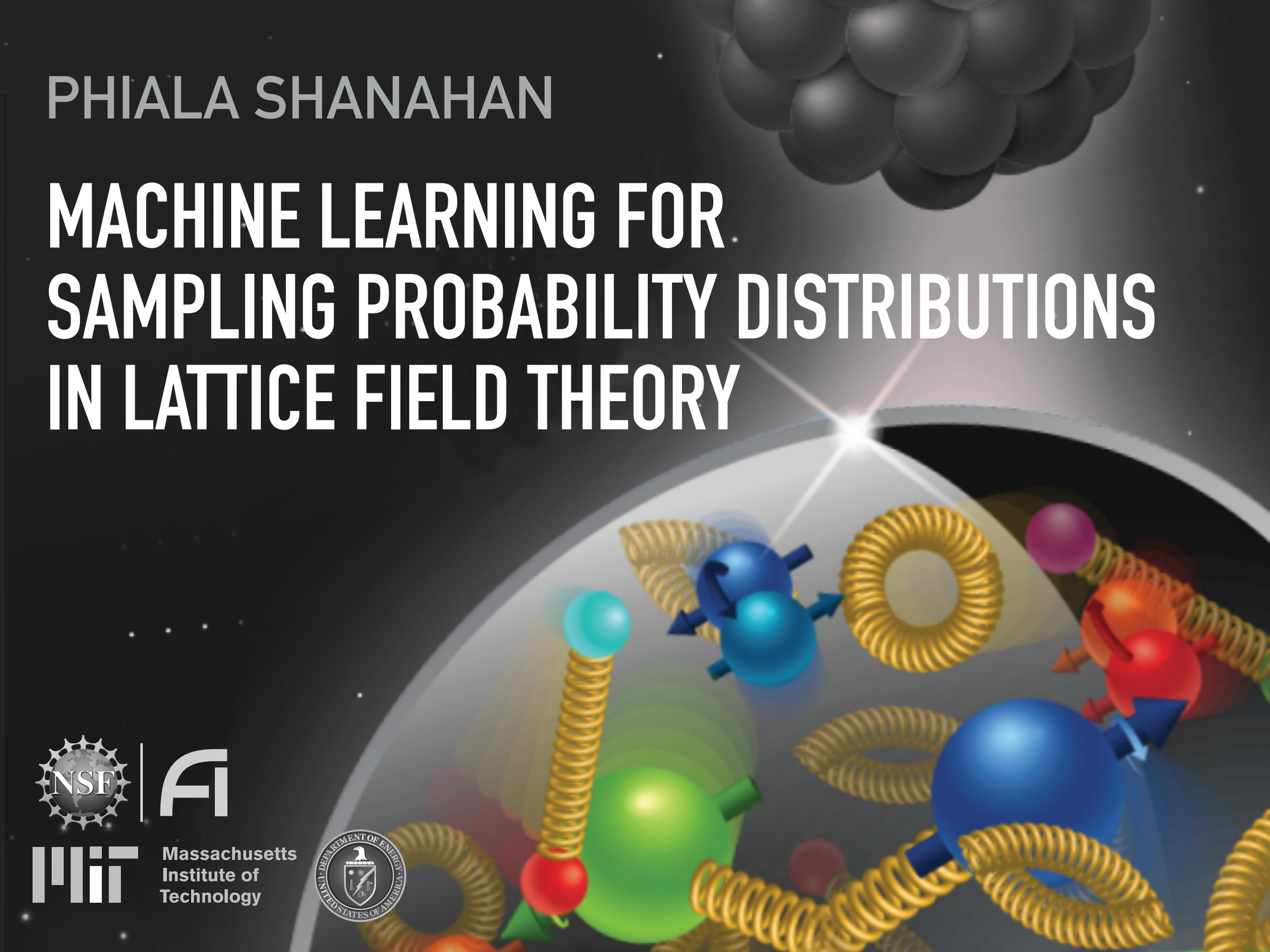
PHIALA SHANAHAN

MACHINE LEARNING FOR SAMPLING PROBABILITY DISTRIBUTIONS IN LATTICE FIELD THEORY



Massachusetts Institute of Technology

Lattice field theory

First-principles calculations of QCD
in the non-perturbative regime

- Lattice field theory is now a well-established **tool for QCD (+QED) calculations**
- Precision results for hadron structure and simple systems
- Beginning of reliable results for nuclear matrix elements and reactions

Algorithmic improvements needed to

- Continue to improve systematics for proton structure studies, e.g.,
  - Composition of the proton
  - Muon g-2
- **Achieve controlled calculations of nuclear systems** for e.g.,
  - Dark matter direct detection
  - Nuclear reactions incl. big bang nucleosynthesis pathway

Lattice field theory

First-principles calculations of QCD
in the non-perturbative regime

- Demand extreme-scale computation

- Require guarantees of exactness, incorporation of complex symmetries

Acceleration via AI/ML algorithms?

Machine learning for first-principles theory

Compute **exact** results from known theory
Use AI/ML to do it **faster**

e.g., lattice QCD calculations, EFT studies, many-body approaches, …

Require **mathematical guarantees of exactness** to preserve rigour of first-principles calculations

No room for approximations, errors, modelling, or any uncertainties which cannot be systematically improved

AI/ML algorithm poorly trained → Results correct, but slower

AI/ML algorithm well trained → Results correct, but faster

Machine learning for first-principles theory

Compute **exact** results from known theory
Use AI/ML to do it **faster**

1. Do the calculation the "same way" but faster

e.g., Tune parameters of existing algorithm using ML
[Free parameters of algebraic multigrid for solving linear systems, automatic differentiation rather than stochastic optimisation]

Machine learning for first-principles theory

Compute **exact** results from known theory
Use AI/ML to do it **faster**

1. Do the calculation the "same way" but faster

e.g., Tune parameters of existing algorithm using ML
[Free parameters of algebraic multigrid for solving linear systems, automatic differentiation rather than stochastic optimisation]

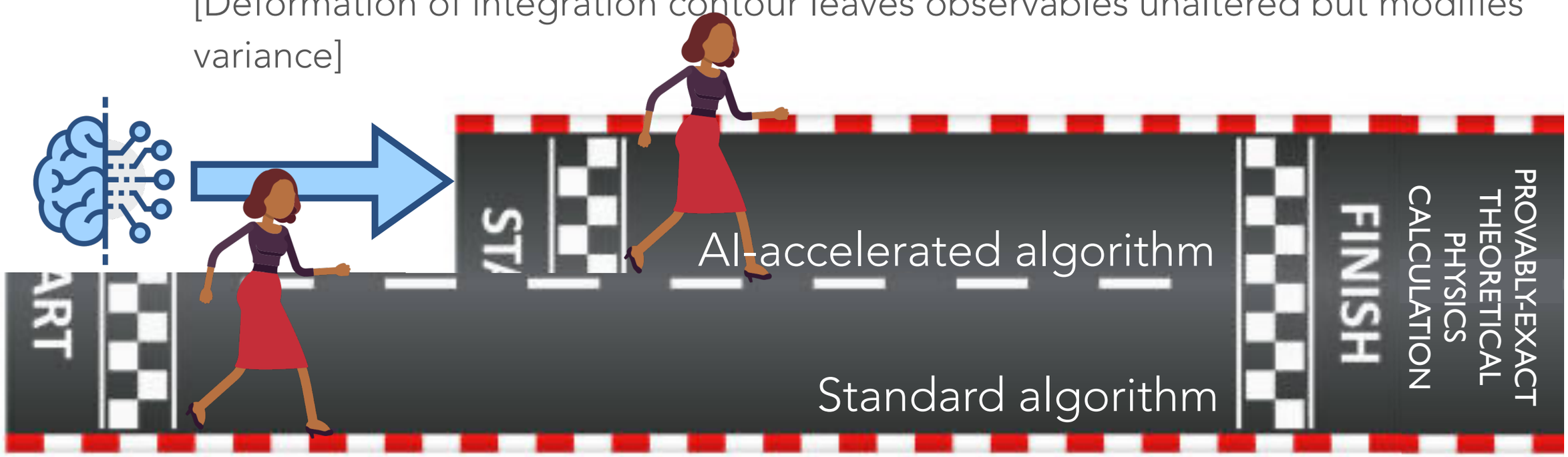
Machine learning for first-principles theory

Compute **exact** results from known theory
Use AI/ML to do it **faster**

2. Transform the problem into a different one with better properties (computationally easier) but same solution

e.g., Preconditioning of any type
[Numerical solver e.g., matrix inversion, faster convergence after preconditioning]

e.g. Change-of-variables
[Deformation of integration contour leaves observables unaltered but modifies variance]



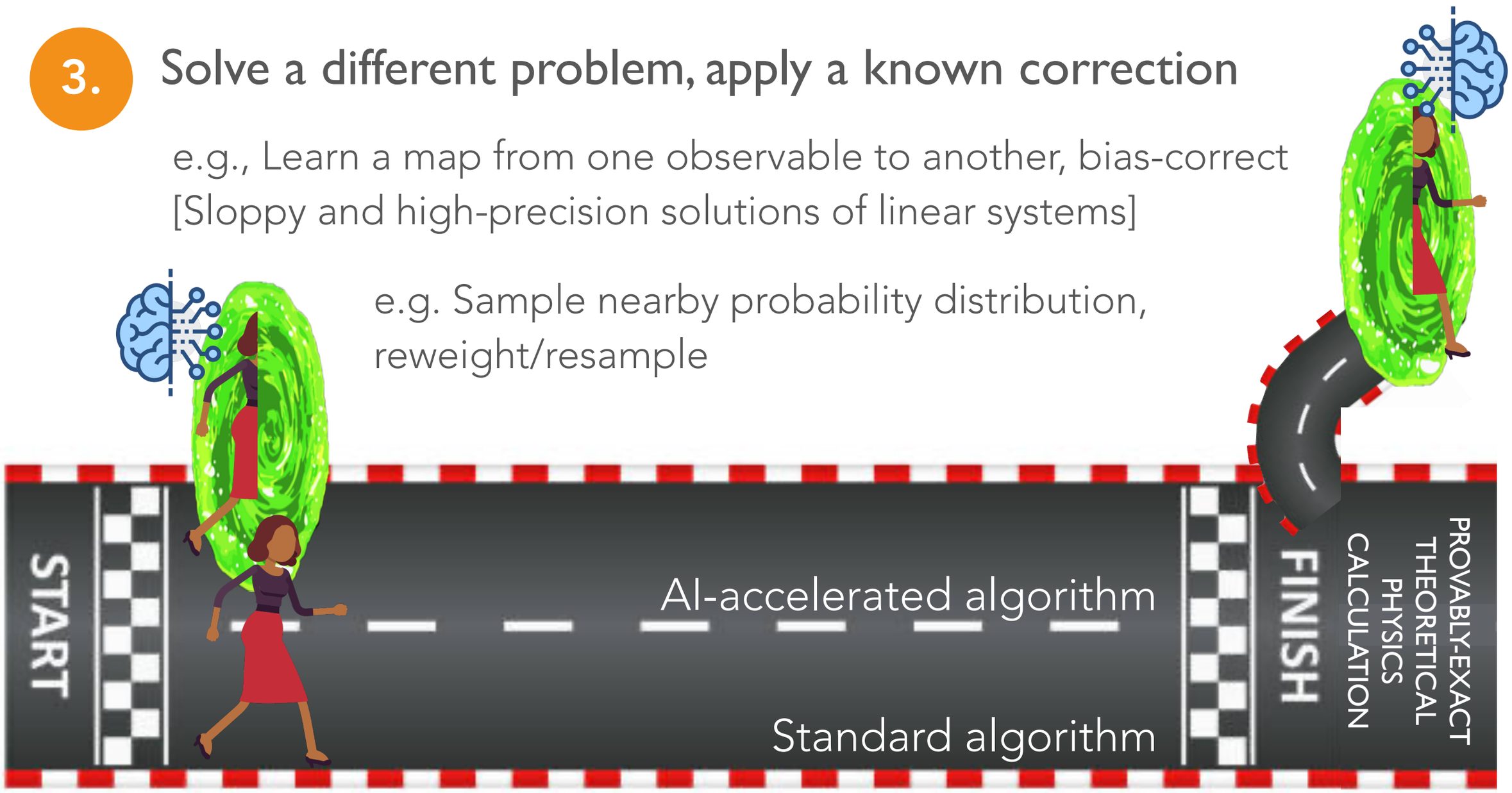
Machine learning for first-principles theory

Compute **exact** results from known theory
Use AI/ML to do it **faster**

3. Solve a different problem, apply a known correction

e.g., Learn a map from one observable to another, bias-correct [Sloppy and high-precision solutions of linear systems]

e.g. Sample nearby probability distribution, reweight/resample

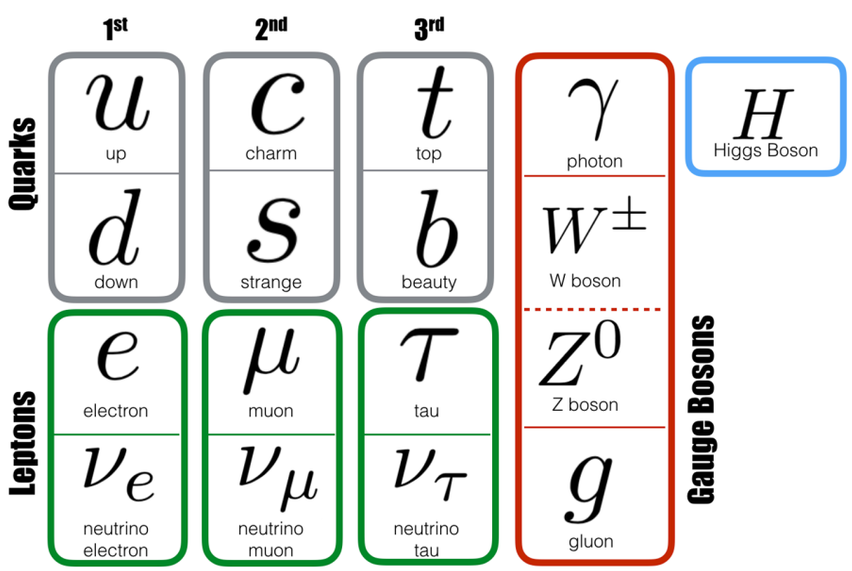


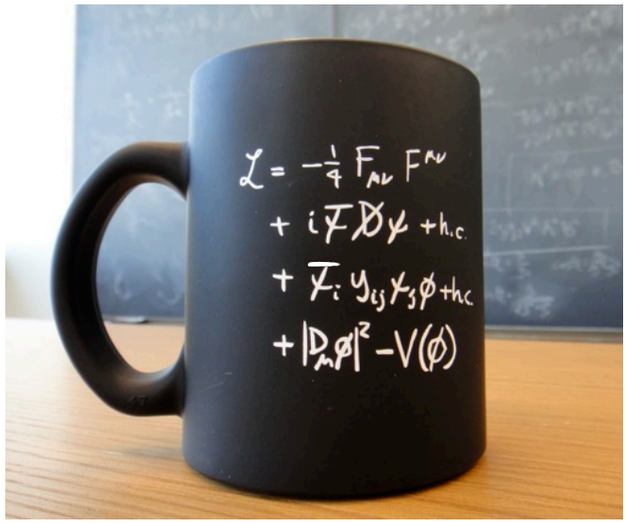
Machine learning for first-principles theory

```
Compute exact results from known theory
Use AI/ML to do it faster
```

3. Solve a different problem, apply a known correction

Now: Details of one example of this approach for lattice field theory calculations

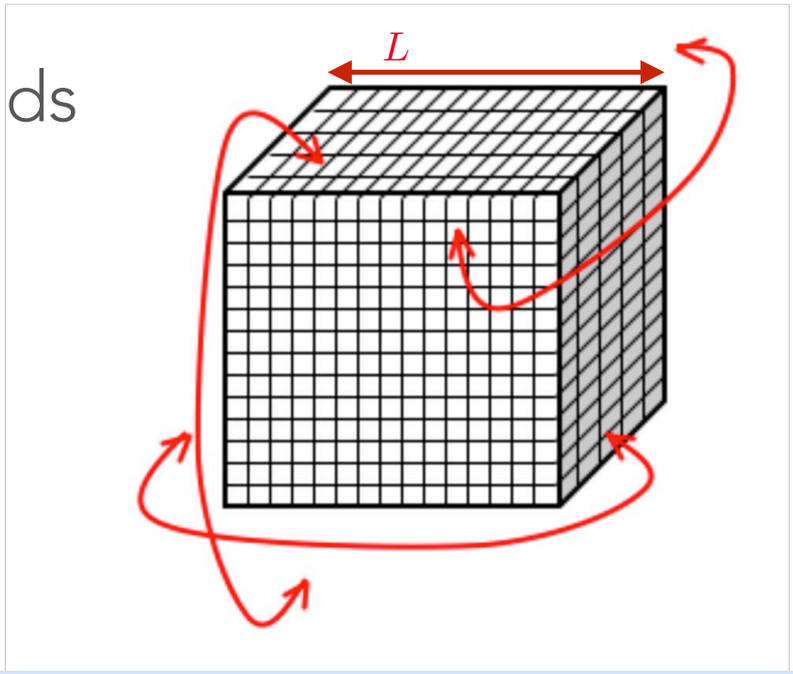




Lattice QCD

Numerical first-principles approach to non-perturbative QCD

- QCD equations ⟷ integrals over fields on lattice (QCD path integral)
- Weighted by action (function encoding dynamics of the theory)
- ~10^{12} variables (for state-of-the-art)



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

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

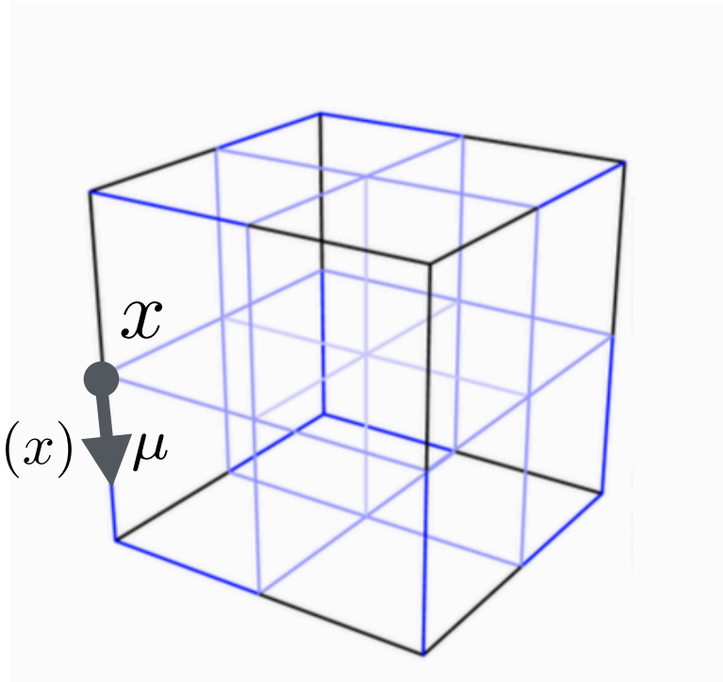
with field configurations U^i distributed according to $e^{-S[U]}$

Generate QCD gauge fields

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

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

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



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

  ➡ Weighted averages over configurations determine physical observables of interest

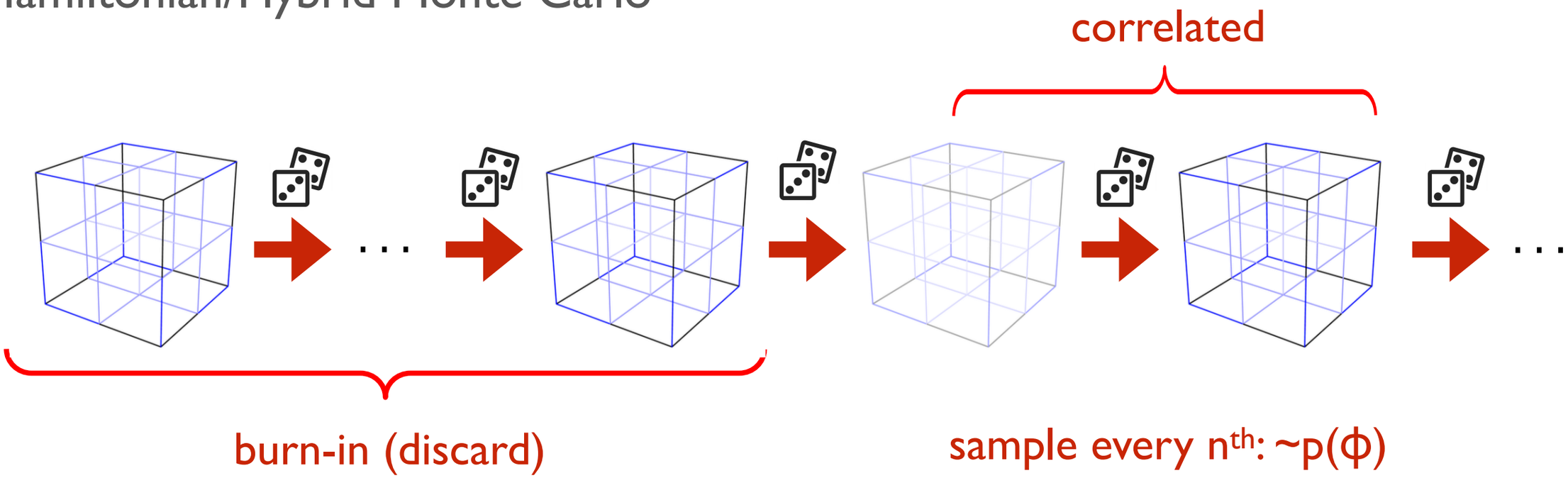
- Calculations use ~10^3 configurations

Generate QCD gauge fields

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

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

Hamiltonian/Hybrid Monte Carlo

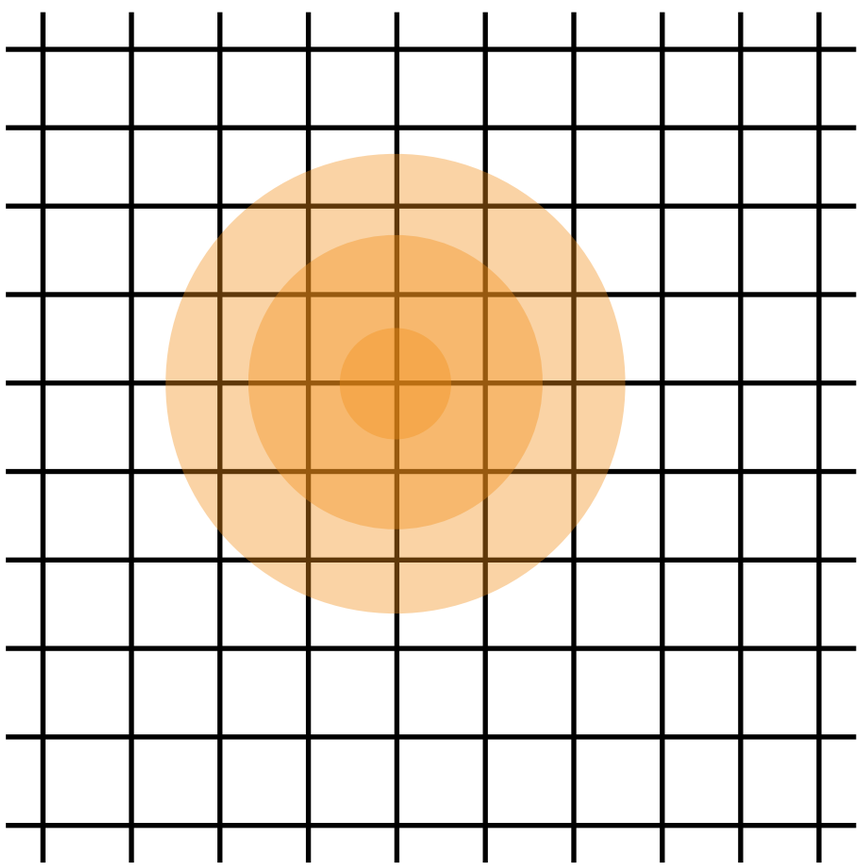


Burn-in time and correlation length dictated by Markov chain '**auto-correlation time**': shorter autocorrelation time implies less computational cost

Generate QCD gauge fields

QCD gauge field configurations sampled via

Hamiltonian dynamics + Markov Chain Monte Carlo



Updates diffusive

Lattice spacing → 0

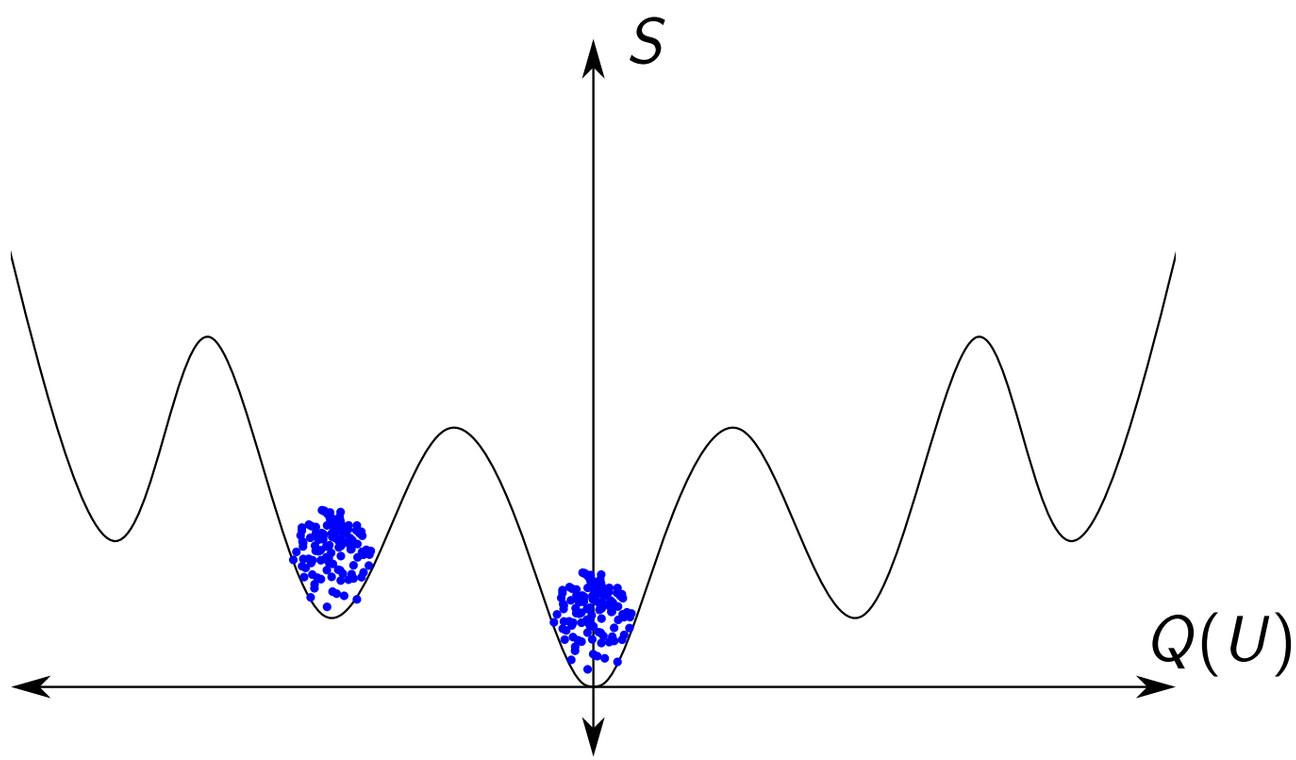
Number of updates to change fixed physical length scale → ∞

"Critical slowing-down" of generation of uncorrelated samples

Generate QCD gauge fields

QCD gauge field configurations sampled via

Hamiltonian dynamics + Markov Chain Monte Carlo



Related problem:

"Freezing" of sampling into modes of the distribution (e.g., topological sectors)

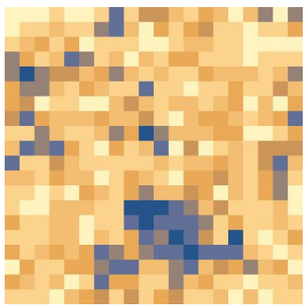
Generate QCD gauge fields

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

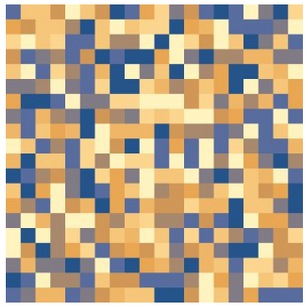
$$P[\phi(x)] \sim e^{-S[\phi(x)]}$$

Parallels with image generation problem

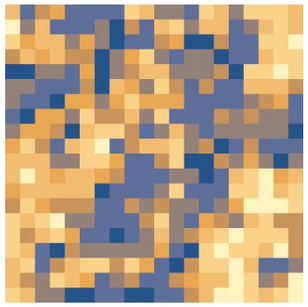
likely
(log prob = 22)



unlikely
(log prob = -6107)



likely
(log prob = 5)

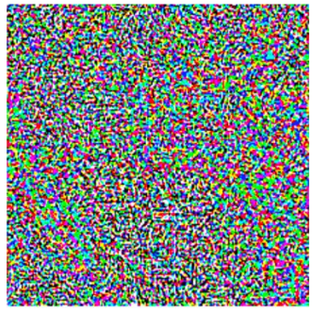


likely

[Karras, Lane, Aila / NVIDIA 1812.04948]

likely

unlikely



Generate QCD gauge fields

Ensemble of lattice QCD gauge fields

- Ensemble of gauge fields has meaning
- $64^3 \times 128 \times 4 \times 9 \times 2$ $\simeq 10^9$ 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

Generate QCD gauge fields

Ensemble of lattice QCD gauge fields

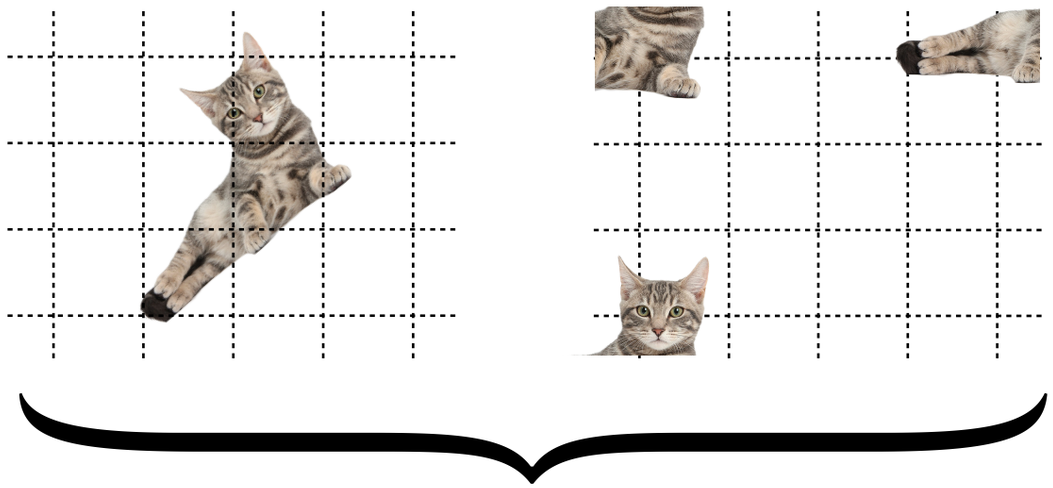
- Ensemble of gauge fields has meaning
- $64^3 \times 128 \times 4 \times 9 \times 2$ $\simeq 10^9$ numbers
- ~1000 samples
- Long-distance correlations are important
- Gauge and translation-invariant with periodic boundaries

Physics is invariant under specific field transformations

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

Gauge field configuration

Transformed gauge field configuration



Encode same physics

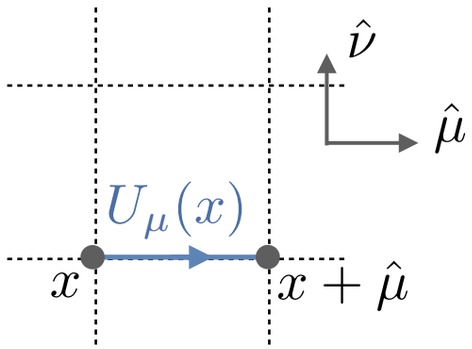
Generate QCD gauge fields

Ensemble of lattice QCD gauge fields

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

Physics is invariant under specific field transformations

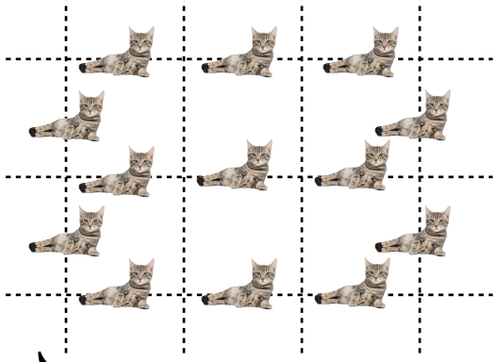
- **Gauge transformation**



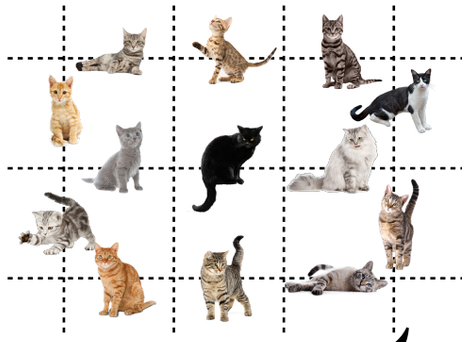
$$U_\mu(x) \to \Omega(x) U_\mu(x) \Omega^\dagger(x+\hat{\mu})$$

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

Gauge field configuration



Transformed gauge field configuration



Encode same physics

Generate QCD gauge fields

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
  000 numbers

tran

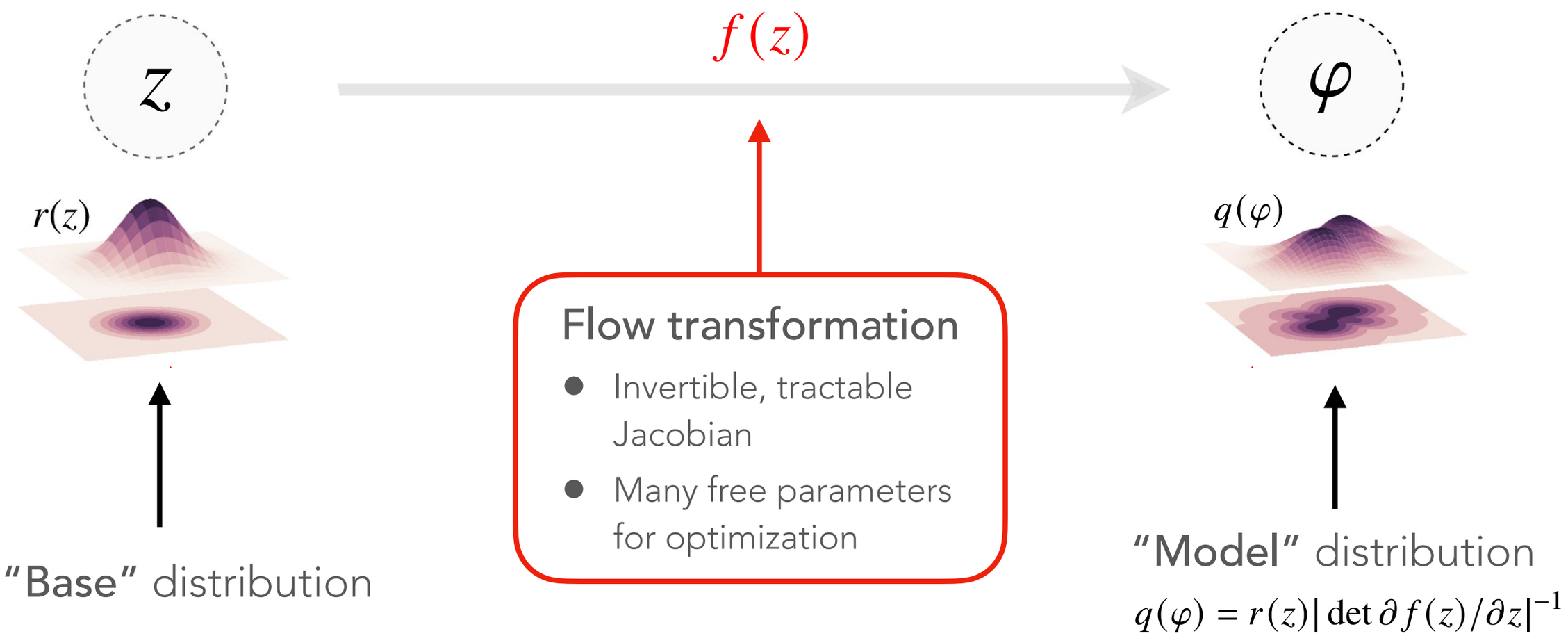
Need custom ML for physics from the ground up

AB-INITIO AI

https://iaifi.org/

Flow models for lattice QCD

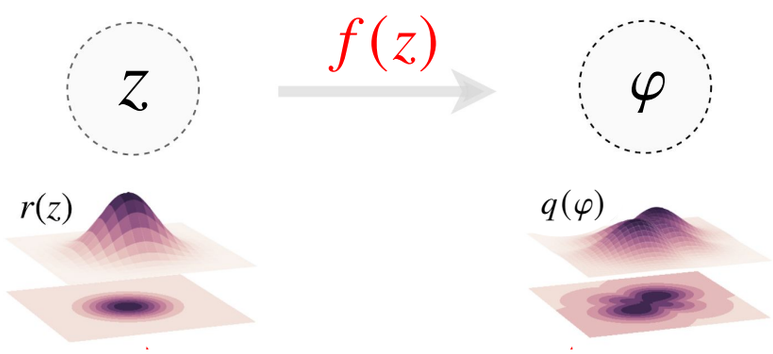
Flow models: Machine-learned maps between probability distributions
[Rezende & Mohamed 1505.05770]



Fields via flow models

Example application: Embarrassingly parallel direct sampling

Flow model as an approximate trivialising map



"Base" distribution:
Efficient to sample
e.g., Haar-uniform

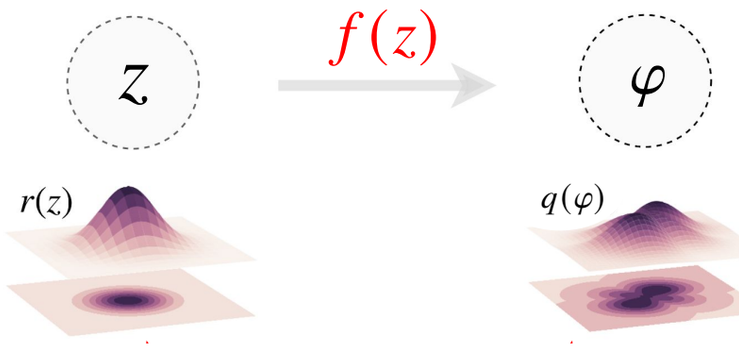
"Model" distribution
$q(\phi) \approx \frac{1}{Z} e^{-S(\phi)}$

- Independent samples of the base distribution map to independent samples of the model distribution

Fields via flow models

Example application: Embarrassingly parallel direct sampling

Flow model as an approximate trivialising map



"Base" distribution:
Efficient to sample
e.g., Haar-uniform

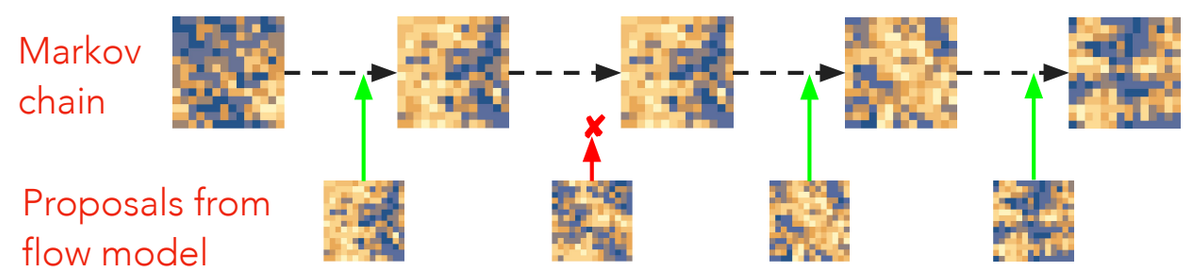
"Model" distribution
$q(\phi) \approx \frac{1}{Z} e^{-S(\phi)}$

- Independent samples of the base distribution map to independent samples of the model distribution

- **Train the model:**
  Gradient descent to minimise "loss function" with minimum at $q(\phi) = \frac{1}{Z} e^{-S(\phi)}$

$$L(q) = \int \underbrace{d\phi\, q(\phi)}_{\text{Estimate stochastically by sampling from the model, i.e., "self training"}} [\log q(\phi) + S(\phi)]$$

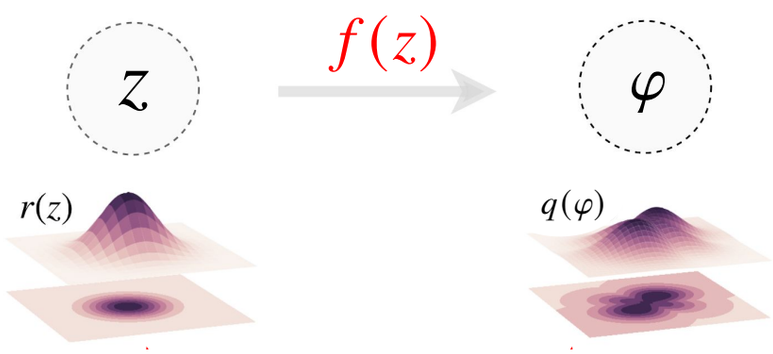
- **Guarantee exactness:**
  Reweight or form a Markov chain with Metropolis-Hastings accept/reject step



Fields via flow models

Example application: Embarrassingly parallel direct sampling

Flow model as an approximate trivialising map



"Base" distribution:
Efficient to sample
e.g., Haar-uniform

"Model" distribution
$q(\phi) \approx \frac{1}{Z} e^{-S(\phi)}$

- Independent samples of the base distribution map to independent samples of the model distribution

Proof-of-principle applications to simple lattice field theories reveal many potential advantages c.f. HMC

- Mitigation of **critical slowing-down** and **topological freezing**
- Efficient **parameter-space exploration** (by re-tuning trained models)
- Direct access to the **partition function**

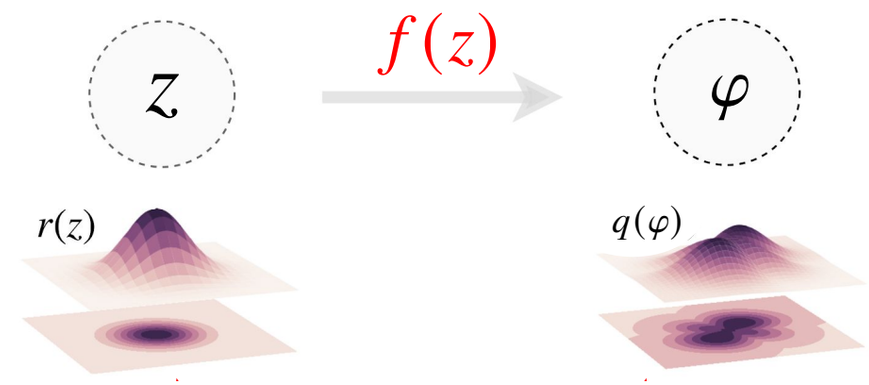
Direct sampling is only one of many approaches to using flow models for lattice QCD!

Flow models for lattice QCD

Flow models: Machine-learned maps between probability distributions
[Rezende & Mohamed 1505.05770]

Many possible applications of flow models in lattice QCD

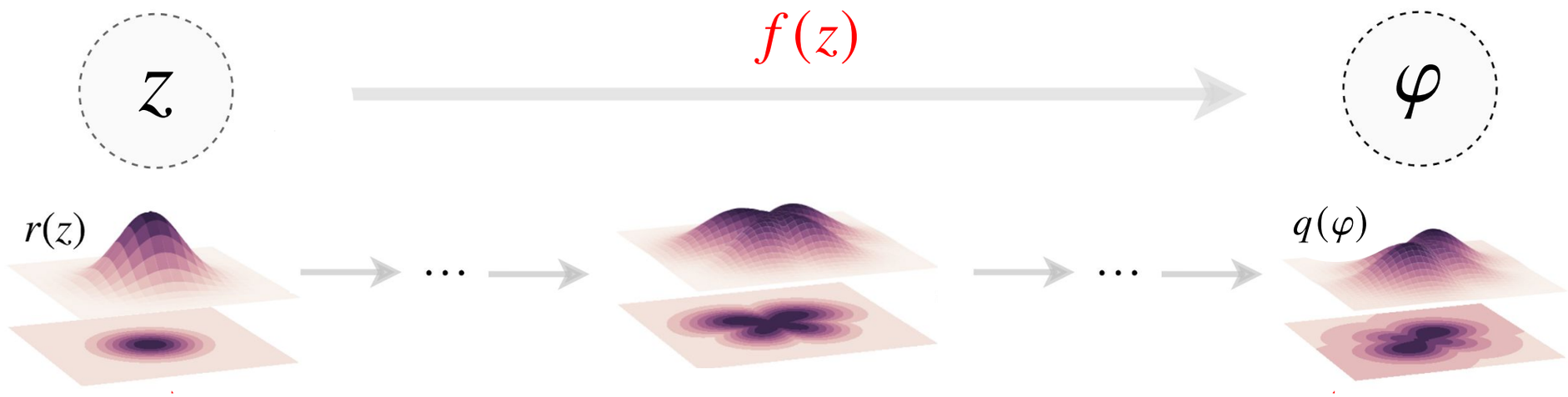
- Direct sampling i.e., $r(z)$ is a trivial distribution and $q(\phi) \approx \frac{1}{Z} e^{-S(\phi)}$
  Generalisation of [Lüscher 0907.5491]
- Hybrid sampling approaches
  e.g.,generalize the proposal distribution in HMC [Foreman et al., 2112.01582]; flows in lattice subdomains [Finkenrath 2201.02216]
- Map from one theory / set of parameters to another
- Contour deformation and density-of-states approaches to sign problem
  [Detmold et al., 2101.12668, Pawlowski+Urban 2203.01243, Lawrence et al., 2205.12303, etc]



Flow architectures designed for QCD gauge fields can be trained and applied in many different ways!

Flow models for lattice QCD

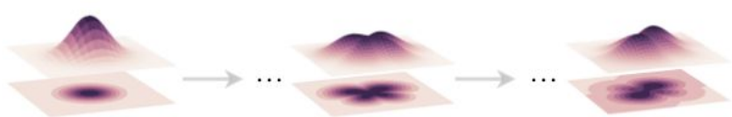
Flow models: Machine-learned maps between probability distributions
[Rezende & Mohamed 1505.05770]

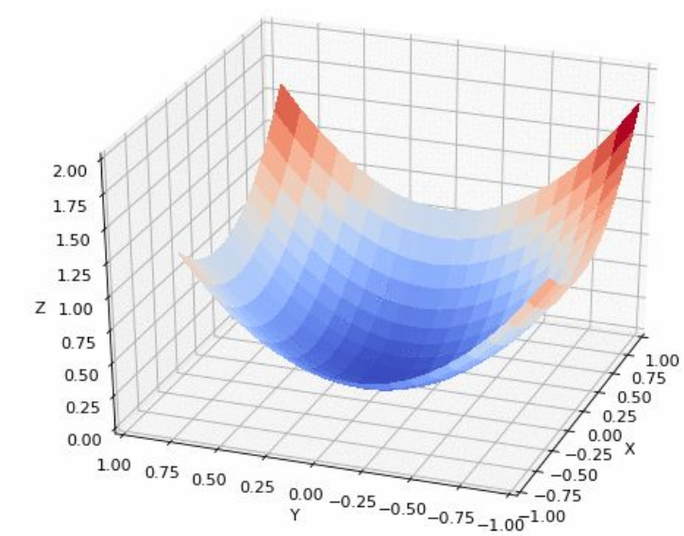


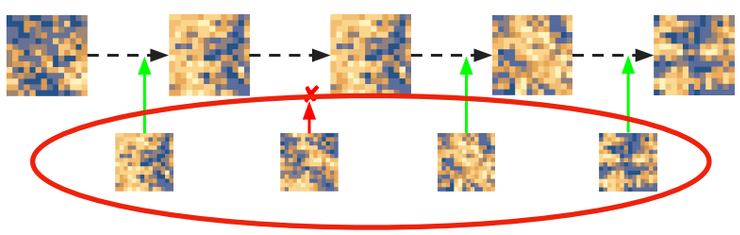
Goal: engineer flow architectures that effectively parameterise transformations of lattice gauge fields

- Diffeomorphisms on lattice field degrees of freedom
- Encode symmetries, e.g., gauge symmetry
- Flexible/expressive/can encode correlations at physically-relevant scale etc

Flow models for fields







Generating samples is
"embarrassingly parallel"

Parameterize flow

Each layer contains
arbitrary neural networks
with many free parameters

Training step

Desired accuracy?

Draw samples from model

Compute loss function

Gradient descent

Save trained model

Markov chain using samples from model

Flow models for lattice QCD

- **Ongoing program** to develop flow model architectures for applications across lattice QCD

✓ **First flow architectures for lattice field theory** (scalar field theory) [Albergo et al., 1904.12072]

✓ **Gauge field theories**

- Flow transformations on compact, connected manifolds [Rezende et al., 2002.02428]
- Gauge-equivariant architectures: Abelian field theories [Kanwar et al., 2003.06413, 2101.08176]
- Gauge-equivariant architectures: non-Abelian field theories [Boyda et al., 2008.05456]

✓ **Theories with fermions**

- Architectures for theories with fermions [Albergo et al., 2106.05934]
- Combining architectures for gauge fields and fermions [Albergo et al., 2202.11712]
- Techniques to incorporate pseudofermions [Abbott et al., 2207.08945]

✓ **Initial application to QCD in 4D**
[Abbott et al., 2208.03832, 2211.07541]

🏁 **Architectures for QCD at scale** [ongoing; Aurora Early Science Project]

AURORA | EARLY SCIENCE PROGRAM FOR DATA & LEARNIN

[see also tutorial notebook 2101.08176, work on multimodal distributions 2107.00734]

Flow models for lattice QCD

- **Ongoing program** to develop flow model architectures for applications across lattice QCD

✓ **First flow architectures for lattice field theory** (scalar field theory) [Albergo et al., 1904.12072]

✓ Gauge field theories

- Flow transformations on compact, connected manifolds [Rezende et al., 2002.02428]
- Gauge-equivariant architectures: Abelian field theories [Kanwar et al., 2003.06413, 2101.08176]
- Gauge-equivariant architectures: non-Abelian field theories [Boyda et al., 2008.05456]

✓ Theories with fermions

- Architectures for theories with fermions [Albergo et al., 2106.05934]
- Combining architectures for gauge fields and fermions [Albergo et al., 2202.11712]
- Techniques to incorporate pseudofermions [Abbott et al., 2207.08945]

✓ Initial application to QCD in 4D
[Abbott et al., 2208.03832, 2211.07541]

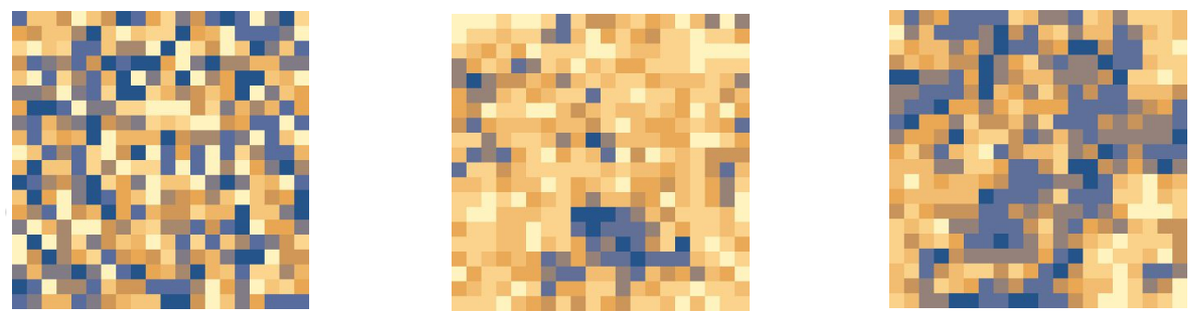
🏁 Architectures for QCD at scale [ongoing; Aurora Early Science Project]

AURORA | EARLY SCIENCE PROGRAM FOR DATA & LEARNIN

[see also tutorial notebook 2101.08176, work on multimodal distributions 2107.00734]

Flow models for scalar fields

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



- Action: kinetic terms and quartic coupling

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

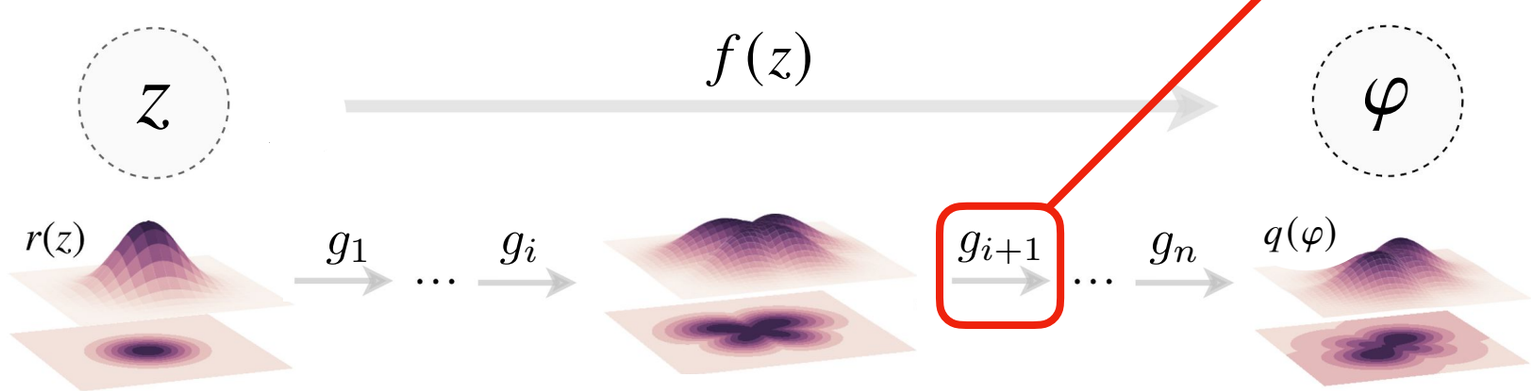
Generate field configurations $\phi(x)$ with probability

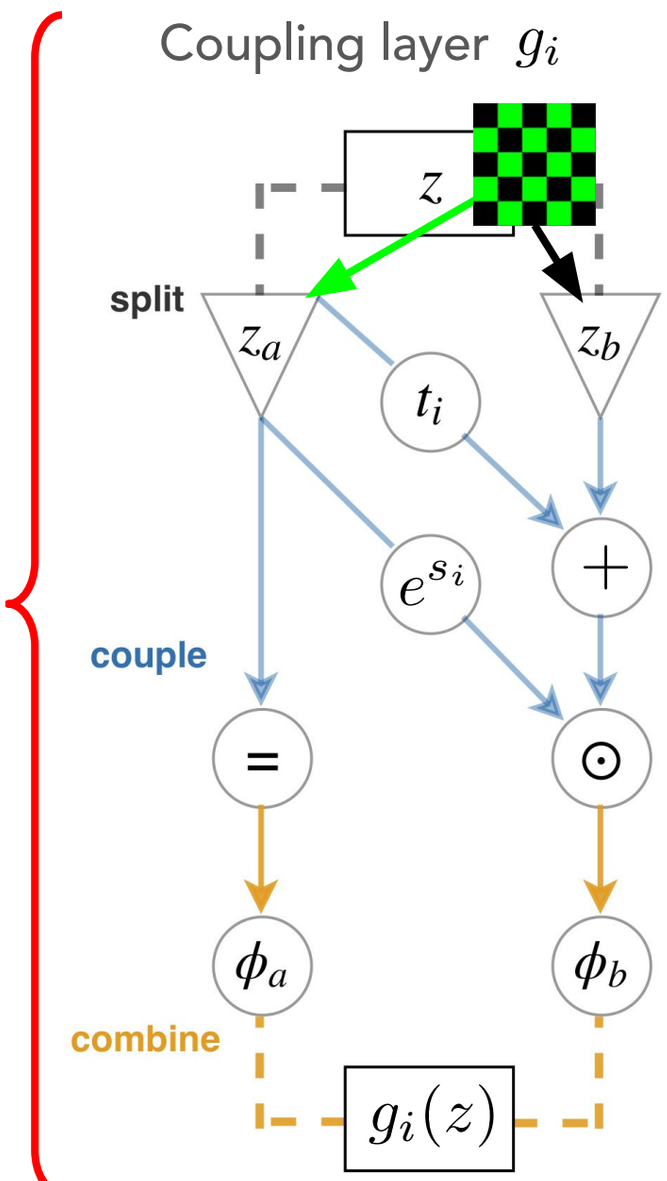
$$P[\phi(x)] \sim e^{-S[\phi(x)]}$$

Flow models for scalar fields

- First demonstration of flows as non-sequential samplers for lattice field theory [Albergo et al., 1904.12072]
- Variation of **"real non-volume-preserving flows"** developed for image generation [Dinh et al., 1605.08803]
  - Update field via sequential "coupling layers" g_i
  - Each layer transforms half of the degrees of freedom conditioned on the other half

    $z_a \to \phi_a = z_a \qquad z_b \to \phi_b = z_b e^{s(z_a)} + t(z_a)$
  - Transformations parameterised by arbitrary neural networks s_i, t_i

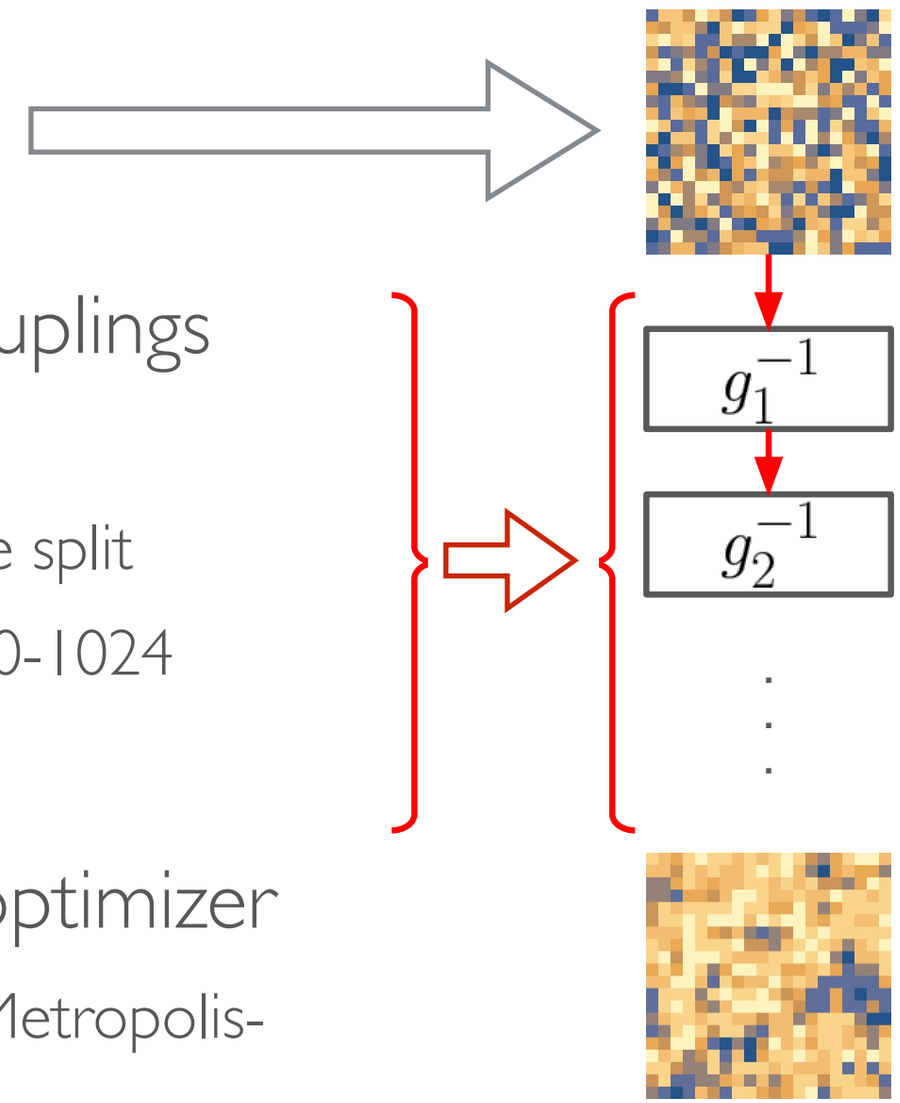




Flow models for scalar fields

- Prior distribution chosen to be uncorrelated Gaussian:

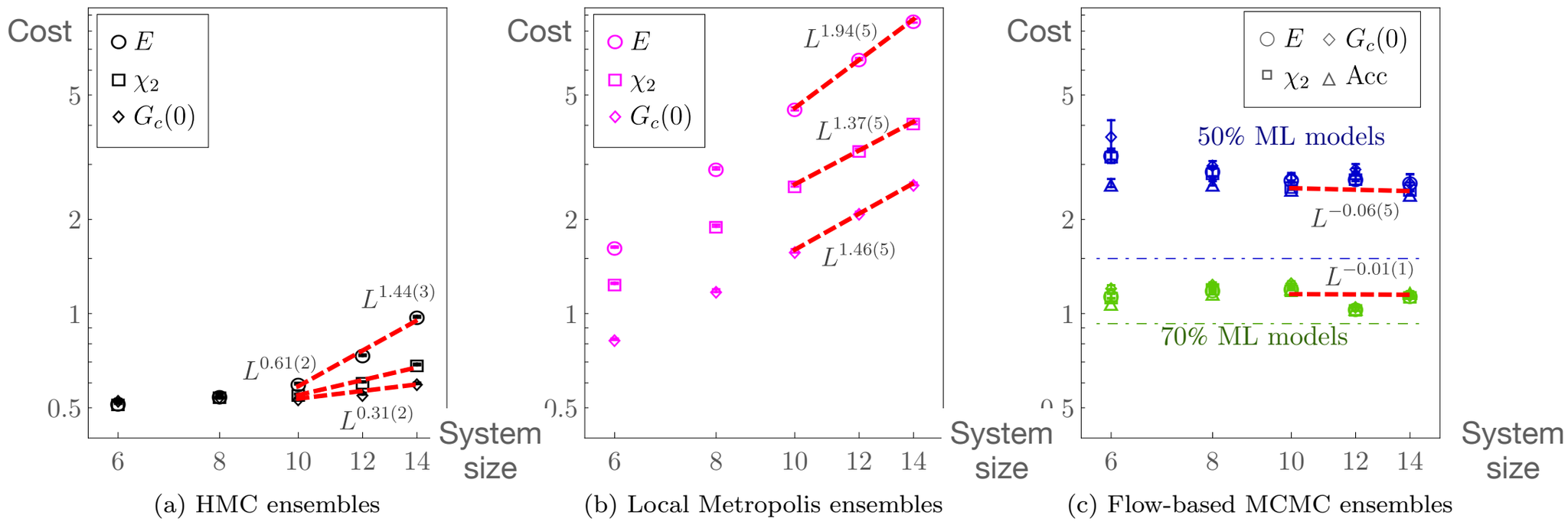
$$\phi(x) \sim \mathcal{N}(0,1)$$



- 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

Flow models for scalar fields

Demonstration of accelerated sampling at the cost of model engineering and training (ϕ^4 theory, 2D, parameters tuned for constant $m_p L$) [Albergo et al., 1904.12072]



- **Many choices in architecture design** (e.g., prior distribution, variable splitting, neural network structure); further work by our group and others [e.g., Nicoli et al., 2007.07115, 2111.11303; Del Debbio et al., 2105.12481; Singha et al., 2207.00980; +…]
- Current best implementations by our group **orders of magnitude more efficient** than 2019 approach! → Architecture development matters

Flow models for lattice QCD

- **Ongoing program** to develop flow model architectures for applications across lattice QCD

✓ First flow architectures for lattice field theory (scalar field theory) [Albergo et al., 1904.12072]

✓ Gauge field theories

- Flow transformations on compact, connected manifolds [Rezende et al., 2002.02428]
- Gauge-equivariant architectures: Abelian field theories [Kanwar et al., 2003.06413, 2101.08176]
- Gauge-equivariant architectures: non-Abelian field theories [Boyda et al., 2008.05456]

✓ Theories with fermions

- Architectures for theories with fermions [Albergo et al., 2106.05934]
- Combining architectures for gauge fields and fermions [Albergo et al., 2202.11712]
- Techniques to incorporate pseudofermions [Abbott et al., 2207.08945]

✓ Initial application to QCD in 4D

[This talk+upcoming manuscripts on scaling and 4D]

🏁 Architectures for QCD at scale [ongoing; Aurora Early Science Project]

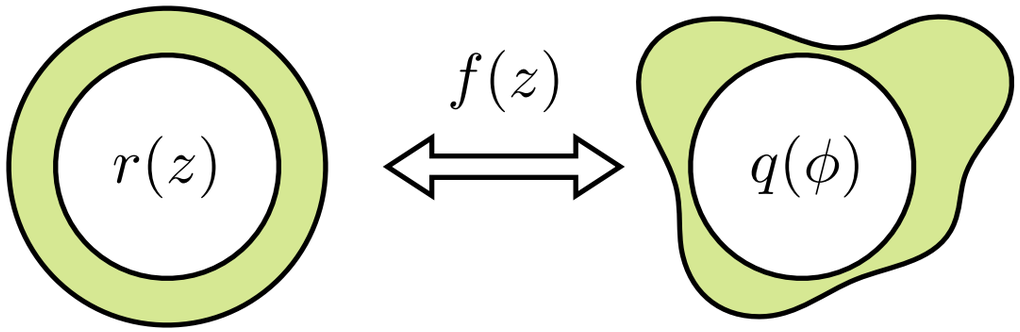
AURORA | EARLY SCIENCE PROGRAM FOR DATA & LEARNIN

[see also tutorial notebook 2101.08176, work on multimodal distributions 2107.00734]

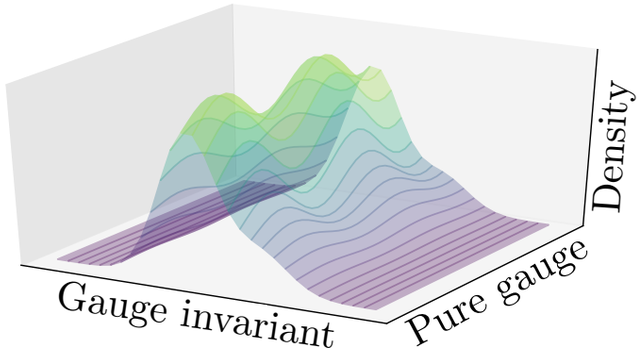
Flow models for gauge field theories

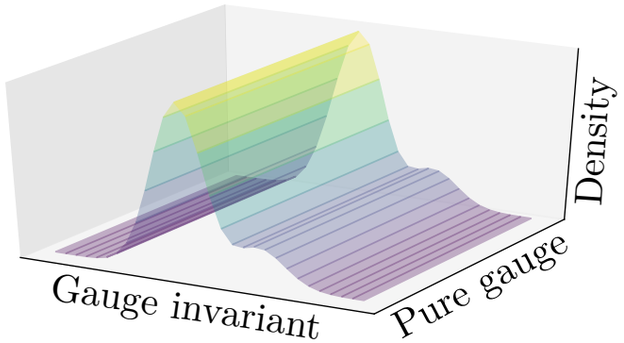
Flow models for gauge field theories require additional developments:

- Definition of flow transformations on **compact connected manifolds** (unlike real transformations relevant for images, scalar field theory) [Rezende et al., 2002.02428]



- Encoding complex **symmetries** of probability distribution (spatial, gauge, …) [Kanwar et al., 2003.06413, Boyda et al., 2008.05456; Related ideas in Favoni et al., 2012.12901,2111.04389; Luo et al., 2012.05232]
  - Not essential for correctness
  - Crucial for practical training of high-dimensional models with high-dimensional symmetries

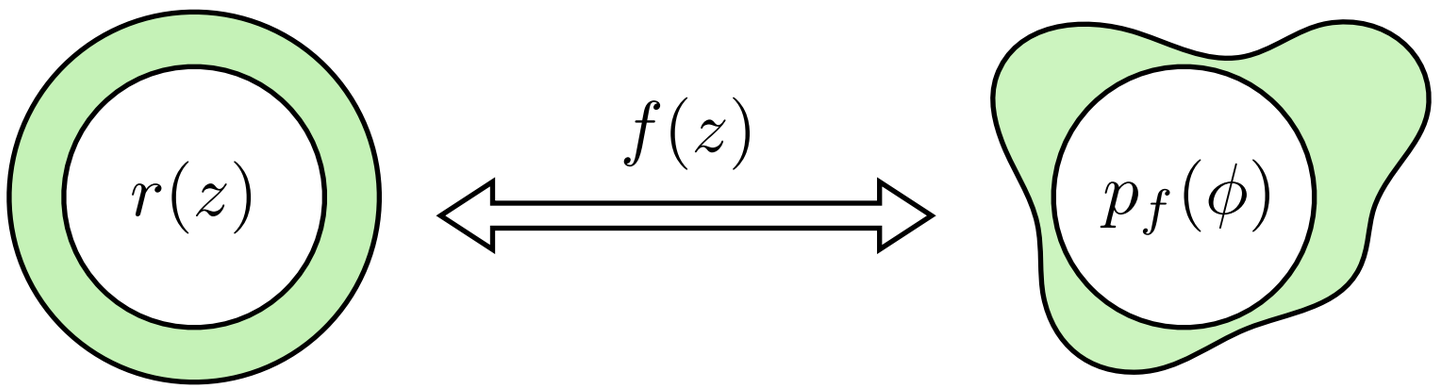




Flows on spheres and tori

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

$$f(\theta) = \sum_i \rho_i f_i(\theta) \quad \begin{array}{l} \rho_i \geq 0 \\ \sum_i \rho_i = 1 \end{array}$$

Flows on spheres and tori

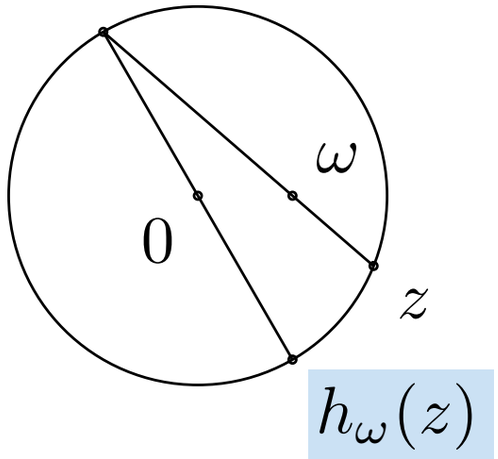
Test case: Flows on the circle

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

- 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

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

Gauge-equivariant flows

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, i.e., variable splitting

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

Links updated by coupling layer

Links frozen in coupling layer

Gauge-equivariant flows

Generative flow architecture that is ***gauge-equivariant***

Define invertible, equivariant coupling layer

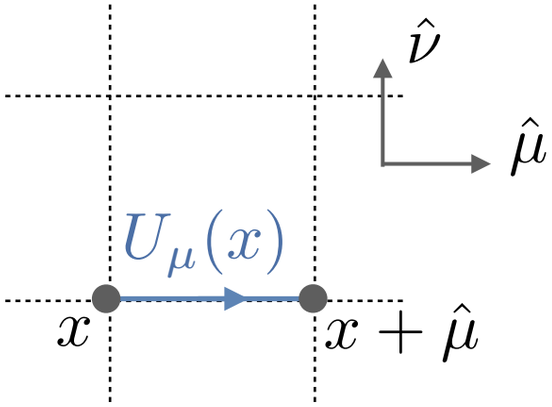
Spacetime dimension

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

Lattice volume

Many choices for coupling layer construction, incl:

- **Residual flows** [Abbott et al., 2305.02402]
  - Transform links $U_\mu(x)$ through parameterised "smearing" steps
- **Spectral flows** [Kanwar et al., 2003.06413, Boyda et al., 2008.05456]
  - Transform links via transformations of plaquettes/Wilson loops (products of links starting and ending at the same lattice site)



[Other related work in e.g., Favoni et al., 2012.12901,2111.04389; Luo et al., 2012.05232, Bacchio et al., 2212.08469, + …]

Residual flows

Inspired by Lüscher's trivialising maps [Lüscher 0907.5941]

- Transform links $U_\mu(x)$ through parameterised "smearing" steps

Link updates via

Link updated by coupling layer

$$U'_\mu(x) = e^{i\epsilon \partial_{x,\mu} \phi(U)} U_\mu(x)$$

Lie-algebra-valued derivative

Gauge-invariant "potential function" depending on frozen links i.e., elements of U^B.

Potential may be defined e.g., in terms of traced plaquettes (or more general loops)

$$\phi(U) = \sum_x \sum_{\mu \neq \nu} c_{\mu\nu}(x; U^B) \operatorname{Re} \operatorname{Tr}(P_{\mu\nu})$$

For invertibility, require $\epsilon \lesssim 1/8$

Spectral flows

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

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

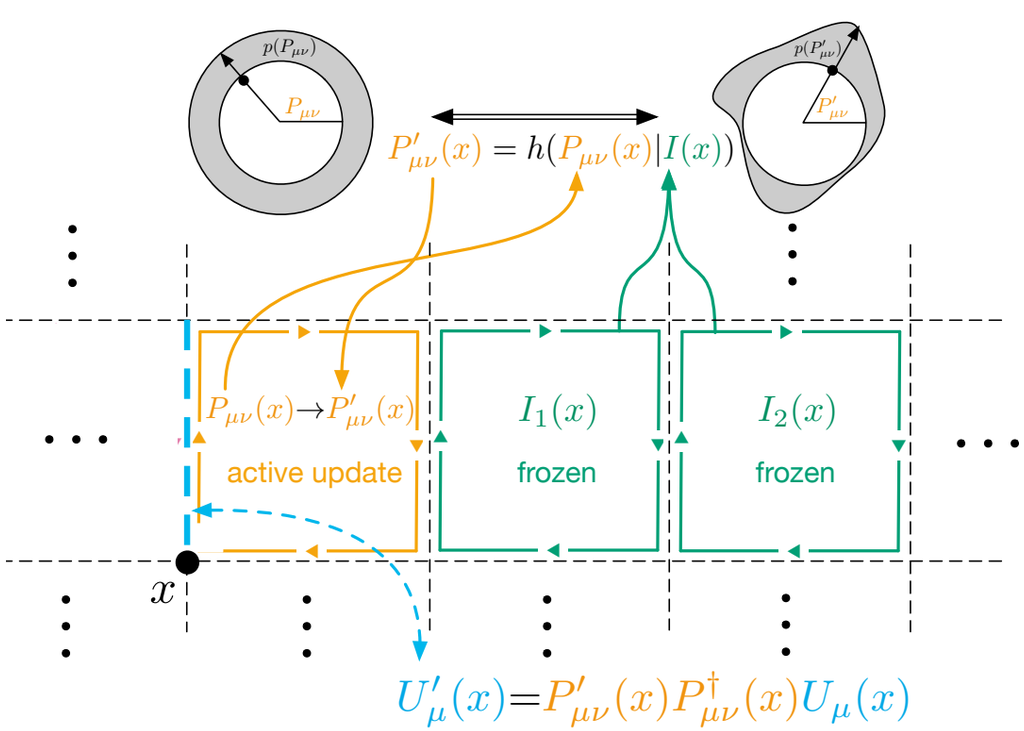
Link updated by coupling layer

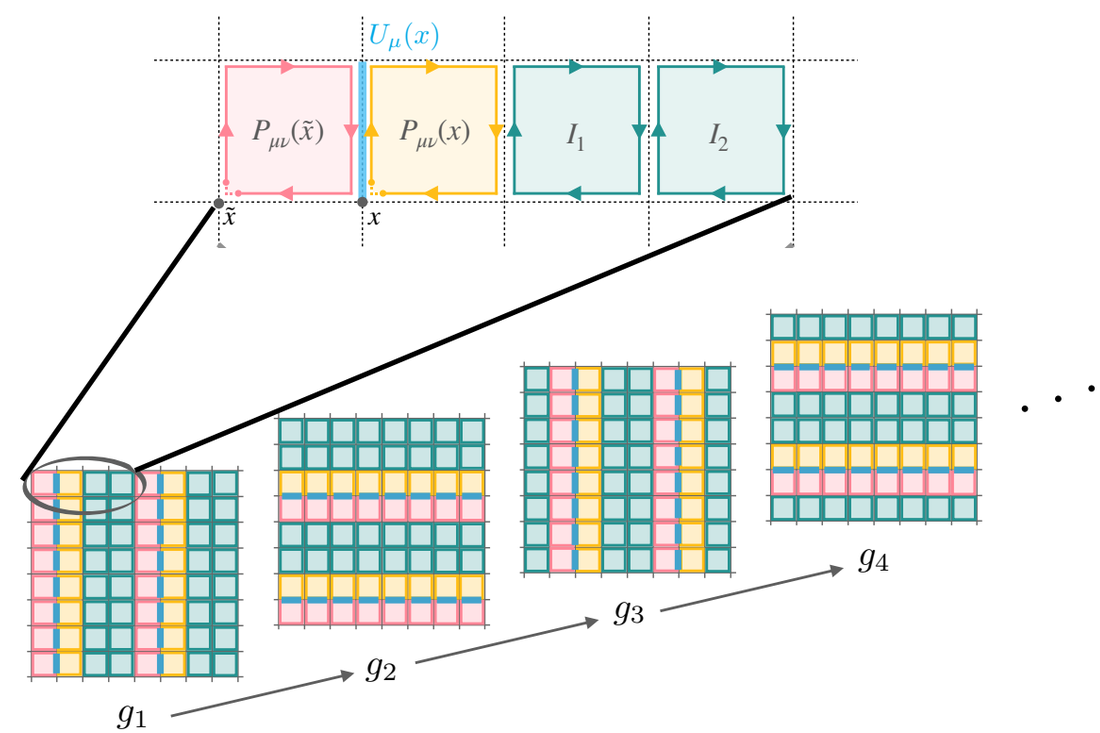
$$U'^i = h(U^i S^i | I^i) S^{i\dagger}$$

Loop that starts and ends at same point

Coupling layer equivariant under the condition

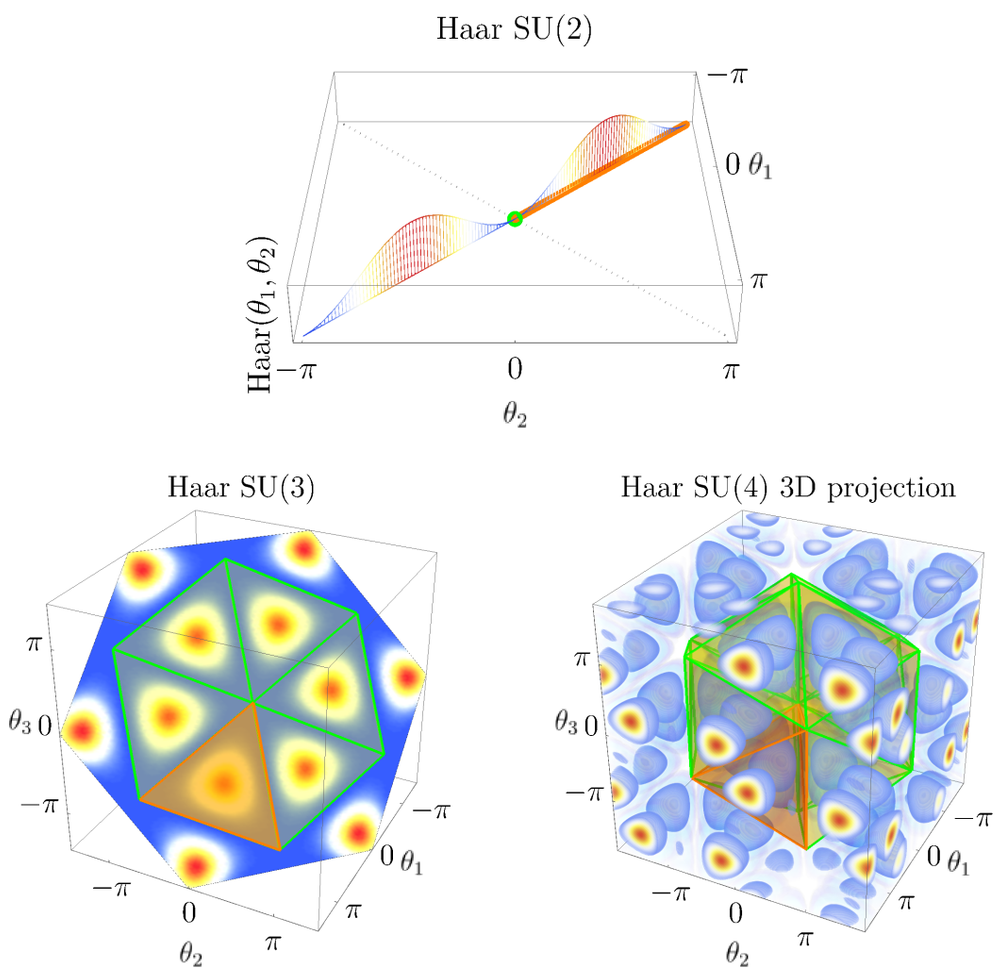
$$h(XWX^\dagger) = X\, h(W)\, X^\dagger, \quad \forall X, W \in G$$

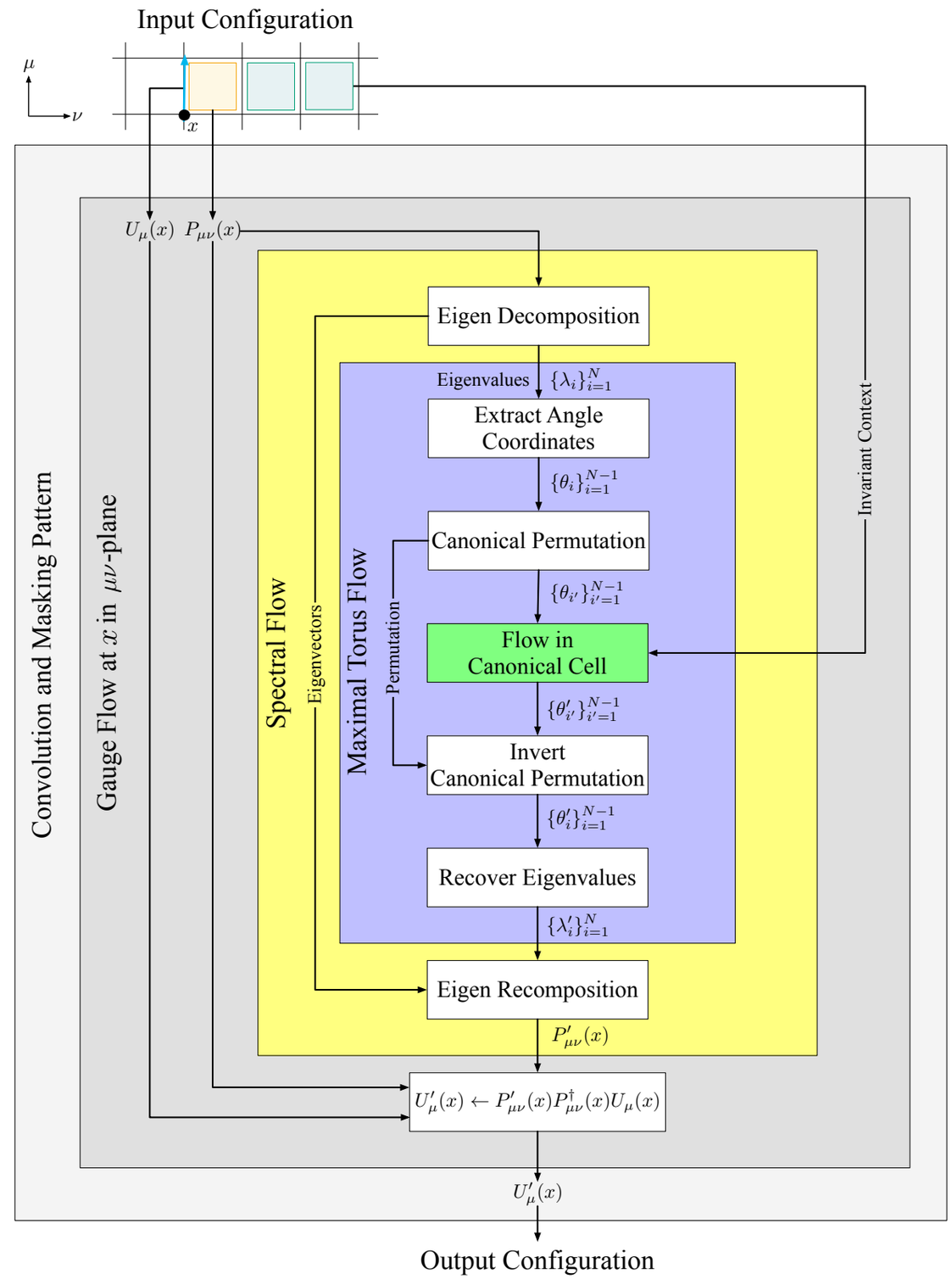




Spectral flows

- Can be applied to non-Abelian theories by stepping via eigendecomposition of (untraced) loops: transform eigenvalues while keeping eigenvectors fixed

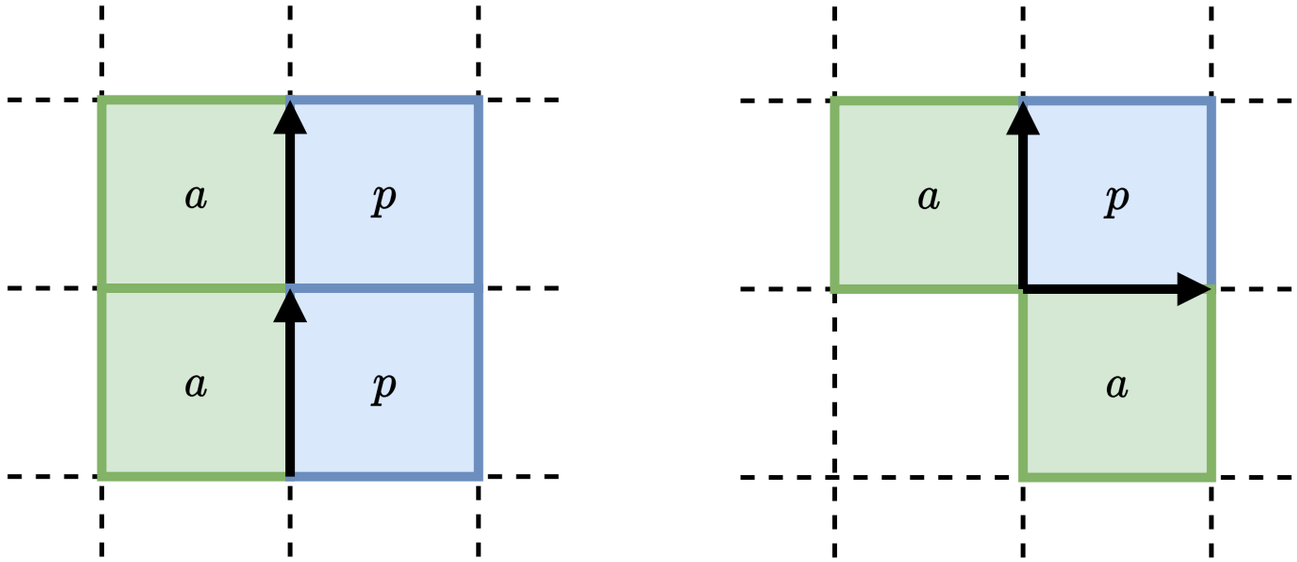




Gauge-equivariant flows

Many choices for coupling layer construction, incl:

- **Residual flows** [Abbott et al., 2305.02402]
  - Transform links $U_\mu(x)$ through parameterised "smearing" steps
  - Limited step size & harder to invert (requires fixed-point iteration)
- **Spectral flows** [Kanwar et al., 2003.06413, Boyda et al., 2008.05456]
  - Transform links via transformations of "active" plaquettes/Wilson loops (products of links starting and ending at the same lattice site)
  - Limited by "passive" transformations of plaquettes/loops (p)



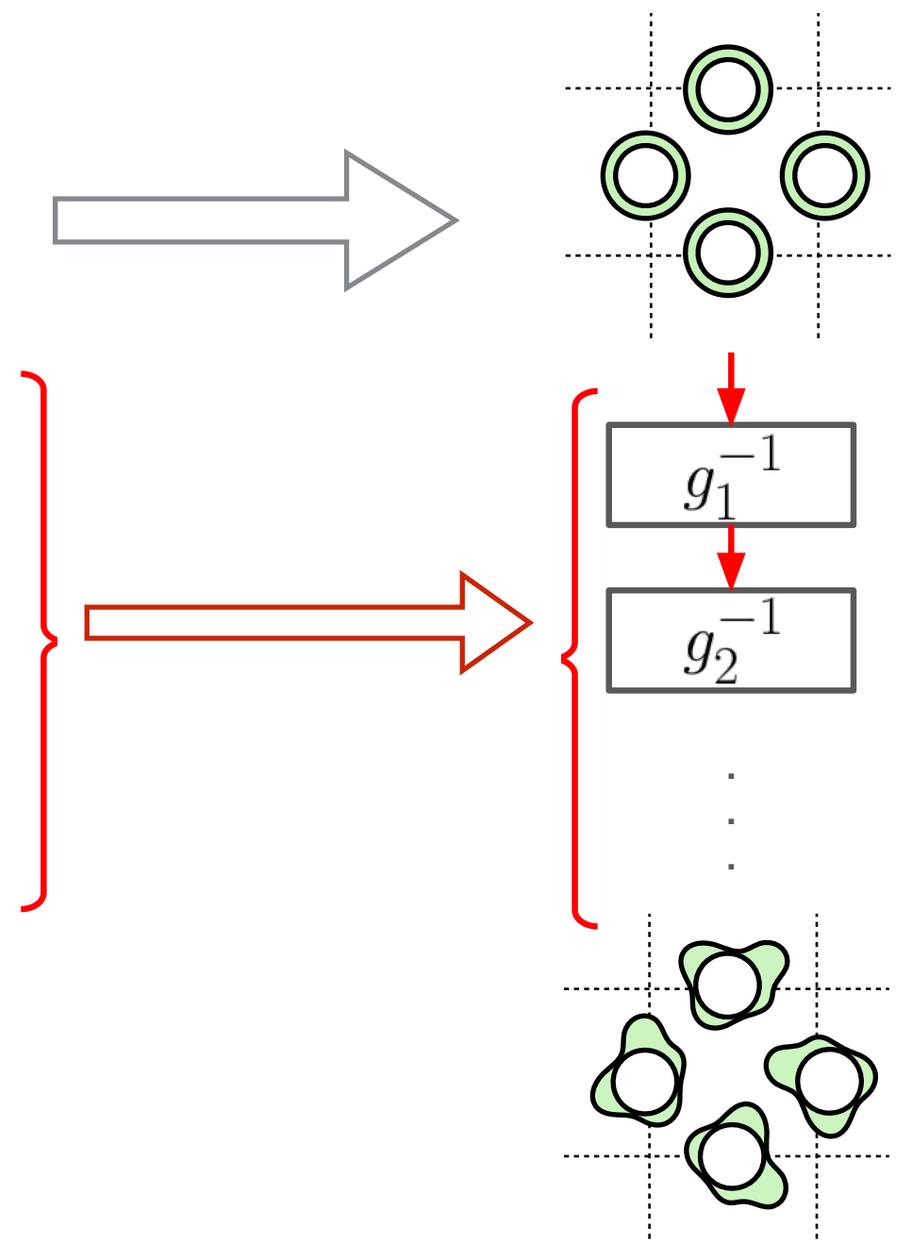
Flow models for gauge field theories

Demonstration of accelerated sampling in U(1) field theory

(2D, L=16) [Kanwar et al., 2003.06413]

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

- Prior distribution chosen to be uniform
- Gauge-equivariant spectral 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

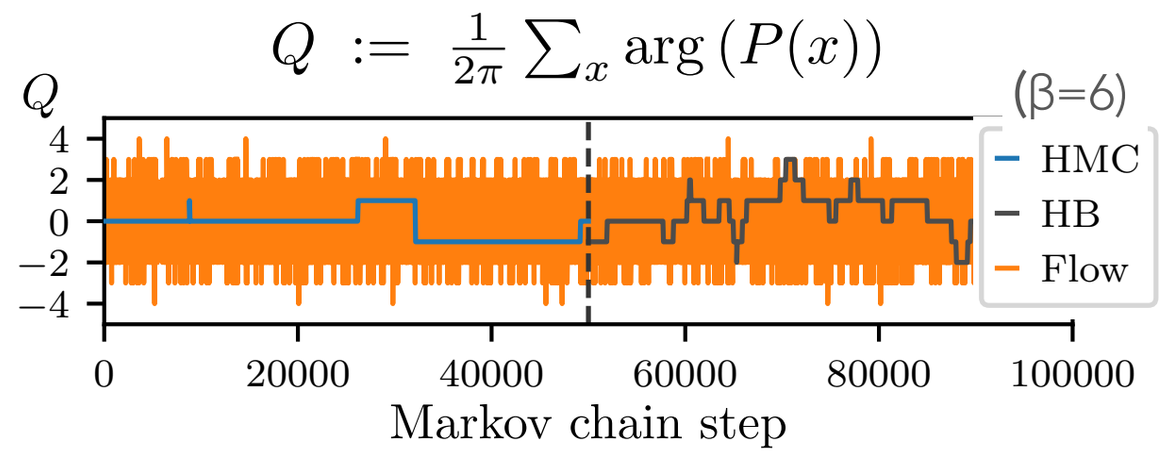
Demonstration of accelerated sampling in U(1) field theory

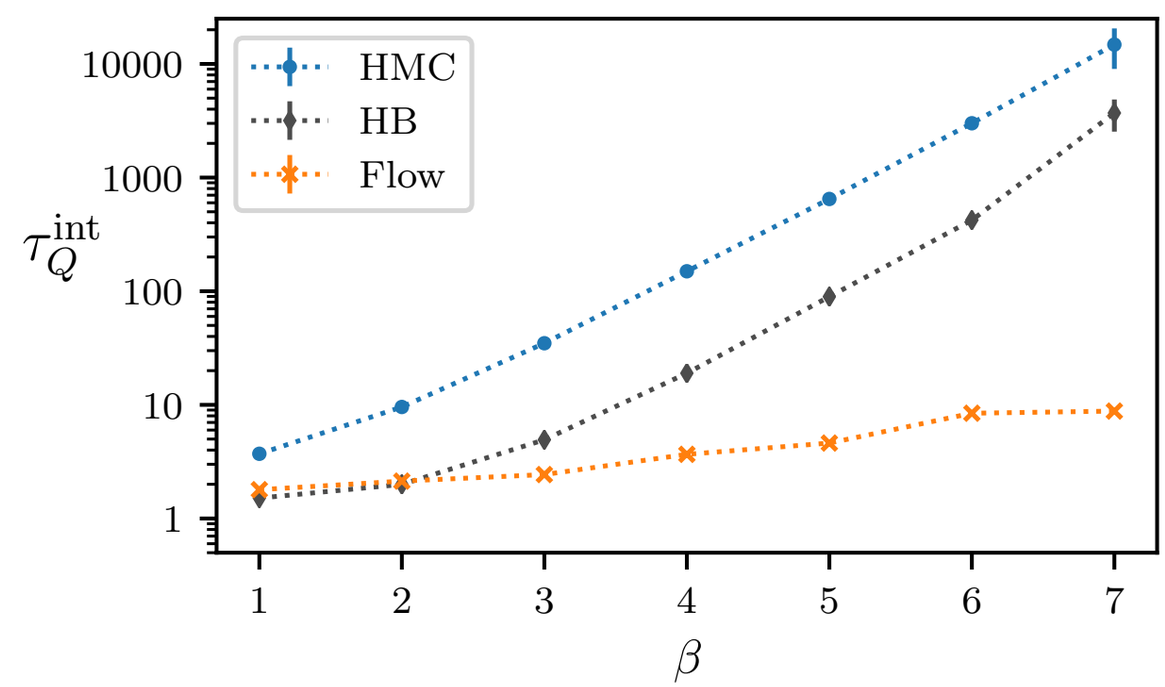
(2D, L=16) [Kanwar et al., 2003.06413]

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

- Efficient sampling of different topological sectors
- Cost of sample from flow model ~ cost of HMC trajectory
  i.e., **flow model orders of magnitude more efficient at large coupling**
- Increase in autocorrelation time in flow samples at large coupling resulting from lower model quality

  → **illustrates trade-off between sampling cost and model development/training**





Flow models for lattice QCD

- **Ongoing program** to develop flow model architectures for applications across lattice QCD

✓ First flow architectures for lattice field theory (scalar field theory) [Albergo et al., 1904.12072]

✓ Gauge field theories

- Flow transformations on compact, connected manifolds [Rezende et al., 2002.02428]
- Gauge-equivariant architectures: Abelian field theories [Kanwar et al., 2003.06413, 2101.08176]
- Gauge-equivariant architectures: non-Abelian field theories [Boyda et al., 2008.05456]

✓ Theories with fermions

- Architectures for theories with fermions [Albergo et al., 2106.05934]
- Combining architectures for gauge fields and fermions [Albergo et al., 2202.11712]
- Techniques to incorporate pseudofermions [Abbott et al., 2207.08945]

✓ Initial application to QCD in 4D

[This talk+upcoming manuscripts on scaling and 4D]

🏁 Architectures for QCD at scale [ongoing; Aurora Early Science Project]

AURORA | EARLY SCIENCE PROGRAM FOR DATA & LEARNIN

[see also tutorial notebook 2101.08176, work on multimodal distributions 2107.00734]

Flow models with fermions

Need new architectures to efficiently handle theories with fermions

[Albergo et al., 2106.05934; Albergo et al., 2202.11712]

- Integrating out fermions

  → expensive fermion determinant

$$S_E(U) = -\beta \sum \underbrace{\mathrm{Re}\, P(x)}_{\text{(Plaquette)}} - \underbrace{\log \det D[U]^\dagger D[U]}_{\text{(Fermion determinant)}}$$

Flow models with fermions

Need new architectures to efficiently handle theories with fermions
[Albergo et al., 2106.05934; Albergo et al., 2202.11712]

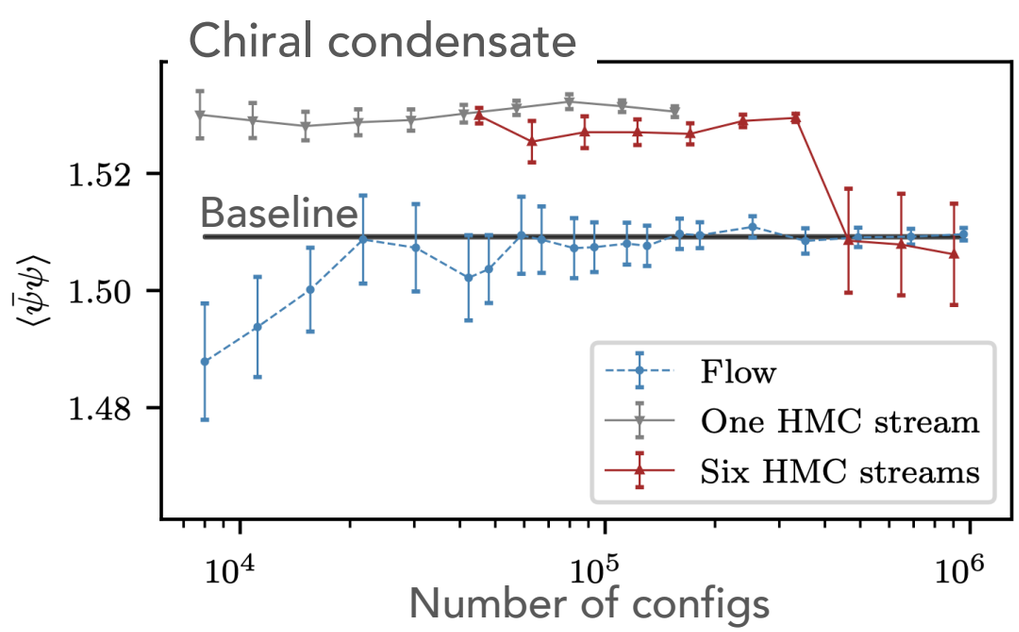
- Integrating out fermions

  → expensive fermion determinant

$$S_E(U) = -\beta \sum \underbrace{\mathrm{Re}\, P(x)}_{\text{(Plaquette)}} - \underbrace{\log\det D[U]^\dagger D[U]}_{\text{(Fermion determinant)}}$$

1. Flows with exact determinant evaluation work [Albergo et al., 2202.11712]

- Existing gauge-equivariant architectures
- Application to Schwinger model at near-critical parameters
  [2D, N_f=2, β=2, L=16, κ=0.276]
  - ✳ HMC biased with underestimated errors
  - ✳ Flow-based sampling gives correct results and error estimates



Flow models with fermions

Need new architectures to efficiently handle theories with fermions

[Albergo et al., 2106.05934; Albergo et al., 2202.11712]

- Integrating out fermions

  → expensive fermion determinant

$$S_E(U) = -\beta \sum \mathrm{Re}\, P(x) - \log\det D[U]^\dagger D[U]$$

(Plaquette) (Fermion determinant)

2. Scalable approach (needs new arch.): stochastic determinant estimators

- ✳ Evaluate determinant using auxiliary (pseudofermion) degrees of freedom

$$\det DD^\dagger = \frac{1}{Z_{\mathcal{N}}} \int \mathcal{D}\phi\, e^{-\phi^\dagger (DD^\dagger)^{-1} \phi}$$

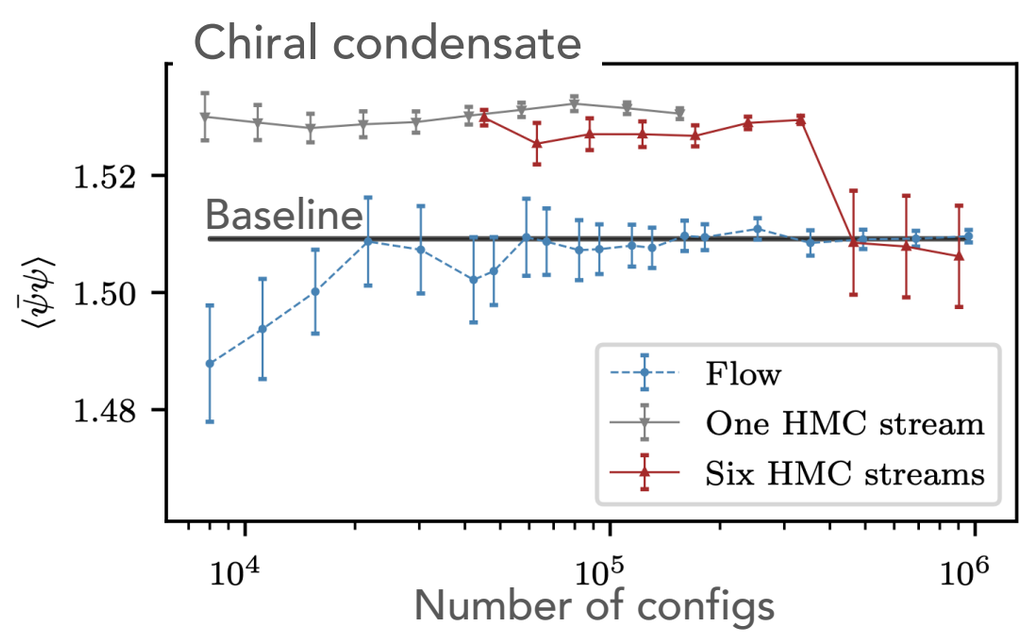
Pseudofermions

1. Flows with exact determinant evaluation work [Albergo et al., 2202.11712]

- Existing gauge-equivariant architectures
- Application to Schwinger model at near-critical parameters

  [2D, N_f=2, β=2, L=16, κ=0.276]

  - ✳ HMC biased with underestimated errors
  - ✳ Flow-based sampling gives correct results and error estimates



Flow models with fermions

Joint architectures for gauge and pseudofermion fields

[Abbott et al., 2207.08945]

$$p(U, \phi) = p(U)p(\phi \mid U)$$

$$p(U) \propto \det DD^\dagger(U) e^{-S_g(U)}$$

"marginal"

$$p(\phi|U) \propto \frac{1}{\det DD^\dagger(U)} e^{-S_{PF}(\phi|U)}$$

"conditional"

Different flow models to approximate marginal and conditional distributions

Flow models with fermions

Joint architectures for gauge and pseudofermion fields

[Abbott et al., 2207.08945]

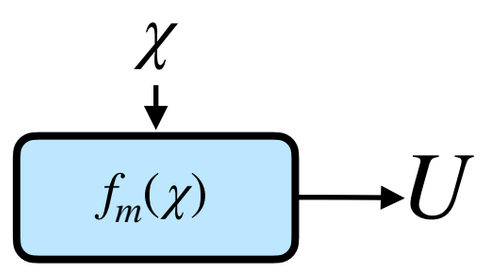
$$p(U, \phi) = p(U)p(\phi \mid U)$$

$$p(U) \propto \det DD^\dagger(U) e^{-S_g(U)}$$

"marginal"

$$p(\phi|U) \propto \frac{1}{\det DD^\dagger(U)} e^{-S_{PF}(\phi|U)}$$

"conditional"



Same gauge architectures
as for pure gauge theories

[Kanwar et al., 2003.06413, Boyda et al., 2008.05456]

Different flow models to approximate marginal and conditional distributions

Flow models with fermions

Joint architectures for gauge and pseudofermion fields

[Abbott et al., 2207.08945]

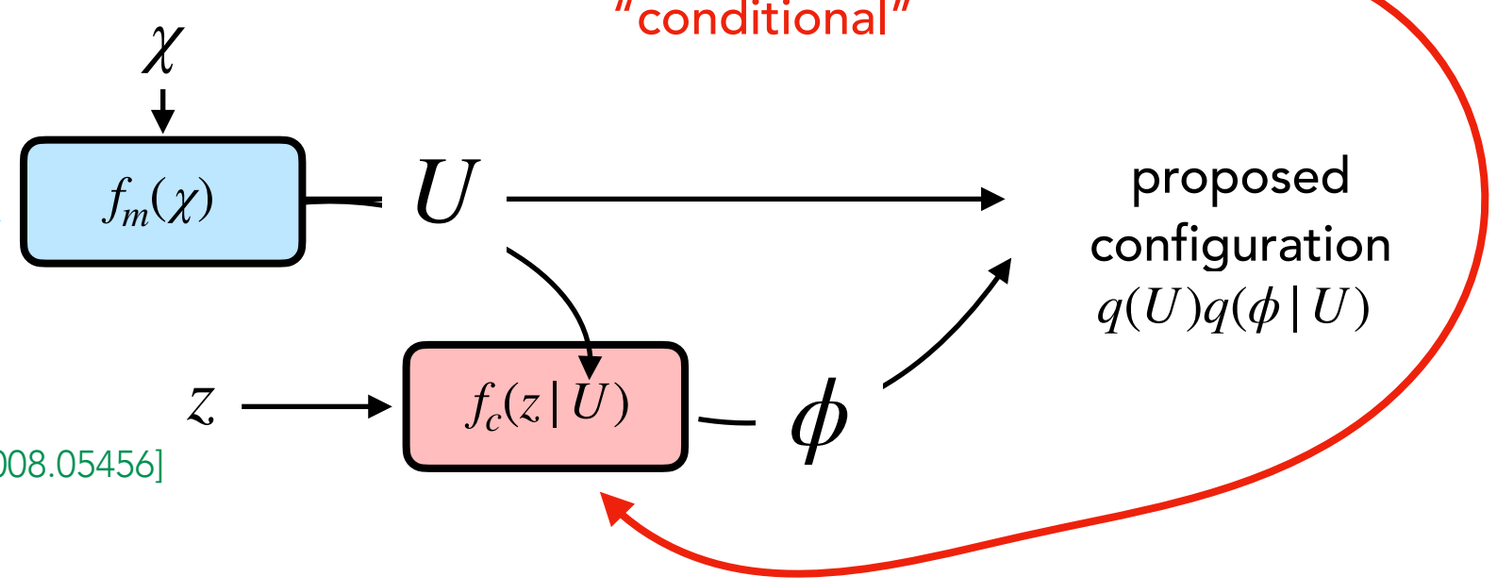
$$p(U, \phi) = p(U)p(\phi \mid U)$$

$$p(U) \propto \det DD^\dagger(U) e^{-S_g(U)}$$

"marginal"

$$p(\phi|U) \propto \frac{1}{\det DD^\dagger(U)} e^{-S_{PF}(\phi|U)}$$

"conditional"



Same gauge architectures as for pure gauge theories

[Kanwar et al., 2003.06413, Boyda et al., 2008.05456]

New conditional architectures

[Abbott et al., 2207.08945]

Different flow models to approximate marginal and conditional distributions

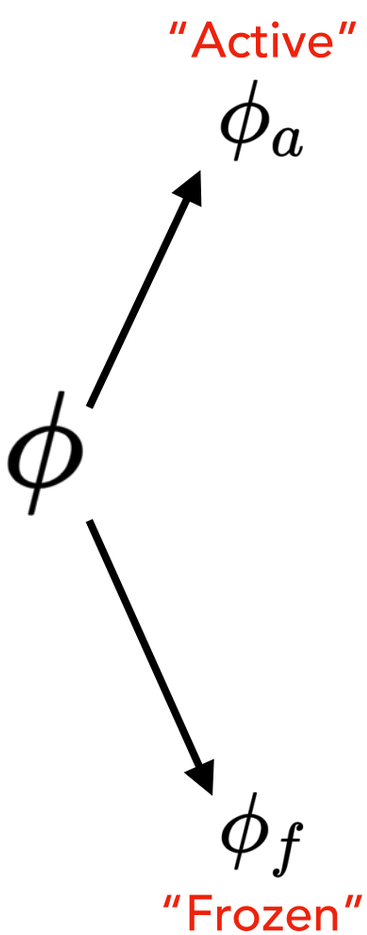
Flow models with fermions

Conditional model maps uncorrelated Gaussian to correlated Gaussian

[Abbott et al., 2207.08945]

$$r(z) \propto e^{-z^\dagger z} \xrightarrow{f_c(z\,|\,U)} \boxed{q(\phi\,|\,U)} \propto e^{-\phi^\dagger A(U)\phi} \simeq e^{-\phi^\dagger \left(D(U)D^\dagger(U)\right)^{-1}\phi}$$

Approximates



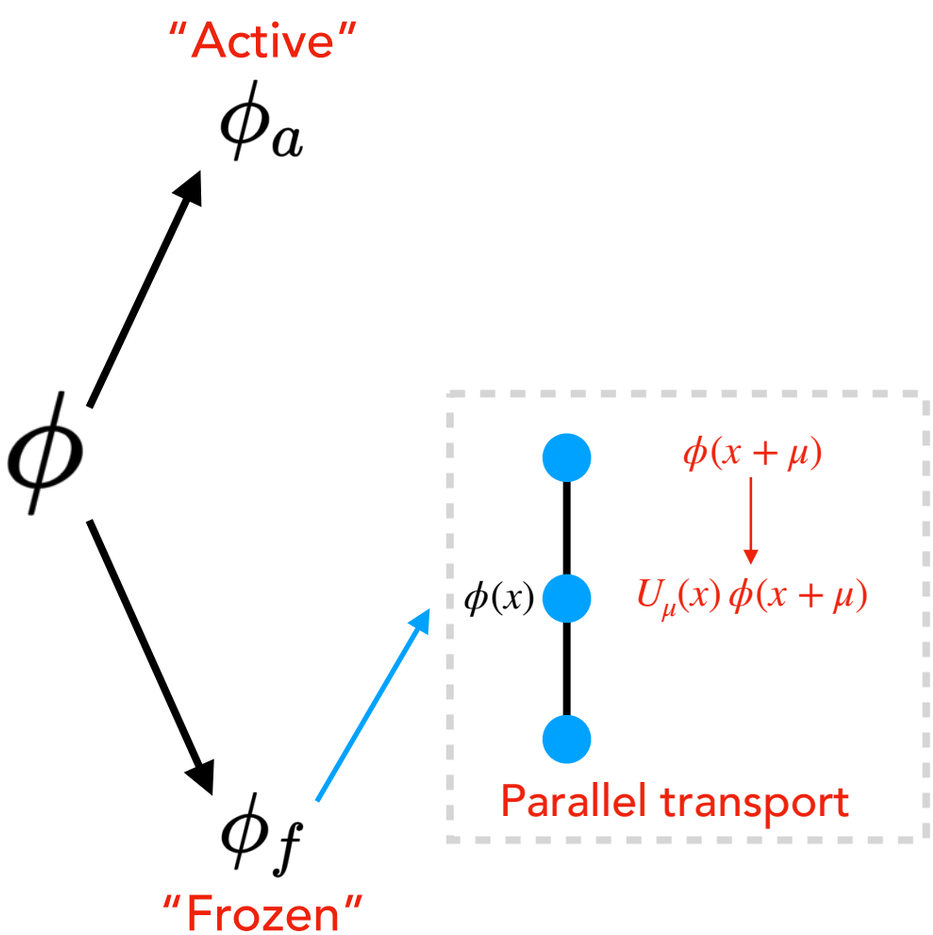
Flow models with fermions

Conditional model maps uncorrelated Gaussian to correlated Gaussian

[Abbott et al., 2207.08945]

$$r(z) \propto e^{-z^\dagger z} \xrightarrow{f_c(z\,|\,U)} \boxed{q(\phi\,|\,U)} \propto e^{-\phi^\dagger A(U)\phi} \simeq e^{-\phi^\dagger \left(D(U)D^\dagger(U)\right)^{-1}\phi}$$

Approximates



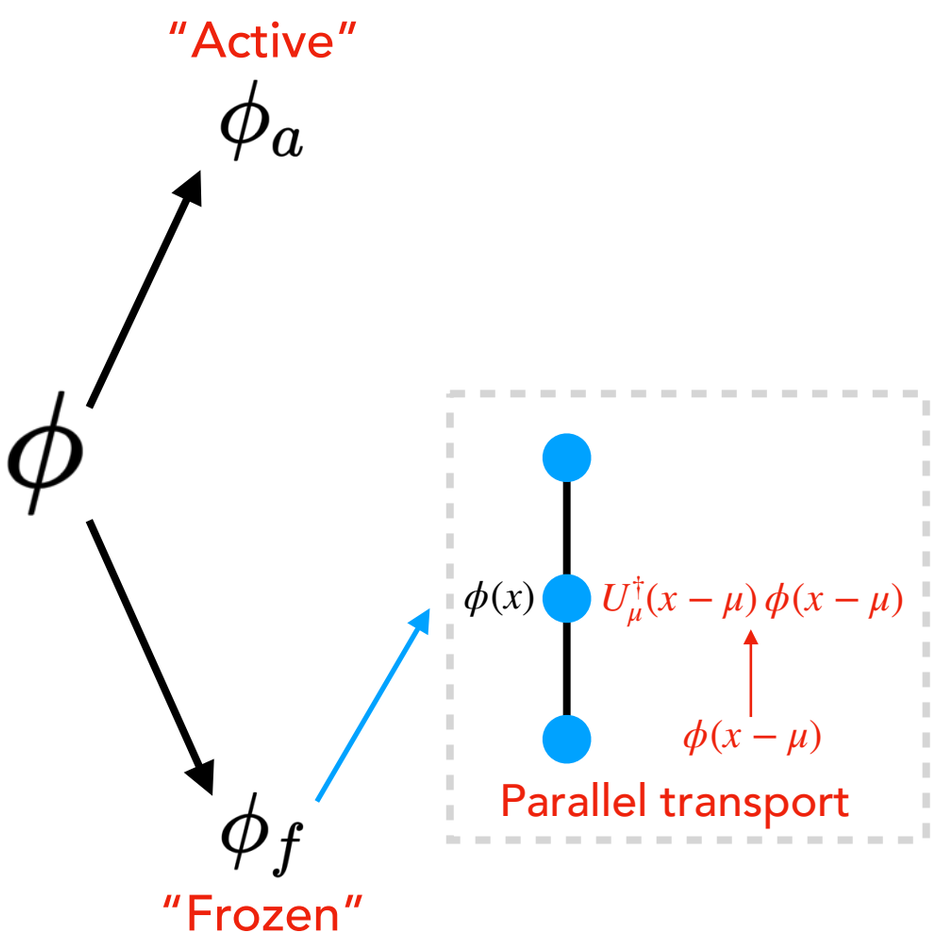
Flow models with fermions

Conditional model maps uncorrelated Gaussian to correlated Gaussian

[Abbott et al., 2207.08945]

$$r(z) \propto e^{-z^\dagger z} \xrightarrow{f_c(z\,|\,U)} \boxed{q(\phi\,|\,U)} \propto e^{-\phi^\dagger A(U)\phi} \simeq e^{-\phi^\dagger \left(D(U)D^\dagger(U)\right)^{-1}\phi}$$

Approximates



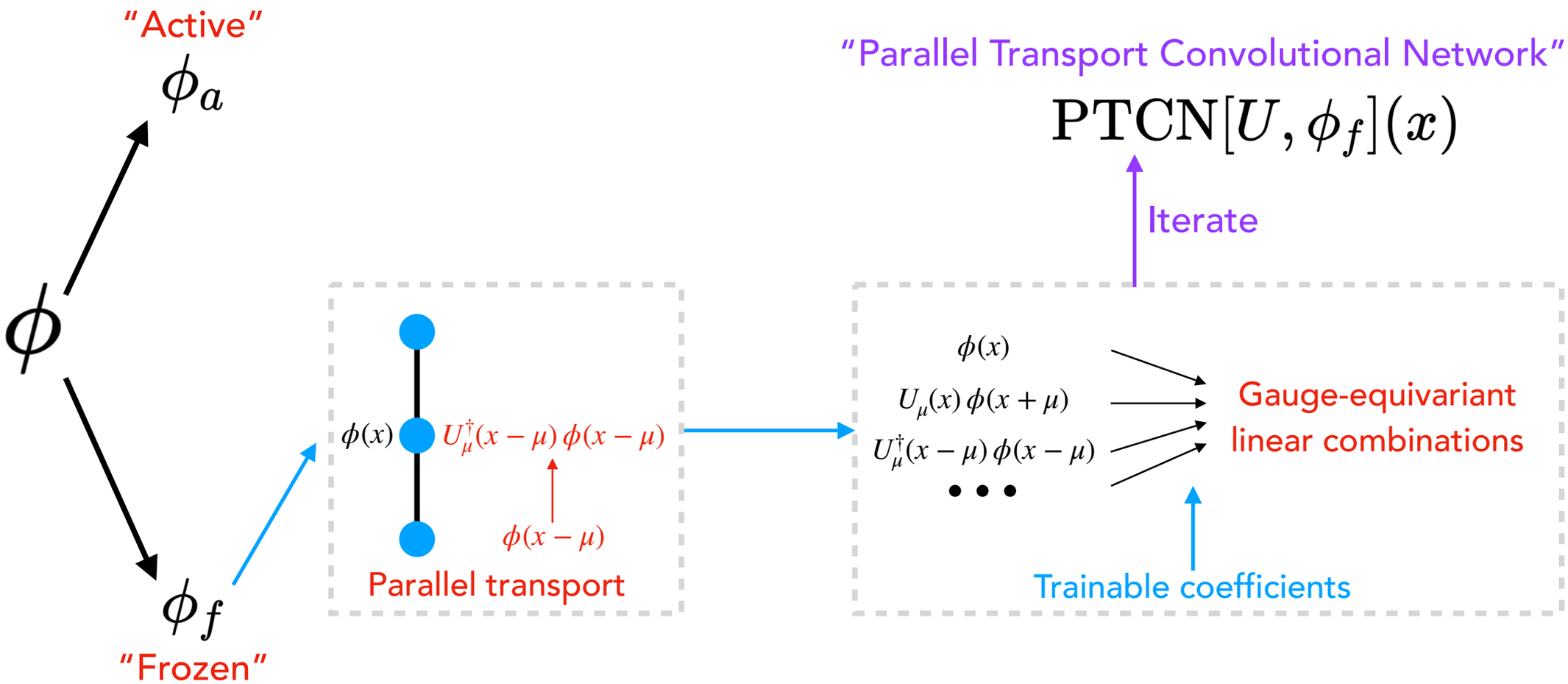
Flow models with fermions

Conditional model maps uncorrelated Gaussian to correlated Gaussian

[Abbott et al., 2207.08945]

$$r(z) \propto e^{-z^\dagger z} \xrightarrow{f_c(z\,|\,U)} \boxed{q(\phi\,|\,U)} \propto e^{-\phi^\dagger A(U)\phi} \simeq e^{-\phi^\dagger \left(D(U)D^\dagger(U)\right)^{-1}\phi}$$

Approximates



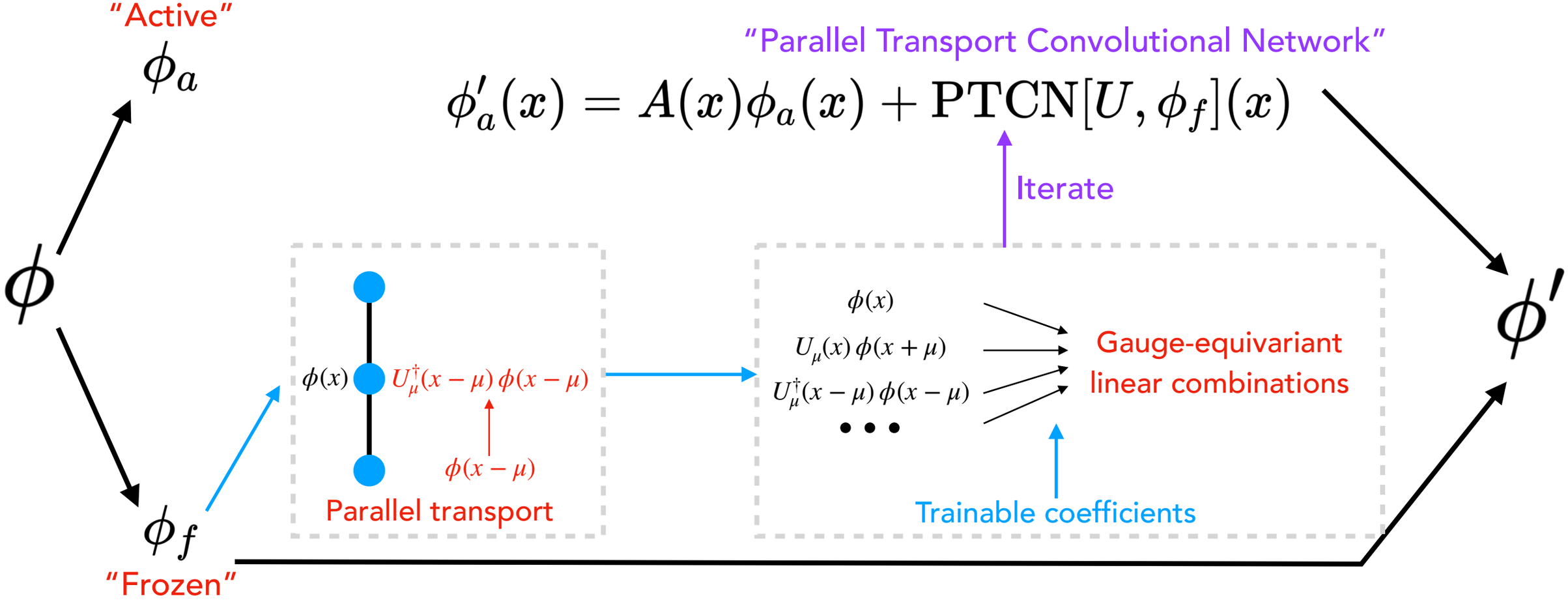
Flow models with fermions

Conditional model maps uncorrelated Gaussian to correlated Gaussian

[Abbott et al., 2207.08945]

$$r(z) \propto e^{-z^\dagger z} \xrightarrow{f_c(z\,|\,U)} \boxed{q(\phi\,|\,U)} \propto e^{-\phi^\dagger A(U)\phi} \simeq e^{-\phi^\dagger \left(D(U)D^\dagger(U)\right)^{-1}\phi}$$

Approximates



Flow models for lattice QCD

- **Ongoing program** to develop flow model architectures for applications across lattice QCD

✓ First flow architectures for lattice field theory (scalar field theory) [Albergo et al., 1904.12072]

✓ Gauge field theories

- Flow transformations on compact, connected manifolds [Rezende et al., 2002.02428]
- Gauge-equivariant architectures: Abelian field theories [Kanwar et al., 2003.06413, 2101.08176]
- Gauge-equivariant architectures: non-Abelian field theories [Boyda et al., 2008.05456]

✓ Theories with fermions

- Architectures for theories with fermions [Albergo et al., 2106.05934]
- Combining architectures for gauge fields and fermions [Albergo et al., 2202.11712]
- Techniques to incorporate pseudofermions [Abbott et al., 2207.08945]

✓ Initial application to QCD in 4D
[Abbott et al., 2208.03832, 2211.07541]

🏁 Architectures for QCD at scale [ongoing; Aurora Early Science Project]

AURORA | EARLY SCIENCE PROGRAM FOR DATA & LEARNIN

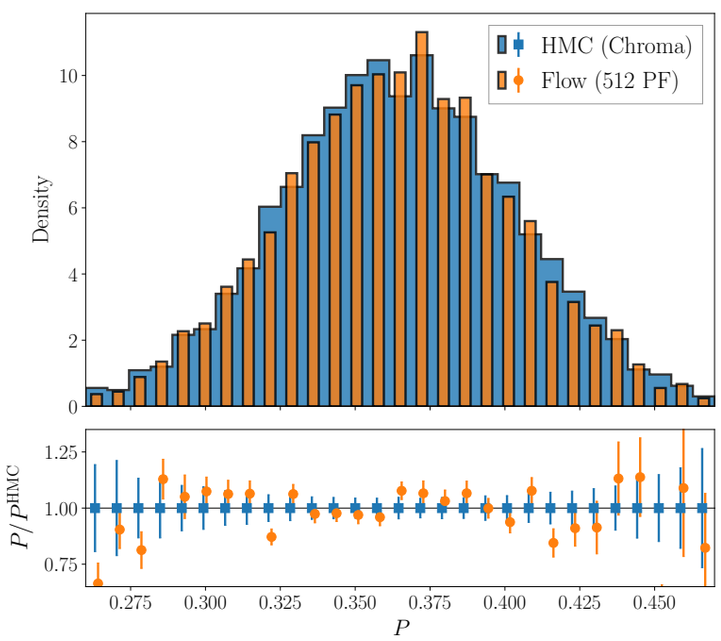
[see also tutorial notebook 2101.08176, work on multimodal distributions 2107.00734]

Flow models for lattice QCD

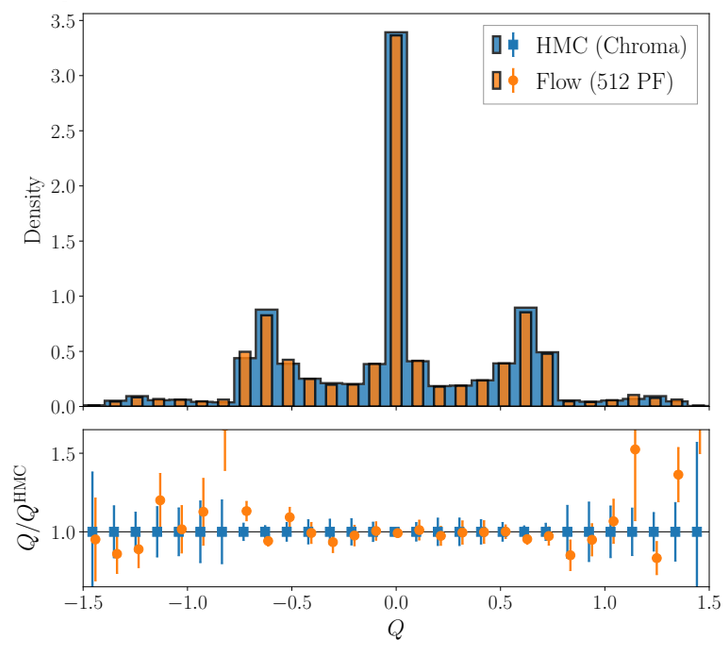
Initial QCD demonstration [Abbot et al., 2208.03832, 2211.07541, 2305.02402]

- Direct combination of published results on gauge-equivariant flows and pseudofermions [Boyda et al., 2008.05456, Abbott et al., 2207.08945]
- Illustration at straightforward parameters $V=4^4$, $N_f=2$, $\beta=1$, $\kappa=0.1$
- Observables from flow ensemble in precise agreement with HMC at high statistics (65k samples)
- Development and scaling of QCD-specific architectures in full swing — stay tuned!

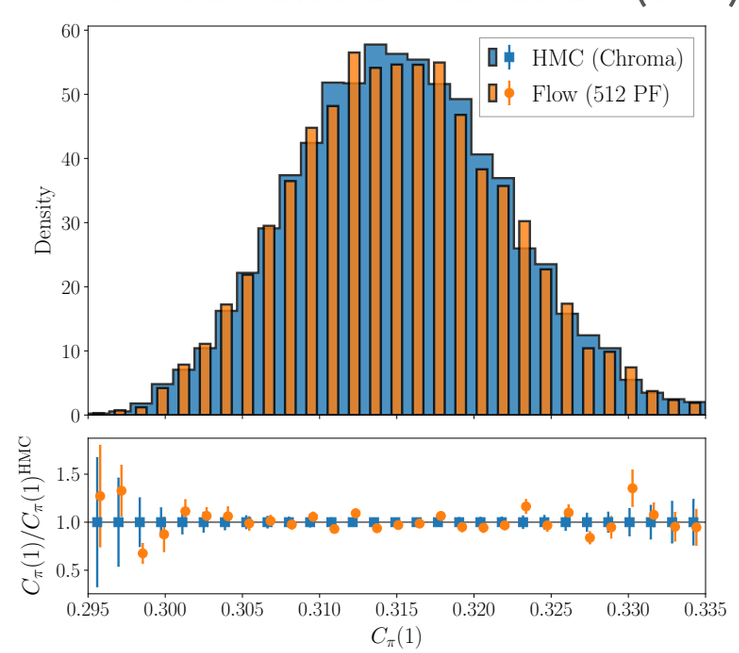
Plaquette



Topological charge (Wilson flowed)



Pion correlation function (t=1)

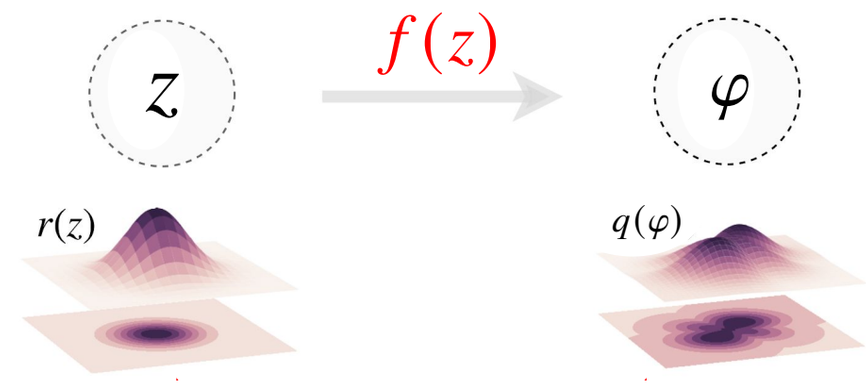


Outlook: Flow models for lattice QCD

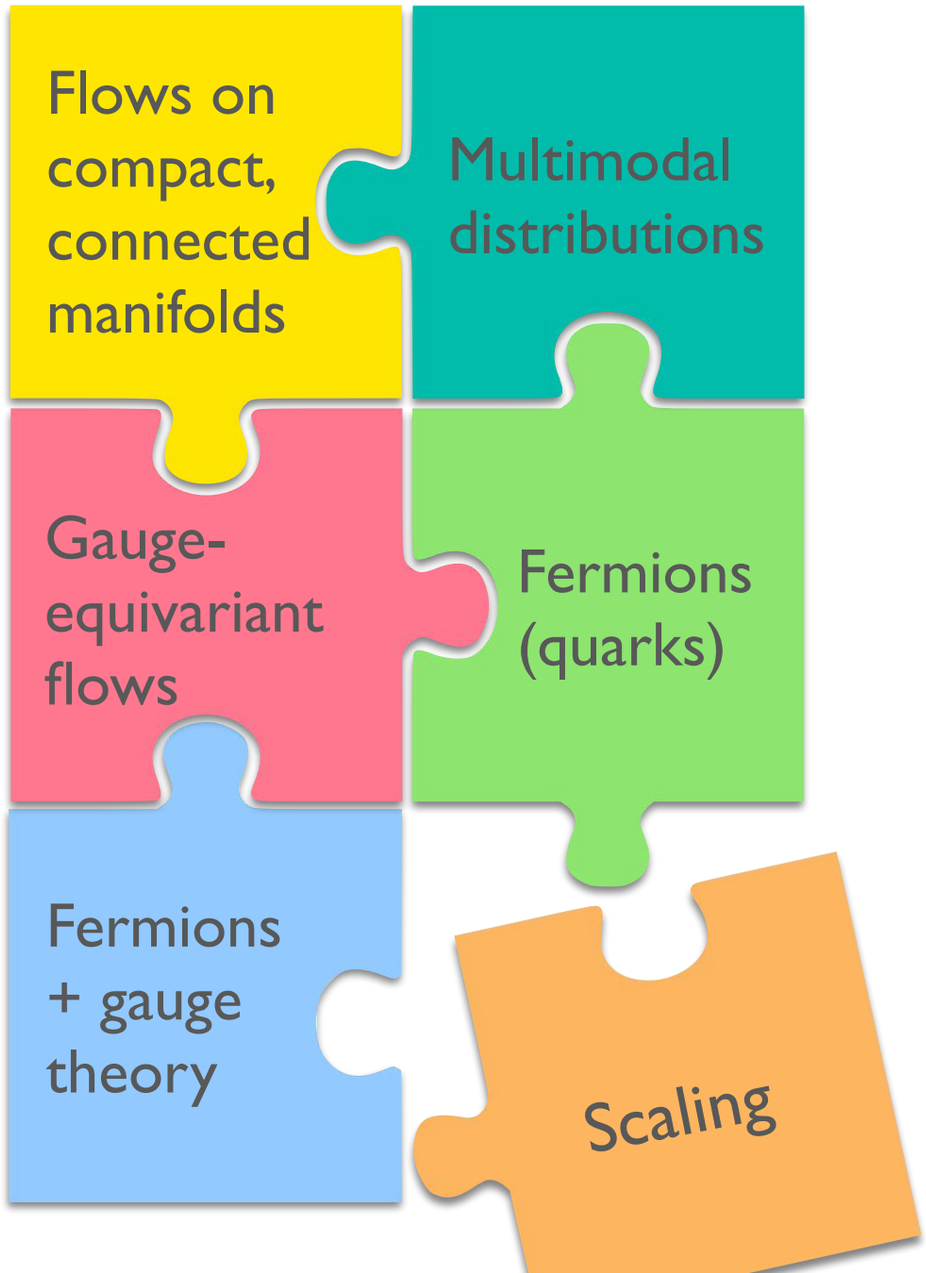
All fundamental components in place to begin exploration of flow models for lattice QCD!

Significant efforts still required to exploit potential of flow models for lattice QCD

- QCD-specific engineering and development only just beginning
- Scaling to state-of-the-art requires engineering custom ML architectures to similar scale as largest industrial ML models



Outlook: Flow models for lattice QCD



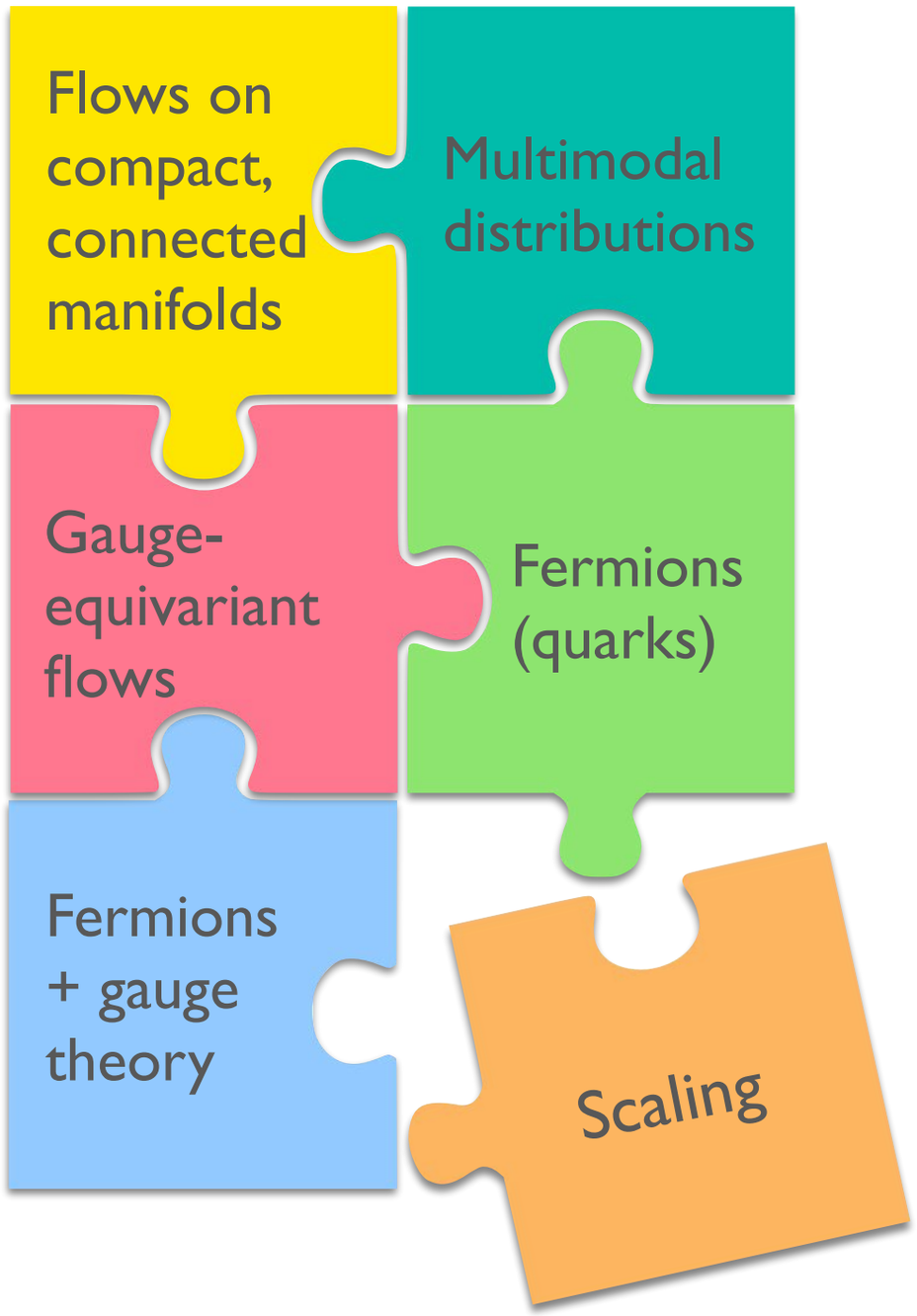
Machine learning for QCD

- **Provably-exact** machine-learning-accelerated sampling algorithm
- Orders of magnitude more **efficient** than conventional algorithms overcoming critical slowing-down
- **Unbiased** results where traditional approaches fail

Deployment for state-of-the-art QCD scheduled for Aurora 2023 first science time

[2107.00734; 2101.08176; *Phys.Rev.D* 104, 114507; *Phys.Rev.D* 103, 074504 (2021); *Phys.Rev.Lett.* 125, 121601; PMLR 8083-8092 (2020); *Phys.Rev.D* 100, 034515 (2019); *Phys.Rev.D* 97, 094506 (2018)]

Outlook: Flow models for lattice QCD



- Costs of fine lattice spacings contribute to systematic limitations for high-precision QCD predictions, e.g.,
  - first-principles proton structure
  - muon g-2
- Finer spacings needed for first controlled calculations of nuclei
  - theory input for intensity-frontier experiments, e.g., dark matter direct detection, DUNE
  - nuclear reactions incl. big bang nucleosynthesis pathway

Machine learning for first-principles theory

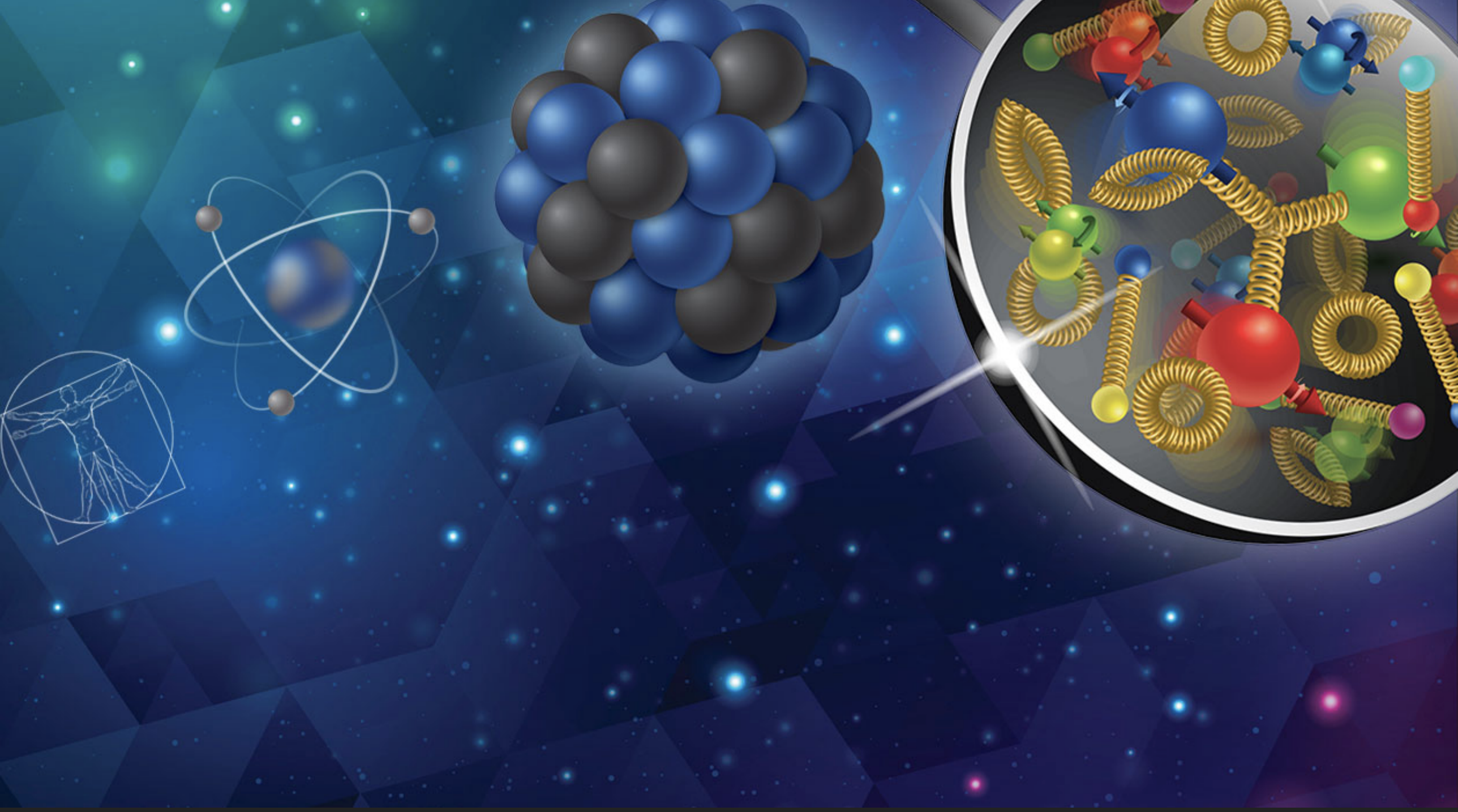
Compute **exact** results from known theory
Use AI/ML to do it **faster**

e.g., lattice QCD calculations, EFT studies, many-body approaches, …

Require **mathematical guarantees of exactness** to preserve rigour of first-principles calculations

Potentially high-impact applications in development across theory areas, often require significant investment

- Developing new ML approaches from the ground up
- Engineering: scaling up to state-of-the-art HPC facilities



Massachusetts Institute of Technology