

AI-assisted design of experiments at the frontiers of computation

methods and new perspectives

ALPACA Workshop, ECT^{*}, Trento

Dr. Pietro Vischia

pietro.vischia@cern.ch

[@pietrovischia](https://twitter.com/pietrovischia)



**Supported by project
RYC2021- 033305-I
funded by**

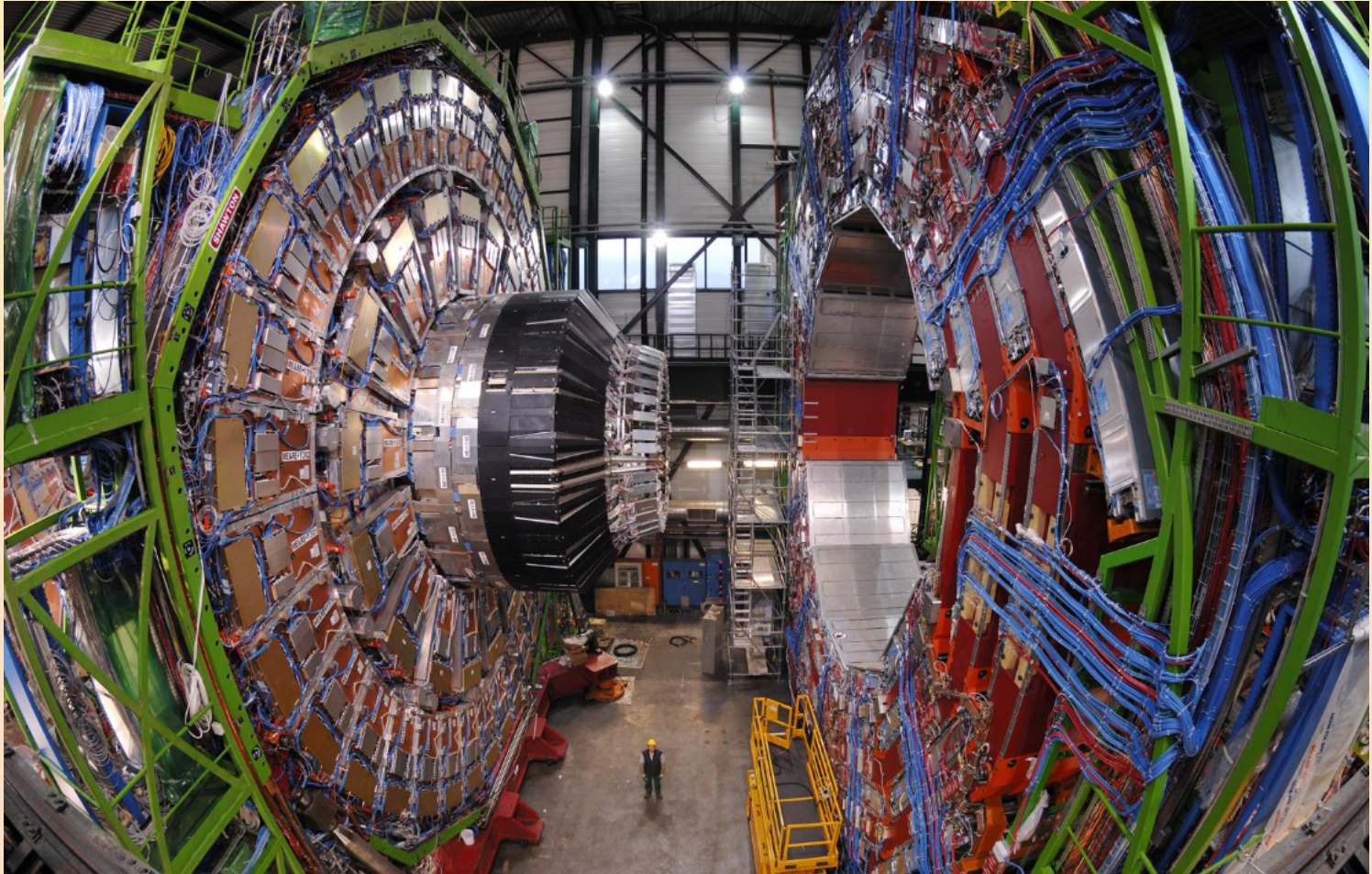


If you are reading this as a web page: have fun! If you are reading this as a PDF:
please visit

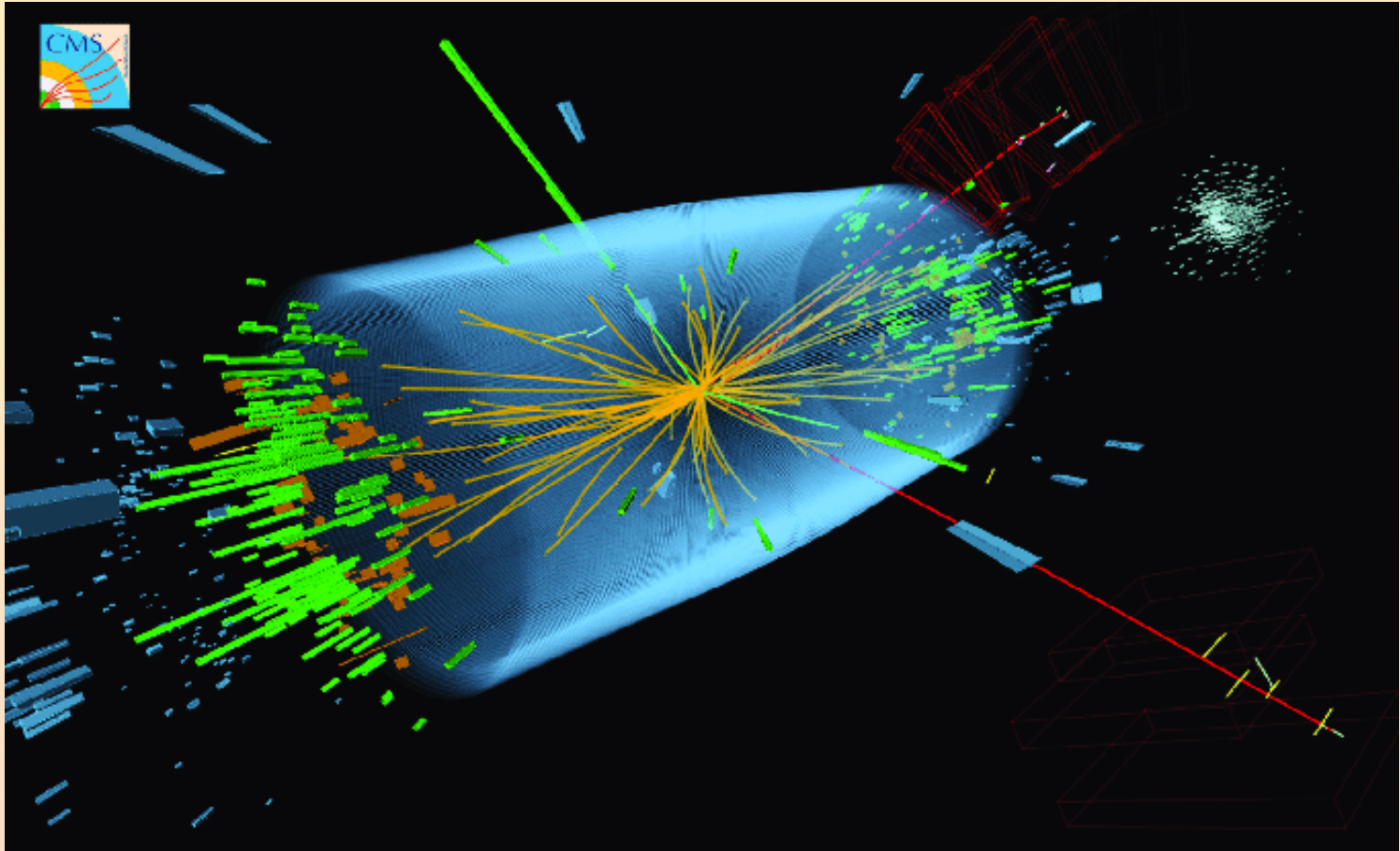
https://www.hep.uniovi.es/vischia/persistent/2023-11-22_AIAssistedDesignAtFrontiersOfComputationAtALPACAWorkshop_vischia.html

to get the version with working animations

Complex Experiments



Complex Data



Likelihood and information

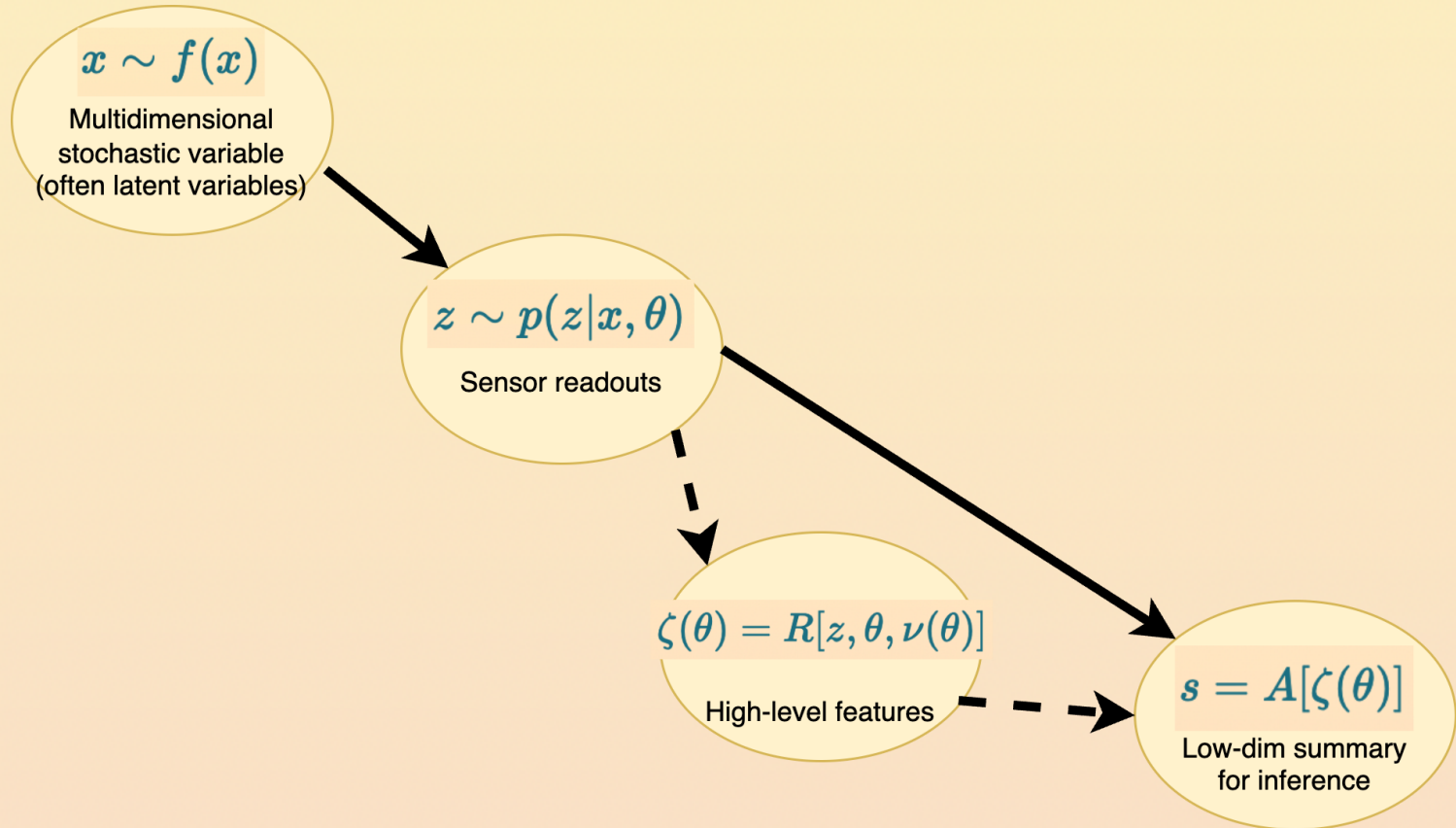
- Data sample X_{obs}

$$\mathcal{L}(X; \theta) := P(X | \theta) |_{X_{obs}}$$

- **The Likelihood Principle:** The likelihood function $L(\vec{x}; \theta)$ contains **all the information** available in the data sample **relevant for the estimation of θ**
 - ✓ Bayesian statistics
 - ✗ Frequentist statistics

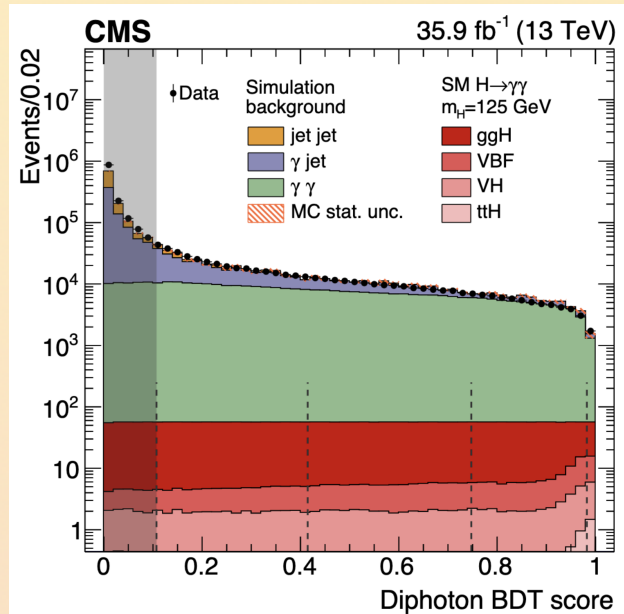
$$I(\theta) = -E \left[\left(\frac{\partial^2}{\partial \theta^2} \ln L(X; \theta) \right)^2 | \theta_{true} \right]$$

Typical analysis pipeline

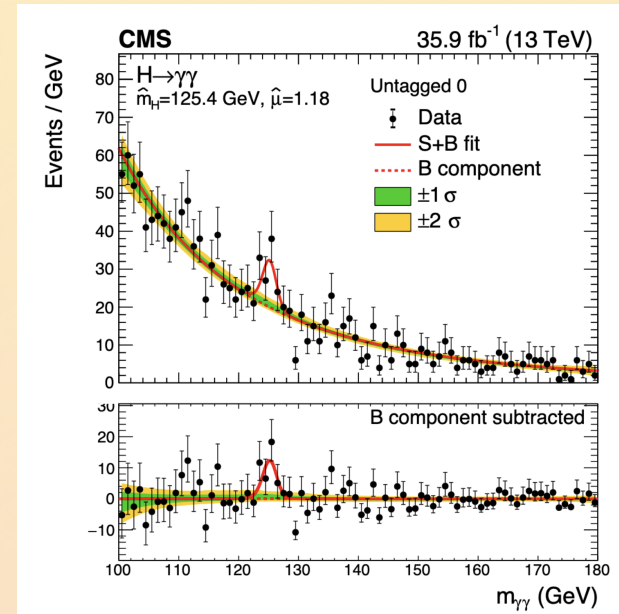


We like low-dim summaries

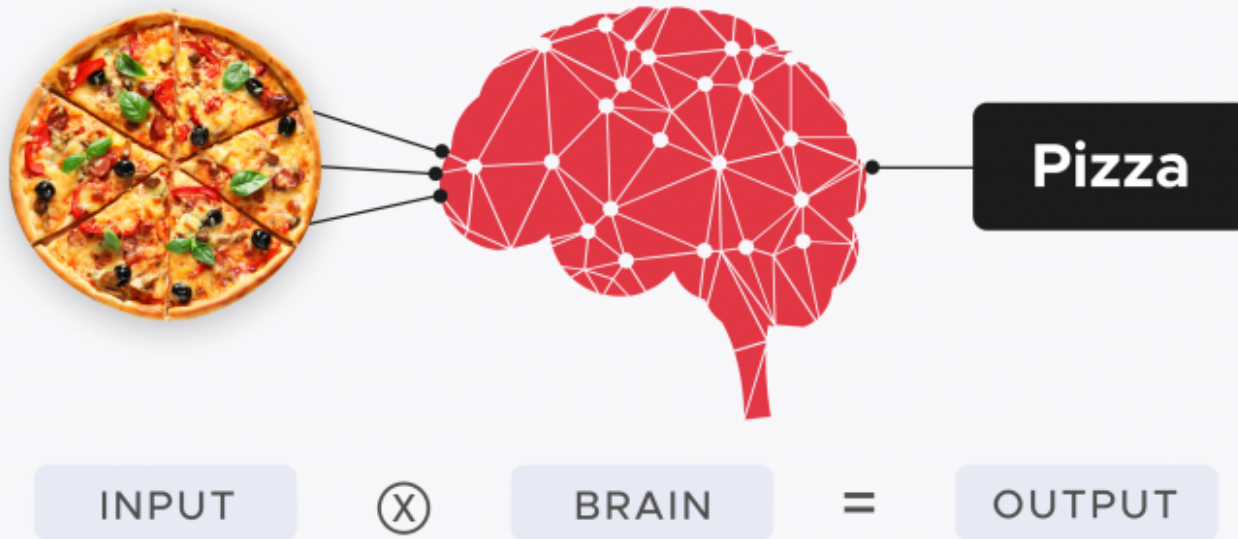
- Discard uninteresting regions



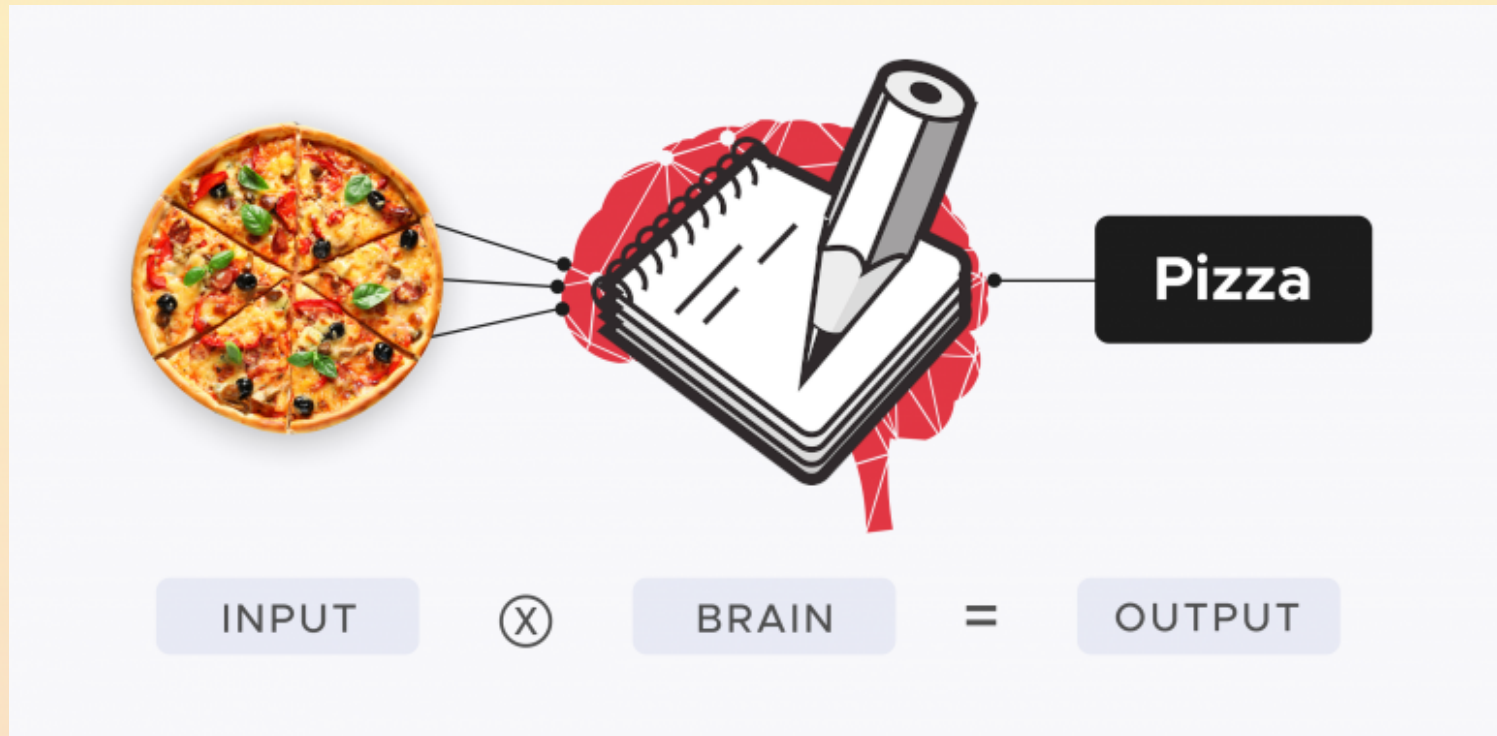
- Physical observable for inference



Brain activity...



...approximated...



...using computers



INPUT

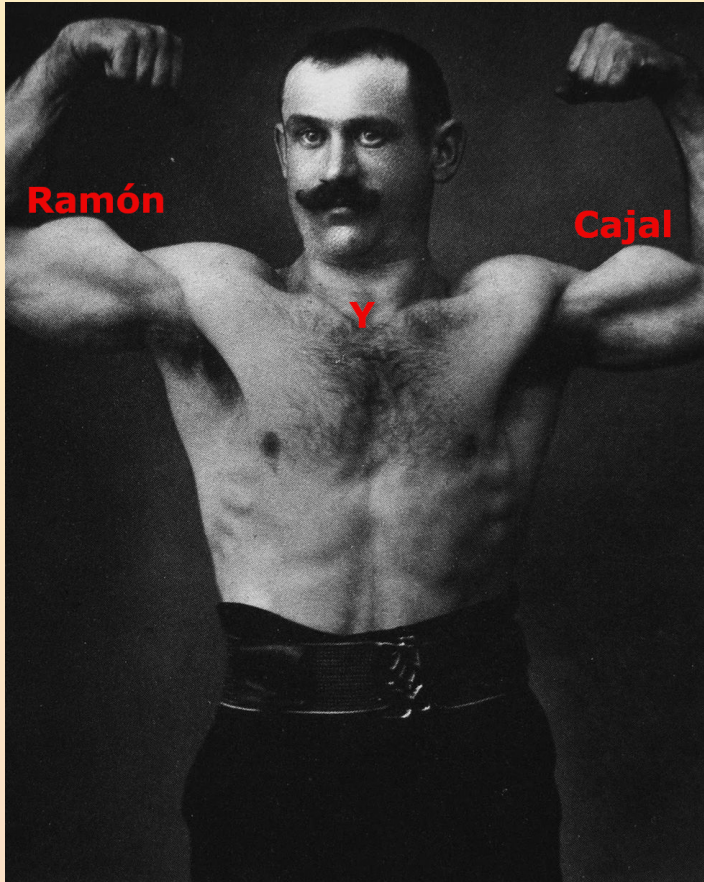


BRAIN

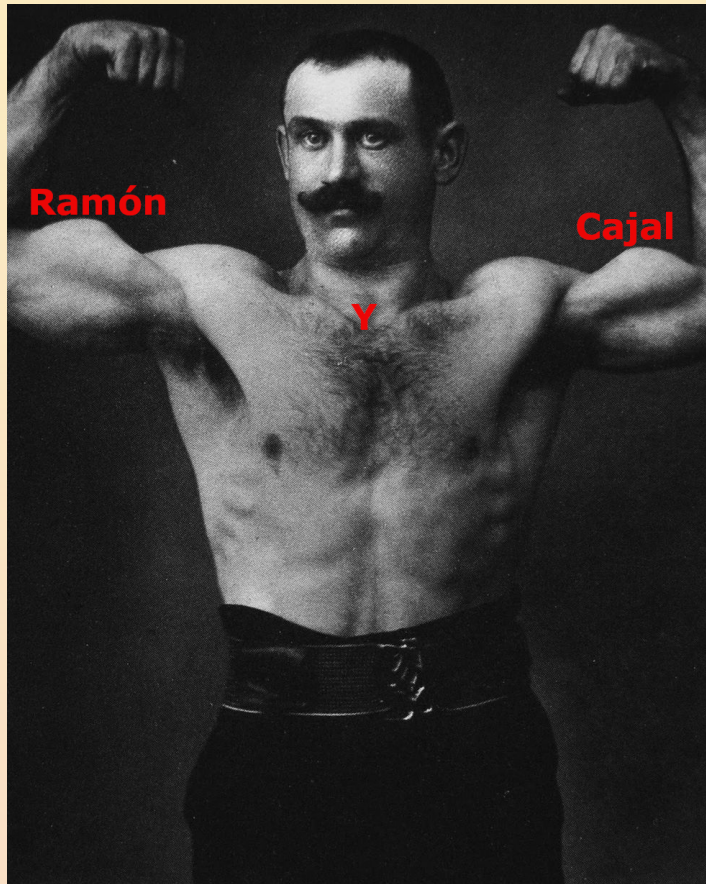
=

OUTPUT

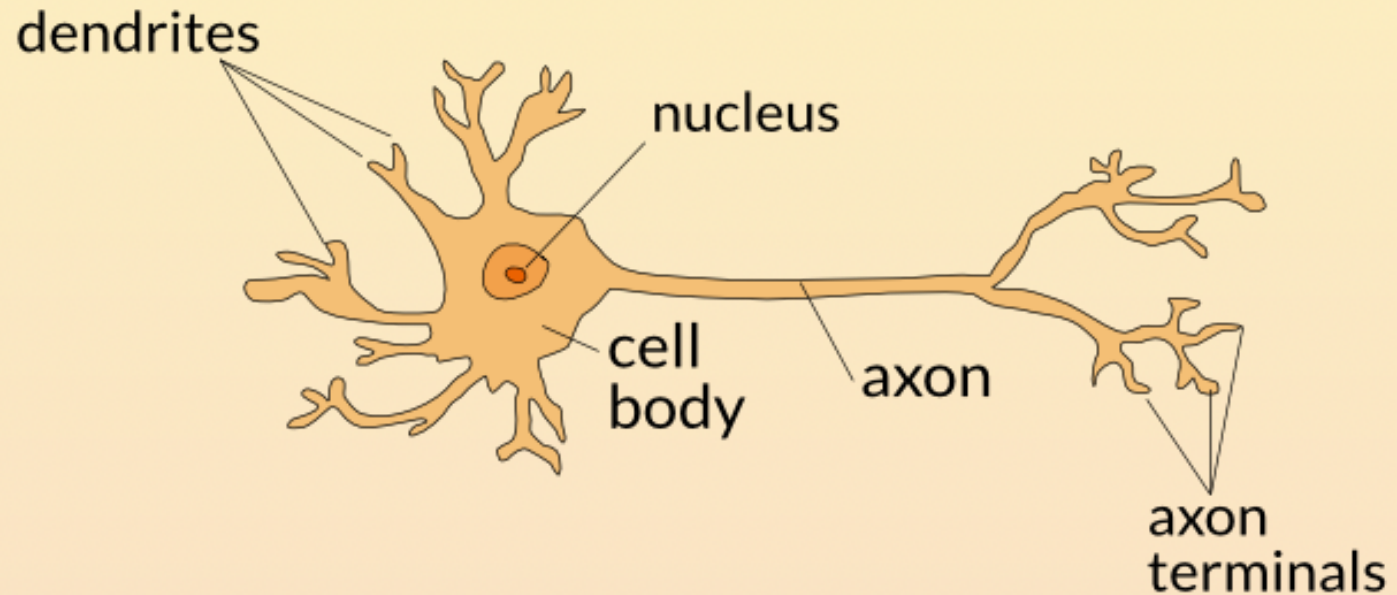
Santiago Ramón y Cajal



Santiago Ramón y Cajal



Real neurons



Computationally heavy



Simplified Neurons

Bulletin of Mathematical Biology Vol. 52, No. 1/2, pp. 99–115, 1990.
Printed in Great Britain.

0092-8240/90\$3.00 + 0.00
Pergamon Press plc
Society for Mathematical Biology

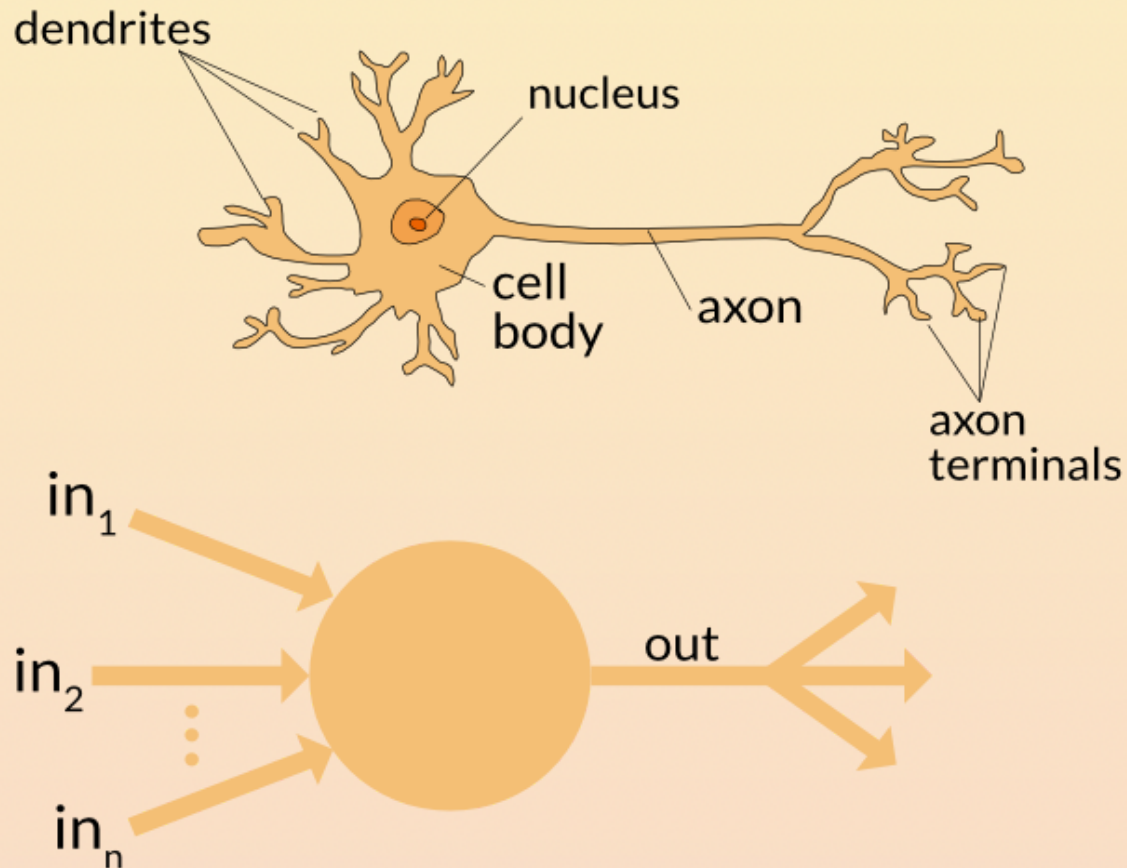
A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY*

■ WARREN S. MCCULLOCH AND WALTER PITTS
University of Illinois, College of Medicine,
Department of Psychiatry at the Illinois Neuropsychiatric Institute,
University of Chicago, Chicago, U.S.A.

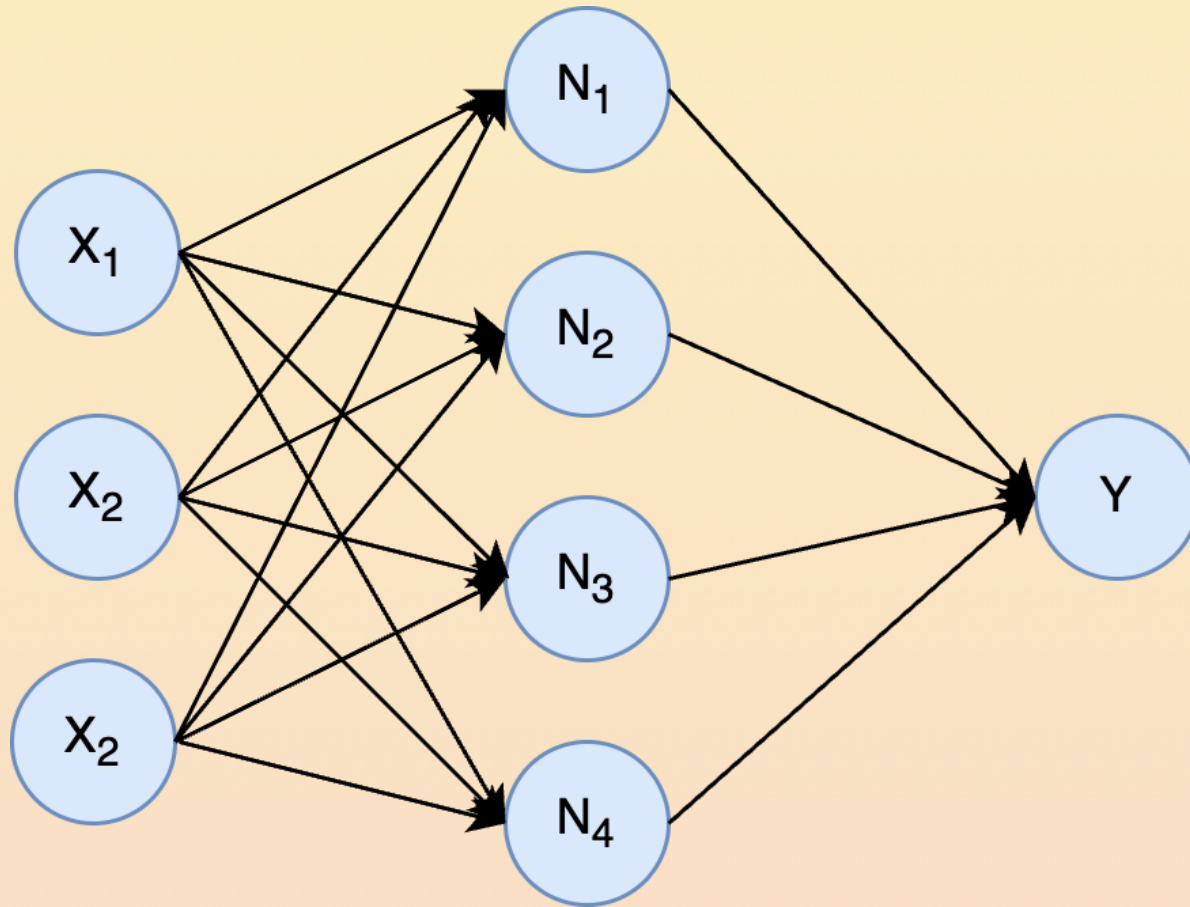
Because of the “all-or-none” character of nervous activity, neural events and the relations among them can be treated by means of propositional logic. It is found that the behavior of every net can be described in these terms, with the addition of more complicated logical means for nets containing circles; and that for any logical expression satisfying certain conditions, one can find a net behaving in the fashion it describes. It is shown that many particular choices among possible neurophysiological assumptions are equivalent, in the sense that for every net behaving under one assumption, there exists another net which behaves under the other and gives the same results, although perhaps not in the same time. Various applications of the calculus are discussed.

Perceptrons

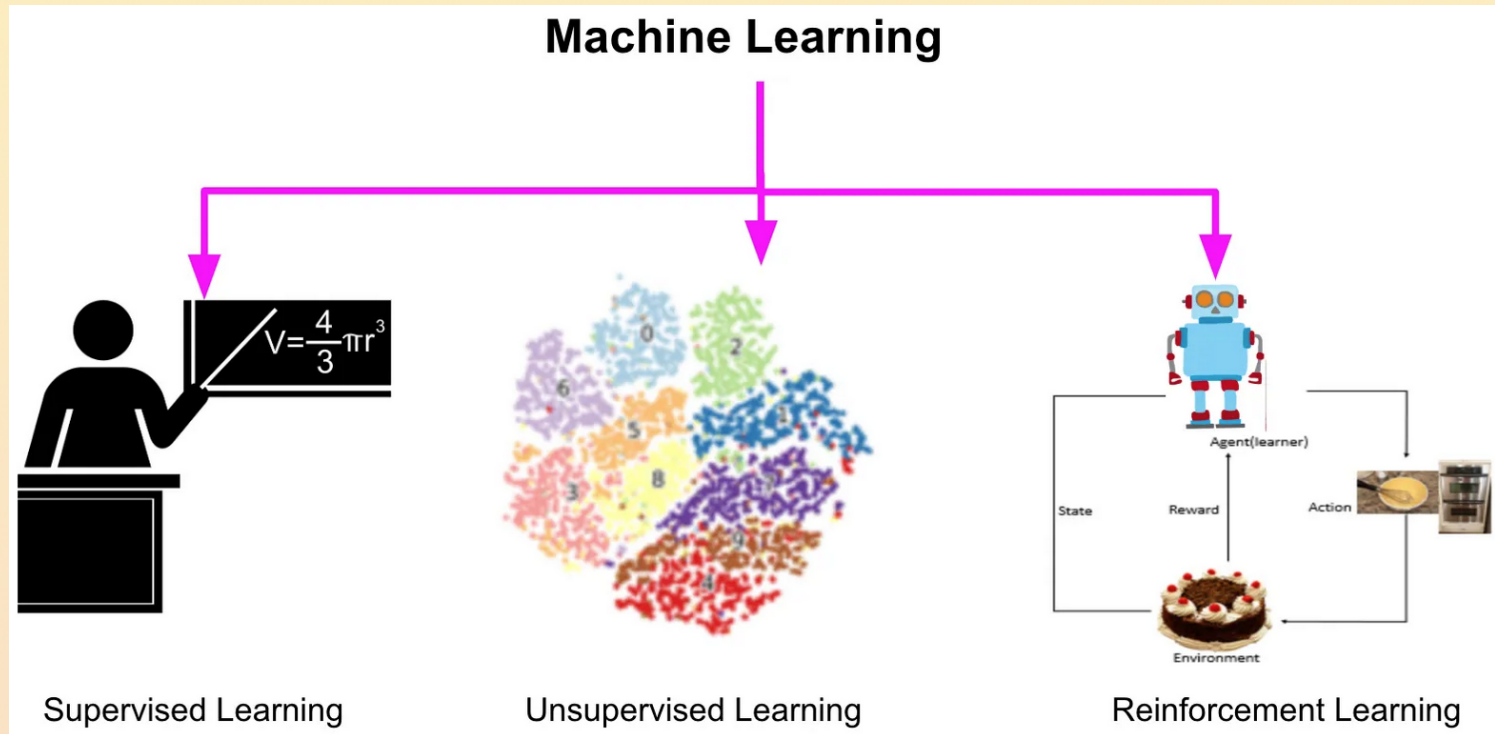
$$y = f\left(b_i + \sum w_i x_i\right)$$



Artificial Neural Networks



Learn in different ways



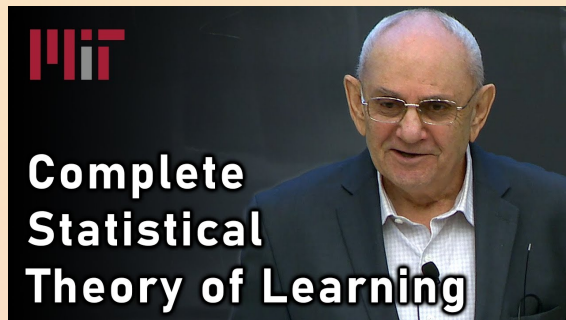
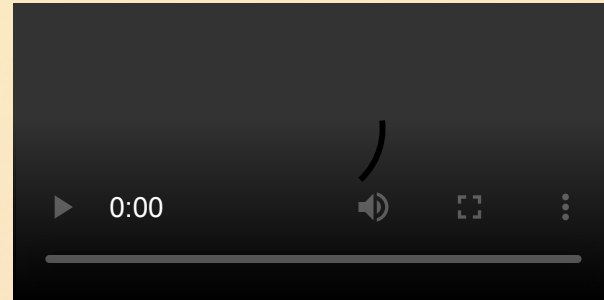
Gradient Descent

- Optimize/learn by finding the minimum of a function $\mathcal{L} : \mathbb{R}^n \rightarrow \mathbb{R}$
- Nonconvex problems: saddle points, manifolds of minima

- Empirical risk minimization

$$\hat{L}(f) = \frac{1}{n} \sum_{i=1}^n |f(x_i) - f^*(x_i)|^2$$

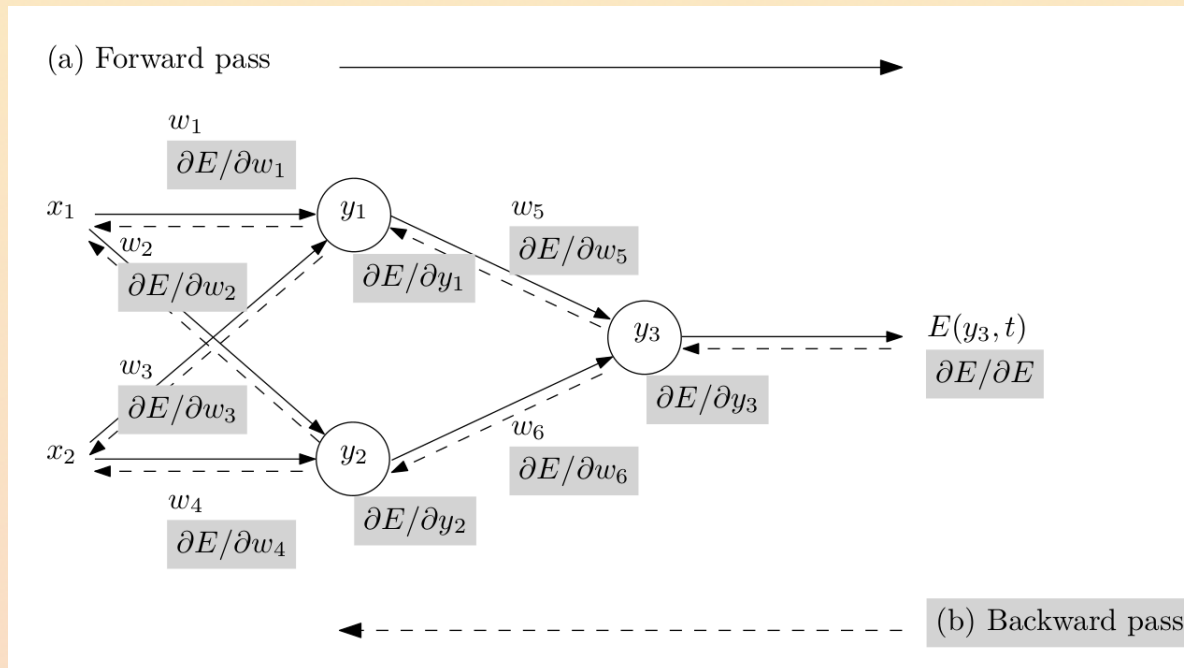
- Generalization
(for learning problems)



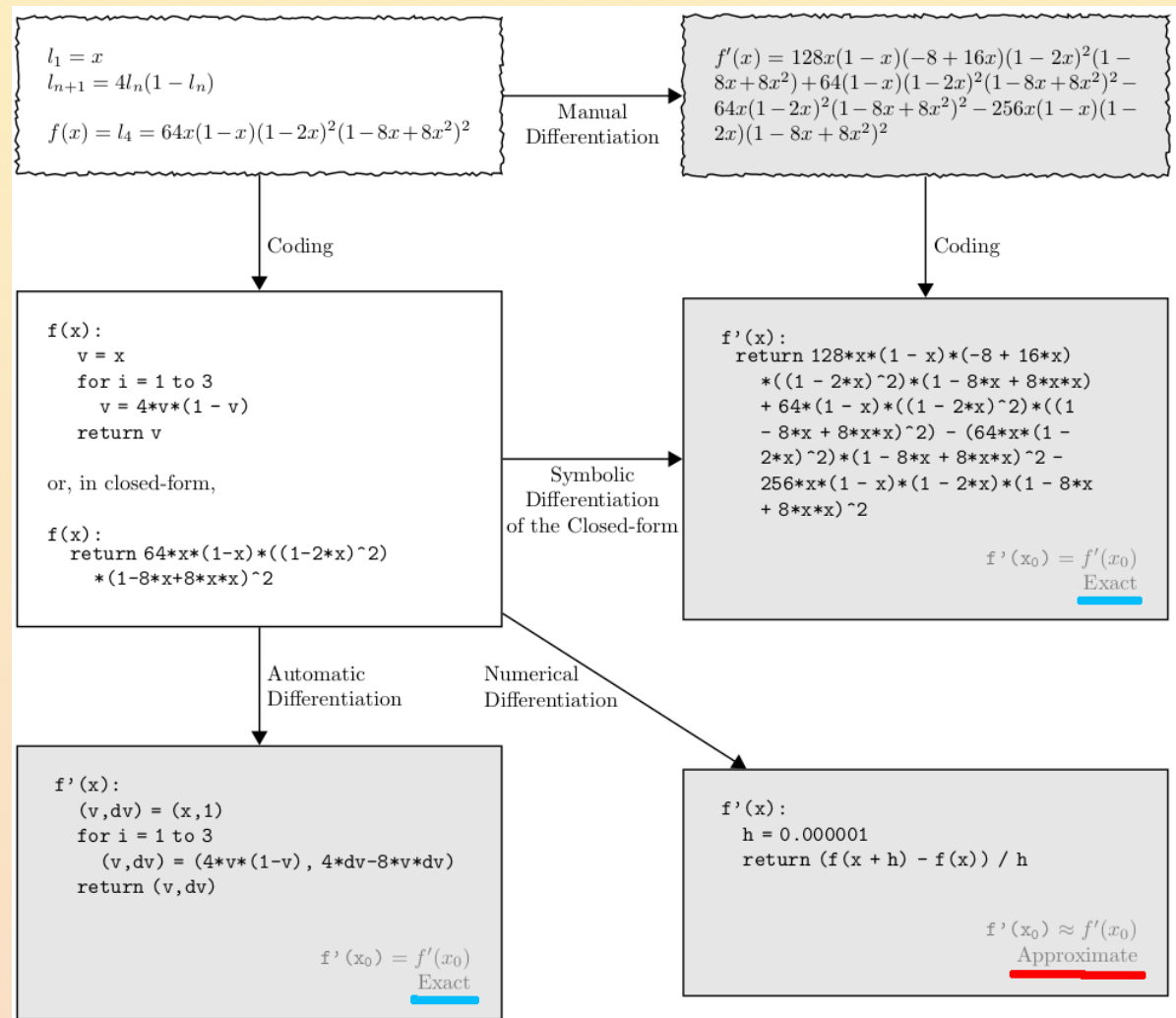
Backpropagation

$$\mathbf{J}(\mathbf{W}) = \frac{1}{n} \sum_{i=1}^n \mathcal{L}(f(x^{(i)}; \mathbf{W}), y^{*(i)}), \quad \mathbf{W}^0 = \operatorname{argmin}_{\mathbf{W}} \mathbf{J}(\mathbf{W})$$

$$\mathbf{W} \leftarrow \mathbf{W} + \eta \frac{\partial \mathbf{J}(\mathbf{W})}{\partial \mathbf{W}}$$



Derive



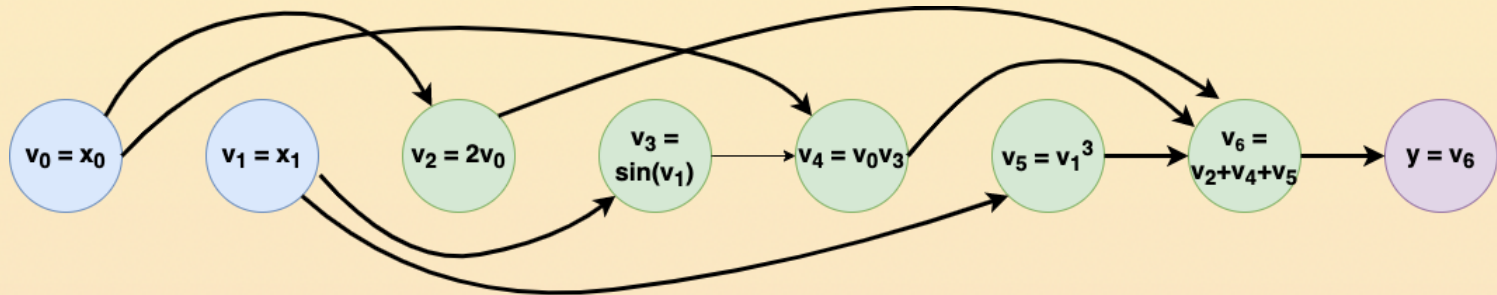
Automatic differentiation

has many names

- Automatic differentiation
- Algorithmic differentiation
- AD
- Autodiff
- Algodiff
- Autograd

Automatic differentiation

$$z(x, y) = 2x + x \sin(y) + y^3$$



Forward mode

- To the extreme, $f : \mathbb{R} \rightarrow \mathbb{R}^m$
- Evaluates $(\frac{\partial f_1}{\partial x}, \dots, \frac{\partial f_m}{\partial x})$
- Computational cost of calculating $\mathbf{J}_f(\mathbf{x})$ for $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ in $\mathbb{R}^n \times \mathbb{R}^m$

$\mathcal{O}(n \text{ time}(f))$

Reverse mode

- To the extreme, $f : \mathbb{R}^n \rightarrow \mathbb{R}$
- Evaluate $\nabla f(\mathbf{x})(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n})$

$\mathcal{O}(m \text{ time}(f))$

Forward and reverse (==backprop) modes

Primal: independent to dependent

Adjoint (derivatives): dependent to independent

$$y(\mathbf{x}) = 2x_0 + x_0 \sin(x_1) + x_1^3$$

Fwd Primal Trace Atomic operation	Value in (1, 2)	Fwd Tangent Trace (set $\dot{x}_0 =$ 1 to compute $\frac{\partial y}{\partial x_0}$) Atomic operation	Value in (1, 2)
$v_0 = x_0$ $v_1 = x_1$	1 2	$\dot{v}_0 = \dot{x}_0$ $\dot{v}_1 = \dot{x}_1$	1 0
$v_2 = 2v_0$ $v_3 =$ $\sin(v_1)$ $v_4 =$ $v_0 v_3$ $v_5 = v_1^3$ $v_6 =$ $v_2 +$ $v_4 + v_5$	2 0.9093 0.9093 8 10.9093	$\dot{v}_2 = 2\dot{v}_0$ $\dot{v}_3 = \dot{v}_1 \cos(v_1)$ $\dot{v}_4 = \dot{v}_0 v_3 +$ $v_0 \dot{v}_3$ $\dot{v}_5 = 3\dot{v}_1 v_1^2$ $\dot{v}_6 = \dot{v}_2 + \dot{v}_4 +$ $\dot{v}_5$	2×1 $0 \times$ -0.41 $1 \times$ $0.9093 +$ 1×0 $3 \times 0 \times 4$ $2 +$ $0.9093 +$ 0
$y = v_6$	10.9093	$\dot{y} = \dot{v}_6$	2.9093

Fwd Primal Trace Atomic operation	Value in (1, 2)	Rev Adjoint Trace (set $\bar{y} =$ 1 to compute $\frac{\partial v}{\partial y}$) Atomic operation	Value in (1, 2)
$v_0 = x_0$ $v_1 = x_1$	1 2	$\bar{x}_0 = \bar{v}_0$ $\bar{x}_1 = \bar{v}_1$	2.9093 11.5839
$v_2 = 2v_0$ $v_3 =$ $\sin(v_1)$ $v_4 =$ $v_0 v_3$ $v_5 = v_1^3$ $v_6 =$ $v_2 +$ $v_4 + v_5$	2 0.9093 0.9093 8 10.9093	$\bar{v}_0 = \bar{v}_0 +$ $\bar{v}_2 \partial v_2 / \partial v_0$ $\bar{v}_0 =$ $\bar{v}_4 \partial v_4 / \partial v_0$ $\bar{v}_1 = \bar{v}_1 +$ $\bar{v}_3 \partial v_3 / \partial v_1$ $\bar{v}_1 =$ $\bar{v}_5 \partial v_5 / \partial v_1$ $\bar{v}_2 =$ $\bar{v}_6 \partial v_6 / \partial v_2$ $\bar{v}_3 =$ $\bar{v}_4 \partial v_4 / \partial v_3$ $\bar{v}_4 =$ $\bar{v}_6 \partial v_6 / \partial v_4$ $\bar{v}_5 =$ $\bar{v}_6 \partial v_6 / \partial v_5$	$\bar{v}_0 + \bar{v}_2 \times$ $2 =$ 2.9093 $\bar{v}_4 \times v_3 =$ 0.9093 $\bar{v}_1 + \bar{v}_3 \times$ $\cos(v_1) =$ 11.5839 $\bar{v}_5 \times$ $3v_1^2 = 12$ $\bar{v}_6 \times 1 =$ 1 $\bar{v}_4 \times v_0 =$ 1 $\bar{v}_6 \times 1 =$ 1 $\bar{v}_6 \times 1 =$ 1
$y = v_6$	10.9093	$\bar{v}_6 = \bar{y}$	1

Designed to be simple in software

```
import torch, math
x0 = torch.tensor(1., requires_grad=True)
x1 = torch.tensor(2., requires_grad=True)
p = 2*x0 + x0*torch.sin(x1) + x1**3
print(p)
p.backward()
print(x0.grad, x1.grad)
```

yielding

```
Primal: tensor(10.9093, grad_fn=<AddBackward0>)
Adjoint: tensor(2.9093) tensor(11.5839)
```

- Torch (and similar software) will correctly differentiate only when the atomic operations are supported within it
 - Common operations are overloaded (`__mul__` rewritten by `torch._mult_`)
 - Operations from libraries (`math.sin()`) must be replaced by their differentiation-aware equivalents (`torch.sin()`)

Differentiable Programming

Execute **differentiable functions (programs)** via **automatic differentiation**

**Yann LeCun** ✓
January 5, 2018 · 🌐

OK, Deep Learning has outlived its usefulness as a buzz-phrase.
Deep Learning est mort. Vive Differentiable Programming!

Yeah, Differentiable Programming is little more than a **rebranding** of the modern collection Deep Learning techniques, the same way Deep Learning was a rebranding of the modern incarnations of neural nets with more than two layers.

But the important point is that people are now **building a new kind of software** by assembling networks of parameterized functional blocks and by training them from examples using some form of gradient-based optimization.

An increasingly large number of people are defining the networks procedurally in a data-dependent way (with loops and conditionals), allowing them to change dynamically as a function of the input **data fed to** them. It's really very much like a regular program, except it's parameterized, **automatically differentiated**, and trainable/optimizable. Dynamic networks have **become** increasingly popular (particularly for NLP), thanks to deep learning frameworks that can handle them such as PyTorch and Chainer (note: our old deep learning framework Lush could handle a particular kind of dynamic nets called Graph Transformer Networks, back in 1994. It was needed for text recognition).

People are now actively working on compilers for **imperative differentiable programming languages**. This is a very exciting avenue for the development of learning-based AI.

Important note: this won't be sufficient to take us to "true" AI. Other concepts will be needed for that, such as what I used to call predictive learning and now decided to call Imputative Learning. More on this later...

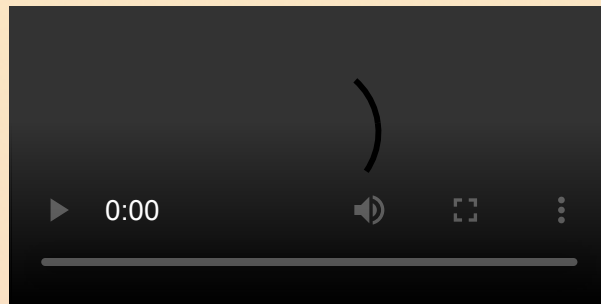
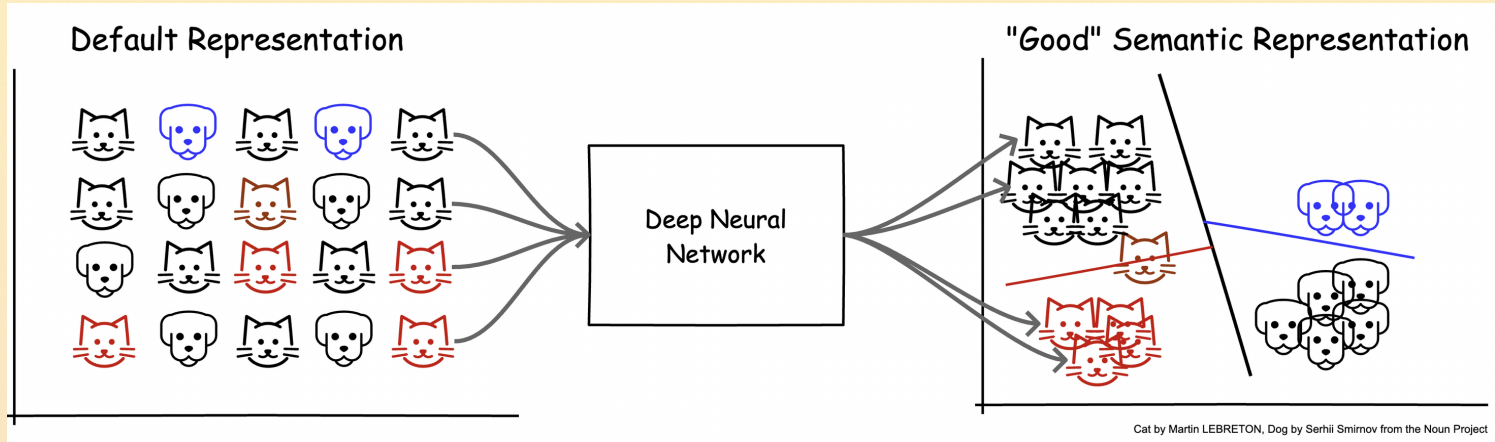
 1.8K

186 Comments 464 Shares

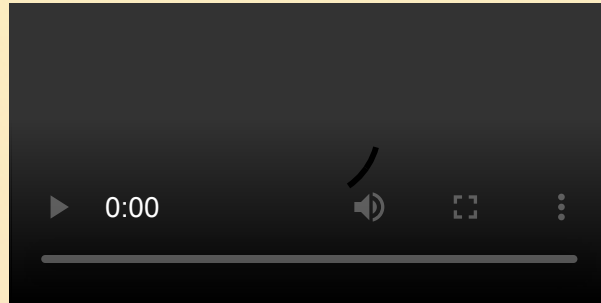
Maps as a Tool of Understanding



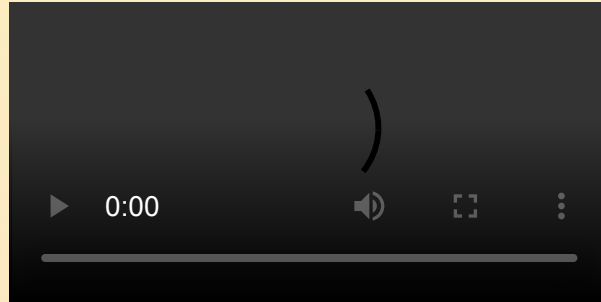
"Representation" simplifies tasks



Impressive results

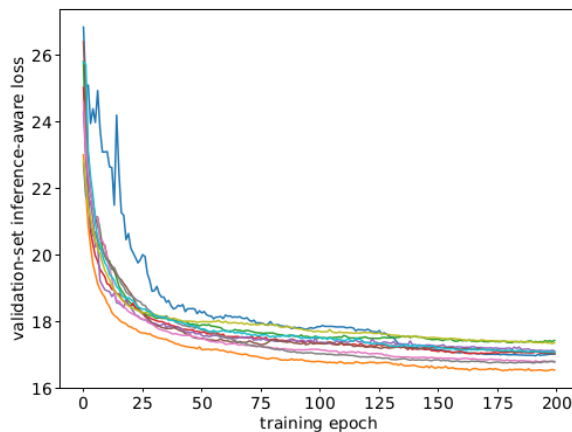
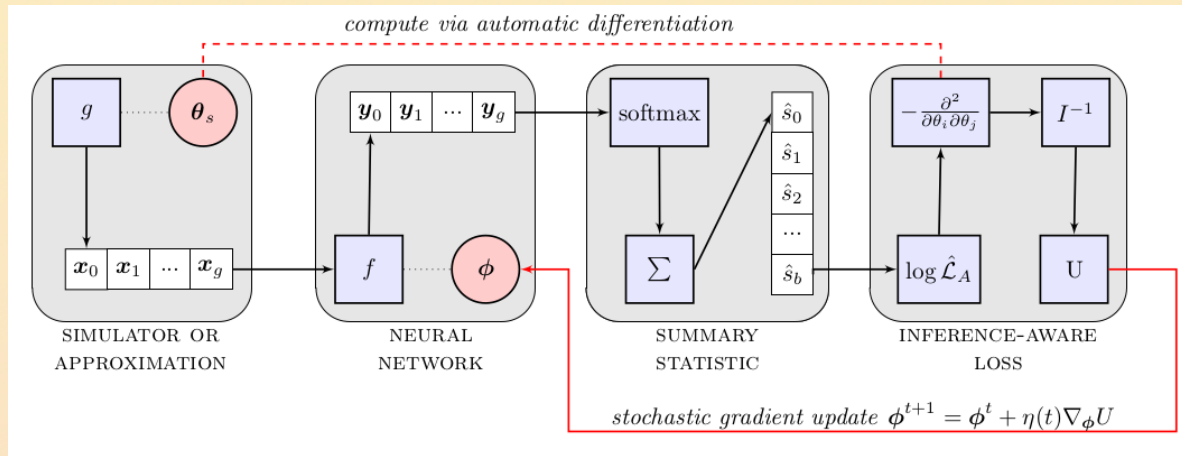


Impressive results

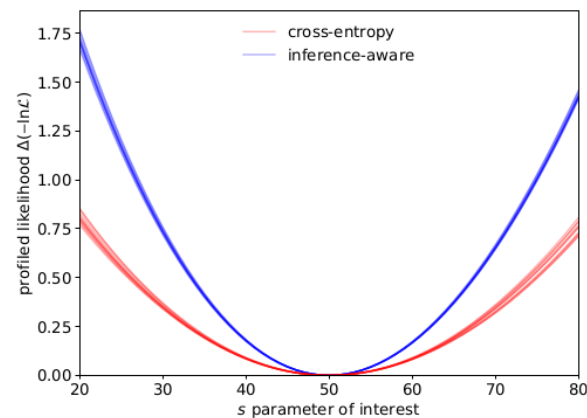


- Busco colaboraciones para aplicaciones médicas de inteligencia artificial

Go to INFERNO: syst-aware inference opt.



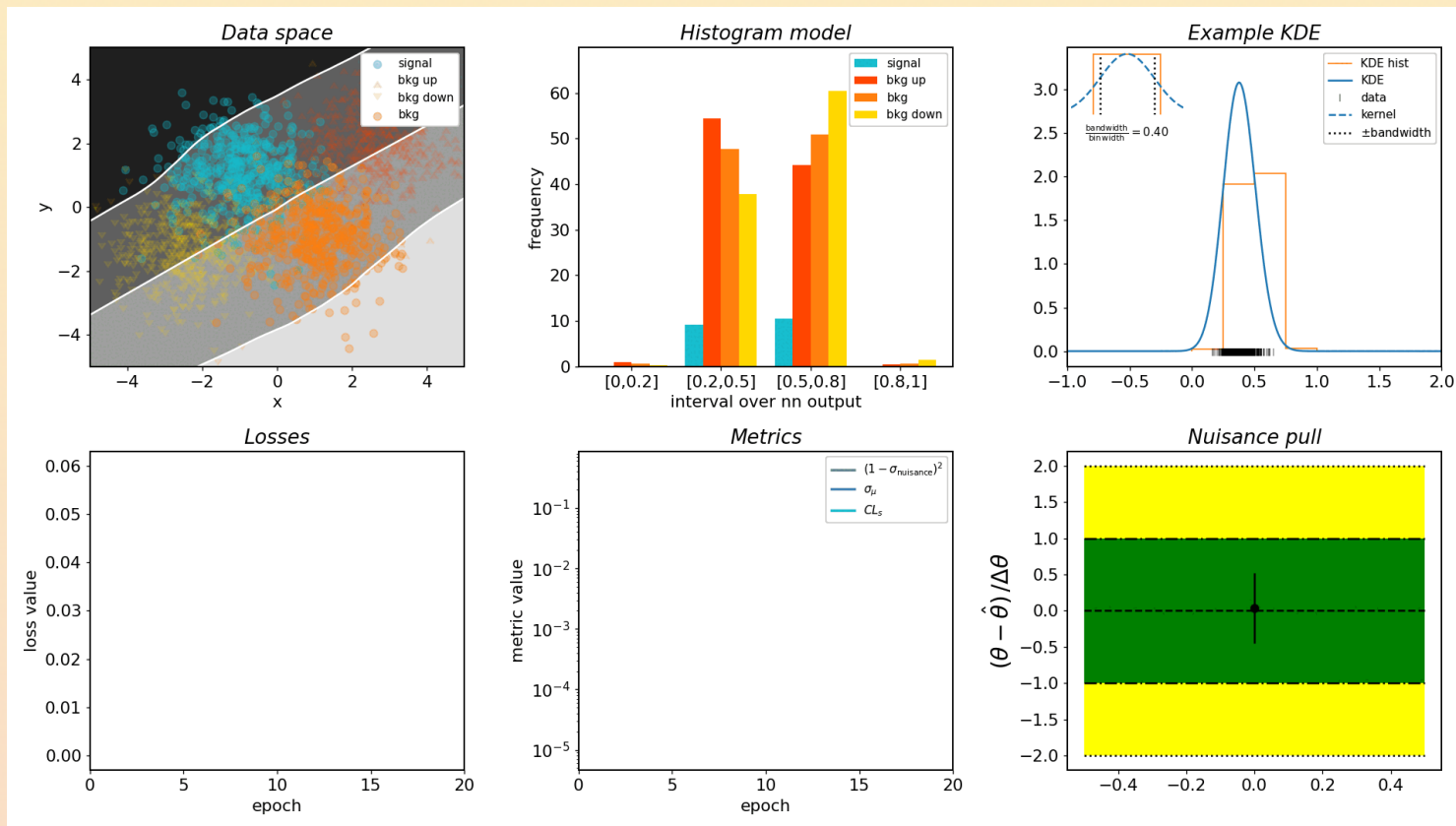
(a) inference-aware training loss



(b) profile-likelihood comparison

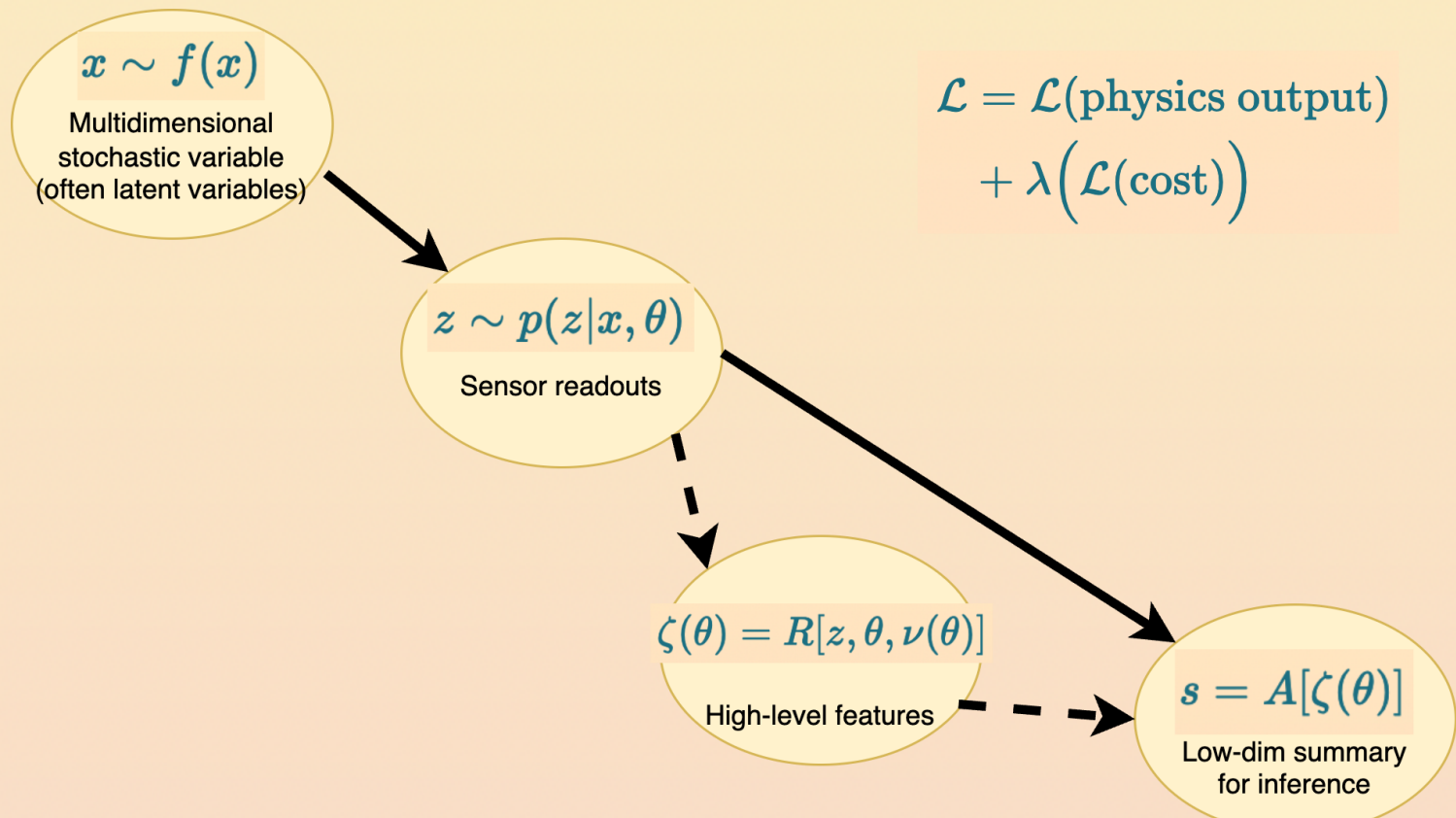
Measurement-aware analysis opt.

neos



Measurement-aware detector opt.!

- Joint optimization of design parameters w.r.t. inference made with data
- MODE White Paper, [10.1016/j.revip.2023.100085](https://arxiv.org/abs/2203.13818) (2203.13818), 117-pages document, physicists + computer scientists (published last month!!!)



Guarantee feasibility within constraints

- Monetary cost
- Case-specific technical constraints

$$\mathcal{L}_{\text{cost}} = c(\theta, \phi)$$

- θ : local, specific to the technology used (e.g. active components material)
- ϕ : global, describing overall detector conception (e.g. number, size, position of detector modules)
- Fixed costs can be added separately to the loss function

In general

Depends on z and nuisances

Cost of the layout with parameters θ

Closed form

$$\hat{\theta} = \arg \min_{\theta} \int L[A(\zeta), c(\theta)] p(z|x, \theta) f(x) dx dz ,$$

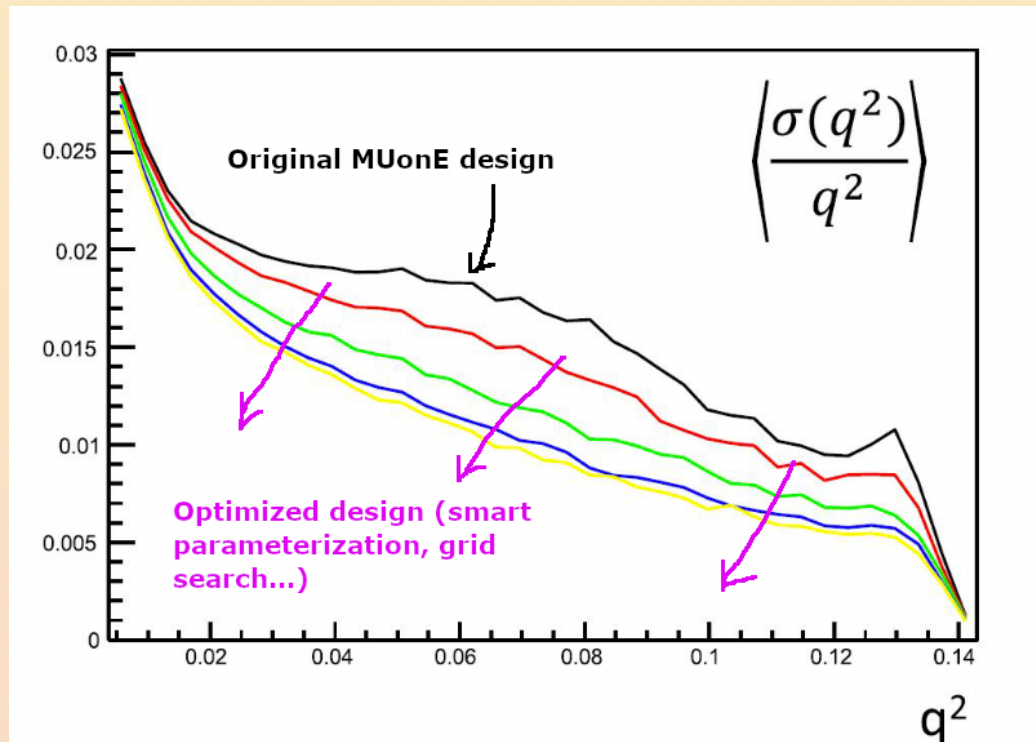
Weight desirable goals while obeying cost constraints

Thrive in asymmetries



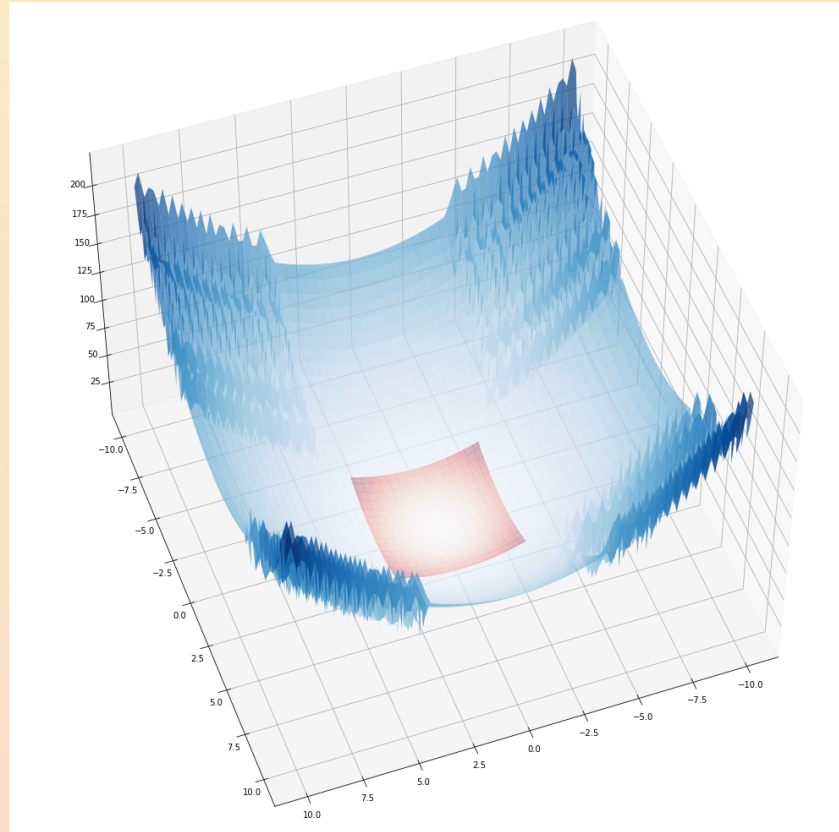
Large gains to be had

- MUonE: proposed 150 GeV muon beam experiment to be built at CERN
 - Measure precisely the q^2 differential cross section in electron-muon scattering
 - 40 tracking stations and a calorimeter
- **Dramatic improvement** in the resolution on q^2 even from a simple grid search



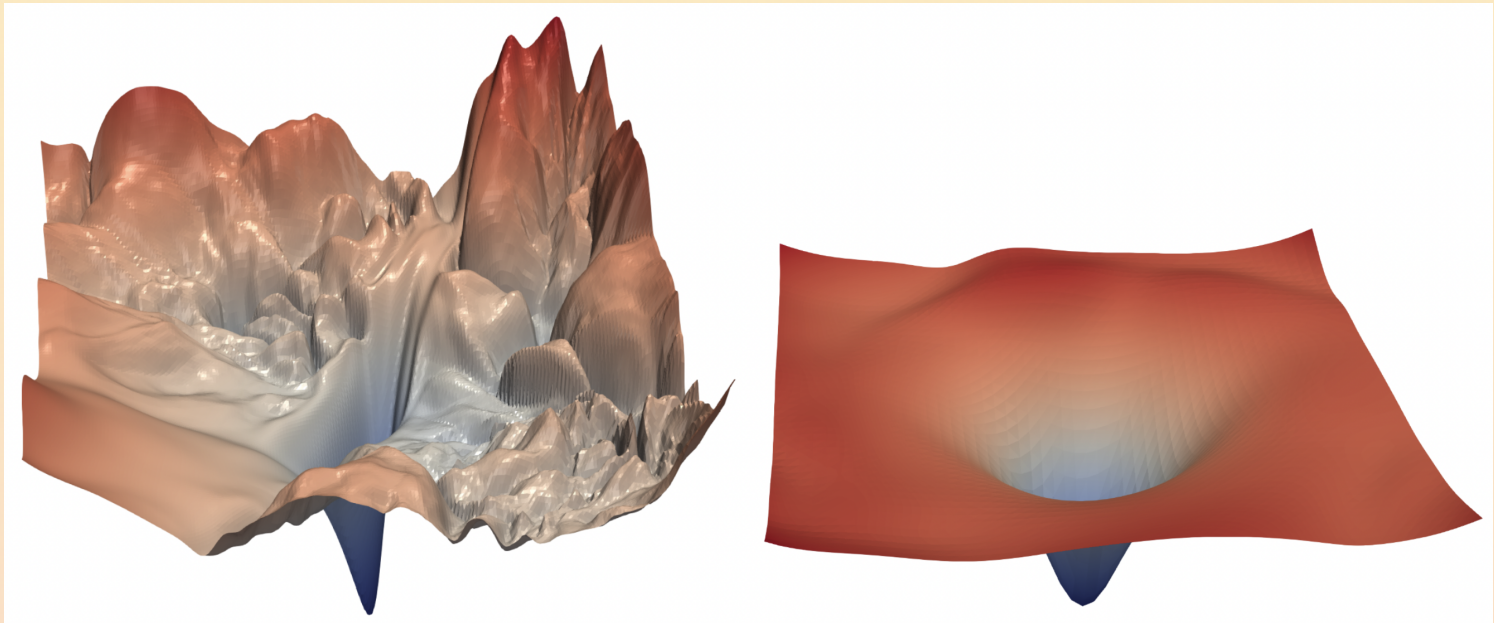
Assist the physicist with a landscape of solutions

- Cannot parameterize everything
- The optimal solution: unrealistic
- Provide feasible solutions near optimality
- The physicist will fine tune

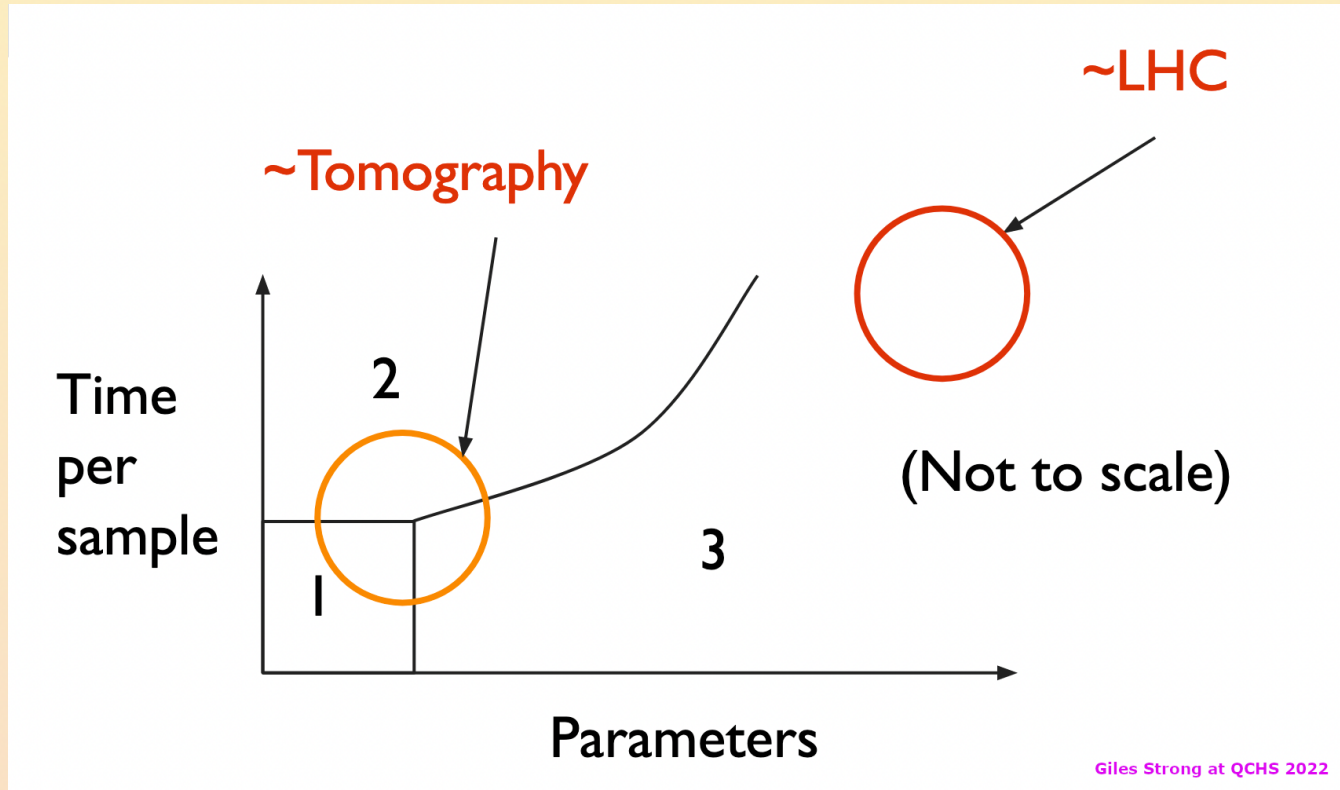


How far from optimality?

- Can we define in a general way an acceptable **increase in loss**?
 - Tradeoff performance/cost
- For sure we can regularize the loss landscape to select our scale of interest



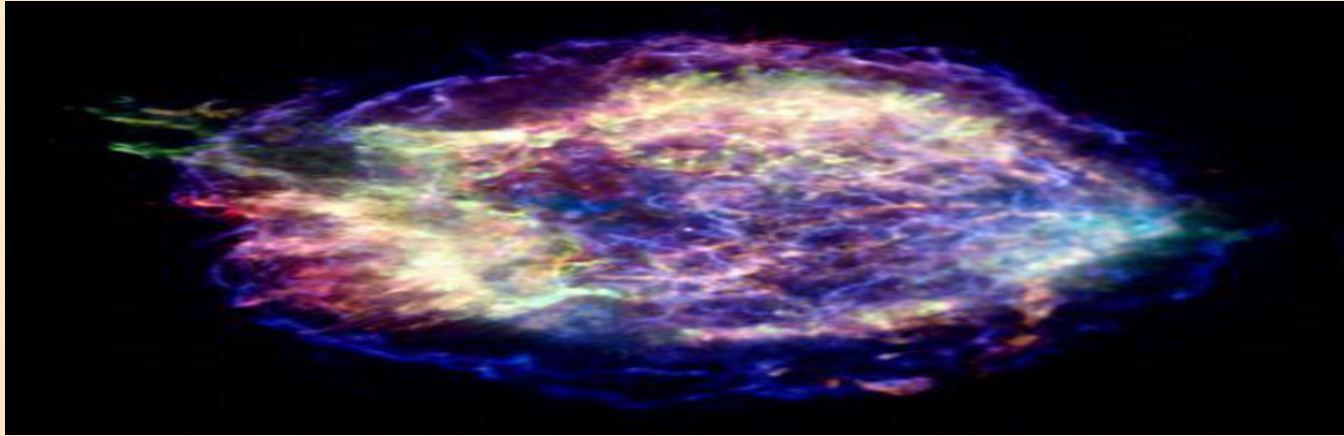
Method of choice depends on scale



1. Grid/random search
2. Bayesian opt, simulated annealing, genetic algos, ...
3. Gradient-based optimization (Newton, BFGS, gradient descent, ...)

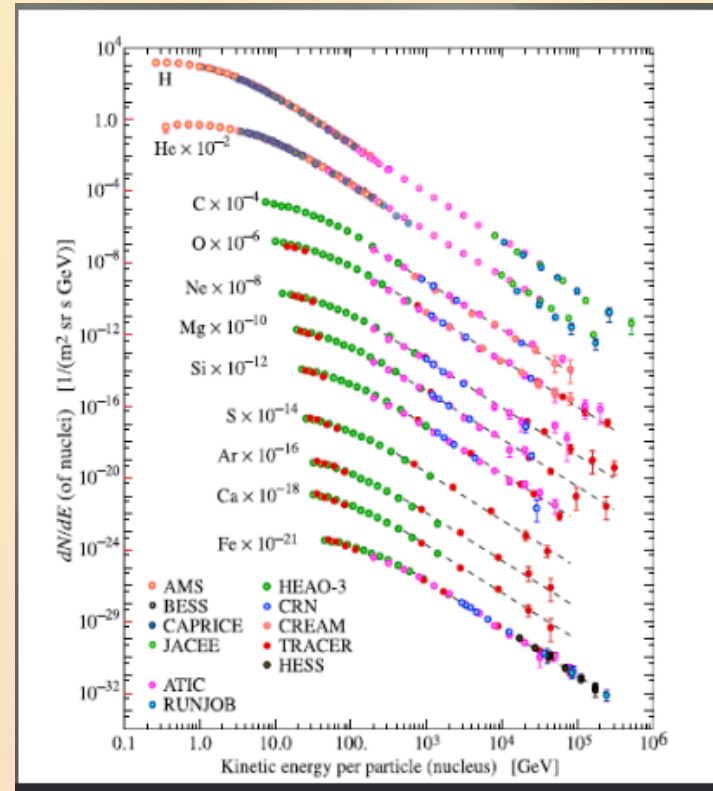
Cosmic rays from supernovae

- High-energy primary cosmic rays produced by supernovae



Primary cosmic rays

- 89% hydrogen nuclei (protons)
- the rest is Helium, Carbon, Oxygen, and other less abundant elements



Muons from cosmic rays

- Cosmic ray muons produced when primary cosmic rays impact with earth's atmosphere
- 1990, [Gaisser](#) formula for flux at sea level

$$\frac{dI_\mu}{dE_\mu} = 0.14 \left(\frac{E_\mu}{\text{GeV}} \right)^{-2.7} \left[\frac{1}{1 + \frac{1.1 E_\mu \cos \theta}{115 \text{ GeV}}} + \frac{0.054}{1 + \frac{1.1 E_\mu \cos \theta}{850 \text{ GeV}}} \right]$$

- Valid only if:
 - Earth curvature negligible ($\theta < 70 \text{ deg}$)
 - Muon decay negligible ($E_\mu > 100/\cos \theta \text{ GeV}$)

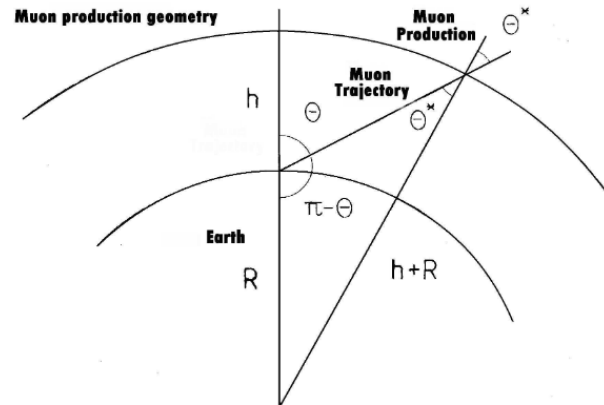


Fig. 1. The relation of the observed zenith angle of muons, θ^* , to the zenith angle at the muon production point in the atmosphere, θ . R is the radius of the Earth. Adopted from [\[3\]](#) [\[4\]](#)

Muons from cosmic rays

- Improved formula by [Guan et al. \(2015\)](#)
- Account for Earth's curvature
- θ at ground and θ^* at production differ

$$\cos \theta^* = \sqrt{\frac{(\cos \theta)^2 + P_1^2 + P_2(\cos \theta)^{P_3} + P_4(\cos \theta)^{P_5}}{1 + P_1^2 + P_2 + P_4}}$$

P_1	P_2	P_3	P_4	P_5
0.102573	-0.068287	0.958633	0.0407253	0.817285

- Correction at low energies

$$\frac{dI_\mu}{dE_\mu} = 0.14 \left[\frac{E_\mu}{\text{GeV}} \left(1 + \frac{3.64 \text{ GeV}}{E_\mu (\cos \theta^*)^{1.29}} \right) \right]^{-2.7} \times \left[\frac{1}{1 + \frac{1.1 E_\mu \cos \theta^*}{115 \text{ GeV}}} + \frac{0.054}{1 + \frac{1.1 E_\mu \cos \theta^*}{850 \text{ GeV}}} \right]$$

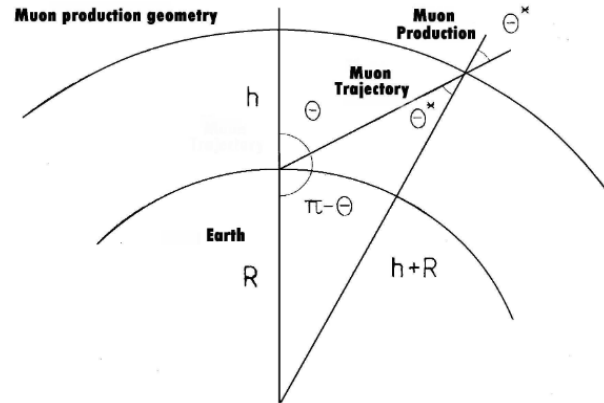
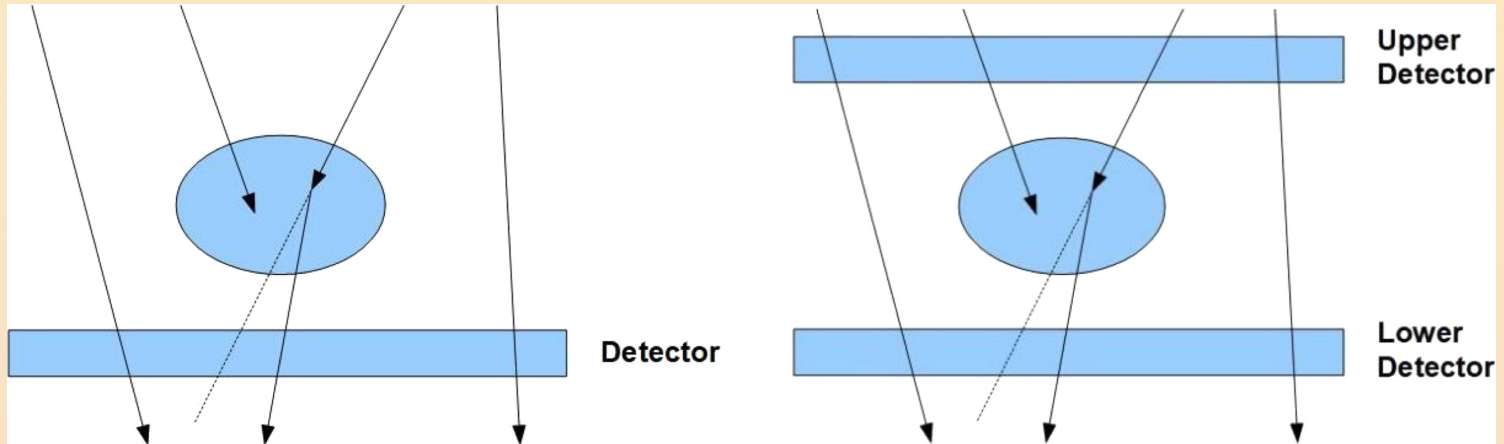


Fig. 1. The relation of the observed zenith angle of muons, θ^* , to the zenith angle at the muon production point in the atmosphere, θ . R is the radius of the Earth. Adopted from [3][4]

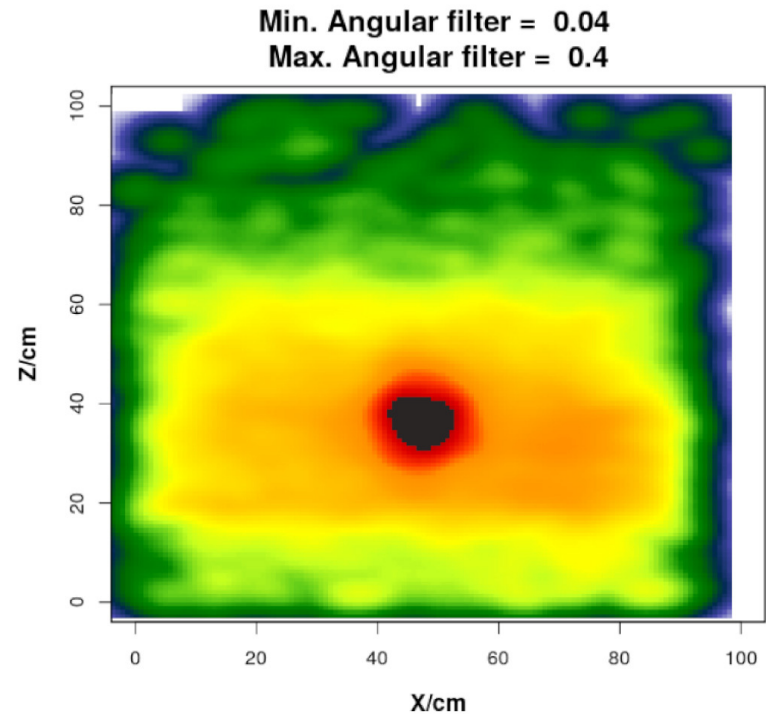
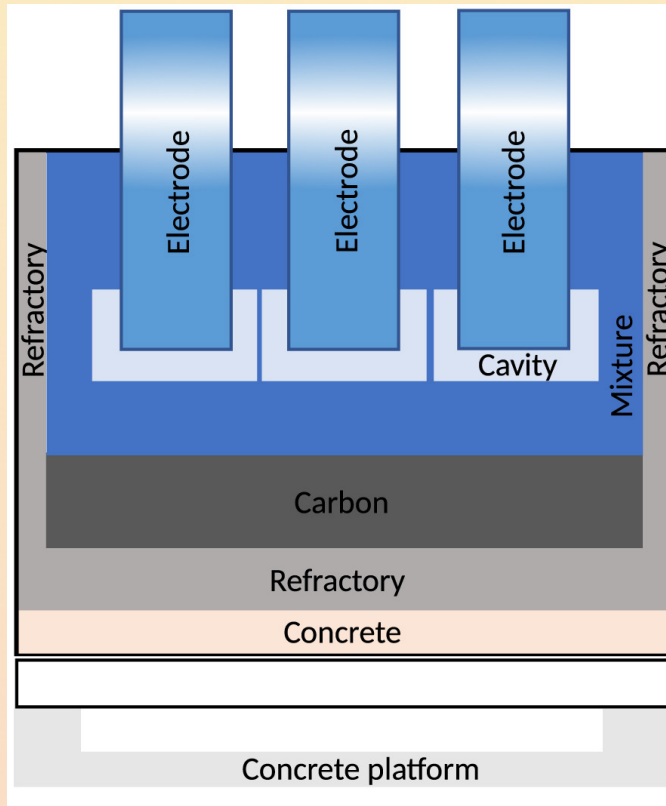
Absorption or scattering

- Absorption: measure missing flux
 - Pyramids, volcanoes...
- Scattering: measure deflection of muon trajectories
 - Containers, furnaces, statues...



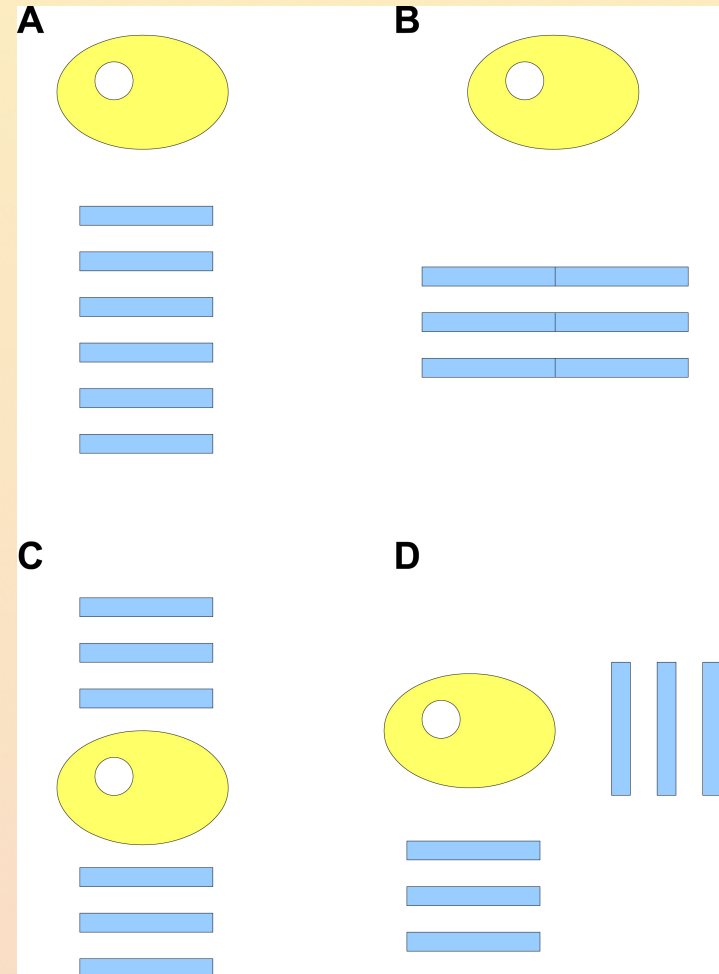
Heart of steel

- Map electrode position in electric furnaces
- Wear of inner walls of pipes



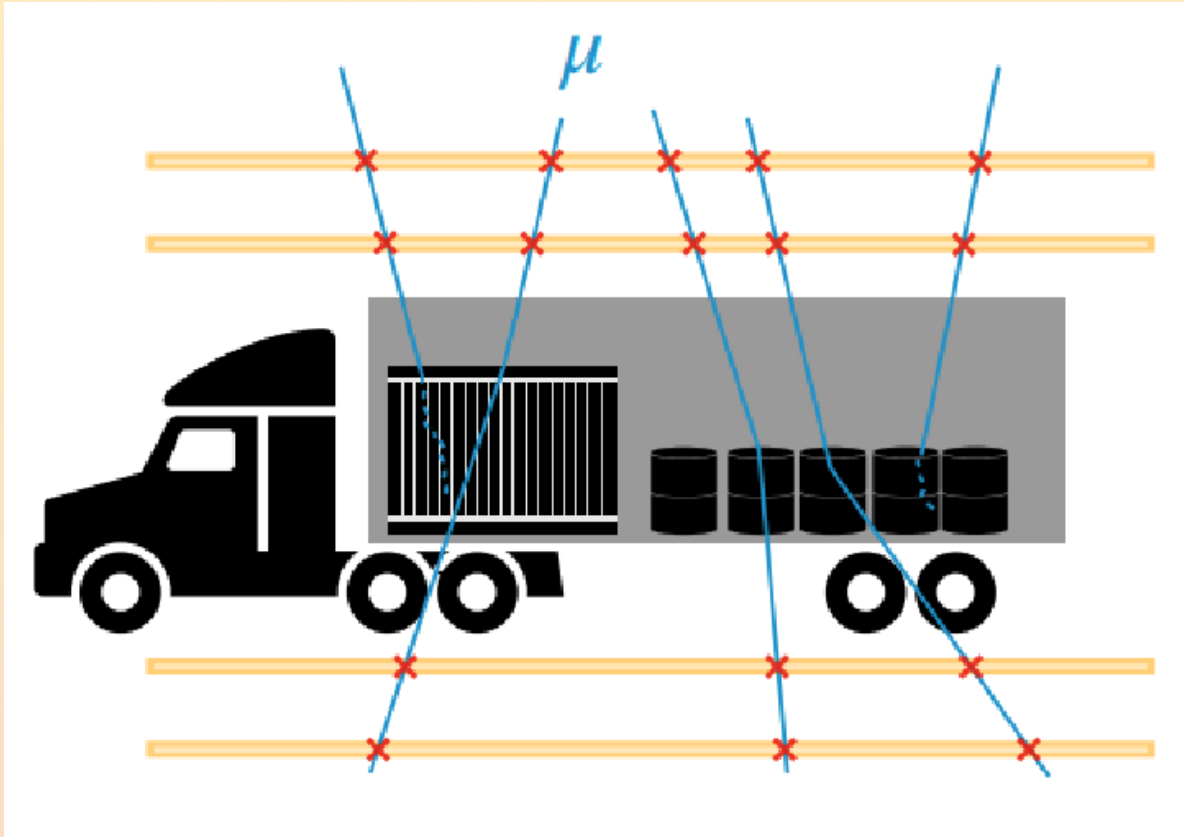
Portable muon detectors

- (A) Maximize angular resolution
- (B) Maximize acceptance
- (C) 3D tomography
- (D) 2D projections in orthogonal planes



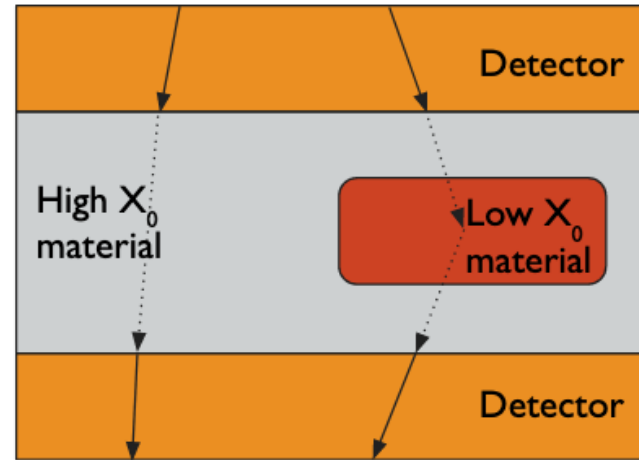
Border control

$$L = \left(1 + e^{k(c_\theta - c_0)}\right) \sum_z \left[w_{imp}(Z) m_{50,\alpha}^{\text{concealed}}[s(Z)] \right]$$



Scanning a passive volume

- Want to infer properties (e.g. 3D map of elemental composition) of unknown volume
 - Shipping container, archeological site, nuclear waste dump, industrial machinery, etc.
- Muons from cosmic rays traverse us all the time
 - On average, 1 muon per cm^2 per minute
 - Change in kinematics provides handle for inference on X_0



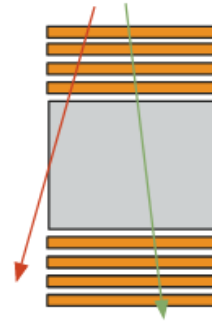
High X_0 = low scattering

Low X_0 = high scattering

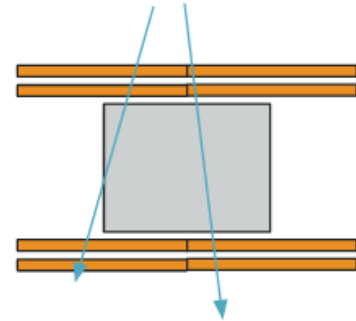
X_0 = average distance between scatterings

Domain knowledge is not enough

- Domain knowledge typically provides heuristics based on proxy objectives
- Will likely have a budget
 - Money, heat, power, positioning of detectors, imaging time...
- Will likely have varying purposes
 - Today want to spot uranium, tomorrow e.g. drugs



Example 1:
Muons
measured
precisely but
less efficiently

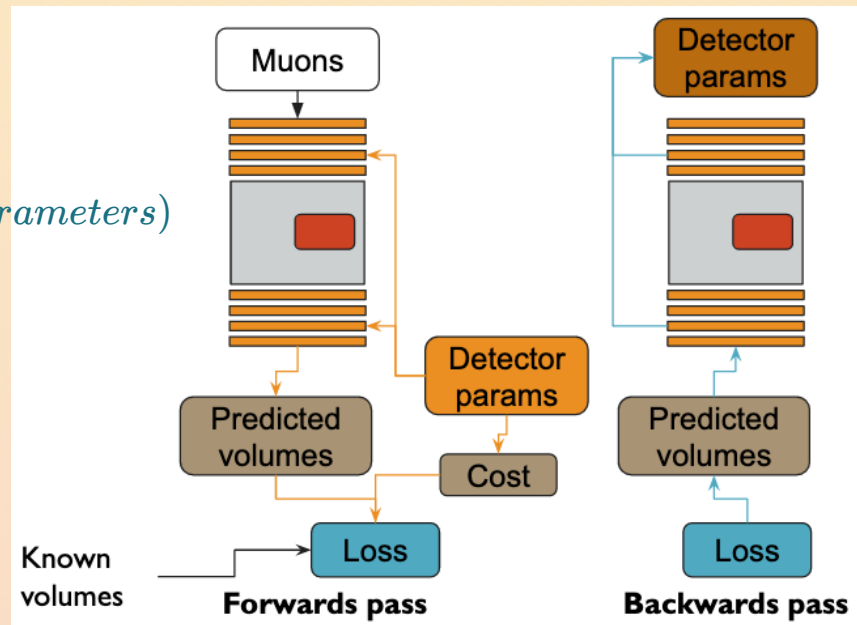


Example 2:
Muons
measured less
precisely but
more
efficiently

TomOpt

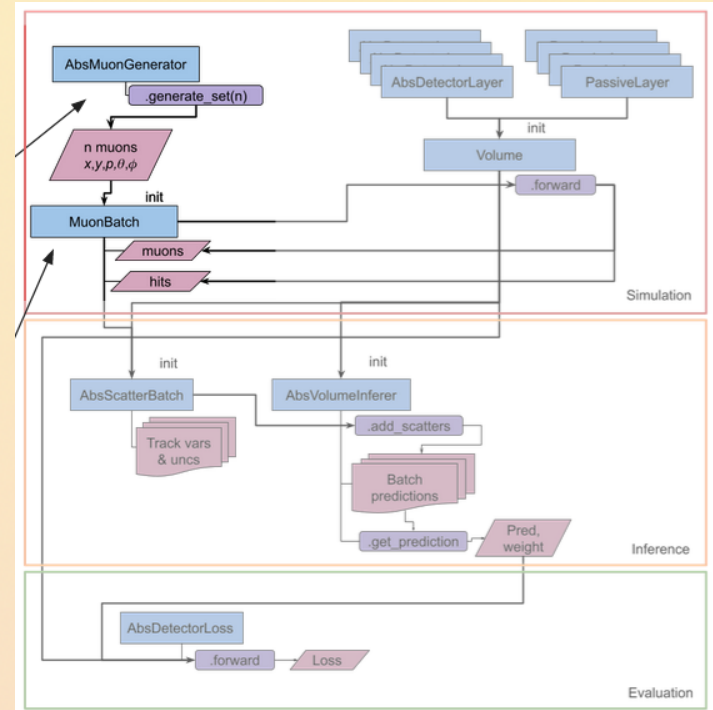
- Differential optimization of muon-tomography detectors (ongoing project)
 - [Giles C. Strong](#), Maxime Lagrange, Aitor Orio, Anna Bordignon, Florian Bury, Tommaso Dorigo, Andrea Giammanco, Mariam Heikal, Jan Kieseler, Max Lamparth, Pablo Martínez Ruíz del Árbol, Federico Nardi, Pietro Vischia, Haitham Zaraket
 - [2309.14027](#) submitted to journal, shorter version accepted by NeurIPS MLPS Workshop!
 - Modular design in python, autodiff via PyTorch

- Inference chain as differentiable pipeline
 - Can compute $p(\Delta_{output} | \Delta_{detector\ parameters})$
- Task as loss function
 - Including target (e.g. prediction uncertainty), costs, constraints
- Backpropagate and optimize as usual



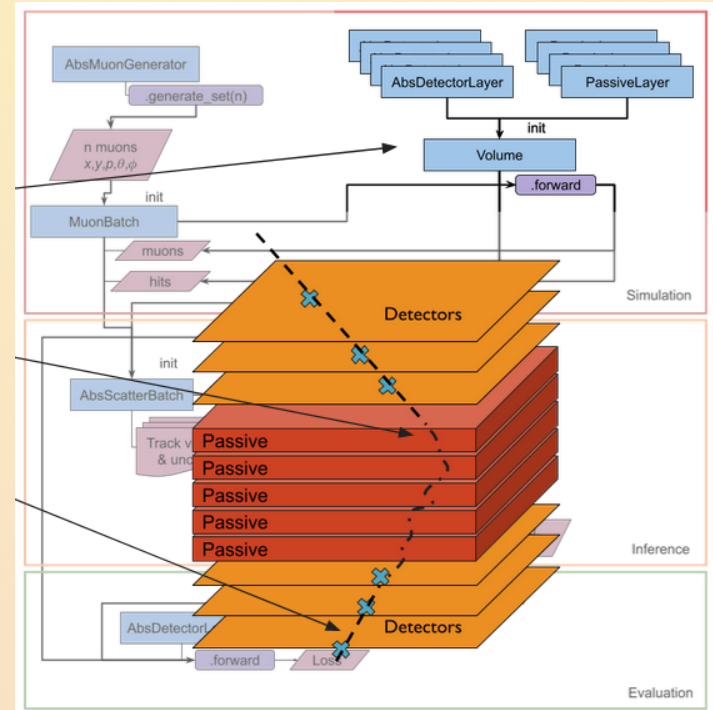
Muon Generation

- Formulas by 2015 and 2016 models
- Tensor of muons (x, y, p, θ, ϕ)
 - θ, p from flux model
 - x, y from ranges
 - ϕ uniformly in $[0, 2\pi]$
- Code handles many muons at once (MuonBatch)
 - Propagate the muon position (can snapshot to track)
 - Scatter ($dx y, d\theta d\phi$) at each step



Volume Specification

- Volume made up of stacked layers in z
- Passive layers scatter muon
 - PDG and GEANT models both available
 - Voxelized passive layers (x, y)
- Active layers record muon hits
 - Parameterized efficiency and resolution (cost per m^2 , physics constraints)
 - Budget is a volume attribute, and is assigned to detector layers



Panel specification

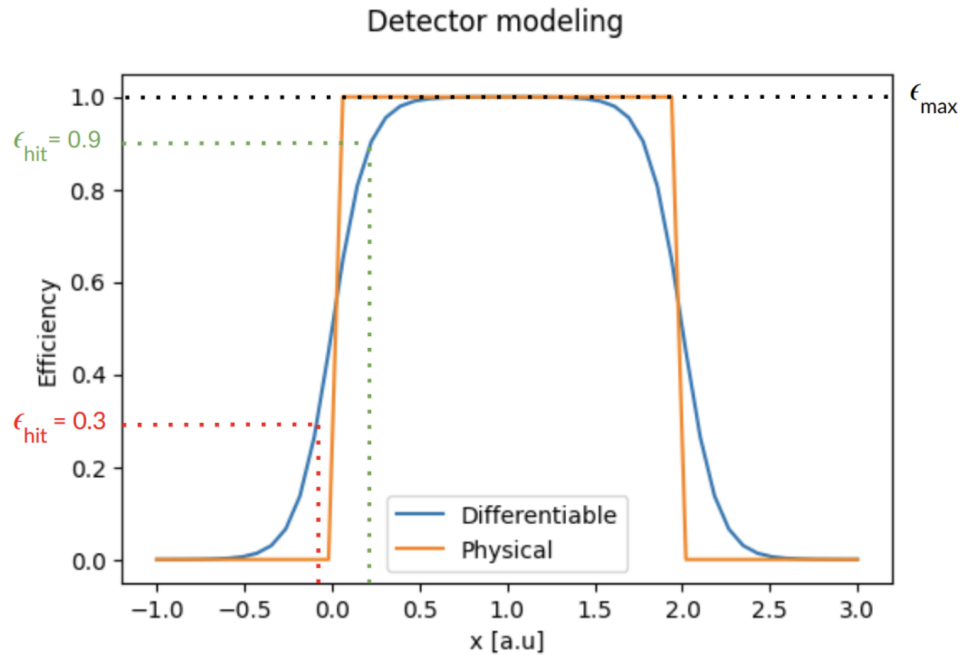
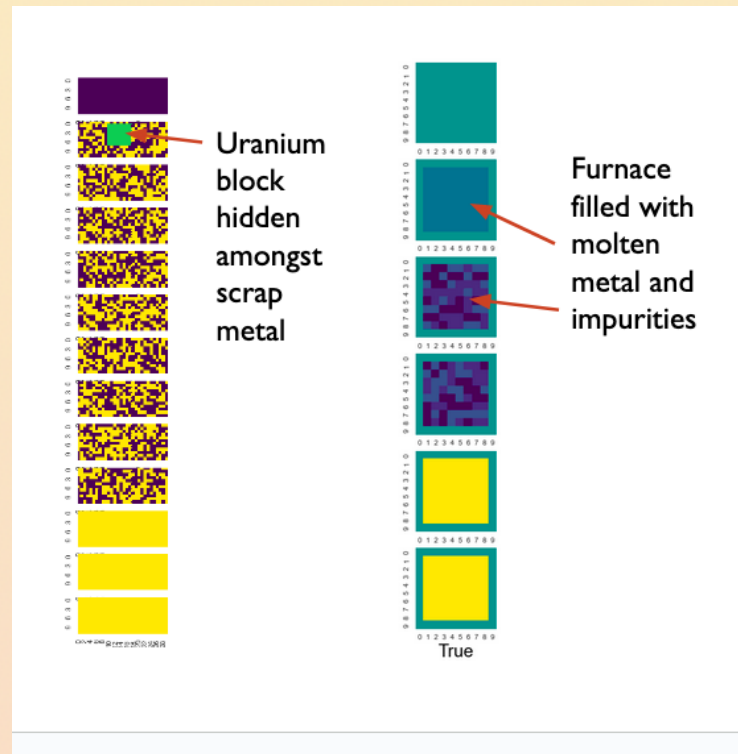


Figure 2: Example of detector panel modeling with sigmoid function used during optimisation (blue) and with rectangle function used during validation (orange).

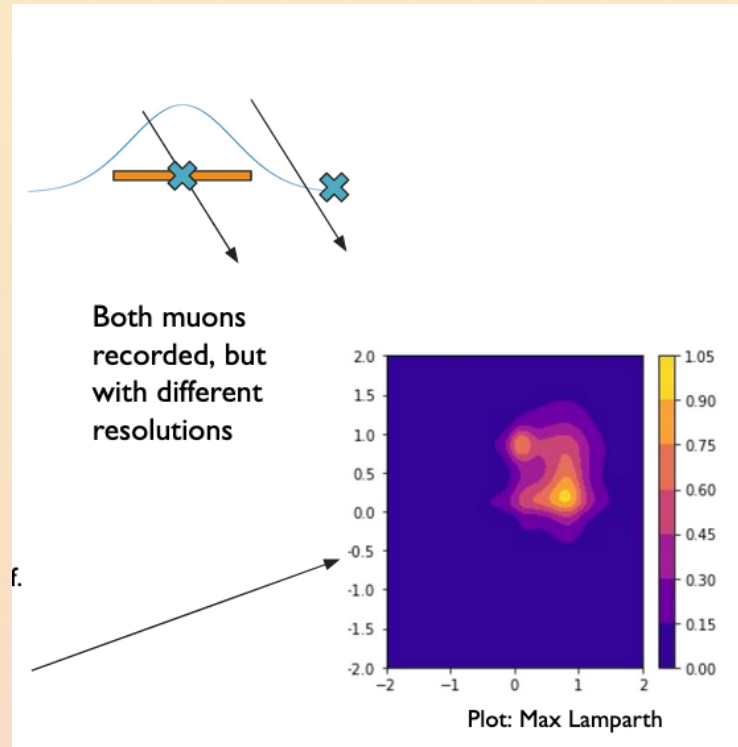
Monte Carlo Truth

- Per each scenario, can build voxelized random volumes
 - Each voxel can be a different material
 - Next: material mixture per voxel



Make muon hits differentiable

- Associate a distribution to resolution and efficiency
 - e.g. Gaussian centered on panel and width equal to panel span
 - p.d.f. of the muon position is now differentiable
- Further generalization: Gaussian Mixture models



From hits to tracks

- Analytic maximum likelihood fit
 - considering uncertainty and efficiency of hits
 - fully differentiable w.r.t. detector parameters
- Provides track parameters and their uncertainties

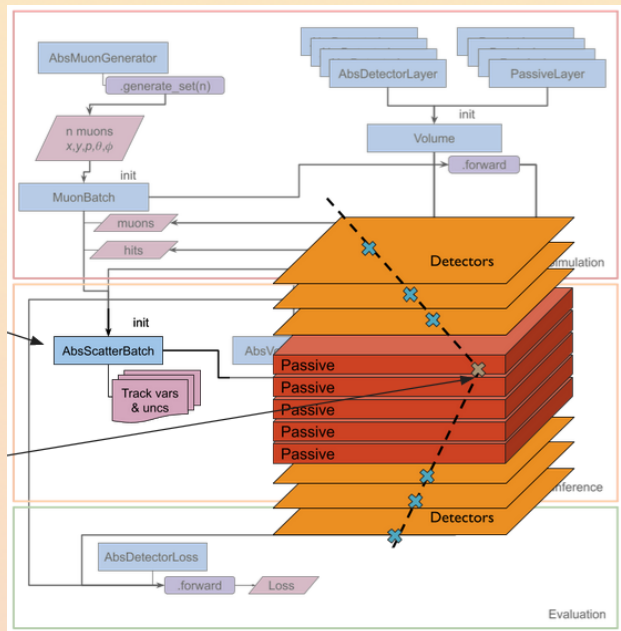
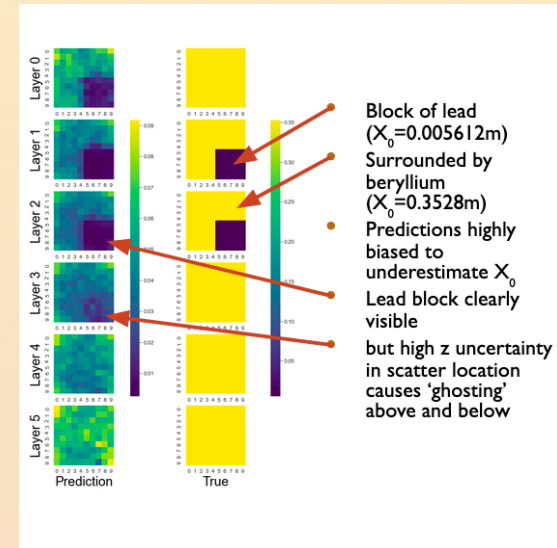


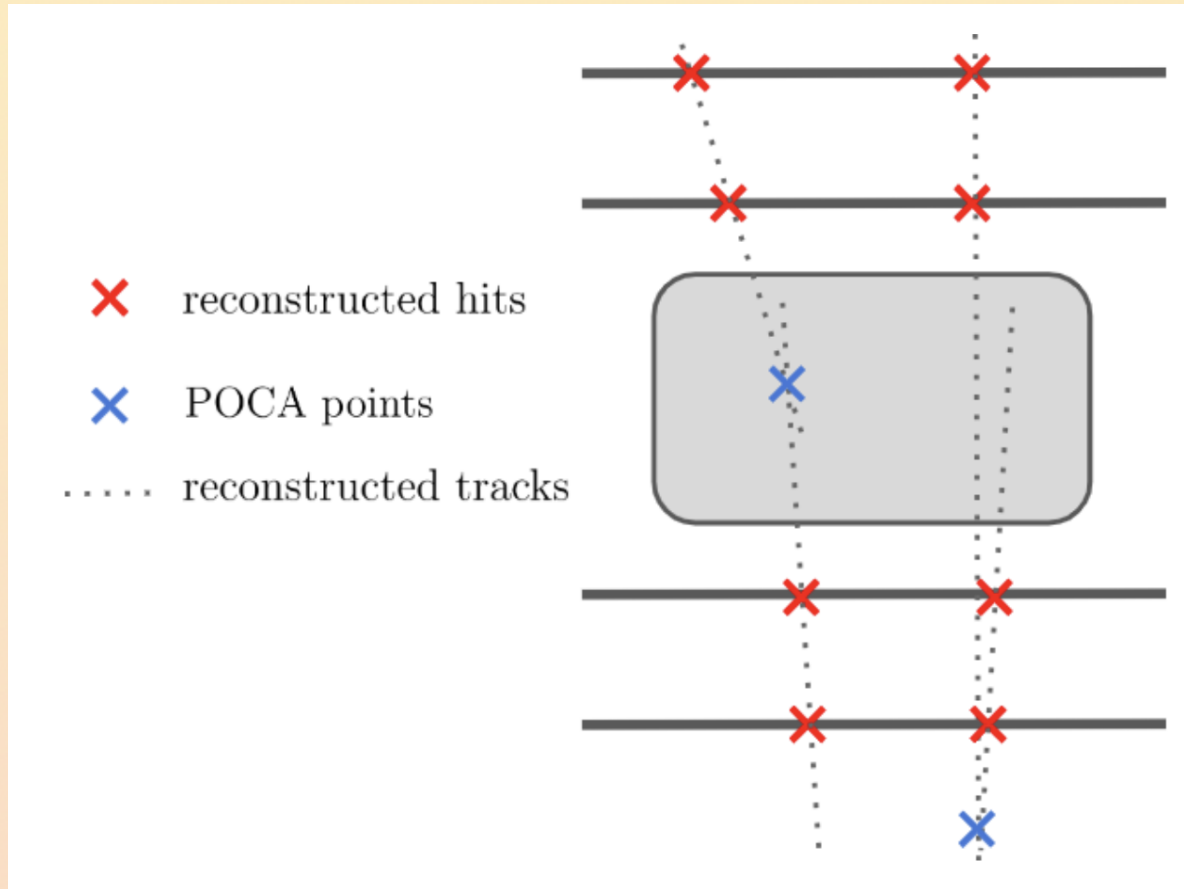
Figure by G.C.Strong

- POCA (POint of Closest Approach)
 - assume one scattering in one point
 - invert model to compute X_0
 - average X_0 per voxel



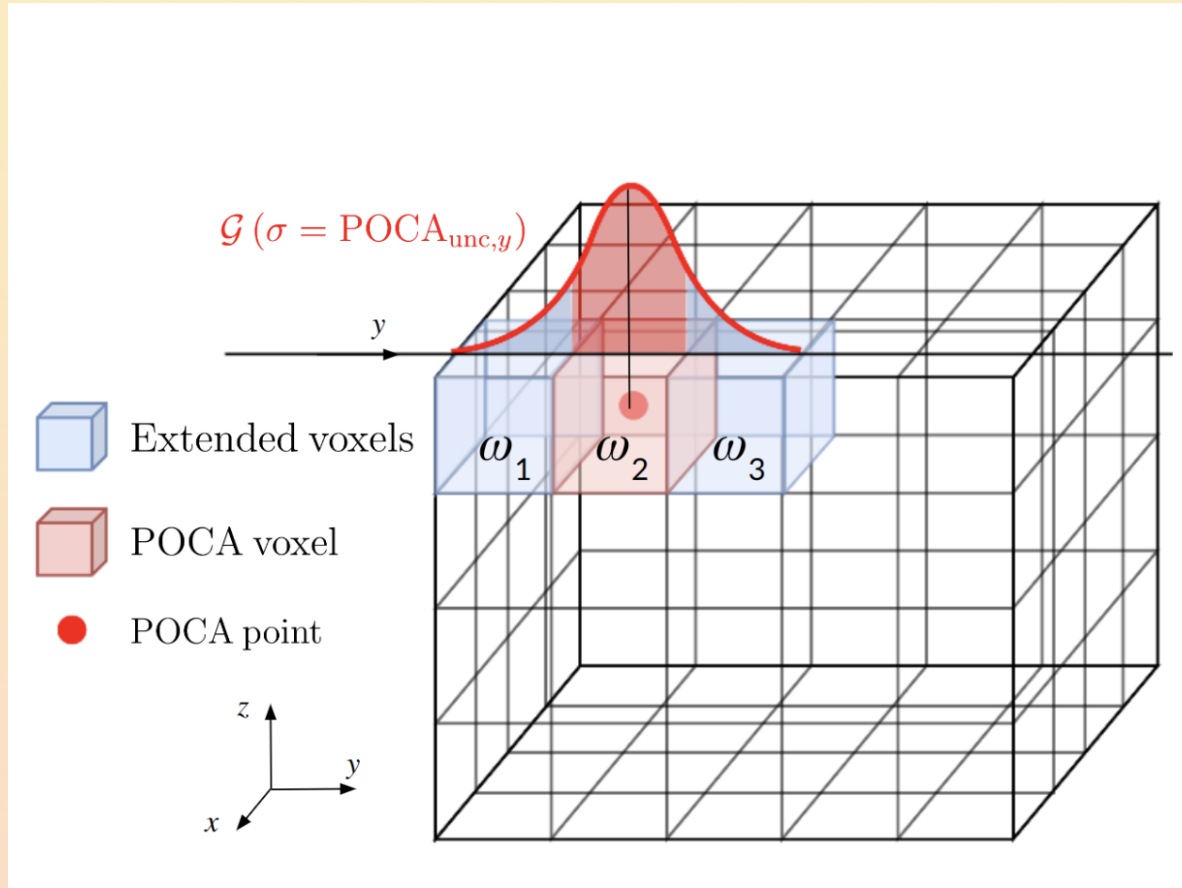
POCA

- Assume one scattering $\rightarrow$ bias!

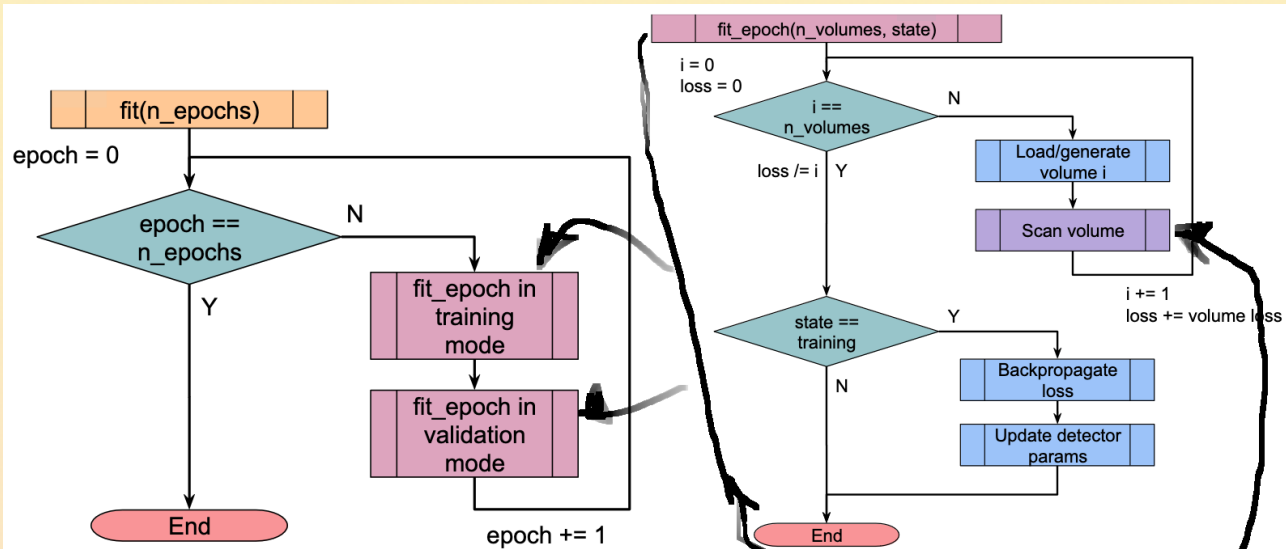


Extended POCA

- Assume one scattering $\rightarrow$ bias!

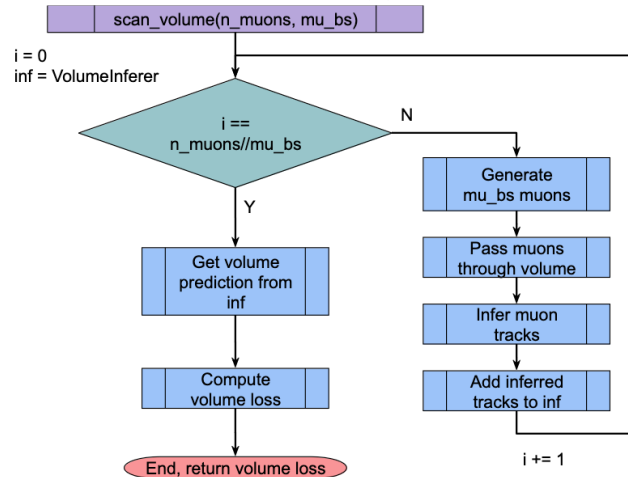


Detector Optimization in TomOpt



(a) Complete fit loop.

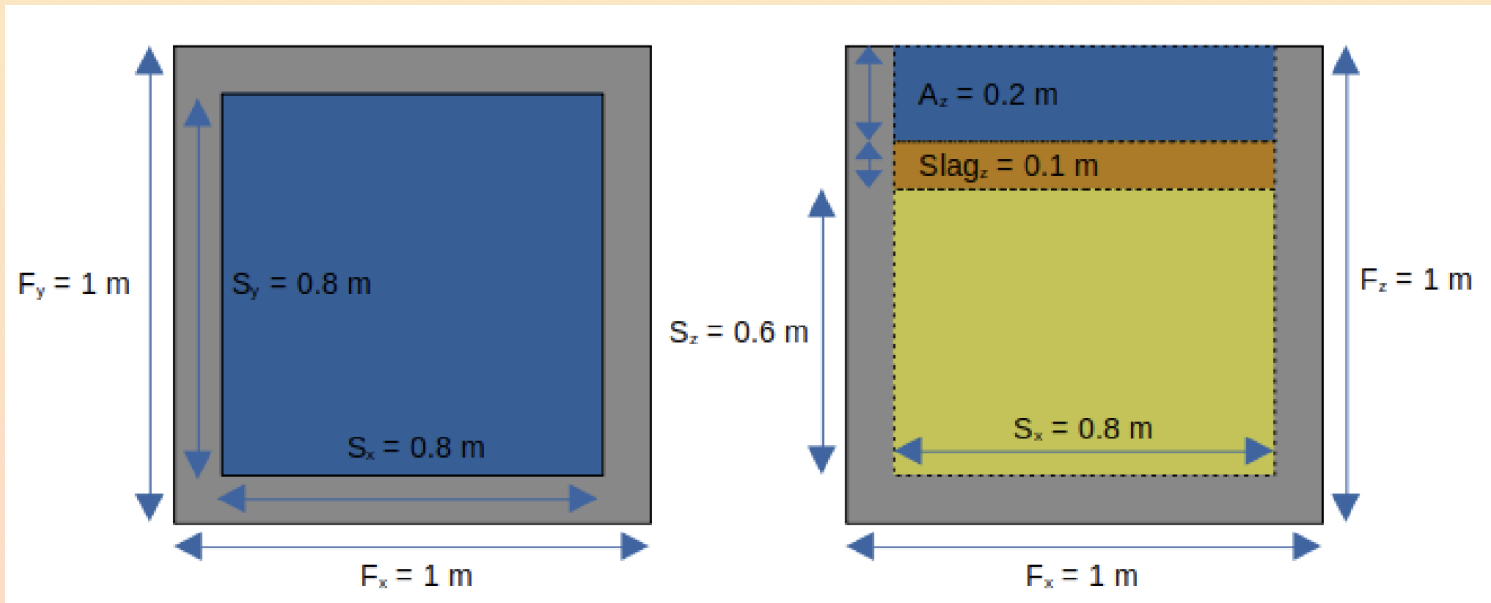
(b) Scan loop for a batch of passive volumes.



(c) Scan loop for muons over a single passive volume.

Heavy Metal

- Transport liquid steel to fill moulds
 - Lack of enough metal: moulds not filled
 - Too much metal: remnants, scraps
 - Slags hides metal from optical inspection

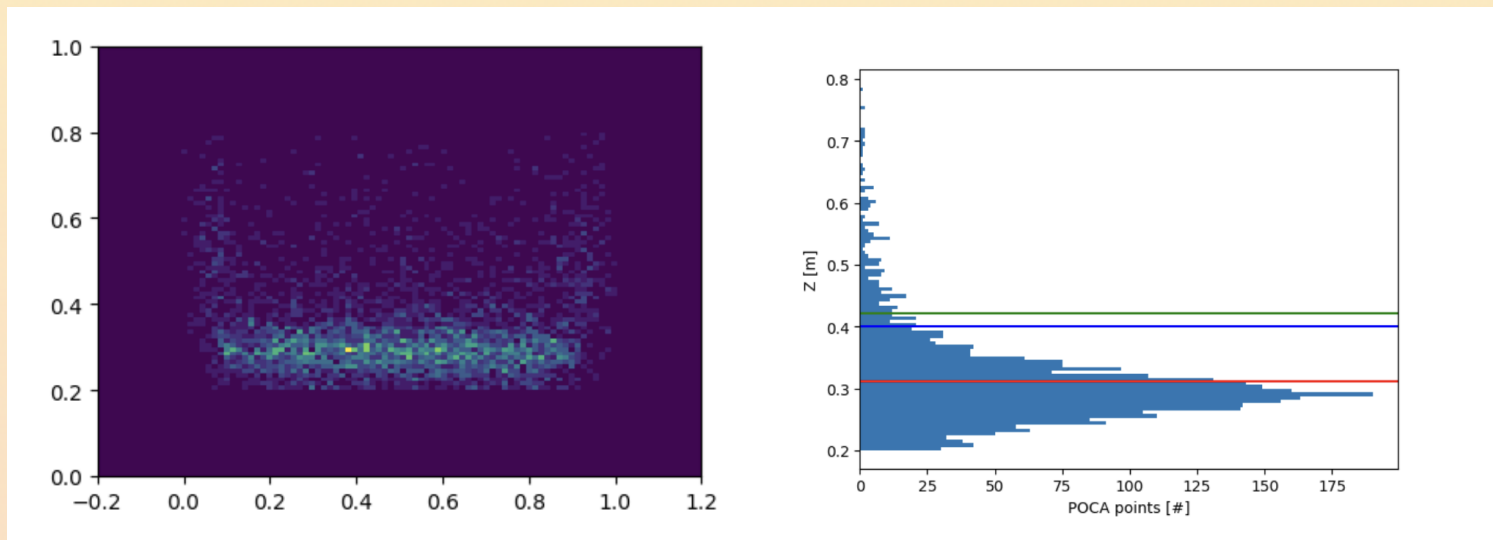


Heavy Metal

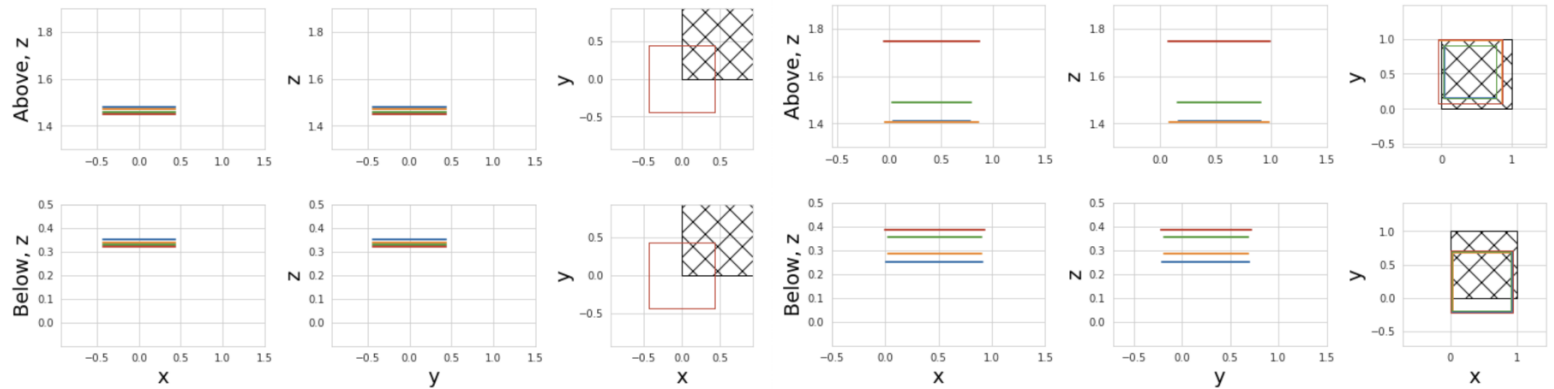


Encouraging results

- But POCA provides a biased estimator, with bias increasing with fill level (more metal, more bias)
- Debiasing via parametric correction



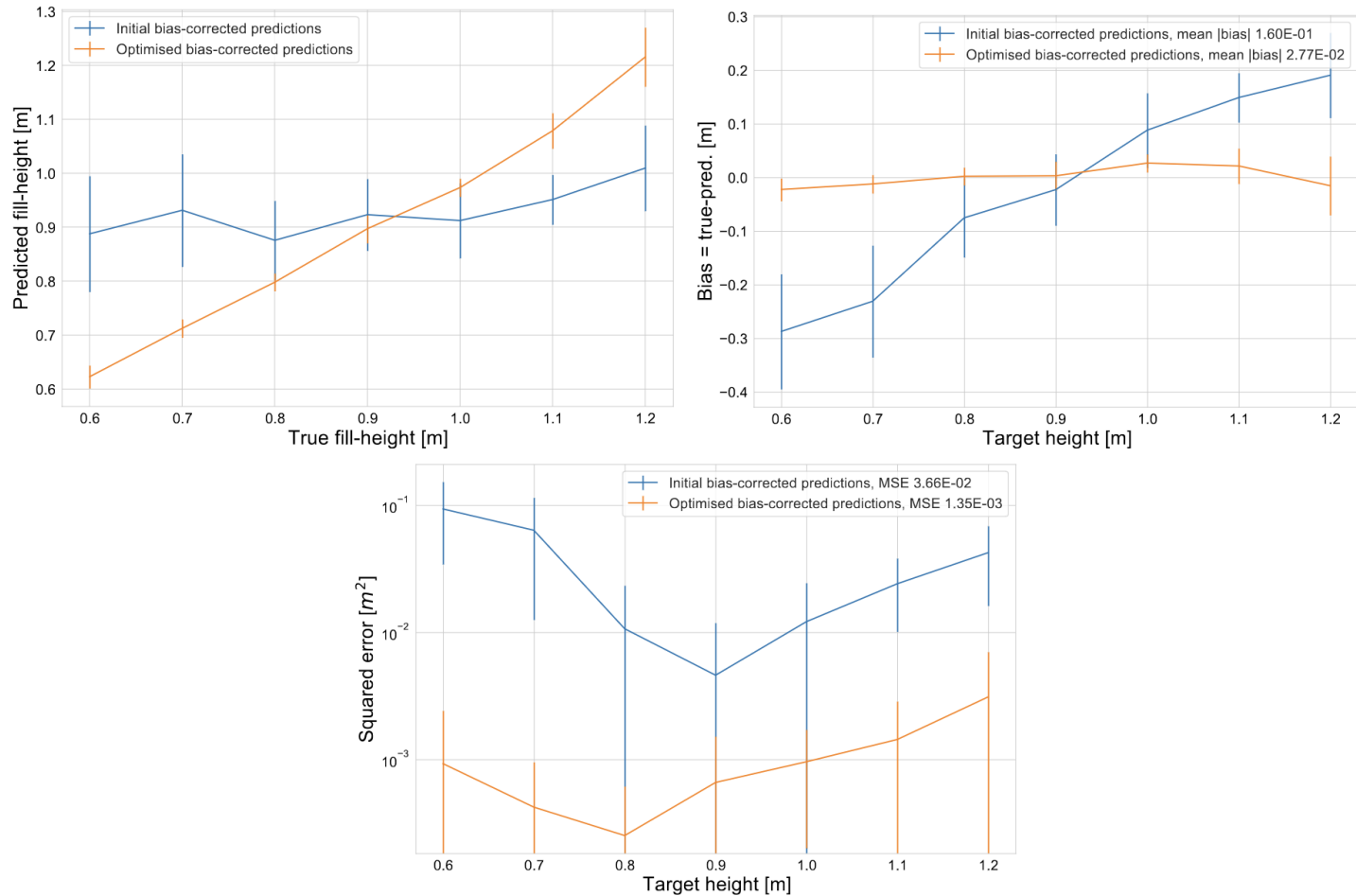
Optimize crappy detector



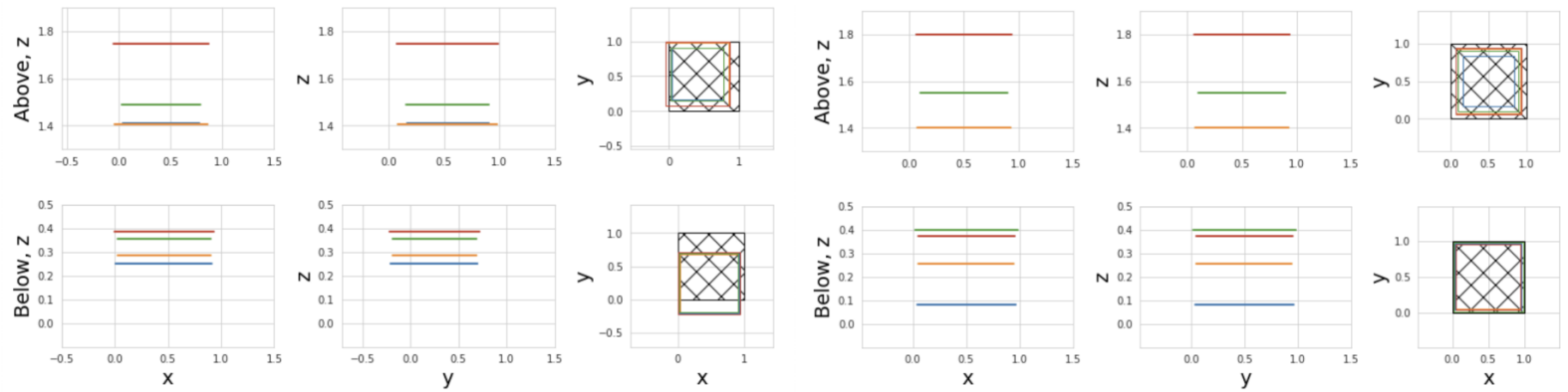
(a) Initial detector configuration.

(b) Detector configuration after stage one optimisation process.

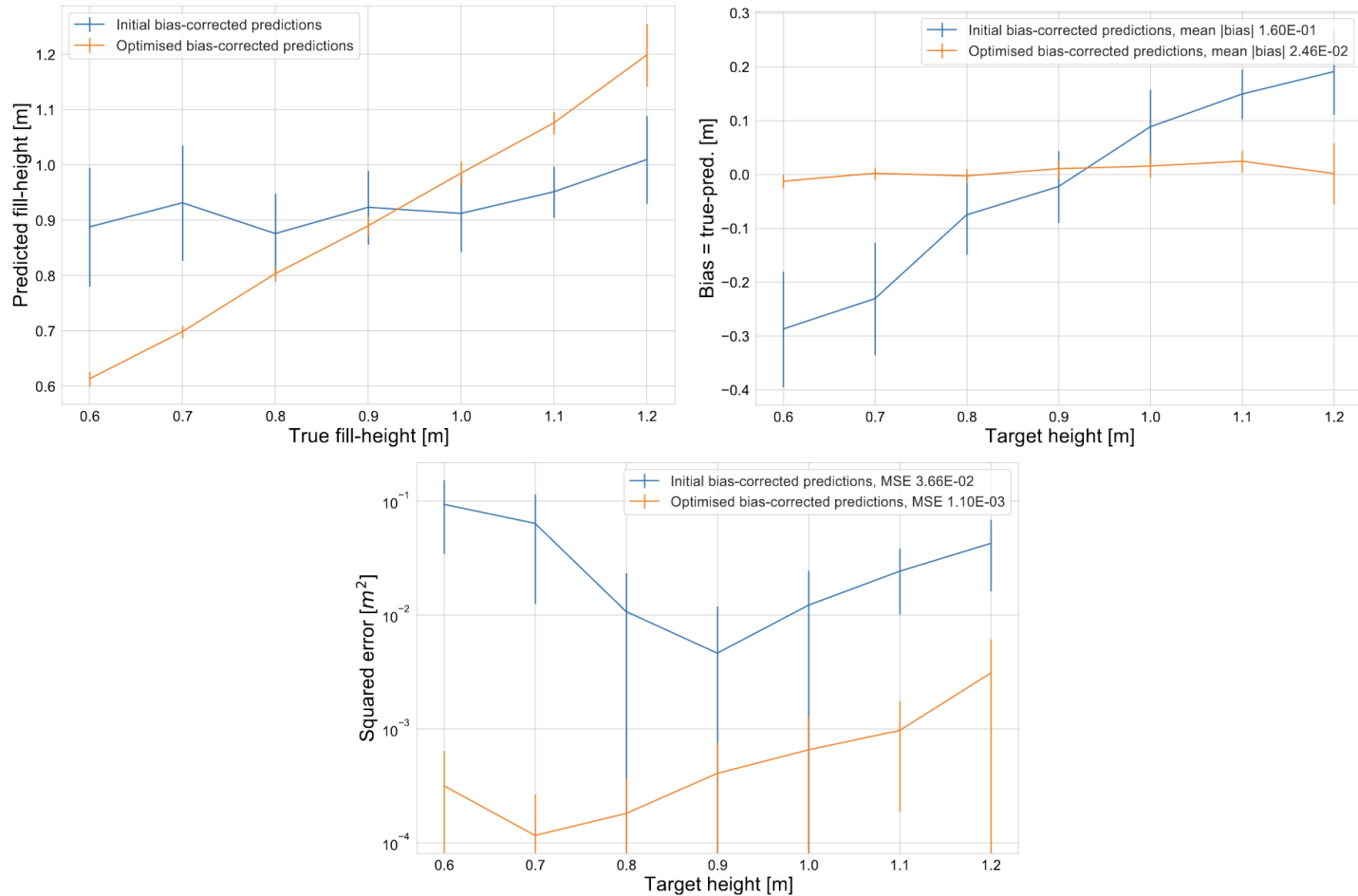
Optimize crappy detector



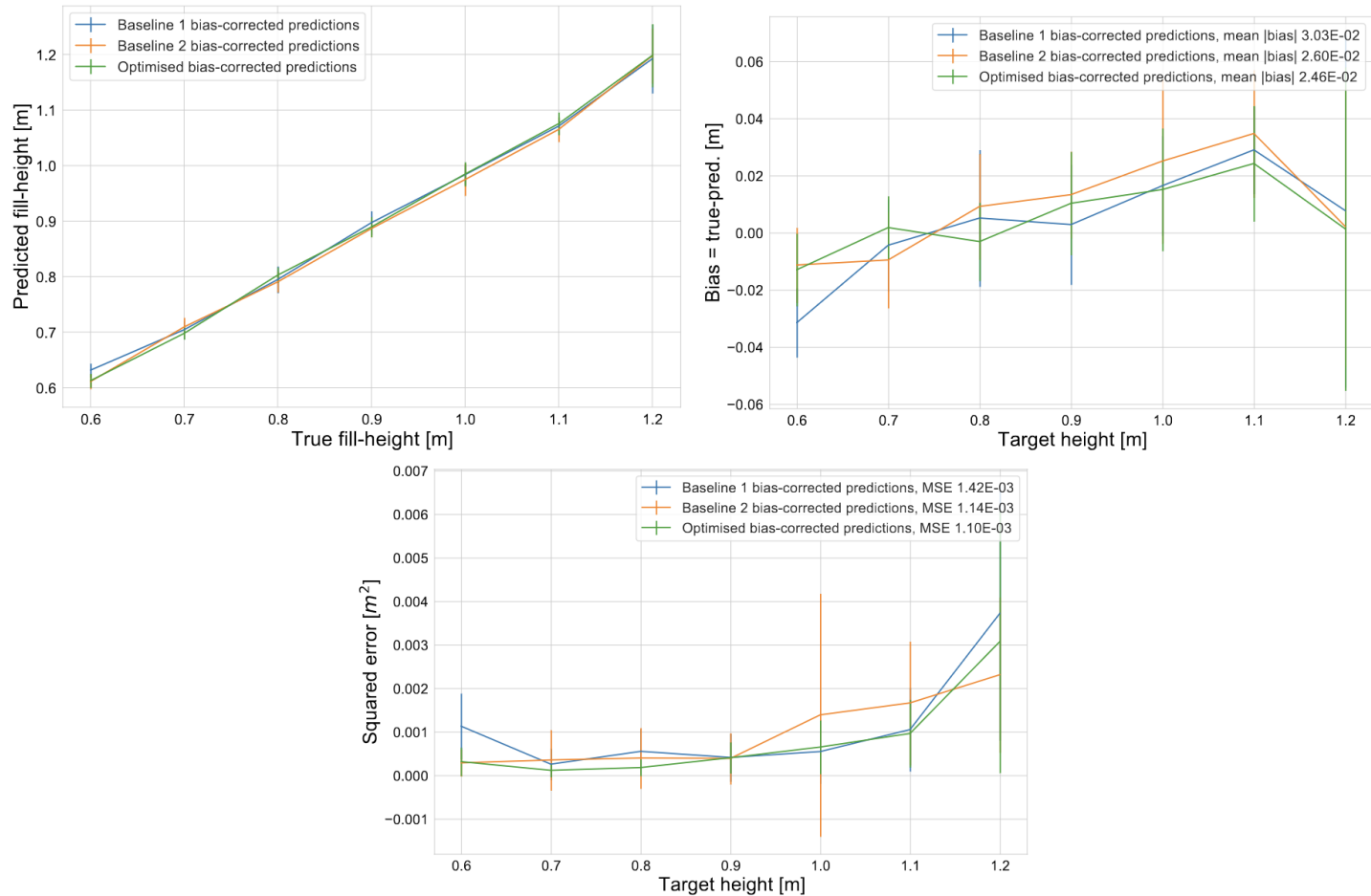
Refine a good detector



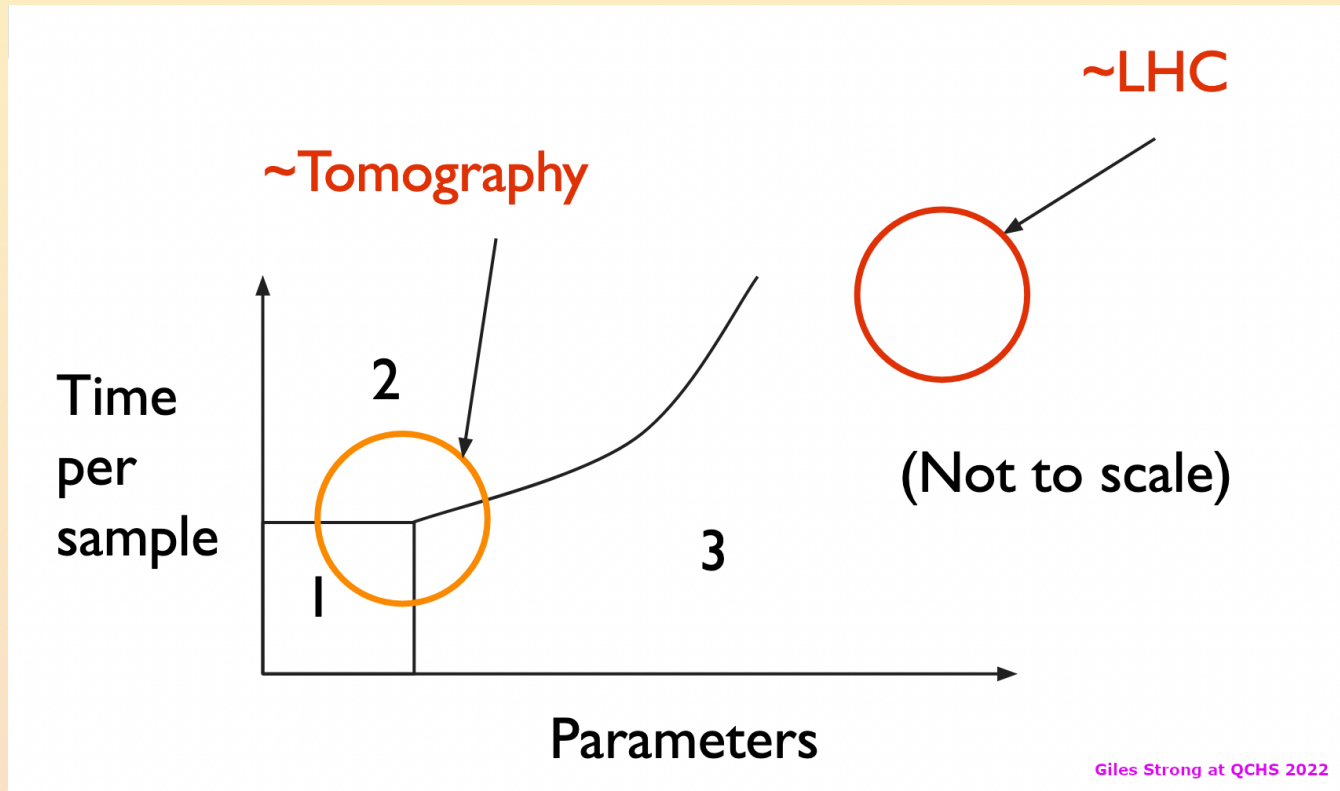
Refine a good detector



Recover human baselines



Method of choice depends on scale



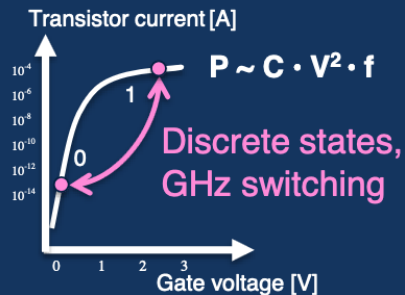
1. Grid/random search
2. Bayesian opt, simulated annealing, genetic algos, ...
3. [Gradient-based optimization](#) (Newton, BFGS, [gradient descent](#), ...)

Need for new paradigm

- If you are interested in Neuromorphic computing or Quantum computing, drop me a line!

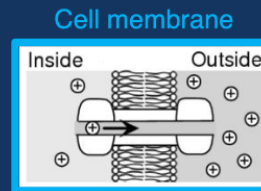
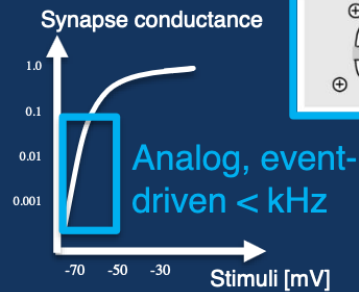
Conventional computers

*mimic
logical and
analytical
thinking*

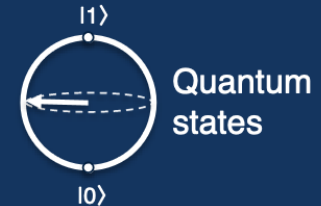


Neuromorphic processors

*mimic
the senses,
learning and
perception*



Quantum processors
*use quantum
superpositions
for probabilistic
inference*



Technology readiness?

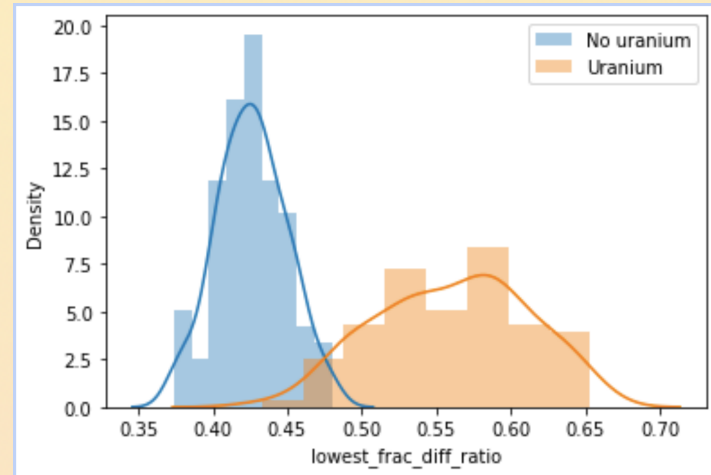
Thank you!



Backup

It's all about summaries

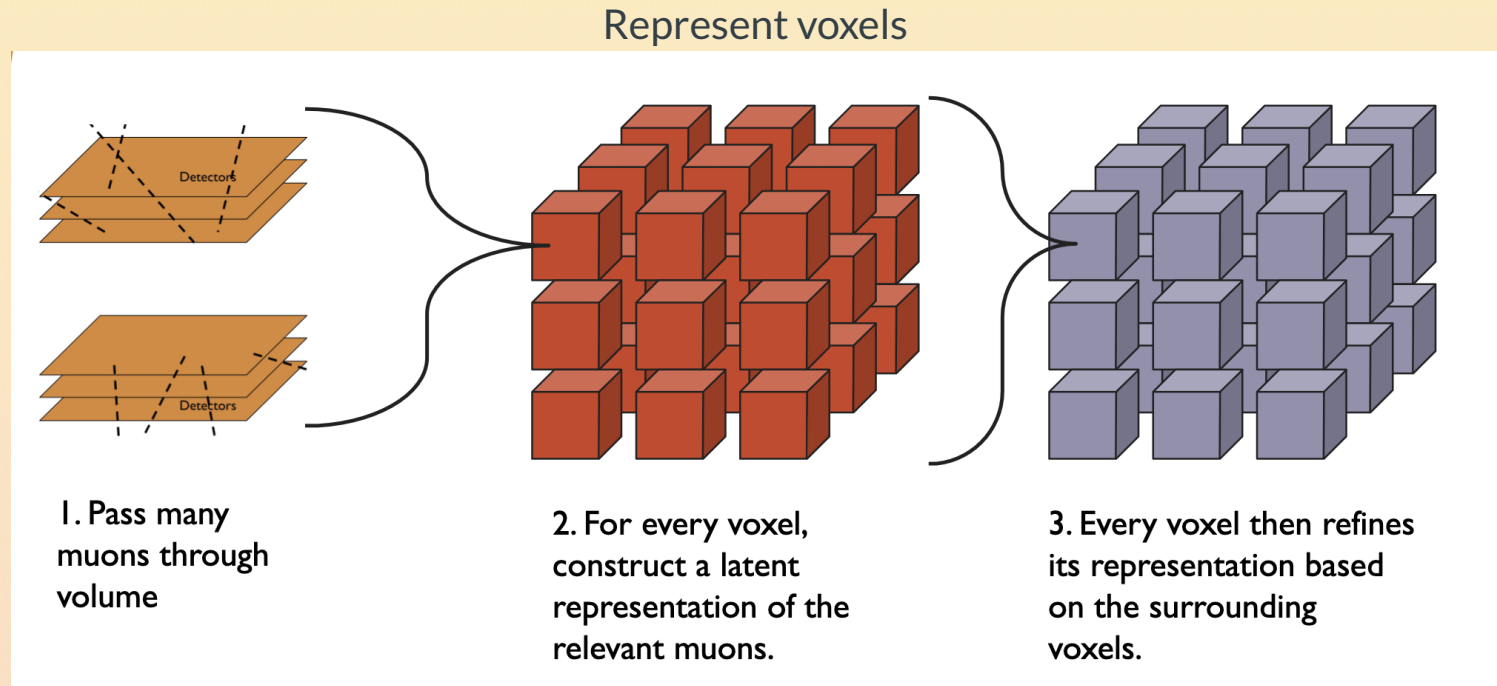
- Promising performance
- Muon track quantities differentiable
 - can compute uncertainties due to spatial resolution
 - useful for aggregating
- X_0 predictions biased and inaccurate
 - Uranium is denser than other materials
 - Under U hypothesis, large difference between mean of the n lowest X_0 voxels and mean X_0 of remaining voxels
 - One-dimensional summary statistic!



- Ultimately, the summary **can be learned** with **graph neural networks**)

Diff. representation with GravNet

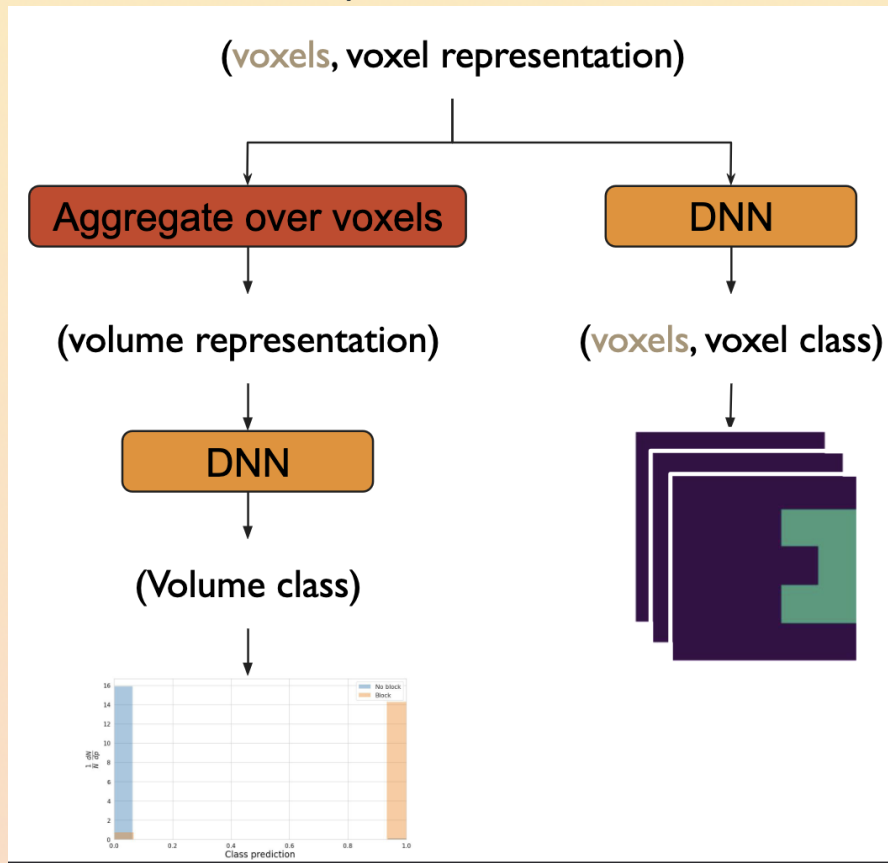
- Use [Quasim et al](#): distance-weighted representations, information-sharing along learnable connections



Diff. representation with GravNet

- Use [Quasim et al](#): distance-weighted representations, information-sharing along learnable connections

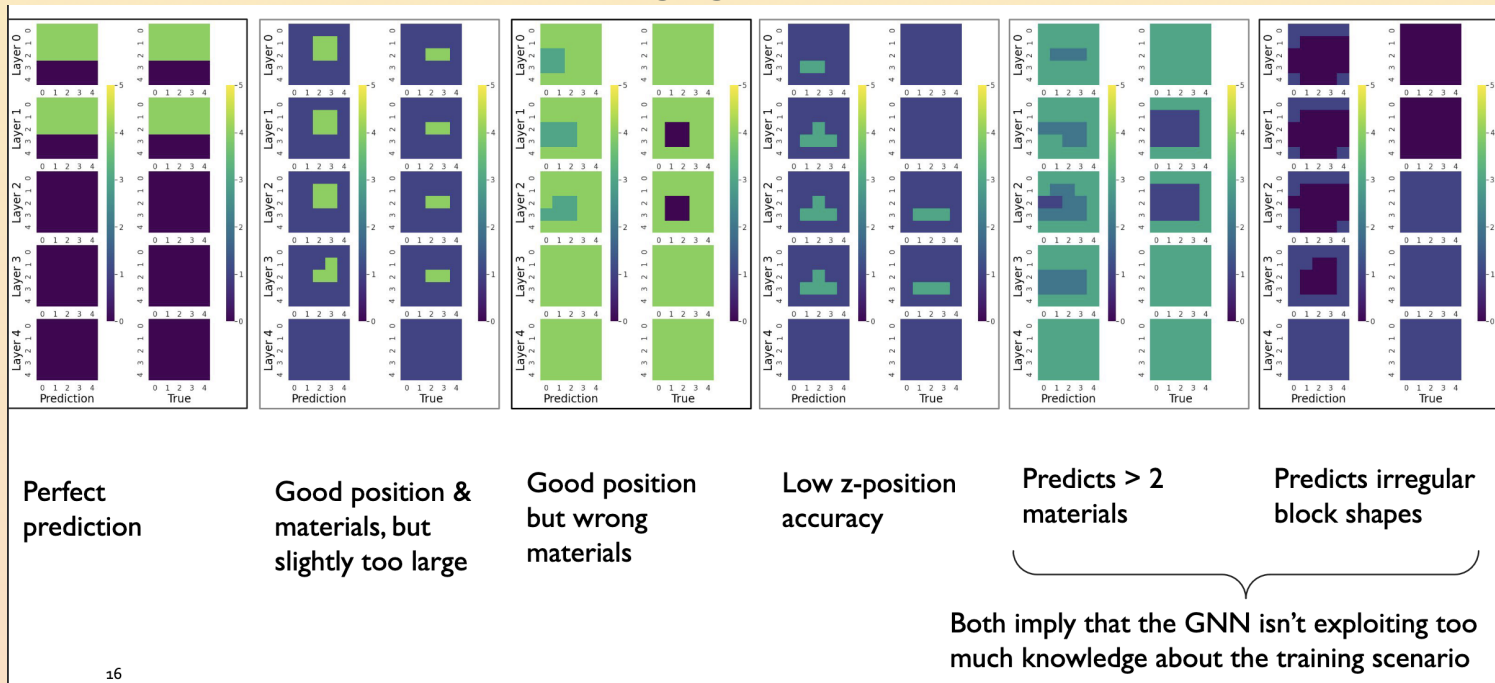
Predict volume (binary classification: there or not there)



Diff. representation with GravNet

- Use [Quasim et al](#): distance-weighted representations, information-sharing along learnable connections

Imaging behaviour



16

Infer the volume

- Expectation maximization:
iteratively
 - estimate scatter density based on current estimate of the image
 - update estimate of the image based on the estimated scatter density

- 1) Gather measurements of scattering and momentum for each muon $i=1$ to M : $(\Delta\theta_x, \Delta\theta_y, \Delta x, \Delta y, p_x^2)_i$.
- 2) Estimate geometry of interaction of each muon with each voxel $j=1$ to N : $(L, T)_{ij}$.
- 3) For each muon voxel pair, compute the weight matrix: W_{ij} using (24).
- 4) Initialize the scattering density in each voxel with a guess: $\lambda_{j,old}$.
- 5) Do until (stopping criteria)
 - a) For each muon, compute $\Sigma_{D_i}^{-1}$ using (29) and taking the inverse.
 - b) For each muon voxel pair, compute the conditional expectation terms: S_{ij} using (43).
 - c) Compute $\lambda_{j,new}$ using (38)
 - d) Set $\lambda_{j,old} = \lambda_{j,new}$ for all voxels.
- 6) End do

Summary of the EM algorithm for muon tomography. Schultz et al. 2007

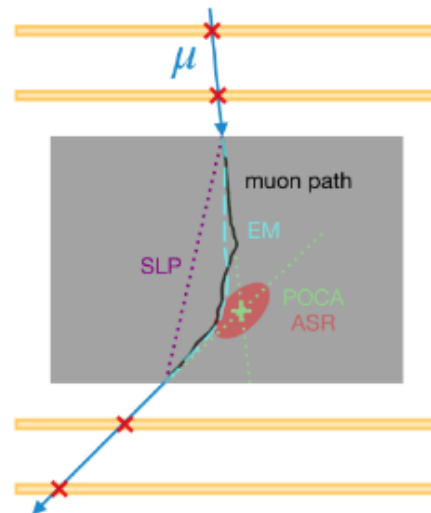
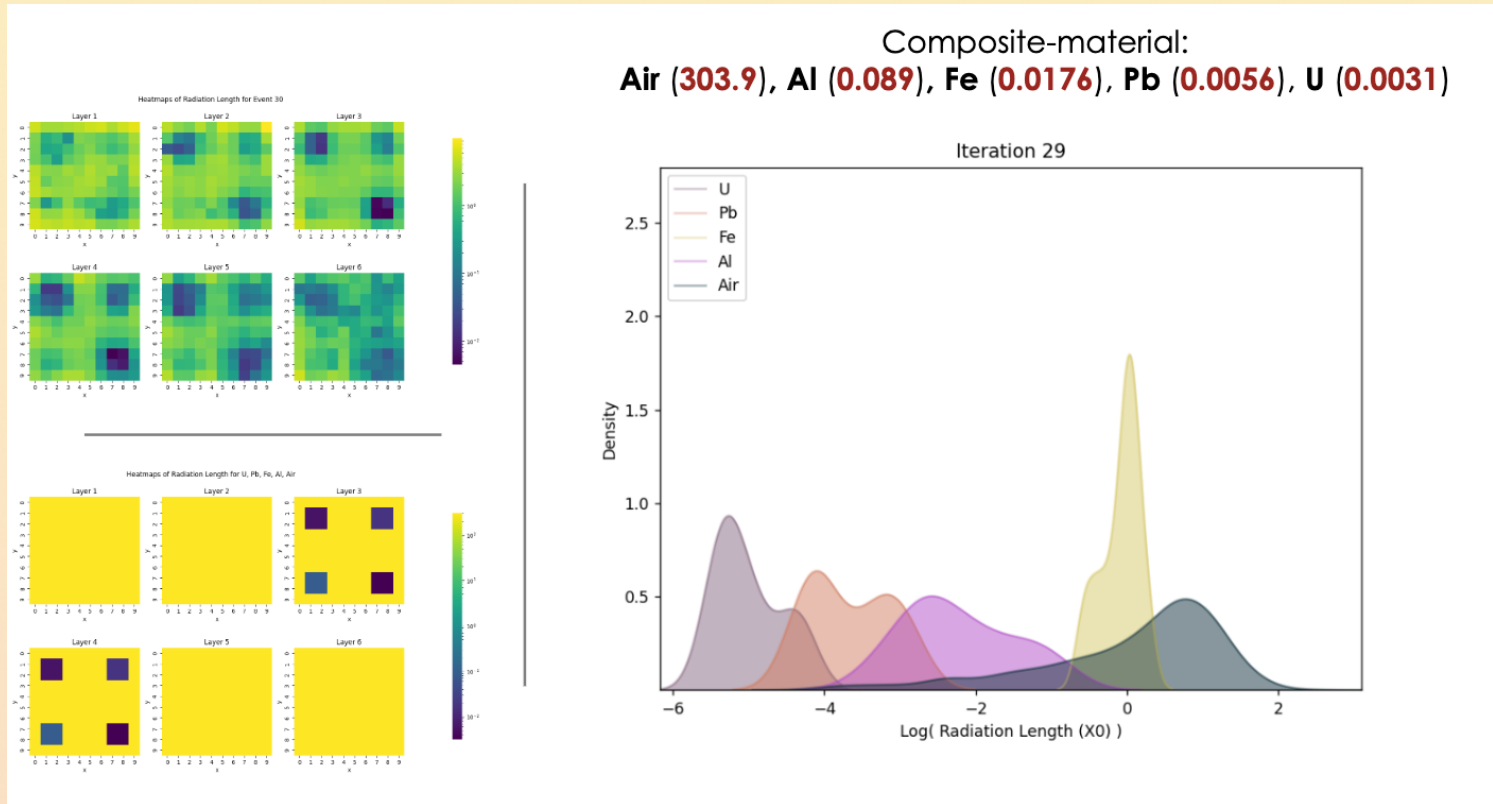


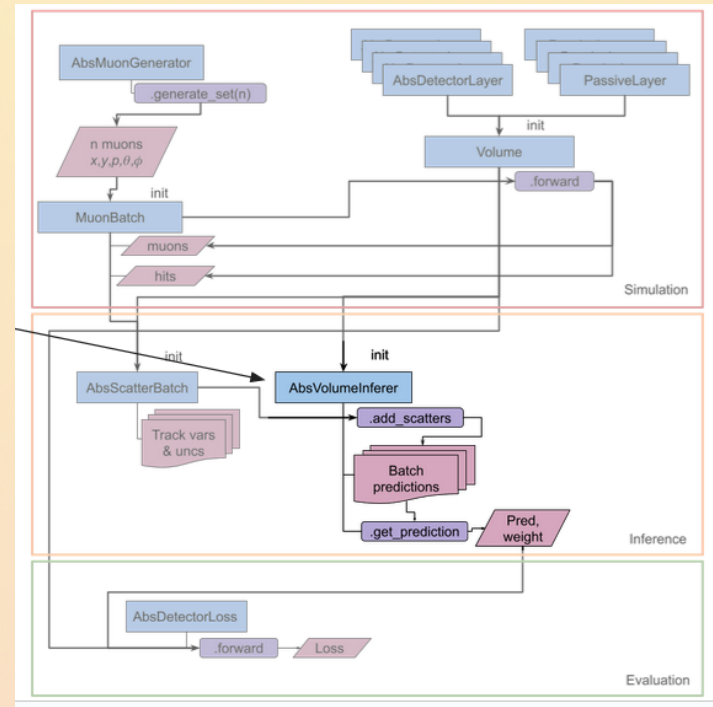
Figure 2: Illustration of commonly used algorithms for reconstructing images in muon scattering tomography. Barnes et al. 2023

EM preliminary performance

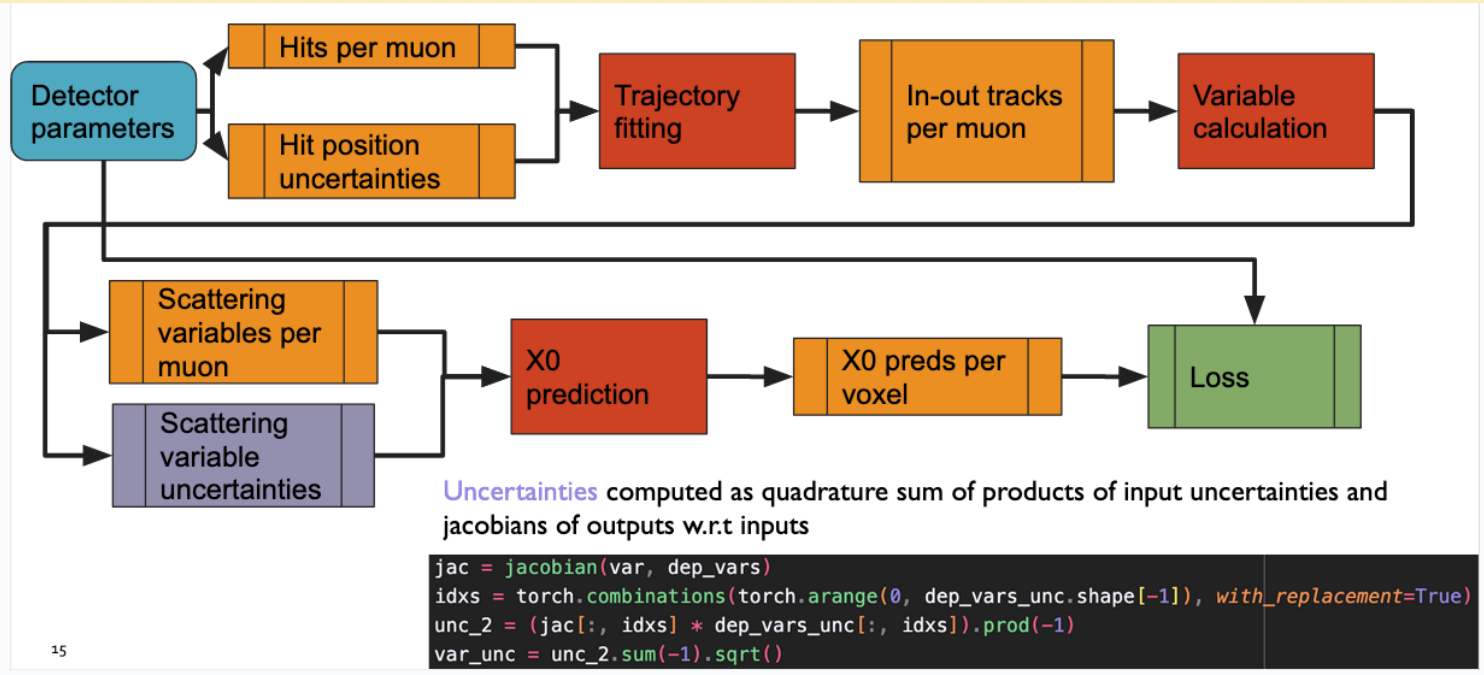


Volume inference

- The main boundary is that inference algorithm must be differentiable
- Basic approach of inverting scatter model to compute X_0 is highly biased
- Maybe a task-specific summary statistic would work better

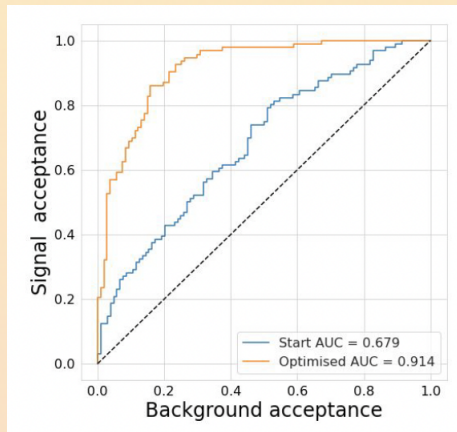
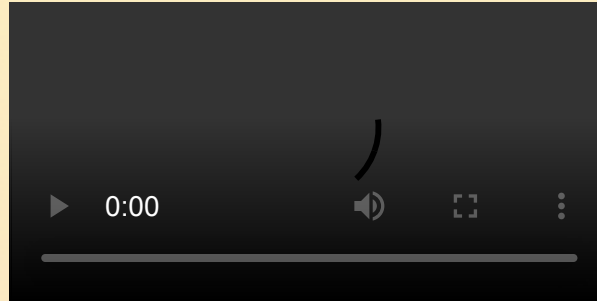


Pipeline



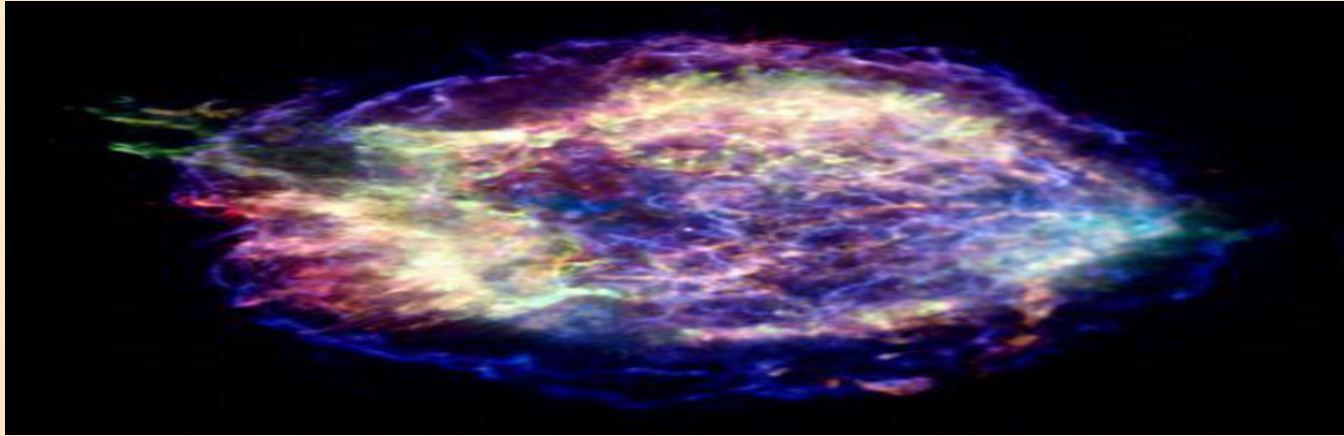
Encouraging results

- Inference via low-dim summaries
- E.g. identify uranium in container



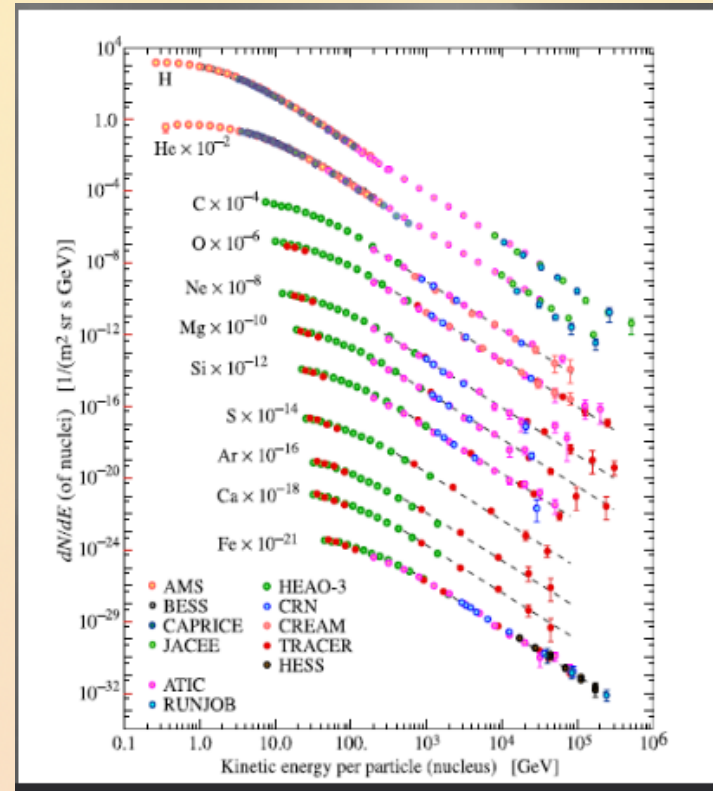
Cosmic rays from supernovae

- High-energy primary cosmic rays produced by supernovae



Primary cosmic rays

- 89% hydrogen nuclei (protons)
- the rest is Helium, Carbon, Oxygen, and other less abundant elements



Muons from cosmic rays

- Cosmic ray muons produced when primary cosmic rays impact with earth's atmosphere
- 1990, [Gaisser](#) formula for flux at sea level

$$\frac{dI_\mu}{dE_\mu} = 0.14 \left(\frac{E_\mu}{\text{GeV}} \right)^{-2.7} \left[\frac{1}{1 + \frac{1.1 E_\mu \cos \theta}{115 \text{ GeV}}} + \frac{0.054}{1 + \frac{1.1 E_\mu \cos \theta}{850 \text{ GeV}}} \right]$$

- Valid only if:
 - Earth curvature negligible ($\theta < 70 \text{ deg}$)
 - Muon decay negligible ($E_\mu > 100/\cos \theta \text{ GeV}$)

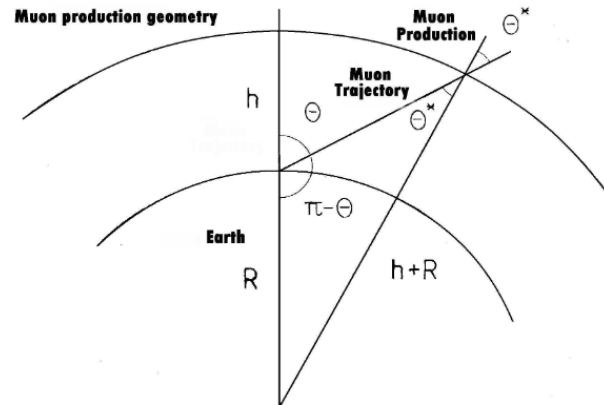


Fig. 1. The relation of the observed zenith angle of muons, θ^* , to the zenith angle at the muon production point in the atmosphere, θ . R is the radius of the Earth. Adopted from [\[3\]](#) [\[4\]](#)

Muons from cosmic rays

- Improved formula by [Guan et al. \(2015\)](#)
- Account for Earth's curvature
- θ at ground and θ^* at production differ

$$\cos \theta^* = \sqrt{\frac{(\cos \theta)^2 + P_1^2 + P_2(\cos \theta)^{P_3} + P_4(\cos \theta)^{P_5}}{1 + P_1^2 + P_2 + P_4}}$$

P_1	P_2	P_3	P_4	P_5
0.102573	-0.068287	0.958633	0.0407253	0.817285

- Correction at low energies

$$\frac{dI_\mu}{dE_\mu} = 0.14 \left[\frac{E_\mu}{\text{GeV}} \left(1 + \frac{3.64 \text{ GeV}}{E_\mu (\cos \theta^*)^{1.29}} \right) \right]^{-2.7} \times \left[\frac{1}{1 + \frac{1.1 E_\mu \cos \theta^*}{115 \text{ GeV}}} + \frac{0.054}{1 + \frac{1.1 E_\mu \cos \theta^*}{850 \text{ GeV}}} \right]$$

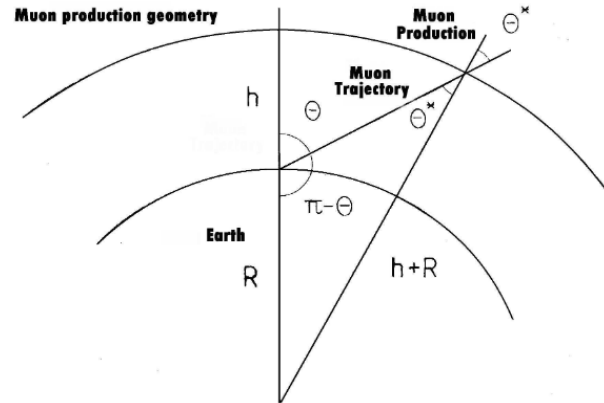
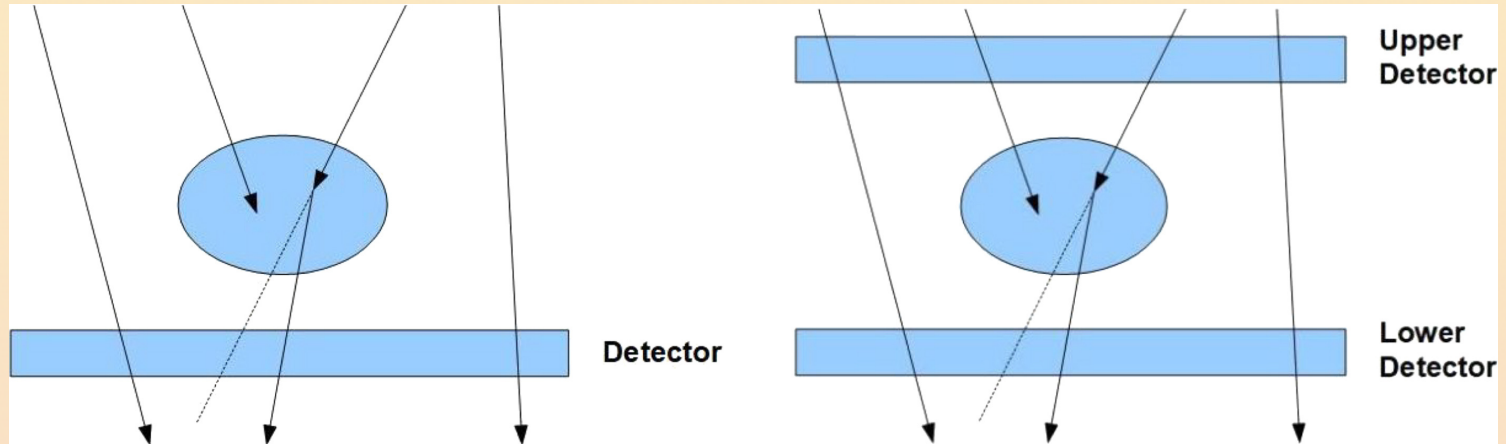


Fig. 1. The relation of the observed zenith angle of muons, θ^* , to the zenith angle at the muon production point in the atmosphere, θ . R is the radius of the Earth. Adopted from [3][4]

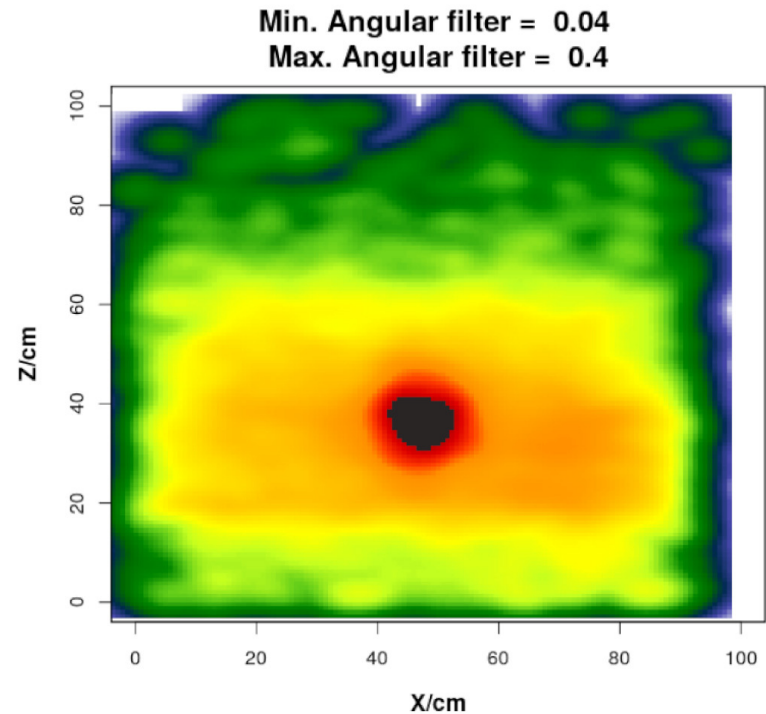
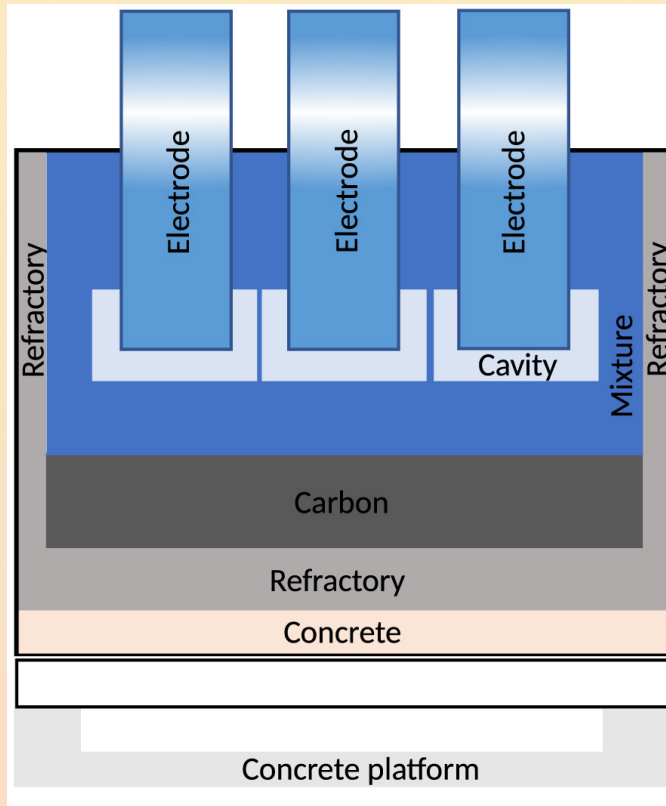
Absorption or scattering

- Absorption: measure missing flux
 - Pyramids, volcanoes...
- Scattering: measure deflection of muon trajectories
 - Containers, furnaces, statues...



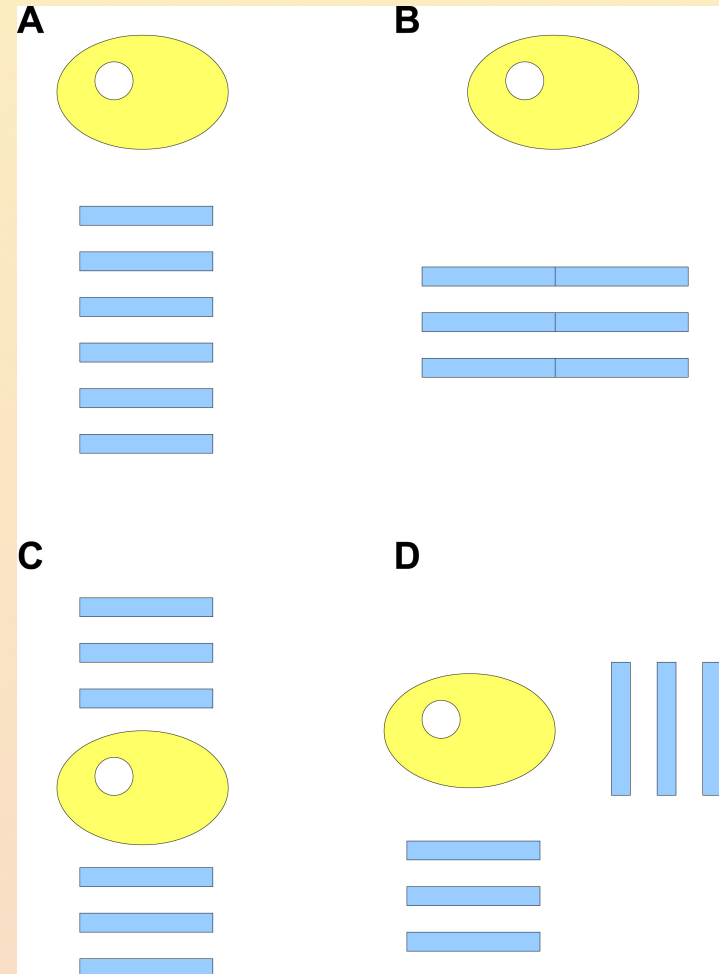
Heart of steel

- Map electrode position in electric furnaces
- Wear of inner walls of pipes



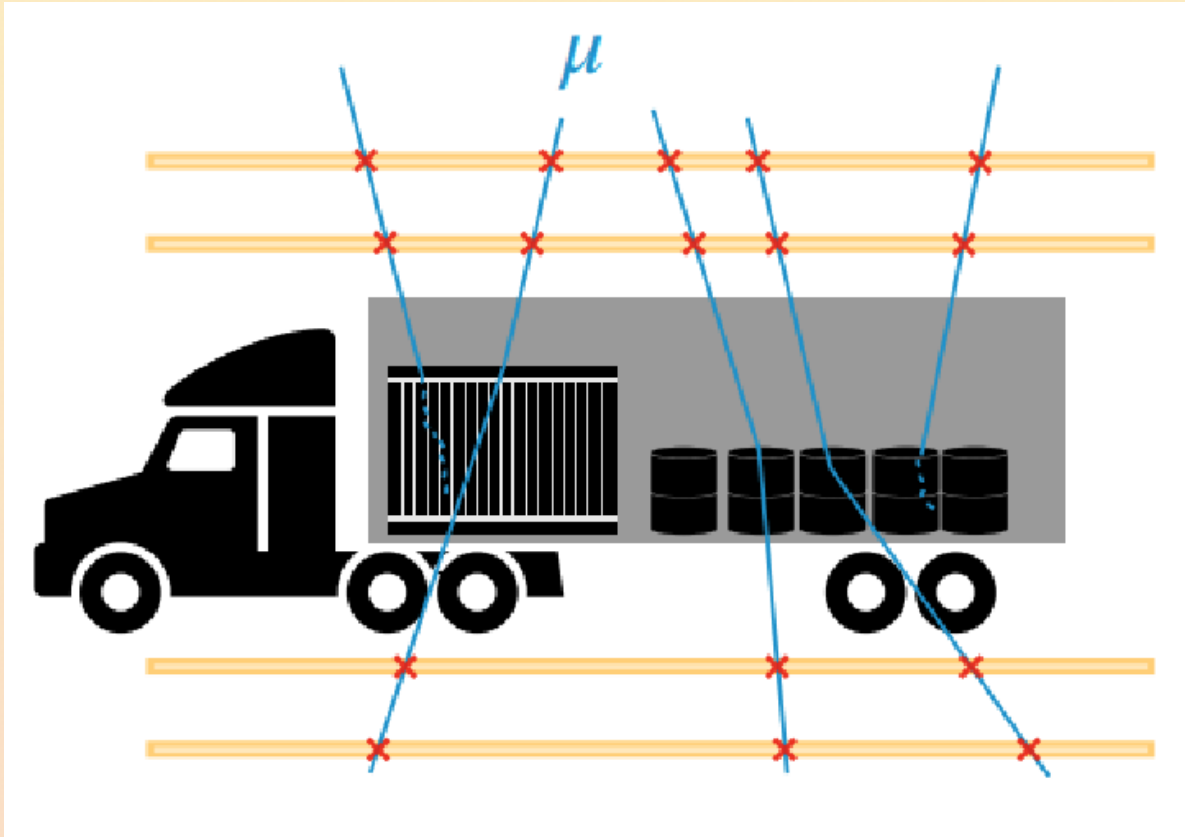
Portable muon detectors

- (A) Maximize angular resolution
- (B) Maximize acceptance
- (C) 3D tomography
- (D) 2D projections in orthogonal planes



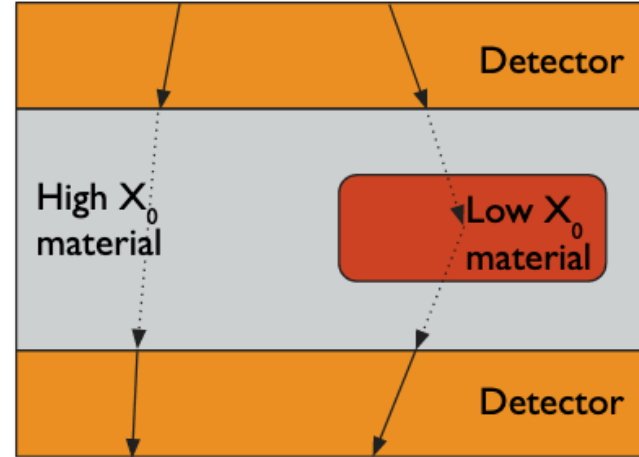
Border control

$$L = \left(1 + e^{k(c_\theta - c_0)}\right) \sum_z \left[w_{imp}(Z) m_{50,\alpha}^{\text{concealed}}[s(Z)] \right]$$



Scanning a passive volume

- Want to infer properties (e.g. 3D map of elemental composition) of unknown volume
 - Shipping container, archeological site, nuclear waste dump, industrial machinery, etc.
- Muons from cosmic rays traverse us all the time
 - On average, 1 muon per cm^2 per minute
 - Change in kinematics provides handle for inference on X_0



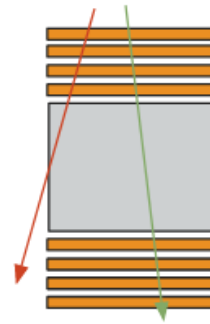
High X_0 = low scattering

Low X_0 = high scattering

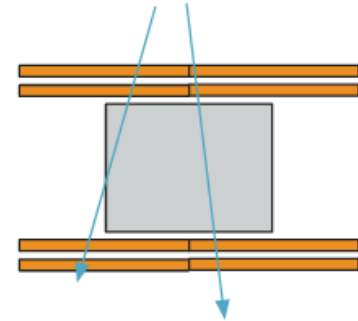
X_0 = average distance between scatterings

Domain knowledge is not enough

- Domain knowledge typically provides heuristics based on proxy objectives
- Will likely have a budget
 - Money, heat, power, positioning of detectors, imaging time...
- Will likely have varying purposes
 - Today want to spot uranium, tomorrow e.g. drugs



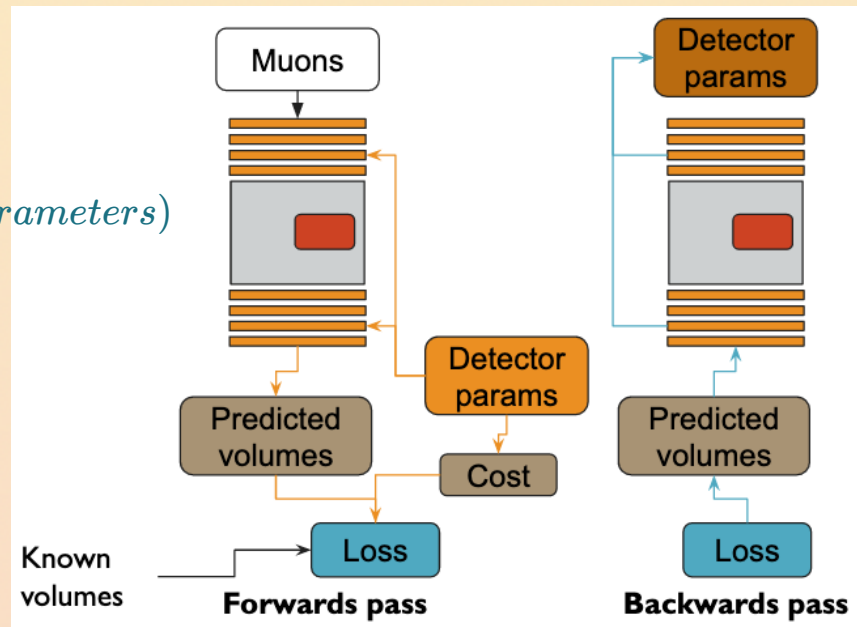
Example 1:
Muons
measured
precisely but
less efficiently



Example 2:
Muons
measured less
precisely but
more
efficiently

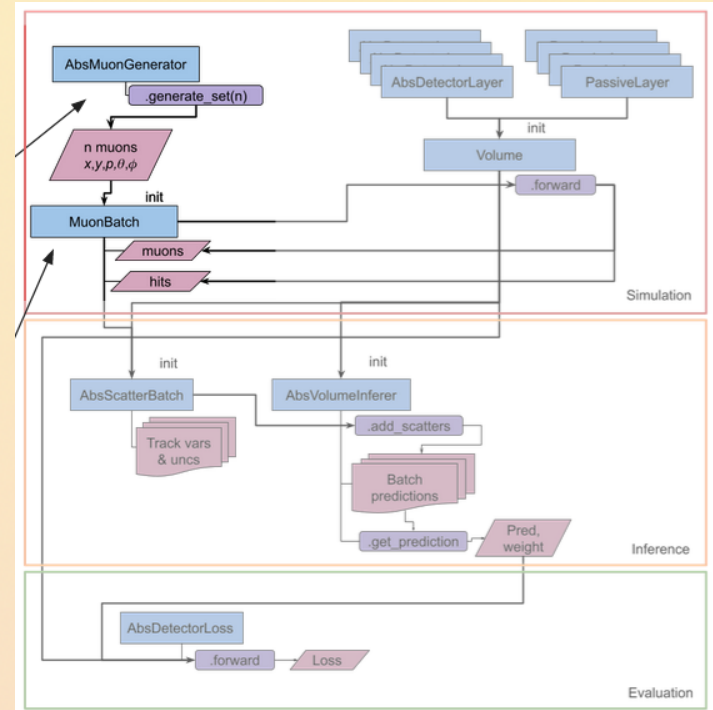
TomOpt

- Differential optimization of muon-tomography detectors (ongoing project)
 - [Giles C. Strong](#), Tommaso Dorigo, Andrea Giammanco, Pietro Vischia, Jan Kieseler, Maxime Lagrange, Mariam Safiedin, Federico Nardi, Anna Bordignon, Haitham Zaraket, Max Lamparth, Federica Fanzago, Oleg Savchenko, Nitesh Sharma
 - Modular design in python, autodiff via PyTorch
- Inference chain as differentiable pipeline
 - Can compute $p(\Delta output | \Delta detector\ parameters)$
- Task as loss function
 - Including target (e.g. prediction uncertainty), costs, constraints
- Backpropagate and optimize as usual
 - Gradient descent



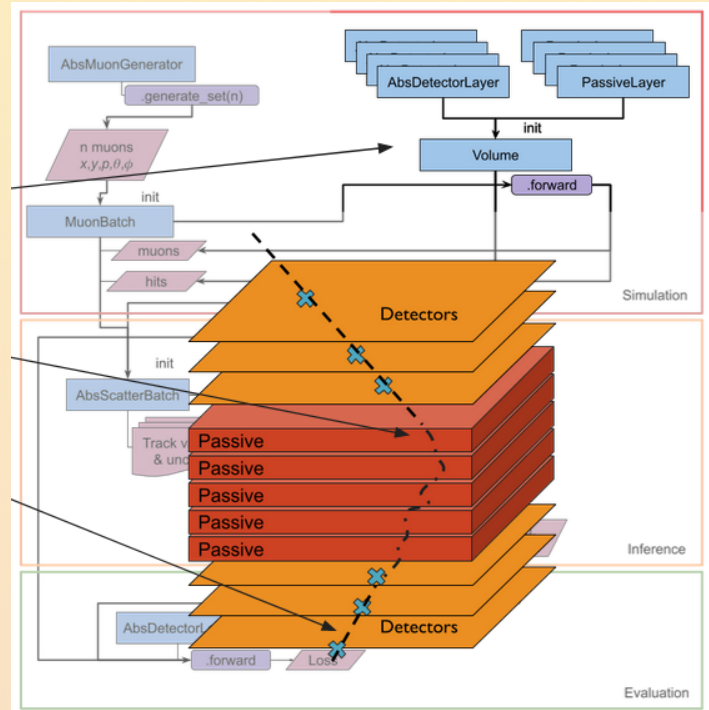
Muon Generation

- Formulas by 2015 and 2016 models
- Tensor of muons (x, y, p, θ, ϕ)
 - θ, p from flux model
 - x, y from ranges
 - ϕ uniformly in $[0, 2\pi]$
- Code handles many muons at once (MuonBatch)
 - Propagate the muon position (can snapshot to track)
 - Scatter ($dx y, d\theta d\phi$) at each step



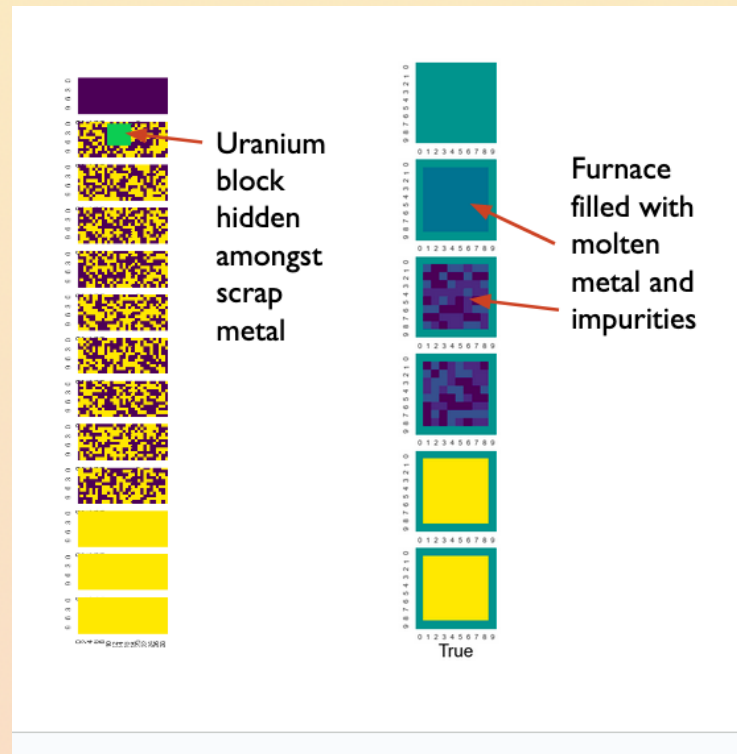
Volume Specification

- Volume made up of stacked layers in z
- Passive layers scatter muon
 - PDG and GEANT models both available
 - Voxelized passive layers (x, y)
- Active layers record muon hits
 - Parameterized efficiency and resolution (cost per m^2 , physics constraints)
 - Budget is a volume attribute, and is assigned to detector layers



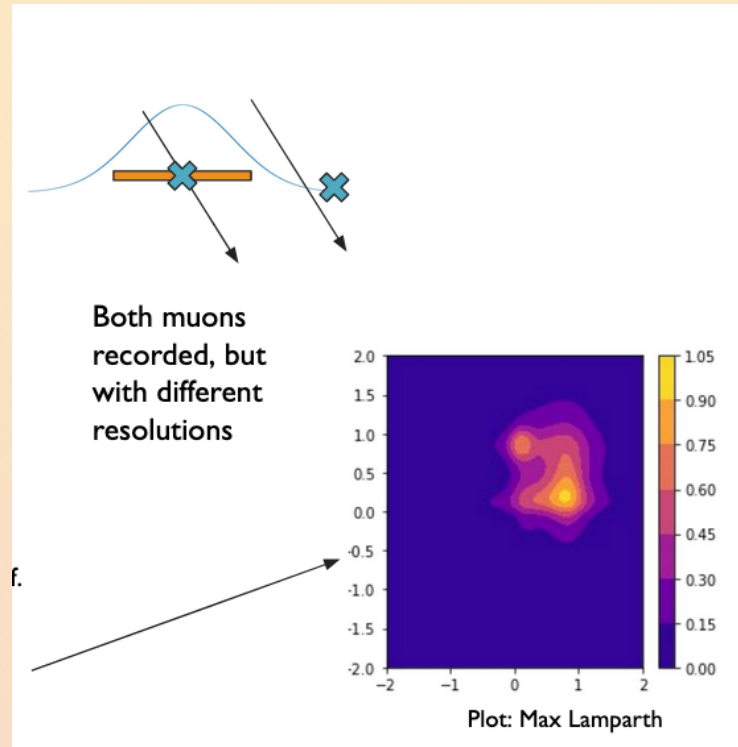
Monte Carlo Truth

- Per each scenario, can build voxelized random volumes
 - Each voxel can be a different material
 - Next: material mixture per voxel



Make muon hits differentiable

- Associate a distribution to resolution and efficiency
 - e.g. Gaussian centered on panel and width equal to panel span
 - p.d.f. of the muon position is now differentiable
- Further generalization: Gaussian Mixture models



From hits to tracks

- Analytic maximum likelihood fit
 - considering uncertainty and efficiency of hits
 - fully differentiable w.r.t. detector parameters
- Provides track parameters and their uncertainties

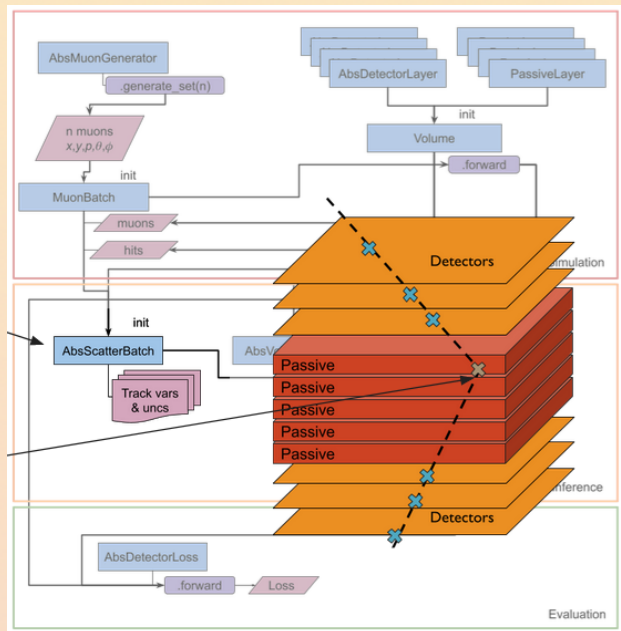
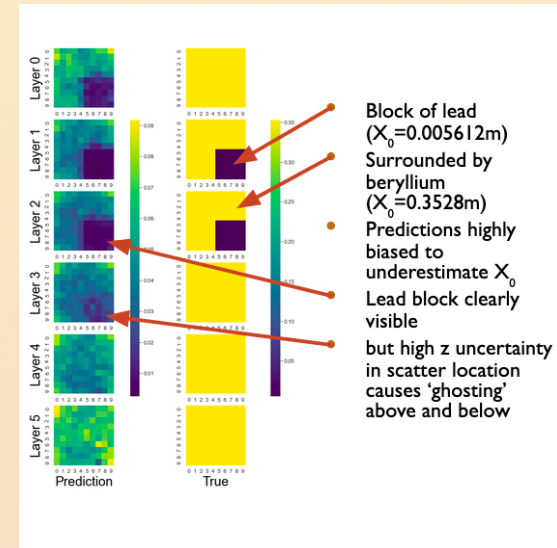


Figure by G.C.Strong

- POCA (POint of Closest Approach)
 - assume one scattering in one point
 - invert model to compute X_0
 - average X_0 per voxel



Infer the volume

- Expectation maximization:
iteratively
 - estimate scatter density based on current estimate of the image
 - update estimate of the image based on the estimated scatter density

- 1) Gather measurements of scattering and momentum for each muon $i=1$ to M : $(\Delta\theta_x, \Delta\theta_y, \Delta x, \Delta y, p_x^2)_i$.
- 2) Estimate geometry of interaction of each muon with each voxel $j=1$ to N : $(L, T)_{ij}$.
- 3) For each muon voxel pair, compute the weight matrix: W_{ij} using (24).
- 4) Initialize the scattering density in each voxel with a guess: $\lambda_{j,old}$.
- 5) Do until (stopping criteria)
 - a) For each muon, compute $\Sigma_{D_i}^{-1}$ using (29) and taking the inverse.
 - b) For each muon voxel pair, compute the conditional expectation terms: S_{ij} using (43).
 - c) Compute $\lambda_{j,new}$ using (38)
 - d) Set $\lambda_{j,old} = \lambda_{j,new}$ for all voxels.
- 6) End do

Summary of the EM algorithm for muon tomography. Schultz et al. 2007

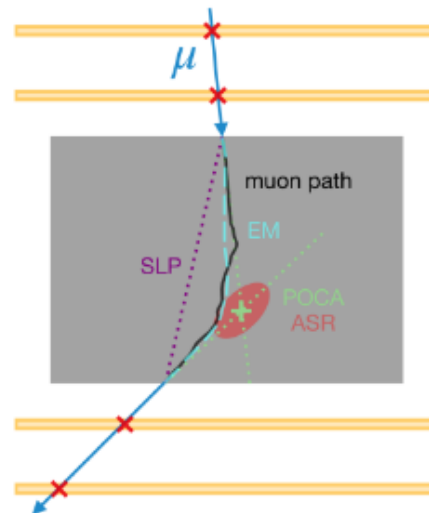
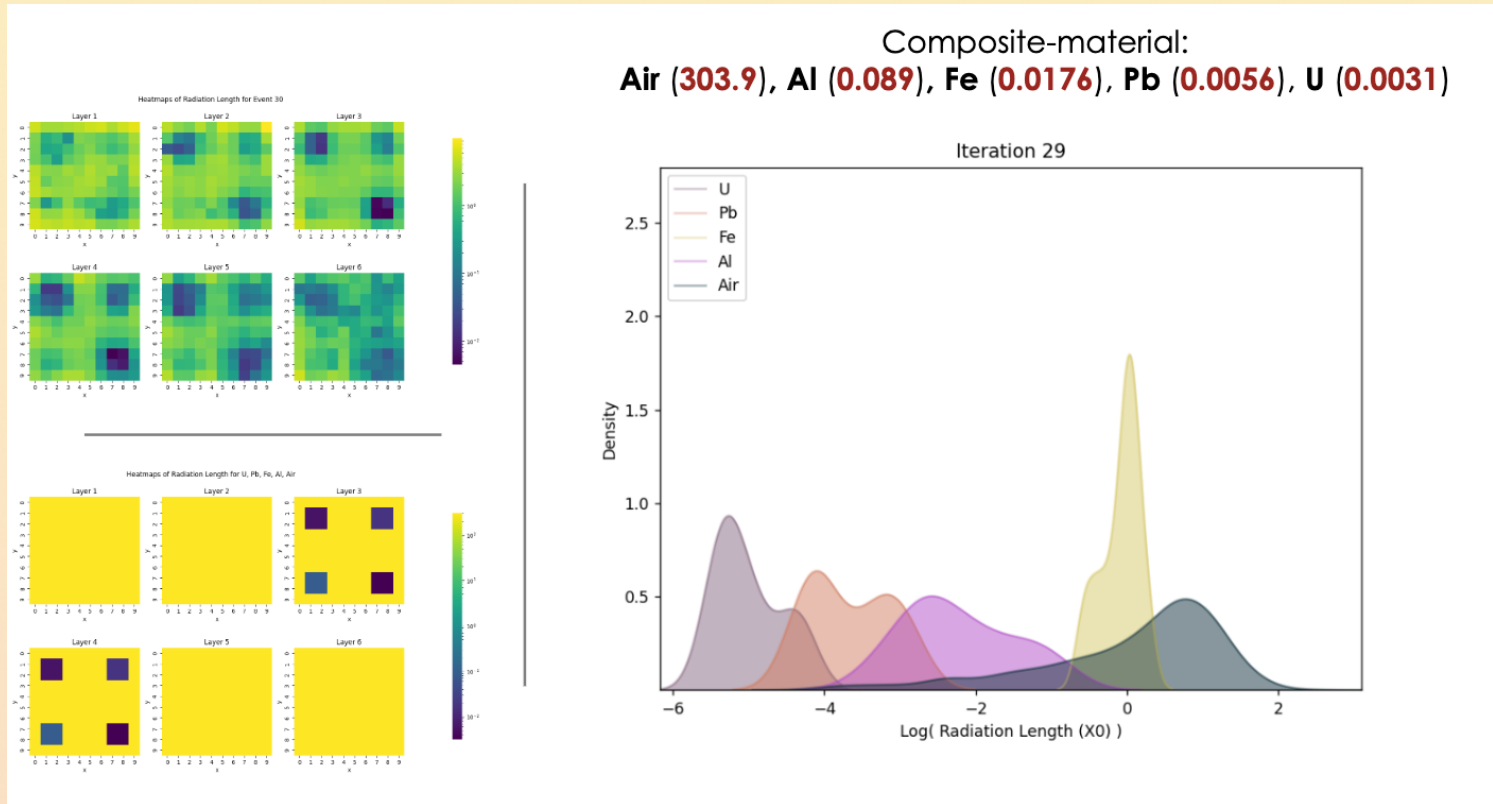


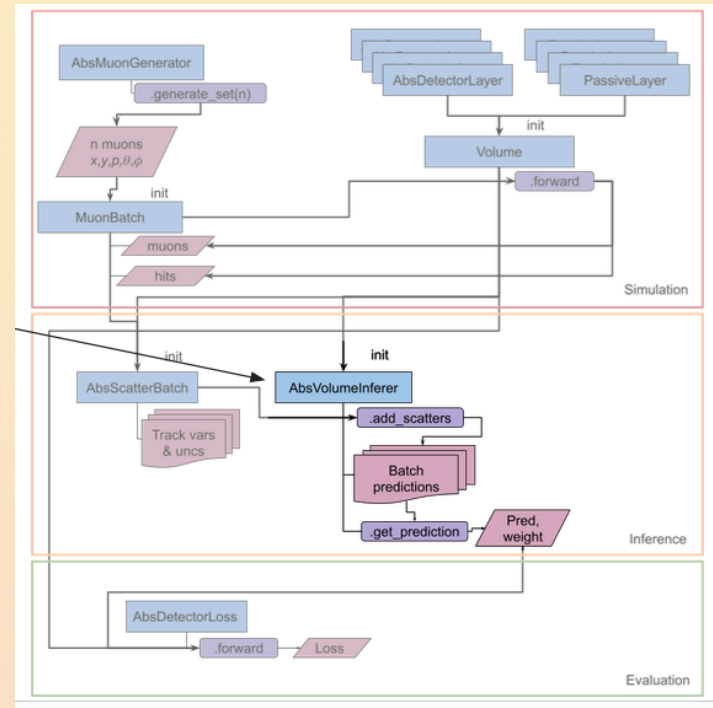
Figure 2: Illustration of commonly used algorithms for reconstructing images in muon scattering tomography. Barnes et al. 2023

EM preliminary performance



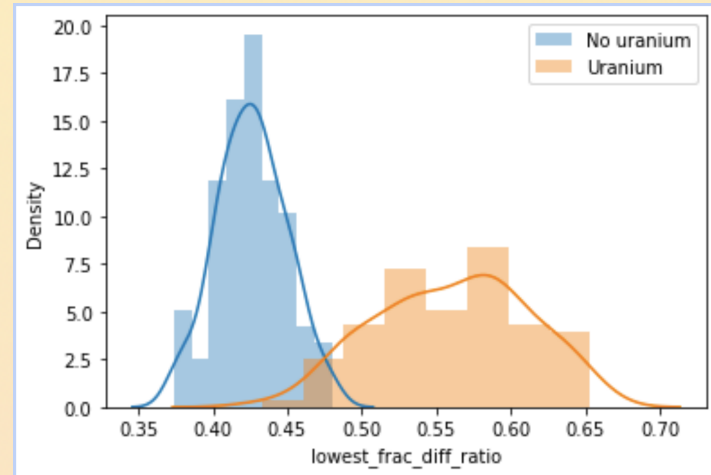
Volume inference

- The main boundary is that inference algorithm must be differentiable
- Basic approach of inverting scatter model to compute X_0 is highly biased
- Maybe a task-specific summary statistic would work better



It's all about summaries

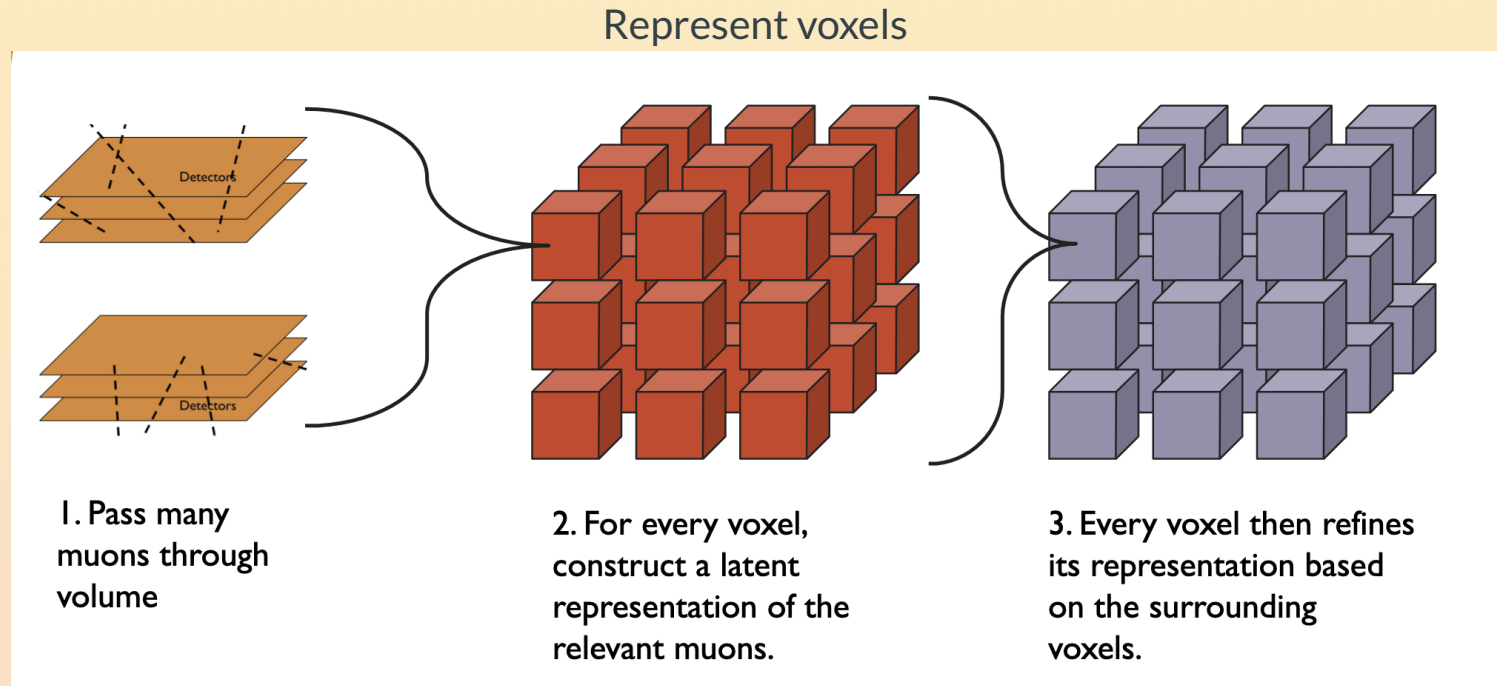
- Promising performance
- Muon track quantities differentiable
 - can compute uncertainties due to spatial resolution
 - useful for aggregating
- X_0 predictions biased and inaccurate
 - Uranium is denser than other materials
 - Under U hypothesis, large difference between mean of the n lowest X_0 voxels and mean X_0 of remaining voxels
 - One-dimensional summary statistic!



- Ultimately, the summary **can be learned** with **graph neural networks**)

Diff. representation with GravNet

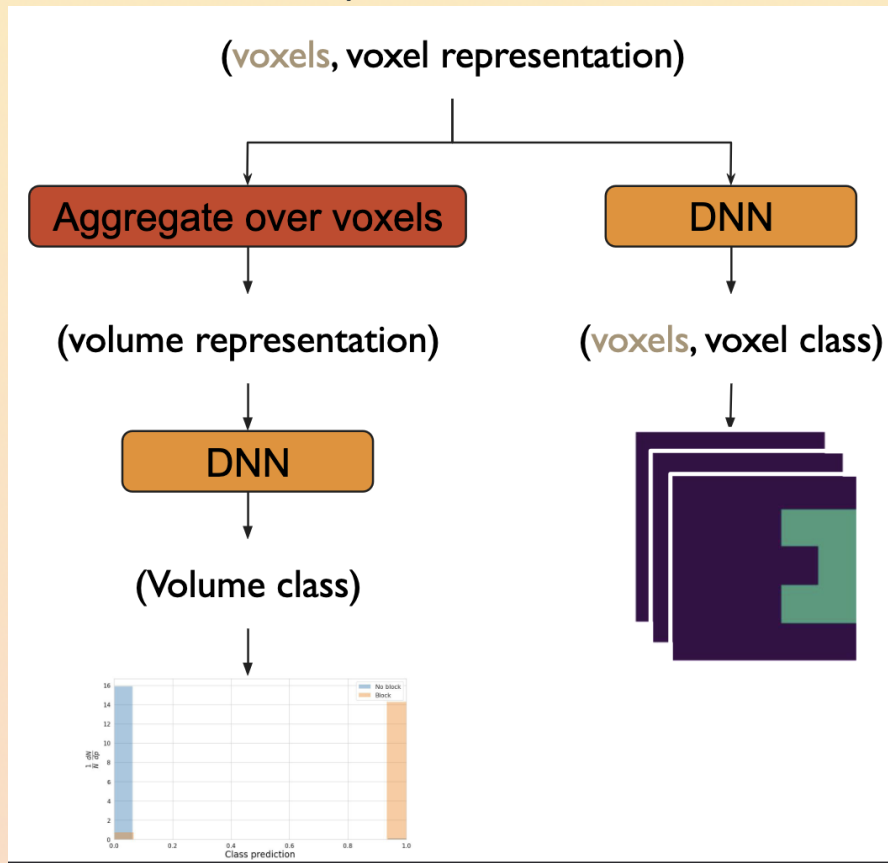
- Use [Quasim et al](#): distance-weighted representations, information-sharing along learnable connections



Diff. representation with GravNet

- Use [Quasim et al](#): distance-weighted representations, information-sharing along learnable connections

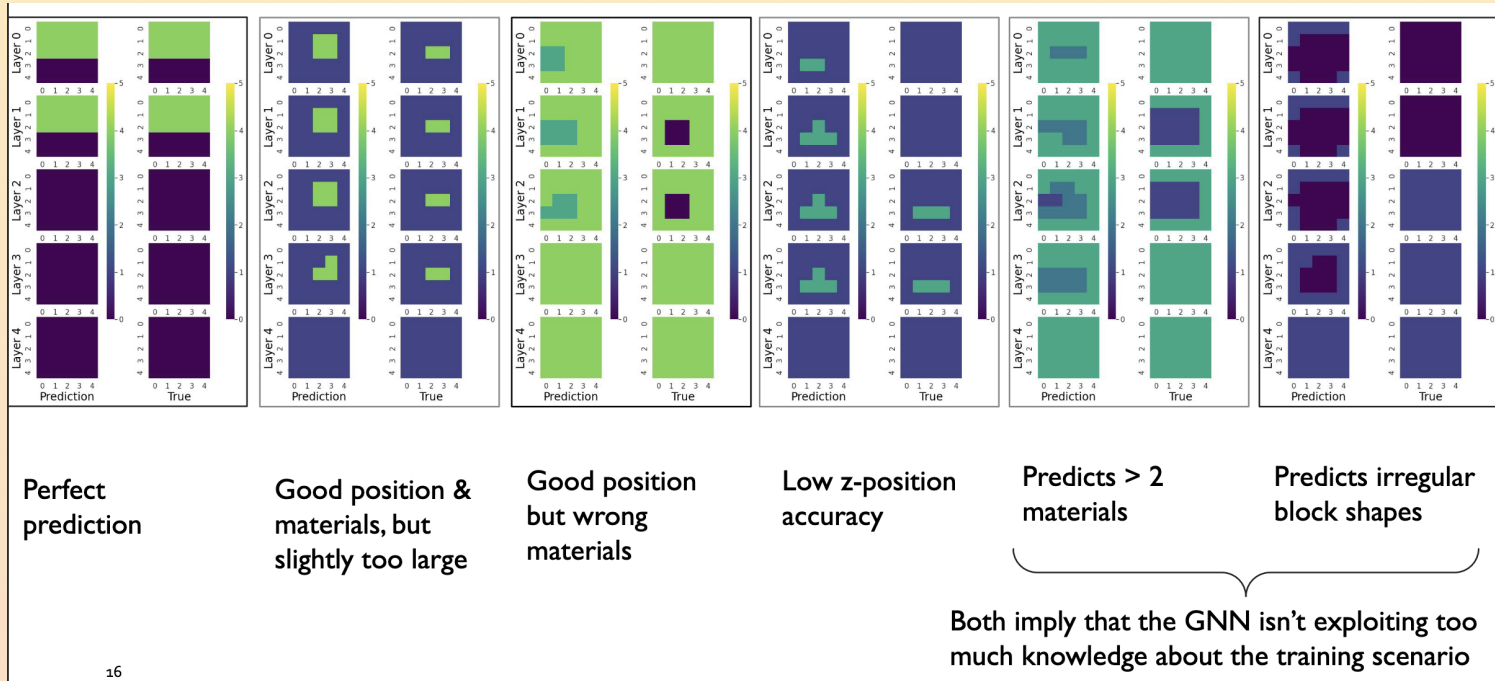
Predict volume (binary classification: there or not there)



Diff. representation with GravNet

- Use [Quasim et al](#): distance-weighted representations, information-sharing along learnable connections

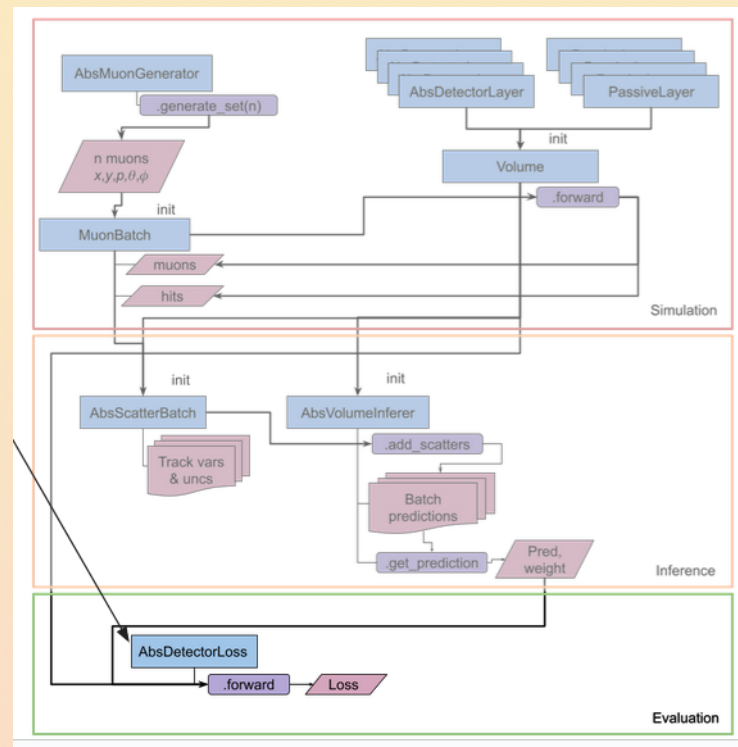
Imaging behaviour



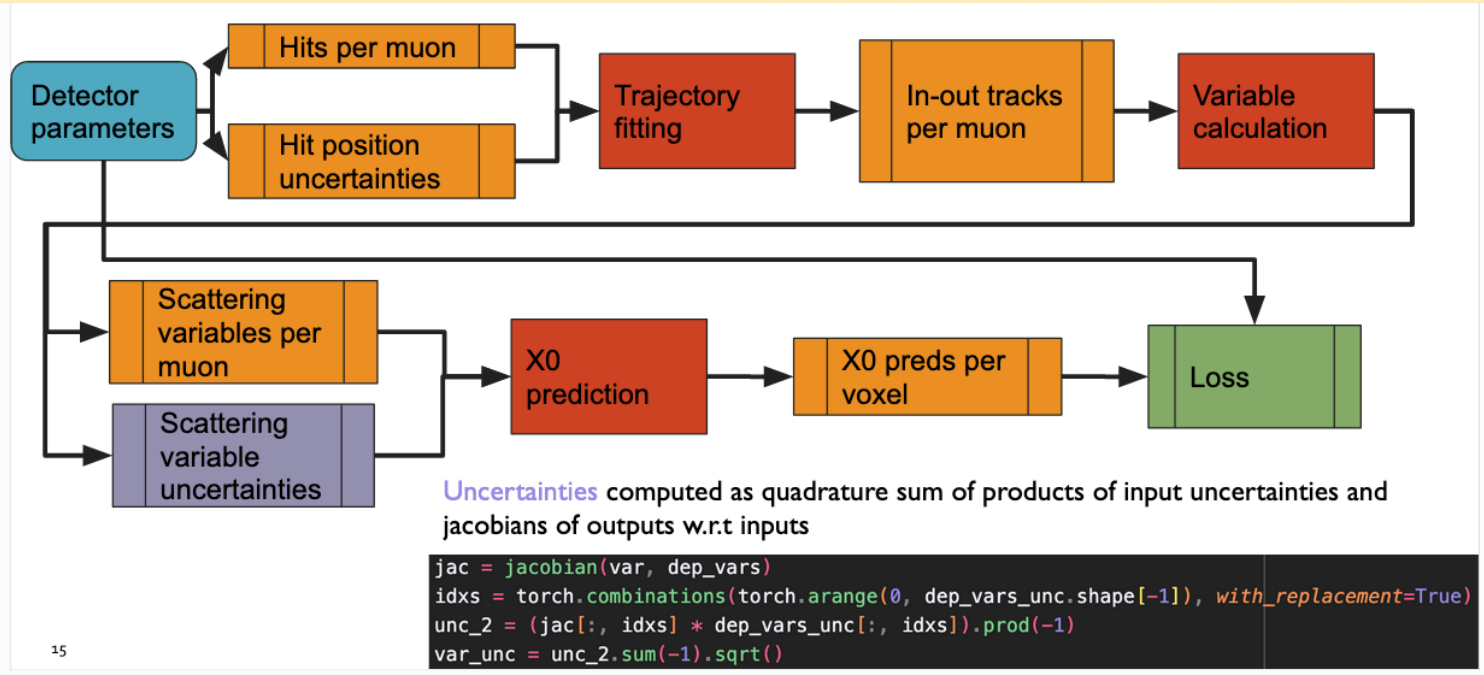
16

Optimization

- Regular gradient descent of a loss function
 - account for cost of the detector and other constraints
 - standard optimisers to update detector parameters



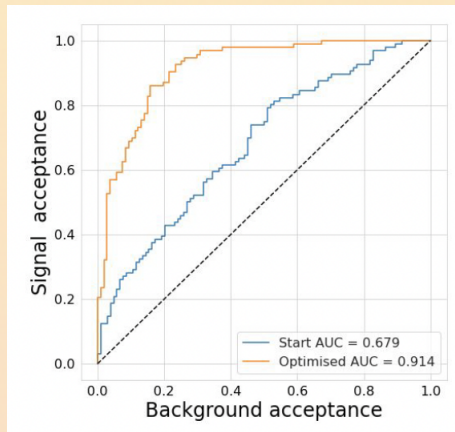
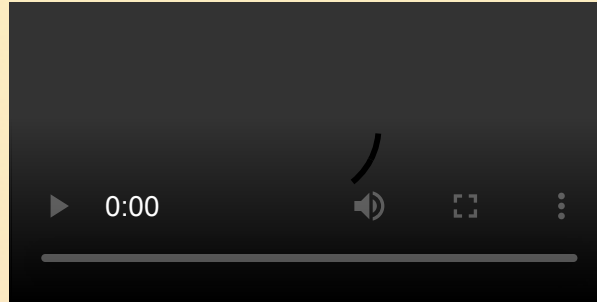
Pipeline



15

Encouraging results

- Inference via low-dim summaries
- E.g. identify uranium in container



The MODE Collaboration

<https://mode-collaboration.github.io/>

- Joint effort (created 11.2020) of particle physicists, nuclear physicists, astrophysicists, and computer scientists

At INFN and Università di Padova Dr. **Tommaso Dorigo**, Dr. **Pablo De Castro Manzano**, Dr. **Federica Fanzago**, Dr. **Lukas Layer**, Dr. **Giles Strong**, Dr. **Mia Tosi**, and Dr. **Hevjin Yarar**
At Université catholique de Louvain Dr. **Andrea Giammanco**, Prof. **Christophe Delaere**, Mr. **Maxime Lagrange**, and Dr. **Pietro Vischia**
At Université Clermont Auvergne, Prof. **Julien Donini**, and Mr. **Federico Nardi** (joint with Università di Padova)
At the Higher School of Economics of Moscow, Prof. **Andrey Ustyuzhanin**, Dr. **Alexey Boldyrev**, Dr. **Denis Derkach**, and Dr. **Fedor Ratnikov**
At the Instituto de Física de Cantabria, Dr. **Pablo Martínez Ruíz del Árbol**
At CERN, Dr. **Jan Kieseler**, Dr. **Sofia Vallecrosia**
At University of Oxford Dr. **Atilim Gunes Baydin**
At New York University Prof. **Kyle Cranmer**
At Université de Liège Prof. **Gilles Louppe**
At GSI/FAIR Dr. **Anastasios Belias**
At Rutgers University Dr. **Claudius Krause**
At Uppsala Universitet Prof. **Christian Glaser**
At TU-München, Prof. **Lukas Heinrich** and Mr. **Max Lamparth**
At Durham University Dr. **Patrick Stowell**
At Lebanese University Prof. **Haitham Zaraket**
At Technische Universität Kaiserslautern Mr. **Max Ahle**, Prof. **Nicolas Gauger**, Dr. **Lisa Kusch** At Technische Universität Worms Prof. **Ralf Keidel**
At Princeton University Prof. **Peter Elmer**
At University of Washington Prof. **Gordon Watts**
At SLAC Dr. **Ryan Roussel**

The Scientific Coordinator of the MODE Collaboration is Dr. **Tommaso Dorigo**, INFN-Sezione di Padova
The Steering Board of the MODE Collaboration includes:

- Prof. **Julien Donini**, UCA
- Dr. **Tommaso Dorigo**, INFN-PD
- Dr. **Andrea Giammanco**, UCLouvain
- Dr. **Fedor Ratnikov**, HSE
- Dr. **Pietro Vischia**, UCLouvain

Series of yearly workshop

- [First installment](#) in Louvain-la-Neuve (Belgium)
- [Second installment](#) in Kolymbari (Greece)
 - [37 talks](#), [9 posters](#), one data challenge with prizes, recordings will be online soon
- You are all invited to the [Third MODE Workshop](#), to be held in Princeton (USA)



If you are interested, join us!

<https://mode-collaboration.github.io/>