Differentiable Programming in HEP

ACAT 2024

Lukas Heinrich, TUM

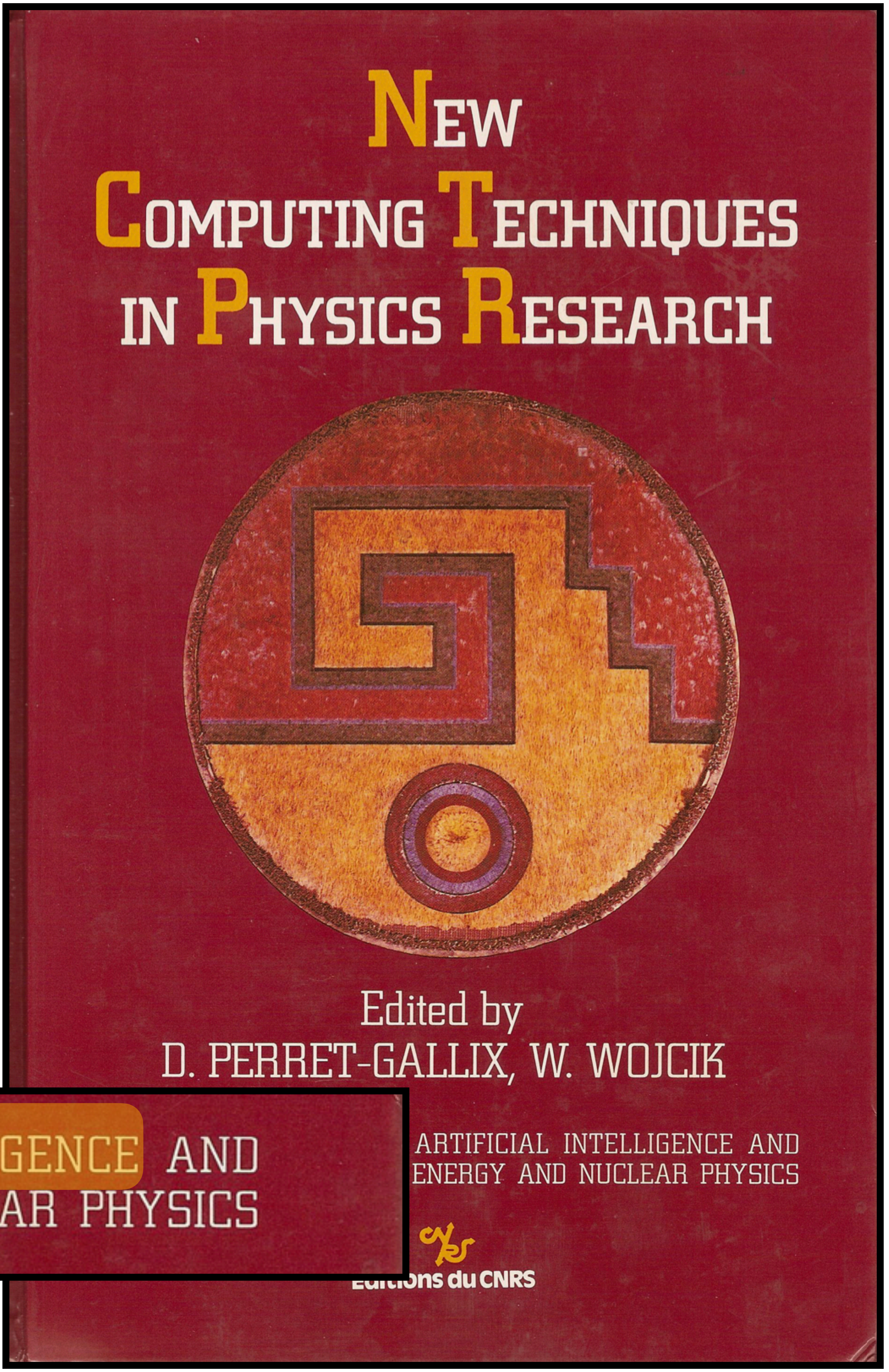
Technische Universität München

ODSL ORIGINS Data Science Lab

Big Picture Questions

Development of AI methods for Particle Physics

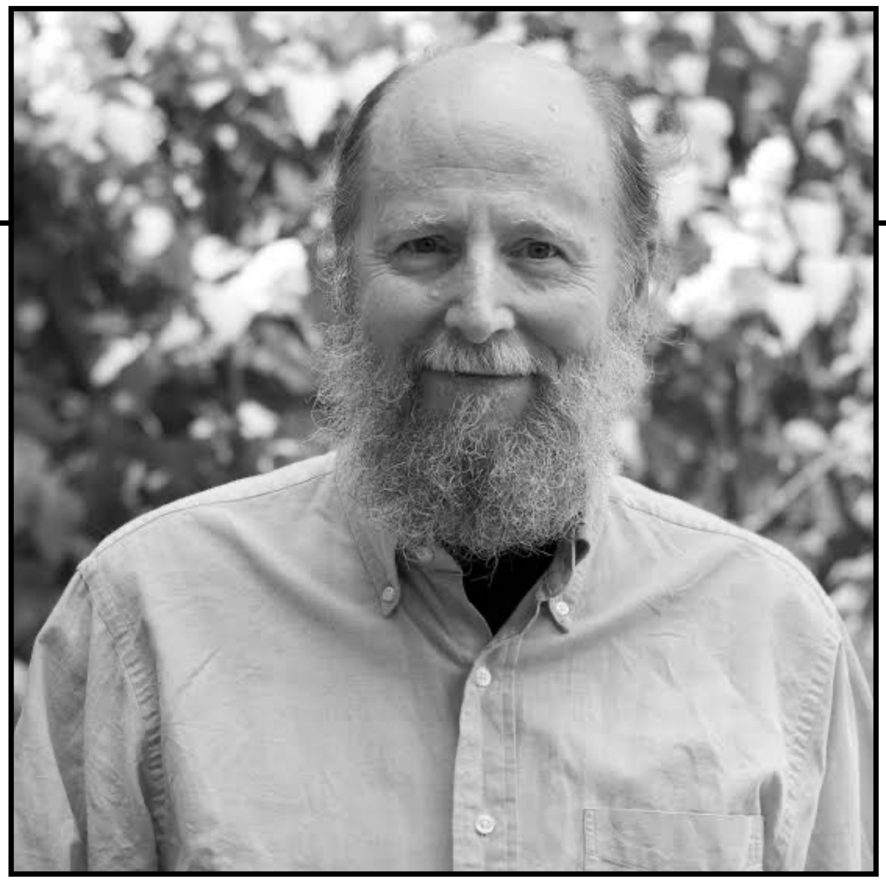
Does this mean we should put Physics Knowledge into AI ?



AINHEP 1990

The Bitter Lesson

Rich Sutton



March 13, 2019

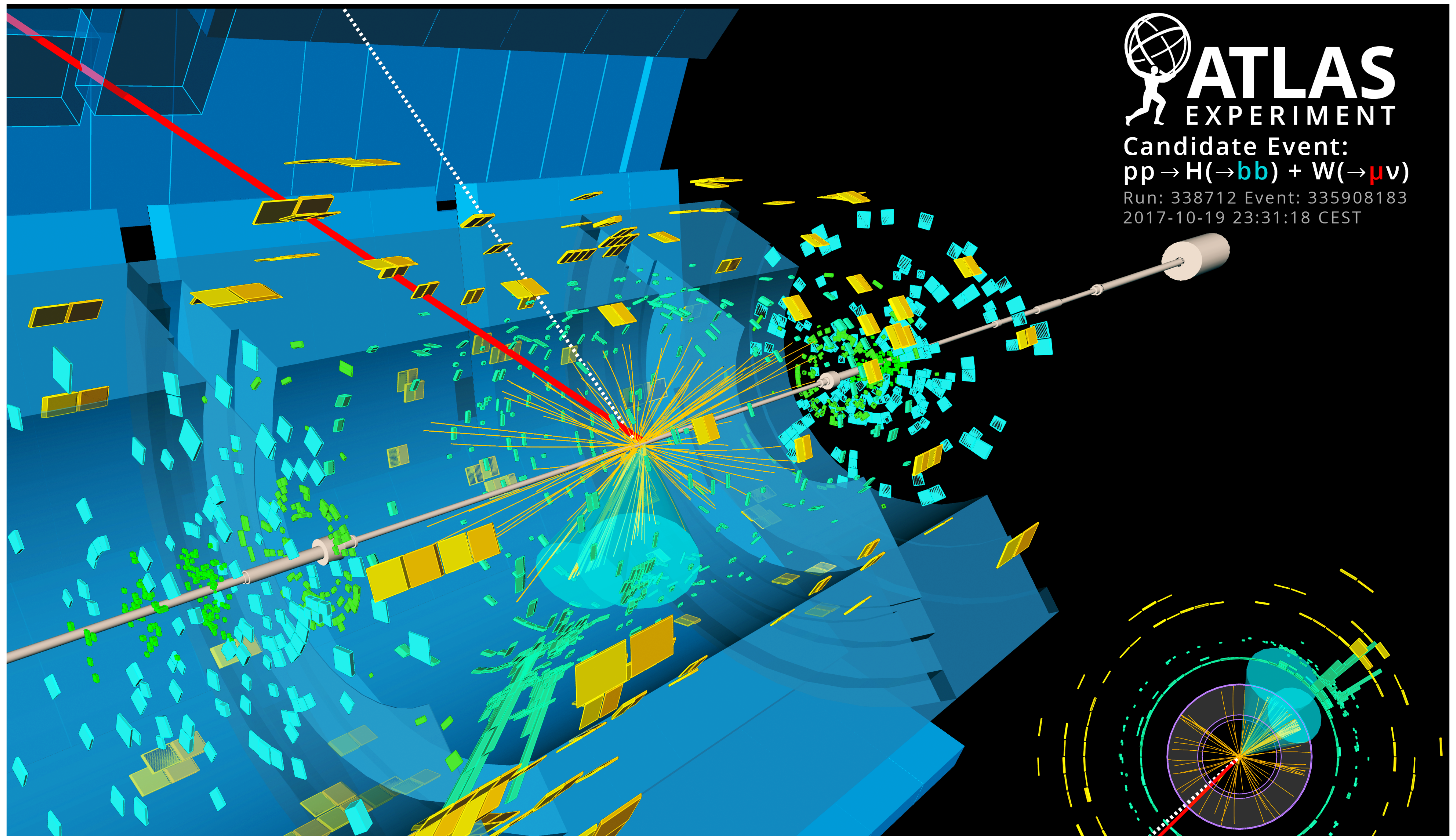
Researchers seek to leverage their human knowledge […],
but the only thing that matters in the long run
is the leveraging of computation

The biggest le ... methods that leverage computation are ultimately the
most effective ... v, or rather its generalization of continued
exponentially ... nducted as if the computation available to the agent
were constan ... 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 cha ... eep search. At the
time, this was looked upon with dismay by the majority of c ... ds that leveraged
human understanding of the special structure of chess. Whe ... ware and software
proved vastly more effective, these human-knowledge-base ... hat ``brute force''
search may have won this time, but it was not a general stra ... s. These researchers
wanted methods based on human input to win and were disa...

… many examples of AI researchers' belated
learning of this bitter lesson

Is this what we want ?



Transformer or whatever comes next.

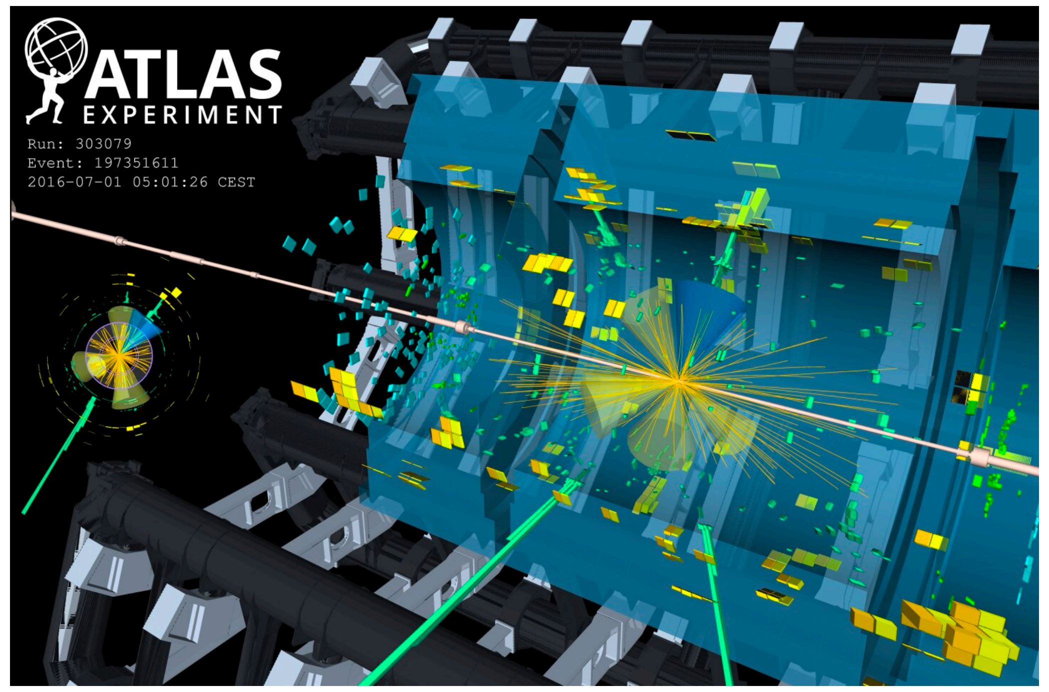
AI: Trust me,
This is a Higgs Decay

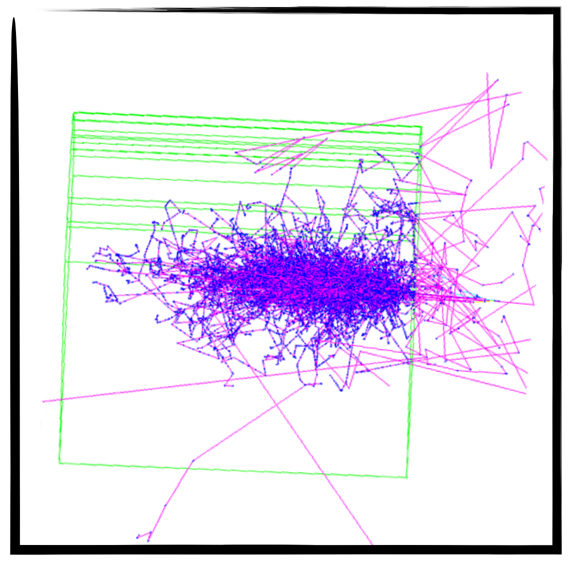
Human: Ok.

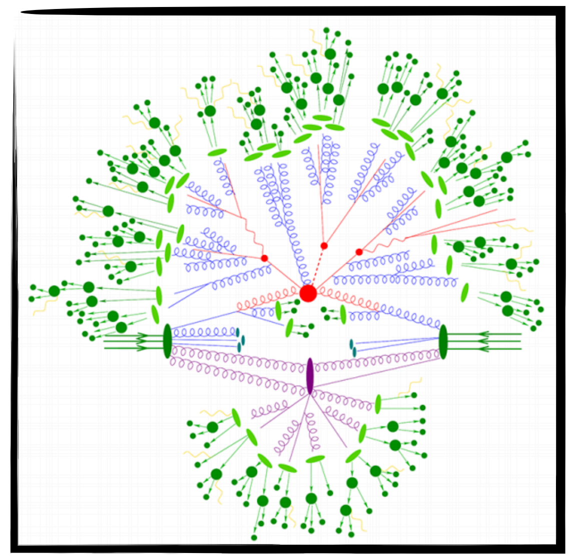
A end to end “Hits to Higgs” system seems neither feasible (today) nor particularly desirable. As scientists we usually want more.

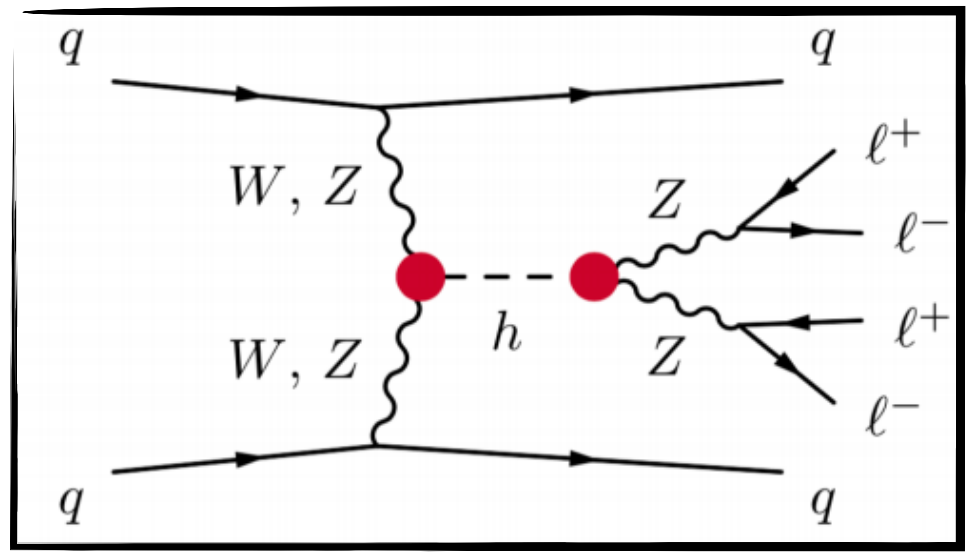
Probably not.

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.***



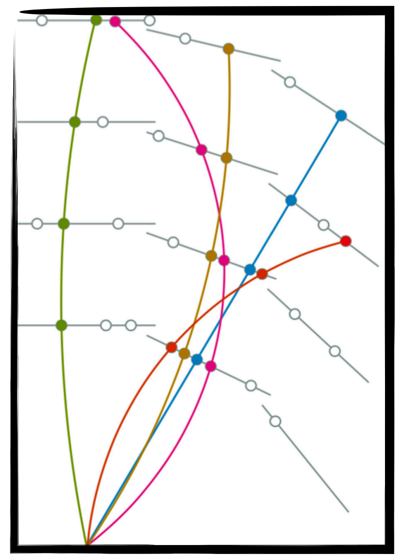


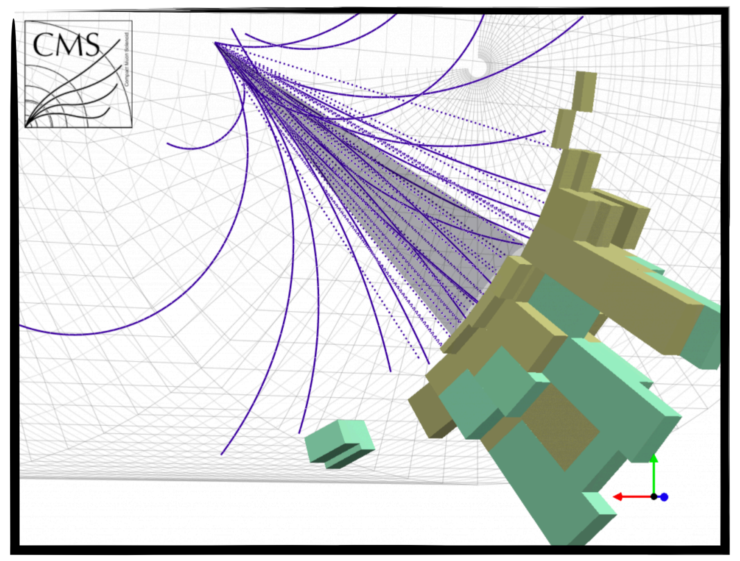


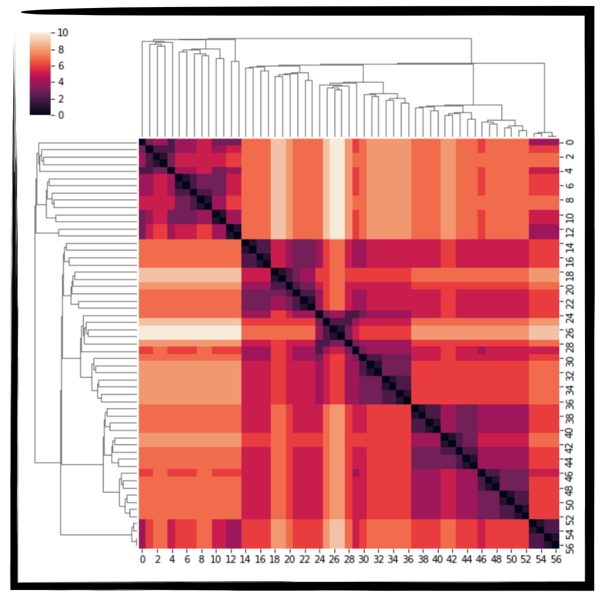


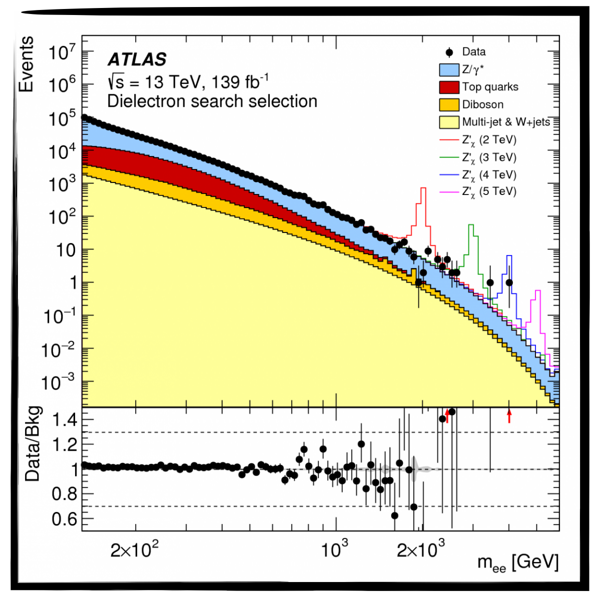
$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu}$$
$$+ i\bar{\psi} \not{D} \psi + h.c.$$
$$+ \bar{\psi}_i y_{ij} \psi_j \phi + h.c.$$
$$+ |D_\mu \phi|^2 - V(\phi)$$

Data









Theory

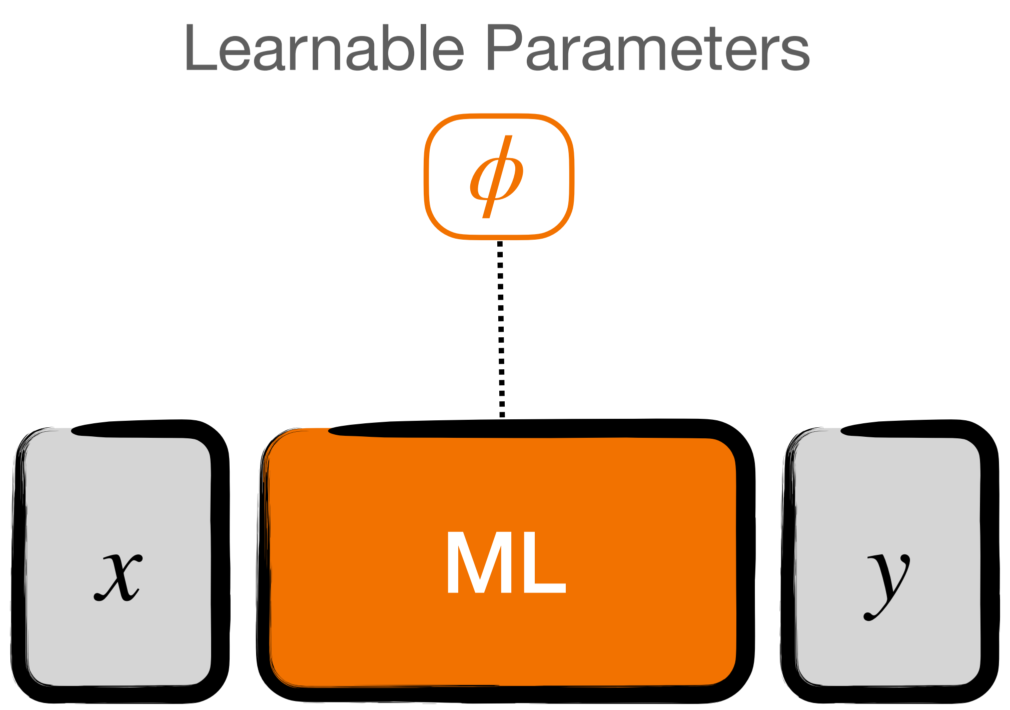
What do we want?

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

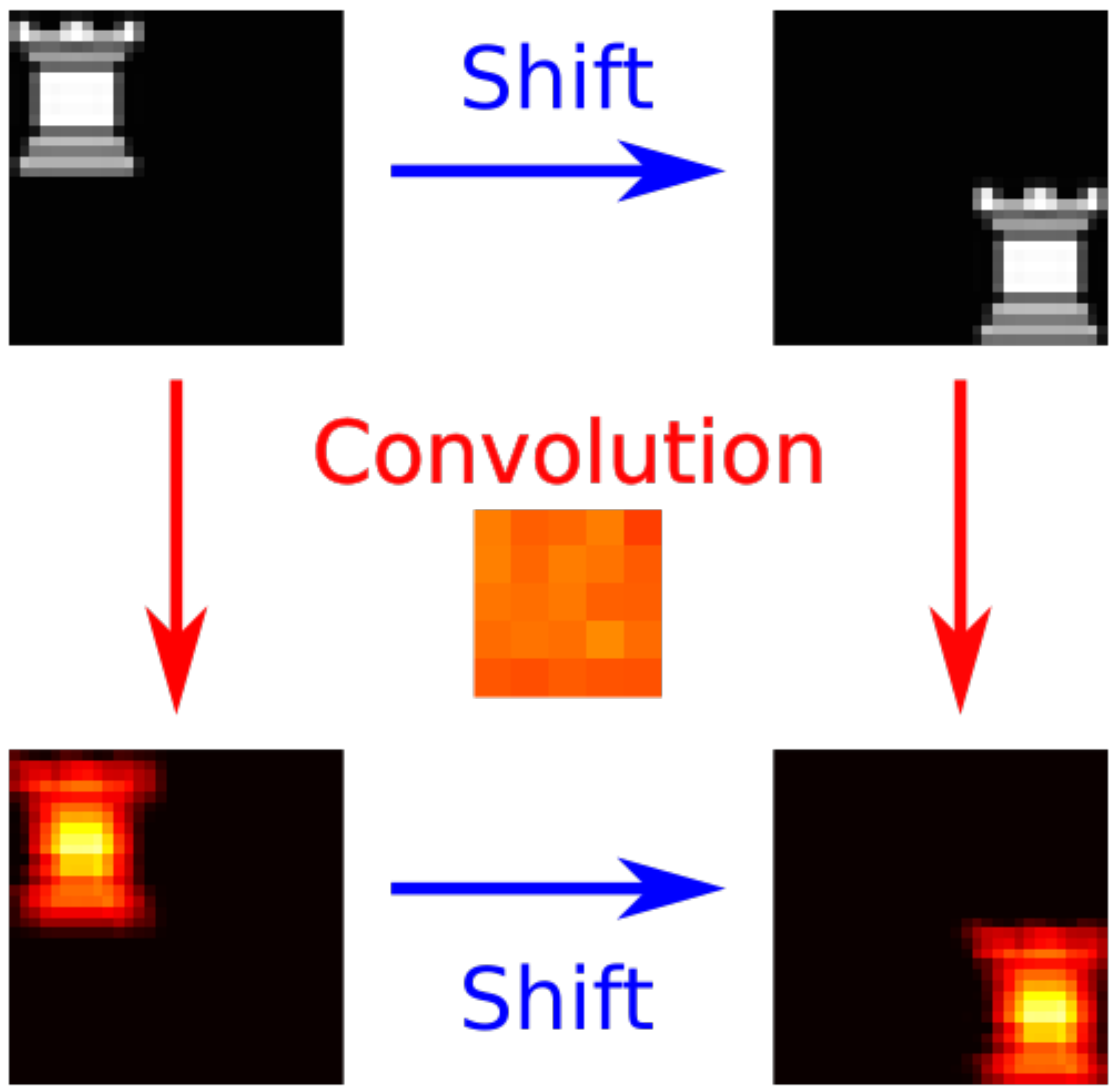
Need framework to build hybrid Physics-AI systems, learnable but with flexible control where & how much physics to put in.

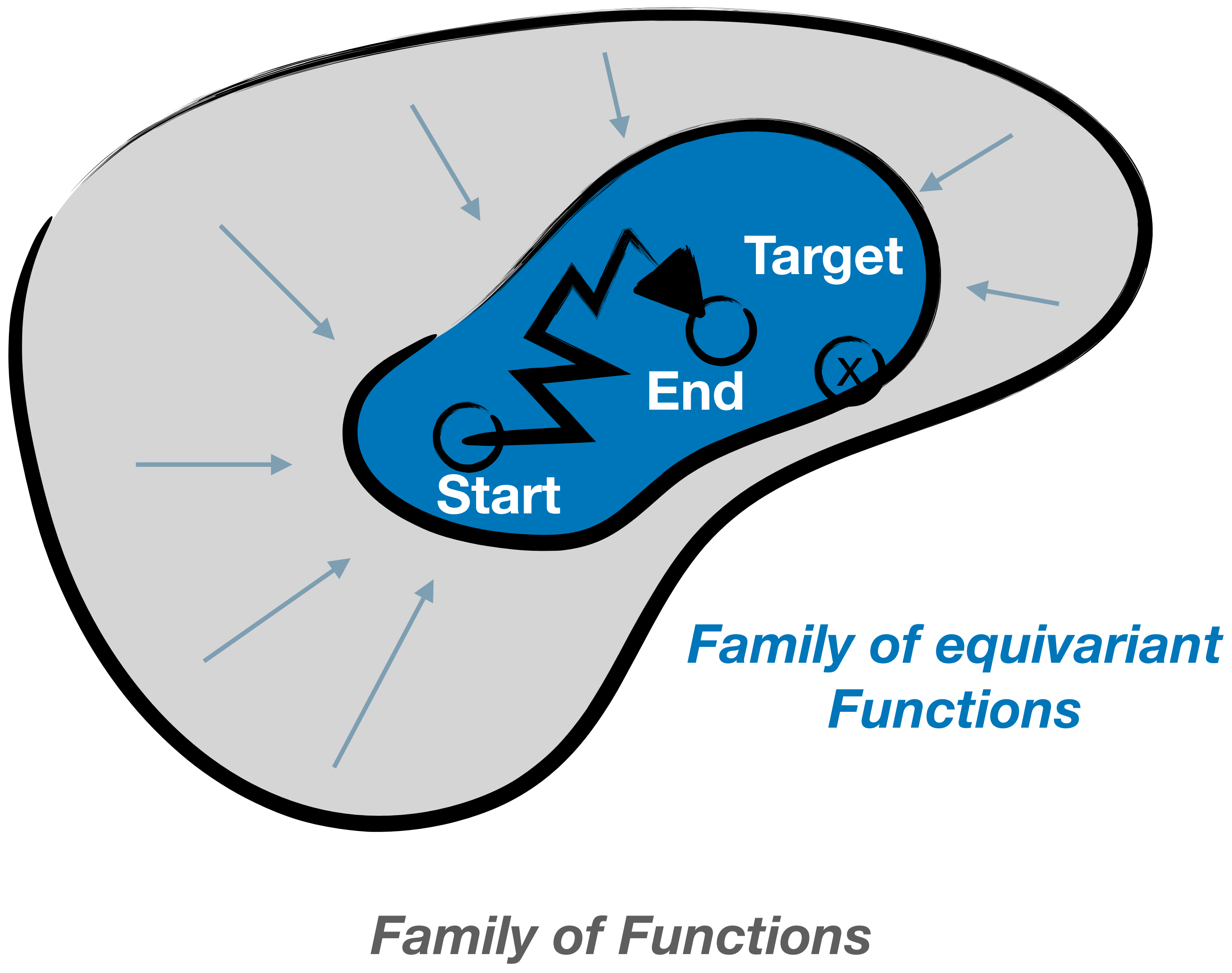
Architectures and the Inductive Bias Story

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

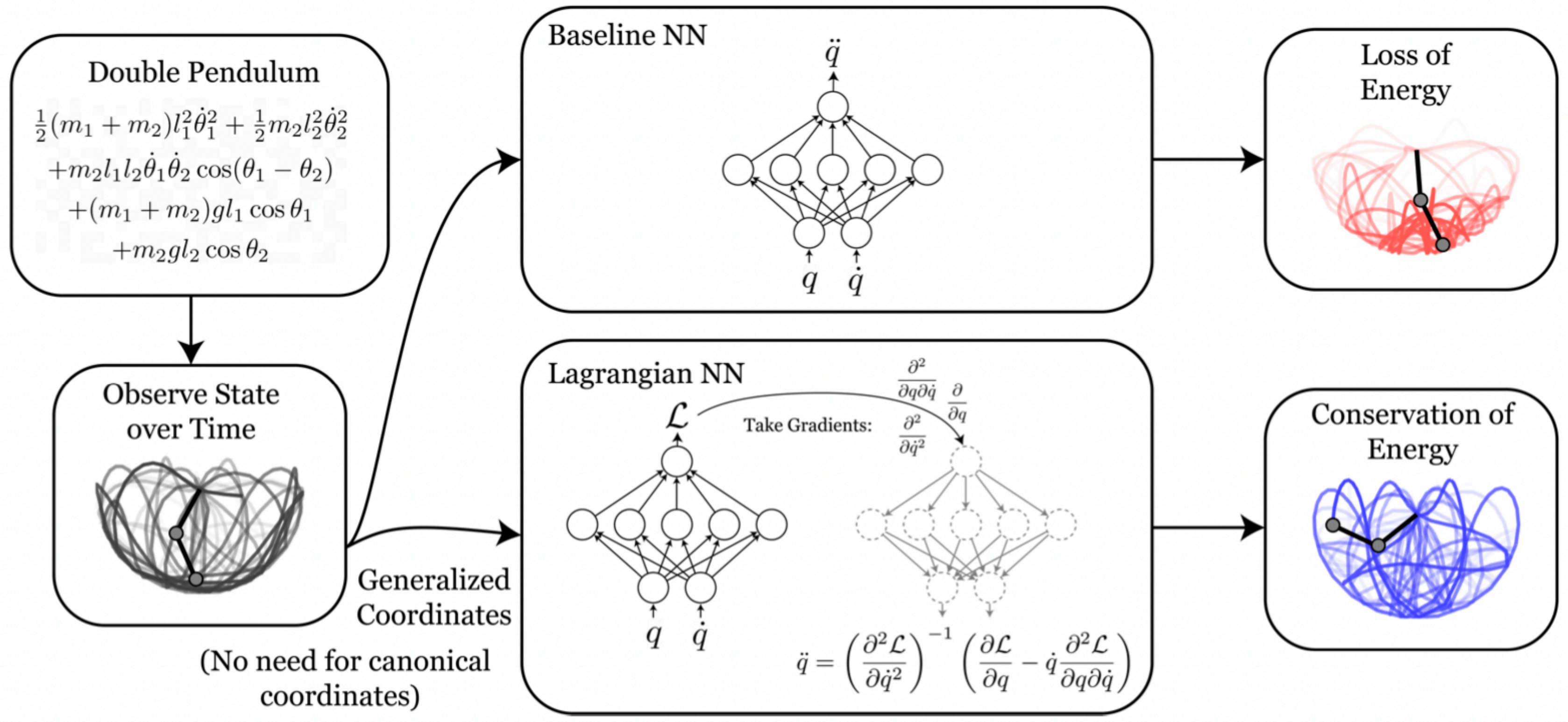


$$f_\phi(r_x(g)x) = r_y(g)f_\phi(x)$$

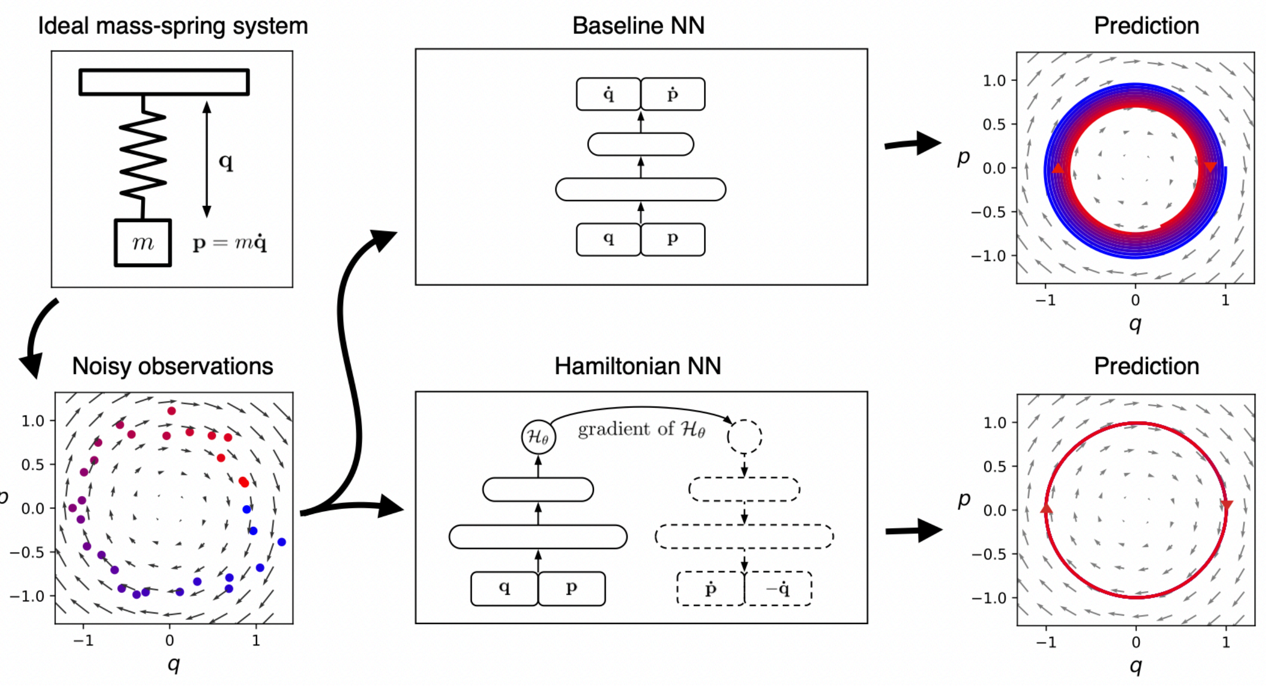




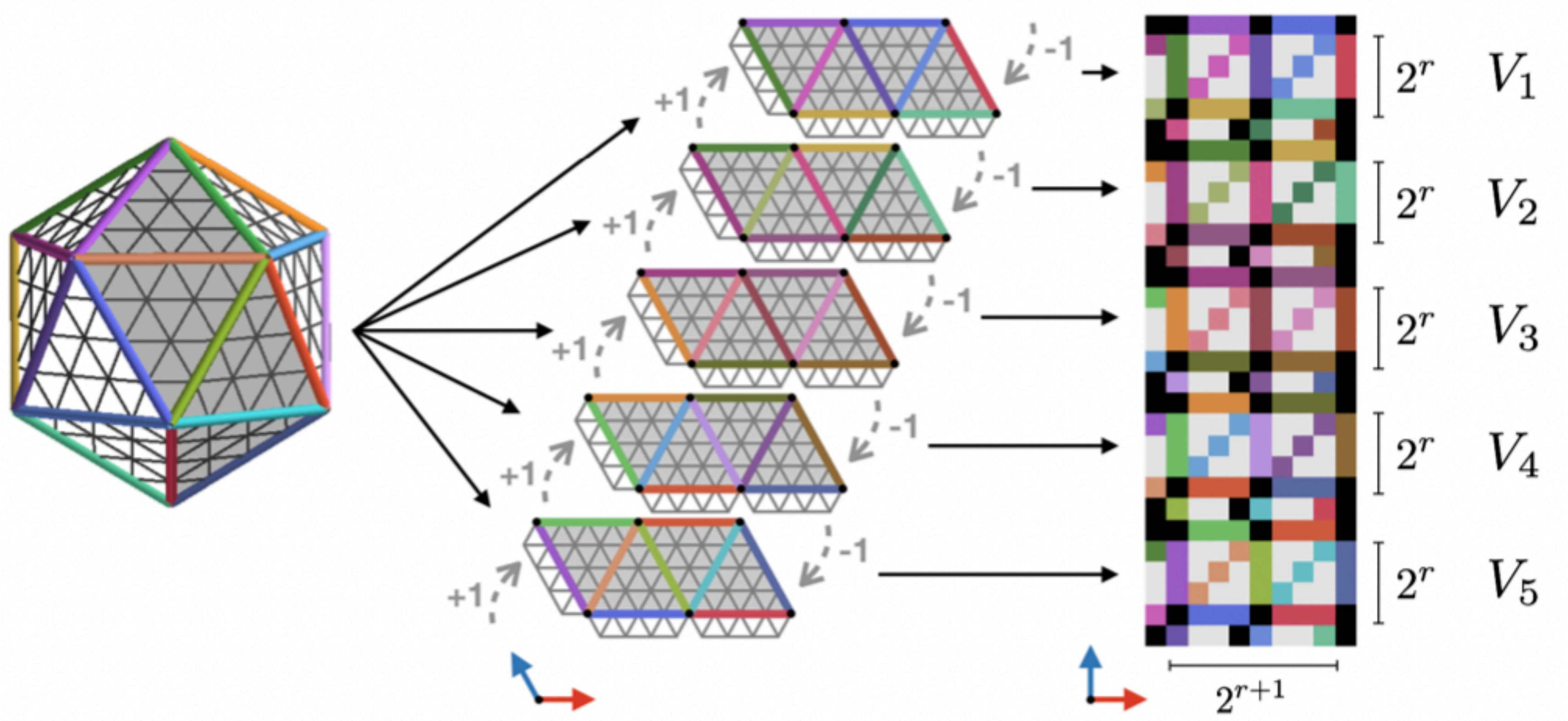
Can’t keep a Physicist from Symmetries



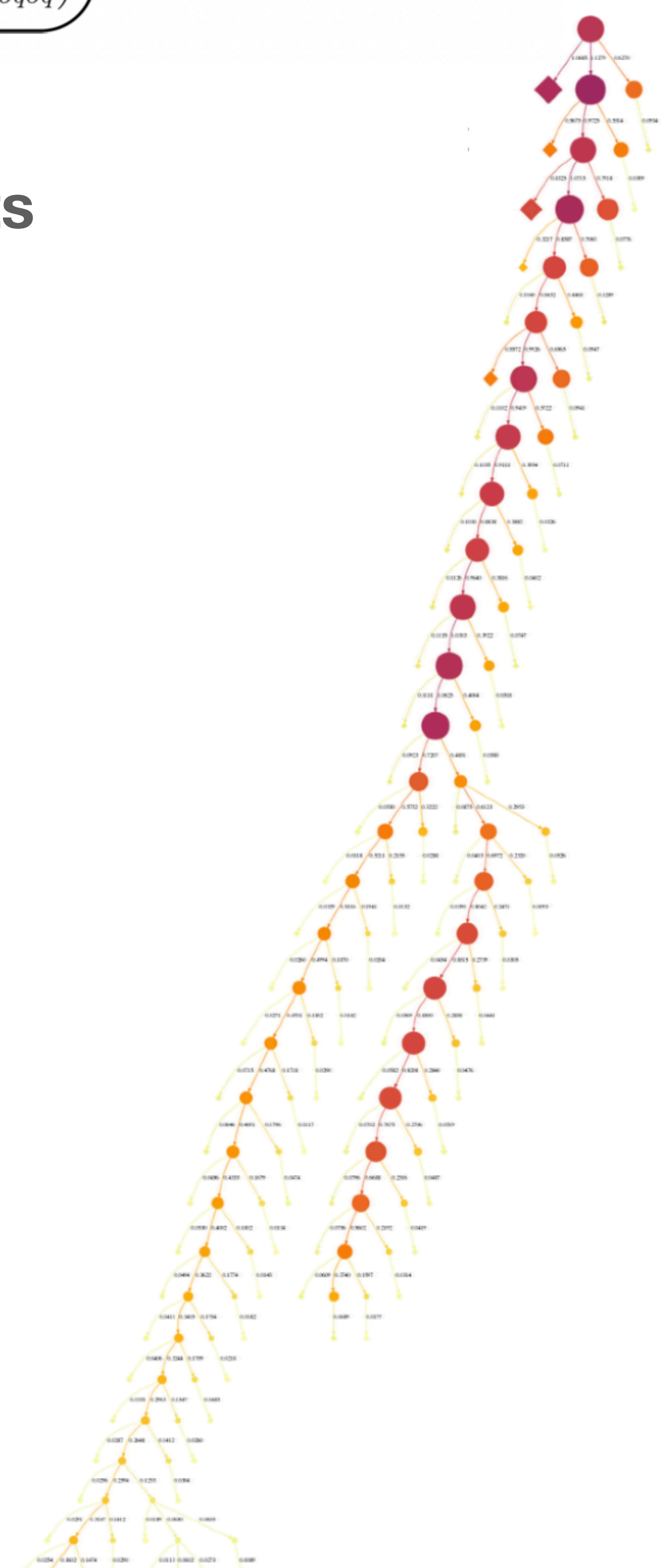
Lagrangian Neural Nets
arXiv: 2003.04630



Hamiltonian Neural Nets
arXiv:1906.01563

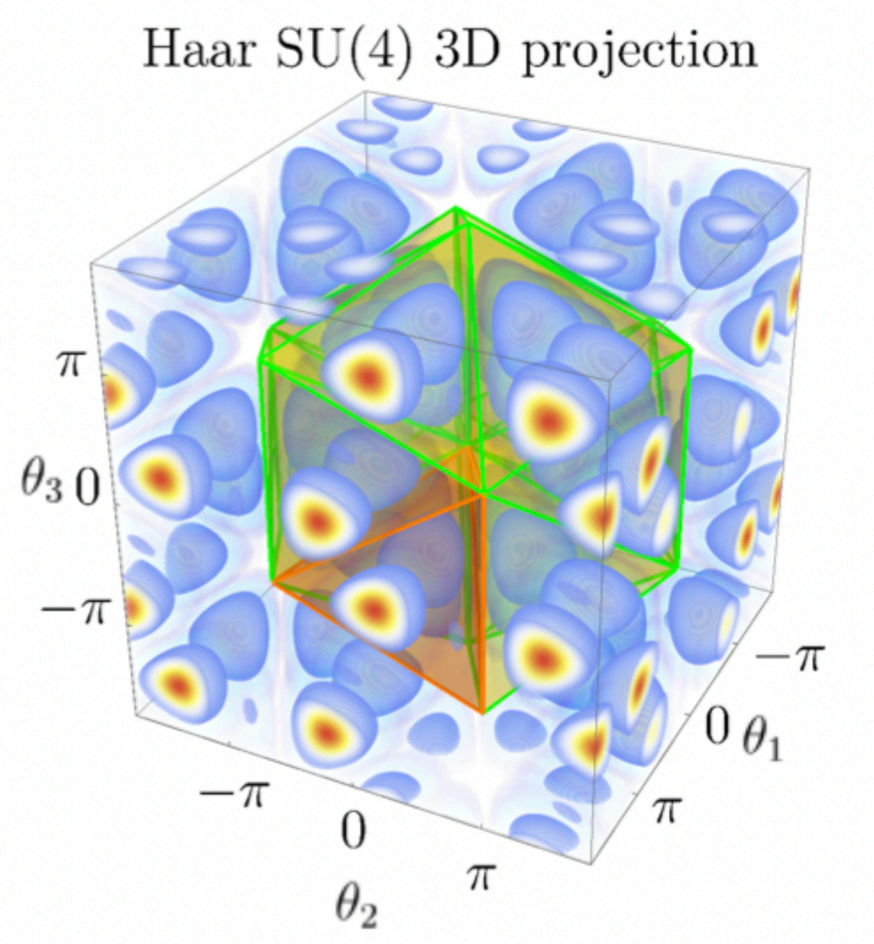


Gauge-Equivariant Convolutional Neural Networks

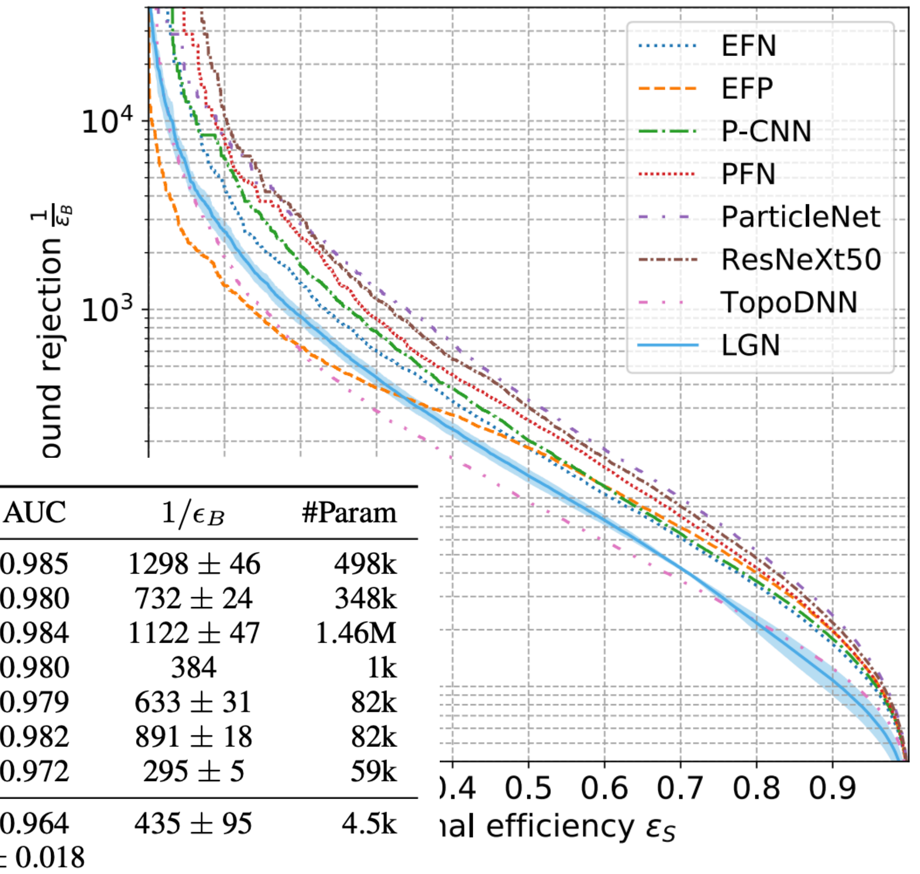


Neural Nets with QCD-like Structure

arXiv:1702.00748



SU(N)-Equivariant Normalizing Flows



Architecture	Accuracy	AUC	$1/\epsilon_B$	#Param
ParticleNet	0.938	0.985	1298 ± 46	498k
P-CNN	0.930	0.980	732 ± 24	348k
ResNeXt	0.936	0.984	1122 ± 47	1.46M
EFP	0.932	0.980	384	1k
EFN	0.927	0.979	633 ± 31	82k
PFN	0.932	0.982	891 ± 18	82k
TopoDNN	0.916	0.972	295 ± 5	59k
LGN	0.929 $\pm$ .001	0.964 $\pm$ 0.018	435 ± 95	4.5k

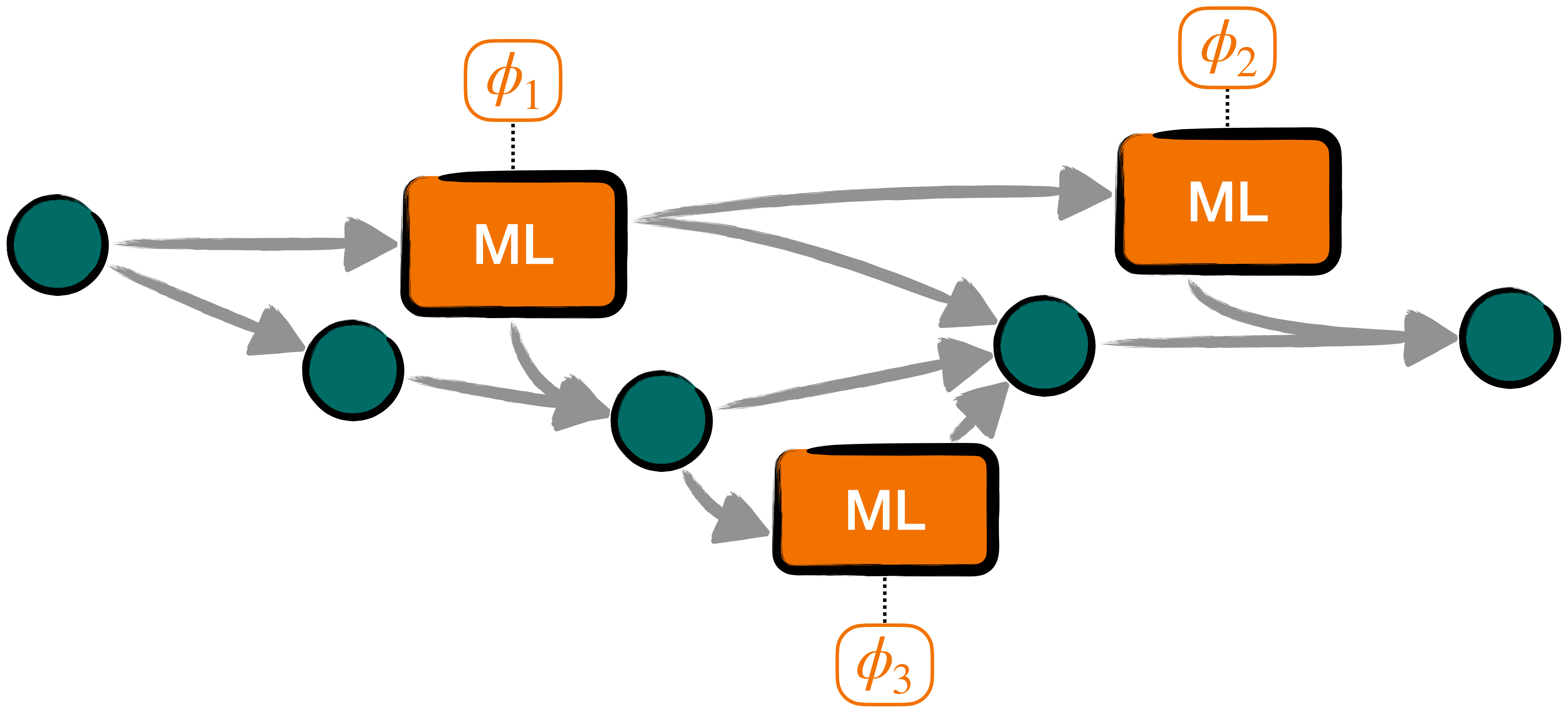
Lorentz-Invariance

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

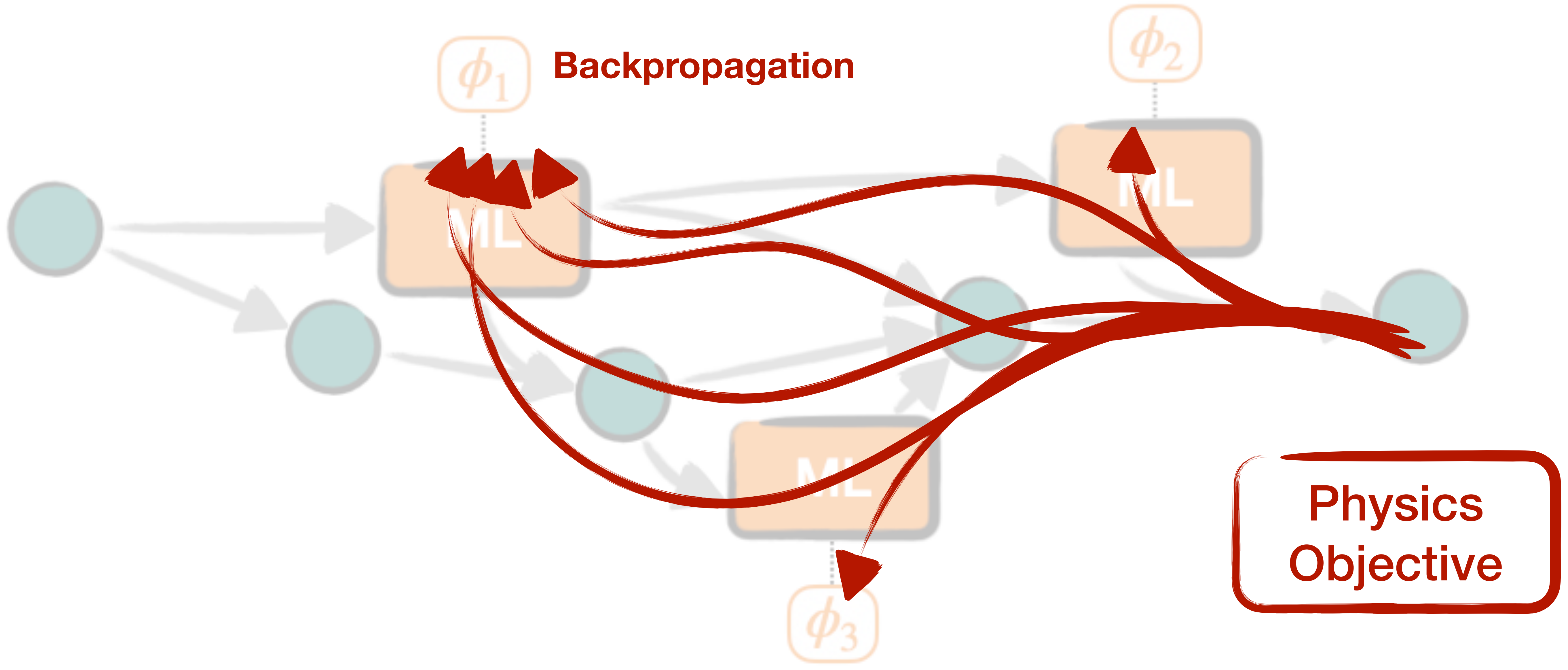
meaning of internal states **data flow between states**



Physics Computation **ML modules**

But there is more

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

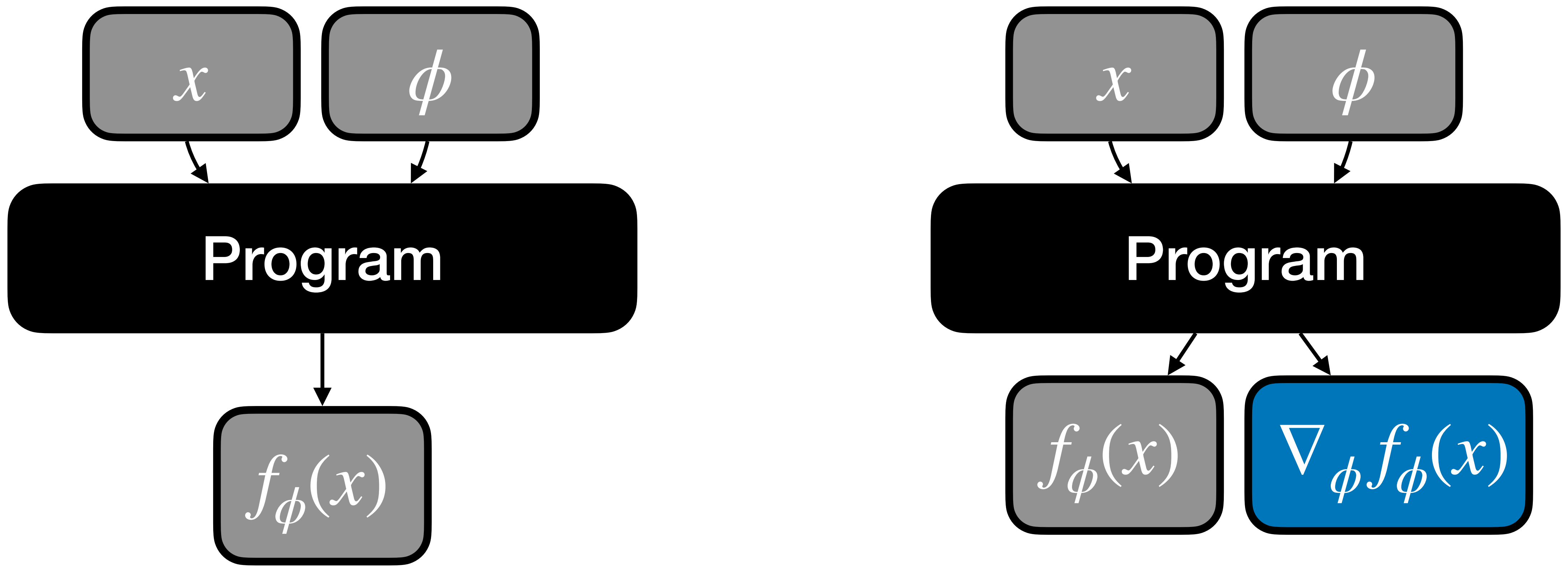


For this to work, the physics components must play nice gradient descent → differentiable programming

Differentiable Programming

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

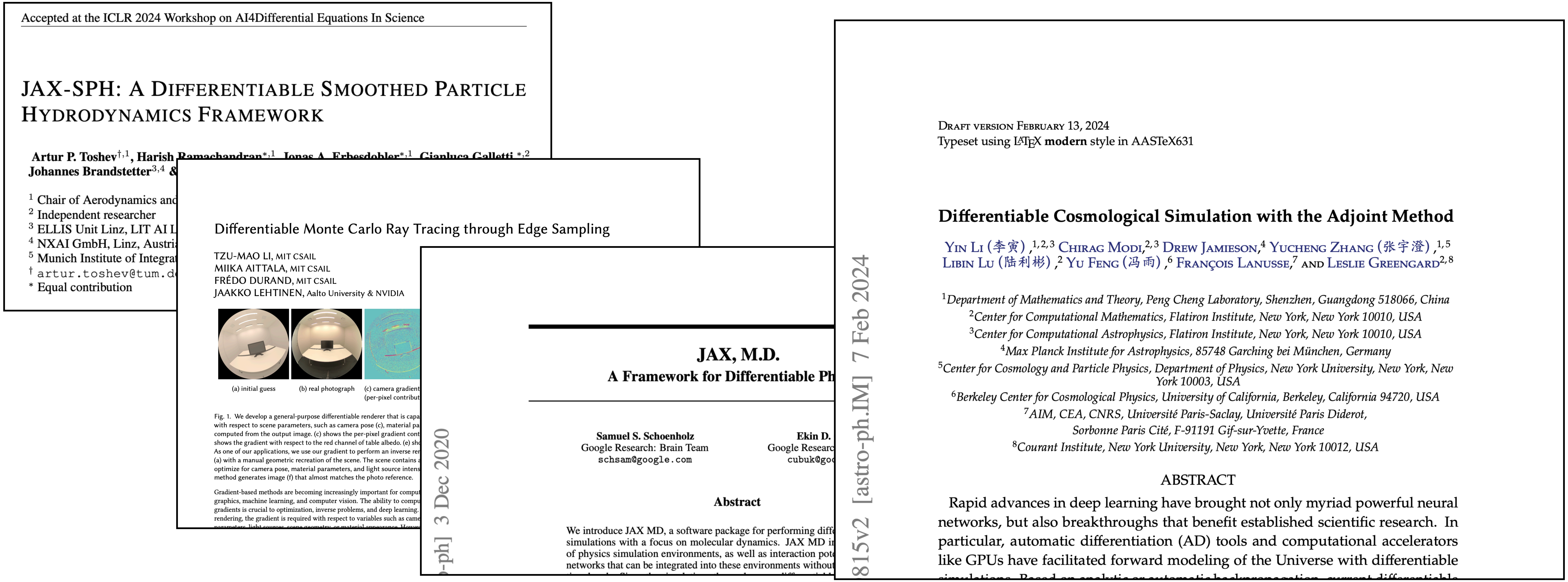
General purpose, ready for physics



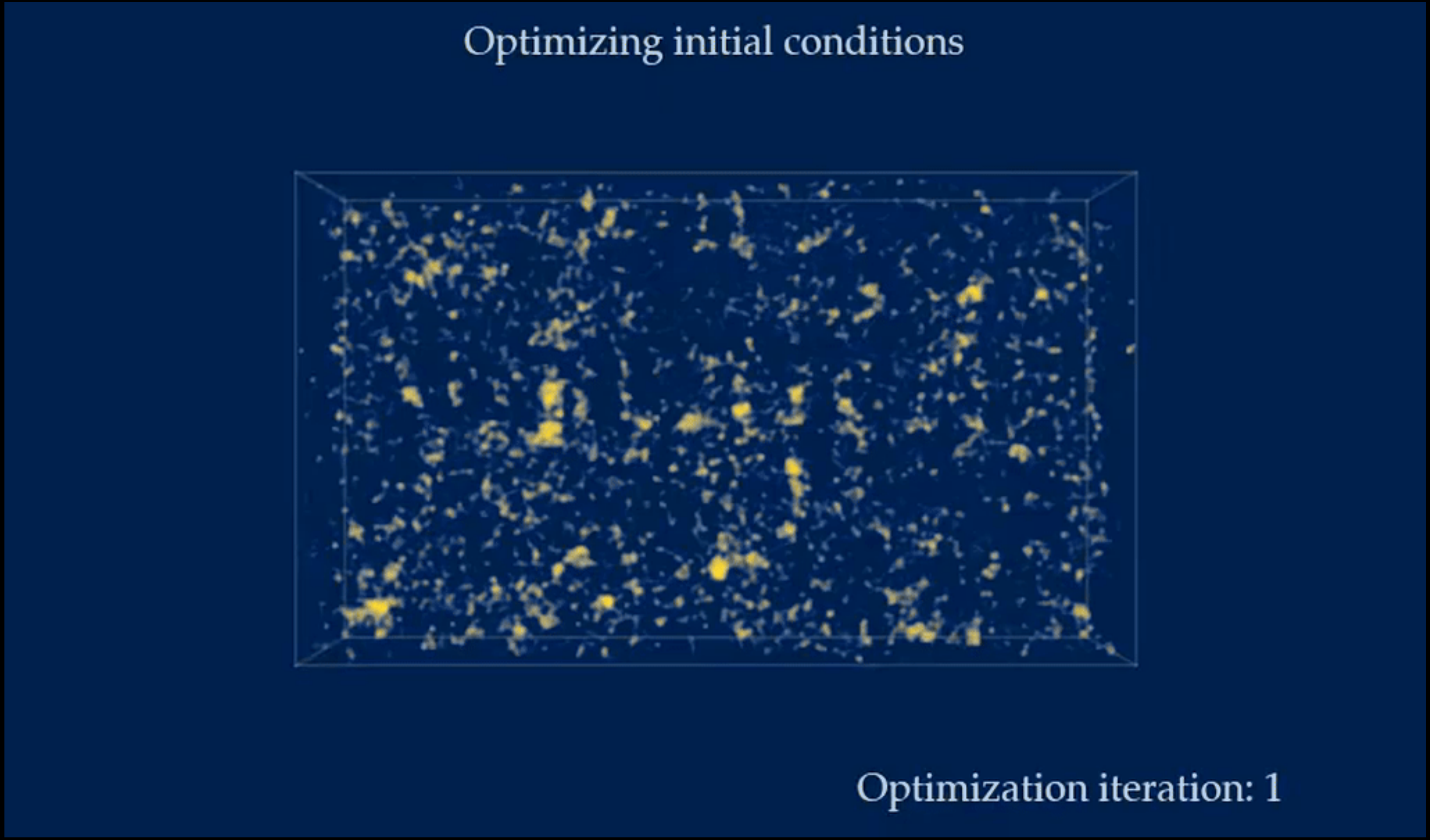
TensorFlow PYTORCH

From beyond HEP

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

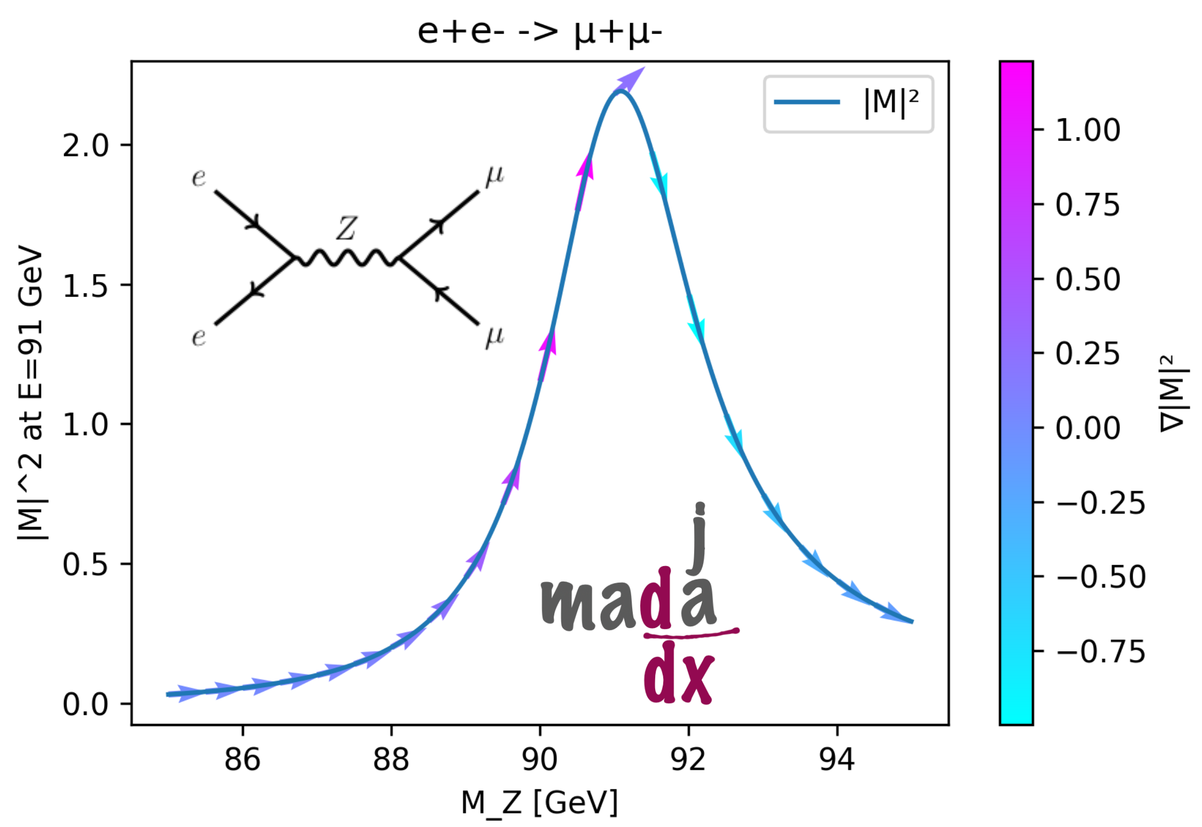


Optimizing a Differentiable Simulator



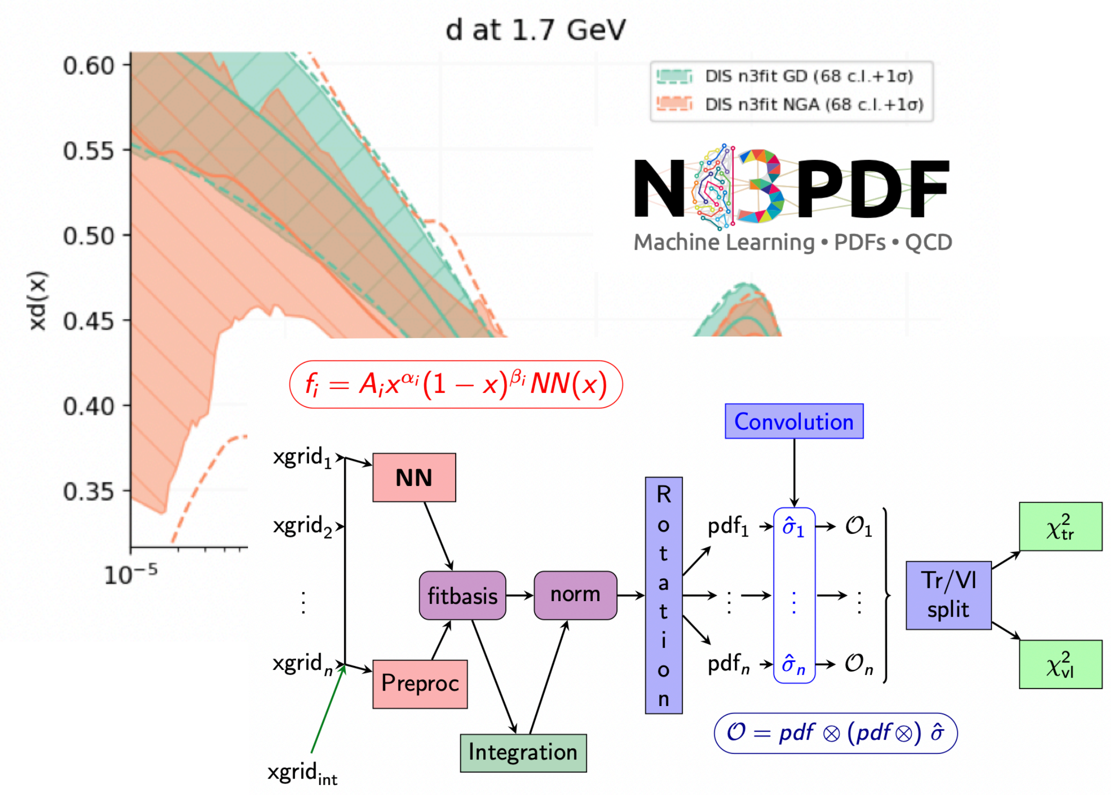
arXiv:2211.09815

In HEP & al, we are pushing as well



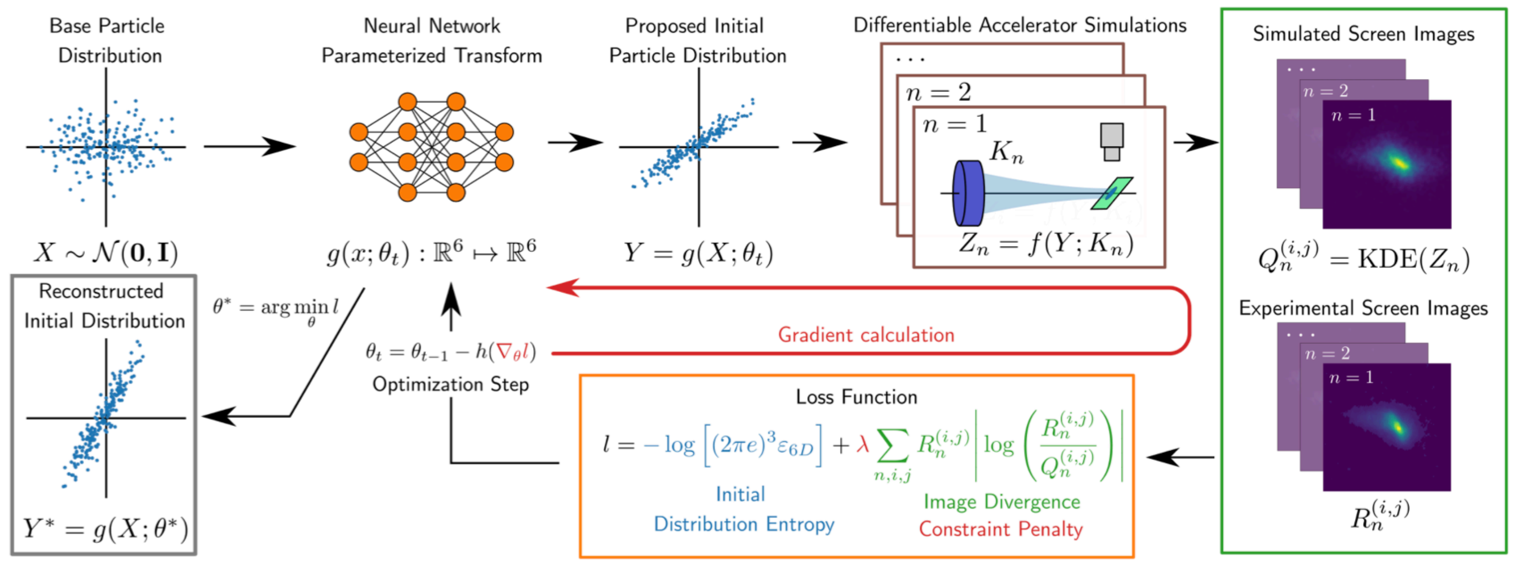
Differentiable Matrix Elements

2203.00057



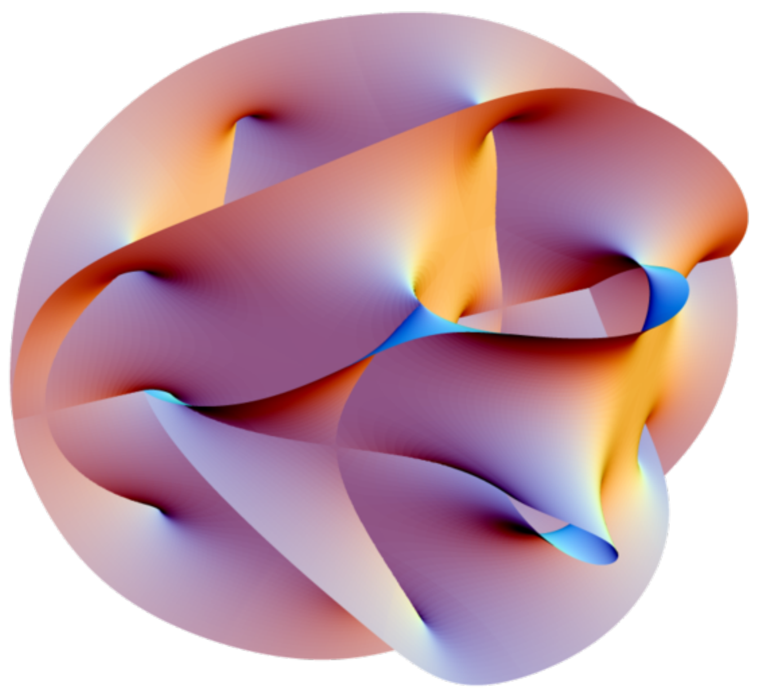
Differentiable Proton Structure

1907.05075



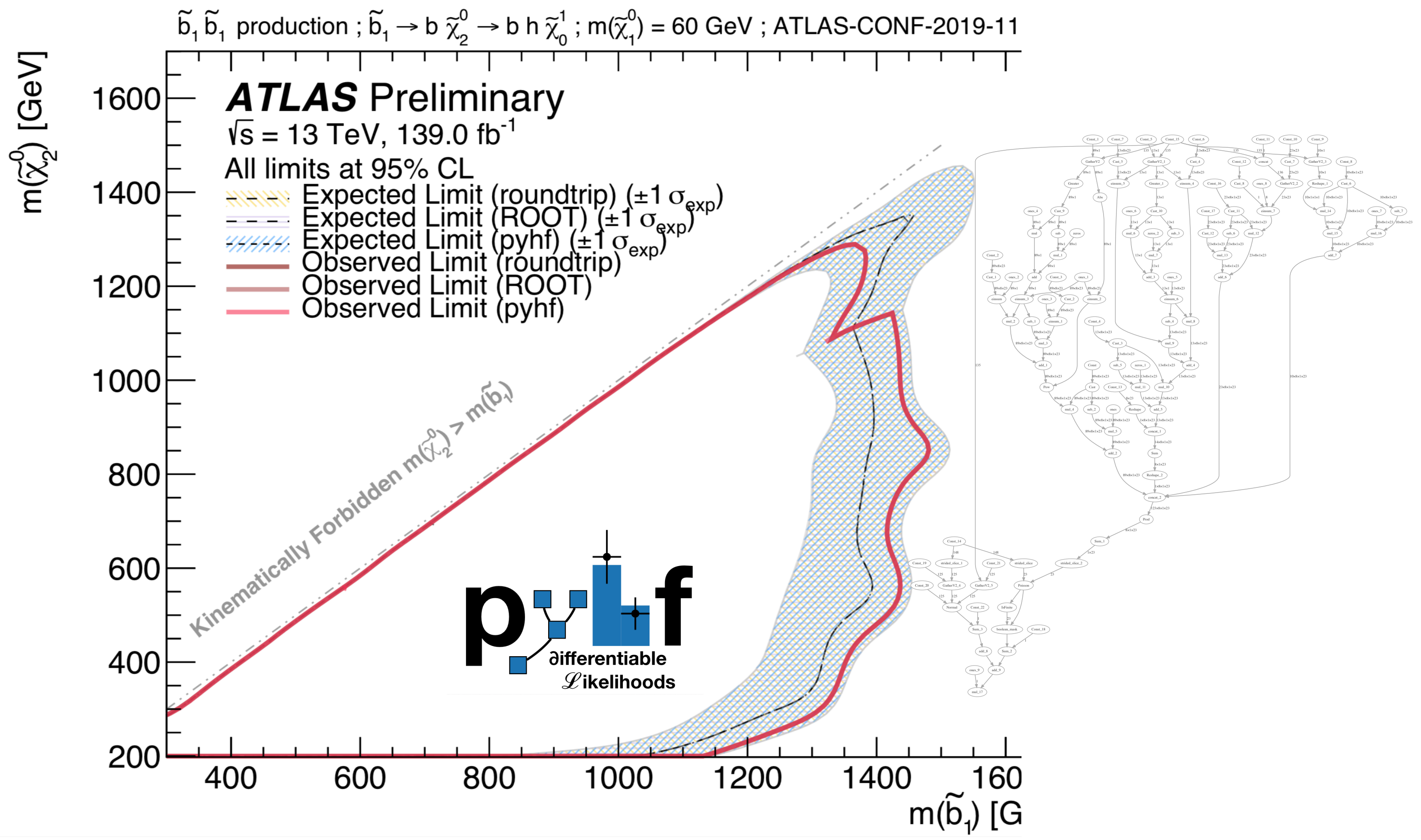
Differentiable Accelerator Simulation

2211.09077



Differentiable Calabi-Yau Manifolds

2211.12520

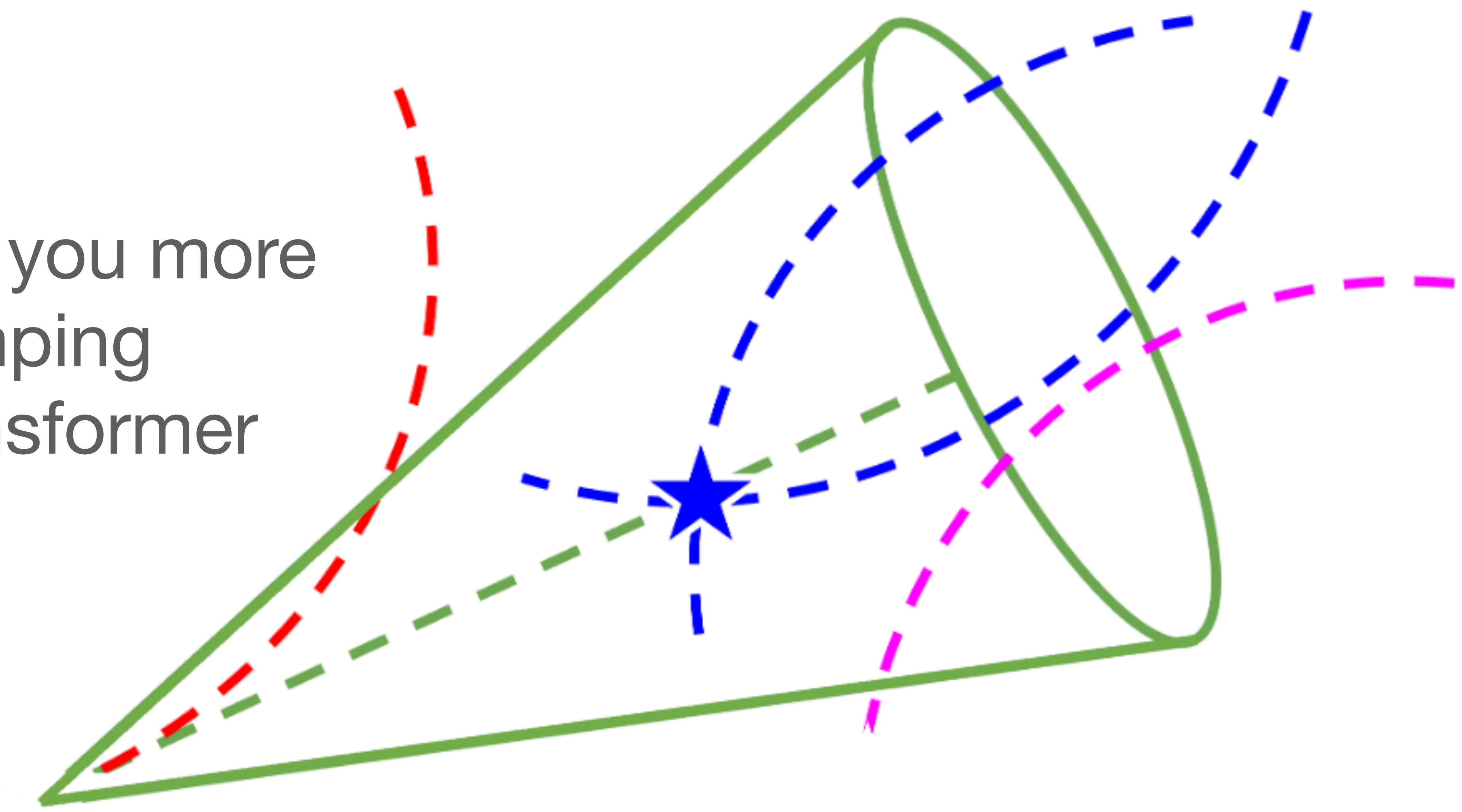


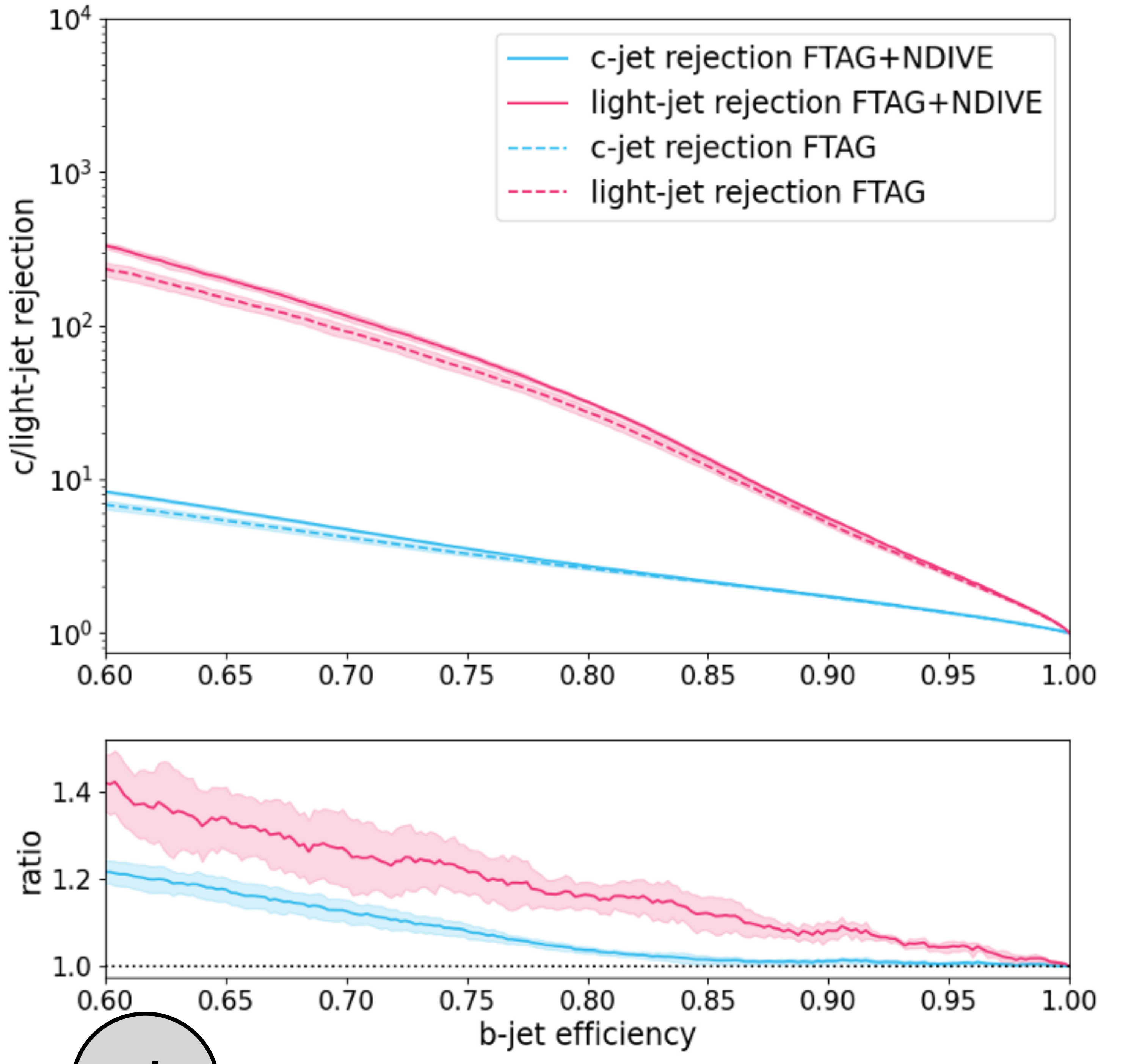
Differentiable Inference

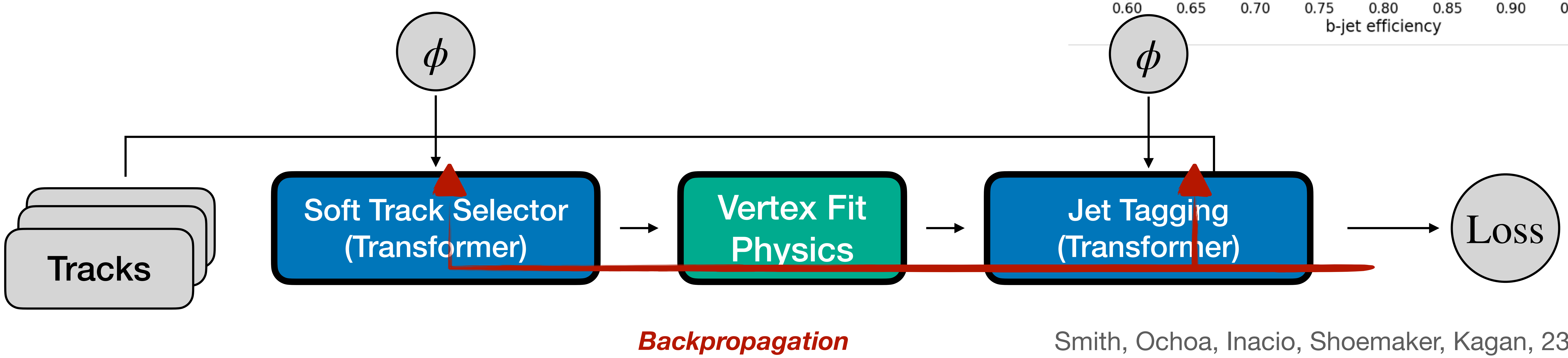
10.21105/joss.02823

Recent Example of a true Hybrid:

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





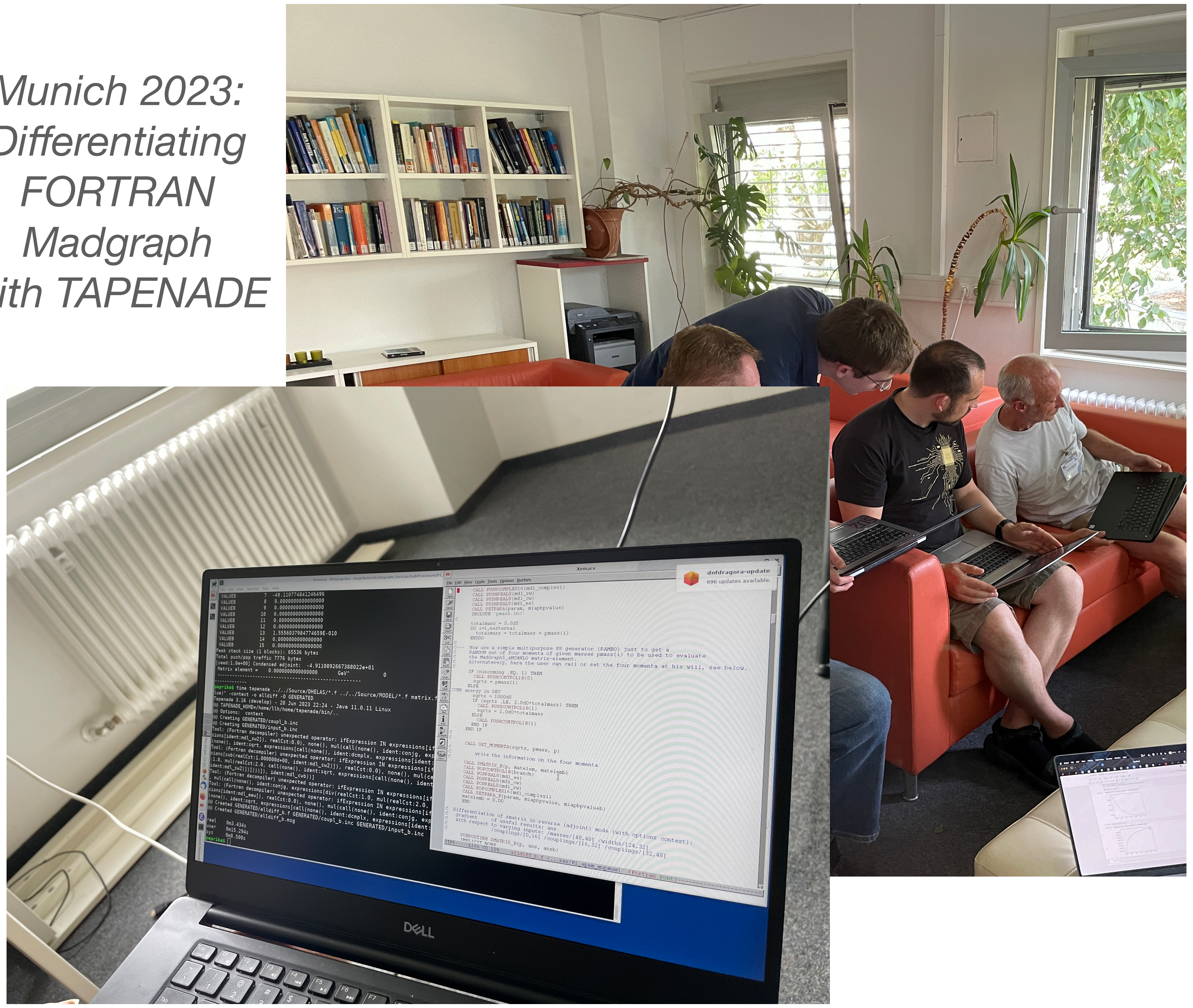


Smith, Ochoa, Inacio, Shoemaker, Kagan, 2310.12804

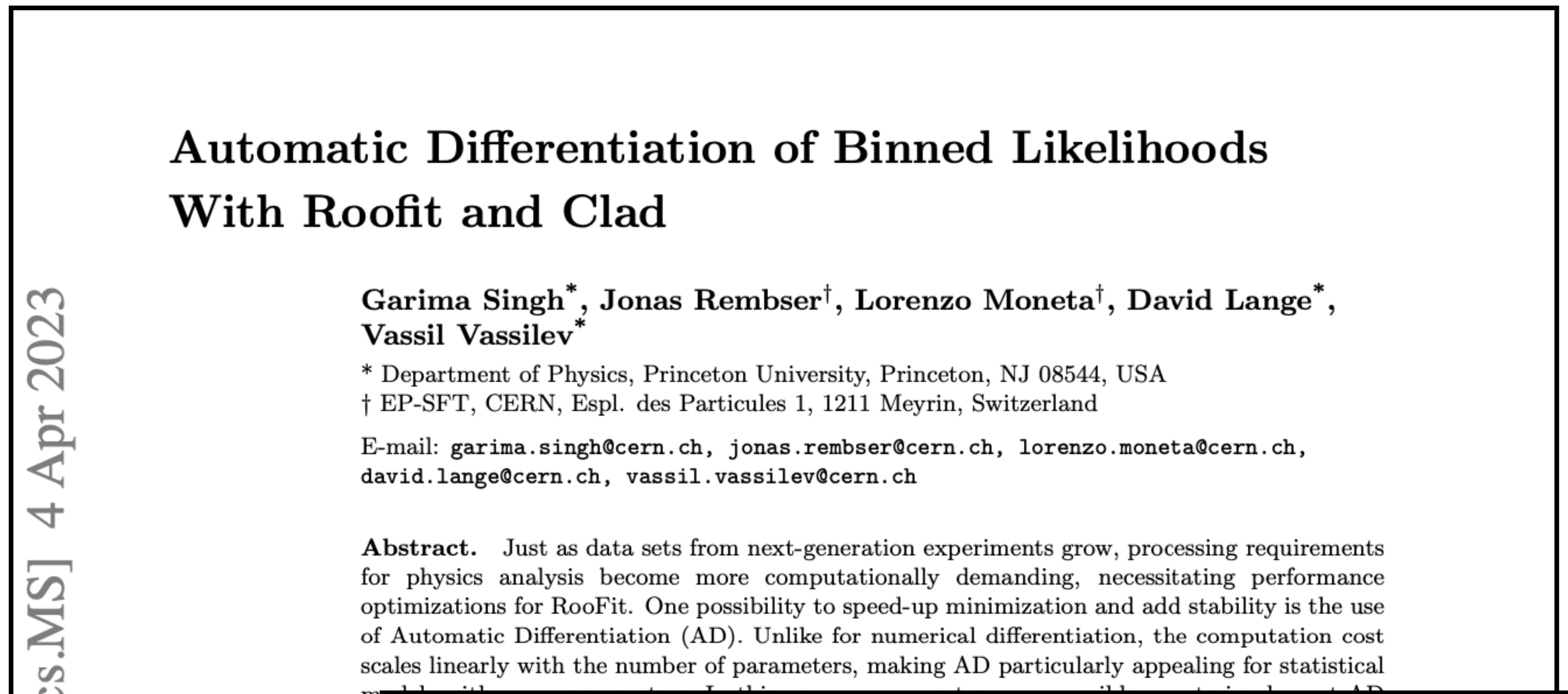
Not too worried about Integrations

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

Munich 2023: *Differentiating FORTRAN Madgraph with TAPENADE*



G. Singh, V. Vassilev et al - Clad



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

Automatic Differentiation of Binned Likelihoods With Roofit and Clad

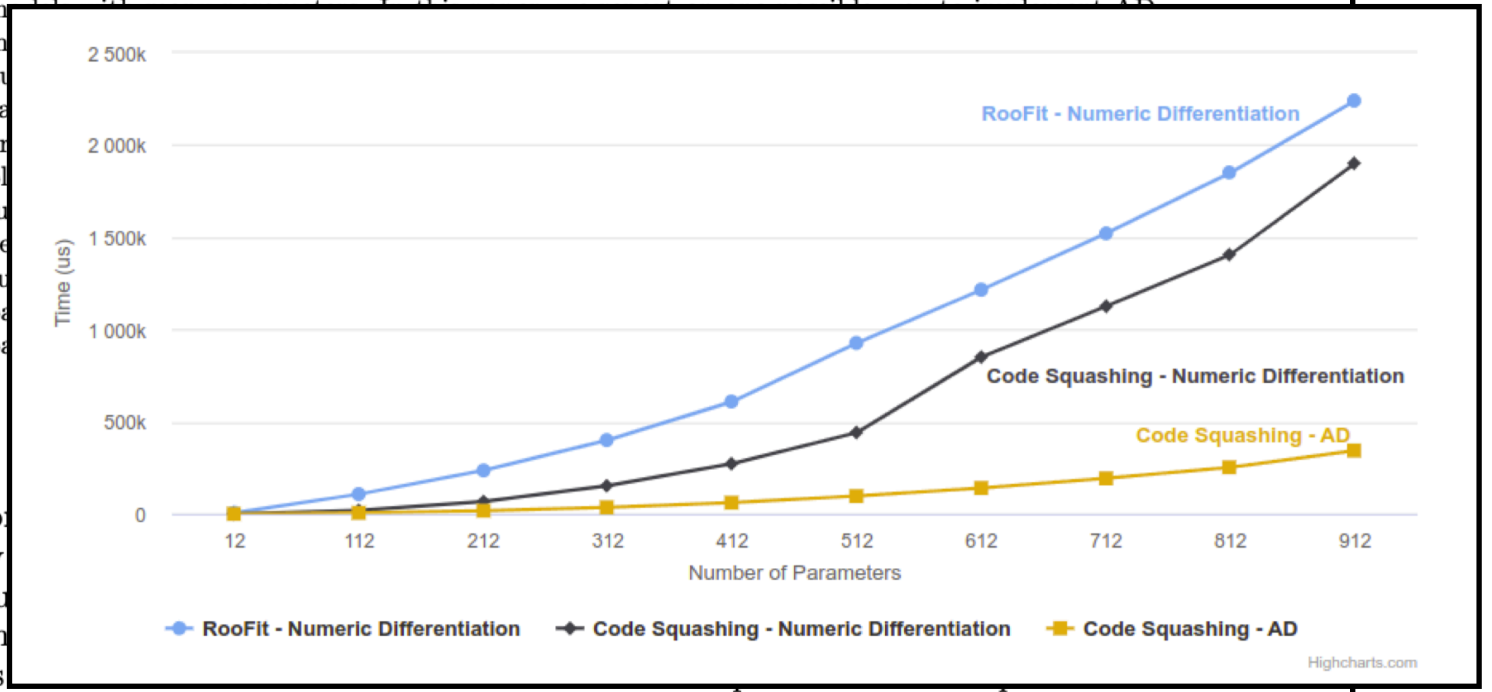
Garima Singh*, Jonas Rembser†, Lorenzo Moneta†, David Lange*, Vassil Vassilev*

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

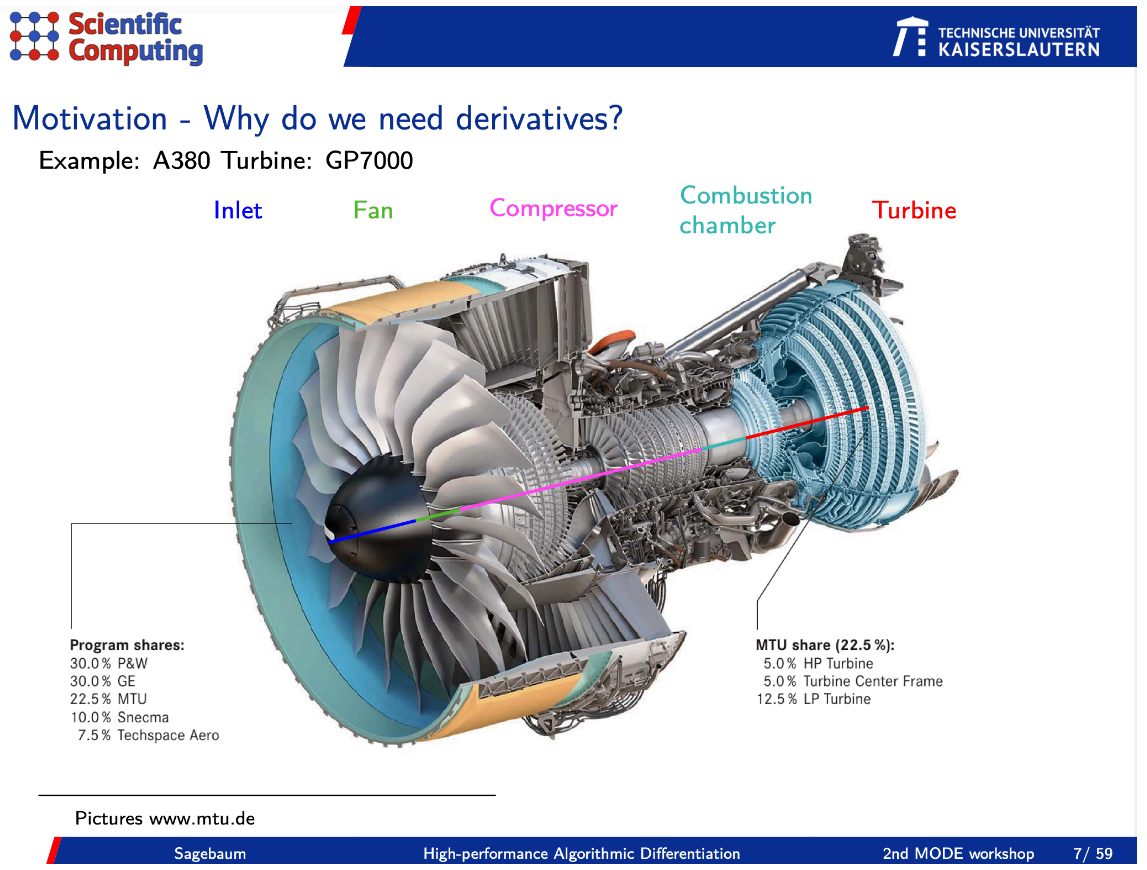
E-mail: garima.singh@cern.ch, jonas.rembser@cern.ch, lorenzo.moneta@cern.ch, david.lange@cern.ch, vassil.vassilev@cern.ch

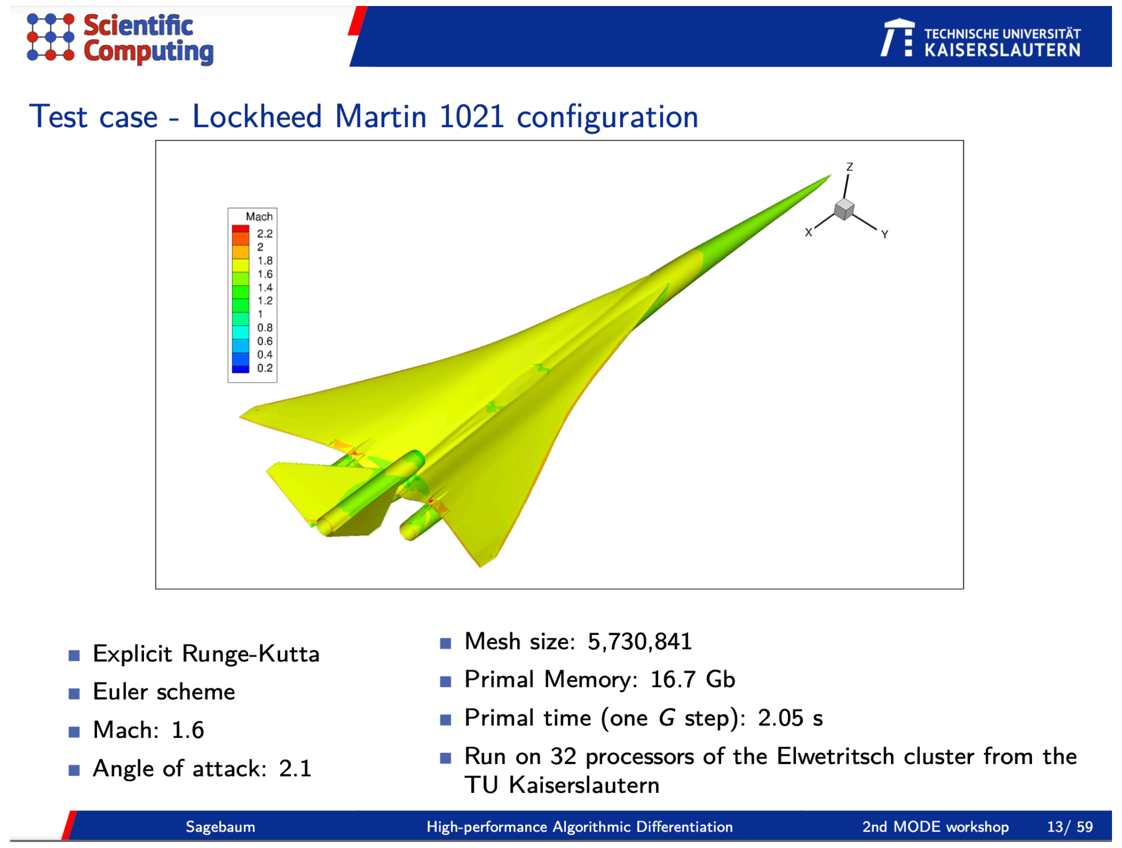
Abstract. Just as data sets from next-generation experiments grow, processing requirements for physics analysis become more computationally demanding, necessitating performance optimizations for RooFit. One possibility to speed-up minimization and add stability is the use of Automatic Differentiation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

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



M. Sagebaum - CoDIPack





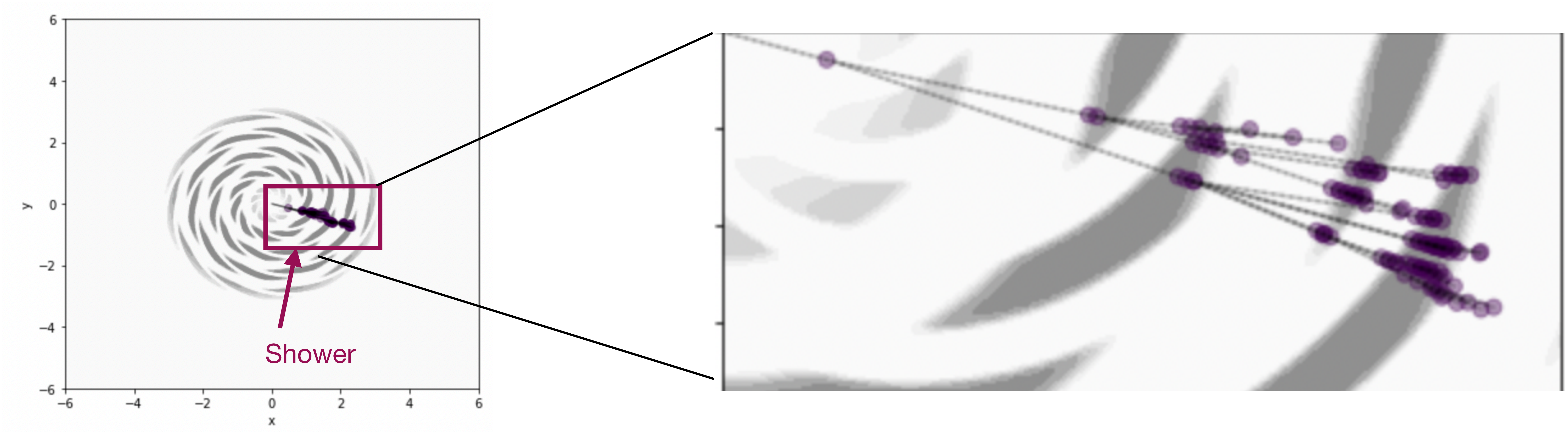
Challenges

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.

Challenges

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

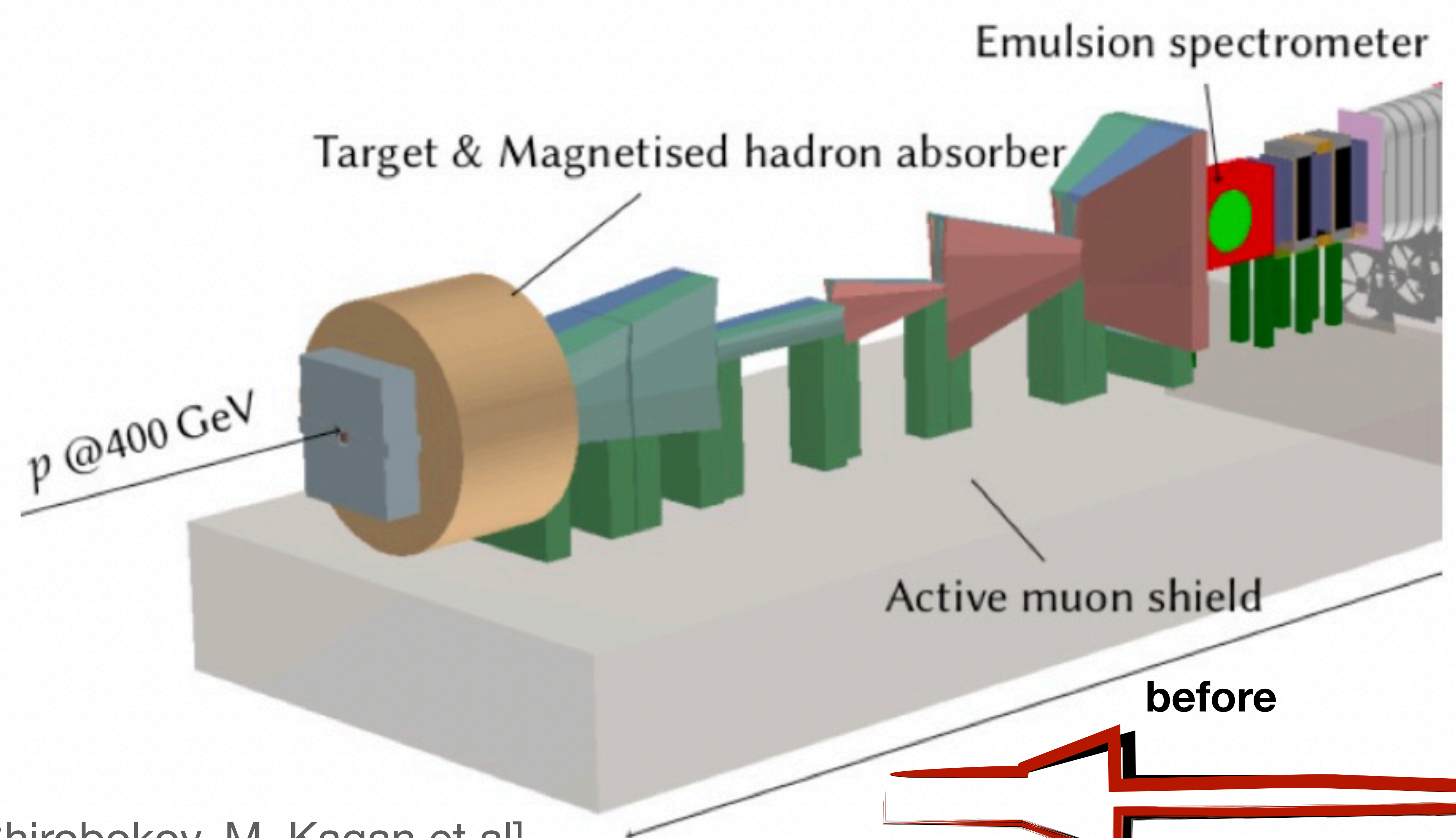


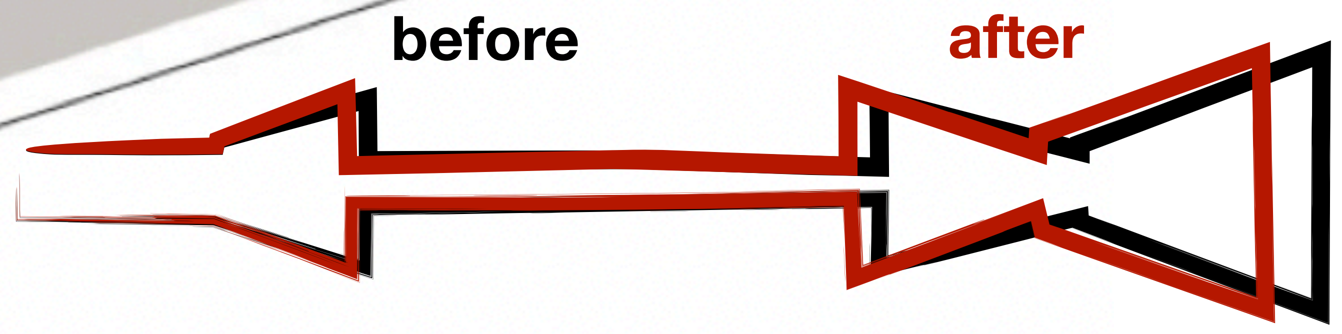
Key: Adapt methods from e.g. Reinforcement Learning or newer "Stochastic Automatic Differentiation"

arxiv:2308.16680 [LH, M. Kagan] arxiv:2210.08572 [G. Arya, et al]

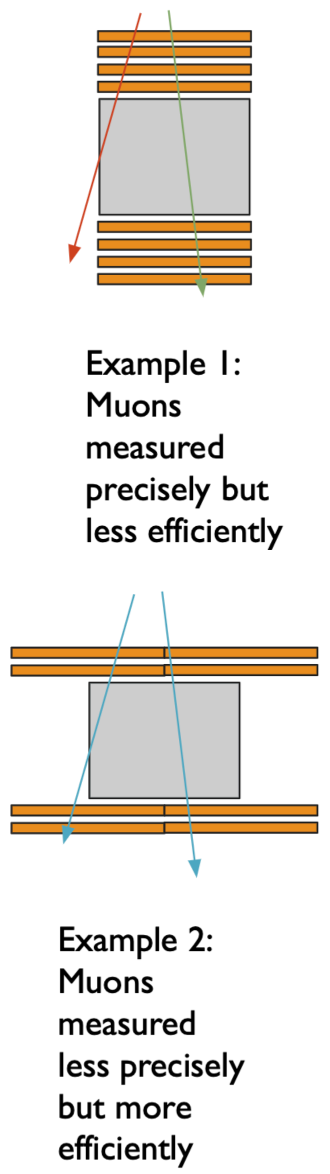
Application: Detector Design

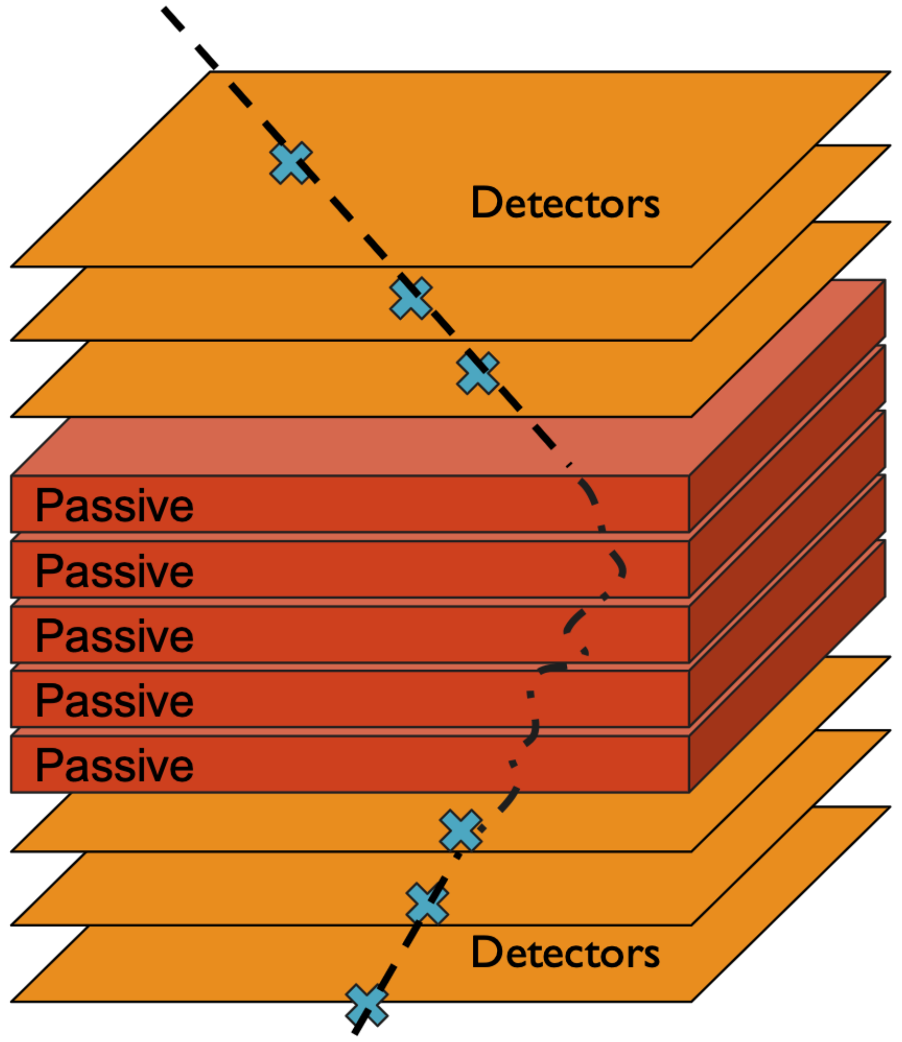
With a differentiable detector simulation or a neural network surrogate, we can optimize the detector design





[S. Shirobokov, M. Kagan et al]
arxiv:2002.04632





G. Strong, T. Dorigo et al
2309.14027

Design

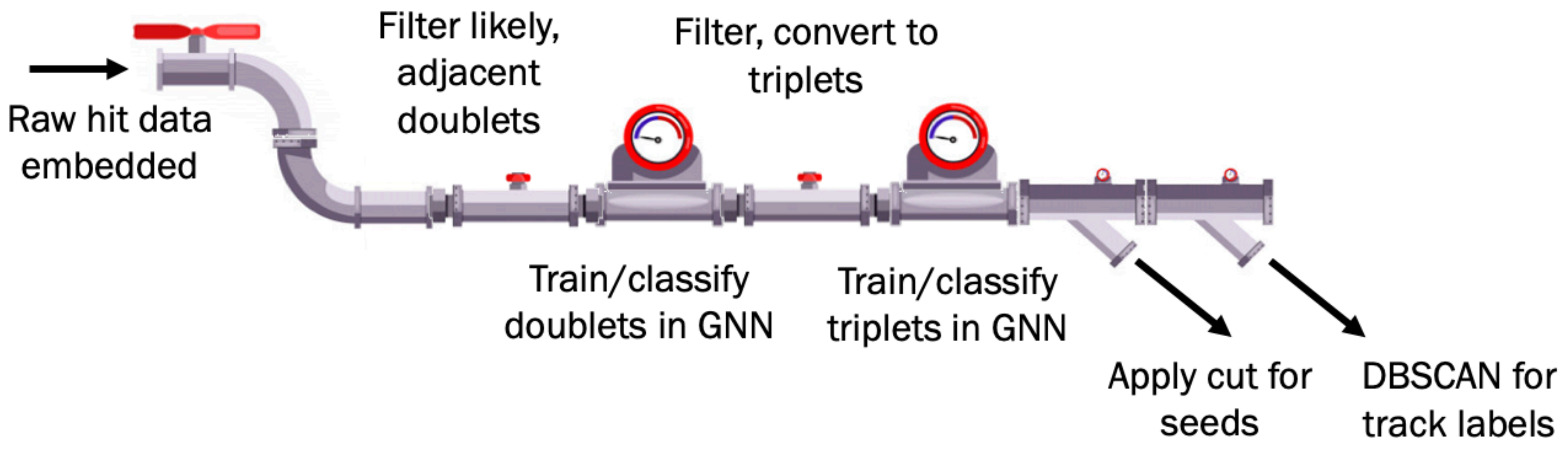
Physics Simulator (or Surrogate) → Possibly ML Reco & Analysis

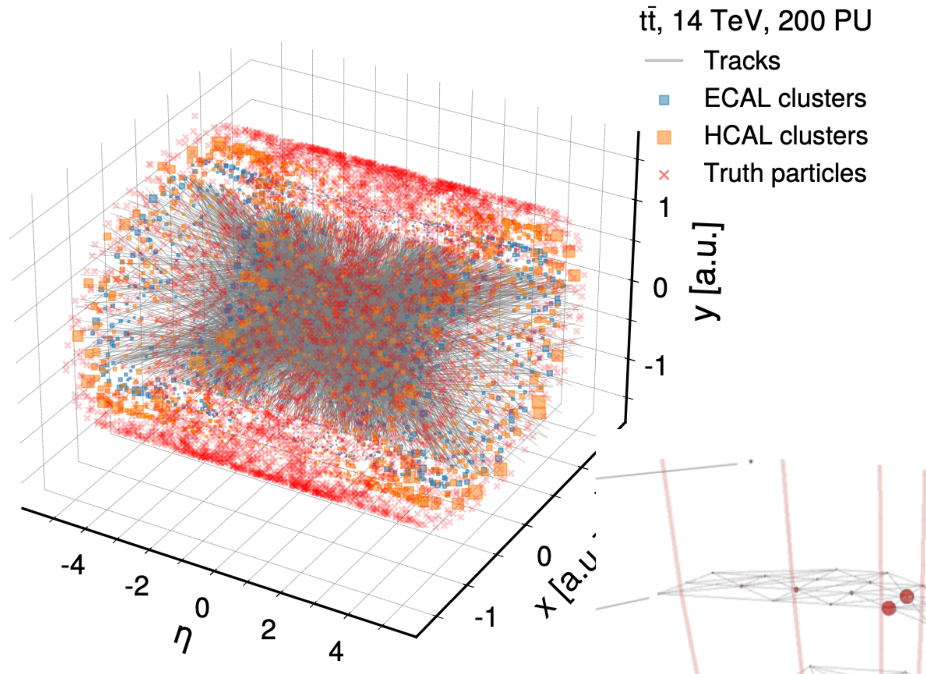
Performance

Backpropagation

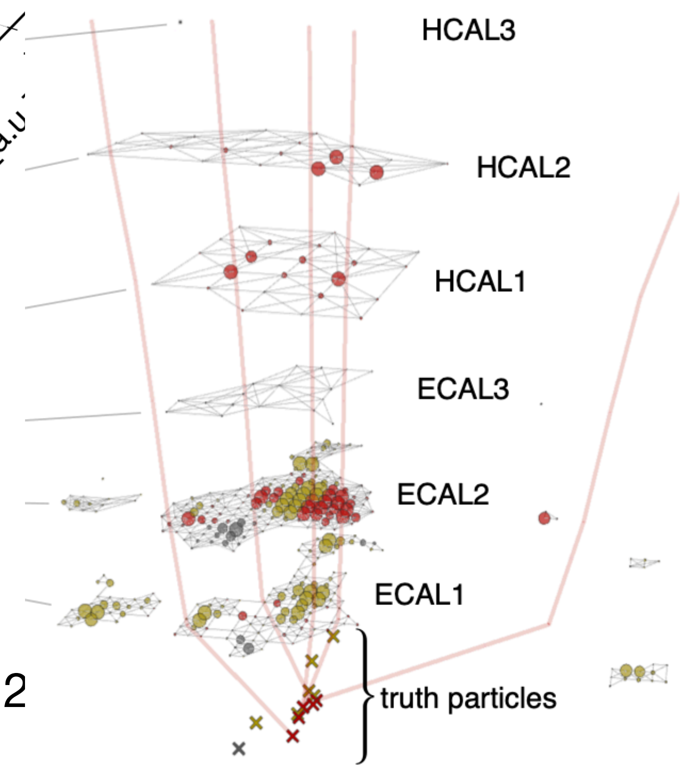
Challenges

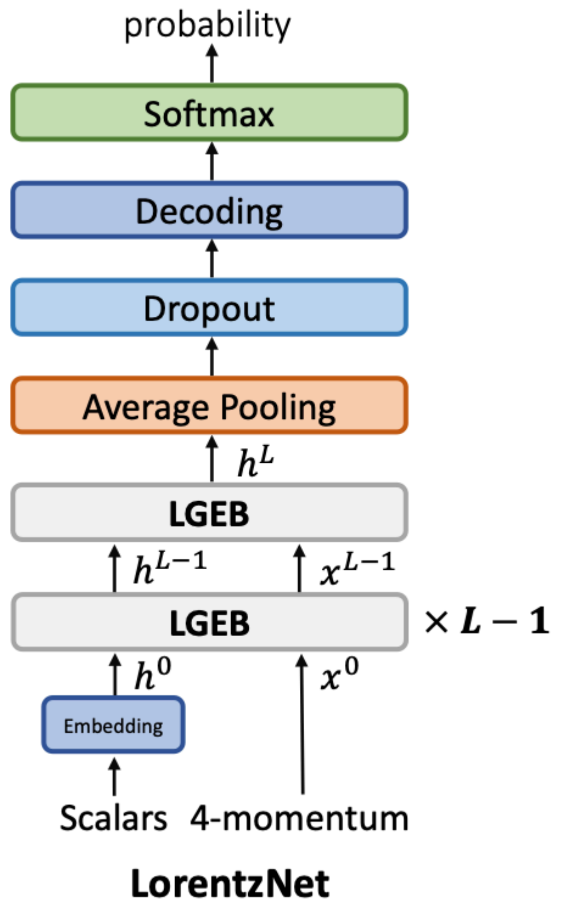
Gradients are not the only secret sauce of ML. Over-parametrization is important. Hybrid Systems likely work best with fairly big neural components connected by some physics



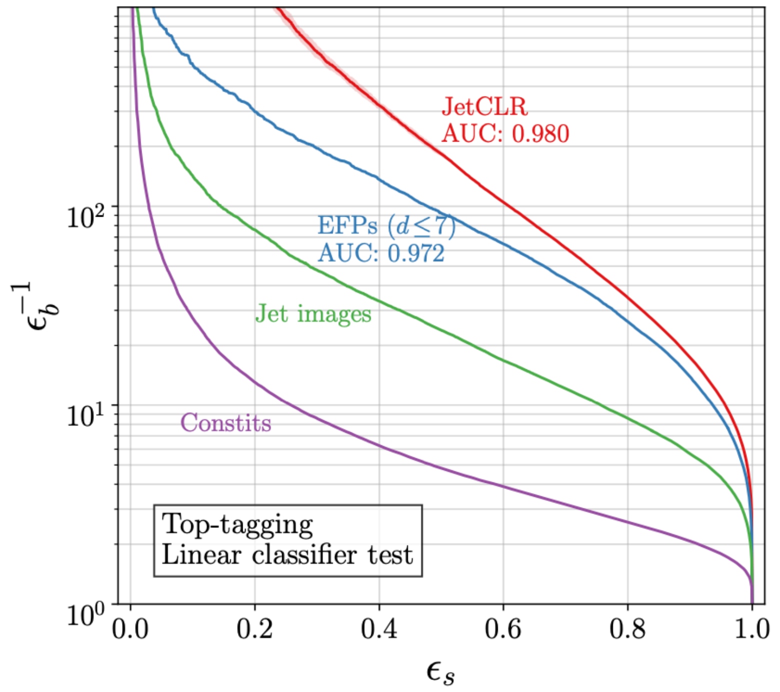


arxiv:2101.08578





arxiv: 2201.08187



arxiv: 2108.04253

ML Tracking (Graphs Neural Nets)

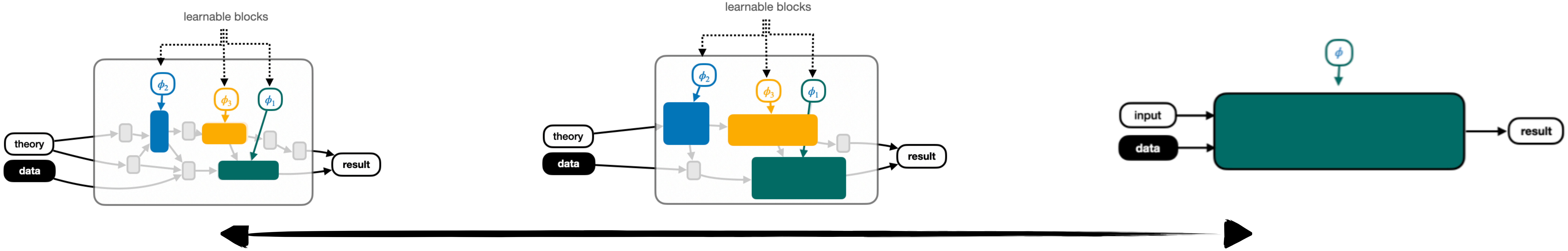
ML Particle Flow (Graphs & Transformers)

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?



intricate differentiable programming pipeline
small ML + lots of physics

few big ML blocks /
connected by little physics

maximalist ML
(black box from raw data)