Learning the Simplicity of Scattering Amplitudes

ML4Jets 2024

Clifford Cheung, **Aurélien Dersy**, Matthew D.Schwartz

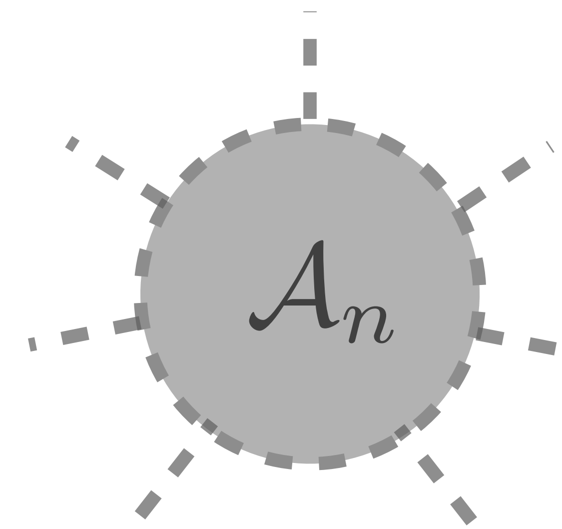
2408.04720

November 5, 2024

What do we care about and why ?

- **Basic building block** in hep-th ?

→ Amplitudes



Why useful ?

- Calculate observables

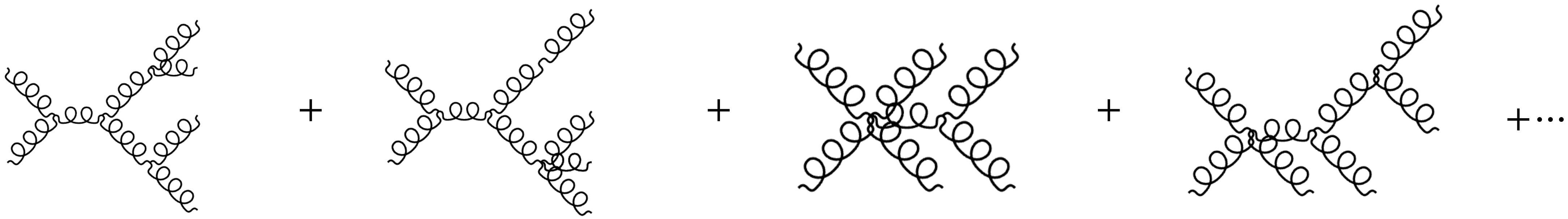
- Gain theoretical insight

→ Staring (intensely) at a **simple** amplitude can teach us a lot !

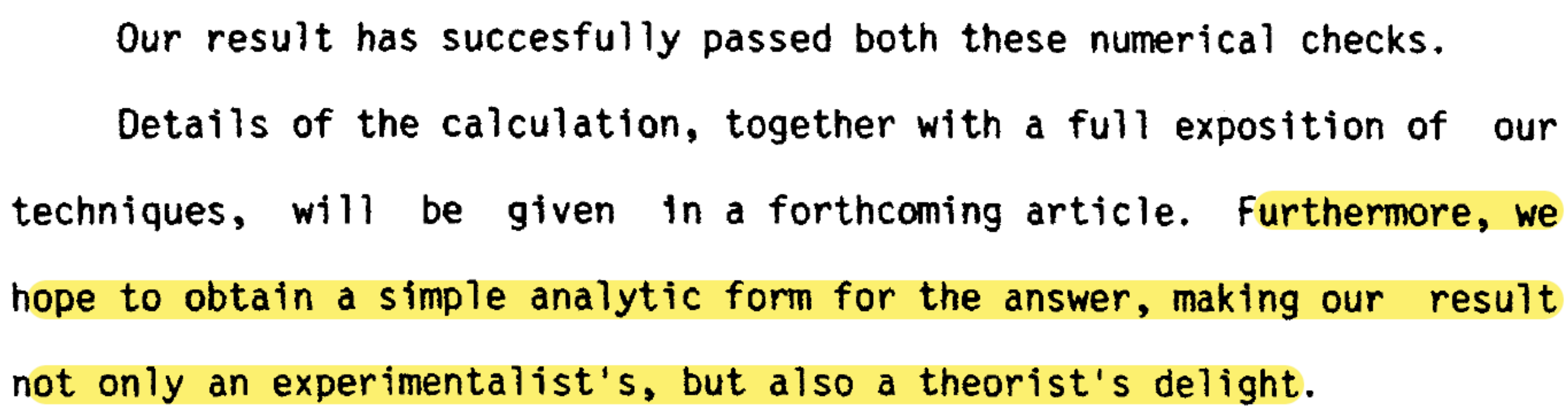
e.g : Double Copy, Amplituhedron, Twistor Grassmanian

Why should amplitudes be simple ?

Process: $gg \to gggg$



220 Feynman diagrams - pages and pages of results !



Our result has succesfully passed both these numerical checks.

Details of the calculation, together with a full exposition of our techniques, will be given in a forthcoming article. Furthermore, we hope to obtain a simple analytic form for the answer, making our result not only an experimentalist's, but also a theorist's delight.

Parke,Taylor (1985)

A theorist's and experimentalist's delight

Process: $gg \to gggg$

Simple !

$$|\mathcal{M}_6(1^-2^-3^+4^+5^+6^+)|^2 = \frac{(p_1 \cdot p_2)^3}{(p_2 \cdot p_3)(p_3 \cdot p_4)(p_4 \cdot p_5)(p_5 \cdot p_6)(p_6 \cdot p_1)}$$

Equation (3) has the correct dimensions and symmetry properties for this n-particle scattering amplitude squared. Also it agrees with the known results[4,5] for $n = 4$, 5, and 6. The agreement for $n = 6$ is numerical.[5,6]

Parke, Taylor (1986)

→ Via **guesswork**: complicated calculation matches a **simple** result !

Spinor-Helicity Formulation

Process: $gg \to gggg$

"Modern" formula

Helicity spinors brackets

$$\mathcal{M}_6(1^-2^-3^+4^+5^+6^+) = \frac{\langle 12\rangle^3}{\langle 23\rangle\langle 34\rangle\langle 45\rangle\langle 56\rangle\langle 61\rangle}$$

$$\langle ij\rangle = \sqrt{2p_i \cdot p_j}\, e^{i\phi}$$

$$[ij] = \sqrt{2p_i \cdot p_j}\, e^{-i\phi}$$

→ Spinor helicity formalism expedites calculations

→ Reduction of spinor helicity amplitudes still requires **guesswork** or **clever ansatz** constructions

ML as an oracle

<u>Goal</u>: Do the guesswork with ML

$$
\begin{aligned}
\mathcal{M} = {} & \frac{\langle 12\rangle^3 [13]}{\langle 23\rangle\langle 24\rangle\langle 35\rangle\langle 45\rangle [23]} + \frac{\langle 12\rangle^3 [14] [25]}{\langle 13\rangle\langle 24\rangle\langle 35\rangle\langle 45\rangle [12] [45]} \\
& - \frac{\langle 12\rangle^2 [13] [24] [35]}{\langle 24\rangle\langle 35\rangle\langle 45\rangle [12] [23] [45]} + \frac{\langle 12\rangle^2 [13] [25] [34]}{\langle 24\rangle\langle 35\rangle\langle 45\rangle [12] [23] [45]} \\
& - \frac{\langle 12\rangle^2 \langle 23\rangle [13] [45]}{\langle 15\rangle\langle 24\rangle\langle 34\rangle\langle 35\rangle [14] [15]} - \frac{\langle 12\rangle^2 [45]}{\langle 15\rangle\langle 34\rangle\langle 35\rangle [15]} \\
& + \frac{\langle 12\rangle^2 \langle 34\rangle [13] [34] [45]}{\langle 15\rangle\langle 23\rangle\langle 24\rangle\langle 35\rangle [12] [15] [23]} - \frac{\langle 12\rangle^2 \langle 15\rangle [15]}{\langle 13\rangle\langle 24\rangle\langle 35\rangle\langle 45\rangle [12]} \\
& + \frac{\langle 12\rangle^2 \langle 23\rangle [15] [24] [34]}{\langle 13\rangle\langle 24\rangle\langle 35\rangle\langle 45\rangle [12] [14] [45]} + \frac{\langle 12\rangle^2 \langle 23\rangle [14] [23] [25]}{\langle 13\rangle\langle 24\rangle\langle 35\rangle\langle 45\rangle [12]^2 [45]} \\
& + \frac{\langle 12\rangle^2 \langle 23\rangle [15] [24]}{\langle 13\rangle\langle 24\rangle\langle 34\rangle\langle 35\rangle [12] [14]} + \frac{\langle 12\rangle^2 [24]}{\langle 13\rangle\langle 35\rangle\langle 45\rangle [12]} \\
& + \frac{\langle 12\rangle^2 [24] [45]}{\langle 13\rangle\langle 15\rangle\langle 35\rangle [12] [15]} + \frac{\langle 12\rangle^2 \langle 34\rangle [13] [24] [34] [45]}{\langle 13\rangle\langle 15\rangle\langle 24\rangle\langle 35\rangle [12] [14] [15] [23]} \\
& + \frac{\langle 12\rangle\langle 13\rangle\langle 23\rangle [13] [34]}{\langle 15\rangle\langle 24\rangle\langle 34\rangle\langle 35\rangle [12] [14]} + \frac{\langle 12\rangle\langle 13\rangle [34]}{\langle 15\rangle\langle 34\rangle\langle 35\rangle [12]} \\
& + \frac{\langle 12\rangle\langle 14\rangle [14] [45]}{\langle 15\rangle\langle 34\rangle\langle 35\rangle [12] [15]} + \frac{\langle 12\rangle\langle 14\rangle [13] [34] [45]}{\langle 15\rangle\langle 24\rangle\langle 35\rangle [12] [15] [23]} \\
& - \frac{\langle 12\rangle\langle 23\rangle [34] [35]}{\langle 24\rangle\langle 35\rangle\langle 45\rangle [12] [45]} + \frac{\langle 12\rangle\langle 23\rangle [15] [34]^2}{\langle 24\rangle\langle 35\rangle\langle 45\rangle [12] [14] [45]} \\
& + \cdots
\end{aligned}
$$

→

ML Model

→

$$\overline{\mathcal{M}} = -\frac{\langle 12\rangle^3}{\langle 15\rangle\langle 23\rangle\langle 34\rangle\langle 45\rangle}$$

Check answer numerically !

Can ML do a symbolic simplification of spinor-helicity amplitudes ?

Transformer for translation

Idea introduced for integration + ODE's

[1912.01412 by Lample, Charton]

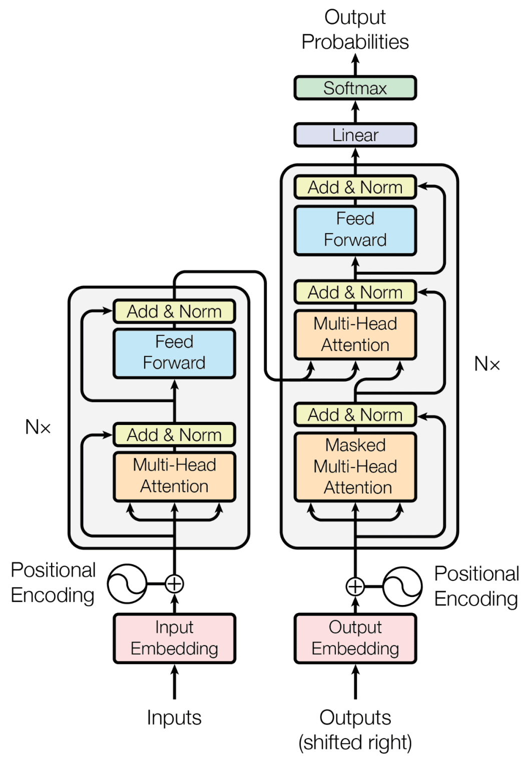
Symbolic translation task

$$\mathcal{M} = \frac{\langle 12\rangle^3\,[13]}{\langle 23\rangle\langle 24\rangle\langle 35\rangle\langle 45\rangle\,[23]} + \frac{\langle 12\rangle^3\,[14]\,[25]}{\langle 13\rangle\langle 24\rangle\langle 35\rangle\langle 45\rangle\,[12]\,[45]} - \frac{\langle 12\rangle^2\,[13]\,[24]\,[35]}{\langle 24\rangle\langle 35\rangle\langle 45\rangle\,[12]\,[23]\,[45]} + \frac{\langle 12\rangle^2\,[13]\,[25]\,[34]}{\langle 24\rangle\langle 35\rangle\langle 45\rangle\,[12]\,[23]\,[45]} - \frac{\langle 12\rangle^2\langle 23\rangle\,[13]\,[45]}{\langle 15\rangle\langle 24\rangle\langle 34\rangle\langle 35\rangle\,[14]\,[15]} - \frac{\langle 12\rangle^2\,[45]}{\langle 15\rangle\langle 34\rangle\langle 35\rangle\,[15]}$$

"French"

Encoder - Decoder

Transformer



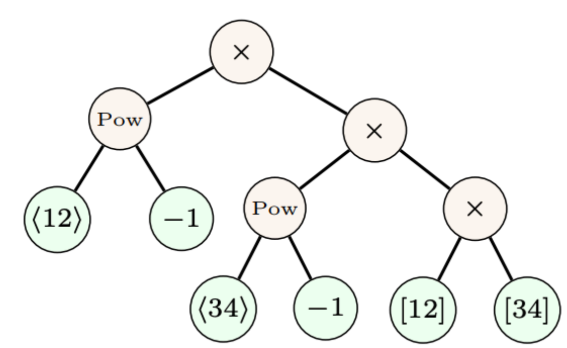
[1706.03762]

$$\overline{\mathcal{M}} = -\frac{\langle 12\rangle^3}{\langle 15\rangle\langle 23\rangle\langle 34\rangle\langle 45\rangle}$$

"English"

A mathematical expression is language !

$$\frac{[12][34]}{\langle 12\rangle\langle 34\rangle} =$$



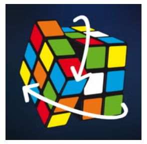
$= [\text{'mul'}, \text{'pow'}, \text{'ab12'}, \text{'-1'}, \text{'mul'}, \cdots]$

Data generation starting from simple amplitudes

Backward generation

$$\overline{\mathcal{M}} = \frac{\langle 12\rangle\,[14]\,[46]^2 - \langle 25\rangle\,[45]\,[46]^2}{\langle 13\rangle\langle 45\rangle^2\,[12]\,[15]\,[16]\,[56]}$$

Simple amplitude



$$\mathcal{M} = \frac{\langle 15\rangle\langle 26\rangle\,[14]\,[46]^2}{\langle 13\rangle\langle 45\rangle^2\langle 56\rangle\,[12]\,[15]\,[16]\,[56]} + \frac{\langle 25\rangle\,[24]\,[46]^2}{\langle 13\rangle\langle 45\rangle^2\,[12]\,[15]\,[16]\,[26]} - \frac{\langle 25\rangle\,[25]\,[46]^3}{\langle 13\rangle\langle 45\rangle^2\,[12]\,[15]\,[16]\,[26]\,[56]} + \frac{\langle 16\rangle\langle 25\rangle\,[13]\,[14]\,[46]^2}{\langle 12\rangle\langle 45\rangle^2\langle 56\rangle\,[12]^2\,[15]\,[16]\,[56]} - \frac{\langle 16\rangle\langle 25\rangle\,[14]\,[45]\,[46]^2}{\langle 12\rangle\langle 13\rangle\langle 45\rangle\langle 56\rangle\,[12]^2\,[15]\,[16]\,[56]} + \frac{\langle 16\rangle\langle 23\rangle\langle 25\rangle\,[14]\,[23]\,[46]^2}{\langle 12\rangle\langle 13\rangle\langle 45\rangle^2\langle 56\rangle\,[12]^2\,[15]\,[16]\,[56]} - \frac{\langle 16\rangle\langle 25\rangle\langle 46\rangle\,[14]\,[46]^3}{\langle 12\rangle\langle 13\rangle\langle 45\rangle^2\langle 56\rangle\,[12]^2\,[15]\,[16]\,[56]} - \frac{\langle 16\rangle\langle 25\rangle\,[14]\,[46]^2}{\langle 12\rangle\langle 13\rangle\langle 45\rangle^2\,[12]^2\,[15]\,[16]}$$

"Scrambled" version

Simple $\overline{\mathcal{M}}$ guidelines

- 4-6 external momenta
- 0-3 numerators
- Monomial denominator
- $[ij]$, $\langle ij\rangle$ only

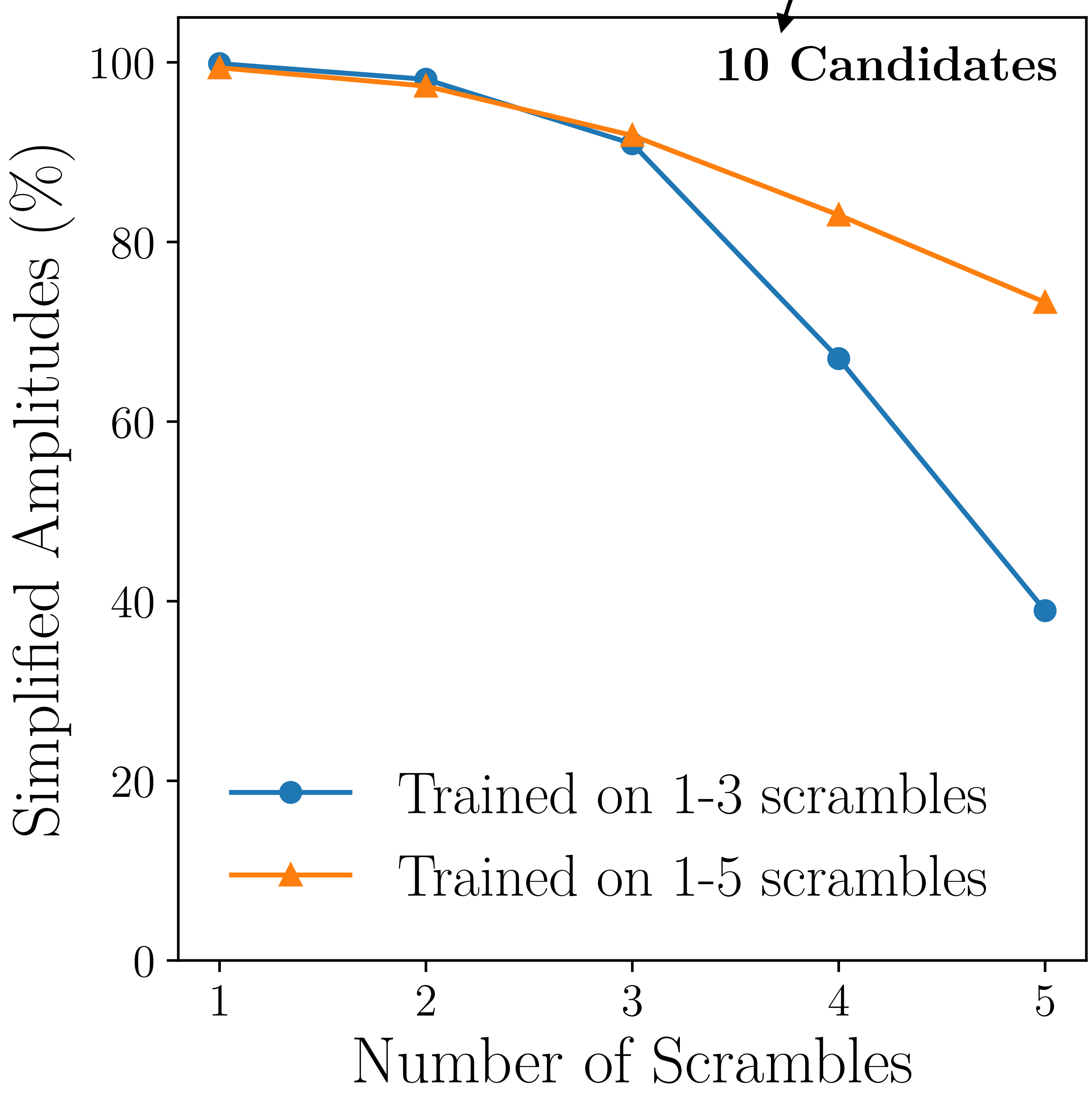
Scrambling moves

Schouten $\langle ij\rangle\langle kl\rangle = \langle il\rangle\langle kj\rangle + \langle ik\rangle\langle jl\rangle$

Momentum conservation $\sum_{j=1}^{n}\langle ij\rangle[jk] = 0$

Results for simple simplification: 5-pt amplitudes

Beam Search and
Nucleus Sampling

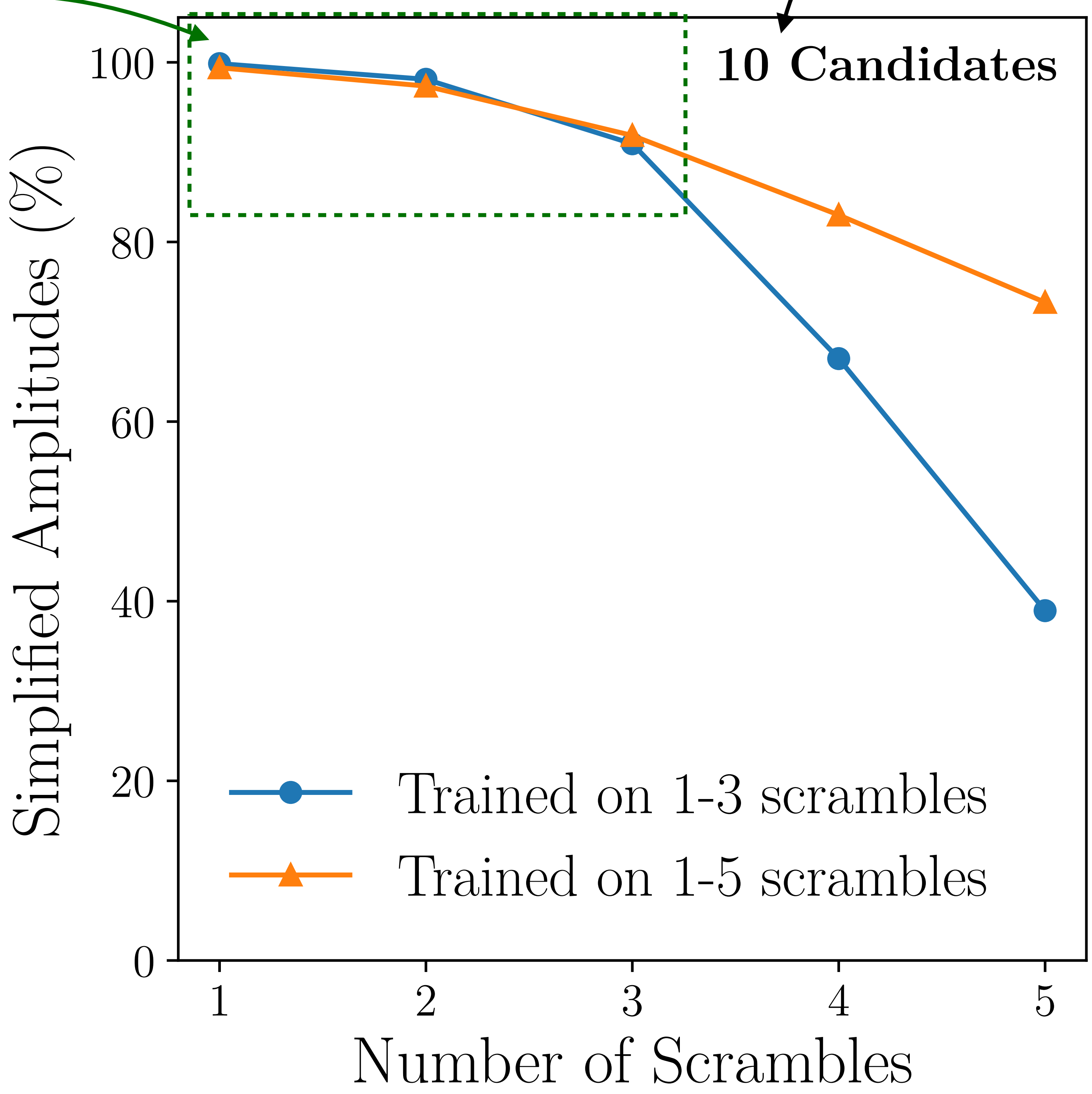


Great for simple simplifications

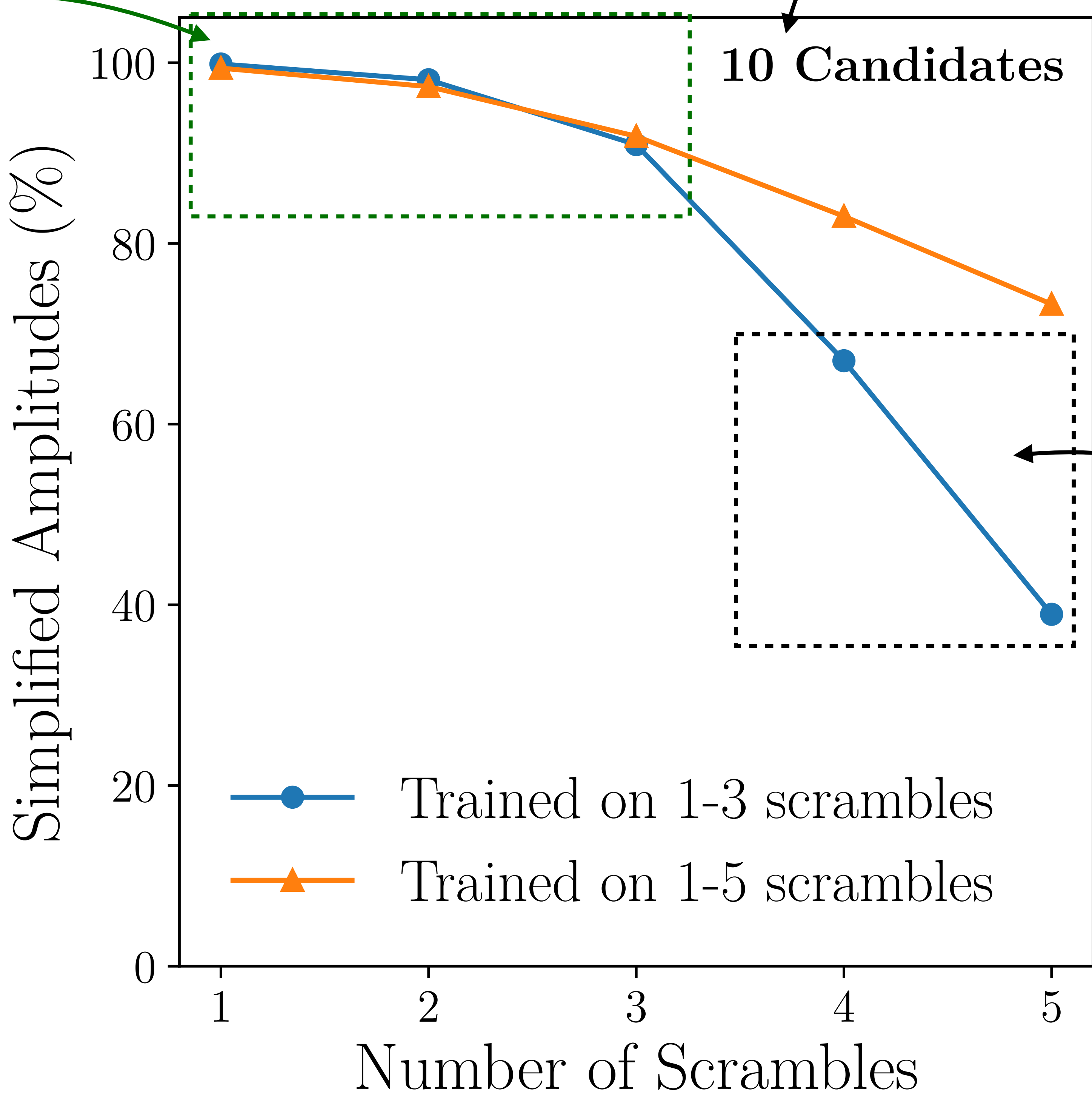
Beam Search and
Nucleus Sampling

Model performs great
for simple inputs

~ 20 terms



Poor generalization for longer inputs



Beam Search and Nucleus Sampling

Model performs great for simple inputs

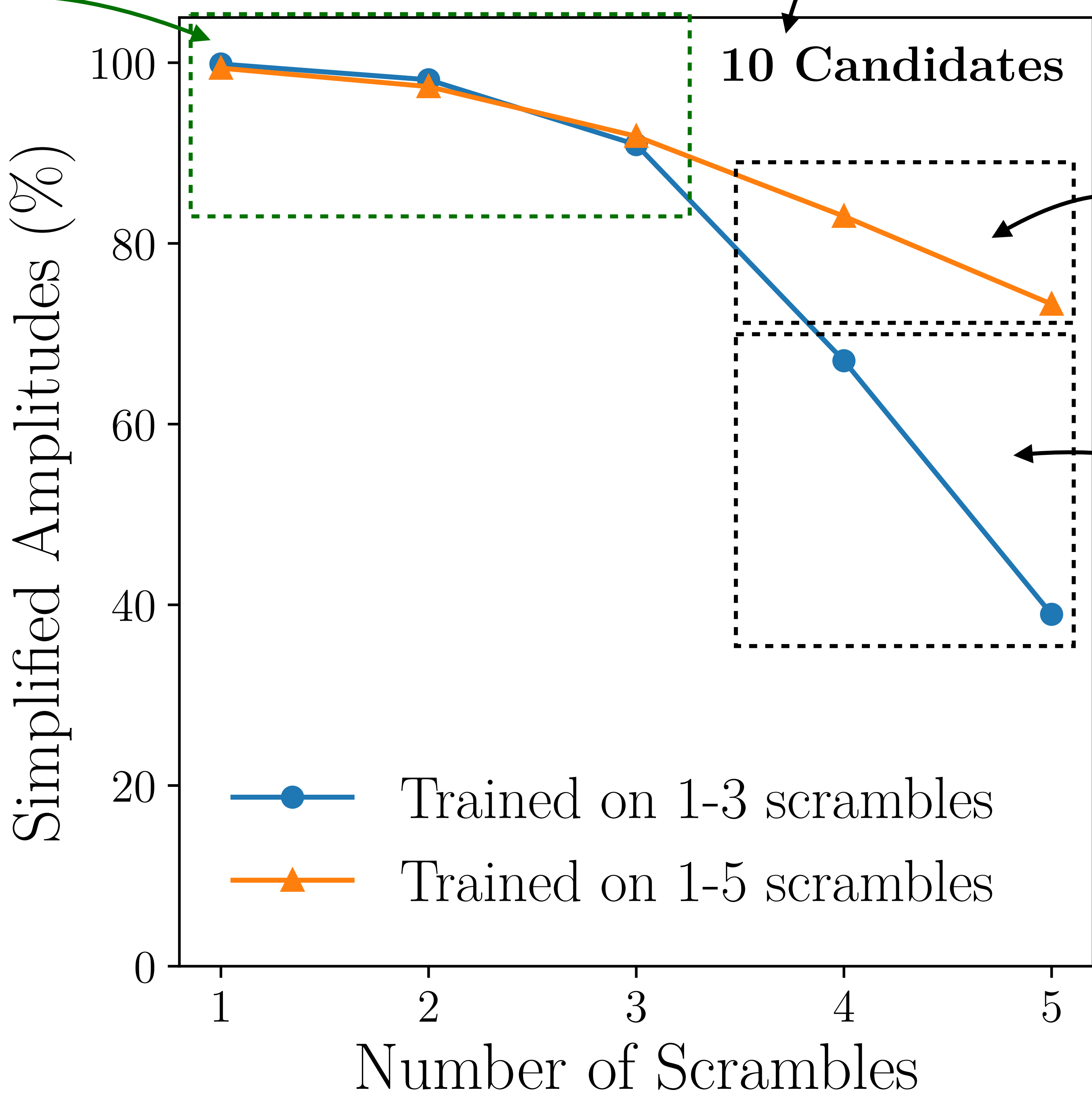
~ 20 terms

Poor generalization out of training domain

Poor generalization for longer inputs

Beam Search and
Nucleus Sampling

Model performs great
for simple inputs
~ 20 terms



Better but under
90 % accurate

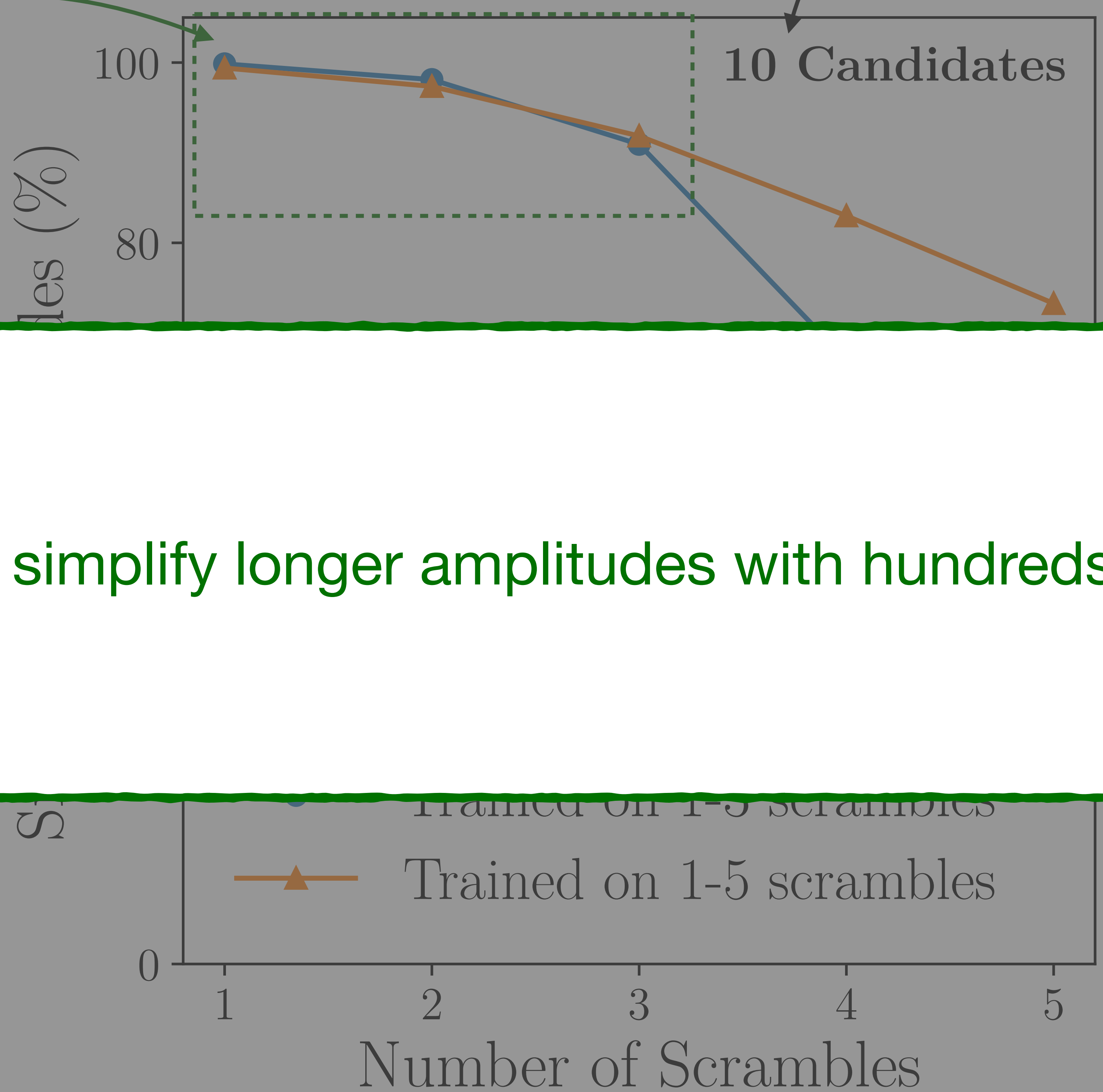
Poor generalization
out of training domain

Poor generalization for longer inputs

Beam Search and
Nucleus Sampling

Model performs great
for simple inputs

~ 20 terms

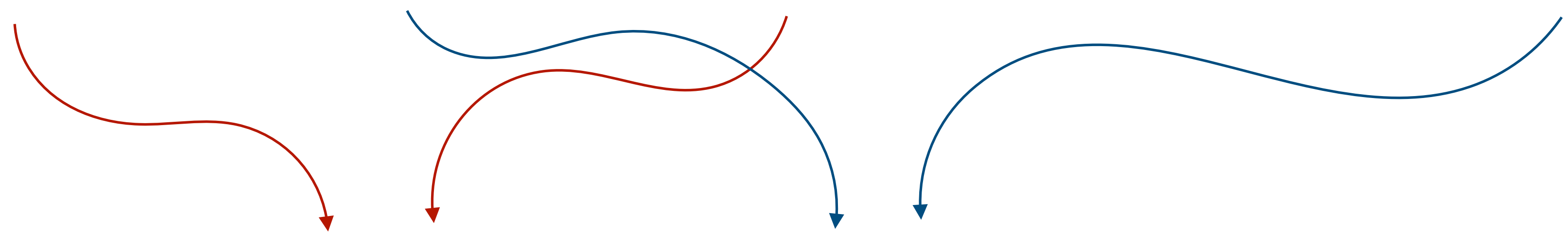


How can we simplify longer amplitudes with hundreds of terms ?

Human-like simplification

Human : Focus attention on terms likely to simplify

$$\langle 13\rangle\langle 24\rangle[13][25] + \langle 14\rangle\langle 34\rangle[25][34] - \langle 14\rangle\langle 23\rangle[13][25] + \langle 14\rangle\langle 24\rangle[12][45] + \langle 14\rangle\langle 34\rangle[23][45] + \cdots$$



$$\langle 12\rangle\langle 34\rangle[13][25] + \langle 14\rangle\langle 34\rangle[13][45] + \langle 14\rangle\langle 24\rangle[12][45] + \cdots$$

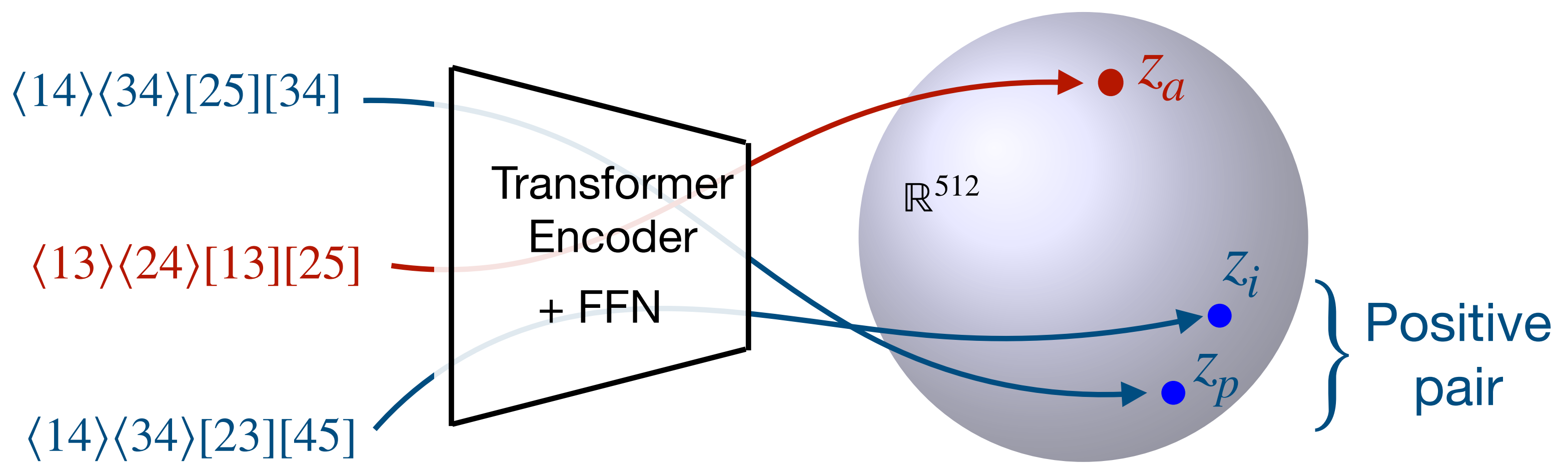
$\rightarrow$ We use a <u>sparse</u> attention pattern

How can we mimic this and learn to group similar terms ?

Supervised Contrastive Learning

Goal: Pull together similar terms in embedding space

→Train this embedding space

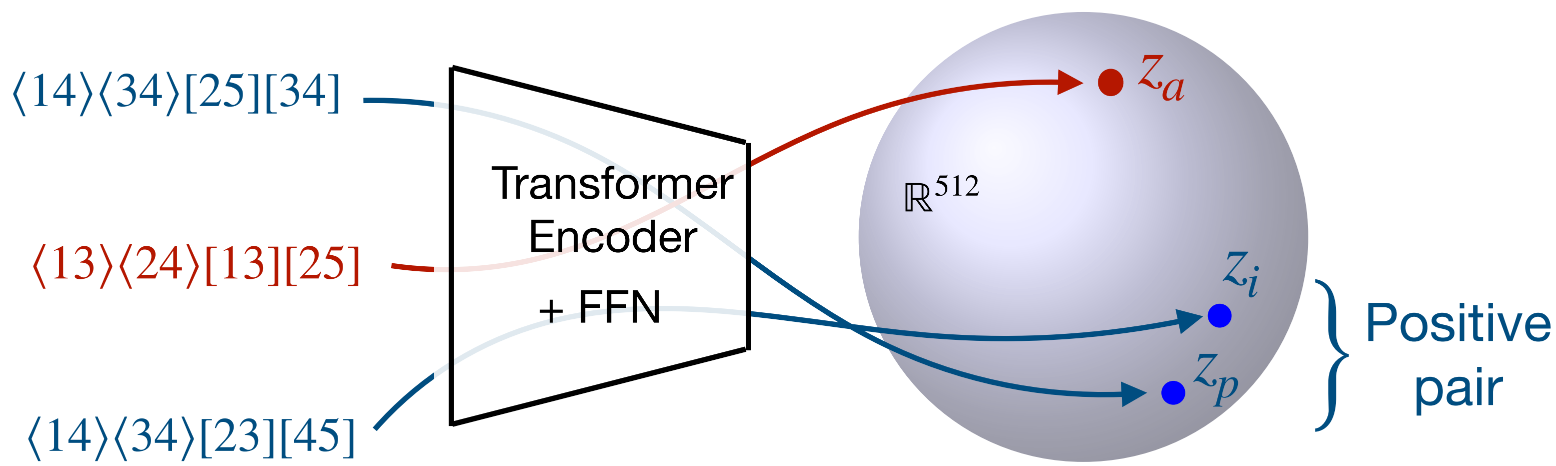


Use supervised contrastive learning !

Supervised Contrastive Learning

<u>Goal</u>: Pull together *<u>similar</u>* terms in embedding space

Terms appearing in same identity



<u>Contrastive Loss</u>

Promote alignement

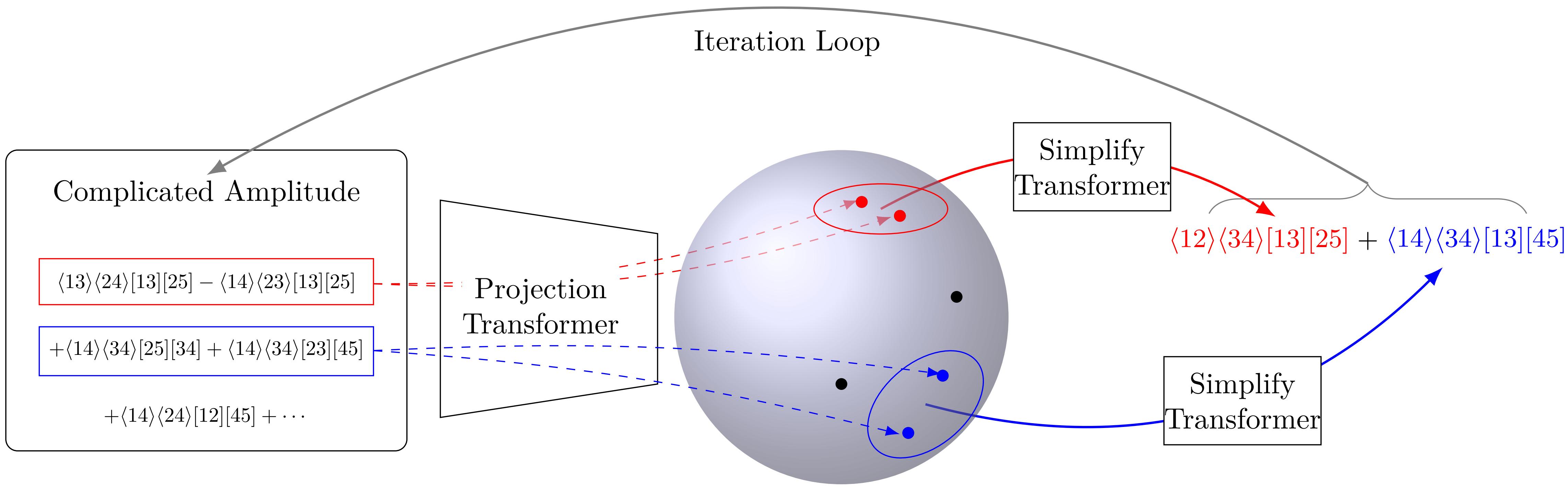
$$\mathcal{L}_{\text{con}} = -\frac{1}{|I|}\sum_{i\in I}\frac{1}{|P(i)|}\sum_{p\in P(i)}\log\frac{\exp(s(\boldsymbol{z}_i,\boldsymbol{z}_p)/\tau)}{\sum_{a\in A(i)}\exp(s(\boldsymbol{z}_i,\boldsymbol{z}_a)/\tau)}$$

Encourage uniformity

Cosine similarity : $s(\boldsymbol{z}_1,\boldsymbol{z}_2) = \frac{\boldsymbol{z}_1\cdot\boldsymbol{z}_2}{||\boldsymbol{z}_1||\,||\boldsymbol{z}_2||}$

Complete Simplification Pipeline: our Oracle

We group terms that have a high cosine similarity in the embedding space

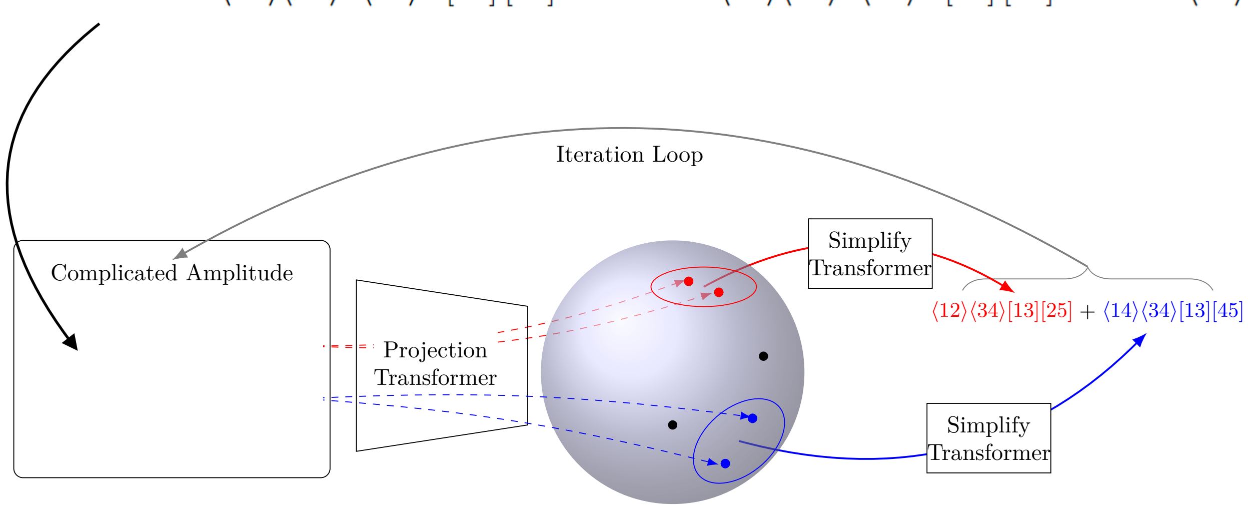


Complete ML pipeline deals with parts of the full amplitude at a time

A drastic simplification: $\mathscr{M}(\phi\phi\phi h^+ h^+)$

$$
\begin{aligned}
\mathcal{M} = {} & \frac{\langle 12\rangle\langle 13\rangle^2\langle 15\rangle\langle 25\rangle\langle 34\rangle\,[12]\,[15]^2\,[24]\,[34]}{\langle 14\rangle\langle 23\rangle\langle 35\rangle^3\langle 45\rangle^3\,[23]\,[35]\,[45]} + \frac{\langle 12\rangle\langle 13\rangle^2\langle 15\rangle\langle 34\rangle\,[12]\,[15]^2\,[34]^2}{\langle 14\rangle\langle 23\rangle\langle 35\rangle^2\langle 45\rangle^3\,[23]\,[35]\,[45]} + \frac{\langle 12\rangle\langle 13\rangle^2\,[12]\,[15]\,[34]}{\langle 14\rangle\langle 35\rangle^2\langle 45\rangle^2\,[35]} \\
& + \frac{\langle 12\rangle\langle 13\rangle^2\langle 25\rangle\langle 34\rangle\,[12]\,[15]\,[24]\,[34]}{\langle 14\rangle\langle 23\rangle\langle 35\rangle^3\langle 45\rangle^2\,[23]\,[35]} + \frac{\langle 12\rangle\langle 13\rangle^2\langle 34\rangle\,[12]\,[15]\,[34]^2}{\langle 14\rangle\langle 23\rangle\langle 35\rangle^2\langle 45\rangle^2\,[23]\,[35]} - \frac{\langle 12\rangle\langle 13\rangle\,[12]\,[15]\,[24]\,[34]}{\langle 35\rangle\langle 45\rangle^3\,[23]\,[45]} - \frac{\langle 12\rangle\langle 13\rangle\,[12]\,[15]\,[34]^2}{\langle 25\rangle\langle 45\rangle^3\,[23]\,[45]} \\
& - \frac{\langle 12\rangle\langle 13\rangle\langle 23\rangle\,[12]\,[34]}{\langle 15\rangle\langle 25\rangle\langle 34\rangle\langle 45\rangle^2} - \frac{\langle 12\rangle\langle 13\rangle\,[12]\,[24]\,[34]}{\langle 15\rangle\langle 35\rangle\langle 45\rangle^2\,[23]} - \frac{\langle 12\rangle\langle 13\rangle\,[12]\,[34]^2}{\langle 15\rangle\langle 25\rangle\langle 45\rangle^2\,[23]} - \frac{\langle 12\rangle\langle 13\rangle\langle 15\rangle^2\langle 34\rangle\,[12]\,[14]\,[15]\,[25]\,[34]}{\langle 14\rangle\langle 35\rangle^3\langle 45\rangle^3\,[23]\,[35]\,[45]} \\
& - \frac{\langle 12\rangle\langle 13\rangle\langle 15\rangle^2\langle 34\rangle\,[12]\,[15]^2\,[34]}{\langle 14\rangle\langle 23\rangle\langle 35\rangle^2\langle 45\rangle^3\,[23]\,[35]} + \frac{\langle 12\rangle\langle 13\rangle\langle 15\rangle\langle 25\rangle\langle 34\rangle\,[12]\,[15]\,[24]\,[25]\,[34]}{\langle 14\rangle\langle 35\rangle^3\langle 45\rangle^3\,[23]\,[35]\,[45]} + \frac{\langle 12\rangle\langle 13\rangle\langle 15\rangle\langle 34\rangle\,[12]\,[15]\,[34]}{\langle 14\rangle\langle 35\rangle^2\langle 45\rangle^3\,[35]} \\
& + \frac{\langle 12\rangle\langle 13\rangle\langle 15\rangle\langle 34\rangle\,[12]\,[15]\,[24]\,[34]}{\langle 14\rangle\langle 35\rangle^2\langle 45\rangle^3\,[23]\,[45]} + \frac{\langle 12\rangle\langle 13\rangle\langle 15\rangle\langle 34\rangle\,[12]\,[15]\,[25]\,[34]^2}{\langle 14\rangle\langle 35\rangle^2\langle 45\rangle^3\,[23]\,[35]\,[45]} + \frac{\langle 12\rangle\langle 13\rangle\langle 15\rangle\langle 34\rangle\,[12]\,[15]\,[34]^2}{\langle 14\rangle\langle 25\rangle\langle 35\rangle\langle 45\rangle^3\,[23]\,[45]} \\
& - \frac{\langle 12\rangle\langle 13\rangle\langle 15\rangle\langle 25\rangle\langle 34\rangle^2\,[12]\,[15]\,[24]\,[34]}{\langle 14\rangle\langle 23\rangle\langle 35\rangle^3\langle 45\rangle^3\,[23]\,[35]} - \frac{\langle 12\rangle\langle 13\rangle\langle 15\rangle\langle 34\rangle^2\,[12]\,[15]\,[34]^2}{\langle 14\rangle\langle 23\rangle\langle 35\rangle^2\langle 45\rangle^3\,[23]\,[35]} - \frac{\langle 12\rangle\langle 13\rangle\langle 15\rangle\langle 34\rangle\,[12]\,[15]\,[34]\,[45]}{\langle 14\rangle\langle 23\rangle\langle 35\rangle^2\langle 45\rangle^2\,[23]\,[35]} \\
& + \frac{\langle 12\rangle\langle 13\rangle\langle 23\rangle\,[12]\,[25]\,[34]}{\langle 14\rangle\langle 35\rangle^2\langle 45\rangle^2\,[35]} + \frac{\langle 12\rangle\langle 13\rangle\langle 23\rangle\,[12]\,[34]}{\langle 14\rangle\langle 25\rangle\langle 35\rangle\langle 45\rangle^2} + \frac{\langle 12\rangle\langle 13\rangle\langle 25\rangle\langle 34\rangle\,[12]\,[24]\,[25]\,[34]}{\langle 14\rangle\langle 35\rangle^3\langle 45\rangle^2\,[23]\,[35]} \\
& - \frac{\langle 12\rangle\langle 13\rangle\langle 34\rangle\,[12]\,[34]\,[45]}{\langle 14\rangle\langle 35\rangle^2\langle 45\rangle^2\,[35]} + \frac{\langle 12\rangle\langle 13\rangle\langle 34\rangle\,[12]\,[24]\,[34]}{\langle 14\rangle\langle 35\rangle^2\langle 45\rangle^2\,[23]} + \frac{\langle 12\rangle\langle 13\rangle\langle 34\rangle\,[12]\,[25]\,[34]^2}{\langle 14\rangle\langle 35\rangle^2\langle 45\rangle^2\,[23]\,[35]} + \frac{\langle 12\rangle\langle 13\rangle\langle 34\rangle\,[12]\,[34]^2}{\langle 14\rangle\langle 25\rangle\langle 35\rangle\langle 45\rangle^2\,[23]} \\
& - \frac{\langle 12\rangle\langle 13\rangle\langle 25\rangle\langle 34\rangle^2\,[12]\,[24]\,[34]\,[45]}{\langle 14\rangle\langle 23\rangle\langle 35\rangle^3\langle 45\rangle^2\,[23]\,[35]} - \frac{\langle 12\rangle\langle 13\rangle\langle 34\rangle^2\,[12]\,[34]^2\,[45]}{\langle 14\rangle\langle 23\rangle\langle 35\rangle^2\langle 45\rangle^2\,[23]\,[35]} + \frac{\langle 12\rangle\langle 15\rangle\langle 23\rangle\,[12]\,[14]\,[25]\,[34]}{\langle 25\rangle\langle 35\rangle\langle 45\rangle^3\,[23]\,[45]} + \frac{\langle 12\rangle\langle 15\rangle\,[12]\,[15]\,[34]}{\langle 25\rangle\langle 45\rangle^3\,[23]} \\
& - \frac{\langle 12\rangle\langle 23\rangle\,[12]\,[34]}{\langle 25\rangle\langle 45\rangle^3} + \frac{\langle 12\rangle\,[12]\,[34]\,[45]}{\langle 25\rangle\langle 45\rangle^2\,[23]} - \frac{\langle 12\rangle\langle 15\rangle^2\langle 23\rangle\langle 34\rangle\,[12]\,[14]\,[25]^2\,[34]}{\langle 14\rangle\langle 35\rangle^3\langle 45\rangle^3\,[23]\,[35]\,[45]} - \frac{\langle 12\rangle\langle 15\rangle^2\langle 23\rangle\langle 34\rangle\,[12]\,[14]\,[25]\,[34]}{\langle 14\rangle\langle 25\rangle\langle 35\rangle^2\langle 45\rangle^3\,[23]\,[45]} \\
& + \frac{\langle 12\rangle\langle 15\rangle^2\langle 34\rangle^2\,[12]\,[14]\,[25]\,[34]}{\langle 14\rangle\langle 35\rangle^3\langle 45\rangle^3\,[23]\,[35]} - \frac{\langle 12\rangle\langle 15\rangle^2\langle 34\rangle\,[12]\,[15]\,[25]\,[34]}{\langle 14\rangle\langle 35\rangle^2\langle 45\rangle^3\,[23]\,[35]} - \frac{\langle 12\rangle\langle 15\rangle^2\langle 34\rangle\,[12]\,[15]\,[34]}{\langle 14\rangle\langle 25\rangle\langle 35\rangle\langle 45\rangle^3\,[23]} + \cdots
\end{aligned}
$$

298 terms !



$$\longrightarrow \quad \overline{\mathcal{M}} = \frac{\langle 12\rangle\langle 13\rangle\langle 23\rangle}{\langle 24\rangle\langle 25\rangle\langle 45\rangle}\left(\frac{[14][35]}{\langle 14\rangle\langle 35\rangle} - \frac{[15][34]}{\langle 15\rangle\langle 34\rangle}\right)$$

2 terms !

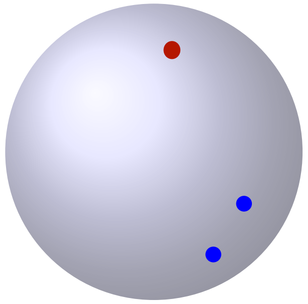
Lessons and Outlook

1) ML is useful for mathematical tasks

+

→ we have numerical checks

2) Need meaningful representations



3) Framework is flexible

→ only need to adapt the training data

At scale, AI assistant for hep-th ?

Backup

Model Architectures and Training Parameters

Hyperparameter Type	Parameter Description	Value
Network architecture	Encoder layers	3
	Decoder layers	3
	Attention heads	8
	Embedding dimension	512
	Maximum input length	2560
Training parameters	Batch size	16
	Epoch size	50,000
	Epoch number	1,500
	Learning Rate	10^{-4}

Encoder-Decoder for simplification

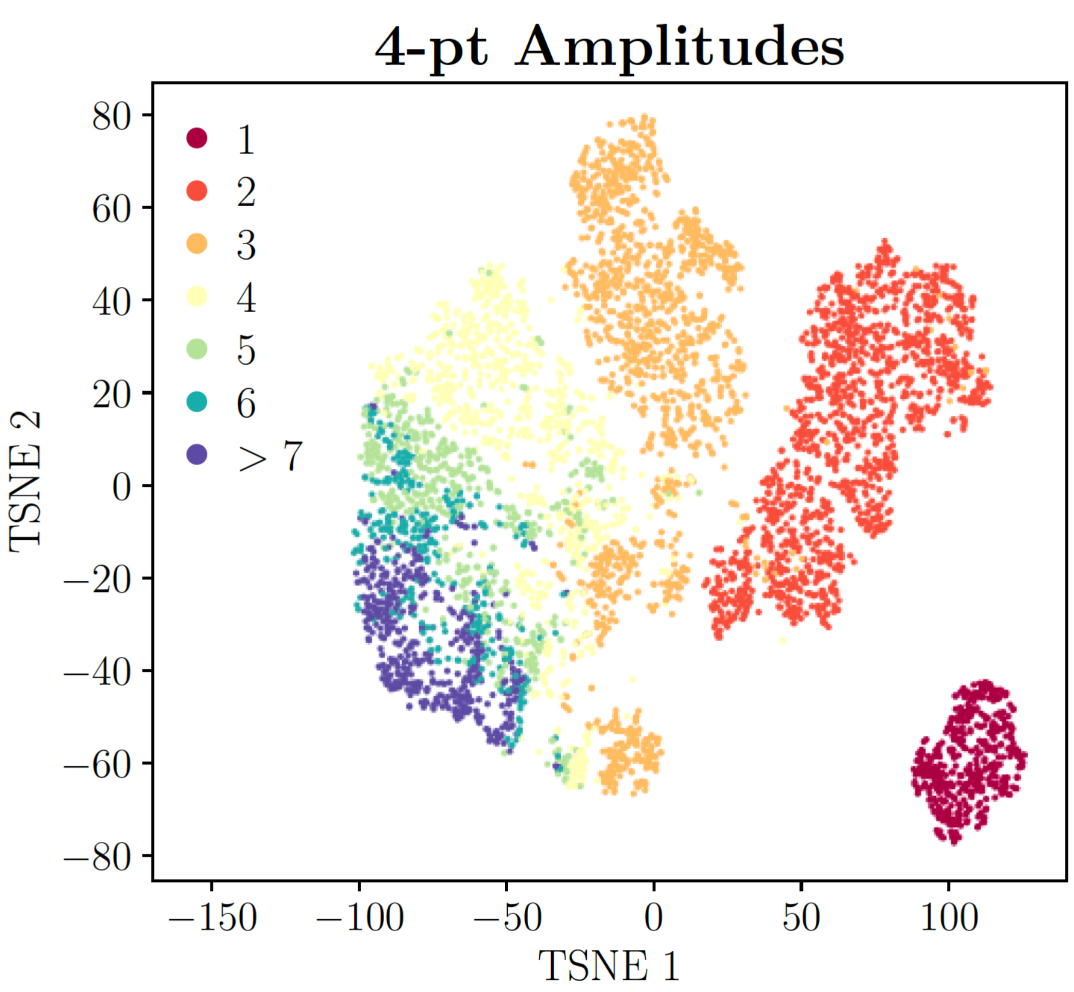
Hyperparameter Type	Parameter Description	Value
Encoder architecture	Num layers	2
	Attention heads	8
	Embedding dimension	512
	Maximum input length	256
Feed-forward architecture	Num layers	2
	Hidden dimension	512
Training parameters	Batch size	128
	Epoch size	10,000
	Epoch number	500
	Learning Rate	10^{-4}
	Temperature	0.15

Embedding network

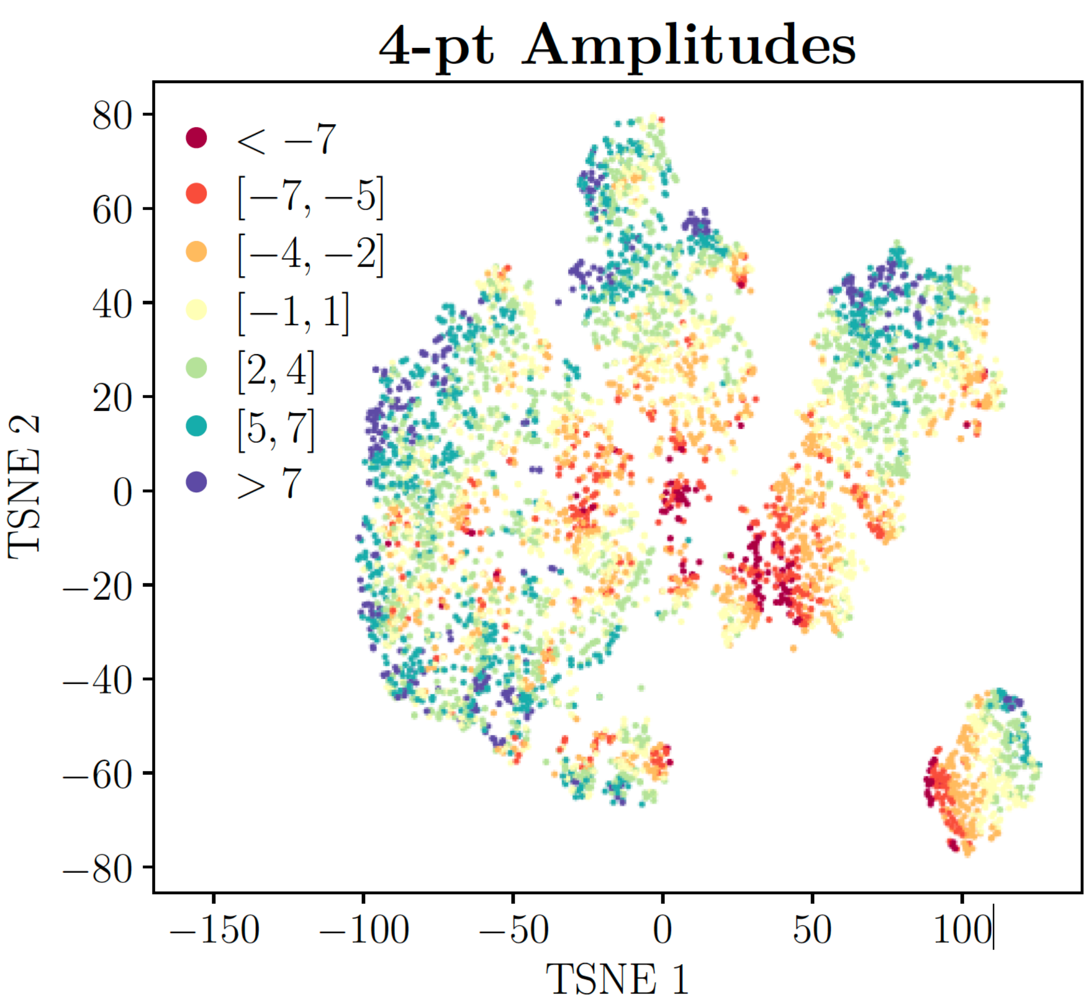
Interesting Representations: t-SNE visualisations

Amplitude embeddings

$$e(\mathcal{M}) = \frac{1}{|\mathcal{P}(\mathcal{M})|} \sum_{w \in \mathcal{P}(\mathcal{M})} \boldsymbol{E}(w|\mathcal{P}(\mathcal{M}))$$



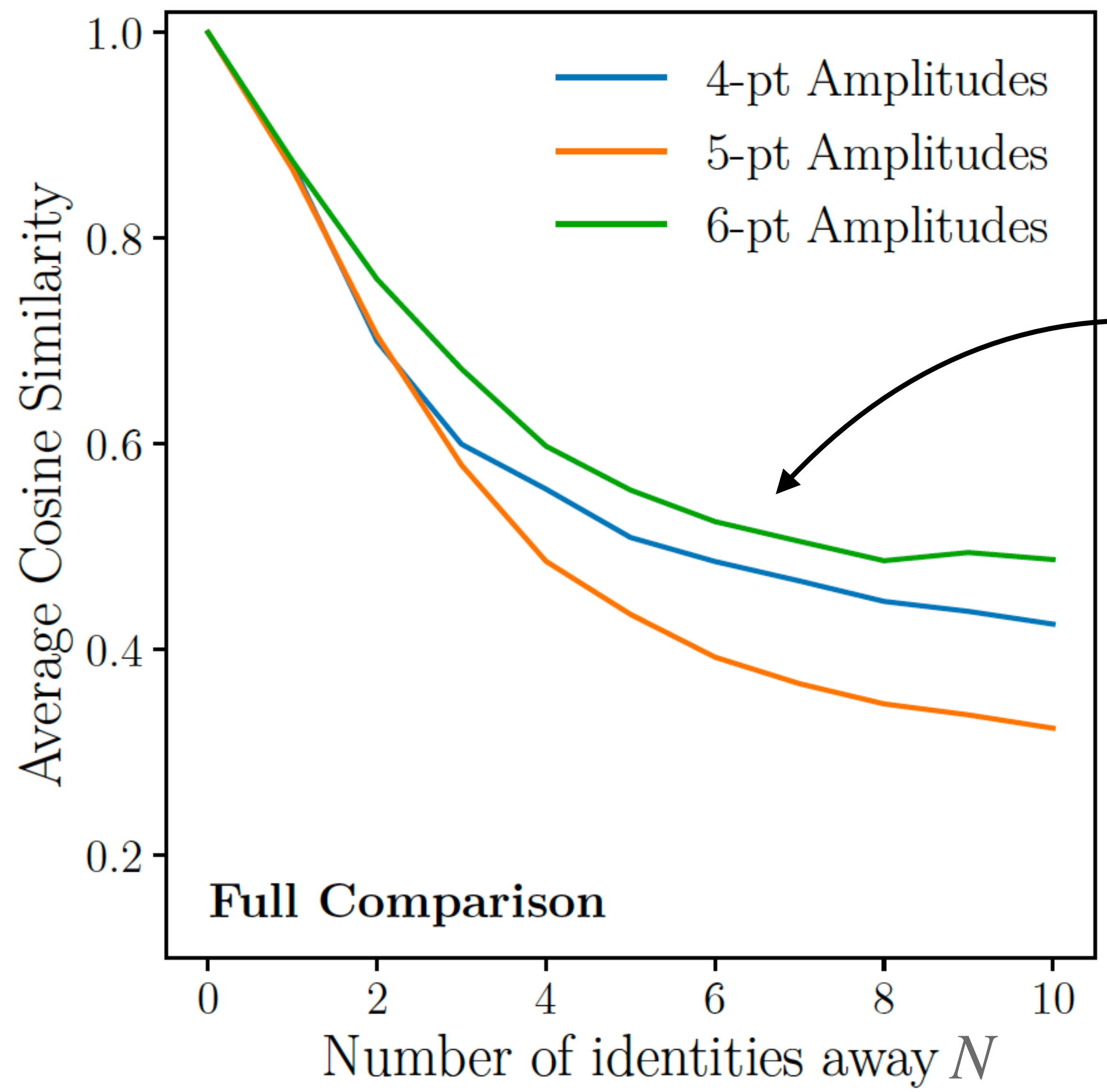
Colour-coded by number of numerators terms



Colour-coded by mass dimension

Similarity Measure

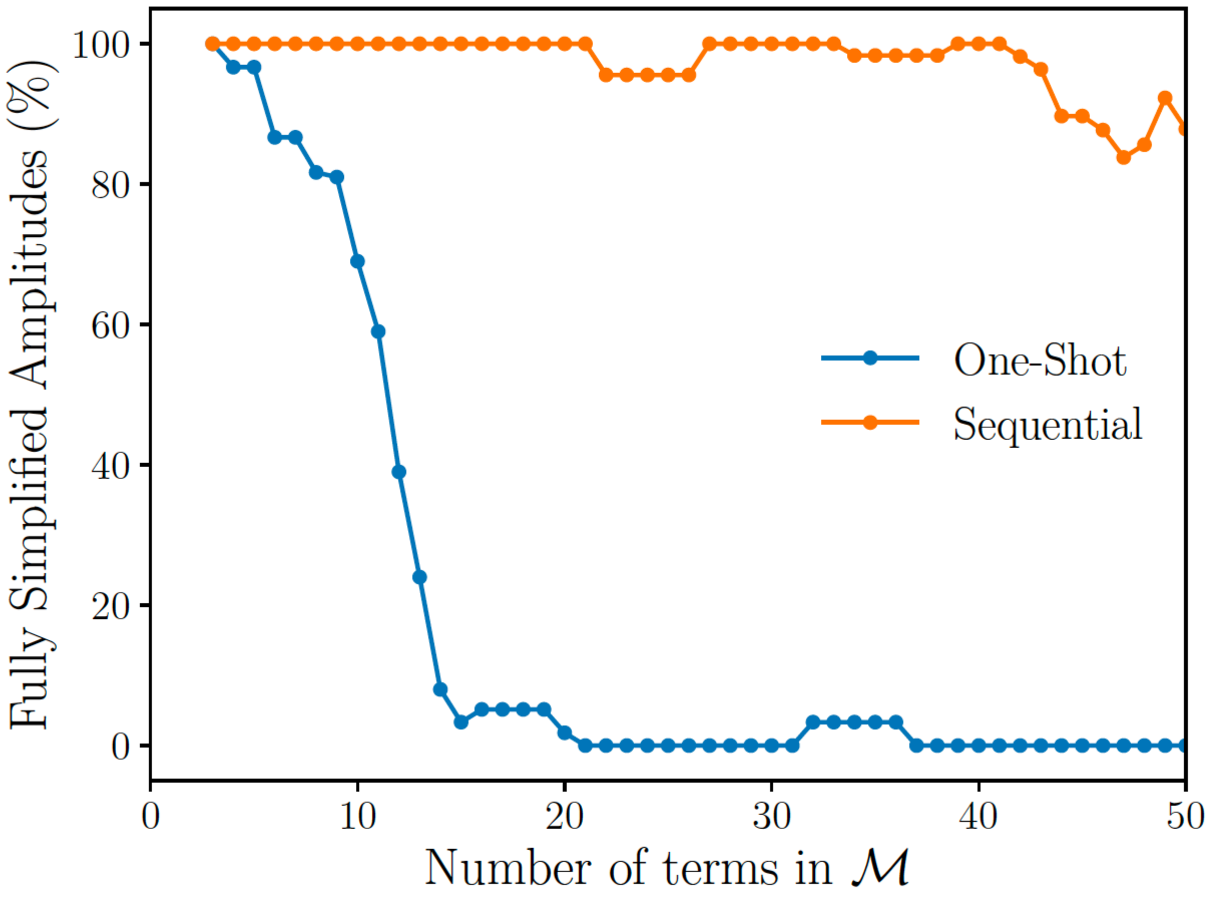
Average cosine similarly between terms that are N identities away

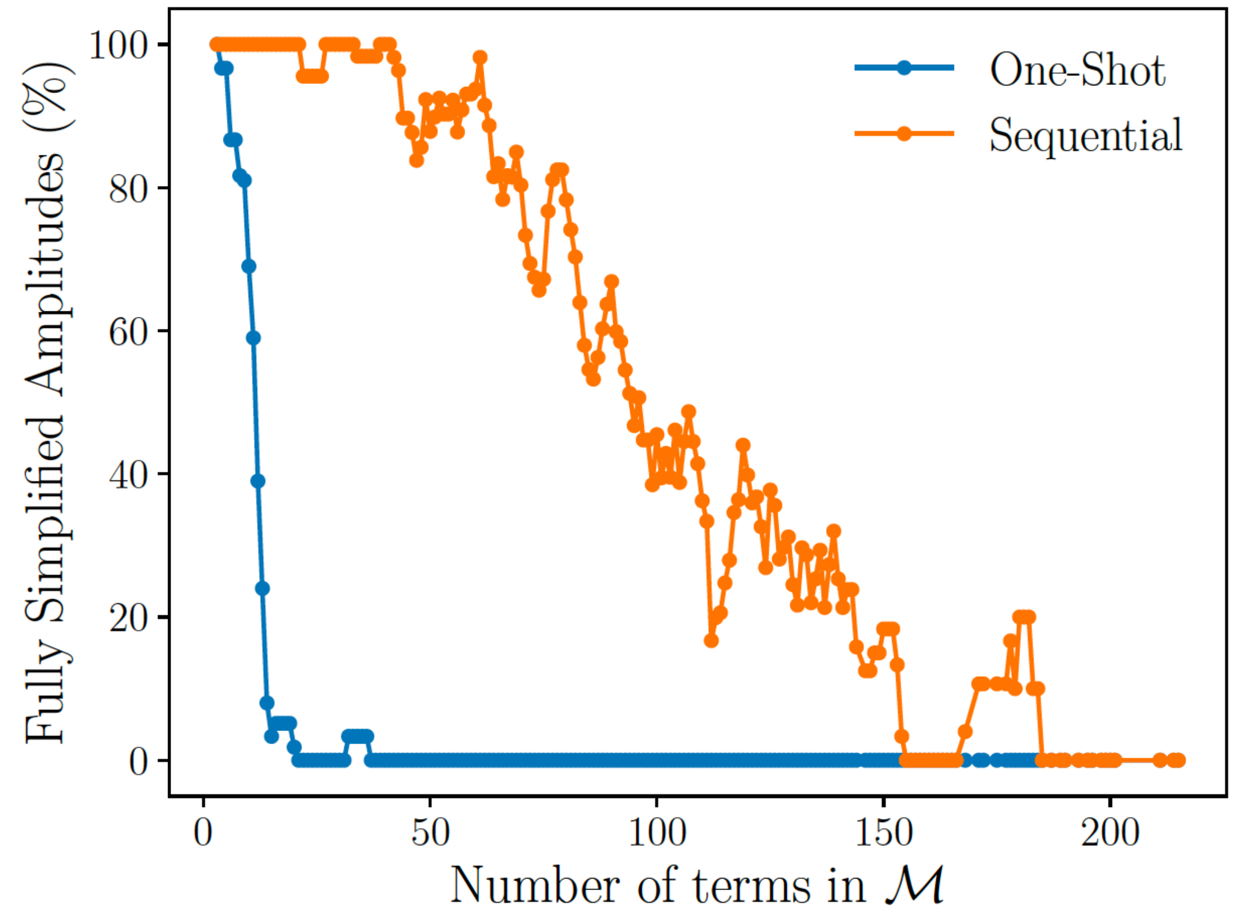


Terms in the embedding space are organised by their similarity !

Simplification results

How often do we fully simplify 4 and 5-pt amplitudes down to a single term following Parke-Taylor ?





Transformer Encoder alone handles 20 terms at most

Full pipeline almost flawless till 50 terms !

Even when fails still reduces to ~13 % of original size