subMIT and Lattice QFT

Yin Lin 林胤
yin01@mit.edu
MIT

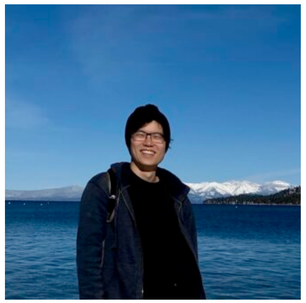
Jan 6, 2023
subMIT Workshop

Massachusetts Institute of Technology

Lattice group members at CTP and IAIFI


Ryan Abbot


Artur Avkhadiev


Joshua Lin

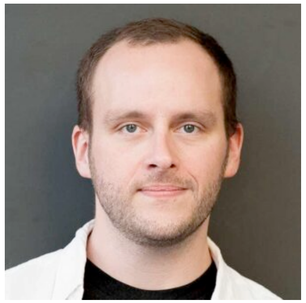

Marianne Moore


Patrick Oare


Dimitra Pefkou

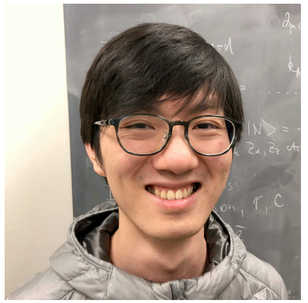

Cagin Yunus


Denis Boyda


Daniel Hackett


William Jay


Jamie Karthein


Yin Lin

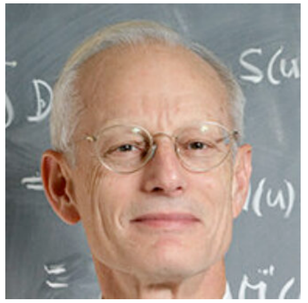

Di Luo


Fernando Romero-Lopez


Julian Urban


Andrew Pochinsky


William Detmold

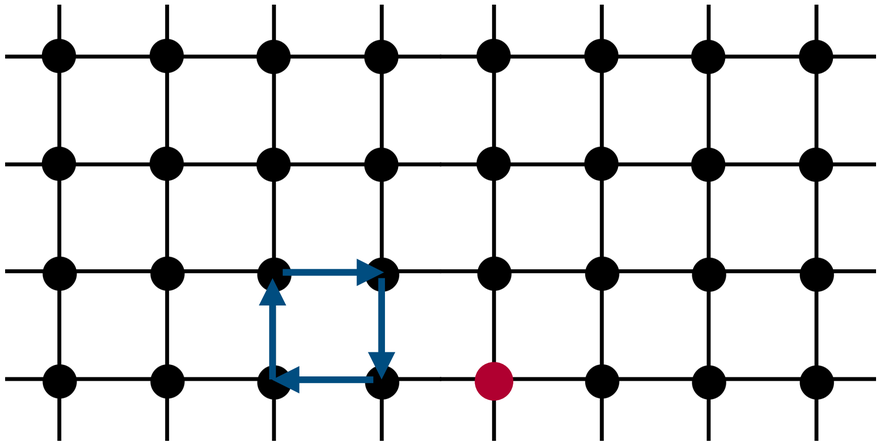

John Negele


Phiala Shanahan

How to compute with LQFT

$$\langle O \rangle = \int \prod_n dU_n d\psi_n d\bar{\psi}_n \frac{e^{-S[U_n,\bar{\psi}_n,\psi_n]}}{Z} O[U_n, \psi_n, \bar{\psi}_n]$$

$$\langle O \rangle = \underbrace{\int \prod_n dU_n}_{\text{measure}} \underbrace{\frac{\det(D) e^{-S_G[U]}}{Z}}_{\text{probability density}} \underbrace{\widetilde{O}[U_n]}_{\text{observable}}$$



$U_\mu(x)$ $\psi(x),\ \bar{\psi}(x)$

yin01@mit.edu Yin Lin 林胤

LQFT Workflow

$$\langle O \rangle = \int \prod_n dU_n \frac{\det(D) e^{-S_G[U]}}{Z} \widetilde{O}[U_n] \approx (1/N) \sum_i^N \widetilde{O}[U_n^{(i)}]$$

[USQCD collaboration, arXiv:1904.09725]

- **Generate configurations**
  - Stochastically estimate determinant → solve Dirac equations many times
  - Output ~10^3 to 10^4 configurations each with size of ~GB
- **Compute correlators $\widetilde{O}[U_n^{(i)}]$ on each configuration**
  - Solve Dirac equations many times, matrix multiplications, …
  - Output a large number of files with negligible sizes
- **Analyze correlators to extract physical observables**
  - chi-squared minimization with bootstrap/jackknife, …

Software and hardware in LQFT

- USQCD software stack (C, C++, CUDA, …) for configuration generation and measurements with openMP and MPI

- Grid (C++) and openQCD (C) are also used in production
- More flexibility for analyses — Python, Julia, C, R …
- Local LQCD cluster (GPU's only, ~20 nodes)
- USQCD collaboration allocates computing resources yearly on various machines across national labs through a proposal process (~million CPU-core-hours, ~100K GPU-hours per project per year)
- More computationally intensive computations are done on national supercomputers

Generating gauge configurations on subMIT

- Most computationally intensive step of most lattice calculations
- Multi-node jobs with strongly coupled nodes needed for production (solve Ax = b)
- Need MPI and openMP with CPU's, but preferably GPU's (Grid, QUDA)
- Only feasible for simple models, testings, or finite temperature on subMIT
- Community resources available → need large storage space (~TB's) that can be shared among all users (globus transfer?)

Example ensembles with physical quark masses

"Very coarse"

"Coarse"

"Fine"

Lattice spacing	Lattice Dimension	Config Available	Config size	Total size
~0.15fm	$32^3 \times 48$	~10^4	433 MB	~5 TB
~0.12fm	$48^3 \times 64$	920	6.8 GB	~6 TB
~0.06fm	$96^3 \times 192$	1449	46 GB	~67 TB

https://github.com/milc-qcd/sharing/wiki/LatticeSharing

Correlator measurements on subMIT

- High throughput — measurements on each configuration can take place independently
- Still need multi-node jobs with strong couplings for large lattices (solve Ax = b)
- Need MPI and openMP with CPU's, but preferably GPU's (Grid, QUDA)
- Small lattices ("very coarse" and "coarse") might be feasible on subMIT with single-node jobs with GPU's

Correlator analyses on subMIT

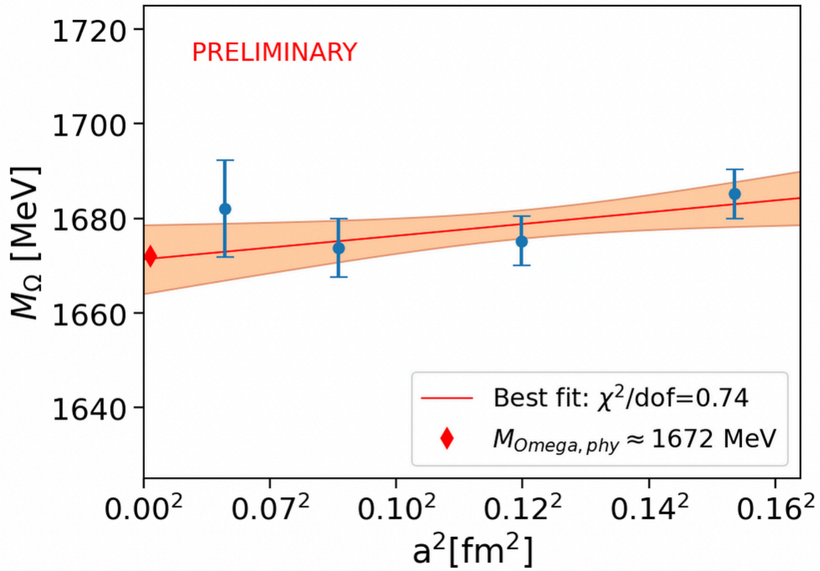
- Most are done in Python with NumPy, SciPy, and other more specialized packages
- Perform fits (optimization) on CPU's with SciPy or the GNU Scientific Library
- Interactive — use of Jupyter notebook is common
- Suitable use case for subMIT after correlators are measured in other resources

```
Least Square Fit:
  chi2/dof [dof] = 0.37 [4]    Q = 0.83    logGBF = 6.9814

Parameters:
              0   1.6713 (73)      [ 1.670 (50) ]
              1     0.29 (26)      [    1 (100) ]
              2   0.03 (1.00)      [  0.0 (1.0) ]

Settings:
  svdcut/n = 1e-12/0    tol = (1e-08*,1e-10,1e-10)    (itns/time = 4/0.1)
  fitter = scipy_least_squares    method = trf

chi2dof = 0.7374082683586785
```



First impressions of subMIT for LQFT

- Have only used it for performing analyses (~5 GB correlator data) with JupyterHub
  - Easy to use web interface with complete file access
  - Would be nice to have more flexibility in deciding the parameters of the server (type of nodes, number of cores, wall time, exclusive access, …)
  - Globus endpoint to transfer data?
- Limited to one-node jobs for generating configurations and measurements due to weak couplings
  - Suitable for preliminary testings before deploying to other resources
- More CPU nodes will be valuable (LQCD clusters have only GPU nodes)
- Database servers and containers (presentation by Denis Boyda)

yin01@mit.edu Yin Lin 林胤