Multilevel Hybrid Monte Carlo for lattice QCD

Leonardo Giusti

University of Milano Bicocca & INFN

WP1@Spoke2@ICSC - March 6^{th} 2023

Outline

- Introduction and motivations:
  - ∗ Signal/noise problem in LQCD, e.g. Hadronic Vacuum Polarization (HVP)
  - ∗ Proposed solution: Multi-Level (ML) integration
  - ∗ Novelty: factorization of the fermion determinant
- First results from large scale R&D simulations
- Proposed collaboration, activities, KPIs, ...
- Resource needed, expected synergies with other WPs, ...

The bottleneck: signal/noise ratio for HVP (HLbL,...)

- The HVP contr. to $a_\mu = (g-2)_\mu/2$ reads

$$a_\mu^{\rm HVP} = \left(\frac{\alpha}{\pi}\right)^2 \int_0^\infty dx_0 K(x_0, m_\mu) G(x_0)$$

  where

$$G(x_0) = \int d^3x \langle J_k^{\rm em}(x) J_k^{\rm em}(0)\rangle$$

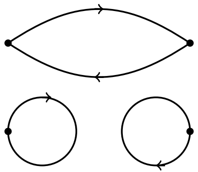
  with $K(x_0, m_\mu)$ being a known function

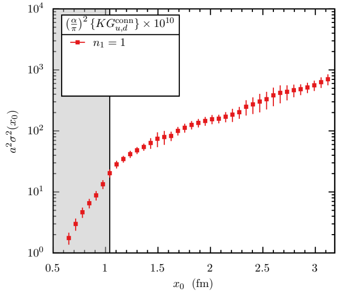
- For the light-connected contribution (by far the largest)

$$\frac{\sigma^2_{G^{\rm conn}_{u,d}}(x_0)}{[G^{\rm conn}_{u,d}(x_0)]^2} \propto \frac{1}{n_0}\, e^{2(M_\rho - M_\pi)|x_0|}$$

  where M_ρ lightest state in that channel.

- For disconnected contribution is worse since variance of correlator is constant in time





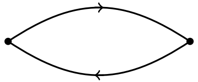
$a = 0.065$ fm , $M_\pi = 270$ MeV

$(V/a^4) = 96 \times 48^3$

Signal/noise ratio: the rôle of pions

- If D is the lattice Dirac operator and

$$W_\pi(y_0, x) = \sum_{\vec{y}} \mathrm{Tr}\left\{ D^{-1}(y,x)[D^{-1}(y,x)]^\dagger \right\}$$



at large time distances the pion propagator and its variance go as

$$C_\pi(y_0, x_0) = \langle W_\pi(y_0, x)\rangle \propto e^{-M_\pi |y_0 - x_0|} \qquad \sigma^2_\pi(y_0, x_0) \propto e^{-2M_\pi |y_0 - x_0|}$$

and therefore the signal/noise ratio is (almost) constant

- Indeed, when $|y - x| \to \infty$, numerical simulations confirm that

$$\mathrm{Tr}\left\{ D^{-1}(y,x)[D^{-1}(y,x)]^\dagger \right\} \propto e^{-M_\pi |y-x|}$$

for every gauge field in the representative ensemble. The size of each quark line, $\exp\{-M_\pi |y - x|/2\}$, is responsible for large fluctuations in other correlators

Signal/noise ratio: very generic problem

[Parisi 84; Lepage 89]

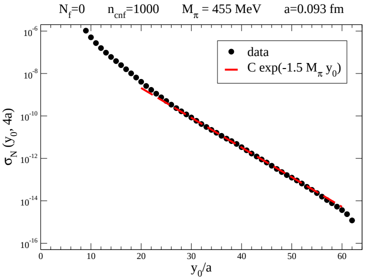
- Nucleon propagator

$$C_N(y_0, x_0) = \langle W_N(y_0, x_0) \rangle \propto e^{-M_N |y_0 - x_0|}$$

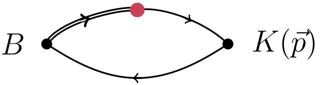
when $|y_0 - x_0| \to \infty$ goes as

$$\sigma_N^2(y_0, x_0) \propto e^{-3M_\pi |y_0 - x_0|}$$

and analogously for other baryonic correlation functions



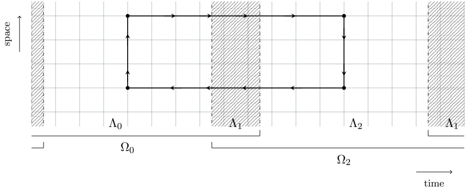
- Semileptonic B decays. Two (noisy) basic building blocks:
  - Mesons with (large) non-zero momentum
  - Static quark line



Multi-level integration for bosons (gluons)

[Parisi, Petronzio, Rapuano 83; Lüscher, Weisz 01; ...; Meyer 02; LG, Della Morte 08 10, ...]

- If the action and the observable can be factorized



$$
\begin{aligned}
S[U] &= S_0[U_{\Omega_0}] + S_2[U_{\Omega_2}] + \ldots \\
O[U] &= O_0[U_{\Omega_0}] \times O_2[U_{\Omega_2}]
\end{aligned}
$$

then

$$\langle O[U] \rangle = \langle \langle\langle O_0[U_{\Omega_0}] \rangle\rangle_{\Lambda_0} \times \langle\langle O_2[U_{\Omega_2}] \rangle\rangle_{\Lambda_2} \rangle_{\Lambda_1}$$

where

$$\langle\langle O_0[U_{\Omega_0}] \rangle\rangle_{\Lambda_0} = \frac{1}{Z_{\Lambda_0}} \int \mathcal{D}U_{\Lambda_0}\, e^{-S_0[U_{\Omega_0}]}\, O_0[U_{\Omega_0}]$$

- Two-level integration:

  - n_0 configurations U_{Λ_1}
  - n_1 configurations U_{Λ_0} and U_{Λ_2} for each U_{Λ_1}

- If $\langle\langle \cdot \rangle\rangle_{\Lambda_i}$ can be computed efficiently with a stat. error comparable to its central value, then prefactor in signal/noise ratio changes as (until S/N problem solved)

$$n_0 \rightarrow n_0 n_1^2$$

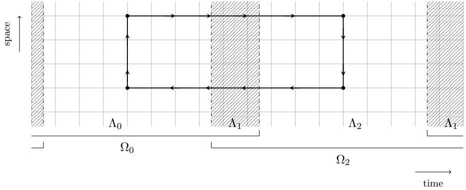
at the cost of generating approximately $n_0 n_1$ level-0 configurations

Multi-level integration for bosons (gluons)

[Parisi, Petronzio, Rapuano 83; Lüscher, Weisz 01; . . . ; Meyer 02; LG, Della Morte 08 10, . . .]

- If the action and the observable can be factorized

$$S[U] = S_0[U_{\Omega_0}] + S_2[U_{\Omega_2}] + \dots$$
$$O[U] = O_0[U_{\Omega_0}] \times O_2[U_{\Omega_2}]$$



then

$$\langle O[U] \rangle = \langle \langle\langle O_0[U_{\Omega_0}]\rangle\rangle_{\Lambda_0} \times \langle\langle O_2[U_{\Omega_2}]\rangle\rangle_{\Lambda_2} \rangle_{\Lambda_1}$$

where

$$\langle\langle O_0[U_{\Omega_0}]\rangle\rangle_{\Lambda_0} = \frac{1}{Z_{\Lambda_0}} \int \mathcal{D}U_{\Lambda_0}\, e^{-S_0[U_{\Omega_0}]}\, O_0[U_{\Omega_0}]$$

- With more active blocks, at the cost of approximatively $n_0 n_1$ level-0 configurations,

$$n_0 \to n_0 n_1^{n_{\rm block}}$$

and gain increases exponentially with the distance since $n_{\rm block} \propto |y_0 - x_0|$. For the same relative accuracy of correlator, computational effort would then increase approximatively linearly with the distance

Factorization of the fermion determinant

[Cè, LG, Schaefer 16; Dalla Brida, LG, Harris, Pepe 20; LG, Saccardi 22]

- Thanks to

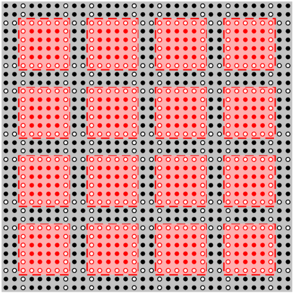
  * Overlapping Domain Decomposition
  * Multi-Boson representation

  factorization of $\det D$ becomes possible

- Indeed

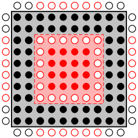
$$\det\ D = \frac{\det W_1}{\det D_{\Lambda_1}^{-1} \prod_{\hat{a}} \left[\det D_{\Phi_1^{\hat{a}}}\ \det D_{\Omega_0^{\hat{a}}}^{-1}\right]}$$

  where Λ_1 is the global (grey) "frame"



- The deviations of W from identity suppressed with thickness of frame. W can be represented by multi-bosons as

$$\det W_1 = \frac{\mathcal{W}_N}{\prod_{k=1}^{N/2} \det \left[W_{u_k}^{\dagger}\ W_{u_k}\right]}$$



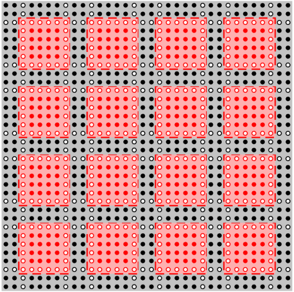
Basic framed domain $\Omega_0^{\hat{a}} = \Lambda_0^{\hat{a}} \cup \Phi_1^{\hat{a}}$

Multi-level integration with fermions

[Cè, LG, Schaefer 16; Dalla Brida, LG, Harris, Pepe 20; LG, Saccardi 22]

- Thanks to
  - ∗ Overlapping Domain Decomposition
  - ∗ Multi-Boson representation

  factorization of $\det D$ becomes possible



- Indeed

$$\det\ D = \frac{\det W_1}{\det D_{\Lambda_1}^{-1} \prod_{\hat{a}} \left[\det D_{\Phi_1^{\hat{a}}} \det D_{\Omega_0^{\hat{a}}}^{-1}\right]}$$

where Λ_1 is the global (grey) "frame"

- The effective gluonic action for 2 flavours, for instance, can be represented as

$$\{\det D^\dagger D\} = \int \mathcal{D}\phi \dots \exp\left\{-\sum_{\hat{a}} S_0^{\hat{a}}[U_{\Omega_0^{\hat{a}}},\dots] - S_1[U_{\Lambda_1}]\right\}$$

where the dependence from the gauge field in $\Lambda_0^{\hat{a}}$ is factorized

Signal/noise ratio for HVP: multi-level solution

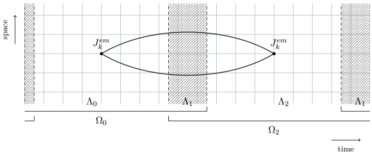
[Dalla Brida, LG, Harris, Pepe 21]

- Wilson glue with $O(a)$-improved Wilson quarks

$$\beta = 5.3\,, \quad (T/a)\times(L/a)^3 = 96\times 48^3$$

$$a = 0.065\ \mathrm{fm}\,, \quad M_\pi = 270\ \mathrm{MeV}$$

$$n_0 = 25\,, \qquad n_1 = 10\,, \qquad n_{tot} = n_0 \cdot n_1$$

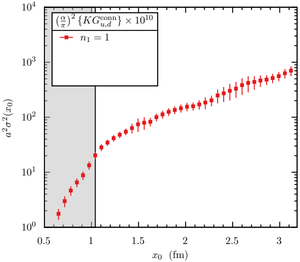


- Domain Decomposition adopted:

$$\Lambda_0 : x_0/a \in [0,39]$$

$$\Lambda_1 : x_0/a \in [40,47] \cup [88,95]$$

$$\Lambda_2 : x_0/a \in [48,87]$$



Signal/noise ratio for HVP: multi-level solution

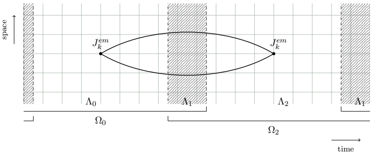
[Dalla Brida, LG, Harris, Pepe 21]

- Wilson glue with $O(a)$-improved Wilson quarks

$$\beta = 5.3\,, \quad (T/a) \times (L/a)^3 = 96 \times 48^3$$

$$a = 0.065\ \mathrm{fm}\,, \quad M_\pi = 270\ \mathrm{MeV}$$

$$n_0 = 25\,, \qquad n_1 = 10\,, \qquad n_{tot} = n_0 \cdot n_1$$

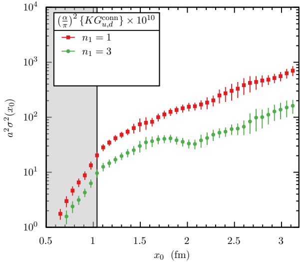


- Domain Decomposition adopted:

$$\Lambda_0 : x_0/a \in [0, 39]$$

$$\Lambda_1 : x_0/a \in [40, 47] \cup [88, 95]$$

$$\Lambda_2 : x_0/a \in [48, 87]$$



Signal/noise ratio for HVP: multi-level solution

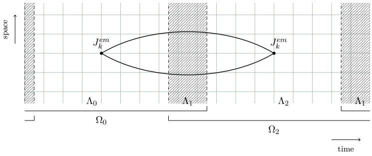
[Dalla Brida, LG, Harris, Pepe 21]

- Wilson glue with $O(a)$-improved Wilson quarks

$$\beta = 5.3\,, \quad (T/a) \times (L/a)^3 = 96 \times 48^3$$

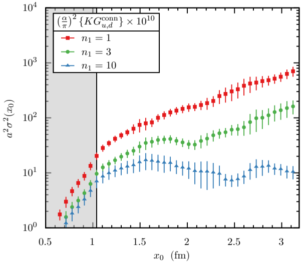
$$a = 0.065\ \mathrm{fm}\,, \quad M_\pi = 270\ \mathrm{MeV}$$

$$n_0 = 25\,, \qquad n_1 = 10\,, \qquad n_{tot} = n_0 \cdot n_1$$



- Sharp rise of σ^2 with x_0 when computed by a standard 1-level integration (red points) is automatically flattened out by the 2-level integration (blue-points)

- Accurate computations can be obtained at large distances: no need for any modeling of the long-distance behaviour of $G_{u,d}^{\mathrm{conn}}$



Use case definition and participants

Use case definition: The Algorithm and the associated highly parallel code for Multilevel Hybrid Monte Carlo for lattice QCD

Participating institutions: INFN, University of Milano - Bicocca

Participants: M. Bruno, M. Cè, L. Giusti, M. Pepe; P. Rescigno, M. Saccardi

External participants: M. Dalla Brida (CERN), T. Harris (ETH), M. Lüscher (CERN)

New resources from ICSC: 1 RTDA (selection completed by the end of April 2023), 1/2 Ph.D

Toward the exaflop: KPIs

Accessible lattice: Wilson glue with $O(a)$-improved Wilson quarks with

$$(T/a) \times (L/a)^3 = 256 \times 256^3 \sim 4 \cdot 10^9$$

$$\#Threads = 4^4 = 256, \qquad \#CPUs = 8^4 = 4096,$$

$$\#cores = 32^4 \sim 10^6$$

```
#define NPROC0 8
#define NPROC1 8
#define NPROC2 8
#define NPROC3 8

#define L0 32
#define L1 32
#define L2 32
#define L3 32

#define L0_TRD 8
#define L1_TRD 8
#define L2_TRD 8
#define L3_TRD 8
```

Selected starting open source code: openQCD-2.4
`https://luscher.web.cern.ch/luscher/openQCD/`

First MPI implementation of multi-level: work already in progress toward MB-DD-HMC-xx.xx, expected to be completed in spring 2024

Final MPI+openMP implementation of multi-level: openQCD-xx.xx expected to be completed and released in summer 2025

Resources needed for R&D and production

R&D:

Computer time per year: 15M corehours

Disk space: (3*50)=150 TB

Tape space: 300 TB

Production:

Computer time per year: hundreds of Millions of corehours (EuroHPC)

Disk space: (3*50)=150 TB

Tape space (repository WP5): (3*1000)=3 PB