Trivializing map as a coarse-graining map

Nobuyuki Matsumoto

RIKEN BNL Research Center
(→ BU from Sep)

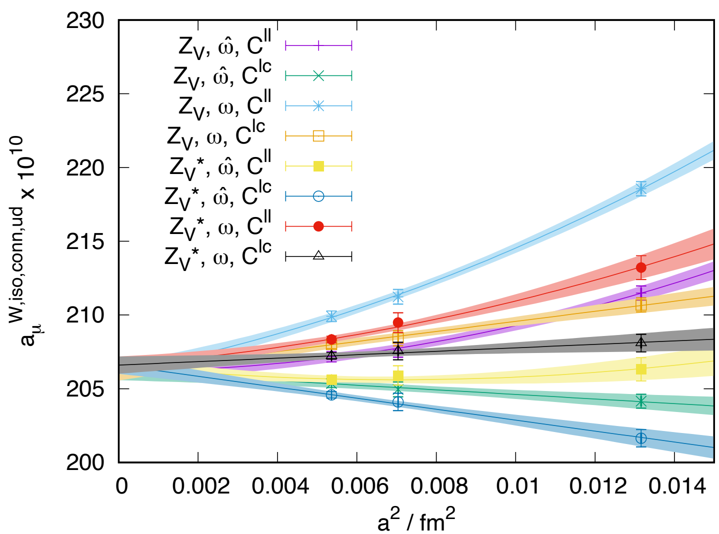
Machine Learning for Lattice Field Theory and Beyond
ECT*, 06.28.2023

Based on collaboration with
P. Boyle, R. Brower, T. Izubuchi, L. Jin, C. Jung

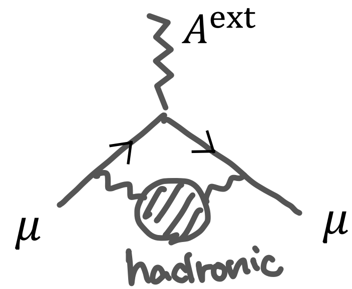
in preparation

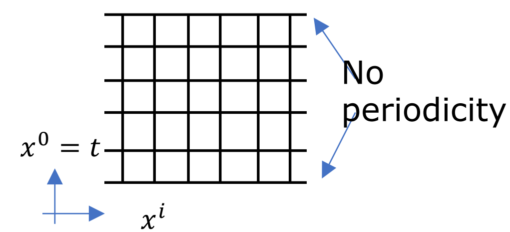
SPDR

Goal Overcome the critical slowing down and topological freezing to perform efficient simulations on fine lattices.



g-2 window update
RBC/UKQCD (incl NM) 2301.08696





- A partial solution may be to introduce the open boundary condition; **Lüscher-Schaefer 1105.4749**

  however, since many statistical techniques assume translational invariance, we want to keep the periodic boundary condition if possible.

- We here develop the field-transformative methods combined with HMC, which was originally proposed by the seminal work of Lüscher. **Lüscher 0907.5491**

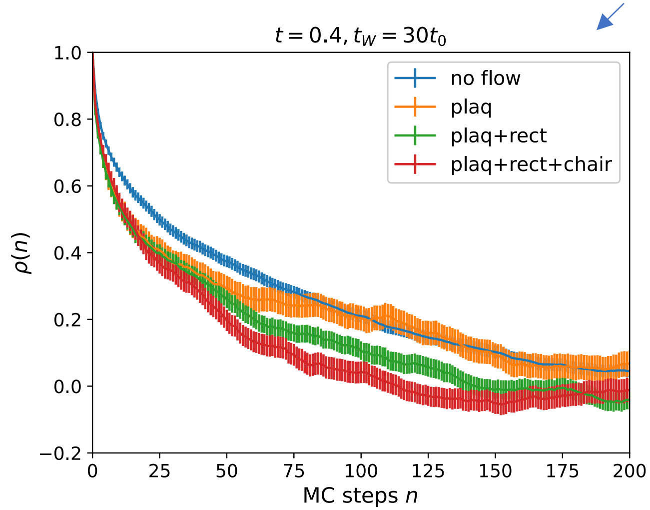
Short timeline

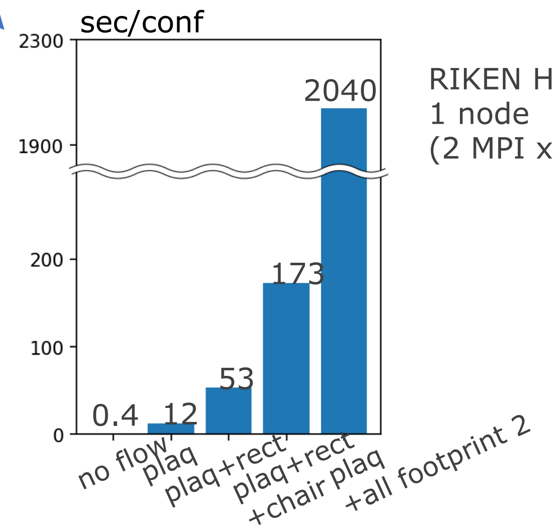
- Original proposal of the algorithm **Lüscher 0907.5491** (Name borrowed from **Nicolai 80**)
- LO approximation in CP^{N-1} model **Engel-Schaefer 1102.1852**

  "The reduction in the forces, …, is compensated by the computational overhead"
- Wilson flow and its generalizations using Schwinger-Dyson eq [4D SU(3) pure gauge]

  Jin LATTICE 2021 poster

  Boyle-Izubuchi-Jin-Jung-NM-Lehner-Tomiya LATTICE2022 [2212.11387]





RIKEN HOKUSAI
1 node
(2 MPI x 40 OpenMP)

Tunneling rate can increase, but still overwhelmed by the overhead.

- Normalizing flow/generative models

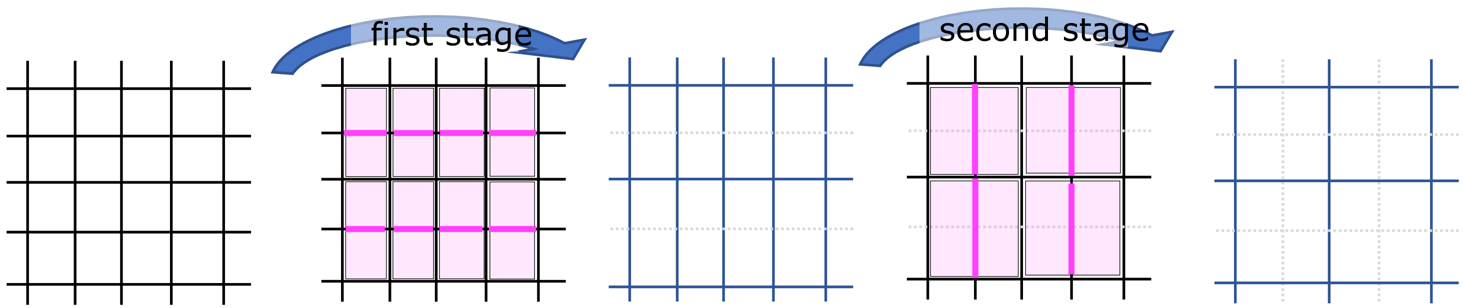
  Albergo-Kanwar-Shanahan 1904.12072,
  Bacchio-Kessel-Schaefer-Vaitl 2212.08469

E.Cellini; in effective string theory, Mon
P.Bialas; w/o action derivative, Tue
M.Gerges; φ^4 transfer learning, Tue
L.Vaitl; continuous flow in 4dYM, Tue
A.Nada; w/ stochasticity, Thu
J.Komijani; improve scaling, Fri
J.Urban; instantaneous generation, Fri

 NM+ in prep

- We develop a variant of the trivializing map, which we refer to as *decimation map.*

The trivialization is divided into several stages
and each stage corresponds to integrating out local degrees of freedom,
which we can interpret as coarse-graining transformations. **cf. Kadanoff 75, Migdal 76**



and repeat

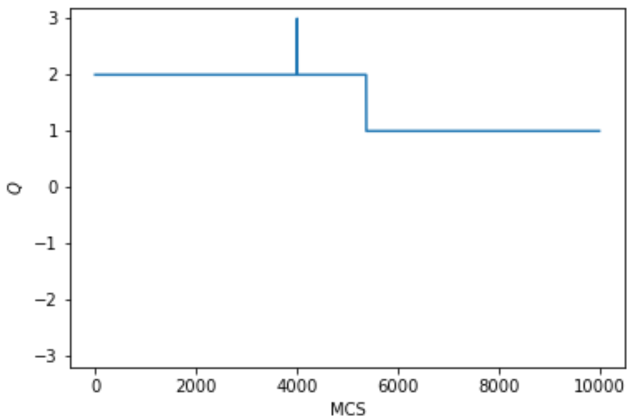
In fact, this map induces the effective action
that exactly corresponds to the one we obtain after integrating out the local DOF.

$$Z = \int \underbrace{(d\mathcal{T})}_{\text{to be trivialized}}\underbrace{(d\mathcal{R})}_{\text{reminder}}\, e^{-S(\mathcal{T},\mathcal{R})} \xrightarrow{\text{integrate } \mathcal{T}} \int (d\mathcal{R})\, e^{-S_{\text{eff}}(\mathcal{R})}$$

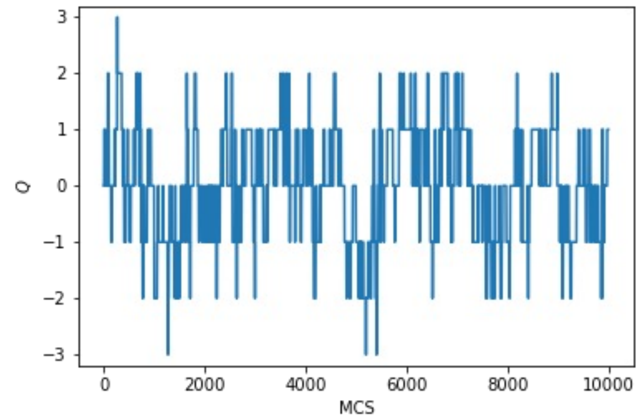
- To show this map can be useful, we apply it to the 2D U(1) model with the guided MC (a generalization of HMC)

<u>Time series of Q</u> ($\beta = 7.1$, 16x16, two-stage transformation)

w/o decimation map

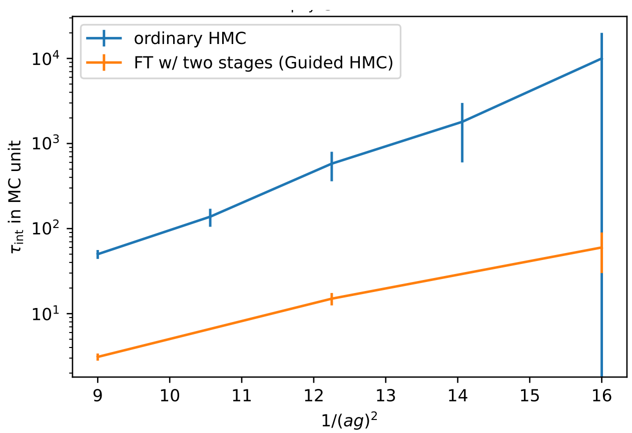


w/ decimation map

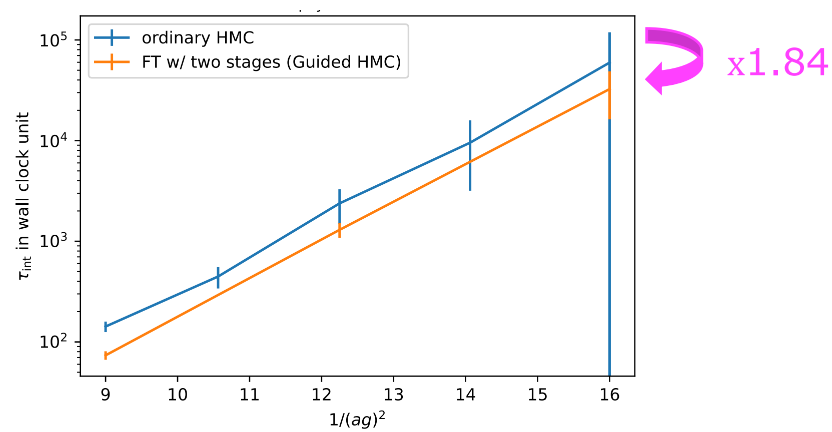


<u>$\tau_{\mathrm{int}}(Q)$</u> (fixed physical volume $V_{\mathrm{phys}}\, g^2 = 4 \times 4$)

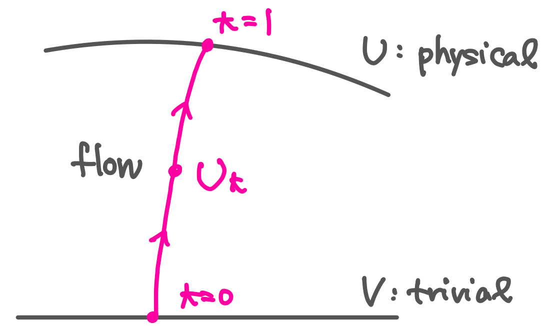
In Monte Carlo (MC) unit



In wall-clock time



- The algorithm is exact, scaling is known and controlled.



- Gradient flow ansatz:

$$\dot{U}_t U_t^{-1} = -\partial K_t \qquad K_t\text{: flow kernel}$$

$$\begin{cases} U_{t=0} \equiv V & \text{for simulation (to be trivial)} \\ U_{t=1} \equiv U & \text{for estimation (physical)} \end{cases}$$

- Require

$$S_{\mathrm{eff},\,t}(V) \stackrel{\Delta}{=} S(U_t) - \ln\det J_t \qquad \left(J_t \equiv \frac{\partial U_t}{\partial V} \right)$$

$$\stackrel{*}{=} (1-t)\, S\big(\mathcal{F}_t(V)\big) \quad \text{requirement}$$

$$\left(\int (dU) e^{-S(U)} = \int (dV) \det J_t \, e^{-S(U_t(V))} \equiv \int (dV) e^{-S_{\mathrm{eff}}(V)} \right)$$

$\Downarrow$ d/dt

Equation for the kernel function K_t:

$$-\partial^2 K_t - t\,\partial S \cdot \partial K_t \stackrel{*}{=} -S$$

(up to const; hereafter as well)

merely a linear differential equation!

- From the properties of the differential operator:

$$\mathcal{L}_t \equiv -\partial^2 - t\,\partial S \cdot \partial \qquad \left(\mathcal{L}_t K_t \stackrel{*}{=} -S \right)$$

the existence of the inverse is proven. **Lüscher 0907.5491**

As for the Fokker-Planck Hamiltonian,
it can be symmetrized to an elliptic nonnegative differential operator.

cf. in RG flow of ϕ^4:
Abe-Fukuma 1805.12094

The spectrum is discrete for compact space, and thus the inverse exists after removing zero-modes.

 Lüscher 0907.5491

- Formally expand in power series:

$$K_t = \sum_{m\geq 0} t^m K^{(m)}$$

Plug into

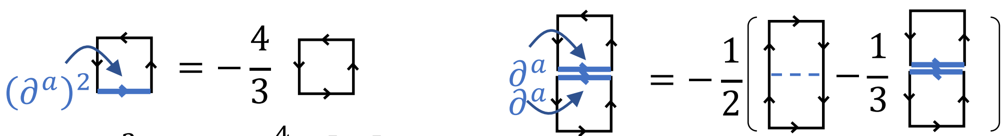
$$-\partial^2 K_t - t\,\partial S \cdot \partial K_t = -S$$

order by order matching

$$\begin{cases} \mathcal{L}_0 K^{(0)} = -S, \\ \mathcal{L}_0 K^{(m)} = \partial S \cdot \partial K^{(m-1)} \; (m \geq 1) \end{cases}$$

NB $\partial^a \sim T^a$

See A.Ipp on visualization; Tue



$$\mathrm{tr}[(T^a)^2 \ldots] = -\tfrac{4}{3}\mathrm{tr}[\ldots] \qquad \mathrm{tr}[T^a A]\,\mathrm{tr}[T^a B] = -\tfrac{1}{2}\left(\mathrm{tr}[AB] - \tfrac{1}{3}\mathrm{tr}A \cdot \mathrm{tr}B\right)$$

- Solution for the Wilson action [4D SU(3)]:

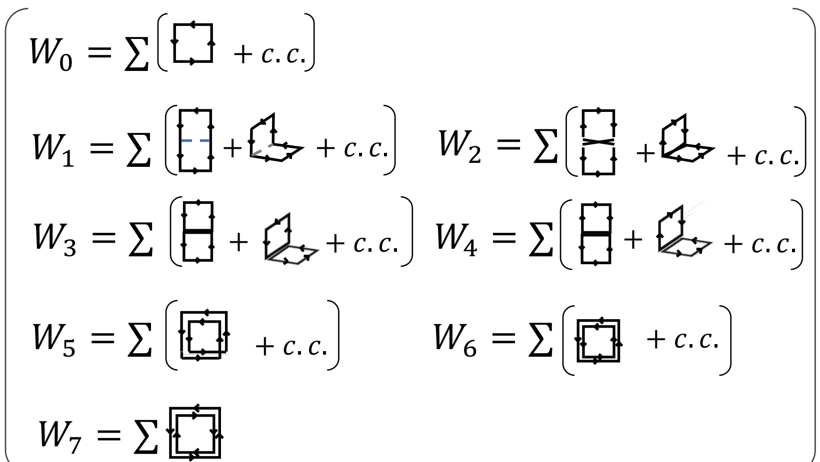
$$K_t = -\frac{\beta}{32} W_0$$

LO: plaquette

$$+t\frac{\beta^2}{192}\left(-\frac{4}{33}W_1 + \frac{12}{119}W_2 + \frac{1}{33}W_3 - \frac{5}{119}W_4 + \frac{3}{10}W_5 - \frac{1}{5}W_6 + \frac{1}{9}W_7\right)$$

NLO: rectangle, chair, twisted rectangle …
"footprint 2 shapes"

$$+O(t^2)$$



- t-expansion can be seen as the power series expansion:

$$K_t = \frac{1}{-\partial^2 - t\partial S \cdot \partial}(-S) = \frac{1}{1 - tM}\cdot(-S) = (1 + tM + t^2M^2 + \cdots)(-S)$$

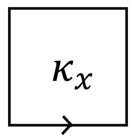
$$M \equiv -\partial^{-2} \circ \partial S \cdot \partial$$

As referred to in **Lüscher 0907.5491,**
the radius of convergence is finite: $t\,\|M\|_2 < 1.0$

($\|M\|_2$: largest eigenvalue of M in magnitude $\sim O(\beta)$)

How good is the t-expansion?

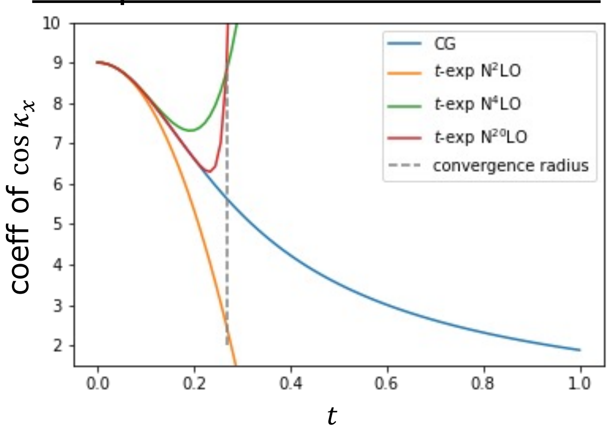
- For the sake of demonstration:



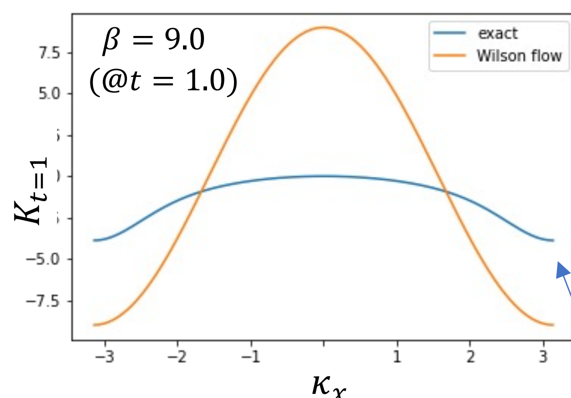
U(1) 1-plaquette system

- Parameterize the function space, solve the linear equation exactly: $K_t = -(\mathcal{L}_t)^{-1}S$

Comparison of the coefficients



Functional form of the kernel



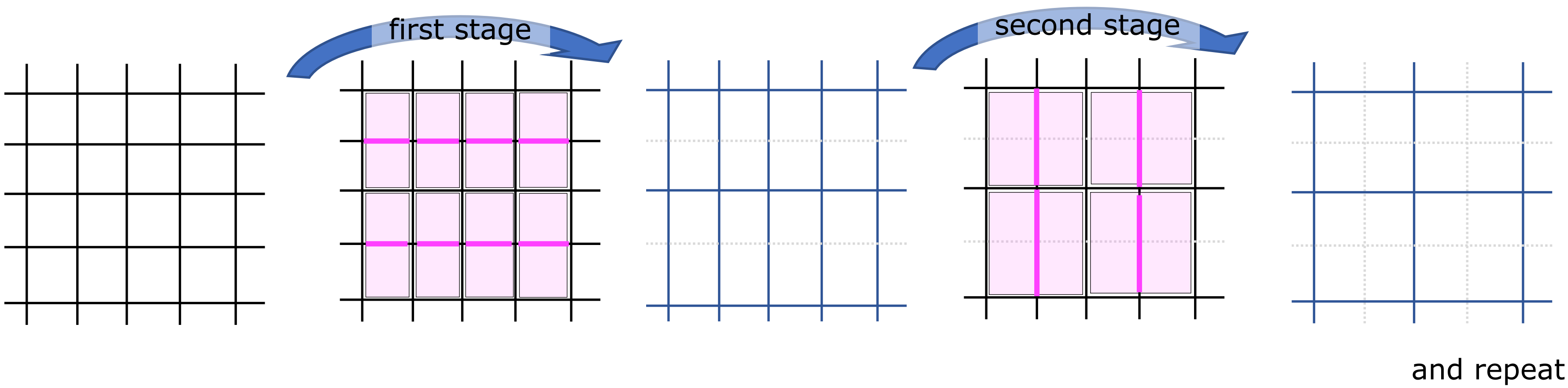
milder angular dependence from the higher irreps

- Radius of convergence: $\beta t < 2.40$ (quite coarse)

Motivated us to reconsider the strategy.

NM+ in prep

Idea trivialize links step by step; *"decimation map"*



- The formulation of Lüscher is applied to each block;
  we construct a differential equation for the kernel K_t:

$$\mathcal{L}_t K_t = -S^{(\mathrm{loc})}$$

and solve it with conjugate gradient (CG).

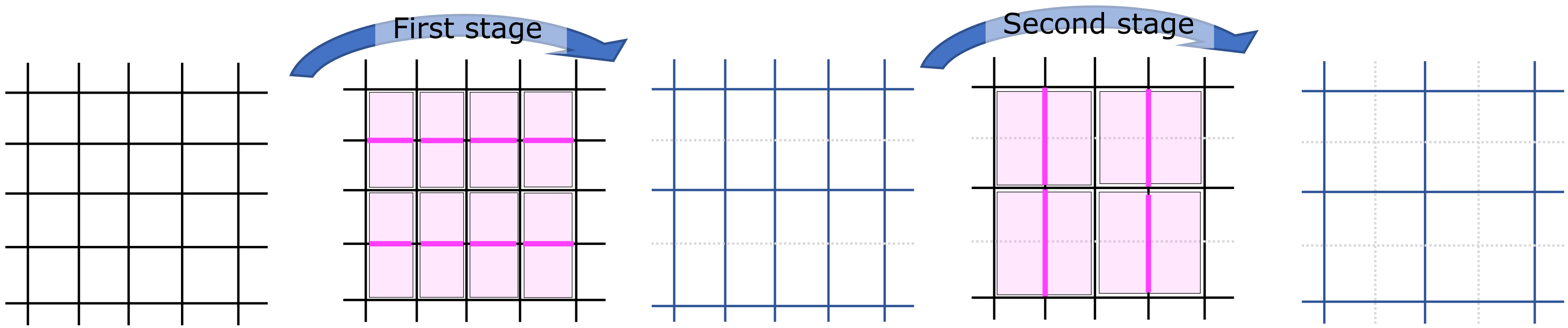
Existence of the inverse follows from basically the same argument.

- The function space is small enough to keep track of.

- Net cost increases only by addition as we increase the stages and thus the trivialized region.

Decimation map (2/4)

NM+ in prep



- When these maps are obtained exactly, the effective action for the remaining links agree with the one we obtain by decimation:

Dividing the system into $\mathcal{T}$ and $\mathcal{R}$,

to be trivialized remainder

local trivialization
(change of variables)

$$Z = \int (d\mathcal{T})(d\mathcal{R})\, e^{-S(\mathcal{T},\mathcal{R})} = \int (d\mathcal{T})(d\mathcal{R})\, e^{-S_{\mathrm{eff}}(\cancel{\mathcal{T}},\mathcal{R})}$$

integrate integrate

$$\int (d\mathcal{R})\, e^{-S_{\mathrm{eff}}(\mathcal{R})} \overset{!}{=} \int (d\mathcal{R})\, e^{-S_{\mathrm{eff}}(\mathcal{R})}$$

must be the same!

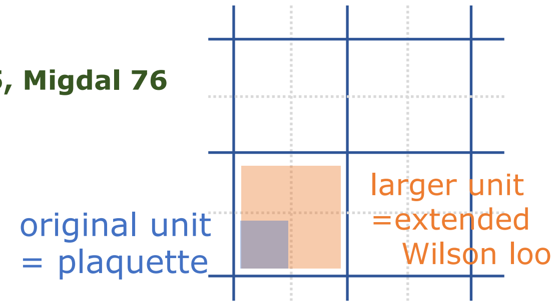
- The effective action can be obtained as a bi-product of the CG via the zero-modes.

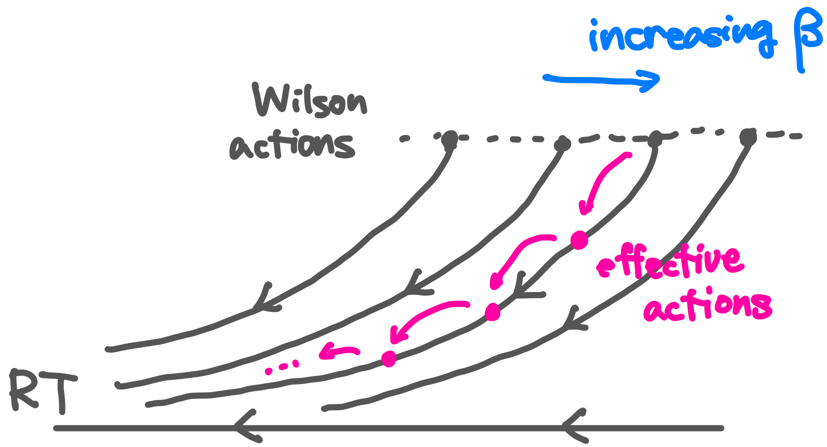
NM+ in prep

- Decimation map leaves the physics in the large units exactly unchanged.
  - Can be seen as a renormalization group transformation.

Wilson-Kogut 74, Kadanoff 75, Migdal 76

M. Ozawa after coffee break



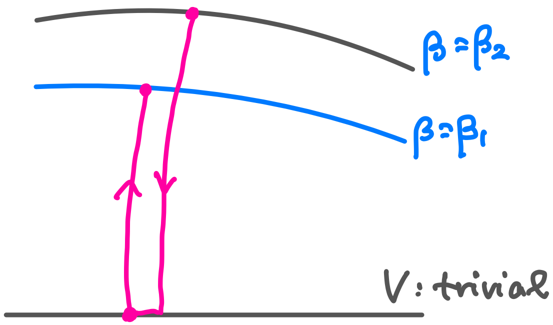


Fixed-pt/Perfect action:
P.Hasenfratz-Niedermayer 93,
DeGrand-A.Hasenfratz-P.Hasenfratz -Niedermayer-Wiese 95,
Niedermayer-Rüfenacht-Wenger 01,
P.Hasenfratz-Hauswirth-Jorg-Niedermayer-Holland 02

U. Wenger, K. Holland next talks

- Using (the inverse of) the decimation map,
  we *integrate in* the decimated links from the coarse-grained system.

- Thus the acceleration happens for the two-fold reasons:
  - Increase of the trivialized region
  - Coarse action for the remaining variables



Comment on Lüscher's remark

- Lüscher considered β-changing maps
  by successively applying trivializing maps and their inverses.

(f) *Renormalization group.* By composing the trivializing map $U = \mathcal{F}_1(V)$ in the Wilson theory with its inverse at another value of the gauge coupling, one obtains a group of transformations whose only effect on the action is a shift of the coupling. The locality properties of these transformations are not transparent, however, and could be quite different from the ones of a Wilsonian "block spin" transformation.

Lüscher 0907.5491; Commun.Math.Phys.293:899-919,2010

- Such maps connect lattice ensembles of the same *lattice* volume and of the same lattice action, and is different from the decimation map.

Simplest lattice gauge system with topology: 2D U(1) pure gauge

Wilson action

$$S(U) \equiv -\beta \sum_x \mathrm{Re}\left(U_{x,0} U_{x+0,1} U^\dagger_{x+1,0} U^\dagger_{x,1}\right)$$
$$= -\beta \sum_x \cos \kappa_x$$

$\kappa_x \equiv \frac{1}{i}\log\left(U_{x,0} U_{x+0,1} U^\dagger_{x+1,0} U^\dagger_{x,1}\right)$: plaquette angle

$-\pi \leq \kappa_x < \pi$

$\beta = \dfrac{1}{(ag)^2}$ Note that $[g] = L^{-1}$

Exact formulas can be easily obtained. **cf. Brower-Nauenberg 80, Rusakov 90**

Characteristic features

- Topological charge: $Q = \frac{-1}{2\pi}\sum_x \kappa_x \in \mathbb{Z}$
- Energy density: UV divergent $\langle e \rangle \equiv \dfrac{1}{2V}\int d^2x\, F_{01}^2 \sim \dfrac{g^2}{2a^2}$.
- Zero correlation length; theory is *ultralocal.*
- topological susceptibility is finite: $\chi_Q \sim \dfrac{g^2}{(2\pi)^2}$.

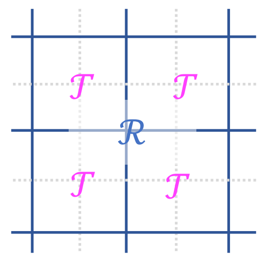
Deal with the topological freezing rather than the entire critical slowing down.

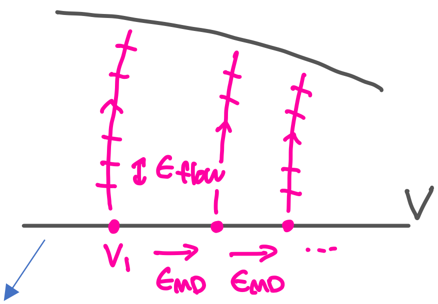
Update algorithm (1/1)

- To show that the map can be useful,
  we use the guided-Hamiltonian Monte Carlo, which is a variant of HMC:
  Horowitz 91 **Duane et al. 87**

  - We replace the action $S_{\mathrm{eff}}(V)$ in H by the approximate one calculated from CG.
  - Detailed balance holds
    thanks to the *Liouville theorem* and the *reversibility* of the Hamiltonian evolution.

Pros
- Simplifies the calculation of force
- Transparent separation of UV/IR
  $\mathcal{T}$ $\mathcal{R}$

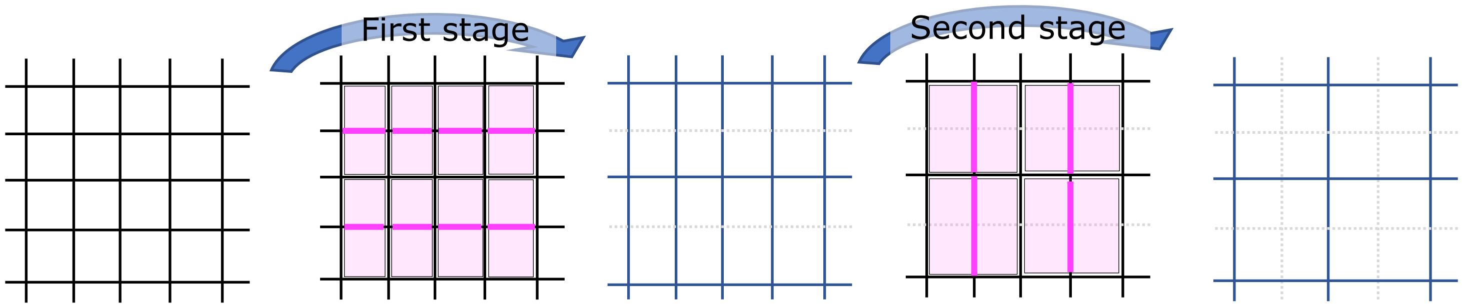




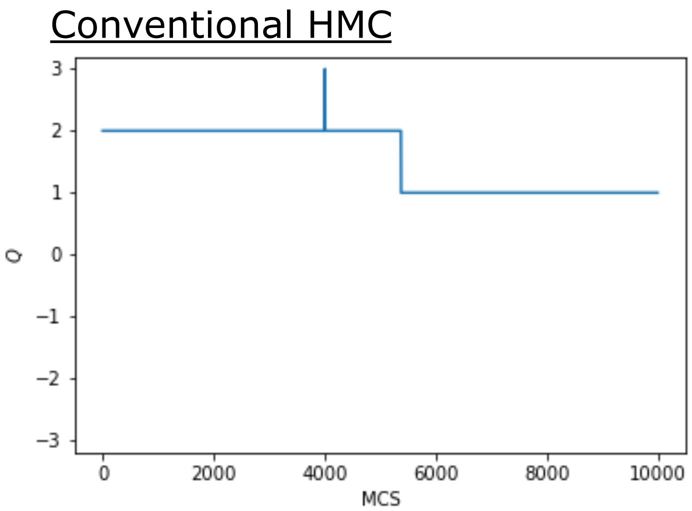
Cons
- Energy conservation is only approximate;
  acceptance rate will be controlled both by MD step size and the flow step size,
  yet they are under control and it scales well.

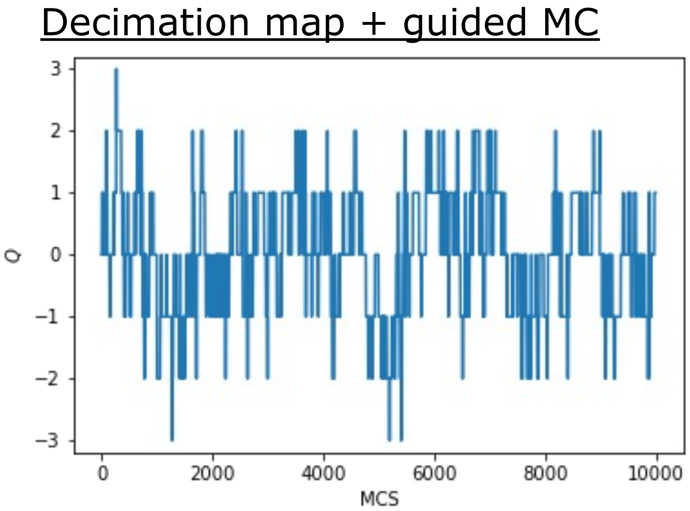
- No gauge fixing, periodic boundary condition.
- Flow equation is solved with the midpoint integrator ($\therefore$ net discretization error $= O(\epsilon^2)$)
- The map is ensured to be bijective by keeping the flow step size within a bound.
  cf. Lüscher 0907.5491
- The flow step size is adaptively changed by referring to the bijection bound.

NM+ in prep



- Time series of Q ($\beta = 7.1$, 16x16)





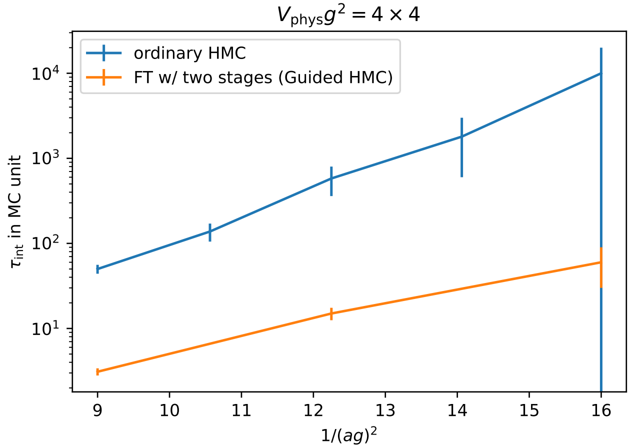
- Whole CG solves take a few seconds to a few minutes using GPU.
  Iteration count @β=7.1: 1st stage≲140, 2nd stage≲1300

 NM+ in prep

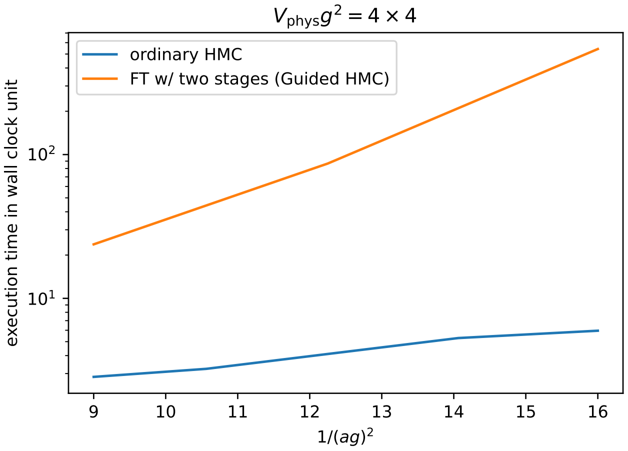
We check the cost scaling towards the continuum limit; physical volume fixed to $4{\times}4/g^2$:

Flow step size is scaled to keep the acceptance rate ~0.8

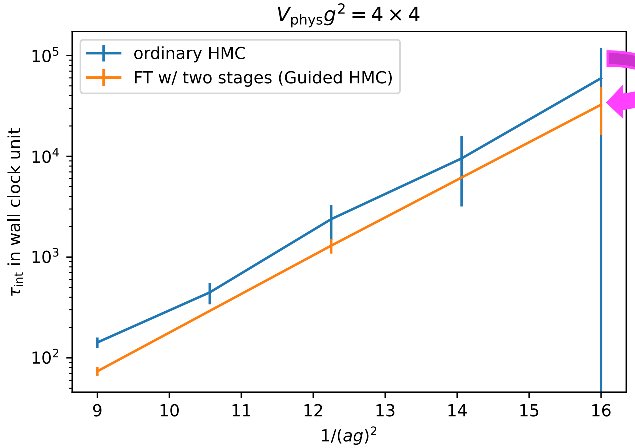
$\tau_{\rm int}(Q)$ in MC unit



Execution time/conf (no parallelization)



$\tau_{\rm int}(Q)$ in wall-clock



Net gain~ x1.84

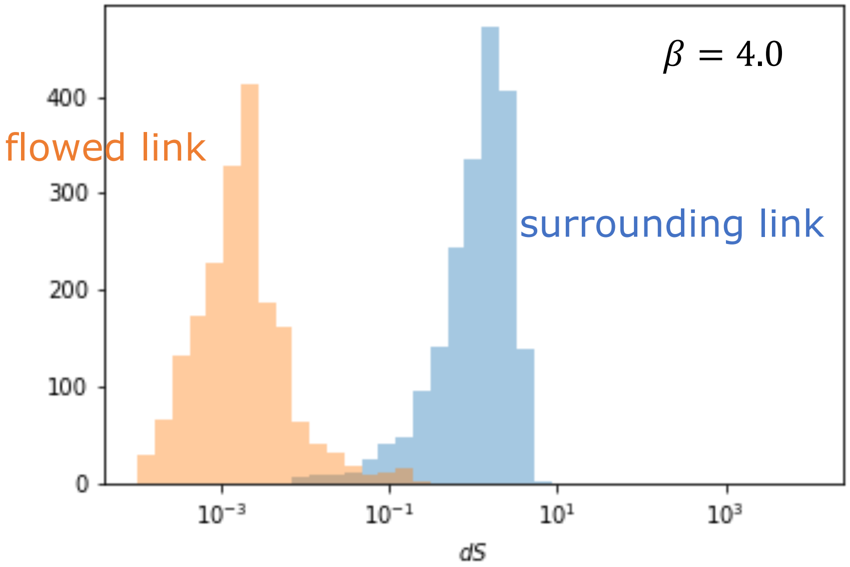
Likely survives in the continuum limit

- We can expect larger gains with more stages.
- With the guided MC, we need to keep the flow step size small to keep the acceptance.

This complication can be circumvented by either
increasing the order of the flow integrator (currently midpoint; $O(\epsilon^2)$)
or switching to *FTHMC,* the HMC with the exact effective action;
the latter will be important anyhow with the fermion.

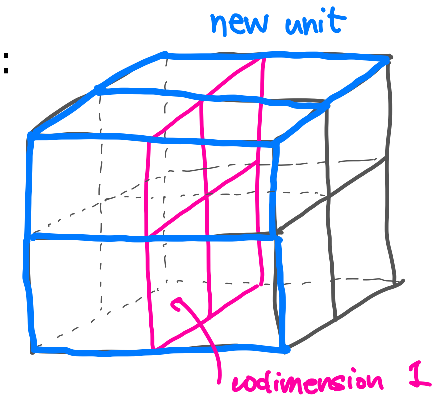
NM+ work in progress
cf. Lüscher 0907.5491

Force hierarchy (1st stage)



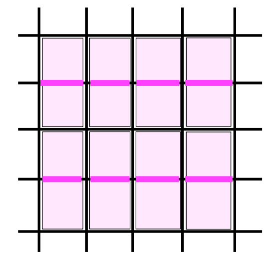
Remarks (2/2)

- The higher-dimensional counterparts of the decimation map can exist:

  - One can define the trivializing map for codimension one surfaces.

    Then thanks to gauge invariance,
    the new minimum unit in the coarse-grained direction
    will be the 1x2 Wilson loops.

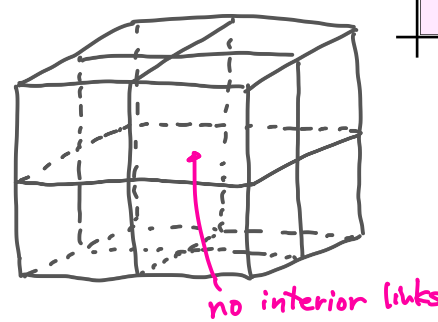
    

    However, it is the special feature of 2D Wilson action
    that the links on the codimension one surface(=line) are independent;

    in higher dimensions, larger function space needs to be prepared.

  - Another possibility may be to introduce
    frozen links surrounding a local volume
    and trivialize the links lying in its interior;

    the resulting effective action will be a "cage action",
    which still may be useful to generate dislocations.

- Representing $\mathcal{L}_t$ for the $SU(N_c)$ cases will be more complicated (though I believe it is doable).

  Note that the basis of the function space can be taken linearly dependent in the CG.

- At least for U(1) and SU(2),
  it is possible to reduce the local trivialization problem to an ODE and avoid using the flow.

  Algorithms keeping exactness are under discussion. **Brower-Taku-NM**
  cf. J.Urban's talk on Fri

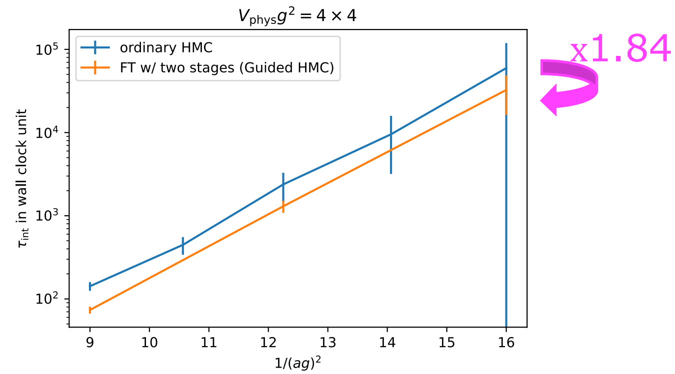
- We considered the use of trivializing maps in such a way that multiple stages of transformation can be regarded as coarse-graining transformations.

"decimation map"

This map allows us to use the effective action we obtain after integrating out the local DOF.

- We showed the results after the two-stage transformation, which reduces $\tau_{\rm int}(Q)$ by a factor of 1.8 in wall clock time.

More gain is expected for the larger number of stages.



- It is true that the current investigation uses special features of 2D and U(1);

however, we believe that having a method that works on this simplest model and has possible generalization directions will be a good starting point for developing algorithms towards QCD.

Active studies around BNL

- gradient flowed HMC w/ fermion in Grid **P. Boyle**
- Machine learning FTHMC **X.Y.Jin 2201.01862**
- L2HMC **Foreman-X.Y.Jin-Osborn 2201.01582**
- Fourier acceleration/Riemannian manifold MC **Sheta-Zhao-Christ 2108.05486, Nguyen-Boyle-Christ-Jang-Jung 2112.04556**
  Parisi 84, Batrouni et al. 85,88,90

Thank you.