To the attention of the chairs of the JENAS group:

Jorgen D'Hondt, ECFA chair, Jorgen.DHondt@vub.be

Marek Lewitowicz, NuPECC chair, Marek.Lewitowicz@ganil.fr

Teresa Montaruli, APPEC chair, teresa.montaruli@cern.ch

Expression of Interest

for a synergic research plan of potential interest of the JENAS group

T. Dorigo, D. Boumediene, C. Delaere, D. Derkach, J. Donini,

A. Giammanco, R. Rossin, M. Tosi, A. Ustyuzhanin, P. Vischia,

December 3, 2019

Project title:

Machine Learning-Optimized Design of

Particle Detector Layout for Future Scientific Experiments

Keywords:

Machine learning, Detector design, Future colliders, High-energy particle physics, High-energy nuclear physics, Astro-particle physics, Detector simulation, Deep neural networks, Pareto optimization

Participating institutions and contacts:

INFN – Sezione di Padova, ITALY:
Université catholique de Louvain, BELGIUM
Université Clermont Auvergne, FRANCE
Higher School of Economics, RUSSIA

Tommaso Dorigo, INFN-Padova – tommaso.dorigo@gmail.com
Andrea Giammanco, UCLouvain – andrea.giammanco@cern.ch
Julien Donini, U. Clermont Auvergne – julien.donini@cern.ch
Denis Derkach, Higher School of Econom. – dderkach@hse.ru
Andrey Ustyuzhanin – andrey.ustyuzhanin@cern.ch

0. Introduction

The present document outlines a proposal for a wide-range study of detector optimization, using machine learning technology for the optimization of the design, and targeting the future capabilities of specialized artificial intelligence technology for the event reconstruction. While in the text below we mostly make use of HEP examples to clarify the goals of the program and its possible deliverables, much of the research is equally of benefit for any effort of detector design for a future experiment in astro-particle physics, muography, or high-energy nuclear physics, which share the physics of the interaction of radiation with matter and the technology of its measurement, as well as the issues connected with the optimized extraction of the relevant information, which in the future will necessarily be based on specialized artificial intelligence systems. Because of this common ground, we believe that this expression of interest should be considered by the JENAS community as a way to create a proficuous bridge between scientists working in the three areas of interest.

1. Background

1.1 An unshaken paradigm

In the course of the past half century, particle detectors used by fixed target and then collider experiments have benefited from incremental as well as paradigm-shifting advances in detector technology. As an example of the former, one may cite the progress in construction technology, operation and readout of wire chambers; as an example of the latter, the introduction of thin silicon sensors for trackers stands out. In contrast with these advancements, a striking observation is the standing paradigm connected to the philosophy of how a stream of particles should be detected and measured: this has mostly remained unchanged across many decades.

We can summarize the paradigm with the sentence “track first, destroy later”: since charged particles can be traced in low-material detection elements such as gaseous devices, while neutral particles such as neutrons, photons, and neutral kaons usually require destructive interactions in a calorimeter if one is to measure their energy, the no-brainer setup has always involved a low-material tracker followed by a thick calorimeter. All the detectors constructed for large particle physics experiments, but also –with differences due to their special constraints –in astro-particle and high-energy nuclear physics experiments, have so far duly complied with it. Further, the need for effective detection and measurement of muons as important probes of electroweak processes has invariably led to the addition, externally to the calorimeter systems, of muon-optimized detector layers, profiting of the high penetration power of those particles, and usually exploiting for additional momentum measurements the return flux of the internal magnets used to bend charged particles in the tracker.

1.2 Pile-up at the LHC and countermeasures

Indeed, the complexity of high-energy collision events in environments such as the one of pp or heavy ion collisions at the LHC, today made worse in the former case by the high pile-up of hundreds of near-simultaneous collisions, are making the requirement of thin trackers even more compelling, if anything:

unwanted nuclear interactions in the material make the reconstruction of charged particles trajectories harder. While machine learning algorithms have been summoned to ease the resulting complex reconstruction tasks, the obtained results are so far still wanting. In the meantime, the need for a 4-dimensional measurement of particle ionizations to disentangle pile-up and primary interaction particles has emerged quite clearly, and consequently, ATLAS and CMS are both getting equipped with timing layers for their phase-II upgrades.

1.3 Hadron calorimetry and jet substructure needs

A different, and partly unexpected, recent input concerns the optimal design of hadron calorimeters. The apparently "simple" task of collectively measuring the energy of jets of hadrons, and the original lack of tight requirements for the spatial resolution of the components of hadronic showers in the medium, led designers of past or present experiments to invest a relatively smaller fraction of their total budgets in the construction of those devices. That status quo has been shaken to the roots during the past decade by the rise of jet substructure studies, which have shown how electroweak decays of heavy objects (top quarks, W and Z bosons, Higgs bosons) can be effectively sorted out of the competing QCD backgrounds in very high-energy jets, thus opening the way to completely revolutionized ways to look for new physics signatures at the very high end of the investigated energy spectrum. Nowadays, high granularity has become a compelling requirement for a hadron calorimeter at a particle collider. The CMS HGCAL detector for the LHC Phase 2 is an example of a step in that direction, and its design will certainly improve by a large margin the amount of usable information about the showers, their development, pointing, and composition. Other similar examples (CALICE for ILC, CaloCube, the SHIP experiment, to which A. Ustyuzhanin participates) exist. Another detector which is being built with these ideas in mind is the LHCb calorimeter for the LHC phase-2 upgrade, where D. Derkach is giving significant contribution with an intensive optimization with machine learning techniques using fast simulation of the apparatus. One thing to note is that in general these designs, while highly improving the situation, have not been optimized for machine-learning-powered reconstruction.

When discussing present-day calorimeters, a note has to be made concerning CMS. Originally endowed with a hadron calorimeter of lesser performance than the corresponding one of the competing ATLAS experiment, CMS regained most of the lost ground by exploiting a very performant "particle flow" reconstruction algorithm. This has been only possible thanks to the high field integral of the inner CMS tracker, which allows particles of different momenta to be "sorted out" by the strong bending. This kind of use of the strong 4-Tesla field, originally rather conceived for compactness of design and high performance of the momentum measurement, is an example of how complex the optimization of a modern particle detector can be, even in the absence of the elephant in the room, machine learning. In that case, the redundancy and strong points of CMS as a whole came to the rescue, thanks to sophisticated new software. But planning ahead for such resilience seems mandatory in the future. In fact, the already mentioned HGCAL project is explicitly accounting for the exploitation by the particle flow algorithm.

1.4 Muon energy measurements at the TeV scale

The R&D of highly-granular calorimeters has further shown –incidentally, also thanks to the application of machine learning in reconstruction of the detected signals– that a sufficiently precise measurement of the longitudinal energy deposition in a high-spatial resolution device can become an effective means of measuring the energy of TeV-range muons, inferring it from the amount of radiation loss they withstand. In that regime, magnetic bending fails to provide a similar performance. It appears very interesting, therefore, to investigate how the measurement of very high-energy muons can be optimized by the targeted detection of soft photons emitted by muons interacting with the thick medium of a calorimeter, by considering designs that allow that information to be extracted with success.

2. Machine learning, the elephant in the room

2.1 Generalities

Machine learning (ML) is ubiquitous, and has redefined performance in a number of human activities and technologies, and reshaped the way we think about optimization. However, in high-energy physics the application of ML-driven solutions to analysis problems has caught up rather slowly, due to a general skepticism of the use of “black boxes” as the algorithms of choice for the very common classification or regression tasks in measurements and searches. As an example, while already in 1989 the first proposals for application of artificial neural networks in hadron collider searches were put forth at the Tevatron, it took over a decade for these tools to become a routinely used tool in particle physics analyses. Algorithms such as Random Forests, Boosted Decision Trees, and in particular Neural Networks, also in their “deep” form, have operated a paradigm shift, by improving the performance of our measurements by large amounts. Nowadays the old standard cut-and-count analysis for a new physics search is usually considered, at most, as a cross-check to a more advanced ML-powered one.

2.2 Innovation

The next step of embracing this new technology for fundamental physics research should be obvious: **allow machines to inform the design of future detectors**. What is at stake is the real meaning we give to the word “optimization”. If we construct today a serious plan of R&D aimed at the design of a detector which will operate, say, at a future hadron collider 30 years from the time of writing, we cannot omit to consider the elephant in the room. The reason is simple. When we optimize the sensors for a tracking device, when we choose layouts for the cells of a calorimeter, when we operate choices on budget allocation for the different components, when we define requirements for the various resolutions of detection elements, we are trying to optimize hundreds, if not thousands, of independent variables in a vastly under-constrained space, and the task is clearly super-human. Worse still, we are usually trying to optimize against the wrong cost function, as we short-cut our goals to be “achieve highest resolution for isolated photons”, e.g., forgetting about the rest of the parameter space: while we want our target to be “the highest precision on the Higgs boson self-coupling that available funds can buy”, we have no precise idea of what compromises between the various design choices could reach the desired result, so we stick to past experience –e.g., high photon resolution will bag us a better identifiable Higgs decay signal– thus ignoring how vastly different is the actual challenge. The result is a potentially enormous performance loss. It is worth pointing out that the inclusion of the cost of detector components in a complete cost function is a very complex problem, but if successfully performed it gives access to a Pareto optimal boundary, which gives the management the power to make more realistic choices on the overall design.

The innovative character of the proposed plan –the one of **achieving a true and complete Pareto optimization of our detector design**– should at this point be obvious. But there is one further step we have to take to see how important it is to change the way we think at optimization. It is easy to predict that if machine learning is already with us today, in 30 years it will look prehistoric to reconstruct charged particle trajectories with, e.g., a presently advanced tool such as a pixel-seeded Kalman filter. While it is hard to foresee exactly how deep and wide will the use of ML tools be in particle reconstruction, it is quite reasonable to expect that the task of extracting high-level variables from the millions of readout channels of the detector components will be entirely commanded by deep neural networks. And the view that those artificial algorithms have at the reconstruction task is sufficiently different from ours to beg us to pause and rethink the whole system.

2.2 Design strategy

Is it reasonable to design a detector without considering how the relevant information will be extracted from its readouts? Of course it is not. Hence, we must start the R&D for a future detector – be it for a HEP, an astro-HEP, or a Heavy ions application all set in the future – by trying to attack the problem in a different way. The idea is to rethink at the layout of detection elements (initially based on present-day technology), such that they produce the best result on some well-thought-over, high-level goals, once all nuisance parameters affecting the measurements are incorporated in the equation, and once we model in a suitable way the different reconstruction capability ML tools can achieve on the task. What this means is that a set of DNNs, tasked with providing answers as independent as possible, must be constructed, and their weights and biases learned, from simulated data coming from a continuously varying set of detector designs and layouts, reaching out to ones starkly in contrast with accepted design paradigms. In other words, there is a highly multi-dimensional space of parameters defining our design choices to investigate, and we must walk out of the local minimum determined by our preconceived notion of "track first, destroy later".

Crucially, the DNNs must contain very elaborate loss functions at their core, which are as precisely connected to our final goals (precision of measurement of some fundamental parameters, discovery reach for a broad set of new physics signatures) as possible, while accounting as well as possible for the systematic uncertainties affecting the measurements in program. In section 3 below a few examples are given of studies we want to undertake in this context.

A seminal study which exemplifies how today's software frameworks, specialized programming languages (python) and ML-design tools (Keras, TensorFlow) allow the construction of neural networks capable of solving similarly complex tasks, is an open-access article recently appeared in Computer Physics Communications (also available as computer code in GitHub and as an arXiv preprint[1]). The algorithm described in the reference (INFERNO, from "inference-aware neural optimization") demonstrated how large improvements in the measurement precision of a parameter can be achieved when the loss function of the NN (see Fig. 1) incorporates the effect on the measurement of nuisance parameters – which "NN optimization" tasks in HEP measurements so far have neglected to include, suffering quite significant performance losses. More to the point, the algorithm demonstrates that such a global, and true, optimization *can at all* be achieved.

[1] P. de Castro Manzano, T. Dorigo, "INFERNO: Inference-Aware Neural Optimization", Comp. Phys. Commun. 244 (2019) 170; Arxiv:1806.04743v2, June 2018.

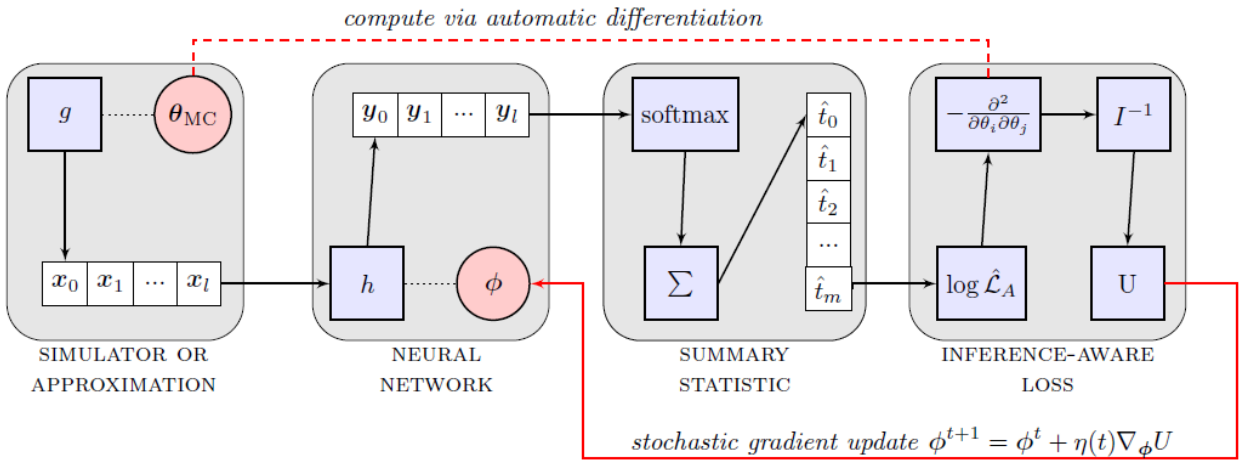

Figure 1: control flow of the INFERNO algorithm, which extracts optimized configuration for a neural network, such that the resulting measurement is maximally robust to the effect of nuisance parameters.

3. Tasks and Deliverables

3.1 Outlook

The research program sketched above is obviously quite ambitious, and because of its wide range is, we believe, beyond the scope that any individual R&D project can encompass. What we are interested to start is a pilot study targeting a few specific, well contained questions of interest to a wide range of experimental endeavours, with a methodology which is however perfectly scalable to a grander goal of full-scale optimization for a future detector at a big experiment (such as the FCC-hh, for example). The **ingredients** are fast simulation of physics processes of interest (MADGRAPH, PYTHIA for collider physics, e.g.), state-of-the-art tools of detector simulation (DELPHES, GEANT4, FLUKA) applied multiple times to the same generated events in order to tame the stochastic nature of detector response, and advanced ML programming in auto-diff frameworks (KERAS, PYTORCH, TENSORFLOW) to sew together the inputs and the desired outputs in a full optimization chain. The **methodology** is a systematical comparison of performance of very different layouts of detection elements of present-day availability, when the metric is the measurement precision on an initially narrow selection of fundamental physics parameters (e.g., the Higgs boson branching fractions) when the accounting of all significant nuisance parameters potentially affecting their estimate is included. Extrapolation to the performance of DNN reconstruction on the considered layouts will be enabled by the benchmarking of selected deep-learning results on a narrow set of tasks. The **focus** will initially be the construction of a software framework similar to the one of the INFERNO algorithm, adapted to the broader task of considering the detector geometry parameters as members of the feature space together with the unknowns systematically affecting the considered measurements. **Results** should be provided in the form of specific guidelines informing the most promising geometry, layouts, and choices of detection technology. These will initially be a narrow set of inputs, eventually scalable to the ultimate aim of shaping the design of a fully optimized future detector.

3.2 Deliverables

3.2.1 Hybridized tracking and calorimetry

The success of the CMS particle flow algorithm, which exploits to some degree the information on the identity of particles generating hits and energy deposits (e.g. photons vs neutral pions or charged hadrons), is a motivation for a broad study of the benefit that a hybrid design of longitudinally overlapping tracker and calorimeter system may obtain. Nuclear interactions, if properly identified, can provide some degree of particle ID information which a powerful algorithm can exploit, as well as detailed extrapolation to the adjacent calorimeter deposits (both upstream and downstream). While complex, an extension of the tracking of particles inside the depth of the calorimeter system can be achieved with specialized machine learning reconstruction tools, if the detector layout provides the required information. Silicon-based calorimetry is an ingredient of this plan, but it is not the only one, as more diversified technologies can better inform the ML-driven extraction of information.

Besides publications in refereed journals, the main deliverable for this task can be defined in the form of selected response functions over geometry space, parametrizing the resolution of specific quantities of interest once a full reconstruction is performed, including the effect of selected systematic uncertainties. As an example again taken from collider physics, one such quantity of interest may be defined as the uncertainty in the cross section of boosted hadronic vector boson decays in a sample extracted from measurements of the invariant mass of boosted jets, accounting for detector-related (jet energy scale and resolution) uncertainties as well as simulation-related (PDF, detector simulation parameters) uncertainties.

3.2.2 Muon energy resolution from radiative losses

As mentioned in Sec. 1.4, it has become evident how the radiative loss of high-energy muons can be exploited to determine their energy better than what the bending of their track can achieve, if the loss is measured with a suitably designed calorimeter, sporting sufficient longitudinal granularity and the most effective detection elements. In this task, which may be of interest for both HEP and Heavy Ion physics applications, the observability of a signal of interest, such as e.g. that of a boosted Higgs boson decay to muon pairs, is studied as a function of the geometry of a calorimeter where finely segmented but conventional hadronic sections are alternated with suitably designed electromagnetic detection elements employing the most promising detection technology for radiation photons in the energy range of interest. The optimization of the design will simultaneously take into account the performance of the calorimeter to measure with high precision boosted decays of W, Z, and H bosons similarly to what is mentioned in Sec. 3.2.1, by constructing a loss function which combines the two research goals.

The main deliverable associated with this task can be formulated as an optimal longitudinal segmentation and identity of the layers of the calorimeter, targeting the above physics cases, together with a map of the performance loss as a function of the variation from the optimal design.

4. Participants

The institutions listed above (INFN, UCLouvain, UCA, HSE) share an interest in investigating the issues discussed in the previous sections. We are in the process of creating a network to share our expertise and organize the work for a first attack at the giant task we have set out for ourselves. We would be very happy to coalesce a much larger group around this task.

INFN-Padova (INFN) includes

Dr. Tommaso Dorigo (First Researcher at INFN)

Dr. Mia Tosi (INFN associate and type-A Researcher at University of Padova)

Prof. Roberto Rossin (INFN associate and Associate Professor at University of Padova)

Dr. Giles Chatam Strong (INFN associate and post-doctoral scientist at University of Padova)

Dr. Hevjin Yarar (INFN associate, Ph. D. student at the University of Padova)

The INFN personnel has ample experience with ML applications in HEP and detector simulation and design.

Université catholique de Louvain (UCLouvain) includes

Dr. Andrea Giammanco (Maitre de Recherches FNRS)

Dr. Pietro Vischia (Post-doctoral fellow at UCLouvain)

Prof. Christophe Delaere (Maitre de Recherches FNRS and associate professor at UCLouvain)

Mr. Hesham el Faham (Ph. D. student at UCLouvain)

All UCLouvain researchers have wide experience with detector simulation, having also developed a fast-simulation package (DELPHES) which will be useful for this project, as well as expertise with ML applications; they are also pursuing studies of muography for which detector optimization is an important issue.

Université Clermont-Auvergne (UCA) includes

Prof. Julien Donini (Full Professor)

Dr. Djamel Boumediene (CNRS Researcher)

The UCA personnel has experience in ML-driven detector design also within the CALICE collaboration, which is pioneering some advanced studies in this direction, and with which we foresee possibilities of useful cooperation.

National Research University Higher School of Economics (HSE), a partner located in Moscow, Russia, associated to YANDEX, will participate to the project offering secondments to the hired post-doctoral scientists, under the supervision of Dr. Denis Derkach (senior researcher at the laboratory of methods for big data analysis). In addition, Dr. Derkach and Prof. Ustyuzhanin have large experience in the topic, and have pioneered studies of detector design for LHCb and SHIP with machine learning.

5. Summary

The design of the geometry and layout of a future detector for, e.g., a higher-energy hadron collider, a new heavy ion facility, a muon collider, or a future space experiment, cannot avoid considering the paradigm-shifting potential of applied artificial intelligence in the reconstruction of particle signals. We propose to start an investigation of the potential gains in physics reach achievable with a full ML-driven optimization of the detector hyper-parameters, performed with deep learning technologies. This optimization must account for the scenario wherein such a device will operate – one in which automated extraction of high-level event features will be executed by ML algorithms; as well as focus specifically on the final goals of the experiment (precision of measurement of fundamental physics parameters, discovery reach in parameter space of new physics theories) factoring in production costs, rather than targeting intermediate and abstract goals which unavoidably lead to sub-optimality. While this is an obviously overambitious goal, we believe that a pilot study that lays down all the necessary software technologies while attacking a small set of well-contained questions will constitute a compelling step in the right direction.

The innovative potential of this project should be obvious, as we plan to demonstrate cutting-edge techniques in the optimization of detector design, extrapolating our reconstruction capabilities to levels presently still not achieved. The demonstrated gains of such an end-to-end optimization approach should become an asset for a wide range of future experiments across different fields of research which all employ particle detection as the source of the information on the physical processes they study.

For the proponents

Tommaso Dorigo