Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

PUNCH 4NFDI

Anomaly detection interpretability and phenomenology

Universal new physics latent space

Based on [2407.20315](https://arxiv.org/abs/2407.20315)

Anna Hallin, G. Kasieczka, S. Kraml, A. Lessa, L. Moureaux, T. von Schwartz, D. Shih
anna.hallin@uni-hamburg.de

Anomaly Detection for High Energy Physics (AD4HEP) Workshop, Columbia University, June 17, 2025

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

PUNCH 4NFDI

Anomaly detection interpretability and phenomenology

Universal new physics latent space

Based on 2407.20315

Anna Hallin, G. Kasieczka, S. Kraml, A. Lessa, L. Moureaux, Tore von Schwartz, D. Shih
anna.hallin@uni-hamburg.de

Anomaly Detection for High Energy Physics (AD4HEP) Workshop, Columbia University, June 17, 2025

A very incomplete selection of…

The BSM zoo and their multitude of different signatures

Theory of everything

Extra dimensions

Twin Higgs

SIMPs

FIMPs

SUSY

Z’

GUT

Leptoquarks

String theory

Right-handed neutrinos

Seesaw mechanism

Technicolor

Quantum gravity

2HDM

Dark photon

Kaluza-Klein

Composite Higgs

Axions

ALPs

W’

Little(st) Higgs

Sterile neutrinos

WIMPs

Supergravity

Magnetic monopoles

…

Very specific searches...

EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH (CERN)

ATLAS EXPERIMENT

Submitted to: EPJC

CERN-EP-2025-055
31st March 2025

Search for electroweak production of vector-like leptons in τ-lepton and b-jet final states in collisions at $\sqrt{s}$ = 13 TeV with the ATLAS de

The ATLAS Collaboration

A search for pair-production of vector-like leptons is presented, considering their decay third-generation Standard Model (SM) quark and a vector leptoquark (U_1) as predicte ultraviolet-complete extension of the SM, referred to as the '4321' model. Given the a decay of U_1 into third-generation SM fermions, the final state can contain multiple τ- and b-quarks. This search is based on a dataset of pp collisions at $\sqrt{s}$ = 13 TeV record the ATLAS detector during Run 2 of the Large Hadron Collider, corresponding to an int luminosity of up to 140 fb^{-1}. No significant excess above the SM background pre is observed, and 95% confidence level limits on the cross-section times branching ra

Model
53

EUROPEAN ORGANIZATION FOR NUCLEAR RESEAR

CMS

CMS-SUS-23-012

CERN-EP-2025-
2025/06/10

Search for dark matter produced in association with a Higgs boson decaying to a τ lepton pair in proton-proton collisions at $\sqrt{s} = 13$ TeV

The CMS Collaboration*

Abstract

A search for dark matter particles produced in association with a Higgs boson decaying into a pair of τ leptons is performed using data collected in proton-proton collisions at a center-of-mass energy of 13 TeV with the CMS detector. The analysis is based on a data set corresponding to an integrated luminosity of 101 fb^{-1} collected in 2017–2018. No significant excess over the expected standard model background is

EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH (CERN)

CMS

CMS-TOP-22-011

CERN-EP-2025-029
2025/04/14

Search for charged-lepton flavour violation in top quark interactions with an up-type quark, a muon, and a τ lepton in proton-proton collisions at $\sqrt{s} = 13$ TeV

The CMS Collaboration*

Abstract

A search for charged-lepton flavour violation (CLFV) in top quark (t) production and decay is presented. The search uses proton-proton collision data corresponding to 138 fb^{-1} collected with the CMS experiment at $\sqrt{s} = 13$ TeV. The signal consists of the production of a single top quark via a CLFV interaction or top quark pair production followed by a CLFV decay. The analysis selects events containing a pair of oppositely

...or very broad searches

PHYSICAL REVIEW D 108, 052009 (2023)

Anomaly detection search for new resonances decaying into a Higgs boson and a generic new particle X in hadronic final states using $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector

G. Aad *et al.*[*]
(ATLAS Collaboration)

(Received 12 June 2023; accepted 29 July 2023; published 18 September 2023)

A search is presented for a heavy resonance Y decaying into a Standard Model Higgs boson H and a new particle X in a fully hadronic final state. The full Large Hadron Collider run 2 dataset of proton-proton collisions at $\sqrt{s} = 13$ TeV collected by the ATLAS detector from 2015 to 2018 is used and corresponds to an integrated luminosity of 139 fb^{-1}. The search targets the high Y-mass region, where the H and X have a significant Lorentz boost in the laboratory frame. A novel application of anomaly detection is used to define a general signal region, where events are selected solely because of their incompatibility with a learned background-only model. It is constructed using a jet-level tagger for signal-model-independent selection of the boosted X particle, representing the first application of fully unsupervised machine learning to an ATLAS analysis. Two additional signal regions are implemented to target a benchmark X decay into two quarks, covering topologies where the X is reconstructed as either a single large-radius jet or two small-radius jets. The analysis selects Higgs boson decays into $b\bar{b}$, and a dedicated neural-network-based tagger provides sensitivity to the boosted heavy-flavor topology. No significant excess of data over the expected background is observed, and the results are presented as upper limits on the production cross section $\sigma(pp \to Y \to XH \to q\bar{q}b\bar{b})$ for signals with m_Y between 1.5 and 6 TeV and m_X between 65 and 3000 GeV.

DOI: 10.1103/PhysRevD.108.052009

I. INTRODUCTION

The Standard Model (SM) provides a framework for understanding fundamental particles and interactions that has been remarkably predictive of experimental results over several decades. The discovery of the Higgs boson in

A search is presented here for a new TeV-scale narrow-width particle Y, which decays into a Standard Model Higgs boson H and a new particle X with a mass near the weak scale. A fully hadronic final state is targeted for both particles. Tagging of the boosted Higgs decay into two b quarks ($H \to b\bar{b}$ tagging) enhances a signal by

OPEN ACCESS
IOP Publishing
Reports on Progress in Physics
Rep. Prog. Phys. 88 (2025) 067802 (37pp)
https://doi.org/10.1088/1361-6633/add762

Model-agnostic search for dijet resonances with anomalous jet substructure in proton–proton collisions at $\sqrt{s} = 13$ TeV

The CMS Collaboration
CERN, Geneva, Switzerland
E-mail: cms-publication-committee-chair@cern.ch

Received 4 December 2024, revised 9 April 2025
Accepted for publication 12 May 2025
Published 3 June 2025

Corresponding editor: Dr Paul Mabey

CrossMark

Abstract
This paper presents a model-agnostic search for narrow resonances in the dijet final state in the mass range 1.8–6 TeV. The signal is assumed to produce jets with substructure atypical of jets initiated by light quarks or gluons, with minimal additional assumptions. Search regions are obtained by utilizing multivariate machine-learning methods to select jets with anomalous substructure. A collection of complementary anomaly detection methods—based on unsupervised, weakly supervised, and semisupervised algorithms—are used in order to maximize the sensitivity to unknown new physics signatures. These algorithms are applied to

PHYSICAL REVIEW LETTERS 125, 131801 (2020)

Dijet Resonance Search with Weak Supervision Using $\sqrt{s} = 13$ TeV pp Collisions in the ATLAS Detector

G. Aad *et al.*[*]
(ATLAS Collaboration)

(Received 7 May 2020; revised 12 July 2020; accepted 4 August 2020; published 21 September 2020)

This Letter describes a search for narrowly resonant new physics using a machine-learning anomaly detection procedure that does not rely on signal simulations for developing the analysis selection. Weakly supervised learning is used to train classifiers directly on data to enhance potential signals. The targeted topology is dijet events and the features used for machine learning are the masses of the two jets. The resulting analysis is essentially a three-dimensional search $A \to BC$, for $m_A \sim \mathcal{O}(\mathrm{TeV})$, $m_B, m_C \sim \mathcal{O}(100\ \mathrm{GeV})$ and B, C are reconstructed as large-radius jets, without paying a penalty associated with a large trials factor in the scan of the masses of the two jets. The full run 2 $\sqrt{s} = 13$ TeV pp collision dataset of 139 fb^{-1} recorded by the ATLAS detector at the Large Hadron Collider is used for the search. There is no significant evidence of a localized excess in the dijet invariant mass spectrum between 1.8 and 8.2 TeV. Cross-section limits for narrow-width A, B, and C particles vary with m_A, m_B, and m_C. For example, when $m_A = 3$ TeV and $m_B \gtrsim 200$ GeV, a production cross section between 1 and 5 fb is excluded at 95% confidence level, depending on m_C. For certain masses, these limits are up to 10 times more sensitive than those obtained by the inclusive dijet search. These results are complementary to the dedicated searches for the case that B and C are standard model bosons.

DOI: 10.1103/PhysRevLett.125.131801

A search for dijet resonances is one of the first analyses performed when a hadron collider reaches a new center-of-mass energy [1–8]. While such searches are sensitive to nearly all resonance decays $A \to BC$, dedicated searches for particular decays will always be more sensitive. This is the motivation for dedicated resonance searches for the case where B and C are τ leptons [9,10], b quarks [11–13], top

While it is crucial to continue searching for particular dijet topologies, the fact that not all SM and BSM possibilities are covered suggests that a complementary generic search effort is required. What is needed is a method for searching for many topologies all at once that ideally does not pay a large statistical trials factor. A variety of existing and proposed model-agnostic searches range

"Something anomalous"

Interpretability

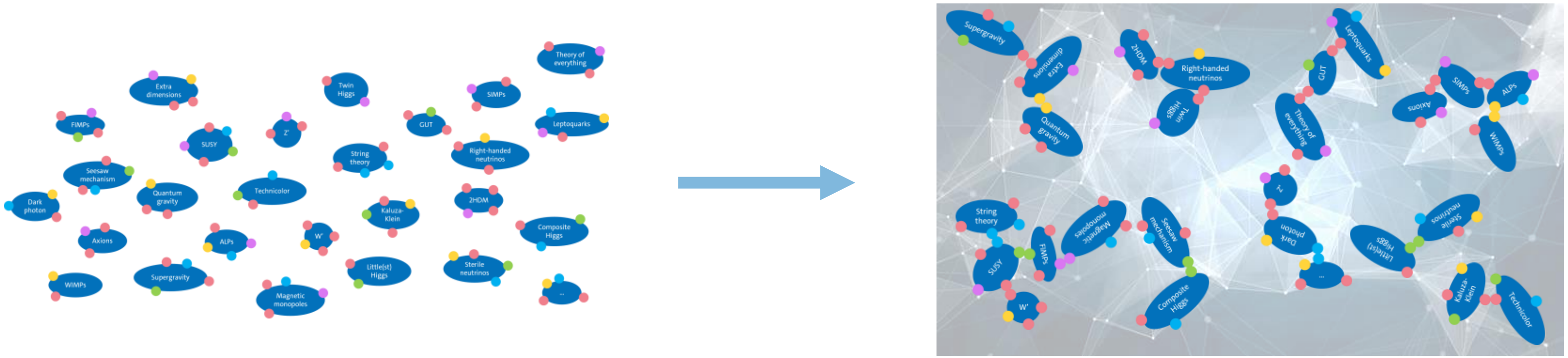
- By design, specific searches focus on small regions of model space
  - If we find something, great!
  - If we don't find anything, how can we use that result in the context of other models?
- Also by design, model-agnostic searches focuses on wide swaths of model space
  - If we find something, great!
  - ...but what is it that we have found?

Grouping models according to their phenomenology

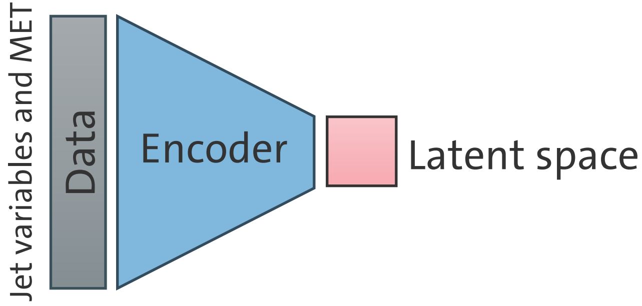
- What if there was a way to figure out which models ”belong together”, phenomenologically?
  - Results of very specific searches could be easier to interpret in the context of other models
  - Further interpretation of results from model-agnostic searches could be made easier
  - Potential blind spots in our search strategies could be identified

Universal new physics latent space

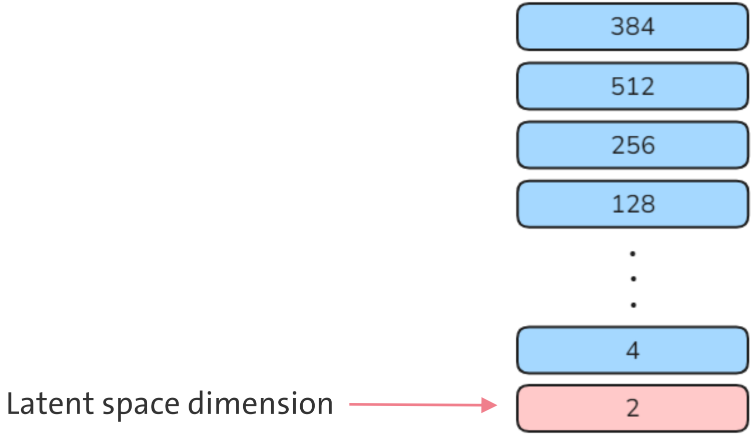
- Idea: use machine learning to embed the phenomenology of different models into a universal latent space
- The distance between the models in this latent space will indicate how close they are to one another, phenomenologically



Model architecture: simple encoder



Fully connected MLP with ReLu activations



Contrastive loss function

Model weights

X_i and X_j are concatenated sets of 10 events

$$\mathcal{L}(\theta) = \sum_{i \neq j} L(\theta, (Y_{ij}, \vec{X}_i, \vec{X}_j))$$

Label $Y = 1$ for points from the same model,
$Y = 0$ for points from different models

where

Parameter for unbounded
latent spaces; we set $d = 1$

$$L(\theta, (Y_{ij}, \vec{X}_i, \vec{X}_j)) = (1 - Y_{ij})\frac{1}{2}D_\theta^2 + Y_{ij}\frac{1}{2}\left(\max\left(0, d - D_\theta\right)\right)^2$$

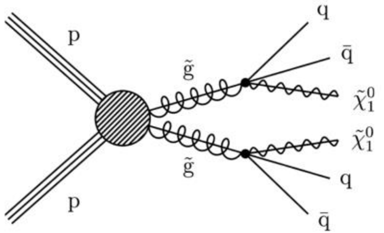
$$D_\theta \equiv D_\theta(\vec{X}_i, \vec{X}_j)$$

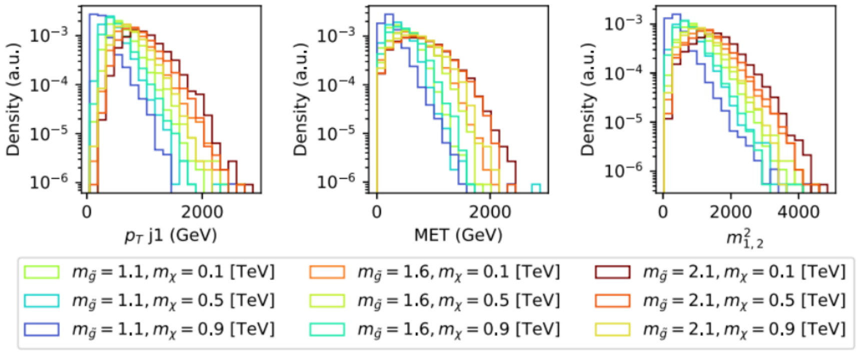
Euclidian distance
in latent space

The loss function is **minimized** when **points from the same model are close together** in the latent space, and **points from different models are far apart**. Since their phase spaces overlap, they cannot be collapsed to single points but must overlap in the latent space.

Dataset 1: MSSM gluino simplified model

- Pair production of gluinos, each gluino decays to a neutralino and a quark-antiquark pair
- Signal: four hard jets and large missing energy (MET)
- Nine mass combinations: $m_{gluino} = 1.1, 1.6, 2.1$ TeV; $m_{neutralino} = 0.1, 0.5, 0.9$ TeV
- Features used for training: $p_T, m, \eta, \Delta\phi$(jet, MET) of the first four jets, MET, and the invariant masses of all possible jet pairs

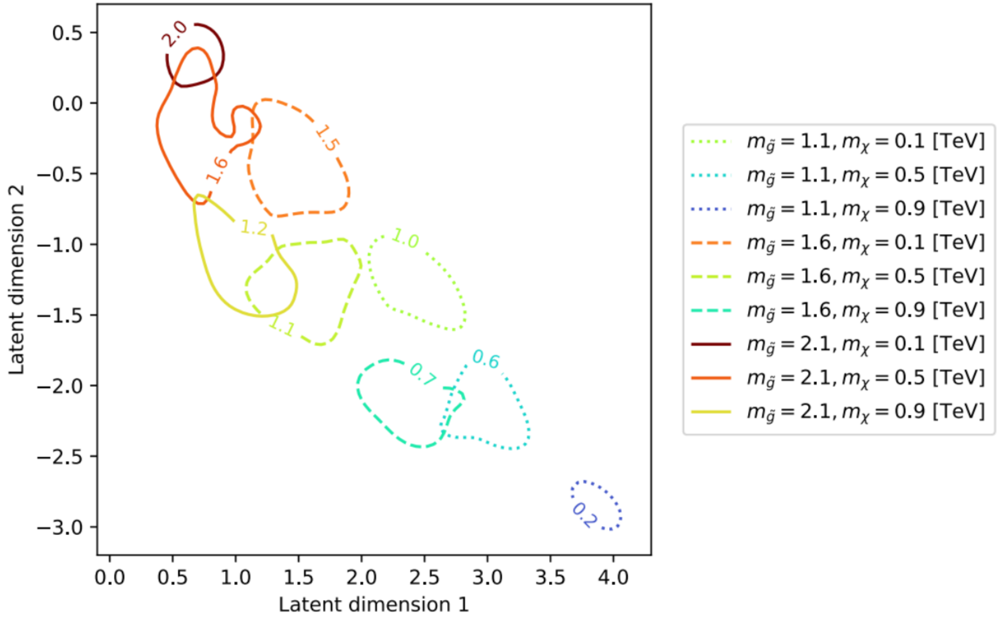




Observations:
- Large mass difference Δm between gluino and neutralino $\rightarrow$ harder jet and MET spectrum
- Models with similar Δm have similar spectra

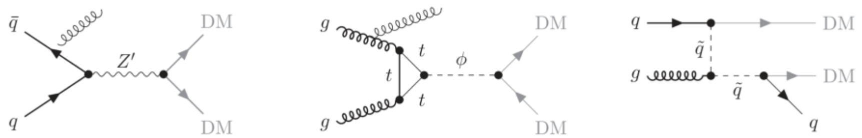
Dataset 1 results

- The model has organized the models in latent space given the mass difference (labeled) between gluino and neutralino
- Agrees with observation about this correlation from the feature plots
- Orthogonal diagonal shows a slight sorting of the models according to the gluino mass, nontrivial as the signal features only depend weakly on this

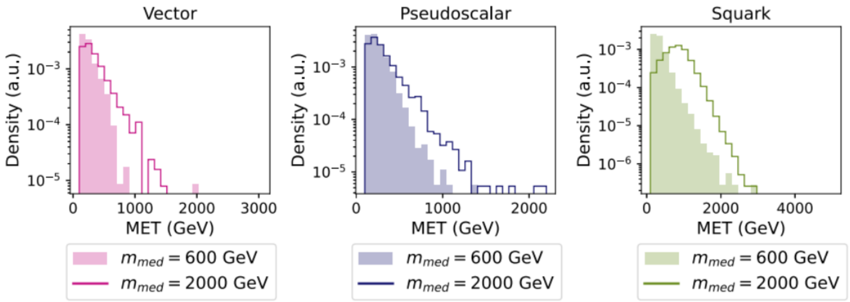


Dataset 2: Dark matter simplified model

- Dark matter production at the LHC, with one jet and missing energy
  - Vector mediator
  - Pseudoscalar mediator
  - Squark mediator



- m_{DM} = 100-900 GeV, in steps of 100 GeV; M_{med} = 600-2000 GeV, in steps of 200 GeV
- Features used for training: $p_T, \eta, \Delta\phi$(jet, MET), MET, and the mass of the leading two jets

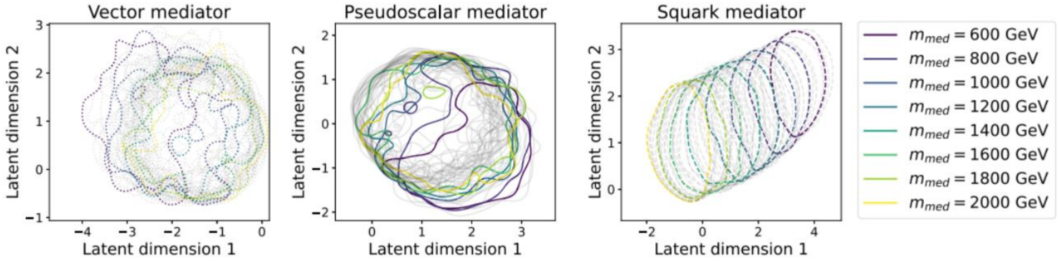


Observations for m_{DM} = 100 GeV:
- MET in vector and pseudoscalar case only weakly depends on the mediator mass
- In the squark case, the MET has a stronger correlation with the mediator mass

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Dataset 2 results

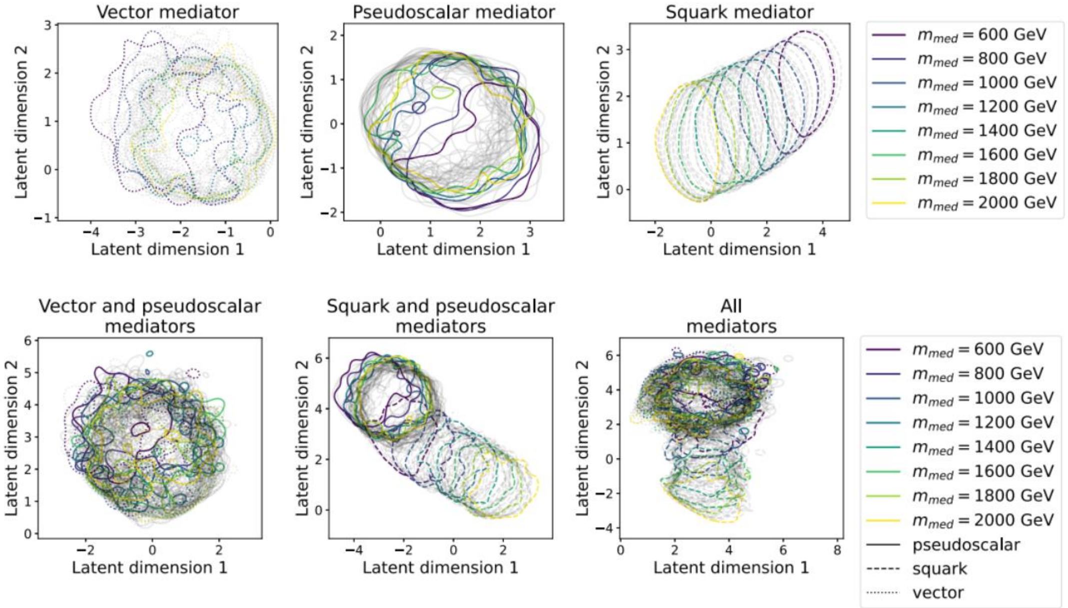
- Separate embeddings
  - Not much dependence on mediator mass in vector and pseudoscalar case
  - Squark case shows dependence on mediator mass



Colorful contours: m_{DM} = 100 GeV; grey contours: all other masses

Dataset 2 results

- Separate embeddings
  - Not much dependence on mediator mass in vector and pseudoscalar case
  - Squark case shows dependence on mediator mass
- Pair embeddings
  - Vector and pseudoscalar overlap
  - Squark case has distinct arrangement
- Embedding of all three
  - Vector and pseudoscalar still overlap
  - Squark case keeps its distinct behavior

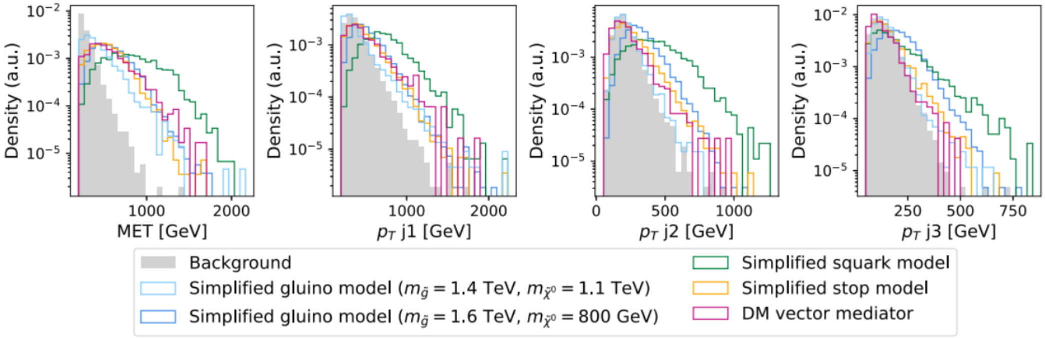


Colorful contours: m_{DM} = 100 GeV; grey contours: all other masses

Dataset 3: Dark machines subset

- A selection of models from the Dark machines anomaly challenge dataset (Dark Machines community 2020, Aarrestad et al 2022)
  - Scenarios resulting in jets + MET
  - SM background
- Features used for training: E, p_T, η, ϕ of the first four jets, MET of the event

BSM scenario	Physical process and model parameters
DM Vector Mediator	$pp \to Z' \to \chi\chi$ • $m_{Z'} = 2$ TeV, $m_{\mathrm{DM}} = 50$ GeV
Gluino Simplified Models	$pp \to \tilde{g}\tilde{g},\ \tilde{g} \to qq + \tilde{\chi}_1^0$ • $m_{\tilde{g}} = 1.4$ TeV, $m_{\chi^0} = 1.1$ TeV • $m_{\tilde{g}} = 1.6$ TeV, $m_{\chi^0} = 0.8$ TeV
Stop Simplified Model	$pp \to \tilde{t}\tilde{t},\ \tilde{t} \to t + \tilde{\chi}_1^0$ • $m_{\tilde{t}} = 1$ TeV, $m_{\chi^0} = 0.3$ TeV
Squark Simplified Model	$pp \to \tilde{q}\tilde{q},\ \tilde{q} \to q + \tilde{\chi}_1^0$ • $m_{\tilde{q}} = 1.8$ TeV, $m_{\chi^0} = 0.8$ TeV

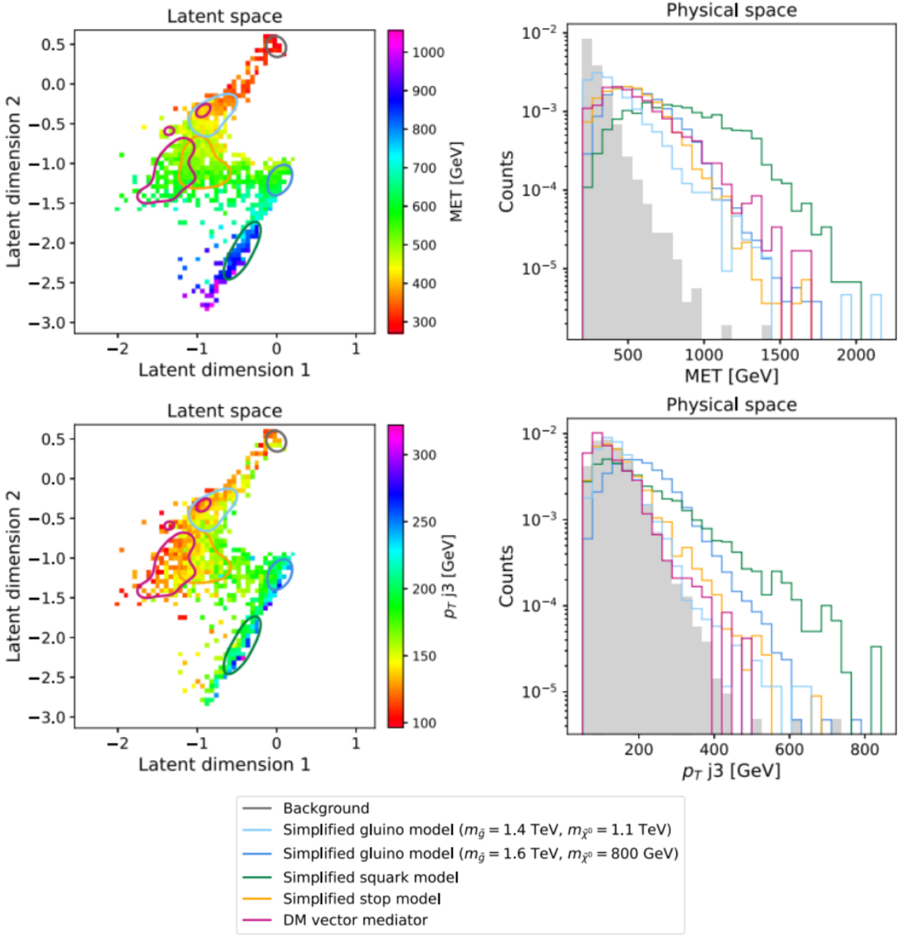


Observations:

- Squark model has the highest $\Delta m \to$ hard MET and p_T spectra
- Heavy gluino model has 4 hard jets $\to$ high p_T of the third jet compared to other models
- Light gluino model has smallest $\Delta m \to$ soft spectra
- Z' model: jets come from ISR, not dependent on its high Δm

Dataset 3 results

- The latent space is organized along two main axes:
  - MET
  - p_T of the third jet
- Light gluino, Z’ and stop models are clustered together, as expected from ISR and soft spectrum of light gluino; light gluino has softer MET, putting it higher in latent dim 2
- Heavy gluino: high p_T of the third jet, intermediate MET
- Squark: highest MET, also hard p_T spectra
- SM background: softest spectra overall



Conclusions

- We developed a machine learning model capable of clustering different BSM scenarios in a latent space according to their phenomenology
- When analyzing the arrangement in latent space, it was obvious that the model had found relevant axes
  - By analyzing such latent spaces, representative models that cover a certain area of the latent space can be selected as benchmarks for searches
  - Correlations between latent space axes and feature space can be used to identify observables suitable for distinguishing different models
  - Regions not covered in latent space could indicate gaps that need to be filled
- Future work: include cross section; explore more expressive architectures...