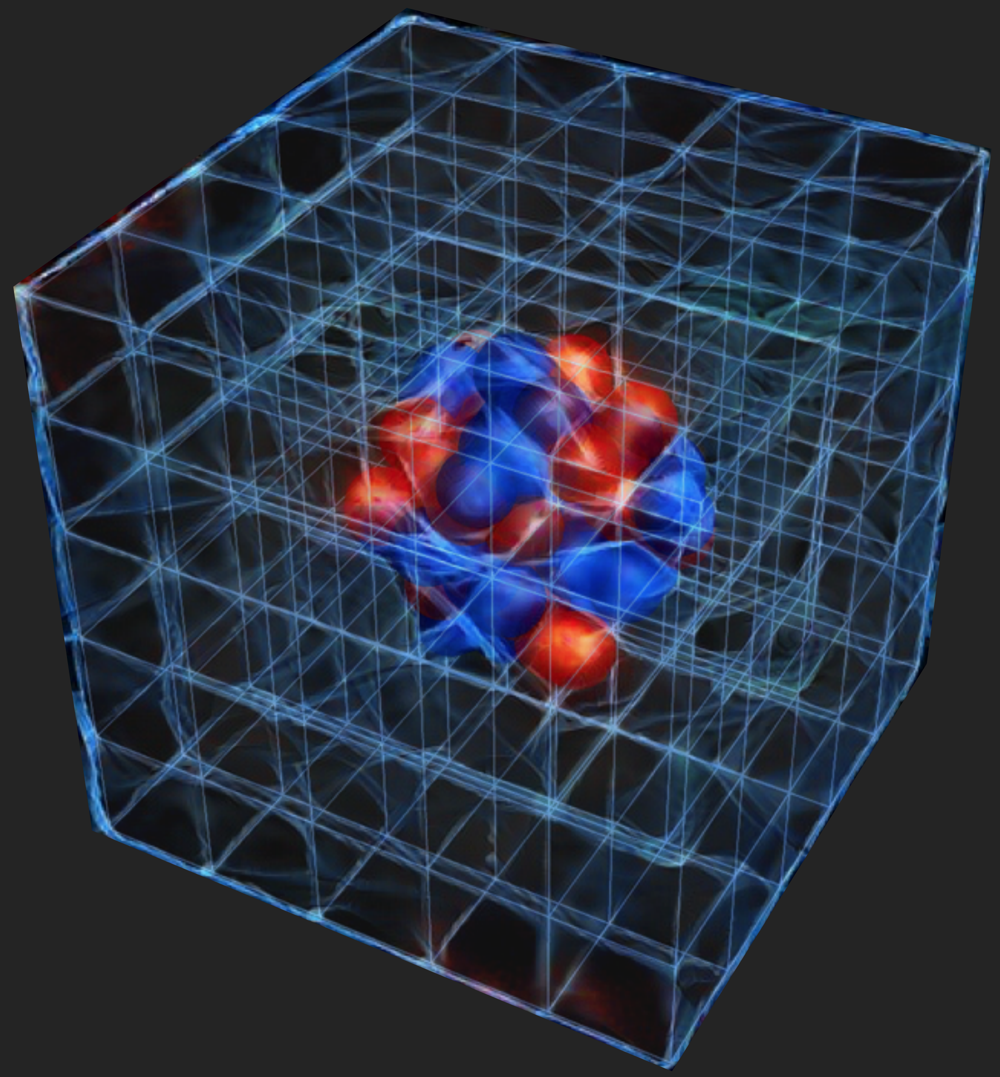
Machine learning for LQCD: ensemble generation



Phiala Shanahan

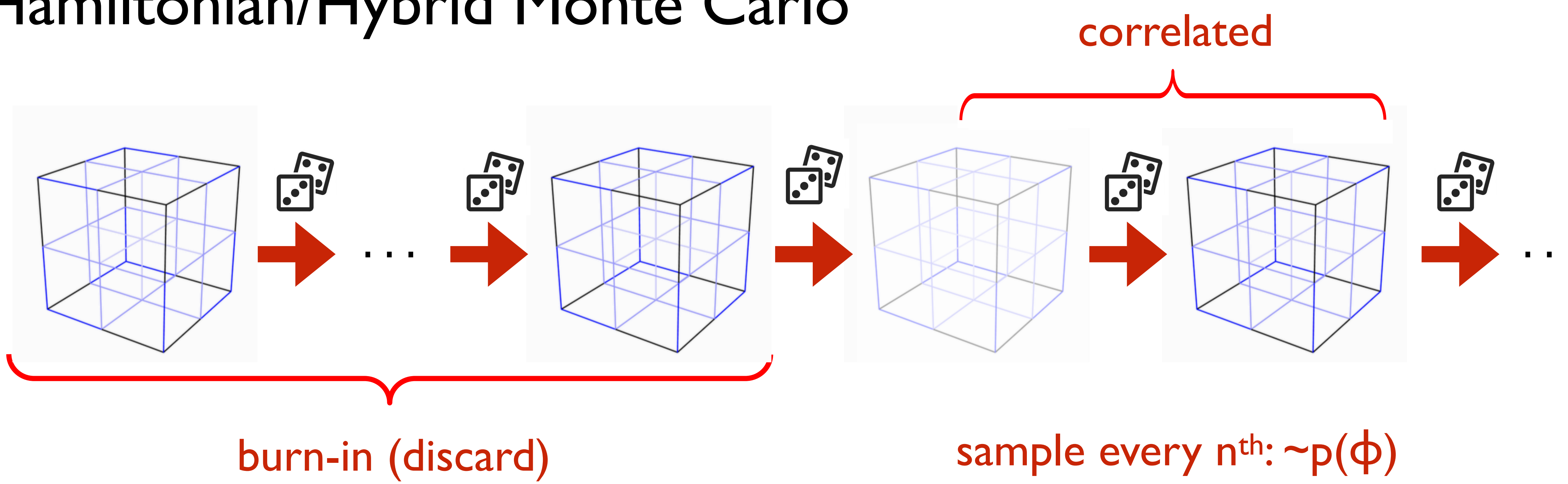
Massachusetts Institute of Technology

Generate QCD gauge fields

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

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

Hamiltonian/Hybrid Monte Carlo

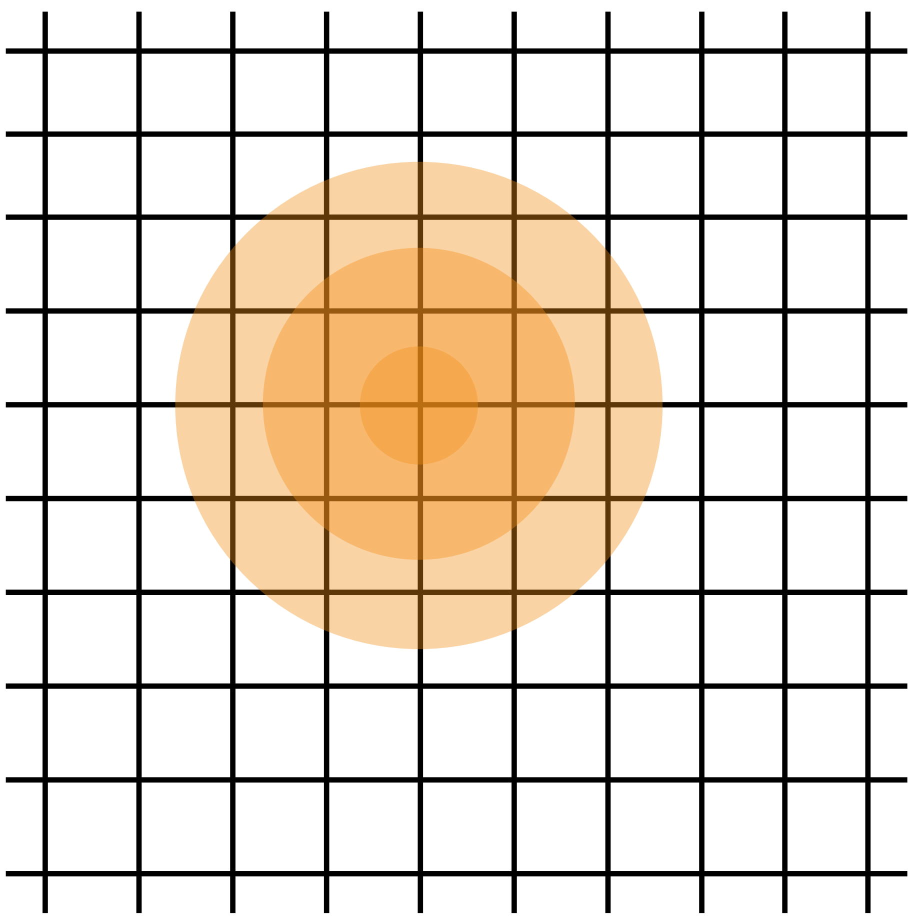


Burn-in time and correlation length dictated by Markov chain **‘autocorrelation time’**: shorter autocorrelation time implies less computational cost

Generate QCD gauge fields

QCD gauge field configurations sampled via

Hamiltonian dynamics + Markov Chain Monte Carlo



Updates diffusive

Lattice spacing $\rightarrow$ 0

Number of updates to change fixed physical length scale $\rightarrow$ ∞

"Critical slowing-down" of generation of uncorrelated samples

Generate QCD gauge fields

QCD gauge field configurations sampled via

Hamiltonian dynamics + Markov Chain Monte Carlo

"Critical slowing-down"
of generation of uncorrelated samples

Autocorrelation measure

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

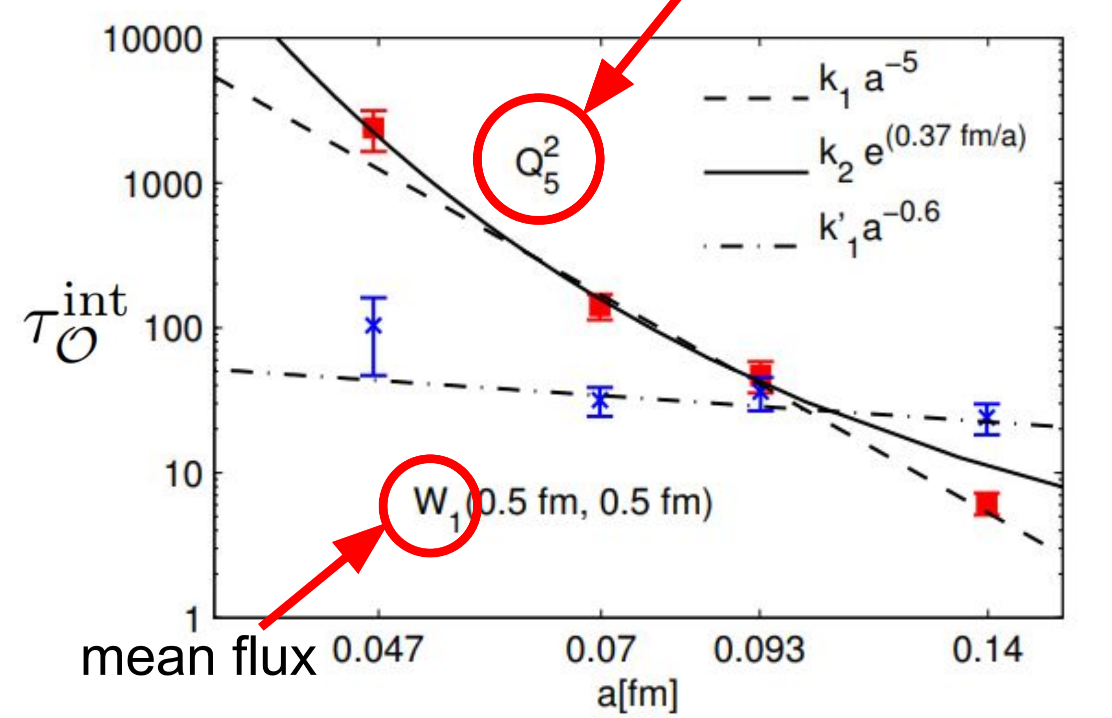
$$\tau_{\mathcal{O}}^{\text{int}} = \alpha_{\mathcal{O}} L^{z_{\mathcal{O}}}$$

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

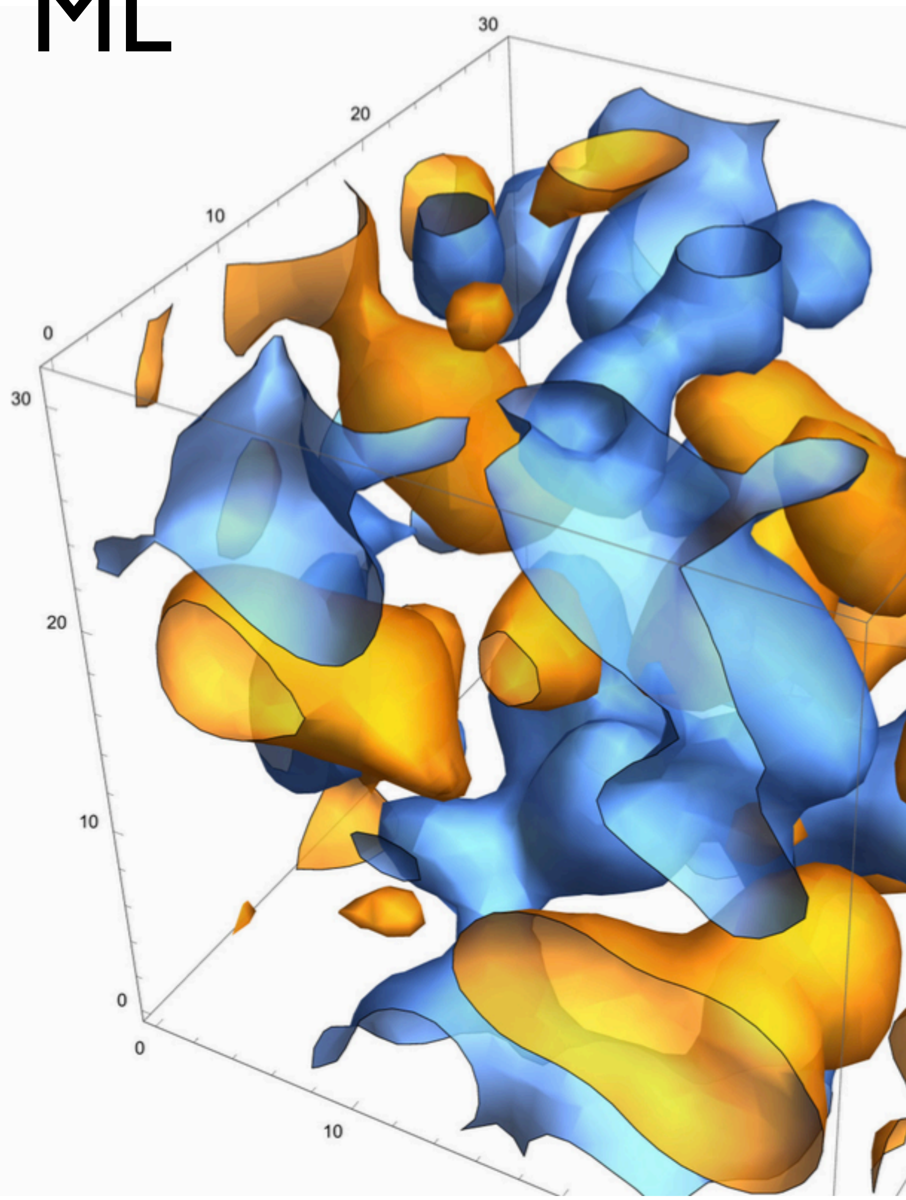
Machine learning QCD

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., PRD 100, 034515 (2019)
   [MIT + Google DeepMind, arXiv:2002.02428]
   Kanwar et al., MIT-CTP/5181

3. Hybrid approaches



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

Machine learning QCD

Generative models for QCD gauge field generation

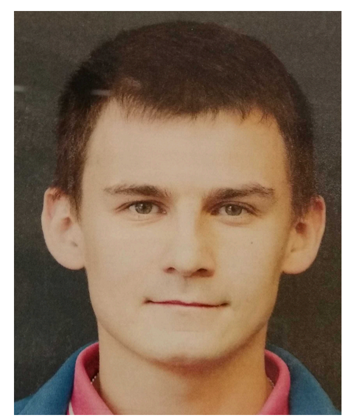
Massachusetts Institute of Technology

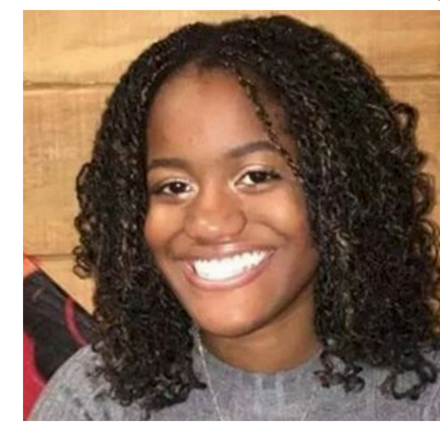
Google

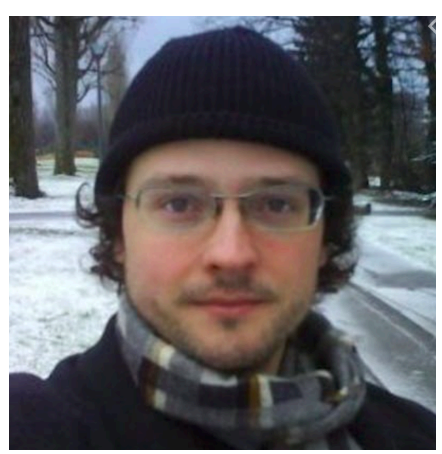
DeepMind

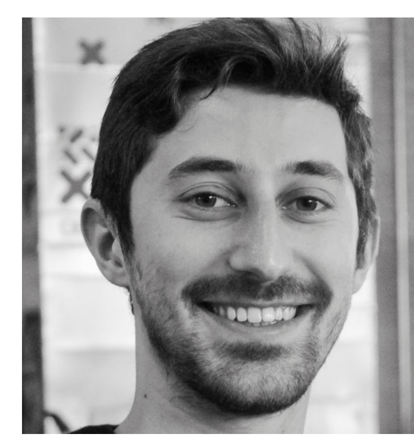
NYU





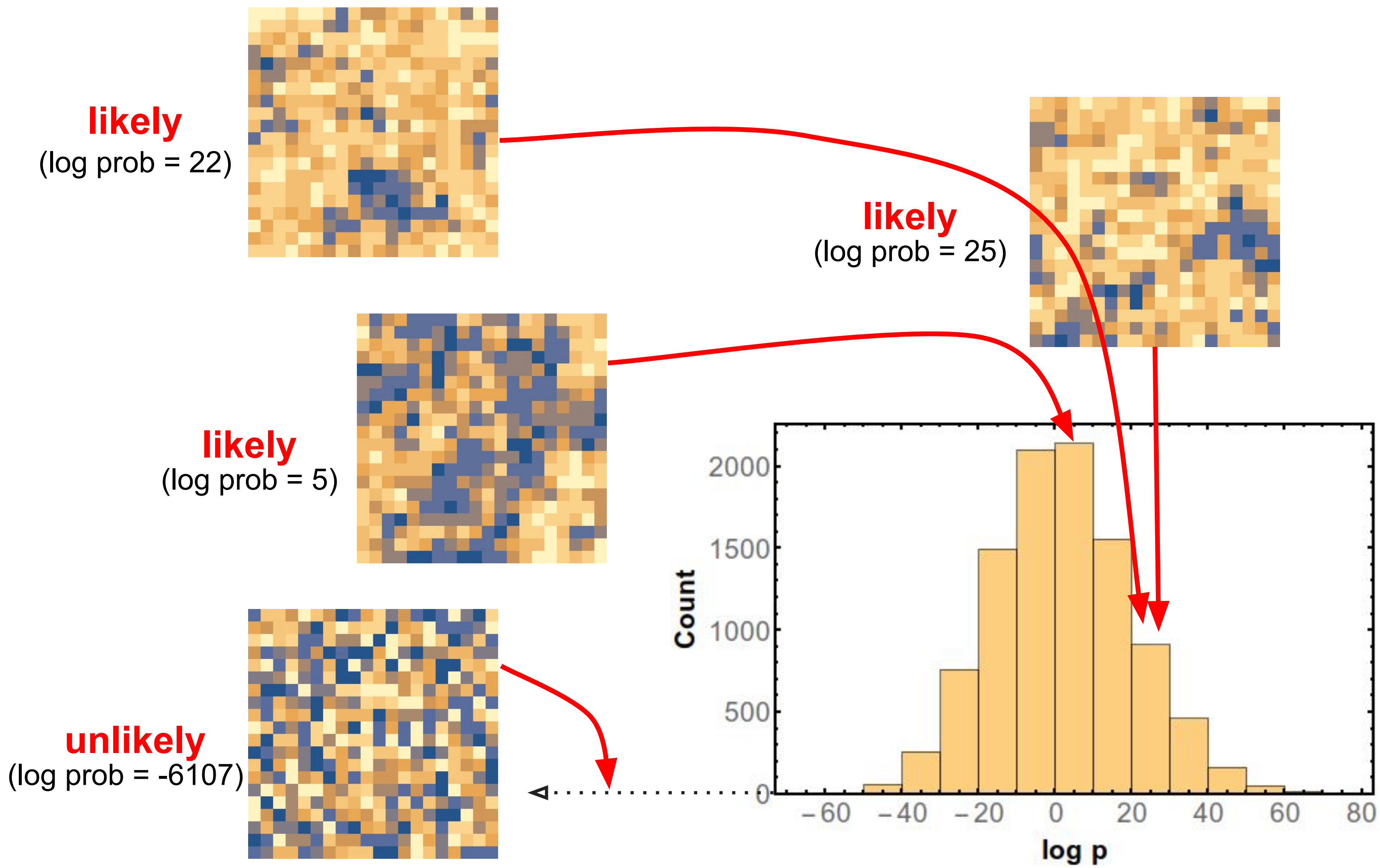



Sampling gauge field configs

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

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



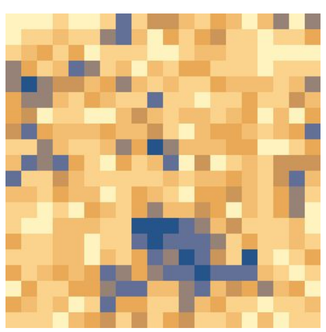
Sampling gauge field configs

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

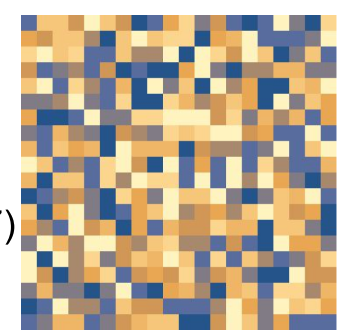
$$P[\phi(x)] \sim e^{-S[\phi(x)]}$$

Parallels with image generation problem

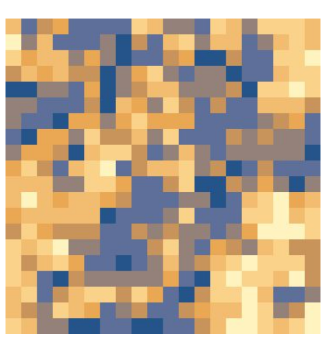
likely
(log prob = 22)



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

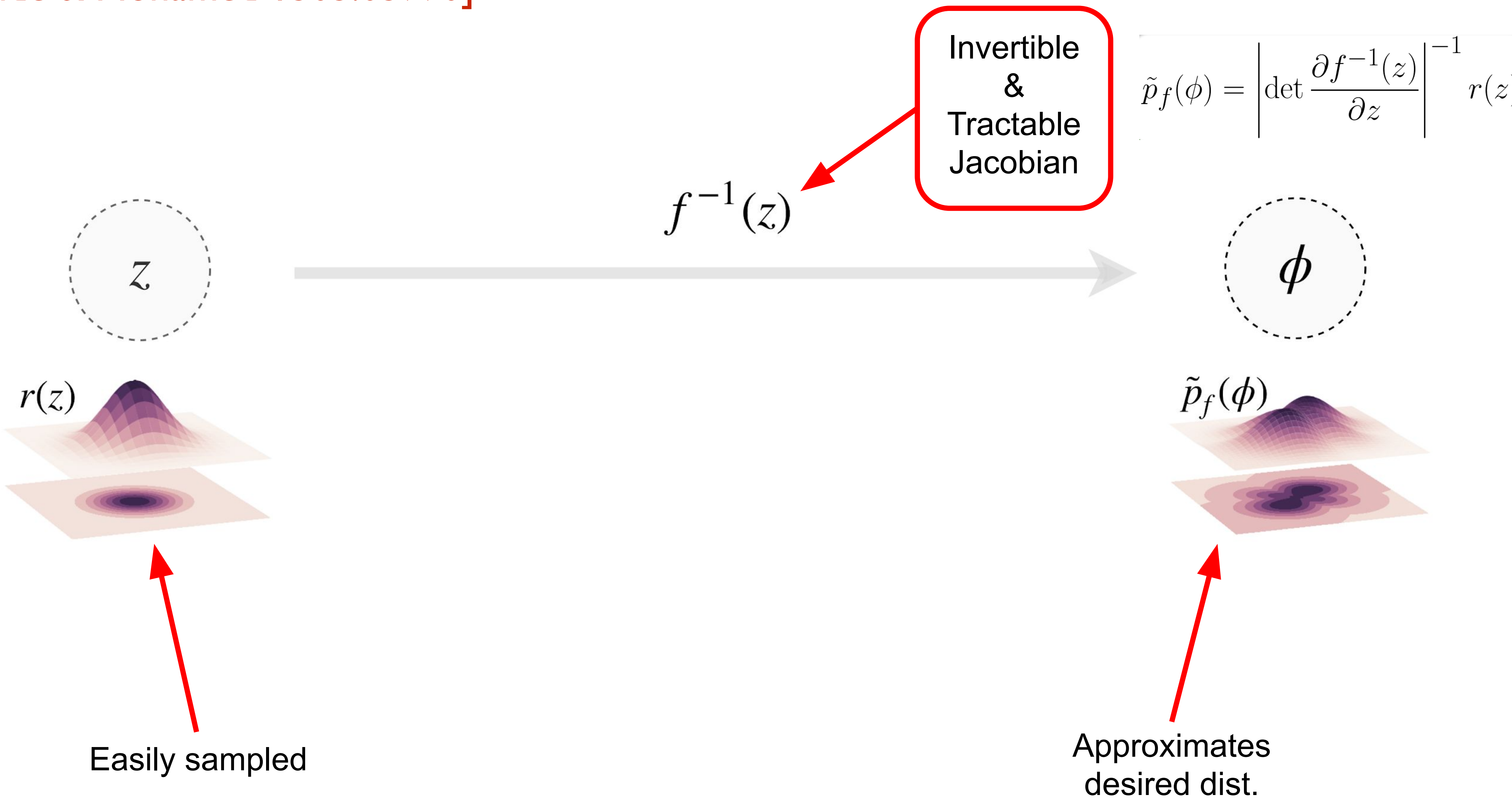
- 32 × 32 pixels × 3 cols ≃3000 numbers
- Translation-inva… within frame

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

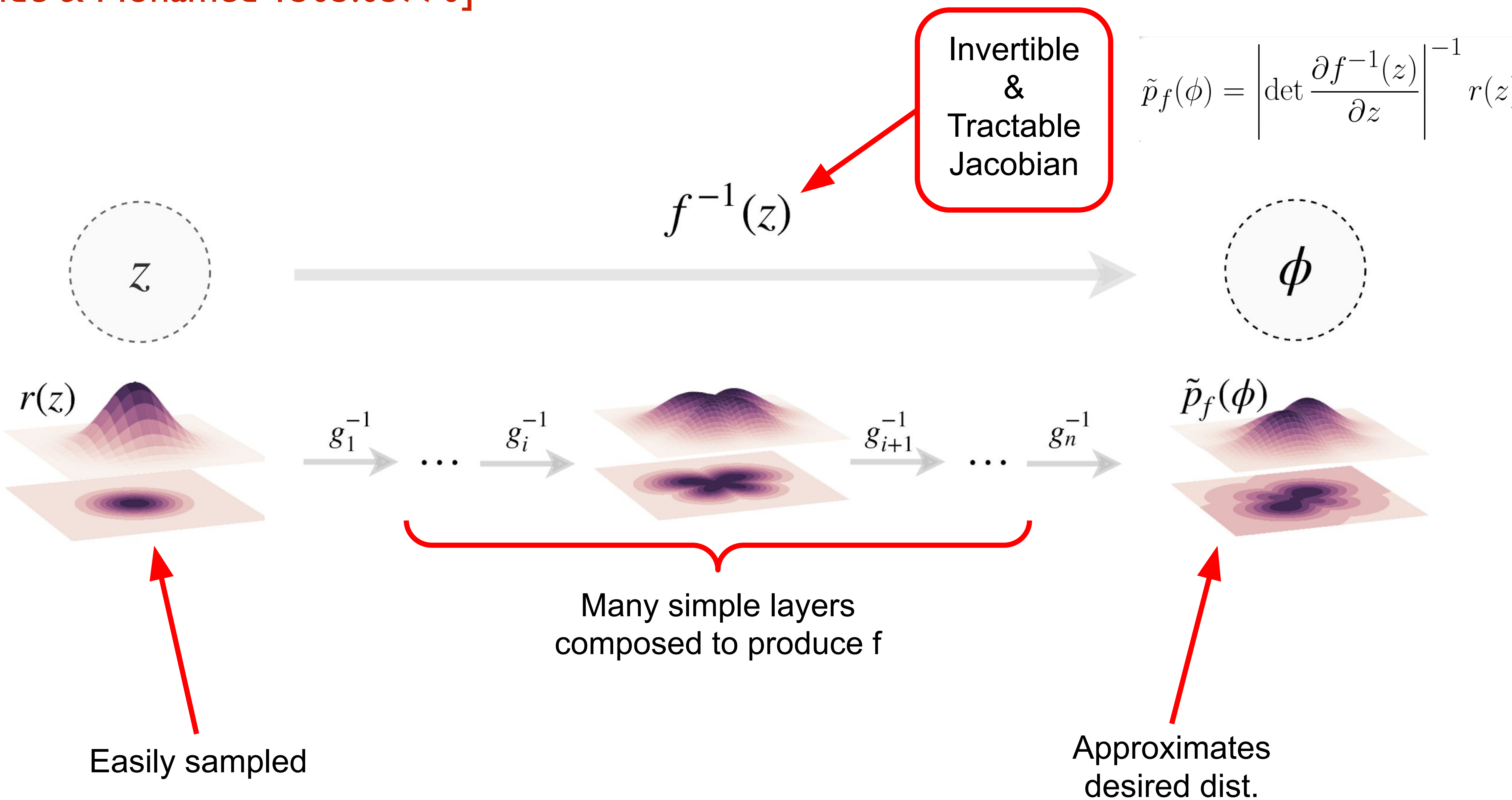
[Rezende & Mohamed 1505.05770]



Generative flow models

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

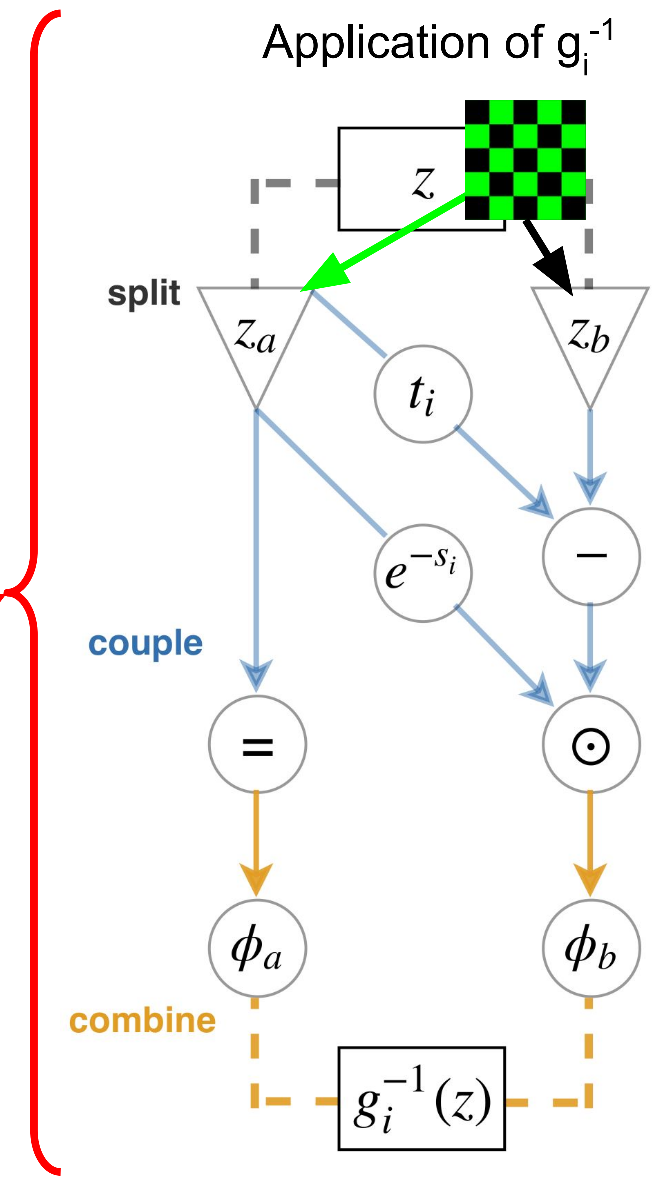
[Rezende & Mohamed 1505.05770]

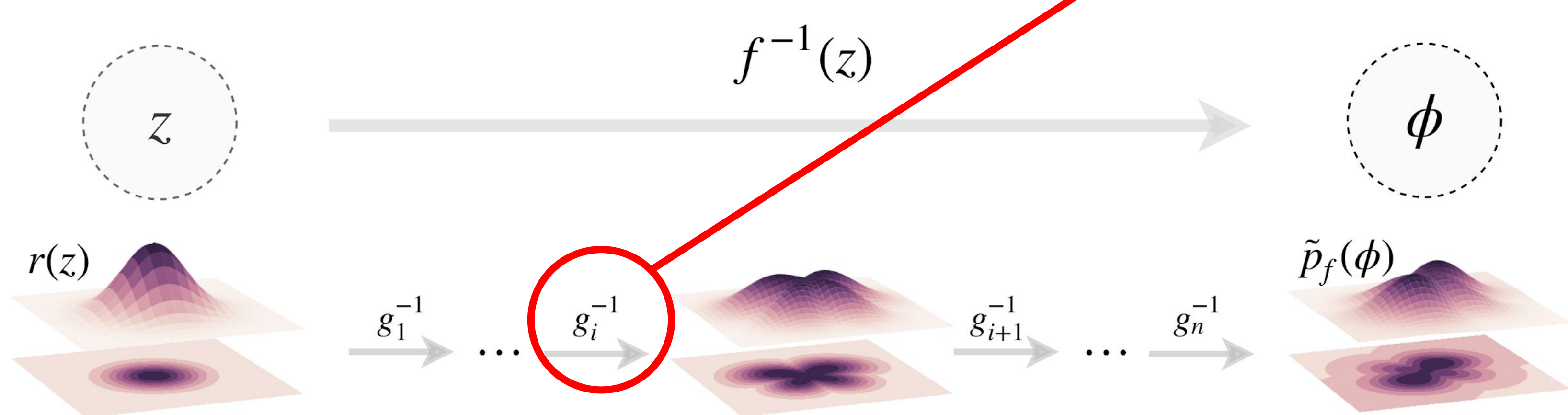


Generative flow models

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



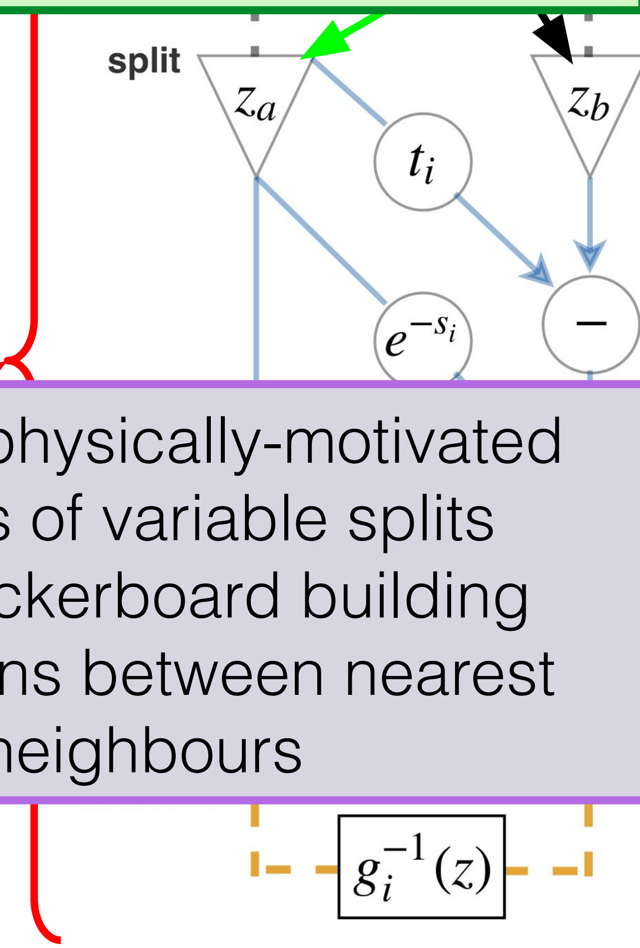


Generative flow models

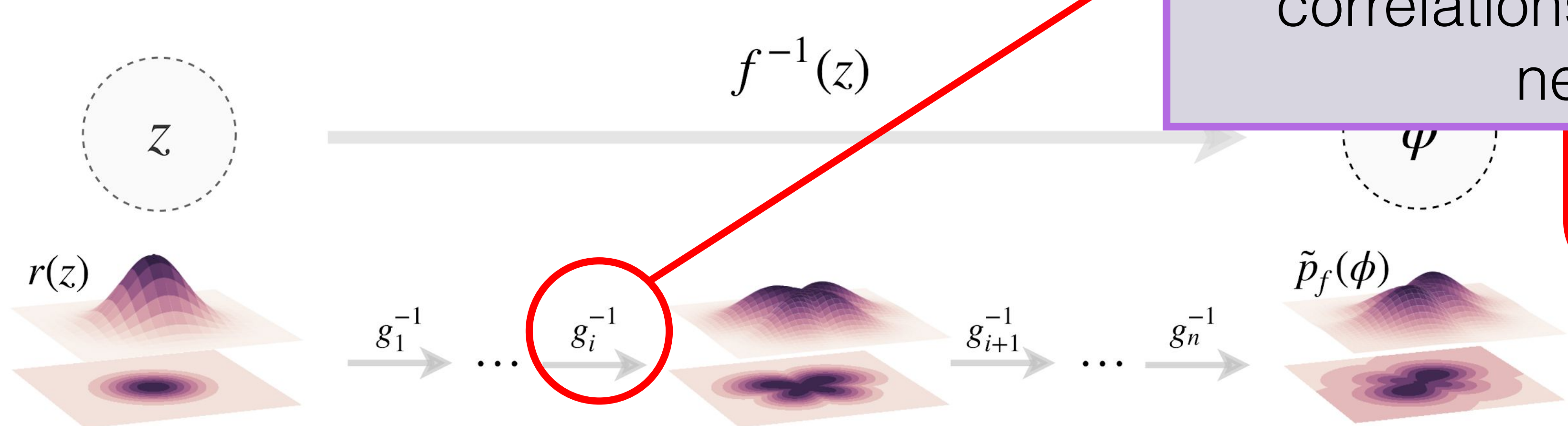
Choose real **non-volume preserving** flows:
[Dinh et al. 1605.08803]

- 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

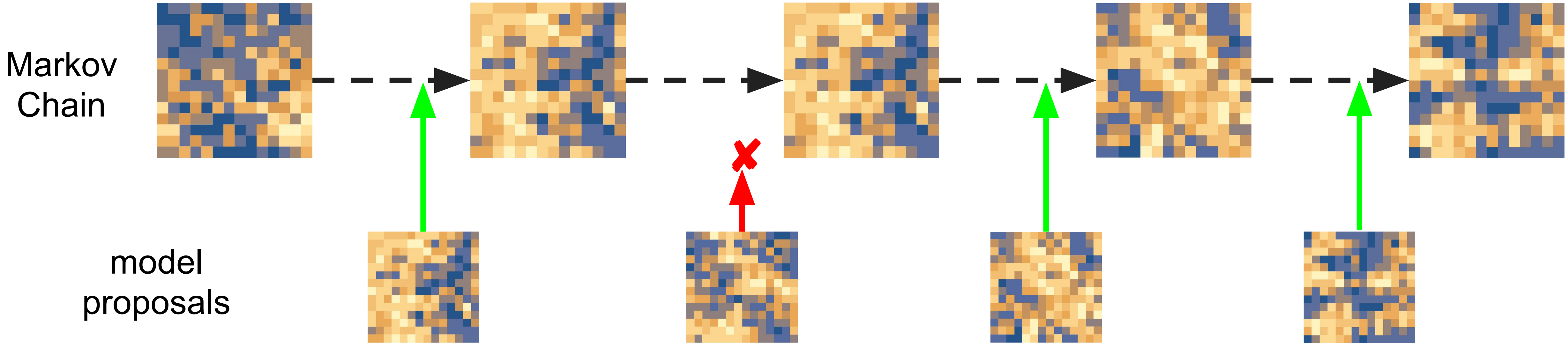
Exactness via Markov chain

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

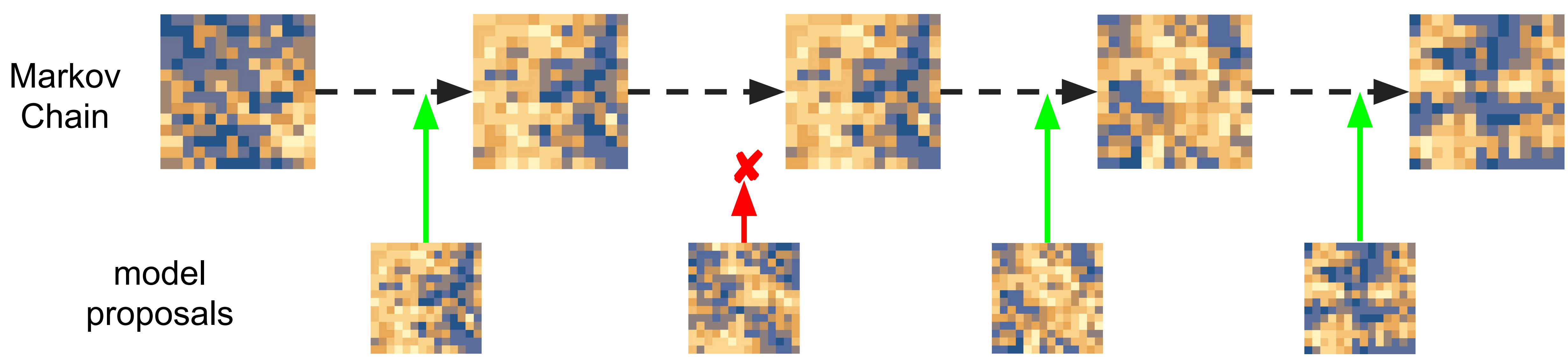
Acceptance probability

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

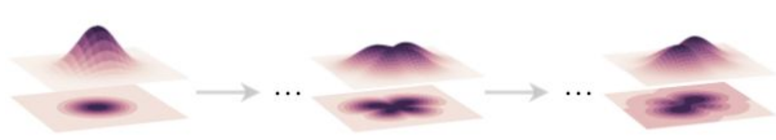
True dist

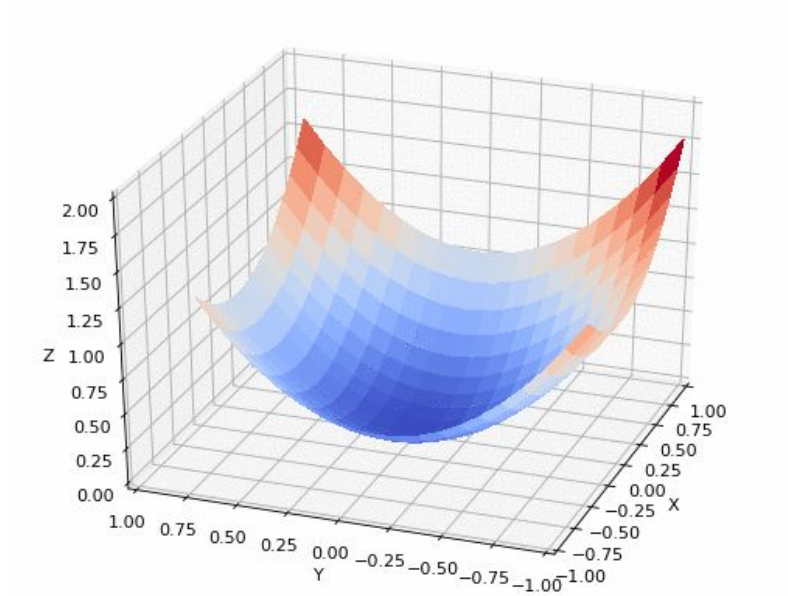
Model dist

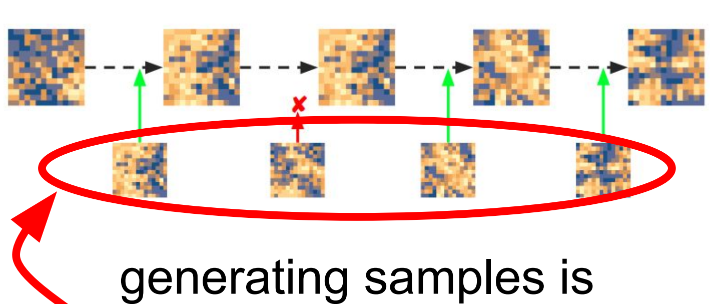
proposal independent
of previous sample



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)

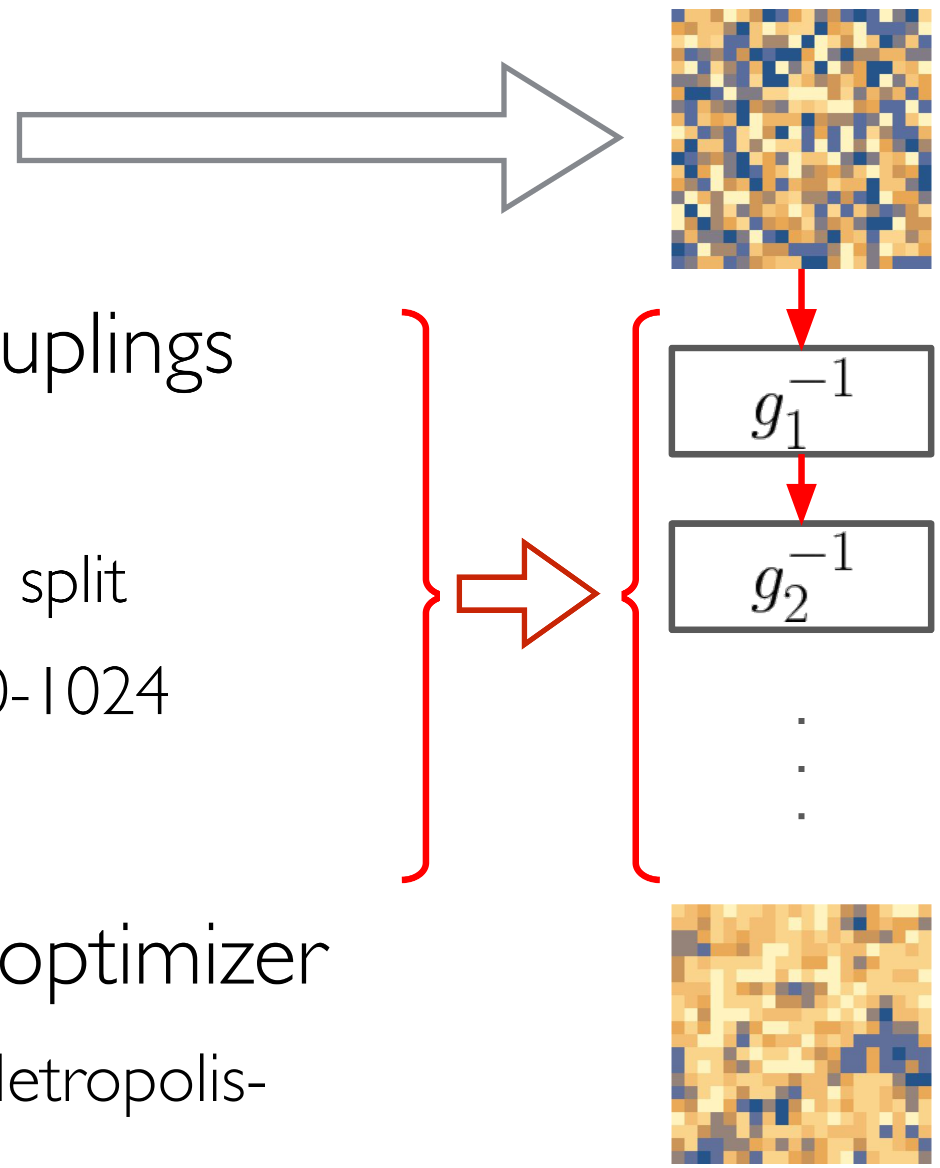
Application: scalar field theory

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

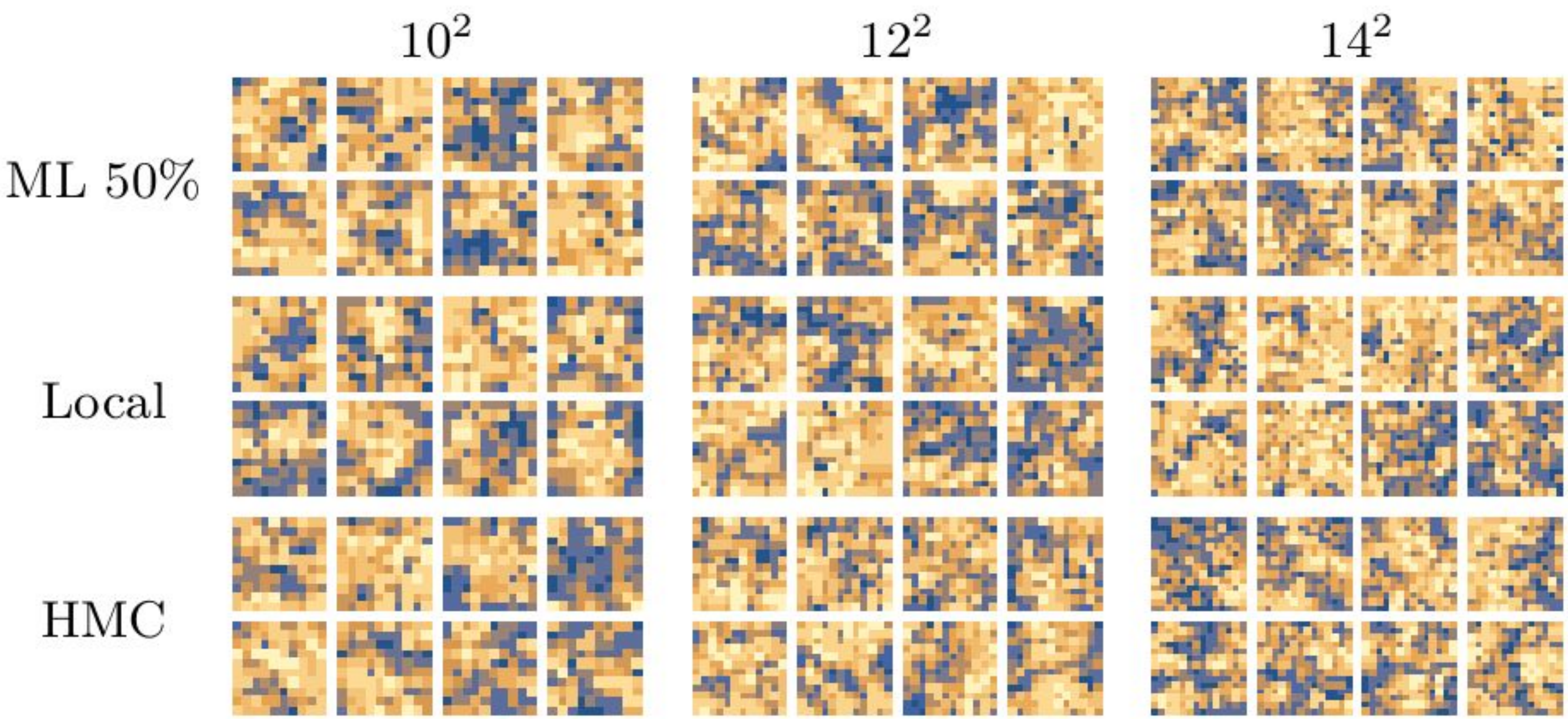


- Train using shifted KL loss with Adam optimizer
  - Stopping criterion: fixed acceptance rate in Metropolis-Hastings MCMC

Application: scalar field theory

First application: scalar lattice field theory

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

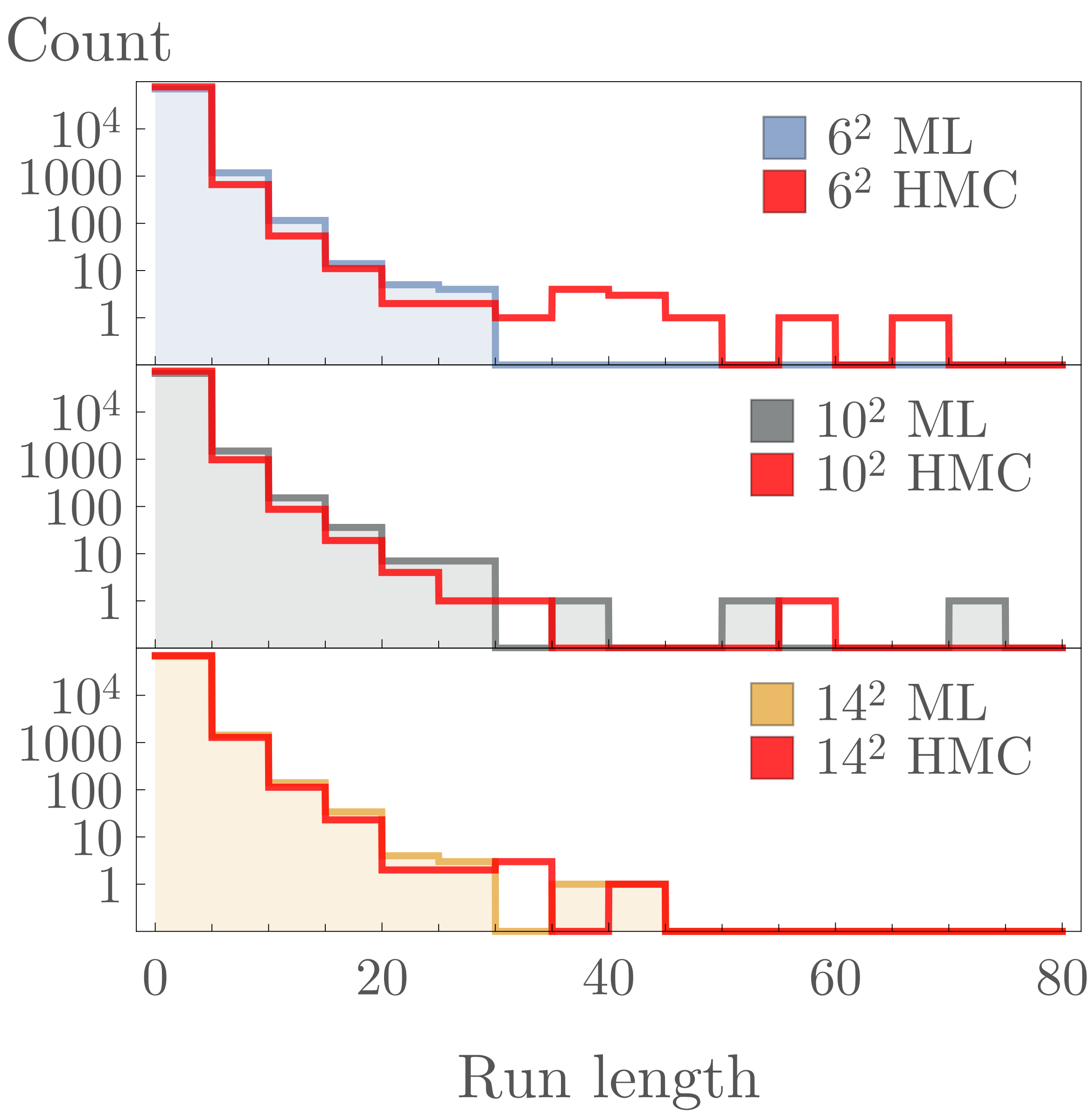


ML model produces **varied samples** and **correlations at the right scale**

Application: scalar field theory

First application: scalar lattice field theory

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

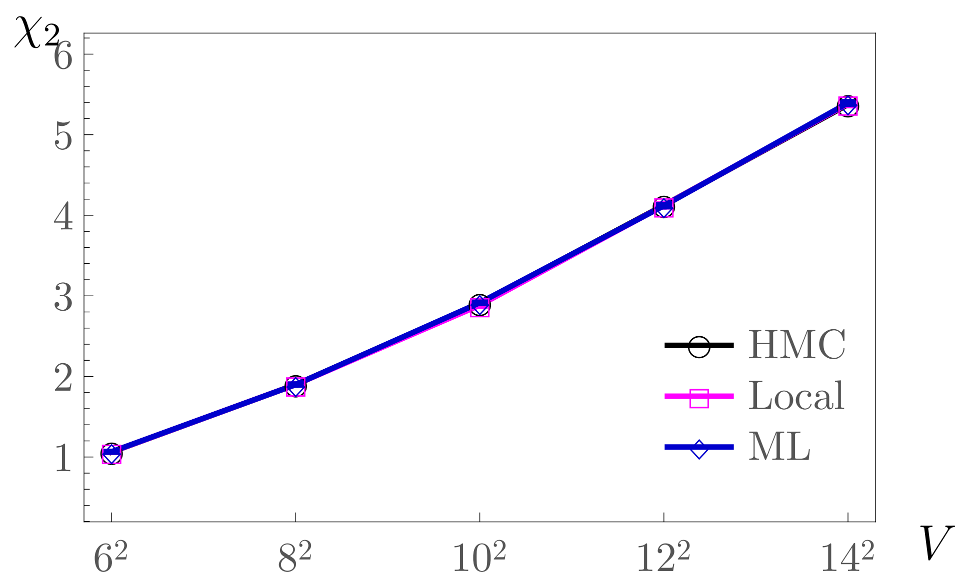


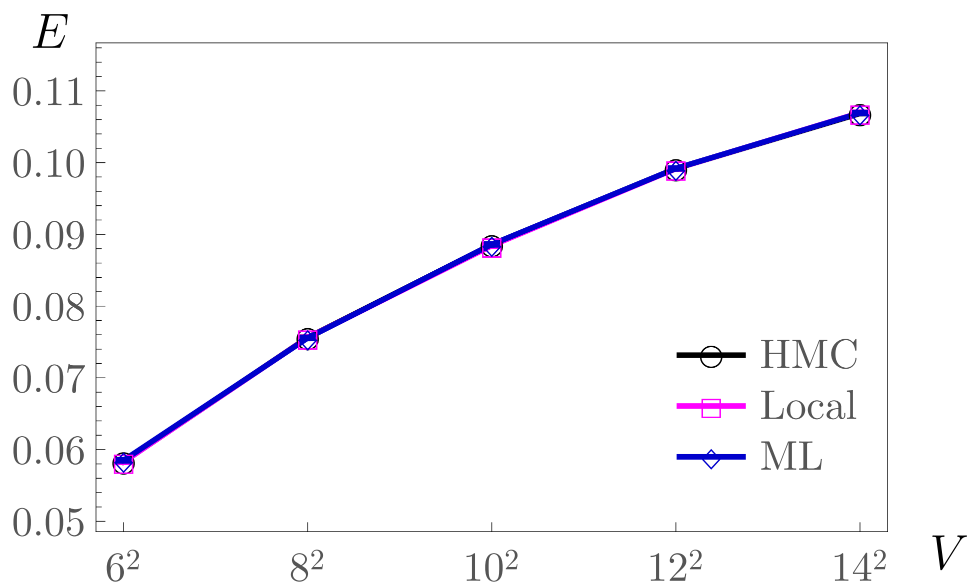
Rejectance runs in the Metropolis-Hastings accept/reject step are **comparable to those in Hamiltonian Monte-Carlo** tuned to same acceptance

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

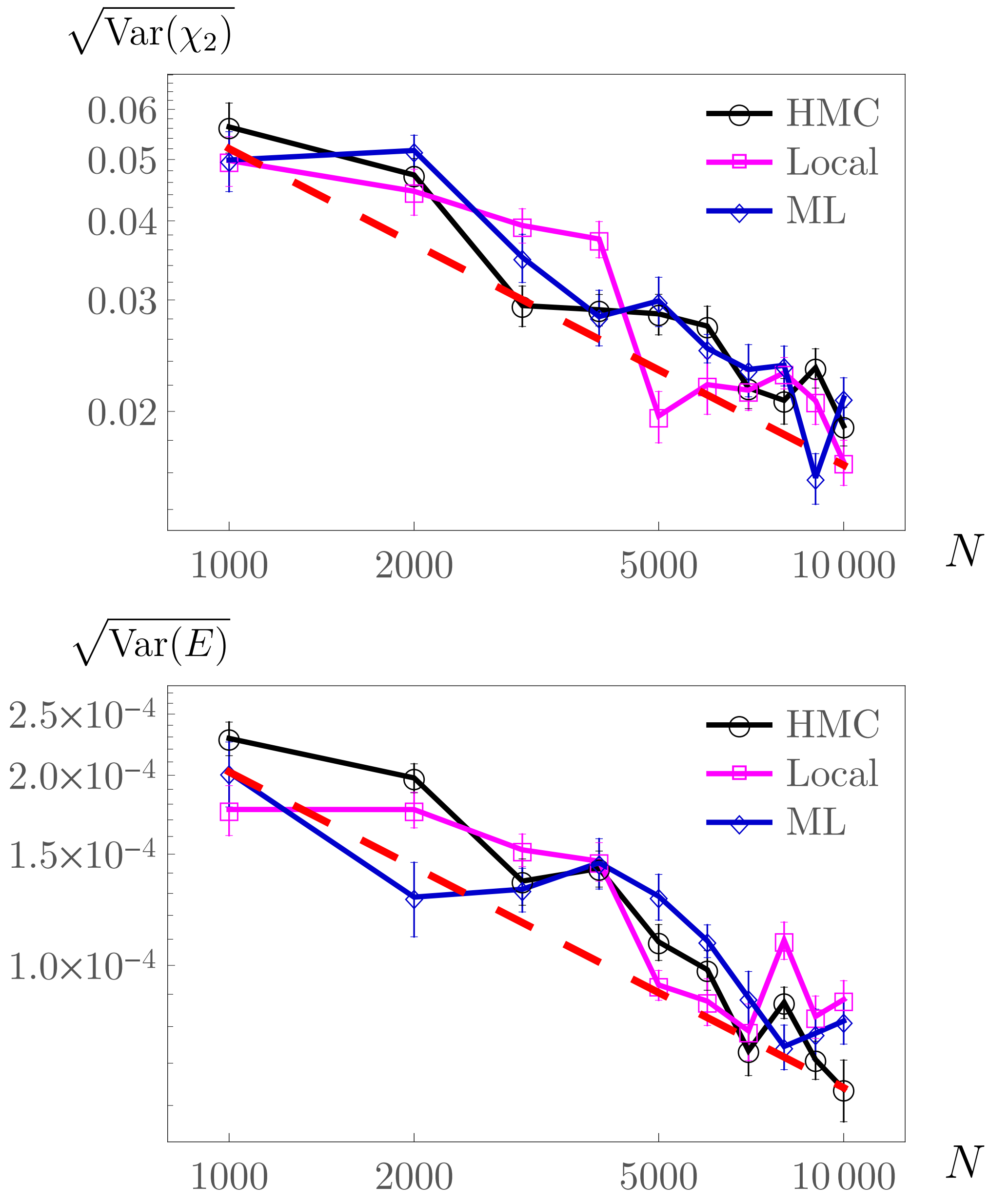
Two-point susceptibility $\chi_2 = \sum_x G_c(x)$

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:

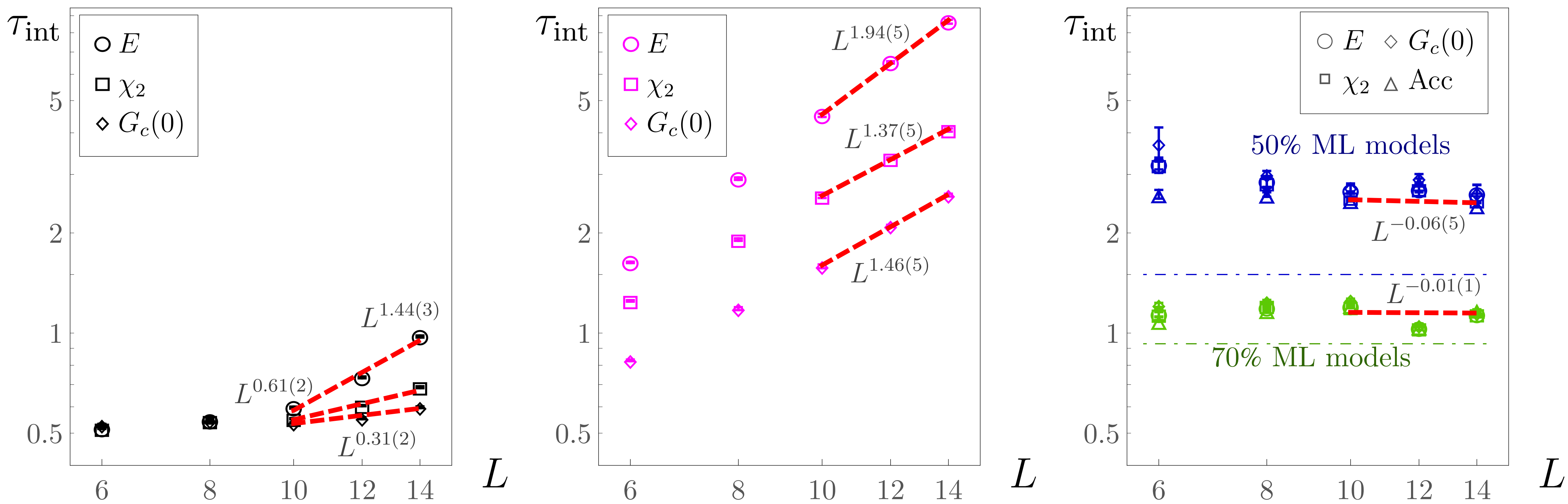
$$\propto 1/\sqrt{N}$$

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

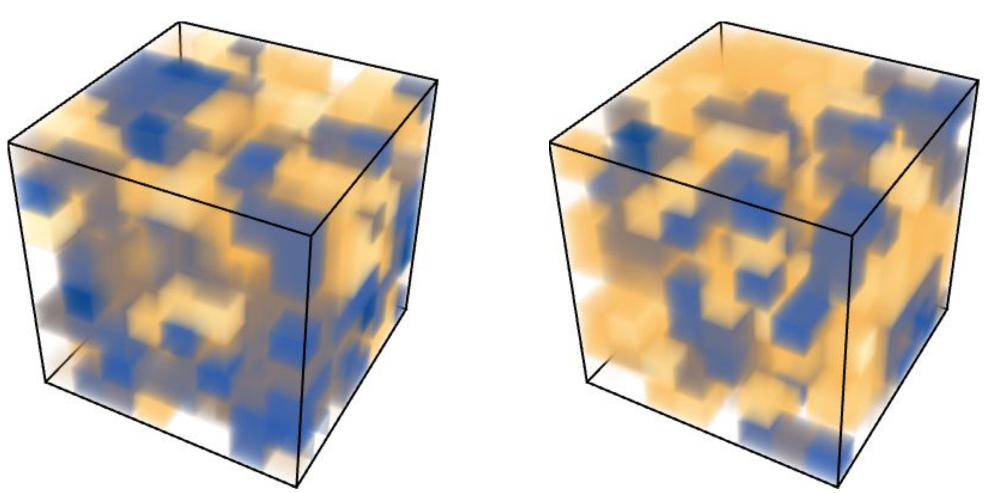
Dynamical critical exponents consistent with zero

Next steps: ML for LQCD

Target application: Lattice QCD for nuclear physics

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

[MIT + Google DeepMind, arXiv:2002.02428]



AURORA | EARLY SCIENCE PROGRAM FOR DATA & LEARNIN

Aurora21 Early Science Project Awarded

Next steps: ML for LQCD

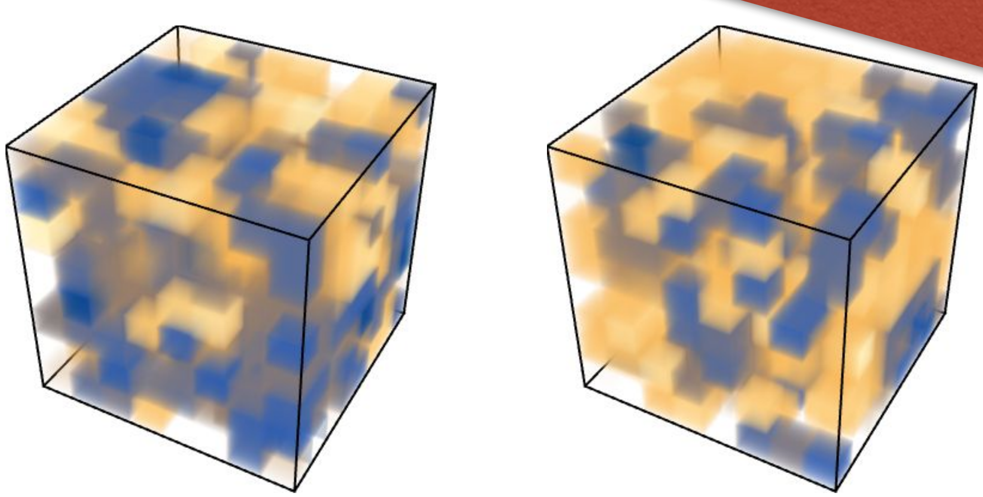
Target application: Lattice QCD for nuclear physics

...mber of dimensions

...grees of freedom

Out-of-the-box ML tools are not appropriate

Need custom ML for physics from the ground up



Aurora21 Early Science Project Awarded

Incorporating symmetries

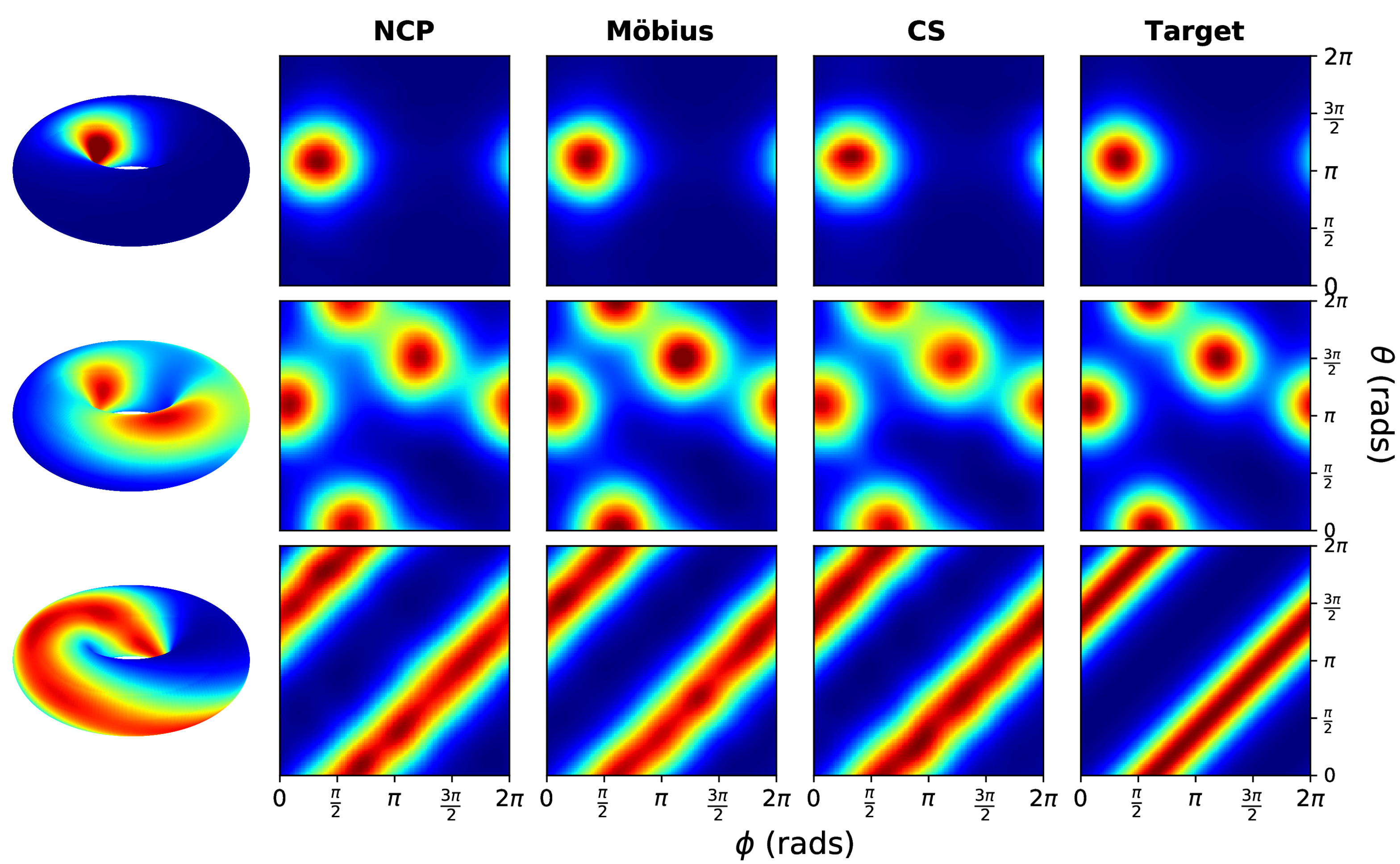
[MIT + Google DeepMind, arXiv:2002.02428]

Normalizing Flows on Tori and Spheres

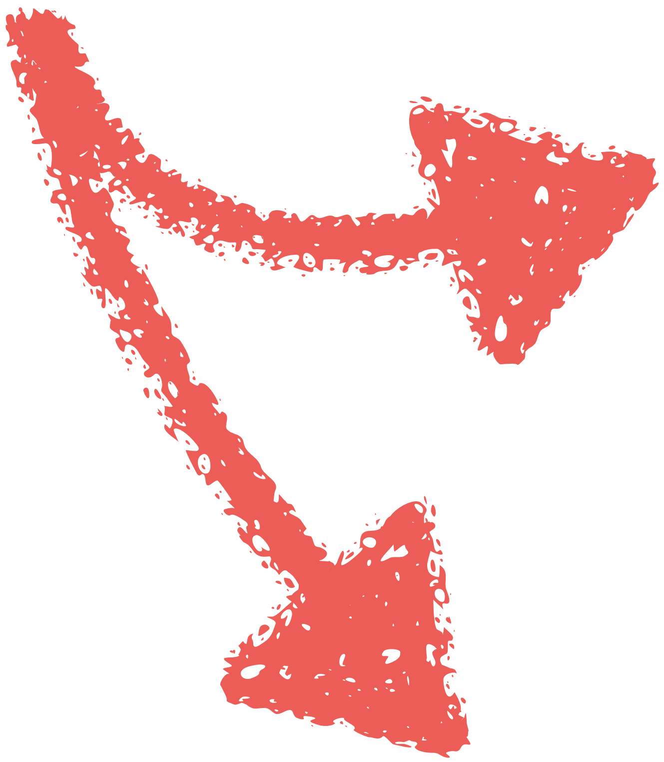
Danilo Jimenez Rezende * 1 **George Papamakarios** * 1 **Sébastien Racanière** * 1 **Michael S. Albergo** 2
Gurtej Kanwar 3 **Phiala E. Shanahan** 3 **Kyle Cranmer** 2

Arbitrarily flexible model architectures designed for compact and connected manifolds

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



Interdisciplinary applications



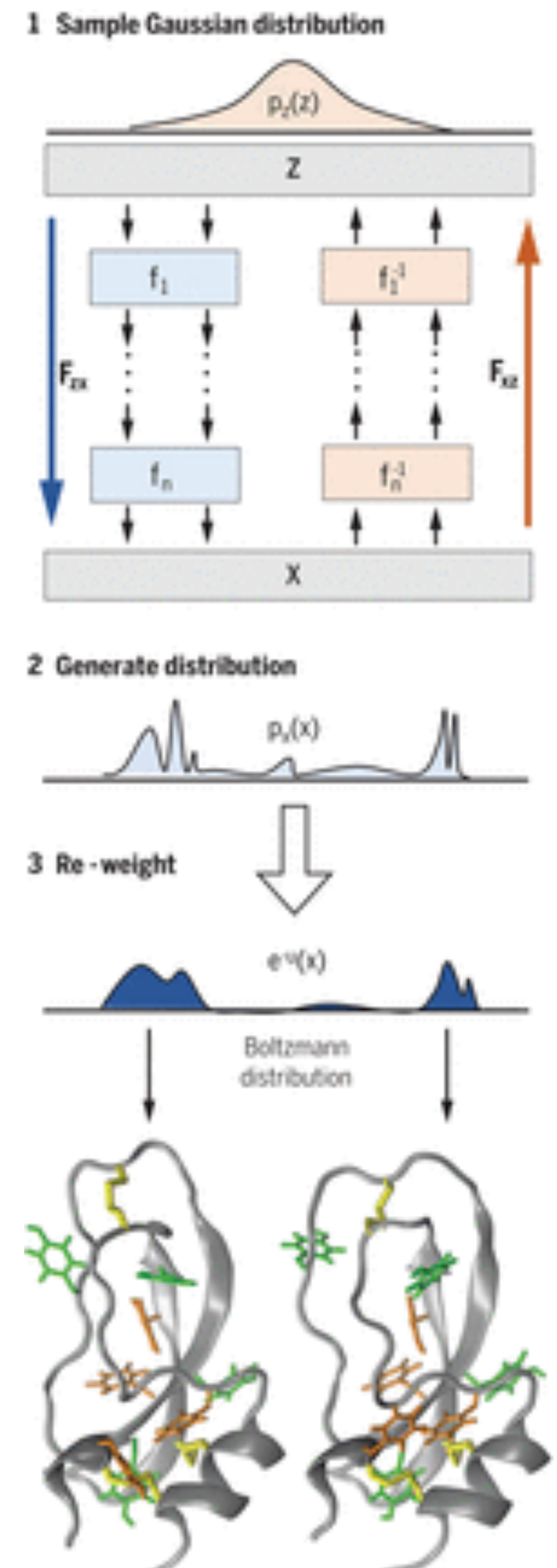
Molecular genetics and drug design

RESEARCH ARTICLE SUMMARY

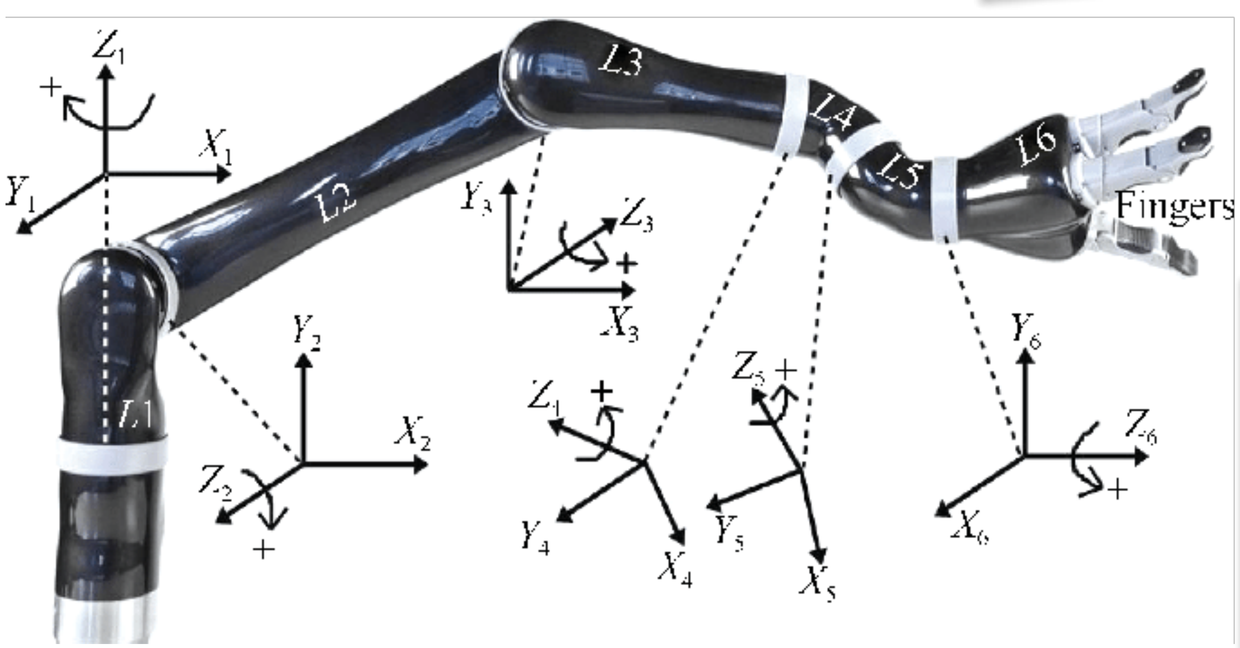
MACHINE LEARNING

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

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

Joint software effort

Our codes exploit and extend existing ML software frameworks

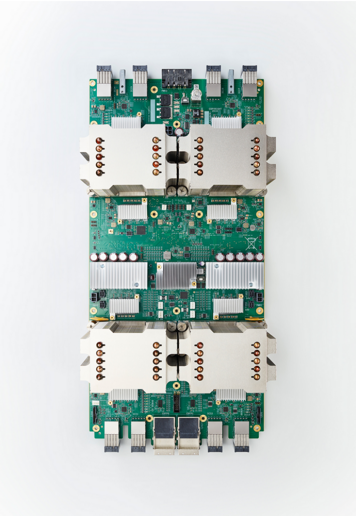
- Tensorflow
- Pytorch
- JAX

Active research projects into training protocols:

- Pruning
- Lottery tickets
- Initialisation frameworks
- …

We run on

- CPUs
- GPUs
- TPUs



Targeting exascale hardware for nuclear physics projects

Outlook

IF a generative flow model can be trained for QCD

After the up-front cost of training the model, it is

- Cheap to generate an arbitrarily large ensemble
- No need to store configurations, only the trained model
- Volume scaling is tractable via hierarchical flow and transfer learning approach
- Cheap to re-train the model to move to nearby parameter values (quark masses, beta)

i.e., if possible, this approach would have significant advantages, even if initial training is expensive

Case study: *Ab-initio* AI Center

CAIFI: The Center for Artificial Intelligence and Fundamental Interactions

"χφ"

LNS (Laboratory for Nuclear Science)	MIT	Harvard	Northeastern University / Tufts
William Detmold Philip Harris Phiala Shanahan Kerstin Perez Tracy Slatyer Washington Taylor Jesse Thaler Mike Williams	Lisa Barsotti Isaac Chuang Marin Soljacic Max Tegmark Pulkit Agrawal Bill Freeman Alexander Rakhlin Justin Solomon	Edo Berger Cora Dvorkin Daniel Eisenstein Doug Finkbeiner Matthew Schwartz Demba Ba Yaron Singer Todd Zickler	James Halverson Brent Nelson Taritree Wongjirad

Senior Investigators: 20 Physicists + 7 AI Experts

Junior Investigators in steady state: ≈ 20 FTE Graduate Students
≈ 7 CAIFI Postdoctoral Fellows

Case study: *Ab-initio* AI Center

CAIFI: The Center for Artificial Intelligence and Fundamental Interactions

"χφ"

Nuclear physics @MIT

LNS
Laboratory for Nuclear Science

William Detmold
Philip Harris
Phiala Shanahan
Kerstin Perez
Tracy Slatyer
Washington Taylor
Jesse Thaler
Mike Williams

MIT

Lisa Barsotti
Isaac Chuang
Marin Soljacic
Max Tegmark

Pulkit Agrawal
Bill Freeman
Alexander Rakhlin
Justin Solomon

HARVARD

Edo Berger
Cora Dvorkin
Daniel Eisenstein
Doug Finkbeiner
Matthew Schwartz

Demba Ba
Yaron Singer
Todd Zickler

NORTHEASTERN UNIVERSITY

James Halverson
Brent Nelson

Taritree Wongjirad

Senior Investigators: 20 Physicists + 7 AI Experts

Junior Investigators in steady state: ≈ 20 FTE Graduate Students
≈ 7 CAIFI Postdoctoral Fellows

Case study: *Ab-initio* AI Center

CAIFI: The Center for Artificial Intelligence and Fundamental Interactions

“χφ”

Our Vision:

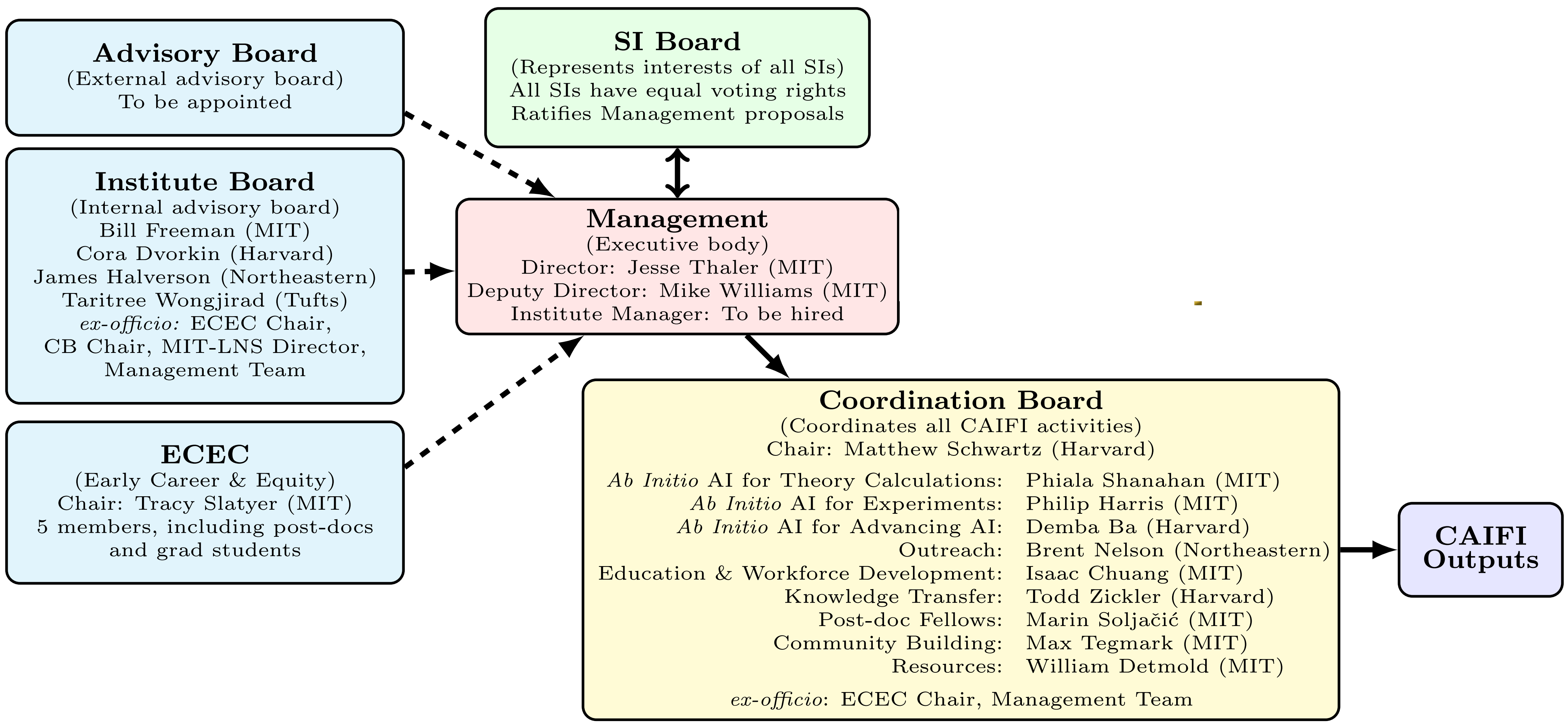
Advance fundamental physics and foundational AI
Build strong multidisciplinary collaborations
Close contact between early-career researchers
Intellectual freedom for CAIFI Fellows
Training and education at intersection of Physics and AI
Access to state-of-the-art computing resources
Advocacy for shared solutions across subfields

- ★ Ab Initio AI for Theory Calculations
- ★ Ab Initio AI for Experiments
- ★ Ab Initio AI for Advancing AI

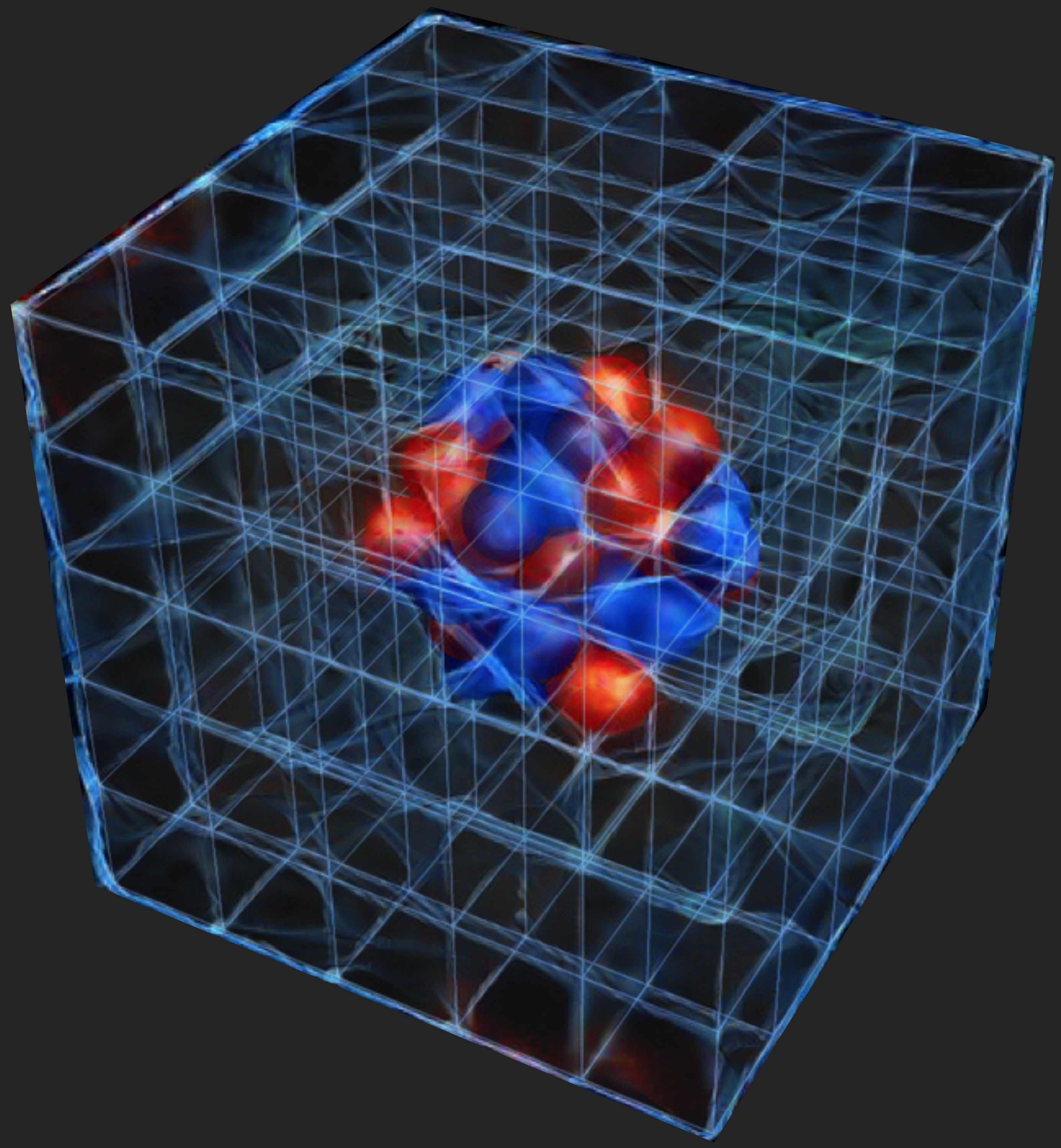
Case study: *Ab-initio* AI Center

CAIFI: The Center for Artificial Intelligence and Fundamental Interactions

"χφ"



Modeled after nuclear/particle experimental collaborations:
Dedicated coordinator for each major activity



Phiala Shanahan

Massachusetts Institute of Technology