AI-assisted design of experiments at the frontiers of computation

methods and new perspectives

LIP Seminars, Lisboa, Portugal

Dr. Pietro Vischia

pietro.vischia@cern.ch

@pietrovischia

Universidad de Oviedo

ICTEA

Supported by project
RYC2021- 033305-I
funded by

GOBIERNO DE ESPAÑA

MINISTERIO DE CIENCIA E INNOVACIÓN

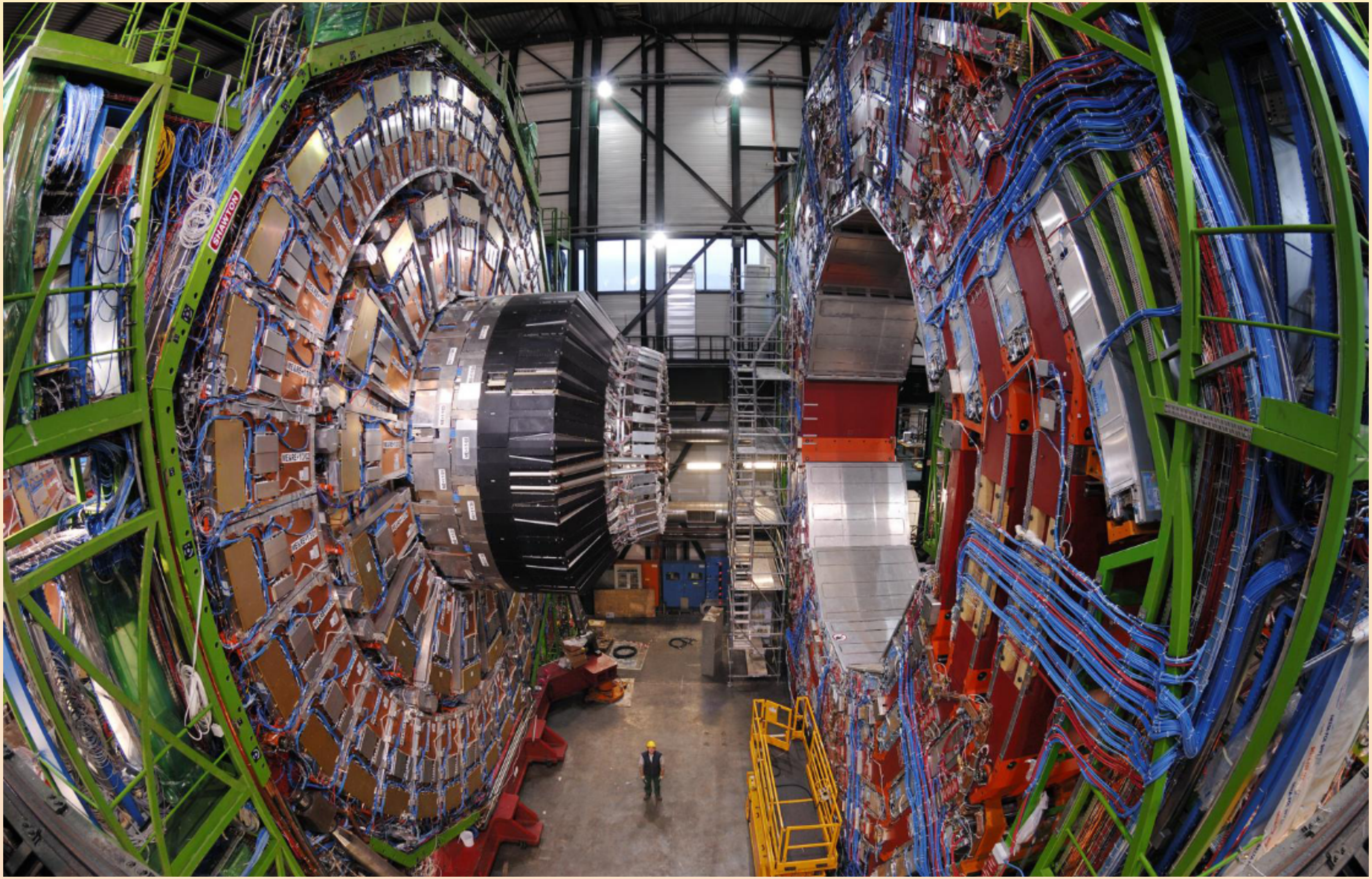
AGENCIA ESTATAL DE INVESTIGACIÓN

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

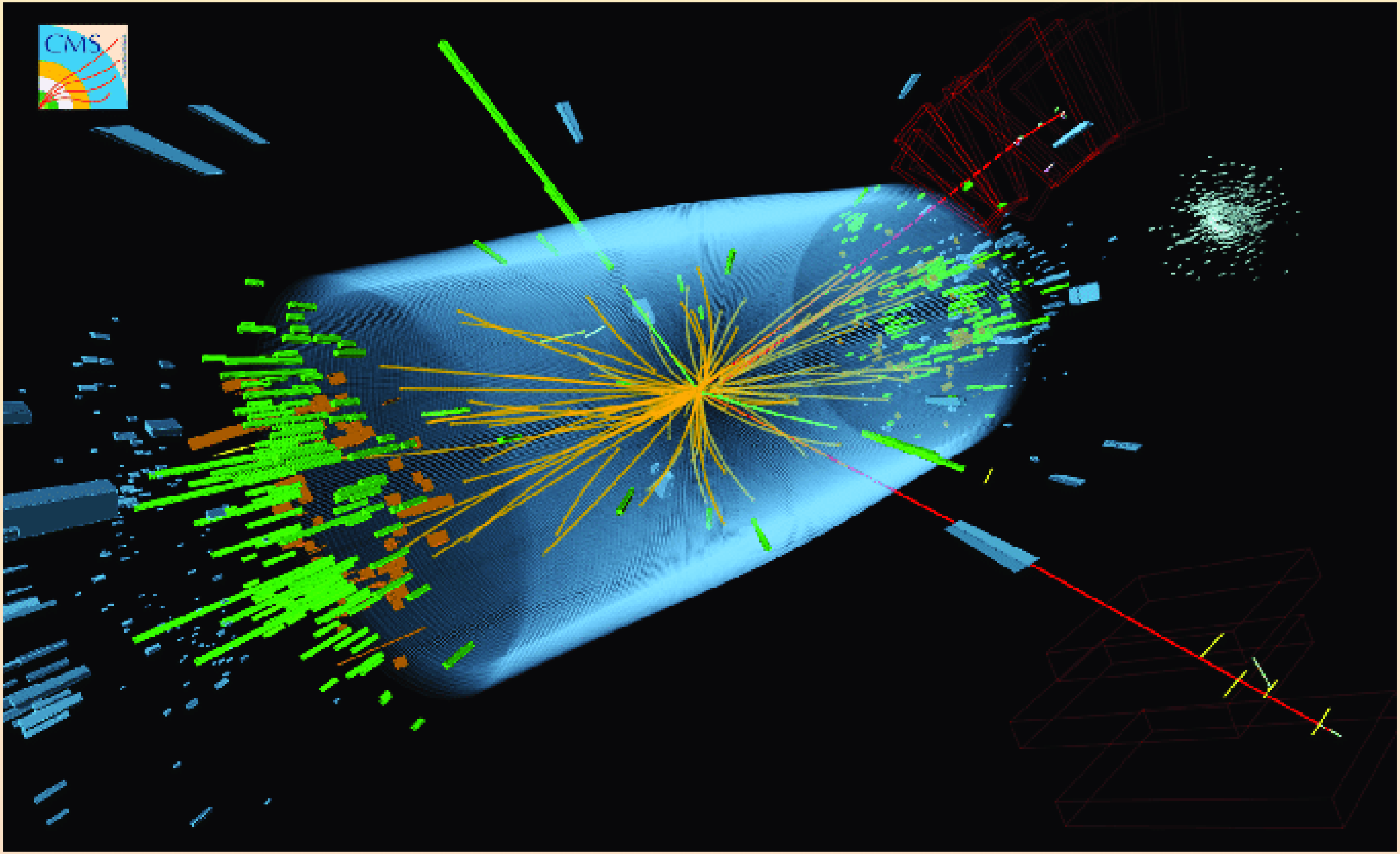
https://www.hep.uniovi.es/vischia/persistent/2024-03-21_AIAssistedDesignAtFrontiersOfComputationAtLIPSeminar_vischia.html

to get the version with working animations

Complex Experiments



Complex Data



Likelihood and information

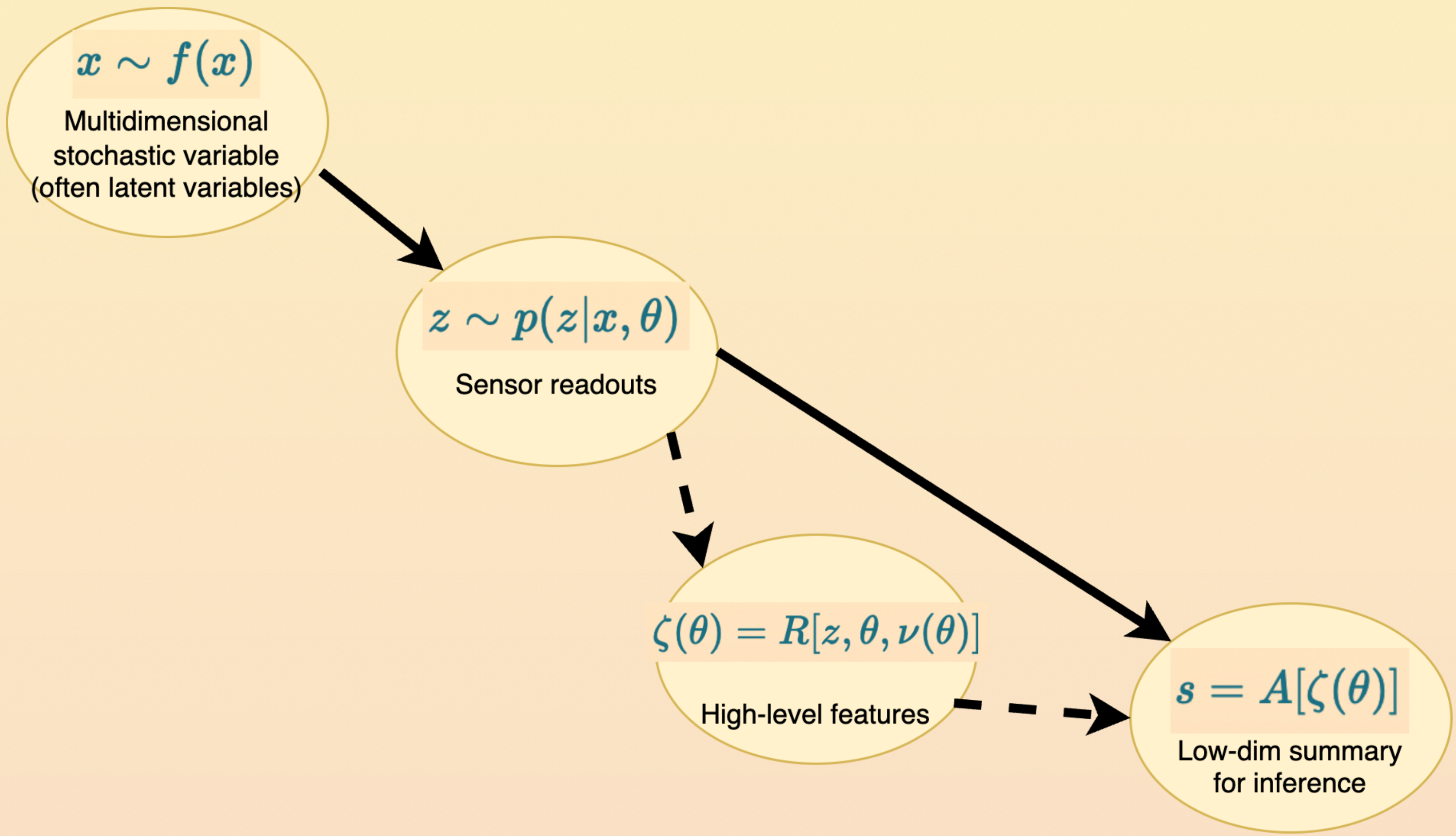
- Data sample X_{obs}

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

- 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

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

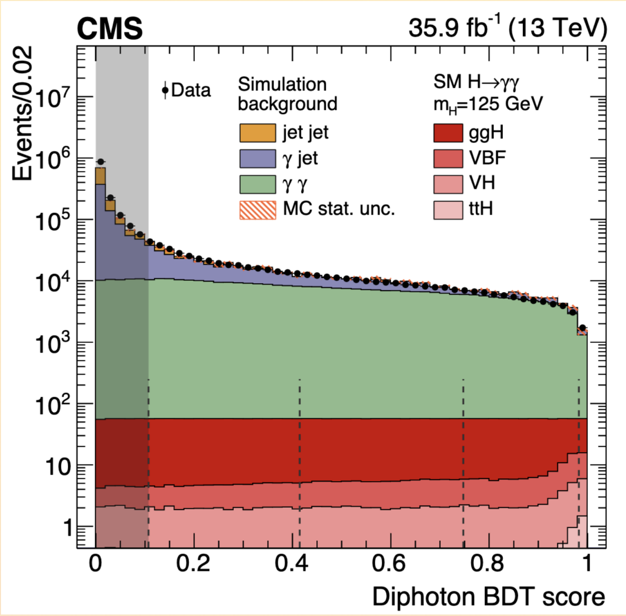
Typical analysis pipeline



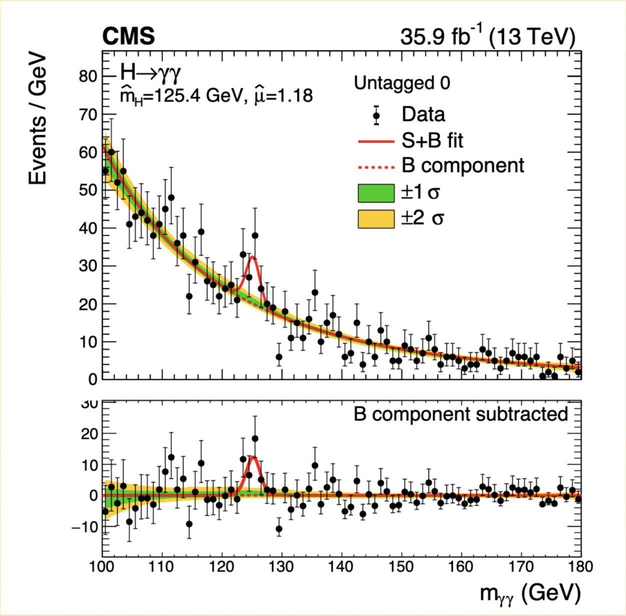
—
Formulas from our white paper

We like low-dim summaries

- Discard uninteresting regions



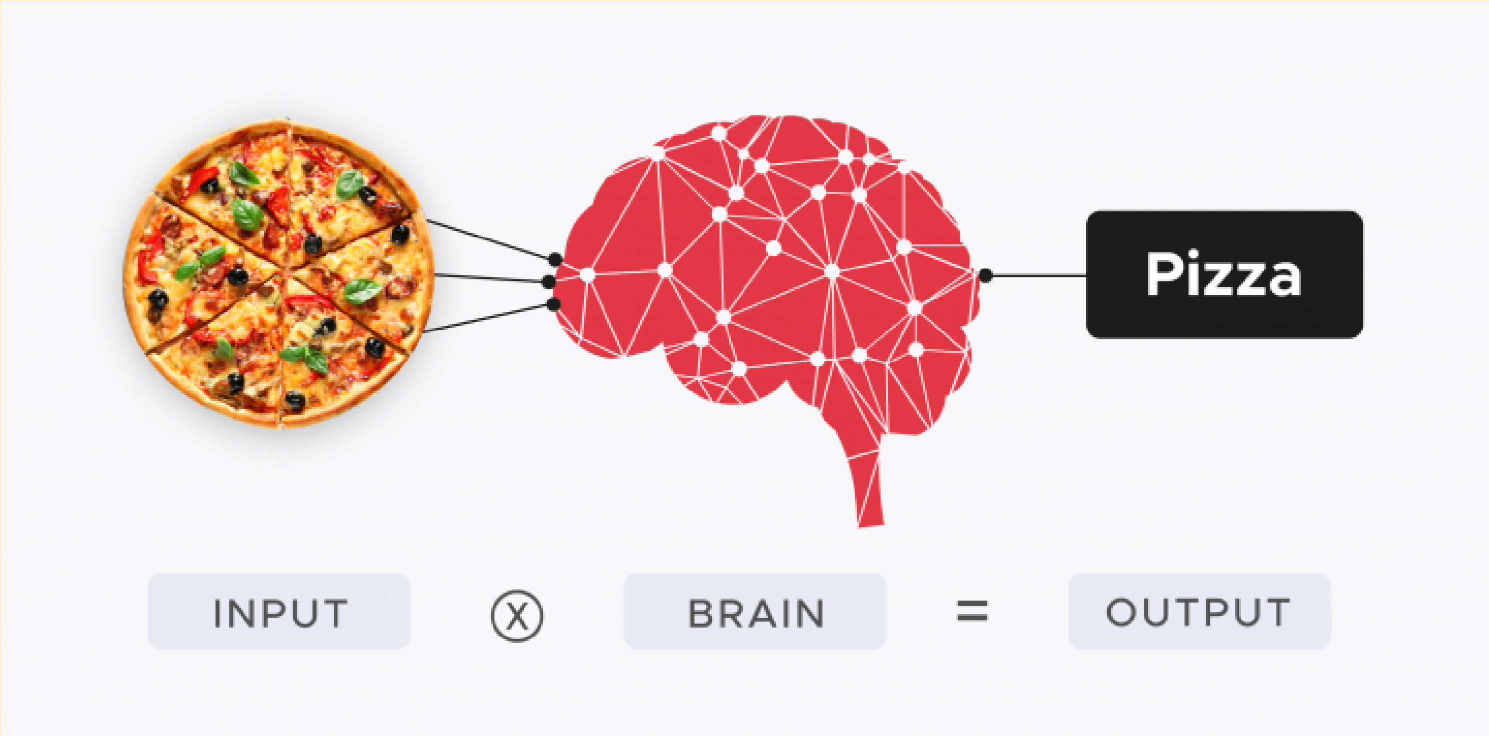
- Physical observable for inference



—

Figure from JHEP 11 (2018) 185

Brain activity...



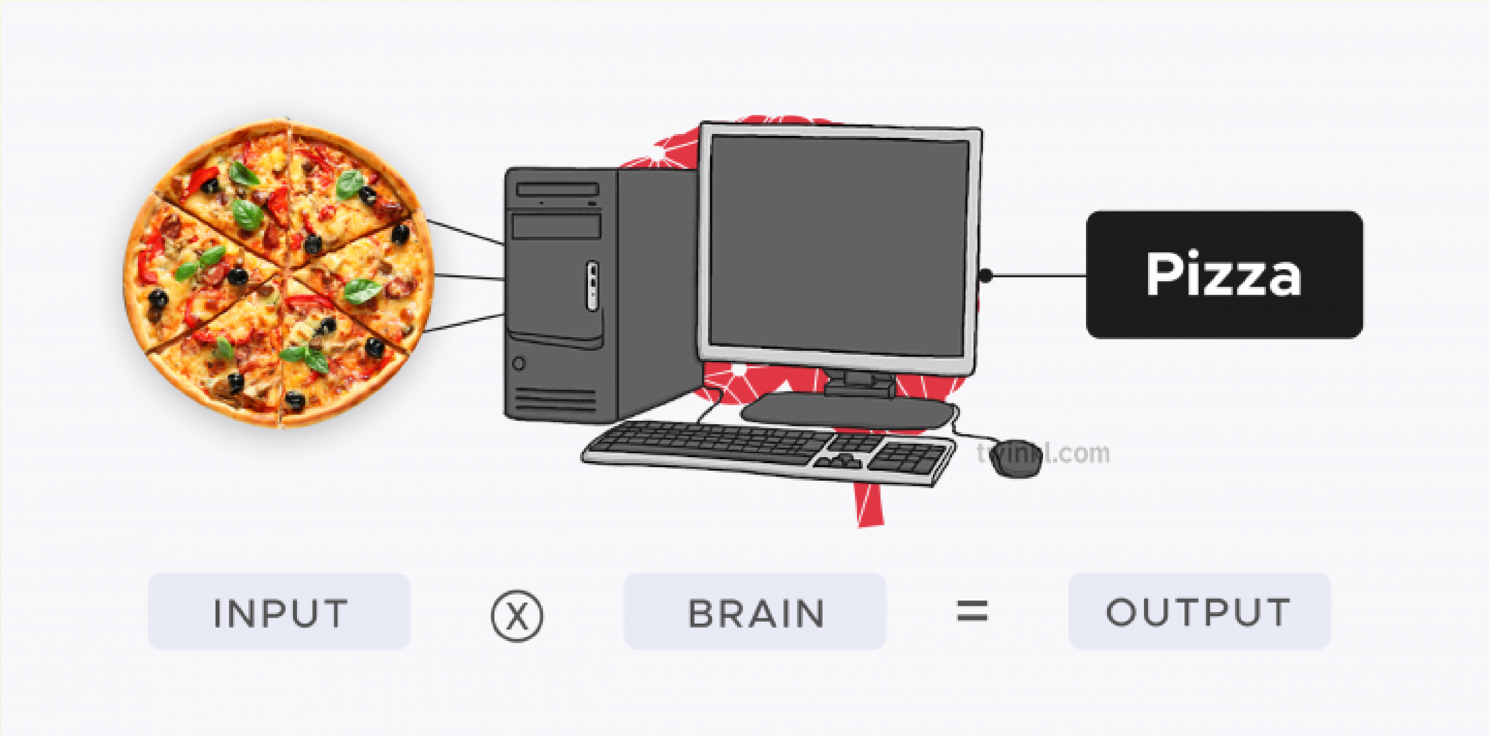
—

Image from zomato

...approximated...

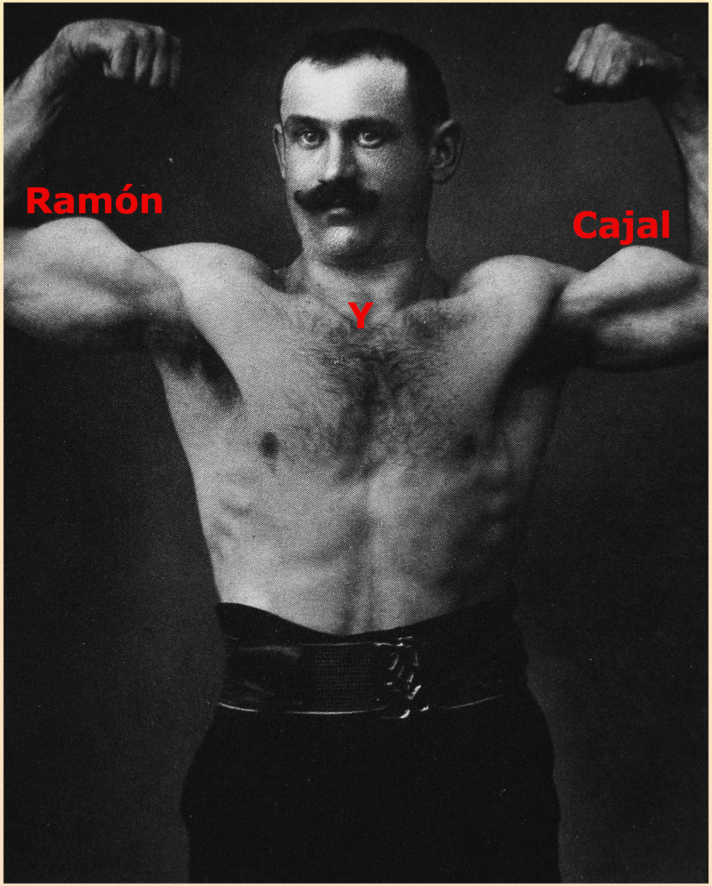
—
Image from zomato, elaborated with image from pixabay

...using computers



—
Image from zomato, elaborated with image from twinkl

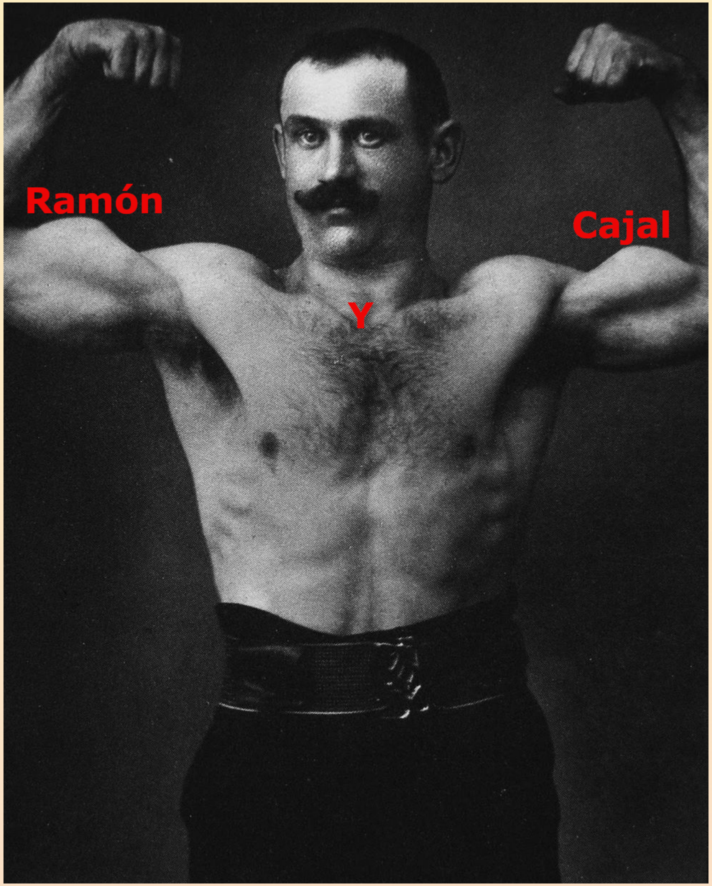
Santiago Ramón y Cajal



—

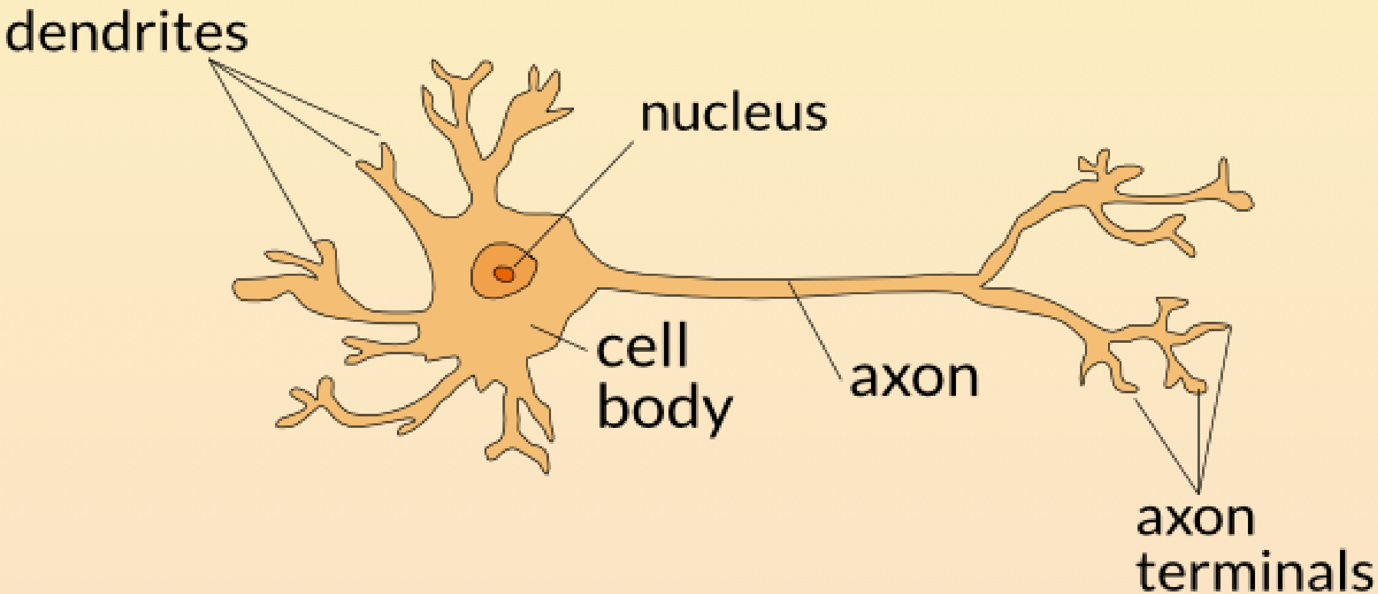
Images from menshealth and The Nobel Prize

Santiago Ramón y Cajal



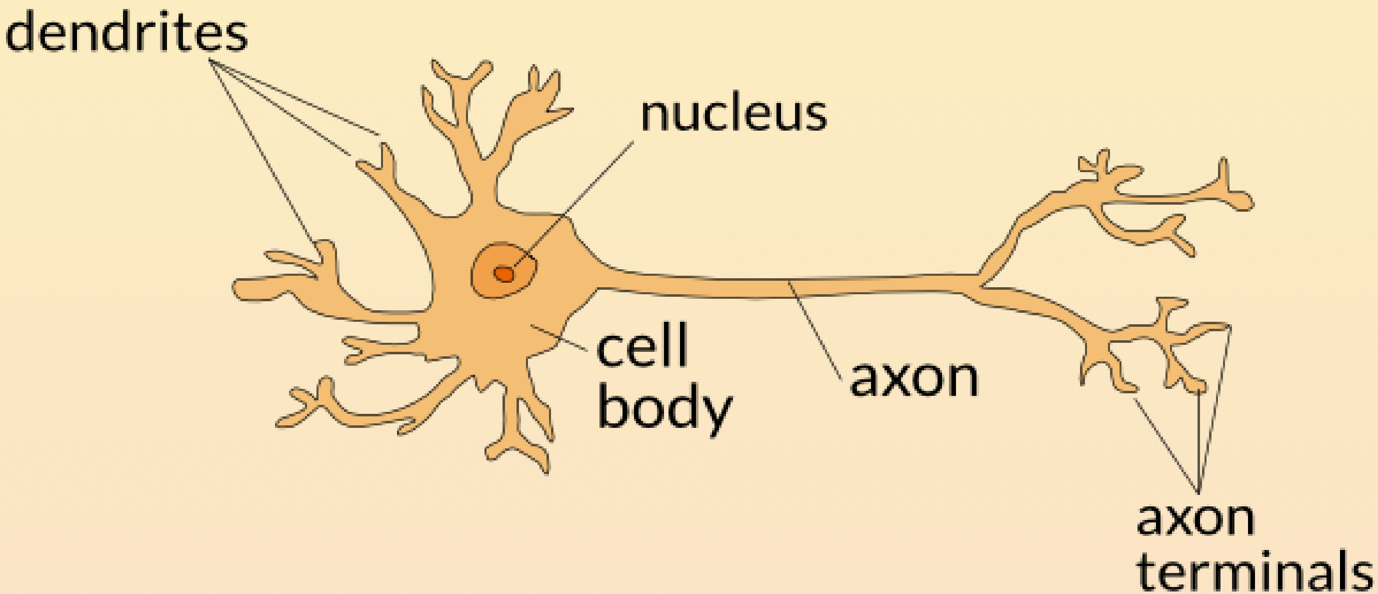
—
Images from menshealth and The Nobel Prize

Real neurons



—
Image from https://appliedgo.net/perceptron/

Real neurons



$$I = C\frac{dV}{dt} + G_{Na}m^3h(V - V_{Na}) + G_K n^4(V - V_K) + G_L(V - V_L)$$

–
Image from https://appliedgo.net/perceptron/

Computationally heavy

—
Animation from giphy.com

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

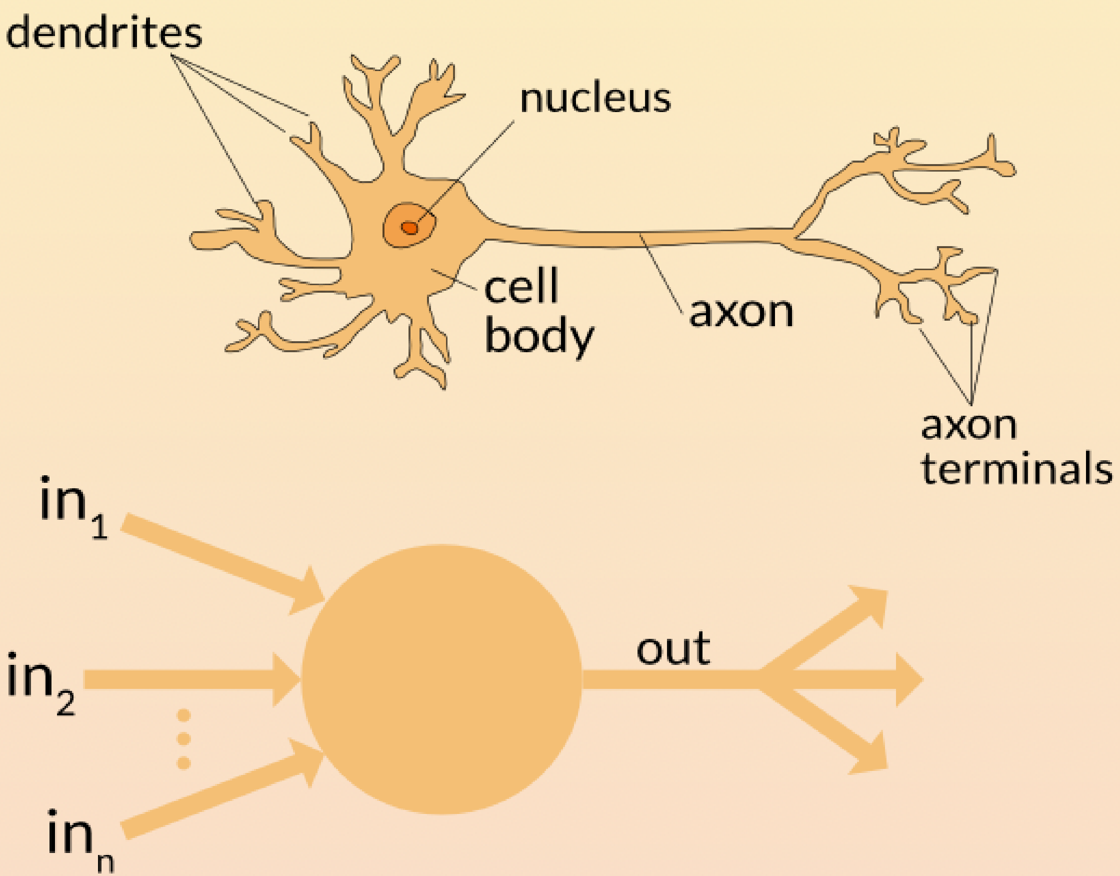
A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY*

■ 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.

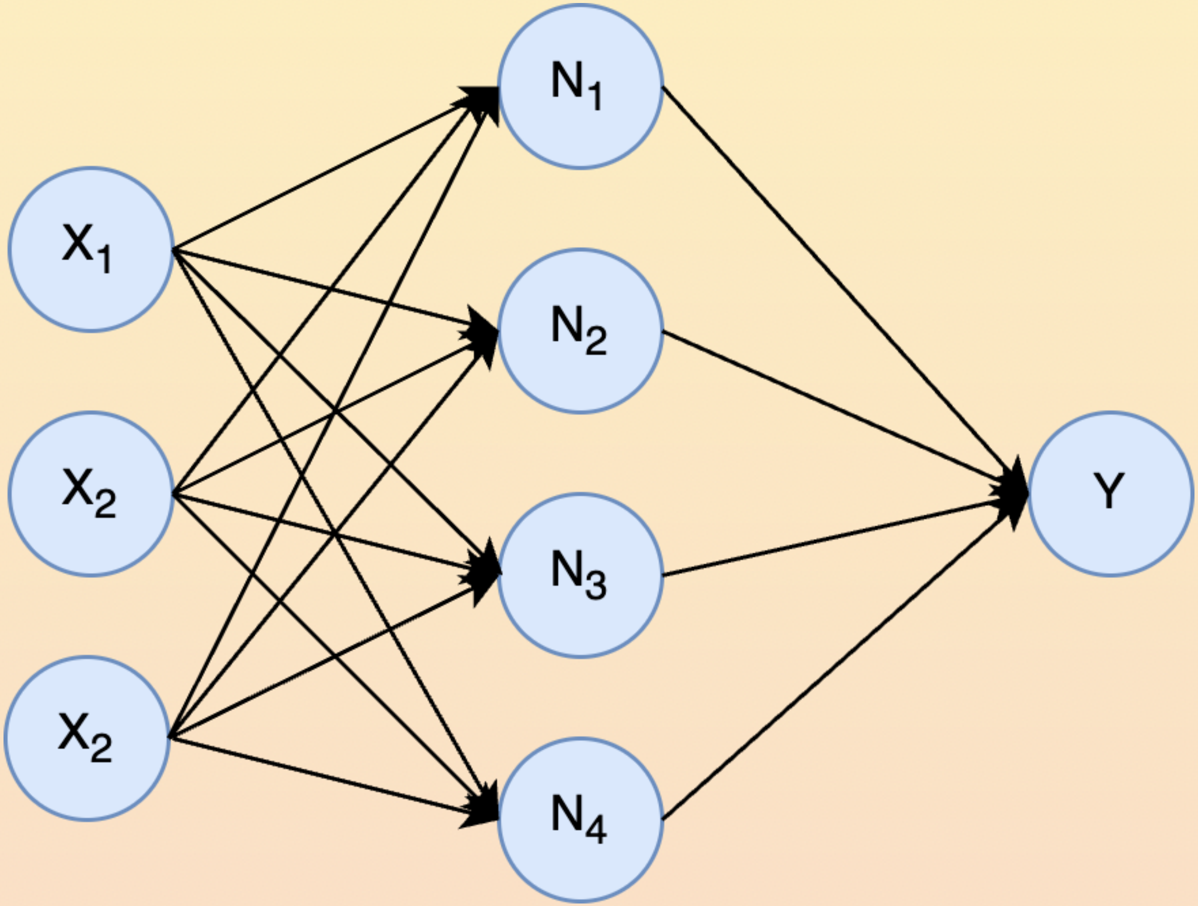
Perceptrons

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



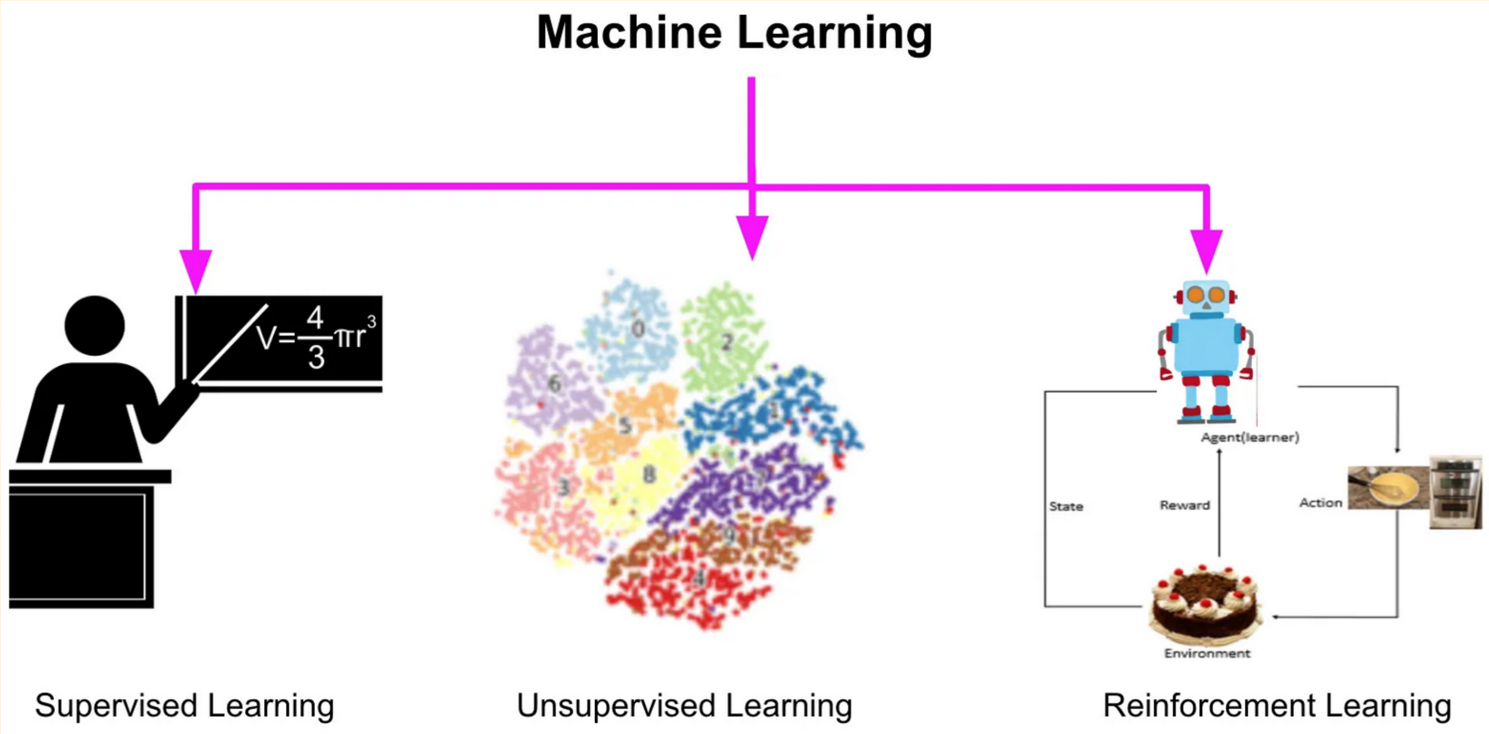
—
Image from https://appliedgo.net/perceptron/

Artificial Neural Networks



—
Illustration (c) P. Vischia, book in preparation

Learn in different ways

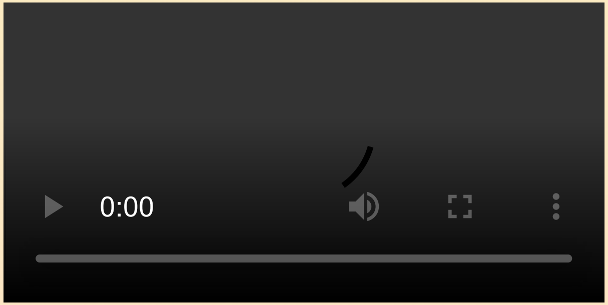


—

Image by Renu Khandelwal

Gradient Descent

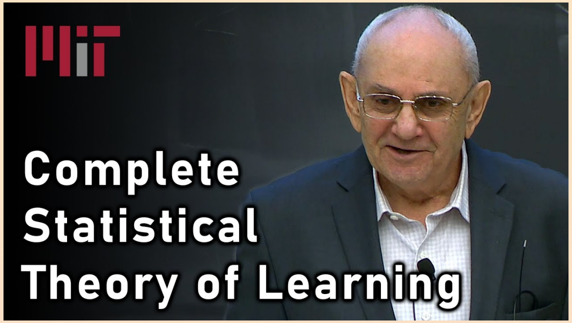
- Optimize/learn by finding the minimum of a function $\mathcal{L} : \mathbb{R}^n \to \mathbb{R}$
- Nonconvex problems: saddle points, manifolds of minima



- Empirical risk minimization

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

- Generalization
  (for learning problems)



—

Vapnik's image from youtube, Gradient descent animation by Alec Radford

Backpropagation

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

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

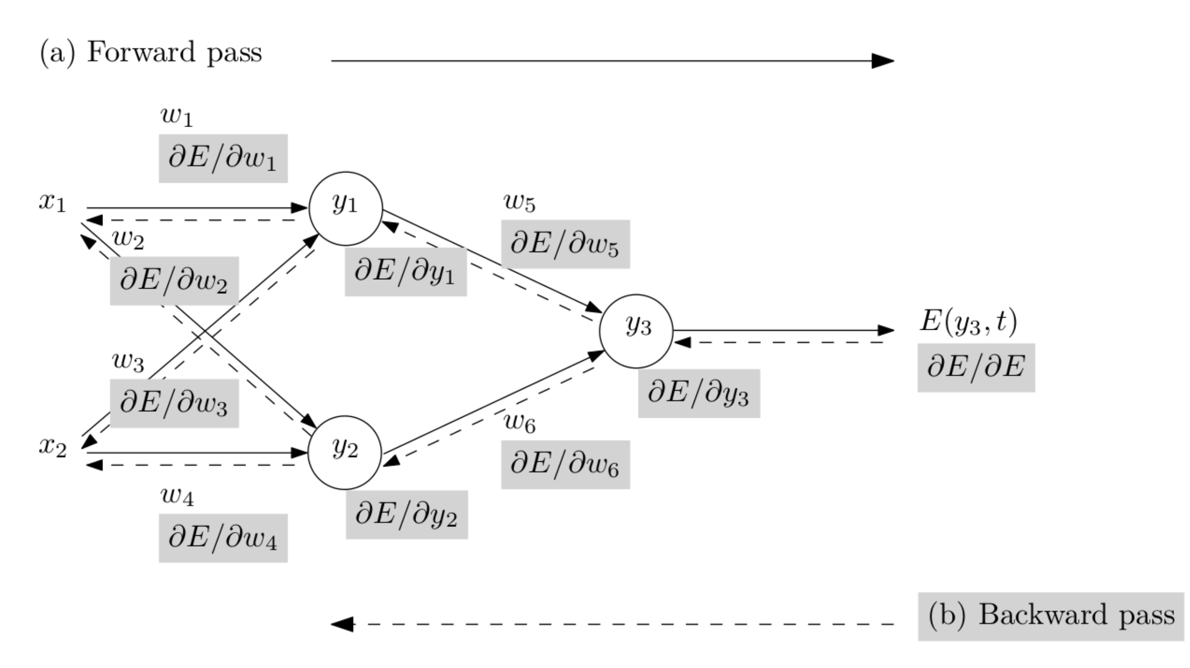
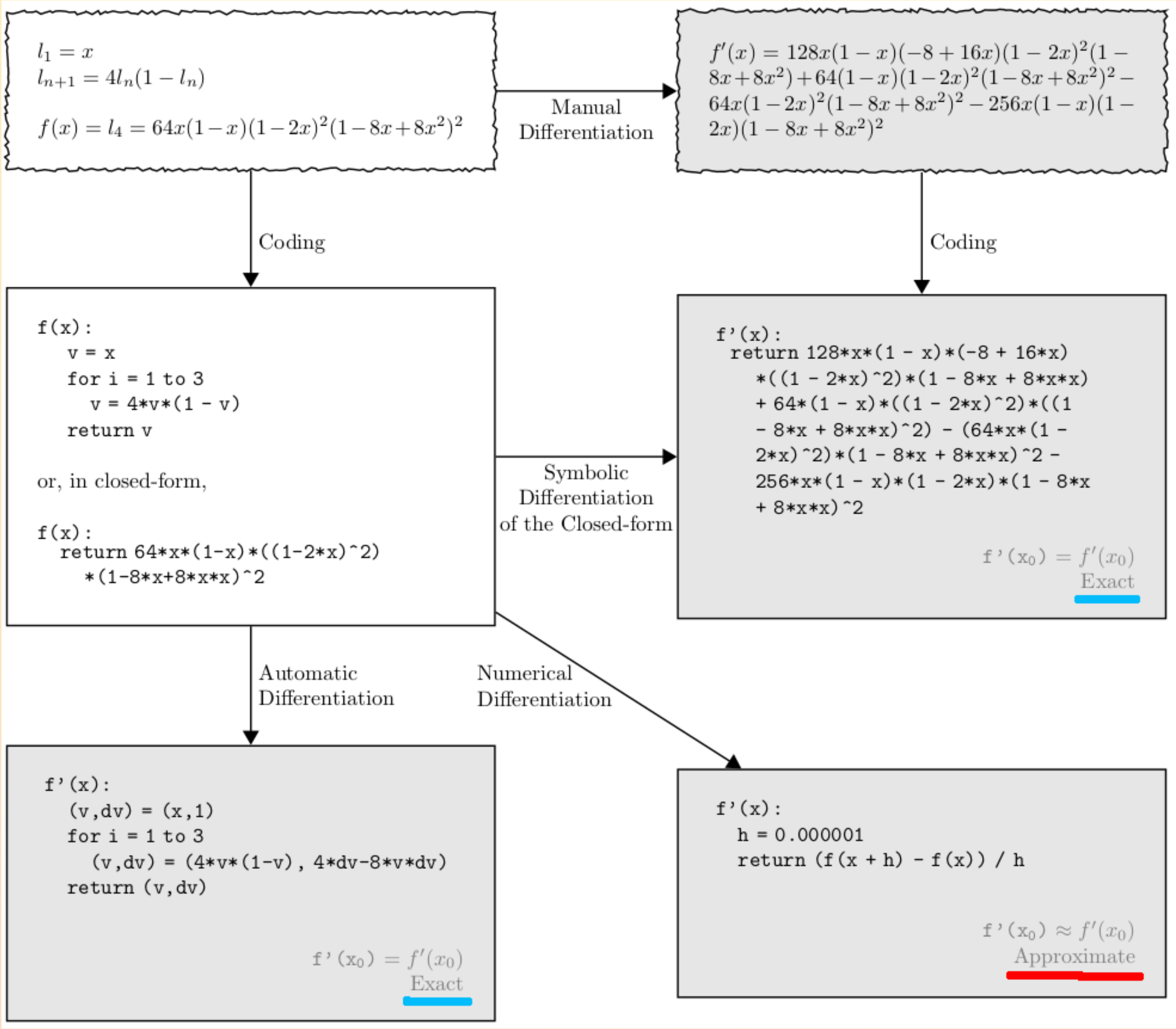


Image from Güneş Baydin et al, JMLR 18 (2018) 1--43

Derive



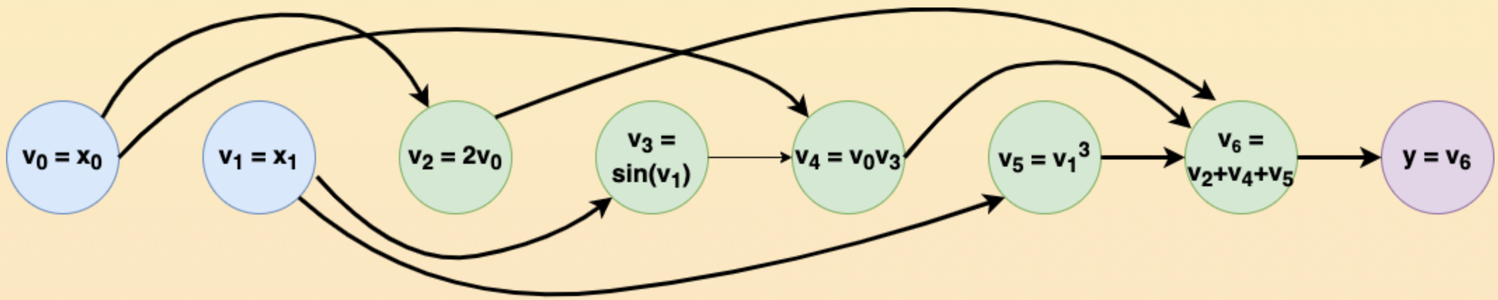
—
Image from Güneş Baydin et al, JMLR 18 (2018) 1--43

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, $f : \mathbb{R} \to \mathbb{R}^m$
- Evaluates $(\frac{\partial f_1}{\partial x}, \ldots, \frac{\partial f_m}{\partial x})$

Reverse mode

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

- Computational cost of calculating $\mathbf{J}_f(\mathbf{x})$ for $f : \mathbb{R}^n \to \mathbb{R}^m$ in $\mathbb{R}^n \times \mathbb{R}^m$

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

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×-0.41 $1 \times 0.9093 + 1 \times 0$ $3 \times 0 \times 4$ $2 + 0.9093 + 0$
$y = v_6$	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)
```

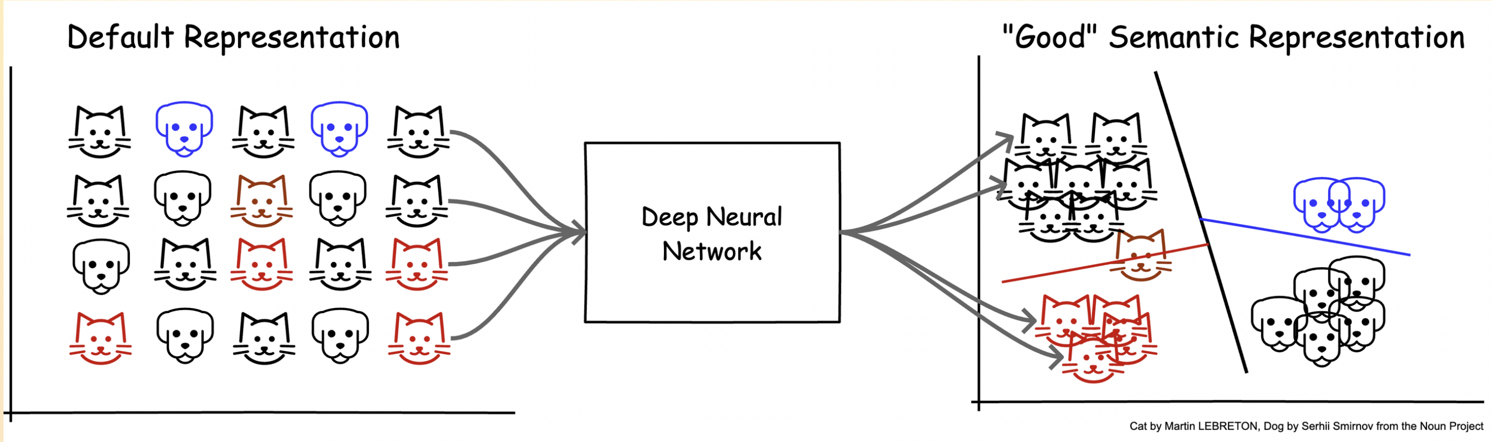
- `Torch` (and similar software) will correctly differentiate only when the atomic operations are supported within it
    - Common operations are overloaded (`__mul__` rewritten by `torch._mult_`)
    - Operations from libraries (`math.sin()`) must be replaced by their differentiation-aware equivalents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

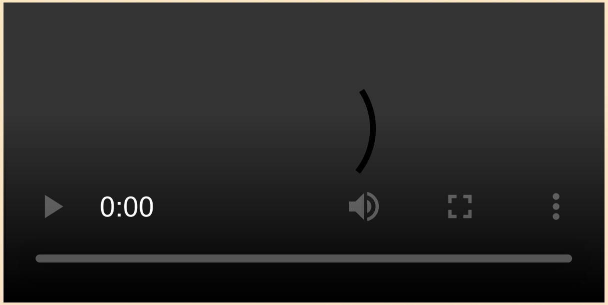
Maps as a Tool of Understanding

—

Tabula Rogeriana by Al Idrisi (1154) 1929 reproduction

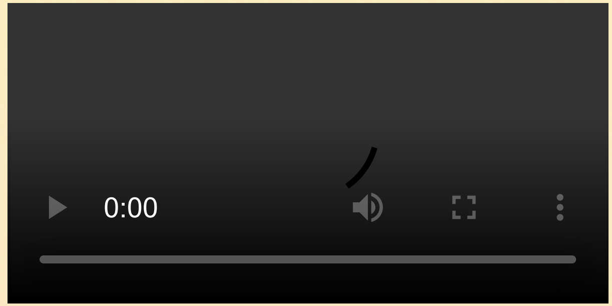
"Representation" simplifies tasks





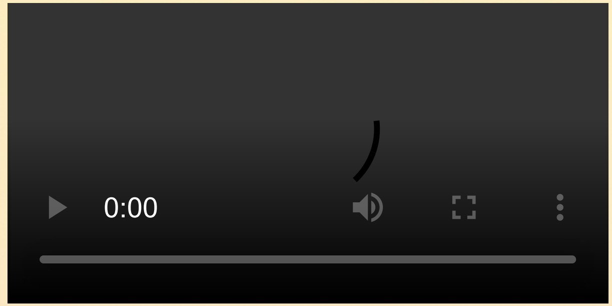
—
Animation and picture from FastForward Labs

Impressive results



—
Videos from YouTube, autonomous driving

Impressive results

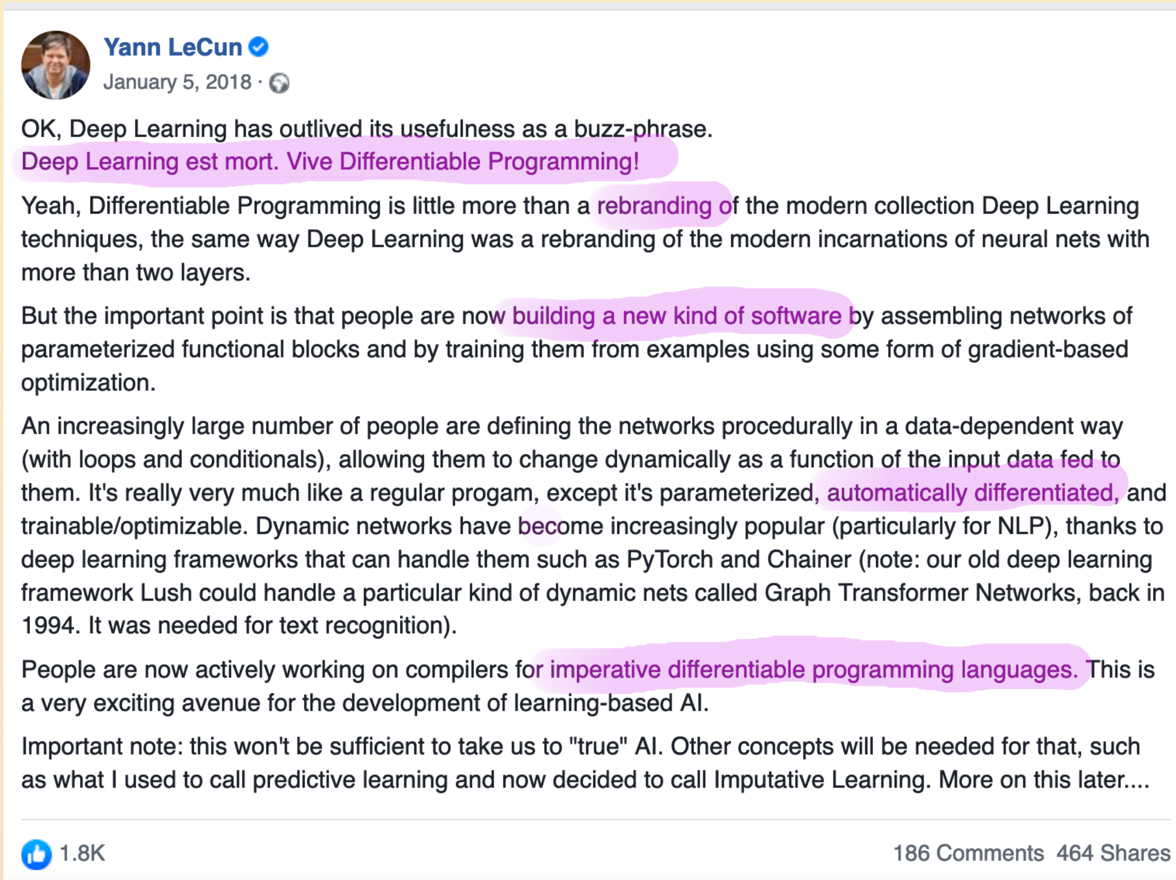


- Busco colaboraciones para aplicaciones médicas de inteligencia artificial

—
Video de YouTube, cancer research

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 progam, 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).

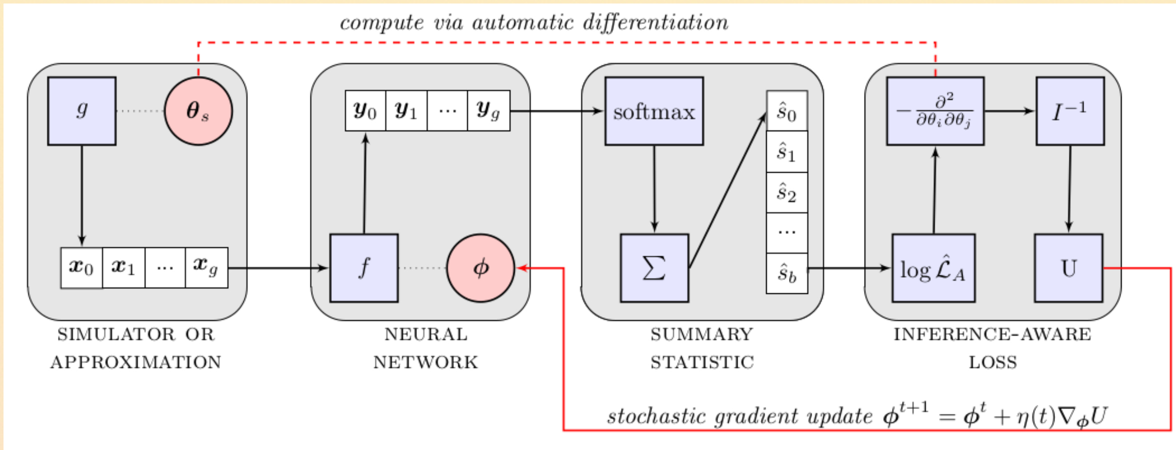
People are now actively working on compilers for imperative differentiable programming languages. This is a very exciting avenue for the development of learning-based AI.

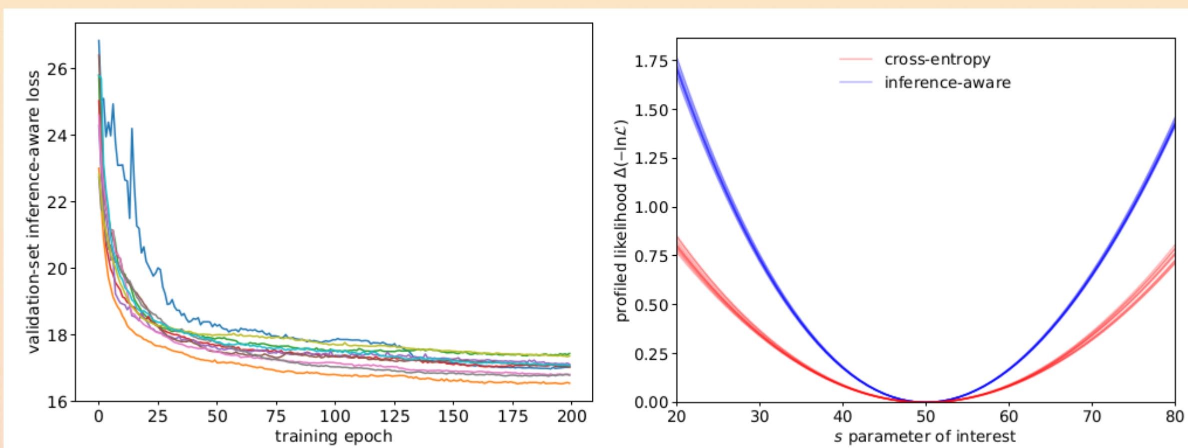
Important note: this won't be sufficient to take us to "true" AI. Other concepts will be needed for that, such as what I used to call predictive learning and now decided to call Imputative Learning. More on this later....

1.8K   186 Comments 464 Shares

—
Screenshot of Yann LeCun's facebook post

Go to INFERNO: syst-aware inference opt.





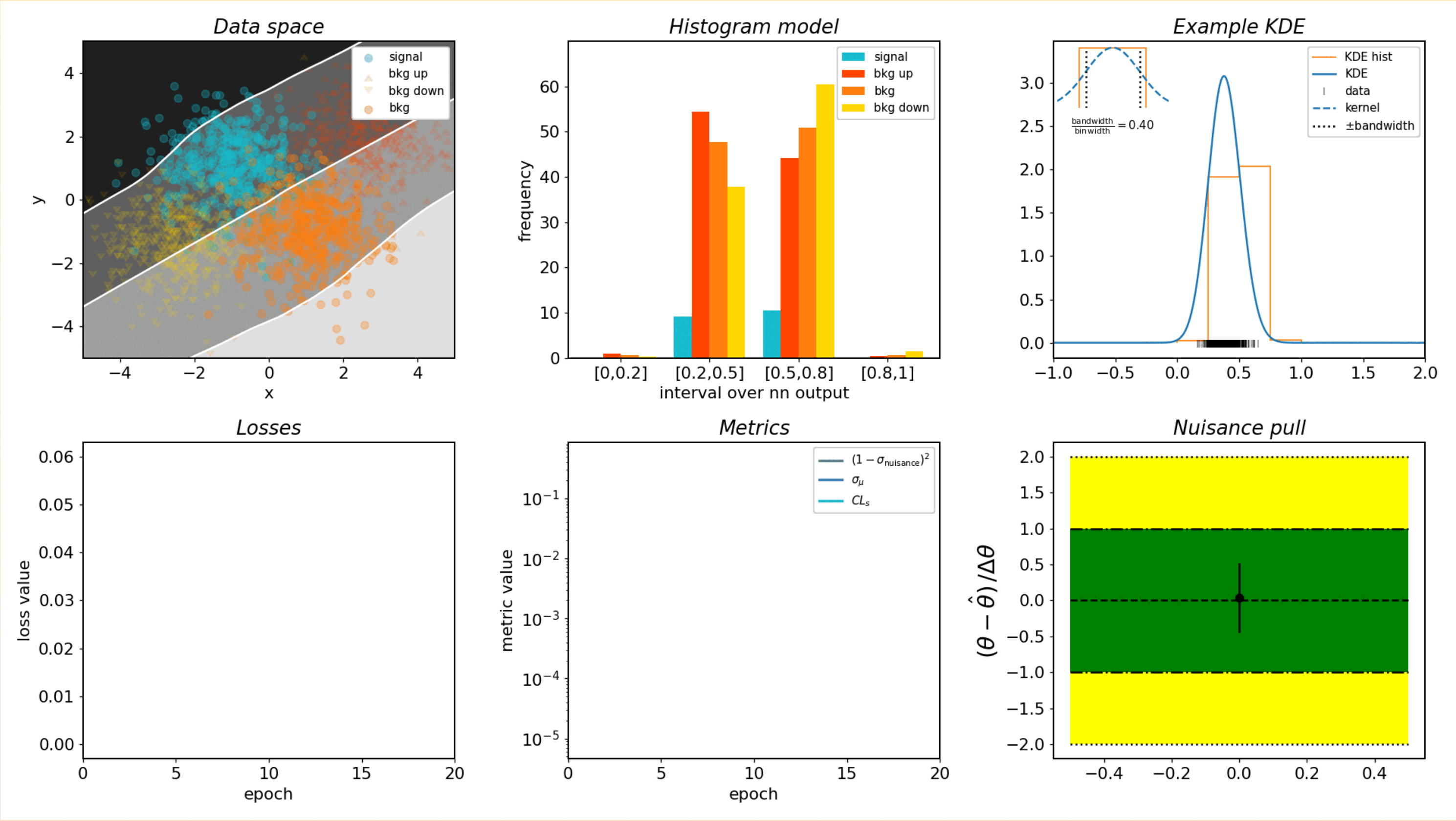
(a) inference-aware training loss

(b) profile-likelihood comparison

Figures from 10.1016/j.cpc.2019.06.007

Measurement-aware analysis opt.

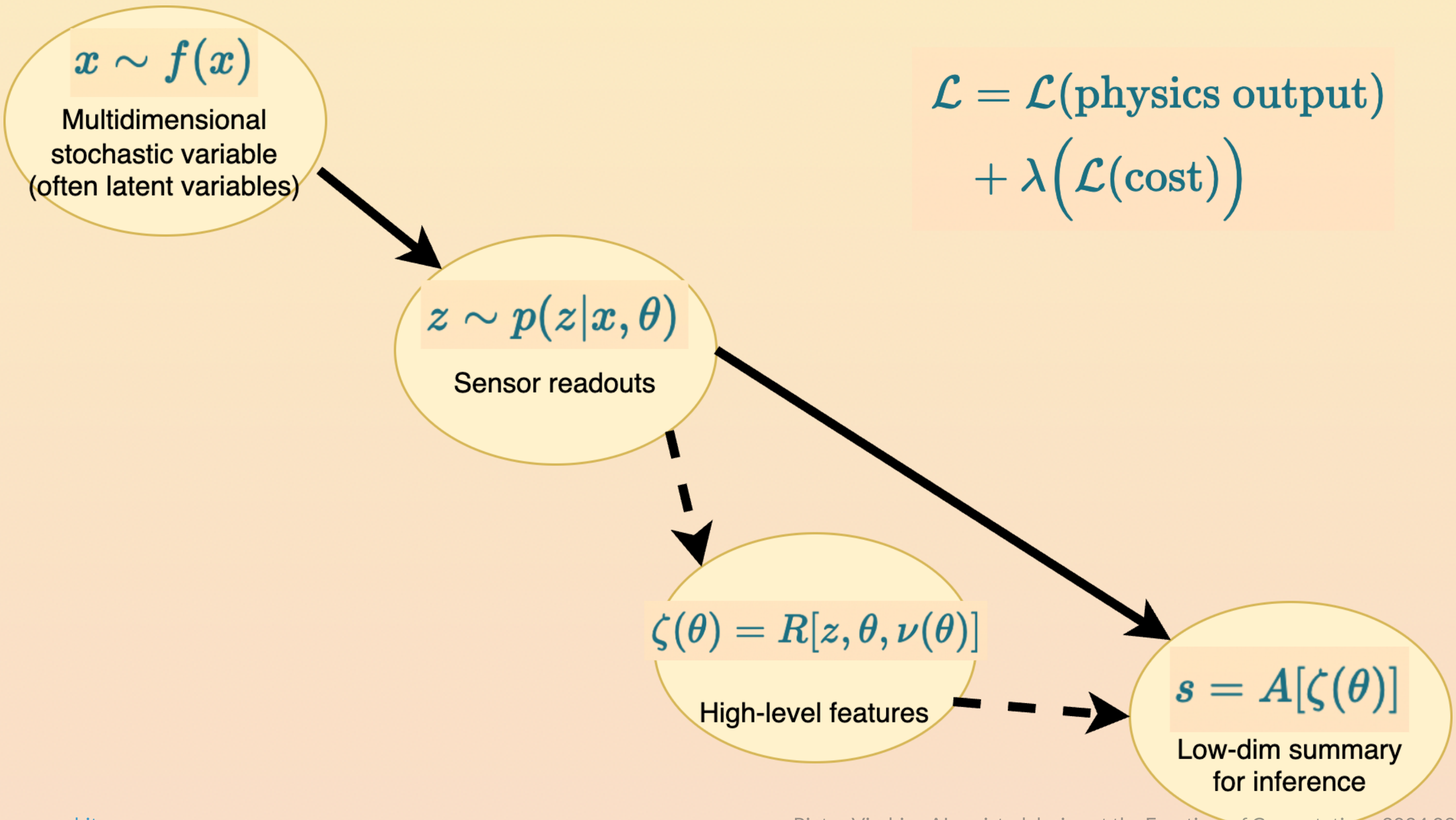
neos



—
Animation from 2203.05570

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 (2203.13818), 117-pages document, physicists + computer scientists



$$\mathcal{L} = \mathcal{L}(\text{physics output}) + \lambda\Big(\mathcal{L}(\text{cost})\Big)$$

—
Formulas from our white paper

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

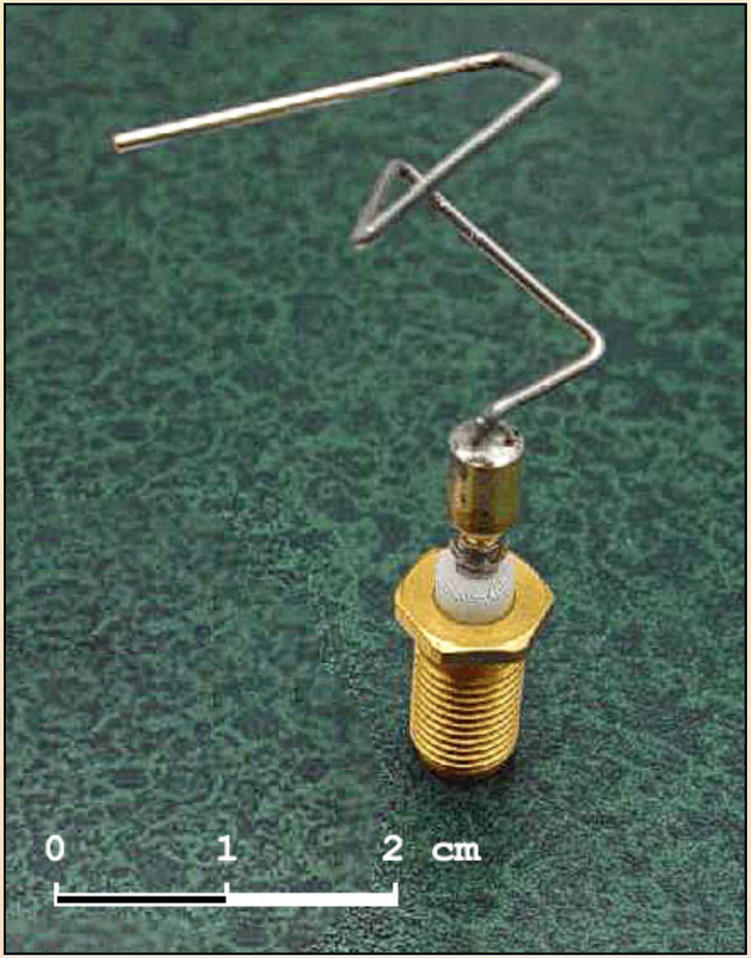
Depends on z and nuisances

$$\hat{\theta} = arg\ min_{\theta} \int \underbrace{L[A(\zeta), \overbrace{c(\theta)}^{\text{Cost of the layout with parameters theta}}]}_{\text{Weight desirable goals while obeying cost constraints}} \overbrace{p(z|x,\theta)}^{\text{Closed form}} f(x)dxdz \ ,$$

—

From our white paper and internal TomOpt presentation

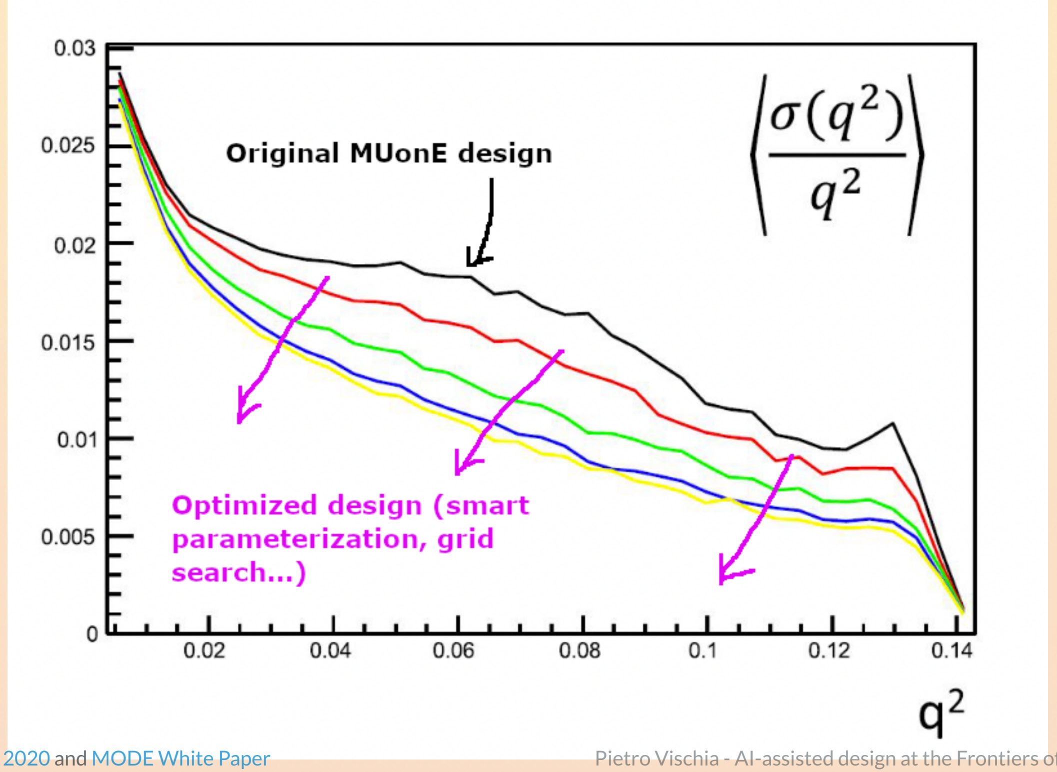
Thrive in asymmetries



—
Photo from doi:10.2514/6.2006-7242

Large gains to be had

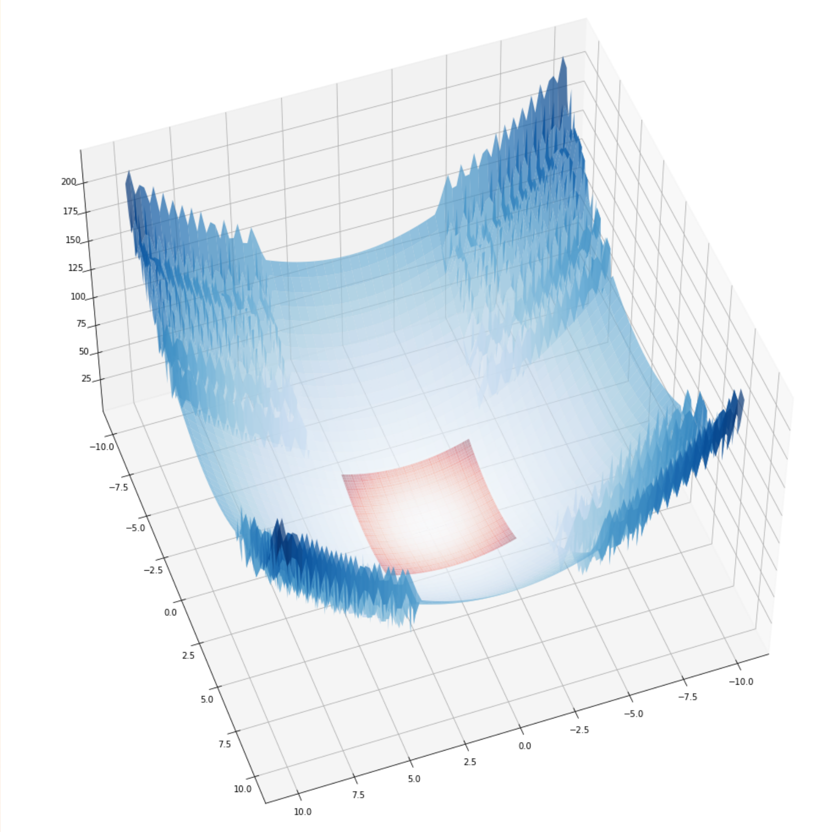
- MUonE: proposed 150 GeV muon beam experiment to be built at CERN
  - Measure precisely the q^2 differential cross section in electron-muon scattering
  - 40 tracking stations and a calorimeter
- Dramatic improvement in the resolution on q^2 even from a simple grid search



—
Image from Physics Open, 4:100022, 2020 and MODE White Paper

Assist the physicist with a landscape of solutions

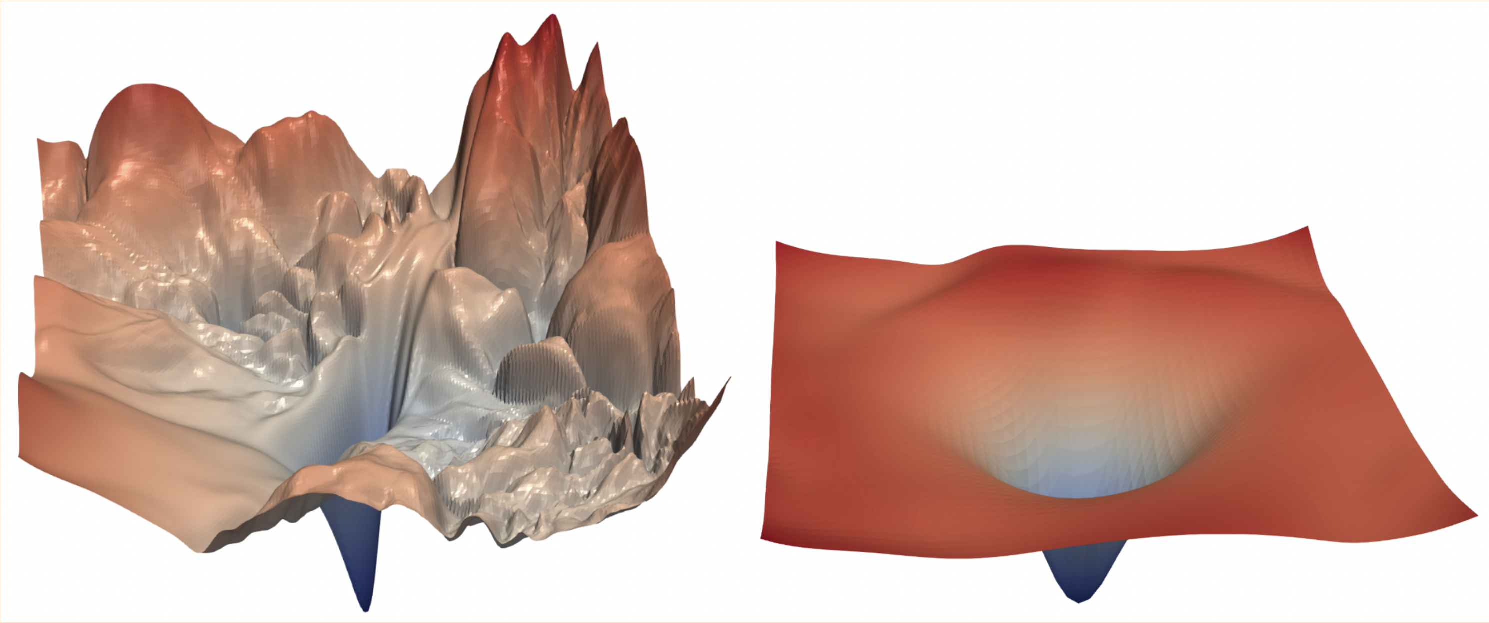
- Cannot parameterize everything
- The optimal solution: unrealistic
- Provide feasible solutions near optimality
- The physicist will fine tune



—
Illustration (c) P. Vischia, book in preparation

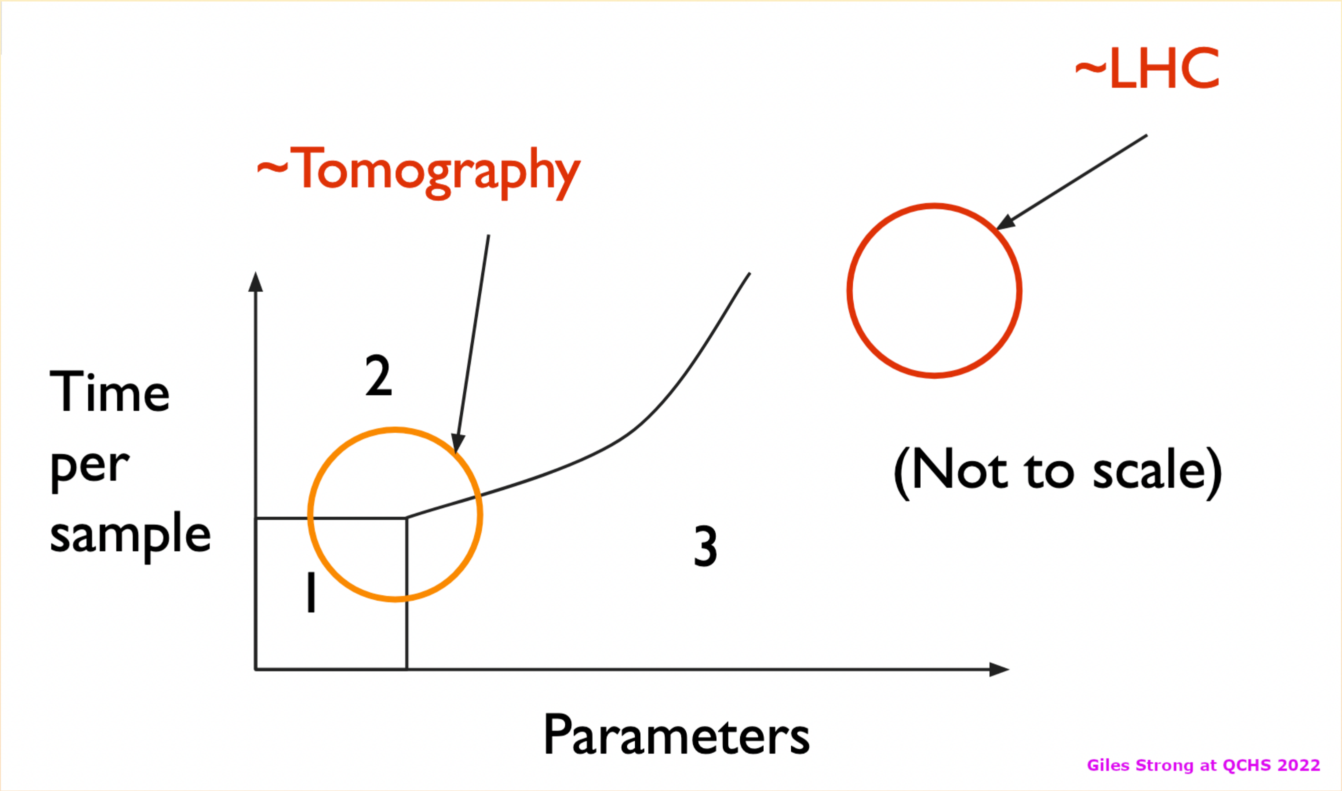
How far from optimality?

- 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



—
Image from Li et al (taken as pictorial representation out of its original context)

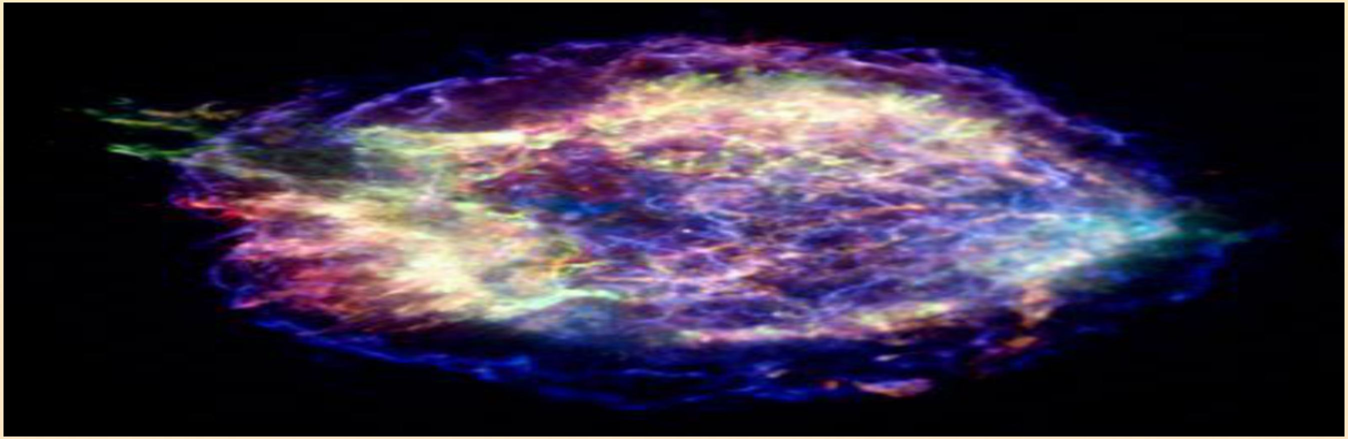
Method of choice depends on scale



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

Cosmic rays from supernovae

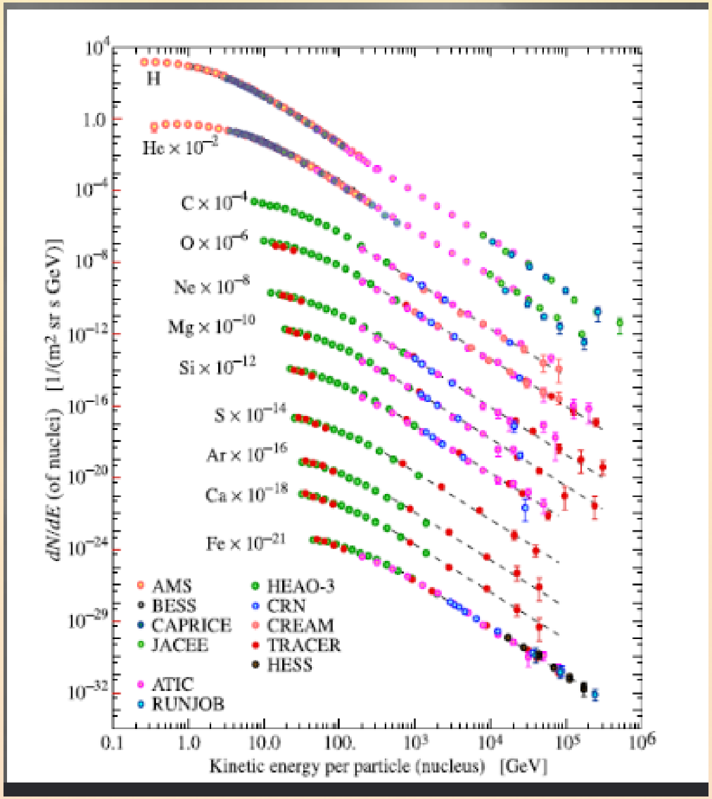
- High-energy primary cosmic rays produced by supernovae



—
Image from uncw.edu

Primary cosmic rays

- 89% hydrogen nuclei (protons)
- the rest is Helium, Carbon, Oxigen, 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

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

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

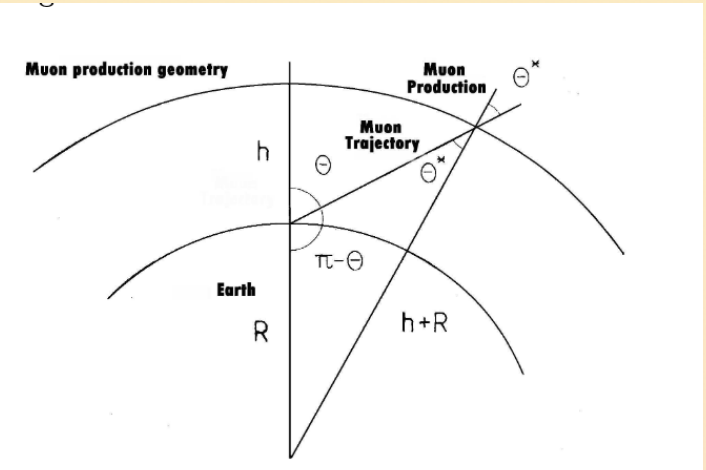


Fig. 1. The relation of the observed zenith angle of muons, θ^*, to the zenith angle at the muon production point in the atmosphere, θ. R is the radius of the Earth. Adopted from [3][4]

—
Image from 1509.06176

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

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

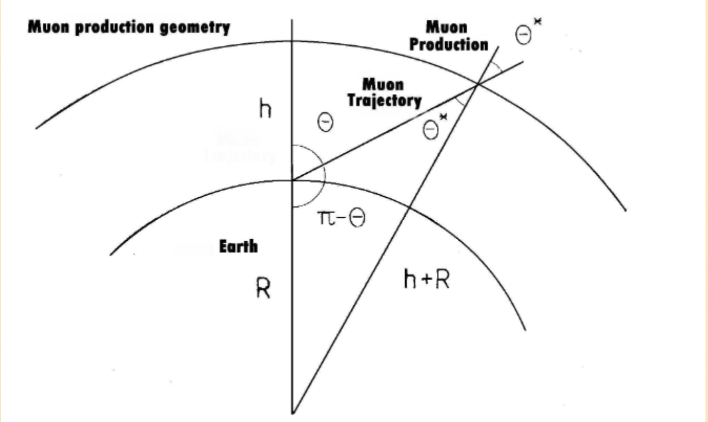
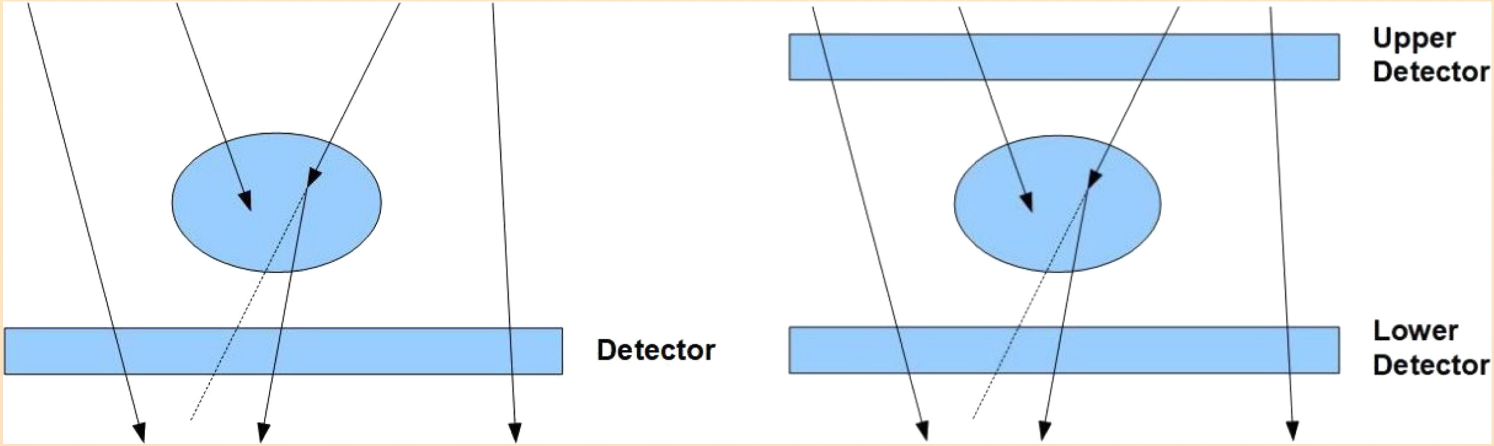


Fig. 1. The relation of the observed zenith angle of muons, θ^*, to the zenith angle at the muon production point in the atmosphere, θ. R is the radius of the Earth. Adopted from [3][4]

Absorption or scattering

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



—

Image from our white paper

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

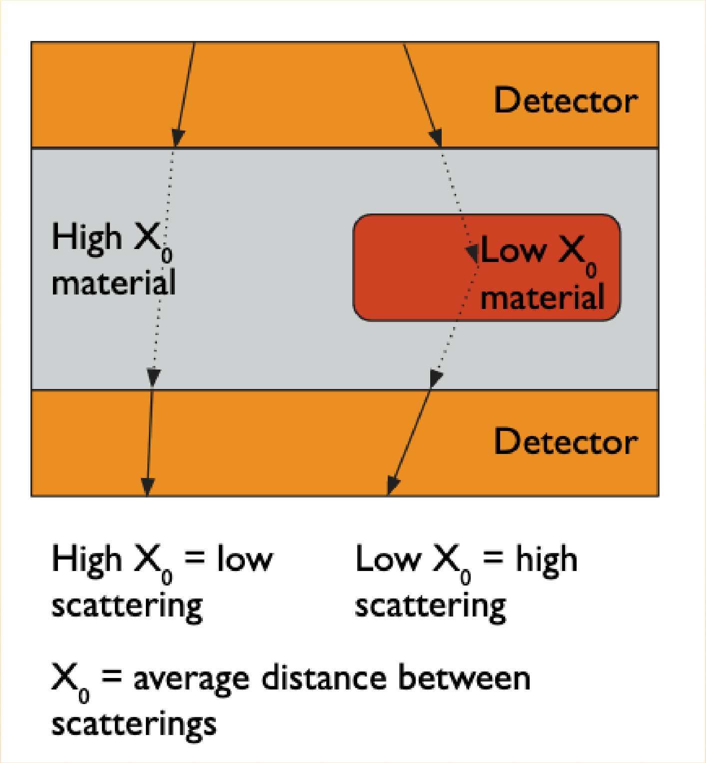
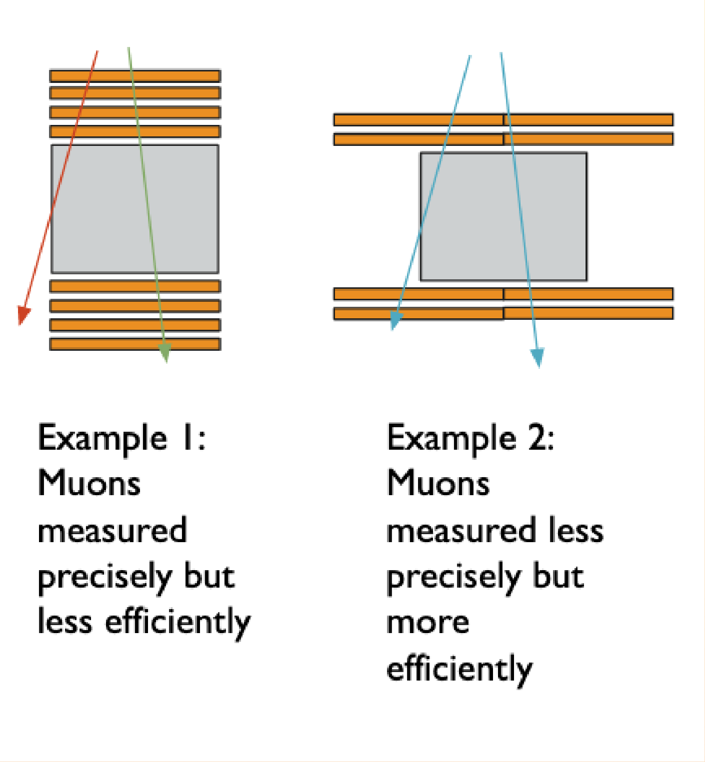


Figure from the TomOpt project

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



—
Figure from the TomOpt project

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
  - Gradient descent

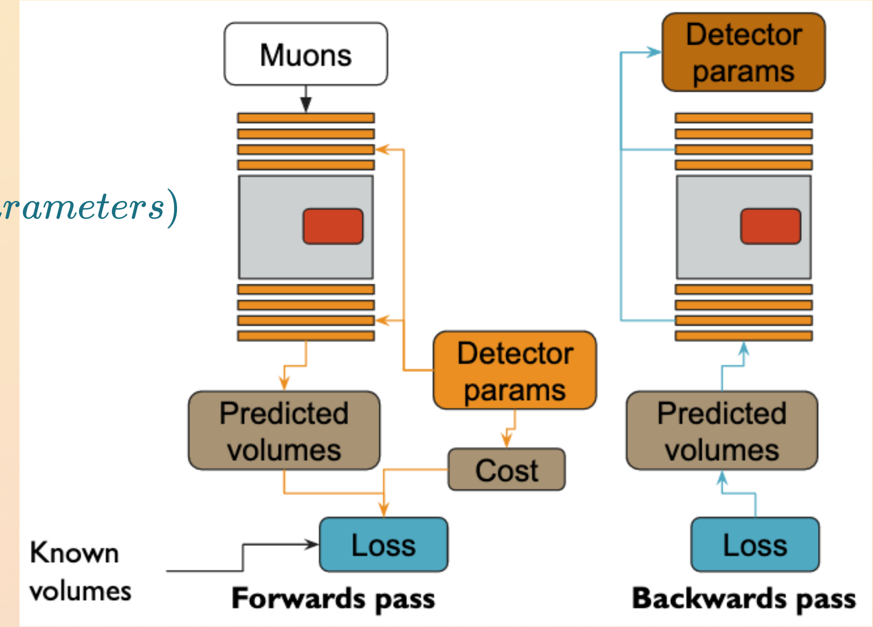
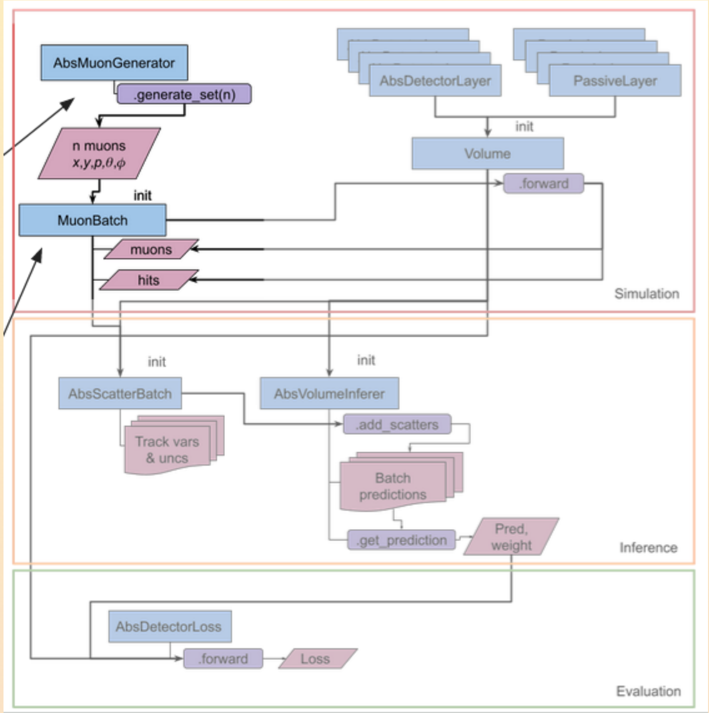


Figure from the TomOpt project

Muon Generation

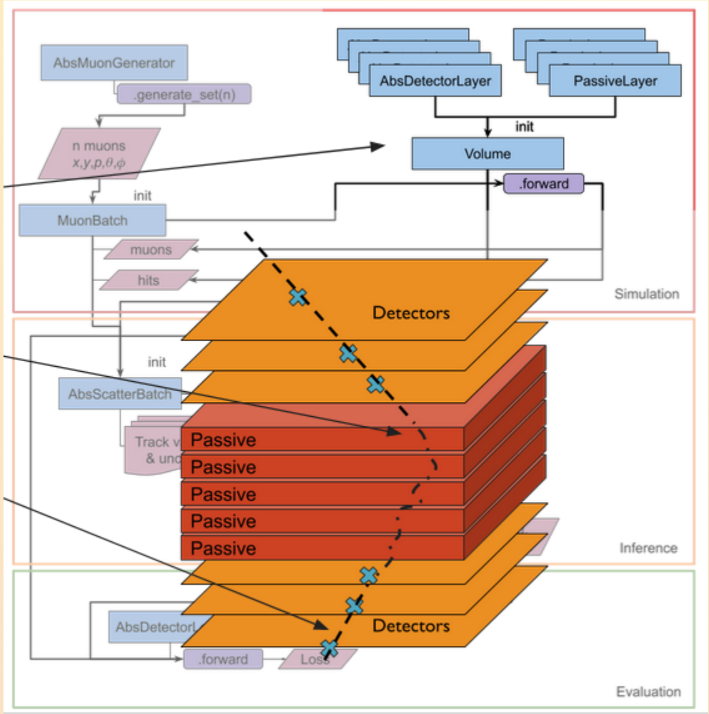
- Formulas by 2015 and 2016 models
- Tensor of muons (x, y, p, θ, ϕ)
    - θ, p from flux model
    - x, y from ranges
    - ϕ uniformly in $[0, 2\pi]$
- Code handles many muons at once (MuonBatch)
    - Propagate the muon position (can snapshot to track)
    - Scatter $(dxy, d\theta d\phi)$ at each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



—
Left: Guan et al, Right: G.C.Strong

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



—

Figure by G.C.Strong

Panel specification

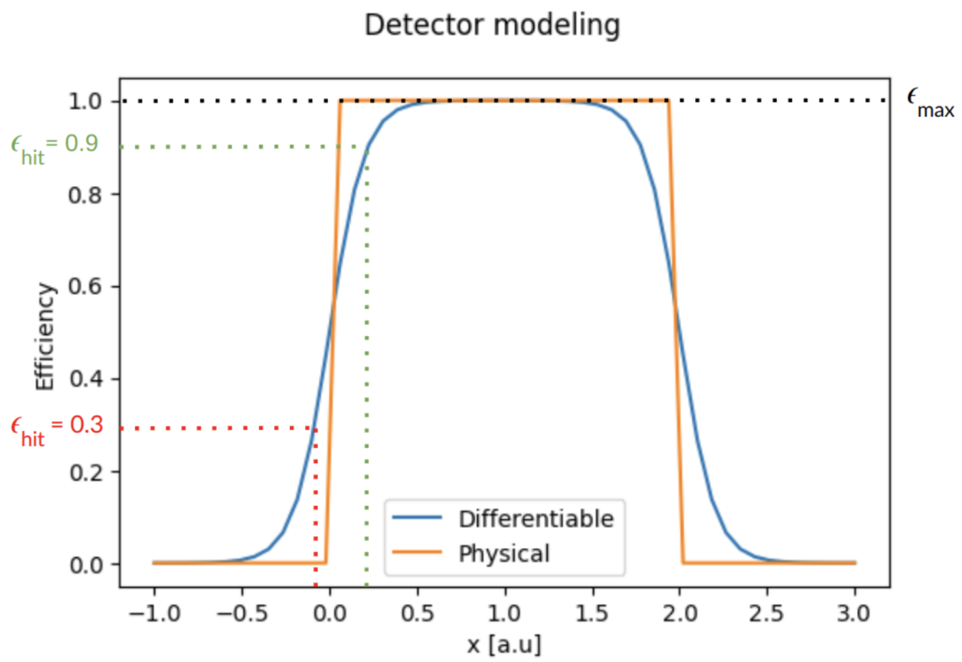
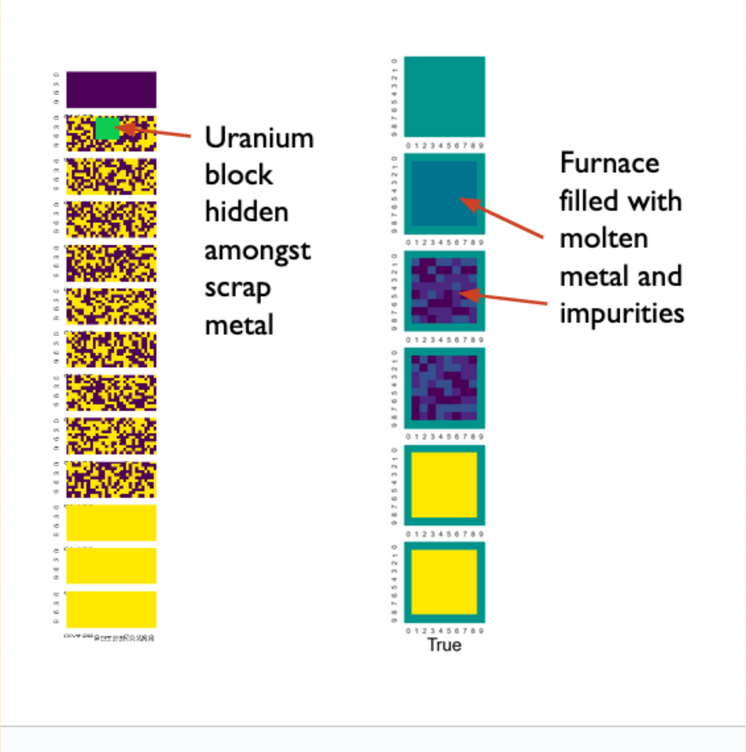


Figure 2: Example of detector panel modeling with sigmoid function used during optimisation (blue) and with rectangle function used during validation (orange).

—
Figure from 2309.14027

Monte Carlo Truth

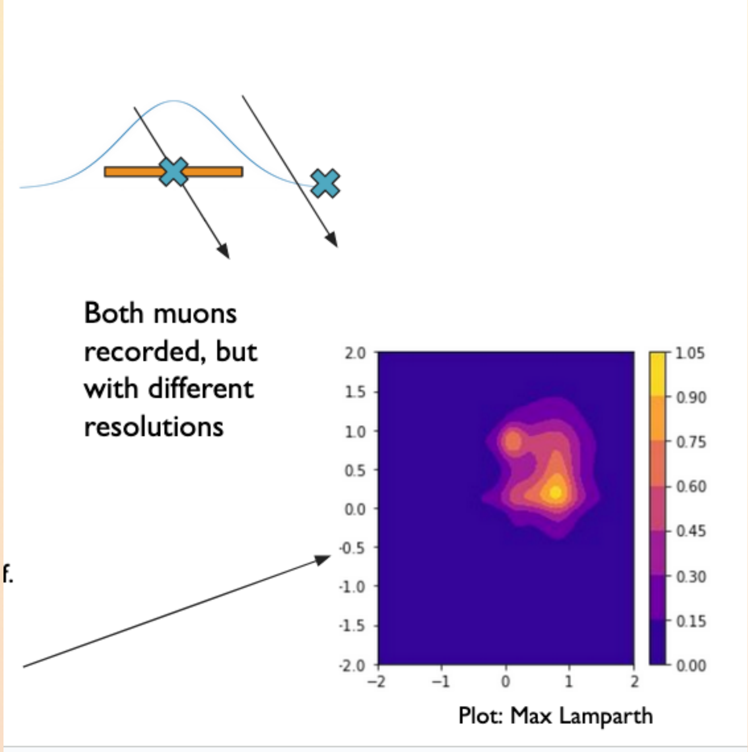
- Per each scenario, can build voxelized random volumes
  - Each voxel can be a different material
  - Next: material mixture per voxel



—
Figure by G.C.Strong

Make muon hits differentiable

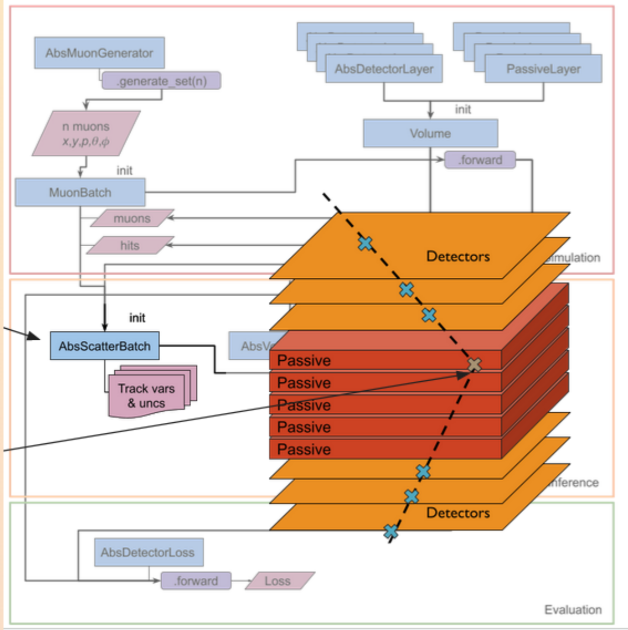
- Associate a distribution to resolution and efficiency
  - e.g. Gaussian centered on panel and width equal to panel span
  - p.d.f. of the muon position is now differentiable
- Further generalization: Gaussian Mixture models



—
Figure by Max Lamparth

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



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

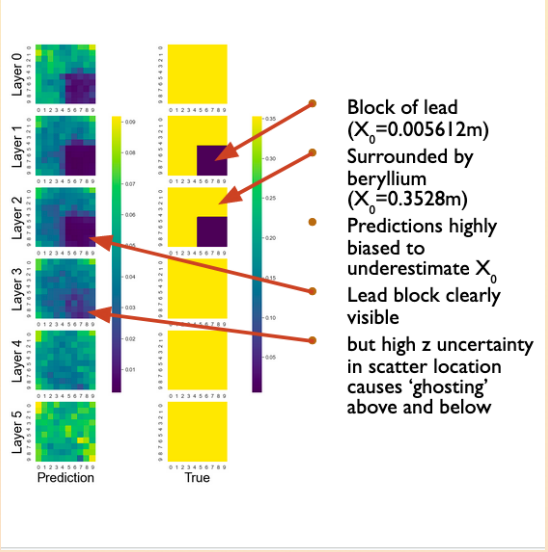
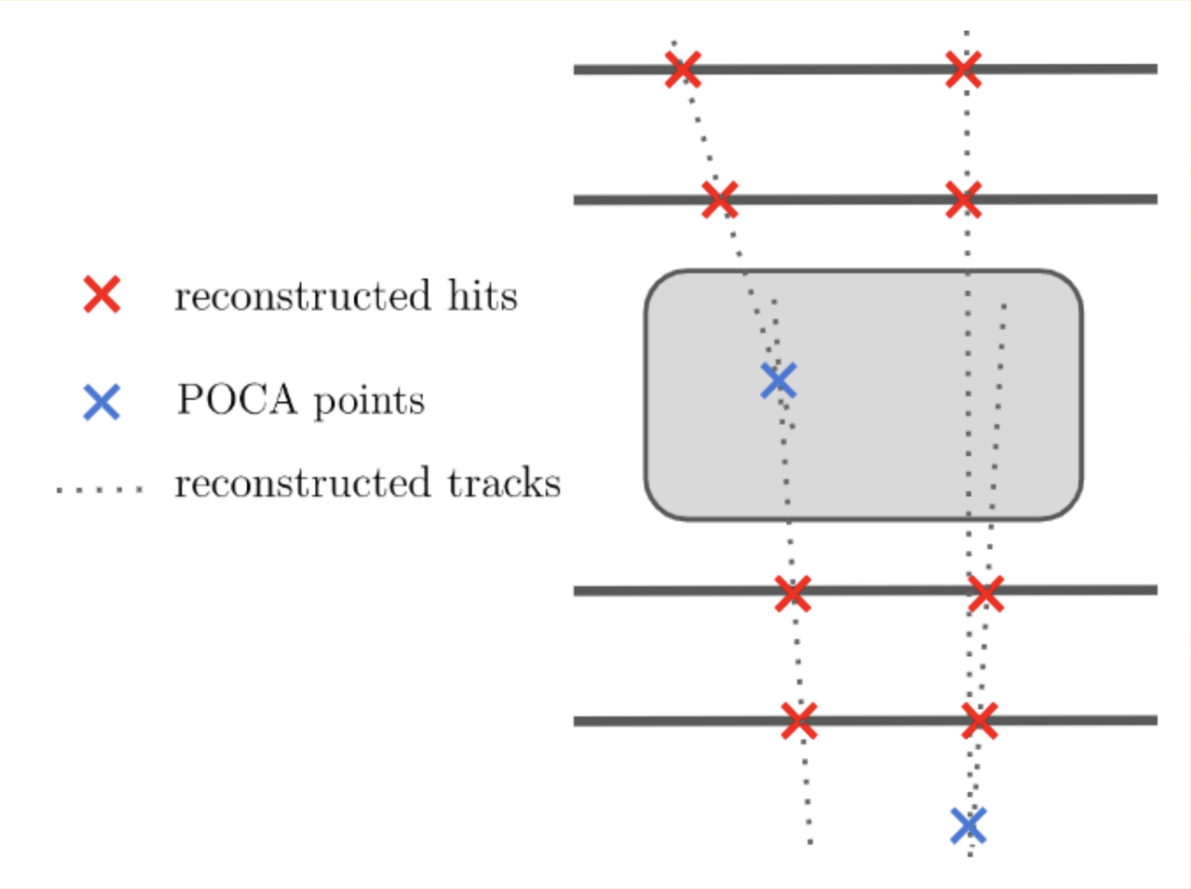


Figure by G.C.Strong

POCA

- Assume one scattering $\longrightarrow$ bias!

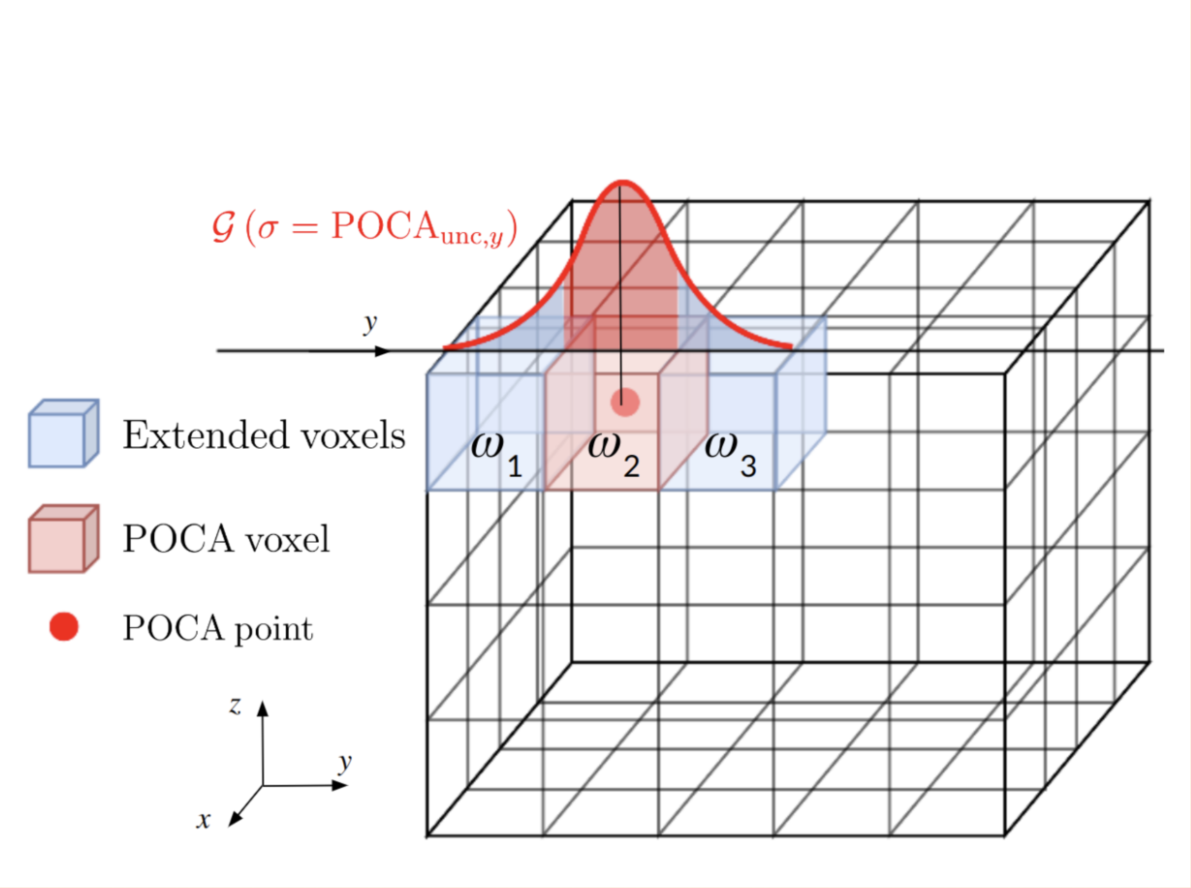


—

Figure from 2309.14027

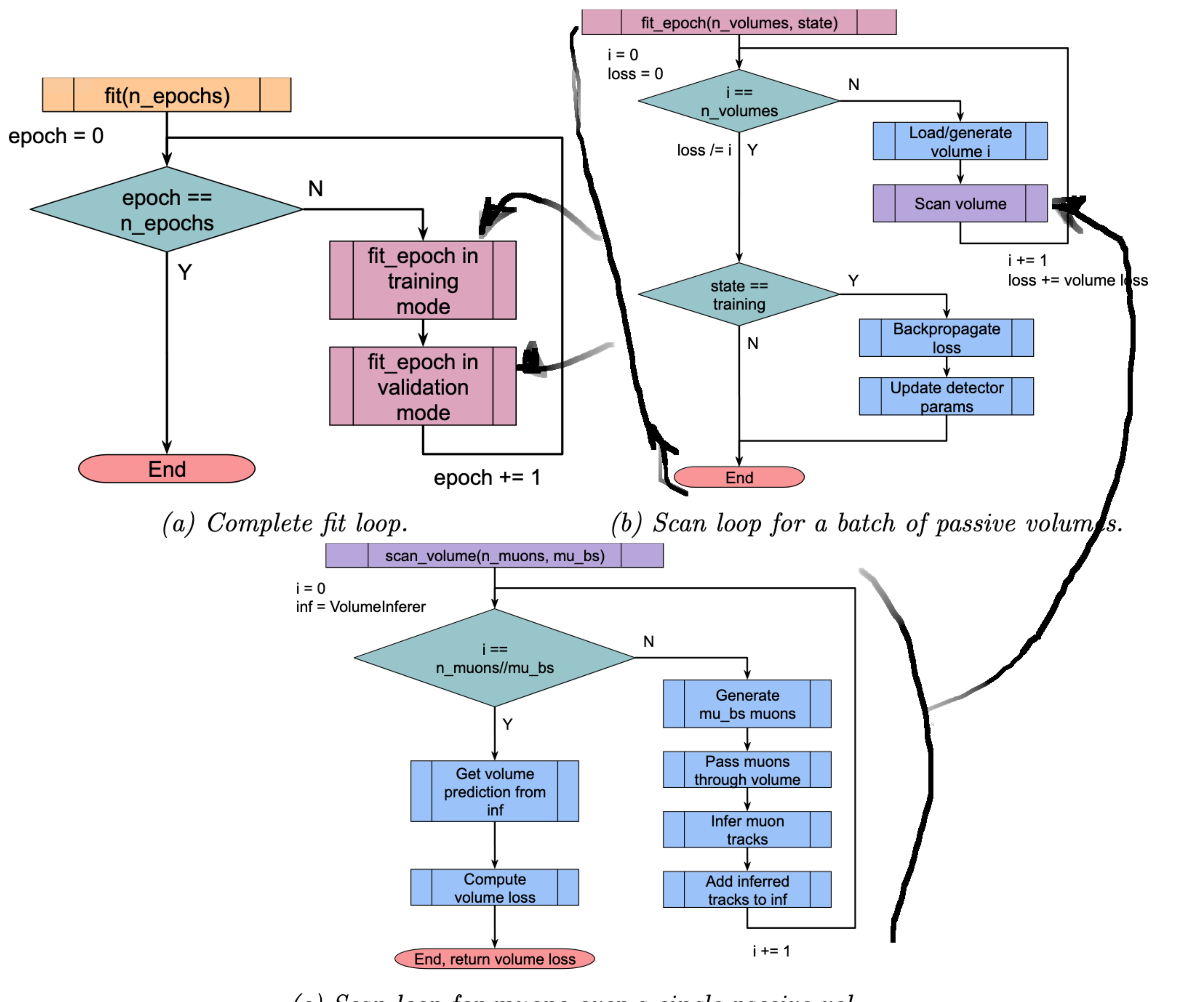
Extended POCA

- Assume one scattering $\longrightarrow$ bias!



—
Figure from 2309.14027

Detector Optimization in TomOpt

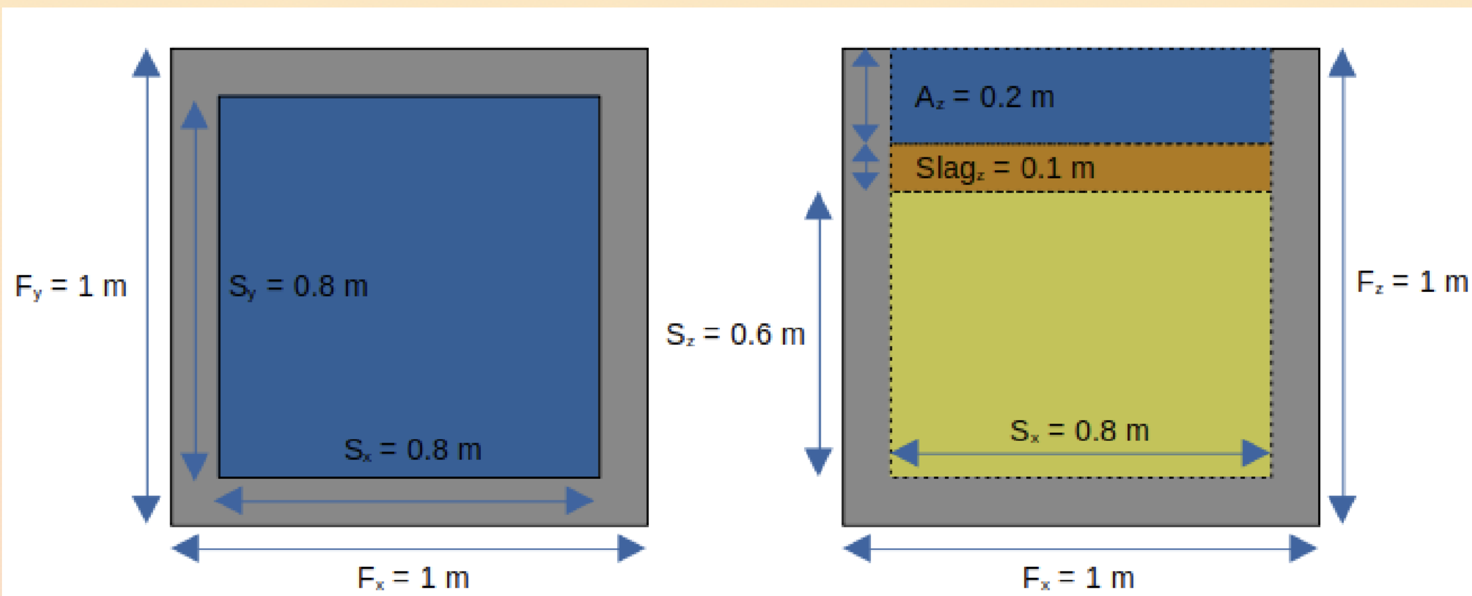


(a) Complete fit loop. *(b) Scan loop for a batch of passive volumes.*

(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



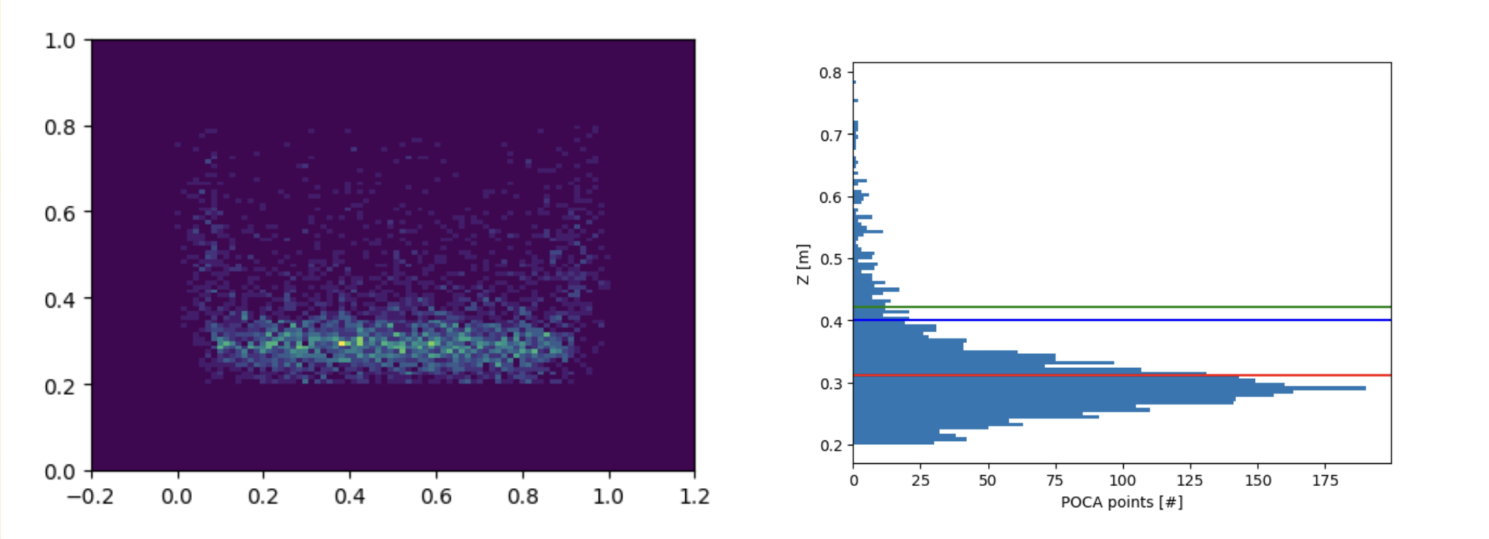
—
Figure from 2309.14027, and dolcidasogno.com

Heavy Metal

—
Figure from 2309.14027, Boris Bukovsky, viaPixabay

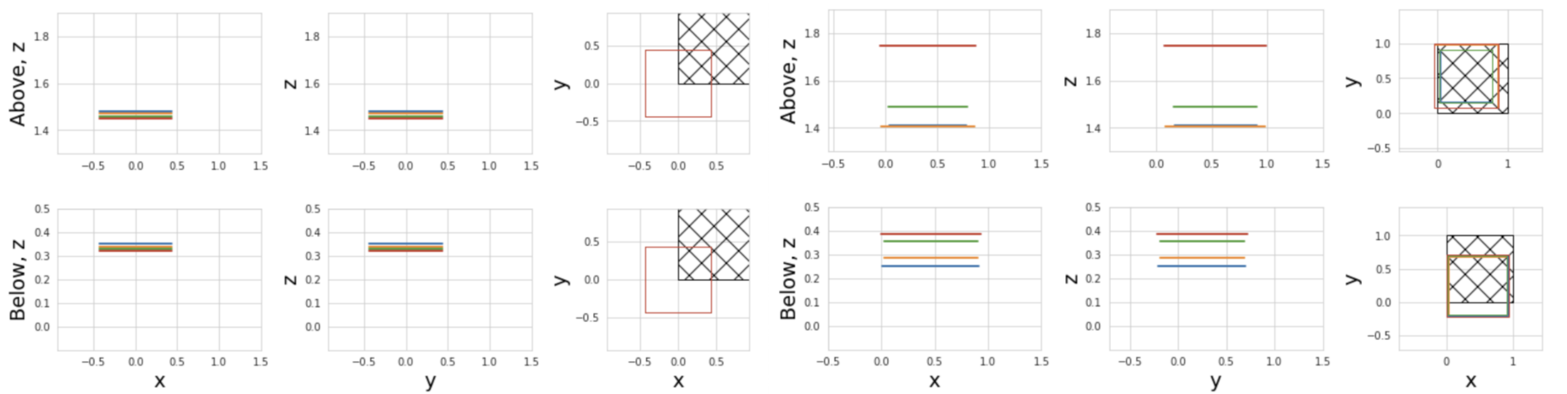
Encouraging results

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



(Blue: true steel level. Red: prediction. Green: bias-corrected prediction)

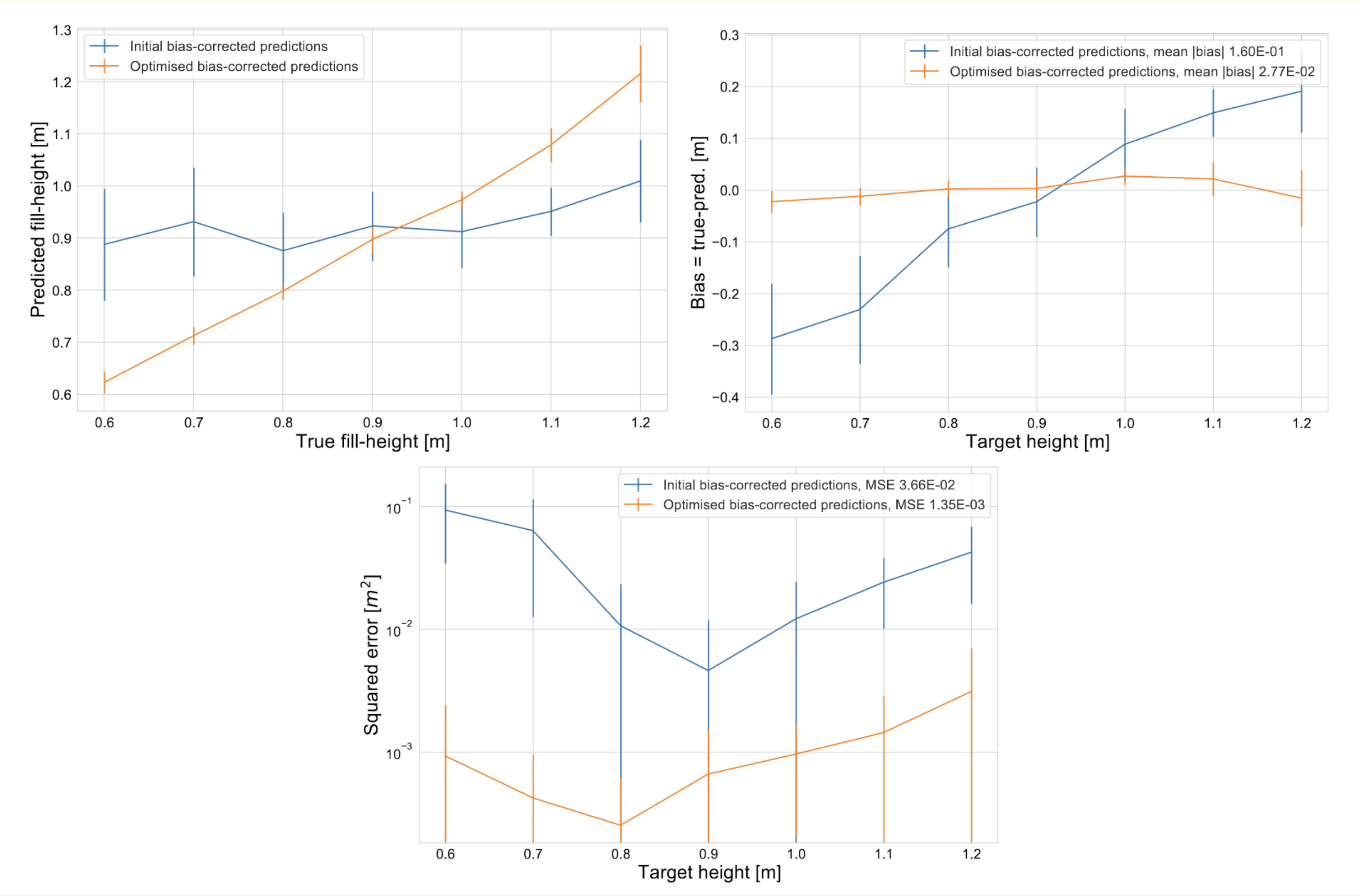
Optimize crappy detector



(a) Initial detector configuration. *(b) Detector configuration after stage one optimisation process.*

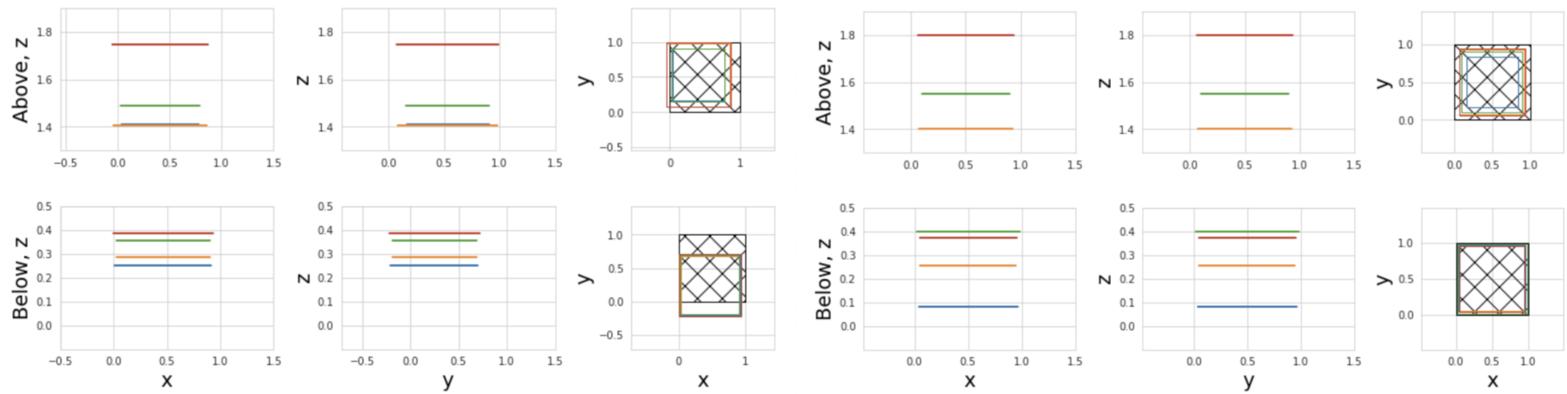
—
Figure from 2309.14027

Optimize crappy detector



—
Figure from 2309.14027

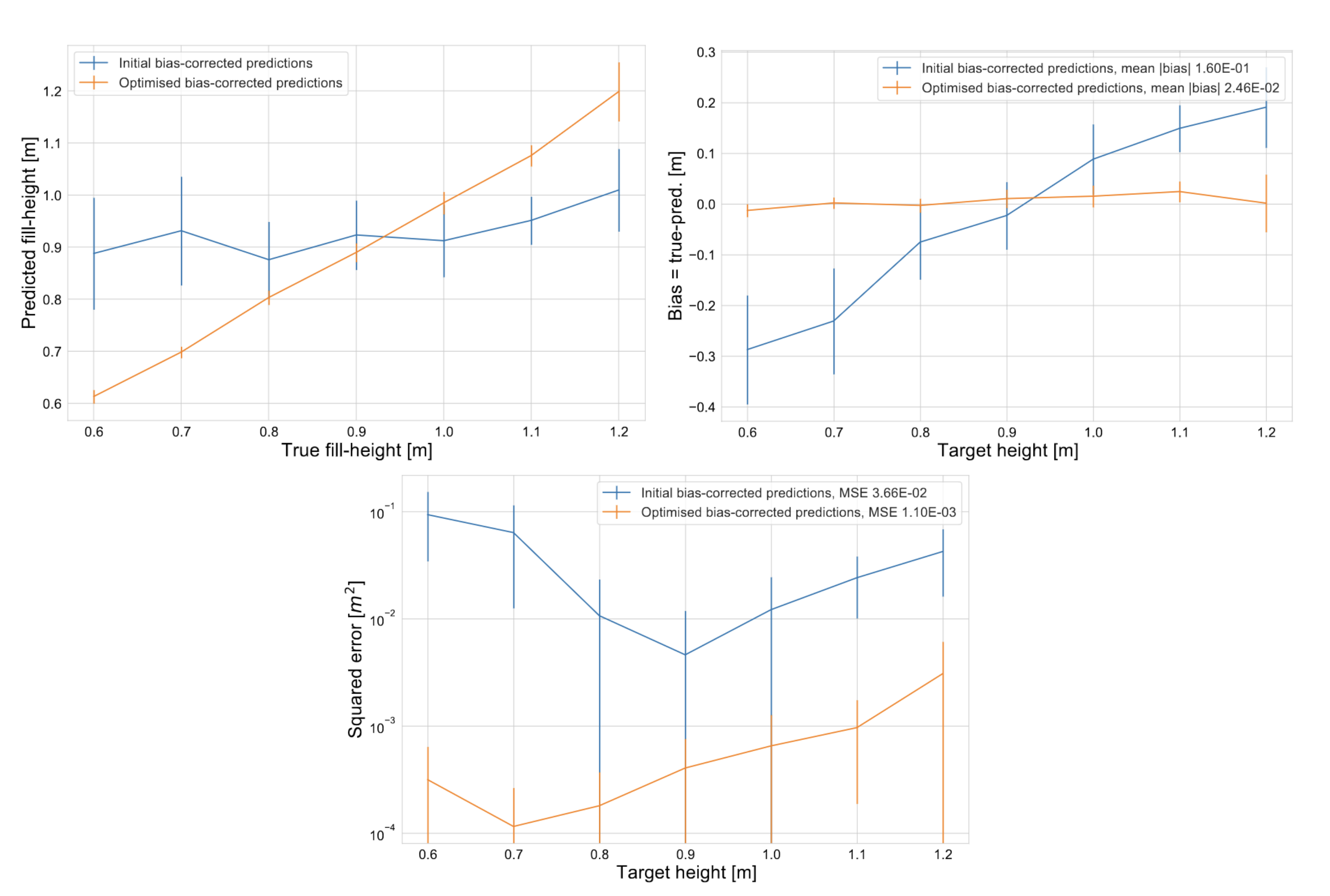
Refine a good detector



(a) Detector configuration after stage one optimisation process. *(b) Detector configuration after stage two optimisation process.*

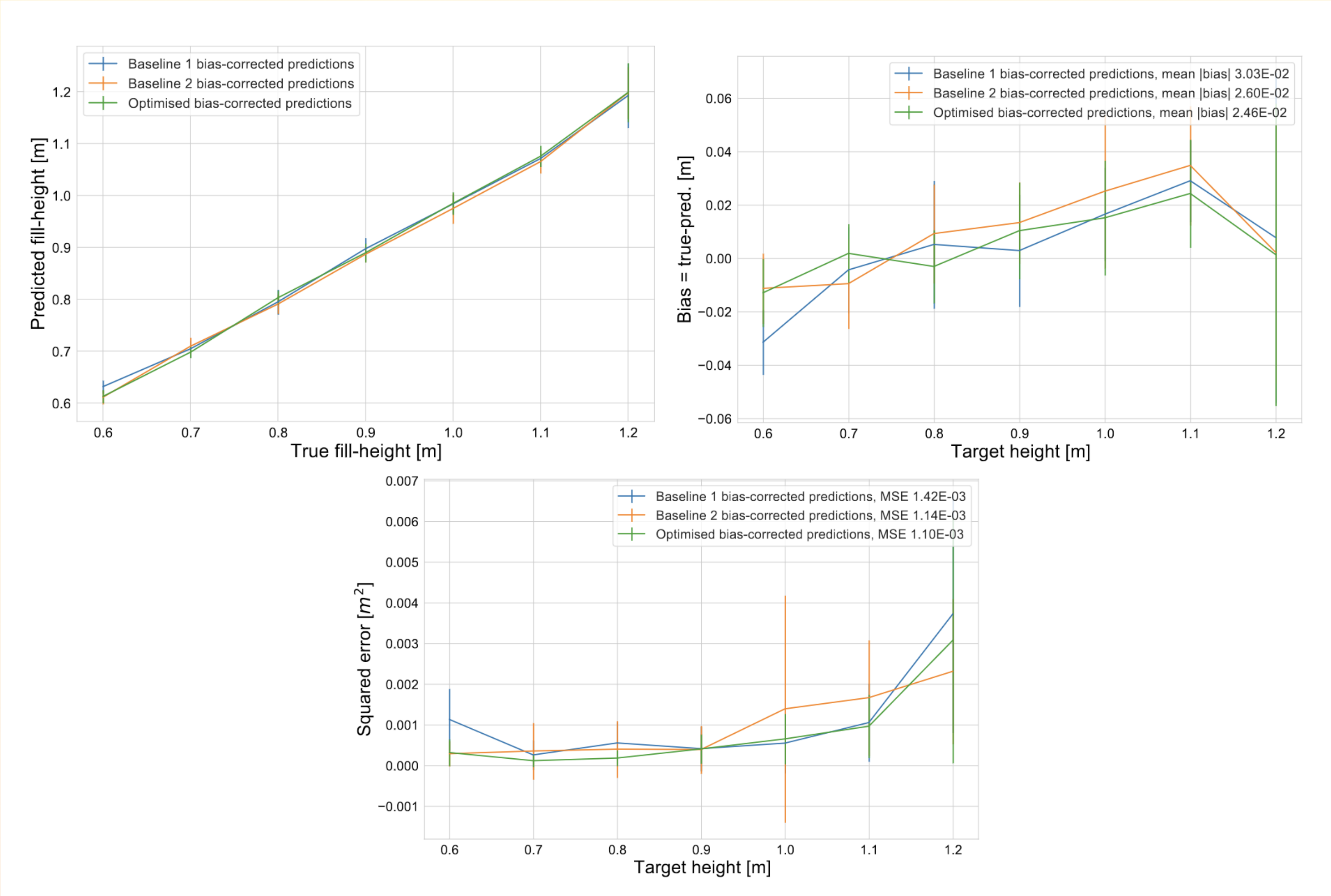
—
Figure from 2309.14027

Refine a good detector



—
Figure from 2309.14027

Recover human baselines



—

Figure from 2309.14027

Let's talk about Gamma Ray

—
Figure from spirit-of-metal.com

Let's talk about Gamma Rays

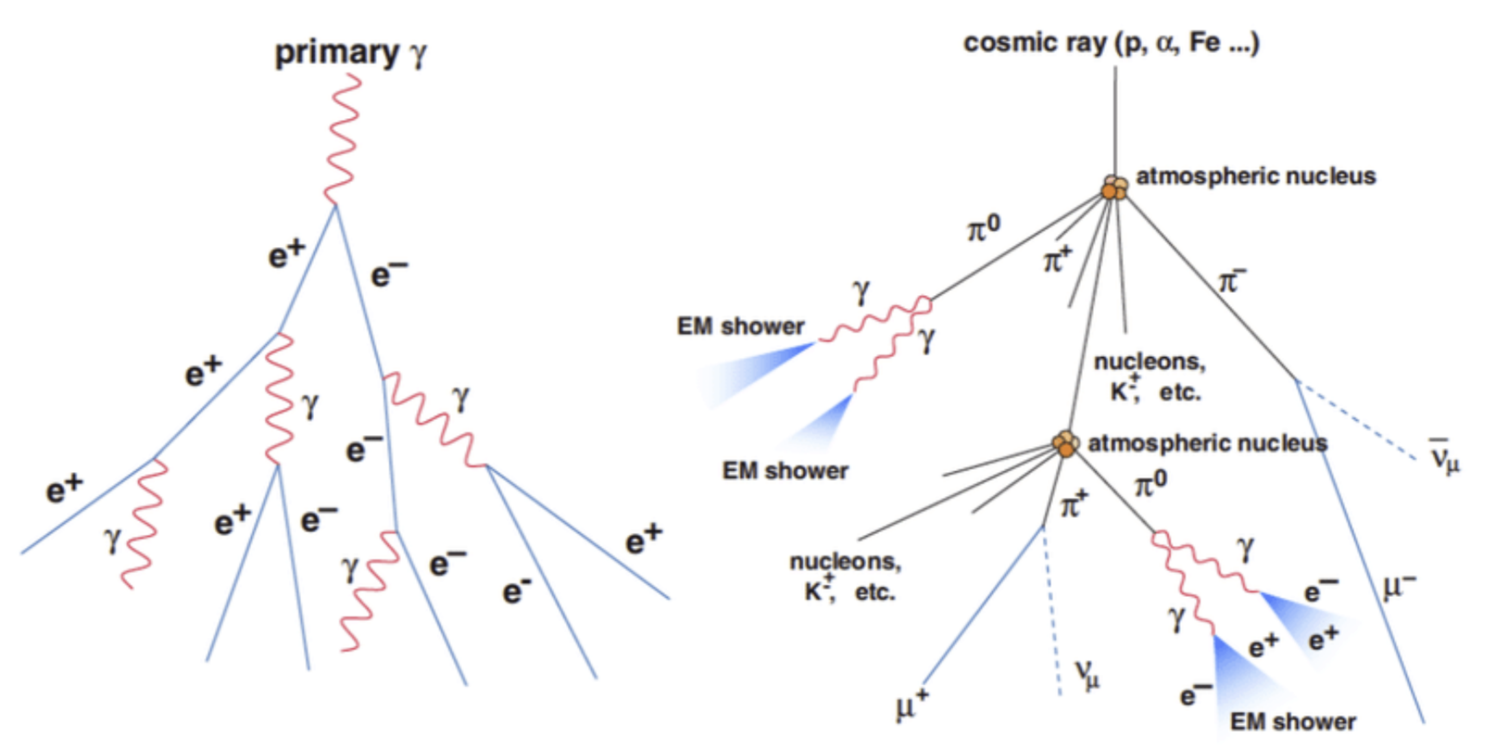


Fig. 1 Electromagnetic and hadronic showers

—

Figure from 2310.01857

SWGO: a telescope for gamma rays

- Southern Wide-Field Gamma-ray Observatory (SWGO)
  - Gamma-ray fluxes in the TeV-PeV range in the southern hemisphere
- Footprint size on the ground depends on energy of primary
  - via position of maximum development of the shower

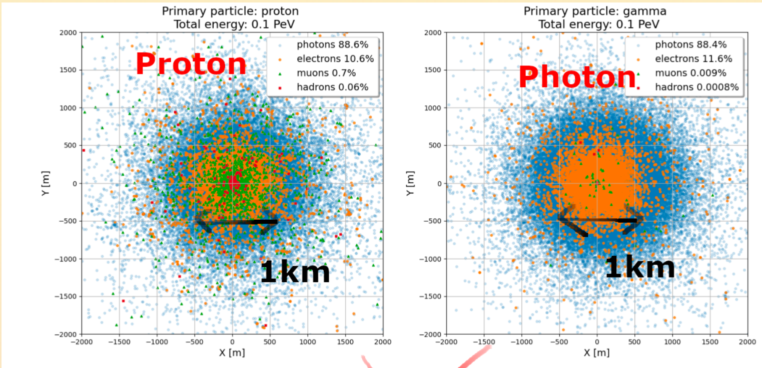


Fig. 2 Radial distribution of secondary particles on the ground, originated at 4.8km altitude by a 100 TeV proton (left) or photon sho (right).

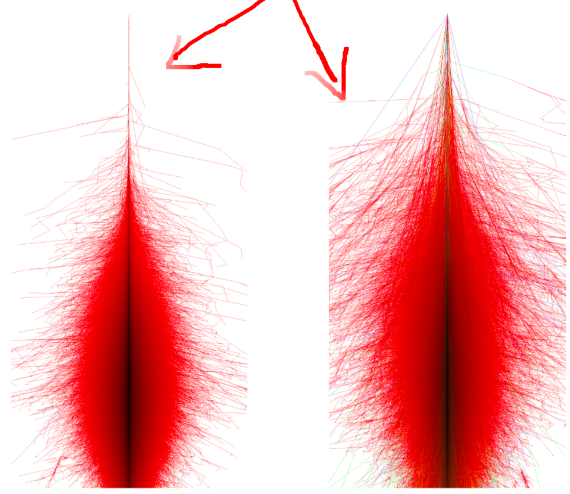


Fig. 3 Lateral development of 100 TeV showers from a primary photon (left) and proton (right). Source: Corsika web page[6]

Cherenkov tanks...

- Varying performance in separating EM, muonic, and hadronic shower constituents

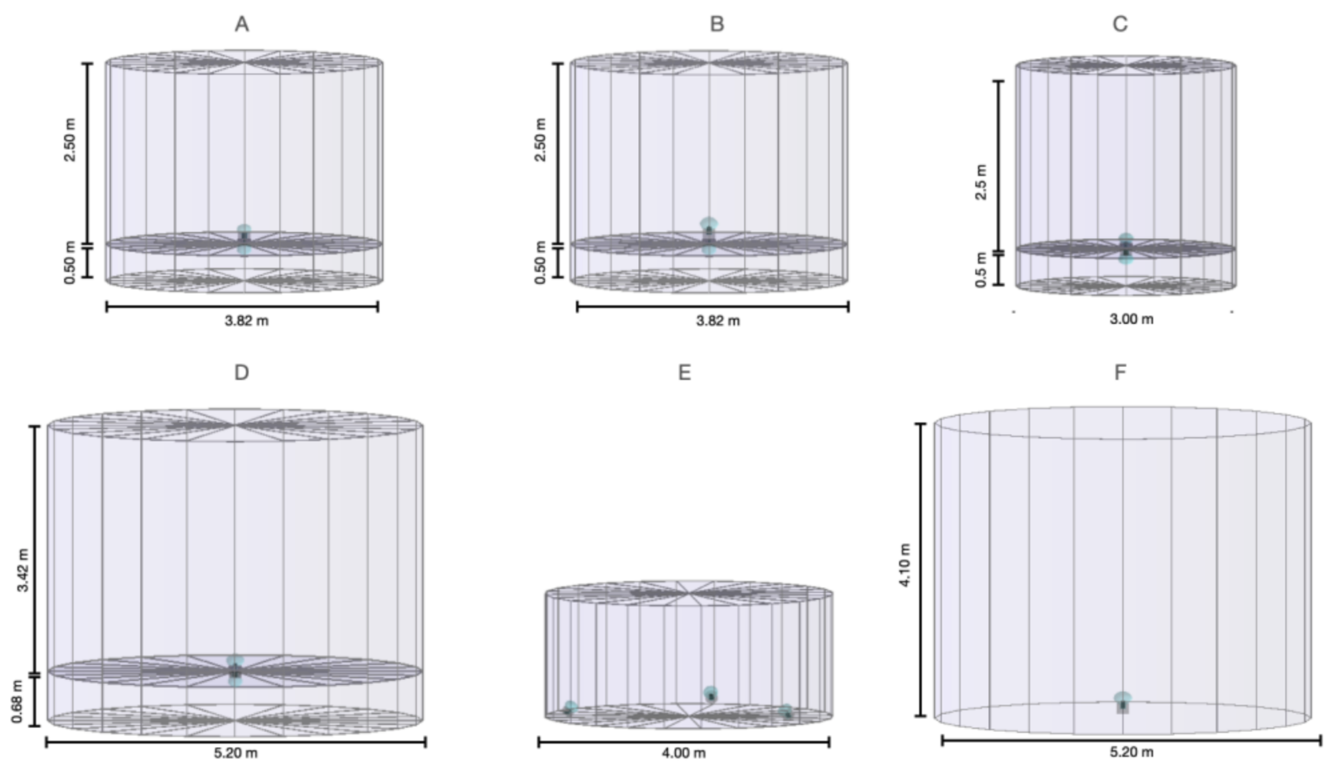
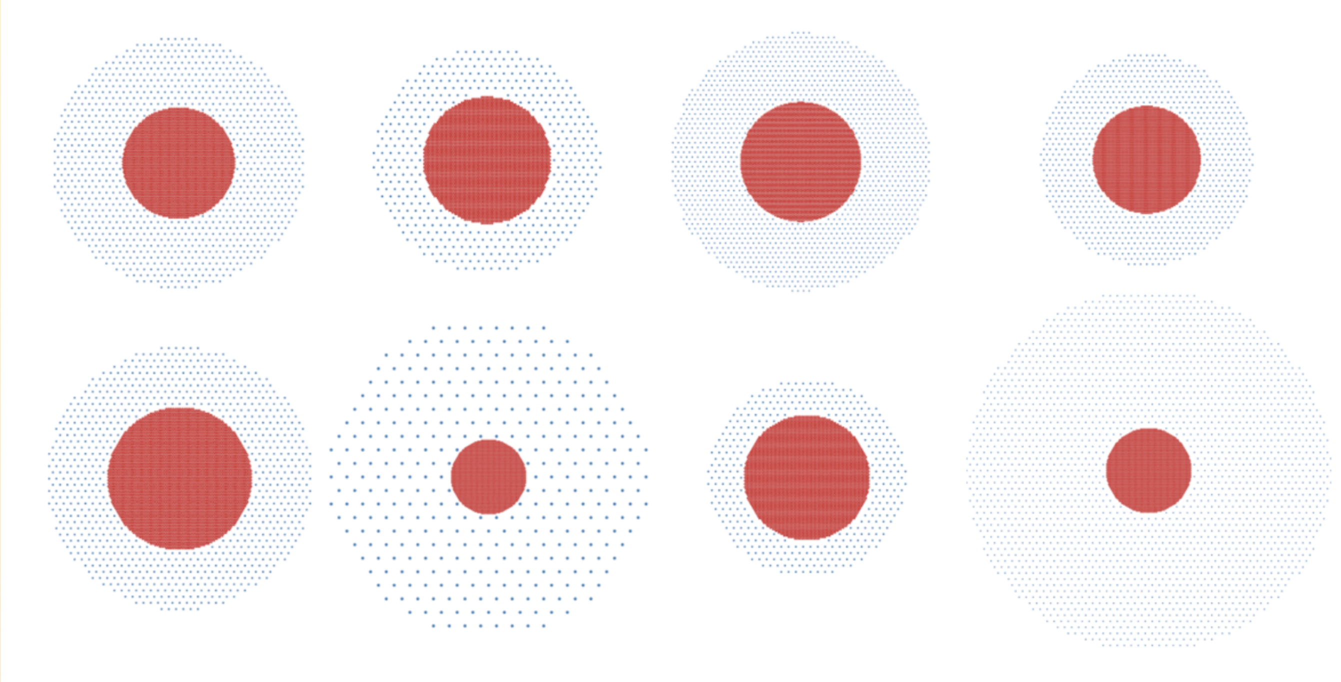


Fig. 4 Schematic design of different SWGO tank design. REF

—
Figure from 2310.01857

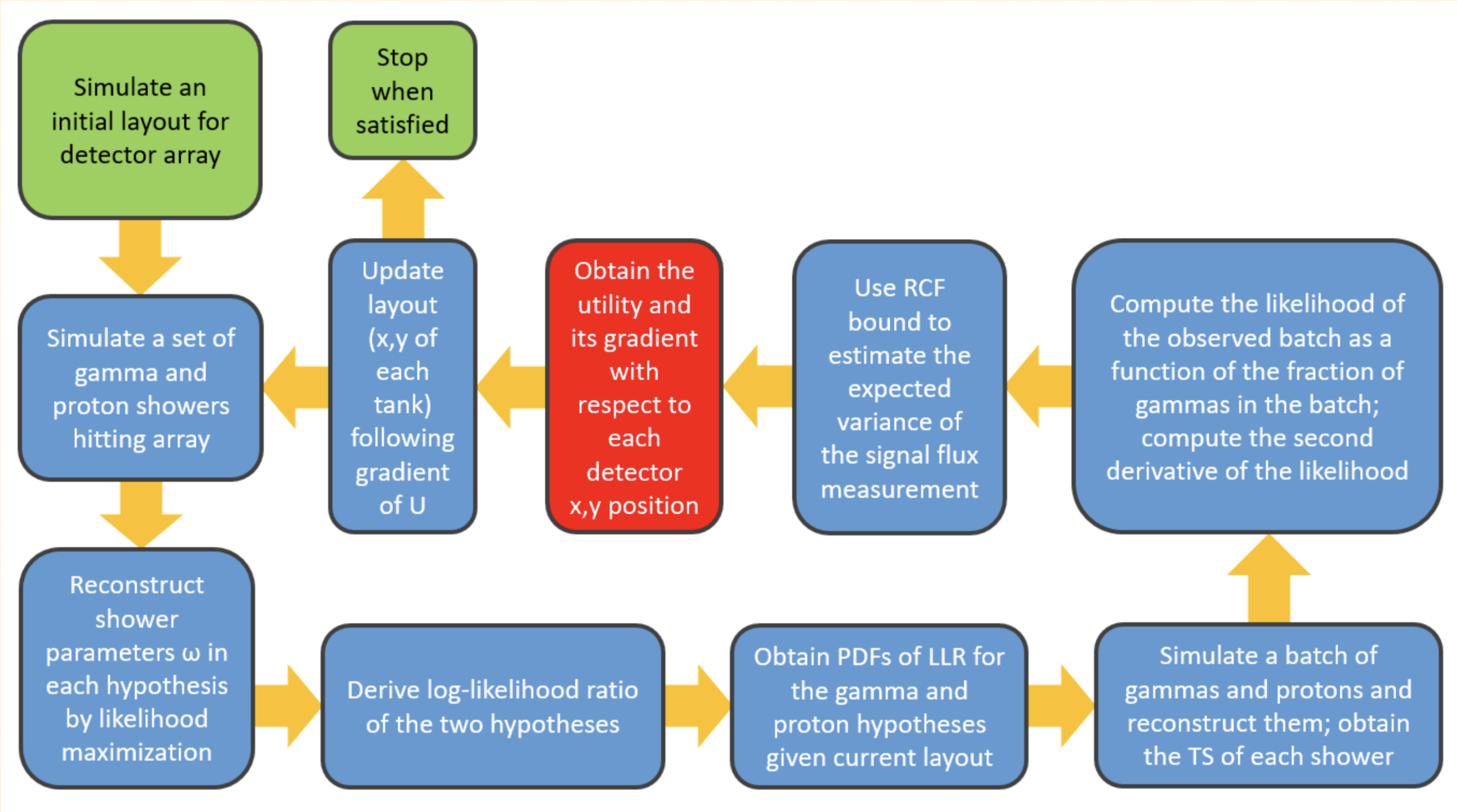
...and where to ~~find~~ place them

- Constraints on the total budget
    - High fill factor: better energy resolution, low sensitivity to $> 1PeV$ photons
    - Low fill factor: higher sensitivity to $> 1PeV$ photons, but poor energy reconstruction



—
Figure from 2310.01857

SWGO array optimization pipeline



—
Figure from 2310.01857

SWGO array optimization

- Stable and virtually identical results regardless of starting point
  - Very clear minimum reached

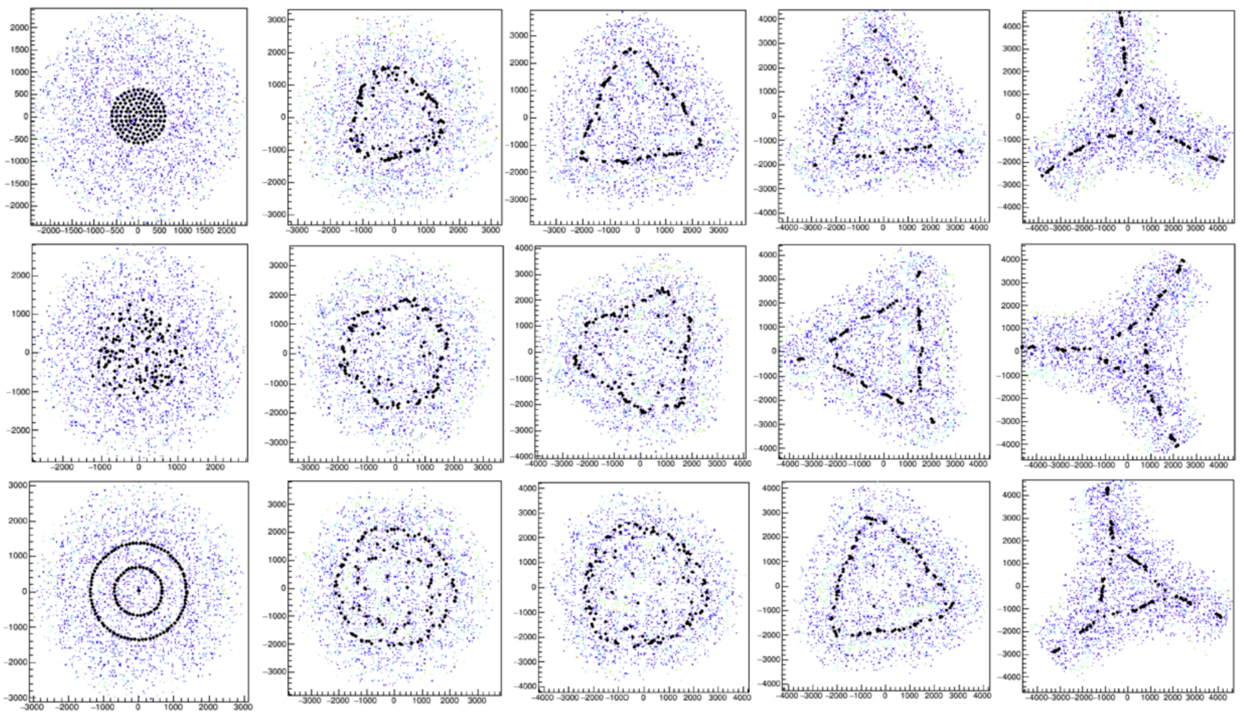


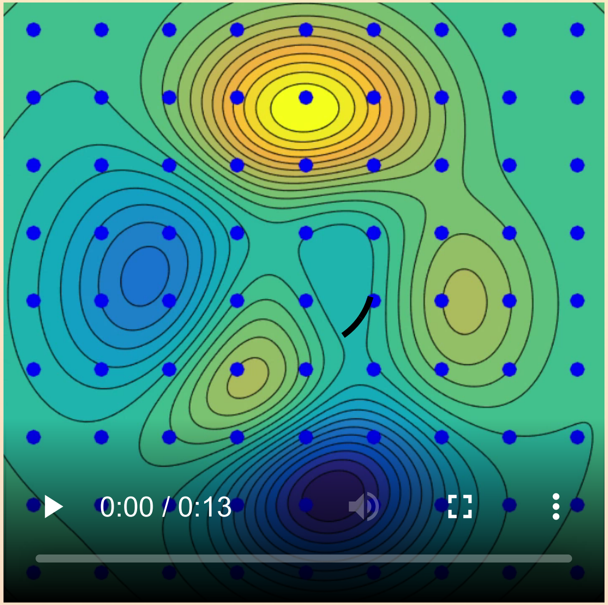
Figure 14: Convergence of three initial layouts (top to bottom: packed ball, wide random ball, two annuli) during a 500-epochs training. From left to right, the configurations of 126 units (129 in the bottom one) are shown at epoch 1,50,150,300, 500. See the text for more detail.

—

Figure from 2310.01857

Experimental design: present and future

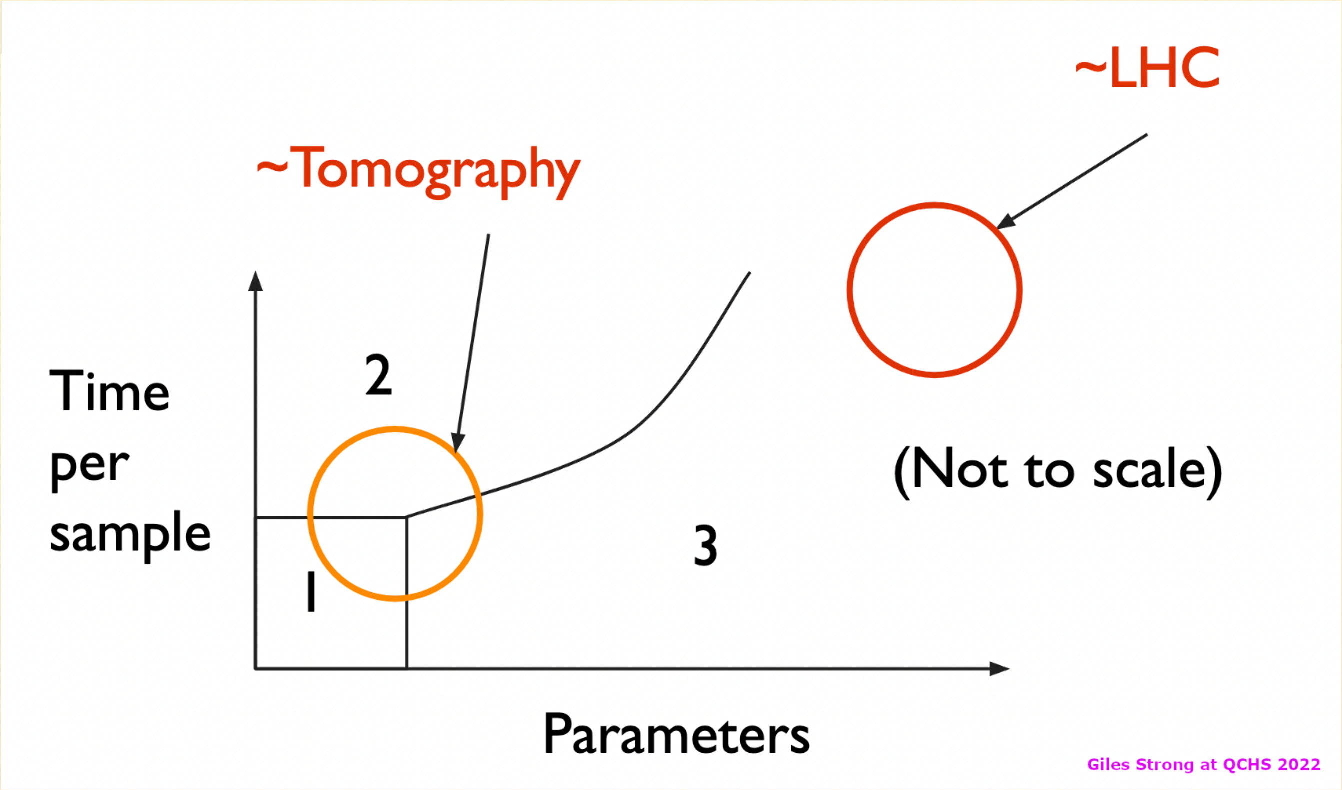
- Gradient descent applied to experiment design works!!!
    - Discreteness and stochasticity mostly solvable or avoidable
- What now?



—

Figure by Gpeyre on wikipedia

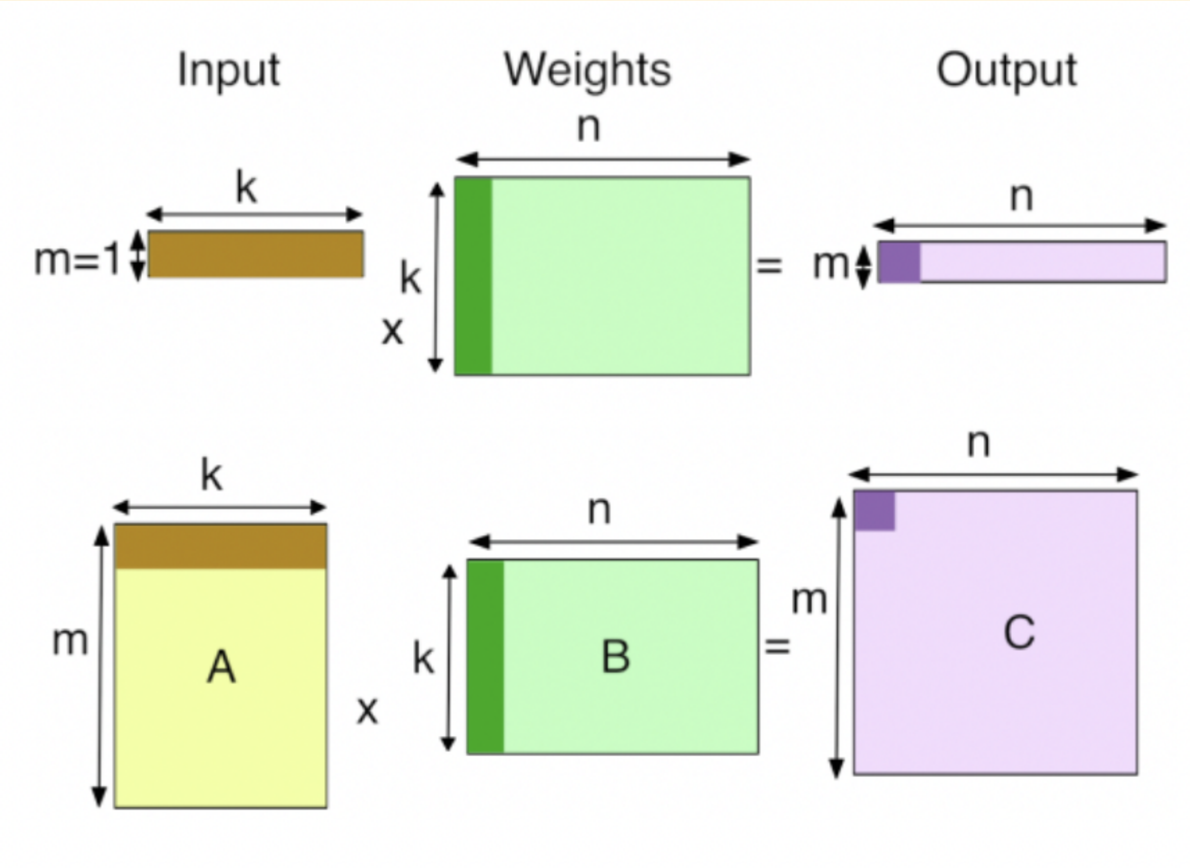
Method of choice depends on scale



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

From perceptron-based networks...

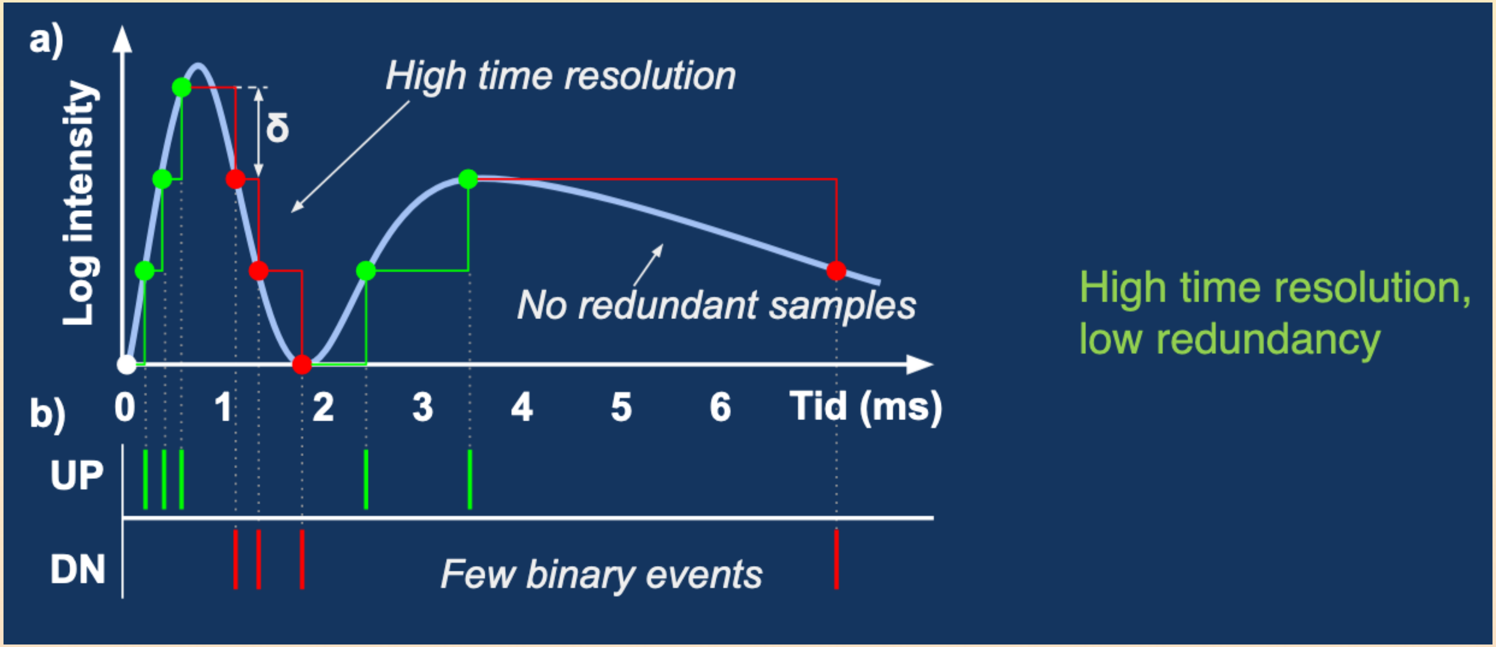
- Matrix multiplication



—
Figures by Gilles Louppe

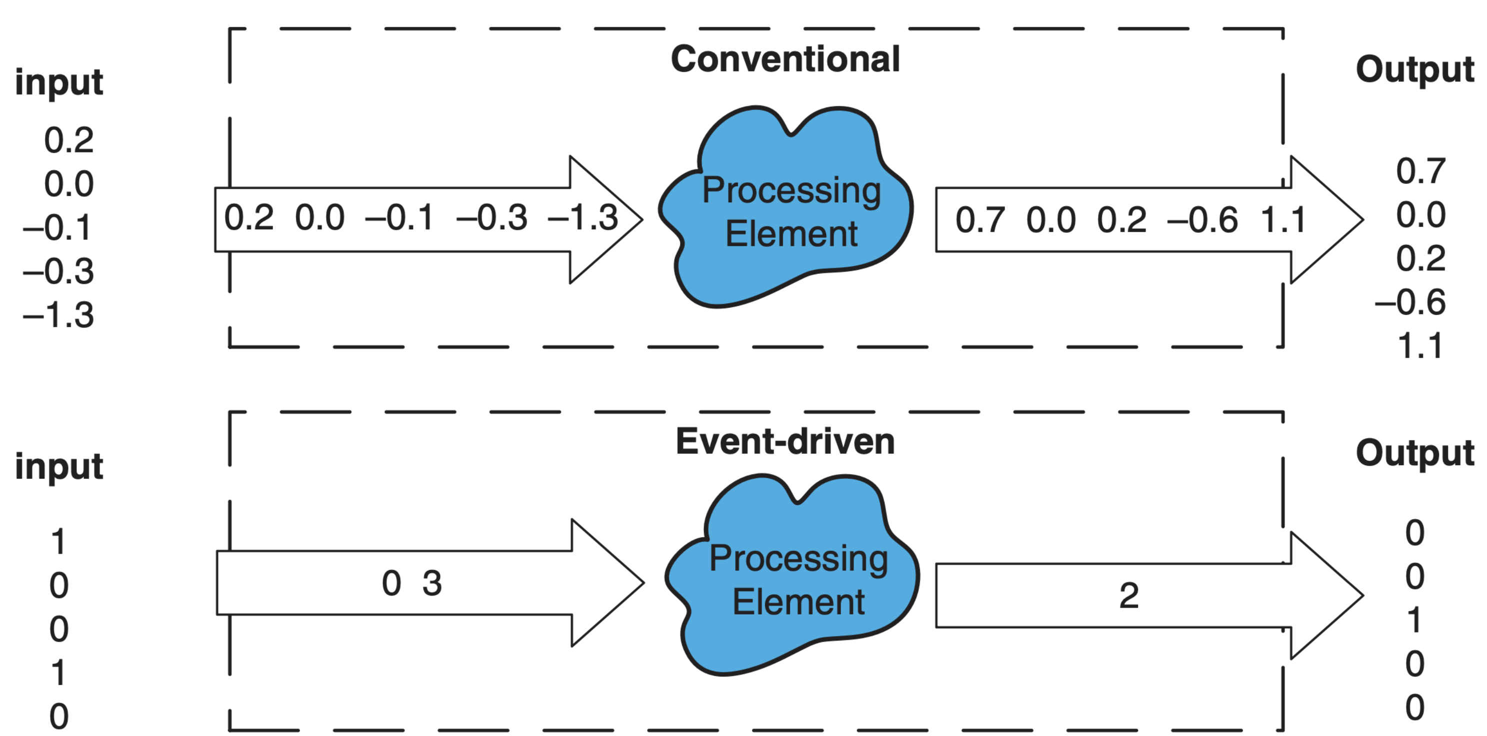
...to spiking neural networks

- Event-driven computations
  - "when a spike occurs, compute something"



The energy advantage

- Perceptron-based networks: matrix multiplication
  - Sparsity doesn't affect much the throughput and energy consumption
- Spiking neural networks: event-driven computations
  - Sparser inputs require less computations, therefore less time and energy



—

Image from Zheng, Mazumder (Wiley, 2019)

Encode information with Qubits

- Random bit (Bernoulli random variable) whose description is not governed by classical probability theory but by quantum mechanics
- Not only "because it can take real values in $[0, 1]$": complex numbers as coefficients α and β create interference
    - Interference is not reproducible with classical bits

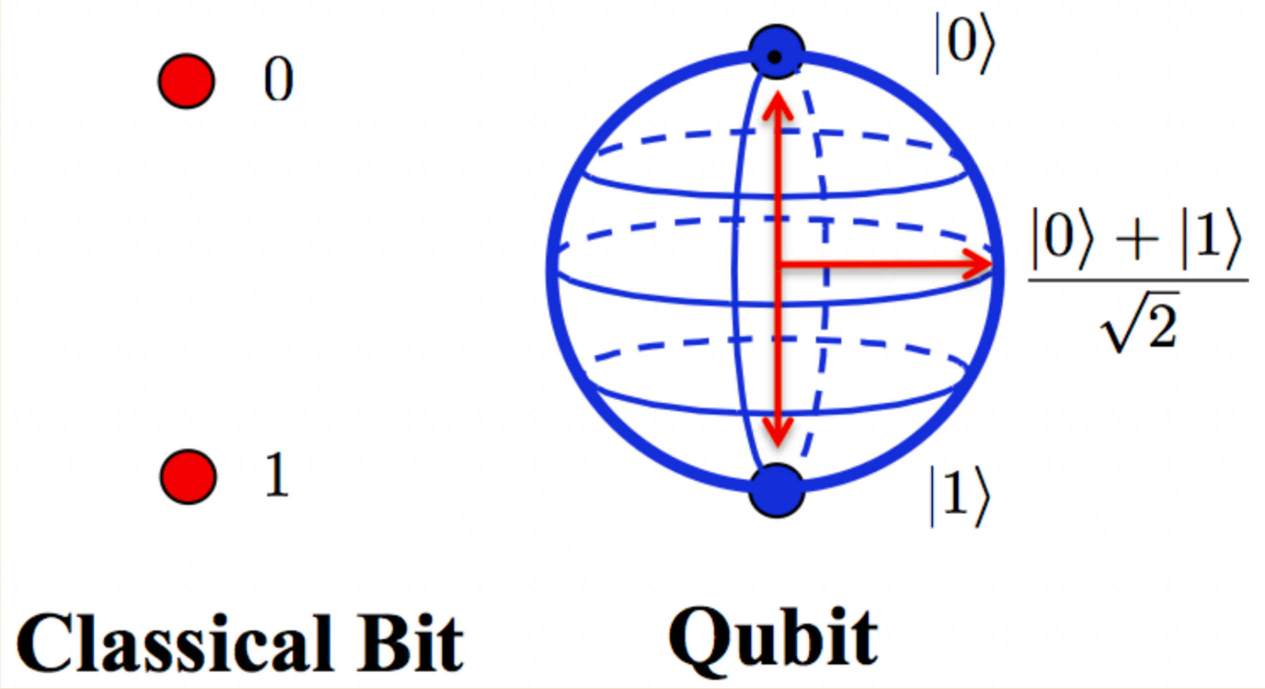
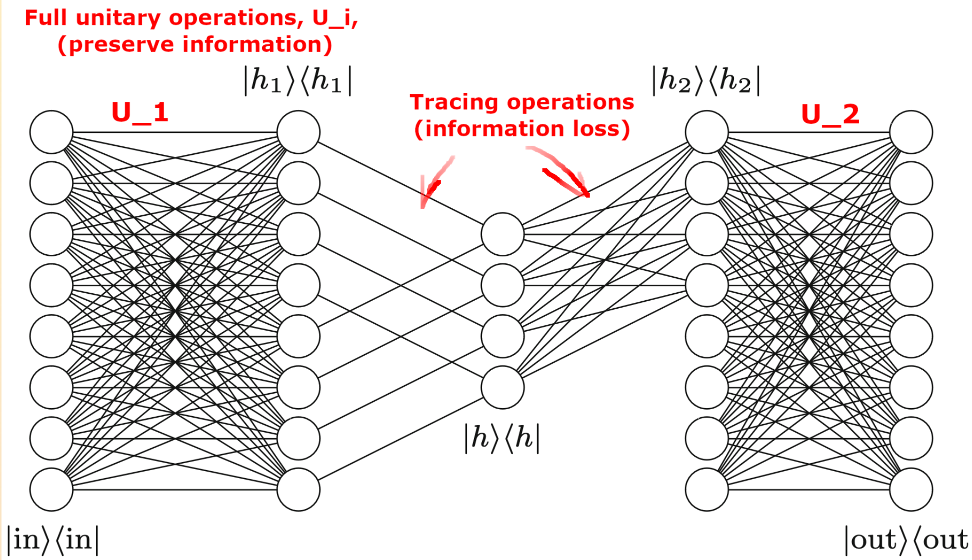


Image by prateekvjoshi

Represent neural networks

- Qubit operations can represent rather naturally neural networks

$$|\mathrm{in}\rangle\langle\mathrm{in}| = |q_1, q_2, q_3\rangle\langle q_1, q_2, q_3|.$$



- Gradient descent exploits intrinsic analytic differentiability of quantum circuits

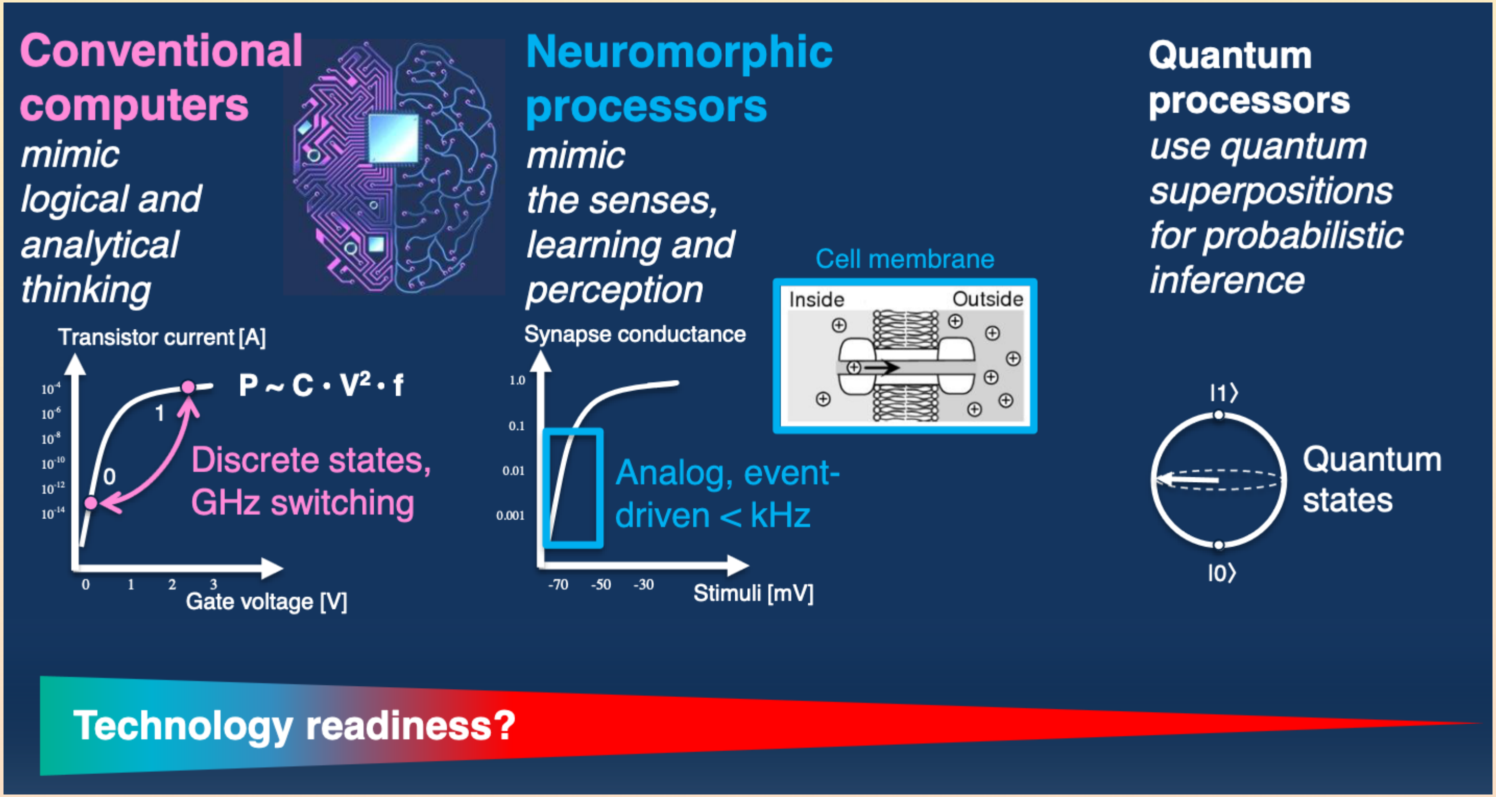
$$\begin{aligned}
\partial_\mu \langle\psi(x,\theta)|\sigma_z|\psi(x,\theta)\rangle &= \langle 0|\ldots\partial_\mu \mathrm{e}^{-i\mu\sigma}\ldots\sigma_z\ldots\mathrm{e}^{i\mu\sigma}\ldots|0\rangle \\
&\quad + \langle 0|\ldots\mathrm{e}^{-i\mu\sigma}\ldots\sigma_z\ldots\partial_\mu \mathrm{e}^{i\mu\sigma}\ldots|0\rangle \\
&= \langle 0|\ldots(-i\sigma)\mathrm{e}^{-i\mu\sigma}\ldots\sigma_z\ldots\mathrm{e}^{i\mu\sigma}\ldots|0\rangle \\
&\quad + \langle 0|\ldots\mathrm{e}^{-i\mu\sigma}\ldots\sigma_z\ldots(i\sigma)\mathrm{e}^{i\mu\sigma}\ldots|0\rangle \\
&= \langle 0|\ldots(1-i\sigma)\mathrm{e}^{-i\mu\sigma}\ldots\sigma_z\ldots(1+i\sigma)\mathrm{e}^{i\mu\sigma}\ldots|0\rangle \\
&\quad + \langle 0|\ldots(1+i\sigma)\mathrm{e}^{-i\mu\sigma}\ldots\sigma_z\ldots(1-i\sigma)\mathrm{e}^{i\mu\sigma}\ldots|0\rangle
\end{aligned}$$

—

Images from M. Schuld, F. Petruccione (2018)

Need for new paradigma

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



—
Image by Fredrik Sandin, Lulea University

The MODE Collaboration

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

- Joint effort
  - Particle physicists
  - Nuclear physicists
  - Astrophysicists
  - Computer scientists
  - Mathematicians
- If you are interested, join us!!!

At INFN and Università of 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**, and Mr. **Maxime Lagrange**

At Universidad de Oviedo and ICTEA Dr. **Pietro Vischia**

At Université Clermont Auvergne, Prof. **Julien Donini**, and Mr. **Federico Nardi** (joint with Universitá 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. **Sofia Vallecorsa**

At Karlsruher Institut für Technologie, Dr. **Jan Kieseler**

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 HEPHY Vienna (OeAW) 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 University of Kaiserslautern-Landau Mr. **Max Aehle**, Prof. **Nicolas Gauger**, Dr. **Lisa Kusch**

At University of Applied Sciences Worms Prof. **Ralf Keidel**

At Princeton University Prof. **Peter Elmer**

At University of Washington Prof. **Gordon Watts**

At SLAC Dr. **Ryan Roussel**

At Lulea University of Technology Prof. **Fredrik Sandin** and Prof. **Marcus Liwicki**

At IGFAE and Universidad de Santiago de Compostela Prof. **Xabier Cid Vidal**

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**, UniOvi

Series of yearly workshop

- First installment in Louvain-la-Neuve (Belgium)
- Second installmentnt in Kolymbari (Greece)
  - 37 talks, 9 posters, one data challenge with prizes, recordings will be online soon
- Third installment in Princeton (USA)
- You are all invited to the Fourth installment, to be held in Valencia (Spain), 23-25 September 2024!!!

Thank you!

—
Photo by me