ML/AI ACTIVITIES IN ATLAS

Glance at Recent Developments

Geoffrey Gilles
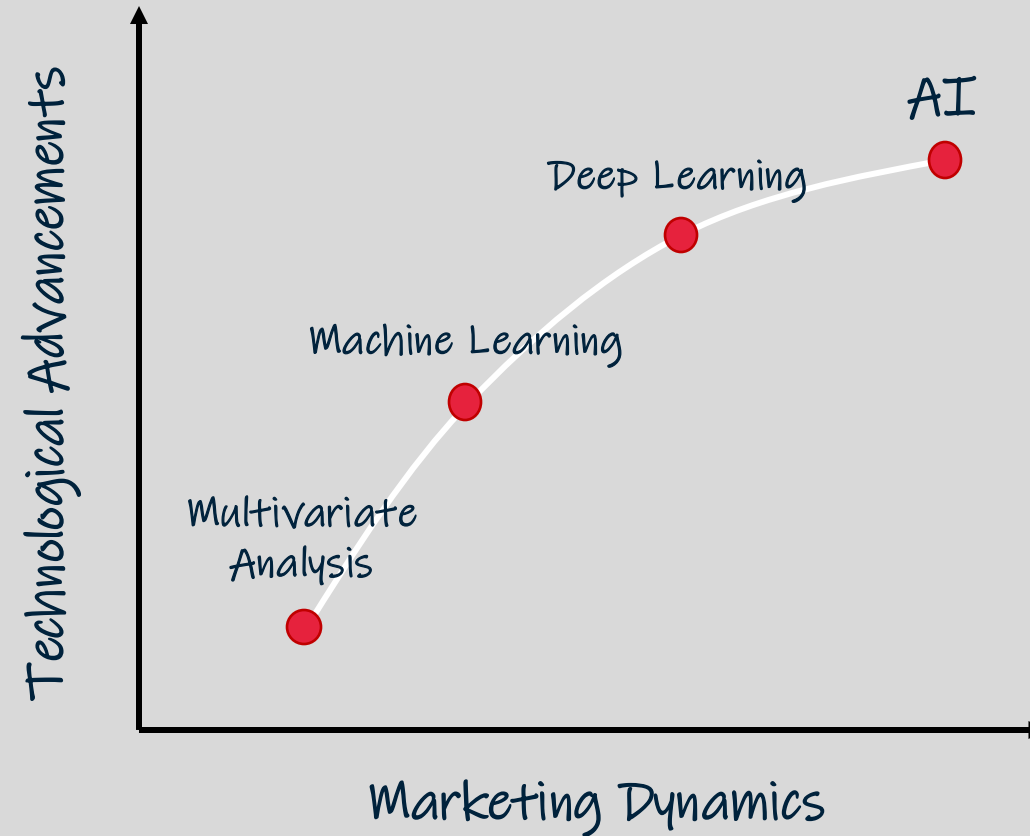
NIKHEF - Dutch National Institute for Subatomic Physics

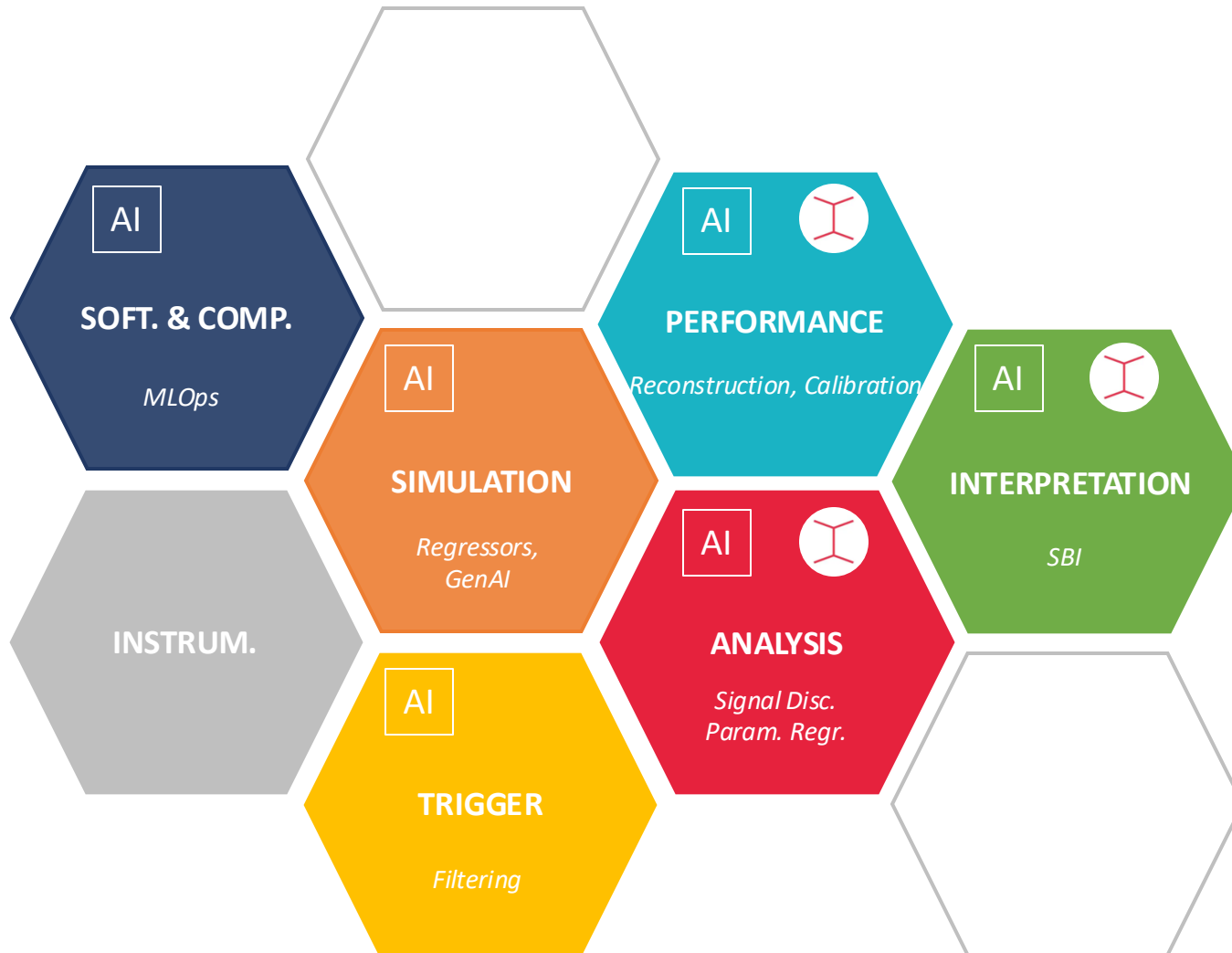*Thanks to Zef for the inputs*

The AI Hype Arc

INTRODUCTION

AI in ATLAS - NIKHEF's Group Activity

AI IN ATLAS

SOFT. & COMP.

AI

MLOps

SIMULATION

AI

Regressors, GenAI

INSTRUM.

PERFORMANCE

AI

Reconstruction, Calibration

INTERPRETATION

AI

SBI

**ANALYSIS**

AI

Signal Disc. Param. Regr.

TRIGGER

AI

Filtering

- Present in almost all sectors
- Collaboration forum supporting developments & tools
- Nikhef's AI activities focused on
  - Combined performance
  - Physics analysis
  - Interpretation

Nik|hef

STEADY PROGRESS

e.g. Flavour Tagging

1st Transformer-based tagger

1st Deep Feed-Forward NN-based tagger

1st GNN-based tagger
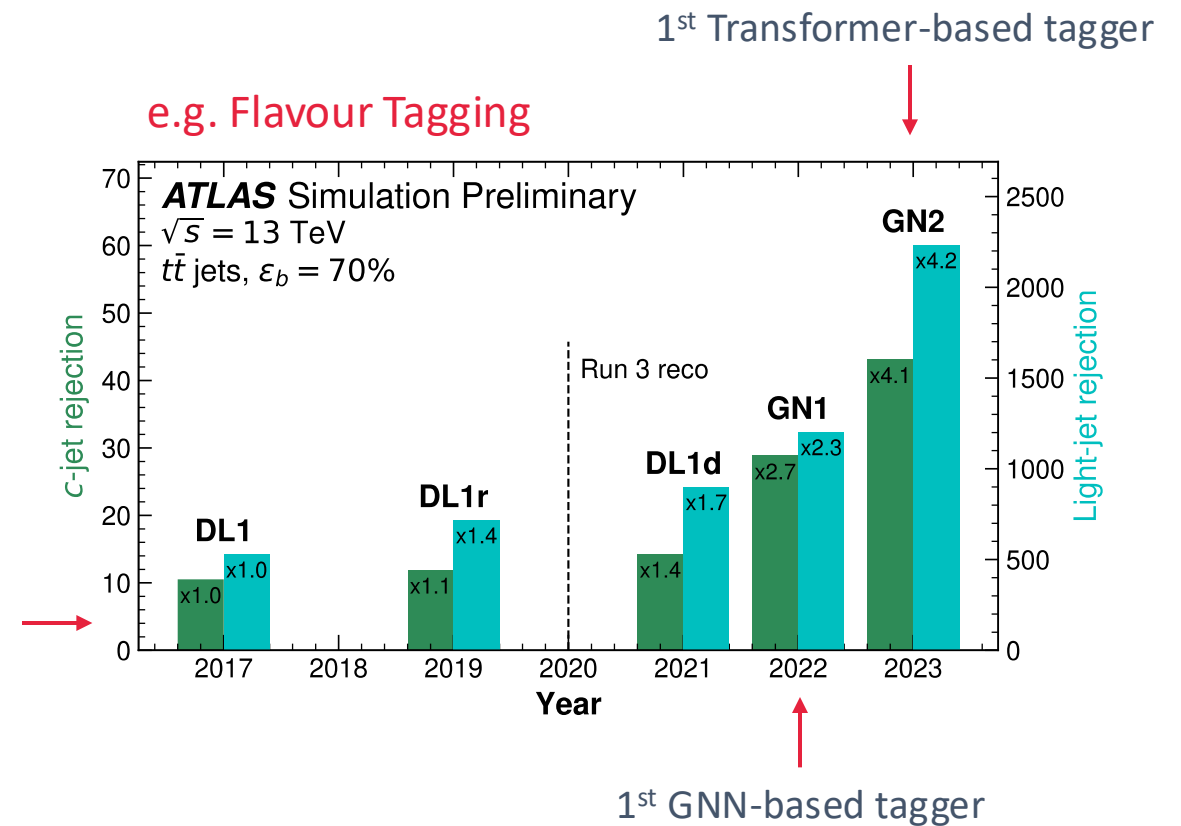
- ATLAS builds on pioneering AI history to achieve groundbreaking physics results
- Deep Learning advances unlocked new applications and pushed performance boundaries
- Cutting-edge and classical AI techniques still coexists, reflecting diverse group expertise
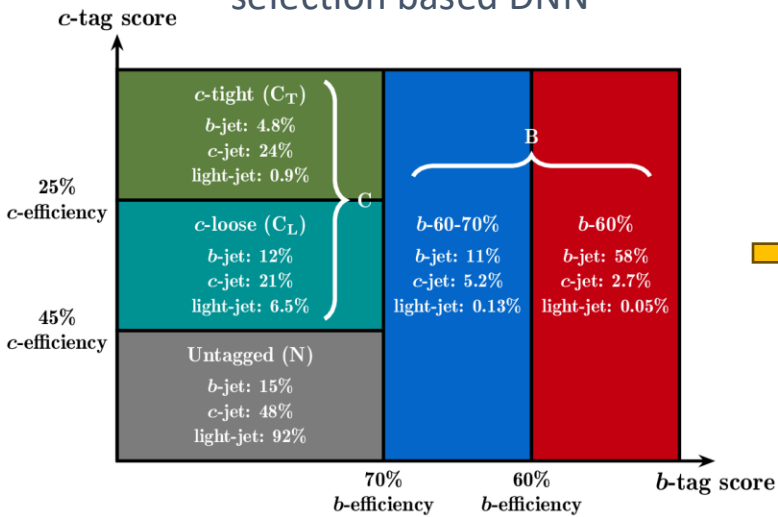
GLANCE AT RECENT DEVELOPMENTS

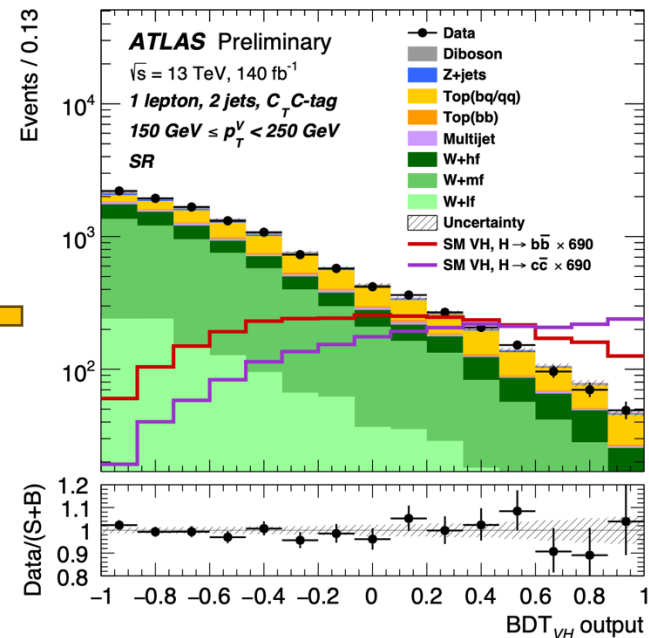Combined Performance, Physics Analysis & Interpretation

DNN+BDT TO DELIVER TOP-TIER RESULTS

- Recent joint $VH(b\bar{b}/c\bar{c})$ measurement, probing 2nd gen. fermion Higgs couplings among others
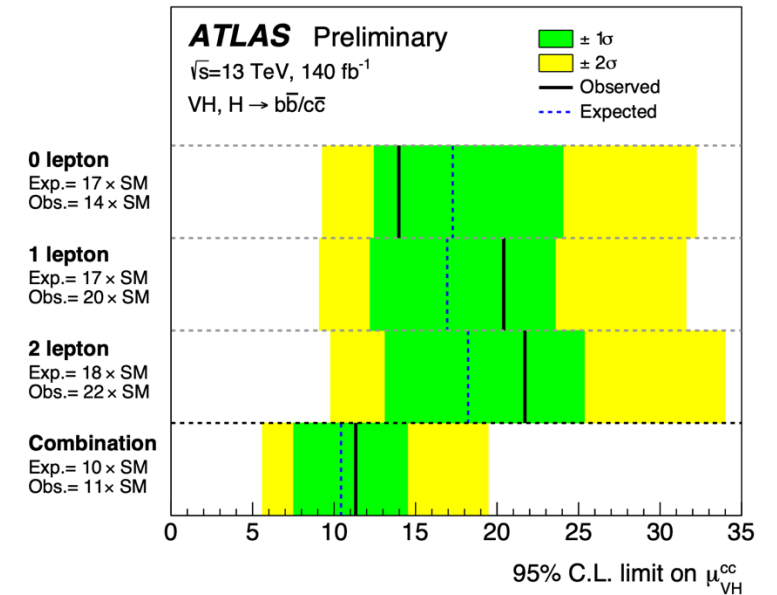- Complex analysis, fitting ~50 SR and ~100 CR defined by tagging and kinematic requirements

Orthogonal b- and c-tagging selection based DNN

Signal vs. bkg discrimination using BDT

Strongest observed limit to date

THE TRANSFORMER EXPLOSION

- Transformer introduced by a team at Google Brain in 2017 [link]
  - Emerged as a dominant paradigm in Machine Learning across various applications
  - Ability to model complex relationships and deliver outstanding performance

- Powerful technique across various ML applications
  - Based on "Attention Mechanism" (see backup)
  - Excel at capturing contextual relationships
  - High capacity for learning intricate features
  - Parallelizability & Reduced computing resources
  - Scalability

- ATLAS Nikhef particularly focused on Transformer applications
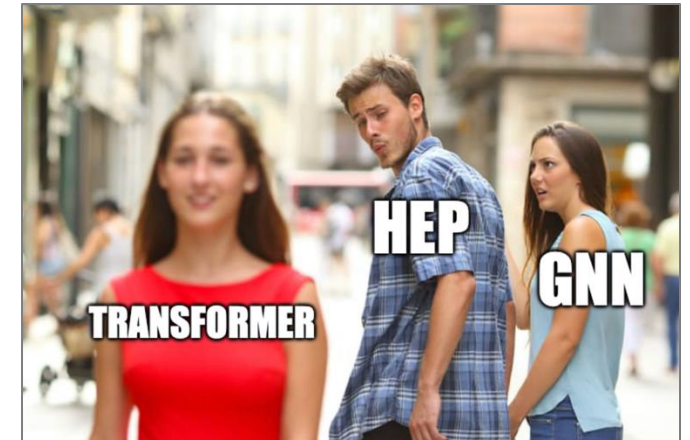
# APPLICATION TO BOOSTED HIGGS-BOSON TAGGING

- **High-performance tagging of boosted Higgs-boson decaying to Heavy Flavour crucial in**
  - Crucial in improving sensitivity to New Physics
  - Precise measurement Higgs properties

- **New cutting-edge algorithm based on Transformer**
  - Learning for sequence of particle tracks within Large-R jet proxying Higgs-boson candidate
  - Remarkable performance improvements
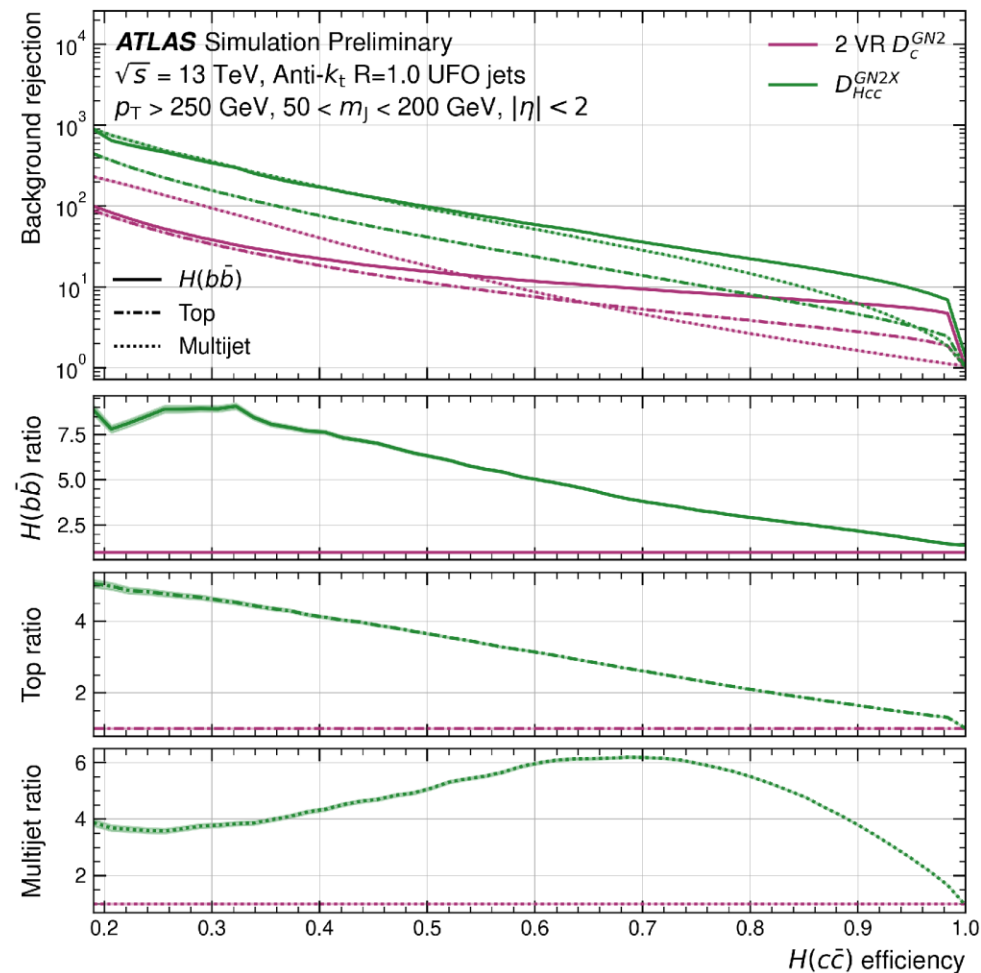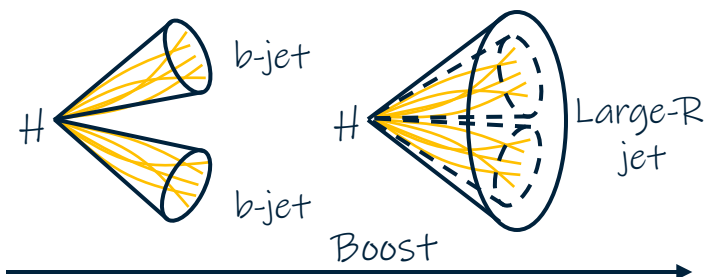
SMARTER TRANSFORMER FOR 4-TOP SEARCH

- 4-top production is the rarest top production and heaviest final state ever observed

- First observation in 2023, possible source of New Physics to now be precisely measured

- Introducing pair-wise features & SM interaction matrix into Particle Transformer architecture
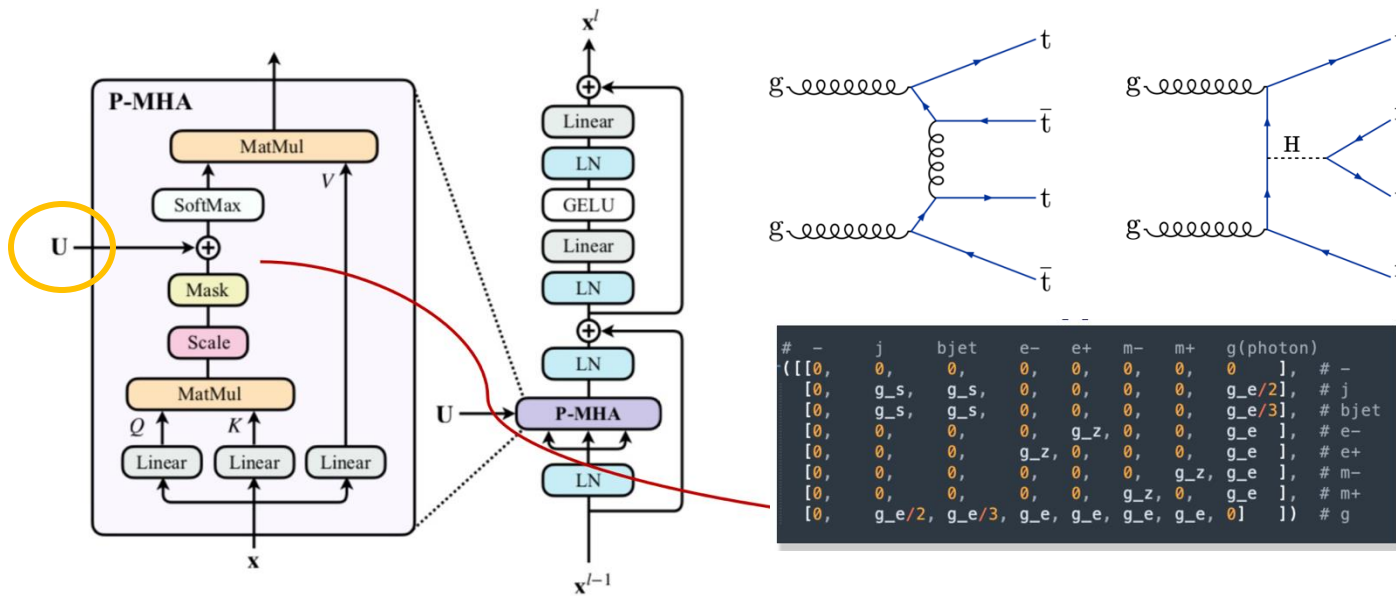
Improved signal efficiency and background regression

More details: arXiv:2211.05143

Polina, Clara, Sascha
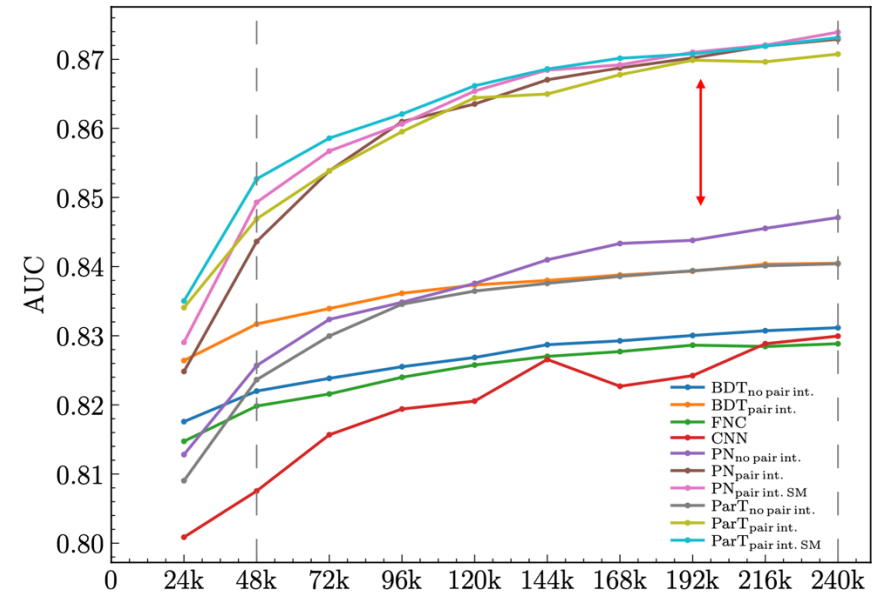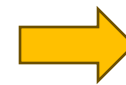
EXPLORING SBI FOR SMEFT INTERPRETATION

- Discriminate SM from EFT and bkg using Transformer learning from entire event information

- Allow to construct likelihood functions that capture intricacies of SMEFT interactions and detector responses in unprecedented details

- Prospective study exploiting $gg \rightarrow ZH$ events shows $c_{\varphi t}$ coefficient to be constrained to [-5,10]

TRANSFORMER

SBI

Geoffrey, Karel, Marcel, Wouter

Nikhef
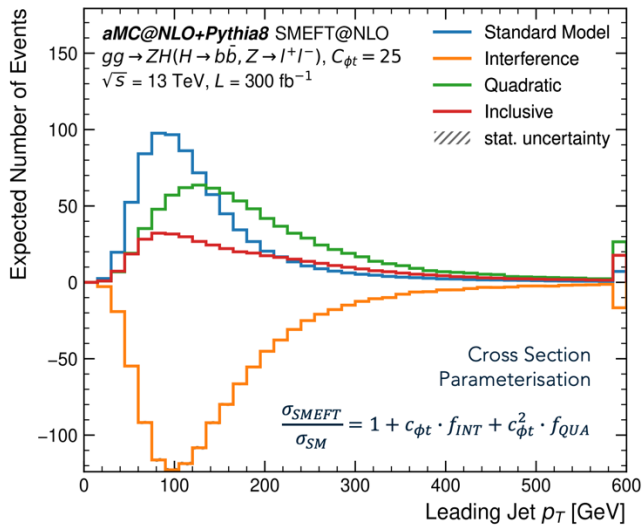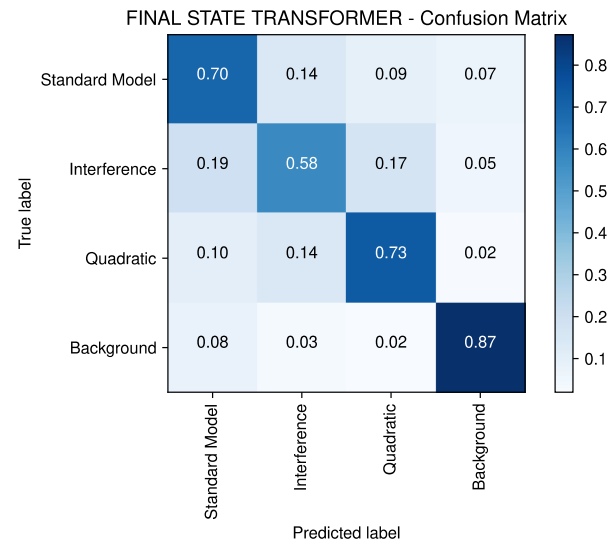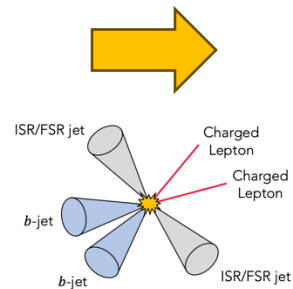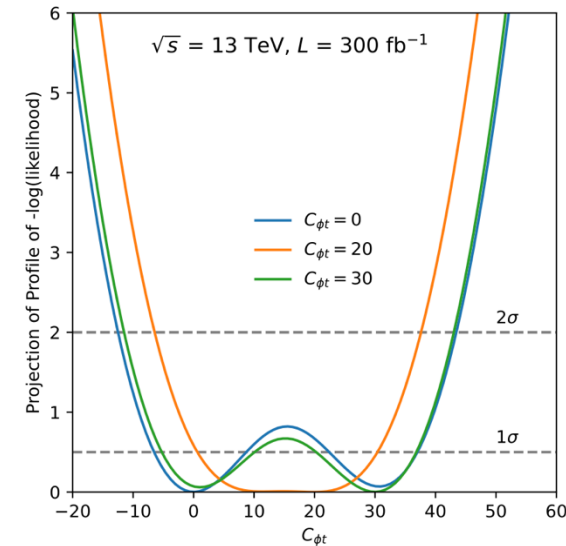
TRANSFORMER-BASED PARTICLE TRACKING

- Tracking becomes increasingly challenging in the context of HL-LHC upgrade

- "TrackFormers" project investigated Transformer architectures to perform hit association prediction, inspired by LLM

- Best models, achieving sub-second full event reconstruction on public Kaggle dataset

Example of Encoder Regressor Design

Zef, Uraz, Slav, Sascha
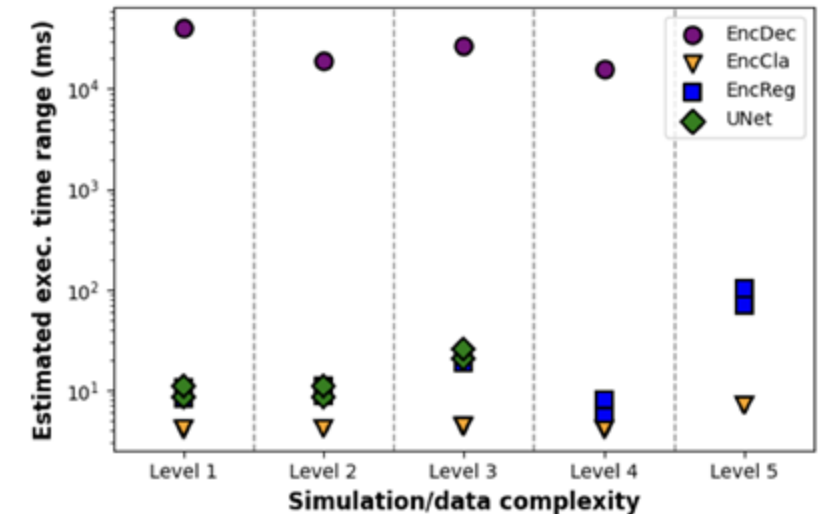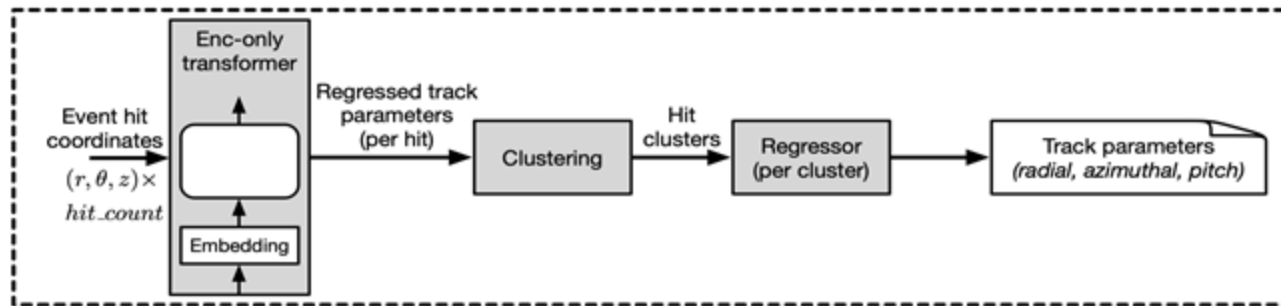
Nikhef

CONCLUSIONS & OUTLOOKS

CONCLUSIONS & OUTLOOKS

- ATLAS Nikhef's active role in AI developments
  - Strong involvement in AI advancements for enhanced performance and physics analysis
  - Contributing to high-impact, top-tier LHC results
  - Training next generation in cutting-edge AI techniques

- Emphasis on Transformer networks in recent research
  - Driving numerous innovative
  - Achieving outstanding performance in various tasks

- Enthusiastically supporting the Nikhef AI initiative
  - Committed to sharing knowledge and experience, particularly in Soft.& Comp. challenges

FINAL STATE TRANSFORMER

- ML development toolkit built upon Transformer encoders, tailored for HEP
- Leverage multi-head mechanism attention to improve classification and regression tasks
- User-friendly interface for integrating Transformers into research workflows

Input Layer   Embedding Layer   Three Multi-Head Attention Layers   Flatten Layer   Dense Layers   Output Layer
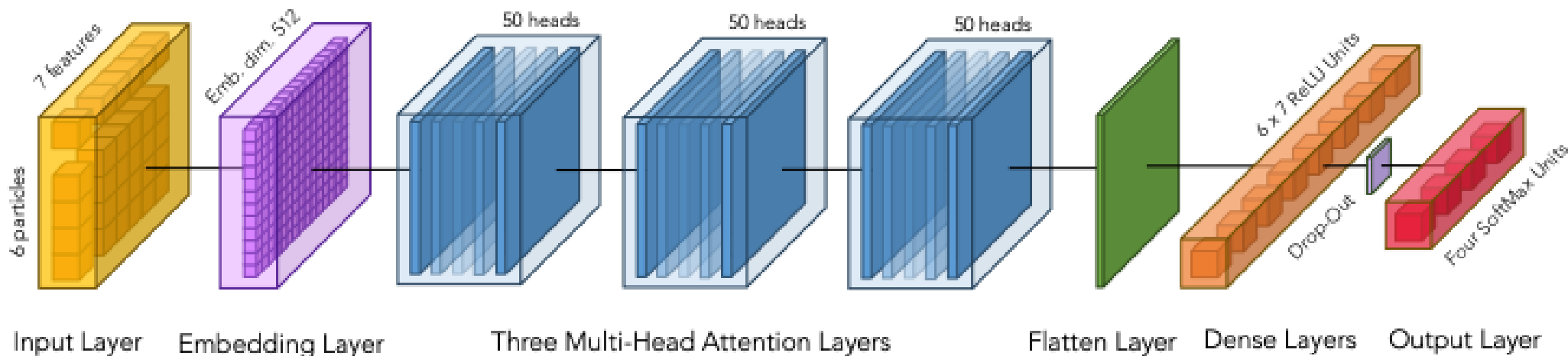
BACK UP

THE SELF-ATTENTION MECHANISM

IN A NUTSHELL
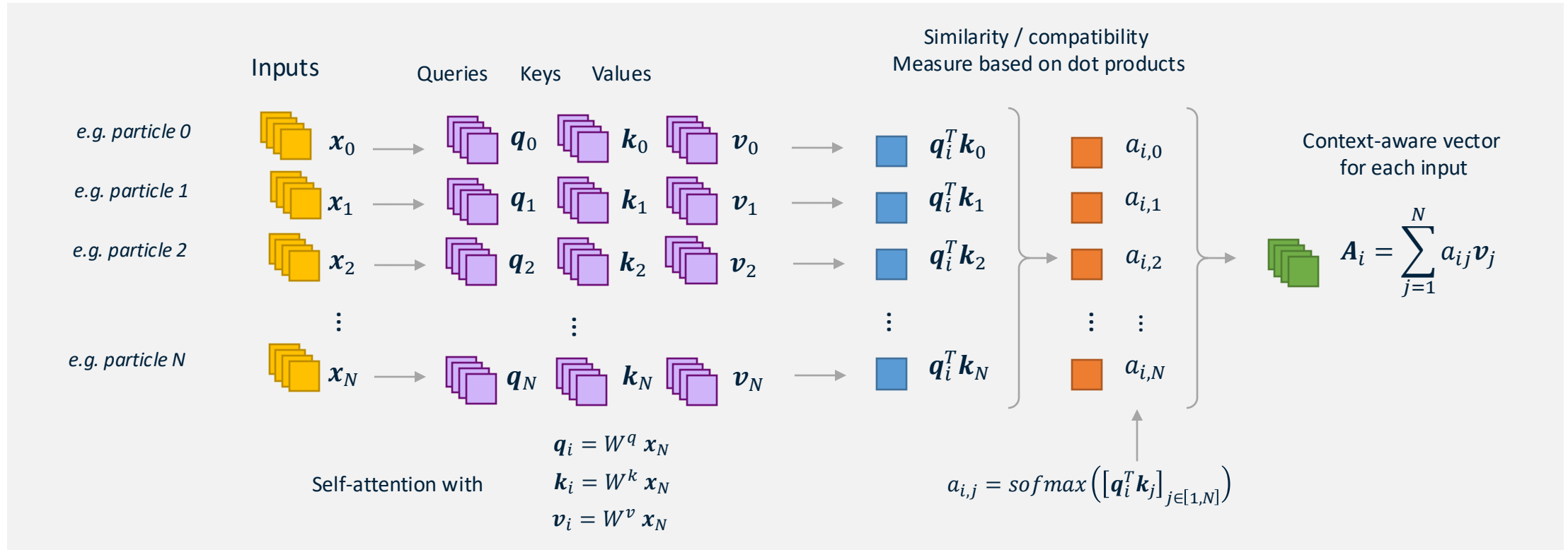
Inputs   Queries   Keys   Values

Similarity / compatibility
Measure based on dot products

e.g. particle 0 x_0 q_0 k_0 v_0 $q_i^T k_0$ $a_{i,0}$

e.g. particle 1 x_1 q_1 k_1 v_1 $q_i^T k_1$ $a_{i,1}$

e.g. particle 2 x_2 q_2 k_2 v_2 $q_i^T k_2$ $a_{i,2}$

e.g. particle N x_N q_N k_N v_N $q_i^T k_N$ $a_{i,N}$

Context-aware vector for each input

$$A_i = \sum_{j=1}^{N} a_{ij} v_j$$

Self-attention with
$$q_i = W^q x_N$$
$$k_i = W^k x_N$$
$$v_i = W^v x_N$$

$$a_{i,j} = sofmax\left(\left[q_i^T k_j\right]_{j\in[1,N]}\right)$$

NB: Self attention can be seen as an example of message passing on a fully connected graph

Nik|hef

THE SELF-ATTE... ...SM

IN A NUTSHELL

MULTI-HEAD ATTENTION
PRINCIPLE

Inputs

Que...

...y / compatibility
...sed on dot products
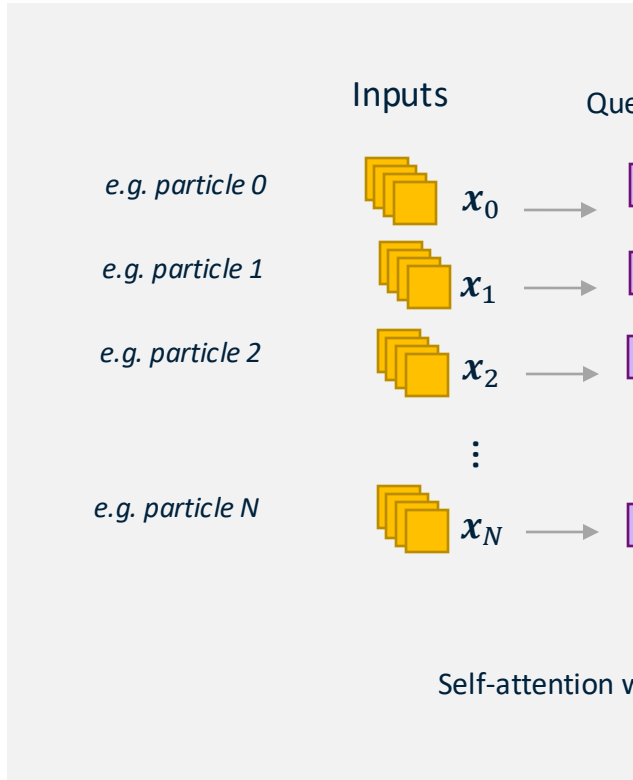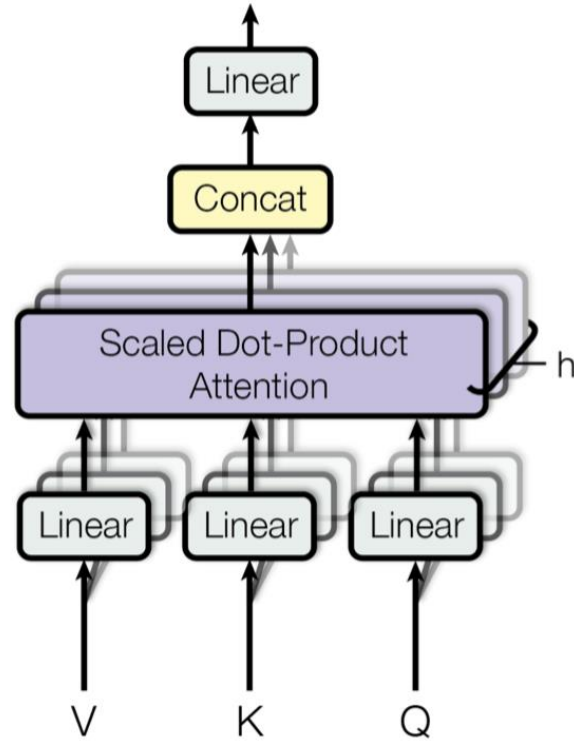
e.g. particle 0 x_0

e.g. particle 1 x_1

e.g. particle 2 x_2

x_N

e.g. particle N

$a_{i,0}$

$a_{i,1}$

$a_{i,2}$

$a_{i,N}$

Context-aware vector
for each input

$$A_i = \sum_{j=1}^{N} a_{ij} v_j$$

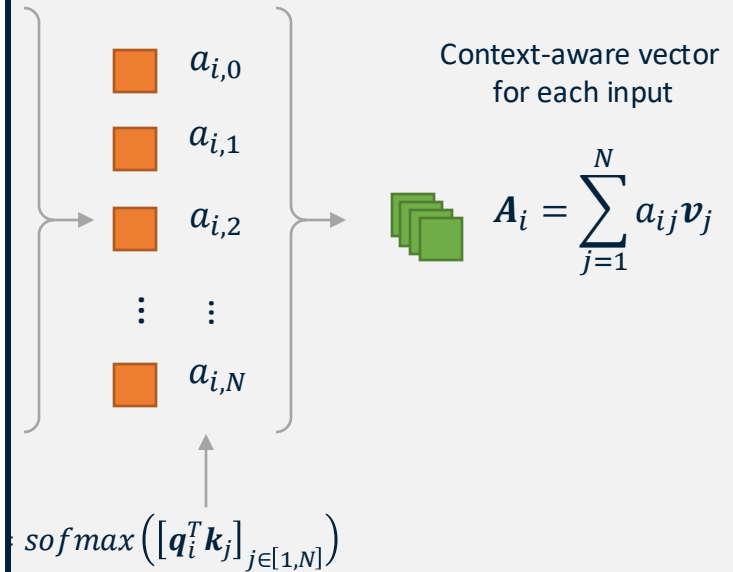Self-attention ...

... $softmax\left(\left[q_i^T k_j\right]_{j \in [1,N]}\right)$

Using scale dot-product attention

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

NB: Self attention can be seen as an example of message passing on a fully connected graph

APPLICATION TO BOOSTED HIGGS-BOSON TAGGING

Large-R Jet classification as primary task

Seq. of 100 tracks / jet (~20 kinematics) and jet kinematics (p_T, η, mass)

Transformer
Encoder

Pooled graph representation

Jet flavour prediction

Track Initialiser

Intial track representation

Combined Inputs

Conditional track representation

Track origin predictions

Auxiliary tasks

Dense layers projecting inputs into embedding representation

6 multi-head attention layers (with 4 attention heads)
Using scale dot-product attention

Vertex predictions

FD FW Network

Feed Forward Network