

DOIs for MILC asqtad ensembles Future Roles of ILDG

J. Simone and the MILC Collaboration

Heath O'Connell, FNAL

Melissa Clegg, FNAL

Annette Holtkamp, CERN

Laura Rueda Garcia, CERN

Patricia Sigrid Herterich, CERN

Sunje Dallmeier-Tiessen, CERN

Jannean Elliott, DOE OSTI

April 23, 2015

Status

- DOIs have been issued for all of the interesting MILC asqtad ensembles
- Working with Inspire-HEP developers (FNAL + CERN) to ingest DOIs and the related metadata.
- I will assist Inspire-HEP with adding data references by hand to Inspire records of existing key publications using asqtad.
- Use standardized data citation format in future publications. Goal is for Inspire to automatically index.

Why DOIs?

- Persistent identification, in which each DOI name permanently and unambiguously identifies the object to which it is associated.
- A DOI differs from, e.g., a URL, in that it identifies an object itself as a first-class entity, rather than the specific place where the object is located at a certain time.
- DOIs are already used by publishers and Inspire-HEP to identify publications.
- Inspire-HEP can link publications to data: ensembles vastly more useful to share if the whole body of related publications is readily available.
- Inspire-HEP used worldwide by HEP and NP researchers.
- NERSC Gauge Connection will be able to generate a current bibliography of publications for ensembles by querying Inspire.
- Easily report to funding agencies metrics concerning data sharing and scientific impact of their investments.

Publication links to data



The image shows a screenshot of the INSPIRE website. At the top, there is a navigation bar with the INSPIRE logo and a welcome message. Below that is a secondary navigation bar with links for HEP, HEP NAMES, INSTITUTIONS, CONFERENCES, JOBS, EXPERIMENTS, and JOURNALS. A 'HELP' link is also present. The main content area features a search bar and a list of tabs: Information, References (373), Citations (286), Files, Plots, and Data. The 'Data' tab is highlighted with a red circle. Below the tabs, the title of a publication is displayed: 'Nonperturbative QCD simulations with 2+1 flavors of improved staggered quarks'. The authors are listed as MILC Collaboration (A. Bazavov et al.) with a link to show all 13 authors. The publication date is Mar 2009, and it is 164 pages long. The journal information is Rev.Mod.Phys. 82 (2010) 1349-1417. The DOI is 10.1103/RevModPhys.82.1349. The FERMILAB-PUB-09-674-T identifier is also present. The e-Print is arXiv:0903.3598 [hep-lat] and the PDF is available. The experiment is LATTICE-MILC. The abstract is also visible, starting with 'Abstract (arXiv) For several years the MILC collaboration has carried out nonperturbative simulations of full QCD with two degenerate flavors of light quarks, up and down, and with one heavier flavor, the strange quark. Several light quark masses, down to about three times the physical light quark mass, and several lattice spacings have been used. These allow for controlled continuum and chiral extrapolations of many low energy QCD observables. Use of an improved staggered quark formalism, 'asqtad' fermions, has been crucial in achieving this goal. Here we review the improved staggered formalism, emphasizing both advantages and drawbacks. In particular, we review the procedure, known as the 'fourth root trick' for removing unwanted staggered species in the continuum limit. We then describe the lattice ensembles created so far, and the physics results obtained on them. These include the heavy quark potential, spectrum of light hadrons, quark masses, decay constants of light and heavy-light pseudoscalar mesons, semileptonic form factors, computation of the strong coupling constant, spectroscopy of quarkonia, neutral meson mixing, and more. We illustrate the impact of some of these results on the determination of CKM matrix elements. All MILC lattice ensembles are publicly available. Some of the results mentioned were obtained by other groups using these MILC ensembles, some were obtained by MILC in collaboration with other groups, and some by the MILC collaboration alone.'

A new **Data** tab will appear in the Inspire full record for publications having data citations. It will link to a page listing the datasets.

Note: I drew the Data box in this image!

Here's a live ATLAS example
<https://inspirehep.net/record/1241574/data>

Data sharing blog
<http://blog.inspirehep.net/2013/10/enabling-data-sharing-citation-and.html>

Inspire citation recommendation

INSPIRE provides a "citation recommendation" for data. It is just a suggestion on how to cite a dataset from a publication. Do you have any preference?

For example, other recommendations look like this:

*Cite as:*ATLAS Collaboration (2013)
HEPData,<http://doi.org/10.7484/INSPIREHEP.DATA.26B4.TY5F>

As data citation is an emerging practice, it is key to provide guidelines. We use DOIs to track citations to data and we also include them as part of the author's profiles, our goal is that every time a dataset is reused, it gets referenced correctly.

Getting a DOI

- Regional grids remain curators of the data and guarantee continual public access.
- Assume responsibility to keep the DOI live!
- Find a DOI issuing organization and registration authority
 - USQCD: DOE Office of Scientific and Technical Information (OSTI) via DataCite.
 - GlobusOnline paid service includes DOI registration
 - CERN: <http://doi.web.cern.ch/>
 - DataCite members <https://www.datacite.org/about-datacite/members>
- Markup data sets with required metadata.
- Register dataset with issuing organization.
- Dataset repository should have markup data plus any extended metadata

Some OSTI XML Markup

<code>title:</code>	and description:	Give the non-experts a clue
<code>site_url:</code>		landing page for the data on NERSC Gauge Connection
<code>creators:</code>		Authors of dataset + ORCID, if known
<code>contract_nos:</code>		Funding agency (DOE) contracts
<code>subject_categories_code:</code>		“72 Physics Of Elementary Particles and Fields”
<code>dataset_type:</code>		“ND” – numeric data
<code>software_needed:</code>		MILC code (should be assigned a DOI)
<code>other_identifying_nos:</code>		I used this field for key physics parameters
<code>keywords:</code>		search terms, e.g. QCD; SU(3); . . .
<code>contact:</code>		Technical contact person

I use Python to generate the XML from a template. A simple 'CURL' shell script interacts with the OSTI registration web service.

What does a MILC DOI look like?

title: "MILC asqtad QCD SU(3) gauge ensemble;
series=a; a=0.043fm; Ls=2.8fm; Nf=2+1; u0.m0=(0.0028,0.014)"

[10.15484/milc.asqtad.en24a/1177873](https://dx.doi.org/10.15484/milc.asqtad.en24a/1177873)

- The prefix 10.15484 is registered to USQCD. All DOIs currently begin with 10.
- A namespace controlled by USQCD. Encodes collaboration, common name of the collection of ensembles (asqtad), type of dataset ('en' \Rightarrow gauge ensemble) and a unique identifier ('24a' \Rightarrow asqtad ensemble number 24, series a).
- OSTI's unique id for this dataset.
- Goal is to keep DOI short. Encourage researchers to use USQCD controlled infix part as the preferred label for an ensemble?

Any DOI is converted to a URL with addition of a prefix
<http://dx.doi.org/10.15484/milc.asqtad.en24a/1177873>

OSTI DOE Data Explorer

(/dataexplorer/)

Search DOE Data Explorer for Energy and Science Data

Lattice QCD gauge ensemble

X Q

+ Advanced Search

DOE Data Explorer (/dataexplorer/) / Search Results / Page 1 of 4

Detail View

Search for: Lattice QCD gauge ensemble

Sort by Relevance

- +

- Lattice QCD gauge ensemble:** USQCD/MILC/asqtad/641192f21b781m0028m014
(/dataexplorer/biblio/1177873)
Aubin, Christopher Alan; Bernard, Claude W.; Burch, Tommy; Datta, Saumen; DeGrand, Thomas Alan; DeTar, Carleton E.; Gottlieb, Steven A.; Gregory, Eric Brittain; Heller, Urs M.; Hetrick, James Edward; et al (Jan. 2015)
MILC asqtad QCD SU(3) gauge ensemble; series-a; a=0.043fm; Ls=2.8fm; Nf=2+1; u0.m0=(0.0028,0.014)
- Lattice QCD gauge ensemble:** USQCD/MILC/asqtad/64144f21b746m0018m018a
(/dataexplorer/biblio/1178033)
Aubin, Christopher Alan; Bernard, Claude W.; Burch, Tommy; Datta, Saumen; DeGrand, Thomas Alan; DeTar, Carleton E.; Gottlieb, Steven A.; Gregory, Eric Brittain; Heller, Urs M.; Hetrick, James Edward; et al (Jan. 2015)
MILC asqtad QCD SU(3) gauge ensemble; series-a; a=0.058fm; Ls=3.7fm; Nf=2+1; u0.m0=(0.0018,0.018)
- Lattice QCD gauge ensemble:** USQCD/MILC/asqtad/64144f21b746m0018m018b
(/dataexplorer/biblio/1178034)
Aubin, Christopher Alan; Bernard, Claude W.; Burch, Tommy; Datta, Saumen; DeGrand, Thomas Alan; DeTar, Carleton E.; Gottlieb, Steven A.; Gregory, Eric Brittain; Heller, Urs M.; Hetrick, James Edward; et al (Jan. 2015)
MILC asqtad QCD SU(3) gauge ensemble; series-b; a=0.058fm; Ls=3.7fm; Nf=2+1; u0.m0=(0.0018,0.018)
- Lattice QCD gauge ensemble:** USQCD/MILC/asqtad/56144f21b7465m0025m018
(/dataexplorer/biblio/1178035)
Aubin, Christopher Alan; Bernard, Claude W.; Burch, Tommy; Datta, Saumen; DeGrand, Thomas Alan; DeTar, Carleton E.; Gottlieb, Steven A.; Gregory, Eric Brittain; Heller, Urs M.; Hetrick, James Edward; et al (Jan. 2015)
MILC asqtad QCD SU(3) gauge ensemble; series-a; a=0.058fm; Ls=3.2fm; Nf=2+1; u0.m0=(0.0025,0.018)
- Lattice QCD gauge ensemble:** USQCD/MILC/asqtad/48144f21b747m0036m018a
(/dataexplorer/biblio/1178036)
Aubin, Christopher Alan; Bernard, Claude W.; Burch, Tommy; Datta, Saumen; DeGrand, Thomas Alan; DeTar, Carleton E.; Gottlieb, Steven A.; Gregory, Eric Brittain; Heller, Urs M.; Hetrick, James Edward; et al (Jan. 2015)
MILC asqtad QCD SU(3) gauge ensemble; series-a; a=0.058fm; Ls=2.8fm; Nf=2+1; u0.m0=(0.0036,0.018)
- Lattice QCD gauge ensemble:** USQCD/MILC/asqtad/48144f21b747m0036m018b
(/dataexplorer/biblio/1178037)
Aubin, Christopher Alan; Bernard, Claude W.; Burch, Tommy; Datta, Saumen; DeGrand, Thomas Alan; DeTar, Carleton E.; Gottlieb, Steven A.; Gregory, Eric Brittain; Heller, Urs M.; Hetrick, James Edward; et al (Jan. 2015)
MILC asqtad QCD SU(3) gauge ensemble; series-b; a=0.058fm; Ls=2.8fm; Nf=2+1; u0.m0=(0.0036,0.018)

Result of the search for “Lattice QCD gauge ensemble” from OSTI’s DOE Data Explorer <http://www.osti.gov/dataexplorer/>

Then select on the first ensemble. . .

(/dataexplorer/)

Search DOE Data Explorer for Energy and Science Data

Start new search - Place phrase in "double quotes"



+ Advanced Search

DOE Data Explorer (/dataexplorer/) / Search Results

Dataset: Lattice QCD gauge ensemble: USQCD/MILC/asqtad/64192f21b781m0028m014

Lattice QCD gauge ensemble:

USQCD/MILC/asqtad/64192f21b781m0028m014

Citation Details

MILC asqtad QCD SU(3) gauge ensemble; series=a; a=0.043fm; Ls=2.8fm; Nf=2+1; u0.m0=(0.0028,0.014)

Authors:

Aubin, Christopher Alan [Fordham U.] (/dataexplorer/search/author/"Aubin, Christopher Alan"); Bernard, Claude W. [Washington U. St. Louis] (/dataexplorer/search/author/"Bernard, Claude W."); Burch, Tommy [U. Regensburg] (/dataexplorer/search/author/"Burch, Tommy"); Datta, Saumen [Tata Institute] (/dataexplorer/search/author/"Datta, Saumen"); DeGrand, Thomas Alan [Colorado U., Boulder] (/dataexplorer/search/author/"DeGrand, Thomas Alan"); DeTar, Carleton E. [Utah U.] (/dataexplorer/search/author/"DeTar, Carleton E."); Gottlieb, Steven A. [Indiana U., Bloomington] (/dataexplorer/search/author/"Gottlieb, Steven A."); Gregory, Eric Brittain [Wuppertal U.] (/dataexplorer/search/author/"Gregory, Eric Brittain"); Heller, Urs M. [American Physical Society] (/dataexplorer/search/author/"Heller, Urs M."); Hetrick, James Edward [U. Pacific, Stockton] (/dataexplorer/search/author/"Hetrick, James Edward"); Orinos, Kostas Nikolaou [William-Mary Coll.] (/dataexplorer/search/author/"Orinos, Kostas Nikolaou"); Osborn, James C. [Argonne National Laboratory, ALCF] (/dataexplorer/search/author/"Osborn, James C."); Toussaint, W. Doug [Arizona U.] (/dataexplorer/search/author/"Toussaint, W. Doug"); Sugar, Robert L. [U. C., Santa Barbara] (/dataexplorer/search/author/"Sugar, Robert L.")

Publication Date:

2015-01-01

OSTI Identifier:

1177873

DOE Contract Number:

AC02-06CH11357; SC0012704; AC02-07CH11359; AC02-05CH11231; AC05-00OR22725; PHY07-57035; FG02-91ER40661; FC02-06ER14442; AC1-1053575; PHY-0552234; NSF05-55243; NSF07-57333

Resource Type:

Dataset

Data Type:

Numeric Data

Resource Relation:

Related Information: Bazavov, A., et al., Rev.Mod.Phys. 82 (2010) 1349-1417, DOI: 10.1103/RevModPhys.82.1349; Aubin, C., et al., Phys.Rev. D70 (2004) 094505, DOI: 10.1103/PhysRevD.70.094505; Bernard, C., et al., Phys.Rev. D64 (2001) 054506, DOI: 10.1103/PhysRevD.64.054506

Research Org:

US Lattice Quantum Chromodynamics Collaboration (USQCD);

To view the full OSTI record for this ensemble.

Inspire developers in discussions with OSTI on how to import these records. Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH) interface?

Next steps

- Import asqtad dataset metadata into Inspire-HEP
- Make links among datasets and existing key publications
- Generate publication bibliographies for datasets on NERSC site using Inspire
- Educate authors about citing datasets in future publications
- Register DOIs for MILC HISQ ensembles

Role of the ILDG

Maximize physics productivity through unfettered sharing of lattice datasets.

Inability to easily access deep, extensive information on physics properties in publications can limit the utility of a dataset.

Open access to LQCD results and data to the whole HEP and NP communities

- Regional grids maintain full responsibility for maintaining persistent public access to registered datasets
- Regional grids maintain web landing pages for their datasets
- Each regional ILDG group obtains a unique DOI prefix
- Regional ILDG develops infrastructure needed to request a DOI
- Each regional ILDG sets standards for their DOI “namespace”
- Regional ILDG representatives become points of contact for collaborations seeking DOIs