Plan Overview

A Data Management Plan created using DMP Tool

Title: Calculation of nucleon axial form factors, proton decay amplitudes, and nucleon EDMs using domain wall fermions

Creator: Sergey Syritsyn

Affiliation: Stony Brook University (stonybrook.edu)

Principal Investigator: Tom Blum, Michael Abramczyk, Taku Izubuchi, Hiroshi Ohki

Data Manager: Tom Blum, Michael Abramczyk, Taku Izubuchi, Hiroshi Ohki

Project Administrator: Tom Blum, Michael Abramczyk, Taku Izubuchi, Hiroshi Ohki

Funder: Digital Curation Centre (dcc.ac.uk)

Template: Digital Curation Centre

Project abstract:

We propose calculation of neutron electric dipole moment (EDM) induced by theta-term in the QCD Lagrangian using background electric field method using low-lying modes of the Dirac operator. Nucleon EDM is one of the most important indications of $\CP$ violation if discovered, and the theta-term is one of the candidates for such interaction. Calculations involving theta-term are extremely challenging at the physical point, and we propose exploring methodology that is likely to be specifically efficient at examinig its effect. We propose to augment our earlier high-statistics calculations performed at mpi=340 and 420 MeV, and explore the efficiency of this technique at the physical point. The bulk of our request will be used in computing low-lying eigenmodes of the Dirac operator in QCD with uniform background field.

Start date: 07-01-2021

End date: 06-30-2022

Last modified: 07-08-2024

Calculation of nucleon axial form factors, proton decay amplitudes, and nucleon EDMs using domain wall fermions

Data Collection

What data will you collect or create?

Data sets (will) include bulk raw output for

- neutron correlation functions (CFs), topological charge density, and their correlation saved from lattice production code with reliable checksums to be preserved for perpetuity
- eigenvectors of the Dirac operator in QCD plus uniform background electric field

In addition to that, "stripped" and gathered data sets resulting from intermediate analysis stages (IAS) will be included for quick access to specific observables.

How will the data be collected or created?

Raw hadron CF data are generated by parallel lattice production code run on large partitions of CPU, GPU, and KNL clusters. Similarly, top.charge data are generated by lattice production code but represent incomplete CFs that are analyzed afterwards. Post-processing (IAS) data will include selection of data entries relevant for specific observables such as nucleon electric dipole moments, form factors, decay amplitudes, etc as well as best-fit parameters for a range of models.

Documentation and Metadata

What documentation and metadata will accompany the data?

Raw and bulk data will be saved in hierarchical files akin to directories and files, with data sets addressed by key paths that are intended to be self-explanatory metadata in their own kind. In addition, IAS data stored in HDF5 file records will be accompanied by metadata attributes or data sets, which will describe, e.g., axes of multidimensional data sets, ranges of parameter variation along the axes, etc. Raw data are saved as AFF file with strong checksums. IAS data are saved using HDF5 file format. Both file formats are accessible using Python modules. Whenever meaningful, data collections will be accompanied by a README file with description of a particular run, parameters, etc. Description of common storage conventions will be available as a PDF manual upon request.

Ethics and Legal Compliance

How will you manage any ethical issues?

none expected

How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?

none expected

Storage and Backup

How will the data be stored and backed up during the research?

Data will be copied to tapes at USQCD tape facilities at JLab, FNAL, and BNL. No additional backup is envisioned. In case of loss of IAS data, they can be regenerated from raw data; in case of raw data loss, IAS data can still be used for analysis and reproducibility.

How will you manage access and security?

All data will be accessible via standard Unix file system or a dedicated command-line utility. Security will be managed with standard Unix file permissions using ACL wherever possible.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

Raw output CFs data are of very high value because it is not easily recreated, in particular due to manpower involved. Its preservation is essential to reproducibility of analysis. In addition, IAS data will be preserved for reproducibility.

What is the long-term preservation plan for the dataset?

Upon publication, raw hadron CF data may be deleted. Post-processed (IAS) data will be preserved because it is valuable for extending projects to other observables, improving precision, performing cross-check and regression tests, as well as reproducibility.

Data Sharing

How will you share the data?

Data can shared immediately for noncompeting projects within the USQCD collaboration and outside. If there is potential for competition outside of ongoing collaboration, data will be shared on a case by case basis. As a general rule, (a) within USQCD exclusivity for analysis and publication is preserved for 1 year after data are generated and (b) outside USQCD such exclusivity is preserved for 3 years unless agreements with other USQCD groups dictate otherwise. Apart from that, data will be considered publicly accessible.

Data sets will be available through tape repositories, with requesting parties having direct access whenever possible. References to stored data will be included in papers, advertised at conferences, and published on project websites. Data sets will be identified by gauge configuration ensemble and ad hoc "run name" generated during the workflow.

Are any restrictions on data sharing required?

Acknowledgement upon publication of research using the data sets is expected. Early requests to access IAS data will be generally heavily scrutinized since they may contain ready-to-publish research findings.

Responsibilities and Resources

Who will be responsible for data management?

All project members will be responsible for data management.

What resources will you require to deliver your plan?

As of 03/2021, the following data sets have been stored and are planned for preservation:

24ID	-2021	proton decay CFs	10 TB
32IDfine	-2021	proton decay CFs	10 TB
24ID_m010	2021-2023	NEDM CFs, topcharge	5 TB
24ID_m005	2021-2023	NEDM CFs, topcharge	5 TB