A data-driven and model-agnostic approach to solving combinatorial assignment problems in searches for new physics

Javier Montejo Berlingen,

based on 2309.05728, in collaboration with Anthony Badea

4. A **data-driven** and
3. **model-agnostic** approach to solving
2. **combinatorial assignment problems** in
1. **searches for new physics**

Javier Montejo Berlingen,

based on 2309.05728, in collaboration with Anthony Badea

Introduction

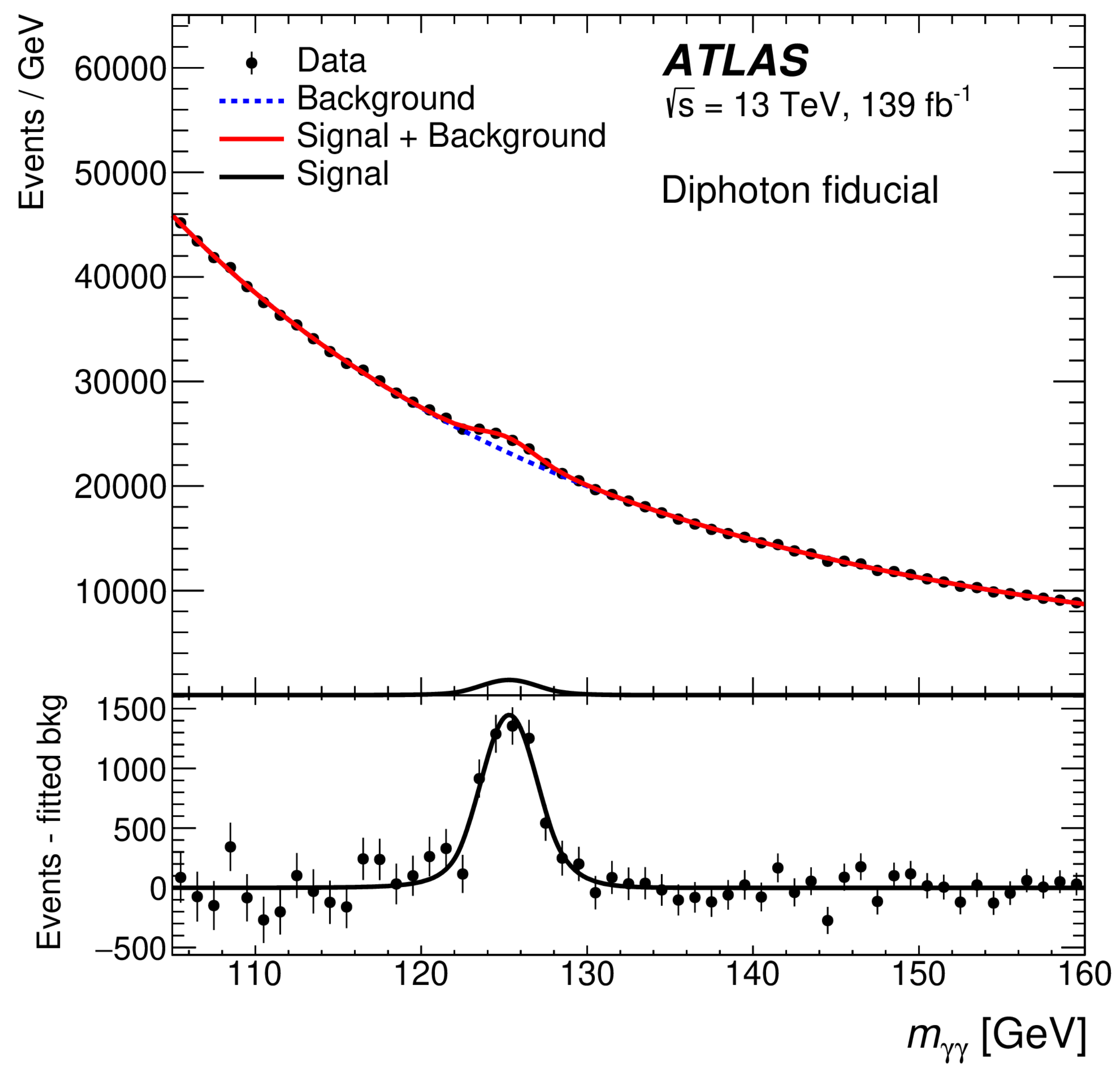
Searches for new physics are key analyses in the LHC program

- The SM is an incomplete theory, new particles are proposed to solve many of its shortcomings

In particular *resonance searches* have a long history of successes

- Strong sensitivity through a wide range of masses
- Model-independent, usually no additional assumptions about production mechanisms or additional objects
- Background can be extracted through a fit to a smooth spectrum

However, the method is restricted to production of a single particle

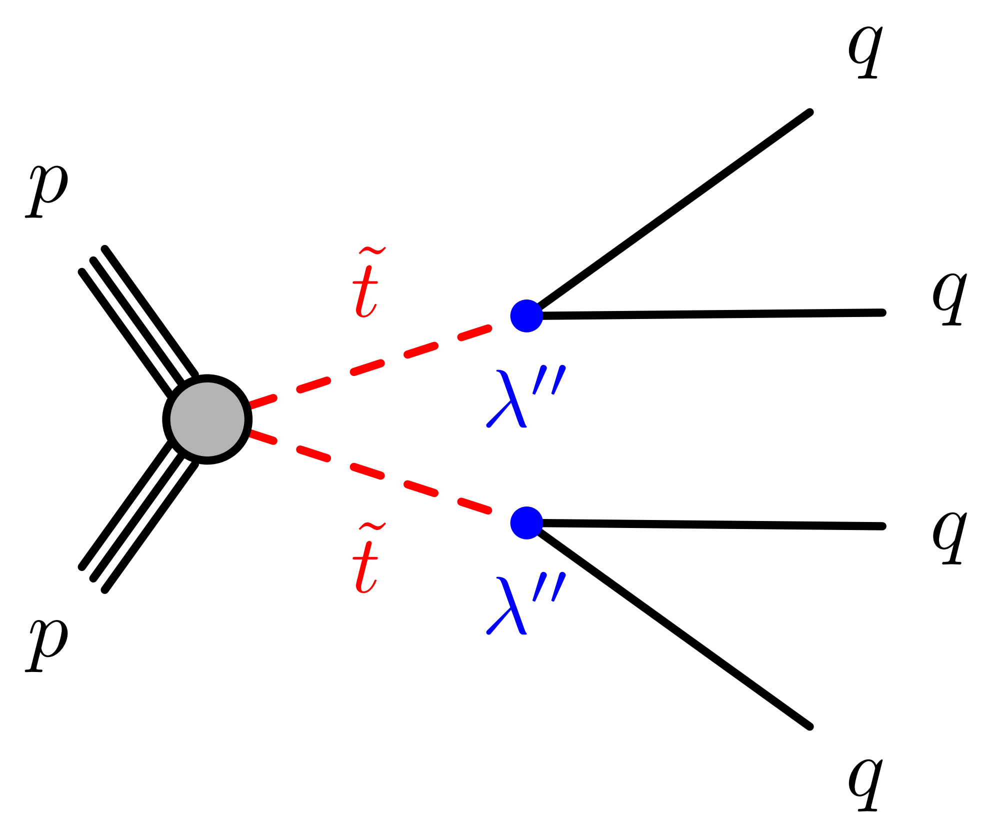


Extending resonance searches

The leading production mechanism for many BSM models is pair production

- An object to parent particle assignment is necessary before reconstructing the mass

This is a **combinatorial assignment problem**

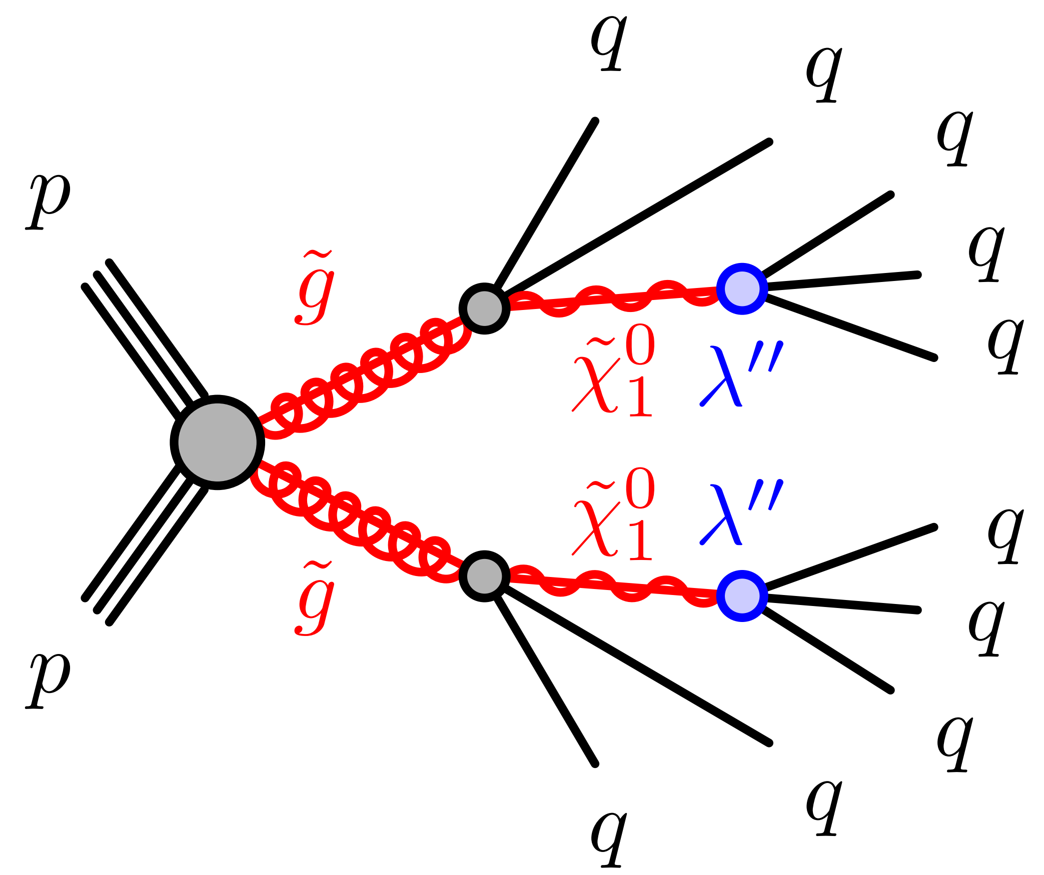


For *low* object multiplicities there are heuristic methods that work nicely

- Iterate over the three permutations, pick the one that minimizes some metric, e.g. the mass asymmetry

For *high* object multiplicities several limitations arise:

- The number of permutations scales in a factorial way, massive CPU consumption
- Possibility for mistakes increases, degrading the mass resolution
- Heuristic methods use only a small portion of the available information



Introducing ML

Multiple examples in the literature using different ML architectures to solve the problem

- SpaNet https://arxiv.org/abs/2106.03898
- Topograph https://arxiv.org/abs/2303.13937
- SaJa https://arxiv.org/abs/2012.03542
- Pelican https://arxiv.org/abs/2211.00454

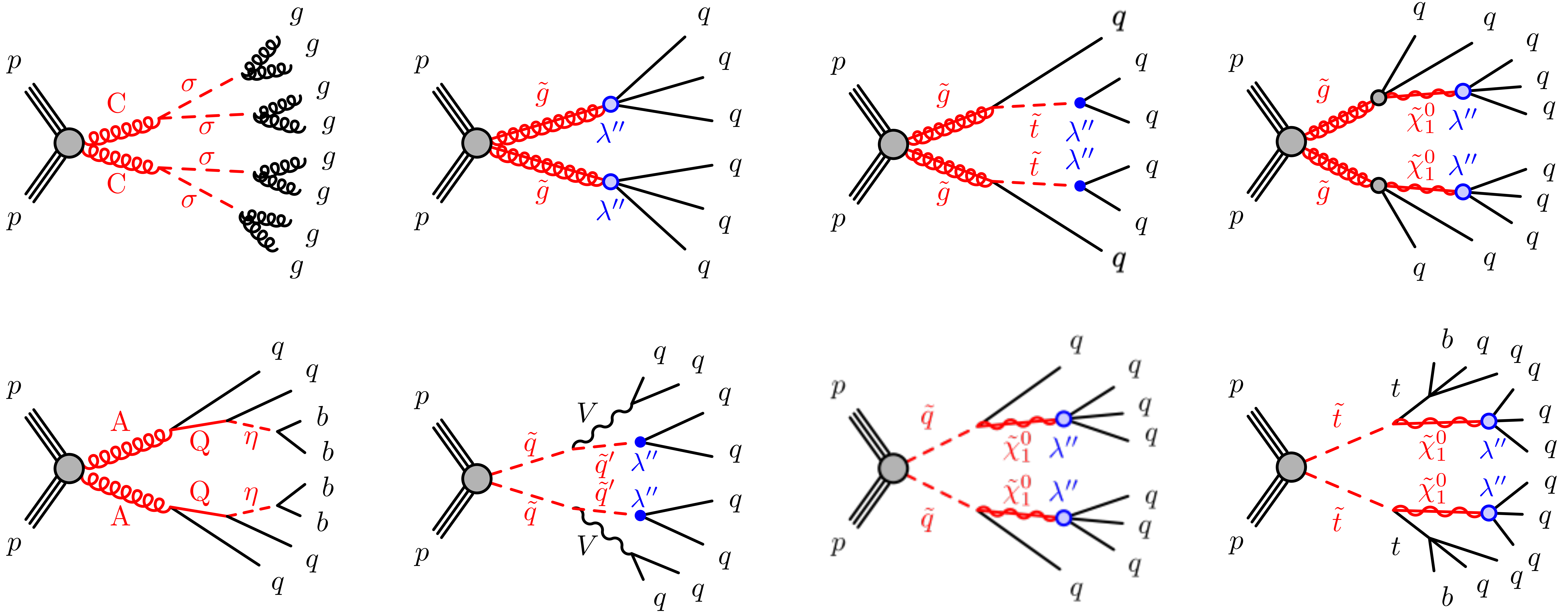
In all cases, they start by assuming a signal model, which brings two problems

- If data contains a different signal it will not be able to reconstruct it
- For background events the permutation that is more signal-like will be chosen

Therefore, we want a **model-agnostic** way of tackling this problem

- But how can we get the correct assignment without a model assumption? i.e., without having a prior definition of what "correct" is?

Which signal to assume?



The minimal assumption

Our solution works for all signals that satisfy two minimal assumptions

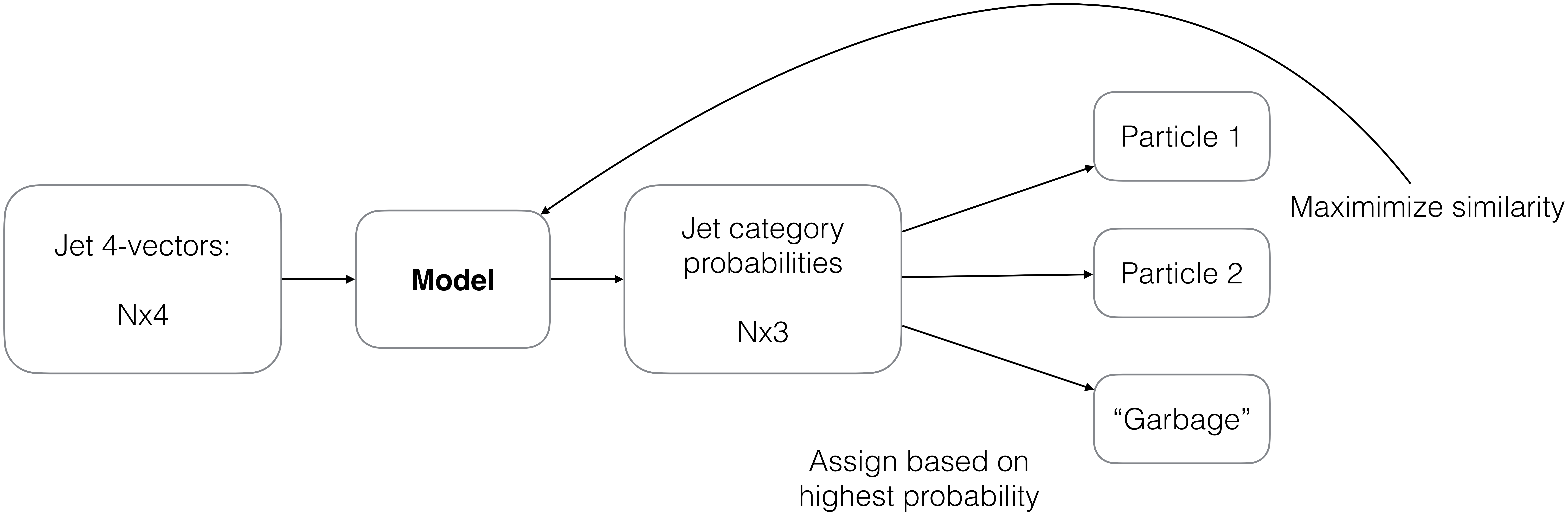
- Particles are pair-produced
- Both particles decay identically

The task of the model is to identify the assignment of objects that maximizes the *similarity* of the two parent particles

- But we don't want to impose an ad-hoc metric of similarity, e.g. not simply restricting to the mass similarity

We want to use a **data-driven** approach, with an unsupervised training strategy to avoid any reliance on simulation, no signal model hypothesis, and no external choice on the metric of similarity

The model: overview



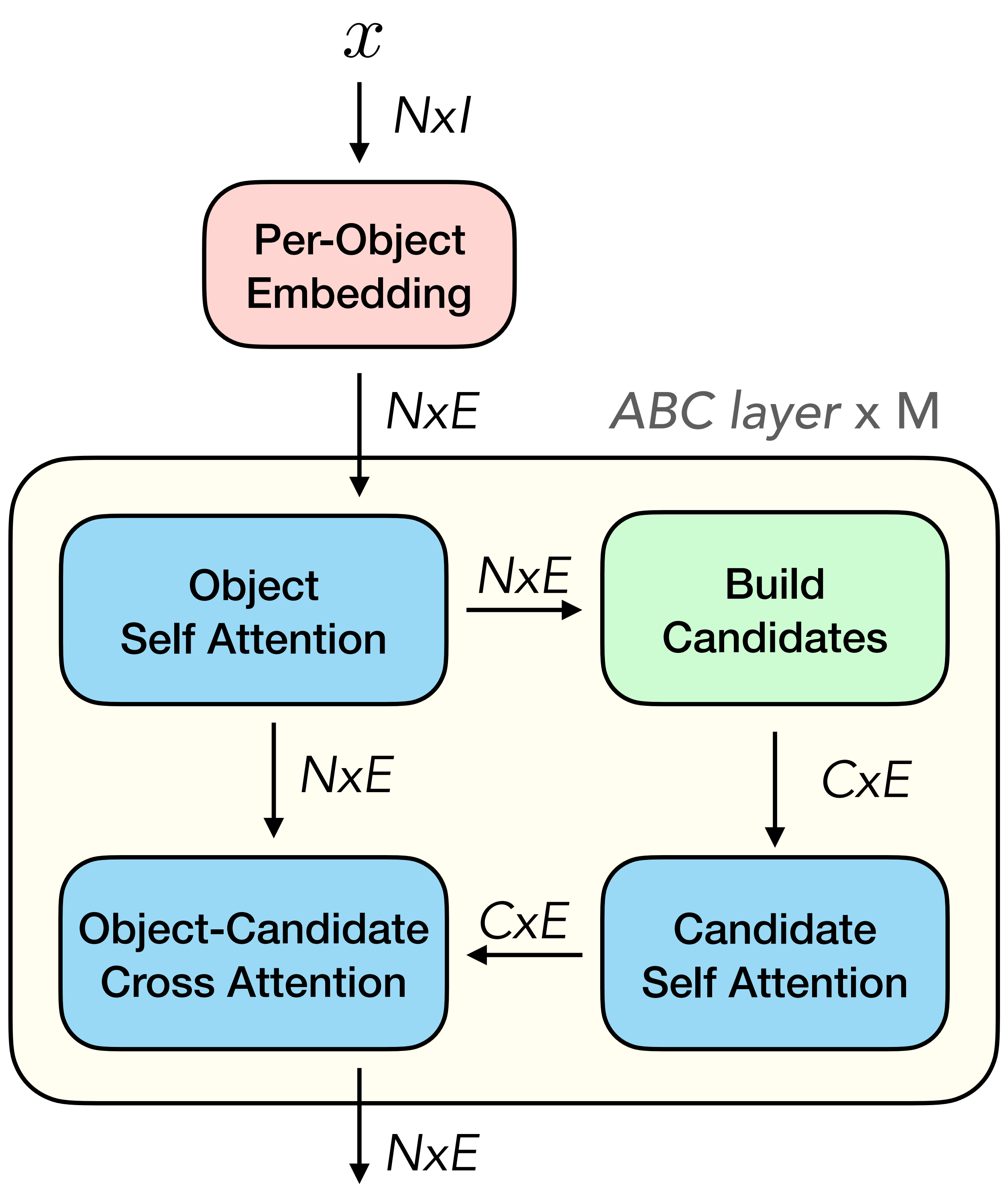
The model: ABC layer

We use the attention mechanism as the basis for a custom layer optimized for combinatorial assignment problems:

- Attention-Based Combinatorial layer, or ABC layer
- Benefits from permutation equivariance and strong expressive power

Building blocks:

- Per-object embedding, every jet is mapped into a larger feature space
    - The three first features are soft-maxed at different steps and interpreted as the category probabilities (more later)
- Object self attention, learns relationships across all jets
- Build candidate particles out of the jets with highest category score (more later)
- Candidate self attention, learns relationships across the parent particles
- Object-candidate cross attention, learns relationships across all jets and parent particles



Building candidates (differentiably)

Particle candidates are built by assigning jets to the particle with the highest probability, but this operation is *not* differentiable

Introducing some math to formalize the problem:

- A combinatorial assignment is a function $f : \mathcal{X} \to A = \{0,1\}^{N \times C}$
- For the correct assignment $a_{ij} = \mathbb{I}[l(x_i) = y_j]$
- And parent particles can be built as $p_j = \sum_i^N a_{ij}^T \cdot x_i$
- Differentiable approximation by using the category probabilities $f \sim f' : \mathcal{X} \to A' = [0,1]^{N \times C}$
- Which can be mapped back to a “hard” decision $a_{ij} = \mathrm{argmax}_{j \in C}(a'_{ij})$

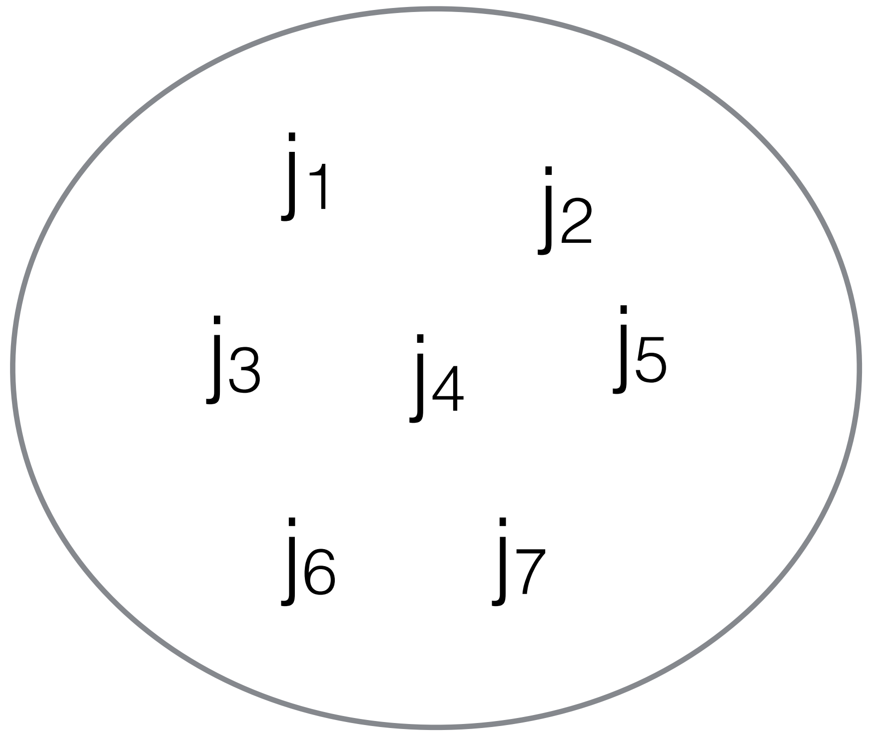
The model provides the probabilities, a'_{ij}, which we can then use to build candidates in Lorentz space, in the embedded feature space, or even mix them

The differentiable probabilities are used during training, the “hard” decision is used during inference

Building candidates (differentiably)

Particle candidates are built by assigning jets to the particle with the highest probability, but this operation is *not* differentiable

$\mathcal{X}$



$f : \mathcal{X} \rightarrow A$

combinatorial assignment

$A = \{0, 1\}^{N \times C}$

1	0	0
0	**1**	0
0	**1**	0
1	0	0
1	0	0
0	**1**	0
0	0	**1**

$p_j = \sum_i^N a_{ij}^T \cdot x_i$

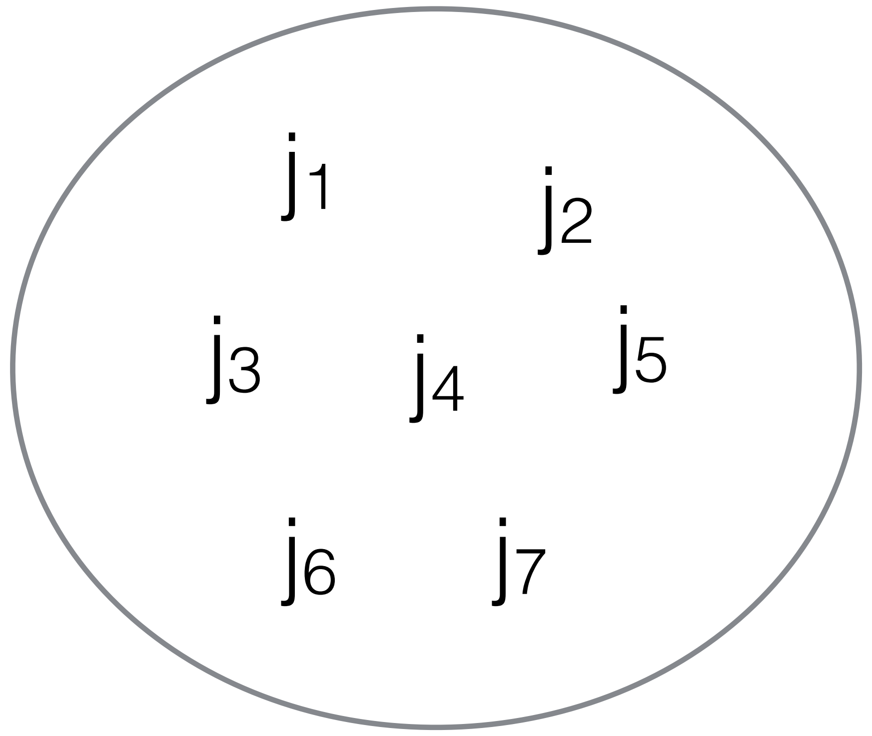
E.g. $p_1 = j_1 + j_4 + j_5$

Building candidates (differentiably)

Particle candidates are built by assigning jets to the particle with the highest probability, but this operation is *not* differentiable

The model provides the differentiable probabilities which are used during training, the “hard” decision is used during inference

$\mathcal{X}$



$$f \sim f' : \mathcal{X} \rightarrow A'$$

(our model)

$$A' = [0, 1]^{N \times C}$$

0.9	0.1	0
0.1	**0.8**	0.1
0.1	**0.9**	0
0.8	0.1	0.1
0.9	0	0.1
0.1	**0.9**	0
0	0.1	**0.9**

Training: $p'_j = \sum_i^N a'^T_{ij} \cdot x_i$

Inference: $p_j = \sum_i^N a^T_{ij} \cdot x_i$

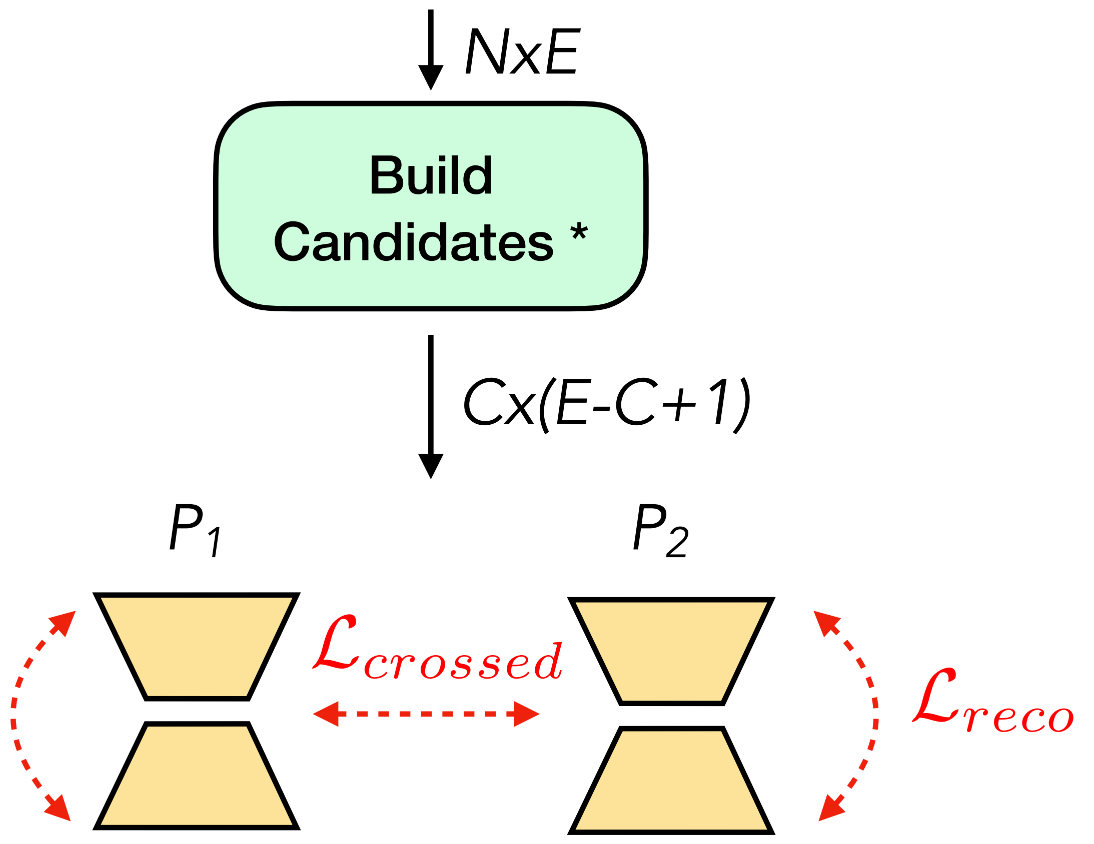
The model: training

After a stack of ABC layers we introduce one last candidate building step:

- Build particle candidates, drop the category probabilities from their features, add the particle mass
- Build two additional particle candidates, ignoring the probabilities and just picking jets randomly. Will be used for contrastive learning.

Process each particle candidate through the same autoencoder

- Reduces the large feature space to a more condensed representation. E.g. for a particle 3-body decay there are just 5 degrees of freedom



Train minimizing a loss including four terms:

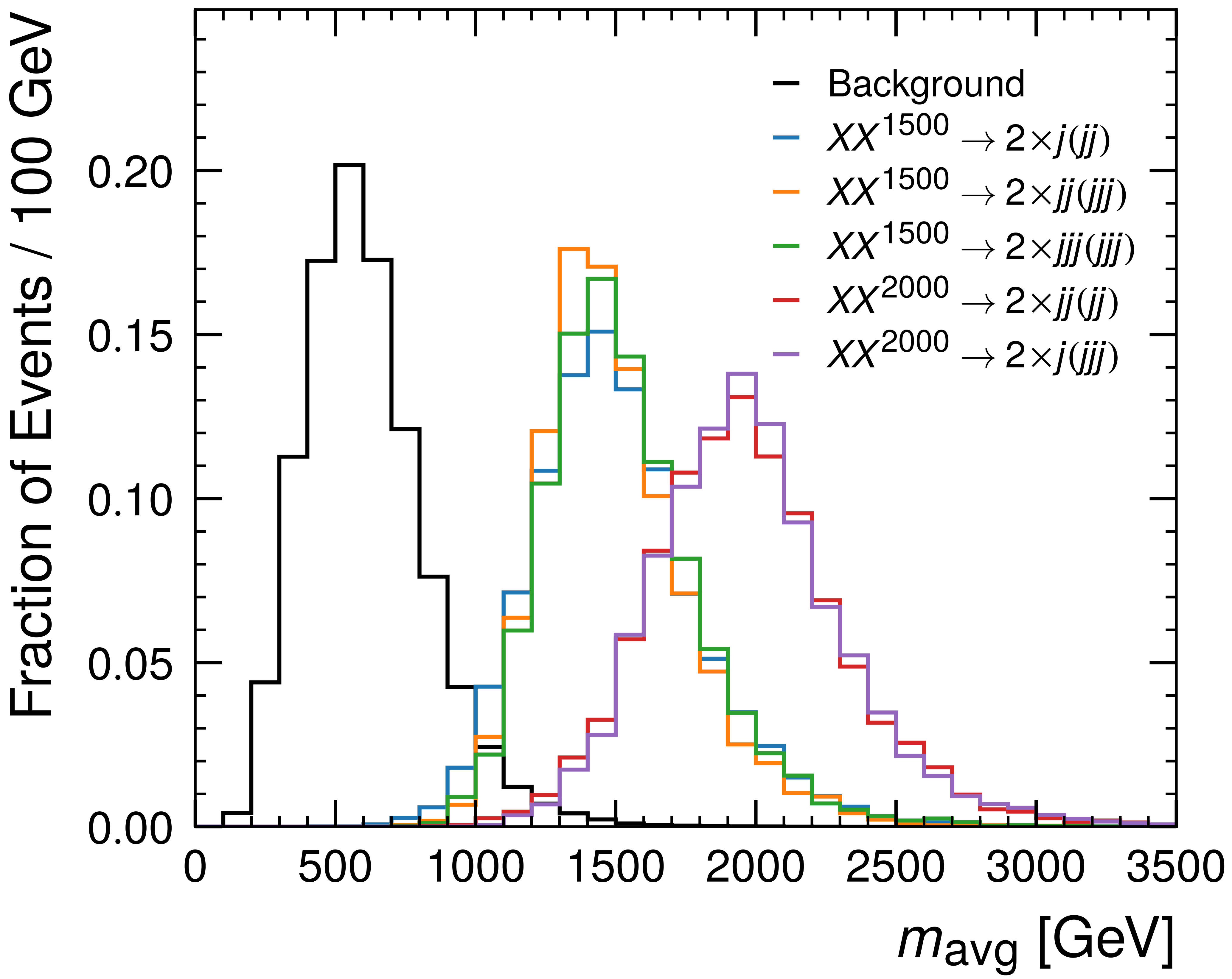
- Autoencoder reconstruction loss $\mathcal{L}_{reco} = \lambda_{reco} \cdot (\|p_1 - \hat{p}_1\| + \|p_2 - \hat{p}_2\|)$
- Minimize distance in the AE latent space between particles $\mathcal{L}_{crossed} = \lambda_{crossed} \cdot \|z_1 - z_2\|$
- Maximize distance between the two random candidates $\mathcal{L}_{rand} = \lambda_{rand} \cdot \max\left(0, 1 + \|z_1 - z_2\| - \|z_1^{rand} - z_2^{rand}\|\right)$
- Minimize transverse energy of the “garbage” system $\mathcal{L}_{ISR} = \lambda_{ISR} \cdot E_T(P_{ISR})/\mathrm{GeV}$

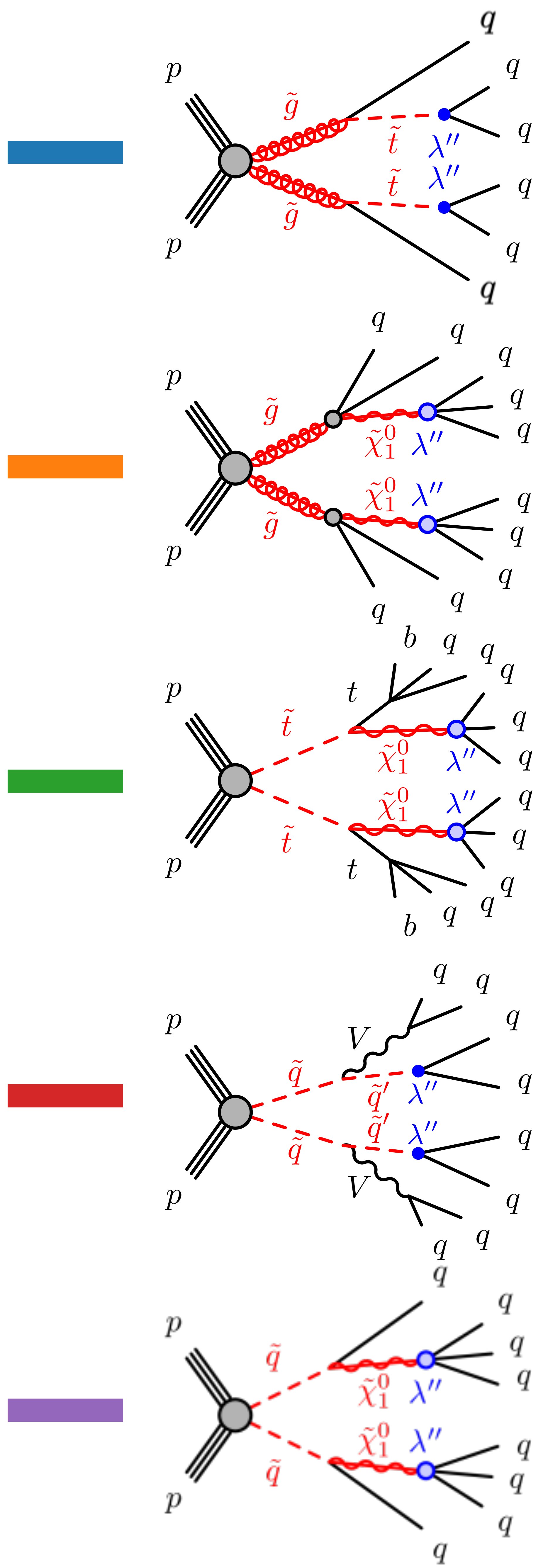
Results

The model is trained on simulated background (would be data on a real analysis)

All signals are reconstructed correctly

- At the right mass
- No dependence on the decay mode or multiplicity
- Despite not having seen any signal at all!



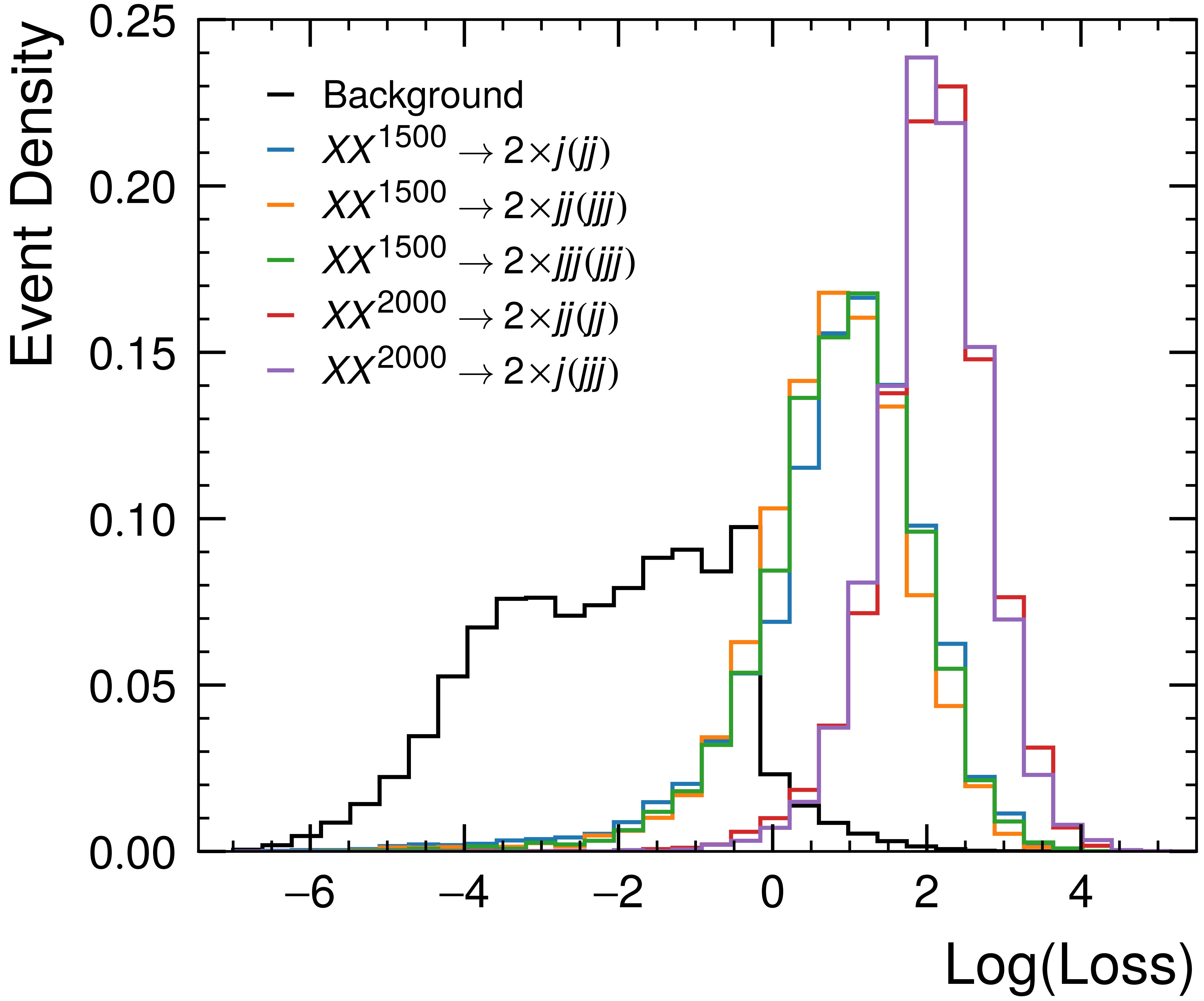


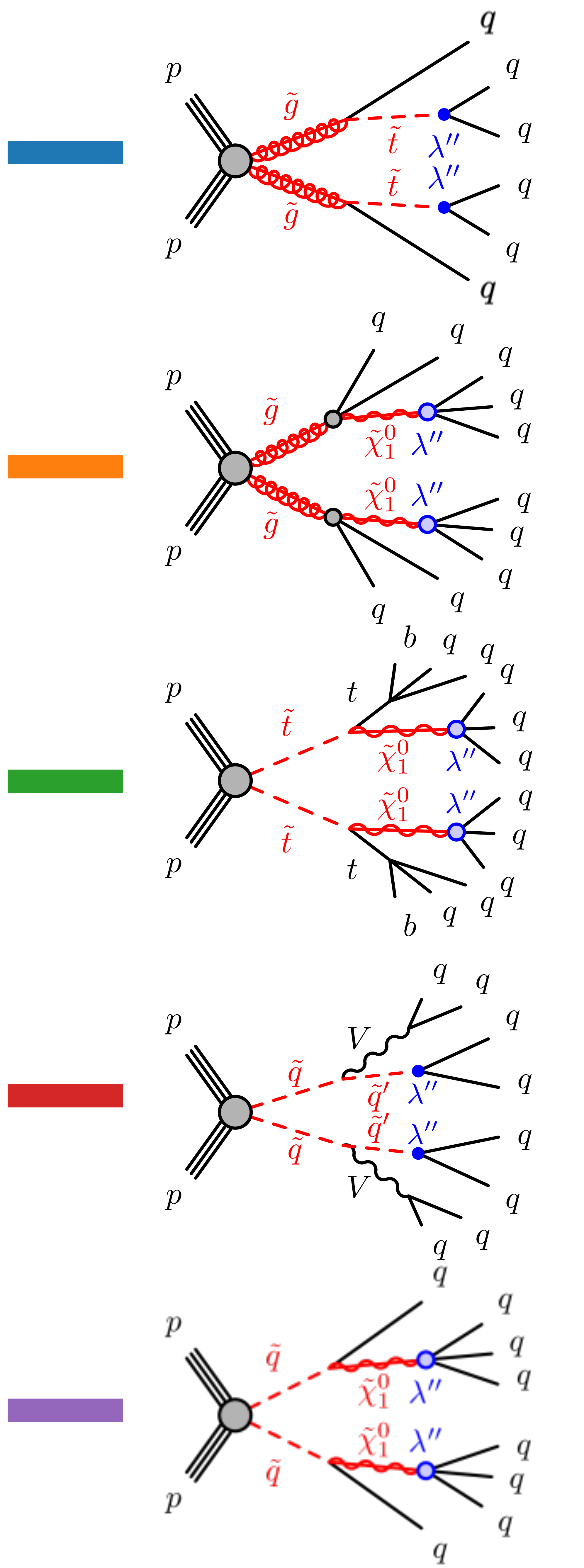
Results

The autoencoder reconstruction loss provides an additional handle to separate signal from background

Loss can be used as anomaly score to further suppress bkg

- However, strongly correlated with mass, left for future work





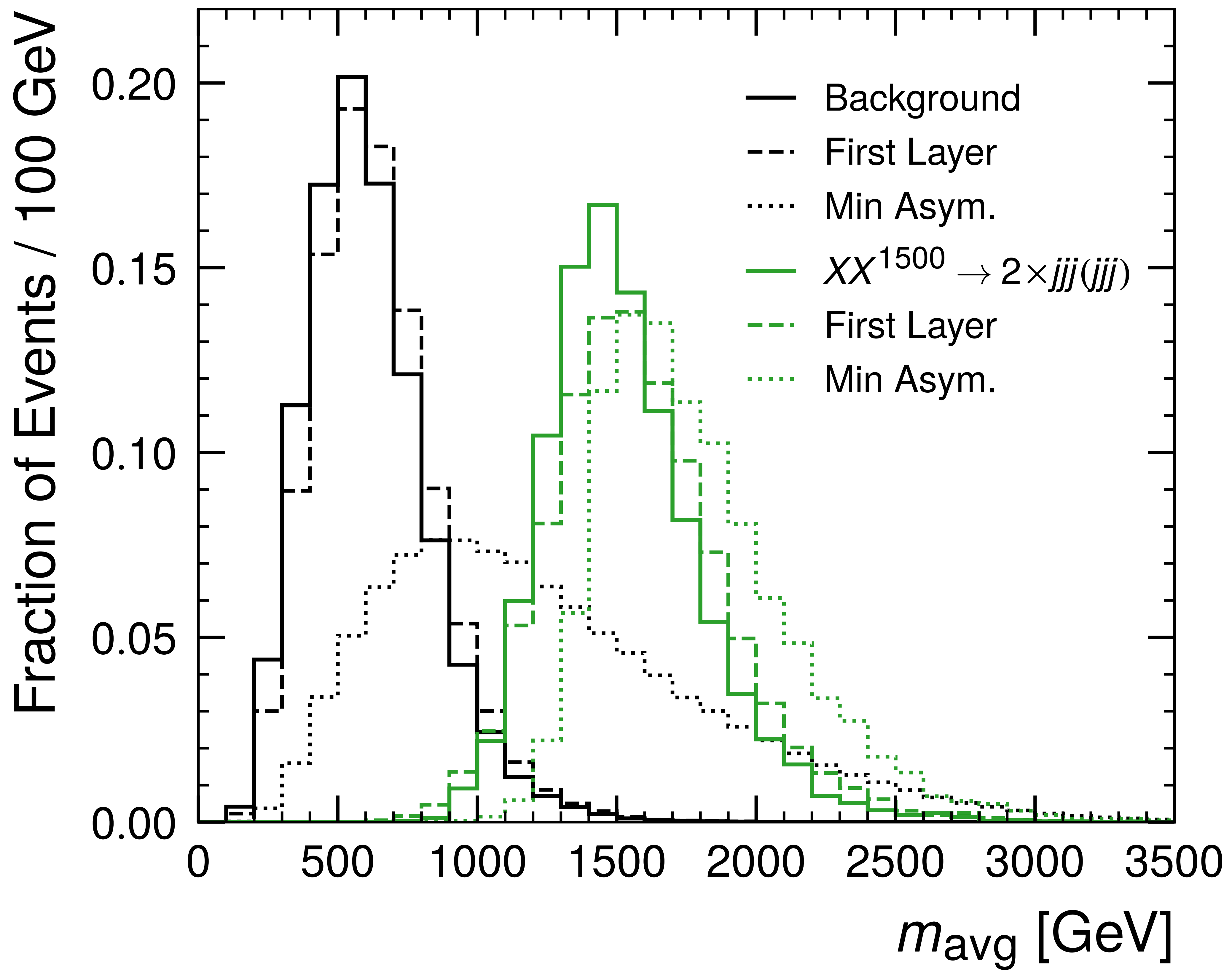
Comparison

Compare the model performance (solid line) with

- Performance after first ABC layer (dashed)
- Iterate over all possible permutations, keep the one that minimizes the mass asymmetry (dotted)

Our model achieves the best performance as seen from:

- Better signal peak
- But much more important, much less background sculpting, leads to factor 200 better signal-to-bkg ratio



Conclusions

Developed a ML model to extend resonance searches to BSM signals with pair-production

- custom ABC layer tailored to combinatorial assignment problems
- model-agnostic, reconstructs correctly any signal regardless of mass and decay topology
- data-driven, the model is trained in an unsupervised approach, no reliance on simulation

More details in 2309.05728

Working ongoing to deploy the model in an upcoming ATLAS analysis, stay tuned