Finding invisible decays with Deep Learning

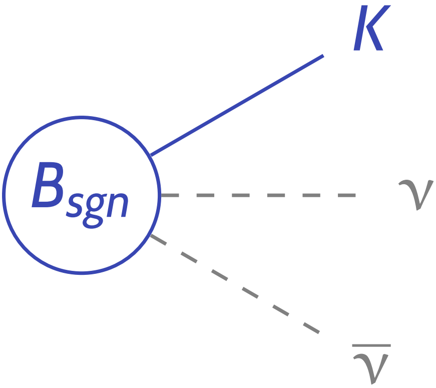
Giulio Dujany

Belle II, IPHC-DRS, CNRS

IPHC Machine Learning day

The problem

We want to study decays with particles that we do not reconstruct like neutrinos



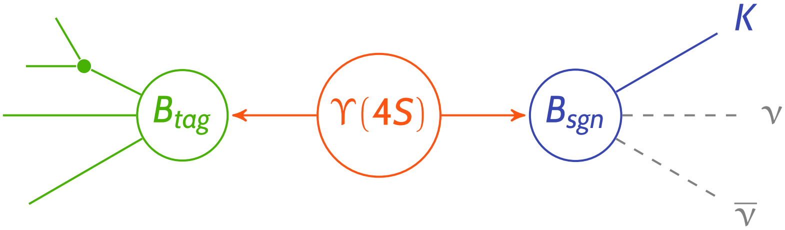
The strategy

Reconstruct all the rest and infer what is missing

©bbcgoodfood.com

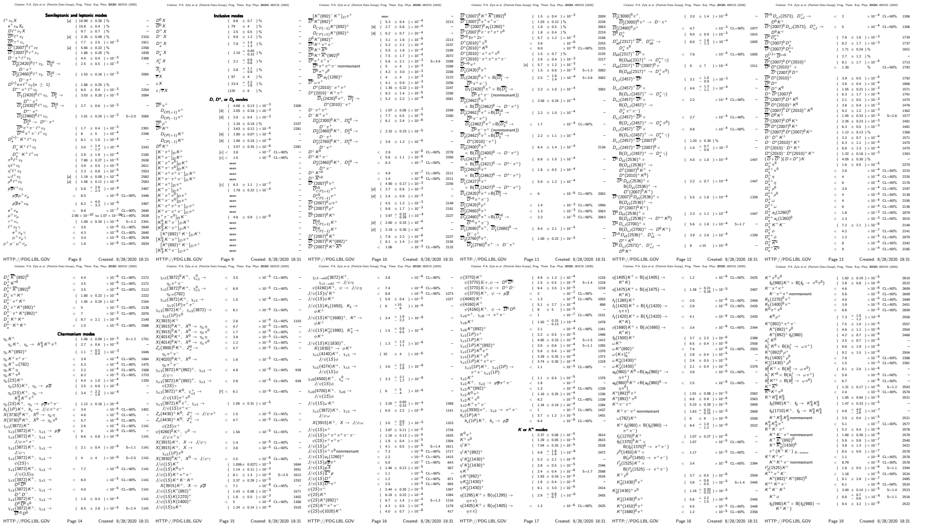
What does it mean at Belle II

What does it mean at Belle II



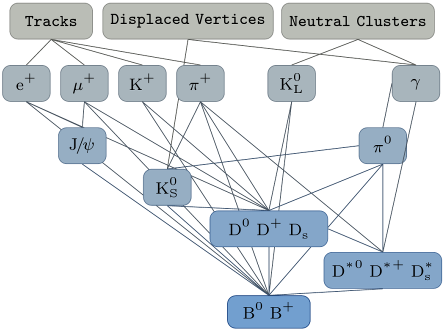
What does it mean at Belle II

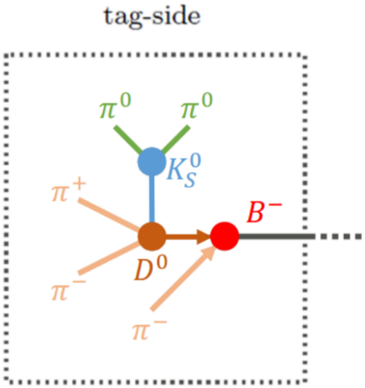
Possible decays of the B_{tag}



The current solution: Full Event Interpretation

- Hierarchical machine learning algorithm
- 6 levels of BDT to reconstruct sequentially the intermediate decays
- more that 10 000 decays considered
- $\mathcal{O}(1\%)$ efficiency

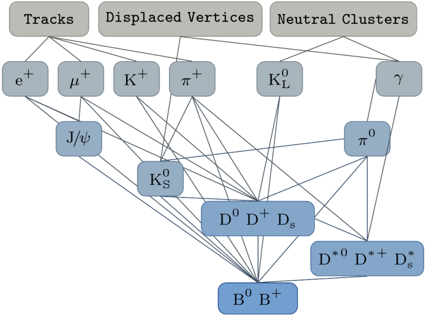


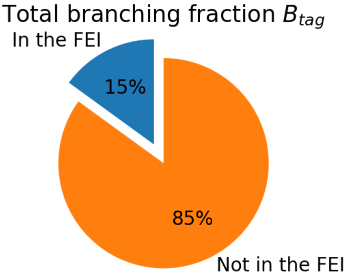


[Comput.Softw.Big Sci. 3 (2019)]

Drawbacks

- 6 stages are disconnected
- Possible sub-decays are hard-coded





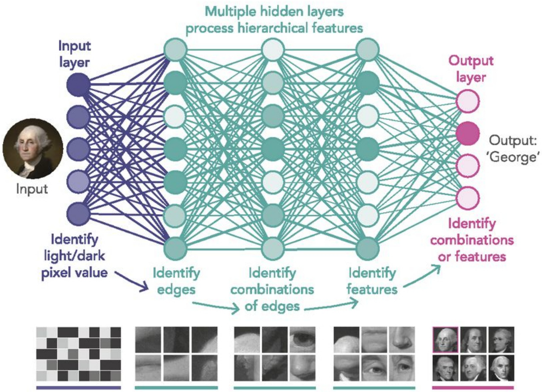
[Comput.Softw.Big Sci. 3 (2019)]

Our goal

Develop a end-to-end trainable method to learn by example how to reconstruct a generic decay from its final state particles

Deep learning

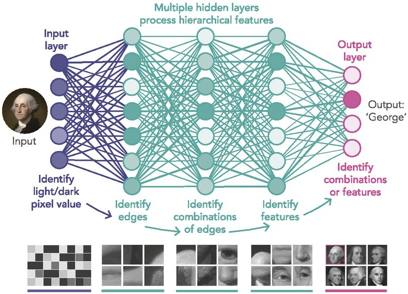
- Deep learning extracts autonomously the high level features without the need of experts intervention
- Successful for complex problems with hierarchical data (image recognition, natural language processing, web searches, ...)



©https://doi.org/10.1073/pnas.1821594116

Deep learning

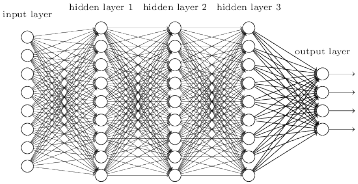
- Deep learning extract autonomously the high level features needed to solve the task
  - Learn autonomously intermediate decays without the need to hard-code them
- Successful for complex problems with hierarchical data (image recognition, natural language processing, web searches, ...)
  - A particle decay is hierarchical and reconstruct a generic decay is surely a complex problem ;)



Which deep learning architecture?

Fully connected neural networks

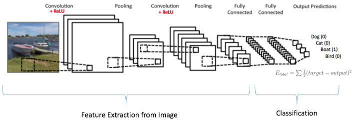
- good for vectors



©neuralnetworksanddeeplearning.com

Convolutional neural networks

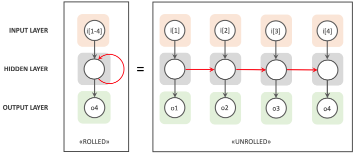
- good for images



©kdnuggets.com

Recurrent neural networks

- good for sequences

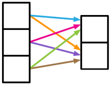


©bouvet.no

A matter of symmetries

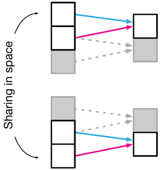
Fully connected neural networks

- good for vectors
- no sharing of learning blocks
- no symmetry



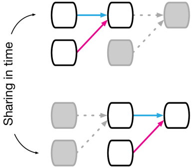
Convolutional neural networks

- good for images
- sharing learning blocks in space
- symmetric for translation in space



Recurrent neural networks

- good for sequences
- sharing of learning blocks in time
- symmetric for translation in time

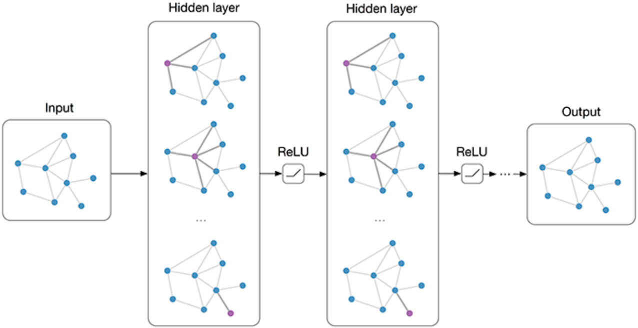


[arXiv:1806.01261]

Graph networks

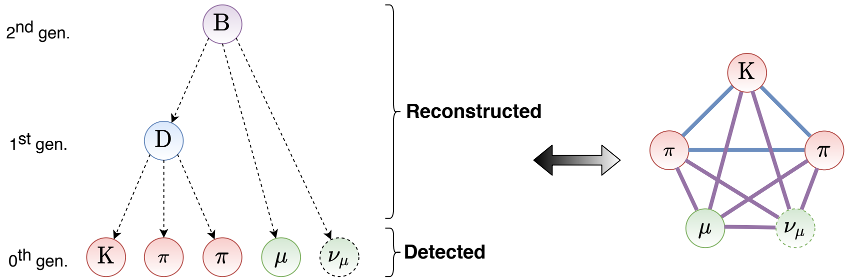
Particles in a decay

- do not have a fixed number
- do not have a clear ordering (symmetric under permutations)
- can be described by a graph → use graph networks

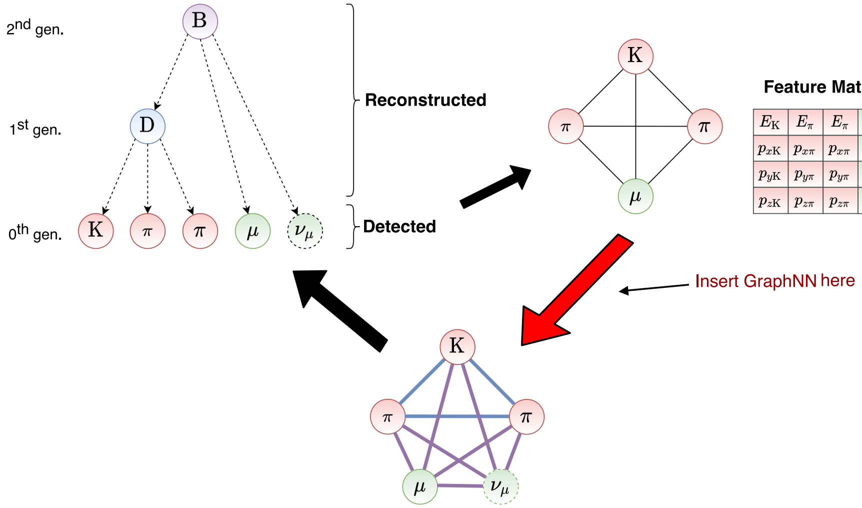


©tkipf.github.io/graph-convolutional-networks

Going from a decay to a graph



Reconstruct a generic decay

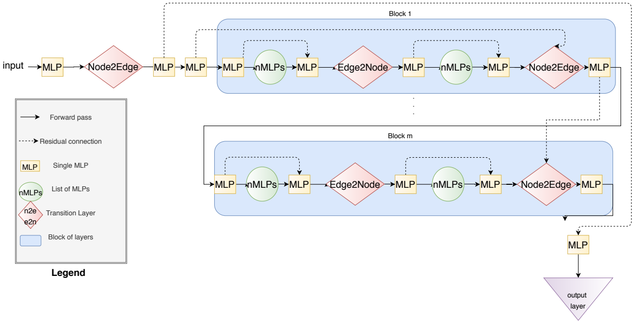


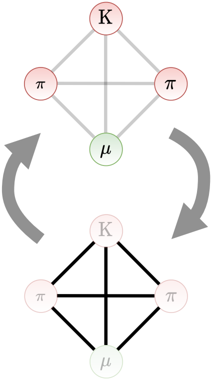
Feature Matrix

E_{K}	E_π	E_π	E_μ
$p_{x\mathrm{K}}$	$p_{x\pi}$	$p_{x\pi}$	$p_{x\mu}$
$p_{y\mathrm{K}}$	$p_{y\pi}$	$p_{y\pi}$	$p_{y\mu}$
$p_{z\mathrm{K}}$	$p_{z\pi}$	$p_{z\pi}$	$p_{z\mu}$

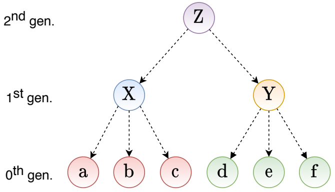
Our Graph Network

- Edge Label prediction using Neural Relational Inference [arXiv:1802.04687]
- Models built using PyTorch and PyTorch Geometric
- Hyperparameter otimisation with Optuna
- About 50 000 free parameters
- Train with millions of simulated decays

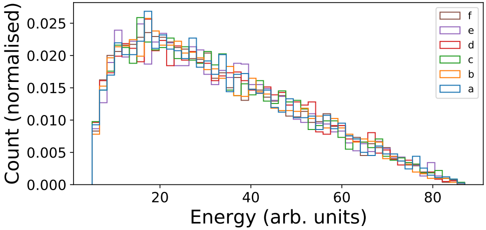


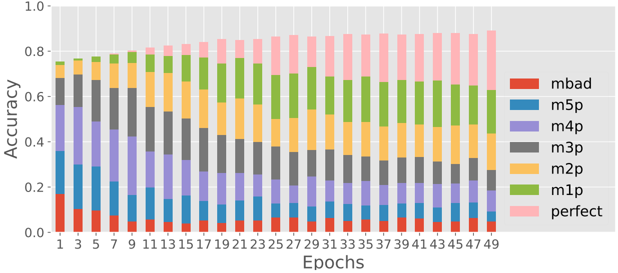


Proof of concept

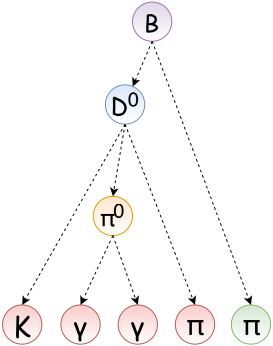


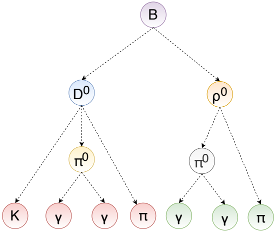
Particle	Z	X	Y	a,b,c	d,e,f
Mass	200	80	60	5	5





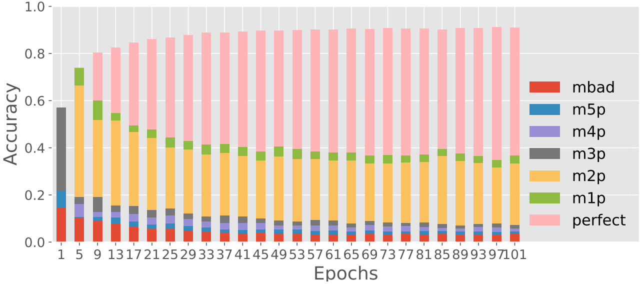
A more realistic test





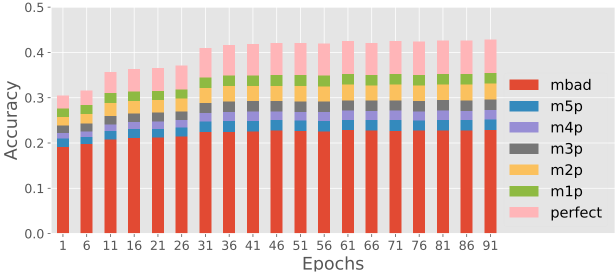
Decay Channels generated with the Belle II software		
Decay Channel	N^{o}FSPs	Motivation
$B^+ \to \overline{D^0}(\to K^+\pi^-\pi^0)\pi^+$	**5**	**benchmark tag side** on T.Keck's PhD thesis on FEI
$B^+ \to D^-(\to \pi^-\pi^+\pi^+)\pi^+\pi^+$	**5**	two 3-body decays, **overlapping spectra**, same FSPs)
$B^+ \to \overline{D^0}(\to K^+\pi^-\pi^0)e^+\nu_e$	**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



Test on generic Belle II MC

- Use MC generated information (not yet reconstructed)
- Overall efficiency $\mathcal{O}(10\%)$



Conclusion

- Showcased a Graph Deep Neural Network to reconstruct a generic decay
- Encouraging preliminary results
- More realistic tests still needed

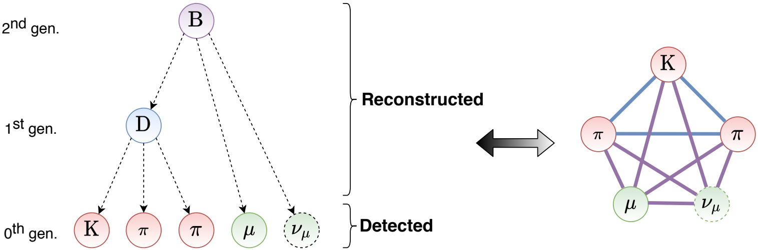
The team

KIT (ETP)	J. Kahn, T. Boeckh
UoS (IPHC)	I. Tsaklidis, G. Dujany
KIT (SCC)	O. Taubert, M Götz

- Most of the work performed by Ilias Tsaklidis during his joint IPHC-KIT internship financed by the Belle2NewPhy Seed Money project
- Principal Investigators: Isabelle Ripp-Baudot (IPHC), Pablo Goldenzweig (KIT)
- Exploited HoreKa computing center at KIT, plan to use also CC-IN2P3

BACKUP

Going from a decay to a graph



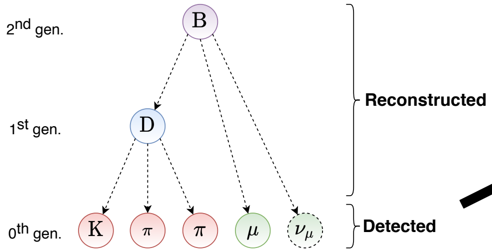
Adjacency Matrix

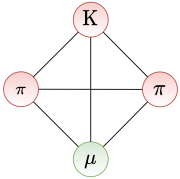
	B	D	K	π	π	μ	ν_μ
B	0	1	0	0	0	1	1
D	1	0	1	1	1	0	0
K	0	1	0	0	0	0	0
π	0	1	0	0	0	0	0
π	0	1	0	0	0	0	0
μ	1	0	0	0	0	0	0
ν_μ	1	0	0	0	0	0	0

Lowest Common Ancestor (LCA) Matrix

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

Reconstruct a generic decay





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}	$p_{z\pi}$	$p_{z\pi}$	$p_{z\mu}$

Insert GraphNN here

Lowest Common Ancestor (LCA) Matrix

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

Hyperparameters' optimisation

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
$B^+ \to \overline{D^0}(\to K^+\pi^-\pi^0)\pi^+$	32	0.001	0.008520	512	4	1	45056
$B^+ \to \overline{D^0}(\to K^+\pi^-\pi^0))e^+\nu_e$	32	0.001	0.008520	512	4	1	45056
$B^+ \to D^-(\to \pi^-\pi^+\pi^+)\pi^+\pi^+$	64	0.00062	0.1883	128	4	12	33792
$B^+ \to \overline{D^0}(\to K^+\pi^-\pi^0)\rho(\to \pi^-\pi^0)$	16	0.00036	0.0624	128	4	12	33792
$B^+ \to \overline{D^0}(\to K^+\pi^-\pi^0)\omega(\to \pi^+\pi^-\pi^0)\pi^+$	16	0.000485	0.0304	128	4	12	33792
$B^+ \to D^-(\to \pi^-\pi^-\pi^+\pi^0)\pi^+\pi^+\pi^0$	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

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