

# TrackML : a tracking Machine Learning challenge

*Thursday, 5 July 2018 14:40 (20 minutes)*

To attain its ultimate discovery goals, the luminosity of the Large Hadron Collider at CERN will increase so the amount of additional collisions will reach a level of 200 interaction per bunch crossing, a factor 7 w.r.t the current (2017) luminosity. This will be a challenge for the ATLAS and CMS experiments, in particular for track reconstruction algorithms. In terms of software, the increased combinatorial complexity will have to be harnessed without any increase in budget.

To engage the Computer Science community to contribute new ideas, we organized a Tracking Machine Learning challenge (TrackML) running on the Kaggle platform from March to June 2018, building on the experience of the successful Higgs Machine Learning challenge in 2014.

The data were generated using [ACTS] (<http://acts.web.cern.ch/ACTS/latest/doc/index.html>), an open source accurate tracking simulator, featuring a typical all silicon LHC tracking detector, with 10 layers of cylinders and disks. Simulated physics events (Pythia tbar) overlaid with 200 additional collisions yield typically 10<sup>7</sup> tracks (10<sup>7</sup> hits) per event.

The task to be performed by participants in the challenge is the “pattern recognition”: associate the hits to tracks corresponding to the original charged particles. The participants are given 10<sup>7</sup> events (including truth information) to train their algorithm, while the evaluation by Kaggle is run on 100 other events. The score used to rank the candidates is the fraction of hits correctly assigned, with a weighting mechanism to favor higher momentum tracks and hits in the innermost and outermost detector layers. In this challenge, there is no CPU constraint, however a second phase of the challenge to be run during the summer will have strong computational constraints.

The emphasis of the challenge is to explore innovative Machine Learning approaches, rather than hyper-optimising known combinatorial approaches. In preliminary discussions with the ML community, Convolutional Neural Network, LSTM, Deep Neural Nets, Monte Carlo Tree Search, geometric Deep Learning have been mentioned. A very simplified 2D version of the challenge (see reference) was successfully run as a two day hackathon in March 2017.

In this talk, the first lessons from the challenge (which will just have been completed) will be discussed. What algorithms have emerged and are the most promising ? How robust is the score, compared to deeper performance evaluations ?

**Primary authors:** CALAFIURA, Paolo (University of California Berkeley (US)); Dr FARRELL, Steven Andrew (Lawrence Berkeley National Lab (US)); GRAY, Heather (LBNL); Dr VLIMANT, Jean-Roch (California Institute of Technology (US)); INNOCENTE, Vincenzo (CERN); SALZBURGER, Andreas (CERN); AMROUCHE, Sabrina (RSA - Universite de Geneve (CH)); GOLLING, Tobias (Universite de Geneve (CH)); KIEHN, Moritz (Universite de Geneve (CH)); GERMAIN, Cecile (Universite Paris Sud); GUYON, Isabelle; ROUSSEAU, David (LAL-Orsay, FR); YILMAZ, Yetkin (Laboratoire Leprince-Ringuet, France); GLIGOROV, Vladimir (Centre National de la Recherche Scientifique (FR)); HUSHCHYN, Mikhail (Yandex School of Data Analysis (RU)); USTYUZHANIN, Andrey (Yandex School of Data Analysis (RU))

**Co-author:** Prof. PLANT, Claudia (Universität Wien)

**Presenter:** GOLLING, Tobias (Universite de Geneve (CH))

**Session Classification:** Computing and Data Handling

**Track Classification:** Computing and Data Handling