Inclusive Flavour Tagging at LHCb

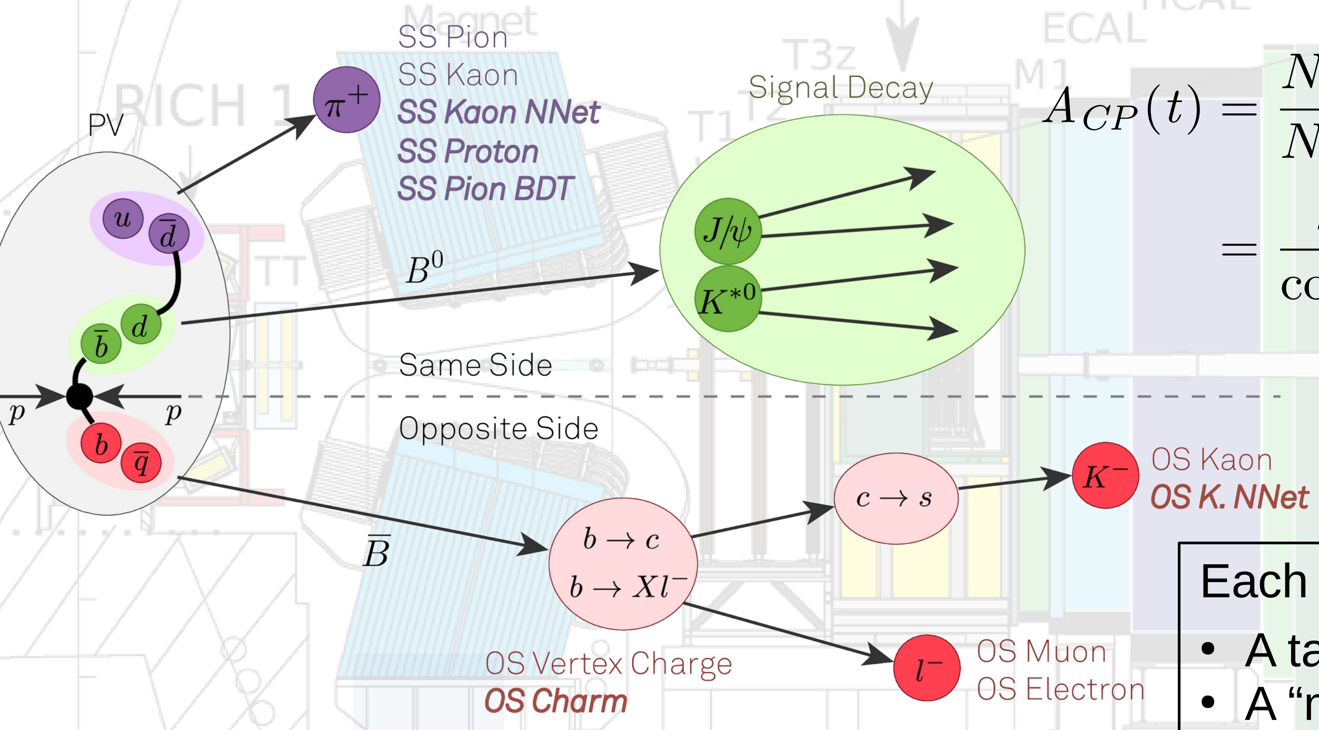
FSP Meeting Bonn
October 6th 2020

Vukan Jevtic
on behalf of the Flavour Tagging working group

Technische Universität Dortmund

Flavour Tagging Technique

Time dependent measurements of mixing parameters / CP violation require knowledge about initial B flavour



$$A_{CP}(t) = \frac{N(\bar{B}^0 \to f)(t) - N(B^0 \to f)(t)}{N(\bar{B}^0 \to f)(t) + N(B^0 \to f)(t)} = \frac{S_f \sin(\Delta m_q t) - C_f \cos(\Delta m_q t)}{\cosh(\frac{1}{2}\Delta\Gamma_q t) + \mathcal{A}_{\Delta\Gamma} \sinh(\frac{1}{2}\Delta\Gamma_q t)}$$

Each tagging algorithm provides:

- A tag decision (b or $\bar{b}$ or unknown)
- A “mistag probability” η

Flavour Tagging Technique

$$A_{CP}^{\text{measured}}(t) = \frac{N(\bar{B}^0 \to f)(t) - N(B^0 \to f)(t)}{N(\bar{B}^0 \to f)(t) + N(B^0 \to f)(t)}$$

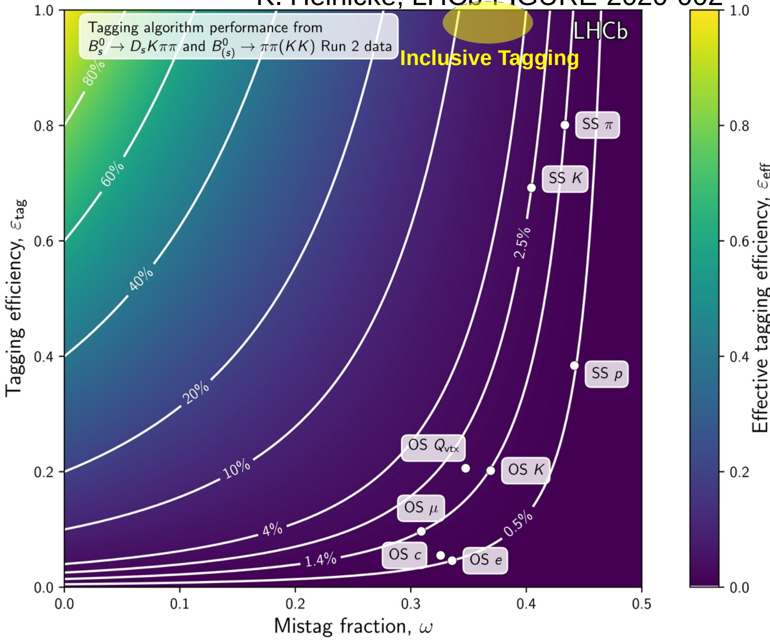
$$= (1 - 2\omega) A_{CP}^{\text{true}}(t)$$

$$\omega = \frac{N_{\text{wrong}}}{N_{\text{wrong}} + N_{\text{right}}} \qquad \varepsilon_{\text{tag}} = \frac{N_{\text{right}} + N_{\text{wrong}}}{N_{\text{tot}}}$$

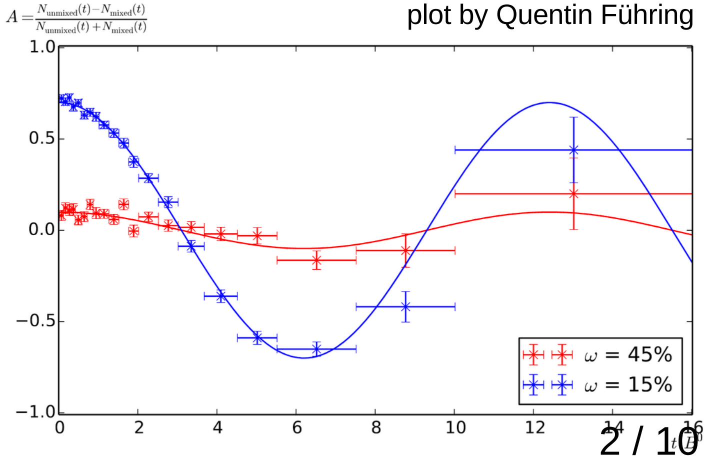
Measurement sensitivity:

$$\sigma^{-1}_{A_{CP}^{\text{true}}} \propto \sqrt{\varepsilon_{\text{tag}}(1 - 2\omega)^2 N} = \sqrt{\varepsilon_{\text{eff}} N}$$

K. Heinicke, LHCb-FIGURE-2020-002

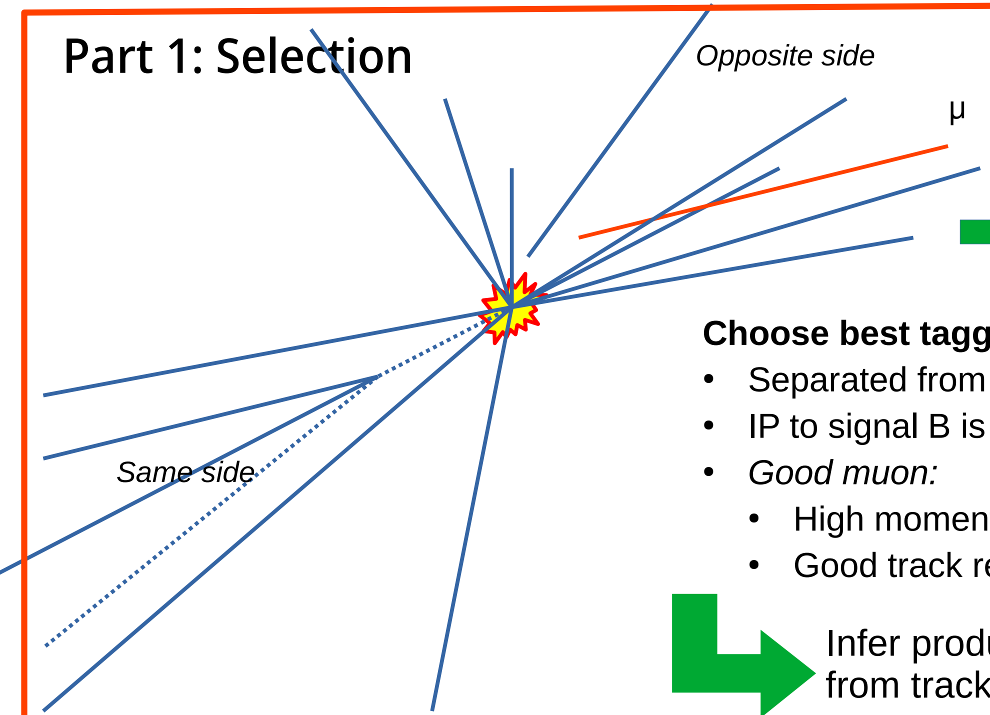


plot by Quentin Führing



Standard Flavour Tagging Approach Example
The opposite side muon tagger

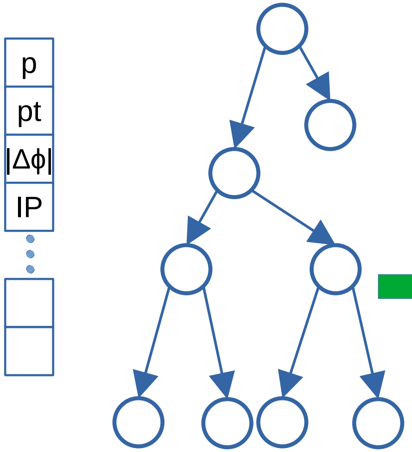
Part 1: Selection



Choose best tagging muon:
- Separated from signal in phi
- IP to signal B is large
- *Good muon:*
  - High momentum
  - Good track reconstruction

Infer production flavour from track charge

Part 2: Tagging via BDT



Mistag probability
η

Inclusive Tagging Approach

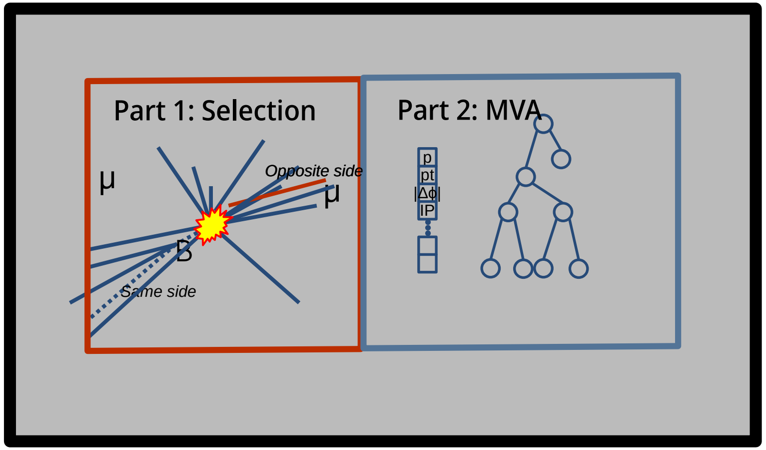
Full event interpretation

Full event

i.e. **all reconstructed tracks** with union of all single tagger track features including

- **Track charge**
- Particle identification
- Track momenta
- Topological information like
  - Phi + eta angle difference w.r.t. B meson

"Inclusive Tagger"



Tag decision

Mistag probability

(d, η)

This idea is not completely new, but has never been fully implemented / optimized for LHCb

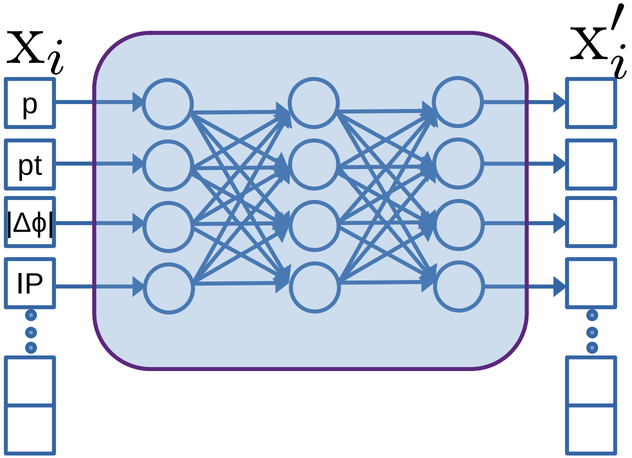
- T. Likhomanenko, D. Derkach, A. Rogozhnikov: "*Inclusive Flavour Tagging Algorithm*", ***J. Phys. Conf. Ser.*** *762 (2016) 012045*

The network architecture was then further optimized by

- *S. Akar, A. Camboni, D. P. O'Hanlon, B. Khanji* ***et al.*** → *Usage of long-short-term-memory units*

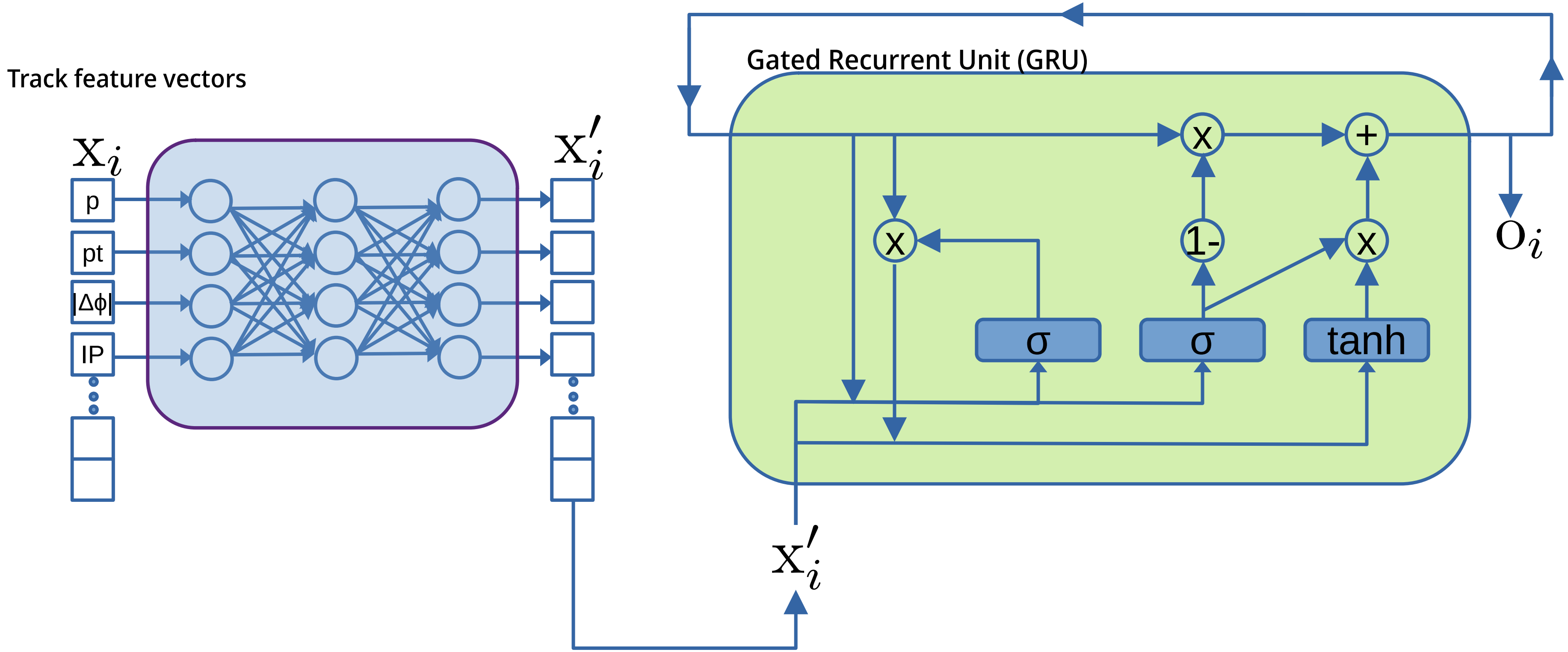
Current Inclusive Tagger Architecture

Track feature vectors

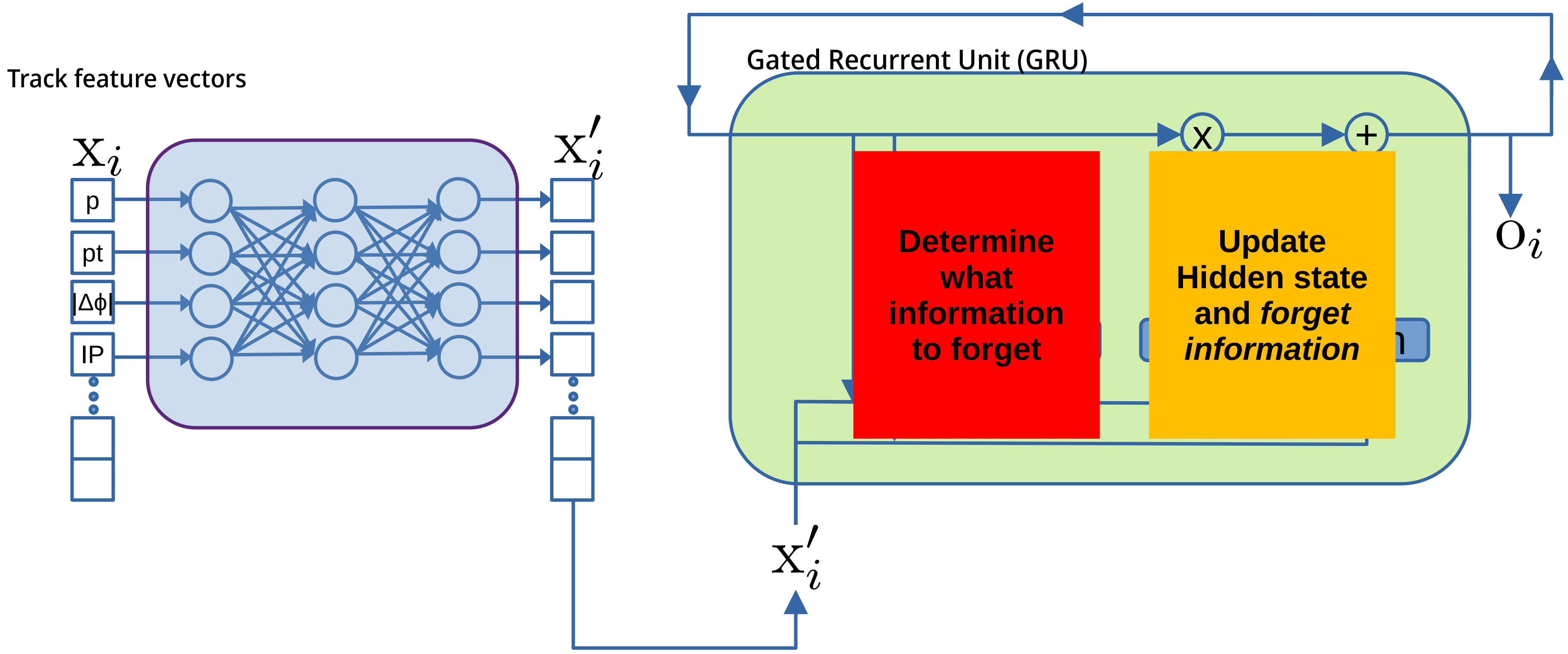


S. Akar
D. P. O'Hanlon
V. Chobanova
Q. Führing
M. Schiller
V. J.

Current Inclusive Tagger Architecture

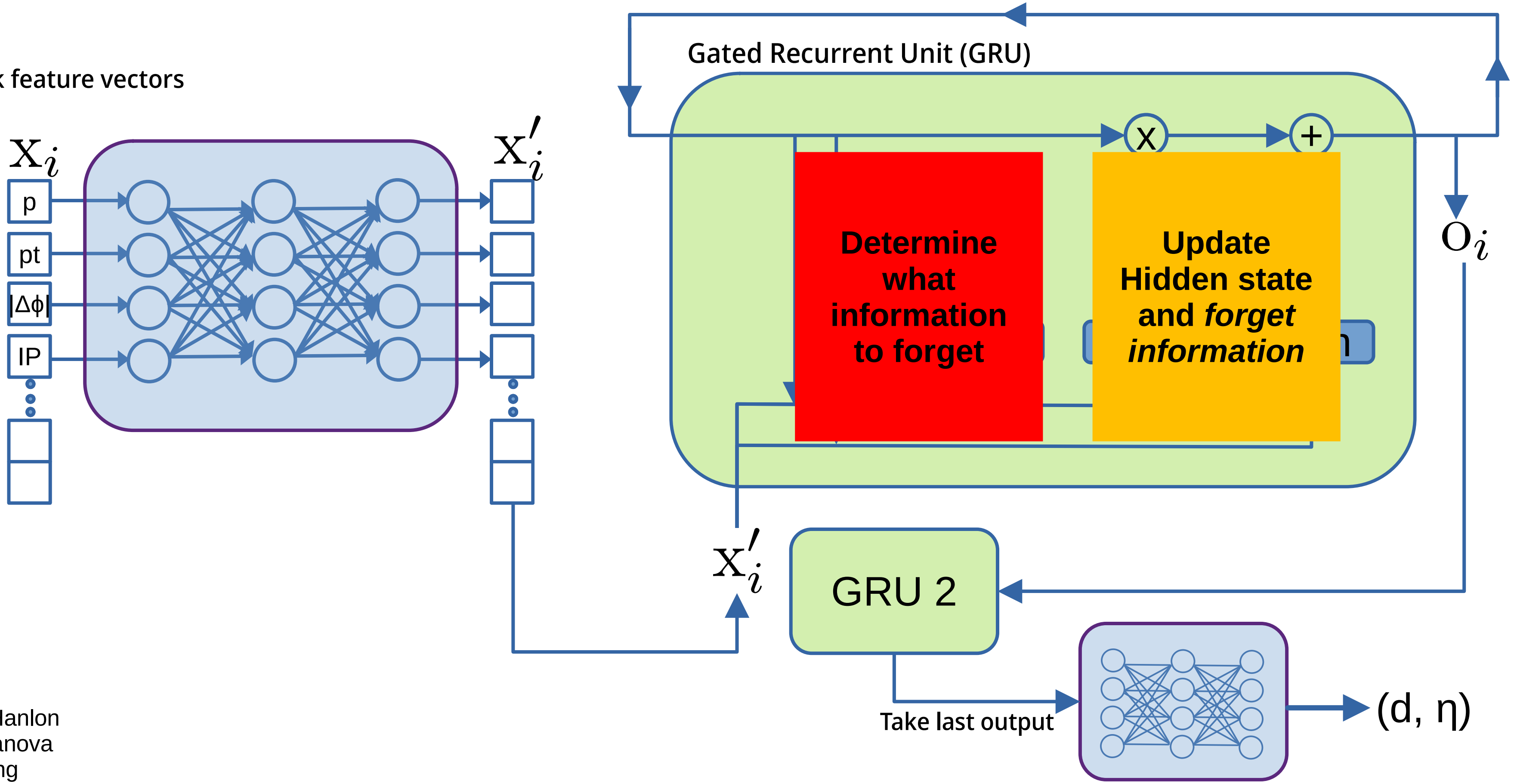


Current Inclusive Tagger Architecture



S. Akar
D. P. O'Hanlon
V. Chobanova
Q. Führing
M. Schiller
V. J.

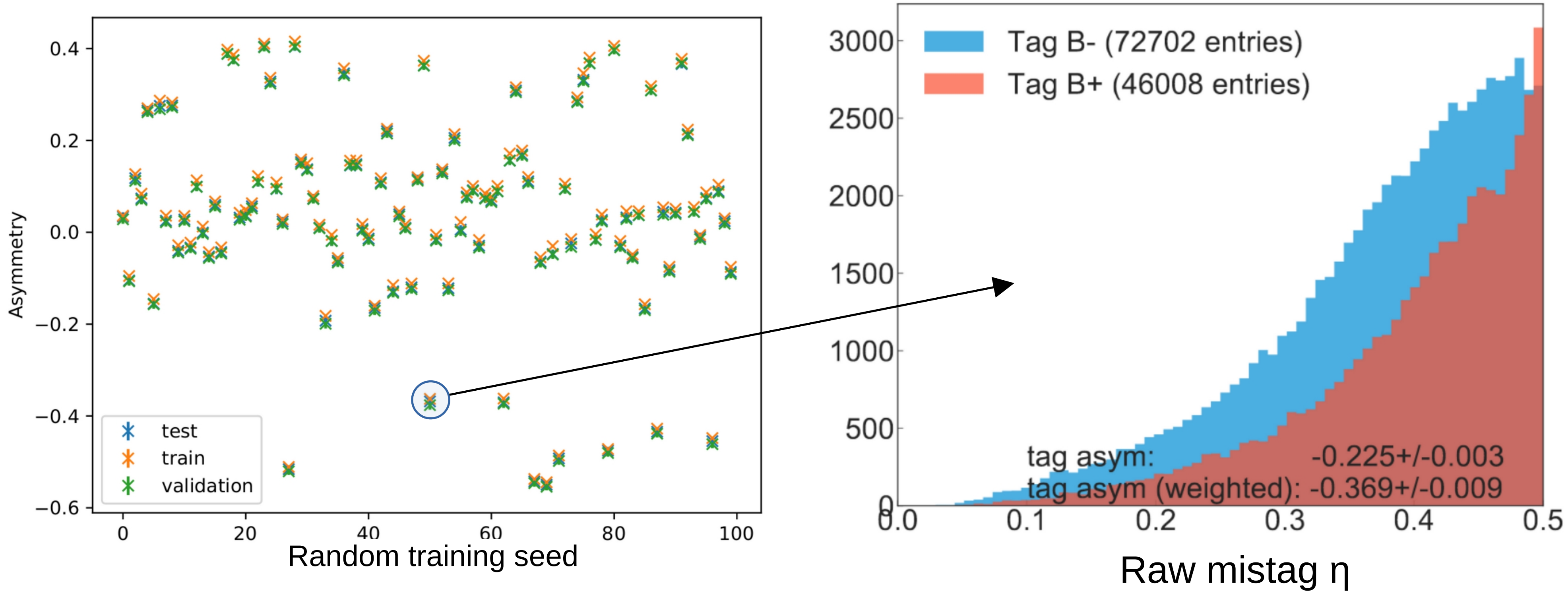
Current Inclusive Tagger Architecture



S. Akar
D. P. O'Hanlon
V. Chobanova
Q. Führing
M. Schiller
V. J.

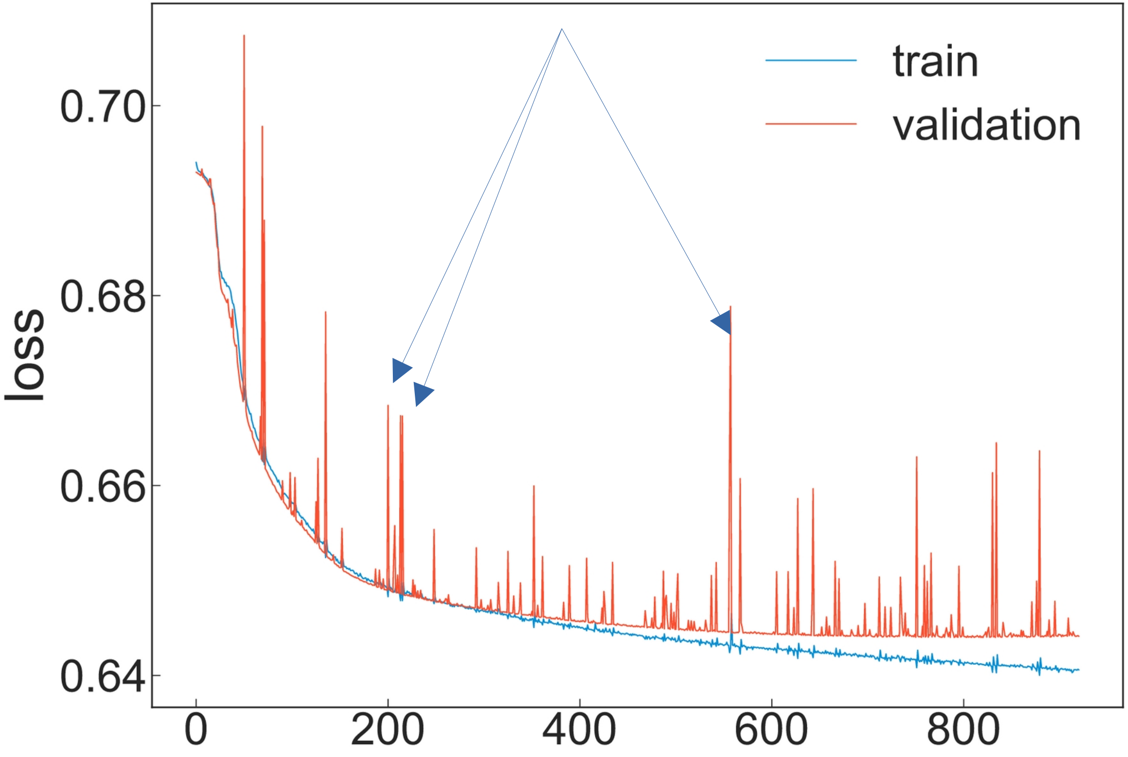
Training the network in practise has been quite a challenge...

Show stopper: Correlation between random seed and tag asymmetry



Happy end of asymmetry issues

Loss function “glitches” are strongly correlated to tag asymmetry



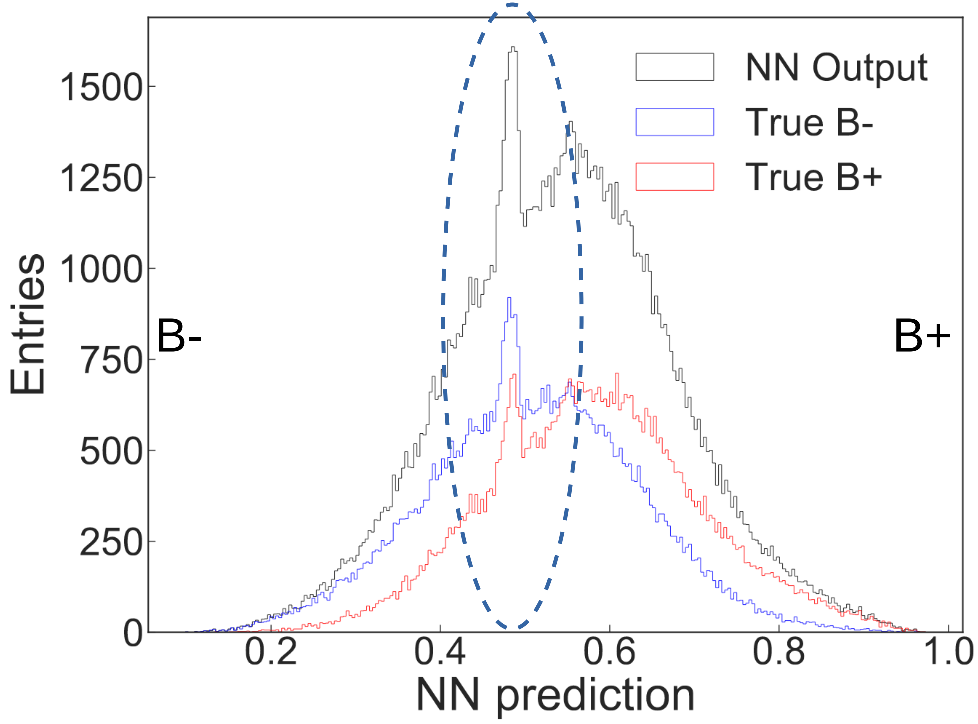
- Issue was that Network chose last training state instead of “best” state.
- Combined with these spikes, this resulted in huge asymmetries (still somewhat surprising to us)

But…

- Choosing best Network state still produced asymmetries of up to 10% though.
- Reduced to up to 1% by choosing perfectly balanced training samples, i.e. 100% balanced instead of 99.9% balanced. Quite surprising as well…

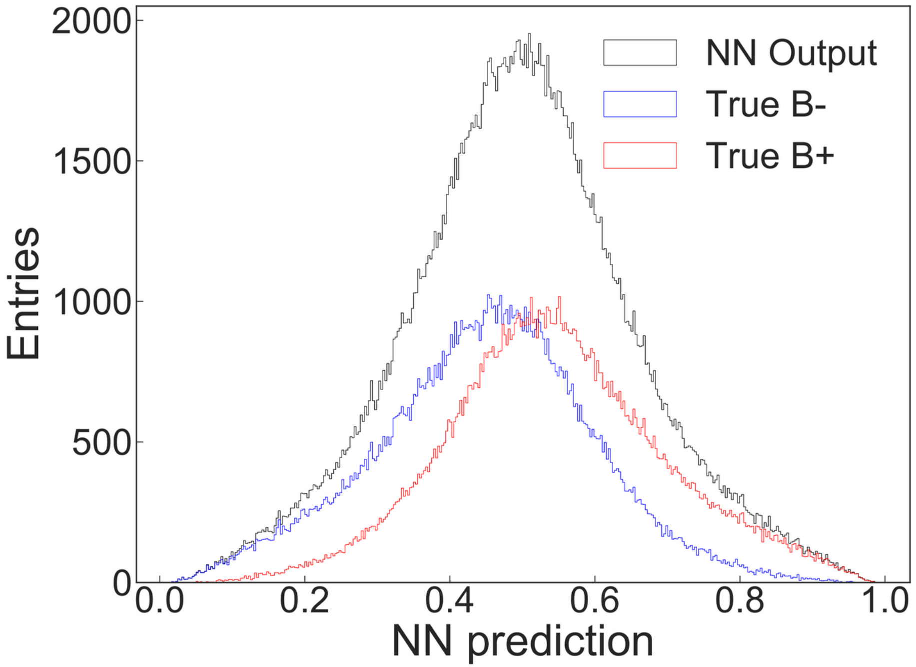
Issue Nr 2: Inconsistent features in the neural network output

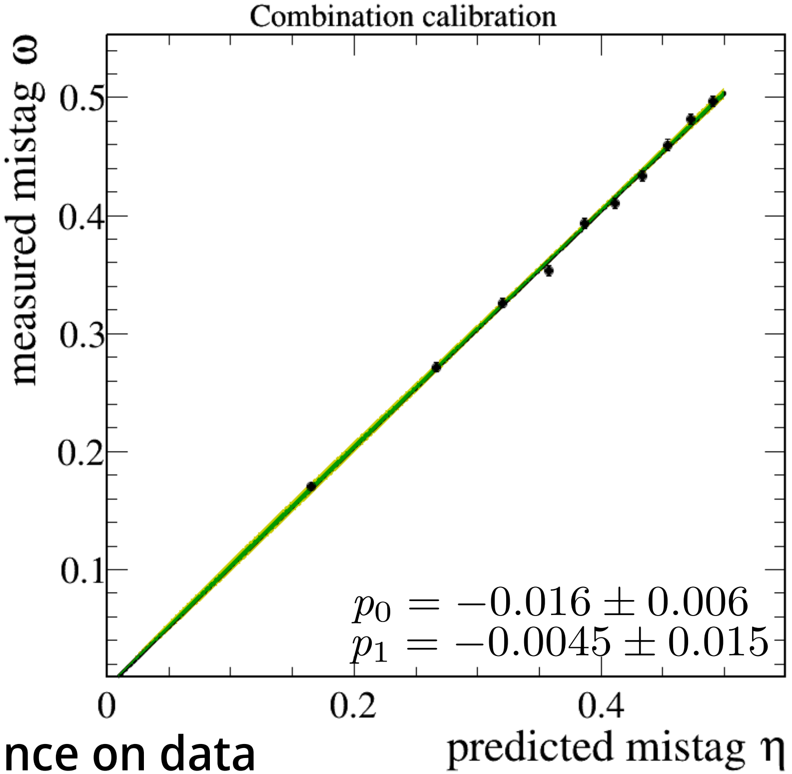
- Neural network output often contained one, sometimes multiple sharper spikes
- The higher the resolution, (i.e. the more statistics), the more such features could be observed
- In addition, these were very inconsistent between trainings



Solved by using smooth activation function instead of ReLU

Results on $B^+ \to J/\psi K^+$ Simulation





Tagging power $\varepsilon_{\text{eff}} \approx 9.7\%$, very similar performance on data

Tagging efficiency: 100%

Classical tagger combination: $\varepsilon_{\text{eff}} \approx 6\%$

Factor 1.6 improvement

Summary

- Inclusive tagging models have been developed and trained successfully
- Observed tagging performance on B+2JpsiK+ improvement is promising
- Major show-stoppers have been understood
  - Development of taggers for neutral modes + Bs mode in progress

Run 3

- Work is ongoing to make the IFT (and classical taggers) run efficiently in the HLT
- If validation studies are convincing, inclusive Tagger could replace classical taggers? → Easier to maintain
- Decision will be made soon