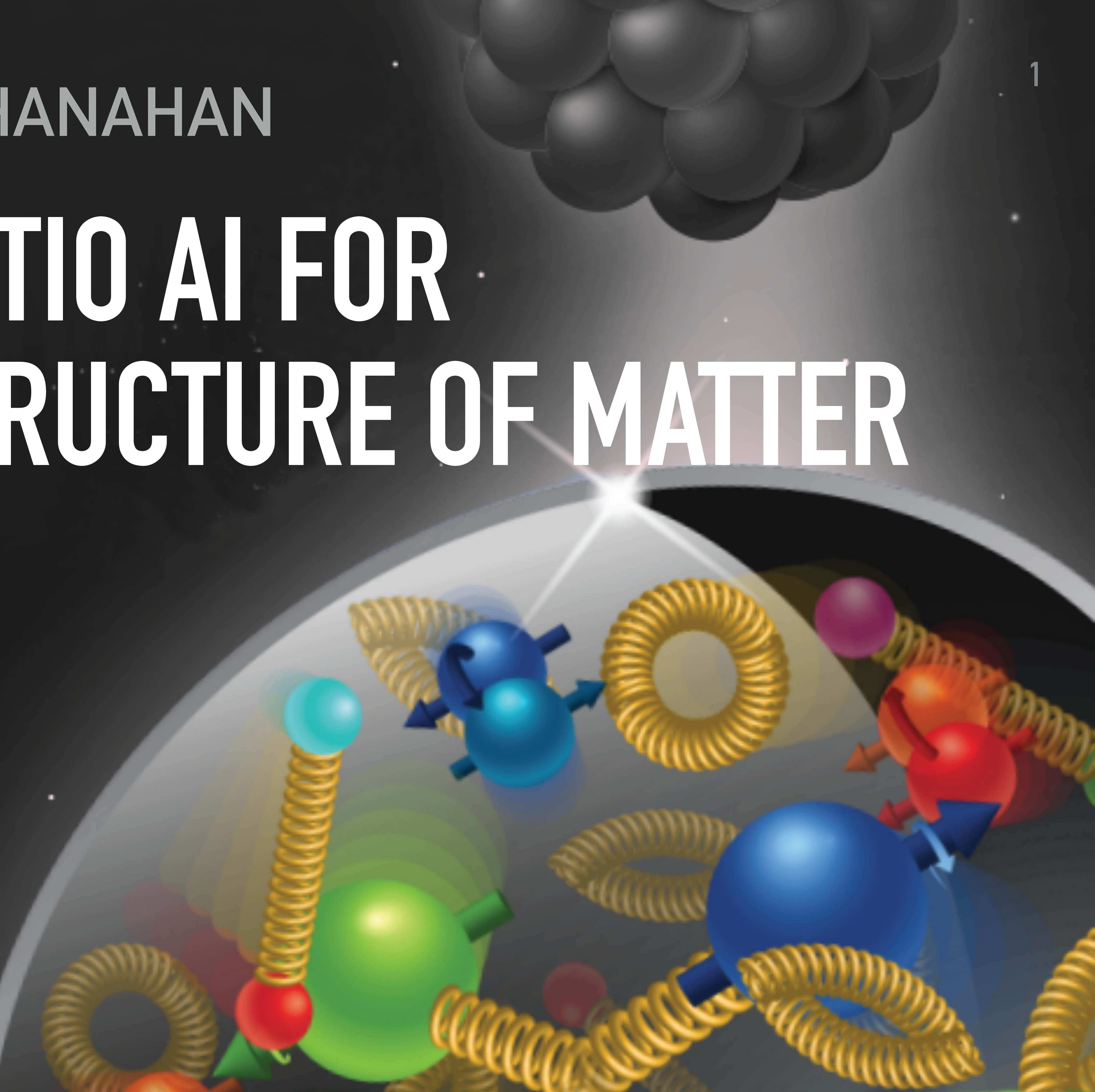
PHIALA SHANAHAN

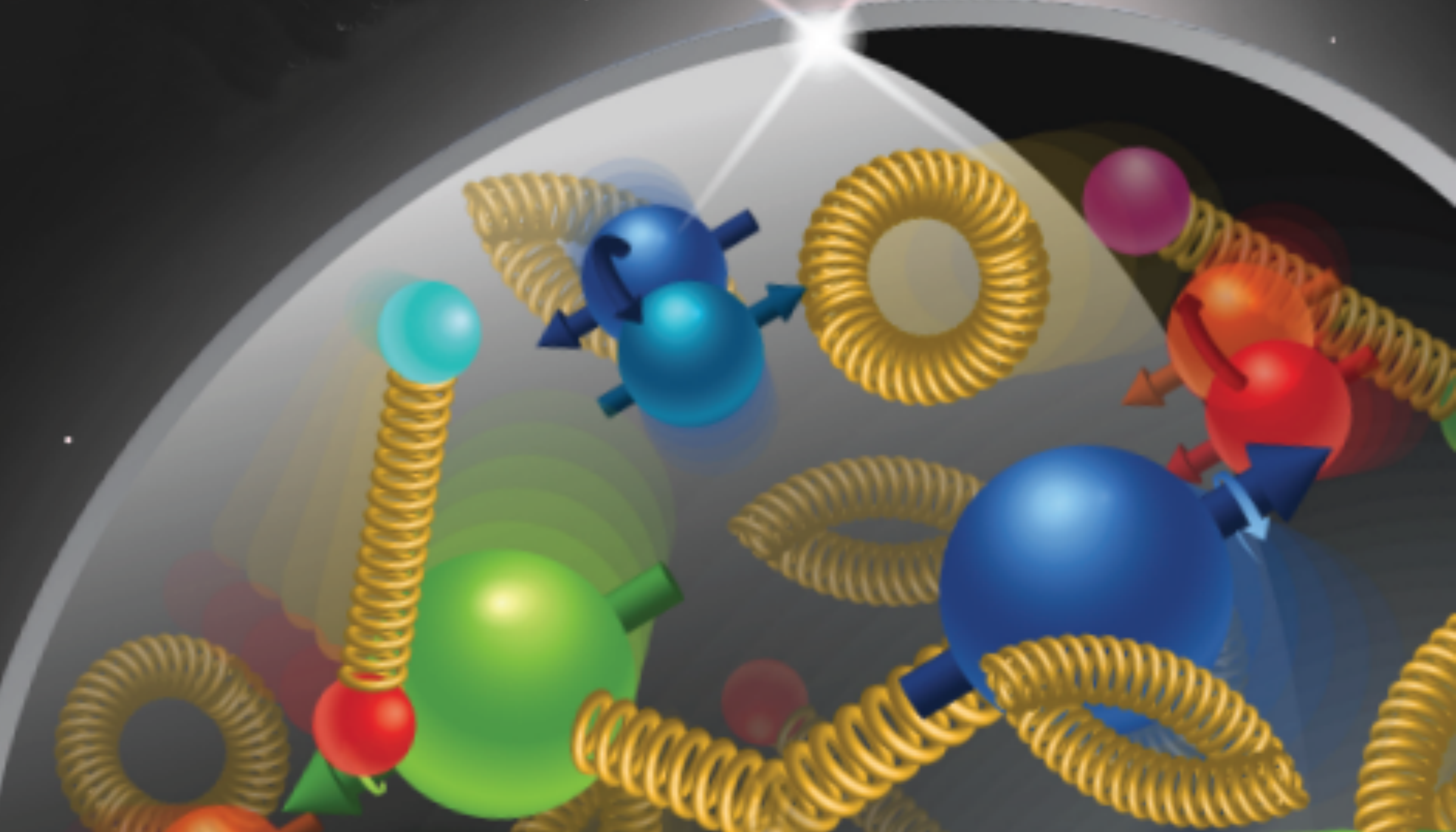
AB-INITIO AI FOR THE STRUCTURE OF MATTER



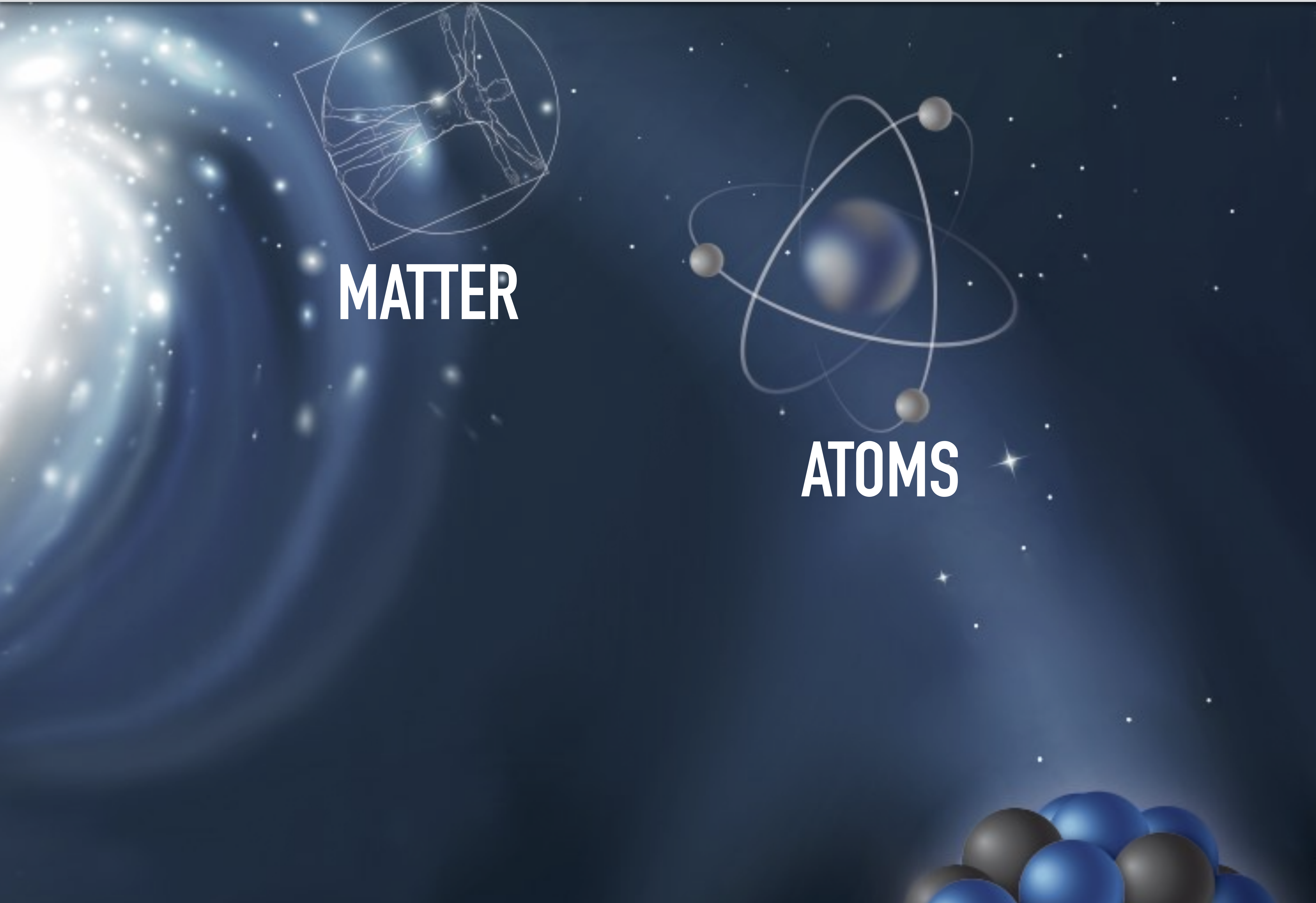
Massachusetts Institute of Technology

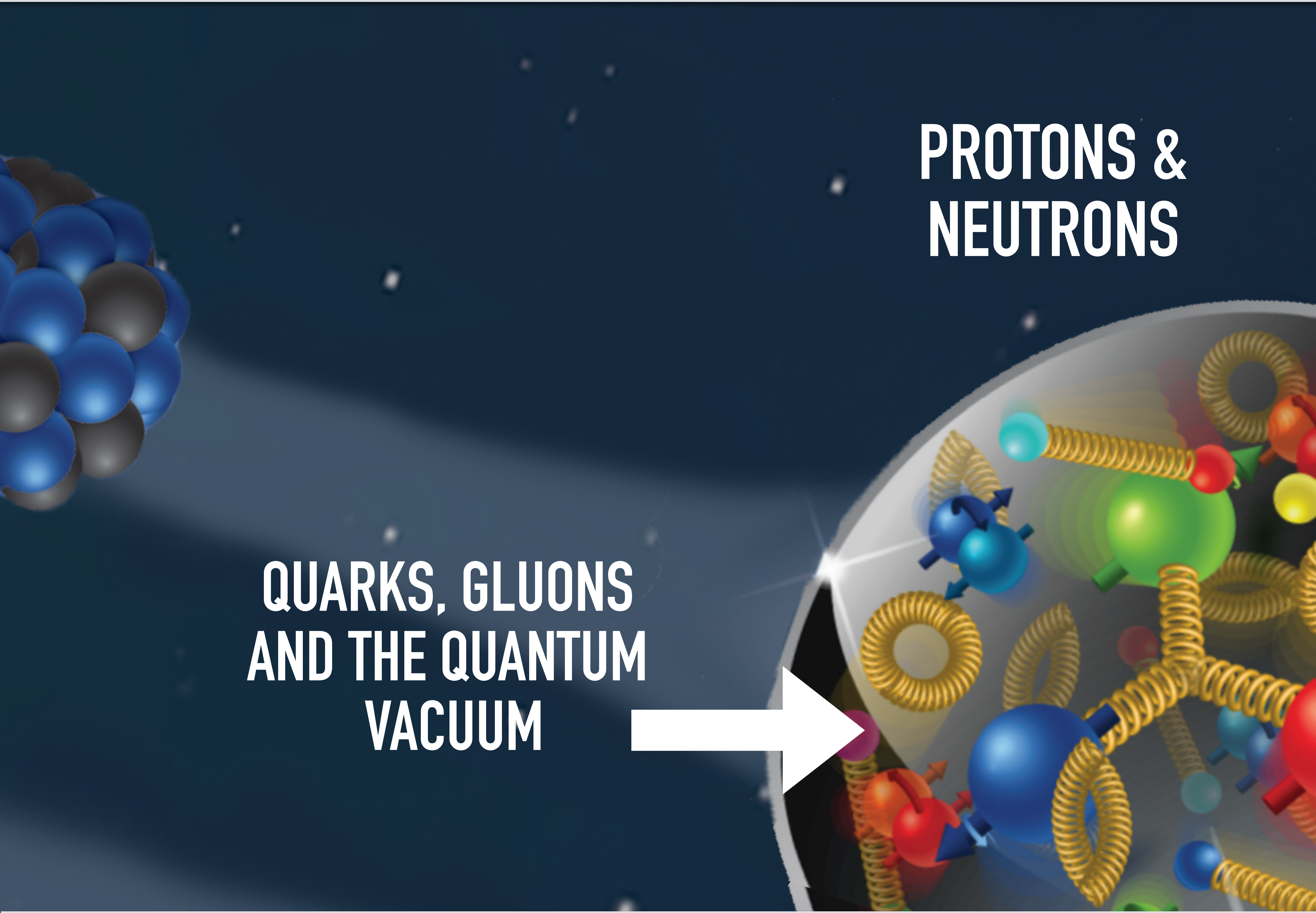
PHIALA SHANAHAN

AB-INITIO AI FOR THE STRUCTURE OF MATTER



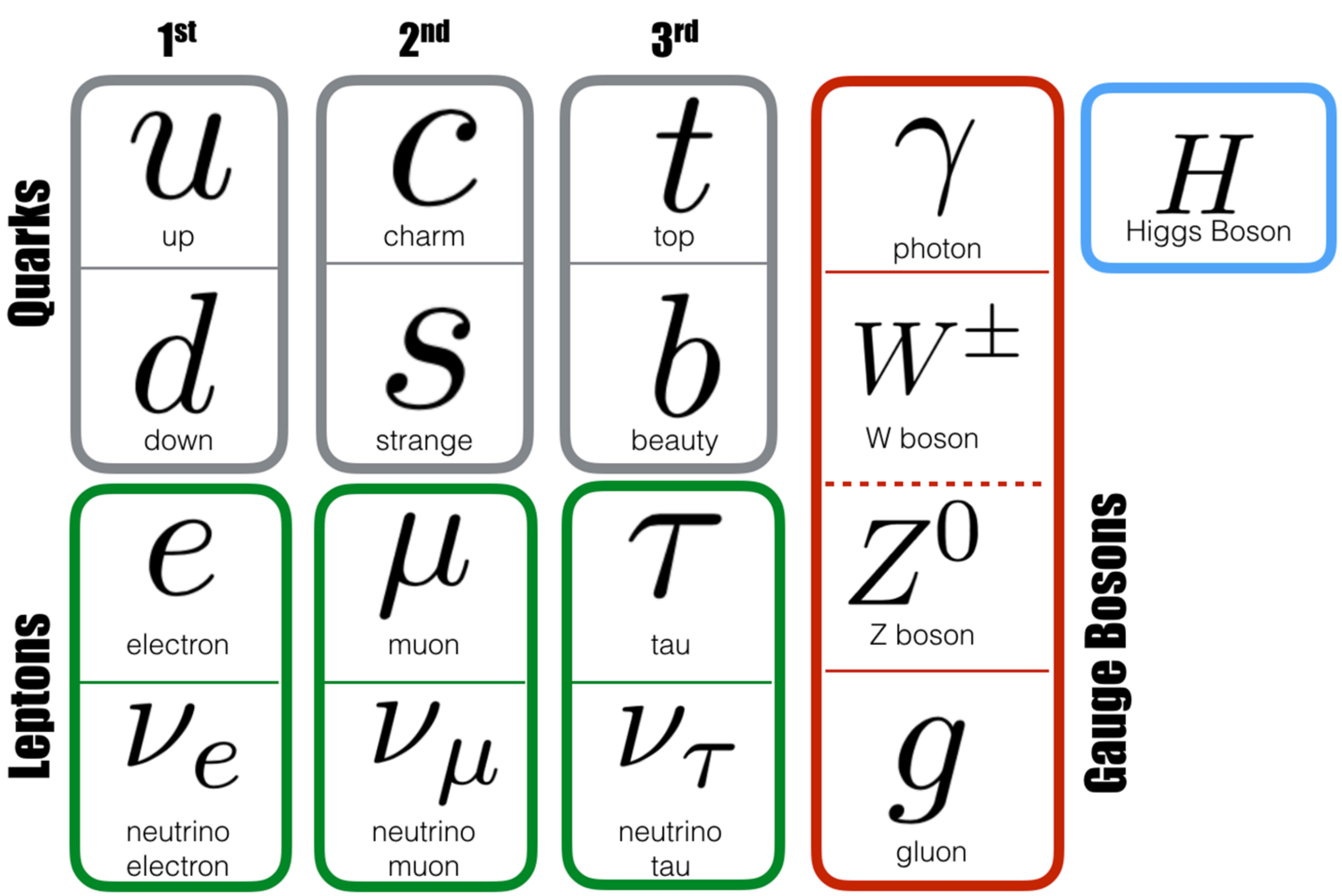
Massachusetts Institute of Technology

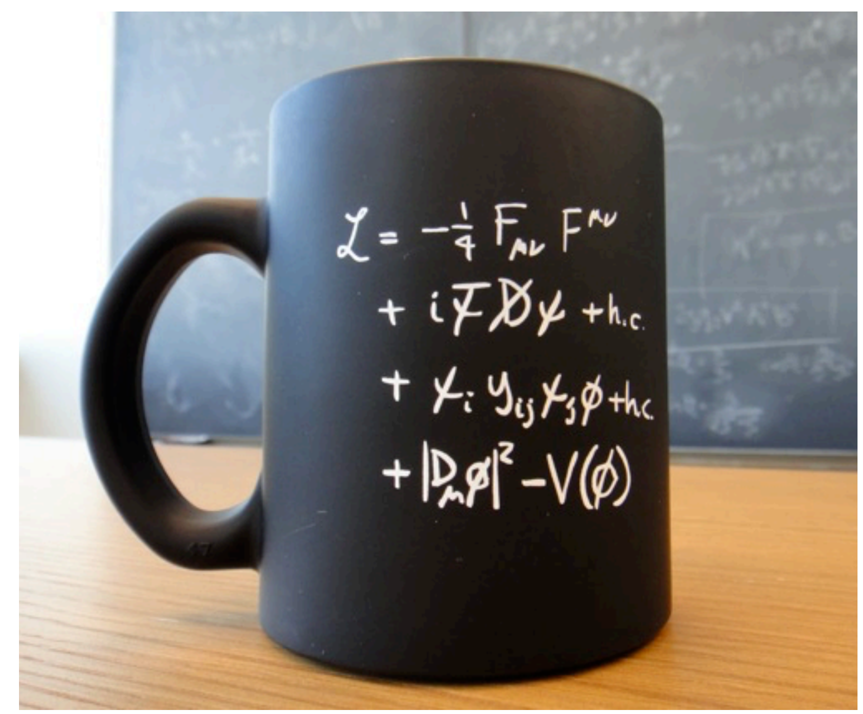

MATTER
ATOMS


PROTONS &
NEUTRONS
QUARKS, GLUONS
AND THE QUANTUM
VACUUM

The structure of matter

The Standard Model of nuclear and particle physics



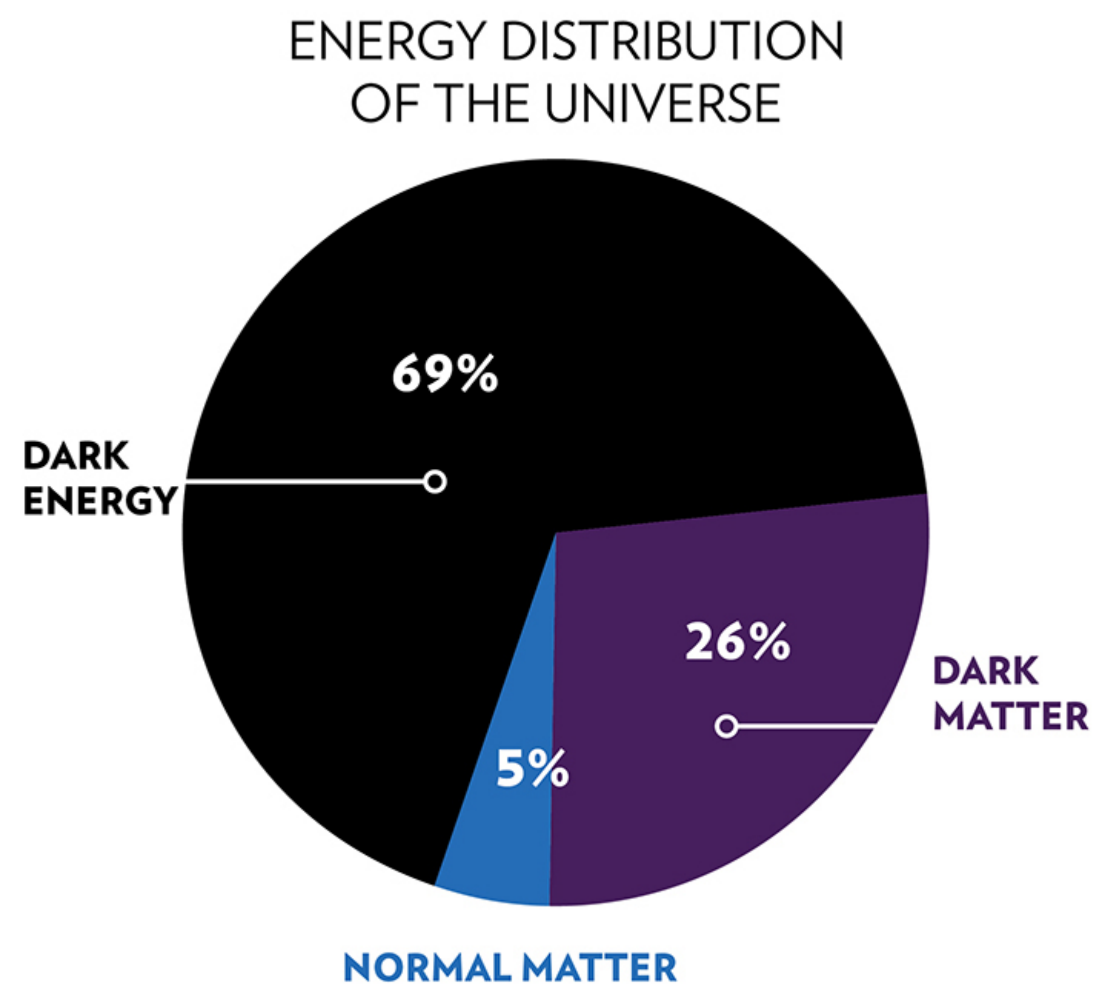


Phiala Shanahan, MIT

The structure of matter

BUT The Standard Model isn't everything

- Dark matter and dark energy
- Neutrino masses
- Matter–antimatter asymmetry
- Gravity
- Naturalness problems
- …

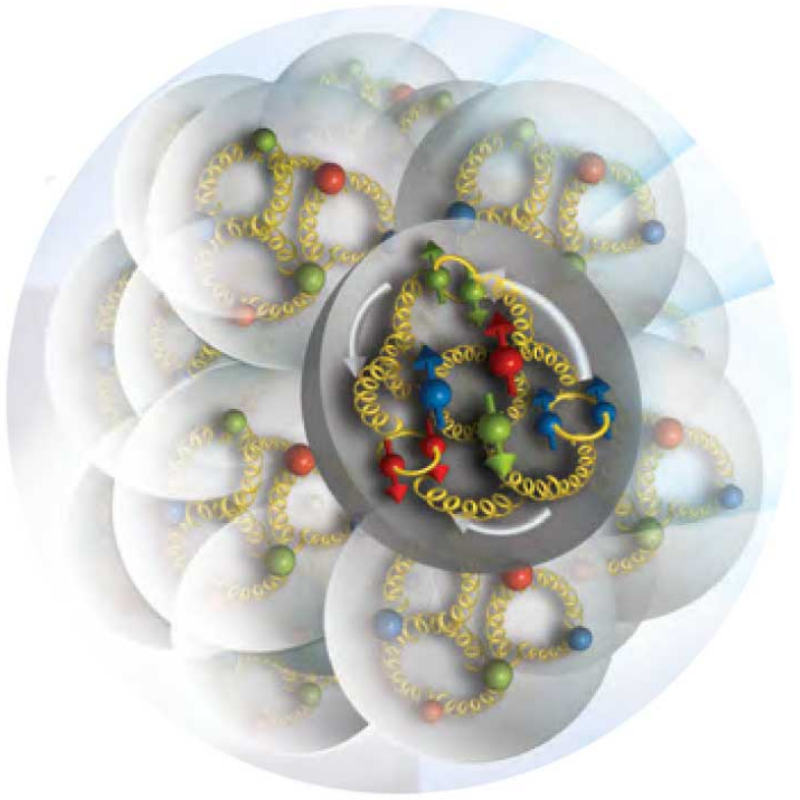


The structure of matter

Understanding the quark and gluon structure of matter

Emergence of complex structure in nature

Backgrounds and benchmarks for searches for new physics

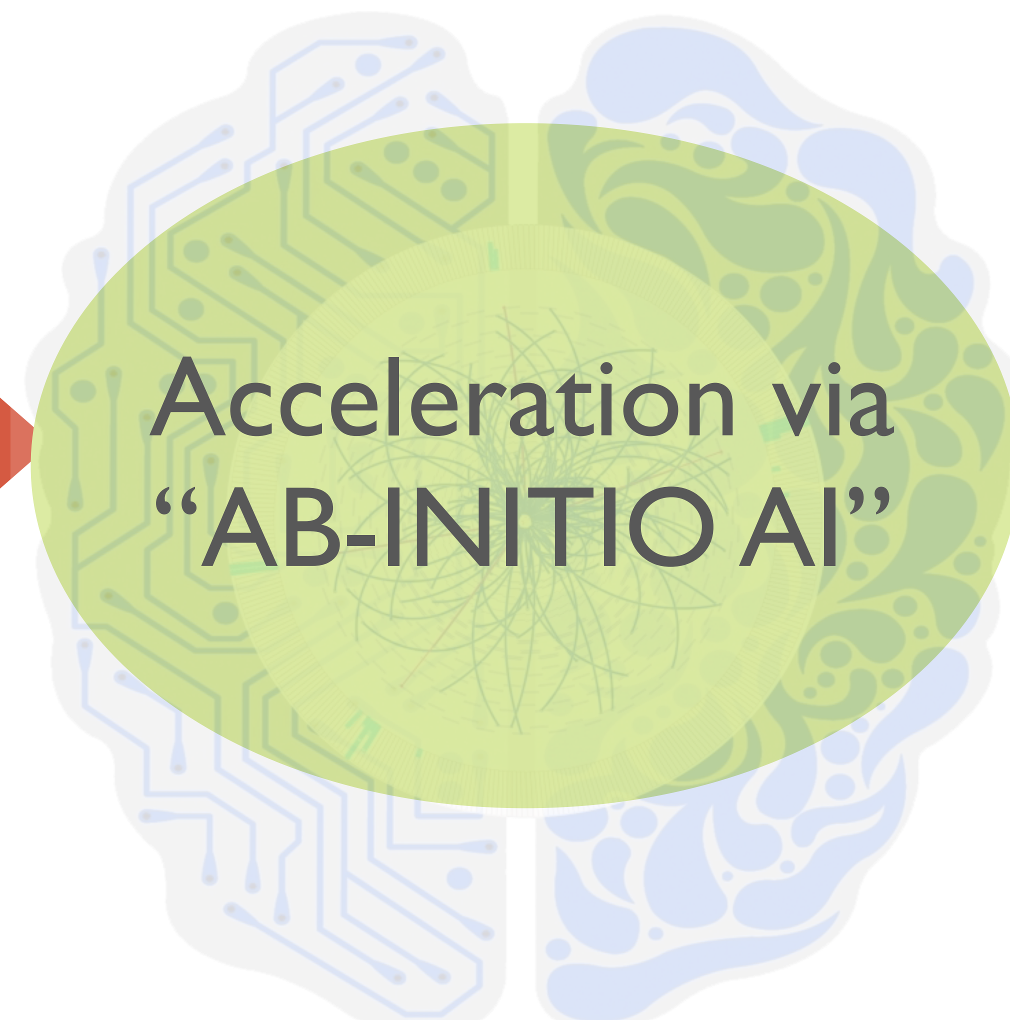


The structure of matter

First-principles studies of the Standard Model of nuclear and particle physics

- Demand extreme-scale computation

- Require guarantees of exactness, incorporation of complex symmetries



Acceleration via "AB-INITIO AI"

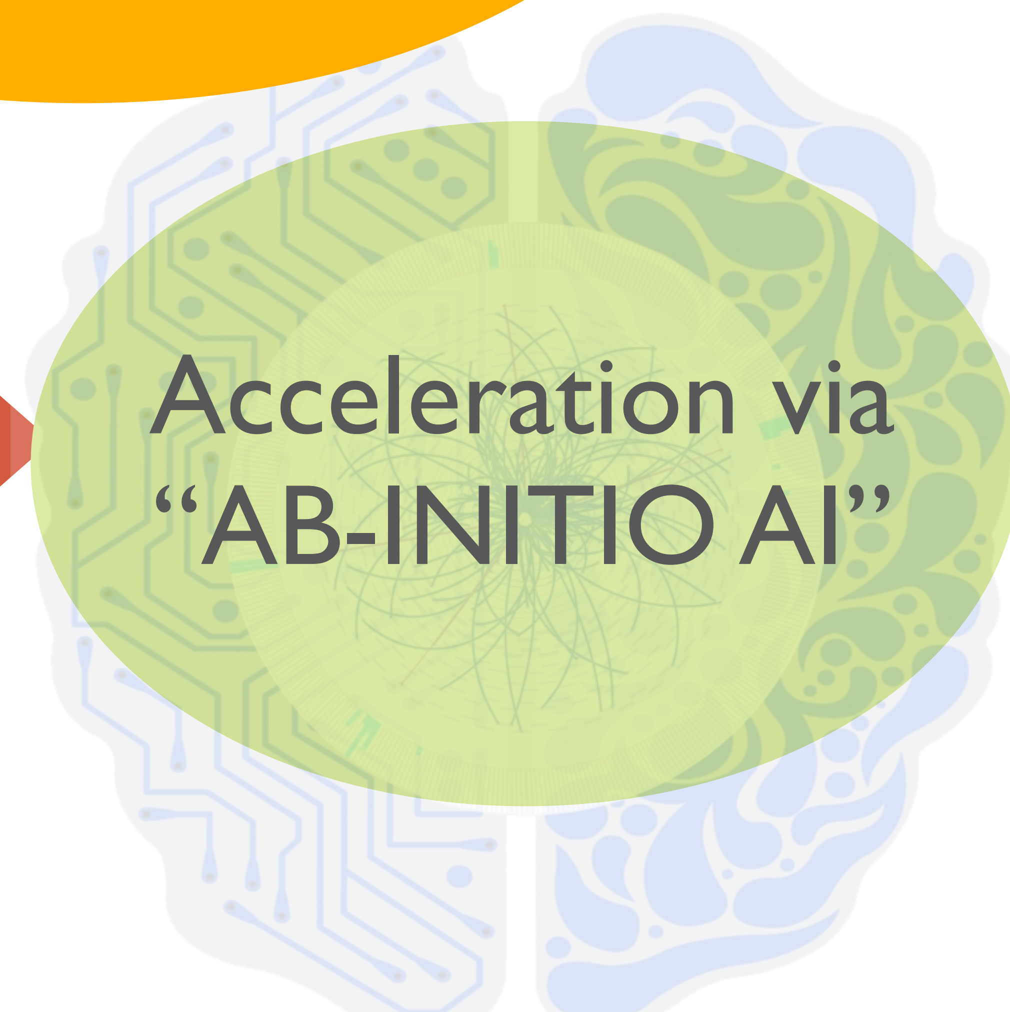
The structure of matter

First-principl... Model of nuc...

E.g., Not enough supercomputing in the world to compute Standard Model prediction for dark matter scattering from detectors!

- Demand extreme-scale computation

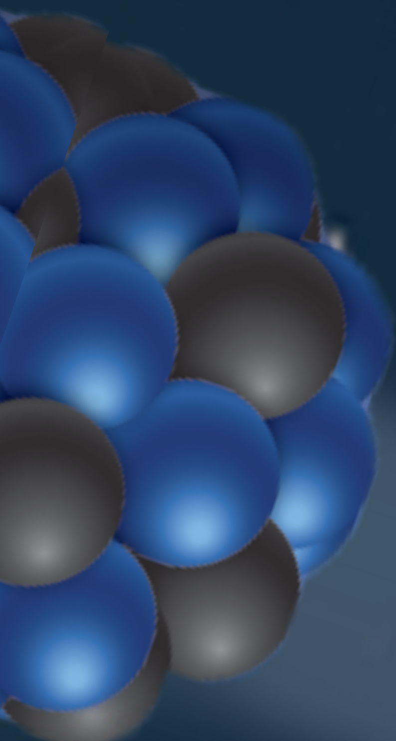
- Require guarantees of exactness, incorporation of complex symmetries



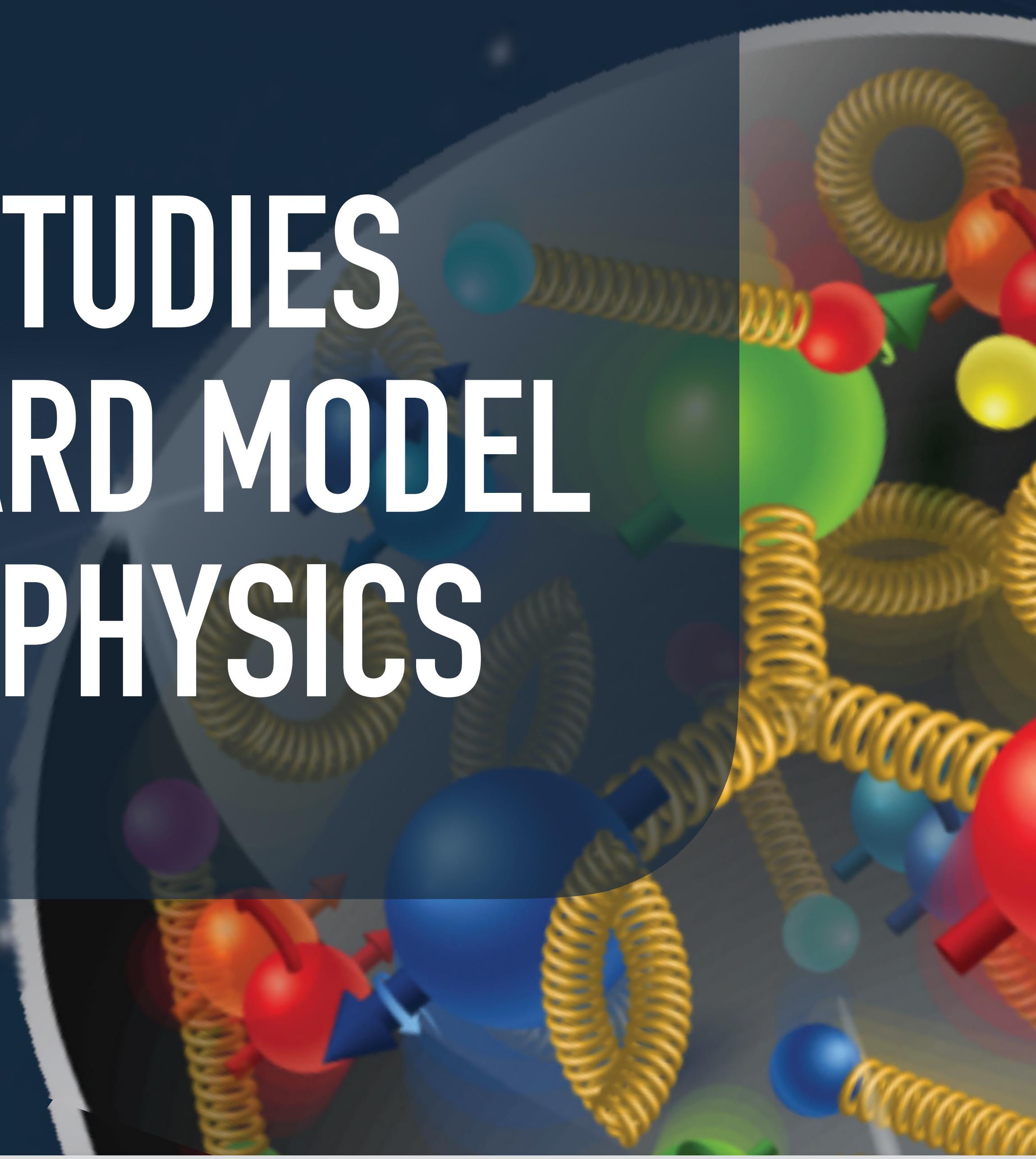
IAIFI: *Ab-initio* AI

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

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



AB-INITO AI
FOR
AB-INITIO STUDIES
OF THE STANDARD MODEL
OF PARTICLE PHYSICS

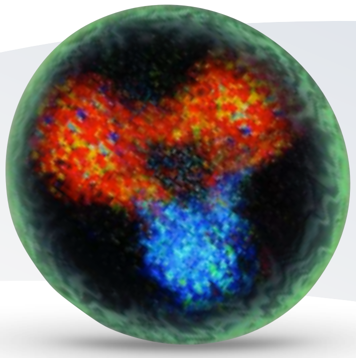


Strong interactions

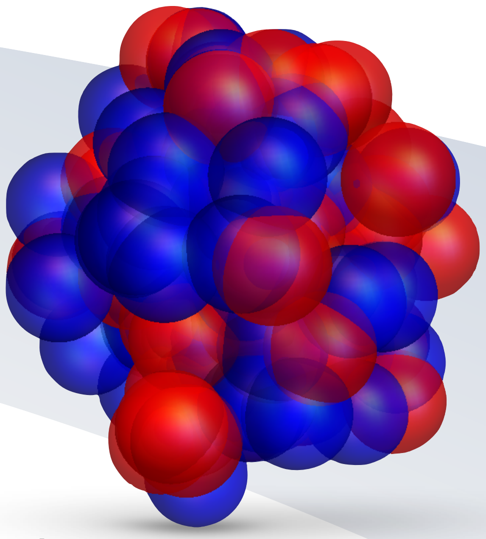
Study nuclear structure from the strong interactions

Quantum Chromodynamics (QCD)

Strongest of the four forces in nature

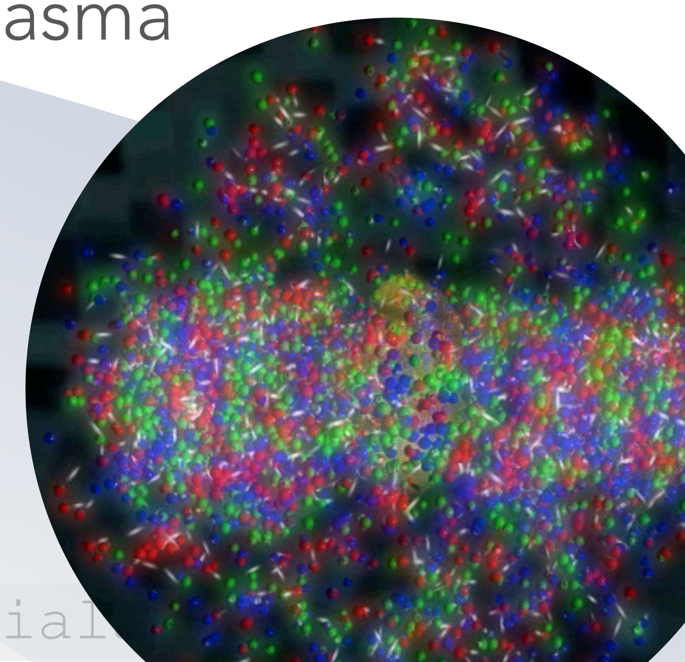


Binds quarks and gluons into protons, neutrons, pions etc.



Binds protons and neutrons into nuclei

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



Strong interactions

Interaction strength depends on energy

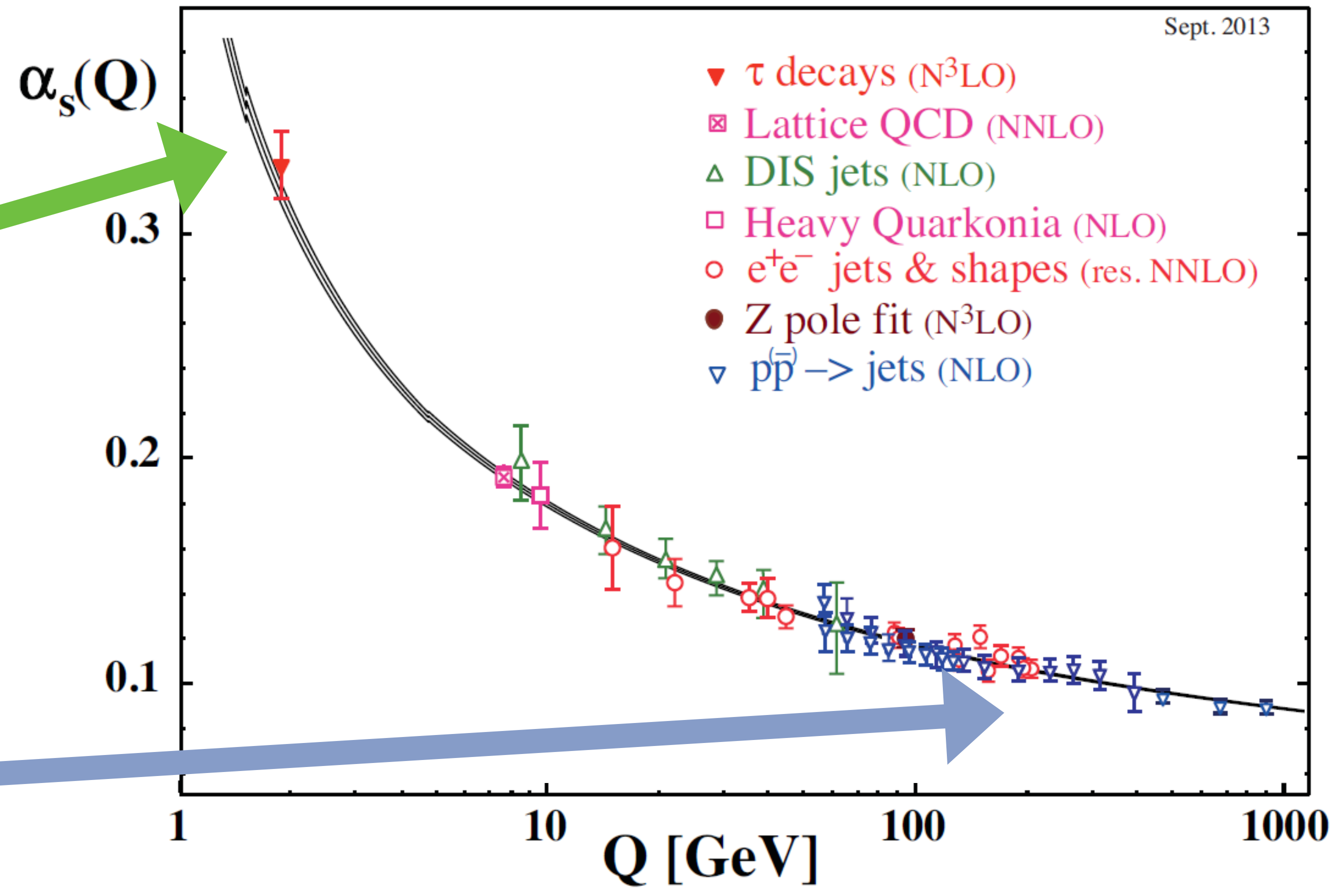
[Gross, Politzer, Wilczek, Nobel 2004]

Low-energy QCD is non-perturbative

Perturbation theory at high energies

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

Strong coupling

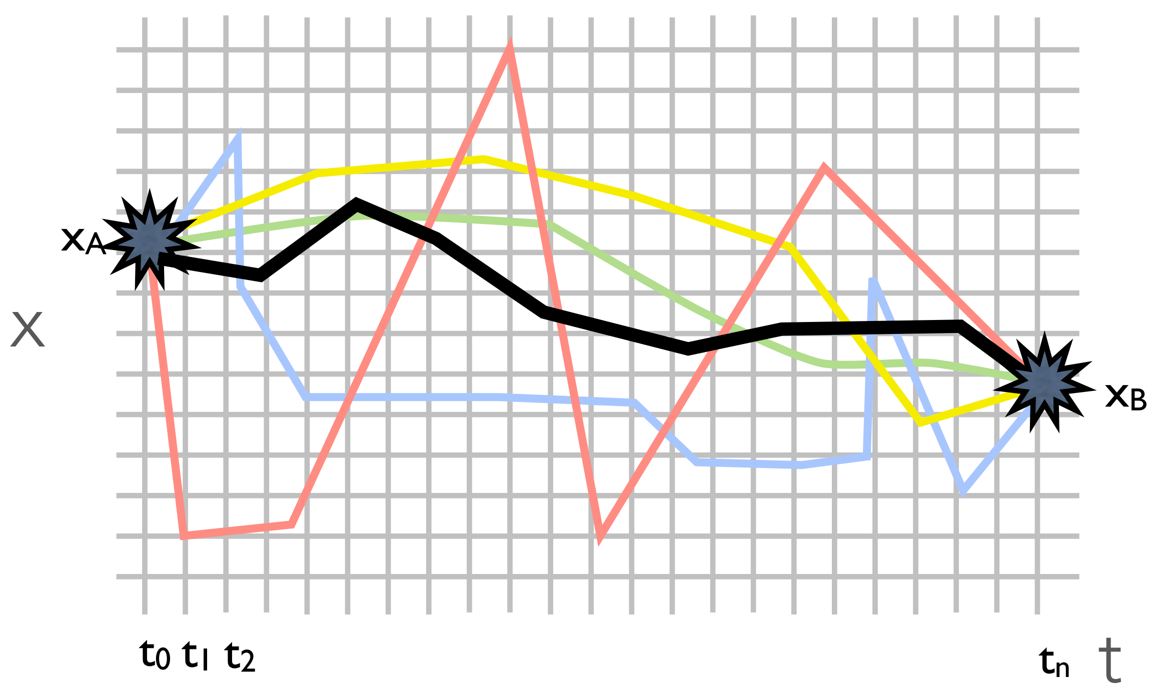


Energy

Lattice QCD

Numerical first-principles approach to non-perturbative QCD

- QCD equations ⟷ integrals over the values of quark and gluon fields on each site/link (QCD path integral)
- ~10^{12} variables (for state-of-the-art)



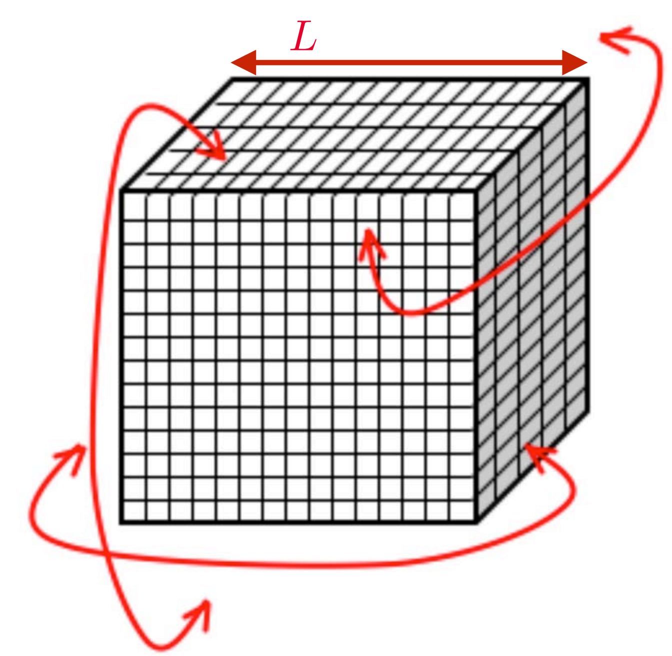
- Evaluate by importance sampling
- Paths near classical action dominate
- Calculate physics on a set (ensemble) of samples of the quark and gluon fields

Lattice QCD

Numerical first-principles approach to non-perturbative QCD

4D Euclidean space-time grid

- Non-zero lattice spacing
  (take limit as spacing becomes small)
- Volume $L^3 \times T \approx 64^3 \times 128$



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

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

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

Phiala Shanahan, MIT

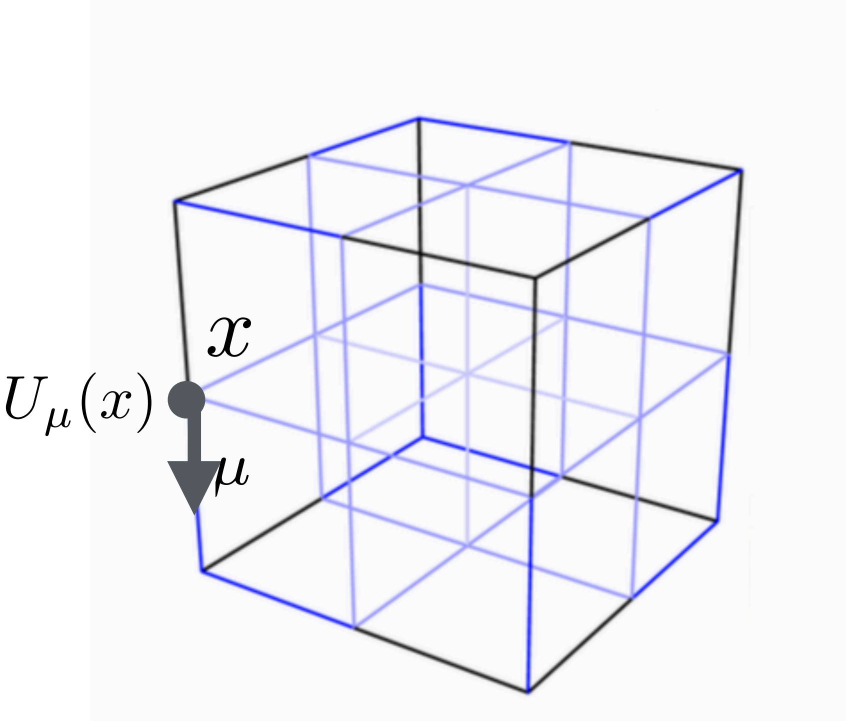
Lattice QCD

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



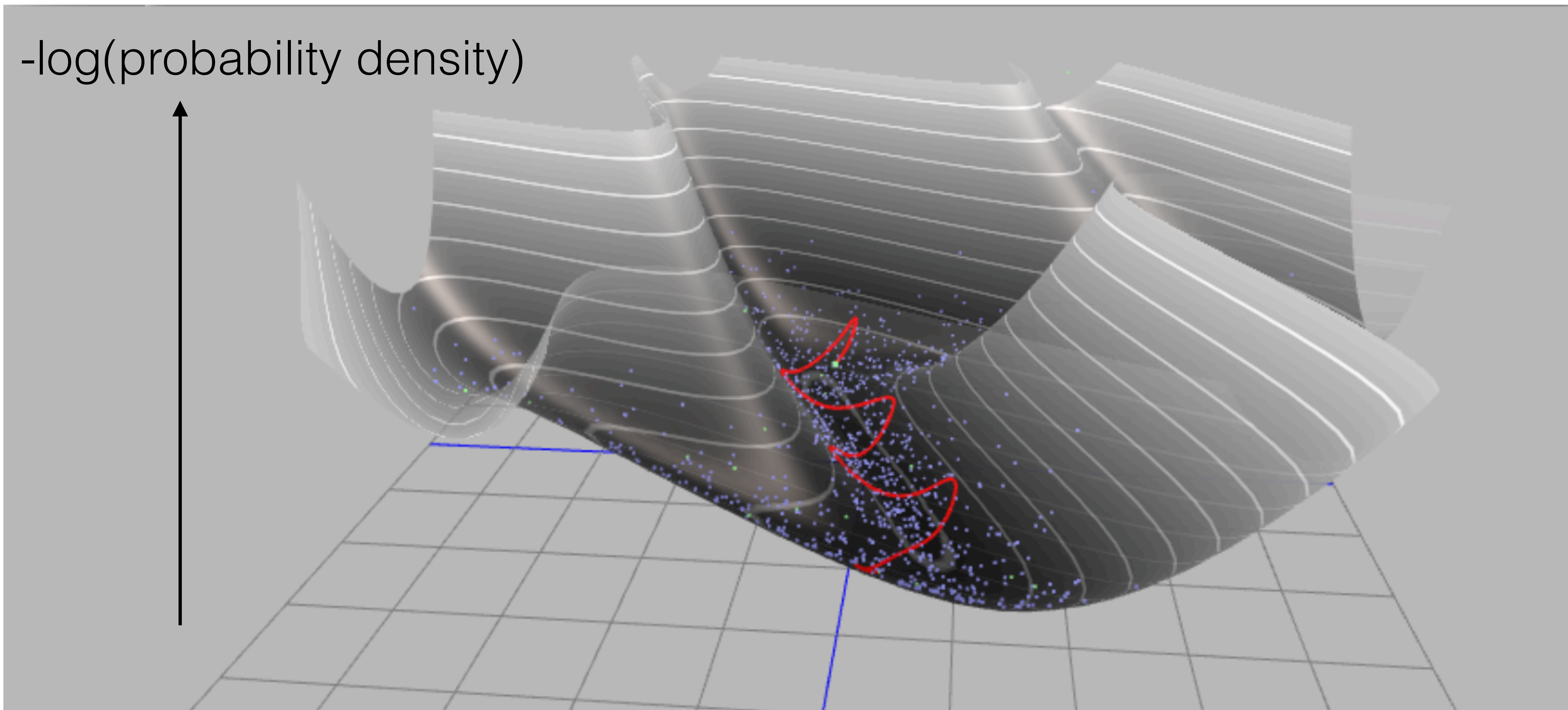
- Calculations use $\sim 10^3$ configurations

Phiala Shanahan, MIT

Generate QCD gauge fields

QCD gauge field configurations sampled via

Hamiltonian dynamics + Markov Chain Monte Carlo

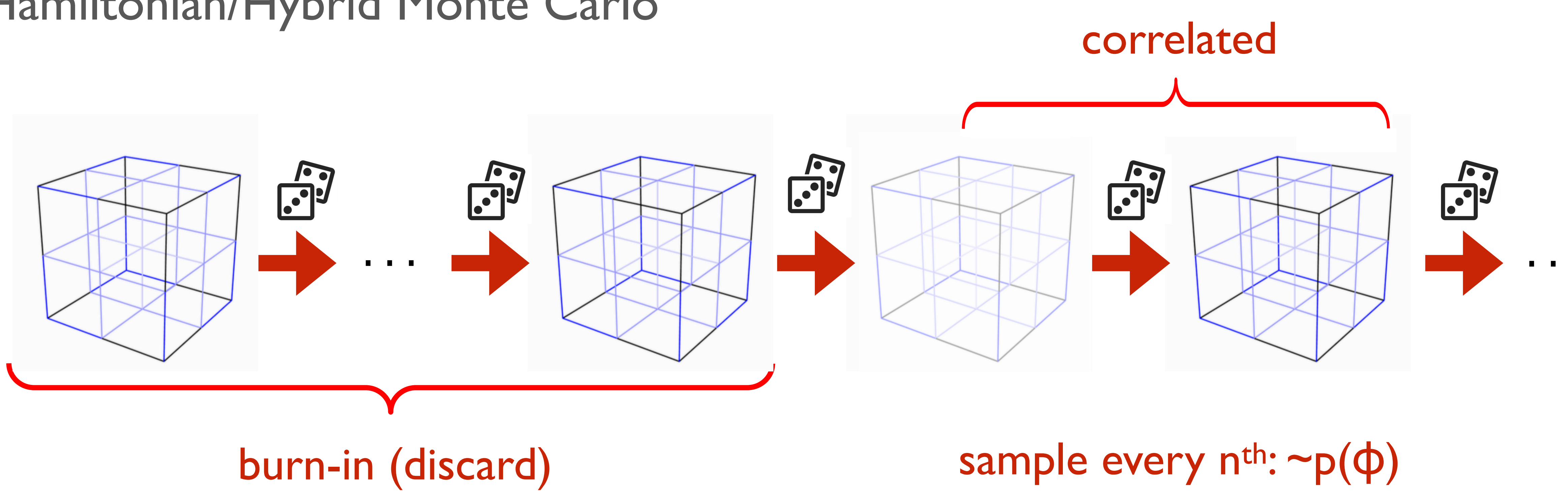


Generate QCD gauge fields

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

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

Hamiltonian/Hybrid Monte Carlo

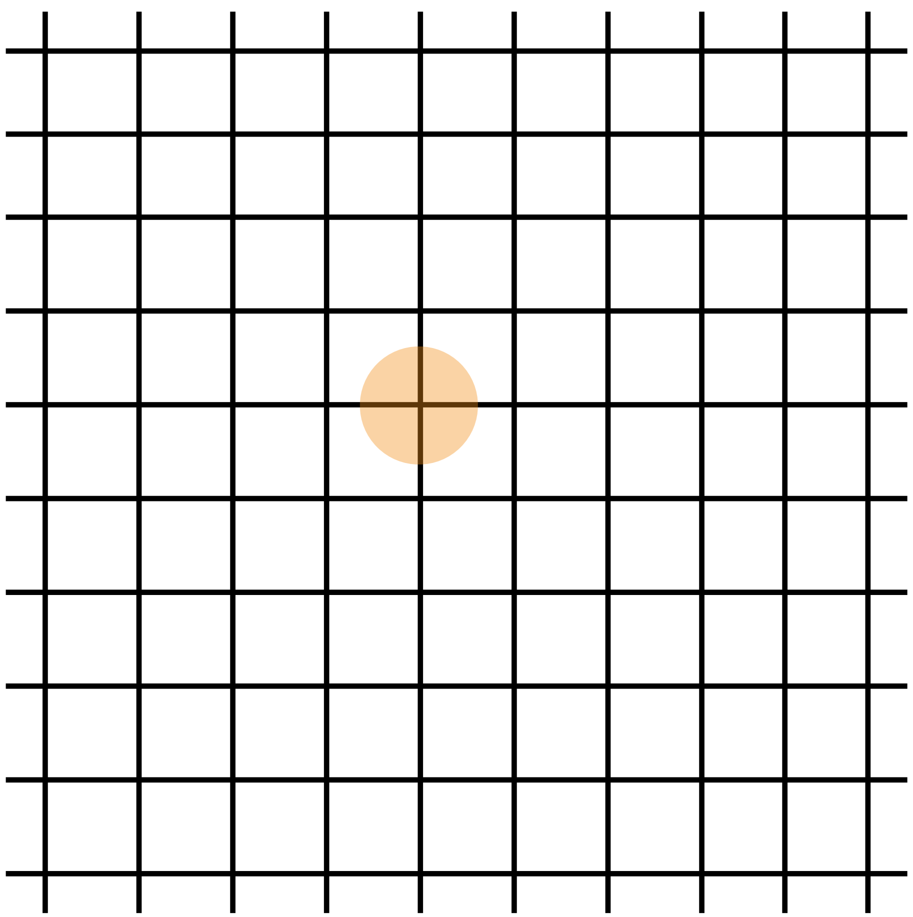


Burn-in time and correlation length dictated by Markov chain '**auto-correlation time**': shorter autocorrelation time implies less computational cost

Generate QCD gauge fields

QCD gauge field configurations sampled via

Hamiltonian dynamics + Markov Chain Monte Carlo



Updates diffusive

Lattice spacing → 0

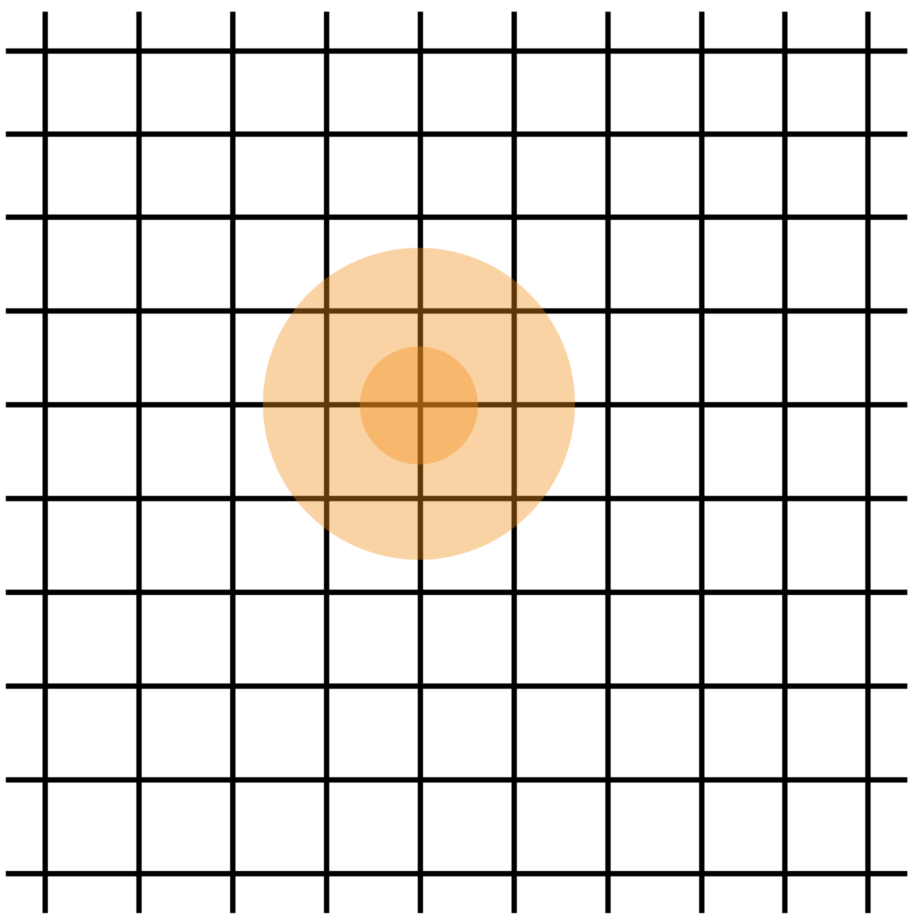
Number of updates to change fixed physical length scale → ∞

"Critical slowing-down" of generation of uncorrelated samples

Generate QCD gauge fields

QCD gauge field configurations sampled via

Hamiltonian dynamics + Markov Chain Monte Carlo



Updates diffusive

Lattice spacing → 0

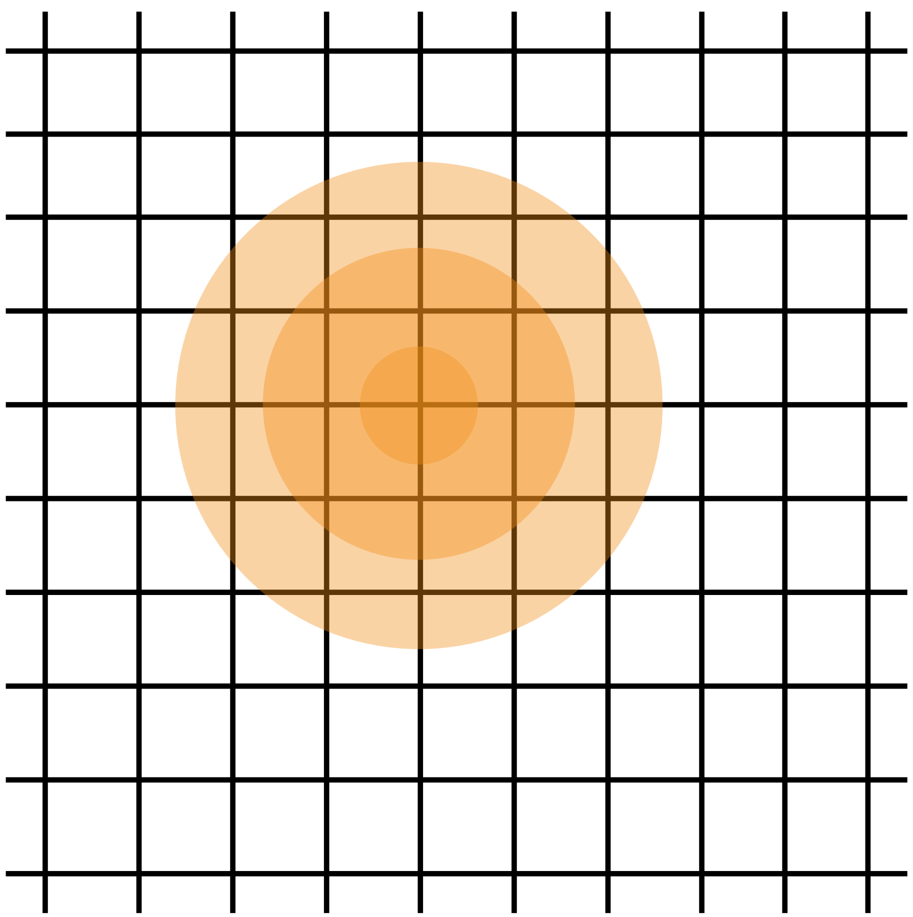
Number of updates to change fixed physical length scale → ∞

"Critical slowing-down" of generation of uncorrelated samples

Generate QCD gauge fields

QCD gauge field configurations sampled via

Hamiltonian dynamics + Markov Chain Monte Carlo



Updates diffusive

Lattice spacing → 0

Number of updates to change fixed physical length scale → ∞

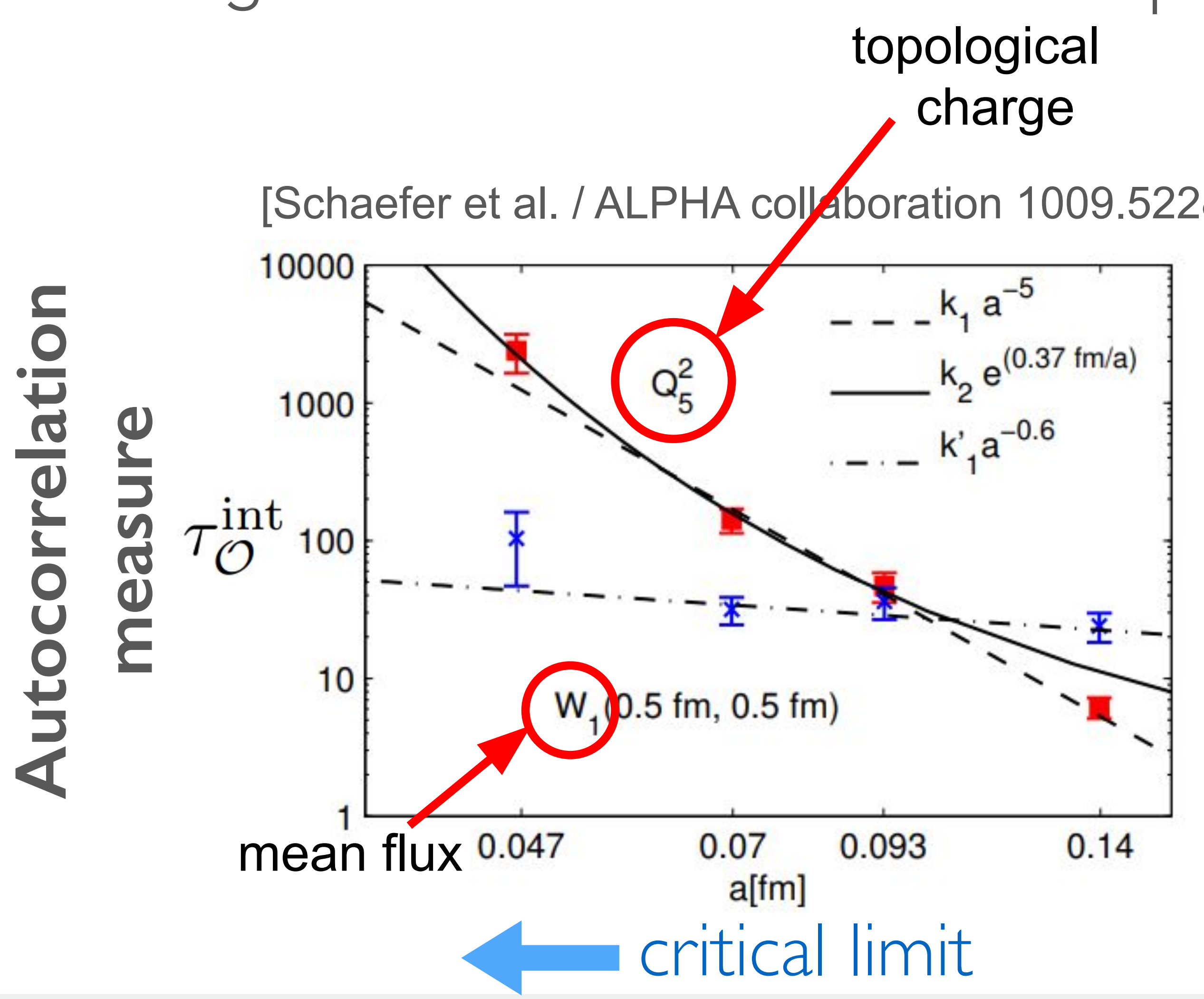
"Critical slowing-down" of generation of uncorrelated samples

Generate QCD gauge fields

QCD gauge field configurations sampled via

Hamiltonian dynamics + Markov Chain Monte Carlo

"Critical slowing-down"
of generation of uncorrelated samples



Machine learning for LQCD

Worldwide 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)
Luo, Clark Stokes, 2012.05232 (2020)
Favoni et al, 2012.12901 (2021)
Luo et al, 2101.07243 (2021)

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

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

F... ...tion by e.g.,

...D 97 (2018)

Tanaka and Tomiya, 1712.03893 (2017)
Zhou et al., Phys.Rev.D 100 (2019)
Li et al., PRX 10 (2020)
...ski and Urban 1811.03533 (2020)
...miya 2010.11900 (2020)
...232 (2020)

Approaches must rigorously preserve quantum field theory in applicable limits

➡ AB-INITIO AI

Efficient compu...
functions/observables
Yoon, Bhattacharya, Gupta, Phys. Rev. D 100, 014504 (201...)
Zhang et al, Phys. Rev. D 101, 034516 (2020)
Nicoli et al., 2007.07115 (2020)

S. Blu...
Boyda et al., 20...

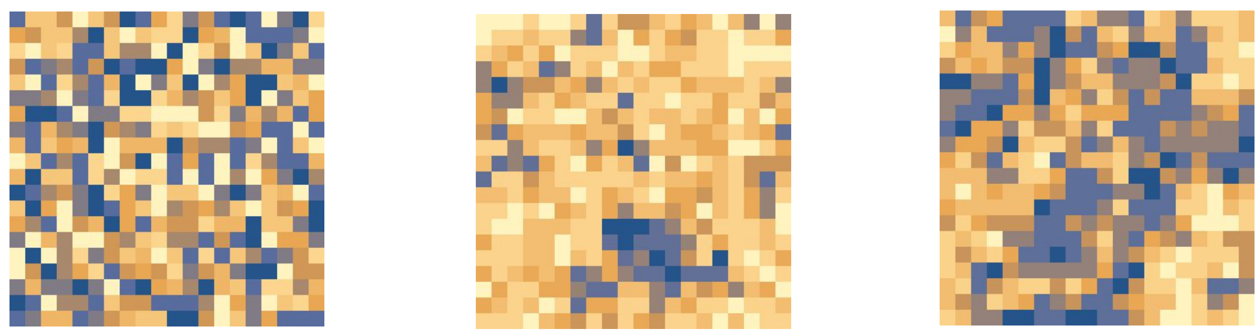
Sign-problem avoidance via contour deformation of path integrals
Alexandruet al., Phys. Rev. Lett. 121 (2020),
Detmold et al., 2003.05914 (2020)

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

Generate QCD gauge fields

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

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

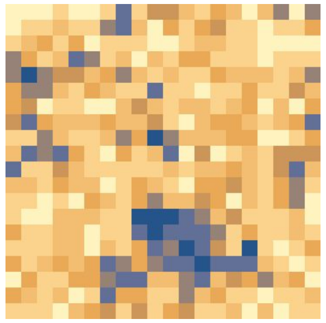
Generate QCD gauge fields

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

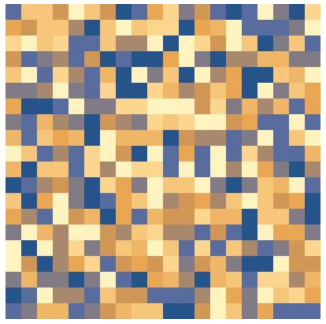
$$P[\phi(x)] \sim e^{-S[\phi(x)]}$$

Parallels with image generation problem

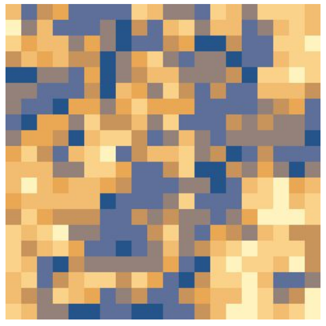
likely
(log prob = 22)



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

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

CIFAR benchmark image set for machine learning

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

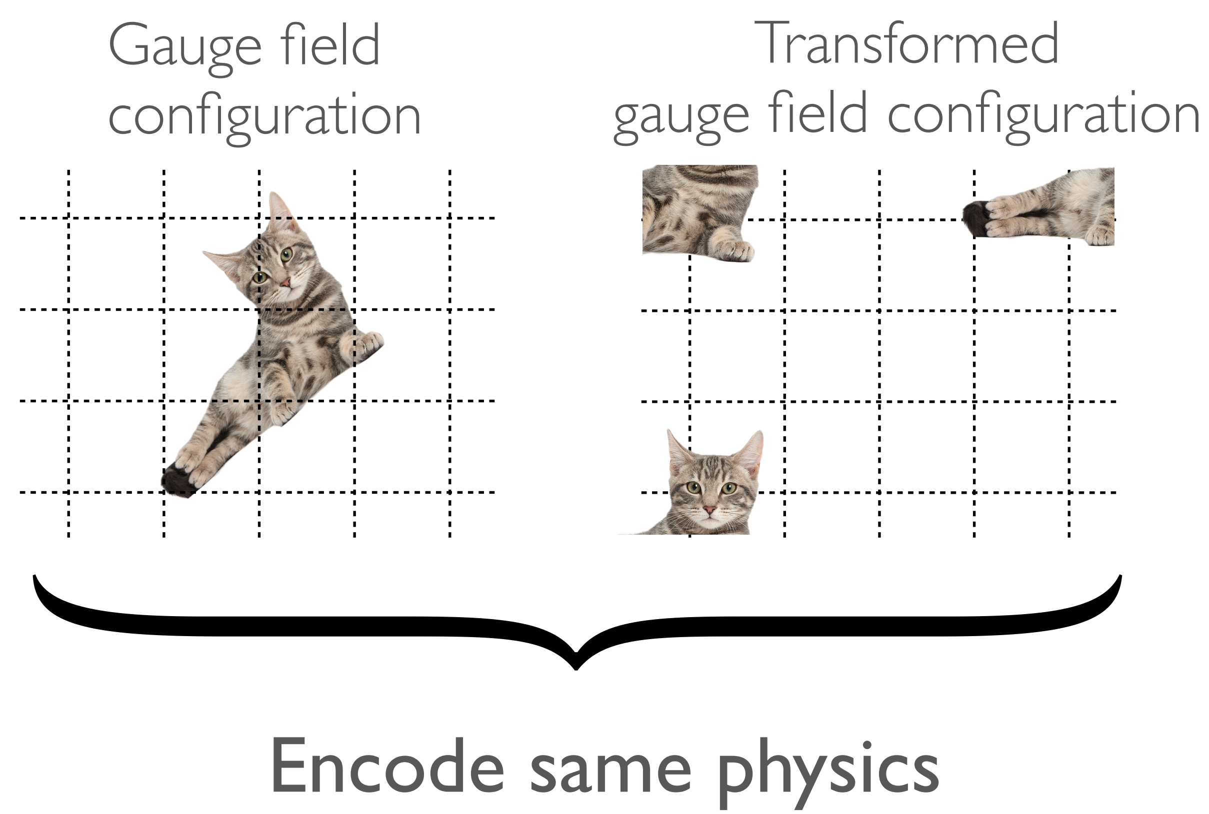
Machine learning QCD

Ensemble of lattice QCD gauge fields

- Ensemble of gauge fields has meaning
- $64^3 \times 128 \times 4 \times N_c^2 \times 2$ $\simeq 10^9$ numbers
- ~1000 samples
- 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**



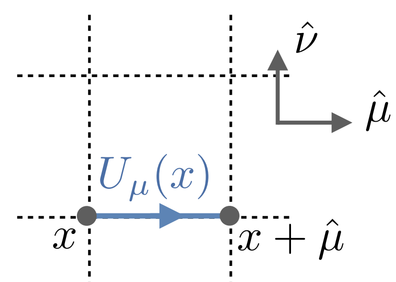
Machine learning QCD

Ensemble of lattice QCD gauge fields

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

Physics is invariant under specific field transformations

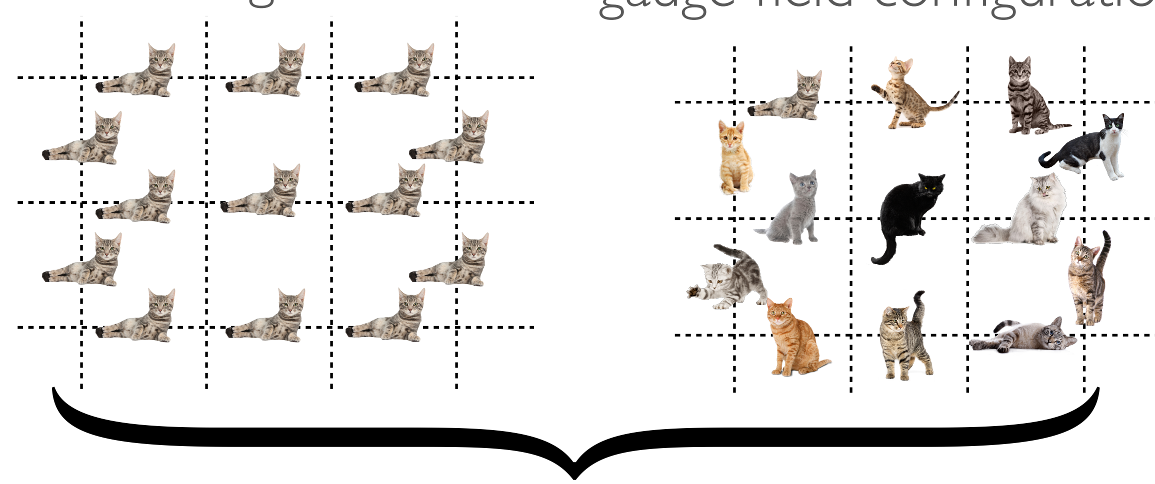
- **Gauge transformation**



$$U_\mu(x) \to \Omega(x) U_\mu(x) \Omega^\dagger(x+\hat{\mu})$$
$$\text{for all } \Omega(x) \in \mathrm{SU}(3)$$

Gauge field configuration

Transformed gauge field configuration



Encode same physics

Machine learning QCD

Ensemble of lattice QCD gauge fields

- ... f gauge fields has

- ... important

- Gauge and translation-invariant with periodic boundaries

CIFAR benchmark image set for machine learning

- Each image has meaning

- 32 x 32 pixels x 3 cols ... numbers

- ... frame

Need custom ML for physics from the ground up

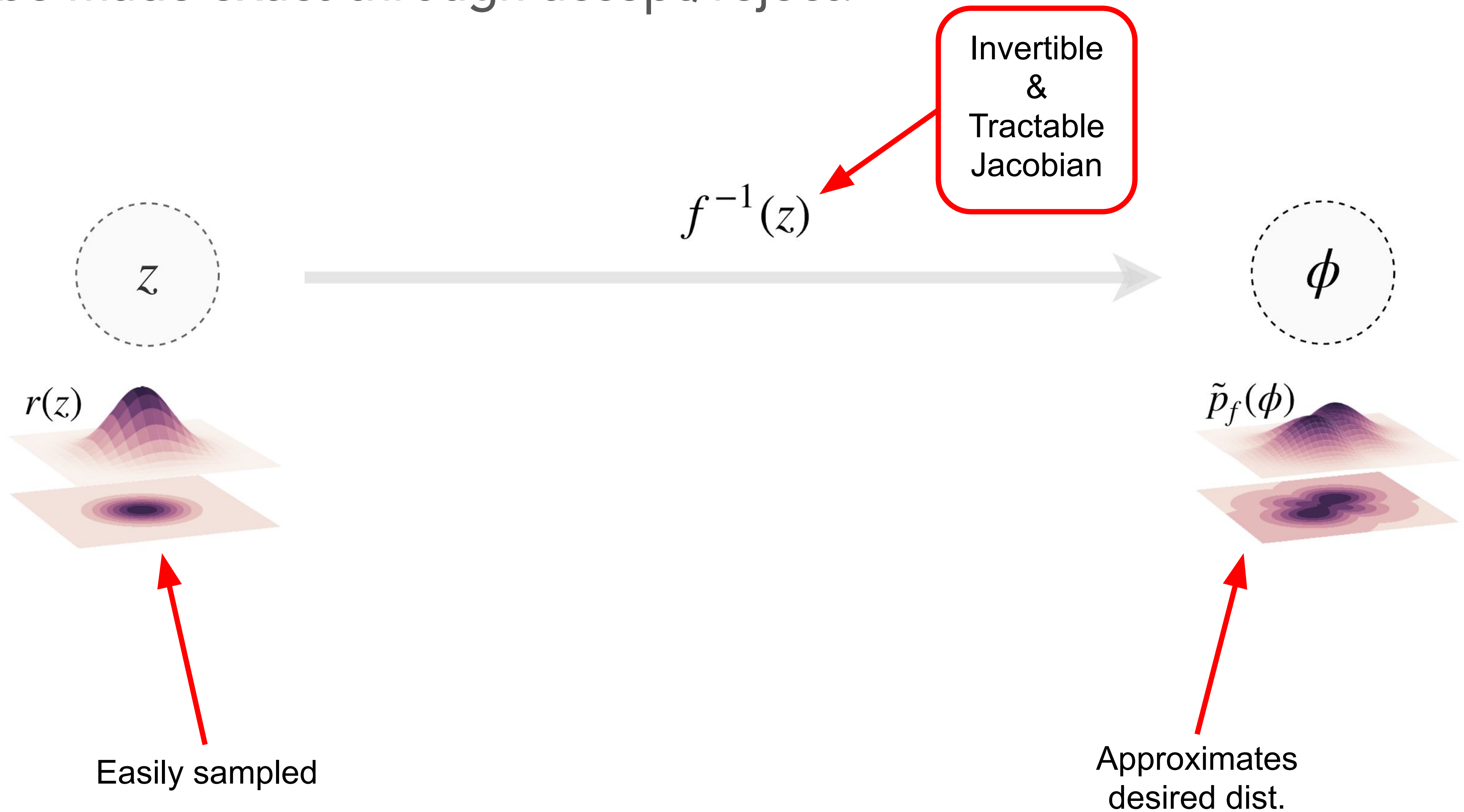
AB-INITIO AI

https://iaifi.org/

Generative flow models

Flow-based models learn a change-of-variables that transforms a known distribution to the desired distribution [Rezende & Mohamed 1505.05770]

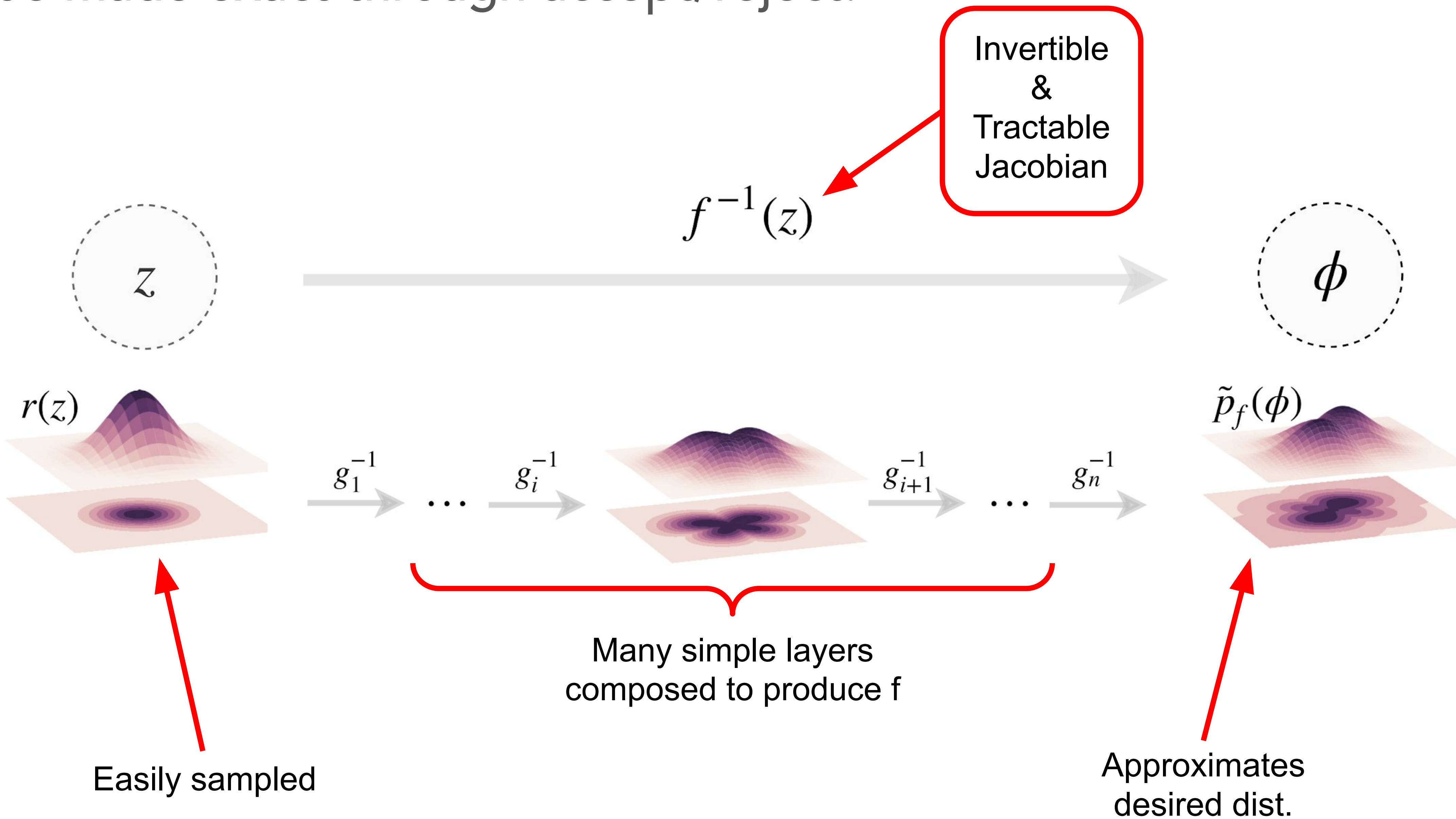
Can be made exact through accept/reject!



Generative flow models

Flow-based models learn a change-of-variables that transforms a known distribution to the desired distribution [Rezende & Mohamed 1505.05770]

Can be made exact through accept/reject!



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

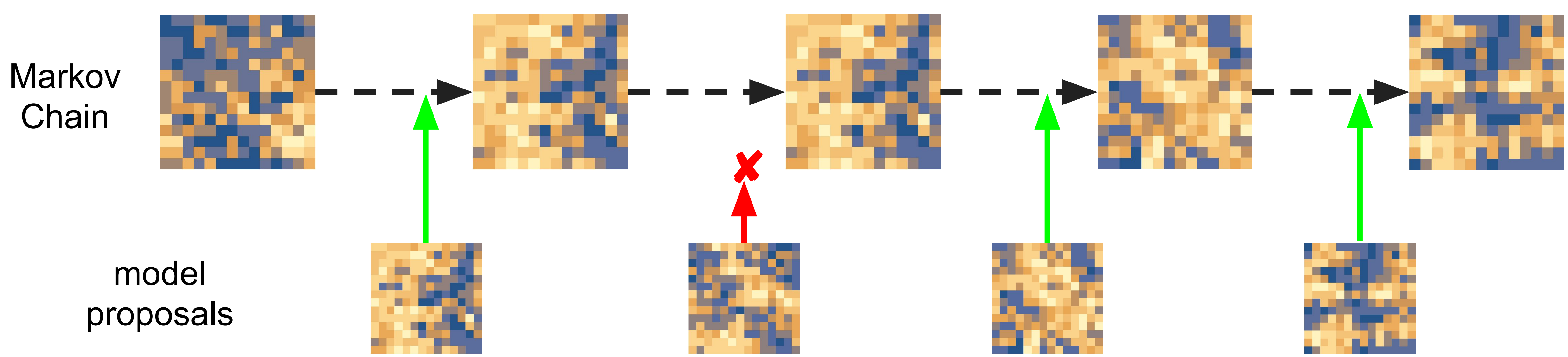
Acceptance probability

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

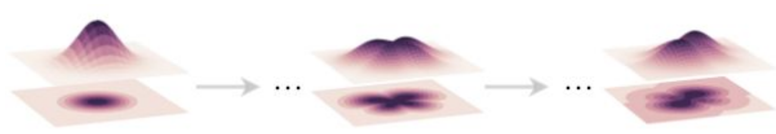
True dist

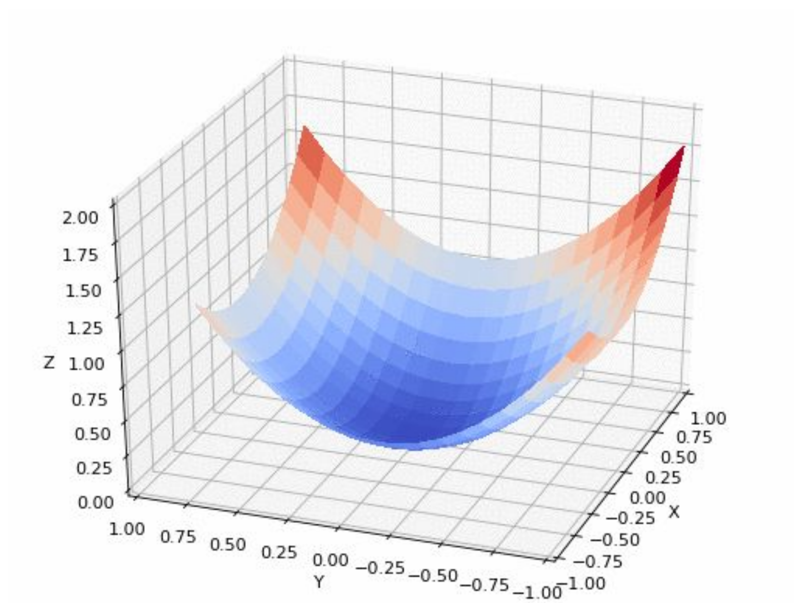
Model dist

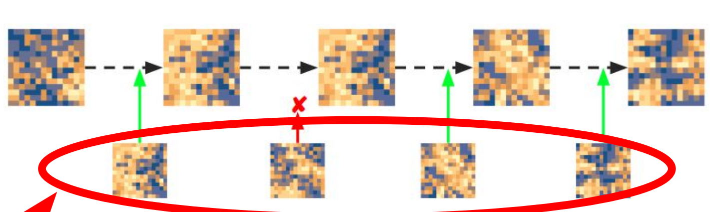
proposal independent of previous sample



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

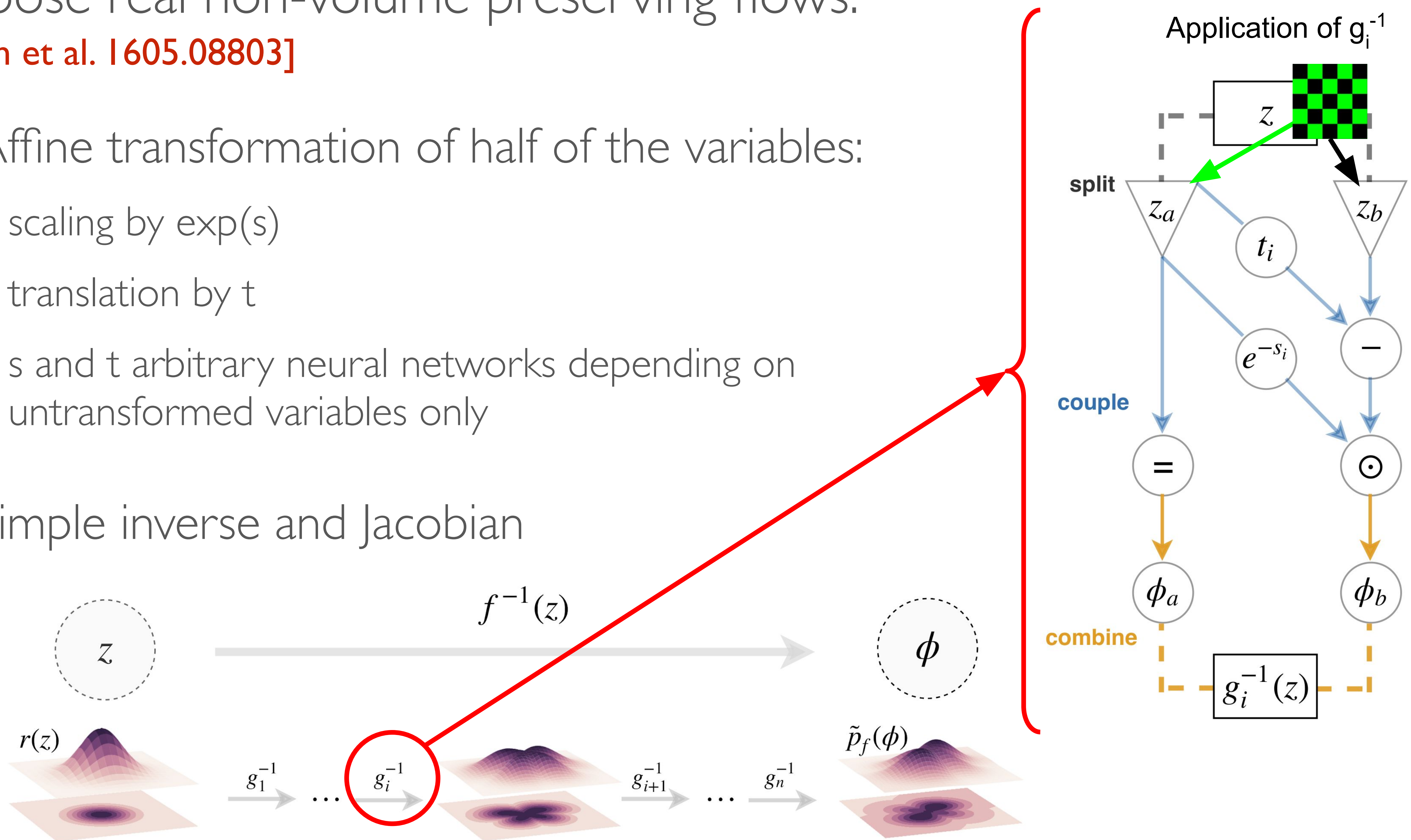
	E1	E2	E3	E4	E5
L	6	8	10	12	14
m^2	-4	-4	-4	-4	-4
λ	6.975	6.008	5.550	5.276	5.113
$m_p L$	3.96(3)	3.97(5)	4.00(4)	3.96(5)	4.03(6)

Application: scalar field theory

First application: scalar lattice field theory

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



Phiala Shanahan, MIT

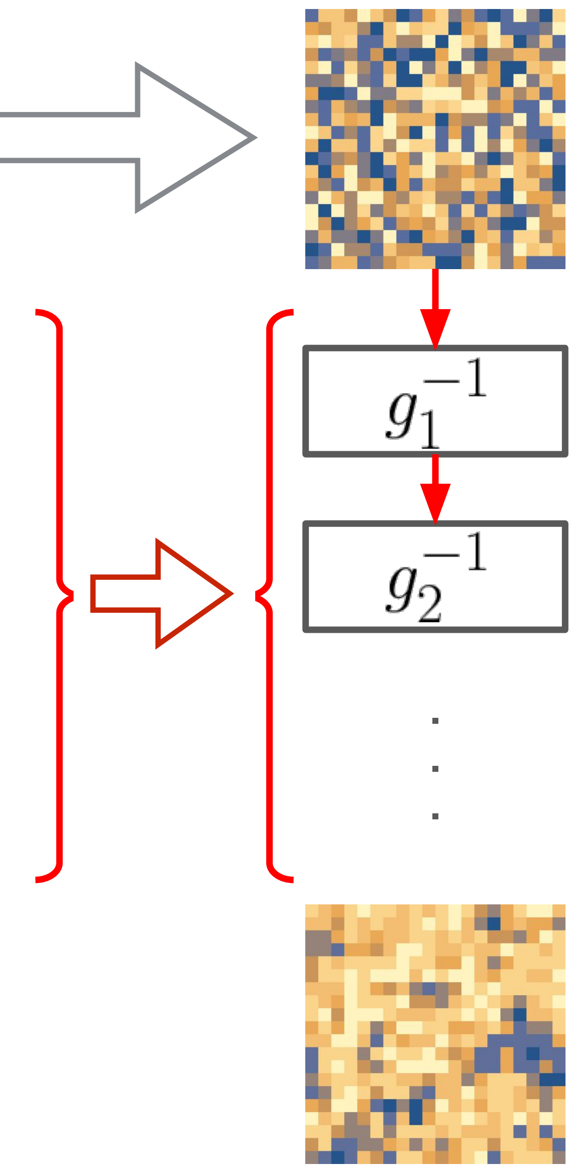
Application: scalar field theory

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



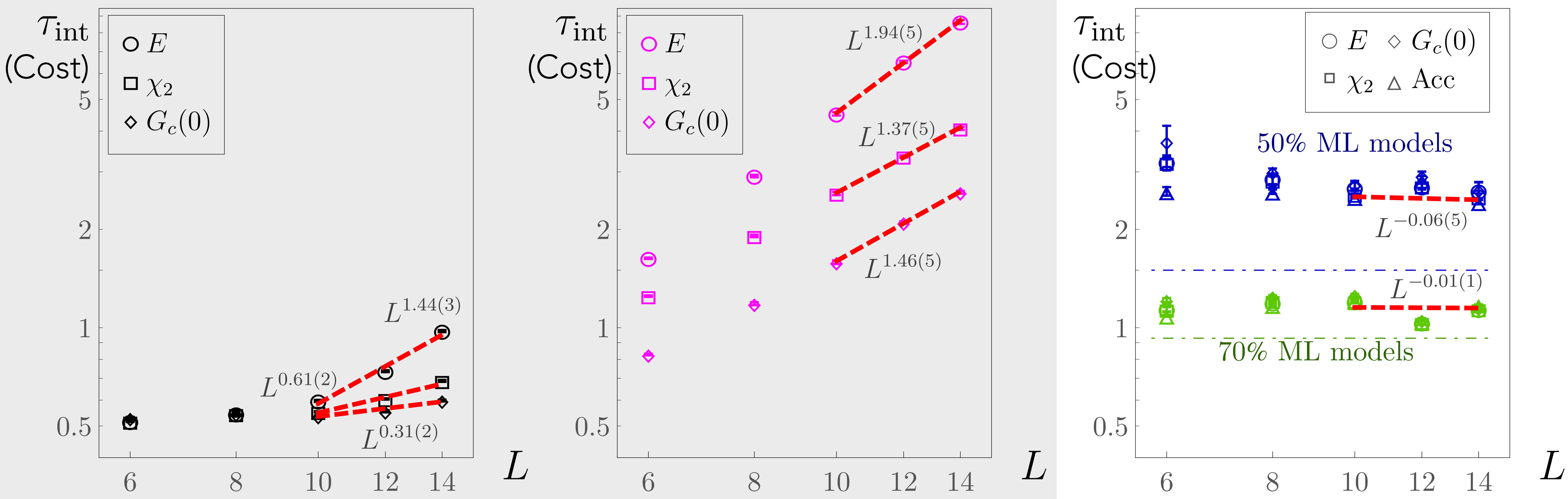
- Train using shifted KL loss with Adam optimizer
  - Stopping criterion: fixed acceptance rate in Metropolis-Hastings MCMC

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

Conventional approaches slow down

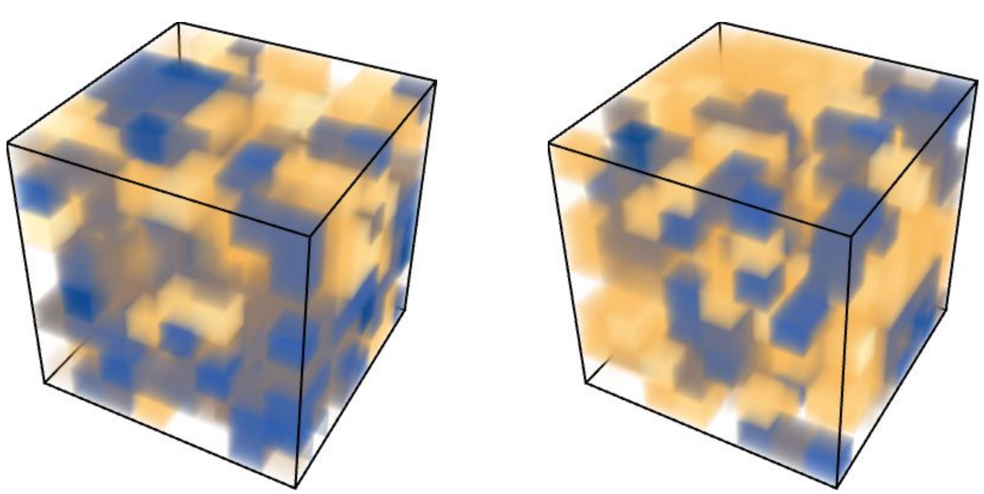
No slow-down

From toy models to QCD

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

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



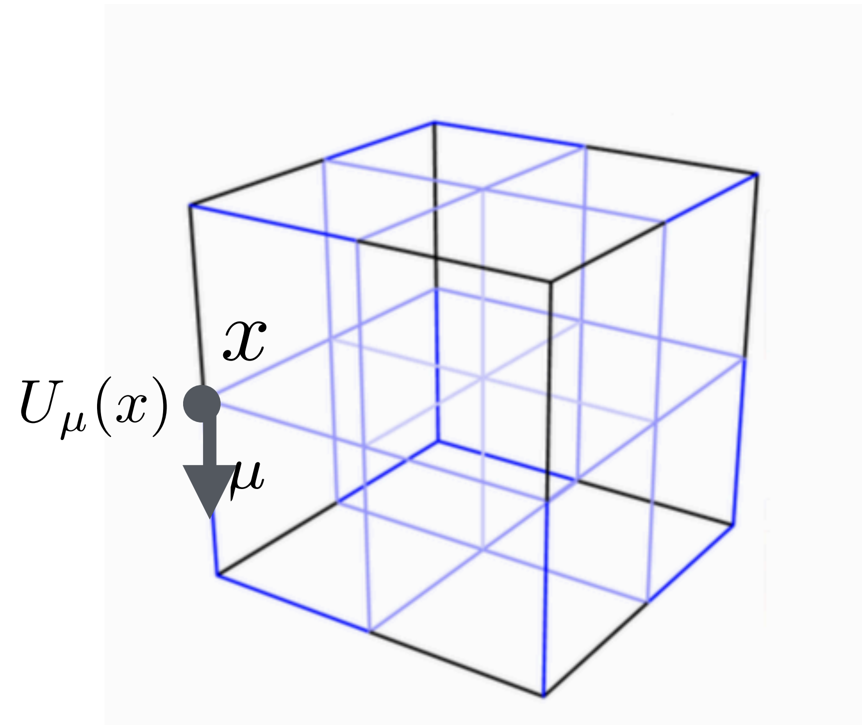
AURORA | EARLY SCIENCE PROGRAM FOR DATA & LEARNIN

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

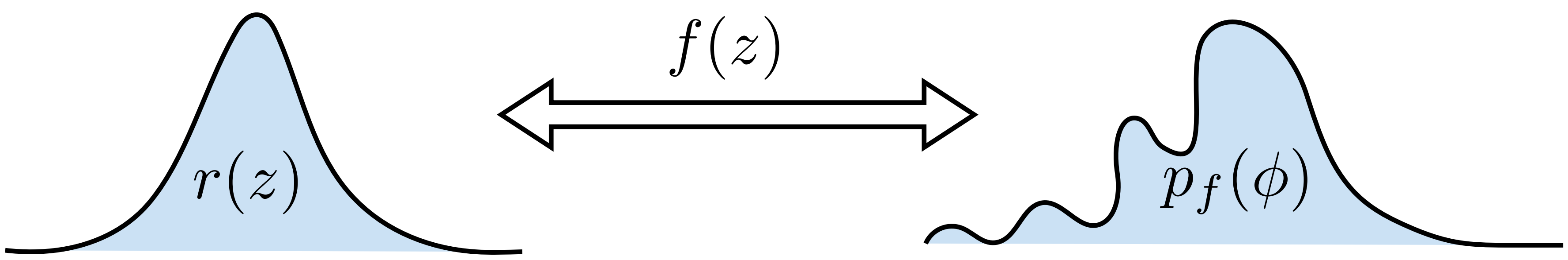


1. Flows on compact, connected manifolds
2. Incorporate symmetries: gauge-equivariant flows

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

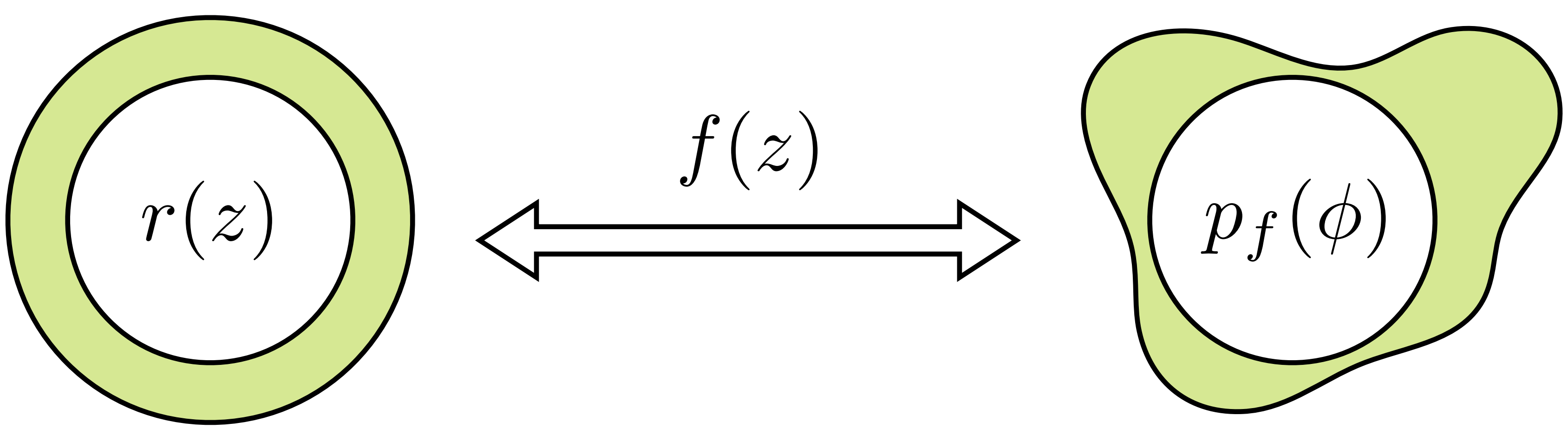
Flows on spheres and tori

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



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

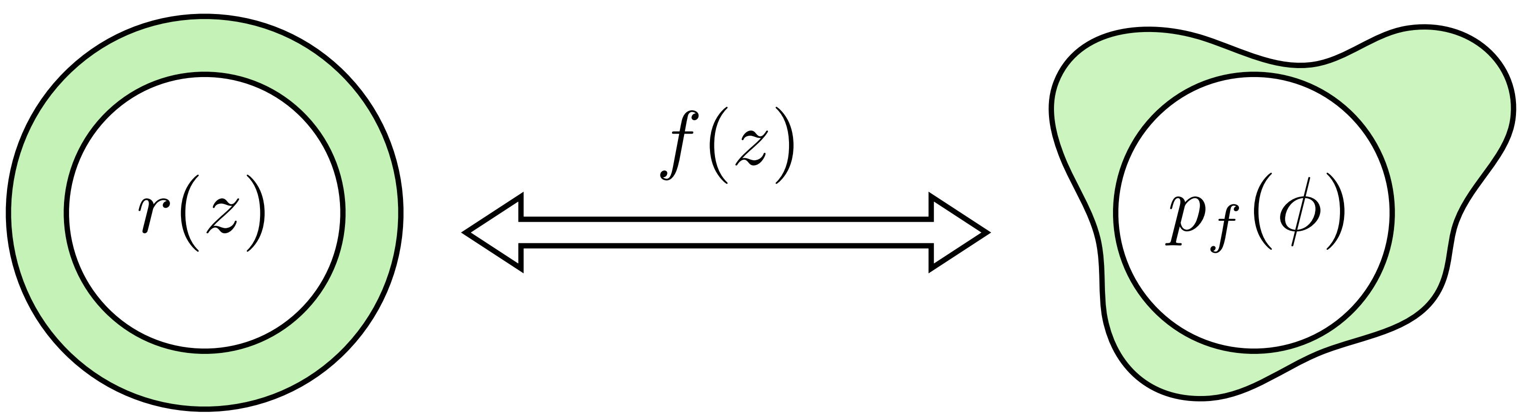
e.g., circles, tori, spheres



Flows on spheres and tori

Test case: Flows on the circle

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



Diffeomorphism if:

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

Ensures transformation is monotonic → invertible

Expressive transformations through:

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

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

Phiala Shanahan, MIT

Flows on spheres and tori

Normalizing Flows on Tori and Spheres

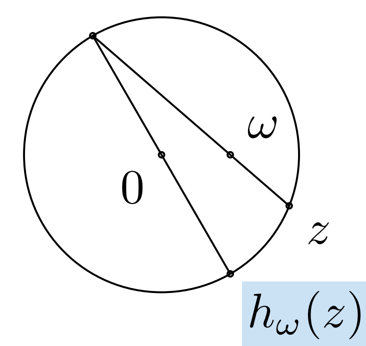
[arXiv:2002.02428]

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**

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

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



- **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

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

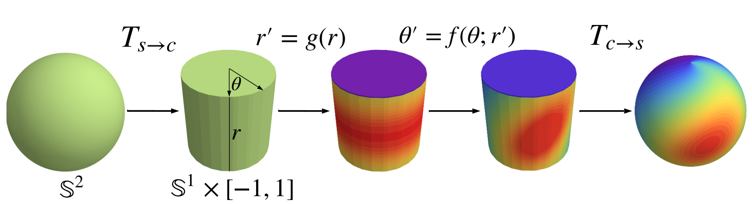
Flows on spheres and tori

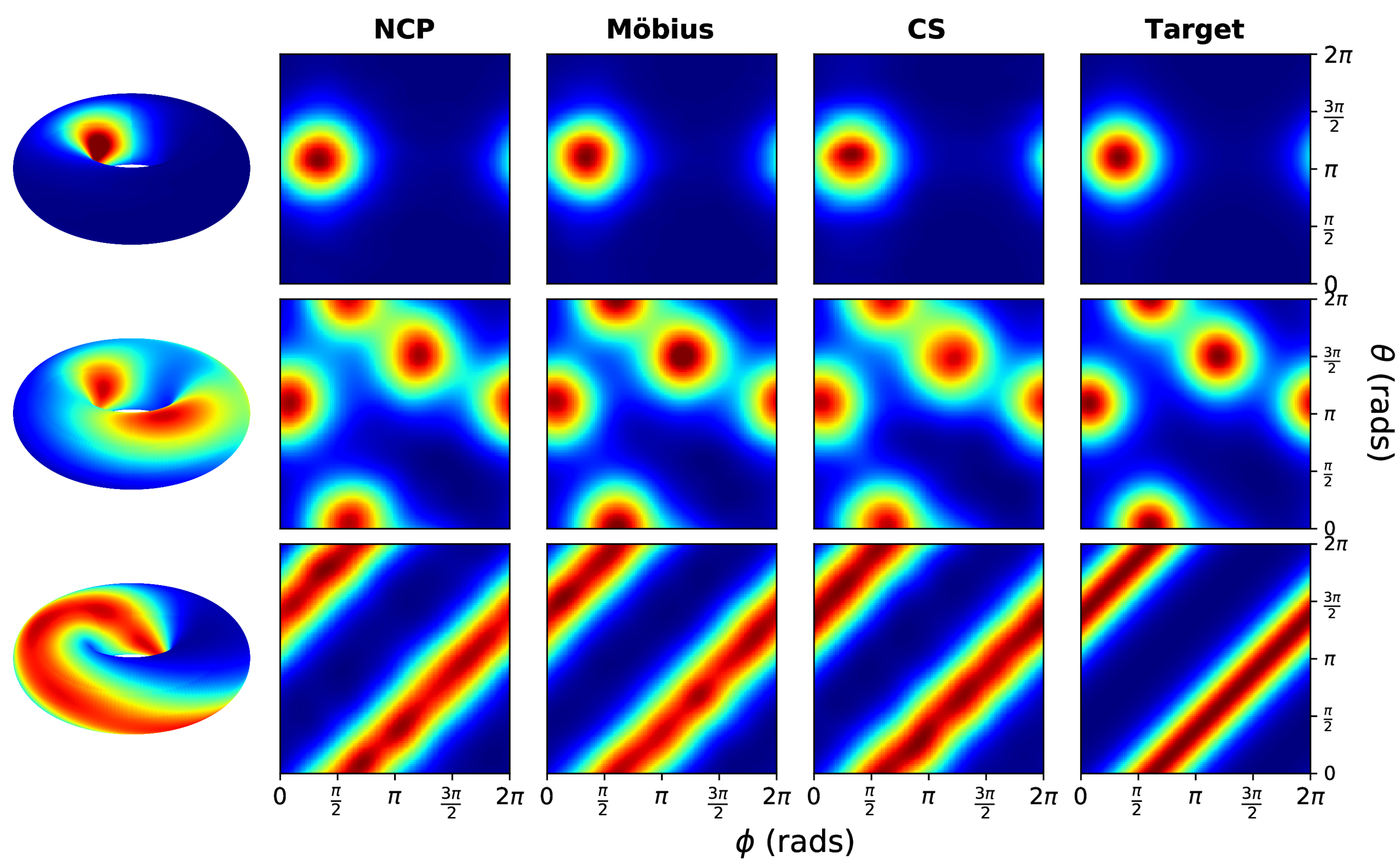
[arXiv:2002.02428]

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

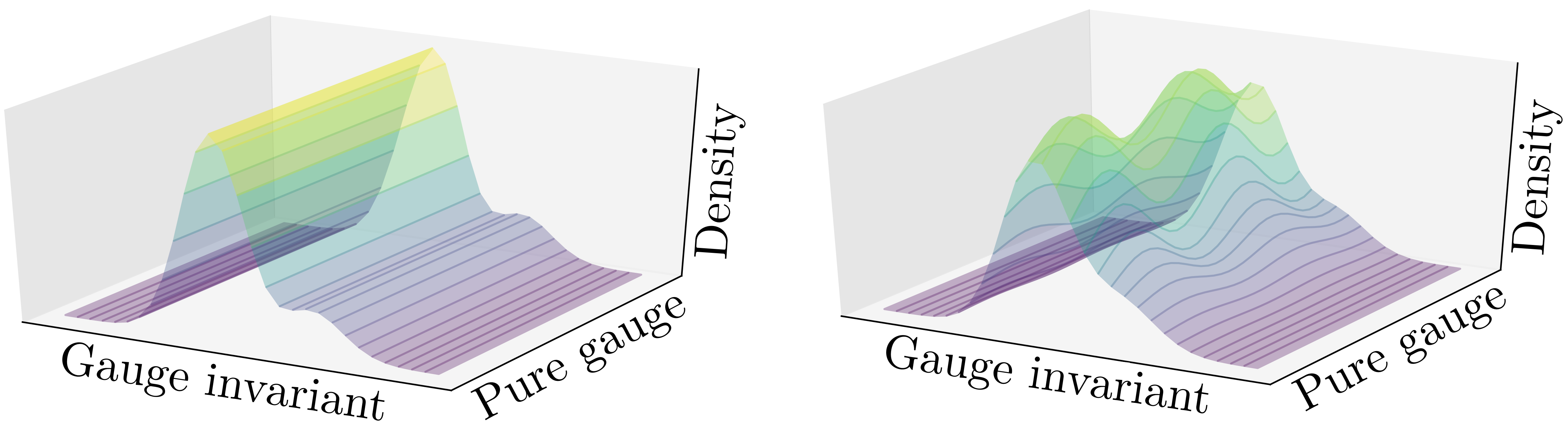




Incorporating symmetries

Incorporating symmetries

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



Incorporating symmetries

Incorporating symmetries

- Not essential for correctness of ML-generated ensembles
- BUT: Crucially 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

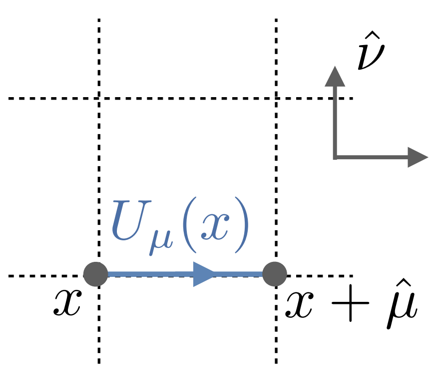
Gauge field theory

First gauge theory application: U(1) field theory

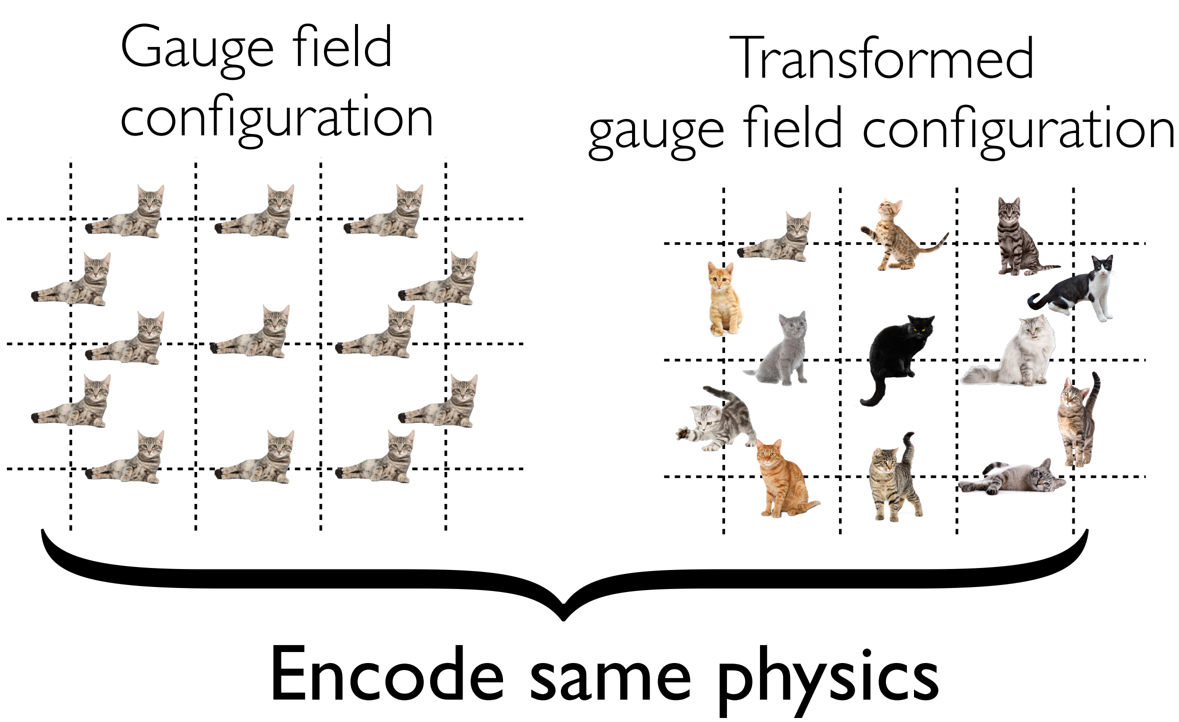
Generative flow architecture that is *gauge-equivariant*

Gauge transformation

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



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



Phiala Shanahan, MIT

Gauge-equivariant flows

First gauge theory application: U(1) field theory

Generative flow architecture that is *gauge-equivariant*

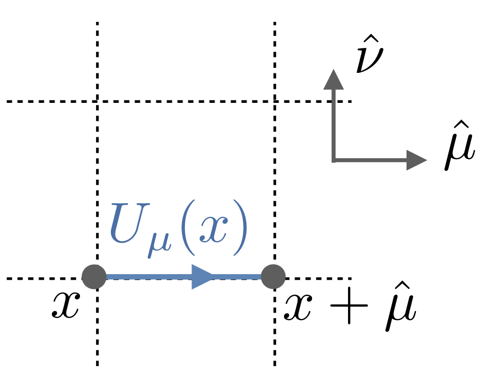
Define invertible, equivariant coupling layer

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

Spacetime dimension

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., PRL 125, 121601 (2020)]

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. → I^i

Loop that starts and ends at same point → $U^i S^i$

Coupling layer equivariant under the condition

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

[Kanwar et al., PRL 125, 121601 (2020)]

Gauge-equivariant flows

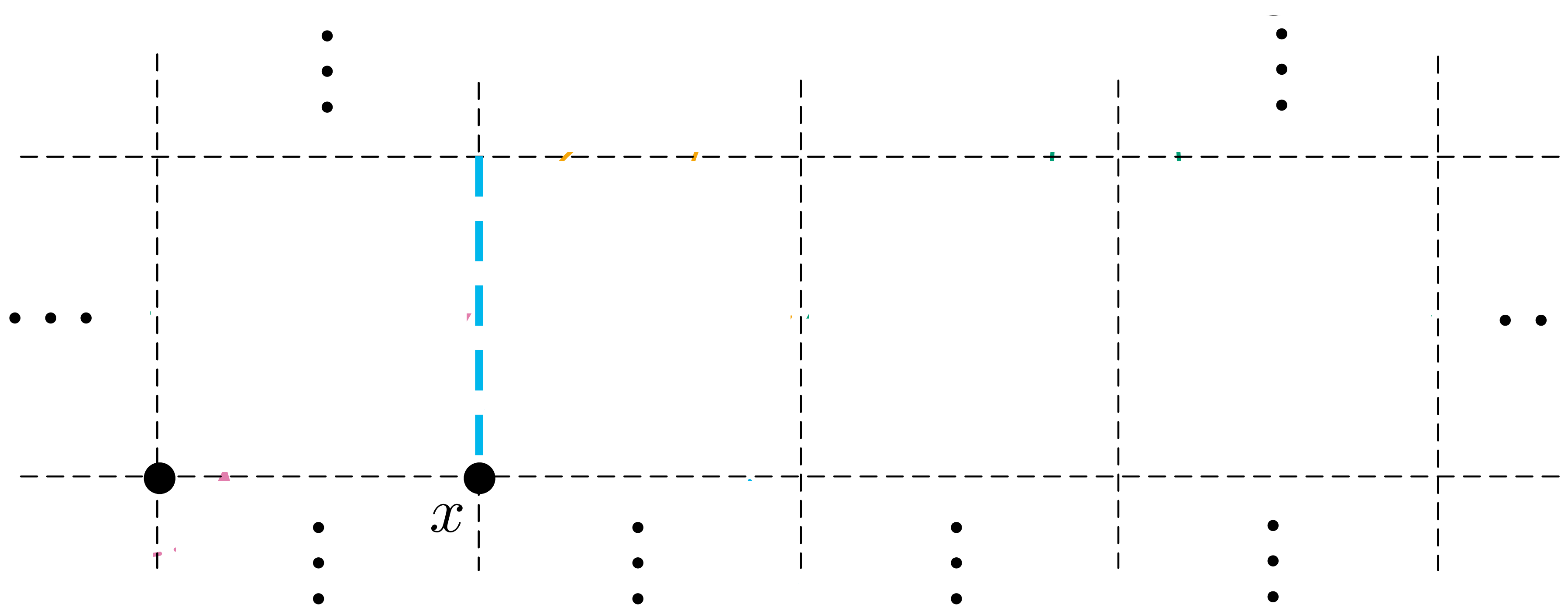
First gauge theory application: U(1) field theory

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.

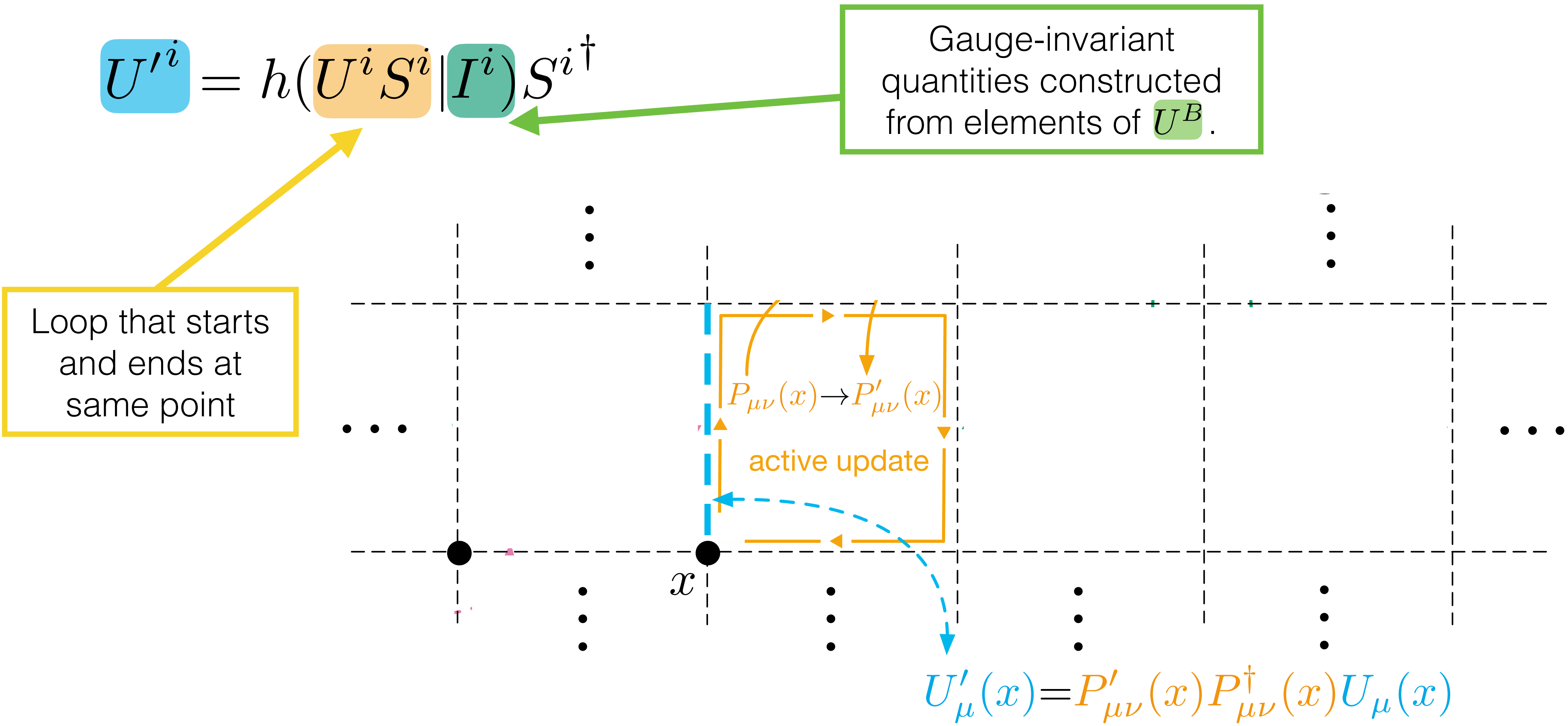
Loop that starts and ends at same point



Gauge-equivariant flows

First gauge theory application: U(1) field theory

Generative flow architecture that is *gauge-equivariant*



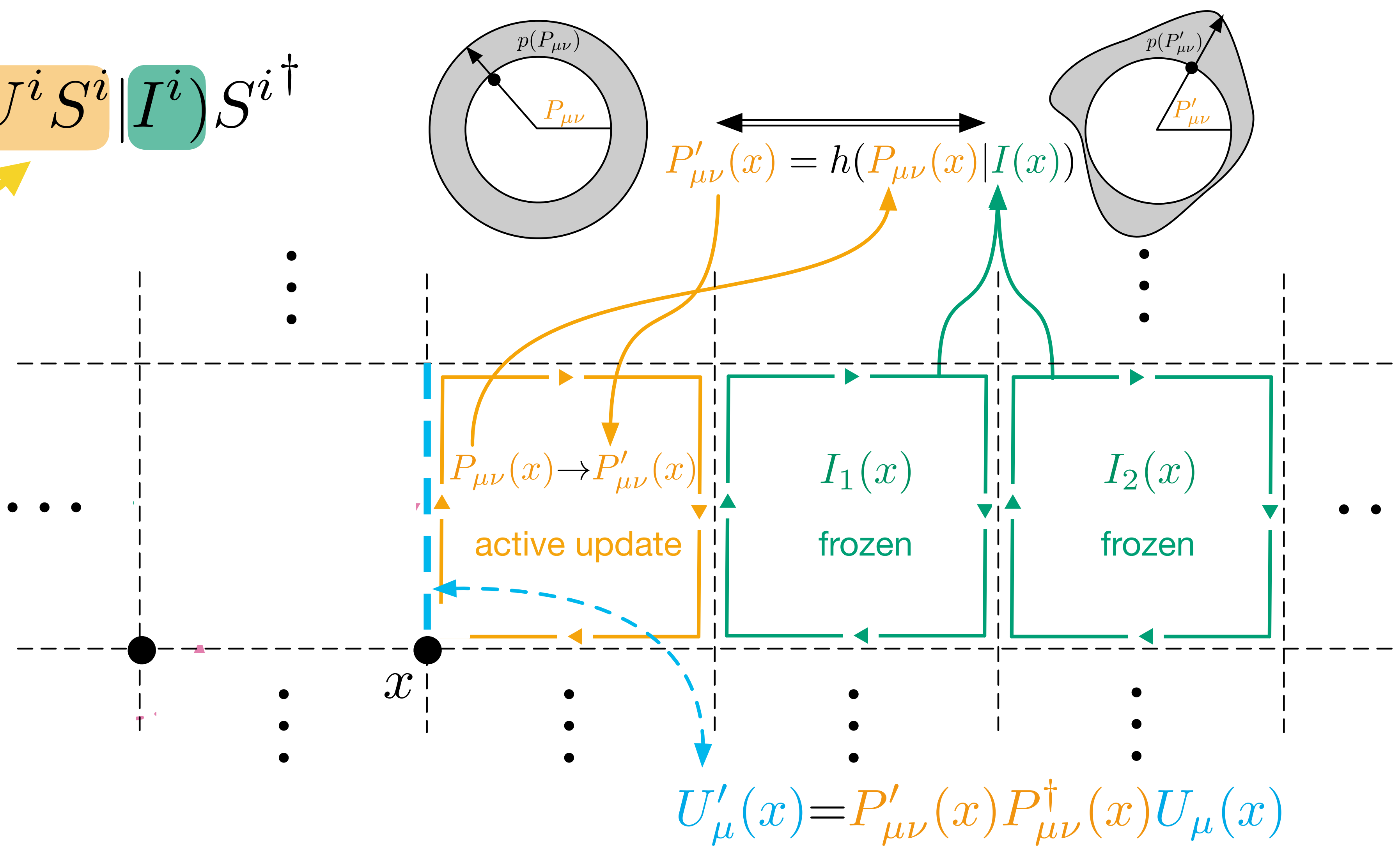
Gauge-equivariant flows

First gauge theory application: U(1) field theory

Generative flow architecture that is *gauge-equivariant*

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

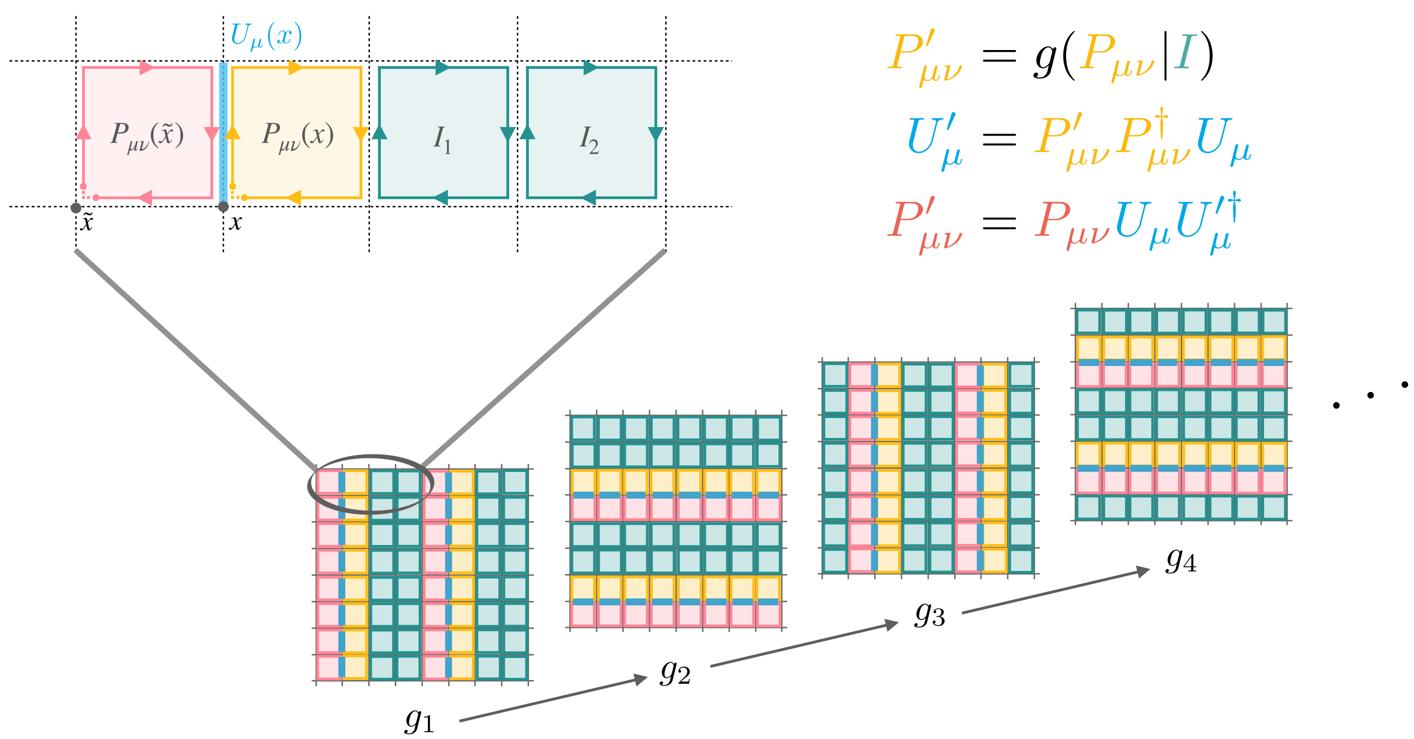
Loop that starts and ends at same point

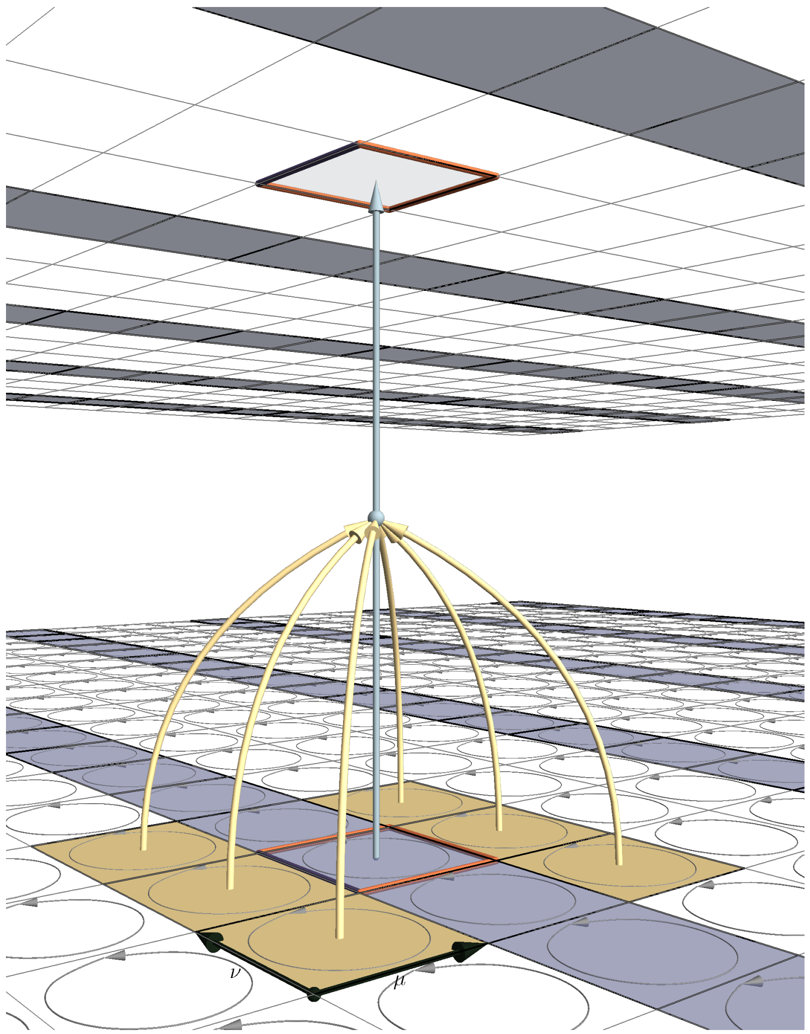


Gauge-equivariant flows

First gauge theory application: U(1) field theory

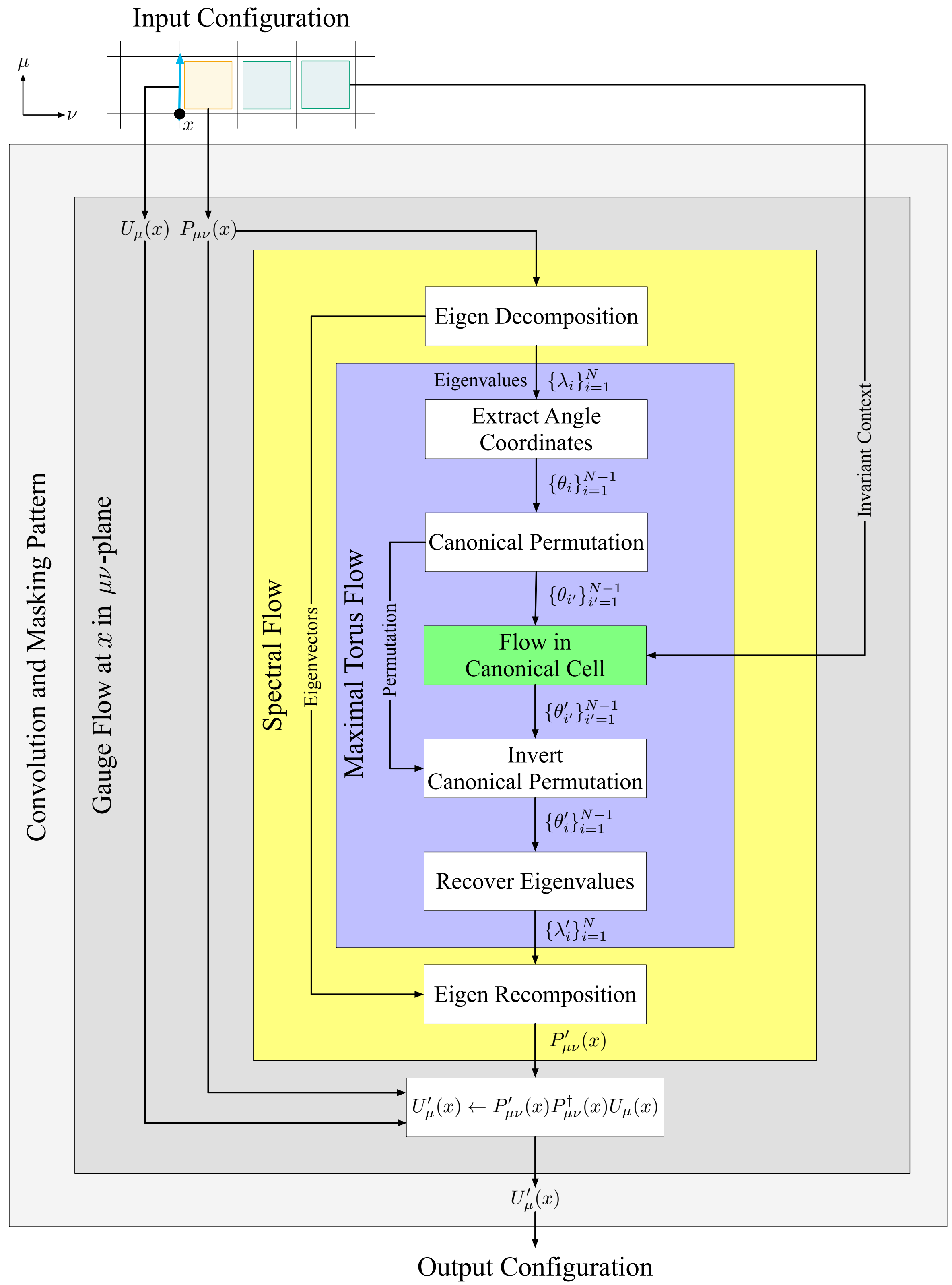
Generative flow architecture that is *gauge-equivariant*

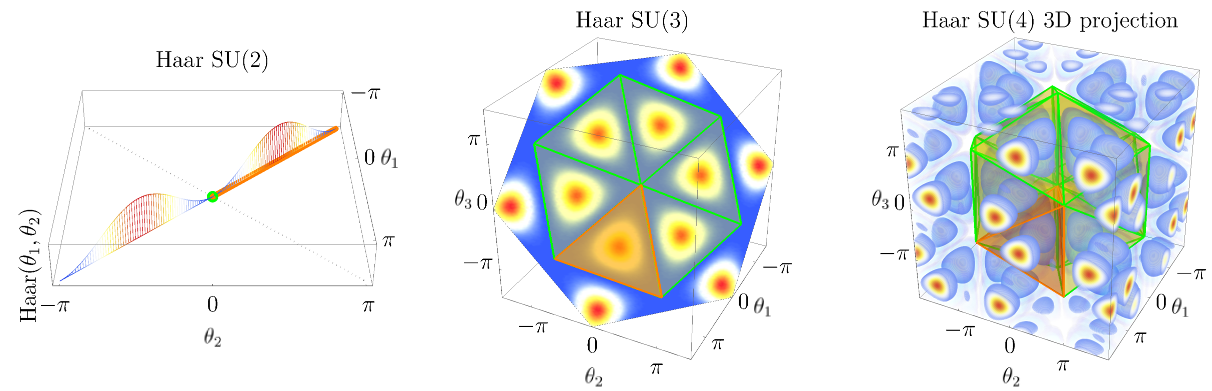




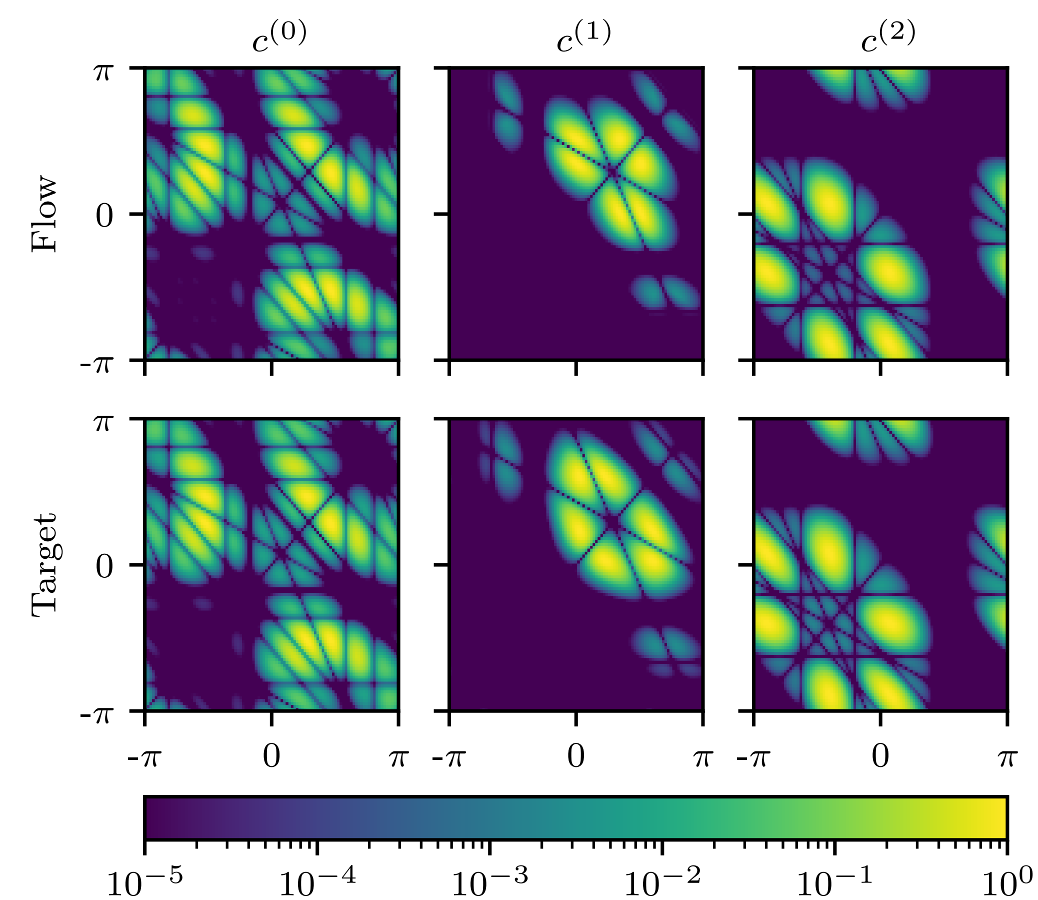
Other recent related work:
Luo, Clark Stokes, 2012.05232 (2020)
Favoni et al, 2012.12901 (2021)
Luo et al, 2101.07243 (2021)

Gauge-equivariant flows





SU(9) flows



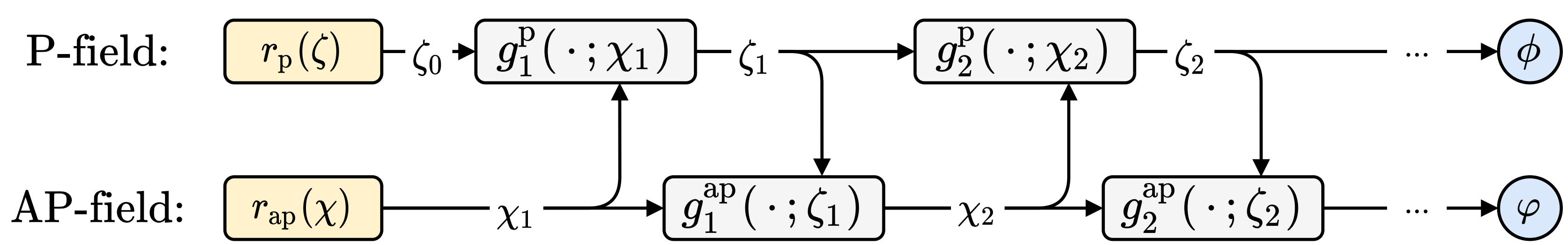
Fermions

QCD action includes Grassman-valued (anticommuting) fermion fields

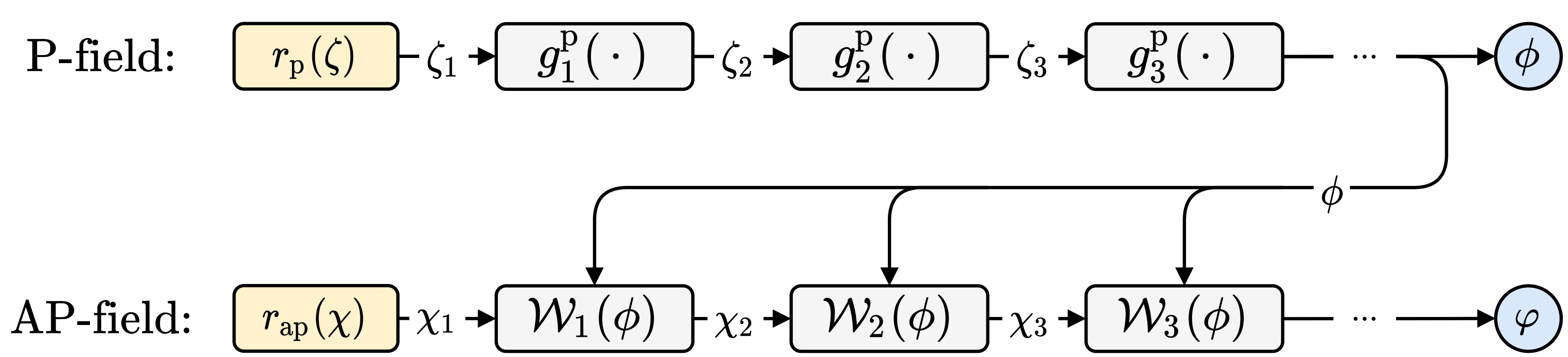
- Analytic integration leaves S[ϕ] expensive to compute
- Conventionally auxiliary "pseudofermion" fields are introduced and marginalised over
- Target density invariant under translations of the fields with anti-periodic boundary conditions on the pseudofermion field
- Design, translation-equivariant convolutions, equivariant linear operators

[arXiv:2106.05934]

Joint architecture based on coupling layers



Autoregressive model based on coupling layers and linear flows



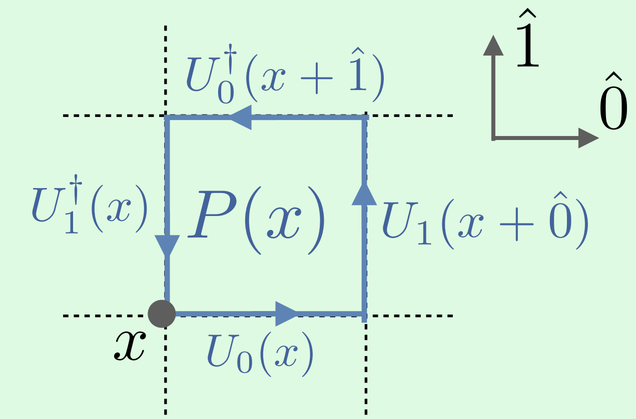
Application: U(1) field theory

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

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

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



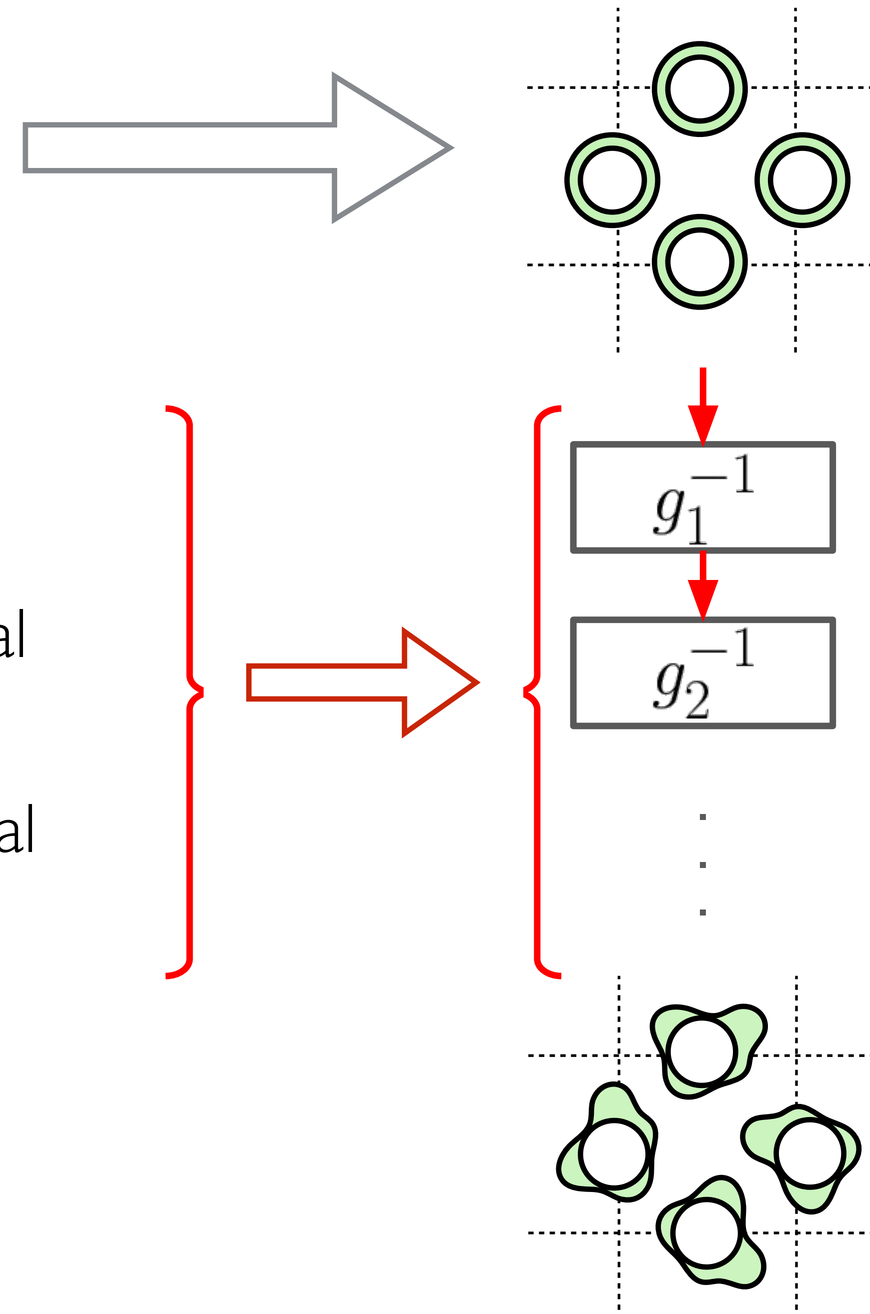
- Fixed lattice size: $L^2 = 16$ with couplings $\beta = \{1, 2, 3, 4, 5, 6, 7\}$
- Continuum limit (critical slow-down) as $\beta \to \infty$

[Kanwar et al., PRL 125, 121601 (2020)]

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



Application: U(1) field theory

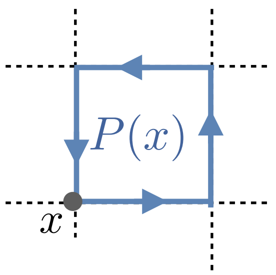
First gauge theory application: U(1) field theory

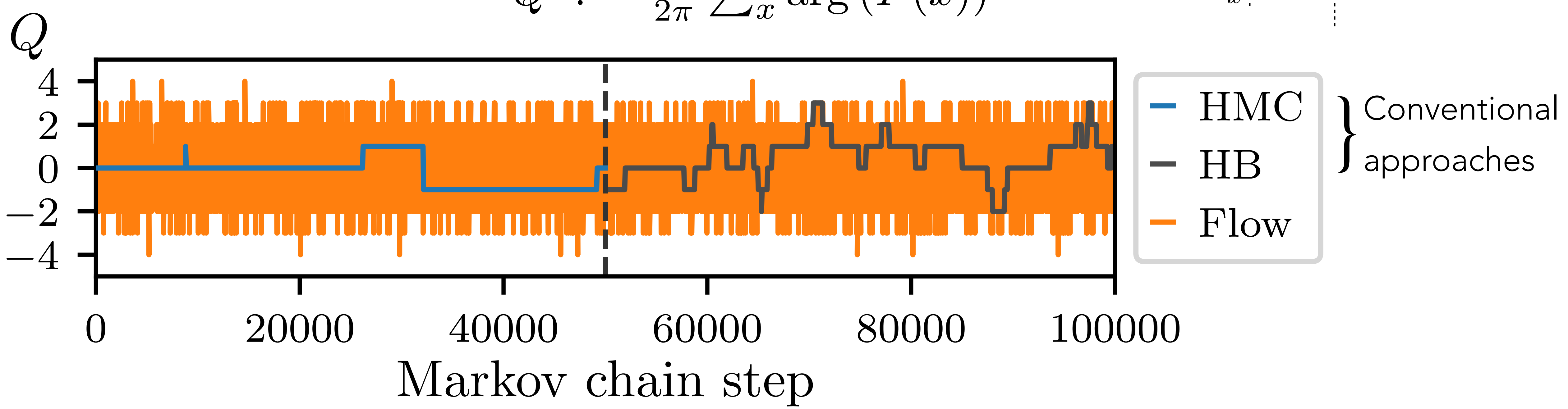
Success: Critical slowing down is significantly reduced

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)]

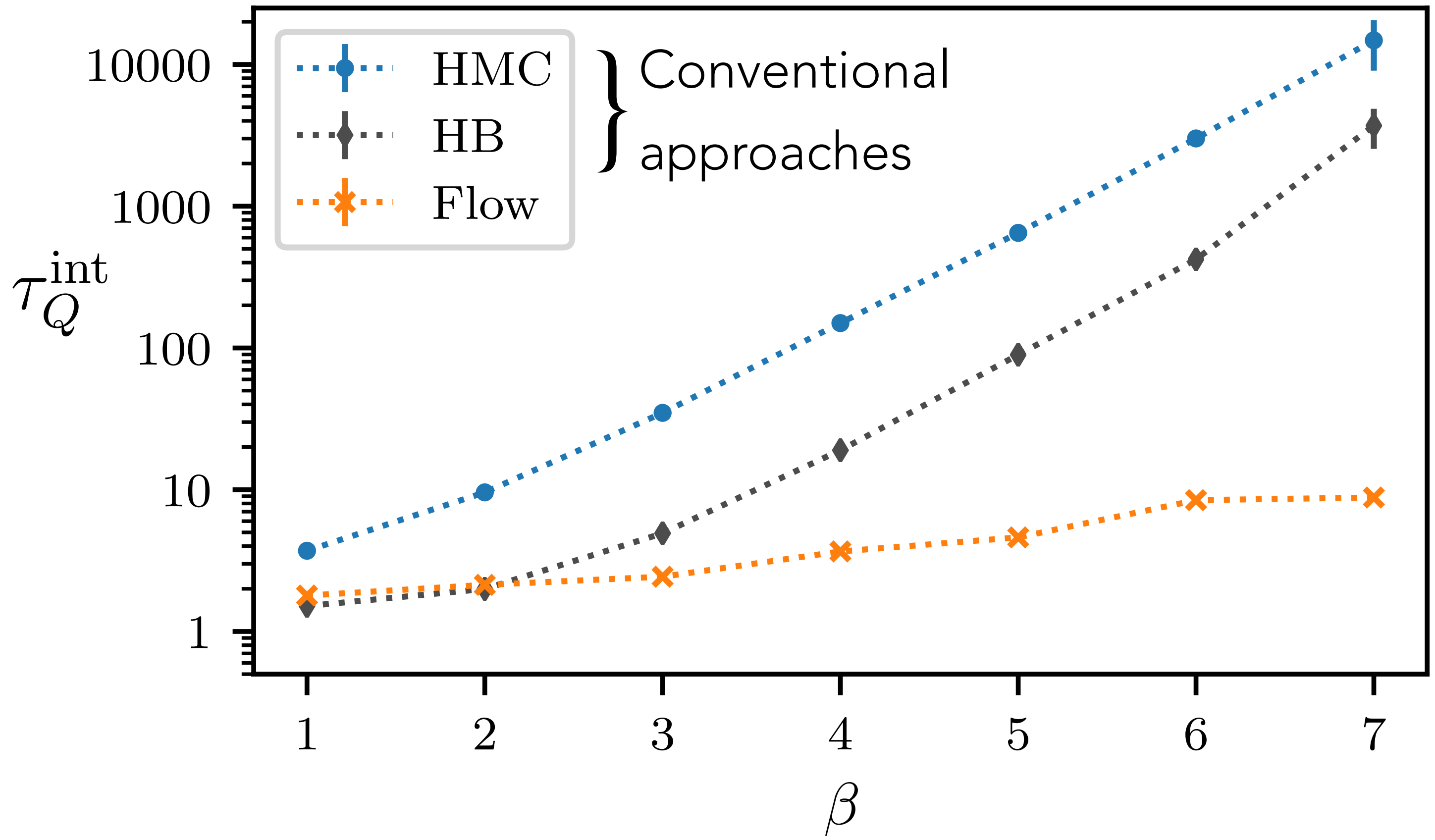
Application: U(1) field theory

First gauge theory application: U(1) field theory

Success: Critical slowing down is significantly reduced

Cost: Up-front training of the model

Cost per independent sample



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

Succe ... ed

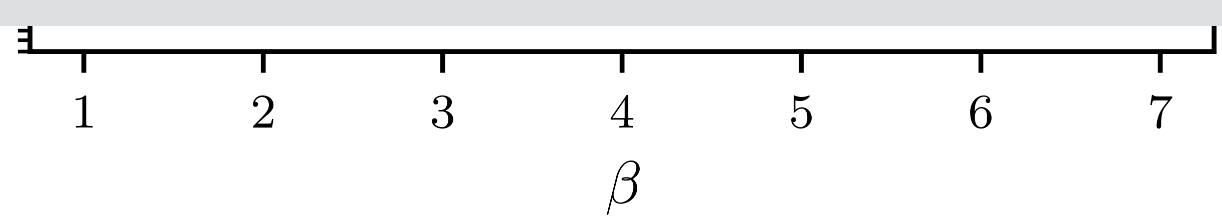
Cost:

SUCCESS!

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

Jupyter notebook tutorial: arXiv:2101.08176

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

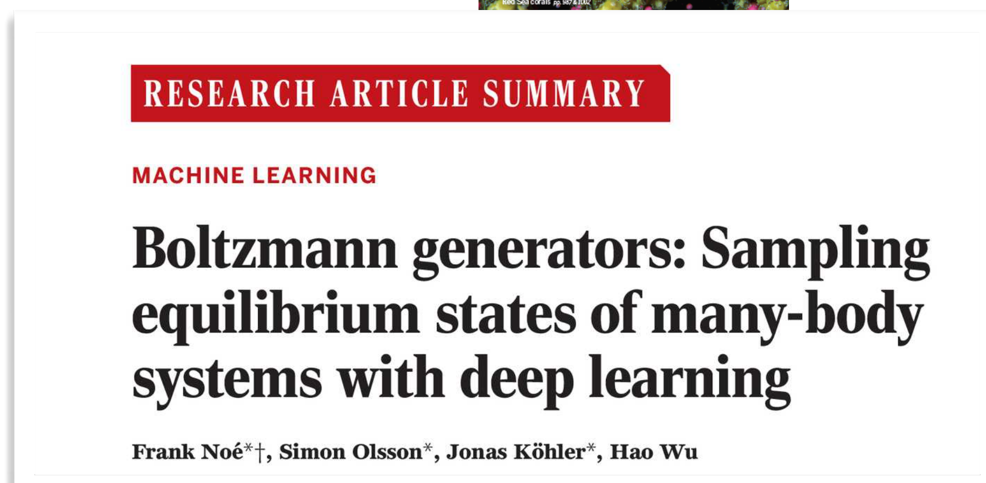


.05456 (2020), PRL 125, 121601 (2020), 2002.02428 (2020)]

2D, L=16

Interdisciplinary relevance

Molecular genetics and drug design

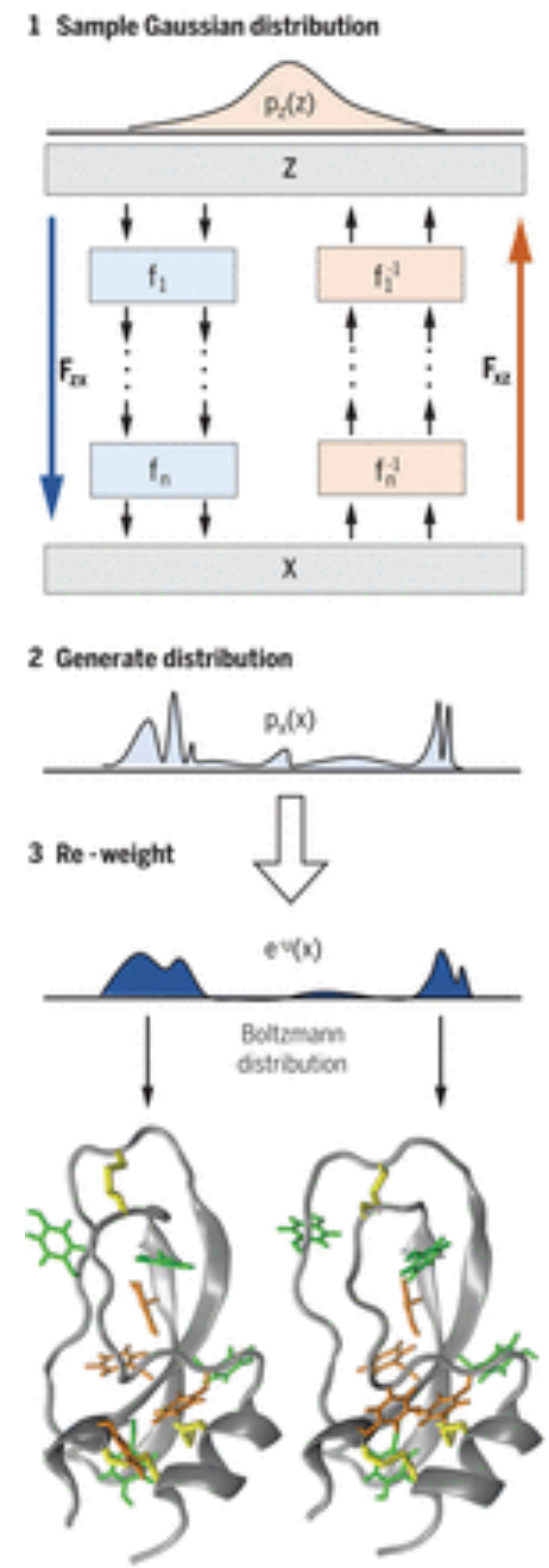


RESEARCH ARTICLE SUMMARY

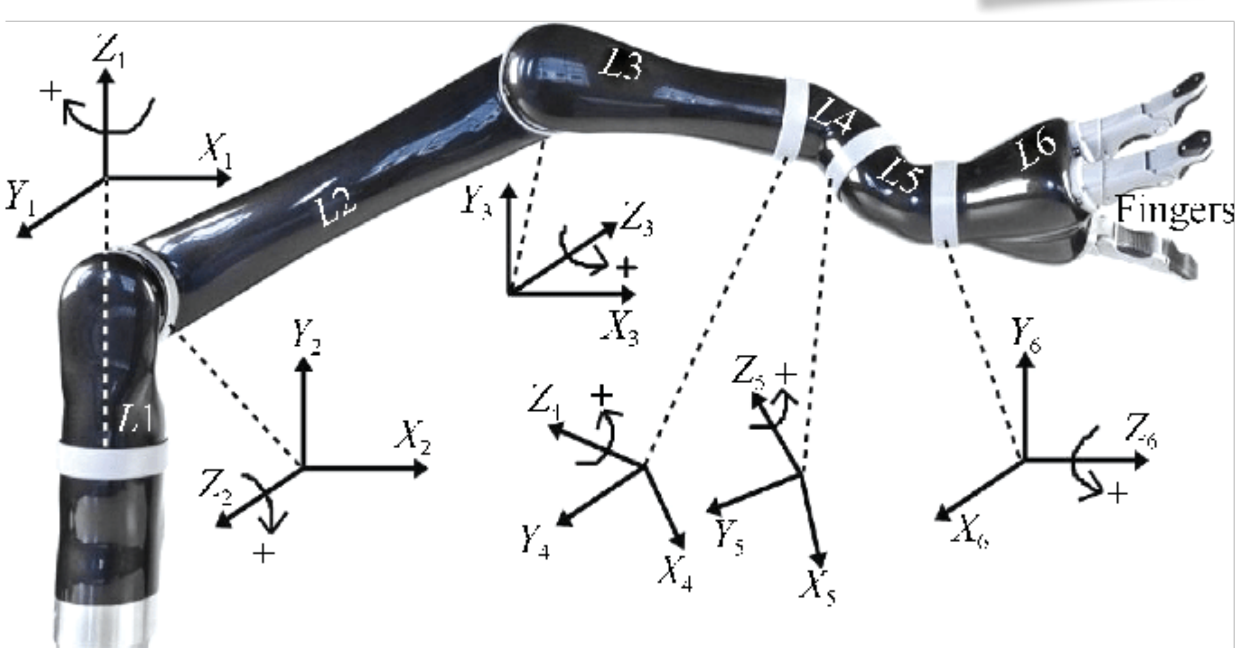
MACHINE LEARNING

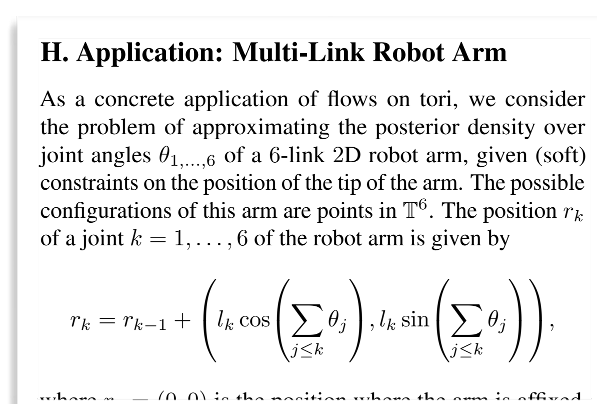
Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning

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

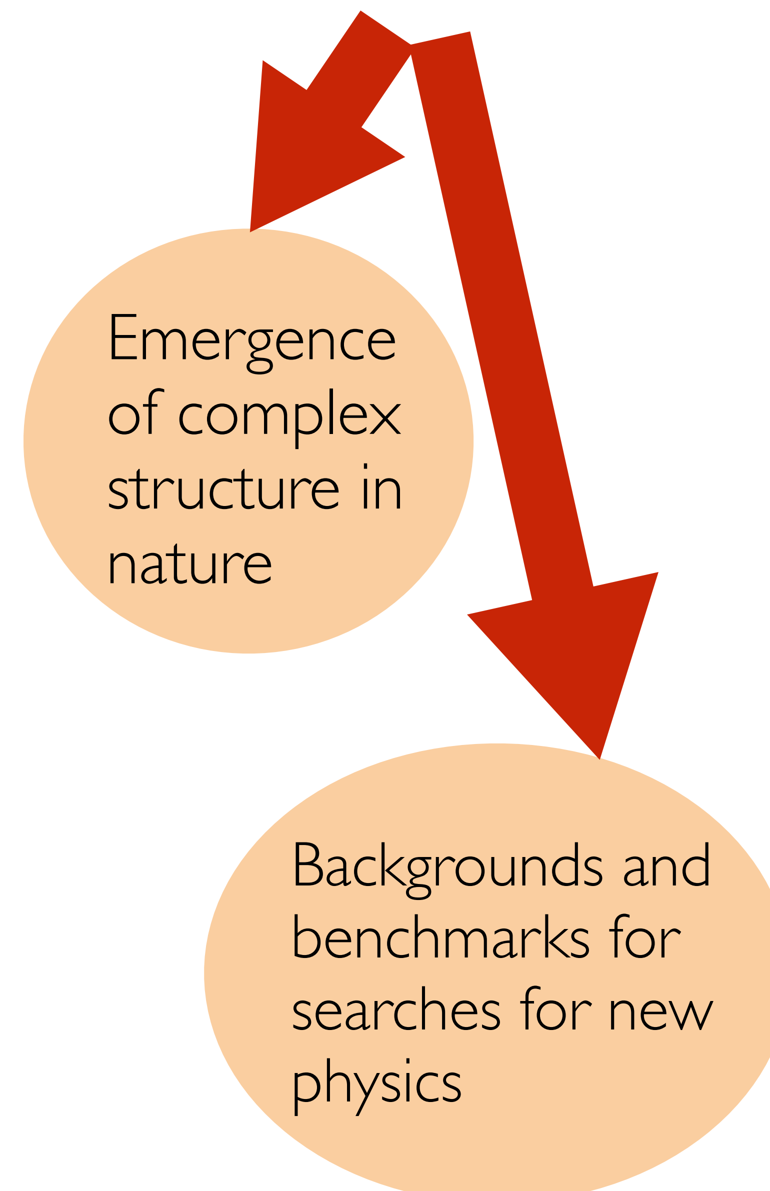
$$r_k = r_{k-1} + \left(l_k \cos\left(\sum_{j \le k} \theta_j \right), l_k \sin\left(\sum_{j \le k} \theta_j \right) \right),$$

Outlook

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

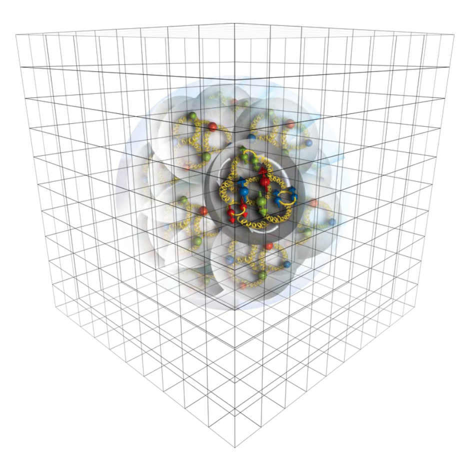
Flow-based generation of QCD gauge fields at scale would

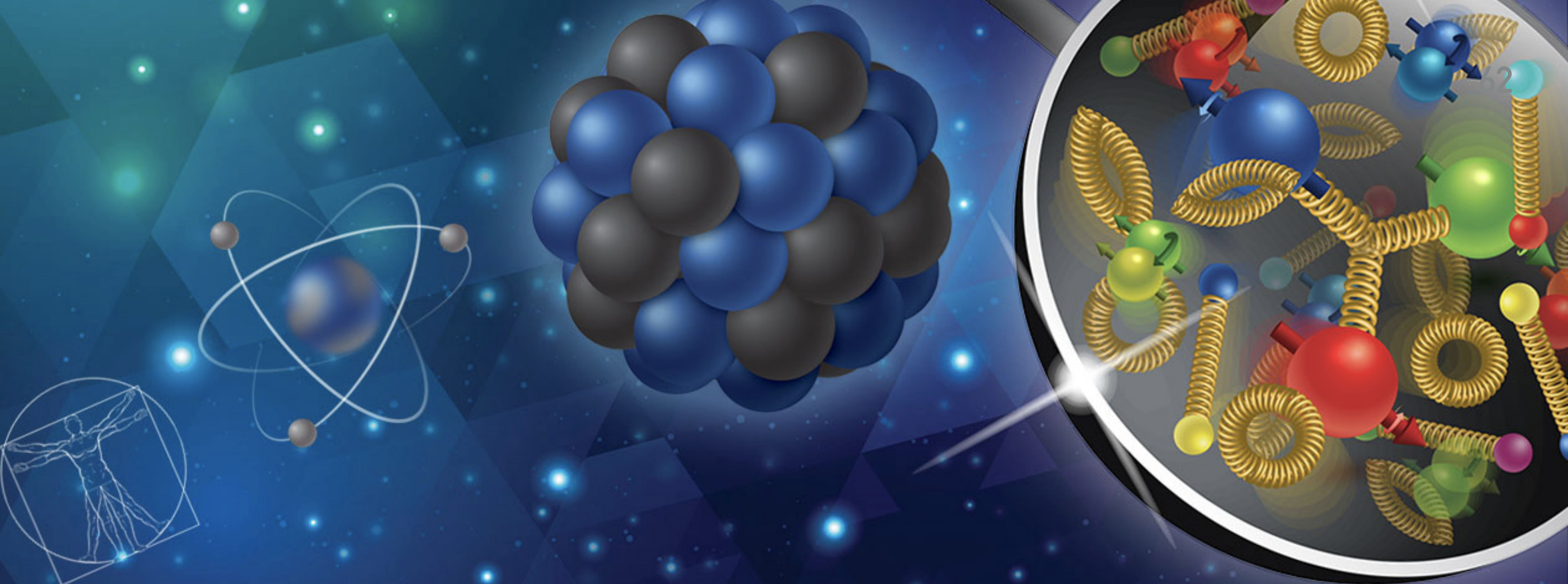
- Enable fast, embarrassingly parallel sampling
  → high-statistics calculations
- Allow parameter-space exploration (re-tune trained models)
- Reduce storage challenges (store only model, not samples)



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
- ….





HUGE POTENTIAL FOR AB-INITO AI

NSF

Massachusetts Institute of Technology

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


NSF
Massachusetts
Institute of
Technology