“Learning” Feynman Integrals

SIMIS-APCTP 2026

Yuanche Liu, Dept. of Mod. Phys., **USTC**

2026.08.27

Based on **arXiv:2510.10099**, in collaboration with **Yingxuan Xu** & **Yang Zhang**

arXiv:2609.xxxxx, in collaboration with **Dongshan Jian, Yanqing Ma, Boxuan Shi, Yu Wu, Zihao Wu and Yang Zhang**

01

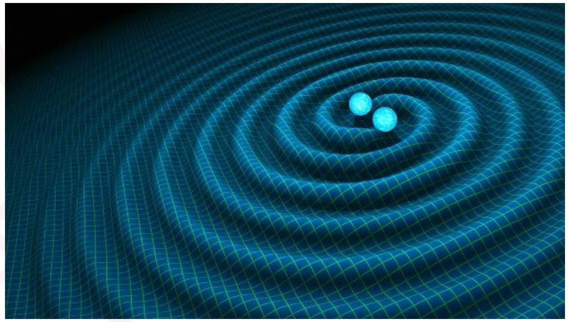
Introduction on Feynman Integrals and Symbology

About Feynman integrals, canonical differential equations and their symbol letters.

Feynman Integrals Everywhere

Gravity: $\int \prod_{i=1}^{L} \frac{dp_i^D}{(2\pi)^D} \frac{(\kappa\, p_j^\mu p_j^\nu)\cdots}{\text{propagators}}$

Gauge theory: $\int \prod_{i=1}^{L} \frac{d^D p_i}{(2\pi)^D} \frac{(g\, p_j^\nu)\cdots}{\text{propagators}}$



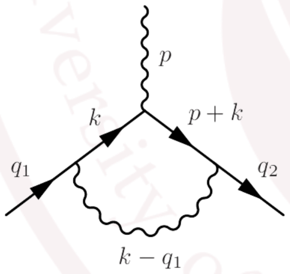
Precision Physics

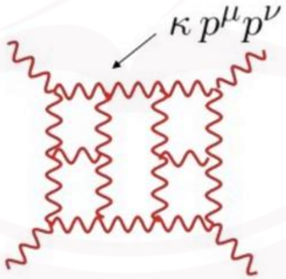
Formal Theory

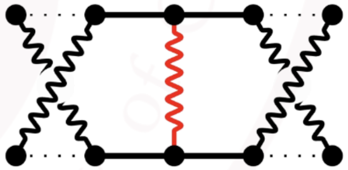
[Bern, Carrasco, Johansson, 2010]

Gravitational Wave

[Driesse, Jakobsen, Mogull, Nega, Plefka, Sauer, Usovitsch, 2026]





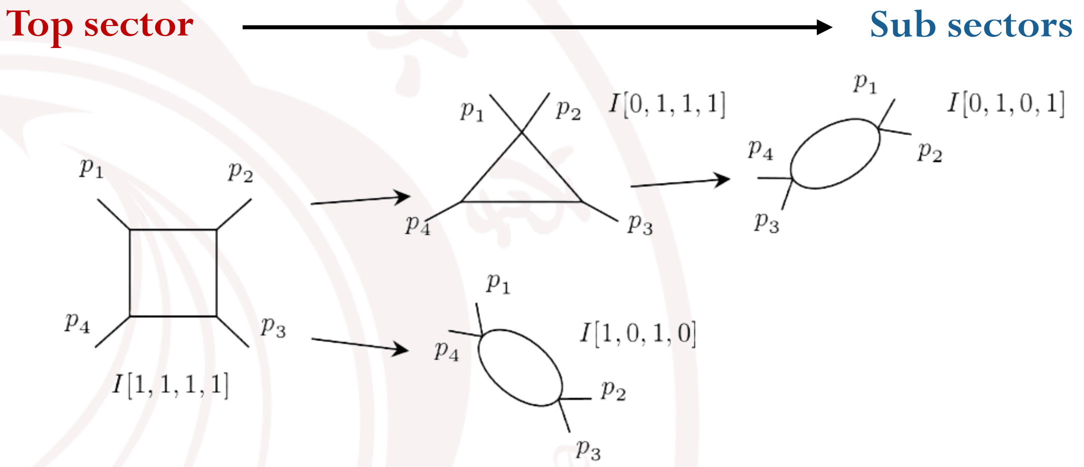


Loop Corrections: Feynman Integrals

Feynman Integrals and their Family Structure

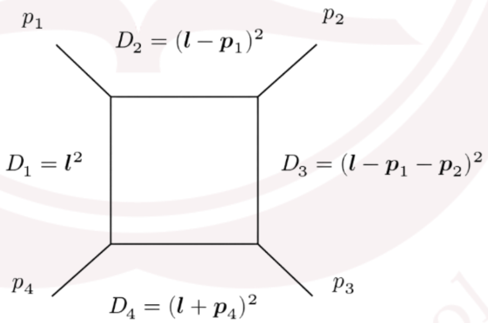
Feynman Integrals can be **recursively built** from top sector to sub sectors

- Form a **family structure**, in which integrals are determined by combination of propagators.



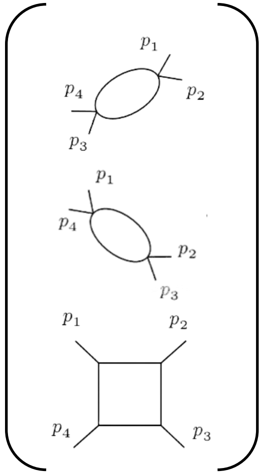
The well-known **IBP identities** [Tkachov, 1981] uncover the **linear relations** between different integrals

- One can determine a **basis** in respect to linear relations – the **Master Integrals**

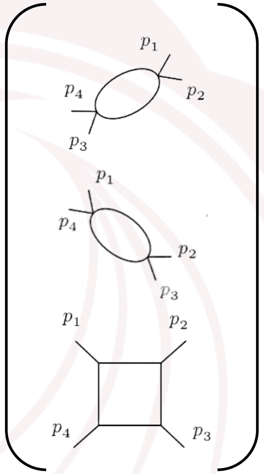


$$
\begin{aligned}
&-\alpha_4 I_{\text{box}}[-1+\alpha_1,\alpha_2,\alpha_3,1+\alpha_4] - \alpha_3 I_{\text{box}}[-1+\alpha_1,\alpha_2,1+\alpha_3,\alpha_4]\\
&-\alpha_2 I_{\text{box}}[-1+\alpha_1,1+\alpha_2,\alpha_3,\alpha_4] + \alpha_3 s I_{\text{box}}[\alpha_1,\alpha_2,1+\alpha_3,\alpha_4]\\
&+(-2\alpha_1-\alpha_2-\alpha_3-\alpha_4+D) I_{\text{box}}[\alpha_1,\alpha_2,\alpha_3,\alpha_4] = 0
\end{aligned}
$$

Canonical Differential Equations

∂_s 

$$= \begin{pmatrix} 0 & 0 & 0 \\ 0 & -\frac{\varepsilon}{s} & 0 \\ \frac{4\varepsilon-2}{st(s+t)} & \frac{-4\varepsilon+2}{s^2(s+t)} & -\frac{\varepsilon t+s+t}{s(s+t)} \end{pmatrix}$$



Taking **derivatives** for master integrals in respect to **kinematic variables** lead to a **differential equation** system.

$$\boldsymbol{I}(s,t,\varepsilon) = \{I_{\text{box}}[0,1,0,1], I_{\text{box}}[1,0,1,0], I_{\text{box}}[1,1,1,1]\}$$

$$\boldsymbol{I} = \varepsilon^2 \exp(\varepsilon\gamma_E)(-s)^{\varepsilon} \begin{pmatrix} \frac{1-2\varepsilon}{\varepsilon} I_{\text{box}}[0,1,0,1] \\ \frac{1-2\varepsilon}{\varepsilon} I_{\text{box}}[1,0,1,0] \\ st I_{\text{box}}[1,1,1,1] \end{pmatrix}$$

$$\frac{\partial}{\partial x_i}\boldsymbol{I}(\boldsymbol{x},\varepsilon) = \varepsilon \boldsymbol{M}_i(\boldsymbol{x})\boldsymbol{I}(\boldsymbol{x},\varepsilon)$$

$$\boldsymbol{M}(x) = \begin{pmatrix} -\frac{1}{x} & 0 & 0 \\ 0 & 0 & 0 \\ \frac{2}{x(x+1)} & -\frac{2}{x+1} & -\frac{1}{x(x+1)} \end{pmatrix}$$

Change to **Canonical Basis**, one obtain ε-factorized **Canonical Differential Equations** [Henn, 2013]

- Sometimes this can be very complicated because of **complicated function space**, see [ε-collaboration, 2025]

Symbol Letters

$$\frac{\partial}{\partial x_i}\boldsymbol{I}(\boldsymbol{x},\varepsilon) = \varepsilon \boldsymbol{M}_i(\boldsymbol{x})\boldsymbol{I}(\boldsymbol{x},\varepsilon)$$

$$\boldsymbol{M}(x) = \begin{pmatrix} -\frac{1}{x} & 0 & 0 \\ 0 & 0 & 0 \\ \frac{2}{x(x+1)} & -\frac{2}{x+1} & -\frac{1}{x(x+1)} \end{pmatrix}$$

$$M(x) = \sum_i a_i \,\mathrm{d}\log \boxed{W_i}, \qquad a_i \in \mathbb{Q}^{n\times n}$$

$$W_i = P \in \mathbb{Q}[x_i]$$

parity **even letter**

$$W_i = \frac{\sqrt{P}-Q}{\sqrt{P}+Q}, \qquad P, Q \in \mathbb{Q}[x_i]$$

parity **odd letter**

Symbol letters

- **Symbol Alphabet** = {All **Symbol Letters**}
- Very **limited** for a certain family!

CDE Matrix contains different **Symbol letters**, revealing the **analyticity**(poles, regions) of Feynman Integrals.

- Ingredients for **analytical bootstrap** on Feynman Integrals or Amplitudes [Carrolo, Chicherin, Henn, Yang, Zhang, 2025]
- **Cluster Algebra** structures [Pokraka, Spradlin, Anastasia, Weng, 2025] from Mathematical Physics

Some previous algorithms to generate **Symbol letters**:

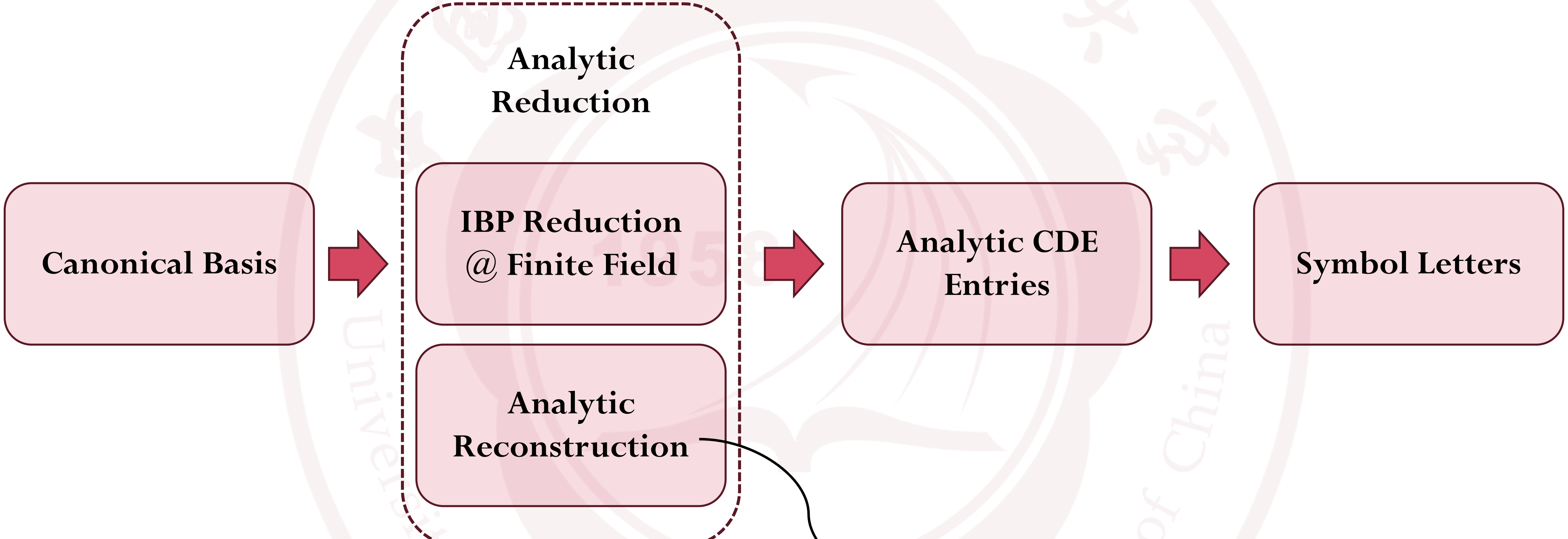
- `Effortless` [Antonela, 2024] and `SOFIA` [Correia, Giroux, Mizera, 2025], using **Landau analysis** and some conjectures.
- `BaikovLetter` [Jiang, Lian, Yang, 2024] using **Baikov Rep.+ Maximal Cut**.

02

Machine Learning And Symbolic Regression

An introduction about Machine Learning and Symbolic Regression.

Why Symbolic Regression?



Extremely **Time Consuming!**

- 2l6p planar [Henn, Matijasic, Miczajka, Peraro, Xu, Yang, 2025]
- **2 hrs/pt** for **numerical** CDE matrix
- **Unable to reconstruct** all letters!

Why Symbolic Regression?

Canonical Basis

Black Box

Input:
Master Integrals

Output:
Analytic Coefficients

Hint:
Easy for **numeric**
Hard for **analytic**

Analytic CDE Entries

Symbol Letters

This is a typical Symbolic Regression task!

Machine Learning: Symbolic Regression

Numerical Data Set

$x = 0.3333, f(x) = 2.1111$
$x = 0.0000, f(x) = 2.0000$
$x = 1.0000, f(x) = 3.0000$
...

Search & Verification

No	Expr	Comp	Error
1	0.4exp(x)	3	0.132
2	x^2+2.000	4	0.000
3	x+ln(x)	4	0.143
4	x^3+x^2	5	0.475

Symbolic Expressions

$$f(x) = x^2 + 2.000$$

Symbolic regression: search interpretable expressions from numerical samples.
PySR [Cranmer, 2023] -- genetic programming
SymbolicGPT [Valipour, Panju, You, Ghodsi, 2021] -- LLM prediction
IdeaSearchFitter [Song, Cai, Zhang, Wei, Pan, Qiu, Cao, Hou, Liu, Luo, Zhu, 2025] -- GP + LLM

Machine Learning: Symbolic Regression

Input Data
`x, y, f(x,y)`

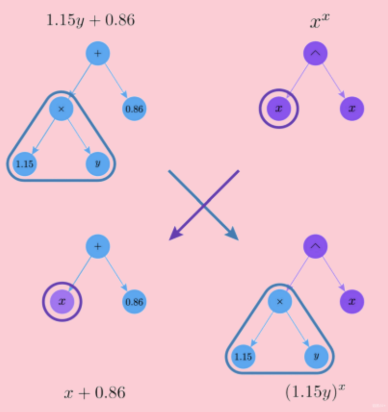
numbers

Symbolic Regression
PySR

Selected Equation
“`f=x+ln(y)`”

symbols

Searching Tree



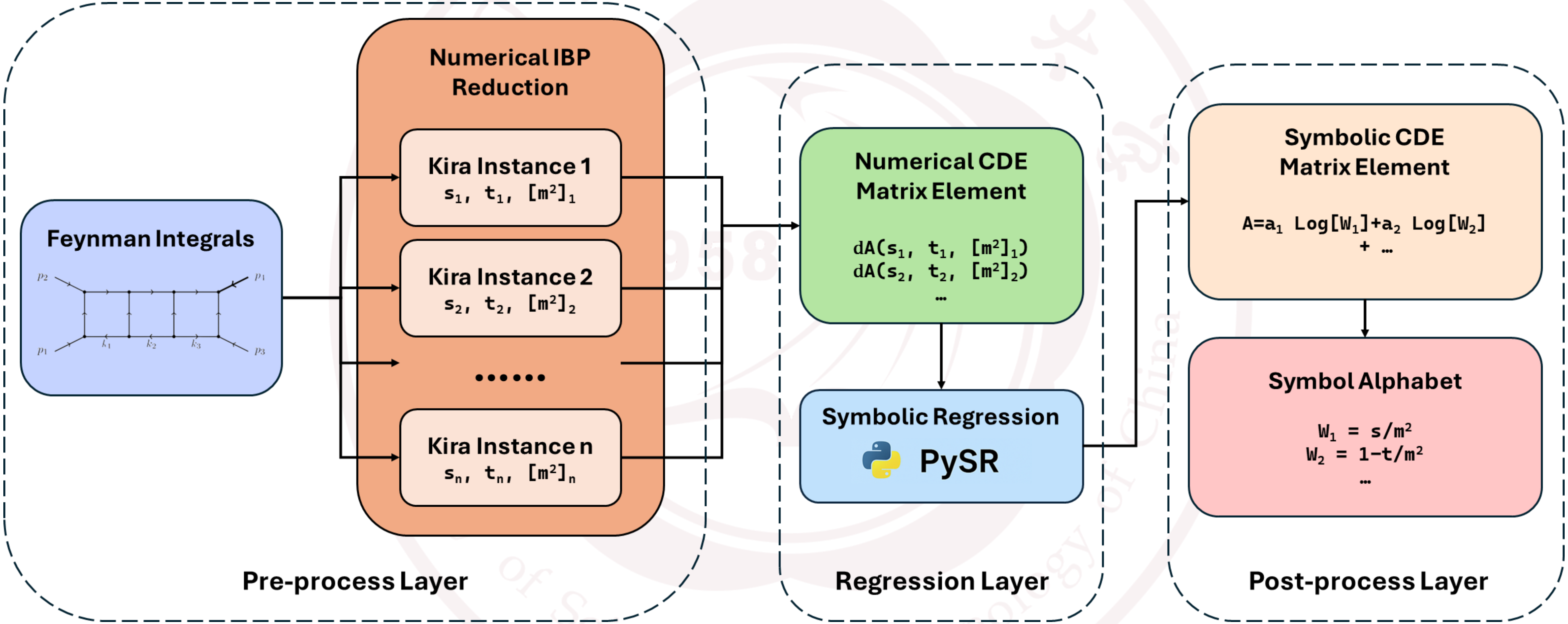
Pareto Frontier

No	Expr	Comp	Error
1	x+y	3	0.132
2	x²+0.1y	5	0.329
3	x+ln(y)	4	0.003
4	x³+y²	5	0.475

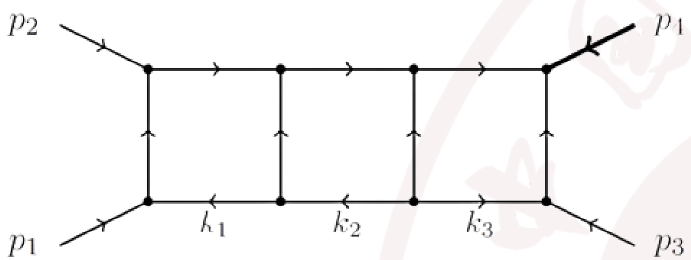
PySR is a powerful **Symbolic Regression** toolkit.

- **No pre-train**: Based on Genetic Programming. Easy to use, no GPU requirement.
- **Efficiency** & **Accuracy**: Pareto Frontier from complexity and accuracy.
- **Sufficient Extension**: Support Julia & Python user-defined functions, and many advanced math operations…

Workflow Introduction



Examples: 3l4p1m & 2l3p1m-np



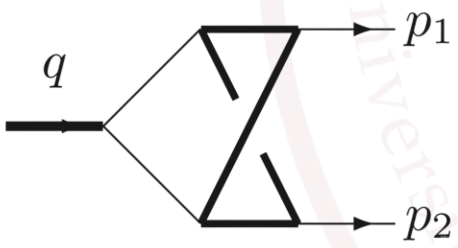
p_4 has mass m^2, others are massless.

3l4p1m [Vita, Matrolia, Schubert, Yundin, 2014]

- **83MIs**, **2**(+1) **scales**: $x=s/m^2$, $y=t/m^2$
- 6 different Symbol letters—**all successfully built**!
- KIRA reduction:
  - **Num**:~20mins 12 threads of i9-13950HX CPU @ 5.30GHz
  - **Ana**: ~100mins 12 threads

$$W_i = \{x, 1-x, y, 1-y, x+y, 1-x-y\}$$

Success



q has mass s, others are massless.
4 propagators with coincide mass.

2l3p1m-np [Manteuffel, Tancredi, 2017]

- **9MIs**(sub-topology), **1**(+1) **scales**: $x=s/m^2$
- 5 different Symbol letters—**all successfully built**!
- Letters have square roots!
- KIRA reduction:
  - **Num**:~10mins 12 threads of i9-13950HX CPU @ 5.30GHz
  - **Ana**: ~60mins 12 threads

$$W_i = \{\sqrt{x}, \sqrt{x+4}, \sqrt{x-4}, \frac{1}{2}(\sqrt{x}+\sqrt{x+4}), \frac{1}{2}(\sqrt{x}+\sqrt{x-4})\}$$

Success

Trials on more Diagrams…

#Loop \ Family	1 Scale 3p1m,4p0m	2 Scales 4p1m	3 Scales 4p2m	5 Scales 5p0m	5+ Scales 5p1m, 6p0m
1	✓✓	✓✓	✓✓	✓	✓
2	✓✓	✓✓	✓✓	✓	✗
3	✓✓	✓✓	✓✓	✓	—
4	✓✓	—	—	—	—

✓✓ : **Complete Symbol Alphabet** reconstructed;

✓ : **All Even Letters** obtained, remaining ones (**Odd Letters**) can be **constructed manually**;

✗ : Some letters not found;

— : Not tested because of incomplete reference.

315p-PBB family

[**Liu**, Matijasic, Miczjka, YX. Xu, YQ. Xu, Zhang, 2024]

316MIs, 5 scales

20+1 **Even Letters Found.**

10 Odd Letters manually built.

Discussions and Future…

Unlike traditional Machine Learning project…

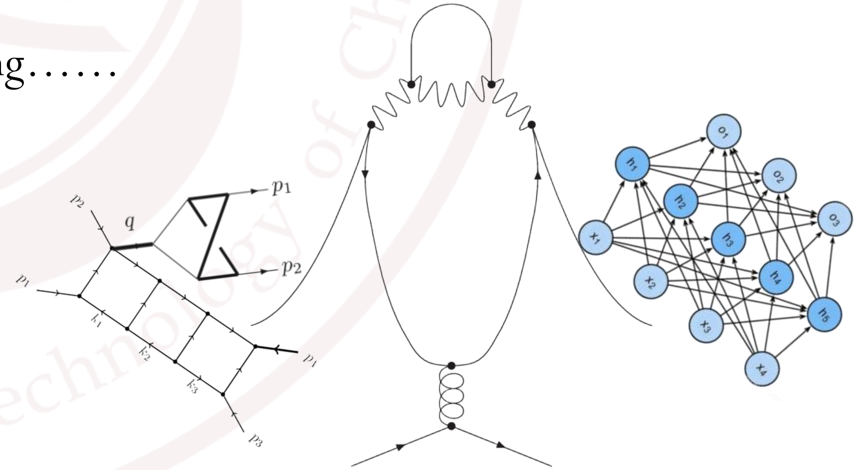
- We pursue a **completely exact solution**, instead of **approximate computation**.

To achieve this…

- We **carefully tune** Hyper-parameter to **"force" the model over-fitting**;
- We also perform an **early-terminate** as soon as **exact solution comes up**.

In the future…

- Improve **accuracy and efficiency**?
- New areas for **AI-ML** to obtain **exact & closed form solution**?
- Symbolic Regression, Reinforcement Learning, LLM-based Searching……



03

AI Agent Makes Canonical Basis Simple

A stateful, sector-by-sector search for exact UT bases

Physics Agents: Moving Down the Stack

From workflow orchestration to physics-directed, long-horizon research

2025 → 2026

HEPTAPOD

Schema-validated orchestration for HEP toolchains

[Menzo, Roman, Gleyzer, Matchev, Fleming, Hoeche, Mrenna, Shyamsundar, 2025]

GRACE

Simulation-native detector and experiment design

[Hill, Ryoo, 2026]

ColliderAgent

Lagrangian → simulation → phenomenological outputs

[Qiu, Cai, Wei, Li, Yin, Cao, Liu, Luo, Yuan, Zhu, 2026]

SMEFT-Pheno-Agent

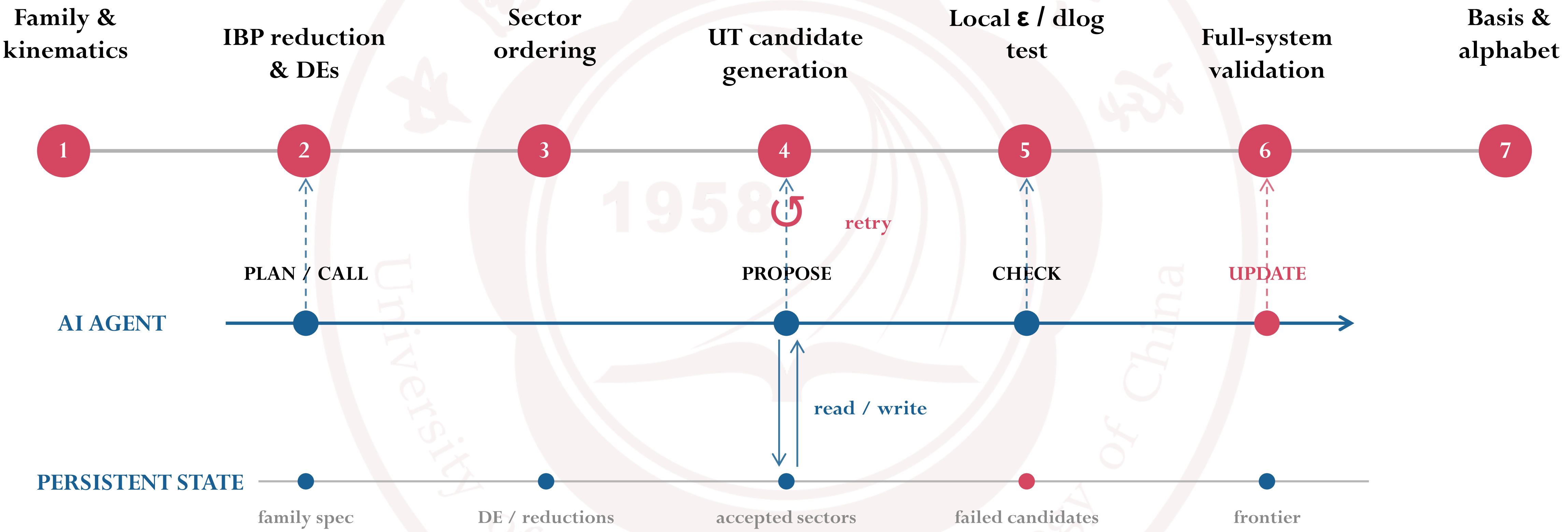
Natural language → 12-phase SMEFT phenomenology

[Guo, Wang, Yang, 2026]

Common denominator: deterministic physics tools + explicit state + verifiable artifacts.

CANON: From an Integral Family to a Canonical Basis

The physics workflow stays familiar; the agent supplies the loop and the memory.

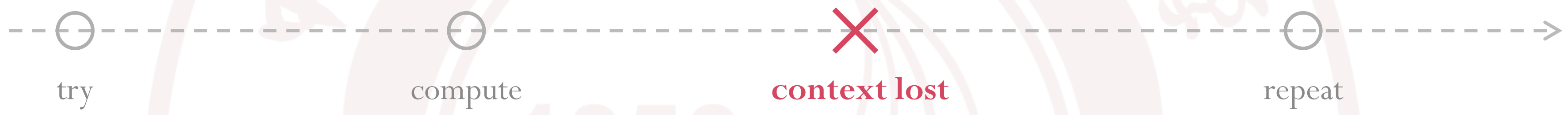


Failed checks become recorded evidence—and the starting point of the next attempt.

State as Memory of the Scientific Process

The hard part is not one clever prompt — it is surviving hundreds of dependent decisions.

Stateless chat



Stateful research

State is not a transcript. It is a scientific ledger of claims and evidence.

One Goal, Many Routes to a UT Basis

Different sectors reward different ideas; no single construction dominates.

D-log integrands
DlogBasis / integrand analysis

leading singularities
maximal cuts, constant residues

Baikov / Gram
loop-by-loop or full Baikov

UT candidate

DE transformations
direct algorithms to epsilon form

Single-UT seed
propagate a known pure integral

known subtopologies
normalization & dimension shifts

Method card : plugin-like multi-strategy integration of all known methods

[Henn, 2013] [Meyer, 2017] [Dlapa, Henn, Yan, 2020]
[Dlapa, Li, Zhang, 2021] [Chen, Jiang, Ma, Xu, Yang, 2022] [epsilon collaboration, 2025]

Human Iteration v.s. Agentic Search

The algebraic check is deterministic; repeated proposal and repair is the bottleneck.

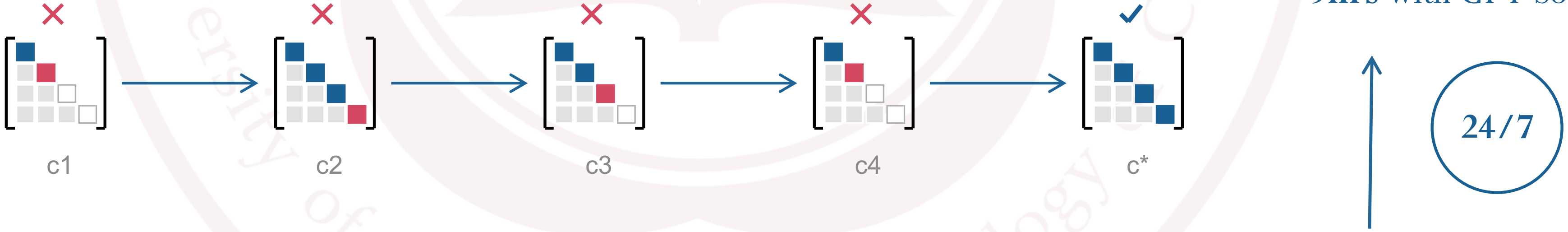
Unpublished family, **40 integrals**
1+ days with human

HUMAN

A good candidate may be nearby, but the loop consumes attention and time.

AGENT

The same family, **40 integrals**
9hrs with GPT Sol 5.6 high



failure signatures cached → next candidates reprioritized

same verifier, no fatigue

Two Ways of "Learning" Feynman Integrals

01

Learn the analytic answer

Symbolic regression reconstructs exact symbol letters from numerical CDE data.

[Liu, Xu, Zhang, 2025]

02

Learn the construction process

A stateful agent coordinates domain methods to build and verify a UT / canonical basis.

CANON ---- work in progress

AI reliably proposes. Human easily decides.

Thank you!

Yuanche Liu, Dept. of Mod. Phys., **USTC**

2026.08.27

MICA: Shared Memory for Mathematica Agents

A local MCP bridge that lets an agent work inside the real Wolfram Desktop notebook.

MICA

Mathematica Interactive Control Agent

REAL NOTEBOOKS

Definitions, nearby cells, outputs and messages remain in the working notebook.

VISIBLE EVIDENCE

Every insertion, evaluation and error stays visible and reviewable by the human.

SAFE CONTROL

Notebook targeting, explicit permissions, local-only communication.

AI → MICA →

coding agent

local MCP bridge

discover / read / edit

run / inspect / continue

Not another detached raw-eval endpoint.

CANON.nb kernel: local

```
In[37]: candidate = GramDet[...]

Out[37]: epsilon-form?  False

message: retry with another method

In[38]: candidate = DlogBasis[...]
Out[38]: epsilon-form?  True
```

The next call continues from here.

`npm install -g @aliceshimada/mica` | `mica install` | `mica mcp`

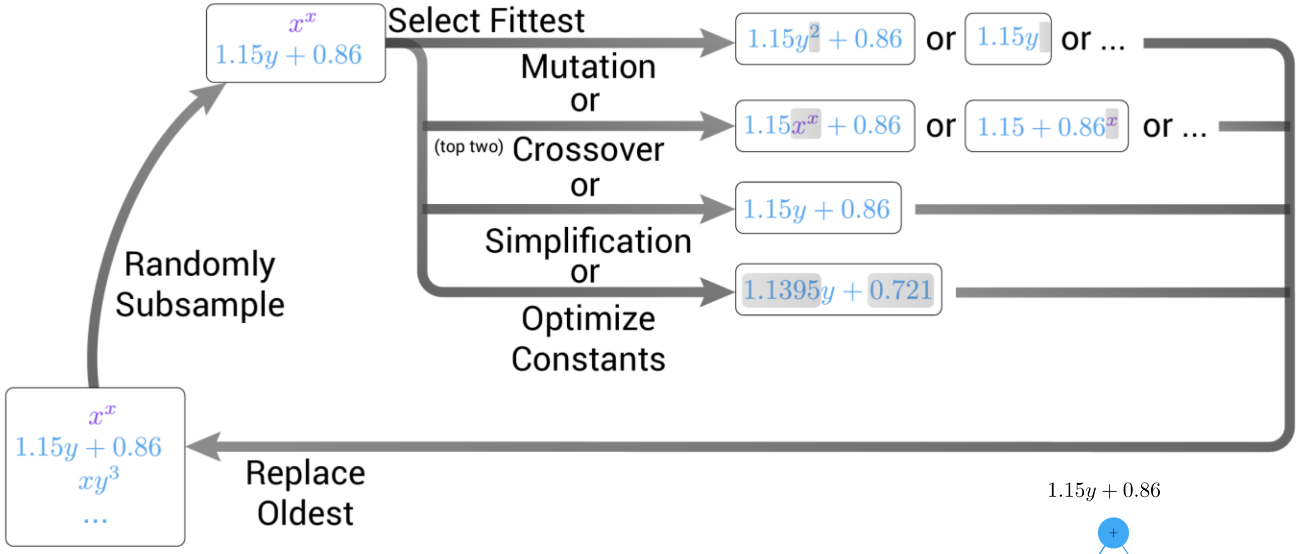
https://github.com/Alice-Shimada/mica [Liu, 2026]

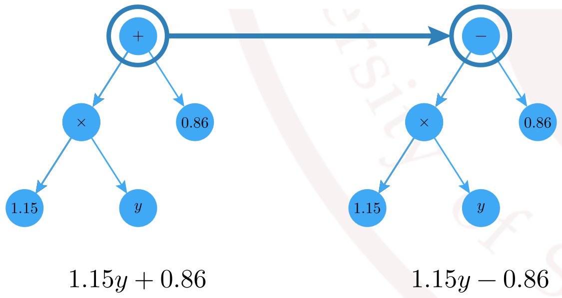
Summary: Symbolic Regression

- **Symbolic Regression**
  - **Machine-Learning** technique for **translating "numerical" back to "analytical"**
  - **Pros: Explainable, feasible**
  - **Cons: Multi-scales difficulties (all ML methods have!)**
- **Numerical CDE + Symbolic Regression = Full Symbol Letters**
  - **Completely successful** in **<5 scales problems !**
  - **Cutting-Edge** 3l5p problems: **All even letters obtained!**

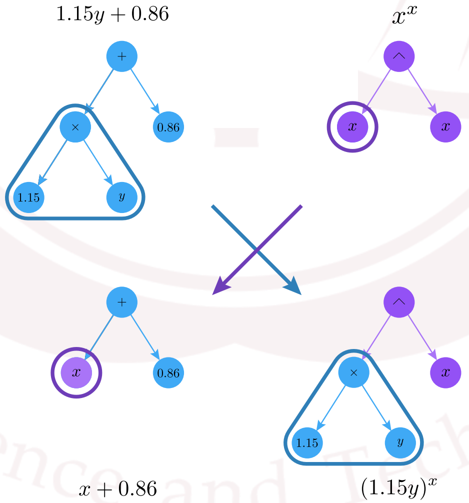
#Loop \ Family	1 Scale `3p1m,4p0m`	2 Scales `4p1m`	3 Scales `4p2m`	5 Scales `5p0m`	5+ Scales `5p1m, 6p0m`
1	✓✓	✓✓	✓✓	✓	✓
2	✓✓	✓✓	✓✓	✓	✗
3	✓✓	✓✓	✓✓	✓	——
4	✓✓	——	——	——	——

PySR: Symbolic Regression Toolkit

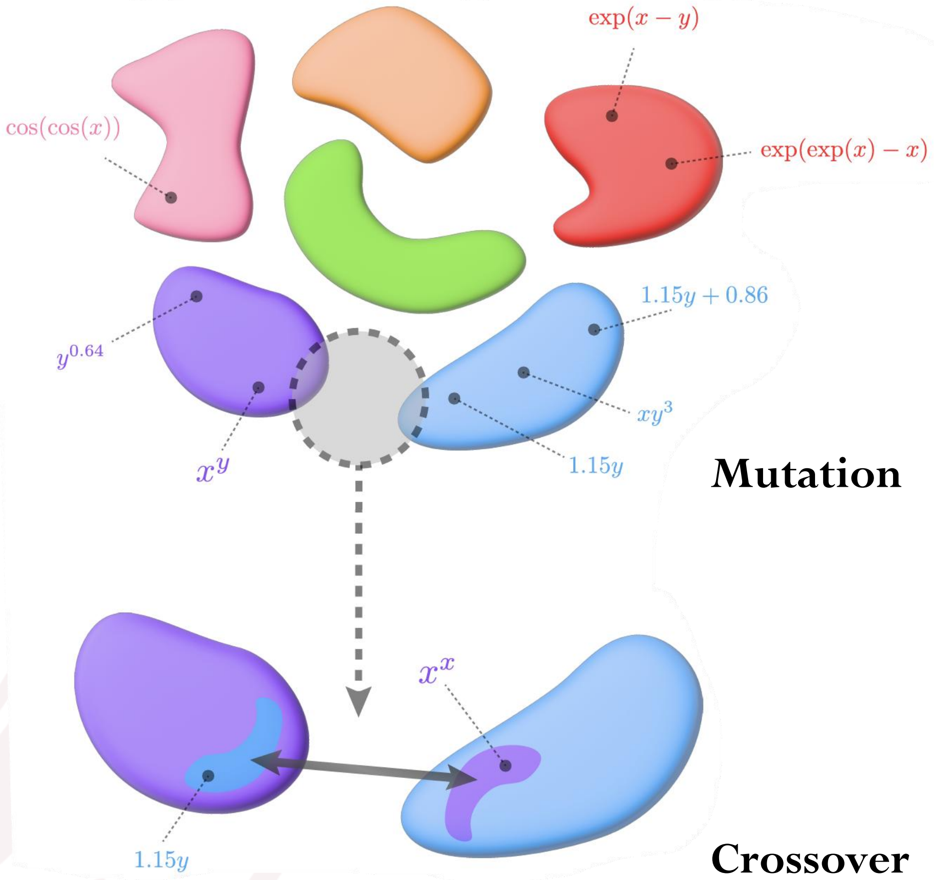




Mutation on an island



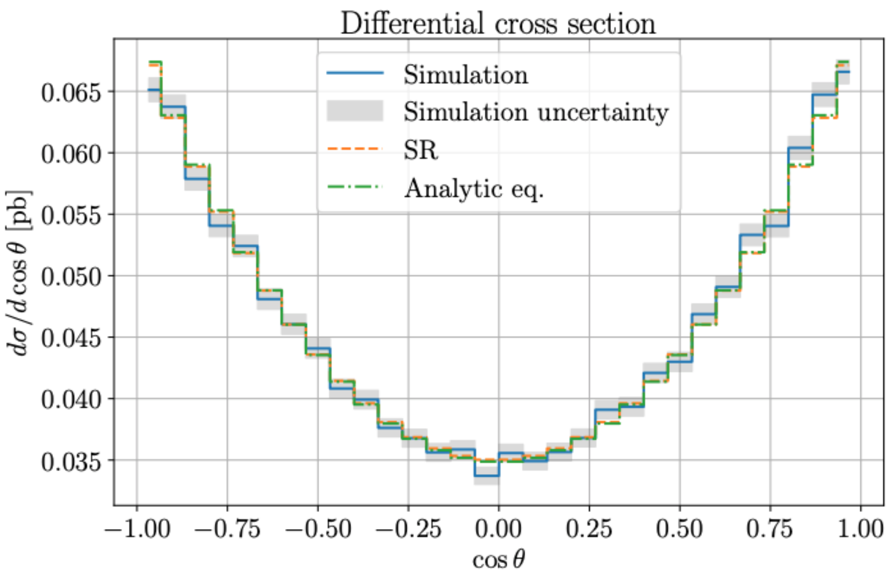
Crossover between islands

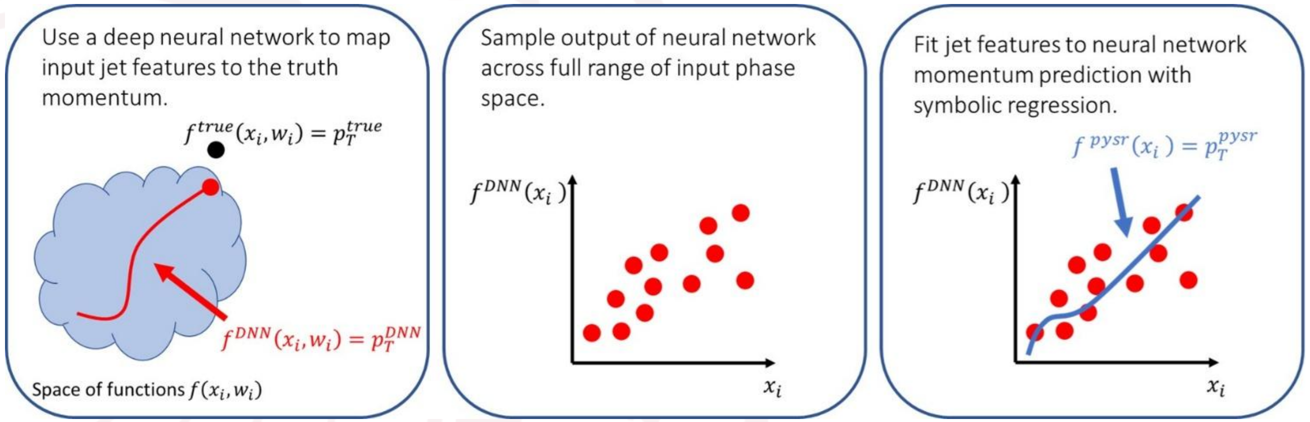


PySR Working Procedures

Diagrams from [Cranmer, 2023]

Working With Symbolic Regression





[Symbolic regression for precision LHC physics]

[Interpretable machine learning methods applied to jet background subtraction in heavy-ion collisions]

See also: Research | PySR