Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

Universal New Physics Latent Space

(2407.20315)

Anna Hallin, Gregor Kasieczka, Sabine Kraml, André Lessa,
Louis Moureaux, **Tore von Schwartz**, David Shih

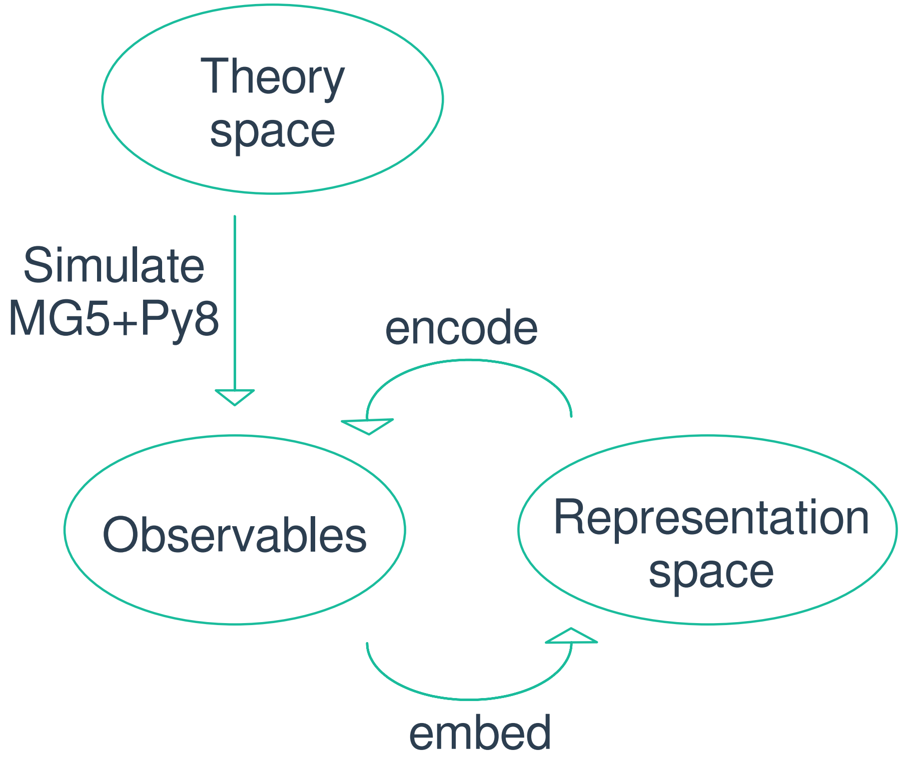
16.12.24

tore.von.schwartz@cern.ch

Glühweinworkshop 2024

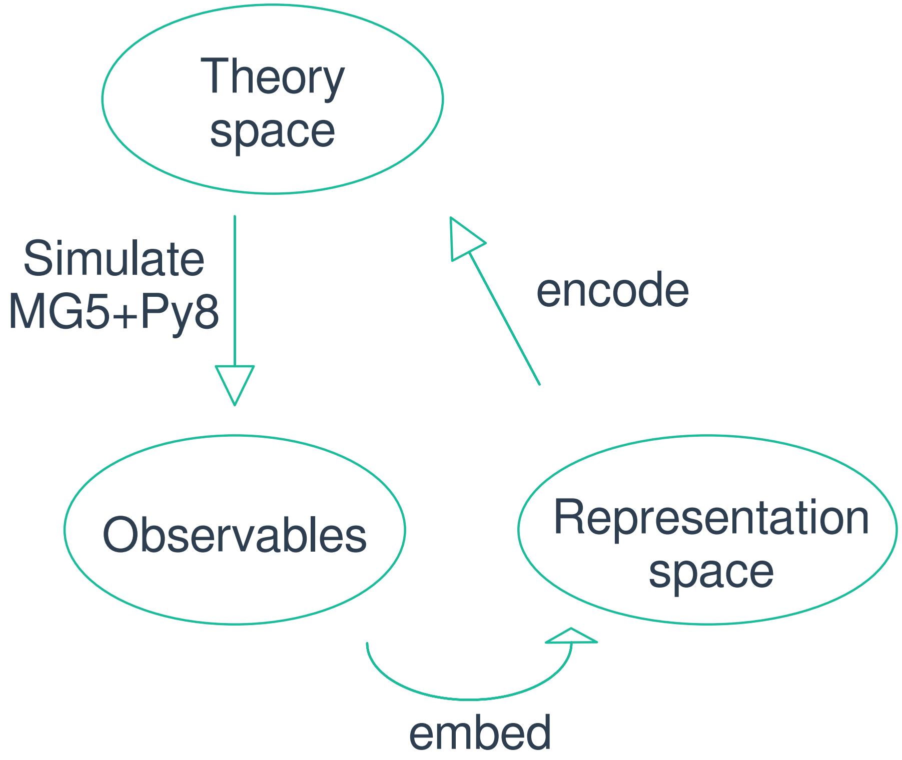
Dimensionality Reduction for Physical Data

- „*Neural Embedding: Learning the Embedding of the Manifold of Phyics Data*", S. E. Park et al. (2208.05484)
- **Embedding in lower dimensional latent space while conserving energy mover's distance between events**



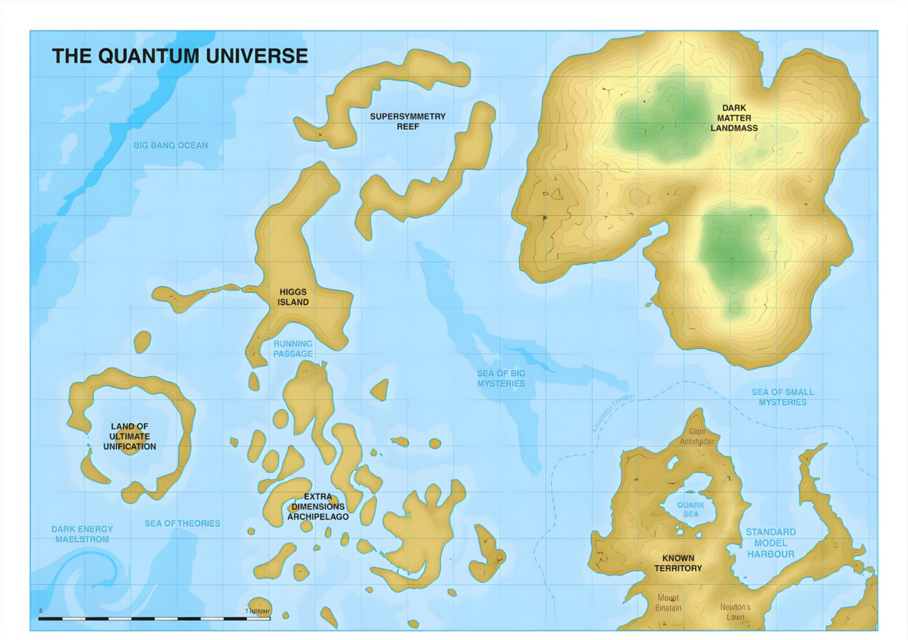
Dimensionality Reduction for Physical Data

- **Encode information about underlying theory instead of the actual events**
- **Learn embedding based on phenomenological similarities**



Universal New Physics Latent Space

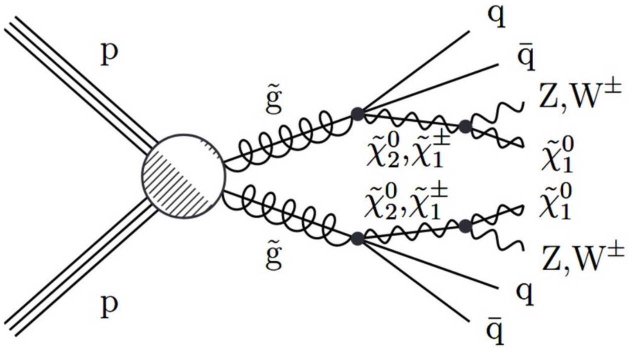
- **New physics data available from simulations**
- **Embed data from different theories in same latent space**
- **Investigate phenomenological similarities in low-dimensional space**

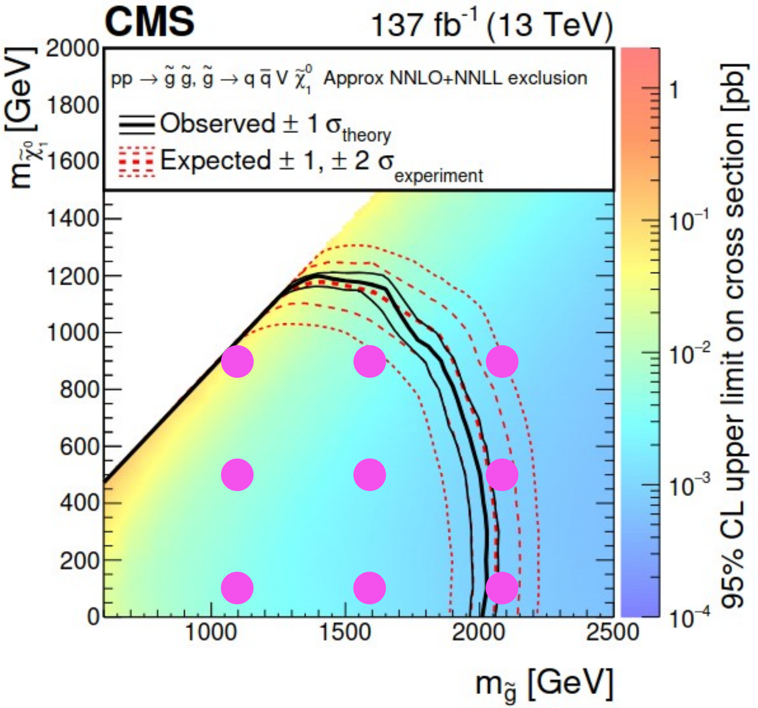


https://cds.cern.ch/record/1601971/files/ILCTDR-OUTREACH.pdf

Data Set

- **Simulation of SUSY events for different mass parameters**
- **Events are gluino decays in proton-proton collisions at 13 TeV**
- **Use kinematic features of leading four jets, missing transverse energy and invariant dijet masses**





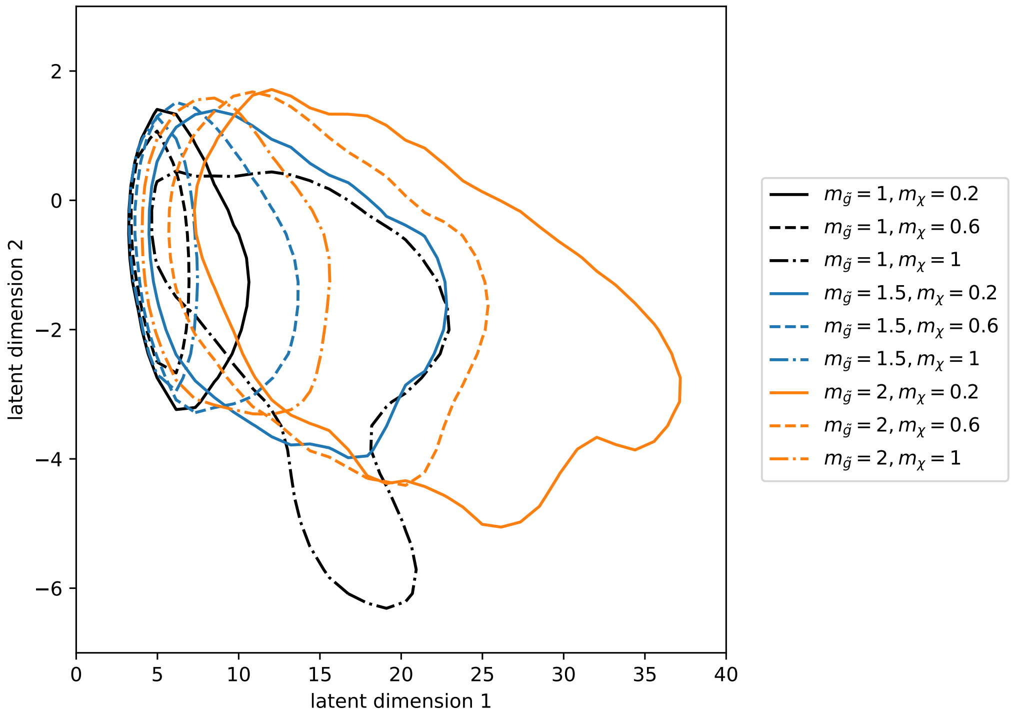
arXiv:1908.04722

Baseline

- **Embed data with autoencoder trained on mean squared error loss**

$$\mathcal{L}_{MSE} = \frac{1}{N}\sum_{i}^{N}(x_i - \hat{x}_i)^2$$

- **No clustering but slight shift in resulting structure**
- **Overlapping feature distributions → overlapping distributions in latent space**

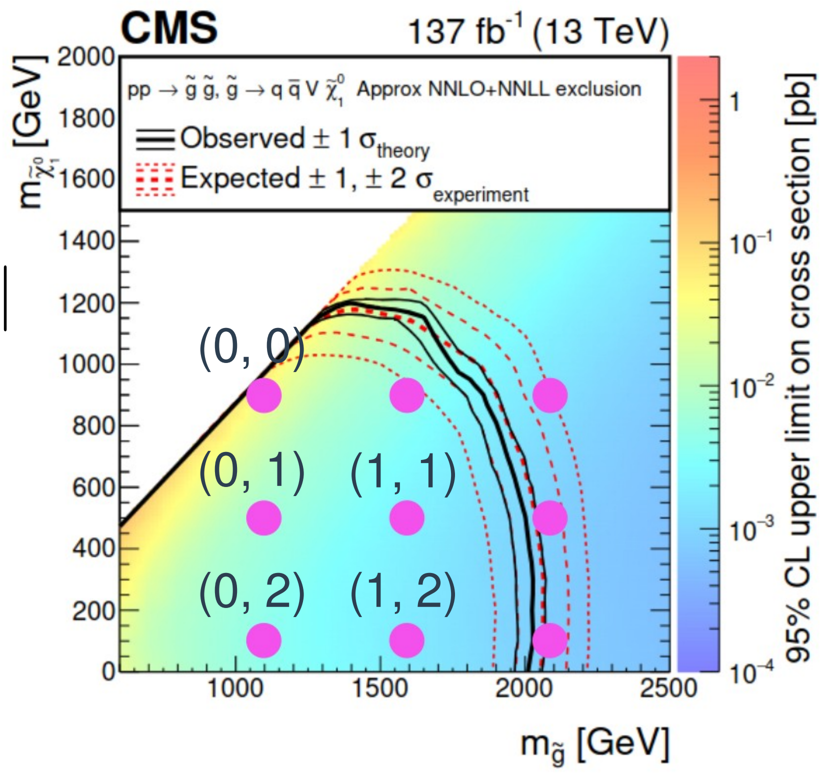


Neural Embedding (NE)

- **Additional loss term to ensure clustering**
- **Conserve metric between pair of events in latent space (arXiv:2208.05484)**

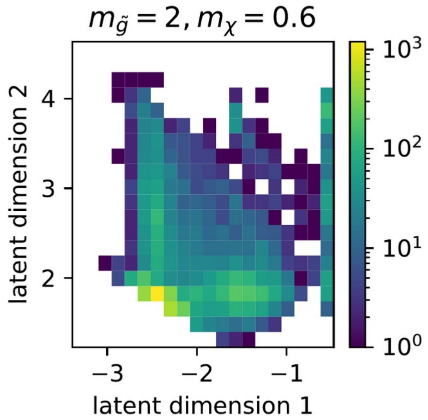
$$\mathcal{L}_{NE} \sim \left| d_{\mathcal{Y}}\left(\phi(u_i), \phi(v_i)\right) - d_{\mathcal{X}}(u_i, v_i) \right|$$

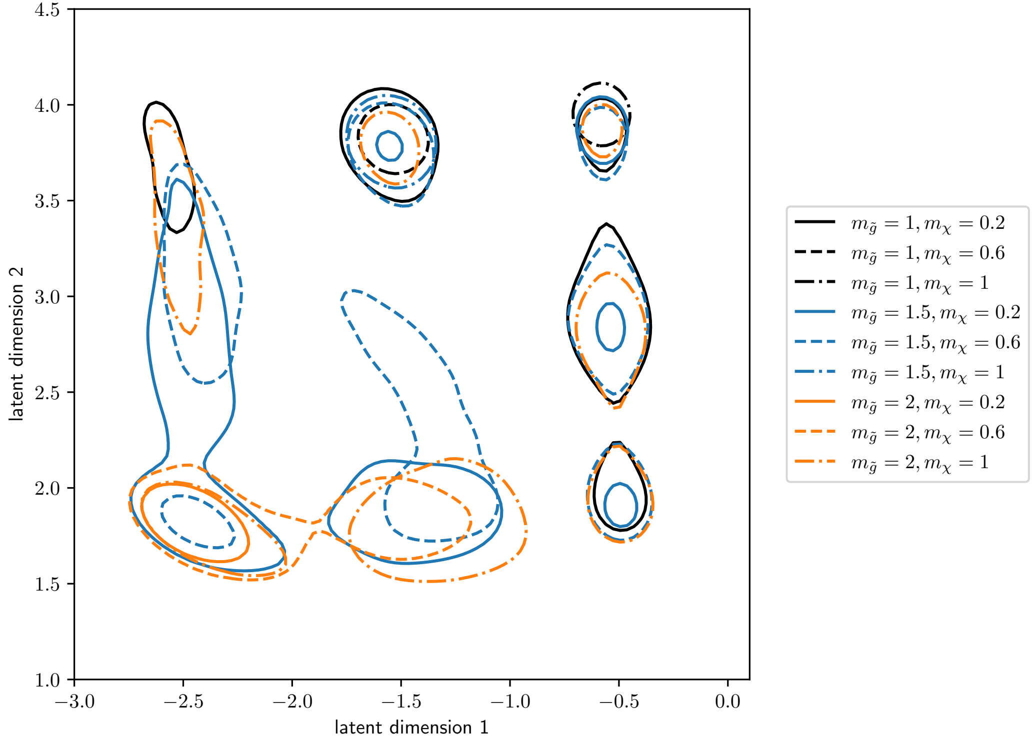
- **Here: conserve distance between data set labels**



Results Neural Embedding

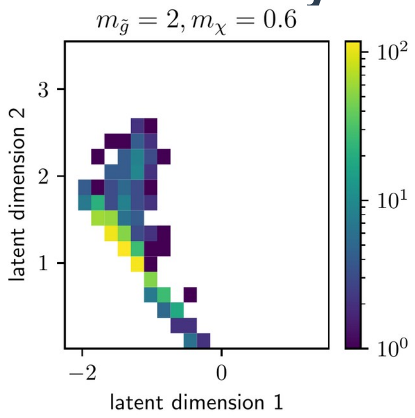
- **Clustering not perfect but overall structure visible**
- **Arrangement of clusters similar to mass space**
- **Embedding accuracy of ~43%**

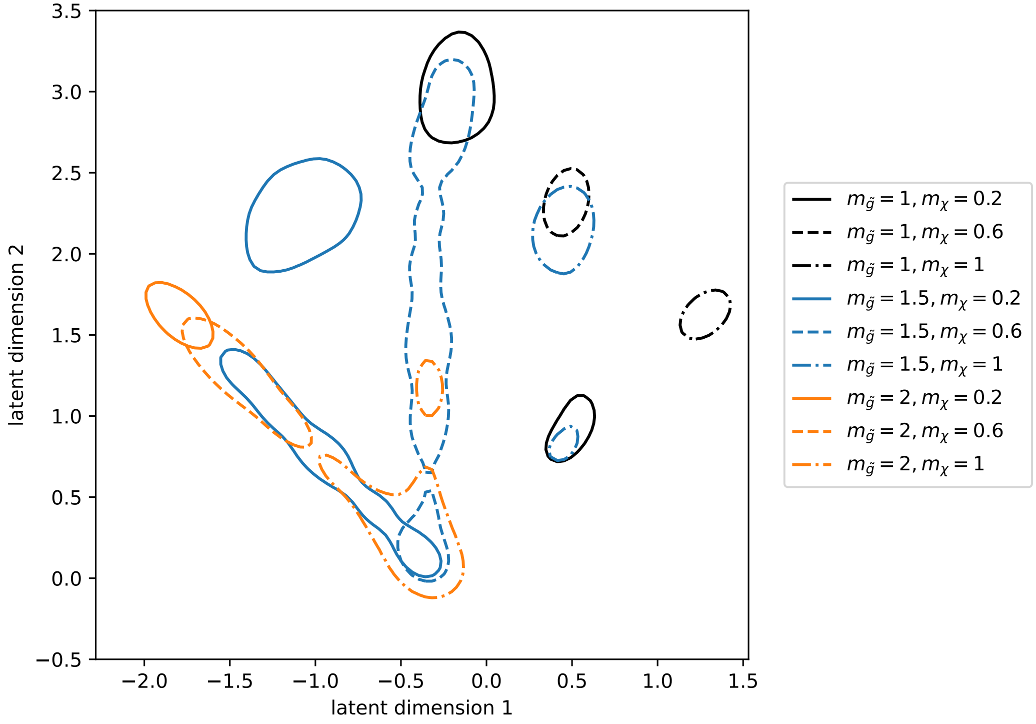




Training on Sets of Events

- **Take sets of events from same theory as input**
- **Increases probability of the model to see distinguishable events**
- **Leads better structured latent space and accuracy of ~70%**





Contrastive Learning

- **Explicit and comparable data set labels not always given**

  → replace NE loss term with contrastive loss term

  (Dimensionality reduction by learning an invariant mapping, R. Hadsell et al.)

$$\mathcal{L}_i = (1 - Y)\mathcal{L}_S(D_i) + Y\mathcal{L}_D(D_i)$$

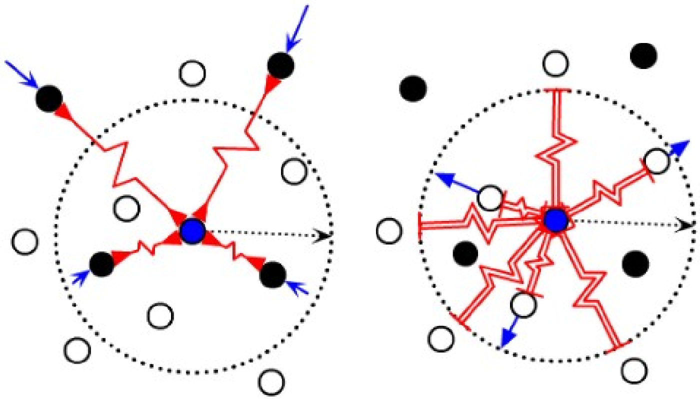
- **Goal: cluster similar points and separate dissimilar points without knowing exact arrangement of data sets**

Contrastive Learning

- **Exact choice of loss function**

$$\mathcal{L}_i = (1-Y)\frac{1}{2}D_i^2 + Y\frac{1}{2}\max(0, m - D_i)^2$$

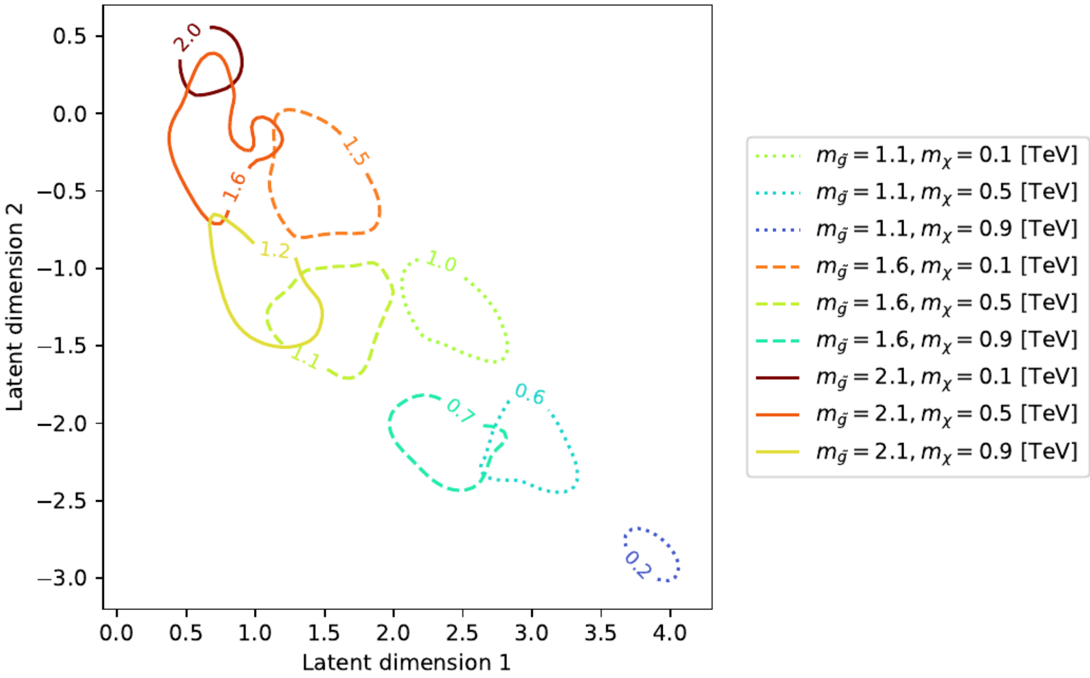
- **New margin parameter needed to deal with unbounded latent spaces**
- **Events from different theories with latent distance larger than margin parameter not longer repelled**



Dimensionality reduction by learning an invariant mapping, R. Hadsell et al.

Results Contrastive Learning

- **Clearly visible clustering**
- **Latent space organisation based on mass difference**
- **Arrangement of clusters stable over multiple trainings**
  → based on physical properties



Dark Machines Data Set

- **Application of this method to Dark Machines data set including a larger variety of processes (2105.14027)**
- **Chosen signals focus on hadronic activity with high missing energy**
  - At least 4 (b-)jets with $p_T > 50$ GeV
  - One (b-)jet with $p_T > 200$ GeV
  - $H_T > 600$ GeV
  - MET > 200 GeV and MET / $H_T \geq 0.2$
- **Background data set containing SM events with same trigger requirements**

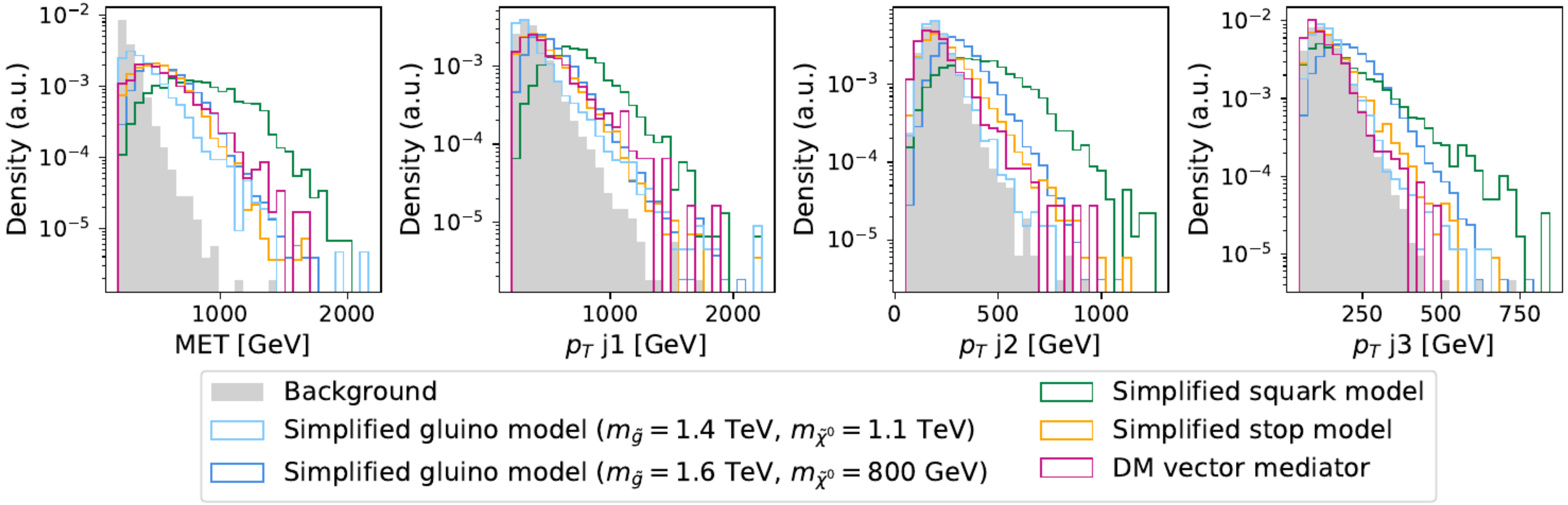
Dark Machines Data Set

- **Five BSM models**
- **Including two gluino models with different mass configurations**
- **Third jet originating from ISR for all models except for gluino models**

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

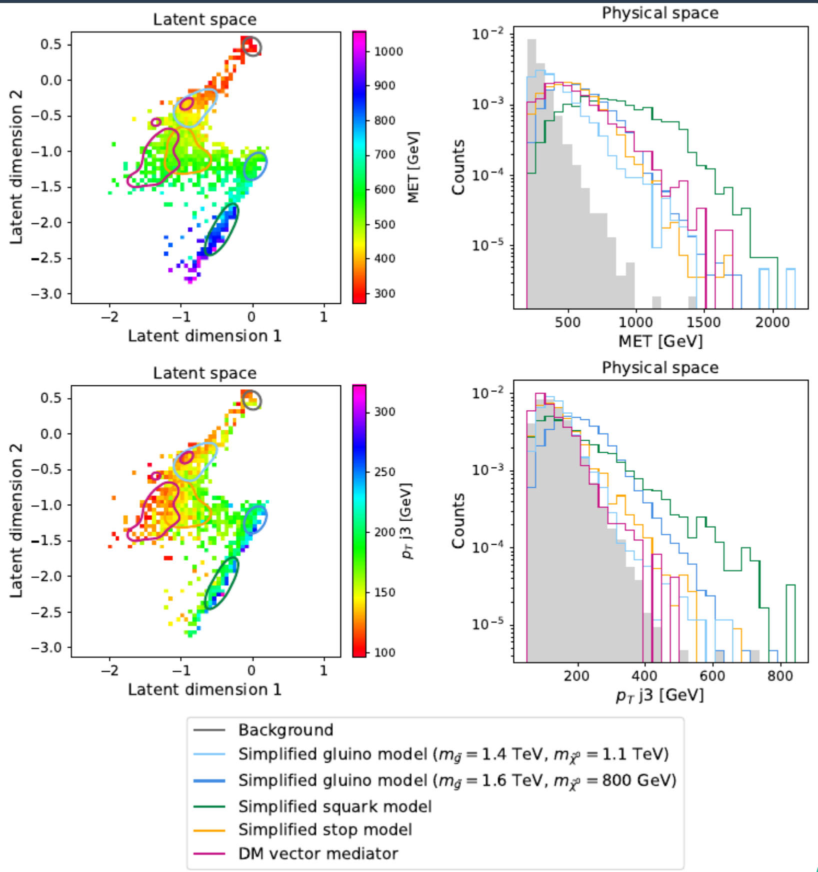
Dark Machines Data Set

- **SM background MET distribution different compared to BSM models**
- **Hardest MET distribution for squark model, softest for light gluino**



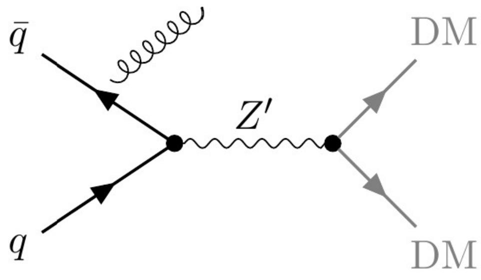
Results Dark Machines Data Set

- **Embedding based on MET and p_T of third jet**
- **SM background, squark model and light gluino model separated best**
- **Similar models are clustered in the latent space**

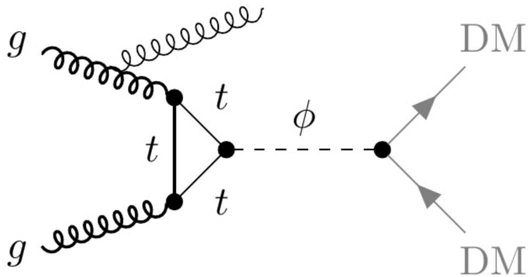


Mediator Data Set

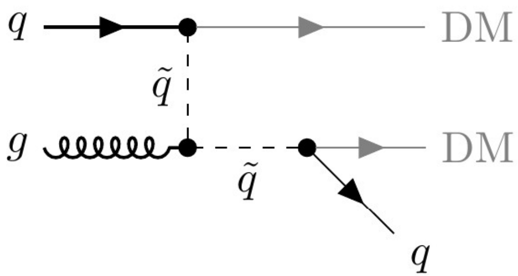
- **Simplified BSM theories with dark matter particle and BSM mediator particle**
- **Kinematic features of leading two jets used for training**
- **Simulation for different mass combinations**



Vector mediator



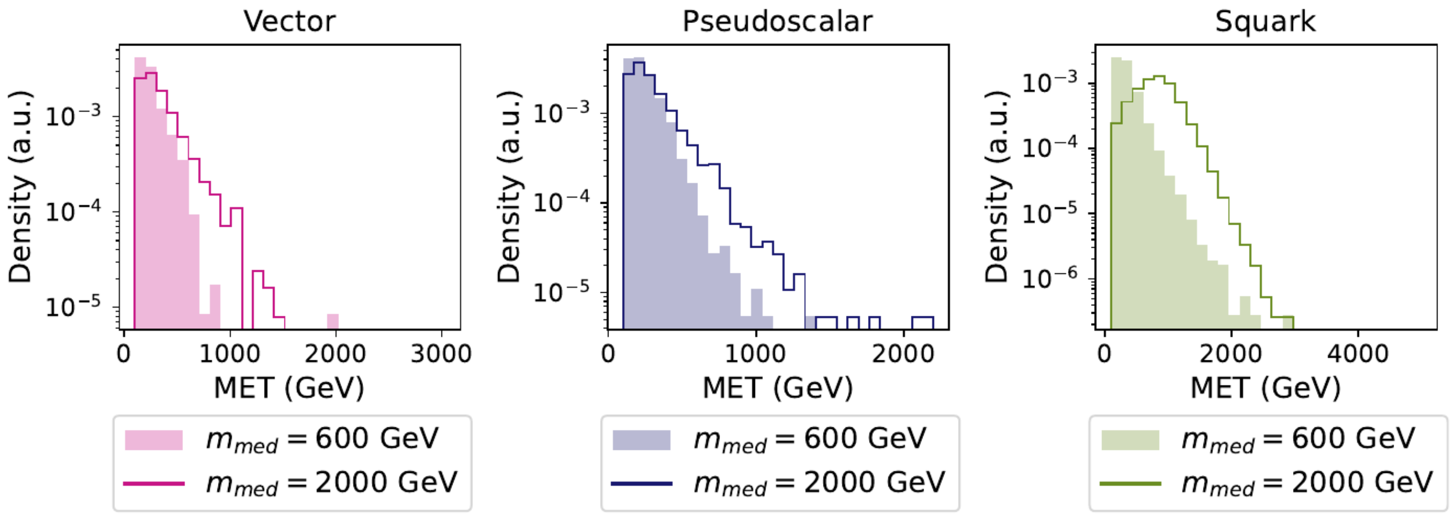
Pseudocalar mediator



Colored mediator

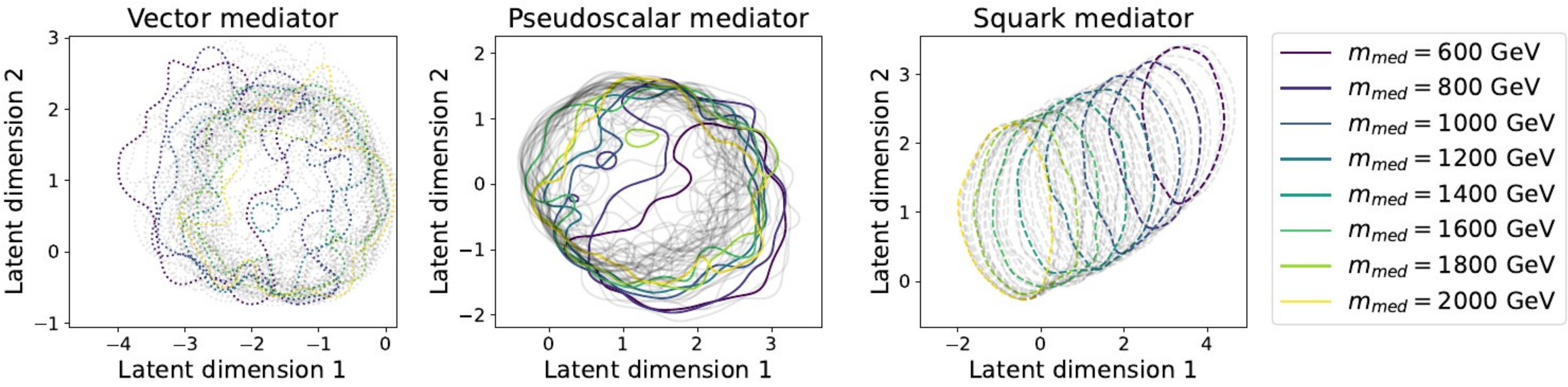
Mediator Data Set

- **Similar MET peak for vector and pseudoscalar mediator for all mediator masses**
- **Shifting peak towards higher MET for squark mediator particles with larger mass**



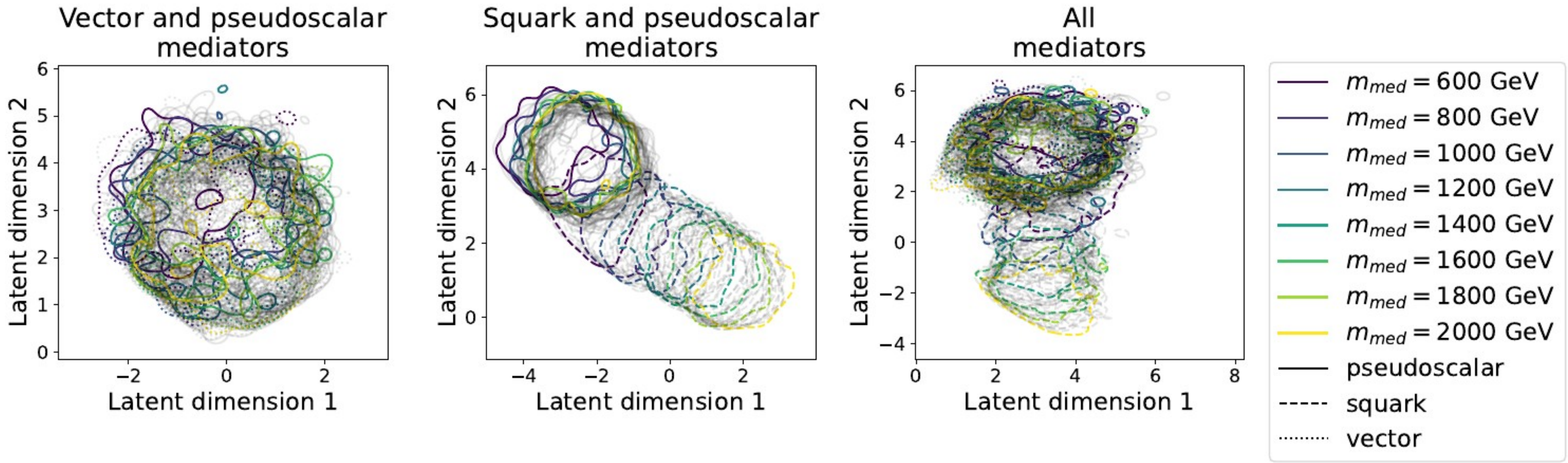
Results Mediator Data Set

- Separate trainings for different mediator particles
- Distributions of individual mass configurations heavily overlapping for vector and pseudoscalar mediator
- Better separation for squark mediator



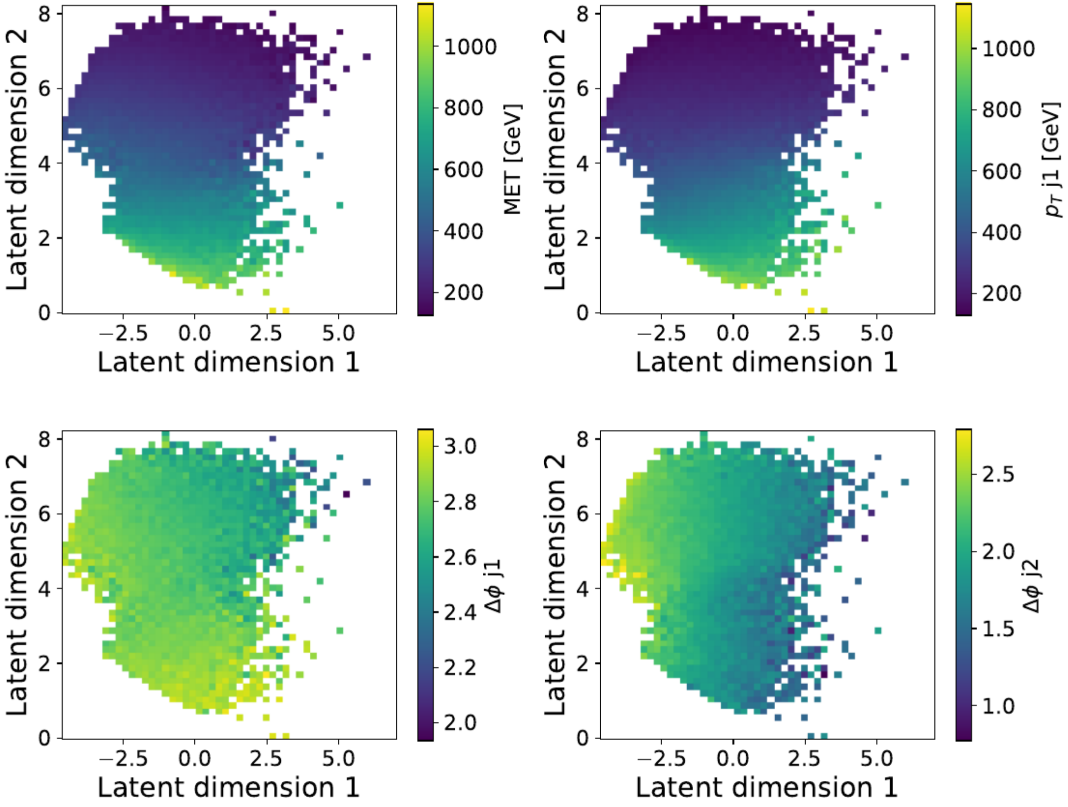
Results Mediator Data Set

- **Embedding of different mediator types into one common latent space**
- **Global structure of different theories while maintaining internal arrangement**



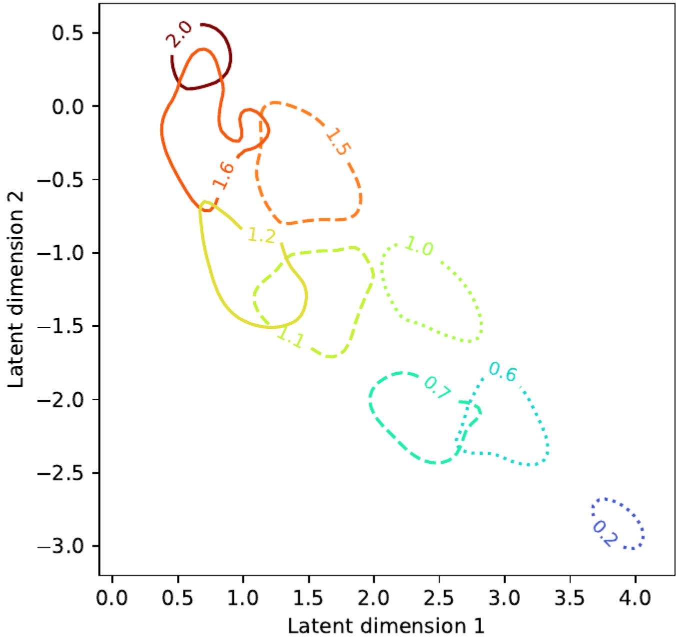
Comparing to Input Features

- **Shown latent spaces for embedding of all mediator particles**
- **Connection between input features and latent space visible**
- **Embedding organized based on MET and $\Delta\phi_2$**



Conclusion

- Learning similarities of theories based on their phenomenology
- Sets as input to abstract from individual events
- Setup applicable to various data sets even without explicit relation in theory space
- Latent space allows to connect low-dimensional representation with theory and feature space
- For more details check out 2407.20315



Reconstruction of Input Features

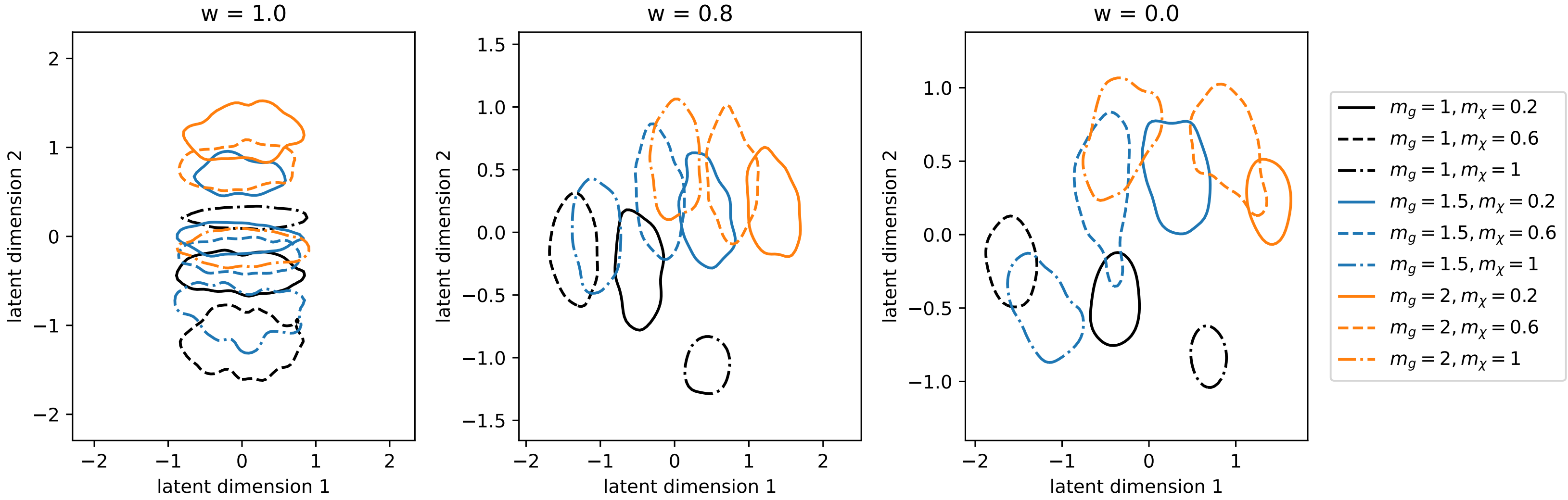
- **Prevention of sparsely populated latent space regions for latent space interpolation**
- **Change model to variational autoencoder with Kullback-Leibler divergence**

$$\mathcal{L} = w \cdot \mathcal{L}_{MSE} + (1 - w) \cdot \mathcal{L}_{CL} + \beta \cdot D_{KL}$$

- **Forces a Gaussian latent space structure**

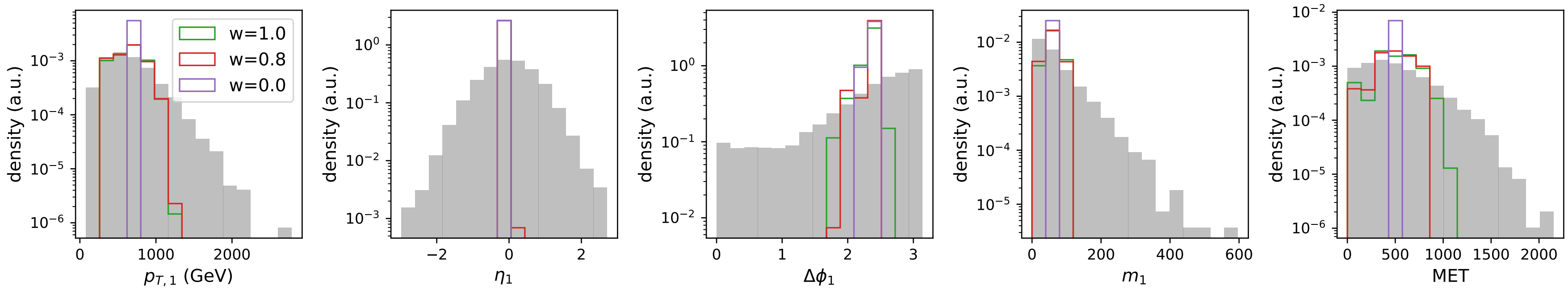
Reconstruction of Input Features

- **Results for SUSY data with different loss weightings**
- **Lower weight improves separation and embedding accuracy**
- **Clustering already visible for higher weight**



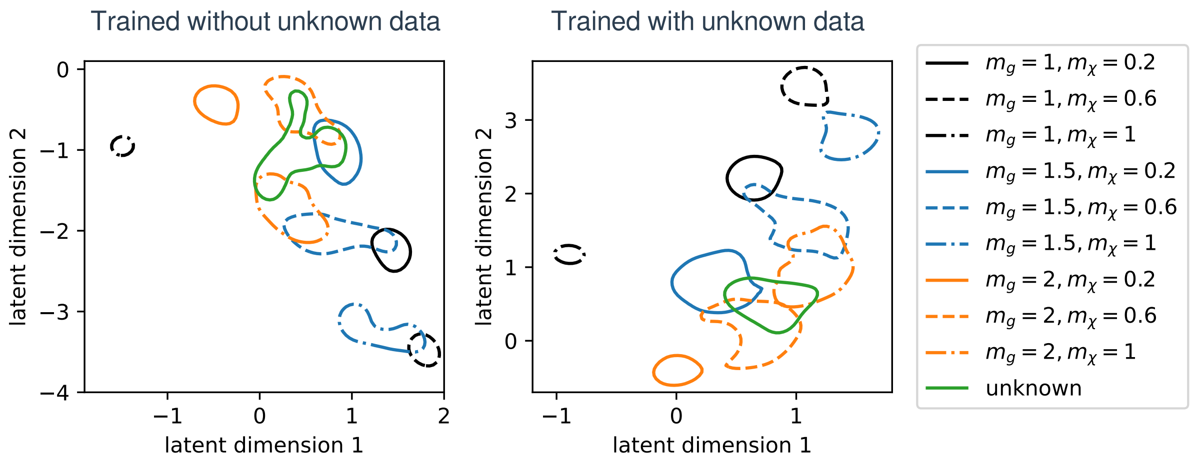
Reconstructing input features

- No reconstruction of the tails of the distributions for all weightings
- Highest weight of the MSE loss term results in best reconstruction
- Could be improved by changing model architecture or number of events per set



From Latent to Theory Space

- **Reconstruct theory parameters from latent space distributions for unknown theory**
- **Prediction for every unknown sample based on kernel density estimation**



From Latent to Theory Space

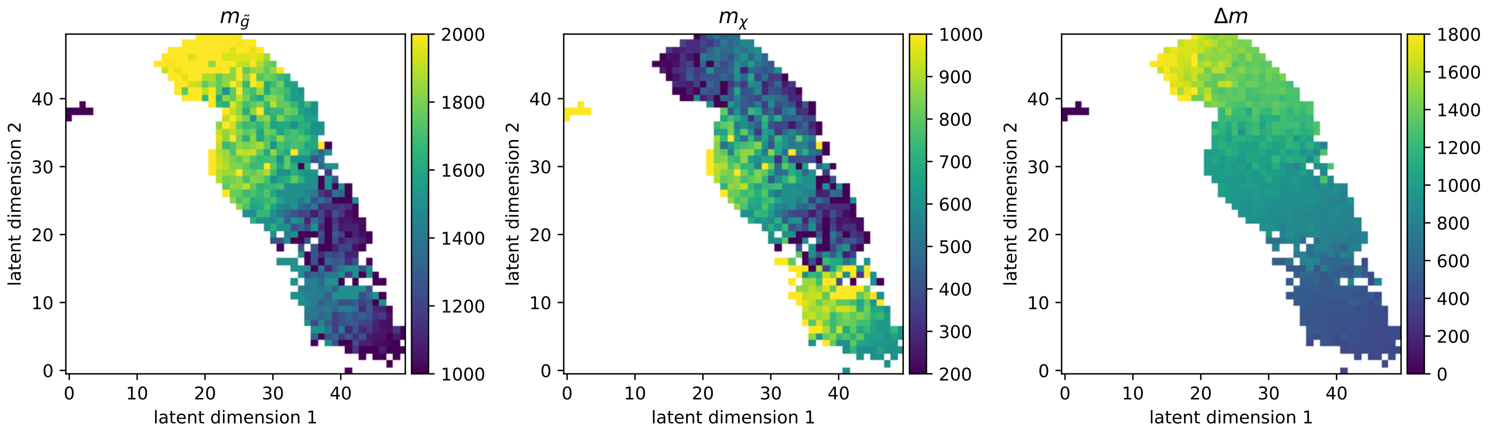
- **Gluino mass prediction extremely precise**
- **For a better understanding of the method further investigations necessary**

	Gluino mass (TeV)	Neutralino mass (TeV)	Mass difference (TeV)
True value	1.83	0.8	1.03
Estimation based on all data	1.82 ± 0.12	0.58 ± 0.15	1.25 ± 0.14
Estimation based on known data	1.81 ± 0.12	0.57 ± 0.17	1.25 ± 0.15

Results Contrastive Learning

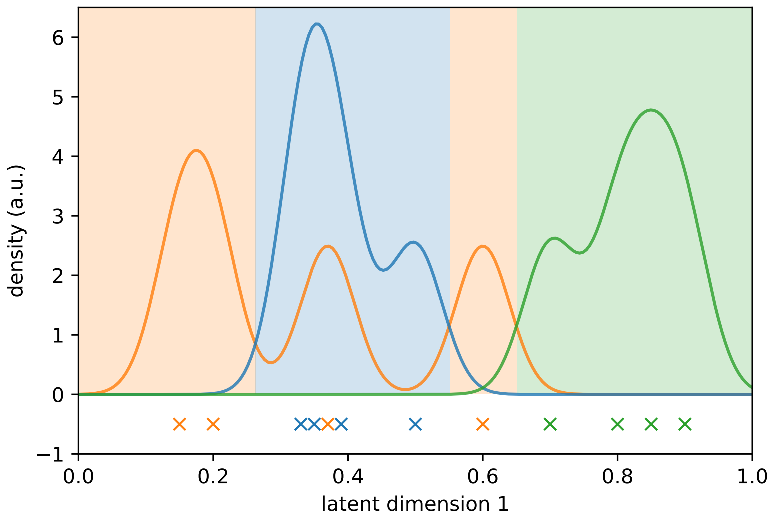
- **Apply binning and color every bin based on mean of gluino mass, neutralino mass or mass difference of events mapped into it**

  → smoothest transition for mass difference



Performance Measurement

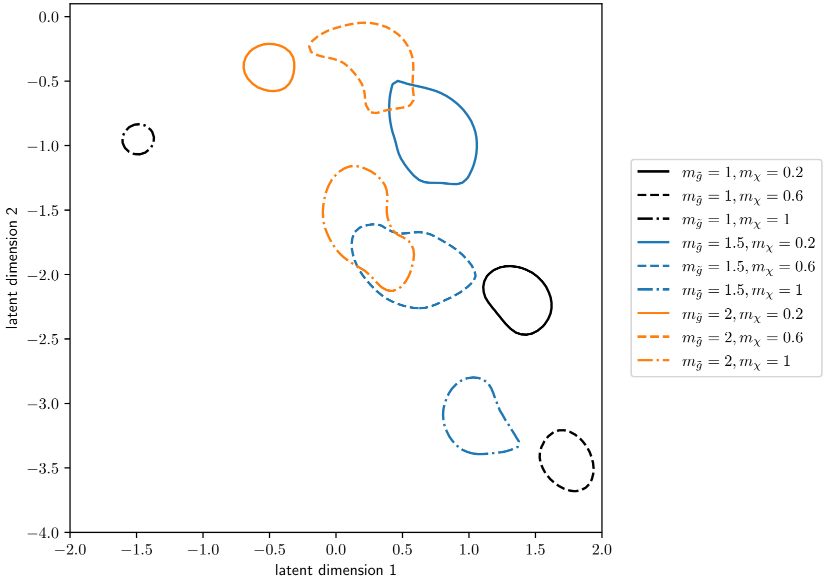
- Comparison with correct output not possible
- Assignment of latent space regions to specific classes based on kernel density estimation
- Allows calculation of “accuracy“



Results Contrastive Learning

- **Comparison of latent space to accuracies of binary classifiers**
- **Theories with lowest accuracy closest together and vice versa**

  → latent space based on phenomenological similarity of the theories

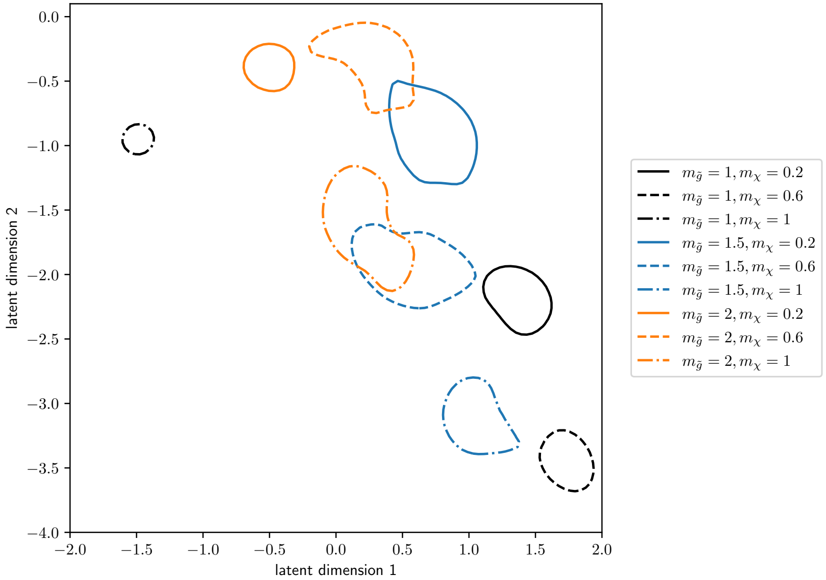


	$m_{\tilde{g}}=1, m_\chi=0.2$	$m_{\tilde{g}}=1, m_\chi=0.6$	$m_{\tilde{g}}=1, m_\chi=1$	$m_{\tilde{g}}=1.5, m_\chi=0.2$	$m_{\tilde{g}}=1.5, m_\chi=0.6$	$m_{\tilde{g}}=1.5, m_\chi=1$	$m_{\tilde{g}}=2, m_\chi=0.2$	$m_{\tilde{g}}=2, m_\chi=0.6$	$m_{\tilde{g}}=2, m_\chi=1$
$m_{\tilde{g}}=1, m_\chi=0.2$	0.500	0.969	1.000	0.990	0.868	0.904	1.000	1.000	0.964
$m_{\tilde{g}}=1, m_\chi=0.6$	0.969	0.500	1.000	1.000	1.000	0.826	1.000	1.000	1.000
$m_{\tilde{g}}=1, m_\chi=1$	1.000	1.000	0.500	1.000	1.000	1.000	1.000	1.000	1.000
$m_{\tilde{g}}=1.5, m_\chi=0.2$	0.990	1.000	1.000	0.500	0.894	1.000	0.943	0.822	0.801
$m_{\tilde{g}}=1.5, m_\chi=0.6$	0.868	1.000	1.000	0.894	0.500	0.977	1.000	0.974	0.760
$m_{\tilde{g}}=1.5, m_\chi=1$	0.904	0.826	1.000	1.000	0.977	0.500	1.000	1.000	1.000
$m_{\tilde{g}}=2, m_\chi=0.2$	1.000	1.000	1.000	0.943	1.000	1.000	0.500	0.823	0.990
$m_{\tilde{g}}=2, m_\chi=0.6$	1.000	1.000	1.000	0.822	0.974	1.000	0.823	0.500	0.907
$m_{\tilde{g}}=2, m_\chi=1$	0.964	1.000	1.000	0.801	0.760	1.000	0.990	0.907	0.500

Results Contrastive Learning

- **Comparison of latent space to accuracies of binary classifiers**
- **Theories with lowest accuracy closest together and vice versa**

  → latent space based on phenomenological similarity of the theories

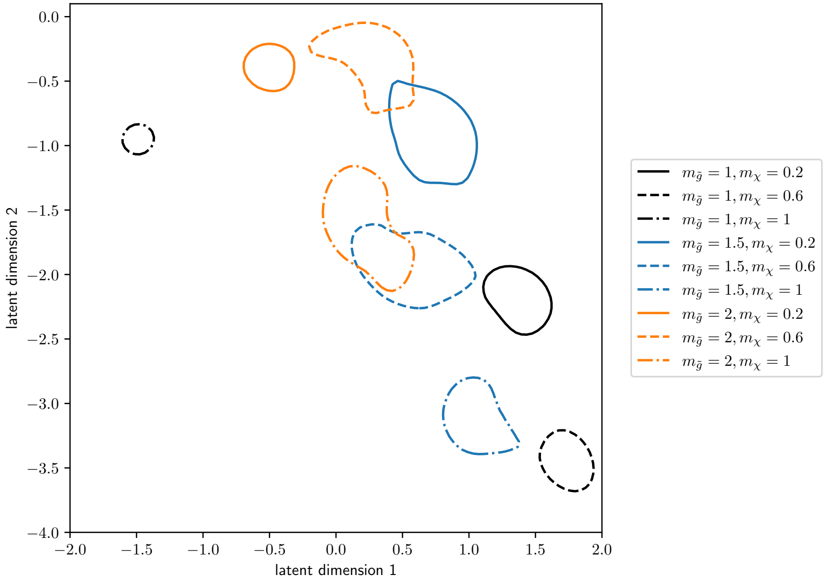


	$m_{\tilde g}=1, m_\chi=0.2$	$m_{\tilde g}=1, m_\chi=0.6$	$m_{\tilde g}=1, m_\chi=1$	$m_{\tilde g}=1.5, m_\chi=0.2$	$m_{\tilde g}=1.5, m_\chi=0.6$	$m_{\tilde g}=1.5, m_\chi=1$	$m_{\tilde g}=2, m_\chi=0.2$	$m_{\tilde g}=2, m_\chi=0.6$	$m_{\tilde g}=2, m_\chi=1$
$m_{\tilde g}=1, m_\chi=0.2$	0.500	0.969	1.000	0.990	0.868	0.904	1.000	1.000	0.964
$m_{\tilde g}=1, m_\chi=0.6$	0.969	0.500	1.000	1.000	1.000	0.826	1.000	1.000	1.000
$m_{\tilde g}=1, m_\chi=1$	1.000	1.000	0.500	1.000	1.000	1.000	1.000	1.000	1.000
$m_{\tilde g}=1.5, m_\chi=0.2$	0.990	1.000	1.000	0.500	0.894	1.000	0.943	0.822	0.801
$m_{\tilde g}=1.5, m_\chi=0.6$	0.868	1.000	1.000	0.894	0.500	0.977	1.000	0.974	0.760
$m_{\tilde g}=1.5, m_\chi=1$	0.904	0.826	1.000	1.000	0.977	0.500	1.000	1.000	1.000
$m_{\tilde g}=2, m_\chi=0.2$	1.000	1.000	1.000	0.943	1.000	1.000	0.500	0.823	0.990
$m_{\tilde g}=2, m_\chi=0.6$	1.000	1.000	1.000	0.822	0.974	1.000	0.823	0.500	0.907
$m_{\tilde g}=2, m_\chi=1$	0.964	1.000	1.000	0.801	0.760	1.000	0.990	0.907	0.500

Results Contrastive Learning

- **Comparison of latent space to accuracies of binary classifiers**
- **Theories with lowest accuracy closest together and vice versa**

→ latent space based on phenomenological similarity of the theories



	$m_{\tilde g}=1, m_\chi=0.2$	$m_{\tilde g}=1, m_\chi=0.6$	$m_{\tilde g}=1, m_\chi=1$	$m_{\tilde g}=1.5, m_\chi=0.2$	$m_{\tilde g}=1.5, m_\chi=0.6$	$m_{\tilde g}=1.5, m_\chi=1$	$m_{\tilde g}=2, m_\chi=0.2$	$m_{\tilde g}=2, m_\chi=0.6$	$m_{\tilde g}=2, m_\chi=1$
$m_{\tilde g}=1, m_\chi=0.2$	0.500	0.969	1.000	0.990	0.868	0.904	1.000	1.000	0.964
$m_{\tilde g}=1, m_\chi=0.6$	0.969	0.500	1.000	1.000	1.000	0.826	1.000	1.000	1.000
$m_{\tilde g}=1, m_\chi=1$	1.000	1.000	0.500	1.000	1.000	1.000	1.000	1.000	1.000
$m_{\tilde g}=1.5, m_\chi=0.2$	0.990	1.000	1.000	0.500	0.894	1.000	0.943	0.822	0.801
$m_{\tilde g}=1.5, m_\chi=0.6$	0.868	1.000	1.000	0.894	0.500	0.977	1.000	0.974	0.760
$m_{\tilde g}=1.5, m_\chi=1$	0.904	0.826	1.000	1.000	0.977	0.500	1.000	1.000	1.000
$m_{\tilde g}=2, m_\chi=0.2$	1.000	1.000	1.000	0.943	1.000	1.000	0.500	0.823	0.990
$m_{\tilde g}=2, m_\chi=0.6$	1.000	1.000	1.000	0.822	0.974	1.000	0.823	0.500	0.907
$m_{\tilde g}=2, m_\chi=1$	0.964	1.000	1.000	0.801	0.760	1.000	0.990	0.907	0.500