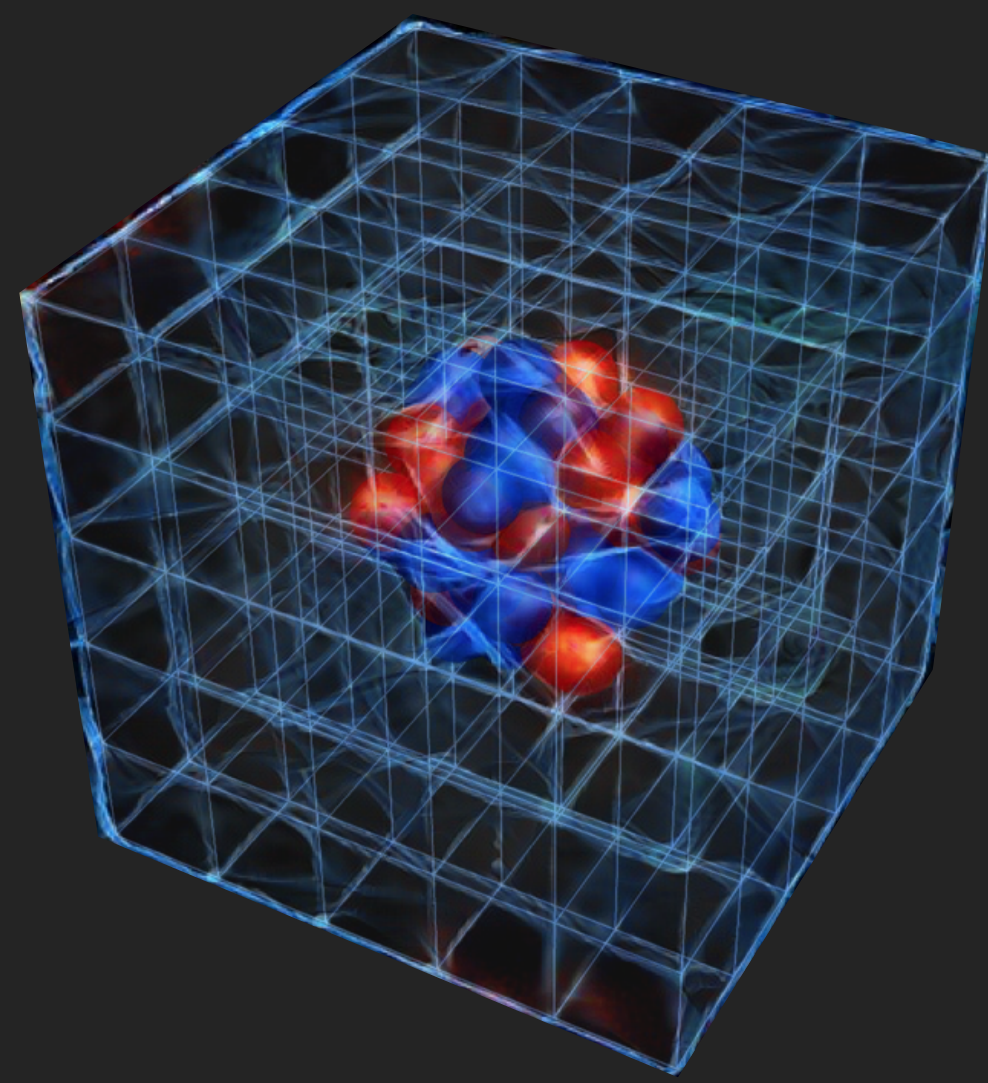
Machine learning for lattice field theory



Phiala Shanahan

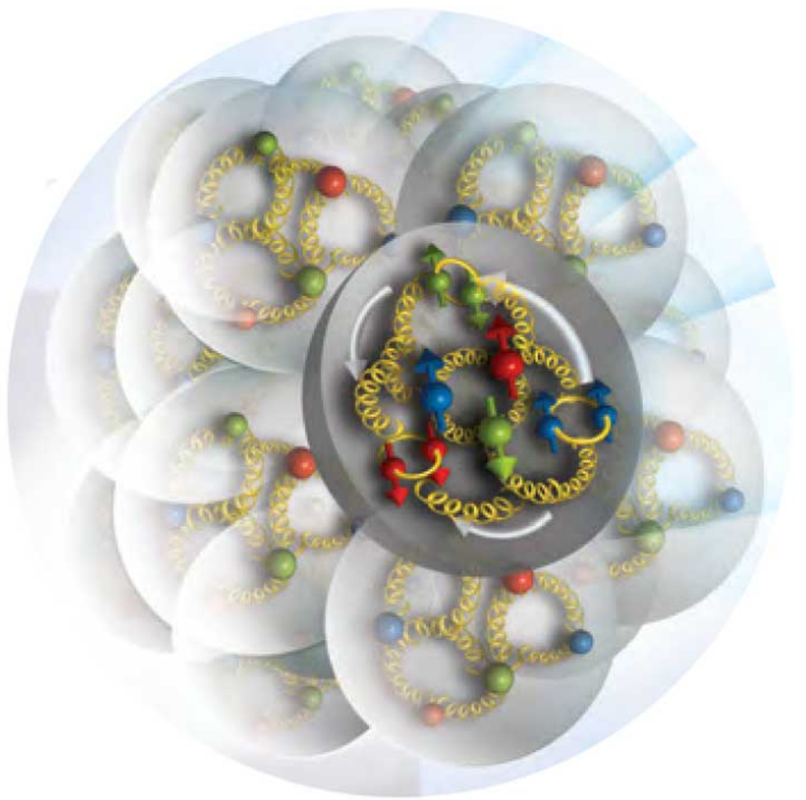
Massachusetts Institute of Technology

The structure of matter

Nuclear structure from the Standard Model

Emergence of complex structure in nature

Backgrounds and benchmarks for searches for new physics

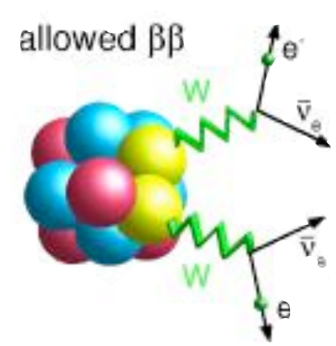


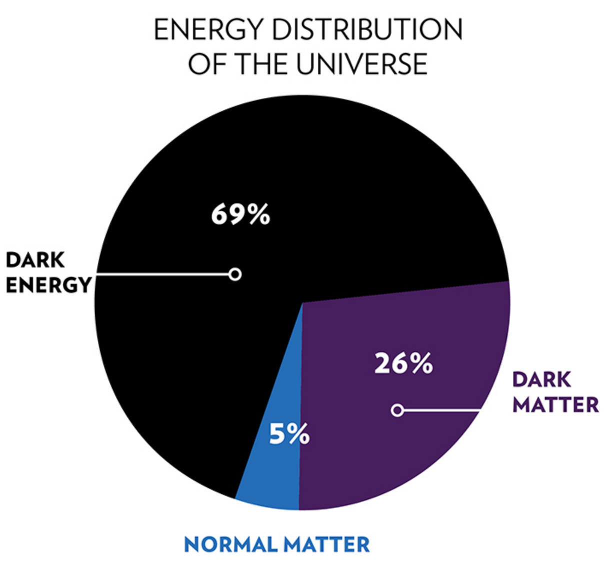
The search for new physics

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…





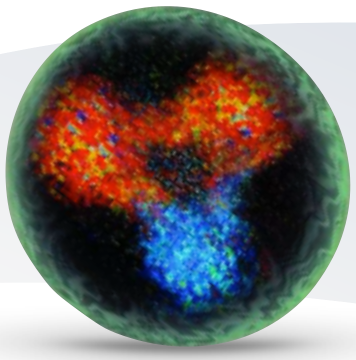
CHALLENGE: **understand the physics of nuclei used as targets**

Strong interactions

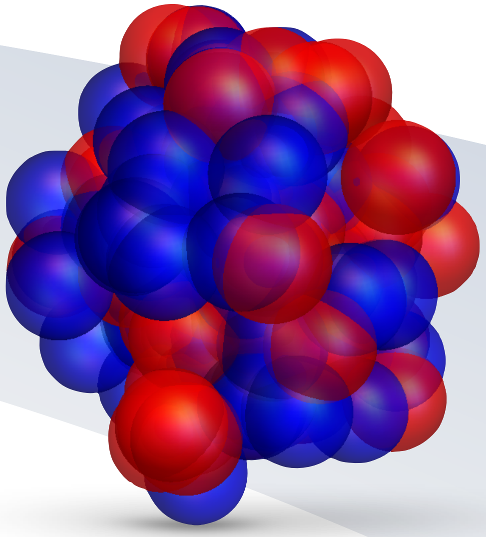
Study nuclear structure from the strong interactions

Quantum Chromodynamics (QCD)

Strongest of the four forces in nature

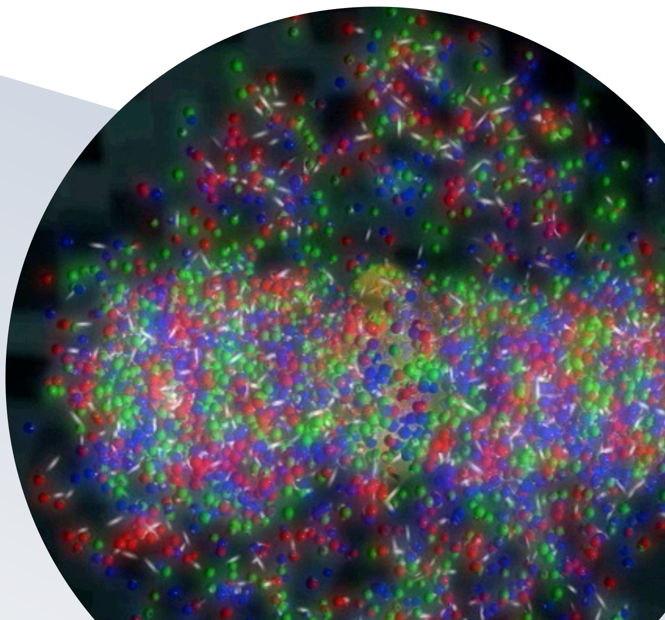


Binds quarks and gluons into protons, neutrons, pions etc.



Binds protons and neutrons into nuclei

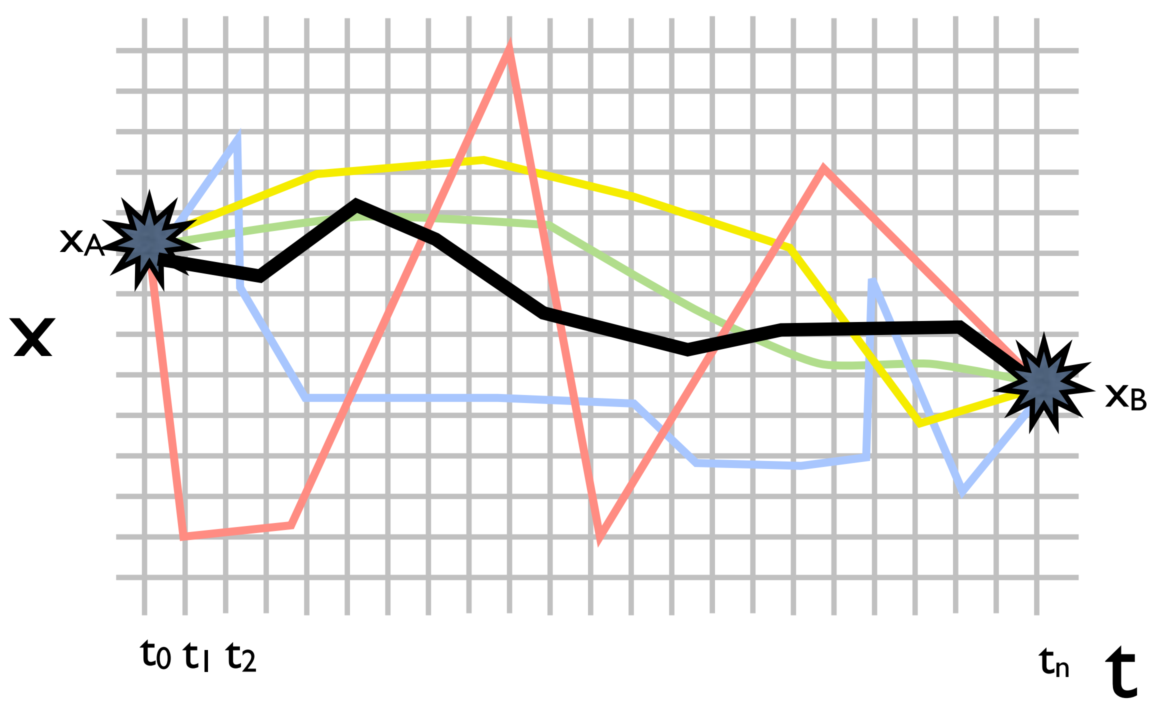
Forms other types of exotic matter e.g., quark-gluon plasma



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)
- $\sim 10^{12}$ variables (for state-of-the-art)

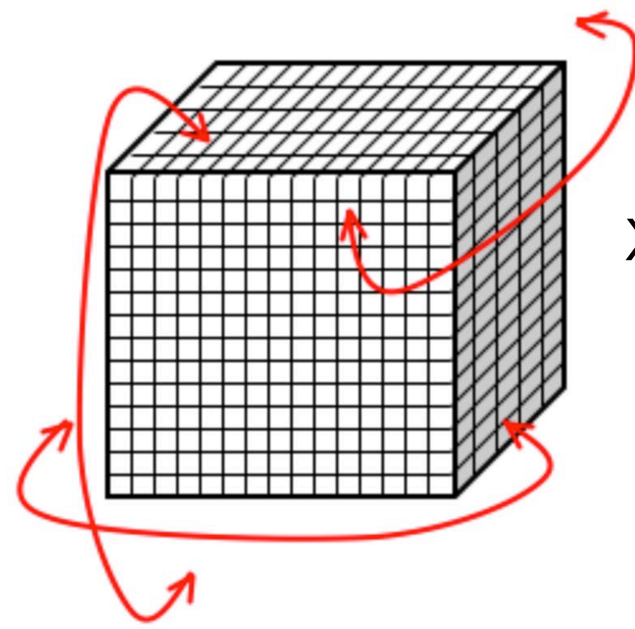


- Evaluate by importance sampling
- Paths near classical action dominate
- Calculate physics on a set (ensemble) of samples of the quark and gluon fields

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

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

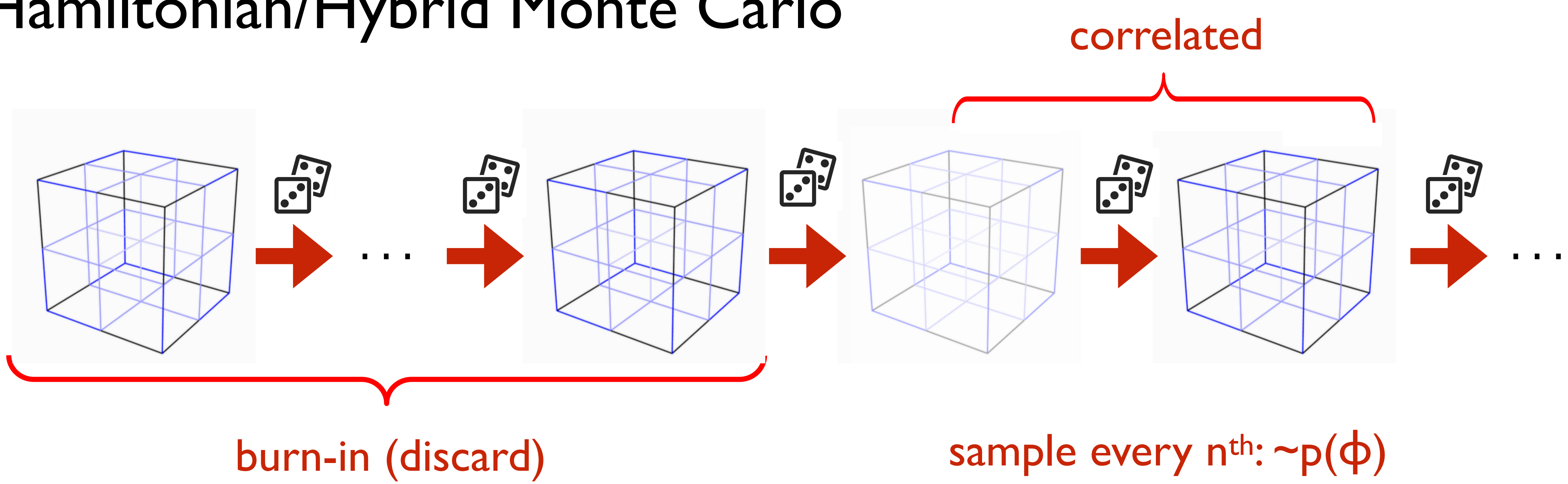
Computational cost grows exponentially with size of nuclear system

Generate QCD gauge fields

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

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

Hamiltonian/Hybrid Monte Carlo

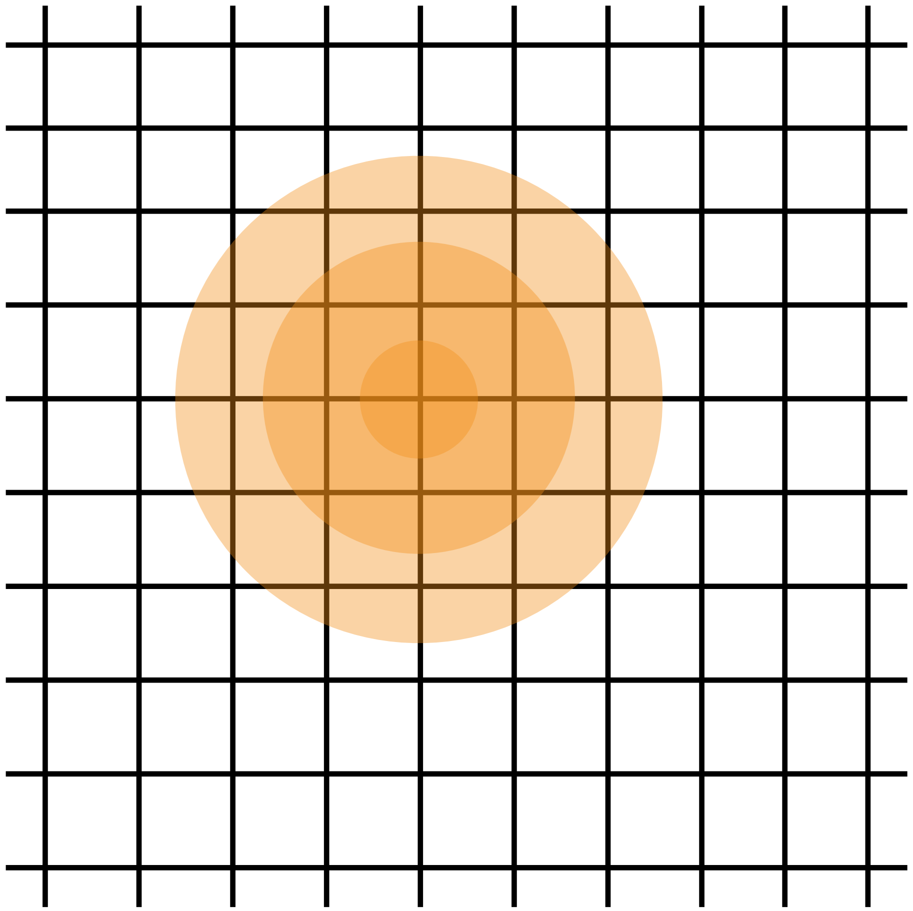


Burn-in time and correlation length dictated by Markov chain **'autocorrelation time'**: shorter autocorrelation time implies less computational cost

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

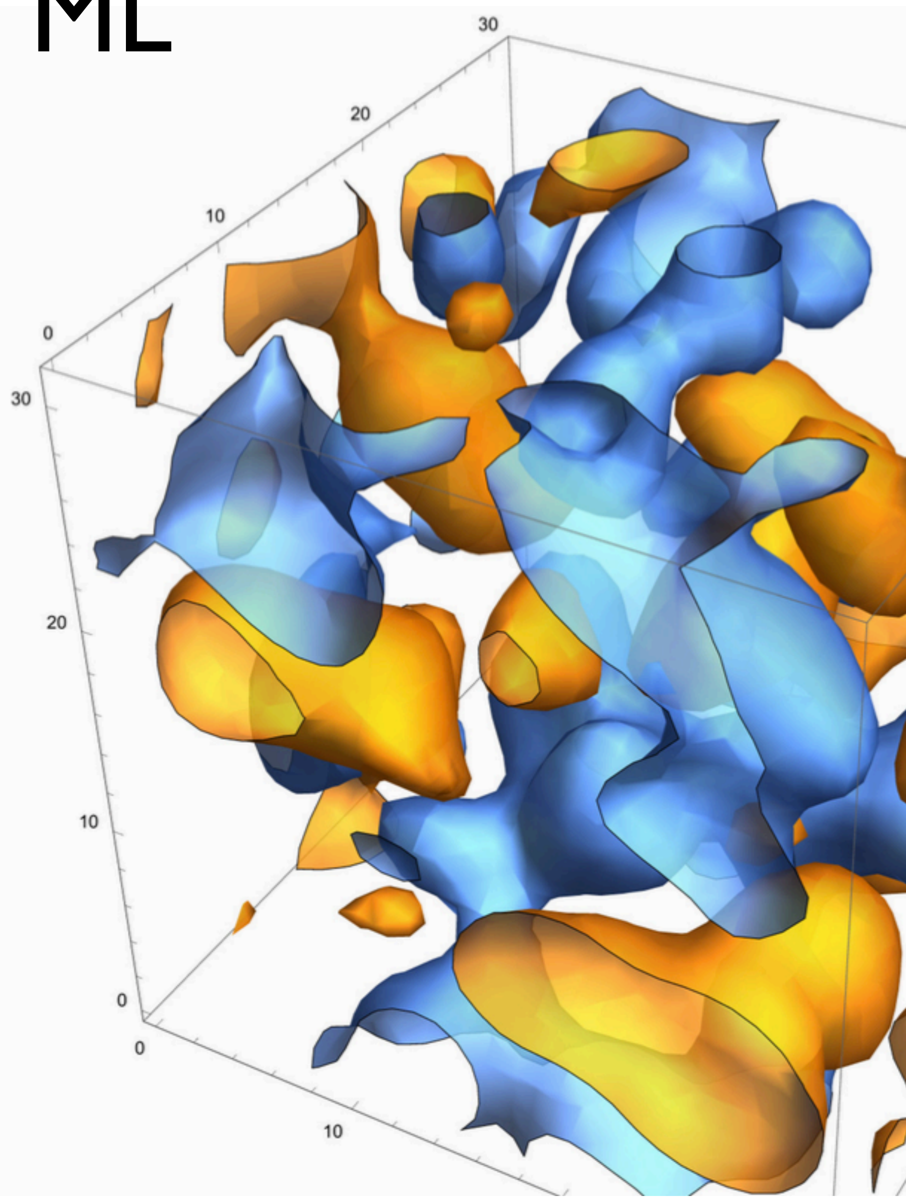
Machine learning QCD

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., PRD 100, 034515 (2019)
   MIT +NYU + Google DeepMind, arXiv:2002.02428
   Kanwar et al., arXiv:2003.06413

3. Hybrid approaches



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

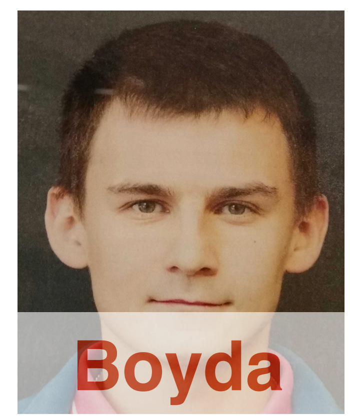
Machine learning for LQCD

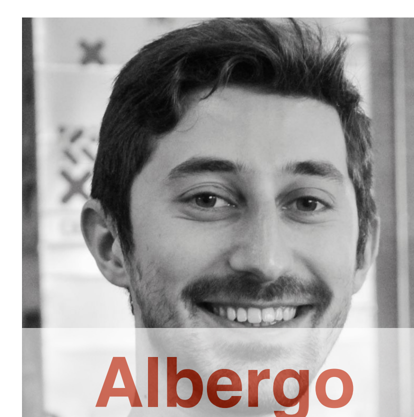
Generative models for QCD gauge field generation

Massachusetts Institute of Technology

Google
DeepMind

NYU

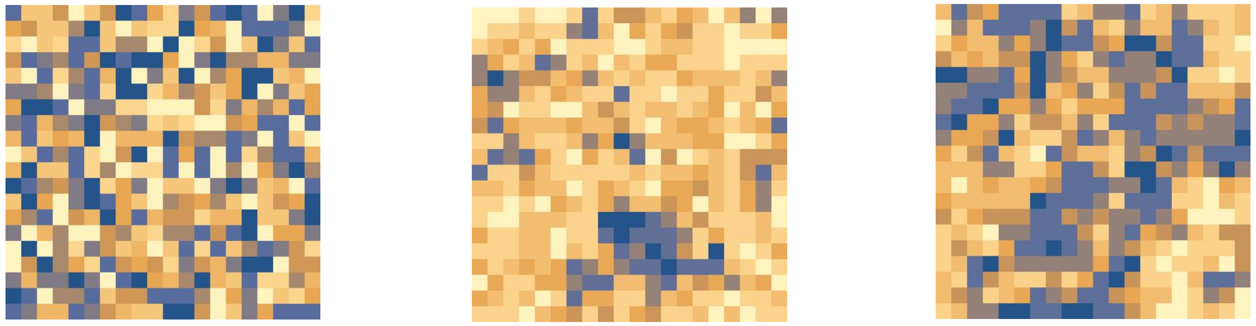





Scalar lattice field theory

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

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

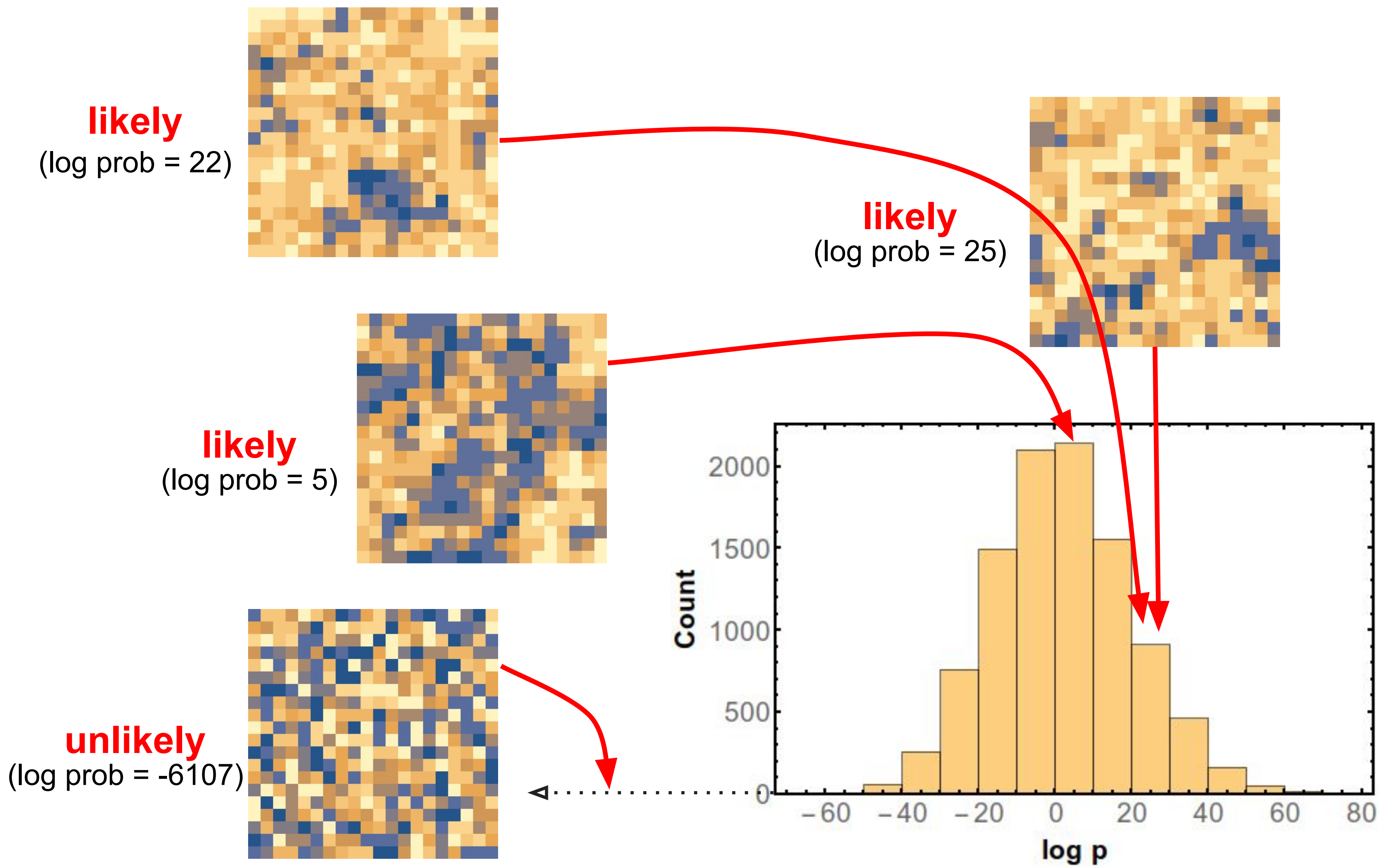
Generate field configurations $\phi(x)$ with probability

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

Sampling gauge field configs

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

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



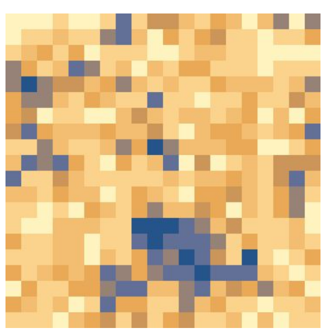
Sampling gauge field configs

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

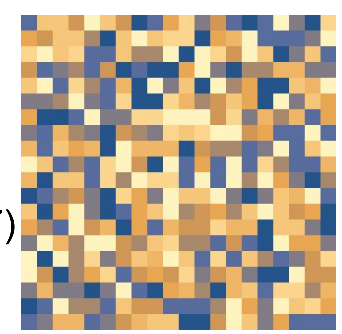
$$P[\phi(x)] \sim e^{-S[\phi(x)]}$$

Parallels with image generation problem

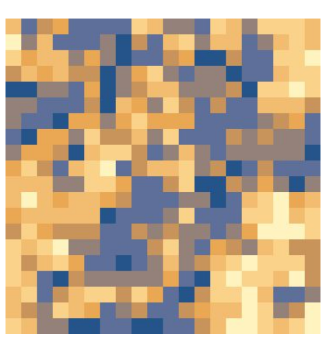
likely
(log prob = 22)



unlikely
(log prob = -6107)



likely
(log prob = 5)

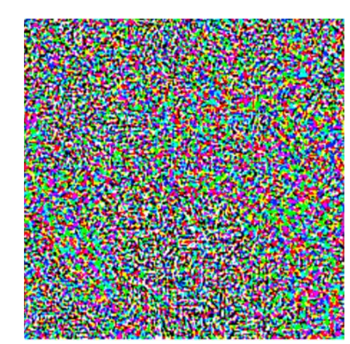


likely

[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

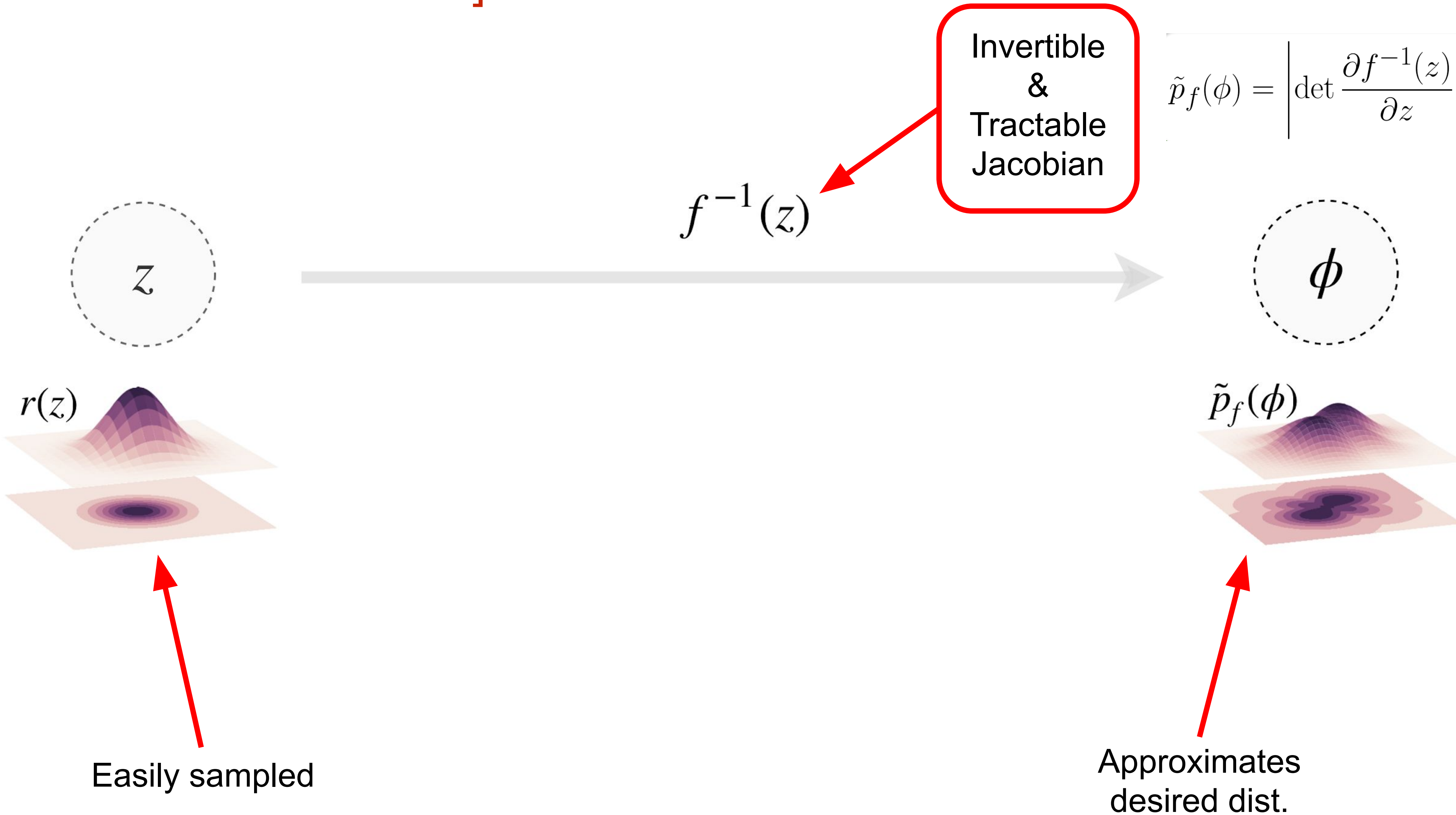
- 32×32 pixels $\times$ 3 cols $\simeq 3000$ numbers
- 60000 samples
- Each image has meaning
- Local structures are important
- Translation-invariance within frame

Generative flow models

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

[Rezende & Mohamed 1505.05770]

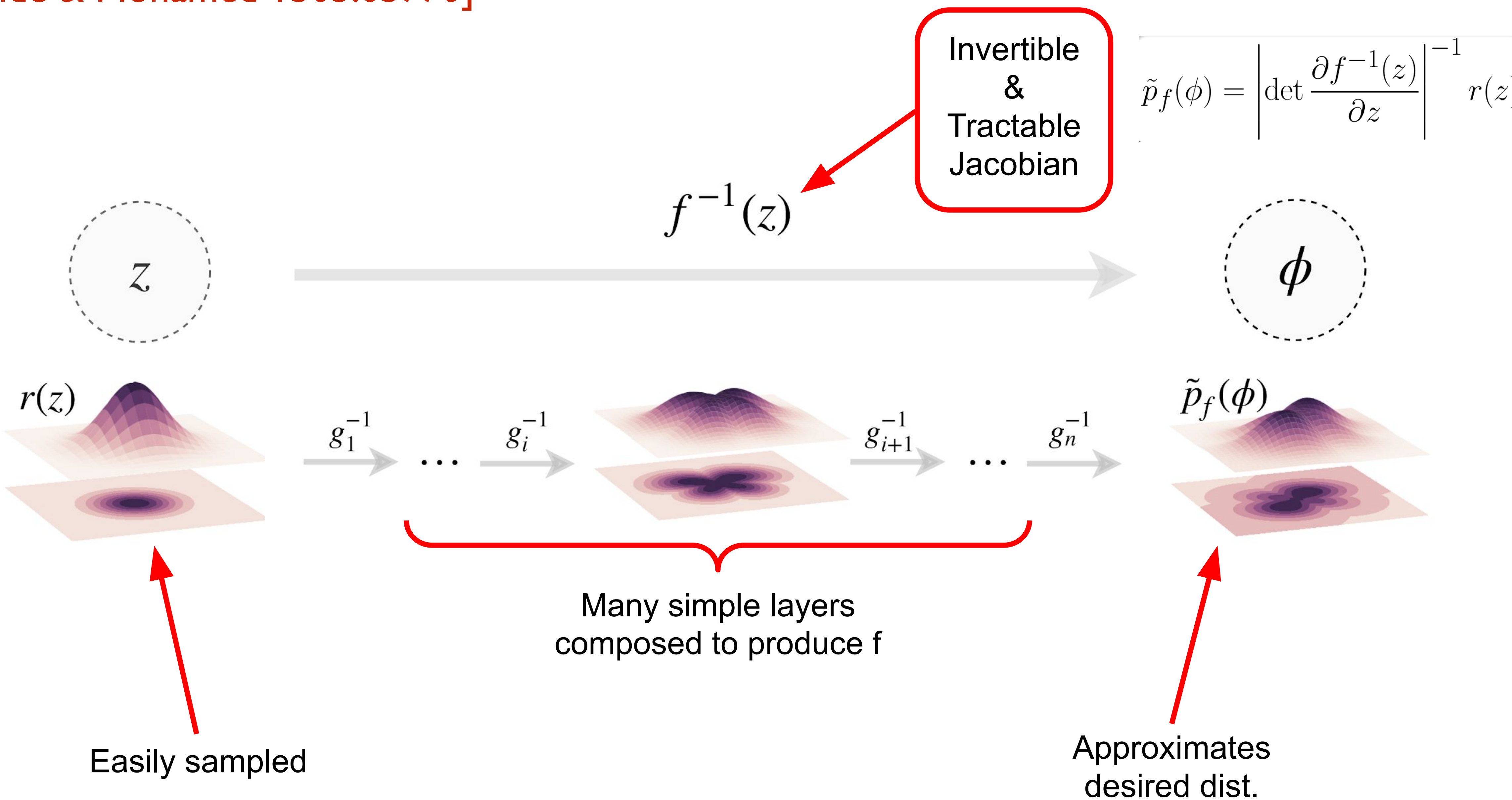
$$\tilde{p}_f(\phi) = \left|\det \frac{\partial f^{-1}(z)}{\partial z}\right|^{-1} r(z)$$



Generative flow models

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

[Rezende & Mohamed 1505.05770]

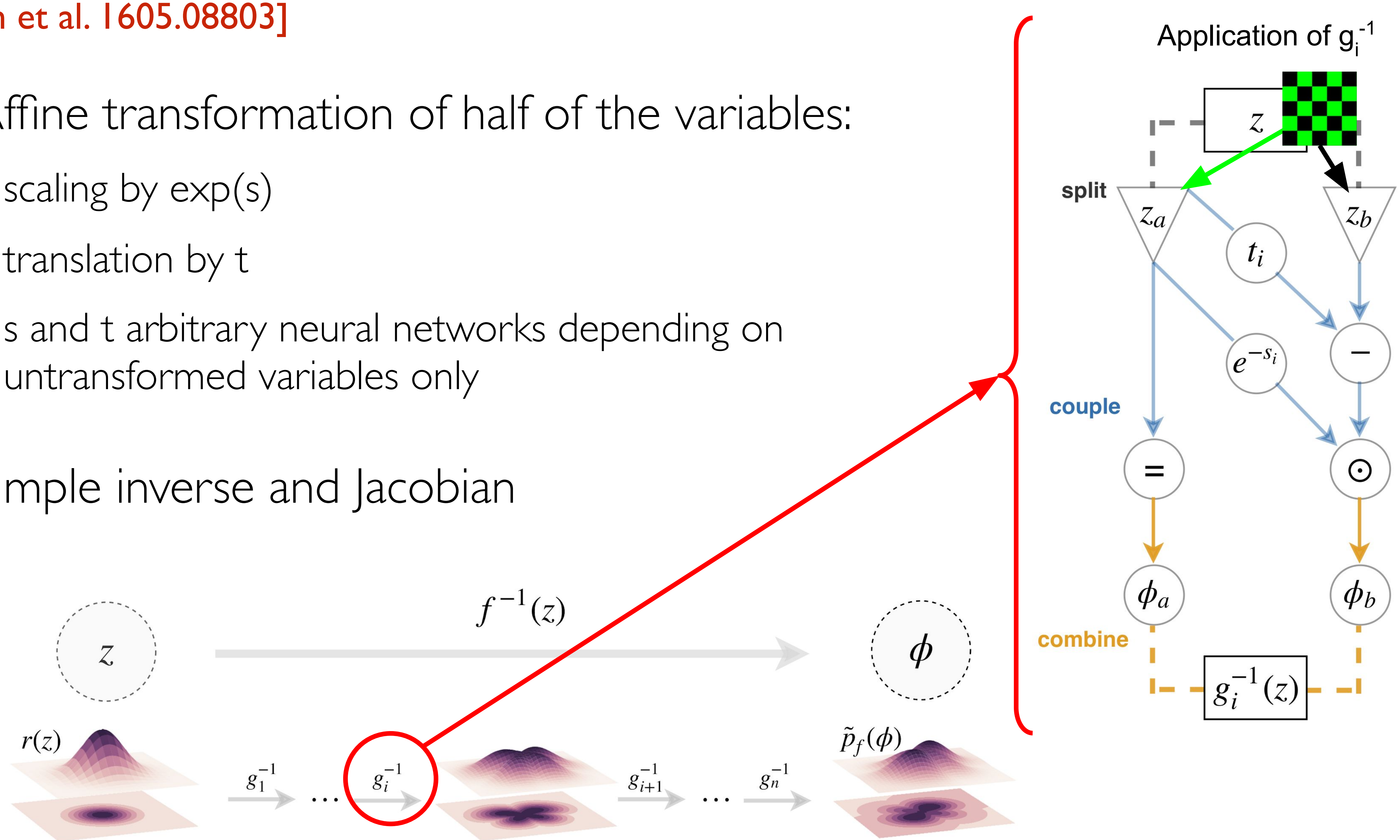


$$\tilde{p}_f(\phi) = \left| \det \frac{\partial f^{-1}(z)}{\partial z} \right|^{-1} r(z)$$

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



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]

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

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

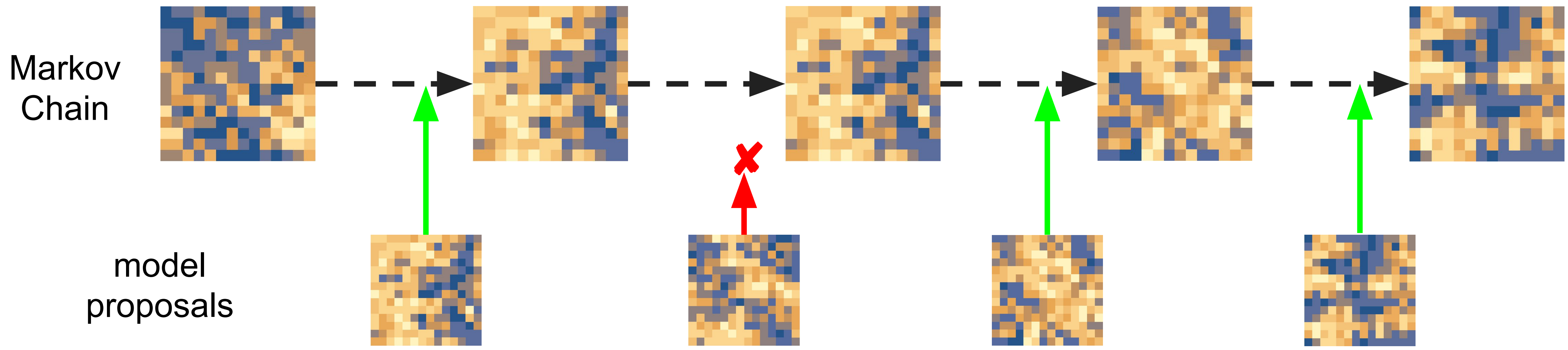
Acceptance probability

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

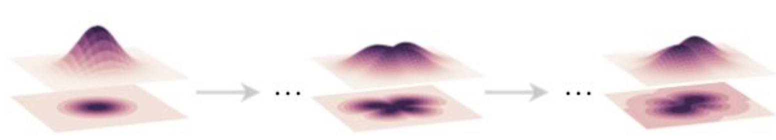
True dist

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

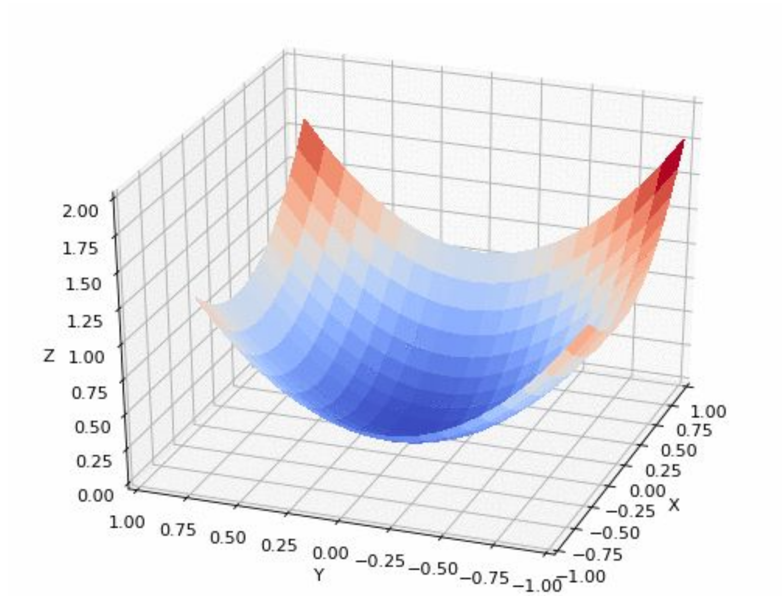
Draw samples from model

Compute loss function

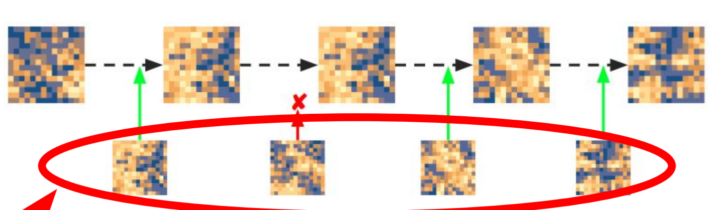
Gradient descent

Desired accuracy?

Save trained model



Markov chain using samples from model



generating samples is "embarrassingly parallel"

Summary chart: Tej Kanwar

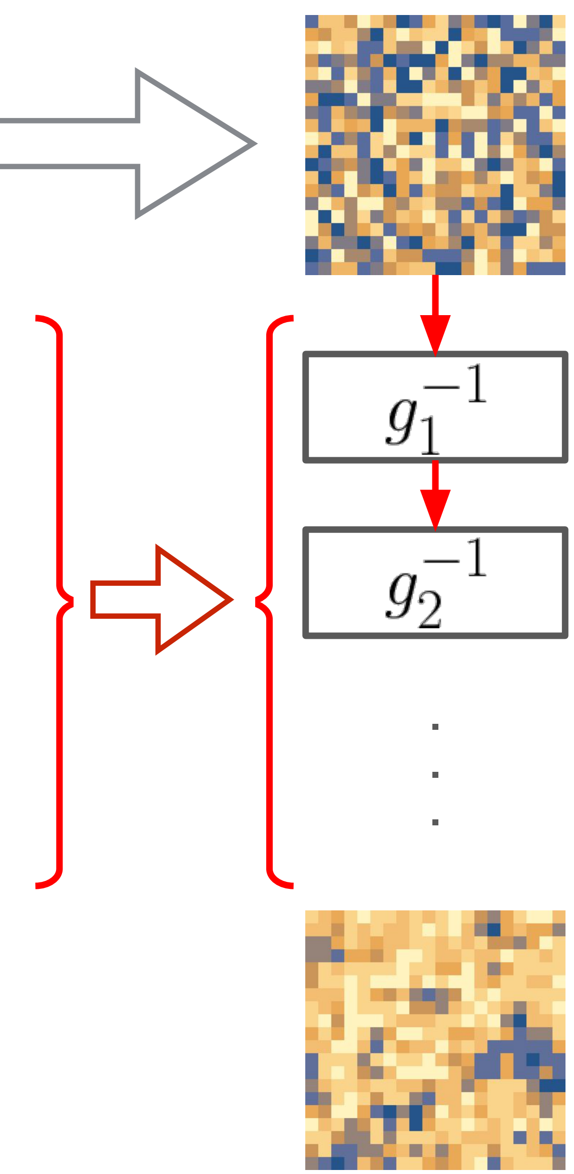
Application: scalar field theory

First application: scalar lattice field theory

- 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

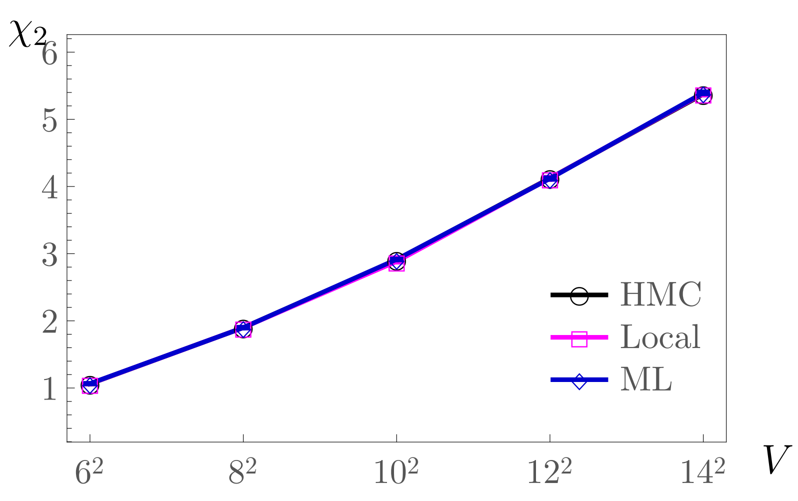
First application: scalar lattice field theory

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

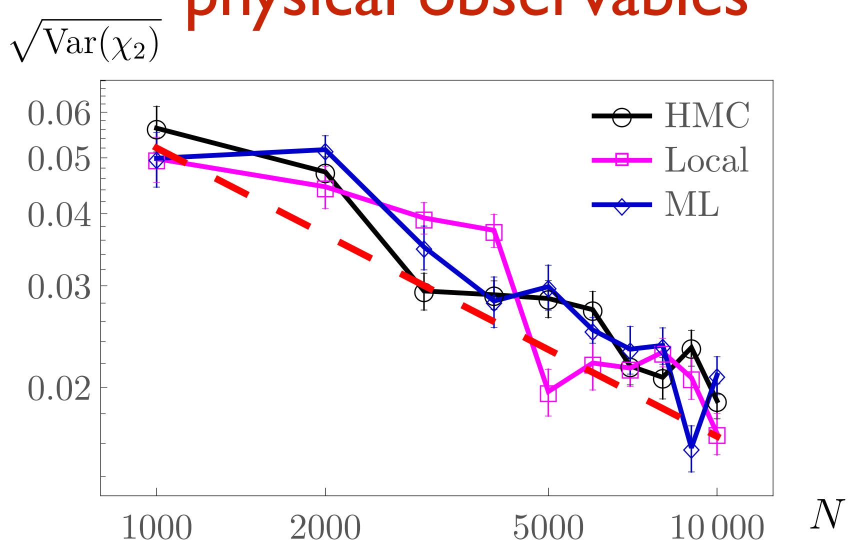
ML model produces **varied samples** and **correlations at the right scale**

Compare:

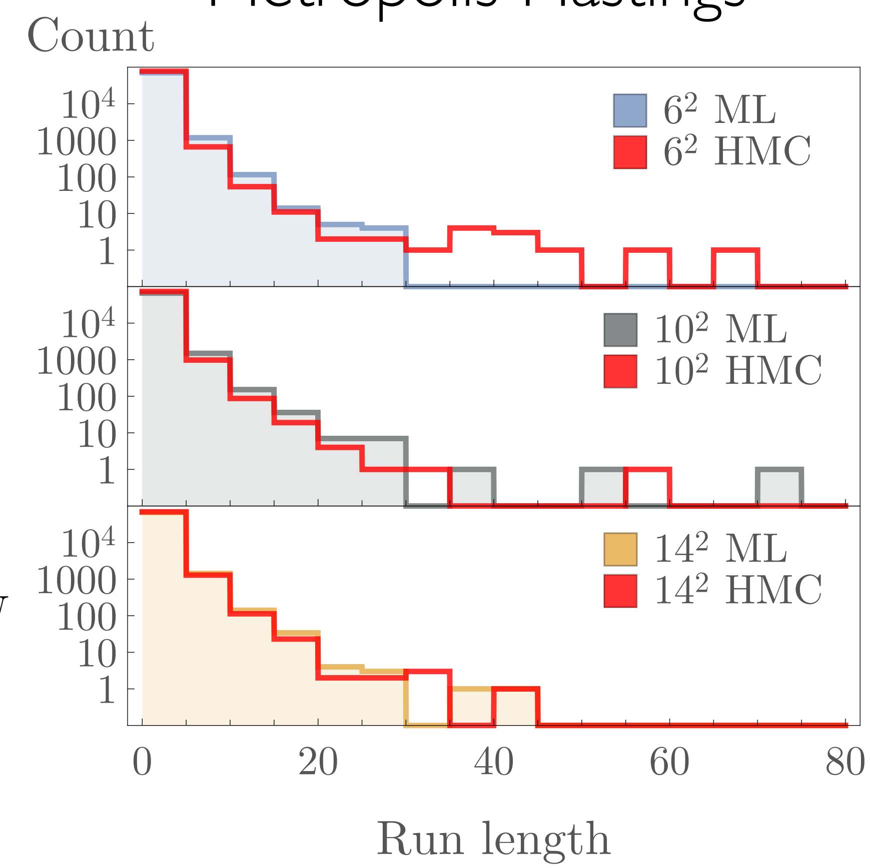
Physical observables



Uncertainties in physical observables



Rejection runs in Metropolis-Hastings

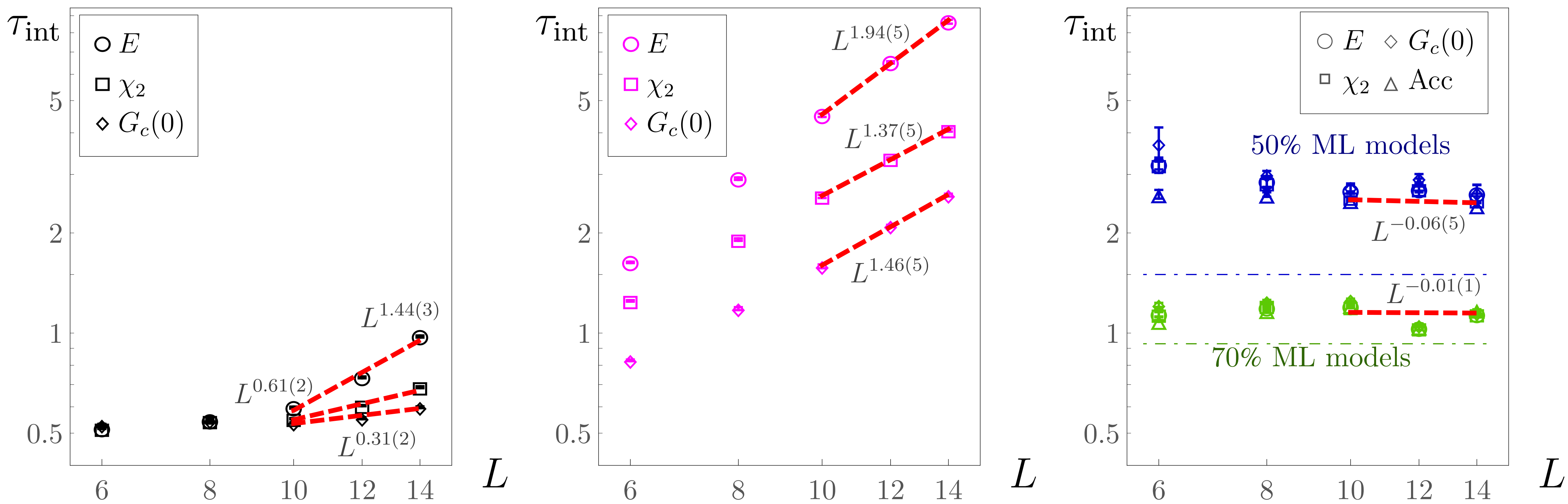


Application: scalar field theory

First application: scalar lattice field theory

Success: Critical slowing down is eliminated

Cost: Up-front training of the model



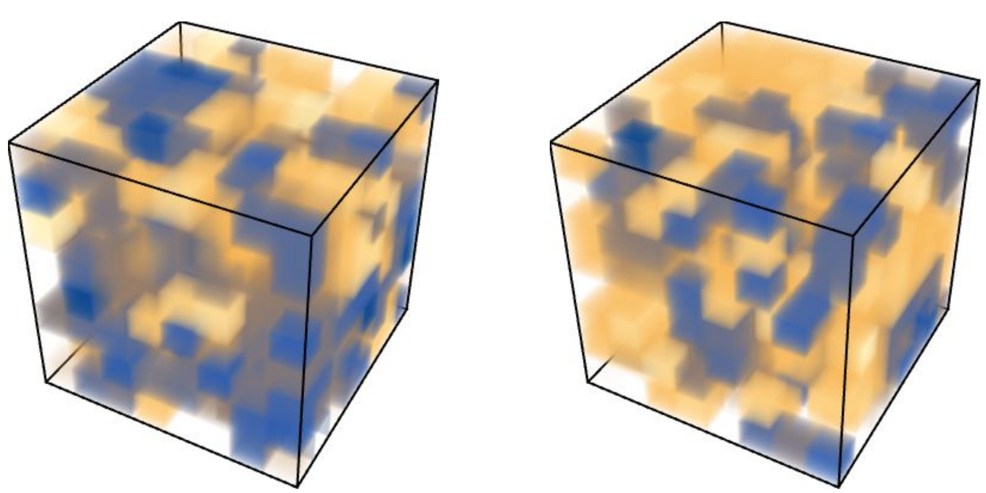
(a) HMC ensembles (b) Local Metropolis ensembles (c) Flow-based MCMC ensembles

Dynamical critical exponents consistent with zero

Next steps: 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
   [MIT, NYU, DeepMind, arXiv:2002.02428, arXiv:2003.06413]

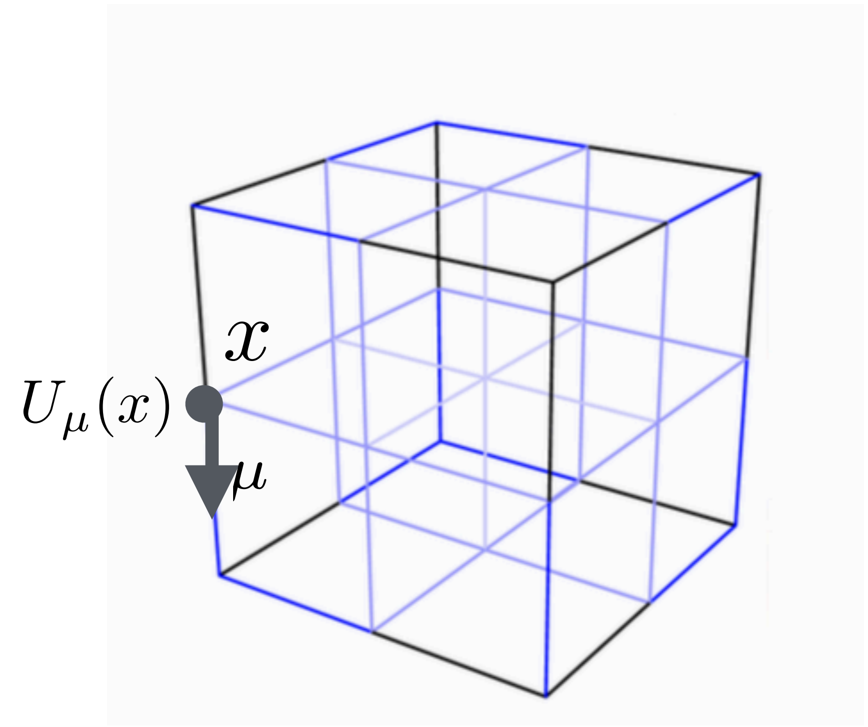


Aurora21 Early Science Project

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** [ICML, arXiv:2002.02428]
2. **Incorporate symmetries: gauge-equivariant flows**

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

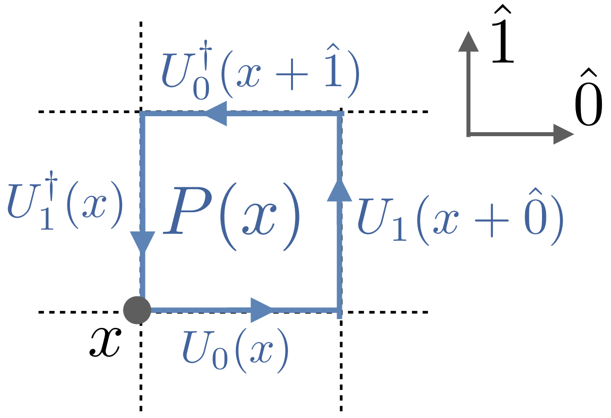
Application: U(1) field theory

First gauge theory application: U(1) field theory

- One complex number $U = e^{i\theta}$ per link on a 2D lattice
- Action: expressed in terms of plaquettes (products of links around closed loops) with a single coupling

$$S(U) := -\beta \sum_x \operatorname{Re} P(x)$$

$$P(x) := U_0(x) U_1(x+\hat{0}) U_0^\dagger(x+\hat{1}) U_1^\dagger(x)$$



- Fixed lattice size: $L^2 = 16$ with couplings $\beta = \{1,2,3,4,5,6,7\}$
- Continuum limit (critical slow-down) as $\beta \to \infty$

[Kanwar et al., arXiv:2003.06413]

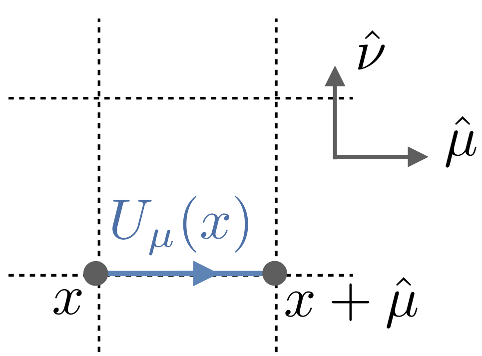
Gauge field theory

First gauge theory application: U(1) field theory

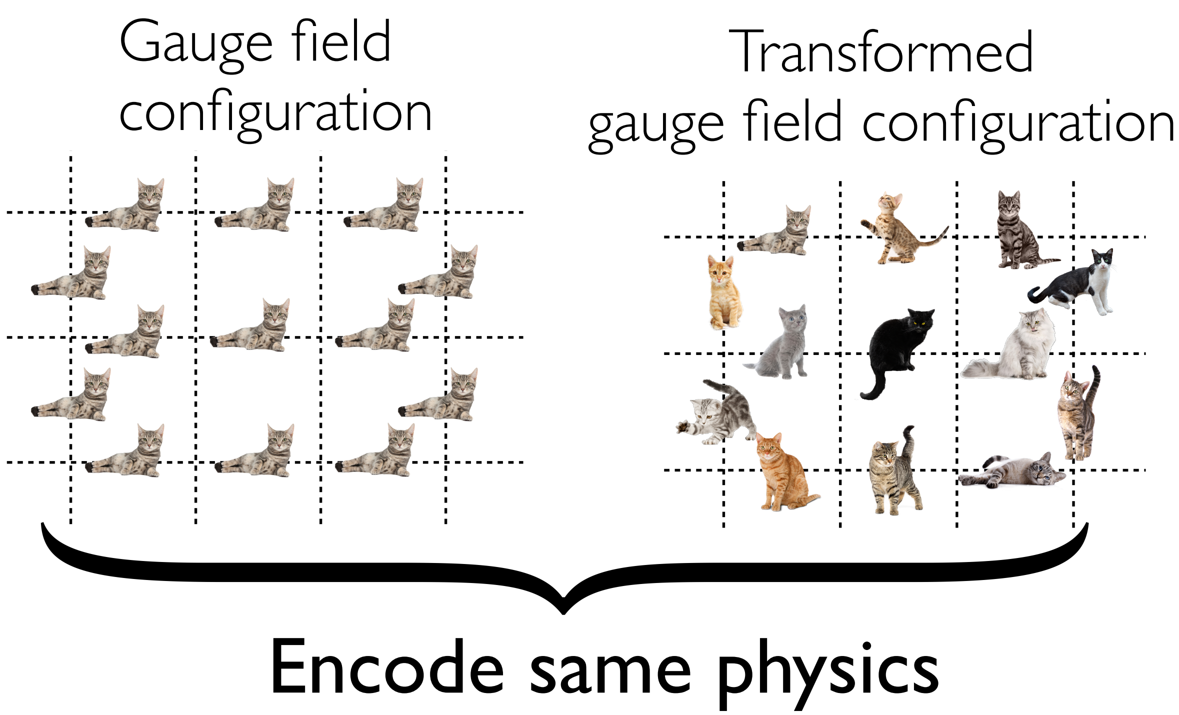
Generative flow architecture that is *gauge-equivariant*

Gauge transformation

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



$$U_\mu(x) \to U'_\mu(x) = \Omega(x) U_\mu(x) \Omega^\dagger(x + \hat{\mu})$$
$$\text{for all } \Omega(x) \in U(1)$$



Gauge-equivariant flows

First gauge theory application: U(1) field theory

Generative flow architecture that is *gauge-equivariant*

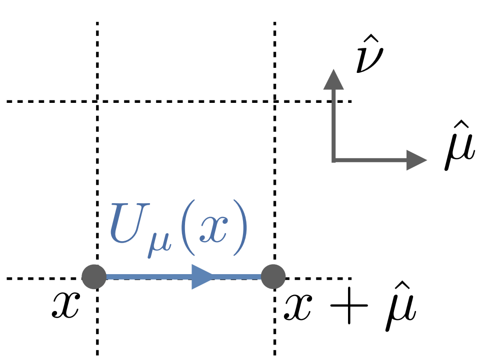
Define invertible, equivariant coupling layer

Spacetime dimension

$$g \; : \; G^{N_d V} \;\rightarrow\; G^{N_d V}$$

Lattice volume

Act on a subset of the variables in each layer



$$g(U^A, U^B) = (U'^A, U^B)$$

Links updated by coupling layer

Links frozen in coupling layer

[Kanwar et al., arXiv:2003.06413]

Gauge-equivariant flows

First gauge theory application: U(1) field theory

Generative flow architecture that is *gauge-equivariant*

Define invertible, equivariant coupling layer $g(U^A, U^B) = (U'^A, U^B)$

Link updates via a kernel $h : G \to G$

Gauge-invariant quantities constructed from elements of U^B.

Link updated by coupling layer

$$U'^i = h(U^i S^i \,|\, I^i)\, S^{i\dagger}$$

Loop that starts and ends at same point

Coulpling layer equivariant under the condition

$$h(XWX^\dagger) = X\,h(W)\,X^\dagger, \quad \forall X, W \in G$$

[Kanwar et al., arXiv:2003.06413]

Gauge-equivariant flows

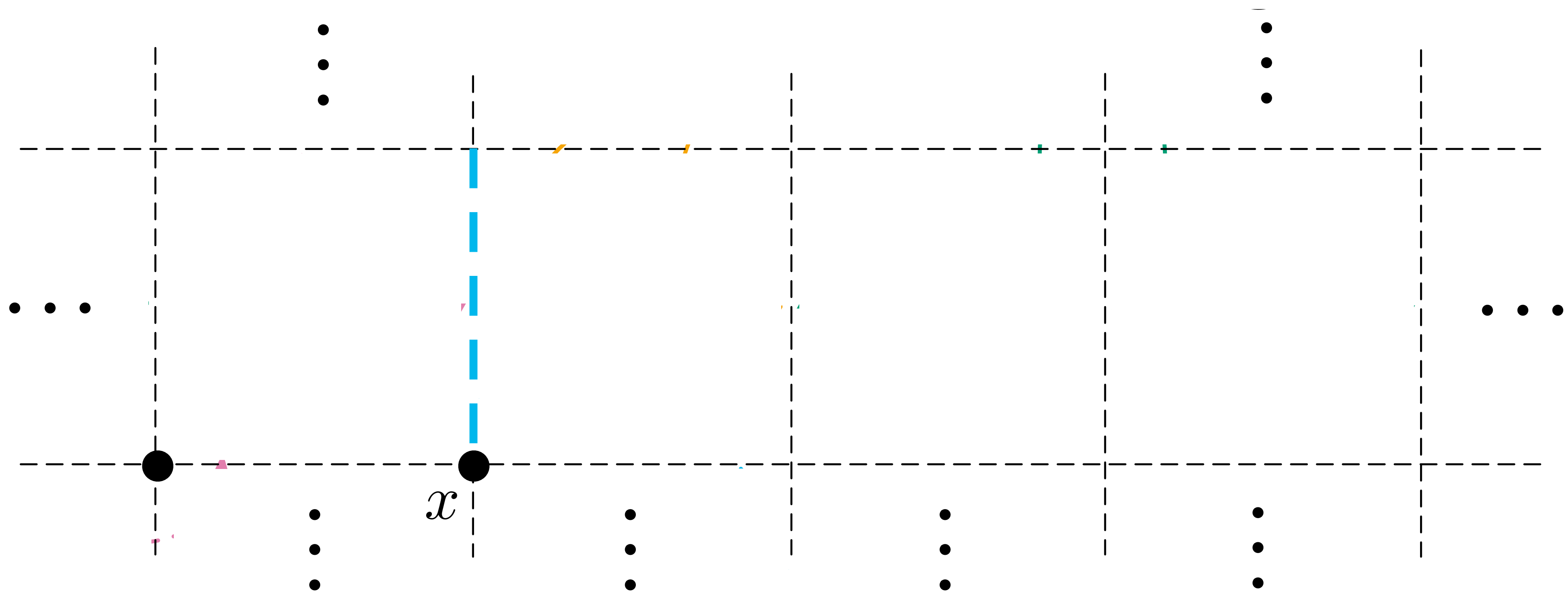
First gauge theory application: U(1) field theory

Generative flow architecture that is *gauge-equivariant*

$$U'^{i} = h(U^{i} S^{i} | I^{i}) S^{i\dagger}$$

Gauge-invariant quantities constructed from elements of U^B.

Loop that starts and ends at same point



[Kanwar et al., arXiv:2003.06413]

Gauge-equivariant flows

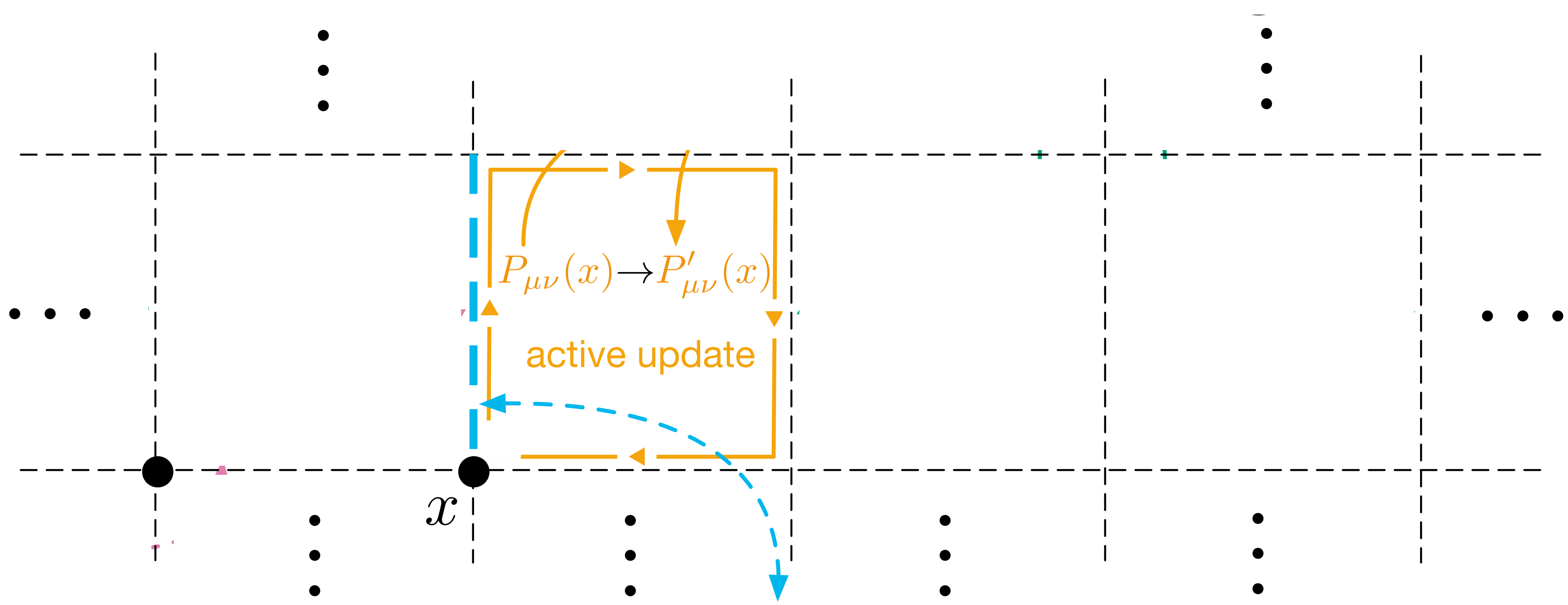
First gauge theory application: U(1) field theory

Generative flow architecture that is *gauge-equivariant*

$$U'^{i} = h(U^{i} S^{i} | I^{i}) S^{i\dagger}$$

Gauge-invariant quantities constructed from elements of U^B.

Loop that starts and ends at same point



$$U'_\mu(x) = P'_{\mu\nu}(x) P^\dagger_{\mu\nu}(x) U_\mu(x)$$

[Kanwar et al., arXiv:2003.06413]

Gauge-equivariant flows

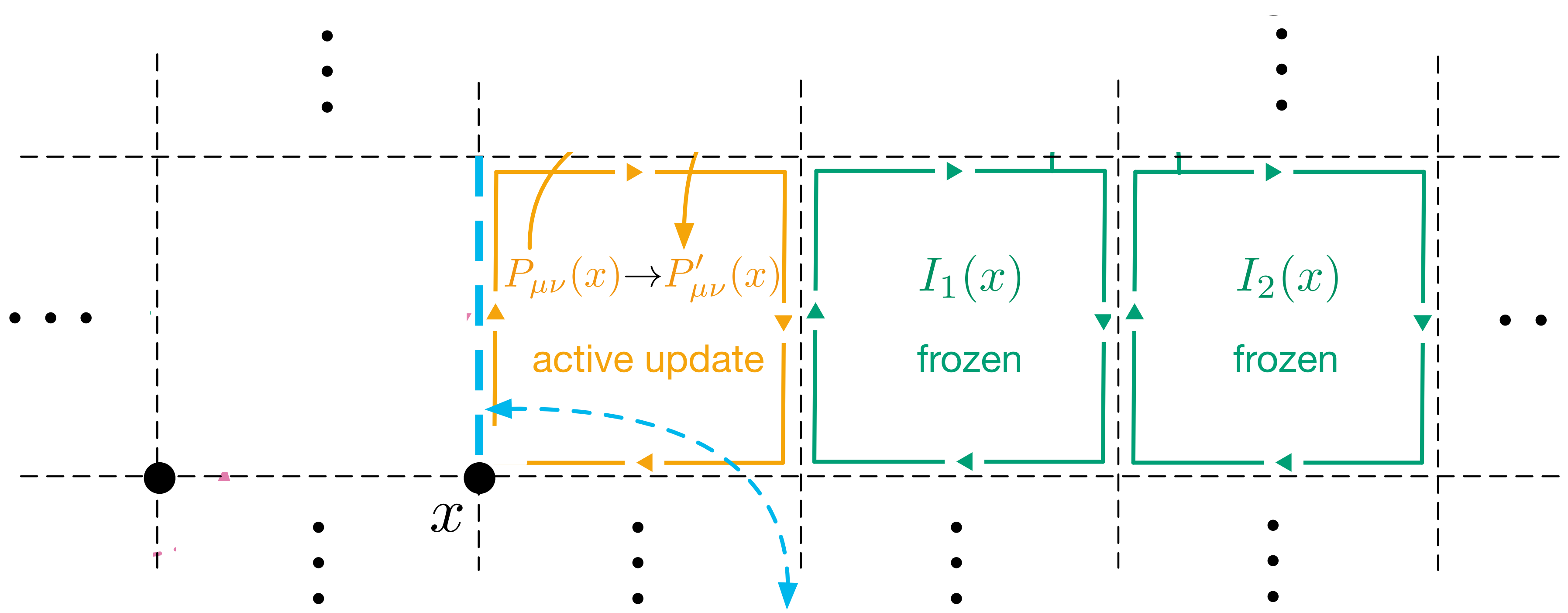
First gauge theory application: U(1) field theory

Generative flow architecture that is *gauge-equivariant*

$$U'^i = h(U^i S^i | I^i) S^{i\dagger}$$

Gauge-invariant quantities constructed from elements of U^B.

Loop that starts and ends at same point



$$U'_\mu(x) = P'_{\mu\nu}(x) P^\dagger_{\mu\nu}(x) U_\mu(x)$$

[Kanwar et al., arXiv:2003.06413]

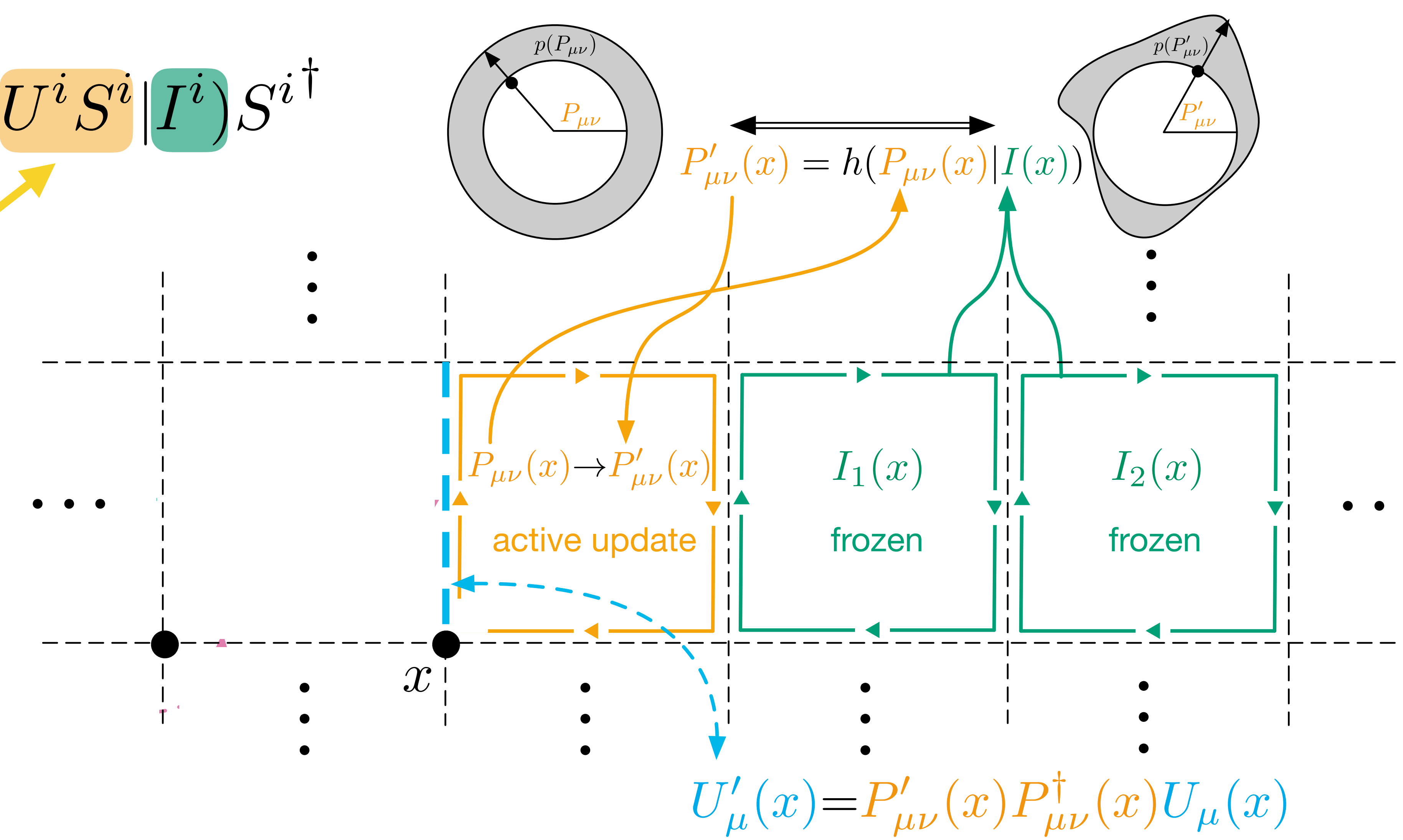
Gauge-equivariant flows

First gauge theory application: U(1) field theory

Generative flow architecture that is *gauge-equivariant*

$$U'^{i} = h(U^{i} S^{i} | I^{i}) S^{i\dagger}$$

Loop that starts and ends at same point



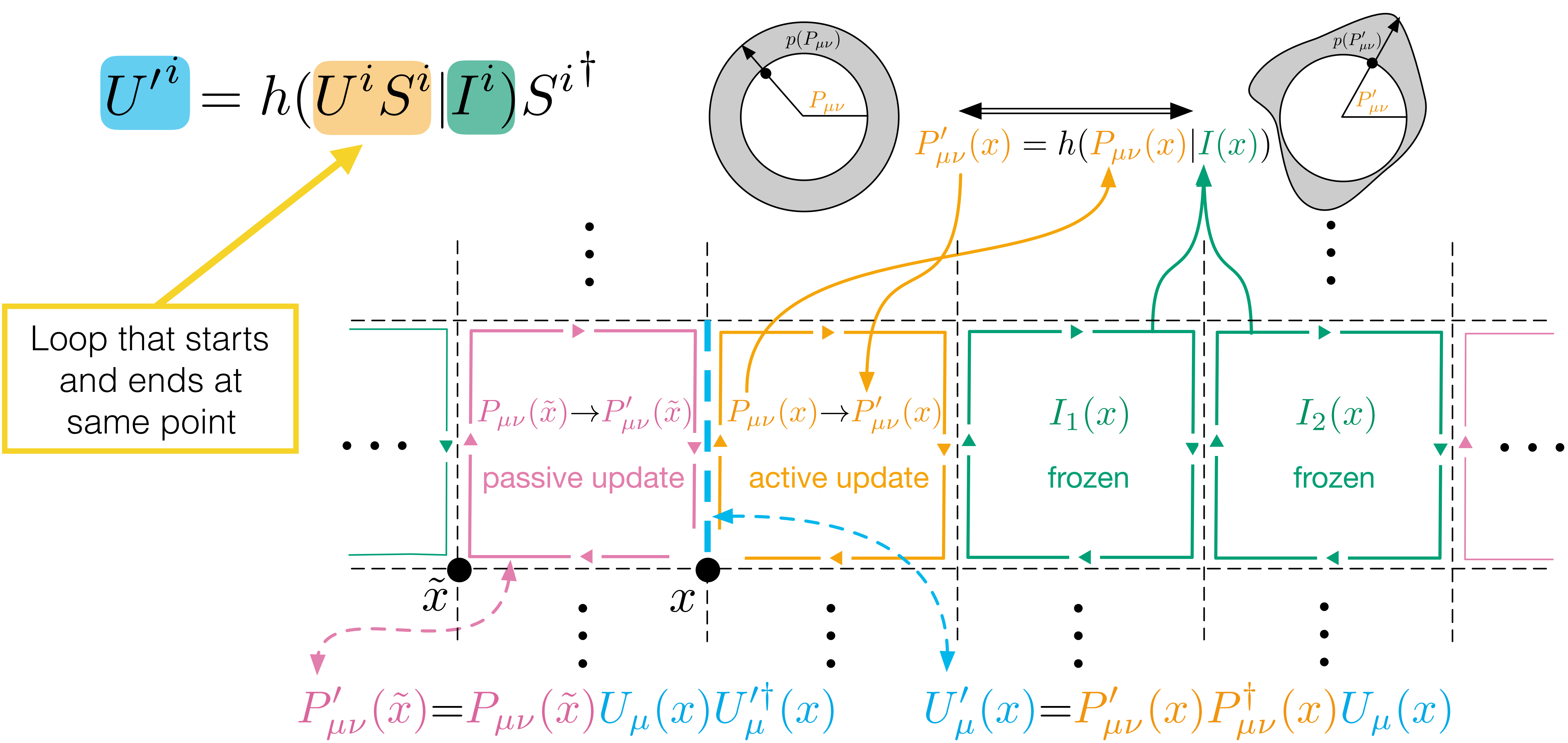
[Kanwar et al., arXiv:2003.06413]

Gauge-equivariant flows

First gauge theory application: U(1) field theory

Generative flow architecture that is *gauge-equivariant*

$$U'^{i} = h(U^i S^i | I^i) S^{i\dagger}$$

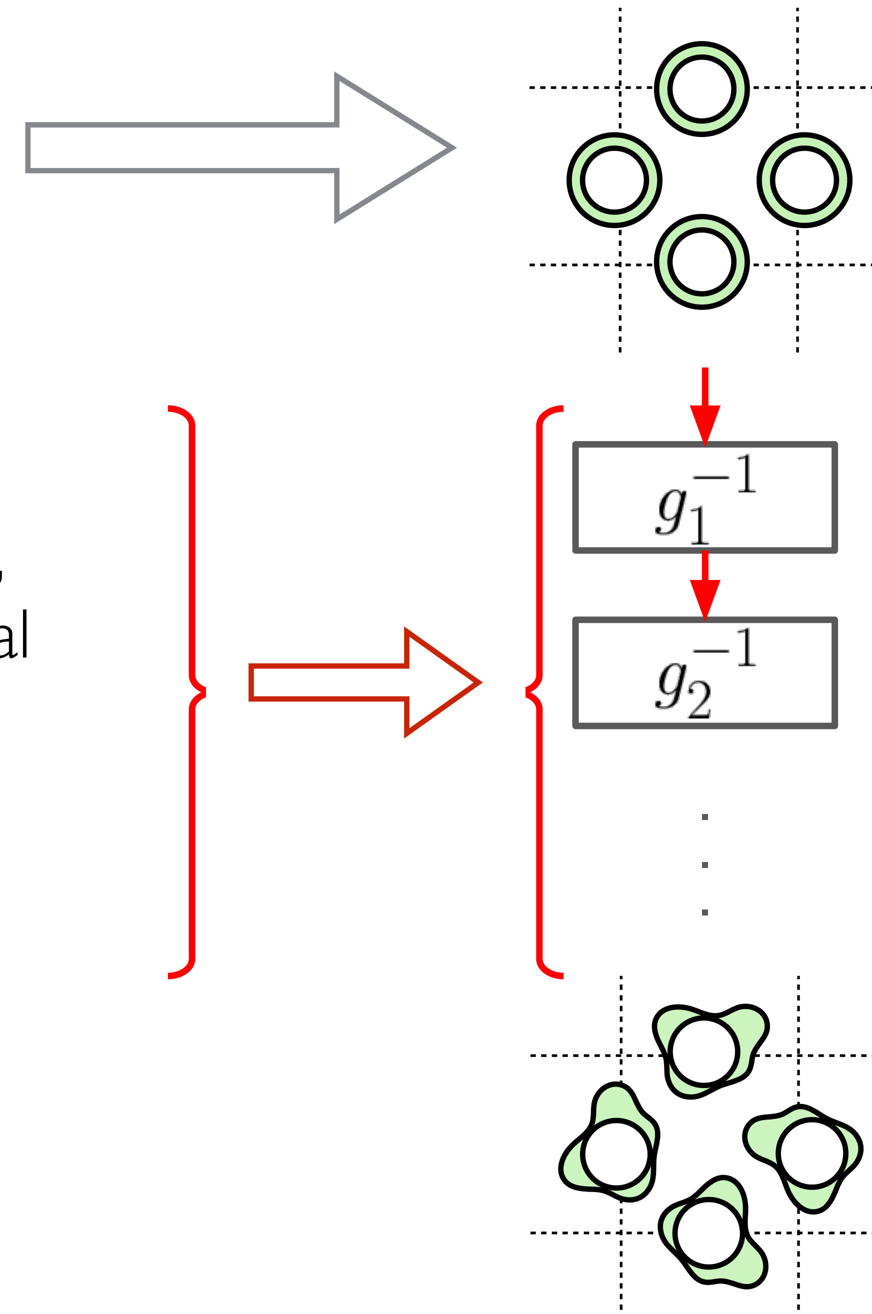


[Kanwar et al., arXiv:2003.06413]

Application: U(1) field theory

First gauge theory application: U(1) field theory

- Prior distribution chosen to be uniform
- Gauge-equivariant coupling layers
  - 24 coupling layers
  - Kernels h: mixtures of non-compact projections, 6 components, parameterised with convolutional NNs (i.e., NN output gives params. of NCP)
  - NNs with 2 hidden layers with 8x8 filters, kernel size 3
- Train using shifted KL loss with Adam optimizer
  - Stopping criterion: loss plateau



[Kanwar et al., arXiv:2003.06413]

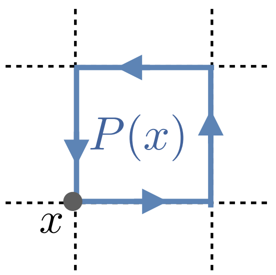
Application: U(1) field theory

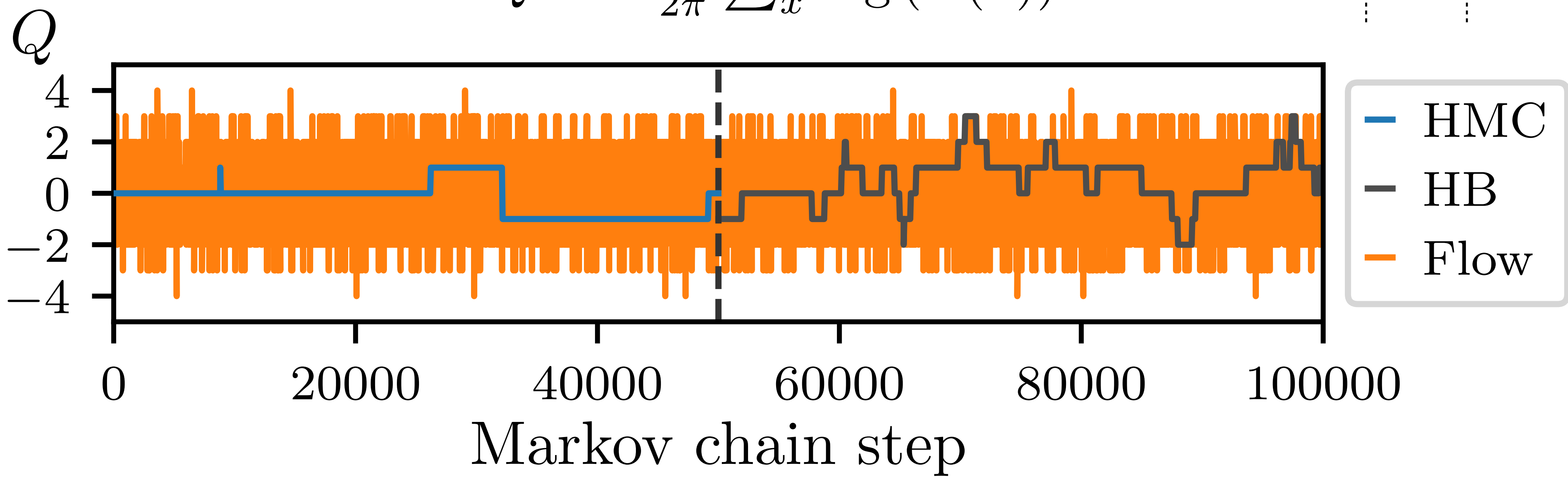
First gauge theory application: U(1) field theory

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., arXiv:2003.06413]

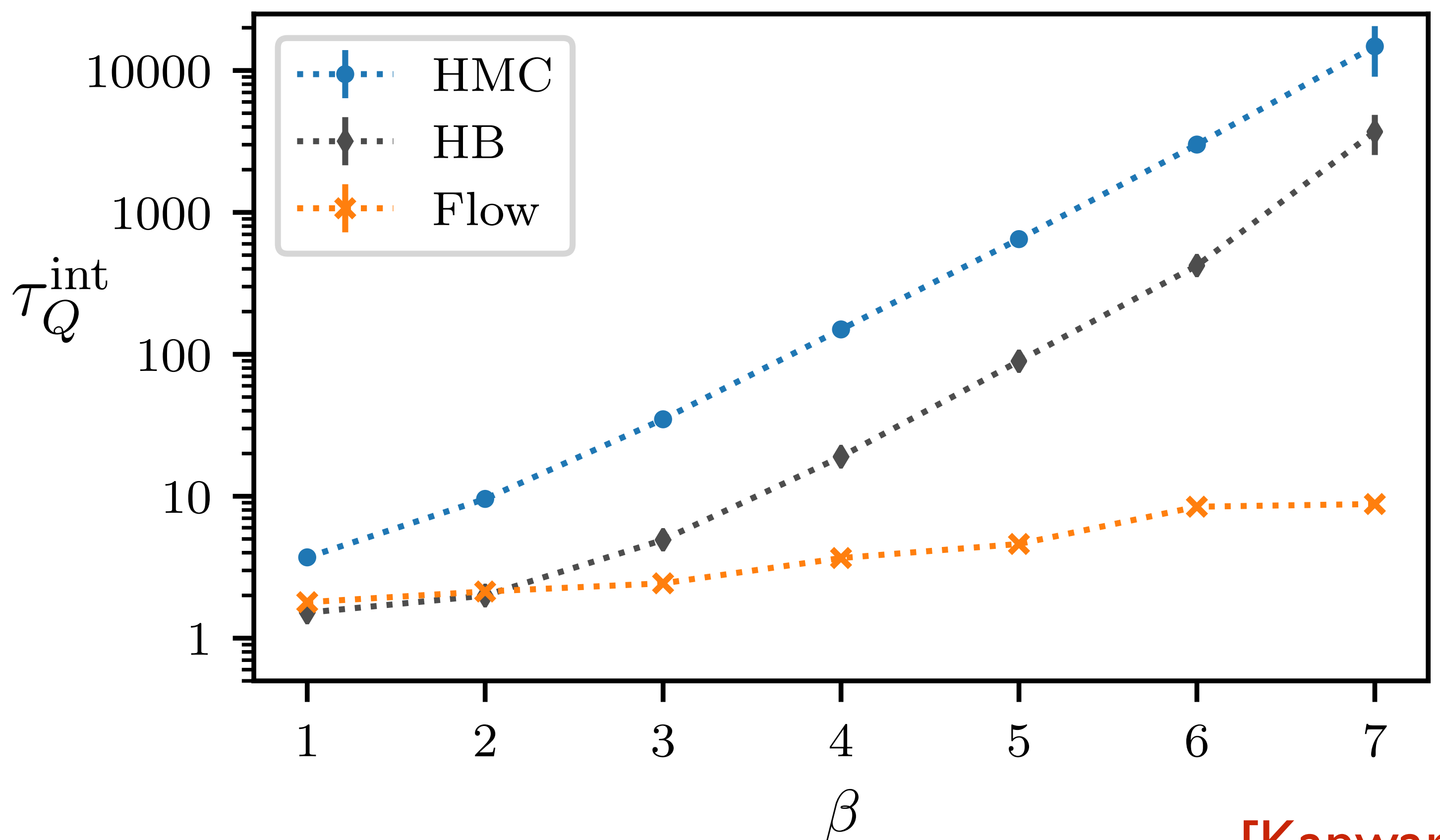
Application: U(1) field theory

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., arXiv:2003.06413]

Application: U(1) field theory

First gauge theory application: U(1) fi

Su

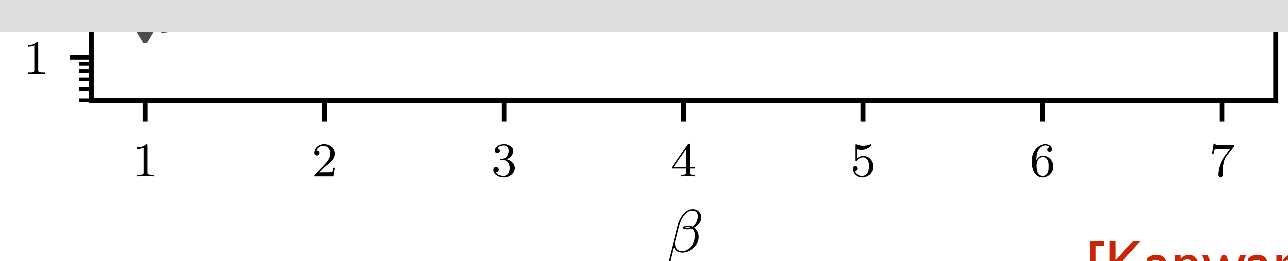
Co

SUCCESS!

Proof-of-principle of efficient, exact, ML algorithm for LQFT

Significant work required to scale to 4D, SU(3), fermions, large lattice volumes

Paper on 2D SU(2), SU(3) pure gauge coming soon



2D, L=16

[Kanwar et al., arXiv:2003.06413]

Outlook

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

Emergence of complex structure in nature

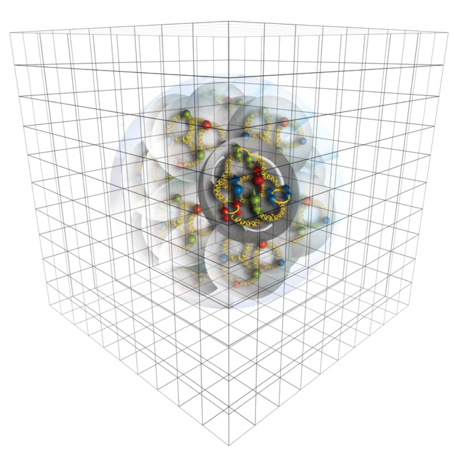
Backgrounds and benchmarks for searches for new physics

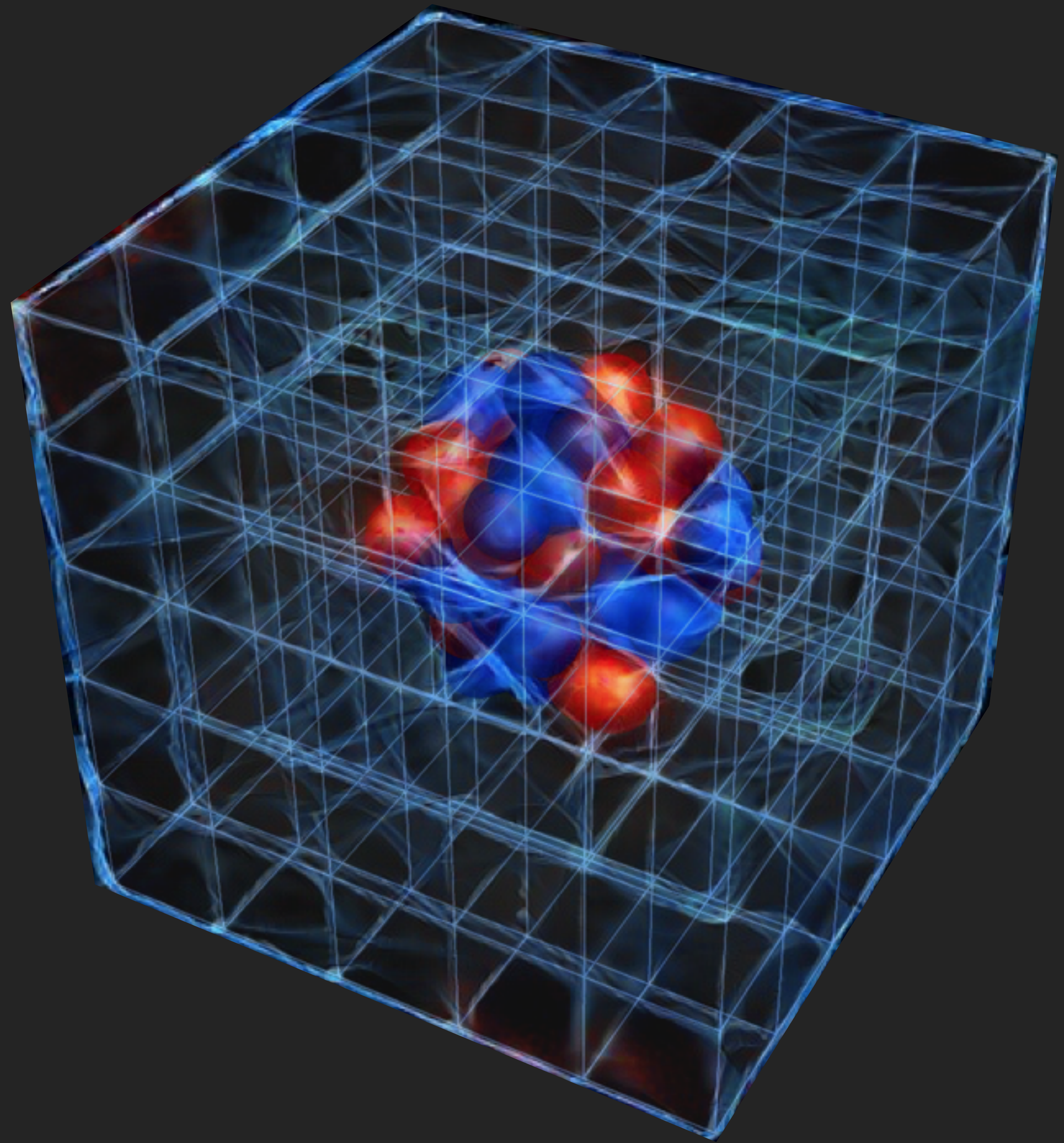
Flow-based generation of QCD gauge fields at scale would

- 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
- ….





Phiala Shanahan

Massachusetts Institute of Technology