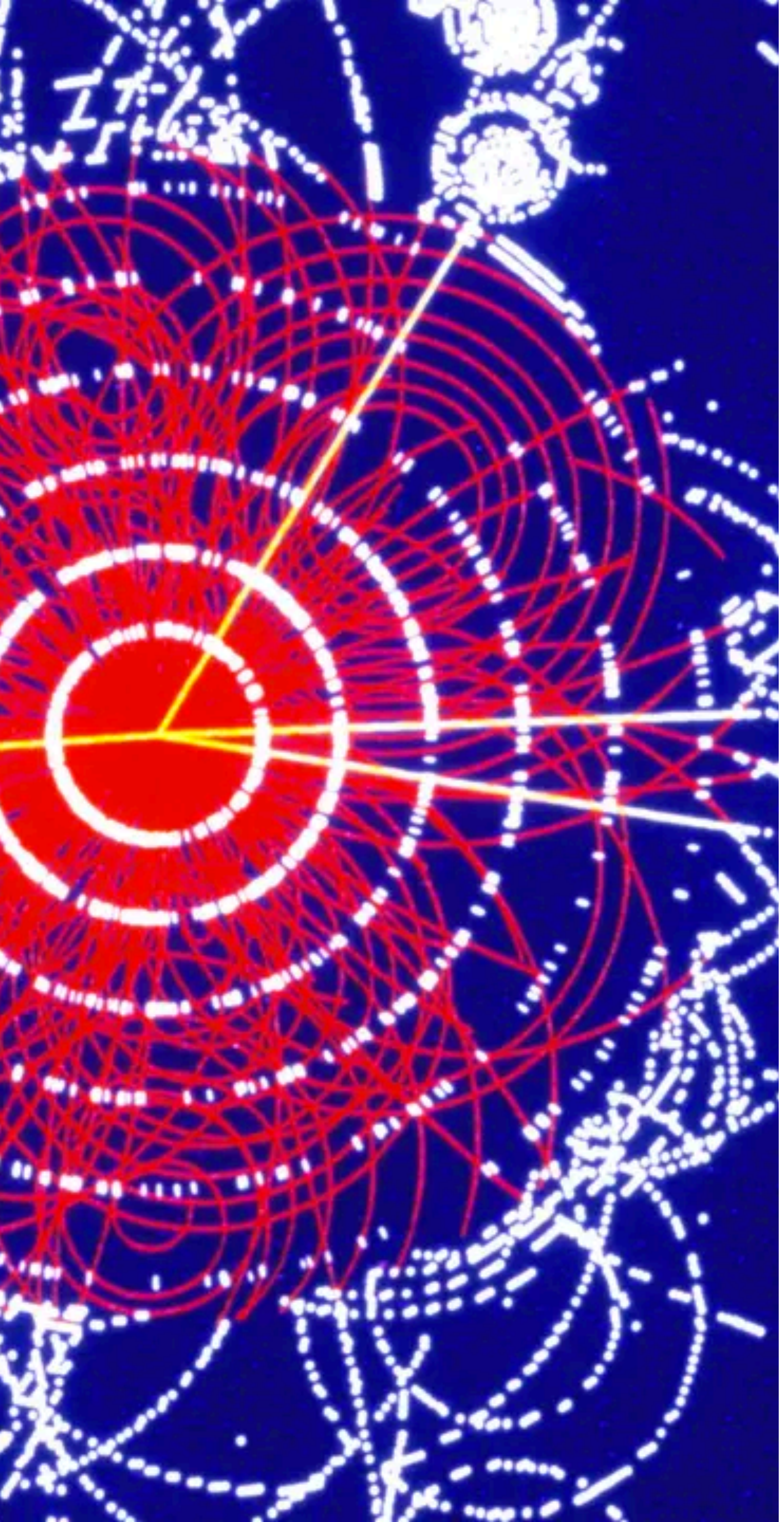


LPSC Grenoble

NATIONAL SCIENCE CENTRE POLAND

FACULTY OF PHYSICS UNIVERSITY OF WARSAW

Enhancing LHC searches for Dark Matter with Graph Neural Networks

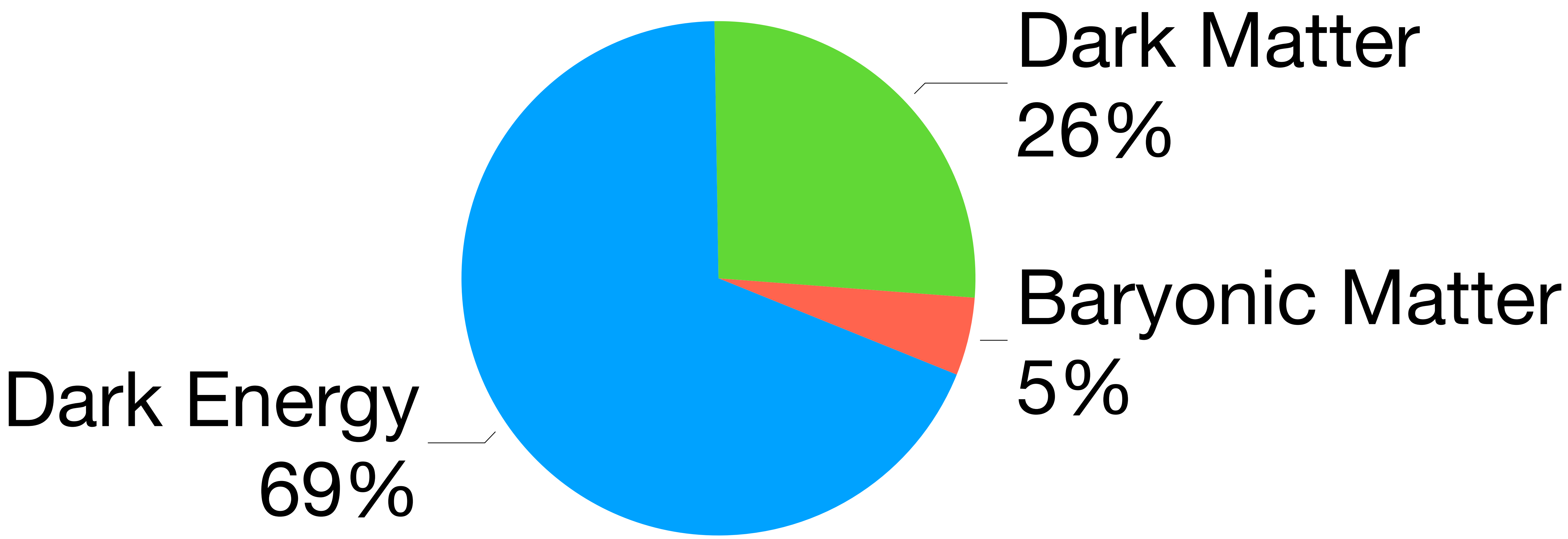
Rafał Masełek

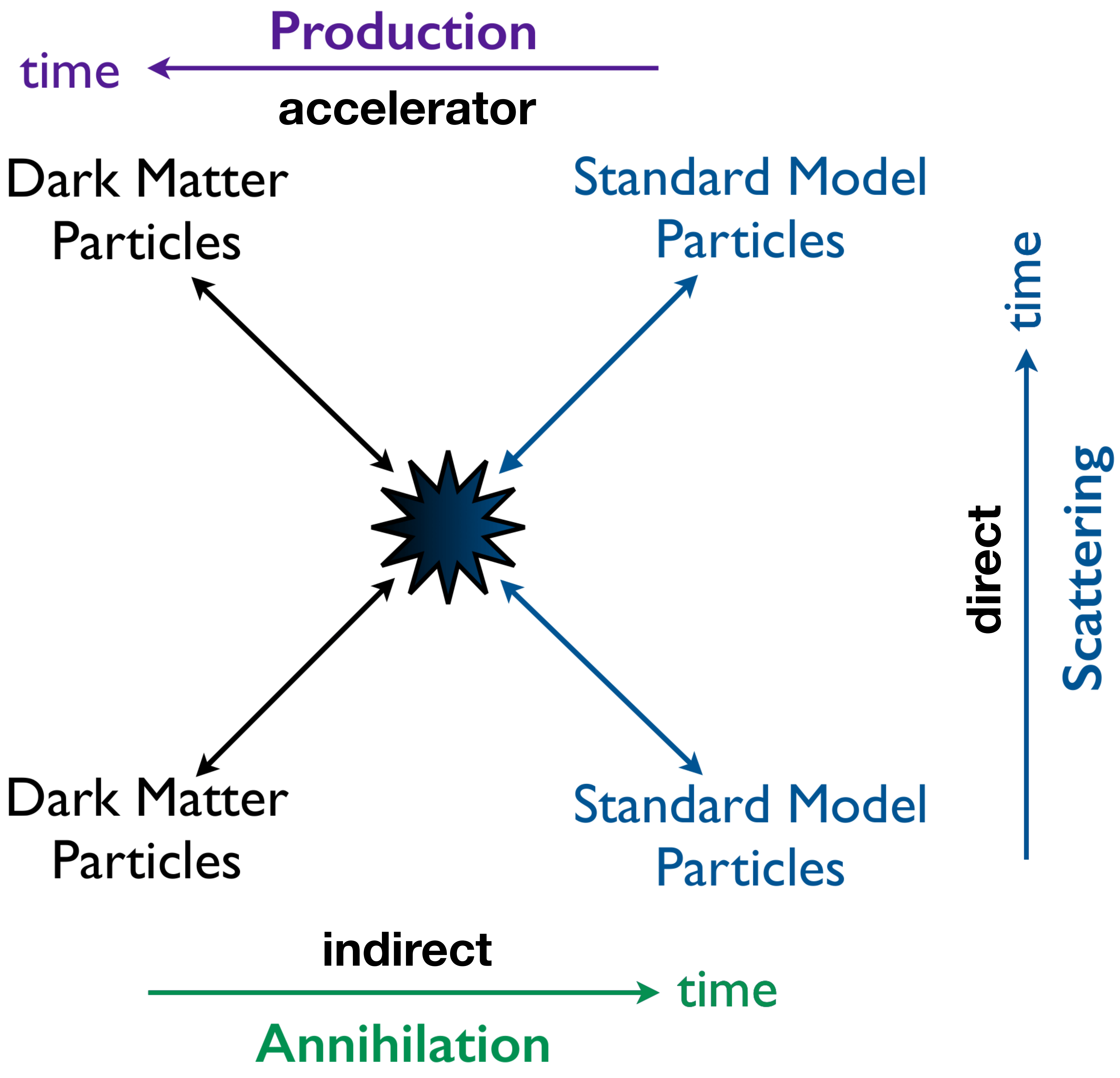
in collaboration with K. Sakurai and M. Nojiri

BOOST2024 31-07-2024

RM is supported by IN2P3 theory master project DataMATTER; Polish National Science Centre grants: 2020/38/E/ST2/00243 and 2021/41/N/ST2/00972; Polish National Agency For Academic exchange grant BPN/BEK/2022/1/00253/DEC/1.

Dark Matter



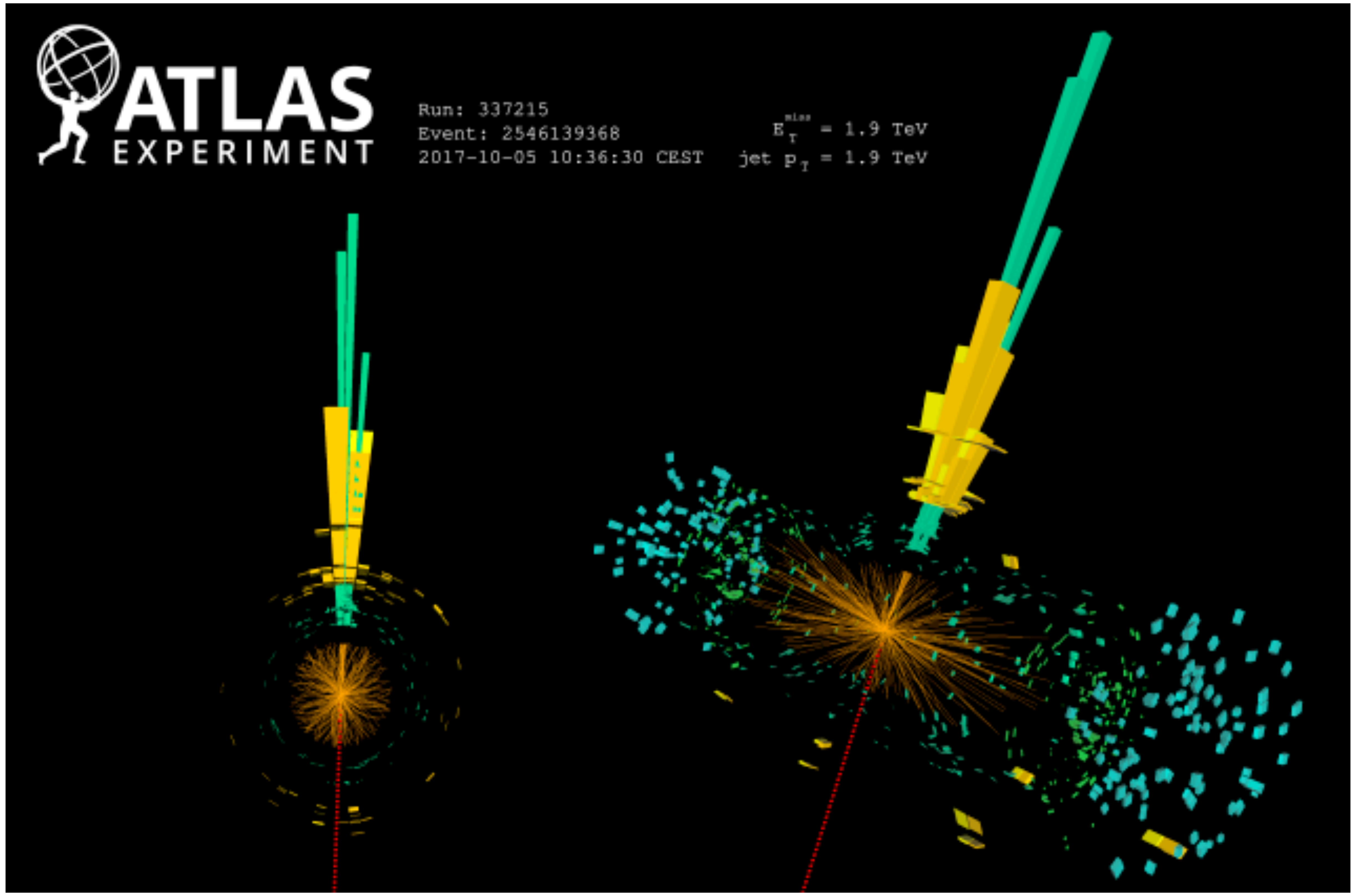


- long-lived over the age of the Universe
- feebly-interacting with photons and baryons
- not too hot

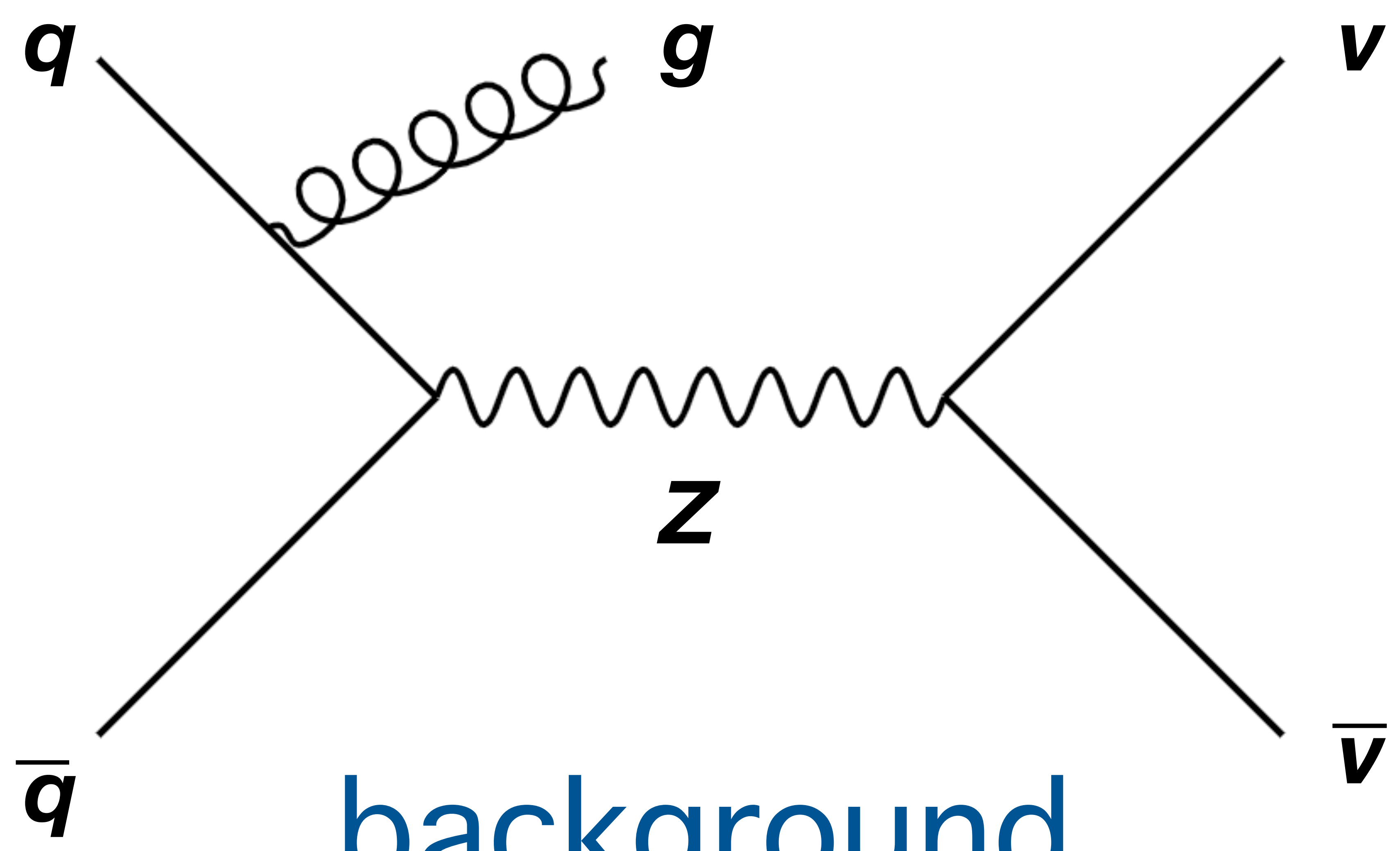
img source: https://particleastro.brown.edu/dark-matter/

DM searches @ LHC — Monojet

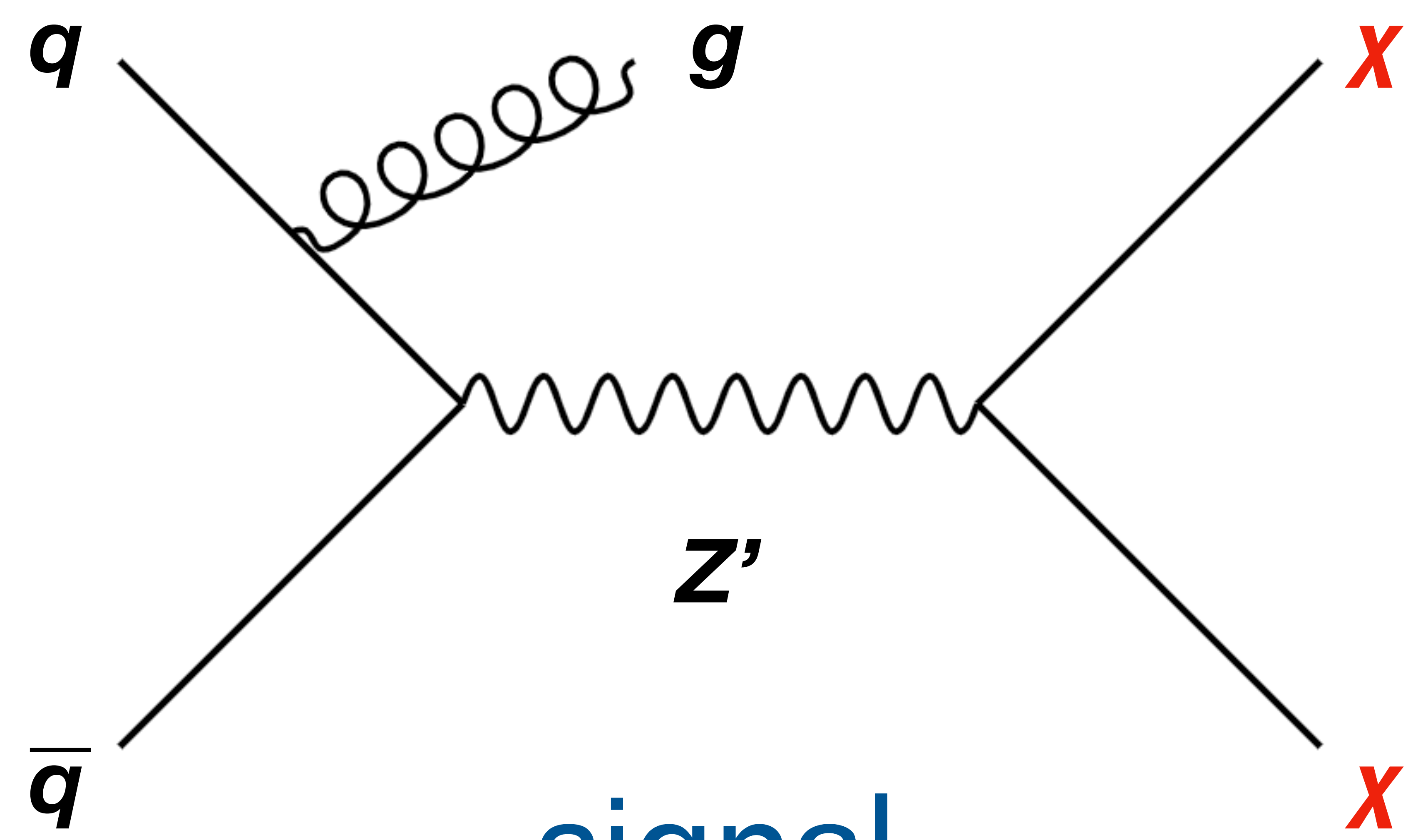
Monojet channel = 1 or more hard jets recoiling against a missing transverse momentum and no isolated leptons



img source: https://cds.cern.ch/record/2725235



background



signal

The idea

- One of the challenges for the Monojet searches is that we observe very similar jets for both signal and background

- Analysis of jet substructure is needed

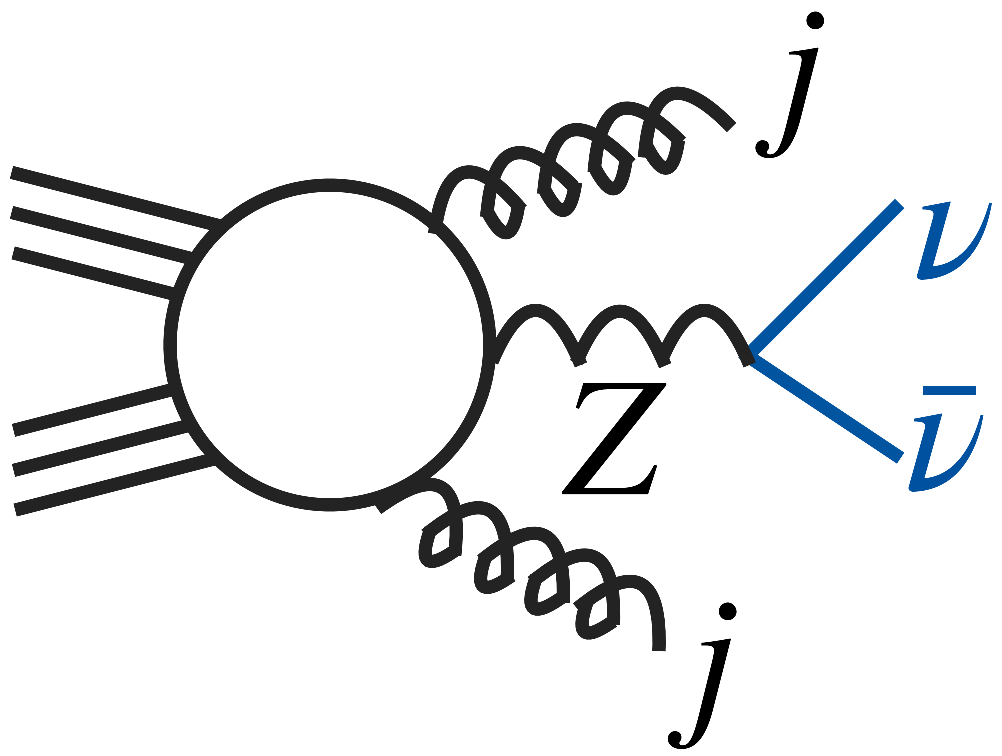
- With Machine Learning we can effectively analyse low-level data

- ML can learn both local and global correlations

- GOAL: Design new analysis using ML

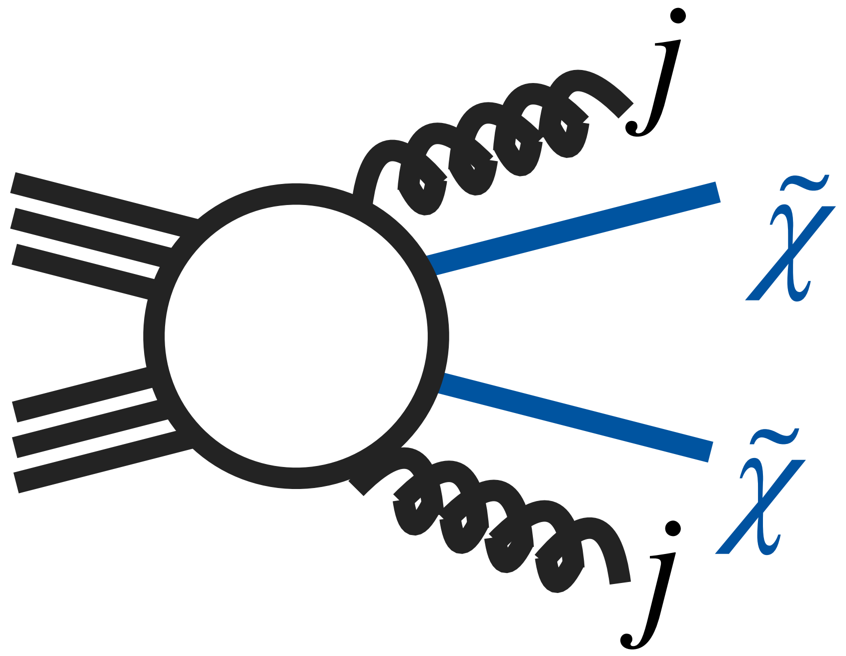
Benchmark model

SM background



$(Z \to \nu\bar{\nu})$ + jets

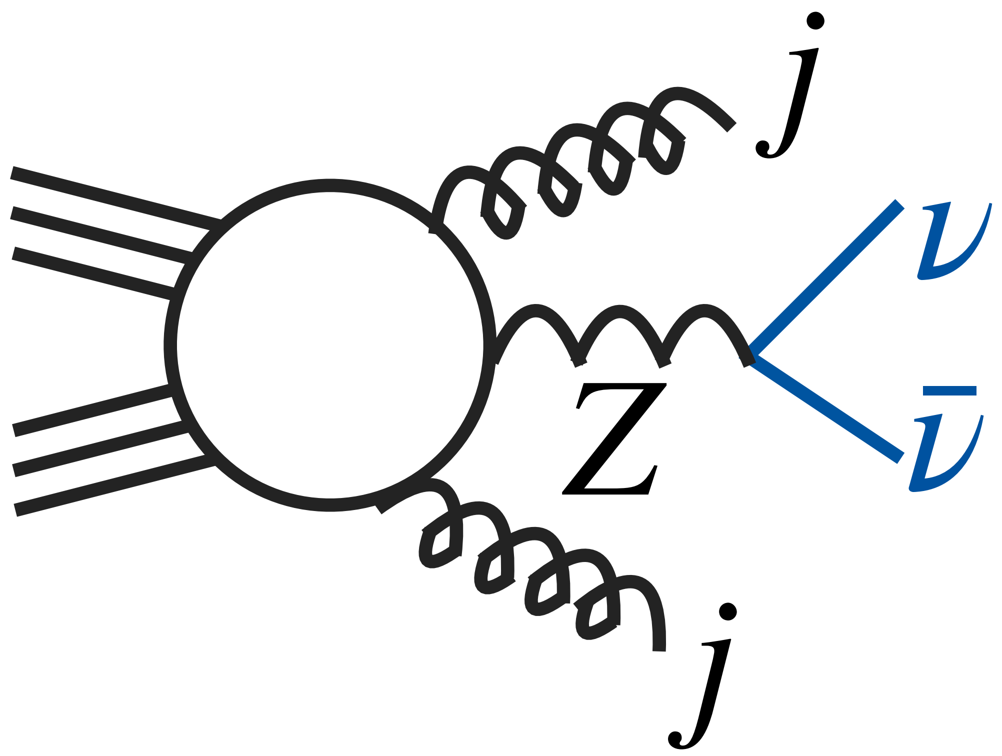
Contributing signal process



EWKino-EWKwino

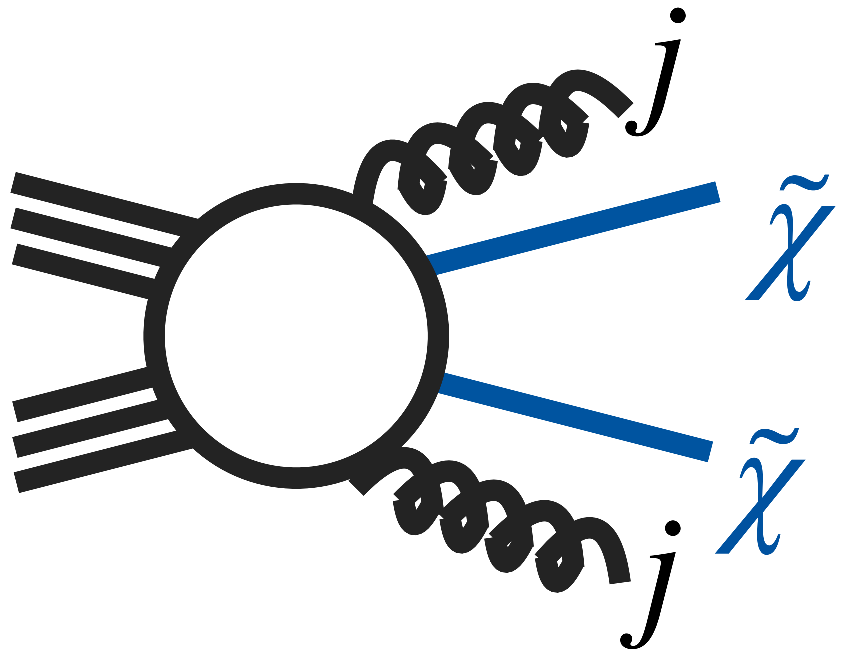
Benchmark model

SM background

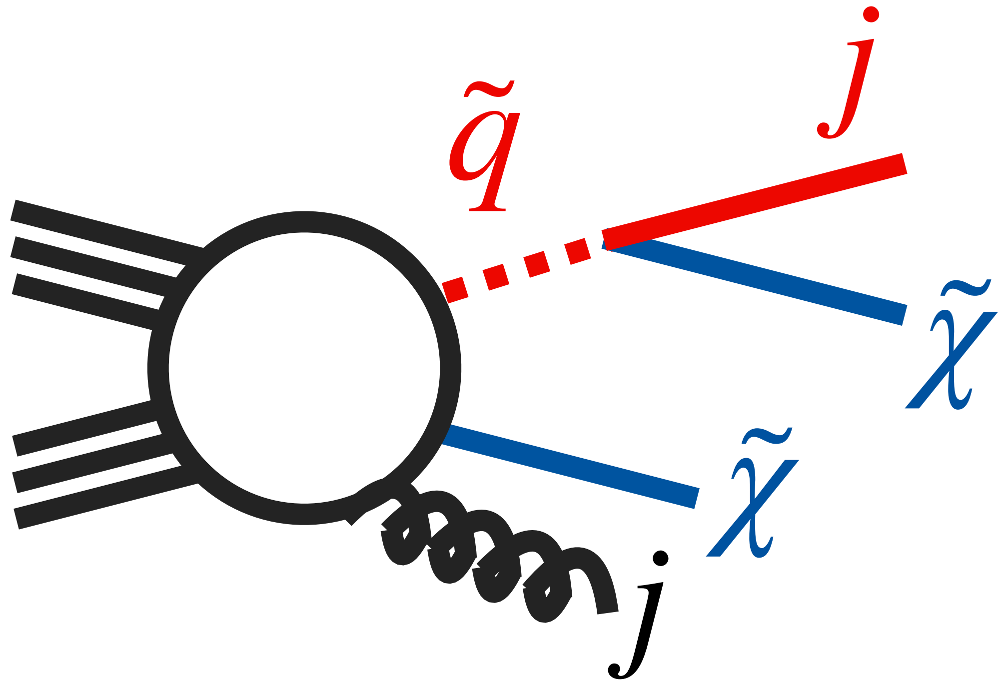


$(Z \to \nu\bar{\nu}) + \text{jets}$

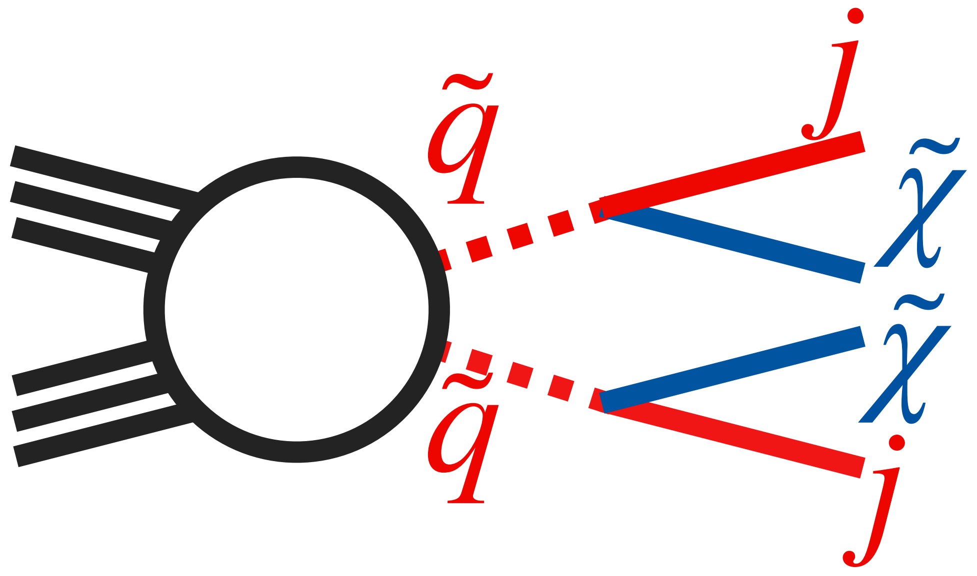
Contributing signal processes



EWKino-EWKwino



EWKino-squark



squark-squark

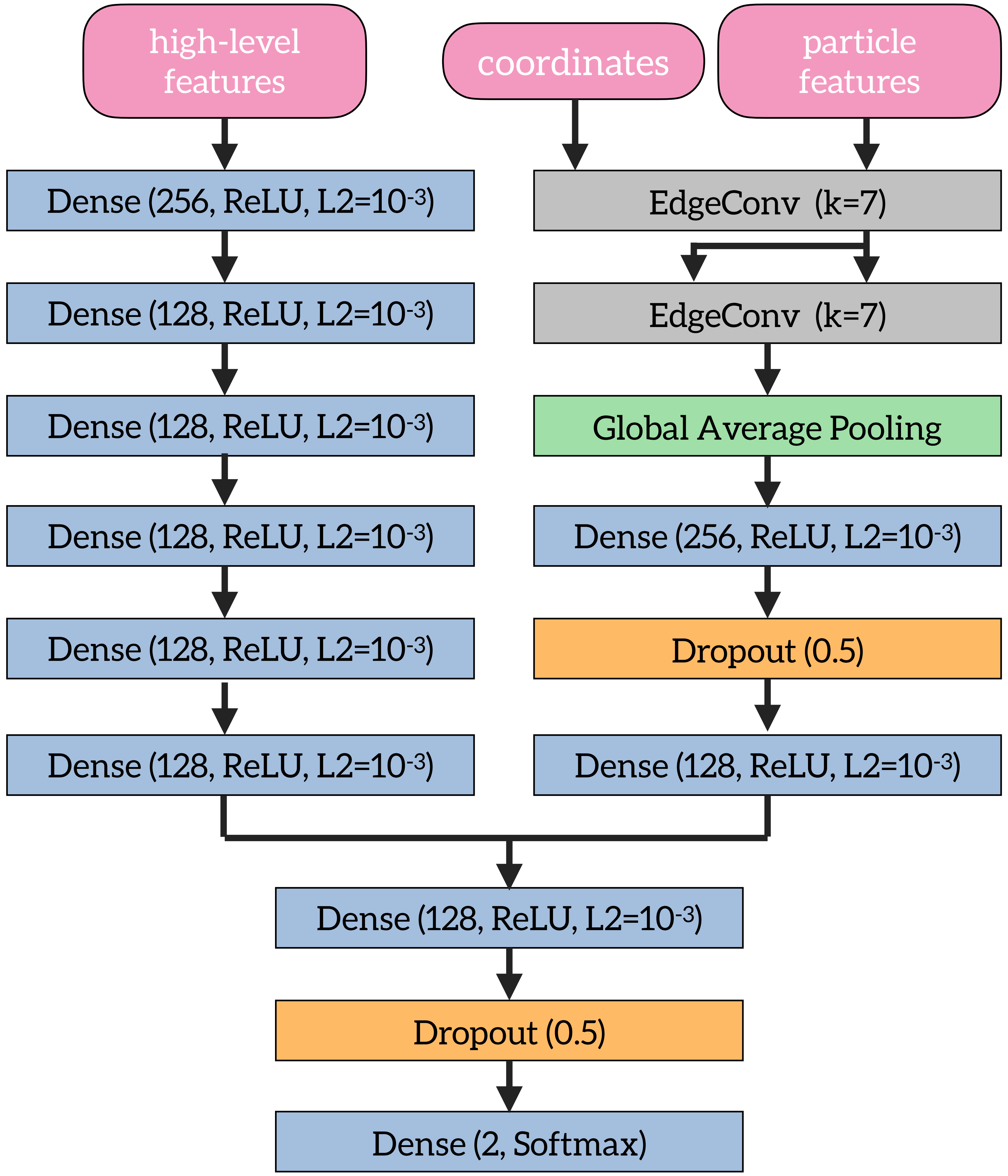
$m_{\tilde{\chi}} \in \{200,\ 300,\ 500,\ 700,\ 900\}$ GeV
$m_{\tilde{q}} \in \{2.0,\ 2.2,\ 2.3,\ 2.7,\ 3.0\}$ TeV
$m_{\tilde{g}} = 10$ TeV

Analysis

Preanalysis

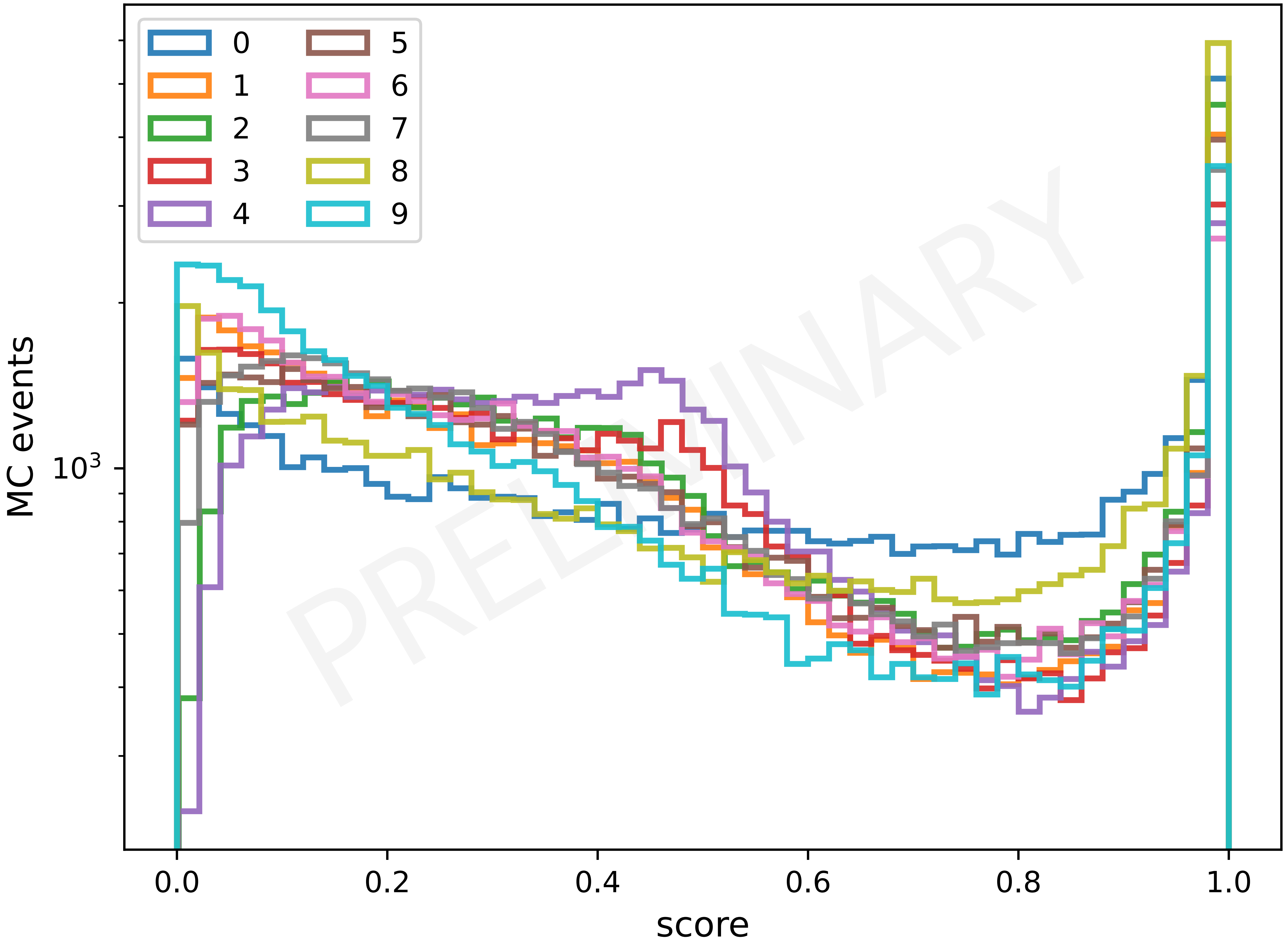
- $4 \geq \#\text{jets} \geq 2$
- $p_T^{1j} > 520$ GeV
- $\eta^{1j} < 2.0$
- $p_T^{2j} > 320$ GeV
- $\eta^{2j} < 2.0$
- MET > 820 GeV
- lepton veto
- $\Delta\phi\left(j^{1,2,3}, p_T^{miss}\right) > 0.8$
- $\Delta\phi\left(j^{4}, p_T^{miss}\right) > 0.4$
- $\text{MET}/\sqrt{\text{H}_\text{T}} > 16\ \sqrt{\text{GeV}}$
- $M_{\text{eff}} > 1600$ GeV
- particle $p_T > 1$ GeV

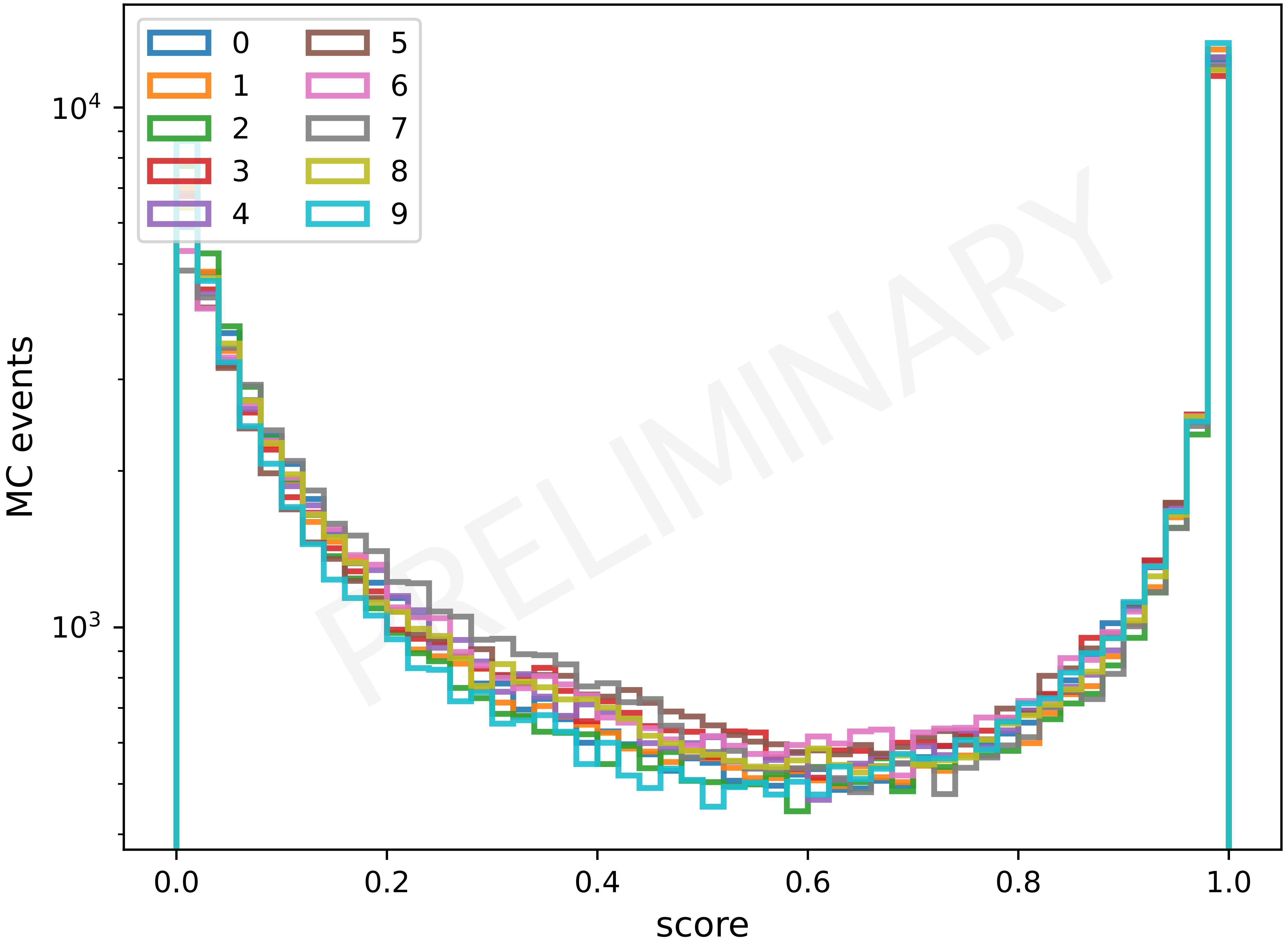
GNN model



Ensemble training

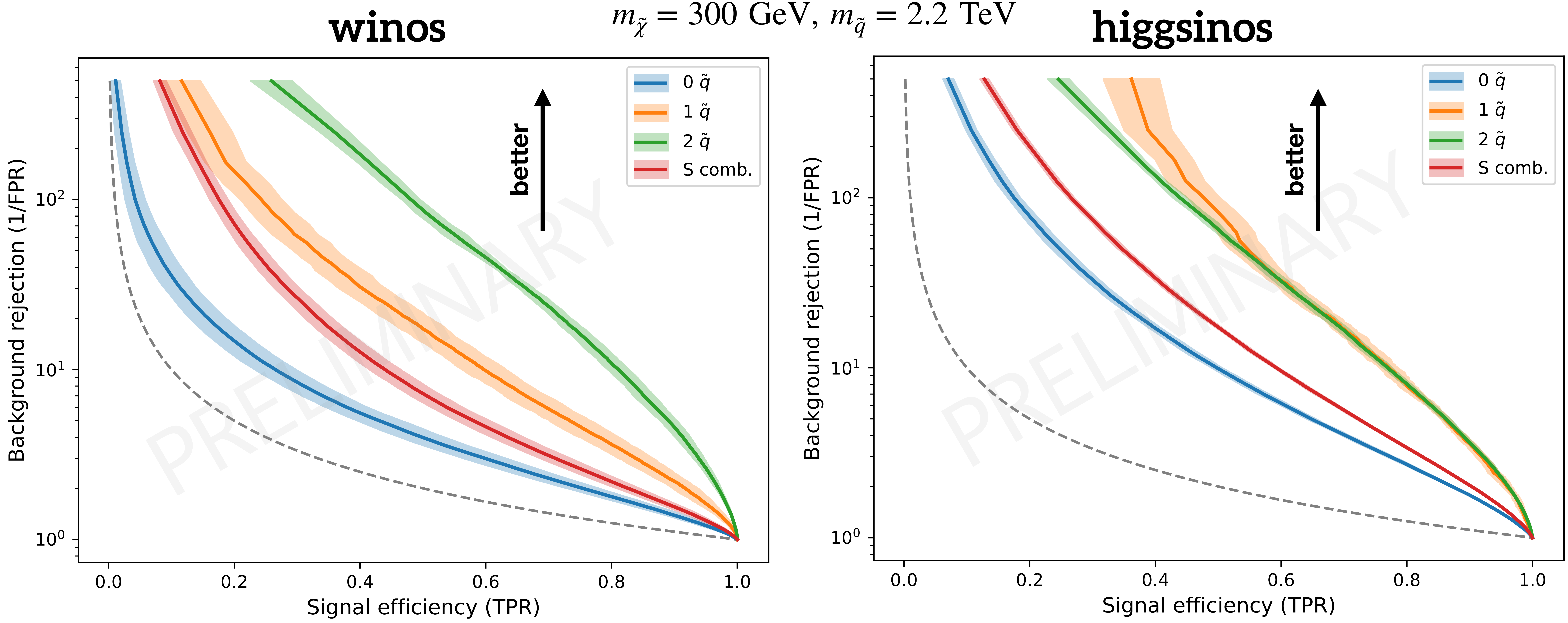
winos $m_{\tilde{\chi}} = 300$ GeV, $m_{\tilde{q}} = 2.2$ TeV **higgsinos**





Evaluation

$m_{\tilde{\chi}} = 300$ GeV, $m_{\tilde{q}} = 2.2$ TeV

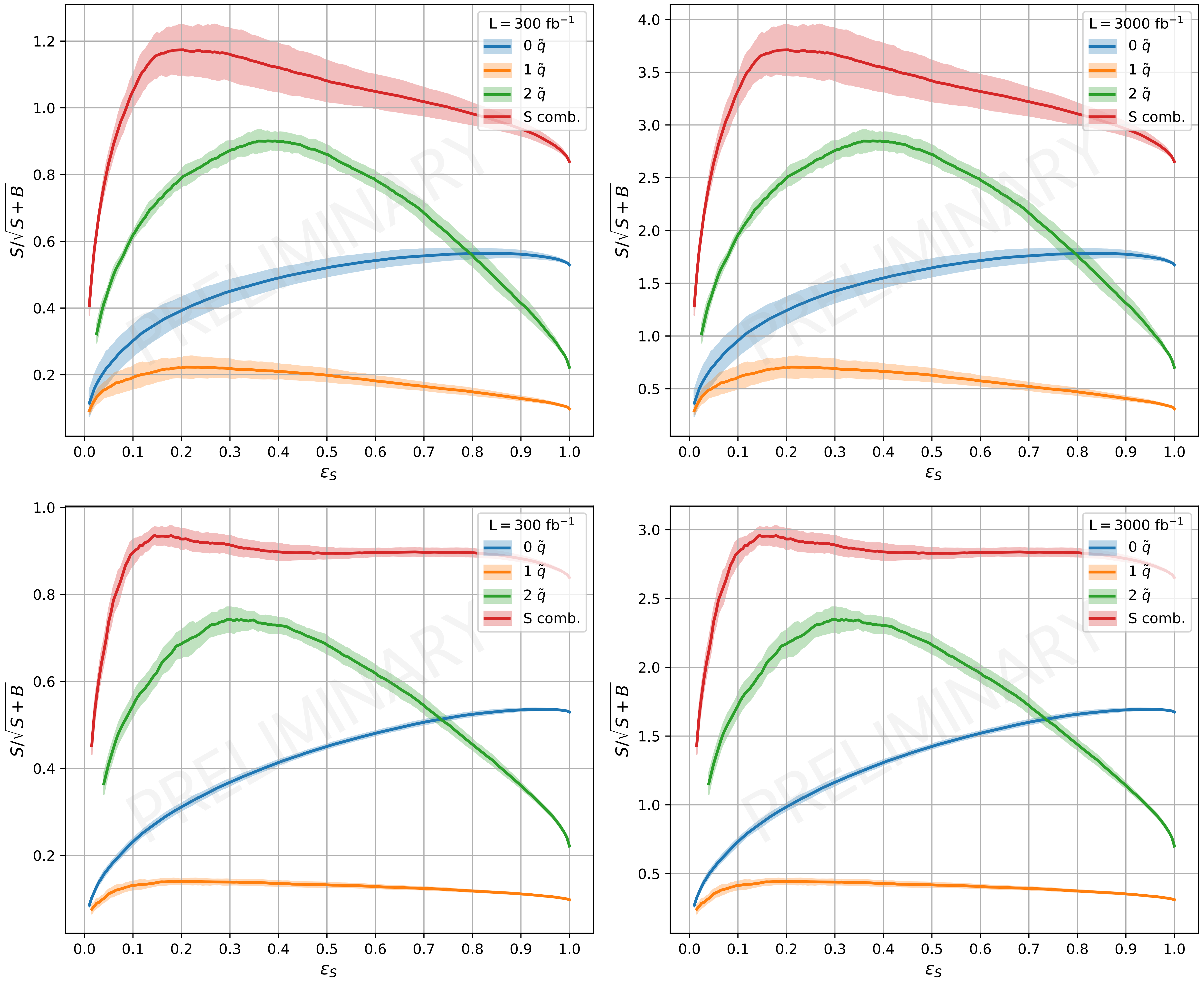


Signal efficiency vs. naive significance

Run-3 | **HL-LHC**

winos

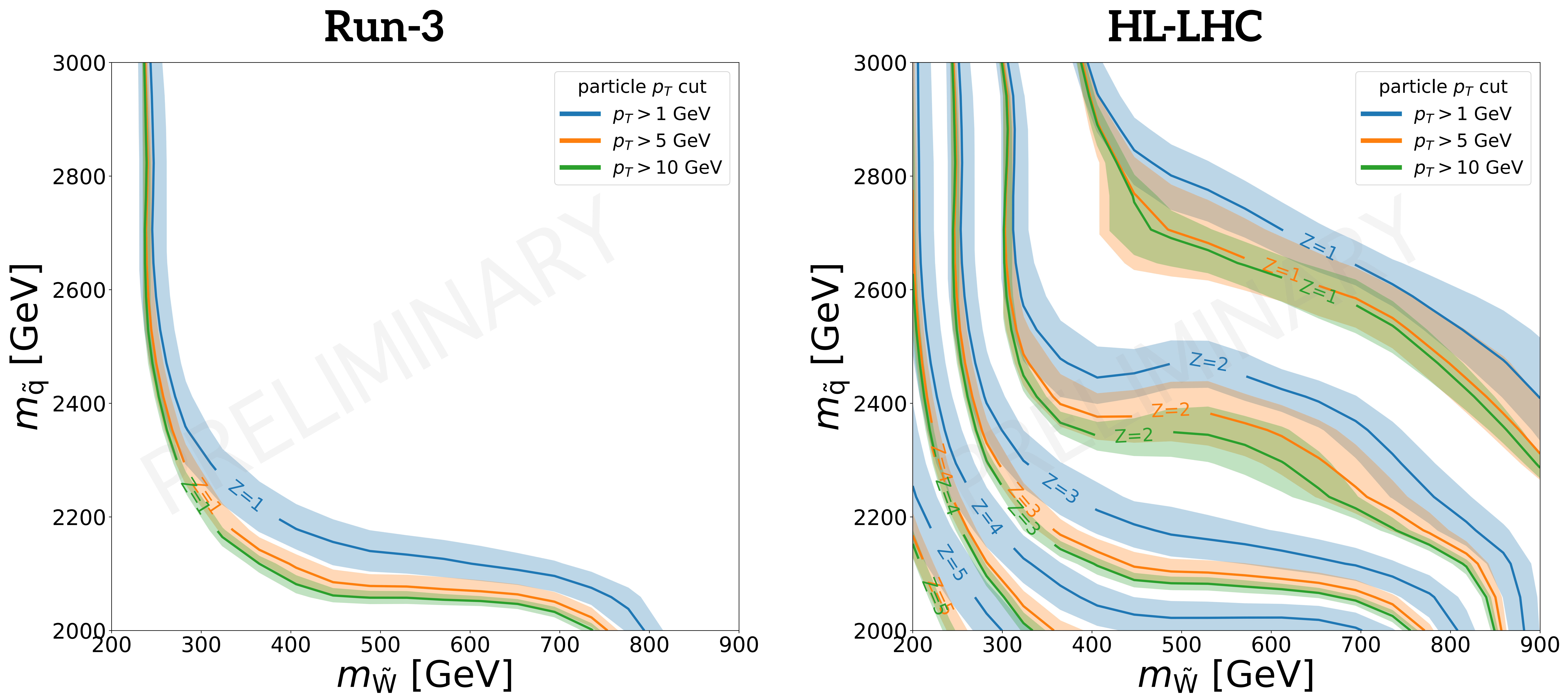
higgsinos



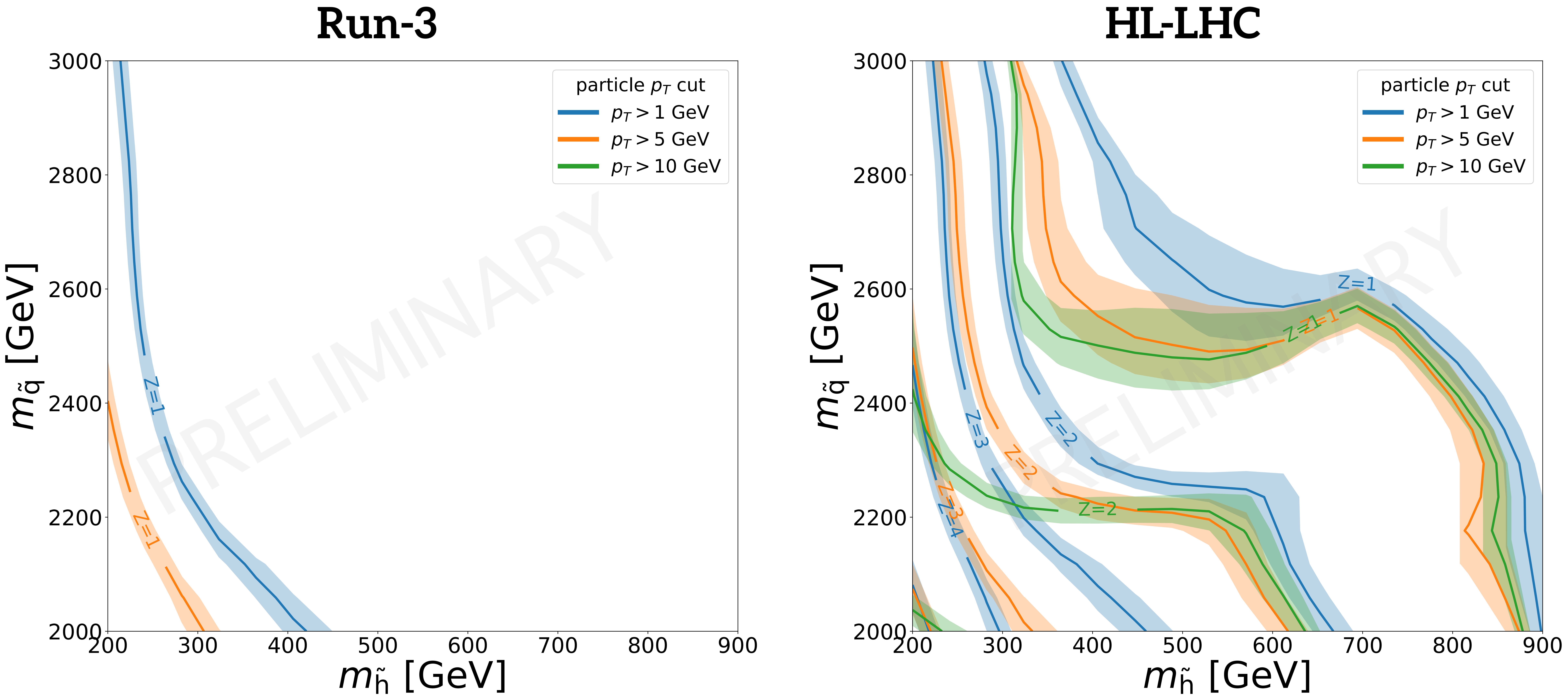
$$Z = S/\sqrt{S+B}$$

$m_{\tilde{\chi}} = 300$ GeV, $m_{\tilde{q}} = 2.2$ TeV

Naive significance for winos



Naive significance for higgsinos



R. Masełek BOOST2024 31-07-2024

Upcoming paper

- [x] new data set
- [x] new architecture
- [x] evaluation
  - [x] ROC curves
  - [ ] AUC
  - [ ] comparison with BDT
  - [ ] impact of cut on particles' pT
  - [ ] cross-evaluation
- [x] limits
  - [x] wino
  - [x] higgsino
  - [ ] bino
- [ ] interpretation
  - [ ] input and output correlations
  - [ ] feature importance
  - [ ] jet characteristics

Machine learning squark and electroweakino production

Rafał Masełek[a,b], Mihoko M. Nojiri[c,d,e] and Kazuki Sakurai[a]

[a] Institute of Theoretical Physics, Faculty of Physics, University of Warsaw, Pasteura 5, 02-093, Warsaw, Poland

[b] *Laboratoire de Physique Subatomique et de Cosmologie (LPSC), Université Grenoble-Alpes, CNRS/IN2P3, 53 Avenue des Martyrs, F-38026 Grenoble, France*

[c] Theory Center, IPNS, KEK, 1-1 Oho, Tsukuba, Ibaraki 305-0801, Japan

[d] The Graduate University of Advanced Studies (Sokendai), 1-1 Oho, Tsukuba, Ibaraki 305-0801, Japan

[e] Kavli IPMU (WPI), University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8583, Japan

Abstract

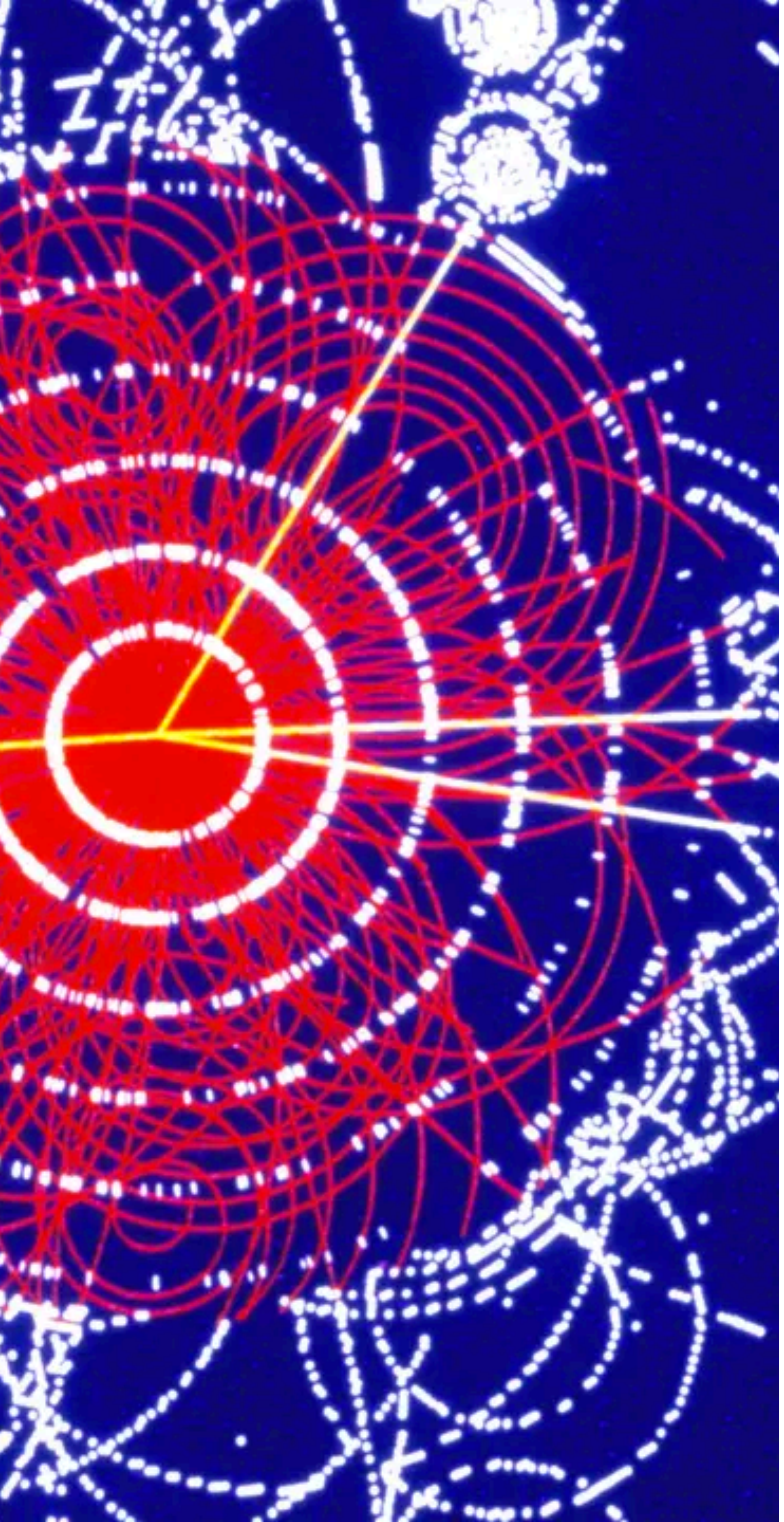
The system of light electroweakinos and heavy squarks gives rise to one of the most challenging signatures to detect at the LHC. It consists of missing transverse energy (MET) recoiled against a few hadronic jets originating either from QCD radiation or squark decays. The analysis generally suffers from the large irreducible $Z+$jets ($Z \to \nu\bar{\nu}$) background. In this study, we explore Machine Learning (ML) methods for efficient signal/background discrimination. Our best attempt uses both reconstructed (jets, MET, etc.) and low-level (particle-flow) objects. We find that the discrimination performance improves as the p_{T} threshold for soft particles is lowered from 10 GeV to 1 GeV, at the expense of larger systematic uncertainty. In many cases, the ML method provides a factor 2 enhancement in $S/\sqrt{S+B}$ from a simple kinematical selection. The sensitivity on the squark-elecroweakino mass plane is derived with this method, assuming the Run-3 and HL-LHC luminosities.

1

Summary

- Dark Matter can be searched @ LHC
- We introduce new analysis based on GNN
- SUSY as benchmark model:
  - EWKino pair production
  - EWKino-squark associated production
  - squark-pair production
- We evaluate our model:
  - sample composition is crucial: the more squarks the easier classification
  - high robustness against a change of the EWKino type
  - high robustness against change of the masses of sparticles
- We derive the limits:
  - For Run-3 LHC, there is always $Z < 2$
  - For HL-LHC, Z=5 for light sparticles, and Z=2 even for $m_{\tilde{\chi}} = 900$ GeV
  - Limits on higgsinos are a little weaker
- Paper should be on arXiv soon.


Thank you for attention!
rafal.maselek@lpsc.in2p3.fr
Dolina Chochołowska, Poland
photo by Piotr Kałuża



FACULTY OF PHYSICS
UNIVERSITY OF WARSAW

Backup slides

Rafał Masełek

BOOST2024 31-07-2024

AUC values

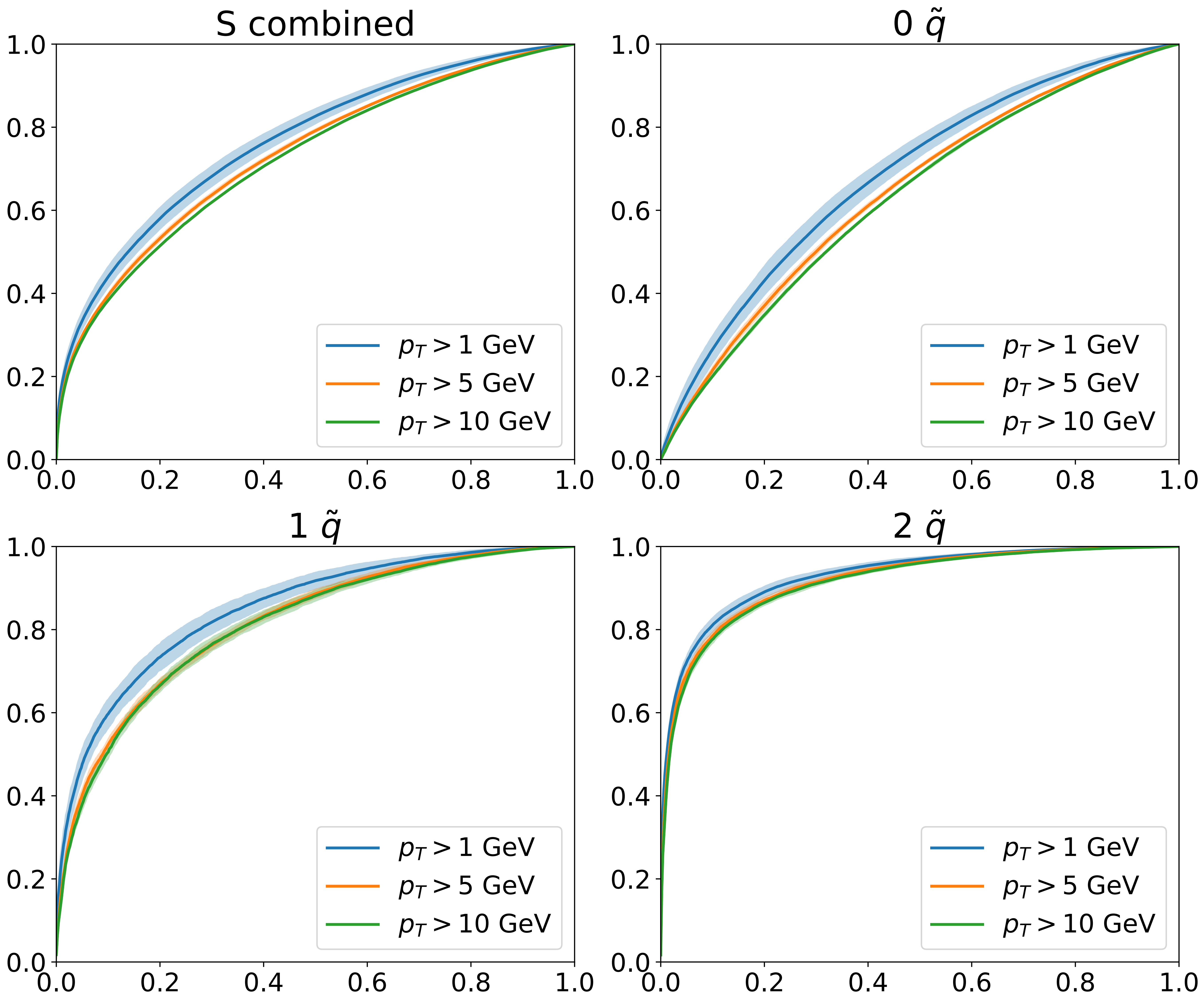
Table 4: Area Under Curve for ensemble of NN models trained and evaluated on $m_{\tilde{W}}$ = 300 GeV, $m_{\tilde{q}}$ = 2.2 TeV, and $m_{\tilde{g}}$ = 10 TeV.

Signal class	mean AUC	standard deviation
0 $\tilde{q}$	0.6844	0.0215
1 $\tilde{q}$	0.8494	0.0173
2 $\tilde{q}$	0.9299	0.0058
combined	0.7646	0.0156

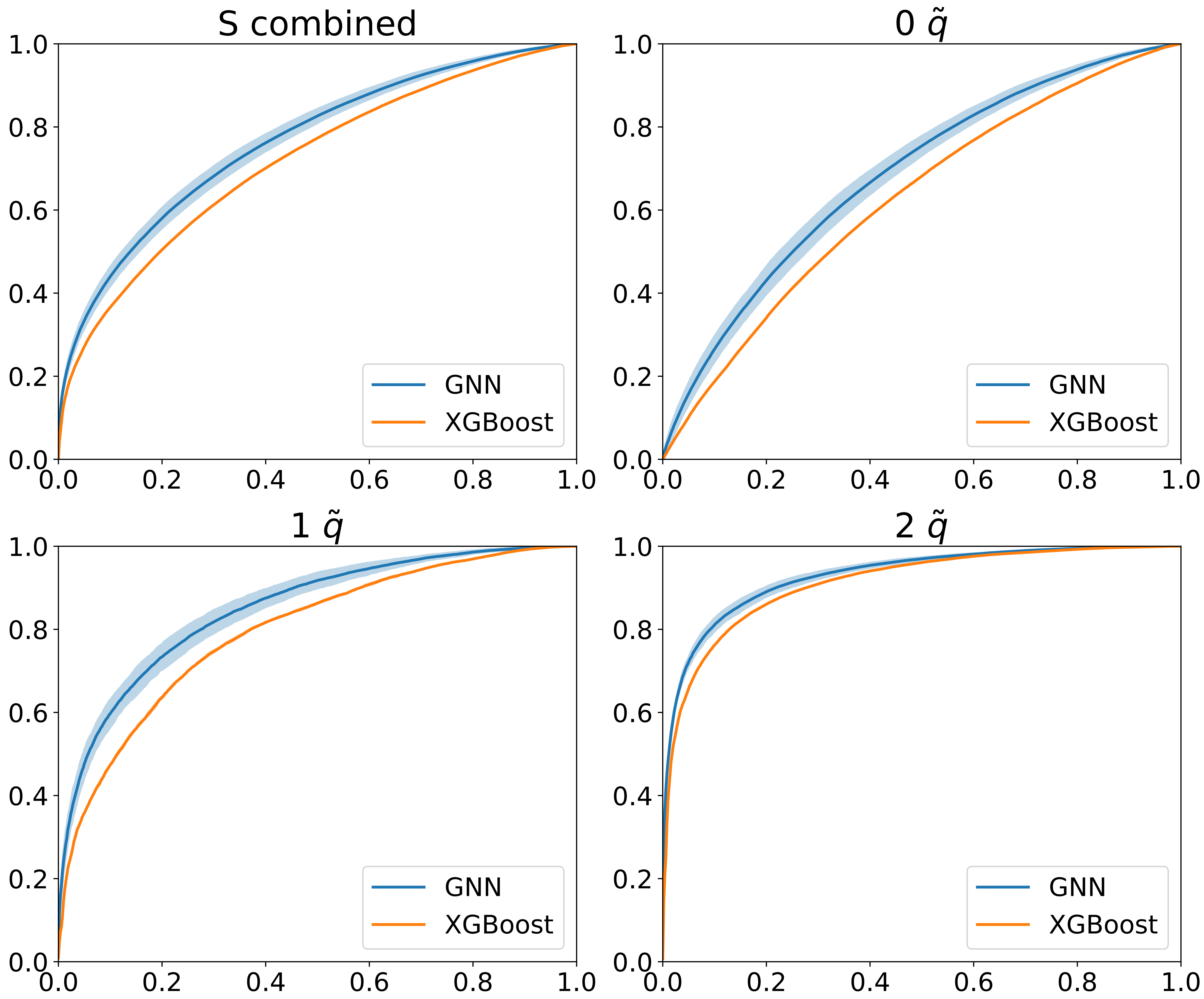
Table 5: Area Under Curve for ensemble of NN models trained and evaluated on $m_{\tilde{h}}$ = 300 GeV and $m_{\tilde{q}}$ = 2.2 TeV.

Signal class	mean AUC	standard deviation
0 $\tilde{q}$	0.8018	0.0038
1 $\tilde{q}$	0.9146	0.0057
2 $\tilde{q}$	0.9157	0.0028
combined	0.8410	0.0019

particles' pT cut



GNN vs. BDT



S combined
$0\ \tilde{q}$
$1\ \tilde{q}$
$2\ \tilde{q}$
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
GNN
XGBoost

cross-evaluation