Istituto Nazionale di Fisica Nucleare

ML4HEP INFN school
Erice, April 12 2023

Deep-Learning-Optimized Design of Experiments: Challenges and Opportunities

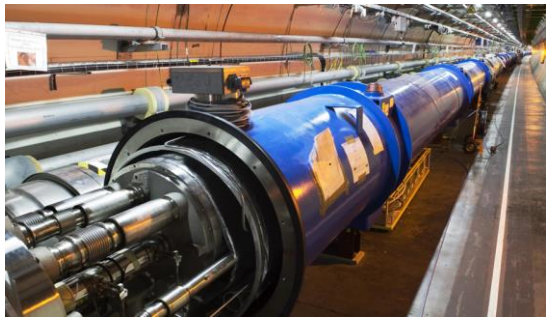
Tommaso Dorigo

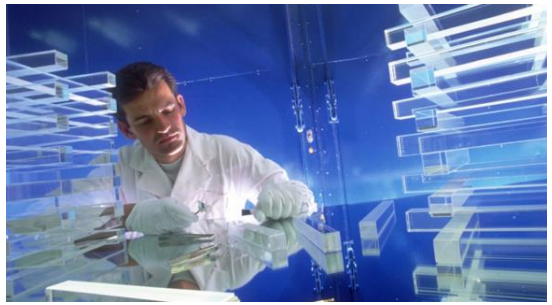
INFN, Sezione di Padova

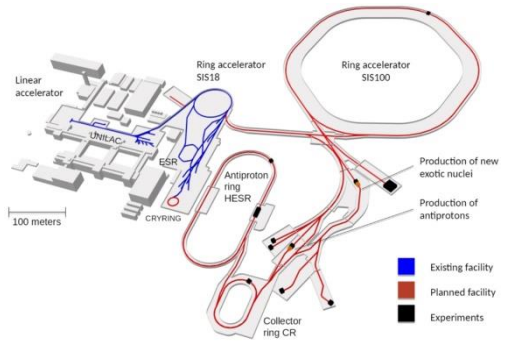
Contents

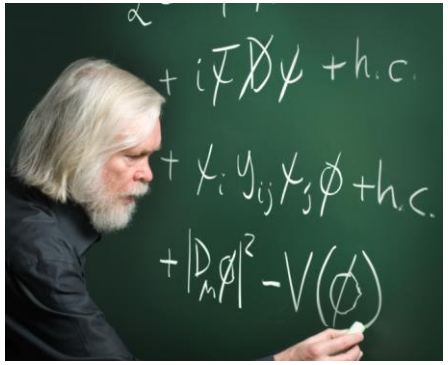
- The context: deep learning for particle physics today
- One step forward: Inference awareness
- The challenging future of HEP
- Many steps forward: Optimization of experimental design
- A few examples from ongoing studies
- A future application: hybrid granular calorimetry
- Concluding remarks

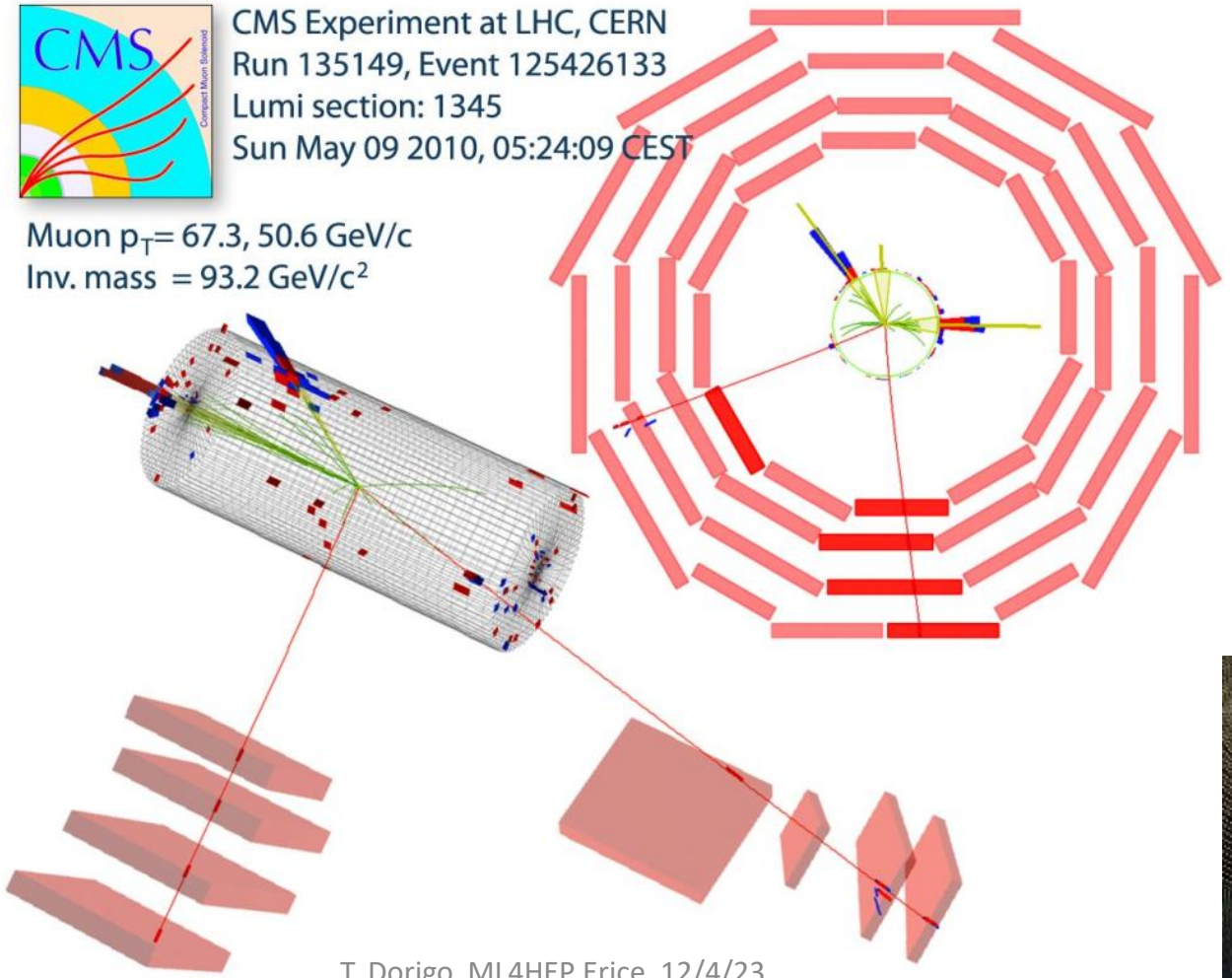
The Context






Linear accelerator
Ring accelerator SIS18
Ring accelerator SIS100
UNILAC
ESR
Antiproton ring HESR
CRYRING
100 meters
Production of new exotic nuclei
Production of antiprotons
Collector ring CR
Existing facility
Planned facility
Experiments

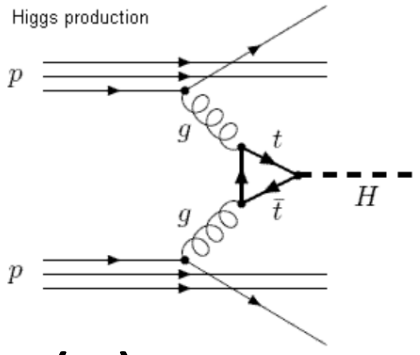



CMS
Compact Muon Solenoid
CMS Experiment at LHC, CERN
Run 135149, Event 125426133
Lumi section: 1345
Sun May 09 2010, 05:24:09 CEST
Muon p_T= 67.3, 50.6 GeV/c
Inv. mass = 93.2 GeV/c^2

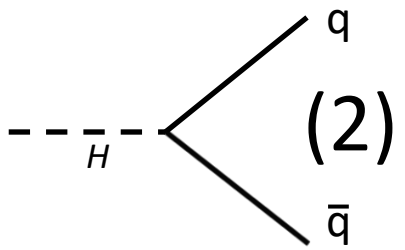

SOUTH

From Collisions to Observable Features

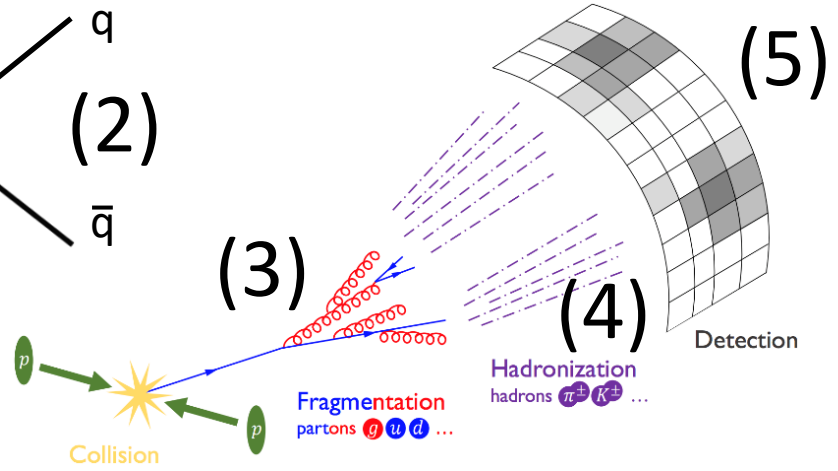
1) A proton-proton collision may produce a Higgs particle, together with hundreds of others
2) The Higgs boson decays immediately to, e.g., a pair of quarks
3) Quarks fragment into streams of quarks and gluons...
4) which create unstable hadrons
5) Hadrons decay into «stable» particles: protons, electrons, photons, neutrons, muons, pions, kaons...
6) which reach the detectors, leaving information on their energy and direction
7) From detected signals one may try to infer the nature of the original process, on a statistical basis

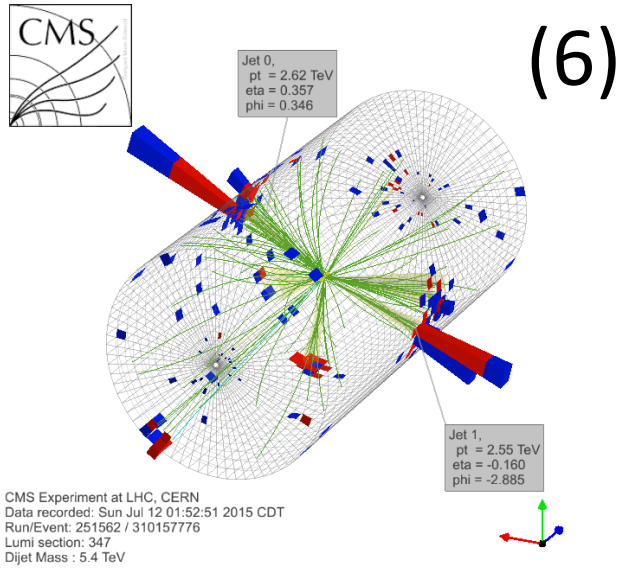


(1)



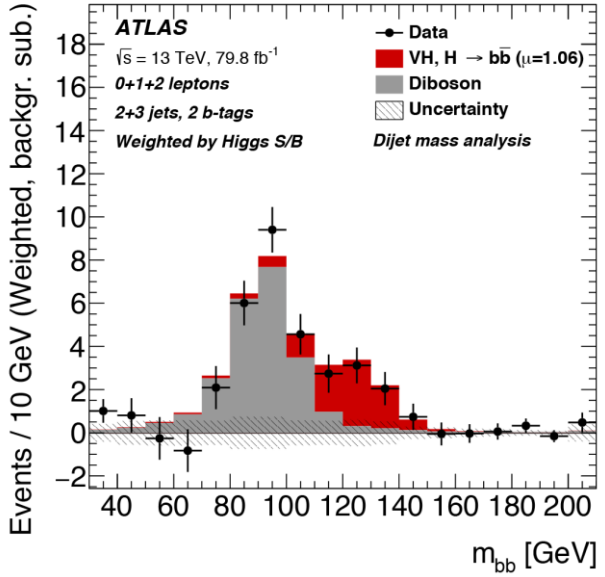
(2)





(6)

(7)



The Rise of Deep Learning in HEP

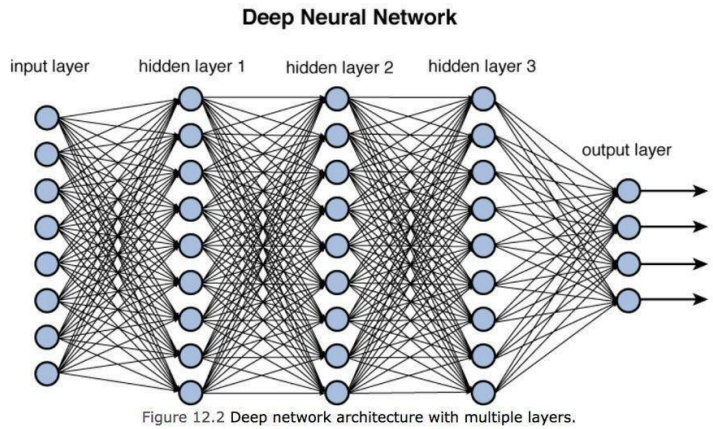


Figure 12.2 Deep network architecture with multiple layers.

DNNs use skyrocketed in HEP after 2012, when ML tools were used for the Higgs discovery. A **true paradigm change**!

Further evidence of the benefit of ML tools for HEP: Kaggle Higgs challenge [Kaggle 2014]

→ 1800 teams (physicists, statisticians, computer scientists), 13k euro prize

→ Task: separate H→ττ decays from backgrounds in LHC simulated data

Most effective solution was based on pool of DNNs, with emphasis on cross-validation

Alternative methods commonly used in HEP (xgboost, Bayesian NN, etc.) were beaten soundly

Solution	Pseudo-significance
Gabor Melis (DNN pooling)	**3.806**
MultiBoost	3.405
TMVA boosted trees	3.200
Naive Bayesian classifier	2.060
1D cut-based selection	1.535

equiv. to 6 times more data!

Outstanding Problems in Fundamental Science

- Formulating new theories of Nature
- Extracting sufficient statistics from high-D data
- Ensuring complete control of our type-1 error rates
- Explore higher energy / higher intensity frontiers, ensuring we do not miss new physics

And, more of relevance to this talk:

- Aligning detector design to final experimental goals
- Achieving full optimization in presence of finite resources and external constraints
- Aligning detector design to future software capabilities

Current Trends of DL in HEP

Current trends of DL applications in particle and astroparticle physics include:

- **Classification** for particle ID and signal discrimination
- **Regression** for measurement of physics parameters from multidimensional data
- **Image-based** methods: e.g. classification of particle jets from calorimeter images with CNNs
- **Optimal inference** through robustness to systematics uncertainties (see below)
- Use of NNs for **online data acquisition**
- **Anomaly detection** for searches of new physics
- **Generative methods** for fast MC simulation

But **we must** look further, as time from blueprint to commissioning is O(20) years!

In market-driven human activities, **co-design** of hardware and software is already happening. So in the following we will give a look at where the next paradigm shift will occur. Let us start with the first logical step forward: optimization of data analysis.

Inference-Aware Analysis Methods

Signal/noise discrimination is an intermediate step in HEP analysis chains: using signal-enriched data we measure production rates and other physical parameters

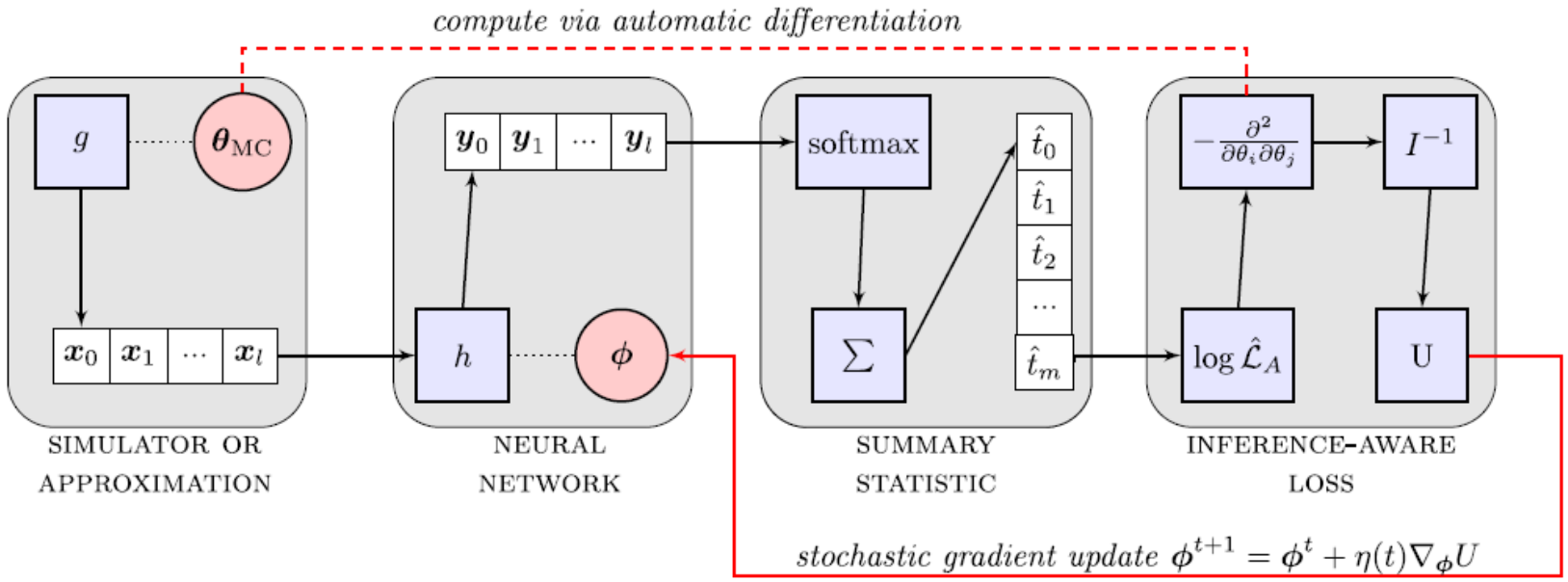
Imprecise models used in training affect the measurement with nuisance parameters

These are only dealt with **after** the training is completed

→ misalignment between the classifier objective and the experimental goal!

Recent development [De Castro & Dorigo 2019] employing differentiable programming techniques: **INFERNO**

- Complete model of inference
- Feedback loop of loss function constructed w/ Fisher matrix of final measurement
- DNN learns to account for nuisances in training phase!



Above: *flow chart of INFERNO, showing the main elements in the optimization loop*

Inference-Aware Analysis Methods / cont'd

Signal/noise discrimination is an intermediate step in HEP analysis chains: using signal-enriched data we measure production rates and other physical parameters

Imprecise models used in training affect the measurement with nuisance parameters

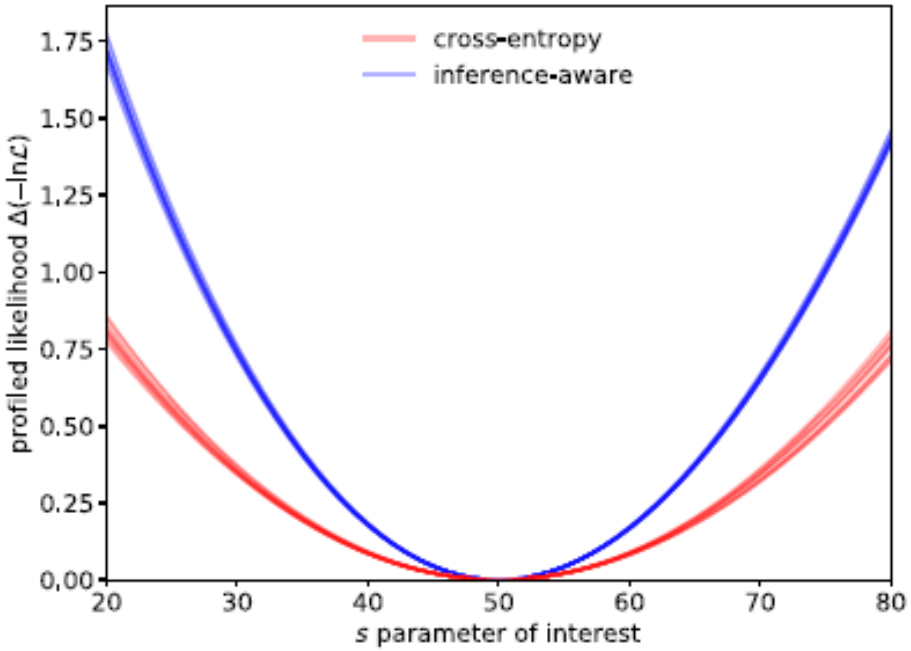
These are only dealt with **after** the training is completed

→ misalignment between the classifier objective and the experimental goal!

Recent development [De Castro & Dorigo 2019] employing differentiable programming techniques: **INFERNO**

→**Successful realignment** w/ experimental goals

→large gains in precision of resulting inference



(b) profile-likelihood comparison

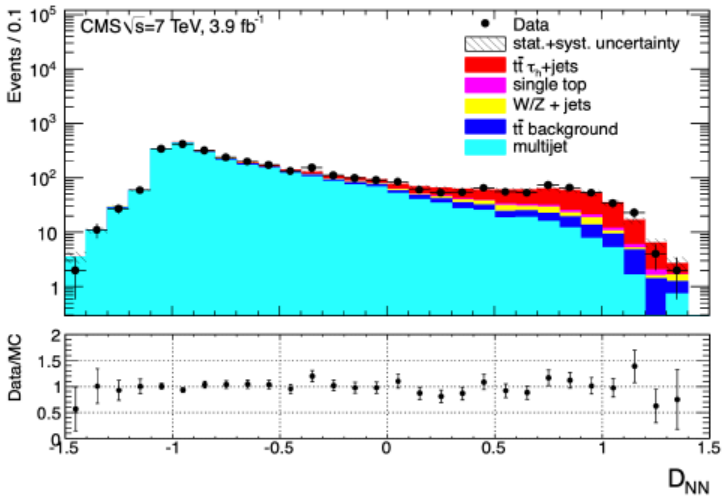
Left: profile likelihood on the parameter of interest for a neural network with (blue) and without (red) the feedback on effect of nuisances provided by INFERNO

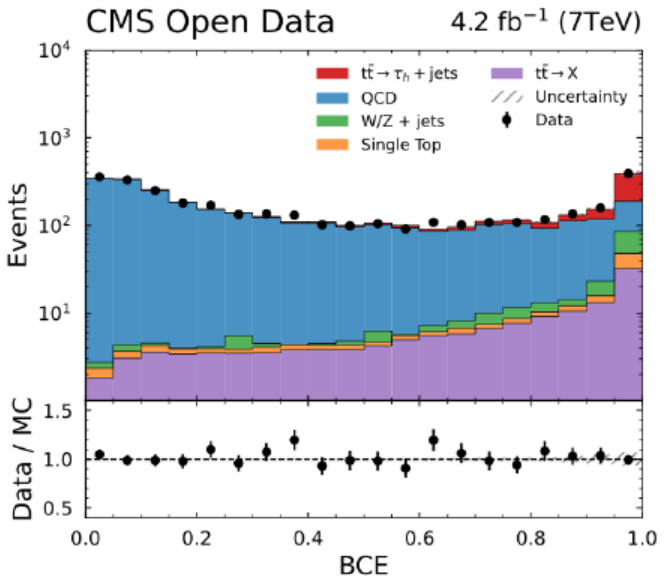
Test of INFERNO on a Real Physics Analysis

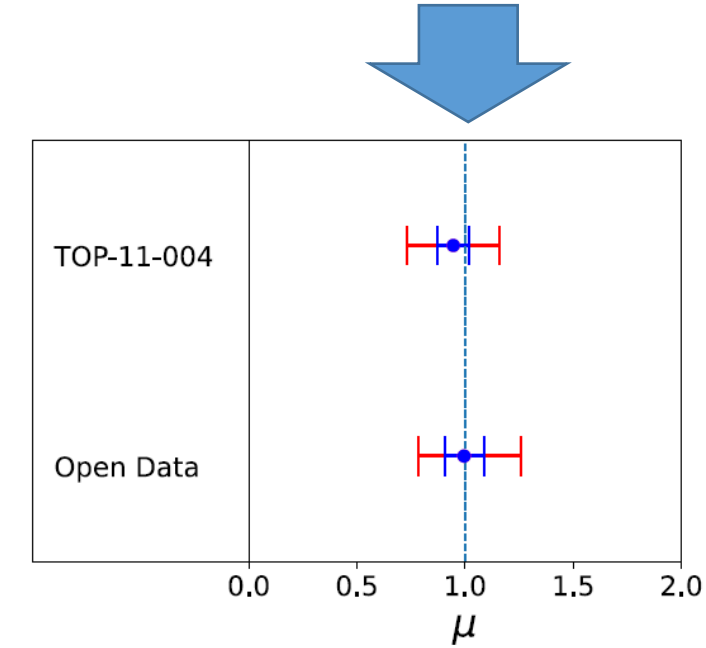
To test the applicability of INFERNO on a real HEP analysis and benchmark it, we chose a resul reproducible on open data, based on NN classification, and with large systematic uncertainties: CMS cross section of tt → τh + jets (TOP-11-004)

An initial test with normal BCE loss produces consistent results on the cross section

https://arxiv.org/abs/1301.5755



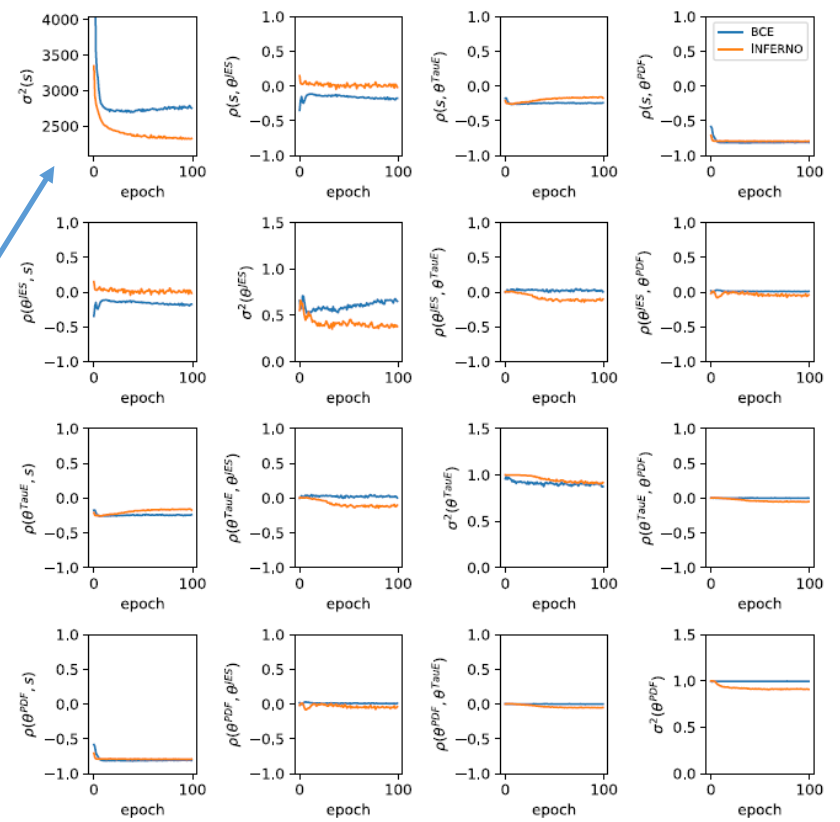


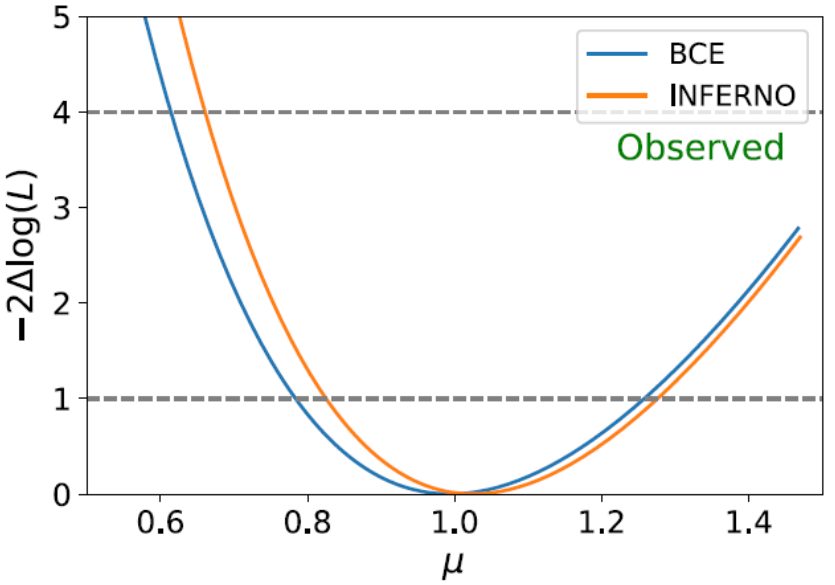


Results and Lessons Learned

INFERNO produces a moderate improvement: it constrains the jet energy scale systematic (by decorrelating the effect of that uncertainty from the summary statistic); yet it cannot reduce systematic uncertainties affecting only the data normalization

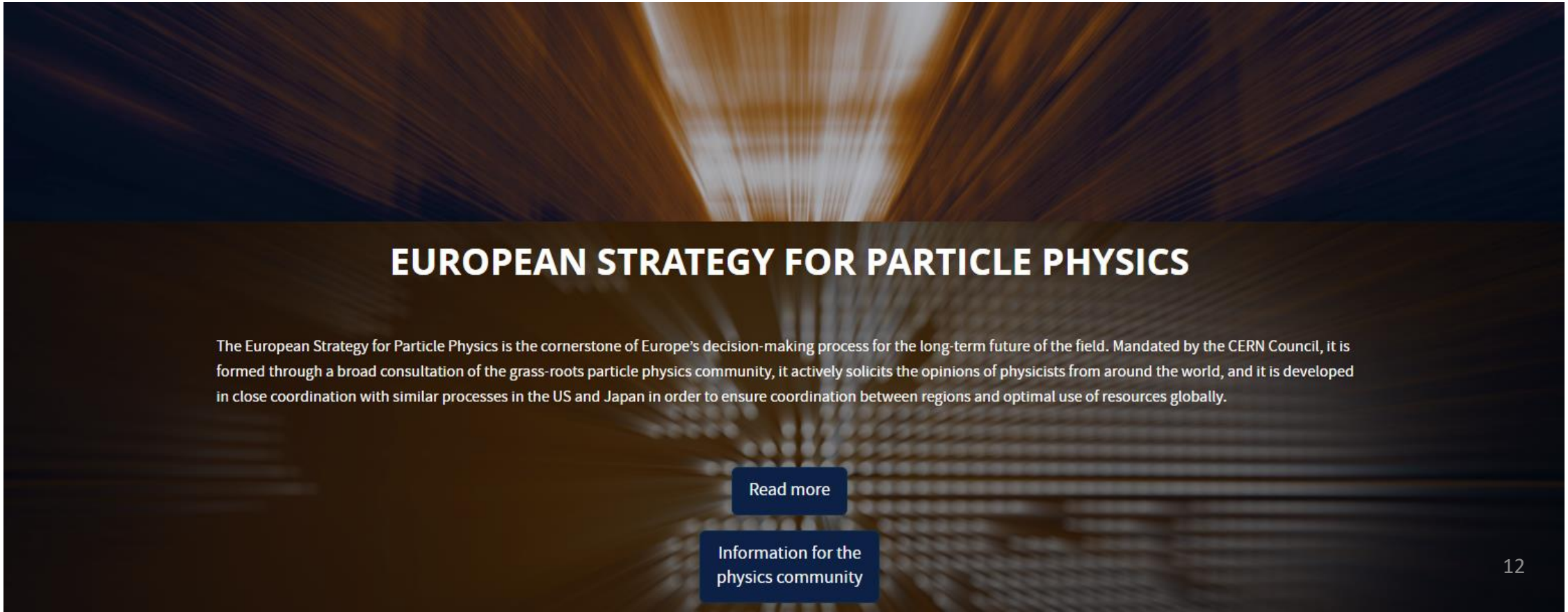
The reproduced analysis is a valid benchmark where to test other approaches. For details see [Layer and Dorigo 2023]





Future Challenges

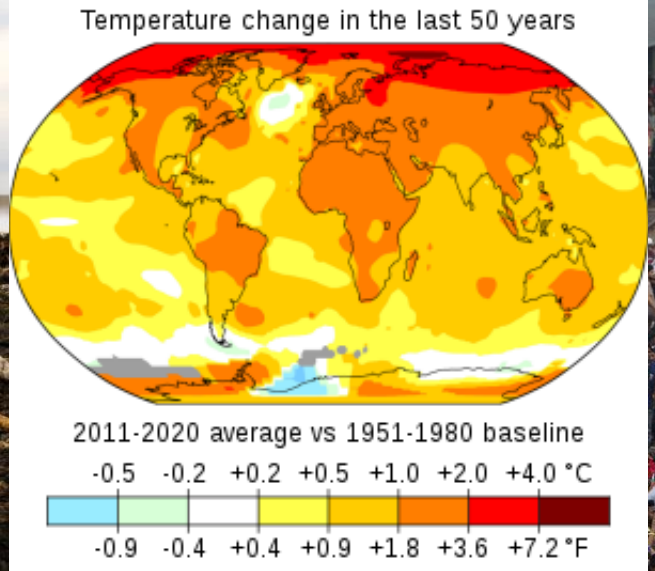
The 2020 update of the European Strategy for Particle Physics (EUSUPP) encourages feasibility studies for new large, long-term projects which will once again push our technological skills to their limits.



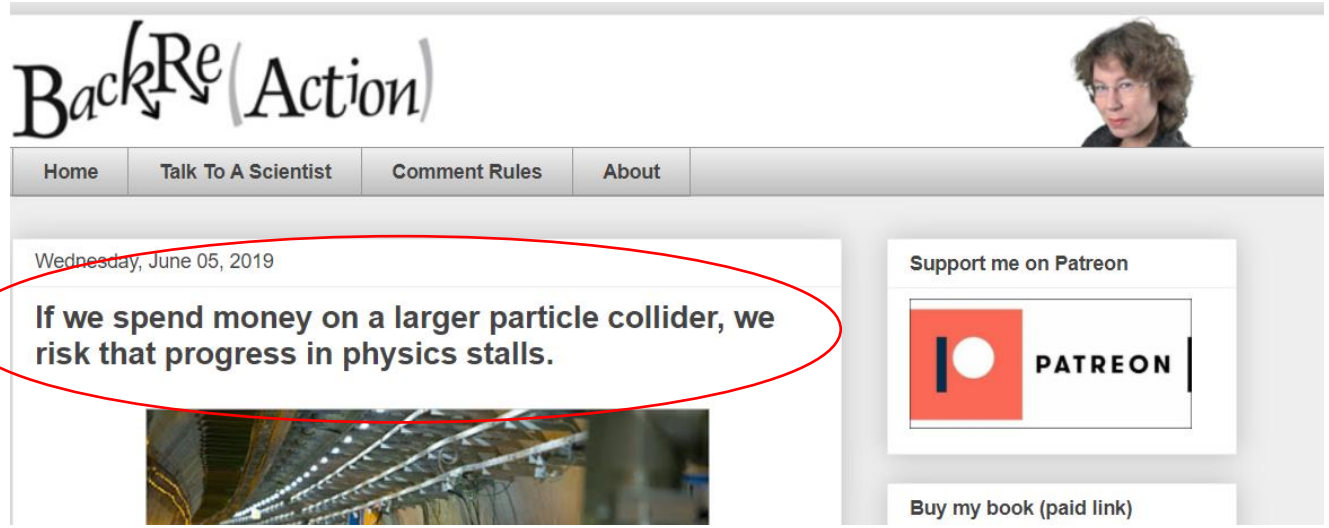
Future Challenges/2

The 2020 update of the European Strategy for Particle Physics (EUSUPP) encourages feasibility studies for new large, long-term projects which will once again push our technological skills to their limits.

At the same time, humanity faces unprecedented global challenges (climate change, pandemics, overpopulation, pollution) which demand resources to seek solutions through applied science innovations, rather than investing in fundamental research.



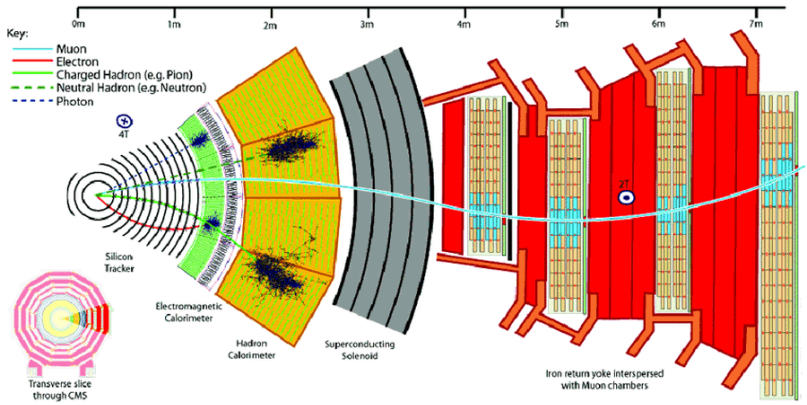
Future Challenges/3



Furthermore, there are indications that wide sectors of society no longer consider the furthering of our understanding of matter at the smallest distance scales a top priority.

In this situation, **ensuring the maximum exploitation of any resources spent on fundamental research is a moral imperative,** and it may be key to ensure that the long-term projects envisioned by the EUSUPP may be undertaken and sustained.

Toward End-to-End Optimization: The *Status Quo* in HEP



Above: *a present-day detector (CMS)*
Below: *a 30-years-old detector for LEP*

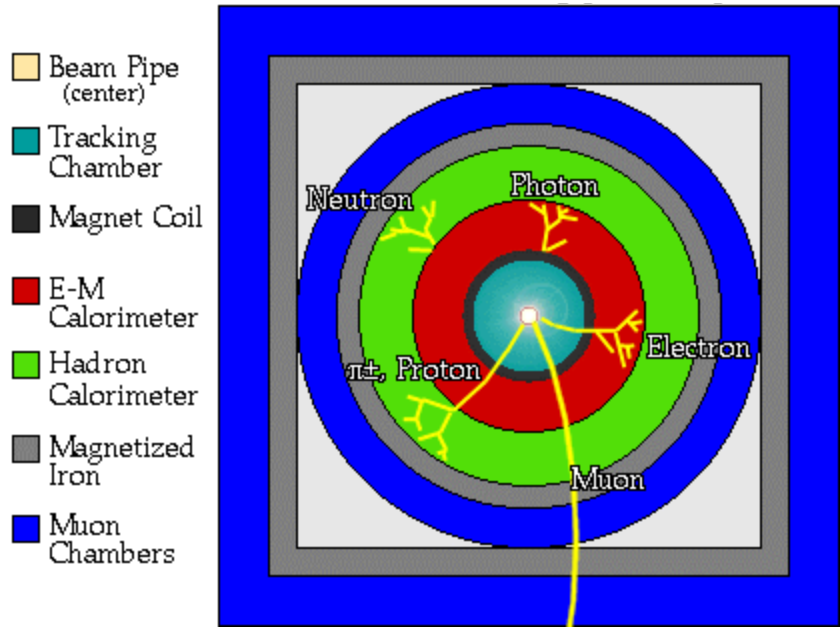
In the past 50+ years the design of new particle detectors leveraged cutting-edge technologies,

yet **a few crucial underlying global paradigms** of experimental design have **remained mostly unchallenged** across decades:

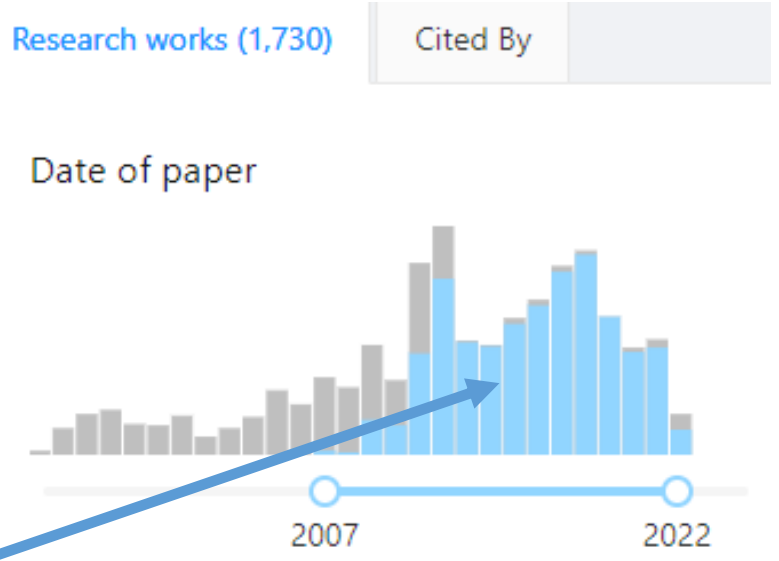
- "Track first, destroy later"
- Redundancy in detection systems, robustness
- Symmetrical layouts

→ No guarantee of optimality!

Those choices **do not directly maximize a high-level utility function**, such as the highest discovery reach for a physical process, or measurement precision



Optimal for What?



The reason why detectors are complex is not only that the studied physics is complex: Science is a demanding job. Physicists want to study *everything* and do it *better* than previously

CMS has over 4000 members, who use the data for a LARGE number of *different* measurements and searches...

***Above**: publication time of **1125** articles by the CMS collaboration (in blue).*
***Below**: a small fraction of the CMS members*

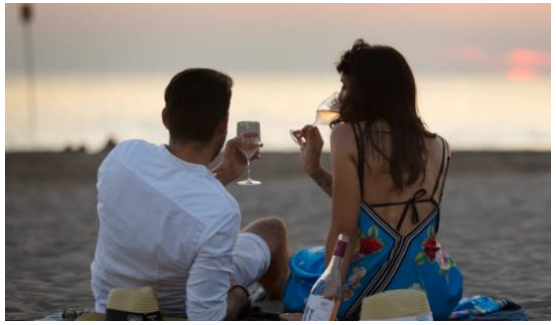
So, what does it mean for a detector to be *optimal*?

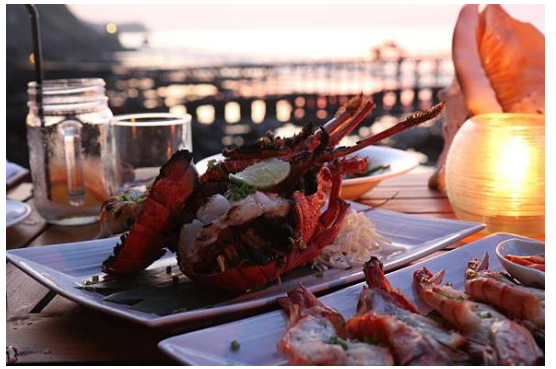
What loss function do we aim to minimize?

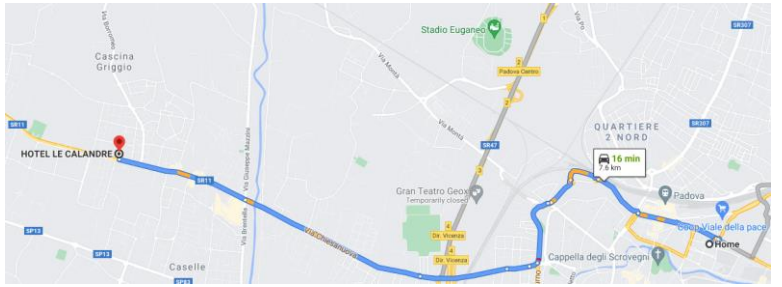
Does it make sense to speak of an experiment-wide utility function?

Concerning the last question: I will convince you that it does

Recipe for a Perfect Dinner







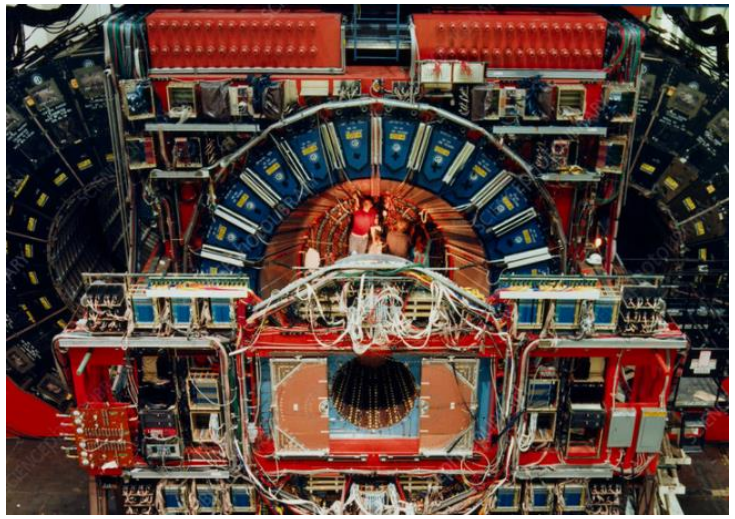
We are not alien to confidently taking complex decisions in a **multi-objective space**. We actually do it routinely...

Of course, we are not deterred by knowing that the exact form of our optimization target is arbitrary

Recipe for a Perfect Trigger

Similarly, we are actually *used* to create multi-target optimization strategies, e.g. when we allocate resources for the trigger menu of a collider detector.

Consider CDF, Run 1 (1992-96): taking in a rate of 300 kHz of proton-antiproton collisions and having to select 50 Hz of writable data created some of the most heated scientifically-driven, rationally motivated, painfully well argumented, and littered with 4-letter words debates I ever listened to.

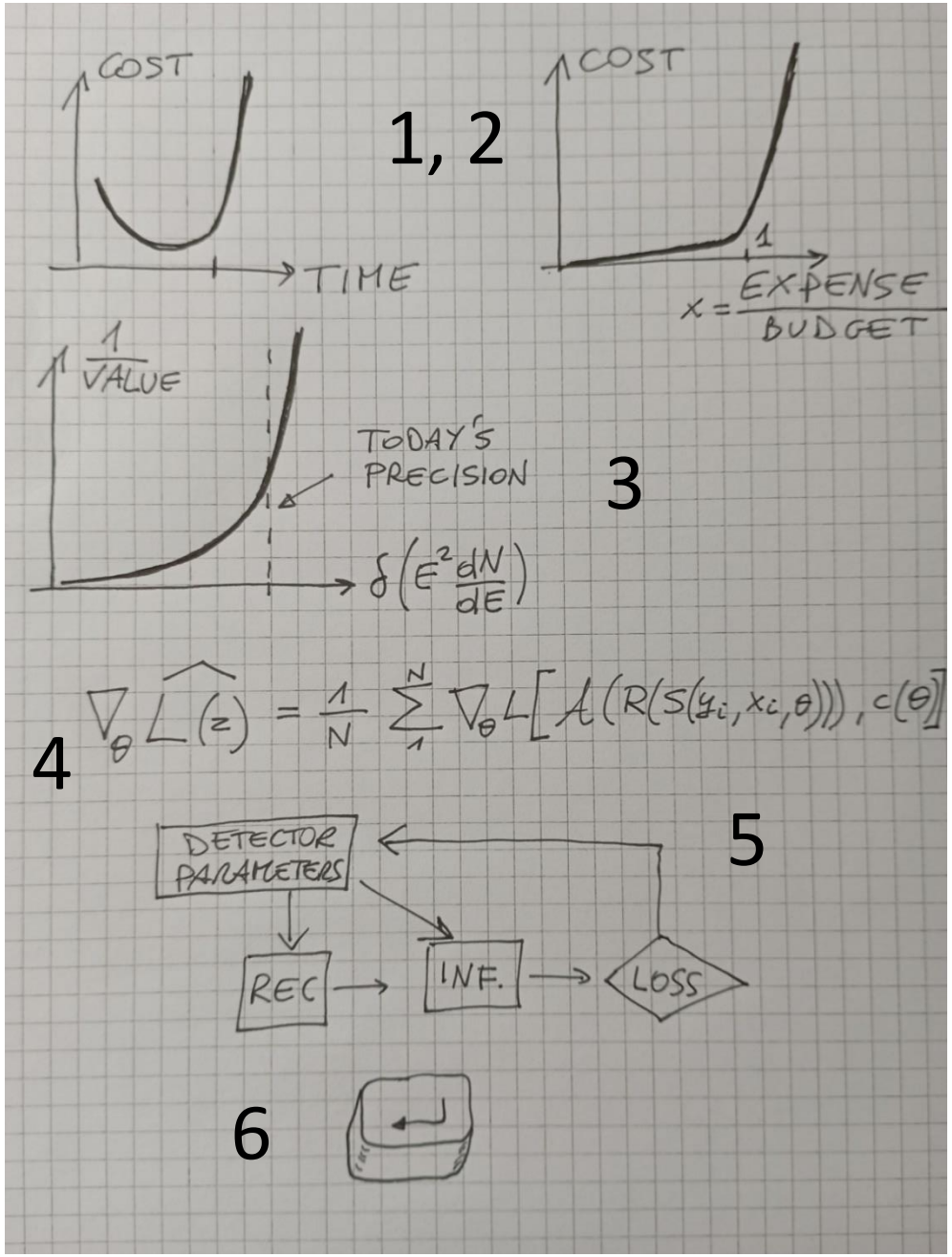


The top quark had to be discovered, but it was not the only goal of the experiment...

Recipe for a Perfect Detector

1. Assess your total **budget** and **time**-to-completion
2. Model as a steep function the **cost** of overriding budget or time
3. Assess the **scientific impact** of each achievable scientific results
4. **Create a differentiable model** of the geometry, the components, the information-extraction procedures, and the utility function
5. Construct a pipeline with those modules, enabling backpropagation and gradient descent functionality
6. Let the chain rule of differential calculus do the hard work for you

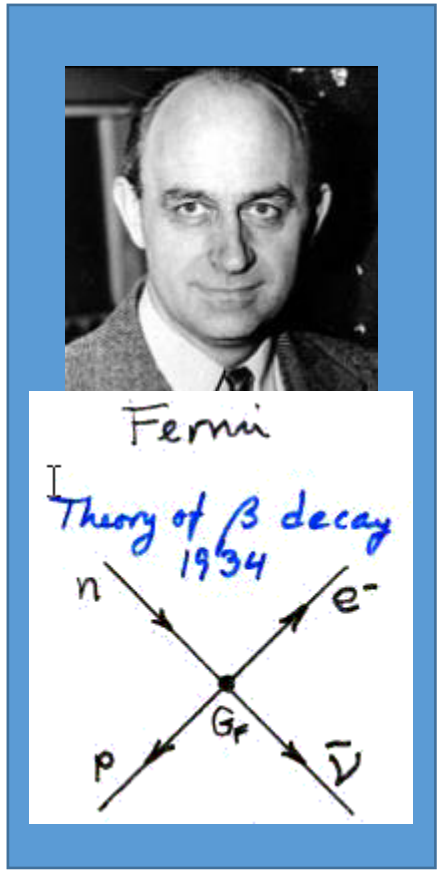
We will discuss this in more detail below, but first let us see what are the typical modi operandi for experiment design



Theory-Driven Experiment Design

In a theory-driven scheme of scientific investigations, we have a theory of the world which produces predictions of observable phenomena A, A', A''...

E.g.: Fermi theory, existence of the electron neutrino



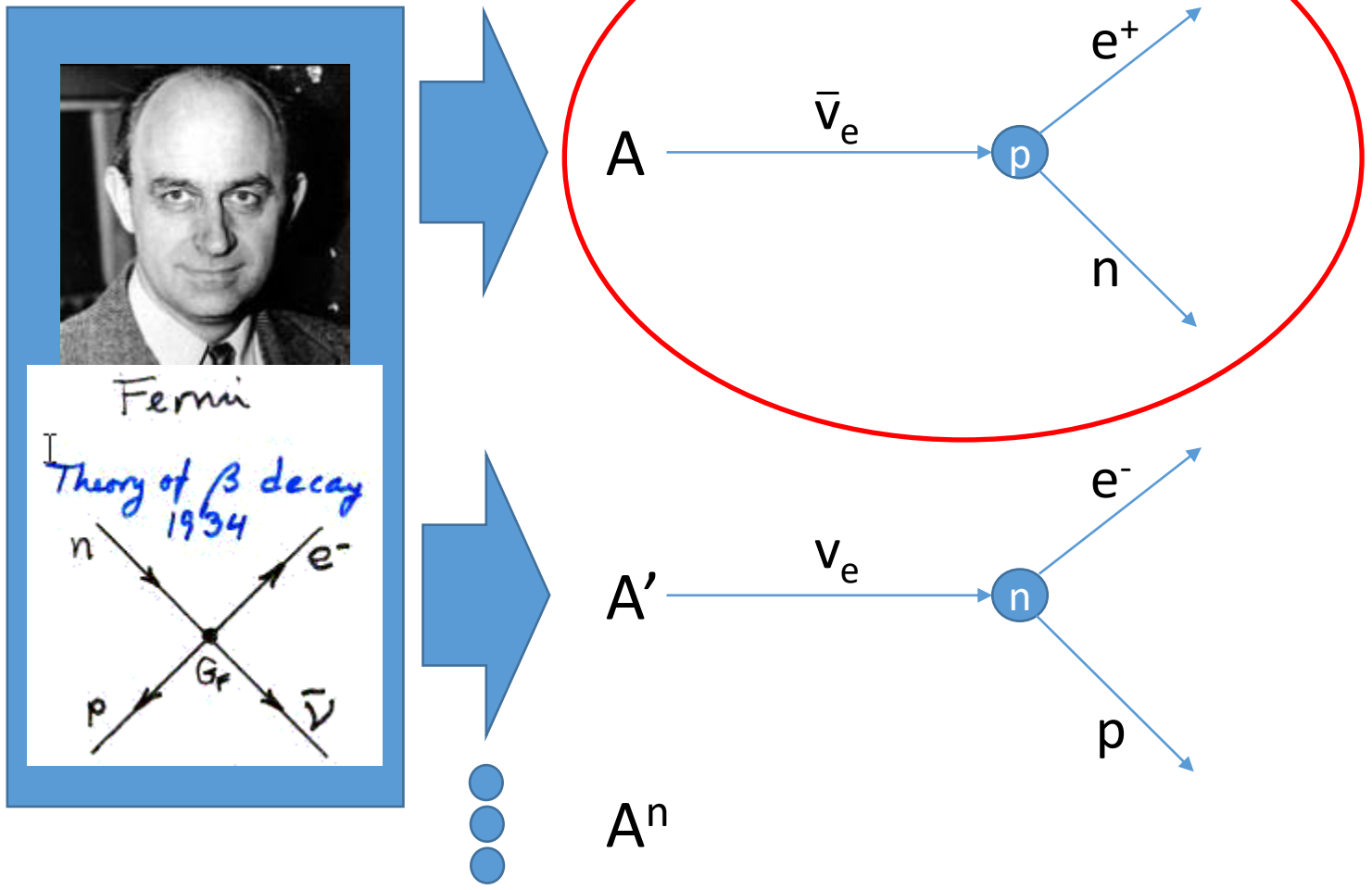
A → $\bar{\nu}_e$ + p → e^+ + n

A' → ν_e + n → e^- + p

A^n

Theory-Driven Experiment Design / 2

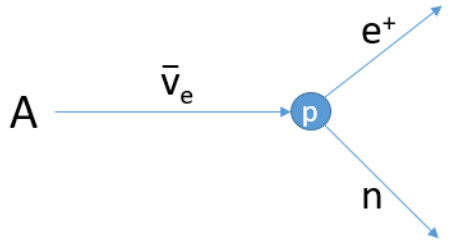
We proceed to test a prediction of the theory: the existence of the neutrino. This implies many reactions must exist. We may focus on the lowest-hanging fruit, say A, and devise a means to study it.



Theory-Driven Experiment Design / 3

Given the well-defined goal of discovering a specific phenomenon, it is relatively easy to at least sort out what technology and mechanisms we intend to use to produce and evidence it

In thie Reines-Cowan experiment, one early question was how to produce an intense flux of neutrinos

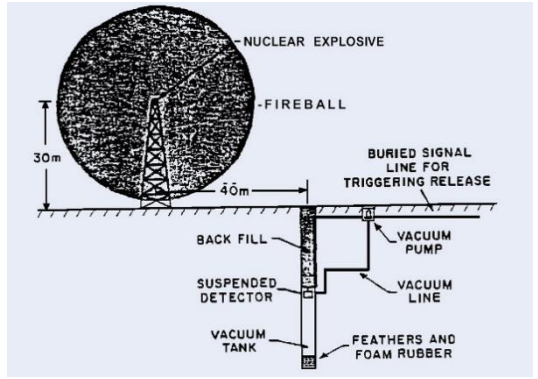


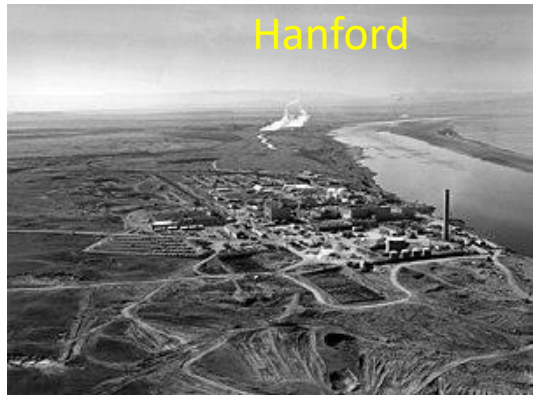
$\sigma = 6 * 10^{-44}\ cm^{-2}$

Want many neutrinos!

Better be repeatable

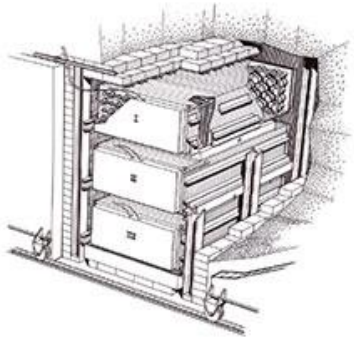
Background is a concern

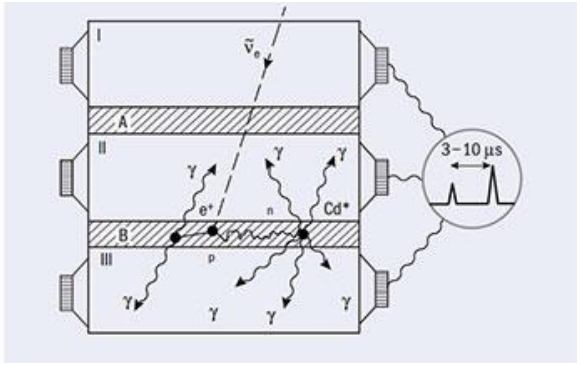




Theory-Driven Experiment Design / 4

The remaining job is to connect the killer observables that the process yields to suitable layouts





Importance of optimization here is limited to some aspects of the problem

- want a massive, transparent target (to see the photons) → use water
- need to find a way to detect neutrons → exploit material with high capture rate of neutrons, emission of photons, delayed → dissolve $CdCl_2$ in water

Once these basic ideas are sorted out, the rest is not crucial.

Precision Measurements

A connected use case to the previous one is that of measuring some important fundamental parameter with the utmost precision

→ help comparisons with theoretical predictions

→ input to future studies

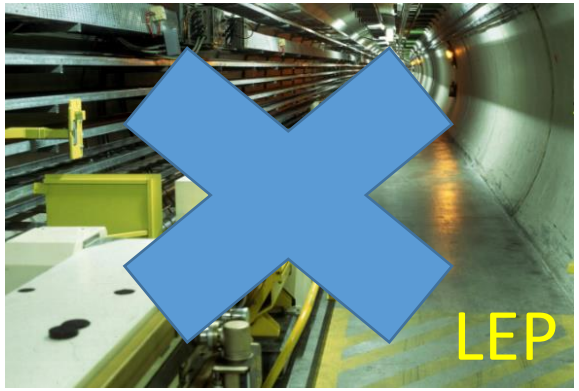
LEP is a great example of that modus operandi
Here optimization may target specific metrics (e.g., precision of M_Z, δN_ν) and an «experiment-wide» loss function may then be relatively easy to agree upon.

The Third Way: Fishing Expeditions

In experiment-driven science, we aim for observing new phenomena, incrementally acquiring information that may later be studied to improve our theories.

We thus focus on exploring Nature as widely as possible. Precision is no longer the single most important bit

→ Our greed for diverse, high-quality data dictates the need for multi-purposedness:



Makeshift Surrogates of Objectives

When physicists design sensors and electronics, operate choices on budget allocations, define requirements for the performance of detection elements, or choose composition and layout of a complex instrument, **they are implicitly trying to find an optimal working point in a loosely-constrained feature space of hundreds of dimensions**.

Such a task is clearly **super-human**.

Because of the implicit nature of the models (likelihood is intractable), the only way forward until recently has been to set their aim on **makeshift surrogates** of their real objectives.

- Simulations are high-fidelity in HEP, but they only allow to probe the result of specific choices, not to map interdependencies and find extrema of an utility.

Makeshift Surrogates of Objectives

When physicists design sensors and electronics, operate choices on budget allocations, define requirements for the performance of detection elements, or choose composition and layout of a complex instrument, **they are implicitly trying to find an optimal working point in a loosely-constrained feature space of hundreds of dimensions**.

Such a task is clearly **super-human**.

Because of the implicit nature of the models (likelihood is intractable), the only way forward until recently has been to set their aim on **makeshift surrogates** of their real objectives.

- Simulations are high-fidelity in HEP, but they only allow to probe the result of specific choices, not to map interdependencies and find extrema of an utility.

Evolving from this *modus operandi* to directly goal-informed decisions enables potentially enormous performance gains.

The Design Space is Large

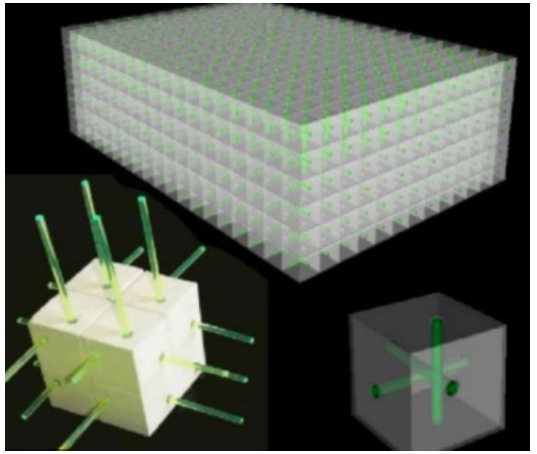
New technological advancements are enabling better designs of our instruments by reducing the cost of complex layouts.

- 3D printing of scintillation detectors is being explored for neutrino physics
- Very thin layouts of resistive AC-coupled silicon detector elements provide large gains in spatial and temporal resolution

→ **The geometry space has become larger and more complex to explore.**

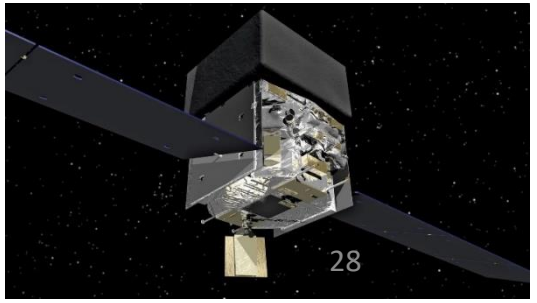
Higher-performance demands are also rising as, *e.g.*,

- Tracking at high collider luminosity requires AI solutions
- As we move fundamental physics research to space, payload and power consumption become driving constraints
- Higher fluxes and energies, and studies of high-mass particles, demand us to invest in more granular, higher-performance hadron calorimeters (more on this later)



Above: *3D printed scintillation detector*

Above: *tracks of particles in a simulated high-luminosity collision*
Below: *the FERMI satellite is a calorimeter detecting gamma rays*



How Large Can Gains of Full Optimization Get?

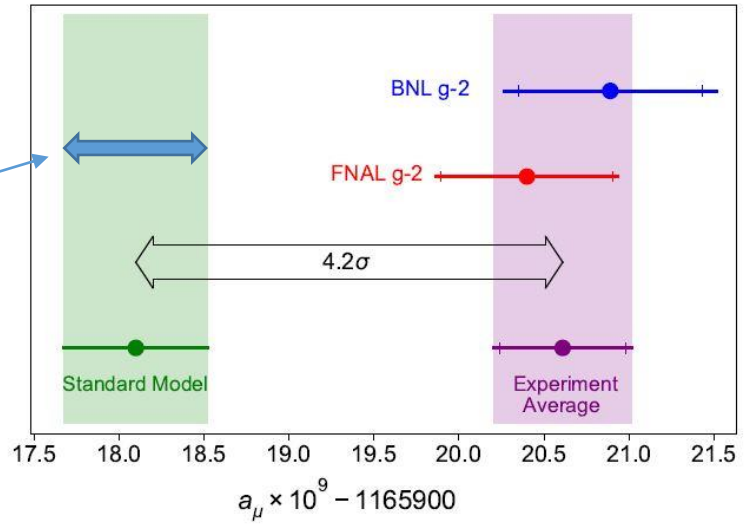
I recently provided an example [Dorigo 2020] of how experimental design as is carried out today leaves **ample room for improvement** from systematic study of geometry and materials

The chance of doing so was offered by refereeing the detector proposed by the MUonE collaboration [Abbiendi et al. 2018], which aims at determining with high precision the cross section of elastic muon-electron scattering

Through the direct exploration of the parameter space of detector geometry, I could show how **large gains** are possible by moving away from choices dictated by past experience

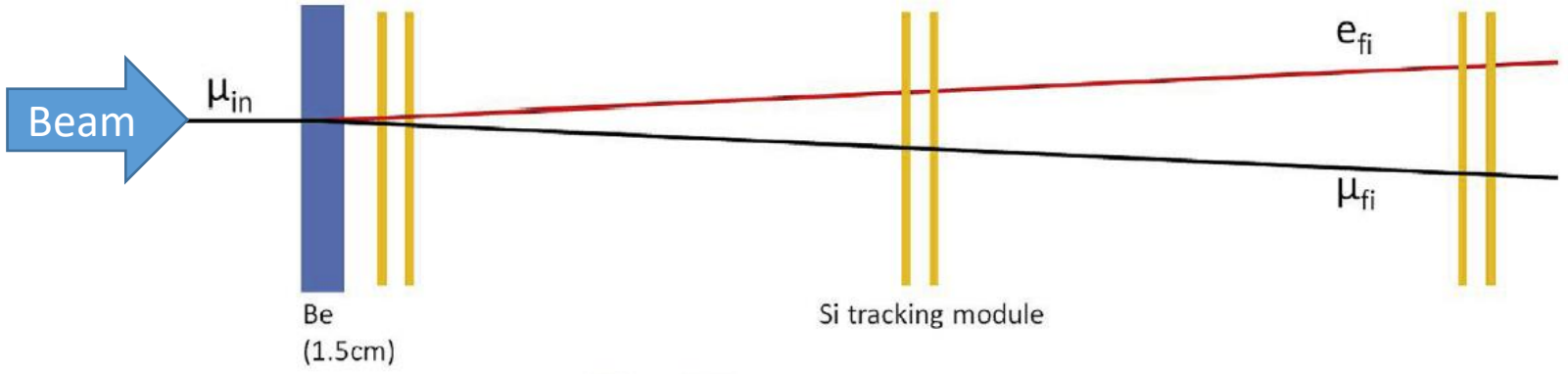
One Example of Geometry Optimization: MUonE

MUonE aims to determine with high precision the probability of elastic muon-electron scattering, as this number may reduce the theory systematics of the g-2 muon anomaly

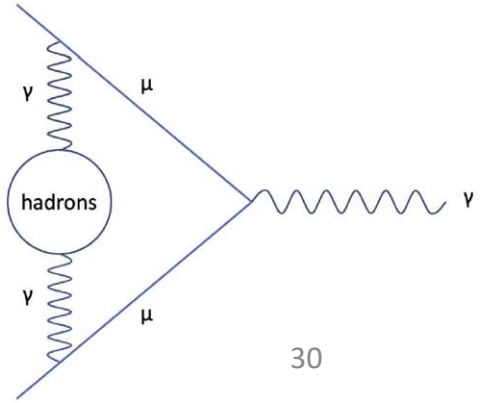


***Above**: a long-standing anomaly in the Standard Model, the muon g-2 value*
***Below:** a muon-photon interaction, with a hadronic quantum loop*

The experiment must be sensitive to the rate of interactions as a function of momentum transfer, with 10^{-4} precision



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

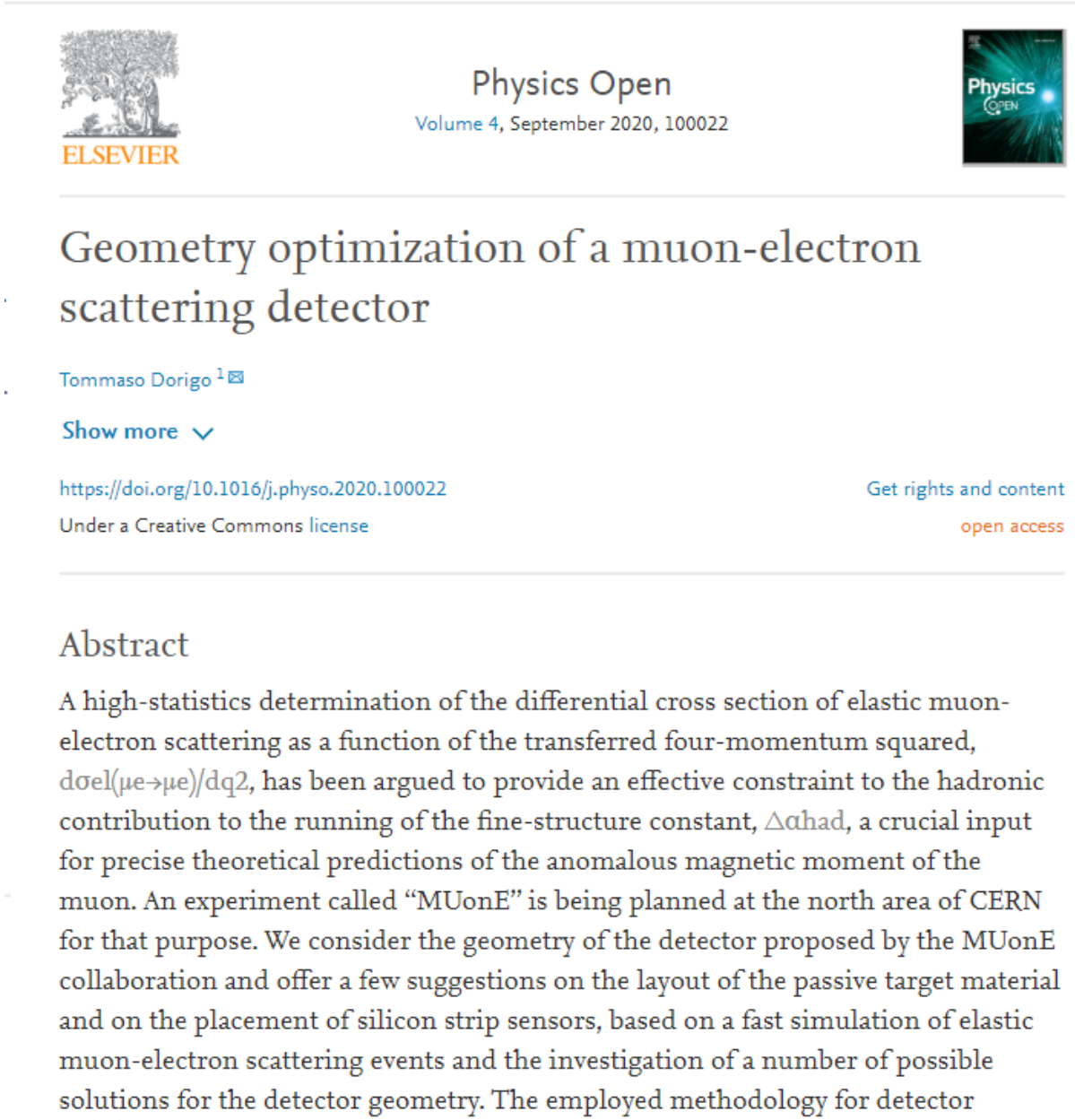


MUonE Optimization

To study alternative layouts, I wrote a simulation of muon scatterings and event reconstruction, then optimized the geometry of the apparatus accounting for every part of the problem affecting the precision of the inference

This identified a layout different from the one chosen by detector experts, with **a factor of 2 improvement in the relevant metric** without increase in detector cost

After some digestive trouble, MUonE adopted most of my proposed improvements



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

Abstract

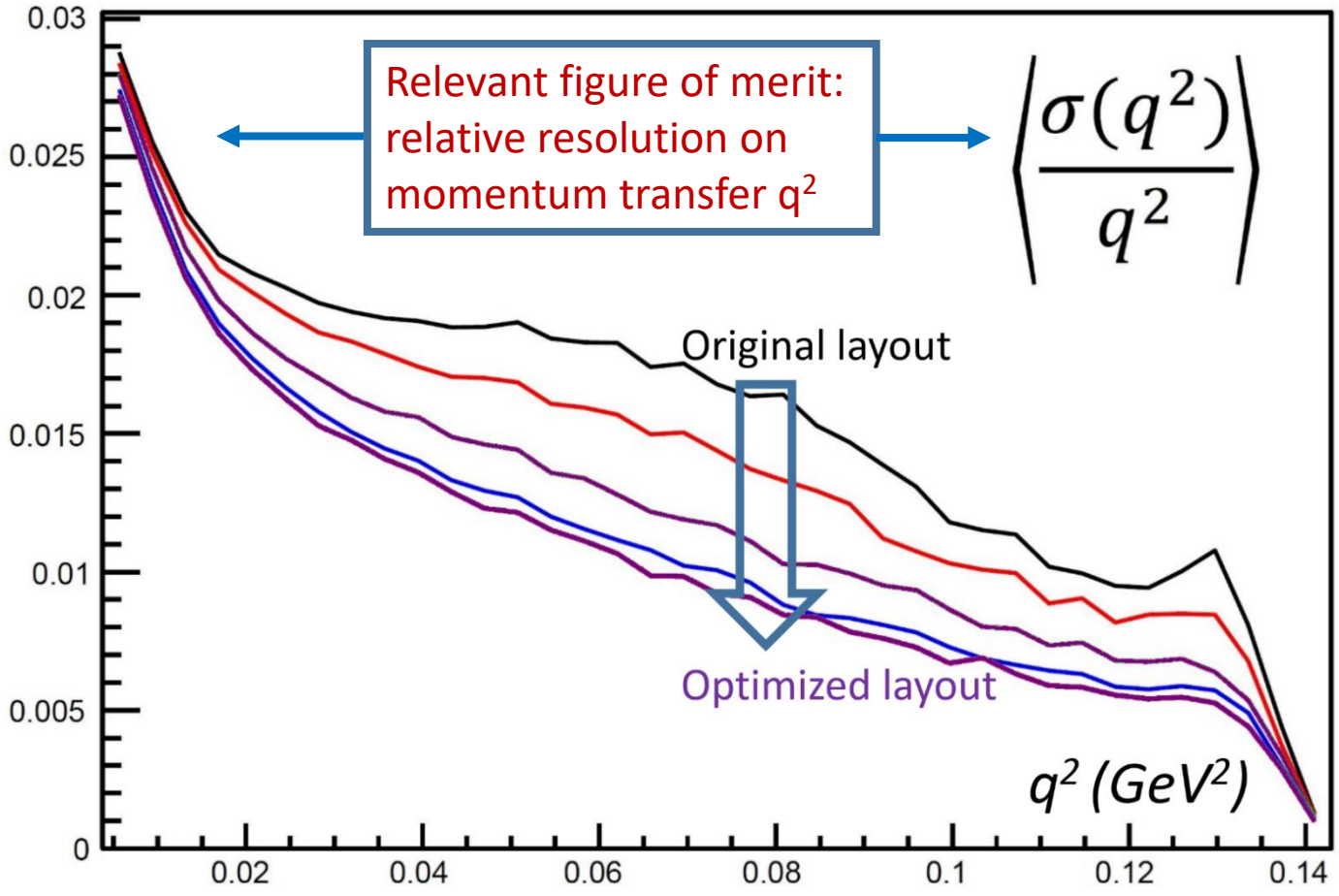
A high-statistics determination of the differential cross section of elastic muon-electron scattering as a function of the transferred four-momentum squared, $d\sigma_{el}(\mu e \to \mu e)/dq^2$, has been argued to provide an effective constraint to the hadronic contribution to the running of the fine-structure 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

MUonE Optimization

A factor of 2 in HEP is HUGE

Incidentally, a similar improvement was found in a different study [Shirobokov20] when optimizing magnet design for shielding the SHIP experiment (again, a limited complexity use case)

We can only guess how large are the gains possible if a fully differentiable model is used for detectors of significantly higher complexity



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

Note: the MUonE optimization was «easy», because there are 3 particles involved and only few Be and Si layers... Discrete optimization scans were sufficient there. **How to scale?**

Event Reconstruction in MUonE

The scattering involves an incoming muon and outgoing muon and electron → it seems proficuous to fit the trajectories to a common vertex

However, if one is not thinking about the potential of alternative layouts (read: end-to-end optimization) to the original proposal, the advantage of a full 3D vertex fit looks like a detail worth leaving to final tweaks. It is instead crucial!

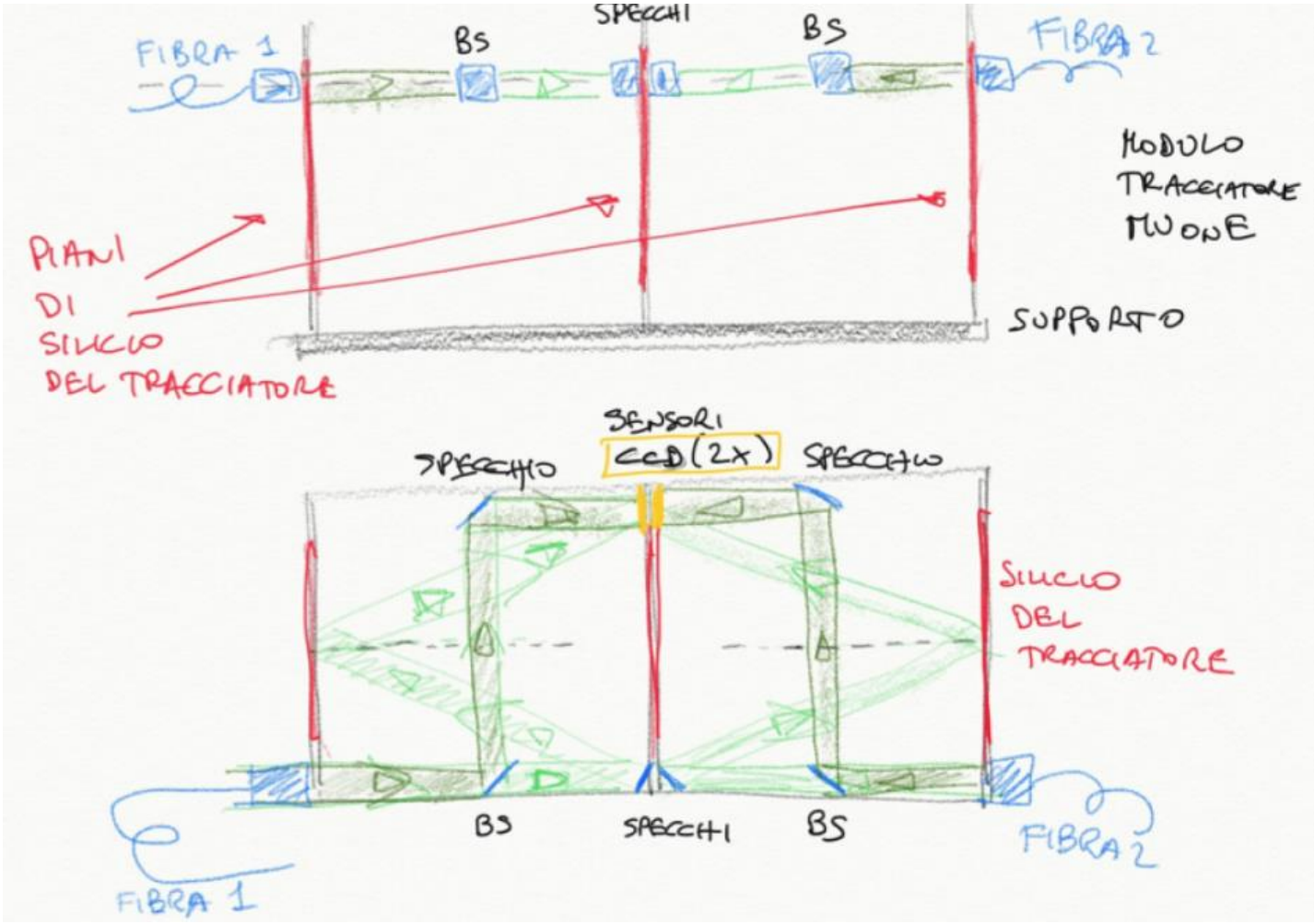
Here is a **correct «holistic» way** of thinking at the matter:

- Interactions mostly take place in Be target, which is thick so it does not offer z-position constraints
- ~~hence a fit involving the z coordinate is useless~~ Hence we should rethink the layout to exploit the constraint: use thin layers spaced in vacuum or air!
- Large systematics on measured angles are lurking, due to relative z positioning errors; hard to position elements with sufficient precision. ~~Hence thin layers layouts are impractical~~ By exploiting vertex fit to many events we should manage to reduce effect of positioning systematics

Reducing Positioning Errors

The MUonE collaboration envisioned a holographic laser system to measure the position along z of modules and targets with 10µm precision (roughly required threshold for wanted q^2 resolution).

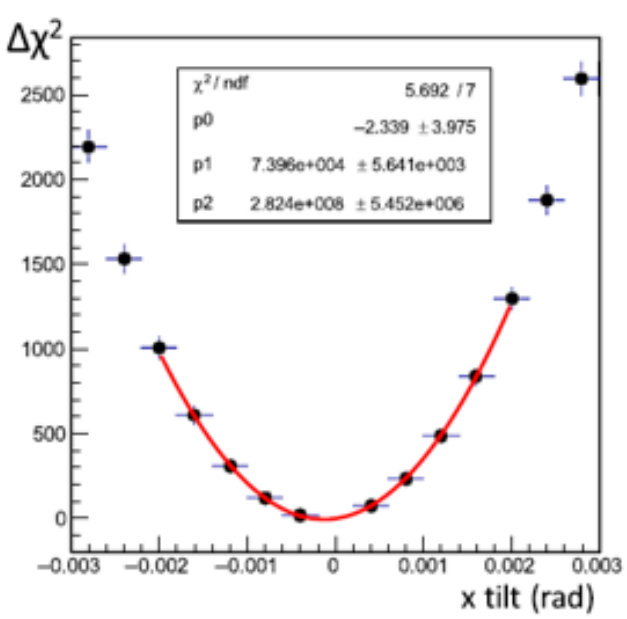
Cost: one 10k euro system per meter-length muon station, 40 stations → 400k euro budget

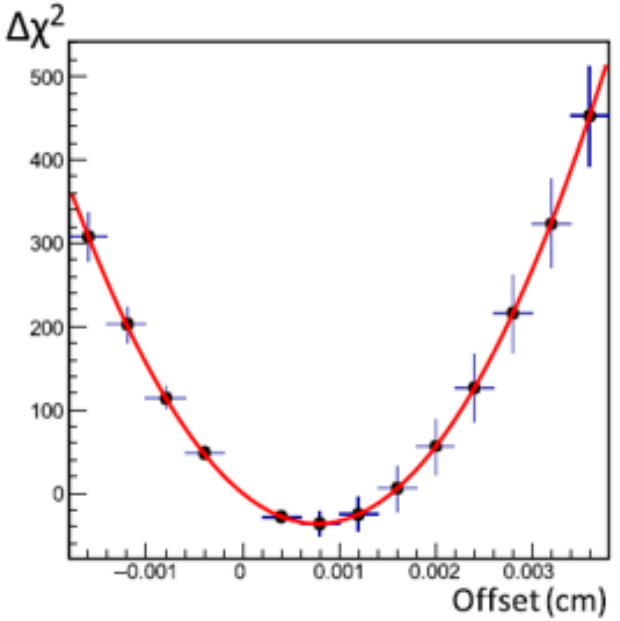


Effect of Thin Layers and Vertex Fit

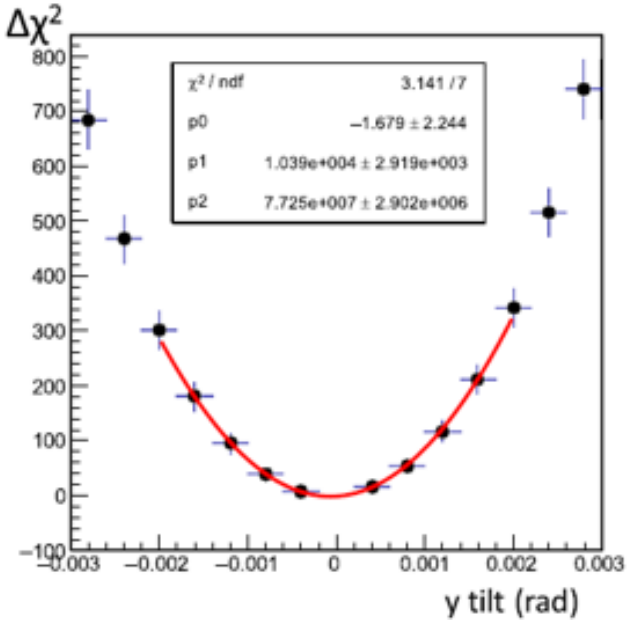
Once you realize the benefit of thin layers, a simulation proves that the holographic shebang is useless:

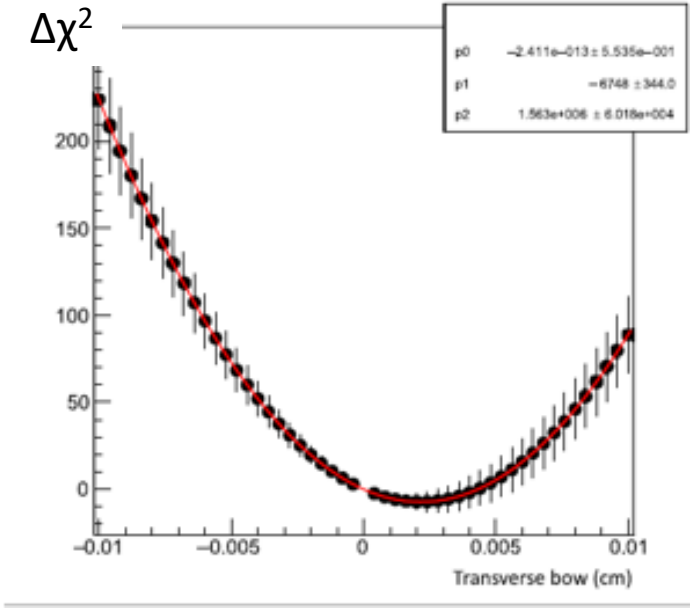
by using 5 minutes worth of beam data, layouts that involve thin target layers allow to fit for the position z of individual elements, as well as for their tilt and bow, to sub-micron accuracy!





Top left: global chi2 vs x tilt; right: vs z offset;
Bottom left: vs y tilt; right: vs transverse bow.





What Do We Learn from this Example?

Could an automatic scanning of layouts come to appraise the benefits of thin target layers?

- the parametrization of the system needs to have the flexibility of considering the segmenting of targets along z in the first place

  → this is possible but might be overlooked (in fact it was)!

- the reconstruction model needs to consider all physical constraints ab initio

  → this is reasonable, but it again highlights the need of optimal reconstruction in searching for optimal geometry

- the model of detector-related systematic uncertainties must include provisions for exploiting the data for self-calibration → this is extremely hard to conceive building in the model if we don't know what we are trying to shoot at!

→ Even in this simple use case, expert insight is required

→ Software and hardware need to be optimized together!

Expert-in-the-Middle Schemes

In today's way of doing business, we strongly leverage prior experience in conceiving a new detector, and exploit most advanced technologies, but do not explore the param space fully:

Expert(s) proposes design → simulation tests it → performance assessed on reachable proxies → possible improvements identified in discrete space → new simulation probes them → evaluation → modification → big simulation campaign, analysis results on valuable end targets (but at that stage there is little that can be tweaked) → Final configuration.

Building a **pipeline for end-to-end optimization** is thus already a big change of paradigm:

- we imply the intention of specifying a loss function for the whole experiment
- we force ourselves to specify how cost, timeline, and other factors play in the loss
- we take responsibility to model inference extraction already at a design stage
- we state the intent of considering how design choices affect the whole chain

Note how the end-to-end approach involves much more than detector experts and post-hoc validation: computer scientists, analysis experts, and a global loss composition are called for.

→ Ironic how end-to-end optimization be perceived as an attempt to substitute experts with AI!

HOW to Do It

Computer Science to the Rescue

Progress in CS redefined performance standards of our technologies in, e.g.,

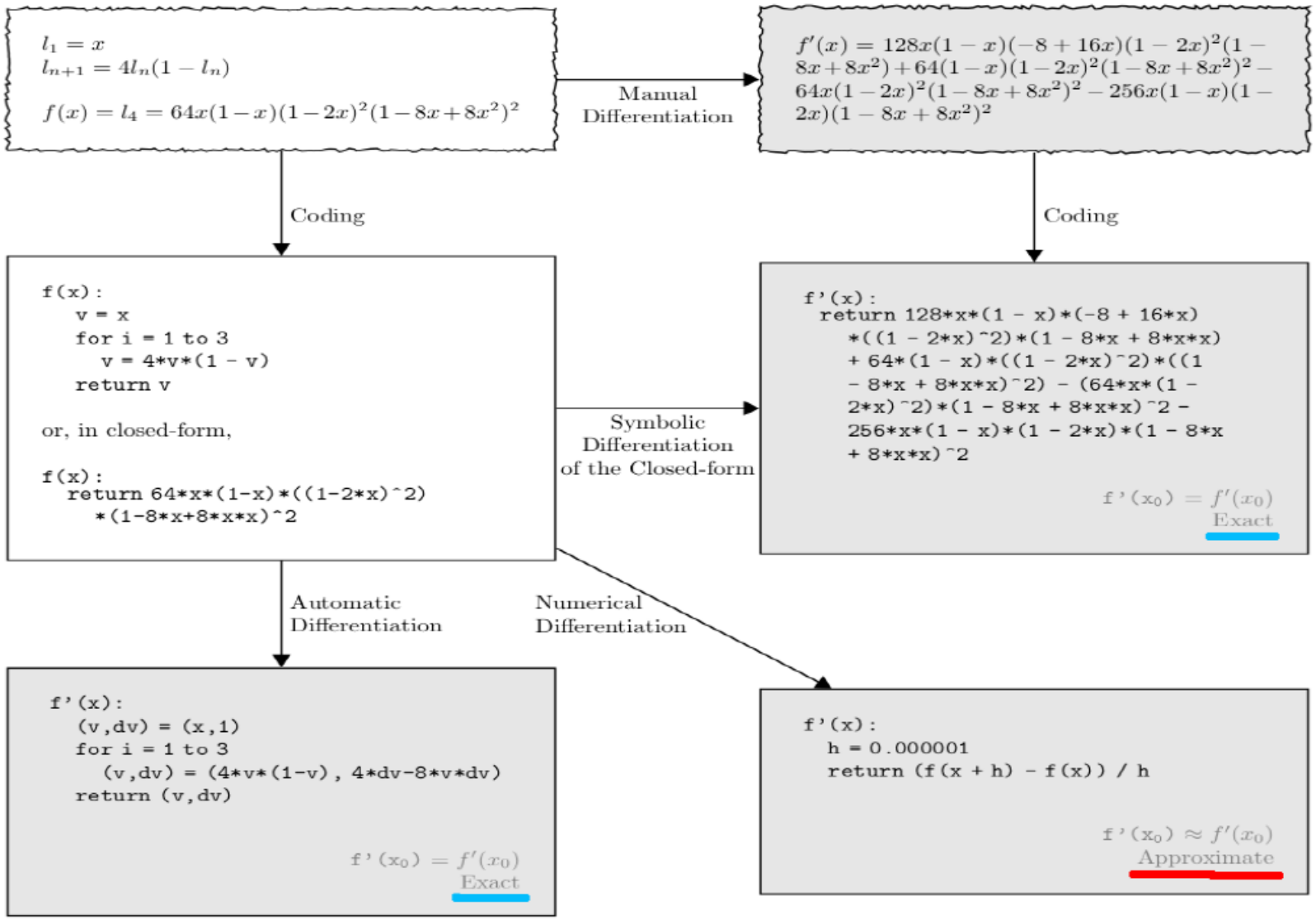
- language translation
- speech recognition
- self-driving vehicles

The developed AI tools are application-specific: the AI potential in providing new solutions to old tasks **depends on our ability to create the right interfaces**.

A new paradigm shift is offered by differentiable programming, which eases the systematic search of minima of arbitrarily complex multi-dimensional functions

→ By casting the whole problem in a differentiable framework a full end-to-end optimization becomes possible.

Automatic Differentiation



The compiler figures out how functions vary depending on parameters, and carries out the complicated task of propagating derivatives around.

Those of us who have done this manually for a long while can't be happier by seeing the rise of Pytorch, TensorFlow, etc.

Left: *different ways of obtaining derivatives of a coded function (Graph courtesy A.G. Baydin)*

Four-Slide Description of a Differentiable Model for Design Optimization - 1

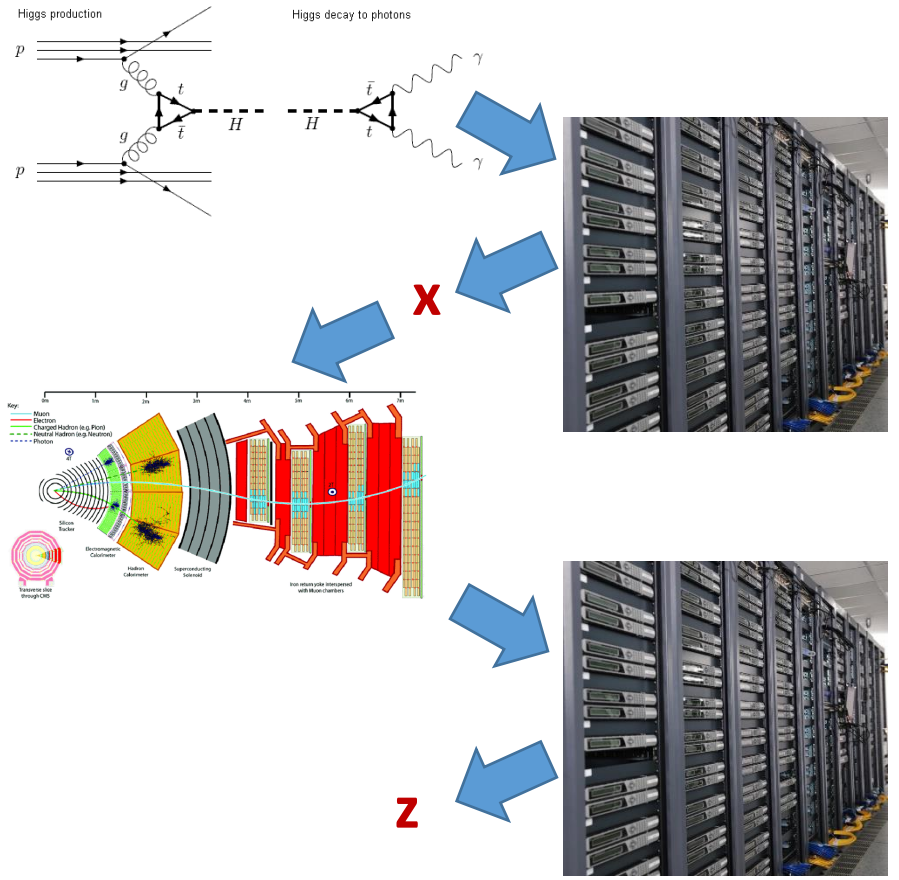
An end-to-end detector design optimization task can be briefly formalized in the following way.

(1) We start with a simulation of the physics processes of relevance, which generates a multi-dimensional, stochastic input variable **x**, distributed with a PDF **f(x)**

(2) The input is turned by the simulation of the detection apparatus into sensor readouts **z** distributed with PDF

$$p(z|x,\theta).$$

z are observed low-level features of the physical process; they depend through **p( )** on parameters **θ** describing physical properties of detector and geometry.



Four-Slide Description of a Differentiable Model for Design Optimization - 2

(3) Detector readouts **z** are used by a reconstruction model **R( )** that produces high-level features

$$\zeta(\theta) = R[z,\theta,\nu(\theta)]$$

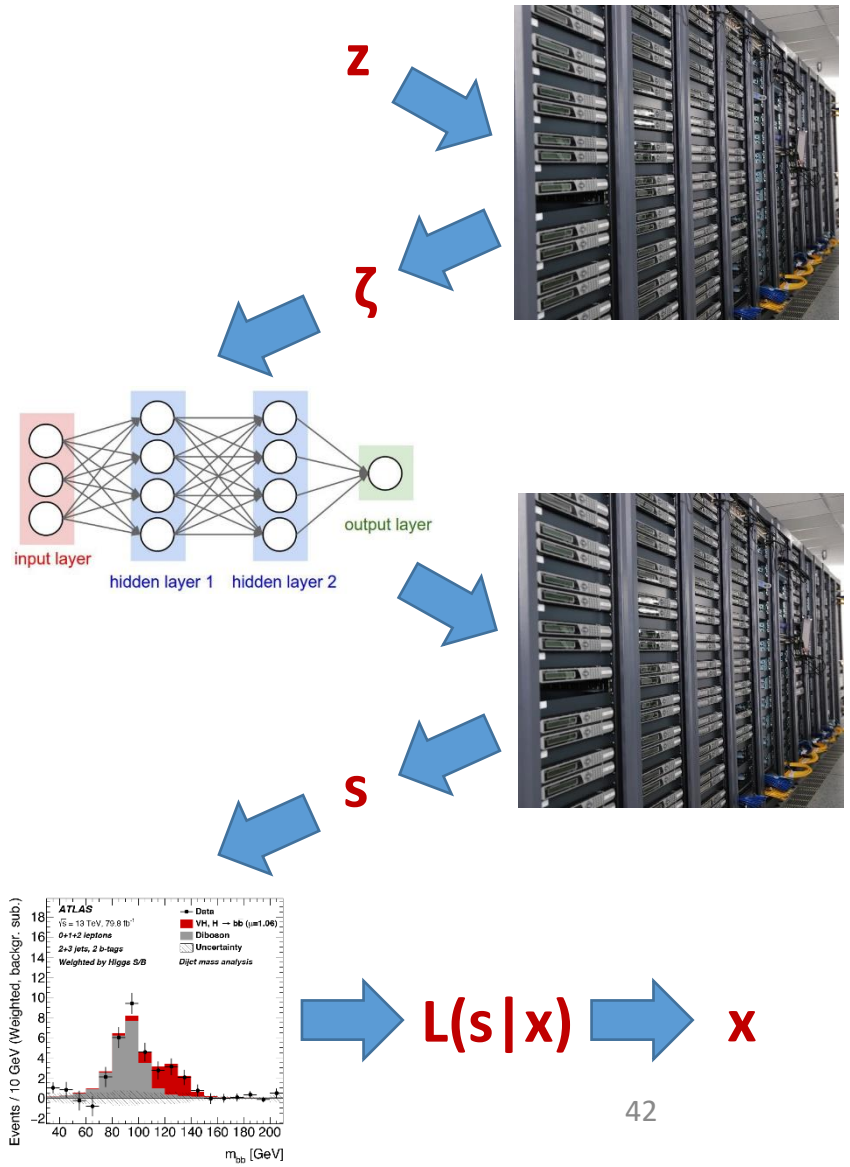
(*e.g.* particle four-momenta), by employing knowledge of detector parameters **θ** as well as a model of nuisance parameters **ν(θ)** which affect the pattern recognition task.

(4) In turn, high-level features **ζ(θ)** constitute the input of a further dimensionality reduction, typically powered by a neural network **NN( )**

Once trained for the task at hand, the network produces a low-dimensional summary statistic

$$s = NN[\zeta(\theta)].$$

(5) With **s**, inference can finally be carried out to infer back **x**. This completes our model of the inference process...



Four-Slide Description of a Differentiable Model for Design Optimization - 3

The problem of detector optimality is that of finding estimators $\widehat{\boldsymbol{\theta}}$ that satisfy

$$\widehat{\boldsymbol{\theta}} = \boldsymbol{arg}\min_{\boldsymbol{\theta}} \int \boldsymbol{L}[\boldsymbol{NN}(\boldsymbol{\zeta}), \boldsymbol{c}(\boldsymbol{\theta})]\boldsymbol{p}(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{\theta})\boldsymbol{f}(\boldsymbol{x})\boldsymbol{dxdz}$$

c(θ) models the cost of a detector of parameters **θ**, and the loss function **L[NN,c]** appraises the result, obeying cost constraints and other limitations.

The PDF **p(z|x,θ)** is not available in closed form –models are implicit–, so we rely on forward simulation: we approximate $\widehat{\boldsymbol{\theta}}$ with a sample of n events:

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{a}} = \boldsymbol{arg}\min_{\boldsymbol{\theta}} \frac{\boldsymbol{1}}{\boldsymbol{n}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{L}\big[\boldsymbol{NN}\big(\boldsymbol{R}(\boldsymbol{z_i})\big), \boldsymbol{c}(\boldsymbol{\theta})\big]$$

where $\mathbf{z_i}$ is distributed as $\mathbf{F(x_i,\theta)}$ to emulate **p(z|x,θ)** as $\mathbf{x_i}$ is sampled from its PDF **f( )**. We thus obtain detector parameters which minimize the loss.

Four-Slide Description of a Differentiable Model for Design Optimization - 4

In some cases we will need to approximate the non-differentiable stochastic simulator **F( )** with a local surrogate model:

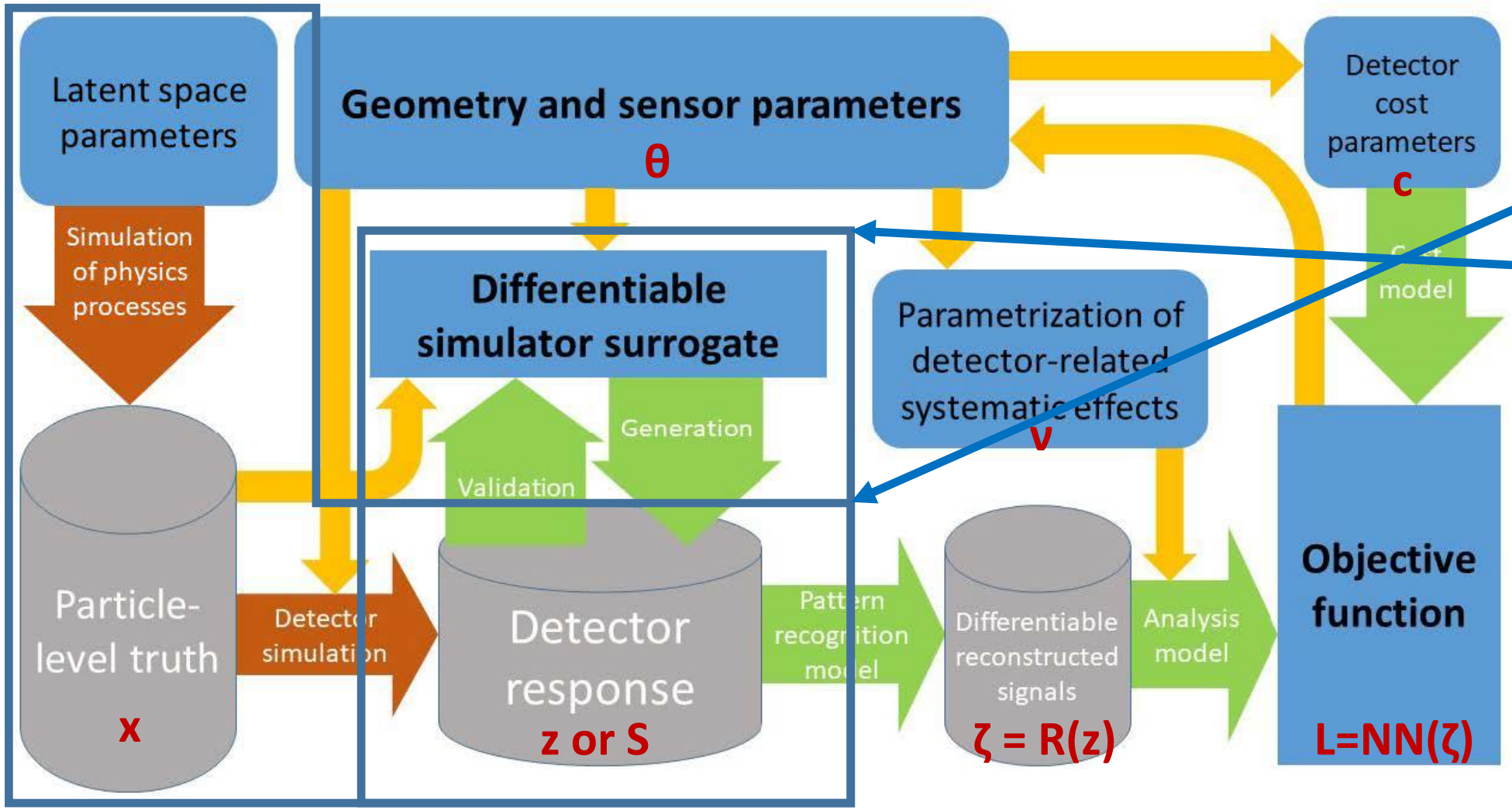
z = S(y,x,θ),

that depends on a parameter **y** describing the stochastic variation of the approximated distribution. This allows to descend to the minimum of the approximated loss $\widehat{L(z)}$ by following its surrogate gradient

$$\nabla_\theta \widehat{L(z)} = \frac{1}{n}\sum_{i=1}^{n} \nabla_\theta L[NN\left(R\left(S(y_i, x_i, \theta)\right)\right), c(\theta)] .$$

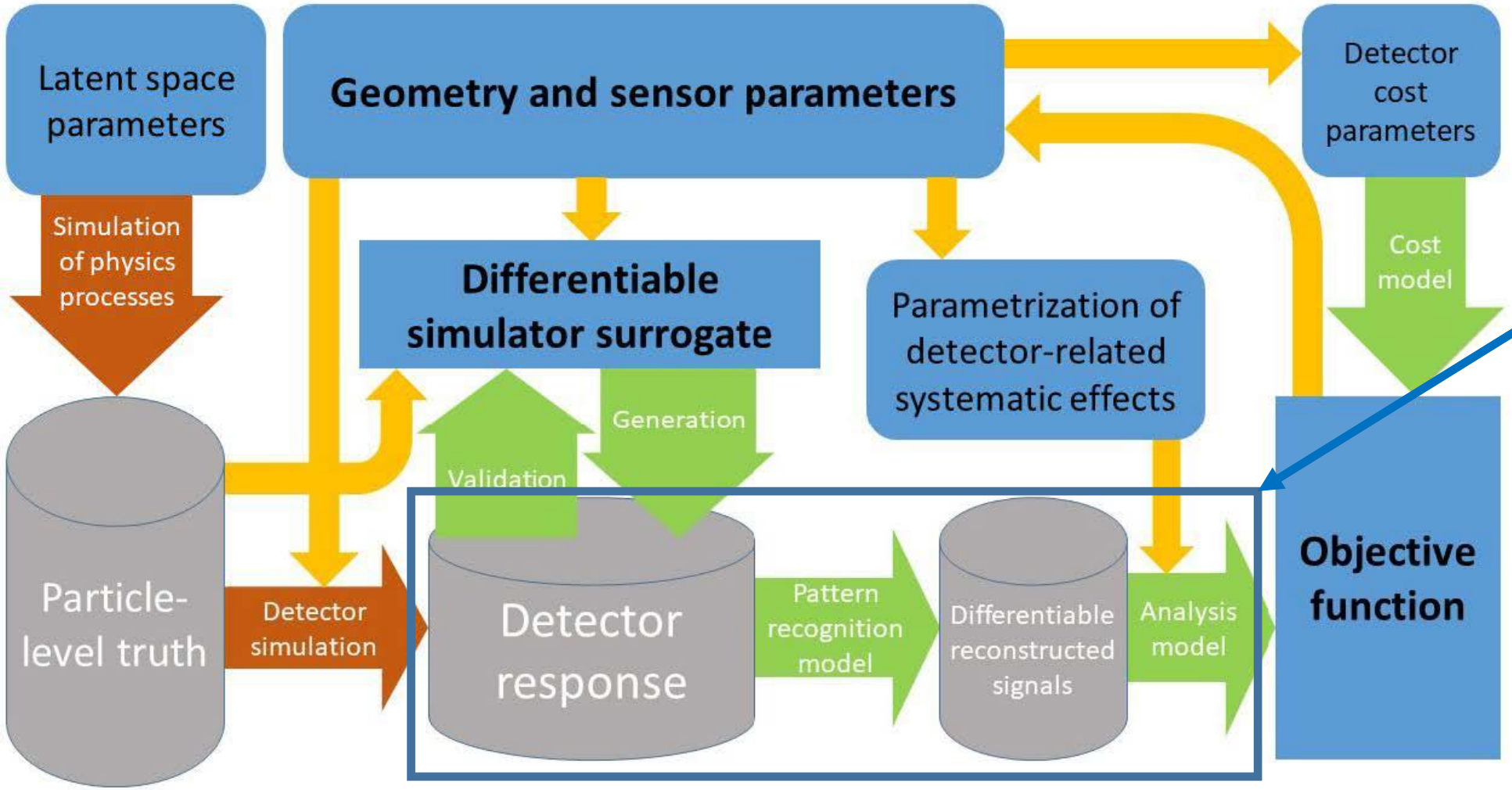
The above recipe requires one to learn the differentiable surrogate **S( )**: this is a task liable to be carried out independently from the optimization procedure.

Putting It All Together



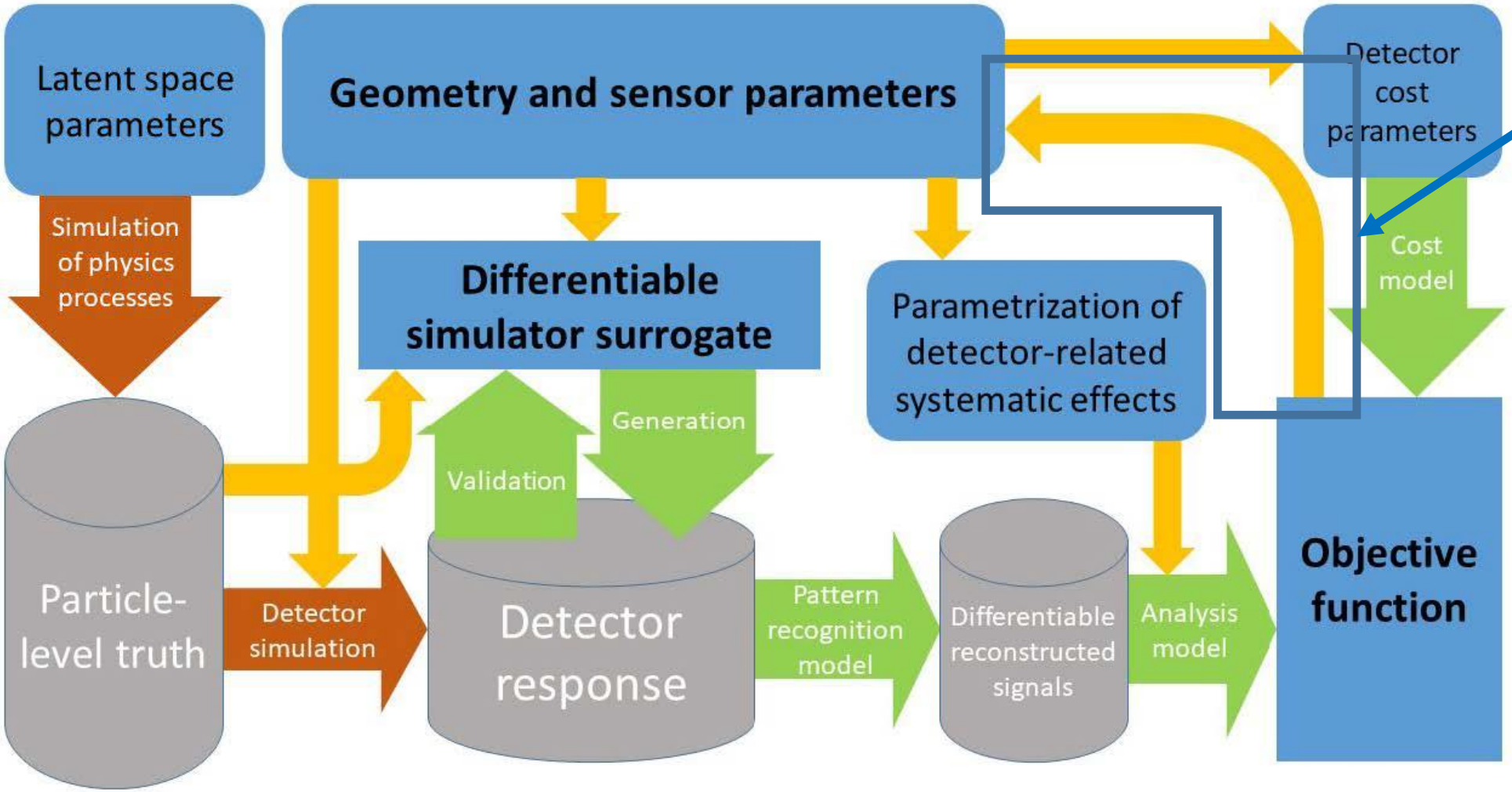
If the **simulation** can be bypassed by a **differentiable surrogate**, we remove the stochasticity of the physics and strongly simplify the problem

Putting It All Together / 2



The model must include a **model** of the absolute state-of-the-art (or even extrapolated future performance!) of reconstruction and inference to avoid any **misalignment**

Putting It All Together / 3



Backpropagation of the gradient of the objective function then allows to find optimal parameters **θ** → obtain end-to-end optimality of the instrument

A Few Additional Details - 1 / Symmetries

Symmetries are a very significant attribute of systems liable to optimization

A system's symmetry implies

1) that there will be degeneracy in the optimal solutions
   - may be a good thing (allows reduction of phase space)

2) that some features of the optimal configuration are easy to guess
   - also good, provides useful starting point
   - questions the need of wrestling with complex optimization tools!

... but the symmetry must be exact, which is rarely the case!

If you break the symmetry by enforcing detector geometry choices, this may propagate to interesting non-symmetrical solutions elsewhere in the problem

E.g. in MUonE: cylindrical symmetry of Physics along incoming muon axis, but once you place a target somewhere, this automatically implies the existence of optimal placing of silicon planes downstream – NOT equispaced!

2 - Parametrization

Care is required when defining the coordinate system and the parameters describing a detector layout.
In many cases, we already know that advantageous configurations impose functional dependencies between parameters

→ silly to force the system to discover these subspaces by brute force

→ much better: choose parametrization that naturally accounts for interdependence at soft spot

Quick example: should we parametrize a calorimeter depth in meters or in X_0 units?
Tradeoff: easier connection to external constraints vs seamless transfer of optimal solutions as one varies the composition

2 – Parametrization / cont'd

Another example: MUonE tracker. The goal is to minimize q^2 resolution for high-q^2 interactions

→ small angle expected for muon track

→ soft spot ties staggering length s to transverse tilt θ for near-horizontal tracks (the most important ones); relation changes for changes in pitch/sensor distance p/d ratios

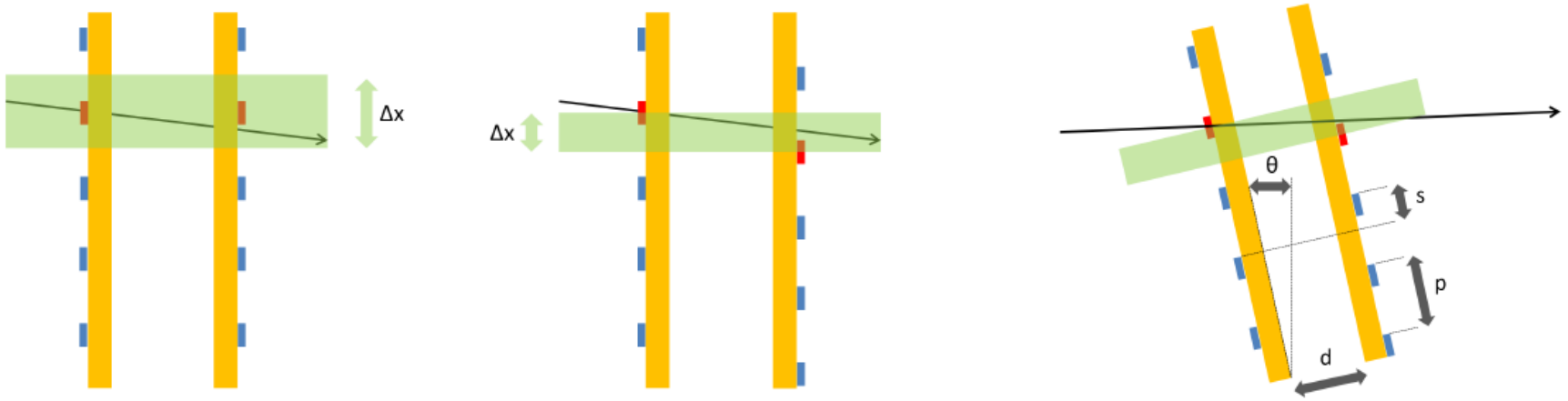


Figure 4: Left and center: a double-sided silicon strip sensor produces twice smaller resolution Δx on single-strip hit position for an orthogonally incident particle if strips on the two sides are staggered by half the strip pitch. Right: the four parameters affecting single-strip hit position resolution (tilt angle θ, strip pitch p, sensor distance d, staggering s).

→ advantageous to study what parameters to describe detector with, e.g. fix p/d = f(θ), thus initially reducing dimensionality of problem; later can release once optimality is found, to fine-tune

3 - How to Deal with Discreteness

Discreteness of design choices requires special handling

E.g., the number of detection elements of a system is something we cannot take derivatives on

- Even the simple translation of a detection element in space is an operation that breaks the differentiability (a particle hits or misses it)

In relatively simple problems (with only few discrete parameters, taking on few possible values) we may consider all combinations

+ easy parallelization
- quickly unmanageable as D grows

One natural way to deal with it is to use reinforcement learning techniques (not discussed further here)

4 – Monetary Cost Considerations

1. Cost is always a driving factor in a project, so it is essential to model it consistently
   - can be tricky when comparing wholly different technologies
2. The impact of cost in the loss may be modulated by the performance of the apparatus, rather than as a fixed constraint
   - e.g. we might be happy to discover that with 50% more than the planned budget we would have access to large improvements in performance
3. Subsystem cost in large collaborative projects is not necessarily a parameter!
   - Imagine you told the ATLAS Calo people to give 30% of their upgrade moneys to the tracking guys... «it's for the common good!» – good luck with that

Machine-Learning Optimized Design of Experiments

MODE Collaboration

https://mode-collaboration.github.io

M. Aehle[17], A. G. Baydin[5], A. Belias[10], A. Boldyrev[4], K. Cranmer[8], P. de Castro Manzano[1], T. Dorigo[1,14], C. Delaere[2], D. Derkach[4], J. Donini[3], P. Elmer[18], F. Fanzago[1], N.R. Gauger[17], A. Giammanco[2], C. Glaser[11], L. Heinrich[12], R. Keidel[17], J. Kieseler[22], C. Krause[13], L. Kusch[17], M. Lagrange[2], M. Lamparth[12], M. Liwicki[21], G. Louppe[6], L. Layer[1], F. Nardi[3,14], P. Martinez Ruiz del Arbol[9], F. Ratnikov[4], R. Roussel[20], F. Sandin[21], P. Stowell[15], G. Strong[1], M. Tosi[1,14], A. Ustyuzhanin[4], S. Vallecorsa[7], P. Vischia[2], G. Watts[19], H. Yarar[1] ,H. Zaraket[16]

1 INFN, Italy
2 Université Catholique de Louvain, Belgium
3 Université Clermont Auvergne, France
4 Laboratory for big data analysis of the HSE, Russia
5 University of Oxford
6 Université de Liege
7 CERN
8 New York University
9 IFCA, Spain
10 GSI, Germany
11 Uppsala Universitet, Sweden
12 TU Munchen, Germany
13 Rutgers University, US
14 Università di Padova, Italy
15 Durham University, UK
16 Lebanese University, Lebanon
17 Kaiserslautern-Landau University, Germany
18 Princeton University, US
19 University of Washington, US
20 SLAC, US
21 Lulea University of Technology, Sweden
22 Karlsruhe Institute of Technology, Germany

Sponsored by iris hep JENAS

Realigning Design Choices and Ultimate Goals

The target of **MODE** [Mode 2020] is to design a scalable, versatile architecture that can provide end-to-end optimization of particle detectors, proving it on a number of different applications

Idea: if we "solve" a few problems we may construct a library of solutions and exploit the universality of the architecture and its modularity, re-using modeling efforts

Initial study cases:

- LHCb EM calorimeter optimization → preliminary results out
- Muon tomography detector optimization → in progress
- Muon collider EM calorimeter → in progress
- Hybrid calorimeter design integrating tracking layers → in progress
- Optimization of detectors for air Cherenkov showers (SWGO) → in progress
- plus many more envisioned, see [Dorigo 2022]

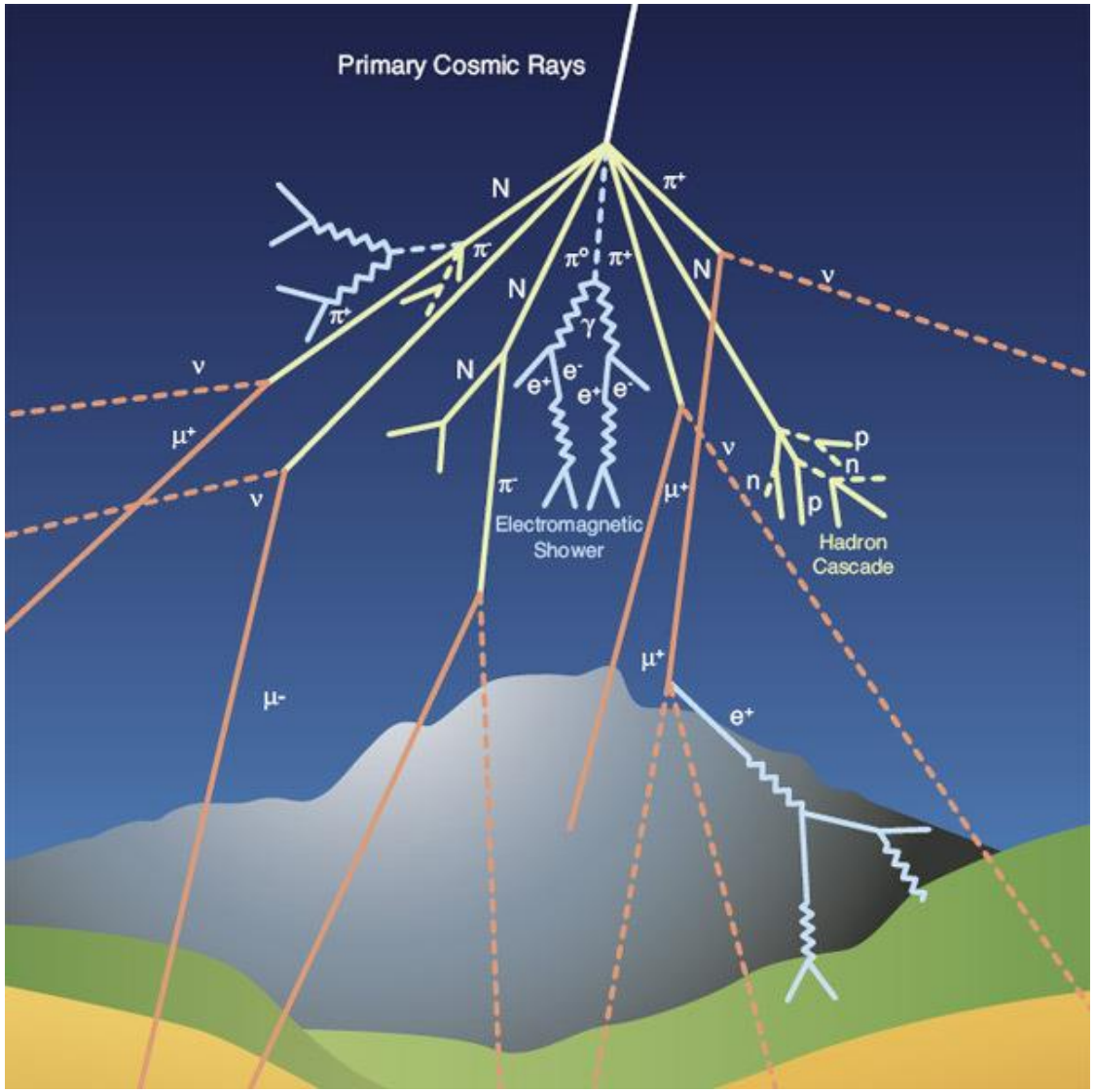
Muon Tomography Optimization

Perhaps the «simplest» use case with elementary particles: muon tomography.

Exploit flux of cosmic rays impinging on Earth's surface as a source of data with which to image composition of unknown volumes

Applications are countless:

- study core of volcanos
- prevent smuggling of dangerous materials in containers
- study wear of industrial apparata (e.g. steel pipes), interior of furnaces, plasma in reactors
- archaeological prospections
- and many more



Above: *a proton or light nucleus hits a nucleus of Nitrogen and produces a shower of unstable hadrons. The latter decay to muons that reach the ground*

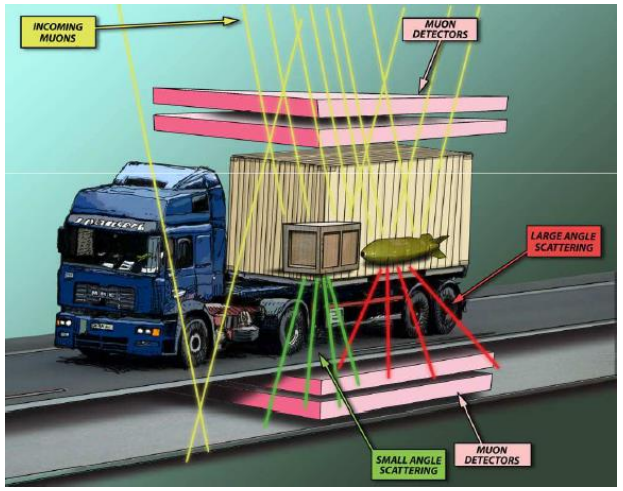
Muon Tomography Optimization / 2

Perhaps the «simplest» use case with elementary particles: muon tomography.

Exploit flux of cosmic rays impinging on Earth's surface as a source of data with which to image composition of unknown volumes

Applications are countless:

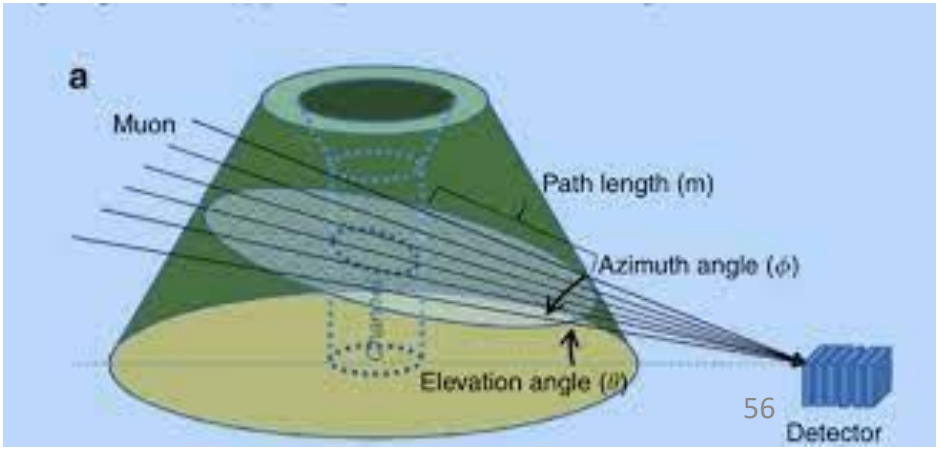
- study core of volcanos
- prevent smuggling of dangerous materials in containers
- study wear of industrial apparata (e.g. steel pipes), interior of furnaces, plasma in reactors
- archaeological prospections
- and many more



Above: *exploit dependence of scattering angle of muons on atomic number of material*

Two main methods for inference

Below: *use incoming momentum spectrum of muons and absorption of parts of it to infer thickness in 3D*



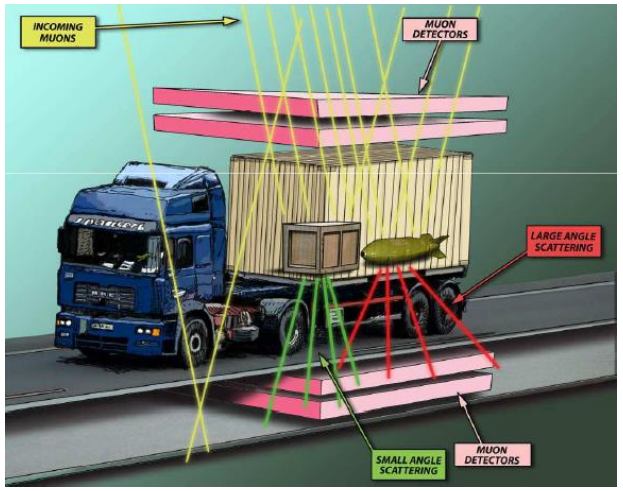
Muon Tomography Optimization / 3

Scattering tomography is not hard to model:

- need no surrogate of a simulator;
- one particle at a time (or even many...)
- parametric scattering model

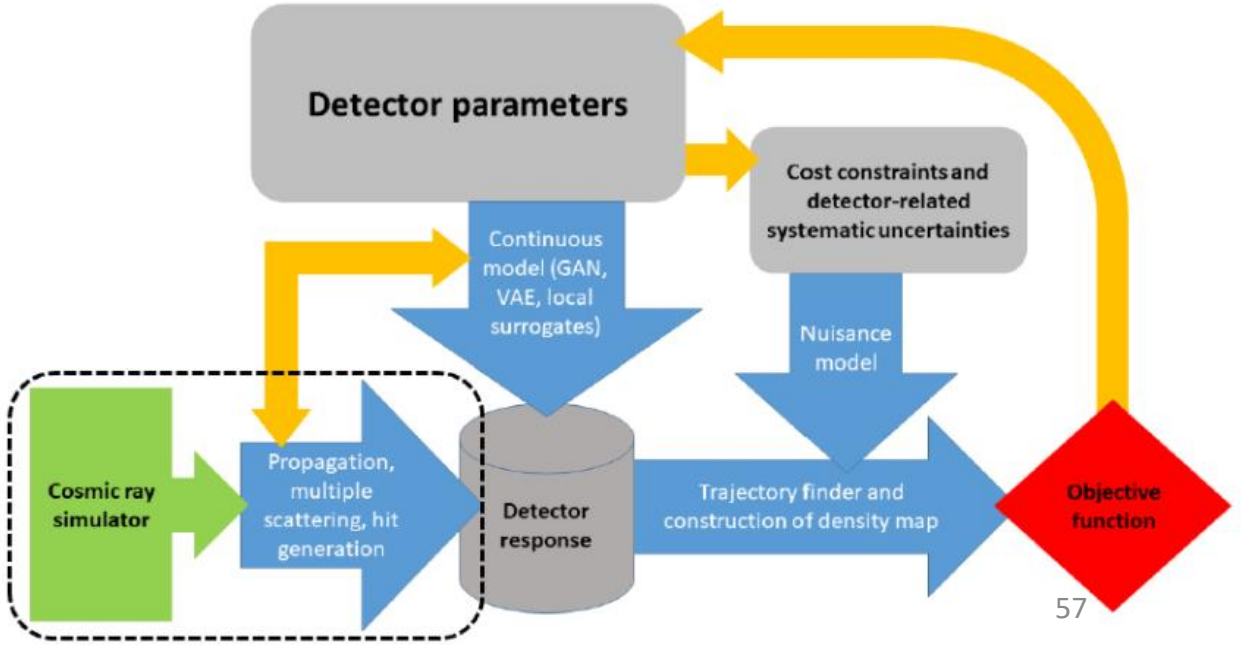
Still, quite complex to put together simulation, inference, and optimization in one package

→ great testing ground for future, more ambitious projects



Above: *exploit dependence of scattering angle of muons on atomic number of material*

Below: *Optimization pipeline of TomOpt*



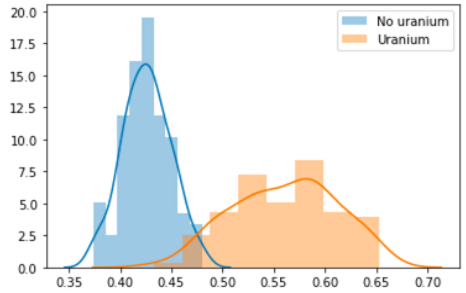
A Simple Example

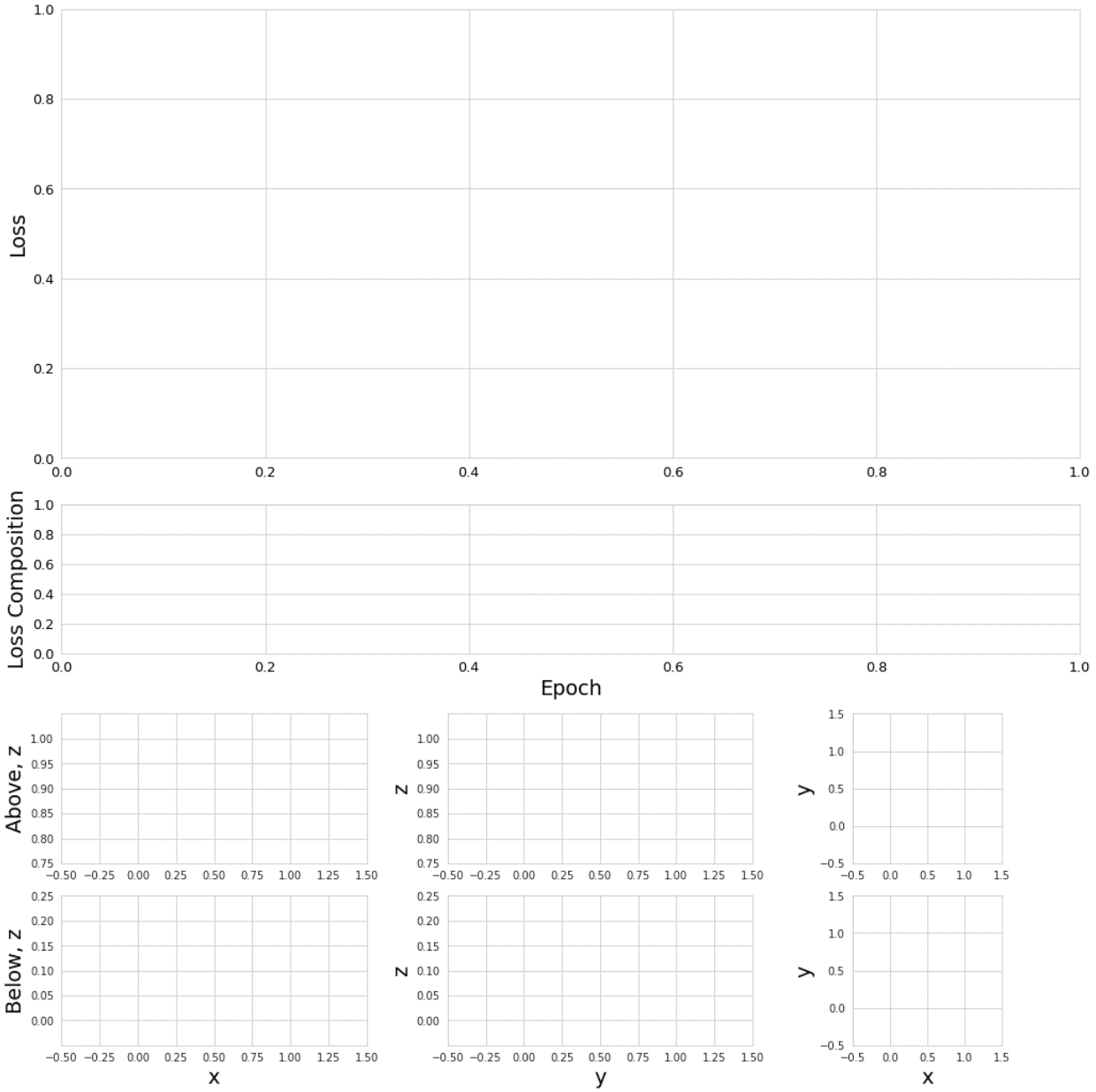
Identification of an U block in a lorry filled with scrap metal and air

U blocks are generated at random in the volume

The system infers the X_0 of voxels, and builds a test statistic on which it does inference on presence of U

Loss is BCE with a monetary cost penalization





Above: test statistic for U / no U volumes

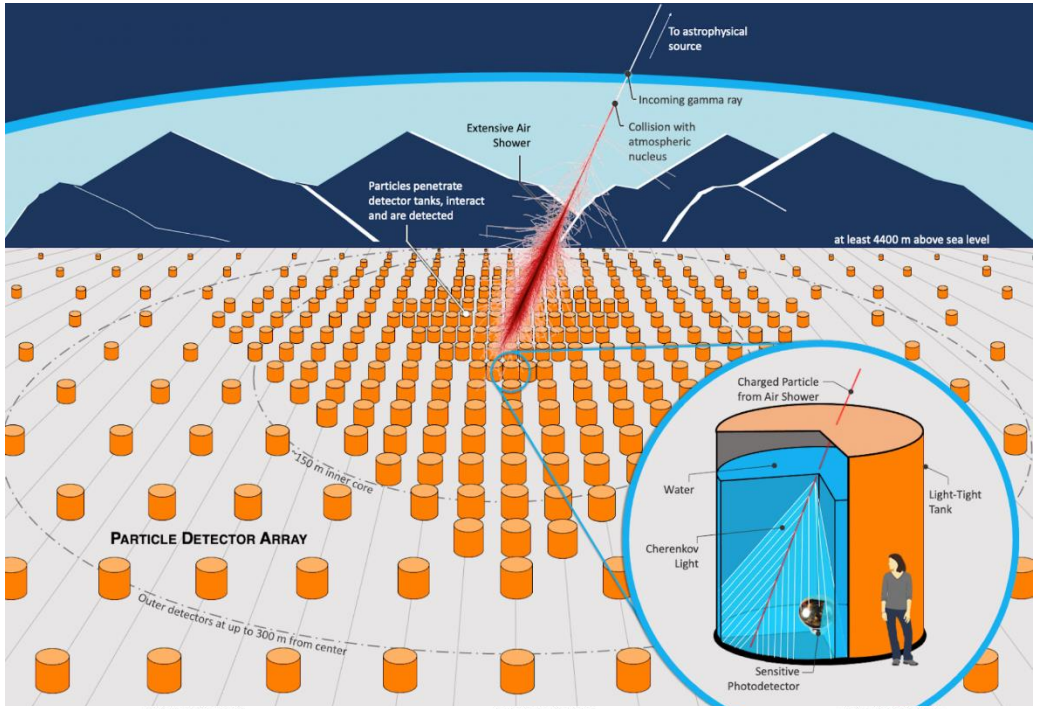
SWGO

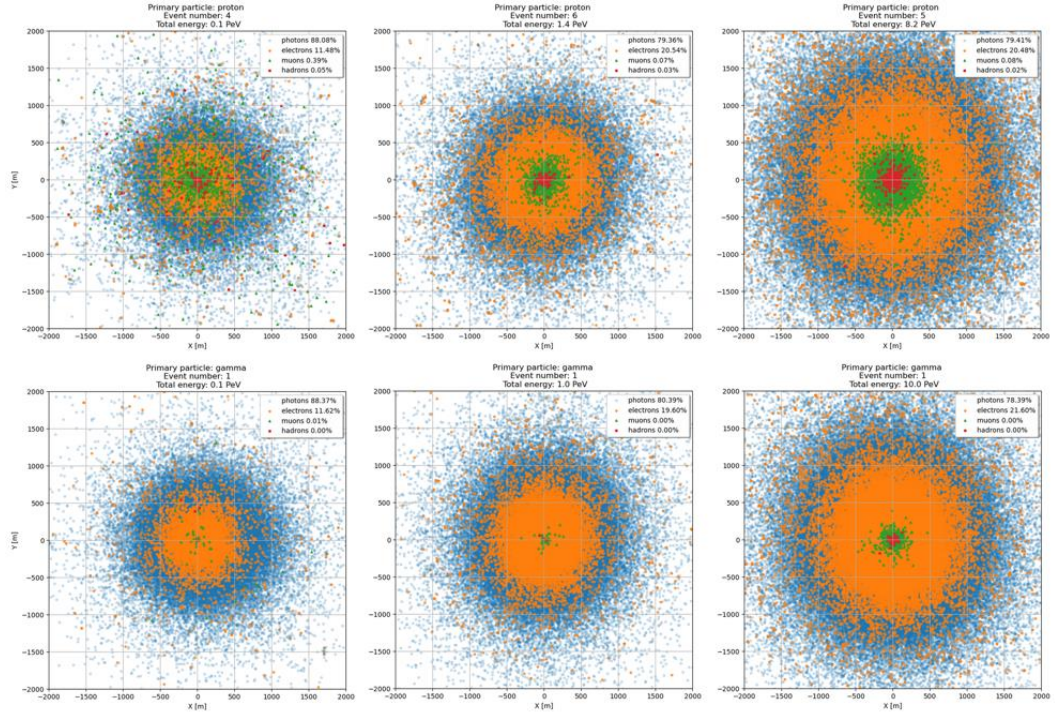
Ground-based array for ultra-high-energy gamma showers, proposed to be installed at high altitude in northern Chile

4000 water Cherenkov tanks, discrimination between e/γ and μ to select muon-poor showers (remove proton background primaries)

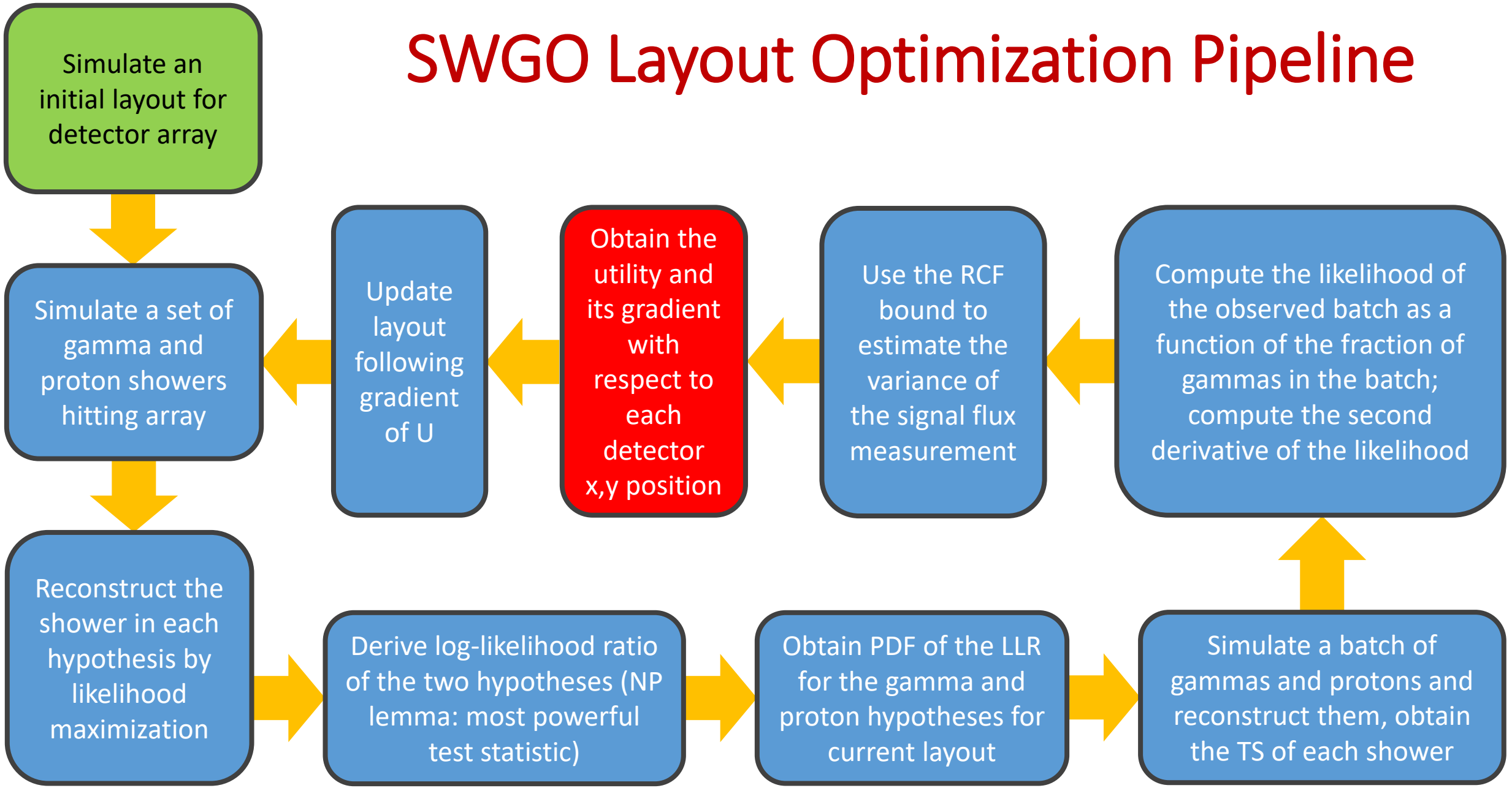
Tank optimization already a big problem

Detector layout simpler to handle, good exercise of DP for full optimization; low coupling of geometry to tank characteristics → can study separately; let us give it a shot...





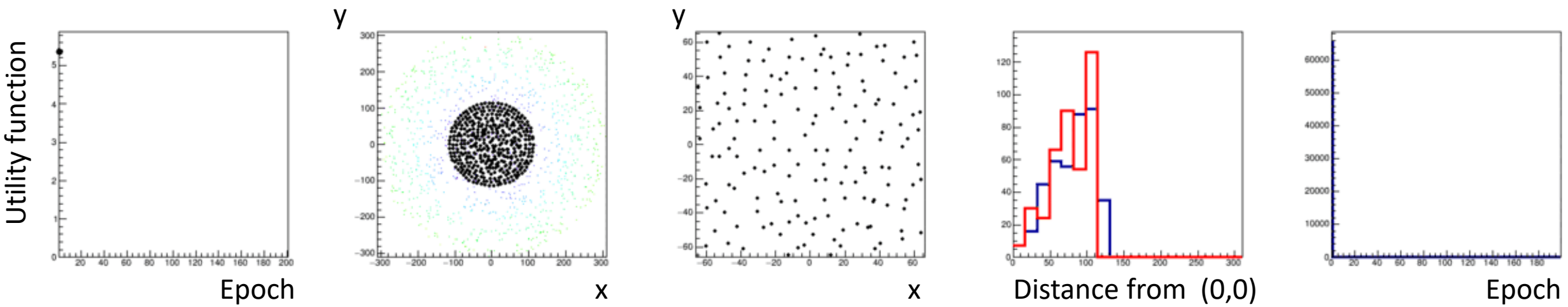
SWGO Layout Optimization Pipeline



Sample Run

$$U = \sum_{E=E_1}^{E_n} w_{Ei} \frac{\Phi_i}{\sigma_{\Phi_i}}$$

Below is shown a run where the initial position of detector elements (black points in second and third diagram) gets updated as the system learns to place them in more advantageous positions, maximizing a weighted sum of precision in flux measurements at various energies



The center of generated showers is shown with a colour in the second diagram, to indicate the precision of its reconstruction.

Another WiP: Hadron Calorimetry

For decades, hadron calorimeters only focused on achieving a good measurement of collective incident hadron energy

→ relevant lengths: Moliére radius, λ_0, X_0

→ Building calorimeter with highly segmented cells seemed unjustified

→ Lateral and longitudinal distribution not the focus

E.g. the CMS central hadron calorimeter only took 3% of the experiment budget!

But today we very much need high-res images for boosted tagging / particle flow
→ this has forced a paradigm shift in the design of these instruments

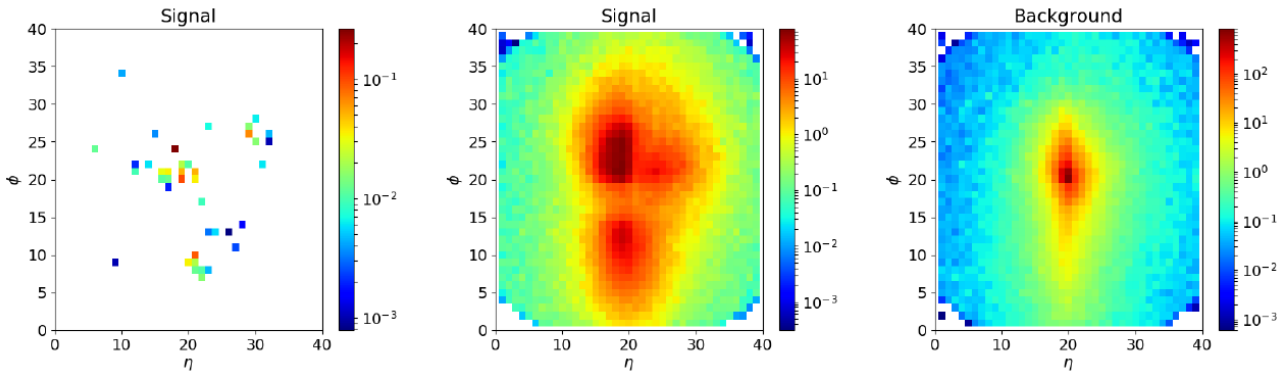
→ Granular calorimetry must be investigated further

Fill-me-up Slide: CNNs for Boosted Jets

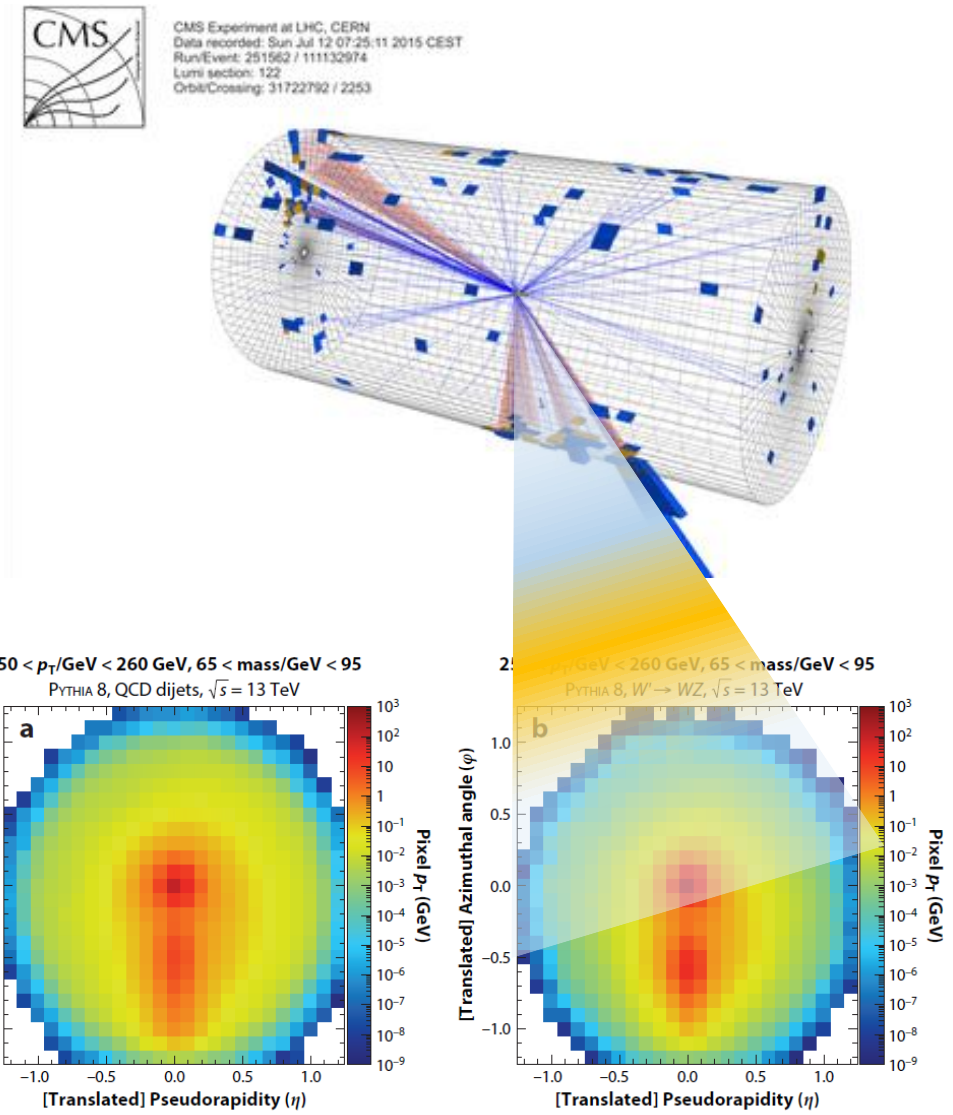
The W boson (1983) and the top quark (1995) were discovered using decays to electrons and muons, as the more frequent quark decays are difficult to distinguish from backgrounds

We later realized that if top quark, W, Z, H bosons have high energy, the jets of hadrons they produce in hadronic decays are distinguishable in highly granular calorimeters from backgrounds through their inner structure

Today the most sensitive searches for new massive particles benefit from CNN-powered imaging techniques



***Left**: true pattern of particles in a top quark jet. Center: observed average distribution for top quark decays. Right: average shape of QCD jets*



***Above**: a background jet image (left) and a W-quark jet image (right)*

A Problem for Future Colliders - What to Do with Muons Above a Few TeV?

In the past, muons were crucial to discover, e.g., the Upsilon, the W/Z, the top, the Higgs...

Future colliders of energy higher than that of the LHC might produce new particles whose signature involves decay to ultra-high-energy muons. But above a few TeV we cannot rely on magnetic bending to measure their momentum:

- CMS has $\delta E/E = (0.06 : 0.17)\ E/\text{TeV}$
- ATLAS has $\delta E/E = (0.08 : 0.20)\ E/\text{TeV}$

→ above 3-4 TeV muon energy becomes unmeasurable

Can we increase the bending power of our magnets? Hardly...

Can we use radiative losses? Not by ordinary means...

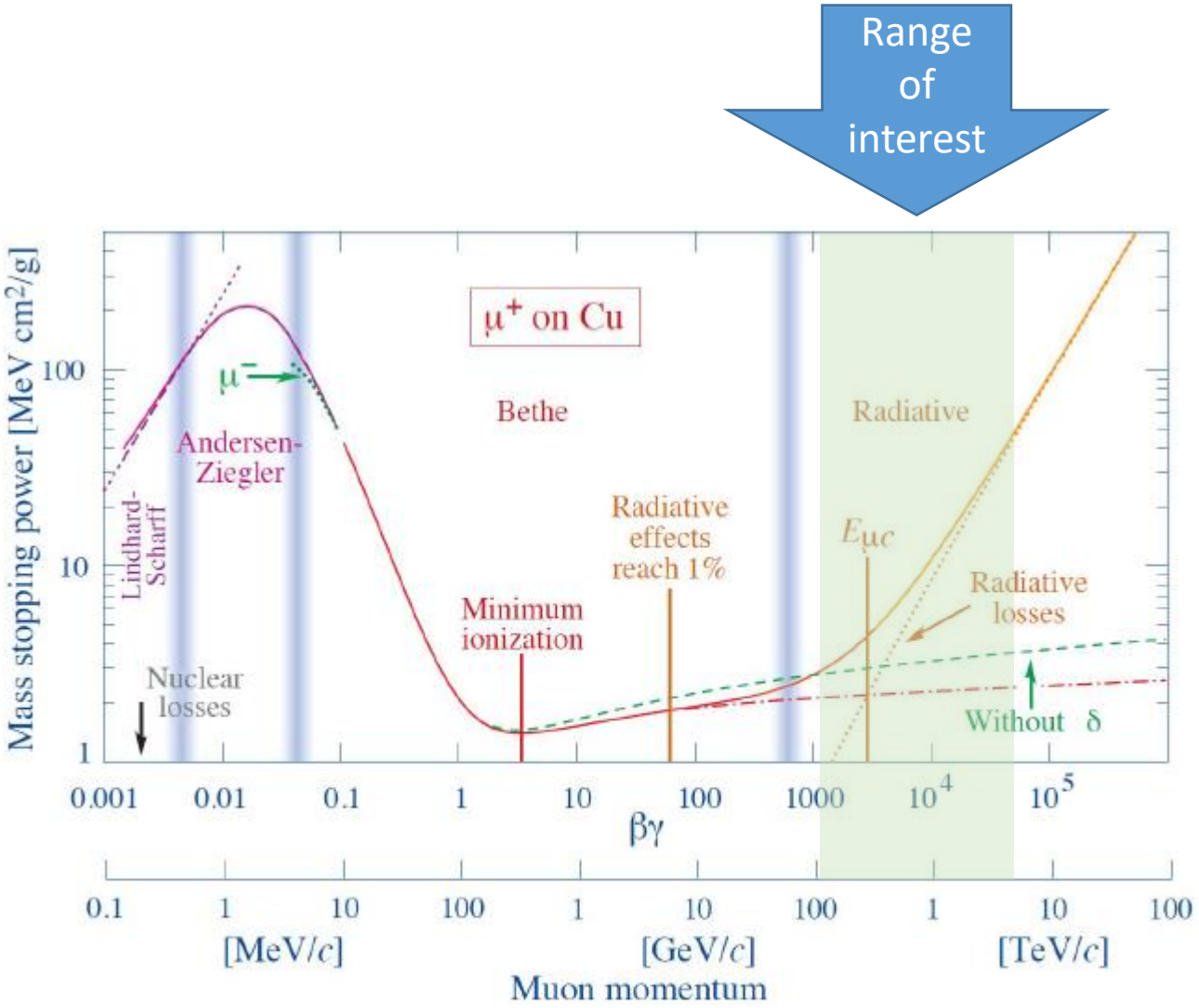
→ Maybe DL can help, but we need a granular calorimeter. Let us see how.

Measuring Muon Energy in a Calorimeter

Muons are minimum-ionizing particles, and they do not produce EM showers as they traverse dense materials

Their behavior changes above a few 100 GeVs, when they start to radiate soft photons in significant amounts – but still, even then the energy loss in a thick calorimeter is typically of the order of a few percent

→ hardly usable – or is it?



Simulation of a High-Granularity Calorimeter

We produced with GEANT4 high-statistic samples of muons interacting in a high-granular, homogeneous $PbWO_4$ calorimeter

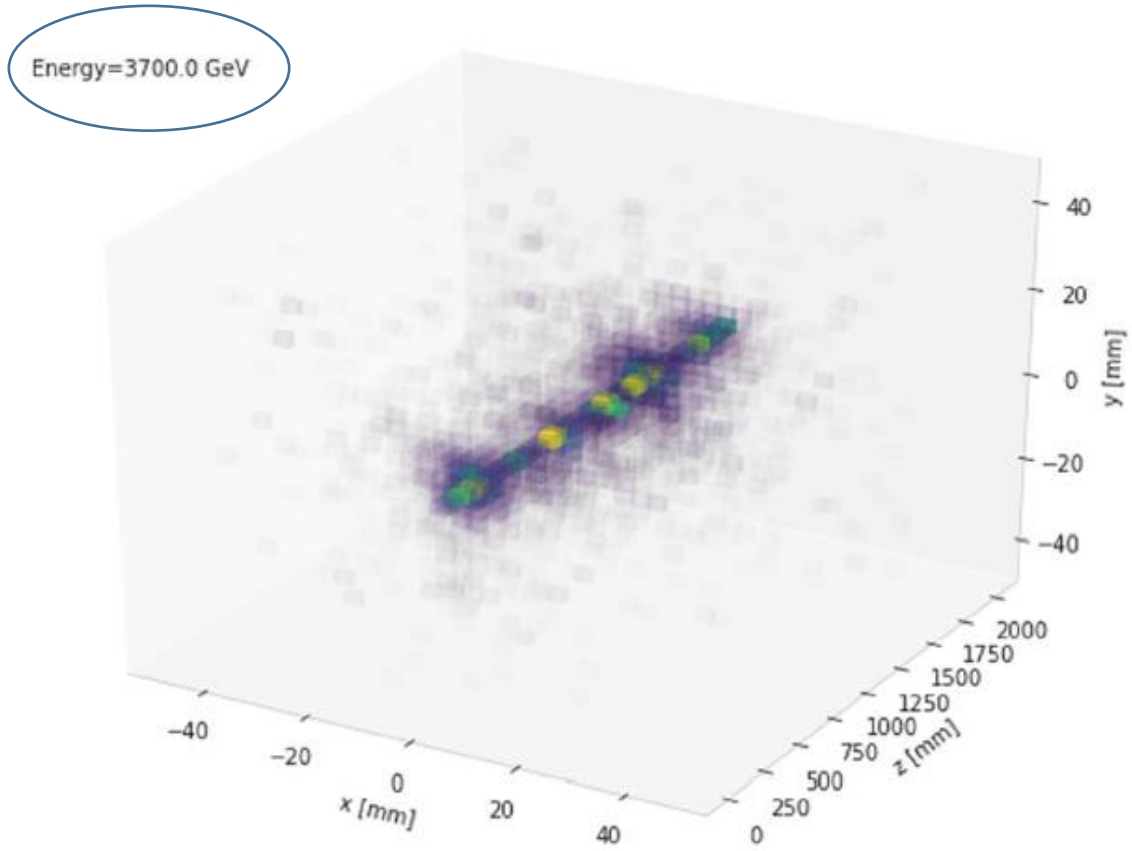
total depth = 2032 mm = 10 λ_0

50 layers of 32x32 cells –

51,200 channels in total

cell size = 3.7 x 3.7 x 39.6 mm

We simulated about 900k muons in the [0.1-8 TeV] range to train and validate our models, and some further fixed-energy samples for testing.

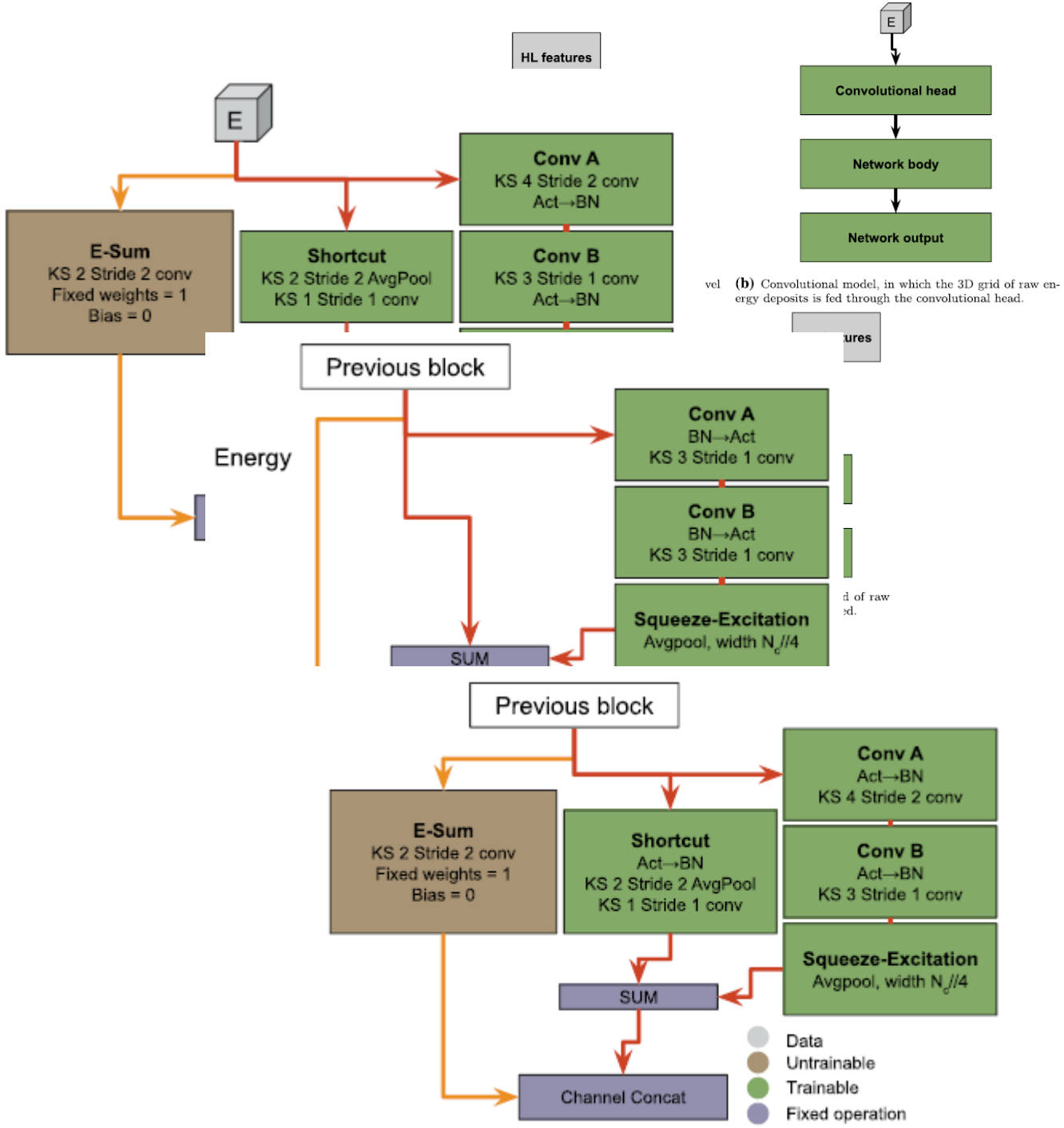


Note: this looks like a shower, but most photon deposits shown are of few MeV only

CNN Model

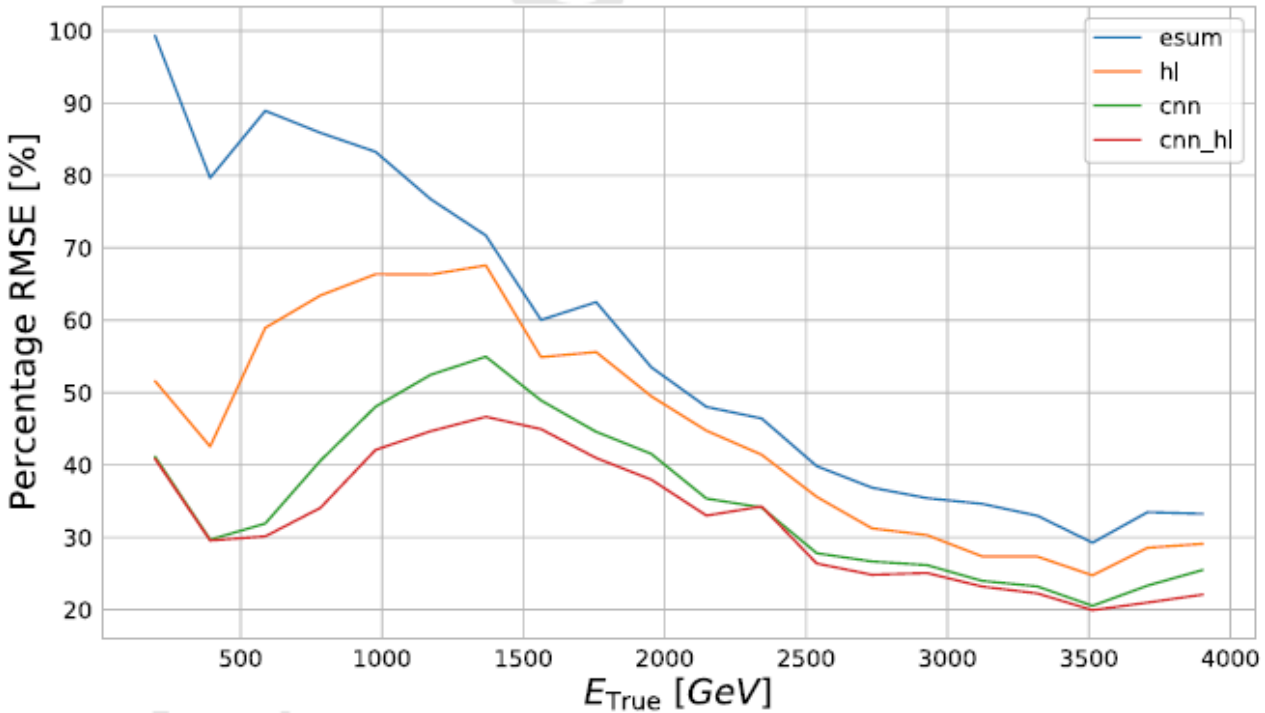
The CNN we originally deployed uses convolutional layers to reduce the dimensionality of the pattern of energy deposits, and employs, in addition to the full xyzE information, 28 high-level features in a dense layer.

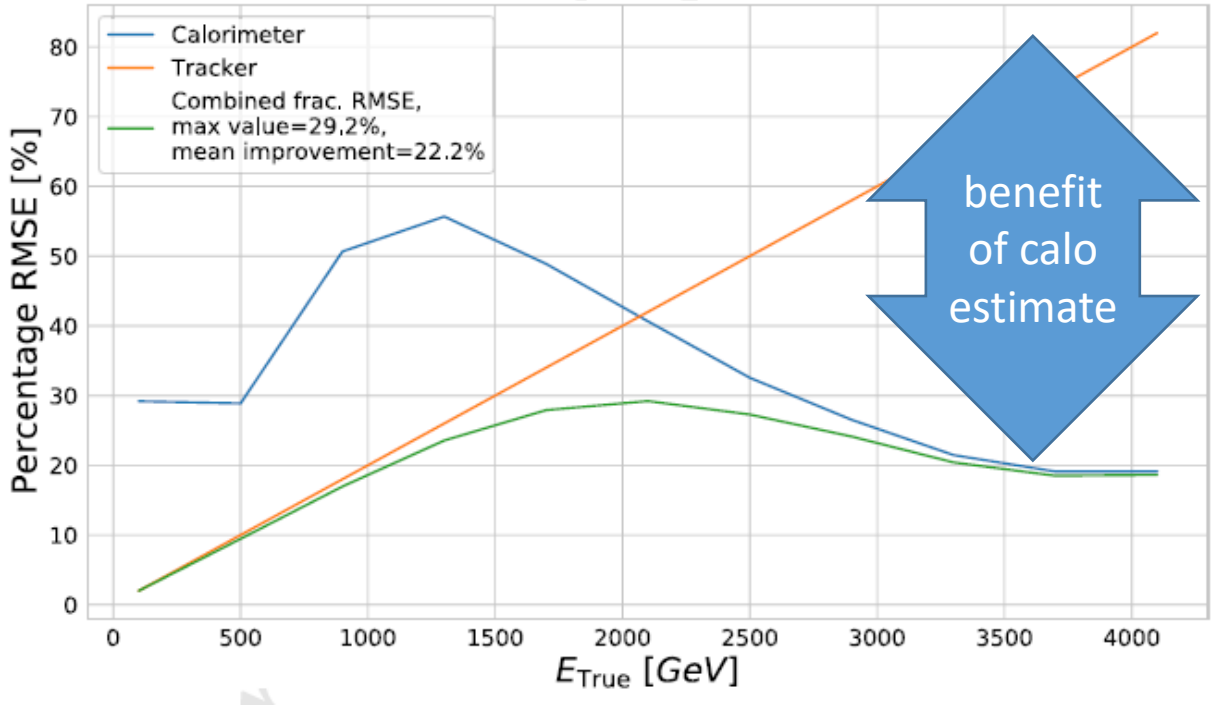
The 28 event features are the result of human-based «domain knowledge» about possible ways to aggregate the information on the spatial distribution of energy deposits



CNN Results

The CNN model [Kieseler 2022] manages to recover 20% resolution for 4-TeV muons. It also demonstrates how there is information in the spatial distribution of photon deposits. A combination with a tracking measurement (δE/E=0.2E/TeV) is shown on the right.



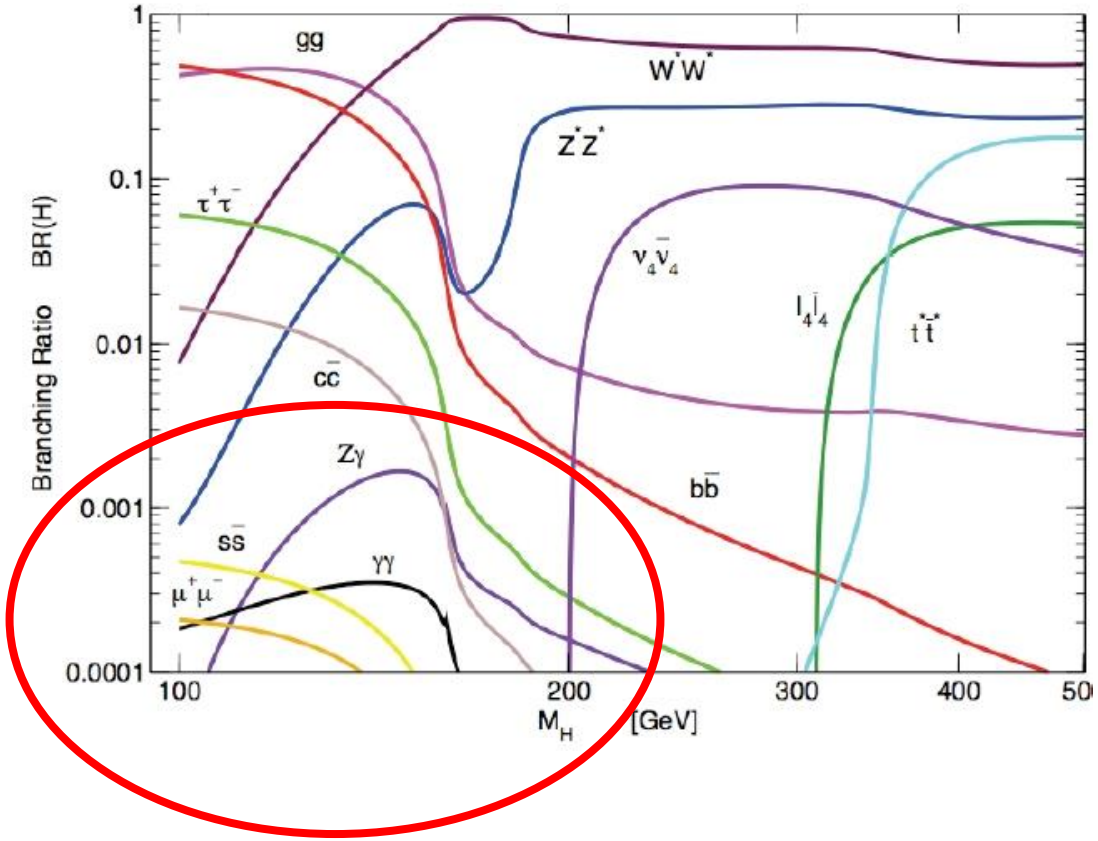


Asking for More to Calorimeters: Particle ID

Charged pions, kaons, and protons constitute the bulk of the hadrons flowing into a hadron calorimeter

Being able to distinguish them would bring in **very large gains**:

- to flavour tagging (killer app: H→ss at a future collider, where you need to tag the fast kaon from s hadronization)
- to energy reconstruction (improved through particle flow techniques)
- to boosted-jet tagging (from improved inner structure reconstruction of jet cores)
- and more

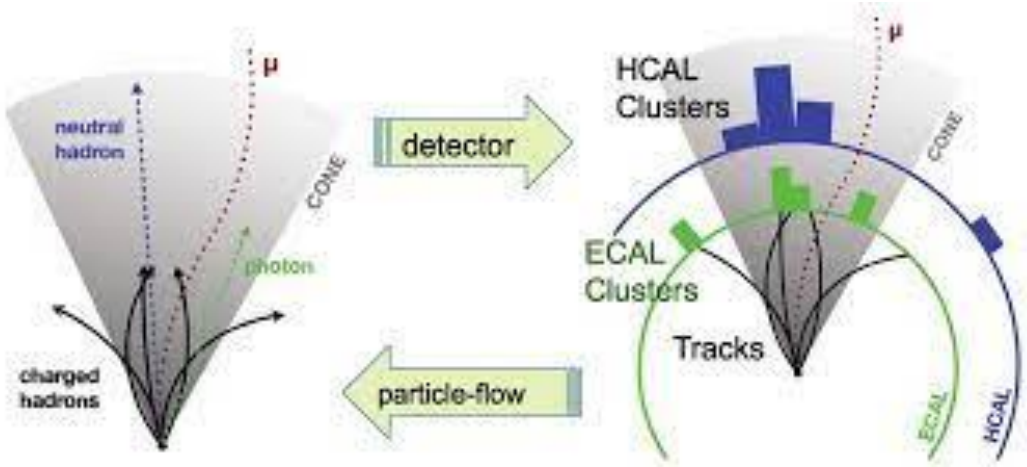


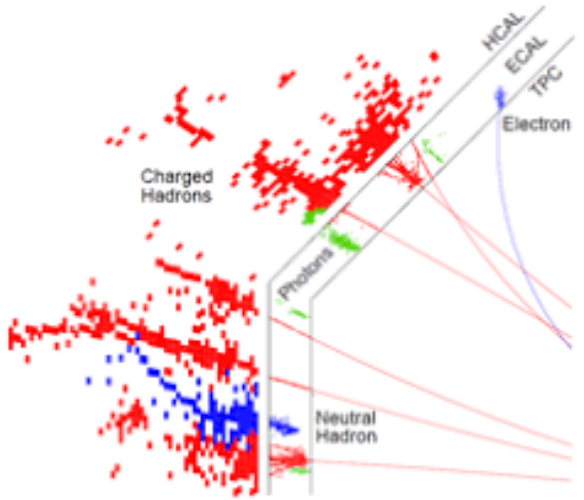
Asking for More to Calorimeters: Particle ID

Charged pions, kaons, and protons constitute the bulk of the hadrons flowing into a hadron calorimeter

Being able to distinguish them would bring in **very large gains**:

- to flavour tagging (killer app: H→ss at a future collider, where you need to tag the fast kaon from s hadronization)
- to energy reconstruction (improved through particle flow techniques)
- to boosted-jet tagging (from improved inner structure reconstruction of jet cores)
- and more



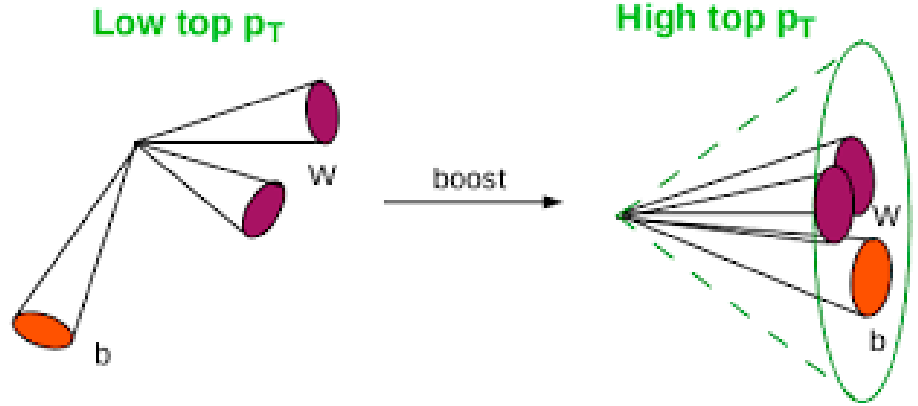


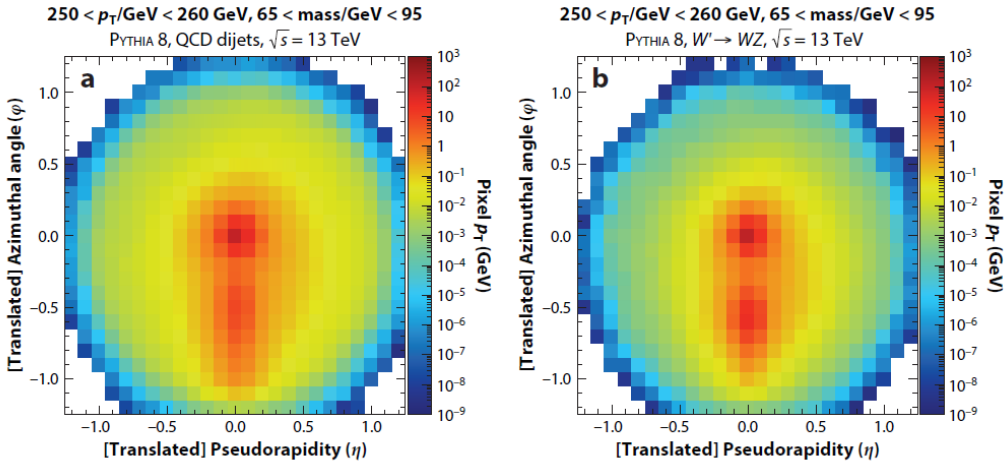
Asking for More to Calorimeters: Particle ID

Charged pions, kaons, and protons constitute the bulk of the hadrons flowing into a hadron calorimeter

Being able to distinguish them would bring in **very large gains**:

- to flavour tagging (killer app: H→ss at a future collider, where you need to tag the fast kaon from s hadronization)
- to energy reconstruction (improved through particle flow techniques)
- to boosted-jet tagging (from improved inner structure reconstruction of jet cores)
- and more





Asking for More to Calorimeters: Particle ID

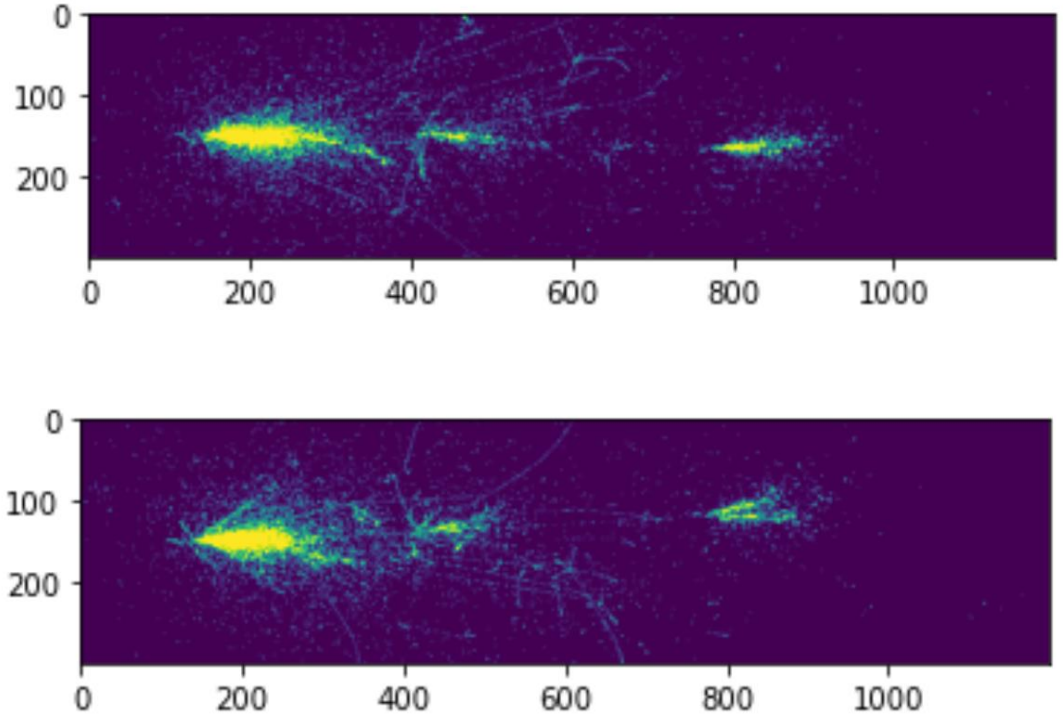
Charged pions, kaons, and protons constitute the bulk of the hadrons flowing into a hadron calorimeter

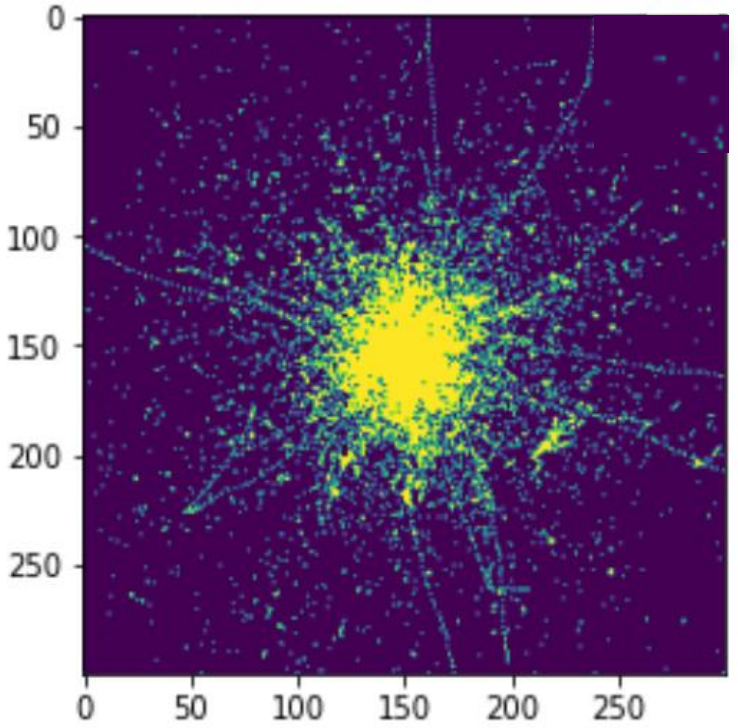
Being able to distinguish them would bring in **very large gains**:

- to flavour tagging (killer app: H→ss at a future collider, where you need to tag the fast kaon from s hadronization)
- to energy reconstruction (improved through particle flow techniques)
- to boosted-jet tagging (from improved inner structure reconstruction of jet cores)
- and more

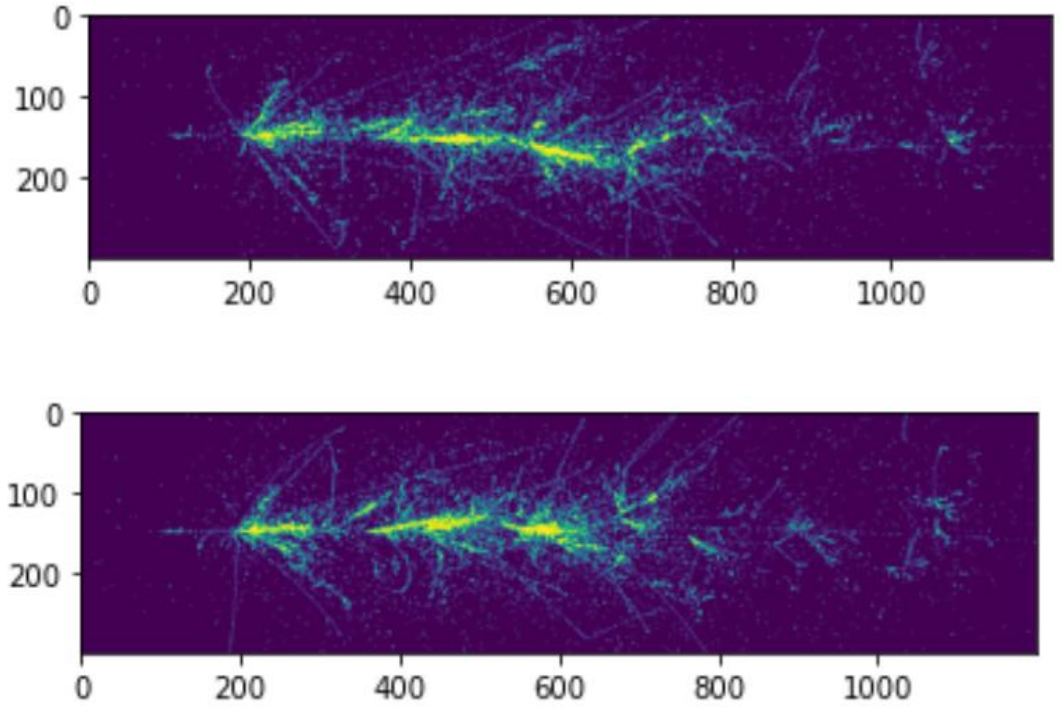
But **can it be done?**

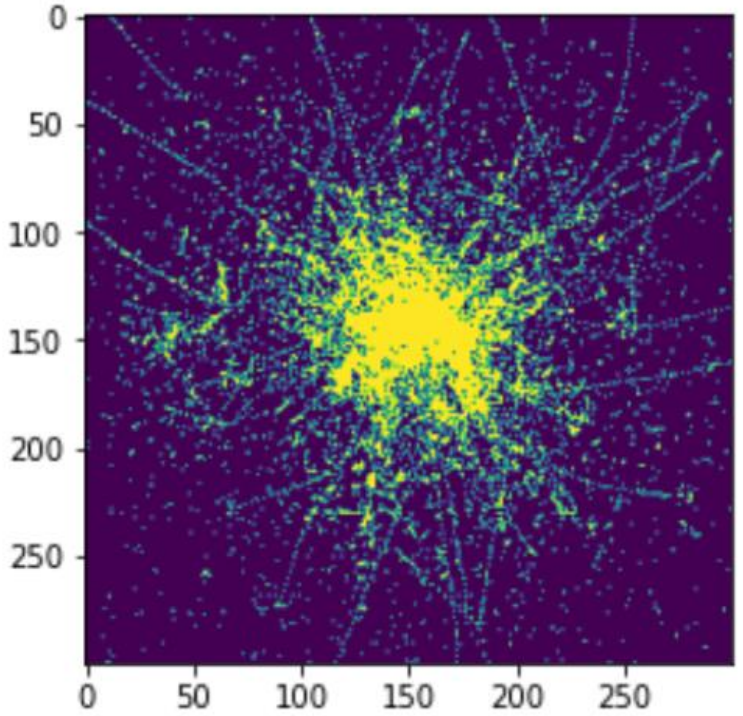
How Hadron Showers Look Like – 100 GeV p



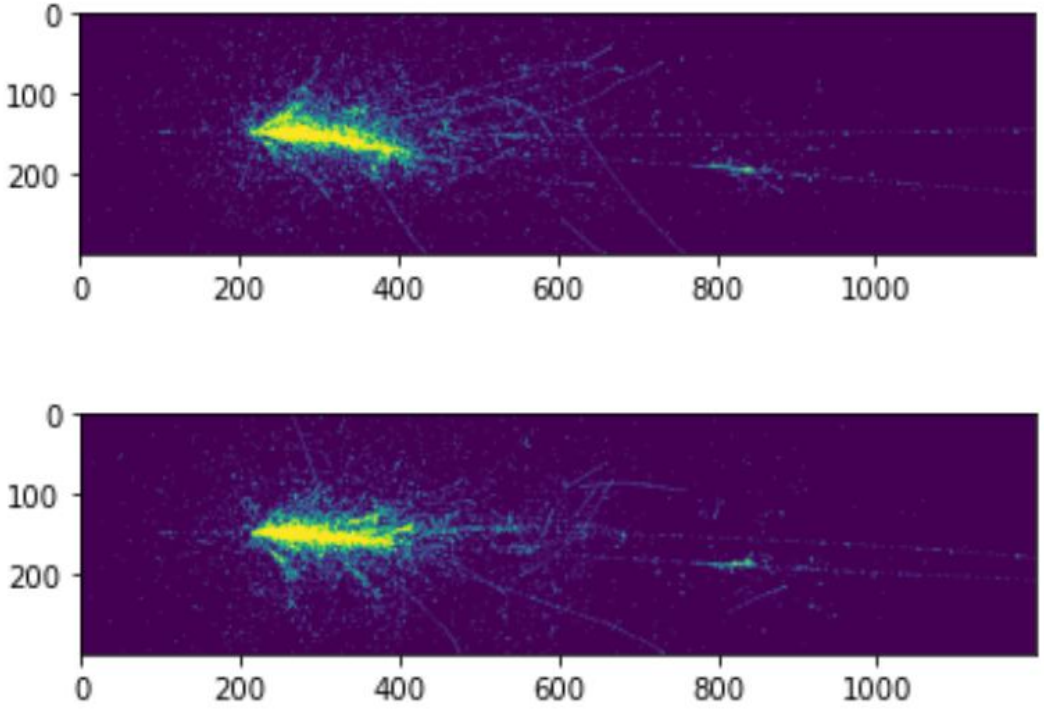


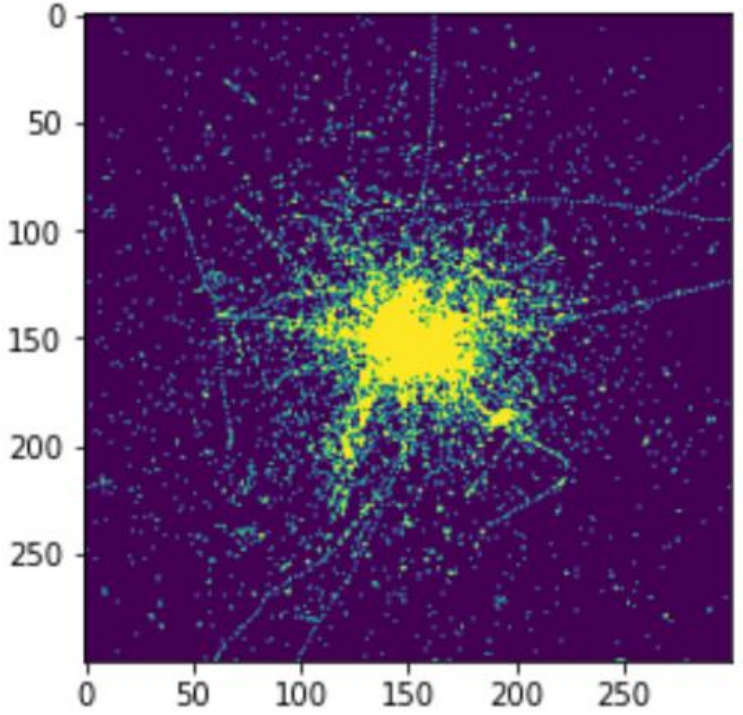
How Hadron Showers Look Like – 100 GeV K^+





How Hadron Showers Look Like – 100 GeV π^+





But Wait a Minute...

«... My grandma could tell them apart!»

That's because they are examples from a hugely varied sample. Hadronic showers are **highly stochastic** – they all look different from one another, regardless of whether the primary particle is the same or not.

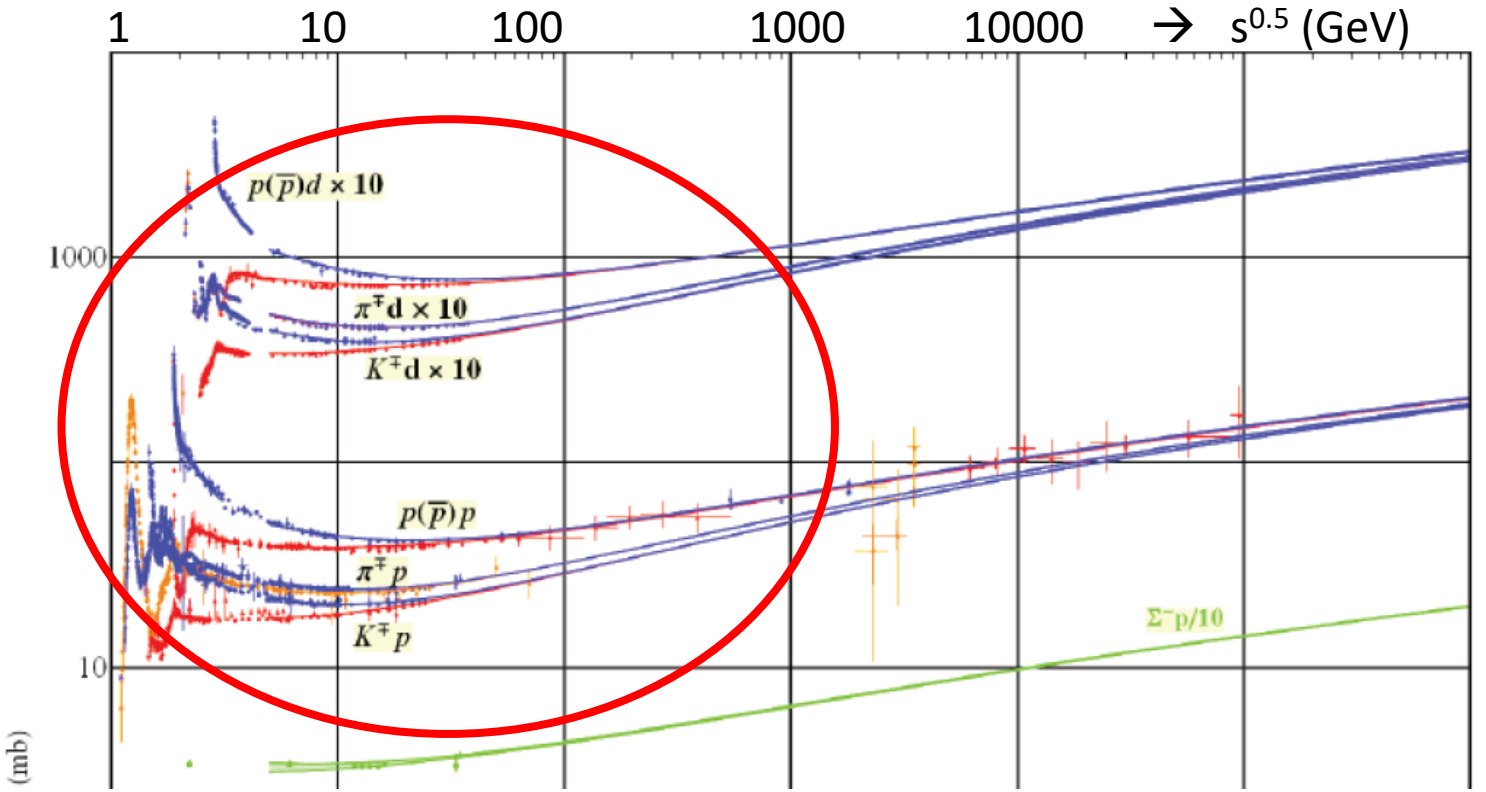
→ the task of distinguishing them is extremely hard!

What Information Are We After?

Physical facts:

- Protons are larger than pions and kaons, in fact the nuclear interaction cross sections of protons, pions, kaons are significantly different

  →this should be «easy» to exploit



What Information Are We After?

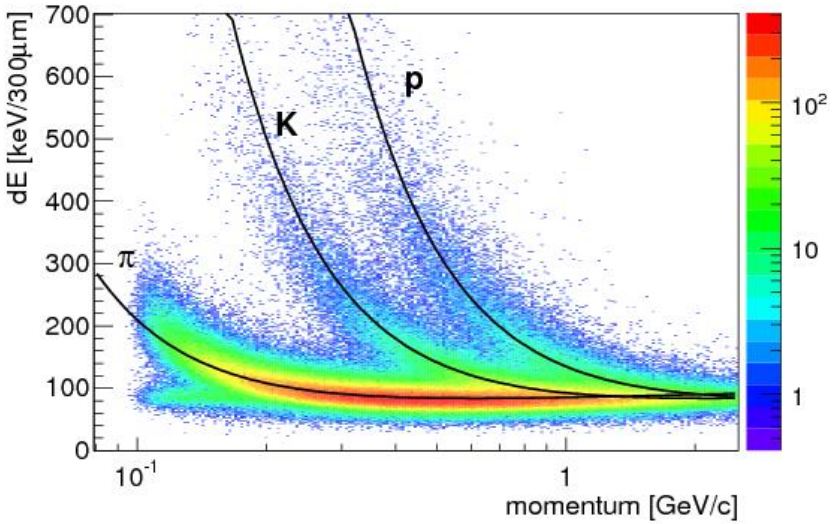
Physical facts:

- Protons are larger than pions and kaons, in fact the nuclear interaction cross sections of protons, pions, kaons are significantly different

  →this should be «easy» to exploit

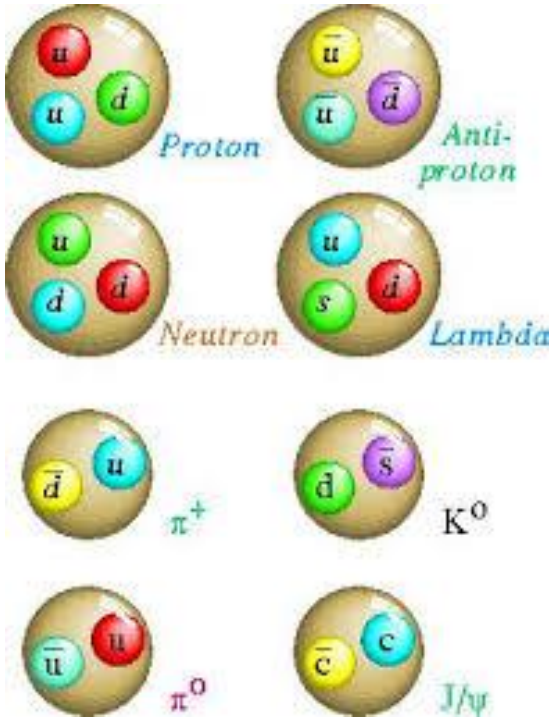
- Ionization power is also different (we only used this in tracking so far)

  → if we have sufficient granularity we can single out the ionization of each particle, at least away from the bulk of the shower



$$\left\langle -\frac{dE}{dx} \right\rangle = K z^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \ln \frac{2 m_e c^2 \beta^2 \gamma^2 W_{\max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right]$$

What Information Are We After?



Physical facts:

- Protons are larger than pions and kaons, in fact the nuclear interaction cross sections of protons, pions, kaons are significantly different

  →this should be «easy» to exploit

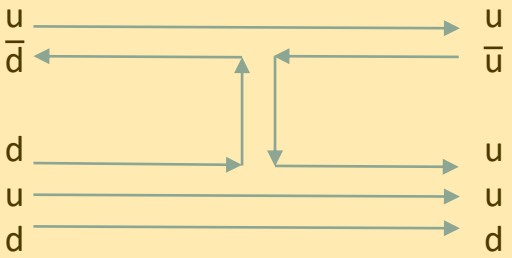
- Ionization power is also different (we only used this in tracking so far)

  → if we have sufficient granularity we can single out the ionization of each particle, at least away from the bulk of the shower

- Kaons contain one unit of strangeness, pions (and protons) do not

  → the daughters in nuclear collisions are different

E.g.: Charge exchange (pions do it, kaons don't)



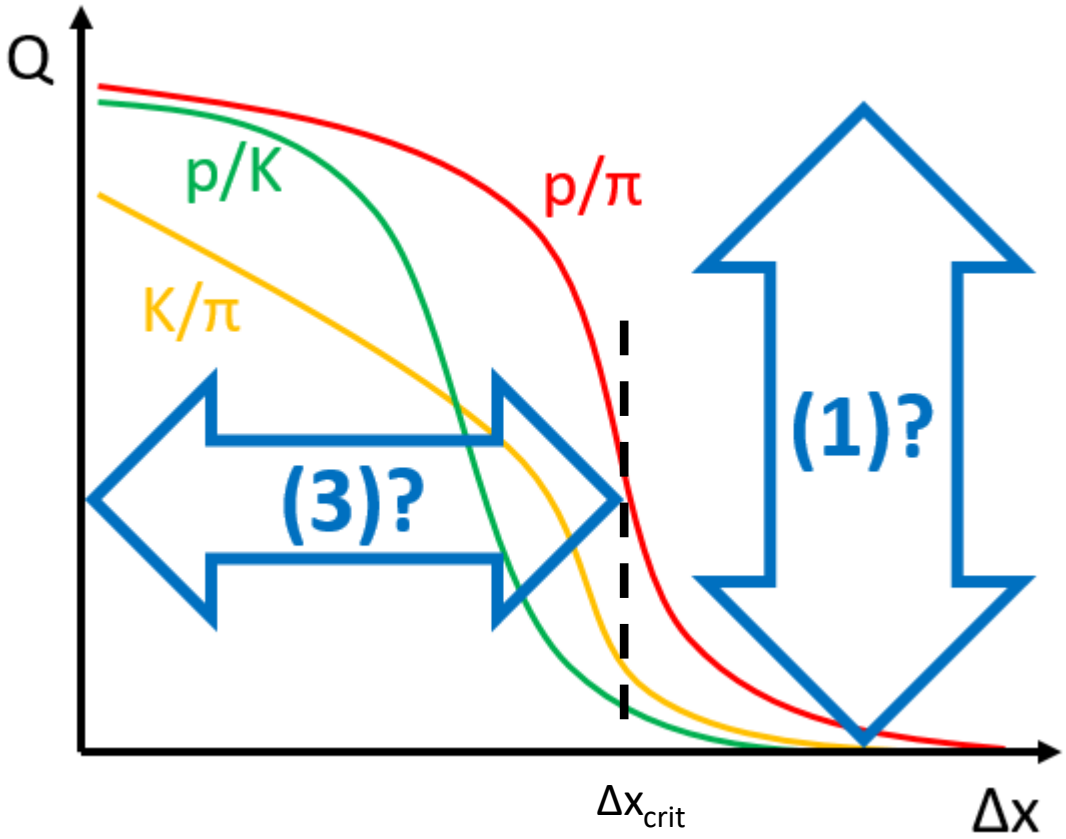
Research Questions and Money Plot

(1) What are the ultimate particle ID capabilities of a granular hadron calorimeter, assuming no limit on size Δx of readout cells?

(2) By how much would that information improve the performance of hadronic jet reconstruction in specific benchmarks of interest (e.g., H->bb, H→ss)?

(3) How does particle ID capability degrade as Δx is increased, and for what value Δx_{crit} does it get lost in conceivable setups?

(4) How much further gain is possible by exploiting timing information?



Longer-Term Plan: Hybridization

Standard setup in particle detectors: tracker → calorimeter

Why abrupt change of density?

- allow measurement of charged particle parameters undisturbed by nuclear interactions

But today we have AI reconstruction...

Plan: investigate coupled system of tracker and calorimeter, slowly vary density in z from step function to smoother solution, see effect

→Requires high-performance reconstruction of nuclear interactions in pattern reco step

→May discover new ways / overcome standing paradigm

The Future

The Future

LHC will run for 15+ more years, but we are already designing new, more powerful accelerators and detectors

The facilities will take O(20) years to build

→ It makes no sense to optimize them for the reconstruction methods available today

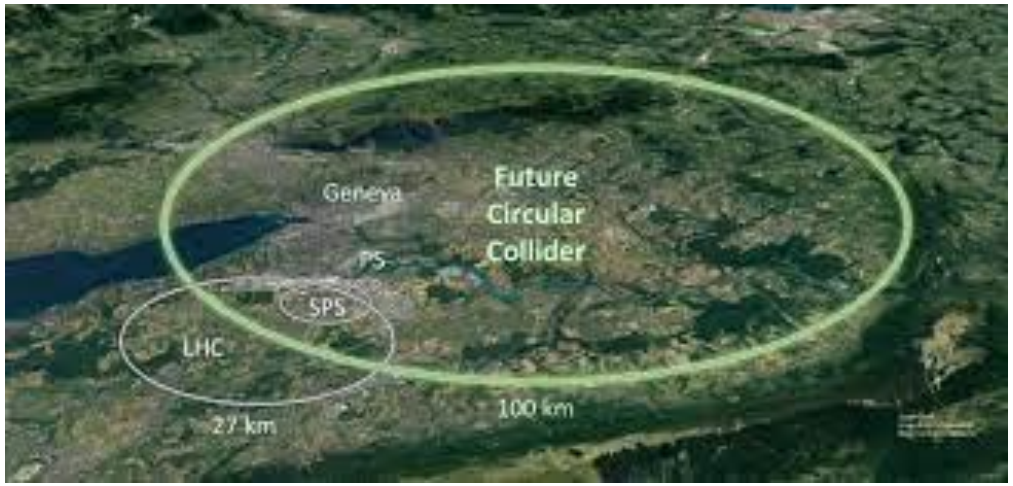
→ We should develop optimization pipelines for these new facilities, including parameters tweaking inference extraction performance from today's standards to AI-powered methods that will be available in 20 years

→ Moving away from experience-driven, robustness-inspired methods is the way to go

→Yield power to computer science in problems where humans cannot compete!

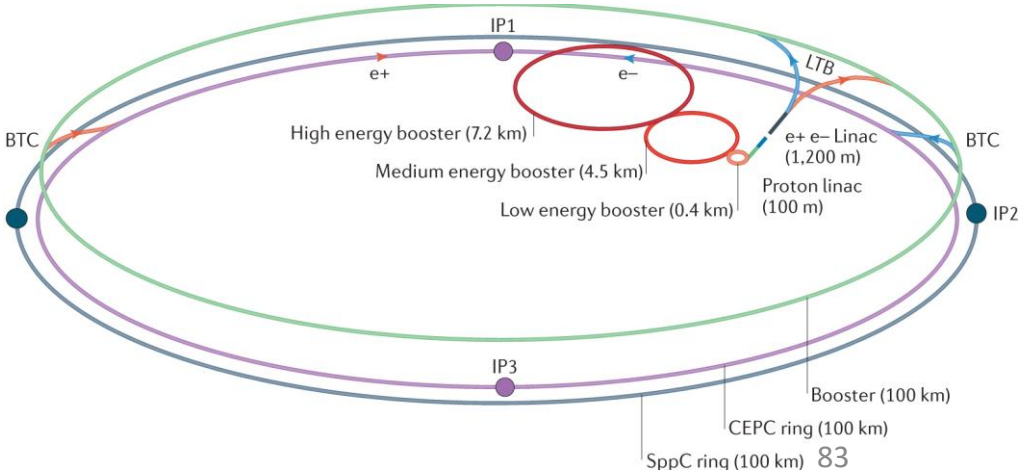
→Selling point is «human-in-the-middle, AI-assisted» design

- This avoids pissing off two generations of detector builders
- And is in fact what we really need



Above: *footprint of a future circular collider in the Geneva area. The LHC is shown for comparison*

Below: *CEPC design schematic*

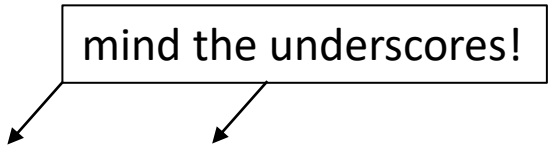


THANK YOU FOR YOUR ATTENTION!

For feedback, late questions, interest in joining MODE:

Please contact me at tommaso.dorigo@gmail.com

Also see my personal blog,
www.science20.com/quantum_diaries_survivor



and my research web page, www.pd.infn.it/%7Edorigo/index.html

References

[Kaggle 2014] https://www.kaggle.com/c/higgs-boson.

[De Castro & Dorigo 2019] P. de Castro Manzano and T. Dorigo, "*INFERNO: Inference-Aware Neural Optimization*", Computer Physics Comm. 244 (2019) 170, 10.1016/j.cpc.2019.06.007.

[Layer & Dorigo 2023] L. Layer, T. Dorigo, and G.C. Strong, "*Application of Inferno to a Top Pair Cross Section Measurement with CMS Open Data*", arXiv:2301.10358 [hep-ex] (2023).

[Dorigo 2020] T. Dorigo, "*Geometry Optimization of a Muon-Electron Scattering Detector*," Physics Open 4 (2020) 100022, arXiv:200200973[physics.ins-det], doi: 10.1016/j.physo.2020.100022.

[Abbiendi 2019] G. Abbiendi *et al.*, "*Letter of Intent: The MUonE Project*", CERN-SPSC-2019-026/SPSC-I-252 (2019).

[Shirobokov 2020] S. Shirobokov, A. Ustyuzhanin, A. Güneş Badyin *et al.*, "*Differentiating the Black-Box: Optimization with Local Generative Surrogates*", arXiv:2002.04632v1[cs.LG] (2020).

[MODE 2020] https://mode-collaboration.github.io

[Dorigo 2022] T. Dorigo et al., *"Toward the End-to-End Optimization of Particle Physics Instruments with Differentiable Programming: a White Paper", arXiv:2203.13818 [physics.ins-det] (2022).*

[Kieseler 2022] J. Kieseler, G. Strong, F. Chiandotto, T. Dorigo, and L. Layer, "*Calorimetric Measurement of Multi-TeV Muons via Deep Regression*", Eur. Phys. Journ. C82 (2022) 79, https://doi.org/10.1140/epjc/s10052-022-09993-5 .

[Dorigo and Guglielmini 2022] T. Dorigo, S. Guglielmini, J. Kieseler, L. Layer, and G.C. Strong, "*Deep Regression of Muon Energy with a K-Nearest Neighbor Algorithm*", arXiv:2203.02841 [hep-ex] (2022).

Backup

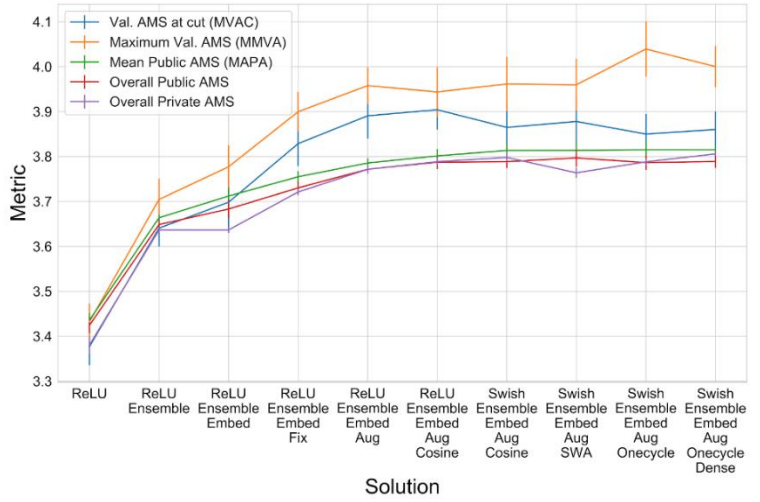
A Side Note: Looking Deeper Into the Best Solution

In a recent study [Strong 2020] the performance of Gabor Melis' solution was reproduced with a similar but much more GPU-efficient setup

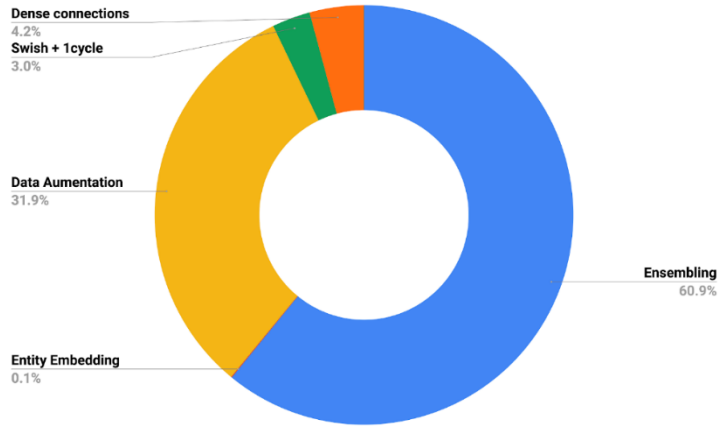
The study focuses on the ingredients which were the most useful to improve the solution: Data augmentation + Swish activation + Densely connected layers = more performant model

- Require fewer models in ensemble to achieve same performance (10 versus 70)
- 1-cycle schedule, quicker to train
- Accounting for difference in GPU processing power, new solution:
  - 13x quicker to train on GPU, 7x quicker on laptop CPU
  - 33x quicker to apply on GPU, 3x quicker on laptop CPU

	Our solution	1st place	2nd place	3rd place
Method	10 DNNs	70 DNNs	Many BDTs	108 DNNs
Train-time (GPU)	8 min	12 h	N/A	N/A
Train-time (CPU)	14 min	35 h	48 h	3 h
Test-time (GPU)	15 s	1 h	N/A	N/A
Test-time (CPU)	3 min	???	???	20 min
Score	3.806 ± 0.005	3.80581	3.78913	3.78682



Above: *figure of merit as a function of successive improvements*
Below: *relative importance of the improvements*



[Strong 2020] G. Strong, "*On the impact of selected modern deep-learning techniques to the performance and celerity of classification models in an experimental high-energy physics use case",* Mach. Learn.: Sci. Technol. 1 (2020) 045006, https://doi.org/10.1088/2632-2153/ab983a.

A Study of Muon Shielding in SHIP

In another seminal work [Shirobokov et al. 2020], local generative surrogates of the gradient of the objective function allowed for SGD minimization

→ Strong reduction in muon background fluxes (relevant metric) in the SHIP experiment (see below)

[Submitted on 11 Feb 2020 (v1), last revised 15 Jun 2020 (this version, v2)]

Black-Box Optimization with Local Generative Surrogates

Sergey Shirobokov, Vladislav Belavin, Michael Kagan, Andrey Ustyuzhanin, Atılım Güneş Baydin

We propose a novel method for gradient-based optimization of black-box simulators using differentiable local surrogate models. In fields such as physics and engineering, many processes are modeled with non-differentiable simulators with intractable likelihoods. Optimization of these forward models is particularly challenging, especially when the simulator is stochastic. To address such cases, we introduce the use of deep generative models to iteratively approximate the simulator in local neighborhoods of the parameter space. We demonstrate that these local surrogates can be used to approximate the gradient of the simulator, and thus enable gradient-based optimization of simulator parameters. In cases where the dependence of the simulator on the parameter space is constrained to a low dimensional submanifold, we observe that our method attains minima faster than baseline methods, including Bayesian optimization, numerical optimization, and approaches using score function gradient estimators.

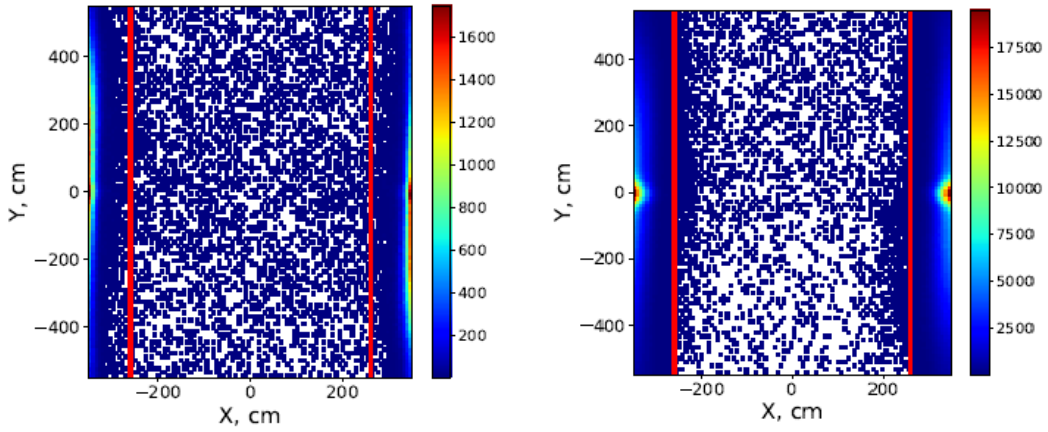
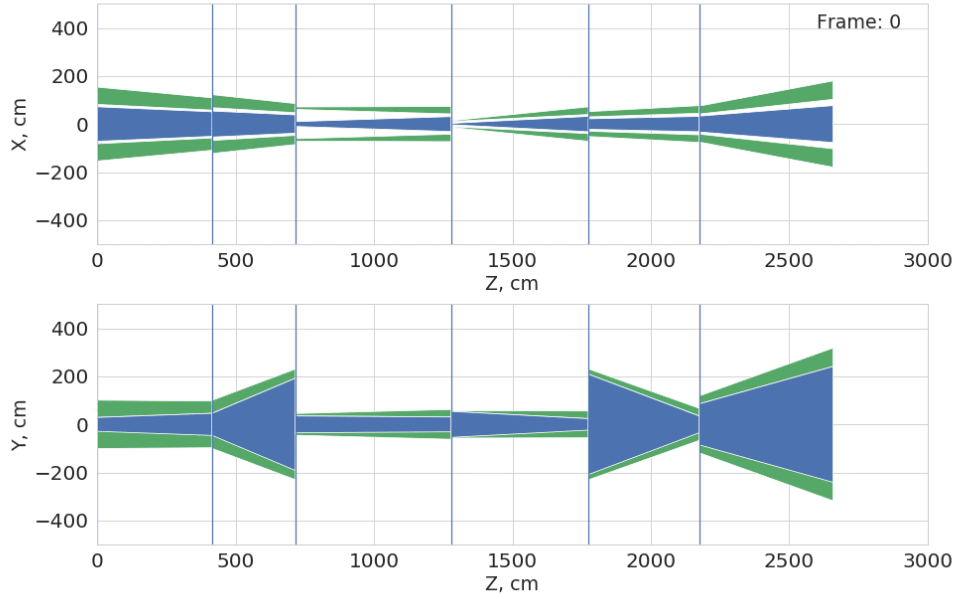


Figure 7. Muon hits distribution in the detection apparatus (depicted as red contour) obtained by Bayesian optimization (Left) and by L-GSO (Right), showing better distribution. Color represents number of the hits in a bin.



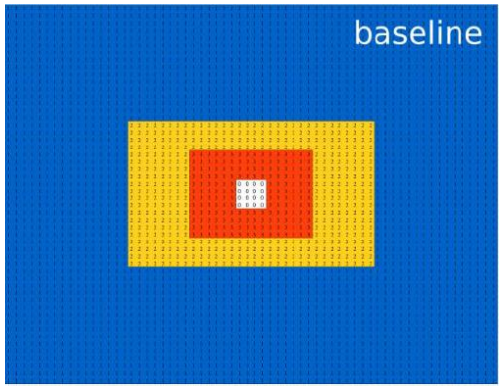
Above: *Geometry optimization of the shape of shielding magnets for SHIP*

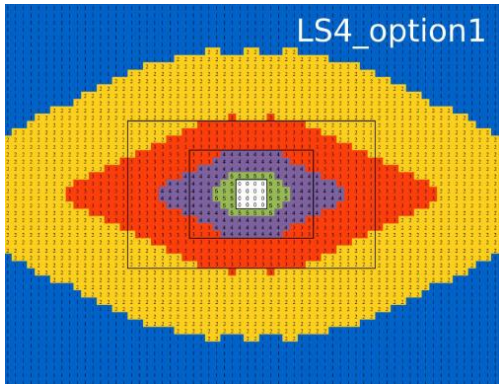
LHCb EM Calorimeter Optimization

MODE members who collaborate with LHCb (**A. Boldyrev, F. Ratnikov, A. Ustyuzhanin, D. Derkach**) recently optimized the electromagnetic calorimeter for the LHCb upgrade using a differentiable model

Creating a model of particle showers is a hard problem, which can be solved w/ GANs

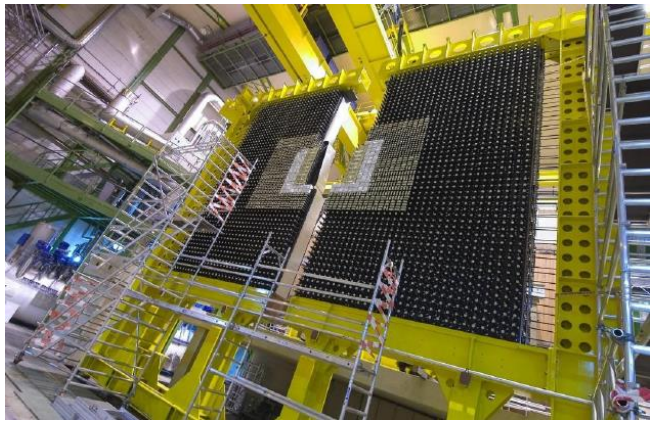
The developed model allows to investigate how to best arrange modules of three different kinds, optimizing cost and final performance metrics.



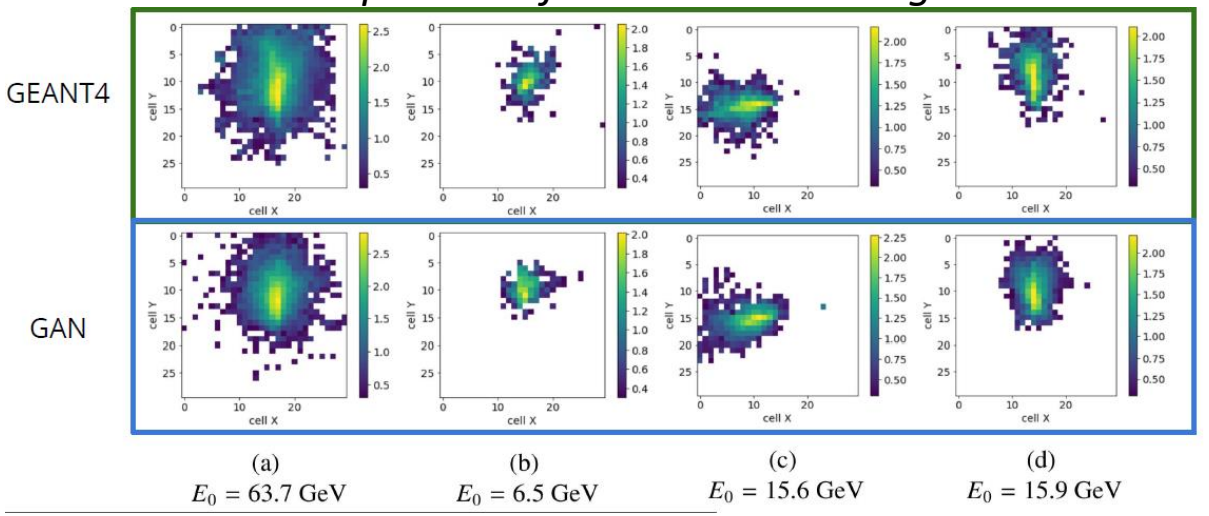


Above: *PMT configurations*

Left: *the LHCb EM calorimeter front face*
Right: *the three photomultiplier modules*



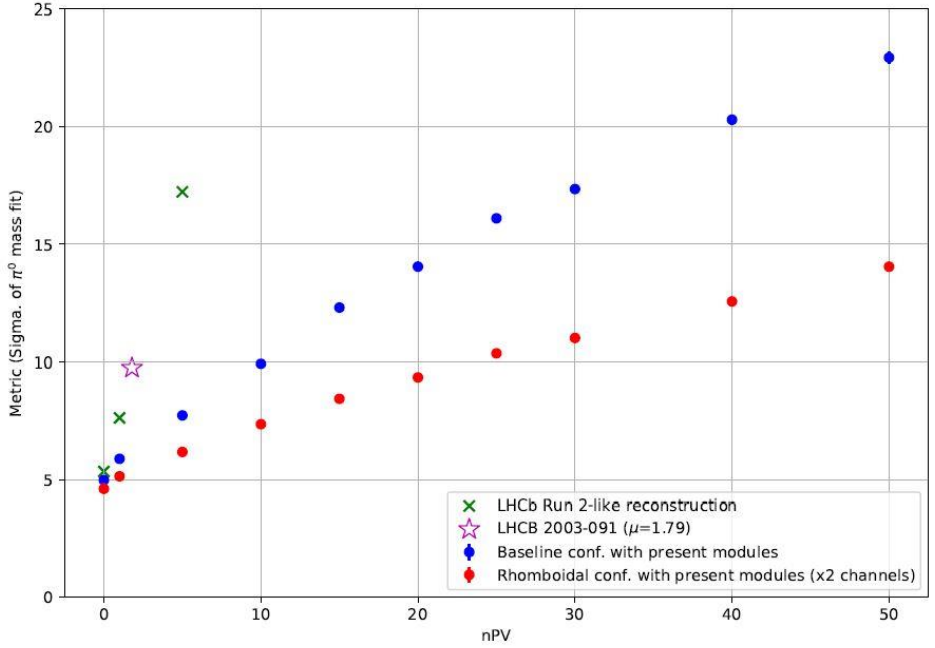
Below: *comparison of showers validating GAN model*



Cost Options for LHCb EM Calorimeter

Shown below (right) is the tradeoff of significance to a relevant physics signal (B_s meson decays to $J/\psi\ \pi^0$) versus the detector complexity, for independently optimized configurations

Below: optimization metric as a function of event complexity (number of primary vertices)



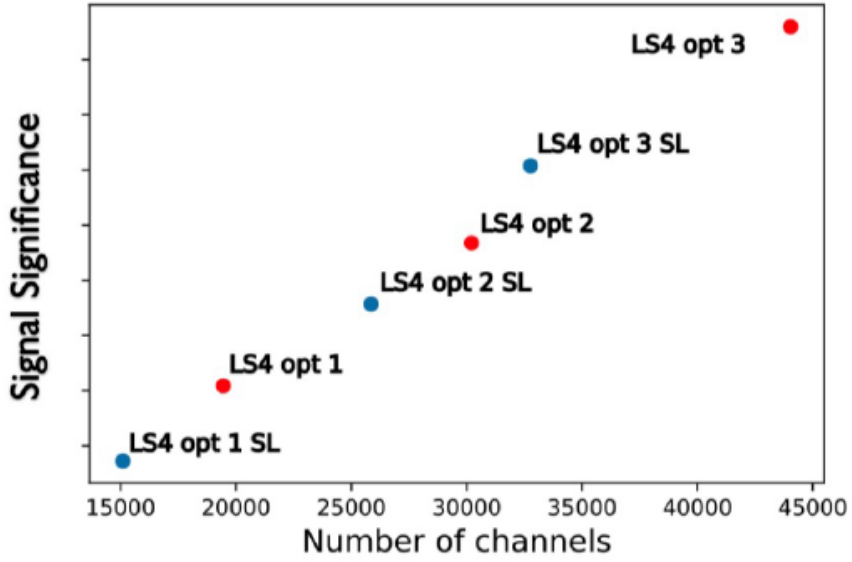
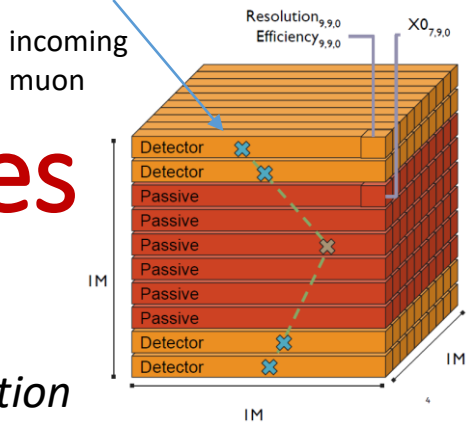


Figure 12: Dependency of the optimized Physics signal significance for different calorimeter configurations on the total number of readout channels used in those configuration. Dependencies like this help to make an educated decision about the optimal detector configuration which provides the best balance between the ultimate physics performance and construction costs.

Example: Optimize Layout of Detector Planes

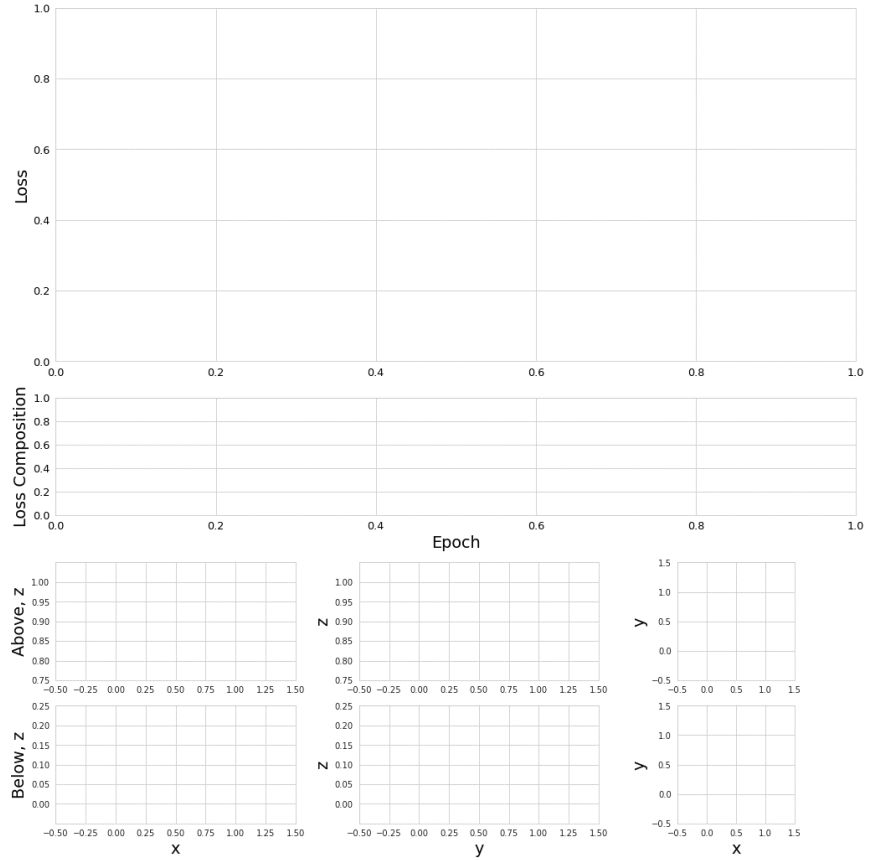


Above: *concept of scattering detection*
Below: *loss minimization and placement of detector panels*

The animation shows the result of a run of the algorithm, which finds optimal placement of detection layers for a muographer searching for U blocks in scrap metal (a $1m^3$ container) using batches of 250 muons.

Loss is combination of detector cost and precision of inference on existence of U block in set of randomly generated volumes

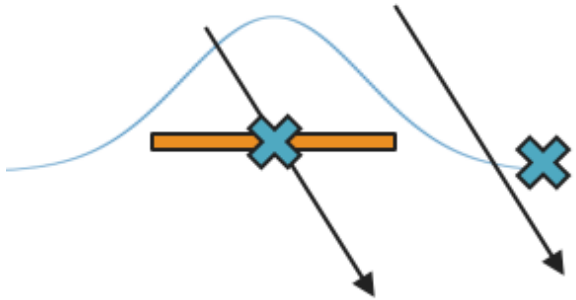
→ A first proof of principle of correct training of a differentiable model of a schematic muon tomography apparatus!



Detector Discreteness in TomOpt

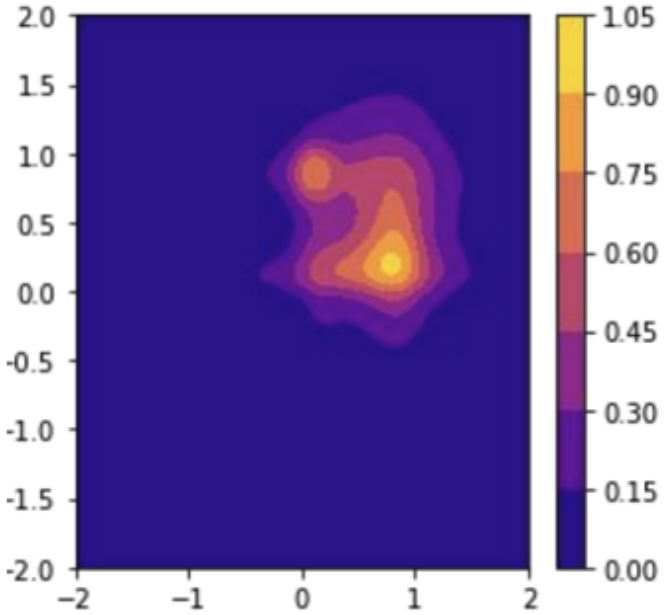
Moving a detector plane is not a differentable operation, as a muon in a batch may go from being detected to not being detected

Fix with Gaussian model of position resolution or efficiency when doing SGD (real dimensions considered at eval time)



Both muons recorded, but with different resolutions

Can model arbitrary situations with mixtures



Plot: Max Lamparth

Bonus Track:
Can We Do Deep Learning with a k-NN ?

The CNN results shown were actually produced several months *after* we first tried a k-NN algorithm on the problem, because I was curious about how far one could go with an old-fashioned statistical tool

The k-NN was eventually re-run on the data used for the CNN publication, after many improvements. Results in [Dorigo and Guglielmini 2022].

Is it competitive with NNs and boosted methods ? Let us have a quick look...

High-Level Features

A k-NN cannot possibly make sense of a 51,200-dimensional space, so we cooked up 28 high-level features to aggregate information in various ways:

- total energy in cells for various values of min cell energy
- moments of energy distribution around track in xy space, in z slices
- number of clusters of cells (above energy threshold)
- number of towers in clusters
- energy of clusters
- imbalance in x and y of energy distribution
- fit to curvature from energy depositions (useful for E in few-100 GeV range)

kNN Construction

The kNN estimate was constructed as the average of several pools of weak learners, each pool trained independently on partially disjunct subsets of training data.

To address the curse of dimensionality, which makes even 28 dimensions too many for local averaging, we define the distance in subspaces by ignoring some of the features through an indicator function, I(d) = 1/0 if feature d is considered or not:

$$\Delta(i,j)^2 = \sum_{d=1}^{D} I(d)(x_i^{(d)} - x_j^{(d)})^2$$

Features in the definition above are standardized to have unit variance and zero mean.

The prediction for test event j can be written as

$$E(j) = \frac{\sum_{m=1}^{k} E(i_{\mathrm{kNN}}(m))}{k}$$

where $i_{kNN}(m)$ is the index of the m-th closest training event to j, according to Δ(m,j).

Pooling of Weak Learners

The mentioned solution of the curse of dimensionality brings a loss of information and turns the problem into one of identifying advantageous subspaces through a good choice of I(), or to combine them.

To reduce information loss, we consider N_{wl} weak learners, each performing a kNN average in a different subspace through different indicators $I_{wl}()$.

The regressors are then combined in a weighted average:

$$E(j) = \sum_{i=1}^{N_{wl}} W_{wl}(i) E_i(j)$$

Weights $W_{wl}(i)$ are optimized by gradient descent.

The loss of information remains, but its effect can be tamed by the added flexibility of the model, and optimized weights W_{wl}.

Overparametrization

One of the aces up the sleeve of DNNs is overparametrization. To inject some in a kNN one may only rely on training data.

Each training event affects the prediction by its position in space (fixed) and by its muon energy (also fixed). One can still inject flexibility by biasing the energy value, and altering the weight of the event in the averaging.

We introduce two sets of parameters b(wl,i), w(wl,i) for each weak learner:

$$E_{wl}(j) = \frac{\sum_{m=1}^{k} E_{wl}(i_{\mathrm{kNN},wl}(m))w(wl,m)}{\sum_{m=1}^{k} w(wl,m)} + \sum_{m=1}^{k} b(wl,m))$$

with O(600k) training events, O(10) weak learners per pool, and O(5) pools, this will boil down to about **66M free parameters** in the final model.

But can they be trained ??

Weights and Biases Initialization and Learning

W() and b() need to be initialized. b() is set to zero, W() to a value that is =1.0 below 5 TeV (4 TeV is the upper limit of the regression range), and smoothly decays to 0 for E=8 TeV with a sigmoid (same approach was used by CNN paper).

During training, the weights and biases get tweaked by the learning process. For W() you still see some remainder of the initial trend, but with large spread

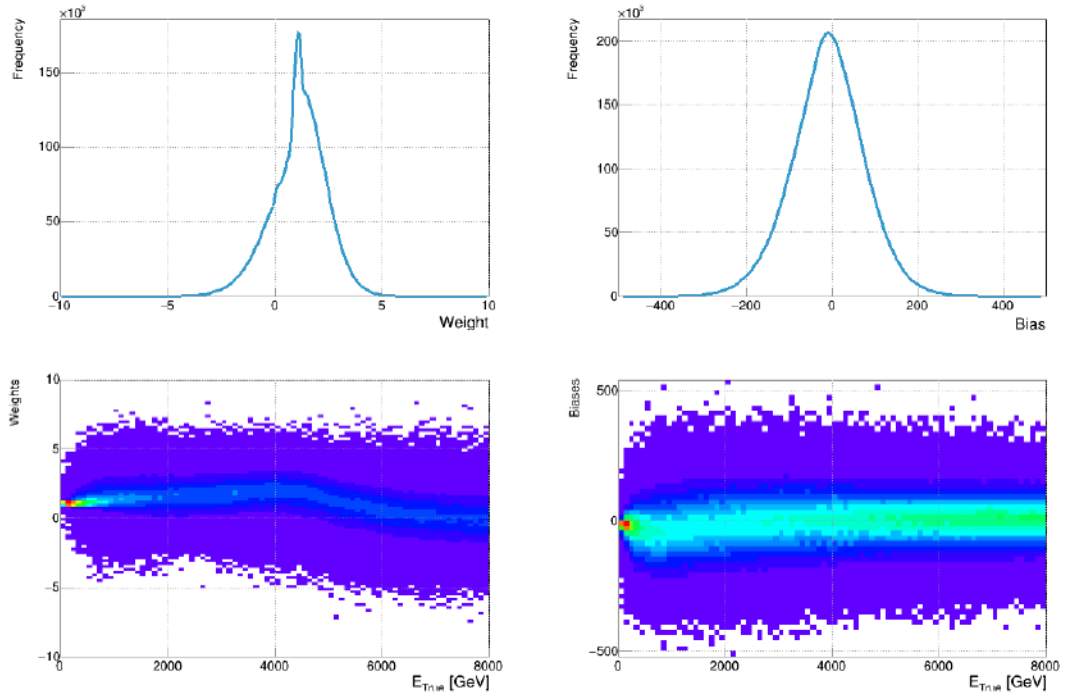


Figure 7: Example of the distribution of weights obtained by an optimization run with 400,000 training events and 5 learners. Top left: distribution of event weights; top right: distribution of event biases (in GeV). Bottom left: weights versus true muon energy; bottom right: biases (in GeV) versus true muon energy.

Results

For a comparison, we look at results of a NN (orange), a default kNN (pink), and Xboost.

The k-NN result outperforms the standard k-NN, is overall similar to those of the ML methods, and is slightly better at high E.

But the CPU and analysis load is non comparable!

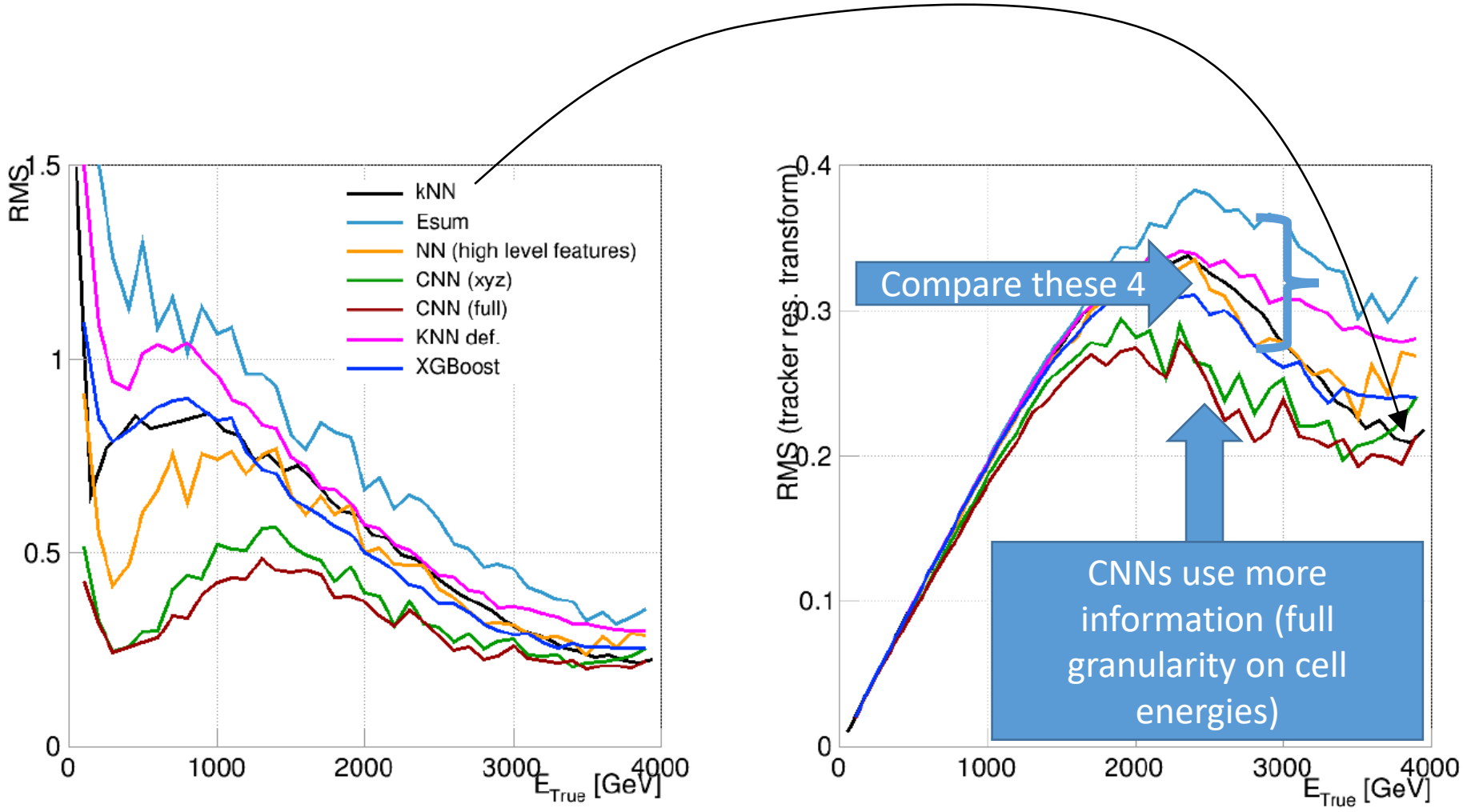


Figure 13: Left: comparison of the mean squared error of predictions of different algorithms employing the same training and test data. Right: comparisons of the mean squared error of the combinations of tracker and calorimeter regressed predictions. Black: deep regression kNN (described in this article); light blue: energy sum model; orange: neural network with high level features; green: CNN with spatial features; red: full CNN; magenta: classical kNN; blue: XGBoost.