

# Differentiable Programming in HEP

ACAT 2024

Lukas Heinrich, TUM

Technische  
Universität  
München





# Big Picture Questions

Development of AI methods  
for Particle Physics

Does this mean we should put  
Physics Knowledge into AI ?

SOFTWARE ENGINEERING, ARTIFICIAL INTELLIGENCE AND  
EXPERT SYSTEMS IN HIGH ENERGY AND NUCLEAR PHYSICS

AINHEP 1990

NEW  
COMPUTING TECHNIQUES  
IN PHYSICS RESEARCH



Edited by  
D. PERRET-GALLIX, W. WOJCIK

ARTIFICIAL INTELLIGENCE AND  
ENERGY AND NUCLEAR PHYSICS



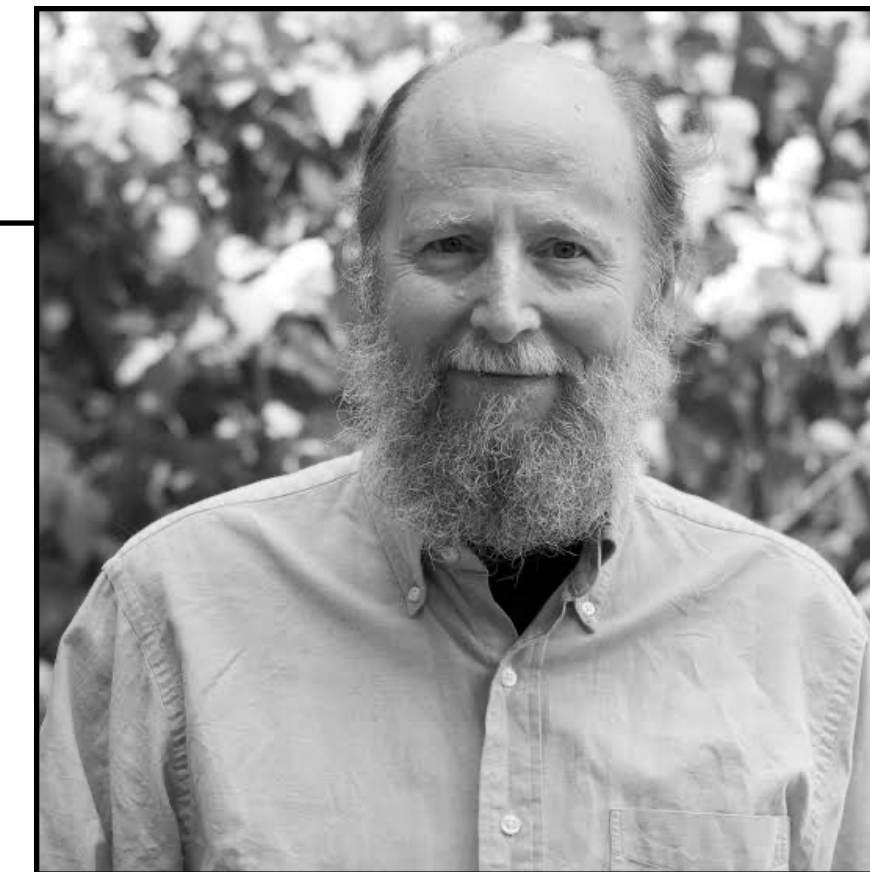
EDITIONS du CNRS



# The Bitter Lesson

Rich Sutton

March 13, 2019

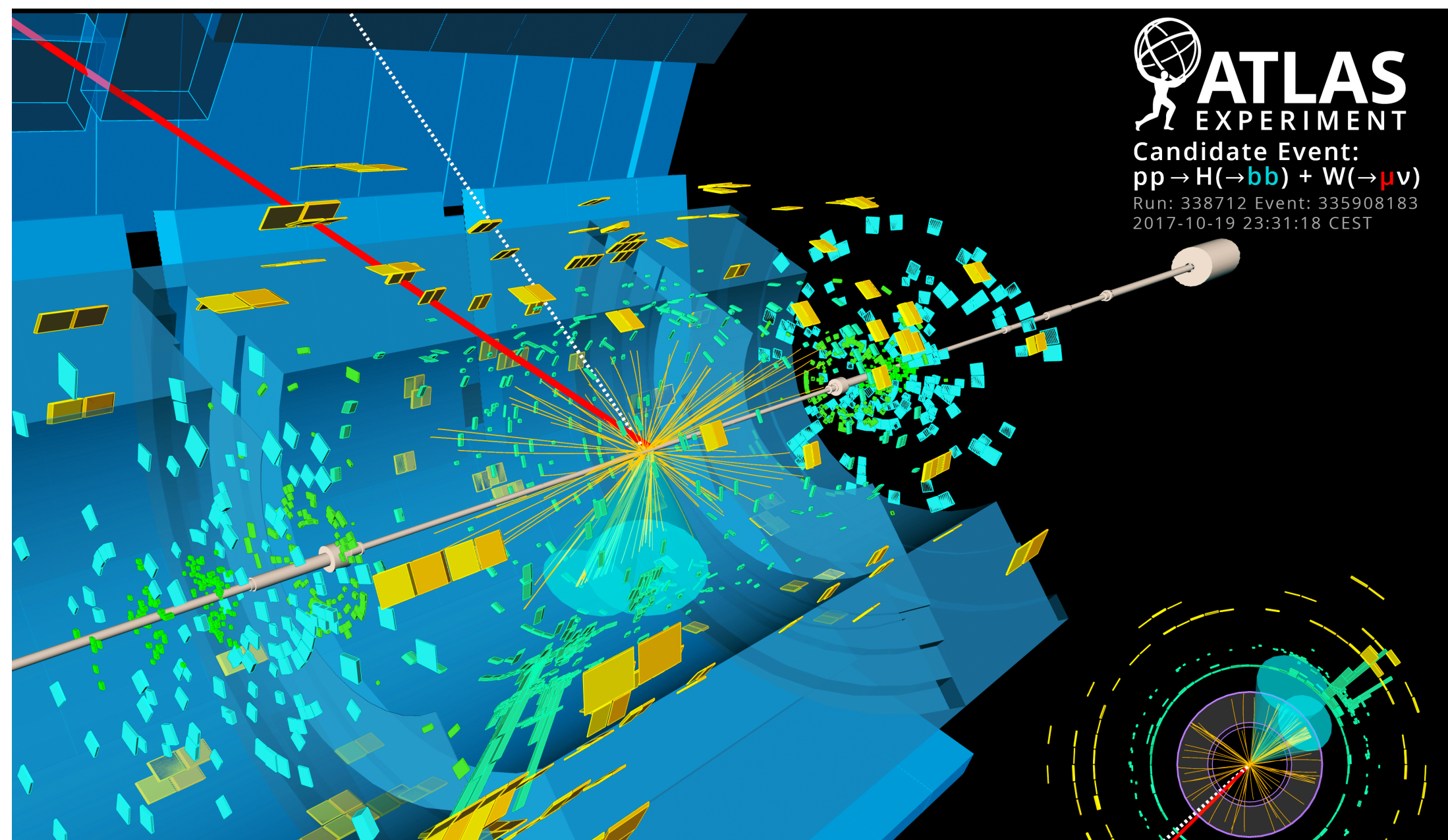


The biggest lesson is that researchers seek to leverage their human knowledge [...], but the only thing that matters in the long run is the leveraging of computation. Methods that leverage computation are ultimately the most effective, or rather its generalization of continued exponentially, conducted as if the computation available to the agent were constant (and the only ways to improve performance) but, over a slightly longer time than a typical research project, massively more computation inevitably becomes available. Seeking an improvement that makes a difference in the shorter term, researchers seek to leverage their human knowledge of the domain, but the only thing that matters in the long run is the leveraging of computation. These two need not run counter to each other, but in practice they tend to. Time spent on one is time not spent on the other. There are psychological commitments to investment in one approach or the other. And the human-knowledge approach tends to complicate methods in ways that make them less suited to taking advantage of general methods leveraging computation. There were many examples of AI researchers' belated learning of this bitter lesson, and it is instructive to review some of the most prominent.

In computer chess, the methods that defeated the world champion Kasparov in 1997 were based on massive deep search. At the time, this was looked upon with dismay by the majority of chess players who had human understanding of the special structure of chess. When these methods proved vastly more effective, these human-knowledge-based methods were abandoned. ... many examples of AI researchers' belated learning of this bitter lesson. These researchers wanted methods based on human input to win and were disappointed when they did not.



# Is this what we want ?



Transformer  
or whatever  
comes next.

AI: Trust me,  
This is a Higgs Decay

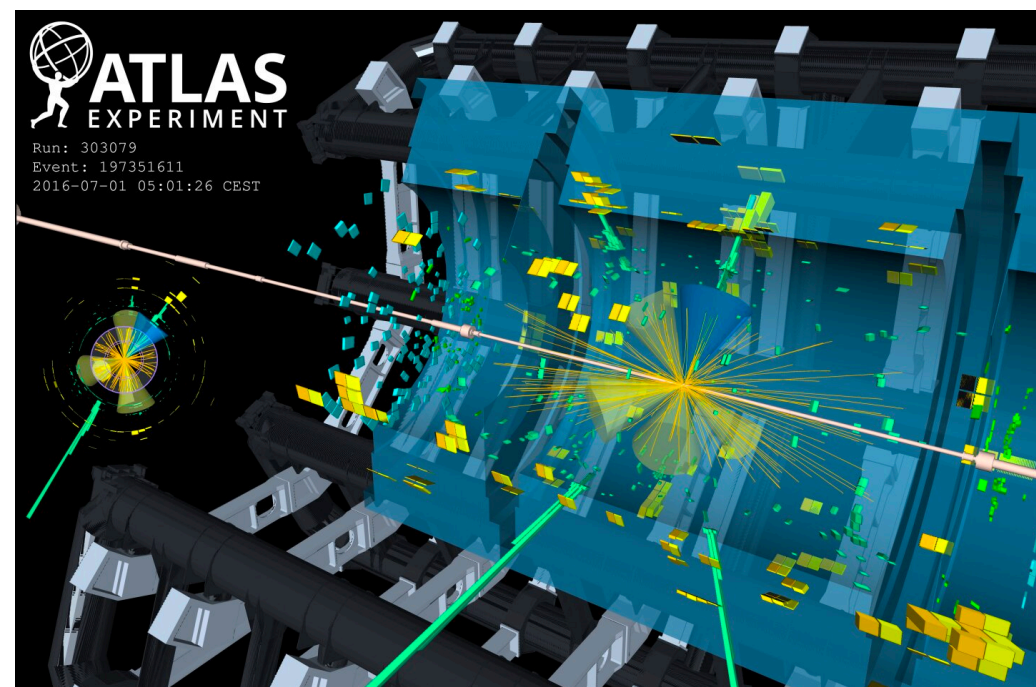
Human: Ok.

**A end to end “Hits to Higgs” system seems neither feasible (today) nor particularly desirable. As scientists we usually want more.**

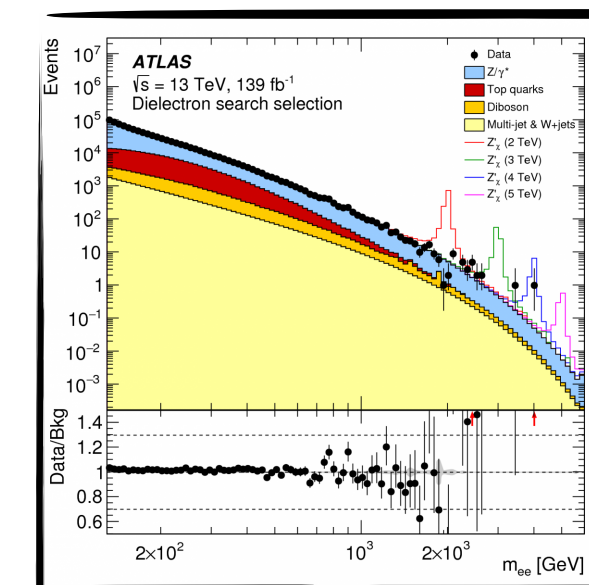
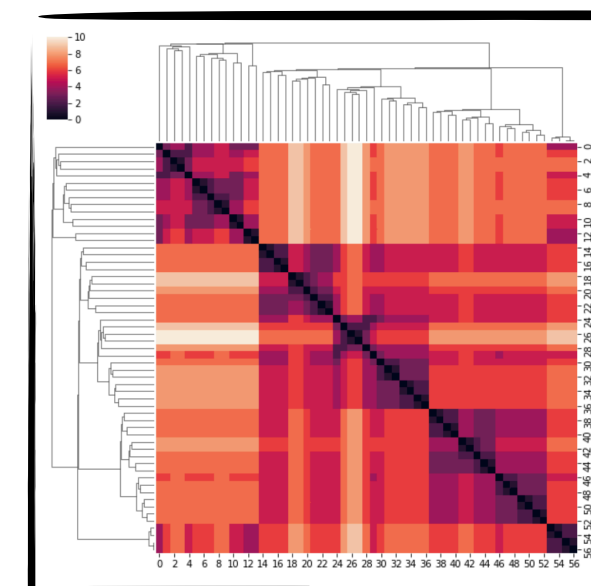
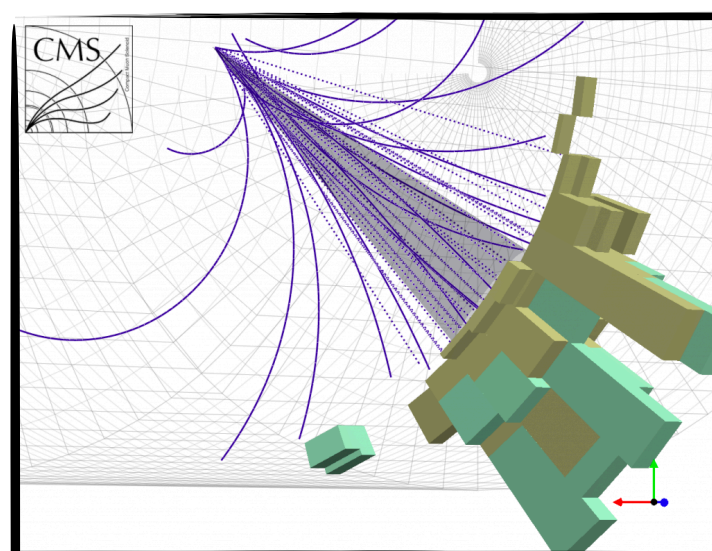
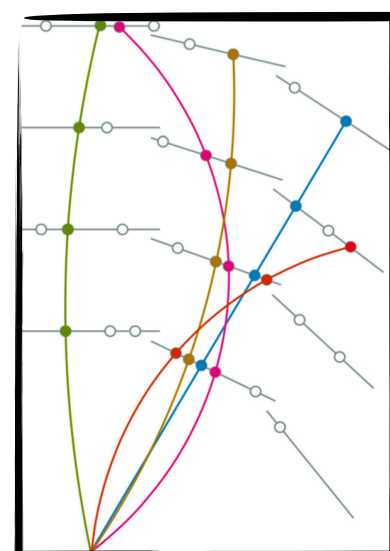
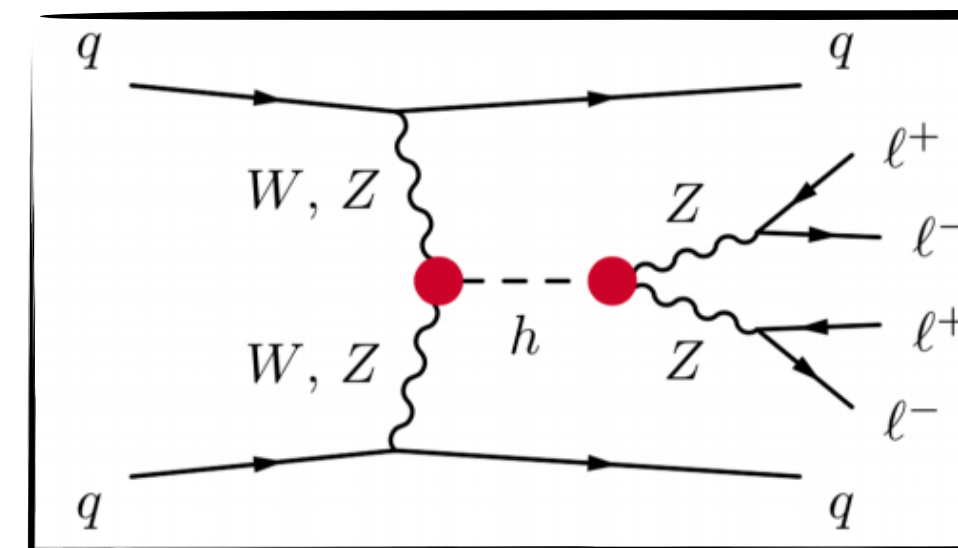
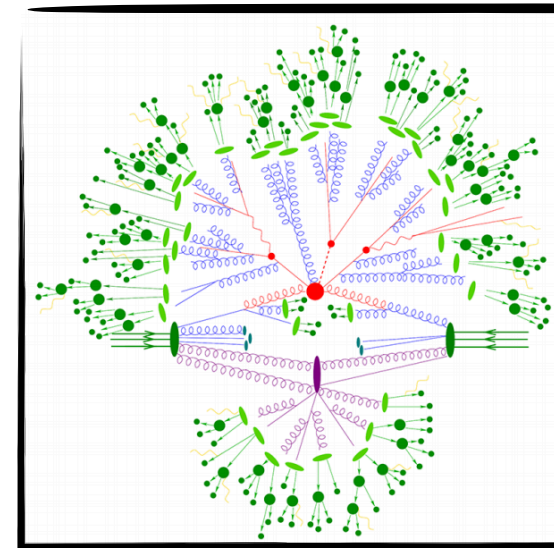
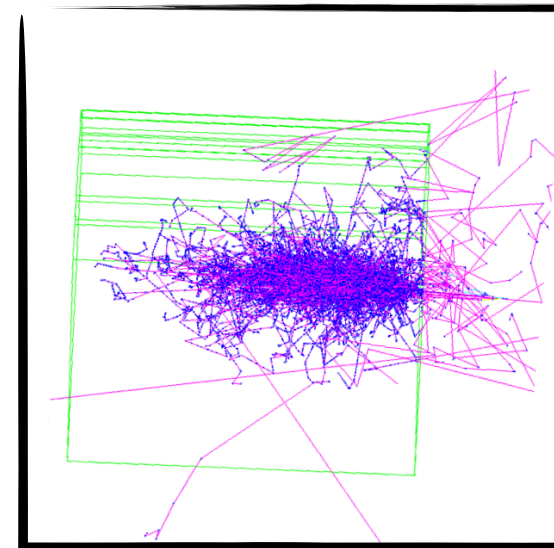


# Probably not.

The compositional & hierarchical nature of the data is core to our understanding. Sacrificing all of it for a non-descript “latent space”? *Some of it yes, but probably not everything.*



Data



$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + i\bar{\Psi}\not{D}\Psi + h.c. + \bar{\Psi}_i y_{ij} \Psi_j \phi + h.c. + \frac{1}{2} \mu \phi^2 - V(\phi)$$

Theory



# What do we want?

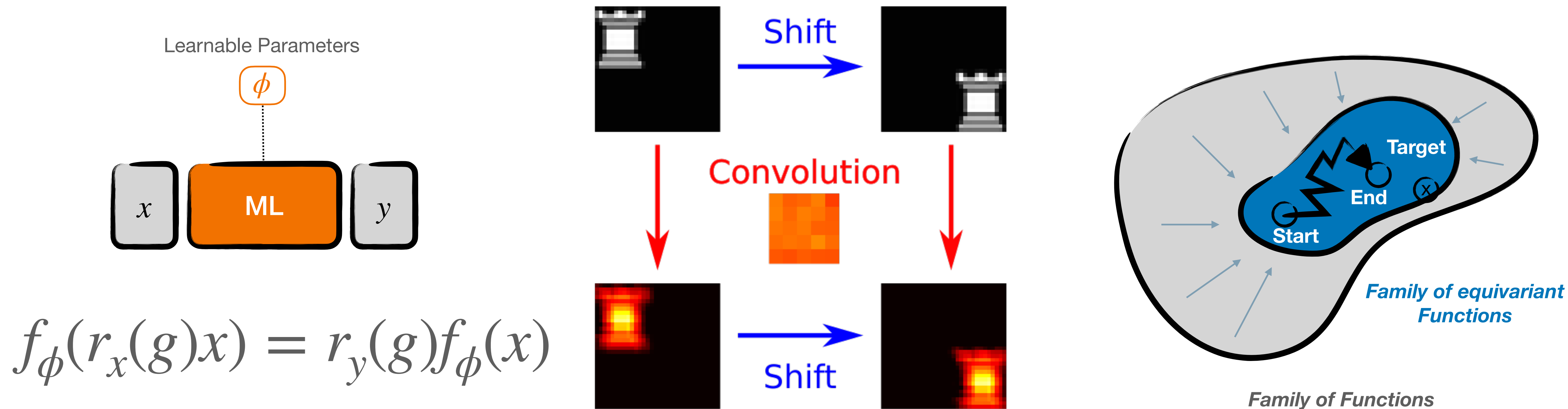
**We don't know (yet) how much domain knowledge will remain important and how much we can leave up to the machine**

**Need framework to build hybrid Physics-AI systems, learnable but with flexible control where & how much physics to put in.**



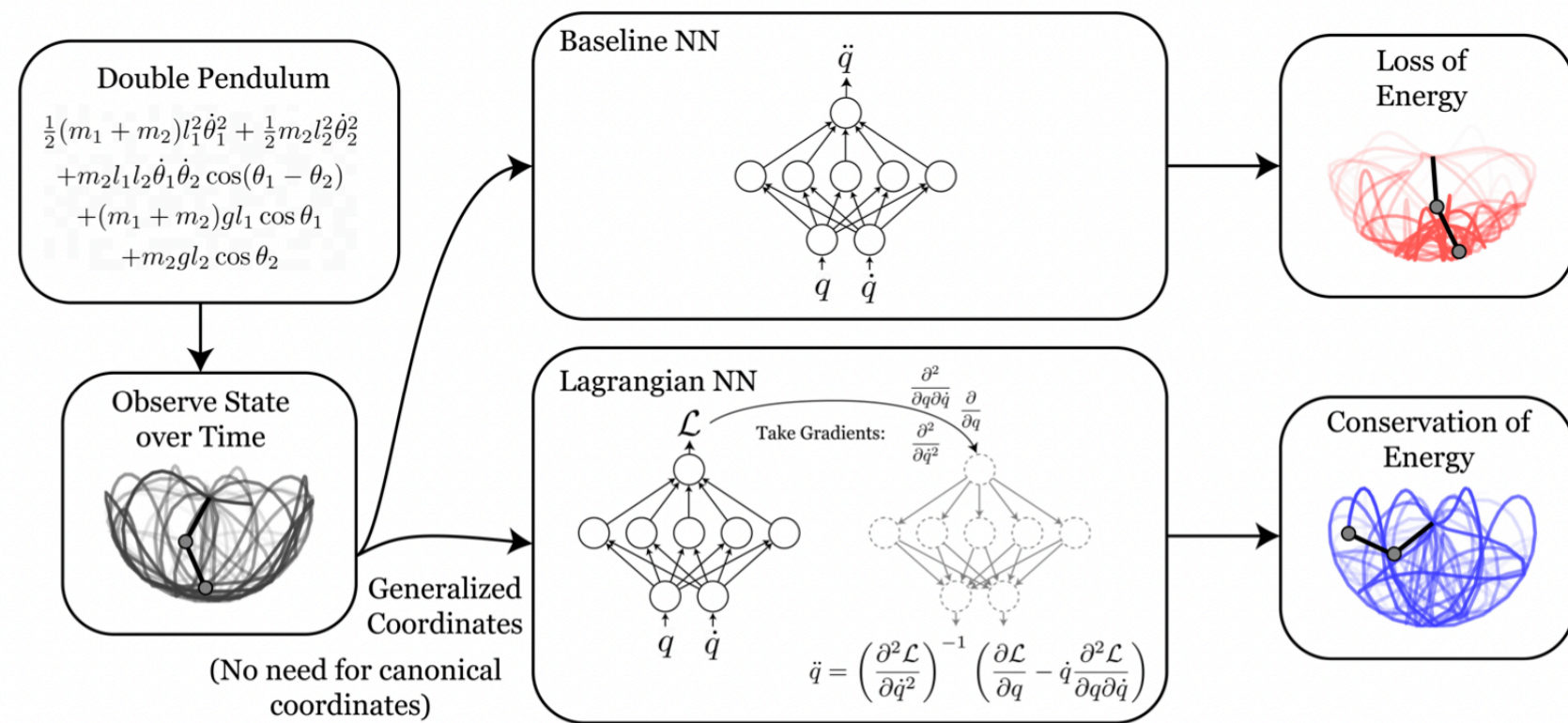
# Architectures and the Inductive Bias Story

Initially the simplest way to integrate physics into AI systems is through adding constraints: **inductive bias**



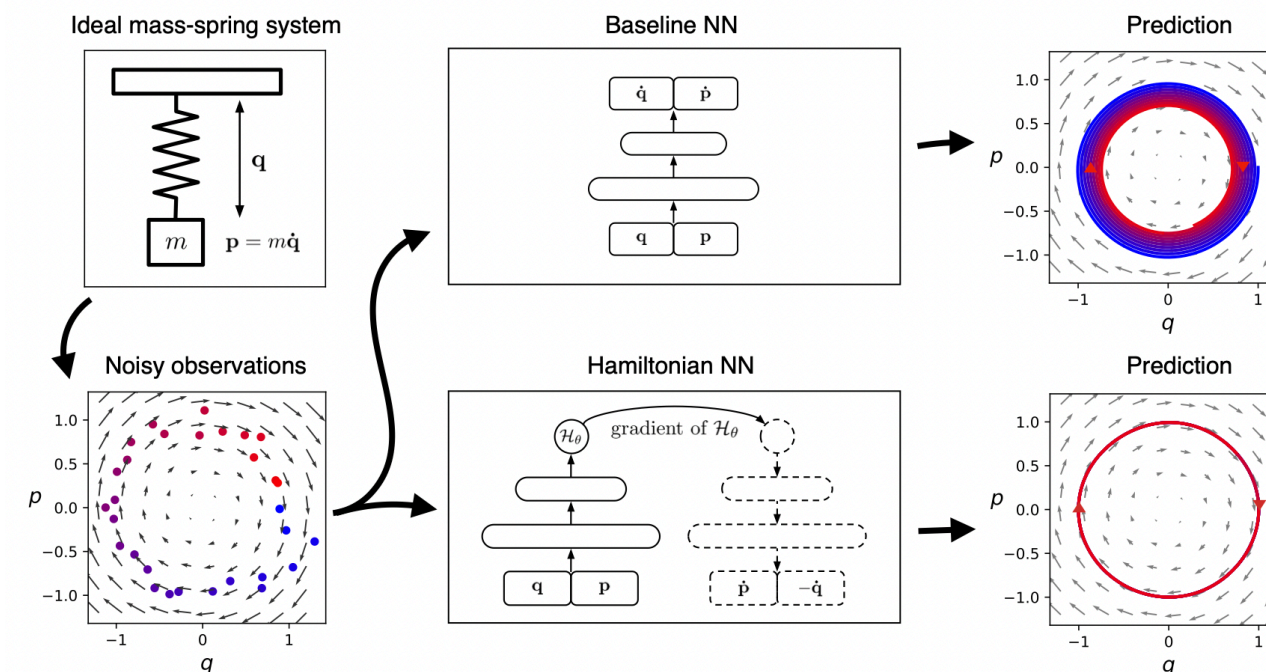
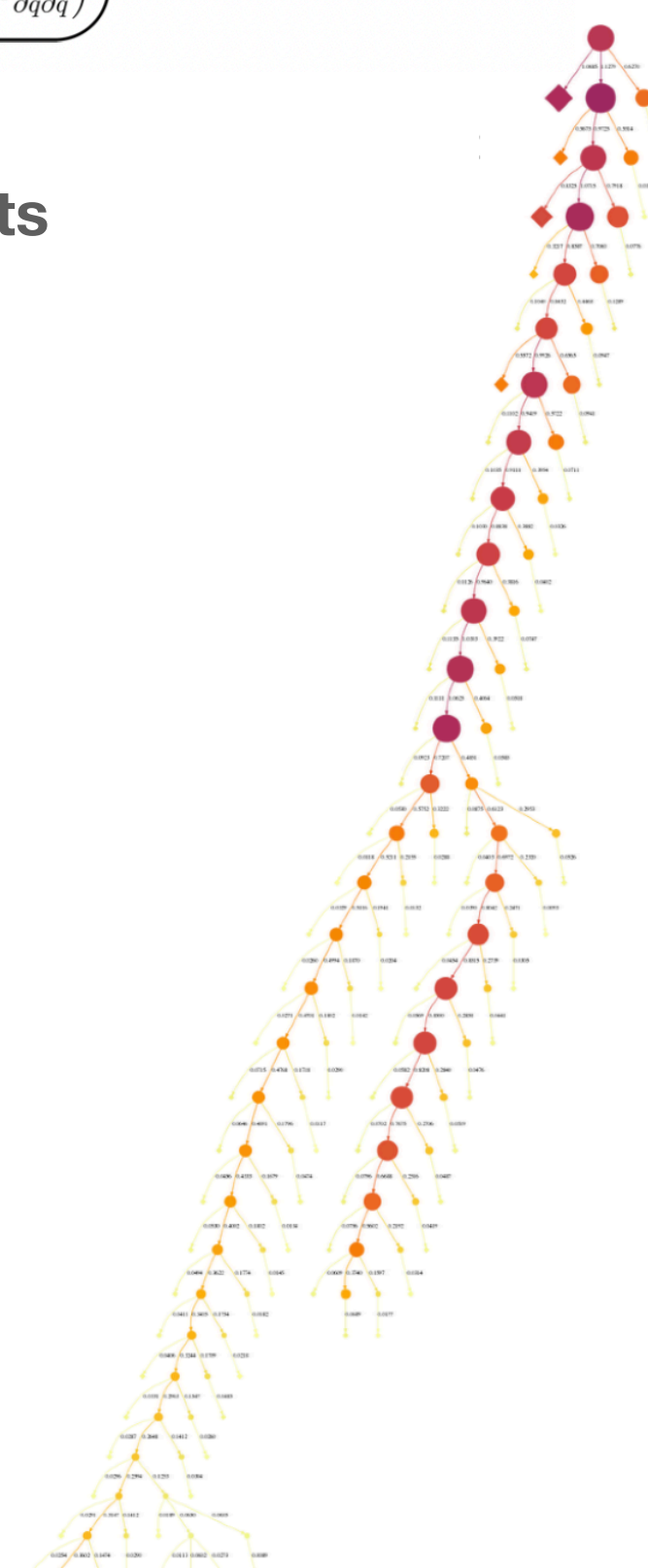


# Can't keep a Physicist from Symmetries



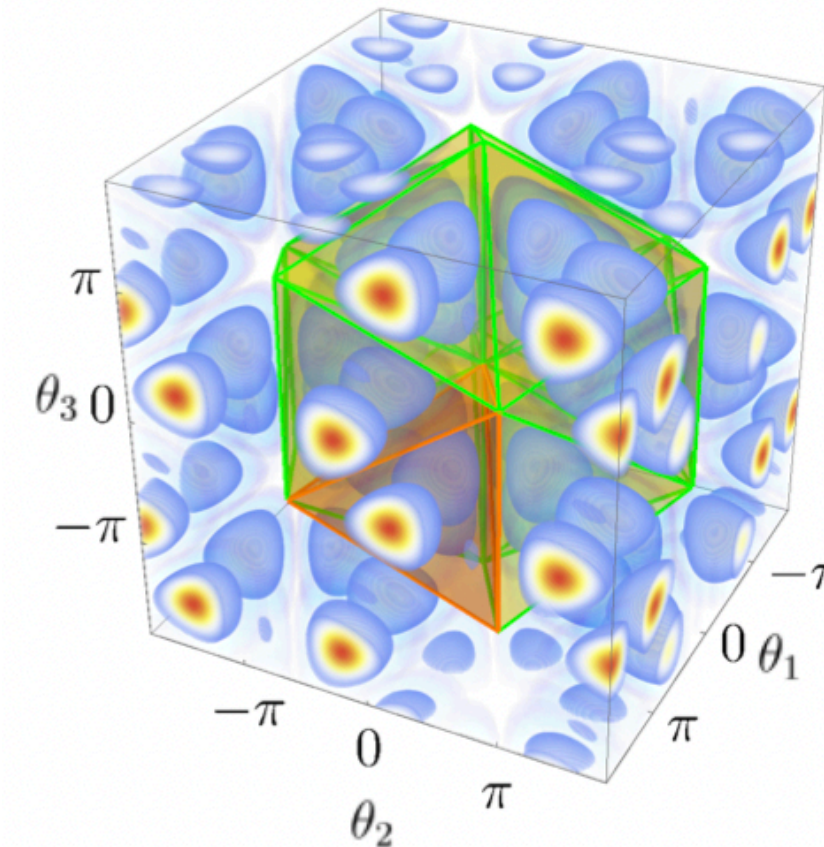
**Lagrangian Neural Nets**  
 arXiv: 2003.04630

**Neural Nets with QCD-like Structure**  
 arXiv:1702.00748

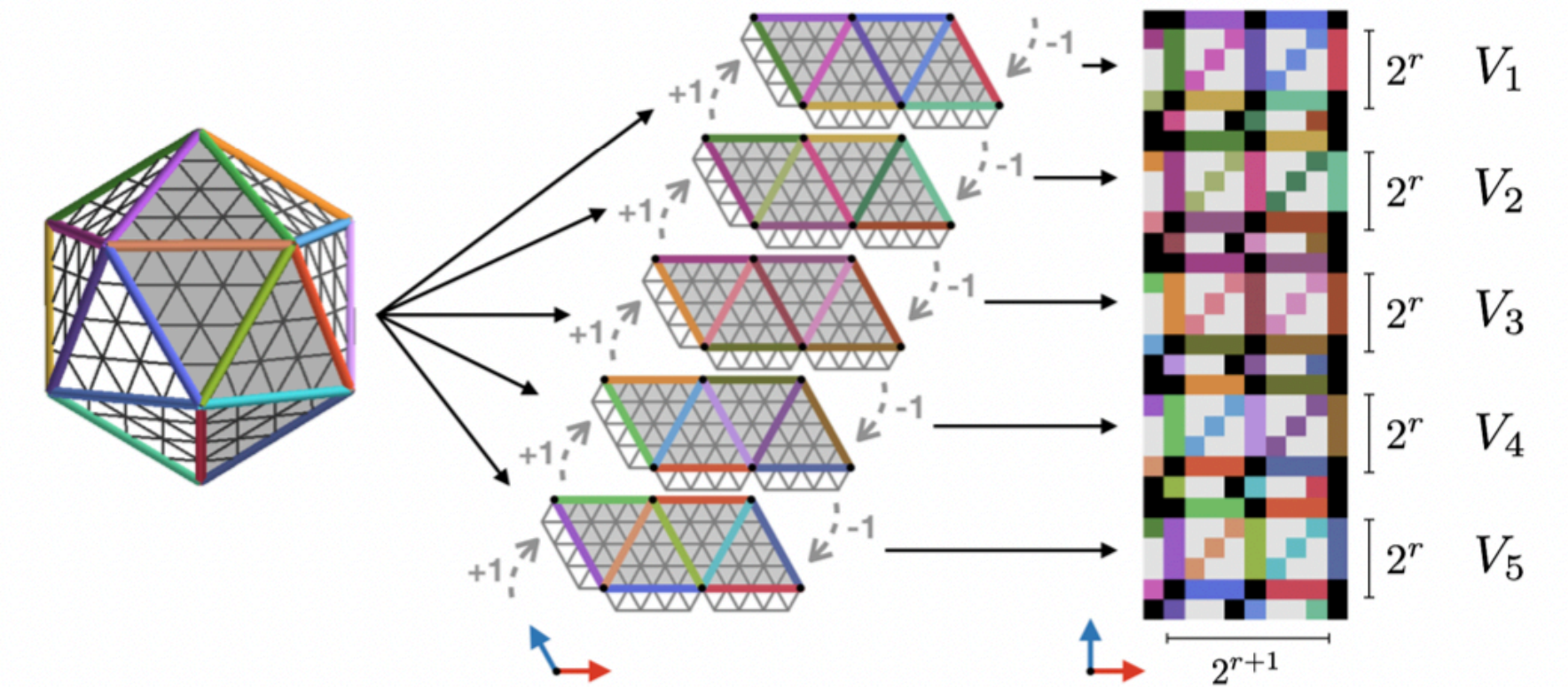


**Hamiltonian Neural Nets**  
 arXiv:1906.01563

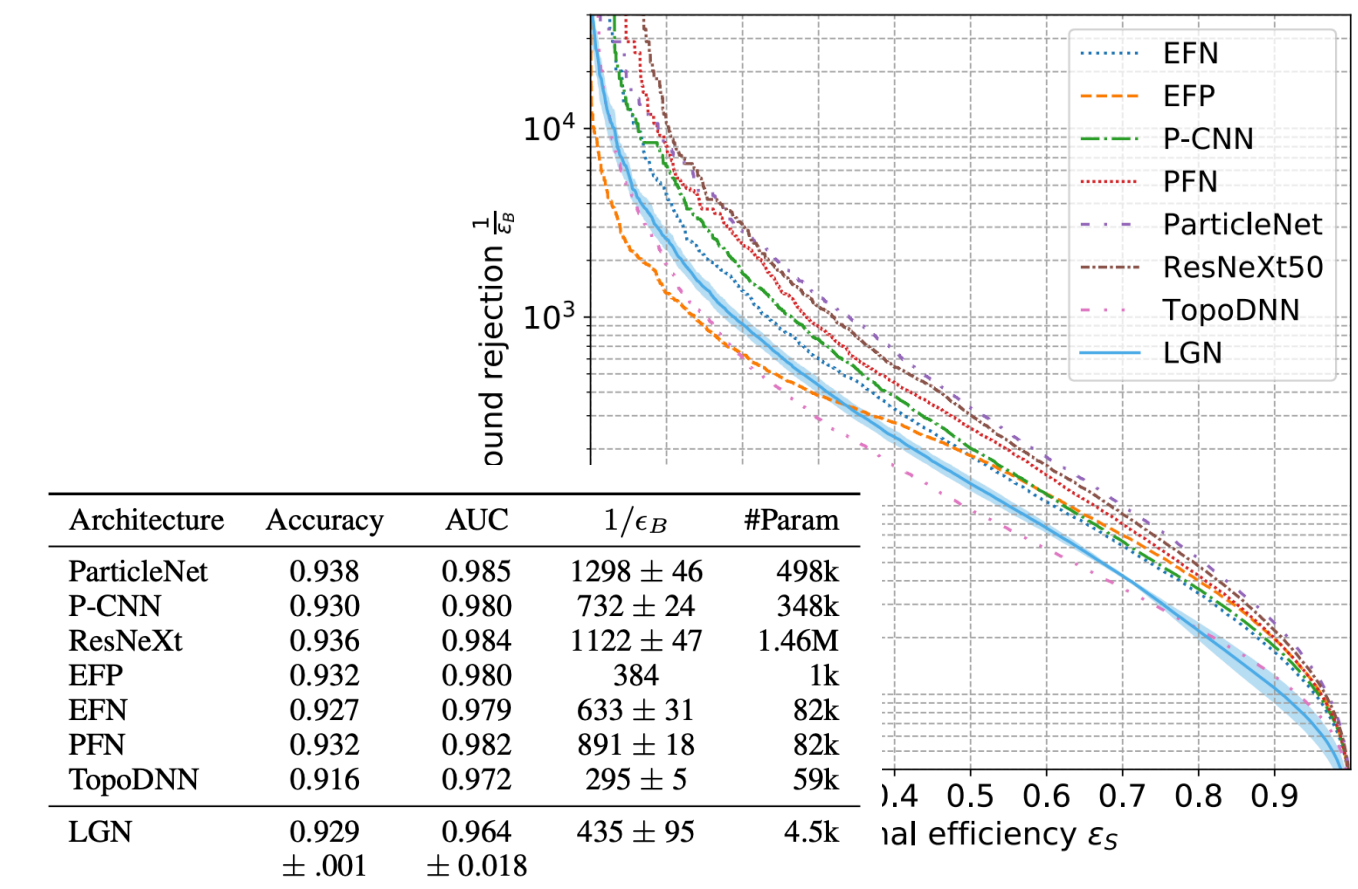
Haar SU(4) 3D projection



**SU(N)-Equivariant Normalizing Flows**



**Gauge-Equivariant Convolutional Neural Networks**



**Lorentz-Invariance**

arXiv:2006.04780

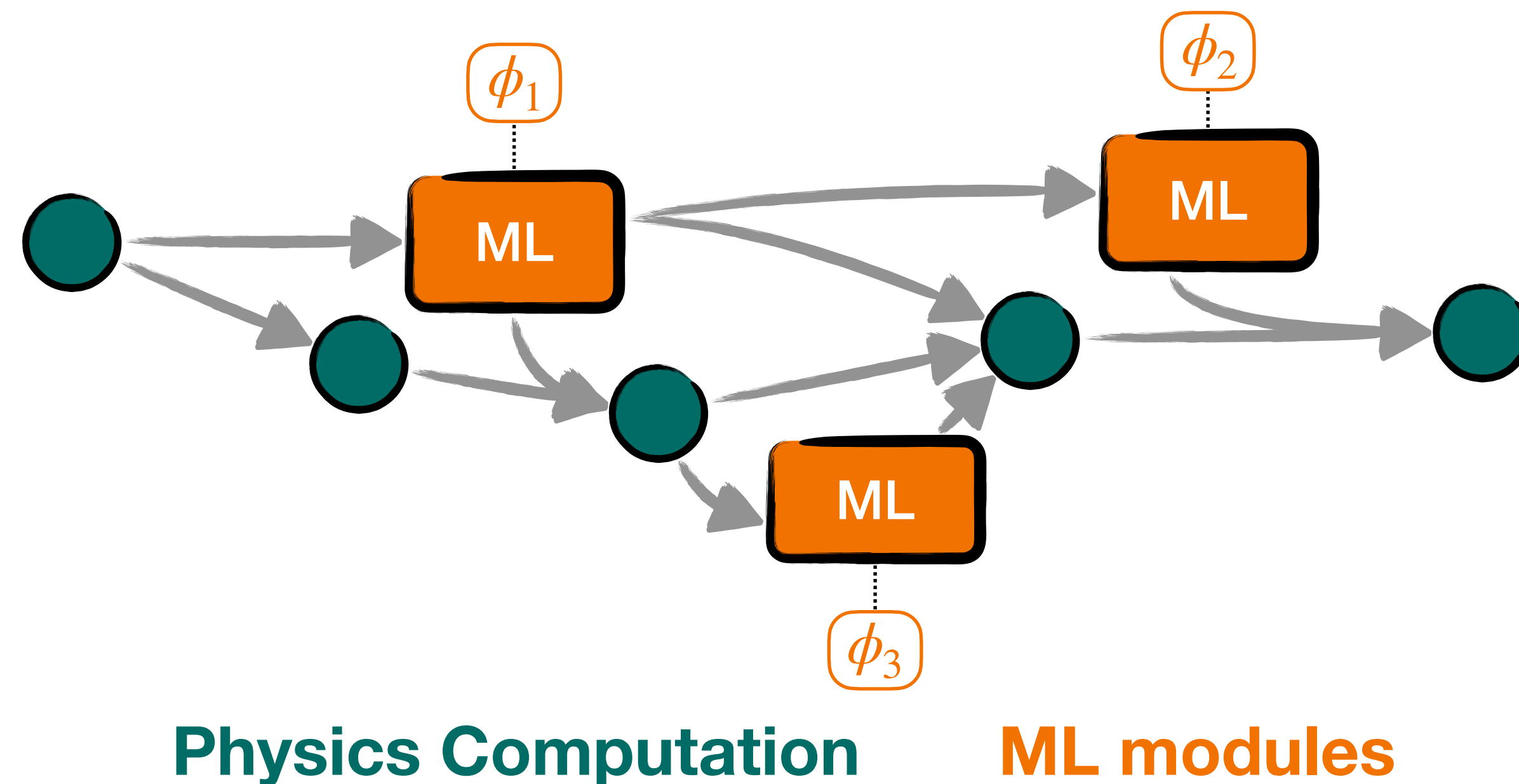


# But there is more

In (experimental) HEP what we think of as physics knowledge is often less about symmetries but more through

meaning of internal states

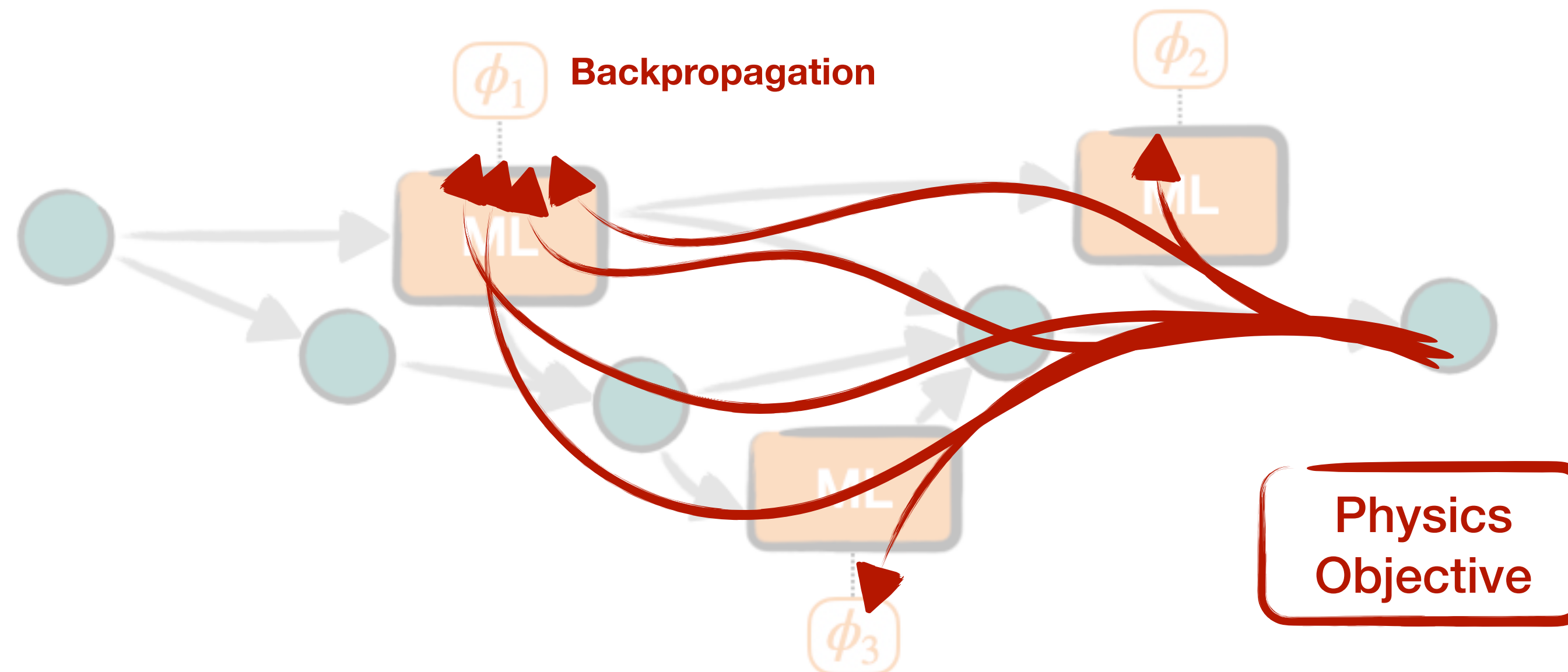
data flow between states





# But there is more

**Optimize this hybrid system directly:** Instead of training separately before assembling a pipeline → train the **ML in-situ**



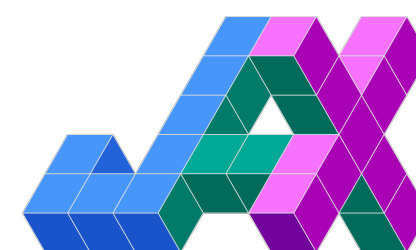
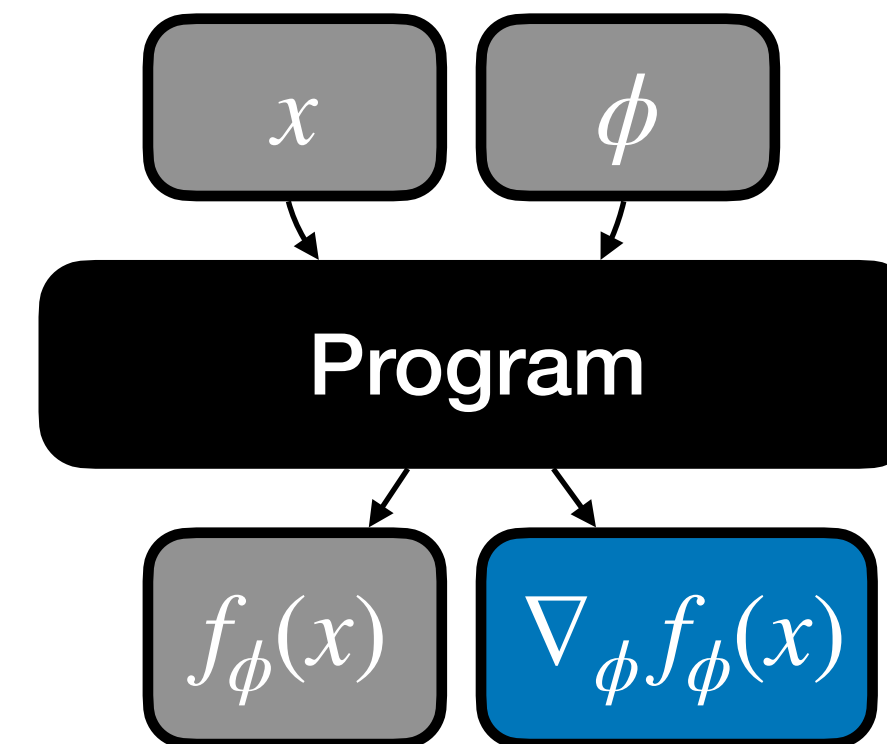
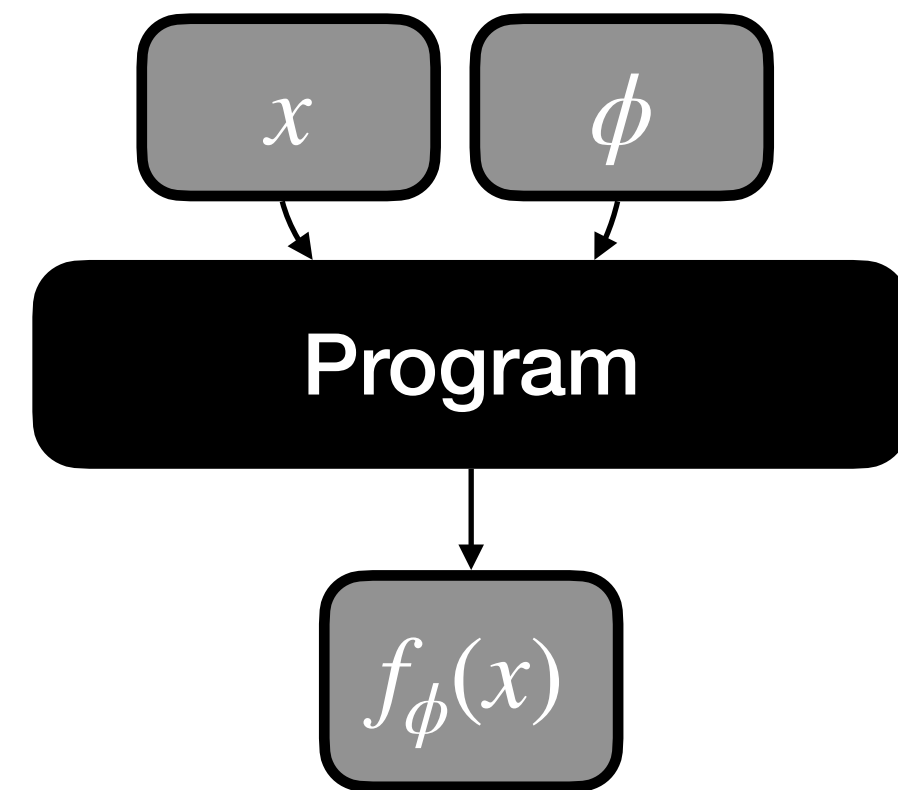
***For this to work, the physics components must play nice gradient descent → differentiable programming***



# Differentiable Programming

At the core, ML frameworks are programming languages that produce “enhanced programs” → add gradient information

General purpose, ready for physics



PYTORCH



# From beyond HEP

## Many project that (re-)write major software components in differentiable languages for us in hybrid AI systems

Accepted at the ICLR 2024 Workshop on AI4Differential Equations In Science

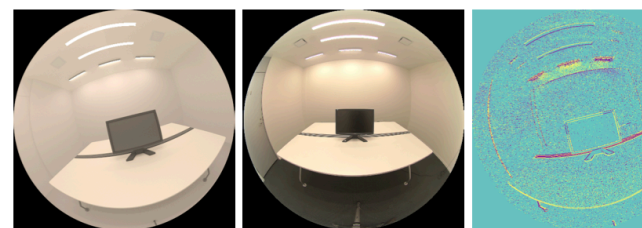
### JAX-SPH: A DIFFERENTIABLE SMOOTHED PARTICLE HYDRODYNAMICS FRAMEWORK

Artur P. Toshev<sup>†,1</sup>, Harish Ramachandran<sup>\*,1</sup>, Jonas A. Erbesdobler<sup>\*,1</sup>, Gianluca Colletti<sup>\*,2</sup>, Johannes Brandstetter<sup>3,4</sup> &

- <sup>1</sup> Chair of Aerodynamics and
- <sup>2</sup> Independent researcher
- <sup>3</sup> ELLIS Unit Linz, LIT AI L
- <sup>4</sup> NXAI GmbH, Linz, Austria
- <sup>5</sup> Munich Institute of Integrat
- <sup>†</sup> artur.toshev@tum.d
- <sup>\*</sup> Equal contribution

### Differentiable Monte Carlo Ray Tracing through Edge Sampling

TZU-MAO LI, MIT CSAIL  
MIIKA AITTALA, MIT CSAIL  
FRÉDO DURAND, MIT CSAIL  
JAAKKO LEHTINEN, Aalto University & NVIDIA



(a) initial guess (b) real photograph (c) camera gradient (per-pixel contribut

Fig. 1. We develop a general-purpose differentiable renderer that is capa with respect to scene parameters, such as camera pose (c), material pa computed from the output image. (c) shows the per-pixel gradient cont shows the gradient with respect to the red channel of table albedo. (e) sho As one of our applications, we use our gradient to perform an inverse rer (a) with a manual geometric recreation of the scene. The scene contains a optimize for camera pose, material parameters, and light source intens method generates image (f) that almost matches the photo reference.

Gradient-based methods are becoming increasingly important for compu graphics, machine learning, and computer vision. The ability to compu gradients is crucial to optimization, inverse problems, and deep learning. rendering, the gradient is required with respect to variables such as came parameters, light sources, scene geometry, or material properties. Howev

-ph] 3 Dec 2020

### JAX, M.D. A Framework for Differentiable Ph

Samuel S. Schoenholz  
Google Research: Brain Team  
schsam@google.com

Ekin D.  
Google Research  
cubuk@goo

#### Abstract

We introduce JAX MD, a software package for performing diffi simulations with a focus on molecular dynamics. JAX MD in of physics simulation environments, as well as interaction pote networks that can be integrated into these environments without

DRAFT VERSION FEBRUARY 13, 2024  
Typeset using L<sup>A</sup>T<sub>E</sub>X **modern** style in AAS<sub>T</sub>E<sub>X</sub>631

### Differentiable Cosmological Simulation with the Adjoint Method

YIN LI (李寅)<sup>1,2,3</sup>, CHIRAG MODI<sup>2,3</sup>, DREW JAMIESON<sup>4</sup>, YUCHENG ZHANG (张宇澄)<sup>1,5</sup>, LIBIN LU (陆利彬)<sup>2</sup>, YU FENG (冯雨)<sup>6</sup>, FRANÇOIS LANUSSE<sup>7</sup>, AND LESLIE GREENGARD<sup>2,8</sup>

- <sup>1</sup>Department of Mathematics and Theory, Peng Cheng Laboratory, Shenzhen, Guangdong 518066, China
- <sup>2</sup>Center for Computational Mathematics, Flatiron Institute, New York, New York 10010, USA
- <sup>3</sup>Center for Computational Astrophysics, Flatiron Institute, New York, New York 10010, USA
- <sup>4</sup>Max Planck Institute for Astrophysics, 85748 Garching bei München, Germany
- <sup>5</sup>Center for Cosmology and Particle Physics, Department of Physics, New York University, New York, New York 10003, USA
- <sup>6</sup>Berkeley Center for Cosmological Physics, University of California, Berkeley, California 94720, USA
- <sup>7</sup>AIM, CEA, CNRS, Université Paris-Saclay, Université Paris Diderot, Sorbonne Paris Cité, F-91191 Gif-sur-Yvette, France
- <sup>8</sup>Courant Institute, New York University, New York, New York 10012, USA

#### ABSTRACT

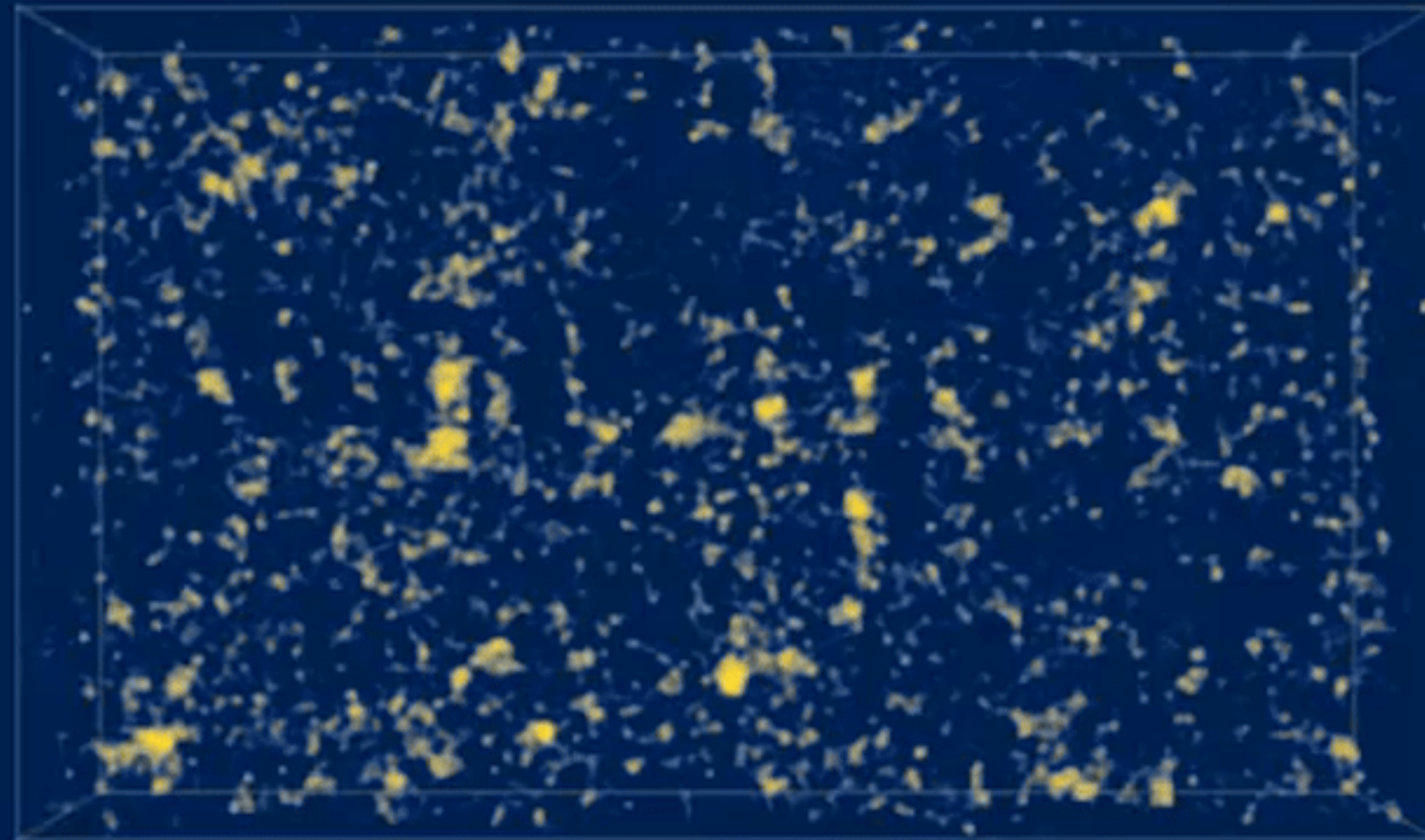
Rapid advances in deep learning have brought not only myriad powerful neural networks, but also breakthroughs that benefit established scientific research. In particular, automatic differentiation (AD) tools and computational accelerators like GPUs have facilitated forward modeling of the Universe with differentiable simulations. Based on analytic gradient-based backpropagation, current differentiable

815v2 [astro-ph.IM] 7 Feb 2024



# Optimizing a Differentiable Simulator

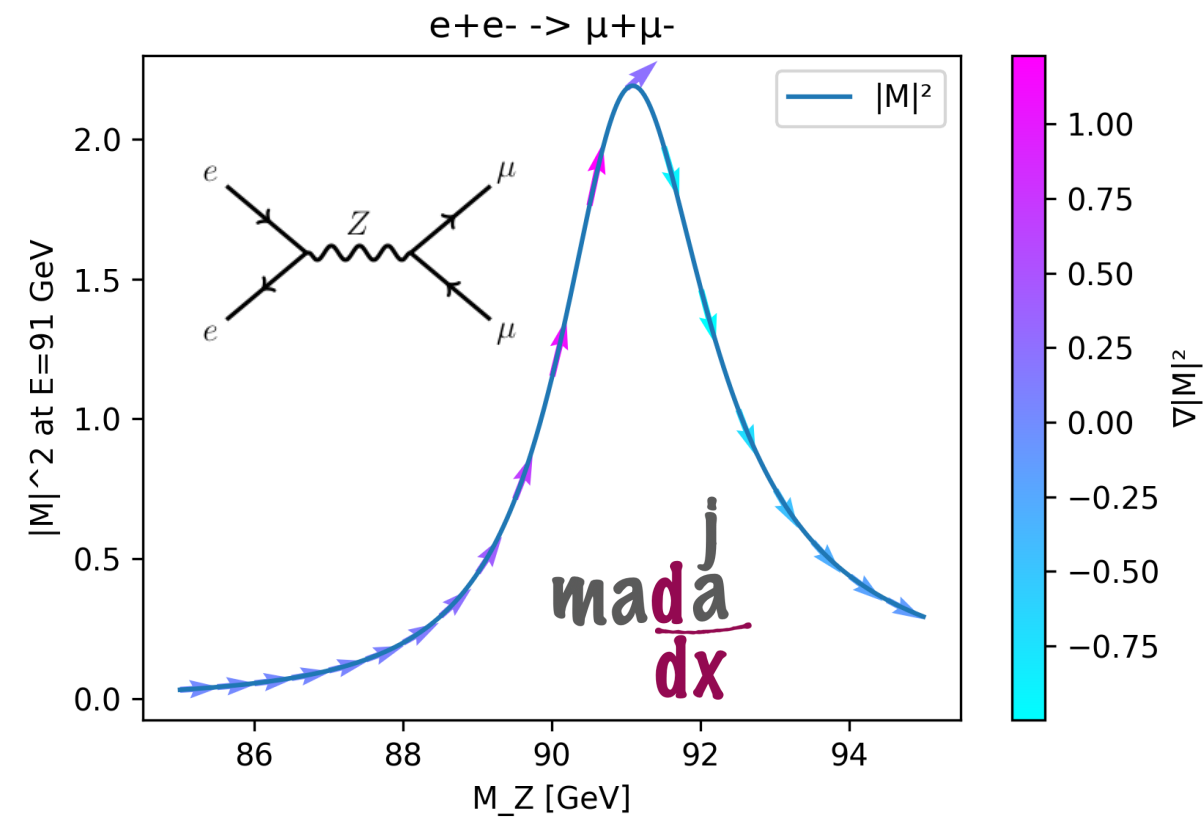
Optimizing initial conditions



Optimization iteration: 1

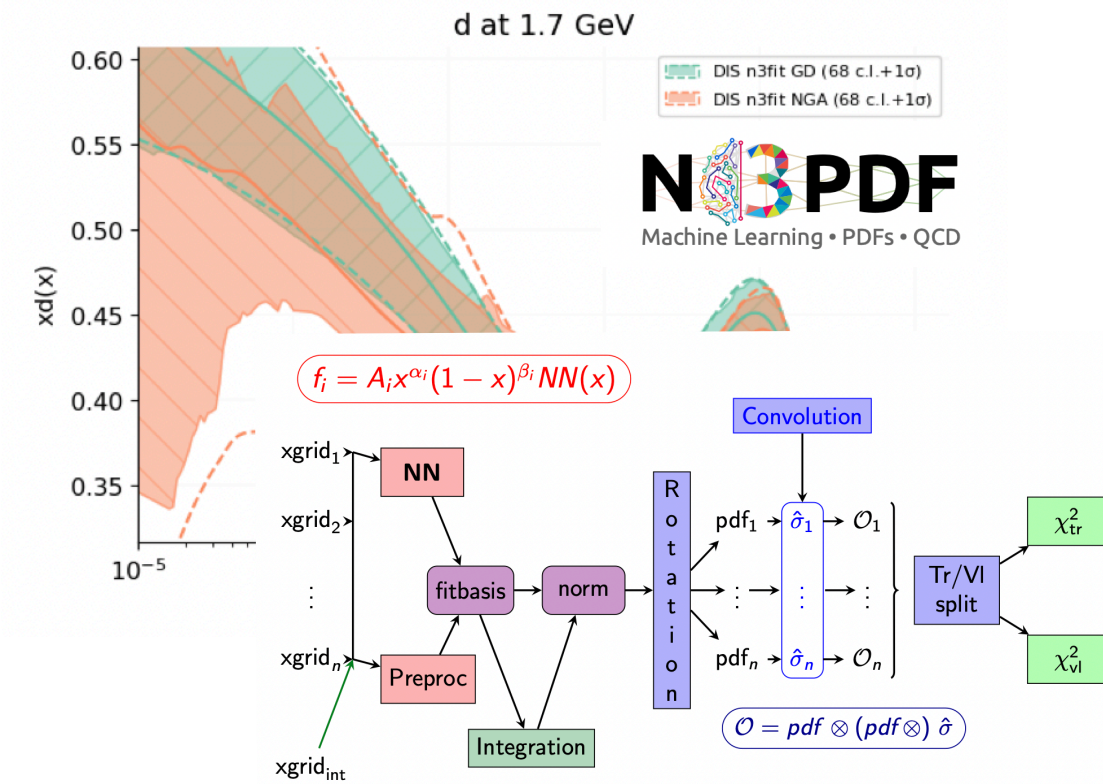


# In HEP & ai, we are pushing as well



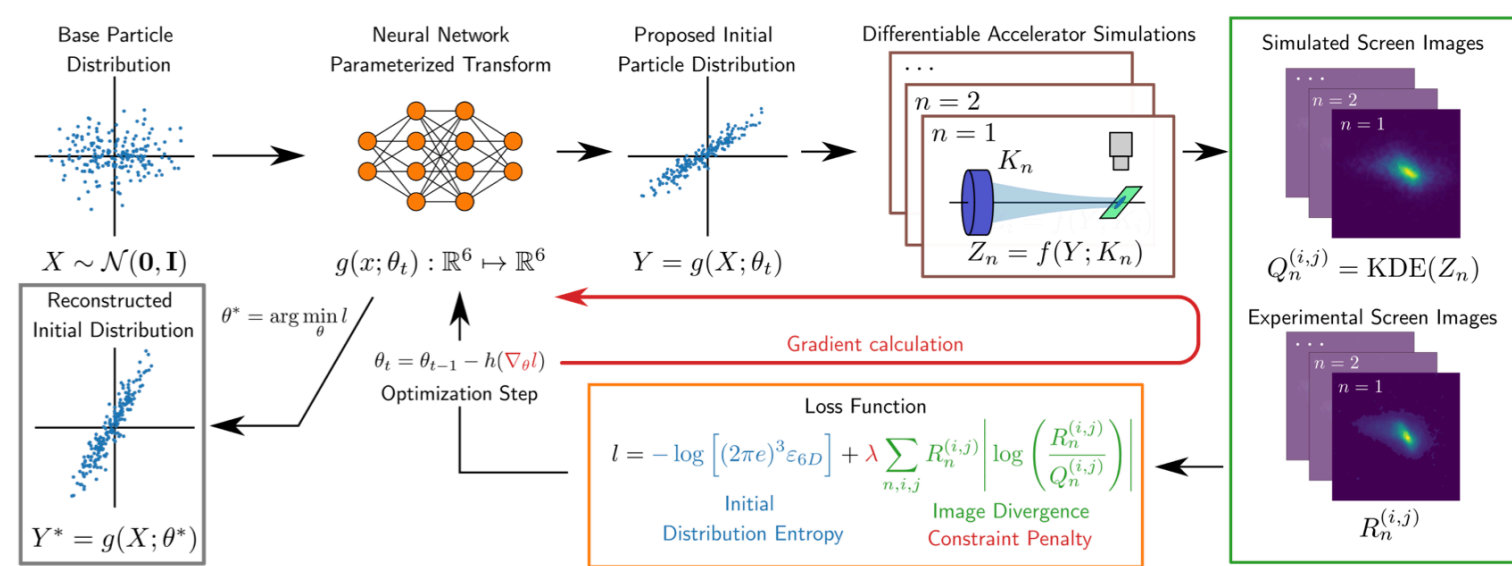
Differentiable Matrix Elements

2203.00057



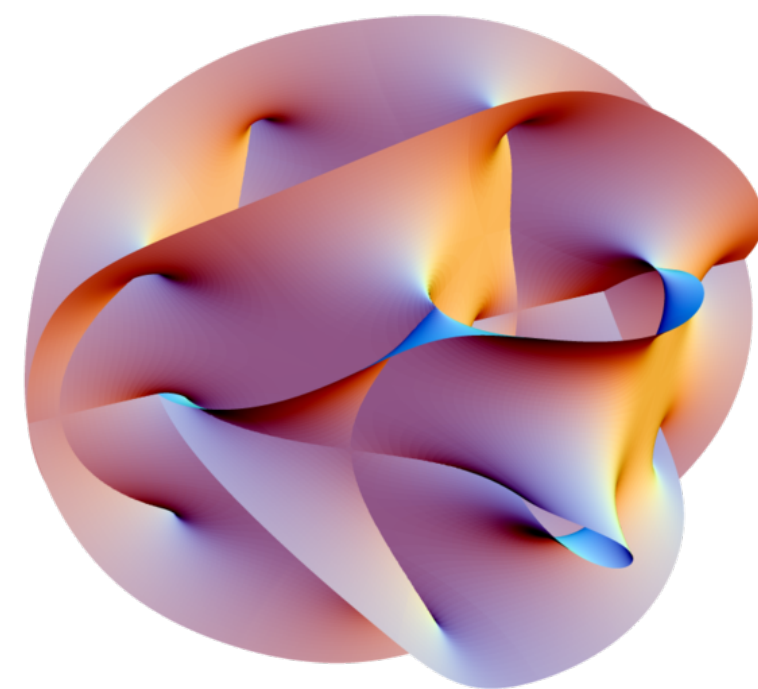
Differentiable Proton Structure

1907.05075



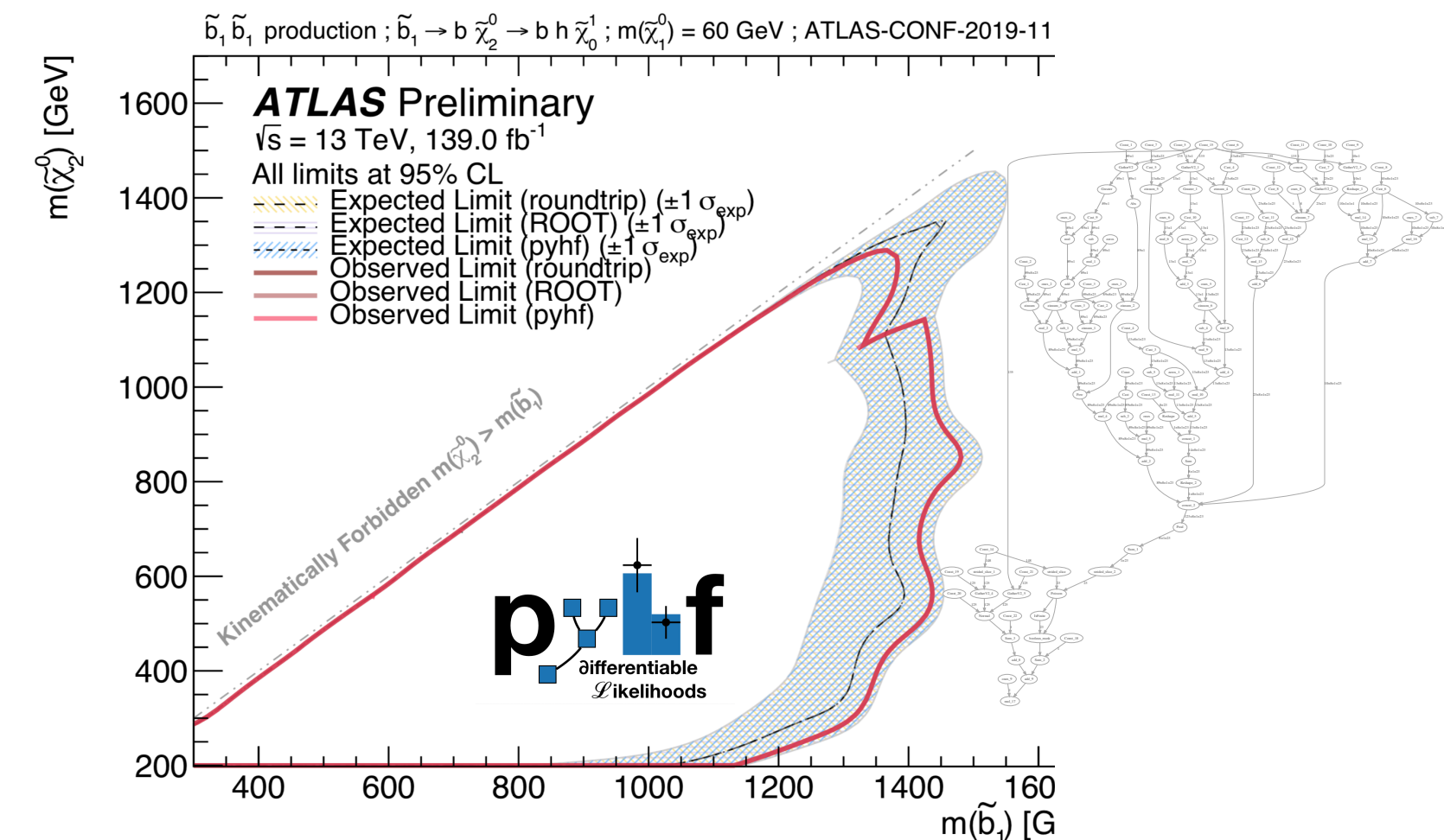
Differentiable Accelerator Simulation

2211.09077



Differentiable Calabi-Yau Manifolds

2211.12520

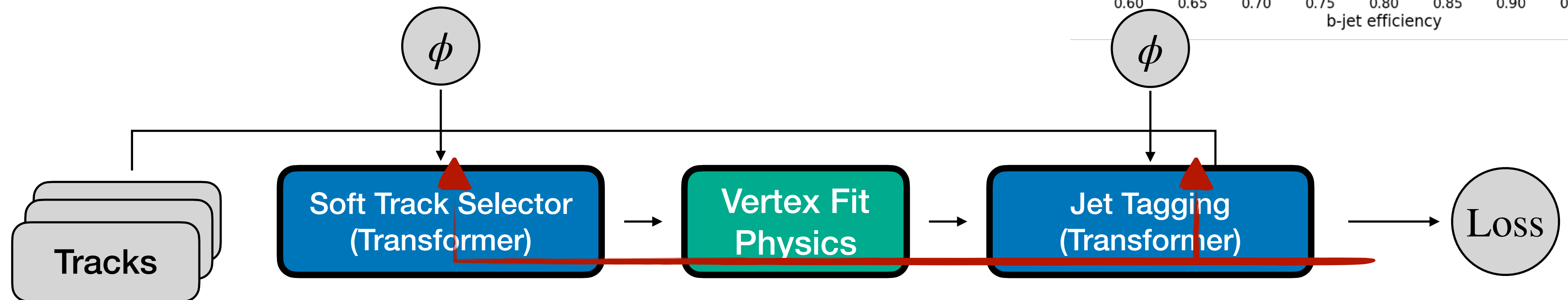
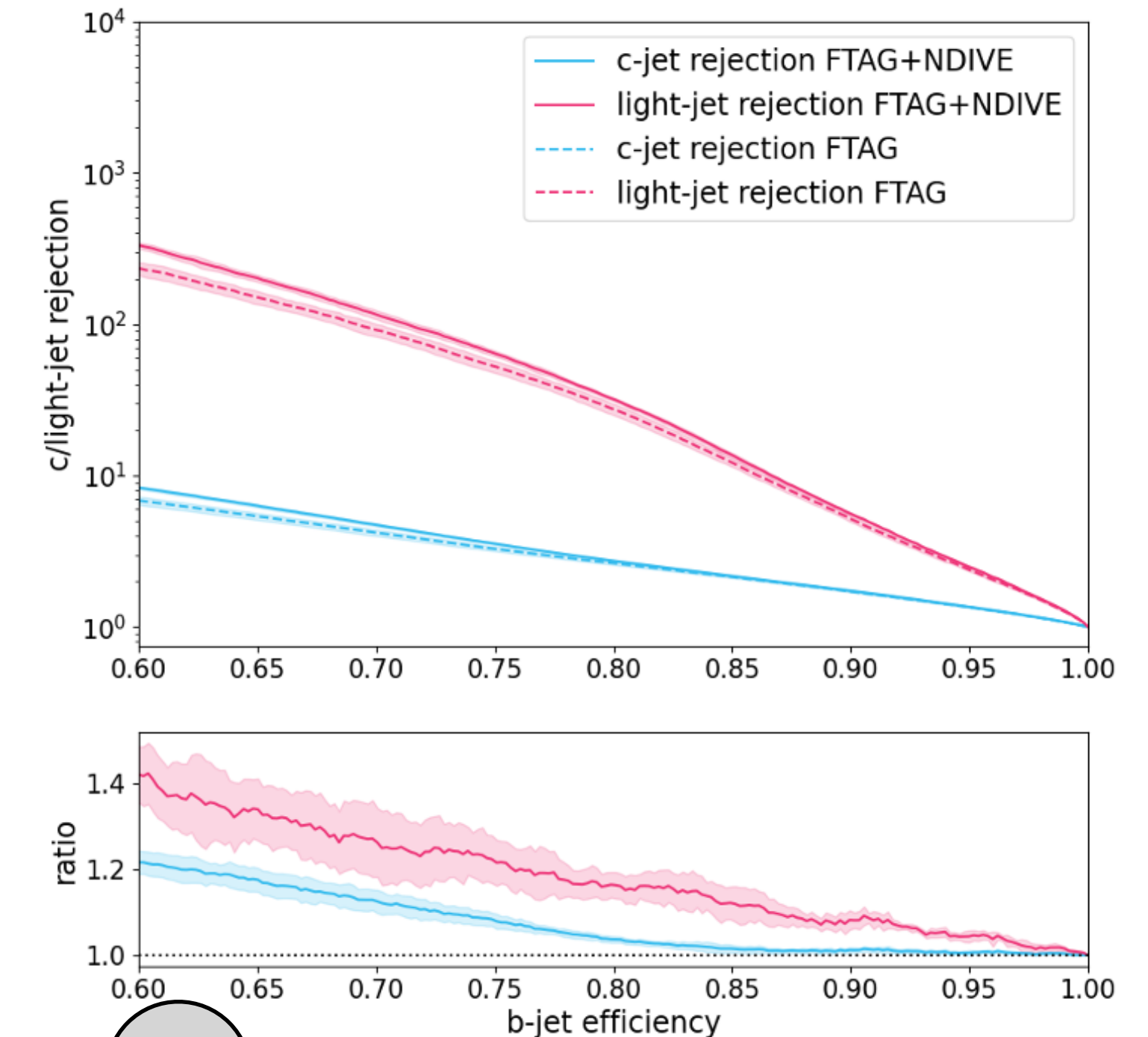
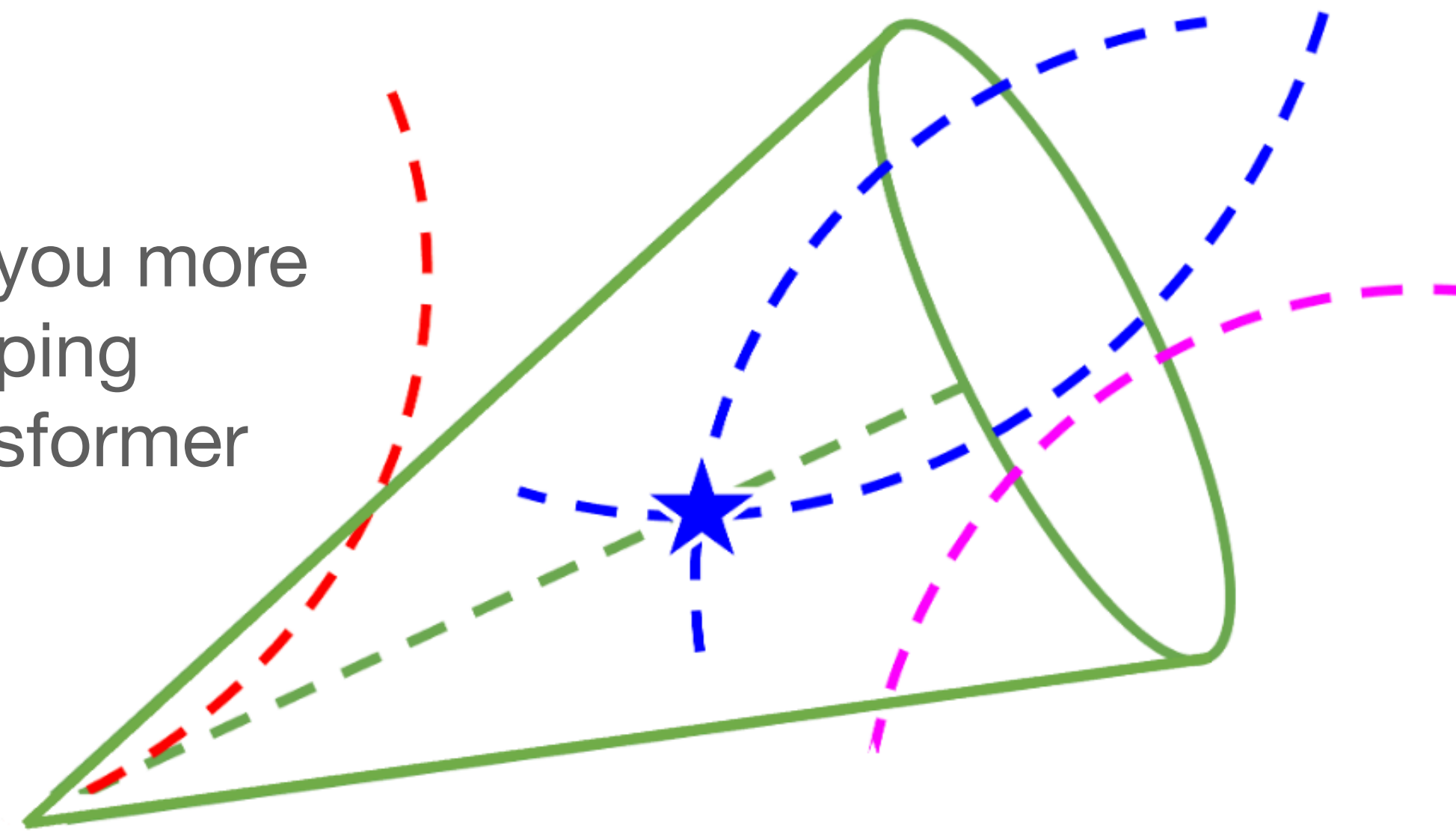


Differentiable Inference

10.21105/joss.02823

# Recent Example of a true Hybrid:

Adding Physics **does give** you more performance than just dumping the raw data into a big transformer



*Backpropagation*



# Not too worried about Integrations

Not only a ML thing. A huge community with experience in automatic differentiation in large C++, Julia, FORTRAN projects: **Even RooFit has become differentiable**

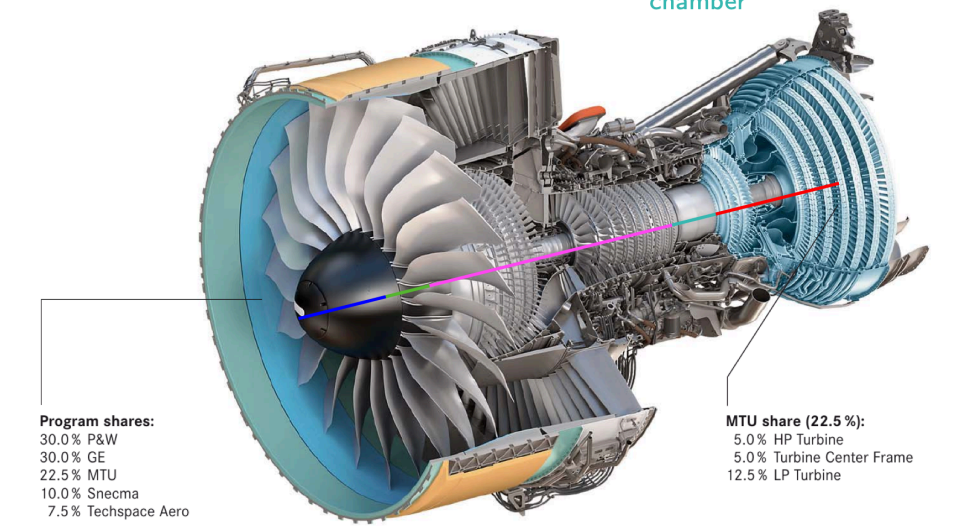
M. Sagebaum - CoDIPack



Motivation - Why do we need derivatives?

Example: A380 Turbine: GP7000

Inlet Fan Compressor Combustion chamber Turbine



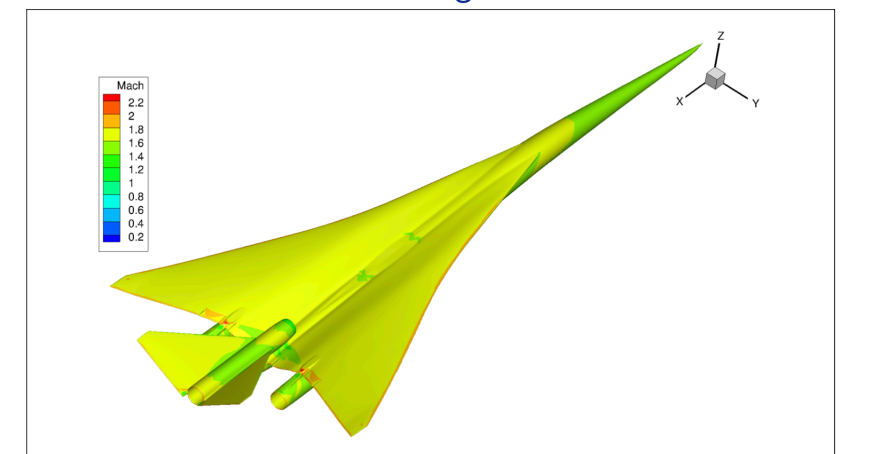
Program shares:  
30.0% P&W  
30.0% GE  
22.5% MTU  
10.0% Snecma  
7.5% Techspace Aero

MTU share (22.5%):  
5.0% HP Turbine  
5.0% Turbine Center Frame  
12.5% LP Turbine

Pictures www.mtu.de



Test case - Lockheed Martin 1021 configuration



- Explicit Runge-Kutta
- Euler scheme
- Mach: 1.6
- Angle of attack: 2.1
- Mesh size: 5,730,841
- Primal Memory: 16.7 Gb
- Primal time (one G step): 2.05 s
- Run on 32 processors of the Elwetritsch cluster from the TU Kaiserslautern

G. Singh, V. Vassilev et al - Clad

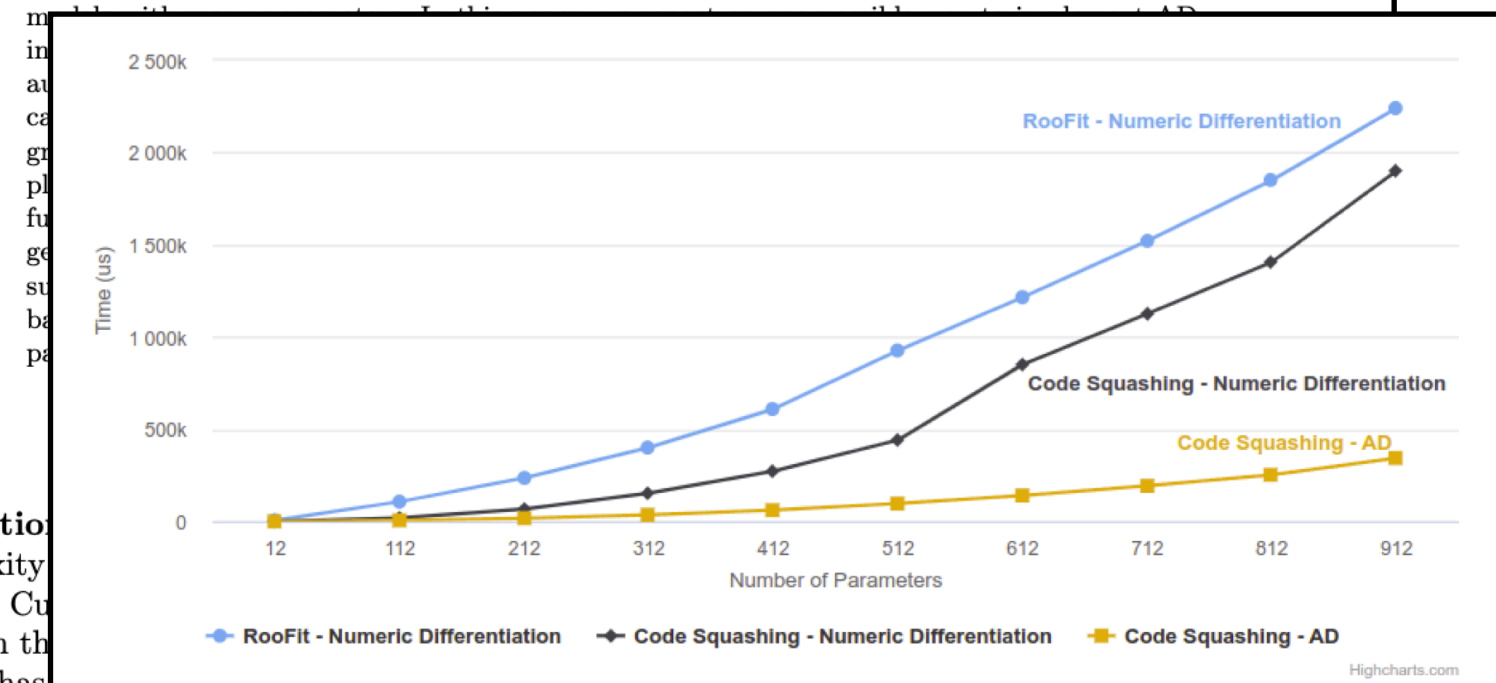
## Automatic Differentiation of Binned Likelihoods With RooFit and Clad

Garima Singh\*, Jonas Rembser†, Lorenzo Moneta†, David Lange\*, Vassil Vassilev\*

\* Department of Physics, Princeton University, Princeton, NJ 08544, USA  
† EP-SFT, CERN, Espl. des Particules 1, 1211 Meyrin, Switzerland

E-mail: garima.singh@cern.ch, jonas.rembser@cern.ch, lorenzo.moneta@cern.ch, david.lange@cern.ch, vassil.vassilev@cern.ch

**Abstract.** Just as data sets from next-generation experiments grow, processing requirements for physics analysis become more computationally demanding, necessitating performance optimizations for RooFit. One possibility to speed-up minimization and add stability is the use of Automatic Differentiation (AD). Unlike for numerical differentiation, the computation cost scales linearly with the number of parameters, making AD particularly appealing for statistical

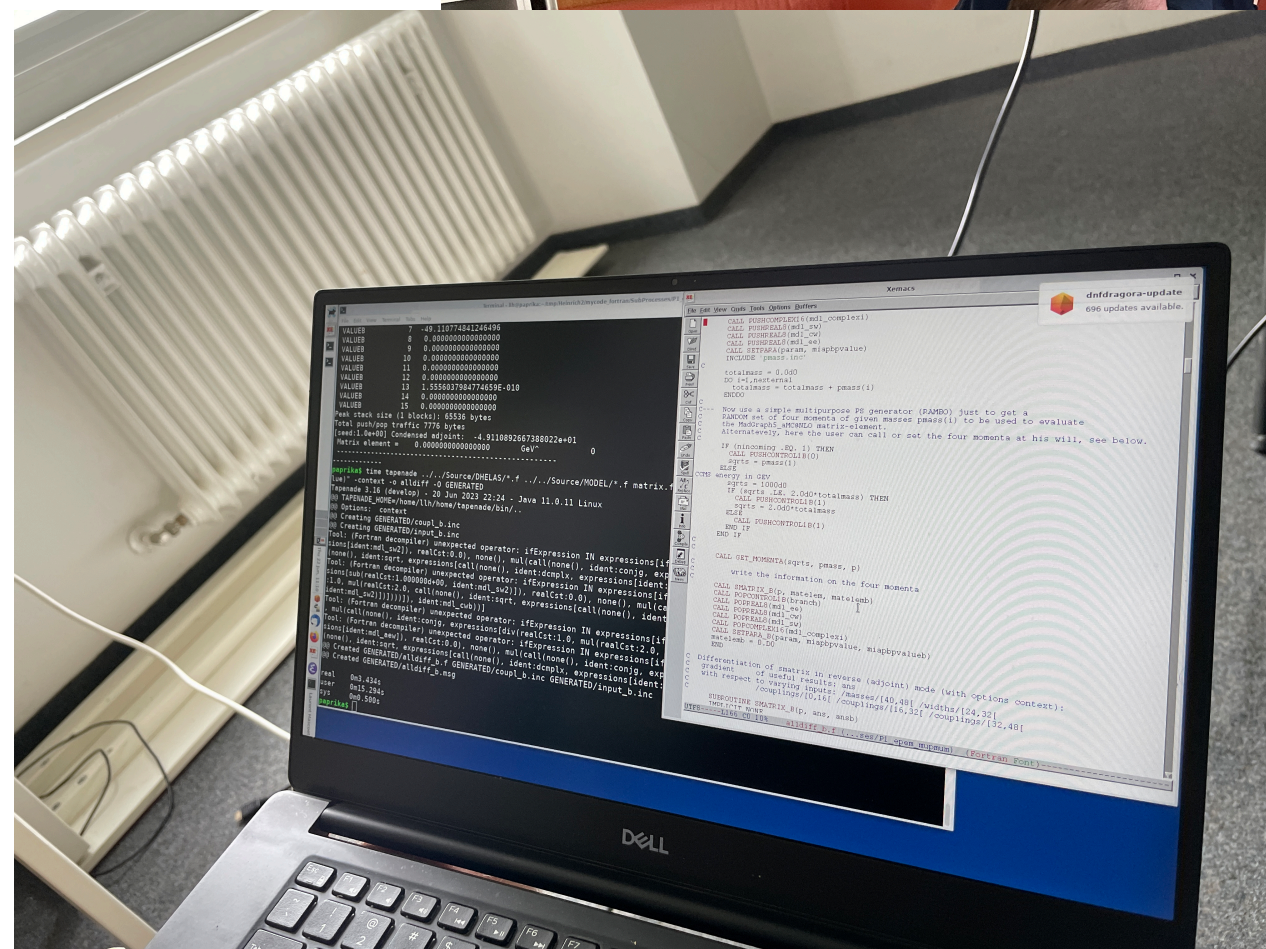
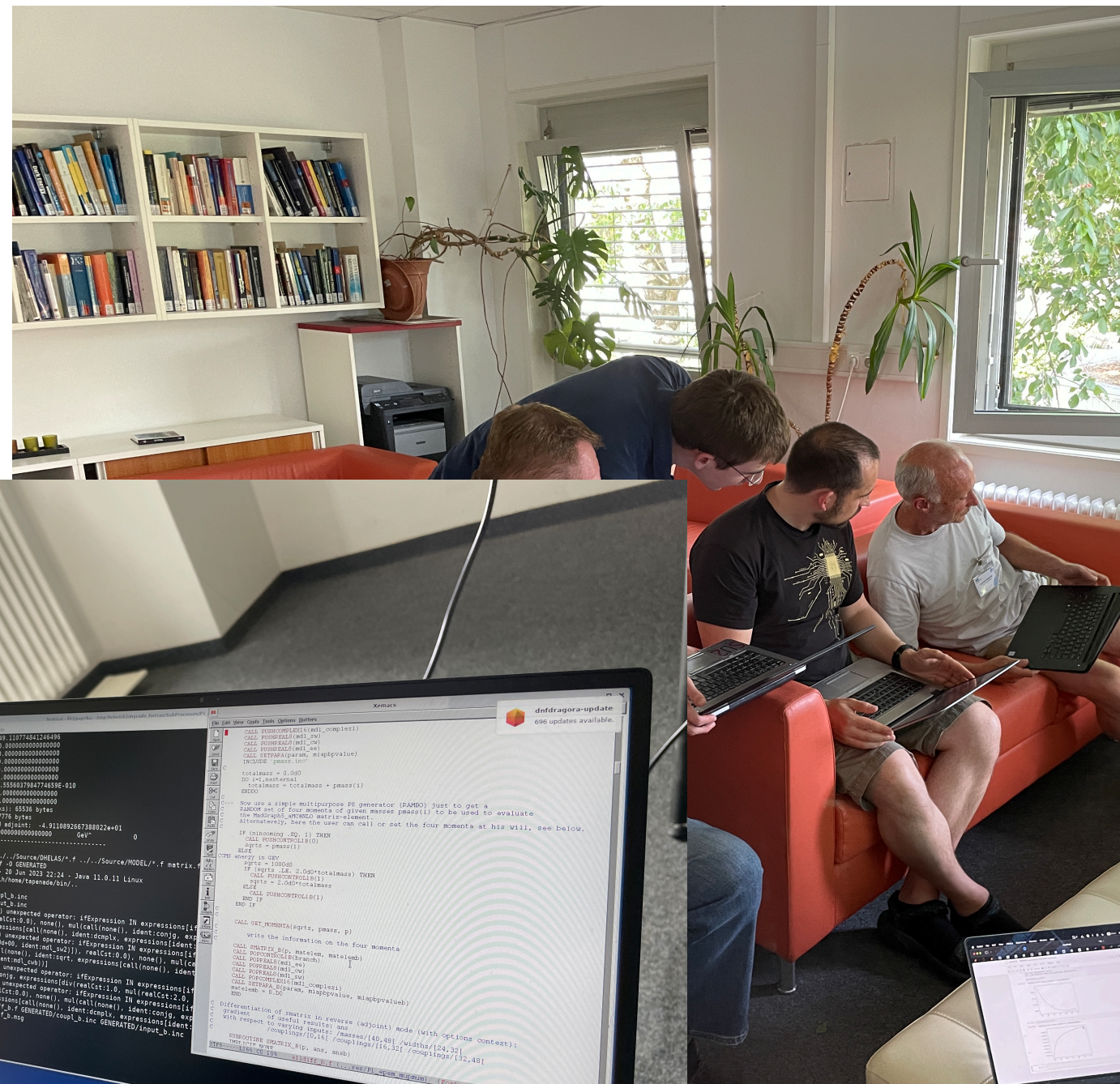


### 1. Introduction

The complexity continuously. Cu executed with the studies, one has hundreds of likelihood components, each representing a different measurement channel. For the

arXiv:2304.02650v1 [cs.MS] 4 Apr 2023

Munich 2023:  
Differentiating  
FORTRAN  
Madgraph  
with TAPENADE

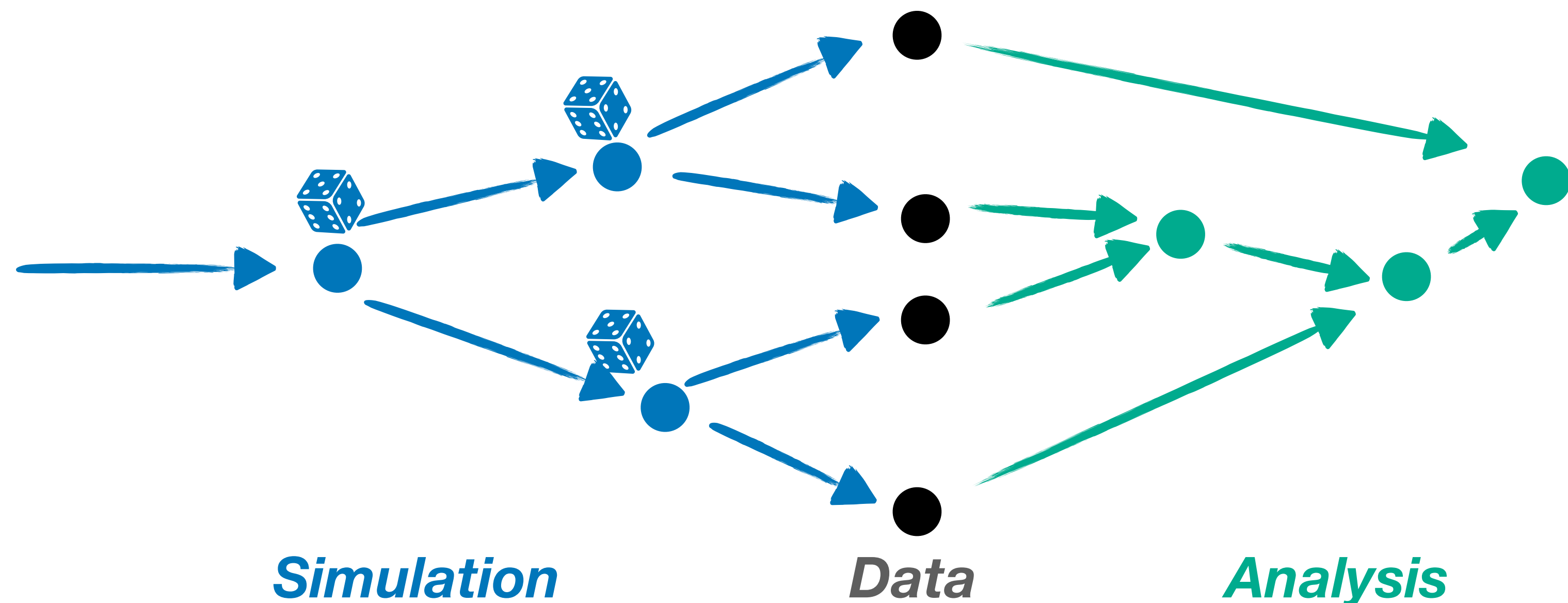




# Challenges

One of the key challenges for differentiable physics in HEP is that we have **non-differentiable operations at the core**

To zero-th order: HEP = splitting & clustering!  
→ a conceptual not a technical challenge.

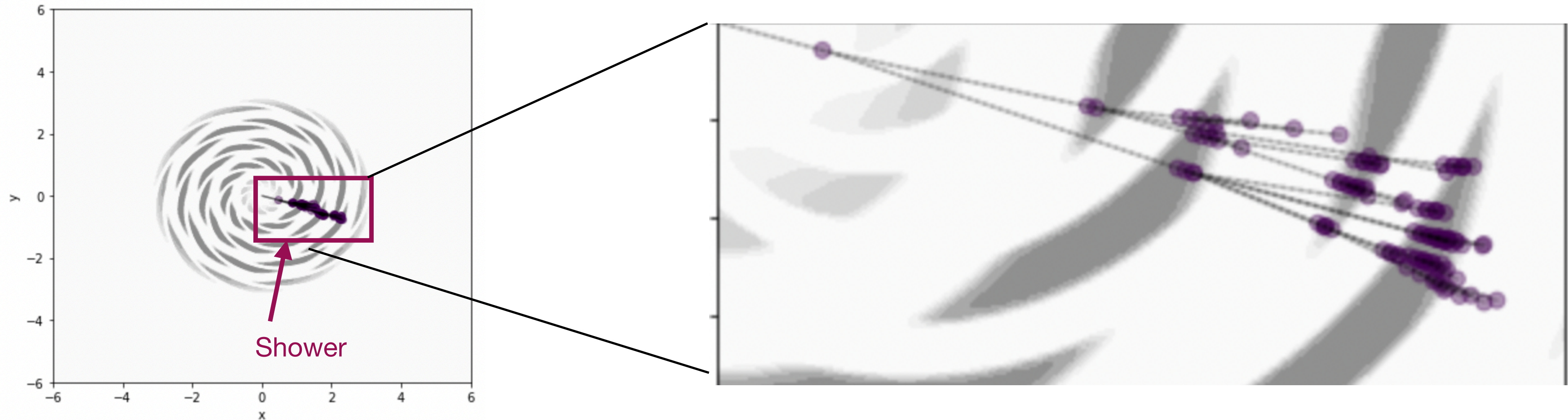


**Care Needed to  
become differentiable**



# Challenges

**With some thought, it's possible to differentiate even discrete processes e.g. as particle showers & event selections**

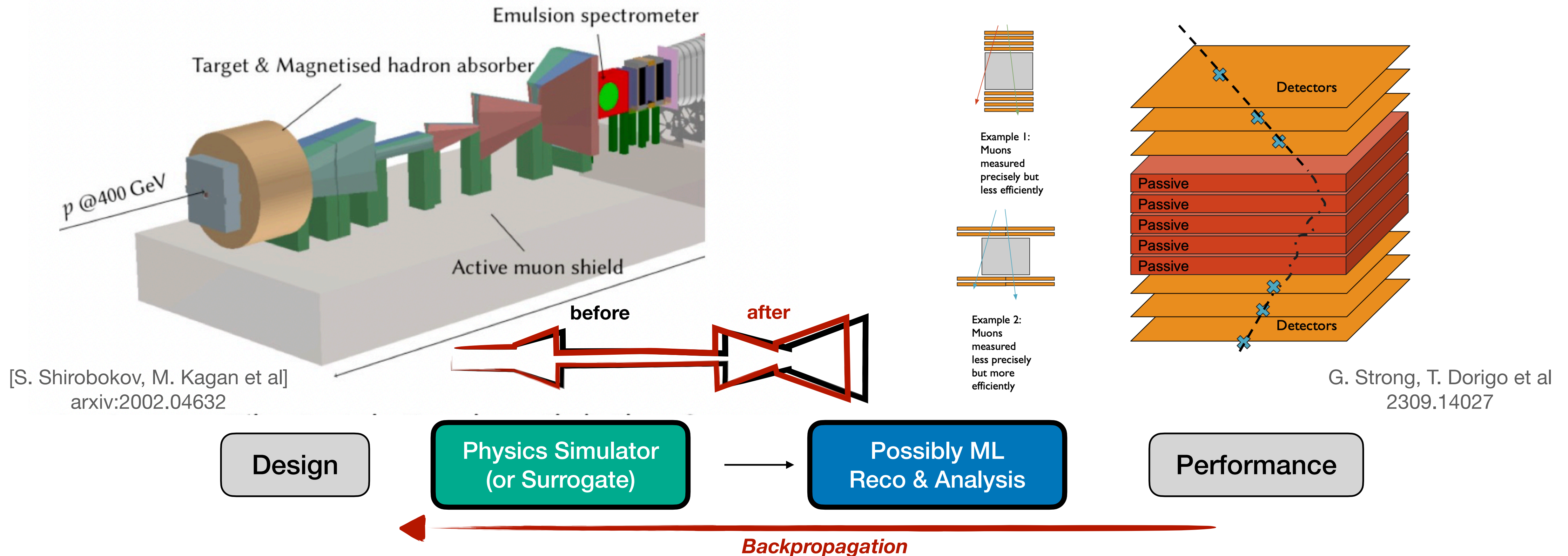


**Key: Adapt methods from e.g. Reinforcement Learning or newer “Stochastic Automatic Differentiation”**



# Application: Detector Design

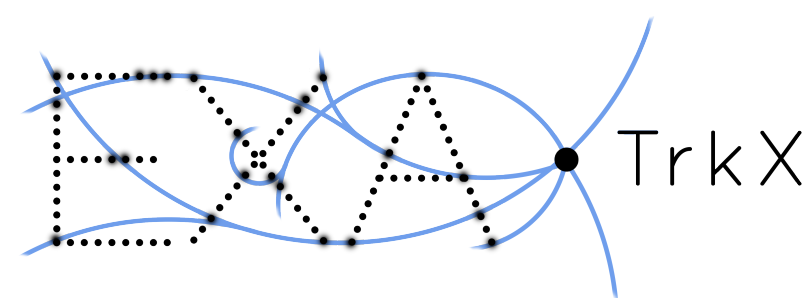
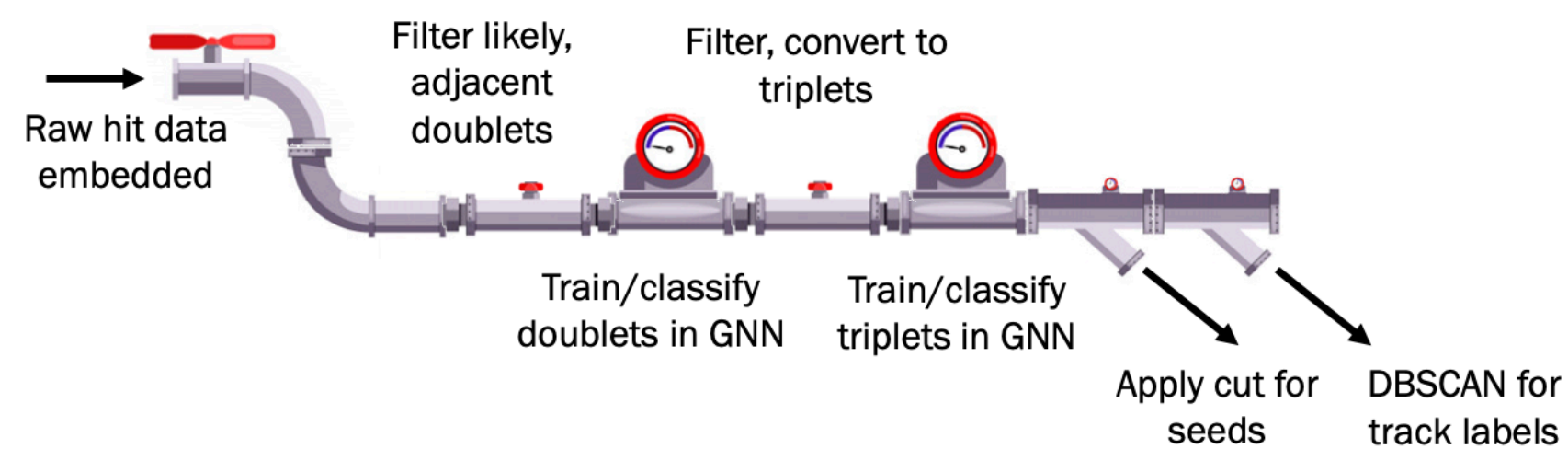
With a differentiable detector simulation or a neural network surrogate, we can optimize the detector design



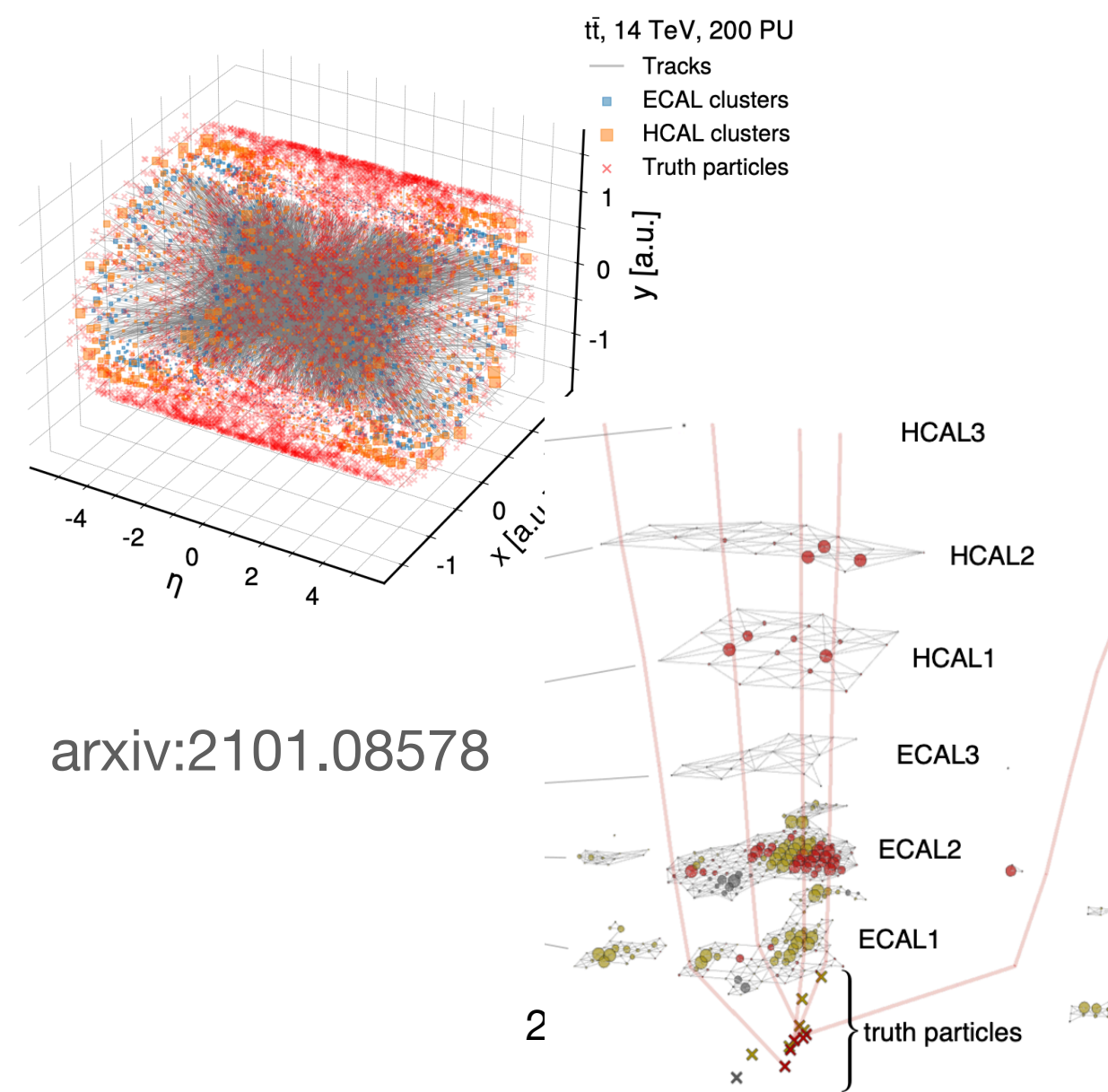


# Challenges

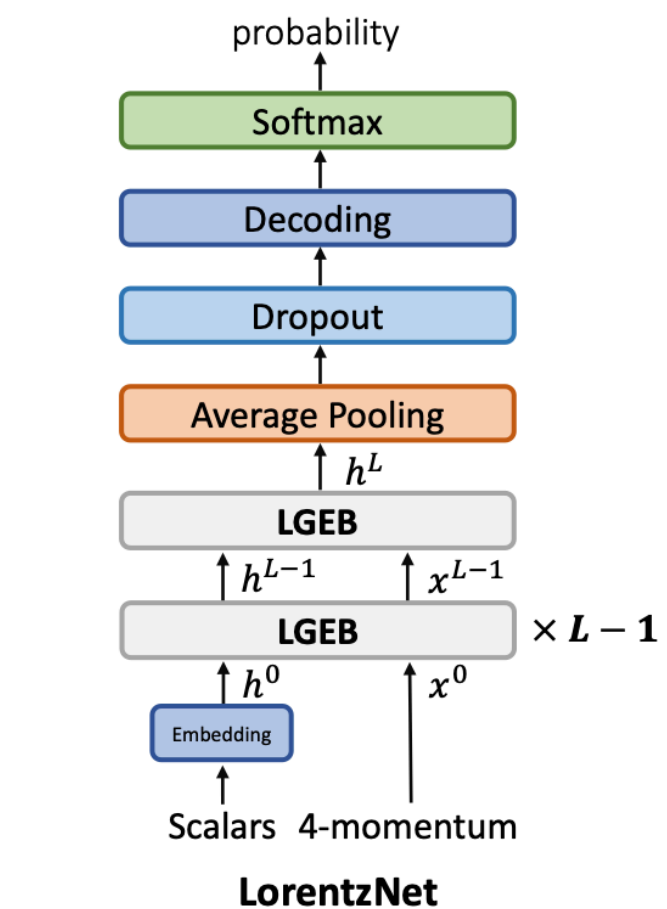
Gradients are not the only secret sauce of ML. **Over-parametrization is important.** Hybrid Systems likely work best with fairly big neural components connected by some physics



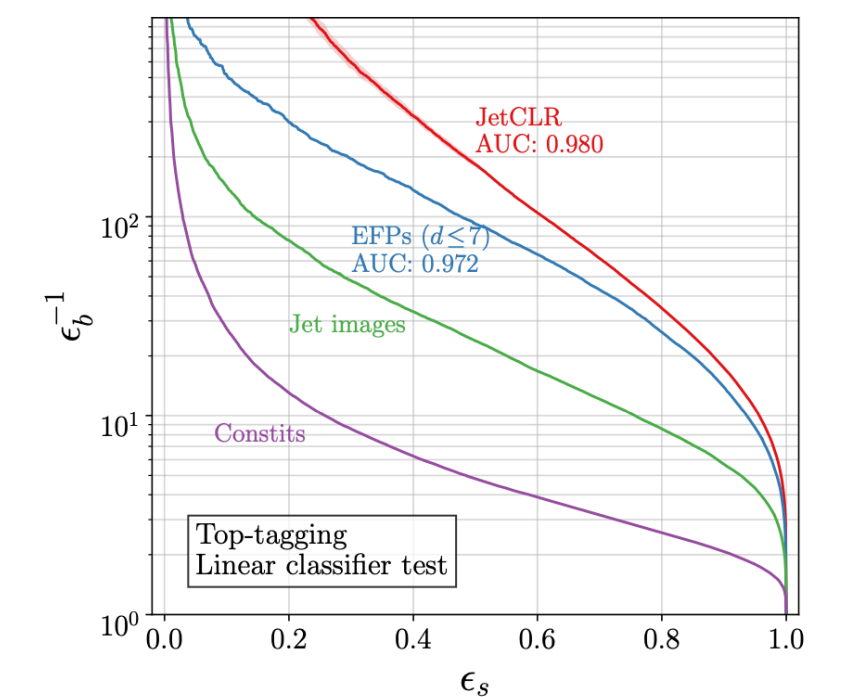
ML Tracking  
(Graphs Neural Nets)



ML Particle Flow  
(Graphs & Transformers)



arxiv: 2201.08187



arxiv: 2108.04253

Jet Representations  
(Transformers)

# Looking Forward

**Differentiable Programming** we have a tool that allows us to inject physics into the data-flow. A more nuanced picture re: role of physics in AI models. Worth investing in R&D.

Where will we land?

