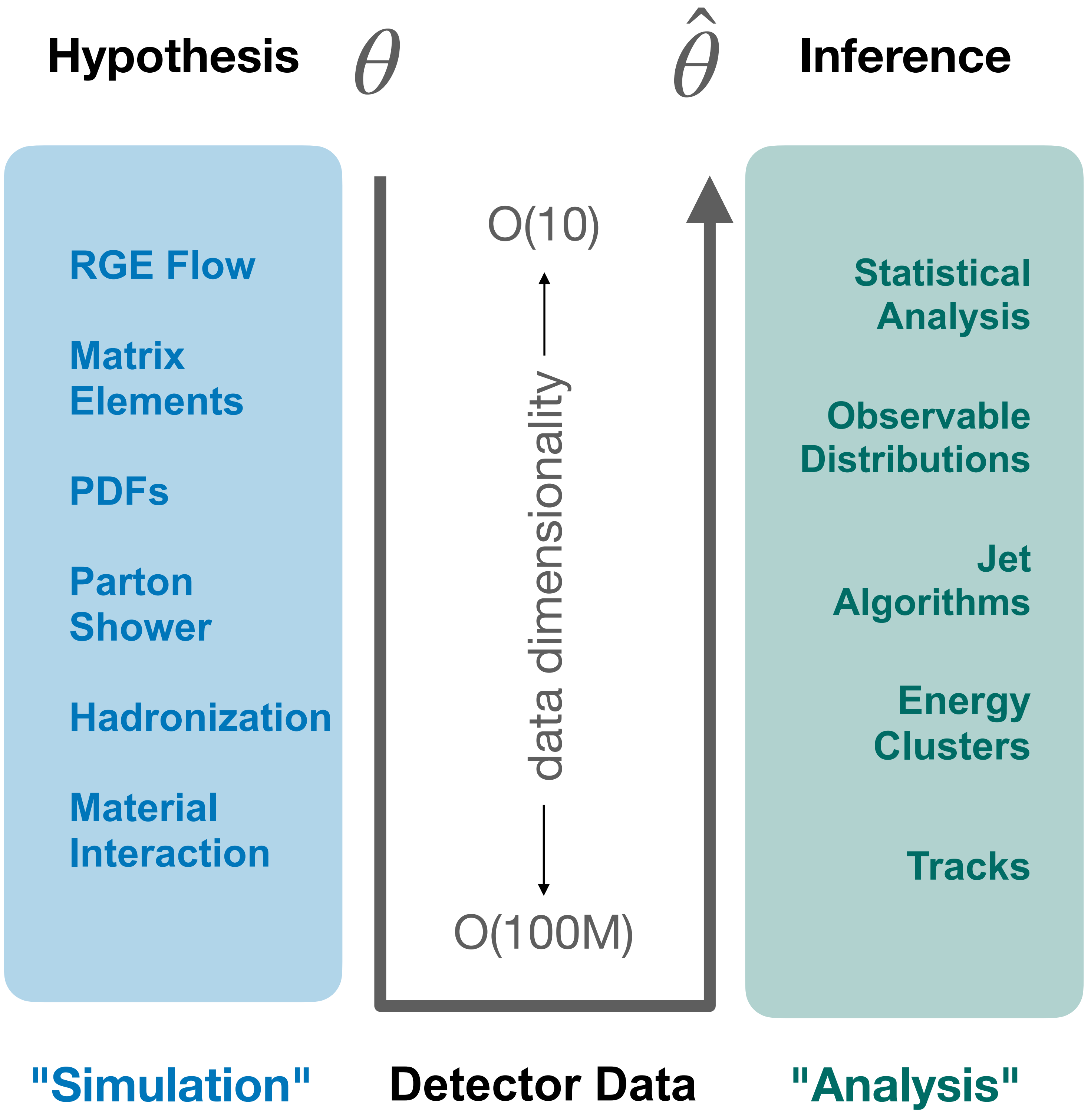
Deep Learning in Analysis

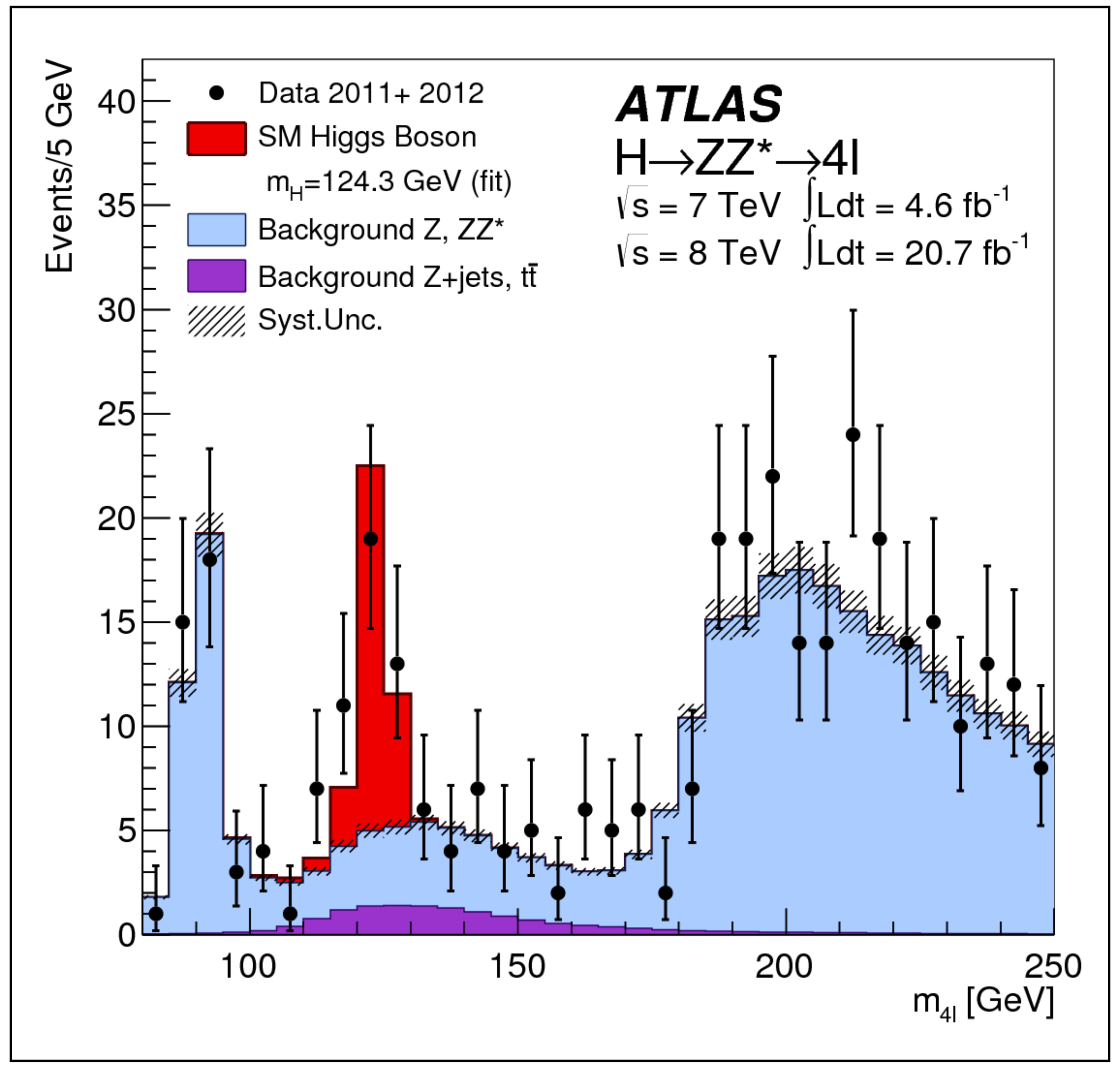
Lukas Heinrich, CERN
on behalf of the ALICE, ATLAS, CMS, and LHCb collaborations

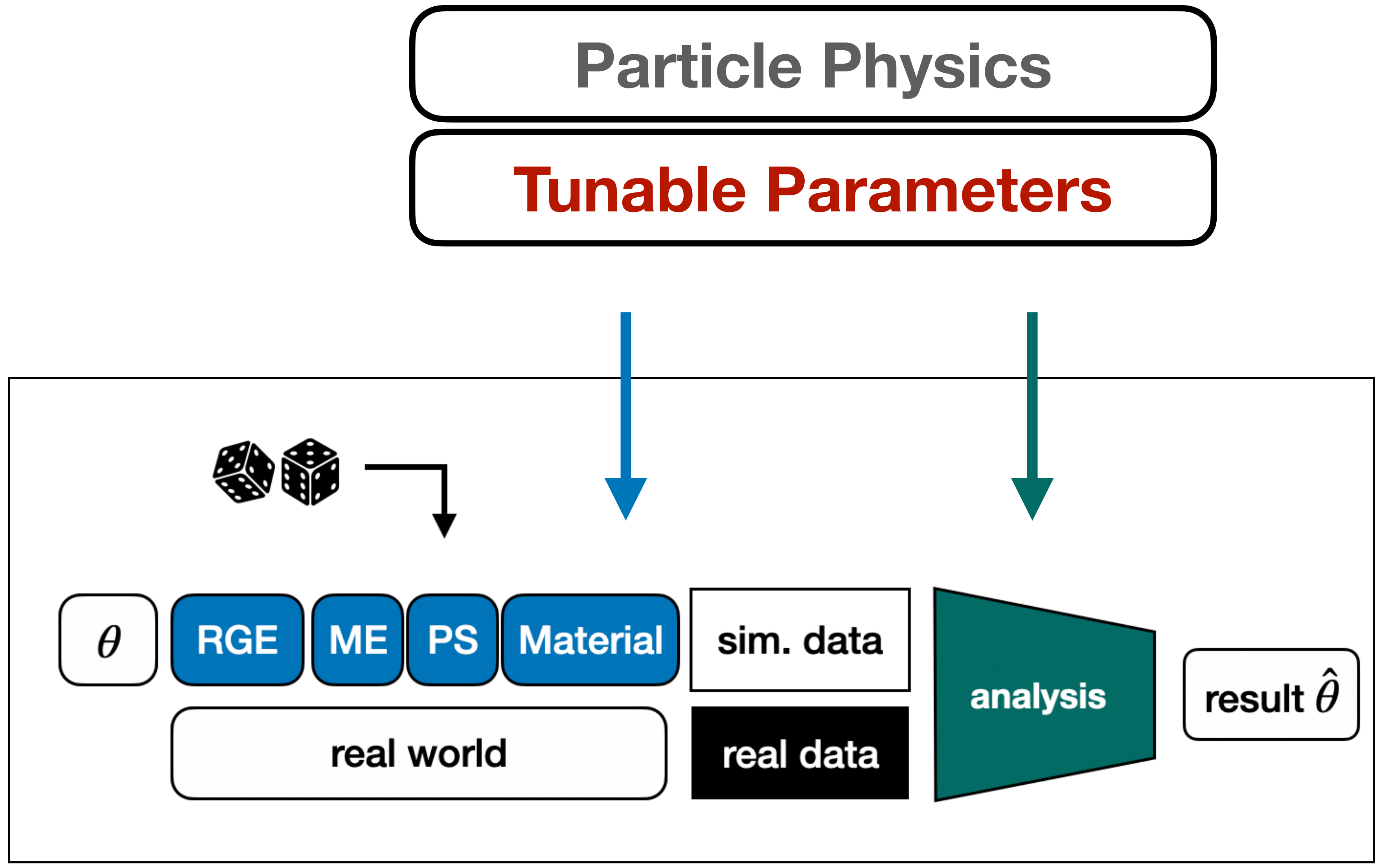
LHCP 2021

Why Deep Learning, Again?



size, data complexity & reliance on simulators
→ analyses pipelines have **many knobs** to tune





Main promise of Deep Learning in HEP

Help to efficiently explore configuration space of scientific pipeline to optimally extract information

Deep Learning Advantages: data representations & gradients

Why is it important now & why is it hard?

LHC @ 13 TeV: the era beyond the low-hanging fruit

→ focus on ever more challenging signals & phase-spaces

Good: High-velocity research cycles inside (IML) outside of HEP (CS)

Key Challenge: transfer from R&D to actual **production usage**

Today's Focus:

Methods in analysis going beyond vanilla Deep Learning with successful R&D → In Production @ LHC experiments

(lots of interesting ML on simulation side discussed elsewhere)

Shaping Discriminants

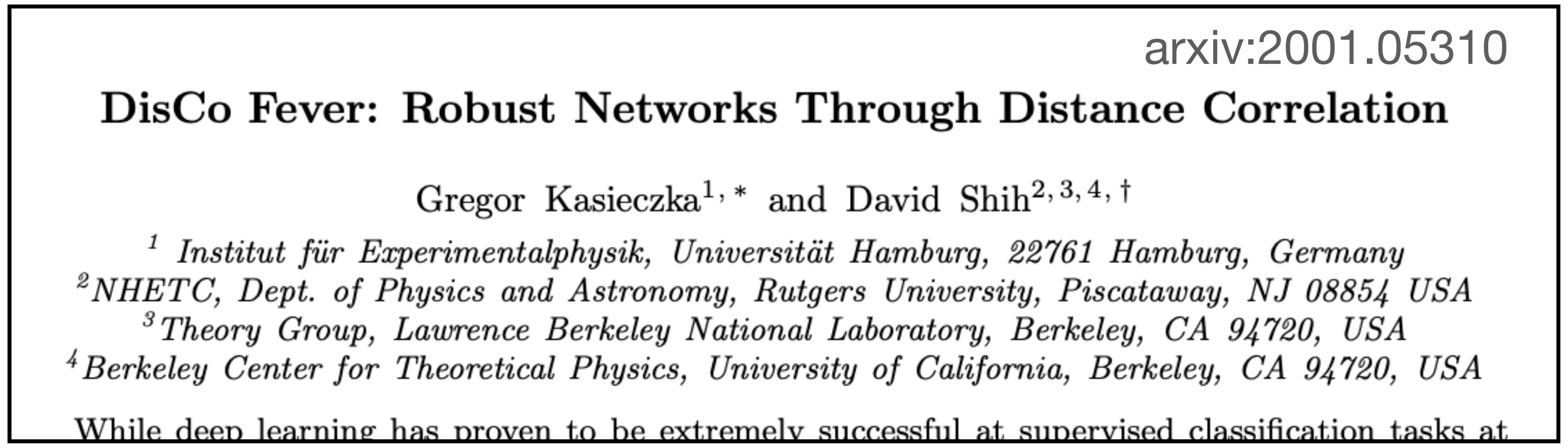
Beyond pure supervised training for challenging SUSY signals with high $n_{\rm jet}$ and low $\sigma_{\rm SUSY}$ (RPV)

Control discriminant shape to be **invariant** through loss engineering

Here: **Distance Correlation**

NN invariant to b-jet multiplicity

- first sensitivity since LEP!


arxiv:2001.05310

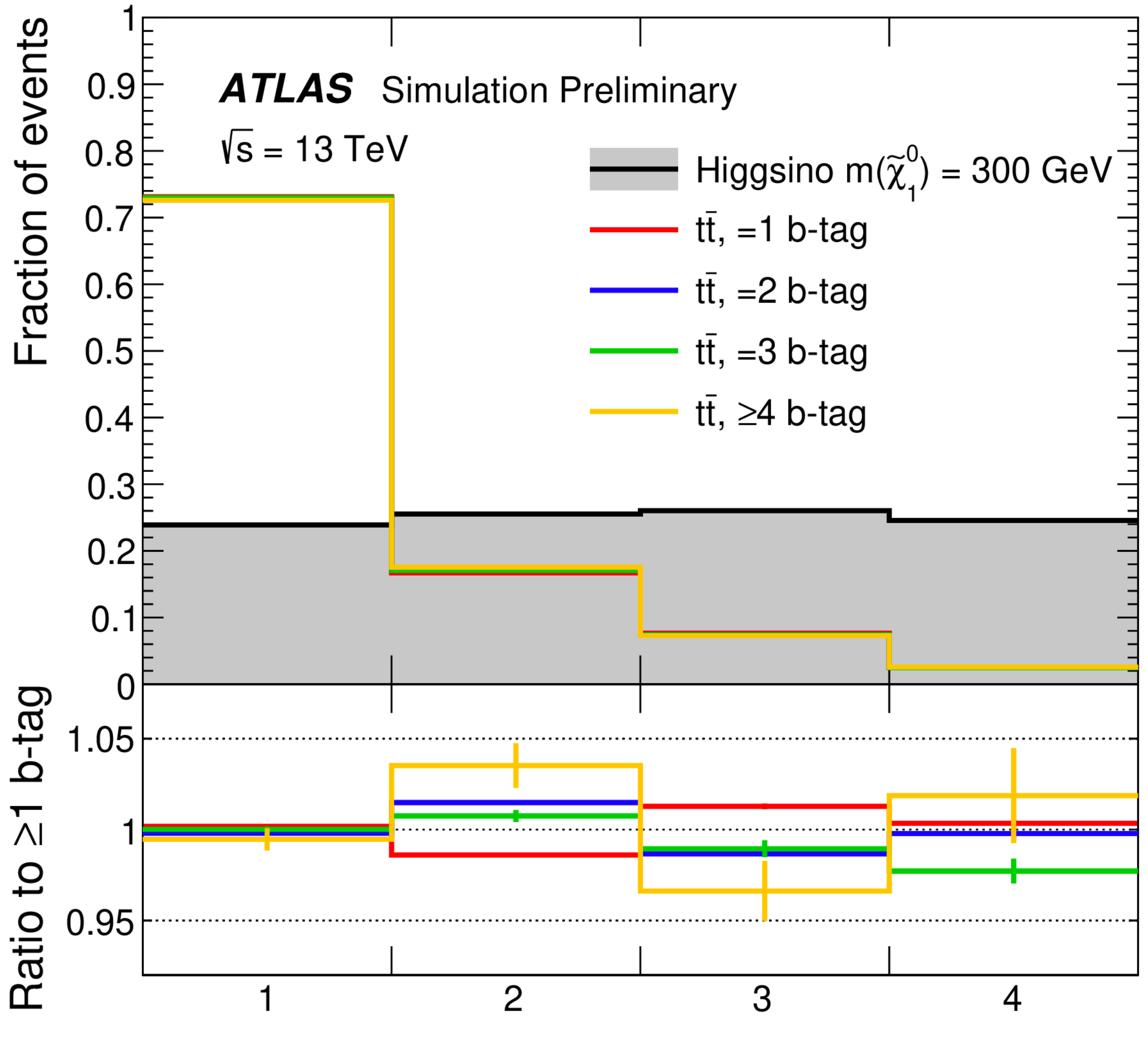
DisCo Fever: Robust Networks Through Distance Correlation

Gregor Kasieczka[1,*] and David Shih[2,3,4,†]

[1] *Institut für Experimentalphysik, Universität Hamburg, 22761 Hamburg, Germany*
[2] *NHETC, Dept. of Physics and Astronomy, Rutgers University, Piscataway, NJ 08854 USA*
[3] *Theory Group, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*
[4] *Berkeley Center for Theoretical Physics, University of California, Berkeley, CA 94720, USA*

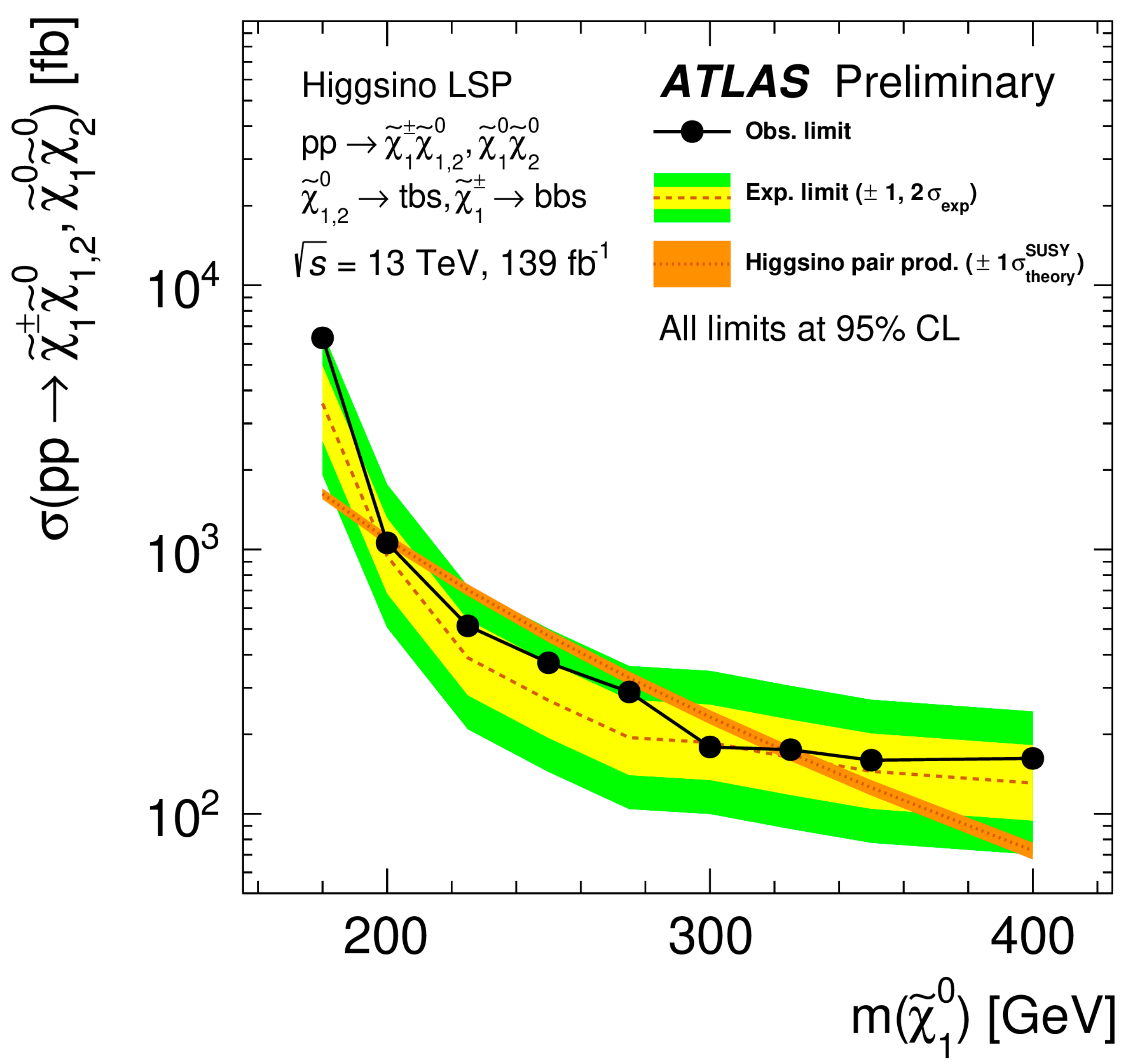
While deep learning has proven to be extremely successful at supervised classification tasks at

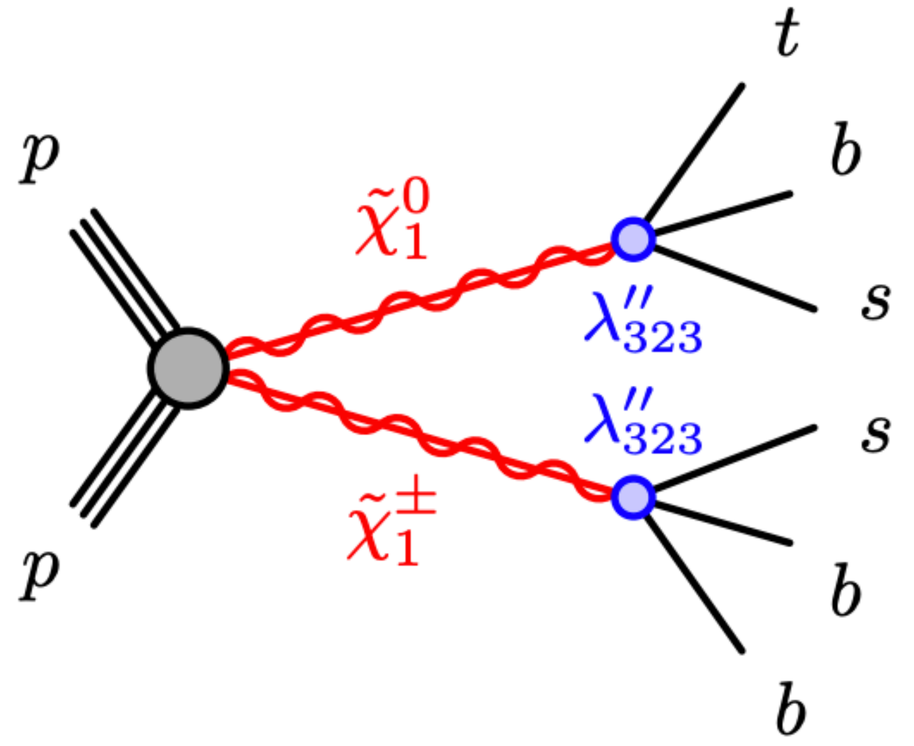
R&D Paper: Jan 2020 → Deployed in Production: Mar 2021



ATLAS-CONF-2021-007

NN bin





$$L(\phi) = L_{\rm BCE}(\hat{y}, y; \phi) + \lambda {\rm dCorr}_{y=0}(\hat{y}, n_b; \phi)$$

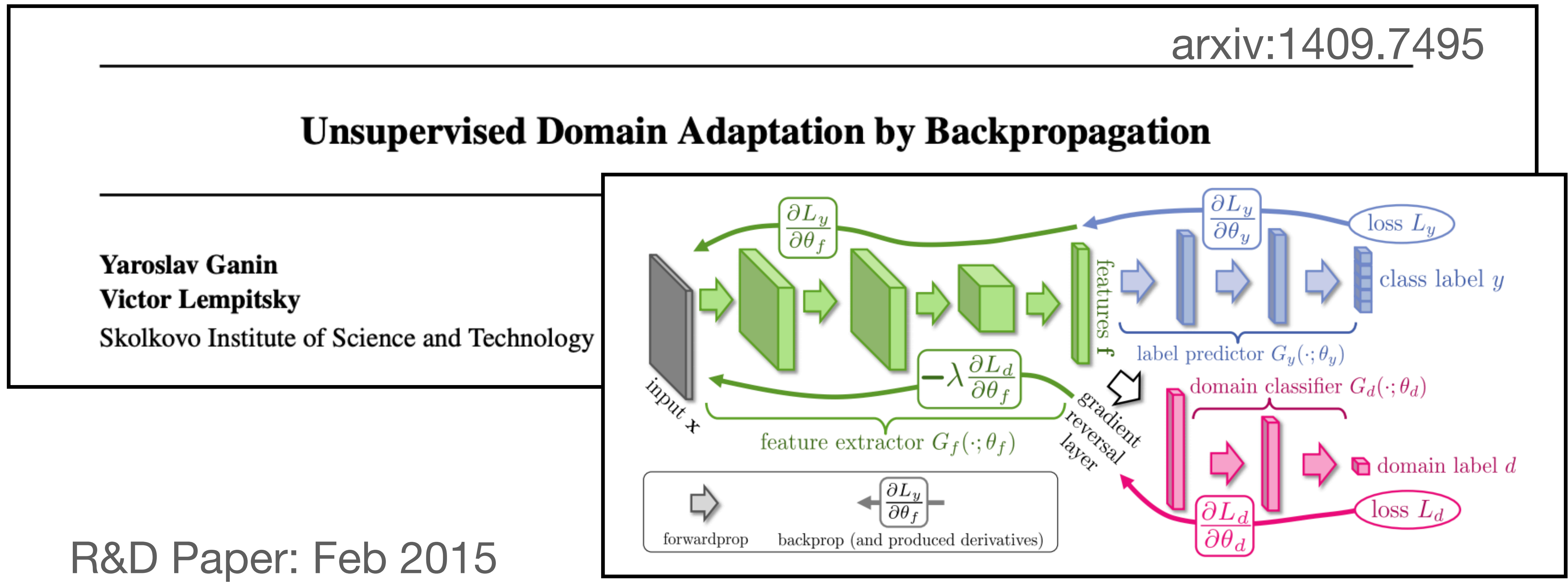
Shaping Discriminants

Beyond pure supervised training for challenging SUSY signals with high n_{jet} and low σ_{SUSY} (RPV)

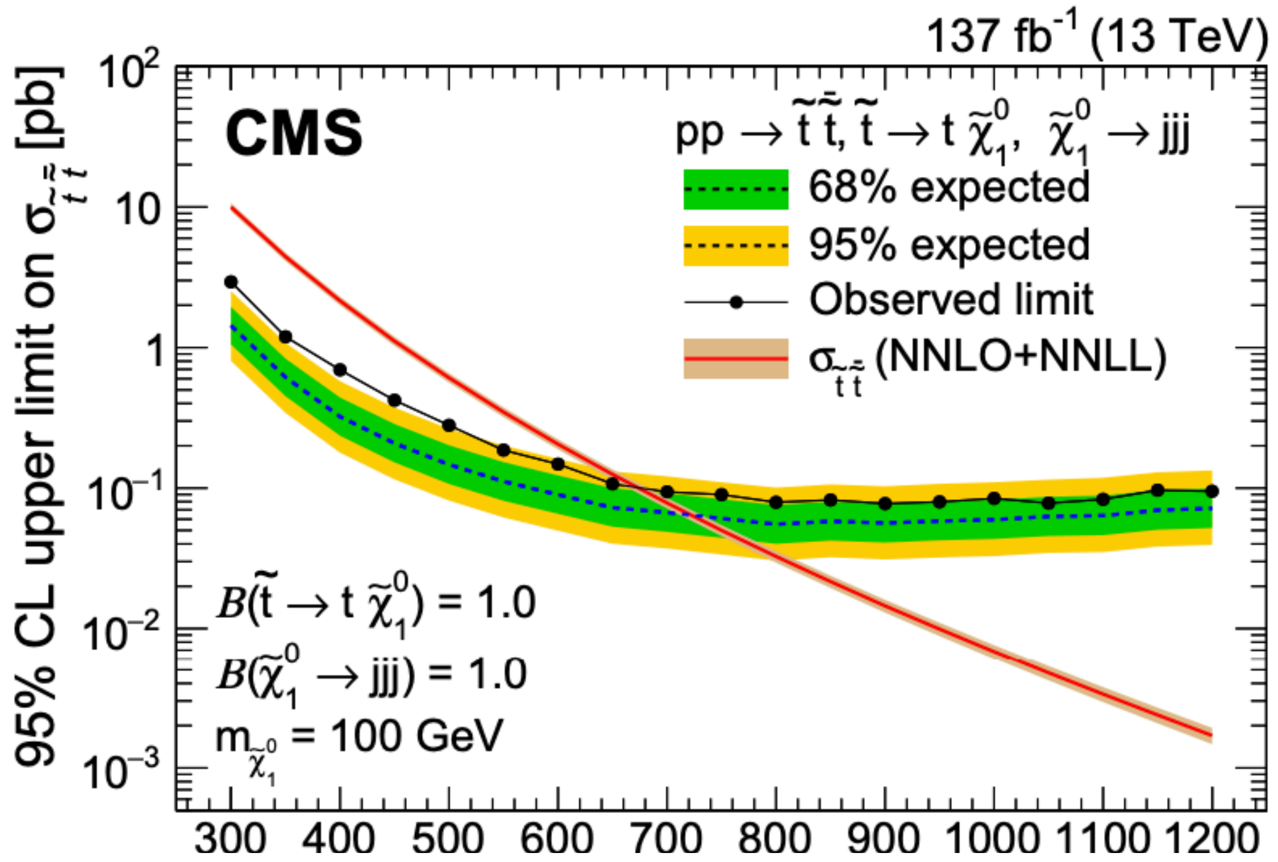
Control discriminant shape to be **invariant** through loss engineering

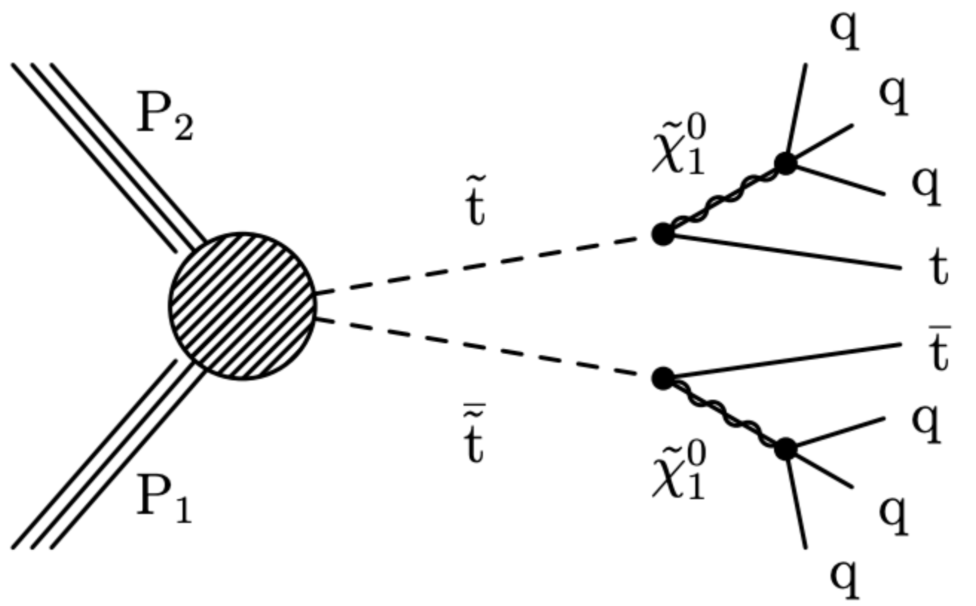
Here: **Gradient Reversal**

NN invariant to jet multiplicity

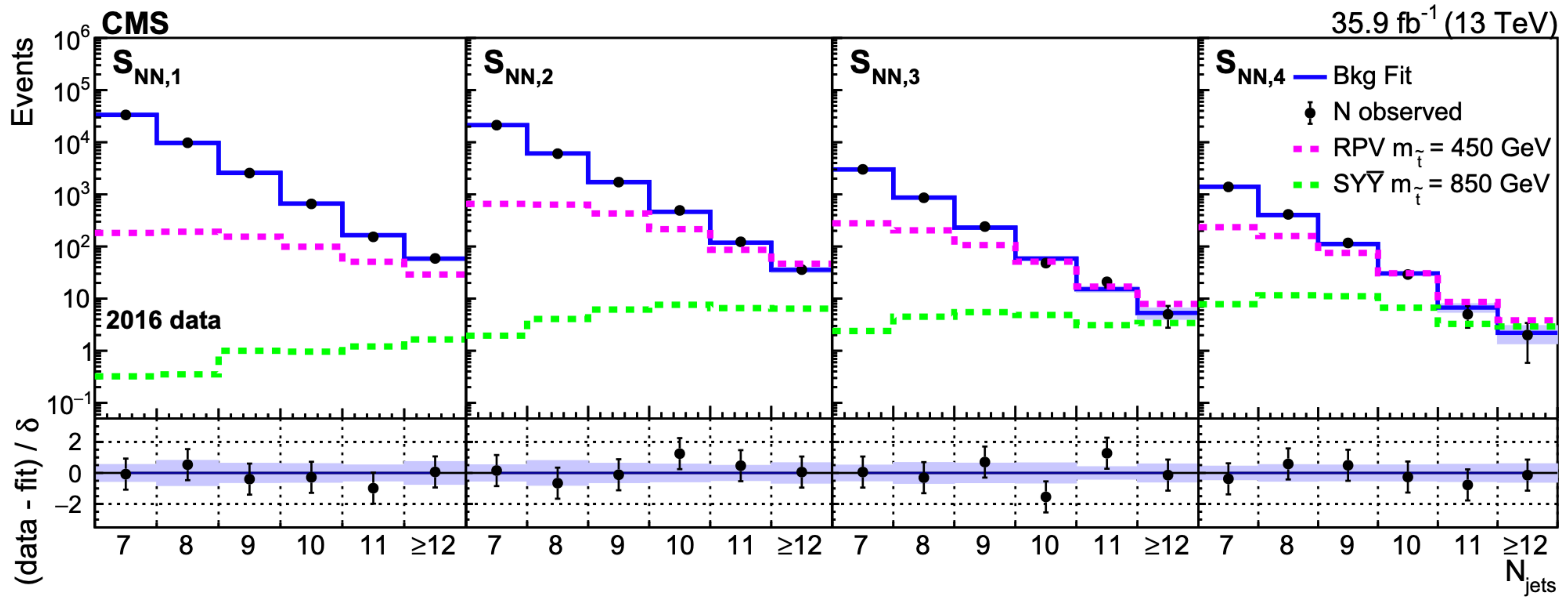


R&D Paper: Feb 2015
→ Deployed in Production: Feb 2021





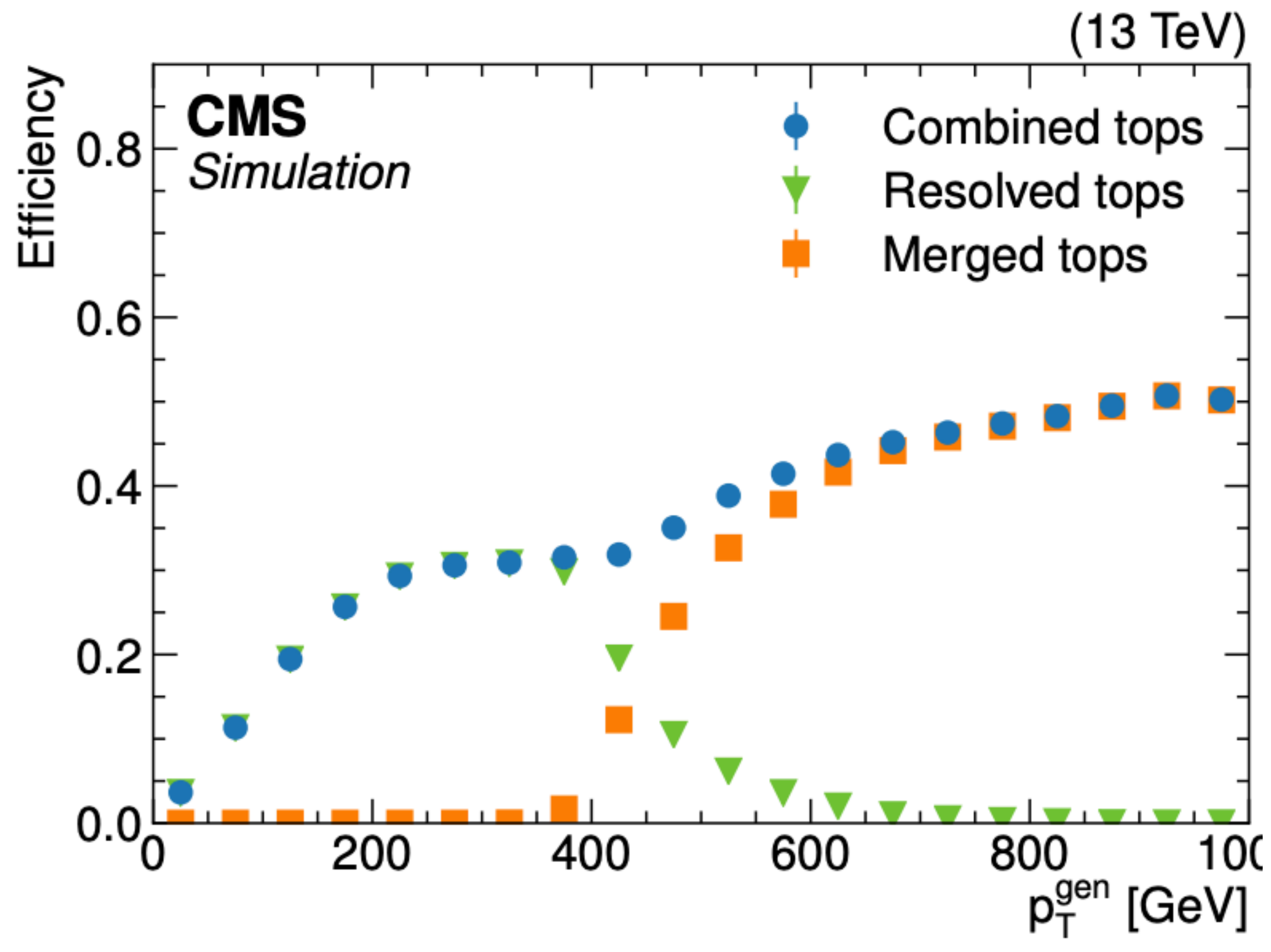
arxiv:2102.06976

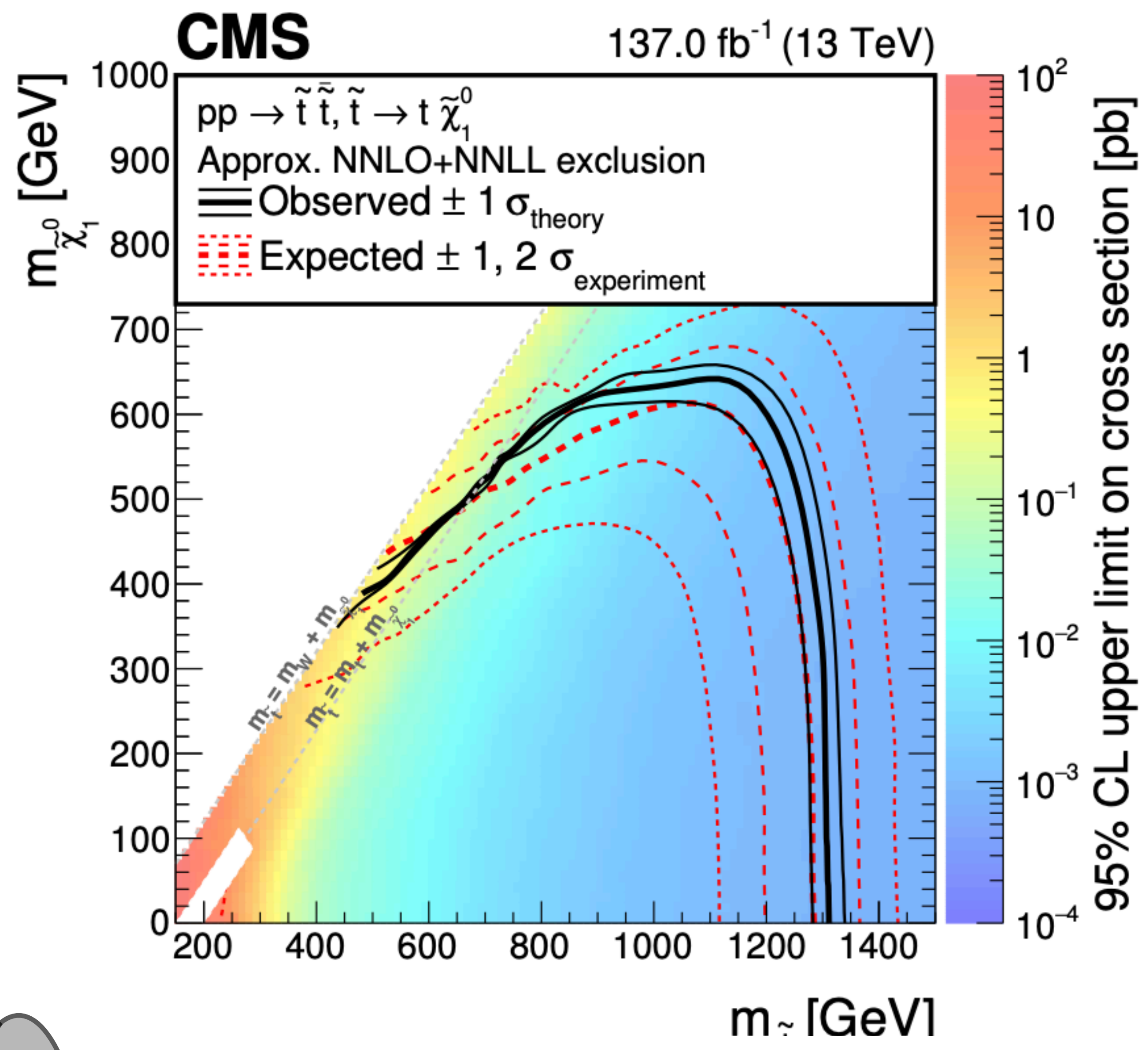


arxiv:2103.01290

System Tagging

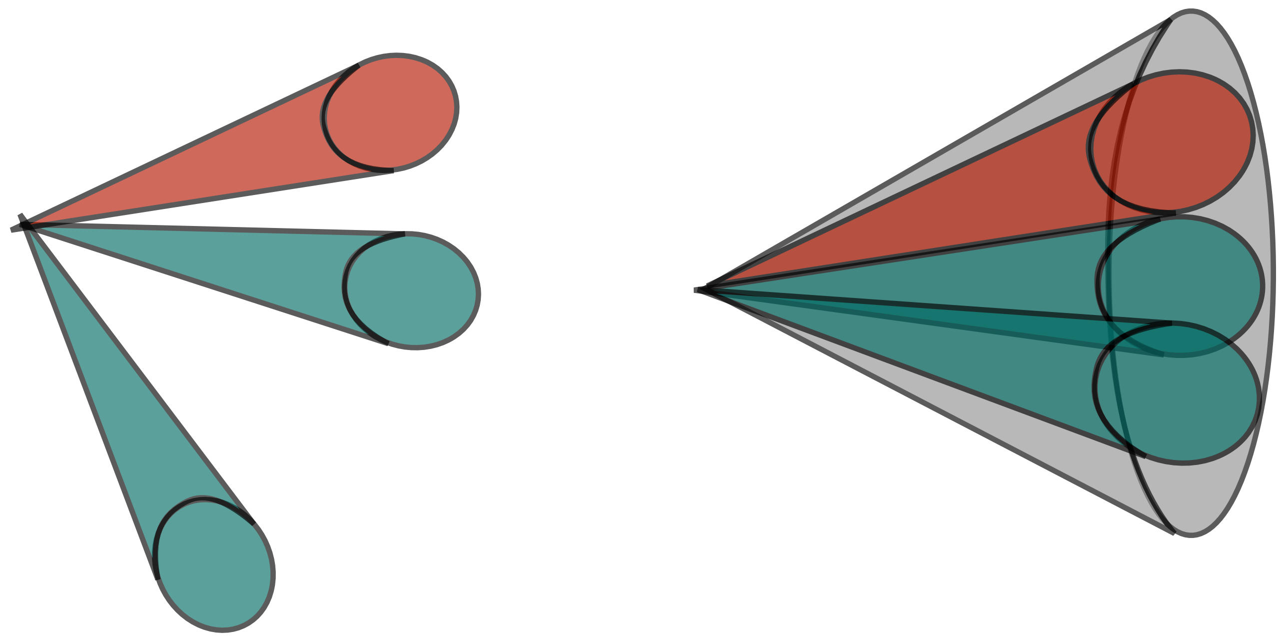
Boosted, hadronic topologies: ideal proving ground for DL





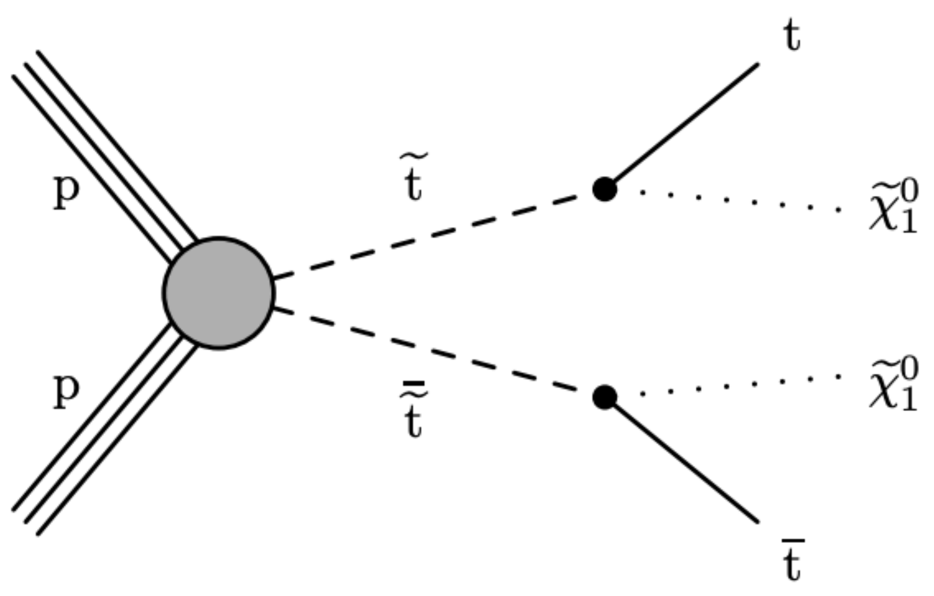
SUSY stop search

Dedicated networks for boosted & resolved multi-jet systems

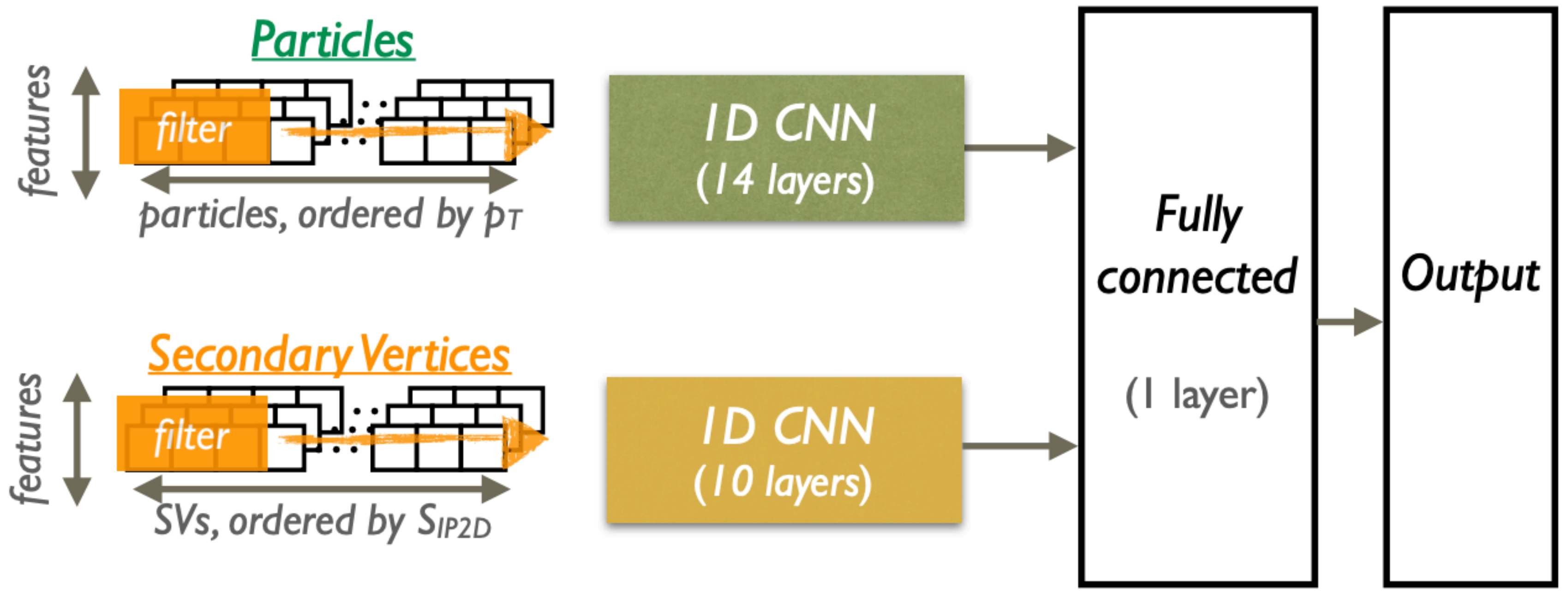


DeepResolved

DeepAK8



DeepAK8-MD: adversarially tuned to be decorrelated with jet mass



Likelihood Ratio Estimation

Approximating Likelihood Ratios with Calibrated Discriminative Classifiers

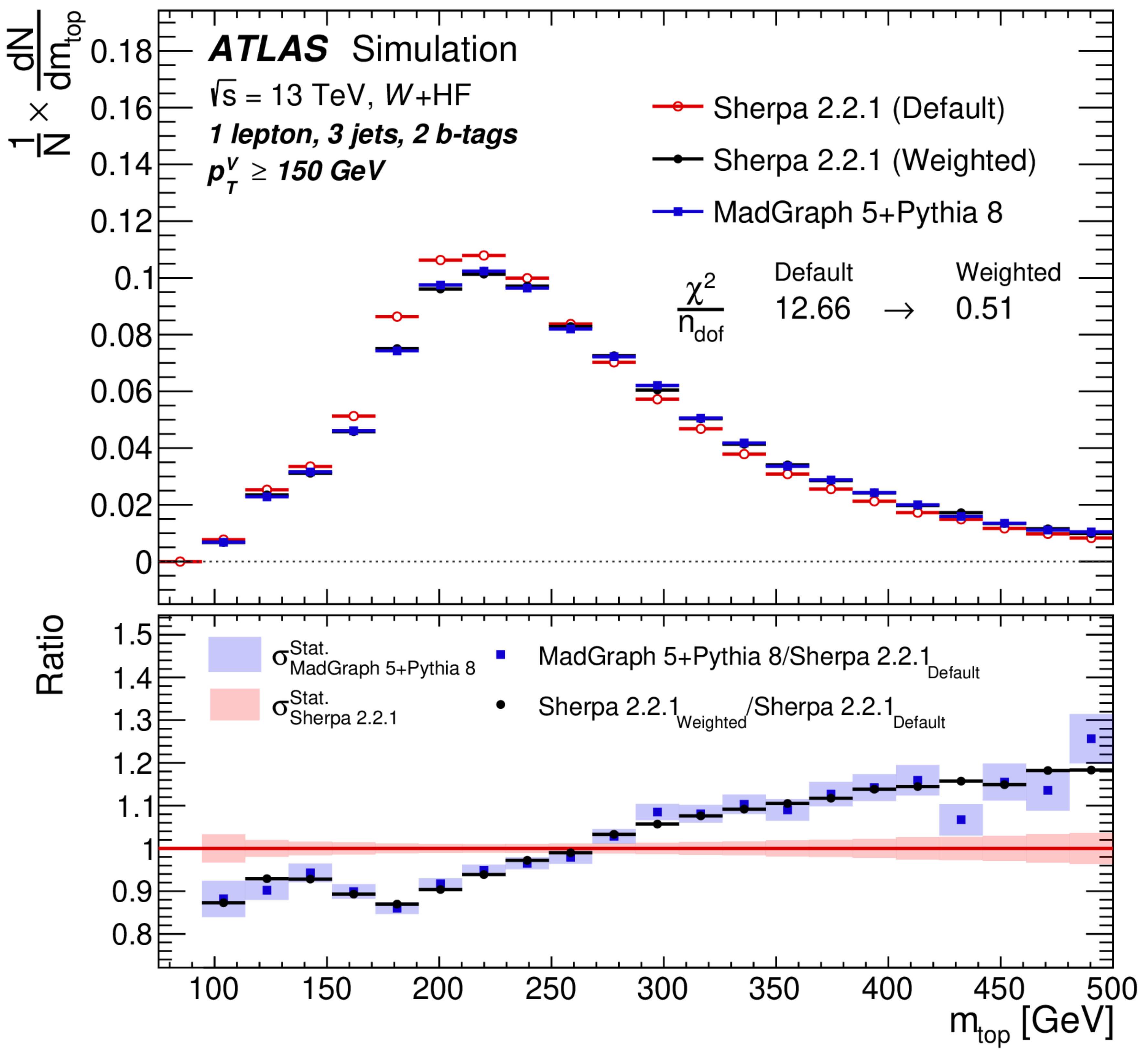
Kyle Cranmer[1], Juan Pavez[2], and Gilles Louppe[1]
[1]New York University
[2]Federico Santa María University

arXiv:1506.02169

Classifiers (e.g. NNs, BDTs) trained to separate two populations

are efficient per-event **L'hood ratio estimators**

- powerful for high-dimensional reweighting



Key for leading shape uncertainties in VH channel with H→bb

- train classifier on Sherpa vs Madgraph
- can reweight nominal (high-stats) samples using classifier for data-efficient but precise systematics modelling

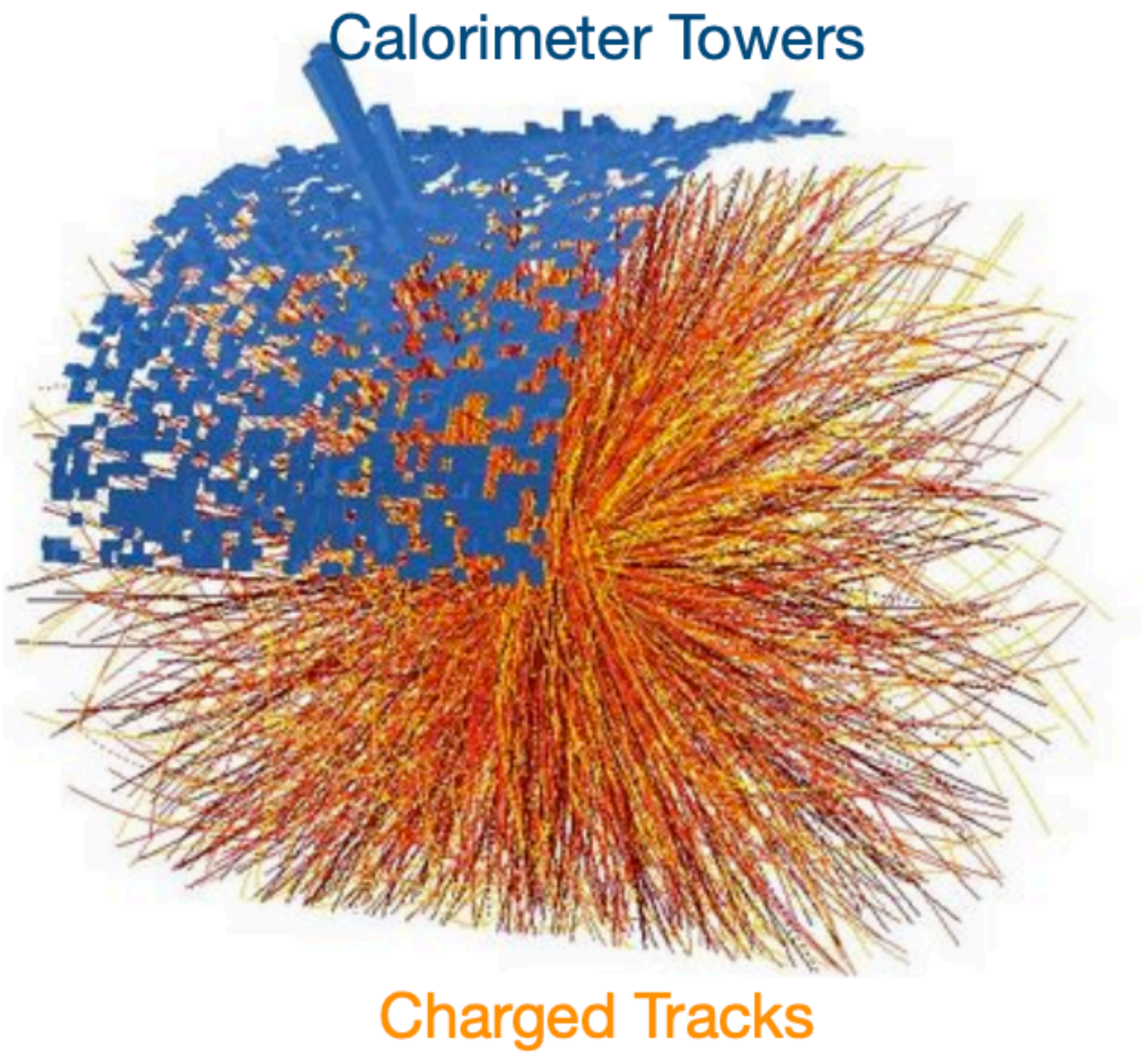
R&D Paper: Nov 2015 → Deployed for V(H→bb): July 2020
(similar technique used in e.g. arXiv:1806.04030)

Jet Reconstruction

Jet p_{T} reconstruction challenging in high-density environment of ALICE

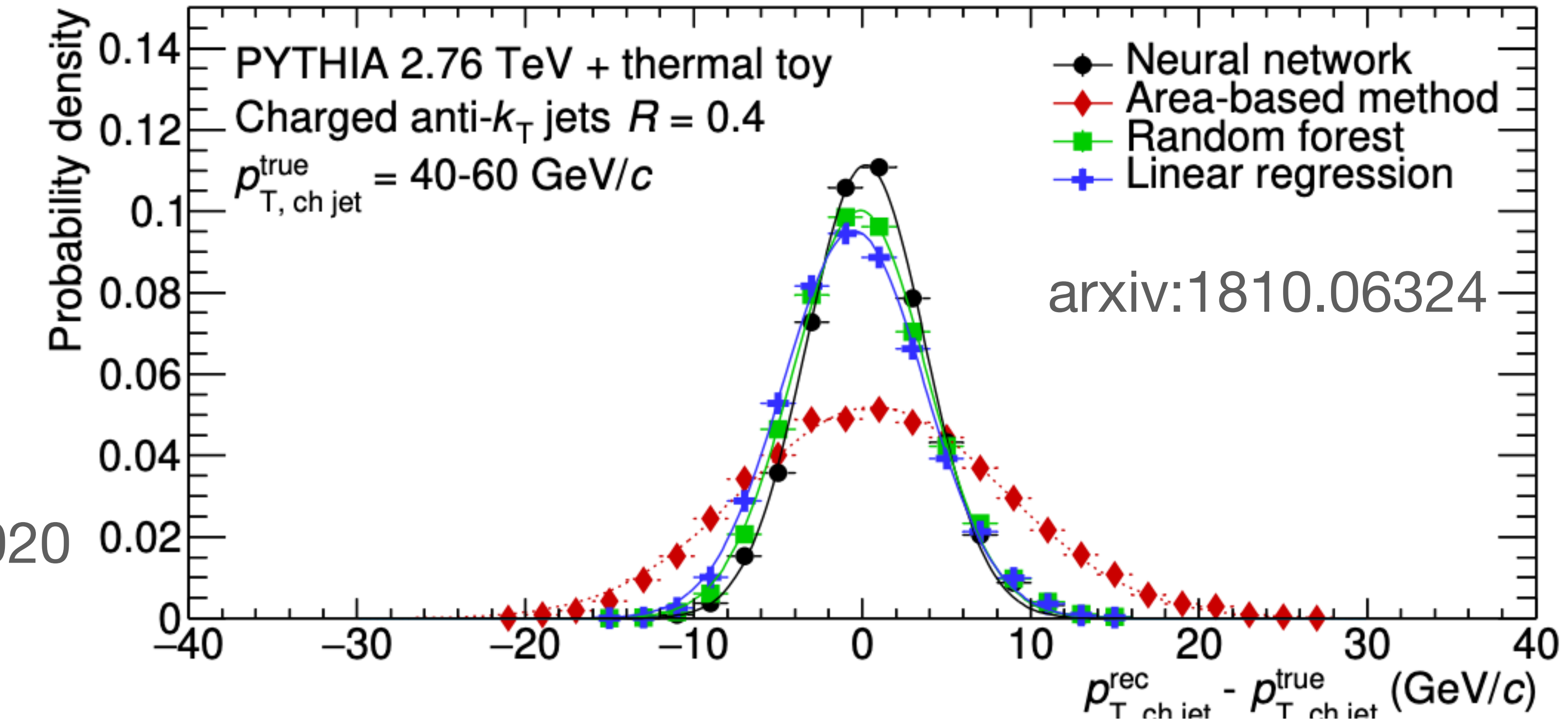
- traditional method: pedestal-subtraction with event-global background estimate
  → limited at low p_{T} & large radii

$$p_{\mathrm{T}}^{\mathrm{rec}} = p_{\mathrm{T}}^{\mathrm{raw}} - \rho A$$



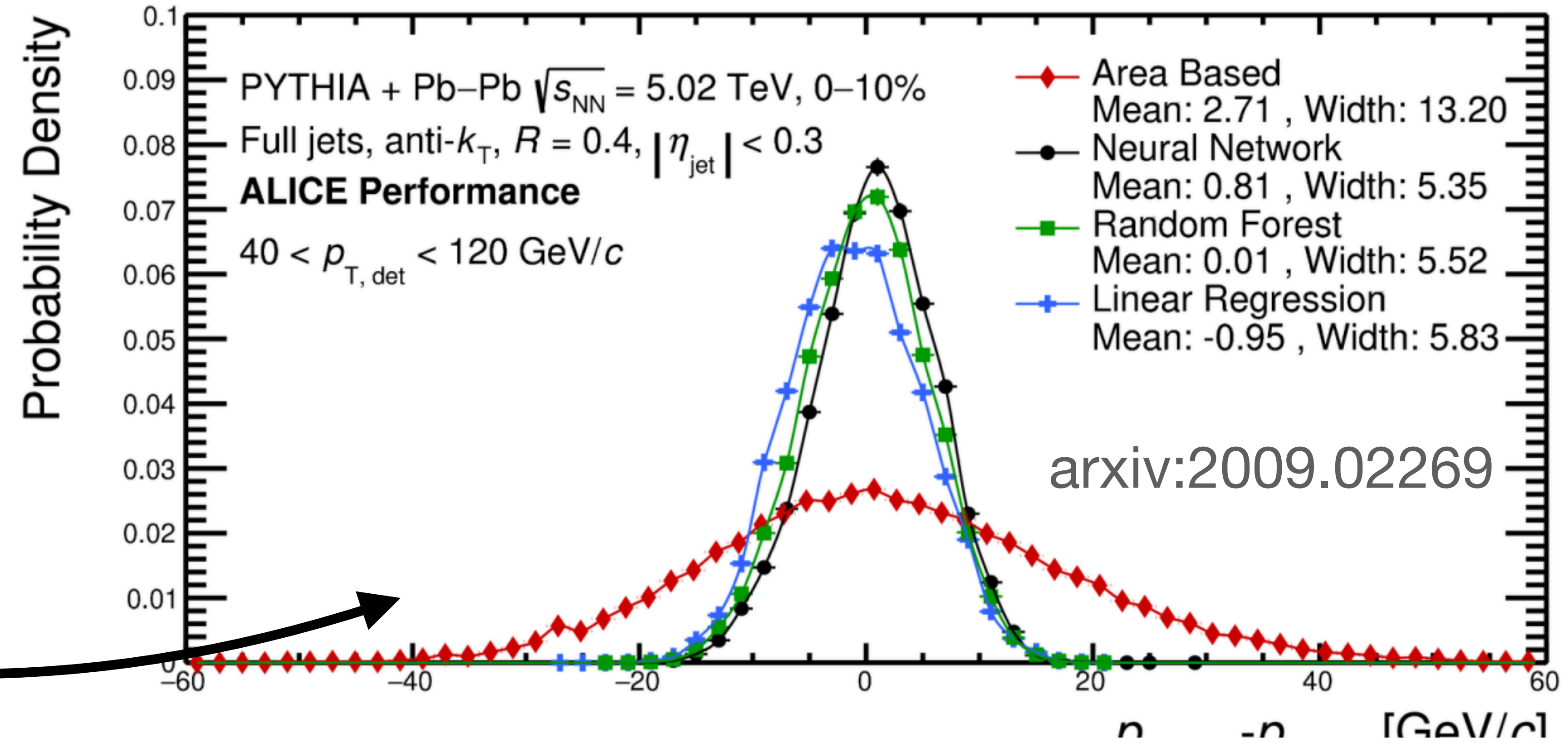
Jet Reconstruction

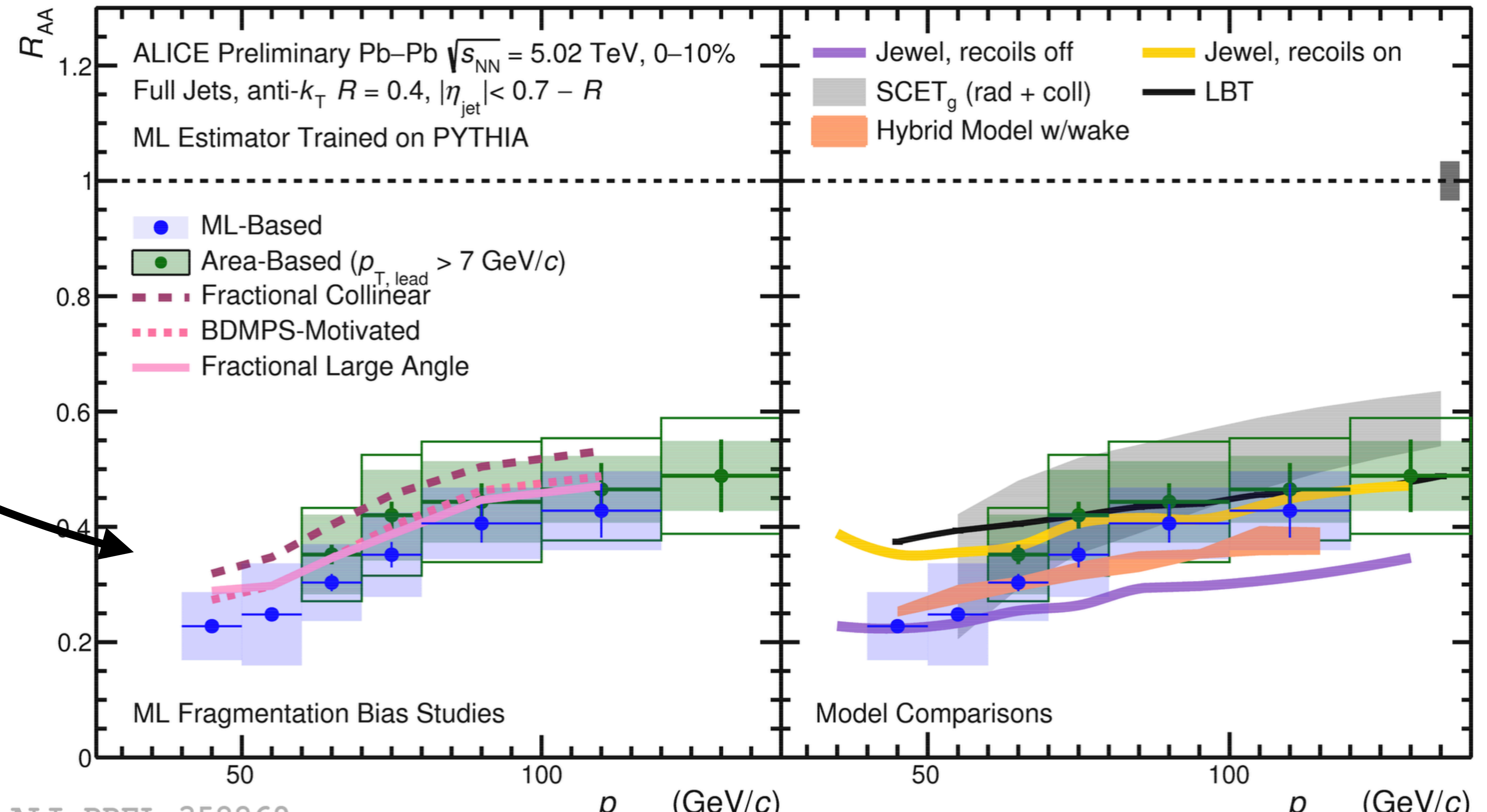
R&D Paper: Oct 2018
→ Deployed in Production: Sep 2020



New ML method:

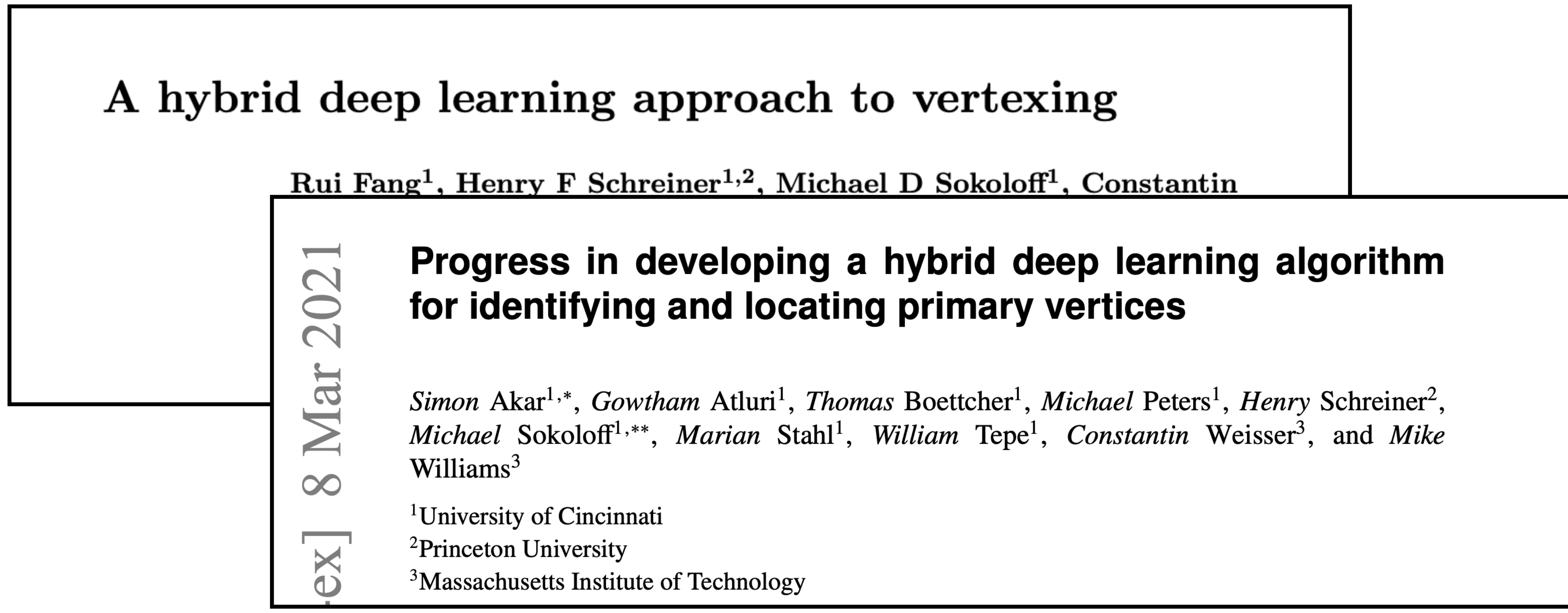
- neural net for jet-by-jet p_{T} correction (based on constituent, jet and event inputs)
- much more precise measurements
- enables measurement of previously out-of-reach phase-space: jet p_{T} as low as 40 GeV





$$p_{\mathrm{T}}^{\mathrm{rec}} = \mathrm{NN}(p_{\mathrm{T}}^{\mathrm{raw}}, p_{\mathrm{T}}^{\mathrm{raw}} - \rho A, n_{\mathrm{const}}, \text{jet shape}, \{p_{\mathrm{T,i}}\}, \ldots)$$

Primary Vertex Finding


A hybrid deep learning approach to vertexing

Rui Fang[1], Henry F Schreiner[1,2], Michael D Sokoloff[1], Constantin

[...ex] 8 Mar 2021

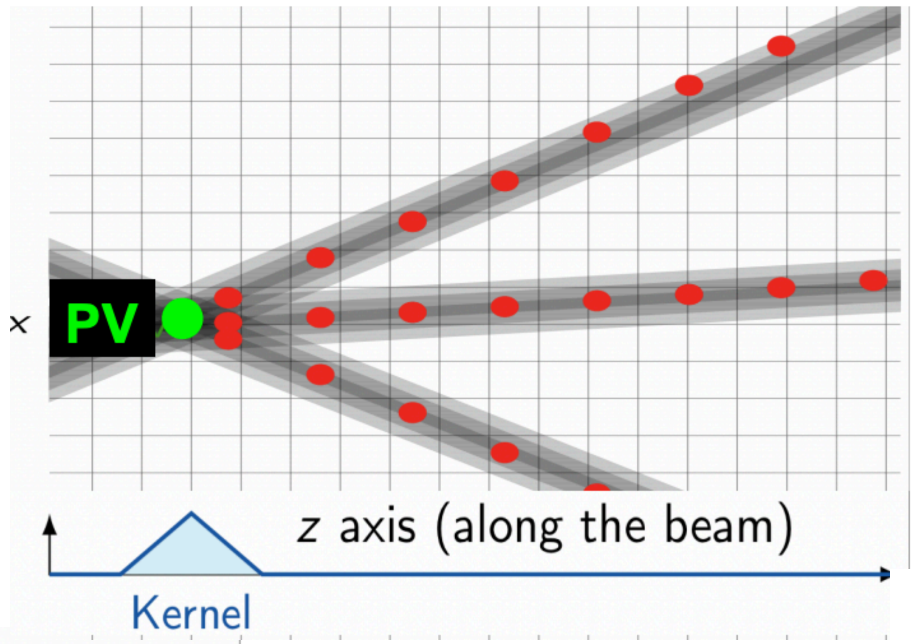
Progress in developing a hybrid deep learning algorithm for identifying and locating primary vertices

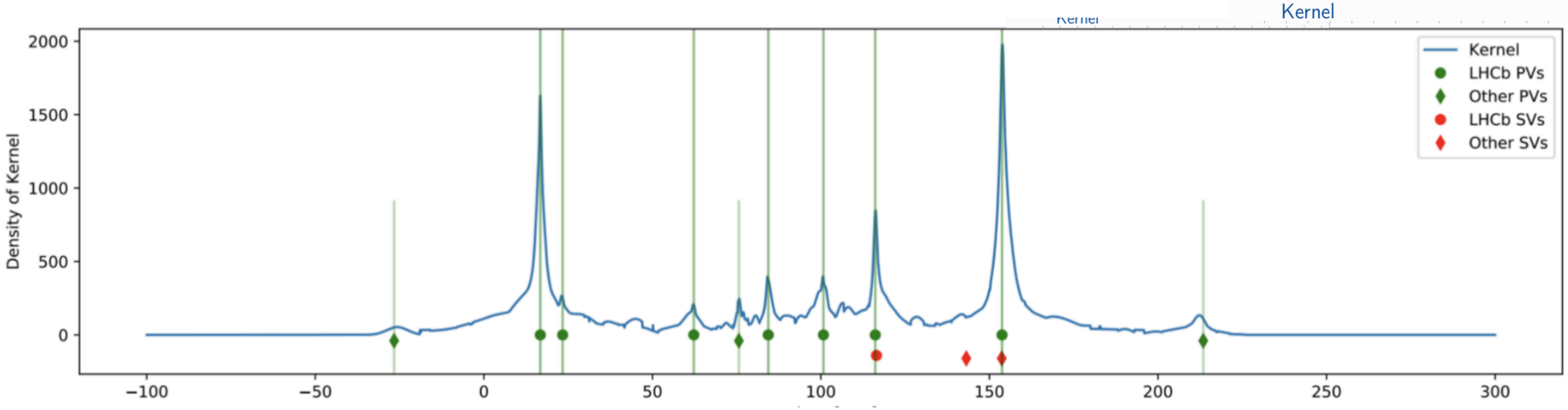
Simon Akar[1,*], *Gowtham* Atluri[1], *Thomas* Boettcher[1], *Michael* Peters[1], *Henry* Schreiner[2], *Michael* Sokoloff[1,**], *Marian* Stahl[1], *William* Tepe[1], *Constantin* Weisser[3], and *Mike* Williams[3]

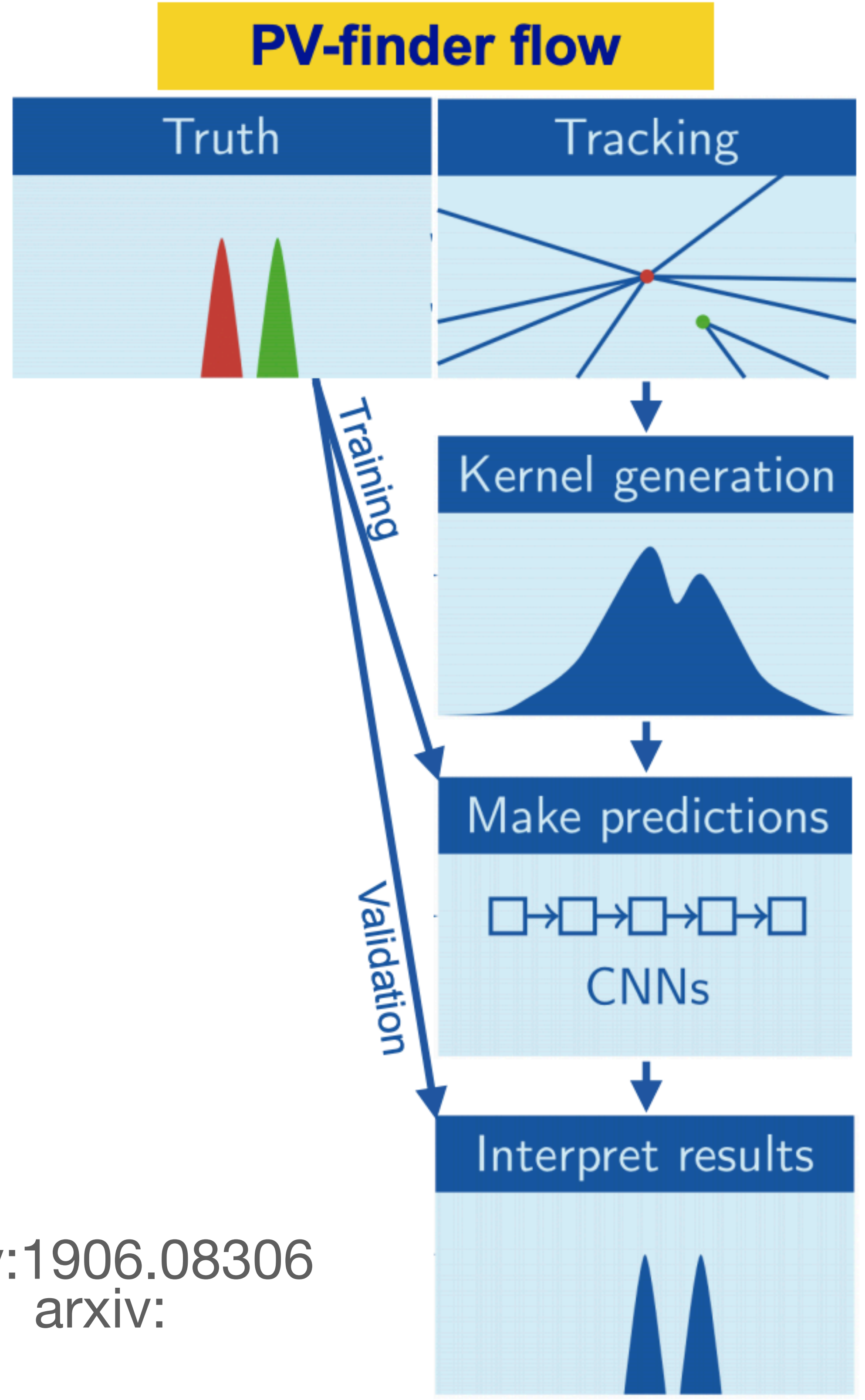
[1]University of Cincinnati
[2]Princeton University
[3]Massachusetts Institute of Technology

Idea:

- project sparse 3D tracking data to 1D distribution via kernel density estimation (KDE)
- Deep Learning (CNN) to process peaky distribution to find primary vertiecs

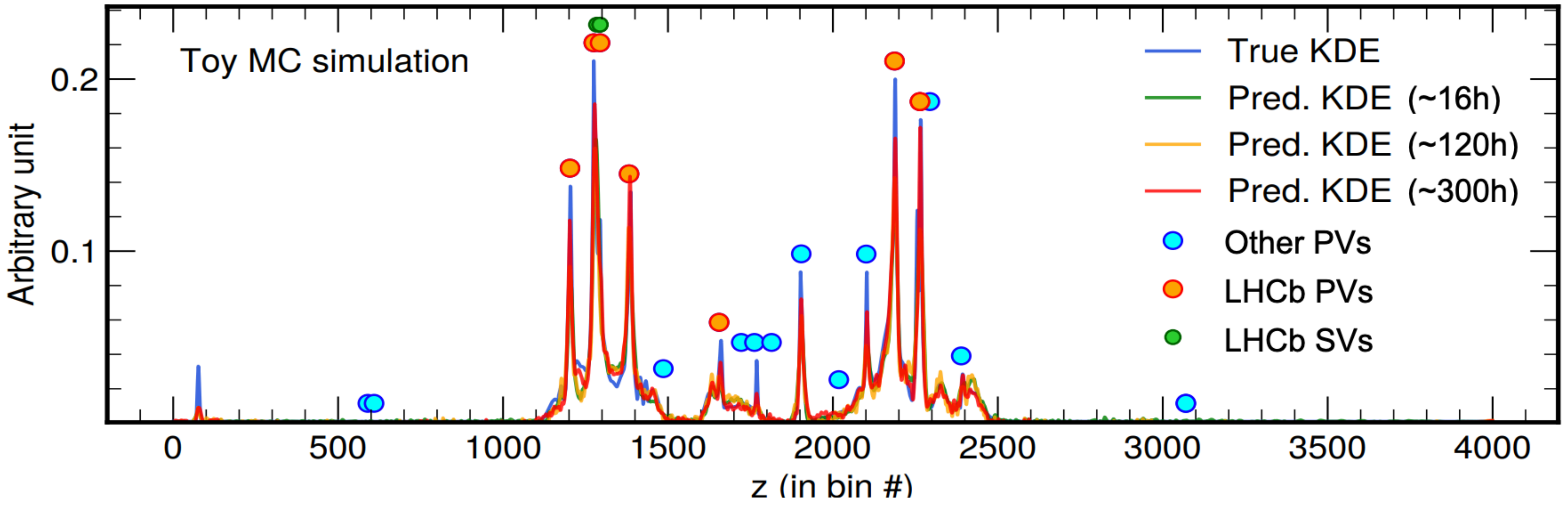






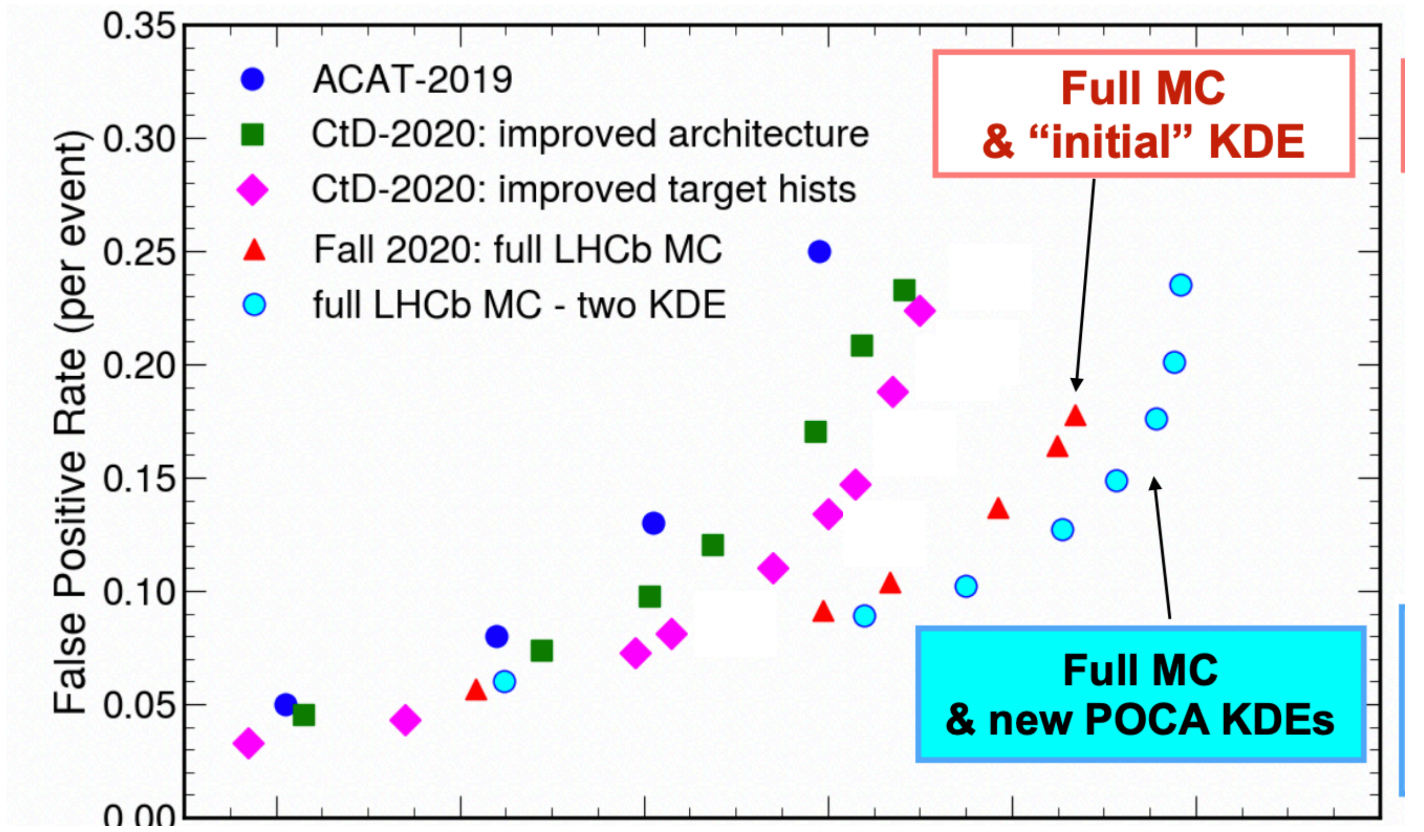
arxiv:1906.08306
arxiv:

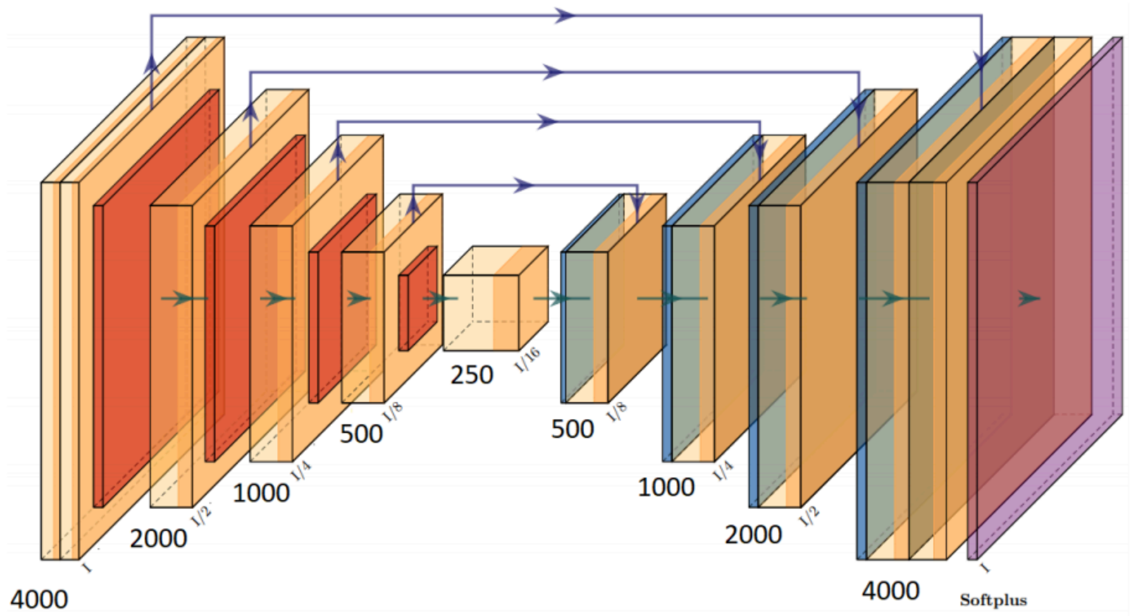
Primary Vertex Finding



Moving towards Production:

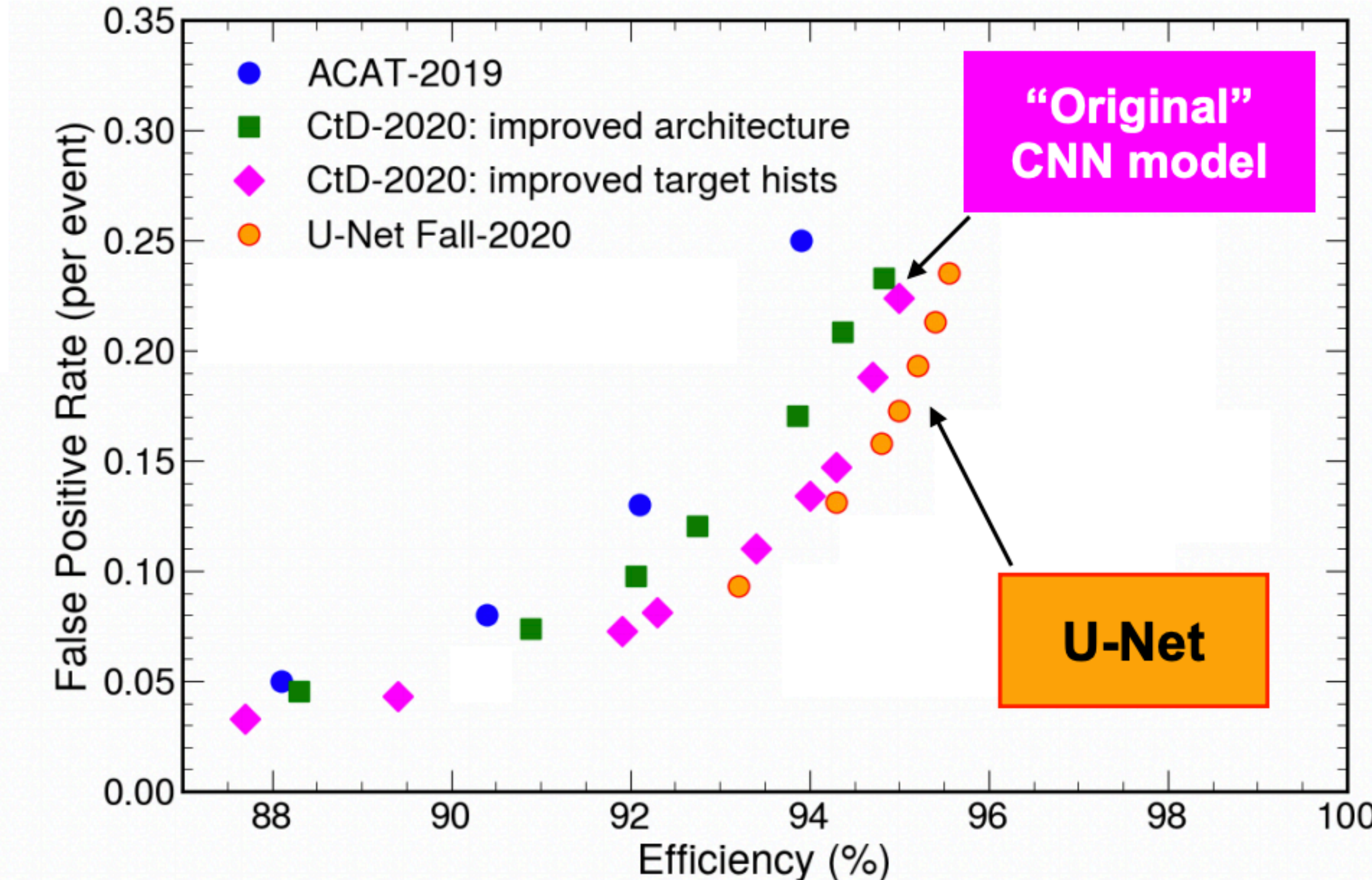
- Replace expensive KDE building with **two DNN-based KDE predictions** from track parameters
- method improved by new KDEs & moving to full LHCb simulation





Ongoing R&D

- replace CNNs with U-Net improves performance on Toy MC (expected to transfer to real LHCb MC)



Future Directions I

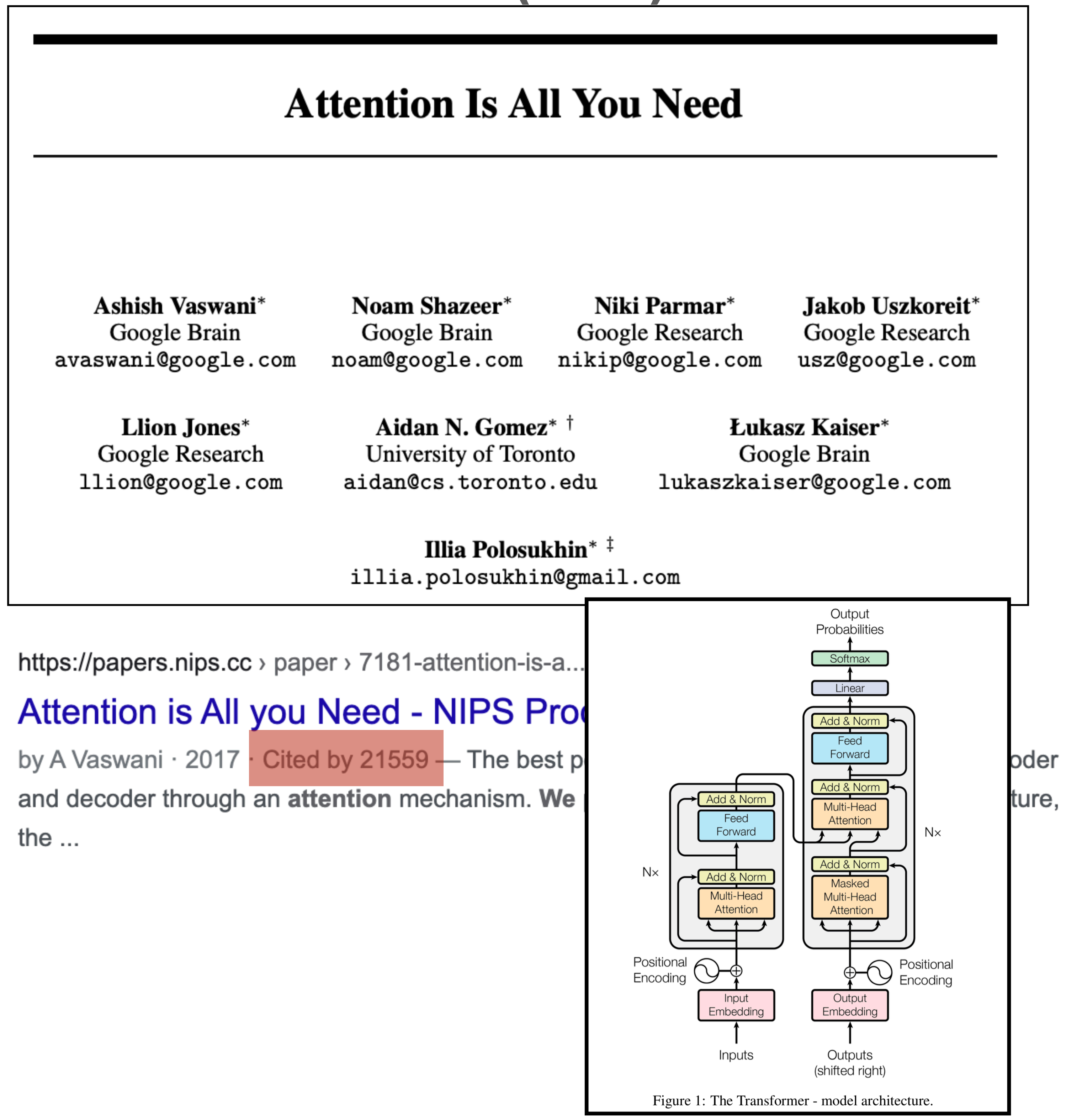
Attention Mechanisms

- interpretable, self-referencing networks
- foundation of recent impressive advances in natural language models

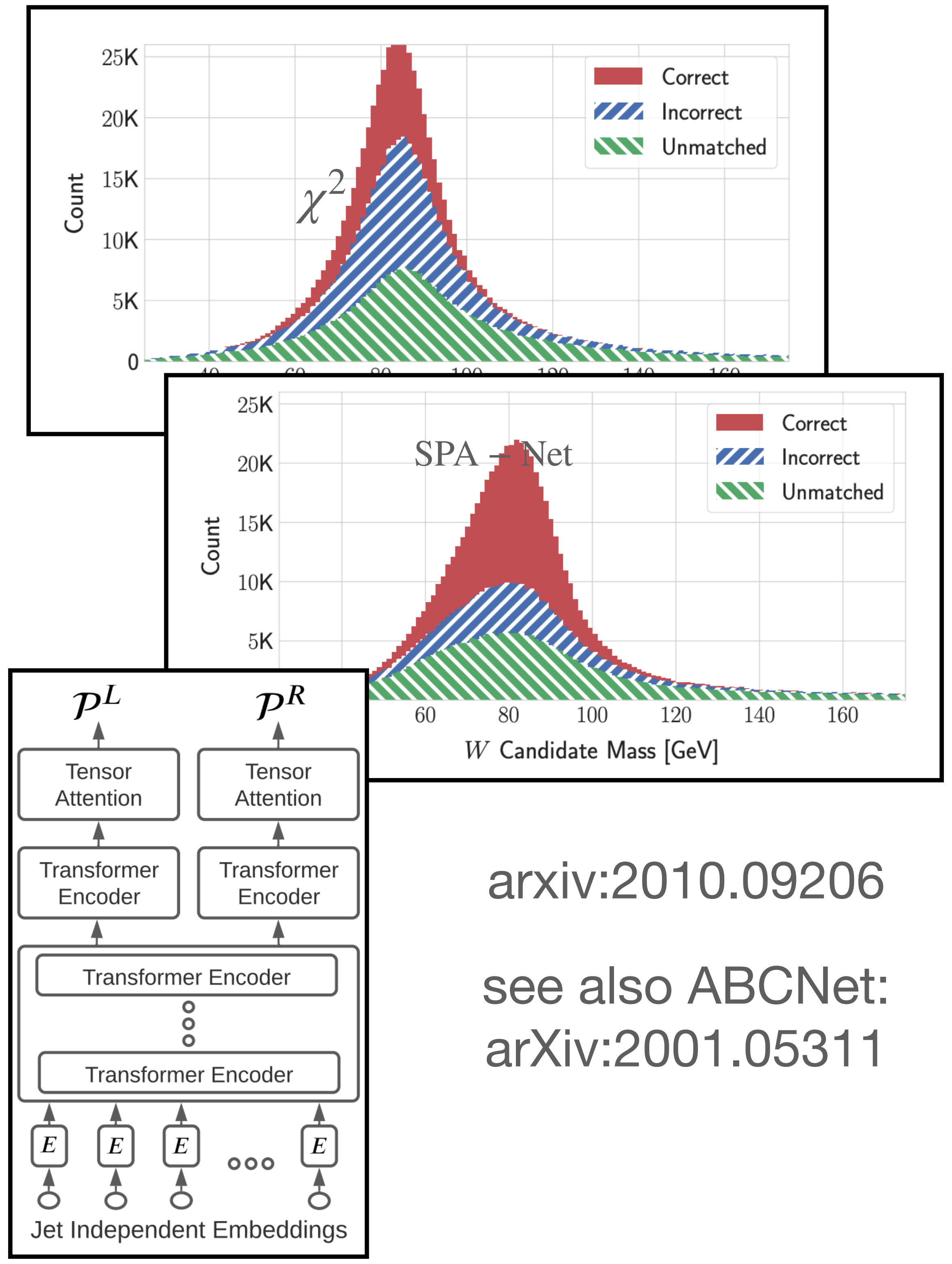
Graph Networks

- natural representation inter-acting variable-size inputs
- heavy R&D in tracking tagging, analysis

Pure ML (2017)

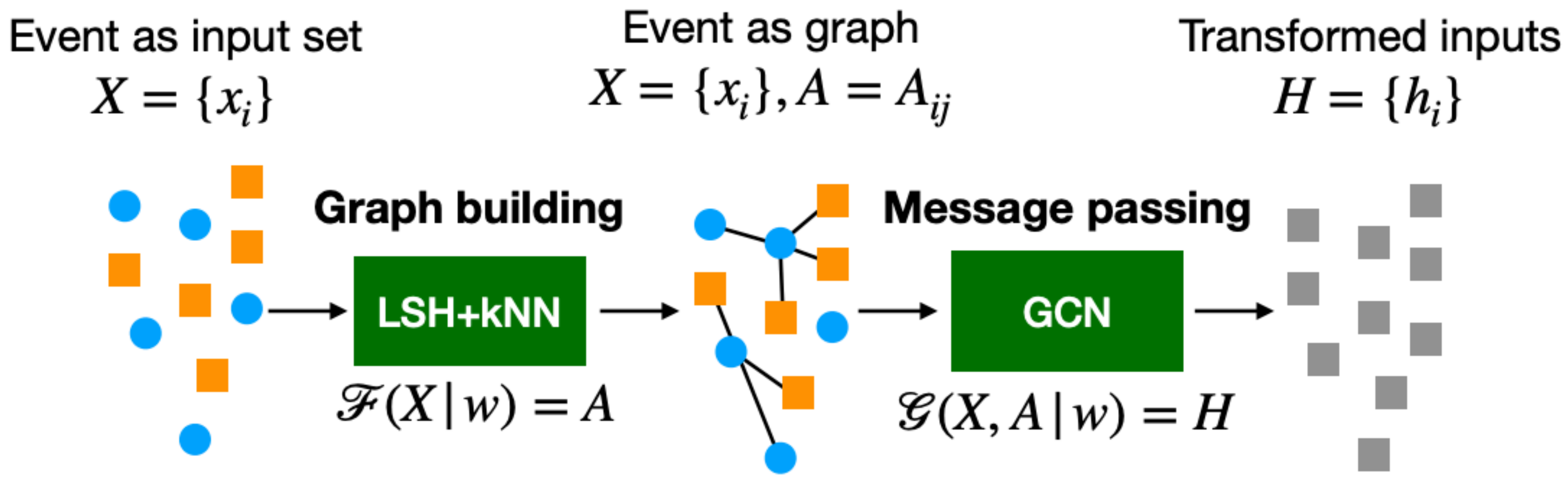


HEPML (2021)



arxiv:2010.09206

see also ABCNet: arXiv:2001.05311

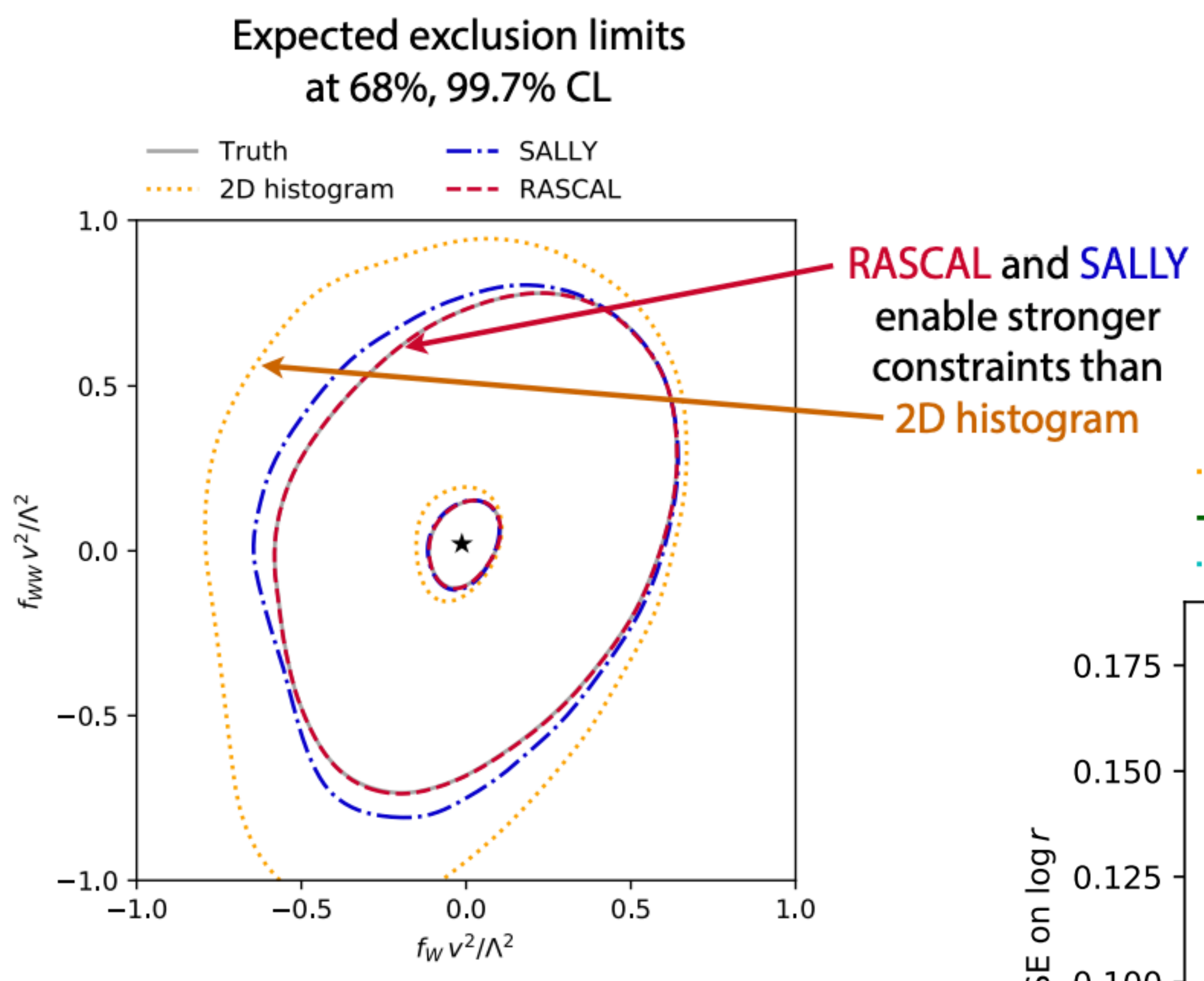


arxiv:2101.08578

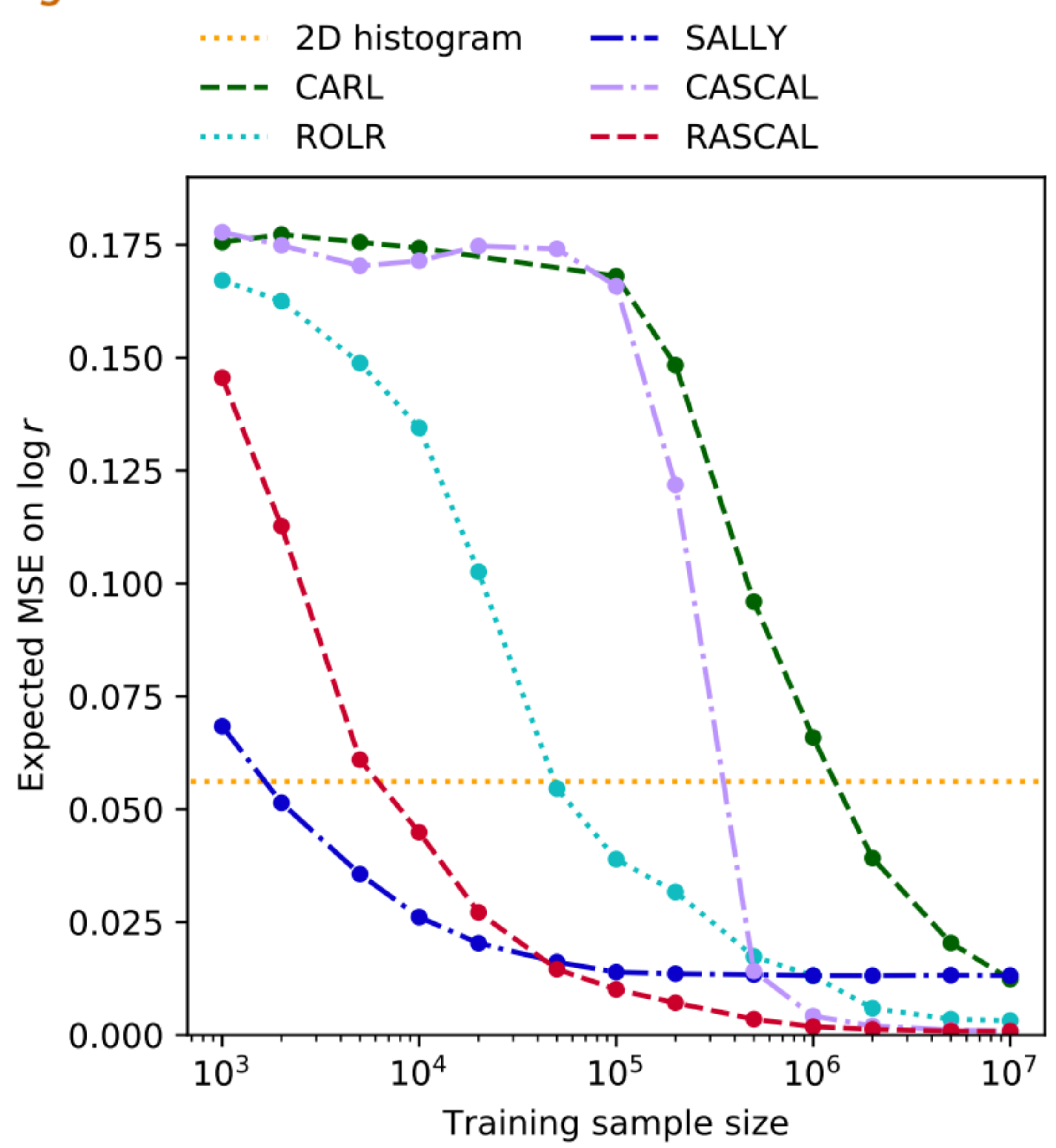
Future Directions II

Simulation-aided L'hood Ratio

- Internal Simulator states can make ratio estimation 100x more data efficient
- general trend of co-designing ML with simulator code

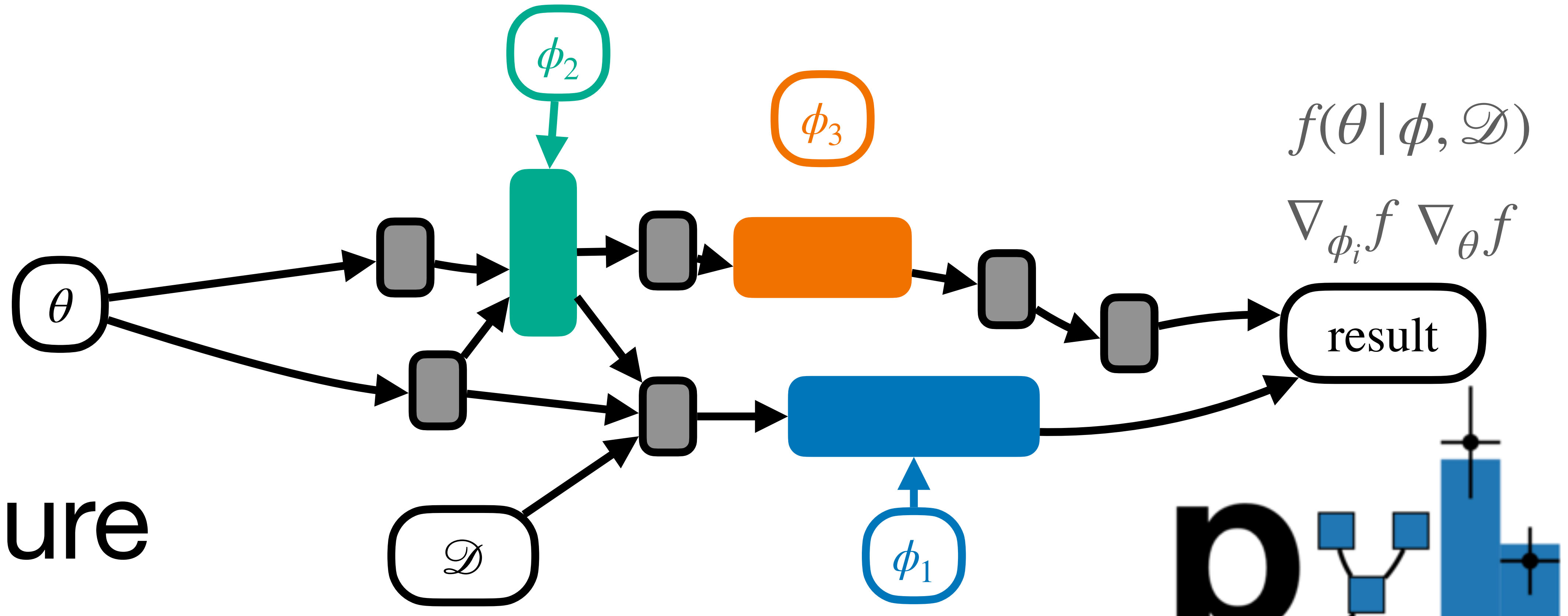


arxiv:1805.00020



Differentiable Programming

- generalization from architecture zoo to new programming paradigm
- mix learnable blocks & physics structure in end-to-end manner



doi:10.21105/joss.02823

Outlook

Deep Learning is (becoming) mainstream in HEP in LHC analyses.
→ deployed at each stage of analysis pipeline

Now focus on going beyond standard ML: **robust representations**, **system classificaton**, **neural reweighting**, **detailed reco**, ...

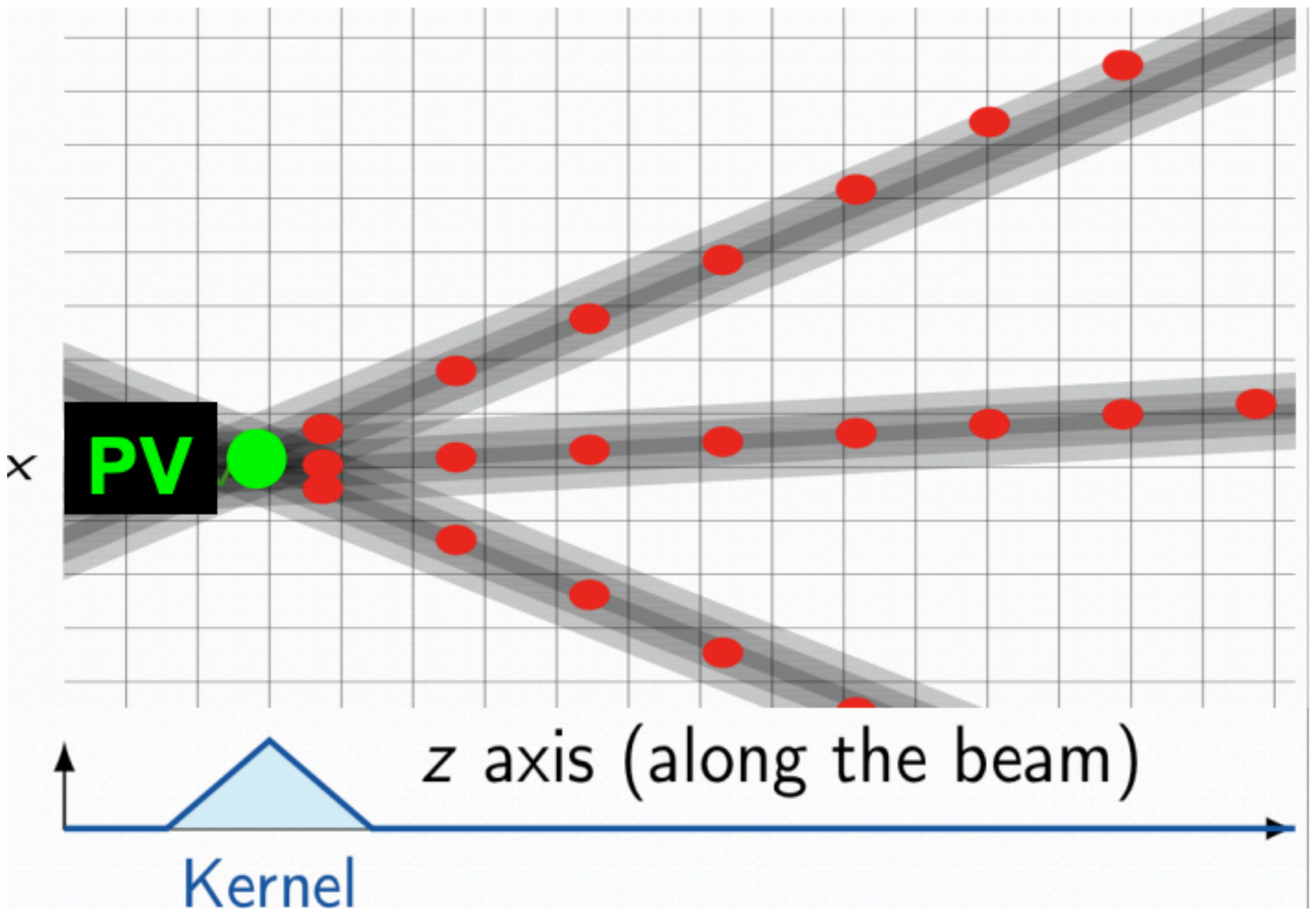
Impressive turnaround from R&D → production at LHC experiments

Many methods in R&D: Attention, Graphs, Simulation-based Inference
→ integrating them pays off, reaches otherwise inaccessible physics
→ **we're only at the beginning of Deep Learning in physics analysis**

Backup

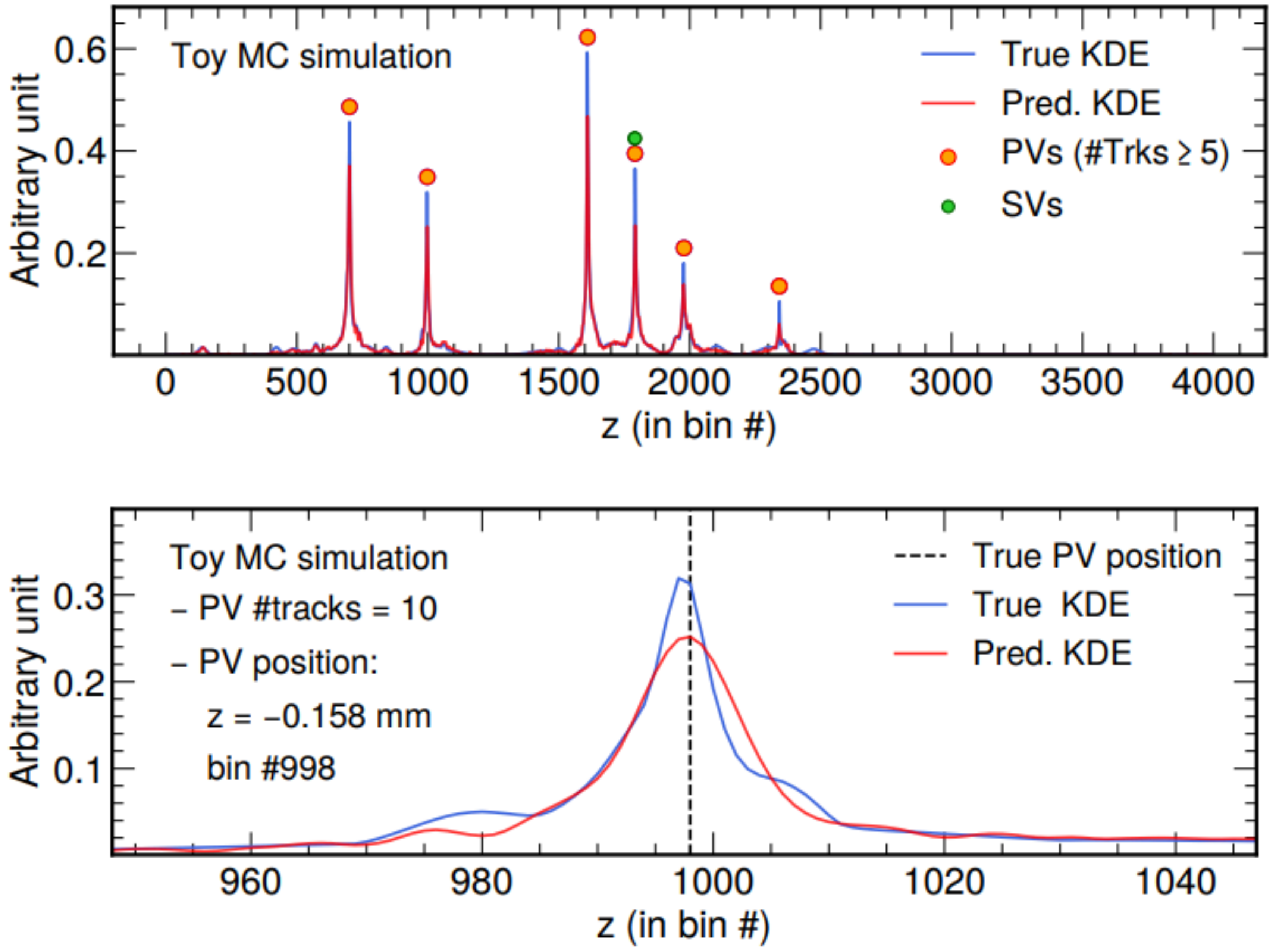
Original LHCb Kernel: $\sum \frac{p^2}{p} - p$

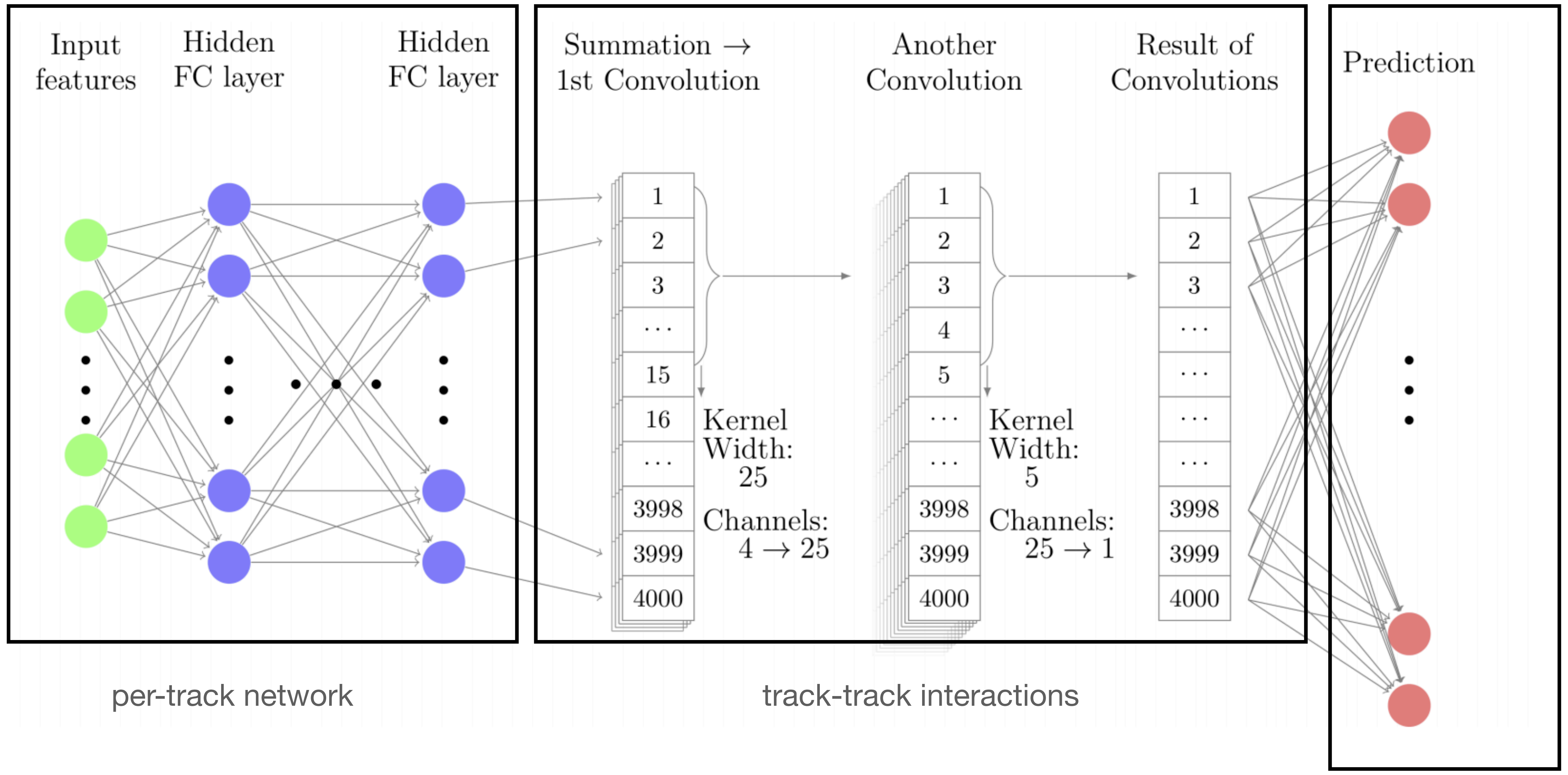
- 1 MINUIT call per z position

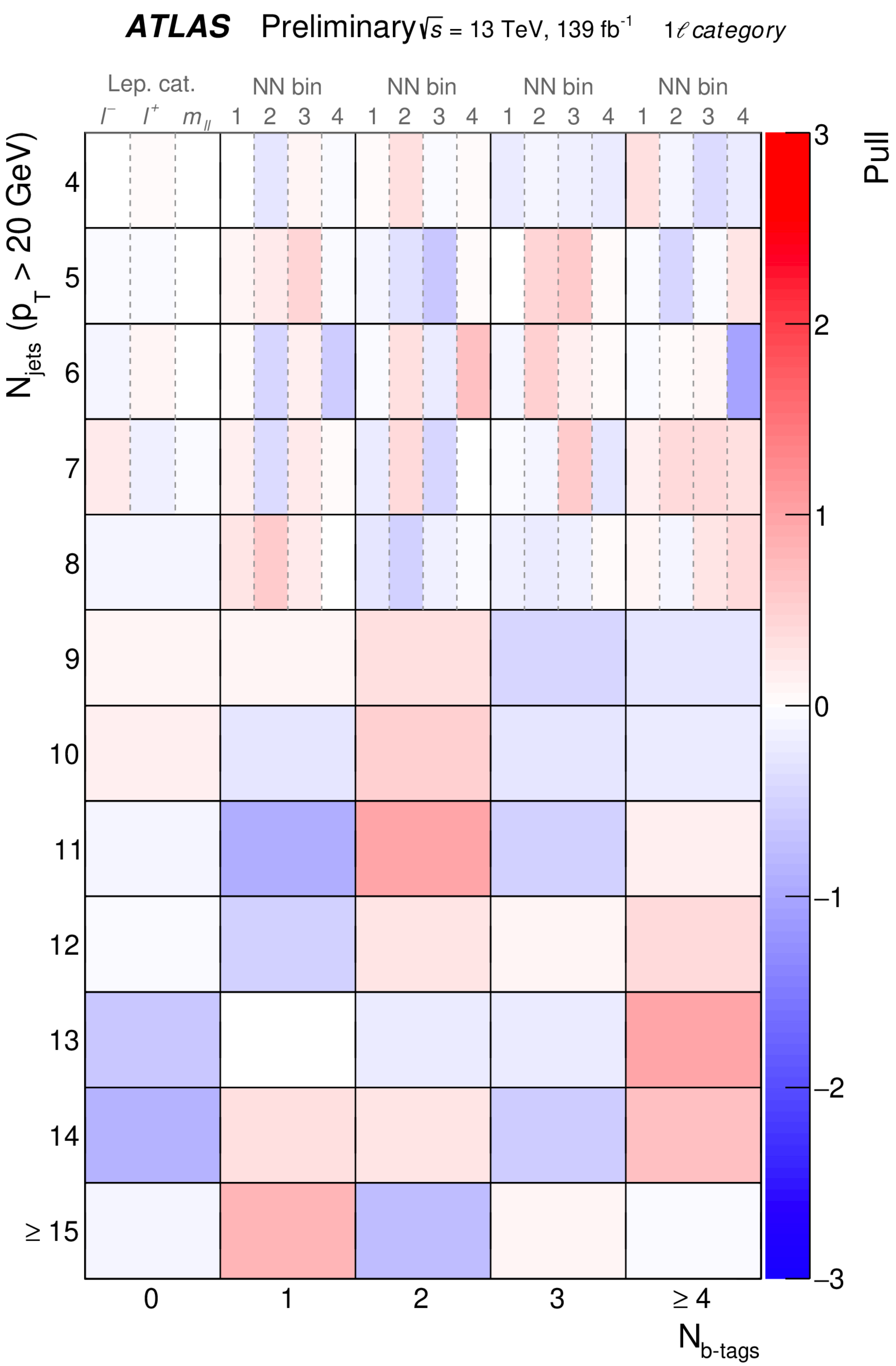


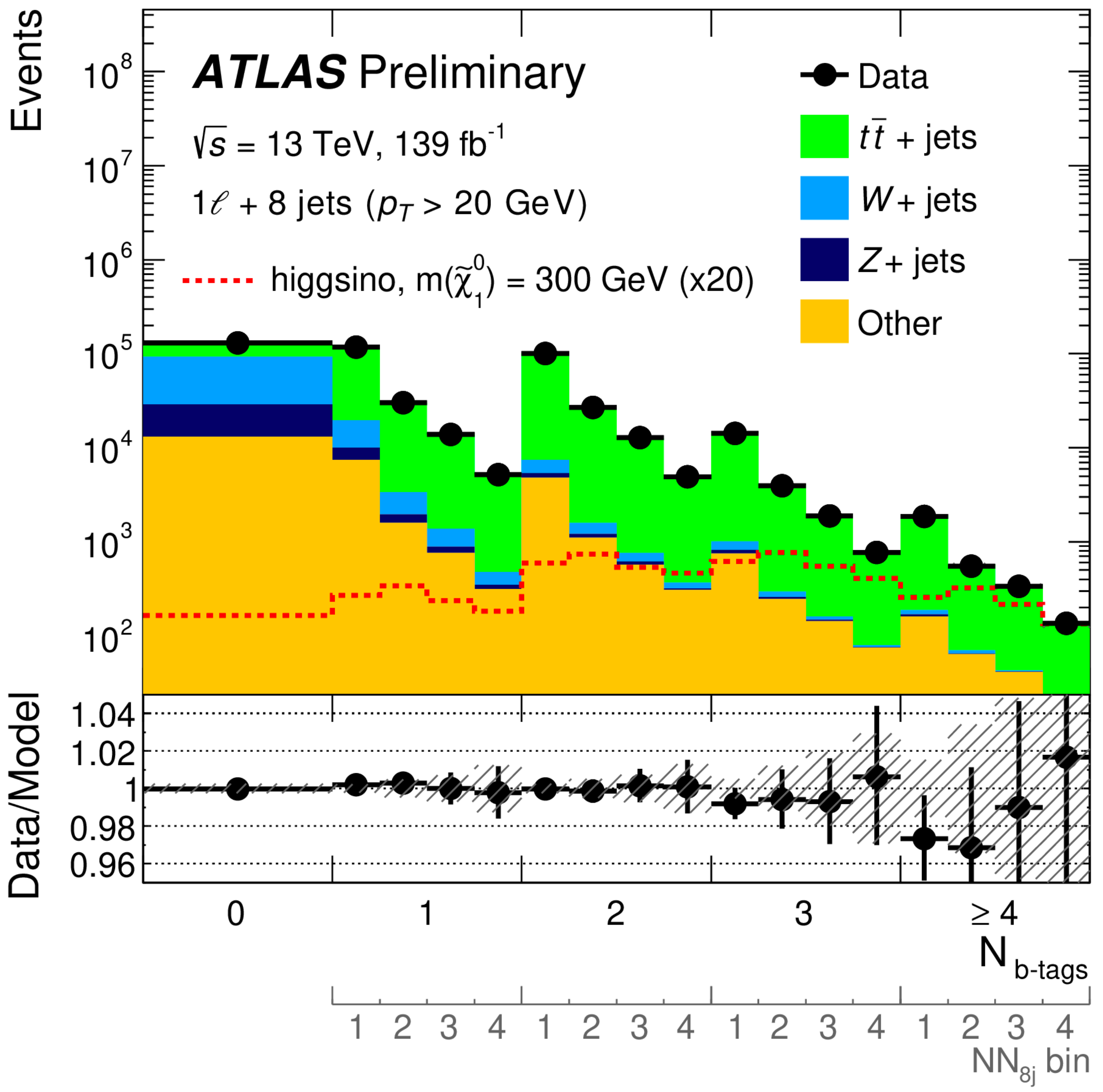
Now two kernels

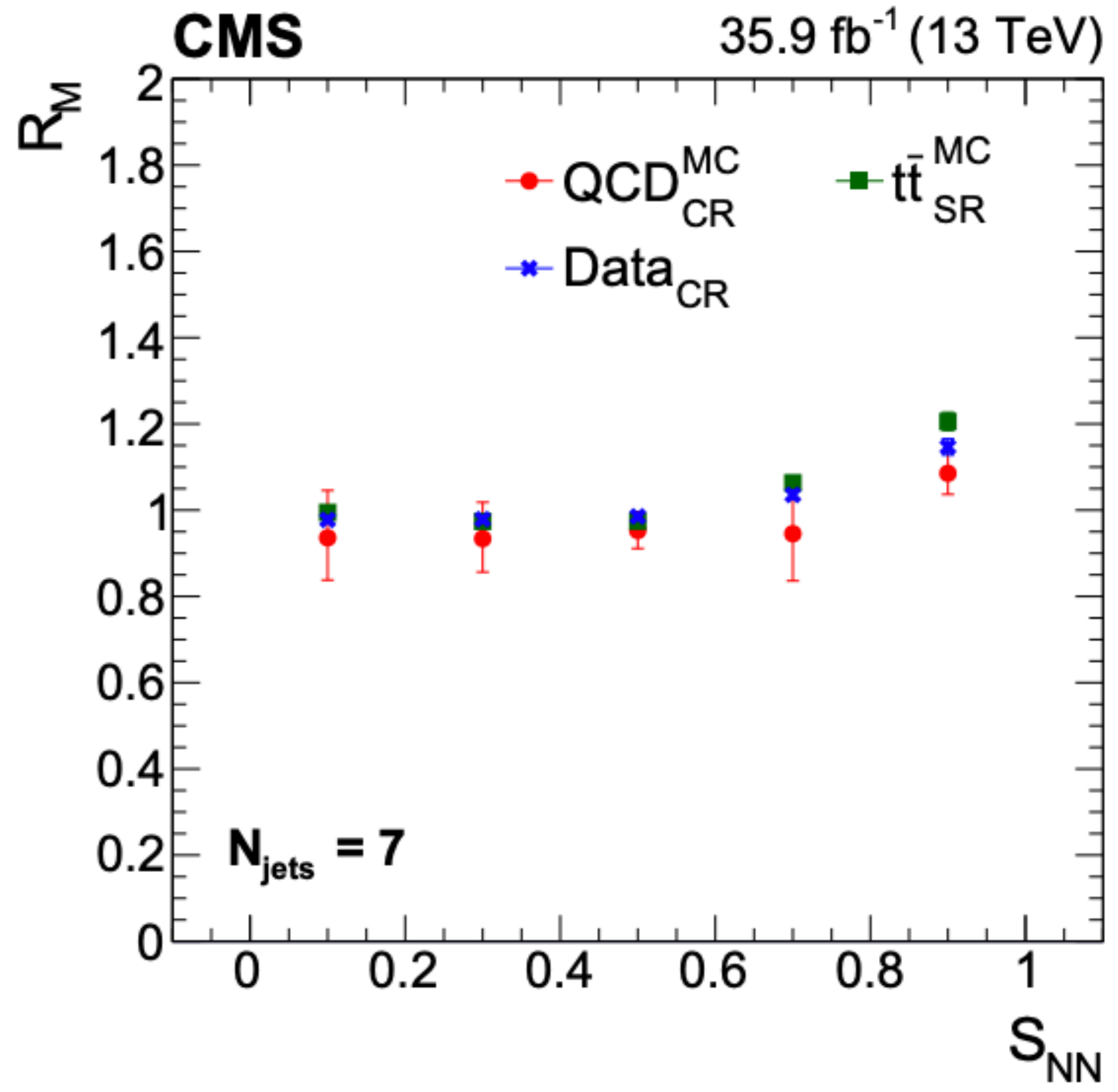
Kernel prediction: $\sum p^2 \quad \sum p$

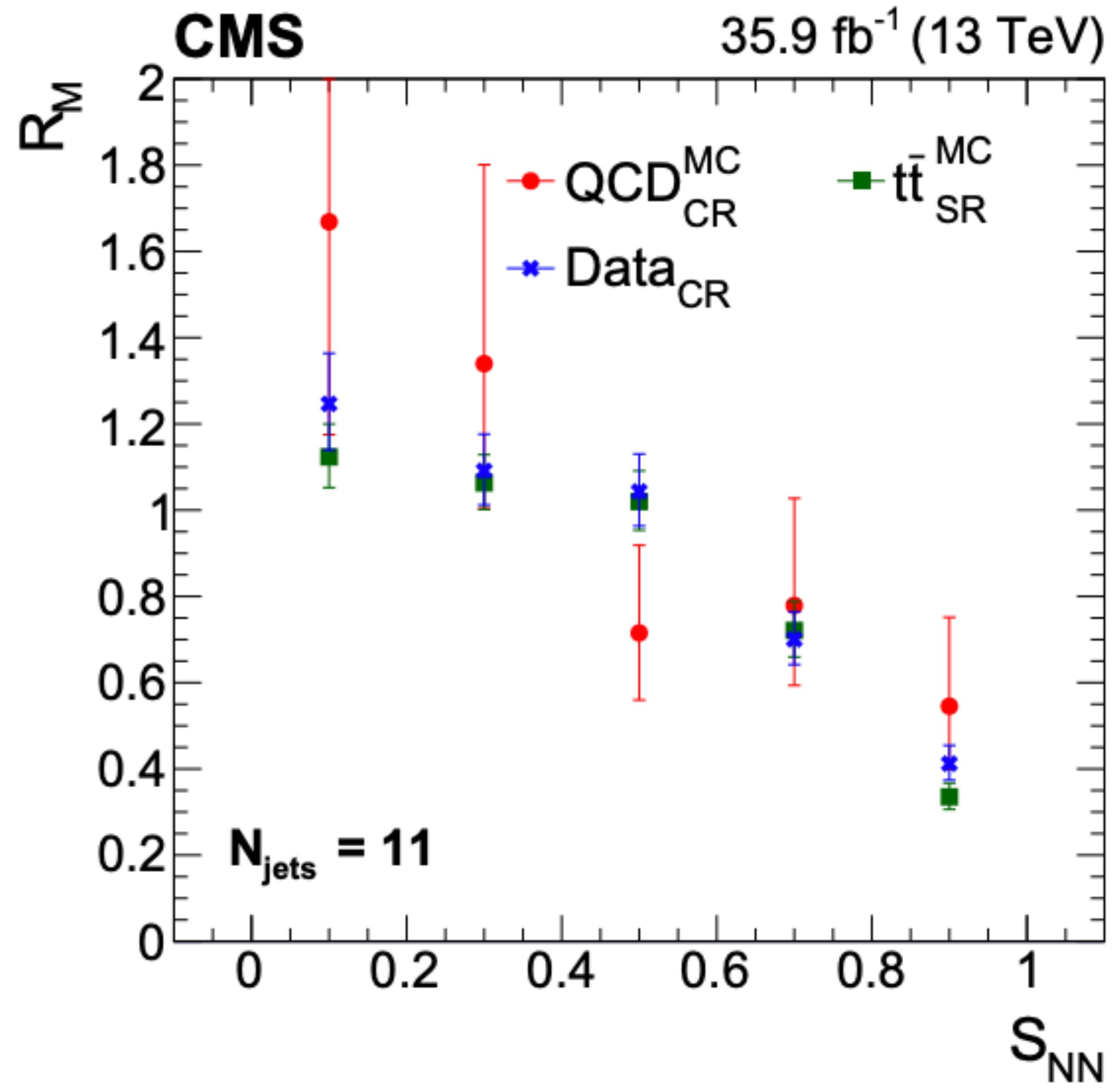





ATLAS Preliminary √s = 13 TeV, 139 fb-1 1ℓ category
Lep. cat.
NN bin
Pull
N_jets (p_T > 20 GeV)
N_b-tags


ATLAS Preliminary
√s = 13 TeV, 139 fb-1
1ℓ + 8 jets (p_T > 20 GeV)
higgsino, m(χ̃₁⁰) = 300 GeV (x20)
Data
tt̄ + jets
W+ jets
Z+ jets
Other
Events
Data/Model
N_b-tags
NN_8j bin


CMS
35.9 fb-1 (13 TeV)
QCD MC CR
tt̄ MC SR
Data CR
N_jets = 7
R_M
S_NN


CMS
35.9 fb-1 (13 TeV)
QCD MC CR
tt̄ MC SR
Data CR
N_jets = 11
R_M
S_NN