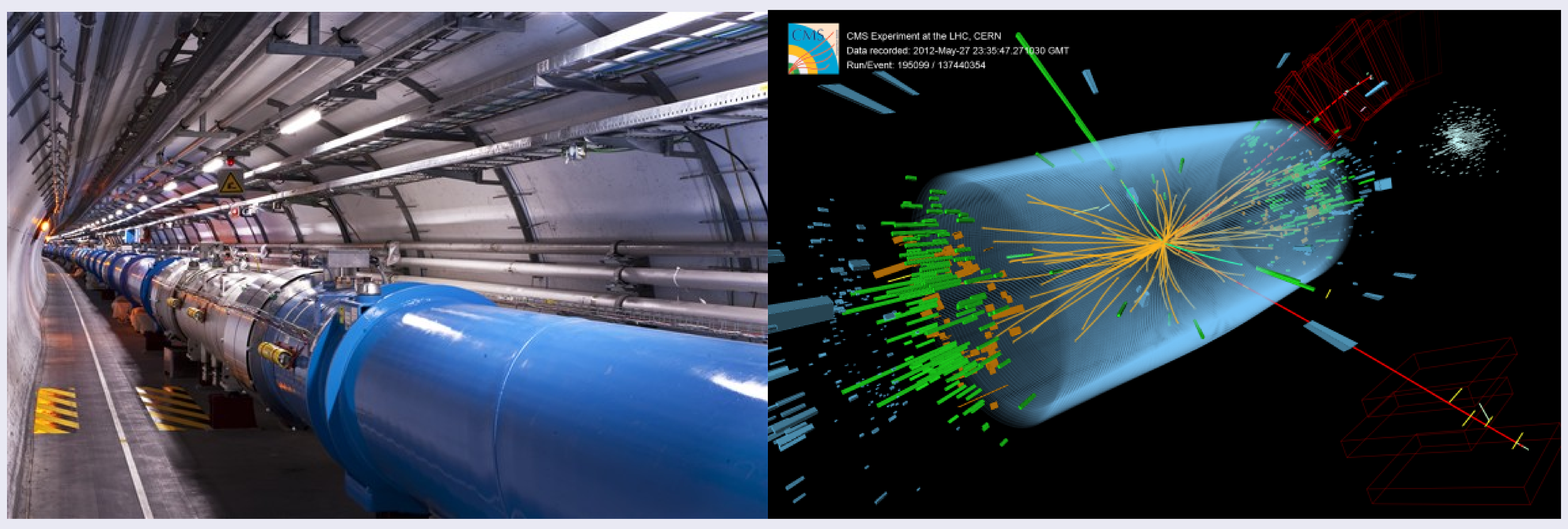
Data-Intensive Analysis for High Energy Physics (DIANA/HEP)

dianahep
http://diana-hep.org

PIs: Peter Elmer (Princeton U.), Brian Bockelman (U.Nebraska-Lincoln), Kyle Cranmer (NYU), Mike Sokoloff (U.Cincinnati)

High Energy Physics (HEP)

The quest to understand the fundamental building blocks of nature, and their interactions, is one of the longest running and most ambitious of human endeavors. Facilities such as the Large Hadron Collider (LHC), where we do our research, represent a huge step forward in our ability to answer these questions. The discovery of the Higgs boson, the observation of exceedingly rare decays of B mesons, and exclusion of countless theories beyond the Standard Model (SM) of particle physics demonstrate that these experiments deliver results. However, the most interesting fundamental physics questions remain wide open, amongst them: What is the dark matter which pervades the universe? Does space-time have additional symmetries or extend beyond the 3 spatial dimensions we know? What is the mechanism stabilizing the Higgs mass from enormous quantum corrections? Are neutrinos, whose only SM interactions are weak, their own anti-particles? Can the theories of gravity and quantum mechanics be reconciled?



The DIANA/HEP Project

The primary goal of DIANA/HEP is to develop state-of-the-art tools for experiments which acquire, reduce, and analyze petabytes of data. Improving performance, interoperability, and collaborative tools through modifications and additions to ROOT and other packages broadly used by the community will allow users to more fully exploit the data being acquired at CERN's LHC and other facilities. The LHC experiments, for example, use nearly 0.5 Exabyte of storage today, and planned upgrades through the 2020s will increase this by more than a factor of 100.

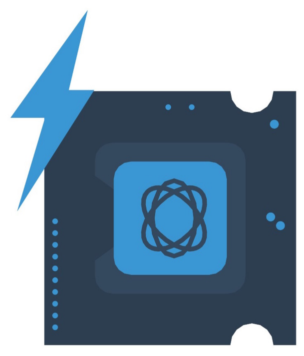
The HEP Analysis Software Ecosystem

ROOT (`https://root.cern.ch`) is home for most community analysis software developed in particle physics and related fields. Begun at CERN in 1995, it provides a sophisticated data format and serialization technology as well as key software tools for data modeling, likelihood fitting, statistics and multivariate data analysis. It also has a broader range of functionalities, not strictly tied to the data-intensive aspects, including interactive C++ analysis, histogramming, graphics, math libraries, image manipulation, and tools for distributed computing. Given the challenges from technology evolution and analysis complexity, DIANA/HEP is building on and improving these community libraries, moving other existing software elements into community libraries, and developing additional new tools. We are also working to strengthen and grow the wider HEP software development community by building durable collaborations which extend beyond the DIANA/HEP project.

Project Status

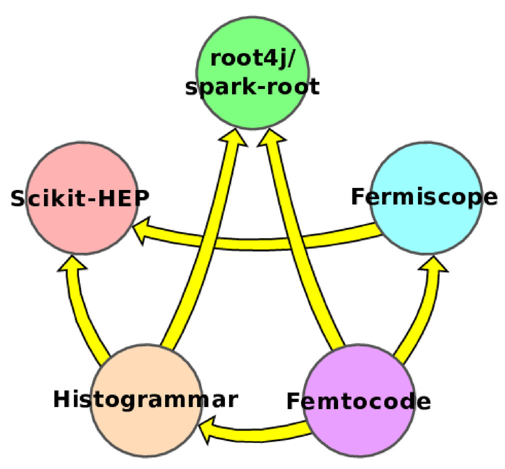
Improved Performance

In order to accelerate existing analysis techniques, we are improving the speed of ROOT IO. We are working to modernize the API and provide faster data rates for simpler objects. By decreasing the time-to-delivery of results, we aim to decrease the time to science.



Bridging to Big Data

The "Big Data" ecosystem is full of clever techniques and ideas of value to data-intensive science. Many build on techniques that are in use in HEP but are better-supported or have larger communities. Unfortunately, the tooling to access HEP data in these frameworks non-existent or immature; DIANA is working to remedy this through tools like Spark-ROOT and Histogrammar. With Femtocode and Fermiscope we are exploring how HEP and Big Data tools together can enable new and powerful ways of working.



New Techniques

We are developing of new tools and methods for high-level statistical analysis in particle physics. Our activities include research and tools for simulator-based inference (Carl), machine learning for particle physics (Scikit-Optimize), high-level software for efficient numerical computations, and education efforts in these respective domains.

High-Level Tools

Improve the interoperability of HEP tools with the larger scientific software ecosystem, incorporating best practices and algorithms from other disciplines into HEP. We're attempting to pull together myriad HEP software packages written in Python into a common interface called Scikit-HEP. The goal is to make Pythonic HEP software easier to discover and thus build a broader HEP Python development community.

Project Team

- Peter Elmer (Lead PI) - *Princeton Univ., Dept. of Physics*
- Brian P. Bockelman (PI) - *Univ. of Nebraska-Lincoln, Dept. of Computer Science and Engineering*
- Kyle Cranmer (PI) - *New York Univ., Dept. of Physics & Center for Data Science*
- Michael D. Sokoloff (PI) - *Univ. of Cincinnati, Dept. of Physics*
- Jinyang Li (Senior Personnel) - *New York Univ., Computer Science Dept.*
- David Abdurachmanov - *Univ. of Nebraska-Lincoln, Dept. of Computer Science and Engineering*
- David Lange - *Princeton Univ., Dept. of Physics*
- Gilles Louppe - *New York Univ., Dept. of Physics & Center for Data Science*
- James Pivarski - *Princeton Univ., Dept. of Physics*
- Eduardo Rodrigues - *Univ. of Cincinnati, Dept. of Physics*
- Zhe Zhang - *Univ. of Nebraska-Lincoln, Dept. of Computer Science and Engineering* (Ph.D. Student)
- Chien-Chin Huang - *New York Univ., Computer Science Dept.* (Ph.D. Student)

Advisory Board

- Amber Boehnlein - *CIO, Thomas Jefferson National Accelerator Facility*
- Katherine Copic - *Director of Growth, Insight Data Science*
- Jacob VanderPlas - *Director of Research, Physical Sciences, eScience Institute, Univ. of Washington*
- Fernando Perez - *Staff Scientist, Data Science and Technology Division, Lawrence Berkeley National Laboratory; Associate Researcher, Berkeley Institute for Data Science, UC Berkeley.*
- Attanagoda Santha - *Architect, Fannie Mae*

Acknowledgement

This project is supported by National Science Foundation grants ACI-1450310, ACI-1450319, ACI-1450323, and ACI-1450377. Any opinions, findings, conclusions or recommendations expressed in this material are those of the developers and do not necessarily reflect the views of the National Science Foundation.

Online version of this poster with clickable links: http://goo.gl/KxSovq