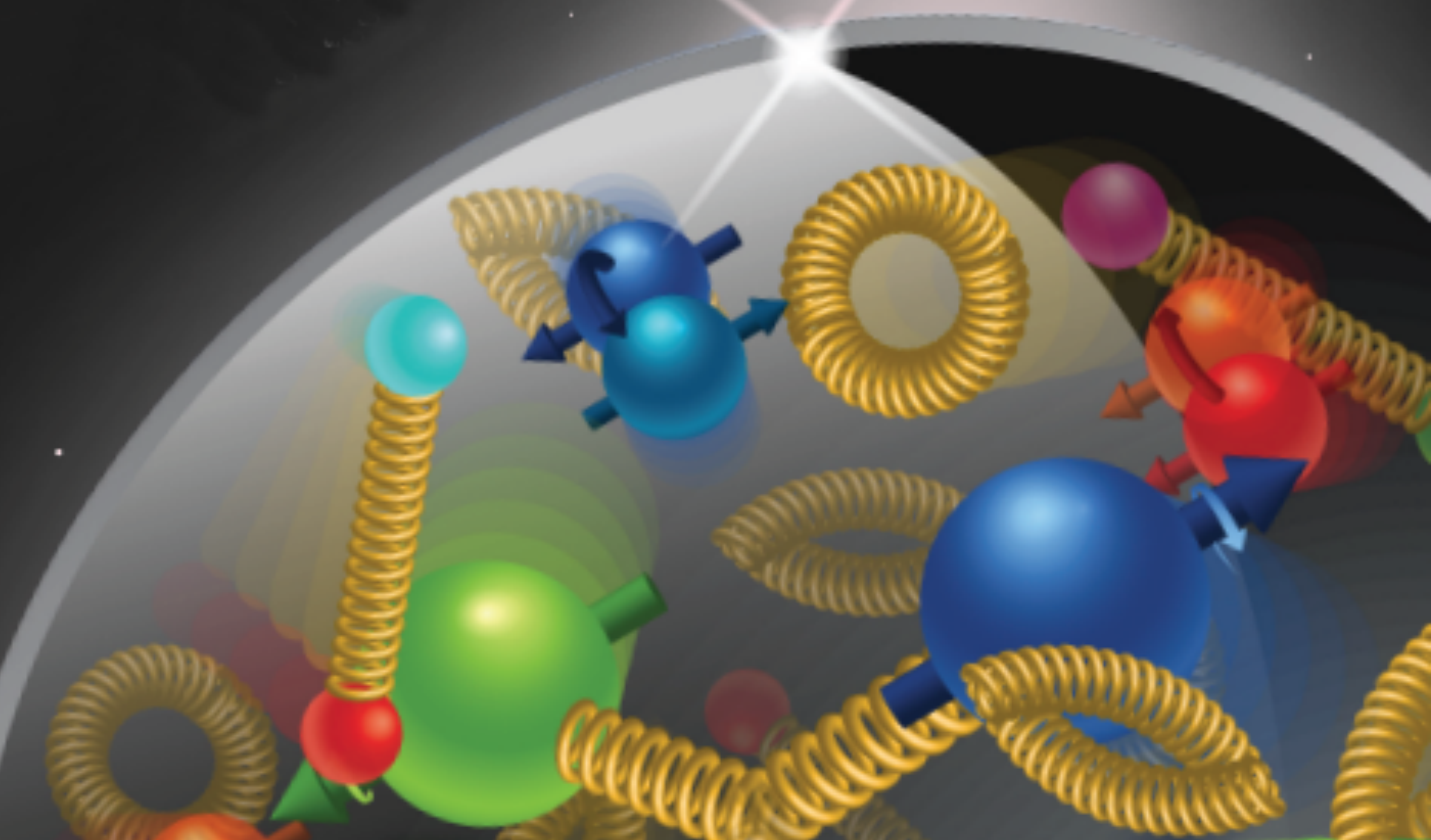
PHIALA SHANAHAN

FLOW MODELS FOR SAMPLING IN LATTICE QCD

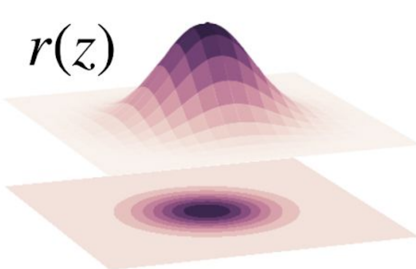


Massachusetts Institute of Technology

Flow models for lattice QCD

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

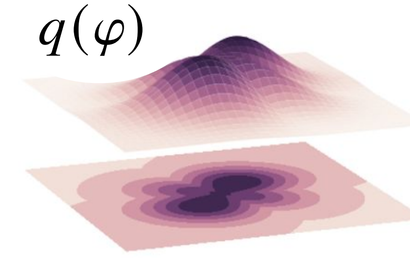
z $\xrightarrow{f(z)}$ φ



"Base" distribution

Flow transformation

- Invertible, tractable Jacobian
- Many free parameters for optimization



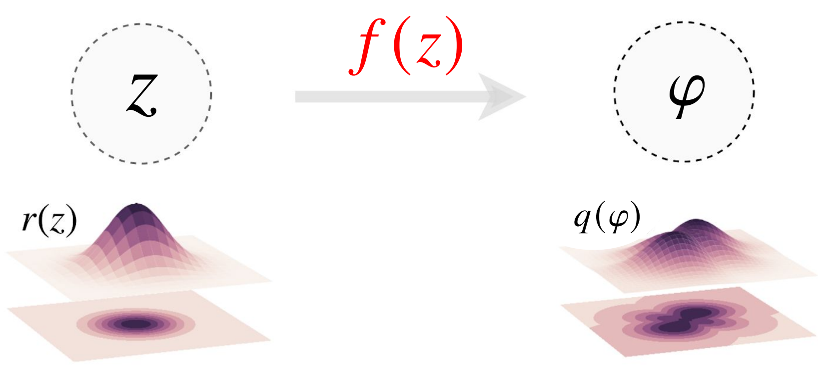
"Model" distribution

$$q(\varphi) = r(z)|\det \partial f(z)/\partial z|^{-1}$$

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

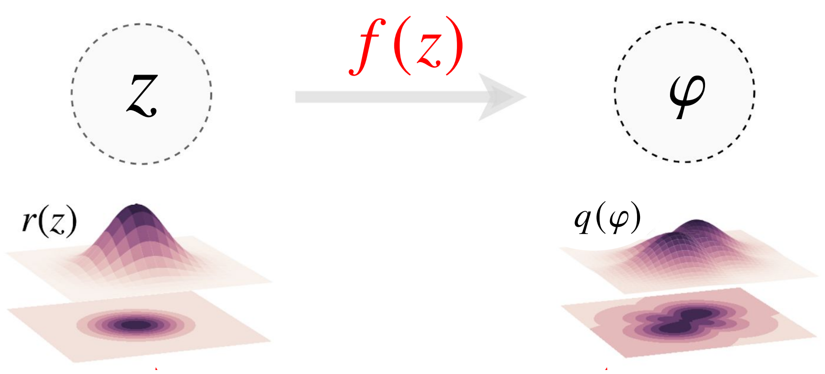
$$q(\phi) \approx \frac{1}{Z} e^{-S(\phi)}$$

- 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

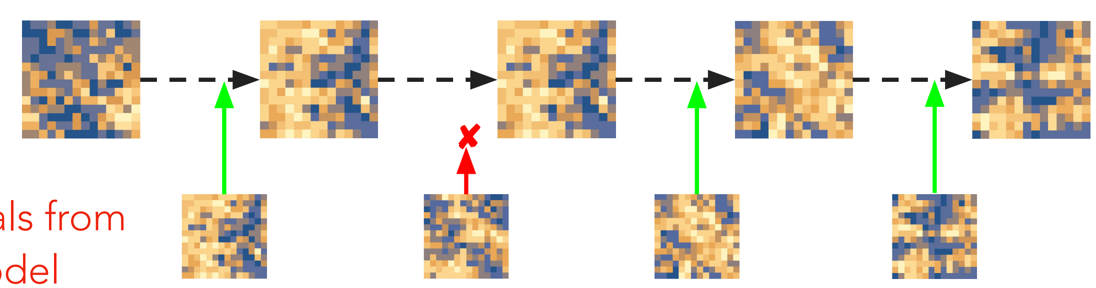
- **Train the model:**
Gradient descent to minimise "loss function" with minimum at $q(\phi) = \frac{1}{Z} e^{-S(\phi)}$

$$L(q) = \int d\phi \, q(\phi) [\log q(\phi) + S(\phi)]$$

Estimate stochastically by sampling from the model, i.e., "self training"

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

Markov chain

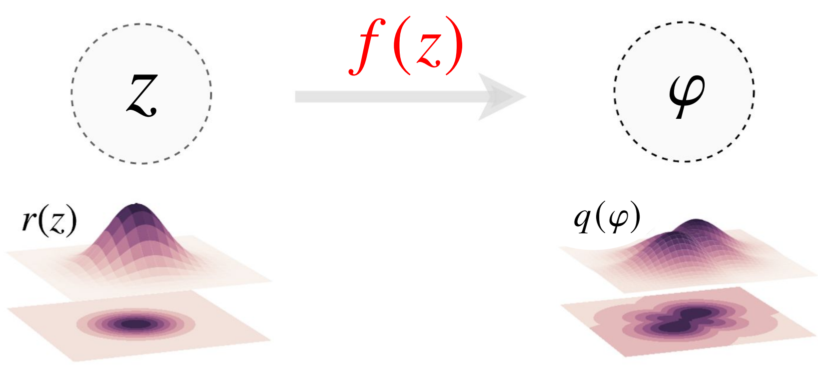


Proposals from flow model

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**

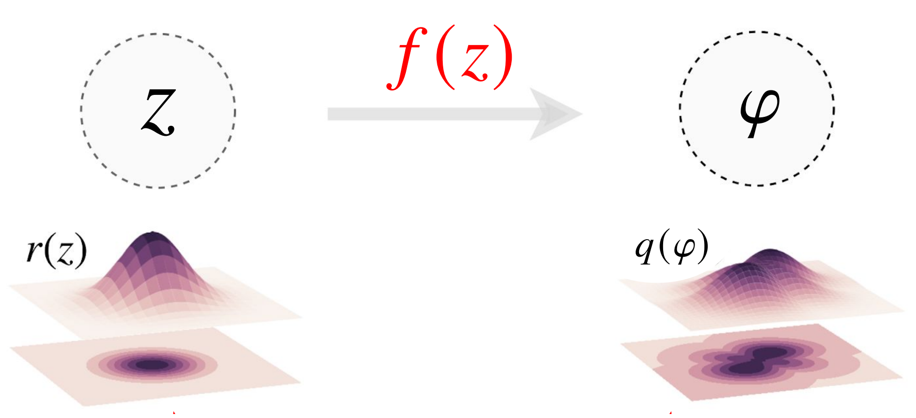
Direct sampling is only one of many approaches to using flow models for lattice QCD!

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

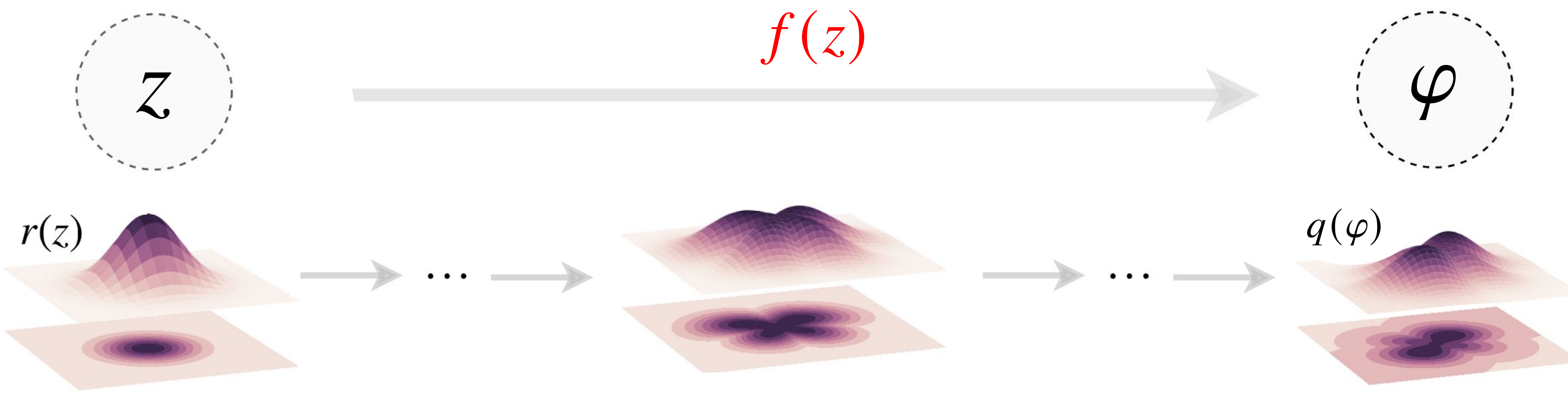
- Direct sampling i.e., $r(z)$ is a trivial distribution and $q(\phi) \approx \frac{1}{Z} e^{-S(\phi)}$
  Generalisation of [Lüscher 0907.5491]
- Hybrid sampling approaches
  e.g.,generalize the proposal distribution in HMC [Foreman et al., 2112.01582]; flows in lattice subdomains [Finkenrath 2201.02216]
- Map from one action/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

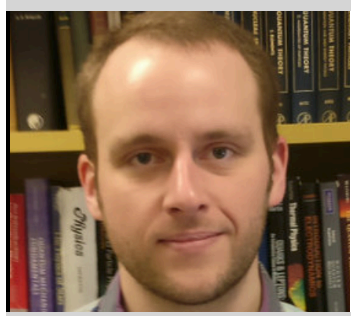
- Diffeomorphisms on lattice field degrees of freedom
- Encode symmetries, e.g., gauge symmetry
- Flexible/expressive/can encode correlations at physically-relevant scale etc

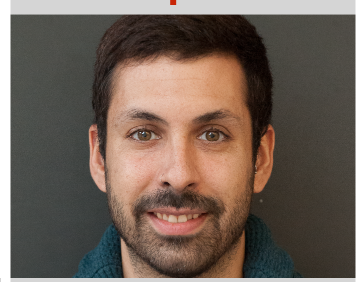
Flow models for lattice QCD

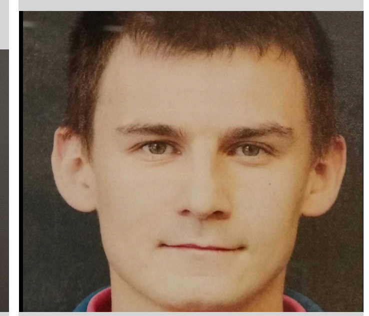
- **Ongoing program** to develop flow model architectures for applications across lattice QCD
- **Long-term industry collaboration** w/ Google DeepMind

Abbott

Massachusetts Institute of Technology

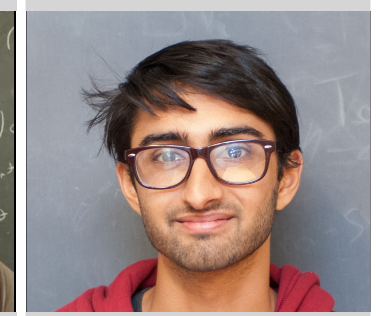
Hackett

Massachusetts Institute of Technology

Romero-López

Massachusetts Institute of Technology

Boyda

Argonne National Laboratory → MIT

Urban

Massachusetts Institute of Technology

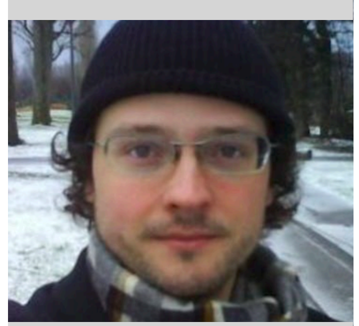
Kanwar

MIT → u^b BERN

Botev

DeepMind

Matthews

DeepMind

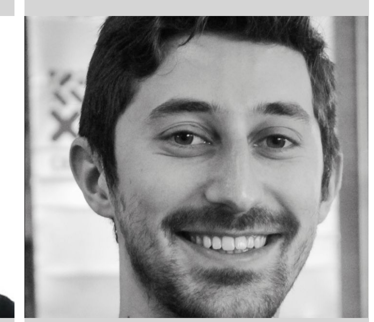
Rezende

DeepMind

Racanière

DeepMind

Razavi

DeepMind

Albergo

NYU

Cranmer

WISCONSIN University of Wisconsin–Madison

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]

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]

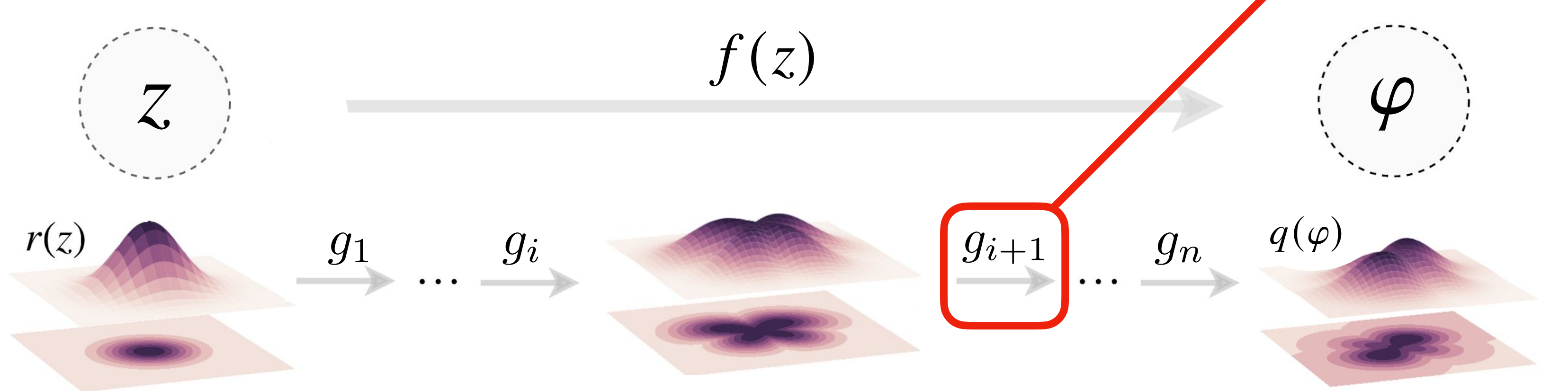
Architectures for QCD at scale [ongoing; Aurora Early Science Project]

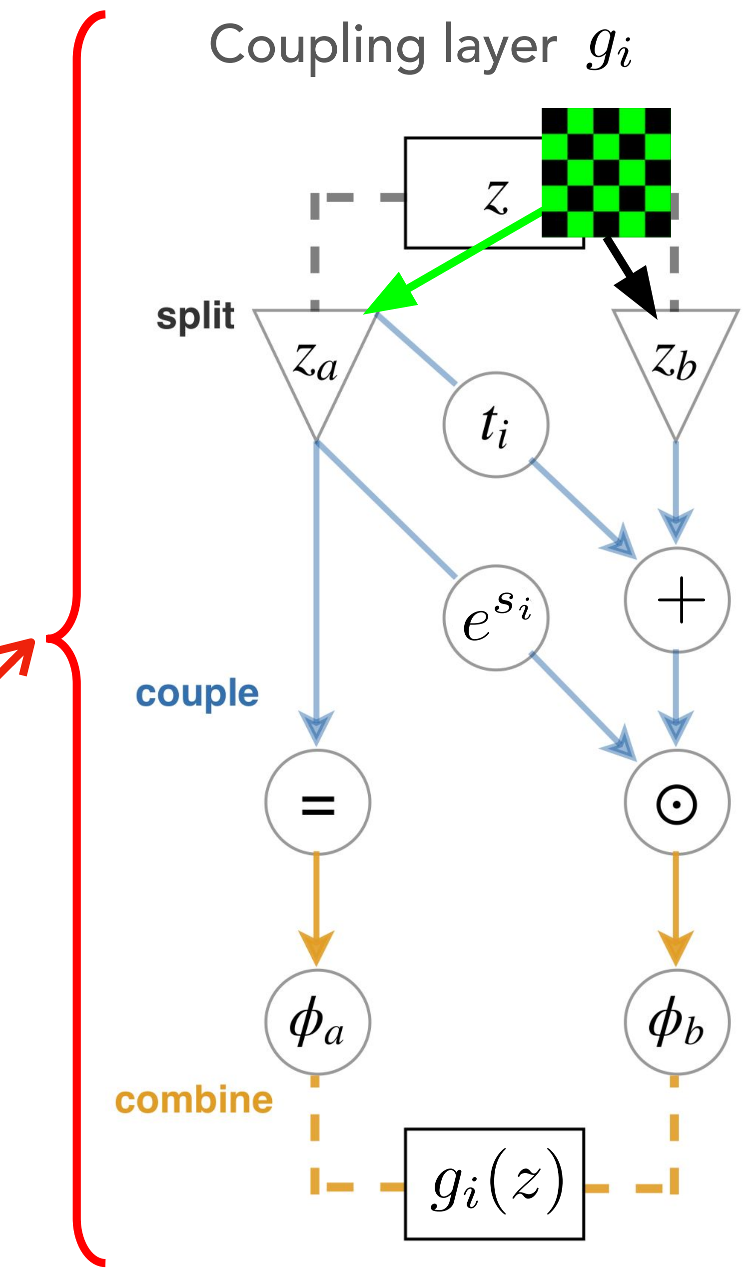
AURORA | EARLY SCIENCE PROGRAM FOR DATA & LEARNIN

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

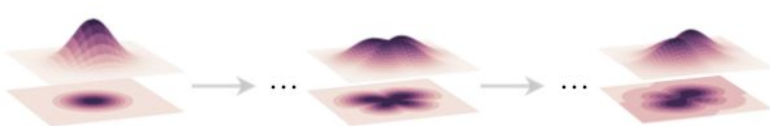
Flow models for scalar fields

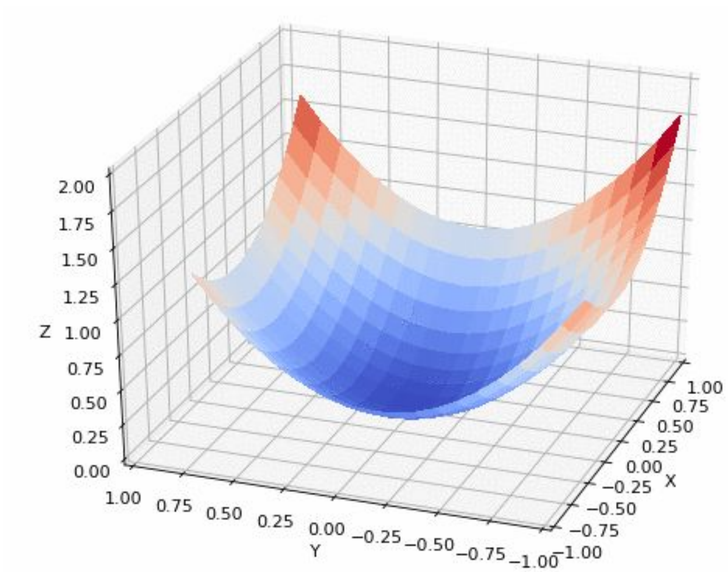
- First demonstration of flows as non-sequential samplers for lattice field theory [Albergo et al., 1904.12072]
- Variation of "**real non-volume-preserving flows**" developed for image generation [Dinh et al., 1605.08803]
  - Update field via sequential "coupling layers" g_i
  - Each layer transforms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

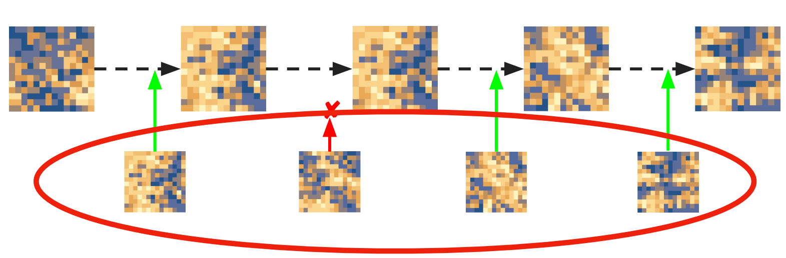




Flow models for scalar fields







Generating samples is "embarrassingly parallel"

Parameterize flow

Each layer contains **arbitrary neural networks** with many free parameters

Training step

- Draw samples from model
- Compute loss function
- Gradient descent

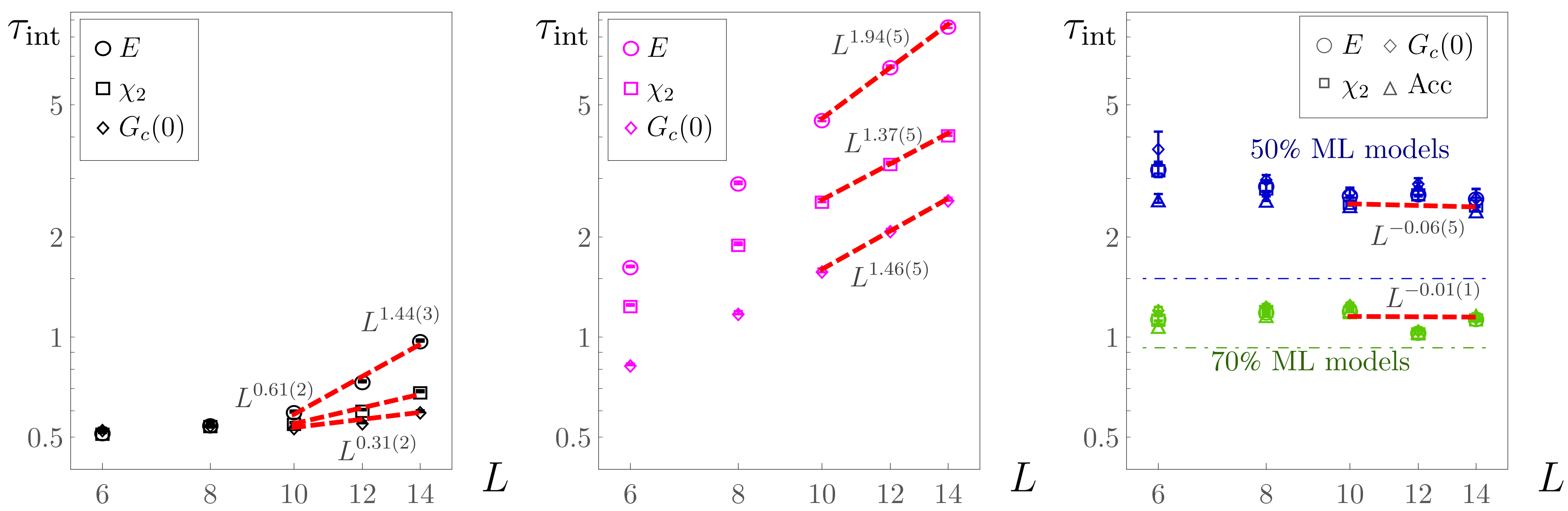
Desired accuracy?

Save trained model

Markov chain using samples from model

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_pL) [Albergo et al., 1904.12072]



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

- **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 [Abbott et al., 2208.03832]

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

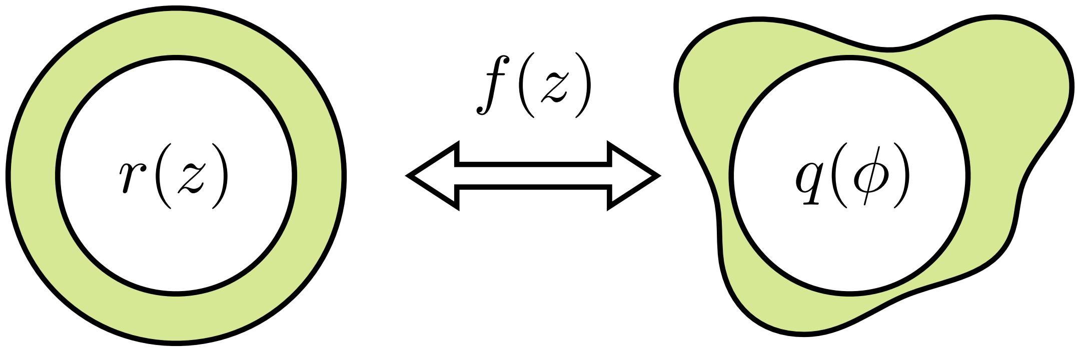
AURORA | EARLY SCIENCE PROGRAM FOR DATA & LEARNIN

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

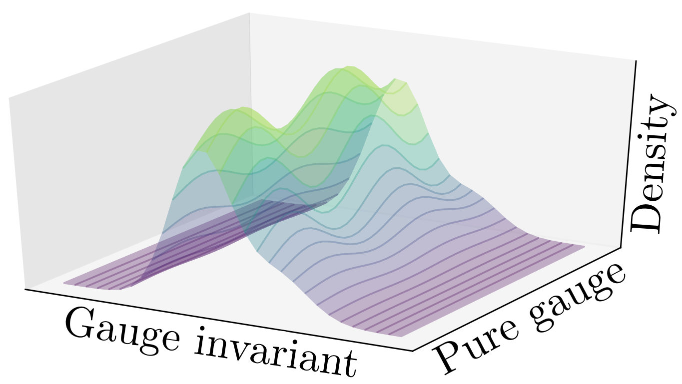
Flow models for gauge field theories

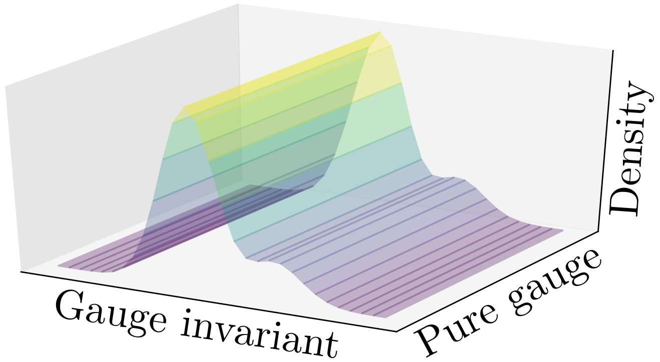
Flow models for gauge field theories require additional developments:

- 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





Gauge-equivariant flows

First gauge theory application: U(1) field theory

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

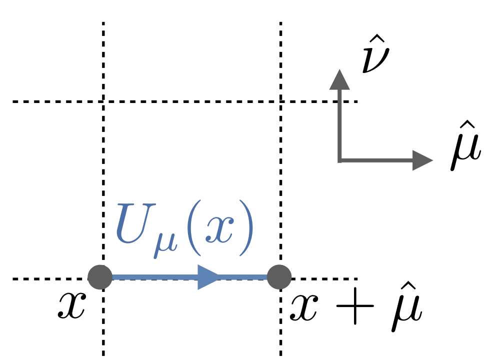
Define invertible, equivariant coupling layer

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

Spacetime dimension (N_d)

Lattice volume (V)

Act on a subset of the variables in each layer



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

Links updated by coupling layer (U'^A)

Links frozen in coupling layer (U^B)

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

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

Gauge-equivariant flows

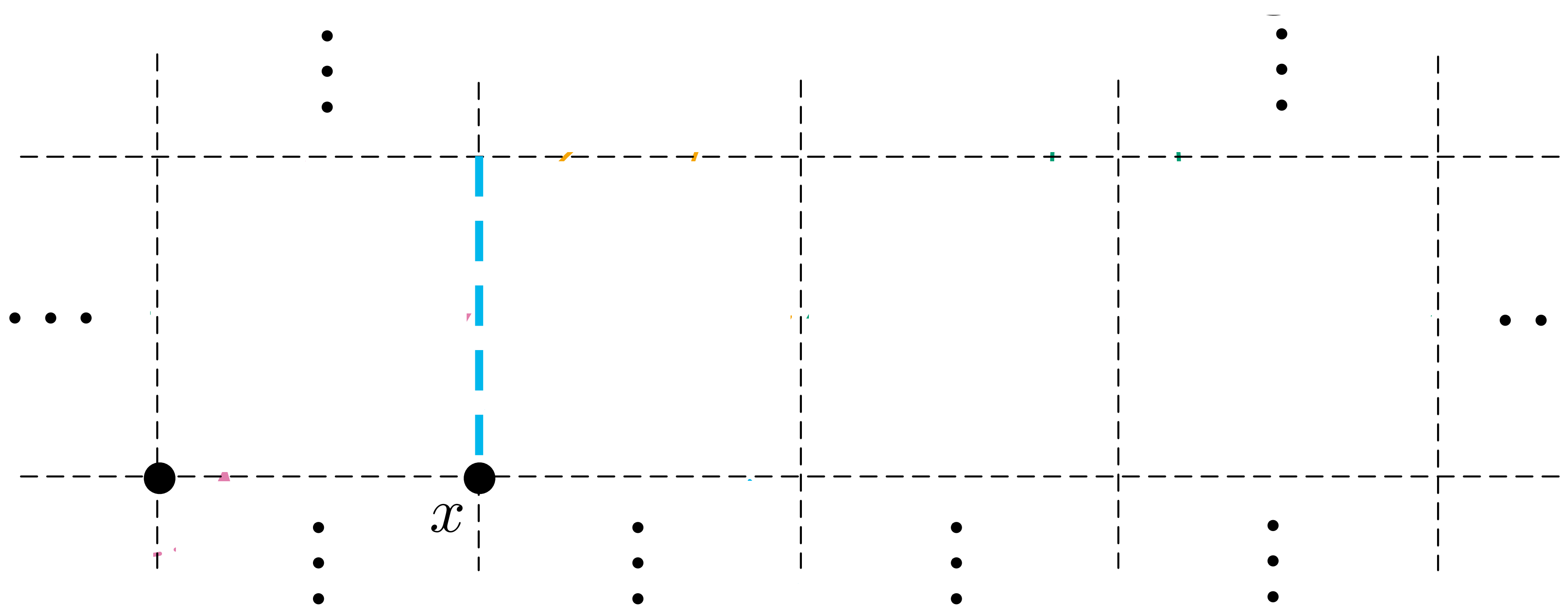
First gauge theory application: U(1) field theory

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



Phiala Shanahan, MIT

Gauge-equivariant flows

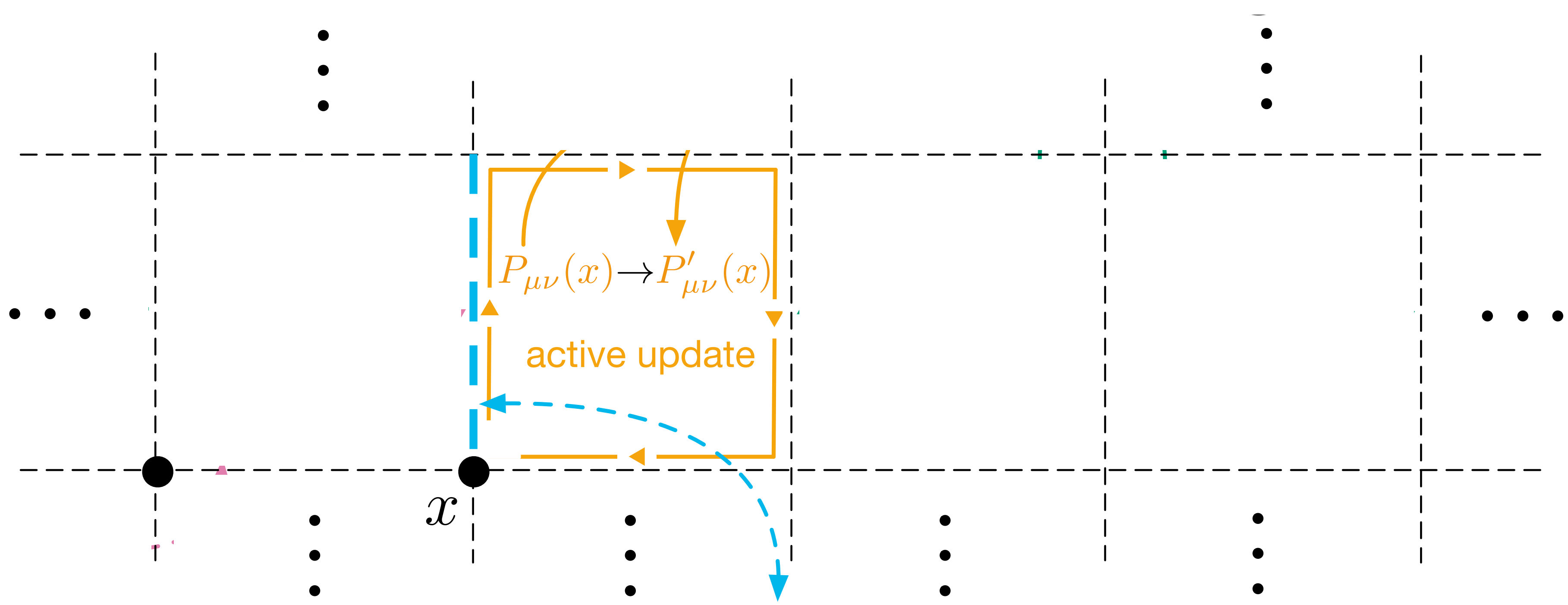
First gauge theory application: U(1) field theory

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



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

Gauge-equivariant flows

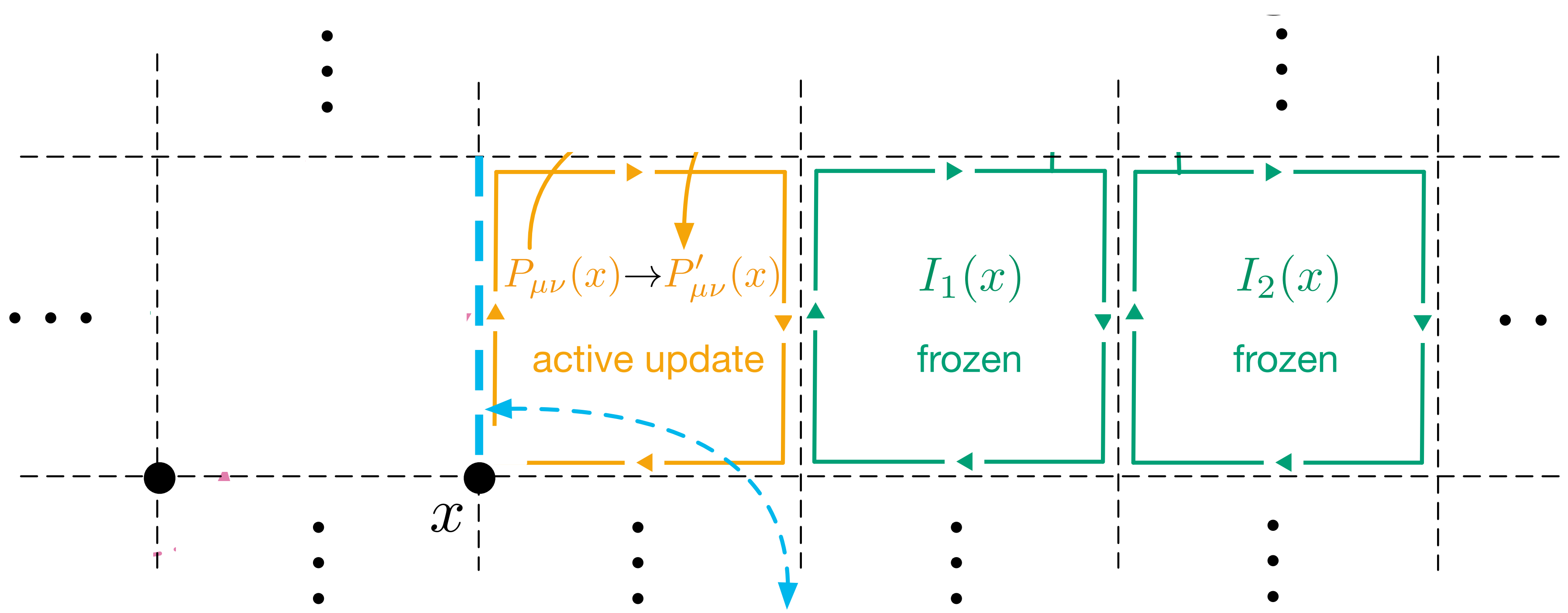
First gauge theory application: U(1) field theory

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

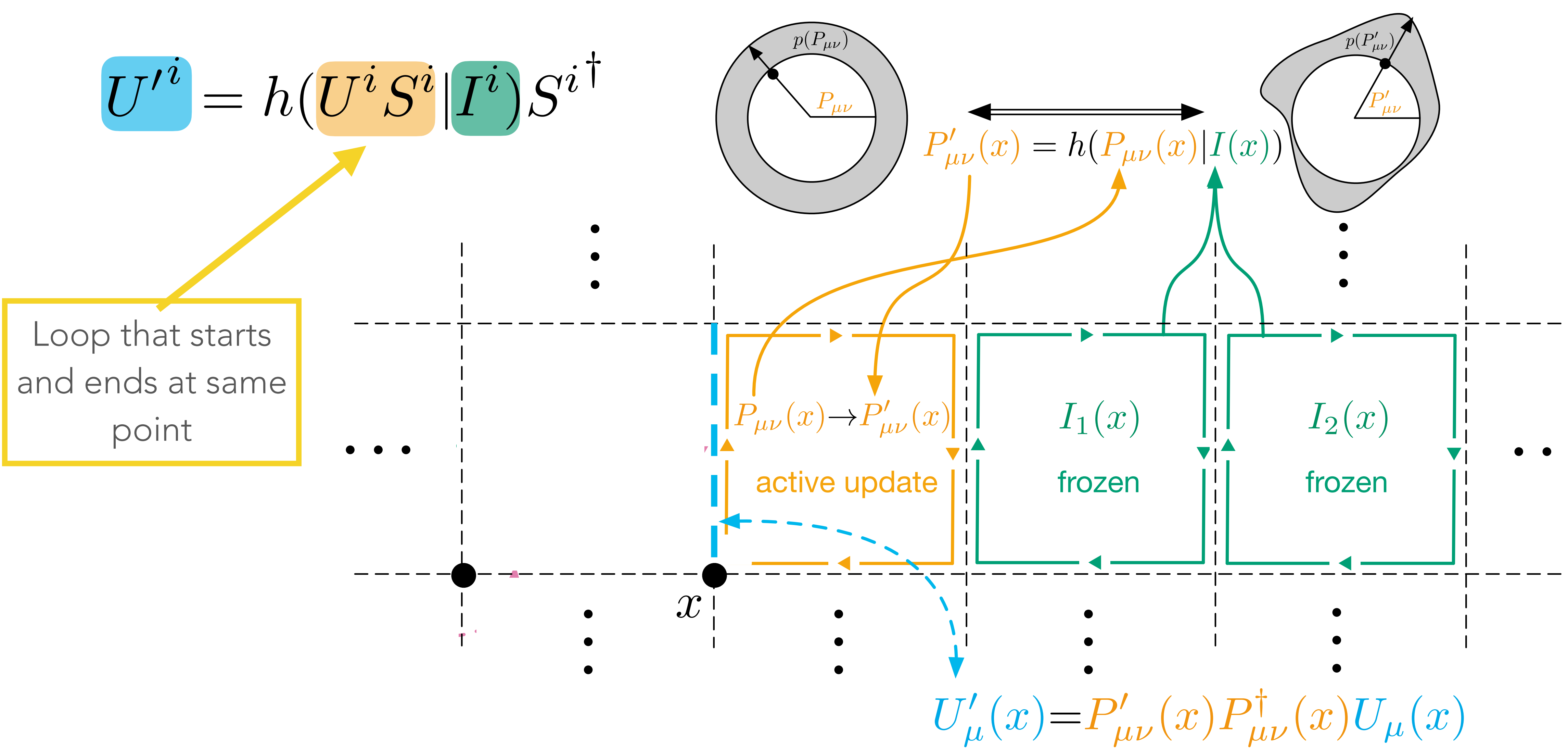


$$U'_\mu(x) = P'_{\mu\nu}(x) P^\dagger_{\mu\nu}(x) U_\mu(x)$$

Gauge-equivariant flows

First gauge theory application: U(1) field theory

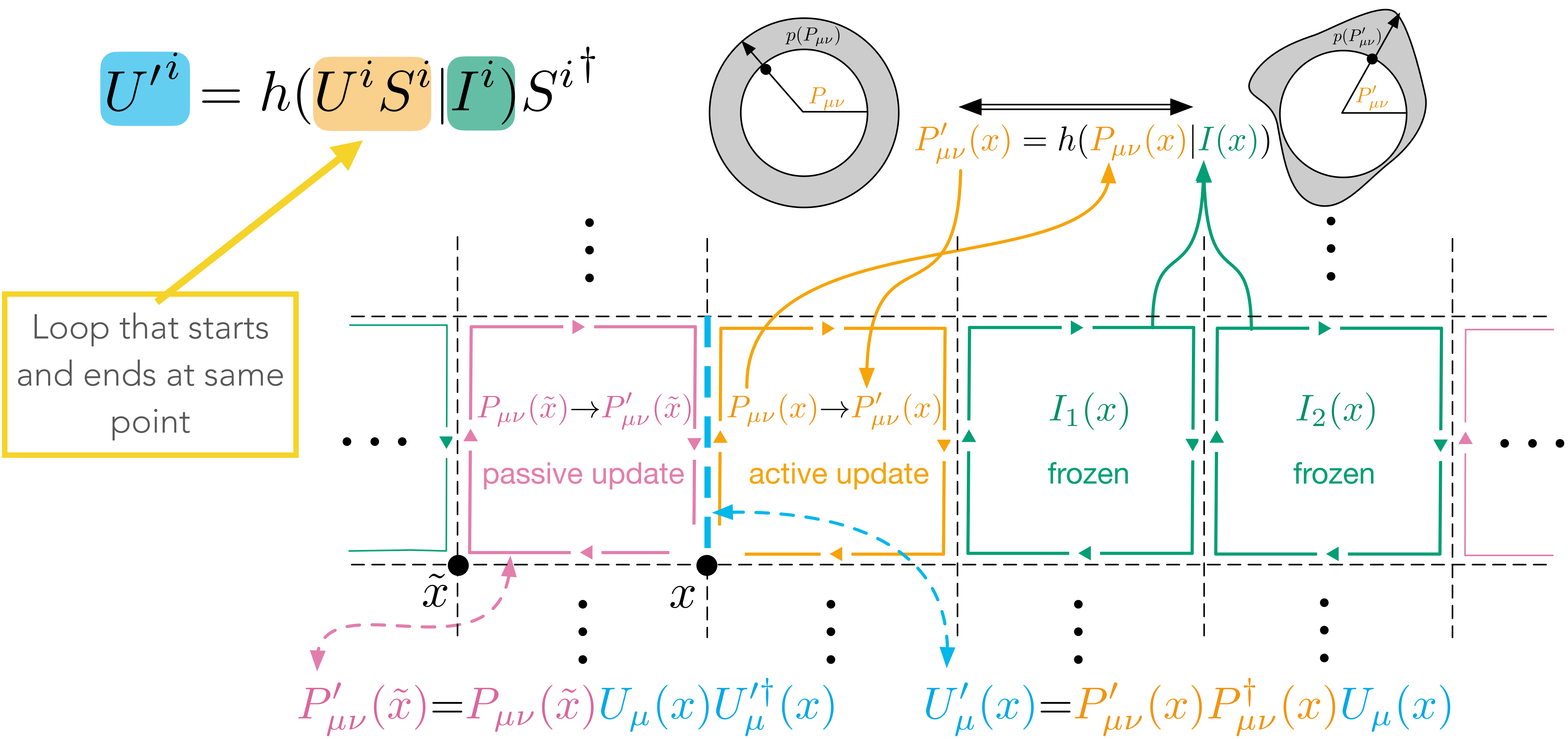
Generative flow architecture that is *gauge-equivariant*



Gauge-equivariant flows

First gauge theory application: U(1) field theory

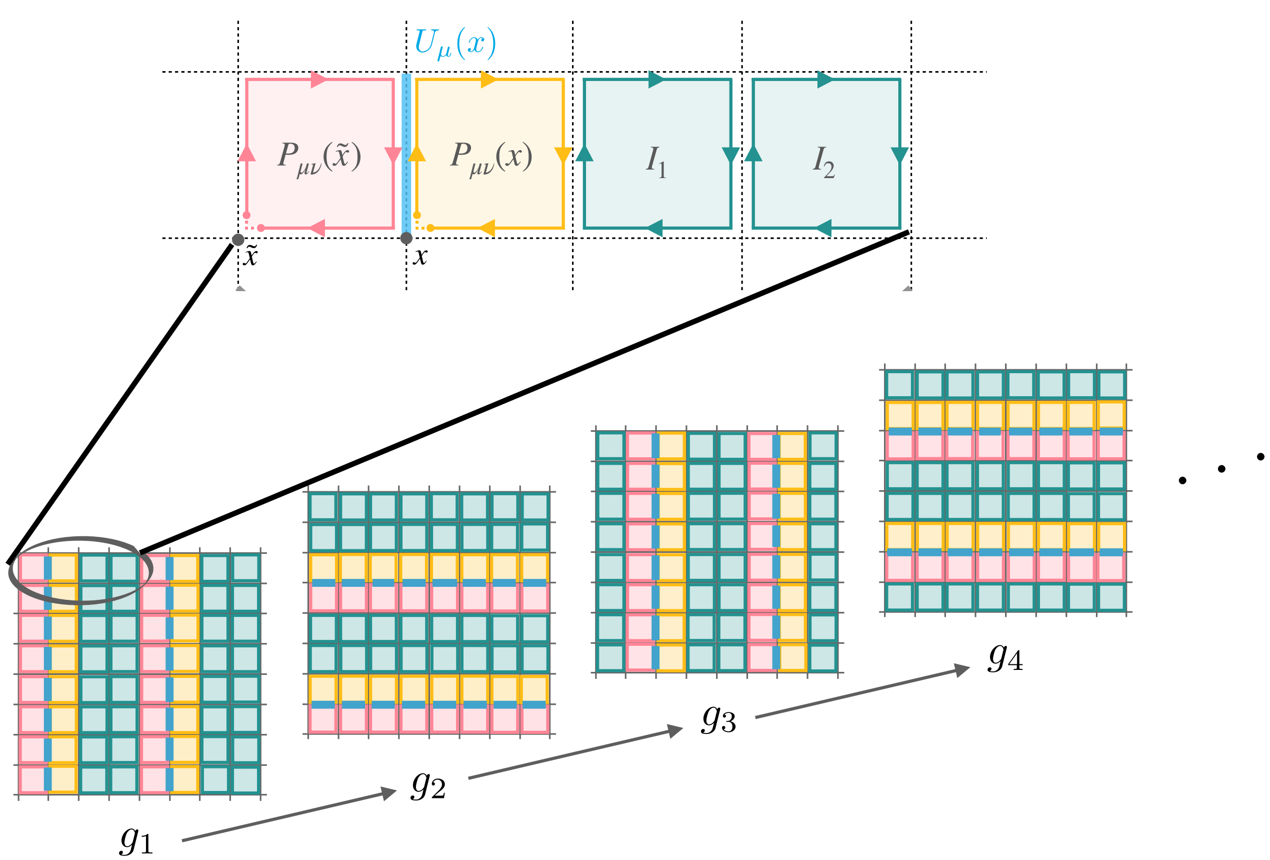
Generative flow architecture that is *gauge-equivariant*

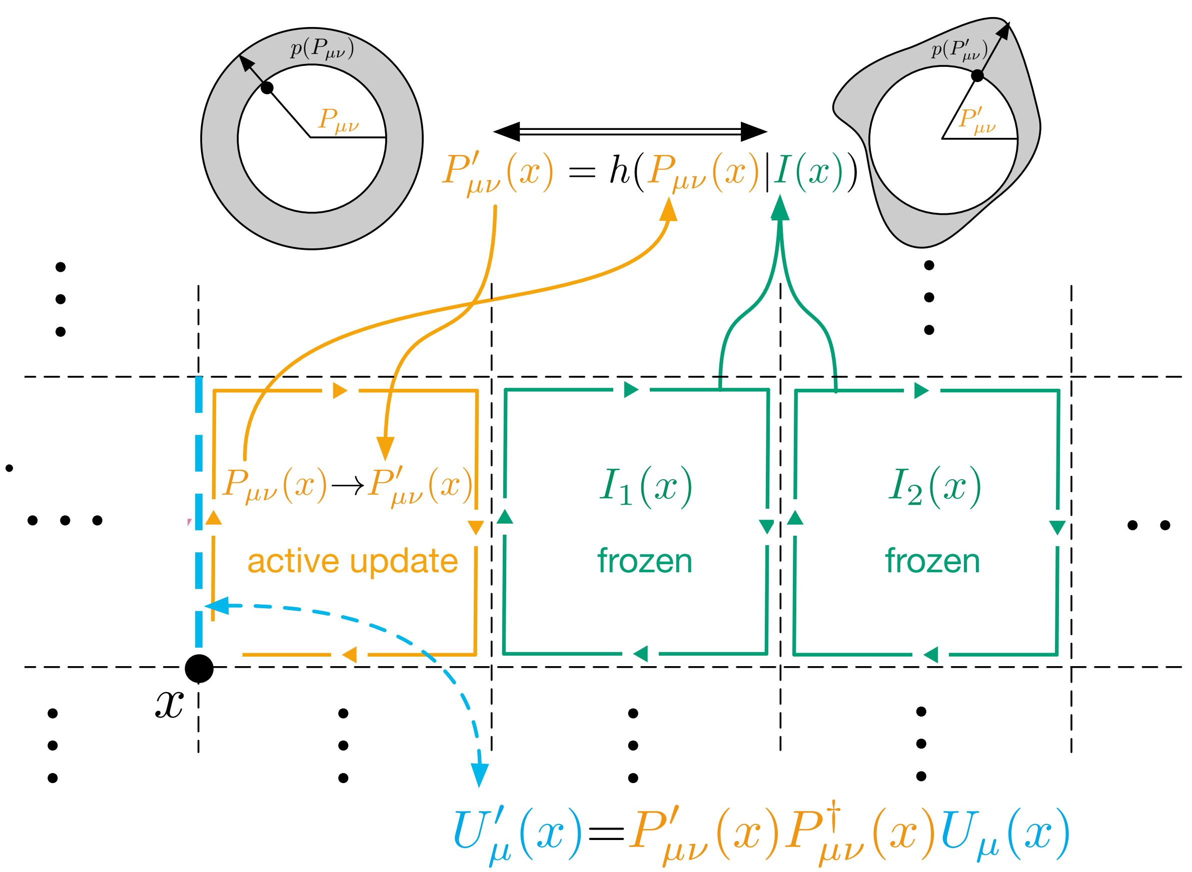


Flow models for gauge field theories

Development of gauge-equivariant flow models for Abelian and non-Abelian theories [Kanwar et al., 2003.06413, Boyda et al., 2008.05456; Favoni et al., 2012.12901,2111.04389; Luo et al., 2012.05232]

- Transform subset of links conditioned on the remaining subset
- Create gauge-equivariant layers by acting via transformations of (untraced) loops

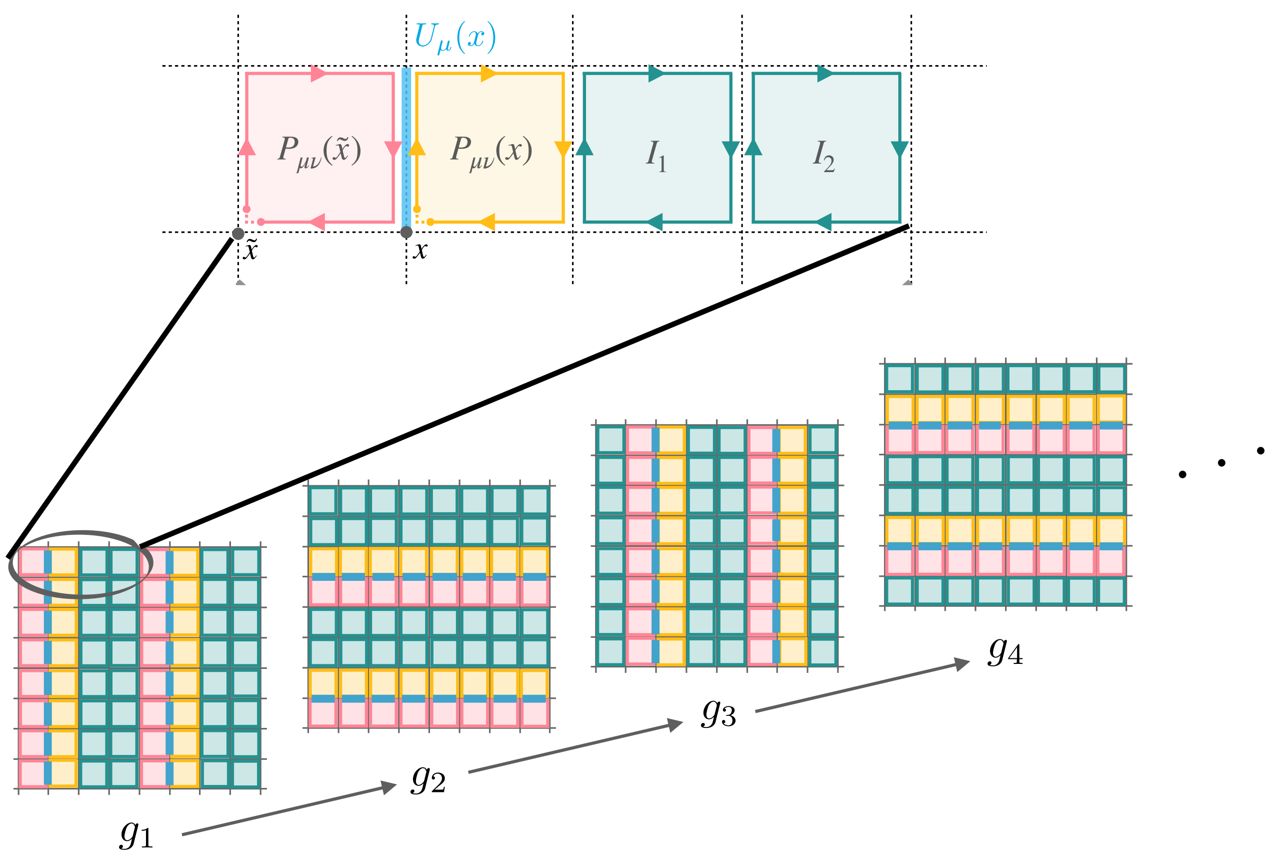


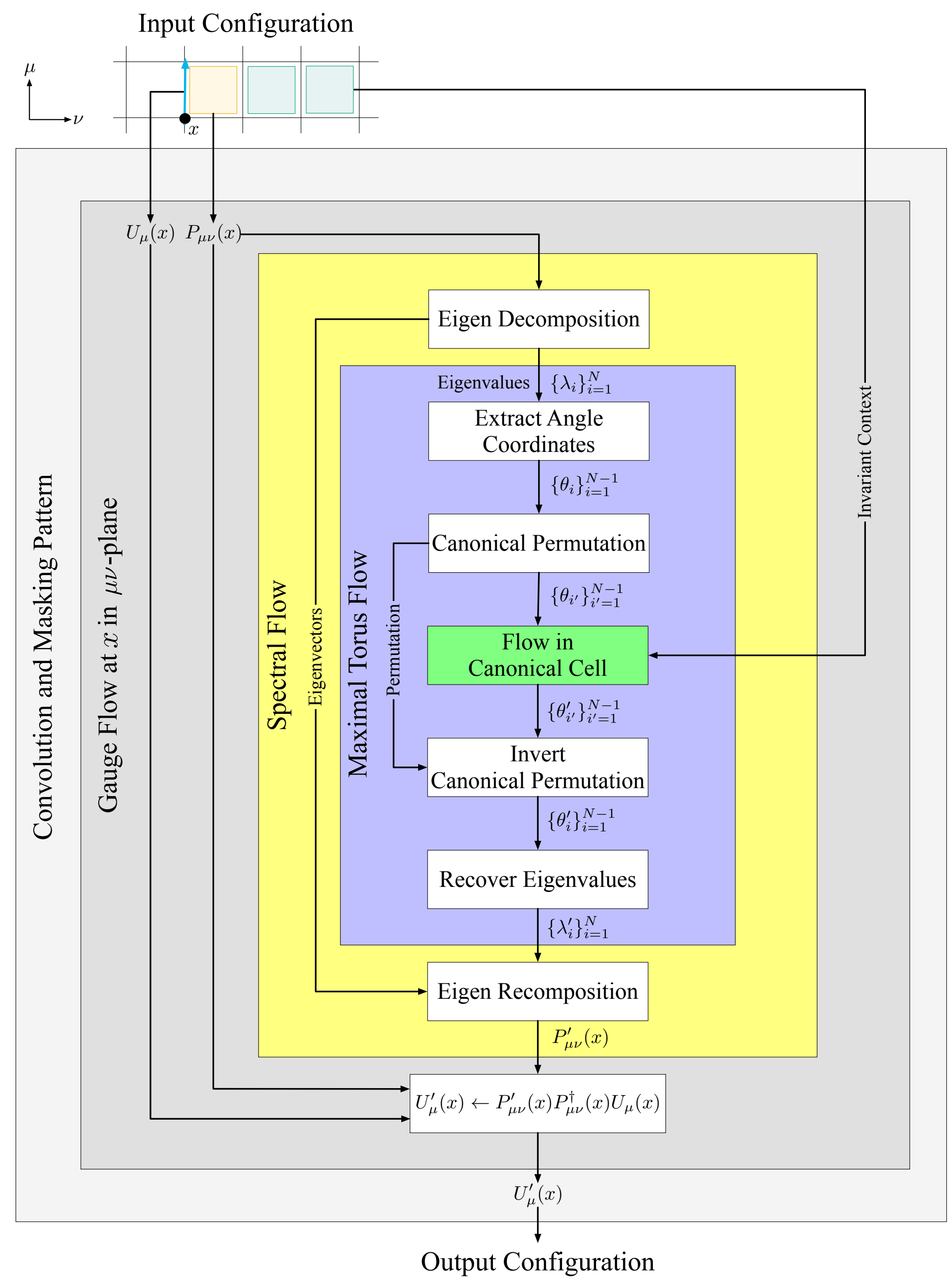


Flow models for gauge field theories

Development of gauge-equivariant flow models for Abelian and non-Abelian theories [Kanwar et al., 2003.06413, Boyda et al., 2008.05456; Favoni et al., 2012.12901,2111.04389; Luo et al., 2012.05232]

- Gauge-equivariant layers for Non-Abelian theories by stepping via eigendecomposition of (untraced) loops





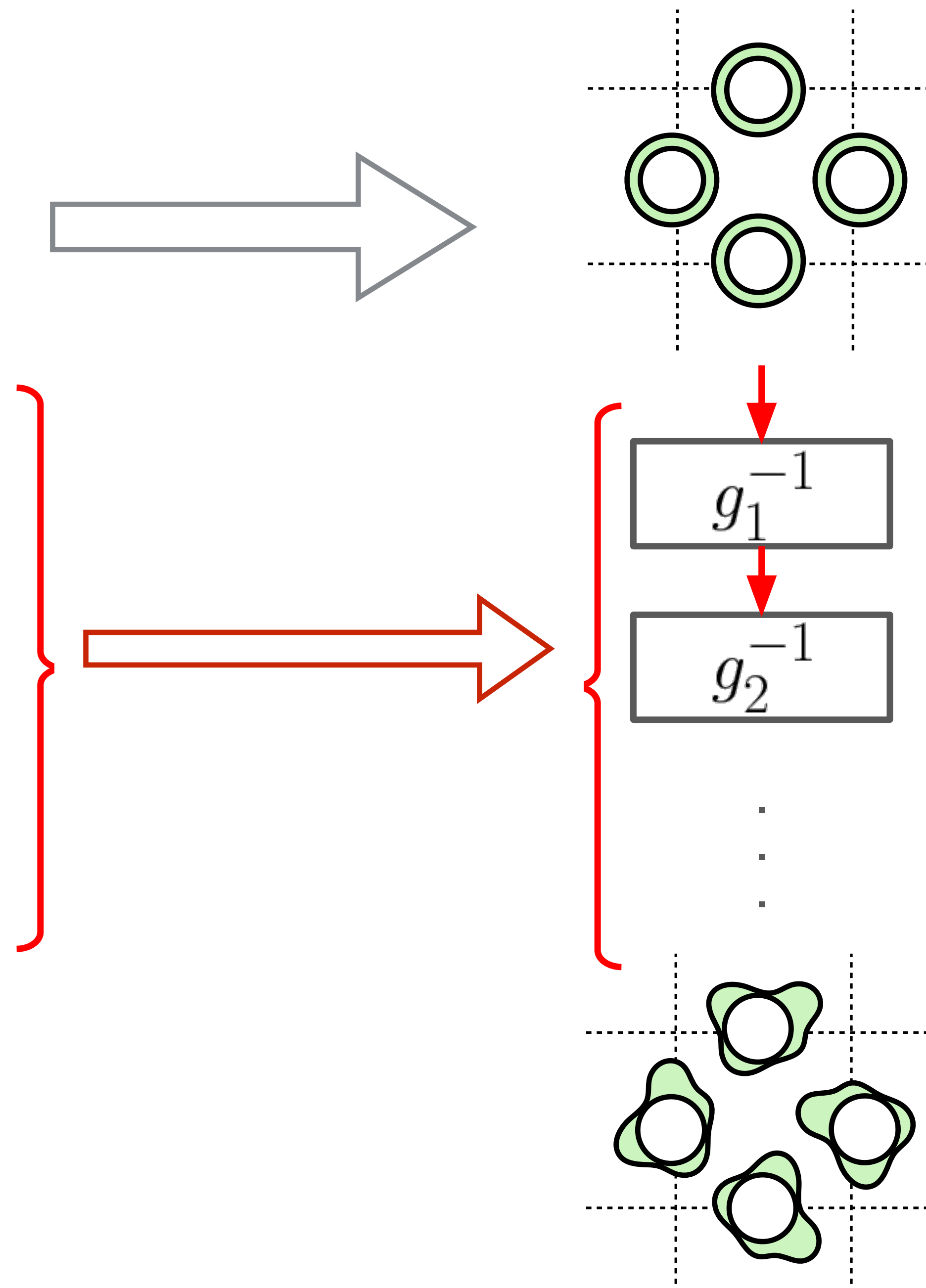
Application: U(1) field theory

Demonstration of accelerated sampling in U(1) field theory

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

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

- 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

Demonstration of accelerated sampling in U(1) field theory

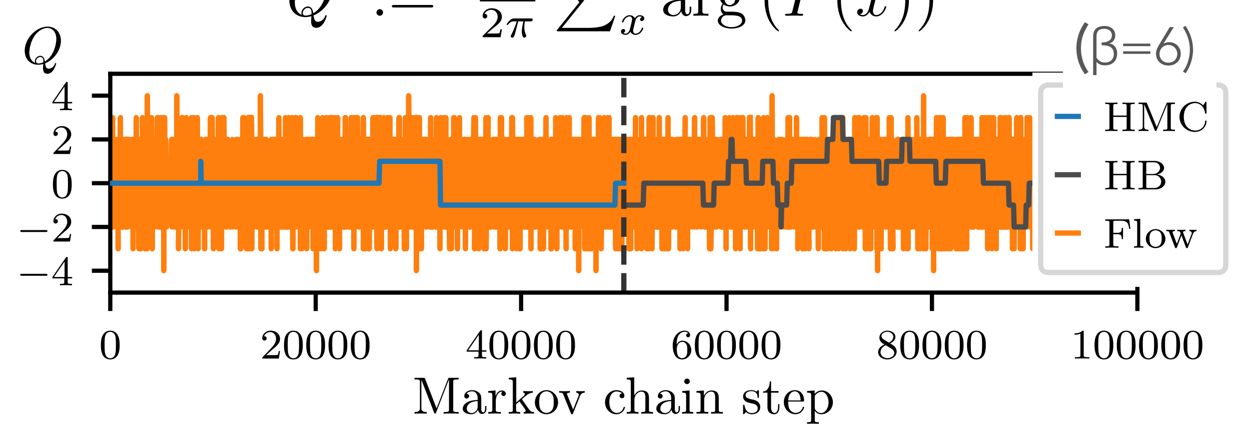
(2D, L=16) [Kanwar et al., 2003.06413]

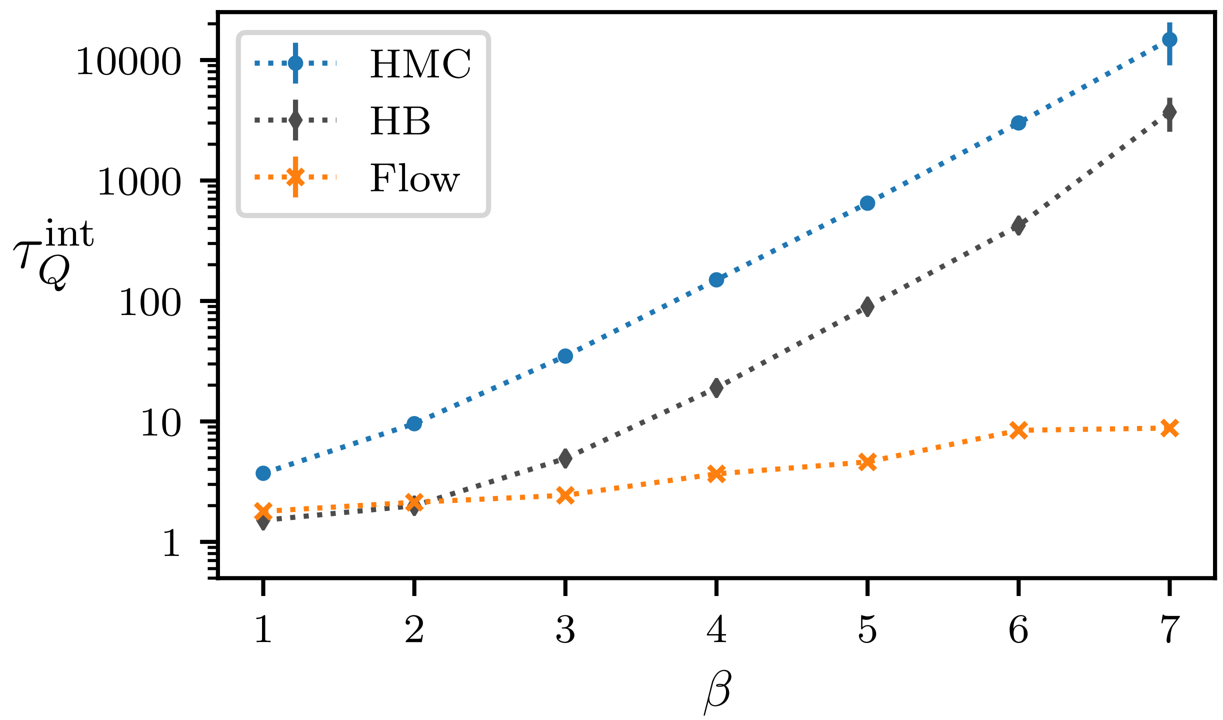
$$S(U) := -\beta \sum_x \mathrm{Re}\, P(x)$$

- **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

$$Q := \frac{1}{2\pi} \sum_x \arg\left(P(x)\right)$$





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]

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

[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

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

(Plaquette) (Fermion determinant)

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

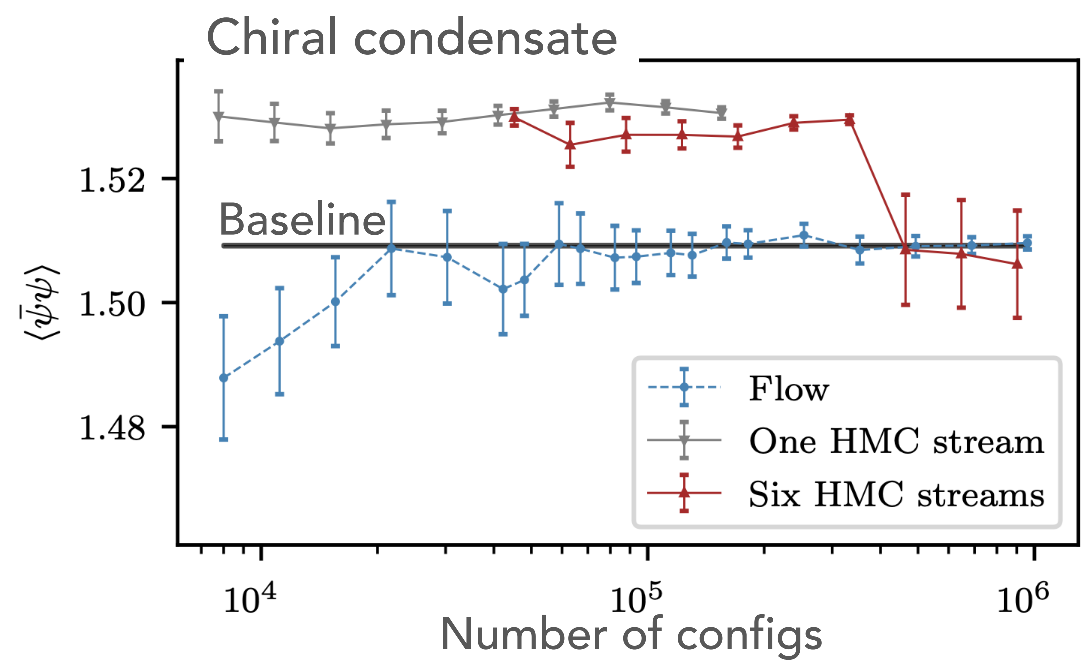
✳ Evaluate determinant using auxiliary (pseudofermion) degrees of freedom

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

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]

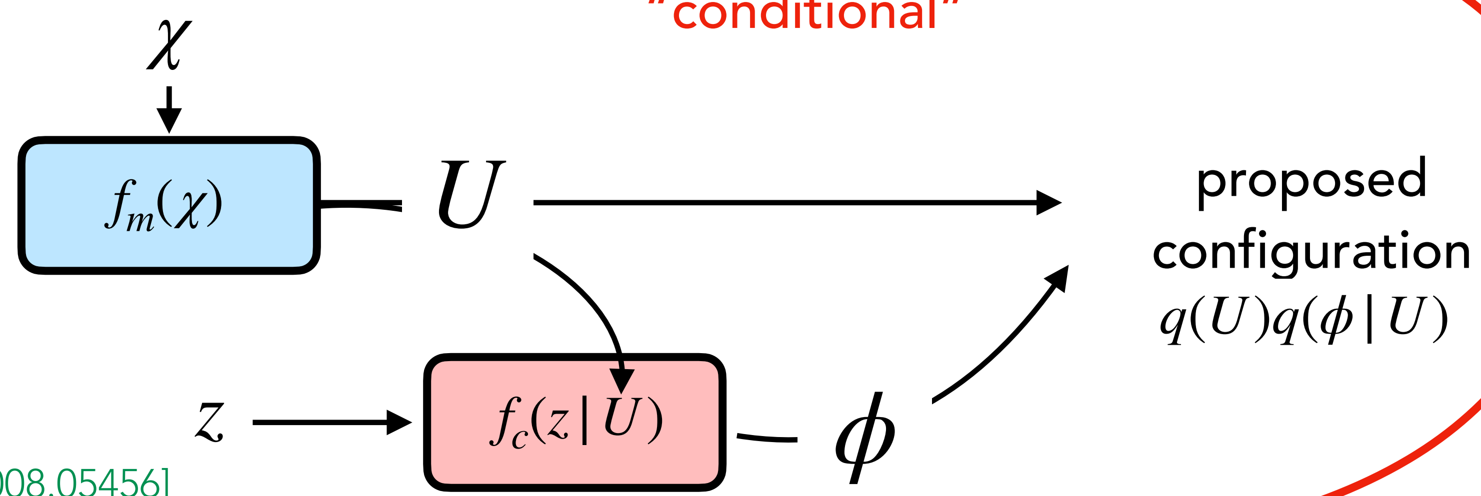
$$p(U,\phi) = p(U)p(\phi \mid U)$$

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

"marginal"

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

"conditional"



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

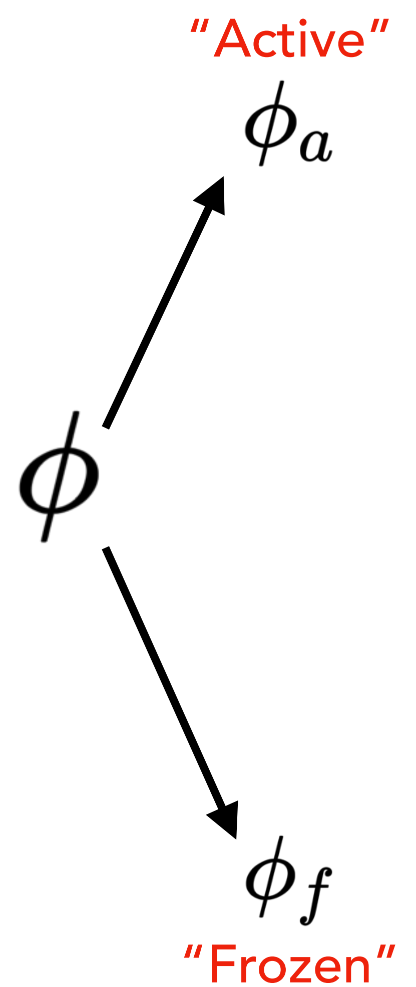
Flow models with fermions

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



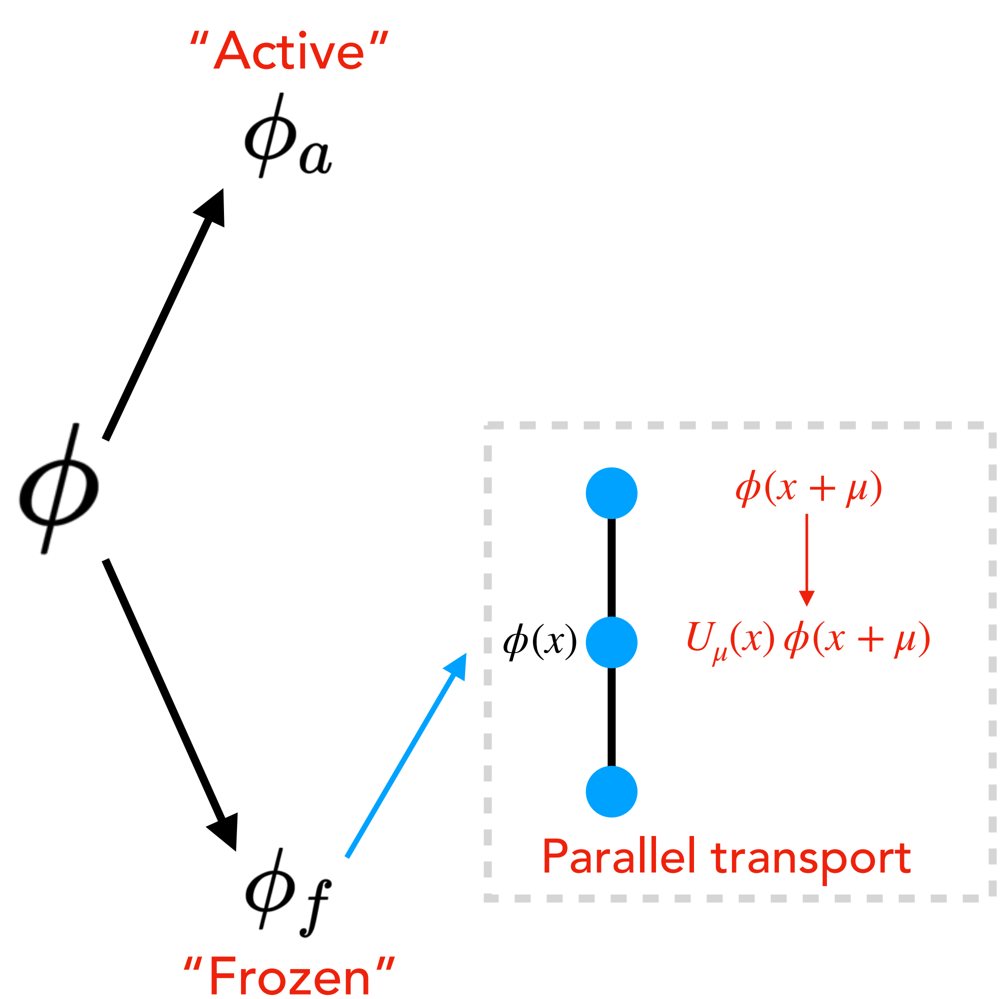
Flow models with fermions

Conditional model maps uncorrelated Gaussian to correlated Gaussian

[Abbott et al., 2207.08945]

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

Approximates



Phiala Shanahan, MIT

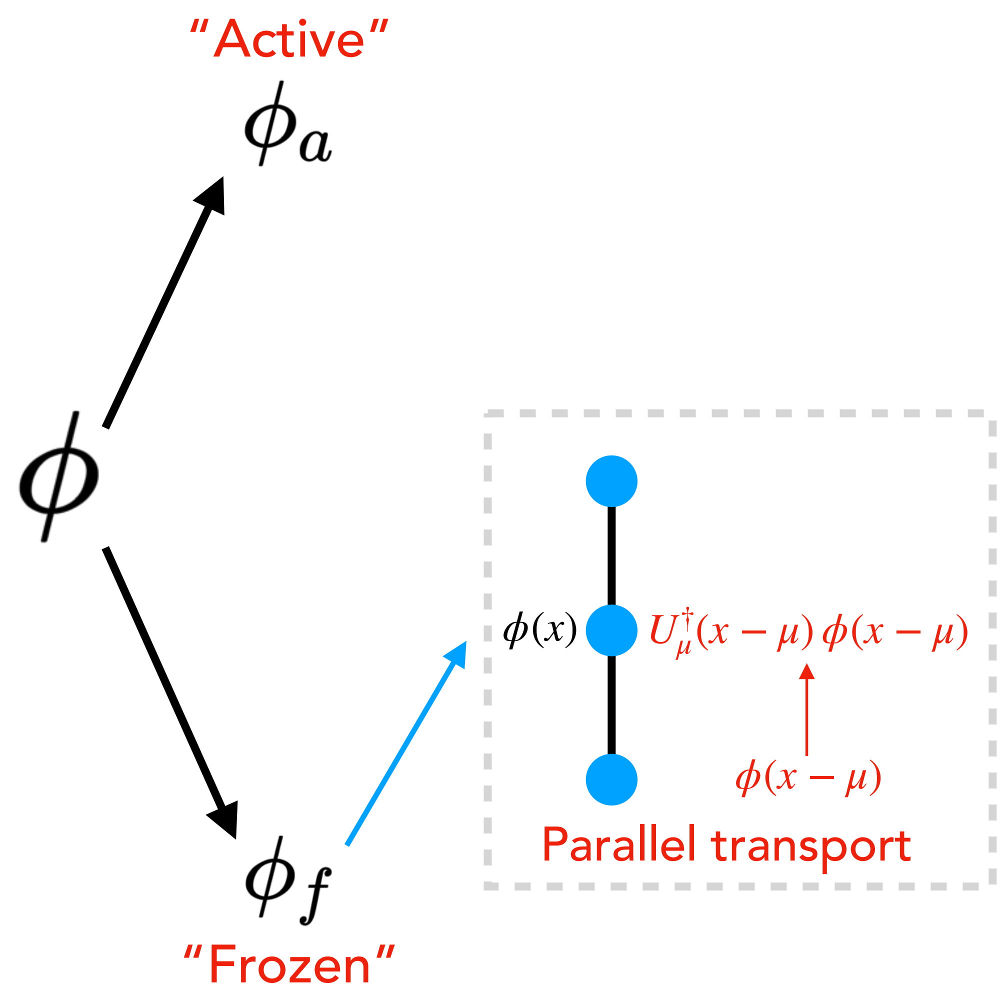
Flow models with fermions

Conditional model maps uncorrelated Gaussian to correlated Gaussian

[Abbott et al., 2207.08945]

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

Approximates



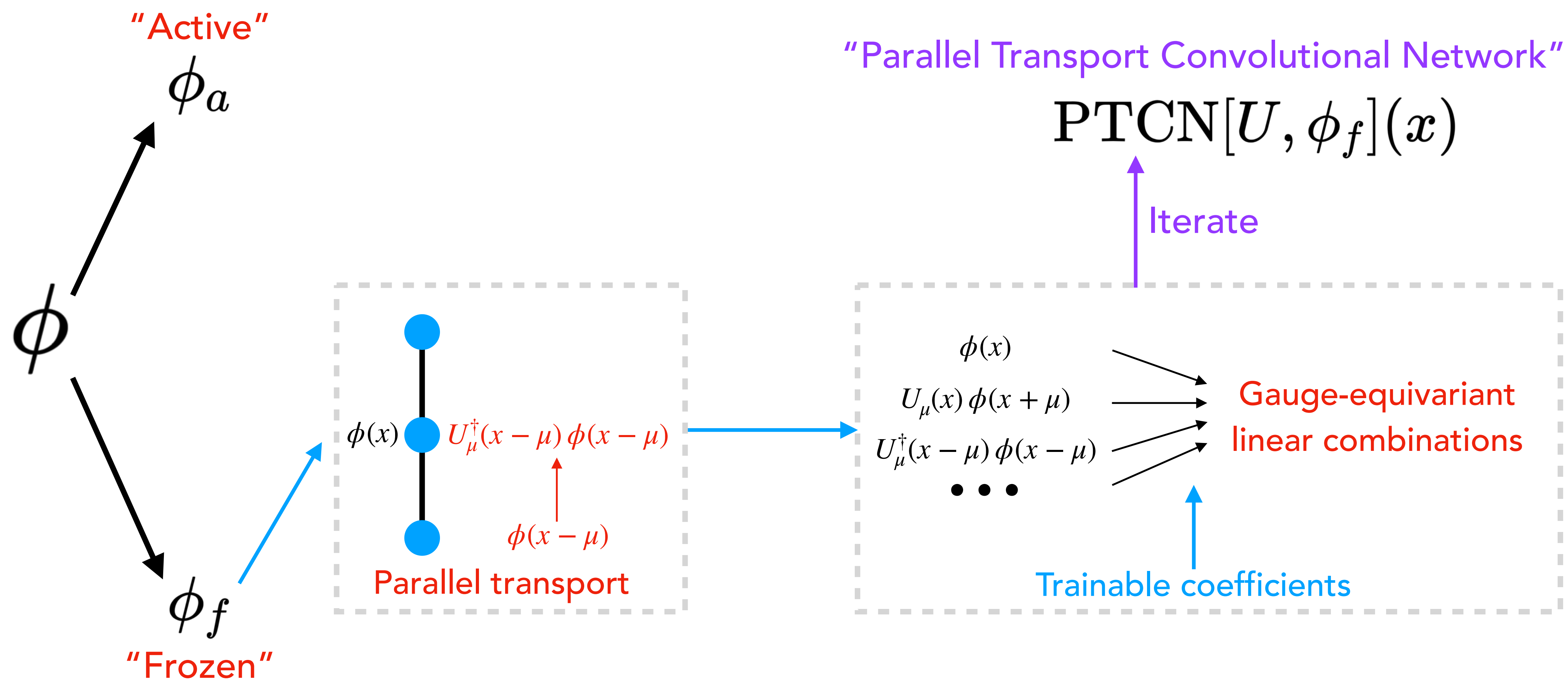
Flow models with fermions

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

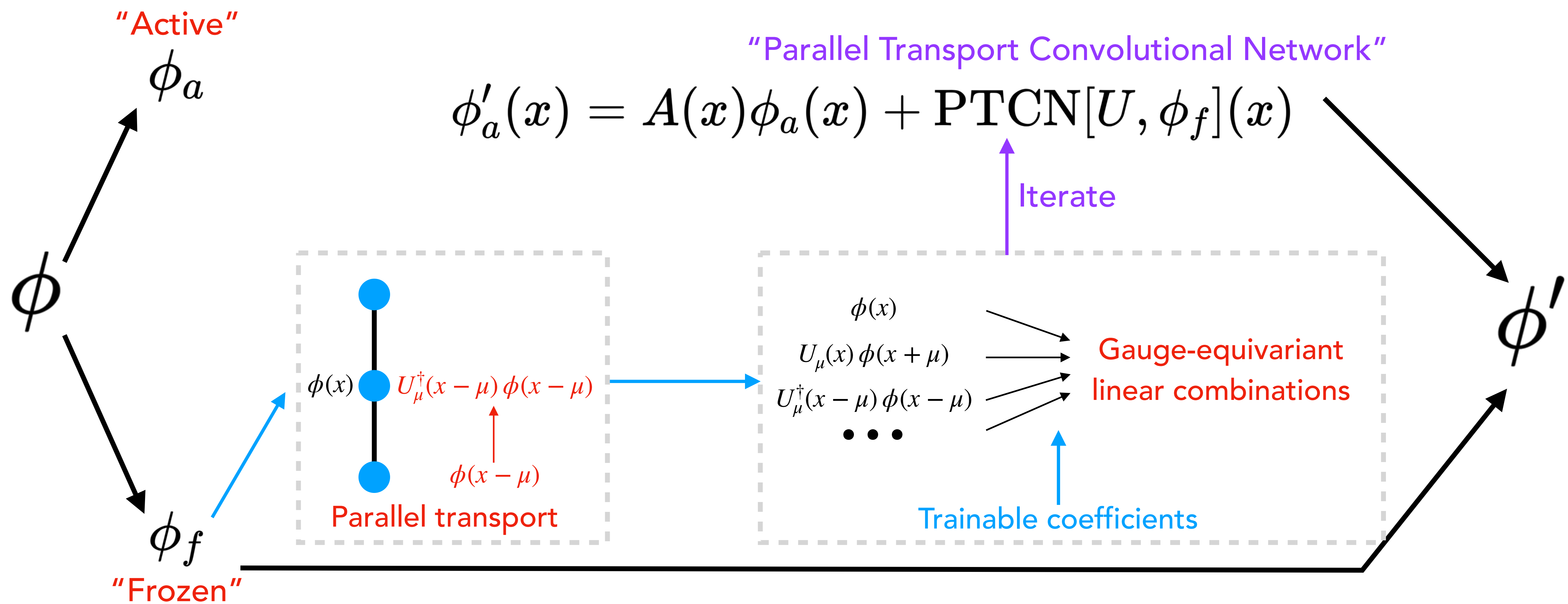
[Abbott et al., 2207.08945]

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

Approximates

"Parallel Transport Convolutional Network"

$$\phi'_a(x) = A(x)\phi_a(x) + \mathrm{PTCN}[U, \phi_f](x)$$



Phiala Shanahan, MIT

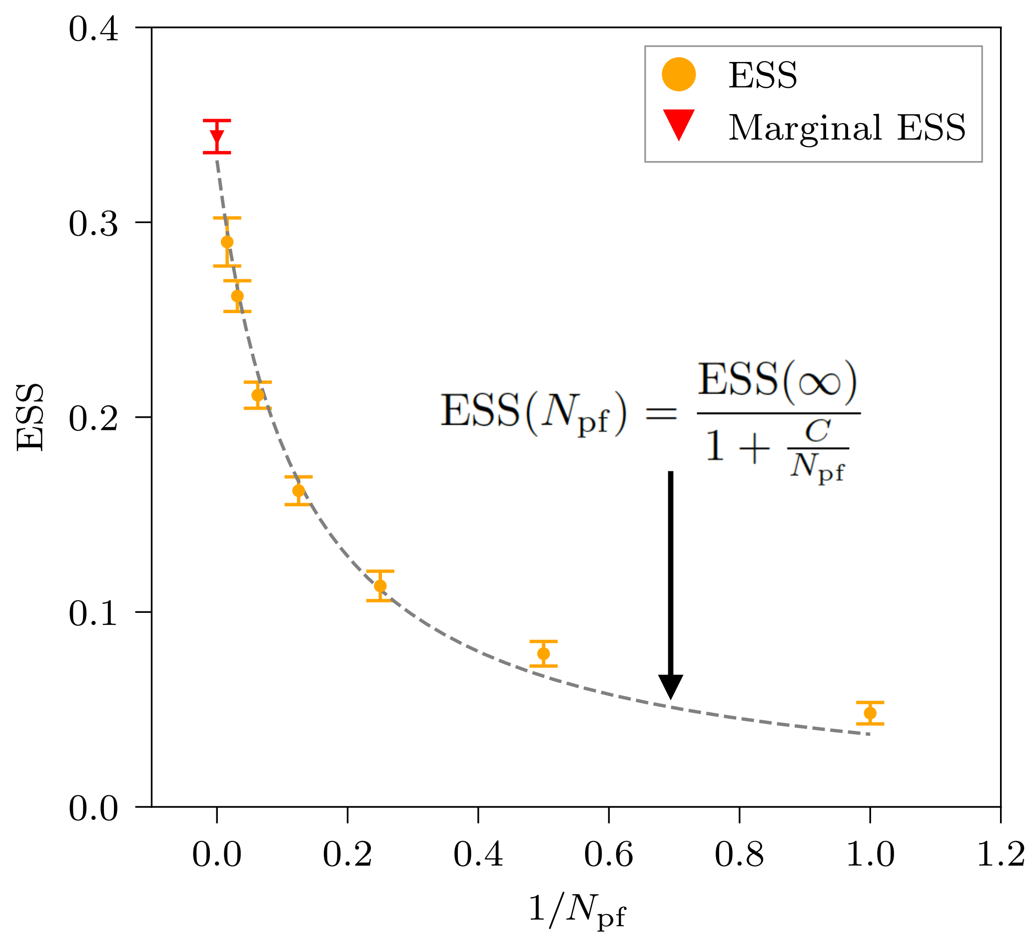
Flow models with fermions

Use joint models more efficiently:

- Draw multiple pseudofermion samples at fixed gauge field to improve stochastic estimate of weights:

$$w_{N_{\rm pf}}(U) = \frac{1}{N_{\rm pf}} \sum_{i=1}^{N_{\rm pf}} \frac{p(\phi^{(i)}, U)}{q(\phi^{(i)}, U)}$$

- Does not require re-evaluating observables

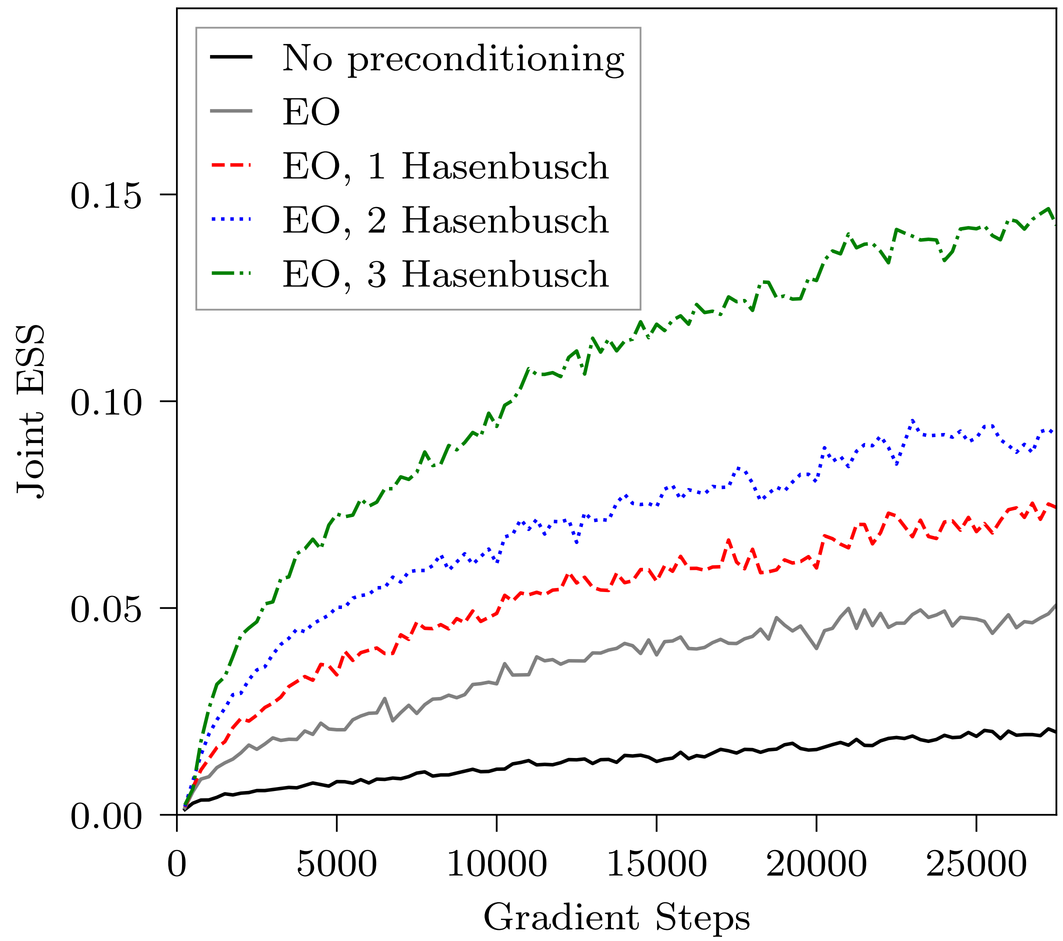


Reduce problem difficulty:

- Combine with preconditioning e.g., even-odd or Hasenbusch factorisation

$$\det M = \left[\frac{\det M}{\det(M+\mu)}\right] \det(M+\mu)$$

Schwinger model $\beta = 2.0,\ \kappa = 0.265\ \ L = 8$



Phiala Shanahan, MIT

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]

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

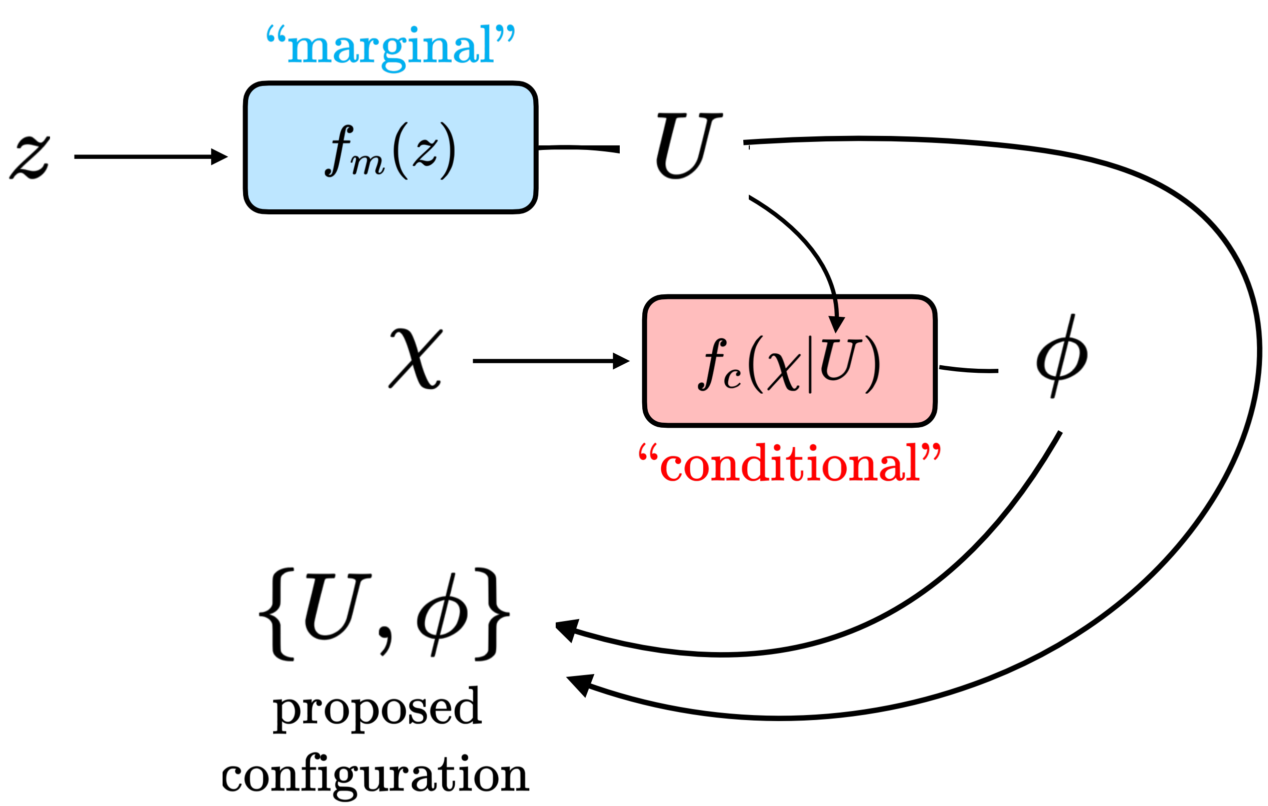
AURORA | EARLY SCIENCE PROGRAM FOR DATA & LEARNIN

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

Flow models for QCD in 4D

Initial QCD demonstration [Abbott et al., 2208.03832]

- 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!**



Marginal:

- Haar-uniform base distribution
- 48 gauge-equivariant spline coupling layers
- Spatially separated convolutions in spectral flow to define spline parameters

Conditional:

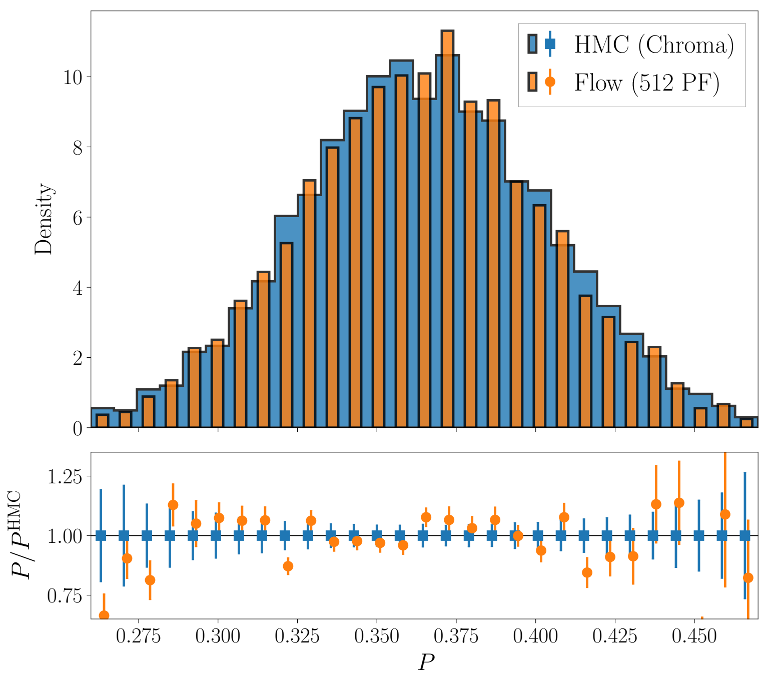
- Gaussian base distribution
- 36 pseudofermion coupling layers built from parallel transport convolutional networks
- Alternating spin and spatial masking pattern

Flow models for QCD in 4D

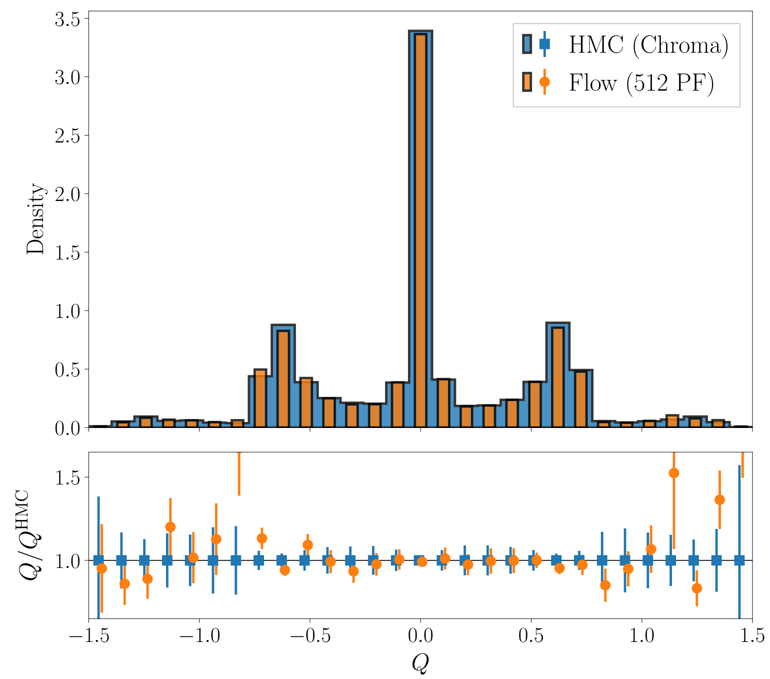
Initial QCD demonstration [Abbott et al., 2208.03832]

- 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!

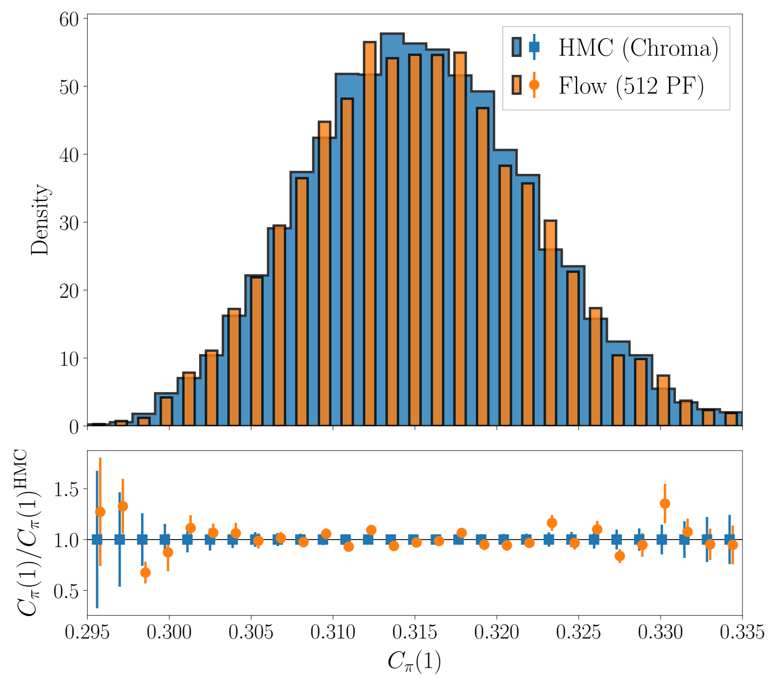
Plaquette



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