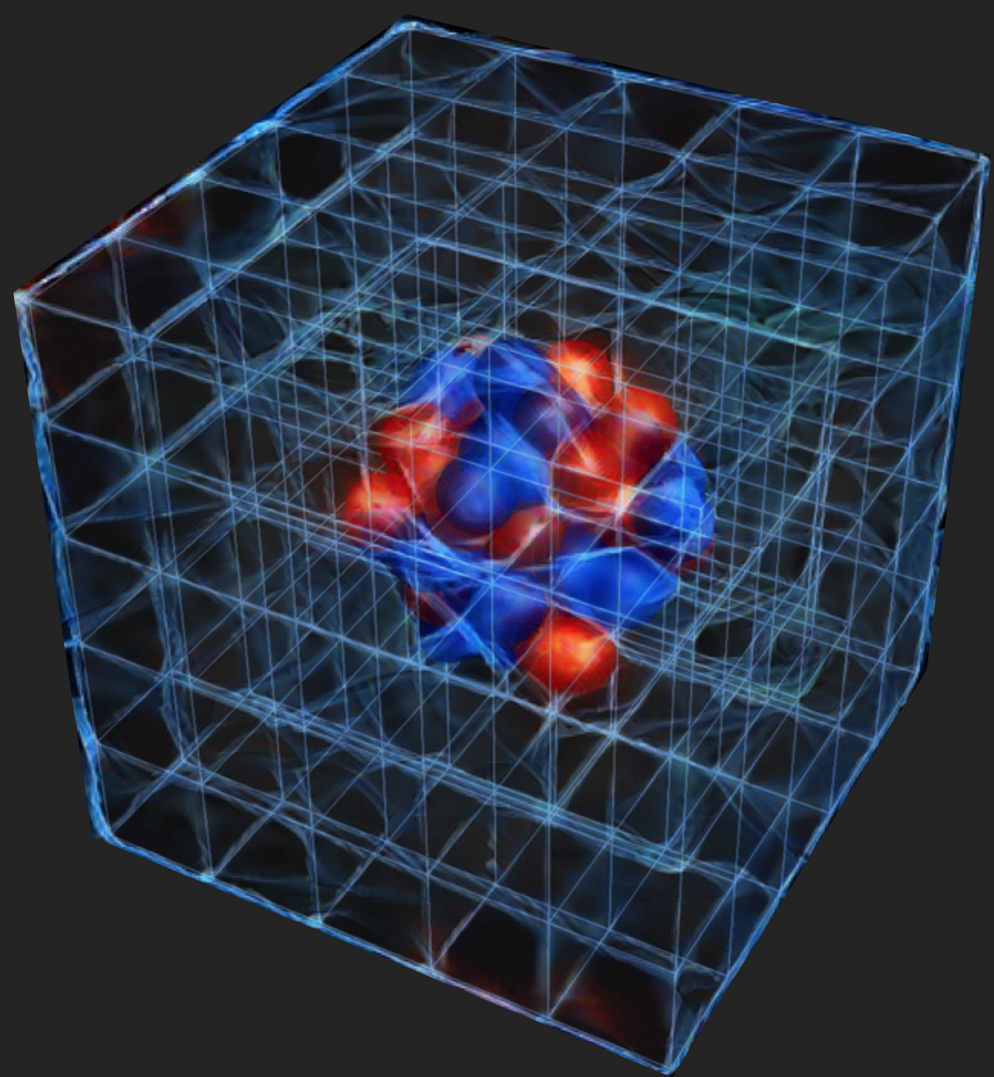
Machine learning for sampling in lattice field theory



Phiala Shanahan

http://iaifi.org/

Massachusetts Institute of Technology

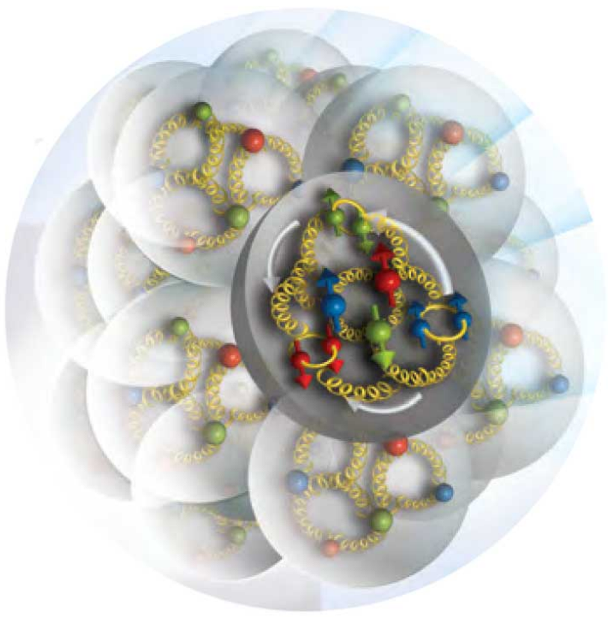
The structure of matter

One application of lattice field theory:

Hadron/nuclear structure and interactions from the Standard Model of particle physics

Emergence of complex structure in nature

Backgrounds and benchmarks for searches for new physics



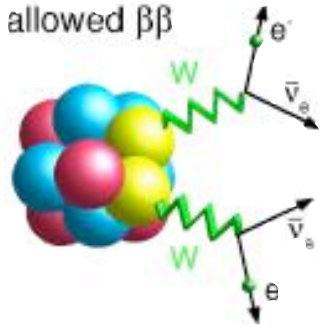
Many studies limited by available computation

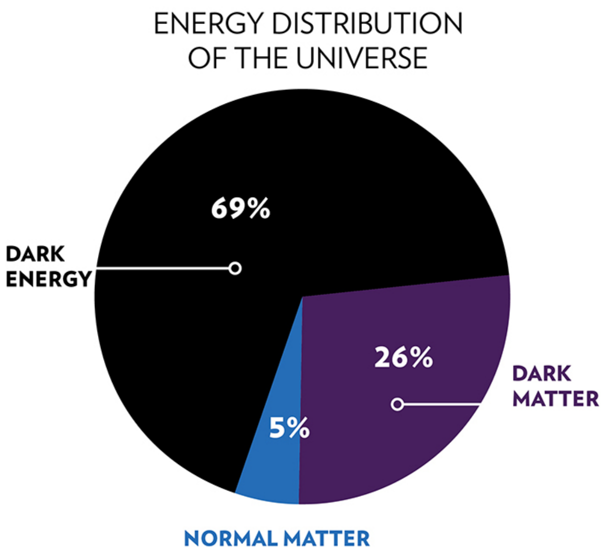
The search for new physics

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…





Exponential*factorial growth in computational cost with A

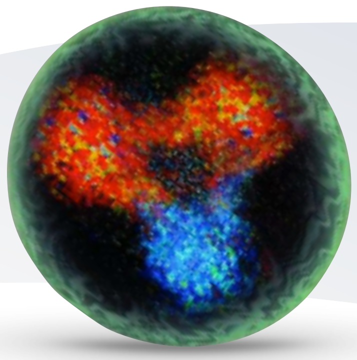
Strong interactions

Study nuclear structure from the strong interactions

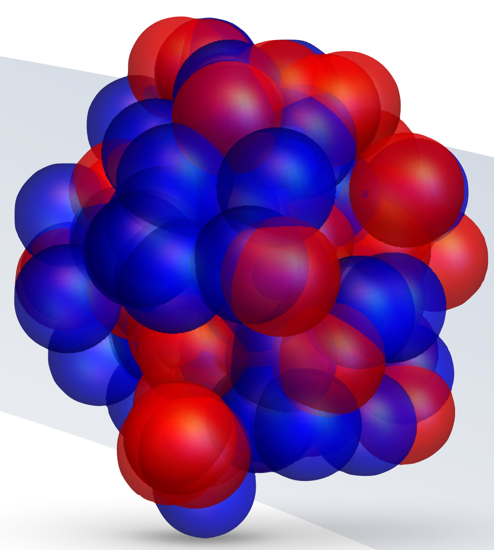
Quantum Chromodynamics (QCD)

Strongest of the four forces in nature
Non-perturbative in low-energy regime

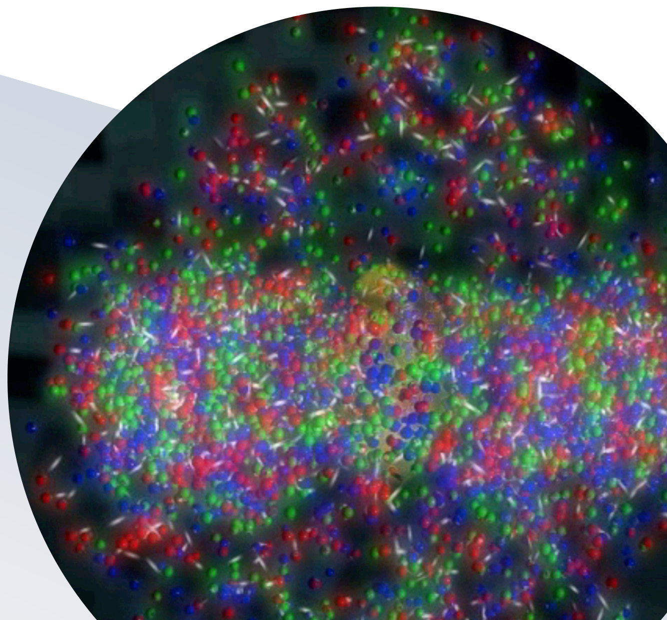
Forms other types of exotic matter e.g., quark-gluon plasma



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



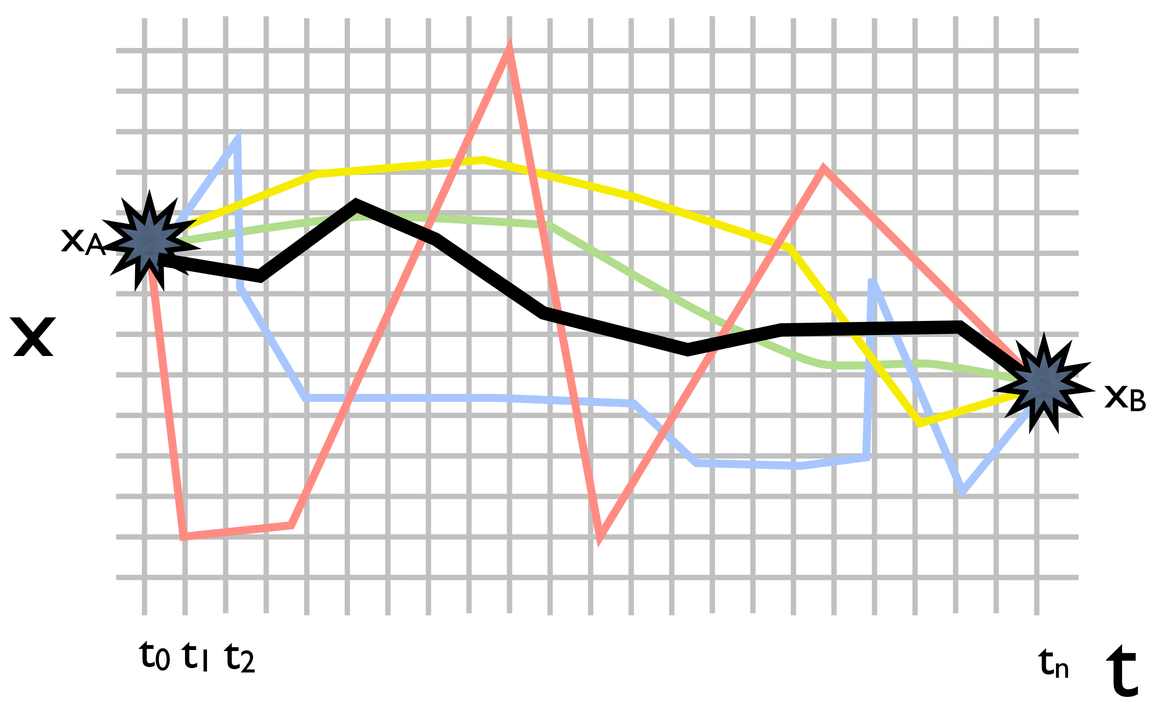
Binds protons and neutrons into nuclei



Lattice QCD

Numerical first-principles approach to
non-perturbative QCD

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

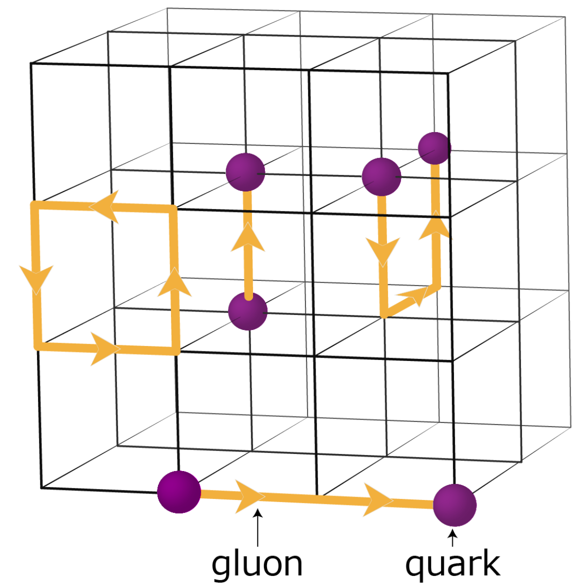


- Evaluate by importance sampling
- Paths near classical action dominate
- Calculate physics on a set (ensemble) of samples of the quark and gluon fields

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

Machine learning for LQCD

MACHINE LEARNING IS

E.g., 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

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

Machine learning for LQCD

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

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

E... ...n by e.g.,

...7 (2018)

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

•

Efficient computati...
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)

S. Blücher...
Boyda et al., 2009.1...

Sign-problem avoidance via contour deformation of path integrals
Alexandruet al., Phys. Rev. Lett. 121 (2020),
Detmold et al., 2003.05914 (2020)

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

Lattice QCD

Workflow of a lattice QCD calculation

1 Generate field configurations via Hybrid Monte Carlo

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

2 Compute propagators

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

3 Contract into correlation functions

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

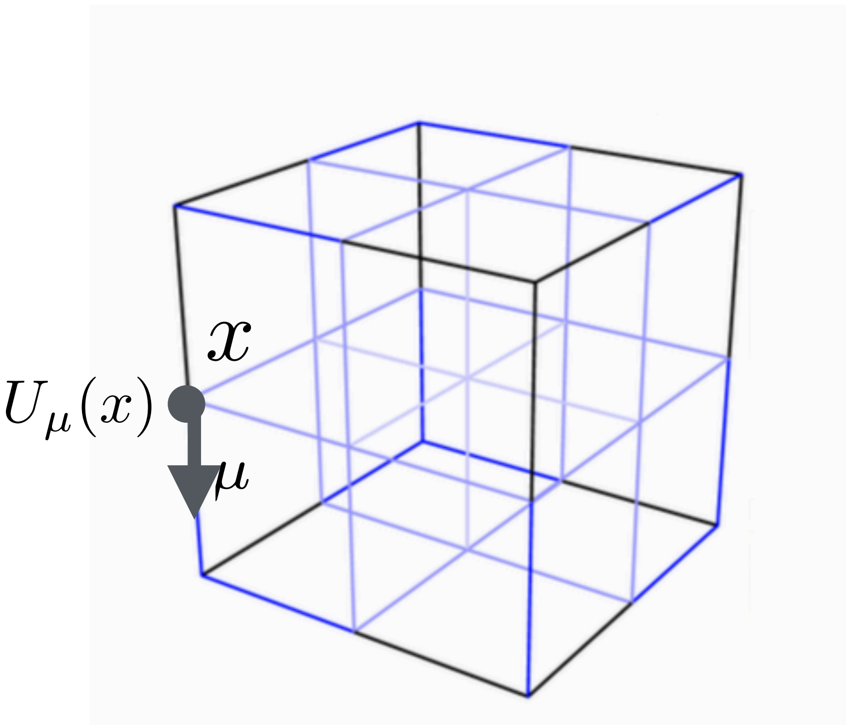
Computational cost grows exponentially with size of nuclear system

Lattice QCD

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

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

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



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

  ➡ Weighted averages over configurations determine physical observables of interest

- Calculations use $\sim 10^3$ configurations

Generate QCD gauge fields

QCD gauge field configurations sampled via

Hamiltonian dynamics + Markov Chain Monte Carlo

Molecular dynamics

Classical motion with

Randomly sampled

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

- 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

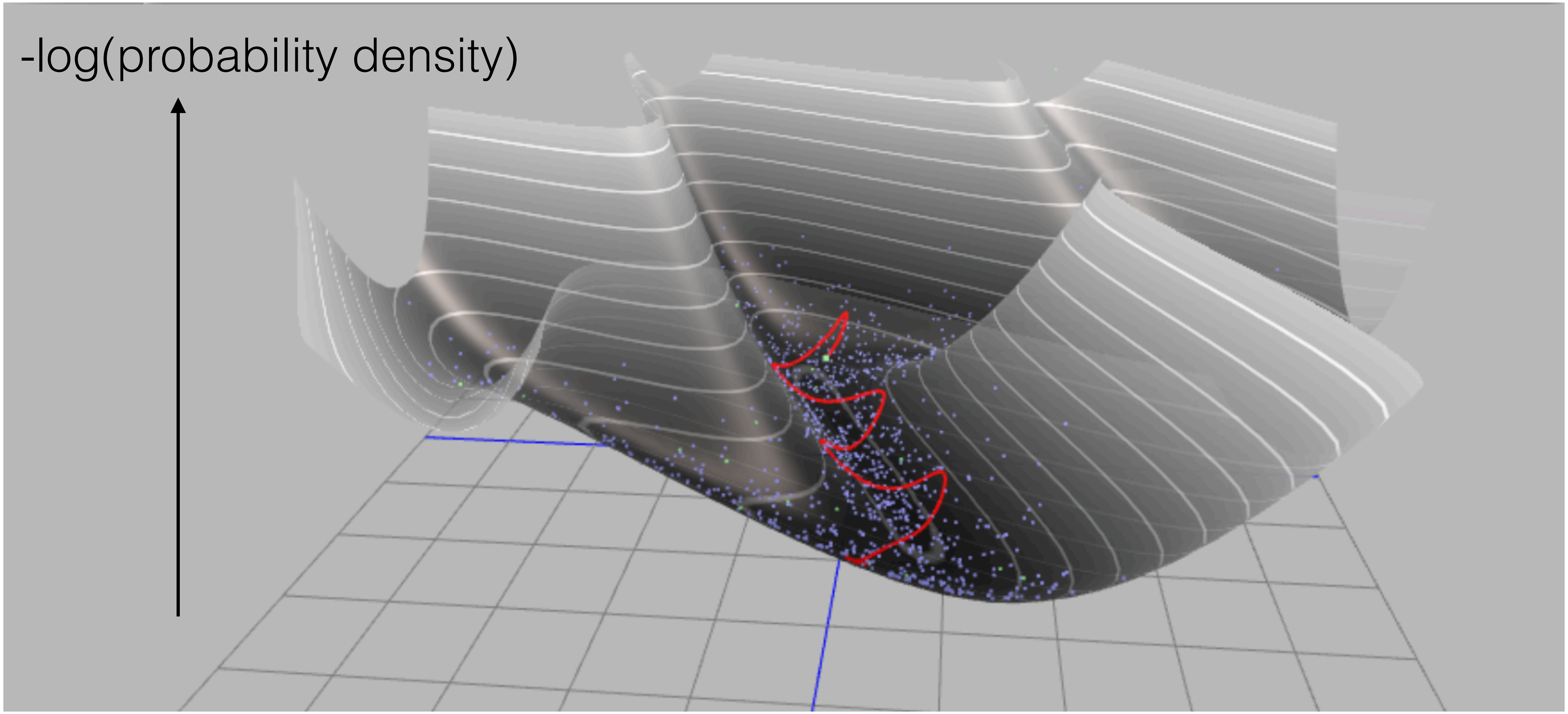
BUT

- Short trajectories for high acceptance

Generate QCD gauge fields

QCD gauge field configurations sampled via

Hamiltonian dynamics + Markov Chain Monte Carlo

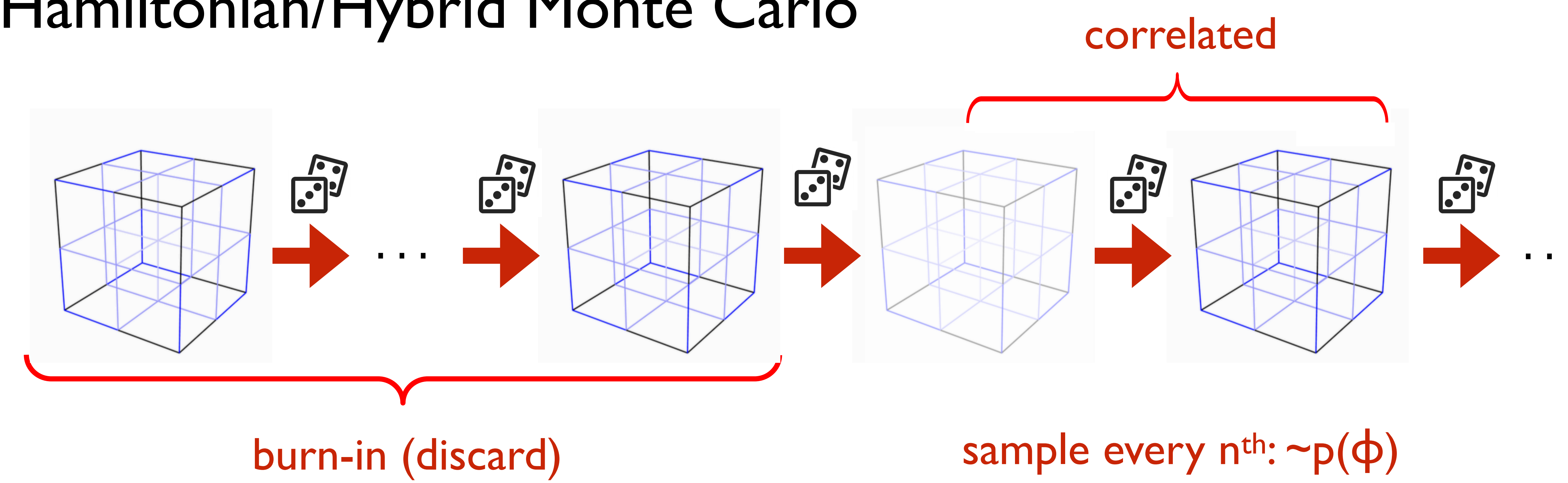


Generate QCD gauge fields

QCD gauge field configurations sampled via

Hamiltonian dynamics + Markov Chain Monte Carlo

Hamiltonian/Hybrid Monte Carlo

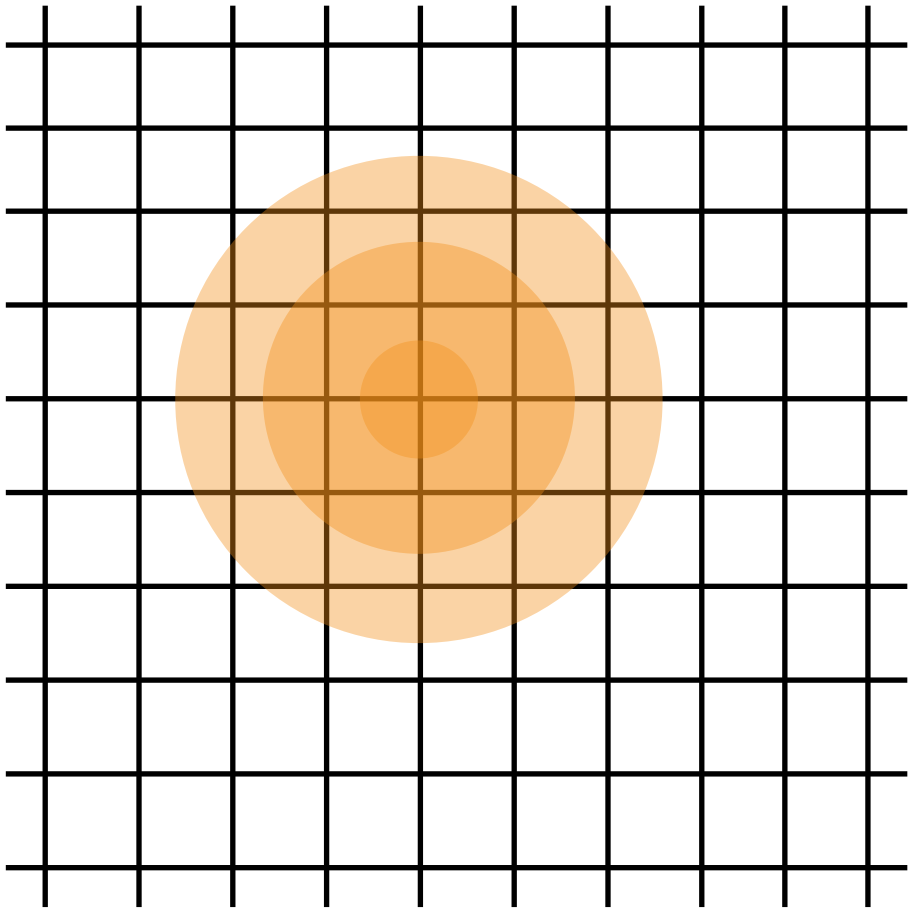


Burn-in time and correlation length dictated by Markov chain **'autocorrelation time'**: shorter autocorrelation time implies less computational cost

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

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

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

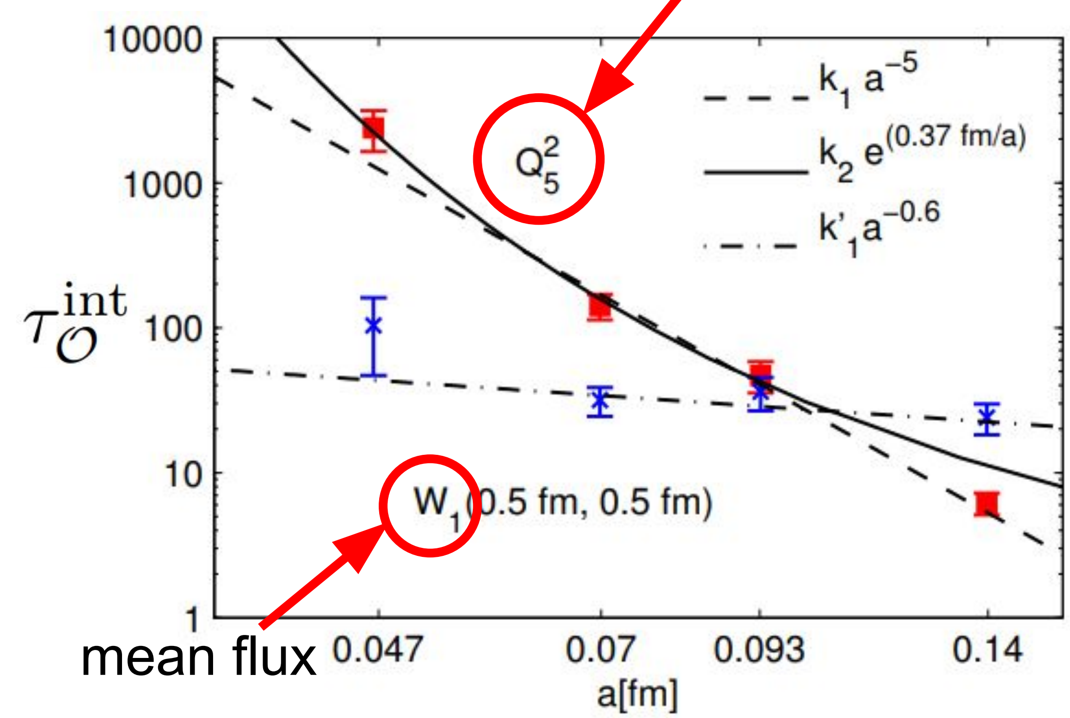
Correlation of observable $\mathcal{O}$ on configurations separated by τ Markov Chain steps

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

Critical exponent

topological charge

[Schaefer et al. / ALPHA collaboration 1009.5228]

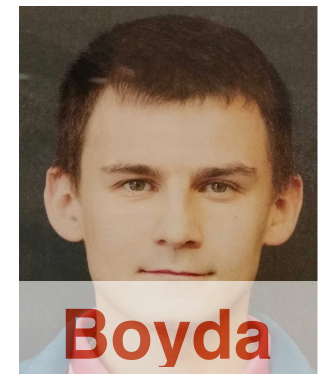


critical limit

Machine learning for LQCD

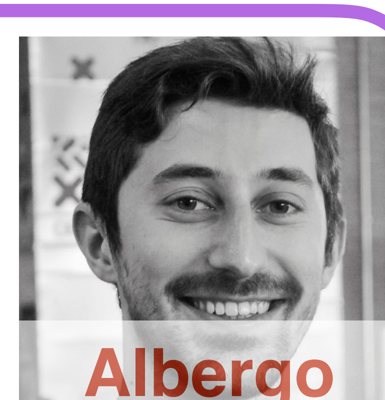
Generative models for QCD gauge field generation

Massachusetts Institute of Technology

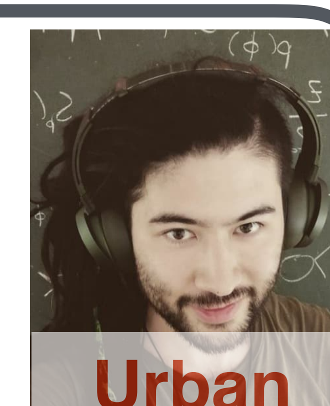


Google DeepMind

NYU



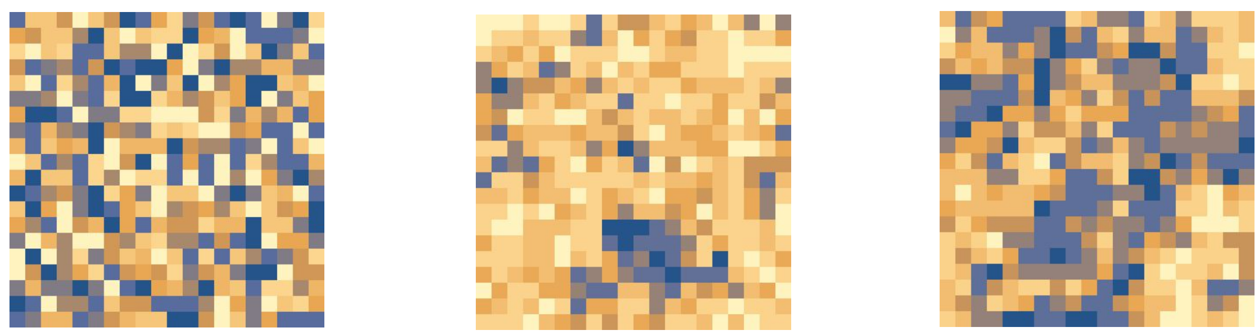
UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386



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

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

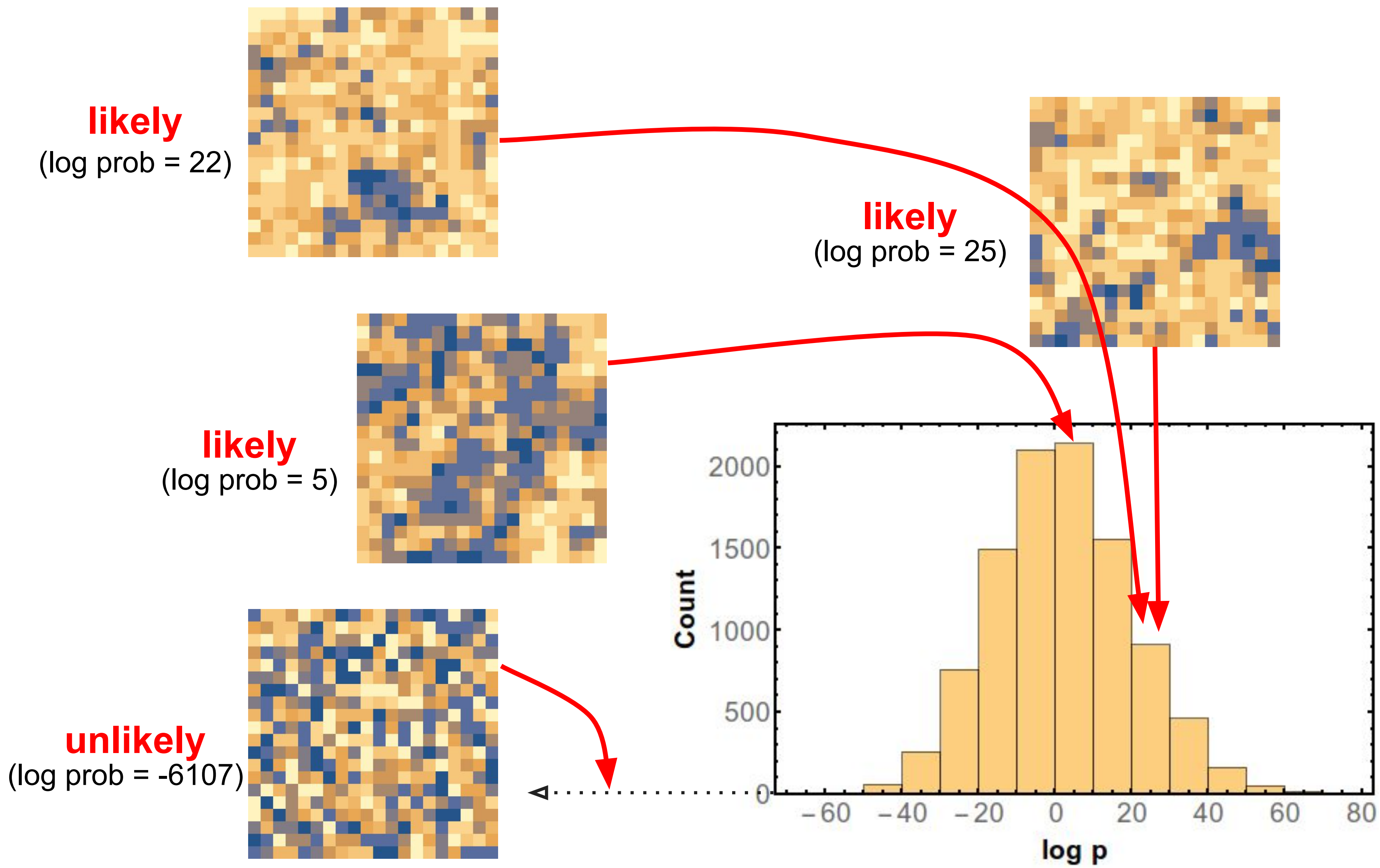
Generate field configurations $\phi(x)$ with probability

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

Sampling gauge field configs

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

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



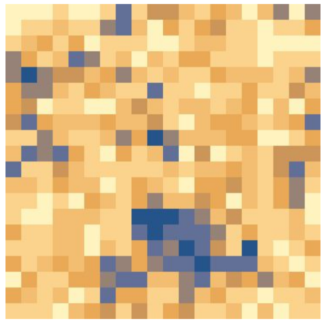
Sampling gauge field configs

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

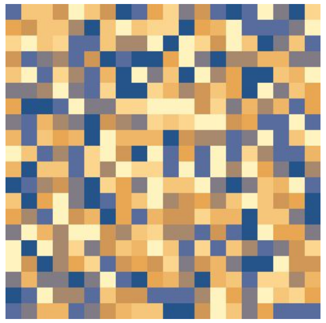
$$P[\phi(x)] \sim e^{-S[\phi(x)]}$$

Parallels with image generation problem

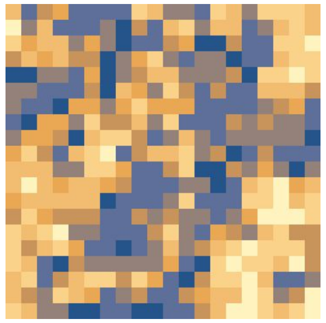
likely
(log prob = 22)



unlikely
(log prob = -6107)



likely
(log prob = 5)

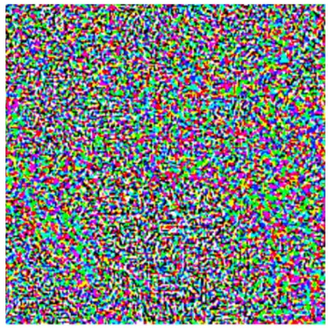


likely

[Karras, Lane, Aila / NVIDIA 1812.04948]

likely

unlikely



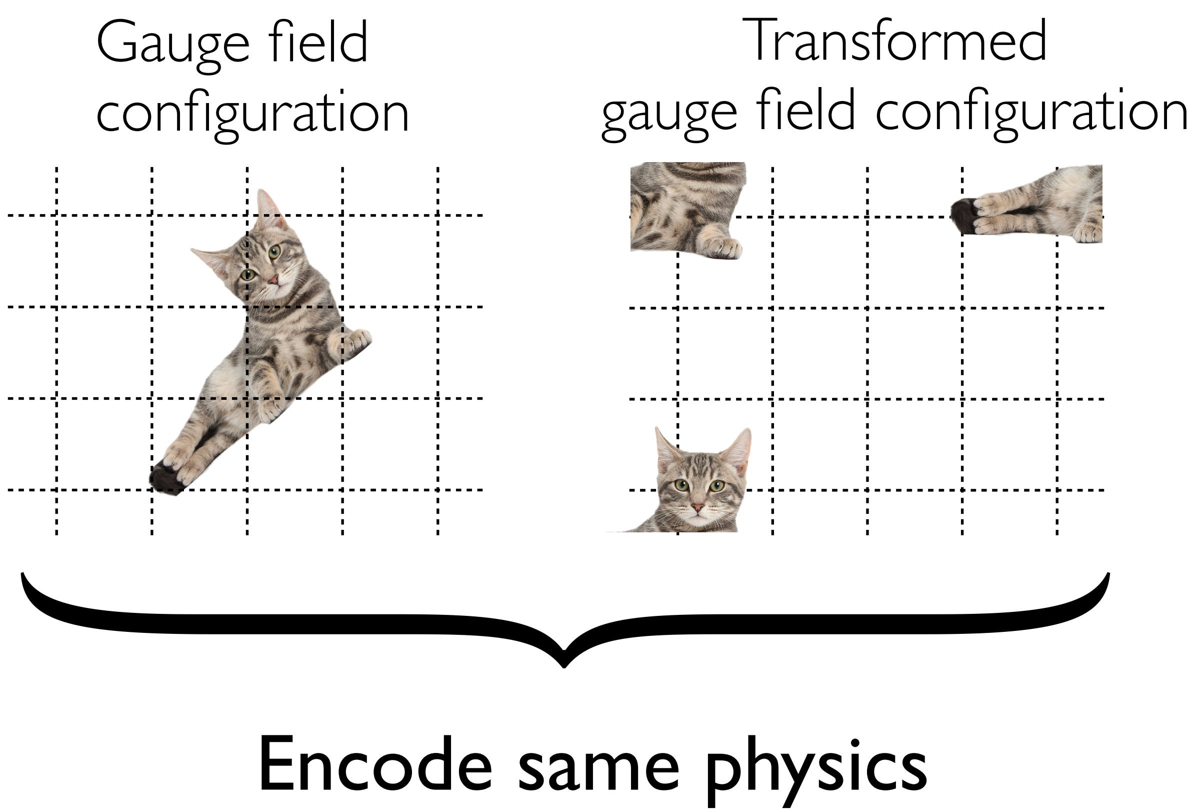
Sampling gauge field configs

Ensemble of lattice QCD gauge fields

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

Physics is invariant under specific field transformations

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



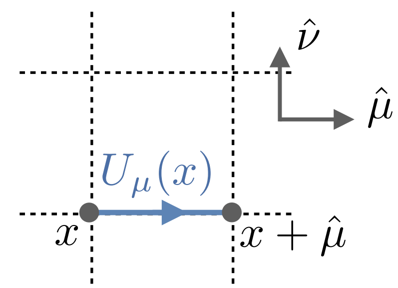
Sampling gauge field configs

Ensemble of lattice QCD gauge fields

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

Physics is invariant under specific field transformations

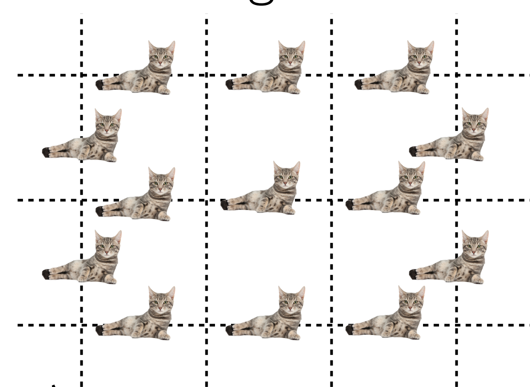
- **Gauge transformation**



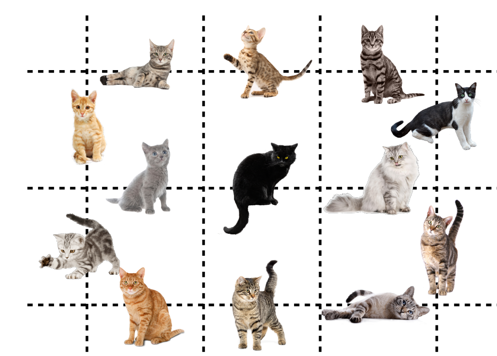
$$U_\mu(x) \to \Omega(x) U_\mu(x) \Omega^\dagger(x+\hat{\mu})$$

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

Gauge field configuration



Transformed gauge field configuration



Encode same physics

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 x 32 pixels x 3 cols ≃3000 numbers
- 60000 samples
- Each image has meaning
- Local structures are important
- Translation-invariance within frame

Machine learning QCD

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

Need custom ML for physics from the ground up

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

Sampling gauge field configs

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

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

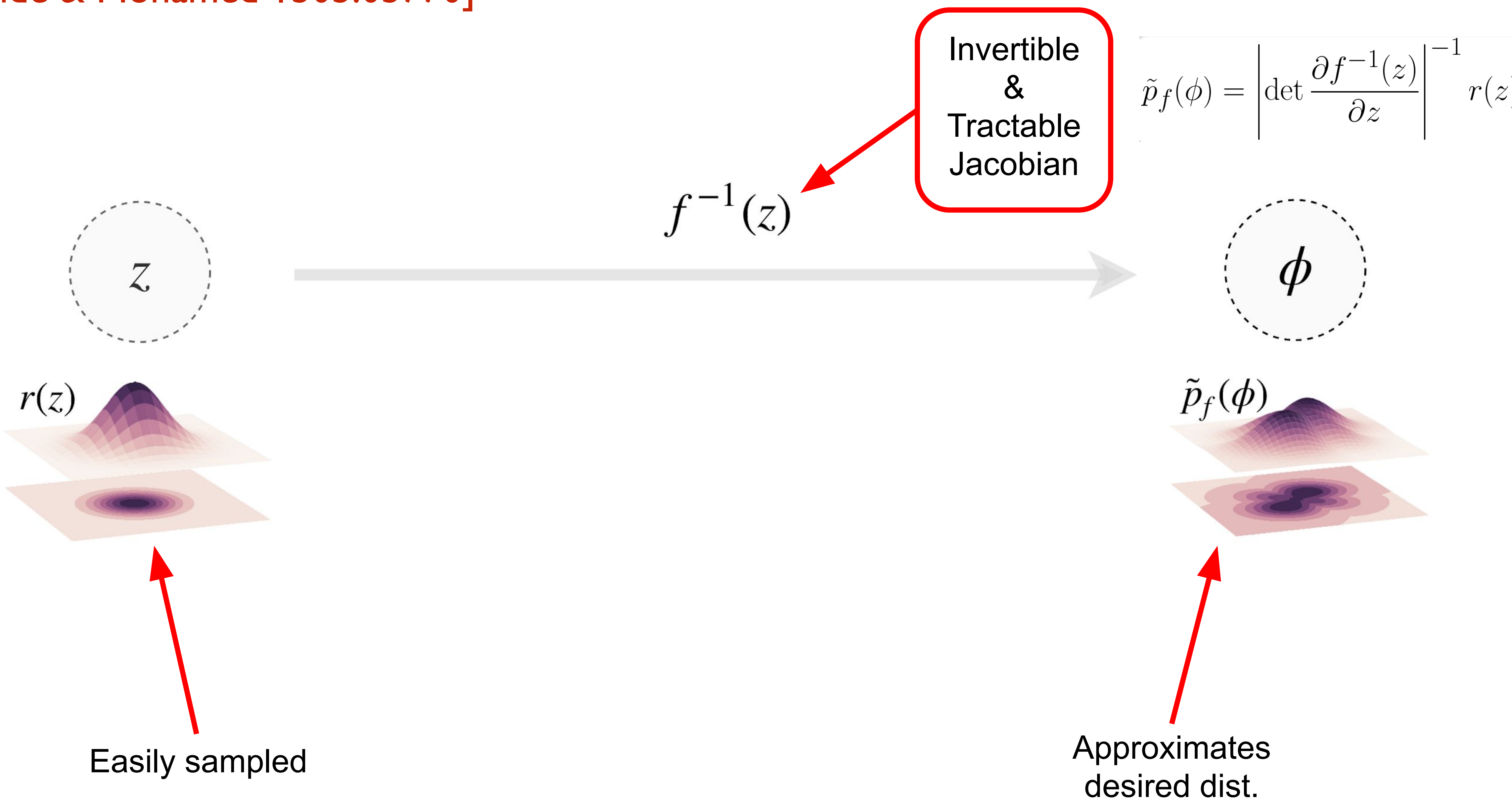
- **Physics distributions have precise symmetries**
  - Lattice symmetries (translation, rotation, reflection)
  - Internal symmetries (gauge symmetries mixing field components)
- **Data hierarchies are challenging**
  - 10^9 to 10^{12} variables per configuration
  - O(1000), samples available (fewer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

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

Generative flow models

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

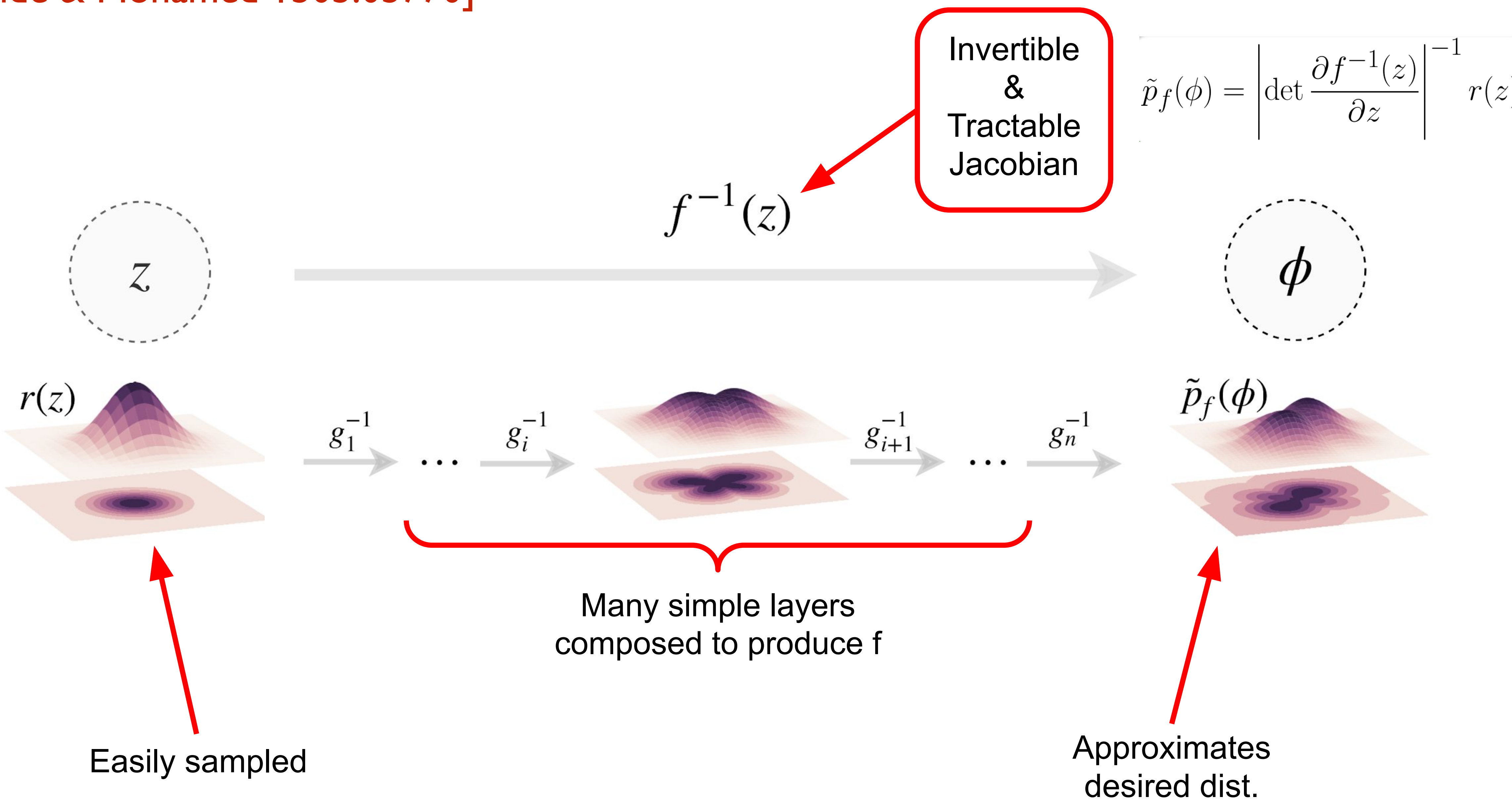
[Rezende & Mohamed 1505.05770]



Generative flow models

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

[Rezende & Mohamed 1505.05770]

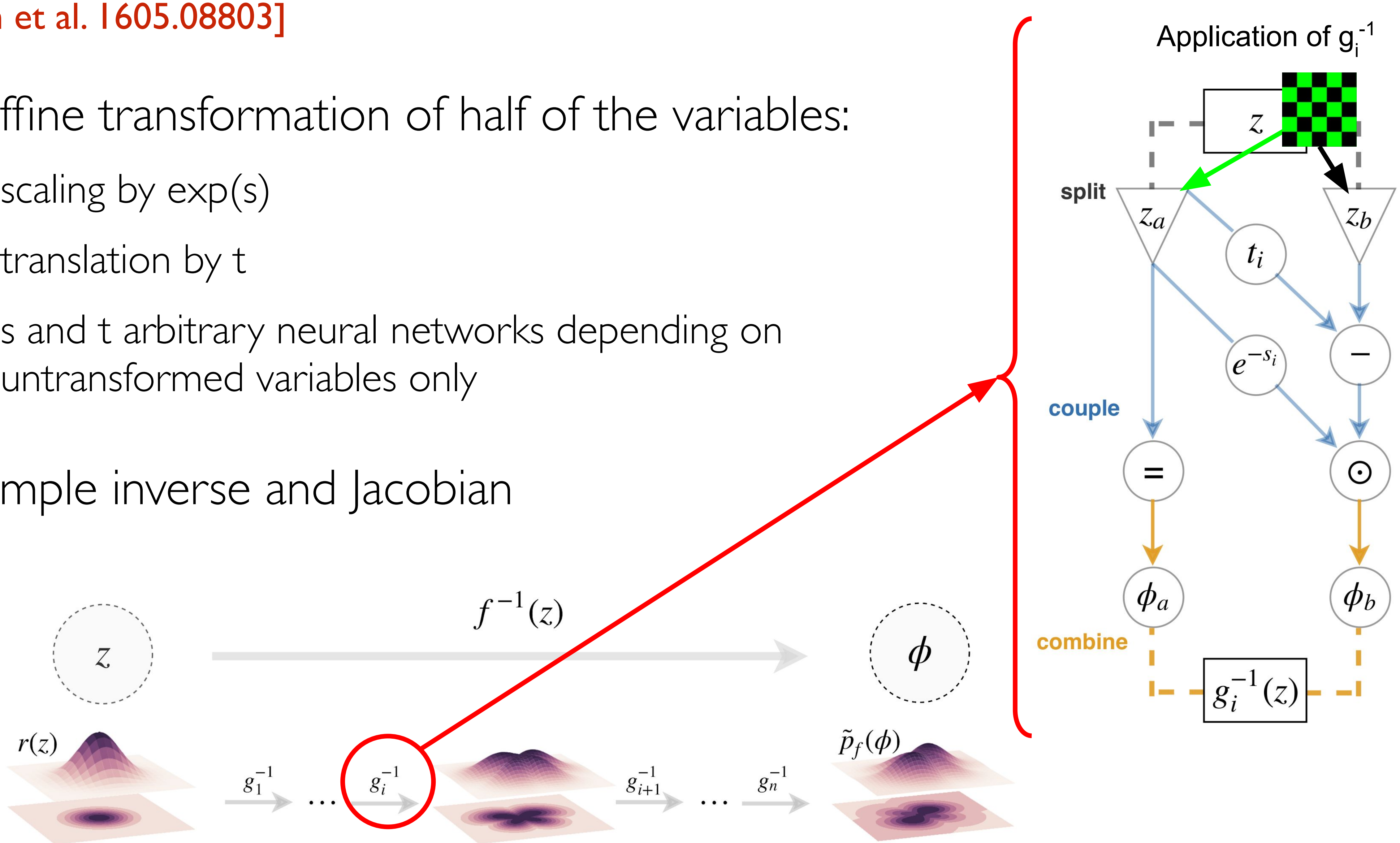


$$\tilde{p}_f(\phi) = \left|\det \frac{\partial f^{-1}(z)}{\partial z}\right|^{-1} r(z)$$

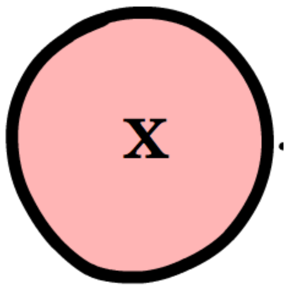
Generative flow models

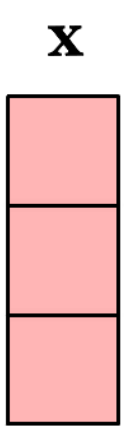
Choose real non-volume preserving flows:
[Dinh et al. 1605.08803]

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



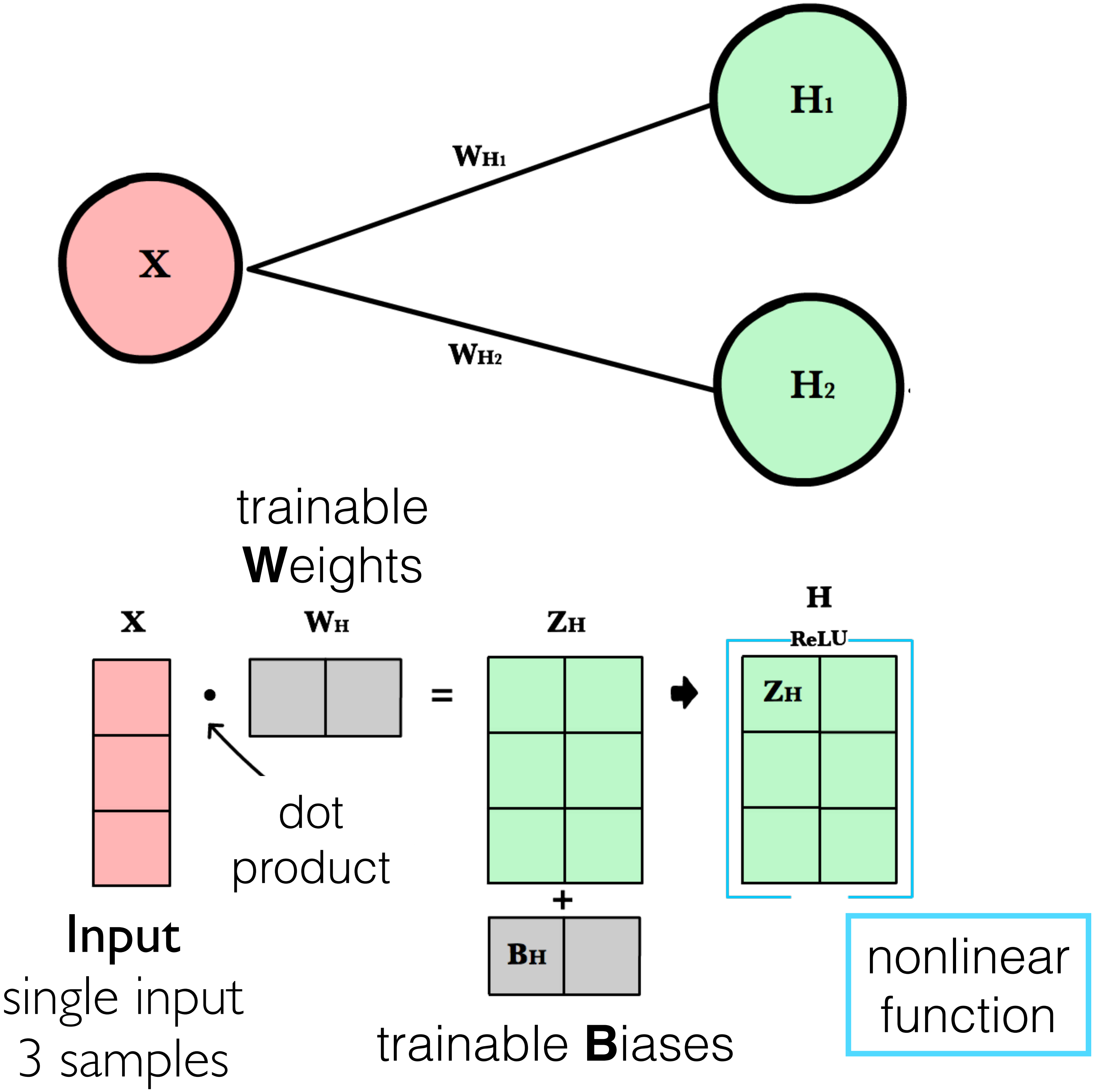
Simple neural network



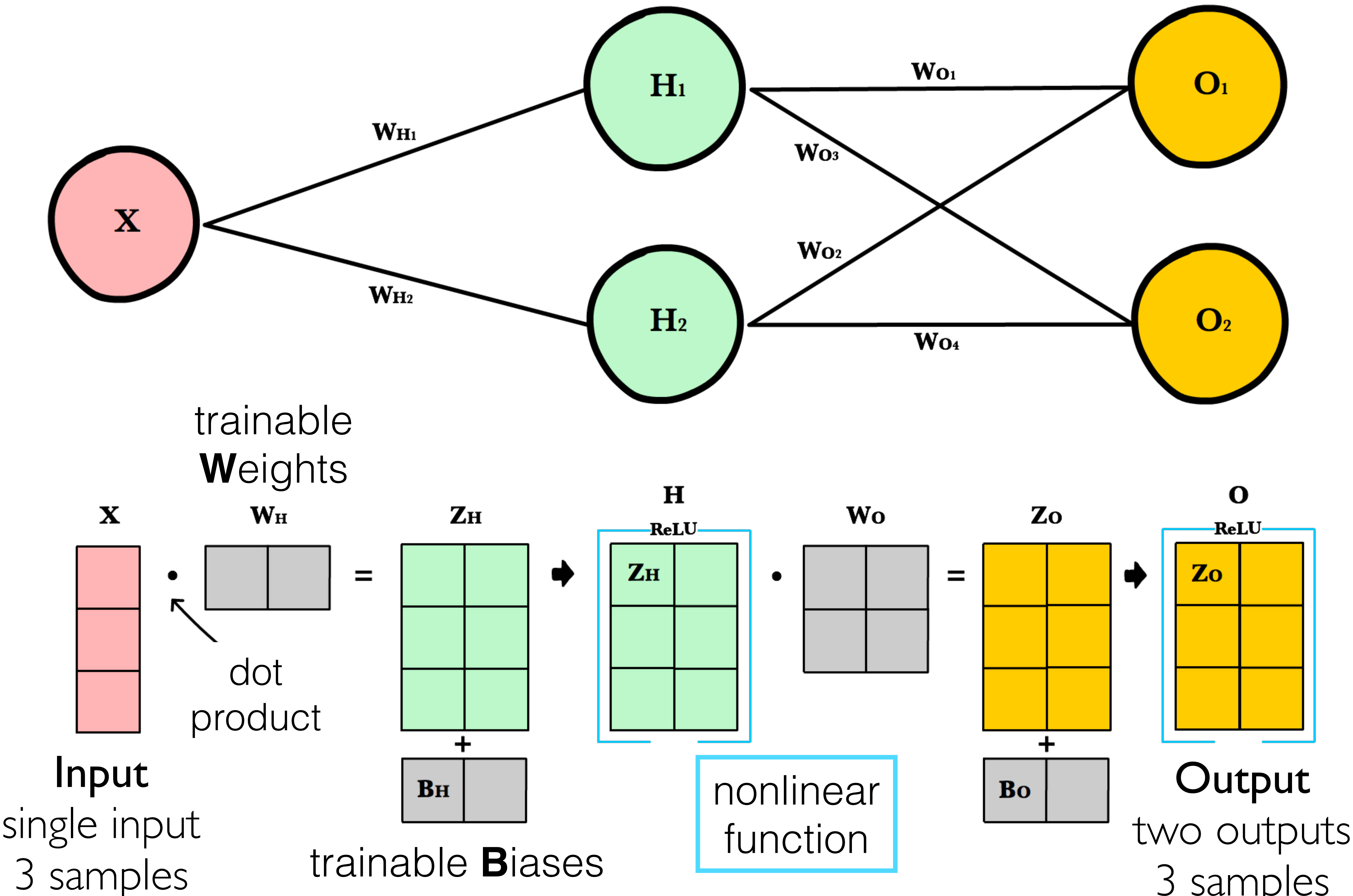


Input
single input
3 samples

Simple neural network



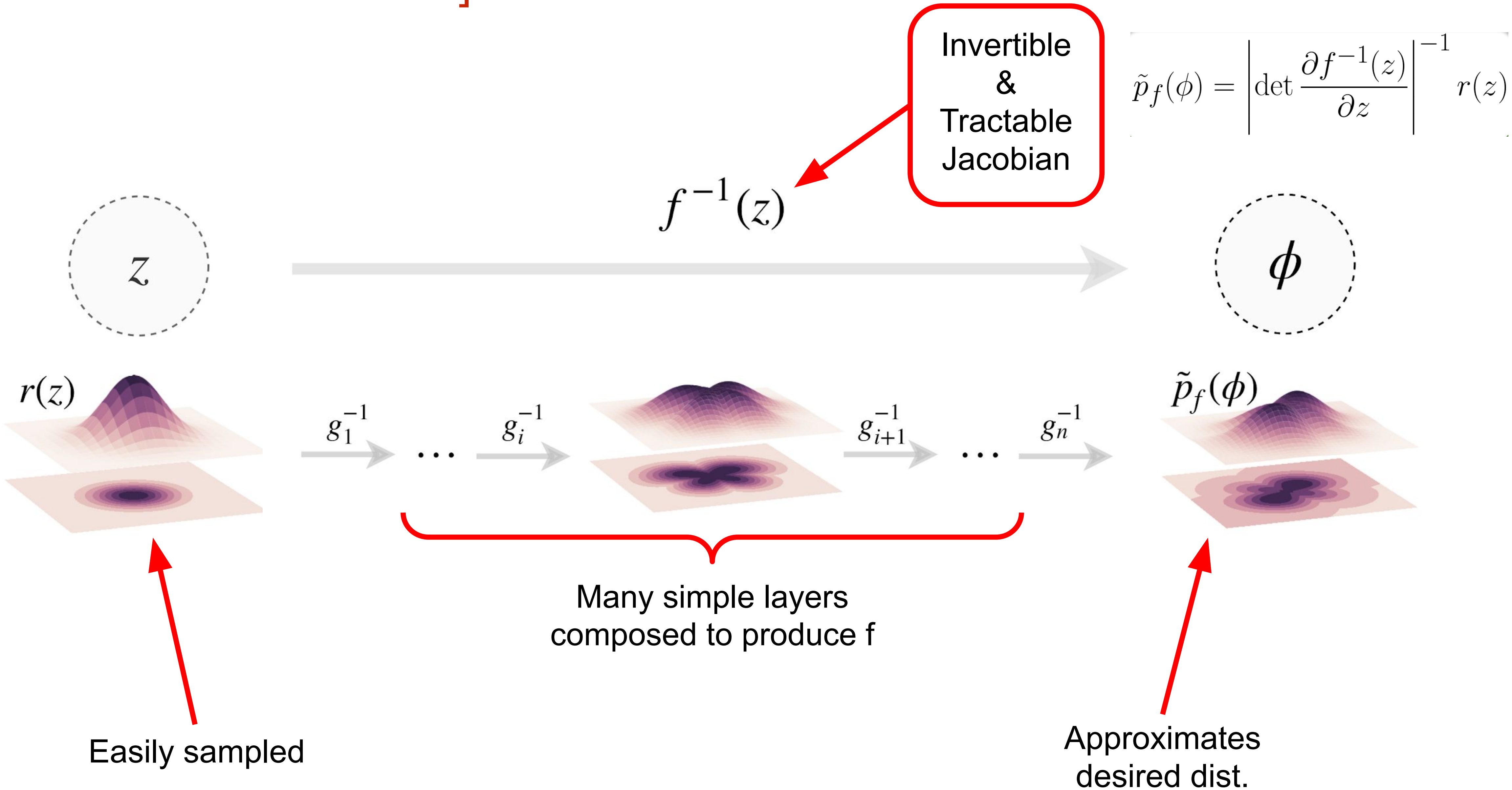
Simple neural network



Generative flow models

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

[Rezende & Mohamed 1505.05770]



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]

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

shift removes unknown normalisation Z

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

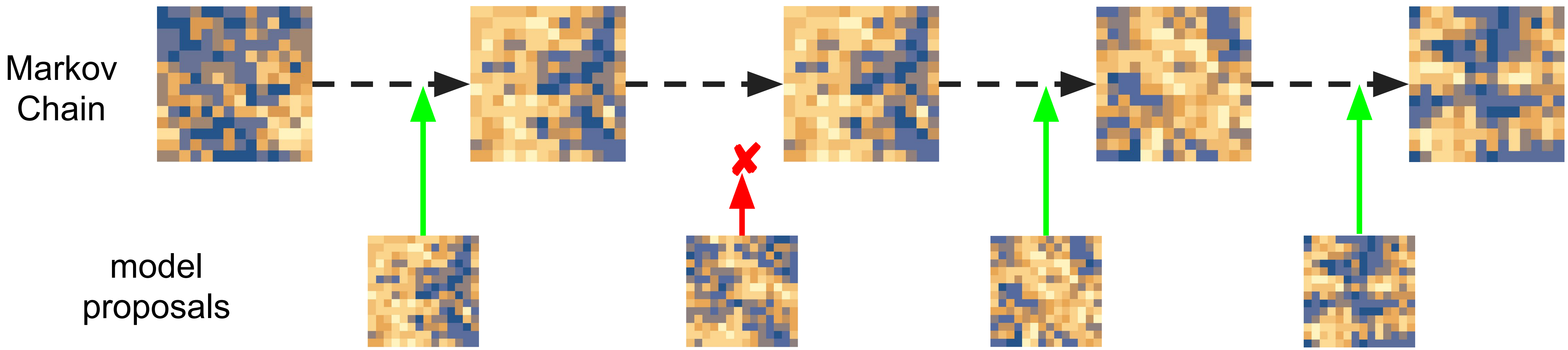
Exactness via Markov chain

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

Acceptance probability

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

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

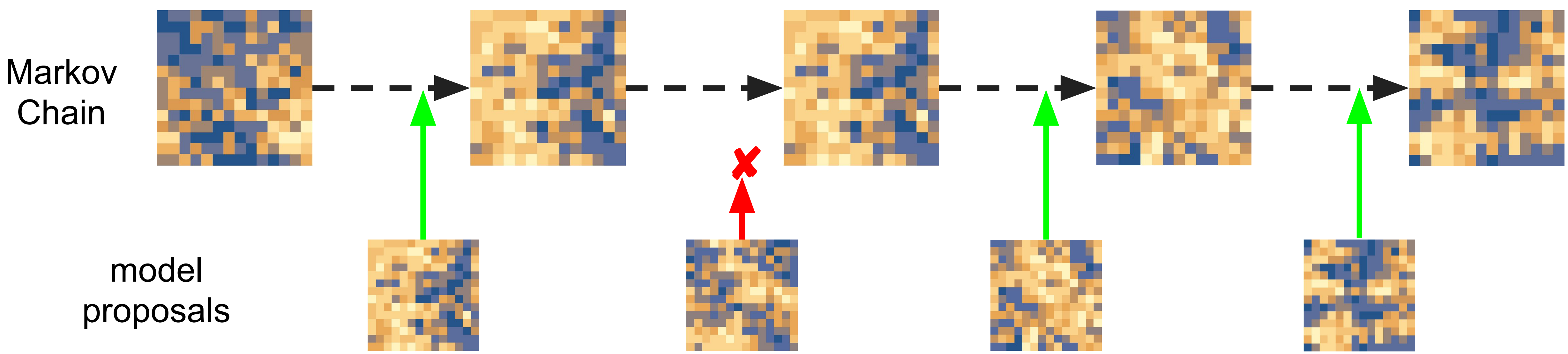
Acceptance probability

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

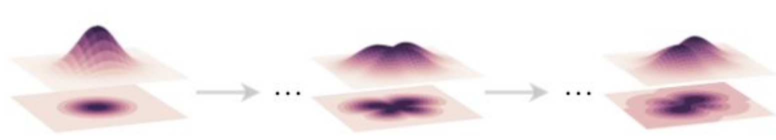
True dist

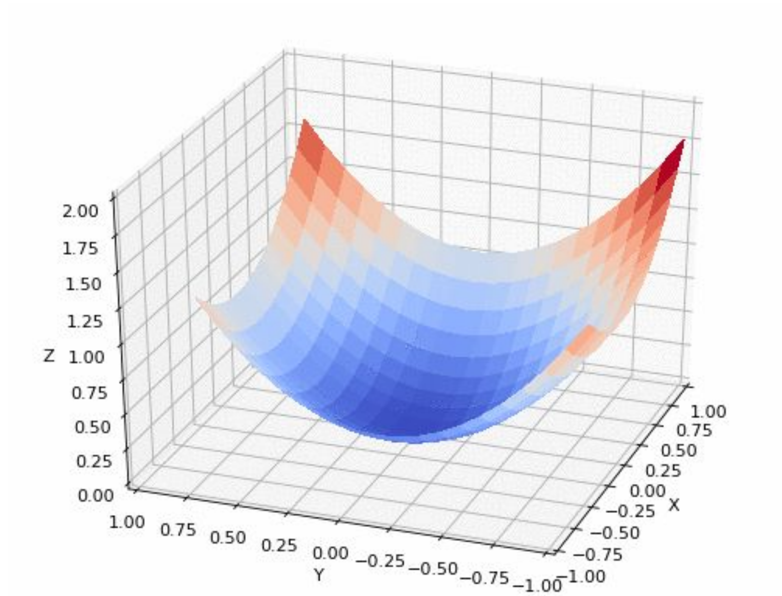
Model dist

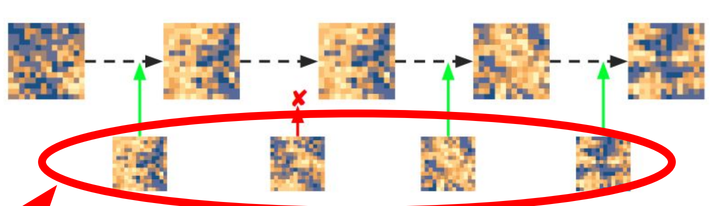
proposal independent of previous sample



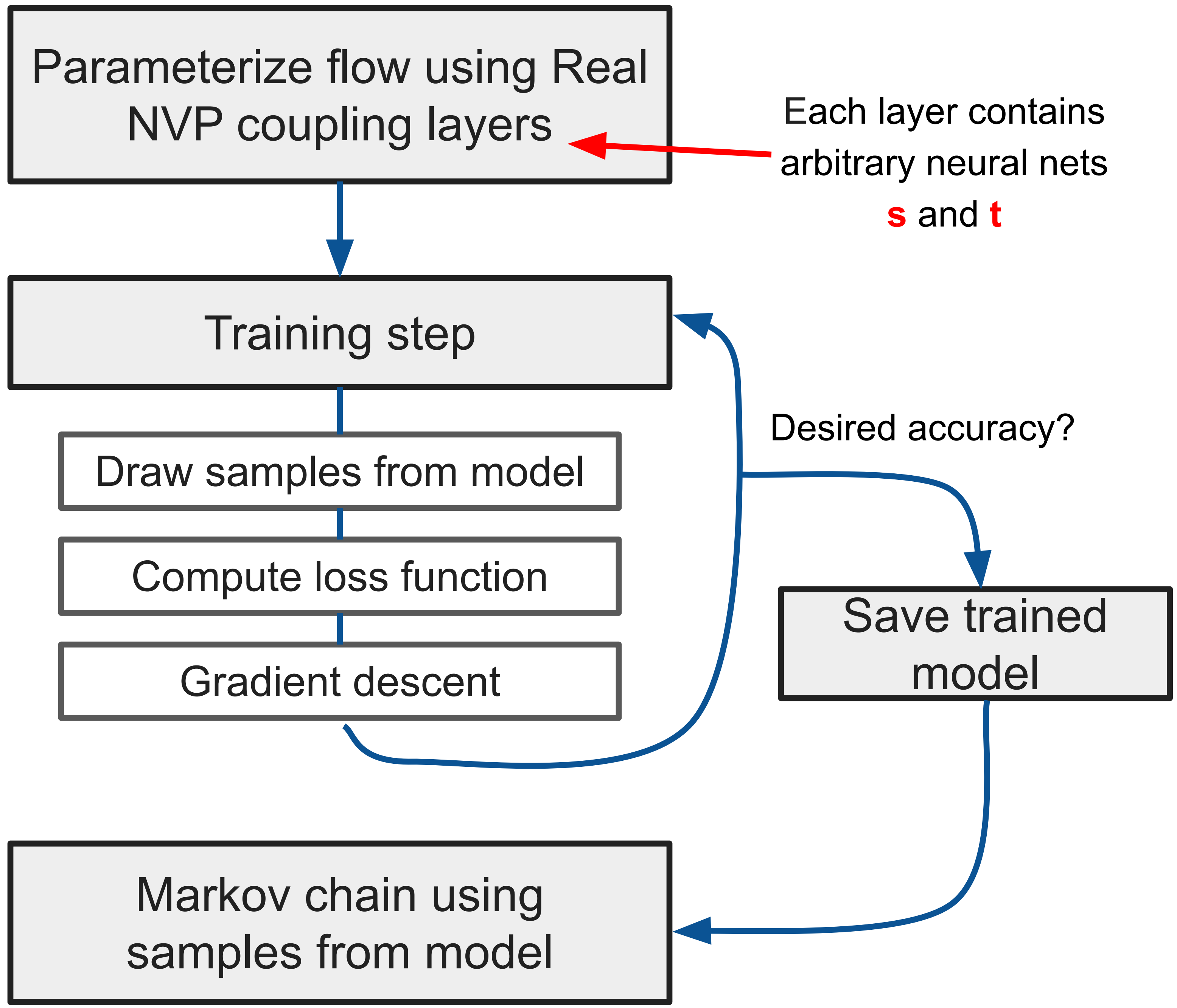
Fields via flow models







generating samples is
"embarrassingly parallel"



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)

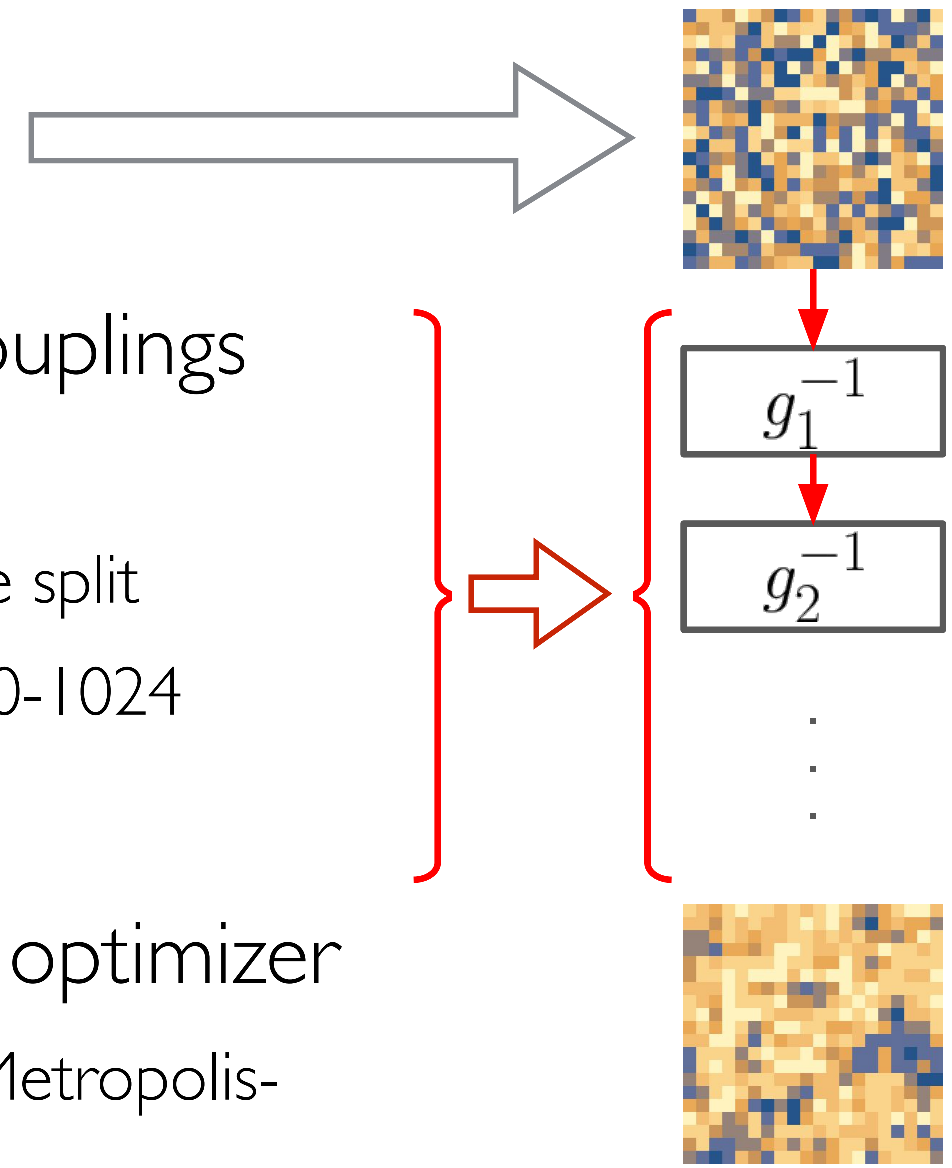
Application: scalar field theory

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



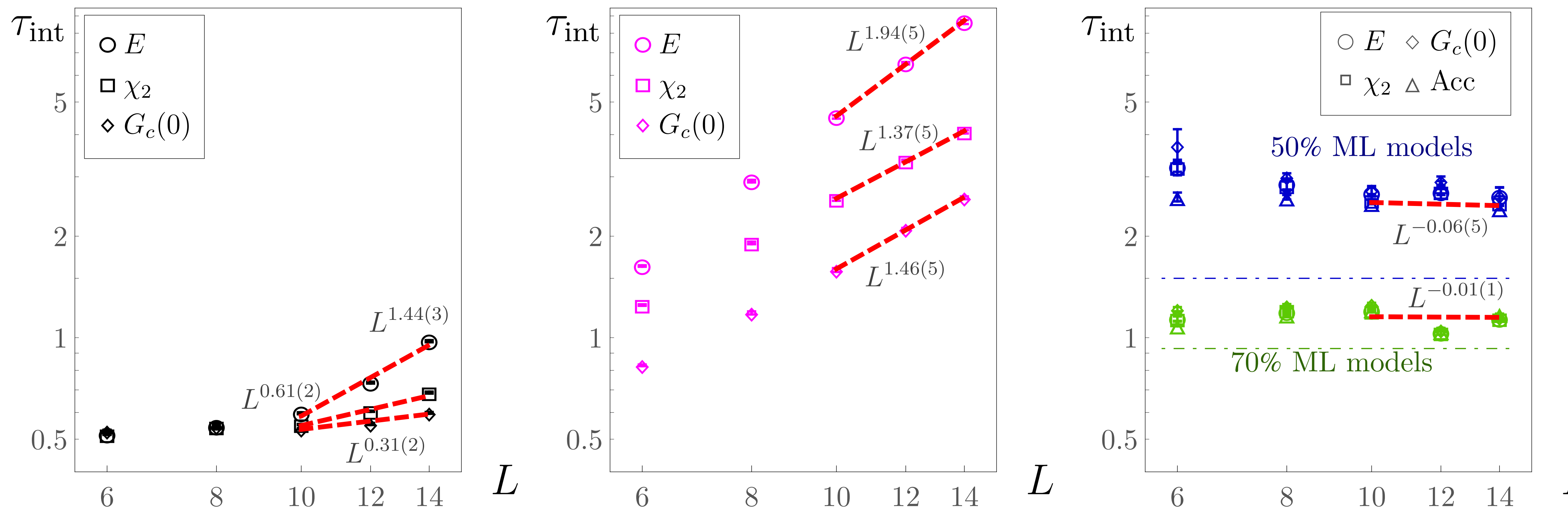
- Train using shifted KL loss with Adam optimizer
  - ∗ Stopping criterion: fixed acceptance rate in Metropolis-Hastings MCMC

Application: scalar field theory

First application: scalar lattice field theory

Success: Critical slowing down is eliminated

Cost: Up-front training of the model



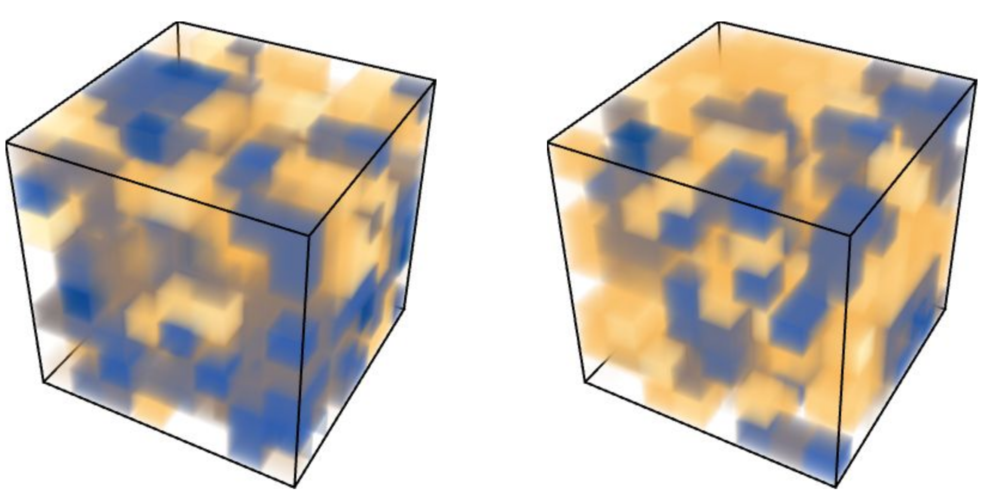
(a) HMC ensembles (b) Local Metropolis ensembles (c) Flow-based MCMC ensembles

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]

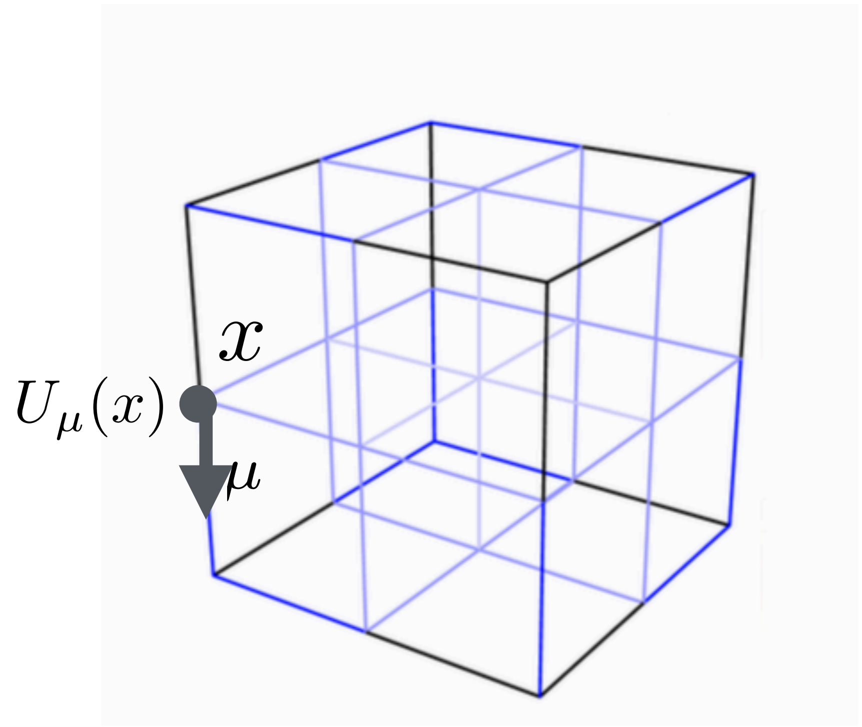


Aurora21 Early Science Project

Incorporating symmetries

Gauge field theories

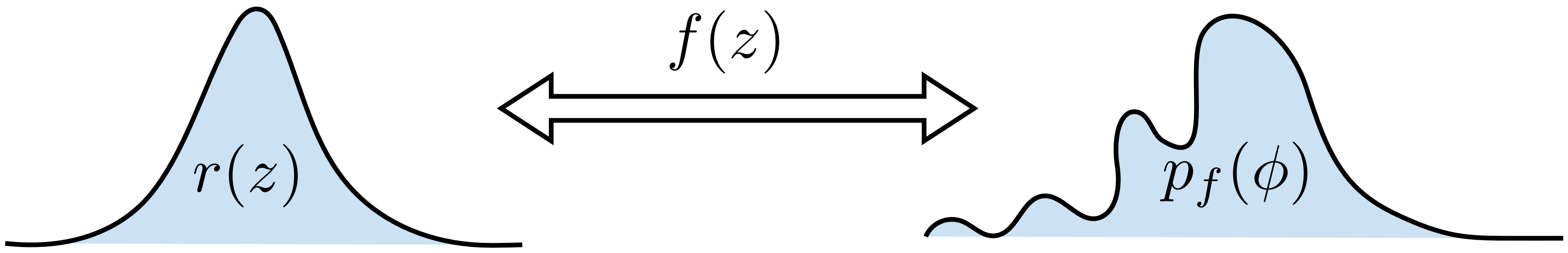
- Field configurations represented by links $U_\mu(x)$ encoded as matrices
- e.g., for Quantum Chromodynamics, SU(3) matrices (3x3 complex matrices M with $det[M] = 1$, $M^{-1} = M^\dagger$)
- Group-valued fields live not on real line but on compact manifolds
- Action is invariant under group transformations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



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

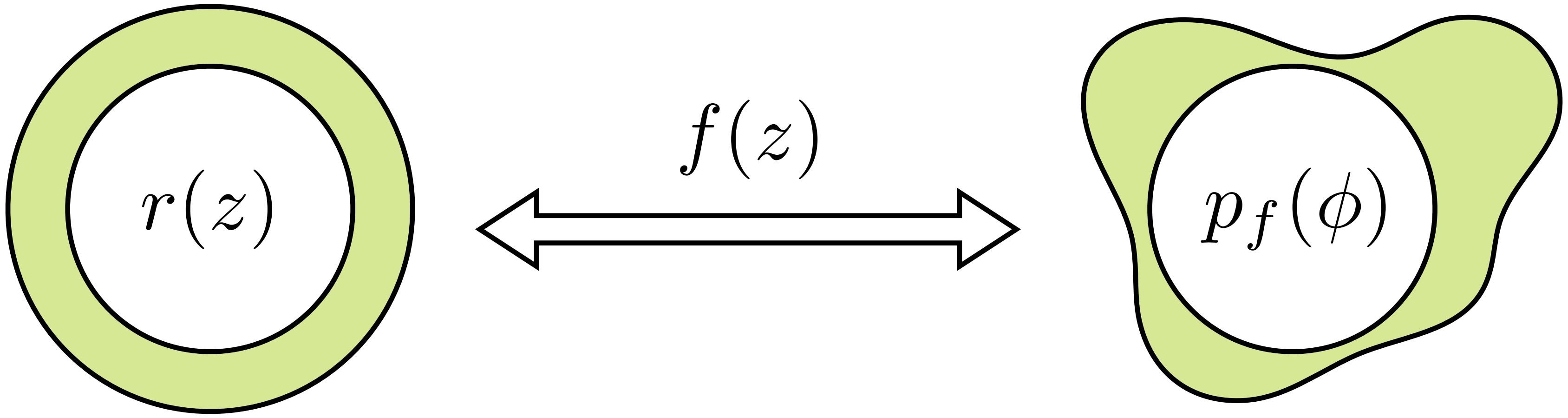
Flows on spheres and tori

Previously: **Real non-volume preserving flows**



Need: **Flows on compact, connected manifolds**

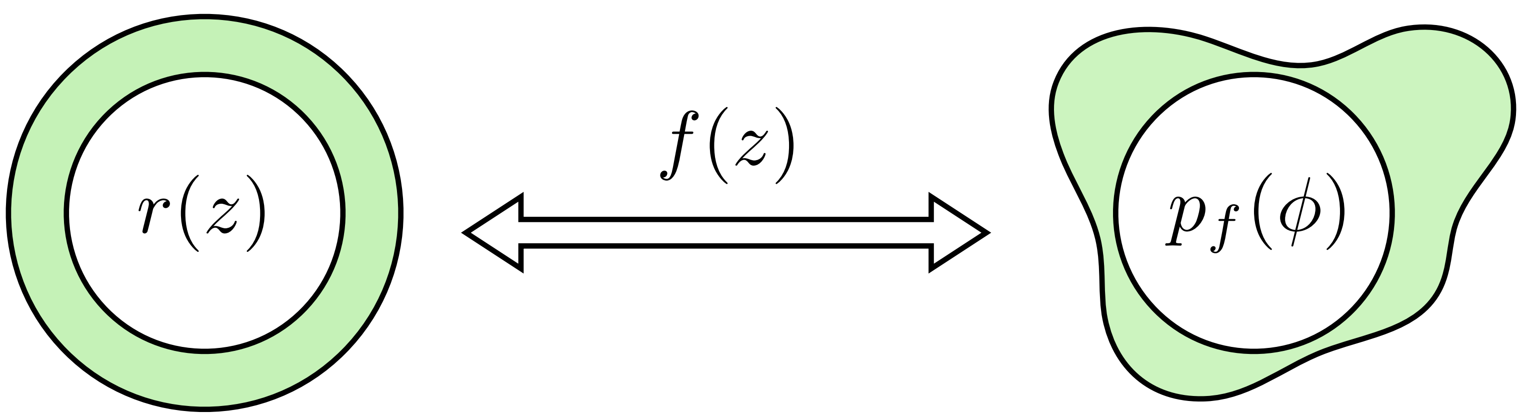
e.g., circles, tori, spheres



Flows on spheres and tori

Test case: Flows on the circle

e.g., U(1) field theory, robot arm positions



Diffeomorphism requires:

$$f(0) = 0,$$
$$f(2\pi) = 2\pi,$$
$$\nabla f(\theta) > 0,$$
$$\nabla f(\theta)|_{\theta=0} = \nabla f(\theta)|_{\theta=2\pi}$$

Ensures transformation is monotonic → invertible

Expressive transformations through:

- Composition $f = f_K \circ \cdots \circ f$
- Convex combination

$$f(\theta) = \sum_i \rho_i f_i(\theta) \qquad \rho_i \geq 0, \quad \sum_i \rho_i = 1$$

Flows on spheres and tori

[arXiv:2002.02428]

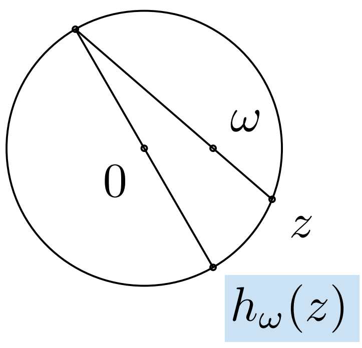
Normalizing Flows on Tori and Spheres

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

- Mobius transformation

$$f_\omega(\theta) = R_\omega \circ h_\omega(z)$$

Rotation to fix $f(\theta = 0)$



- Circular splines
  - Rational quadratic function of θ on each of K segments
  - Several conditions on coefficients to guarantee diffeomorphism

$$f(\theta) = \frac{\alpha_{k2}\theta^2 + \alpha_{k1}\theta + \alpha_{k0}}{\beta_{k2}\theta^2 + \beta_{k1}\theta + \beta_{k0}}$$

- Non-compact projection
  - Project to the real line and back: careful with numerical instabilities at endpoints

$$f(\theta) = 2\tan^{-1}\left(\alpha \tan\left(\frac{\theta}{2} - \frac{\pi}{2}\right) + \beta\right) + \pi$$

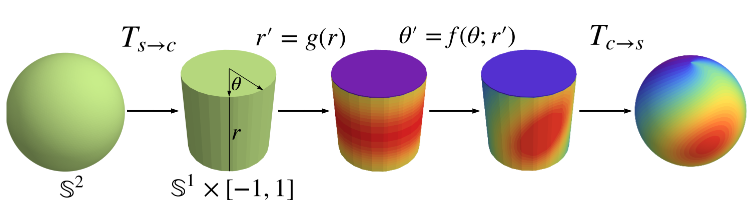
Flows on spheres and tori

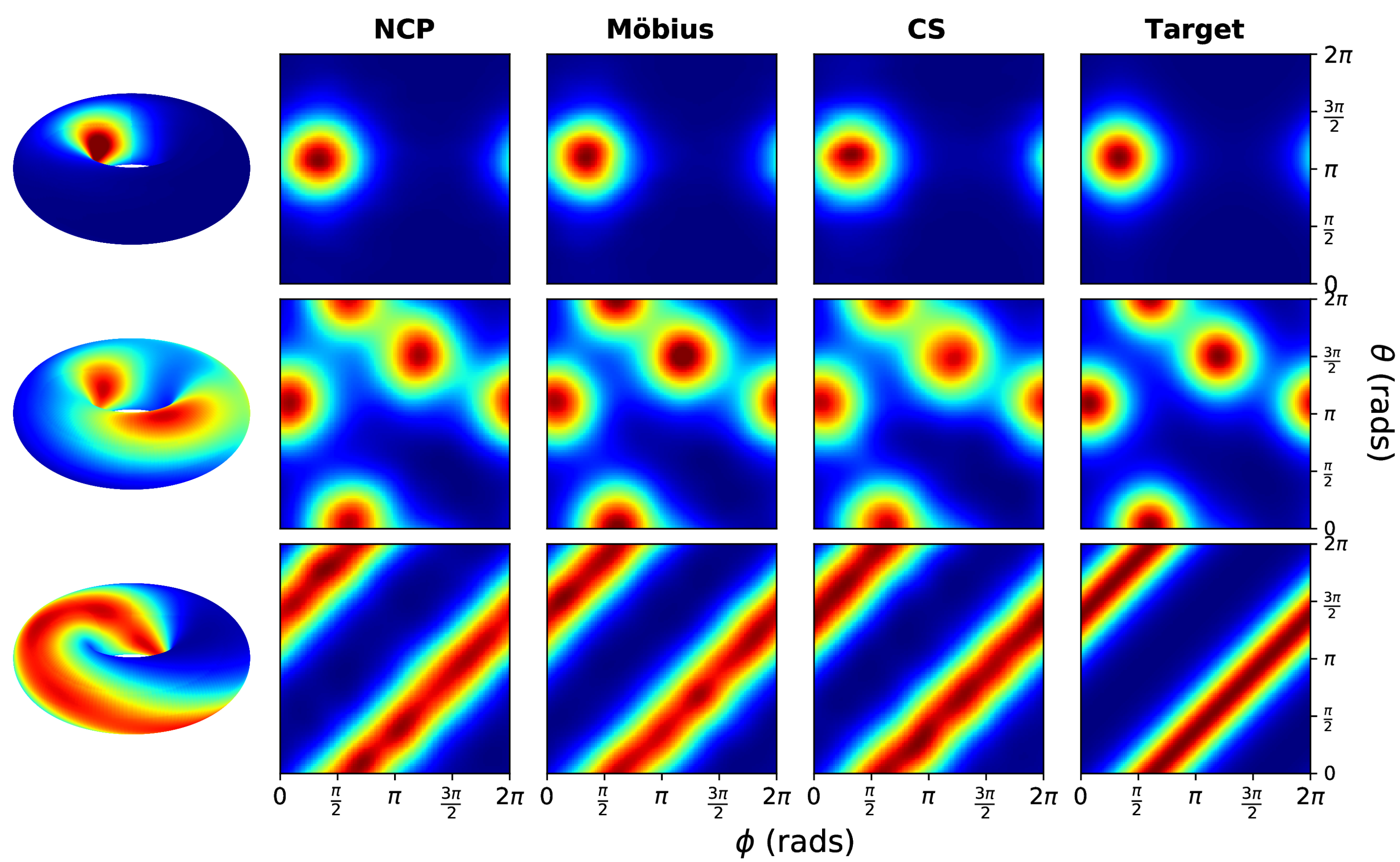
[arXiv:2002.02428]

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]

- Extend straightforwardly to cartesian products of circles and intervals (e.g., tori)
- Extend recursively to D-dimensional spheres





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

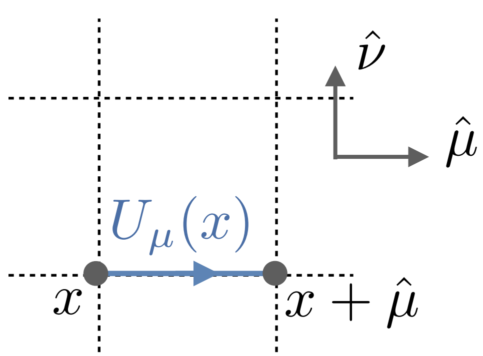
Gauge field theory

First gauge theory application: U(1) field theory

Generative flow architecture that is *gauge-equivariant*

Gauge transformation

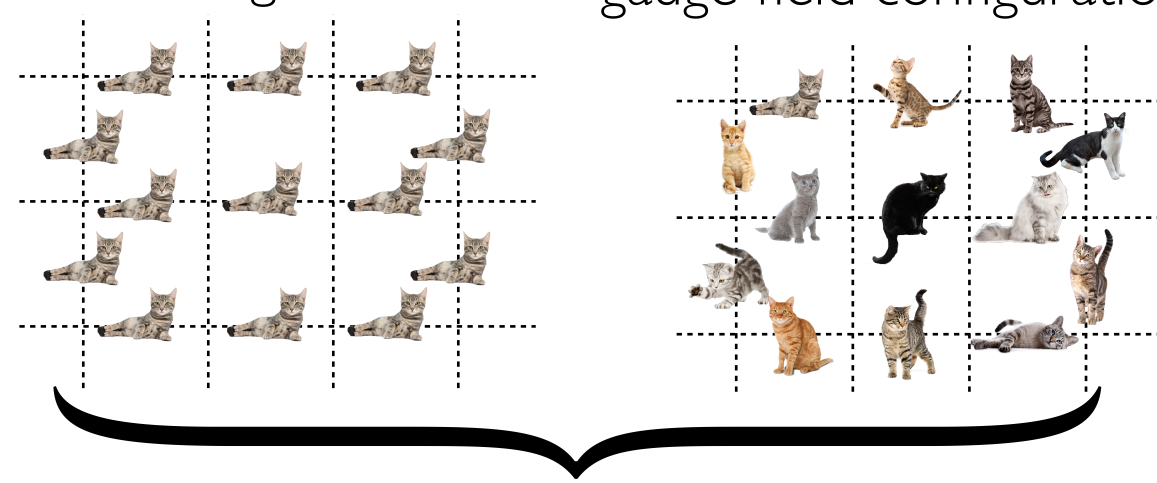
Separate group transformation of each link matrix $U_\mu(x)$



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

Gauge field configuration

Transformed gauge field configuration



Encode same physics

Gauge-equivariant flows

First gauge theory application: U(1) field theory

Generative flow architecture that is *gauge-equivariant*

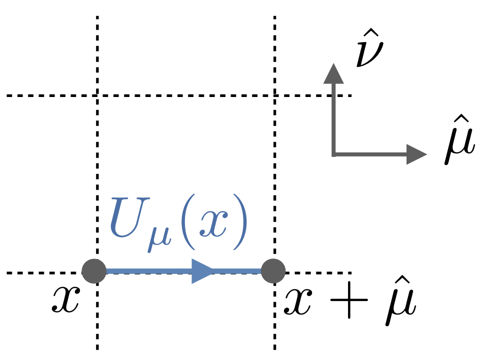
Define invertible, equivariant coupling layer

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 $h : G \to G$

Link updated by coupling layer

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

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

Loop that starts and ends at same point

Coupling 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

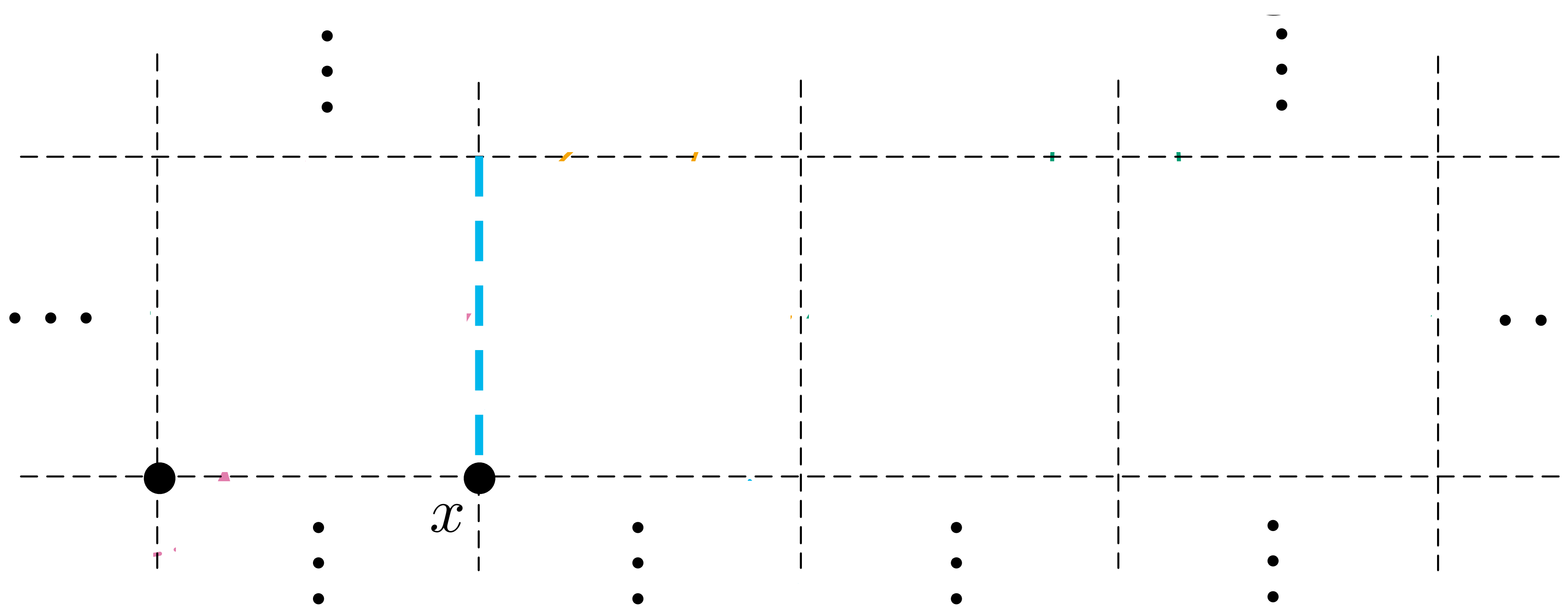
First gauge theory application: U(1) field theory

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

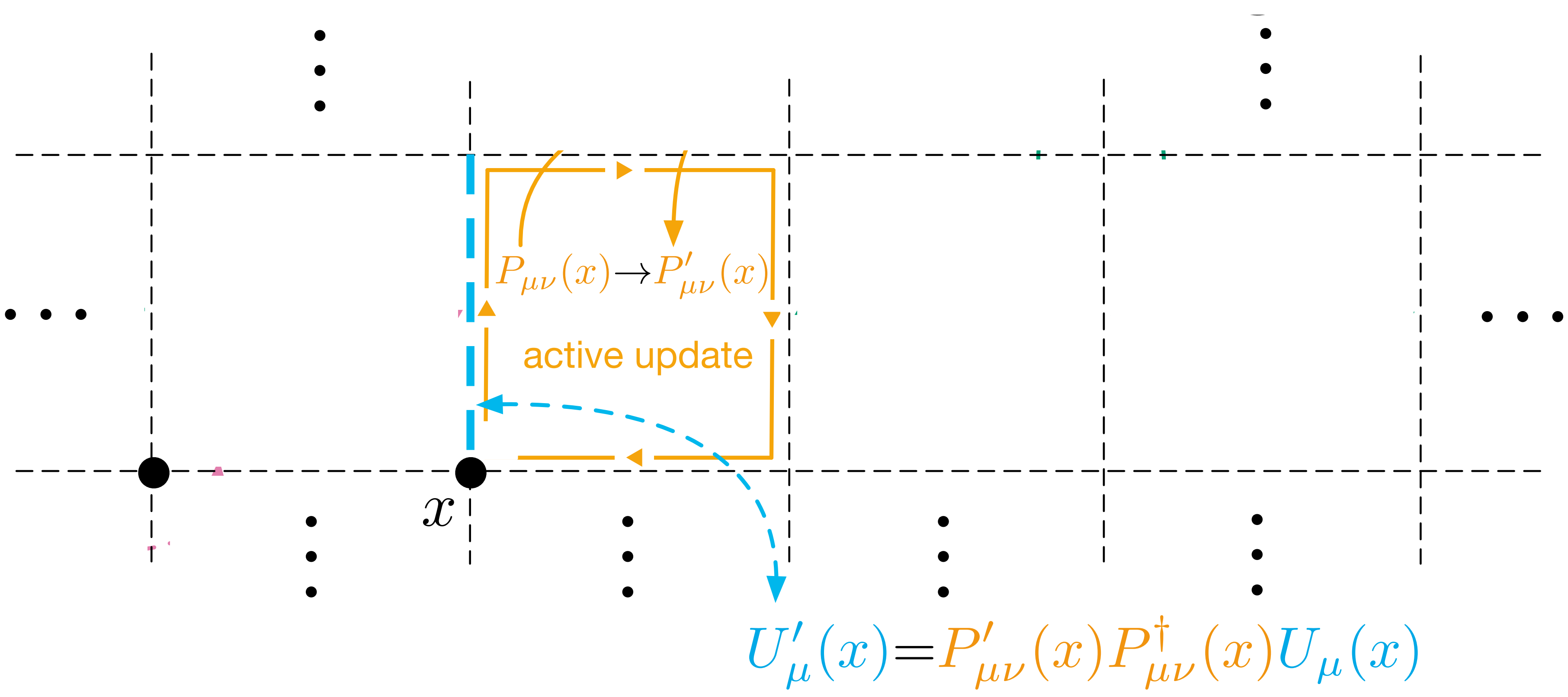
First gauge theory application: U(1) field theory

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



[Kanwar et al., arXiv:2003.06413]

Gauge-equivariant flows

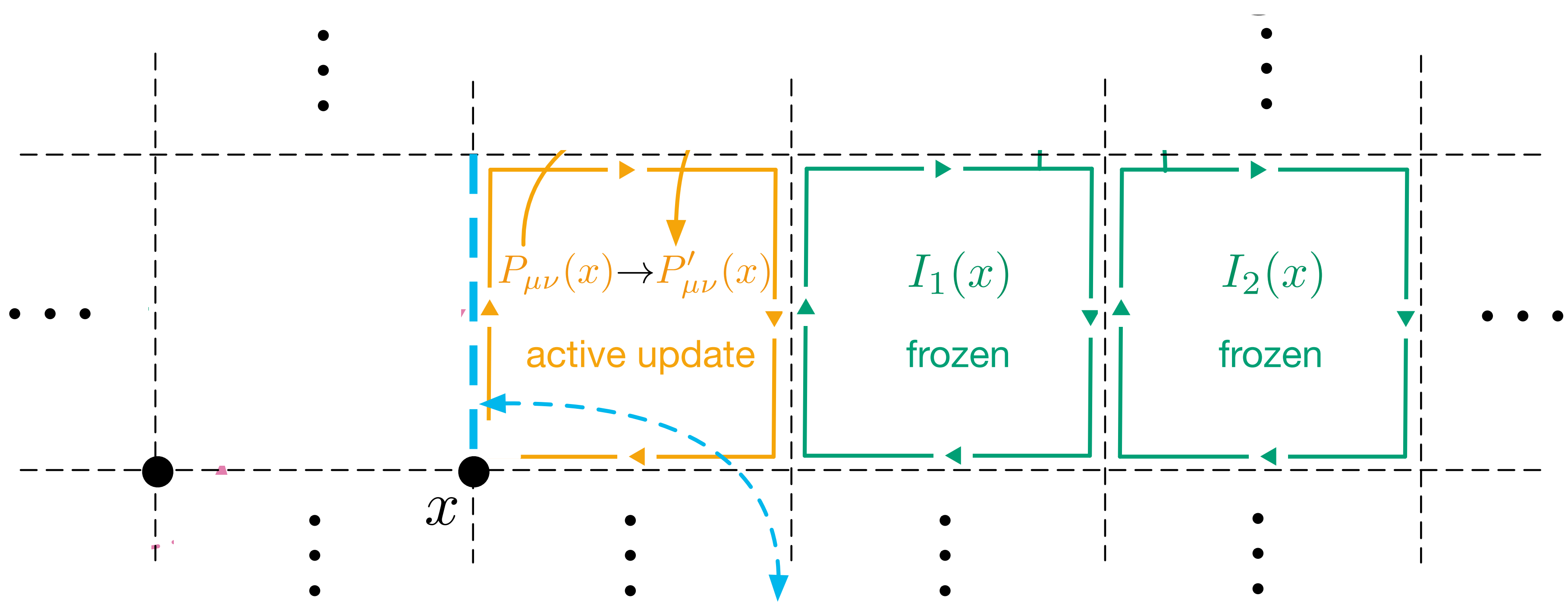
First gauge theory application: U(1) field theory

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



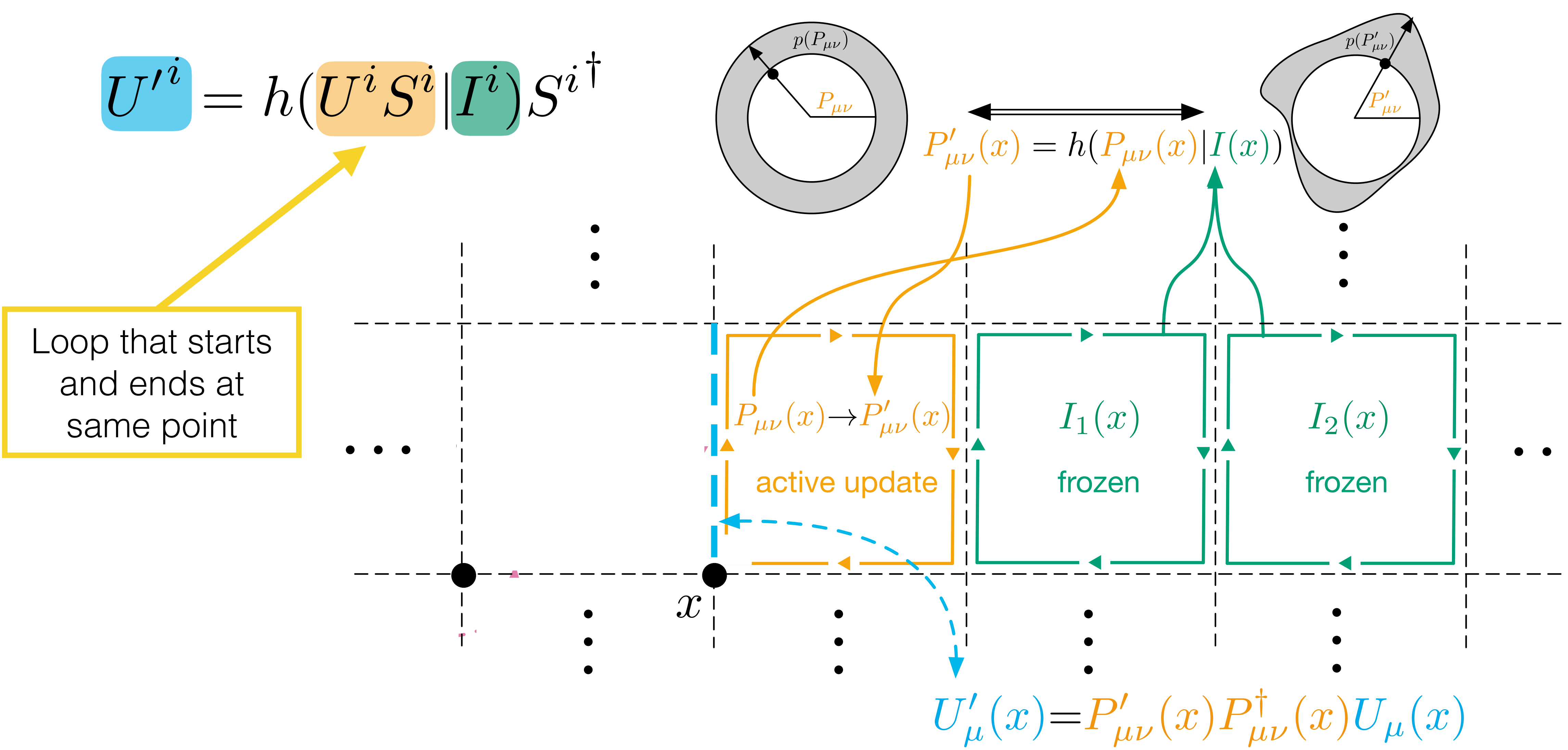
$$U'_{\mu}(x) = P'_{\mu\nu}(x) P^{\dagger}_{\mu\nu}(x) U_{\mu}(x)$$

[Kanwar et al., arXiv:2003.06413]

Gauge-equivariant flows

First gauge theory application: U(1) field theory

Generative flow architecture that is *gauge-equivariant*

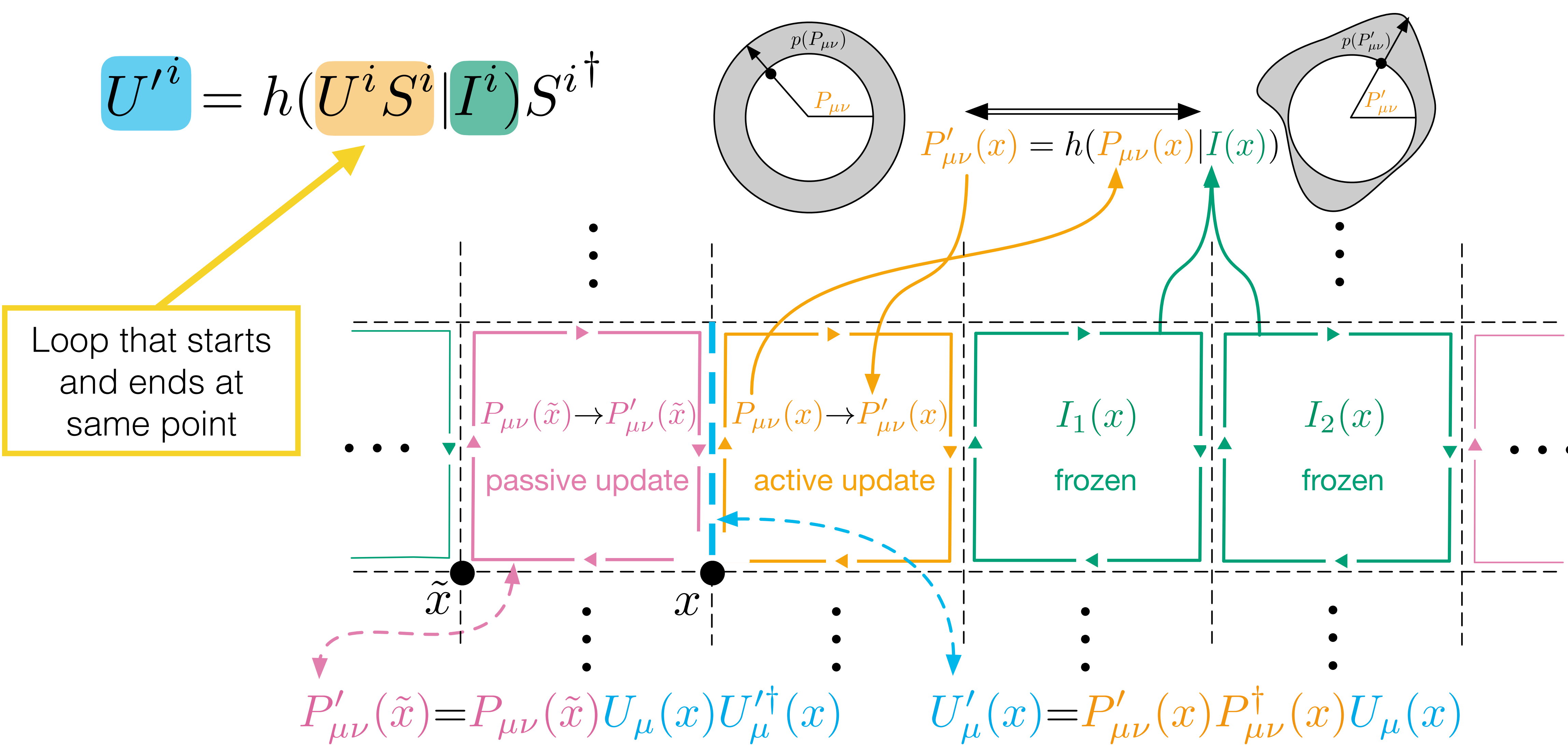


[Kanwar et al., arXiv:2003.06413]

Gauge-equivariant flows

First gauge theory application: U(1) field theory

Generative flow architecture that is *gauge-equivariant*



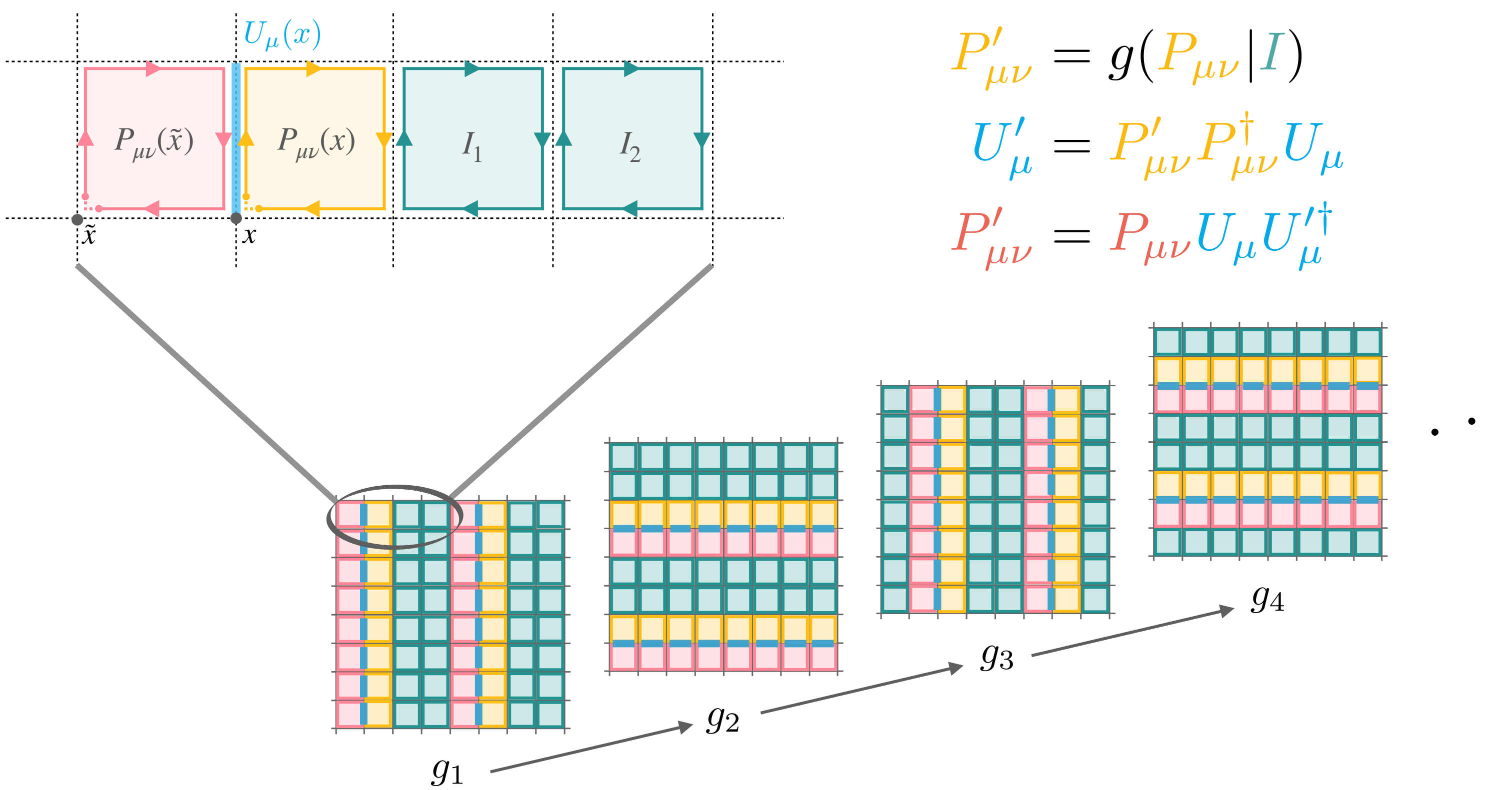
[Kanwar et al., arXiv:2003.06413]

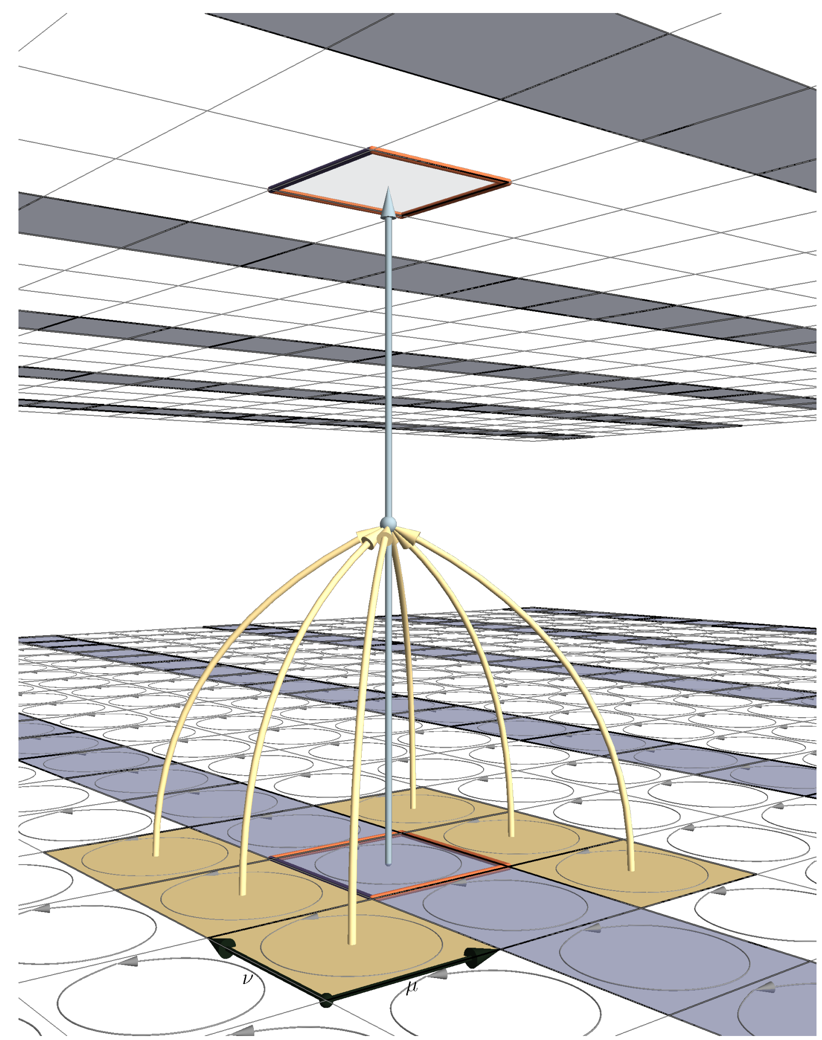
Gauge-equivariant flows

First gauge theory application: U(1) field theory

Generative flow architecture that is *gauge-equivariant*

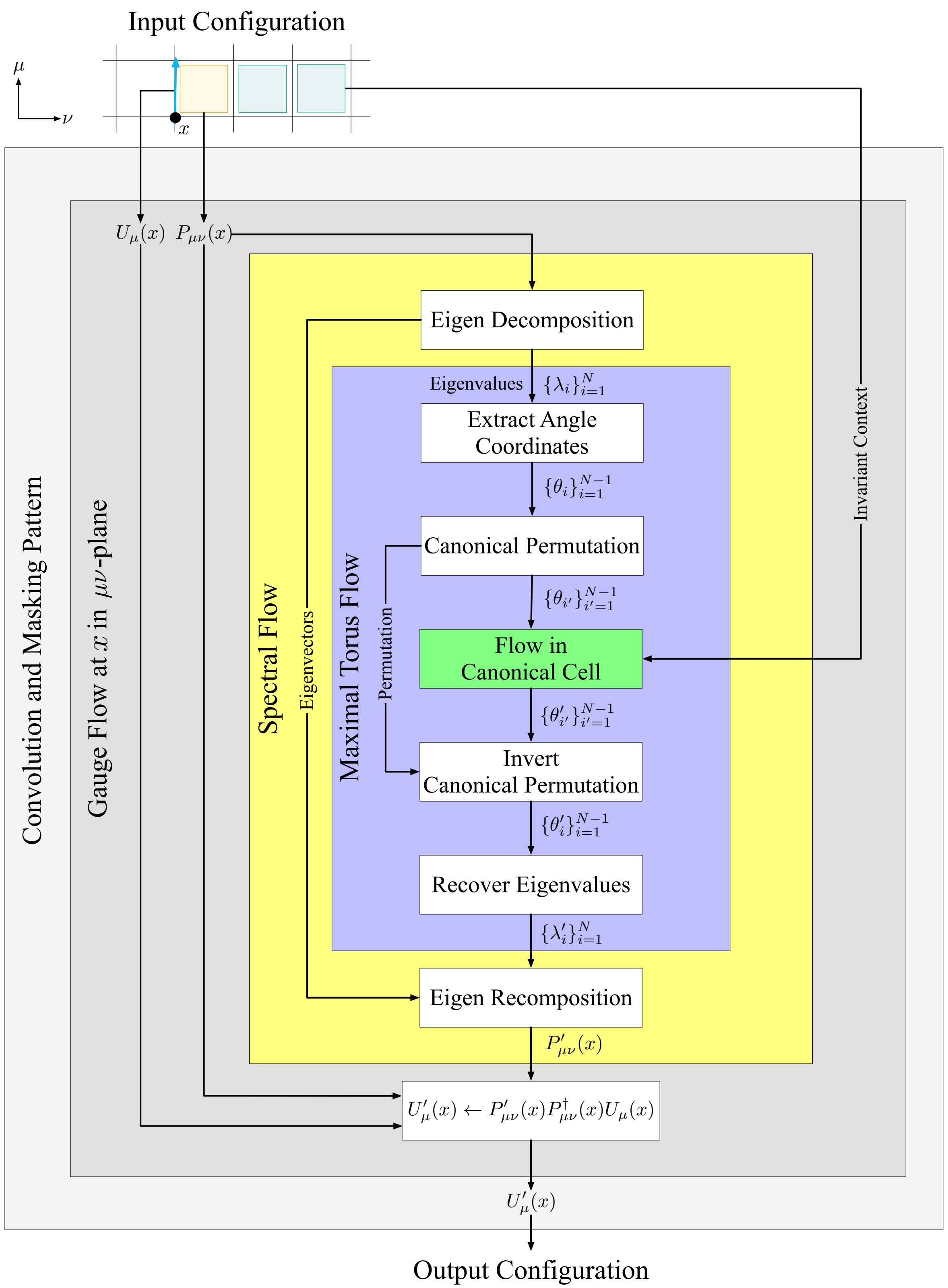
$$P'_{\mu\nu} = g(P_{\mu\nu}|I)$$
$$U'_{\mu} = P'_{\mu\nu}P^{\dagger}_{\mu\nu}U_{\mu}$$
$$P'_{\mu\nu} = P_{\mu\nu}U_{\mu}U'^{\dagger}_{\mu}$$

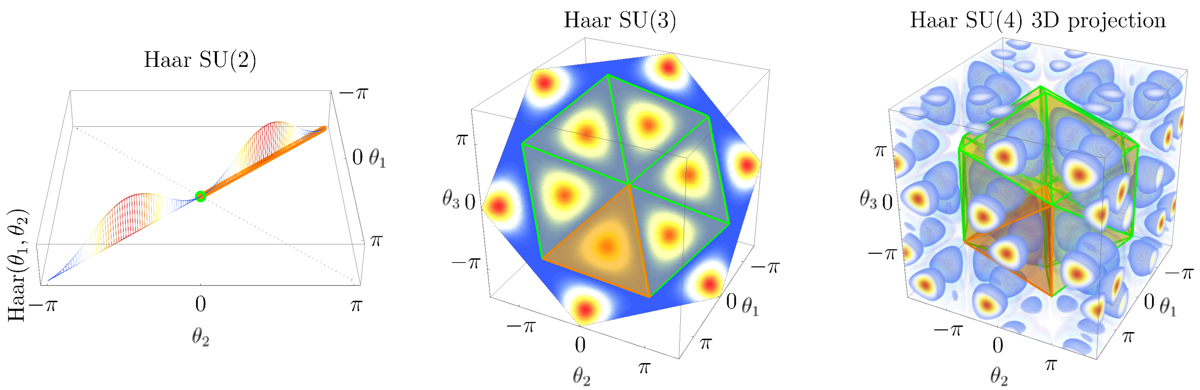




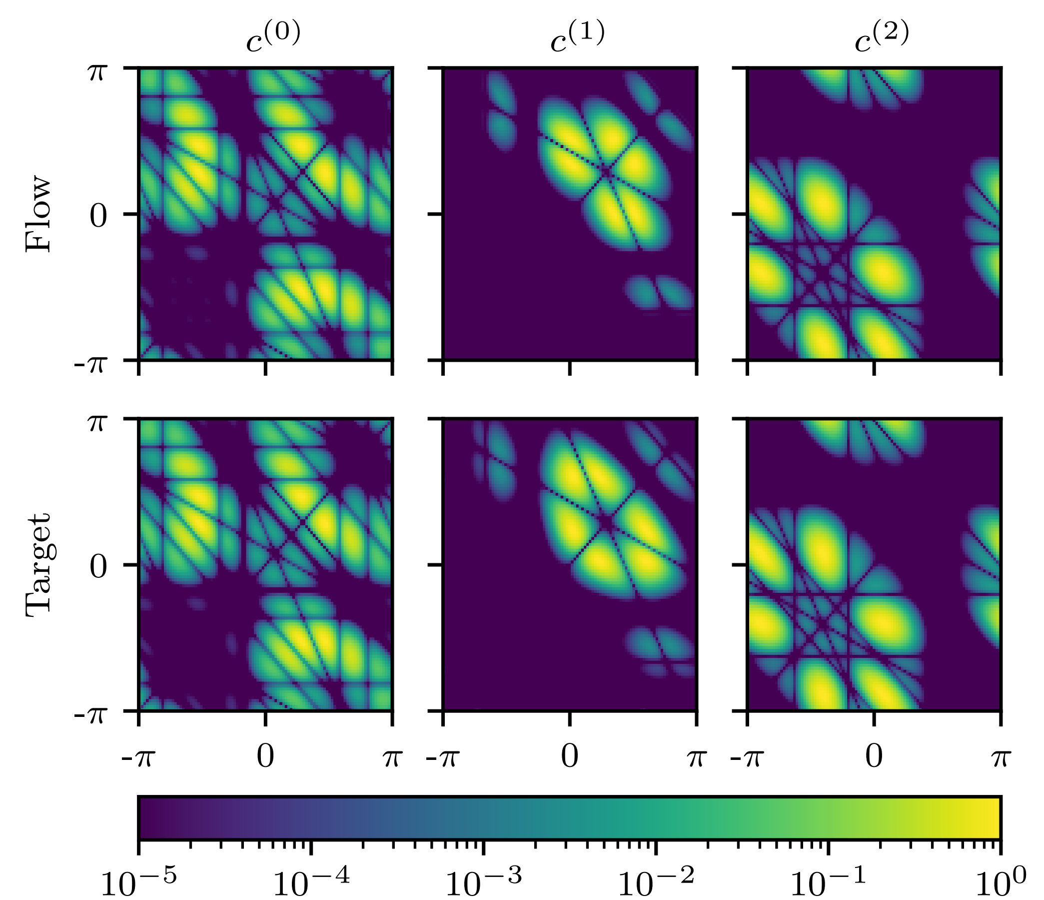
[Kanwar et al., arXiv:2003.06413]

Application: SU(N) field theory





SU(9) flows

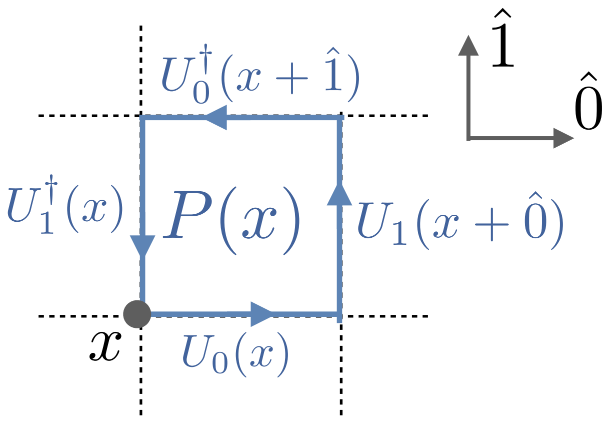


[2008.05456 (2020), PRL 125, 121601 (2020), 2002.02428 (2020)]

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



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

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

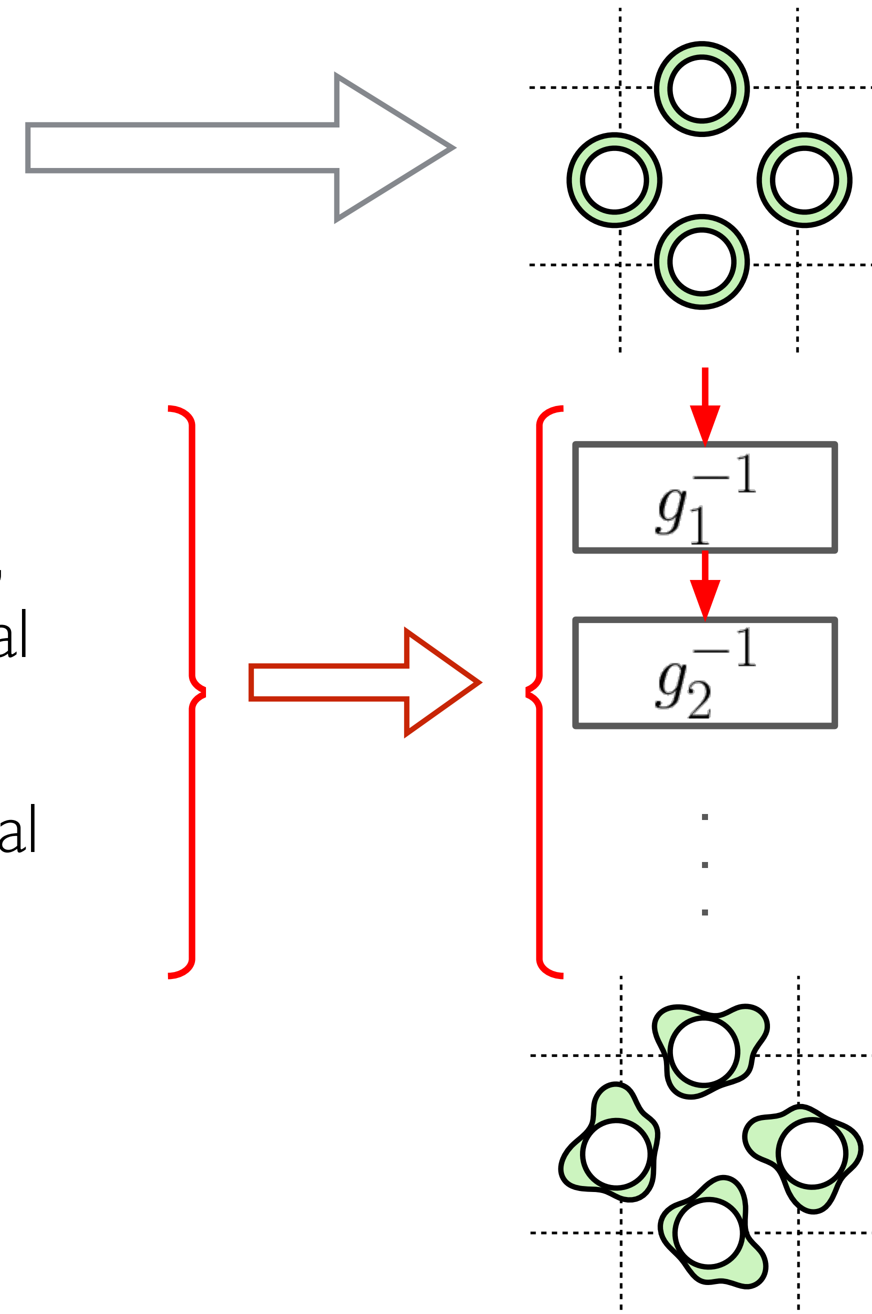
- Fixed lattice size: $L^2 = 16$ with couplings $\beta = \{1, 2, 3, 4, 5, 6, 7\}$
- Continuum limit (critical slow-down) as $\beta \to \infty$

[Kanwar et al., arXiv:2003.06413]

Application: U(1) field theory

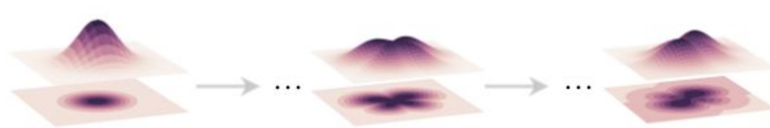
First gauge theory application: U(1) field theory

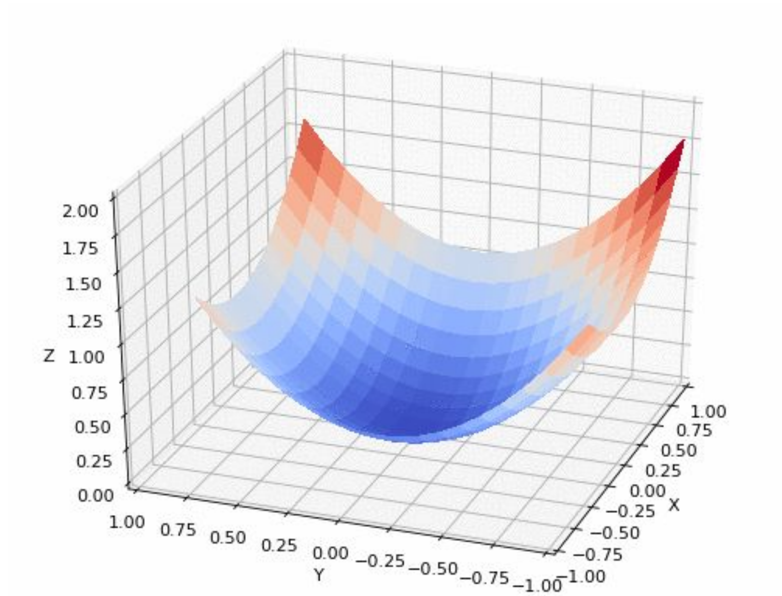
- Prior distribution chosen to be uniform
- Gauge-equivariant coupling layers
  - 24 coupling layers
  - Kernels h: mixtures of non-compact projections, 6 components, parameterised with convolutional NNs (i.e., NN output gives params. of NCP)
  - NNs with 2 hidden layers with 8x8 convolutional filters,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

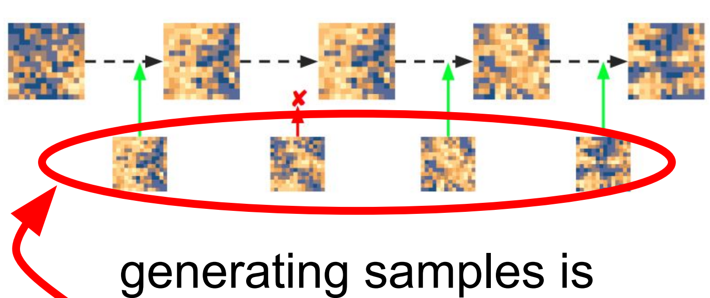


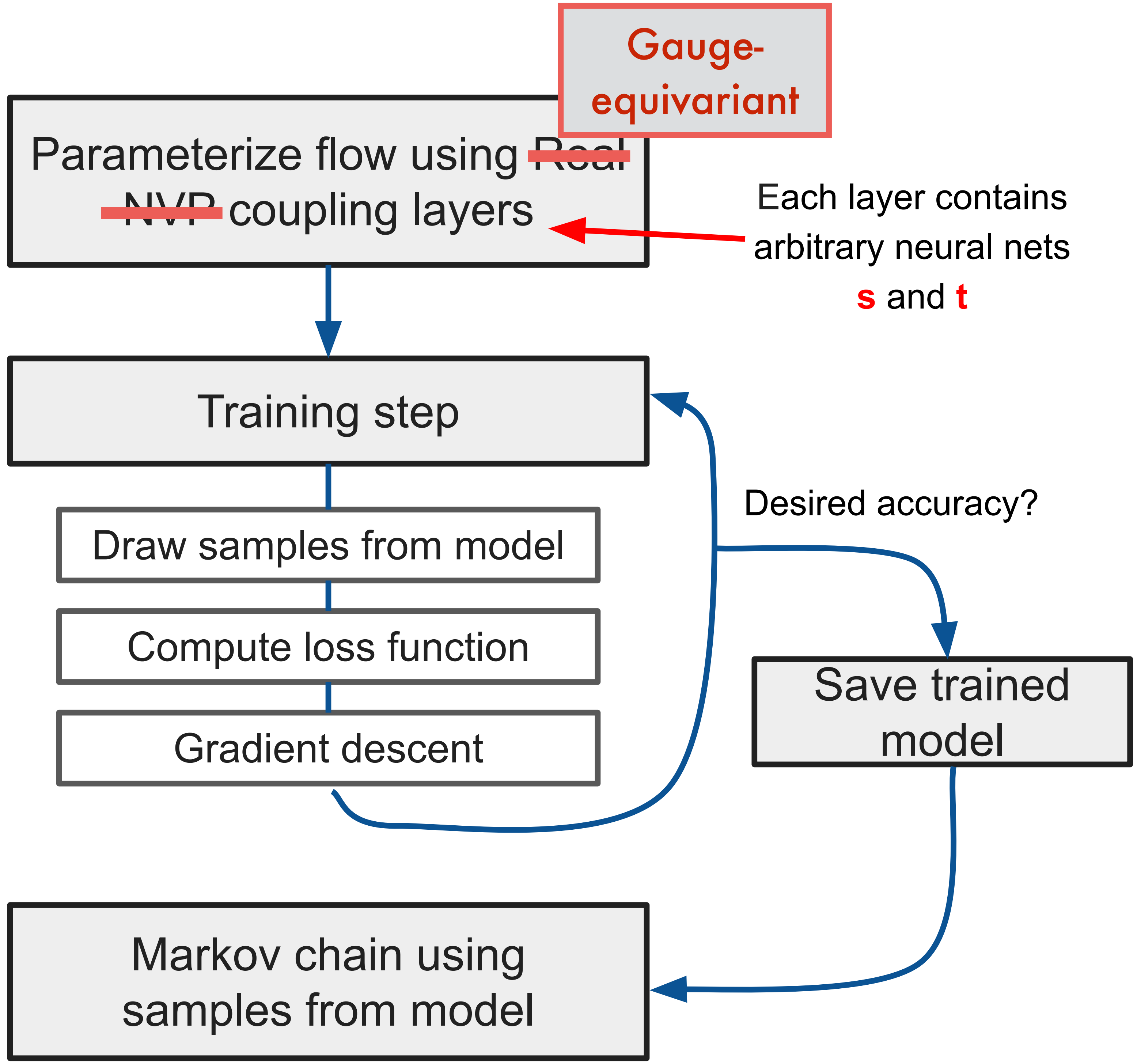
[Kanwar et al., arXiv:2003.06413]

Fields via flow models









Summary chart: Tej Kanwar

Application: U(1) field theory

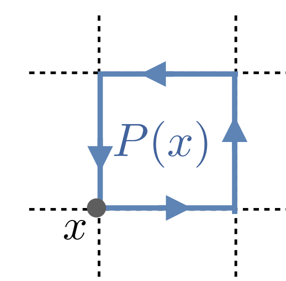
First gauge theory application: U(1) field theory

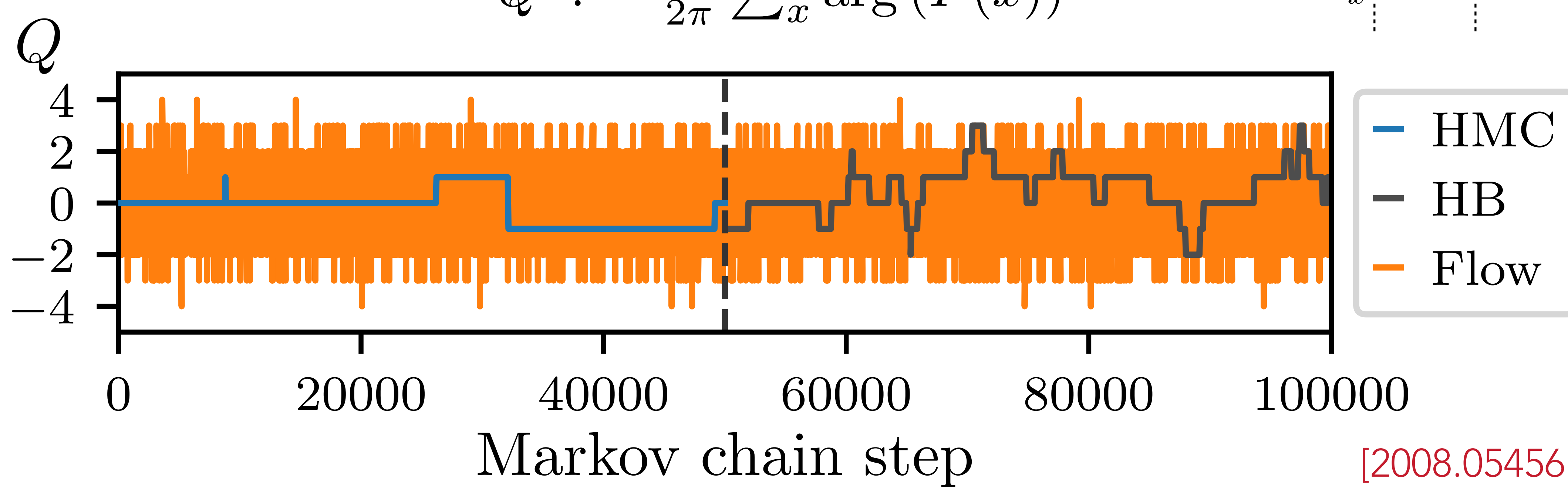
Success: Critical slowing down is significantly reduced

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

[2008.05456 (2020), PRL 125, 121601 (2020), 2002.02428 (2020)]

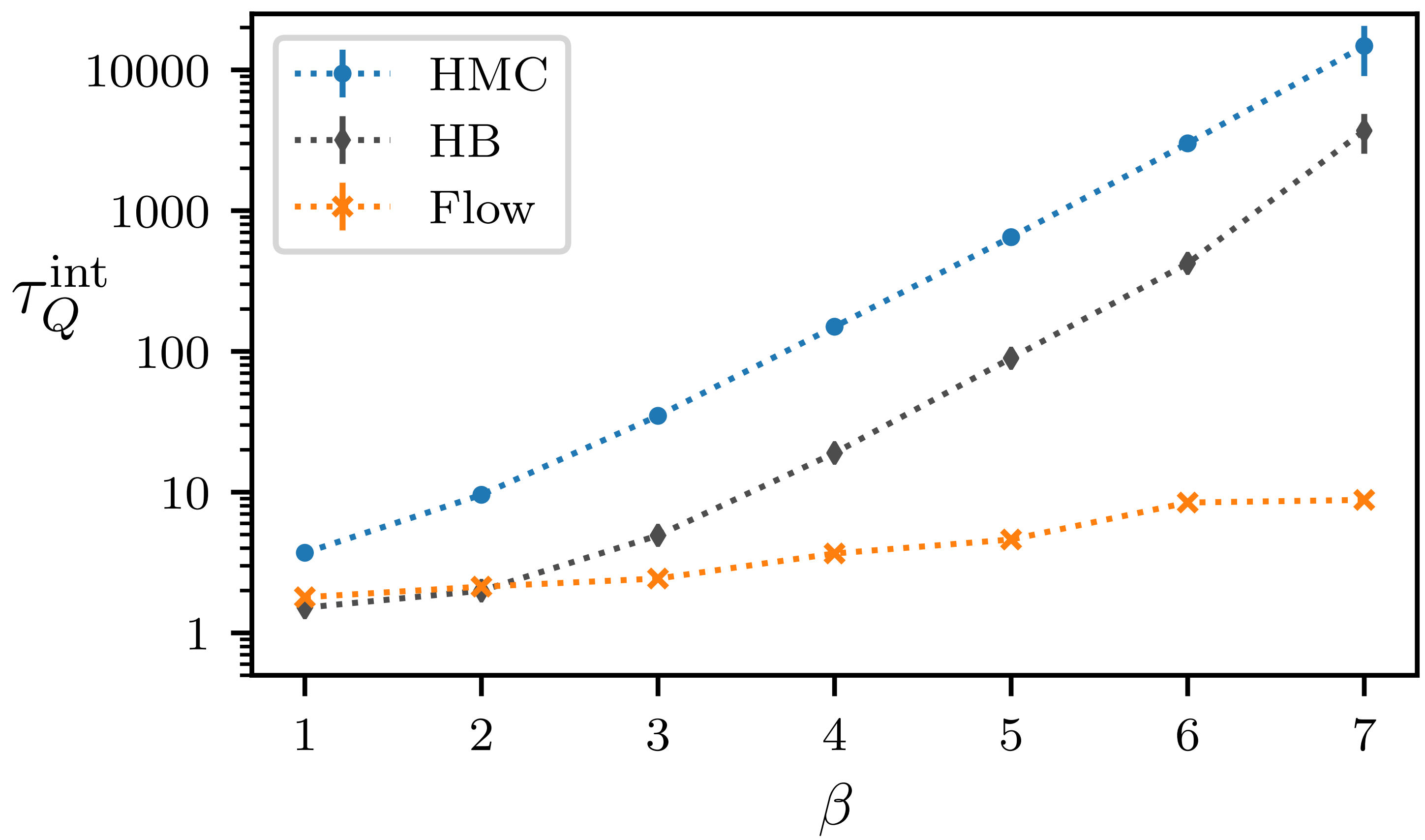
Application: U(1) field theory

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

[2008.05456 (2020),
PRL 125, 121601 (2020),
2002.02428 (2020)]

Application: U(1) field theory

First gauge theory application: U(1) field theory

Success: Critical slowing down is significantly reduced

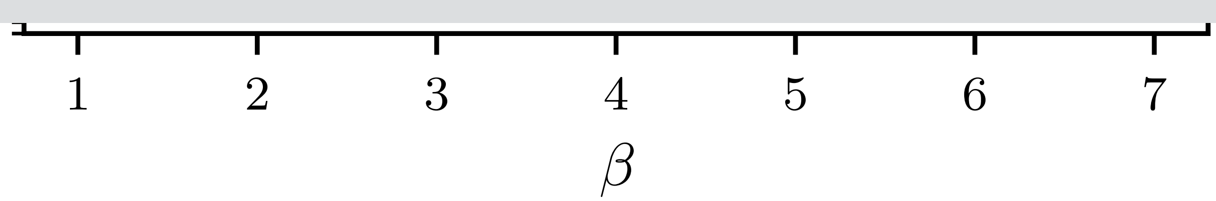
Co

SUCCESS!

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

↓

Significant work required to add fermions, scale to state-of-the-art

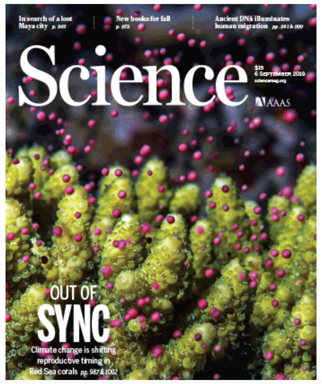


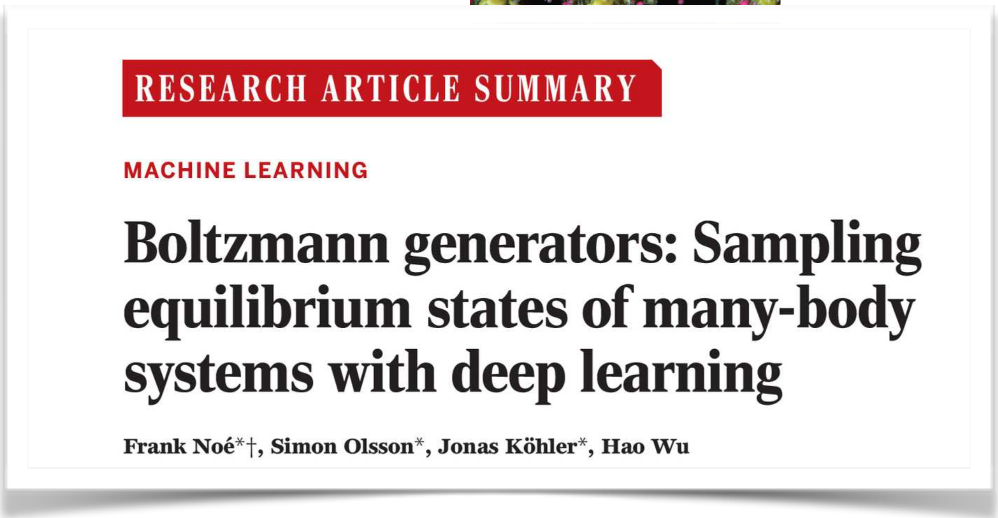
2D, L=16

[2008.05456 (2020), PRL 125, 121601 (2020), 2002.02428 (2020)]

Interdisciplinary applications

Molecular genetics and drug design



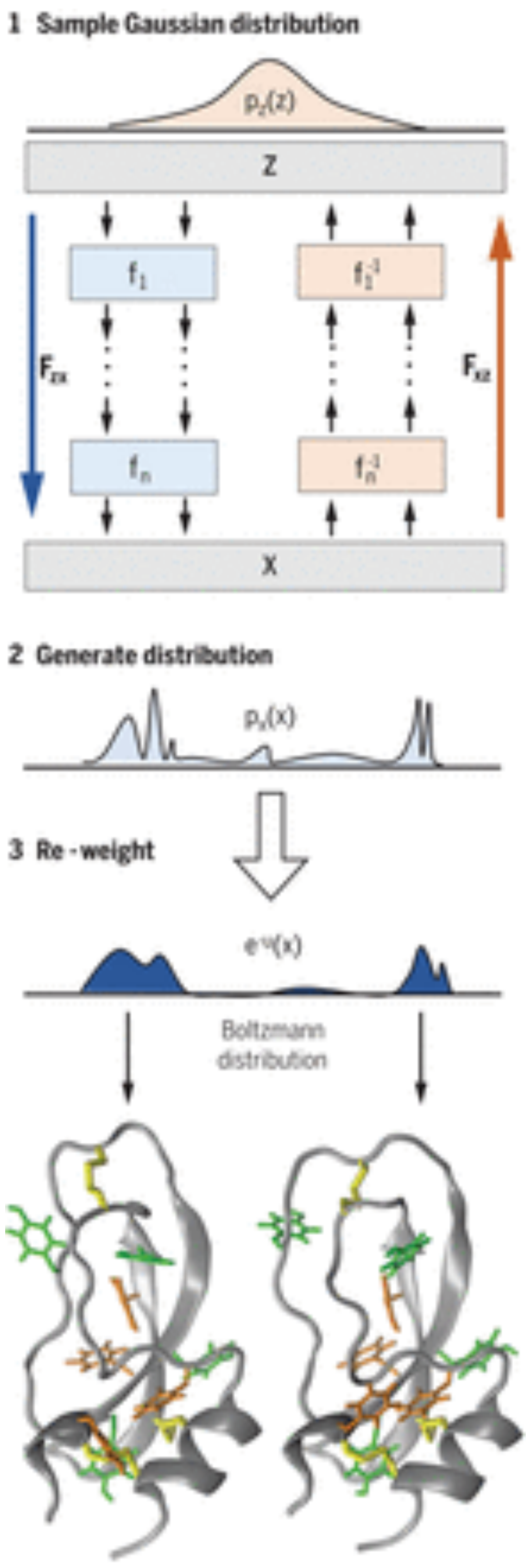


RESEARCH ARTICLE SUMMARY

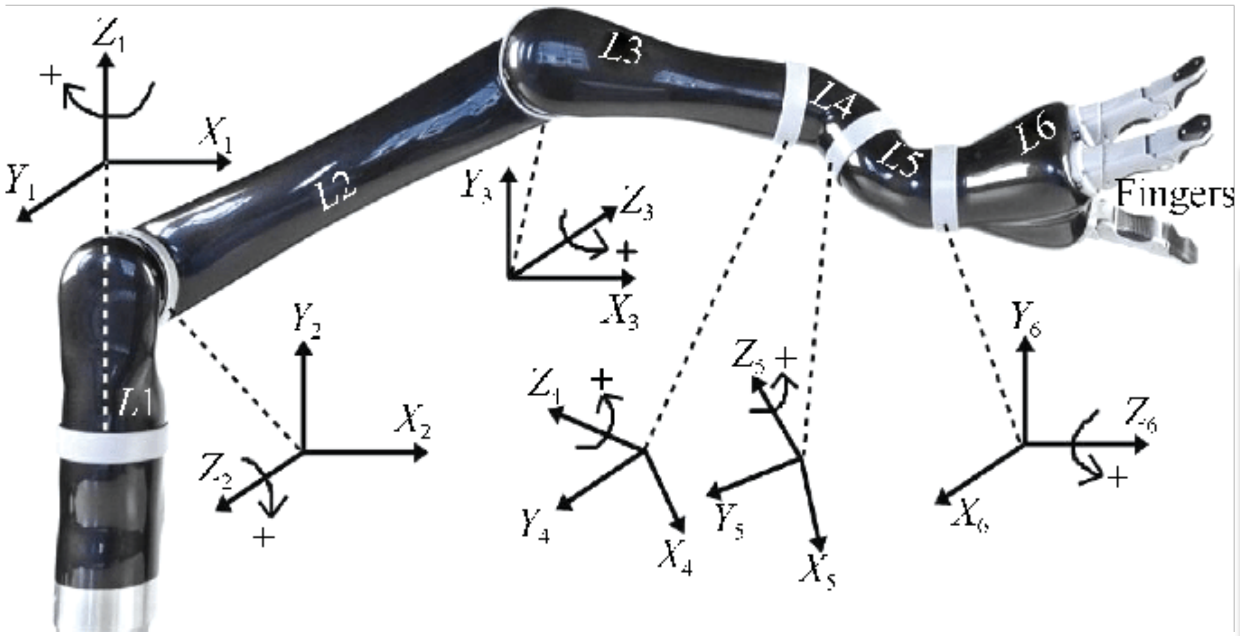
MACHINE LEARNING

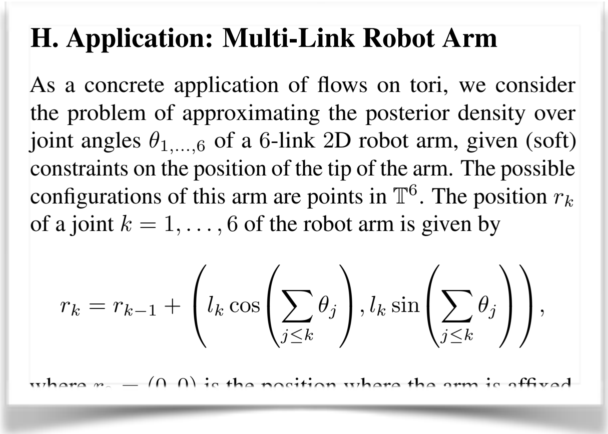
Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning

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 $\theta_{1,\dots,6}$ 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 $k = 1, \dots, 6$ of the robot arm is given by

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

Outlook

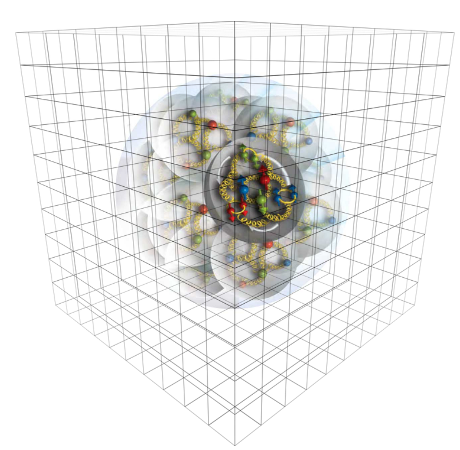
ML-accelerated algorithms have huge potential to enable first-principles nuclear physics studies

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
- ….



Emergence of complex structure in nature

Backgrounds and benchmarks for searches for new physics

Joint software effort

Our codes exploit and extend existing ML software frameworks

- Tensorflow
- Pytorch
- JAX

TensorFlow

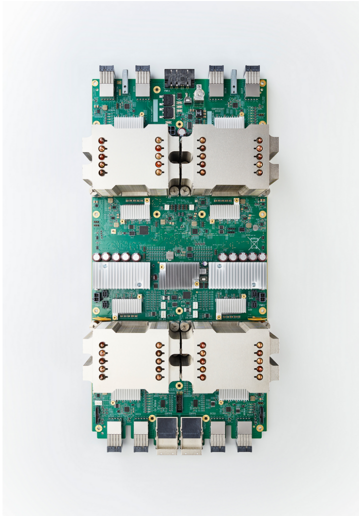
PYTORCH

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