Louvain-la-Neuve, September 6, 2021

Differentiable Programming for Experimental Design

Opening of the Workshop

Tommaso Dorigo

INFN, Sezione di Padova

Welcome!

Am very happy to be here!

Heartfelt thanks to Pietro Vischia for making this possible,

and to the organizing committee at UcLouvain!

Many thanks also to IRIS-HEP and JENAA (Appec-NuPecc-ECFA), who graciously provided funding that made this possible!

Why another ML workshop?

There are lots of ML workshops around (I myself organized a couple which took place last month, and spoke at two more in July!)

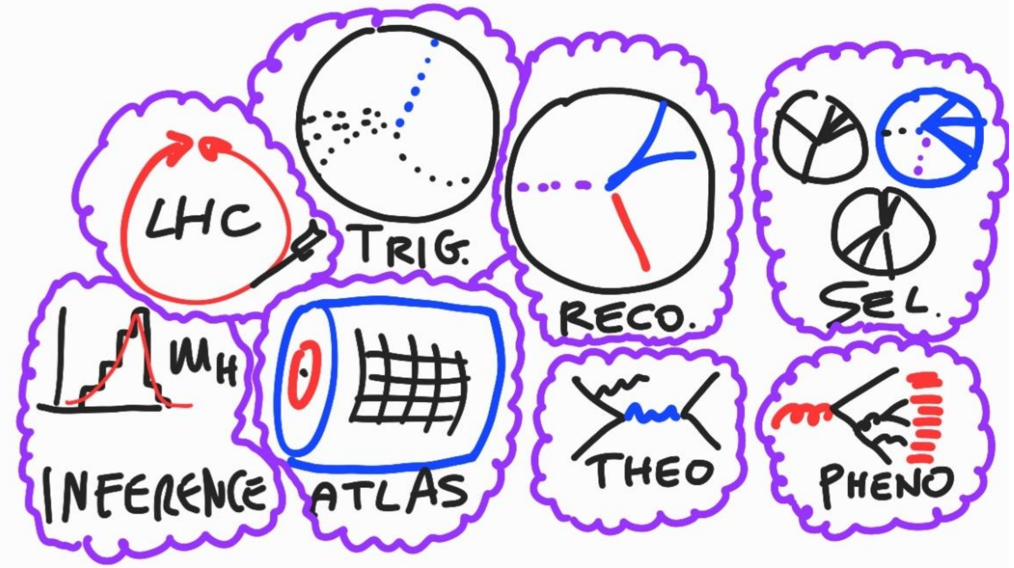
The ML revolution has changed the way we do our data analysis. But a shift of interest is taking place from maxxing the ROC curve to harder problems, e.g. (in HEP):

- Evolve pattern recognition following tougher demands (e.g. mitigating PileUp at the LHC, finding substructure in high-pT jets, extracting directionality in DM emulsion signals, ...)
- Use for online applications, to improve triggering, data collection, data compression
- Improve inference by realigning experimental techniques to goals; rotate systematics away

What do all of the above have in common? They aim to **get more bang for the buck**: they are meant to IMPROVE SENSITIVITY

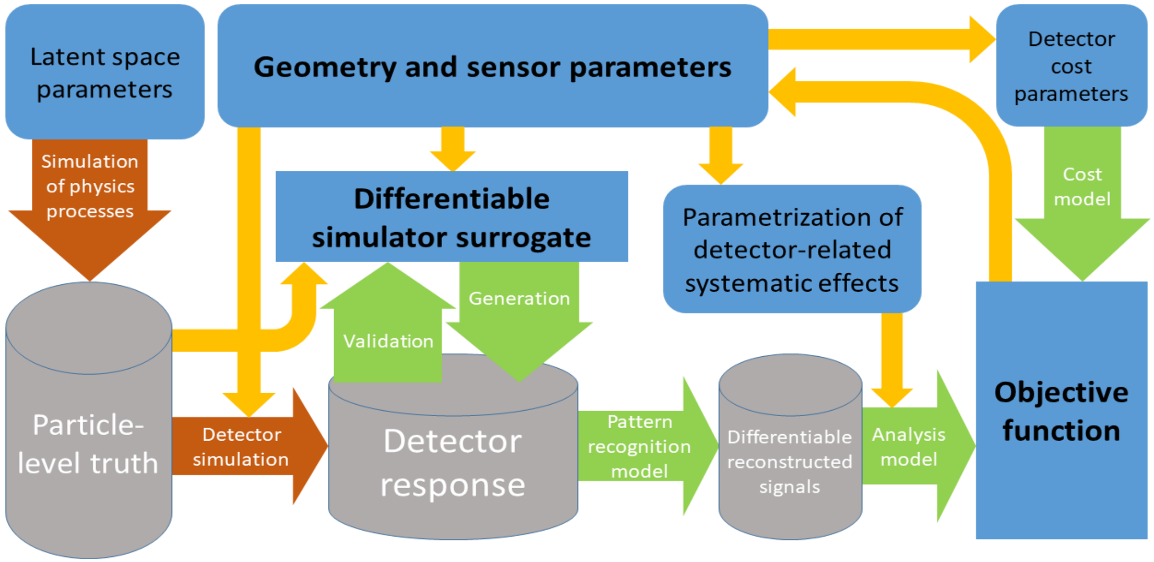
That's all good, but...

Applications such as the cited ones only look at a part of the system or two at a time



Great work, but can we do better?

Yet computer science today allows us to take a more holistic view at the bang-buck problem...



A cognitive dissonance

Should close-to-retirement detector builders lead the design of instruments that will take data in 20 years from now?

Not to mention that the AI which will reconstruct trajectories in a future tracker will laugh at (or worse, be annoyed by) the Kalman filter they designed our detector around...

Detector development has a life of its own, independent on pure research goals. That’s ok, but – shouldn’t goal-oriented research leverage the knowledge of detector experts, but lead the design choices for specific apps?

Note to self: The time of the Rutherfords, the Fermis, the Goldhabers has gone. Today it is impossible to master theory, experimental design, and data analysis at the same time...

Well-consolidated paradigms

- Be robust. You’re studying invisible, minuscle particles – so aim for redundant, cross-calibrating setups
  - Ok, but it comes at a price!
- Track first, destroy later. This one is older than I am! But in the meantime silicon pixel detectors, boosted jet tagging, particle flow happened.
  - Is a lightweight tracker followed by a thick calorimeter really our only option?
- Prefer symmetrical layouts. Fine, it eases construction, simulation, and everything in between
  - ....But is the physics symmetric? Not really, more often than not...
- As for stacking up 2D structures to build 3D detectors, this was a no-brainer until recently, but today 3D printing and sub-micron technology allow for alternatives.
  - Or are we so lucky that the optimal solution always lays on the 2D subspace?

Computer Science to the rescue

Computer science offers new solutions to our tasks today – but it is up to us to create the required interfaces.

Because of that, I'm very excited to see that we are succeeding at putting together a mixed community of researchers bringing tough problems (the physicists) and computer scientists who have the expertise to help them find solutions

What's in the plate for the latter? Highly-cited papers, new challenging problems to solve (possibly fostering advances in unrelated areas and spin-offs), and long-lasting, fruitful collaborations

MODE

This workshop is organized by the MODE collaboration (mode-collaboration.github.io)

- As per its title, we focus explicitly on **detector optimization using differentiable programming**

Why detector optimization? Can't we keep fiddling with reconstruction and inference?

- No – there is strong coupling between construction choices and ultimate sensitivity, and the gold mine is in realigning the two.

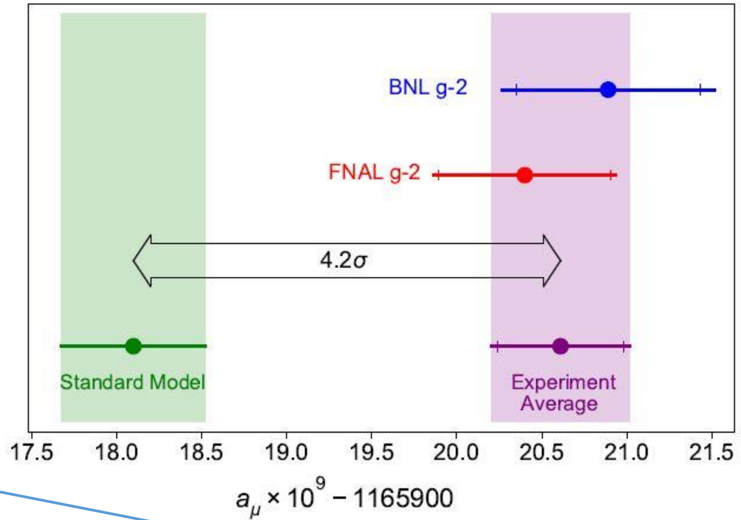
Why DP? Optimization is not necessarily reliant on a differentiable model. My gut feeling: the latter is more likely to allow for the learning of new, innovative ways to exploit the complexity of the design space. We only need our models to be accurate enough...

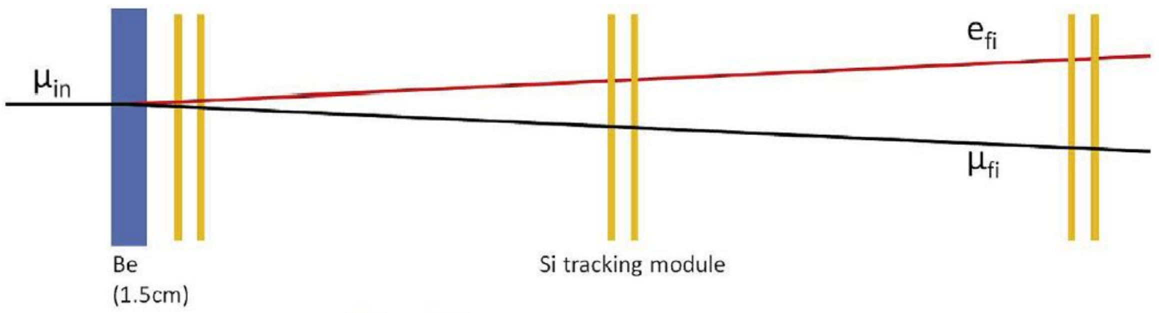
Is it worth taking this on? In early attempts with O(10) free parameters, we see **x2 improvements** in the relevant metrics as typical returns. What will a full optimization in a 100-1000D space bring? I, for one, want to know!

One example of geometry optimization: MUonE

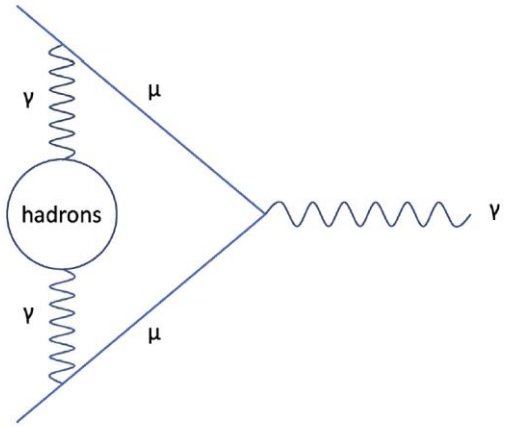
The MUonE experiment aims to determine with high precision the muon-electron elastic scattering differential cross section, to extract hadronic contributions and reduce the systematics of the g-2 muon anomaly

The experiment must be sensitive to hadronic loop effects particularly at **high q^2**, where a 10^{-4} measurement may substantially improve the theoretical understanding of the g-2 value

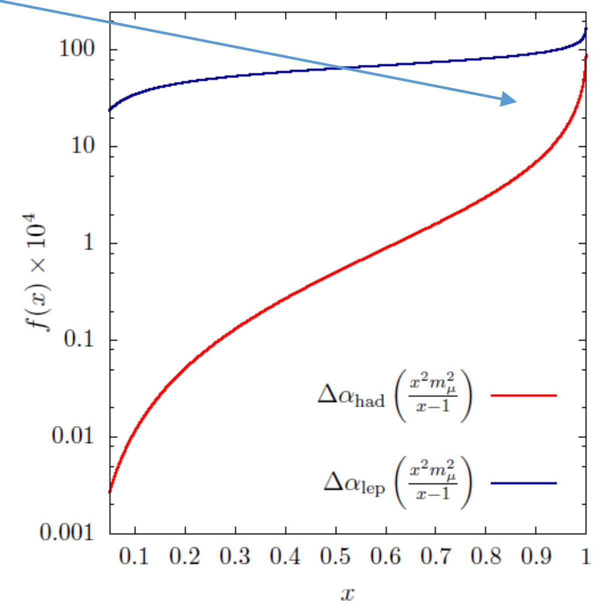




Above: layout of one of 40 1m-long stations



Virtual hadronic loop



MUonE optimization

By optimizing layout with a discrete sampling, I proved how an improvement of **a factor of 2** in the relevant metric could be achieved without increase in detector cost

The study also proved how dreaded systematic effects from positioning uncertainties could be nullified by software means

ELSEVIER

Physics Open
Volume 4, September 2020, 100022

Geometry optimization of a muon-electron scattering detector

Tommaso Dorigo [1]

Show more

https://doi.org/10.1016/j.physo.2020.100022

Get rights and content

Abstract

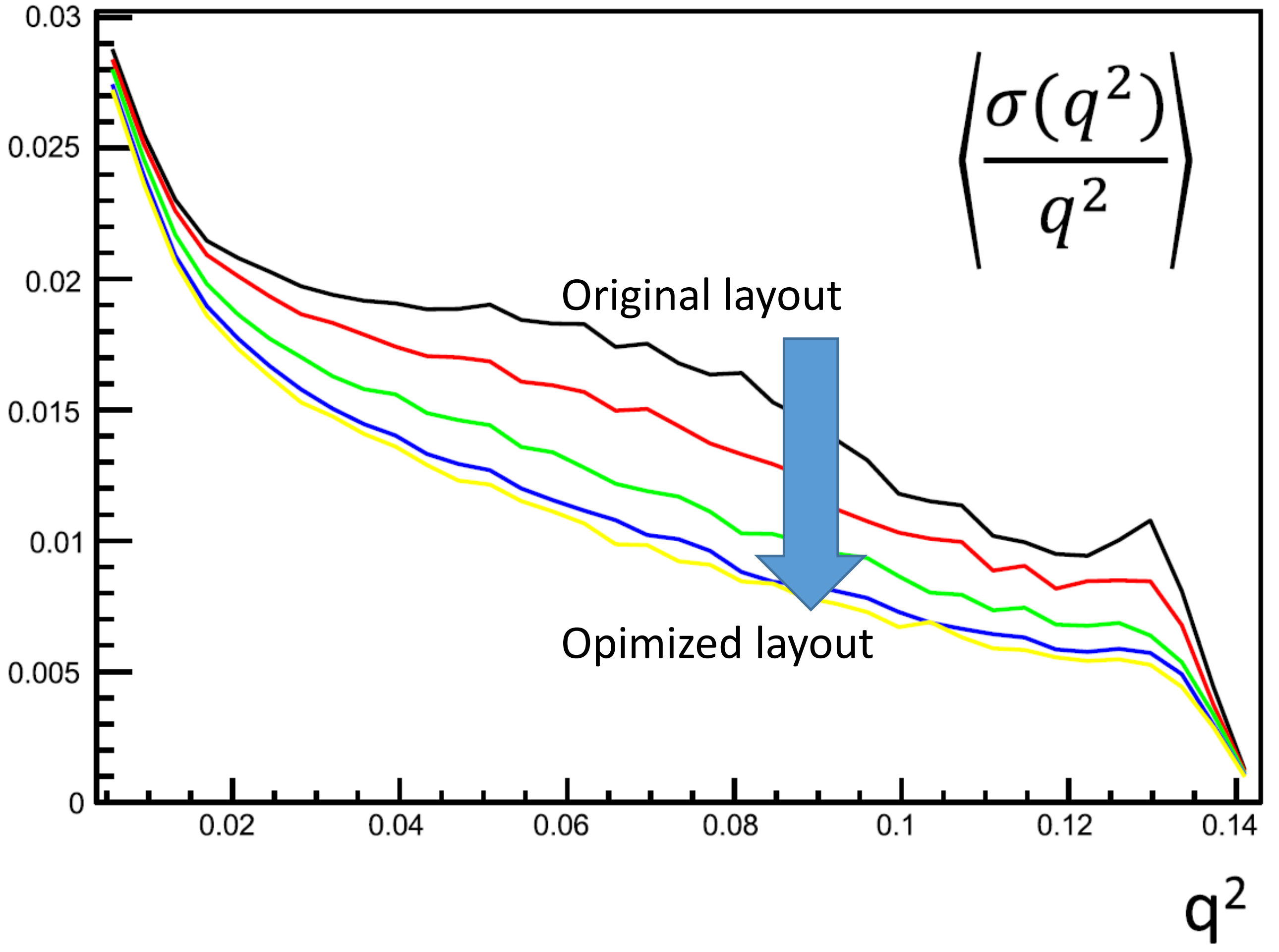
A high-statistics determination of the differential cross section of elastic muon-electron scattering as a function of the transferred four-momentum squared, $d\sigma_{el}(\mu e \to \mu e)/dq^2$, has been argued to provide an effective constraint to the hadronic contribution to the running of the fine-structure constant, $\Delta\alpha_{had}$, a crucial input for precise theoretical predictions of the anomalous magnetic moment of the muon. An experiment called "MUonE" is being planned at the north area of CERN for that purpose. We consider the geometry of the detector proposed by the MUonE collaboration and offer a few suggestions on the layout of the passive target material and on the placement of silicon strip sensors, based on a fast simulation of elastic muon-electron scattering events and the investigation of a number of possible solutions for the detector geometry. The employed methodology for detector

Sample results

The study was *not* performed with deep learning technologies, as it was not strictly necessary given the reduced space of design choices I wished to investigate.

By optimizing layout with a discrete sampling, I proved how an improvement of **a factor of 2** in the relevant metric could be achieved without increase in detector cost

The results prove that design optimization is not something alien to our reach, but rather, something we should pay more attention to!



Above: relative resolution in event q^2 for different configurations (the higher, black line is the original proposal by the MUonE coll.)

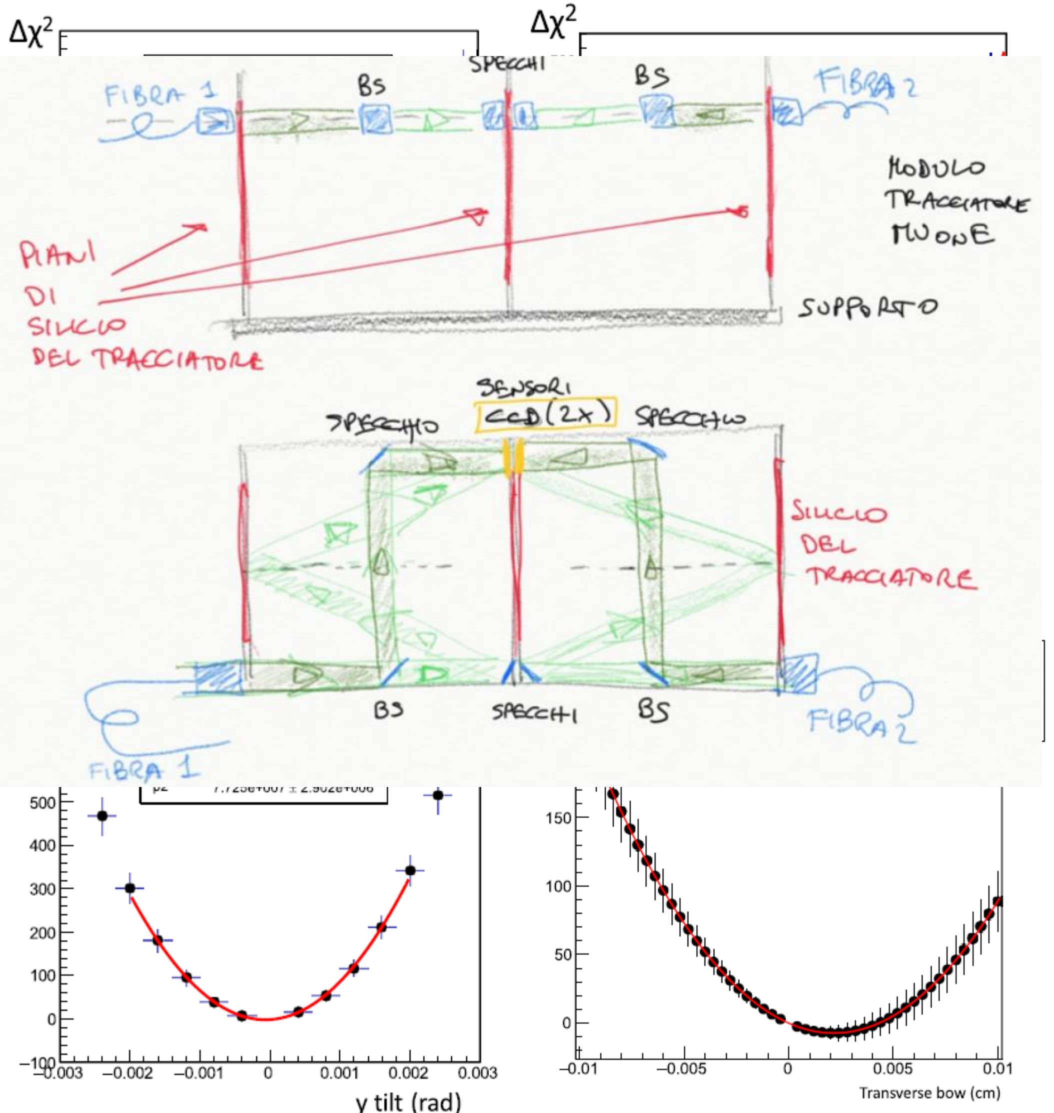
Speaking of systematic uncertainties,

MUonE correctly identified the need for locating the scattering vertex to within 10µm along the beam axis (it has a strong impact on the q^2 resolution), and proceeded to design a very fancy holographic laser system, to be mounted on each station (=40 systems) to monitor the sensors locations

Cost: several hundred kEuro

The optimization study showed that with 5’ of muon beam data, the location, tilt and bow of all detector and target elements can be determined with O(1µ) accuracy by a global fit to the vertex

This is an example of the dividends that the study of a full model of (physics)+(detector)+(reco method)+(inference extraction) can provide



Another example: Muon shielding in SHIP

In a seminal work by MODE members, local generative surrogates of the gradient of the objective function were proven to allow for the minimization by SGD and a **x2 reduction in muon background fluxes** in the SHIP experiment

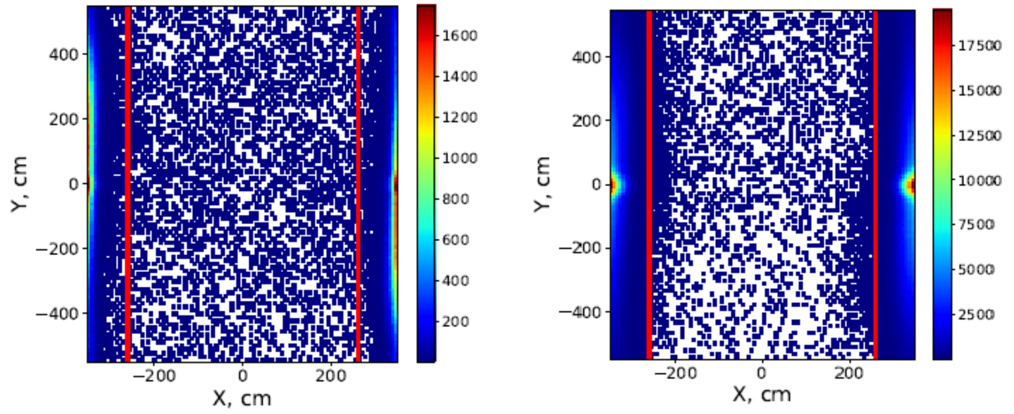
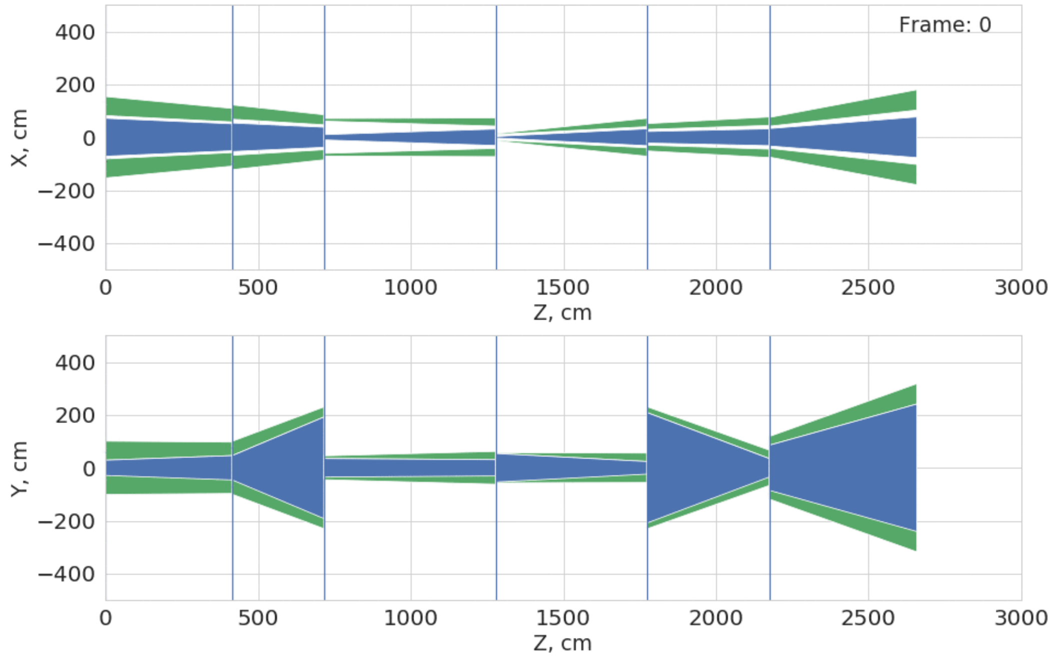


Figure 7. Muon hits distribution in the detection apparatus (depicted as red contour) obtained by Bayesian optimization (Left) and by L-GSO (Right), showing better distribution. Color represents number of the hits in a bin.



Geometry optimization at work in real time!

100D - Can we pull it off?

For a more complex particle physics application than the two examples above, things might rapidly become untreatable. Or maybe not. Can we pull it off?

No answer to this question twarts the risk of being dubbed wishful thinking.

But innovation requires investing efforts, believing in our odds, and trusting our foresight.

I submitted an ERC advanced grant request on using DP for detector design last year. They laughed me down. How naive! After all, particle detectors are maddeningly complex – and he wants to automatize their design?? He's nuts, or deluded, or both.

But the software tools are there! What is missing is the infrastructure, and a lot of sweating... And the dividends are potentially huge.

What MODE is about

We aim to create a versatile, scalable, customizable infrastructure, where a generic detector design task can be encoded, along with all the players (pattern reco, nuisances, cost constraints, a well constructed objective function).

Then **automatic scanning** of the space of design solutions becomes possible!

MODE will tackle a variety of use cases, from easy to hard ones, within multiple domains (HEP, astro-HEP, nuclear, neutrino physics)

→we will build a library of solutions, made of modules which can be reused

→gain expertise

→increase capability to solve harder tasks

→reduce time to result for any given new task

Every application will result in high-visibility publications, open-source code

Final goal: **making optimal use of research funds, improve discovery and measurement potential**

In addition…

What we are after has loads of spin-off potential, e.g.

→Muon tomography: a wealth of industrial applications

→Hadron therapy: let's put these particles to good use for human health!

→Radiation shielding: good for nuclear plants and waste management, as well as for space travel

The kind of solutions we want to work out can further generalize to a number of other complex systems optimization tasks, if we are good enough to create versatile infrastructures.

Want to join us? The door is open!

You are here today (in person or by avatar) because you worked at topics related to our research plan, or because you have an interest in doing so in the future.

So you look as close to the right audience as I can think of, when I say we need your help to improve the effectiveness of MODE.

I hope you will consider joining MODE, sharing research time and resources with us (e.g. help with coding, if you're still capable of that, or direct your students to our research topics, if you're a low-bandwidth senior!)

According to our Statute, becoming a collaboration member requires you to

→be interested in our research plan, and to produce research in that area

→bring competence of relevance, or vow to acquire it

→aim to contribute to it within your possibilities

Just send to the MODE steering board (Dorigo, Donini, Giammanco, Vischia, Ratnikov) a confirmation of the above and a short bio (or CV, or pointer to google scholar) and chances are we'll get you in!

Enjoy the workshop!