Towards conquering critical slowing down

Nobuyuki Matsumoto
(RIKEN BNL Research Center)

RIKEN BNL Research Center

HU Berlin/DESY Zeuthen 09.12.2023
(online)

work in collaboration with
P. Boyle, T. Izubuchi, L. Jin, C. Jung, C. Lehner, and A. Tomiya
BNL/Edinburgh BNL/RBRC UConn BNL Regensburg IPUT Osaka

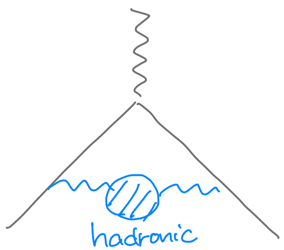
PoS(LATTICE2022)229 [arXiv:2212.11387] (work in progress)

- Lattice calculation is giving important inputs in the precision test of the standard model.

E.g.: Contribution of
hadronic vacuum polarization to muon g-2

$$a_\mu^{HLO} = \int_0^\infty dt\, w_t\, C(t)$$

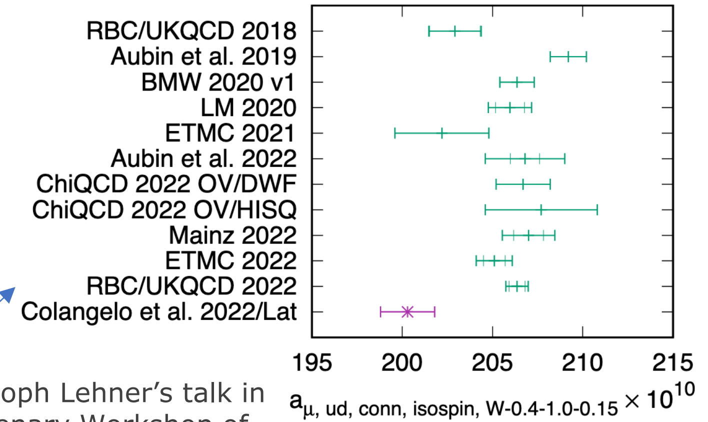
Bernecker-Meyer 2011



$$C(t) \equiv \frac{1}{3} \sum_{\mathbf{x},k} \langle J_k(\mathbf{x},t) J_k(0) \rangle$$

J_μ: vector current

w_t encodes the information of the momentum loop



From Christoph Lehner's talk in The Fifth Plenary Workshop of the Muon g-2 Theory Initiative

RBC/UKQCD (incl NM) in prep

One of the major sources of systematical error is the continuum extrapolation (evaluation of the $a \to 0$ limit).

- However, as we reach the continuum limit, we encounter the infamous *critical slowing down* when generating configurations, which adds extreme computational cost to the simple volume scaling.

Algorithms aimed for accelerating Monte Carlo (MC) sampling:

- Overrelaxation
  Adler 81, Whitmer 84, Creutz 87
- Multigrid MC
  Parisi 84, Goodman-Sokal 86 (see also Wolff 90)
- Fourier acceleration/Riemannian manifold MC
  Parisi 84, Batrouni et al. 85,88,90 / Nguyen et al. 2112.04556
- Parallel tempering
  Swendsen-Wang 86, Geyer 91, Hukushima-Nemoto 96

  with defects:
  Hasenbusch 1706.04443, Berni-Bonanno-D'Elia 1911.03384, Bonanno-Bonati-D'Elia 2012.14000
- Cluster algorithm
  Swendsen-Wang 87, Wolff 89
- Trivializing map/normalizing flow
  Luscher 0907.5491 / Rezende-Mohamed 15

  stochastic:
  Wu-Kohler-Noe 20, Caselle-Cellini-Nada-Panero 2201.08862
- L2HMC, winding HMC, …
  Foreman-X.Y.Jin-Osborn 2105.03418, Albandea, et al. 2106.14234, …

Short timeline on trivializing map

- Original proposal **Lüscher 0907.5491**
- Test in CP^{N-1} model ⇨ acceleration rather negative **Engel-Schaefer 1102.1852**
- Machine learning approaches **Albergo-Kanwar-Shanahan 1904.12072, Foreman et al. 2112.01586 Bacchio-Kessel-Schaefer-Vaitl 2212.08469**

This work

- We attempt to improve the flow kernel $\tilde{S}_t$ (=generating function) of the map using a *Schwinger-Dyson* (SD) *equation*.
  Gonzalez-Arroyo, Okawa 87, de Forcrand et al. hep-lat/9806008

  $$U \to e^{-\epsilon\, \partial \tilde{S}_t(U)} U$$

- We perform the HMC using the resulting effective action in the MD Hamiltonian.
  cf. L Jin LATTICE 2021

Advantages of this method

- basis functions for the flow kernel can be chosen by hand
- can be applied to general actions of interest without analytic calculation
- the coefficients in the kernel are determined by lattice estimates of the observables

- We apply our method to Wilson and DBW2 actions and show that:
  - With the SD method, we can have better control of the effective action than the known (perturbative-type) t-expansion.
  - In particular cases, faster decorrelation (in MC step unit) is observed for long-ranged observables by adding rectangle and chair to the flow.
  - However, we have large algorithmic overhead, and need to check the scaling with larger statistics to confirm the actual benefits at large β.

We here report preliminary results in this direction.

Table of contents

Table of contents

Critical slowing down (1/2)

Goal Make physical predictions from the lattice path integral:

$$\langle \mathcal{O} \rangle \equiv \frac{\int (dU)\, e^{-S(U)} \mathcal{O}(U)}{\int (dU)\, e^{-S(U)}}$$

$U_{x,\mu}$: link variable

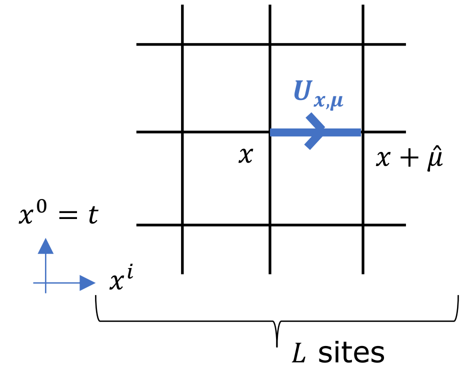
(dU): Haar measure

$\mathcal{O}$: observable

S: action

e.g., Wilson action **Wilson 74**

$$S(U) \equiv -\frac{\beta}{6} \sum_{x,\mu<\nu} \mathrm{Re\,tr}\left[U_{x,\mu} U_{x+\mu,\nu} U^{\dagger}_{x+\nu,\mu} U^{\dagger}_{x,\nu}\right]$$



- We give input values in physical units (e.g., in GeV) for the scales that will be dynamically generated in the system (e.g., correlation length).

  ➡ enables us to introduce lattice spacing in physical units (e.g., $a = 0.1\ \mathrm{fm}$) for a given β.

- Fixing the physical lattice volume (e.g., $La = 5\ \mathrm{fm}$), we take the $a \to 0$ limit by tuning β towards $\beta \to \infty$.

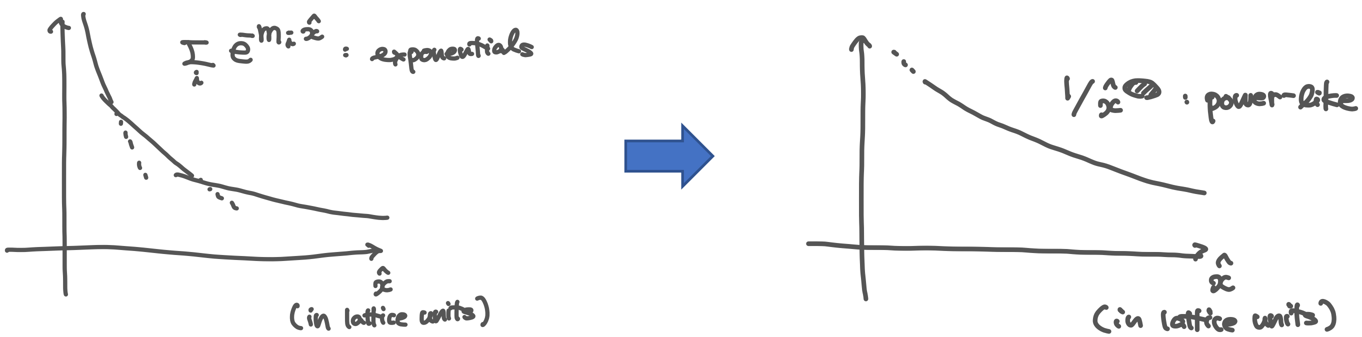
  ➡ Make predictions about the continuum theory.

- We expect to have a finite correlation length *in physical units* in the continuum.

infinite correlation length *in lattice units* (since $a \to 0$),
which is a property of 2nd order phase transition
regarding the lattice system as a statistical system.

Wilson 74

- Generically, as we approach the critical point,
more and more modes contribute to the correlator to give the quasi-long-range correlation.

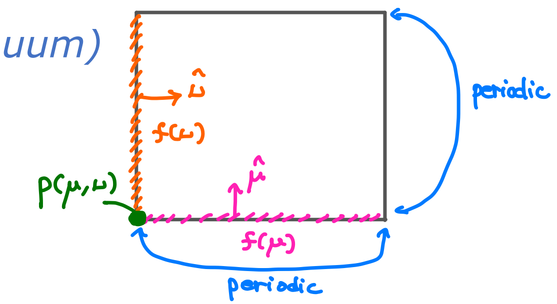


Such long correlation makes the Monte Carlo simulation inefficient. *critical slowing down*

(We expect fermions add another nonlocal structure in the theory;
however, in the following we basically concentrate on the gauge DOF.)

Further complication in QCD: *topological freezing*

Cause: nontrivial topological sectors of gauge field on T^4 *(in the continuum)*



- Gauge field A_μ is periodic *up to gauge transformation*:   **'t Hooft 81, cf. Dirac monopole**

$$A_\mu(x_\nu = L) = v_\nu(x)\left(\partial_\mu + A_\mu(x_\nu = 0)\right) v_\nu^{-1}(x).$$

- The gauge function (or *transition function*) $v_\mu(x)$ completely encodes the topological information of the gauge field: **Luscher 82, van Baal 82, Phillips-Stone 86, see also Kronfeld 88**

$$Q \equiv \frac{-1}{16\pi}\int d^4x \,\mathrm{tr}\, F_{\mu\nu}\tilde{F}_{\mu\nu}$$

$$= \frac{1}{24\pi^2}\left\{ \begin{array}{l} \sum_\mu \int_{f(\mu)} \mathrm{tr}\left(v_\mu d v_\mu^{-1}\right)^3 \\ -3\sum_{\mu\neq\nu}\int_{p(\mu,\nu)} \mathrm{tr}\,[d v_\nu^{-1}(x_\mu = L) v_\nu(x_\mu = L) v_\mu(x_\nu = 0) d v_\mu^{-1}(x_\nu = 0)] \end{array} \right\}$$

Solely expressed with $v_\mu(x)$!

- One can show that $Q \in \mathbb{Z}$ by, e.g., taking the pure gauge:

$$A_\mu dx^\mu = g^{-1}dg \qquad \left(\text{constraint: } g^{-1}(x_\mu = L) g(x_\mu = 0) = v_\mu\right) \Rightarrow Q \equiv \frac{1}{24\pi^2}\int_{\partial V} \mathrm{tr}\,(g^{-1}dg)^3 \in \mathbb{Z}$$

  Nontrivial $v_\mu(x)$ can give nontrivial Q.

- Topological sectors are disconnected $\because$ they have $v_\mu(x)$ that cannot be continuously deformed to one another.

As the continuum limit is reached, the lattice gauge field acquires continuum-like nature. Correspondingly, configurations will be trapped in the emerging disconnected sectors during Monte Carlo simulation (topological freezing).

More mathematical way to see the freezing is through the *geometrical definition of the lattice topological charge*:   **Luscher 82**

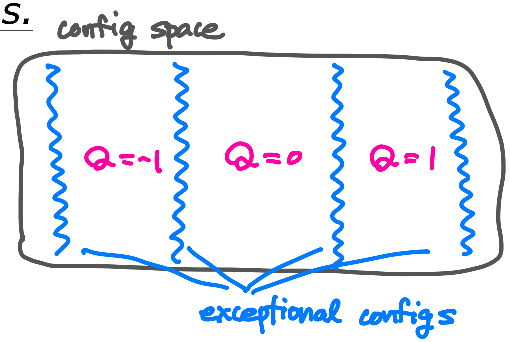
Simpler example: $U(1)$ on T^2 **Phillips 85, see also Fujiwara et al. hep-lat/0001029**

- Lattice topological charge: winding in the *plaquette angles* κ_x*:*

$$Q^{(\mathrm{lat})} = \frac{-1}{2\pi}\sum_x \kappa_x \qquad \left(\kappa_x \equiv \frac{1}{i}\log\left(U_{x,0}U_{x+0,1}U^{\dagger}_{x+1,0}U^{\dagger}_{x,1}\right), \quad \text{take a single branch s.t. } \log 1 = 0. \right)$$

Q is defined unambiguously except for the *exceptional configurations.*

for which $\exists x,\ \kappa_x = \pi$
($\therefore$ measure zero in path integral).

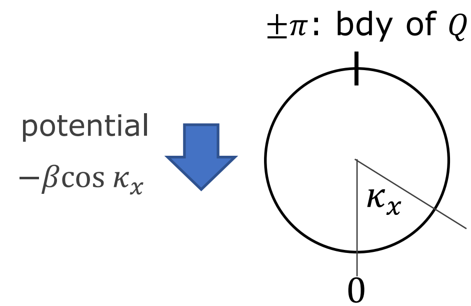


➡ Boundary of Q sectors are the exceptional configurations.

- Tunneling only occurs when the fluctuation becomes so large that the plaquette angle goes around the S^1 penetrating the potential barrier at $\pm\pi$.

However, such large fluctuation will be directly suppressed for the Wilson action at large β:

$$S(U) = -\beta\sum_x \cos\kappa_x\,.$$



➡ *Emergence of disconnected topological sectors.*

Similarly for SU(2) on T^4, exceptional configurations (= boundary of Q) consists of $\exists(\text{local Wilson loop}) = -1$, which will be suppressed at large β. **Luscher 82**

(Except for these configurations, one can define Q via the transition functions $v_\mu(x)$.)

A detour for the topological freezing: *open boundary condition* **Luscher-Schaefer 1105.4749**

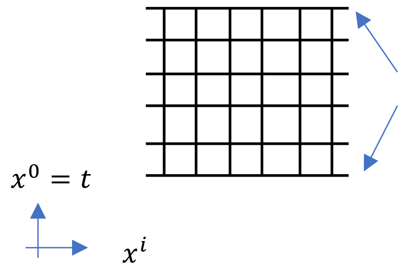
Pros

No more topological sectors in the continuum!

Cons

Need to consider the boundary effects.
In particular, translational invariance will be violated.

we want to avoid this if possible
∵ many statistical techniques assume the translational invariance



Regarding both the critical slowing down and the topological freezing,
they are rather intrinsic to the lattice simulation near the continuum (at large β).

➡ it will be advantageous if one can use small β simulation to generate large β configurations.

➡ *trivializing map!*

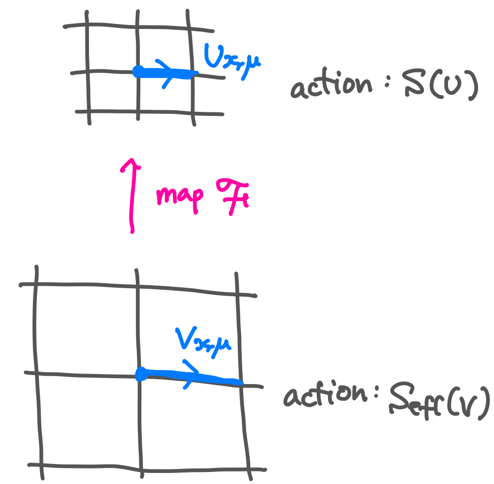
Table of contents

Idea **Luscher 0907.5491**

- With a field transformation (or a change of the integration variable), we can generate a new action for the transformed variable:

For $U = \mathcal{F}(V)$, U: original (or "physical") variable V: new (or "artificial") variable

$$Z \equiv \int dU\, e^{-S(U)} = \int dV\, \underbrace{\det \mathcal{F}_*(V)}_{\text{Jacobian}}\, e^{-S(\mathcal{F}(V))} \equiv \int dV\, e^{-S_{\text{eff}}(V)}.$$



i.e., the action for the V-field is

$$S_{\text{eff}}(V) \equiv S\big(\mathcal{F}(V)\big) - \ln \det \mathcal{F}_*(V).$$

- We can perform the ordinary Monte Carlo sampling (e.g., HMC **Duane et al. 87**) in the V-space with the action $S_{\text{eff}}(V)$.

  ➡ Prepare $\mathcal{F}$ so that the sampling in V-space becomes efficient!

- Ultimate $\mathcal{F}$: *trivializing map* **cf. Nicolai map in SUSY theory [Nicolai 80]**

  $S_{\text{eff}}(V) = \text{const}$ (for which $Z = \int dV\ \text{const}$.)

  Such $\mathcal{F}$ will map the theory to the strong coupling limit ($\beta = 0$), which is the opposite of where the critical slowing down occurs ($\beta = \infty$).

Trivializing map (2/4)

Some ~~boring~~ mathematics

- We need to write down the Jacobian matrix $\mathcal{F}_*(V)$:
  - Introduce a local parametrization $(\theta^a_{x,\mu})$ of the field space around a configuration $U_{x,\mu}$:

$$e^{\theta^a_{x,\mu} T^a} U_{x,\mu}.$$

T^a: $\mathrm{su}(3)$ generators. $\mathrm{tr}\big(T^a T^b\big) = -\frac{1}{2}\delta^{ab}$

  - Haar measure: $(dU) \propto \prod_A d\theta^A$

    $A \equiv (x, \mu, a)$ labels the DOF

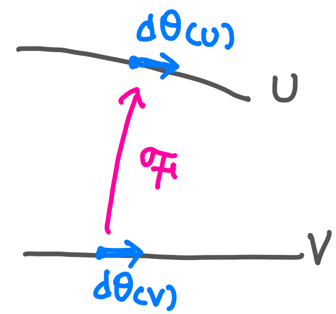
[right-invariant metric: $\|\delta U\|^2 \propto \sum_{x,\mu} \mathrm{Re\,tr}\big(U^\dagger_{x,\mu}\delta U_{x,\mu}\big)^2 \propto \sum_A (\delta\theta^A)^2$

for $\delta U_{x,\mu} \equiv \big(e^{\delta\theta^a_{x,\mu}T^a} - 1\big)U_{x,\mu}$.

e.g., Polyakov 87]

$\mathcal{F}_*(V) = (\mathcal{F}_*(V)^{AB})$ can be read off from the infinitesimals:

$$d\theta^A_{(U)} = \mathcal{F}_*^{AB}(V)\, d\theta^B_{(V)}.$$



- For later convenience, we also define the right-invariant derivative:

$$\partial^a_{x,\mu} U_{x,\mu} \equiv \lim_{t\to 0} \frac{\big(e^{tT^a}-1\big)U_{x,\mu}}{t} = T^a U_{x,\mu}.$$

[In other words, $\partial^a_{x,\mu} = \partial_{\theta^a_{x,\mu}}|_{\theta=0}$.]

<u>Comment on the convention</u>

In Lüscher 0907.5491, the symbol $\theta^a_{x,\mu}$ is used for the *Maurer-Cartan form* $\Theta^a_{x,\mu}$:

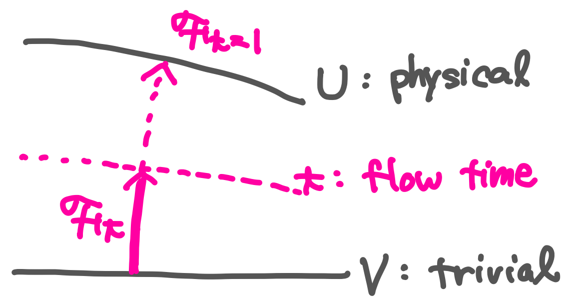
$$\Theta^a_{x,\mu} = \big(1 + O(\theta)\big)d\theta^a_{x,\mu} \qquad \text{(at each point } U_{x,\mu} \text{ on the group manifold).}$$

[$\Theta^a_{x,\mu}$ is the dual of $\partial^a_{x,\mu}$: $\langle\Theta^A, \partial^B\rangle = \delta^{AB}$.]   **See, e.g., Chevalley 46**

Trivializing map (3/4)

- Lüscher particularly considered the gradient flow form:

Luscher 0907.5491



$$\dot{\mathcal{F}}_t(U)_{x,\mu} = -T^a \partial^a_{x,\mu} \tilde{S}_t(U) \cdot U_{x,\mu} \,.$$

- Requiring that $\mathcal{F}_t$ trivializes the theory at $t = 1$:

$$S_{\mathrm{eff},\,t}(V) \stackrel{\Delta}{=} S\big(\mathcal{F}_t(V)\big) - \ln\det \mathcal{F}_{t*}(V) \overset{*\ \text{requirement}}{=} (1-t)\, S\big(\mathcal{F}_t(V)\big)$$

$$\Downarrow \ d/dt$$

NB
$S(U)$: original action
$S_{\mathrm{eff},t}(V)$: effective action
$\tilde{S}_t(U)$: flow kernel

Equation for the kernel function $\tilde{S}_t$:

$$\underbrace{-(\partial^A)^2 \tilde{S}_t}_{\text{from Jacobian}} + \underbrace{t\, \partial^A S\, \partial^A \tilde{S}_t}_{\text{from action}} \overset{*}{=} S$$

(up to const; ignored hereafter)

Solving the map has boiled down to solving a linear differential equation!

- For convenience we define

$$\mathcal{L}_t \equiv -(\partial^A)^2 + t\, \partial^A S\, \partial^A$$

$$\boxed{\therefore\ \mathcal{L}_t \tilde{S}_t \overset{*}{=} S}$$

Existence (with another math)

Luscher 0907.5491

- The differential operator $\mathcal{L}_t = -(\partial^A)^2 + t\,\partial^A S\,\partial^A$ is
  - elliptic ($\therefore$ bounded from below)
  - symmetric with respect to the inner product: $(\psi, \phi) \equiv \int (dU)\, e^{-t\,S(U)}\, \psi^*(U)\, \phi(U)$

    i.e., $(\psi, \mathcal{L}_t\phi) = (\mathcal{L}_t\psi, \phi)$.

$\therefore$ $\mathcal{L}_t$ shares almost the same properties with the Hamiltonian in QM.

$\Rightarrow$ $\mathcal{L}_t$ is diagonalizable and the eigenvectors form a complete set. **e.g., Lee 81**

- For a normalized eigenvector ψ_n, $\lambda_n = (\psi_n, \mathcal{L}_t\psi_n) = \int (dU)\, e^{-tS} |\partial^A \psi_n|^2 \geq 0$.

$\Rightarrow$
- eigenvalues of $\mathcal{L}_t$ are nonnegative
- zero-mode is only constant ($\because$ we need $\partial^A \psi_n = 0$)

$\Rightarrow$ $\mathcal{L}_t^{-1}$ can be taken modulo constant.

$\therefore$ The solution of $\mathcal{L}_t \tilde{S}_t = S$ exists!

t-expansion (1/2)

Luscher further gave a way to construct the map as a t-expansion: **Luscher 0907.5491**

- Expand $\tilde{S}_t$ as a Taylor series:

  $$\tilde{S}_t = \sum_{k\geq 0} t^k\, \tilde{S}^{(k)}.$$

  Plug into the equation:

  $$-(\partial^A)^2\tilde{S}_t + t\,\partial^A S\,\partial^A\tilde{S}_t = S.$$

  ⇒ Matching the powers of t,

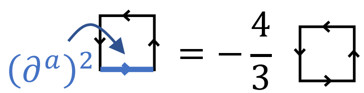
  $$\begin{cases} \mathcal{L}_0\tilde{S}^{(0)} = S, \\ \mathcal{L}_0\tilde{S}^{(k)} = -\partial S\cdot\partial\tilde{S}^{(k-1)}\ (k\geq 1). \end{cases}$$

- This recurrence equation can be inverted order by order.

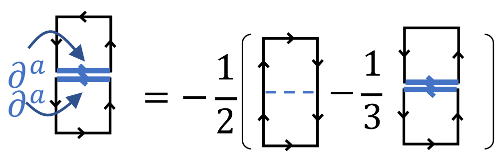
  ∵ Operator $\mathcal{L}_0$ can be represented as a matrix using the involved Wilson loops as basis functions.

Basic machinery

Recall:
$\partial^a_{x,\mu} \sim T^a$ insertion

- 

  $$\mathrm{tr}[(T^a)^2 \ldots] = -\tfrac{4}{3}\mathrm{tr}[\ldots]$$

- 

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

- Radius of convergence is proven to be finite.

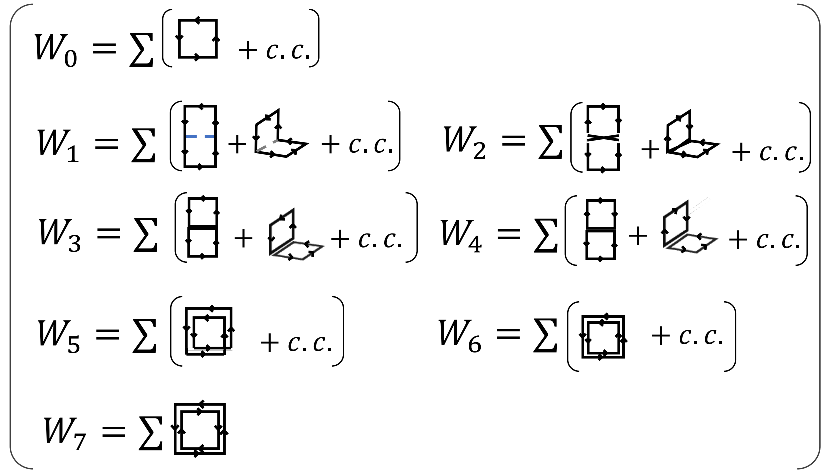
- Solution for the Wilson action case: **Luscher 0907.5491**

$$\tilde{S}_t = -\frac{\beta}{32} W_0$$ LO: plaquette

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

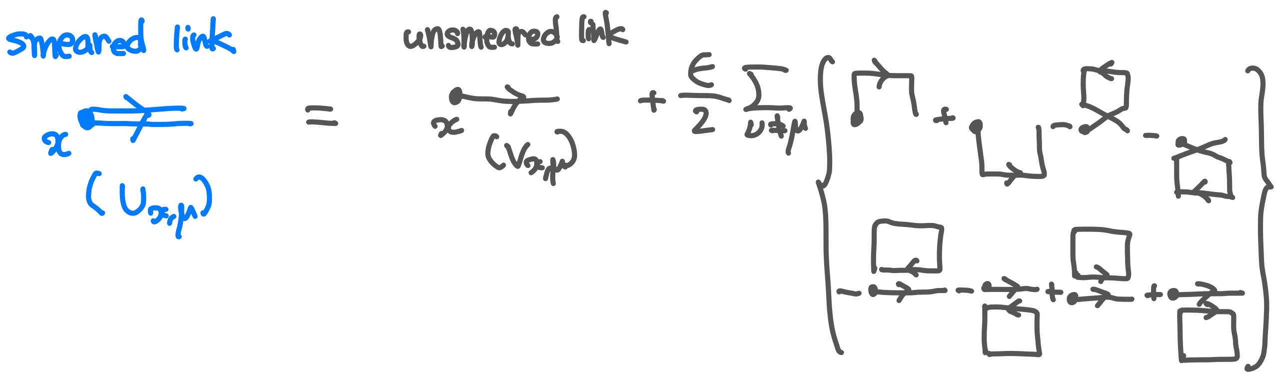
$$+O(t^2)$$

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



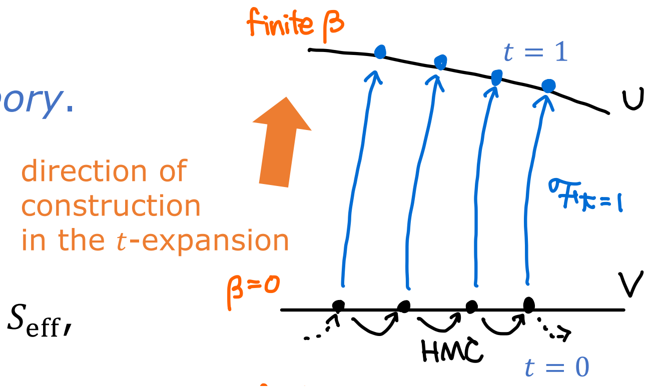
- Leading order: Wilson flow = stout smearing

Morningstar-Peardon hep-lat/0311018



Improving the map = Adding more complicated shapes in RHS

Boyle-Izubuchi-L.Jin-Jung-NM-Lehner-Tomiya 2212.11387 (work in progress)



- Note that the t-expansion is performed around $t = 0$;
  this corresponds to the *expansion around the trivial $\beta = 0$ theory*.

  Correspondingly, the expansion admits
  a similarity to the strong coupling expansion.

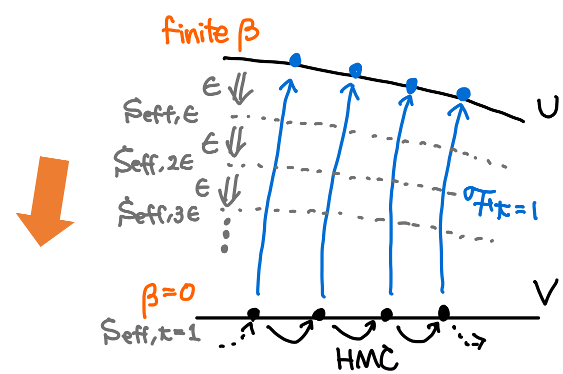
  In particular, at $t = 0$, one only needs to add the plaquette in S_{eff},
  ∴ the expansion begins with the plaquette.

- However, our primary target is to decrease β,
  for which the information of the nontrivial theory should be necessary.

  Because of the asymptotic freedom,
  one can expect that the relevant modes become rather wave-like for large β.

  ∴ it is possible that the exact solution $\tilde{S}_t$ is
  different from that around $\beta = 0$.

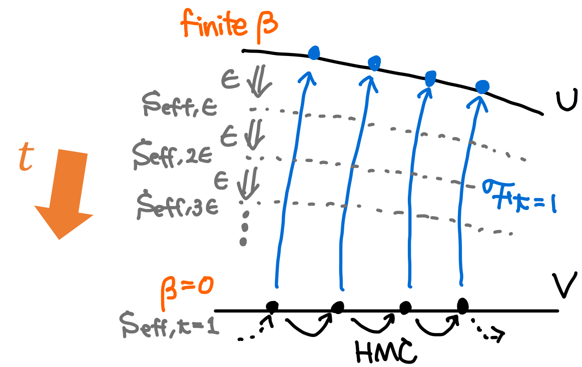
  ➡ We try to decrease the action
  step by step from the large β theory.



Boyle-Izubuchi-L.Jin-Jung-NM-Lehner-Tomiya 2212.11387 (work in progress)

- We require that at each flow time t:
  - $V \to \mathcal{F}_{t,\epsilon}(V)$
  - $S_{\text{eff},\,t+\epsilon}(V) \equiv S_{\text{eff},\,t}\big(\mathcal{F}_{t,\epsilon}(V)\big) - \ln\det \mathcal{F}_{t,\epsilon,*}(V)$

$$\overset{*}{=} S_{\text{eff},\,t}(V) - \epsilon\, S(V).$$



This suggests $S_{\text{eff},\,t}(V) = (1-t)\, S(V).$

use to be $\mathcal{F}_t(V)$ instead of V

Comment

Composition ordering needs to be reversed:

$$\mathcal{F}_{t=m\epsilon} = \mathcal{F}_{0,\epsilon} \circ \mathcal{F}_{\epsilon,\epsilon} \circ \mathcal{F}_{(m-1)\epsilon,\epsilon} \circ \mathcal{F}_{(m-1)\epsilon,\epsilon}.$$

$$\begin{aligned}
\because\quad & S_{\text{eff},\,t=m\epsilon}(V) \\
&= S_{\text{eff},\,(m-1)\epsilon}\big(\mathcal{F}_{(m-1)\epsilon,\epsilon}(V)\big) - \ln\det \mathcal{F}_{(m-1)\epsilon,\epsilon,*}(V) \\
&= S_{\text{eff},(m-2)\epsilon}\big(\mathcal{F}_{(m-2)\epsilon,\epsilon}(\mathcal{F}_{(m-1)\epsilon,\epsilon}(V))\big) \\
&\quad -\ln\det \mathcal{F}_{(m-2)\epsilon,\epsilon,*}\big(\mathcal{F}_{(m-1)\epsilon,\epsilon}(V)\big) - \ln\det \mathcal{F}_{(m-1)\epsilon,\epsilon,*}(V) \\
&= \cdots \\
&= S_{\text{eff},\,t=0}(\mathcal{F}_{0,\epsilon} \circ \cdots \circ \mathcal{F}_{(m-1)\epsilon,\epsilon}(V)) - \textstyle\sum_\ell \ln\det \mathcal{F}_{\ell\epsilon,\epsilon,*}\big(\mathcal{F}_{(\ell+1)\epsilon,\epsilon} \circ \cdots \circ \mathcal{F}_{(m-1)\epsilon,\epsilon}(V)\big) \\
&= S(\mathcal{F}_{m\epsilon}) - \ln\det \mathcal{F}_{m\epsilon,*}
\end{aligned}$$

Boyle-Izubuchi-L.Jin-Jung-NM-Lehner-Tomiya 2212.11387 (work in progress)

- The trivializing map with the requirement

$$S_{\text{eff},\, t+\epsilon}(V) \overset{*}{=} S_{\text{eff},\, t}(V) - \epsilon\, S(V).$$

can be related to Lüscher's $\mathcal{F}_t^{(L)}$.

(L) stands for Lüscher.

- We again assume the gradient form:

$$\mathcal{F}_{t,\epsilon}(U)_{x,\mu} = e^{-\epsilon\, T^a \partial^a_{x,\mu} \tilde{S}_t(U)} U_{x,\mu}.$$

⇒ equation for $\tilde{S}_t$:

$$[-(\partial^A)^2 + (1-t)\partial^A S\, \partial^A]\, \tilde{S}_t \overset{*}{=} S$$

we used to have:

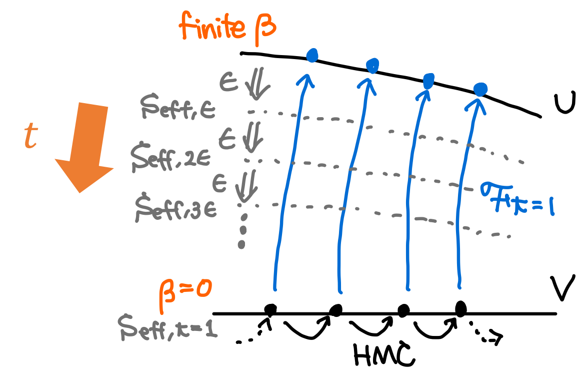
$$[-(\partial^A)^2 + t\, \partial^A S\, \partial^A] \tilde{S}_t^{(L)} = S$$

We notice that t is replaced by $1-t$.

∴ Writing $t = m\epsilon$ $(0 \le t \le 1)$ and $1 \equiv n\epsilon$,

$$\mathcal{F}_{t=m\epsilon} = \mathcal{F}_{0,\epsilon} \circ \cdots \circ \mathcal{F}_{(m-1)\epsilon,\epsilon} = \mathcal{F}^{(L)}_{(n-1)\epsilon,\epsilon} \circ \cdots \circ \mathcal{F}^{(L)}_{(n-m)\epsilon,\epsilon}.$$

In particular $\mathcal{F}_{t=1} = \mathcal{F}_{t=1}^{(L)}$ (for $\epsilon \to 0$).



∴ The above * corresponds to
constructing the same map from the opposite direction.

Table of contents

Complication in the 2nd approach: all Wilson loops can be relevant in principle

➡ Need a systematic way to truncate the function space

➡ Use *"action tomography"* with a Schwinger-Dyson equation:

Gonzalez-Arroyo, Okawa 87, de Forcrand et al. [QCD-TARO] hep-lat/9806008

- Expand $S_{\rm eff}(V)$ with *the effective couplings β_j:*

(we omit the t-dependence momentarily)

$$S_{\rm eff}(V) = \sum_j \beta_j W_j$$

$\{W_j\}$: Wilson loop basis

- β_j obey the linear equation:

$$\sum_j \beta_j \left\langle \partial^A W_j\, \partial^A W_i \right\rangle_{S_{\rm eff}} = \langle (\partial^A)^2 W_i \rangle_{S_{\rm eff}}$$

$\langle \cdot \rangle_{S_{\rm eff}}$: expectation value with respect to $S_{\rm eff}$

$\because$

- Consider a variation using W_i as the flow kernel:

$$\delta V = -\epsilon T^A \partial^A W_i \cdot V$$

- The path integral is invariant under this variation *(Schwinger-Dyson equation)*:

$$0 = \delta \int (dV)\ e^{-S_{\rm eff}(V)} = \int (dV)\ e^{-S_{\rm eff}(V)} \epsilon\ [\underbrace{-(\partial^A)^2\, W_i}_{\text{from Jacobian}} + \underbrace{\partial^A S_{\rm eff}\, \partial^A W_i}_{\text{from action}}]$$

- Combining this formula with the expansion of $S_{\rm eff}(V)$:

$$\sum_j \beta_j \left\langle \partial^A W_j\, \partial^A W_i \right\rangle_{S_{\rm eff}} = \langle (\partial^A)^2 W_i \rangle_{S_{\rm eff}}.$$

We can tell β_j from the expectation values!

cf. Gonzalez-Arroyo, Okawa 87, de Forcrand et al. [QCD-TARO] hep-lat/9806008

- Generically, we need infinite number of couplings to parametrize $S_{\mathrm{eff}}(V)=\sum_j \beta_j W_j$.

$\therefore \quad \sum_j \beta_j \left\langle \partial^A W_j\, \partial^A W_i \right\rangle_{S_{\mathrm{eff}}} = \langle (\partial^A)^2 W_i \rangle_{S_{\mathrm{eff}}}$: infinite-dimensional matrix (practically unusable)

- Instead, we can try to mimic $S_{\mathrm{eff}}(V)$ with a finite basis:

$S'_{\mathrm{eff}}(V)=\sum'_j \beta'_j W_j$ ← prime symbols indicate truncation (j runs a finite range)

and determine β'_j by:

$\sum'_j \beta'_j \left\langle \partial^A W_j\, \partial^A W_i \right\rangle_{S_{\mathrm{eff}}} = \langle (\partial^A)^2 W_i \rangle_{S_{\mathrm{eff}}}$: finite-dimensional matrix! (i is also restricted to the finite range)

Such β'_j turn out to give the best approximation of $S_{\mathrm{eff}}(V)$ in the sense that it minimizes the norm:

$\|S_{\mathrm{eff}}-S'_{\mathrm{eff}}\|_{S_{\mathrm{eff}}}$, where $\underline{\|S\|^2_{S_{\mathrm{eff}}} \equiv \langle (\partial^A S)^2 \rangle_{S_{\mathrm{eff}}}}$

L2 norm of the force
∴ The truncation error calculable.

∵ Subtracted equation

$$\sum'_j (\beta_j-\beta'_j) \left\langle \partial^A W_j\, \partial^A W_i \right\rangle_{S_{\mathrm{eff}}} = 0$$

is the stationary condition:

$$\begin{aligned}
&\frac{\partial}{\partial \beta'_i} \|S_{\mathrm{eff}}-S'_{\mathrm{eff}}\|^2_{S_{\mathrm{eff}}} \\
&= \frac{\partial}{\partial \beta'_i} \langle [\partial^A (S_{\mathrm{eff}}-S'_{\mathrm{eff}})]^2 \rangle_{S_{\mathrm{eff}}} \\
&= -2\sum'_j (\beta_j-\beta'_j) \left\langle \partial^A W_j\, \partial^A W_i \right\rangle_{S_{\mathrm{eff}}} \\
&\equiv 0.
\end{aligned}$$

This Schwinger-Dyson method gives us a way to truncate effective actions.

Boyle-Izubuchi-L.Jin-Jung-NM-Lehner-Tomiya 2212.11387 (work in progress)

Given a way to project the action onto a finite-dimensional function space, we can construct the flow in this subspace:

- Parametrize $\tilde{S}_t$ with the finite basis:

$$\tilde{S}_t(V) = \sum_k' \gamma_{k,t} W_k.$$

- Differentiate the equation for $\beta'_{j,t}$:

$$\left\langle -(\partial^A)^2 W_i + \sum_j' \beta'_{j,t}\, \partial^A W_j\, \partial^A W_i \right\rangle_{S_{\mathrm{eff},\,t}} = 0$$

d/dt

$$\sum_k' \gamma_{k,t} \underbrace{\left\langle \partial^B W_k \partial^B[-(\partial^A) W_i + \partial^A S'_{\mathrm{eff},\,t}\, \partial^A W_i] \right\rangle_{S_{\mathrm{eff},\,t}}}_{\text{from the Boltzmann weight}} = \underbrace{-\sum_j' \dot{\beta}'_{j,t} \left\langle \partial^A W_j\, \partial^A W_i \right\rangle_{S_{\mathrm{eff},\,t}}}_{\text{from explicit } t\text{-dependence of } \beta'_{j,t}}$$

This equation gives the coefficients $\gamma_{k,t}$ for a given $\dot{\beta}'_{j,t}$ (thus a trajectory of $S'_{\mathrm{eff},\,t}$)!

- We particularly take $\dot{\beta}'_{j,t} = -\frac{\beta'_{j,t}}{1-t}$ so that $\beta'_{j,t} = (1-t)\beta'_{j,t=0} = (1-t)\beta_{j,t=0}$.

Linear equation for $\gamma_{k,t}$

$$\sum_k' \gamma_{k,t} \left\langle \underline{\partial^B W_k \partial^B[-(\partial^A) W_i + \partial^A S'_{\mathrm{eff},\,t}\, \partial^A W_i]} \right\rangle_{S_{\mathrm{eff},\,t}} = \frac{1}{1-t} \left\langle \partial^A S'_{\mathrm{eff},\,t} \partial^A W_i \right\rangle_{S_{\mathrm{eff},\,t}}$$

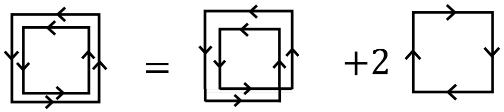
In practice, we use the numerical derivative with the five-point formula to calculate this matrix

Technical comment: Mandelstam constraints (1/1)

- Some of the basis functions are not linearly independent *("Mandelstam constraints")*

Mandelstam 79
See also Giles 81, Loll 93, Watson hep-th/9311126

Relevant example:



$$\left(W_6 = W_5 + 2W_0 \right)$$

$$\because \ (\mathrm{tr}U)^2 = \mathrm{tr}(U^2) + 2\mathrm{tr}U^\dagger \qquad \left(U \in SU(3) \right)$$

Further relation can be obtained by the Cayley–Hamilton eq:

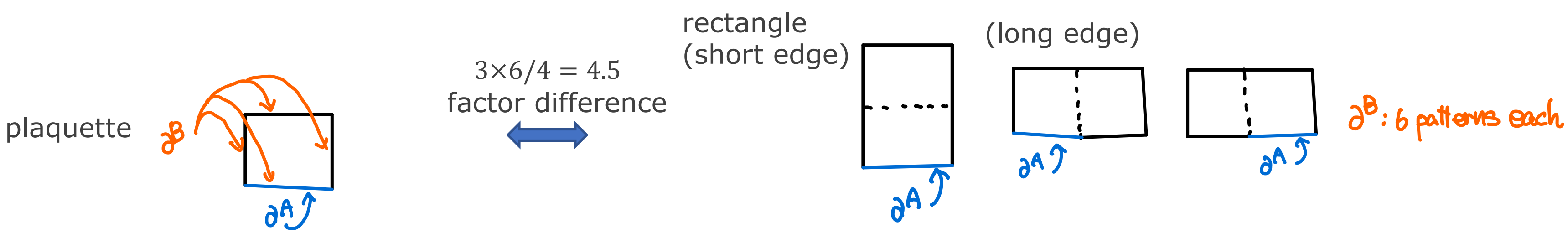
$$U^3 = (\mathrm{tr}\,U)U^2 - \frac{1}{2}[(\mathrm{tr}\,U)^2 - \mathrm{tr}\,U^2]U + \mathbb{I}.$$

- We need to pick a linearly independent basis to perform the inversions.

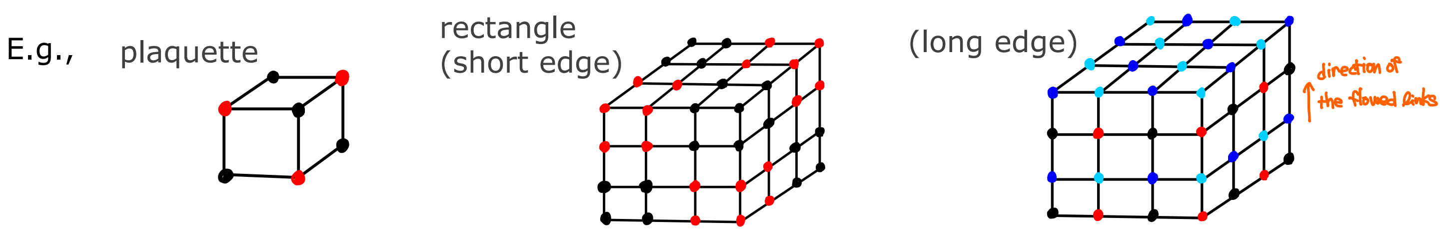
- We use the HMC with the exact transformed action $S_{\text{eff}}(V)$. Detailed algorithm: **Luscher 0907.5491**

- Fully parallelized code based on qlat software (C++ codebase) **cf. L Jin LATTICE 2021**
  https://github.com/waterret/Qlattice

- Most costly part: matrix mults including the Hessian $\partial^A \partial^B \tilde{S}_t$ in the force propagation.



- By dividing the directions of the flowed links and appropriately coloring/masking the lattice, we can run the multiplications in parallel. **Lüscher 0907.5491, Boyda et al. 2008.05456**



Coloring is actually mandatory to ensure that the map is one-to-one for a sufficiently small but finite ϵ.

Table of contents

- RIKEN HOKUSAI

HOKUSAI
BIG WATERFALL

- Univ of Tokyo Oakforest-PACS (retired)

picture taken from HP of CCS

- USQCD facility at BNL (KNL)
  funded by US DOE

Brookhaven
National Laboratory

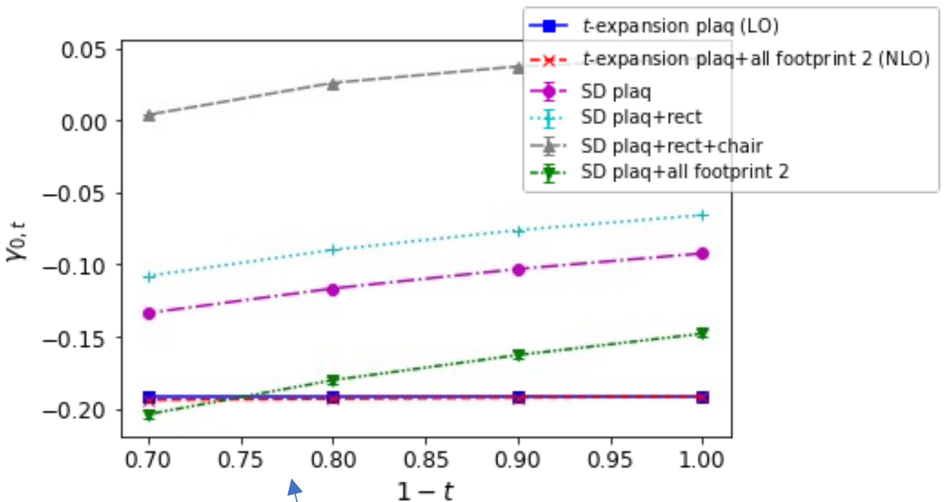
We are grateful for these resources.

Boyle-Izubuchi-L.Jin-Jung-NM-Lehner-Tomiya 2212.11387 (work in progress)

$8^4, \beta = 6.13$ Wilson $(a^{-1} = 2.56\ \mathrm{GeV})$

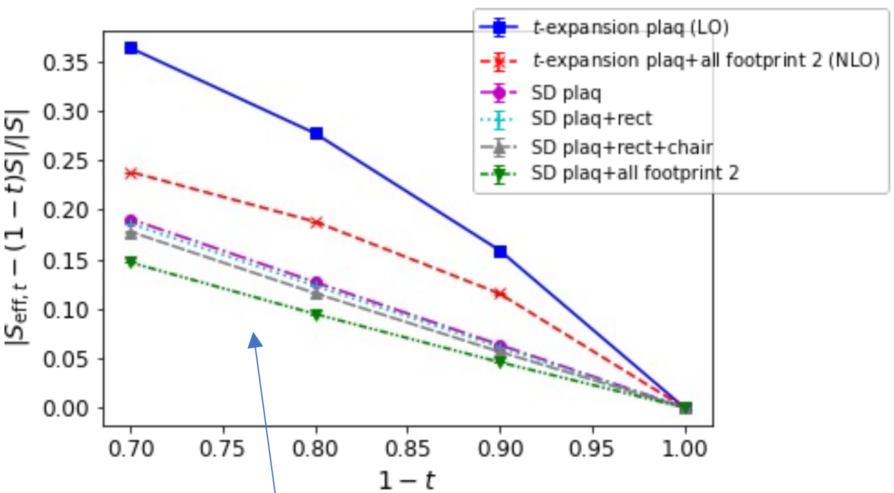
Ce-Consonni-Engel-Giusti 1506.06052

Determined $\gamma_{0,t}$(plaqutte coefficient)



flow obtained by Schwinger-Dyson (SD)
is quite different from that by t-expansion

Difference from the target trajectory



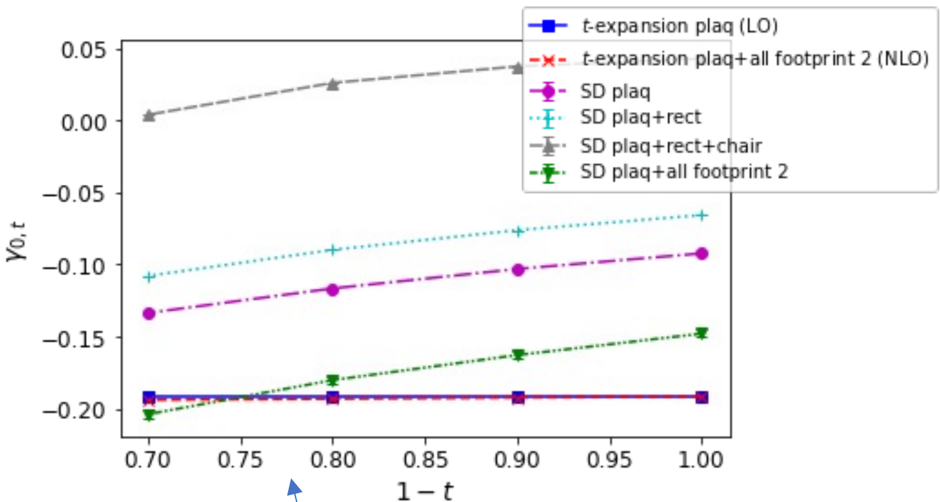
With the SD method,
we can have better control of the effective action

Boyle-Izubuchi-L.Jin-Jung-NM-Lehner-Tomiya 2212.11387 (work in progress)

$8^4, \beta = 6.13$ Wilson $(a^{-1} = 2.56 \text{ GeV})$

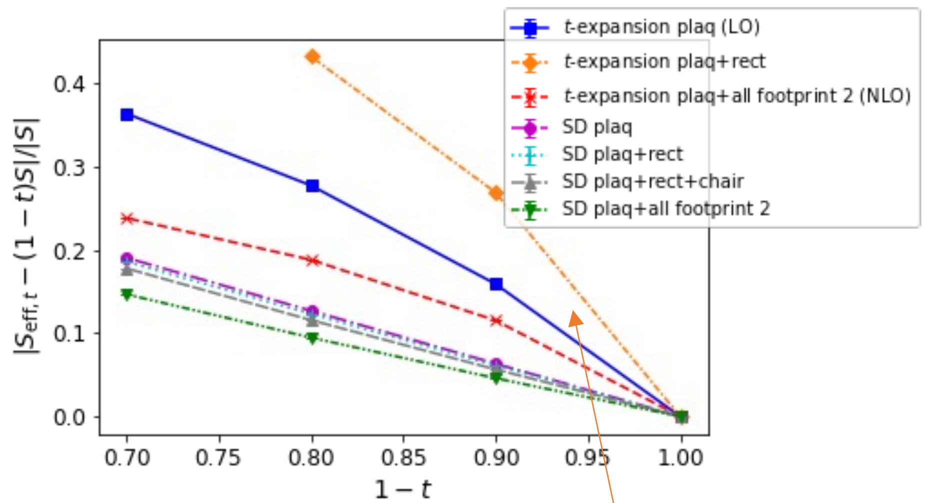
Ce-Consonni-Engel-Giusti 1506.06052

Determined $\gamma_{0,t}$(plaqutte coefficient)



flow obtained by Schwinger-Dyson (SD)
is quite different from that by t-expansion

Difference from the target trajectory

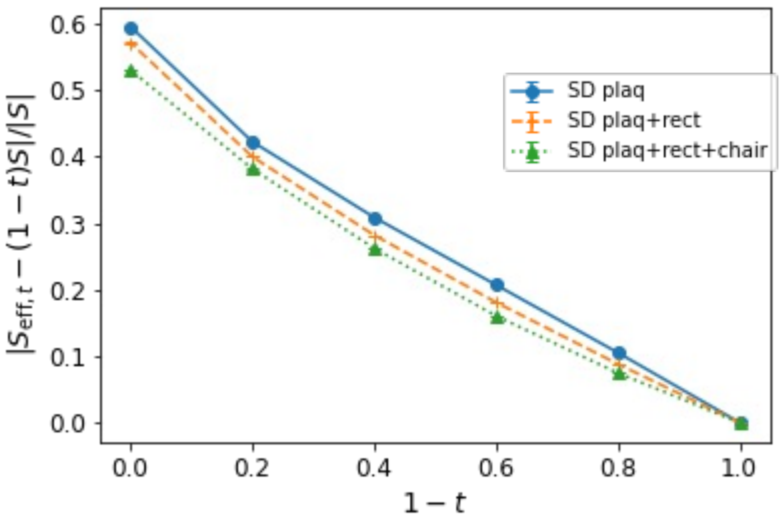


Naively adding the rectangle term
to the LO t-expansion makes
the deviation more significant

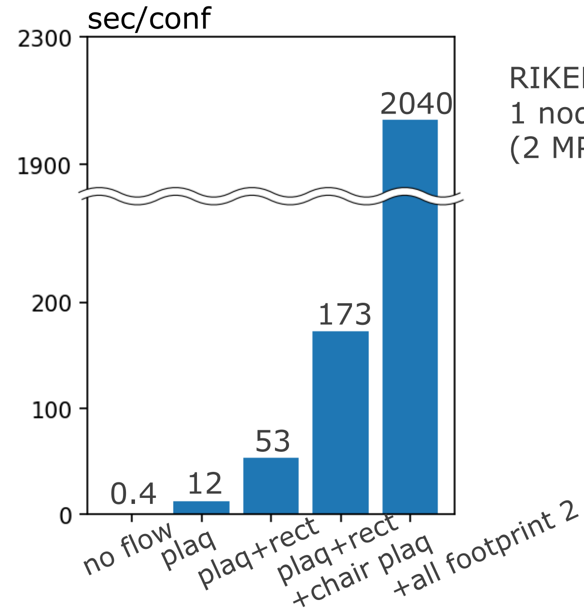
Boyle-Izubuchi-L.Jin-Jung-NM-Lehner-Tomiya 2212.11387 (work in progress)

$8^3 \times 16, \beta = 0.89$ DBW2 $(a^{-1} = 1.49\ \text{GeV})$
$(c_1 = -1.4008)$ Necco hep-lat/0309017

Difference from the target trajectory



Computational cost (1 step flow)



RIKEN HOKUSAI
1 node
(2 MPI x 40 OpenMP)

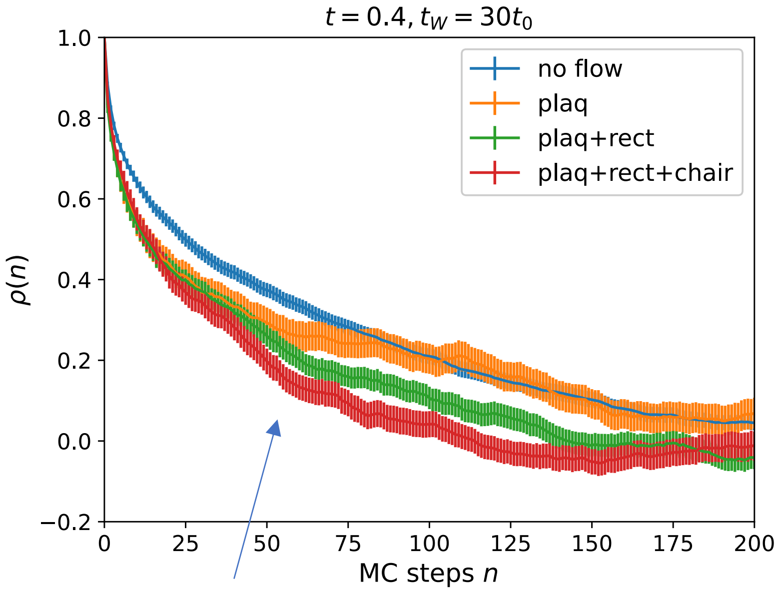
The increase can be understood by the increase of the nonzero matrix elements in $\partial^A \partial^B \tilde{S}_t$.

Boyle-Izubuchi-L.Jin-Jung-NM-Lehner-Tomiya 2212.11387 (work in progress)

$8^3 \times 16, \beta = 0.89$ DBW2 $(a^{-1} = 1.49\ \mathrm{GeV})$

$(c_1 = -1.4008)$ **Necco hep-lat/0309017**

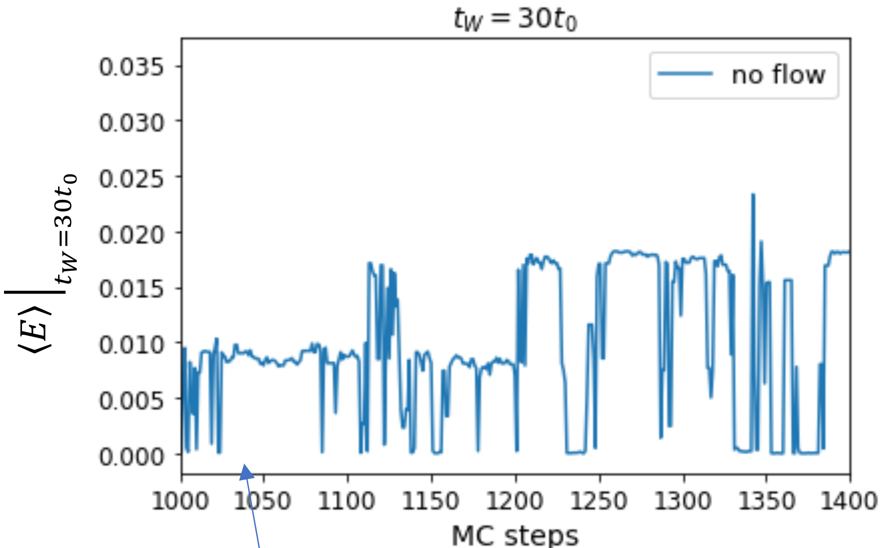
Normalized autocorrelation function $\rho(n)$ for the smeared energy density $(t_w = 30t_0)$



Faster decorrelation (in MC steps) by including extended loops

t_W: Wilson smearing flow time

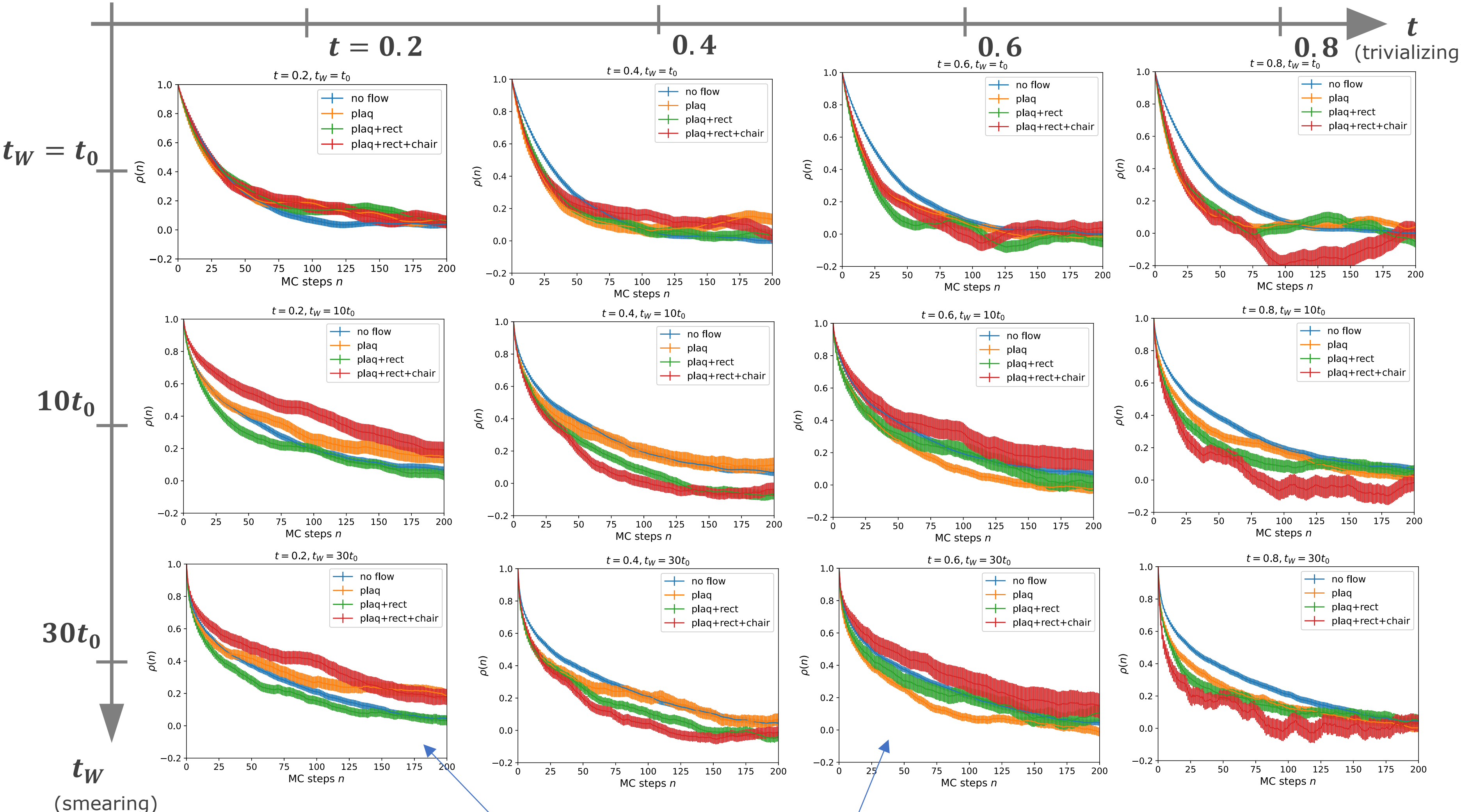
$t_W^2 \langle E \rangle|_{t_W = t_0} = 0.3$ **Luscher 1006.4518**



Smearing is performed so a long time that the energy reflects the instantons

Boyle-Izubuchi-L.Jin-Jung-NM-Lehner-Tomiya 2212.11387
(work in progress)

In most cases, faster decorrelation is observed by the extended loops,
but the autocorrelation is not controlled completely:



decorrelation by large kernels
not clear for these trivializing-flow times
(there seems to be an oscillation-like behavior)

Boyle-Izubuchi-L.Jin-Jung-NM-Lehner-Tomiya 2212.11387
(work in progress)

- From the oscillation-like behavior,
  it can be doubted that the determination of the flow coefficients $\gamma_{j,t}$ is not sufficient.

➡ Statistical errors? $O(\epsilon^2)$ effects?

More direct optimization may also be effective. **cf. Bacchio-Kessel-Schaefer-Vaitl 2212.08469**

- Relatively small improvements suggest that
  the large loops (which are truncated) contribute to increasing autocorrelation.

➡ What does the exact $\tilde{S}_t$ at large β actually look like?

An example that effectively succeeds in decreasing β?

More appropriate basis functions than the Wilson loops?

- MC sampling strategies:

FT-HMC can be numerically costly when including large Wilson loops.

➡ Can we arrange the accept/reject step
s.t. the scaling behavior is better enough for large lattices in 4D?

Discussion in this regard: e.g., Komijani-Marinkovic 2301.01504

Summary

- We proposed a way to design an approximate trivializing map with the Schwinger-Dyson equation

  Advantages of this method
    - the basis for the flow kernel can be chosen arbitrarily by hand
    - can be applied to the general action of interest
    - the coefficients in the kernel are determined by lattice estimates of the observables; no need for analytic calculation such as t-expansion
    - truncation effects and goodness of the flow can be measured by the force norm

- We showed that
  - With the SD method, we can have a better control of the effective action
  - We in some cases have positive effects on the autocorrelation of long-ranged objects by adding rectangles and chairs to the flow

Outlook

- Develop more efficient strategies
- Include fermion / develop algorithm that is capable of it.

Thank you.