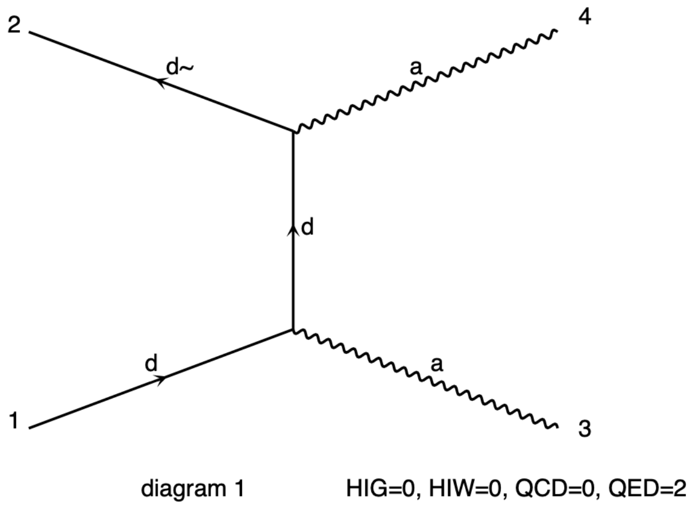
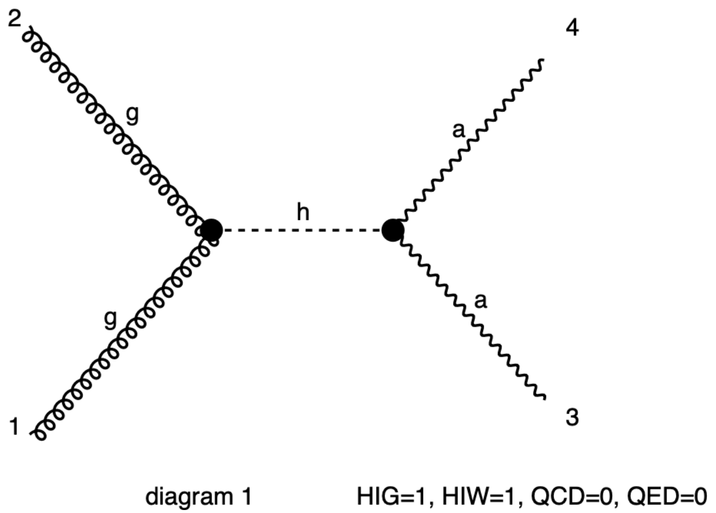
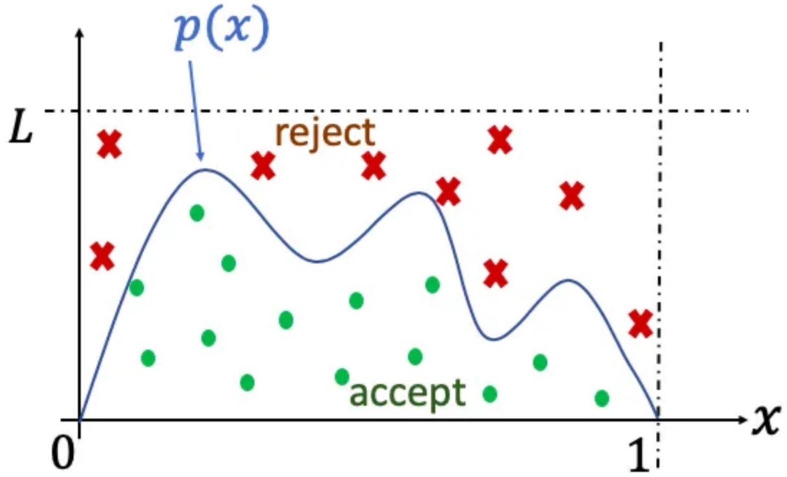






ERROR ESTIMATION OF NUMERICAL INTEGRATIONS AND ITS APPLICATION IN MADGRAPH

JINSEOK YOO (DEPARTMENT OF PHYSICS, KOREA UNIVERSITY)

JULY 12TH, 2024

2024 KOREA-CERN SUMMER STUDENT PROGRAM

INDEX

INTRODUCTION

Personalities

- Undergraduate student

@Korea University, Dept. of physics / Dept. of mathematics

- 2020 ~ 2021

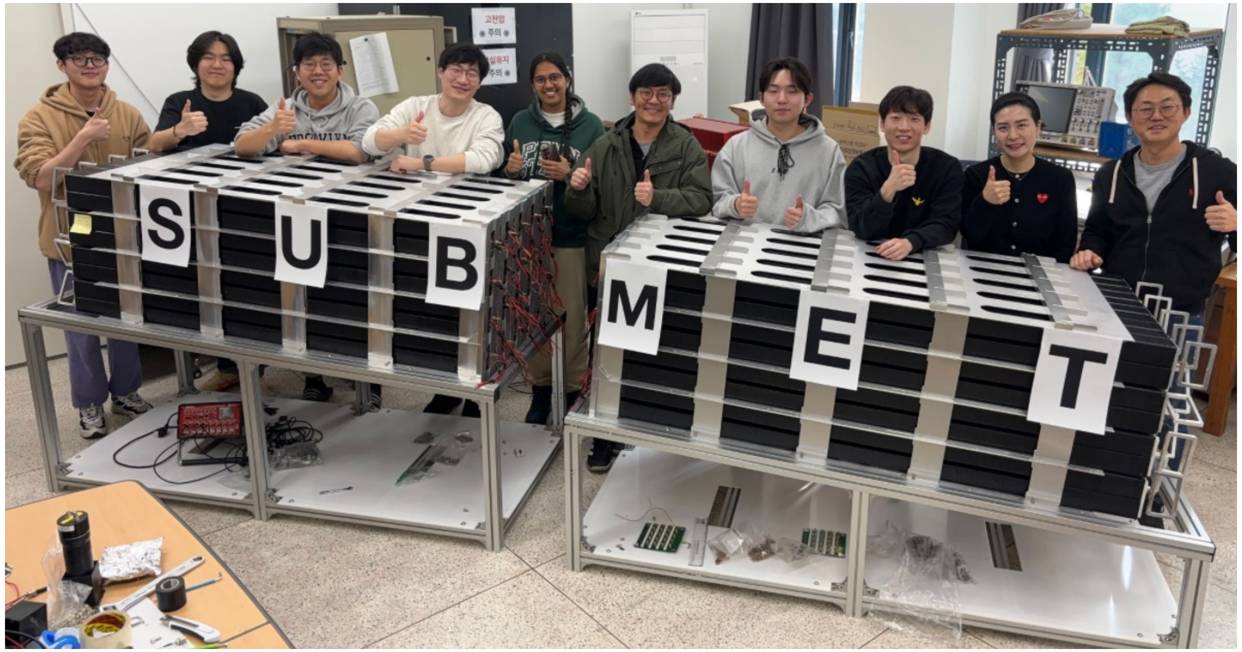
Undergraduate Researcher

@ KU Nuclear Physics Lab. (Prof B. Hong)

- 2023 ~

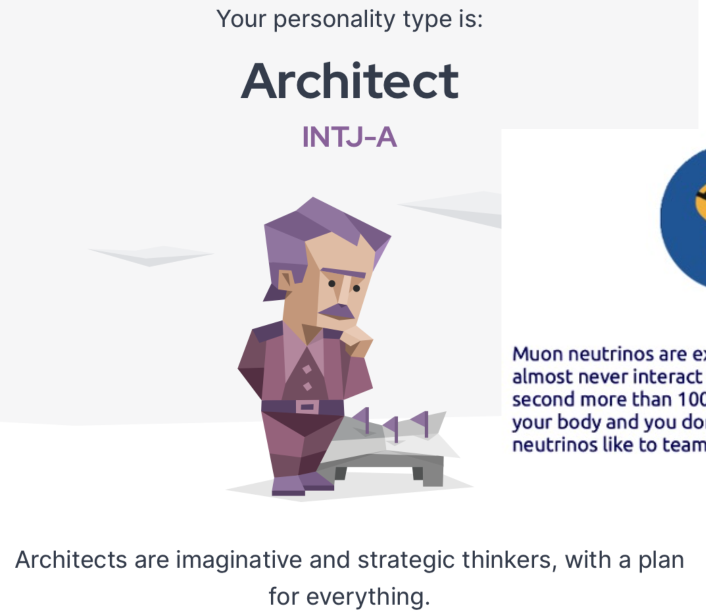
Undergraduate Researcher

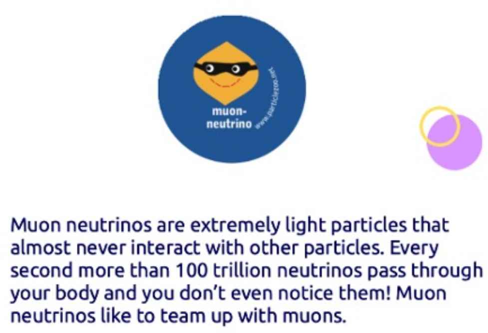
@ KU High Energy Physics Experiment Lab. (Prof J.Yoo)



INTRODUCTION

Personalities





ESTJ -> INFJ -> INTJ (9th July, 2024)

INDEX

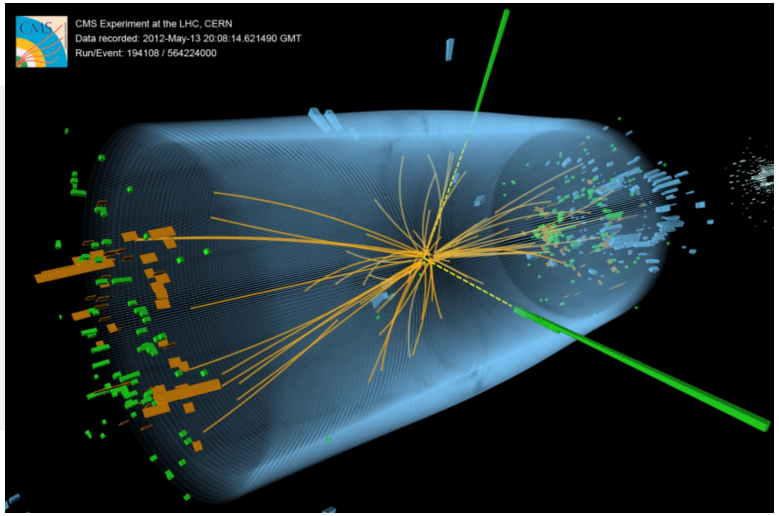
- I. Self Introduction
- II. Motivation
- III. Verification of the error rate for three numerical integration method : MonteCarlo, trapezoidal, simpson
- IV. Confirmation
- V. Brief introduction of MADGraph

MOTIVATION

Why we need numerical method?

- Imaginary

$$\mathscr{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}D\psi + h.c. + \bar{\psi}_i y_{ij}\psi_j\phi + h.c. + |D_\mu\phi|^2 - V(\phi)$$



Beautiful!

- Real

$$\sigma = \frac{1}{2s}\int f(x_1)f(x_2)|M|^2\, d^3P_1 \ldots d^3P_n \delta^4(P - p_1 - p_2 \ldots - p_n)$$

$$M \approx \langle \mu^+\mu^- | H_{\text{int}} | e^+e^- \rangle + \ldots$$

Not easy to calculate analytically…

↓

We need numerical method!

MOTIVATION

3 ways to integrate numerically

- In the case of integration, the error reduces as…
  - $O\left(\frac{1}{\sqrt{N}}\right)$ for MonteCarlo integration
  - $O\left(\frac{1}{N^2}\right)$ for Trapezoidal integration
  - $O\left(\frac{1}{N^4}\right)$ for Simpson's 1/3 rule

Result of $\int_0^1 e^x\,dx = e - 1 \sim 1.71828$

Number of points	Monte Carlo Result	Monte Carlo Error	Trapezoidal Result	Trapezoidal Error
100	1.61237	0.105907	1.71828	1.4319e-07
200	1.66695	0.0513285	1.71828	1.4319e-07
300	1.67652	0.0417618	1.71828	1.4319e-07
400	1.69452	0.0237582	1.71828	1.4319e-07
500	1.71341	0.00486871	1.71828	1.4319e-07

MOTIVATION

Advantage of monte carlo

- But in MADGraph, it uses MonteCarlo integration!
  - Why? : MonteCarlo integration is independent of dimension.

- $O\left(\frac{1}{\sqrt{N}}\right)$ for MonteCarlo integration
- $O\left(\frac{1}{N^{2/n}}\right)$ for Trapezoidal integration → If $n > 8$, MonteCarlo integration is better than the others.
- $O\left(\frac{1}{N^{4/n}}\right)$ for Simpson's 1/3 rule

INDEX

- I. Introduction
- II. Motivation
- III. Verification of the error rate for three numerical integration method : MonteCarlo, trapezoidal, Simpson
- IV. Confirmation
- V. Brief introduction of MADGraph

MONTECARLO INTEGRATION

Method

- $I = \int_\Omega f(\vec{x})d^n\vec{x}$ where $\Omega \subset \mathbb{R}^n$
- $V = \int_\Omega d^n\vec{x}$: Volume
- For given N samples, $\overrightarrow{x_1}, \overrightarrow{x_2}, \dots, \overrightarrow{x_N} \in \Omega$,

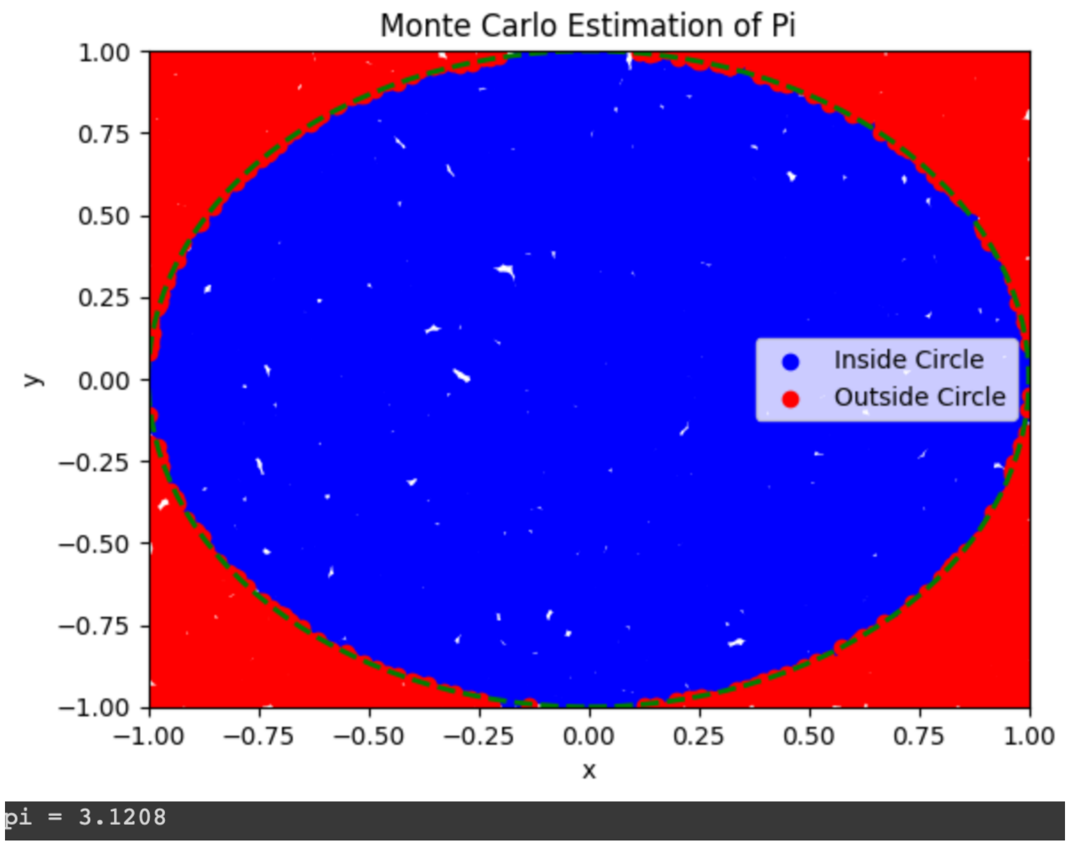
$$I \sim \frac{V}{N} \Sigma_{i=1}^{N} f(\overrightarrow{x_i}) \equiv Q_N$$

Example :

$$\Omega = [-1,1]\times[-1,1] \subset \mathbb{R}^2,$$

$$f(x,y) = \begin{cases} 1, if\ x^2 + y^2 \leq 1 \\ 0, otherwise \end{cases}$$

$$Q_N = \frac{4}{N}\Sigma_{i=1}^{N} f(\overrightarrow{x_i}) \sim \pi$$



MONTECARLO INTEGRATION

Prove of error estimation

Suppose $\vec{x_i}s$ are i.i.d (independent and identically distributed).

Define $y_i \equiv V \times f(x_i)$, where $V = \int_\Omega d^n\vec{x}$. Then

$$Q_N \equiv \frac{V}{N}\Sigma_{i=1}^N f(\vec{x_i}) = \frac{y_1 + y_2 + \cdots + y_N}{N}$$

Define $p(y_i)$ be the probability density function(pdf) of y_i, and $\mathrm{P_N}(Q_N)$ be the distribution of Q_N, then

$$P_N(Q_N) = \int dy_1 \cdots dy_n\, p(y_1)\cdots p(y_N)\delta(Q_N - \frac{1}{N}\Sigma y_i)$$

Define Fourier transform of $p(y)$ as

$$\phi(k) = \int dy\, e^{ik(y - <y>)}p(y)$$

Likewise, for $P_N(Q_N)$,

$$\Phi_N(k)$$

$$= \int dy_1 \cdots dy_N\, e^{i\left(\frac{k}{N}\right)(y_1 - <y> + \cdots + y_N - <y>)}p(y_1)\cdots p(y_N)$$

$$= \left[\phi\left(\frac{k}{N}\right)\right]^N$$

MONTECARLO INTEGRATION

Prove of error estimation

By expansion,

$$\phi\left(\frac{k}{N}\right) = \int dy\, e^{i\left(\frac{k}{N}\right)(y - <y>)} p(y) = 1 - \frac{k^2\sigma^2}{2N^2} + \cdots$$

Then,

$$\Phi_N(k) = \left(1 - \frac{k^2\sigma^2}{2N^2} + O\left(\frac{k^3}{N^3}\right)\right)^N \to e^{-k^2\sigma^2/2N}$$

By using inverse transform, we have

$$P_N(Q_N) = \frac{1}{2\pi}\int dk\, e^{-ik(Q_N - <y>)} \Phi_N(k)$$

$$= \frac{1}{2\pi}\int dk\, e^{ik(Q_N - <y>)} e^{-\frac{k^2\sigma^2}{2N}}$$

$$= \frac{\sqrt{N}}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left(-\frac{N(Q_N - <y>)^2}{2\sigma^2}\right)$$

Taking $N \to \infty$, $P_N(Q_N)$ approaches to gaussian distribution with

$$\sigma_N = \frac{\sigma}{\sqrt{N}}$$

Independent of dimension!

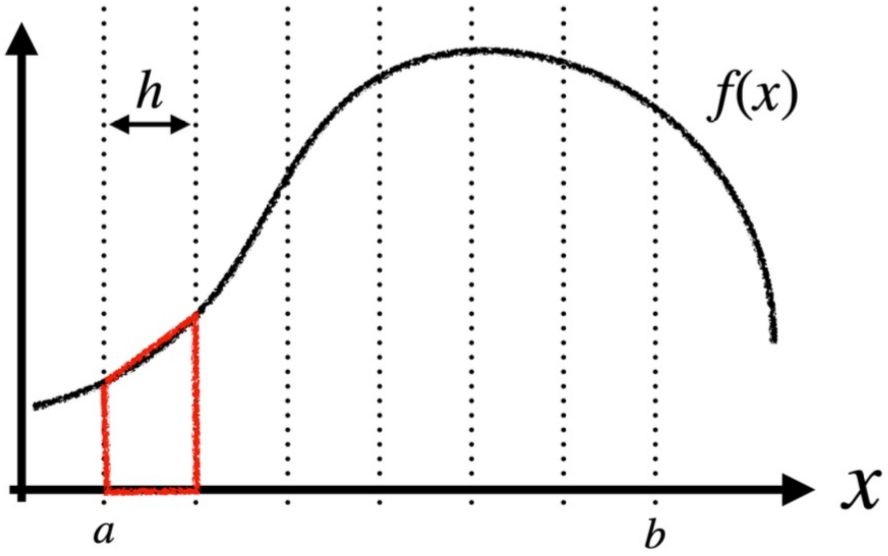
TRAPEZOIDAL RULE

Method

$$\int_a^b f(x)\,dx \approx \sum_{k=1}^{N} \frac{f(x_{k-1}) + f(x_k)}{2} \Delta x_k .$$

$$= h\frac{1}{2}\sum_{k=1}^{N}\left[f(a+(k-1)h) + f(a+kh)\right] \quad \text{where } h = \frac{b-a}{N}$$

$$= h\left[\frac{1}{2}f(a) + \frac{1}{2}f(b) + \sum_{k=1}^{N-1} f(a+kh)\right]$$



TRAPEZOIDAL RULE

Prove of error estimation in 1-D

Define

$$g_k(t) = \frac{1}{2}t[f(a_k) + f(a_k + t)] - \int_{a_k}^{a_k+t} f(x)\,dx$$

where $h = \frac{b-a}{N}$ and $a_k = a + (k-1)h$.

Then,

$$\frac{dg_k}{dt} = \frac{1}{2}[f(a_k) + f(a_k+t)] + \frac{1}{2}t \cdot f'(a_k+t) - f(a_k+t),$$

$$\frac{d^2 g_k}{dt^2} = \frac{1}{2}t \cdot f''(a_k+t).$$

Suppose $f''(x)$ is bounded,
i.e. $\exists \xi$ $s.t.$ $|f''(x)| \leq |f''(\xi)|$,
It follows that

$$-f''(\xi) \leq f''(a_k+t) \leq f''(\xi),$$

$$-\frac{f''(\xi)t}{2} \leq g_k''(t) \leq \frac{f''(\xi)t}{2}.$$

Note that $g_k(0) = 0, g_k'(0) = 0,$
By integrating, we have

$$-\frac{f''(\xi)t^3}{12} \leq g_k(t) \leq \frac{f''(\xi)t^3}{12}$$

TRAPEZOIDAL RULE

Prove of error estimation in 1-D and n-D

By summing, and take $t = h$, we have

$$-\frac{f''(\xi)h^3N}{12} \leq \frac{b-a}{N}\left[\frac{f(a)+f(b)}{2}+\sum_{k=1}^{N-1} f\left(a+k\frac{b-a}{N}\right)\right]-\int_a^b f(x)dx \leq \frac{f''(\xi)h^3N}{12}.$$

Conclusion :

$$\text{error} = \int_a^b f(x)\,dx - \frac{b-a}{N}\left[\frac{f(a)+f(b)}{2}+\sum_{k=1}^{N-1} f\left(a+k\frac{b-a}{N}\right)\right] = \frac{f''(\xi)h^3N}{12} = \frac{f''(\xi)(b-a)^3}{12N^2}.$$

For n-dimensional integration, we have to repeat the calculation n times, therefore

$$error \propto \frac{1}{N^{2/n}}$$

SIMPSON'S RULE

Method

$$P(x) = Ax^2 + Bx + C$$

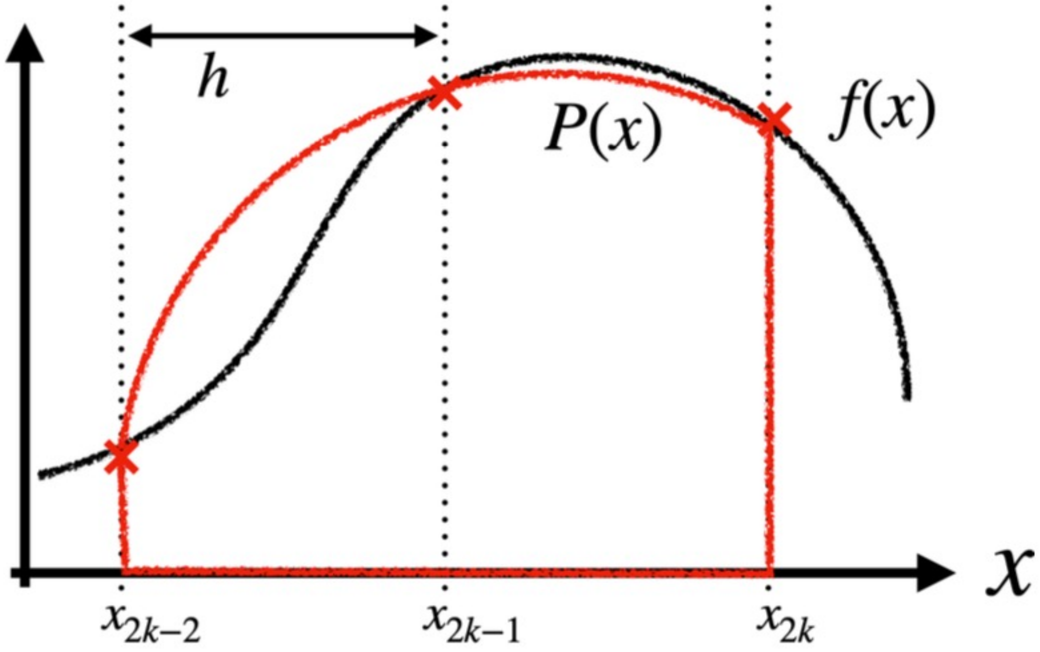
$$= f(\alpha)\frac{(x-m)(x-\beta)}{(\alpha-m)(\alpha-\beta)} + f(m)\frac{(x-\alpha)(x-\beta)}{(m-\alpha)(m-\beta)} + f(\beta)\frac{(x-\alpha)(x-m)}{(\beta-\alpha)(\beta-m)}$$

By integrating (use integration by substitution),

$$\int_\alpha^\beta f(x)dx \approx \int_\alpha^\beta P(x)dx = \frac{\beta-\alpha}{6}\left[f(\alpha) + 4f\left(\frac{\alpha+\beta}{2}\right) + f(\beta)\right]$$

$$x_{2k-2} = \alpha, x_{2k-1} = m, x_{2k} = \beta$$

$$h = \frac{\beta - \alpha}{2}$$



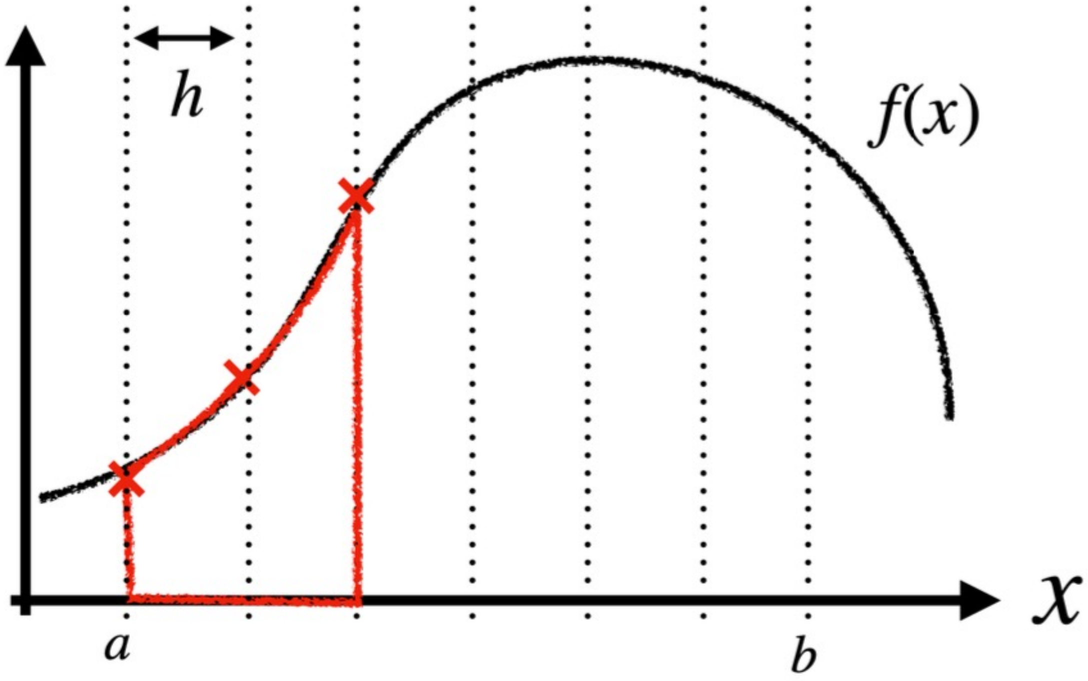
$$\int_{x_{2k-2}}^{x_{2k}} f(x)dx \approx \frac{h}{3}\left[f(x_{2k-2}) + 4f(x_{2k-1}) + f(x_{2k})\right]$$

SIMPSON'S RULE

Method

If we increase the number of slices,

$$I(a,b) \approx \frac{h}{3}\left[f(a) + f(b) + 4\sum_{k=1}^{N/2} f(x_{2k-1}) + 2\sum_{k=1}^{N/2-1} f(x_{2k})\right]$$



SIMPSON'S RULE

Prove of error estimation in 1-D

- Start with the simplest case : only two slices

$$E = I - \frac{h}{3}(f_o + 4f_1 + f_2) \quad \Rightarrow \quad E = \int_0^{2h} f(x)dx - \frac{h}{3}(f_o + 4f_1 + f_2)$$

Where $x_{2k-2} = x_0$, $x_{2k-1} = x_1$, $x_{2k} = x_2$, and

$f_0 = f(x_0), f_1 = f(x_1), f_2 = f(x_2)$

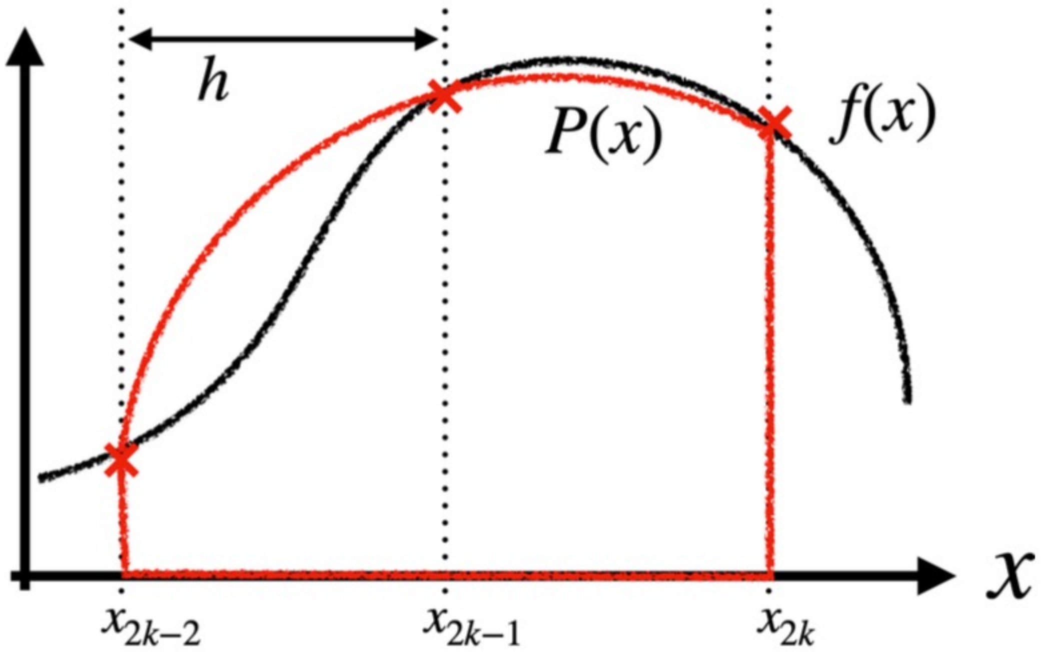
Taylor expansion of $f(x)$ about $x = h$ is

$$f(x) = f_1 + (x-h)f_1^{(1)} + \frac{1}{2}(x-h)^2 f_1^{(2)} + \frac{1}{6}(x-h)^3 f_1^{(3)} + \frac{1}{24}(x-h)^4 f_1^{(4)} + O(x-h)^5$$

$$f_o = f_1 - h f_1^{(1)} + \frac{1}{2}h^2 f_1^{(2)} - \frac{h^3}{6} f_1^{(3)} + \frac{1}{24}h^4 f_1^{(4)} + O(h)^5$$

$$f_1 = f_1$$

$$f_2 = f_1 + h f_1^{(1)} + \frac{1}{2}h^2 f_1^{(2)} + \frac{h^3}{6} f_1^{(3)} + \frac{1}{24}h^4 f_1^{(4)} + O(h)^5$$



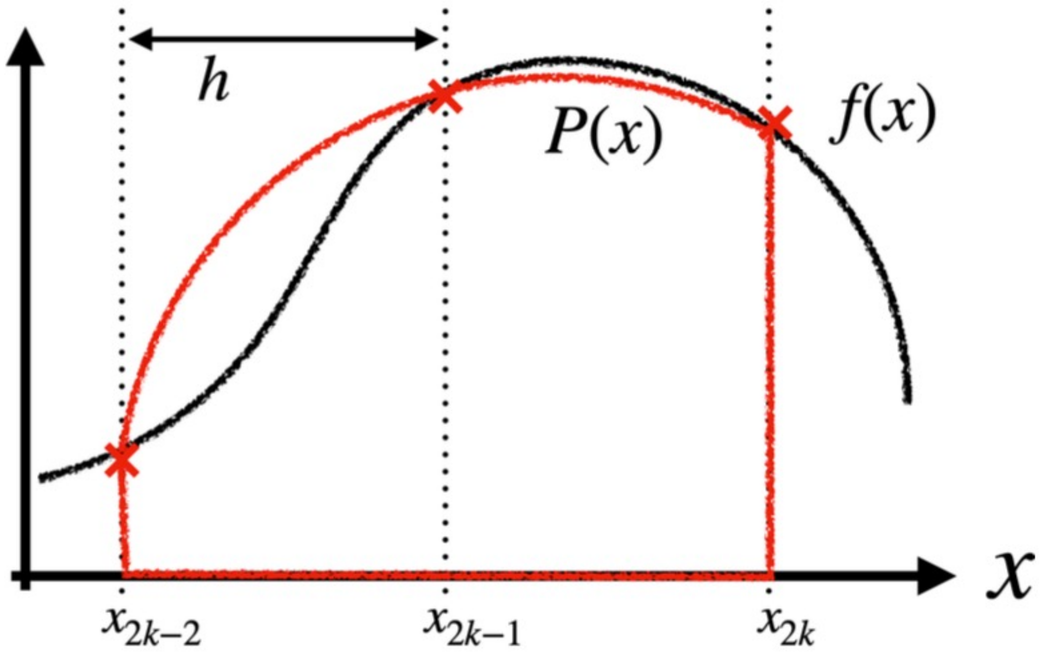
SIMPSON'S RULE

Prove of error estimation in 1-D

$$E = I - \frac{h}{3}(f_o + 4f_1 + f_2) \quad \Rightarrow \quad E = \int_0^{2h} f(x)dx - \frac{h}{3}(f_o + 4f_1 + f_2)$$

If we replace $f(x)$ as above, we have

$$E = -\frac{1}{90} h^5 f_1^{(4)}$$



SIMPSON'S RULE

Prove of error estimation in 1-D and n-D

Note that

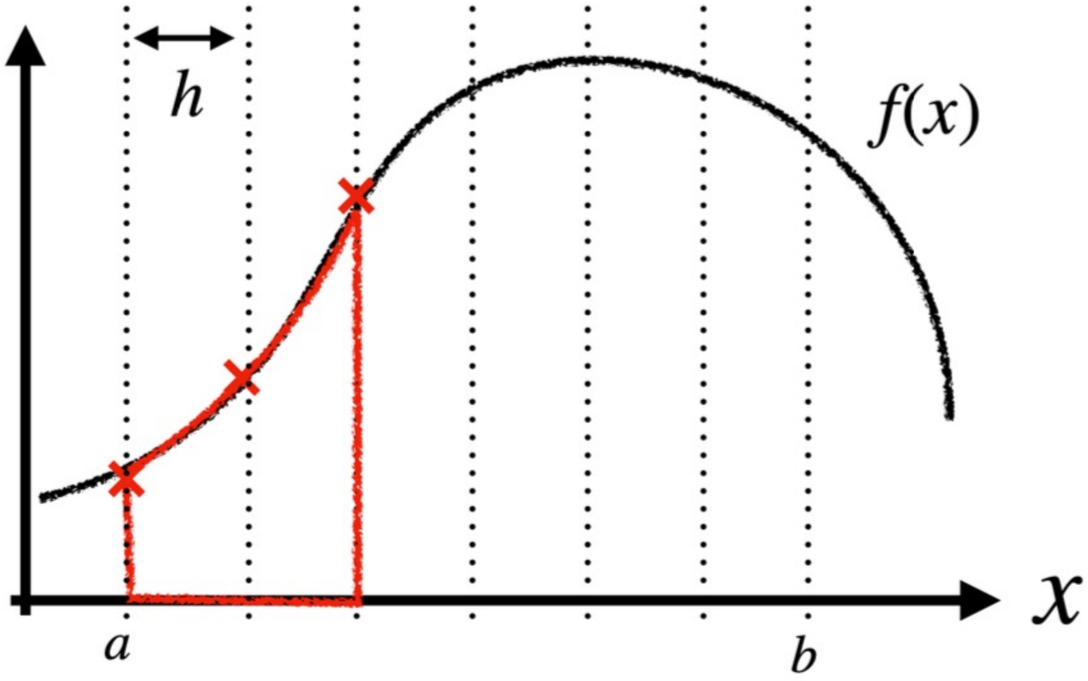
$$E = -\frac{1}{90} h^5 f_1^{(4)}$$

for only two slices.

Now consider N slices, then

$$E_{[a,b]} \approx -\frac{h^5}{90}\left(\frac{N}{2}\right)\left(\frac{2}{N}\sum_{i=1}^{\frac{N}{2}} f_{2i-1}^{(4)}\right) \Rightarrow E_{[a,b]} \approx -\frac{h^5}{90}\frac{N}{2}\overline{f^{(4)}}$$

$$\Rightarrow \quad E_{[a,b]} \approx -\frac{h^4}{180}(b-a)\overline{f^{(4)}}$$



Conclusion :

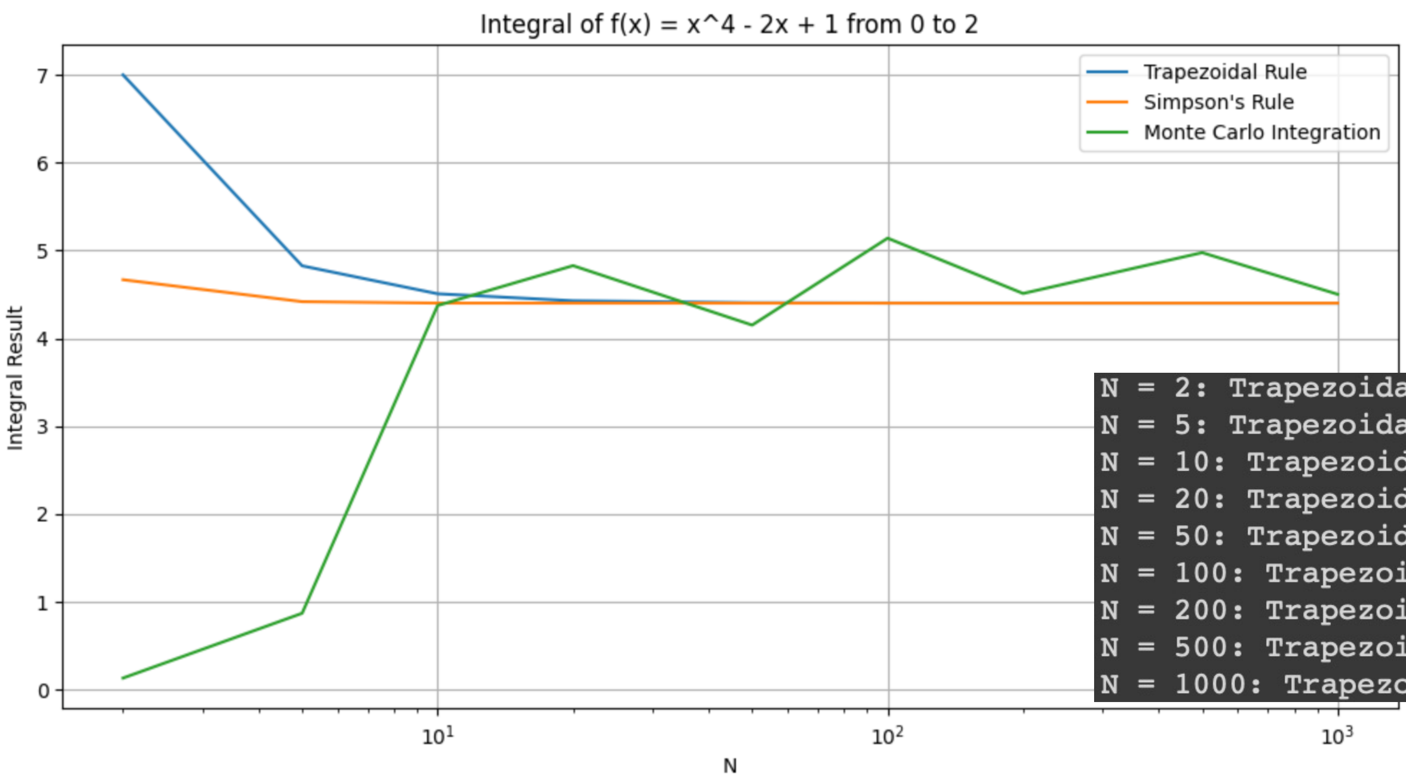
$$error \propto \frac{1}{N^{4/n}}$$ for n-dimensional

INDEX

- I. Introduction
- II. Motivation
- III. Verification of the error rate for three numerical integration method : MonteCarlo, trapezoidal, simpson
- IV. Confirmation
- V. Brief introduction of MADGraph

1D NUMERICAL INTEGRATIONS

1D integration results

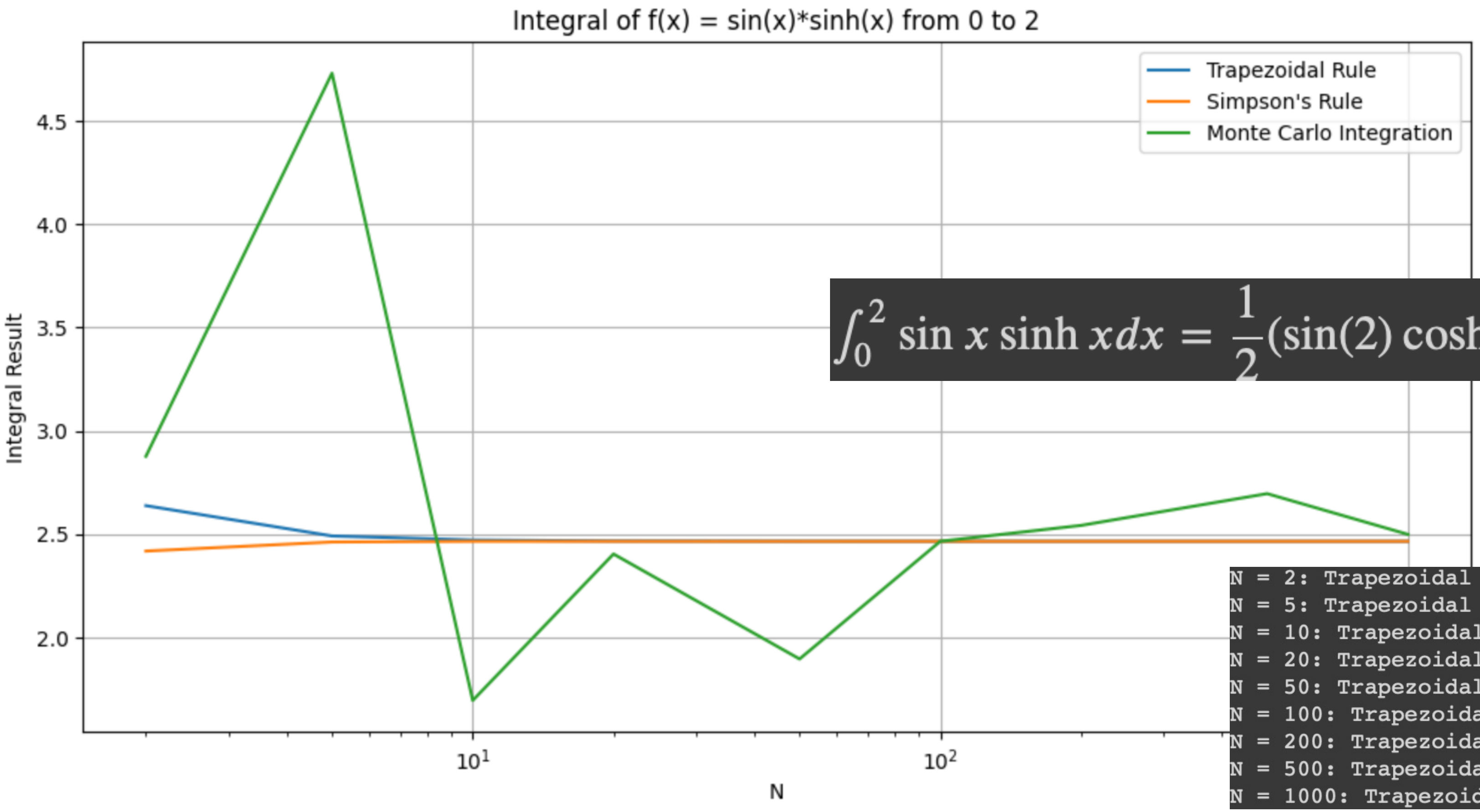


Note that $\int_0^2 x^4 - 2x + 1dx = \frac{22}{5} = 4.4$

```
N = 2: Trapezoidal = 7.00, Simpson's = 4.67, Monte Carlo = -0.14
N = 5: Trapezoidal = 4.82, Simpson's = 4.42, Monte Carlo = 2.42
N = 10: Trapezoidal = 4.51, Simpson's = 4.40, Monte Carlo = 4.40
N = 20: Trapezoidal = 4.43, Simpson's = 4.40, Monte Carlo = 3.67
N = 50: Trapezoidal = 4.40, Simpson's = 4.40, Monte Carlo = 6.55
N = 100: Trapezoidal = 4.40, Simpson's = 4.40, Monte Carlo = 4.25
N = 200: Trapezoidal = 4.40, Simpson's = 4.40, Monte Carlo = 5.02
N = 500: Trapezoidal = 4.40, Simpson's = 4.40, Monte Carlo = 4.36
N = 1000: Trapezoidal = 4.40, Simpson's = 4.40, Monte Carlo = 4.47
```

1D NUMERICAL INTEGRATIONS

1D integration results



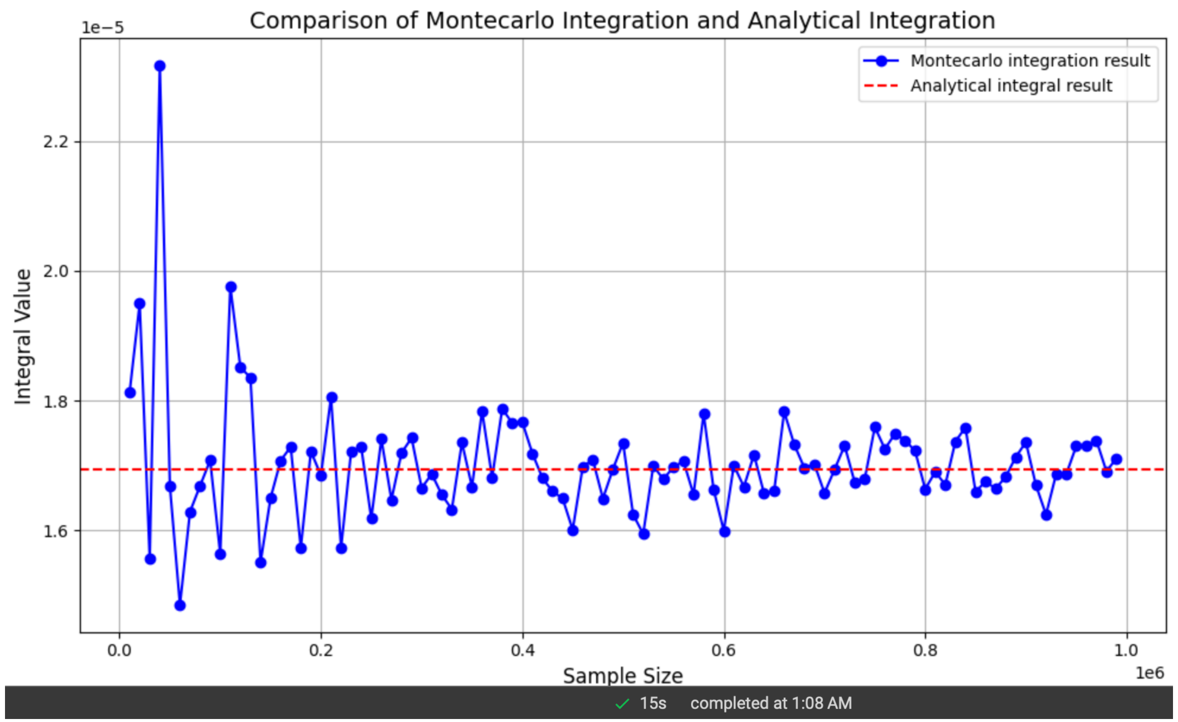
$$\int_0^2 \sin x \sinh x dx = \frac{1}{2}(\sin(2)\cosh(2) - \cos(2)\sinh(2)) \sim 2.4651$$

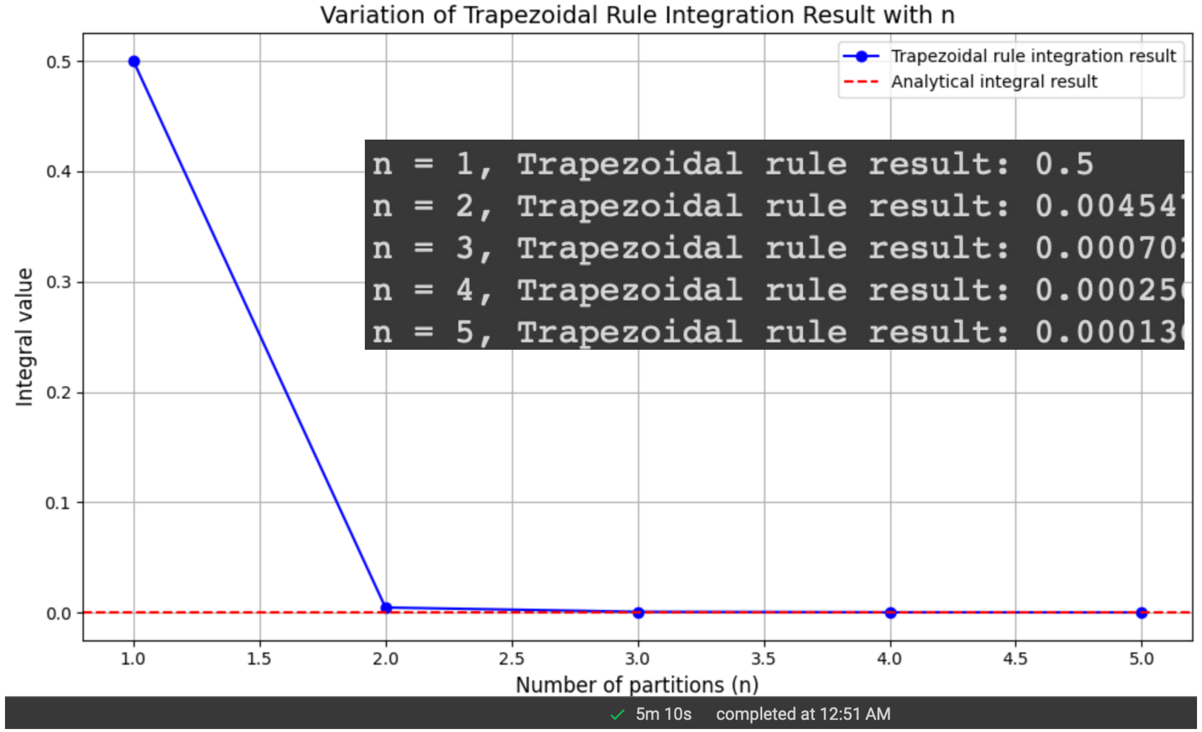
```
N = 2: Trapezoidal = 2.64, Simpson's = 2.42, Monte Carlo = 2.88
N = 5: Trapezoidal = 2.49, Simpson's = 2.46, Monte Carlo = 4.73
N = 10: Trapezoidal = 2.47, Simpson's = 2.47, Monte Carlo = 1.69
N = 20: Trapezoidal = 2.47, Simpson's = 2.47, Monte Carlo = 2.40
N = 50: Trapezoidal = 2.47, Simpson's = 2.47, Monte Carlo = 1.90
N = 100: Trapezoidal = 2.47, Simpson's = 2.47, Monte Carlo = 2.47
N = 200: Trapezoidal = 2.47, Simpson's = 2.47, Monte Carlo = 2.54
N = 500: Trapezoidal = 2.47, Simpson's = 2.47, Monte Carlo = 2.70
N = 1000: Trapezoidal = 2.47, Simpson's = 2.47, Monte Carlo = 2.50
```

10D NUMERICAL INTEGRATIONS

10D integration results

$$\int_0^1 \int_0^1 \cdots \int_0^1 \left(\prod_{i=1}^{10} x_i^2\right) dx_1 dx_2 \cdots dx_{10} = \left(\int_0^1 x^2\, dx\right)^{10} = \left(\frac{1}{3}\right)^{10} = \frac{1}{59049} \approx 0.00001693$$





N-D NUMERICAL INTEGRATIONS

Implementation

We may numerically integrate the cross section such as …

The master formula for $2 \to N$ scattering is

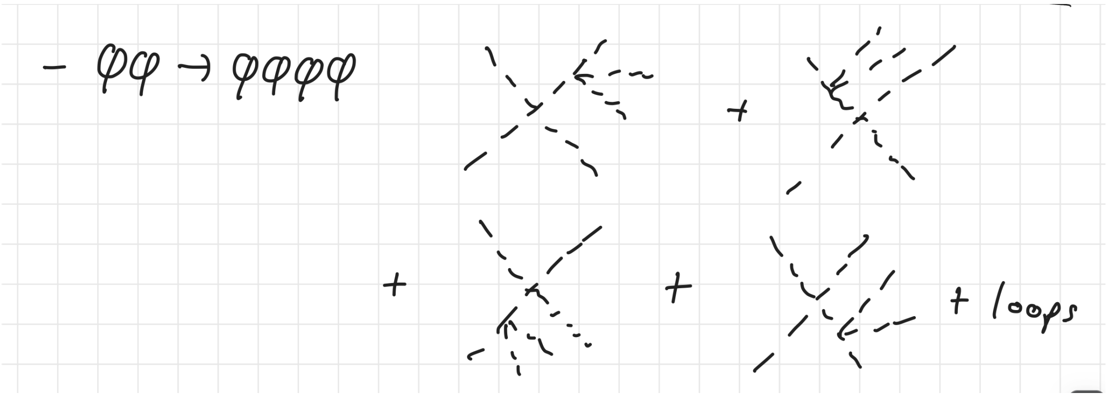
$$d\sigma = \frac{1}{2s} \prod_{i=1}^{N} d\Pi_i (2\pi)^4 \delta^{(4)}(p_A + p_B - \sum_i p_i) \cdot |\mathcal{M}|^2 \qquad d\Pi_i = \frac{d^3\mathbf{p}_i}{(2\pi)^3} \frac{1}{2E_i}. \qquad (3.1)$$

with monte carlo integration.

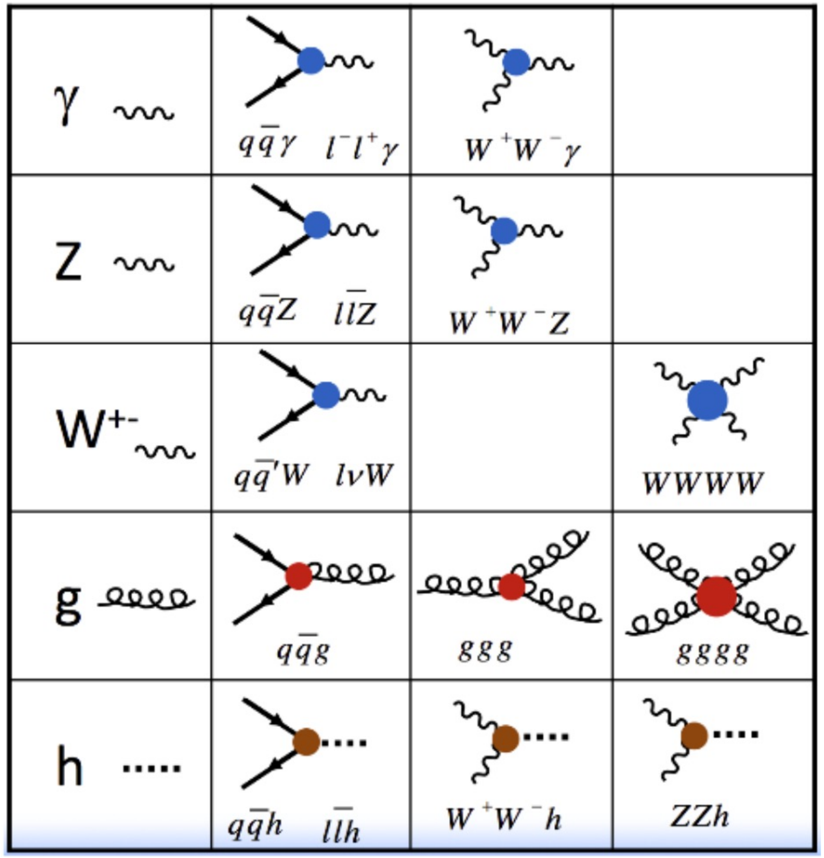
INDEX

BRIEF INTRODUCTION TO MADGRAPH

What can we do with MADGraph?



$\mathbb{Z}_2$ diagram

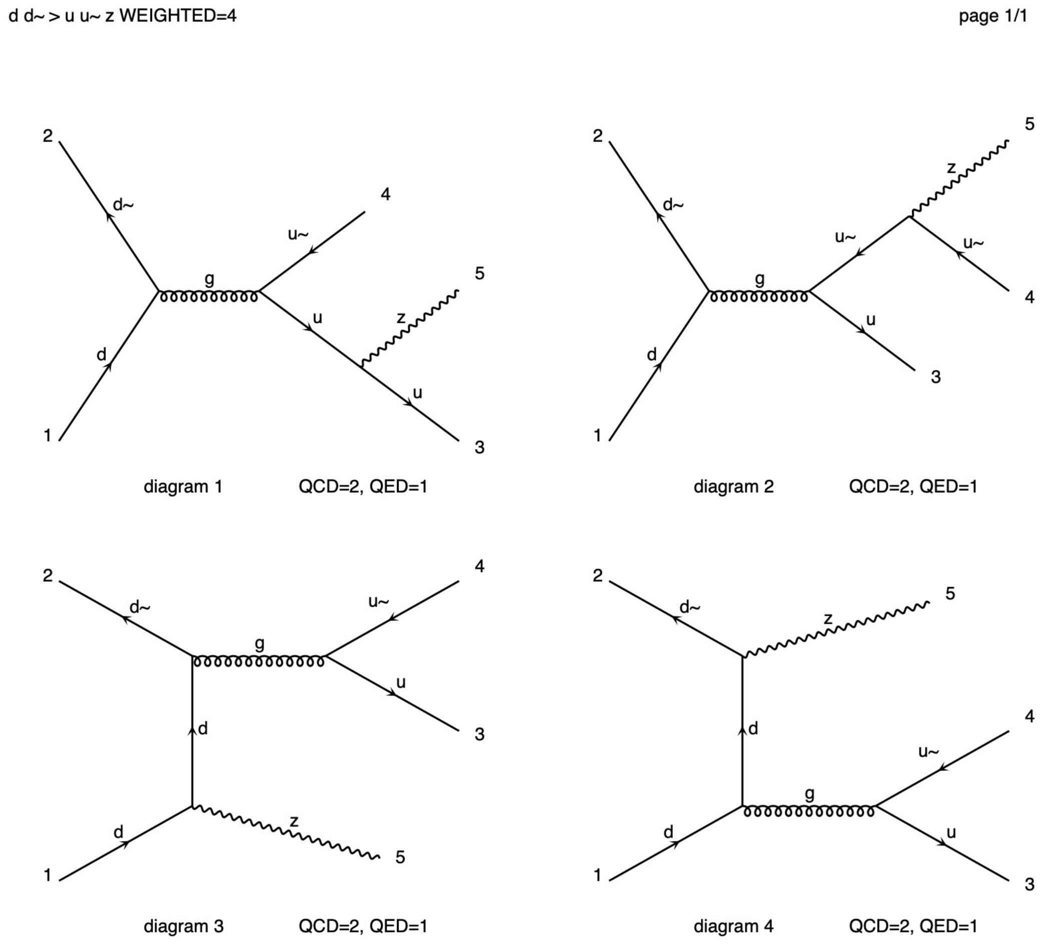


$SU(3)\times SU(2)\times U(1)$ building blocks

BRIEF INTRODUCTION TO MADGRAPH

What can we do with MADGraph?

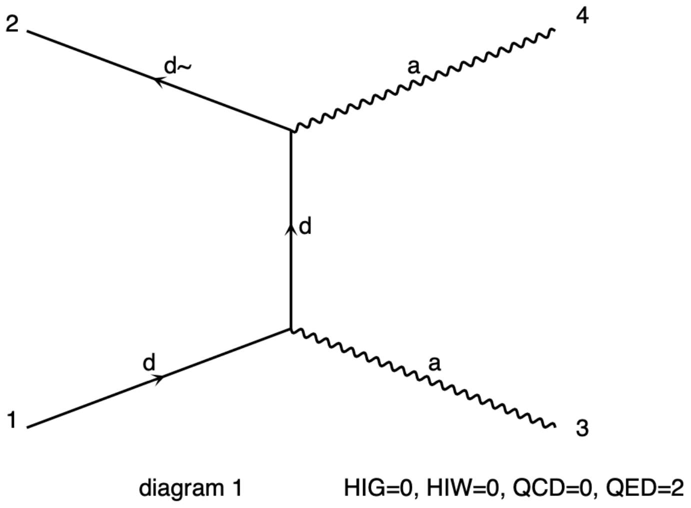
- We can draw possible Feynman diagrams very easily!!
- Example : $d\bar{d} \rightarrow u\bar{u}Z$
- After you run MADGraph, just type two lines:
  - Generate d d~ > u u~ z
  - Display diagrams
- MADGraph's calculation is based on MonteCarlo integration.



BRIEF INTRODUCTION TO MADGRAPH

What can we do with MADGraph?

- We can even see statistics about the process.
- Example : $p - p \rightarrow \gamma\gamma$
- Just type:
  - Import model heft ; generate p p > a a ; launch



2 Datasets

2.1 run_01

- Sample consisting of: signal events.
- Generated events: 10000 events.
- Normalization to the luminosity: 1490785+/- 4400 events.
- Ratio (event weight): 149 - warning: please generate more events (weight larger than 1)!

Path to the event file	Nr. of events	Cross section (pb)	Negative wgts (%)
PROC_heft_6/Events/run_01/-unweighted_events.lhe.gz	10000	149 @ 0.3%	0.0

Dataset	Integral	Entries per event	Mean	RMS	% underflow	% overflow
run_01	1490784	1.0	43.0098	32.72	0.0	0.04

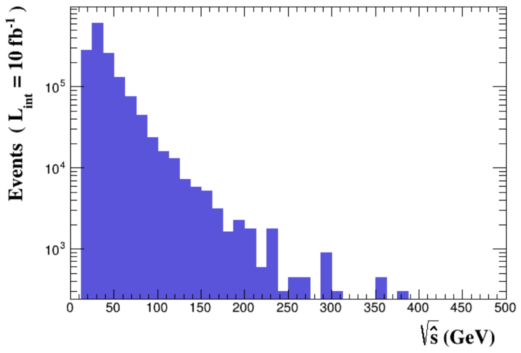


Figure 3.

BRIEF INTRODUCTION TO MADGRAPH

What can we do with MADGraph?

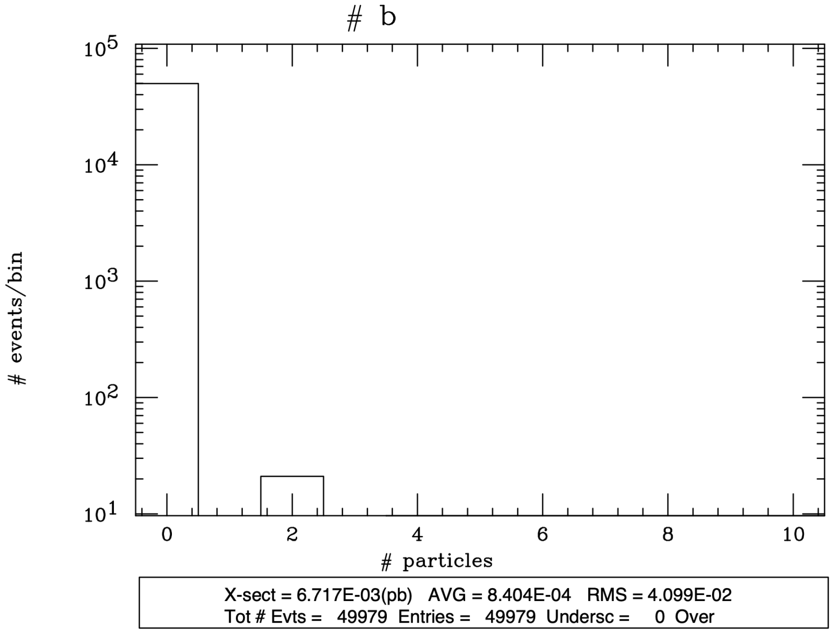
- It seems that we can modify parameters to find a new process.
- Ex :
  - https://cds.cern.ch/record/2791279/files/LHCC-P-020.pdf : Technical Proposal for the milliQan sub-detector
  - https://cds.cern.ch/record/2891837/files/JHEP06(2022)114%20(3).pdf : The Effective Vector Boson Approximation in heigh-energy muon collisions.

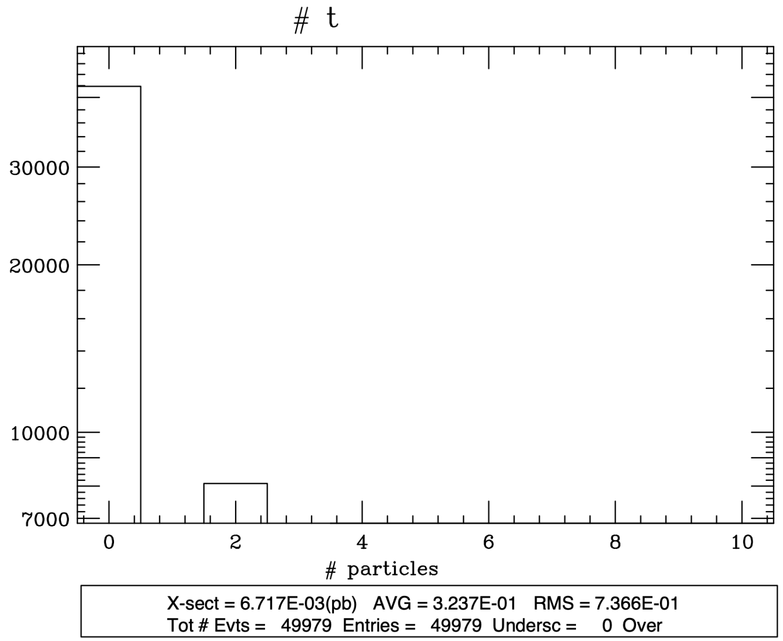
```
###################################
## INFORMATION FOR SMINPUTS
###################################
Block sminputs
    1 1.325070e+02 # aEWM1
    2 1.166390e-05 # Gf
    3 1.180000e-01 # aS (Note that Parameter not used if you use a PDF set)

###################################
## INFORMATION FOR YUKAWA
###################################
Block yukawa
    5 4.200000e+00 # ymb
    6 1.645000e+02 # ymt
   15 1.777000e+00 # ymtau

###################################
## INFORMATION FOR DECAY
###################################
DECAY   6 1.491500e+00 # WT
DECAY  23 2.441404e+00 # WZ
DECAY  24 2.047600e+00 # WW
DECAY  25 6.382339e-03 # WH
DECAY 9000006 6.382339e-03 # WH1
## Dependent parameters, given by model restrictions.
## Those values should be edited following the
## analytical expression. MG5 ignores those values
## but they are important for interfacing the output of MG5
## to external program such as Pythia.
```

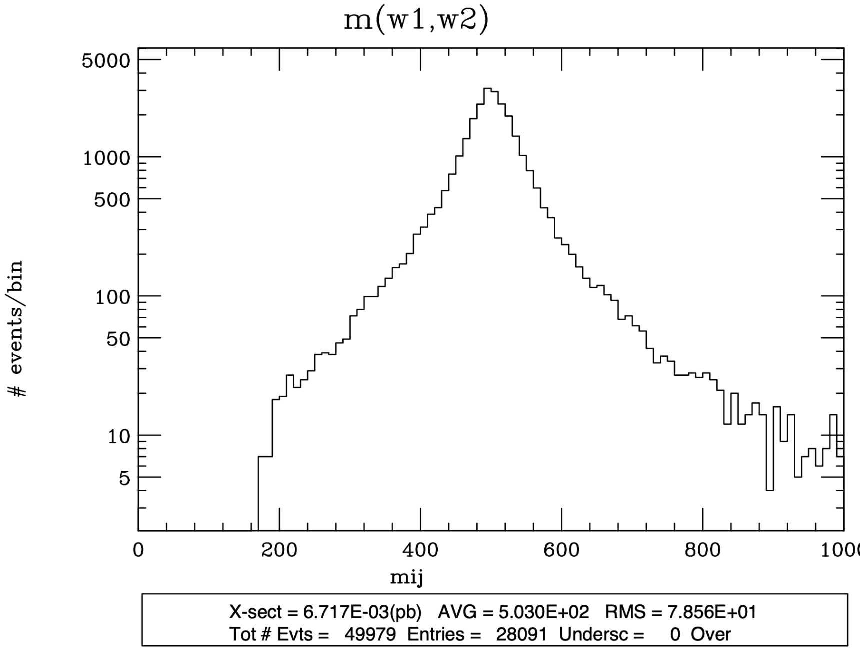
BRIEF INTRODUCTION TO MADGRAPH

What can we do with MADGraph?

- ‘Heavy Higgs like’ particle decay simulation







Preference in top-quark decay (H to qq~)

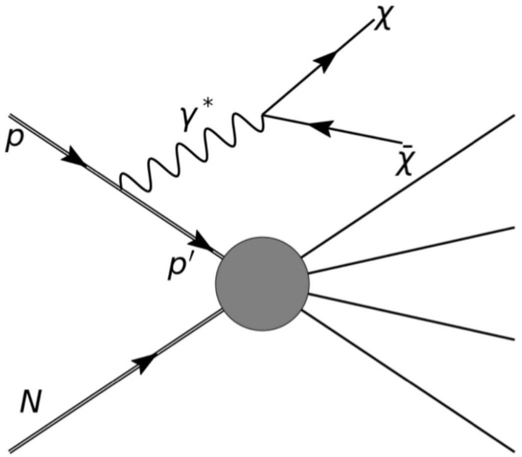
- Mij ~ 500 GeV > 125 GeV

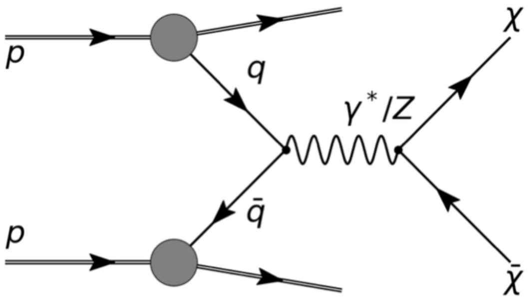
REFERENCE

- Related course @ CERN Summer Student Program:
  - Foundations of Statistics by Glen Cowan
  - MADGraph by Olivier Mattelaer
  - Theoritical Concepts in Particle Physics by Tim Cohen

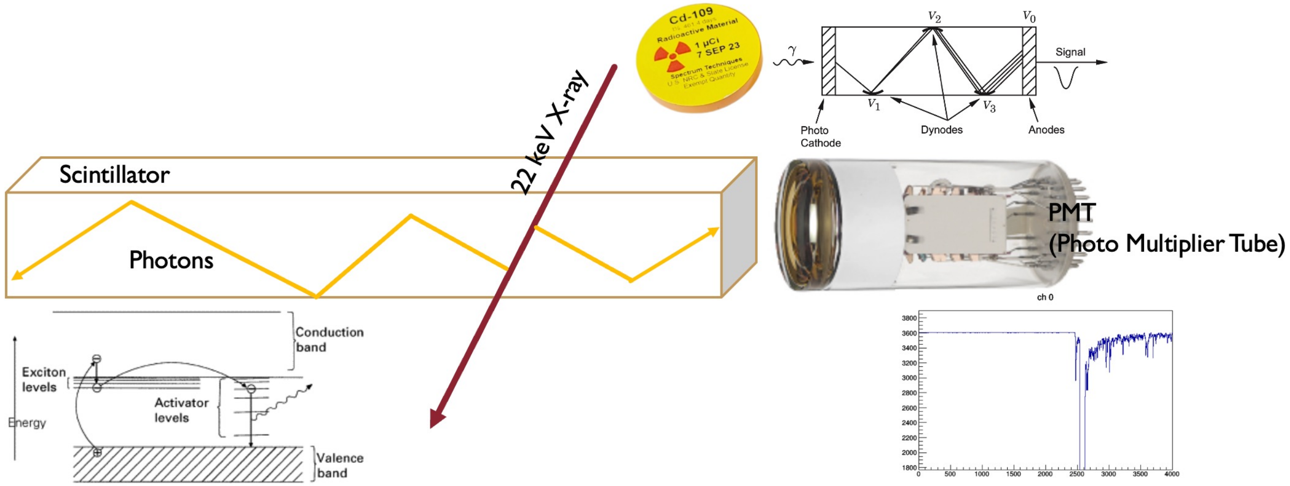
- S. Navas *et al.* (Particle Data Group),to be published in Phys. Rev. D 110, 030001 (2024)
- Caflisch RE. Monte Carlo and quasi-Monte Carlo methods. *Acta Numerica*. 1998;7:1-49. doi:10.1017/S0962492900002804
- Rahman, Q.I., Schmeisser, G. Characterization of the speed of convergence of the trapezoidal rule.*Numer. Math.* 57, 123–138 (1990). https://doi.org/10.1007/BF01386402
- *Atkinson, Kendall E. (1989).An Introduction to Numerical Analysis (2nd ed.). John Wiley & Sons.* ISBN 0-471-50023-2.
- Alwall, J., Frederix, R., Frixione, S. *et al.* The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *J. High Energ. Phys.* 2014, 79 (2014). https://doi.org/10.1007/JHEP07(2014)079
- Cowan, Glen. *Statistical data analysis*. Oxford university press, 1998.

IF TIME ALLOWS…

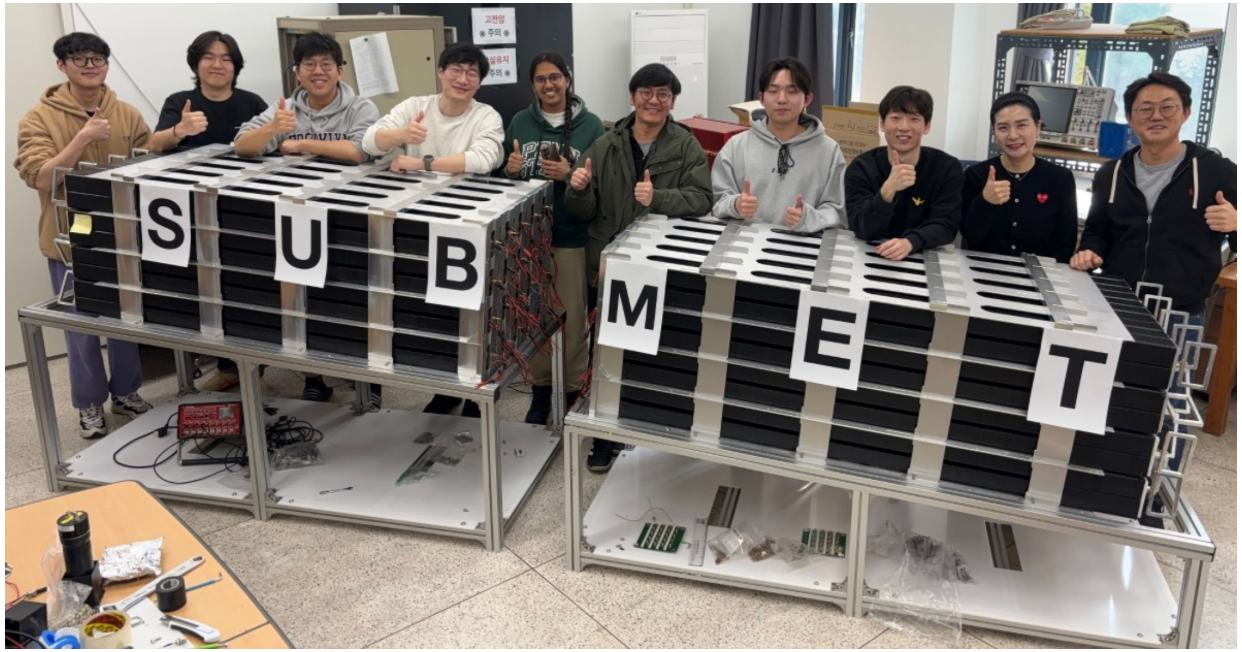


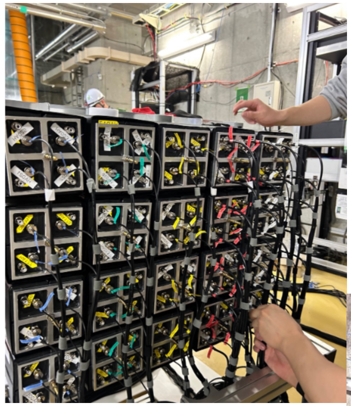


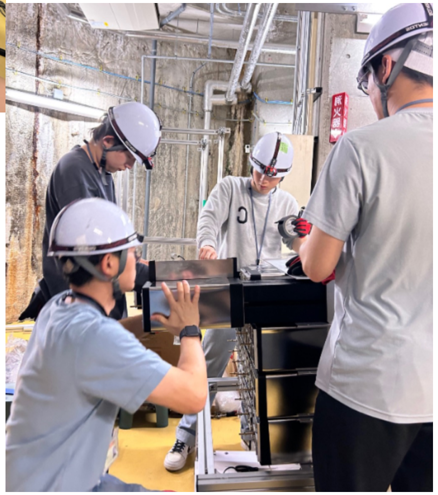
$Q = \epsilon e$ where $\epsilon \in \mathbb{R}$



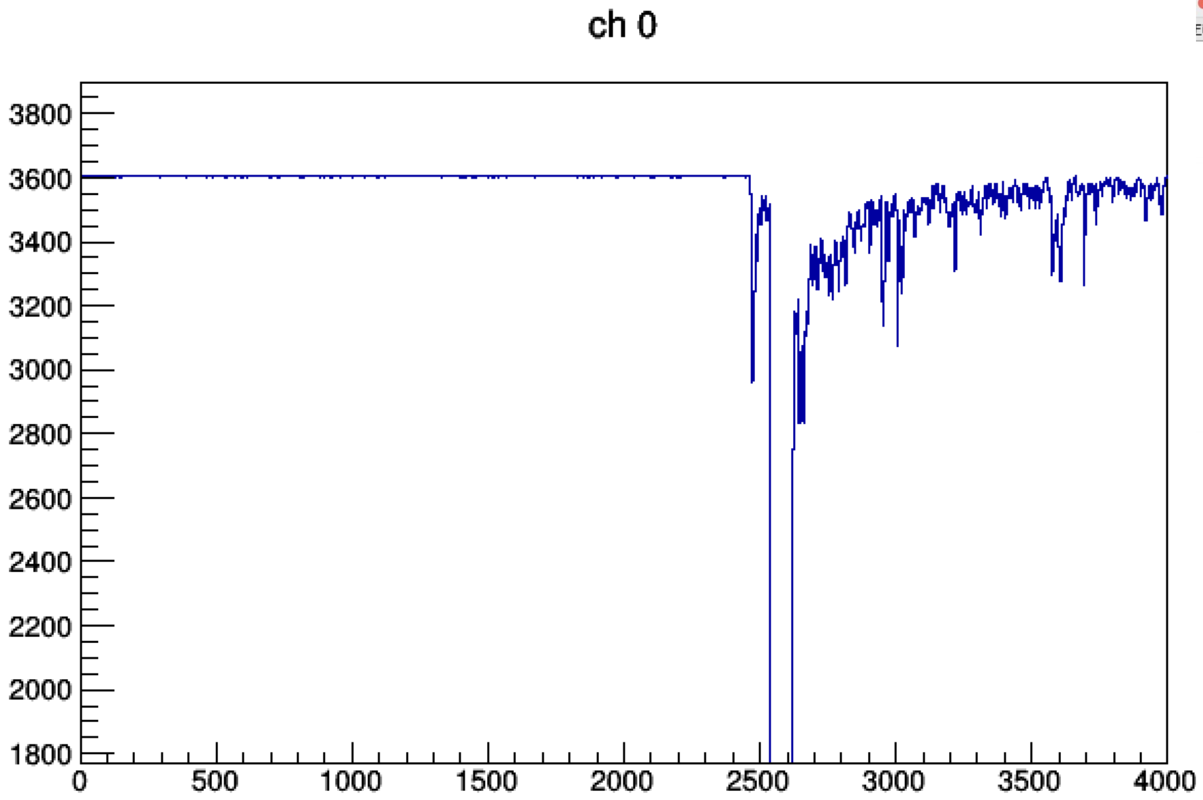
IF TIME ALLOWS…

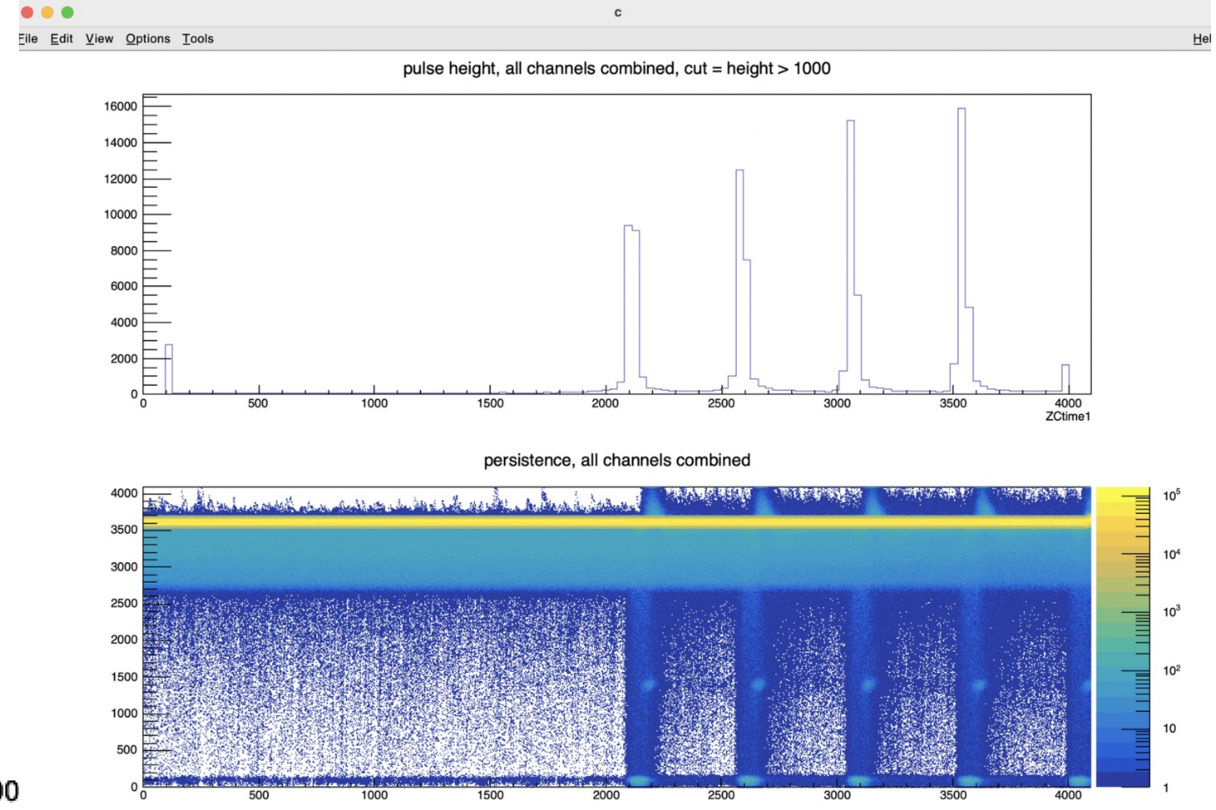




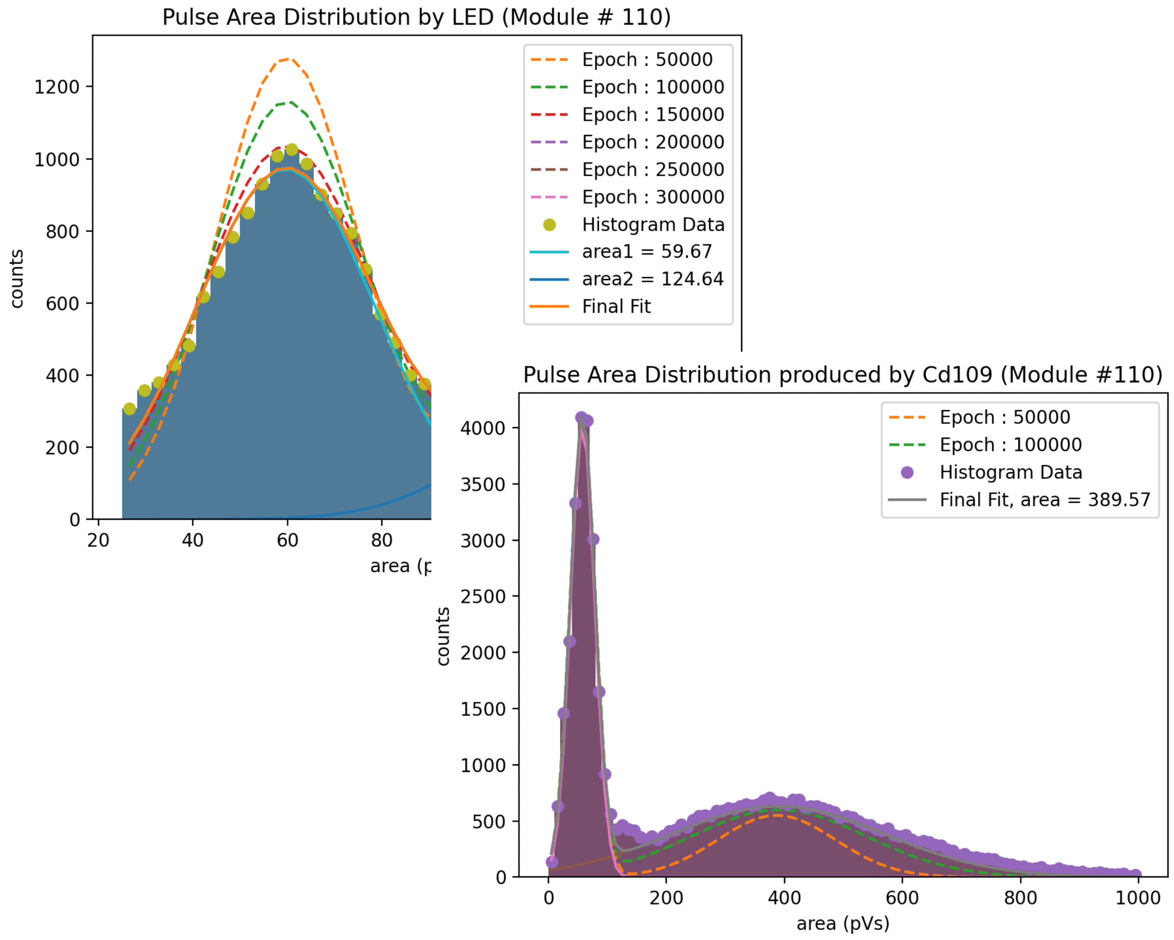


IF TIME ALLOWS…



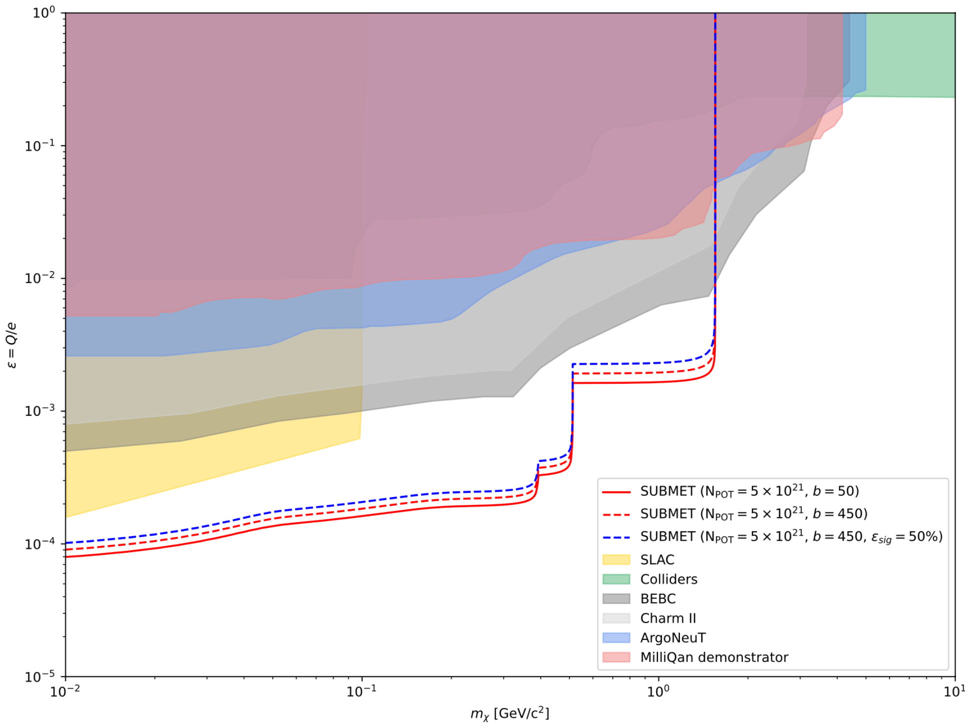


IF TIME ALLOWS…



- S_1 = 59.67 $\pm$ 3 pVs,
  S_2 = 389.57 $\pm$ 10 pVs
- 22 keV ray corresponds to
  $S_2/\ S_1$= 6.53 $\pm$ 0.37 PEs
- N_{PE}/ keV = 0.29 $\pm$ 0.02
  < 1 : probability

IF TIME ALLOWS…



THANK YOU!