Karlsruhe Institute of Technology

P&i Faculté de physique et ingénierie Université de Strasbourg

QMat QUANTUM SCIENCE AND NANOMATERIALS

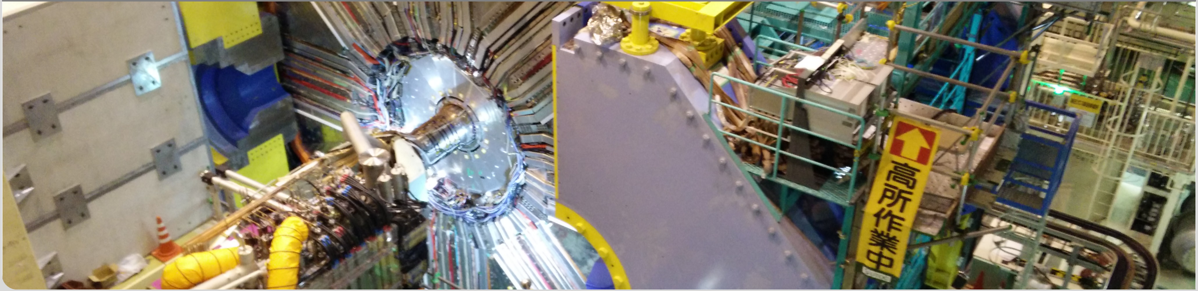
IPHC Institut Pluridisciplinaire Hubert CURIEN STRASBOURG

Belle II

Demonstrating learned particle decay reconstruction with graph neural networks at Belle II

Ilias Tsaklidis | 19/06/2020

SUPERVISORS: PABLO GOLDENZWEIG (KIT), ISABELLE RIPP (IPHC), TUTORS: JAMES KAHN (KIT), GIULIO DUJANY (IPHC)



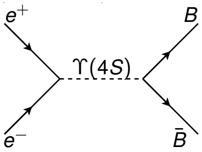
Introduction

- This work: development of Deep Learning (DL) algorithm to improve sensitivity in Belle II.
- The Standard Model is very successful but incomplete (CP violation, dark matter, unification, neutrino masses, etc.).
- Belle II: precision measurements. State-of-the-art of detector, hardware, and software technologies.

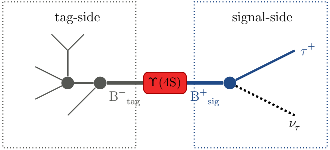
Section 1

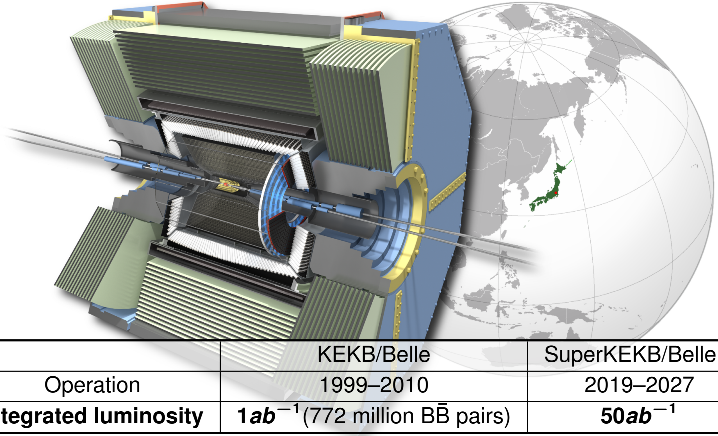
Scientific Context

Belle II Experiment



Focus on B, charm and τ physics. Can measure rare processes $\mathcal{B} < 10^{-5}$.





	KEKB/Belle	SuperKEKB/Belle II
Operation	1999–2010	2019–2027
Integrated luminosity	**$1ab^{-1}$**(772 million $B\bar{B}$ pairs)	**$50ab^{-1}$**

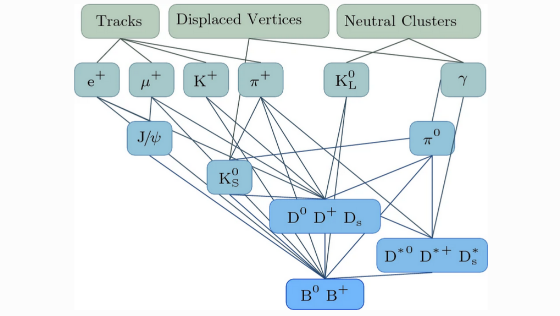
Replacing the Full Event Interpretation (FEI)

Goal: Create an algorithm that supersedes the FEI.

Why: FEI is a hierarchical machine learning algorithm. Design issues:

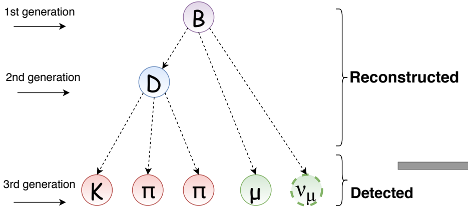
- 6 distinct stages with Fast BDTs.
- Choice of kinematic variables to exploit.
- Hard-coded reconstructed sub-decay processes.

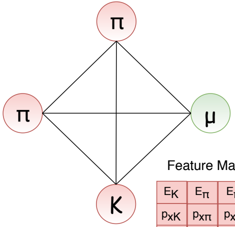
How: This work(graFEI): end-to-end method to reconstruct decays using simple kinematic information by example.



Parent	Tagging	Belle II FEI
$B^{\pm}$	Hadronic	0.61%
	Semi-leptonic	1.45%

Elements of graph theory and strategy





Feature Matrix

E_K	E_π	E_π	E_μ
p_{xK}	$p_{x\pi}$	$p_{x\pi}$	$p_{x\mu}$
p_{yK}	$p_{y\pi}$	$p_{y\pi}$	$p_{y\mu}$
p_{zK}	$p_{z\pi}$	$p_{z\pi}$	$p_{z\mu}$

Lowest Common Ancestor Matrix (LCA)

	K	π	π	μ
K	0	1	1	2
π	1	0	1	2
π	1	1	0	2
μ	2	2	2	0

Adjacency Matrix

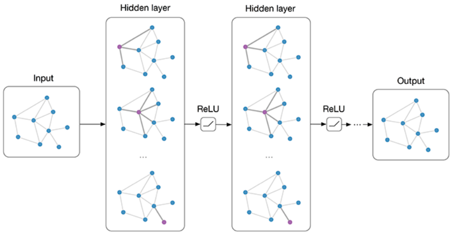
	B	D	K	π	π	μ	ν_μ
B	0	1	0	0	0	1	1
D	1	0	1	1	1	0	0
K	0	1	0	0	0	0	0
π	0	1	0	0	0	0	0
π	0	1	0	0	0	0	0
μ	1	0	0	0	0	0	0
ν_μ	1	0	0	0	0	0	0

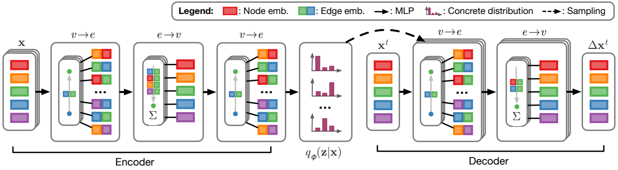
Section 2

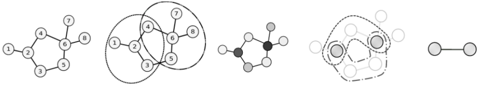
Outline

Search in literature

- GNNs for decay tree reconstruction → novelty.
- no out-of-the-box solution.
- Graph Convolutional Networks, clustering, graph pooling, edge contraction. → inefficient
- Edge Label prediction using NRI[1] → promising.

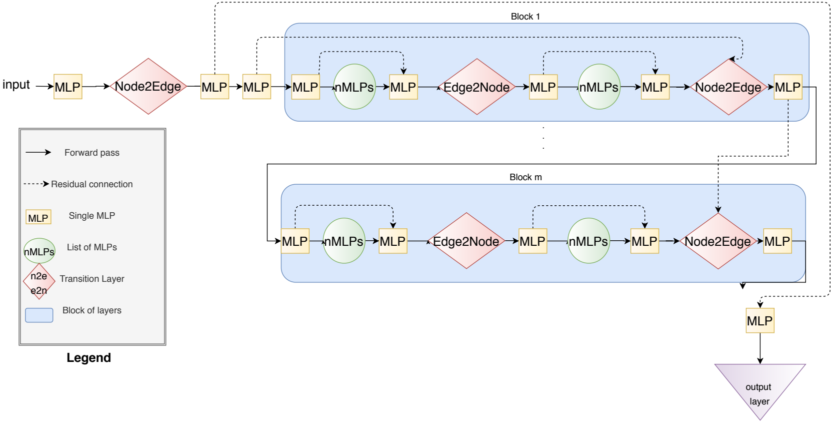






[1] *Neural Relation Inference or Interacting Systems, arXiv:1802.04687v2*

Encoder Architecture



All the models are built using the DL library ***Pytorch***.

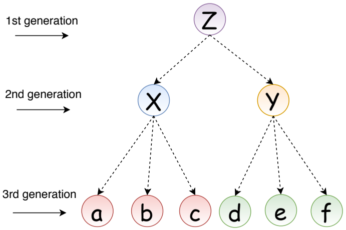
Many changes wrt initial architecture.

Hyperparameters (Optimization using ***Optuna***): **number of MLPs, blocks**, hidden nodes per layer; batch size; **learning rate**; dropout rate; number of epochs.

Data produced with *Phasespace*

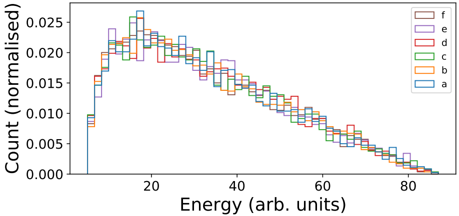
Input features: 4 momentum

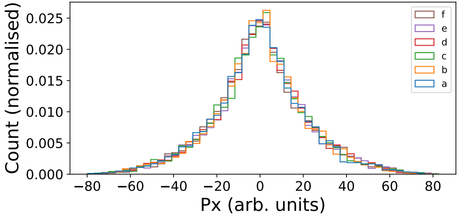
3o3 dataset



Particle	Z	X	Y	a, b, c	d, e, f
Mass (arb units)	200	80	60	5	5

Data is split into training (90%) and validation (10%) sets.





Experiments' overview

Proof of concept → First level reconstruction → Train on 2 datasets → Robustness to noise → Missing particles → Unbalanced dataset → Belle II benchmark → Complex kinematic scenarios → Unbalanced Belle II dataset → Train on generic B-meson decays

1. GNNs for particle decay reconstruction.

Experiments' overview

Proof of concept → First level reconstruction → Train on 2 datasets → Robustness to noise → Missing particles → Unbalanced dataset → Belle II benchmark → Complex kinematic scenarios → Unbalanced Belle II dataset → Train on generic B-meson decays

2. Generalization on different #FSPs

Experiments' overview

Proof of concept → First level reconstruction → Train on 2 datasets → Robustness to noise → Missing particles → Unbalanced dataset → Belle II benchmark → Complex kinematic scenarios → Unbalanced Belle II dataset → Train on generic B-meson decays

3. Identify and separate two different decays.

Experiments' overview

Proof of concept → First level reconstruction → Train on 2 datasets → Robustness to noise → Missing particles → Unbalanced dataset → Belle II benchmark → Complex kinematic scenarios → Unbalanced Belle II dataset → Train on generic B-meson decays

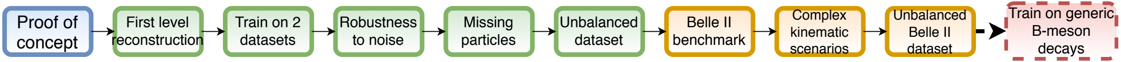
4. Robust to noise. Detector related uncertainties.

Experiments' overview

Proof of concept → First level reconstruction → Train on 2 datasets → Robustness to noise → Missing particles → Unbalanced dataset → Belle II benchmark → Complex kinematic scenarios → Unbalanced Belle II dataset → Train on generic B-meson decays

5. Missing kinematic information (semileptonic events or undetected particles).

Experiments' overview



6. Demonstration on large dataset.

Experiments' overview

Proof of concept → First level reconstruction → Train on 2 datasets → Robustness to noise → Missing particles → Unbalanced dataset → Belle II benchmark → Complex kinematic scenarios → Unbalanced Belle II dataset → Train on generic B-meson decays

7. Indicate competition with FEI

Experiments' overview

Proof of concept → First level reconstruction → Train on 2 datasets → Robustness to noise → Missing particles → Unbalanced dataset → Belle II benchmark → Complex kinematic scenarios → Unbalanced Belle II dataset → Train on generic B-meson decays

8. Include channels not dealt with FEI

Experiments' overview

Proof of concept → First level reconstruction → Train on 2 datasets → Robustness to noise → Missing particles → Unbalanced dataset → Belle II benchmark → Complex kinematic scenarios → Unbalanced Belle II dataset → Train on generic B-meson decays

9. Demonstration on larger Belle II dataset.

Experiments' overview

Proof of concept → First level reconstruction → Train on 2 datasets → Robustness to noise → Missing particles → Unbalanced dataset → Belle II benchmark → Complex kinematic scenarios → Unbalanced Belle II dataset - - → Train on generic B-meson decays

1. GNNs for particle decay reconstruction.
2. Generalization on different #FSPs
3. Identify and separate two different decays.
4. Robust to noise. Detector related uncertainties.
5. Missing kinematic information (semileptonic events or undetected particles).
6. Demonstration on large dataset.
7. Indicate competition with FEI
8. Include channels not dealt with FEI
9. Demonstration on larger Belle II dataset.

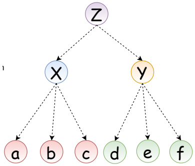
Section 3

Results

Proof of Concept

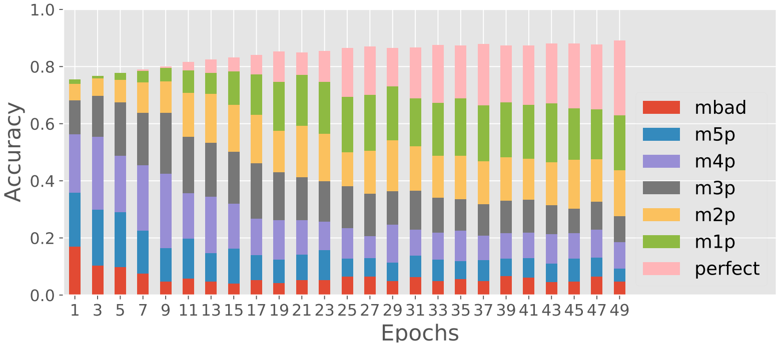
Proof of concept

Train on generic B-meson decays



3o3 dataset

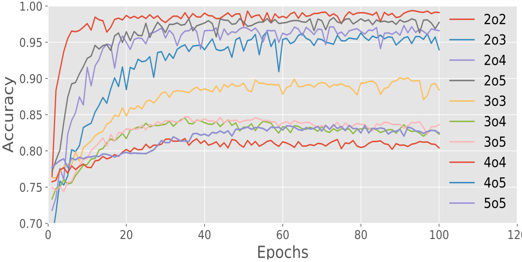
- Accuracy: individual entries.
- perfect: exact decay trees.
- mXp: mistakes per prediction.

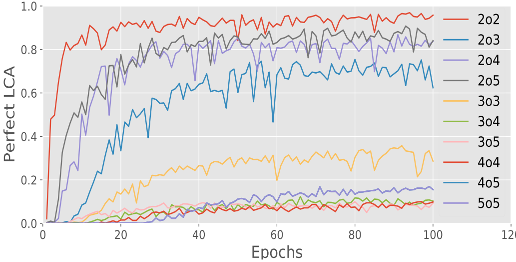


GNNs can achieve particle decay reconstruction.

First Level Reconstruction

Proof of concept → First level reconstruction → - → Train on generic B-meson decays



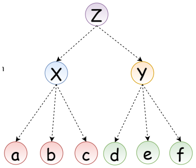


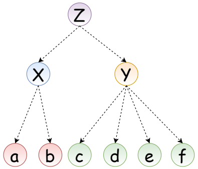
The model can generalize to different datasets.

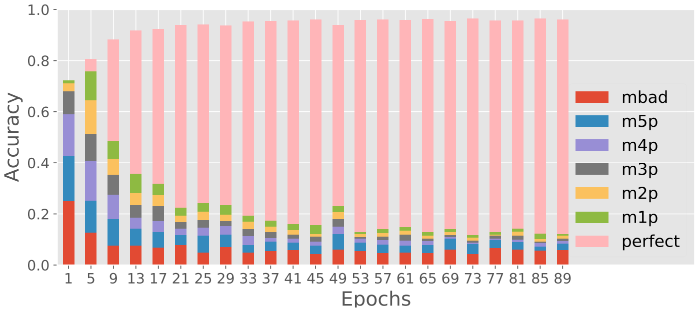
Unstable training, scaling of performance: shallow networks.

Two datasets

Proof of concept → First level reconstruction → Train on 2 datasets → - → Train on generic B-meson decays



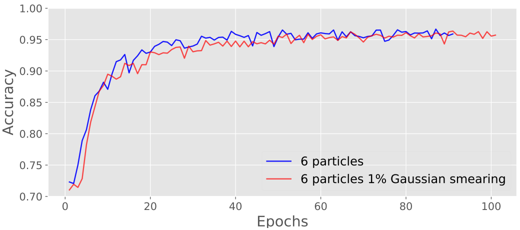


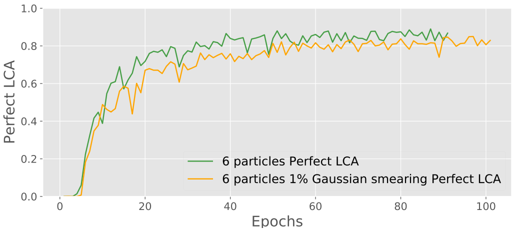


The model can identify different decays.
Deeper model, better training.

Data with noise

Proof of concept → First level reconstruction → Train on 2 datasets → Robustness to noise → → Train on generic B-meson decays





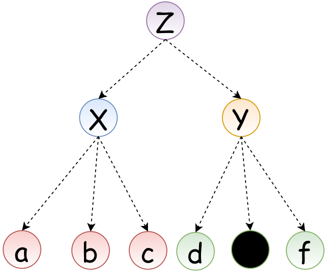
Model trained on unsmeared data applied to smeared:

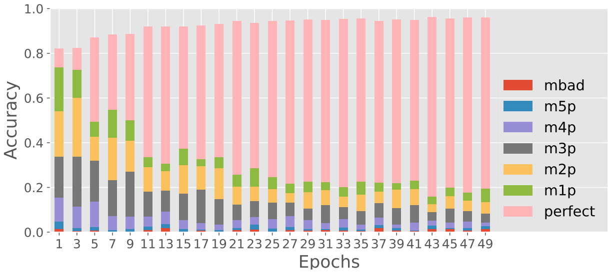
Acurracy: 0.9756, Perfect: 0.8891.

Robust model to random noise.

Missing particles

Proof of concept → First level reconstruction → Train on 2 datasets → Robustness to noise → Missing particles → ··· → Train on generic B-meson decays

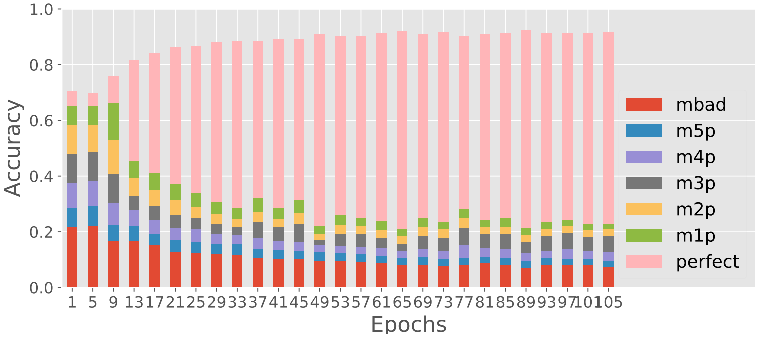




Indication for semileptonic tagging.

Mix of all the *Phasespace* datasets

Proof of concept → First level reconstruction → Train on 2 datasets → Robustness to noise → Missing particles → Unbalanced dataset → … → Train on generic B-meson decays



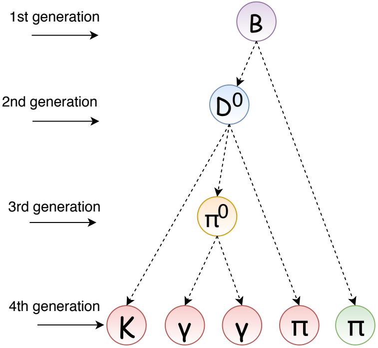
Padding is used to mix the dataset. Masking is used to train on the padded data efficiently. The model can reconustruct numerous decays simultaneously.

Data produced with the Belle II software (basf2)

1. Monte Carlo simulation (no detector simulation yet)
2. Signal side: $B \to \mu\nu_\mu$, easy to separate since only one FSP
3. Input features: 4 momentum + charge

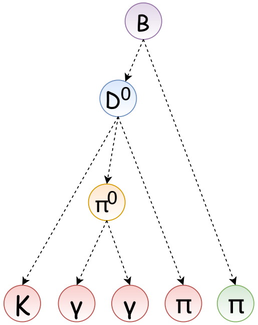
Decay Channels generated with the Belle II software		
Decay Channel	N^{o}FSPs	Motivation
$B^+ \to \overline{D^0}(\to K^+\pi^-\pi^0)\pi^+$	**5**	**benchmark tag side** on T.Keck's PhD thesis on FEI
$B^+ \to D^-(\to \pi^-\pi^+\pi^+)\pi^+\pi^+$	**5**	two 3-body decays, **overlapping spectra**, same FSPs)
$B^+ \to \overline{D^0}(\to K^+\pi^-\pi^0)e^+\nu_e$	**5**	**semileptonic decay** to demonstrate semileptonic tagging
$B^+ \to \overline{D^0}(\to K^+\pi^-\pi^0)\rho(\to \pi^-\pi^0)$	**7**	resonances **not dealt with FEI**, includes 4 photons that need to be assigned to the correct π^0
$B^+ \to \overline{D^0}(\to K^+\pi^-\pi^0)\omega(\to \pi^+\pi^-\pi^0)\pi^+$	**9**	Three 3-body decays, resonances **not dealt with FEI**
$B^+ \to D^-(\to \pi^-\pi^-\pi^+\pi^0)\pi^+\pi^+\pi^0$	**9**	two **4-body decays**

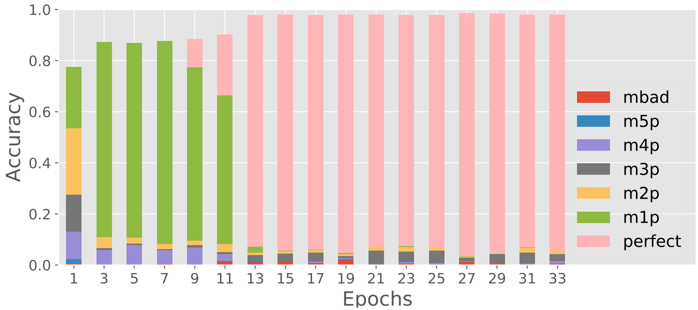
Table 1: Decay channels produced with the Belle II software for this work. All the π^0 decay into two photons. All the datasets contain the decay channel presented here and its charge conjugate



Benchmark Belle II dataset

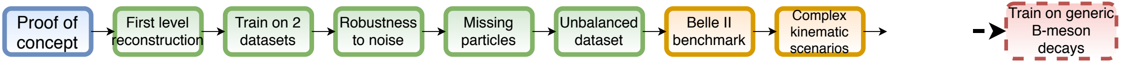
Proof of concept → First level reconstruction → Train on 2 datasets → Robustness to noise → Missing particles → Unbalanced dataset → Belle II benchmark → Train on generic B-meson decays

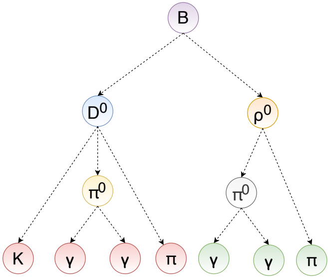


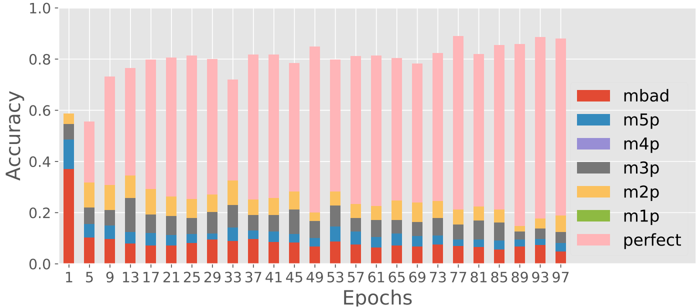


The model can compete with the FEI.

Complex kinematic scenarios

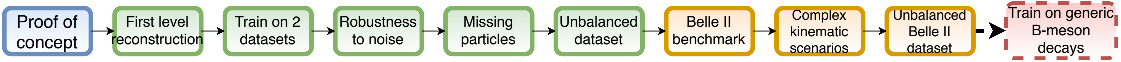


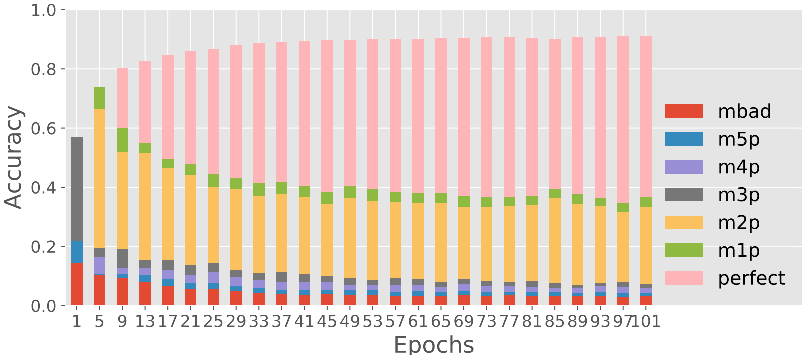




Resonance not dealt with FEI.

Mix of all the Belle II datasets





Last test before training on generic decays.

Section 4

Summary

Conclusions

- Proof of concept of a graph based, end-to-end approach for decay tree reconstruction from example, exploiting simple kinematic variables.
- Lowest Common Ancestor matrix contains the necessary information to capture the structure of a decay tree.
- 75% of perfectly predicted LCAs on unbalanced data (all the *Phasespace* datasets).
- 95% of perfectly predicted LCAs on the benchmark decay tree used by Belle II for B-tagging.
- Efficient predictions on decay channels that FEI doesn't deal with.

Outlook

1. Train on generic B-mesons decays.
2. Test the performance of the model on events with extra particles(beam background like etc.).
3. Train on reconstructed events, after the detector simulation.
4. Understand how the idea depth of the network scales with the #FSPs.

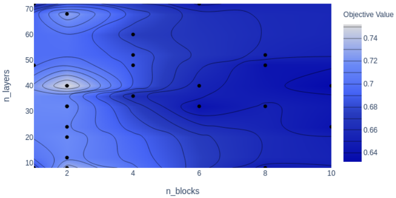


Figure: 6 particles

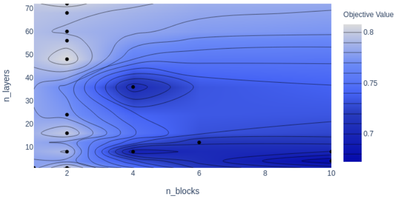


Figure: 7 particles

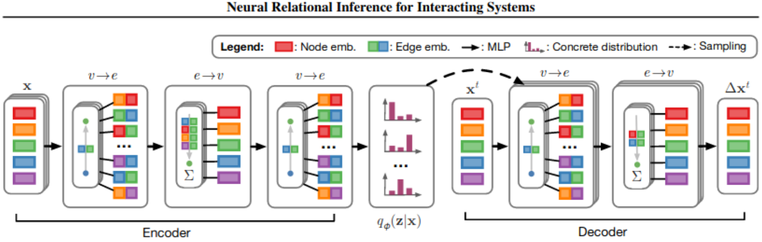
Some -non exhaustive- References

- The Full Event Interpetation
  "Keck, T. and others",
  "arXiv:1807.08680".
- Neural Relational Inference for Interacting Systems
  "Thomas Kipf and Ethan Fetaya and Kuan-Chieh Wang and Max Welling and Richard Zemel"
  "arXiv:1802.04687".
- Variational Graph Auto-Encoders
  "Thomas N. Kipf and Max Welling",
  "arXiv:1611.07308".
- A Comprehensive Survey on Graph Neural Networks
  "Zonghan Wu and Shirui Pan and Fengwen Chen and Guodong Long and Chengqi Zhang and Philip S. Yu",
  "arXiv:1901.00596".

Section 5

Backup

Graph Autoencoder



- Autoencoder: A NN that learns a representation (encoding) typically in a lower dimensional space and then tries to reconstruct the original input (decoding) from this lower representation
- The autoencoder from the paper *Neural Relational Inference for Interacting Systems* is used for learning the law of Physics that governs the interaction of n-body systems
- We use the encoder part for an edge-labelling task. We interpret the learnt edge labels as the entries of the LCA matrix

Proof of Concept: 3o3 Overtraining

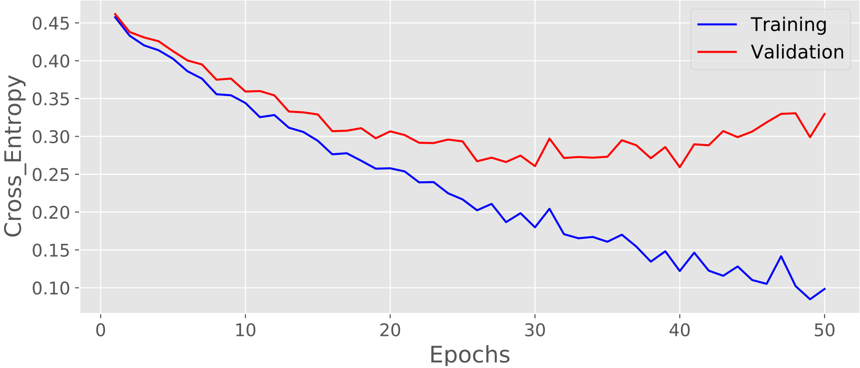


Figure: Demonstration of overtraining for the 3o3 dataset with a shallow network

Elements of Deep Learning

1. Data is split into training and validation set to monitor overtraining.
2. Input tensors with basic kinematic information (4 momentum).
3. random initialization of weights.
4. activation function (ELU in this work) turns off some nodes.
5. Dropout erases some nodes randomly to fight overtraining.
6. calculation of loss of the final predictions (using Cross Entropy in this work).
7. calculation and multiplication of $\frac{d\Phi}{dw_{ij}}$ with the learning rate. Update of all the weights.

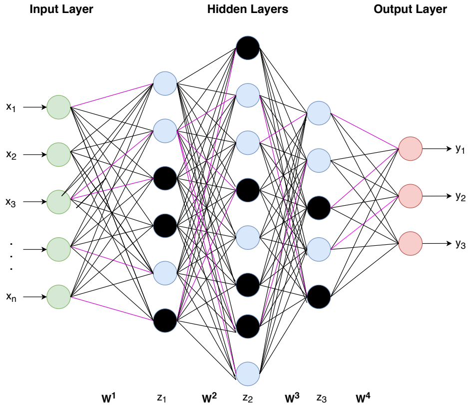


Figure: Typical Multilayer Perceptron (MLP)

Learnable parameters and Hyperparameters

Best tuning for mixed datasets							
Set	bsize	lr	dropout	nhid	nBlocks	nMLPs	DoF
6par	16	0.0011	0.000744	128	8	14	75776
7par	16	0.000072	0.308	128	8	4	34816
8par	16	0.000185	0.133	80	4	14	23680
$B^+ \rightarrow \overline{D^0}(\rightarrow K^+\pi^-\pi^0)\pi^+$	32	0.001	0.008520	512	4	1	45056
$B^+ \rightarrow \overline{D^0}(\rightarrow K^+\pi^-\pi^0))e^+\nu_e$	32	0.001	0.008520	512	4	1	45056
$B^+ \rightarrow D^-(\rightarrow \pi^-\pi^+\pi^+)\pi^+\pi^+$	64	0.00062	0.1883	128	4	12	33792
$B^+ \rightarrow \overline{D^0}(\rightarrow K^+\pi^-\pi^0)\rho(\rightarrow \pi^-\pi^0)$	16	0.00036	0.0624	128	4	12	33792
$B^+ \rightarrow \overline{D^0}(\rightarrow K^+\pi^-\pi^0)\omega(\rightarrow \pi^+\pi^-\pi^0)\pi^+$	16	0.000485	0.0304	128	4	12	33792
$B^+ \rightarrow D^-(\rightarrow \pi^-\pi^-\pi^+\pi^0)\pi^+\pi^+\pi^0$	64	0.00117	0.00551	256	4	12	67584
all Phasespace	128	0.001	0.25	1024	2	4	69632
all Belle	128	0.001	0.25	1024	2	4	69632

$$\text{learnable} = [(4 \cdot 2) + (5 \cdot 2) + (2 \cdot nMLPs \cdot 2)] \cdot nblocks] \cdot nhid$$

2 missing particles 3o3