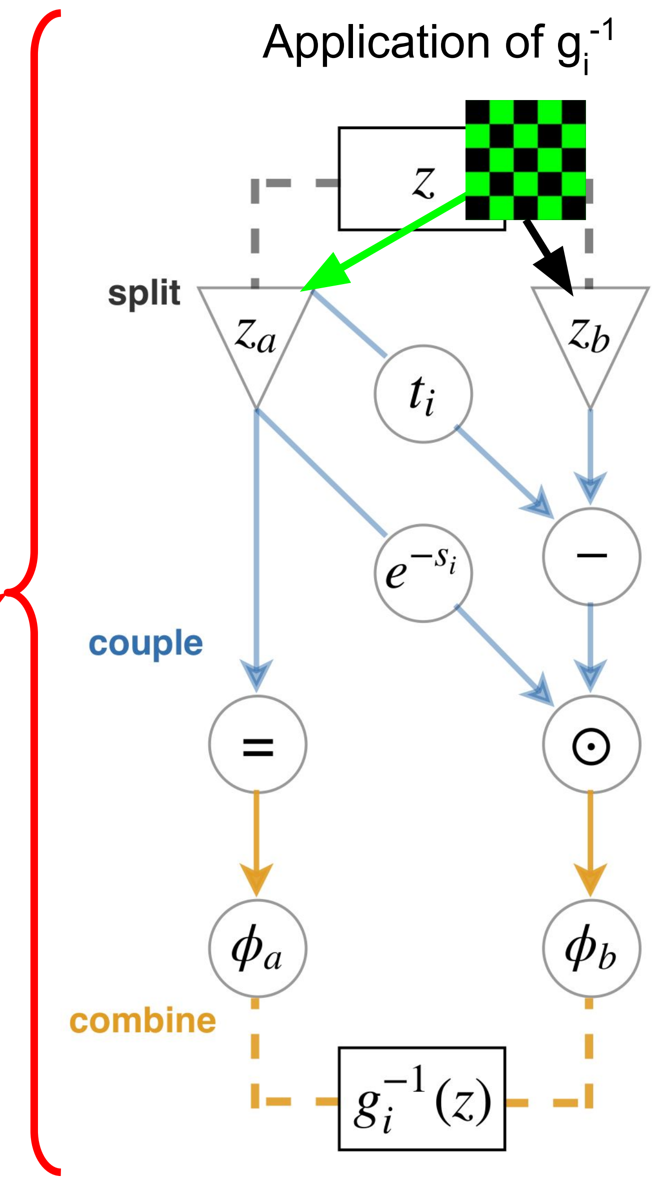


Generative flow models

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





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]

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

shift removes unknown normalisation Z

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

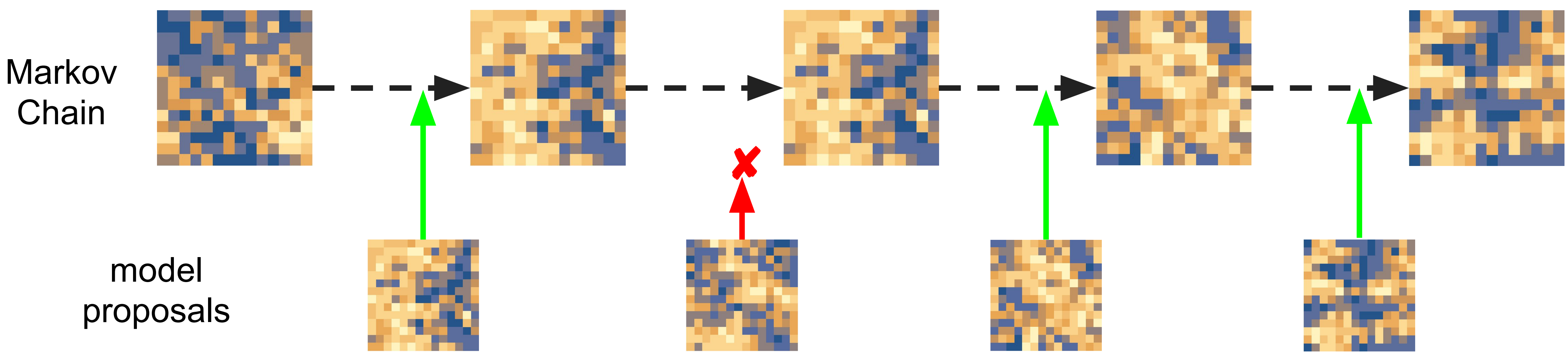
Acceptance probability

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

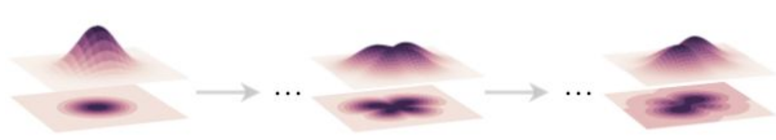
True dist

Model dist

proposal independent of previous sample

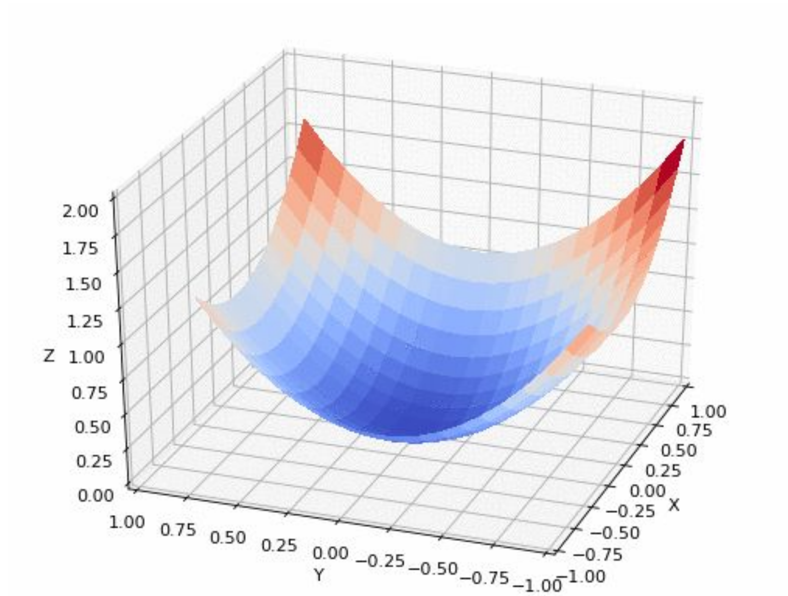


Fields via flow models



Parameterize flow using Real NVP coupling layers

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



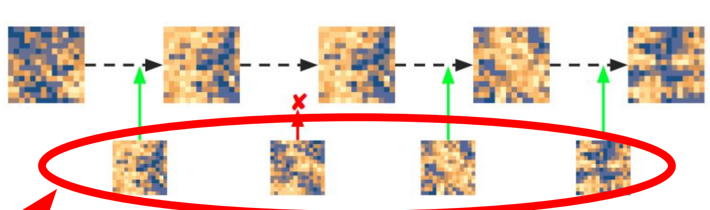
Training step

- Draw samples from model
- Compute loss function
- Gradient descent

Desired accuracy?

Save trained model

Markov chain using samples from model



generating samples is "embarrassingly parallel"

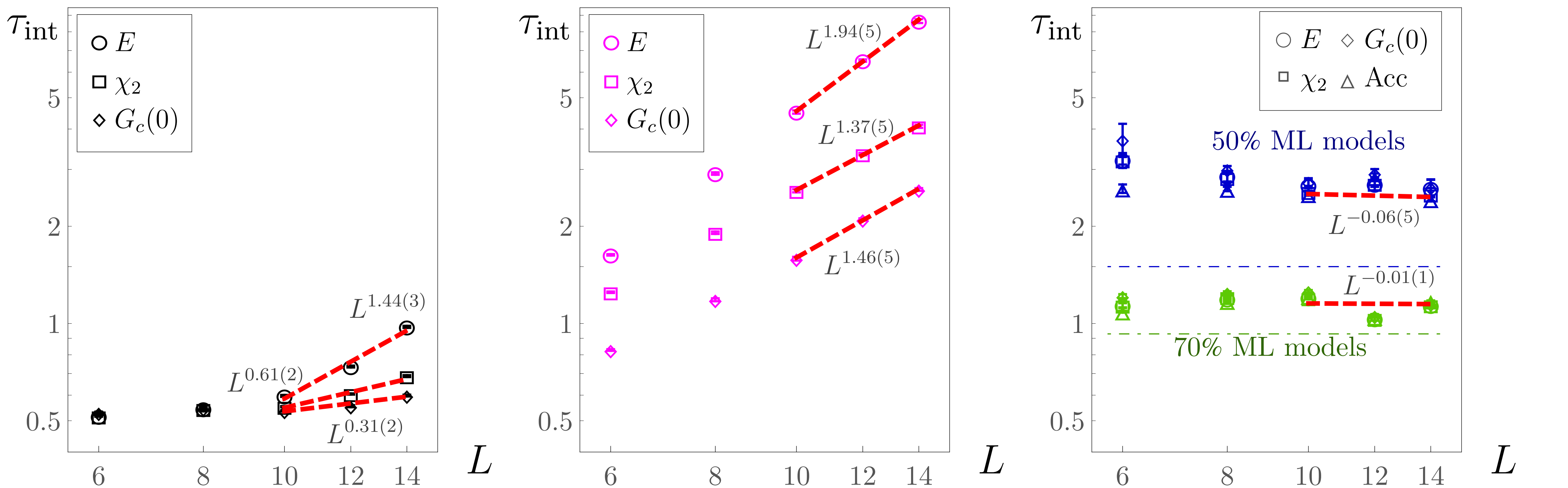
Summary chart: Tej Kanwar

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

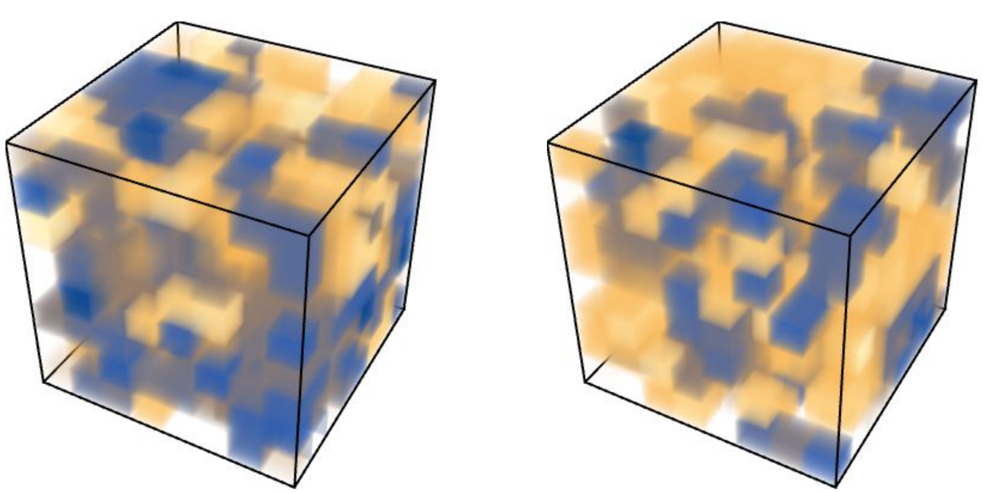
Albergo et al., Phys.Rev.D 100 (2019)

Next steps: ML for LQCD

Target application: Lattice QCD for nuclear physics

1. Scale number of dimensions → 4D
2. Scale number of degrees of freedom → $48^3 \times 96$
3. Methods for gauge theories

[Phys.Rev.D 100 (2019), Phys.Rev.Lett. 125 (2020), 2002.02428, 2008.05456]

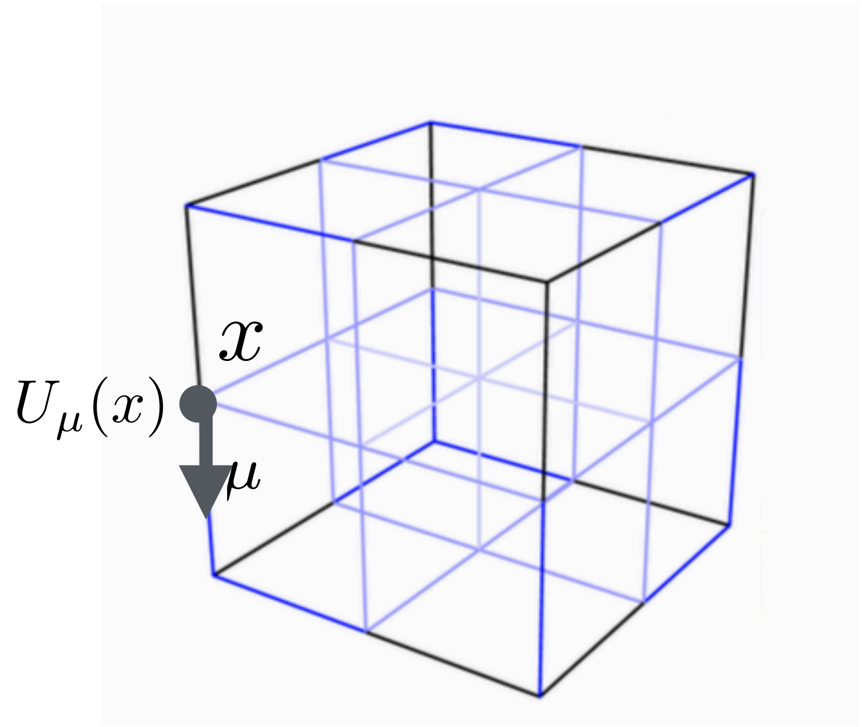


Aurora21 Early Science Project

Incorporating symmetries

Gauge field theories

- Field configurations represented by links $U_\mu(x)$ encoded as matrices
- e.g., for Quantum Chromodynamics, SU(3) matrices (3x3 complex matrices M with $det[M] = 1\,,\ M^{-1} = M^\dagger$)
- Group-valued fields live not on real line but on compact manifolds
- Action is invariant under group transformations on gauge fields



1. **Flows on compact, connected manifolds**
2. **Incorporate symmetries: gauge-equivariant flows**

Incorporating symmetries

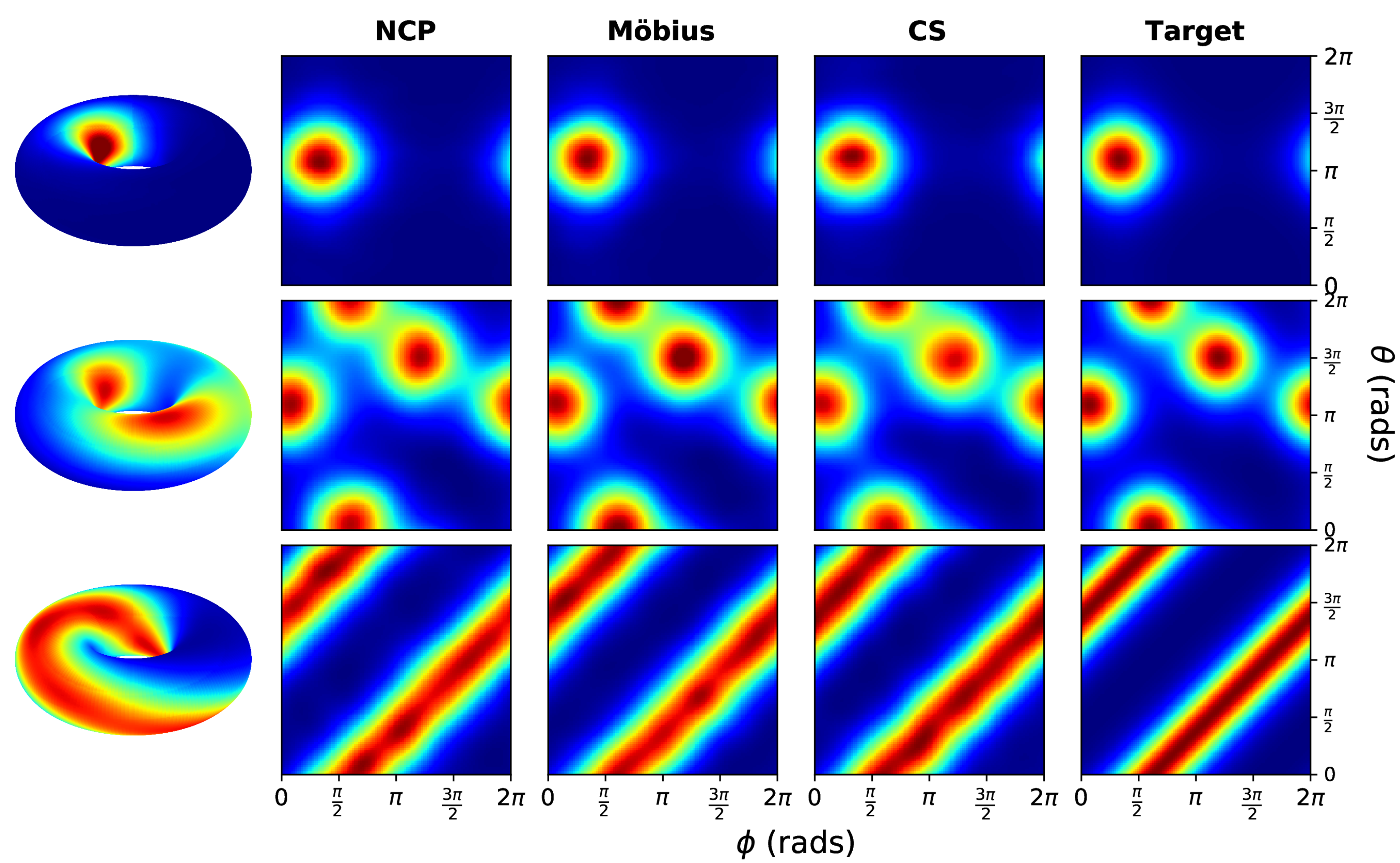
[MIT + Google DeepMind, arXiv:2002.02428]

Normalizing Flows on Tori and Spheres

Danilo Jimenez Rezende [*1] **George Papamakarios** [*1] **Sébastien Racanière** [*1] **Michael S. Albergo** [2]
Gurtej Kanwar [3] **Phiala E. Shanahan** [3] **Kyle Cranmer** [2]

Arbitrarily flexible model architectures designed for compact and connected manifolds

e.g., physics data on compact domains OR robot arm positions



Incorporating symmetries

Incorporating symmetries

- Not essential for correctness of ML-generated ensembles
- BUT: Likely important in training high-dimensional models especially with high-dimensional symmetries

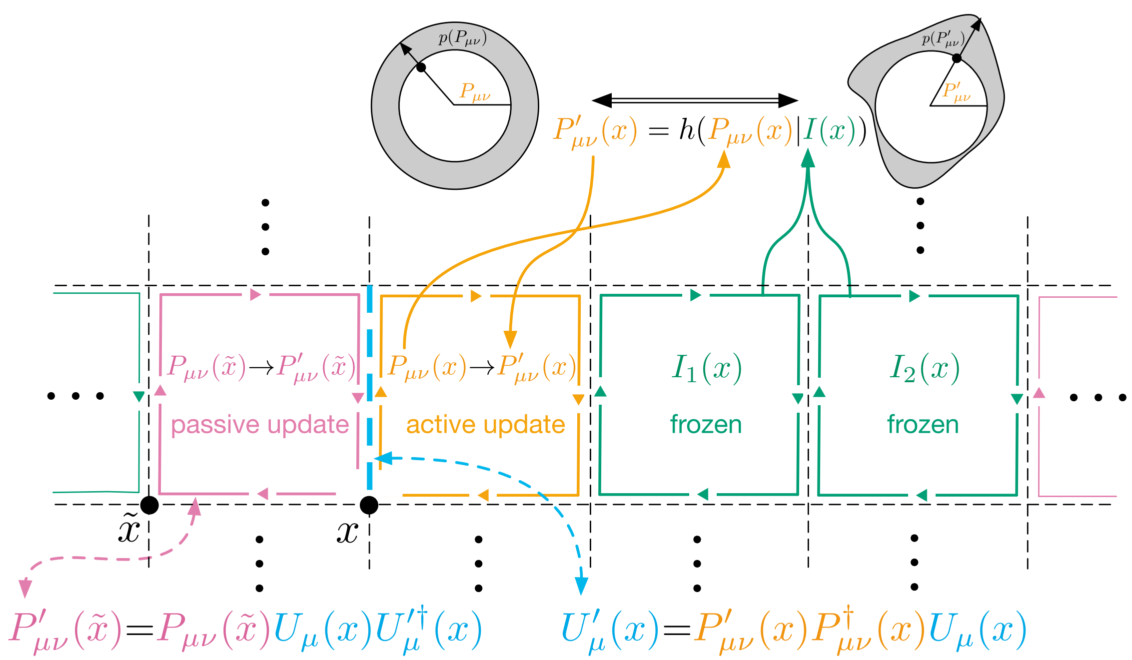
Flow defined from coupling layers will be invariant under symmetry if

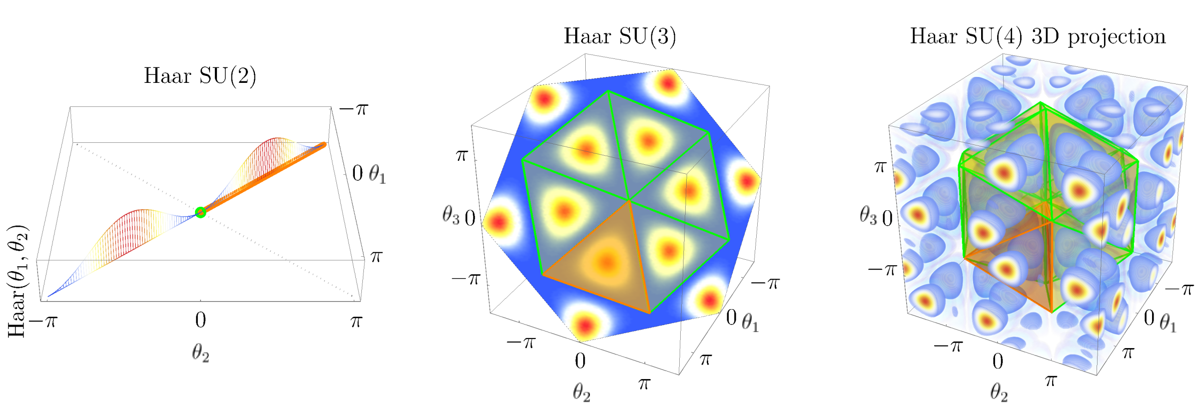
1. **The prior distribution is symmetric**
2. **Each coupling layer is equivariant under the symmetry**
   i.e., all transformations commute through application of the coupling layer

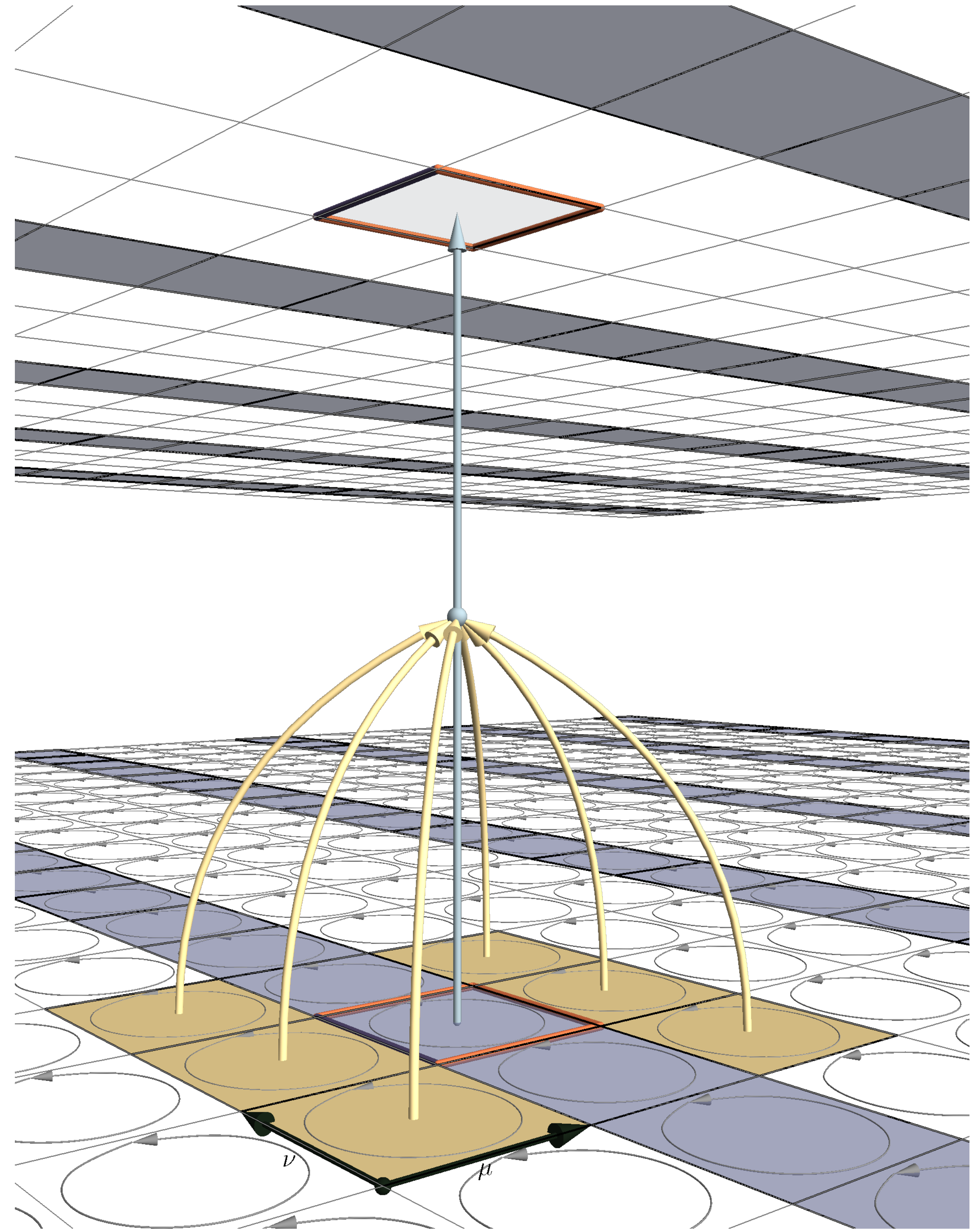
Incorporating symmetries

Generative flow architecture
that is *gauge-equivariant*

Kanwar et al., Phys.Rev.Lett. 125 (2020)
Boyda et al., 2008.05456 (2020)







Application: U(1) field theory

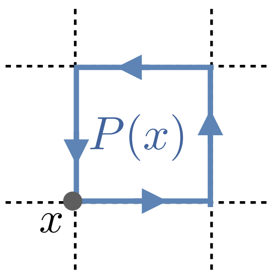
First gauge theory application: U(1) field theory

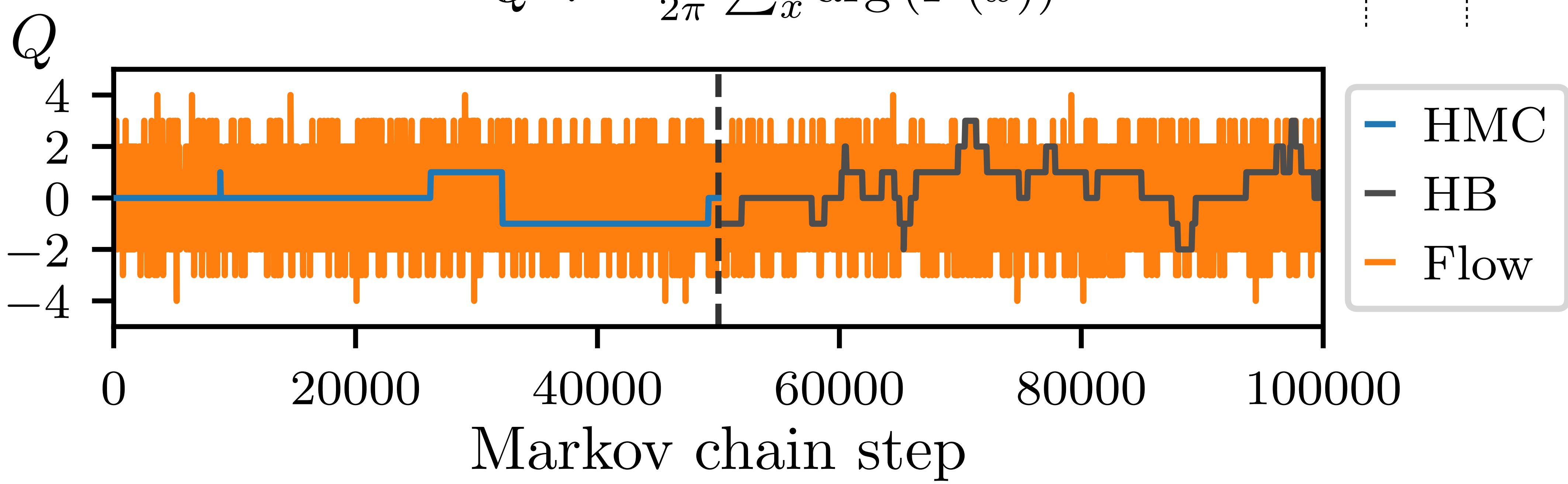
Success: Critical slowing down is significantly reduced

Cost: Up-front training of the model

Sampling of the topological charge

$$Q := \frac{1}{2\pi}\sum_x \arg\left(P(x)\right)$$





2D, L=16, β=6

[Kanwar et al., PRL 125 (2020)]

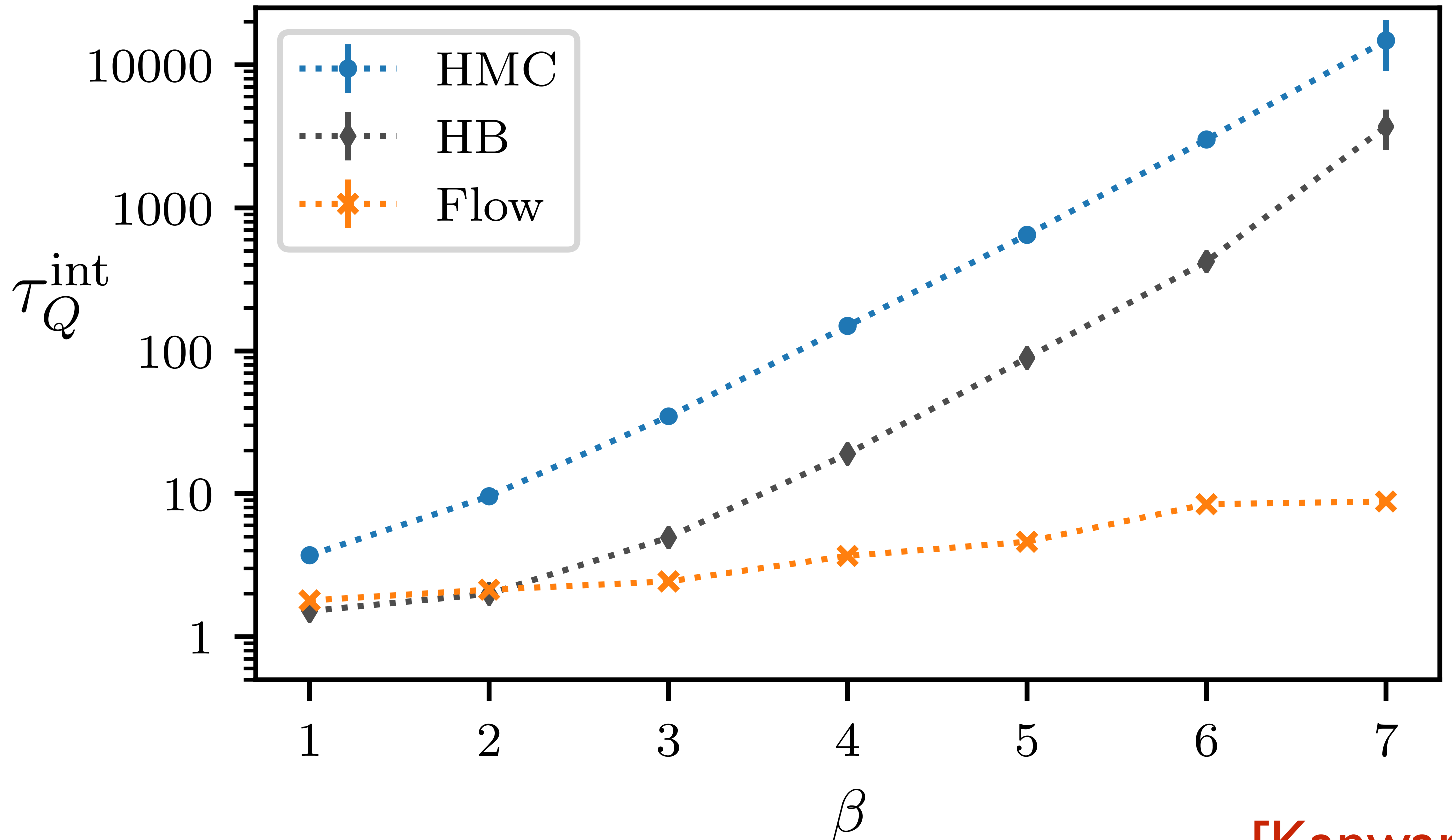
Application: U(1) field theory

First gauge theory application: U(1) field theory

Success: Critical slowing down is significantly reduced

Cost: Up-front training of the model

Integrated autocorrelation time



2D, L=16

[Kanwar et al., PRL 125 (2020)]

Application: U(1) field theory

First gauge theory application: U(1) field theory

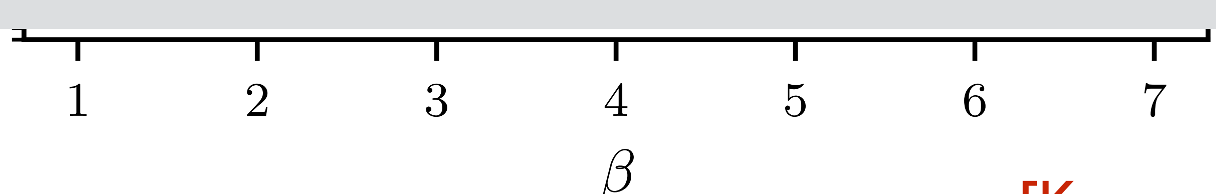
Success: Critical slowing down is significantly reduced

Co

SUCCESS!

Proof-of-principle of efficient, exact, ML algorithm for U(N) and SU(N) LQFT

Significant work required to scale to state-of-the-art

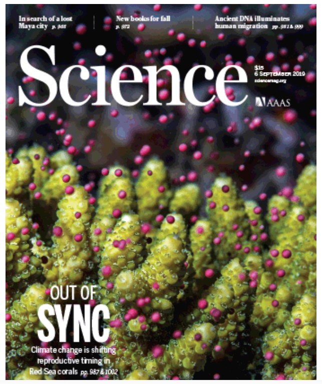


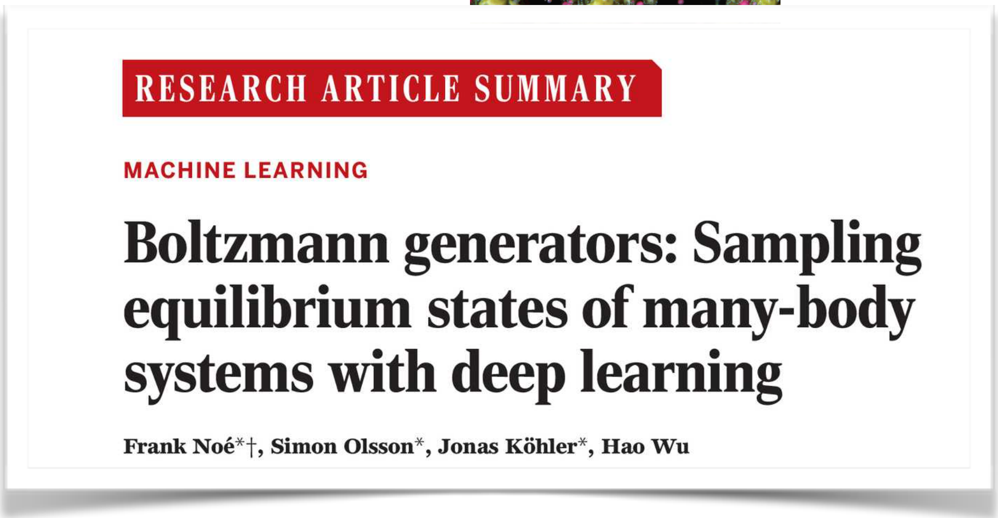
2D, L=16

[Kanwar et al., PRL 125 (2020)]

Interdisciplinary applications

Molecular genetics and drug design



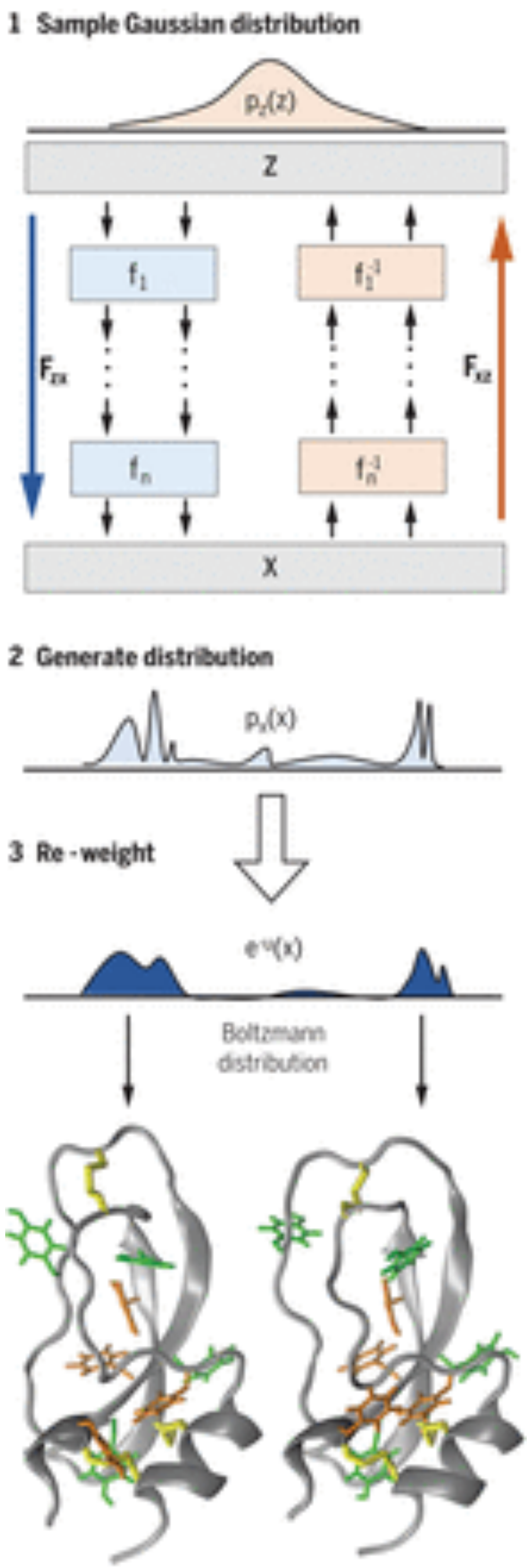


RESEARCH ARTICLE SUMMARY

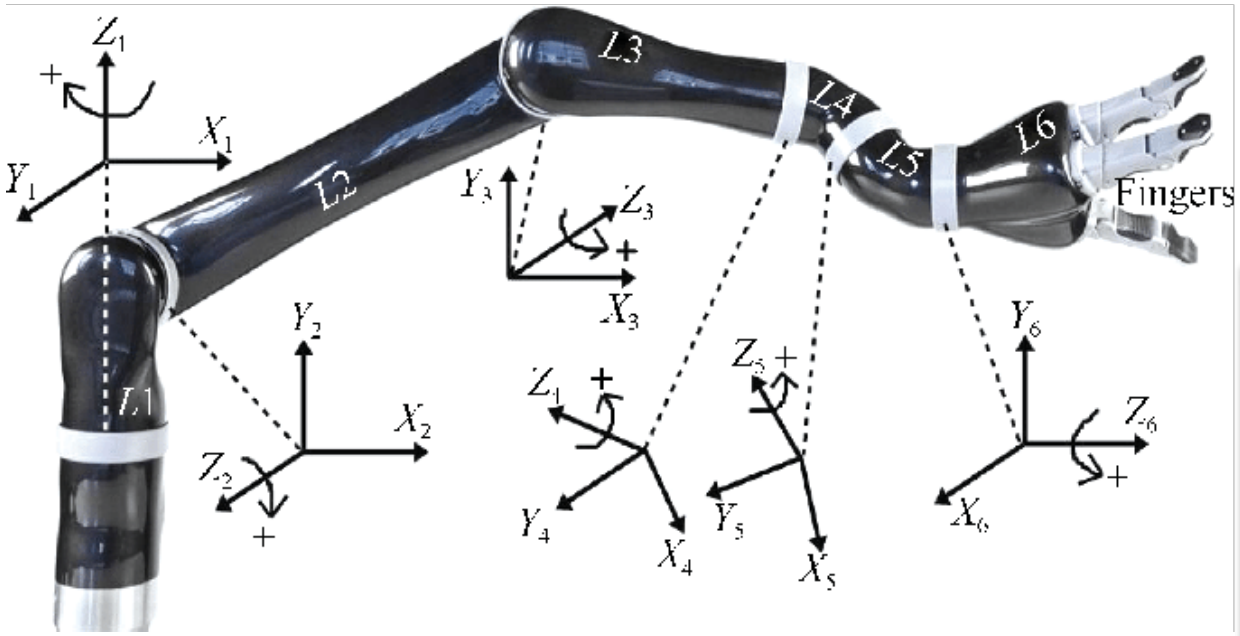
MACHINE LEARNING

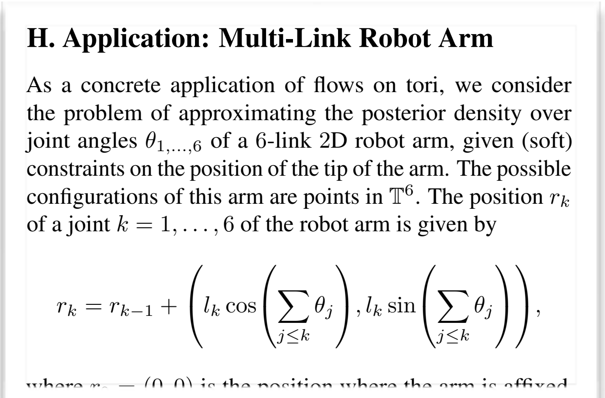
Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning

Frank Noé*†, Simon Olsson*, Jonas Köhler*, Hao Wu



Robotics





H. Application: Multi-Link Robot Arm

As a concrete application of flows on tori, we consider the problem of approximating the posterior density over joint angles $\theta_{1,\ldots,6}$ of a 6-link 2D robot arm, given (soft) constraints on the position of the tip of the arm. The possible configurations of this arm are points in $\mathbb{T}^6$. The position r_k of a joint $k = 1, \ldots, 6$ of the robot arm is given by

$$r_k = r_{k-1} + \left(l_k \cos\left(\sum_{j \leq k} \theta_j \right), l_k \sin\left(\sum_{j \leq k} \theta_j \right) \right),$$

Joint software effort

Our codes exploit and extend existing ML software frameworks

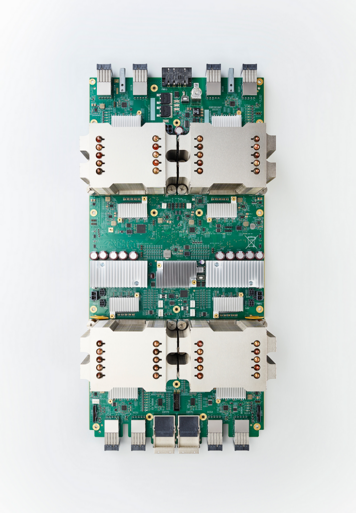
- Tensorflow
- Pytorch
- JAX

Active research projects into training protocols:

- Pruning
- Hyperparameter searches
- Initialisation frameworks
- …

We run on

- CPUs
- GPUs
- TPUs



Targeting exascale hardware for nuclear physics projects

AURORA | EARLY SCIENCE PROGRAM FOR DATA & LEARNIN

Outlook

ML-accelerated algorithms have huge potential to enable first-principles lattice QCD physics studies

Emergence of complex structure in nature

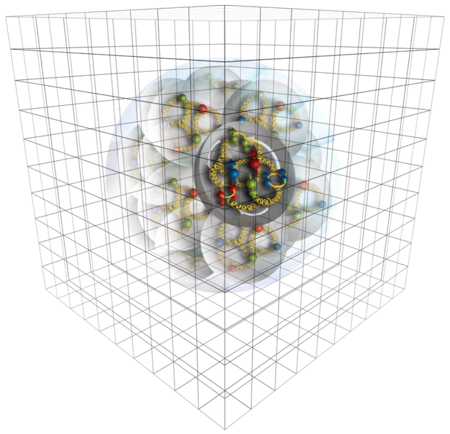
Backgrounds and benchmarks for searches for new physics

Flow-based generation of QCD gauge fields at scale would

- Enable fast, embarrassingly parallel sampling
  → high-statistics calculations
- Allow parameter-space exploration (re-tune trained models)
- Reduce storage challenges (store only model, not samples)

Implementations of flow models at scale (e.g., 4D, $64^3 \times 128$) **conceptually straightforward, but work needed**

- Training paradigms
- Model parallelism
- Exascale-ready implementations
- ….



Outlook

ML-accelerated algorithms have huge potential to enable first-principles lattice QCD physics studies

CAUTION REQUIRED

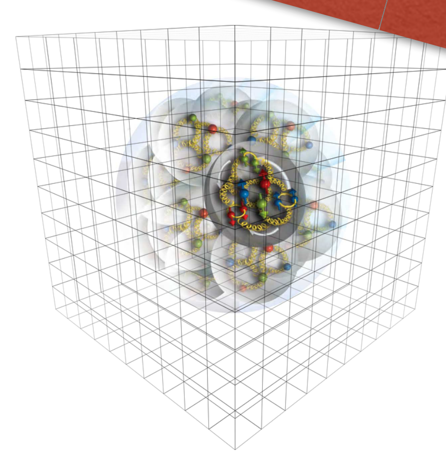
ML can accelerate LQCD while preserving rigour, but out-of-the-box tools aren't always appropriate

F... ...uge fields at scale would

- R...

Implementations of flow mod... conceptually straightforward, but work r...

- Training paradigms
- Model parallelism
- Exascale-ready implementations
- ….



Ab-initio AI Center

The NSF AI Institute for Artificial Intelligence and Fundamental Interactions (IAIFI) *"eye-phi"*

Senior Investigators: 20 Physicists + 7 AI Experts

Junior Investigators: ≈20 PhD Students, ≈7 IAIFI Fellows in steady state

MIT CSAIL

MIT KAVLI INSTITUTE

IDSS

Pulkit Agrawal
Lisa Barsotti
Isaac Chuang
William Detmold
Bill Freeman
Philip Harris
Kerstin Perez
Alexander Rakhlin

Phiala Shanahan
Tracy Slatyer
Marin Soljacic
Justin Solomon
Washington Taylor
Max Tegmark
Jesse Thaler
Mike Williams

Demba Ba
Edo Berger
Cora Dvorkin
Daniel Eisenstein
Doug Finkbeiner
Matthew Schwartz
Yaron Singer
Todd Zickler

James Halverson
Brent Nelson

Taritree Wongjirad

Ab-initio AI Center

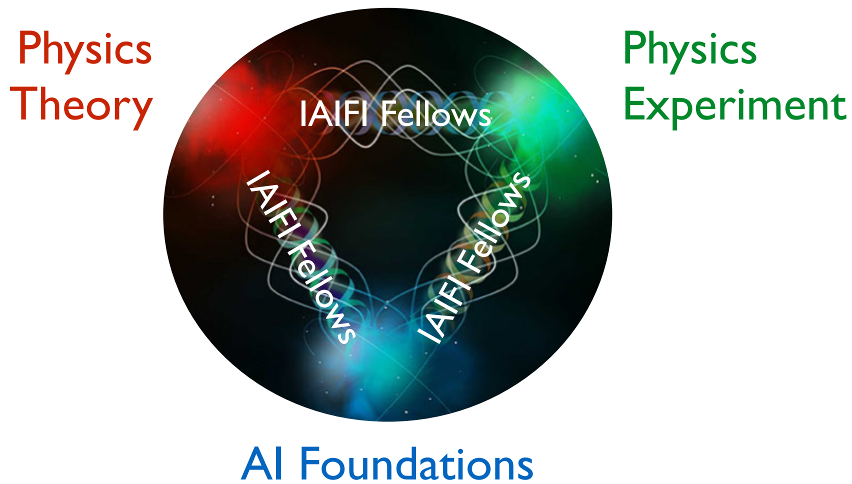
http://iaifi.org/

Machine learning that incorporates
first principles, best practices, and domain knowledge
from fundamental physics

Symmetries, conservation laws, scaling relations, limiting behaviors, locality, causality, unitarity, gauge invariance, entropy, least action, factorization, unit tests, exactness, systematic uncertainties, reproducibility, verifiability, …

Ab-initio AI Center

http://iaifi.org/



AI^2 for Theoretical Physics

- Standard Model of Nuclear & Particle Physics
- String Theory & Physical Mathematics
- Astroparticle Physics
- Automated Discovery of Physics Models

AI^2 for Experimental Physics

- Particle Physics Experiments
- Gravitational Wave Interferometry
- (Multi-Messenger) Astrophysics

AI^2 for Foundational AI

- Symmetries & Invariance
- Speeding up Control & Inference
- Physics-Informed Architectures
- Neural Networks Theory