SYMBA – **Sy**m**b**olic Computation of Squared **A**mplitudes
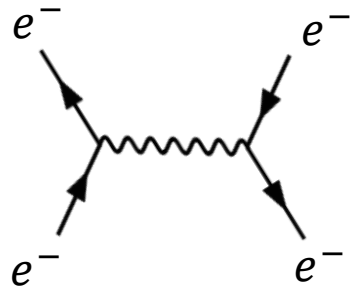
MARCO KNIPFER

UNIVERSITY OF ALABAMA

GOOGLE SUMMER OF CODE 2022

Introduction

- Feynman diagram

- Amplitude $\mathcal{M}$

- squared amplitude $|\mathcal{M}|^2$

- Cross section $\sigma(1 + 2 \rightarrow 1' + 2' + \cdots + n')$

```
M=1/2*i*e^2*gamma_{+%\sigma_73,%gam_55
,%eta_12}*gamma_{-%\sigma_73,%gam_56,%d
el_50}*e_{i_3,%eta_12}*e_{i_3+%n1}
del_50}(p_2)-u*e_{i_3,%gam_55}(p_3)_u^
(*)*e_{i_5,%gam_56}(p_4)_u^(*)+(m_e^2
+ -s_13 + 1/2*reg_prop)
```

$$
\mathcal{M}^2 = 2*e^4*(m_e^4 + \\
- 1/2*m_e^2*s_13 + 1/2*s_14*s_23 \\
+ -1/2*m_e^2*s_24 \\
- 1/2*s_12*s_34)*(m_e^2 \mp -s_13 \\
+ 1/2*reg\_prop)^{\wedge}(-2)
$$

$$\sum_{\sigma_1}(\not{p} u_{1\delta} \bar{u}_{1\alpha}) = (\not{p}_1 + m_1)_{\delta\alpha},$$

$$\sum_{\sigma_2} v_{2\beta} \bar{v}_{2\gamma} = (\not{p}_2 - m_2)_{\beta\gamma},$$

- Sum over polarizations
- Lorentz indices
- Lots of complicated identities needed
- Traces over γ-matrices
- Longest step on PC

Data Generation

- https://marty.in2p3.fr
- A **M**odern **AR**tificial **T**heoretical ph**y**sicist
- QED and QCD data, tree level, up to 3→3
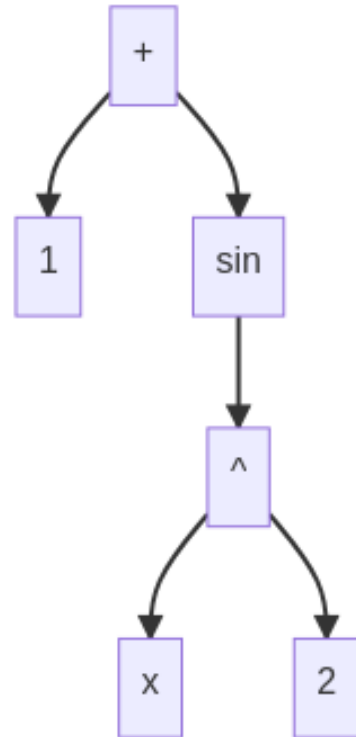
Expression encoding

- New in this project: (hybrid) prefix notation for expressions

- Motivated by

DEEP LEARNING FOR SYMBOLIC MATHEMATICS
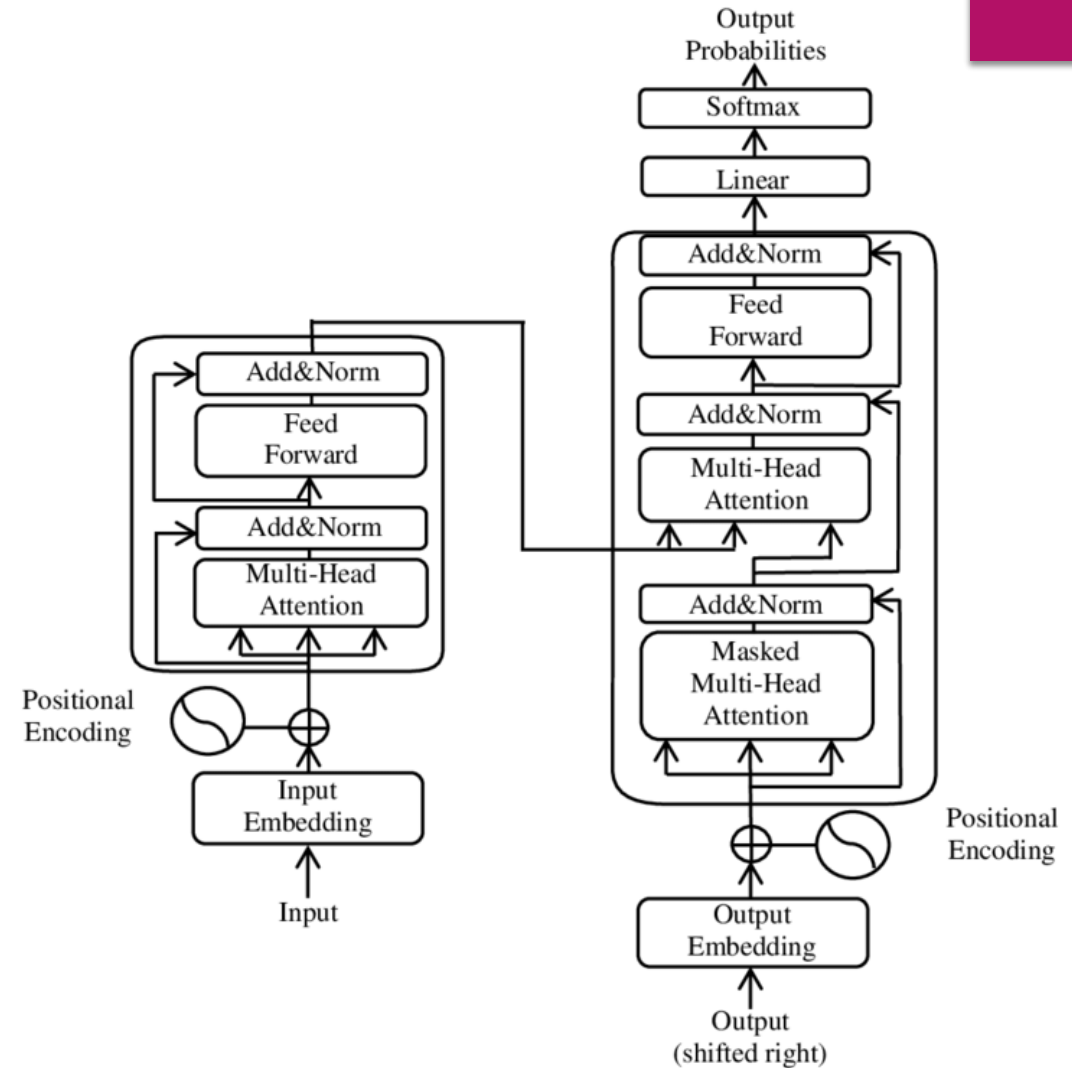
Guillaume Lample[*]
Facebook AI Research
glample@fb.com

François Charton[*]
Facebook AI Research
fcharton@fb.com

- $a * b + c$ → sum prod a b c

- Also: Encoding of Lorentz indices and subscripts

- Wrote sympy package for prefix notation https://github.com/BoGGoG/SympyPrefix

- Blog post: https://boggog.github.io/machine/learning/feynman/physics/symba/2022/07/14/Introduction-Feynman-Amplitudes-Project.html
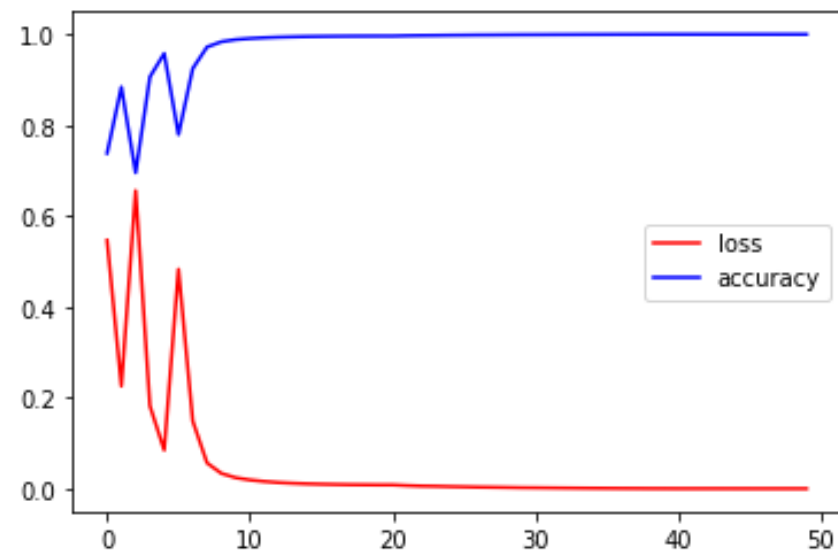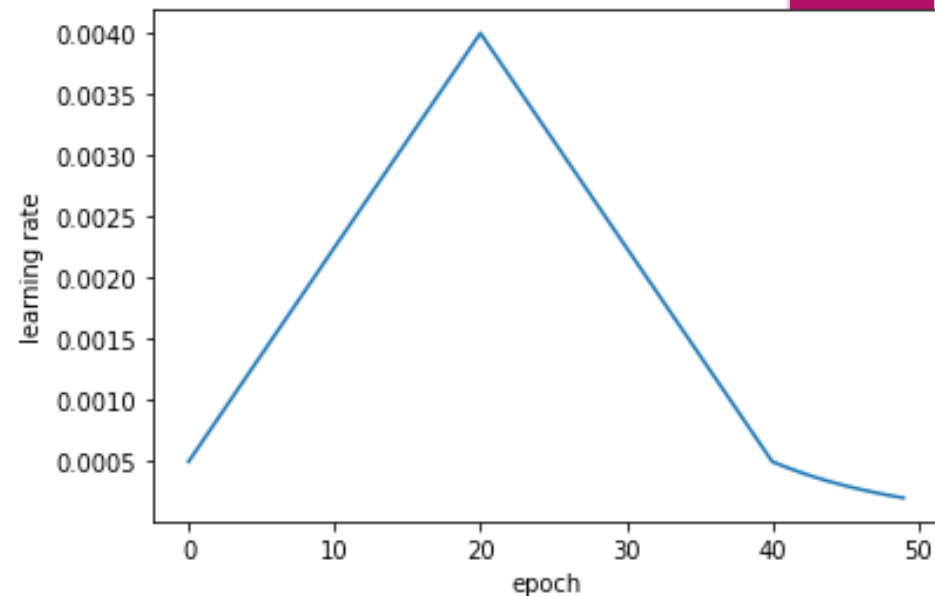
Model: Transformer

▶ predicting one token after the other

▶ maximal sequence length: 350

▶ embed_dim = 256

▶ latent_dim = 2048

▶ num_heads = 8

Training & Results (QED)

- 10h on Google Colab Pro+

- next-token accuracy: 99.98%

- token accuracy (full expressions, ≤ 350 tokens)
  - train: 0.9812
  - test: 0.9655

ToDo until deadline (September 30th)

- ▶ Train on QCD data
- ▶ Implement beam search