# High Energy Physics
# Data Science Toolkit Development

## 1 PI and Co-PI Information


**PI:** Green, Chris, Dr.
Fermi National Accelerator Laboratory
PO Box 500
Batavia IL 60510-0500
greenc@fnal.gov

**Co-PI:** Sehrish, Saba, Dr.
Fermi National Accelerator Laboratory
PO Box 500
Batavia IL 60510-0500
ssehrish@fnal.gov


## 2 Project Summary

### 2.1 Executive Summary


A typical High Energy Physics (HEP) experiment has several categories of computing workflow: production of simulation, collection of real data, physics object identification, and many unique end-user analyses. Even the smaller running and imminent experiments will have total collected and simulated data volumes in the 10 TiB to $\sim$ PiB range, with the attendant challenges for processing CPU demands, IO, networking and storage. Simulated data are usually at least ten times the volume of collected data, and both are processed several times to achieve publication quality physics objects. End-user analyses are run hundreds to thousands of times using these physics objects over the lifetime of an experiment.

By using new-to-the-field HPC facilities and technologies to carry out specific representative HEP tasks, we intend to develop tools and libraries which will allow these tasks to be carried out with maximum efficient use of HPC resources and state-of-the-art data science technologies, and thereby evolve the field's traditional high-throughput computing (HTC) model going forward.

Using advanced features available in the most recent language standard, we have developed a C++ library[1] suitable for writing tabular data in HDF5 format. This makes it available for iterative analysis, taking advantage of the caching and large memory spaces of the ALCF environment. This format must be evaluated and tuned in a big-data HPC environment. We will execute several representative workflows from different experiments, starting from the HDF5-format data, and their performance and scaling characteristics studied. As we develop these applications and workflows, other libraries and tools will be developed to be used by a wide range of HEP experiments who may not have the resources to design such applications from scratch.


[1] The code is available at https://bitbucket.org/fnalscdcomputationalscience/hep_hpc.

## 2.2 Benefit to Community

The field of HEP is looking for ways of leveraging HPC facilities to satisfy more of its computing needs. The statistical independence of the basic units of HEP data (the "event") has led to the traditional HTC model which relies on sequential processing, data analysis and persistence formats: communication between parallel computing tasks is generally not utilized outside a few, "niche" applications within the pantheon of algorithms produced by the community. HEP workflows often have a low calculation to data ratio ("decision-bound"), and many algorithms are not designed to be readily vectorizable.

Relatively recently, increases in the complexity of each event due to factors like collider luminosity or detector complexity have clashed with constraints such as memory or per-unit processing time to mitigate toward higher levels of parallelism, as have new techniques such as deep learning and "image-style" analysis. A set of tools enabling relatively straightforward use of HPC facilities for several common classes of computing task within HEP will give experiments without resources or expertise for computing R & D a lower barrier to leveraging HPC computing resources for their workflows. These tools will also improve the ability of HEP experiments to use externally developed tools such as HDF5, deep learning toolkits and a range of parallel interactive analysis tools. These efforts will ameliorate the deficiencies of the current HTC model.

## 2.3 Impact Statement

HEP experiments study the tiniest constituents of matter and how they interact. HPC systems allow more complex analyses of the huge volumes of data we collect on elusive and poorly understood particles like neutrinos, better to answer questions about our Universe such as, "How did it form and expand?"

## 2.4 Technology Summary

We are working to develop an HPC toolkit that complements the more HEP-traditional, HTC-focused and well-established software system for processing a wide range of HEP simulated and experimental data. This toolkit will comprise components:

- facilitating production of HEP data in HDF5 format for end-user analysis applications,
- facilitating the reading of the data into parallel analysis applications, and
- implementing common HEP analysis tasks using standard HPC technology underpinnings.

We will work with the data handling community to facilitate the physical movement of data between traditional HEP data storage systems and HPC systems.

There are aspects of traditional HEP workflows that employ tabular data structures, such as certain types of raw digitization data or end-user analysis. We have developed a preliminary version of a toolkit to facilitate production of HDF5 files in our common HEP event processing framework (*art*)[2] within standard HEP workflows with a standardized

[2] *art* is described at http://art.fnal.gov.

structure that is conducive to highly parallel, column-wise analyses. We will refine this conversion toolkit and the organization of the data it produces for scaling and performance properties while developing HPC-centric implementations of a selection of HEP-typical analysis tasks.

Additional toolkit components as itemized above will be developed by factorising and packaging code and workflows for use by different experiments with common needs. We are already developing tools to facilitate reading of structured data directly into R, Spark and numpy. We will extend these tools to deal with data sets of the scale involved in the example applications, as described below. We will also explore options for parallel training of deep-learning systems using existing representative analysis tasks and develop tools to facilitate this for common HEP applications.

## 2.5 Application Summary

### 2.5.1 Application Requirements

**Data Facilities (Theta)** We require a fully C++14-compliant compiler and standard C++ library. If necessary, we will provide our own software stack, including the compiler, Intel Threading Building Blocks, and HEP-specific toolkits and libraries. We will need MPI but can link to the system's optimized MPI C library.

**Analysis Suite (Theta, Cooley, Sage)**

- HDF5 (>= v1.10.1) (C) MPI (C), Python (>=2.7), mpi4py, numpy, scipy, h5py, pandas (all).
- Caffe2 (Cooley).
- Spark (>= v2.0) with Scala, Java bindings for HDF5, Dask (Sage).
- OpenMP and OpenACC support in the C++ compiler.
- Eigen, Blaze and BLAS for use in C++.

### 2.5.2 Application Description

The technology will be developed using several applications, described below, for representative workflows for HEP experiments. In all cases the input data shall be converted from their native format to HDF5 for maximum flexibility for use with existing HPC tools and libraries.

An analysis[3] of Dark Matter data from the CMS experiment at CERN involves the application of several multi-dimensional filters (based on event features) to the data using Spark as our analysis framework and Scala API to implement the application with Python / MPI as alternative options. Preliminary performance and scaling results have been obtained on Edison and Cori at NERSC: we intend to study and tune the performance, generalizing the tools developed for use with other, similar workflows from other experiments.

The raw digitized time sampling data from the LArIAT detector at Fermilab is amenable to representation in HDF5 format and is representative of similar data from other neutrino

[3]The code is available at https://github.com/sabasehrish/spark-hdf5-cms.

experiments. We will study the HPC performance of a preliminary processing of the LArIAT raw data (about 5 TiB) to calibrated waveforms, and hit finding on the calibrated waveforms, turning the time-based information into spatial information. We will compare the speed and scalability of an MPI and numpy-based solution with traditional HEP reconstruction software and hardware.

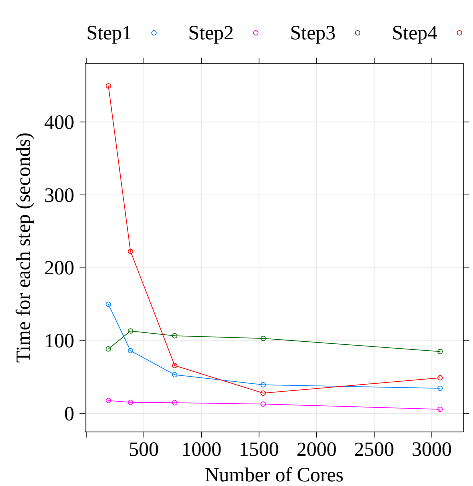
We will develop and characterize convolutional neural nets for discriminating particle species in LArIAT. LArIAT has both simulated data and testbeam data with independent particle species identified by detectors upstream of the liquid argon TPC. In order to fully understand the residual differences between the simulated data and real data, and to improve the realism of our simulations, we will explore the use of generative adversarial networks which show us how to close the gap between existing simulation and real data. We will explore the use of multiple GPUs to accelerate the training.

The protoDUNE experiment at Fermilab, whose detector characteristics are similar to those of LArIAT, will collect approximately 2.5 PiB split between 24 different physics processes between May and June of 2018. Using MPI and numpy, we will implement and refine particle identification techniques, and improve the modeling and measurement of hadronic energy response and pion interaction final state multiplicities using of order 100 TiB of simulated data.

The Muon g-2 experiment at Fermilab has been undertaken to produce a high statistics, high precision measurement of the muon anomalous magnetic moment. To make the measurement, the energy of the positrons that have decayed from stored muons must be determined to a high degree of precision from 1,296 detector elements sampled by waveform digitizers at 800 MHz. Precise analysis of these data to the required level of 140 ppb will require a multiplicity of techniques such as direct or template fitting, and/or machine learning, and a large and realistic simulation sample in addition to many passes on the data as the analysis is refined. For this project, we plan to produce of order 10 TiB simulated data to test different fitting techniques on HPC in order to build the best analysis workflow for Muon g-2.

### 2.5.3 Application Performance

The graph opposite shows the scaling performance of just one of the representative use cases (CMS Dark Matter analysis using Spark) at NERSC. It represents the time taken to read and analyze 360 million events (250 GiB). The first three steps demonstrate the time to read and prepare data for analysis in memory. The fourth step shows the analysis time to create a histogram of the transverse momentum of interesting electrons in the events, including the application of several complex filtering operations with a few global synchronizations and several transformations on the data. Different steps exhibit different or no scaling. We need to work on the read performance.



The performance of the other workflows described herein will be measured and improved as the technology is developed and refined.

### 2.5.4 Application Development Needed

The existing, non-parallel LArIAT waveform calibration application will be adapted to run on Theta using numpy / MPI, taking as input the HDF5 raw digitized sample data converted from legacy format. The results will be verified against waveforms produced with the traditional application, and throughput, scaling and resource use will be compared with same. The HDF5 input and output data structures will be considered as a performance variable and optimized where appropriate. Auxiliary calibration data will be staged where appropriate using local SSD. The size of the raw data is such that an event per hardware thread will fit within the MCDRAM of a single node; since each event can be processed independently 256-way parallelism on the node is straightfoward. Since the output calibrated waveforms are an intermediate stage for the production of physics results, a final application would stage these on SSD to remain available for multiple subsequent final analysis workflows.

Existing LArIAT machine learning particle ID analyses are limited by the size of the training samples that can be used, and the available computing resources for training. We will make use of the HDF file storage already developed. Adapting to the use of mini-batch training will allow the use of multiple nodes in parallel.

Waveform processing for protoDUNE is similar to that for LArIAT, but on a larger scale. A raw data for a single protoDUNE event is $\sim$270 MiB, comprising 2560 wires; this is about 100 times the size of the similarly structured LArIAT data. While protoDUNE events are too large to devote a single hardware thread to each event, the fact that each wire can be processed independently provides options to be explored for node-level parallel processing of wires within an event. Given the expected total size of the experimental data set, scaling studies will be important in determining the optimium configuration for a full analysis.

The Muon g-2 waveform processing is similar to that of LArIAT and protoDUNE, except that the data come from crystal calorimeters rather than TPC wires. The application will build on tools developed for the previous applications, making use of SSD staging of intermediate data to facilitate evaluation and refinement of fitting algorithms.

The conversion toolkit is designed to run in parallel on multiple HEP data files. HDF5 is natively supported by MPI, which is our primary parallelization approach. If a postdoc's time were allocated, we would implement MPI I/O parallel writing to the same HDF5 and study the throughput relative to the many-independent-files approach.

## 3 Estimate of Resources Requested

**Theta Resources:** The development and tuning of the toolkit on Theta will be based on the LArIAT waveform analysis, the protoDUNE particle identification, and the Muon g-2 analysis studies described above.

The breakdown of expected core-hour usage on Theta is:

- Evaluate different data organization in HDF5 files for user application (500K hours)
- Profile and tune the MPI (Python) applications (500K hours)
- Scaling studies (2M hours)

We expect to need 200–300 TiB of Lustre storage space. Data will be imported to Lustre from Fermilab.

**Cooley Resources:** The work on Cooley is based on the LArIAT particle classification application.

This project requires 2000 node-hours for development of training algorithms. This requires interactive use of a single node during development; the time estimate is based on a year's worth of developer effort.

Testing of the algorithms, by training with a million event sample, will be done approximately every two weeks. This would require a total of approximately 2500 node-hours.

Data storage requirements are on the order of 1 TiB.

**Sage Urika-GX Cluster Resources:** We will evaluate Spark for HEP analysis on Sage. We already have an implementation of an analysis use case from the CMS experiment on-going dark matter search. Our input data is in HDF5 format and the analysis is implemented using Spark/Scala. In our experience using Spark on Edison and Cori at NERSC, we have observed good scaling but not performance. We are planning to tune our HEP use case implementation in Spark to optimally run on Sage. We plan to perform the following tasks on Sage:

- Evaluate read performance of the Spark and HDF5 reader and optimize
- Evaluate analysis time and determine bottlenecks in the implementation

Our goal is to determine whether Spark is a viable candidate for HEP analysis running on hardware designed to support big data frameworks. We will need 200K hours for the performance evaluation and optimization study. Our data set is approximately 0.5 TiB in size. We are interested in exploring the use of SSD for faster loading of data into memory for interactive analysis.

## 4 Other Collaborations

We are collaborating with the HDF Group to develop the high performance I/O capable toolkit for our experiments. We have run the prelimanry implementation of the CMS dark matter use case in Spark on Edison. The scaling results are included in this proposal. We are currently collaborating with NERSC to optimize the read performance of our HDF5 Spark reader on Cori Phase I.

## 5 Project Team Members

### 5.1 Names and Levels of Effort

- Chris Green: 40%

- Saba Sehrish: 20%
- Jim Kowalkoswki: 10%
- Adam Lyon: 15%, directing efforts of Muon g-2 postdoc(s) and student(s).
- Andrew Norman: 15%, directing efforts of protoDUNE postdoc(s) and student(s).
- Marc Paterno: 10%
- Jason St. John: 15%, directing efforts of LArIAT postdoc(s) and student(s).

**Chris Green**

Computer Science Researcher
Scientific Computing Division
Fermi National Accelerator Laboratory
Batavia, IL 60510-0500
'Phone: 630-840-2167
`greenc@fnal.gov`

**Education and Training**

- University of Liverpool, Ph.D. in High Energy Physics, 1998.
- University of Liverpool, B.Sc. in Physics and Mathematics, 1994.

**Research and Professional Experience**

- Computer Science Researcher, Fermi National Accelerator Laboratory, Computing Division, since 2010.
- Computer Professional, Fermi National Accelerator Laboratory, Computing Division, 2006–2010.
- Postdoctoral Fellow, Lousiana State University, 2006.
- Postdoctoral Fellow, Los Alamos National Laboratory, 2003–2006.
- Postdoctoral Fellow, University of California Riverside, 2002–2003.
- Consultant, International Business Systems, 2000–2001.
- Postdoctoral Fellow, Purdue University, 1997–2000.

**Synergistic Activities**

- Senior researcher on Fermilab LDRD-2016/010, "Preparing HEP reconstruction and analysis software for exascale-era computing".
- Collaboration with the HDF Group on use of the HDF5 libraries for storage of experimental HEP data.
- Developer on the *art* event processing framework, used by most present and future neutrino and muon experiemnts at Fermilab, and others.
- Previously User Liaison for the Open Science Grid, including working with ITER Fusion simulation group on HPC facilities.

**Doctoral Advisors**

- Dr. P. P. Allport, University of Liverpool (joint).
- Dr. M. Tyndel, Rutherford Appleton Laboratory (joint).

**Postdoctoral Advisors**

- Dr. W. Metcalf, Louisiana State University.
- Dr. W. C. Louis III, Los Alamos National Laboratory.
- Dr. G. Van Dalen, University of California Riverside.
- Dr. D. Bortoletto, Purdue University.

Saba Sehrish
Computer Science Researcher
Fermi National Accelerator Laboratory
ssehrish@fnal.gov

**Research and Professional Experience**

- **Fermi National Accelerator Laboratory - Computer Science Researcher** *(Jul 2013 - Present)*
Member of Tools and Advanced Computing group at Scientific Computing Division of Fermilab. Working on different projects for HEP software frameworks: 1) LArSoft - a toolkit for Liquid Argon experiments, 2) Big Data Project - exploring big data tools and HPC for HEP analytics, and 3) CMSSW - software for the CMS experiment. Worked on a web portal for analysis of cosmology data (PDACS) and a level 3 framework for LSST.

**Education and Training**

- **Postdoctoral Research Fellow**, Northwestern University *(Sep 2010 - Jun 2013)*
Research activities include designing and implementing a Data Model I/O Library for exascale science (DAMSEL), optimizing parallel I/O libraries (MPI-IO) and big data tools for scientific applications (using MapReduce/Hadoop)
Mentor: Alok Choudhary

- **PhD, Computer Engineering**, University of Central Florida*(2010)*
Dissertation Title: *Improving Performance and Programmer Productivity for I/O-Intensive High Performance Computing Applications.*
Advisor: Jun Wang

- **MS, Computer Engineering**, University of Central Florida *(2007)*
Research Project: *Implemented a directory memory predictor and pre-fetcher for distributed shared memory multiprocessors.*
Advisor: Mark Heinrich

- **BS, Computer Systems Engineering**, GIKI, Pakistan *(2003)*

**Selected Publications**

- Saba Sehrish, Jim Kowalkowski, Marc Paterno. *Spark and HPC for High Energy Physics Data Analyses.* The 3rd IEEE International Workshop on High-Performance Big Data Computing (HPBDC), with the 31st IEEE International Parallel and Distributed Processing Symposium (IPDPS 2017).

- Saba Sehrish, Jim Kowalkowski, Marc Paterno. *Exploring Performance of Spark for a Scientific Use case.* The 2nd IEEE International Workshop on High-Performance Big Data Computing (HPBDC), with the 30th IEEE International Parallel and Distributed Processing Symposium (IPDPS 2016).

- Saba Sehrish, Jim Kowalkowski, Marc Paterno. *Exploring Performance of Spark for a Scientific Use case.* The 2nd IEEE International Workshop on High-Performance Big Data Computing (HPBDC), with the 30th IEEE International Parallel and Distributed Processing Symposium (IPDPS 2016).

- Saba Sehrish, Seung Woo-Son, Wei-keng Liao, Alok Choudhary, Karen Schuchardt. *Improving Collective I/O Performance by pipelining Request Aggregation and File Access.* In the Proceedings of the 20th European MPI Users' Group Meeting (EuroMPI 2013).

- Saba Sehrish, Grant Mackey, Pengju Shang, Jun Wang, John Bent. *Supporting HPC Analytics Applications with Access Patterns using Data restructuring and Data-centric scheduling techniques in MapReduce.* IEEE Transactions on Parallel and Distributed Systems (TPDPS 2012).

- Saba Sehrish, Grant Mackey, Jun Wang, John Bent. *MRAP - A Novel MapReduce based Framework to support HPC Analytics Applications with Access Patterns.* The International ACM Symposium on High-Performance Parallel and Distributed Computing (HPDC 2010).

**Related Talks**

- Explaining the Fundamental Secrets of the Universe: How Spark and HPC might help Experimental High Energy Physics Analyses? at Grace Hopper Celebration for Women in Computing Data Science Session (GHC 2016)
- Exploring Spark for High Energy Physics Analysis at Big Data Analytics: Challenges and Opportunities Workshop (in conjunction with Supercomputing 2015)
- Big Data in HPC and Programming abstractions at Grace Hopper Celebration for Women in Computing New Investigators' Session (GHC 2012)

**Synergistic Activities**

- Outreach: Grace Hopper Celebration for Women in Computing (GHC), Data Science Track Program Committee (2015, 2016 and 2017)
- Collaboration: Working with Cray/NERSC/AMPLab collaboration to improve the HDF5-based Spark implementation of the CMS Dark Matter analysis
- Liaison for HEP Computational Center of Excellence (CCE)

## Marc Paterno

Computer Science Researcher
Scientific Computing Division
Fermi National Accelerator Laboratory
Batavia, IL 60510-0500
Phone: 630-840-4532
`paterno@fnal.gov`

**Education and Training**

- State University of New York at Stony Brook, Ph.D. in Physics, 1993
- State University of New York at Stony Brook, M.S. in Physics, 1993
- Massachusetts Institute of Technology, S.B. in Physics, 1985

**Research and Professional Experience**

- Computer Science Researcher, Fermi National Accelerator Laboratory, Computing Division, since 2008.
- Computer Professional, Fermi National Accelerator Laboratory, Computing Division, 2000–2008.
- Postdoctoral Fellow, University of Rochester, 1993-2000.

**Selected Publications**

- Saba Sehrish, Jim Kowalkowski, Marc Paterno, *Spark and HPC for High Energy Physics Data Analyses*. The 3rd IEEE International Workshop on High-Performance Big Data Computing (HPBDC), with the 31st IEEE International Parallel and Distributed Processing Symposium (IPDPS 2017).
- Saba Sehrish, Jim Kowalkowski, Marc Paterno, *Exploring Performance of Spark for a Scientific Use Case*. The 2nd IEEE International Workshop on High-Performance Big Data Computing (HPBDC), with the 30th IEEE International Parallel and Distributed Processing Symposium (IPDPS 2016).

**Synergistic Activities**

- PI for Fermilab LDRD-2016/010, "Preparing HEP reconstruction and analysis software for exascale-era computing".
- Collaboration with the HDF Group on use of the HDF5 libraries for storage of experimental HEP data.
- Technical lead for the *art* event processing framework, used by most present and future neutrino and muon experiments at Fermilab.
- Member of the organizing committee for *HPCMASPA 2015*, the "Workshop on Monitoring and Analysis for High Performance Computing Systems Plus Applications".
- Reviewer for IEEE Nuclear Science Symposium and Medical Imaging Conference, since 2011.
- Core analyst and developer for the CMS event processing framework.
- Previously Fermilab representative to the INCITS C++ Standardization Committee, concentrating on library issue relevant to the scientific computing community. Co-leader in design efforts for the C++0X random number library and for the Mathematical Special Functions standards for both C and C++.

**Doctoral Advisor:** Chang-Kee Jung, State University of New York at Stony Brook

**Postdoctoral Advisor:** Thomas Ferbel, University of Rochester

**Jim Kowalkowski**
Computer Science Research
Scientific Computing Division
Fermi National Accelerator Laboratory
Batavia, IL 60510

**Education and Training**
M.S. in Computer Science, Illinois Institute of Technology, 1993
B.S. in Computer Science and Mathematics, Northern Illinois University, 1987

**Academic/Professional Appointments**

| | |
|---|---|
| 2008-present | Computer Science Researcher, Fermi National Accelerator Laboratory |
| 1998-2008 | Computing Professional, Fermi National Accelerator Laboratory |
| 1997 | Software Engineer, Lucent, Illinois |
| 1992-1996 | Software Engineer, Argonne National Laboratory |

**Current Activities**
As Scientific Computing Division Software R&D Coordinator, I have oversight for the future direction and architecture of software systems. I advise and participate in research and development activities related to large-scale computational frameworks and big data science within the division, including moving towards multi-core and many-core platforms.

**C. Selected Publications**
1. Saba Sehrish, Jim Kowalkowski, Marc Paterno. "Exploring Performance of Spark for a Scientific Use case," The 2nd IEEE International Workshop on High-Performance Big Data Computing (HPBDC), with the 30th IEEE International Parallel and Distributed Processing Symposium (IPDPS 2016).
2. Biery, K.; Green, C.; Kowalkowski, J.; Paterno, M.; Rechenmacher, R., "artdaq: An Event- Building, Filtering, and Processing Framework," Nuclear Science, IEEE Transactions on, vol.60, no.5, pp.3764,3771, Oct. 2013.

**Other Publications:**
1. Chard, R., Sehrish, S., Rodriguez, A., Madduri, R., Uram, T. D., Paterno, M., Heitmann, K., Cholia, S., Kowalkowski, J., and Habib, S. (2014). PDACS: A portal for data analysis services for cosmological simulations. In Proceedings of the 9th Gateway Computing Environments Workshop, GCE '14, pages 30–33, Piscataway, NJ, USA. IEEE Press.
2. Dubey, Abhishek and Piccoli, Luciano and Kowalkowski, James B. and Simone, James N. and Sun, Xian-He and Karsai, Gabor and Neema, Sandeep, "Using Runtime Verification to Design a Reliable Execution Framework for Scientific Workflows," Proceedings of the 2009 Sixth IEEE Conference and Workshops on Engineering of Autonomic and Autonomous Systems, EASE '09, pp. 87-96, 2009.
3. Green, C., Kowalkowski, J., Paterno, M., Fischler, M., Garren, L., et al. (2012). "The art framework." J.Phys.Conf.Ser., 396:022020.

**Synergistic Activitie**s
1. HEP Analysis using Big Data: Project lead for sub-project focused on HPC Big Data technologies, including Spark at NERSC
2. Technical lead and co-investigator for *Extending the ART Framework to Support Large Scale Multiprocessing* for the Intensity Frontier (Computational HEP)
3. Senior software architect for the Fermilab HEPCloud program.

**Presentations and Conferences**

1. CHEP 2013: Invited speaker, talk entitled *Data processing in the wake of massive multi-core processors*.
2. CHEP 2015: Track Convener, Data Analytics & Software, Okinawa, Japan.
3. HPCMASPA 2015,2016: PC member.

**Collaborators and Co-editors over the Last Five Years**

Scott Dodelson (FNAL), Salman Habib (ANL), Katrin Heitmann (ANL), Steve Kent (FNAL), Tom LeCompte (ANL), Adam Lyon (FNAL), Ravi Madduri (ANL), Marc Paterno (FNAL), Saba Sehrish (FNAL), Paolo Calafiura (LBNL), Maria Spiropulu (CalTech), Panagiotis Spentzouris (FNAL), Mayur Mudigonda (LBNL), Steve Farrell (LBNL), Giuseppe Cerati (FNAL), Lindsey Gray (FNAL), Josh Bendavid (CalTech), Jean-Roch Vlimant (CalTech), Prabhat (LBNL)

**Adam L. Lyon**
Senior Scientist / Associate Head, Scientific Computing Division
Fermi National Accelerator Laboratory
MS 234
Batavia, IL 60510-0500
1-630-840-5522, lyon@fnal.gov

## ■ Education and Training

| | | |
|---|---|---|
| North Carolina State University<br>Raleigh, NC, USA | Physics | B.S., 1991<br>*Valedictorian* |
| University of Maryland<br>College Park, MD, USA | Physics | Ph.D., 1997 |
| University of Rochester<br>Rochester, NY, USA | Postdoctoral Research Associate | 1997—2002 |

## ■ Research and Professional Experience

| | |
|---|---|
| 2014—Present | Senior Scientist, Fermilab |
| 2014—Present | Associate Head, Scientific Computing Division,<br>Head of Systems for Scientific Applications |
| 2011—Present | Fermilab staff scientist, Muon $g-2$ experiment,<br>Computing, Software, Simulation head |
| 2007—2014 | Scientist I, Fermilab |
| 2005—2011 | SAMGrid Project Manager, Fermilab |
| 2002—Present | Fermilab staff scientist, D0 experiment |
| 2002—2007 | Associate Scientist, Fermilab |
| 1998—2002 | Postdoctoral Research Associate, University of Rochester, CLEO experiment |
| 1991—1997 | Graduate student, University of Maryland, D0 experiment at Fermilab |

## ■ Publications

*Limits on Anomalous Trilinear Gauge Boson Couplings from WW, WZ, and Wγ production in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV* (8.6 fb$^{-1}$)
V.M. Abazov *et al.* (D0 Collaboration), Phys. Lett. B **718**, 451 (2012).

*Wγ Production and Limits on Anomalous WWγ Couplings* (4.2 fb$^{-1}$)
V.M. Abazov *et al.* (D0 Collaboration), Phys. Rev. Lett. **107**, 241803 (2011).

*First study of the radiation-amplitude zero in Wγ production and limits on anomalous WWγ couplings at $\sqrt{s} = 1.96$ TeV*
V.M. Abazov *et al.* (D0 Collaboration), Phys. Rev. Lett. **100**, 241805 (2008)

*Bounds on CP Asymmetry in $b \to s\gamma$ Decays*
T. Coan *et al.* (CLEO Collaboration), Phys. Rev. Lett. **86**, 5661 (2001)

*Bounds on the CP Asymmetry in Like-sign Dileptons from $B^0\bar{B}^0$ Meson decays*
D.E. Jaffe *et al.* (CLEO Collaboration), Phys. Rev. Lett. **86**, 5000 (2001)

*Search for Squarks and Gluinos in Events Containing Jets and a Large Imbalance in Transverse Momentum*
B. Abbott *et al.* (D0 Collaboration), Phys. Rev. Lett. **83**, 4937 (1999)

## ■ Synergistic Activities

Member, Fermilab Muon $g-2$ Collaboration (E989).

Member, Fermilab D0 Collaboration.

Program committee member, Computing in High Energy Physics, 2015

## ■ Collaborators and Co-Editors

Casey, B. (FNAL); Childers, J. (ANL); Herzog, D. (U Washington); Robert, B.L. (Boston U); Childers, J. (ANL);

## ■ Graduate and Postdoctoral Advisors

Graduate Advisors: Nicholas J. Hadley (University of Maryland).
Postdoctoral Advisor: Edward H.Thorndike (University of Rochester).

# Andrew John Norman

Fermi National Accelerator Laboratory
Wilson Hall, MS 120
P.O. Box 500
Batavia, IL 60510-5011
630 804 4016 (office)

## EDUCATION

| | | |
|---|---|---|
| Ph.D. in Physics | College of William and Mary | 2004 |

Dissertation Title: *Measurement of the Branching Fraction for* $K_L^0 \to \mu^+\mu^-e^+e^-$

| | | |
|---|---|---|
| M.S. in Physics | College of William and Mary | 1998 |
| B.S. in Physics and Mathematics | College of William and Mary | 1995 |

Honors Thesis Title: *A Measurement of the Mass of the Muon Type Neutrino*

## RESEARCH POSITIONS

| | | |
|---|---|---|
| Scientist I | Fermi National Accelerator Lab | 2015-Present |
| Associate Scientist | Fermi National Accelerator Lab | 2010-2015 |
| Research Scientist | University of Virginia | 2008-2010 |
| Post Doctoral Research Associate | University of Virginia | 2004-2007 |
| Graduate Research Assistant | College of William and Mary | 1997-2003 |

## PROFESSIONAL POSITIONS

***Fermilab Computing Div.***

| | | |
|---|---|---|
| Assistant Head, Science Workflows and Operations | Scientific Computing Division | 2016-Present |
| Group Leader | SCD Scientific Programs, Nova | 2016-Present |
| Associate Department Head | Sci. Data Processing Solutions | 2016 |
| Group Leader | Scientific Data Management and Database Applications | 2014-2016 |
| Group Leader (Interim) | Offline Production Operations | 2014 |
| Group Leader | Scientific Data Management | 2013-2014 |

***DUNE Experiment***

| | | |
|---|---|---|
| Software & Computing Coord. | DUNE Experiment | 2016-Present |

***NOvA Experiment***

| | | |
|---|---|---|
| Head | Nova, Data Acquisition | 2014-Present |
| Head | Nova, Trigger Systems | 2012-Present |
| Computing Liaison | Nova Experiment | 2012-2016 |
| Deputy Level 2 Project manager | Nova DAQ hardware/electronics | 2012-2014 |
| Level 3 Project manager | Nova DAQ Integration | 2008-2014 |
| Level 3 Project manager | Nova Detector Controls and Monitoring | 2007-2010 |

## SELECT PUBLICATIONS

**NOvA** Collaboration, P. Adamson *et al.*, "Measurement of the neutrino mixing angle $\theta_{23}$ in NOvA," *Submitted to: Phys. Rev. Lett.* (2017) .

**NOvA** Collaboration, P. Adamson *et al.*, "First measurement of electron neutrino appearance in NOvA," *Phys. Rev. Lett.* **116** no. 15, (2016) 151806.

**NOvA** Collaboration, P. Adamson *et al.*, "First measurement of muon-neutrino disappearance in NOvA," *Phys. Rev.* **D93** no. 5, (2016) 051104.

A. Norman *et al.*, "Performance of the NOA Data Acquisition and Trigger Systems for the full 14 kT Far Detector," *J. Phys. Conf. Ser.* **664** no. 8, (2015) 082041.

**NO$\nu$A** Collaboration, A. Norman, "The NO$\nu$A Experiment, The First 12 Months of Commissioning, Operations and Physics Data," *AIP Conf. Proc.* **1666** (2015) 110001.

A. Norman, E. Niner, and A. Habig, "Timing in the NOvA detectors with atomic clock based time transfers between Fermilab, the Soudan mine and the NOvA Far detector," *J. Phys. Conf. Ser.* **664** no. 8, (2015) 082040.

A. Norman *et al.*, "Large Scale Monte Carlo Simulation of Neutrino Interactions Using the Open Science Grid and Commercial Clouds," *J. Phys. Conf. Ser.* **664** no. 3, (2015) 032023.

A. Habig, A. Norman, and C. Group, "Recent Evolution of the Offline Computing Model of the NOvA Experiment," *J. Phys. Conf. Ser.* **664** no. 3, (2015) 032011.

C. Group *et al.* "Fermilab Computing at the Intensity Frontier," *J. Phys. Conf. Ser.* **664** no. 3, (2015) 032012.

**MIPP Collaboration** Collaboration, J. Paley *et al.*, "Measurement of Charged Pion Production Yields off the NuMI Target," *Phys.Rev.* **D90** no. 3, (2014) 032001.

## SYNERGISTIC ACTIVITIES

**International Conference on Computing in High Energy Physics**

| | | | |
|---|---|---|---|
| Track Convener | Data Analysis & Software | Okinawa, Japan | 2015 |
| Track Convener | Data Acquisition & Triggers | Amsterdam, Netherlands | 2013 |

**Neutrino Factories and Future Neutrino Facilities (NuFACT)**

| | | | |
|---|---|---|---|
| Track Convener | Muon Physics (WG4) | Rio de Janeiro, Brazil | 2015 |
| Track Convener | Muon Physics (WG4) | Glasgow, Scotland | 2014 |
| Track Convener | Muon Physics (WG4) | Beijing, China | 2013 |

## CONFERENCE PRESENTATIONS, COLLOQUIUM, SEMINARS

*Muon Physics Working Group Summary*
**NuFact 2015** — Working Group Summary — Rio de Janeiro, BR — 2015

*Performance of the NO$\nu$A Data Acquisition and Data Driven Trigger Systems*
**CHEP 2015** — Parallel talk — Okinawa, JP — 2015

*Archiving Scientific Data Outside of the Traditional HEP Domain*
**CHEP 2015** — Parallel talk — Okinawa, JP — 2015

*Timing in the NO$\nu$A detectors with atomic clock based time transfers*
**CHEP 2015** — Poster Session — Okinawa, JP — 2015

*Large Scale Monte Carlo Simulations of Neutrino Interactions Using the Open Science Grid and Commercial Clouds*
**CHEP 2015** — Poster Session — Okinawa, JP — 2015

*The Nova Experiment: The First 12 months of commissioning, operations and physics data*
**NEUTRINO 2014** [1] — Invited Plenary — Boston, MA — 2014

*Muon Physics Plans and Questions for 2014*
**NuFact 2014**[2] — Working Group Intro. — Glasgow, Scotland — 2014

*Muon Physics Summary*
**NuFact 2013**[3] — Working Group Summary — Beijing, China — 2013

*A $\nu_e$ World at the Intensity Frontier:*
*The Impact of a Large $\theta_{13}$ on the Long Baseline Neutrino Program at Fermilab*
**Wichita State Univ.** Physics Colloquium Wichita, KS 2012

*Matter, Antimatter and the Ghostly Neutrino*
**Science Cafe Wichita**[4] Invited public talk Wichita, KS 2012

*The Impact of a Large $\theta_{13}$ on the Long Baseline Neutrino Program at Fermilab*
**Wayne State Univ.** HEP/Astro Seminar Detroit, MI 2012

*The Impact of a Large $\theta_{13}$ on the Long Baseline Neutrino Program at Fermilab*
**Univ. of Michigan** HEP/Astro Seminar Ann Arbor, MI 2012

*The Mu2e Experiment and the search for Lepton Favor Violation*
**BEACH2012**[5] Invited Plenary Wichita, KS 2012

*Project X and the physics of the Intensity Frontier*
**BEACH2012** Invited Plenary Wichita, KS 2012

*The NOvA Experiment,*
*A long-baseline neutrino experiment at the intensity frontier*
**HQL2012**[6] Invited Plenary Prague, CZ 2012

*Project X and the Intensity Frontier*
**HQL2012** Invited Plenary Prague, CZ 2012

*And 18 other invited talks at major conferences or universities others since 2003*

**COLLABORATORS**

| | | |
|---|---|---|
| NOvA Collaboration | FNAL-E929 | 2005-Present |
| DUNE Collaboration | FNAL-E1071 | 2016-Present |
| Mu2e Collaboration | FNAL-E973 | 2007-2014 |
| MIPP Collaboration | FNAL-E907 | 2004-Present |
| A. Lyon, R. Ross | SciDAC-Data | 2015-Present |

**RESEARCHERS SUPERVISED**

| | | | |
|---|---|---|---|
| Leonidas Aliaga Soplin | FNAL Research Associate | NOvA | 2016-Present |
| Aristeidis Tsaris | FNAL Research Associate | NOvA | 2016-Present |
| Pengfei Ding | FNAL Research Associate | NOvA | 2014-Present |
| Anna Eng | FNAL Target Program | NOvA | 2011 |
| Chad Materniak | Univ. of Virginia PhD Student | MIPP | 2004-2009 |

**ADVISORS**

| | | | |
|---|---|---|---|
| E. Craig Dukes | Postdoc | University of Virginia | 2004-2010 |
| John Kane | PhD (Primary) | College of William & Mary | 1996-2003 |
| Morton Eckhause | PhD (HEP Group) | College of William & Mary | 1996-2003 |
| Robert Welsh | PhD (HEP Group) | College of William & Mary | 1996-2003 |
| John Kane | Undergraduate | College of William & Mary | 1994-1995 |

# Jason M. St. John
*Curriculum Vitae*


Fermilab - MS 309
P.O. Box 500
Batavia, IL 60510
USA

+1 205 568 6479
`stjohn@fnal.gov`


## Education

Ph.D. Physics, Boston University, 2012.

*Advisor*: Prof. James Rohlf
*Dissertation*:"A Search for New Resonances with the Dijet Angular Ratio Using the Compact Muon Solenoid"

B.A. Physics, Harvard, Cambridge, 2001.

*Advisor*: Prof. Howard Georgi

## Research Experience

### *University of Cincinnati, Fermilab, Batavia, IL*

**Postdoctoral Research Fellow (2012-present)**

**Run Coordinator for the LArIAT experiment**: Responsible 24/7 for handling operational emergencies, coordinating the work of diverse experts inside and outside the collaboration. Responsible for the real-world implementation of the collaboration's data-taking goals. Answerable to the spokespersons and first contact on matters of accelerator operations, facility matters, instrumentation survey, and safety.

**Database Group Co-Convener for the MicroBooNE experiment**: (2012-2016) Technical liaison to the Fermilab Database Hosting and Database Applications groups. Translated the needs of the collaboration for the experts, and the experts' capabilities and recommendations for the collaboration. Designed and implemented Hoot-Gibson, the PostgreSQL database tracking the history of connections among hundreds of cables, electronics units, and TPC sense wires, with provision for alternate histories.

**Data Production Convener for LArIAT experiment**: Worked closely with lab scientific computing staff to implement automated, monitored workflow which met the technical specifications of the experiment. Acted as a technical liaison among experts in software framework, file catalog system, data moving system, and data processing.

**Reconstruction Software Co-Convener for LArIAT experiment**: Mentored and guided the coding efforts of graduate students and colleagues, acquiring and spreading expertise in the software framework and batch processing system. Code design, implementation, and coordination of the efforts of colleagues.

**Technical Coordinator for LArIAT installation**: Led students and colleagues in small, task-driven teams, building their expertise and sense of ownership in the experiment. Worked closely with experts on computing, heavy equipment, magnets, cryogenic filtration system, accelerator facility, and safety.

### *Boston University, Boston, MA*

**Graduate Student Research Associate (2004-2012)**

*CMS Experiment (2006-2012)* **with Jim Rohlf**

**Dijet Angular Ratio**: Search for strongly produced massive resonances decaying to pairs of hadronic jets, deriving sensitivity expectations and integrating over nuisance parameters. Developed a batch system workflow to optimize analysis parameters.

**HCAL Data Quality Monitoring group leader**: Led software development within the CMS experiment DQM framework, producing online and offline monitoring for the entire Hadron Calorimeter subsystem.

**HCAL Electronics Monitoring**: Developed low-level register monitoring within XDAQ framework in close consultation with electronics engineers and software framework experts. Provided a digest of highly technical details enabling unfamiliar shift workers could make useful debugging reports to experts.

*US Particle Accelerator School* **(June 12-23, 2006)**

**Advanced Accelerator Physics and Accelerator Simulations** Instructor: Georg Hoffstaetter and Changsheng Song, Cornell University.

*Novel Materials Laboratory, Boston U (2004-2006)* **with Kevin Smith**

**Material studies** Examined transition metal nitrides, organic 1-D conductors with a variety of techniques: Resonant inelastic X-ray scattering, X-ray and UV photoelectron spectroscopy, Auger electron spectroscopy, scanning tunneling electron microscopy. Became fluently conversant with ultra high vacuum techniques and synchrotron undulator/wiggler light sources.

### *Boston University, Boston, MA*

**Physics Lecture Demonstration Specialist (2001-2004)**

*In this role* 'Demo Guy' to Physics Department providing expert consultation and training to teaching faculty. Maintained, organized, revived, and developed the Department teaching apparatus collection. Active and vocal leader in the Physics Instructional Resource Association, part of team designing curriculum and instructing in the Lecture Demonstration Workshop at the AAPT Annual Summer Meeting.

### *Harvard University, Cambridge, MA*

**Undergraduate Research Assistant (June 1998 – June 2000)**

*Multilayer Fabrication Laboratory* **with Dr. Suzanne Romaine.**

*Harvard-Smithsonian Center for Astrophysics* Developed and characterized nanostructure multilayer films for grazing-incidence X-ray telescope optics (10-100 keV). Designed and implemented a database of experimental procedures and results for exploring a large phase space of film preparation and characterization.

## Service and Outreach

*Officer* Fermilab Students and Postdocs Association. 2012-2013.

*Delegate* Fermilab annual lobbying trip.

## Teaching experience

*College of DuPage, Glen Ellyn, IL* Adjunct Faculty. Physics 1201 – algebra- and trigonometry-based classical linear and rotational kinematics and dynamics (work, energy, impulse, momentum, and collisions), fluids, heat, thermodynamics, periodic motion, and wave motion. Spring 2014.

*Boston University, Boston, MA* Graduate Teaching Assistant. Lagrangian and Hamiltonian Dynamics with practical lab section. Advanced Laboratory. The Physics of Music with practical lab section. Introductory Particle Physics. Fall 2004 – Fall 2007.

## Awards, Invited Conference Talks, and Seminar Presentations

"CAPTAIN + LArIAT" NuFact, Rio de Janeiro, Brazil. Invited talk on behalf of the LArIAT and CAPTAIN Collaborations. 2015 Aug 11.

*Poster*, "LAriaT - Liquid Argon in a Testbeam", Neutrino 2014, Boston University, Boston, MA. on behalf of the LArIAT Collaboration. 2014 June 2-7.

"LAriaT - Liquid Argon in a Testbeam" NuFact, University of Glasgow, Scotland, UK. Invited talk on behalf of the LArIAT Collaboration. 2014 Aug 25.

"MicroBooNE: prospects for making the first neutrino interaction measurements on argon at low energy" NuFact, IHEP, Beijing, China. Invited talk on behalf of the MicroBooNE Collaboration. 2013 August 23.

*Invited talk*, "The Search for Massive Resonances with the Dijet Angular Ratio at CMS", LPC Physics Seminar, November 2011

*Invited talk*, "Hadron Collider Physics: Jets and the Physicists Who Love Them", FNAL Graduate Student Association, March 2011

*Instructor and Curriculum Designer*, Di-Jet Physics Course, CMS Data Analysis School, Fermi National Lab, January 2011

*Poster*, "The Dijet Centrality Ratio and Contact Interactions at CMS", US LHC Users' Organization, October 2010

*Graduate Panelist*, Society of Physics Students convention, Boston University, Boston, 2009

*Teaching Fellow of the Year* for excellence in physics instruction, Boston University, Boston, 2006

*Originator*, "Physics Theater" roadshow, co-originator with Dr. Andrew Duffy, Boston University, 2003

## Selected List of Publications

I list here publications I have had major contribution to in the form of analysis, student mentorship, or analysis review.

CMS Collaboration (Vardan Khachatryan et al.), "Search for New Physics with the Dijet Angular Ratio", *CDS* CMS-PAS-EXO-11-026, (2012).

CMS Collaboration (Vardan Khachatryan et al.), "Search for Quark Compositeness with the Dijet Centrality Ratio in pp Collisions at $\sqrt{s}$=7 TeV", *Phys.Rev.Lett.* 105:262001, (2010). arXiv:1010.4439 [hep-ex]

CMS Collaboration (Vardan Khachatryan et al.), "Search for Dijet Resonances in 7 TeV pp Collisions at CMS", *Phys.Rev.Lett.* 105:211801, (2010). arXiv:1010.0203 [hep-ex]

T. Learmonth, C. McGuinness, P.-A. Glans, B. Kennedy, J. St. John, J.-H. Guo, M. Greenblatt, and K. E. Smith, "Resonant soft x-ray inelastic scattering and soft x-ray emission study of the electronic structure of $\alpha$-$MoO_3$",*Phys. Rev. B***79**, 035110 (2009).

In addition to the above, I have been a member of the CMS collaboration author list since 2009 with over 400 publications and more than 46,000 citations.