

The Past, Present and Future of Machine Learning in Particle Physics

Matthew Schwartz

Harvard University

May 22, 2023



The past



He-Man at a roller rink in the 80's (Nov 2022)

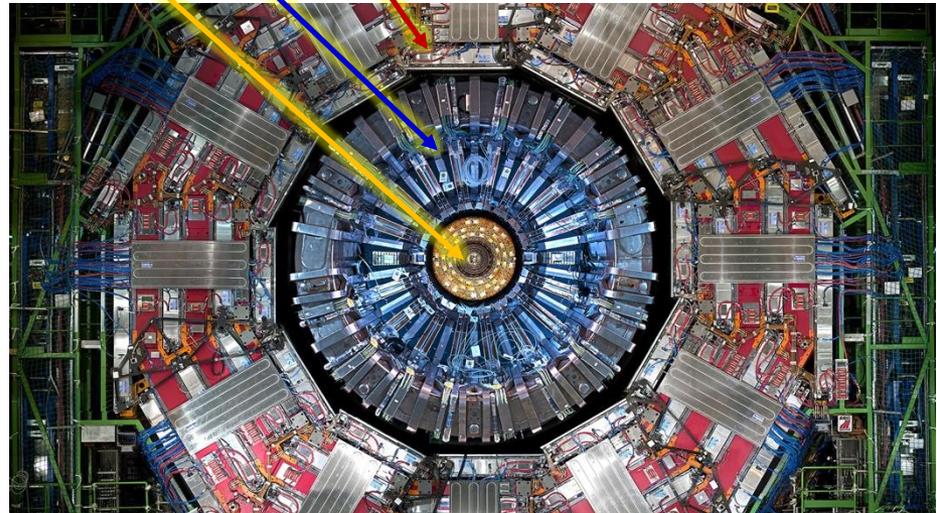
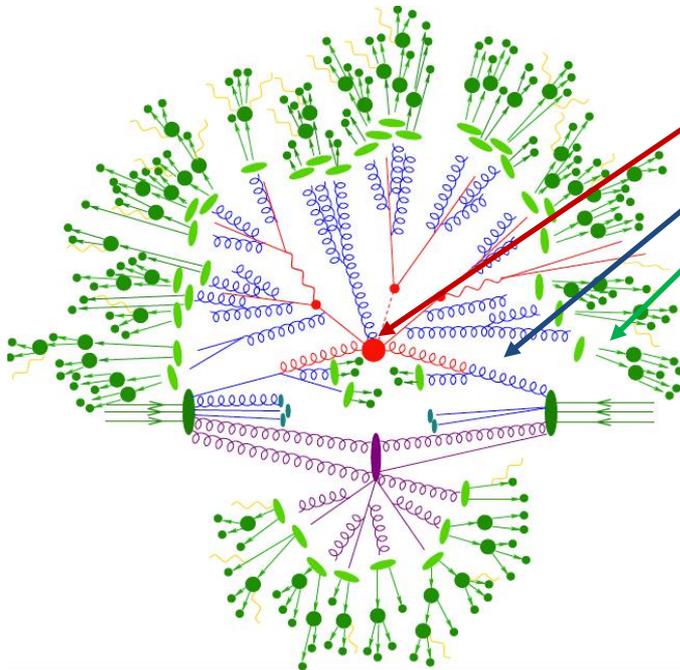
How is ML used in particle physics?

Main use so far is in **collider physics**

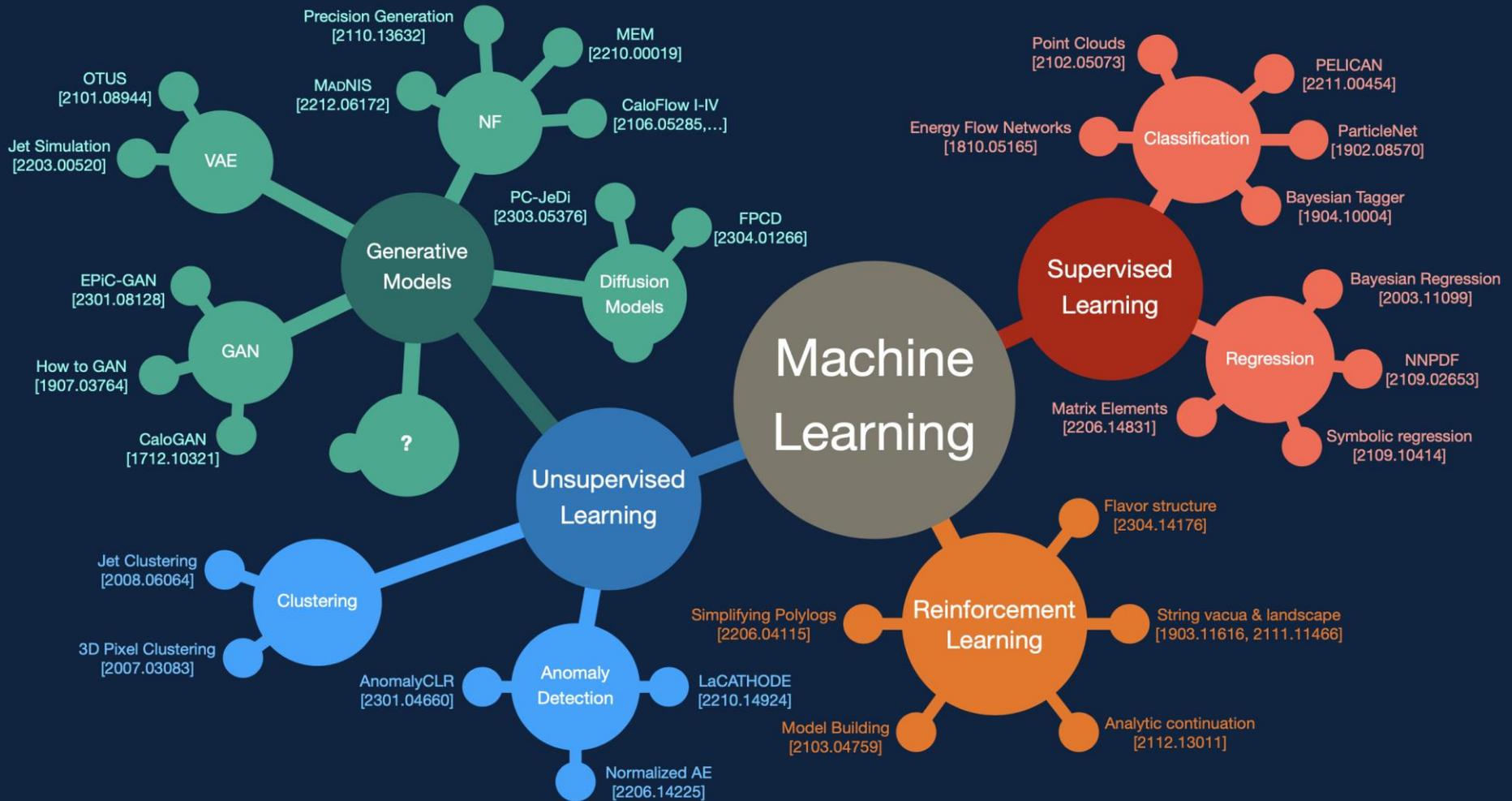
- Can generate billions of simulated events on a laptop
- Simulations accurate over **20 orders of magnitude!**
- Unheard of in any other area of science

Simulation pipeline:

- **Perturbative QCD: $10^{-19}\text{m} - 10^{-16}\text{m}$**
- **Parton shower: $10^{-16}\text{m} - 10^{-14}\text{m}$**
- **Hadronization/fragmentation $10^{-14}\text{m} - 10^{-12}\text{m}$**
- **tracker/ecal/hcal simulation $10^{-12}\text{m} - 10^2\text{m}$**



Most results heavily use these simulations



Current areas of progress

1. Lattice QCD

- Normalizing flows, Monte Carlo sampling, Spectral reconstruction, ...

2. Simulation/unfolding [see Krause's talk]

- Learn to reproduce simulations with a neural network
- Can speed up simulations by factors of $10^3 - 10^5$
- Can be used for unfolding: remove effects of simulation on data

3. Anomaly detection

- Search for deviations from background
- No signal hypothesis necessary (?)

4. Data representation

- Can ML provide a better way to categorize and understand data?
- e.g. optimal transport, graph networks, etc.

5. Classification

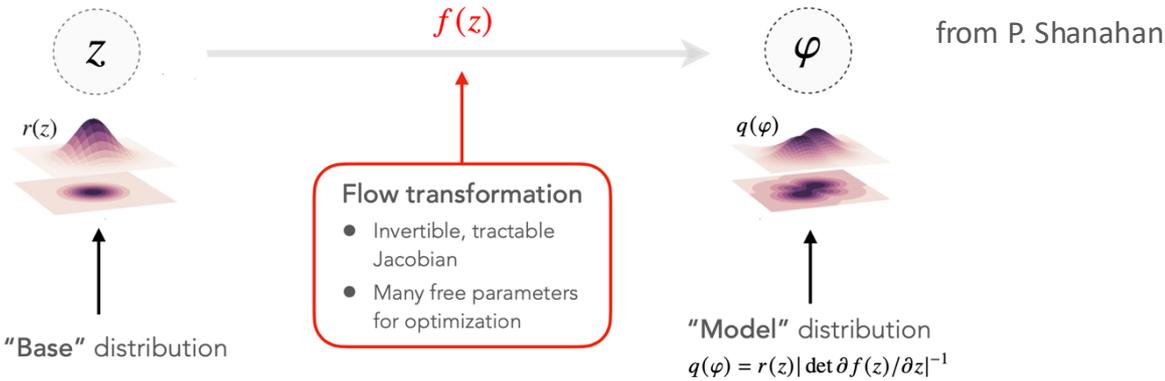
- Top tagging, W tagging, Q v G discrimination, new physics searches

6. Symbolic regression

- Large language models (chatGPT) ?

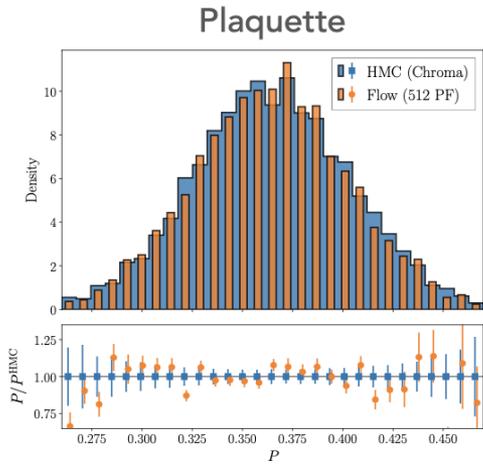
Lattice QCD

Normalizing flows: invertible maps from Gaussians to complicated physical distributions

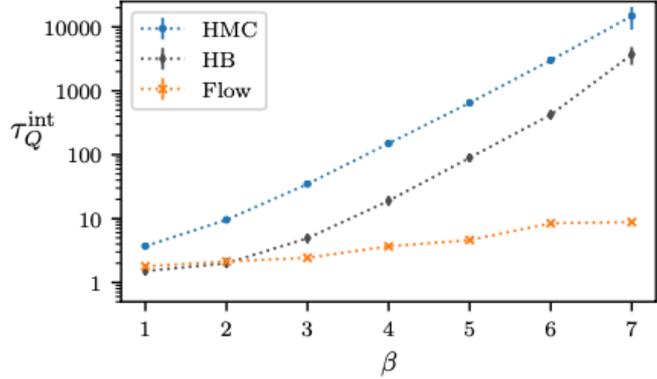


Boyda et al., 2008.05456
Abbott et al., 2207.08945

Precise agreement with Hybrid Monte Carlo (HMC)



Significantly improved speed

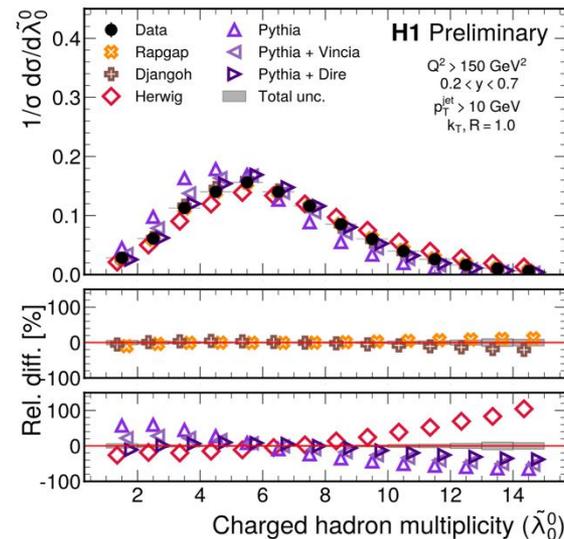
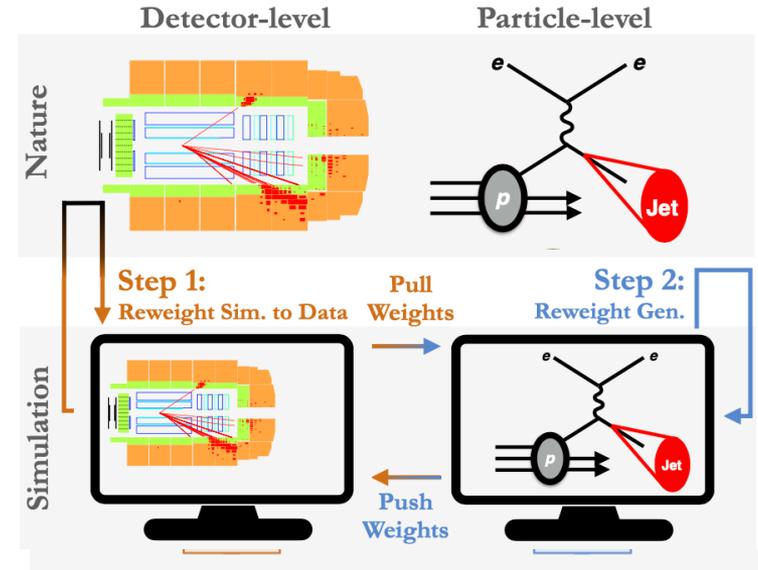
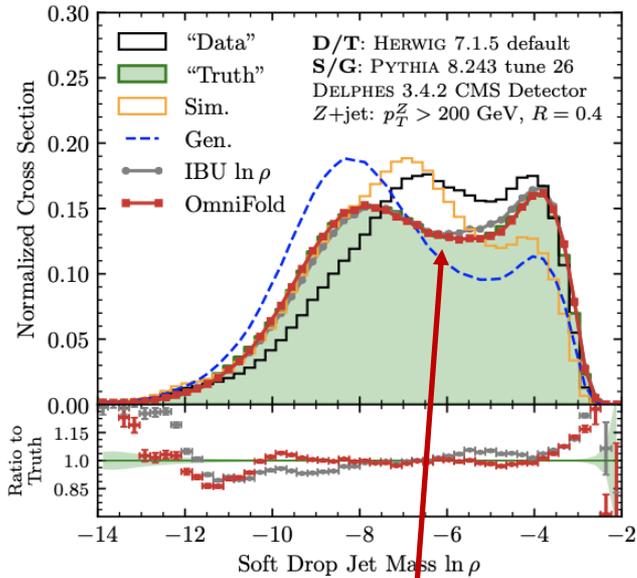


Simulation/Unfolding

OmniFold

Andreassen et al. (arXiv:1911.09107)

- Uses ML to learn mapping from generator to detector
- Can then unfold any distribution to truth level
- Previous unfolding techniques are observable-by-observable



Used on H1 data (HERA)

H1prelim-22-034

- Tested in simulation:
 - Herwig as “truth” + delphes -> “data”
 - pythia as “sim” + delphes -> “gen”
 - Truth and omnifold agree

Anomaly detection

The Dream:

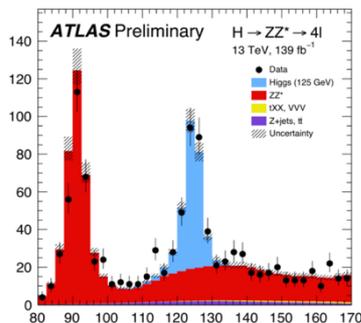
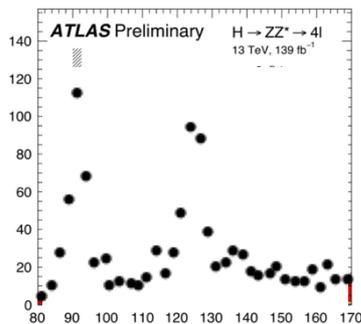
- ML sees something unusual in the data, new physics is found!
- **Unsupervised method:** do not need a signal hypothesis
- Way to find “unknown unknowns”

The main idea:

- **Background is understood well enough by ML that statistical outliers are seen**

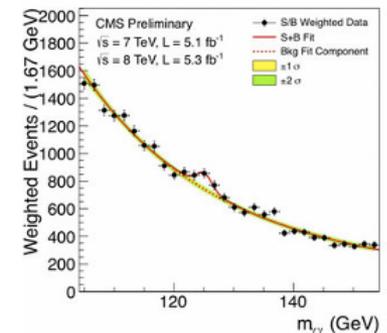
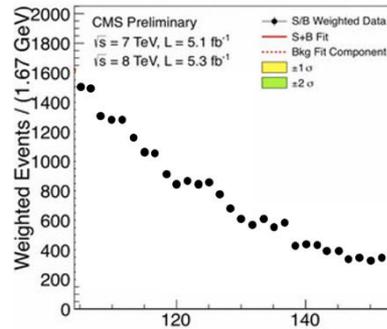
Easy: outliers

$$h \rightarrow e^- e^+ \mu^- \mu^+$$

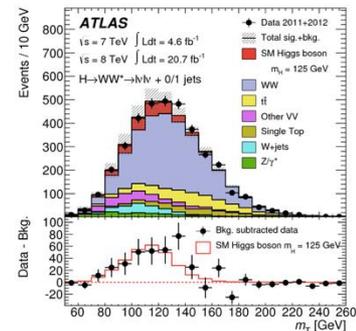
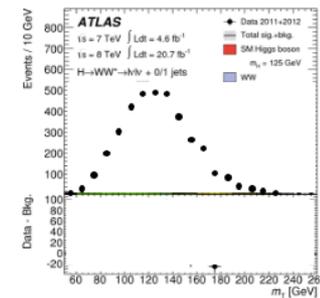


Hard: overdensities

$$h \rightarrow \gamma\gamma$$



$$h \rightarrow e^- \mu^+ \bar{\nu} \nu$$



- Collins et al: 1805.02664
- D'Angelo and Wulzer: 1806.02350
- Collins et al: 1902.02634
- D'Angelo et al: 1912.12155
- Nachman & Shih: 2001.04990
- Stein et al: 2012.11638
- Carron et al: 2106.10164
- Hallin et al: 2109.00546

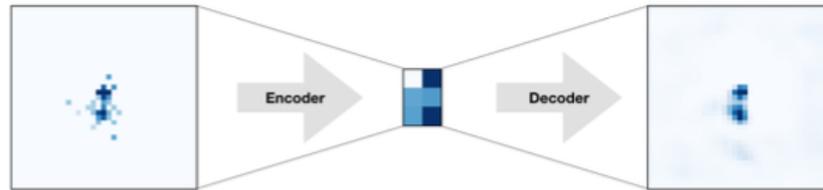
- Hajer et al: 1807.10261
- Heimel et al: 1808.08979
- Farina et al: 1808.08992
- Cerri et al: 1811.10276
- Roy + Vijay: 1903.02032,
- Atkinson et al: 2105.07988
- Carron et al: 2106.10164
- Ngairangbam et al: 2112.04958

...

Autoencoders

Variational autoencoders: [Farina et al: 1808.08992]

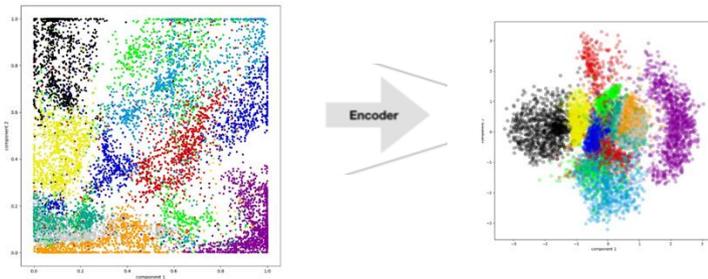
- Compress the background/data to a low-dimensional latent space



- Uncompress back to data space.
- Poorly reconstructed events are anomalies

Fraser, MDS, et al: 2110.06948

- Can look for anomalies directly in low-dimensional latent space



Metric	Number of medoids	Method	Top jet		W jet	
			AUC	$\epsilon_S(\epsilon_B = 0.1)$	AUC	$\epsilon_S(\epsilon_B = 0.1)$
Wass(1)	-	Avg	0.81	0.33	0.62	0.02
	1	Medoid	0.83	0.28	0.63	0.02
	3 (elbow)	Medoids (min)	0.85	0.43	0.67	0.04
	5	Medoids (min)	0.87	0.54	0.60	0.05
Wass(5)	7	Medoids (min)	0.87	0.54	0.61	0.05
	4 (elbow)	Medoids (min)	0.67	0.22	0.41	0.04
MAE	1	Medoid	0.82	0.40	0.71	0.07
	3 (elbow)	Medoids (min)	0.82	0.49	0.61	0.08

- Use k medoids or Wasserstein optimal transport metric
- Use event-to-ensemble distance for anomaly score
- Take home messages:
 - Performance depends on metric and sample
 - Cannot optimize in signal-independent manner

$$d_{Wass}^{(p)} = \left(\min_f \sum_{i,j} f_{ij} (c_{ij})^p \right)^{1/p}$$

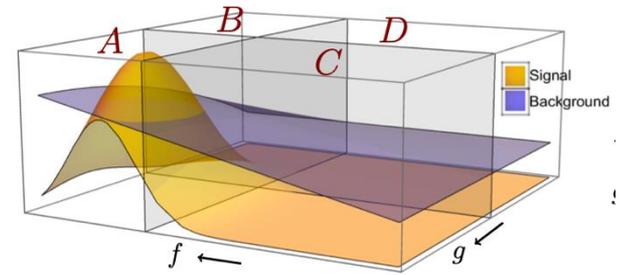
ABCDDisCo: ML the ABCD method

ABCD method:

- Standard experimental sideband technique
- Estimate background in region A via $N_A = \frac{N_B N_C}{N_D}$
- Requires two features f and g to be uncorrelated
 - E.g. f = mass and g = rapidity

Distance Correlation (DisCo): alternative to adversarial networks

- Decorrelates observables, easy to train



Kasieczka and Shih [arXiv: 2001.05310]

- Single DisCo



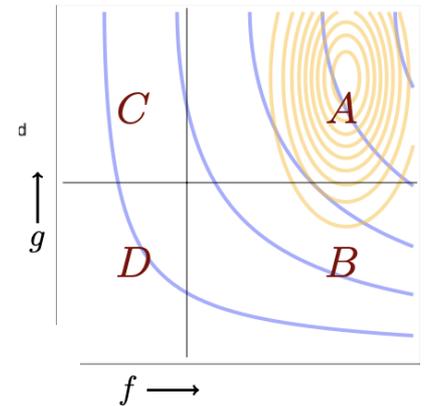
f is fixed (e.g. mass) g is learned

Kasieczka, Nachman, MDS, Shih [arXiv: 2007.14400]

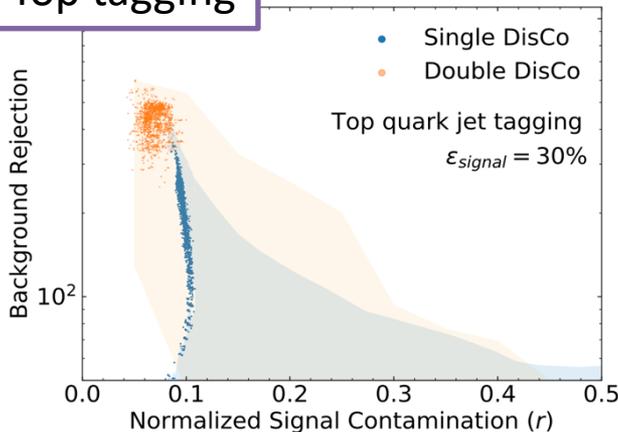
- Double DisCo



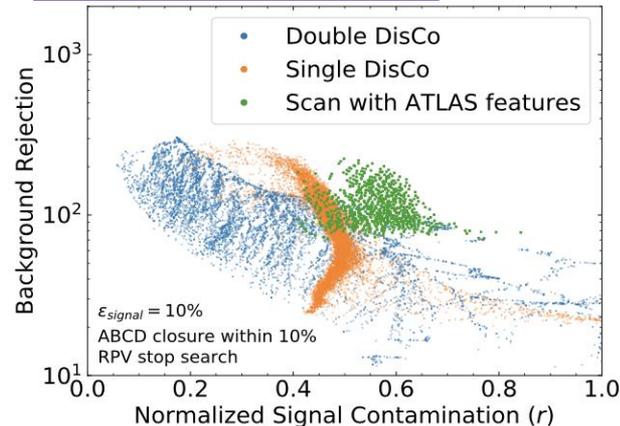
f and g are learned



Top tagging



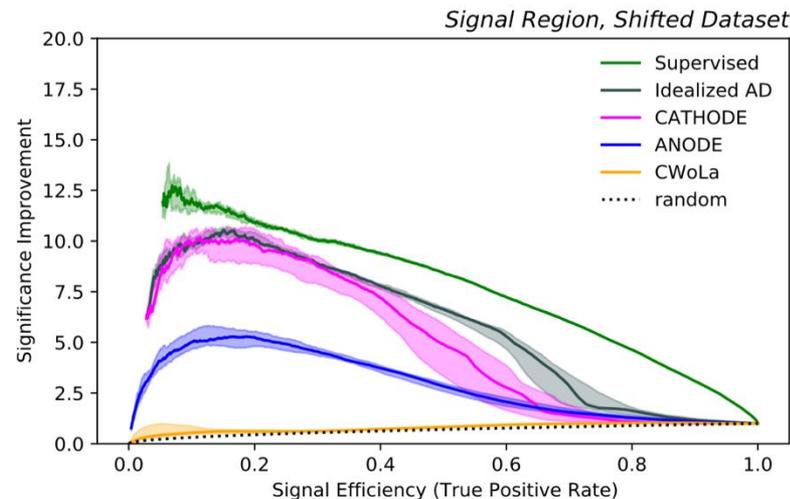
RPV squark search



Challenges for Anomaly Detection

- Background regions are **signal dependent**
 - e.g. if looking for a dijet resonance, need a dijet background
 - No such thing as a signal-independent background
- **Looking under the lampost**
 - We only know how to look for resonances/new particles
 - Different signals are too varied to be pooled together
- Very **sensitive to metric** for what is anomalous
 - Tails of backgrounds are unique in their own way
- Supervised classifiers always do better

[Hallin et al. 2109.00546]



Classification: top tagging

e.g. top tagging

Jet substructure approach (2008-2017):

- Think about physics
- Deconstruct jet
- Look for W within top jet
- Look at helicity angle
- Hopkins Top Tagger (2008)

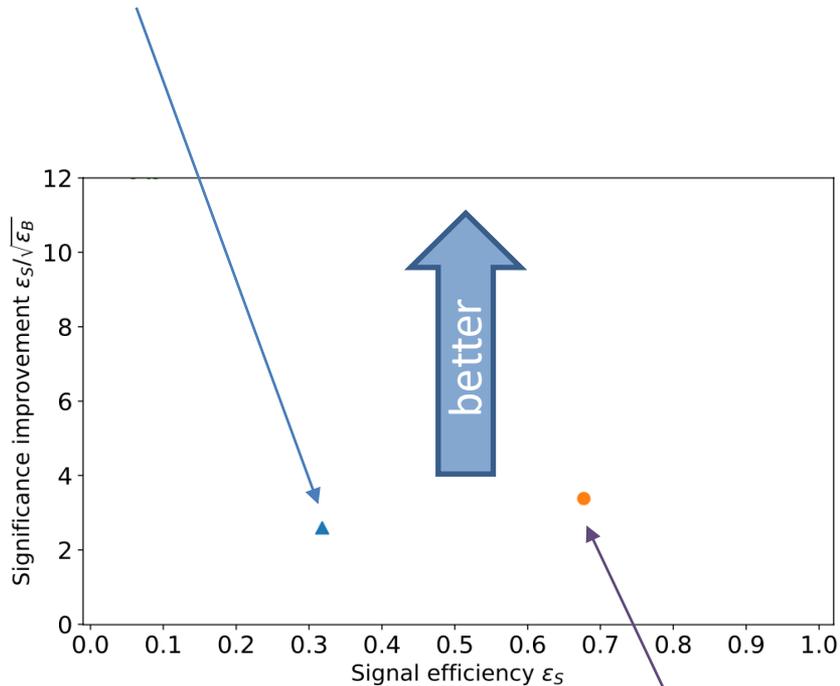
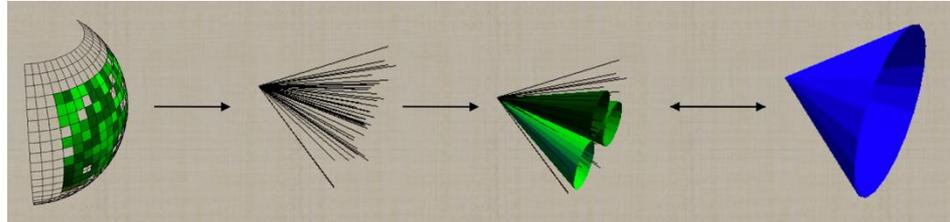


fig from 2103.12226

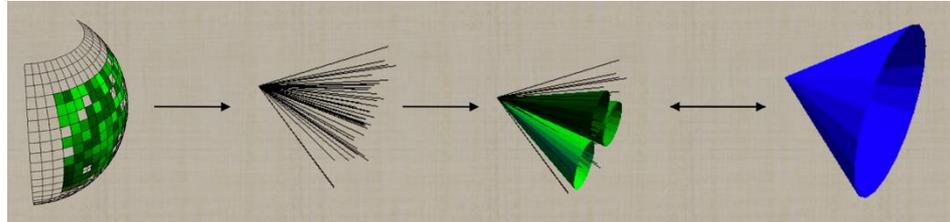
best jet substructure algorithm (2014)

Classification: top tagging

e.g. top tagging

Jet substructure approach (2008-2017):

- Think about physics
- Deconstruct jet
- Look for W within top jet
- Look at helicity angle
- Hopkins Top Tagger (2008)



Machine learning methods are much better

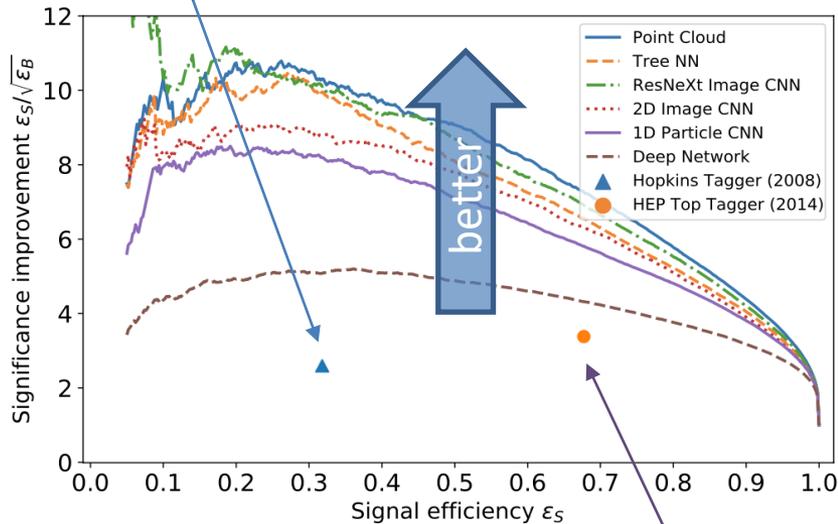
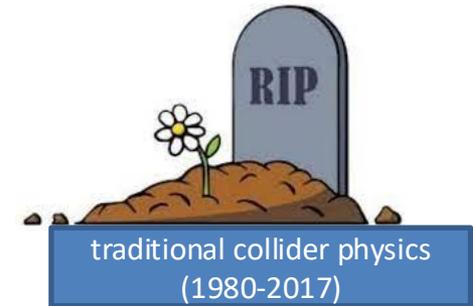


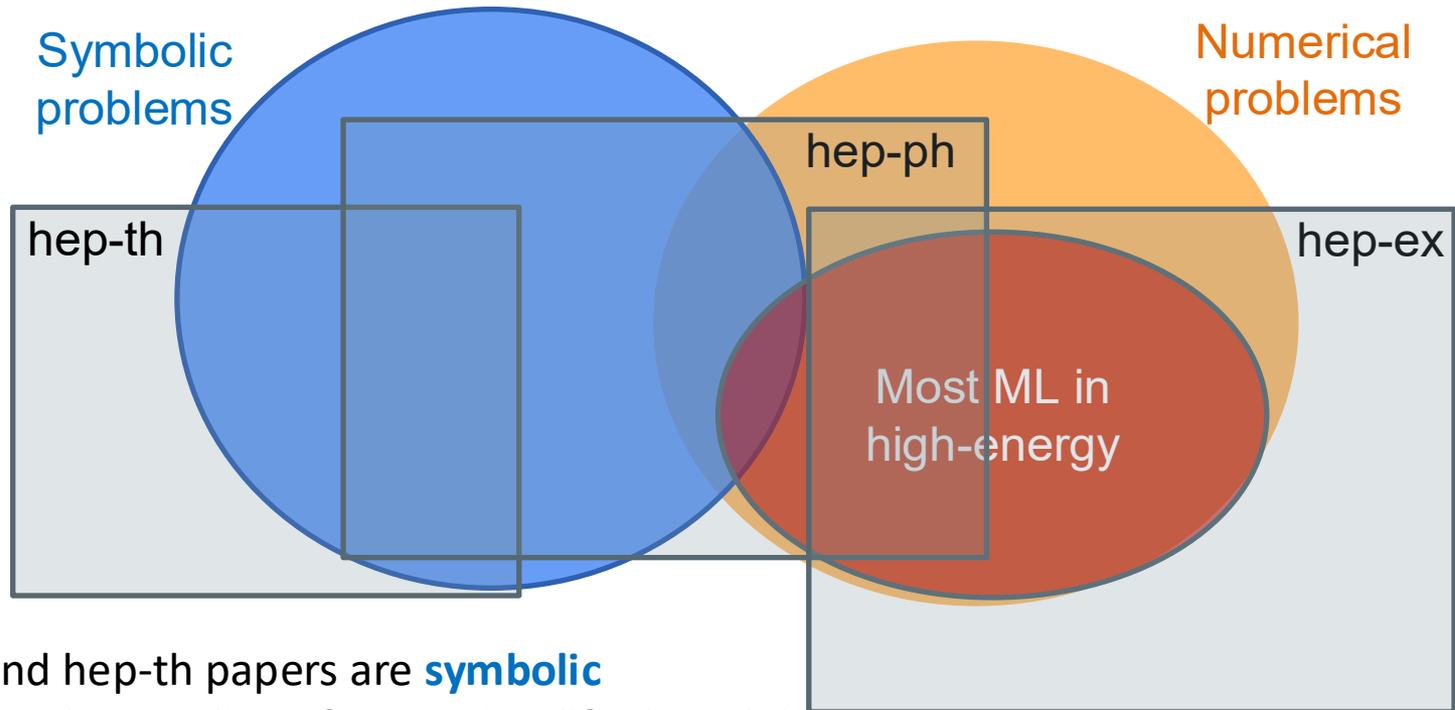
fig from 2103.12226

best jet substructure algorithm (2014)



- ML requires less “thinking”
- Provides less physical insight
- Better performance

What subfield will ML make obsolete next?



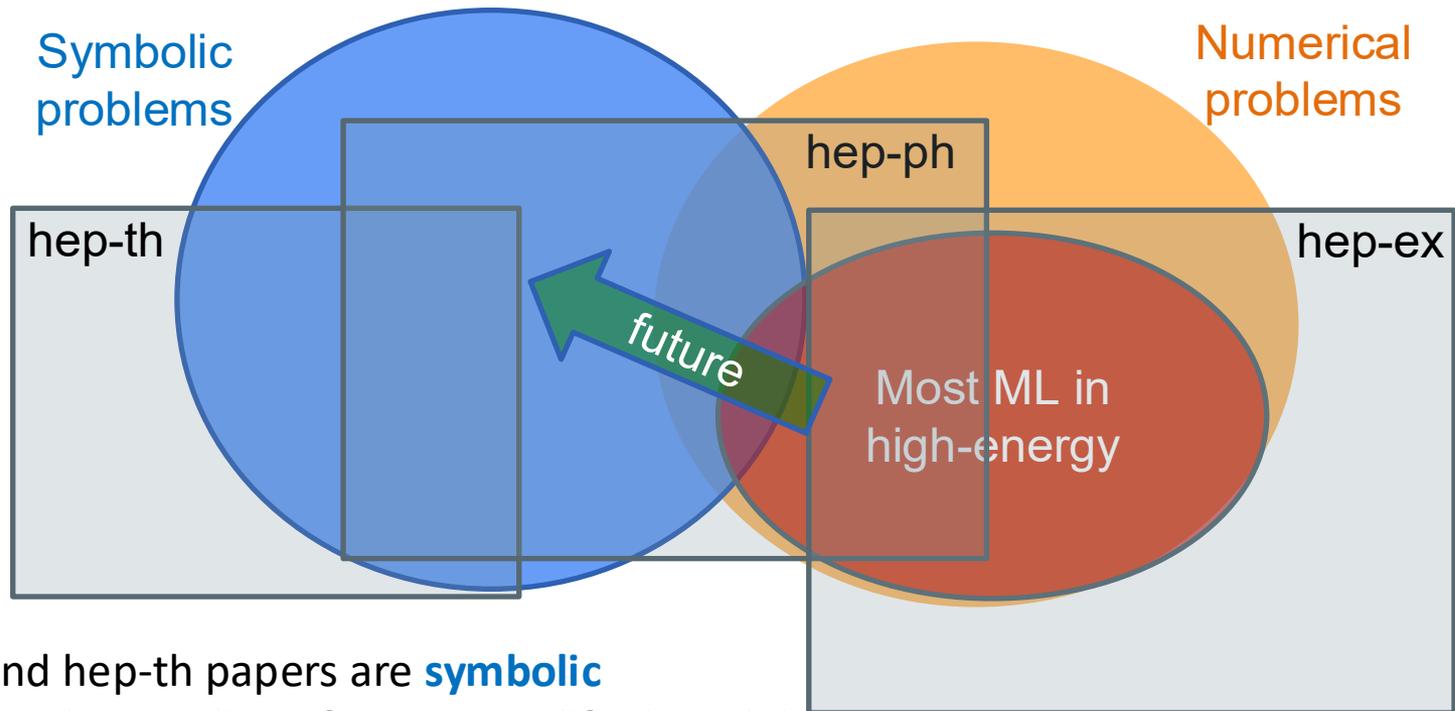
Most hep-ph and hep-th papers are **symbolic**

- Qualitative understanding of some simplified model
- Approximate but exact solutions to some equation
- Analytic computations in some system

So far, most ML in physics is highly **numerical**

- Collider physics data is millions of numbers
- Approximate answers are ok

What subfield will ML make obsolete next?



Most hep-ph and hep-th papers are **symbolic**

- Qualitative understanding of some simplified model
- Approximate but exact solutions to some equation
- Analytic computations in some system

So far, most ML in physics is highly **numerical**

- Collider physics data is millions of numbers
- Approximate answers are ok

Symbolic ML methods will be **essential** for the future of High Energy Physics

- The world is changing because of **symbolic** large language models

The present



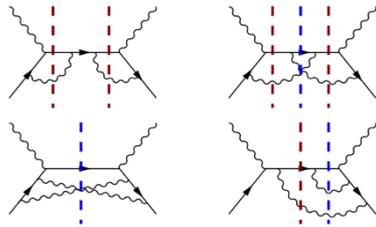
Einstein at the waterpark (April, 2023)

Example: simplifying polylogarithms

e.g. Compton scattering at NLO

Lee, Schwartz, Zhang *PRL* 126 (2021)

Want to compute some Feynman diagrams



Reduce to iterated integrals

$$I(s) = \int_1^s \frac{ds'}{\sqrt{(s'-1)(s'+3)}} \int_1^{s'} \frac{ds''}{s''}$$

Do integrals

$$I(s) = -\frac{1}{3}\text{Li}_2(-r^3) + \text{Li}_2(-r) - \frac{1}{2}\ln^2 r + \frac{\pi^2}{18}$$

simplify

$$\begin{aligned}
 I(s) &= \int_1^{r(s)} d \ln r' \int_1^{r'} d \ln \frac{r''^2 - r'' + 1}{r''} \\
 &= \int_1^{r(s)} \frac{dr'}{r'} \int_1^{r'} dr'' \left(-\frac{1}{r''} + \frac{1}{r'' - z_1} + \frac{1}{r'' - z_2} \right) \\
 &= \frac{2\pi^2}{9} - \frac{\ln^2 r}{2} + \left(-\frac{i\pi}{3} + \ln r \right) \ln(r - z_1) + \ln(r z_1) \ln(r - z_2) \\
 &\quad - \text{Li}_2(1 - z_1) + \text{Li}_2(1 - r z_1) - \text{Li}_2(1 - z_2) + \text{Li}_2(1 - r z_2)
 \end{aligned}$$

$z_1 = \frac{1+i\sqrt{3}}{2}$
 $z_2 = \frac{1-i\sqrt{3}}{2}$

- simplifying full result (1000 terms) took one month

result is complicated

Why is simplification important?

- Removes unphysical singularities
- A lot of physics in analytic structure of Feynman diagrams
- Simple form indicates deeper structure
- Simplification at intermediate steps make full calculation tractable
- In a sense, **all of science is simplification**



Chat GPT 3.5 (2022)

MA

what is dilog(1)?

In[1]:= PolyLog[2, 1]

$$\text{Out[1]} = \frac{\pi^2}{6}$$



The dilogarithm function, denoted $\text{Li}_2(x)$, is a complex function defined for all complex numbers x except for $x = 1$. The value of the dilogarithm at $x = 1$ is infinite, as the series expansion of the dilogarithm function diverges at $x = 1$.



Chat GPT 4 (2023)

MA

What is dilog(1)?



The dilogarithm function, also known as $\text{Li}_2(z)$ or the second-order polylogarithm, has a specific value when evaluated at $z = 1$:

$$\text{Li}_2(1) = \frac{\pi^2}{6}.$$

So, the value of the dilogarithm function at 1 is $\frac{\pi^2}{6}$, which is approximately 1.64493.

Problem statement

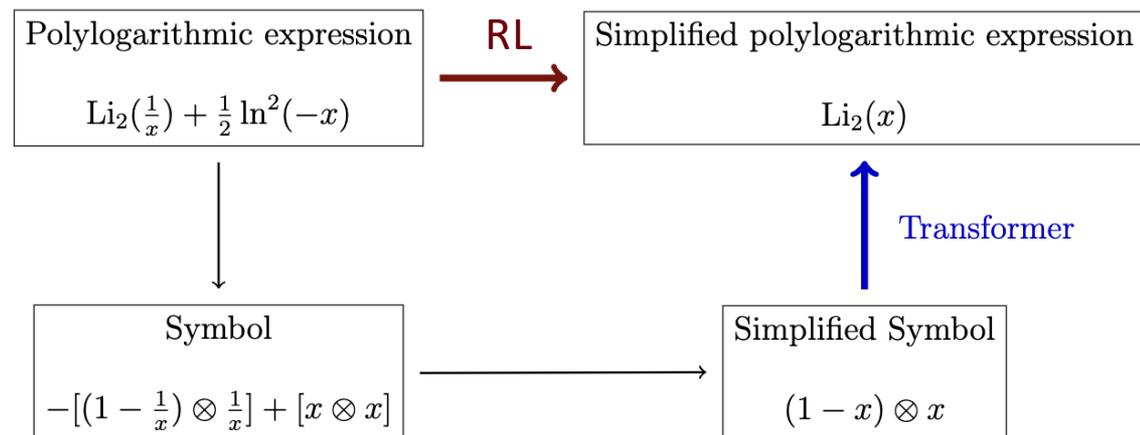
Given some polylogarithmic expression:

$$f(x) = 9 \left(-\text{Li}_3(x) - \text{Li}_3 \left(\frac{2ix}{-i + \sqrt{3}} \right) - \text{Li}_3 \left(-\frac{2ix}{i + \sqrt{3}} \right) \right) \\ + 4 \left(-\text{Li}_3(x) + \text{Li}_3 \left(\frac{x}{x+1} \right) + \text{Li}_3(x+1) - \text{Li}_2(-x) \ln(x+1) \right) \\ - 4 \left(\text{Li}_2(x+1) \ln(x+1) + \frac{1}{6} \ln^3(x+1) + \frac{1}{2} \ln(-x) \ln^2(x+1) \right)$$

1. What is its simplest form?
2. Does it simplify to zero?
3. What identities do we apply in what order to simplify it?

Two approaches

1. Reinforcement learning
2. Transformer networks

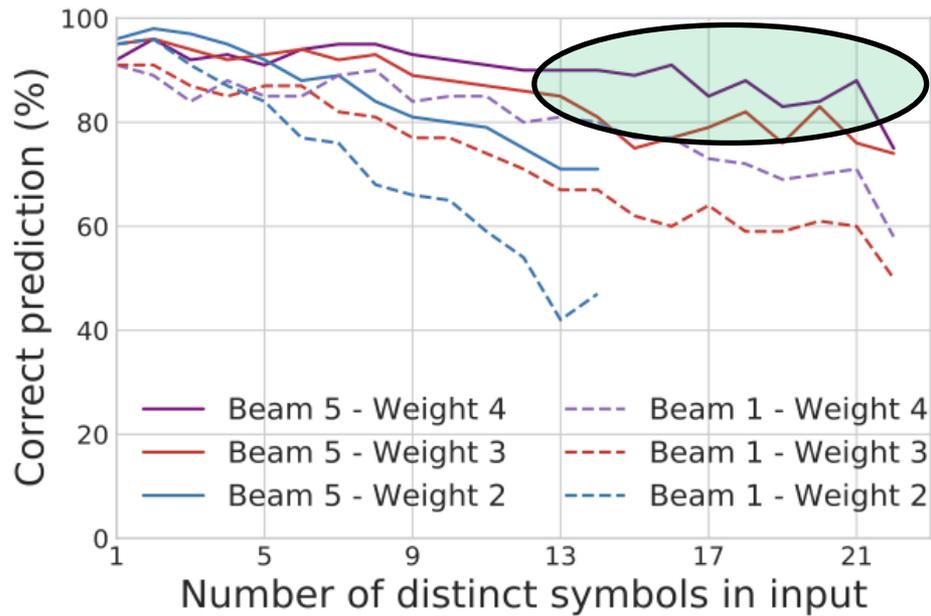


Translate (integrate) the symbol

Training data looks like this

Input symbol \mathcal{S}_i	Simple expression F_i
$-(-x^2 - x + 1) \otimes (1 - x) + (-x^2 - x + 1) \otimes x$ $-(-x^2 - x + 1) \otimes (x + 1) + x \otimes (1 - x) - x \otimes x + x \otimes (x + 1)$	$\text{Li}_2\left(\frac{(1-x)(x+1)}{x}\right)$
$-\frac{23}{4}(1 - x) \otimes x - 6(x^2 + x + 1) \otimes x$	$2 \text{Li}_2(x^3) - \frac{1}{4} \text{Li}_2(x)$
$-40(6 - x^2) \otimes (6 - x^2) - 3(1 - x) \otimes (-x^6 - x^2 + 3)$ $-3(x + 1) \otimes (-x^6 - x^2 + 3) - 3(x^4 + x^2 + 2) \otimes (-x^6 - x^2 + 3)$ $+\frac{1}{4}(5 - 2x) \otimes (2 - x)$	$3 \text{Li}_2(-x^6 - x^2 + 3)$ $-\frac{1}{4} \text{Li}_2(2x - 4)$ $-20 \ln^2(x^2 - 6)$
$8 \frac{x^2 - x - 1}{x - 1} \otimes x - 8((x + 1)(x^2 - x - 1)) \otimes x$ $+8(1 - x) \otimes (-x^3 + x^2 - x - 1) - 8 \frac{1}{x - 1} \otimes x$ $-8(1 - x) \otimes (x(x^3 - x^2 + x + 1))$	$4 \text{Li}_2(x^2)$

Results



		Beam Size 1	Beam Size 5
Weight 2	Transformer	82%	91%
	Classical Algorithm	59%	59%
Weight 3	Transformer	78%	88%
Weight 4	Transformer	80%	89%

Weight 2	Weight 3	Weight 4
$Li_2(x)$	$Li_3(x)$	$Li_4(x)$
$\ln(x) \ln(y)$	$Li_2(x) \ln(y)$	$Li_3(x) \ln(y)$
	$\ln(x) \ln(y) \ln(z)$	$Li_2(x) Li_2(y)$
		$Li_2(x) \ln(y) \ln(z)$
		$\ln(w) \ln(x) \ln(y) \ln(z)$

Can simplify complicated expressions with Li_2, Li_3, Li_4, \dots

- Limited by compute (network size, training time)

Example application

1. Scattering amplitude gives some function of GPLs with complex arguments

$$f(x) = 4\zeta_3 + 9 \left[G(0, 0, 1, x) + G\left(0, 0, \frac{-1 - \sqrt{3}i}{2}, x\right) + G\left(0, 0, \frac{-1 + \sqrt{3}i}{2}, x\right) \right] \\ + 4 \left[-G(-1, -1, -1, x) + G(-1, 0, -1, x) + G(0, -1, -1, x) + G(0, 0, 1, x) - G\left(0, 0, 1, \frac{x}{x+1}\right) \right]$$

2. Express in terms of classical polylogs

$$f(x) = 9 \left(-\text{Li}_3(x) - \text{Li}_3\left(\frac{2ix}{-i + \sqrt{3}}\right) - \text{Li}_3\left(-\frac{2ix}{i + \sqrt{3}}\right) \right) \\ + 4 \left(-\text{Li}_3(x) + \text{Li}_3\left(\frac{x}{x+1}\right) + \text{Li}_3(x+1) - \text{Li}_2(-x) \ln(x+1) \right) \\ - 4 \left(\text{Li}_2(x+1) \ln(x+1) + \frac{1}{6} \ln^3(x+1) + \frac{1}{2} \ln(-x) \ln^2(x+1) \right)$$

3. Compute the symbol and simplify

$$\mathcal{S}[f(x)] = 9(x^2 + x + 1) \otimes x \otimes x + 13(1 - x) \otimes x \otimes x + 4(x + 1) \otimes x \otimes x$$

4. Integrate the symbol with a transformer network

$$f(x) = -\text{Li}_3(x^3) - \text{Li}_3(x^2) + 4\zeta_3 \quad \checkmark$$

The same basic RL or transformer approach **can be applied to many problems**

- Need to be able to generate training data
- Have identities to apply

[Cheung, Dersy, MDS work in progress]

E.g: spinor helicity amplitudes

[23] [35] [46]⁴

$$\begin{aligned} & \frac{(\langle 12 \rangle \langle 15 \rangle \langle 24 \rangle^3 \langle 26 \rangle \langle 35 \rangle [23] [26] [35] [56]^2 + \langle 12 \rangle \langle 15 \rangle \langle 24 \rangle^2 \langle 26 \rangle \langle 34 \rangle \langle 35 \rangle [23] [35] [36] [56]^2 \\ & + \langle 14 \rangle \langle 15 \rangle \langle 23 \rangle \langle 24 \rangle^2 \langle 26 \rangle \langle 35 \rangle [23] [35] [36] [56]^2 + \langle 14 \rangle \langle 15 \rangle \langle 24 \rangle^3 \langle 26 \rangle \langle 35 \rangle [13] [23] [45] [56] \\ & - \langle 14 \rangle \langle 15 \rangle \langle 24 \rangle^3 \langle 26 \rangle \langle 35 \rangle [13] [24] [35] [56] + \langle 14 \rangle \langle 15 \rangle \langle 24 \rangle^3 \langle 26 \rangle \langle 35 \rangle [15] [23] [34] [56] \\ & + \langle 14 \rangle \langle 15 \rangle \langle 24 \rangle^3 \langle 26 \rangle \langle 35 \rangle [23] [35] [46] [56]^2 + \langle 14 \rangle \langle 24 \rangle^3 \langle 26 \rangle \langle 35 \rangle^2 [23] [34] [35] [56] \\ & + \langle 14 \rangle \langle 24 \rangle^3 \langle 26 \rangle \langle 35 \rangle \langle 45 \rangle [24] [34] [35] [56] - \langle 14 \rangle \langle 24 \rangle^3 \langle 26 \rangle \langle 35 \rangle \langle 56 \rangle [24] [35] [36] [56] \\ & + \langle 15 \rangle^2 \langle 24 \rangle^3 \langle 26 \rangle \langle 35 \rangle [23] [35] [56]^3) \end{aligned}$$

simplify this:

Possible “moves”:

- **Schouten:**

$$\langle ij \rangle \rightarrow \frac{\langle il \rangle \langle kj \rangle}{\langle kl \rangle} + \frac{\langle ik \rangle \langle jl \rangle}{\langle kl \rangle}$$

- **Insertion of identity:**

$$\frac{\langle p_1 p_2 \rangle [p_3 p_4]}{\langle p_1 p_3 \rangle} \rightarrow \frac{\langle p_1 p_2 \rangle [p_3 p_4]}{\langle p_1 p_3 \rangle} \frac{\langle ij \rangle}{\frac{\langle il \rangle \langle kj \rangle}{\langle kl \rangle} + \frac{\langle ik \rangle \langle jl \rangle}{\langle kl \rangle}}$$

- **Momentum conservation:**

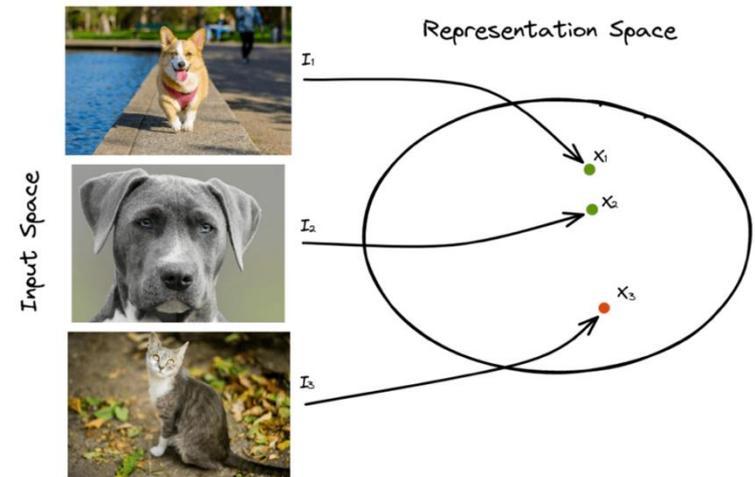
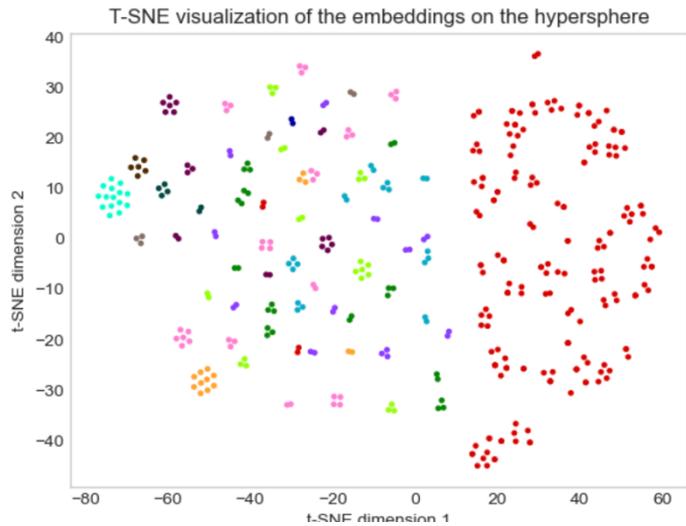
$$\langle ij \rangle \rightarrow - \sum_{i \neq j} \frac{\langle it \rangle [tk]}{[jk]}$$

- **Insertion of zero:**

$$\langle mk \rangle + [ln] \rightarrow \langle mk \rangle + [ln] + \left(\langle ij \rangle - \frac{\langle il \rangle \langle kj \rangle}{\langle kl \rangle} + \frac{\langle ik \rangle \langle jl \rangle}{\langle kl \rangle} \right)$$

Contrastive learning

- Typical expressions involving spinors are very large
 - hundreds of thousands of terms
- Cannot expect ML to process large inputs directly
- Can use **contrastive learning** to group similar terms



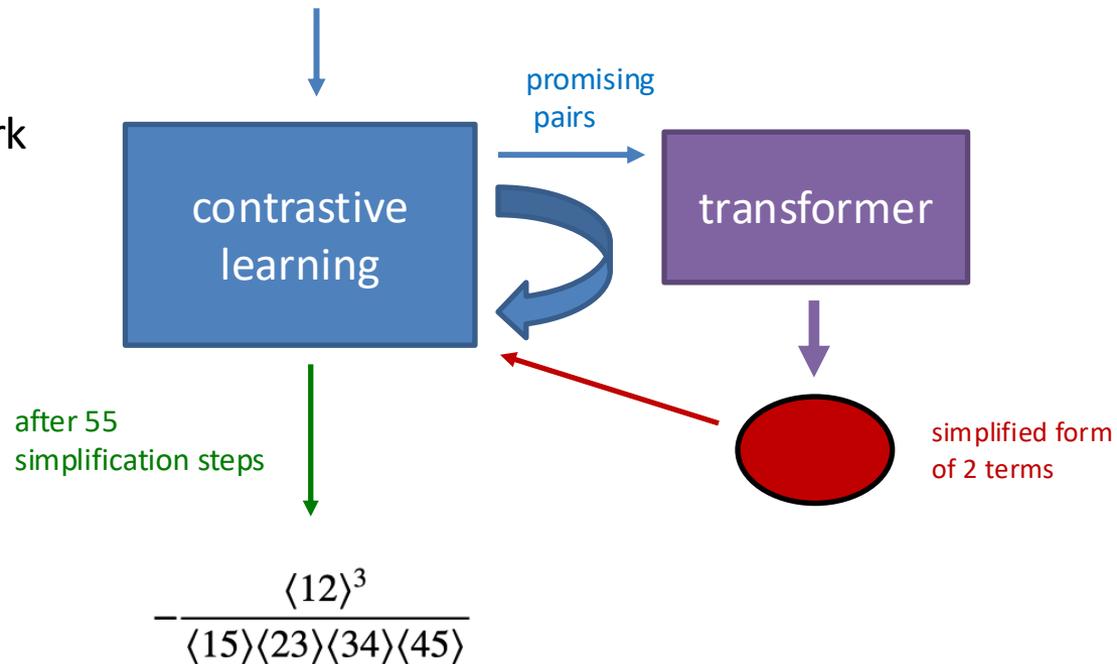
- Learns to classify by mass dimension
- Learns other features as well

Example application: 79 terms \rightarrow 1 term

1. Compute 5-point MHV amplitude with Feynman diagrams: 390 terms
2. Choose some smart reference vector to reduce to 79 terms
(smarter choice can reduce to 17)

$$\frac{\langle 12 \rangle^2 \langle 15 \rangle^2 \langle 24 \rangle \langle 34 \rangle [12] [14] [15] [23] [25] + \langle 12 \rangle^2 \langle 15 \rangle \langle 23 \rangle \langle 34 \rangle \langle 45 \rangle [12] [15] [23] [25] [34] + \dots 77 \text{ terms}}{\langle 15 \rangle^2 \langle 23 \rangle \langle 34 \rangle^2 \langle 45 \rangle^2 [12]^2 [15] [23] [45]}$$

3. Feed to network



4. Output when simplification completes

Can ML help with the non-perturbative S-matrix?

What does this mean?

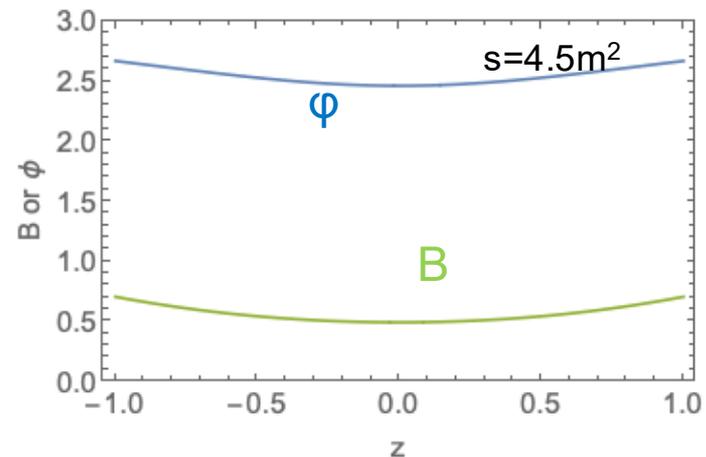
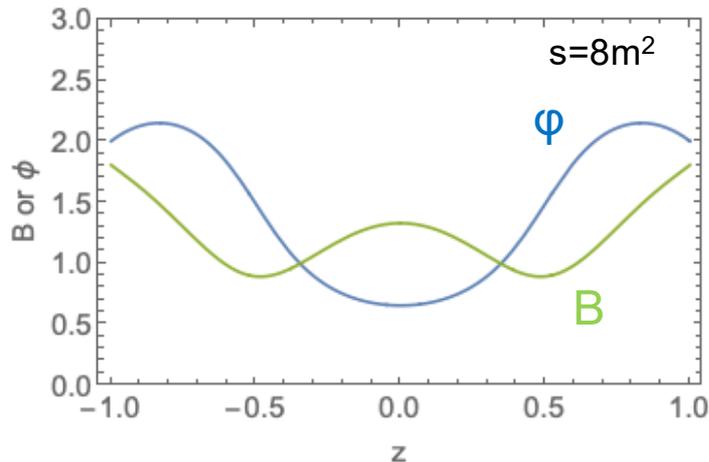
[Dersy, MDS, Niewinski Zhiboedov, work in progress]

- Use analyticity, unitarity, crossing symmetry to determine S

Example questions:

For a given cross section $\sigma \sim |A|^2$

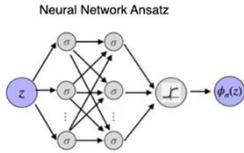
- Does there always exist a phase ϕ so that $A = B e^{i\phi}$?
- e.g. Penedones et al 1708.06756, Fitzpatrick et al 2207.12448



- How do we determine ϕ from B ?
- Can there be many phases ϕ_1, ϕ_2, \dots for the same B ?

Can we find $\varphi(z)$ given $B(z)$? ... Yes!

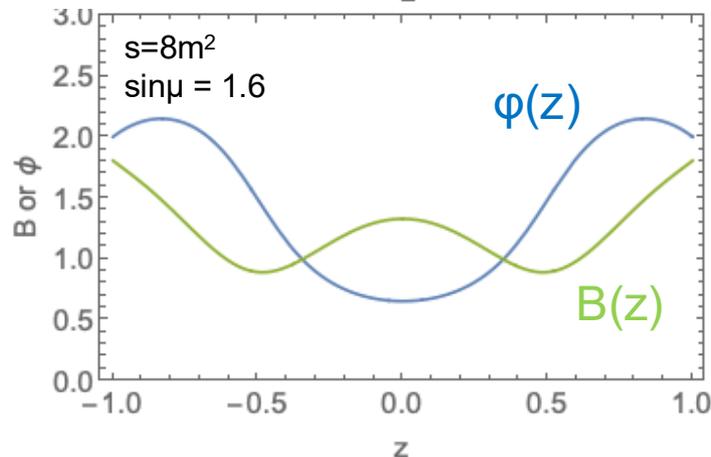
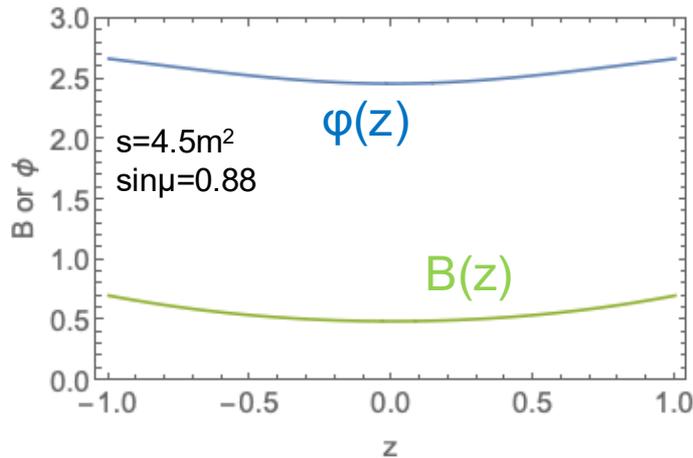
- Parametrize $\varphi(z)$ as a neural network



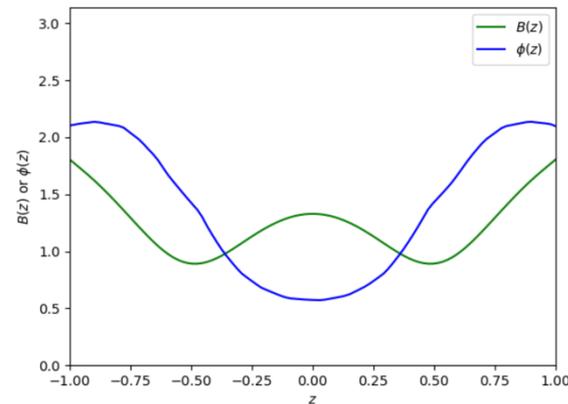
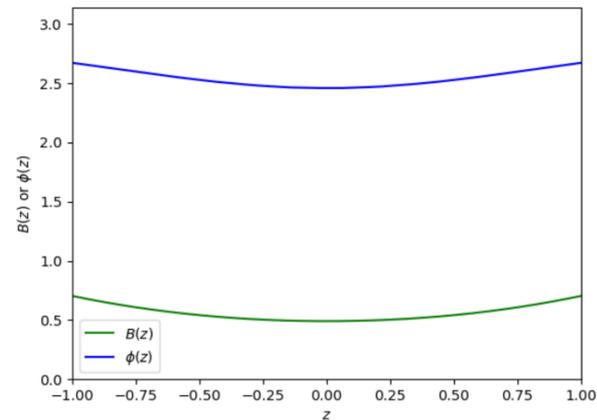
- Loss function is unitarity condition

$$\mathcal{L} = \mathbb{E} \left\| \left| B(z) \sin \phi(z) - \frac{1}{4\pi} \int_{-1}^1 dz_1 \int_0^{2\pi} d\phi_1 B(z_1) B(z_2) \cos(\phi(z_1) - \phi(z_2)) \right| \right\|^2$$

S-matrix bootstrap results



φ from B using ML

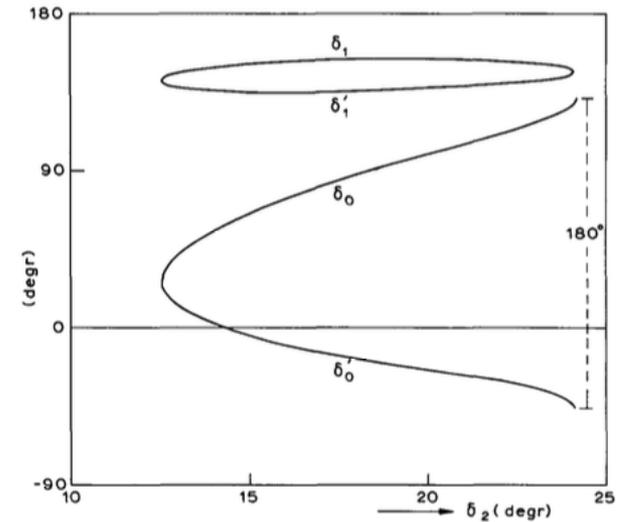
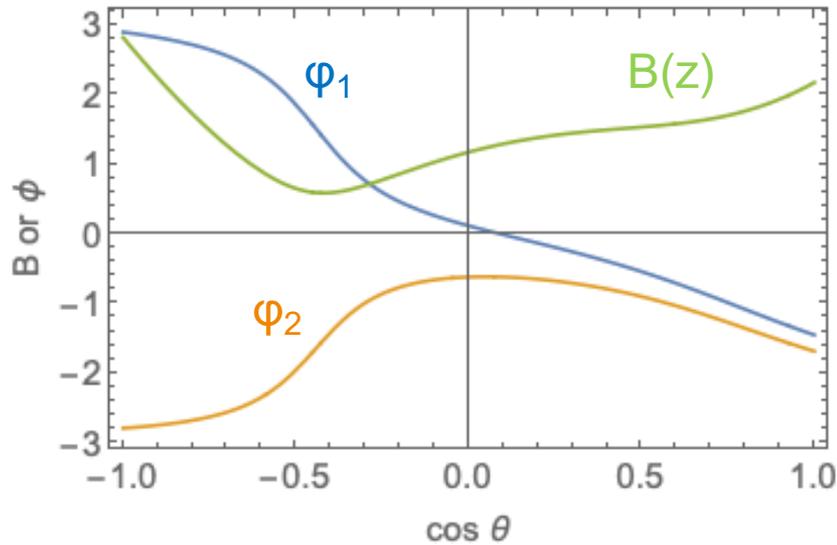


Phase ambiguities

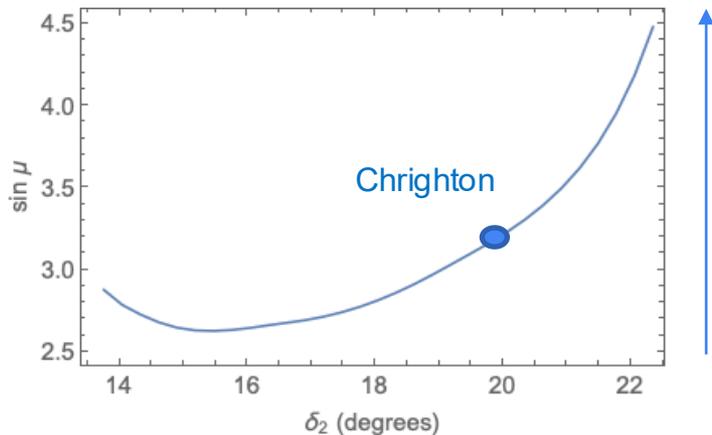
Crichton (1966) found a solution for L=2

$$\delta_2 = \delta'_2 = 20^\circ = 0.349066$$

$$\delta_0 = -0.407703, \quad \delta_1 = -0.758247, \quad \delta'_0 = 1.72571 \quad \delta'_1 = -0.463483$$



Changing δ_2 changes the solution along a curve [Atkinson 1973]



[Atkinson 1973]

Family of solutions

Phase ambiguities with ML

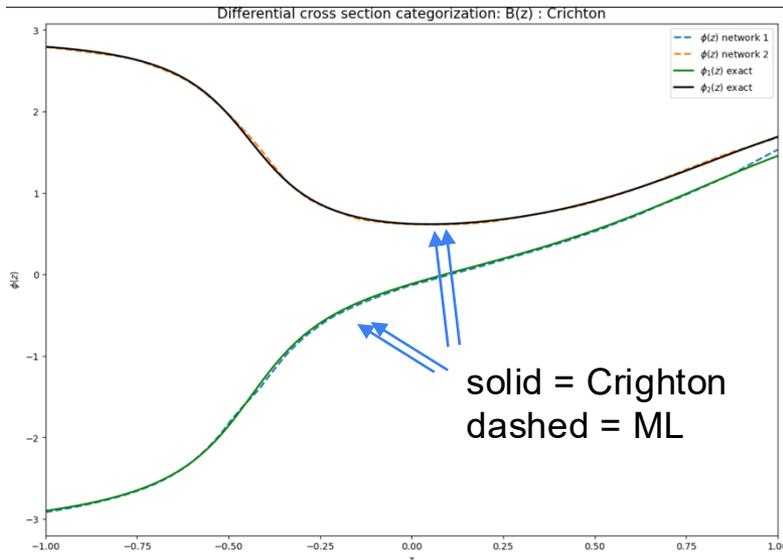
Given $B(z)$, find two phases $\phi_1(z)$ and $\phi_2(z)$

- Unitarity condition loss for each ϕ :

$$\mathcal{L} = \mathbb{E} \left\| \left| B(z) \sin \phi(z) - \frac{1}{4\pi} \int_{-1}^1 dz_1 \int_0^{2\pi} d\phi_1 B(z_1) B(z_2) \cos(\phi(z_1) - \phi(z_2)) \right| \right\|^2$$

- Add repulsive loss to keep solutions apart

$$\mathcal{L}_R = \mathbb{E}_z \|d(\phi_1(z), \phi_2(z))\|^{-p} + \mathbb{E}_z \|d(\pi - \phi_1(z), \phi_2(z))\|^{-p}$$

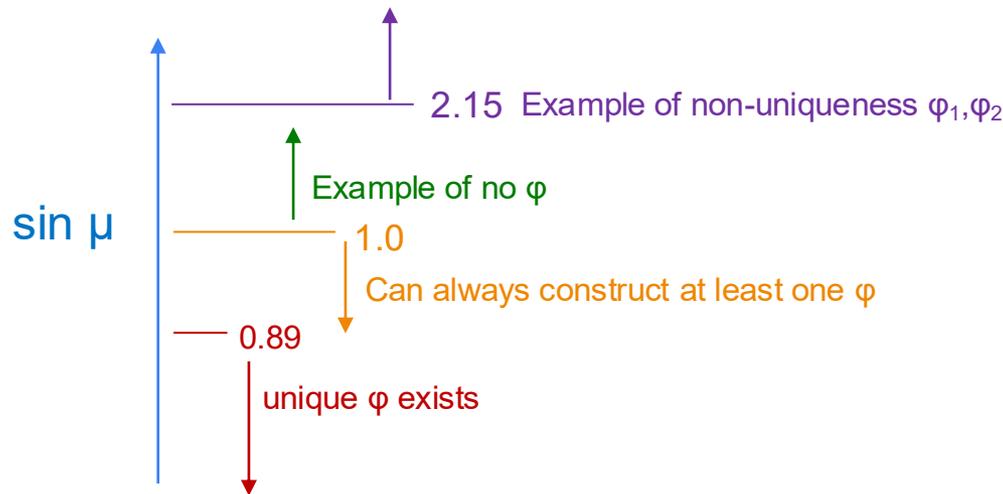


- Two solutions found with ML agree exactly with Crichton's
- We input $B(z)$ and find phases

Martin's $\sin\mu$

$$\sin\phi(z) = \frac{\int_{-1}^1 dz_1 \int_0^{2\pi} d\phi_1 B(z_1) B(z_2) e^{i\phi(z_1) - i\phi(z_2)}}{4\pi B(z)} \leq \max_z \frac{\int_{-1}^1 dz_1 \int_0^{2\pi} d\phi_1 B(z_1) B(z_2)}{4\pi B(z)} \equiv \sin\mu[B]$$

- Martin (1969): If $\sin\mu < 1$ for a given $B(z)$ then there always exists a phase $\phi(z)$
- There exist $B(z)$ with $\sin\mu > 1$ for which no phase exists
- Gangal and Kupsch (1984) $\sin\mu < 0.89$ then the phase is unique
- Atkinson et al. multiple phases exist for with $\sin\mu = 2.15$



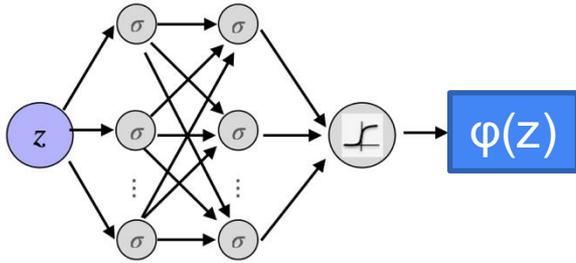
Open questions

- Can 2.15 be lowered?
- Can 0.89 be raised?
- How can we construct ϕ when $\sin\mu > 1.0$?
- Is there an upper bound on μ for **physical** amplitudes?

Machine Learning approach

Represent functions as NNs

Construct two functions using same $F(z)$ and z_1



$$A_1(z) = \frac{z - z_1}{1 - z_1} F(z) \quad A_2(z) = \frac{z - z_1^*}{1 - z_1^*} F(z)$$

Loss = unitarity for ϕ_1 + unitarity for ϕ_2 + ϕ_1 - ϕ_2 repulsive term

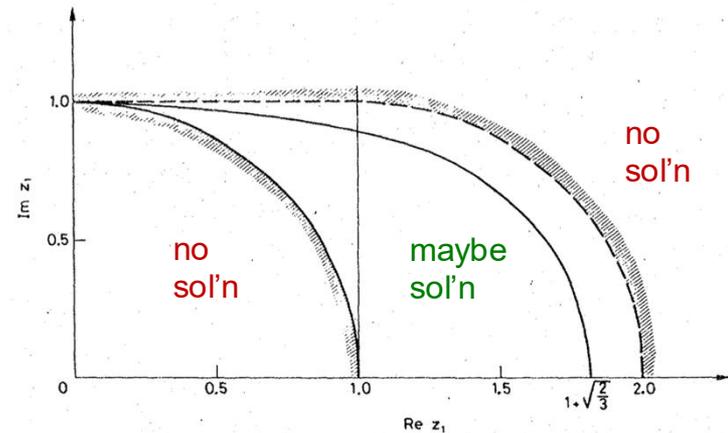
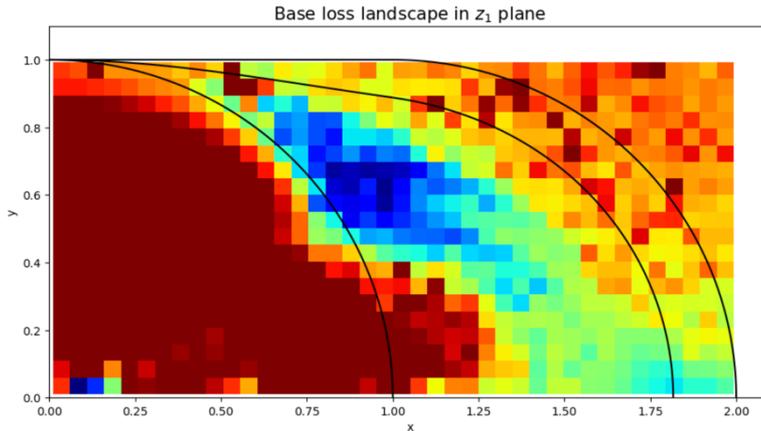
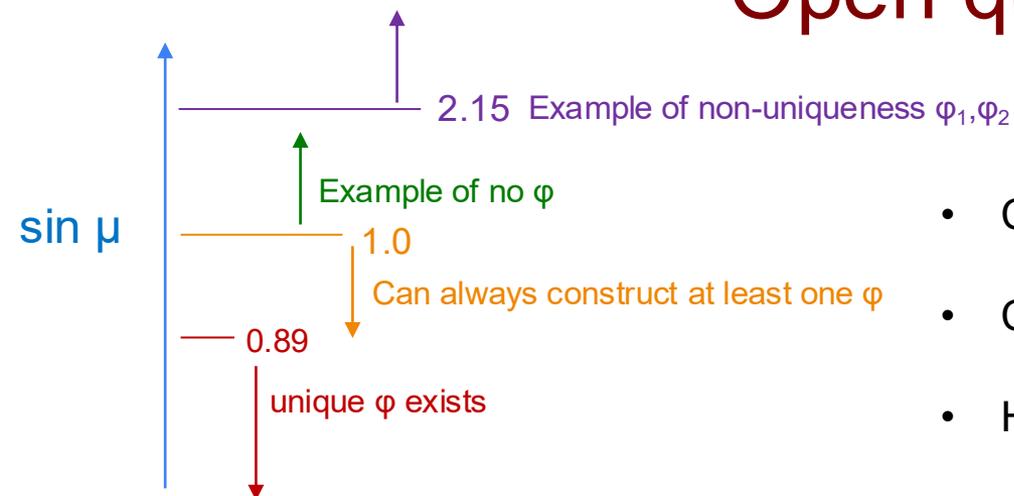


Fig. 1 of Atkinson 1977

- Loss landscape matches Atkinson 1977
- Gradient descent on $\sin \mu$ finds **solution with $\sin \mu = 1.64$**
 - New **lowest known value** of $\sin \mu$ with phase ambiguity

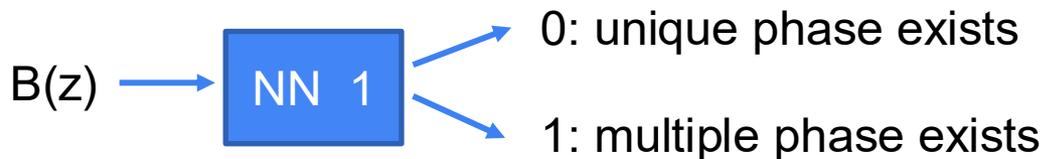
Open questions



- Can 2.15 be lowered?
 - Yes: to 1.64 at least
- Can 0.89 be raised?
 - Uniqueness proof hard to answer with ML
- How can we construct φ when $\sin \mu > 1.0$?
 - Using ML ✓
- Is there a bound on μ for **physical** amplitudes?
 - Need more examples and physical insight

New questions

- What properties to the phase ambiguous solutions have?
- Is there a better way to characterize solutions than $\sin \mu$?
 - Train NN 1



- Train NN 2 to figure out what NN 1 is doing
 - Symbolic regression?
 - Can we learn a better indicator for phase determination than $\sin \mu$?

The future



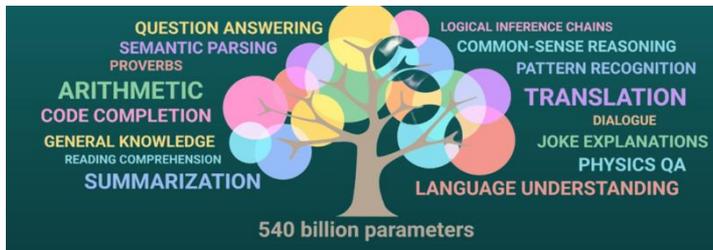
Orc Aerobic class (3/2023)

The future

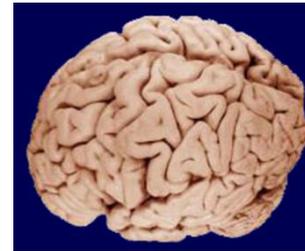


- 3rd generation model (2020)
- 175 billion parameters

Google : PaLM (2022)
540 billion parameters

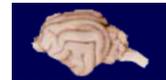


Human brain

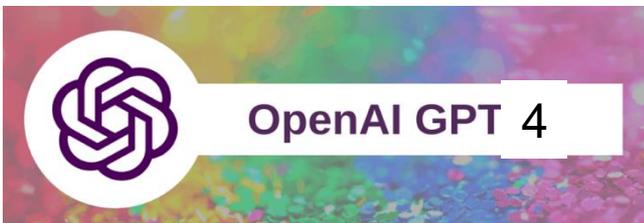


80 billion neurons
150 trillion synapses

Cat brain



0.760 billion neurons
10 trillion synapses

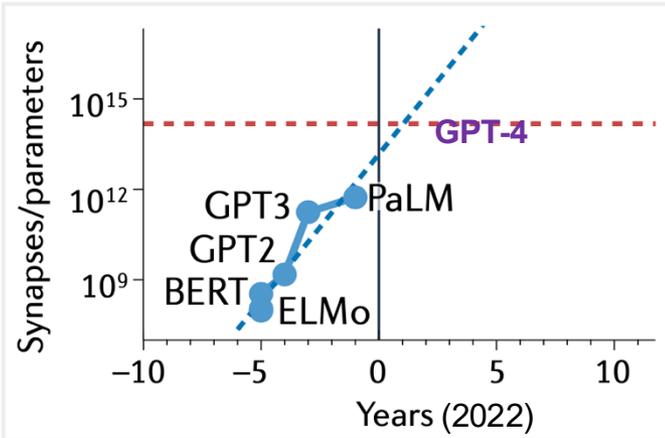


170 trillion parameters!

2023 ... 2040... 2100? ...3000...?

Future of AI

Should artificial intelligence be interpretable to humans? MDS, Nature reviews physics (2022)



- ELMo (94 million parameters, 2018)
- GPT2 (1.5 billion parameters, 2019)
- GPT3 (175 billion parameters, 2020)
- PALM (540 billion parameters, 2022)
- GPT4 (110 trillion parameters, 2023)

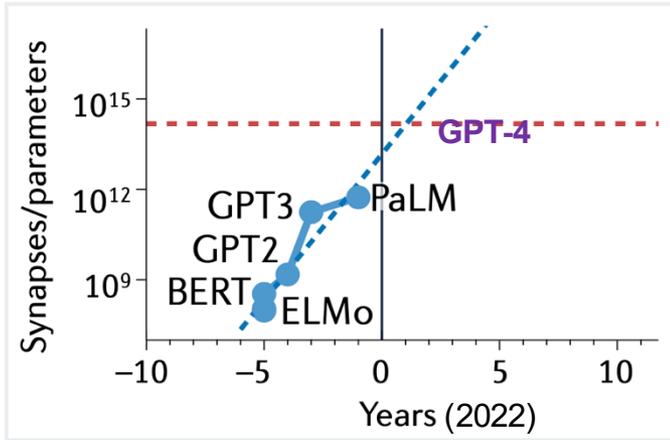
AI grows by factor of ~10/year

Future of AI

Should artificial intelligence be interpretable to humans? MDS, Nature reviews physics (2022)

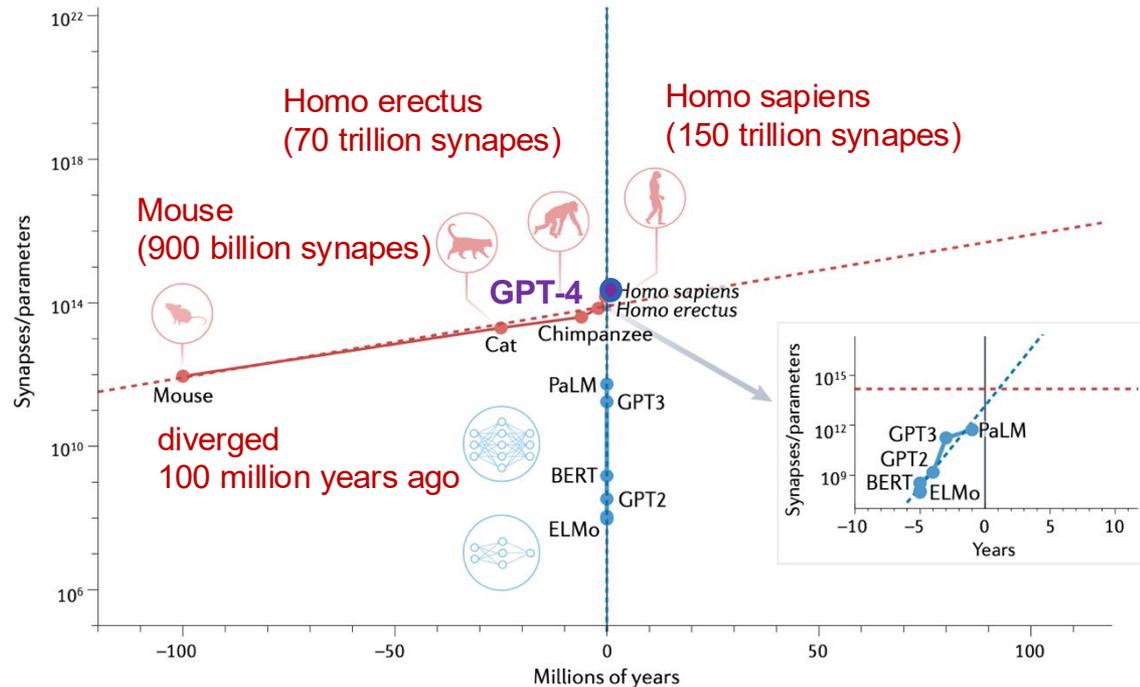
- ELMo (94 million parameters, 2018)
- GPT2 (1.5 billion parameters, 2019)
- GPT3 (175 billion parameters, 2020)
- PALM (540 billion parameters, 2022)
- GPT4 (110 trillion parameters, 2023)

AI grows by factor of ~10/year



Biological intelligence grows by a factor of 2 in 10^6 years

- Both AI and biological intelligence grow exponentially
- Factor of 10^6 difference in exponent (!!)



Does AI look like biological intelligence?

Exam	GPT-4
SAT Math	700 / 800 (~89th)
Graduate Record Examination (GRE) Quantitative	163 / 170 (~80th)
Graduate Record Examination (GRE) Verbal	169 / 170 (~99th)
AP Environmental Science	5 (91st - 100th)
AP Macroeconomics	5 (84th - 100th)
AP Microeconomics	5 (82nd - 100th)
AP Physics 2	4 (66th - 84th)

- GPT4 can do freshman level college physics very well
- In ~ 1 year will do it perfectly
- In ~ 3 years it will have mastered college

Can the improvements continue?

- Already use all available training data (the internet)

Does AI look like biological intelligence?

Exam	GPT-4
SAT Math	700 / 800 (~89th)
Graduate Record Examination (GRE) Quantitative	163 / 170 (~80th)
Graduate Record Examination (GRE) Verbal	169 / 170 (~99th)
AP Environmental Science	5 (91st - 100th)
AP Macroeconomics	5 (84th - 100th)
AP Microeconomics	5 (82nd - 100th)
AP Physics 2	4 (66th - 84th)

- GPT4 can do freshman level college physics very well
- In ~ 1 year will do it perfectly
- In ~ 3 years it will have mastered college

Can the improvements continue?

- Already use all available training data (the internet)

Yes!

- It would be embarassingly foolish to say this is the endpoint
- ML can learn just like we do
 - Data augmentation
 - Create and solve toy problems
 - Fewer sociological pressures than human beings have

Conclusions

Machine learning is having a huge impact on high energy physics

- Lattice QCD
- Anomaly detection
- Detector simulation/unfolding
- Jet substructure/classification

Future of ML in high energy will likely be more symbolic

- Begin with hybrid numerical/symbolic problems
 - (polylogarithms, spinor helicities, S-matrix phase determination)
- Eventually exploit large language models to understand physical systems like we do



AI generated image, April 2023

The future is bright!

ML provides hope at finally solving problems too difficult for human beings