



Contribution ID: 41

Type: Oral presentation to parallel session

## A taxonomy of scientific software applications - HEP's place in the world

*Monday, 14 October 2013 15:45 (21 minutes)*

Modern HEP software stacks, such as those used by the LHC experiments at CERN, involve many millions of lines of custom code per experiment, as well as a number of similarly sized shared packages (ROOT, Geant4, etc.) Thousands of people have made contributions over time to these code bases, including graduate students, postdocs, professional researchers and software/computing professionals. Elaborate software integration, testing and validation systems are used to manage the resulting workflow.

HEP has also been a poster child for "Big Data" science with more than 100 PetaBytes of event data stored around the world. Its applications however typically need event data of order 1MB in memory at a given time plus some tens of MB of calibration data. The resulting data processing, as well as its Monte Carlo simulations, is "embarrassing parallel". The hardware needed for this type of "High Throughput" computing is relatively unspecialized, in the category of "low-end" CPU servers, although great numbers are needed. Most of the code executed is arguably non-numerical, with floating point operations corresponding to only a small fraction of the total time.

These are particular points in a large phase space of scientific applications. Other characteristics like MPI-style parallelism, small code bases, the need for "High Performance" computing (GPU's, specialized interconnects or large memory needs) are not uncommon. Many are true numerical codes and the data requirements vary greatly. This presentation will cover the results of investigations of the software stacks and applications of a variety of other scientific fields. Where are the commonalities and with whom? Or do we inhabit a small niche in the world of scientific computing? Particular attention will be placed on the characteristics and needs of other scientific projects which will require similar or greater amounts of resources than HEP in the next ten years.

**Primary author:** Dr ELMER, Peter (Princeton University (US))

**Presenter:** Dr ELMER, Peter (Princeton University (US))

**Session Classification:** Software Engineering, Parallelism & Multi-Core

**Track Classification:** Software Engineering, Parallelism & Multi-Core