

DOE Final Report

Project: DE-FC02-06ER41441

Massimo Di Pierro

March 15, 2012

DOE Award Number: DE-FC02-06ER41441
Name of Recipient: Massimo Di Pierro
Principal Investigator: Massimo Di Pierro
Consortium Name: USQCD Collaboration

1 License

This report and its content can be distributed without limitation.

The software developed under this grant is already posted online under version control system, always under an OSI approved Open Source License. The specific licence is different for different parts of the code and details about the licenses are included with the code.

2 Executive Summary

Our research project is about the development of visualization tools for Lattice QCD. We developed various tools by extending existing libraries, adding new algorithms, exposing new APIs, and creating web interfaces (including the new NERSC gauge connection web site). Our tools cover the full stack of operations from automating download of data, to generating VTK files (topological charge, plaquette, Polyakov lines, quark and meson propagators, currents), to turning the VTK files into images, movies, and web pages. Some of the tools have their own web interfaces. Some Lattice QCD visualization have been created in the past but, to our knowledge, our tools are the only ones of their kind since they are general purpose, customizable, and relatively easy to use. We believe they will be valuable to physicists working in the field. They can be used to better teach Lattice QCD concepts to new

graduate students; they can be used to observe the changes in topological charge density and detect possible sources of bias in computations; they can be used to observe the convergence of the algorithms at a local level and determine possible problems; they can be used to probe heavy-light mesons with currents and determine their spatial distribution; they can be used to detect corrupted gauge configurations. There are some indirect results of this grant that will benefit a broader audience than Lattice QCD physicists and explained below.

3 Achievements and Comparison of Goals

The goal of this research project is the creation of a visualization toolkit that could be used to aid physicists in the analysis of data from lattice QCD computations. This goal has been successfully achieved as described below.

Links to code, images and video can be found at:

<http://latticeqcd.org>

The original grant proposal stated:

Lattice QCD computations comprise multiple steps, creating very large datasets, but the final result is typically encompassed in a small set of numbers with the analysis performed in an automated way. While an automated procedure may be beneficial in efficiency, the ability to visualize the data being analyzed is important both as an aid to the analysis, and as a means of acquiring insight into the physics. [...]

Crucial to the success of the graphics-visualization initiative will be a close collaboration between physicists to devise and interpret visualization of physically important quantities, and computer scientists to provide the appropriate visualization toolbox. Questions that visualization might address are many: can we understand how flux-tube formation observed with infinitely heavy quarks extends to hadrons where one or more of the quarks is light; what is the distribution of charge within a nucleon; can we display the distribution of spin and magnetism within a hadron? In the longer term, can we visualize the interactions of hadrons? Currently, no general-purpose package is available tailored to the display of lattice data. Thus a software package will be developed with a general GUI capable of reading a set of four-dimensional lattice quantities, and taking their ensemble average; performing a projection into a real four-dimensional vector; interpolating the 4-D vector into a continuous four-dimensional field; taking

three-dimensional slices of a four-dimensional field; displaying the data using density plots, iso-surfaces, and 2-D projections; and displaying the evolution of data, both in simulation time for four-dimensional quantities, and as the evolution of three- and two-dimensional slices in the remaining coordinates. The software will support two types of plug-ins: type-1 plug-ins that perform specific physics measurements and output a real 4-D vector, and type-2 plug-ins that take the interpolated 3-D field and generate specific types of plots.

Most of the research underlying this project will consist of identifying a set of physical measurements suitable to be implemented as type-1 plug-ins. The visualization techniques for the type-2 plug-ins are very similar to standard techniques used for representation of 3-D geophysical data and, when possible, we will incorporate existing libraries into the development of our plug-ins.

The system will be developed in C++ and take advantage of existing graphics and visualization libraries such the Trolltech QT libraries and the Visualization Tool Kit (VTK) library. The plug-ins will be callable from C or C++ code conforming to the QCD API, and will form another component of our Level 4 QCD Toolbox. The system will be capable of reading datasets in the SciDAC/ILDG format and the MILC format.

[...]

DePaul University will lead the design and development of a visualization tool for lattice QCD. Work will be done in collaboration with physicists involved in the project and with computer scientists at the University of North Carolina. The goals for the first year of the project are to identify and catalog the types of datasets to be visualized, identify appropriate smoothing and visualization algorithms, and develop a prototype interface. In subsequent years, plugins will be developed to read in the various types of datasets produced in lattice QCD simulations, and tools for manipulating the data in increasingly sophisticated ways will be created. A total of 1.08 FTE per year is budgeted for this effort.

A QCD physics toolbox will be constructed which will contain sharable software building blocks for inclusion in application codes, performance analysis and visualization tools, and software for automation of physics work flow. New software tools will be created for managing the large data sets generated in lattice QCD simulations, and for sharing them through the International Lattice Data Grid consortium.

Our basic toolkit consists of two parts. The first part has been implemented in the form of C++ libraries which are now included within the FermiQCD toolkit which is part of the USQCD Software Suite. These API allow the project of arbitrary fields into 3D and 4D scalar

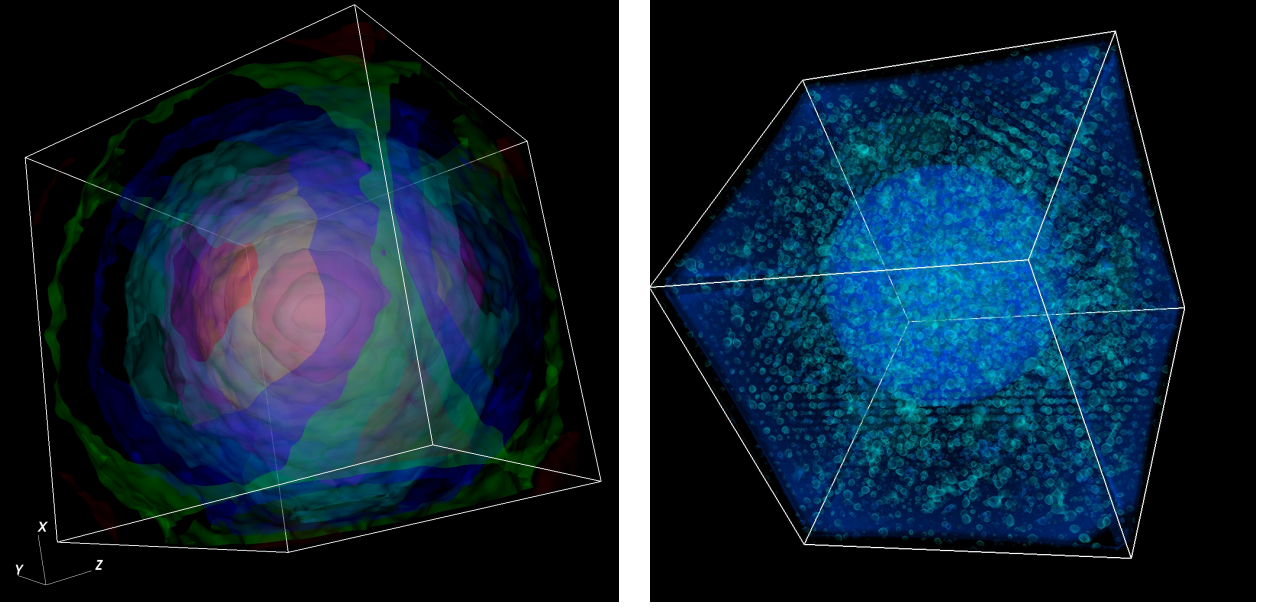


Figure 1: Example Images. The left one shows iso-surfaces for a quark wave function. The right one shows the density of HMC hits in presence of a lump of topological charge.

fields with can be visualized using open source toolkits like VisIt, Paraview, and MayaVi. The toolkit includes algorithms to project topological charge density, plaquette, Polyakov line components, quark propagators, meson propagators, and current insertions. It provides an API to create other custom operators and quark contractions and project/visualize them too. We also modified the minimum residue and the stabilized bi-conjugate inverter so that it is possible to observe the spatial effect of this algorithms and visualize their convergence for arbitrary sources. Our choice of the FermiQCD toolkit is motivated by the critical need to be able to perform visualizations for arbitrary $SU(N)$ gauge groups. FermiQCD is the only lattice QCD code, part of the USQCD Software Suite, that at this time supports arbitrary gauge groups.

The second part of our toolkit consists of a collection of Python programs that interface with the C++ programs and make the system more accessible to scientists by implementing a typical workflow. Specifically we developed 7 different programs.

The first program (as required by the original grant proposal) has the ability to convert MILC/ILDG data (as well other data formats) into the format required for visualization. On top of that, as described below, the same program has the ability to download gauge configurations from the NERSC repository, which is the largest public repository of gauge

configurations in the United States.

The second program is the main interface to the C++ algorithms. The program provides command line options to run physical algorithms such as the computation of the topological charge density, plaquette, Polyakov, lines, quark propagators, meson propagators, currents, 4-quark operators. The program downloads, compiles and runs the requested algorithms. Each algorithm provides the option to save the computation steps in the VTK format for visualization. Not all the C++ algorithm are accessible via this interface and some requires explicit programming. Yet the provided code and documentation should be a sufficient example for the scientists to write their own customized code for other particular cases.

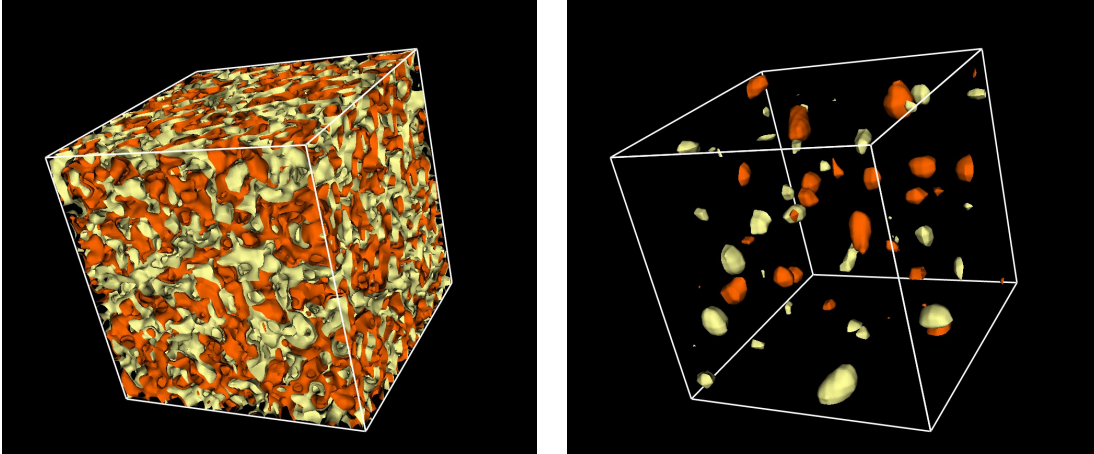


Figure 2: Example Images: They show the visualization of the topological charge at different cooling steps.

The third program performs the tasks of extracting information from the VTK files generated by the previous step, resampling them (to achieve better resolution in visualization), interpolating them (to make smooth visualizations and movies) and generating VisIt scripts. The VisIt visualization toolkit was developed by the Lawrence Livermore National Laboratory and it is a critical part of our workflow. It can be accessed via a GUI or programmatically via Python. Using a GUI to automate a workflow and process many files at once is not practical. It must be done programmatically. Our program generates VisIt scripts using meta-programming techniques so that QCD physicists do not have to code. By running these scripts scientists can directly obtain images and movies without programming.

The fourth program we develop allows to convert 3D VTK files into interactive web pages. The program opens the VTK files, identifies optimal thresholds for the iso-surfaces, computes said iso-surfaces and generates a 3D representation of the polygons in JavaScript. The output

consists of static HTML files which embed the 3D objects. They can be visualized with any browser and the viewer can rotate them with the mouse. While VisIt is a general purpose tool is very powerful which allows many more manipulations of the data, the possibility of generating 3D interactive web pages opens the possibility for scientists to view the data without installing VisIt and to publish the data on the web for other people to see.

The other programs we created are beyond the scope of the original grant proposal but we felt they were necessary and part of a broader interpretation of the concept of visualization. In fact, not all visualizations are spatial visualizations. There is other information that is important to visualize, which often takes the forms of simple 2D plots but often is not looked at because of the extra work involved in doing so. Examples are autocorrelations between physical measurements on different gauge configurations, moving averages, distributions of bootstrap samples. Our other programs serve this purpose. One of the programs read the output of typical QCD algorithms, extract the numerical results for each gauge configuration, and combines them into a bootstrap analysis. The user can specify the expression to bootstrap using command line arguments without programming. The code generates CSV files storing data for the intermediate steps of the computation. The other two programs can read those CSV files and generate plots and fits from them automatically. CSV files can also be read by many third party analysis and visualization programs.

Consider for example the computation of a four-quark matrix element. It involves the computation of a two-point and a three-point correlation function for each gauge configuration and their bootstrap analysis. Our programs can perform this analysis in the standard way but they also generate autocorrelation plots for each two-point and three-point correlation function, moving averages for the ratio on different time-slices, and bin the distribution of the bootstrap samples.

Those programs have been documented in long technical document attached to this report [1].

The main obstacle to this research has been accessibility. Visualization is indeed useful not but the way QCD physicists normally approach lattice QCD computations. We have therefore put lots of extra work in making our programs accessible by creating web interfaces that could simplify the task. Another obstacle is that no computing time was allocated to this project. While this did not prevent us from achieving the task of developing the tools, it did not allow us to move beyond the original stated goal and utilize the tools for obtaining more abitious scientific results which would have been computing intensive.

Nevertheless we have used our toolkit to produce scientific results. Specifically we collaborated with Chulwoo Jung at Columbia University, Mike Clark and Richard Brower at Boston University and looked at the autocorrelation of topological charge density over short HMC trajectories [15]. It is a well known problem that global topological charge has a long autocorrelation. We found that the local topological charge instead has very short autocorrelation

and therefore there is no measurable bias in production gauge configurations.

We also were able to observe the effect of a single instanton on a quark propagator and how its presence gives mass to the quark by increasing the exponential fall off of the propagator [1].

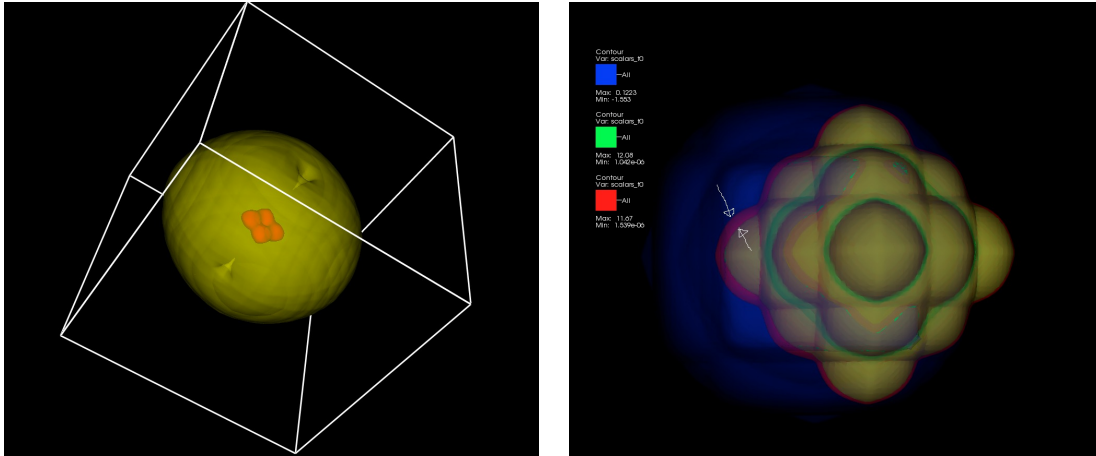


Figure 3: Example Images: left left image shows the density of a current inserted between to meson operators. The right image shows the shrinking of a quark propagator (red) in presence of a localized topological charge (blue).

Anyway, that is beyond the scope of the current grant and more visualizations will be done in the future as computing time becomes available.

In our original proposal we stated that our tools would have a GUI based on the Qt toolkit. In the very early stages of our project we have revised that decision. On the one side Nokia, owner of Qt, decided to cut support for the library. On the other side it became evident that desktop GUI have become an obsolete technology giving way to modern web based interfaces. We have therefore put lots of extra work in this direction and we have created three web applications.

The first web application (nersc) [3] was developed in collaboration with the National Energy Research Scientific Center (NERSC) to replace their previous interface to the lattice QCD archive known as “gauge connection”. The new system allows searching for gauge configuration, visualize statistics, and collaborate online by editing metadata in a wiki format. The system also allows batch downloads of the data using the program we described above. We have used our toolkit to process many ensembles from the NERSC archive and generated movies of the topological charge density. This program is designed to be very general purpose and it can be used by communities other than Lattice QCD to publish their data online.

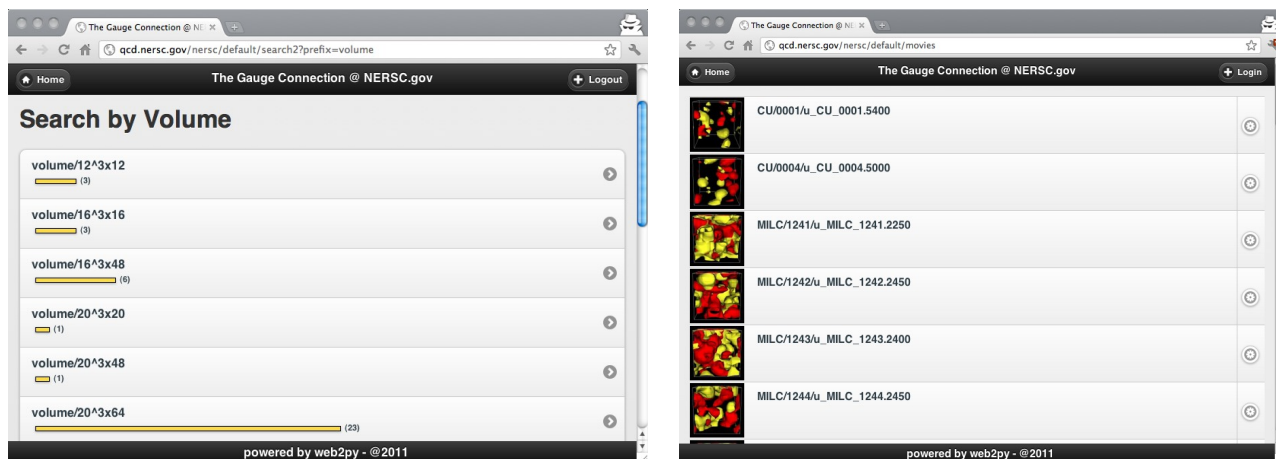


Figure 4: Screenshots from the NERSC “gauge connection” web site.

The second web application (vis) [6, 8] provides an interface to the algorithms and allows upload of gauge configurations and schedule visualization algorithms to run in background. The system provides a web interface to the Portable Batch System (PBS) and to VisIt and. It can schedule both computations and visualizations. The results are displayed in a web page.

The third web application (mc4qcd) [7, 9] is an interface to our analysis and plotting tools. It allows users to upload the log file result of a physics runs, to extract data from it using pattern matching, and to perform bootstrap analysis. The results of the analysis are stored and published online together with the replative plots (including autocorrelations, moving averages, distribution of bootstrap samples, and fits). Scientists in a group can track results and collaborate online by sharing data and comments.

In order to develop these tools, in the early stages of the project we have developed a set of libraries for creating online scientific applications called web2py. This project is not part of the goal of this grant but it turned out to be an important and necessary component to replace the obsolete GUI concept with the modern web based paradigm. This project took a life on its own and found applications beyond this physics project. It was released open source and it is now used by thousands of users and businesses worldwide. It won the Bossie Award for “best open source software development tool” in 2011 and the InfoWorld Best Technology of the Year Award in 2012. This provides an example of unexpected broader impact of DOE funded research. Although we consider this very important and we are proud of the result, since it falls outside the scope of the original grant, we omitted references to it in the rest of this report. Yet references are available upon request.



Figure 5: Screenshots from the VIS and the MC4QCD web applications respectively.

In 2010 we also contributed organize the 6th High End Visualization workshop in Obergugl, Austria.

Some of the visualization created with our tools were used for the opening video for the Lattice 2011 conference: <http://vimeo.com/25242353>

3.1 Summary of Code Created

The code written as part of this grant is published in the following repositories:

- <http://code.google.com/p/qcdutils/> It is the main toolkit, documented in [1].
- <http://code.google.com/p/fermiqcd/> This is a pre-existing C++ library for lattice QCD computations part of the USQCD Software Suite. The core visualization algorithms have been included here and distributed together. They are accessible via QCDUTILS.
- <http://code.google.com/p/nersc-data-publisher/> This is code behind the new NERSC “gauge connection” web interface. Developed in collaboration with James Hetrick (University of Pacific) and David Skinner (NERSC).
- <http://code.google.com/p/qcdvis/> This is a web interface to the visualization algorithms.

- <http://https://launchpad.net/qcdmc> This is a web interface to our analysis and plotting tools.

3.2 Published Web Sites

- <http://latticeqcd.org> This is the main front-end where we have published links to the code and some of our visualizations. More will be published as our tools are put into production. This site includes web interfaces to VIS and MC4QCD.
- <http://qcd.nersc.gov> This is the new NERSC “gauge connection” archive (developed in collaboration with NERSC). It also stores some videos created using our tools.
- <http://tests.web2py.com/ildg> This is a new proposed web site for the International Lattice Data Grid. It provides an interface for searching lattice QCD data in a visual way.

3.3 Fostered Collaborations

During this research project we have collaborated with Prof. James Hetrick from the University of Pacific and David Skinner at NERSC to re-build the new “gauge connection” web site.

We collaborated with Prof. Werner Berger from the Center for Computation and Technology, Louisiana State University and together we organized the 6th High End Visualization Workshop.

We utilized the VisIt software created by the Lawrence Livermore National Laboratory. Although we did not interact directly with the authors we interacted indirectly by using various online resources generated for that project.

Finally we interacted with the rest of the USQCD and the ILDG collaborations, from which we received constant feedback and suggestions.

3.4 Personnel

This grant has funded the PI and many graduate students:

- Nate Wilson
- Yaoquan Zhong
- Brian Schinazi

- Tony Garcia
- Vincent Havery

3.5 Published Papers

In the following bibliography we list all the papers published by the PI and supported directly by this research grant. We omitted papers and books published by the PI on other topics not directly related to the grant scope.

References

- [1] M. Di Pierro “qcdutils“ arXiv:1202.4813 [hep-lat]
- [2] M. Di Pierro “Improving Non-Linear Fits“ arXiv:1202.0988 [cs.NA]
- [3] M. Di Pierro, J. Hetrick, S. Cholia, D. Skinner “Making QCD Lattice Data Accessible and Organized through Advanced Web Interfaces“ PoS **LAT2011** (2011) arXiv:1112.2193 [hep-lat]
- [4] E. T. Neil *et al.* [for the Fermilab Lattice Collaboration and for the MILC Collaboration], “B and D meson decay constants from 2+1 flavor improved staggered arXiv:1112.3978 [hep-lat]
- [5] A. Bazavov *et al.* [Fermilab Lattice and MILC Collaborations], “B- and D-meson decay constants from three-flavor lattice QCD,” arXiv:1112.3051 [hep-lat]
- [6] M. Di Pierro, Y. Zhong and B. Schinazi, “Vis: Online analysis tool for lattice QCD,” PoS **LATTICE2010**, 326 (2010).
- [7] M. Di Pierro and Y. Zhong, “Analysis and visualization tools for Lattice QCD,” PoS **LAT2009**, 038 (2009).
- [8] M. Di Pierro, “Vis: QCD workflow and visualization tool“ in Proceedings of the 6th High End Visualization Workshop, Dec 8th - 12th, (2010), Obergurgl (Austria), Lehmanns Media, ISBN 978-3-86541-361-1
- [9] M. Di Pierro, Y. Zhong and B. Schinazi, “mc4qcd: Online Analysis Tool for Lattice QCD,” PoS A **CAT2010**, 054 (2010) arXiv:1005.3353 [hep-lat]

- [10] C. Bernard *et al.* [Fermilab Lattice Collaboration and MILC Collaboration], “Tuning Fermilab Heavy Quarks in 2+1 Flavor Lattice QCD with Application to Phys. Rev. D **83**, 034503 (2011) arXiv:1003.1937 [hep-lat]
- [11] A. Bazavov *et al.* [Fermilab Lattice and MILC Collaborations], “The Ds and D+ Leptonic Decay Constants from Lattice QCD,” PoS **LAT2009**, 249 (2009) arXiv:0912.5221 [hep-lat]
- [12] T. Burch *et al.*, “Quarkonium mass splittings in three-flavor lattice QCD,” Phys. Rev. D **81**, 034508 (2010) arXiv:0912.2701 [hep-lat]
- [13] J. A. Bailey *et al.* [Fermilab Lattice Collaboration and MILC Collaboration], “Progress on charm semileptonic form factors from 2+1 flavor lattice QCD,” PoS **LAT2009**, 250 (2009) arXiv:0912.0214 [hep-lat]
- [14] T. Burch *et al.*, “Quarkonium mass splittings with Fermilab heavy quarks and 2+1 flavors of PoS **LAT2009**, 115 (2009) arXiv:0911.0361 [hep-lat]
- [15] M. Di Pierro *et al.*, “Visualization as a tool for understanding QCD evolution algorithms,” J. Phys. Conf. Ser. **180**, 012068 (2009).
- [16] C. Bernard *et al.*, “Visualization of semileptonic form factors from lattice QCD,” Phys. Rev. D **80**, 034026 (2009) arXiv:0906.2498 [hep-lat]
- [17] C. Bernard *et al.*, “B and D Meson Decay Constants,” PoS **LATTICE2008**, 278 (2008) arXiv:0904.1895 [hep-lat]
- [18] J. A. Bailey *et al.*, “The B \rightarrow π l $\bar{\nu}$ semileptonic form factor from three-flavor lattice QCD: Phys. Rev. D **79**, 054507 (2009) arXiv:0811.3640 [hep-lat]
- [19] C. Bernard *et al.*, “The Anti-B \rightarrow D* l anti- ν form factor at zero recoil from three-flavor Phys. Rev. D **79**, 014506 (2009) arXiv:0808.2519 [hep-lat]
- [20] M. Di Pierro, “A Visualization Toolkit for Lattice Quantum Chromodynamics“, in Proceedings of the 4th High End Visualization Workshop, June 18th - 22th, (2007), Obergurgl (Austria), Lehmanns Media, ISBN 9783865412164
- [21] C. Bernard *et al.* [Fermilab Lattice, MILC and HPQCD Collaborations], “The decay constants $f(B)$ and $f(D^+)$ from three-flavor lattice QCD,” PoS **LAT2007**, 370 (2007).

- [22] R. Todd Evans, E. Gamiz, A. X. El-Khadra and M. Di Pierro, “A Determination of the $B_0(s)$ and $B_0(d)$ mixing parameters in 2+1 lattice PoS **LAT2007**, 354 (2007) arXiv:0710.2880 [hep-lat]
- [23] M. Di Pierro, “Visualization for Lattice QCD“ PoS **LAT2007**, 331 (2007)
http://pos.sissa.it/archive/conferences/042/031/LATTICE2007_031.pdf