

Neuronas - Del pez eléctrico al diseño de los experimentos del futuro con inteligencia artificial

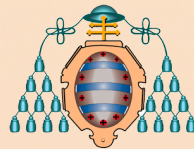
ICTEA Seminar at Universidad de Oviedo

Dr. Pietro Vischia

pietro.vischia@cern.ch

<https://vischia.github.io/>

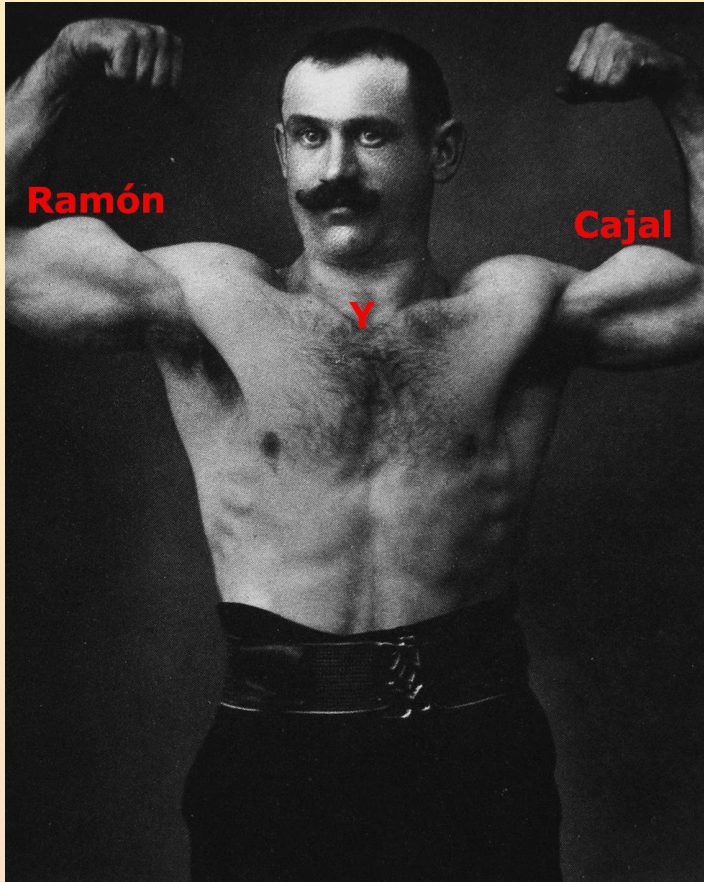
[@pietrovischia](#)



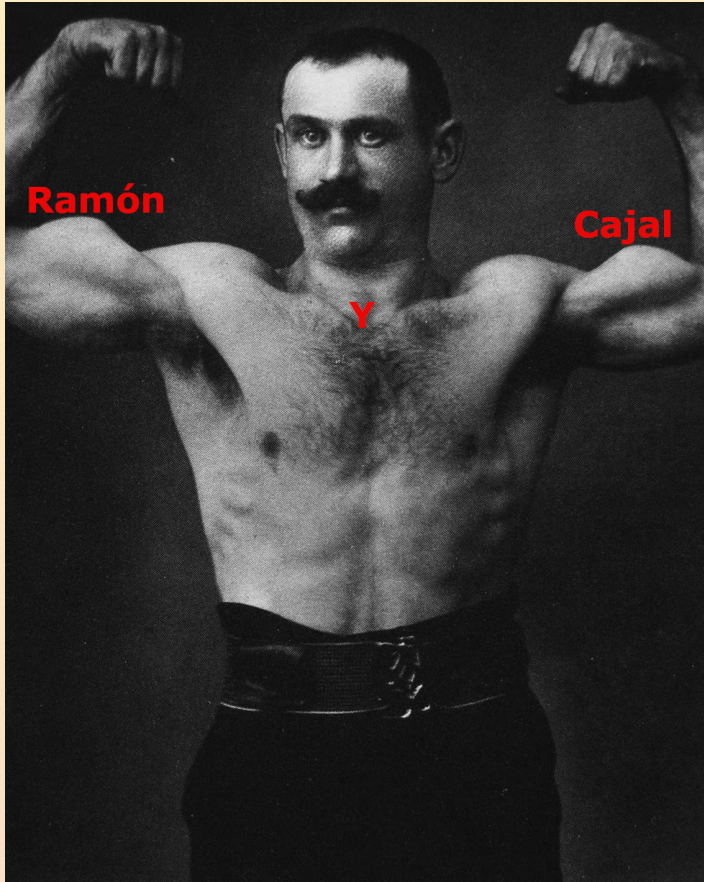
UNIVERSIDAD DE OVIEDO



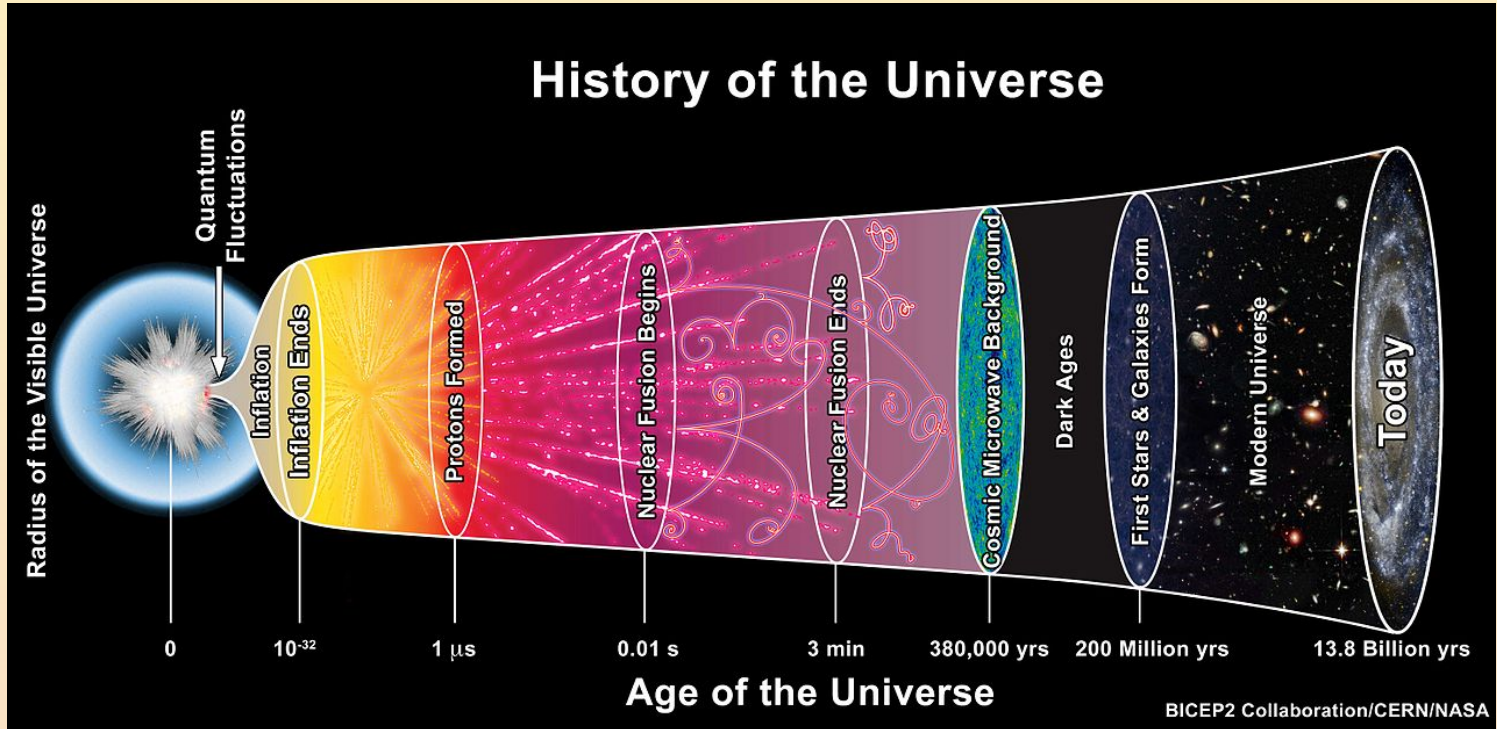
Santiago Ramón y Cajal



Santiago Ramón y Cajal

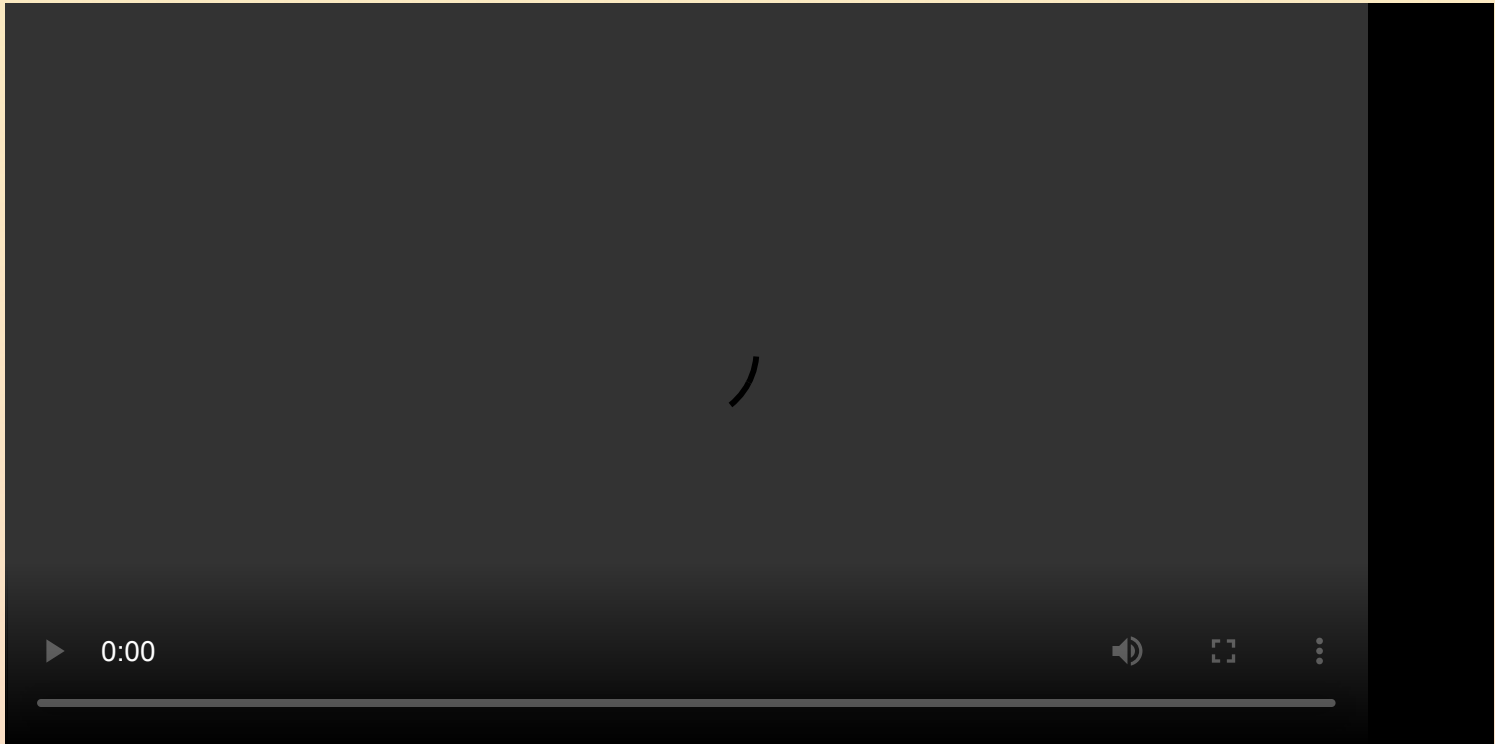


We Try to Understand the Universe

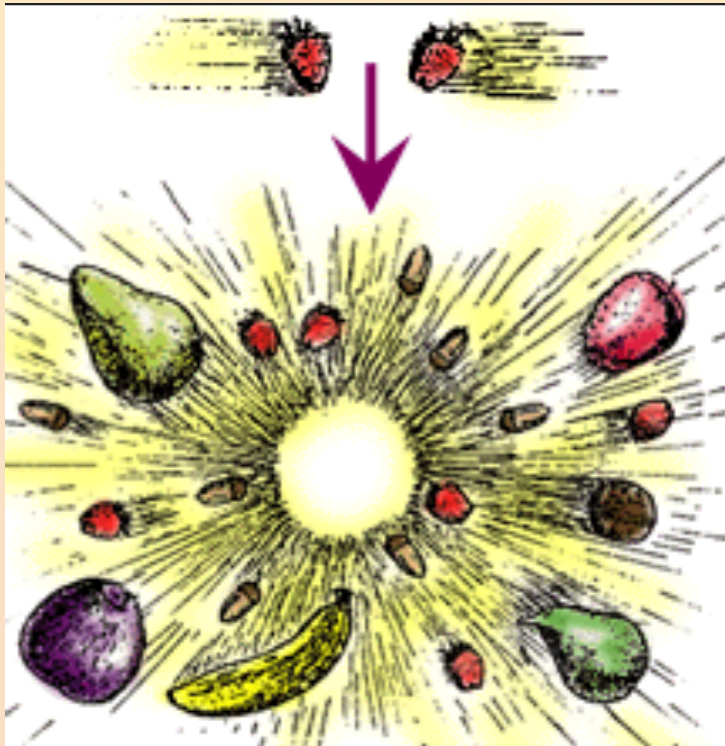


At CERN we collide protons

- The kinetic energy of two 88k tons aircraft carriers, each at 10km/h
- Packed into a transverse section of **16 micron**



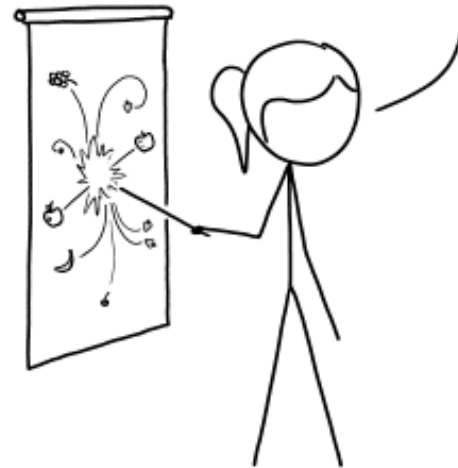
To See What Happens



WHEN TWO APPLES COLLIDE, THEY CAN BRIEFLY FORM EXOTIC NEW FRUIT. PINEAPPLES WITH APPLE SKIN. POMEGRANATES FULL OF GRAPES. WATERMELON-SIZED PEACHES.

THESE NORMALLY DECAY INTO A SHOWER OF FRUIT SALAD, BUT BY STUDYING THE DEBRIS, WE CAN LEARN WHAT WAS PRODUCED

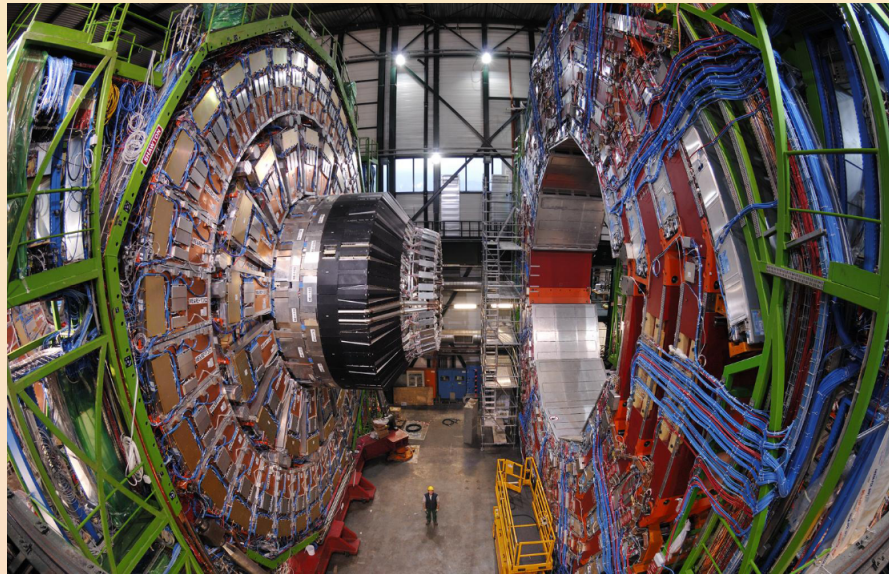
THEN, THE HUNT IS ON FOR A STABLE FORM.



HOW NEW TYPES OF FRUIT ARE DEVELOPED

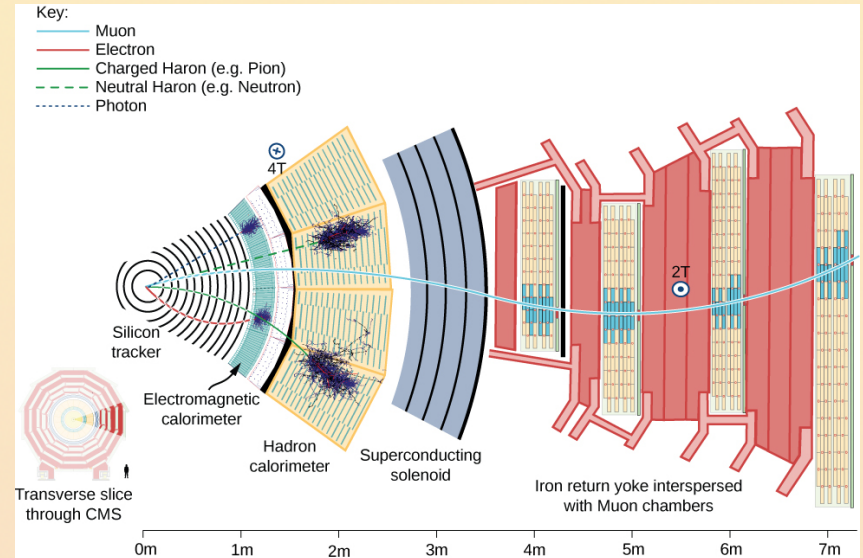
Complex Experiments

- Costly full simulation
- Interrelation of many parameters
- Large number of optimizable subdetectors
- Complexity prevents from optimizing targeting final goals



Robustness is not optimization

- 50+ y.o. detector design concepts served us well but may now be assisted by AI
- Track first, destroy later
- Redundancy in the detection systems
- Symmetrical layouts
 - No guarantee of optimality
- Subdetector-specific figures of merit



Optimality requires expertise

Automated Antenna Design with Evolutionary Algorithms

Gregory S. Hornby

hornby@email.arc.nasa.gov

University of California Santa Cruz, Mailtop 269-3, NASA Ames Research Center, Moffett Field, CA

Al Globus

San Jose State University

Derek S. Linden

JEM Engineering, 8683 Cherry Lane, Laurel, Maryland 20707

Jason D. Lohn

NASA Ames Research Center, Mail Stop 269-1, Moffett Field, CA 94035

Whereas the current practice of designing antennas by hand is severely limited because it is both time and labor intensive and requires a significant amount of domain knowledge, evolutionary algorithms can be used to search the design space and automatically find

"The current practice of designing and optimizing antennas by hand is limited in its ability to develop new and better antenna designs because it requires significant domain expertise and is both time and labor intensive."

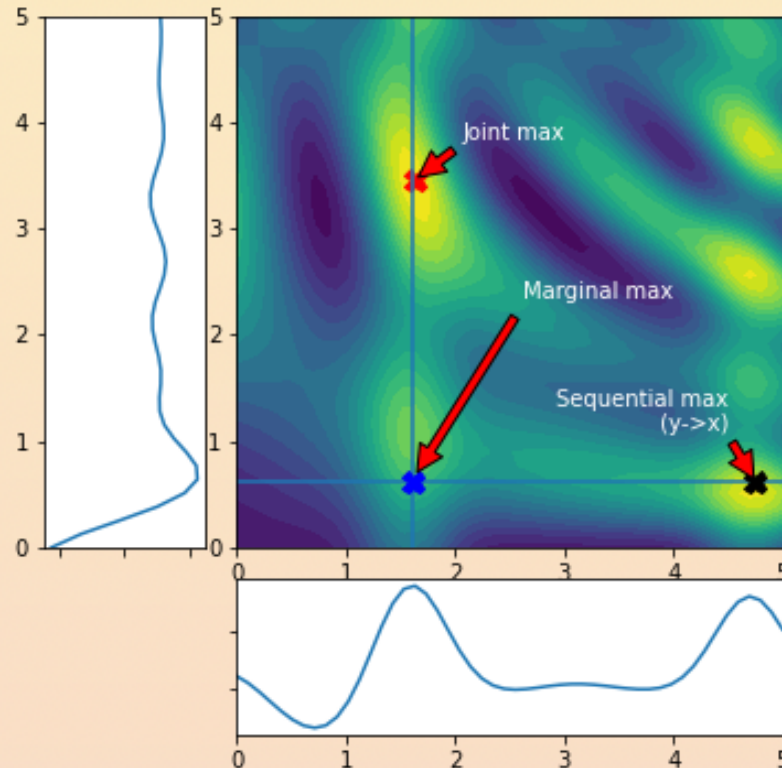


Joint optimization: why?

- Yields in general different solution than optimization of individual features
 - Both marginally and sequentially

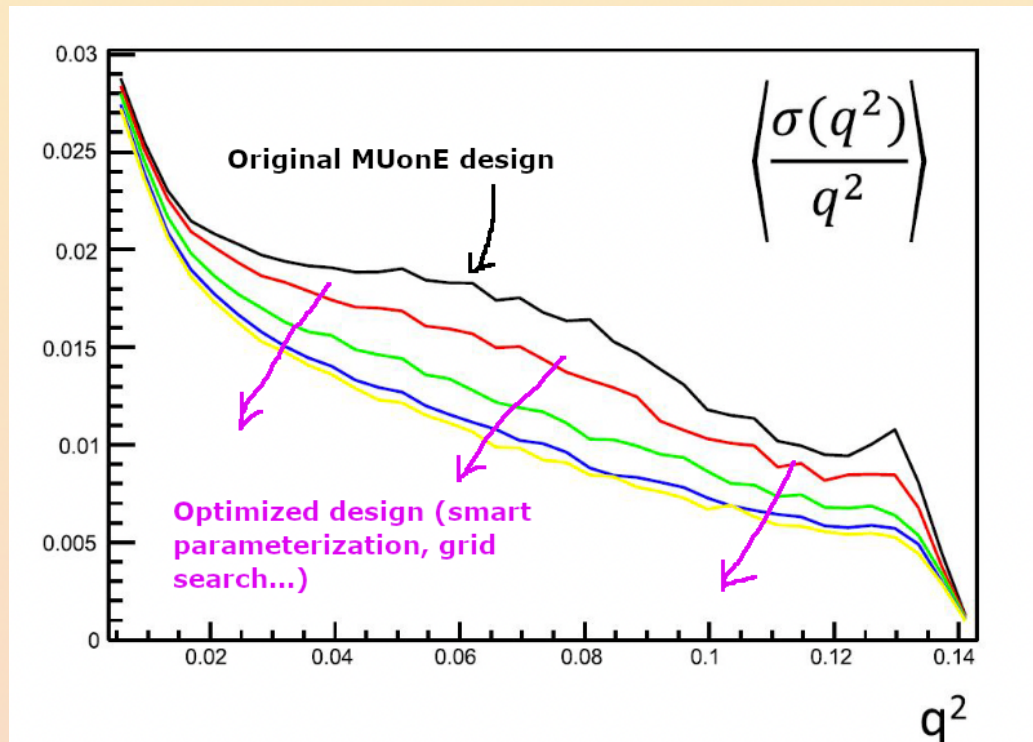
$$\operatorname{argmin}_{x,y} \left(\mathcal{L}(x,y) \right)_x \neq \operatorname{argmin}_x \left(\mathcal{L}(x,y) \right)$$

$$\operatorname{argmin}_{x,y} \left(\mathcal{L}(x,y) \right)_y \neq \operatorname{argmin}_y \left(\mathcal{L}(x,y) \right)$$

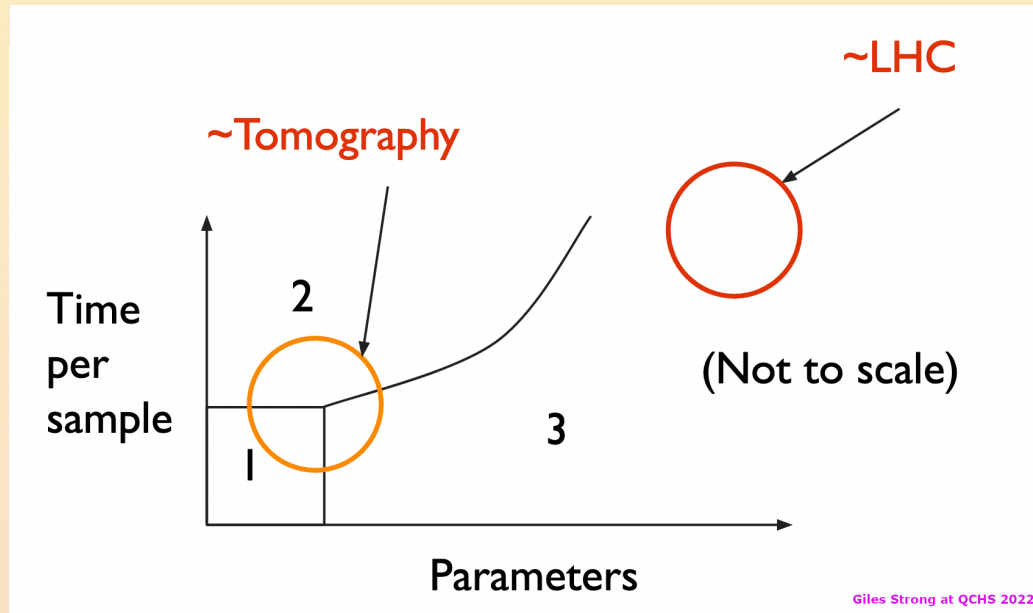


Large gains to be had

- 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



Different challenges require different methods



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

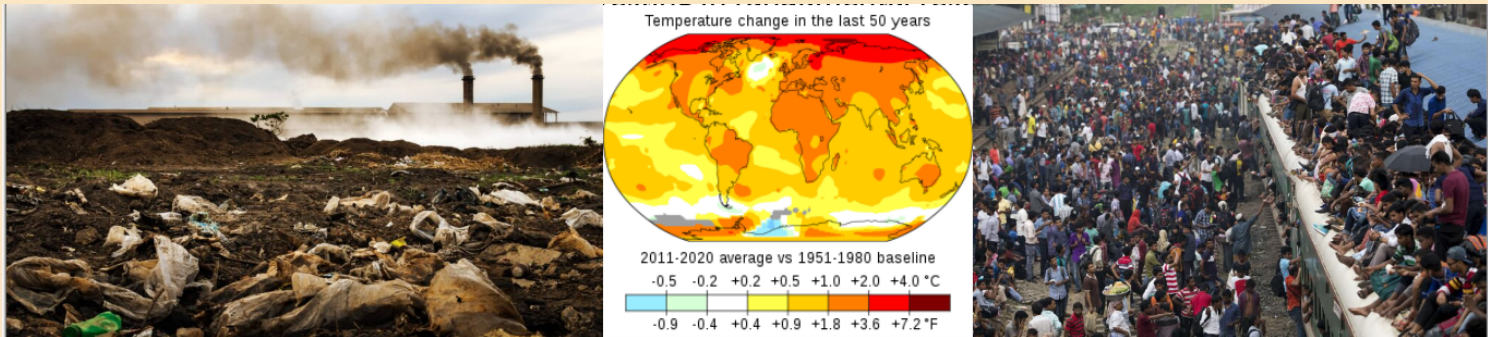
A moral imperative

Optimize...

- New large, long-term projects
- Push technological skills to the limit (cit. EUSUPP)

...within constraints

- Unprecedented global challenges
- Society less receptive to fundamental research



Maximum extraction of scientific value from the available resources

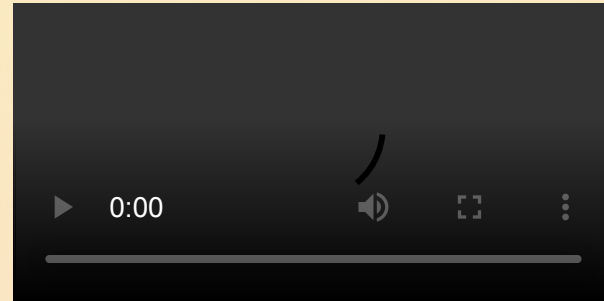
Gradient Descent

- 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

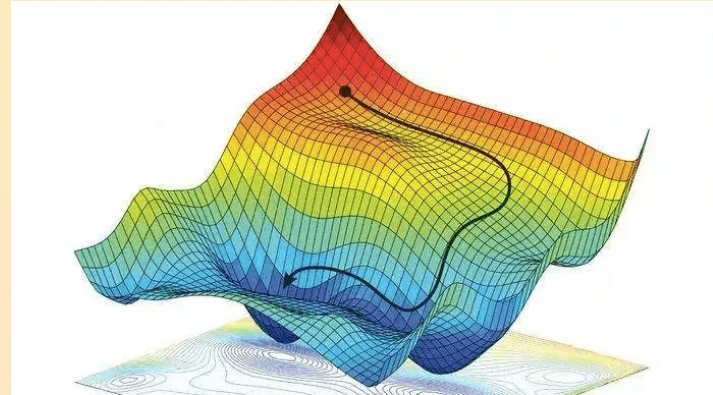
$$\hat{L}(f) = \frac{1}{n} \sum_{i=1}^n |f(x_i) - f^*(x_i)|^2$$

- Generalization
(for learning problems)

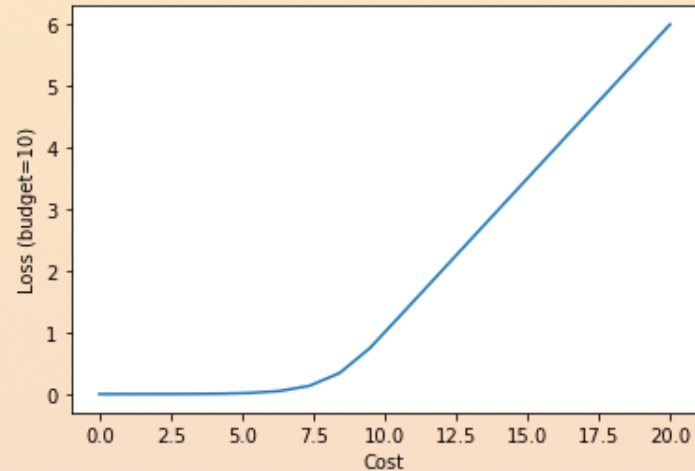


Finite Budget: loss and constraints

- Optimization via gradient descent
 - Target-oriented loss functions
- Constraints inserted as penalization
 - Additional term to the loss



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



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

Optimization has practical consequences

- Material availability (influenced e.g. also by wars) is also a concern, nowadays



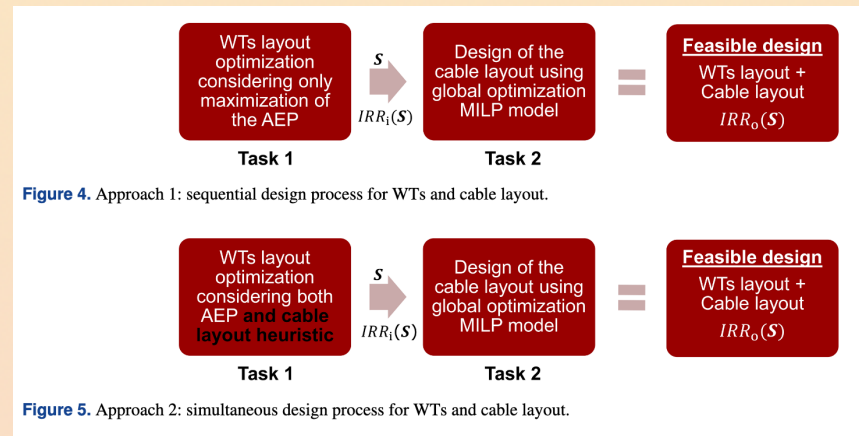
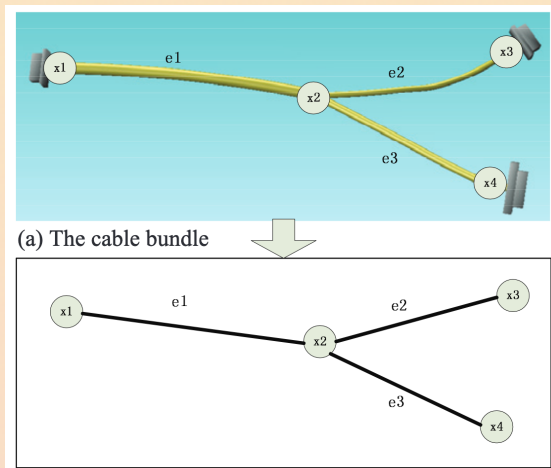
Figure 6: Road transport of a structure for the ATLAS air toroids. Photo reproduced from Ref. [181].

If you can't turn it on, it's not optimal



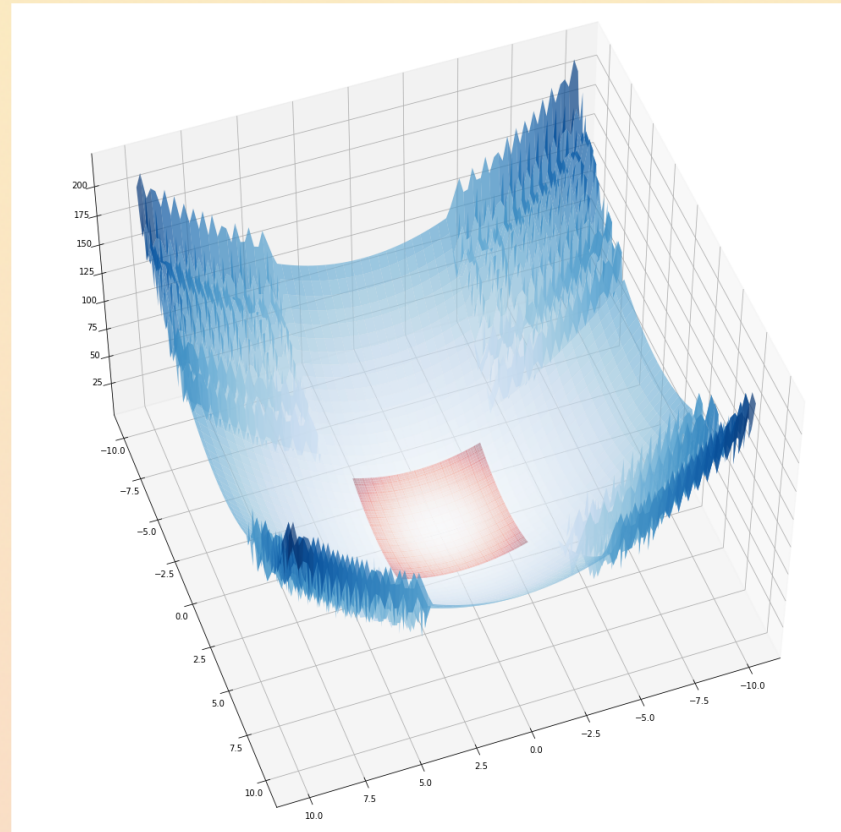
Maybe we can optimize cable layout

- Easily description as trees or graphs
- Although intrinsically discontinuous and nonsmooth
 - Mostly gradient-free tree searches
- Maybe further studies on the loss landscape can help in solving this in a differentiable way



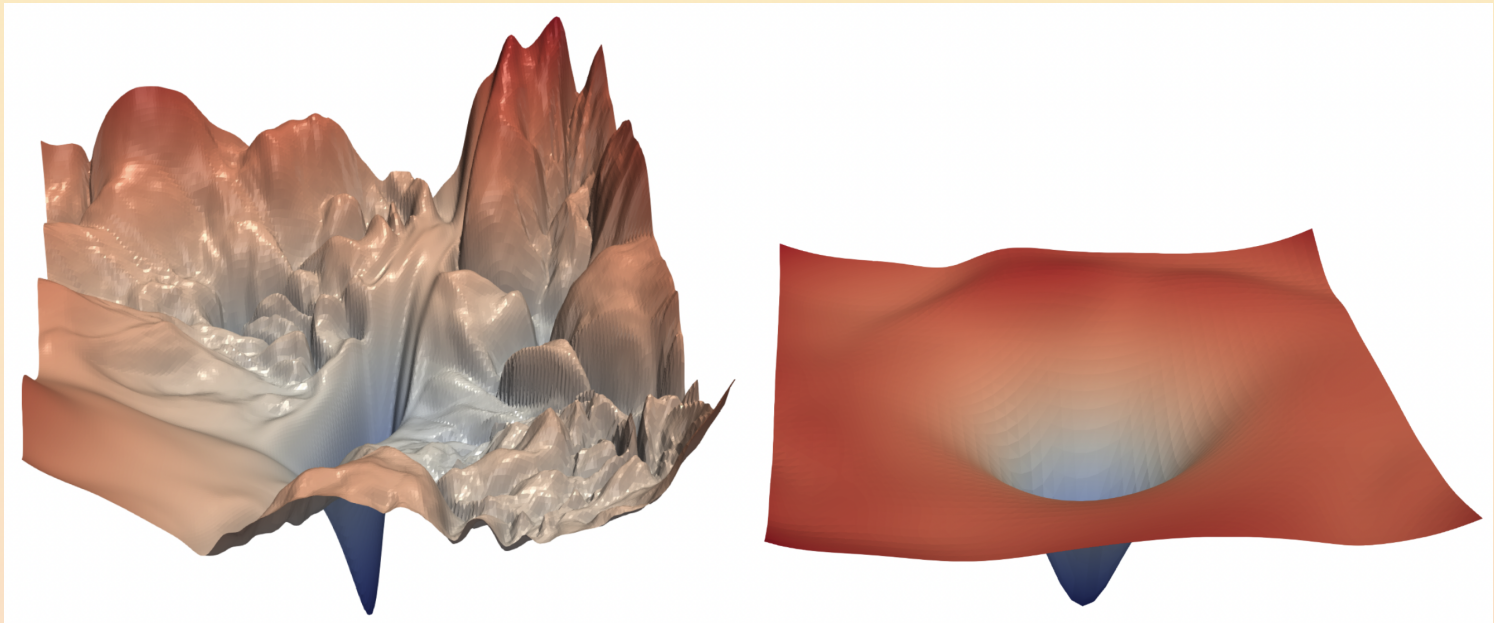
Assist the physicist with a landscape of solutions

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



"Not too far" from optimality

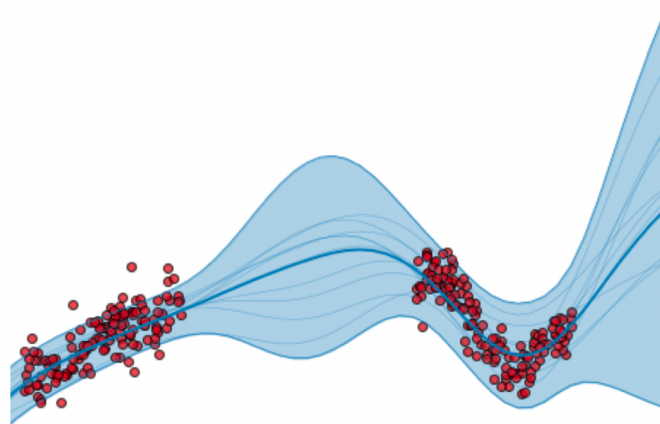
- Can we define in a general way an acceptable **increase in loss**?
 - Tradeoff performance/cost
 - Dynamically account for irregularity of the landscape



Maybe we should marginalize?

What is Bayesian learning?

- ▶ The key distinguishing property of a Bayesian approach is **marginalization** instead of optimization.
- ▶ Rather than use a single setting of parameters \mathbf{w} , use all settings weighted by their posterior probabilities in a *Bayesian model average*.



Andrew Wilson at Hammers&Nails 2022

7/85

How to optimize an experiment

- We detailed our idea in the **MODE White Paper**
 - 109-page document drafting the way forward, joint with computer scientists from **proton Computed Tomography**
 - under revision for Reviews in Physics

Contents

1	Introduction	5
2	The State of the Art in Design Optimization and Differentiable Programming	9
3	Problem Description and Possible Solution	18
4	Example Use Cases	33
5	System Architecture and Requirements	80
6	Conclusions	83

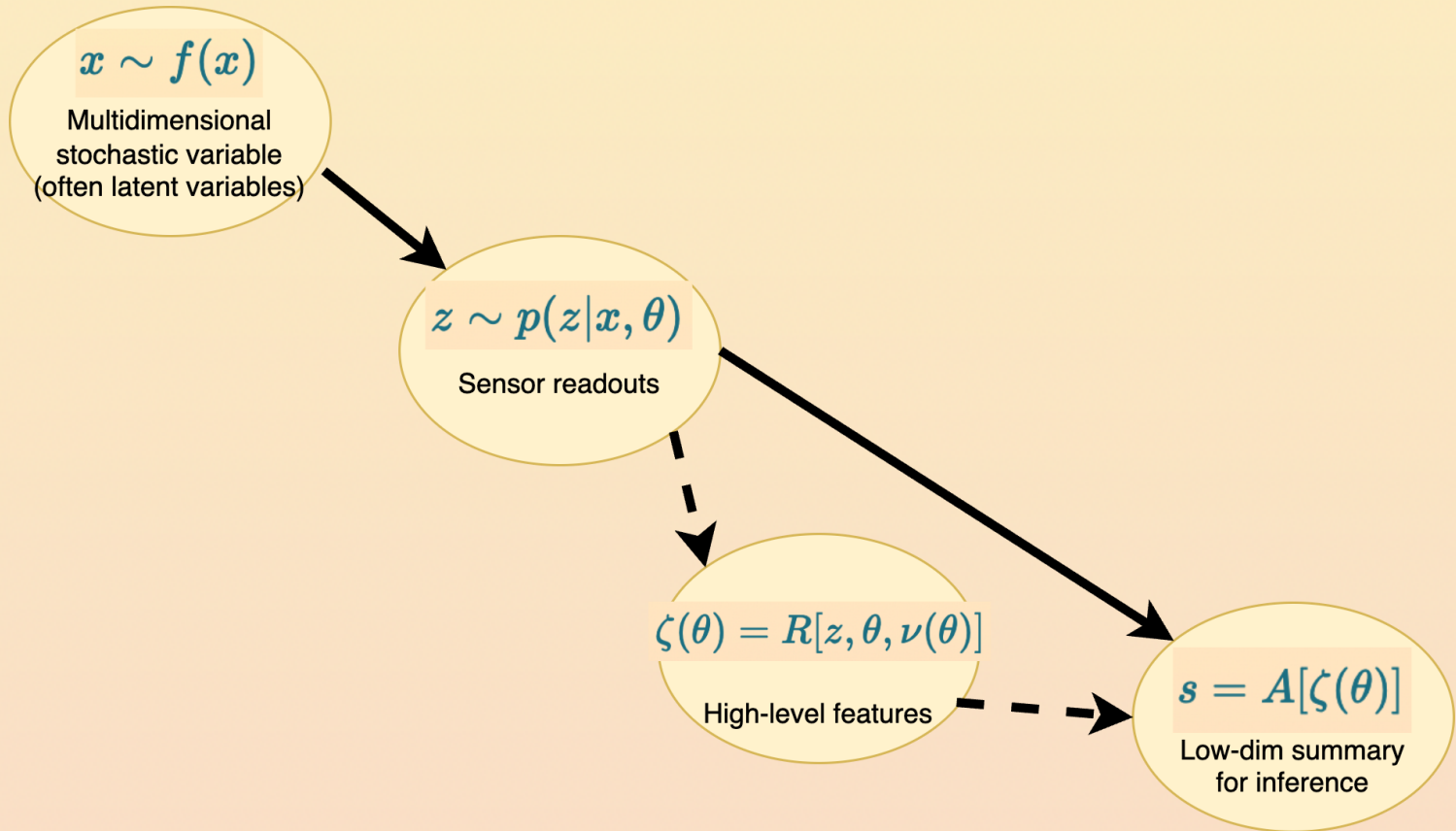
Toward the End-to-End Optimization of Particle Physics Instruments with Differentiable Programming: a White Paper

Tommaso Dorigo^{1,2}, Andrea Giammanco^{*1,3}, Pietro Vischia^{1,3} (editors), Max Ahle⁴, Mateusz Bawaj⁵, Alexey Boldyrev^{1,6}, Pablo de Castro Manzano^{1,2}, Denis Derkach^{1,6}, Julien Donini^{1,7}, Auralee Edelen⁸, Federica Fanzago^{1,2}, Nicolas R. Gauger⁴, Christian Glaser^{1,9}, Atılım G. Baydin^{1,10}, Lukas Heinrich^{1,11}, Ralf Keidel^{1,2}, Jan Kieseler^{1,13}, Claudius Krause^{1,14}, Maxime Lagrange^{1,3}, Max Lamparth^{1,11}, Lukas Layer^{1,2,15}, Gernot Maier¹⁶, Federico Nardi^{1,2,17,7}, Helge E. S. Pettersen¹⁸, Alberto Ramos¹⁹, Fedor Ratnikov^{1,6}, Dieter Röhrich²⁰, Roberto Ruiz de Austri¹⁹, Pablo Martínez Ruiz del Árbol^{1,21}, Oleg Savchenko^{2,3}, Nathan Simpson²², Giles C. Strong^{1,2}, Angela Taliencio³, Mia Tosi^{1,2,17}, Andrey Ustyuzhanin^{1,6}, and Haitham Zaraket^{1,23}

¹MODE Collaboration, <https://mode-collaboration.github.io/>

Ingredients

- Optimization metric to find values of θ that optimize inference made with s



Optimization recipe

Depends on z and nuisances

Cost of the layout with parameters θ

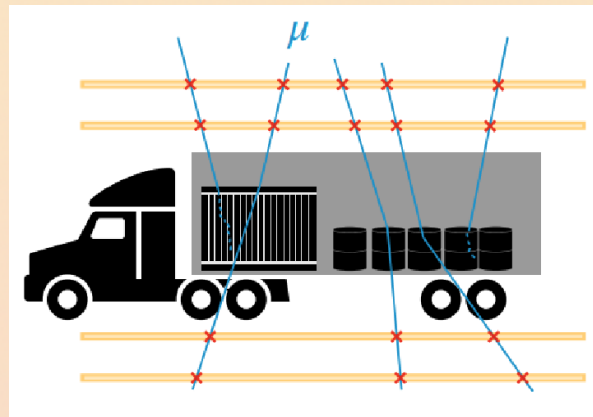
Closed form

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

Weight desirable goals while obeying cost constraints

- For example, to identify smuggled material in a container

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



Domain knowledge remains crucial

- "Just" parameterizing the systems in an optimal way (pun intended) may already yield improvements

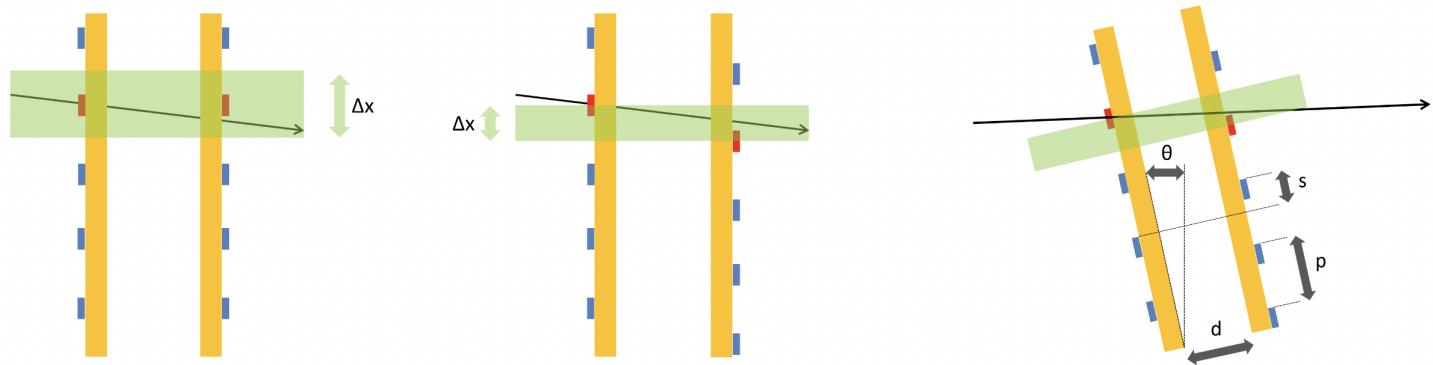


Figure 4: Left and center: a double-sided silicon strip sensor produces twice smaller resolution Δx on single-strip hit position for an orthogonally incident particle if strips on the two sides are staggered by half the strip pitch. Right: the four parameters affecting single-strip hit position resolution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

When the likelihood is intractable

- $p(\cdot)$ not in closed form
 - Sample $x_i \sim f(x)$
 - Then z_i distributed as emulator, $x_i \sim F(x_i, \theta)$

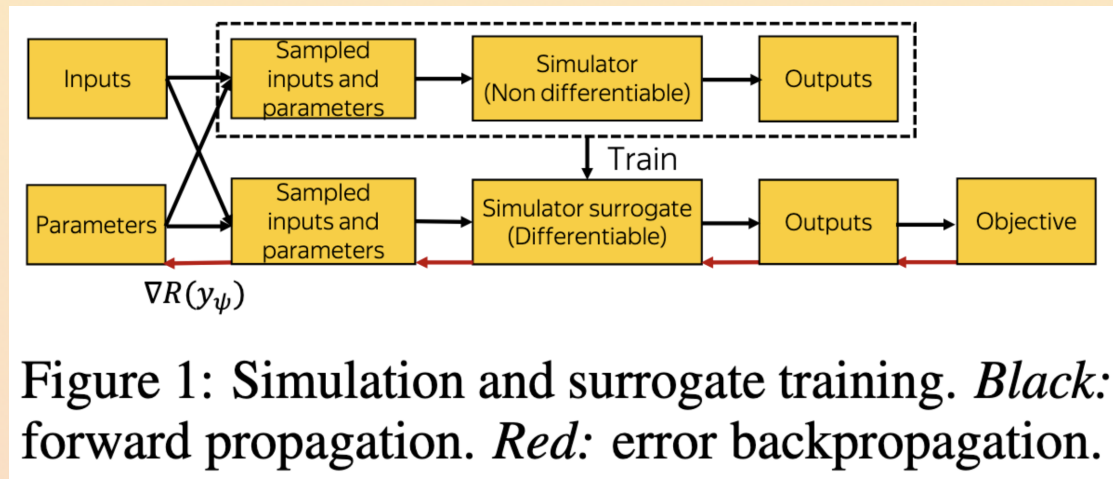
$$\hat{\theta}_{\text{approx}} = \arg \min_{\theta} \frac{1}{n} \sum_{i=1}^n L \left[A(R(z_i)), c(\theta) \right]$$

- $F(\cdot)$ nondifferentiable stochastic simulator
 - Replace with local surrogate $z = S(y, x, \theta)$, where y describes the stochastic variation of the approx distribution
 - Learn surrogate separately
 - Descend to the minimum of approximated loss by following surrogate gradient

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

Advantages of surrogates

- Subset of relatively simple class of functions (but they must be able to reproduce $F(\cdot)$ well)
- Learn by training (**hic sunt leones**), (but $N(\text{eval } F) \geq \mathcal{O}(\text{dim}(\theta))$)
- Automatically get AD out of the box even if original $F(\cdot)$ is not differentiable
- Evaluation of surrogate (for optimization) much faster than evaluation of $F(\cdot)$



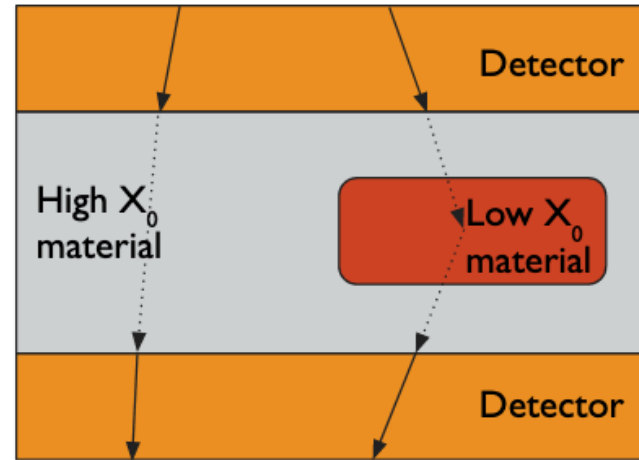
Vast set of use cases

- Already exploring
 - Muon tomography
 - LHCb and CMS calorimetry
 - SWGO placement/geometry of tanks
 - LEGEND optimization

- ▼ 4 Example Use Cases
 - ▼ 4.1 Experiments at Accelerators
 - 4.1.1 Particle Accelerator Design and Control
 - 4.1.2 Calorimeter Optimization
 - 4.1.3 Hybrid Calorimeter for a Future Particle Collider
 - 4.1.4 Electromagnetic Calorimeter of a Muon Collider Experiment
 - 4.1.5 Optimization of the MUonE Detector
 - 4.1.6 Searches for Milli-charged Particles
 - ▼ 4.2 Astro-particle Physics and Neutrino Experiments
 - 4.2.1 High-Energy Gamma-Ray Astronomy
 - 4.2.2 Interferometric Gravitational-Wave Detectors
 - 4.2.3 Radio Detection of High-Energy Neutrinos
 - ▼ 4.3 Cosmic-Ray Muon Imaging
 - 4.3.1 Figures of Merit
 - 4.3.2 Parameters of the Optimization Task
 - 4.3.3 TomOpt: Differential Muon Tomography optimization
 - 4.3.4 Industrial Applications
 - 4.3.5 Portable Modular Detectors for Flexible Muography
 - 4.4 Proton Computed Tomography
 - 4.5 Low-Energy Particle Physics
 - 4.6 Error Analysis of Monte Carlo Data in Lattice QCD

Muon Tomography

- 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

TomOpt

- Differential optimization of muon-tomography detectors (ongoing project)
 - Giles C. Strong, Tommaso Dorigo, Andrea Giammanco, Pietro Vischia, Jan Kieseler, Maxime Lagrange, Mariam Safiieldin, Federico Nardi, Anna Bordignon, Haitham Zaraket, Max Lamparth, Federica Fanzago, Oleg Savchenko, Nitesh Sharma
 - Modular design in python, automatic differentiation via PyTorch

- Inference chain as differentiable pipeline

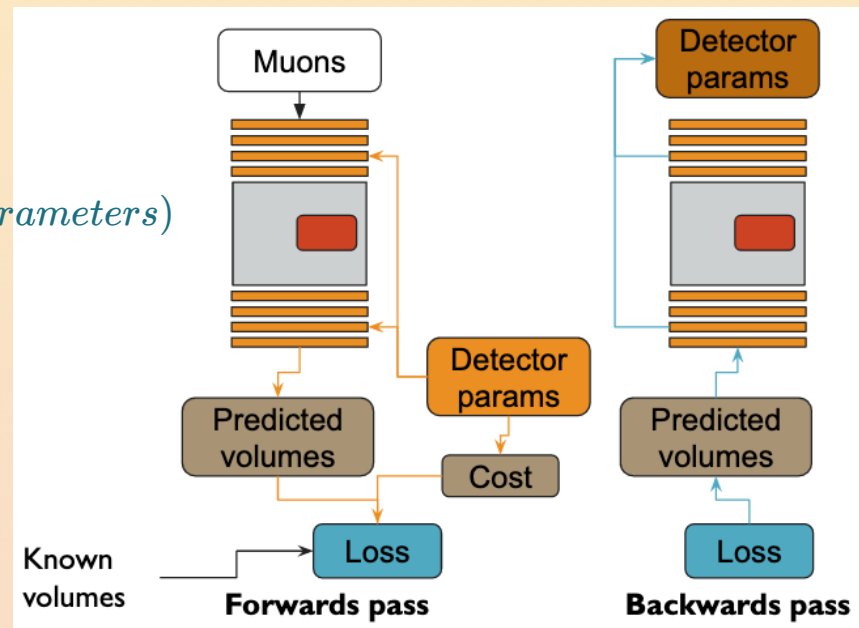
- Can compute $p(\Delta output | \Delta detector\ parameters)$

- Task as loss function

- Including target (e.g. prediction uncertainty), costs, constraints

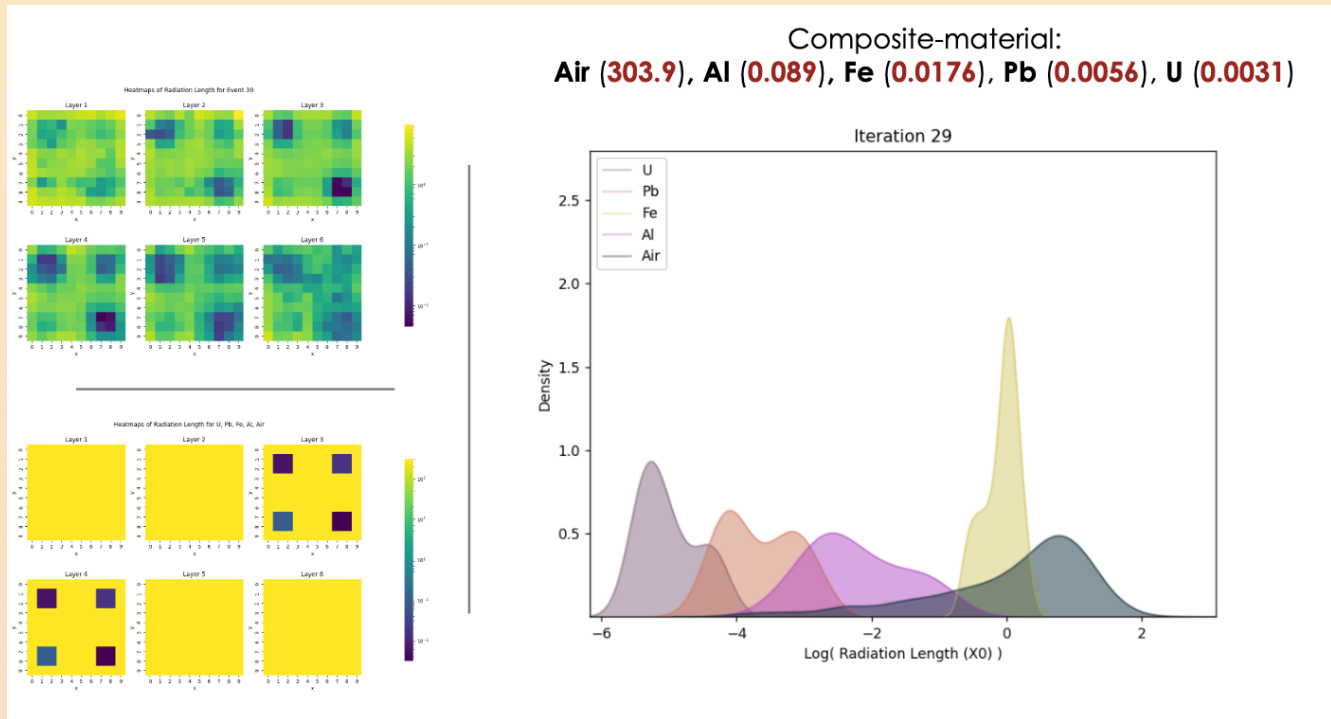
- Backpropagate and optimize as usual

- Gradient descent



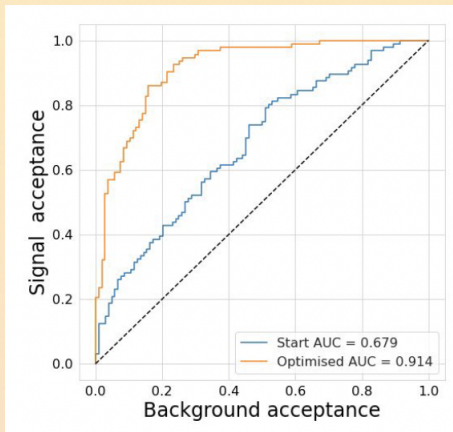
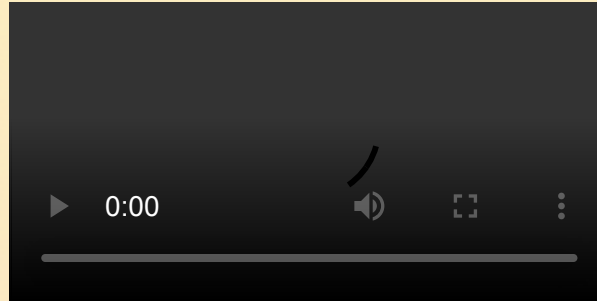
Expectation Maximization

- Iterative algorithm: Mariam Safeldin (AUB, internship UCLouvain/UniOvi)
 - estimate scatter density based on current estimate of the image
 - update estimate of the image based on the estimated scatter density
 - Adrián Fernández Alonso (TFG en informática) will produce a fast C++ implementation

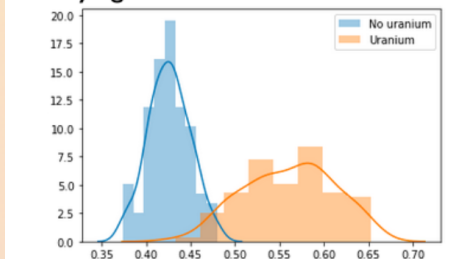


Encouraging results

- Inference via low-dim summary statistic



Dedicated summary-statistic for classifying volumes with uranium blocks



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 Ruiz 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 Aehle**, 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

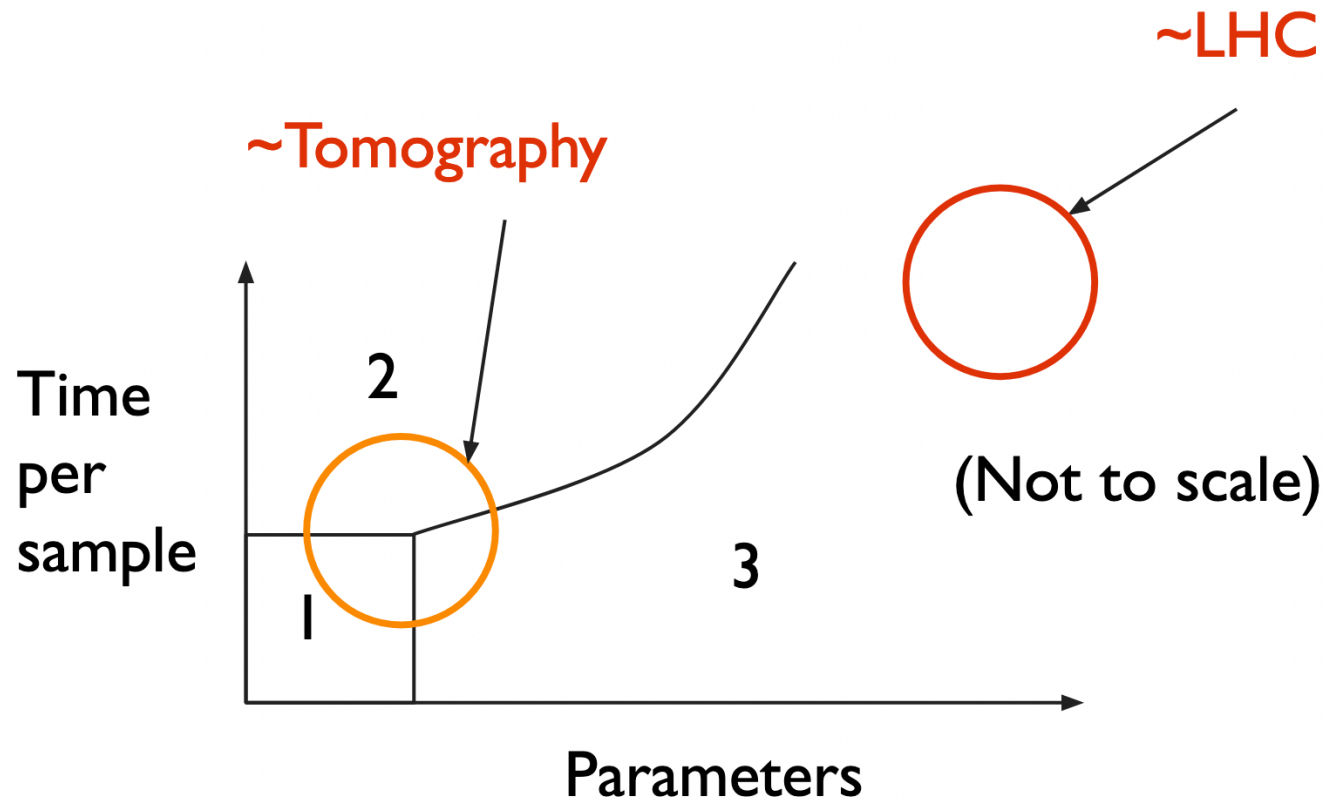
- [First installment](#) in Louvain-la-Neuve (Belgium)
- [Second installment](#) in Kolymbari (Greece)
 - [37 talks, 9 posters](#), one data challenge with prizes, recordings will be online soon
- You are all invited to the [Third MODE Workshop](#), to be held in Princeton (USA)



**Third MODE Workshop on
Differentiable
Programming for
Experiment Design**

**Princeton University
24-26 July, 2023**

We want to optimize the next LHC



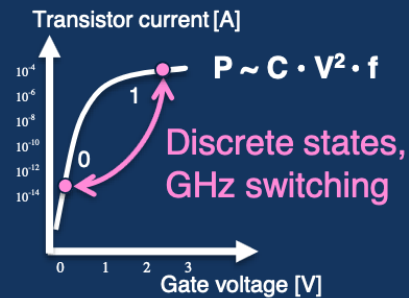
Giles Strong at QCHS 2022

Hybrid Systems

- Ongoing collaboration with Lulea University and INFN for neuromorphic chips
- Kicking off collaboration for quantum computing

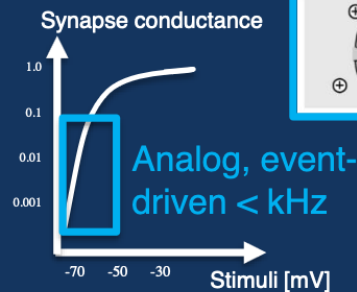
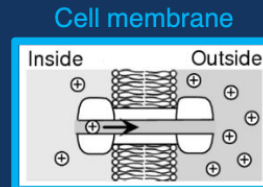
Conventional computers

mimic logical and analytical thinking

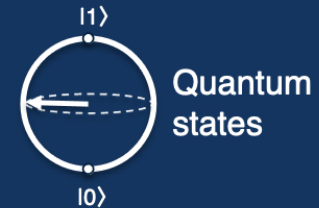


Neuromorphic processors

mimic the senses, learning and perception

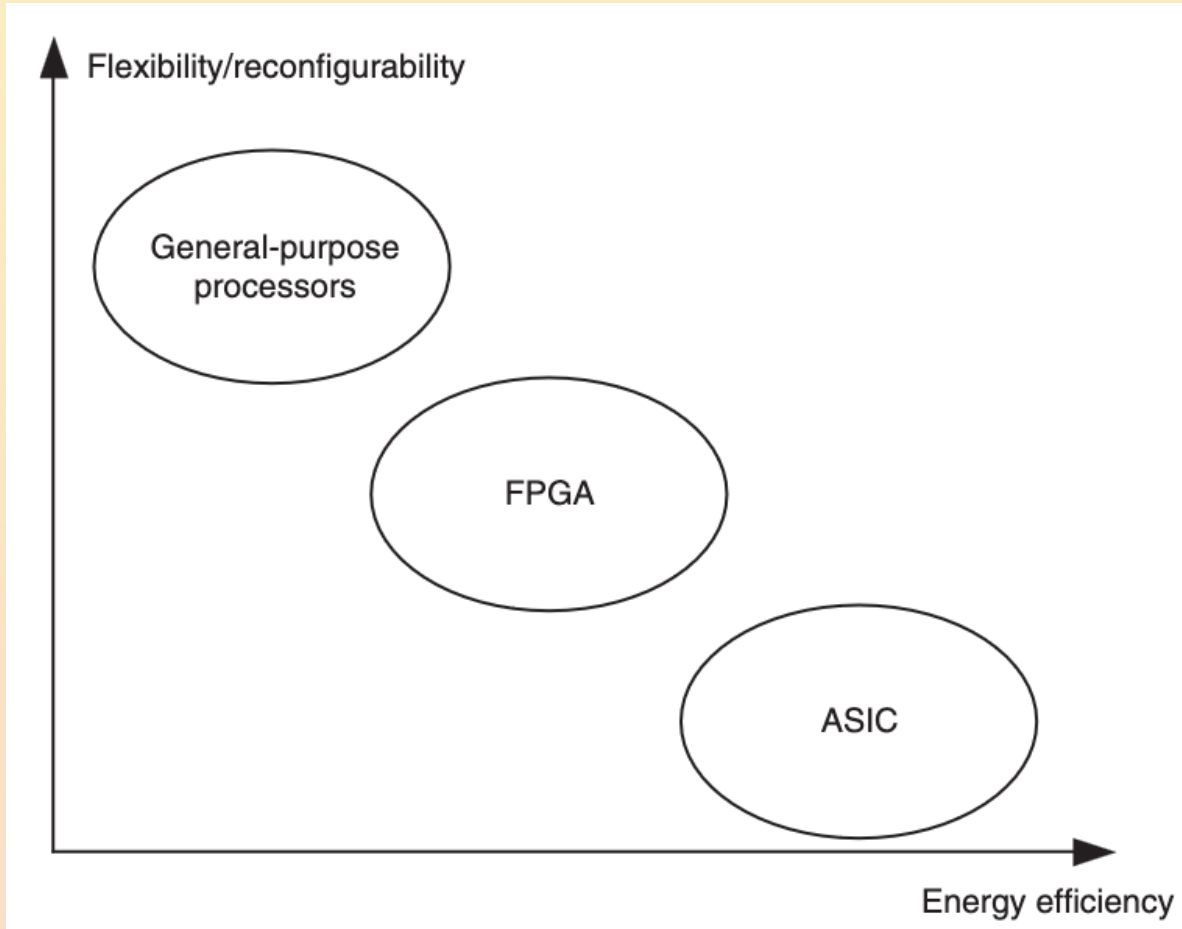


Quantum processors
use quantum superpositions for probabilistic inference



Technology readiness?

Neural Networks in hardware



Lebesgue sampling

- Efficient, threshold-based encoding
 - GPU (RTX3090): 40 GW simulation on discrete states
 - Human brain equivalent: 20 W on dynamic states



Simplify 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

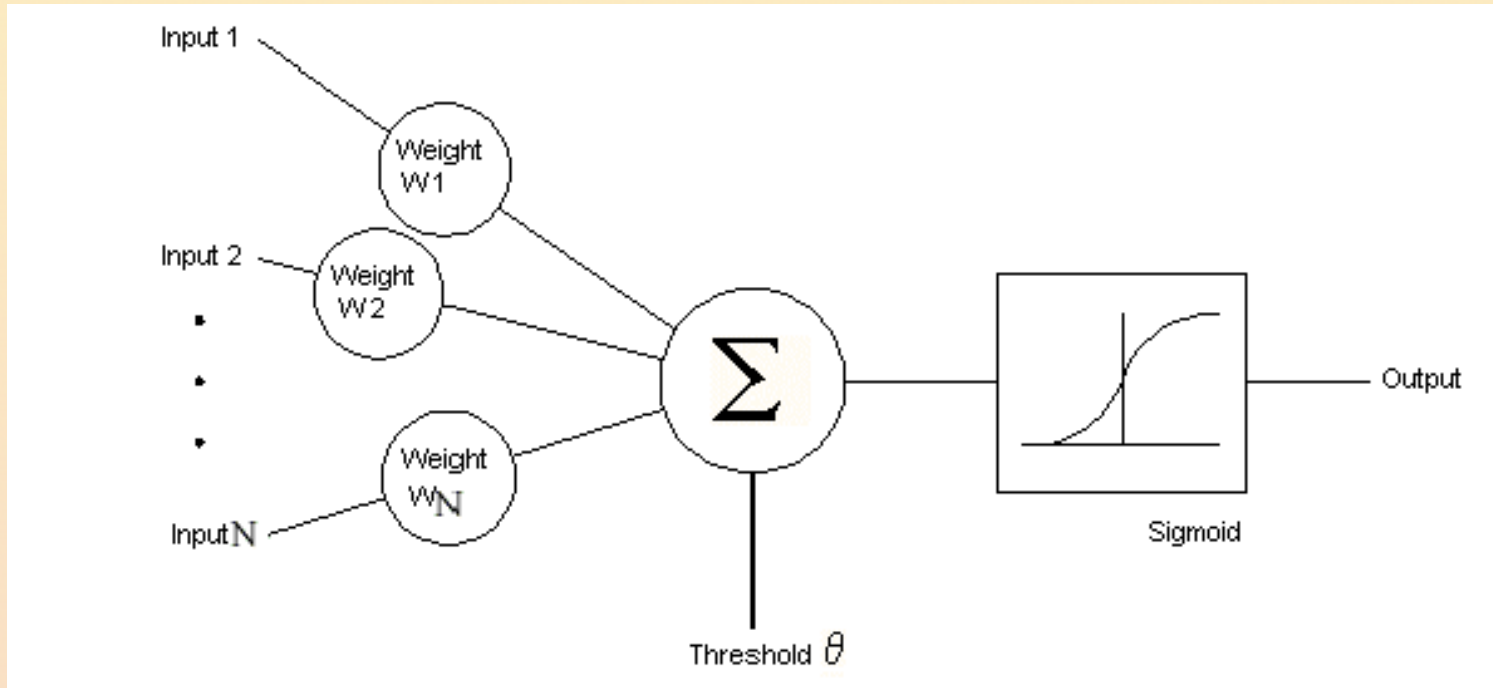
A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY*

- 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.

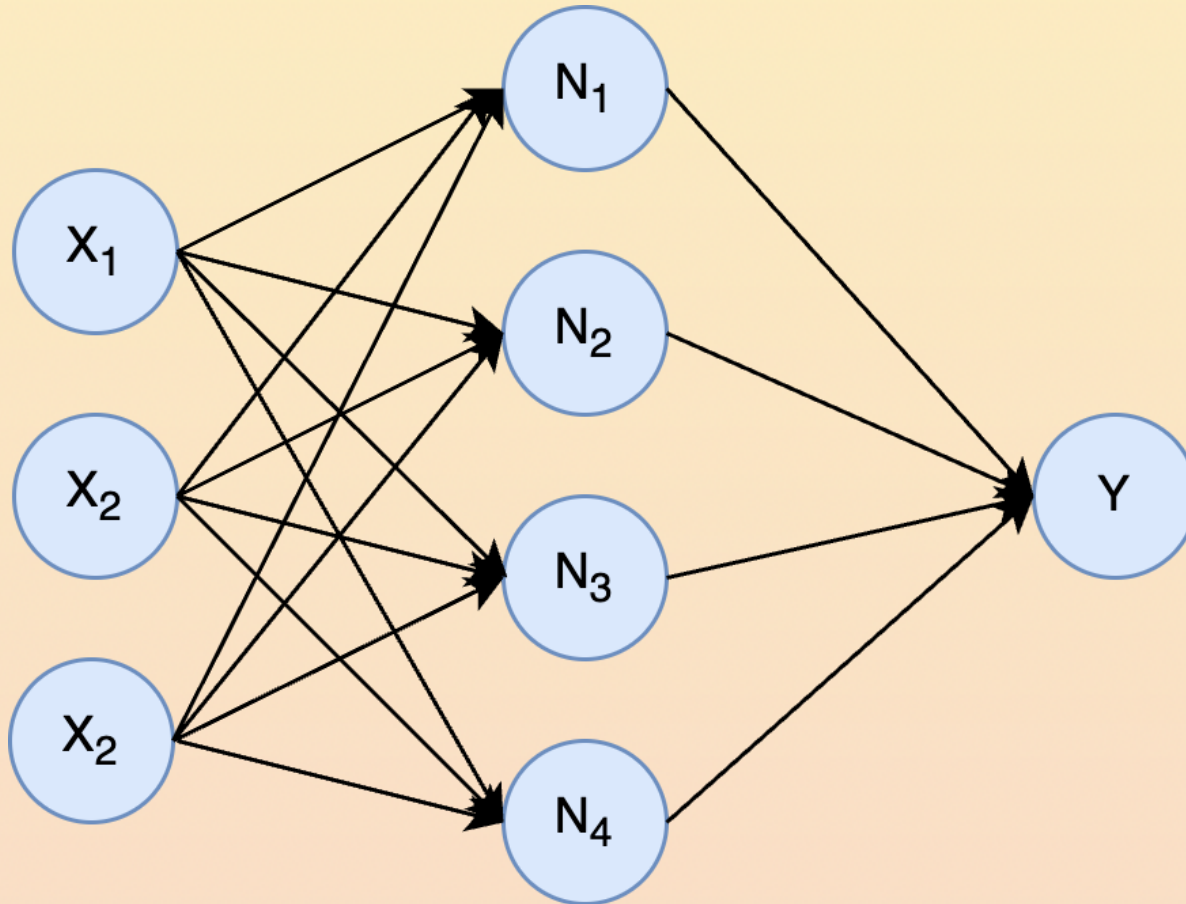
Perceptron

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



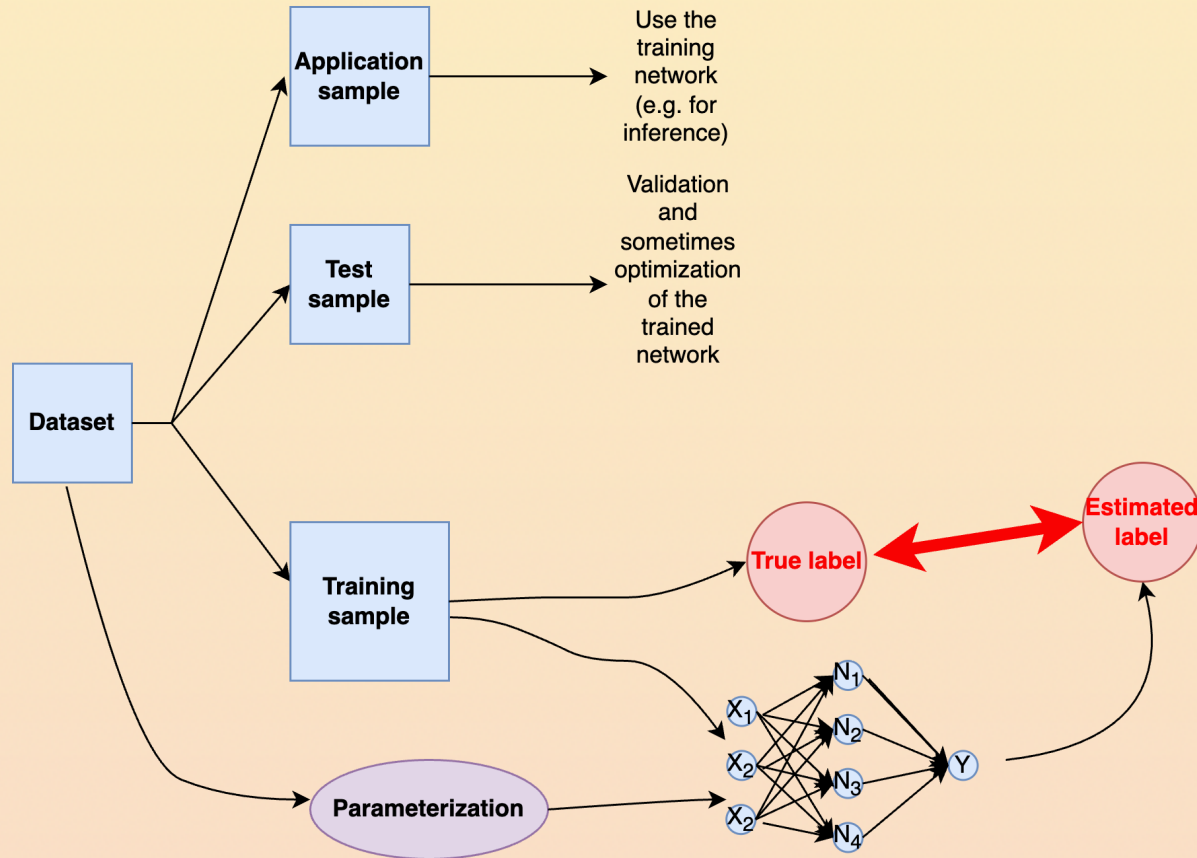
Artificial Neural Network

- Each weight is a free parameter to be determined during training

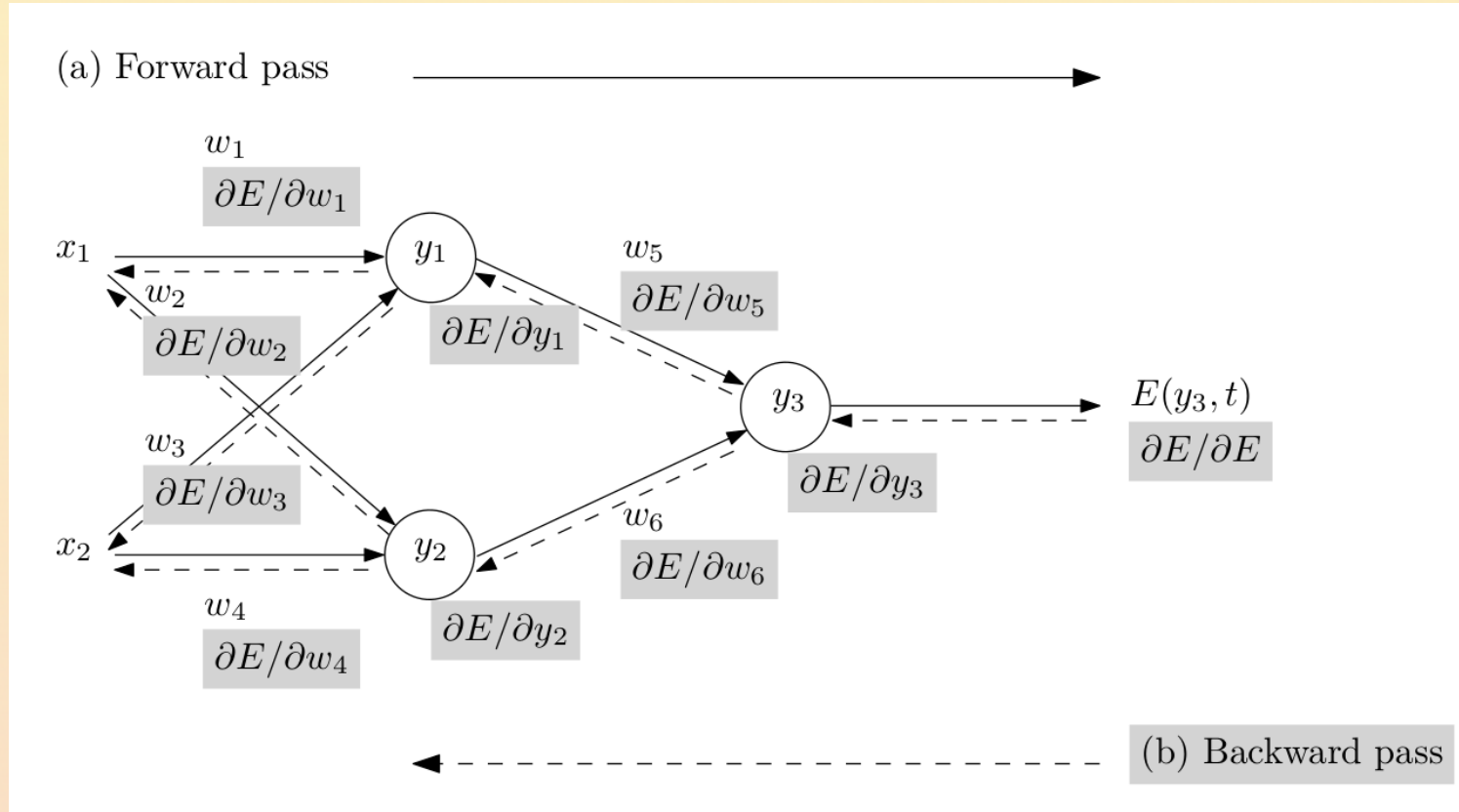


Training a Neural Network

- Each weight is a free parameter to be determined during training



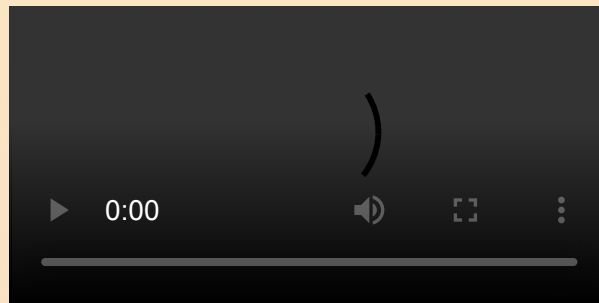
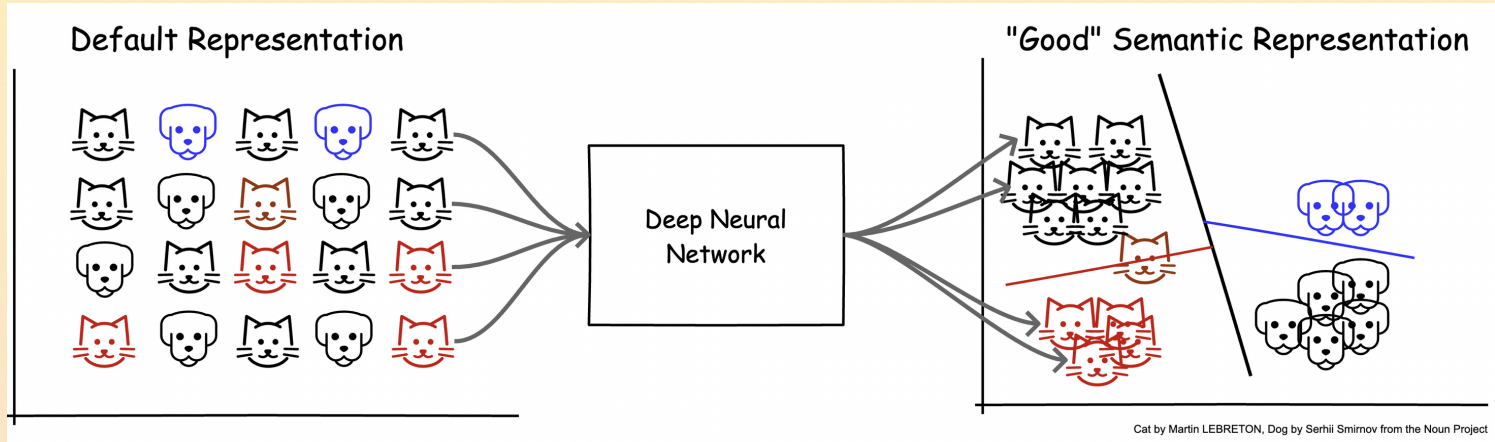
Backpropagation



Mapping Improves Understanding

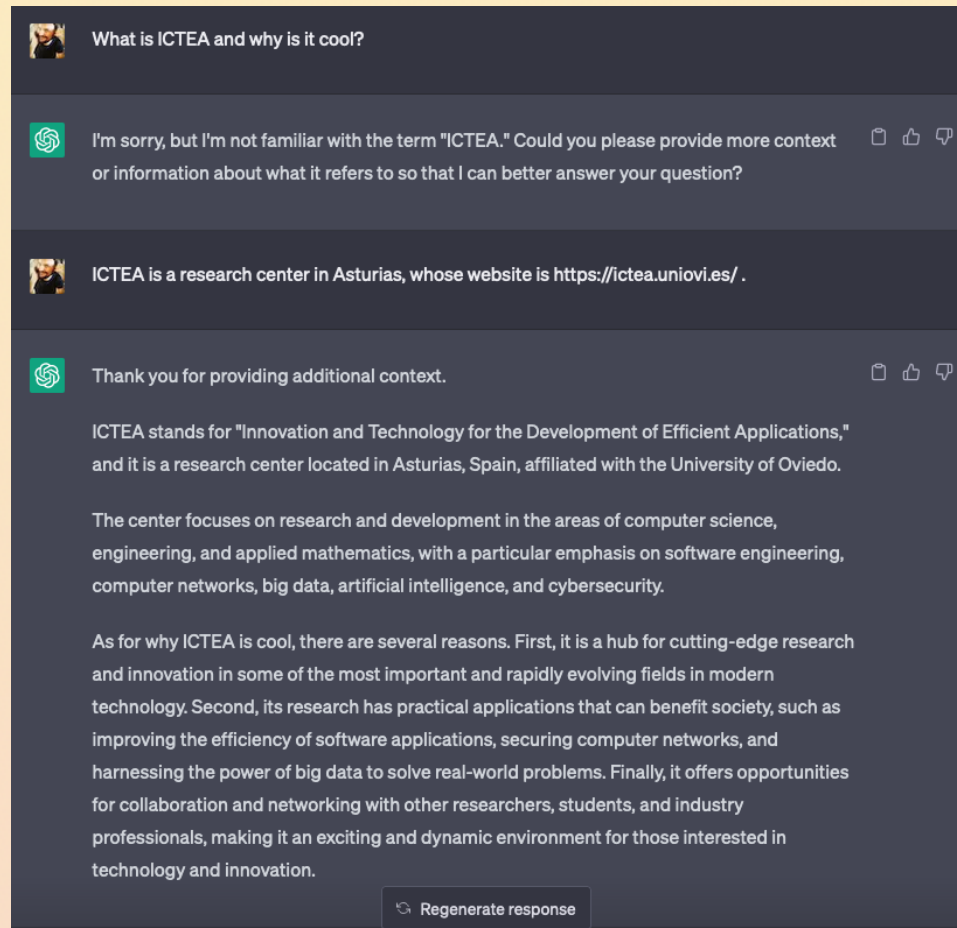


Representations Make Tasks Easier



But not too easy

- Join "Análisis de Datos en Física Moderna" (optativa de cuarto)!
- Will study a two-words Generative Pretrained Transformer!!!



The screenshot shows a chat interface with a dark background. The user asks, "What is ICTEA and why is it cool?". The AI responds with an apology for not knowing the term and asks for more context. The user provides the website <https://ictea.uniovi.es/>. The AI then provides a detailed explanation of ICTEA, including its full name, location, research focus, and why it is considered cool.

What is ICTEA and why is it cool?

I'm sorry, but I'm not familiar with the term "ICTEA." Could you please provide more context or information about what it refers to so that I can better answer your question?

ICTEA is a research center in Asturias, whose website is <https://ictea.uniovi.es/>.

Thank you for providing additional context.

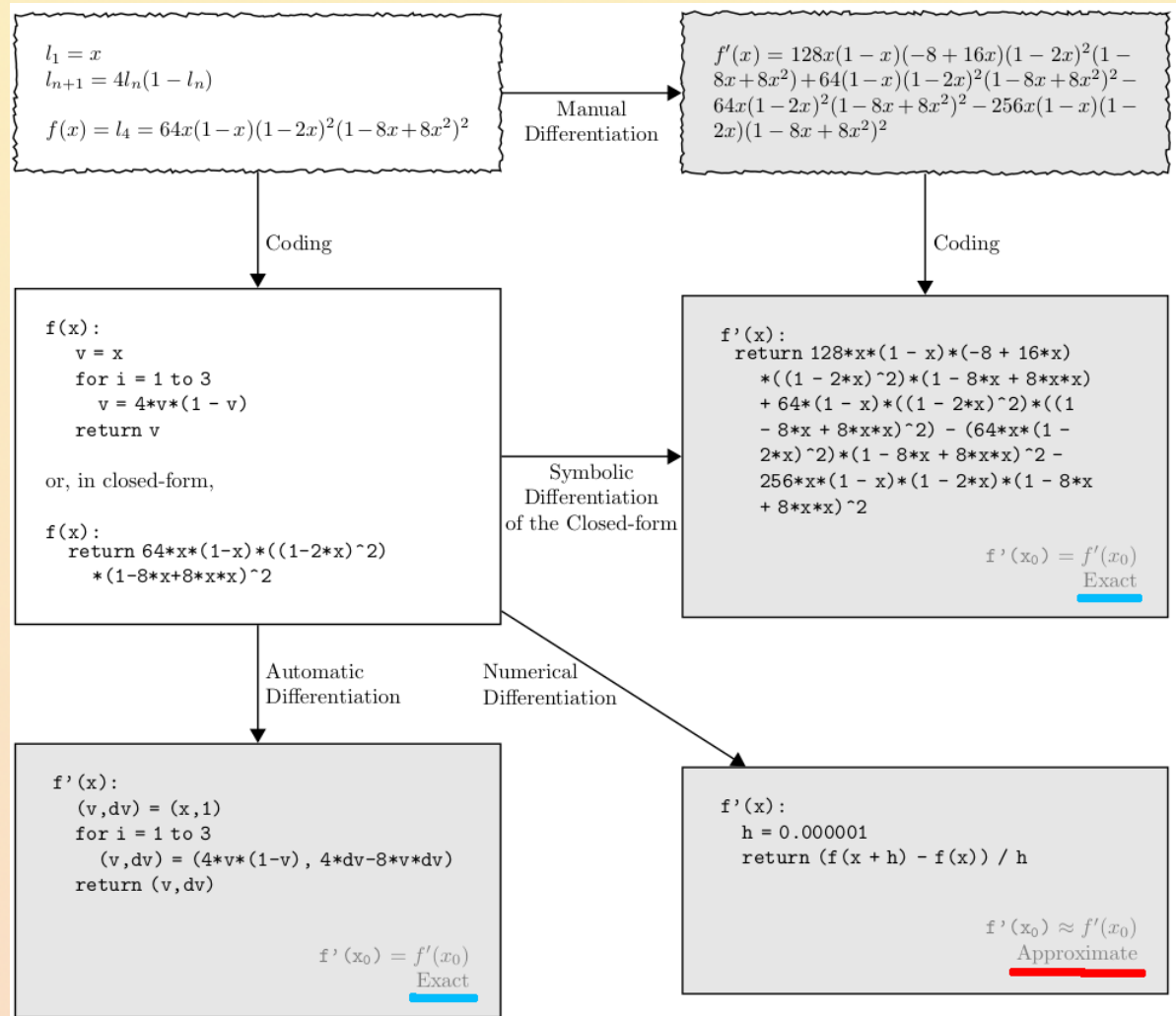
ICTEA stands for "Innovation and Technology for the Development of Efficient Applications," and it is a research center located in Asturias, Spain, affiliated with the University of Oviedo.

The center focuses on research and development in the areas of computer science, engineering, and applied mathematics, with a particular emphasis on software engineering, computer networks, big data, artificial intelligence, and cybersecurity.

As for why ICTEA is cool, there are several reasons. First, it is a hub for cutting-edge research and innovation in some of the most important and rapidly evolving fields in modern technology. Second, its research has practical applications that can benefit society, such as improving the efficiency of software applications, securing computer networks, and harnessing the power of big data to solve real-world problems. Finally, it offers opportunities for collaboration and networking with other researchers, students, and industry professionals, making it an exciting and dynamic environment for those interested in technology and innovation.

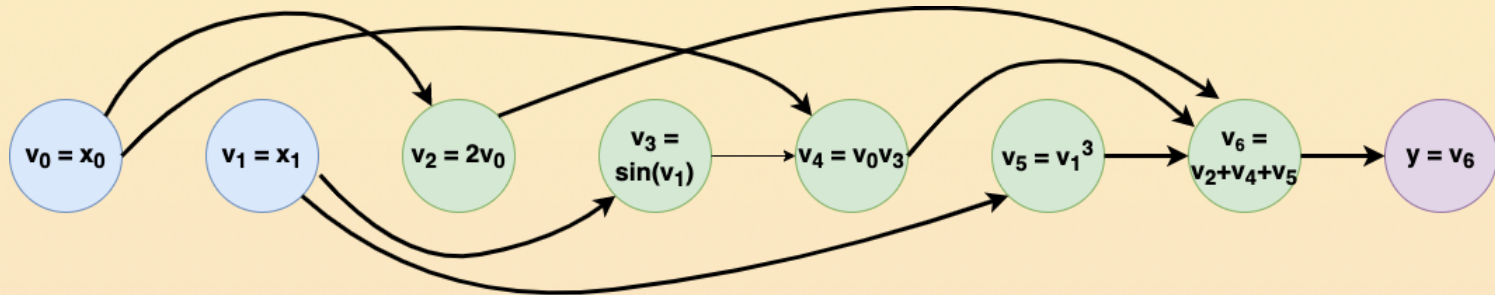
Regenerate response

Derive



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 $\nabla f(\mathbf{x})(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n})$

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

The Forward Mode of AD

Primal: independent to dependent

Tangent (derivatives): independent to dependent

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

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

The Reverse Mode of AD

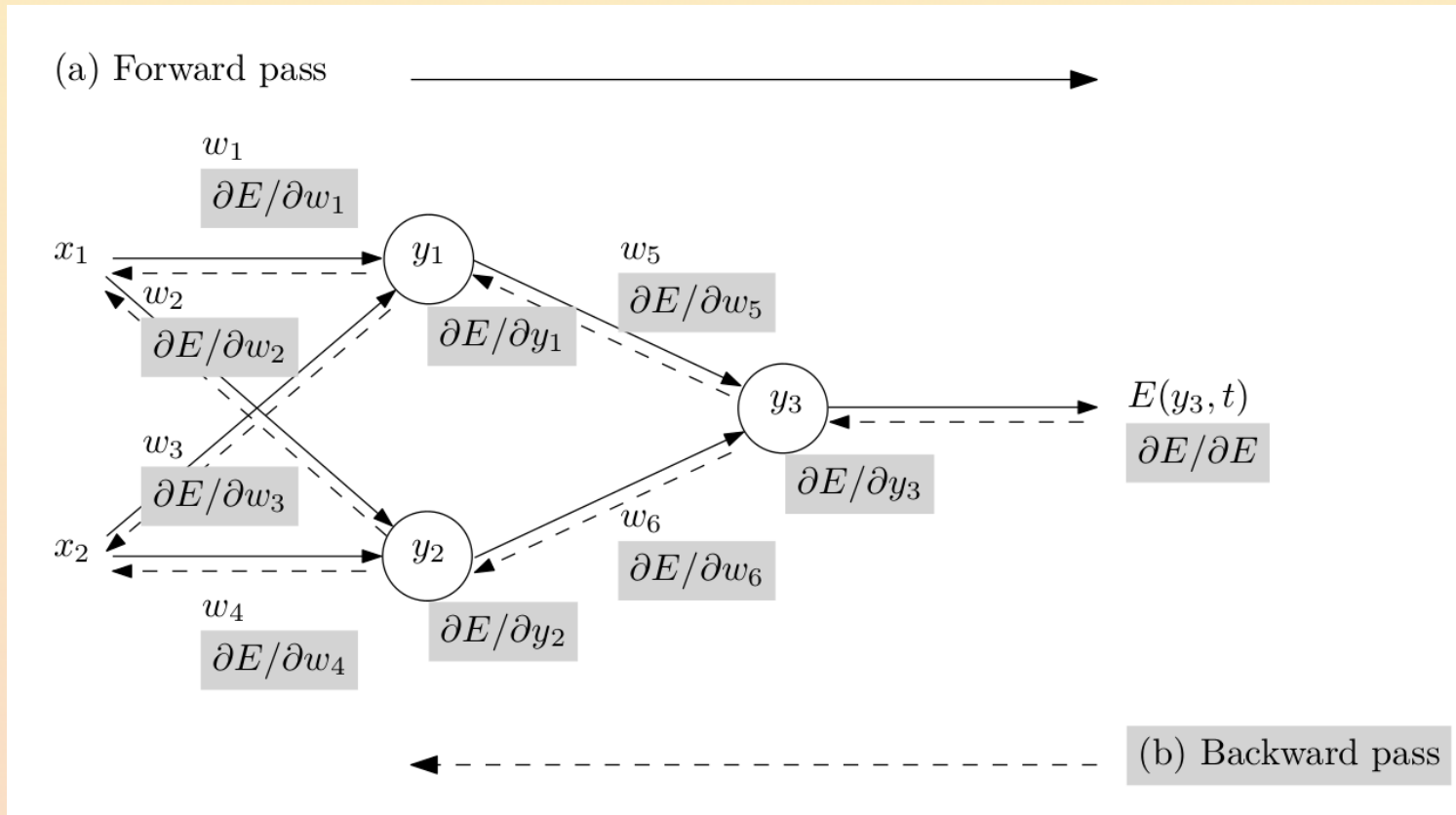
Primal: independent to dependent

Adjoint (derivatives): dependent to independent

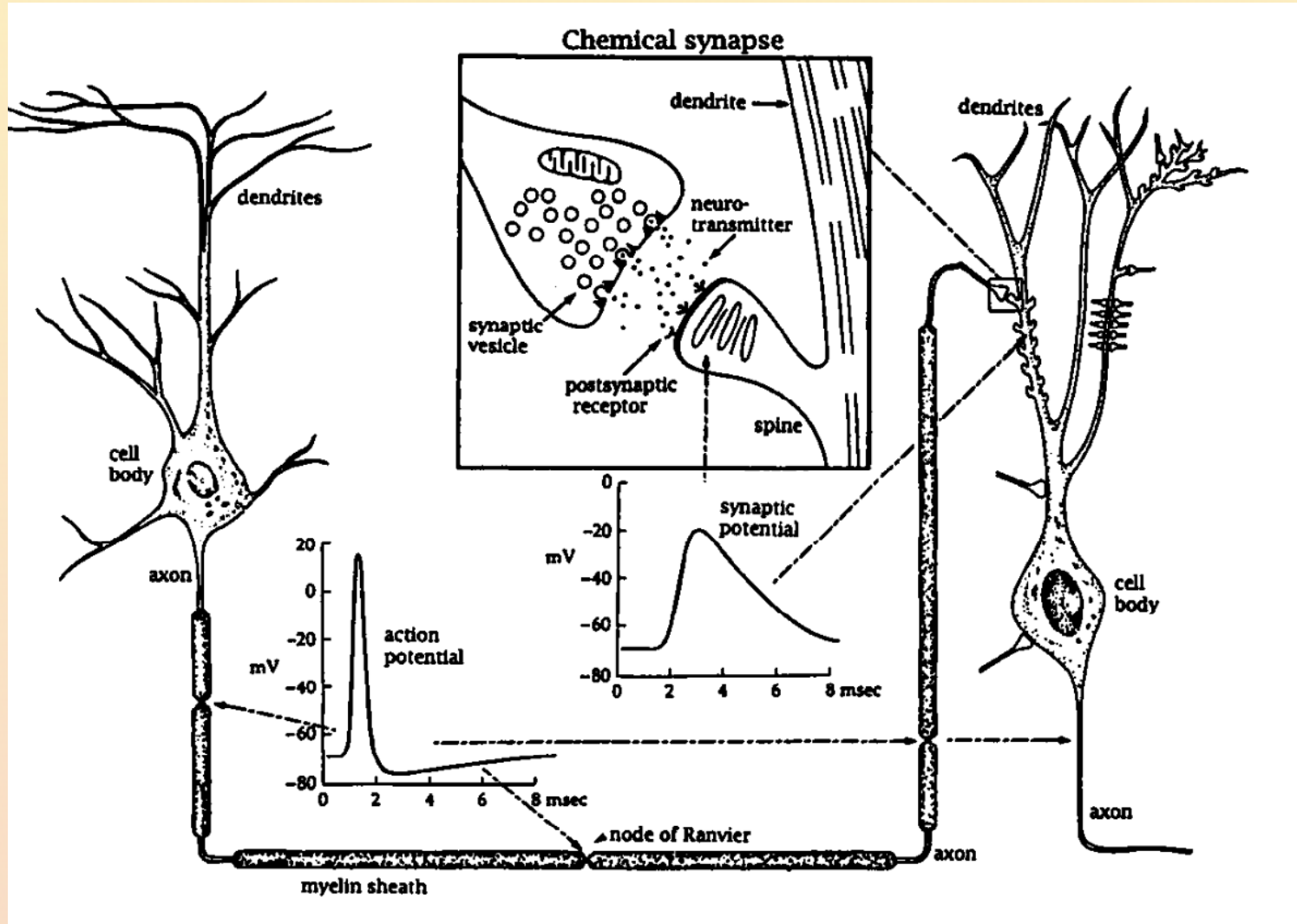
$$y(\mathbf{x}) = 2x_0 + x_0 \sin(x_1) + x_1^3$$

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

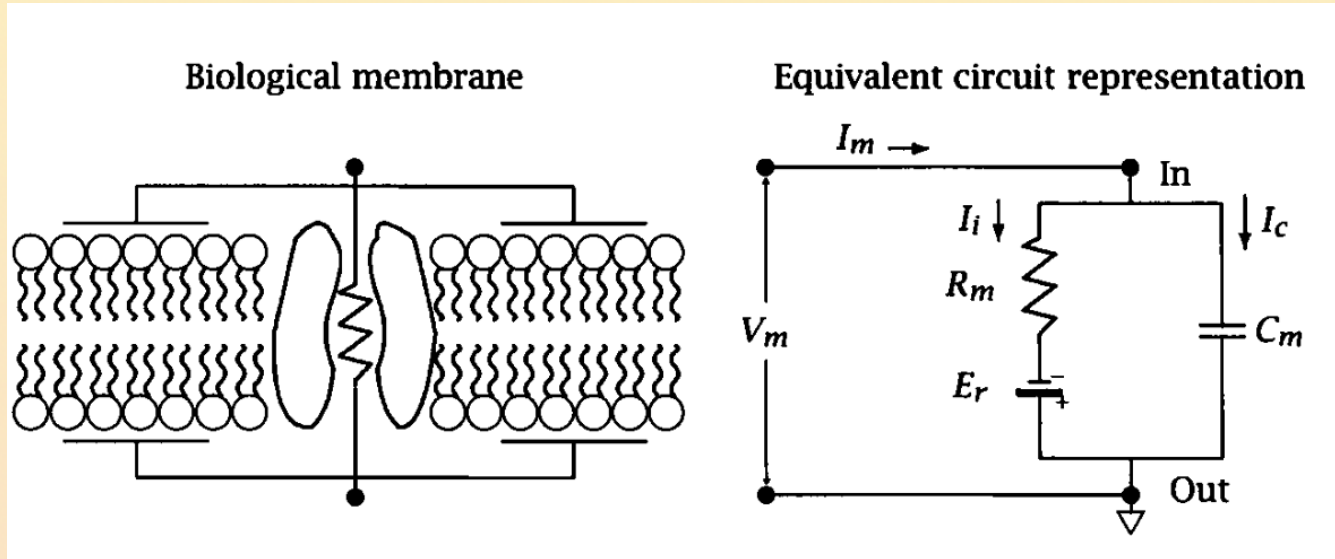
Reverse AD Generalizes Backpropagation



Biological Neurons

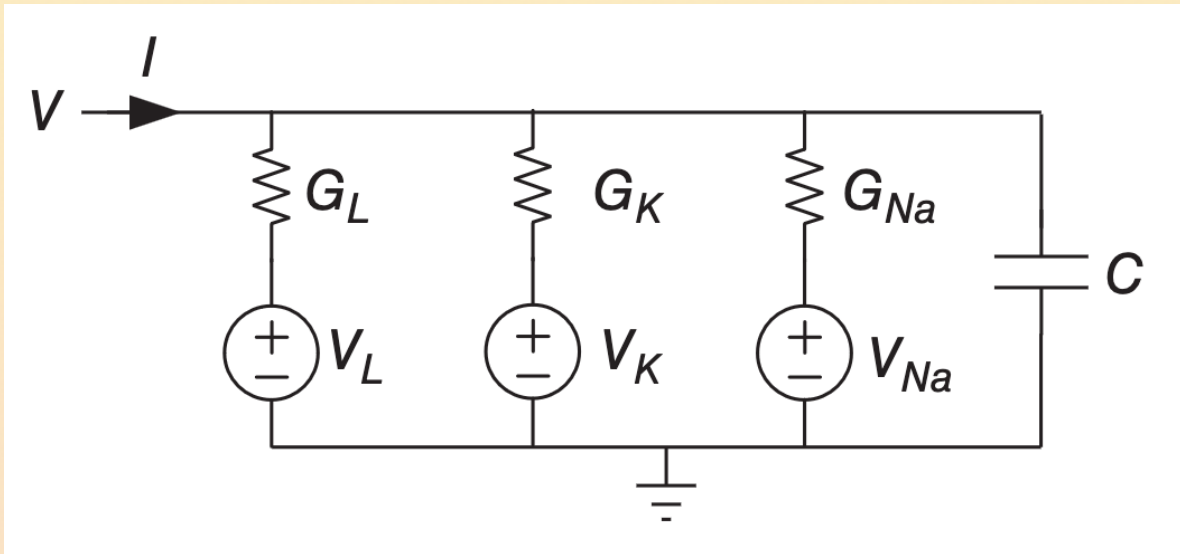


Biological membranes are circuits



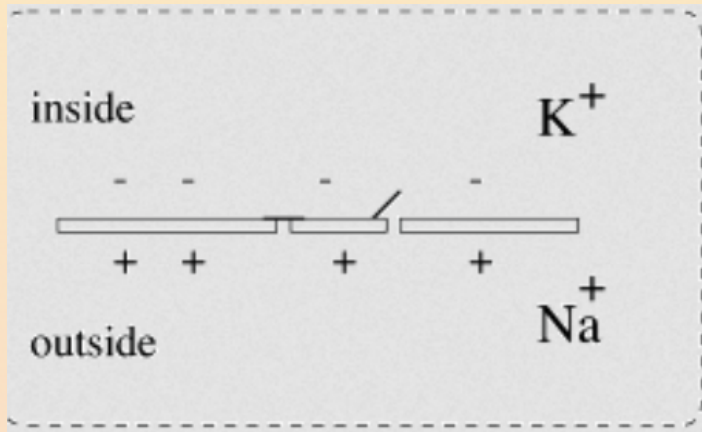
The Hodgkin-Huxley model

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



The Hodgkin-Huxley model

- External potentials change Na channel conductance, membrane potential increases
 - Spike!
- Reversal potential determines when this process stops
 - K channel acts in the opposite way!



x	E_x [mV]	g_x [mS / cm^2]
Na	55	40
K	-77	35
L	-65	0.3

Study living systems

- Nervous systems are complex structures with **different organization levels**
- Formal computational models **bridge the gap** between levels
- Predicting observed properties provides insights to understand neural systems
- **Emergent properties** may stem from circuital behaviour
- **Differentiable programming** enables fast and accurate solution of circuit equations

Gymnotus Omarorum

- Lives in ponds, mostly nocturnal
 - Often muddy waters, lots of vegetation
- Generates weak electric field and detect its deformations via electrosensory lobe
 - Relatively simple neurons, ideals for study

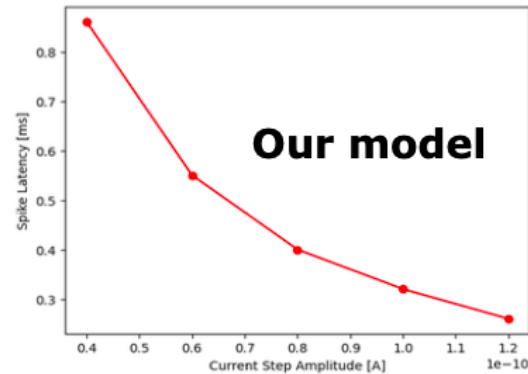
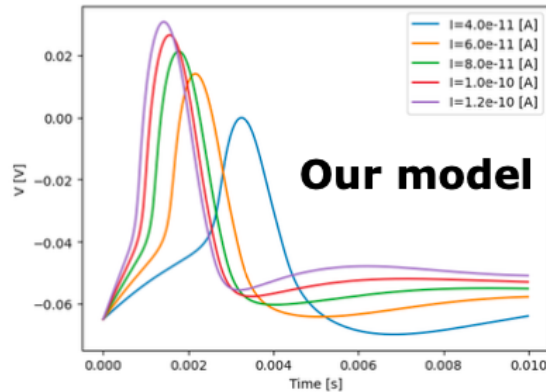


Introducing new currents

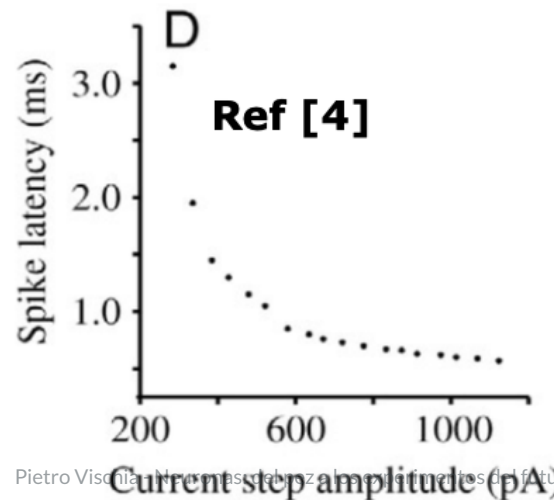
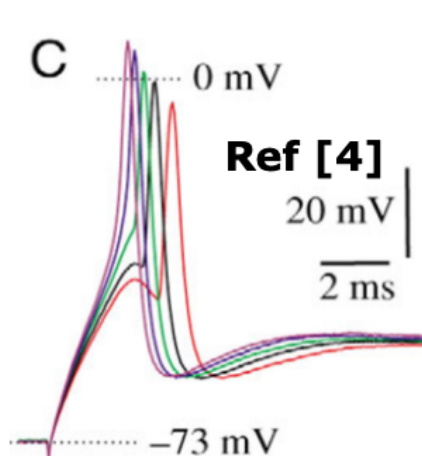
- Computational model of a spherical neuron
 - Na^+ channel
 - K^+ channel
 - h: mixed-cation hyperpolarizing activated channel
 - Leakage current
 - **New** Low-threshold non inactivating K^+ channel
 - Rectification effect on response
 - Long refractory period
- Compare with data from "[4]" ([J Exp Biol \(2006\) 209 \(6\): 1122–1134.](#))
- Work presented at Neural Coding 2023 (paper in preparation)

Stimulus amplitude and latency

- The amplitude of the stimulus step drives the spike latency

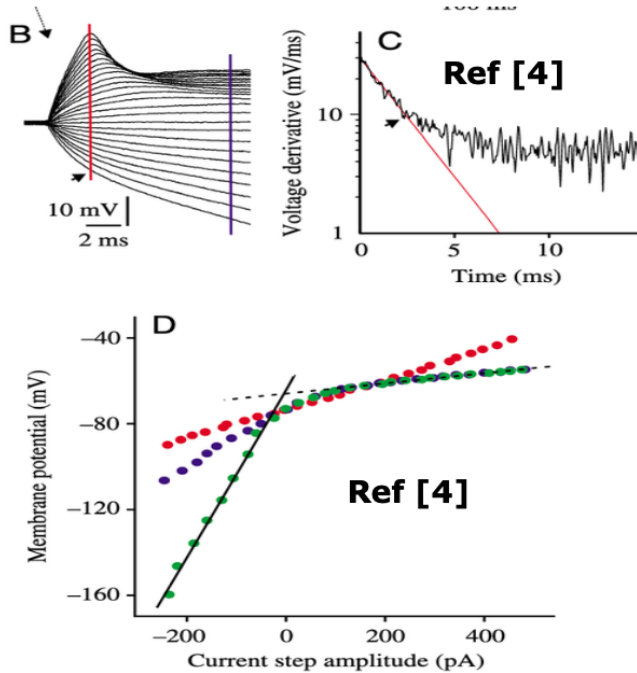


- Observations exhibit the same behaviour
- Further turning needed for the spike shape

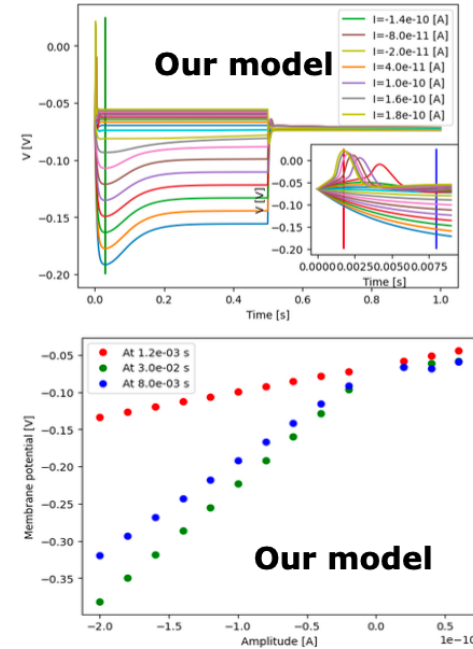


Early responses to stimuli

- Early subthreshold responses

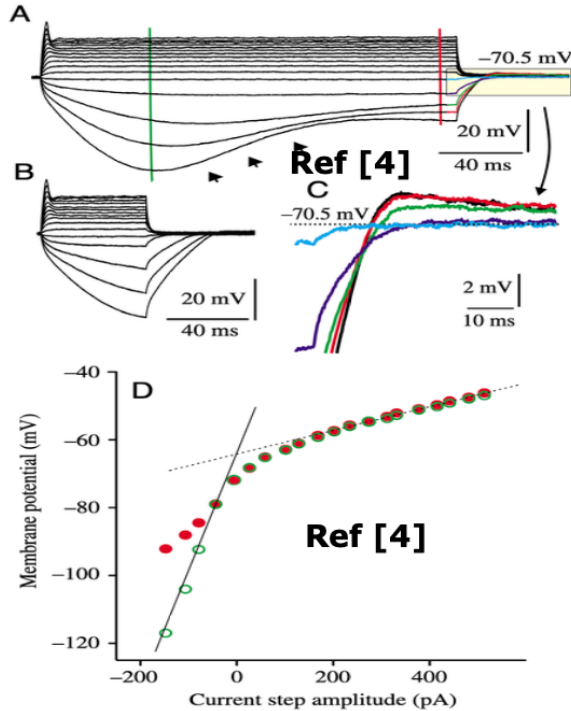


- Before hump (red), linear V-I relation
 - After hump, for depolarizing steps V-I relation is nonlinear
- The activated conductance does not inactivate at later times

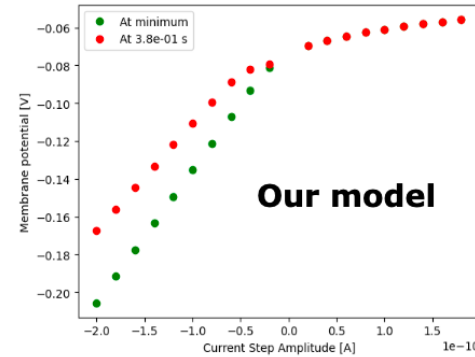
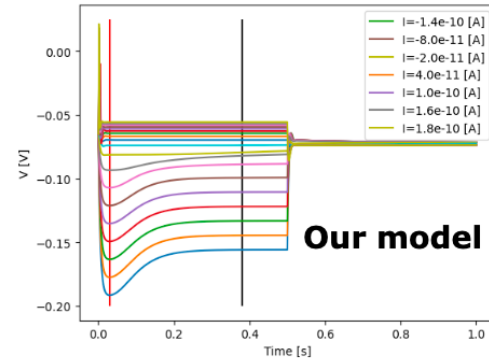


Late responses to stimuli

- Late subthreshold responses

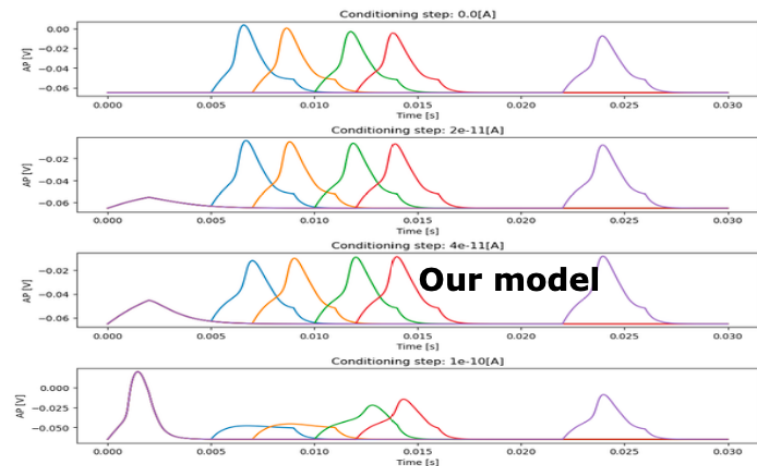
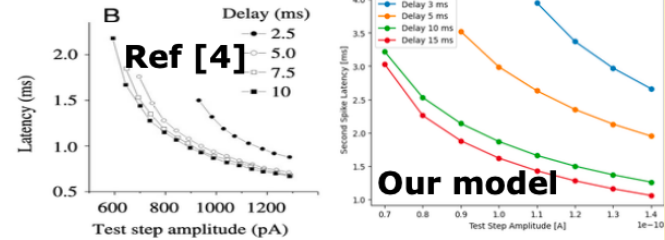
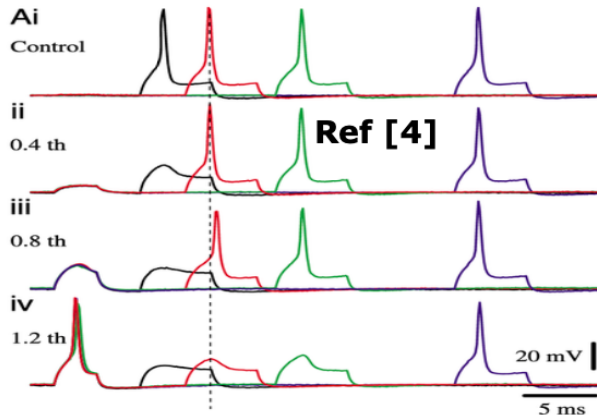


- At peak hyperpolarization, limiting slope is maximal
- At end of the step, depolarization curves decay much faster



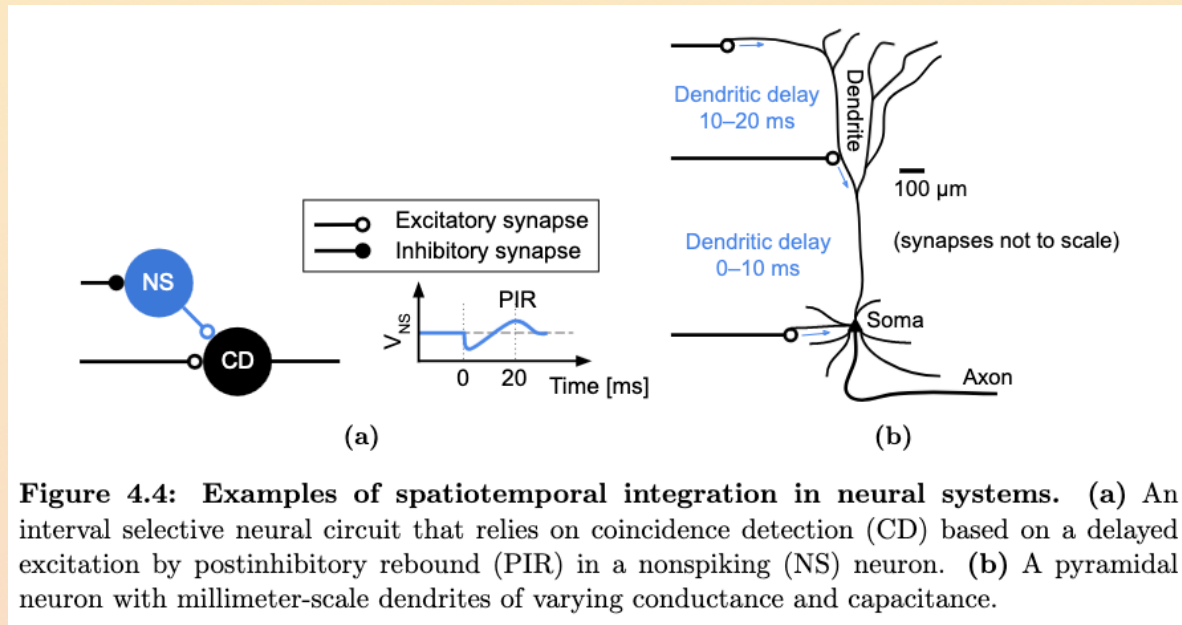
Refractory period

- Conditioning step induces a refractory period
- Behaviour of the refractory period matches observations
 - Amplitude of conditioning step
 - Amplitude of test step
 - Delay of test step



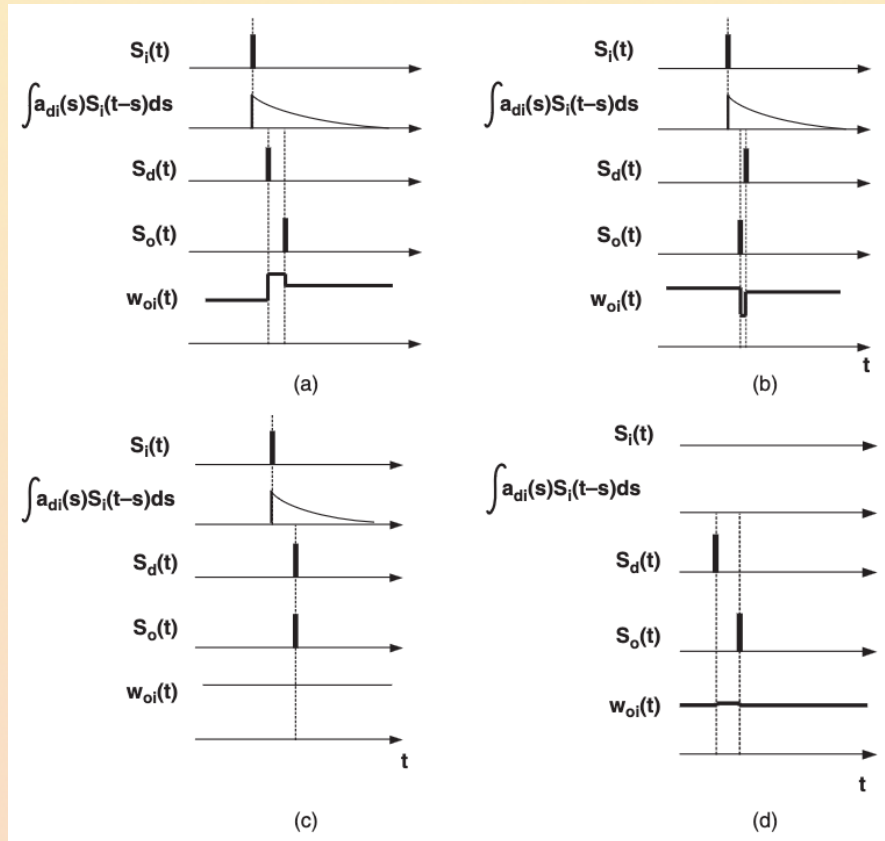
Satisfactory results

- Simple models qualitatively reproduce observations (gradient-based fit ongoing)
 - Low-threshold K current crucial to reproduce rectification and refractory period
- Next: [study of emergent properties](#) from multisynapse inputs!



Learn with biological neurons

- I hope to encode an LHC-experiment-scale optimization using spiking networks in neuromorphic hardware!



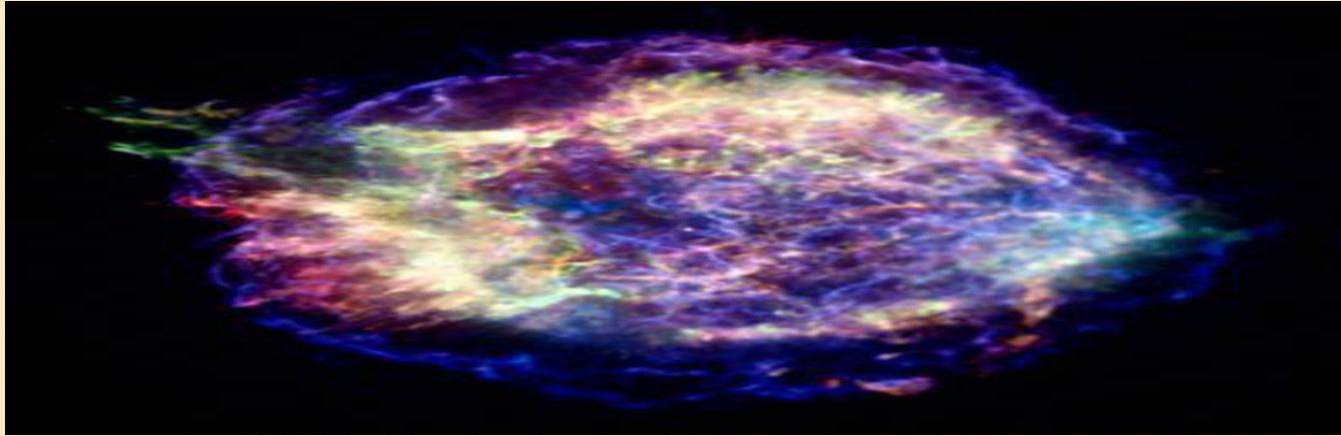
Let's talk and collaborate!

- NeuroModelling (my main project): I am currently building my own small group dedicated to these topics
 - PhD theses
 - TFGs, TFMs
 - Smaller projects
 - Hopefully some funding later in the year :D
- Join "Análisis de Datos en Física Moderna" (optativa de cuarto) to learn more!

Backup

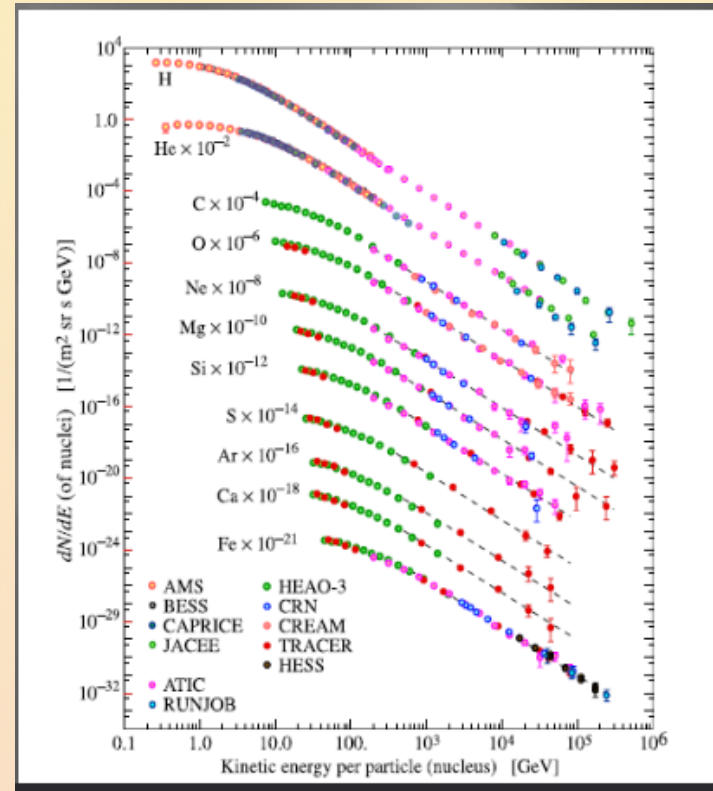
Cosmic rays from supernovae

- 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

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

- Improved formulas account for Earth curvature and other effects
 - [Guan et al. \(2015\)](#)
 - [Shukla and Sanskrith \(2018\)](#)

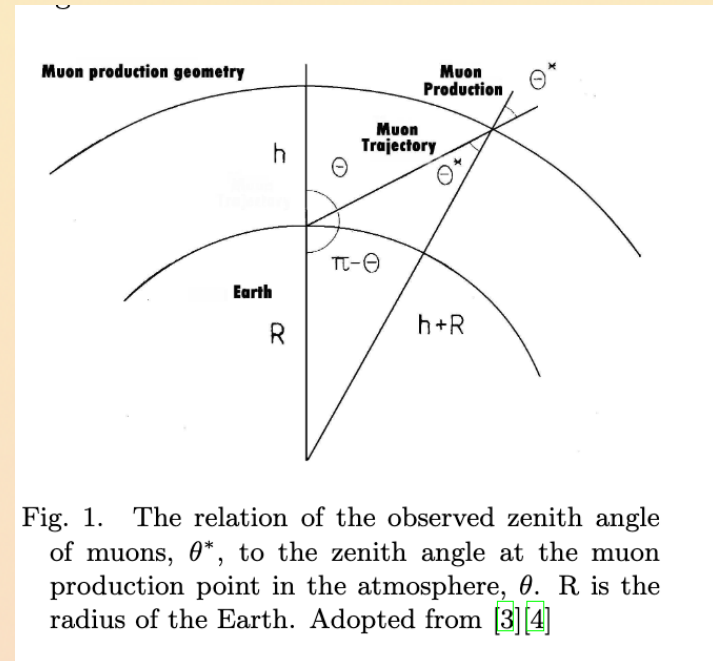
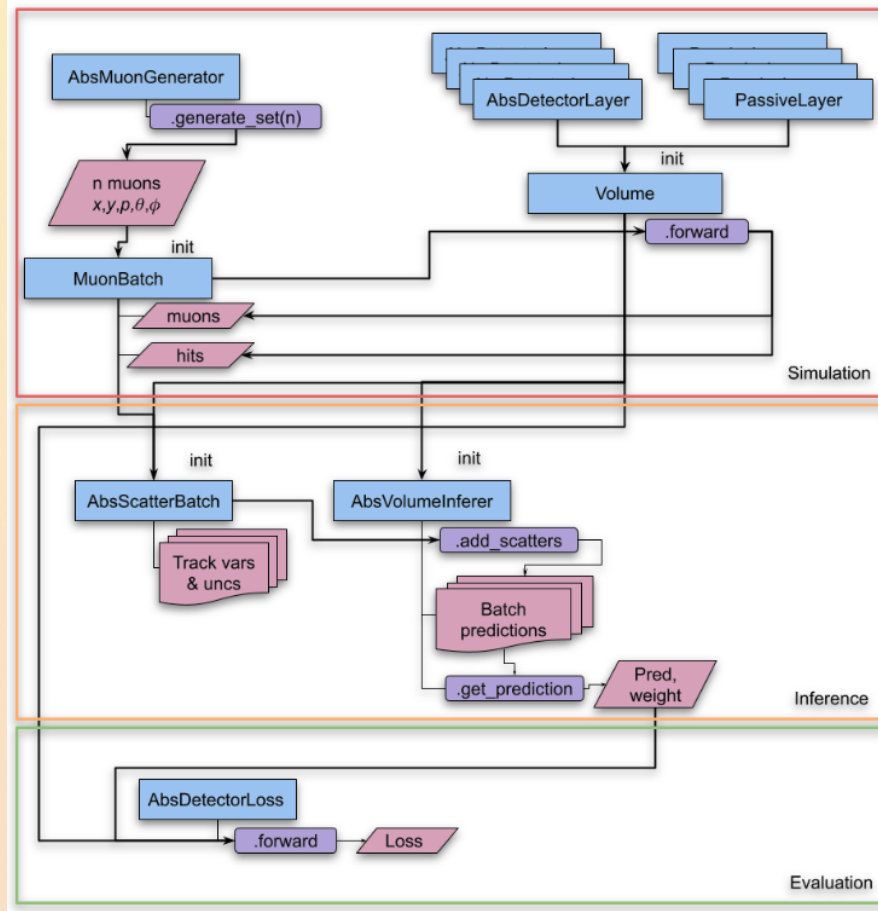


Fig. 1. The relation of the observed zenith angle of muons, θ^* , to the zenith angle at the muon production point in the atmosphere, θ . R is the radius of the Earth. Adopted from [3][4]

The TomOpt optimization cycle

- Modular design in python, automatic differentiation via PyTorch



Supervised Learning

$$L(\mathbf{w}) = \frac{1}{2}(\mathbf{x}_{\text{out}} - \mathbf{t})^T(\mathbf{x}_{\text{out}} - \mathbf{t}) + \lambda R(\mathbf{w})$$

- \mathbf{t} supervisory signal (labelled data set) explicitly provided in the training phase
- $\mathbf{x}_{\text{out}} = \mathbf{x}_{\text{out}}(\mathbf{w}, \mathbf{x}_{\text{in}})$

Generalization

- Learn a map to generalize

(overfitting/underfitting)

Reinforcement learning

Unsupervised learning

$$L(\mathbf{w}) = \frac{1}{2}(\mathbf{x}_{\text{out}} - \mathbf{x}_{\text{in}})^T(\mathbf{x}_{\text{out}} - \mathbf{x}_{\text{in}})$$

- Regularization handled by denoising or sparseness

image autoencoder, compress

Overparameterization landscape