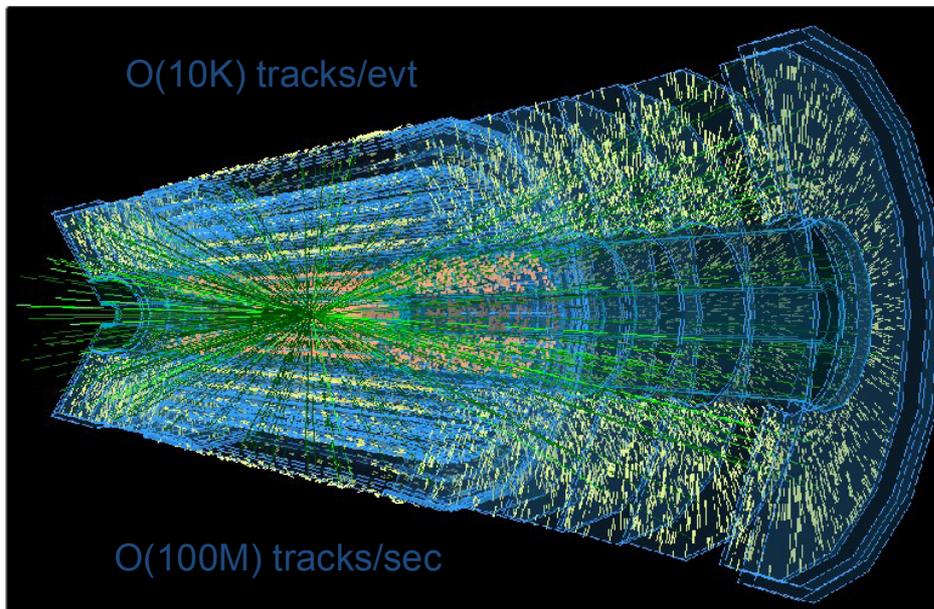


CiML 2016 abstract: Tracking Machine Learning Challenge

Authors : David Rousseau (LAL-Orsay), Cecile Germain (LRI-Orsay), Vincenzo Innocente (CERN), Riccardo Cenci (Pisa), Michael Kagan (SLAC), Isabelle Guyon (ChaLearn), David Clark (UCB), Steve Farrell (LBNL), Rebecca Carney (Stockolm), Andreas Salzburger (CERN), Davide Costanzo (Sheffield), Markus Elsing (CERN), Tobias Golling (UGeneve), Tony Tong (Harvard), Jean-Roch Vlimant (Caltech), Sandro Wenzel (CERN)

The instantaneous luminosity of the Large Hadron Collider at CERN is expected to increase so that the amount of parasitic collisions can reach a level of 200 interaction per bunch crossing, almost a factor of 10 w.r.t the current luminosity. In addition, the experiments plan a 10-fold increase of the readout rate. This will be a challenge for the ATLAS and CMS experiments, in particular for the tracking, which will be performed with a new all Silicon tracker in both experiments. In terms of software, the increased combinatorial complexity will have to be dealt with within flat budget at best.

Preliminary studies show that the CPU time to reconstruct the events explodes with the increased level of parasitic collisions. The increase is dominated by the increase of the CPU time of the tracking, itself dominated by the increase of the CPU time of the pattern recognition stage. In addition to traditional CPU optimisation and better use of parallelism, exploration of completely new approaches to pattern recognition has been started.



To reach out to Computer Science specialists, a Tracking Machine Learning challenge (trackML) is being set up for 2017, building on the experience of the successful Higgs Machine Learning challenge in 2014 (which associated ATLAS and CMS physicists with Computer Scientists). A few relevant points:

- A dataset consisting of a simulation of a typical full Silicon LHC experiments has been created, listing for each event the measured 3D points, and the list of 3D points associated to a true track. The data set is large to allow the training of data hungry Machine Learning methods : the orders of magnitude are : one million event, 10 billion tracks, 1 terabyte. Typical CPU time spent by traditional algorithms is 100s per event.
- The participants to the challenge should find the tracks in an additional test dataset, meaning building the list of 3D points belonging to each track (deriving the track parameters is not the topic of the challenge)
- A figure of merit should be defined which combines the CPU time, the efficiency and the fake rate (with an emphasis on CPU time)
- The challenge platforms should allow measuring the figure of merit and to rate the different algorithms submitted.

The emphasis is to expose innovative approaches, rather than hyper-optimising known approaches. Machine Learning specialists have showed a deep interest to participate to the challenge, with new approaches like Convolutional Neural Network, Deep Neural Net, Monte Carlo Tree Search and others.