Sensitivity Studies for Search of $B^+ \to K^{*+} \nu\bar{\nu}$ using Lorentz Equivariant Neural Networks at the Belle II Experiment

Caspar Schmitt, N. Hartmann, S. Bilokin, H. Hashemi, T. Kuhr

for the Belle II collaboration

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

$b \to s\nu\bar{\nu}$ Transitions

- Probe Flavor Changing Neutral Currents in $b \to s\nu\bar{\nu}$.
- Suppression in standard model allows for precise tests on alternate theories.
- Limits enhanced by precise theory prediction, since no γ exchange.
- Possible at Belle II only due to neutrinos in final state.

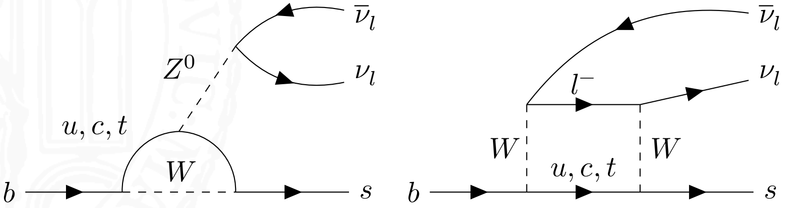
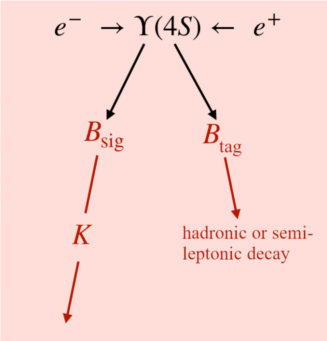
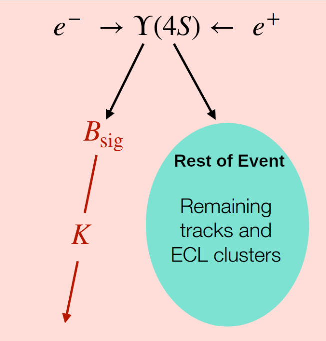


Figure 1: Lowest order diagrams in standard model.

How to Reconstruct $B \to K^* \nu \bar{\nu}$ at Belle II?

Known collision kinematics and large Belle II coverage allow two types of reconstruction:





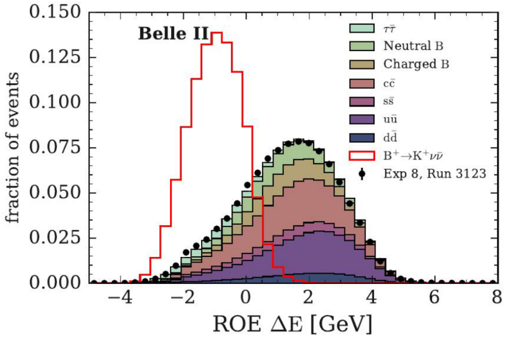
Traditional:
Full reconstruction of both B.
Low efficiency, high purity.

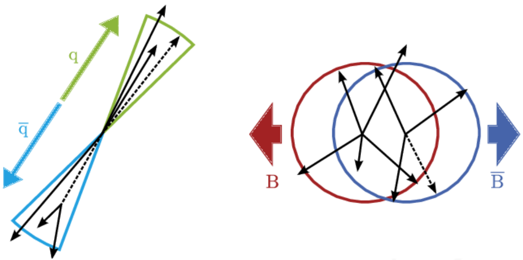
Novel 'inclusive' approach: (arXiv:2104.12624)
ML classification with minimal reconstruction.
High efficiency, low purity.

Input Features in ML classification of $B \to K^{(*)}\nu\bar{\nu}$

- **Secondary kinematic features and event shapes**

  Pro: Easy to interpret. Contra: Loss of information in high-level representation.

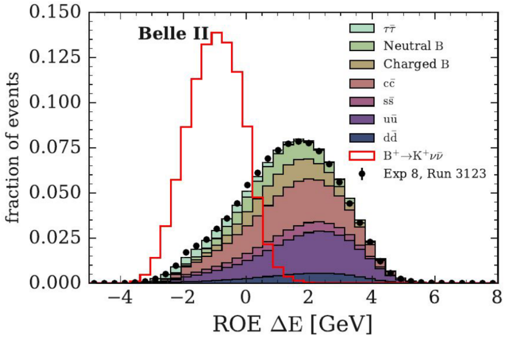


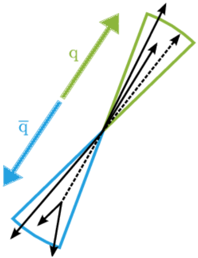


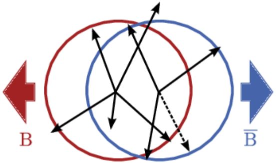
Input Features in ML classification of $B \to K^{(*)}\nu\bar{\nu}$

- **Secondary kinematic features and event shapes**

  Pro: Easy to interpret. Contra: Loss of information in high-level representation.





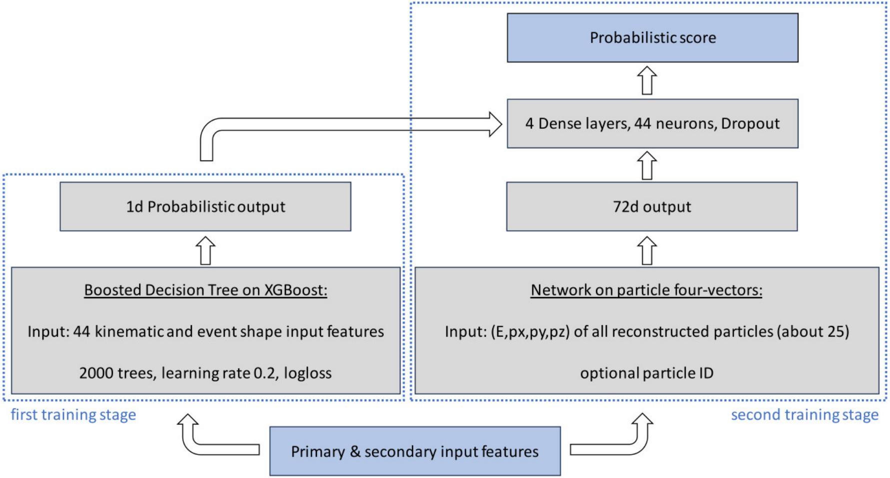


- **Primary momenta, energies**

  Pro: Full information content. Contra: Hard to interpret.

Network Architecture

Two-branch model on primary and secondary features in 2 training stages.



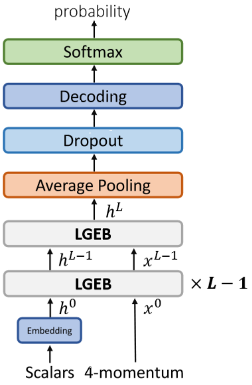
Network on Four-Vectors

Initial deep sets and GNNs (adjacency via angular distance or decay tree relations) performed poorly.

Domain-specific Lorentz-Equivariant GNN respects Lorentz group symmetries of four-momenta. Entire net equivariant under boosts and rotations.

Lorentz Equivariant Net among best performing on top tagging benchmark data set (arXiv:2201.08187).

Learn Lorentz-equivariant embedding of four-vectors, while avoiding costly tensor operations.



LorentzNet

Efficient Lorentz Equivariant Neural Net (arXiv:2201.08187)

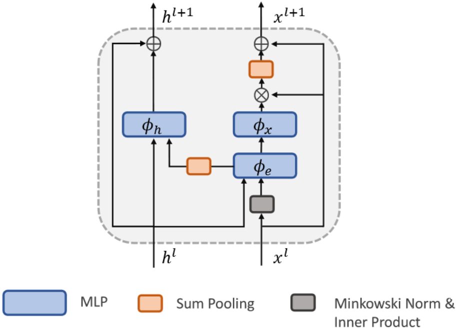
A continuous function $\phi : \mathbb{R}^{4\times N} \to \mathbb{R}^4$ is Lorentz-Equivariant if and only if

$$\phi(p_1, ..., p_N) = \sum_{i=1}^{N} g_i(\langle p_i, p_1\rangle, ..., \langle p_i, p_N\rangle) \cdot p_i \quad \text{for scalar functions/NNs } g_i.$$

Minkowski dot product attention

$$p_i^{l+1} = p_i^l + c\sum_{j\in N} \phi_x(\phi_e(h_i^l, h_j^l, ||p_i^l - p_j^l||^2, \langle p_i^l, p_j^l\rangle)) \cdot p_j^l$$

learns embedding of vectors p_i and scalars h_i through Lorentz group equivariant continuous mappings.

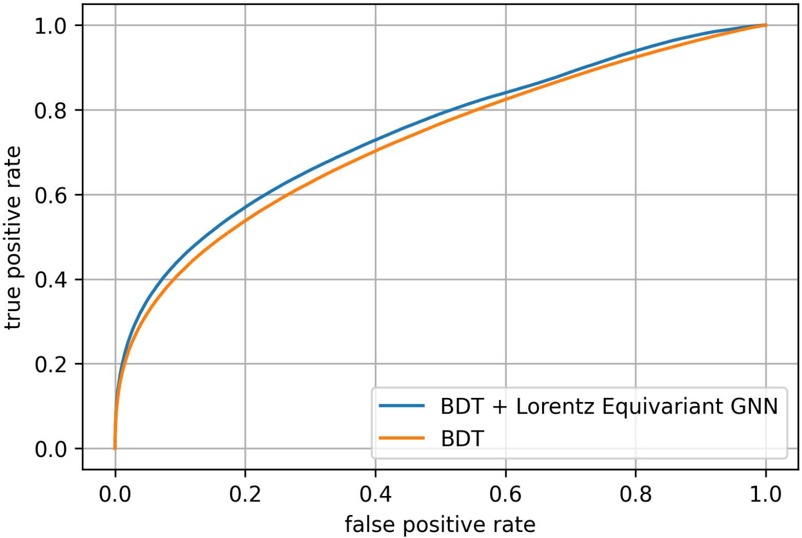


Lorentz Group Equivariant Block (LGEB)

Training Procedure for Second Training Stage

- training on Monte Carlo truth-matched, testing on non-truth-matched data.
- four momenta include intermediate particles and beam as in arXiv:2006.04780.
- batches of size 140 with 50% signal and 7 equi-proportional background types.
- binary crossentropy loss, Adam optimizer with learning rate of 0.001.
- about 10 hours training for 2500 batches and 30 epochs on NVIDIA GeForce GTX 1080 Ti.

ROC Curve


1.0
0.8
0.6
0.4
0.2
0.0
true positive rate
BDT + Lorentz Equivariant GNN
BDT
0.0
0.2
0.4
0.6
0.8
1.0
false positive rate

Outlook

- ML approches beyond BDTs for 'inclusive' reconstruction $B \to K^{*+} \nu \bar{\nu}$.
- Combining the BDT and a Lorentz equivariant architecture performs best.
- In the future scalars $(m_{\mathrm{inv}}, \mathcal{L}_{\mathrm{pid}}, q, \ldots)$ could be added to nodes.

Future model-independent searches:

- Kinematic quantities are model-dependent (four-vectors)!
- NN performance will vary with tested model.
- *Idea for uniform performance:*
  Train model as function of theory parameters.