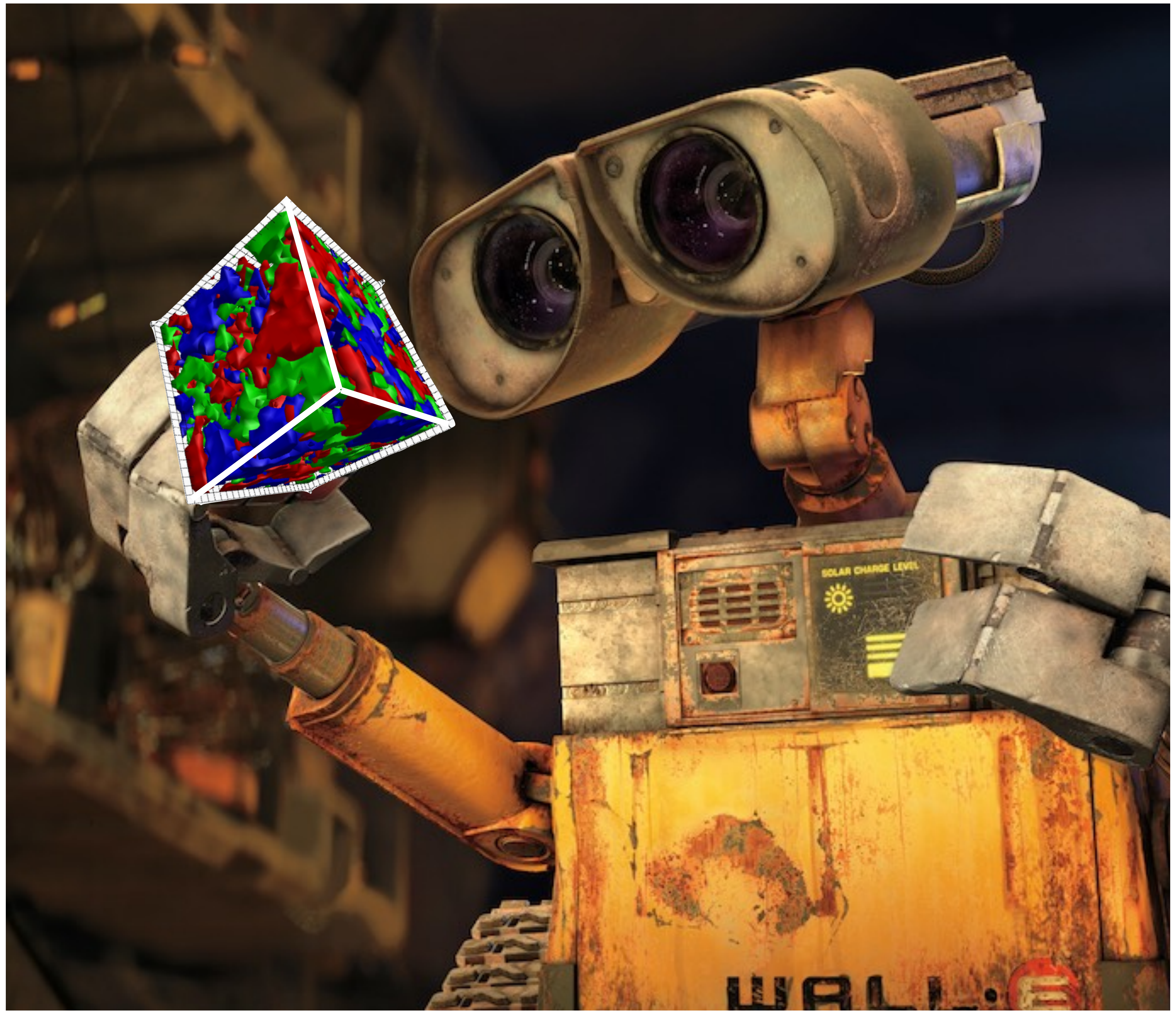
Generative flow models

for

Lattice Quantum Field Theory

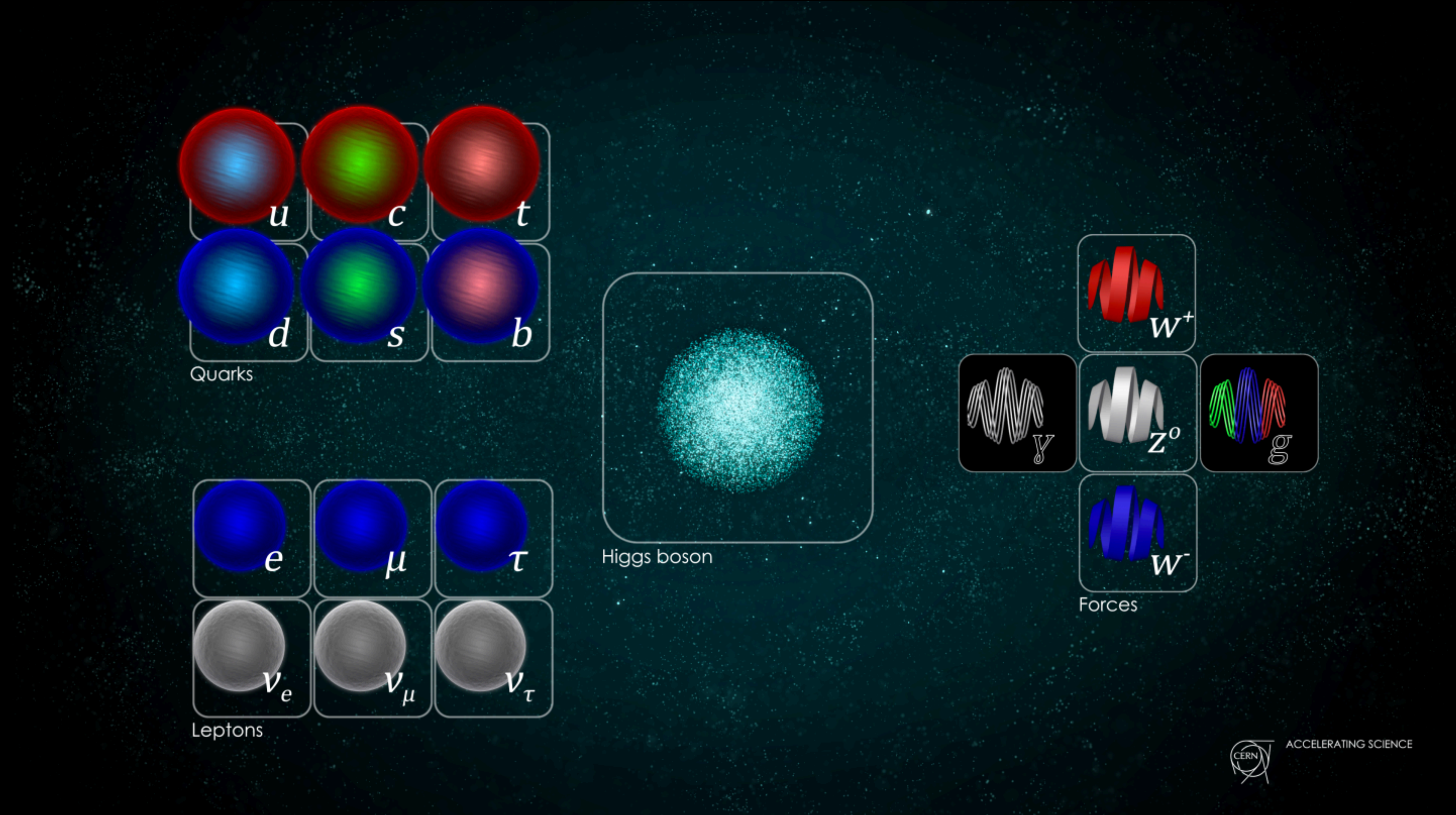


Stokes, Kamleh, Leinweber 1312.0991
WALL-E (2008) *Pixar, please don't sue me*

Gurtej Kanwar
Institute for Theoretical Physics, AEC, U. Bern

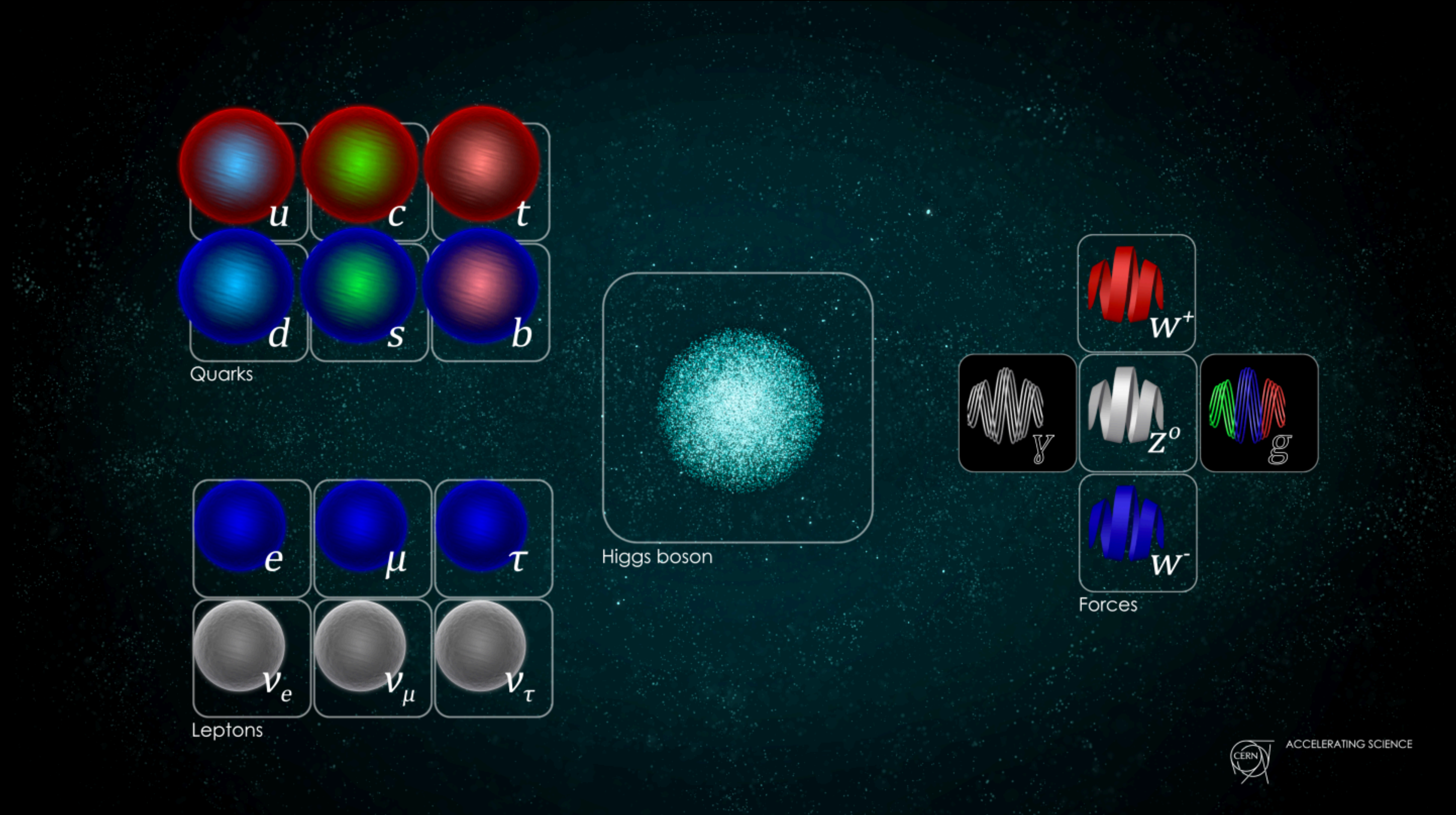
October 23, 2024
Bethe Colloquium

The Standard Model



The Standard Model

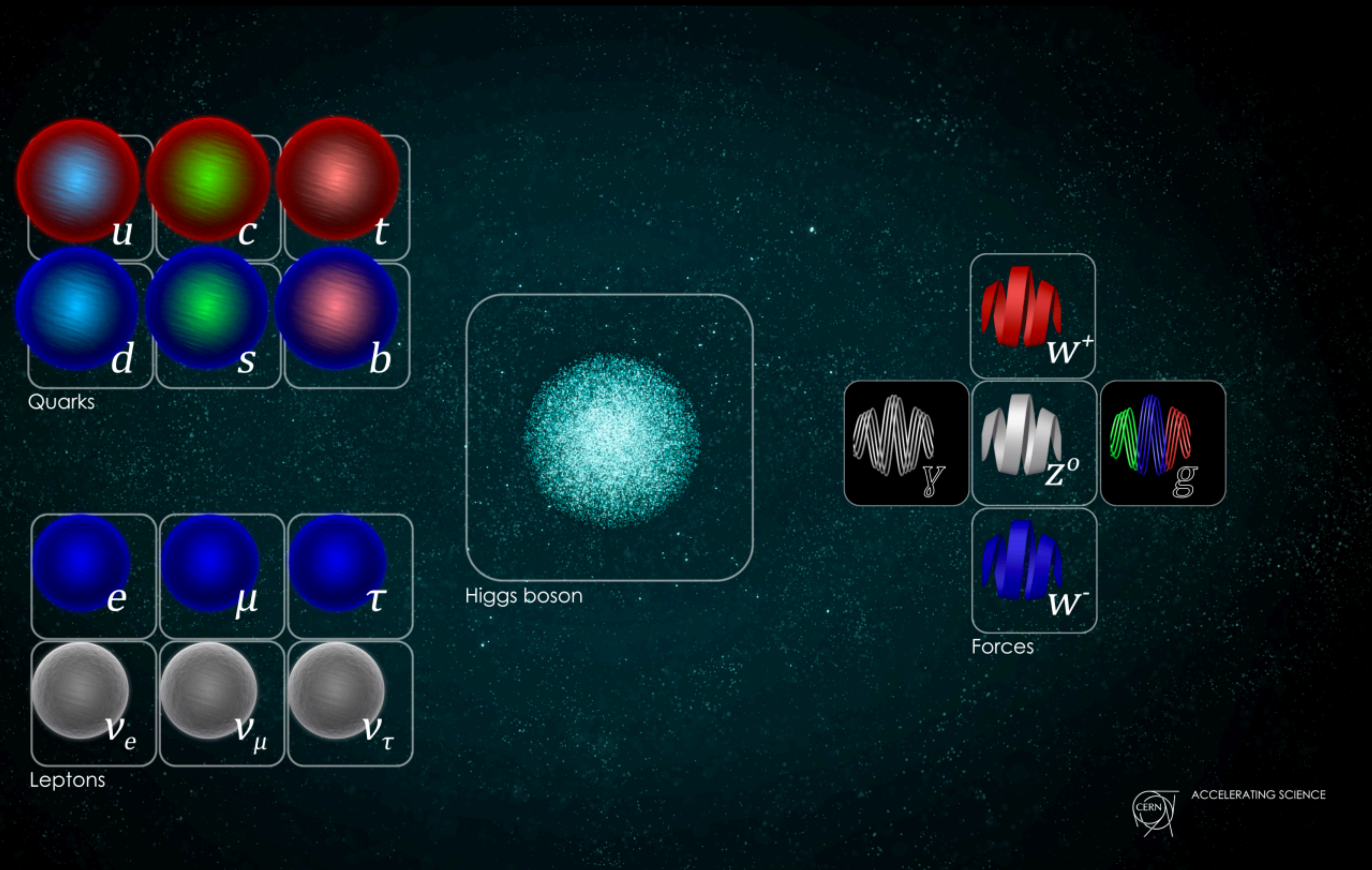
A story of many spectacular successes…



The Standard Model

A story of many spectacular successes…

- Unified description as a **quantum field theory** with only 3 fundamental forces and 19 free parameters

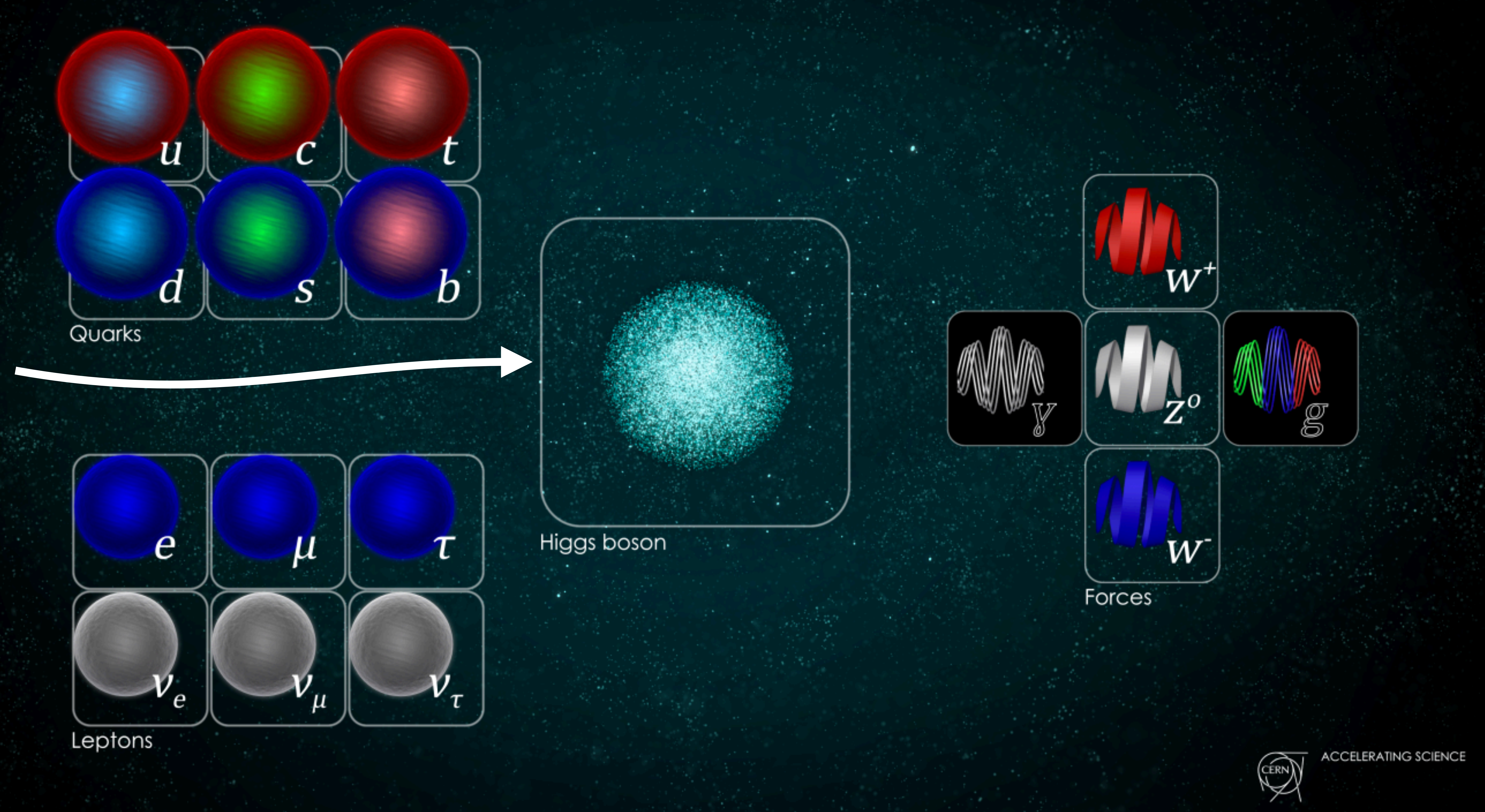


The Standard Model

A story of many spectacular successes…

- Unified description as a **quantum field theory** with only 3 fundamental forces and 19 free parameters

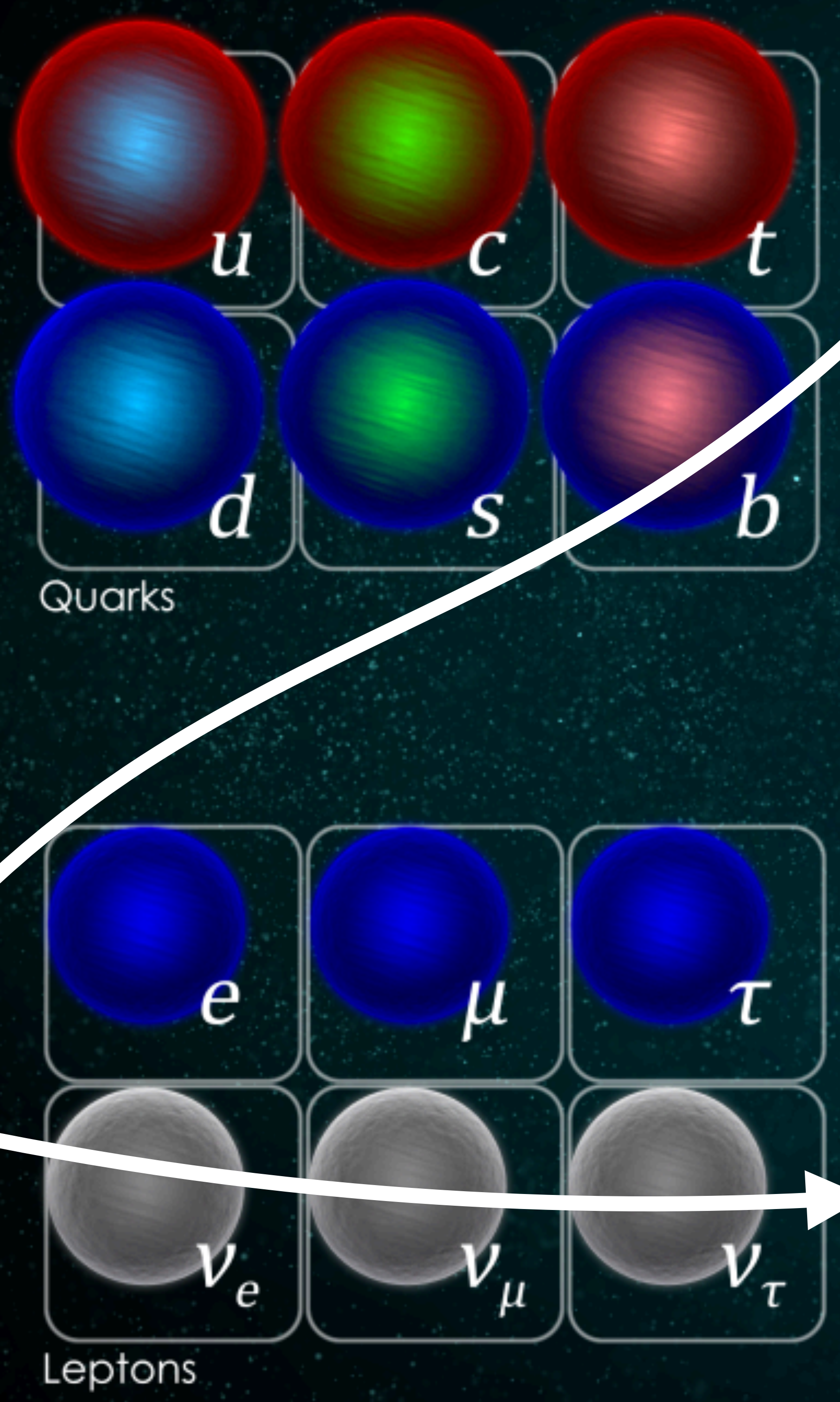
- Discovery of the Higgs in 2012

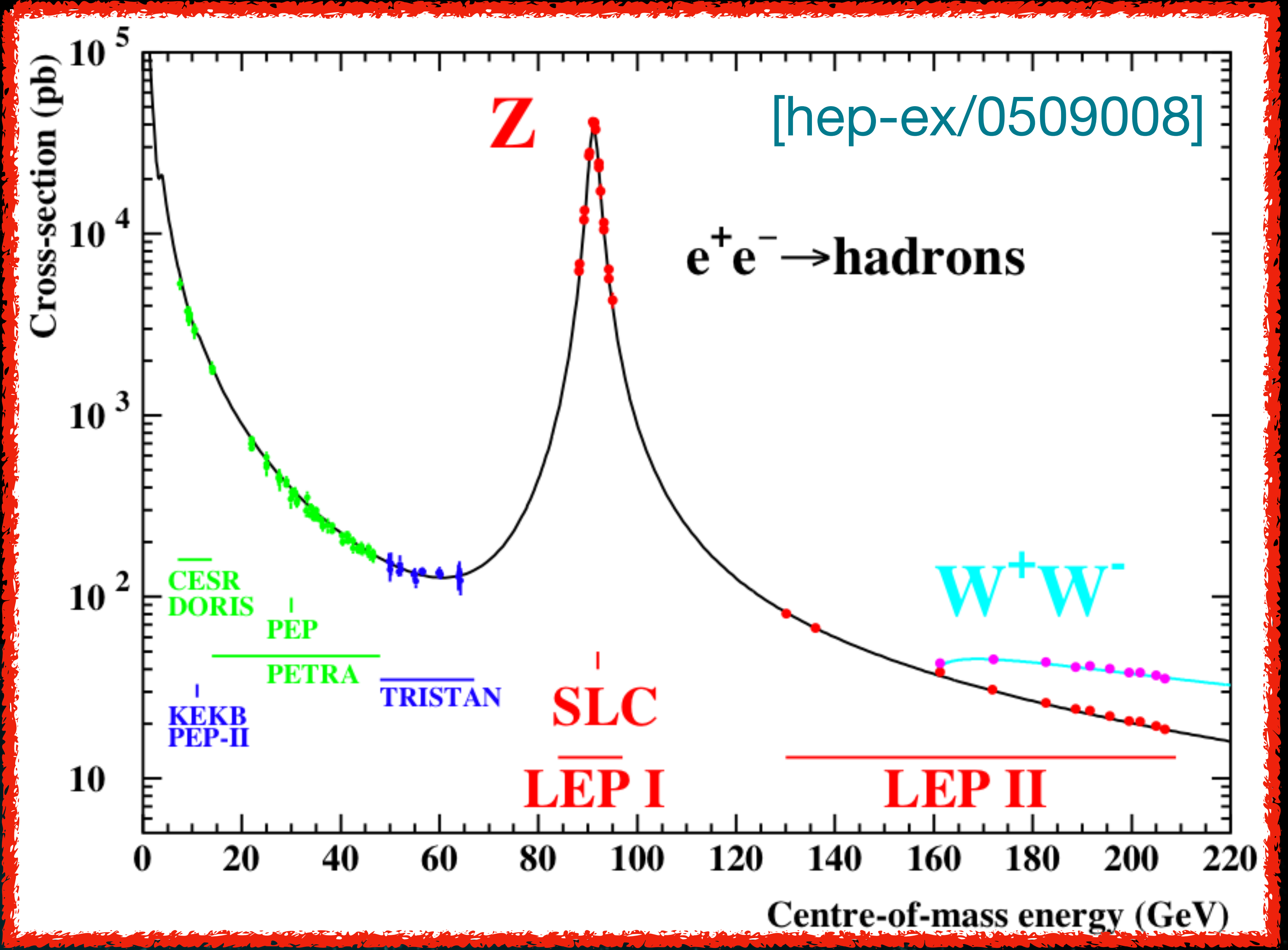


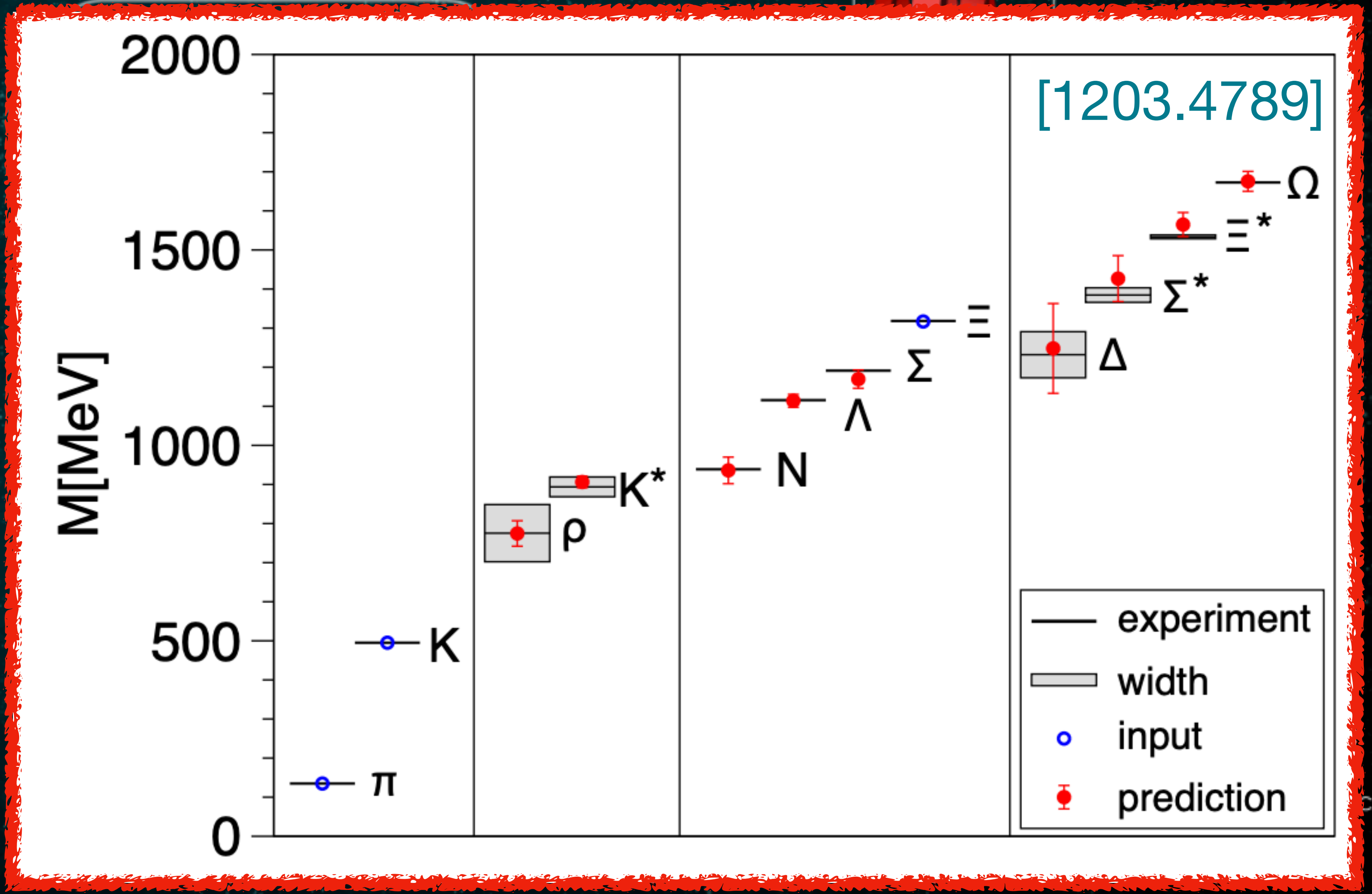
The Standard Model

A story of many spectacular successes…

- Unified description as a **quantum field theory** with only 3 fundamental forces and 19 free parameters
- Discovery of the Higgs in 2012
- **Precisely reproduced** cross-sections, decay rates, resonances



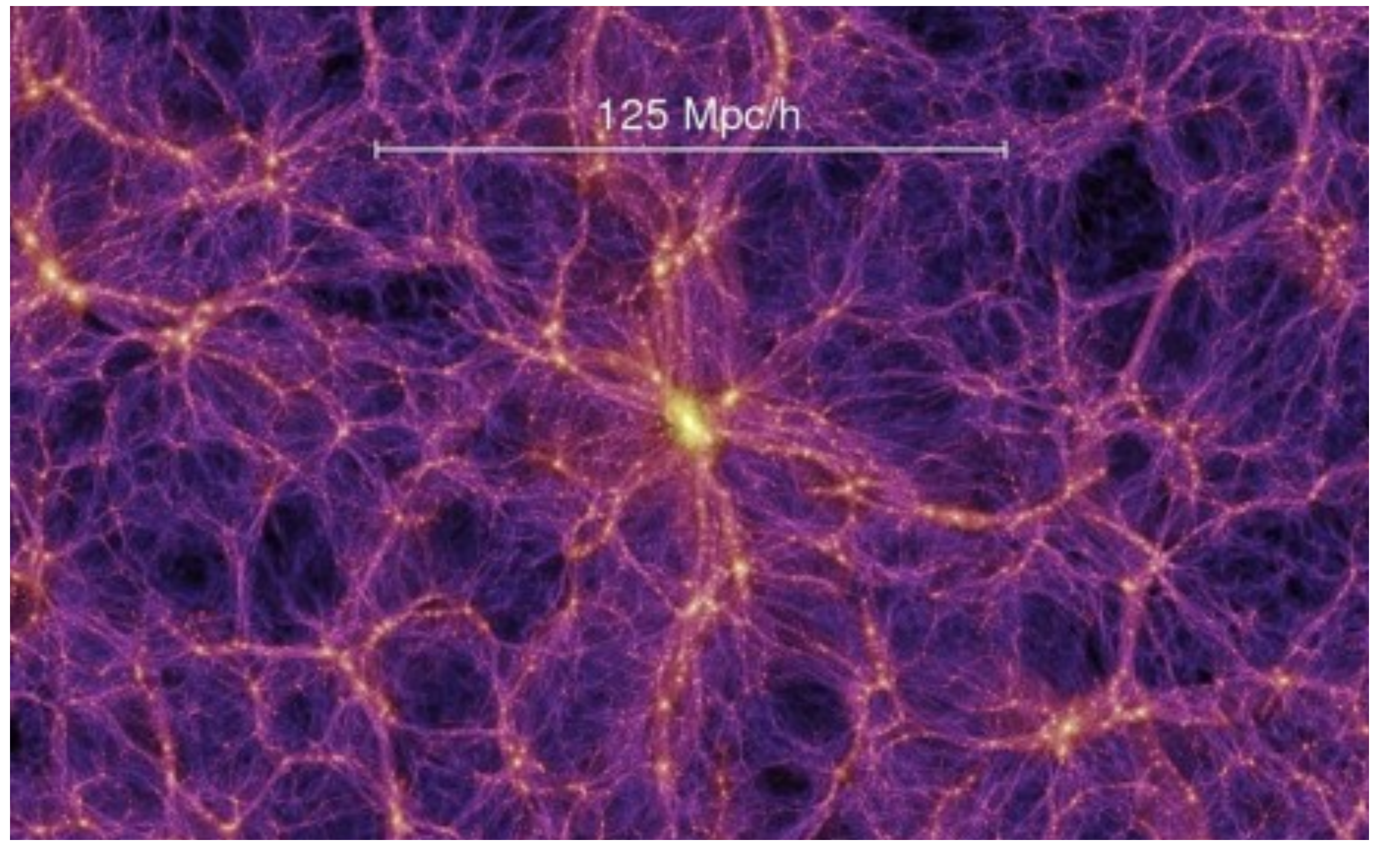




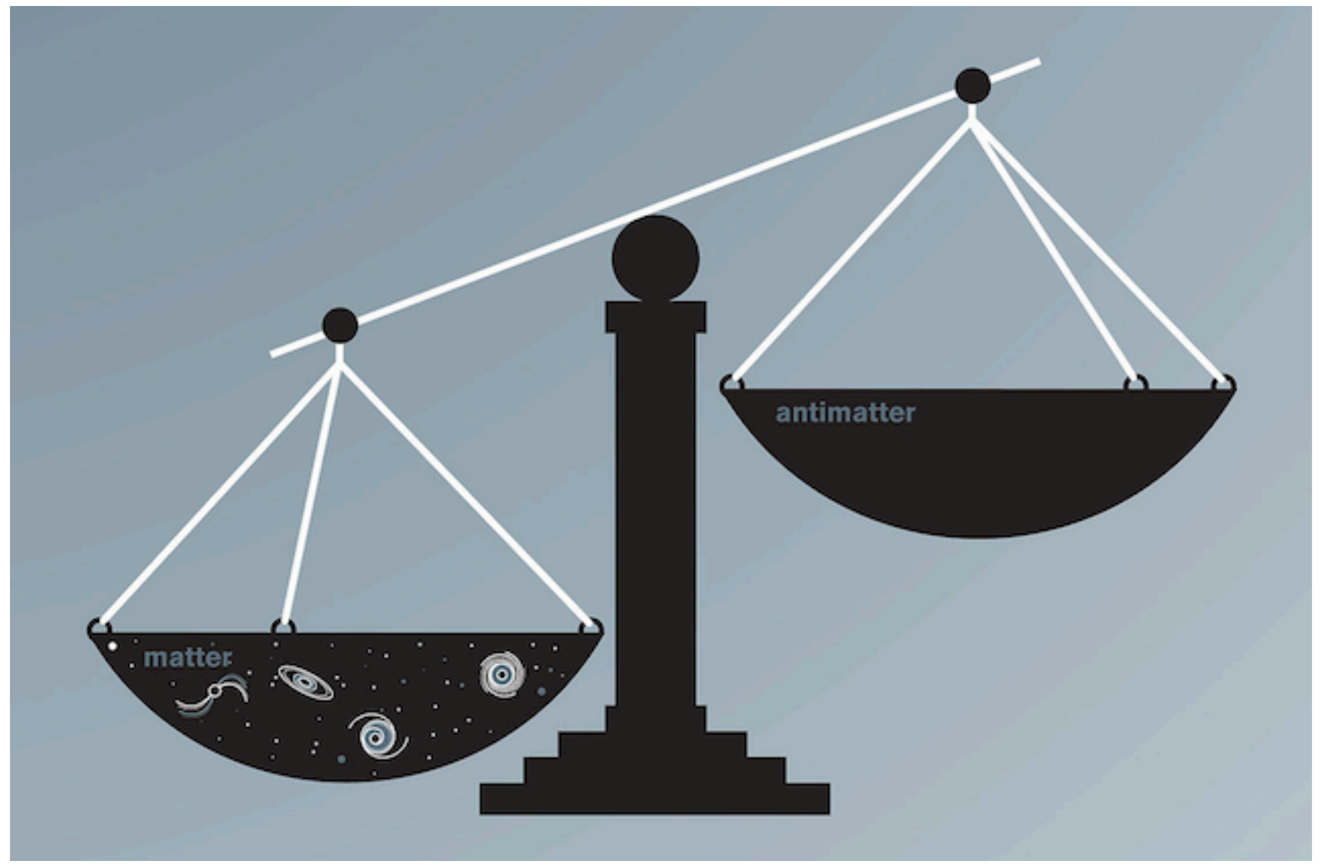
(Beyond) The Standard Model

… yet many unanswered questions remain.

What particles make up Dark Matter?
(Do particles make up Dark Matter?)



How should we account for neutrino masses?

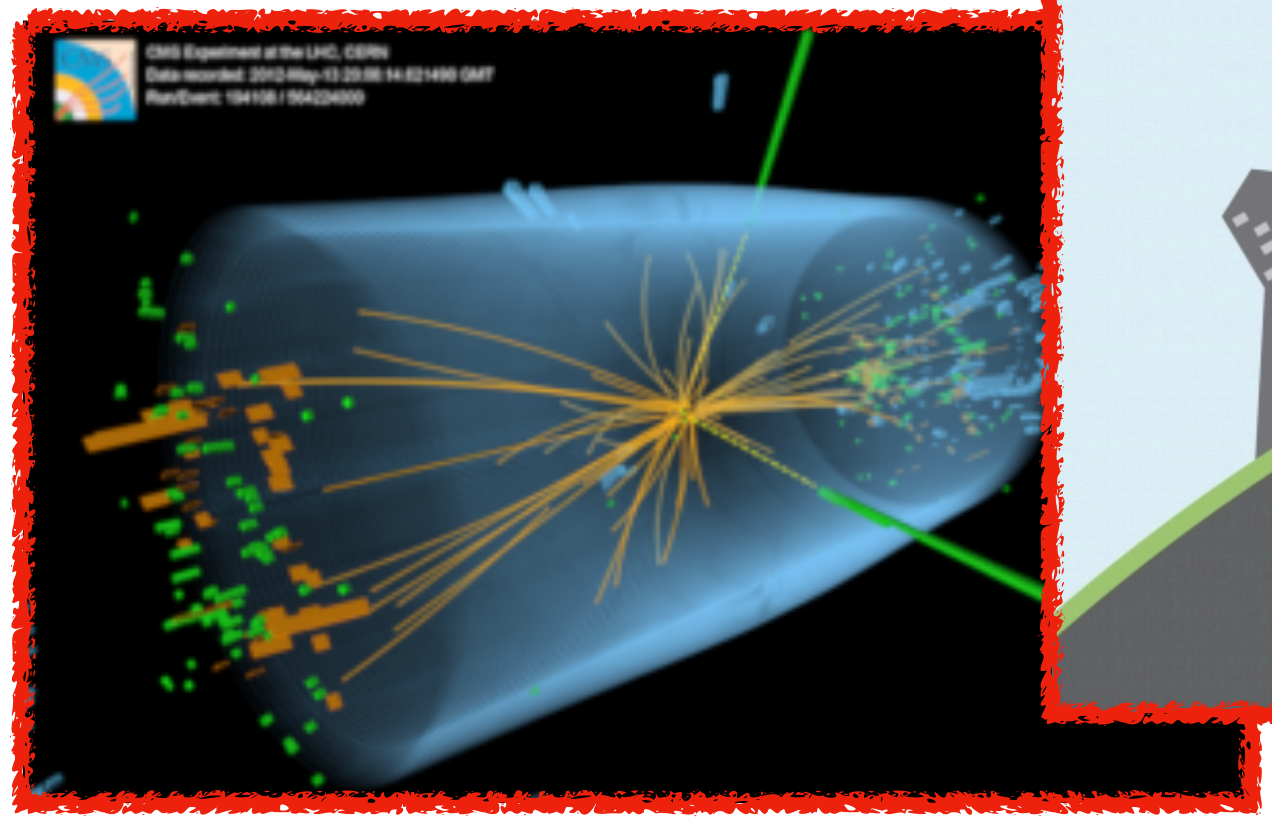


What is the origin of the matter-antimatter asymmetry?

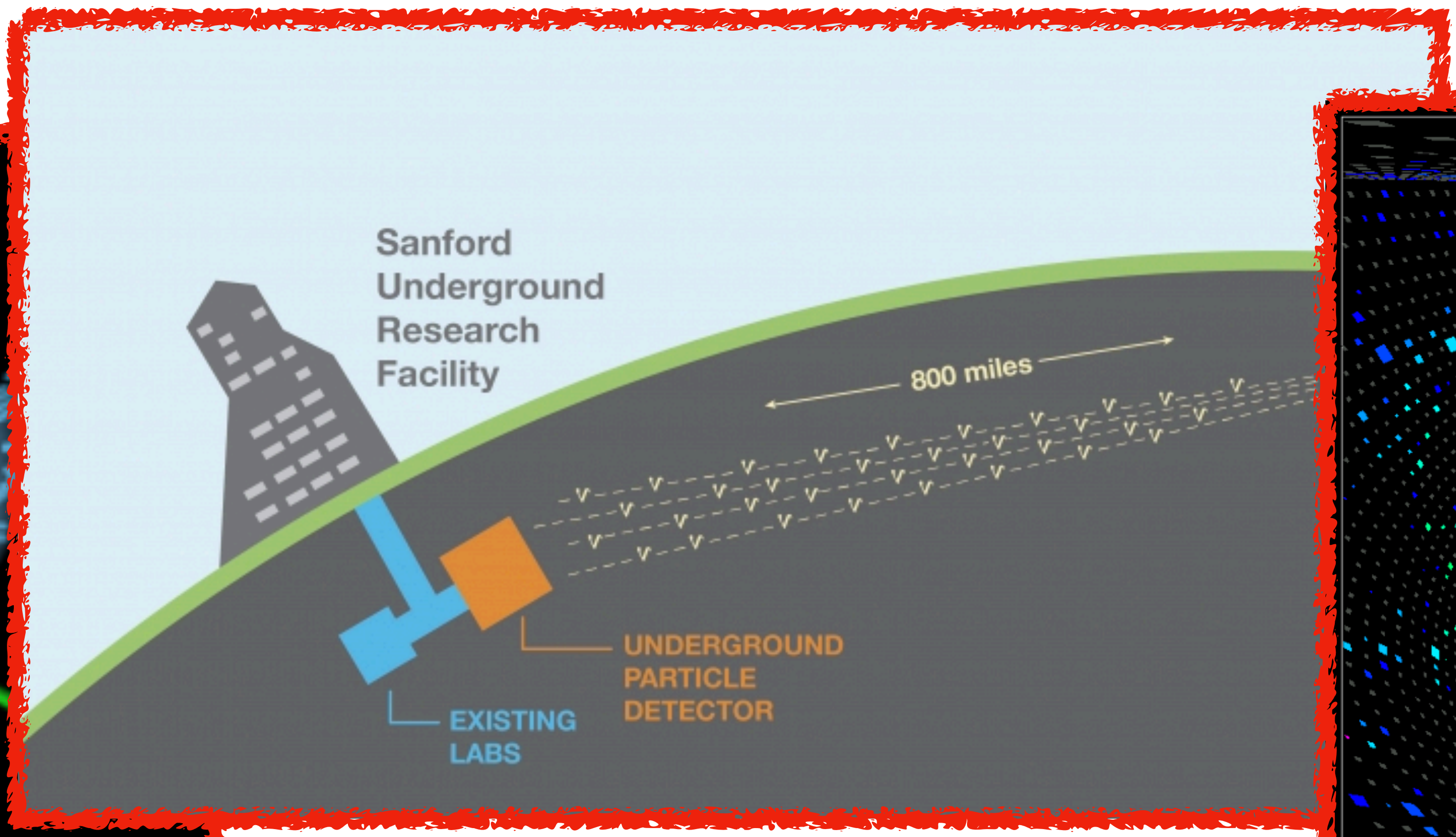
Springel et al, astro-ph/0504097v2
https://www.symmetrymagazine.org/article/how-heavy-is-a-neutrino
https://www.symmetrymagazine.org/article/october-2005/explain-it-in-60-seconds

(Beyond) The Standard Model

Answering these questions involves a joint **theoretical** and **experimental** effort.

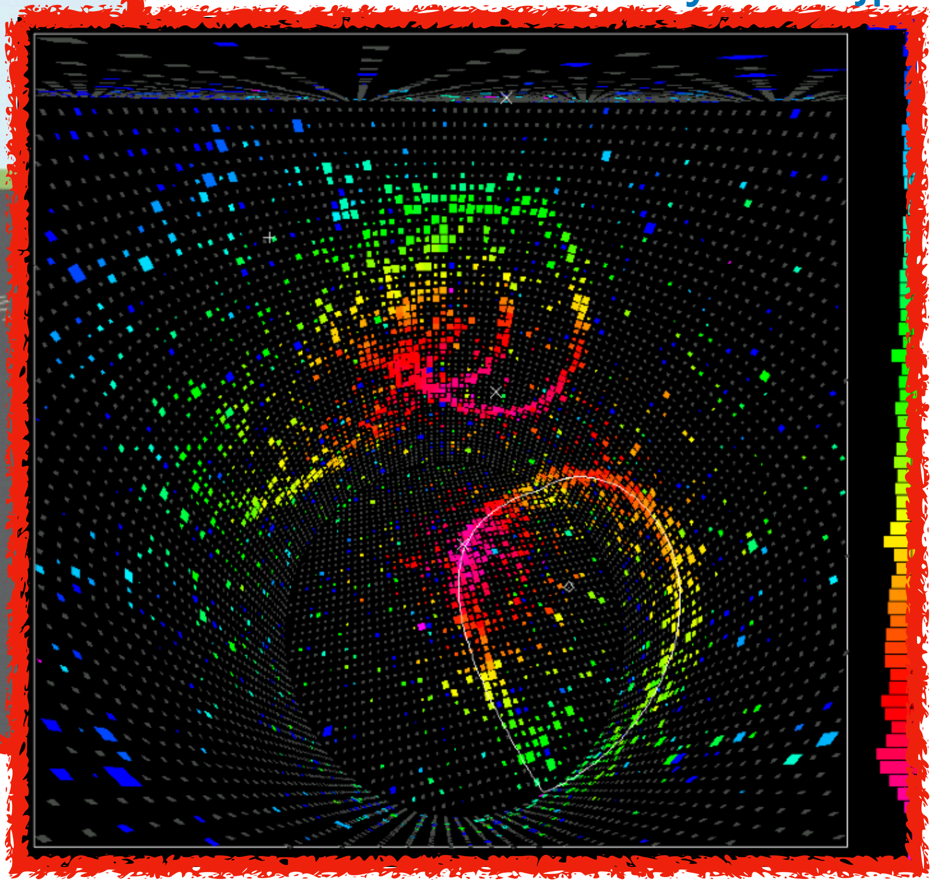


cds.cern.ch/images/OPEN-PHO-EXP-2013-003-1



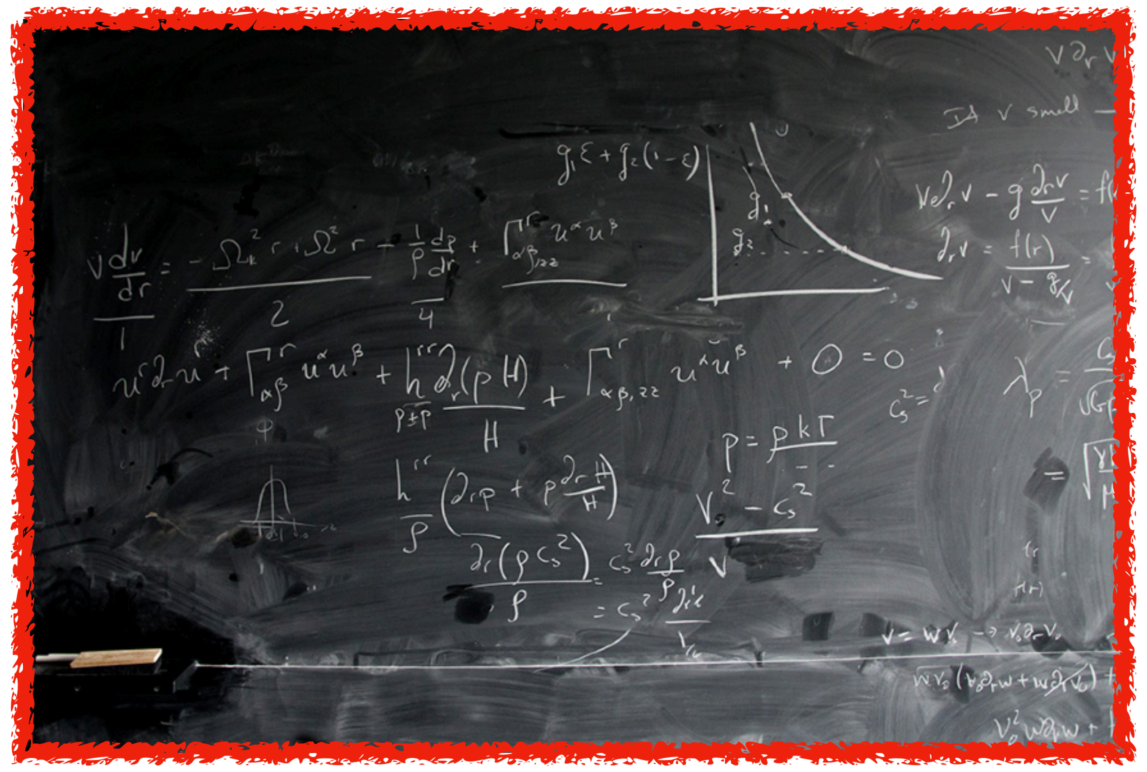
www.dunescience.org

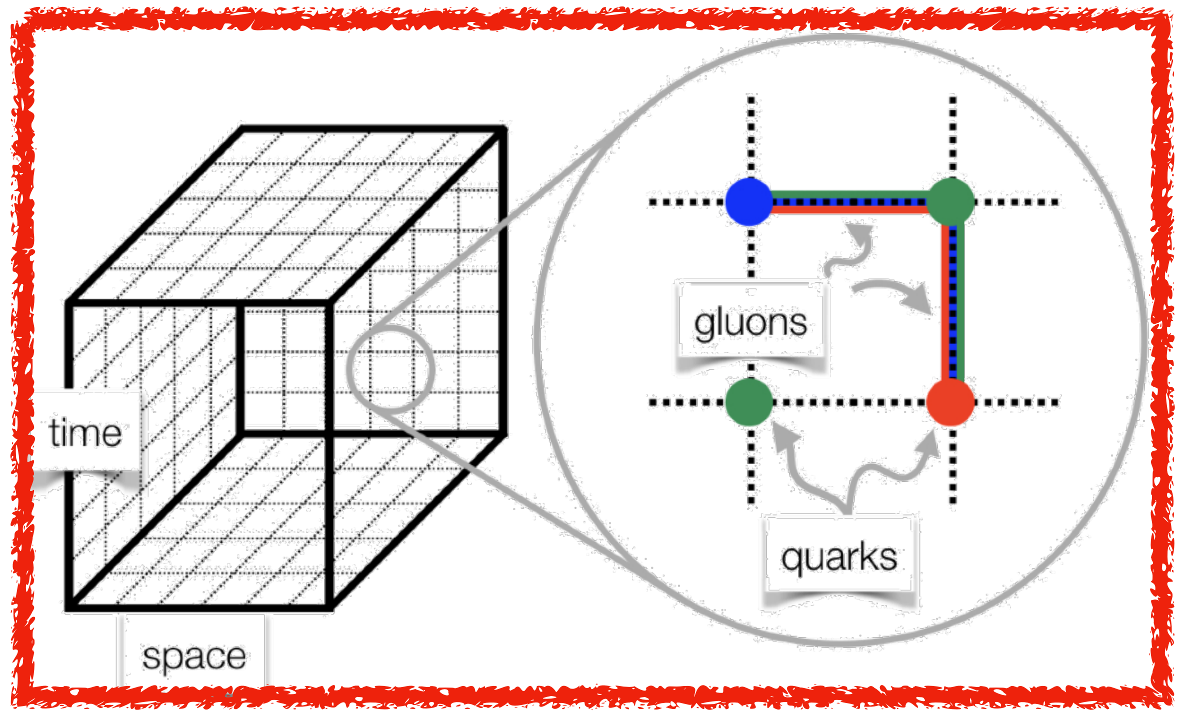
www-sk.icrr.u-tokyo.ac.jp



...

+

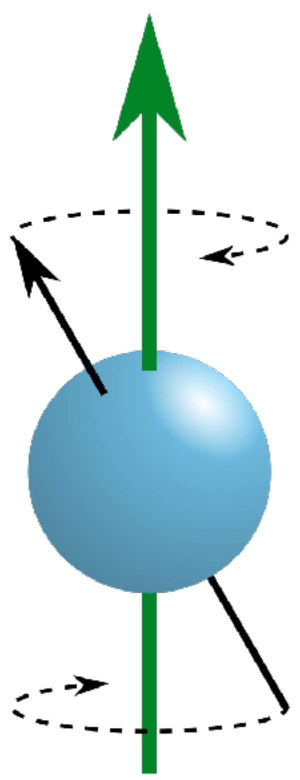




...

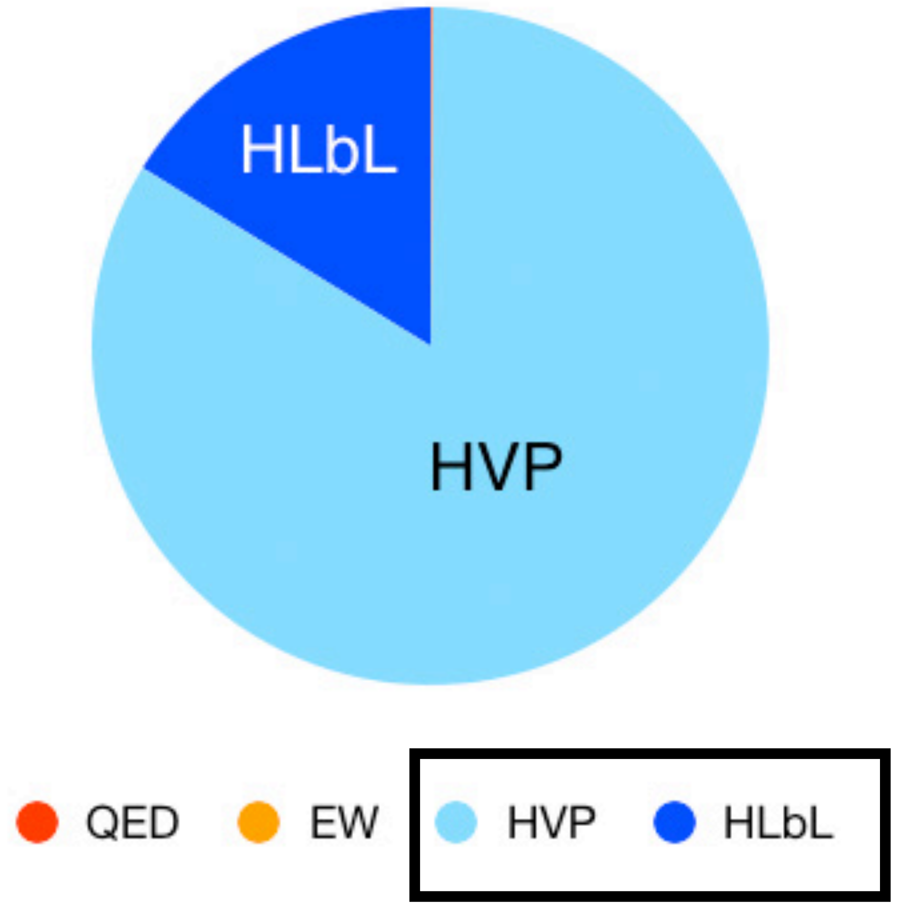
QCD in the search for new physics

QCD predictions are vital to interpreting experiments at the precision frontier.



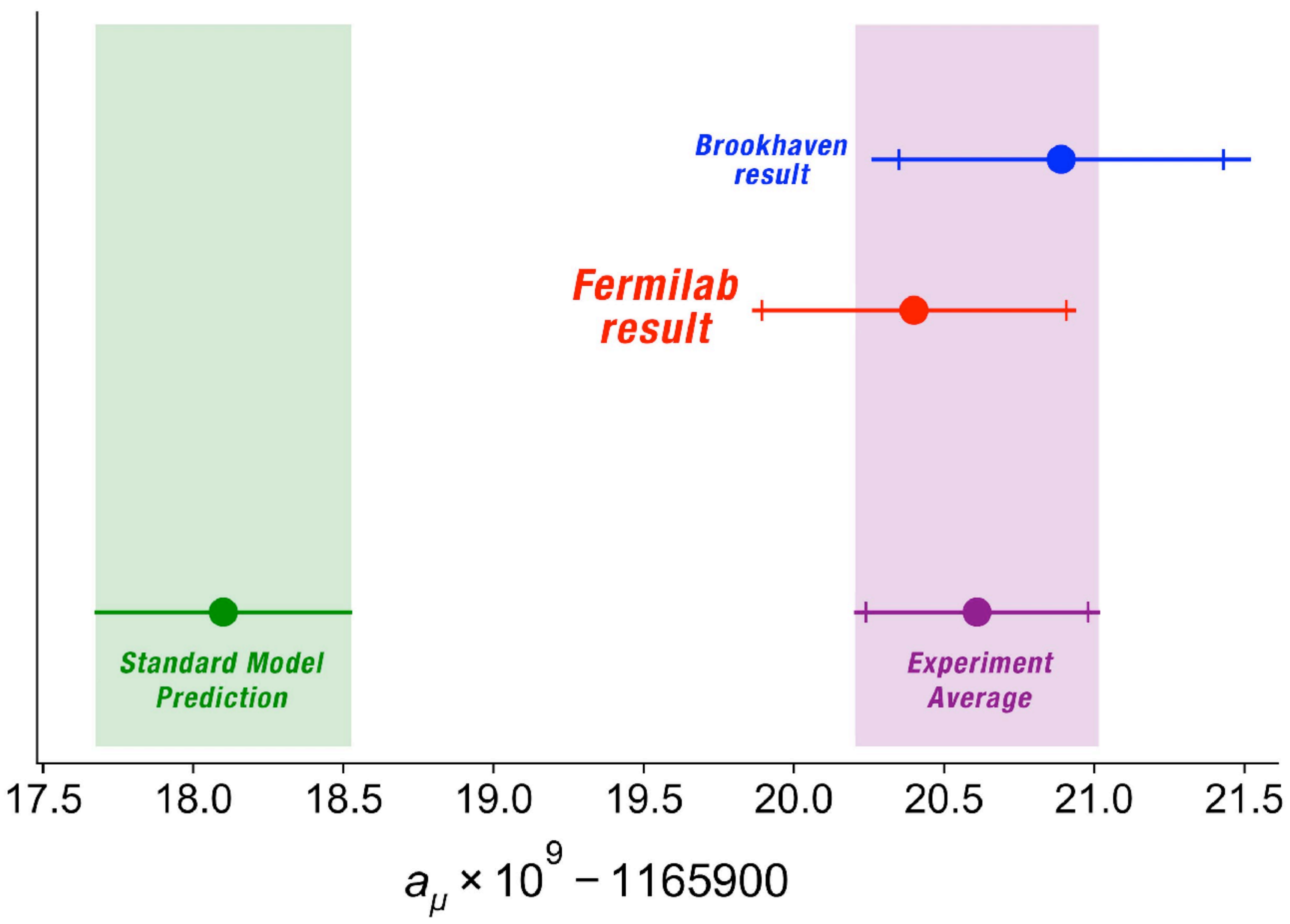
The anomalous magnetic moment (g-2) of the muon is a prime example of searching for new physics at the precision frontier

Standard Model theory error²



QCD contributions

2020

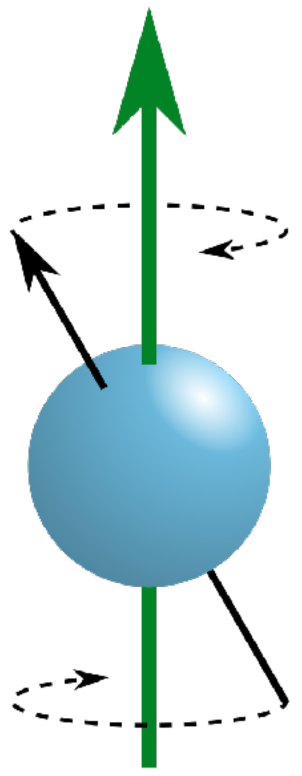


THEORY VS EXPERIMENT

Muon g-2 Collaboration, Press release (2023)
Muon g-2 Theory Initiative, WP (2020) 2006.04822

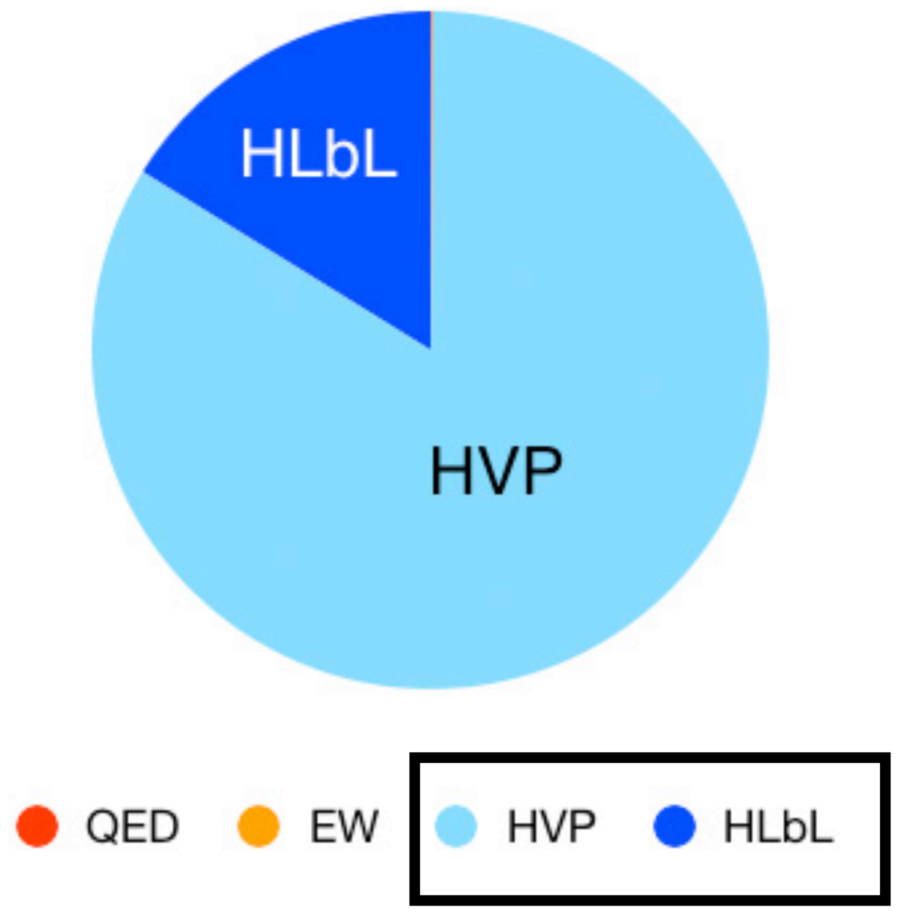
QCD in the search for new physics

QCD predictions are vital to interpreting experiments at the precision frontier.



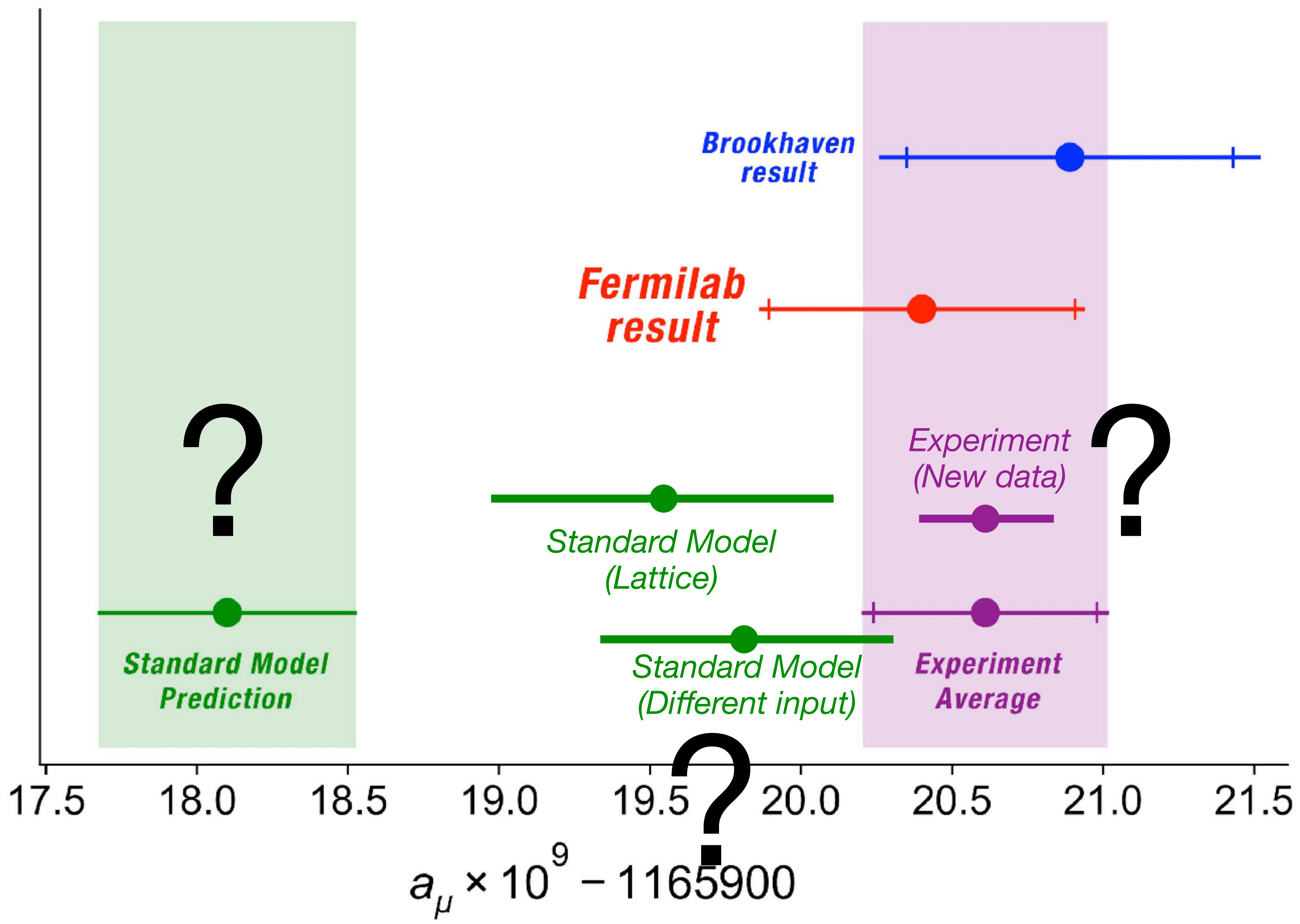
The anomalous magnetic moment (g-2) of the muon is a prime example of searching for new physics at the precision frontier

Standard Model theory error^2



QCD contributions

2023



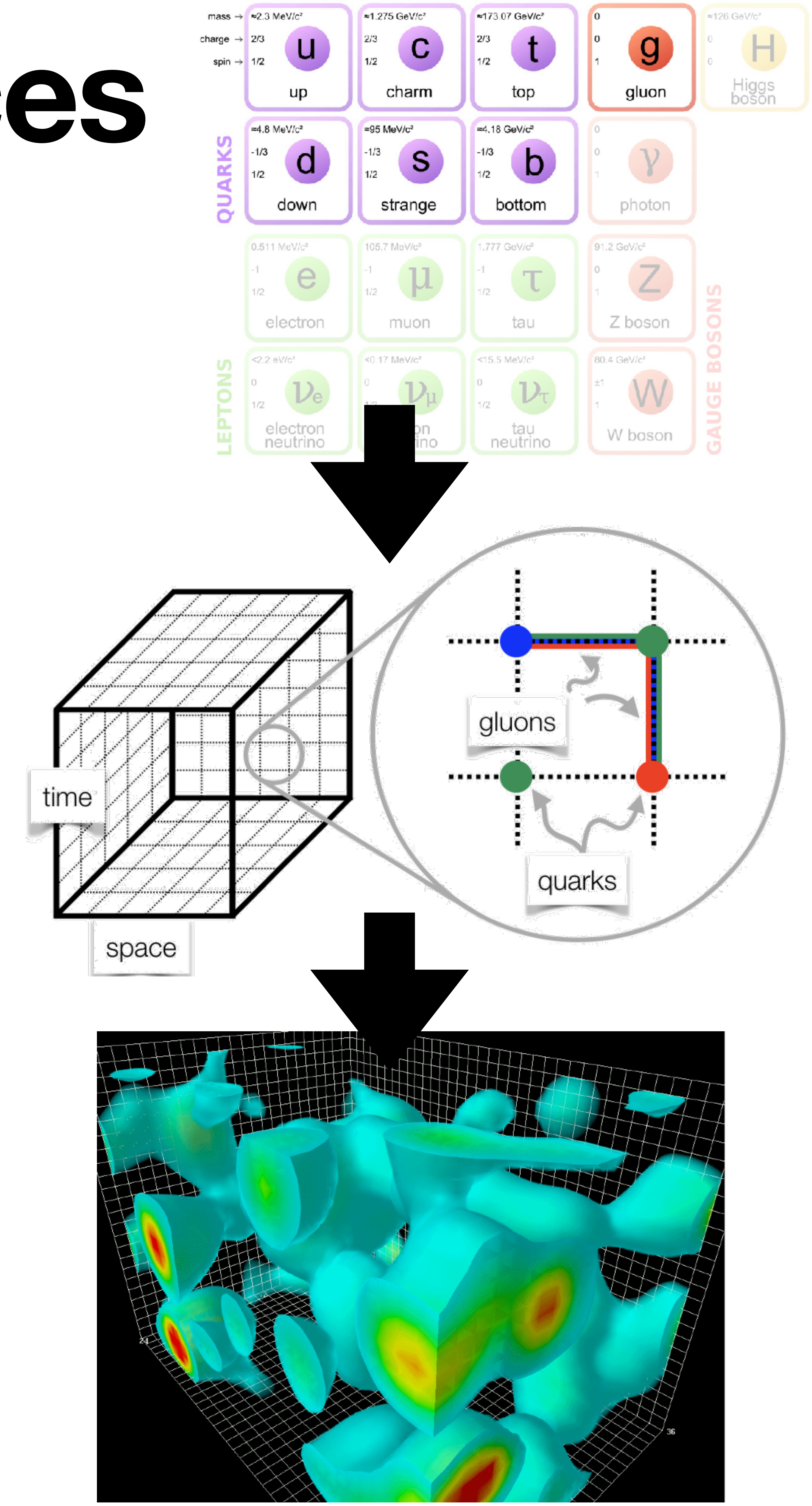
THEORY ? **VS** **EXPERIMENT**

Muon g-2 Collaboration, Press release (2023)
Muon g-2 Theory Initiative, WP (2020) 2006.04822

Quantum field theories on lattices

The Standard Model of particle physics and many candidates for more fundamental theories can be regularized with a spacetime lattice.

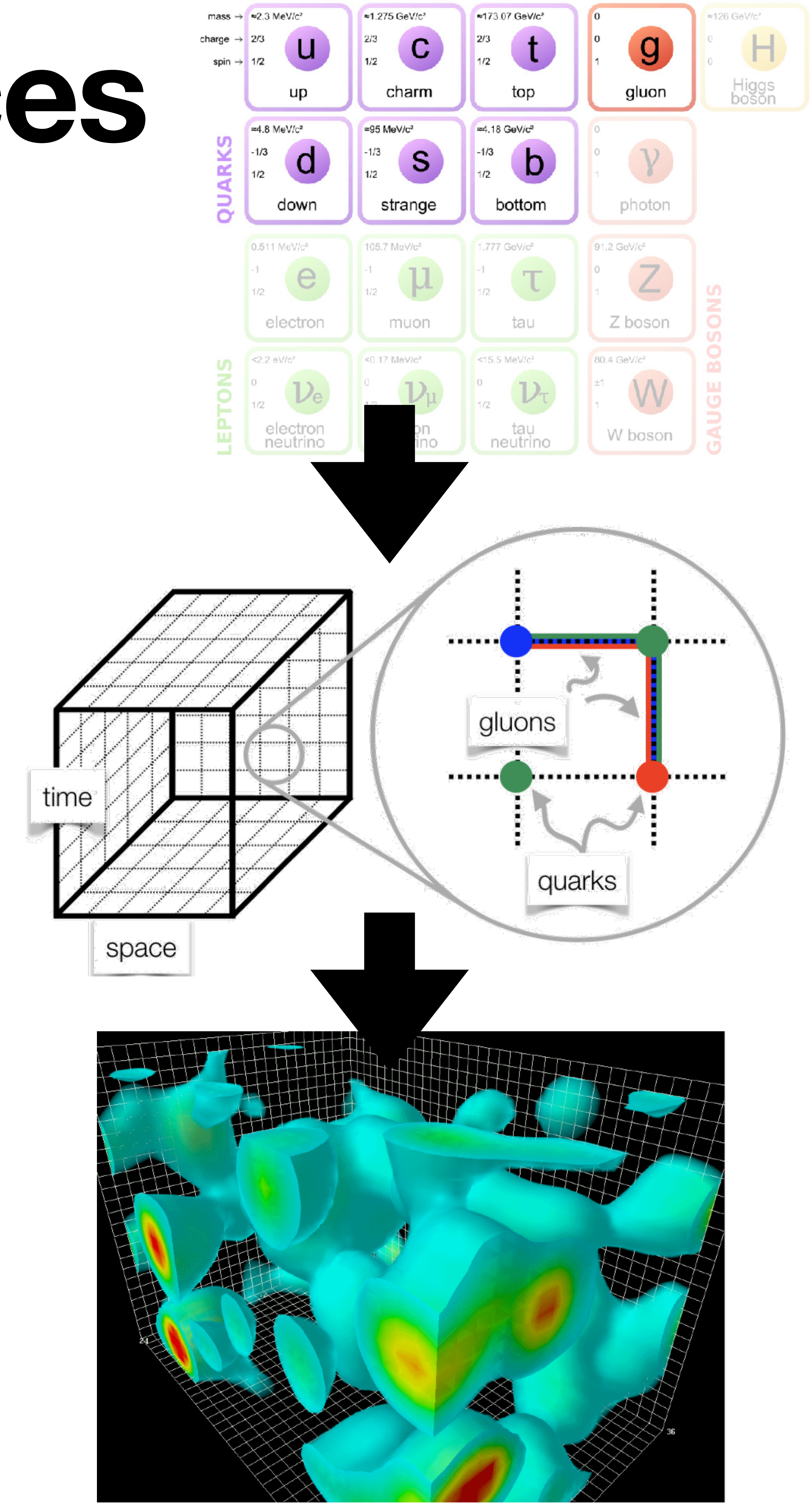
- **Perturbative** methods
  - Electromagnetic and weak force
- **Non-perturbative** methods
  - Strong nuclear force (Quantum Chromodynamics, QCD) at low energies
  - **Lattice QCD**: highly **complex** and **large-scale** numerical simulations



Quantum field theories on lattices

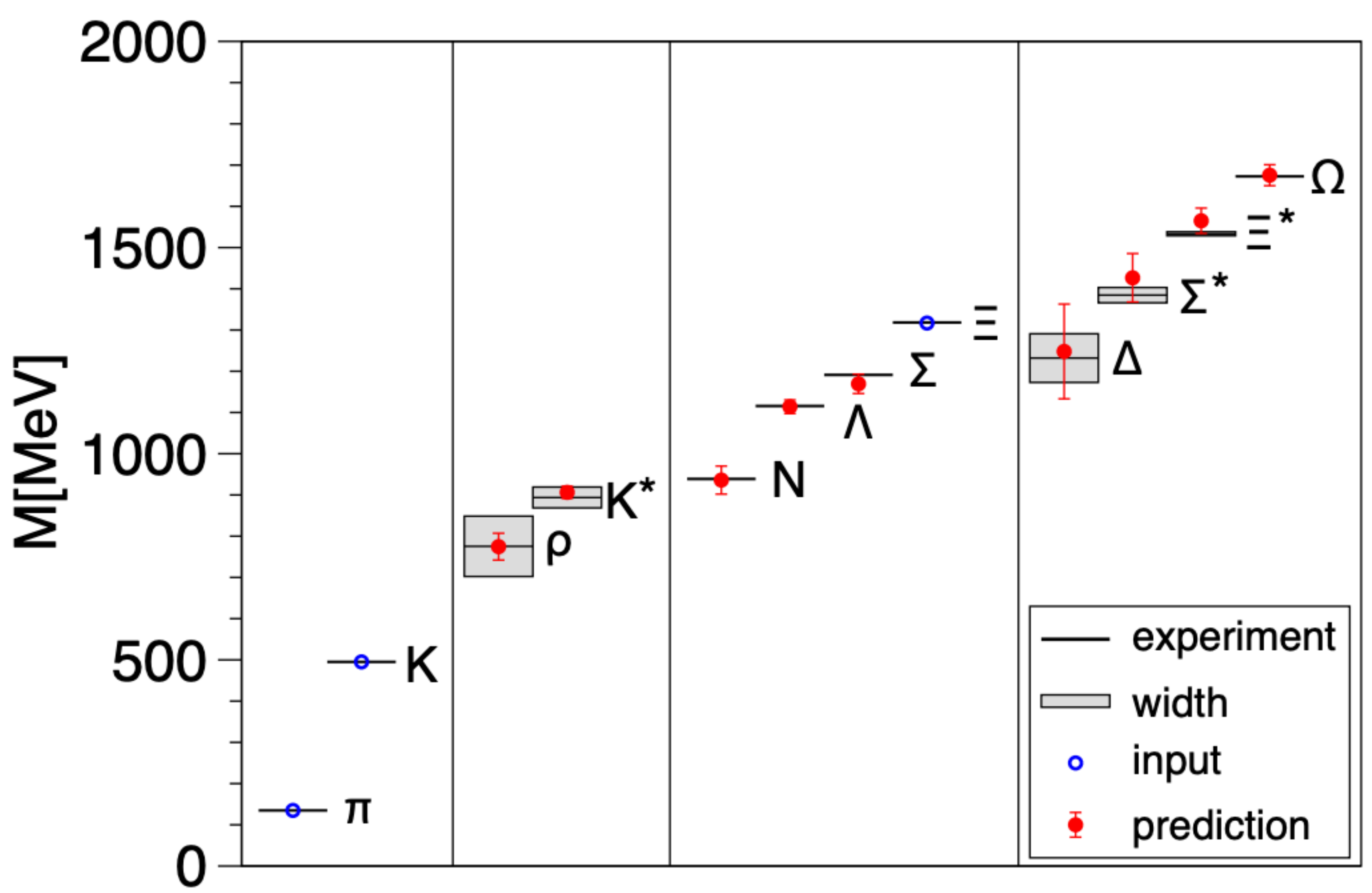
The Standard Model of particle physics and many candidates for more fundamental theories can be regularized with a spacetime lattice.

- **Perturbative** methods
  - Electromagnetic and weak force
- **Non-perturbative** methods
  - Strong nuclear force (Quantum Chromodynamics, QCD) at low energies
  - **Lattice QCD**: highly **complex** and **large-scale** numerical simulations

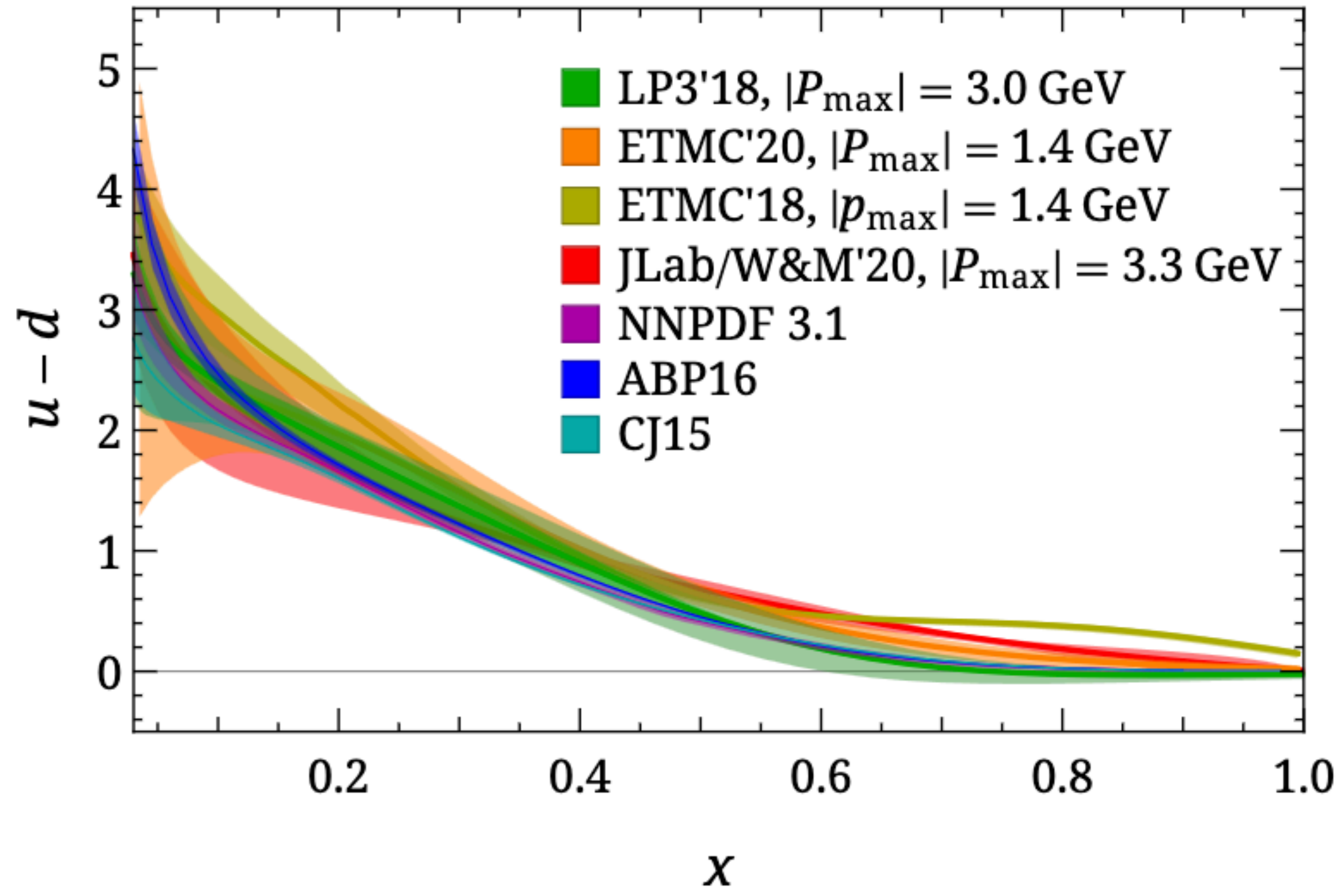


Lattice QCD

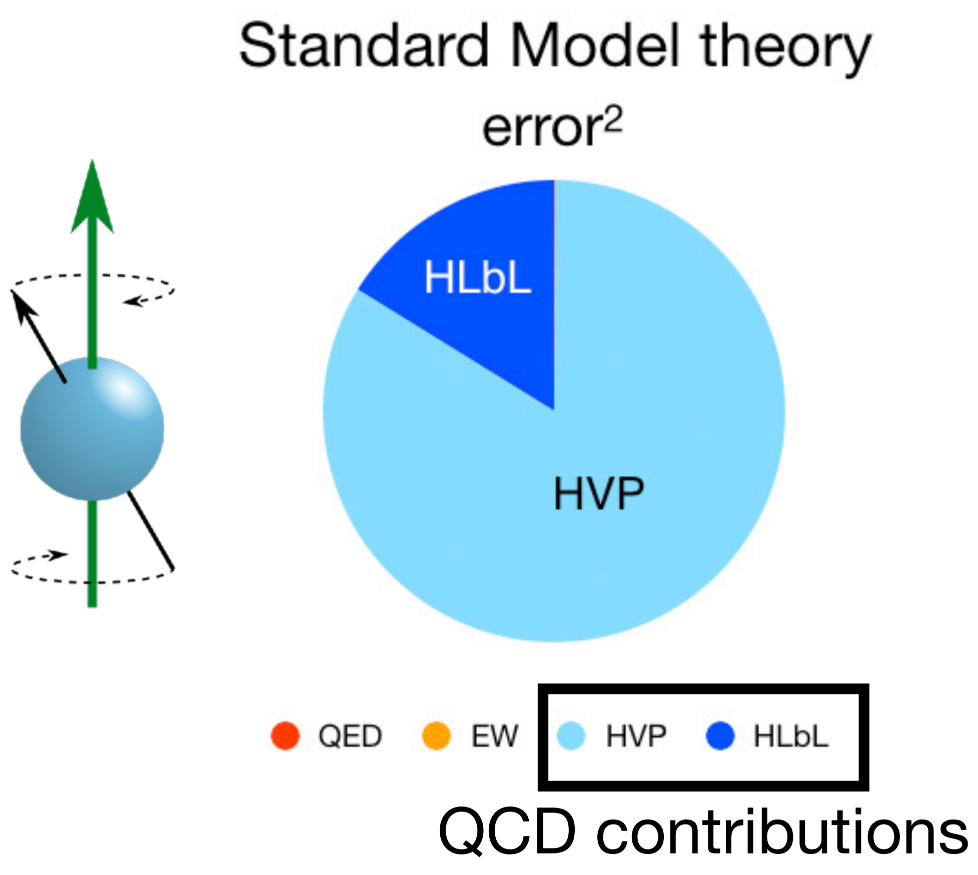
- Hadronic spectrum / structure
  - Heavy resonances
  - PDFs and their generalizations
  - Form factors
- QCD phase diagram
  - Critical point
  - Equation of state
- New physics searches
  - Muon g-2
  - Heavy meson decays
- …



Fodor & Hoelbling RMP84 (2012) 449



Constantinou+ 2006.08636



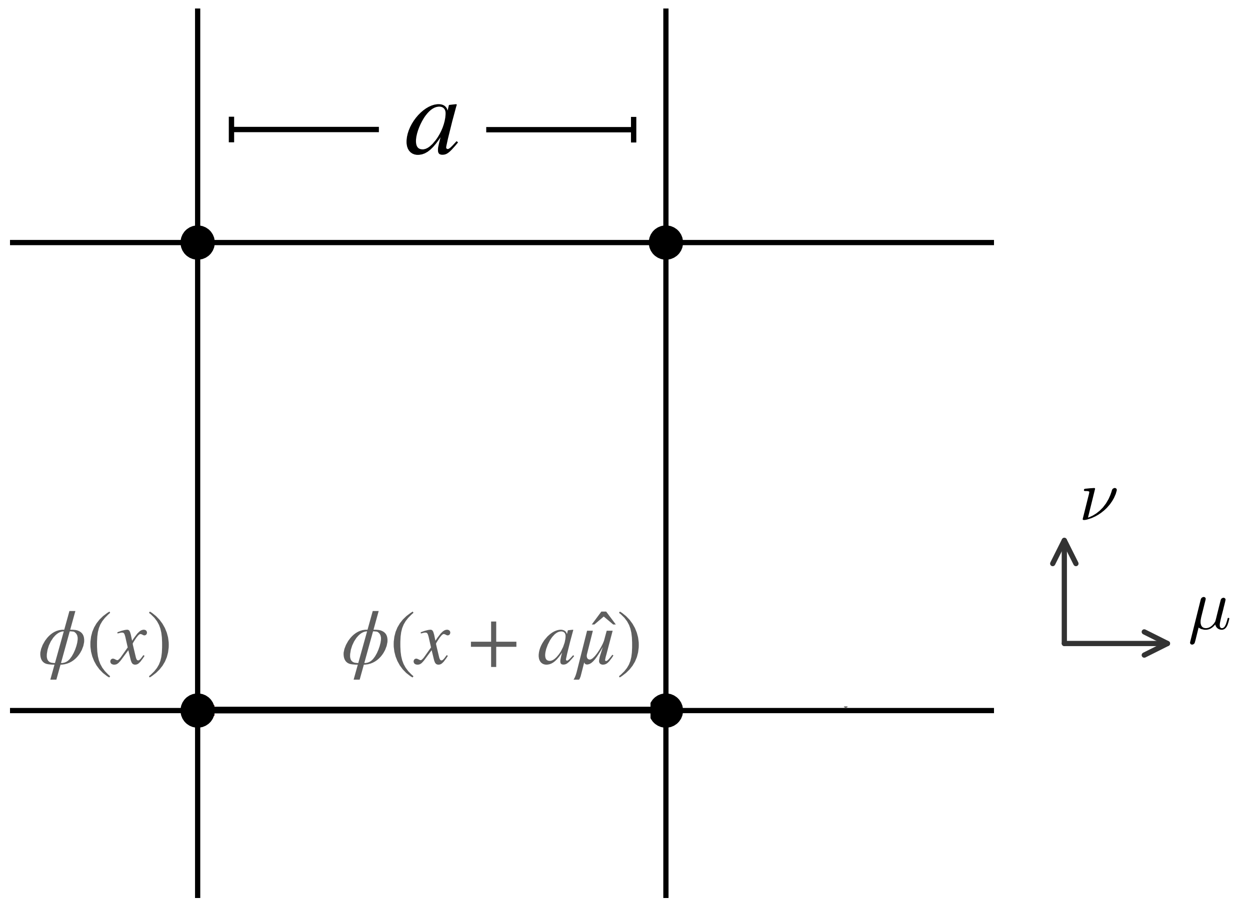
Muon g-2 Press release (2023)

Lattice quantum field theory

Lattice quantum field theory

Path integral definition of physical observables

- Euclidean spacetime $t \to i\tau$
- Discretized action S



$$S_E[\phi] = \int d^4x \, \frac{1}{2}\partial_\mu \phi(x) \partial^\mu \phi(x) + \frac{m^2}{2}\phi^2(x) + V(\phi(x))$$

$$\Rightarrow$$

$$S_E(\phi) = a^4 \sum_x \frac{1}{2} \sum_\mu \frac{\phi(x + a\hat{\mu}) - \phi(x)}{2a} + \frac{m^2}{2}\phi^2(x) + V(\phi(x))$$

Vaccum/thermal expt. value of quantum operator $\mathcal{O}$

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \Big[\prod_x \int_{-\infty}^{\infty} d\phi(x) \Big] \, \mathcal{O}(\phi) \, e^{-S_E(\phi)}$$

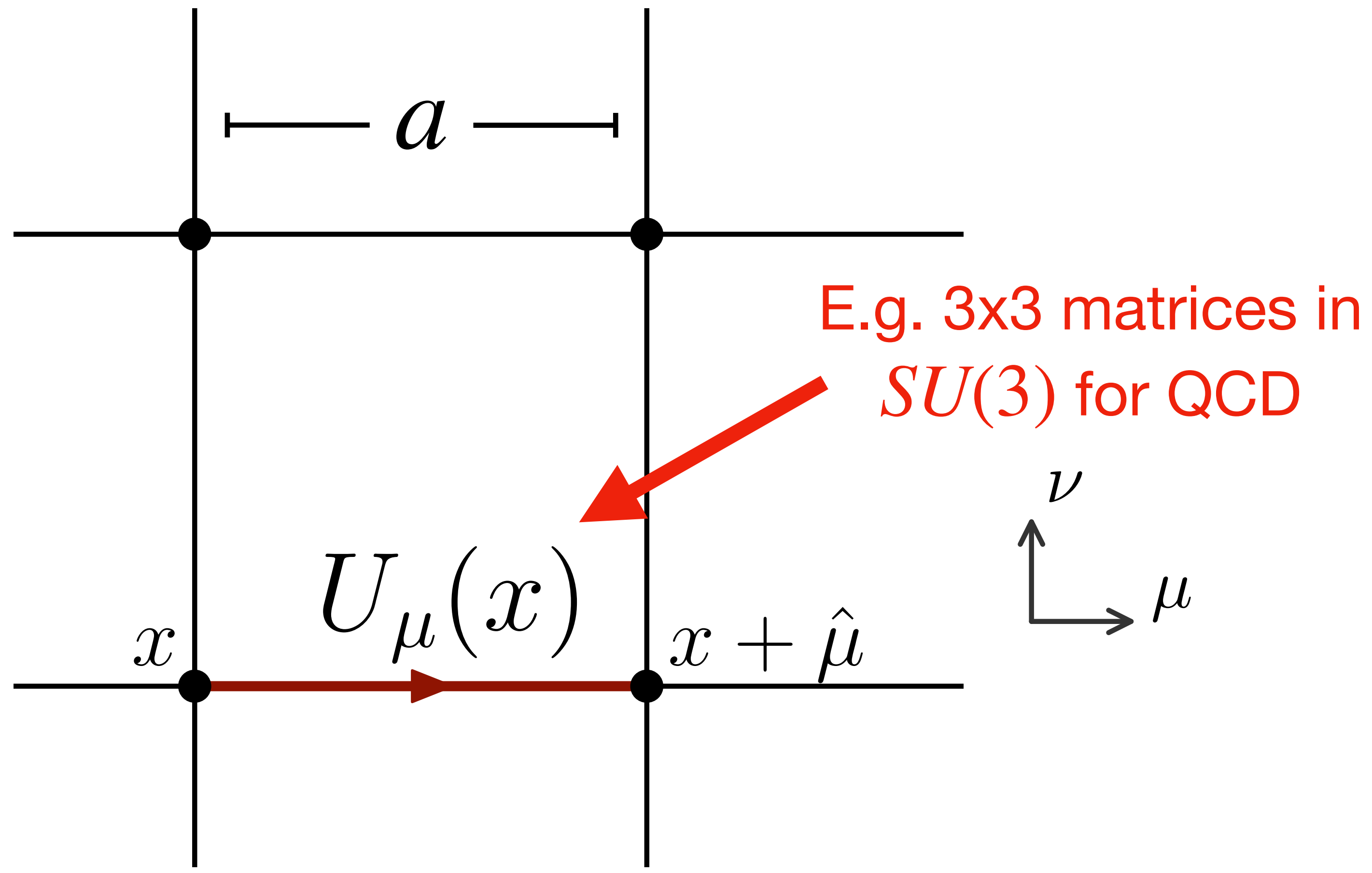
Partition function

$$Z \equiv \Big[\prod_x \int_{-\infty}^{\infty} d\phi(x) \Big] \, e^{-S_E(\phi)}$$

Lattice quantum field theory

Path integral definition of physical observables

- Euclidean spacetime $t \to i\tau$
- Discretized action S



Lattice QCD and other lattice gauge theories:

- Gauge group $G = \mathrm{U}(1)$ or $\mathrm{SU}(N)$ or …
- Gauge field discretized to link variables $U_\mu(x) \in G$

Vaccum/thermal expt. value of quantum operator $\mathcal{O}$

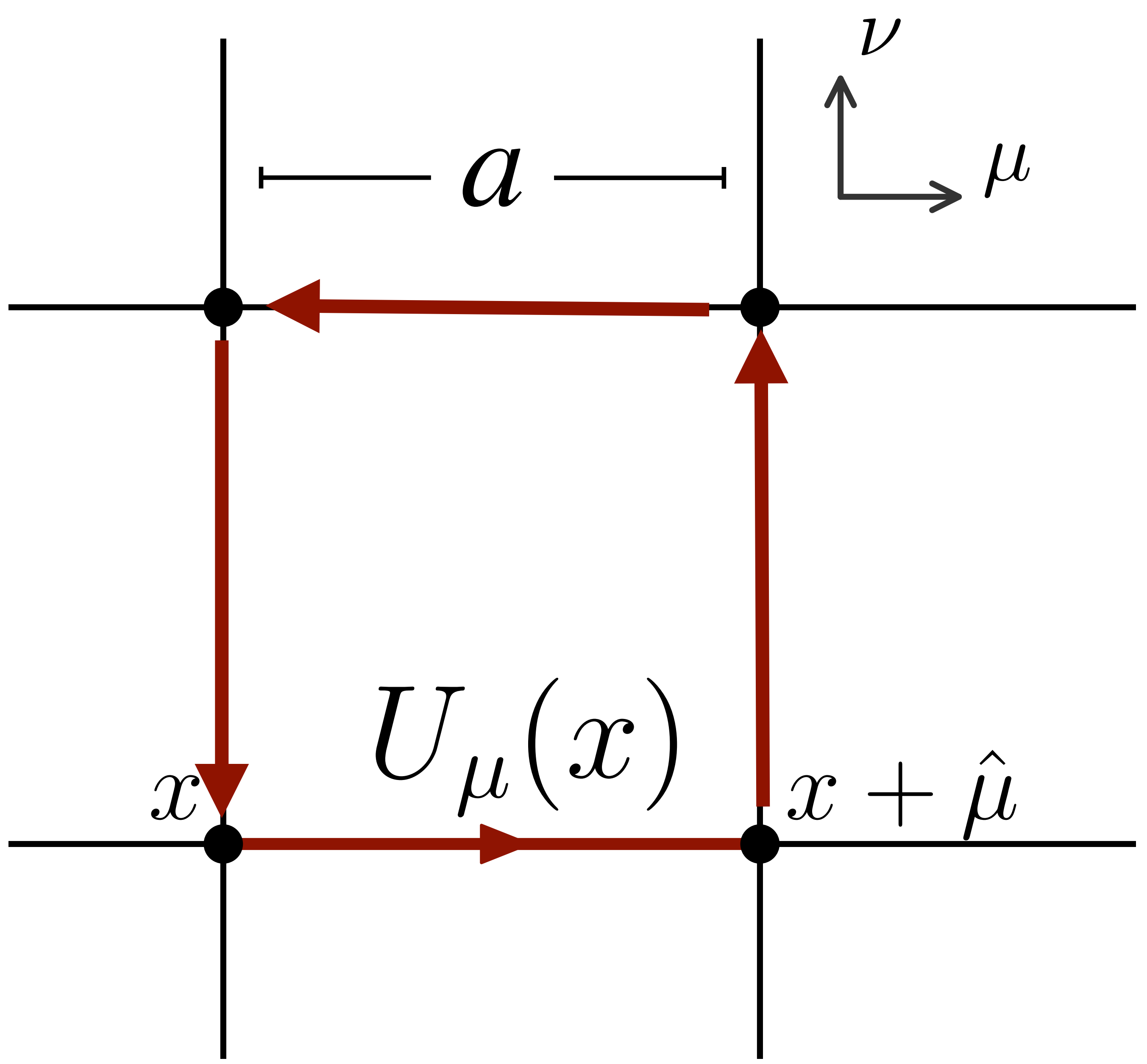
$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int \mathcal{D}U \, \mathcal{O}(U) \, e^{-S(U)}$$

$$Z = \int \mathcal{D}U \, e^{-S(U)}, \quad \int \mathcal{D}U = \prod_{x,\mu} \int dU_\mu(x)$$

Partition function — Path integral measure

Wilson action for gluon dynamics

$$S(U) = -\frac{\beta}{3}\sum_{x}\sum_{\mu<\nu} \mathrm{ReTr}\, P_{\mu\nu}(x)$$



$$P_{\mu\nu}(x) = U_\mu(x)U_\nu(x+\hat{\mu})U_\mu^\dagger(x+\hat{\nu})U_\nu^\dagger(x)$$

- Gluon self-interaction dynamics (Yang-Mills)
- Confinement, topological instantons
- Gauge symmetry $U_\mu(x) \mapsto \Omega(x)U_\mu(x)\Omega^\dagger(x+\hat{\mu})$

Monte Carlo simulation

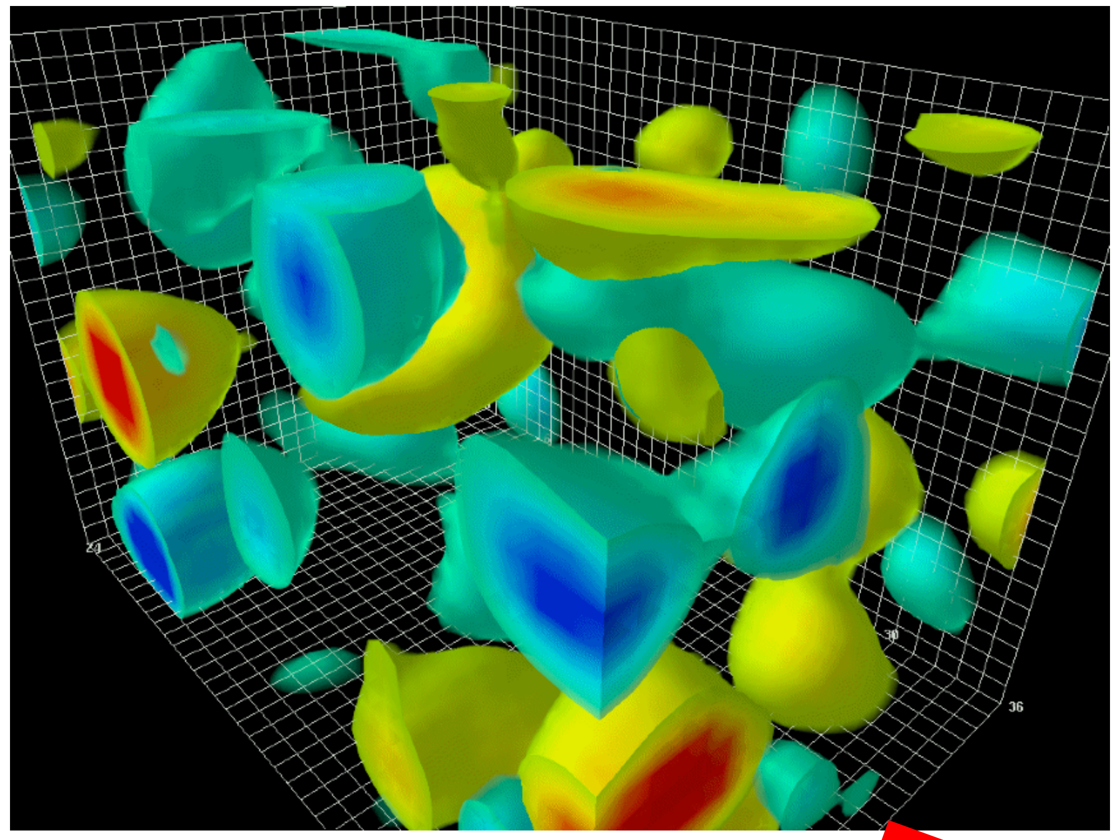
$$\langle \mathcal{O} \rangle = \left[\prod_{x,\mu} \int dU_\mu(x) \right] \mathcal{O}(U)\, e^{-S(U)}/Z$$

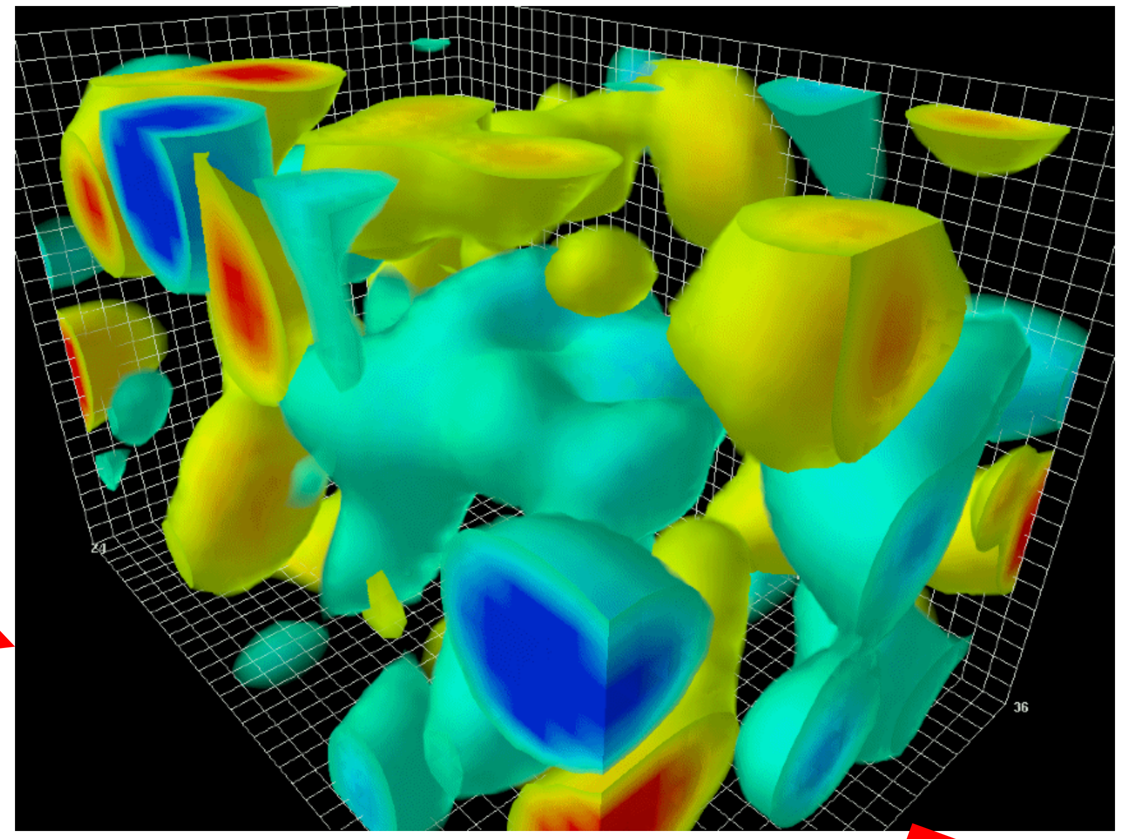
Approximate the path integral using **Markov chain Monte Carlo**

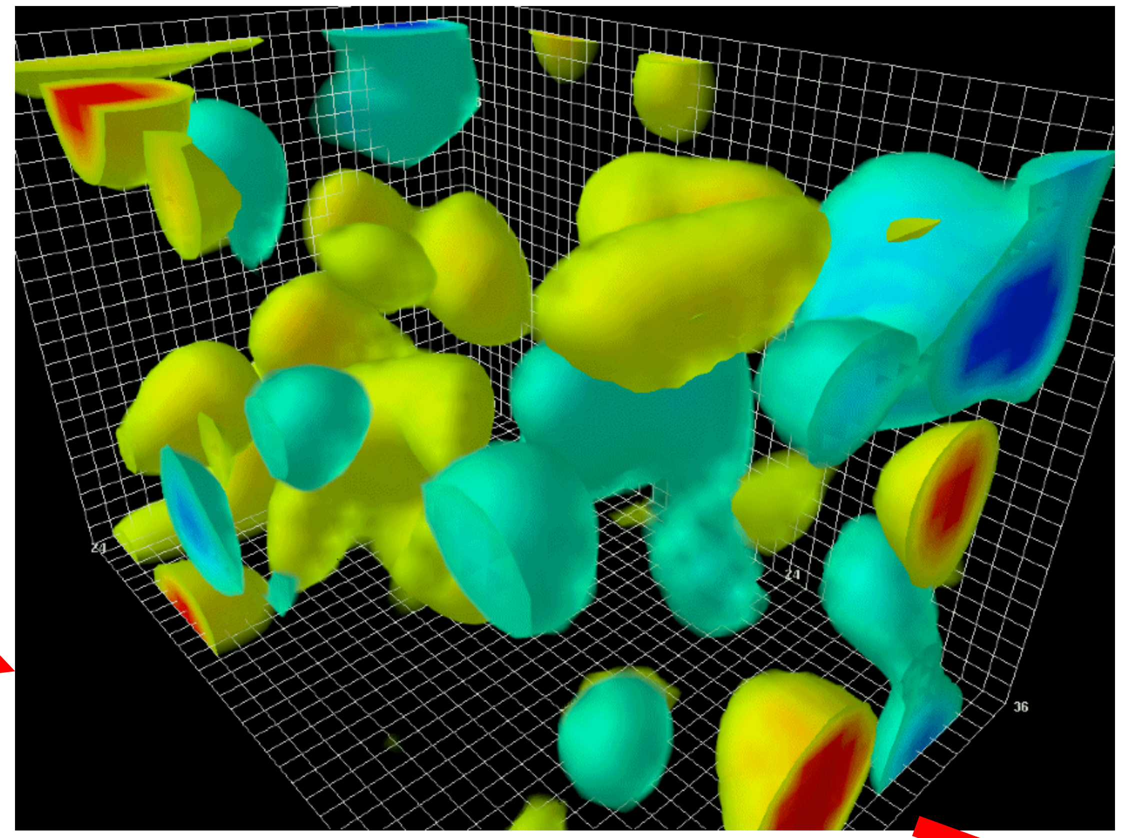
Positive integrand allows interpreting path integral weights as a probability measure:

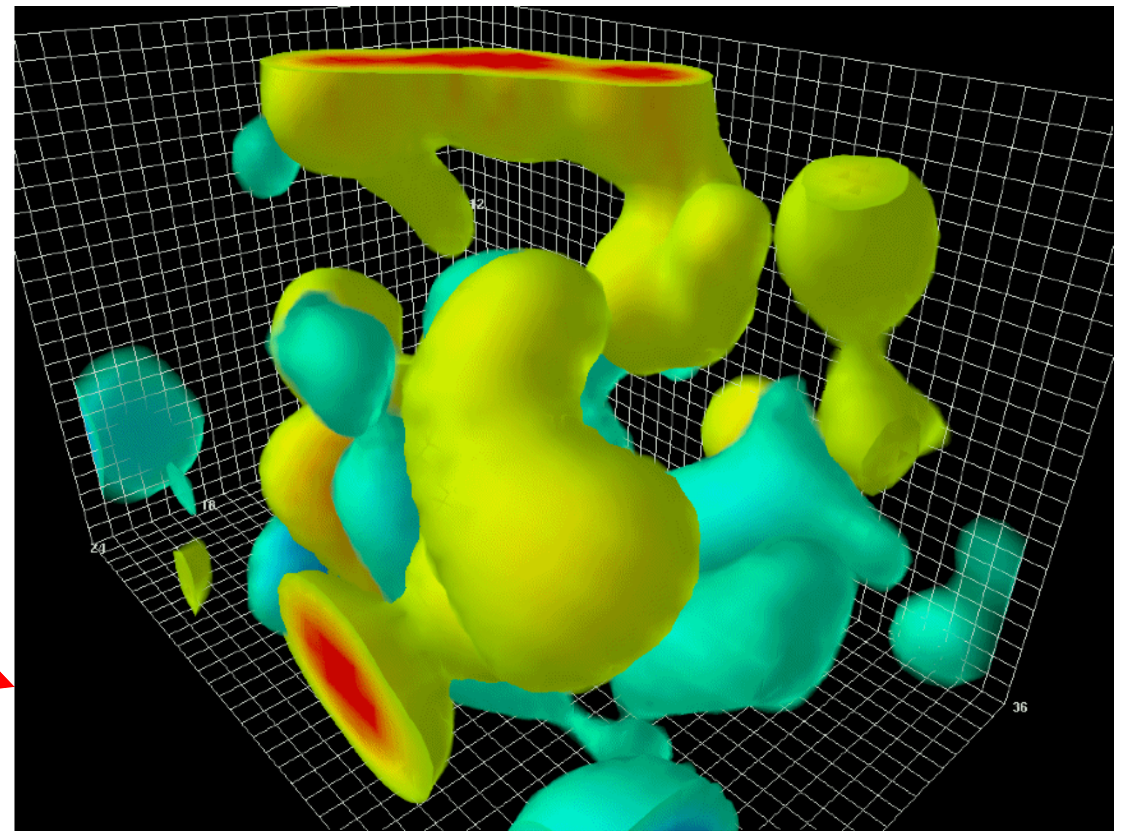
$$U_i \sim p(U) = e^{-S(U)}/Z$$

$$\langle \mathcal{O} \rangle \approx \frac{1}{n} \sum_{i=1}^{n} \mathcal{O}(U_i)$$









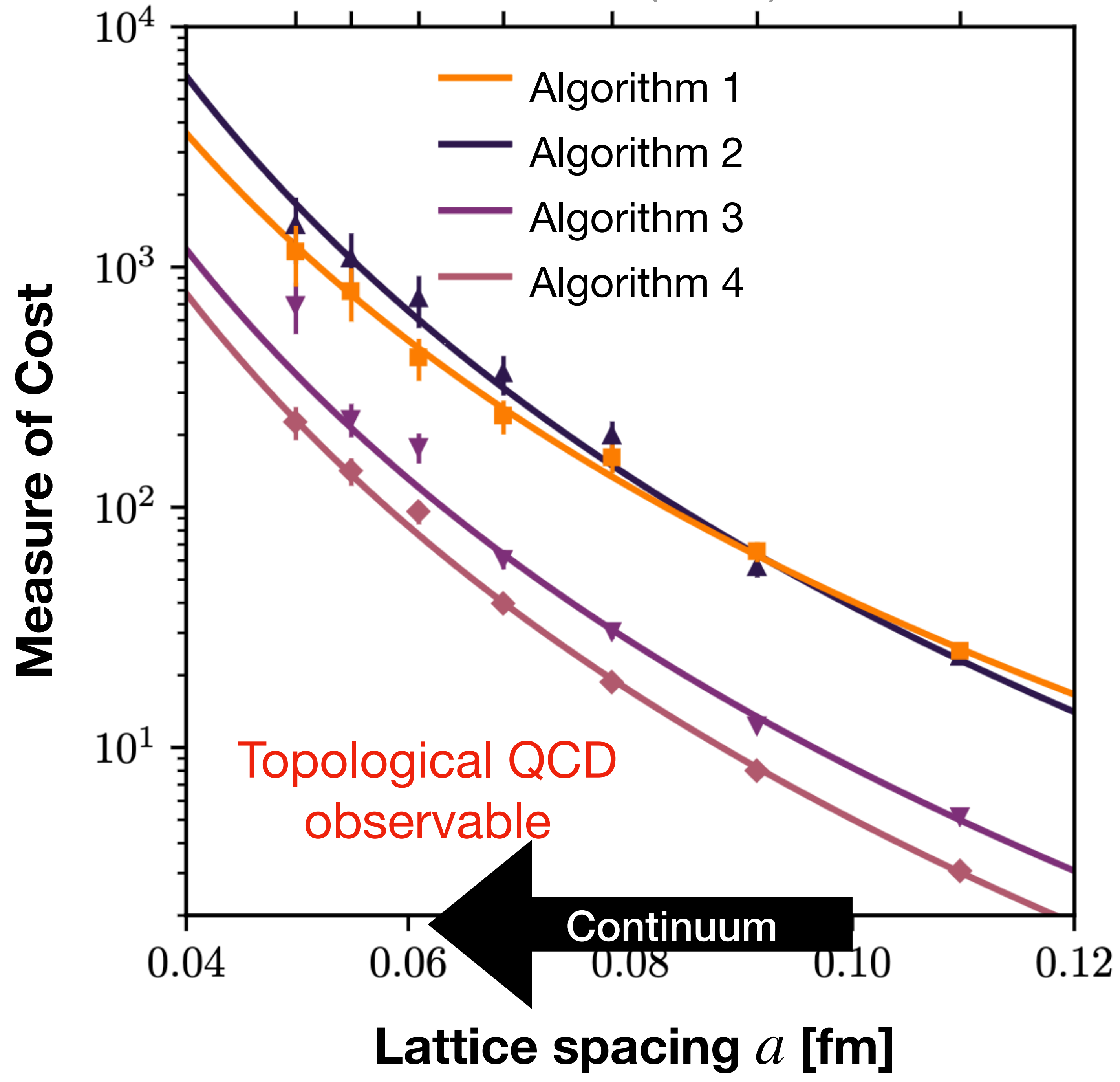
Why generative models?

State-of-the-art LGT calculations require **enormous computational cost.**

- $\gtrsim 10^9$ degrees of freedom
- "Critical slowing down" as $a \to 0$
- Costly matrix inversion for propagators $\langle \psi \bar{\psi} \rangle$ (especially as $m_q \to 0$)

This **limits the precision of physics results** (challenging uncertainties from $a \to 0$, $m_\pi \to \sim 140\text{MeV}$, and $V \to \infty$ limits!)

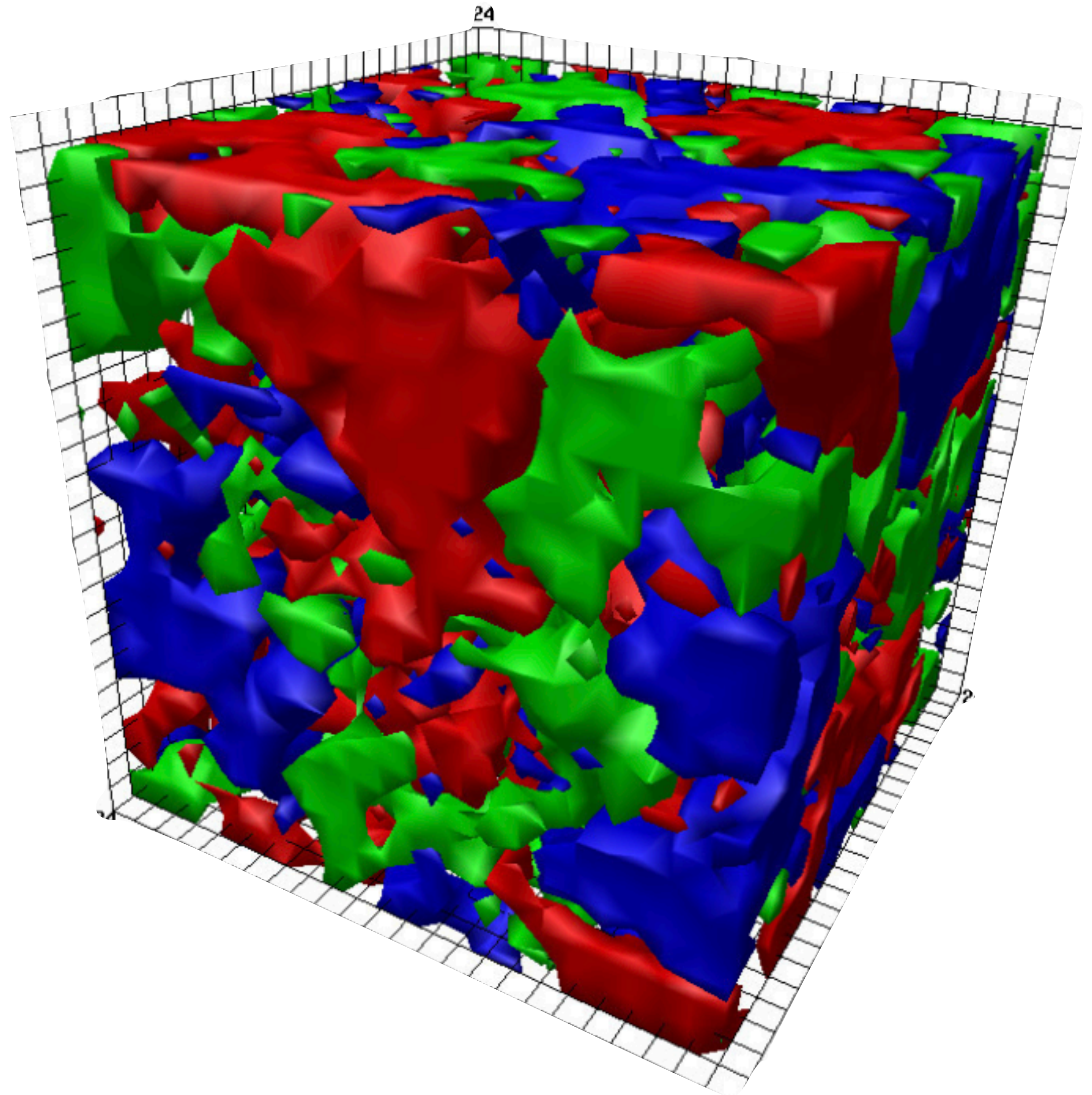
Eichhorn, et al. (2023) 2307.04742



Why generative models?

Lattice field theories may be well-suited for application of ML

- Problem involving **lots** of well-structured data (lattice cfgs ~ images)
- Analytically-known Boltzmann distribution
- Global updates may be possible



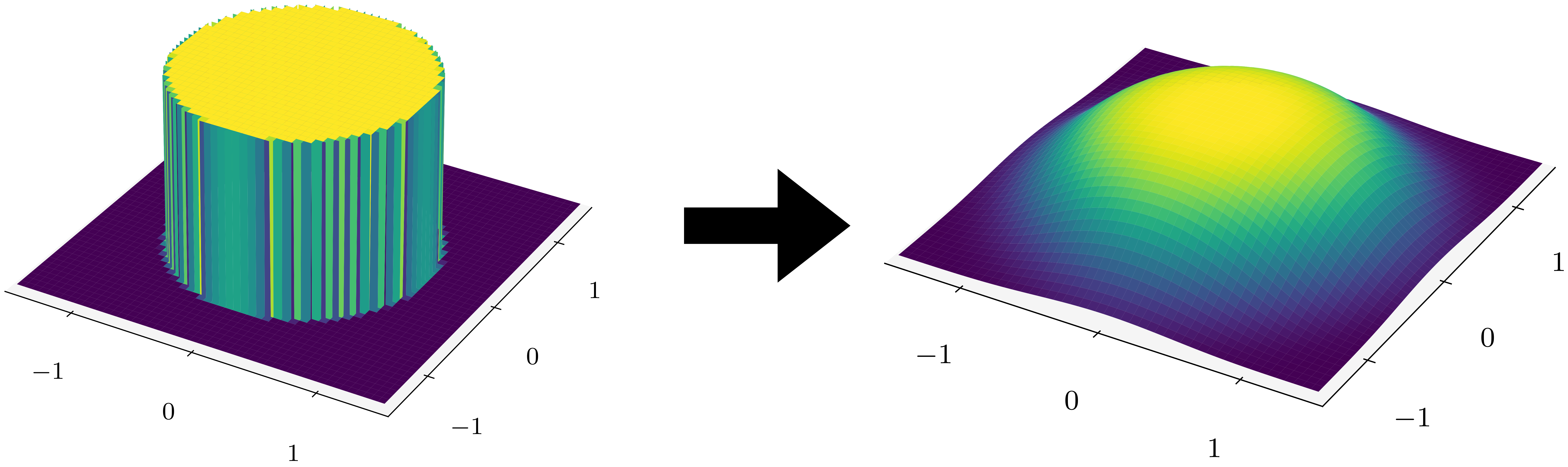
Stokes, Kamleh, Leinweber 1312.0991

Generative flow models

Sampling using flows

Box-Muller transform (Marsaglia polar form)

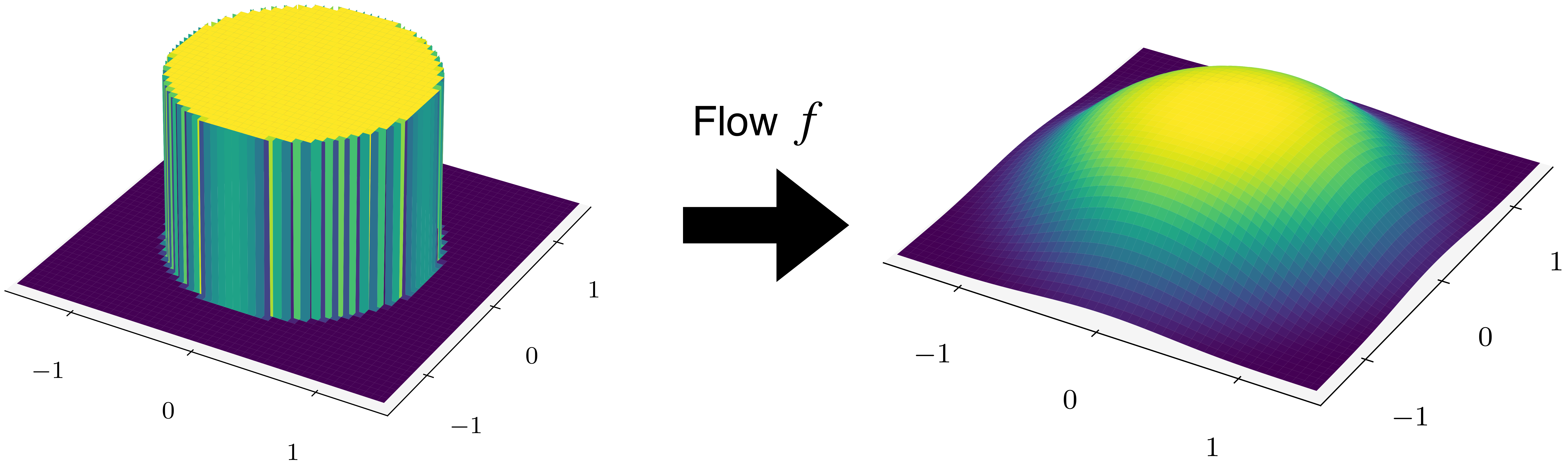
$$x' = \frac{x}{r}\sqrt{-2\ln r^2} \qquad y' = \frac{y}{r}\sqrt{-2\ln r^2}$$



Sampling using flows

Box-Muller transform (Marsaglia polar form)

$$x' = \frac{x}{r}\sqrt{-2\ln r^2} \qquad y' = \frac{y}{r}\sqrt{-2\ln r^2}$$



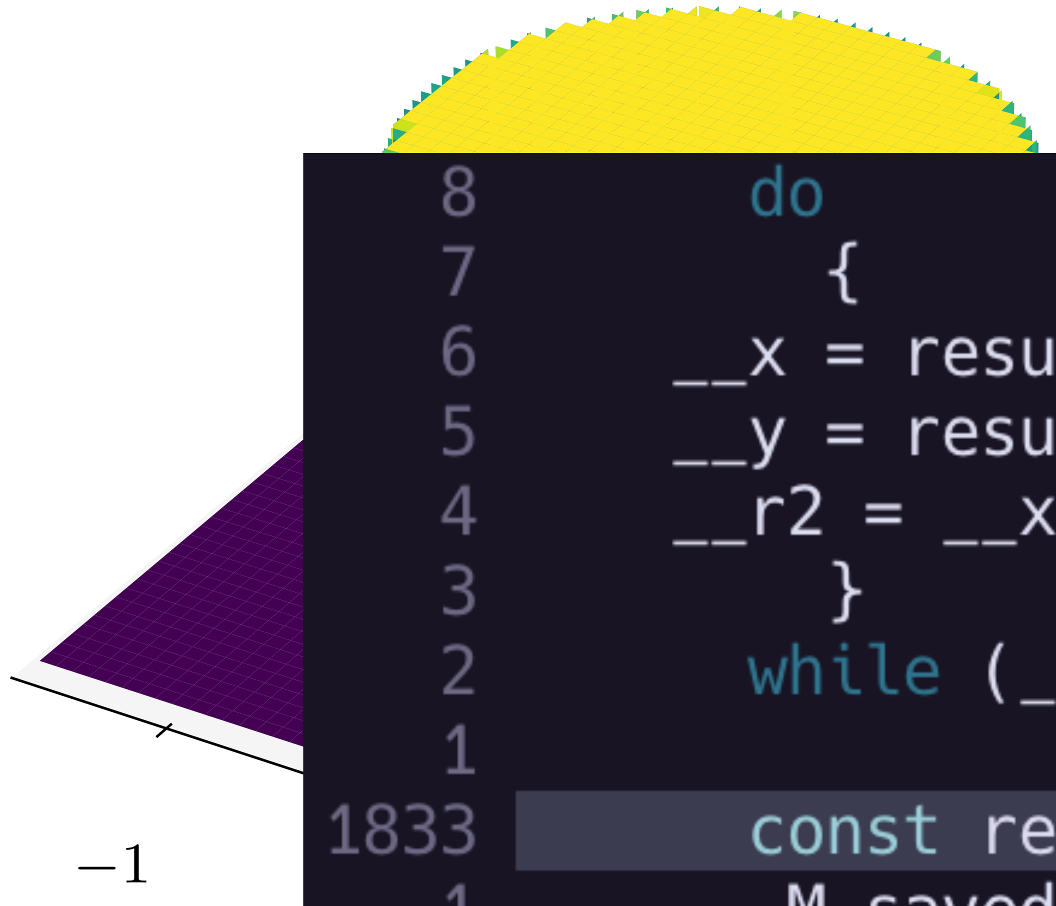
(Simple) Prior density:
$r(x, y)$

(More complex) Output density:
$q(x', y') = r(x, y)\,|\det J|^{-1}$

Sampling using flows

Box-Muller transform (Marsaglia polar form)

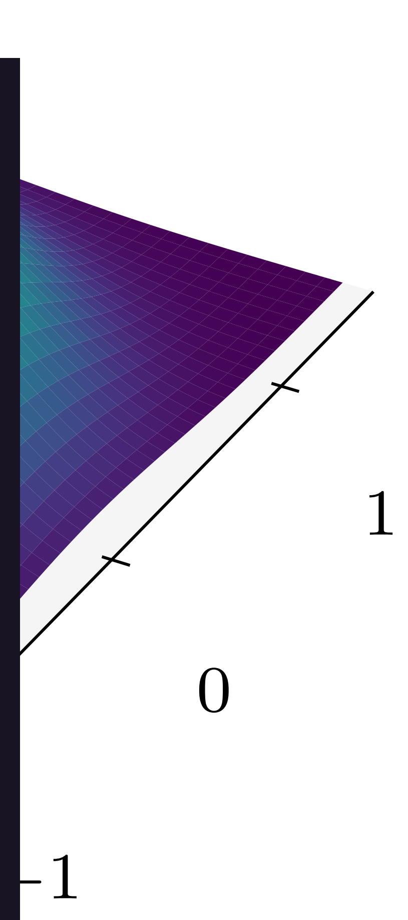
$$x' = \frac{x}{r}\sqrt{-2\ln r^2} \qquad y' = \frac{y}{r}\sqrt{-2\ln r^2}$$



libstdc++ <random>

```
    do
      {
    __x = result_type(2.0) * __aurng() - 1.0;
    __y = result_type(2.0) * __aurng() - 1.0;
    __r2 = __x * __x + __y * __y;
      }
    while (__r2 > 1.0 || __r2 == 0.0);

    const result_type __mult = std::sqrt(-2 * std::log(__r2) / __r2);
    _M_saved = __x * __mult;
    _M_saved_available = true;
    __ret = __y * __mult;
```



(Simple) Prior density:
$r(x, y)$

(More complex) Output density:
$q(x', y') = r(x, y)\,|\det J|^{-1}$

Machine learning + flows

Rezende & Mohamed (2015) PMLR 37, 1530

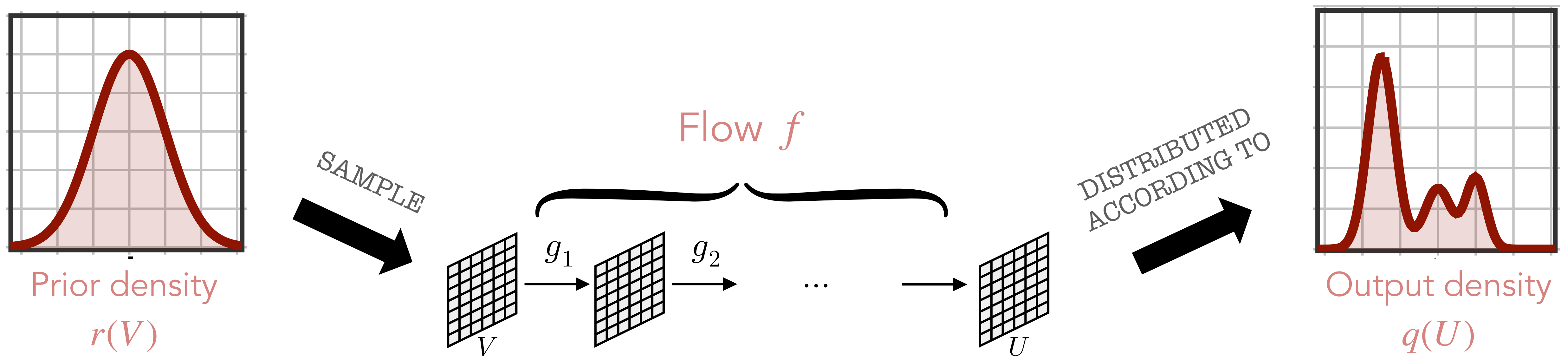
By making f learnable, we can approximate more complicated distributions.

- Must be a **diffeomorphism** with **tractable Jacobian**

- Discrete learnable flows:

Dinh+ (2014) 1410.8516 Dinh+ (2016) 1605.08803

$$f = g_1 \circ \ldots \circ g_n$$
$$\det J = \det J_1 \cdot \ldots \cdot \det J_n$$



- Continuous learnable flows:

Chen+ (2018) 1806.07366 Zhang+ (2018) 1809.10188

$$f(V) = \int_0^T dt\, \nabla \varphi(U(t);t)\Big|_{U(0)=V} + V \qquad \ln \det J = -\int_0^T dt\, \nabla^2 \varphi(U(t);t)$$

The “trivializing map” is a special continuous flow
Lüscher CMP293 (2010) 899

Note: For compact spaces, derivatives and integrals should be appropriately modified to act in the space.

Massachusetts Institute of Technology

The NSF Institute for Artificial Intelligence and Fundamental Interactions


Phiala Shanahan


Denis Boyda


Fernando Romero-López


Julian Urban


Ryan Abbott

HARVARD UNIVERSITY

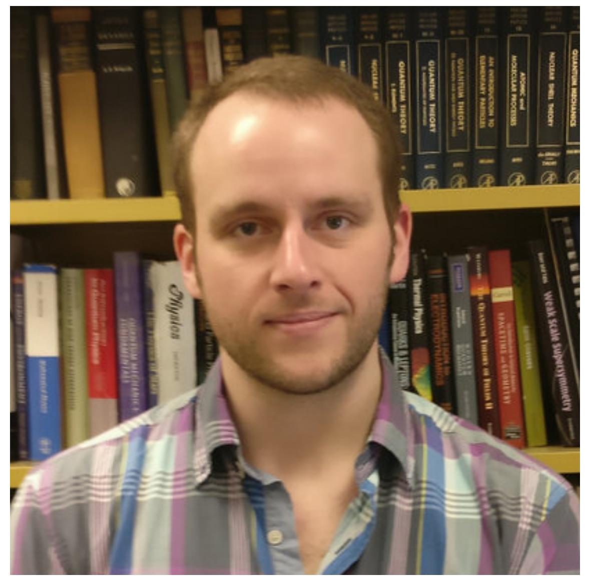
The NSF Institute for Artificial Intelligence and Fundamental Interactions


Michael Albergo

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON


Kyle Cranmer

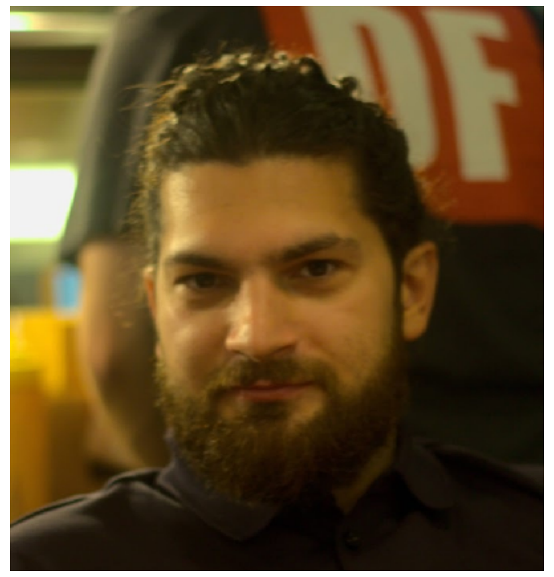
Fermilab


Dan Hackett

DeepMind


Sébastien Racanière


Danilo Rezende


Aleksander Botev


Alexander Matthews


Ali Razavi

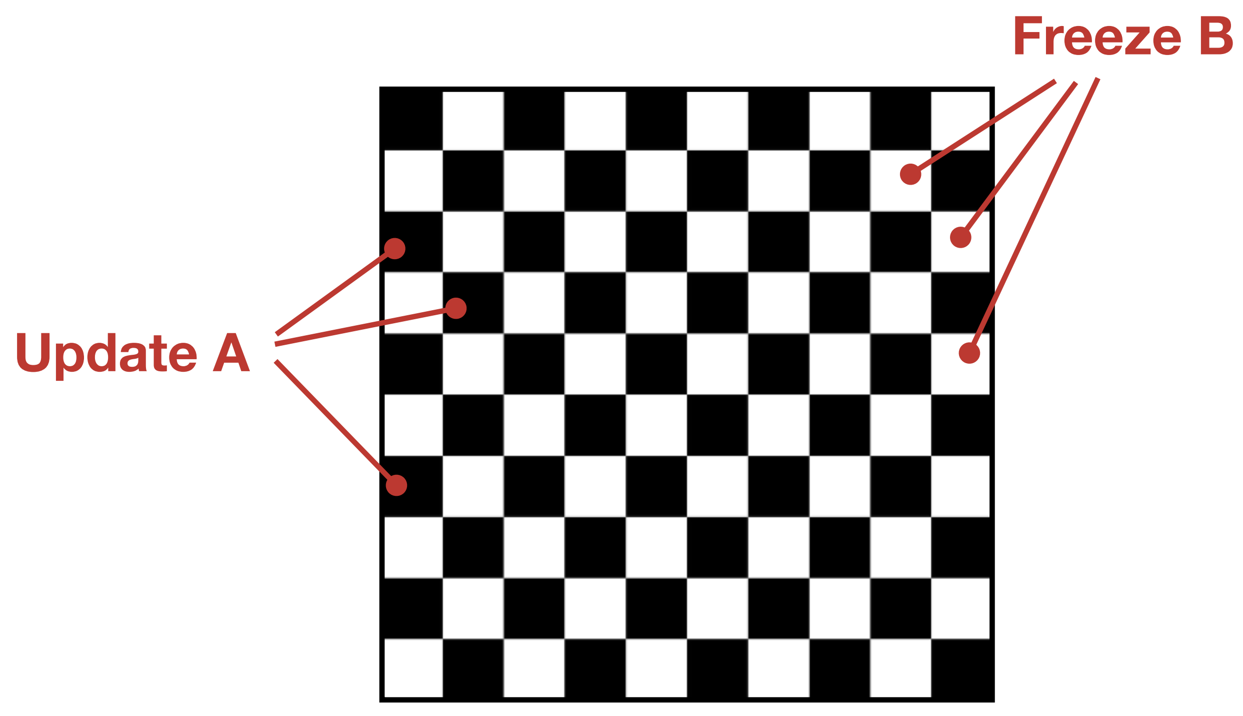
Test for scalar field theory

Scalar field $\phi(x) \in \mathbb{R}$, 1+1D spacetime

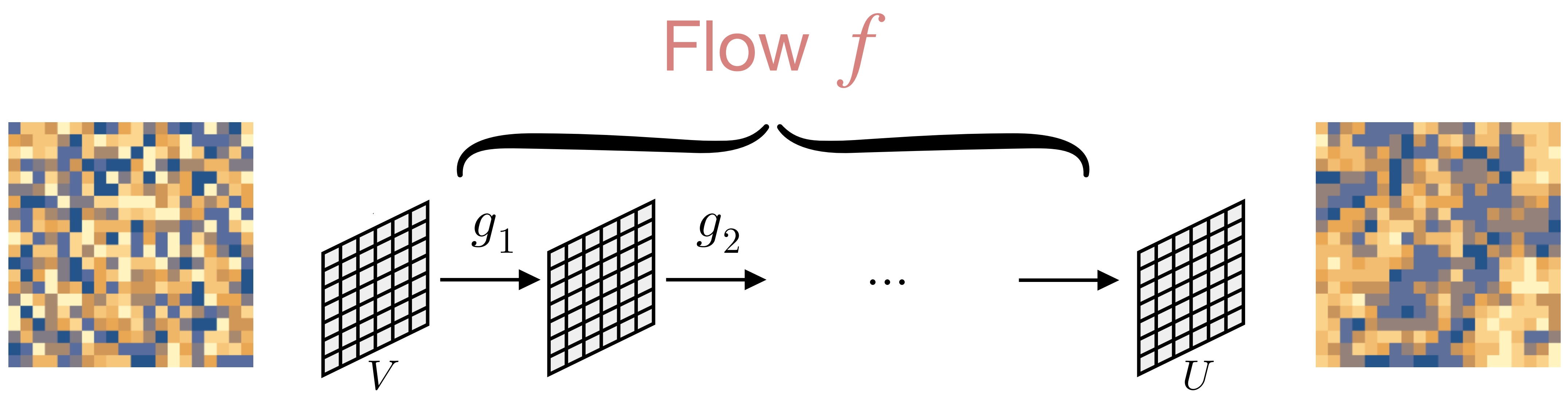
$$S[\phi] = \sum_x \partial_\mu \phi(x) \partial^\mu \phi(x) + \frac{M^2}{2} \phi(x)^2 + \lambda \phi(x)^4$$

Machine learning jargon

Neural network (NN) = highly parameterized function approximator, usually a composition of linear + elementwise non-linear transformations



Checkerboard masking pattern m



Albergo, GK, Shanahan PRD100 (2019) 034515

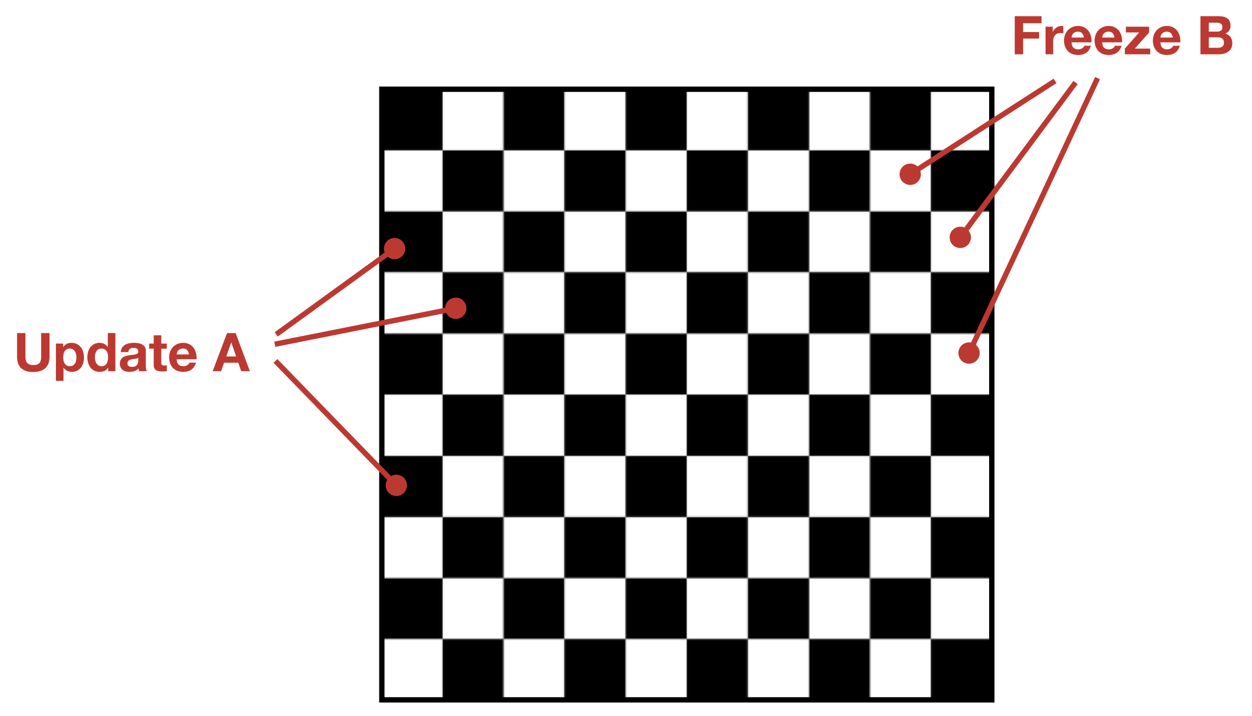
Test for scalar field theory

Scalar field $\phi(x) \in \mathbb{R}$, 1+1D spacetime

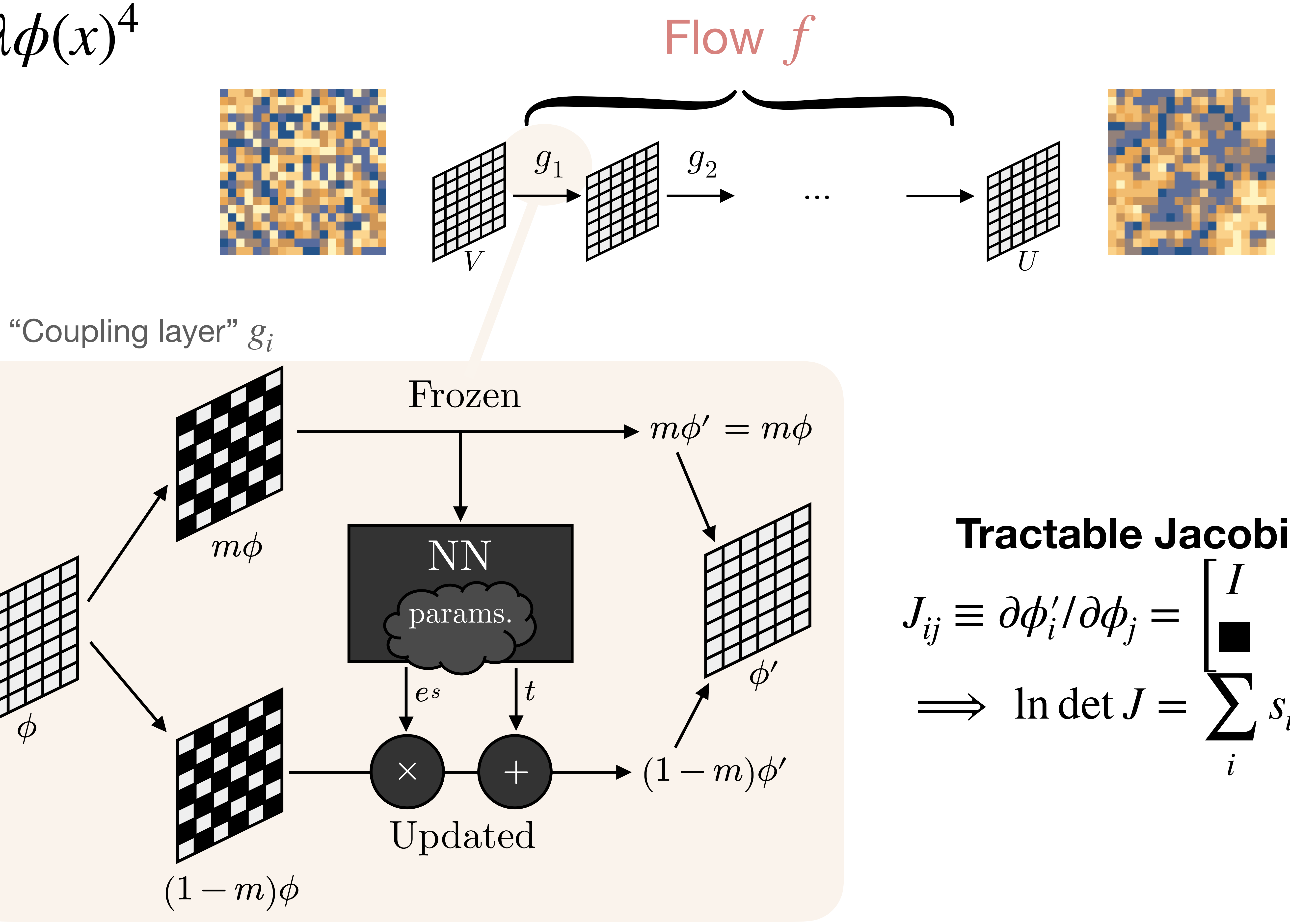
$$S[\phi] = \sum_x \partial_\mu \phi(x) \partial^\mu \phi(x) + \frac{M^2}{2}\phi(x)^2 + \lambda \phi(x)^4$$

Machine learning jargon

Neural network (NN) = highly parameterized function approximator, usually a composition of linear + elementwise non-linear transformations



Checkerboard masking pattern m



Tractable Jacobian

$$J_{ij} \equiv \partial\phi'_i / \partial\phi_j = \begin{bmatrix} I & \\ \blacksquare & \delta_{ij} e^{s_i} \end{bmatrix}$$

$$\implies \ln \det J = \sum_i s_i$$

Albergo, GK, Shanahan PRD100 (2019) 034515

Test for scalar field theory

Kullback & Leibler Ann. Math. Statist. 22 (1951) 79

Self-training using Kullback-Leibler divergence between $p(U) = e^{-S[U]}/Z$ and $q(U)$

$$\mathscr{L} \equiv D'_{\mathrm{KL}}(q\,||\,p) = \int \mathscr{D}U\, q(U)\left[\log q(U) - \log e^{-S[U]}\right]$$

$$= \int \mathscr{D}U\, q(U)\left[\log q(U) + S(U)\right] \geq -\log Z$$

Exactness by reweighting or Metropolis

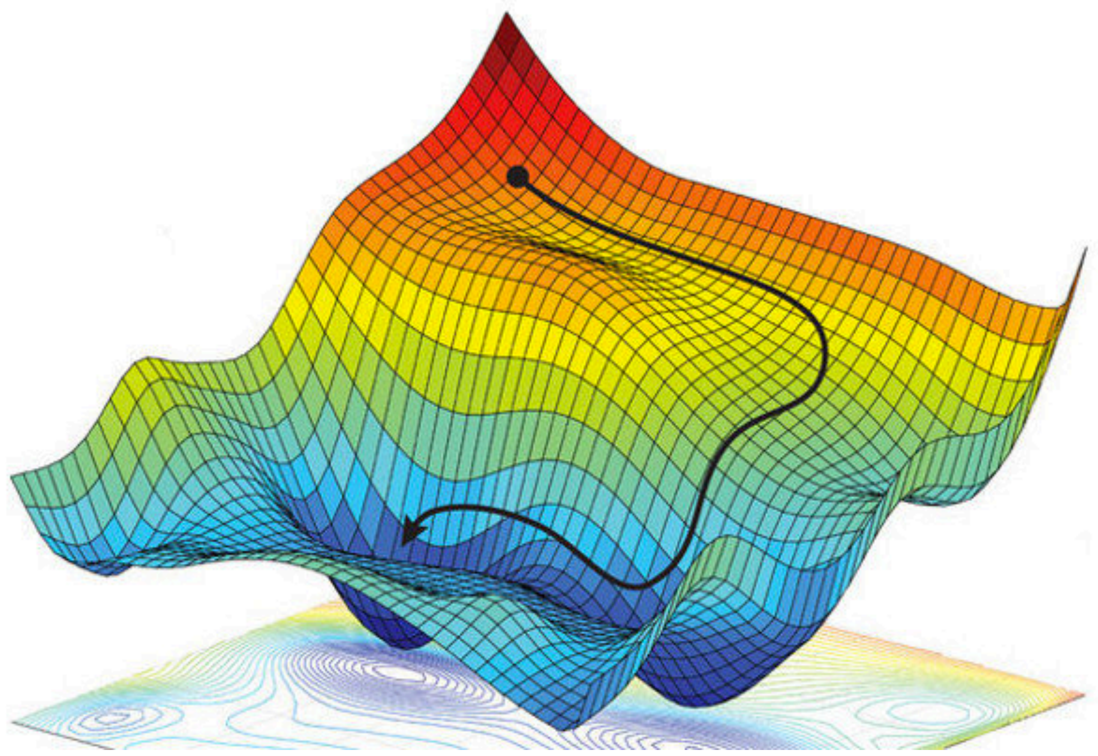
Albergo, GK, Shanahan PRD100 (2019) 034515
Nicoli+ PRE101 (2020) 023304

$$p_{\mathrm{acc}}(U \to U') = \min\left(1, \frac{p(U')}{q(U')}\frac{q(U)}{p(U)}\right)$$

Machine learning jargon

Training = optimization, typically by stochastic gradient descent
Loss function $\mathscr{L}$ = target function to be minimized

$$\vec{\omega}' = \vec{\omega} - \epsilon \nabla_{\vec{\omega}} \mathscr{L}$$



[Image credit: 1805.04829]

Test for scalar field theory

Kullback & Leibler Ann. Math. Statist. 22 (1951) 79

Self-training using Kullback-Leibler divergence between $p(U) = e^{-S[U]}/Z$ and $q(U)$

$$\mathscr{L} \equiv D'_{\mathrm{KL}}(q\,||\,p) = \int \mathscr{D}U\, q(U)\left[\log q(U) - \log e^{-S[U]}\right]$$

$$= \int \mathscr{D}U\, q(U)\left[\log q(U) + S(U)\right] \geq -\log Z$$

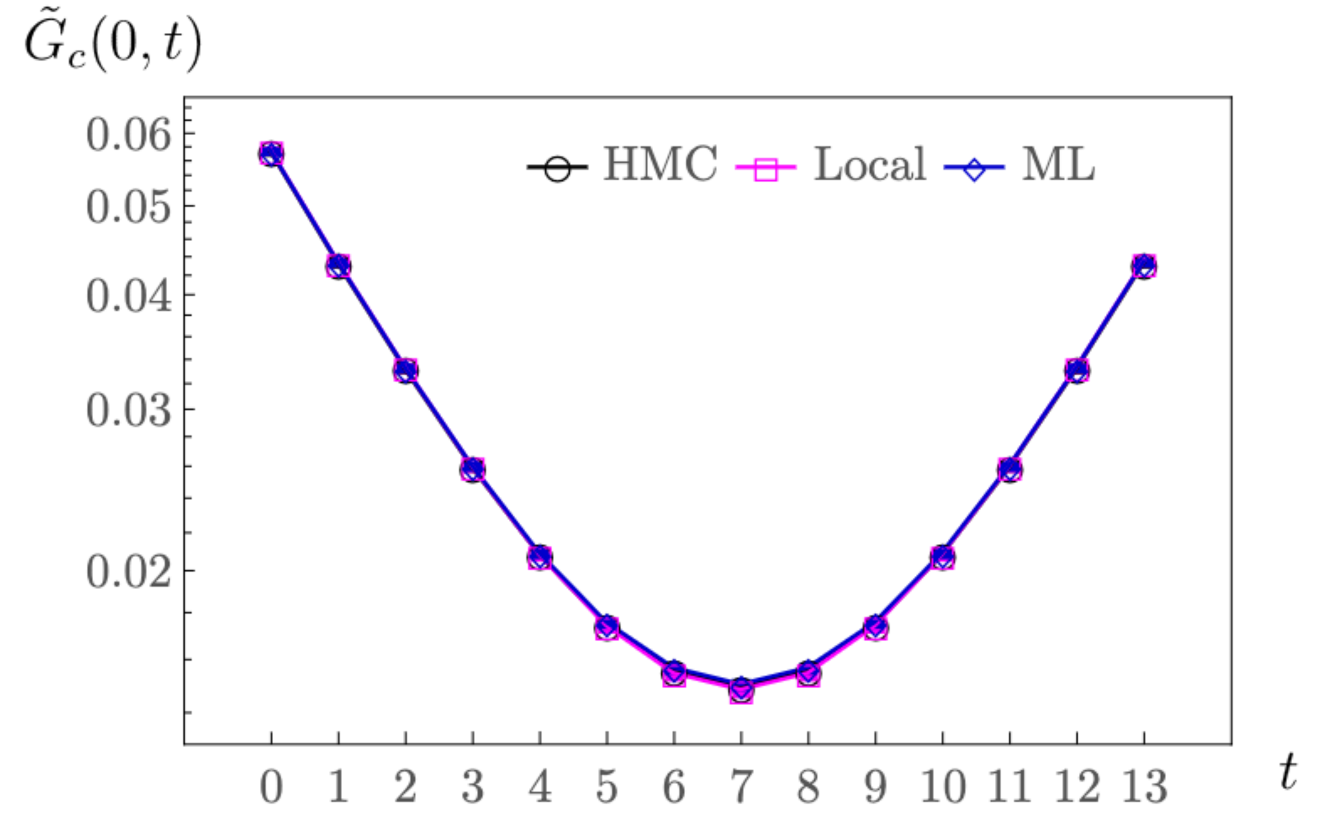
Exactness by reweighting or Metropolis

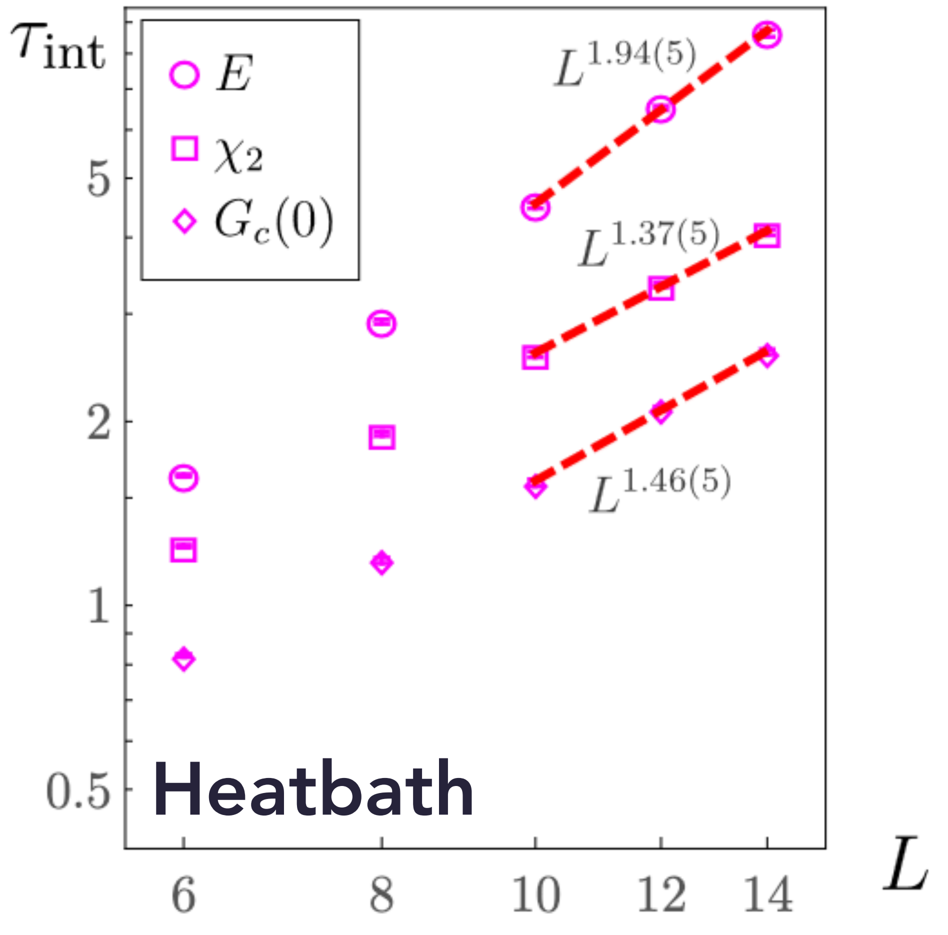
Albergo, GK, Shanahan PRD100 (2019) 034515
Nicoli+ PRE101 (2020) 023304

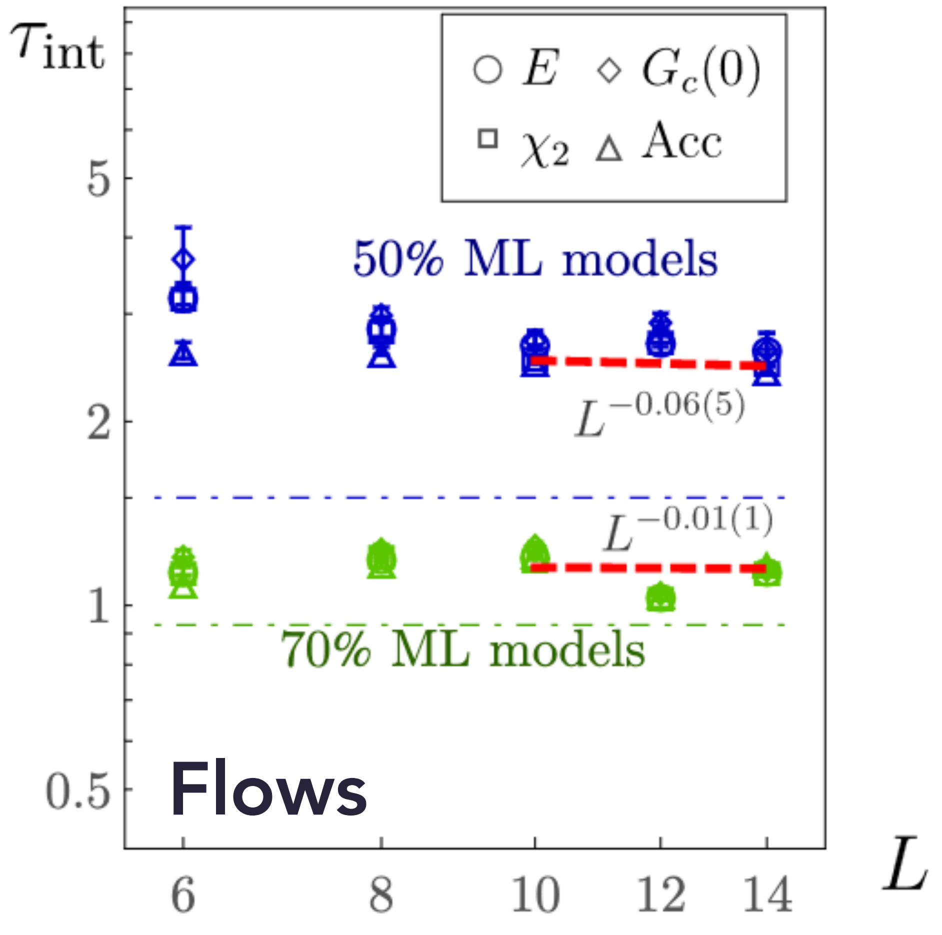
$$p_{\mathrm{acc}}(U \to U') = \min\left(1, \frac{p(U')}{q(U')}\frac{q(U)}{p(U)}\right)$$

Machine learning jargon

Training = optimization, typically by stochastic gradient descent
Loss function $\mathscr{L}$ = target function to be minimized







Test for scalar field theory

Machine learning jargon

astic gradient descent
o be minimized

Self-tra
betwee

$\mathscr{L} \equiv$

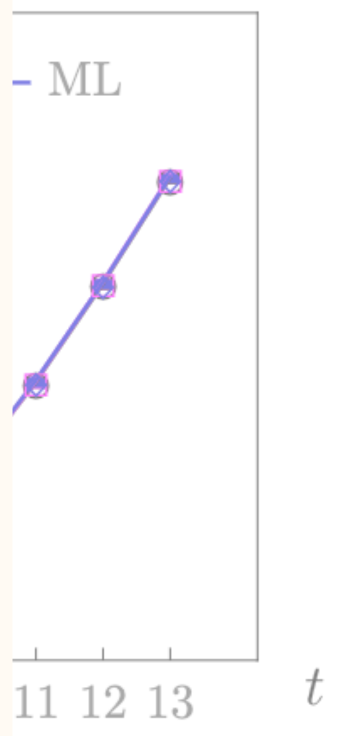
Exactne

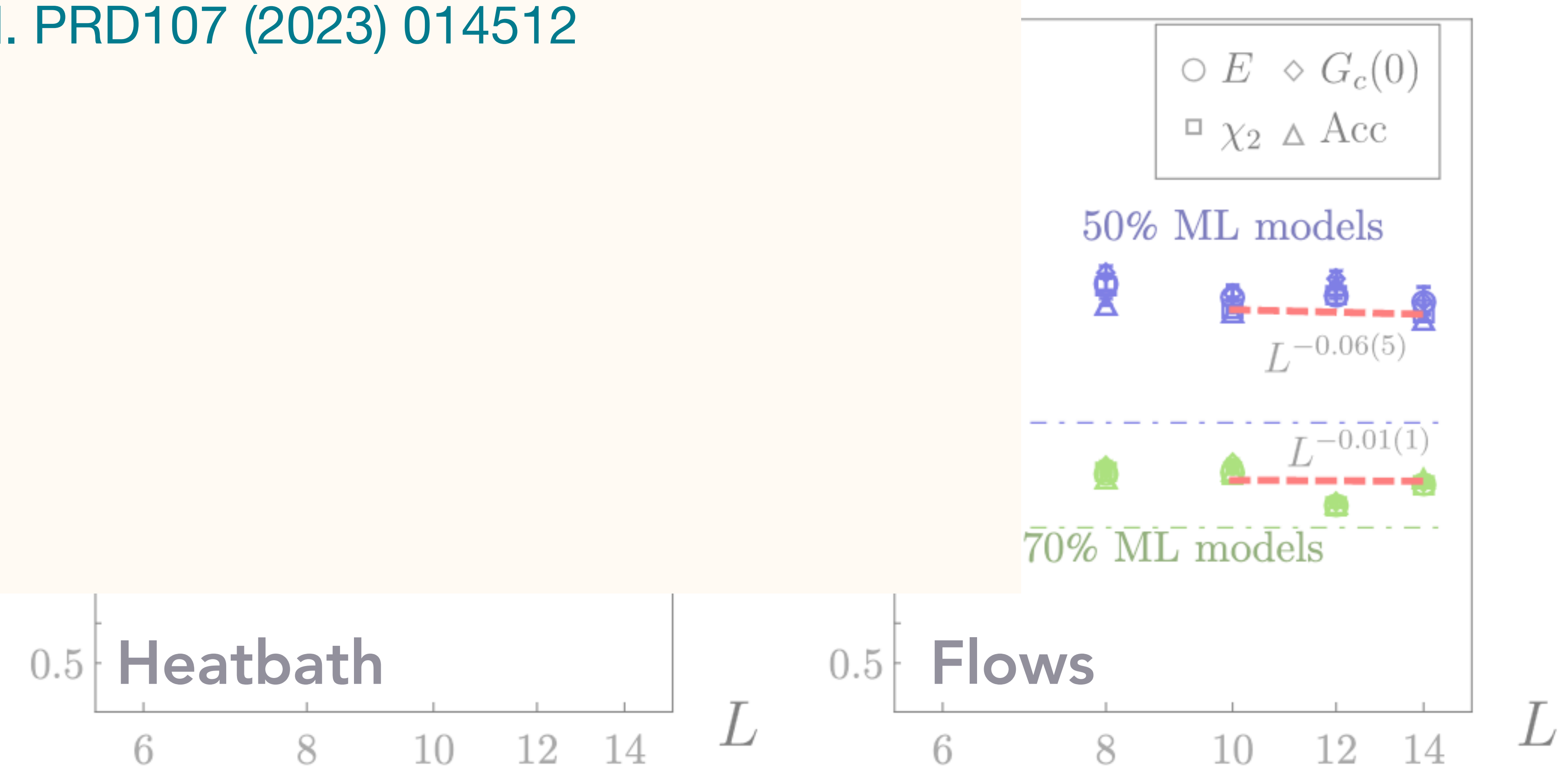
Albergo, G
Nicoli+ PR

Superseded by many recent scalar field theory results:

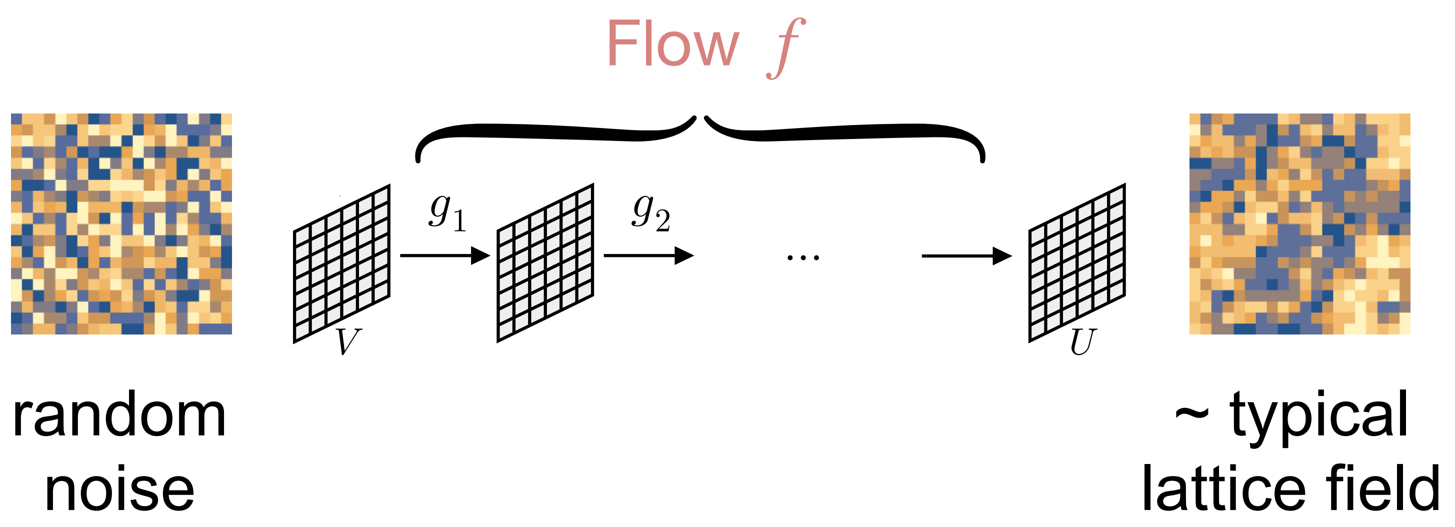
- Lattice size up to $L = 64$
- Smaller lattice spacings
- Broken phase

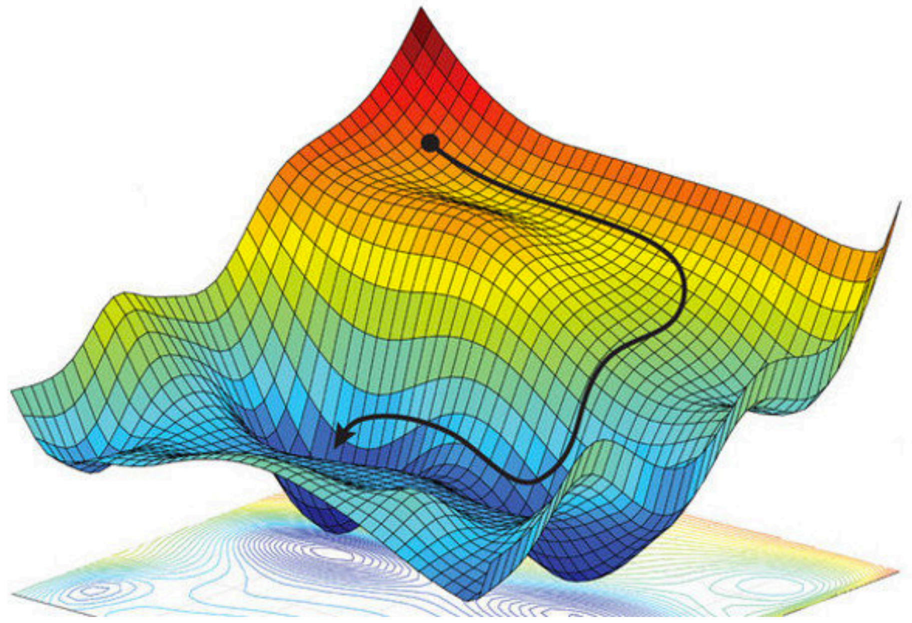
Del Debbio, et al. JHEP07 (2021) 2105.12481
de Haan, et al. NeurIPS (2021) 2110.02673
Nicoli, et al. PRL126 (2021) 032001
Caselle, et al. JHEP07 (2022) 015
Komijani, Marinkovic PoSLATTICE (2022) 019
Gerdes, et al. (2022) 2207.00283
Albandea, et al. (2023) 2302.08408
Singha, et al. PRD107 (2023) 014512
…

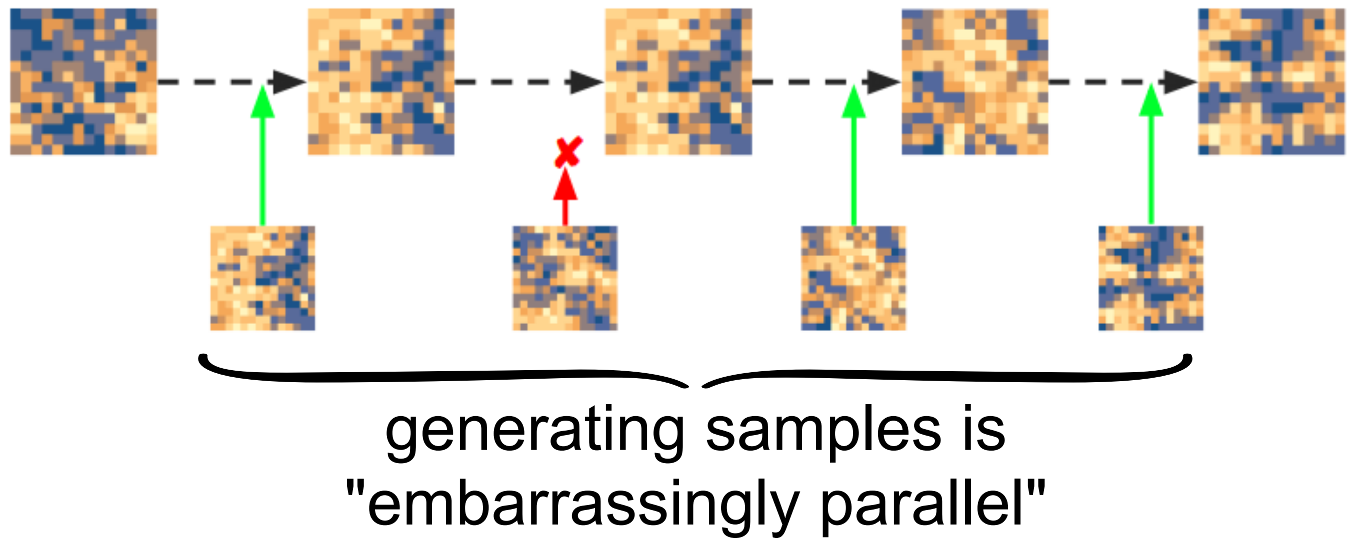


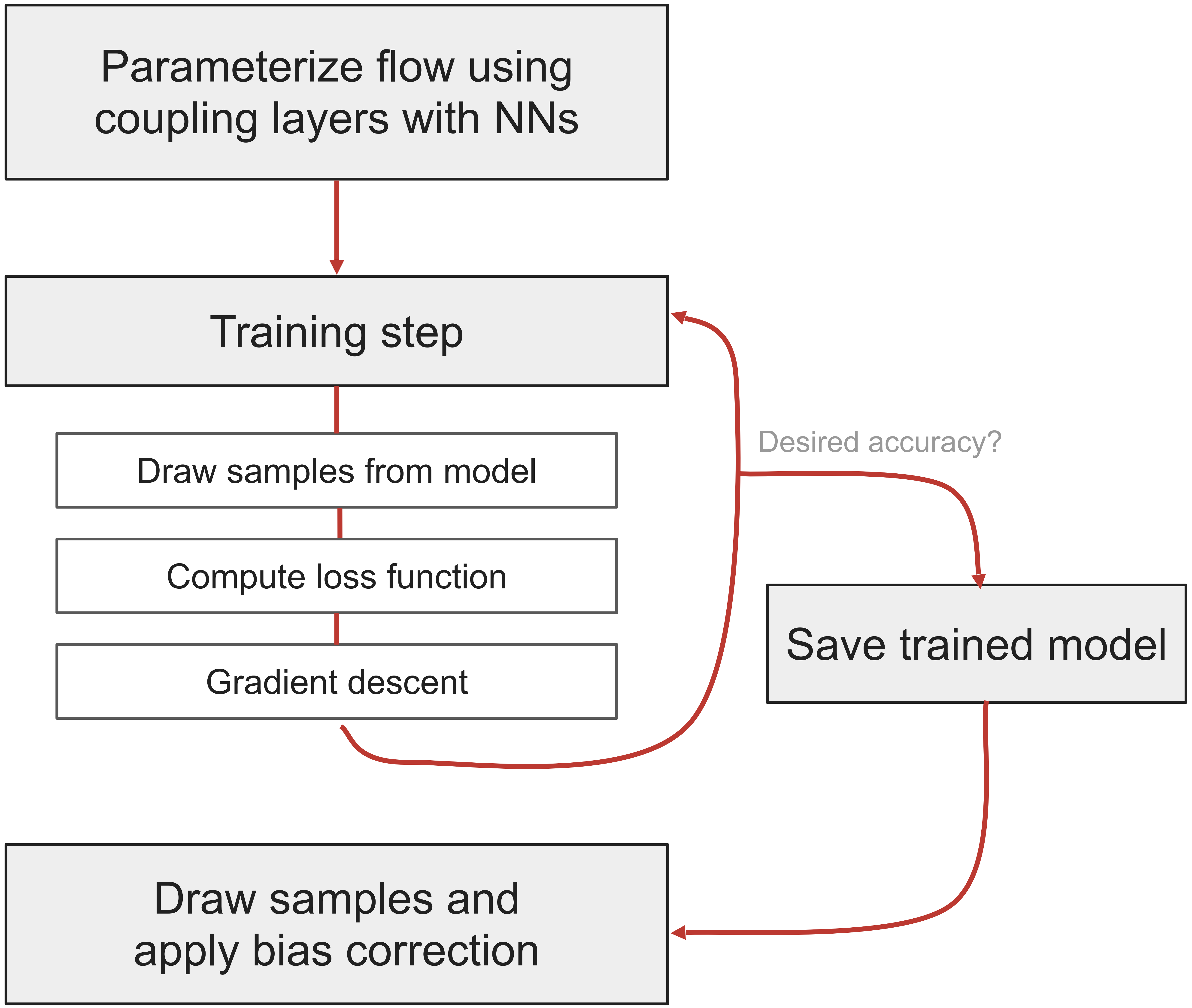


Birds-eye view









Symmetries in flows

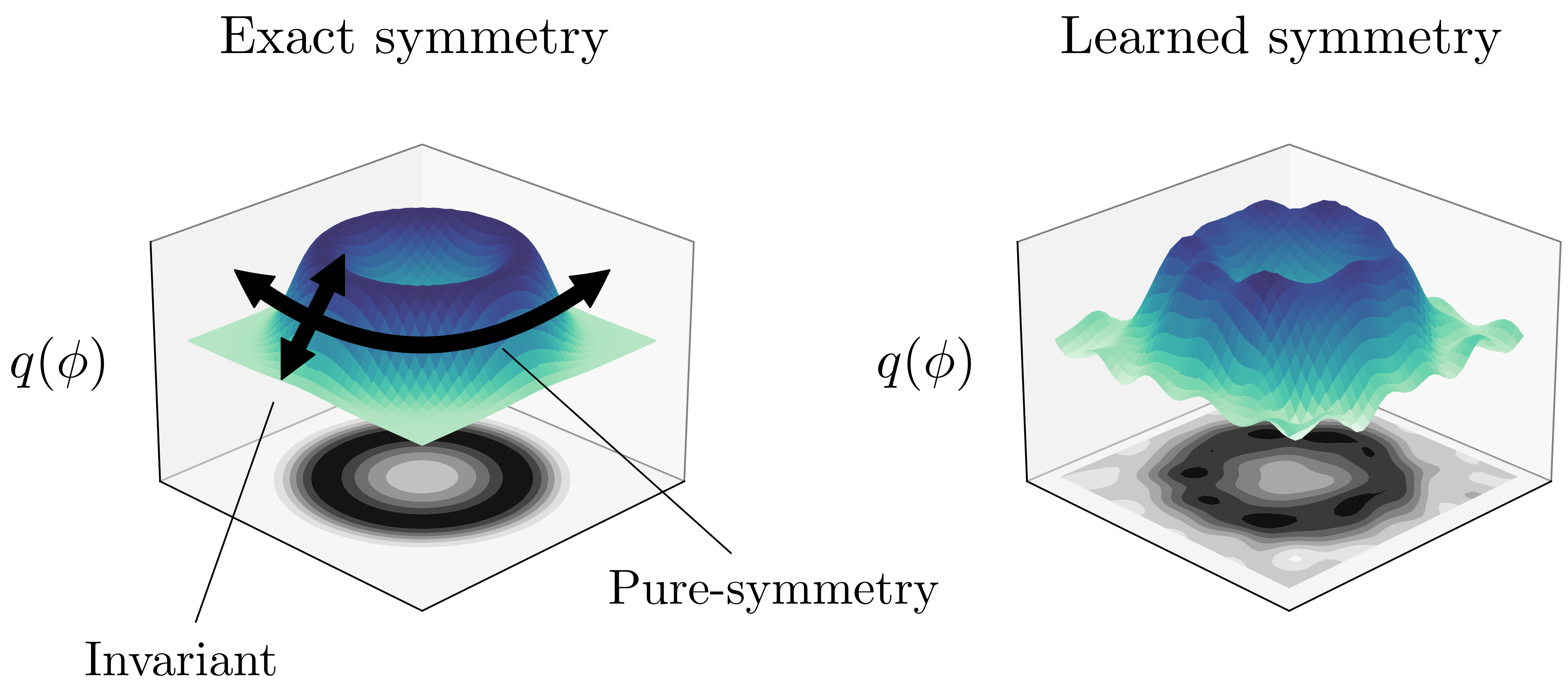
Motivation: Since target $p(\phi)$ is invariant under symmetries, natural to also make $q(\phi)$ invariant.

Invariant prior + **equivariant** flow = symmetric model   Cohen, Welling 1602.07576

$$r(t \cdot \phi) = r(\phi) \qquad f(t \cdot \phi) = t \cdot f(\phi)$$

Symmetries…

- ✓ Reduce data complexity of training
- ✓ Reduce model parameter count
- ✓ May make "loss landscape" easier



Gauge symmetry

Many lattice QFTs possess a large gauge symmetry group.

Gauge symmetry for $\mathrm{SU}(3)$
lattice gauge theory

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

GK, et al. PRL125 (2020) 121601

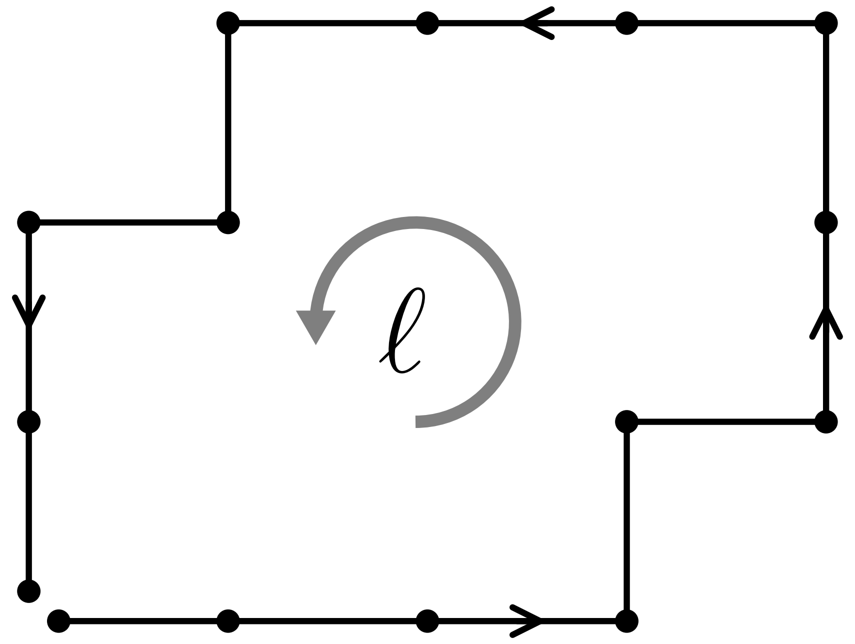
Gauge-invariant prior:

Uniform (Haar) distribution
$r(U) = 1$ works.

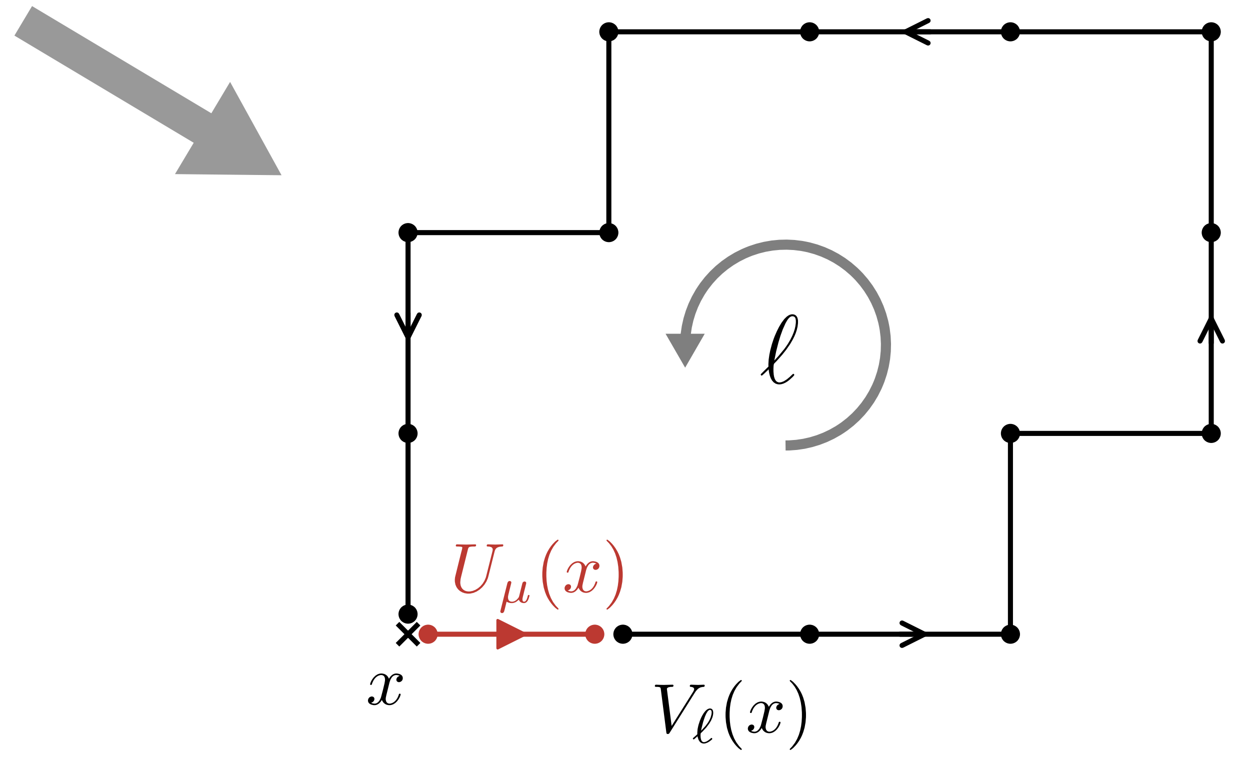
Gauge-equivariant flow:

Coupling layers acting on
(untraced) Wilson loops.

Loop transformation easier to satisfy.



$$W_\ell(x) \xrightarrow{\text{Flow}} W'_\ell(x)$$



$$U'_\mu(x) = W'_\ell(x)\, V^\dagger_\ell(x)$$

Gauge symmetry

Many lattice QFTs possess a large gauge symmetry group.

Gauge symmetry for $\mathrm{SU}(3)$ lattice gauge theory

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

Gauge-invariant prior:

Uniform (Haar) distribution $r(U) = 1$ works.

Gauge-equivariant flow:

Coupling layers acting on (untraced) Wilson loops.

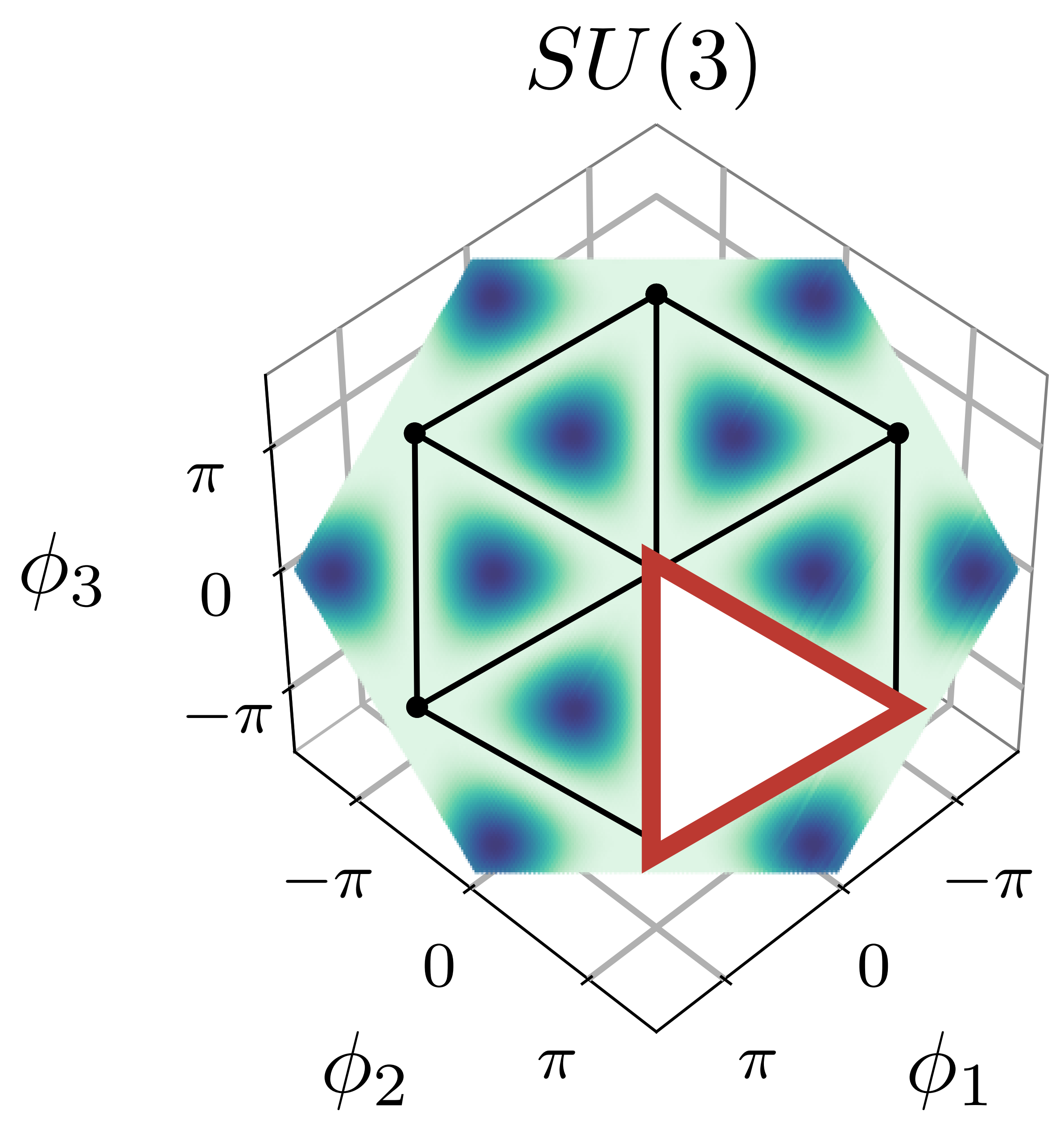
Loop transformation easier to satisfy.

Custom flows designed for $U(1)$ and $SU(N)$ gauge manifolds

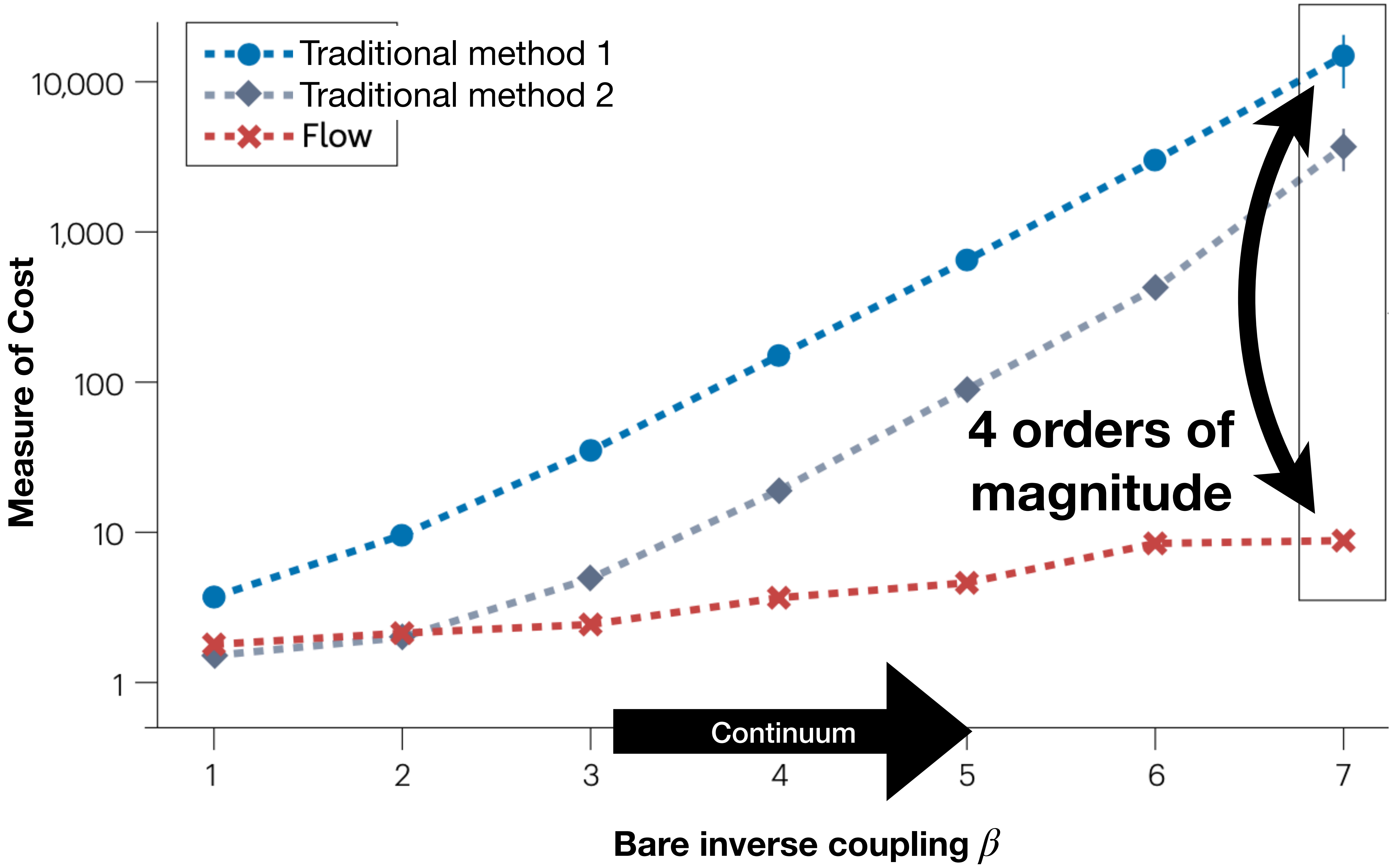
GK, et al. PRL125 (2020) 121601

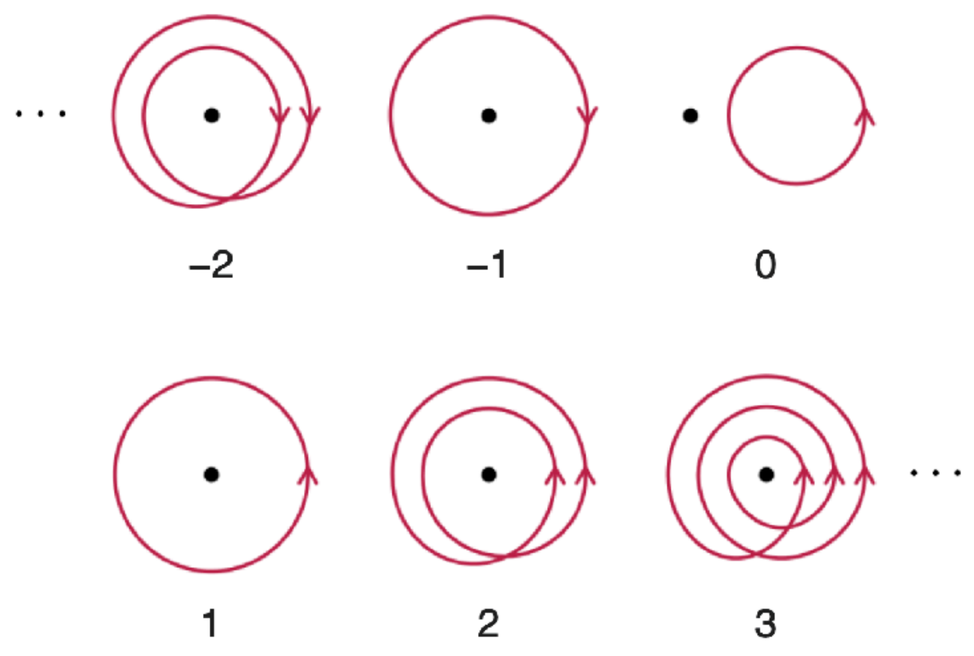
Rezende, et al. PMLR119 (2020) 8083

Boyda, et al. PRD103 (2021) 074504



Flows can solve topological freezing





Cost of MCMC vastly reduced due to better topological mixing.

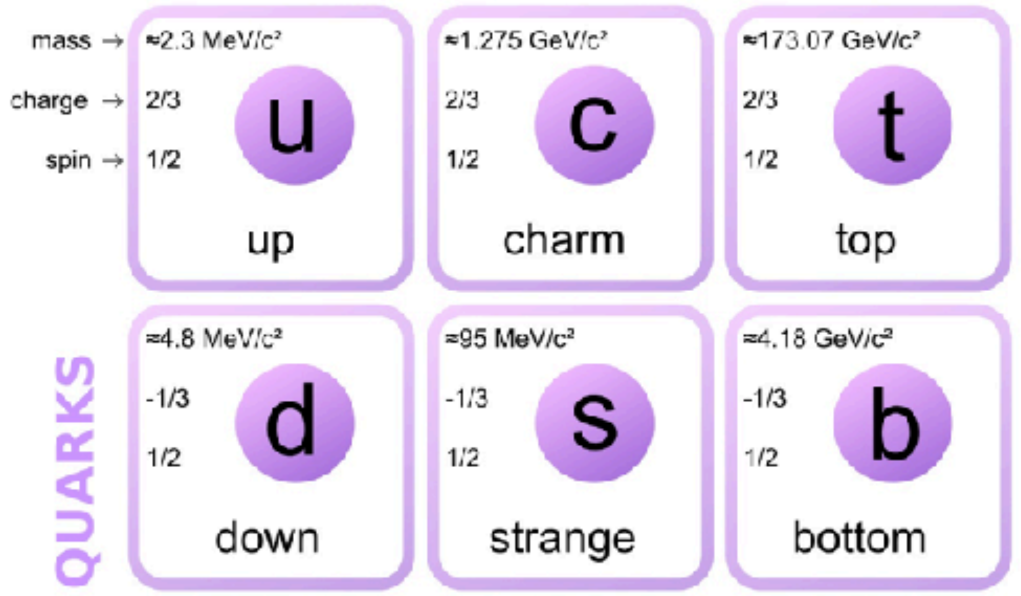
Cranmer, GK, Racanière, Rezende, Shanahan Nature Reviews Physics 5 (2023) 526

Including the quarks

Interaction between all quark flavors ($\psi_u, \psi_d, \ldots$) and gluons (U):

Action $$S_f = \sum_f \bar{\psi}_f D_f[U] \psi_f$$

Path integral $$\int \prod_f [d\bar{\psi} d\psi] e^{-S_f} = \prod_f \det(D_f[U])$$



- D_f is a sparse $O(V) \times O(V)$ matrix
- Typically use the **pseudofermion** representation for pairs of quark flavors

$$\det(D^2) \propto \int d\phi^\dagger d\phi e^{-\phi^\dagger D^{-1} \phi}$$

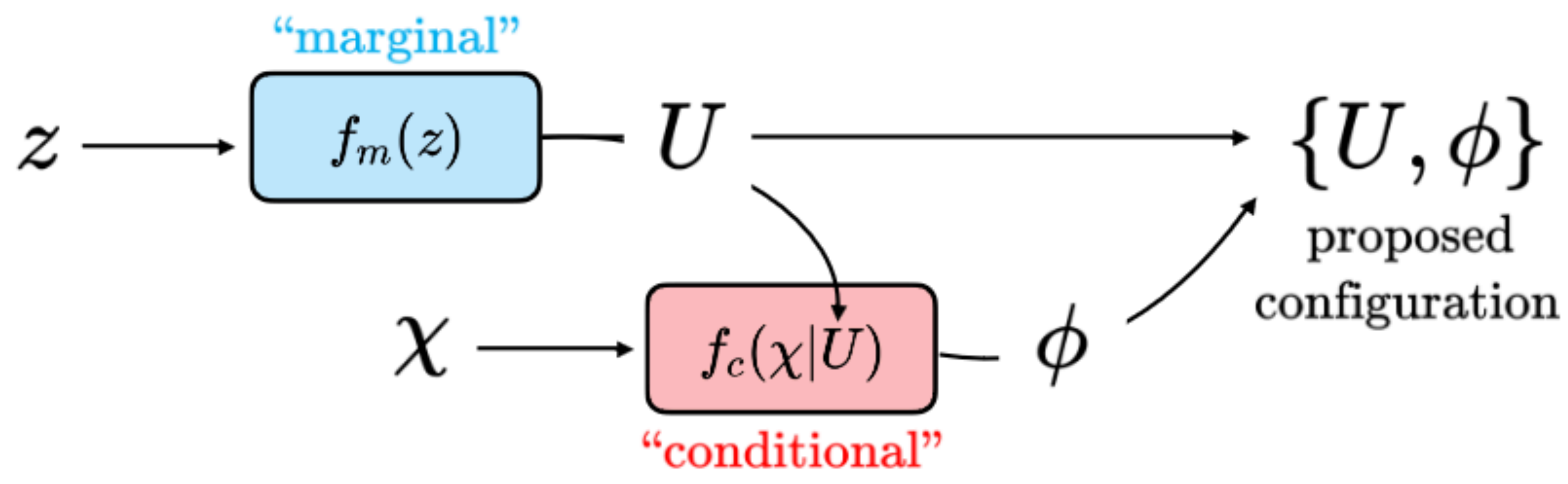
PRD104 (2021) 114507
PRD106 (2022) 014514
PRD106 (2022) 074506

Flows with pseudofermions

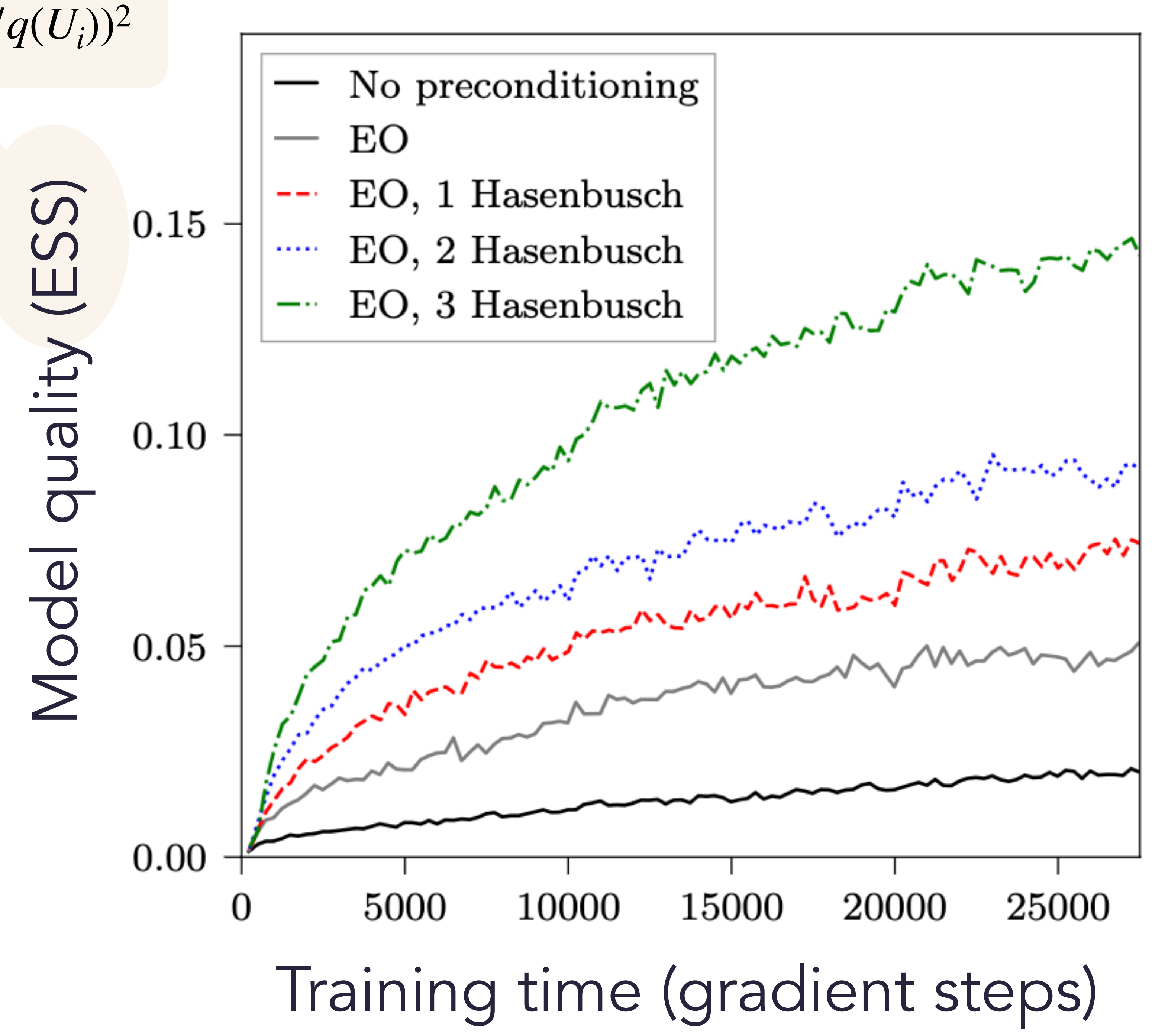
Pseudofermions highly effective in HMC, logical to use for flows also.

Separate coupling layers for gauge field and PFs can be composed arbitrarily

- **Simplest case:** marginal + conditional model



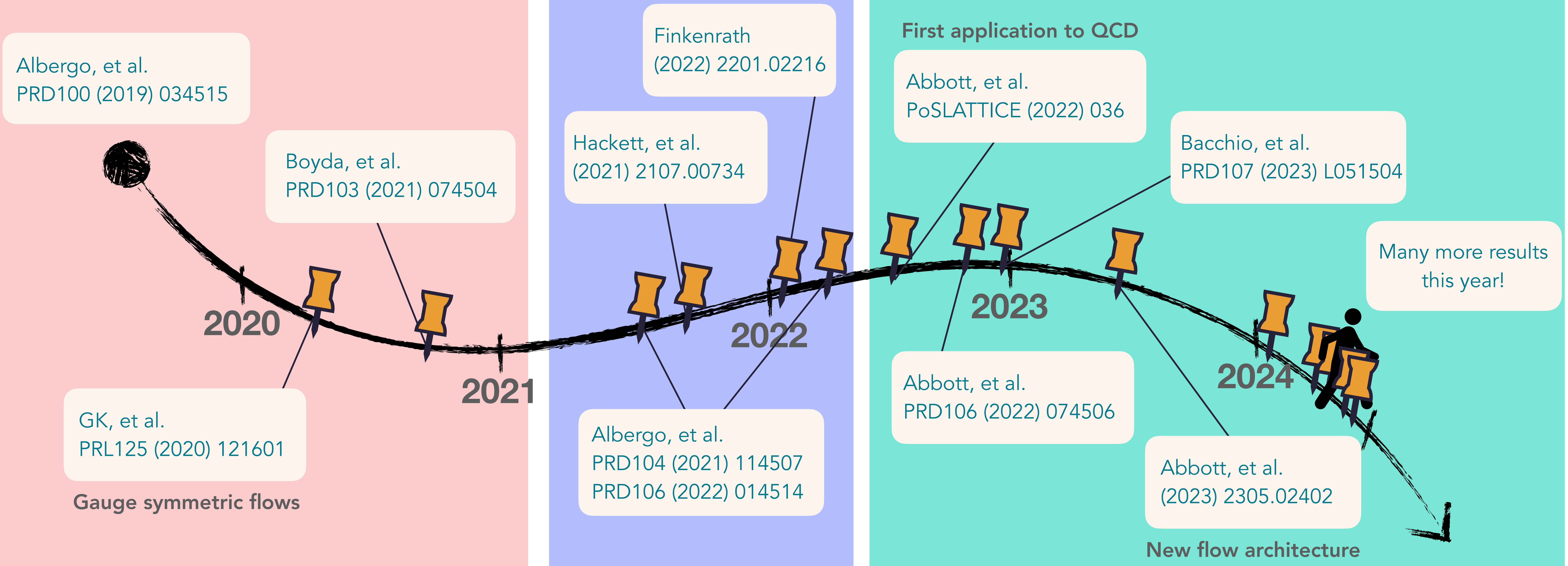
$$\mathrm{ESS} = \frac{\left(\frac{1}{N}\sum_i p(U_i)/q(U_i)\right)^2}{\frac{1}{N}\sum_i \left(p(U_i)/q(U_i)\right)^2}$$



- **Preconditioning** works equally well for flows

- Modified Metropolis allows averaging away noise in the conditional flow

Building up to QCD applications



- Viability to solve CSD
- Gauge theories
- 2D spacetime

- Fermions
- Broken symmetry

- 4D spacetime
- Refined flow models
- Better training

- Integrating flows into practical simulations

Beyond critical slowing down

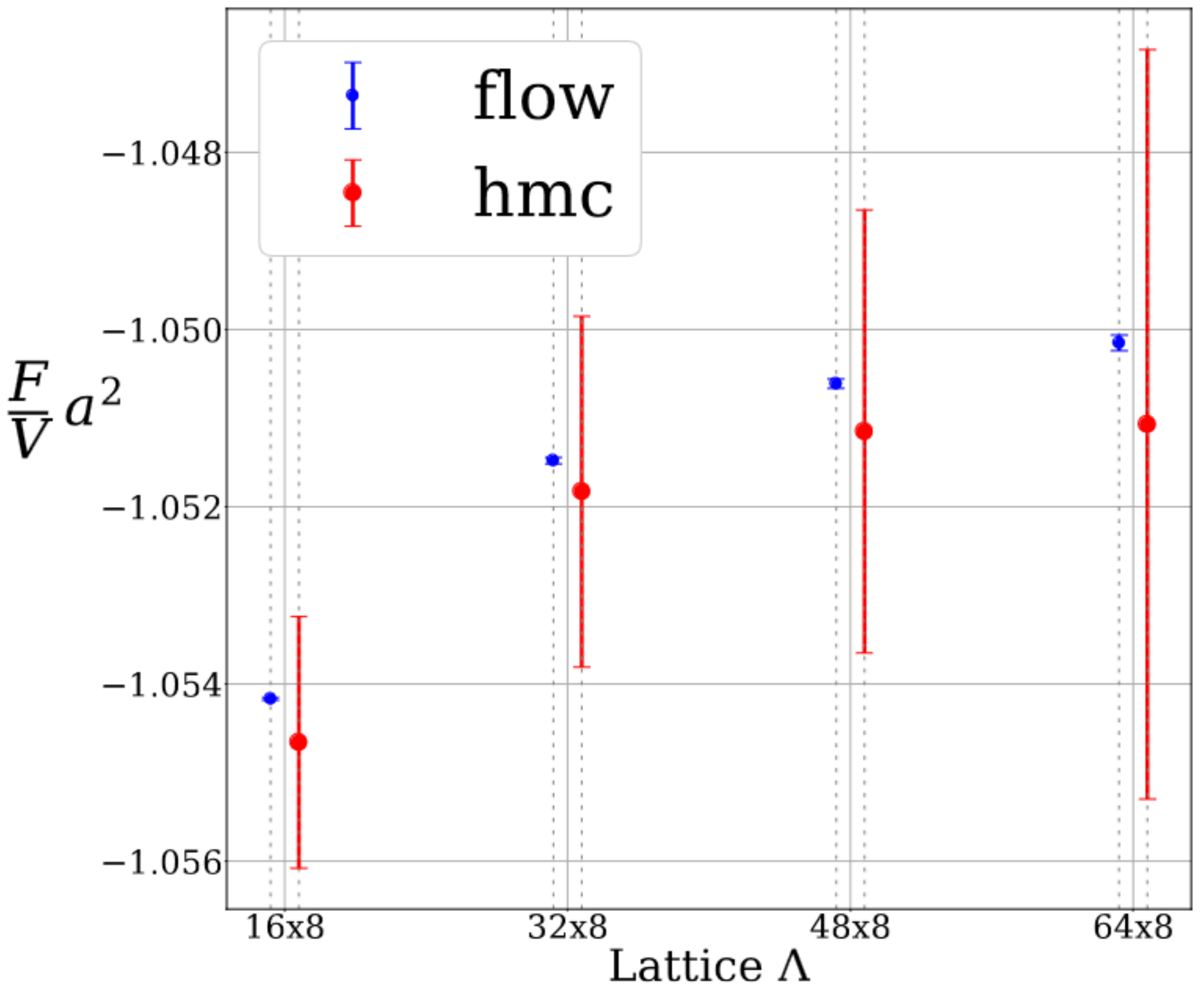
Nicoli+ PRE101 (2020) 023304
Nicoli+ PRL126 (2021) 032001

New paradigms

- Partition functions (e.g. for thermodynamics)
- Parameter dependence
  Gerdes+ (2022) 2207.00283
  Singha+ (2022) 2207.00980
- Correlated samples
- Transformed replica exchange
- Sign problems
  Lawrence+ PRD103 (2021) 114509
  Rodekamp+ PRB106 (2022) 125139
  Pawlowski & Urban (2022) 2203.01243

Practical gains

- Embarrassingly parallel sampling
- Storage-free ensembles



With $U_i \sim q(U)$,

$$\hat{Z} = \frac{1}{N}\sum_{i=1}^{N} e^{-S[U_i]}/q(U_i)$$

and $\hat{F} = -\log \hat{Z}$

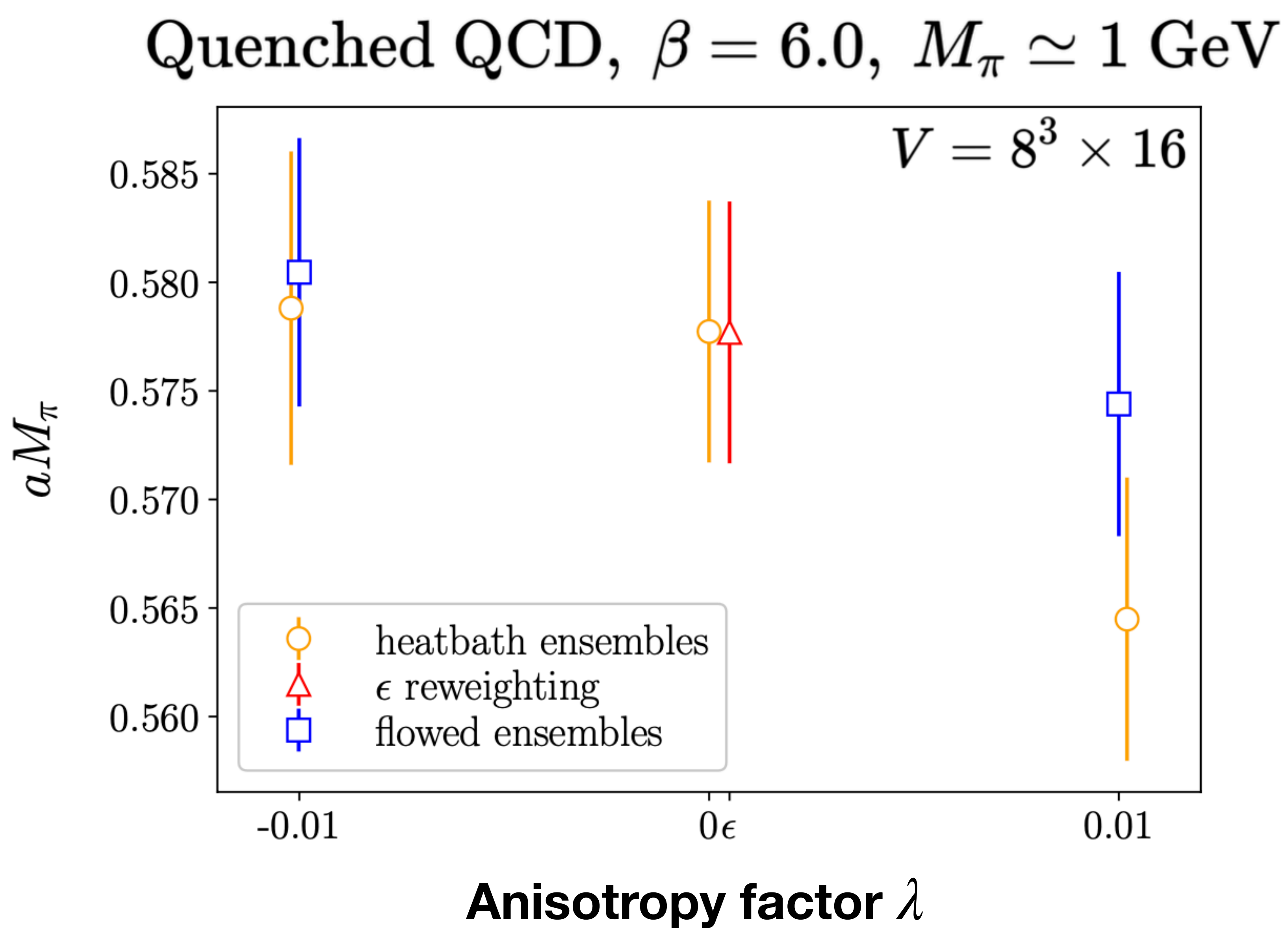
Near-term applications

Correlated sampling (e.g. Feynman-Hellmann) PRD109 (2024) 094514

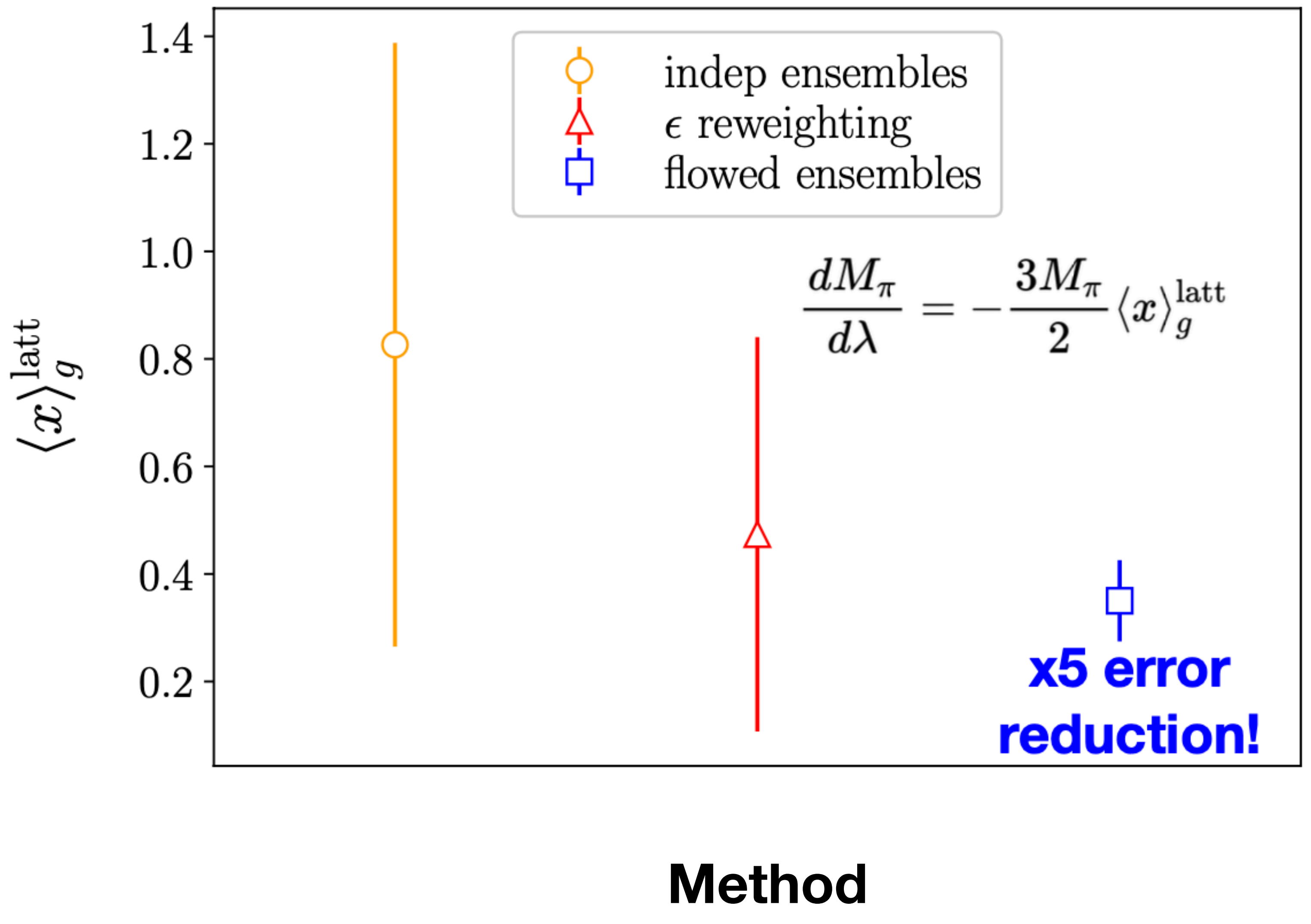
- "Shorter" distance to flow
- Correlations give noise reduction

Replica exchange with flows 2404.11674

- "Shorter" distance to flow
- Flows can be easily inserted into existing PT procedures

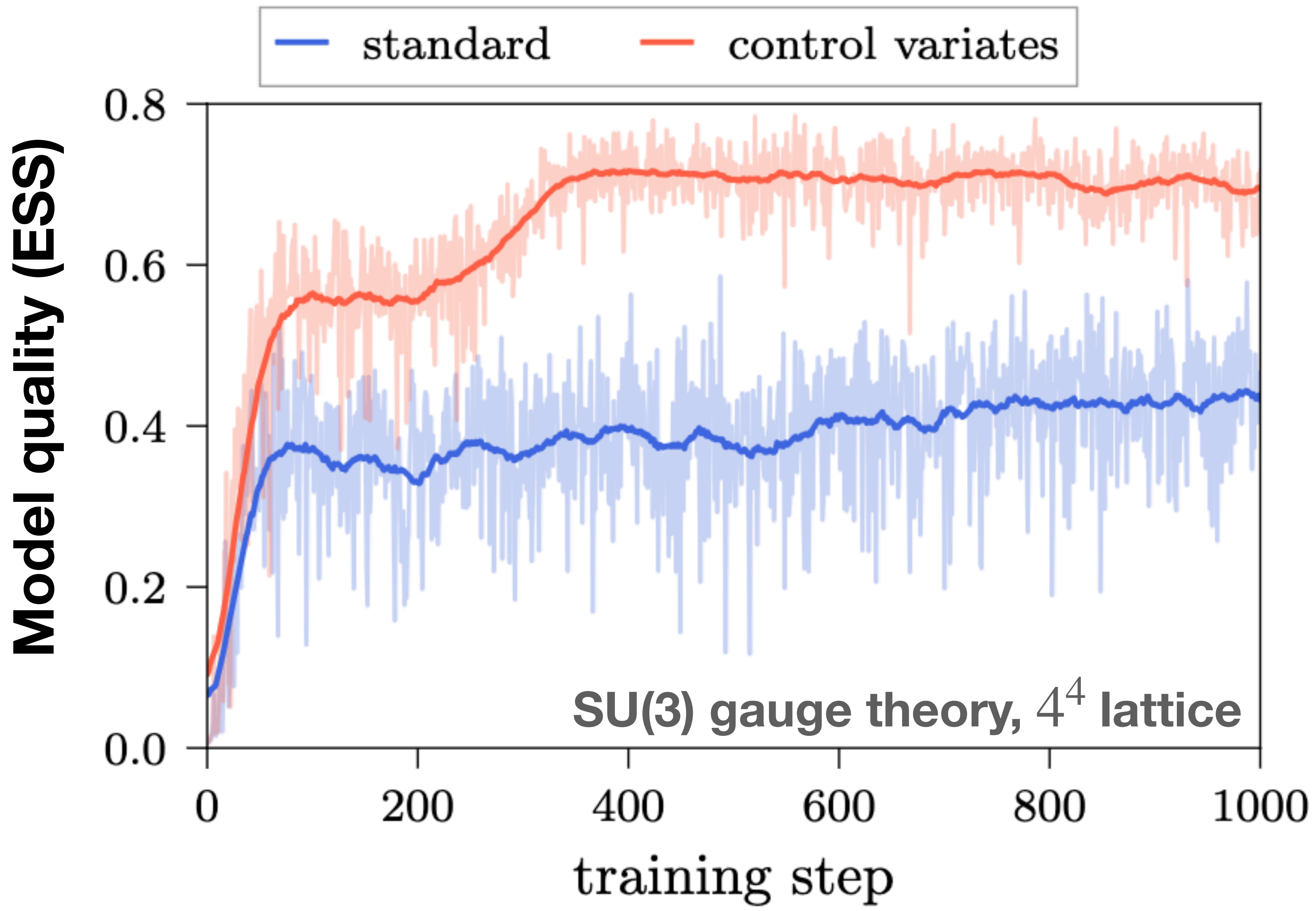


Estimate derivative



Recent developments

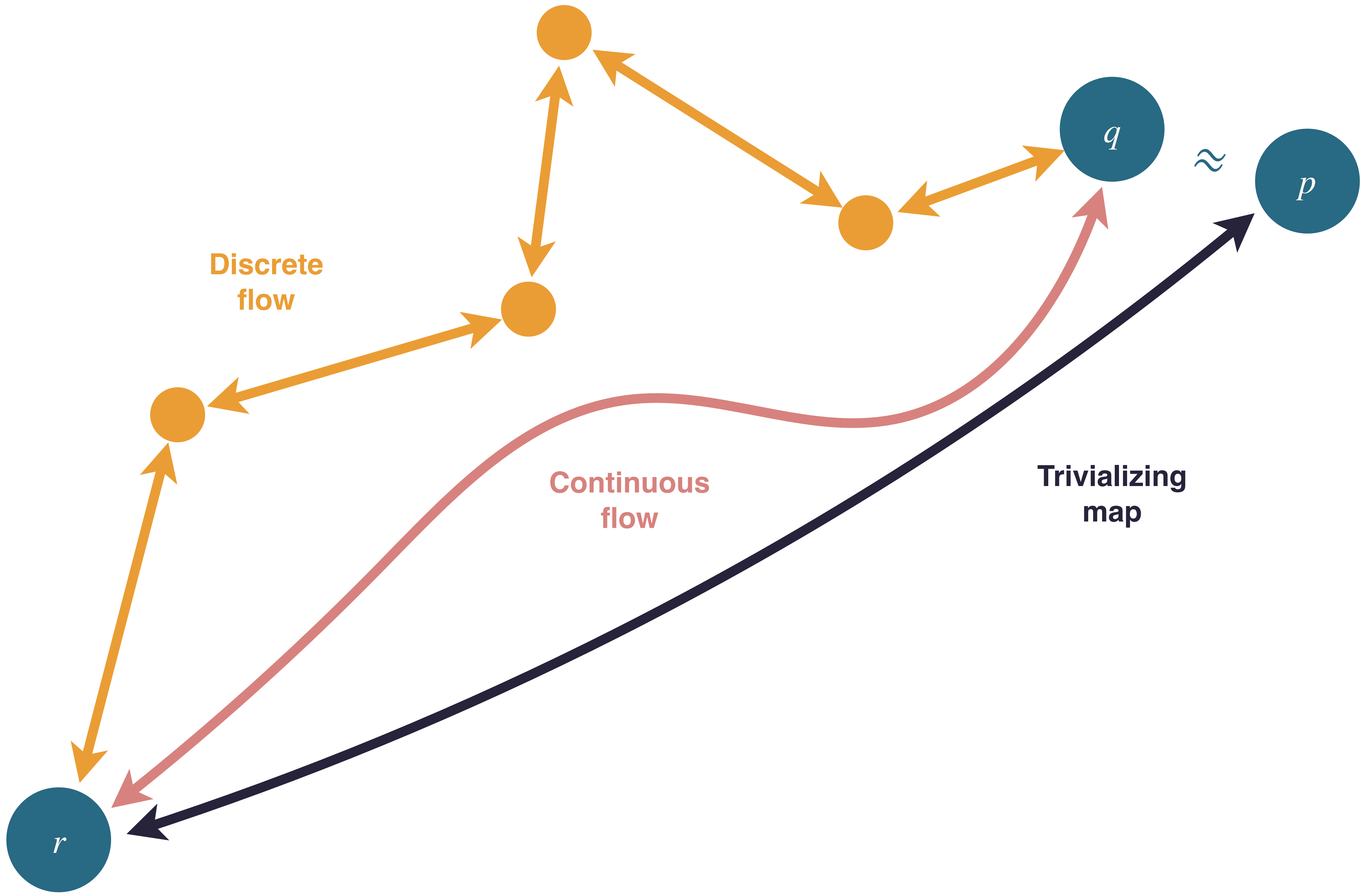
- Better training procedures
  - Minimize gradient noise with control variates or path gradients
    Vaitl, Nicoli, Nakajima, Kessel (2022) 2207.08219
    Białas, Korcyl, Stebel (2022) 2202.01314



Abbott, et al. (2023) 2305.02402

- "Residual flows"
  - Flow = Discrete steps according to gradient of scalar function $\hat{S}(\phi)$
  - Symmetries easier to encode
  - Relation to trivializing map, continuous flows
    Lüscher CMP293 (2010) 899
    Bacchio, Kessel, Schaefer, Vaitl PRD107 (2023) L051504

Continuous vs discrete flows



Open questions

- What is the most efficient path through distribution space?
- What paths are easiest to cast as continuous or discrete?
- When can integrator error be systematically eliminated?

Adding noise?

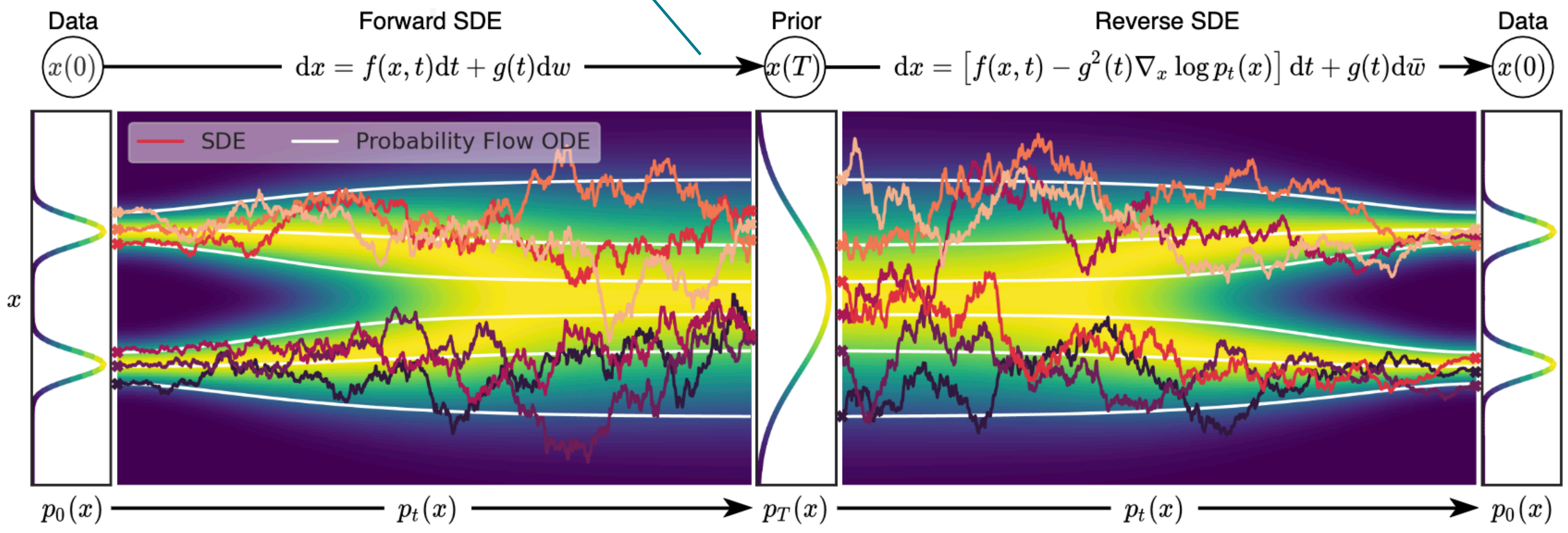
Continuous normalizing flows = ODE

Diffusion models = SDE

Out-of-equilibrium, stochastic NFs, …

See many excellent talks at this workshop

Song, Sohl-Dickstein, Kingma, Kumar, Ermon, Poole, ICLR (2021) 2011.13456
Albergo, Boffi, Vanden-Eijnden 2303.08797



Summary & Outlook

- Generative flow models have the potential to
  - Solve critical slowing down
  - Calculate partition functions
  - Explore parameter dependence
  - …
- Better **systematic control** in Lattice QCD and other lattice field theories by careful use of generative ML
- Many **practical developments** over the last 2 years
  - 3+1D pure-gauge theory
  - Quarks → demos on full QCD
- But many **open questions** still to be answered!

Summary & Outlook

- Generative flow models have the potential to
  - Solve critical slowing down
  - Calculate partition functions
  - Explore parameter dependence
  - …
- Better **systematic control** in Lattice QCD and other lattice field theories by careful use of generative ML
- Many **practical developments** over the last 2 years
  - 3+1D pure-gauge theory
  - Quarks → demos on full QCD
- But many **open questions** still to be answered!

Backup slides

Self-training scheme

Optimization must be designed for inverted data hierarchy in the lattice problem.

1. Define **"Reverse" Kullback-Leibler (KL)** divergence between $q(\phi)$ and $p(\phi) = e^{-S(\phi)}/Z$

$$D_{\mathrm{KL}}(q\,||\,p) := \int \mathcal{D}\phi\, q(\phi)\big[\log q(\phi) - \log p(\phi)\big] \geq 0$$

2. Measure using samples ϕ_i **from the model**

$$D_{\mathrm{KL}}(q\,||\,p) \approx \frac{1}{M}\sum_{i=1}^{M}\big[\log q(\phi_i) + S(\phi_i)\big]$$

3. Minimize by stochastic gradient descent

Inspired by:
- Self-Learning Monte Carlo (SLMC) [Huang, Wang PRB95 (2017) 035105; Liu, et al. PRB95 (2017) 041101; …]

- Self-play reinforcement learning [Silver, et al. Science 362 (2018), 1140]

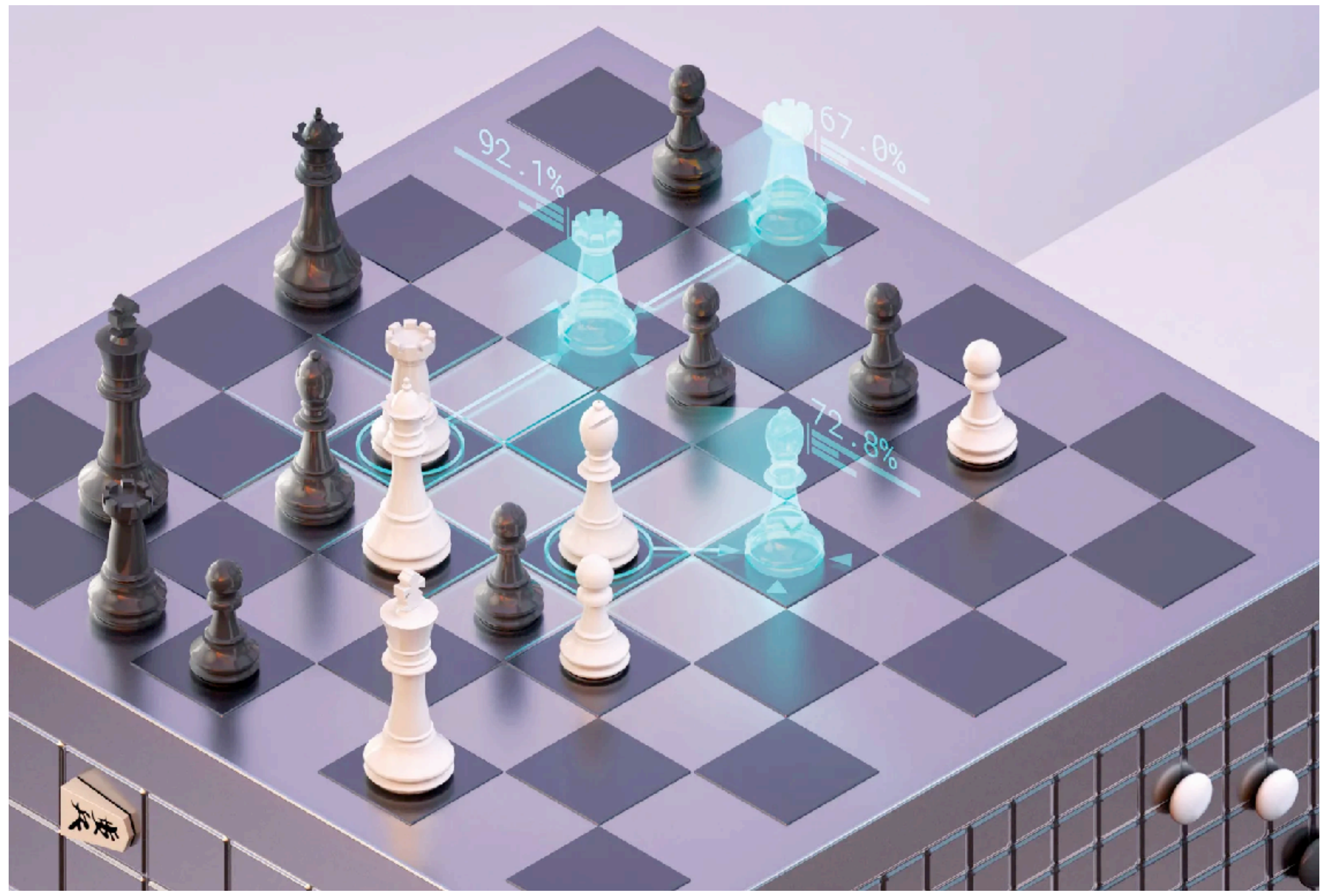


Image credit: DeepMind

Related approaches

Generative Adversarial Networks (GANs):

- Highly expressive
- Work in the direction of GANs for lattice

  Urban, Pawlowski 1811.03533

  Zhou, Endrődi, Pang, Stöcker 1810.12879

Variational AutoEncoders (VAEs):

- Can also learn meaningful directions in the prior variables

However: No access to $q(\phi)$… hard to make exact!

Karras, Lane, Aila / NVIDIA 1812.04948

AI-generated faces (GAN)

Shen & Liu 1612.05363

AI-generated faces (VAE)

What about volume scaling?

Abbott, et al. 2211.07541

Fixed models will always* scale exponentially poorly with the **physical volume**.

* in a direct sampling scheme

- Expect variance of log reweighting factors to scale as $(L/\xi)^d$
  Scaling relation $\mathrm{ESS}(V) = \mathrm{ESS}(V_0)^{V/V_0}$, where $V_0 \sim \xi^d$
- This says nothing about scaling towards the continuum limit!

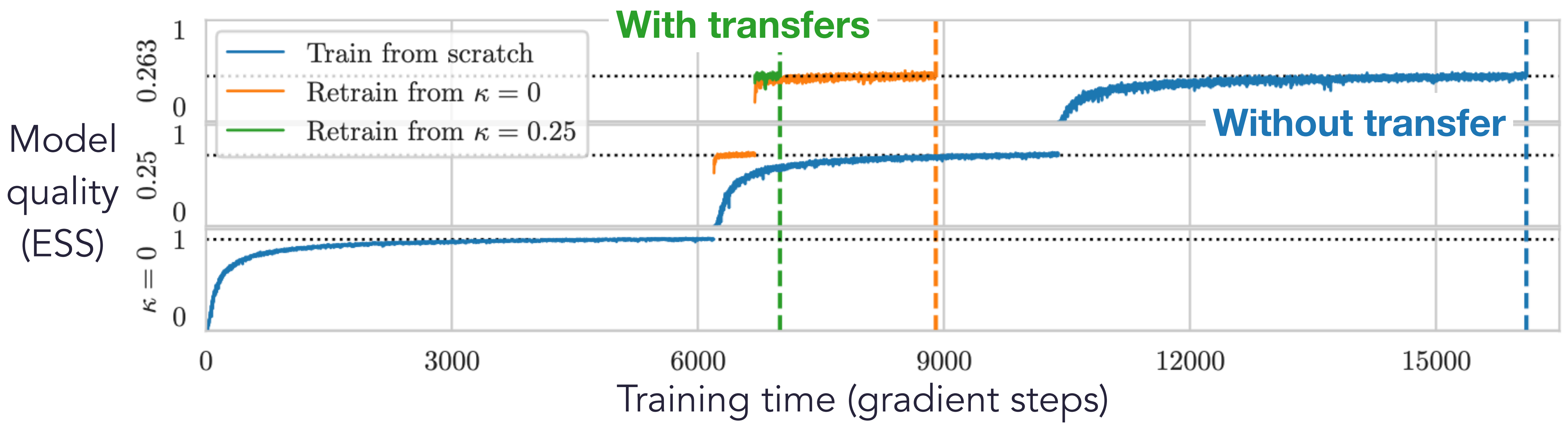
We should be thinking about targeting boxes of size $\approx \xi^d$.

- For larger volumes, hybrid/multilevel sampling schemes should be used

Transfer learning

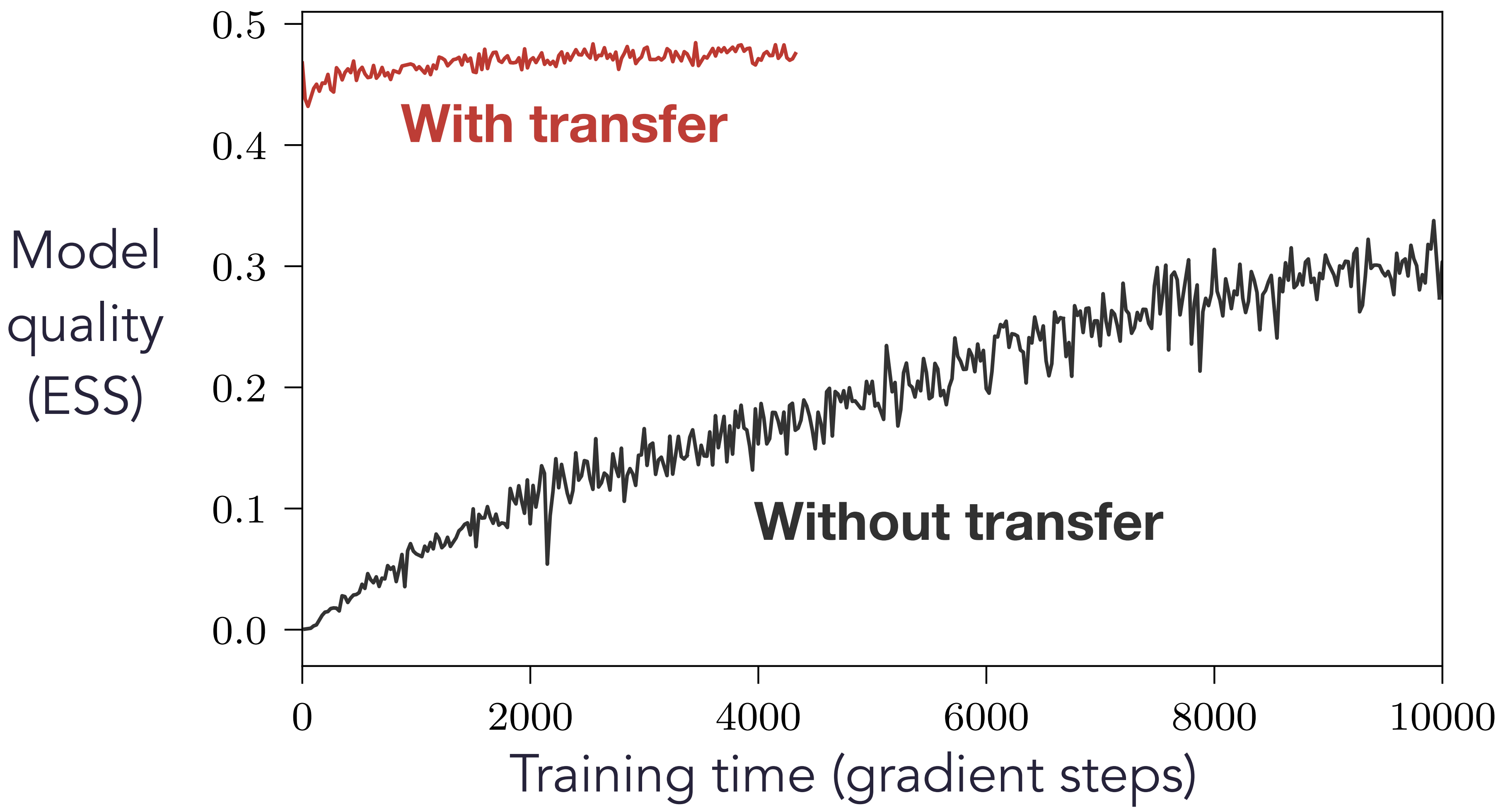
Both parameter transfer and volume transfer are highly effective for lattice field theory.

Abbott, et al. 2211.07541



- Schwinger model [U(1) gauge theory + fermions]
- Parameter transfer $\kappa = 0 \to 0.25 \to 0.263(\kappa_{cr})$

Boyda, GK, … PRD103 (2021) 074504

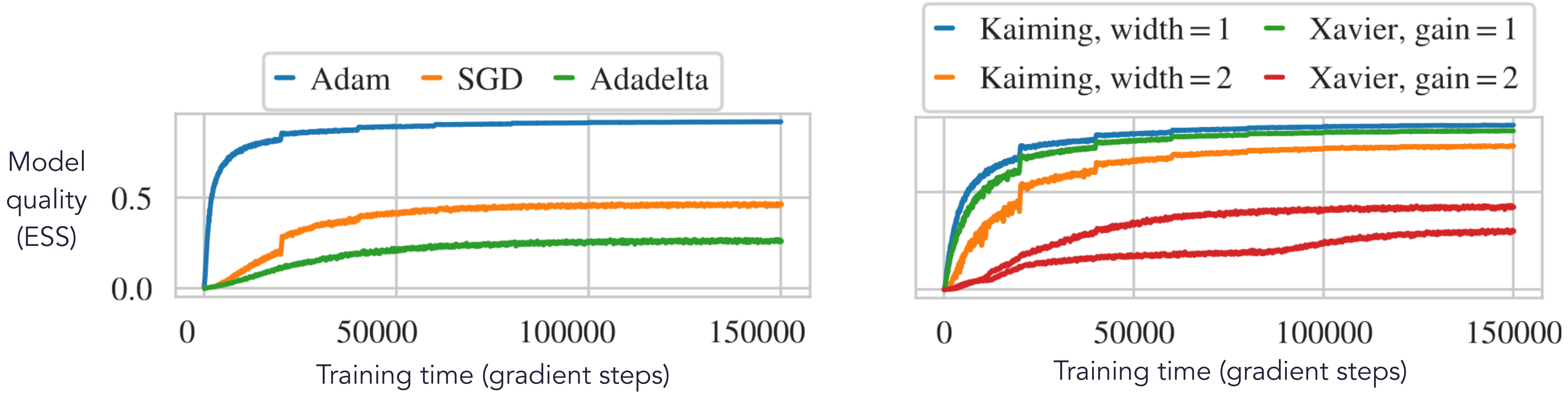


- SU(N) gauge theory
- Volume transfer $8 \times 8 \to 16 \times 16$ (red)
- Directly start at 16×16 (black)

Hyperparameters can make a big difference

Optimization algorithm, hyperparameters,
and initialization have strong effects on training rate.

Abbott, et al. 2211.07541

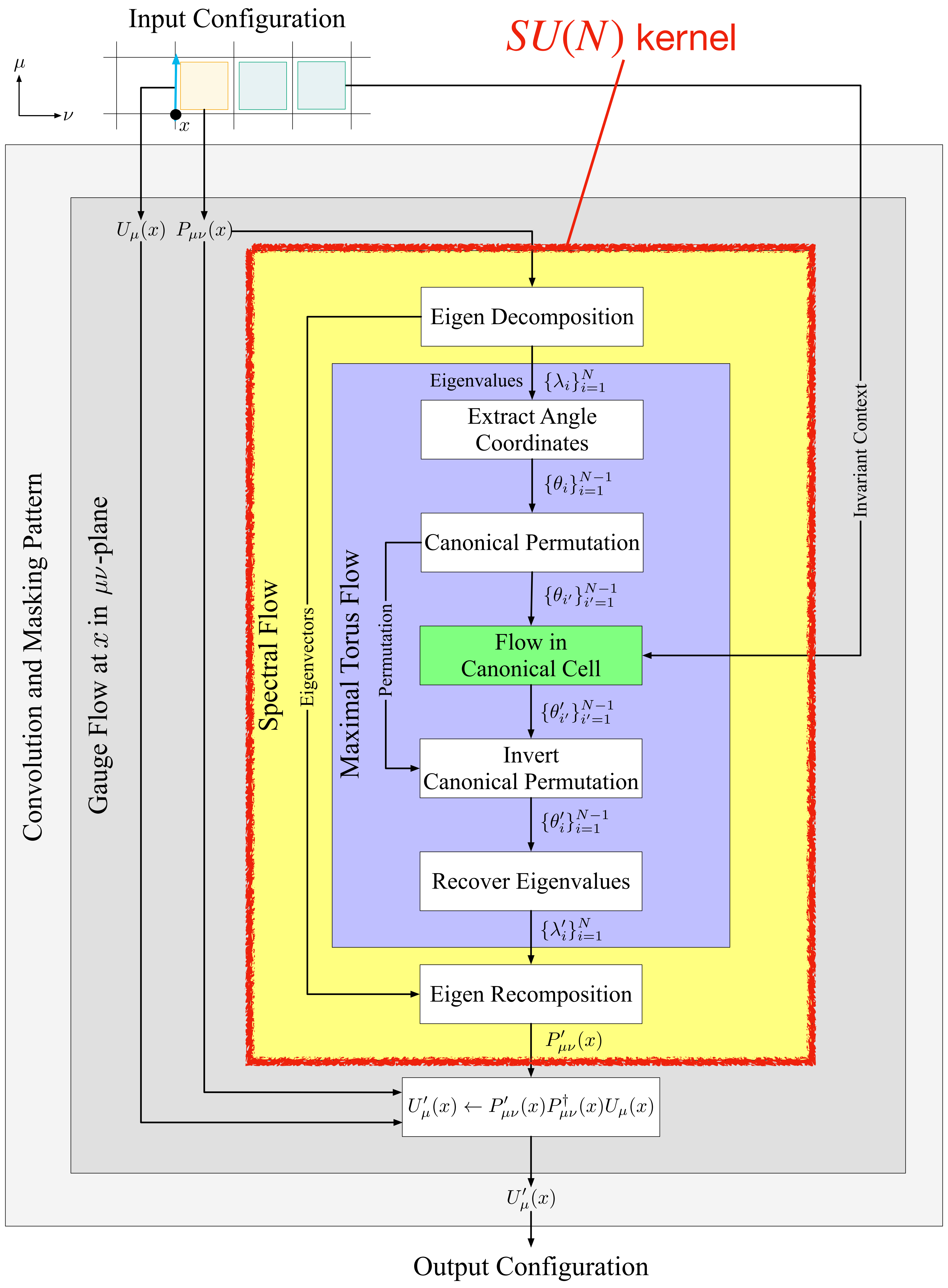


Kernel for $SU(N)$ theories

Intuition: should move points between conjugacy classes, without moving around within CCs

Conjugacy classes for $SU(N)$ described by **spectrum** of the matrix: unordered set of eigenvalues. Kernel should transform spectrum!

- Act on list of eigenvalues
- Equivariant under permutations



SU(N) kernels: strategy

SU(N) matrix-conj. equivariance is **non-trivial**.

$$h(\Omega W \Omega^\dagger) = \Omega h(W) \Omega^\dagger$$

Useful observations:

- Conjugation only rotates eigenvectors.
- Spectrum is invariant.
- Wilson loop spectrum encodes gauge-invariant physics → **This is what we want to transform.**

Strategy: Invertibly transform only the spectrum of W via a "spectral map".

Or, "spectral flow".

[Boyda, GK, Racanière, Rezende, Albergo, Cranmer, Hackett, Shanahan **PRD103 (2021) 074504**]

SU(N) kernels: strategy

SU(N) matrix-conj. equivariance is **non-trivial**.

$$h(\Omega W \Omega^\dagger) = \Omega h(W) \Omega^\dagger$$

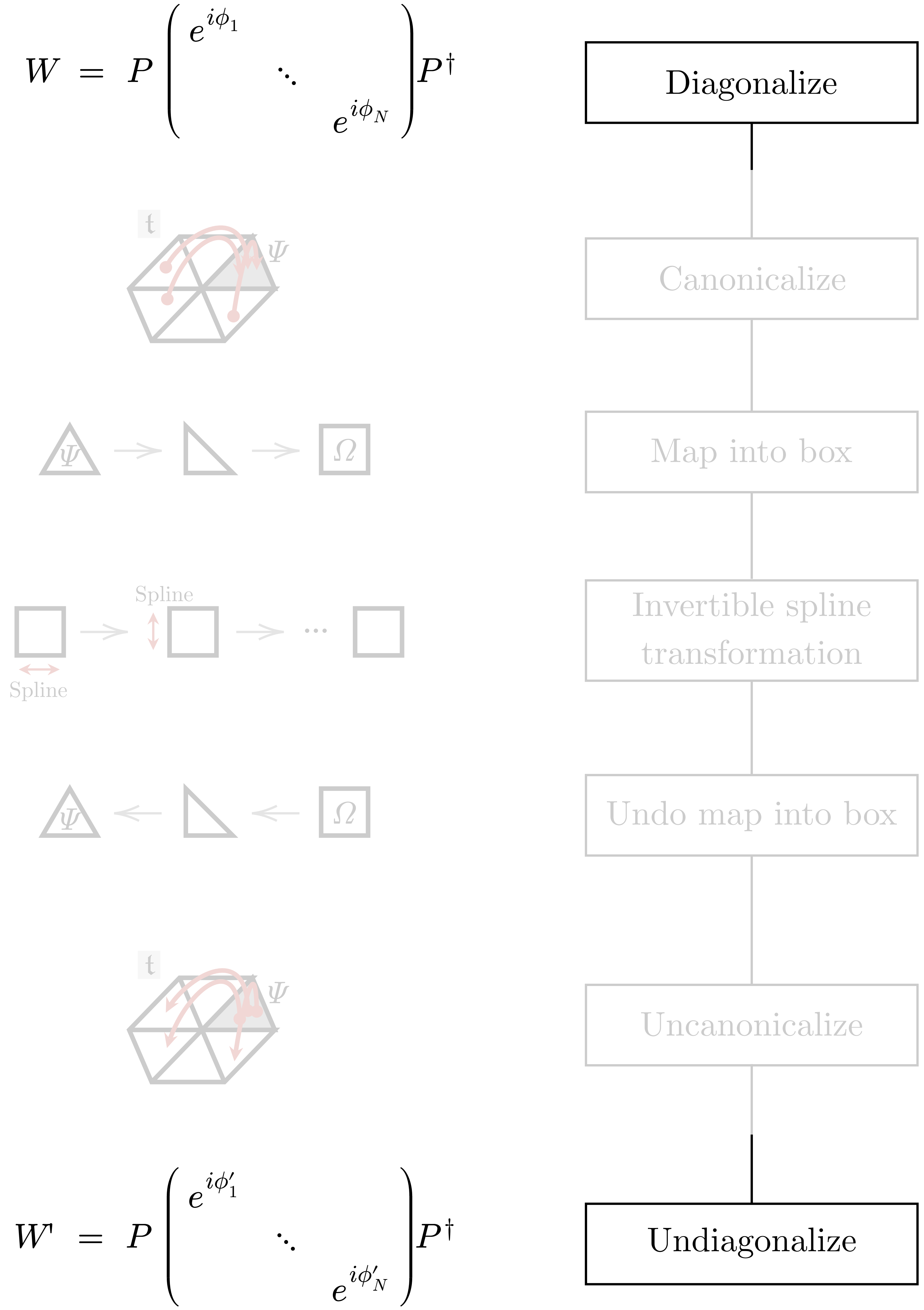
Useful observations:

- Conjugation only rotates eigenvectors.
- Spectrum is invariant.
- Wilson loop spectrum encodes gauge-invariant physics → **This is what we want to transform.**

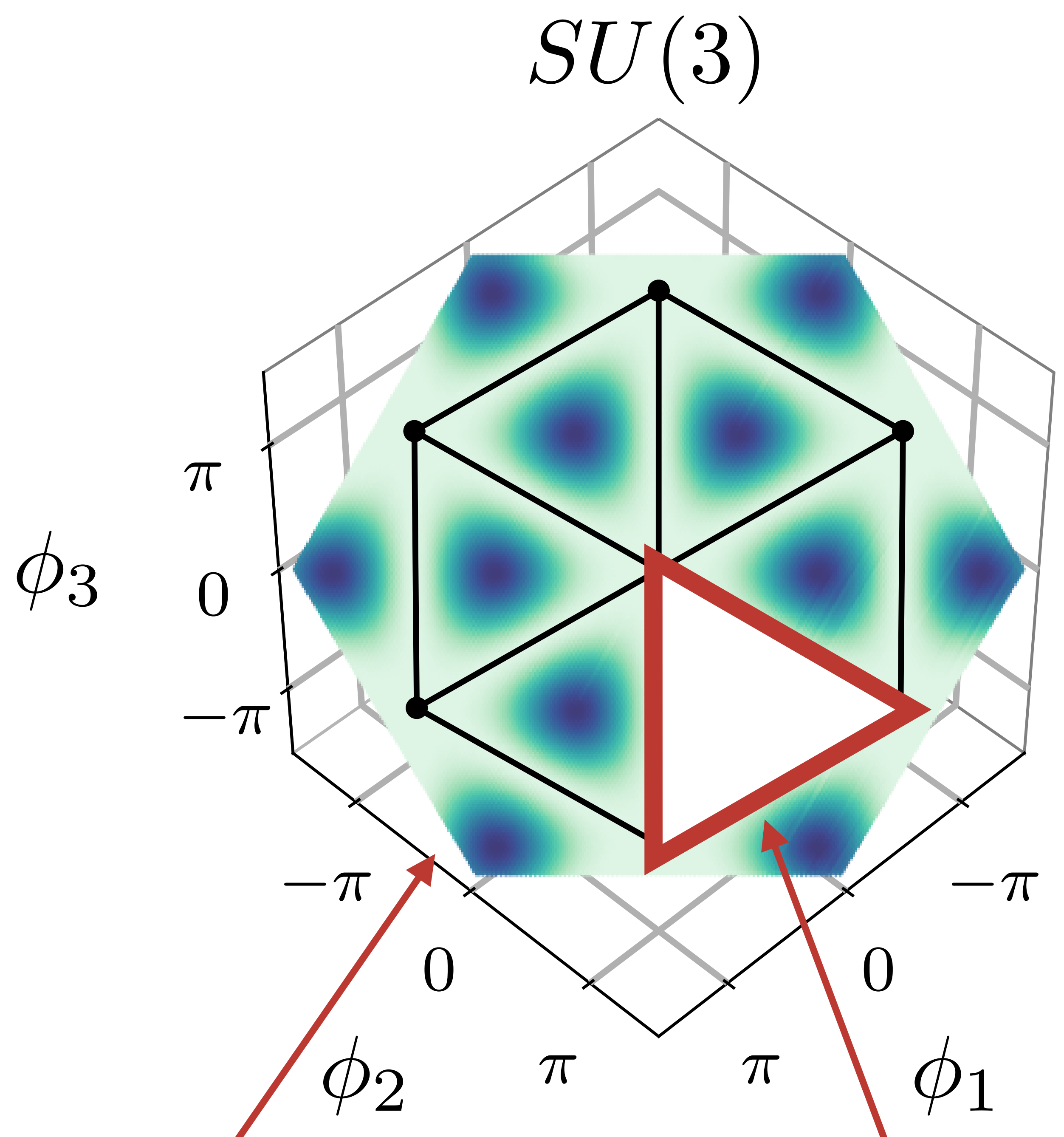
Strategy: Invertibly transform only the spectrum of W via a "spectral map".

Or, "spectral flow".

[Boyda, GK, Racanière, Rezende, Albergo, Cranmer, Hackett, Shanahan **PRD103 (2021) 074504**]

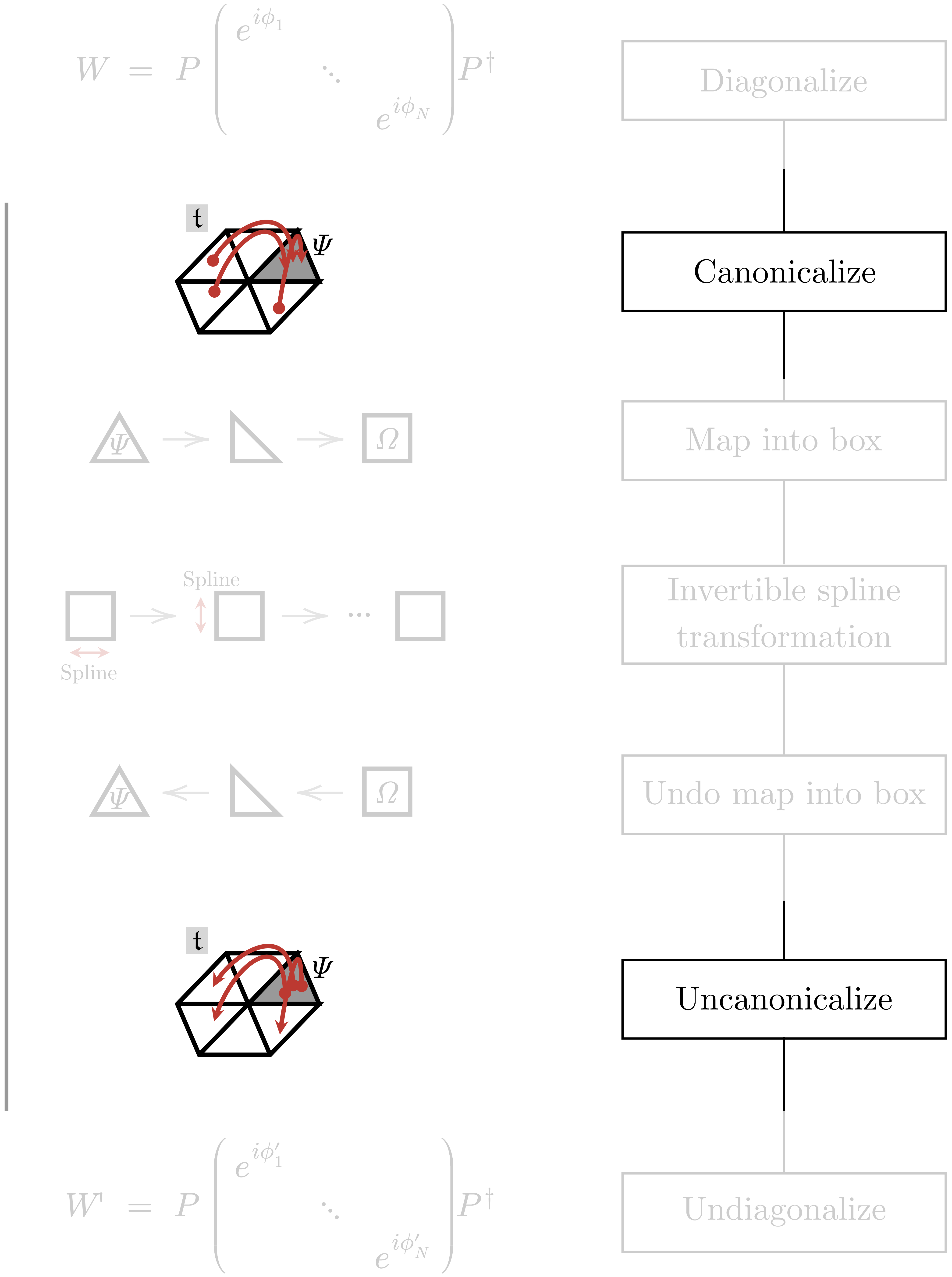


SU(N) kernels: Permutation equivariance



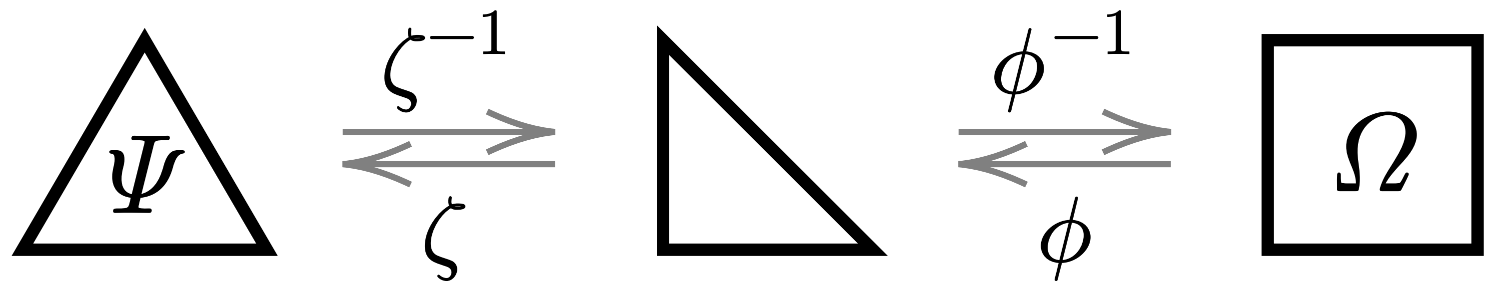
Sub-manifold of $\det = 1$ eigenvalues

"Cell", related to other cells by permutations of $\{\phi_1, \phi_2, \phi_3\}$.



SU(N) kernels:
Transform the canonical cell

Change variables to rectilinear box Ω



Transform by acting on coords of box Ω, either…

Autoregressive … or … Independent