

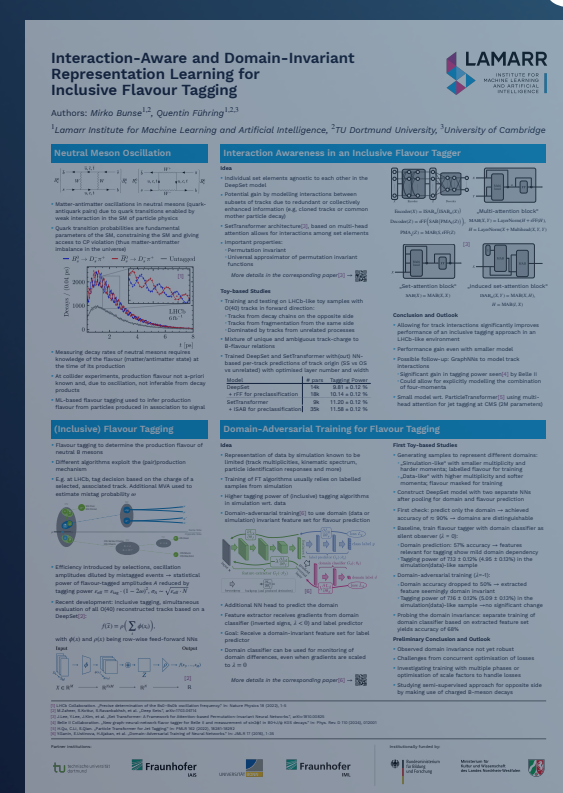
Interaction-Aware and Domain-Invariant Representation Learning for Inclusive Flavour Tagging

EuCAIFCon 2025, Cagliari

Foundations Models – Flash Talk for Poster Session B

Mirko Bunse^{1,2}, Quentin Führung^{1,2,3}

¹TU Dortmund University, ²Lamarr Institute for Artificial Intelligence and Machine Learning, ³University of Cambridge



Partner institutions:



Institutionally funded by:

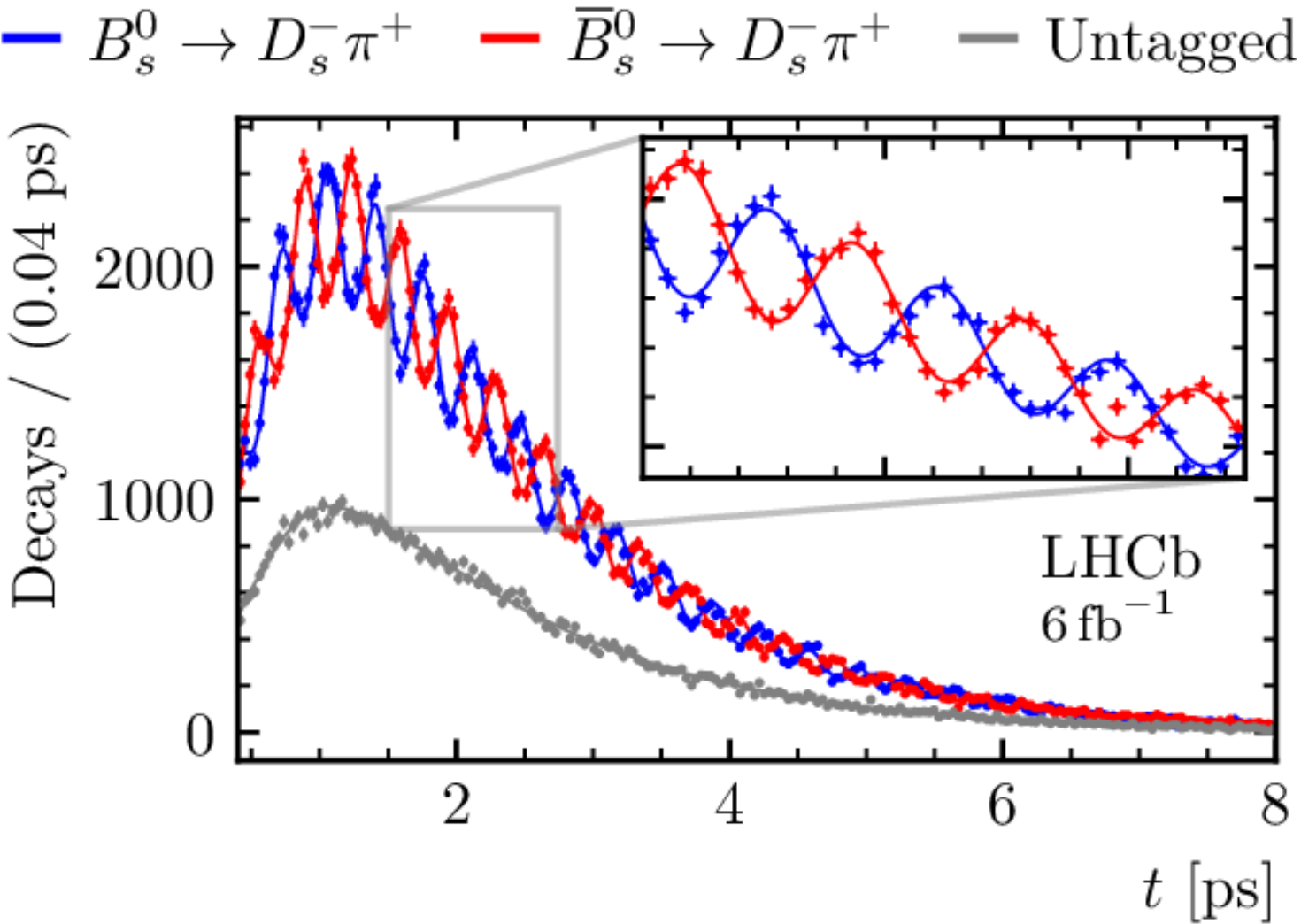
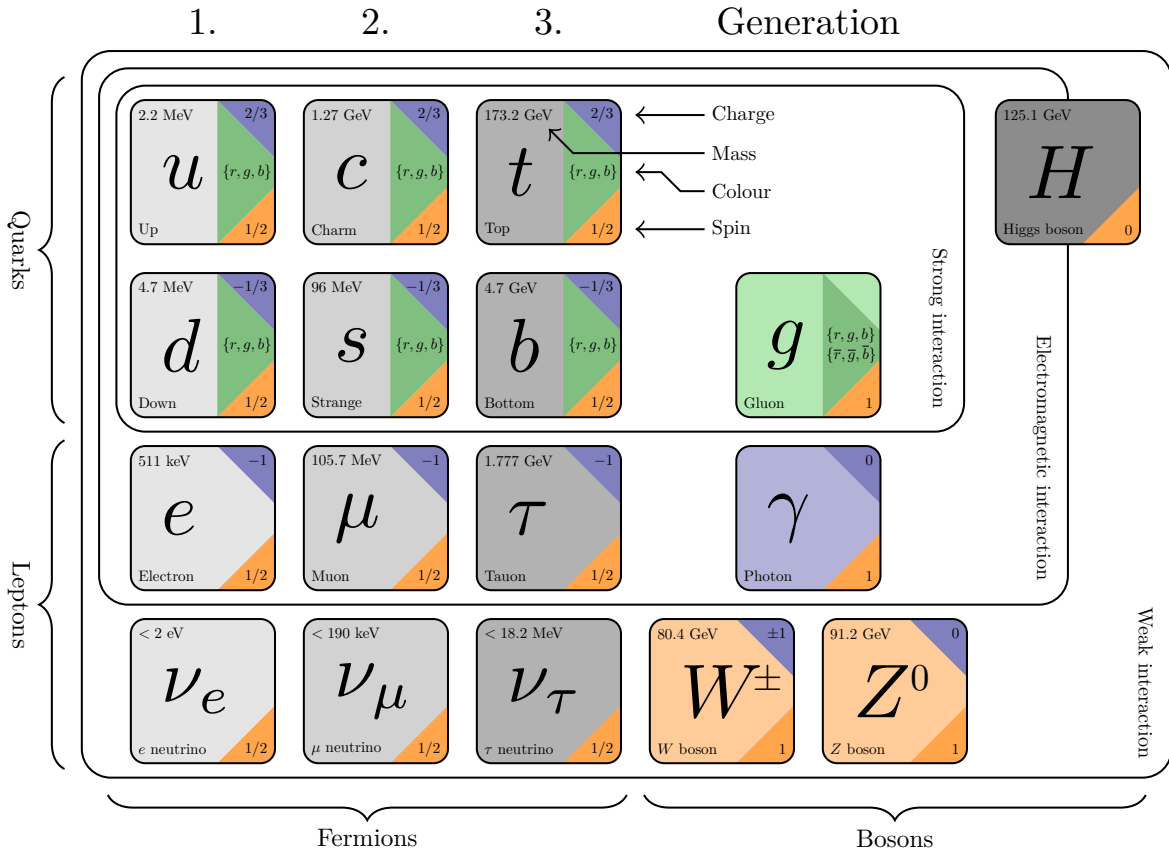
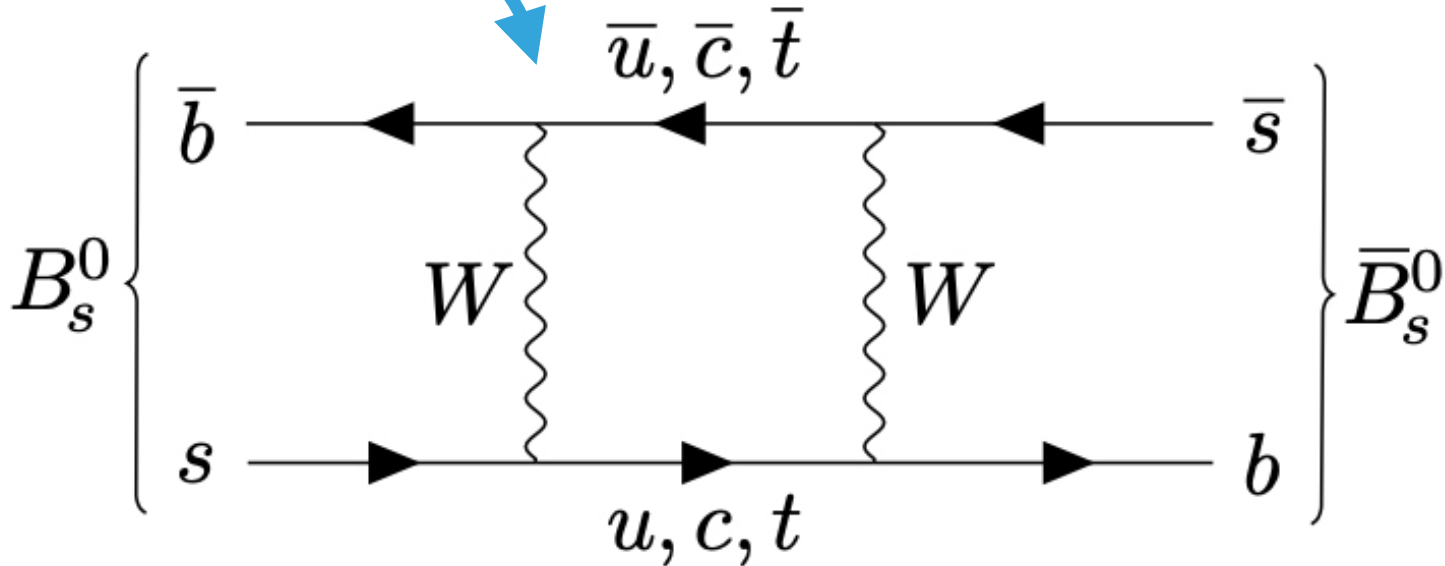


Physics Motivation: Neutral $B_{(s)}^0$ -meson oscillations



Quark-mixing probabilities:
Fundamental properties of the SM

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}$$



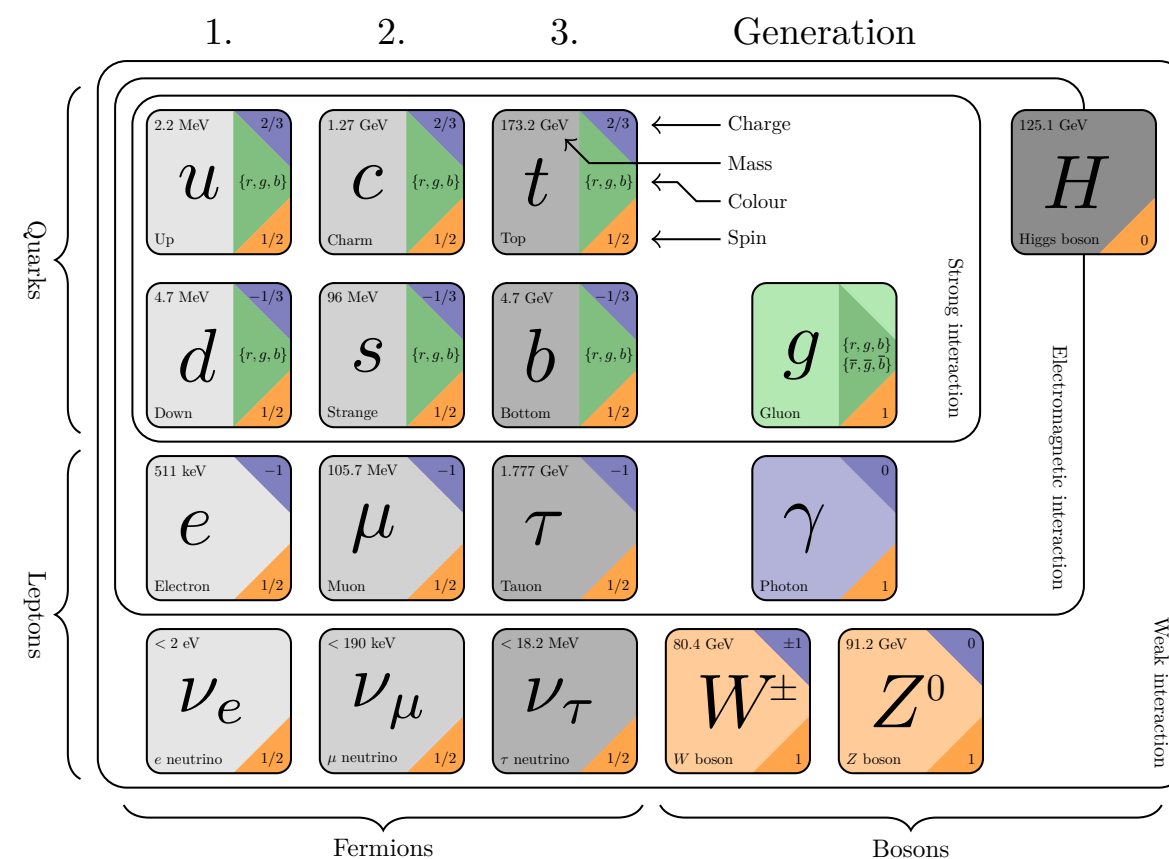
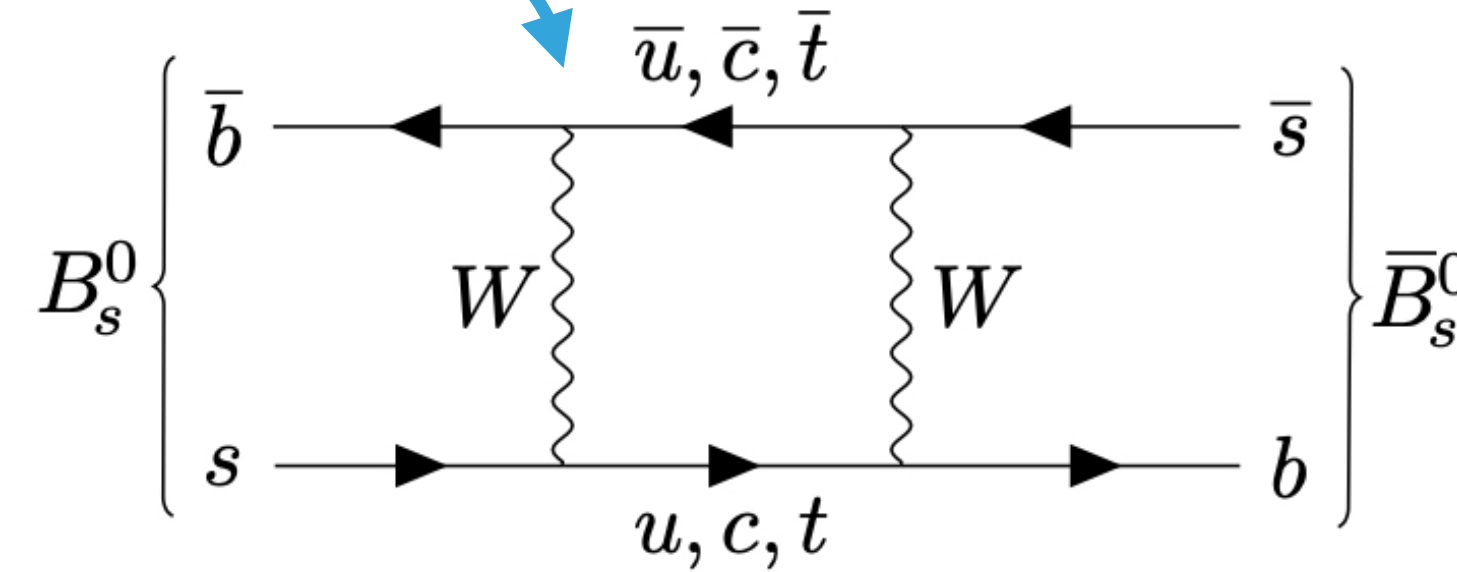
[LHCb collaboration, Precise determination of the B_s - $B_{\bar{s}}$ oscillation frequency, In: Nature Physics 18, 1-5 (2022)]

Physics Motivation: Neutral $B_{(s)}^0$ -meson oscillations



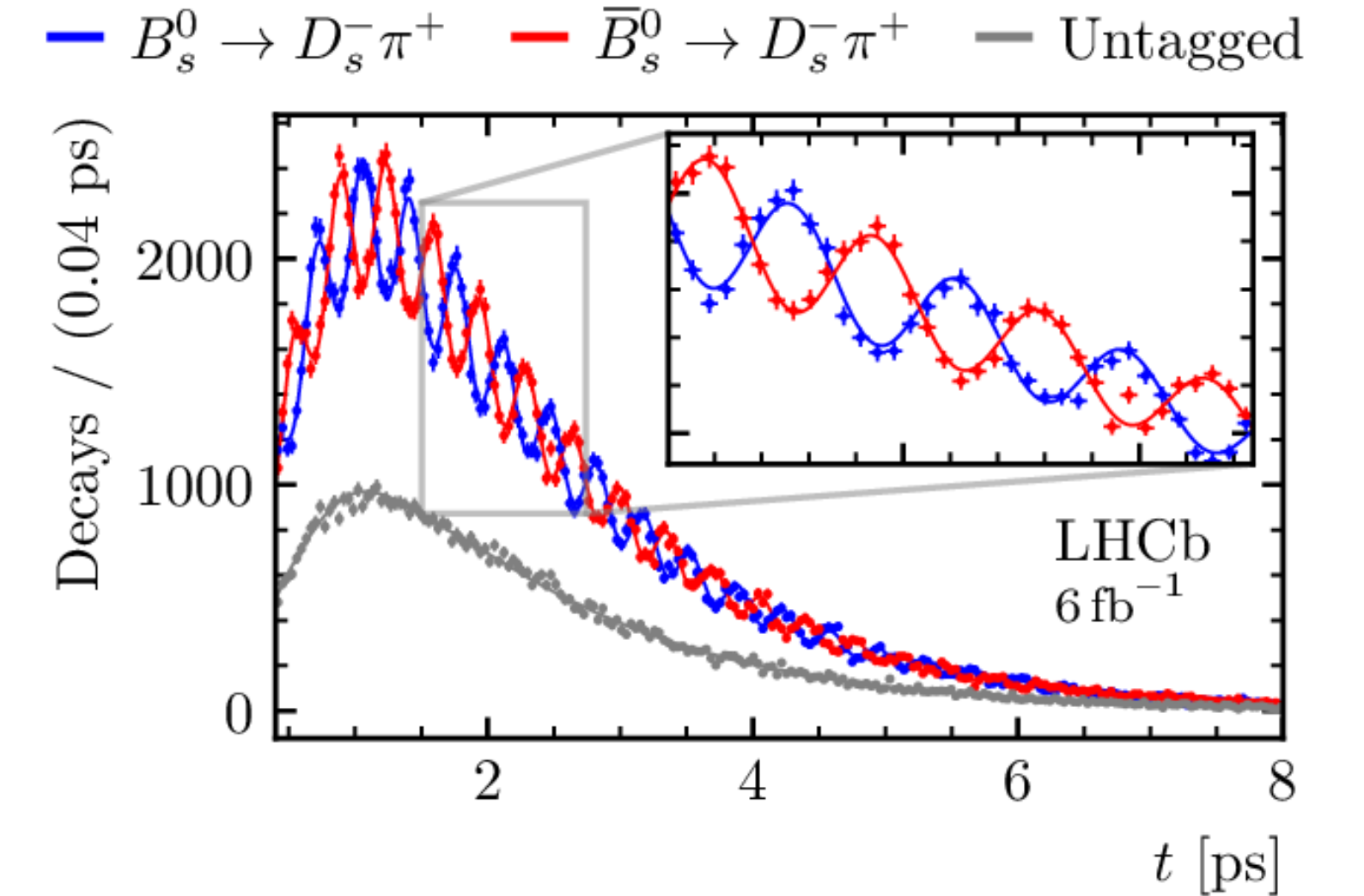
Quark-mixing probabilities:
Fundamental properties of the SM

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}$$



$$B_s^0 \rightarrow f$$

$$\bar{B}_s^0 \rightarrow B_s^0 \rightarrow f$$



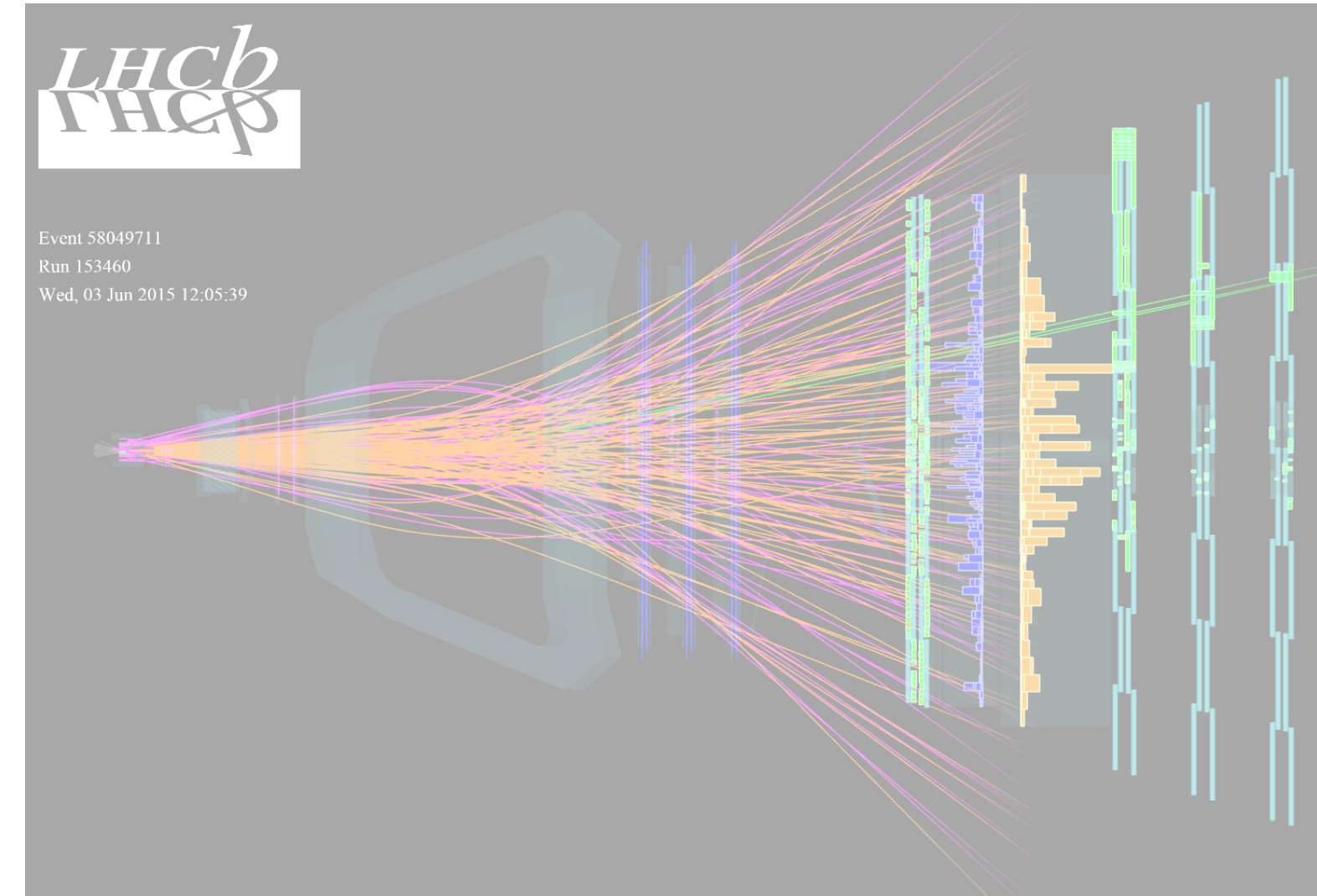
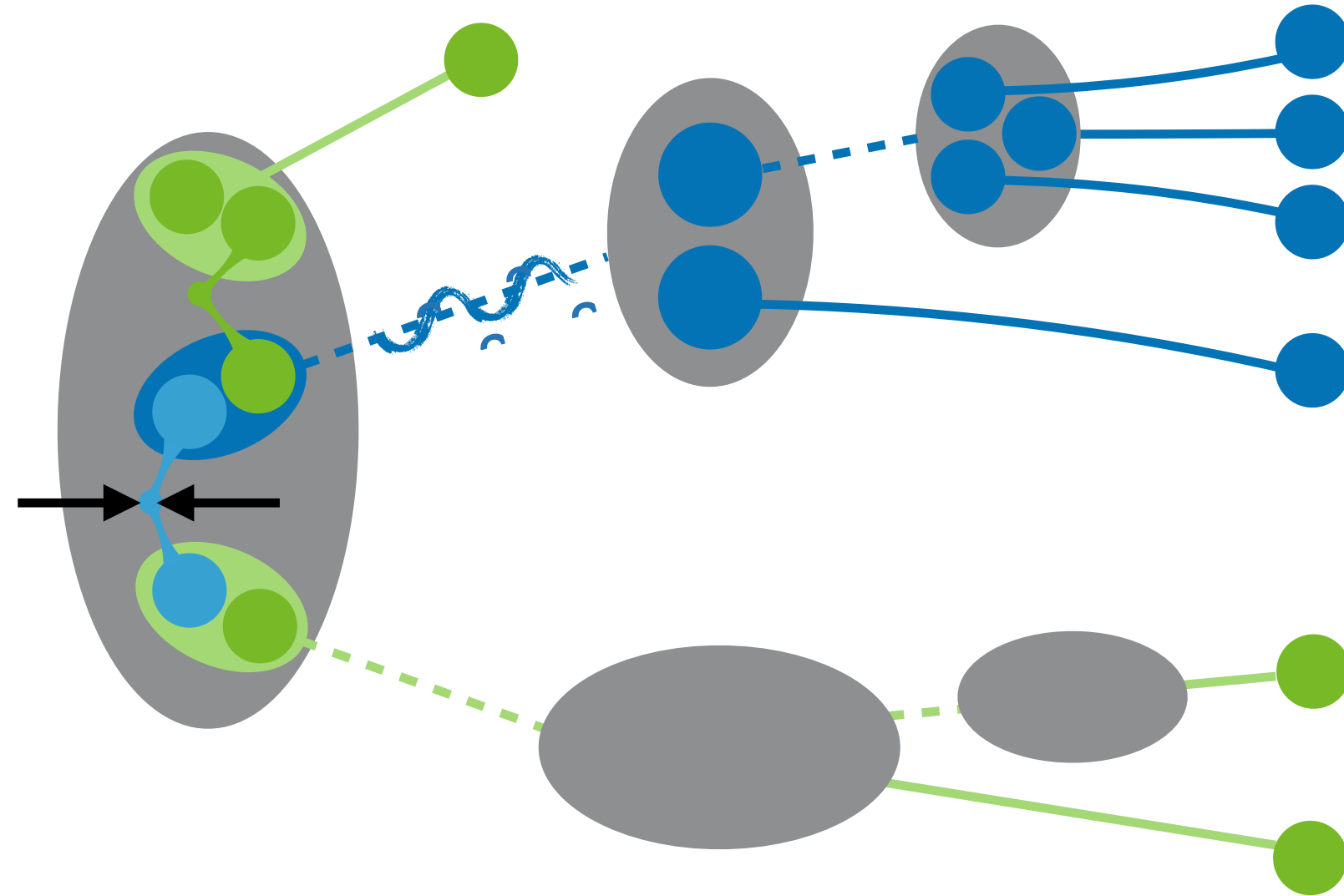
[LHCb collaboration, Precise determination of the B_s - $B_{\bar{s}}$ oscillation frequency, In: Nature Physics 18, 1-5 (2022)]

Oscillation measurements require
knowledge of the initial state (B_s^0 or $\bar{B}_s^0$)

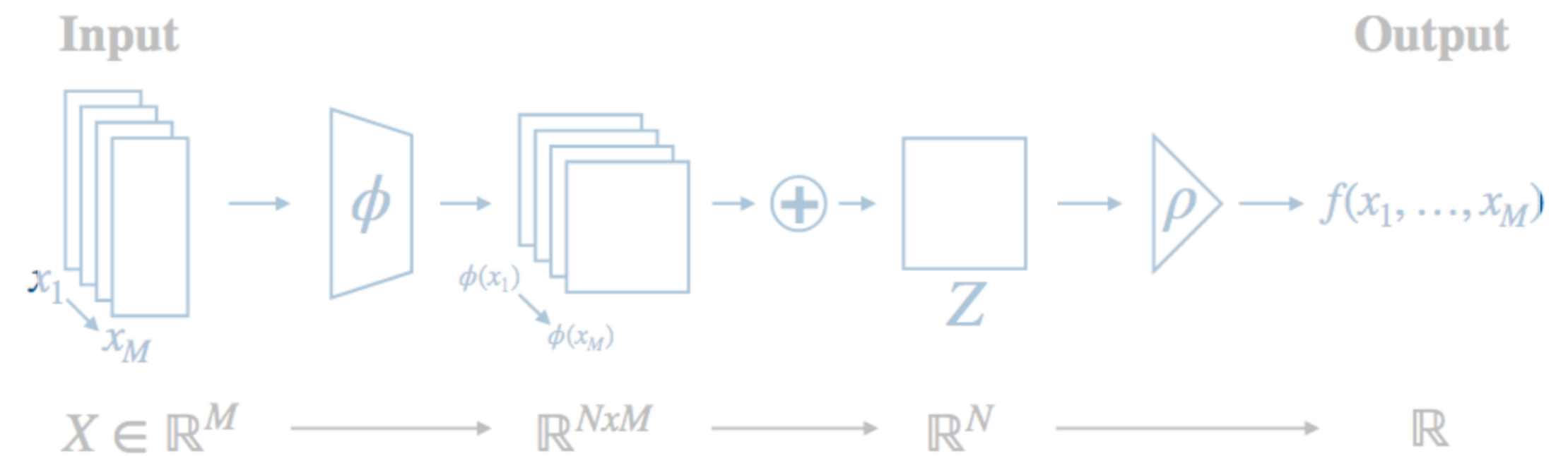
Inclusive Flavour Tagging

Oscillation measurements require knowledge of the initial state (B_s^0 or $\bar{B}_s^0$)

Flavour Tagging:
Algorithms exploiting specific processes



Inclusive Flavour Tagging:
Simultaneous analysis of all tracks with DeepSet NN

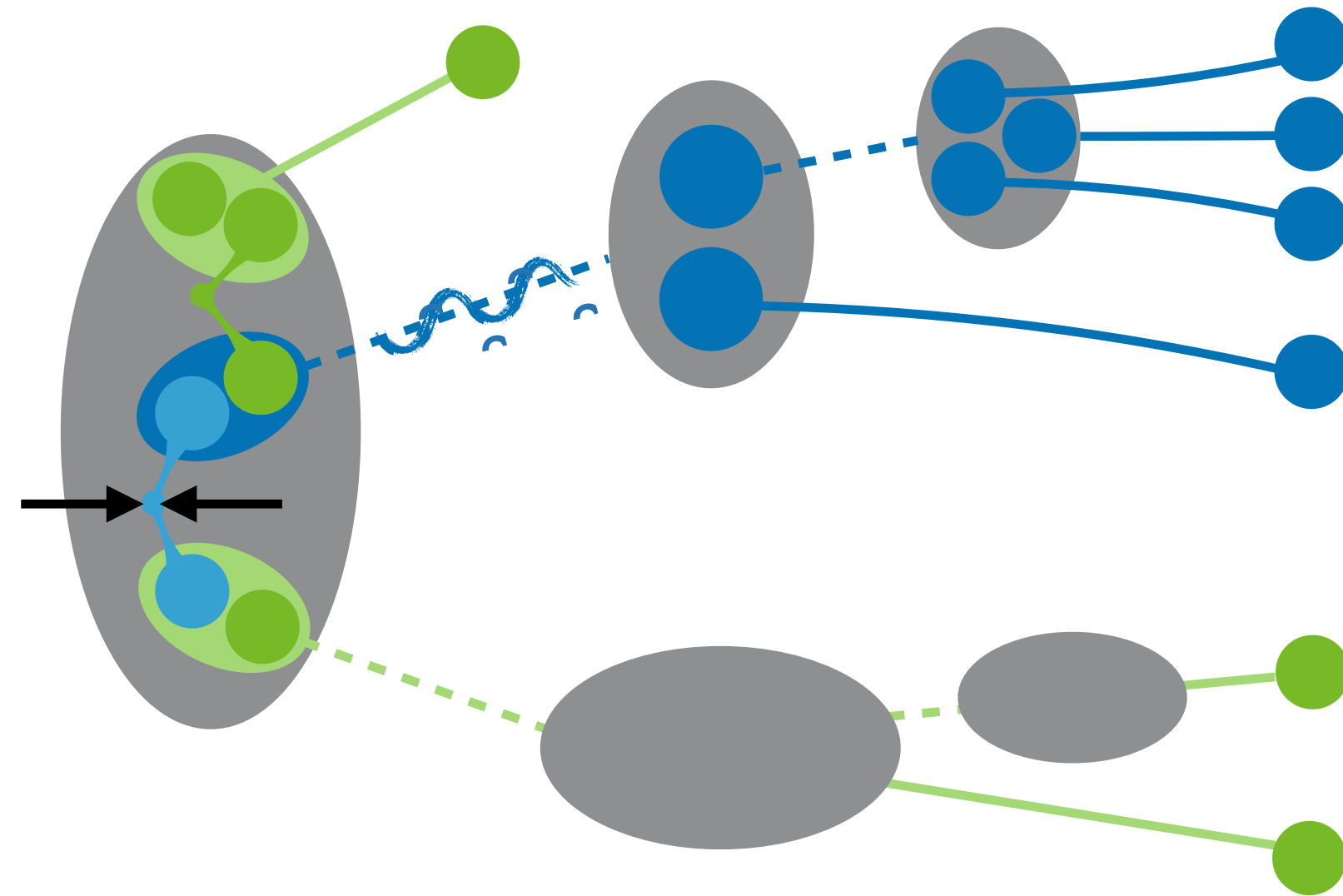


[Zaheer et.al., Deep Sets, arXiv:1703.06114]

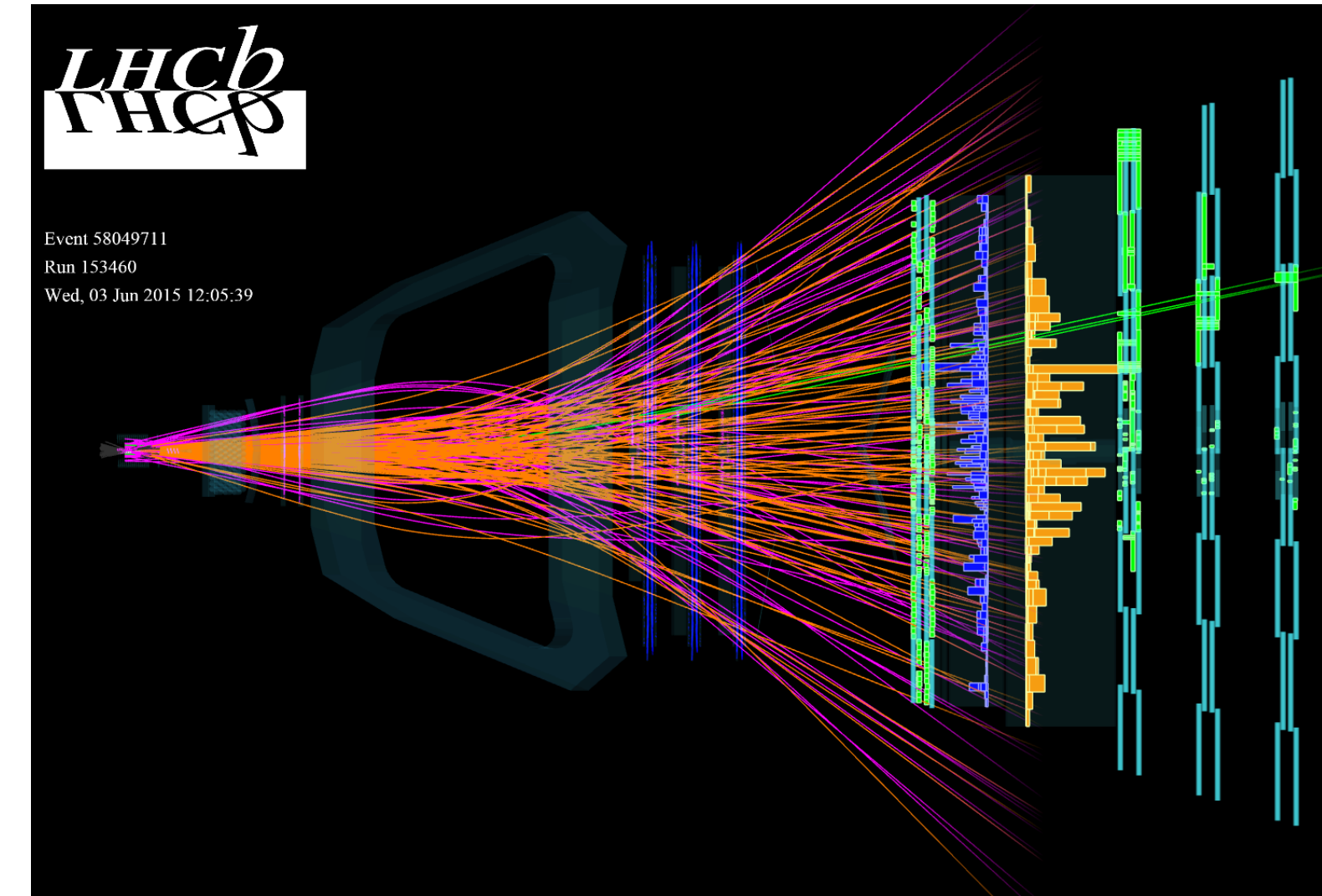
Inclusive Flavour Tagging

Oscillation measurements require knowledge of the initial state (B_s^0 or $\bar{B}_s^0$)

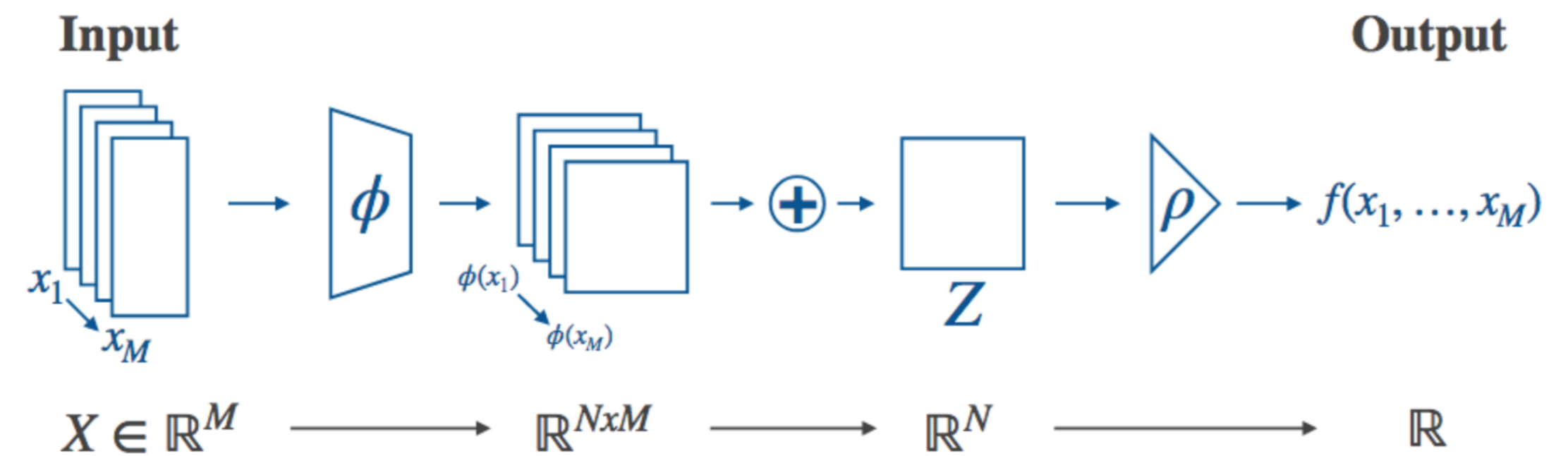
Flavour Tagging:
Algorithms exploiting specific processes



State of the Art
(at LHCb)

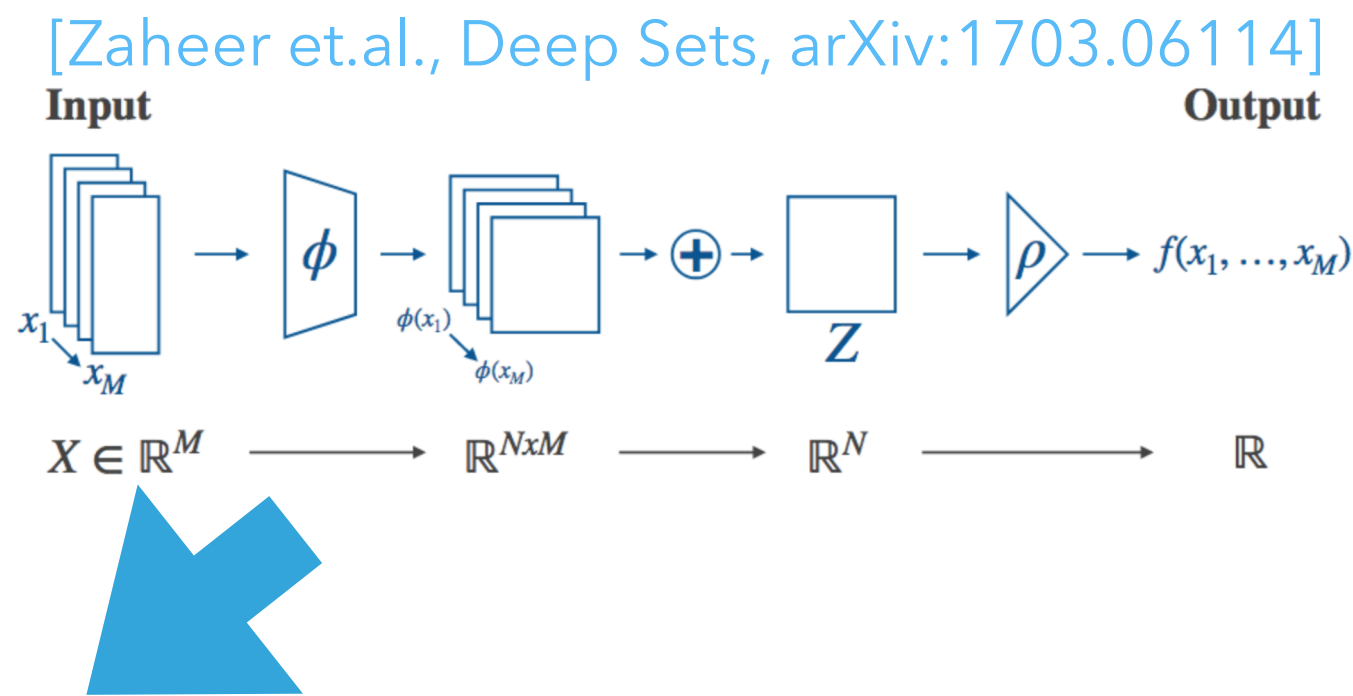


Inclusive Flavour Tagging:
Simultaneous analysis of all tracks with DeepSet NN



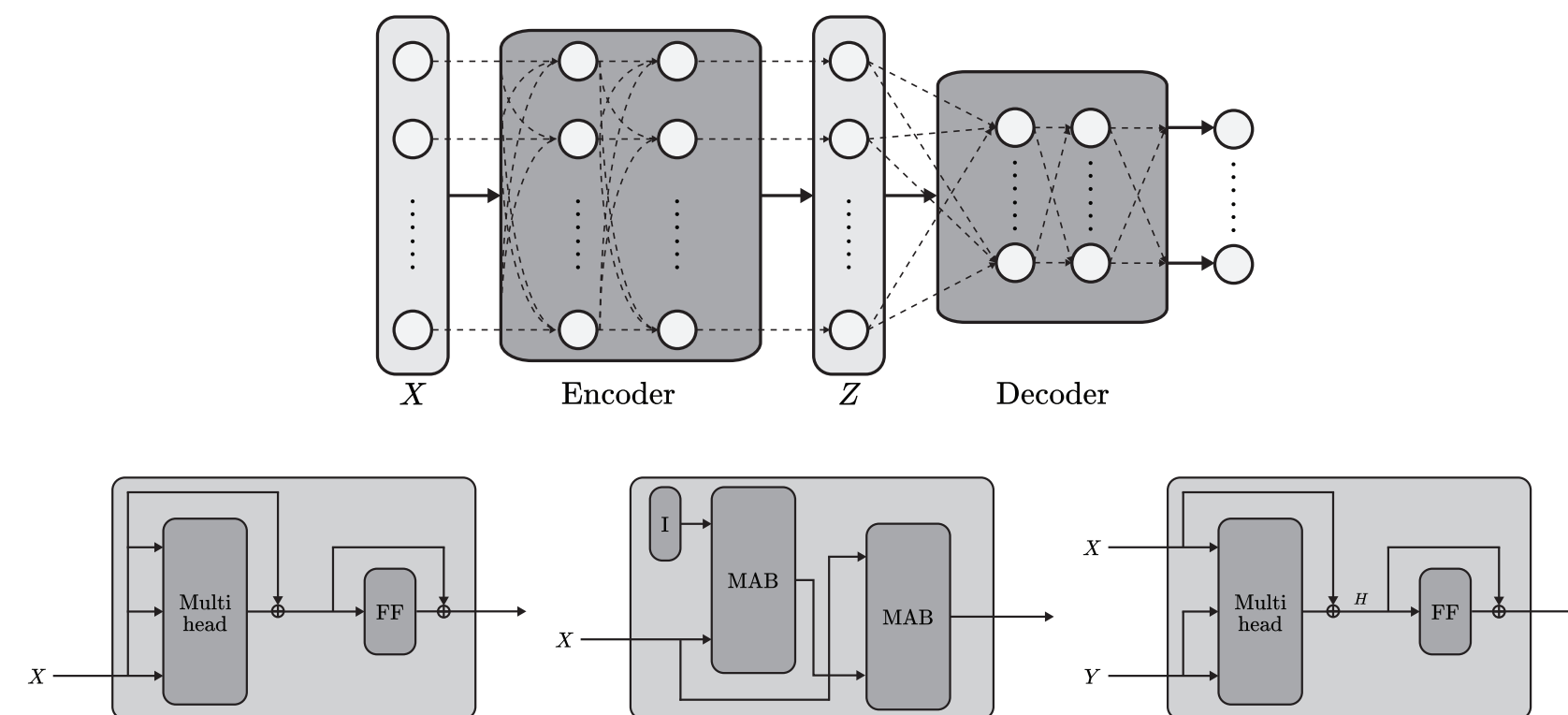
[Zaheer et.al., Deep Sets, arXiv:1703.06114]

Our work: Interaction awareness and domain invariance



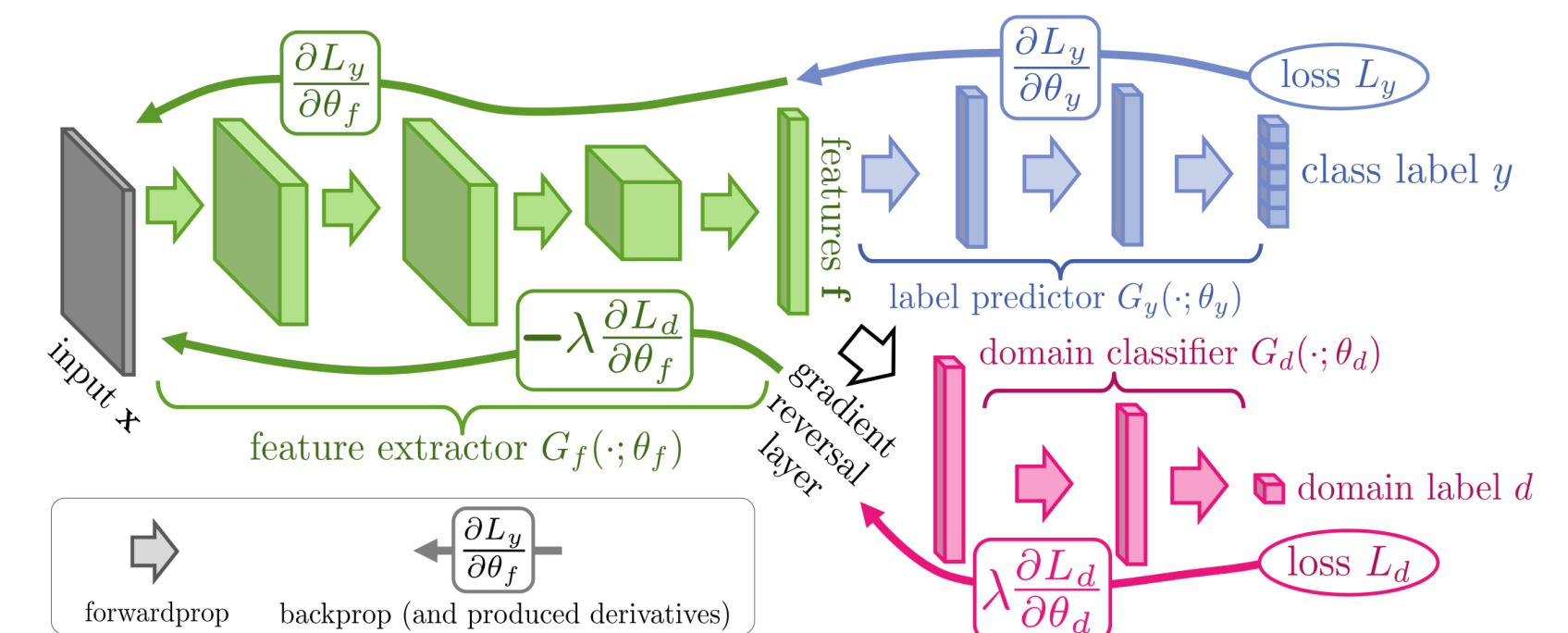
DeepSet NN
trained on simulation

Studying the SetTransformer architecture



[Lee et al., Set Transformer: A Framework for Attention-based Permutation-Invariant Neural Networks, arXiv:1810.00825]

Implementing domain-adversarial training



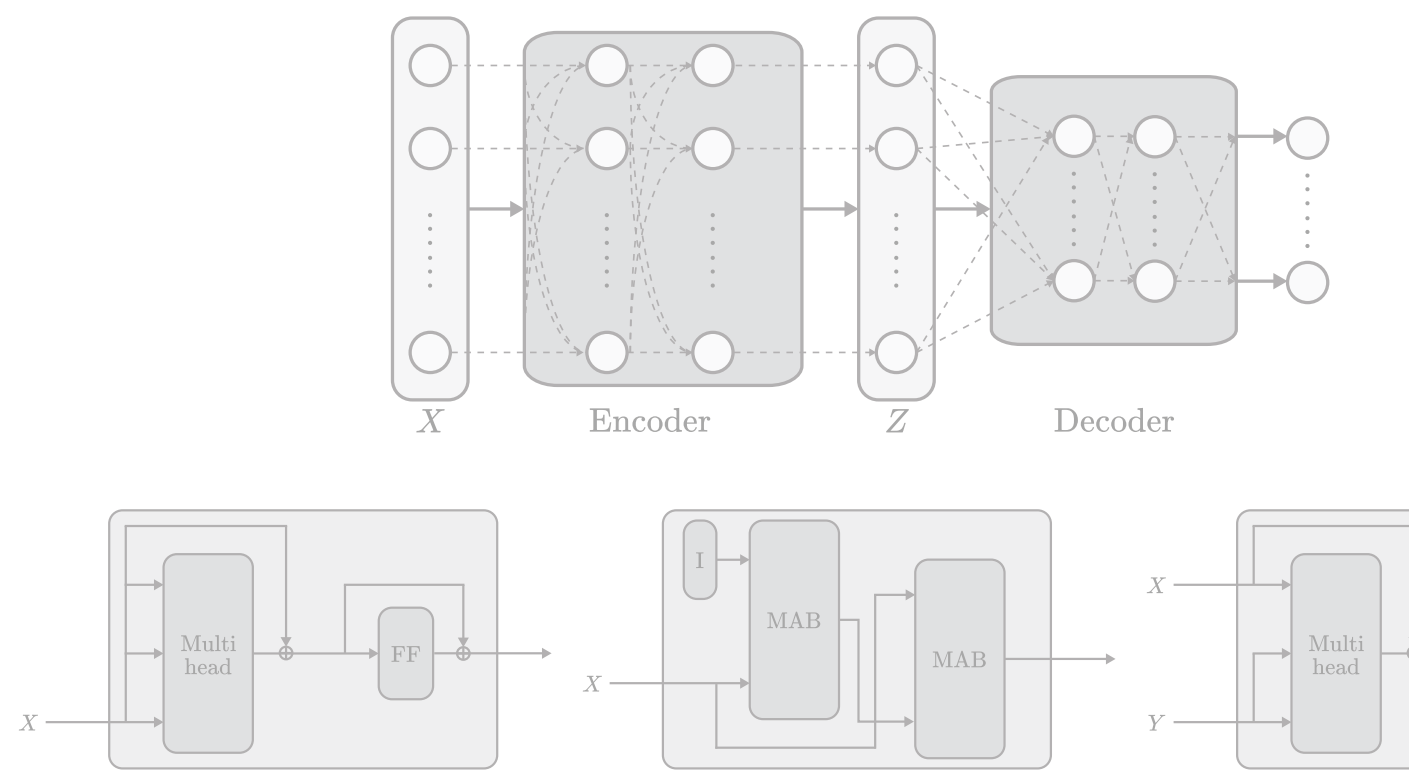
[Ganin et.al., Domain-Adversarial Training of Neural Networks, arXiv:1505.07818]

Our work: Interaction awareness and domain invariance



Studying the SetTransformer architecture

Significant improvement
in toy-based studies



[Lee et al., Set Transformer: A Framework for Attention Permutation-Invariant Neural Networks, arXiv:1810.00825]

Interaction-Aware and Domain-Invariant Representation Learning for Inclusive Flavour Tagging

Authors: Mirko Bunse^{1,2}, Quentin Fühling^{1,2,3}

¹Lamarr Institute for Machine Learning and Artificial Intelligence, ²TU Dortmund University, ³University of Cambridge

Neutral Meson Oscillation

Idea

- Individual set elements agnostic to each other in the DeepSet model
- Potential gain by modelling interactions between subsets of tracks due to redundant or collectively enhanced information (e.g. cloned tracks or common mother particle decay)
- SetTransformer architecture[3], based on multi-head attention allows for interactions among set elements
- Important properties:
 - Permutation invariant
 - Universal approximator of permutation invariant functions

More details in the corresponding paper[3] →

Toy-based Studies

- Training and testing on LHCb-like toy samples with O(40) tracks in forward direction:
 - Tracks from decay chains on the opposite side
 - Tracks from fragmentation from the same side
 - Dominated by tracks from unrelated processes
- Mixture of unique and ambiguous track-charge to B-flavour relations
- Trained DeepSet and SetTransformer with(out) NN-based per-track predictions of track origin (SS vs OS vs unrelated) with optimised layer number and width

Model	# pars	Tagging Power
DeepSet	14k	9.81 ± 0.12 %
+ rFF for preclassification	18k	10.14 ± 0.12 %
SetTransformer	9k	11.20 ± 0.12 %
+ ISAB for preclassification	35k	11.58 ± 0.12 %

(Inclusive) Flavour Tagging

Idea

- Flavour tagging to determine the production flavour of neutral B mesons
- Different algorithms exploit the (pair)production mechanism
- E.g. at LHCb, tag decision based on the charge of a selected, associated track. Additional MVA used to estimate mistag probability w

Efficiency introduced by selections, oscillation amplitudes diluted by mistagged events → statistical power of flavour-tagged amplitudes A reduced by tagging power $\epsilon_{\text{tag}} = \epsilon_{\text{tag}} \cdot (1 - 2w)^2$, $\epsilon_{\text{tag}} \sim \sqrt{\epsilon_{\text{tag}} \cdot N}$

Recent development: Inclusive tagging, simultaneous evaluation of all O(40) reconstructed tracks based on a DeepSet[2]:

$$f(\vec{x}) = \rho(\sum_i \phi(x_i))$$

with $\phi(x)$ and $\rho(x)$ being row-wise feed-forward NNs

Output

$X \in \mathbb{R}^N \rightarrow \mathbb{R}^{N \times M} \rightarrow \mathbb{R}^M \rightarrow \mathbb{R}$

More details in the corresponding paper[6] →

Domain-Adversarial Training for Flavour Tagging

Idea

- Representation of data by simulation known to be limited (track multiplicities, kinematic spectrum, particle identification responses and more)
- Training of FT algorithms usually relies on labelled samples from simulation
- Higher tagging power of (inclusive) tagging algorithms in simulation wrt. data
- Domain-adversarial training[6] to use domain (data or simulation) invariant feature set for flavour prediction

First Toy-based Studies

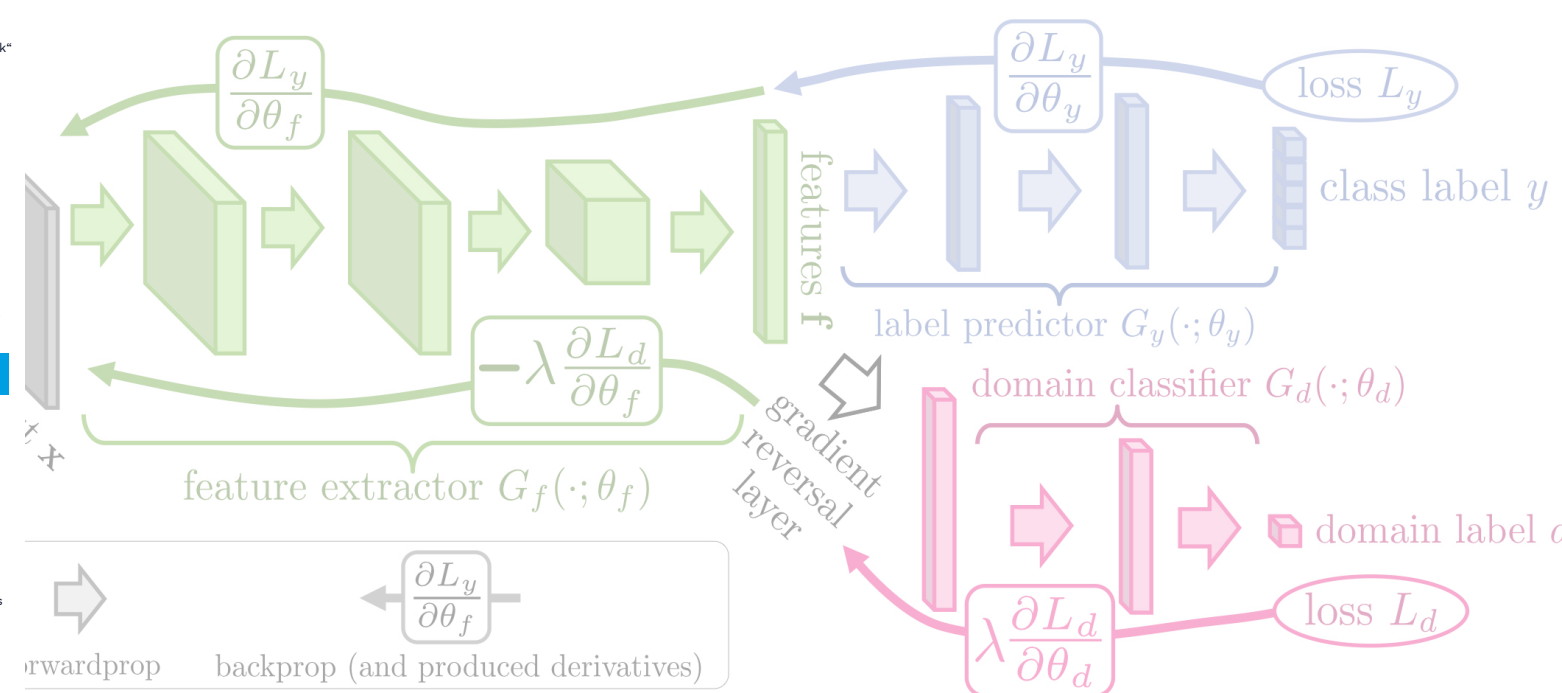
- Generating samples to represent different domains:
 - „Simulation-like“ with smaller multiplicity and harder momenta; labelled flavour for training
 - „Data-like“ with higher multiplicity and softer momenta; flavour masked for training
- Construct DeepSet model with two separate NNs after pooling for domain and flavour prediction
- First check: predict only the domain → achieved accuracy of $\approx 80\%$ → domains are distinguishable
- Baseline, train flavour tagger with domain classifier as silent observer ($\lambda = 0$):
 - Domain prediction: 57% accuracy → features relevant for tagging show mild domain dependency
 - Tagging power of $713 \pm 0.12\%$ (4.95 ± 0.13%) in the simulation(data)-like sample
- Domain-adversarial training ($\lambda > 0$):
 - Domain accuracy dropped to 50% → extracted feature seemingly domain invariant
 - Tagging power of $716 \pm 0.12\%$ (5.09 ± 0.13%) in the simulation(data)-like sample → no significant change
- Probing the domain invariance: separate training of domain classifier based on extracted feature set yields accuracy of 68%

Preliminary Conclusion and Outlook

- Observed domain invariance not yet robust
- Challenges from concurrent optimisation of losses
- Investigating training with multiple phases or optimisation of scale factors to handle losses
- Studying semi-supervised approach for opposite side by making use of charged B-meson decays

Implementing domain-adversarial training

Different challenges encountered



[Ganin et.al., Domain-Adversarial Training of Neural Networks, arXiv:1505.07818]