Using Transformers to Compute Scattering Amplitudes

Matthias Wilhelm

HAMLET Physics, Copenhagen
August 19th, 2024

[2405.06107] and work in progress with
T. Cai, F. Charton, K. Cranmer, L. Dixon, G. Merz, N. Nolte

The Niels Bohr International Academy SDU DANMARKS FRIE FORSKNINGSFOND VILLUM FONDEN

Table of contents

Machine learning to find New Physics

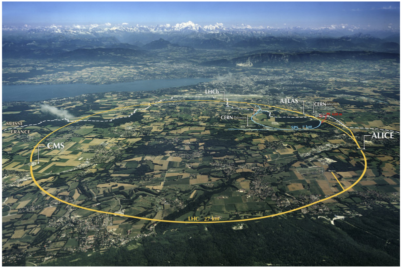


Image: CERN

New Physics at LHC & high-luminosity upgrade?

Machine learning to find New Physics

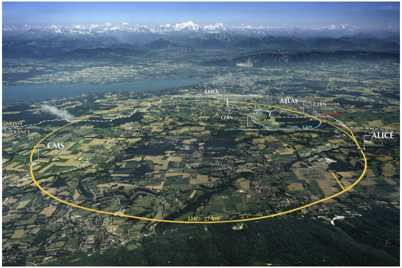


Image: CERN

New Physics at LHC & high-luminosity upgrade?

- Direct detection → Roman's talk
- Indirect detection → this talk
  = compare experimental data to theory predictions

Machine learning to find New Physics

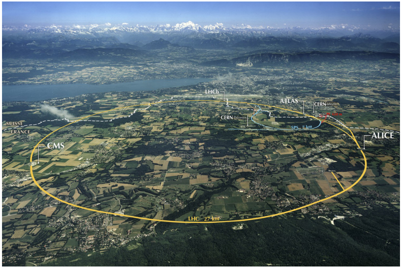


Image: CERN

New Physics at LHC & high-luminosity upgrade?

- Direct detection → Roman's talk
- Indirect detection → this talk
  = compare experimental data to theory predictions

Machine learning to find New Physics

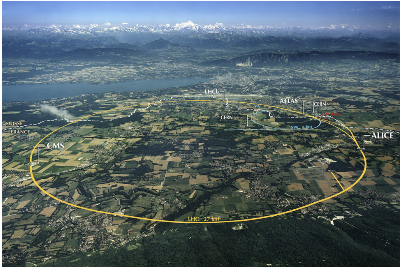


Image: CERN

New Physics at LHC & high-luminosity upgrade?

- Direct detection → Roman's talk
- Indirect detection → this talk
  = compare experimental data to theory predictions

ML for theory predictions?

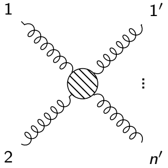
- Analytic / exact results → No noise

↔ ML for math

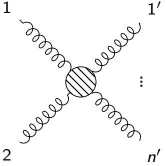
Table of contents

Calculating theory predictions

Scattering amplitude $\mathcal{A}$

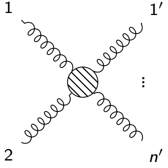


Scattering amplitude $\mathcal{A}$



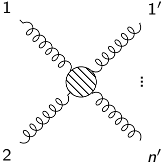
$\xrightarrow{\text{PDFs, ...}}$

Cross section σ

$\sim \int d\Omega \Big|$  $\Big|^2$

Calculating theory predictions

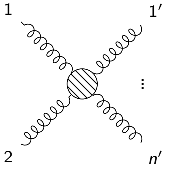
Scattering amplitude $\mathcal{A}$



PDFs, ... $\longrightarrow$

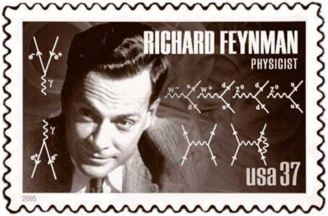
Cross section σ

$\sim \int d\Omega \, \Big| \quad \Big|^2$



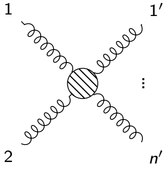
Calculating scattering amplitudes via Quantum Field Theory

$\rightarrow$ **Feynman diagrams** $=$ (Loop) expansion in interaction strenght



Calculating theory predictions

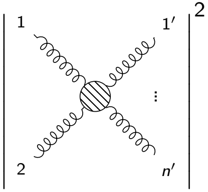
Scattering amplitude $\mathcal{A}$



$\xrightarrow{\text{PDFs, ...}}$

Cross section σ

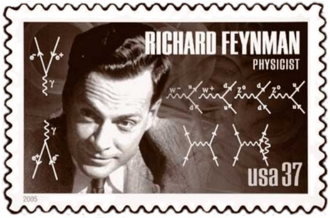
$\sim \int d\Omega$



Calculating scattering amplitudes via Quantum Field Theory

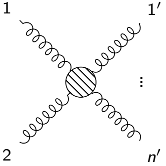
$\rightarrow$ **Feynman diagrams** $=$ (Loop) expansion in interaction strenght

High precision $\rightarrow$ Hard!



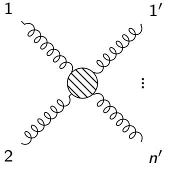
Calculating theory predictions

Scattering amplitude $\mathcal{A}$



PDFs, ... $\longrightarrow$

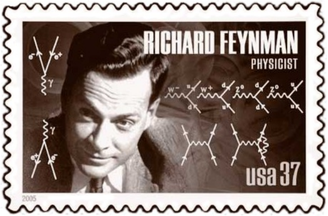
Cross section σ

$\sim \int d\Omega \,\Big|$  $\Big|^2$

Calculating scattering amplitudes via Quantum Field Theory

$\rightarrow$ **Feynman diagrams** $=$ (Loop) expansion in interaction strenght

High precision $\rightarrow$ Hard!



$\rightarrow$ **Can machine learning help?**

Theoretical data

Toy model: Quantum Chromodynamics (QCD)
$\rightarrow$ More symmetric cousin of QCD ("QCD")

Theoretical data

Toy model: Quantum Chromodynamics (QCD)

$\rightarrow$ More symmetric cousin of QCD ("QCD")

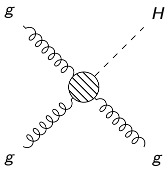
Simpler $\rightarrow$ Results up to eight-loop order ($L = 8$)

[Dixon, McLeod, **MW** (2020)], [Dixon, Gurdogan, McLeod, **MW** (2022)]

Toy model: Quantum Chromodynamics (QCD)
$\rightarrow$ More symmetric cousin of QCD ("QCD")

Simpler $\rightarrow$ Results up to eight-loop order ($L = 8$)

[Dixon, McLeod, **MW** (2020)], [Dixon, Gurdogan, McLeod, **MW** (2022)]



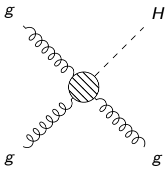
2 gluons $\rightarrow$ gluon + Higgs

Theoretical data

Toy model: Quantum Chromodynamics (QCD)
$\rightarrow$ More symmetric cousin of QCD ("QCD")

Simpler $\rightarrow$ Results up to eight-loop order ($L = 8$)

[Dixon, McLeod, **MW** (2020)], [Dixon, Gurdogan, McLeod, **MW** (2022)]



2 gluons $\rightarrow$ gluon + Higgs

magic $\longrightarrow$

tensor product in letters a,b,c,d,e,f with length $2L$ and integer coefficients

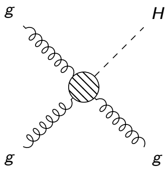
Theoretical data

Toy model: Quantum Chromodynamics (QCD)

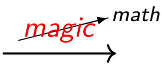
$\rightarrow$ More symmetric cousin of QCD ("QCD")

Simpler $\rightarrow$ Results up to eight-loop order ($L = 8$)

[Dixon, McLeod, **MW** (2020)], [Dixon, Gurdogan, McLeod, **MW** (2022)]



2 gluons $\rightarrow$ gluon + Higgs



tensor product in letters a,b,c,d,e,f with length $2L$ and integer coefficients

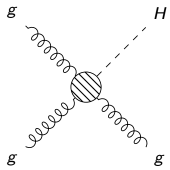
Theoretical data

Toy model: Quantum Chromodynamics (QCD)

$\rightarrow$ More symmetric cousin of QCD ("QCD")

Simpler $\rightarrow$ Results up to eight-loop order ($L = 8$)

[Dixon, McLeod, **MW** (2020)], [Dixon, Gurdogan, McLeod, **MW** (2022)]



2 gluons $\rightarrow$ gluon + Higgs

~~magic~~ math $\longrightarrow$

tensor product in letters a,b,c,d,e,f with length $2L$ and integer coefficients

$L = 1$: $(-2)[b \otimes d + c \otimes e + a \otimes f + b \otimes f + c \otimes d + a \otimes e]$

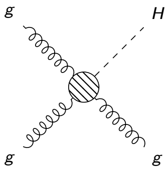
Theoretical data

Toy model: Quantum Chromodynamics (QCD)

$\rightarrow$ More symmetric cousin of QCD ("QCD")

Simpler $\rightarrow$ Results up to eight-loop order ($L = 8$)

[Dixon, McLeod, **MW** (2020)], [Dixon, Gurdogan, McLeod, **MW** (2022)]



2 gluons $\rightarrow$ gluon + Higgs

tensor product in letters a,b,c,d,e,f with length $2L$ and integer coefficients

$L = 1$: $(-2)[b \otimes d + c \otimes e + a \otimes f + b \otimes f + c \otimes d + a \otimes e]$

$L = 8$: 1.7 billion terms $\rightarrow$ **Big data**

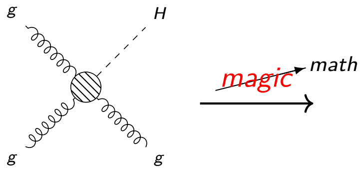
Theoretical data

Toy model: Quantum Chromodynamics (QCD)

$\rightarrow$ More symmetric cousin of QCD ("QCD")

Simpler $\rightarrow$ Results up to eight-loop order ($L = 8$)

[Dixon, McLeod, **MW** (2020)], [Dixon, Gurdogan, McLeod, **MW** (2022)]



2 gluons $\rightarrow$ gluon + Higgs

tensor product in letters a,b,c,d,e,f with length $2L$ and integer coefficients

$L = 1$: $(-2)[b \otimes d + c \otimes e + a \otimes f + b \otimes f + c \otimes d + a \otimes e]$
$L = 8$: 1.7 billion terms $\rightarrow$ **Big data**

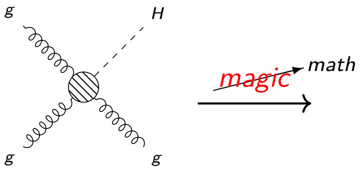
Can a machine learn this and predict higher loops?

Theoretical data

Toy model: Quantum Chromodynamics (QCD)
$\rightarrow$ More symmetric cousin of QCD ("QCD")

Simpler $\rightarrow$ Results up to eight-loop order ($L = 8$)

[Dixon, McLeod, **MW** (2020)], [Dixon, Gurdogan, McLeod, **MW** (2022)]



2 gluons $\rightarrow$ gluon + Higgs

tensor product in letters a,b,c,d,e,f with length $2L$ and integer coefficients

$L = 1$: $(-2)[b \otimes d + c \otimes e + a \otimes f + b \otimes f + c \otimes d + a \otimes e]$
$L = 8$: 1.7 billion terms $\rightarrow$ **Big data**

Can a machine learn this and predict higher loops?

Case for ML: Results are hard to calculate but easy to check!

Table of contents

ML formulation

$L = 1$: $(-2)[b \otimes d + c \otimes e + a \otimes f + b \otimes f + c \otimes d + a \otimes e]$

$L = 6$: $-52800 c \otimes c \otimes e \otimes c \otimes c \otimes c \otimes e \otimes e \otimes c \otimes d \otimes d \otimes d + \ldots$

Tensor product → Pairs of words and numbers

E.g. `bd` → −2, `ce` → −2, `aa` → 0, …,

`cceccceecddd` → −52800, …

ML formulation

$L = 1$: $(-2)[b \otimes d + c \otimes e + a \otimes f + b \otimes f + c \otimes d + a \otimes e]$

$L = 6$: $-52800 c \otimes c \otimes e \otimes c \otimes c \otimes c \otimes e \otimes e \otimes c \otimes d \otimes d \otimes d + \dots$

Tensor product → Pairs of words and numbers

E.g. bd → −2, ce → −2, aa → 0, ...,

cceccceecddd → −52800, ...

Idea: "Translate" words to numbers

→ Transformer architecture!

(ChatGPT, etc.)

Implementation

Tokenizing numbers base 1000 with sign,
e.g. $-52800 \rightarrow$ `-,52,800`

Accuracy: % of numbers correctly predicted

Classification task $\rightarrow$ Minimize cross entropy

Implementation

Tokenizing numbers base 1000 with sign,
e.g. $-52800 \rightarrow$ `-,52,800`

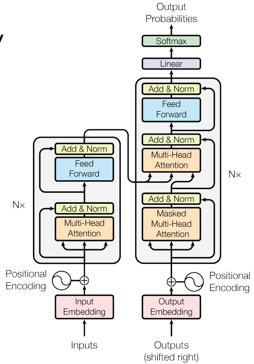
Accuracy: % of numbers correctly predicted

Classification task $\rightarrow$ Minimize cross entropy

Encoder-decoder transformer

- 1–8 layers
- 256–1024 dimensions
- 8–16 attention heads
- 4.5–245 million parameters

$\rightarrow$ Much smaller than ChatGPT! (tens of billions)



[Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser, Polosukhin (2017)]

Implementation

Tokenizing numbers base 1000 with sign,
e.g. $-52800 \to$ `-,52,800`

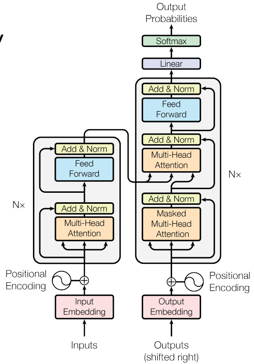
Accuracy: % of numbers correctly predicted

Classification task → Minimize cross entropy

Encoder-decoder transformer

- 1–8 layers
- 256–1024 dimensions
- 8–16 attention heads
- 4.5–245 million parameters

→ Much smaller than ChatGPT! (tens of billions)



[Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser, Polosukhin (2017)]

Training: single V100 (A100)

Table of contents

Can the machine learn the coefficients and structure of the result?

Experiment 1: Vanishing coefficients = Simple structure

Number of total and non-zero coefficients for loops $L = 1, \ldots, 8$

Loop	1	2	3	4	5	6	7	8
Total (6^{2L})	36	1,296	46,656	$1.7 \cdot 10^6$	$6.0 \cdot 10^7$	$2.2 \cdot 10^9$	$7.8 \cdot 10^{10}$	$2.8 \cdot 10^{12}$
Total nonzero	6	12	636	11,208	263,880	$4.9 \cdot 10^6$	$9.3 \cdot 10^7$	$1.7 \cdot 10^9$

Experiment 1: Vanishing coefficients = Simple structure

Number of total and non-zero coefficients for loops $L = 1, \ldots, 8$

Loop	1	2	3	4	5	6	7	8
Total (6^{2L})	36	1,296	46,656	$1.7 \cdot 10^6$	$6.0 \cdot 10^7$	$2.2 \cdot 10^9$	$7.8 \cdot 10^{10}$	$2.8 \cdot 10^{12}$
Total nonzero	6	12	636	11,208	263,880	$4.9 \cdot 10^6$	$9.3 \cdot 10^7$	$1.7 \cdot 10^9$

Simple rules: Zero if

- starts in `d`, `e` or `f`
- ends in `a`, `b` or `c`
- contains (`d` next to `a`, `e` or `f`), (`e` next to `b`, `d` or `f`) or (`f` next to `c`, `d` or `e`)

Experiment 1: Vanishing coefficients = Simple structure

Number of total and non-zero coefficients for loops $L = 1, \ldots, 8$

Loop	1	2	3	4	5	6	7	8
Total (6^{2L})	36	$1,296$	$46,656$	$1.7 \cdot 10^6$	$6.0 \cdot 10^7$	$2.2 \cdot 10^9$	$7.8 \cdot 10^{10}$	$2.8 \cdot 10^{12}$
Total nonzero	6	12	636	$11,208$	$263,880$	$4.9 \cdot 10^6$	$9.3 \cdot 10^7$	$1.7 \cdot 10^9$

Simple rules: Zero if

- starts in `d`, `e` or `f`
- ends in `a`, `b` or `c`
- contains (`d` next to `a`, `e` or `f`), (`e` next to `b`, `d` or `f`) or (`f` next to `c`, `d` or `e`)

Can this be learned?

Experiment 1: Vanishing coefficients = Simple structure

Number of total and non-zero coefficients for loops $L = 1, \ldots, 8$

Loop	1	2	3	4	5	6	7	8
Total (6^{2L})	36	1,296	46,656	$1.7 \cdot 10^6$	$6.0 \cdot 10^7$	$2.2 \cdot 10^9$	$7.8 \cdot 10^{10}$	$2.8 \cdot 10^{12}$
Total nonzero	6	12	636	11,208	263,880	$4.9 \cdot 10^6$	$9.3 \cdot 10^7$	$1.7 \cdot 10^9$

Simple rules: Zero if

- starts in `d`, `e` or `f`
- ends in `a`, `b` or `c`
- contains (`d` next to `a`, `e` or `f`), (`e` next to `b`, `d` or `f`) or (`f` next to `c`, `d` or `e`)

Can this be learned?

Experiment 1

Data: All non-zero terms + equal number zeros at $L = 6$

Test set: 10,000

Results: **99.91%** after 1 epoch (300,000) and **99.97%** after 2

Experiment 1: Vanishing coefficients = Simple structure

Number of total and non-zero coefficients for loops $L = 1, \ldots, 8$

Loop	1	2	3	4	5	6	7	8
Total (6^{2L})	36	$1,296$	$46,656$	$1.7 \cdot 10^6$	$6.0 \cdot 10^7$	$2.2 \cdot 10^9$	$7.8 \cdot 10^{10}$	$2.8 \cdot 10^{12}$
Total nonzero	6	12	636	$11,208$	$263,880$	$4.9 \cdot 10^6$	$9.3 \cdot 10^7$	$1.7 \cdot 10^9$

Simple rules: Zero if

- starts in `d`, `e` or `f`
- ends in `a`, `b` or `c`
- contains (`d` next to `a`, `e` or `f`), (`e` next to `b`, `d` or `f`) or (`f` next to `c`, `d` or `e`)

Can this be learned? ✓

Experiment 1

Data: All non-zero terms + equal number zeros at $L = 6$

Test set: 10,000

Results: **99.91%** after 1 epoch (300,000) and **99.97%** after 2

Experiment 2: Predicting coefficients from words

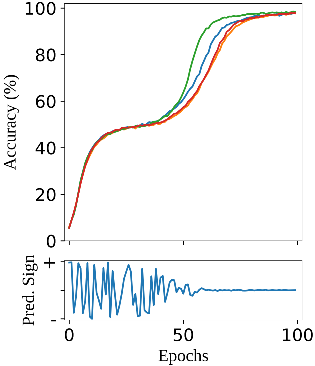
Data: All non-zero terms at $L = 6$
Test set: 100,000
Results: **99.3%** after 199 epochs ✓

Data: All non-zero terms at $L = 6$

Test set: 100,000

Results: **99.3%** after 199 epochs ✓

Learning dynamics:



Double plateau: First magnitude, then sign!

Experiment 2: More advanced structures

Bose symmetry: Invariance under permuting gluons

- Cycle: `a,b,c,d,e,f` $\rightarrow$ `b,c,a,e,f,d`
- Flip: `a,b,c,d,e,f` $\rightarrow$ `a,c,b,d,f,e`

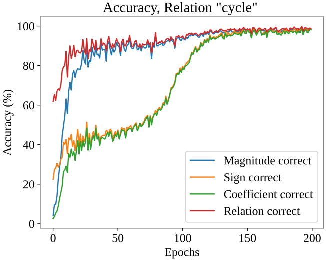
Satyendra Nath Bose, 1925

Experiment 2: More advanced structures

Bose symmetry: Invariance under permuting gluons

- Cycle: `a,b,c,d,e,f` → `b,c,a,e,f,d`
- Flip: `a,b,c,d,e,f` → `a,c,b,d,f,e`

Satyendra Nath Bose, 1925



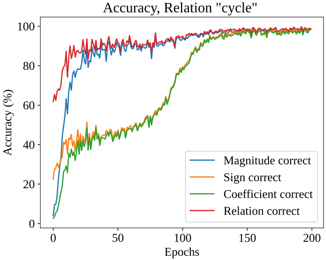
Learning dynamics

⇒ Learned before coefficients! ✓

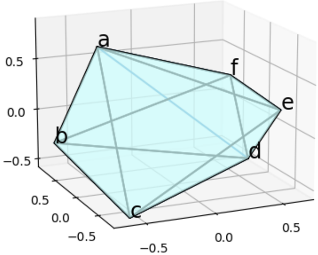
Bose symmetry: Invariance under permuting gluons

- Cycle: `a,b,c,d,e,f` $\to$ `b,c,a,e,f,d`
- Flip: `a,b,c,d,e,f` $\to$ `a,c,b,d,f,e`

Satyendra Nath Bose, 1925



Learning dynamics



Embedding space PCA

$\Rightarrow$ Learned before coefficients! ✓

Table of contents

Conclusion

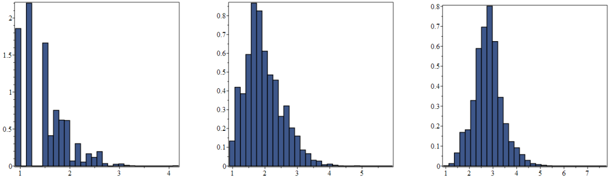
Machine learning for Theory Predictions?

- ML formulation of 2 gluon $\to$ gluon $+$ Higgs in "QCD": words with integer coefficients
- Transformers can learn
  - coefficients from words
  - simple structures such as Bose symmetry
- Hints at relations between loop orders

Outlook and Challenges for loop generalization

To the next loop: Extrapolation, not interpolation!

- Distribution shift of coefficients



$L = 4$ $L = 5$ $L = 6$

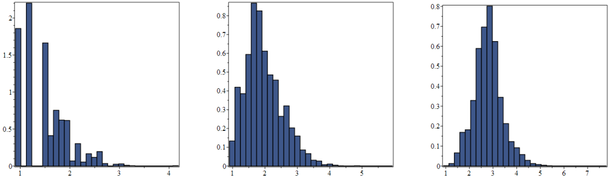
- Discover relations between loops explicitely?
- All-loop coefficients: $\texttt{b}\,\underbrace{\texttt{d...d}}_{2L-1} \to (-1)^L 2^{2L-1}(2L-3)!!$

  $\to$ Can ML learn more?
- Use structure $\to$ Foundation model?
- Accuracy $\lnsim$ 100% $\to$ Error correction

Outlook and Challenges for loop generalization

To the next loop: Extrapolation, not interpolation!

- Distribution shift of coefficients



$L = 4$ $L = 5$ $L = 6$

- Discover relations between loops explicitly?
- All-loop coefficients: $\mathtt{b}\underbrace{\mathtt{d...d}}_{2L-1} \to (-1)^L 2^{2L-1}(2L-3)!!$

  $\to$ Can ML learn more?
- Use structure $\to$ Foundation model?
- Accuracy $\lesssim$ 100% $\to$ Error correction

Thank you!