

Machine Learning and Lattice QCD

Jacob Finkenrath, 29. September 2026, Larnaca, Cyprus

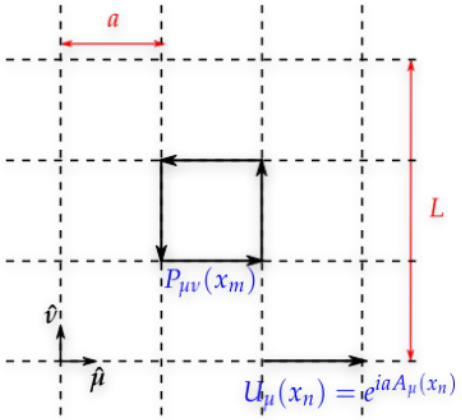
Funded by the European Union

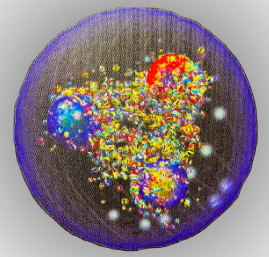
erc
European Research Council
Established by the European Commission

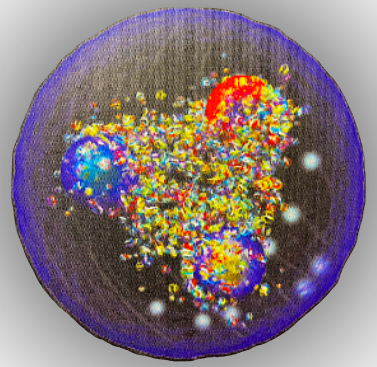
BERGISCHE UNIVERSITÄT WUPPERTAL

Table of contents

- ✦ Motivation
- ✦ Introduction
- ✦ Normalizing flows
- ✦ ML application towards the long range







See more complete picture of the field:

Dan Hackett Lattice 2026 Plenary

ML workshop at Trento 2026

AI in Europe

Lattice QCD calculation depend on large scale resources
European resources are sharing more and more with AI a
EuroHPC: AIFactories

RAISE: in pilot stage to funnel funding to European AI research
- towards AI Gigafactories:Trends towards LLM-requirements

SCAINCE:

- Working group in fundamental physics
  Will draft recommendations for future European policies of AI in Science
  Towards April 27, community white paper on AI in LQFT

Here focus on:
- Does ML in LQFT require large scale computations ?
- And if How ?

AI in HEP and TH

Application in HEP: See e.g. EuCAIF conference: Heidelberg 26

- ML enters the whole production chain
- From running accelerator to the interpretation of measurements

For example

- Detector simulations, speeding up calculations by 100x
- Jet reconstructions, resolve multi-jet events

Theory ex machina: R. Winterhalder, talk EuCAIF26

In theoretical particle physics:

- prediction have to be reliable, ML is not replacing but enhancing/optimizing the common analysis chain

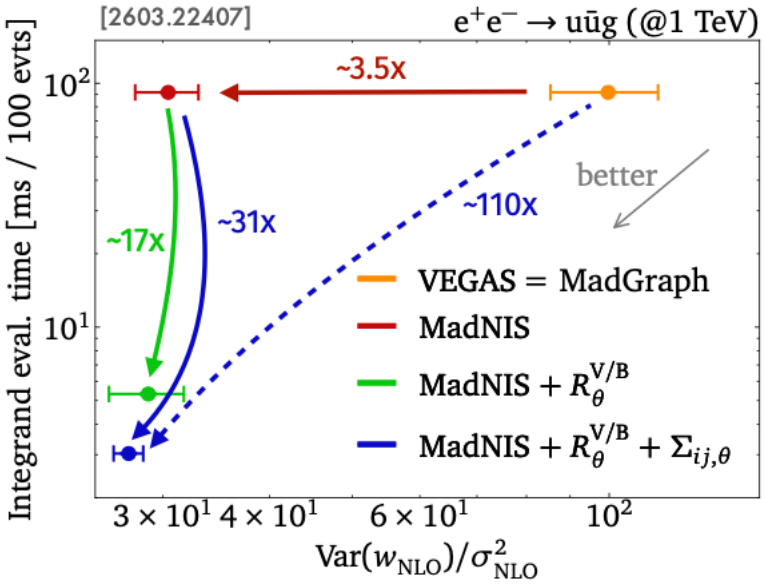
- Trend towards the use of more agentic AI
  - Not here, but …

See talk by Tunstall at CERN 2026

Hugging face: Ultra-scale book

Hugging face: Smol training book

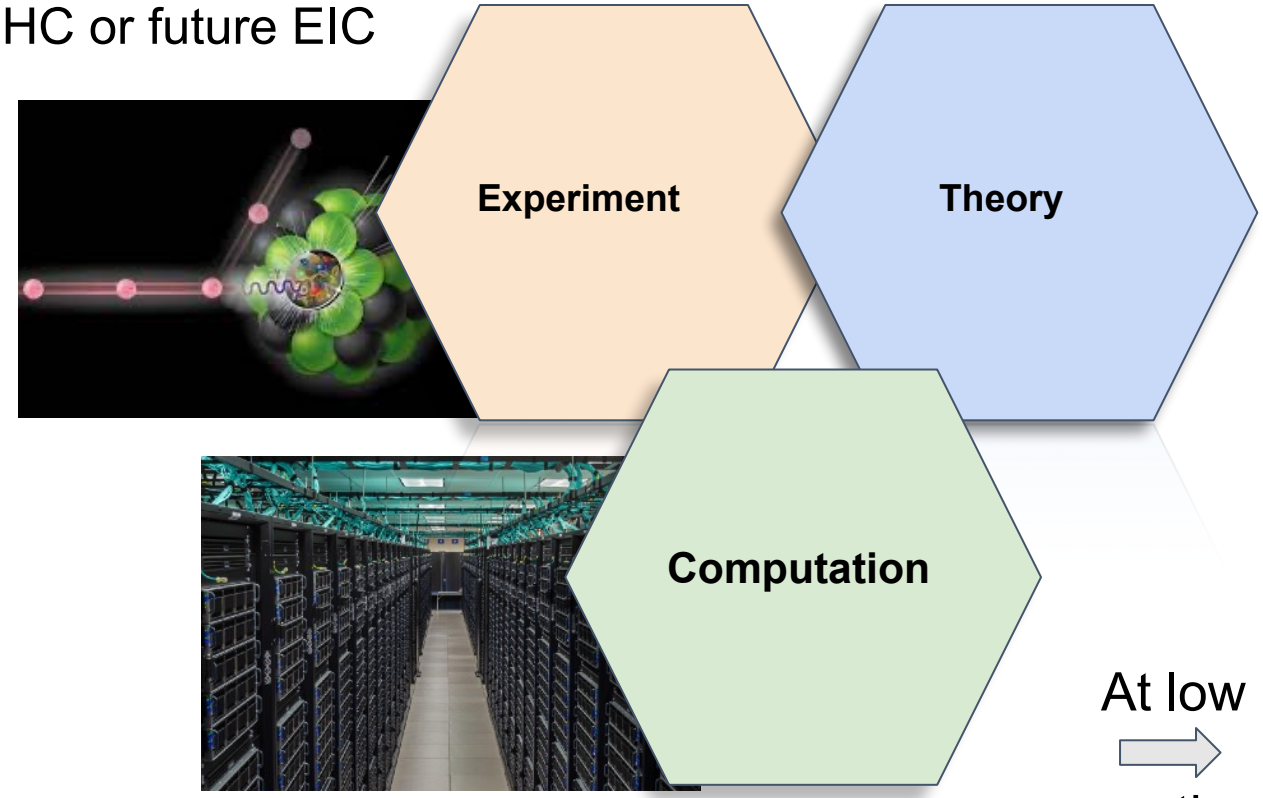
Event generation



Our motivation

Understanding the building blocks of our world

Particle Collider: like LHC or future EIC

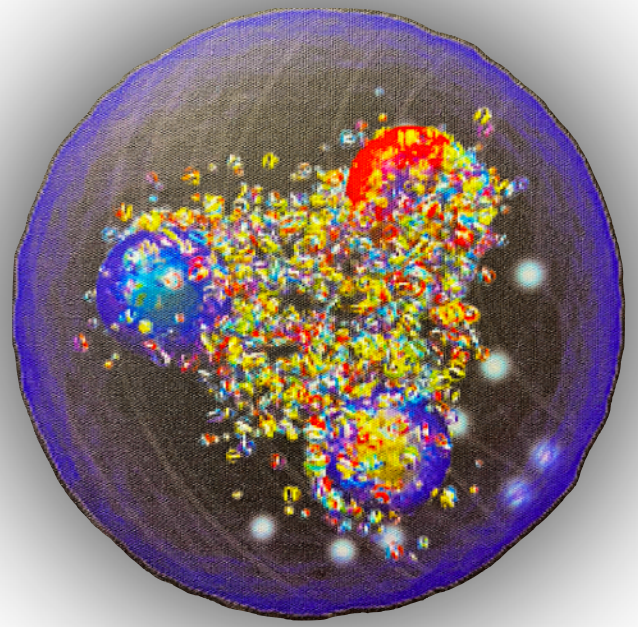


Lagrangian of Standard model - QED, QCD, Weak Interactions - solvable via perturbation theory at high energy

$$\begin{aligned}\mathcal{L} = & -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \\ & + i\bar{\Psi}\not{D}\psi \\ & + D_\mu\Phi^\dagger D^\mu\Phi - V(\Phi) \\ & + \bar{\Psi}_L\hat{Y}\Phi\Psi_R + h.c.\end{aligned}$$

At low energies QCD is non-perturbative ⇨ Lattice QCD only known ab initio method to predict

Strong Interaction



Strong interaction:

❖ Quarks and gluons are glued together via strong force:
Described by Quantum Chromodynamics:

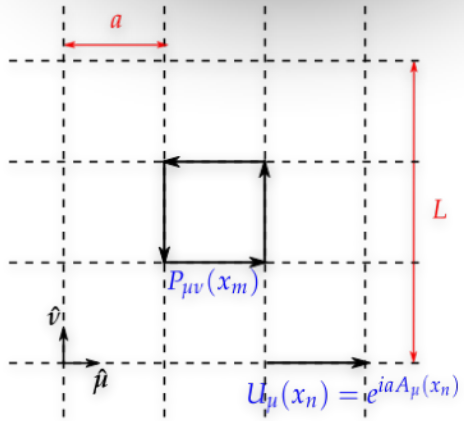
$$\mathscr{L}_{QCD} = -\frac{1}{4} F^a_{\mu\nu} F^{a\,\mu\nu} + \sum_{f=u,d,s,c,b,t} \overline{\psi}_f \left(i\gamma^\mu D_\mu - m_f \right) \psi_f$$

Unique properties:

❖ Confinement

❖ Asymptotic freedom

❖ Mass generation via interaction



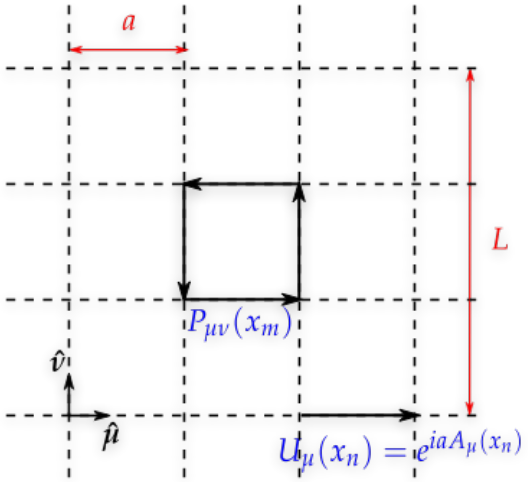
❖ QCD defined on a lattice provide a non-perturbative regularisation

Lattice Gauge Theory

Evaluation of observables via the path integral:

$$\langle \mathcal{O} \rangle = Z^{-1} \int D[U, P, \phi]\, \mathcal{O}(U) \cdot e^{-H(U,P,\phi)}$$

a high dimensional integral of up to
5 500 000 000 degrees of freedom in U



1. Hybrid Monte Carlo (MCMC)

Importance sampling via Boltzmann weight by evolving (U,P) via Hamilton's Equations:

$$\dot{P} = -\frac{\partial H}{\partial U} \quad \text{and} \quad \dot{U} = \frac{\partial H}{\partial P}$$

Force calculation
> 10 000 per Trajectory

Inversion of Dirac operator

$$D(U) \cdot x = b$$

2.) Observable calculation

Primary observables are calculated on a set of configurations:

$$\langle \mathcal{O} \rangle = \frac{1}{N} \sum_{i}^{N} \mathcal{O}(U) + O\left(\sqrt{2\tau_{int}/N}\right)$$

Fermion propagators
> 1 Mi per measurement

Computation in LQCD

See also P. Shanahan talk at CERN 2026

Computational effort:

MCMC Simulations | Measurements

Precision physics requirement:

- Exactness of prediction: statistics and systematics have to be known to scrutinise nature

-> ML as algorithmic tool to speed up computation

$$\langle \mathcal{O} \rangle = \frac{1}{N} \sum_{i}^{N} \mathcal{O} + O\left(\frac{2\tau_{int}}{N}\right)$$

MCMC simulation:

- Critical slowing down of autocorrelation time $\tau_{int} \propto a^{-5}$

Measurements:

- improved operators with better overlap

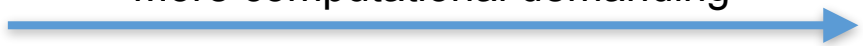
Target of using AI: Accelerating algorithms/computation to obtain result faster
- but without losing **principle of exactness**

AI MCMC Simulations | AI-acc Measurements

Computation in LQCD

Different scales, different challenges

More computational demanding



Short range	Intermediate range	Long range
- e.g. local gauge noise, smearing or higher statistics - mainly require local algorithms **Challenging example:** - Glueballs-operators	- e.g. instantons, disconnected diagrams - common techniques often not separated from long-range **Challenging example:** - critical slowing down of MCMC due to topological freezing	- e.g. small eigenvalues, solution of Dirac equation, - computational challenging and demanding towards physics quark masses **Challenging example:** - Solution of Dirac equation

Machine Learning techniques - Outline

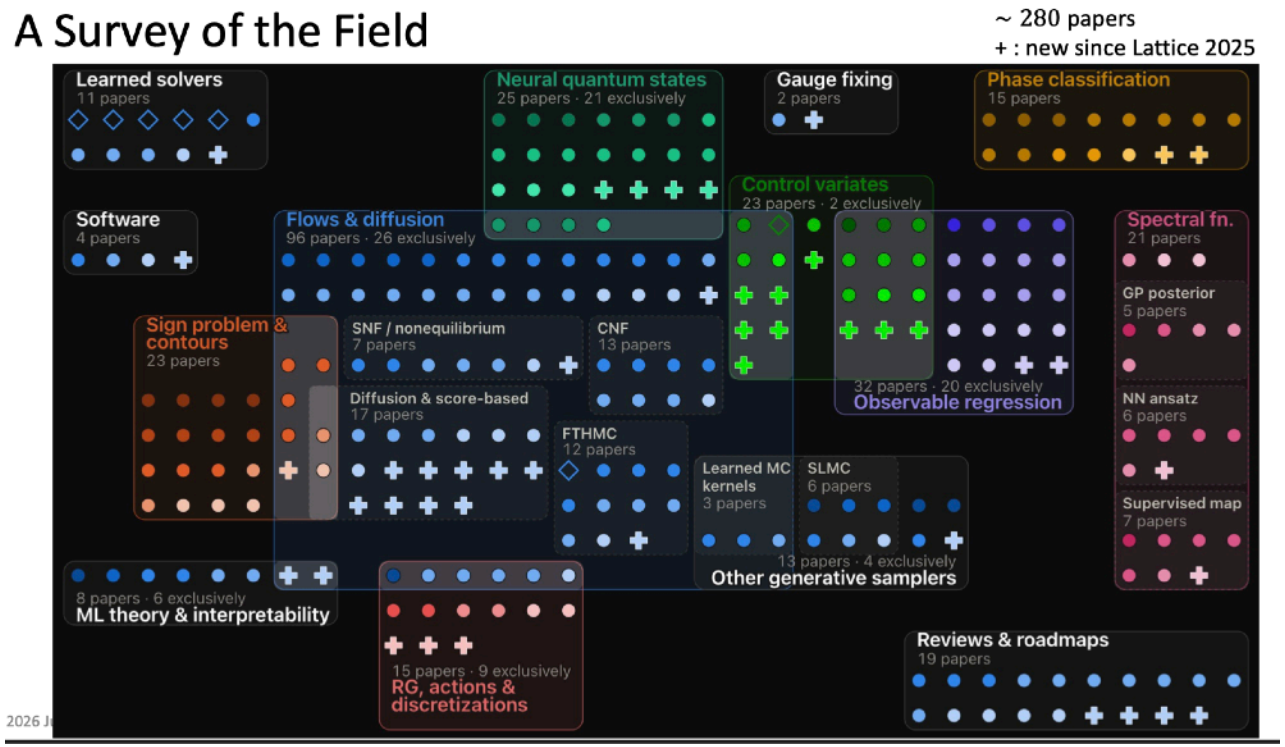
Basics

- Learnable functions: f(U) = V
- Symmetries and network architectures
- Update algorithms via flows

Some applications:

- **From short range towards long range**
  - Learning generative function
  - Multi-level approaches
  - Preconditioners for solvers

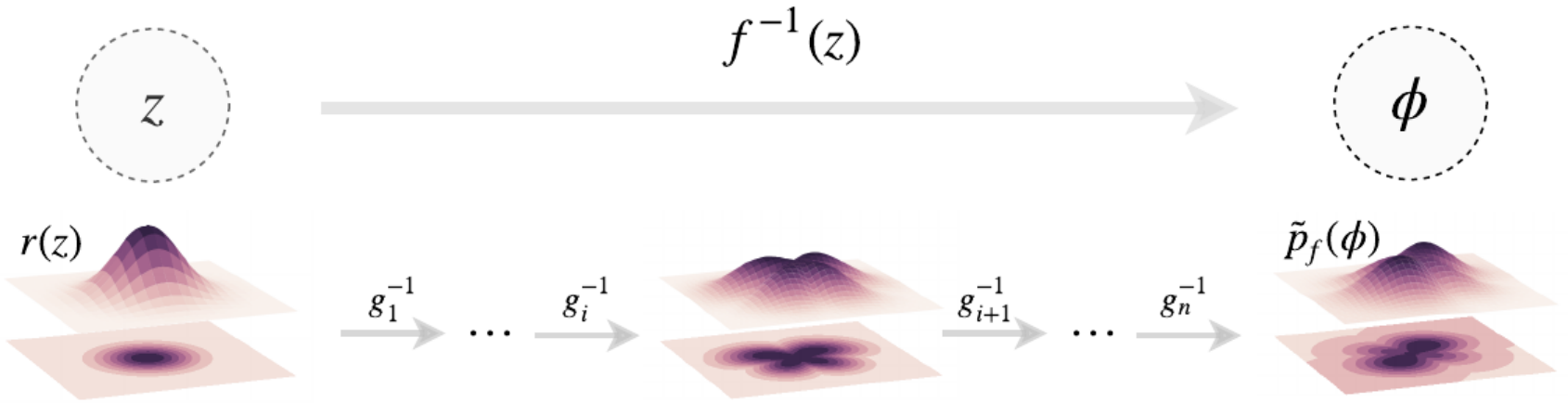
Following: See Dan Hackett Lattice 2026 Plenary



See also for a more complete list of references:

Dan Hackett Lattice 2026 Plenary

Normalizing flows



Base distribution

- Tractable density $r(z)$
- exact sampling

e.g. Haar uniform

Flow transformation

- parametrised by NNs
- Invertible (diffeomorphism)
- Tractable Jacobian

$\det J = |\det \partial V/\partial U|$

Learned model distribution

- function of flow params
- Exact sampling via tractable density

$$q(U) = \tilde{p}(V)\ \det J$$

Albergo et al., arXiv:1904.12072

Kanwar et al., Phys.Rev.Lett.125 (2020) 12, 121601

Boyda et al., Phys.Rev.D 103 (2021) 7, 074504

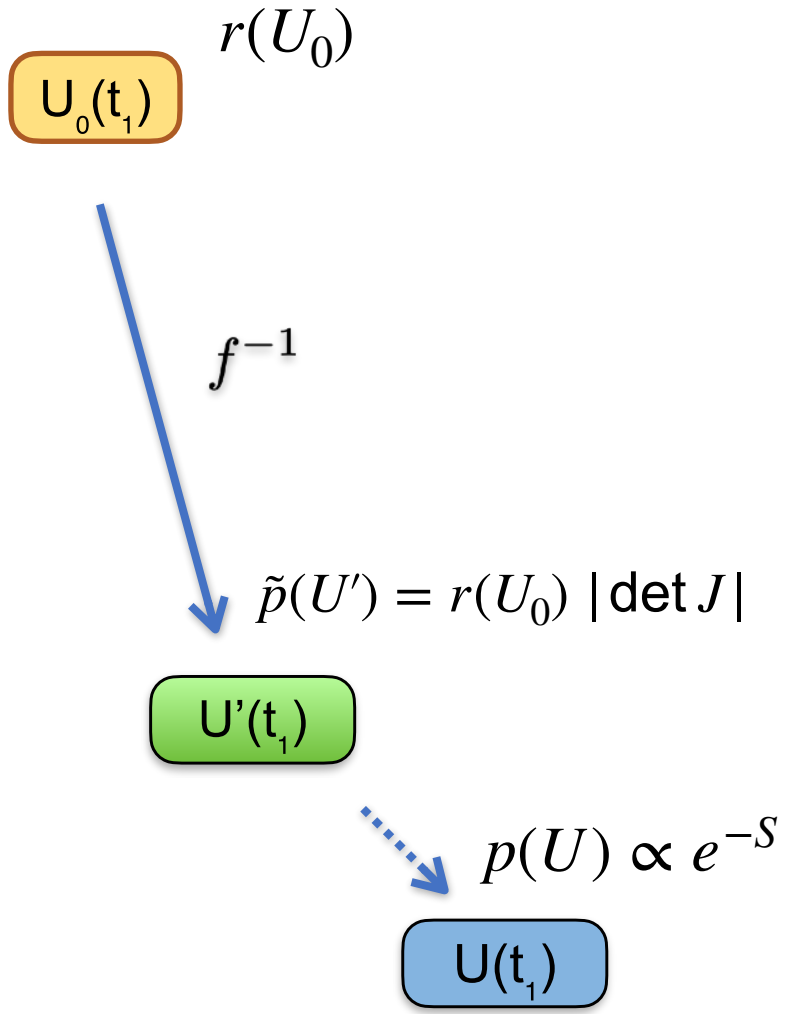
Albergo et al., arXiv:2101.08176

Flow model samplers

Step 1: define a flow model ansatz $\tilde{p} = r\,|\det J|$

Step 2: train the flow by approximating target distribution such $\tilde{p} \approx p \propto e^{-S}$

Step 3: compute exact expectation values in target theory by reweighting / accept-reject steps



Independent samples from $\tilde{p}$
- advantage: no autocorrelations

Cost: extra variance from reweighting

$$\text{ESS} = \frac{1}{\langle w^2 \rangle_{\tilde{p}}} \qquad \text{with} \qquad w = \frac{p}{\tilde{p}}$$

Symmetries

Symmetries: include symmetric in flow maps
- necessity otherwise mismodeling high dimensional symmetry.

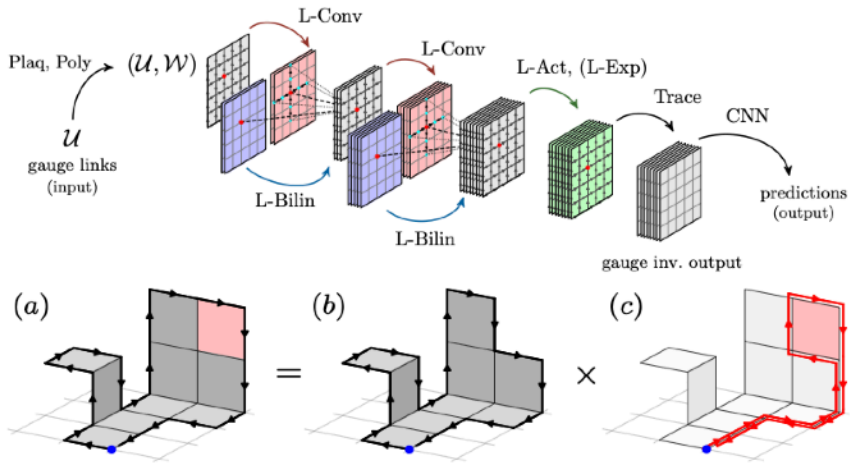
Gauge symmetry Kanwar et al., Phys.Rev.Lett.125 (2020) 12, 121601

- Design gauge invariant map such that:

$$f(\,g(U)\,) = g(\,f(U)\,)$$

Convolutional models:
- include partial translational invariance
- allows volume transfer: invest costs $\sim L^4$

Favoni et al., 2012.12901



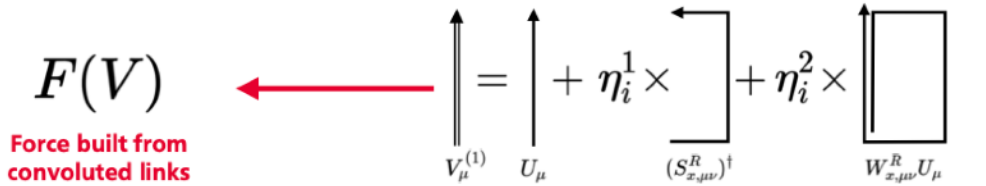
Possible architecture: Residual layers:

$$U'_\mu(x) = e^{F(U)} U_\mu(x) \;\rightarrow F = \sum_i \delta_i P(W^i_{\mu\nu})$$

- With δ_i trainable network and P hermitian trace
- build arbitrary loops "convoluting" the frozen links

Nagai, Tomiya, 2103.11965

Abbott, et al., 2305.02402



Architectures of flows on gauge theories

Luscher, 0907.5491

Bacchio et al., 2212.08469

Gerdes et al. 2410.13161

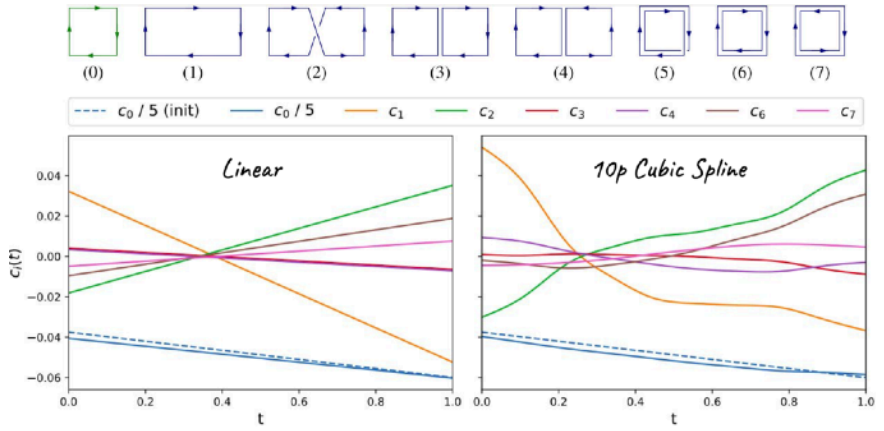
Continuous flows: learned gradient flows

Flow by integrating ODE from $t = 0 \to 1$

$$\dot{U}_{x\mu}(t) = Z_{x\mu}(U(t), t)\ U_{x\mu}(t)$$

Where Z is $su(N)$ algebra-valued

- can be learned, but not necessary $Z \neq \nabla \tilde{S}(t)$



Bacchio et al., 2212.08469

Evaluate $\tilde{p}$ by integrating:

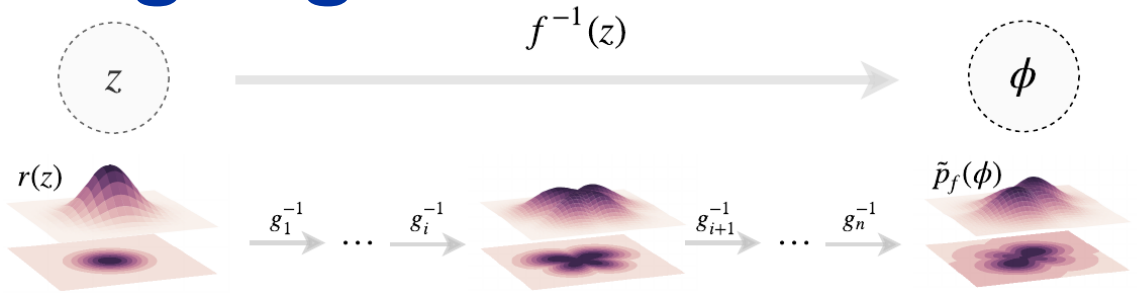
$$\partial_t q + \nabla \cdot (q\ Z) = 0 \qquad \Rightarrow \frac{d}{dt} \log q(t) = -\nabla \cdot Z(t)$$

Generalizes Luschers gradient flow for a learned time-dependent action $\tilde{S}(t)$

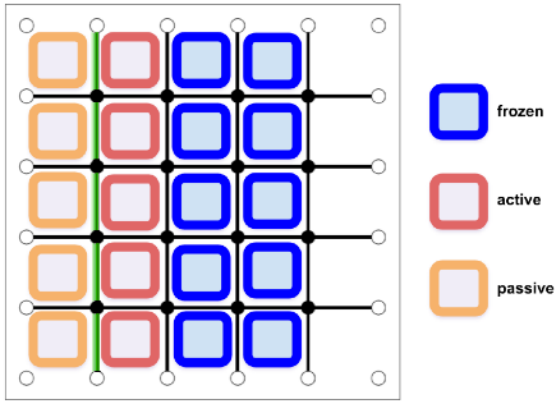
- automatically encode all symmetries of the theory by choosing $\tilde{S}(t)$ invariant

Architectures of flows on gauge theories

Discrete flows:
Build a flow by stacking many simple sub-flows



Variabel partitioning:
- freeze some links
- update other (active links)



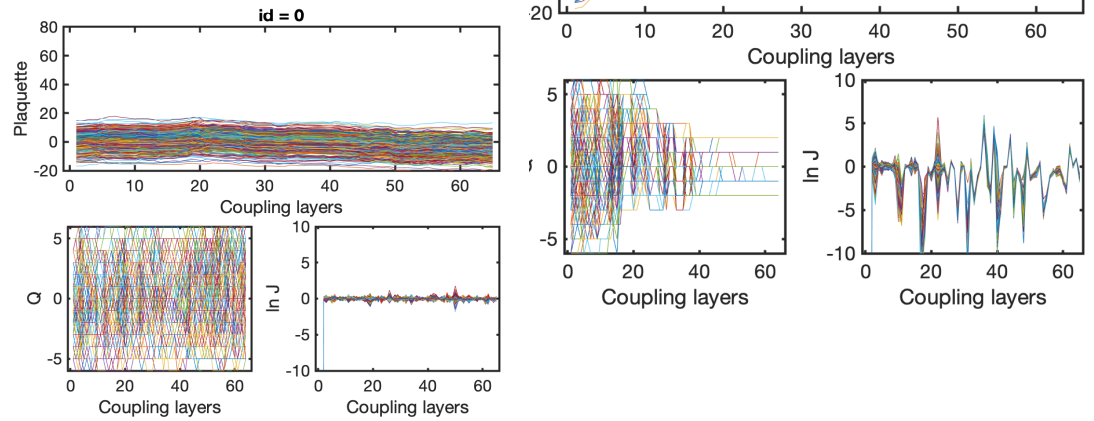
Leads to Triangular Jacobian
Alternate which links are mask per layer

- introduce coupling layers with

$$g_i^{-1}(z) := \begin{cases} \phi_a = z_a \\ \phi_b = (z_b - t_i(z_a)) \odot e^{-s_i(z_a)}. \end{cases}$$

- train the coupling layers and by minimizing the loss-function

$$L(\tilde{\rho}) = \int \prod \phi_j \, \tilde{\rho}(\phi) \ln(\tilde{\rho}(\phi)/\rho(\phi))$$

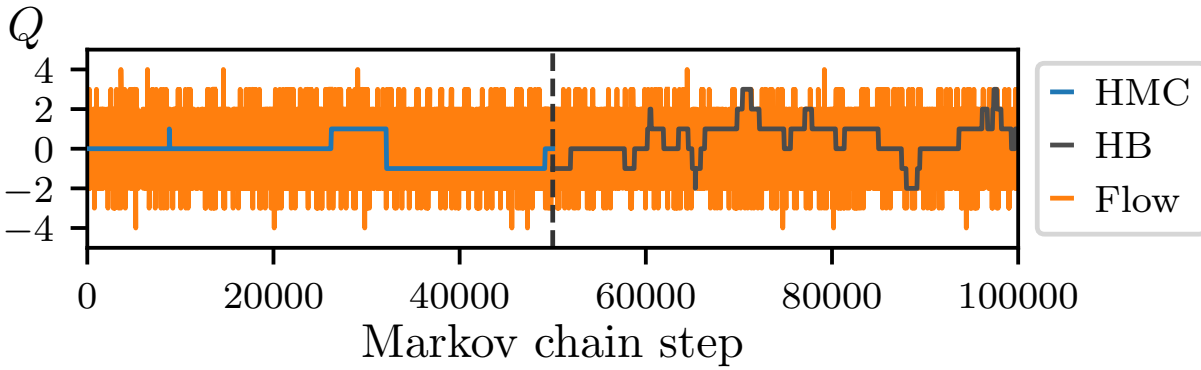


Application within MCMC

MCMC algorithms

1. Propose U' according to $T_0(U \to U')$
2. Correct with $P_{acc}(U \to U') = \min\left[1, \frac{\tilde{\rho}(U)\rho(U')}{\rho(U)\tilde{\rho}(U')}\right]$

Kanwar et al., Phys.Rev.Lett.125 (2020) 12, 121601



Solves critical slowing down in 2D U(1) model for moderate lattice sizes $L \sim 16$

We have a proposal which generates independent configurations with weight: $\tilde{\rho}(U) = r(f(U)) \cdot \left|\det \frac{\partial f(U)}{\partial U}\right|$

Need correction steps to correct towards the right weight:

Acceptance rate:

ratio of distributions $(\tilde{\rho}(U)\rho(U'))/(\rho(U)\tilde{\rho}(U'))$ is log-normal distributed.

Knechtli, Wolff, hep-lat/0303001

$$P_{acc} = \mathrm{erfc}\{\sqrt{\sigma^2(\Delta S)/8}\}$$

$$\Delta S = \ln\{\rho(U')\} - \ln\{\rho(U)\} + \ln\{\tilde{\rho}(U)\} - \ln\{\tilde{\rho}(U')\}$$

Scaling of normalizing flows

Fine tuning problem: $\sigma^2(\Delta S)$
How the variance scales

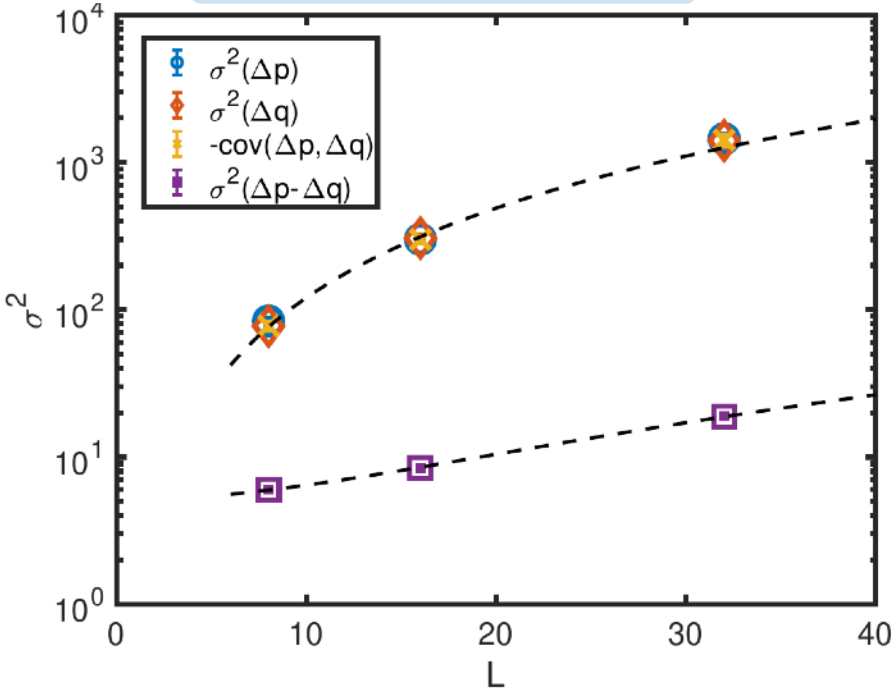
$$\mathrm{var}(\Delta\rho) + \mathrm{var}(\Delta\tilde{\rho}) \approx -2 \cdot \mathrm{cov}(\Delta\rho, \Delta\tilde{\rho})$$

Covariances of distributions scales like variances

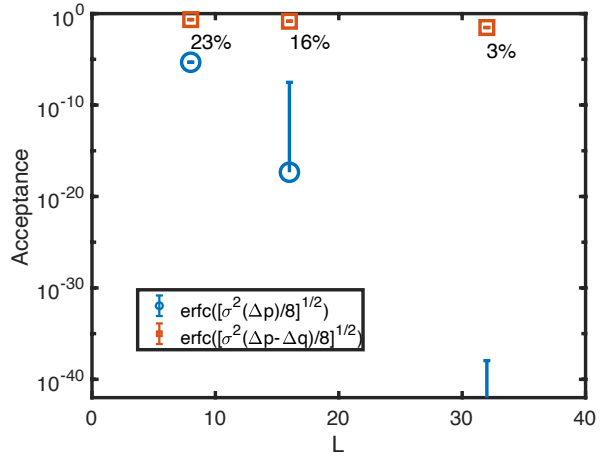
$$\sigma^2 = \mathrm{var}(\Delta\rho) + \mathrm{var}(\Delta\tilde{\rho}) - 2 \cdot \mathrm{cov}(\Delta\rho, \Delta\tilde{\rho})$$

but still grows with the volume

Del Debbio, Phys. Rev. D 104, 094507



Idea: localisation of updates
- training of localised flow J. F., arxiv:2201.02216
- insert a open BC-defect and use flow to connect it to target Bonnano et al., 2510.25704

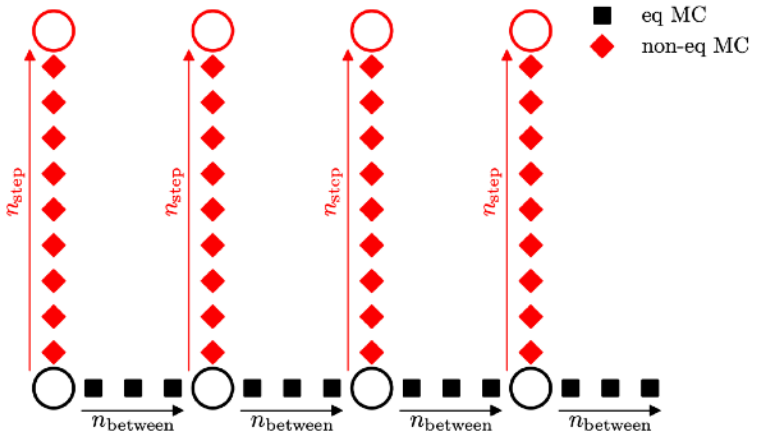


Stochastic Normalizing Flows with defect

Use defect:

Bonnano et al., 2510.25704

- set spacial links of distance d within a time slice to zero
- run MCMC with defect
- Use non-equilibrium MCMC with flow to connect to target theory without defect

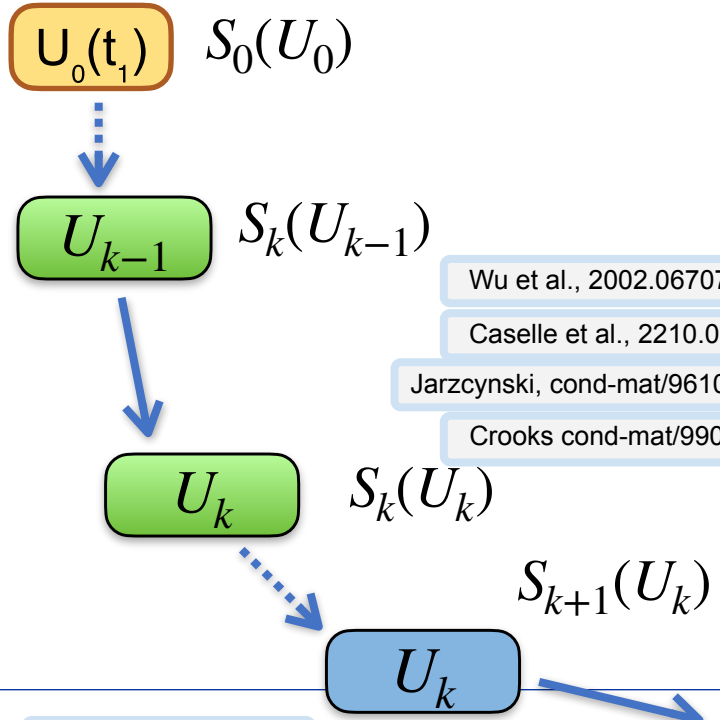


AIS / Jarzcynski / Non-Equilibrium MCMC

At each p_k apply an MCMC hit and track the work

$$\Delta W_k = S_{k+1}(U_k) - S_k(U_k)$$

Unbiased reweighting: $\langle O \rangle_p = \dfrac{\hat{w} O(U_k)\rangle}{\hat{w}}$ with $\hat{w} = e^{-\sum_k \Delta W_k}$



Wu et al., 2002.06707

Caselle et al., 2210.03139

Jarzcynski, cond-mat/9610209

Crooks cond-mat/9901352

Stochastic normalizing flows

Add deterministic flows bridging between theories

Generalized work:

$$\Delta W_k = S_{k+1}(f_k(U_k)) - S_k(U_k) - \log J_k(U_k)$$

Luscher, 0907.5491

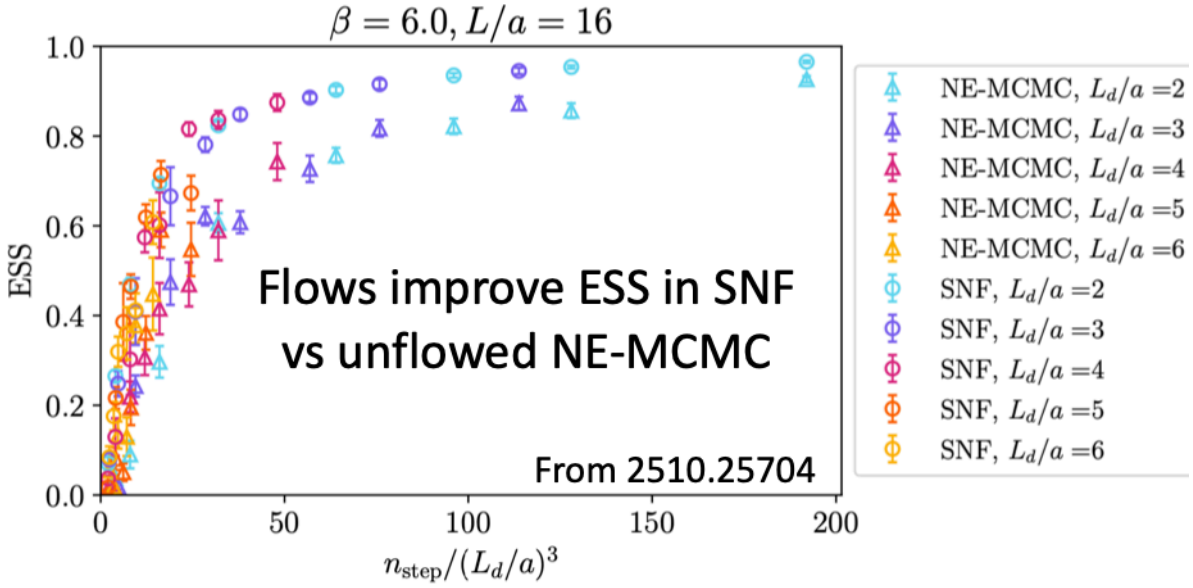
Defect Repair Flows

Bonnano et al., 2510.25704, 2601.20708

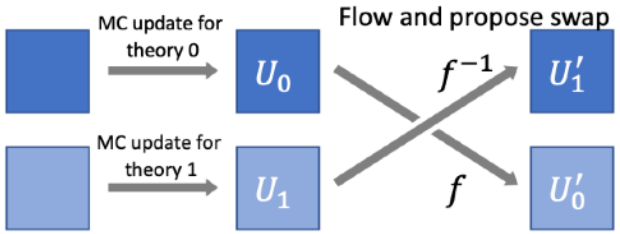
Idea: Localize flows

- with defect which can enhance toological fluctuations

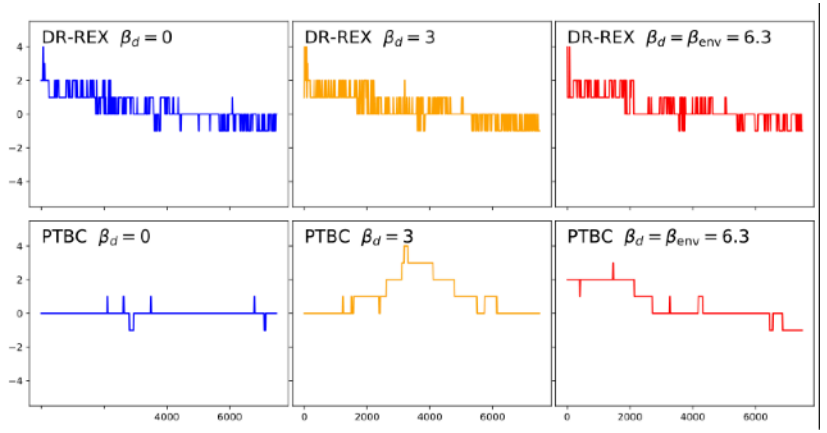
Using Stochastic normalizing flows:
Advantage in pure gauge below a ~ 0.03 fm



Another option:
Using parallel tempering with flows



$$p_{\text{acc}} = \min\left[1, \frac{p_0(U_1')\,p_1(U_0')}{p_0(U_0)\,p_1(U_1)}\,J_f(U_0)\,J_{f^{-1}}(U_1)\right]$$



- study with with fermions missing
- Still requires long chains, O(1k) auto-corr-times remaining

Abbott et al., 2404.11674

Recap - normalizing flows

Localization can overcome volume penalty
- flow out of local defects

Training of flows
- so far not scalable, learning limited to short range

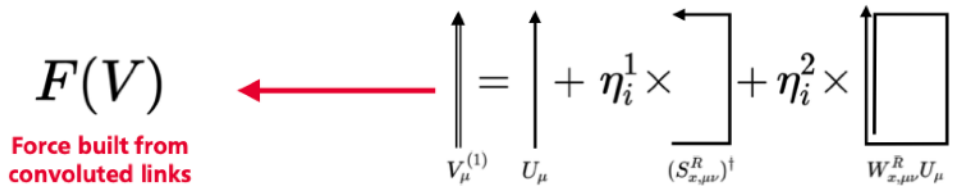
How this can be utilised in other applications ?

Short range

- e.g. local gauge noise, smearing or higher statistics
- mainly require local algorithms

Challenging example:
- Glueballs-operators

Transformation are local



Correlated ensembles and the derivative trick

F. Romero-Lopez, Y. Fu, D. Hackett, G. Kanwar, P. Shanahan

Abbott, et al., 2401.10874

Bacchio, 2305.07932

Catumba, Ramos 2502.15570

Abbott et al., 2606.15986

Idea: Lets perturb theory w/ an operator Q

$$S \to S_\lambda = S - \lambda Q$$

Now: Derivative inserts Q into n-pt functions $\langle O \rangle$

$$\left.\frac{\partial \langle O \rangle_\lambda}{\partial \lambda}\right|_\lambda = \langle OQ \rangle_0 - \langle O \rangle_0 \langle Q \rangle_0$$

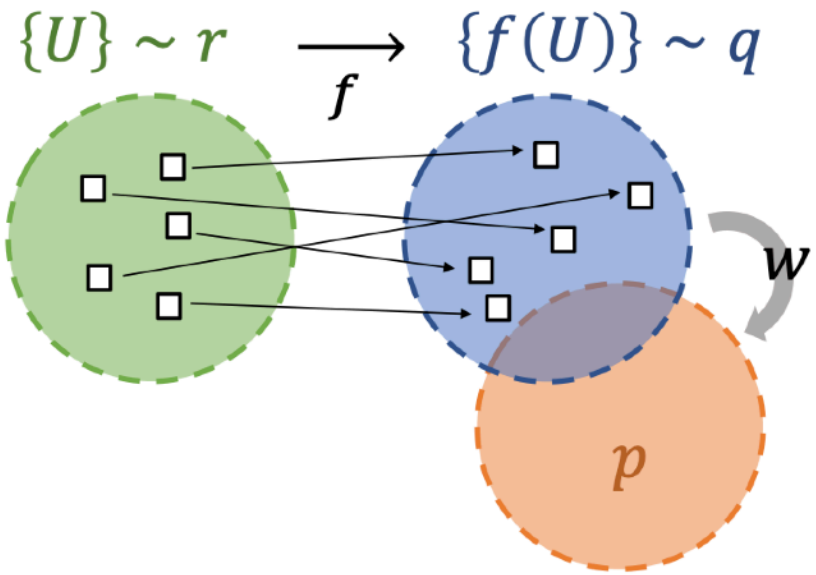
Trick: Finite difference can be approximated:

$$\left.\frac{\partial \langle O \rangle_\lambda}{\partial \lambda}\right|_0 \approx \frac{1}{\lambda}[\langle O \rangle_\lambda - \langle O \rangle_0]$$

Train a flow f to map $S_0 \to S_\lambda$ use to evaluate both expectations on same ensemble

$$= \frac{1}{\lambda}\langle w(f(U))\mathcal{O}(f(U)) - \mathcal{O}(U) \rangle_0$$

A flowed ensemble is correlated w/ the original



- Correlated noise cancelation

Idea also generalizes straightforward to Feynman-Hellmann methods

Example: Glueballs in Yang-Mills

Abbott et al., 2606.15986

F. Romero-Lopez CERN talk

This can be generalized to n-point function:
Learning generating functional

Two point function: $C(t) = \langle G(t)G(0)\rangle_c = \langle G(t)G(0)\rangle - \langle G(t)\rangle\langle G(0)\rangle$

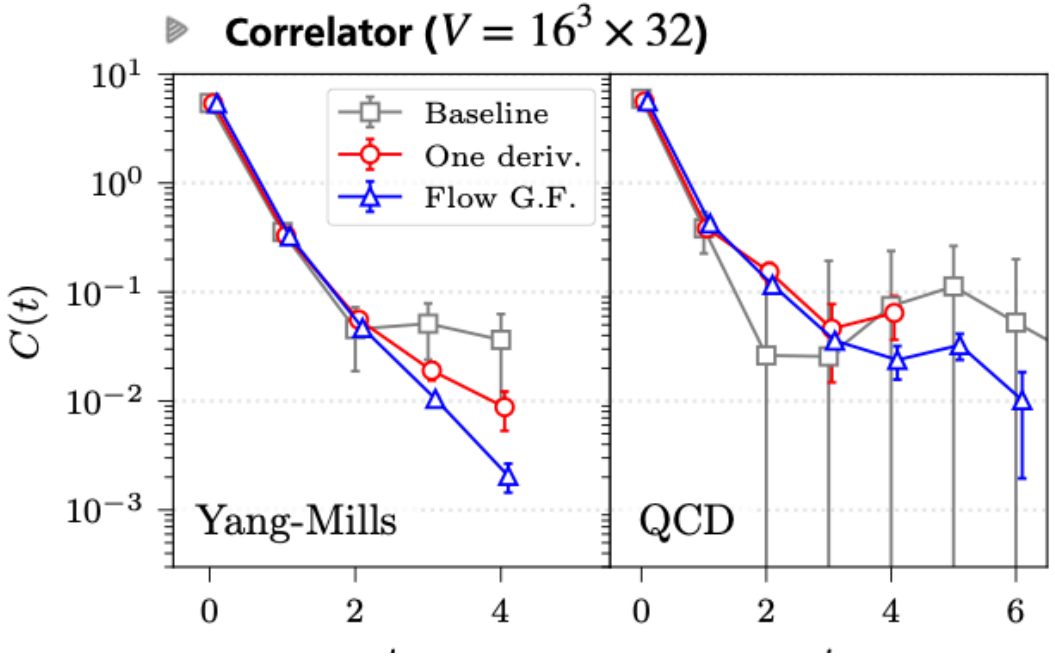
Expressed via generating functional

$$C(t) \simeq \frac{1}{\lambda\lambda'}[\langle\hat{\omega}_{\lambda,\lambda'}\rangle - \langle\hat{\omega}_{\lambda,0}\rangle\langle\hat{\omega}_{0\lambda'}\rangle]$$

with $\langle\hat{\omega}_{\lambda,\lambda'} = \langle e^{\lambda G(0)+\lambda' G(t)}\rangle$

Now: a flow can be learned which significantly reduces the variances

Tested on $V = 16^3 \times 32$
- Variance reduction by 300



Status - flow maps

Applications of Flows (classical or learned) become very versatile
- scale setting, operator optimization, improved distributions …
- first steps on-going to include fermions

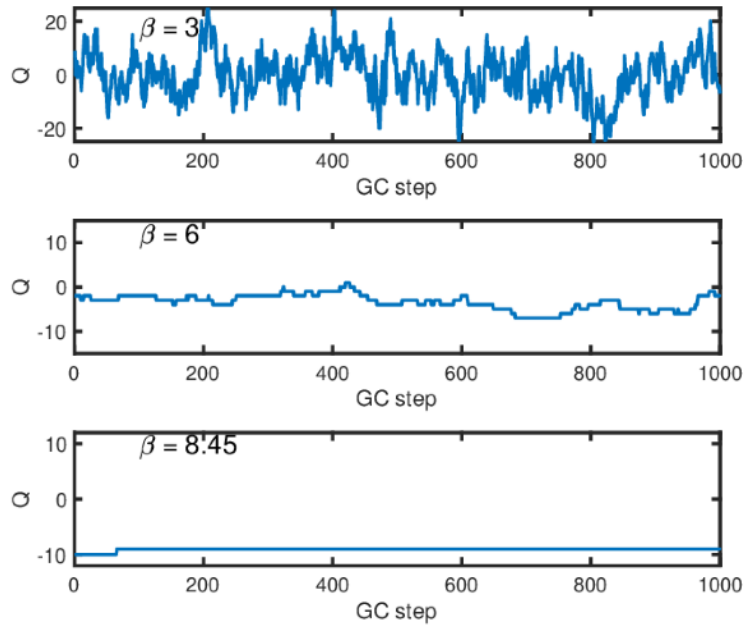
In general: any further advances on network architecture will directly improve application

Intermediate range

- e.g. instantons, disconnected diagrams
- common techniques often not separated from long-range

Challenging example:
- critical slowing down of MCMC due to topological freezing

2D Schwinger Model



severe critical slowing down

Multi-level sampling

A. Hasenfratz, E. Neil, L. Parato,
N. Schwartz

How to deal with long range correlations ?

A. Hasenfratz Lattice 26 talk

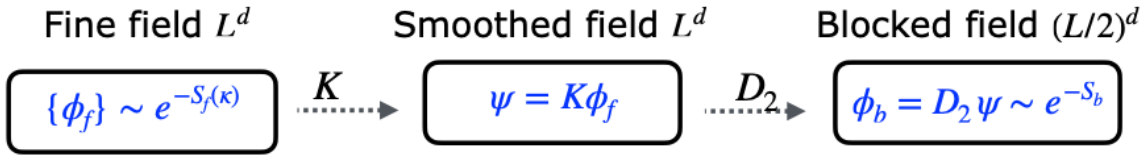
A. Hasenfratz et al., 2608.28581

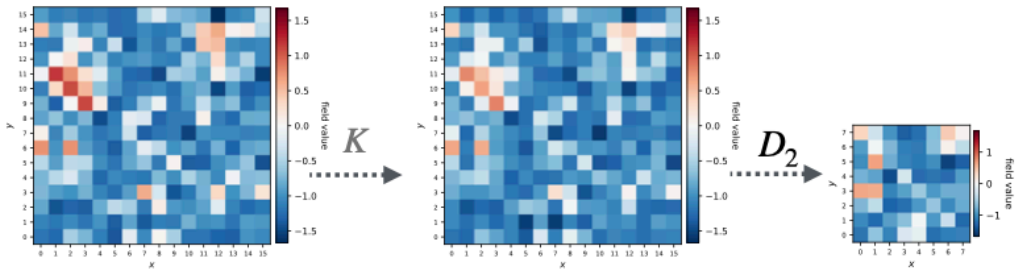
Idea: use renomalization group guided inverse blocking

Long distance IR physics difficult to learn

- networks are local, use them to feed local physics
- start with distribution which includes long range physics

RG blocking





RG blocking:

- Smooth fields $\psi = K\phi$;
  K : linear convolution
- Coarse grain or decimate:
  $\phi_b = D_2\psi = D_2K\phi$

$\{\phi_f\}$, $\{\psi\}$ and $\{\phi_b\}$ have the same mid-range and IR physics

- The coarse blocked S_b action is usually unknown

Multi-level sampling

Perfect blocking:

Select κ that the blocked ensemble is well described $\{\phi_b\} \simeq \{\phi_c\}, \quad \phi_c \sim e^{-S_c}$

At criticallity: $S_c \simeq S_b \simeq S_f$

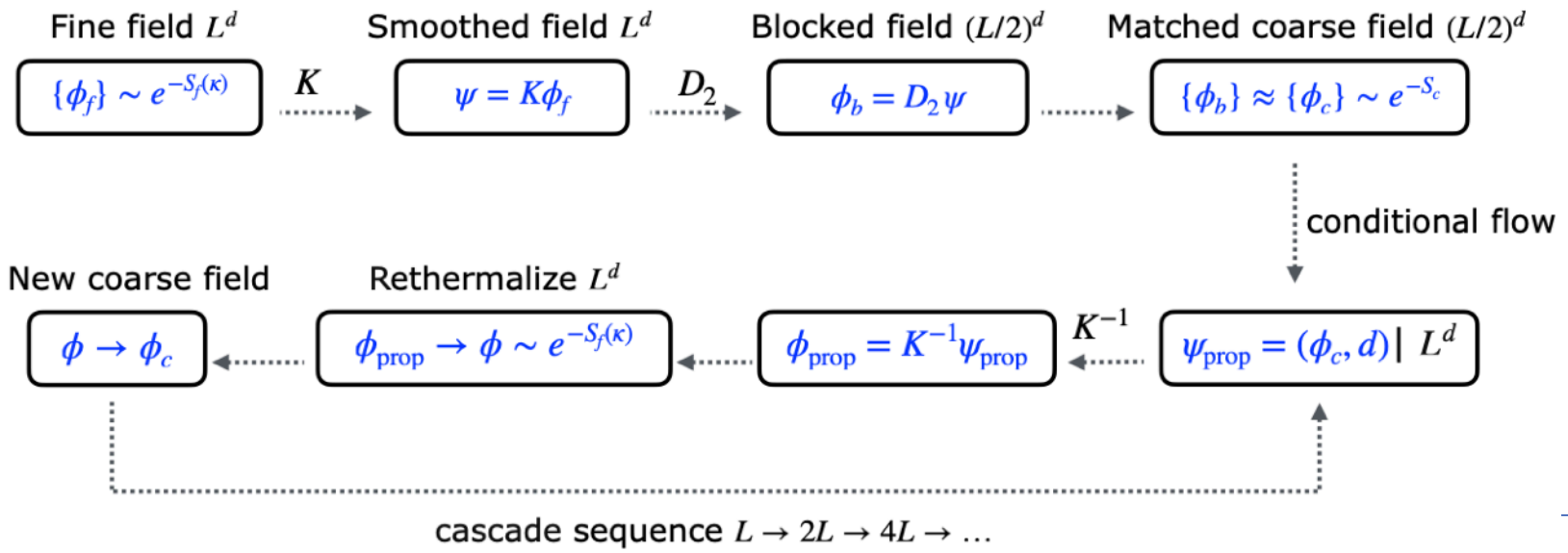
- where ϕ_c contains IR physics

Now: use a conditional flow which: $d_{01} \sim q_\theta(d_{01}\,|\,\phi_c),\ d_{10} \sim q_\theta(d_{10}\,|\,\phi_c, d_{01}),\ d_{11} \sim q_\theta(d_{11}\,|\,\phi_c, d_{01}, d_{10})$

- network learns the missing UV conditional distribution

In practice:

- using rethermalization and iterate



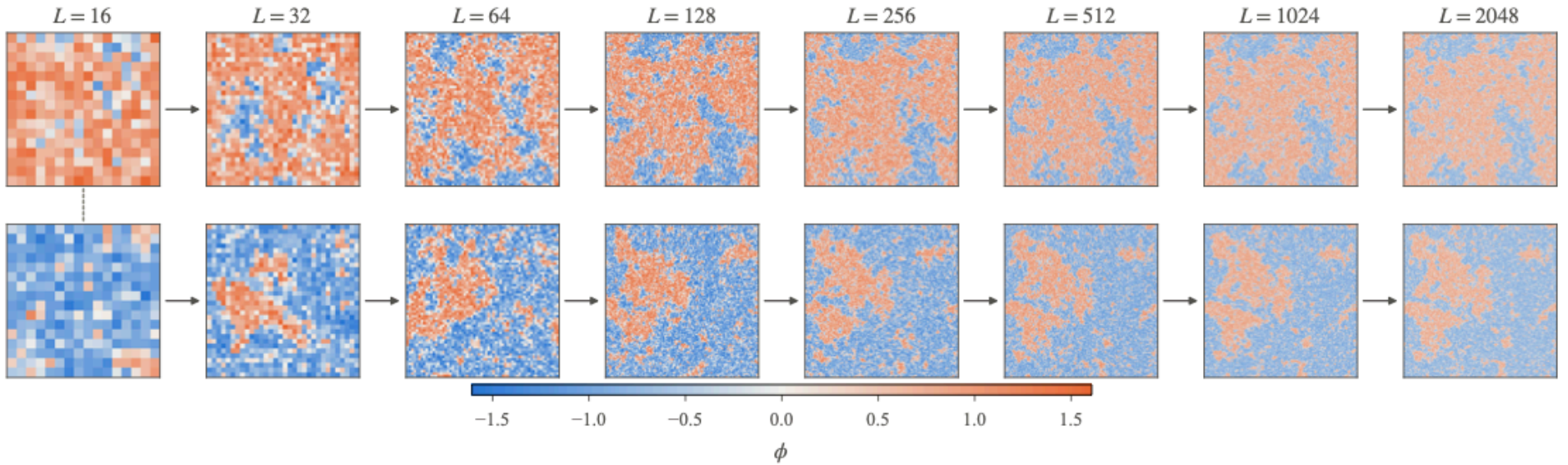
Multi-level

Test-bed: 2D-ϕ^4 model at criticality

- staring from L=16 to generate L=2048 lattice

Rethermalization repairs mismatch from upscaling

$\Longrightarrow$ use the rethermalized field as coarse input for the next step (only at κ_{cr})



- By learning/filling UV information but keep IR modes
- Next steps, gauge theories …

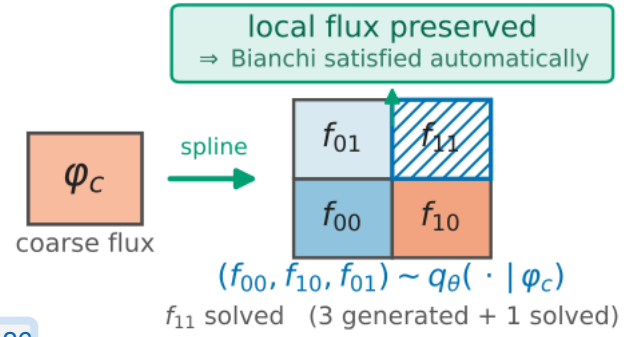
More on multi-level sampling

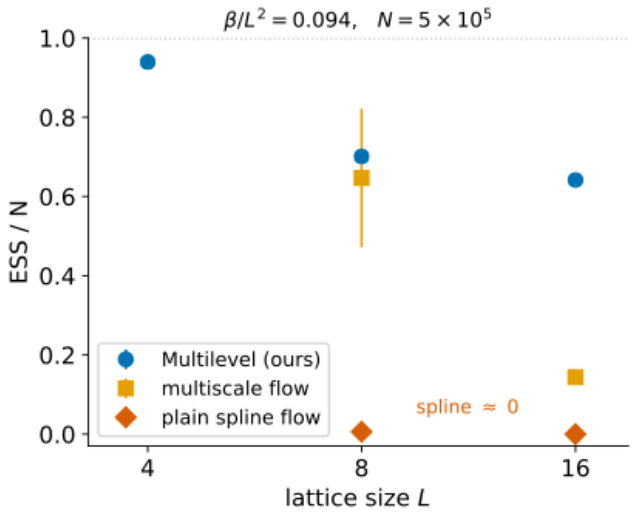
Application to gauge theories using multi-level procedures:

- Multiscale normalizing flow — Abbott et al. 2404.10819
  - Worked well for lower dimensions d=2,3 but became increasingly difficult for d=4, ESS ~ 0
- Another approach by A. Singha, S. Nakajima — See Talk by A. Singha at Lattice 26
  - sampling via plaquettes, keep Bianchi-identity and fill up missing plaquettes
  - recover gauge-links
  - Looks promising for U(1) d=2
  - application to higher dimension and compact groups on-going — Singha et al., 2609.35722

- still need to understand how this can keep up at higher dimensions like d=4 and inclusion of fermions

Singha et al., arXiv 2609.07781





Recap - Status

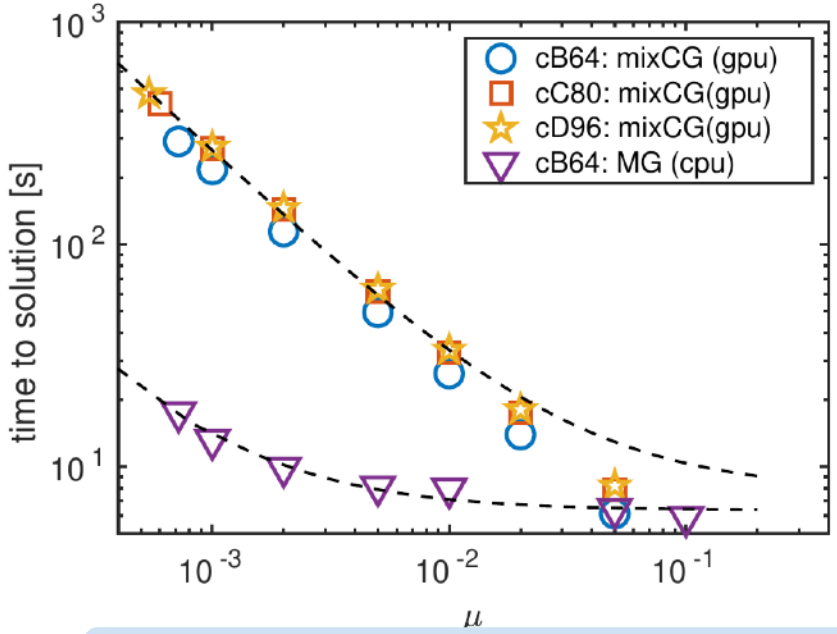
Multi-level procedure:
- still work in progress, but promising and scalable
  - keep IR modes by learning UV modes

Long range

- e.g. small eigenvalues, solution of Dirac equation,
- computational challenging and demanding towards physics quark masses

Challenging example:
- Solution of Dirac equation

Solve Dirac equation: $D \cdot x = b$
Notorious expensive



C. Alexandrou et. al., Phys. Rev. D 94 (2016), 11, 114509

Towards long distances

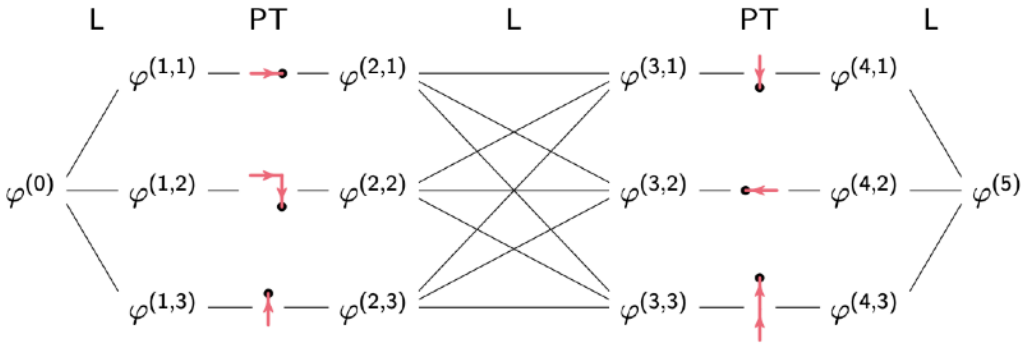
S. Pfahler, D. Knuettel, R. Constantini
C. Lehner, T. Wettig

See S. Pfahler talk at Uni Bonn

Pfahler et al., 2602.23840

Idea: Learn approximation to Dirac operator
- Application to solvers as preconditioner

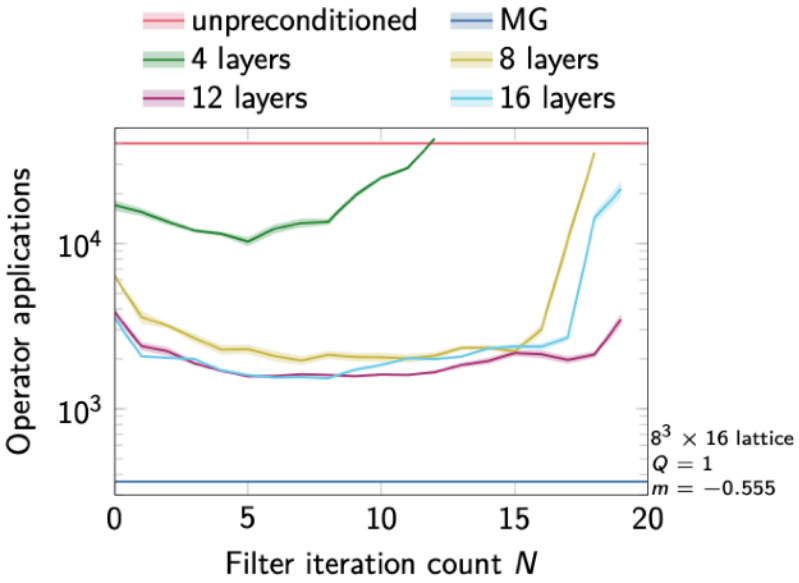
Use covariant network:
Dense linear layer: $\psi_i(x) \overset{L}{=} \sum_{j=1}^{n} W_{ij}\phi_j(x)$
Parallel Transport layer: $\psi_i(x) \overset{PT}{=} T_{P(i}\phi_i(x)$



Now: train network by minimising

$$C = |MD\phi - \phi|^2$$

- where: ϕ random or obtained through N tiers of GMREs



Similar work by M. Choi and H. Ohno

Solver

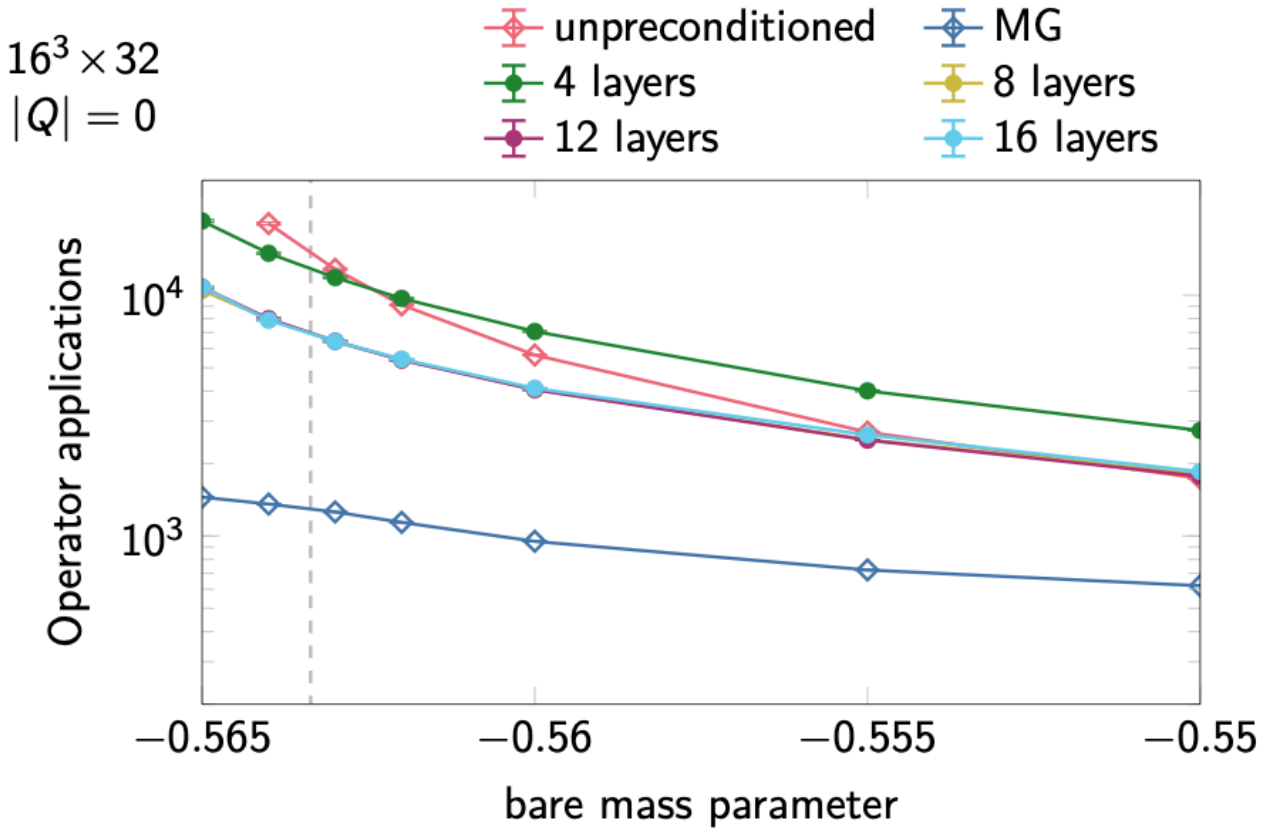


Kernels with higher range:
- Adding more pathes
- Applying more layers

Results:
On quenched Wilson lattice
- reduces iteration count
- Can be transferred to other configuration

Preconditioner might be sensitive to small eigenvalues

Pfahler et al., 2602.23840



Current status

In which case it works …

- flow models show some advantages for short distance physics
- Still have to improve in order to scale up training and require inclusion of fermions
- Possible combination with multi-level approach could help

There are many idea's and a lot of un-covered landscape

- **e.g. Diffusion models, larger interest and effort:**
  - Motivated by success on image generation
  - Learning score and diffusion might help to bridge UV to IR
    - Many studies, still at an earlier stage

See e.g, Bridging ML and Stat FT through generative and diffusion processes,,Trento 2026

- **Towards large scale, its natural to learn low-modes, like MG preconditioner**
  - This would enable to enter IR regime
  - Would have wider impact on large scale utilisation of ML techniques in Lattice QCD

Outlook

Does ML in LQFT require large scale computations ?
and if How ?

Utilization within workflows

- **Shorter range - applications**
  - learning generative functions
  - Learning perfect actions, .. Holland et al., 2504.15870

Possible future path, learning IR-region

  - will speed up our computations
  - will require larger computational resources

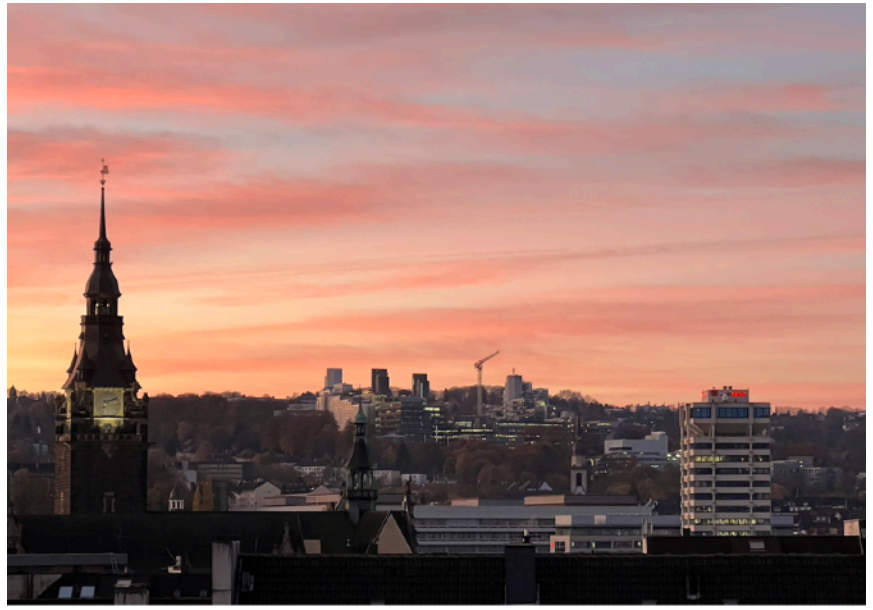
- **requires new ideas, new investigation, natural expensive but high-gain**

… towards Physics for AI

Conclusion

White paper on AI in Lattice QFT
- lead by G. Aarts and A. Nada
- Stayed tuned for CERN workshop beg. April 27

Follow up:
- Upcoming Workshop at Uni Wuppertal
- 15.2 - 19.2.2027
- Algorithmic Advancements in Lattice Field Theory at the Frontier of High-Energy Physics



Thank you

Funded by the European Union
European Research Council
Established by the European Commission



BERGISCHE
UNIVERSITÄT
WUPPERTAL



Funded by
the European Union



European Research Council
Established by the European Commission