1 Description for the general public

The Large Hadron Collider (LHC) sings. Its superconducting magnets hum and whistle as they fling particles at nearly light-speed around a 27 km-long circular tunnel that straddles the Swiss/French border. But this December, the world's largest particle accelerator will go silent [1]. The pause is scheduled as the result of long planned upgrade to the high luminosity version of the experiment. With advancements in hardware, for some experiments, it will bring the possibility of recording up to 3 orders of magnitude more data than currently. Since all of the High Energy Physics experiments heavily rely on statistics, it is of great importance to accumulate as great volume of data as possible. This is why, the scheduled update to the so called "Run 3" brings excitement and enthusiasm to the world of physics. Nevertheless, it also poses some new challenges, one of which is the great need for fast reconstruction methods.

Current reconstruction techniques employed in all of the four major LHC experiments, including ALICE, rely on statistical simulations. Such approach is very accurate, but also extremely slow. Monte Carlo simulations require tremendous resources, hence they already occupy 50-70% of all CERN computing power worldwide. It is evident that with the upcoming update, even such a technical colossus as GRID, won't handle so much computation. The obvious answer for such a problem would be to add more CPUs. Although, how much can we scale the largest grid computing system which is using over half a million computer cores spread around 170 centres in 42 countries?

Maybe, instead of doing so we should focus on speeding up the simulation techniques? Because of the natural shortcomings of Monte Carlo methods, we propose a fundamentally different approach based on machine learning. Current hype on Artificial Intelligence resulted in great number of new generative algorithms e.g Generative Adversarial Networks or Variational Autoencoders. Such techniques proved their potential in numerous applications: from fake celebrities images generation up to drug design. The real-life problem of track simulation in CERN experiments is the perfect environment to test the methods in production.

In this project we propose to implement new algorithms based on GANs or VAE, which will try to "learn" all of the physical laws from already collected examples. This can be achieved thanks to the core of the method – artificial neural network. This machine learning technique is actually commonly used because of its great capabilities in discovering complex patterns in data. Those patterns may be used to predict desired output. In our case this will be complex physical laws, which our network will have to discover to compute the correct output. Thanks to over 24 PB of real exemplar data, we believe that well trained neural networks may be able to properly simulate trajectories of particles which emerge during the collision. Our preliminary results of studies conducted on the most important tracking detector – Time Projection Chamber – are promising. In this proof of concept solution we observed even 4 orders of magnitude speed up in comparison with traditional simulation techniques, although additional research is needed to match the quality of full, computationally complex, sequential techniques. We plan to achieve this goal by incorporating newest advancements in machine learning methods, not only related to the field of generative models. We believe that our sequential problem of predicting following points of particle trajectories may also benefit from different techniques used in the domain of natural language processing; such as very recent *attention* method, which resulted in self-attention GAN model. On the other side we also intend to apply additional constrains based on physical laws, to provide high-fidelity results of simulation.

The LHC will start to sing a new song in two years. This time however, it won't be simple whistling and humming. With extended capabilities the superconducting magnets will play the whole symphony of noises. Simulation methods shouldn't stay behind, and fast machine learning based solutions may be a perfect *preludium* to this masterpiece.

[1] https://www.datacenterdynamics.com/analysis/probing-universe/