Antideutron Analysis

(Velocity Reconstruction with Deep Learning)

Hsin-Yi Chou

Academia Sinica

Oct 31st, 2025

RICH Velocity Reconstruction with Deep Learning (Old Version)

Inputs: hit information + others

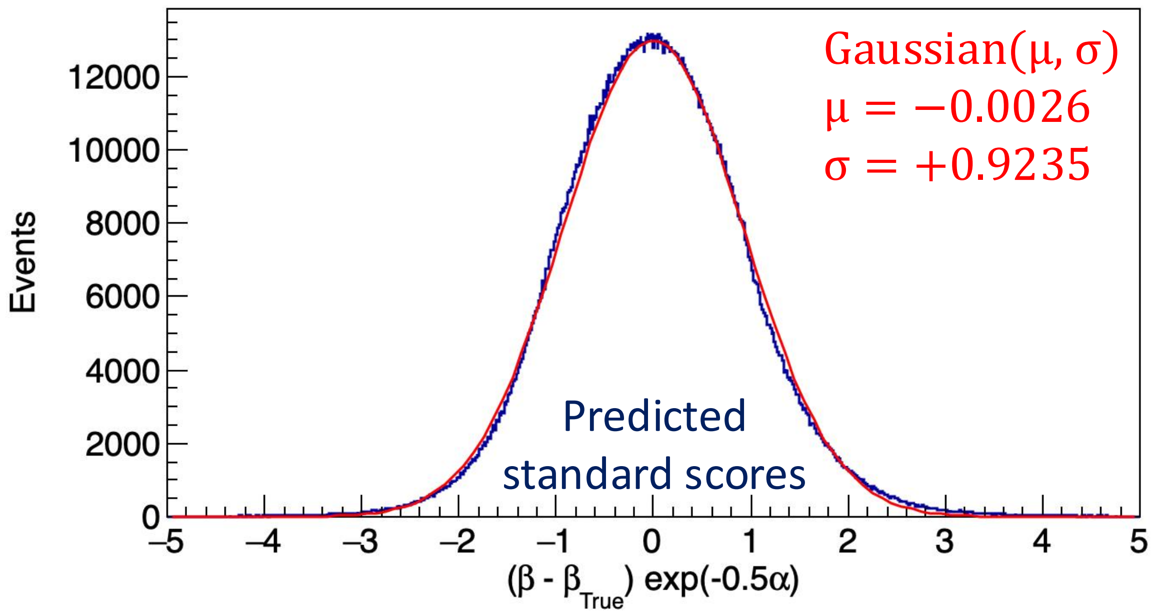
Outputs:

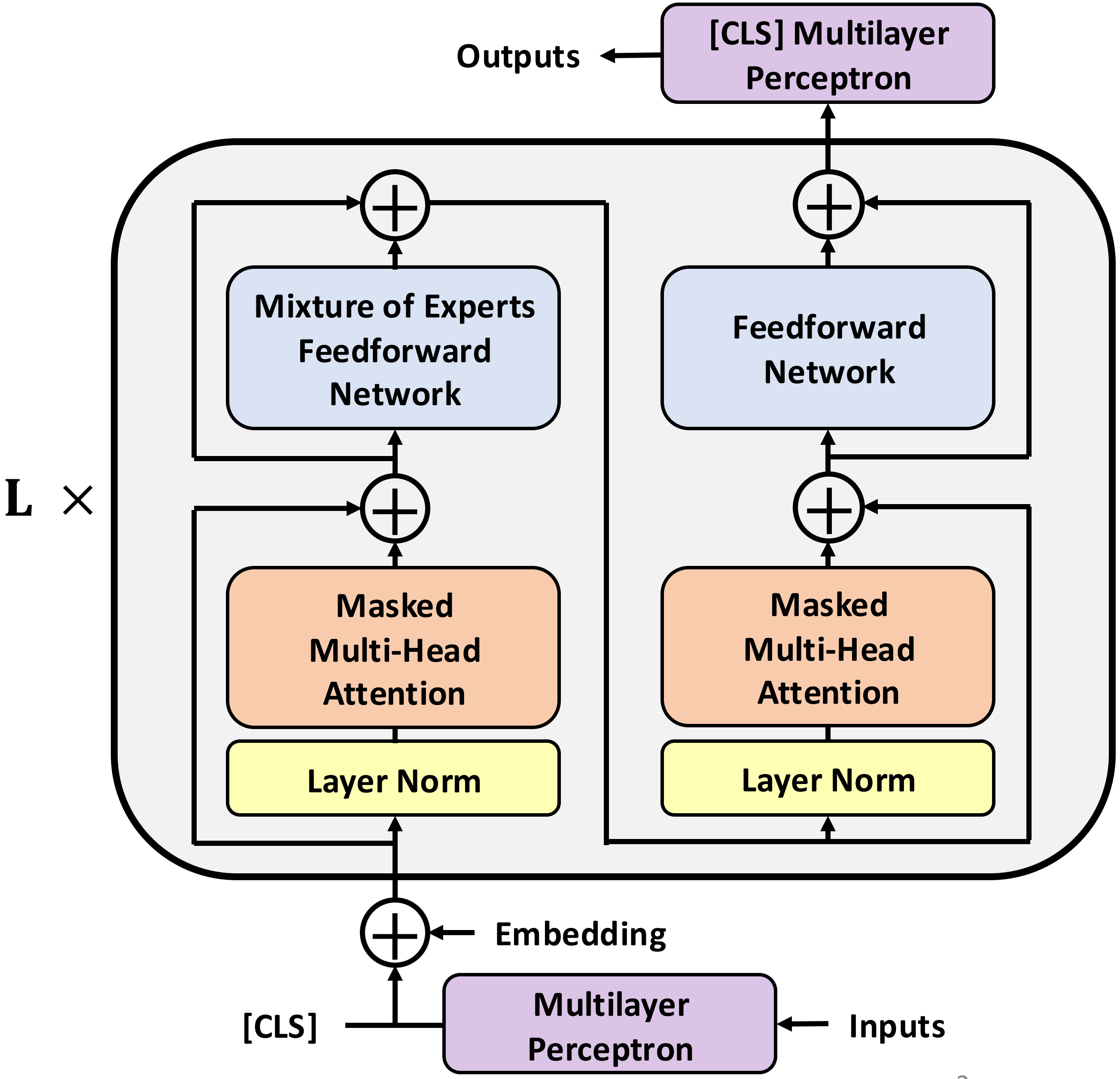
1) $\boldsymbol{\beta}$, predicted velocity

2) $\boldsymbol{\alpha} = \ln \boldsymbol{\sigma}^2$, predicted velocity log variance (uncertainty)

Heteroscedastic Loss (Gaussian Negative Log-Likelihood):

$$\mathcal{L} = \frac{1}{N}\sum\left[\left(\beta_i - \beta_{i,\mathrm{True}}\right)^2 e^{-\alpha_i} + \alpha_i\right]$$





Length of Sequence:

- Inputs N = 512
- Latents M = 8
- Outputs
  1) Gaussian Q = 2
  2) Student's t Q = 3

Number of Layers:

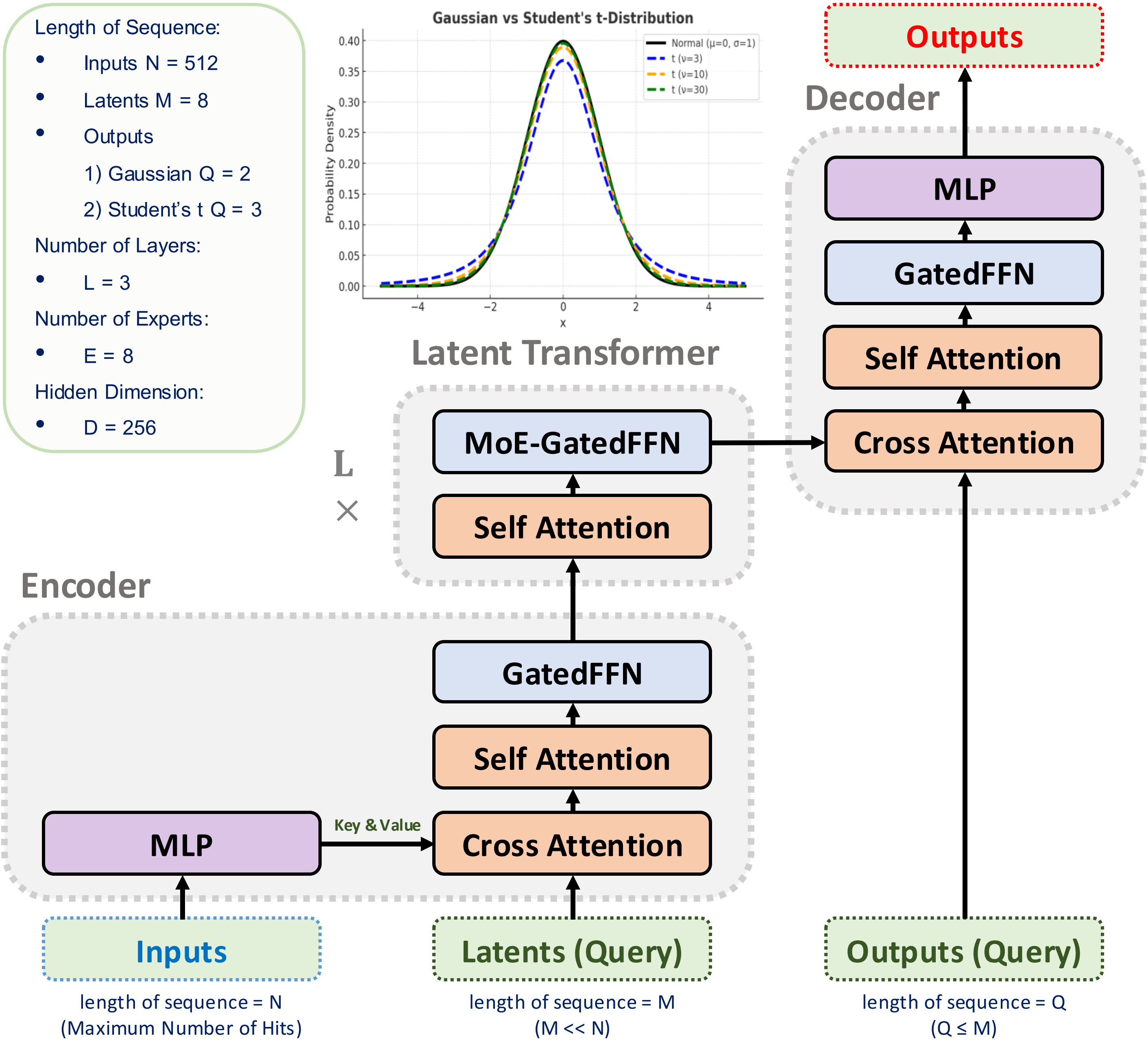
- L = 3

Number of Experts:

- E = 8

Hidden Dimension:

- D = 256



Physical Variable Reconstruction with Deep Learning

- **Inputs:**
  hit relation information

- **Outputs:**
  - <u>Gaussian Distribution</u>
    1) $\boldsymbol{\beta_G}$, predicted mean
    2) $\boldsymbol{\alpha_G} = \ln \sigma^2$, predicted log variance (uncertainty)
  - <u>Student's t Distribution</u>
    1) $\boldsymbol{\beta_T}$, predicted mean
    2) $\boldsymbol{\alpha_T} = \ln \sigma^2$, predicted log variance (uncertainty)
    3) $\boldsymbol{\gamma_T} = \ln \nu$, predicted tail heaviness (robustness to outliers)

- **Physics-informed Objective Function:**
  - <u>Gaussian NLL</u>

$$\mathcal{L}_G = -\frac{2}{N}\sum \ln \boldsymbol{p_{G,i}} = \frac{1}{N}\sum (\mathrm{C}_{G,i} + \mathrm{E}_{G,i})$$

$$\mathrm{C}_G = \ln 2\pi + \alpha_G$$
$$\mathrm{E}_G = (\beta_G - \beta_{\mathrm{True}})^2 e^{-\alpha_G}$$

$$\boldsymbol{p_G} = (2\pi \cdot e^{\alpha_G})^{-1/2} \cdot e^{(-1/2)\cdot(\beta_G - \beta_{\mathrm{True}})^2 e^{-\alpha_G}}$$

  - <u>Student's t NLL</u>

$$\mathcal{L}_T = -\frac{2}{N}\sum \ln \boldsymbol{p_{T,i}} = \frac{1}{N}\sum (\mathrm{C}_{T,i} + \mathrm{E}_{T,i})$$

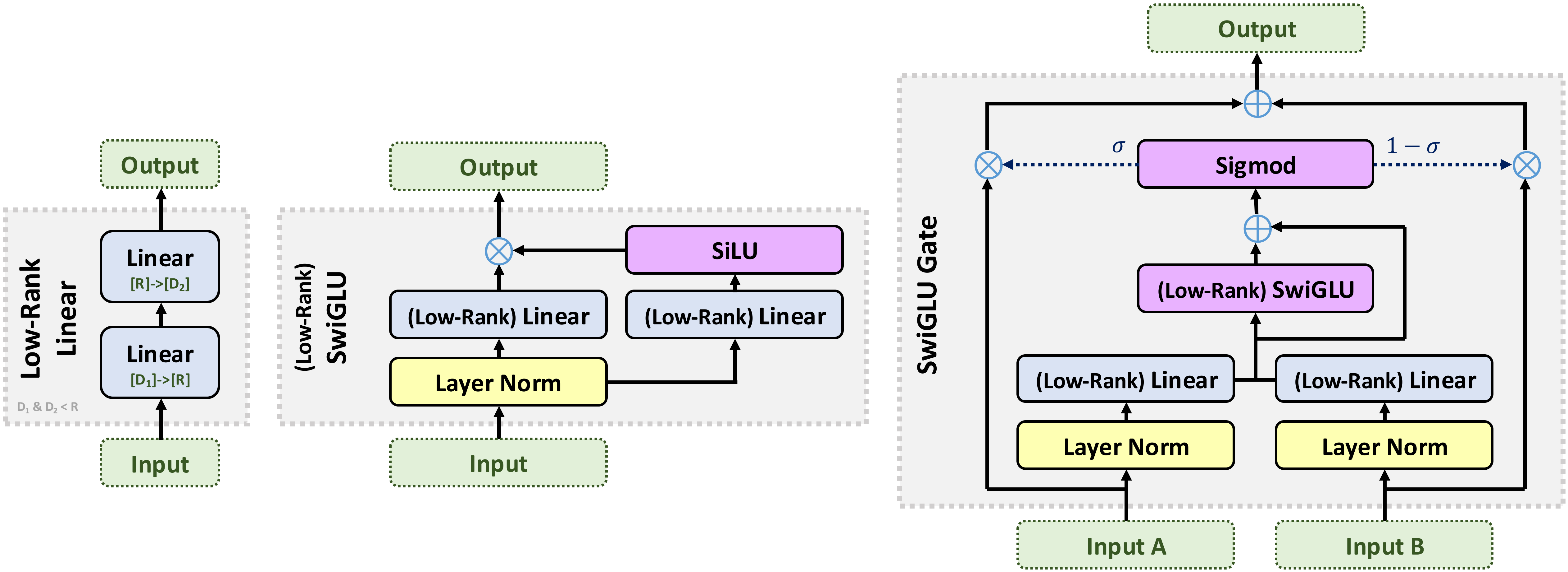
$$\mathrm{C}_T = 2 \cdot [\ln \Gamma(e^{\gamma_T}/2) - \ln \Gamma((1+e^{\gamma_T})/2)] + (\ln \pi + \alpha_T + \gamma_T)$$
$$\mathrm{E}_T = (1 + e^{\gamma_T}) \cdot \ln[1 + (\beta_T - \beta_{\mathrm{True}})^2 e^{-(\alpha_T+\gamma_T)}]$$

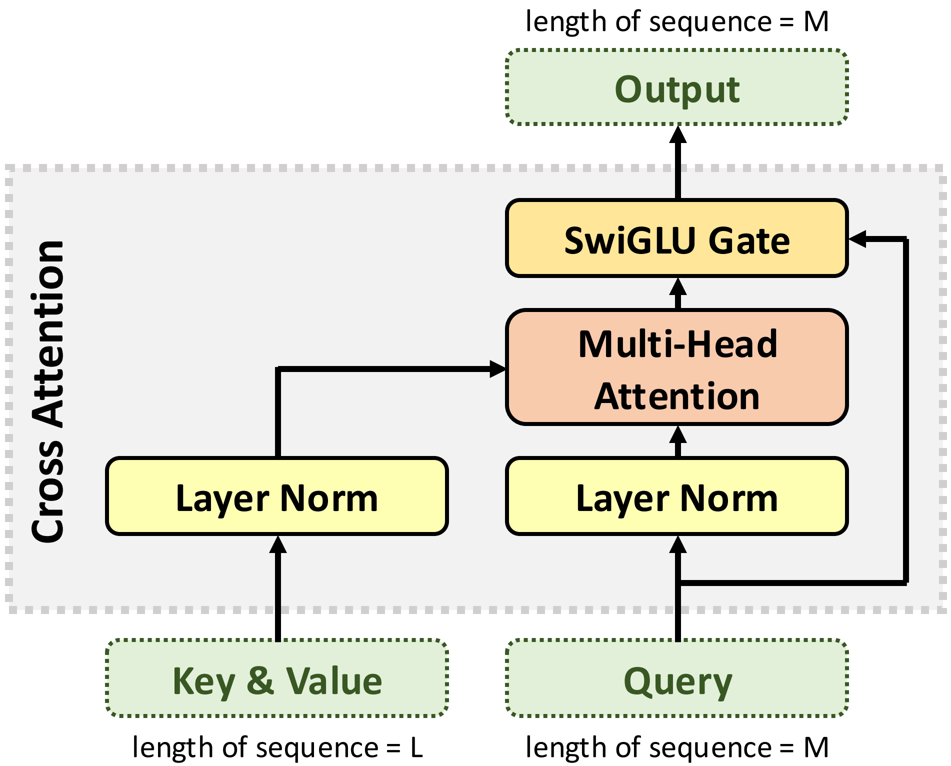
$$\boldsymbol{p_T} = \frac{\Gamma((1+e^{\gamma_T})/2)}{\Gamma(e^{\gamma_T}/2)} (\pi \cdot e^{\alpha_T+\gamma_T})^{-1/2} [1 + (\beta_T - \beta_{\mathrm{True}})^2 e^{-(\alpha_T+\gamma_T)}]^{-(1+e^{\gamma_T})/2}$$

p_T follows a Gaussian distribution when $\gamma_T \to \infty$

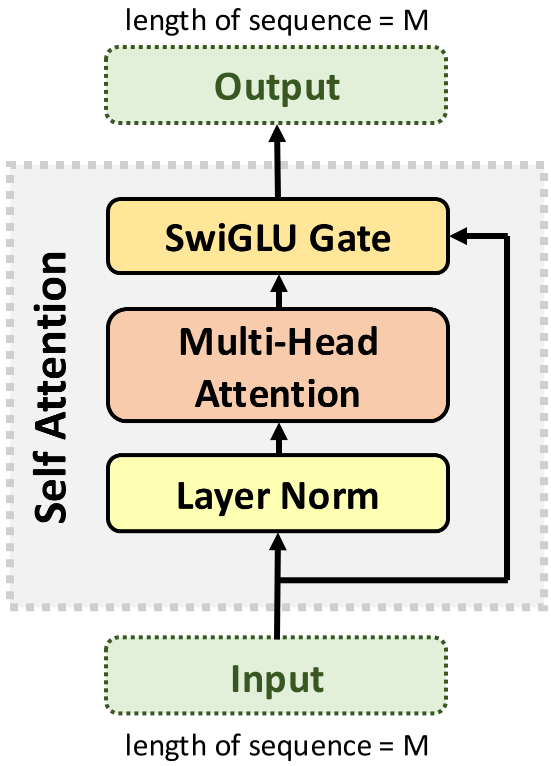
$$\text{Distribution Variance} = \begin{cases} e^{\alpha_T+\gamma_T}/(e^{\gamma_T}-2) & ,\gamma_T > \ln 2 \\ \infty & ,0 < \gamma_T \le \ln 2 \\ \text{undefined} & ,\gamma_T \le 0 \end{cases}$$

Inference rate = ~18,000 events/seconds on RTX 3090 => 1.5 x 10^9 events/days

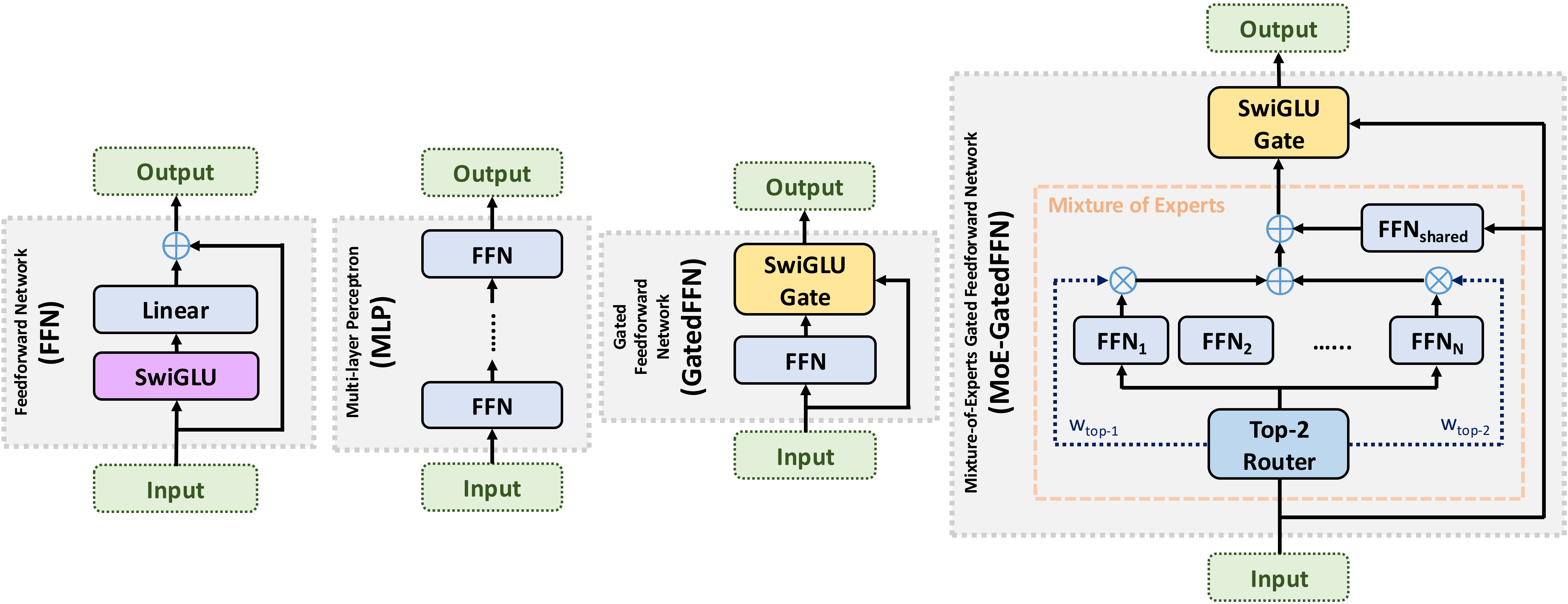

Output
Low-Rank
Linear
Linear
[R]->[D2]
Linear
[D1]->[R]
D1 & D2 < R
Input
Output
(Low-Rank)
SwiGLU
SiLU
(Low-Rank) Linear
(Low-Rank) Linear
Layer Norm
Input
Output
SwiGLU Gate
σ
Sigmod
1 − σ
(Low-Rank) SwiGLU
(Low-Rank) Linear
(Low-Rank) Linear
Layer Norm
Layer Norm
Input A
Input B

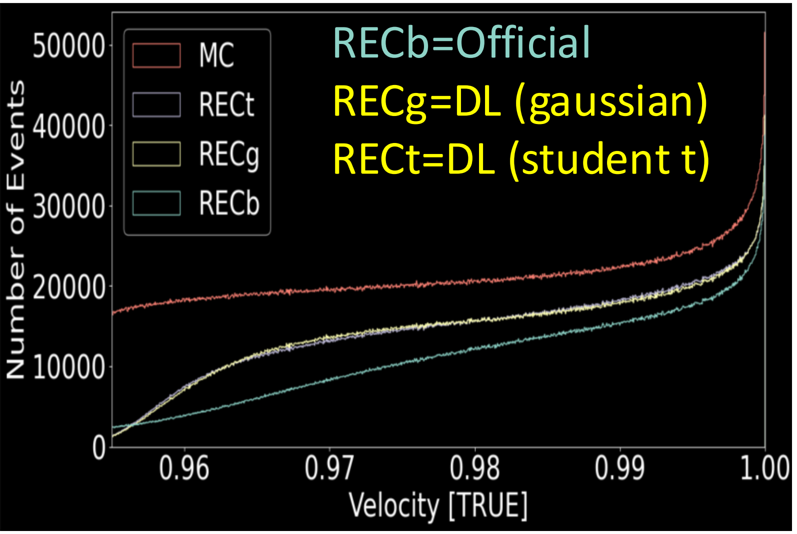


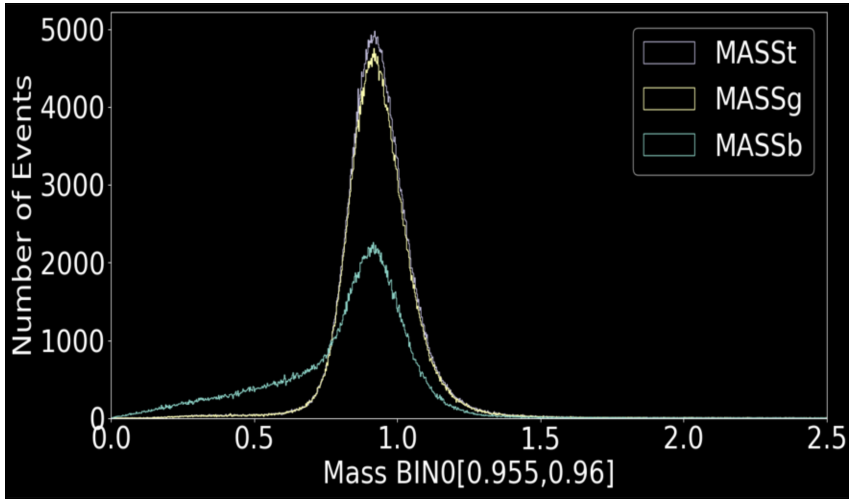
length of sequence = M
Output
Cross Attention
SwiGLU Gate
Multi-Head Attention
Layer Norm
Layer Norm
Key & Value
Query
length of sequence = L
length of sequence = M

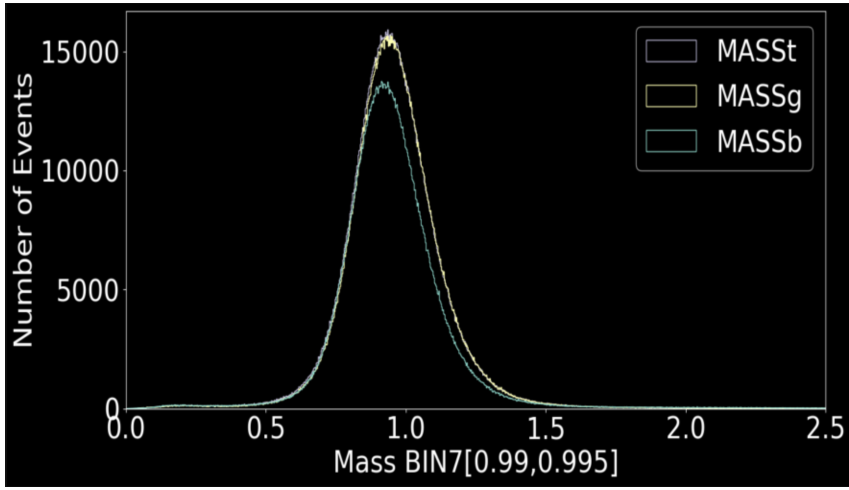


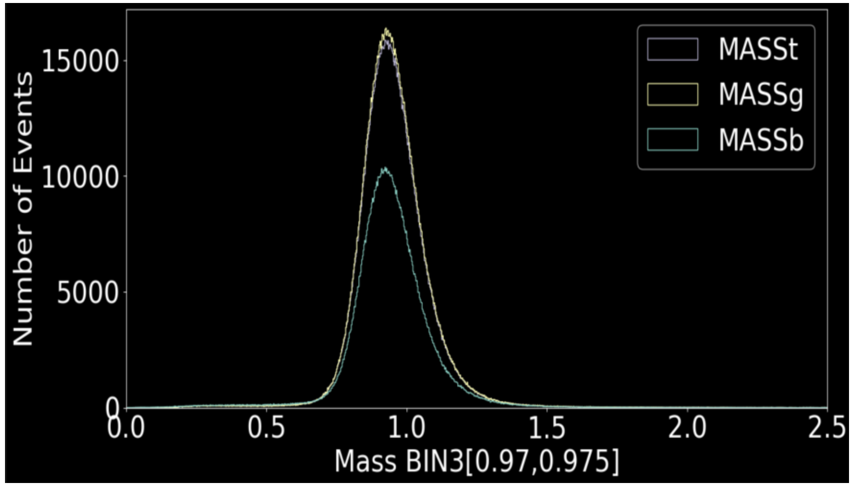
length of sequence = M
Output
Self Attention
SwiGLU Gate
Multi-Head Attention
Layer Norm
Input
length of sequence = M

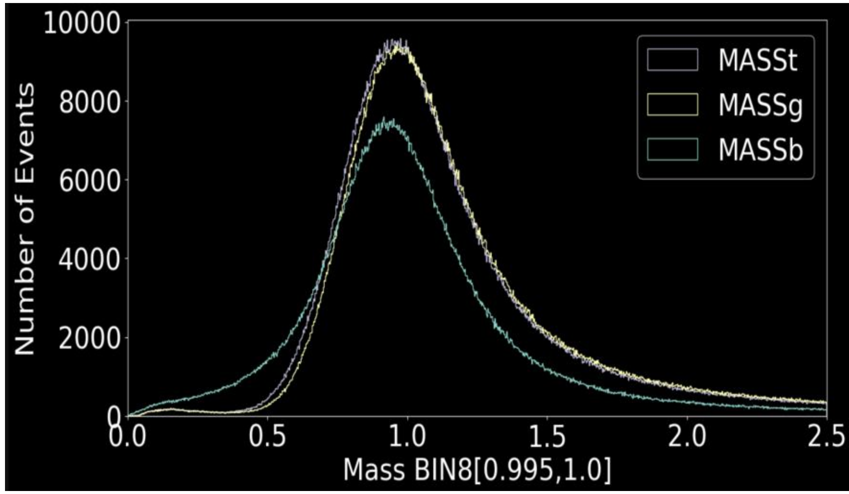

Output
Linear
SwiGLU
Input
Feedforward Network
(FFN)
Output
FFN
......
FFN
Input
Multi-layer Perceptron
(MLP)
Output
SwiGLU Gate
FFN
Input
Gated Feedforward Network
(GatedFFN)
Output
SwiGLU Gate
Mixture of Experts
FFN_{shared}
FFN_1
FFN_2
......
FFN_N
w_{top-1}
w_{top-2}
Top-2 Router
Input
Mixture-of-Experts Gated Feedforward Network
(MoE-GatedFFN)


RECb=Official
RECg=DL (gaussian)
RECt=DL (student t)
MC
RECt
RECg
RECb
Number of Events
Velocity [TRUE]
0
10000
20000
30000
40000
50000
0.96
0.97
0.98
0.99
1.00


MASSt
MASSg
MASSb
Number of Events
Mass BIN0[0.955,0.96]
0
1000
2000
3000
4000
5000
0.0
0.5
1.0
1.5
2.0
2.5


MASSt
MASSg
MASSb
Number of Events
Mass BIN7[0.99,0.995]
0
5000
10000
15000
0.0
0.5
1.0
1.5
2.0
2.5


MASSt
MASSg
MASSb
Number of Events
Mass BIN3[0.97,0.975]
0
5000
10000
15000
0.0
0.5
1.0
1.5
2.0
2.5


MASSt
MASSg
MASSb
Number of Events
Mass BIN8[0.995,1.0]
0
2000
4000
6000
8000
10000
0.0
0.5
1.0
1.5
2.0
2.5