Data Management and Ensemble Generation within CLS

Wolfgang Söldner
on behalf of CLS

Regensburg University

Lattice 2022
The 39th International Symposium on Lattice Field Theory
August 9th, 2022

>cls

Simulation Overview

CLS 2 + 1f simulation program

→ see https://www-zeuthen.desy.de/alpha/public-cls-nf21/

- CLS (Coordinated Lattice Simulations): HU Berlin, CERN, TC Dublin, Krakow, UA Madrid, Mainz, Milano Bicocca, Münster, Odense/CP3-Origins, Regensburg, Roma I, Roma II, Wuppertal, DESY Zeuthen
- lattice action and simulations
  - two degenerate light quarks and one strange quark
  - non-perturbatively improved Wilson action (clover)
  - tree-level improved Symanzik gauge action
  - several improvement coefficients and renormalization constants available
- utilizing open boundaries in time (at small lattice spacings) → avoid topological freezing
  - some ensembles with (anti-)periodic boundary conditions in time ($a \gtrsim 0.06$ fm)
  - always periodic boundary conditions in spatial directions

⇒ main focus of this large scale project: performing a well controlled continuum limit

Overview of ensembles generated and used within CLS

- ensembles available at small lattice spacing (up to $a = 0.039$ fm), at phys. point, high statistics with 1000 - 20000 MDUs
- three different chiral trajectories
  - constant average quark mass: $\overline{m} = m_{\mathrm{symm}}$ ← $\mathrm{Tr}M = const.$
    → main CLS trajectory with most ensembles available at lattice spacings $a \in [0.085\,\mathrm{fm}, 0.039\,\mathrm{fm}]$
  - constant physical strange quark mass: $\widetilde{m}_s = \widetilde{m}_{s,\mathrm{ph}}$
  - symmetric line: $m_s = m_\ell$

Simulation Overview

Reweighting

- strange reweighting factor to correct for error from rational approximation
- twisted mass reweighting $\rightarrow$ add a twisted mass term to light quark action in simulations to stabilize HMC runs

$\Rightarrow$ $\langle O \rangle = \frac{\langle RO \rangle}{\langle R \rangle}$ with observable O and rwt. factor R

Negative reweighting factors

- strange quark is simulated using rational approximation of $\sqrt{D^\dagger D}$ $\leftarrow$ positive sign of $\det D$ was assumed
- nevertheless (quite unexpected) neg. signs of $\det D$ have been observed at coarse lattice spacings $\rightarrow$ see [2003.13359]
- only for very few ensembles the fraction of configurations with a negative reweighting factor was found to be sizable

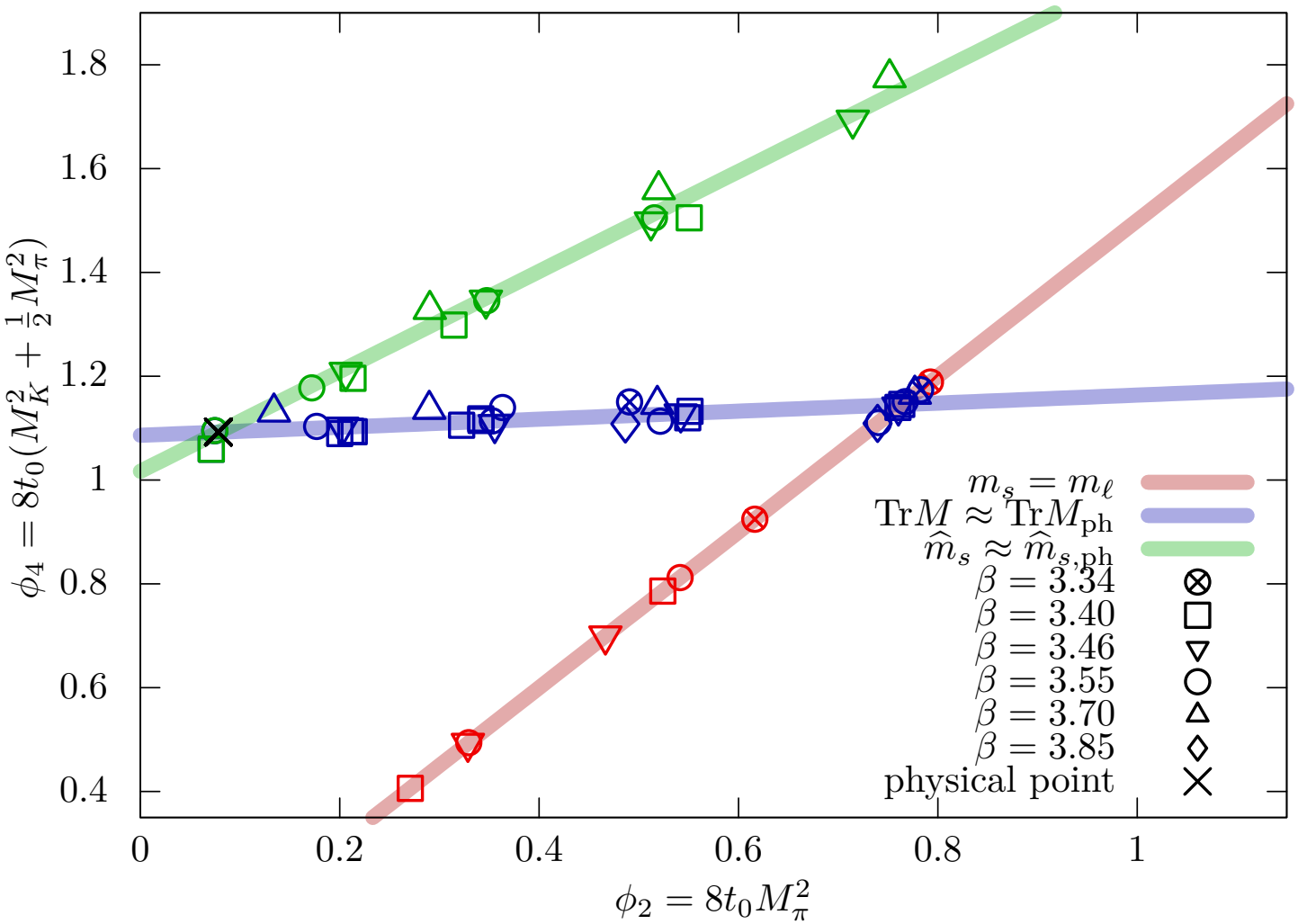
$\Rightarrow$ Note: negative signs have to be included in ensemble averages of observables

Current status of ensemble generation

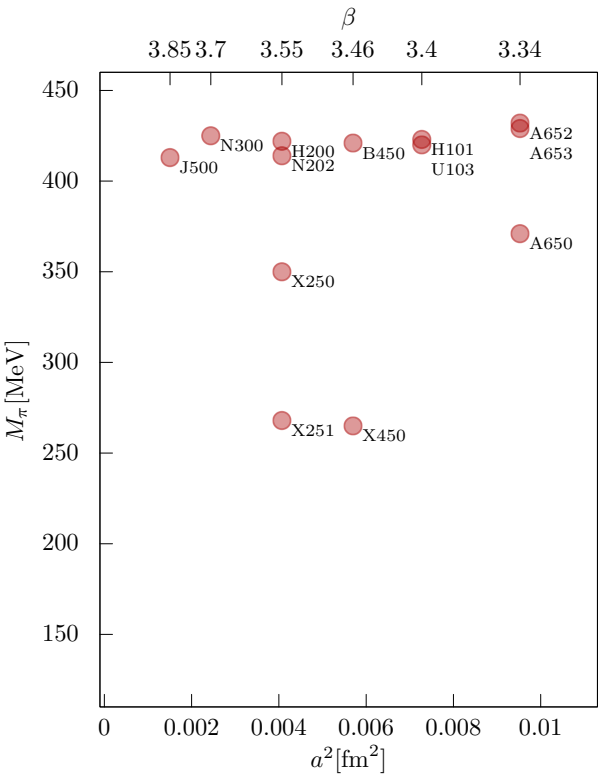
- generation of ensembles still ongoing
- plans for future (expensive) ensembles: under discussion

>cls

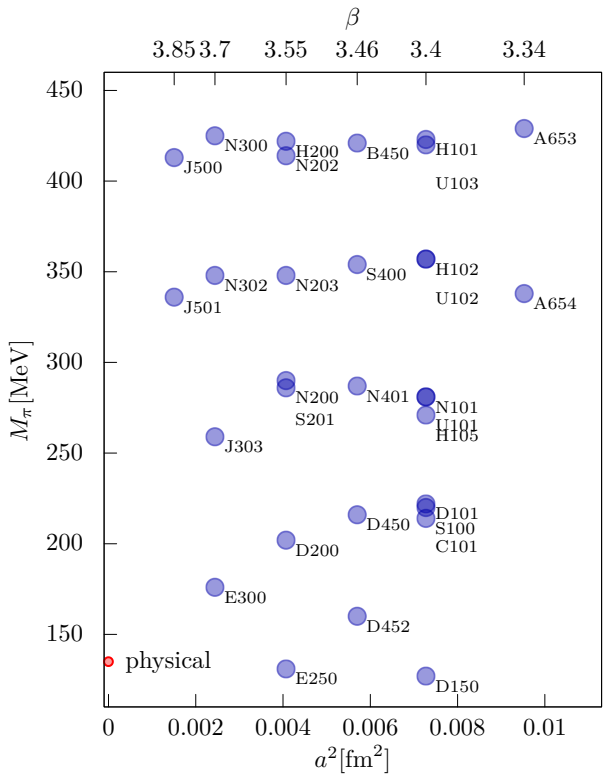
Ensemble overview



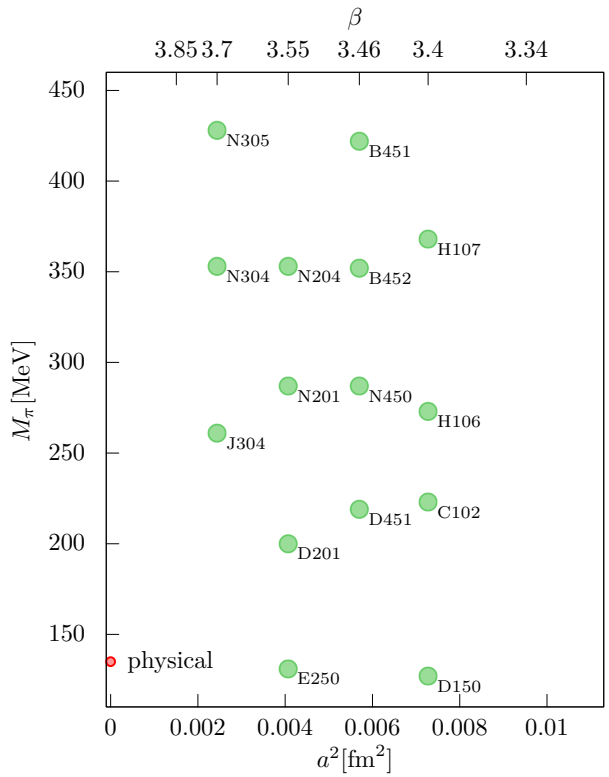
Ensemble overview



$m_s = m_\ell$



$m_s + 2m_\ell \approx$ phys.

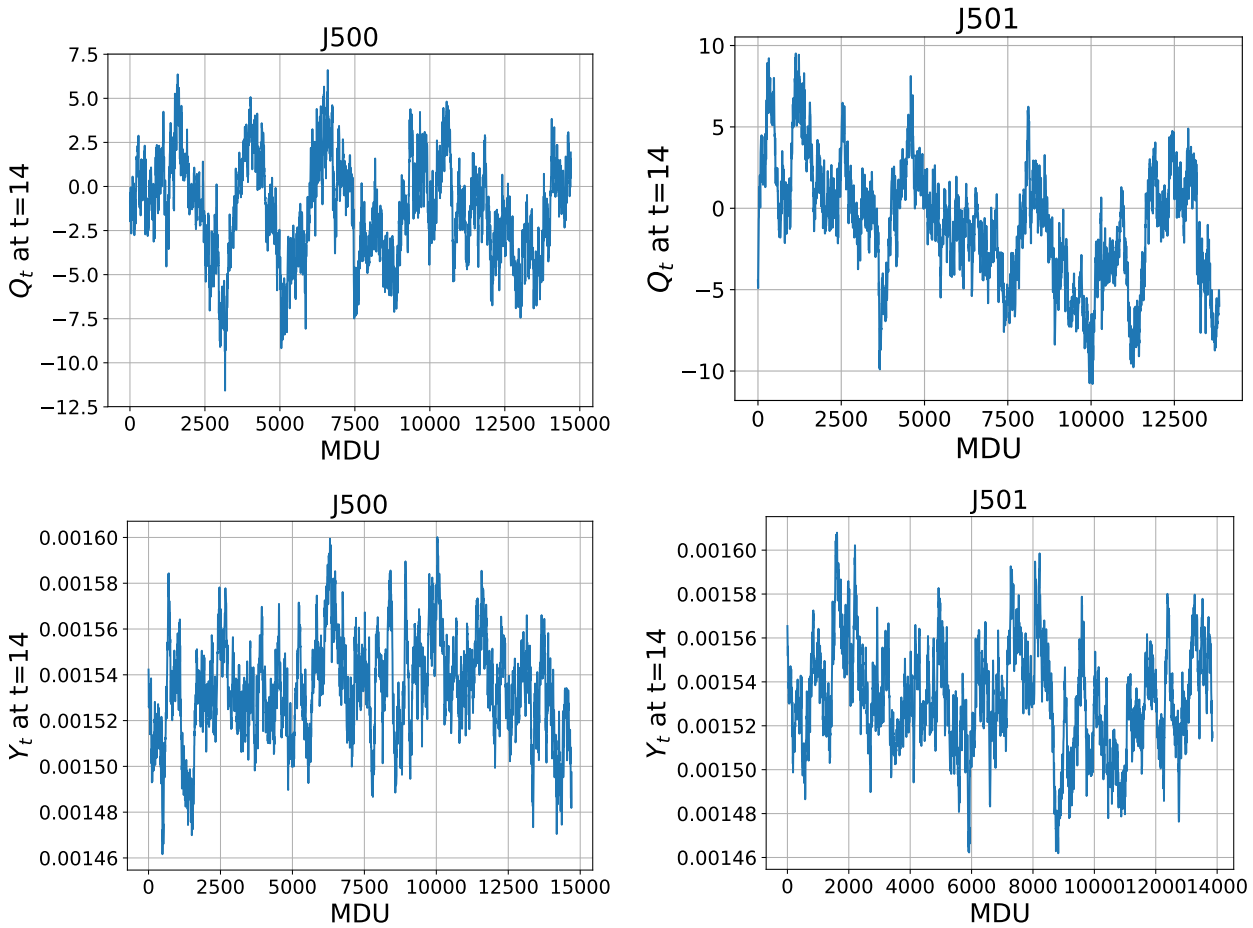


$\widehat{m}_s \approx$ phys.

- 6 different lattice spacings ($a \approx 0.098 - 0.039$ fm), 2 ensembles at the physical point
- geometries range from 48×24^3 to 192×96^3
- note: a few ensembles with heavier masses are not displayed (N500,..)

>cls

Autocorrelations at finest lattice spacing $a = 0.039\,\mathrm{fm}$

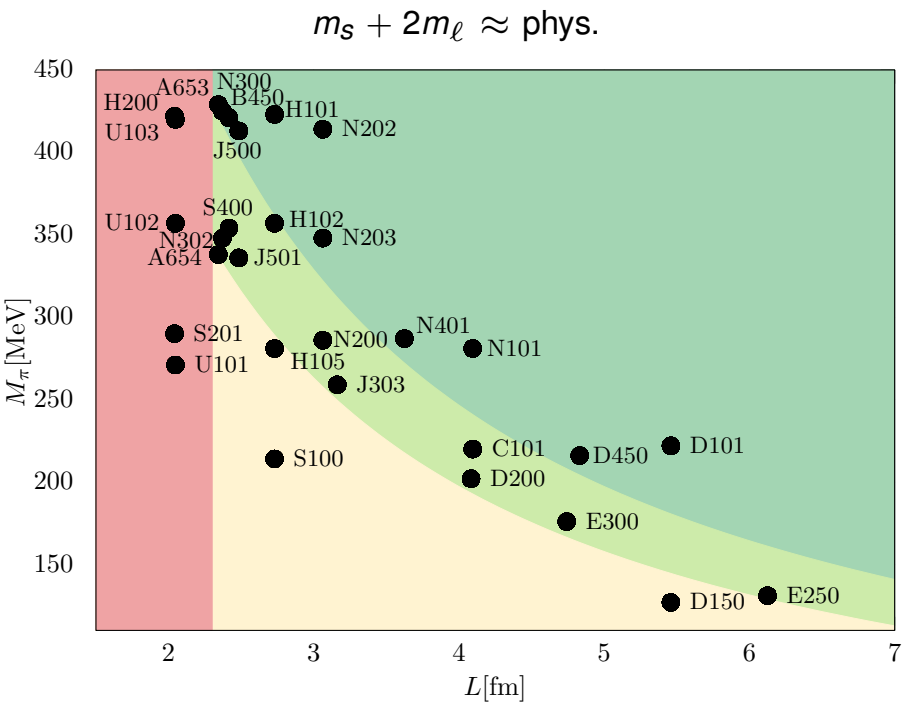


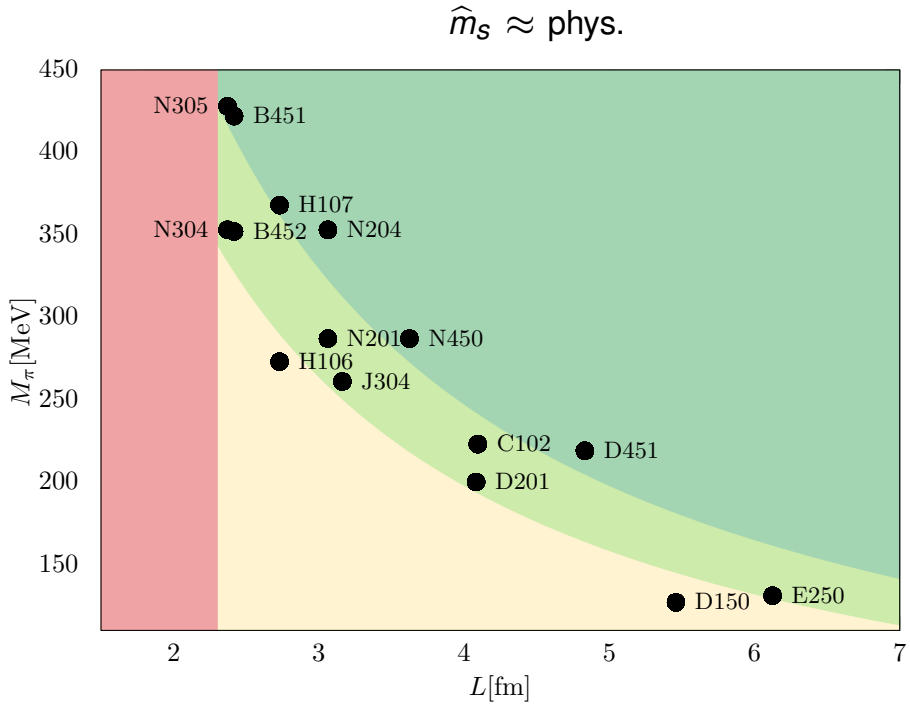
Monte Carlo history of topological charge Q_t and action density Y_t at Wilson flow time $t/a^2 = 14 \approx t_0/a^2$

J500/1: 192×64^3, $M_\pi \approx 420/350$ MeV, $a \approx 0.039$ fm

- $\rightarrow$ large autocorrelations (as expected)
- $\rightarrow$ open boundaries work nicely (no topological freezing)

Finite volume effects





Overview of ensembles: Volumes

- yellow area: $M_\pi L \leq 4$
  light green area: $4 < M_\pi L \leq 5$
  green area: $5 < M_\pi L$
- almost all ensembles are within light green or green area (and also $L \gtrsim 2.3\,\text{fm}$)
  $\rightarrow$ small finite volume effects

Data Management

Data details

- 126,000 configurations (975 TB) stored at Zeuthen and redundantly at Regensburg (on tape)
  → public overview of storage status at `https://www-zeuthen.desy.de/alpha/public-cls-nf21/`
- configurations typically separated by 4 MDU, generated with publicly available code (openQCD)
- openQCD data format (binary format, double precision, layout different to ILDG format)
- multiple checksums for each configuration are available for later data verification

Data management

- (internal) webpage available with overview and details (metadata) of existing ensembles
- metadata collected: runtime environment (machine, code-version, persons that ran the simulation), simulation parameters (input and log files), observables (plaquette, Q, Y_t, Q_t, t_0), checksums
- transfer (via scp) directly after generation to backup destinations
- tools are available for backing up data (reading and writing to tape), scripts are used for automated data handling

Sharing of ensembles

- configurations are available for users outside CLS upon request
- policy: old configurations are available, newer configurations on a project by project basis on request
- transfer of configurations is performed after arrangement with a storage site
- plan: make most ensembles available through LDG (once fully operational again)