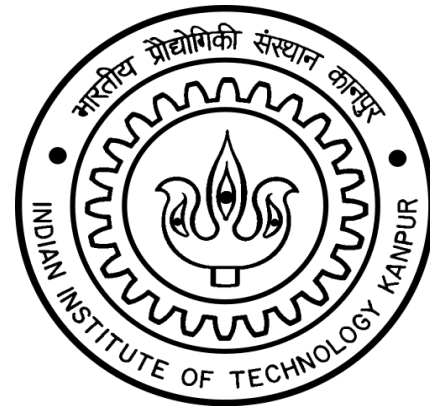


# Machine Learning in High Energy Physics



Swagata Mukherjee (IIT Kanpur)

SWPNIOE 2026, September 11–12



# Machine Learning in High Energy Physics Collider experiments



# ML & HEP: old friendship

DELPHI Collaboration



DELPHI 92-20 PHYS 159

25 February 1992

## B Tagging With Neural Networks An Alternative Use of Single Particle Information for Discriminating Jet Events<sup>1</sup>

P. Branchini, M. Ciuchini

INFN - Sezione Sanità  
Scuola del dottorato di ricerca - Università "La Sapienza" - Roma  
Istituto Superiore di Sanità - Physics Laboratory

P. Del Giudice

Istituto Superiore di Sanità - Physics Laboratory  
INFN - Sezione Sanità

more than three decades ago...

## TUTORIAL ON NEURAL NETWORK APPLICATIONS IN HIGH ENERGY PHYSICS: A 1992 PERSPECTIVE\*

BRUCE DENBY  
Fermi National Accelerator Laboratory  
M.S. 318  
Batavia, Illinois 60510 U.S.A.  
denby@fnal.bitnet

### ABSTRACT

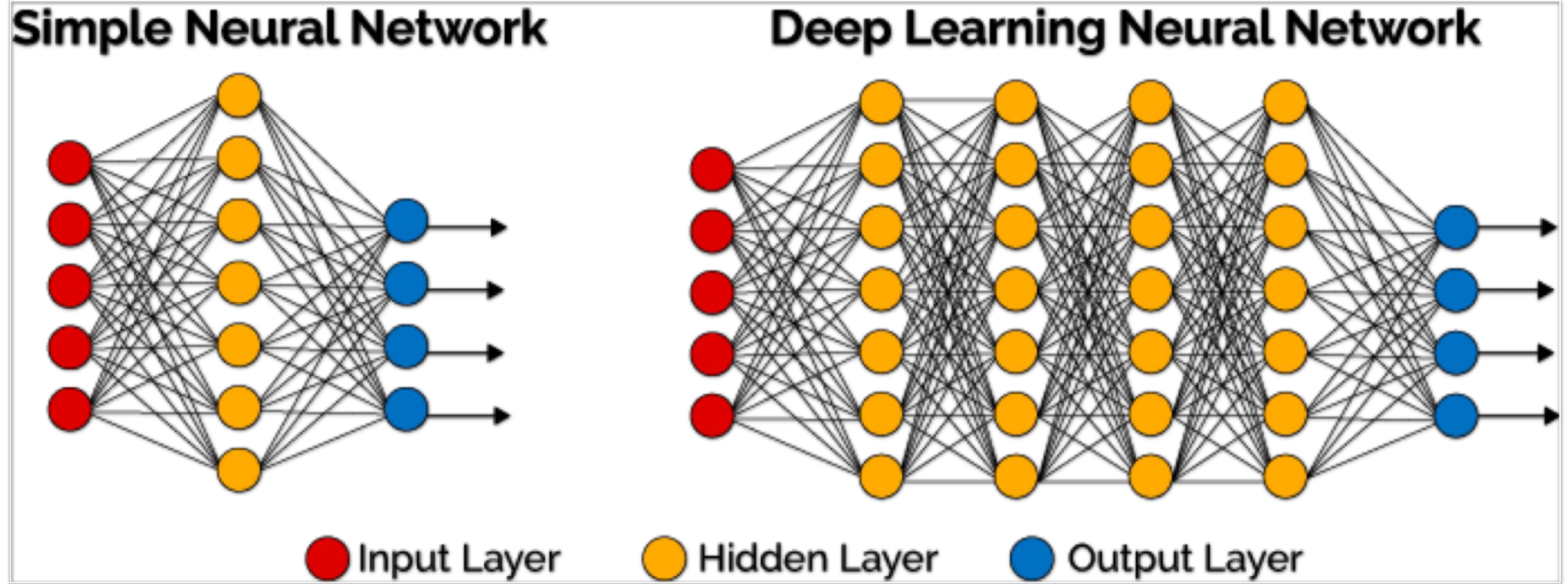
Feed forward and recurrent neural networks are introduced and related to standard data analysis tools. Tips are given on applications of neural nets to various areas of high energy physics. A review of applications within high energy physics and a summary of neural net hardware status are given.

### Outline of Paper

1. Architectures
  - 1.1 Feed Forward Networks
    - 1.1.1 Event Selection
      - 1.1.1.1 Example 1 - Square Root Net
      - 1.1.1.2 Example 2 - Tau Particle Selection
      - 1.1.1.3 Example 3 - Secondary Vertex Selection
    - 1.1.2 Function Mapping
      - 1.1.2.1 Example 1 - Sin(x) Net
      - 1.1.2.2 Example 2 - Drift Chamber z Position
  - 1.2 Recurrent Nets and Tracking
    - 1.2.1 Tracking with a Recurrent Net
    - 1.2.2 Improvements to Tracking
      - 1.2.2.1 Rotor Tracking
      - 1.2.2.2 Elastic Tracking
2. Tips on Applying NN in HEP
  - 2.1 Choice of Variables
    - 2.1.1 Offline Applications
    - 2.1.2 Trigger Applications
  - 2.2 Training
    - 2.2.1 General Comments
    - 2.2.2 Neural Network Software Packages
    - 2.2.3 Low Level Pattern Recognition
    - 2.2.4 Physics Process Determination
    - 2.2.5 Example of a Hybrid Application - Isolation
3. Survey of NN Applications in HEP
  - 3.1 Track Segment and Vertex Finding
  - 3.2 Quark/Gluon Separation

# From shallow to deep ...

more data, more computing power, ML-friendly hardware (e.g. GPUs)



$$\hat{y}_i = f(\sum_j w_{ij}x_j + b_i)$$

# ImageNet Classification with Deep Convolutional Neural Networks

January 2012

Paper by G. Hinton et al. launches  
the Deep CNN revolution

Alex Krizhevsky  
University of Toronto  
kriz@cs.utoronto.ca

Ilya Sutskever  
University of Toronto  
ilya@cs.utoronto.ca

Geoffrey E. Hinton  
University of Toronto  
hinton@cs.utoronto.ca

**Nobel Prize in  
Physics 2024**



## Abstract

We trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than the previous state-of-the-art. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of the convolution operation. To reduce overfitting in the fully-connected layers we employed a recently-developed regularization method called “dropout” that proved to be very effective. We also entered a variant of this model in the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry.

At different times,  
different algorithms were  
at the edge of ML research



January 2012  
Paper by G. Hinton et al. launches  
the Deep CNN revolution

LHC/CERN was busy with Higgs  
discovery. But picked up momentum  
on deep CNN soon after.

HEP-ex community: early adopters of  
Machine Learning. Higgs discovery made use  
of Boosted Decision Tree (BDT)

**Data Science @ LHC 2015**  
Bridging High-Energy Physics and Machine Learning communities  
9 - 13 November 2015, CERN

**Local Organising Committee**

- Xavier Cif (CERN)
- Gilles Louppe (CERN)
- Michelangelo Mangano (CERN)
- Maurizio Pierini (CERN)
- Jean-Roch Vannier (Caltech)

**Program Committee**

- Kyle Cranmer (New York U)
- C.G. G. Germani (CERN)
- Valentin Voevodsky (CERN)
- Gilles Louppe (CERN)
- Andrew Senior (University of Cambridge)
- Maurizio Pierini (CERN)
- David Rosenberg (LAL Orsay)
- Maria Spiropulu (Caltech)
- Jean-Roch Vannier (Caltech)
- Daniel Whiteson (UC Irvine)

**International Advisory Committee**

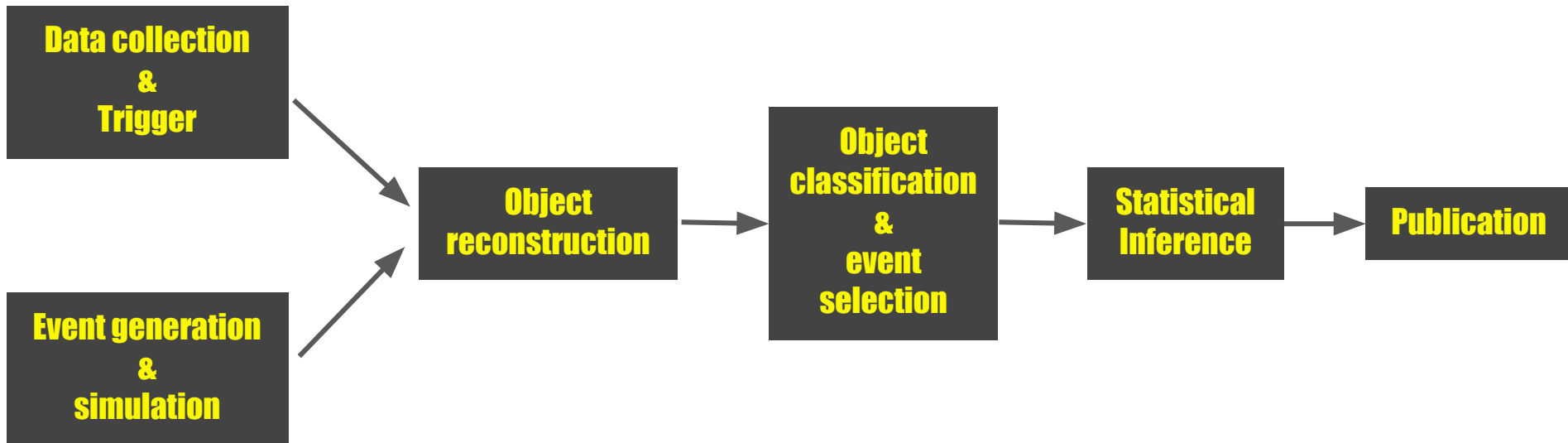
- Roger Barlow (Huddersfield U)
- Tamas Borki (Wigner Inst.)
- Ian Fisk (Siemens Foundation)
- Maria Gromova (CERN)
- Eleni Kousi (Weizmann)
- Balázs Kégl (LAL Orsay)
- Constantin Loupas (LHC)
- Stuart Russell (UC Berkeley)
- Victoria Stodden (U. of Illinois-Chicago)
- Max Welling (Amsterdam U)

sponsored by  
LHC Physics Center at CERN: <http://lpc.web.cern.ch>  
Fermilab National Laboratory: <http://fnl.gov>  
Moore-Sloan Data Science Environment: <http://cds.nyu.edu/moore-sloan>

<http://cern.ch/DataScienceLHC2015>

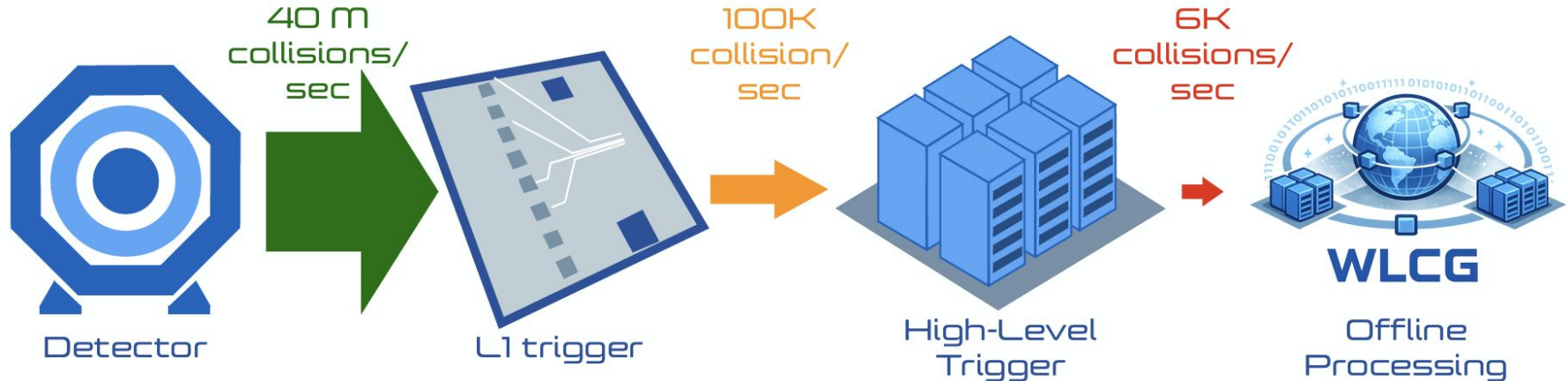
# Full chain of a publication from CMS/ATLAS *in a nutshell*

ML is used in basically every stage



# Triggering @ hadron collider experiments (CMS/ATLAS)

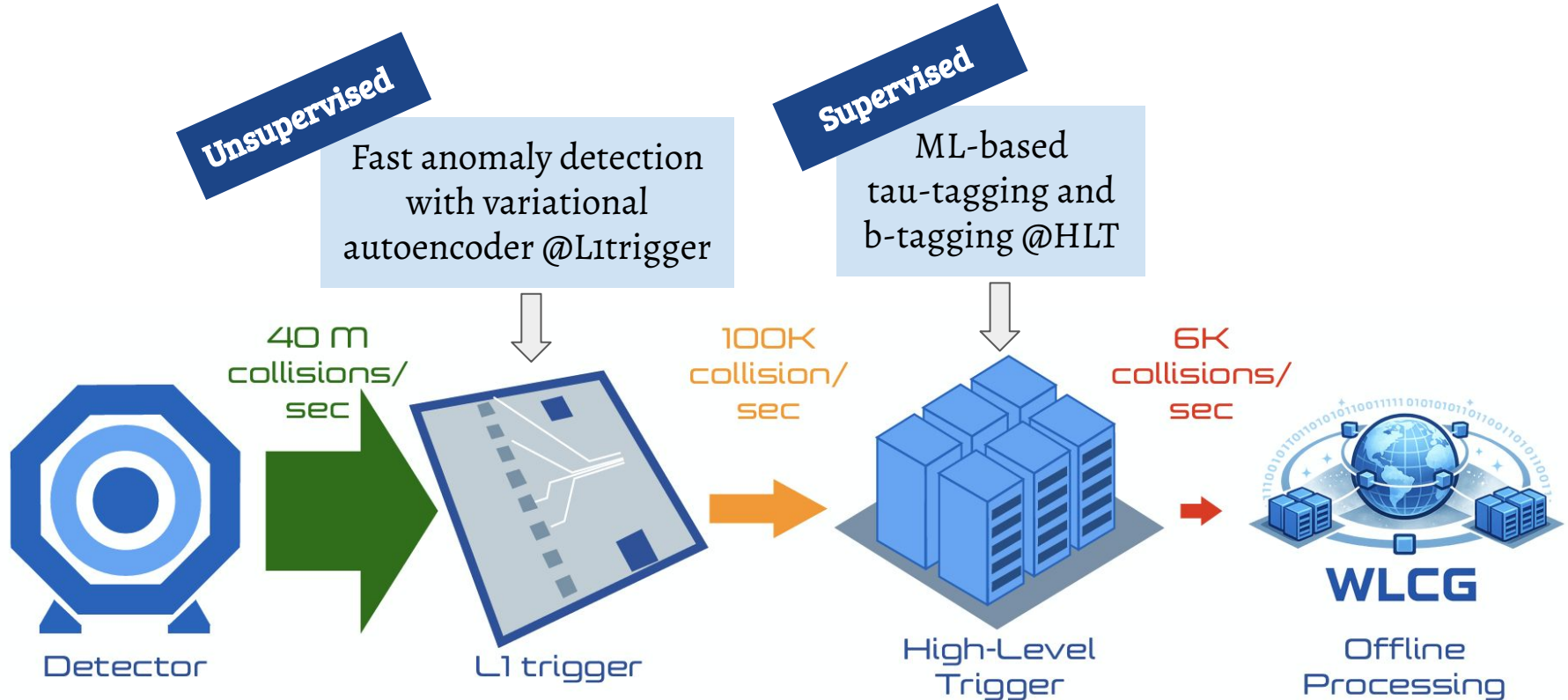
- ❑ One bunch-crossing every 25 ns (i.e, 40M collisions/sec).
- ❑ Hundreds of particles emerging from each bunch-crossing.
- ❑ Each recorded event is O(1) MB of data (CMS/ATLAS).
- ❑ The total amount of data produced by LHC is too large to be stored by CMS/ATLAS.
- ❑ Can only store O(1)% of total events → Real-time selection (trigger)



# ML @ trigger level

Human bias in triggers? Are we throwing away events containing BSM?

CMS introduced unsupervised techniques to mitigate this. Anomaly detector!



# Big Data

<https://www.symmetrymagazine.org/article/10-years-of-lhc-physics-in-numbers> March 2020

# 278

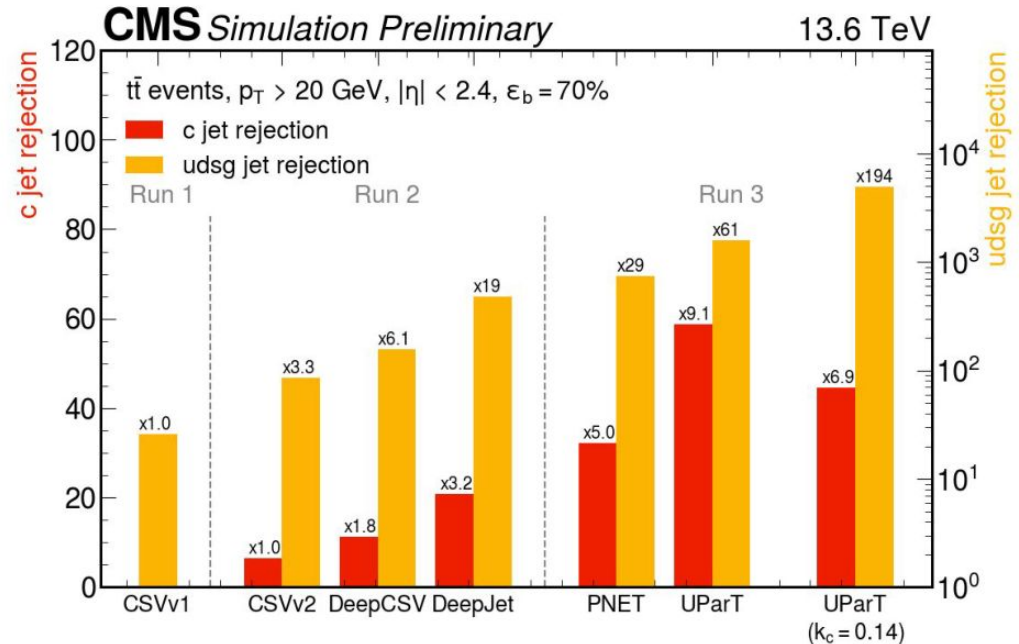
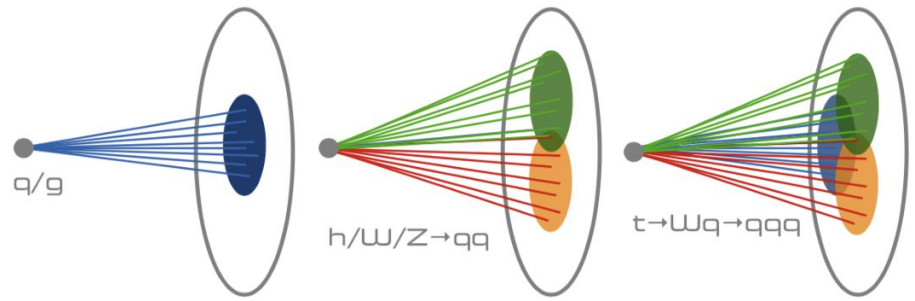
**petabytes of data**

1 PB =  $10^{15}$  Byte

In the last decade, LHC experiments collected almost 280 petabytes of data, which scientists recorded on tape. You would need to stream Netflix 24/7 for more than 15,000 years to eventually use that much data! But from another perspective, platforms like Facebook (which has 2.5 billion users) collect that much data in 70 days!

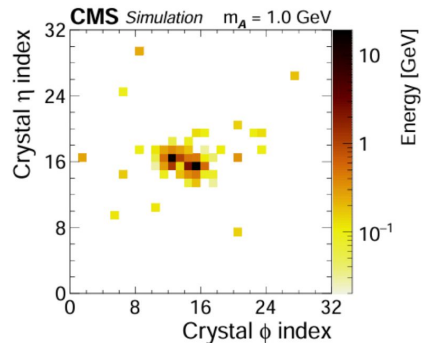
# Jet tagging

- ❑ Jet: spray of (mostly hadronic) particles produced at a collision.
- ❑ Jet tagging: study a jet and guess which particle started the shower. Traditionally a testbed for ML (since LEP era)
- ❑ With modern ML, performance boost beyond expectations
- ❑ ATLAS/ CMS can tag  $c$ -quark jets (and  $H \rightarrow cc$  decays)

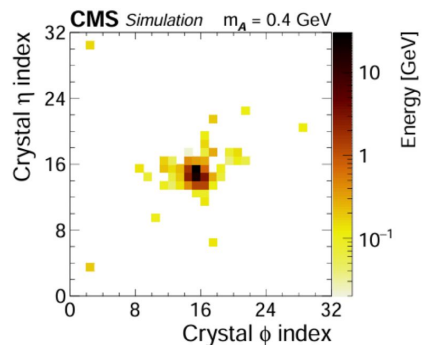


# Experimentally challenging searches for BSM

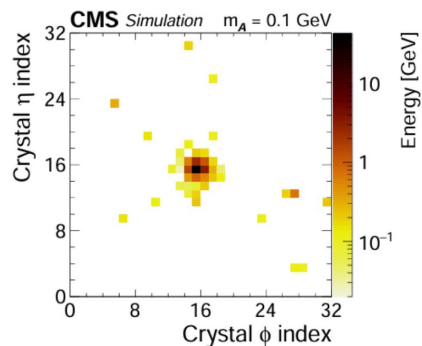
- Low hanging fruits are mostly gone from the BSM landscape.
- Focus shifted towards exotic, experimentally challenging signatures.
  - ♦ Long-lived BSM particle: displaced signatures.
  - ♦ Very low mass BSM particles: merged signatures.
  - ♦ A regime where cut-based tend to fail.
    - Heavy usage of ML.



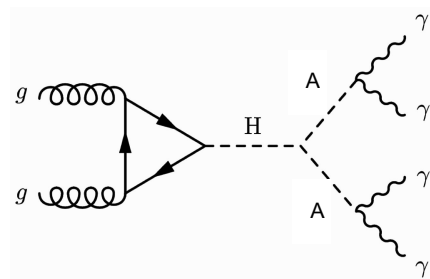
$m_A = 1.0$  GeV, Lorentz boost=50  
Two photons still visible by eye.



$m_A = 0.4$  GeV, Lorentz boost=150  
Looks like one photon.

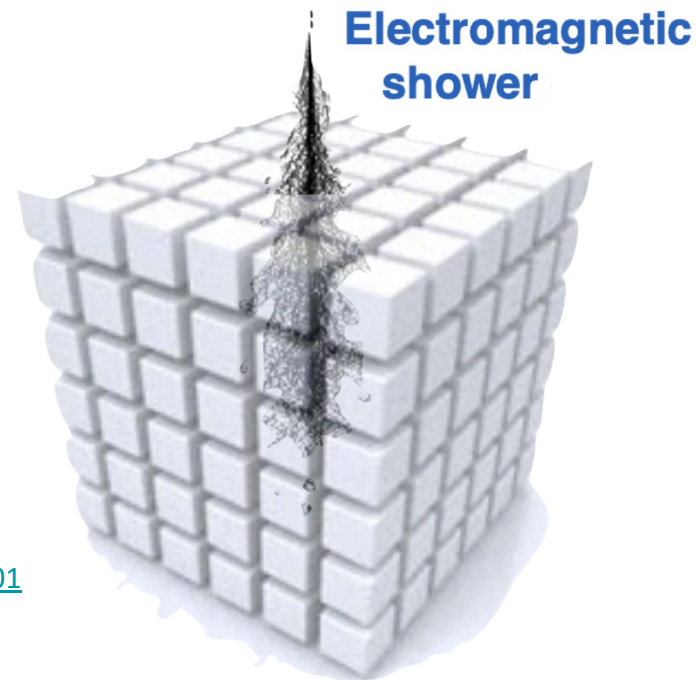


$m_A = 0.1$  GeV, Lorentz boost=625  
Looks like one photon.



Opening angle  
between two  
photons very  
small

How to reconstruct low mass  $A \rightarrow \gamma\gamma$  ?



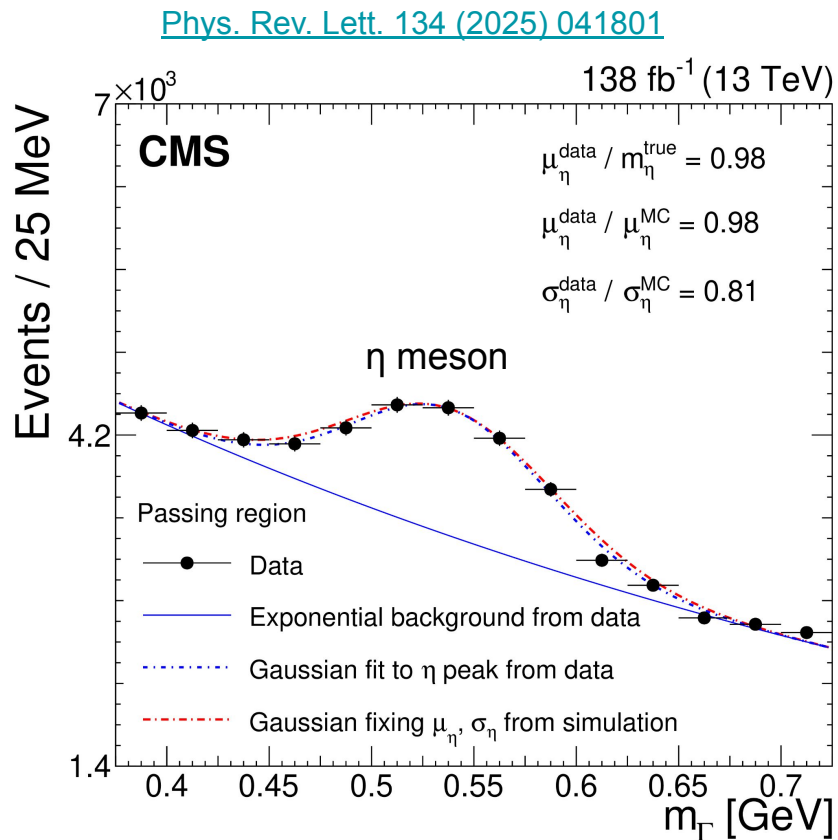
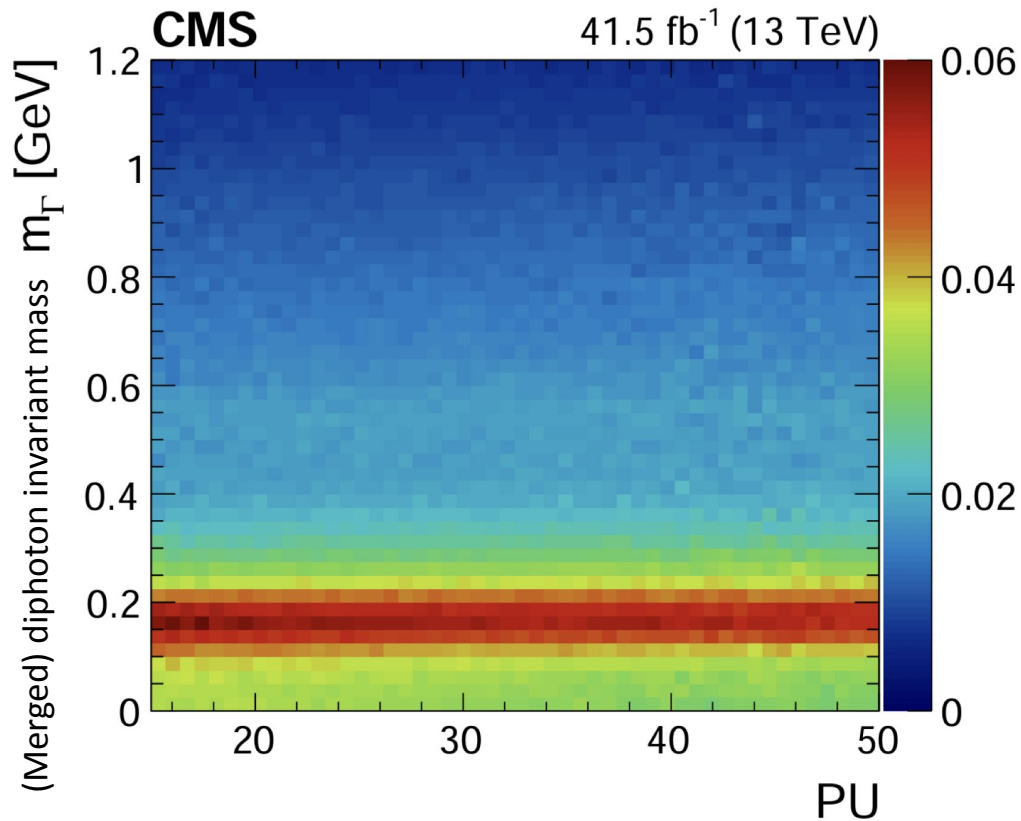
[Phys. Rev. Lett. 131 \(2023\) 101801](#)

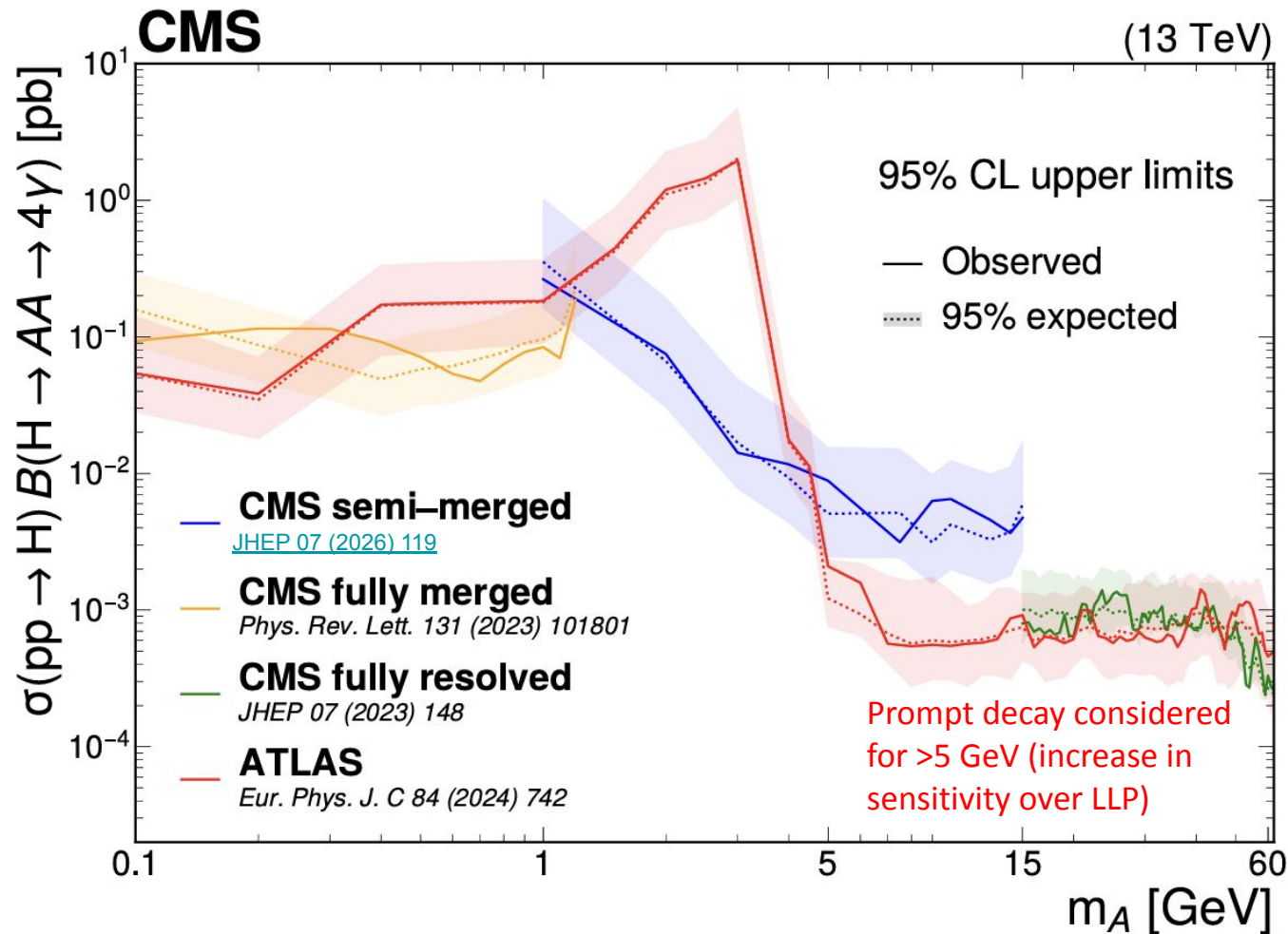
[Phys. Rev. D 108 \(2023\) 052002](#)

# Validation in real data with $\pi^0 \rightarrow \gamma\gamma$ and $\eta \rightarrow \gamma\gamma$

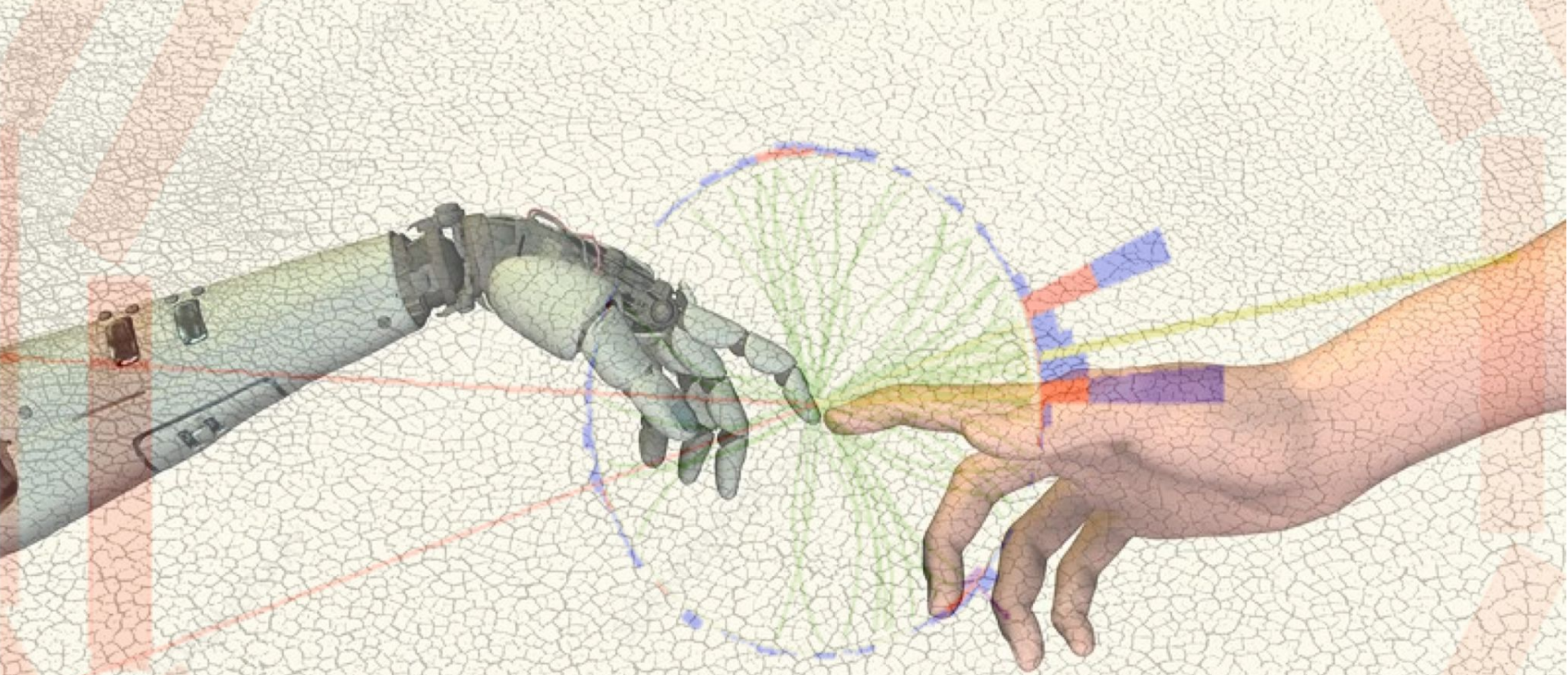
end-to-end  
reconstruction using  
CNN (ResNET)

Robust against pileup





Very low mass searches like this are possible because of ML



**LHC data is valuable and finite.**

**Make the most of it leveraging ML.**

**Extra slides**

