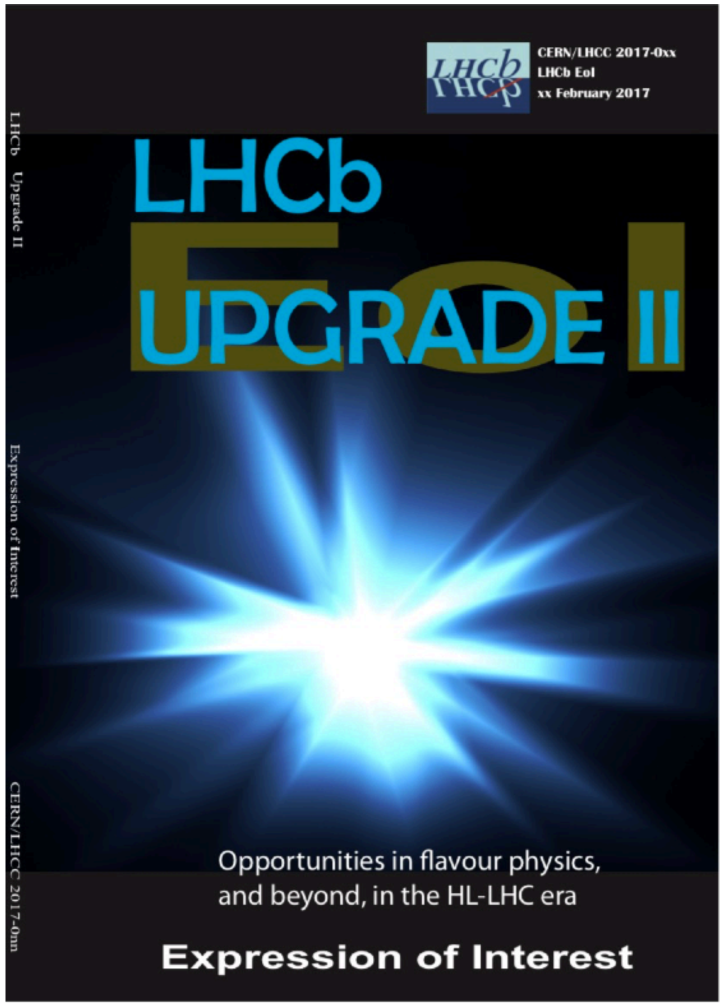
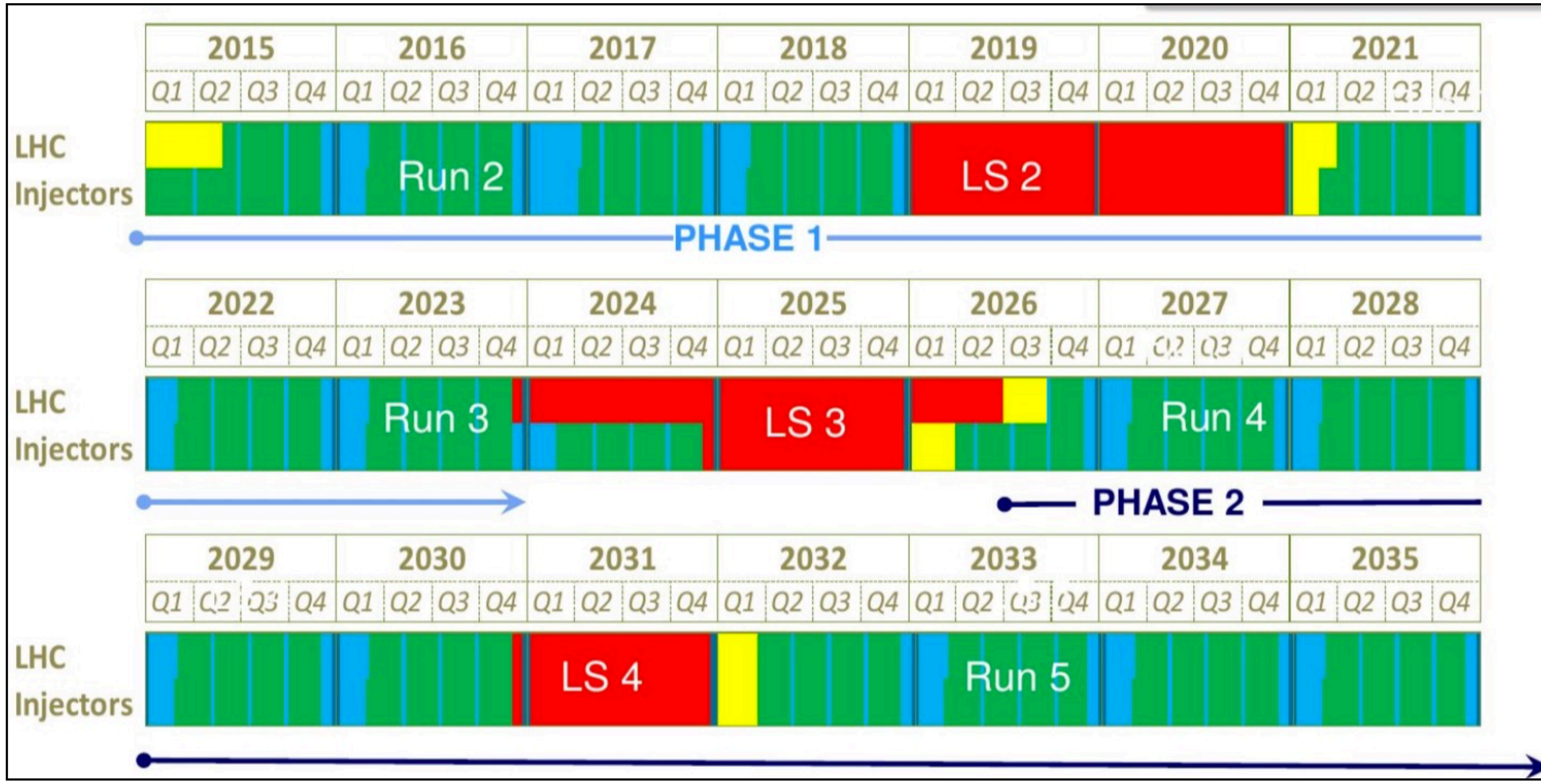
Using Machine Learning to Speed Up Calorimeter R&D

Fedor Ratnikov on behalf of the LHCb Calorimeter Upgrade group
LambdaLab, NRU Higher School of Economics
Yandex School of Data Analysis

SCHOOL OF DATA ANALYSIS

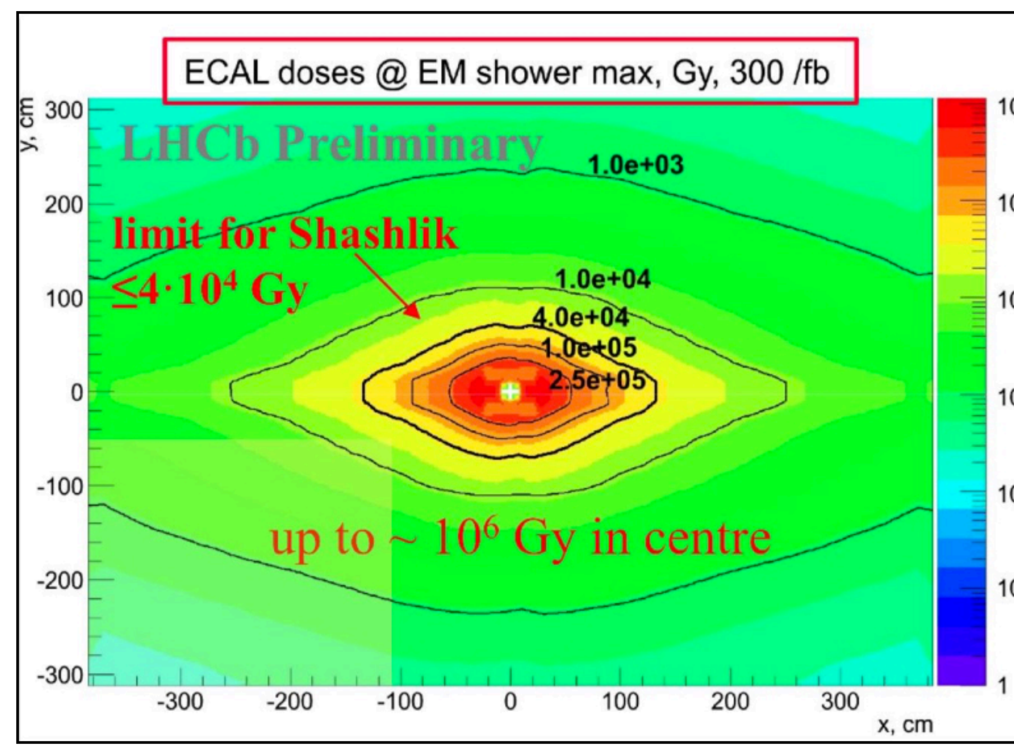
Inspiration





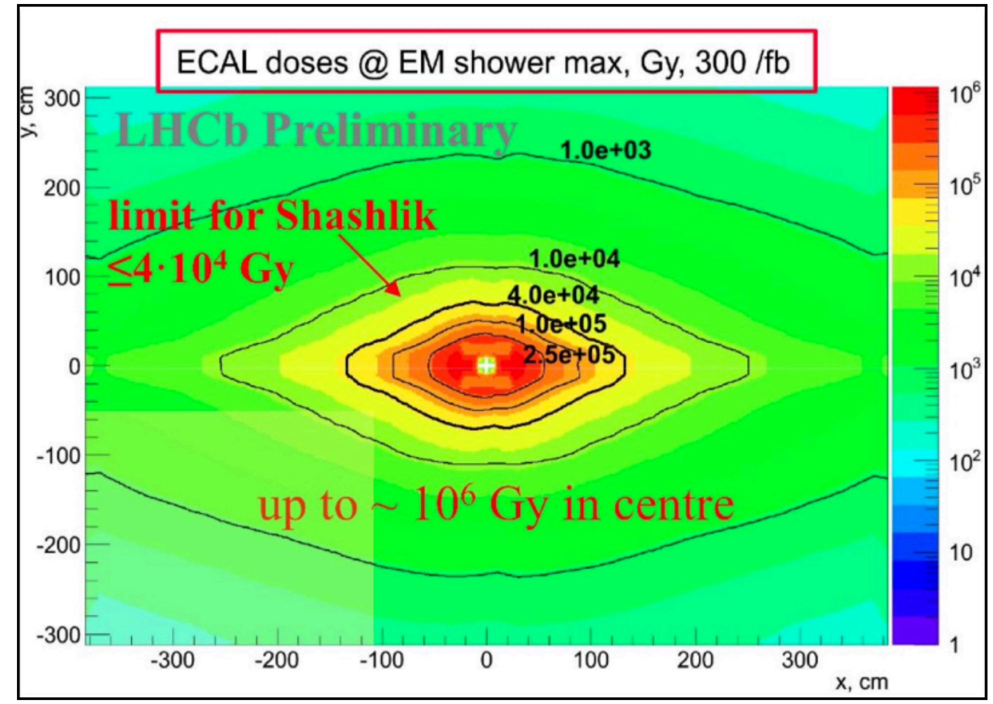
LHCb Upgrade 2 targets Run 5&6

- $1.5e34\ cm^{-2}c^{-2}$ instantaneous luminosity



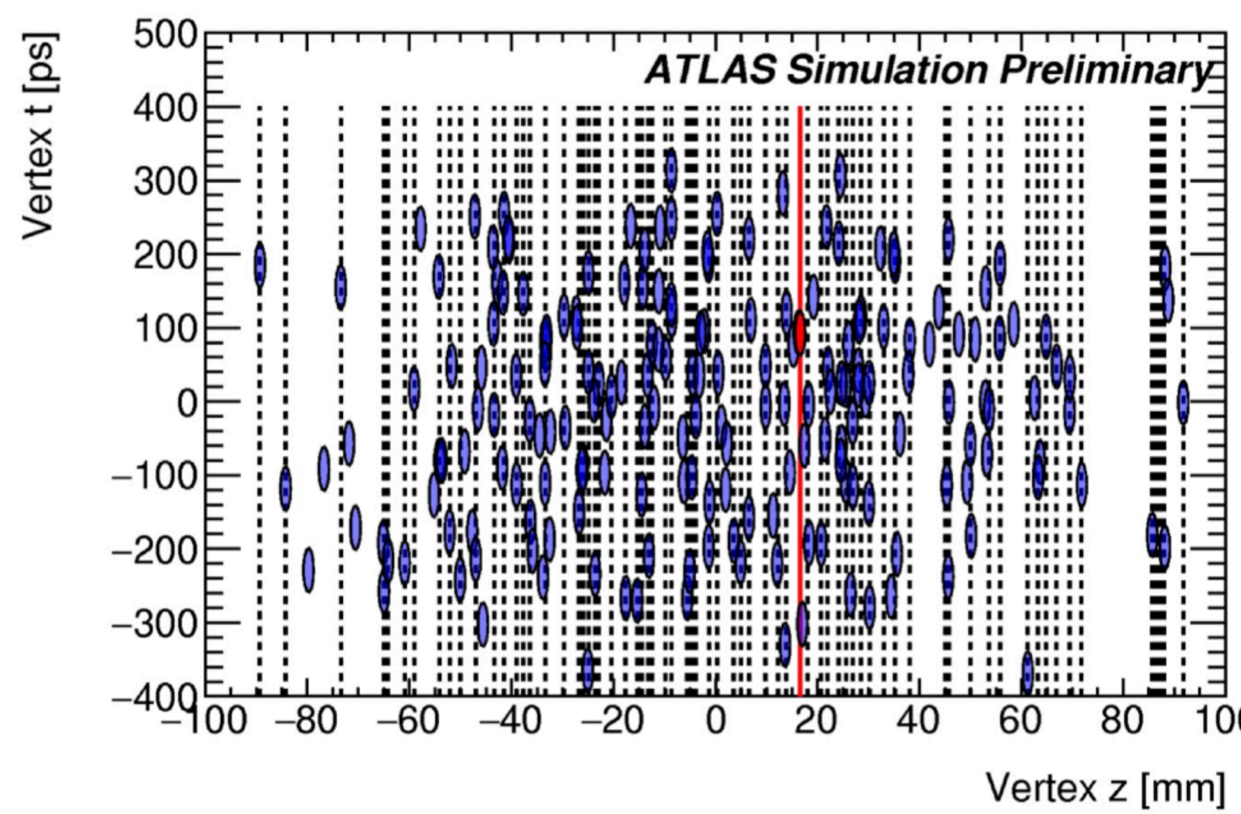
Inspiration

Homogeneous Crystal	Shashlik Module	Spaghetti Module
Requires long crystal (order 40 cm) to contain 25 X_0	Can be made very compact (~ 15-20 cm)	Can be made very compact (~ 15-20 cm)
Fixed Moliere Radius	Tunable Moliere Radius	Tunable Moliere Radius
Good energy resolution, few %/√E (but material budget in front of ECAL should be kept at minimum)	Good energy resolution	Challenging optimization to reach good energy resolution
Very good homogeneity	**No radiation-hard WLS fibers (yet) to transport light**	**Fibers scintillate AND transport light!**
Large volume of crystal ➡ high cost	Some cost optimization possible	Some cost optimization possible

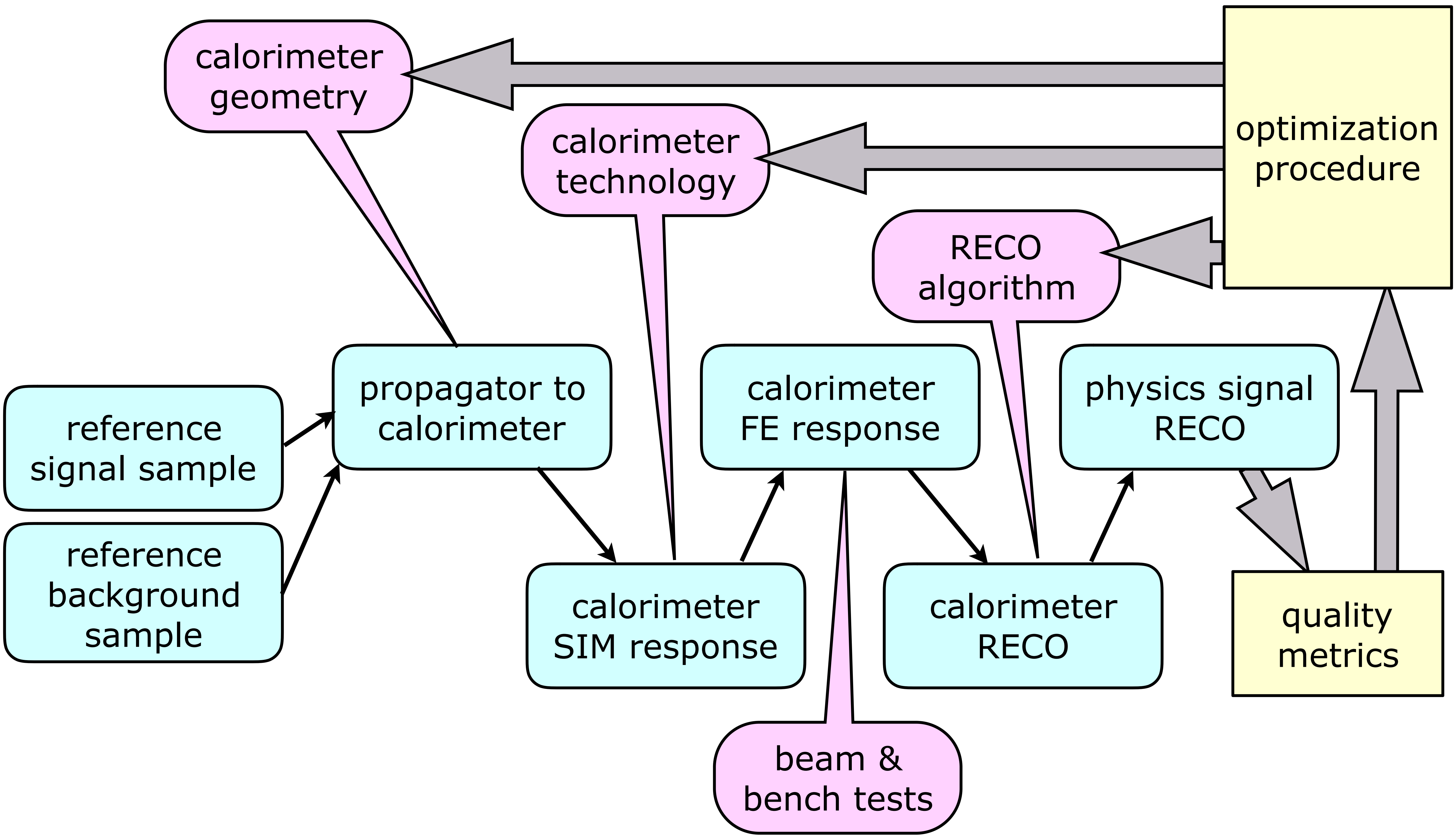


Requires extensive R&D studies for U2 LHCb ECAL

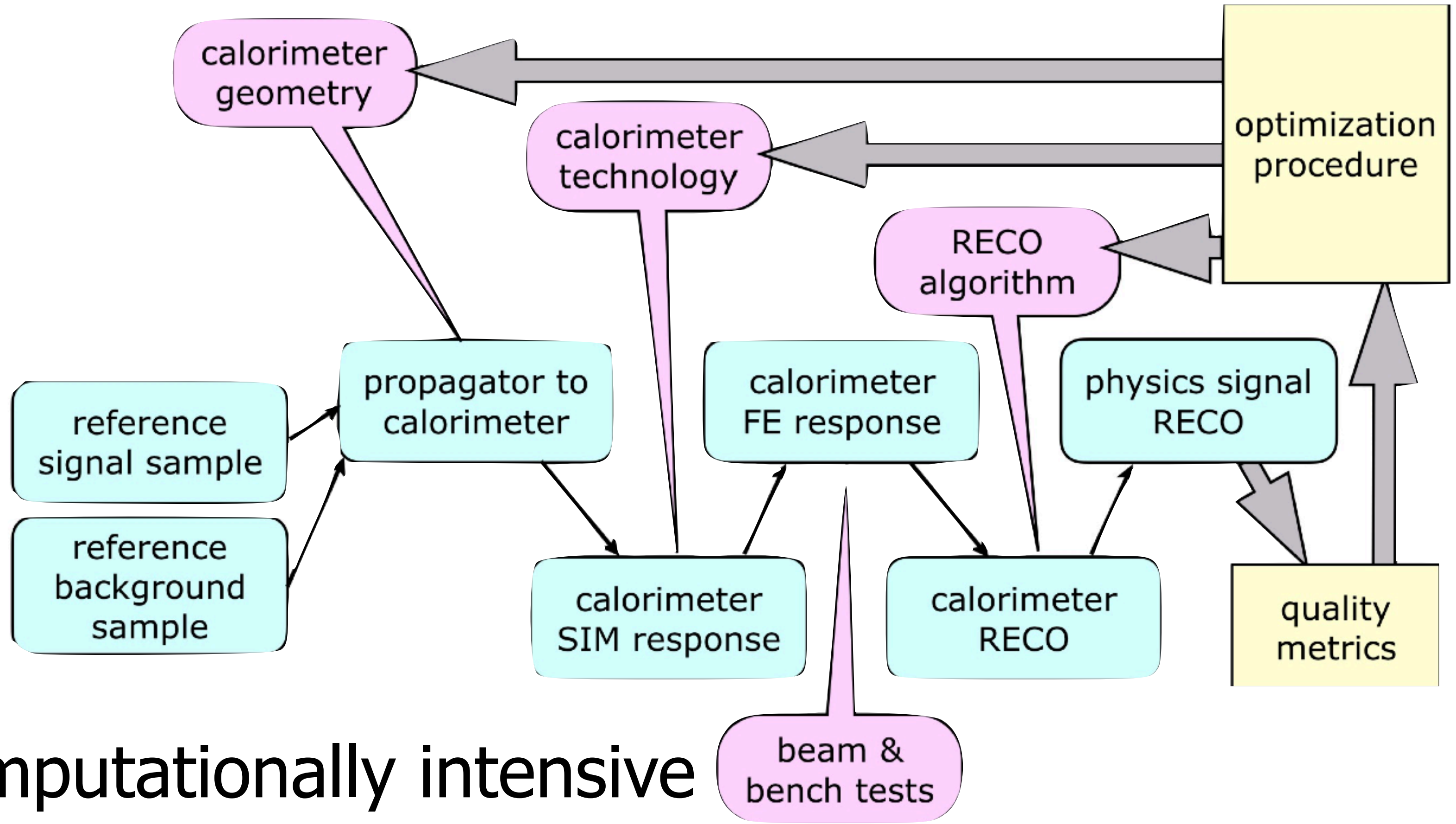
- module technologies
- module configuration
- readout properties
- timing properties
- installation geometry
- …



Optimization Cycle



Optimization Cycle



Bottlenecks:

- calorimeter simulation is computationally intensive
  - shower development
  - photons transport
- direct beam and bench tests hard to directly include into simulation stack
- RECO algorithm needs tuning for the particular module technology/ geometry/configuration
- multi-parametric optimization may be expensive

ML in the Optimization Cycle

Machine Learning provides a set of tools and methods which allow effective fit of multi-dimensional data to non-parametric (generic) functions

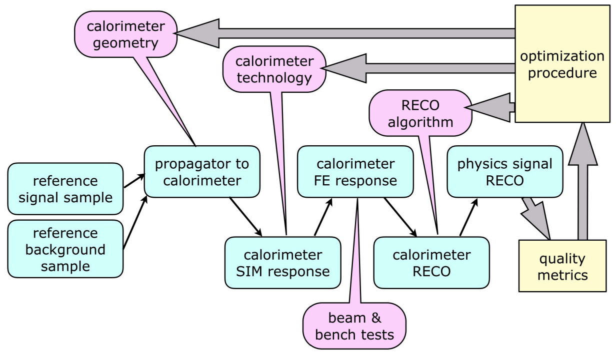
- allow quick turn over for the optimization cycle, when parameters are changed
- eliminate manual work for re-tuning simulation and reconstruction

ML model may be suboptimal comparing to “the best” solution

- however it catches main features, that is usually good enough to estimate physics performance and give feedback to ongoing detector R&D

I’ll demonstrate how it can be used in LHCb U2 ECAL inspired environment

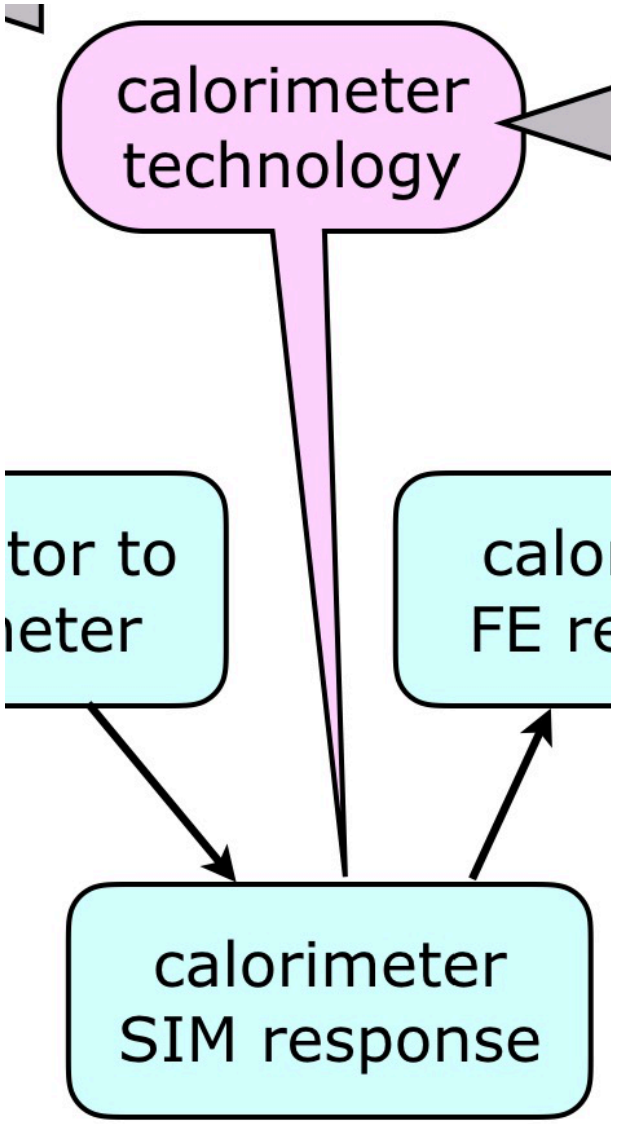
Fast SIM Response



Generate sample of calorimeter local responses to single particles with GEANT

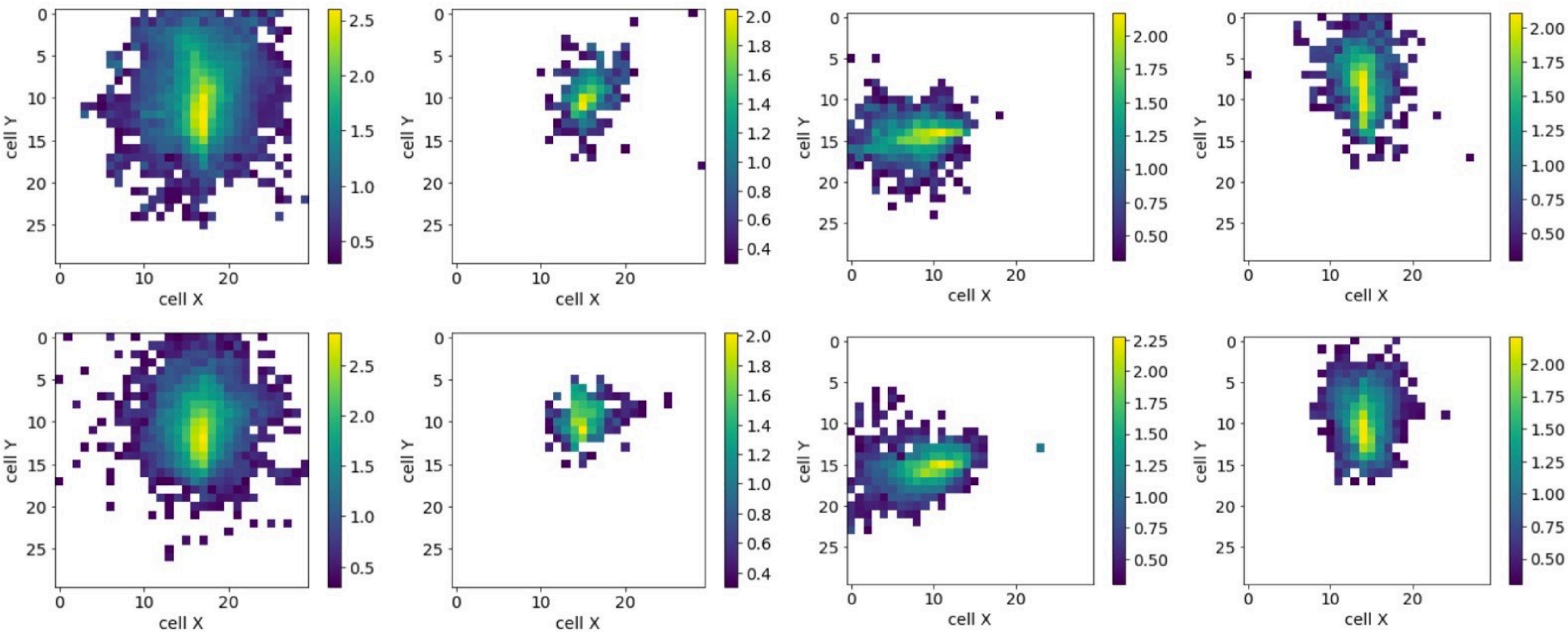
Different approaches may be used then

- use generated sample as an object library and pick the best object in the optimization cycle using fast search techniques
- train a ML generative model which reproduce major features of original response matrix



Detector Response

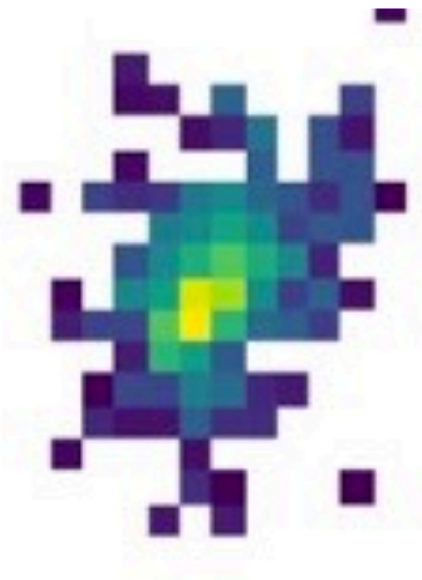
EPJ Web of Conferences 214, 02034 (2019)



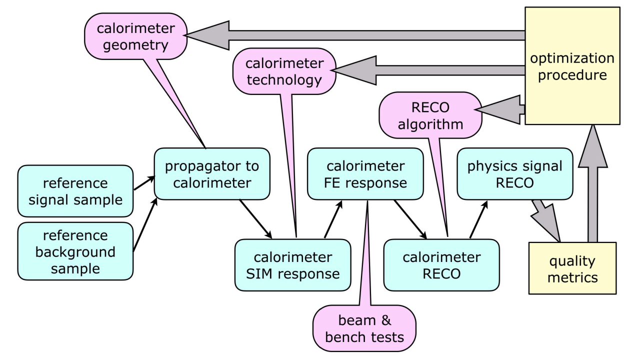
(a) $E_0 = 63.7$ GeV (b) $E_0 = 6.5$ GeV (c) $E_0 = 15.6$ GeV (d) $E_0 = 15.9$ GeV

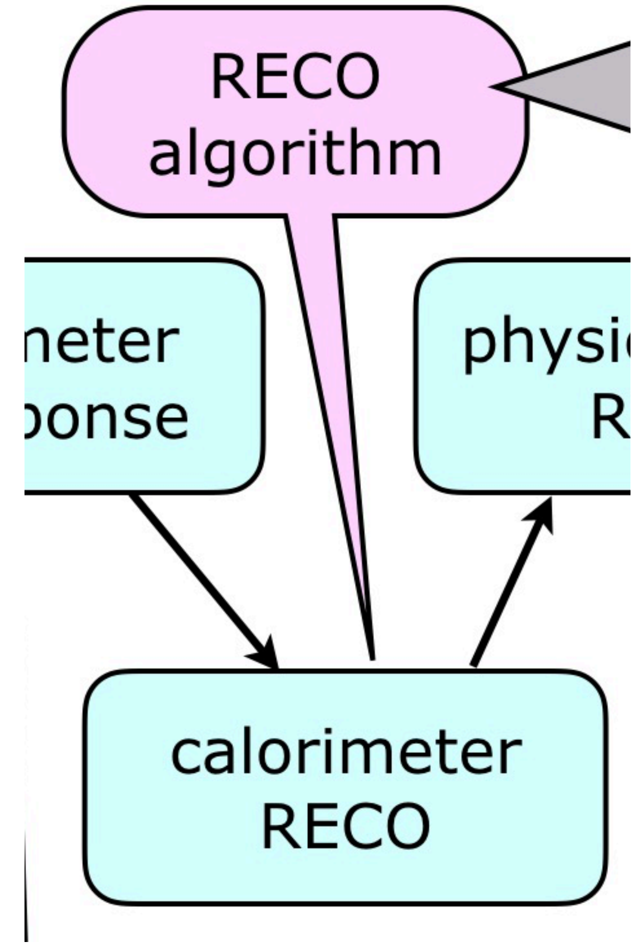
Generative model trained on the sample produced by Geant4 for LHCb-inspired shashlik ECAL technology produces good enough clusters at 1e-5 smaller CPU time

Spatial Reconstruction



$\rightarrow \quad (X, Y)$





Simple regression: $x_{CM} \rightarrow X,\ y_{CM} \rightarrow Y$

More complicated: $\left(x_{CM}, y_{CM}, \frac{dx}{dz}, \frac{dy}{dz}, \sum E_{i,j}\right) \rightarrow (X, Y)$

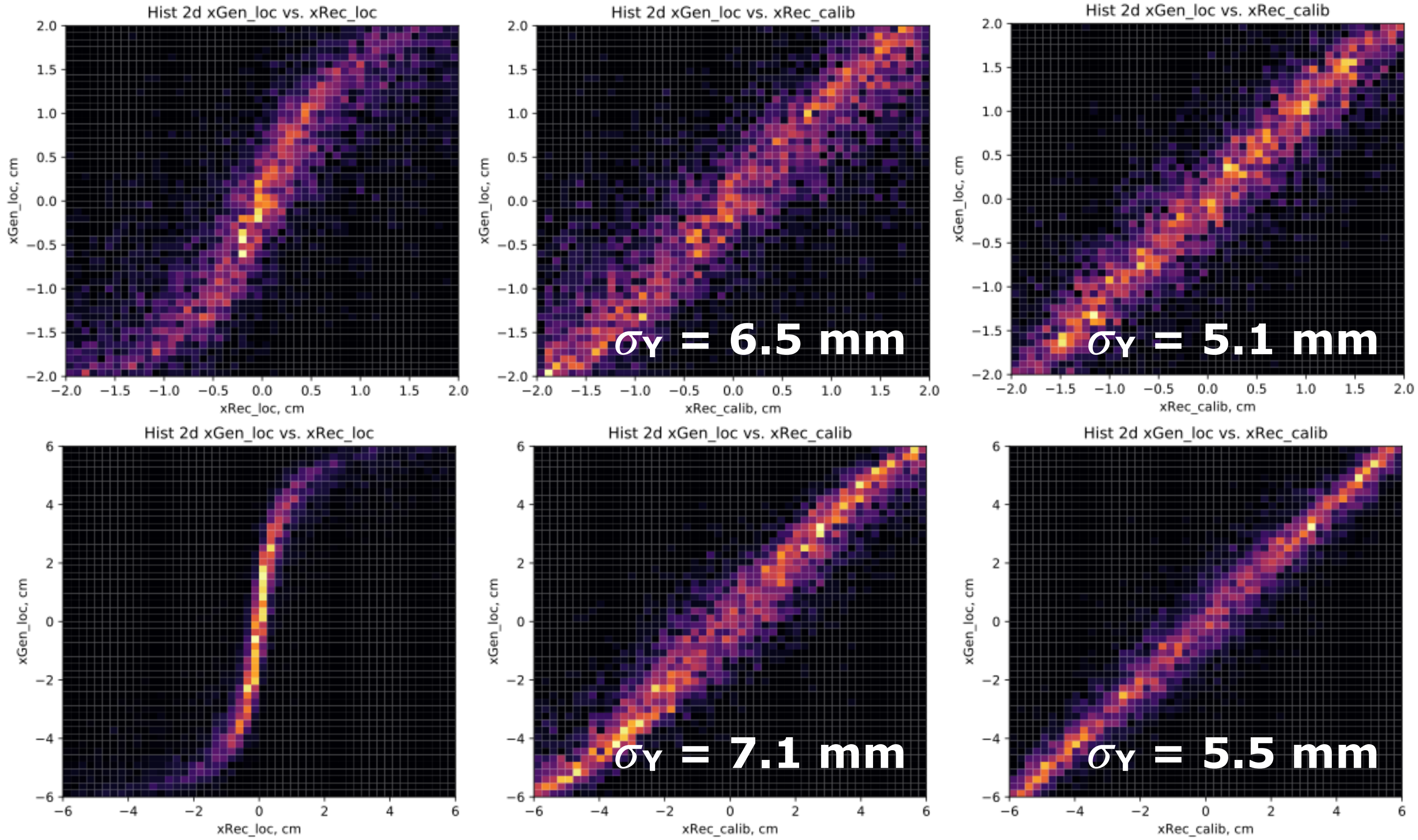
Full regression: $\left(E_{i,j}, \frac{dx}{dz}, \frac{dy}{dz}\right) \rightarrow (X, Y)$

Spatial Reconstruction

original | parametric RECO | non-parametric ML RECORECO

ECAL module 4 cm

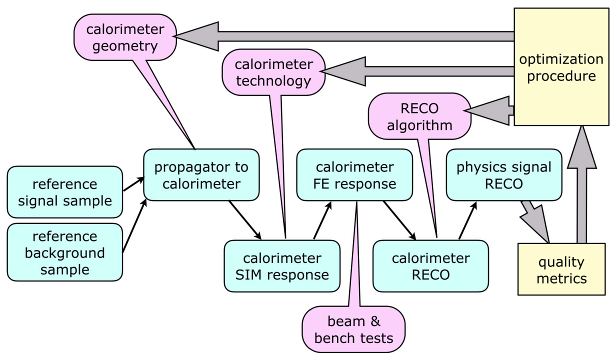
ECAL module 12 cm



Train ML regressor (xgboost) to reconstruct coordinate

- non-parametric blindly trained ML regressor well reproduce manually-tuned parametrized reconstructor
- ML approach is agnostic to various calorimeter properties

Pileup Mitigation with Timing

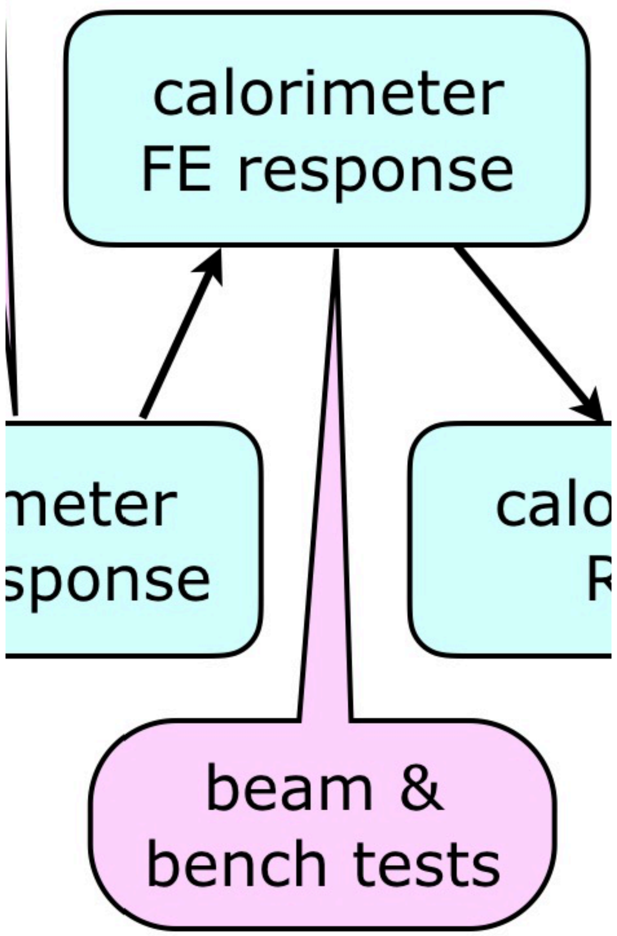


Performance strongly dependent on peculiar details of detector and electronics behaviors

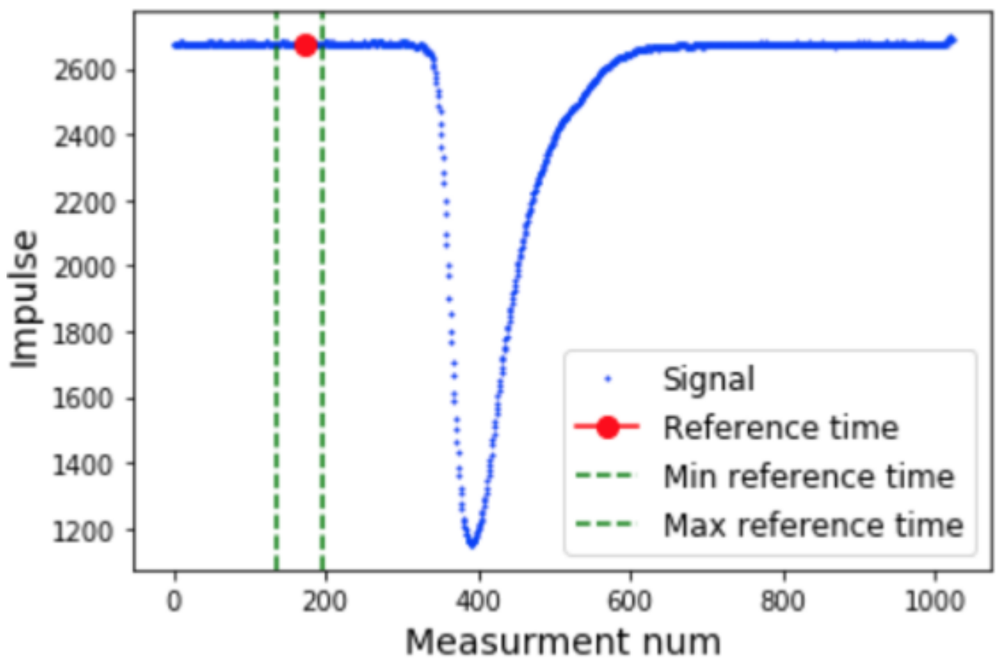
- hard to reproduce in simulation

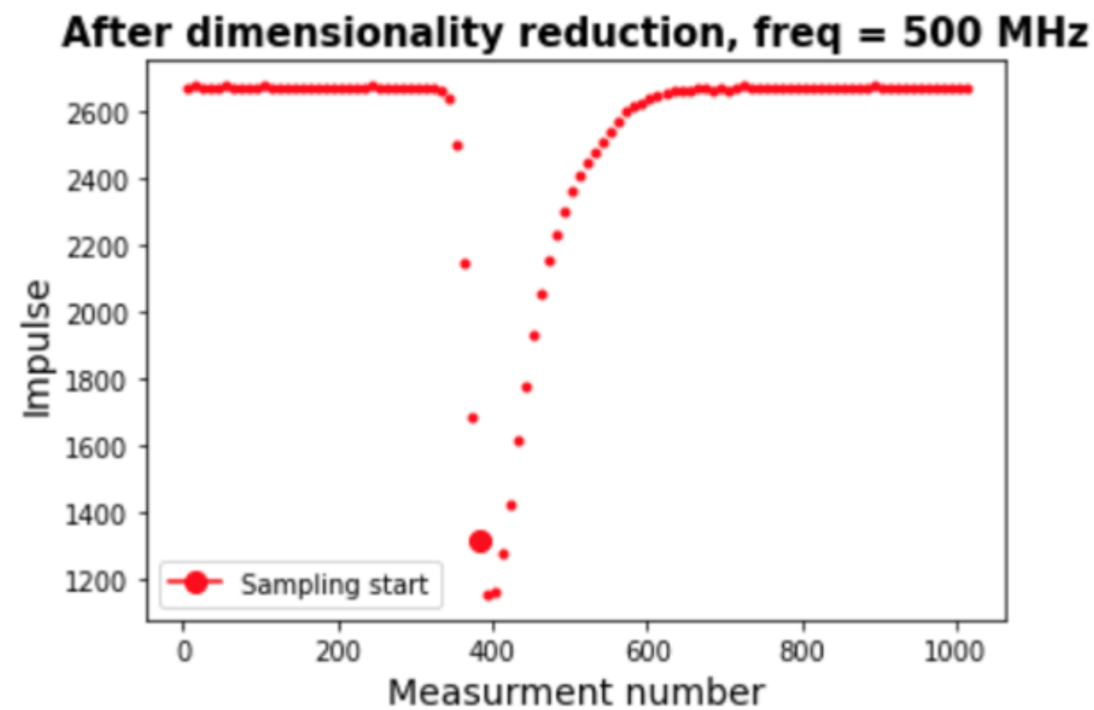
Test beam results provide the latest and greatest information

- evaluate important features
- calibrate simulation on measured points



Data Used





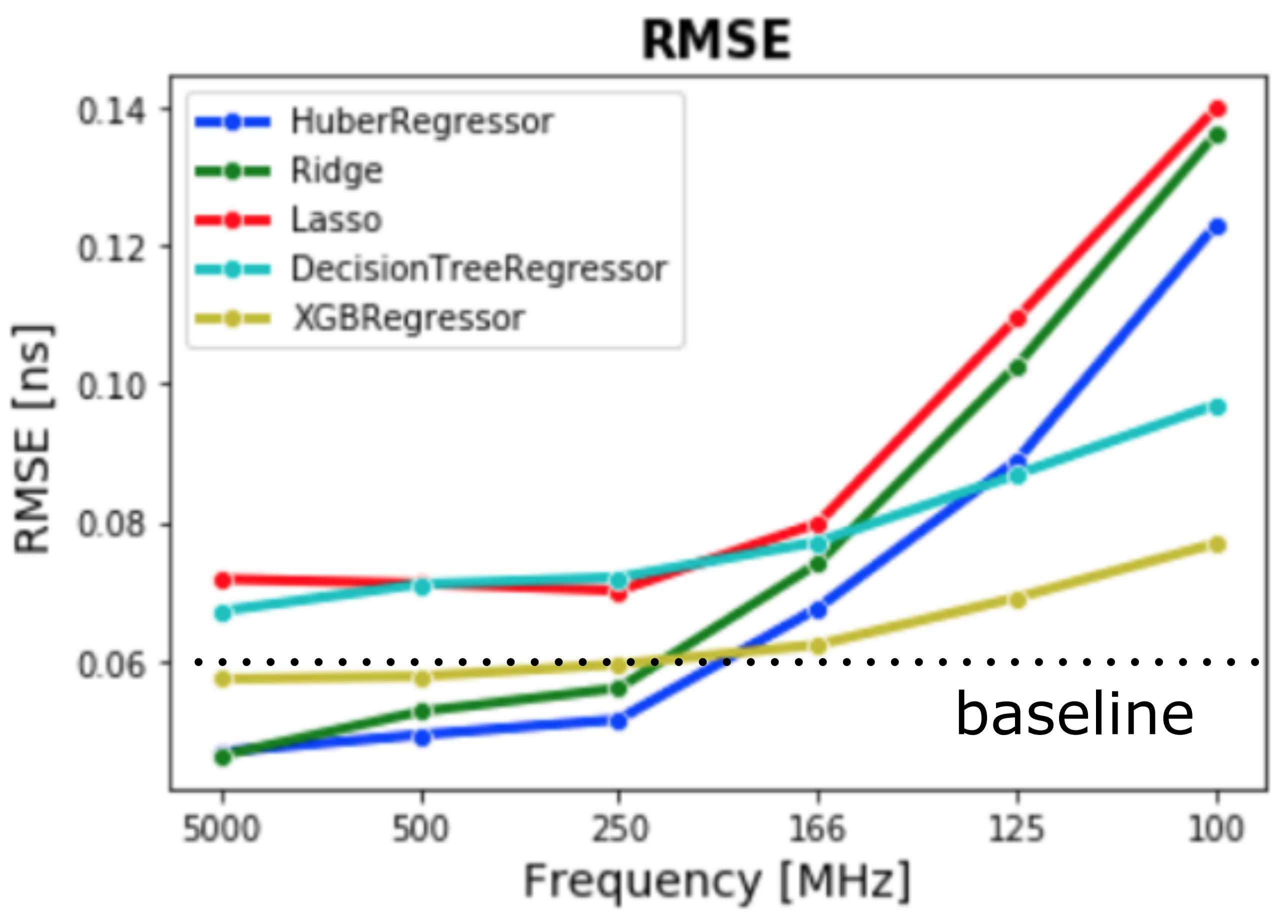
Data obtained from the 30 GeV electron beam @DESY

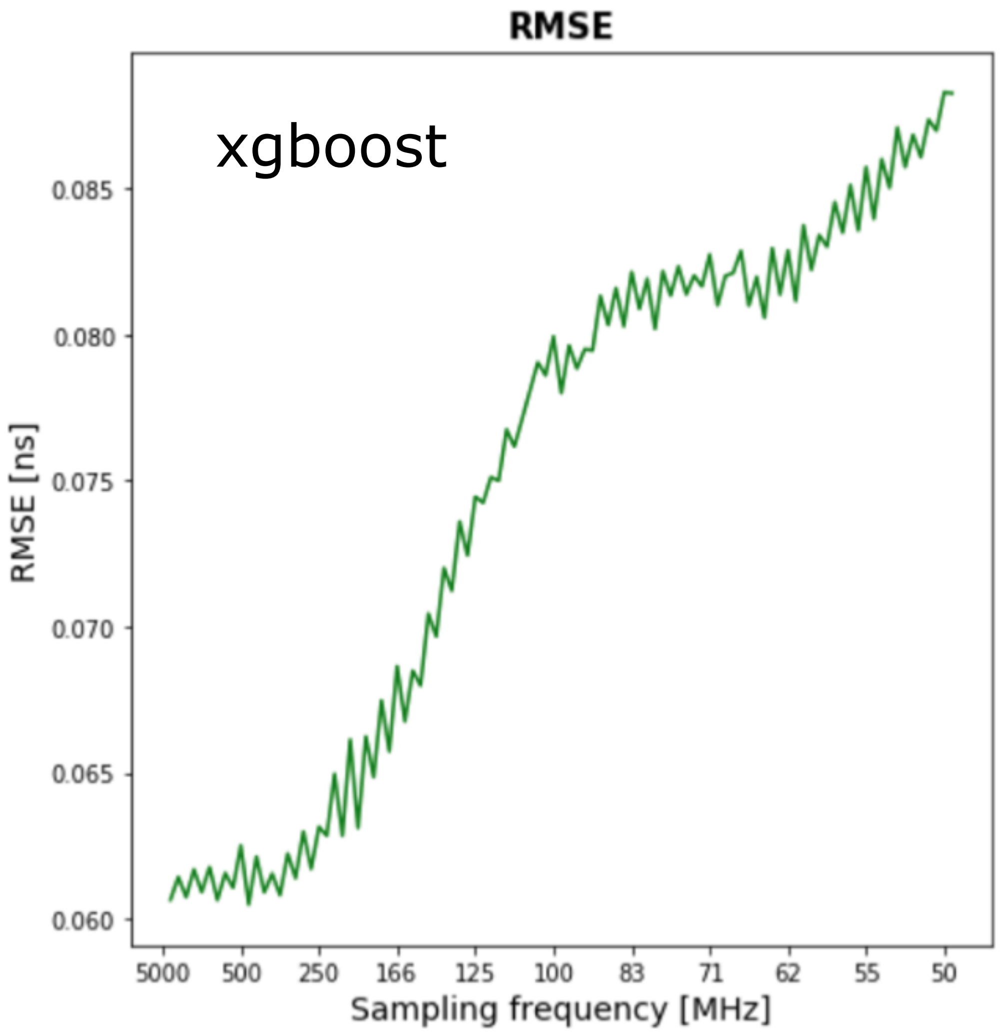
- "output" module of the LHCb electromagnetic calorimeter
- each signal is 1024 impulse measurements

  200 ps sampling (5GHz)

- artificial re-sampling to lower sampling rates

Time Resolution





Different regressors demonstrate similar result

- algorithm-agnostic estimation of the actual signal timing properties
- use xgboost for following studies

Effect of Background Contribution

At high pileup expect overlap of signals from different vertexes

- timing information may be used to mitigate pileup

  for individual calorimeter cell

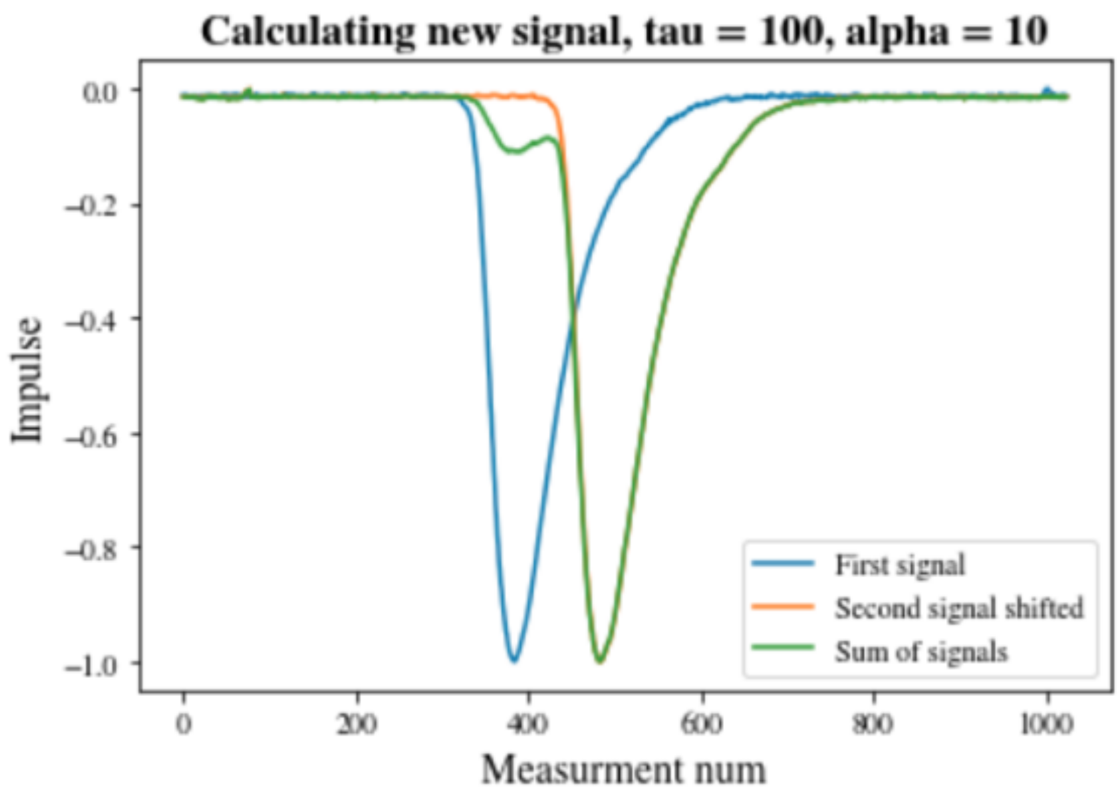
  - as a part of signal processing at readout

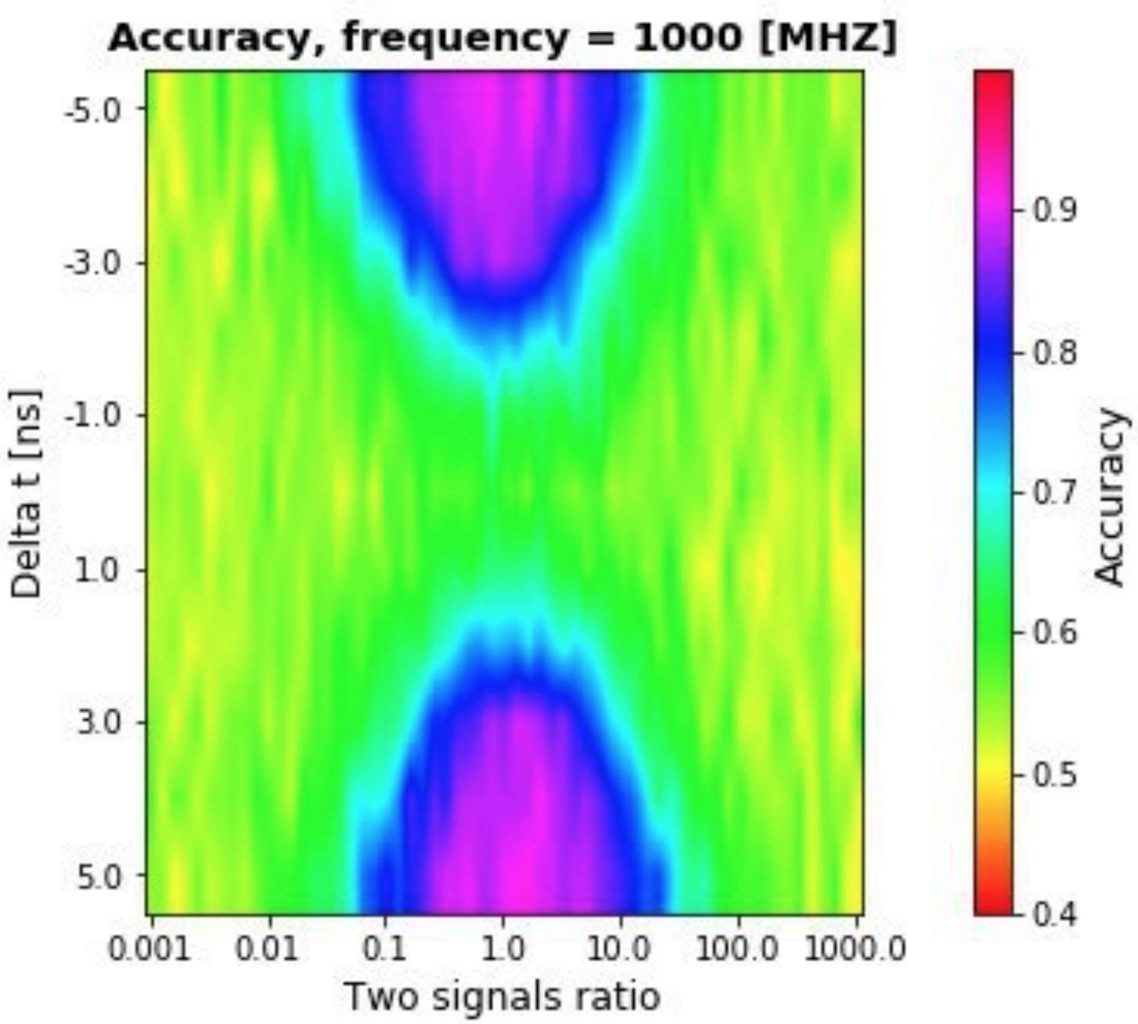
  for several cells of energy clusters

  - at RECO level

- Consider individual cells in the following

Single vs Double Signal Discrimination



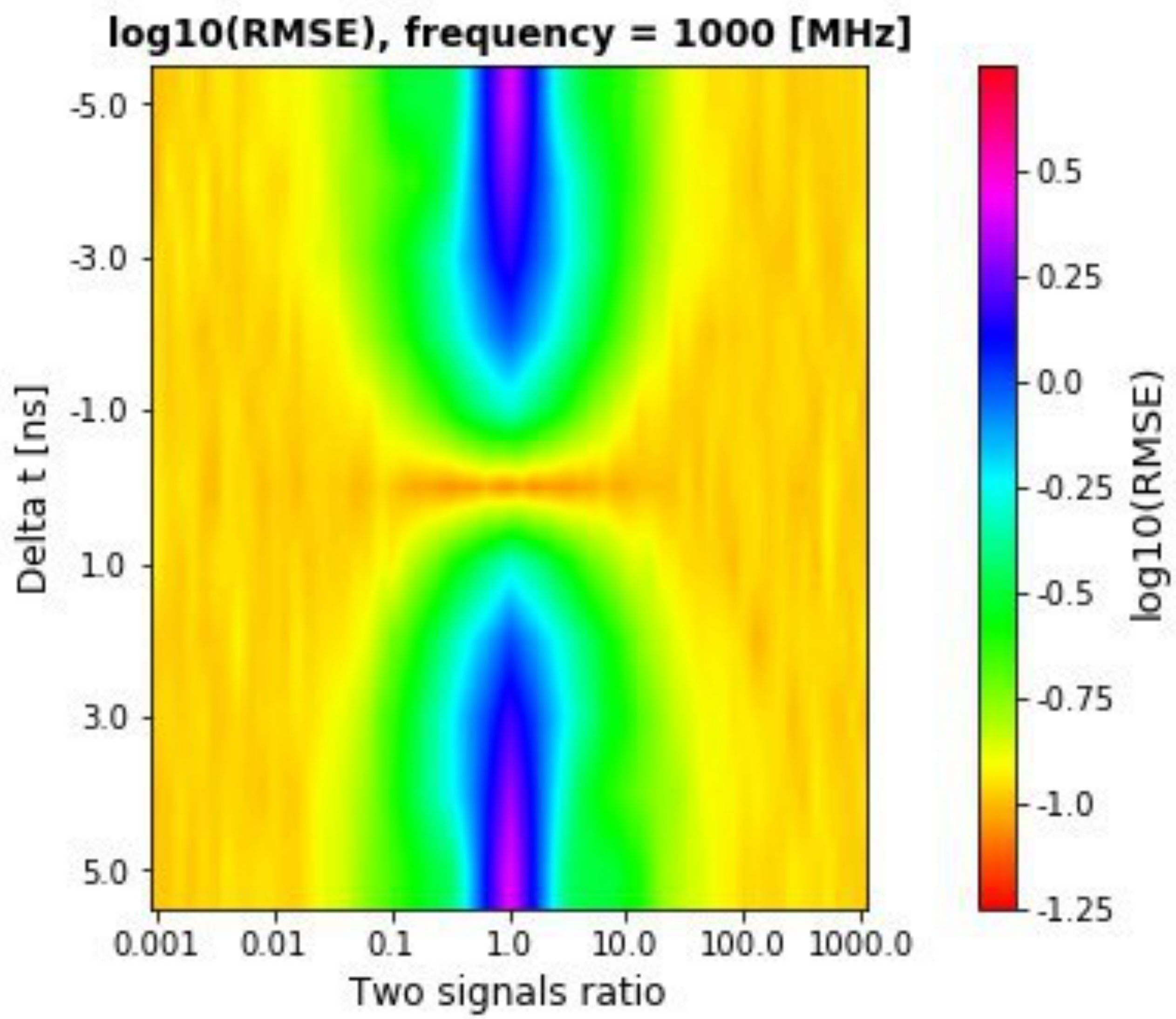


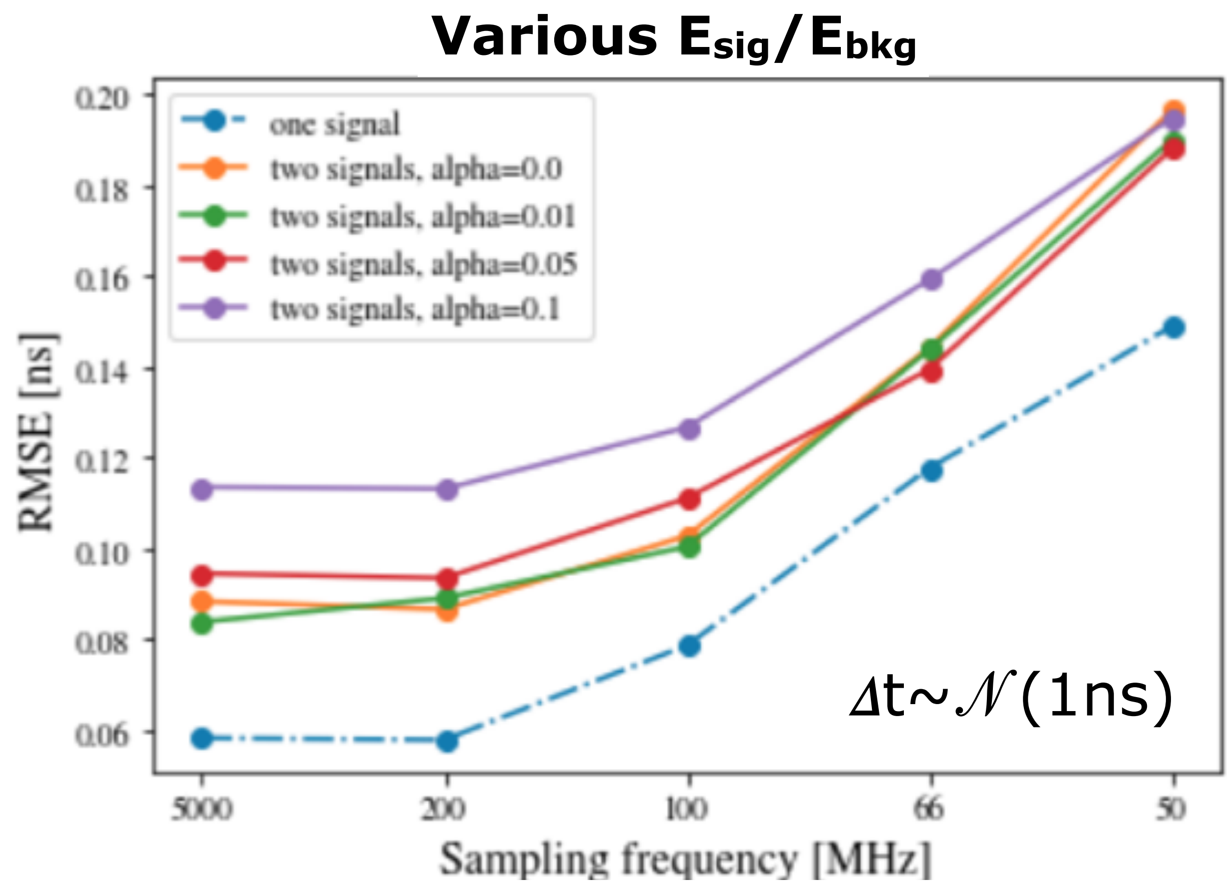
Combine two signals

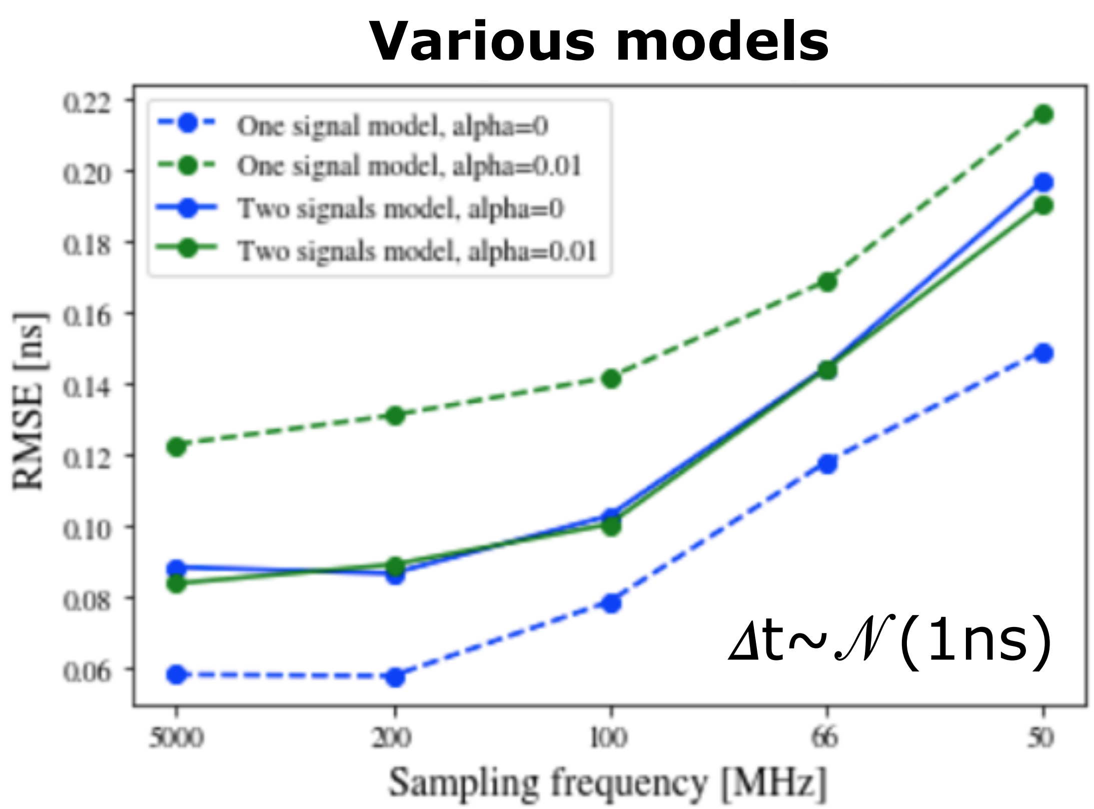
- at given amplitude ratio
- at given time shift

Sensitive to time shifts greater than 1 ns

Effect on Time Resolution







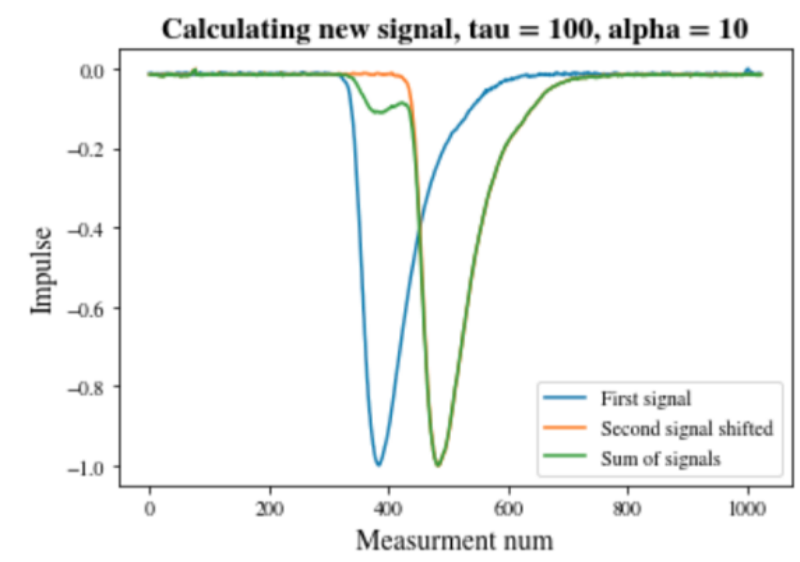
Well defined region of time resolution degradation

- appropriate mixture of 1-signal and 2-signal models should be used for model training

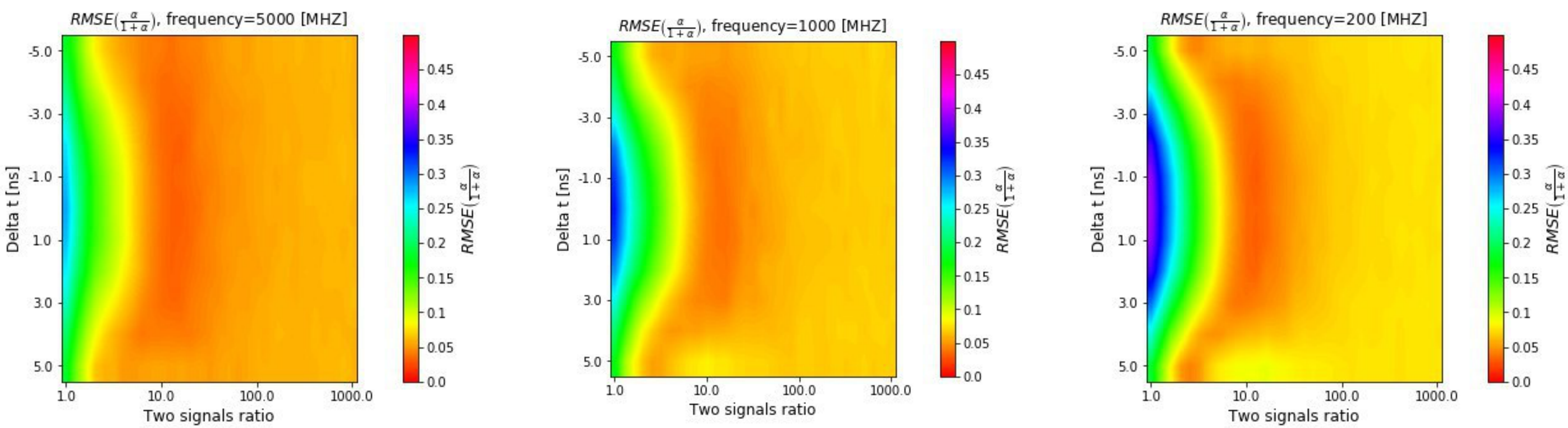
Effect on Amplitude Resolution

How well could we extract signal amplitude on top of the background contribution?

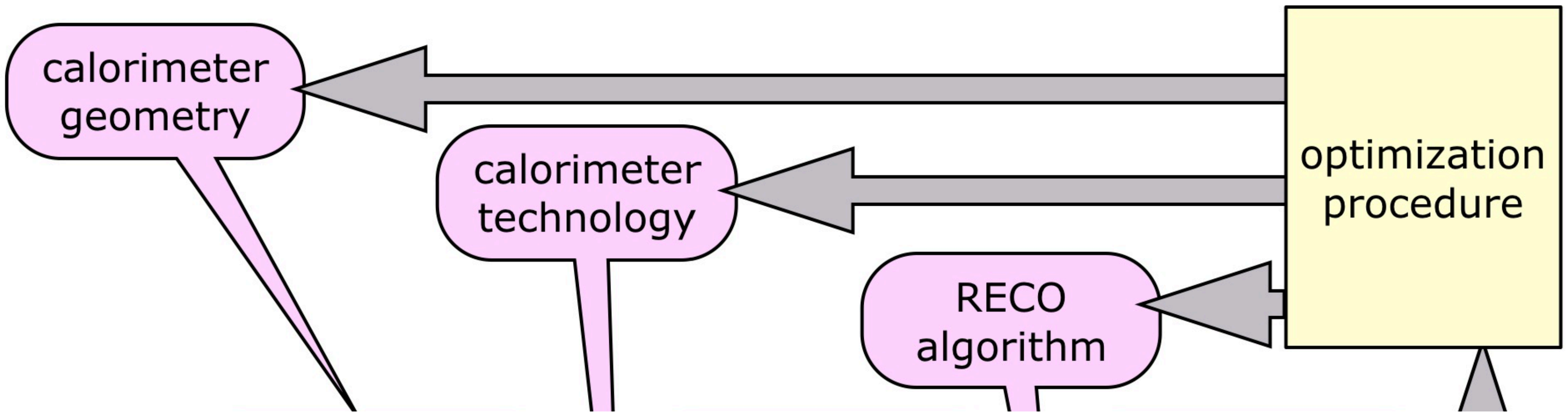
- sampling rate makes difference

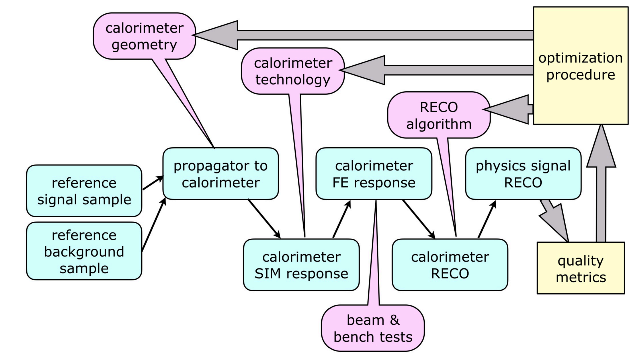


$$E_{sig} = E \times \frac{\alpha}{(1+\alpha)}$$



Global Optimization





Many parameters to optimize simultaneously

- e.g. granularity distribution in LHCb U2 ECAL

Trade between physics performance and costs

- not obvious measure of success
- non-differentiable optimization loss function

Relatively long single iteration loop

ML provides special methods developed for such use cases

- e.g. Bayesian optimization

Summary

Building a bridge between calorimeter detector R&D and its affect on the ultimate physics performance is a time consuming work

Using standard ML approaches for classifications, regressions, domain adaptations, generations, optimization can automatize different steps of the optimization cycle

- this allows quick evaluation of physics performance for the particular calorimeter technology and configuration
  - speed up and steer R&D by quick feedback
  - facilitate global detector optimization