UNIVERSITY OF SUSSEX

MARIE CURIE ACTIONS

UK Research and Innovation

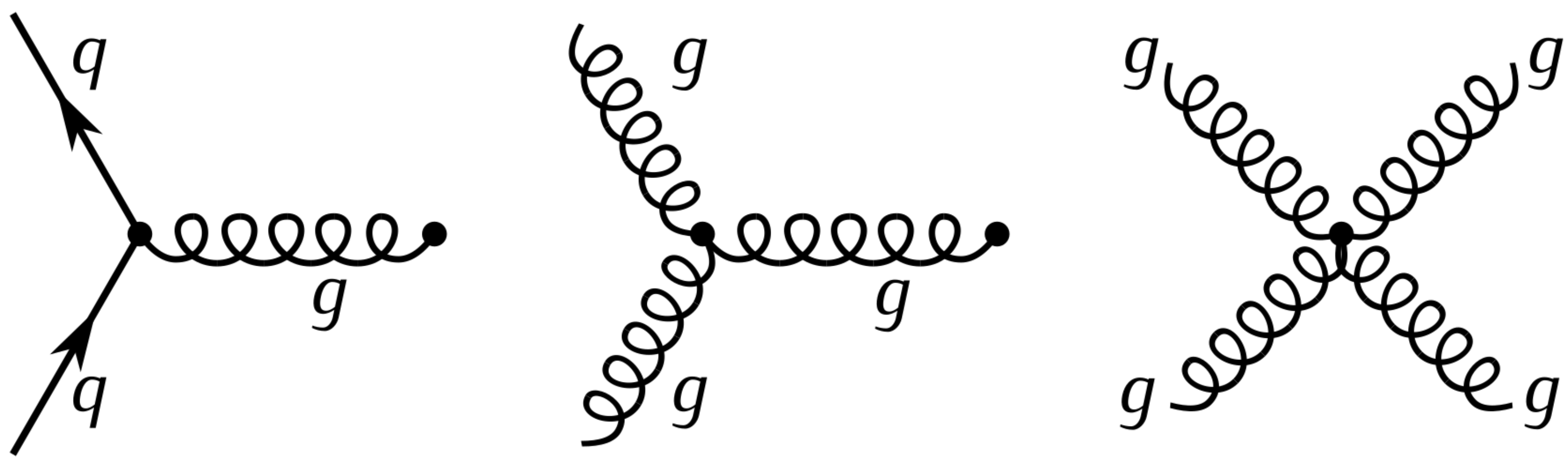
MONTE CARLO EVENT GENERATORS AND COMPUTATIONAL PHYSICS

MATTHEW A. LIM, UNIVERSITY OF SUSSEX

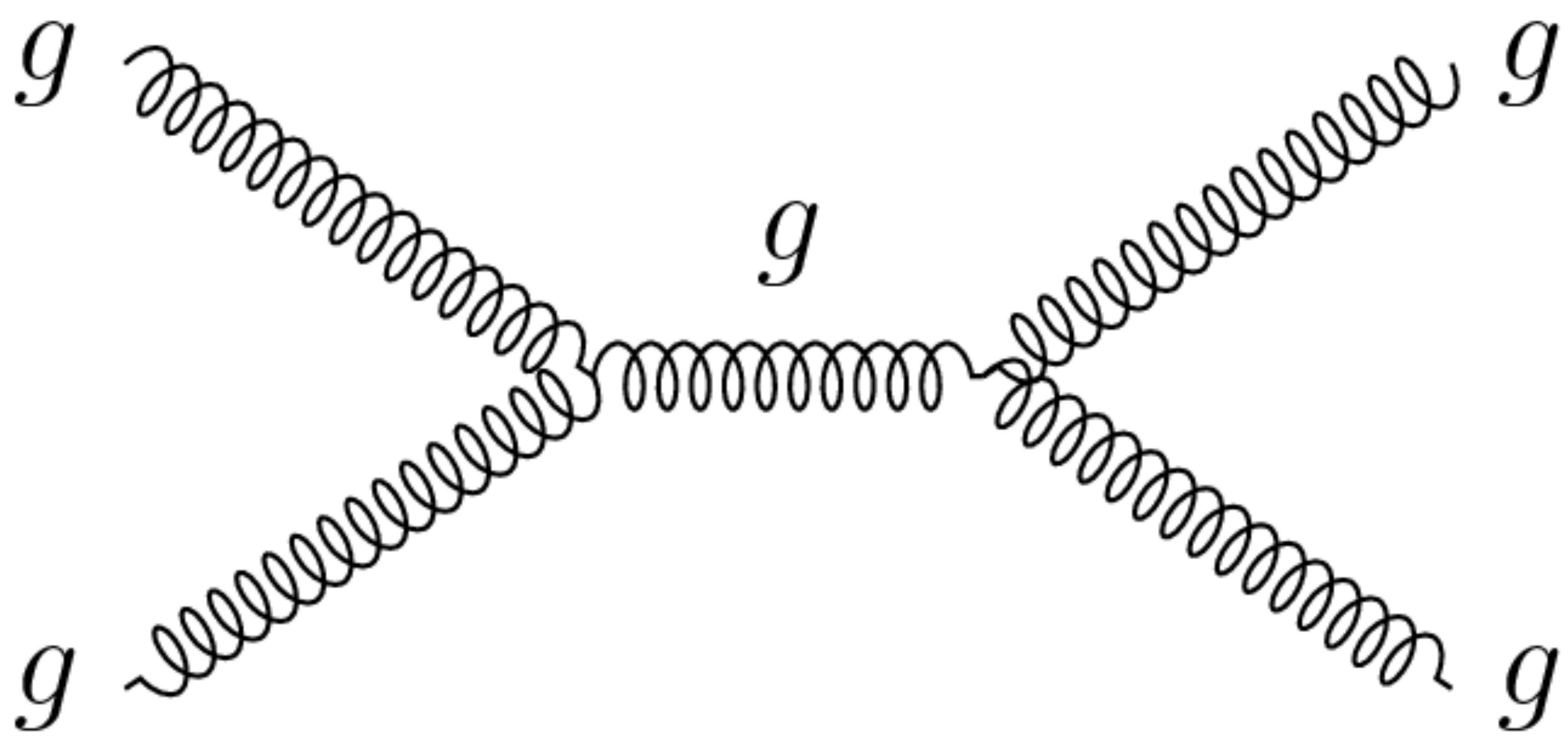
NEXT PHD WORKSHOP, QMUL 12-15 JUNE 2023

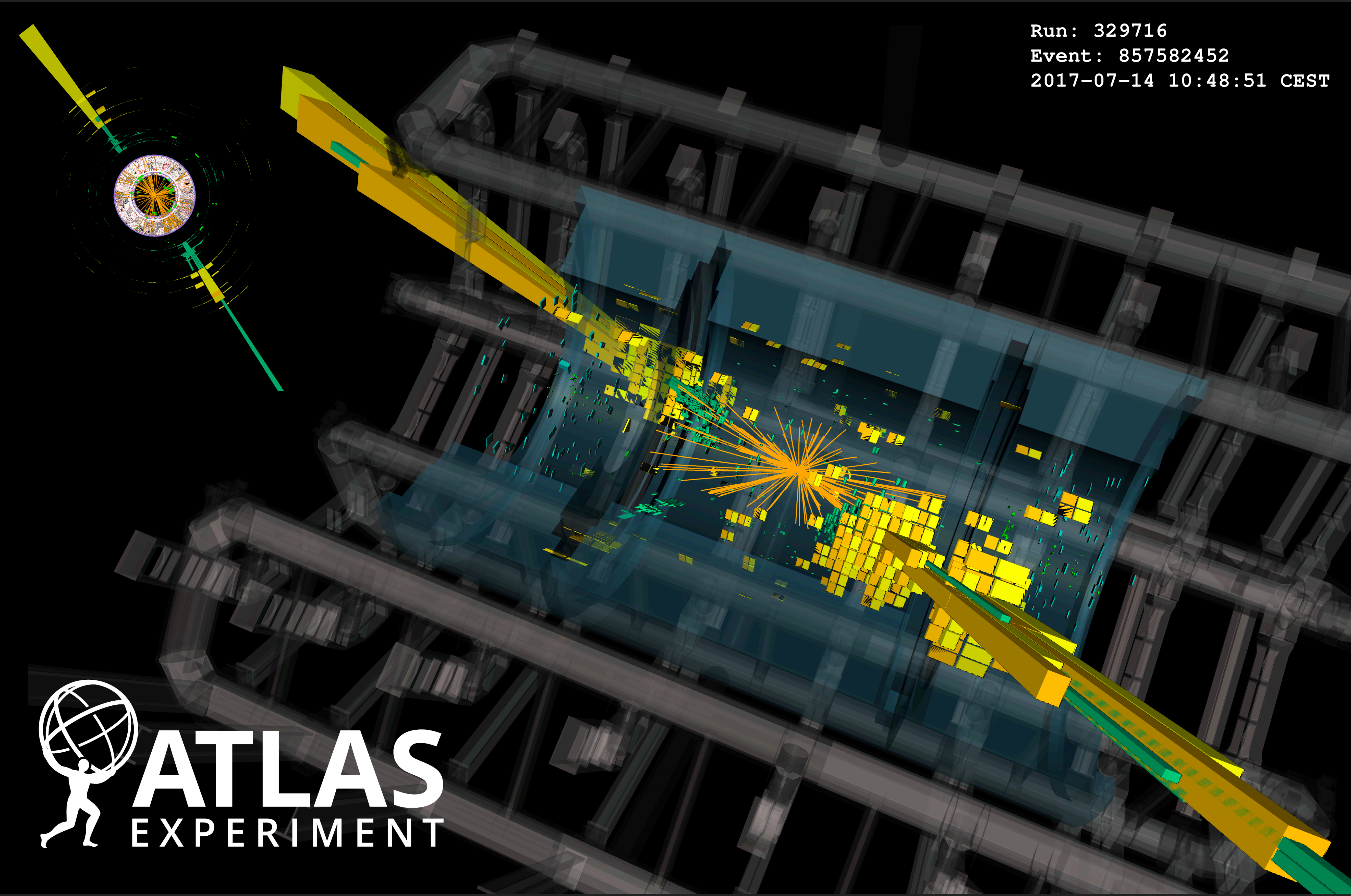
MOTIVATION

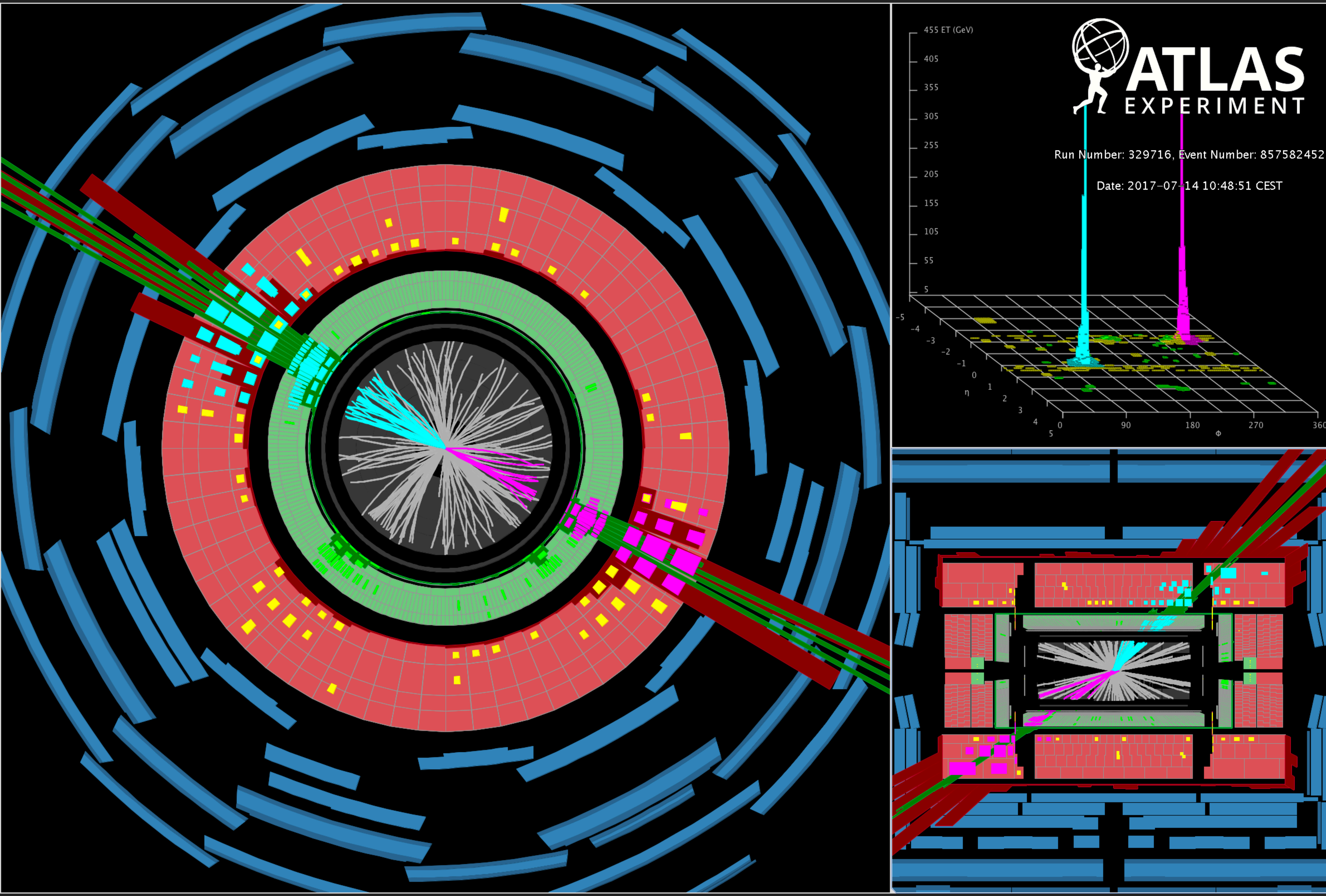
$$\mathscr{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + i\overline{\psi}\gamma^{\mu}\partial_{\mu}\psi$$



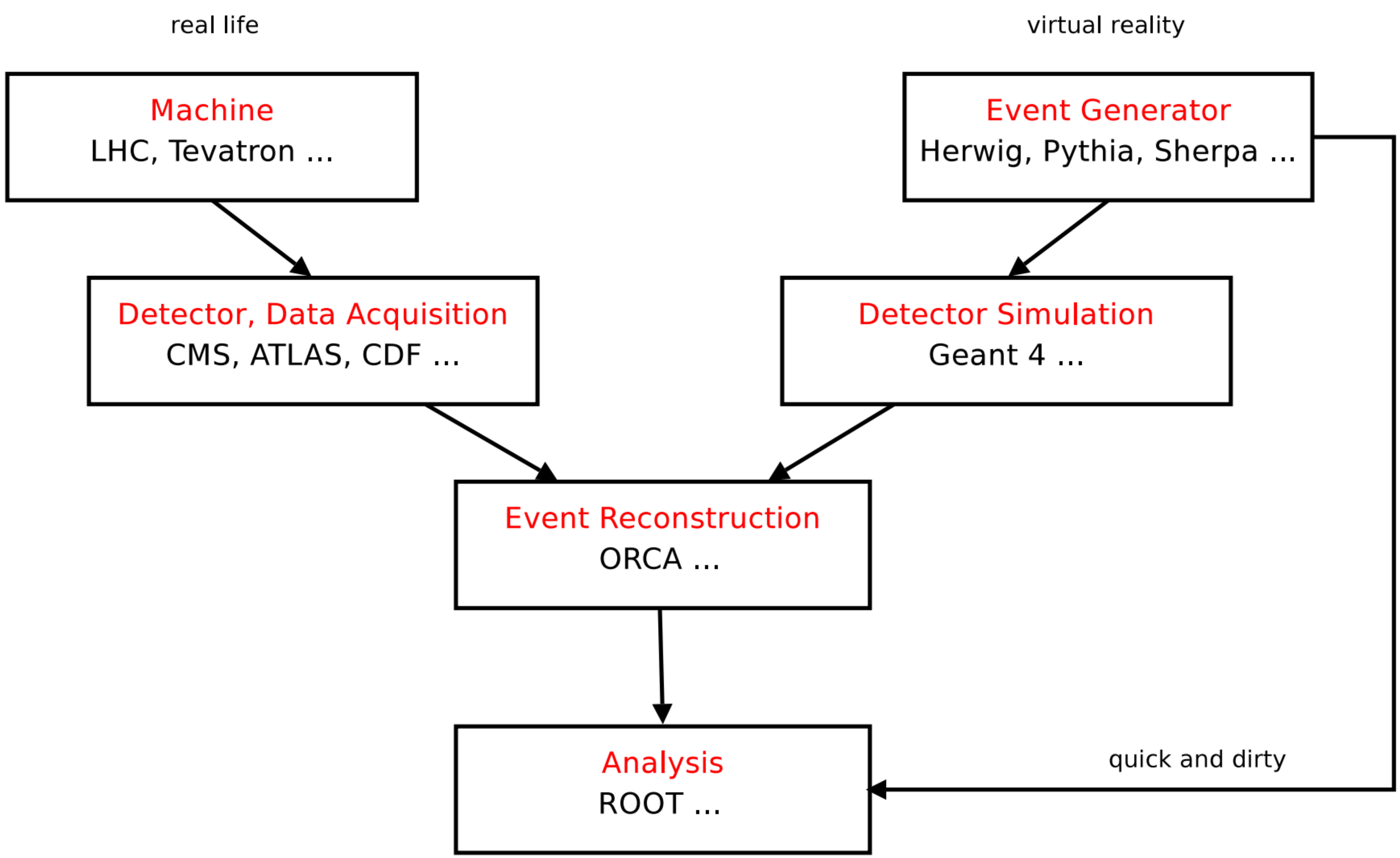
MOTIVATION




Run: 329716
Event: 857582452
2017-07-14 10:48:51 CEST
ATLAS
EXPERIMENT

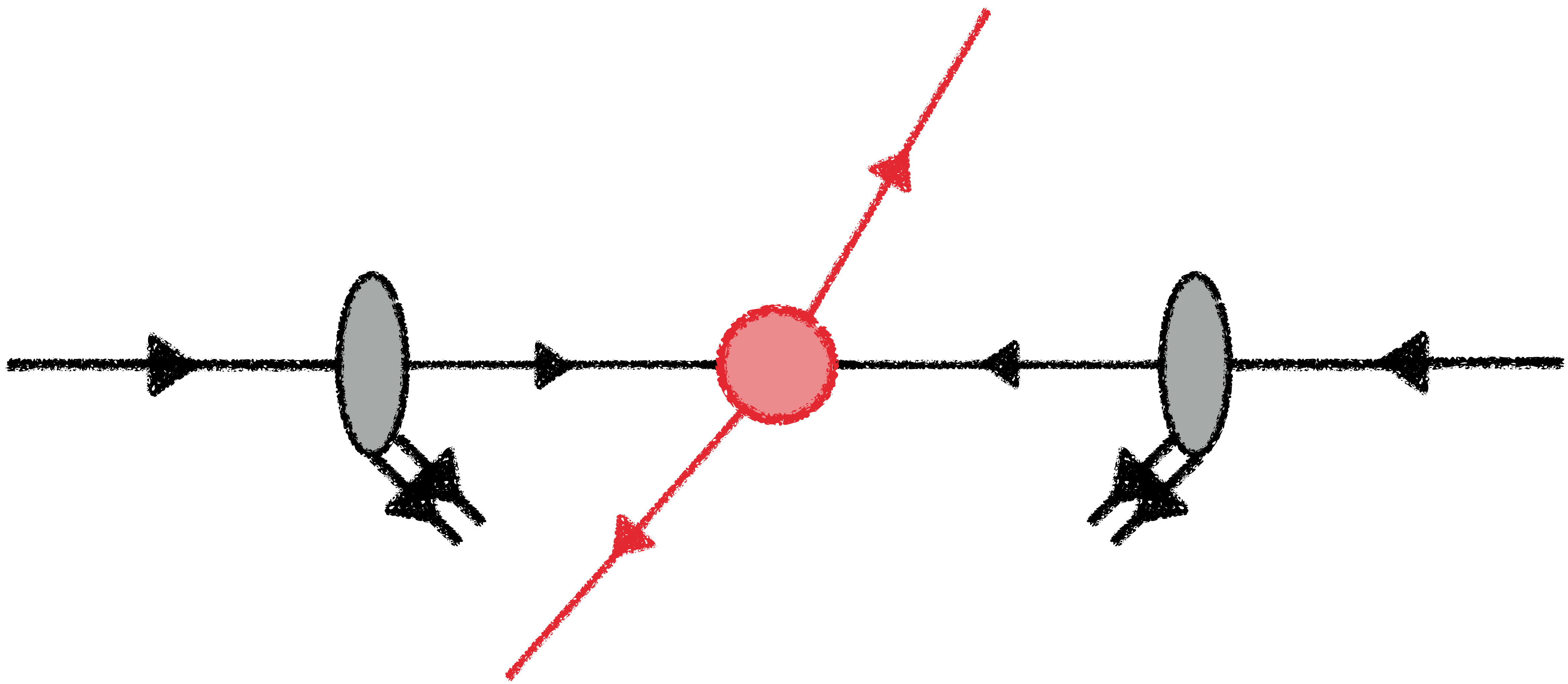

ATLAS
EXPERIMENT
Run Number: 329716, Event Number: 857582452
Date: 2017-07-14 10:48:51 CEST
455 ET (GeV)
405
355
305
255
205
155
105
55
5
η
Φ
0
90
180
270
360

MOTIVATION

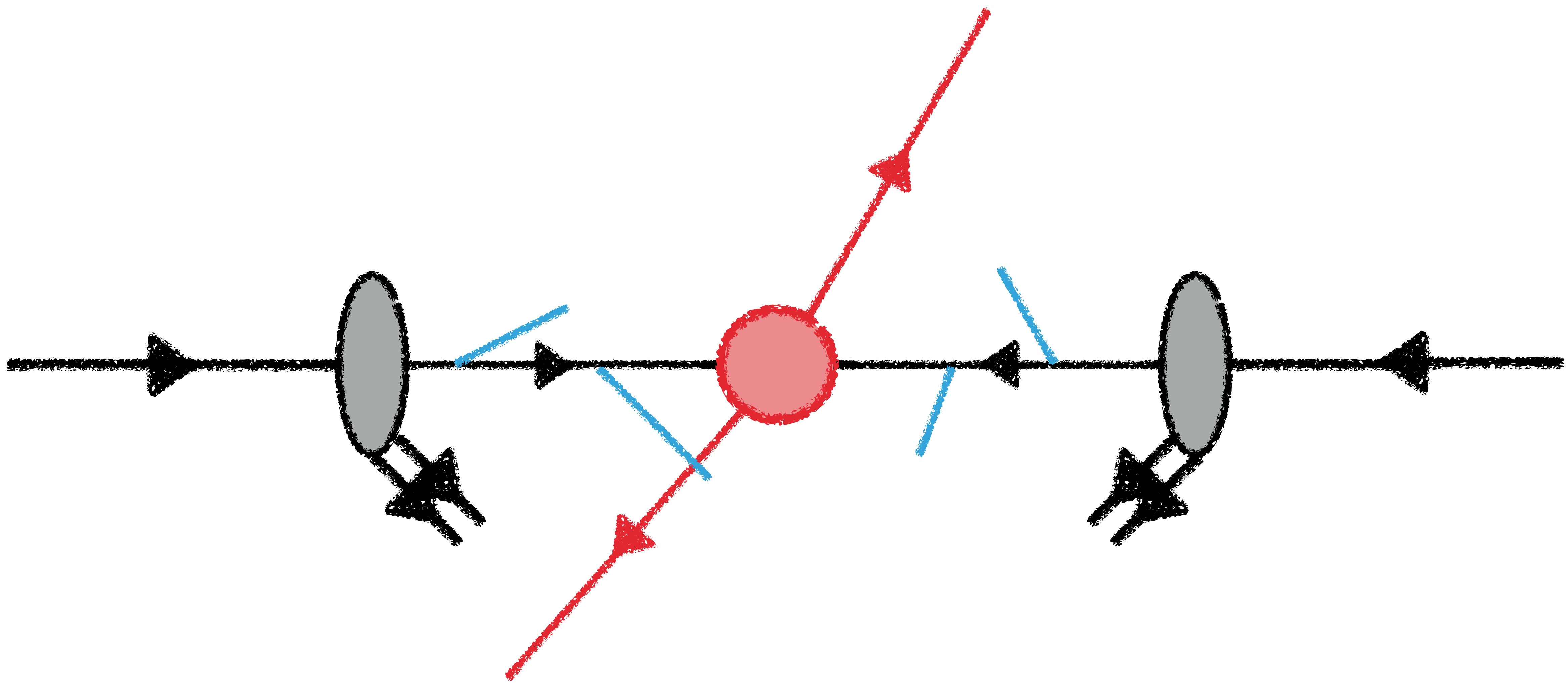


Credit: Stefan Gieseke

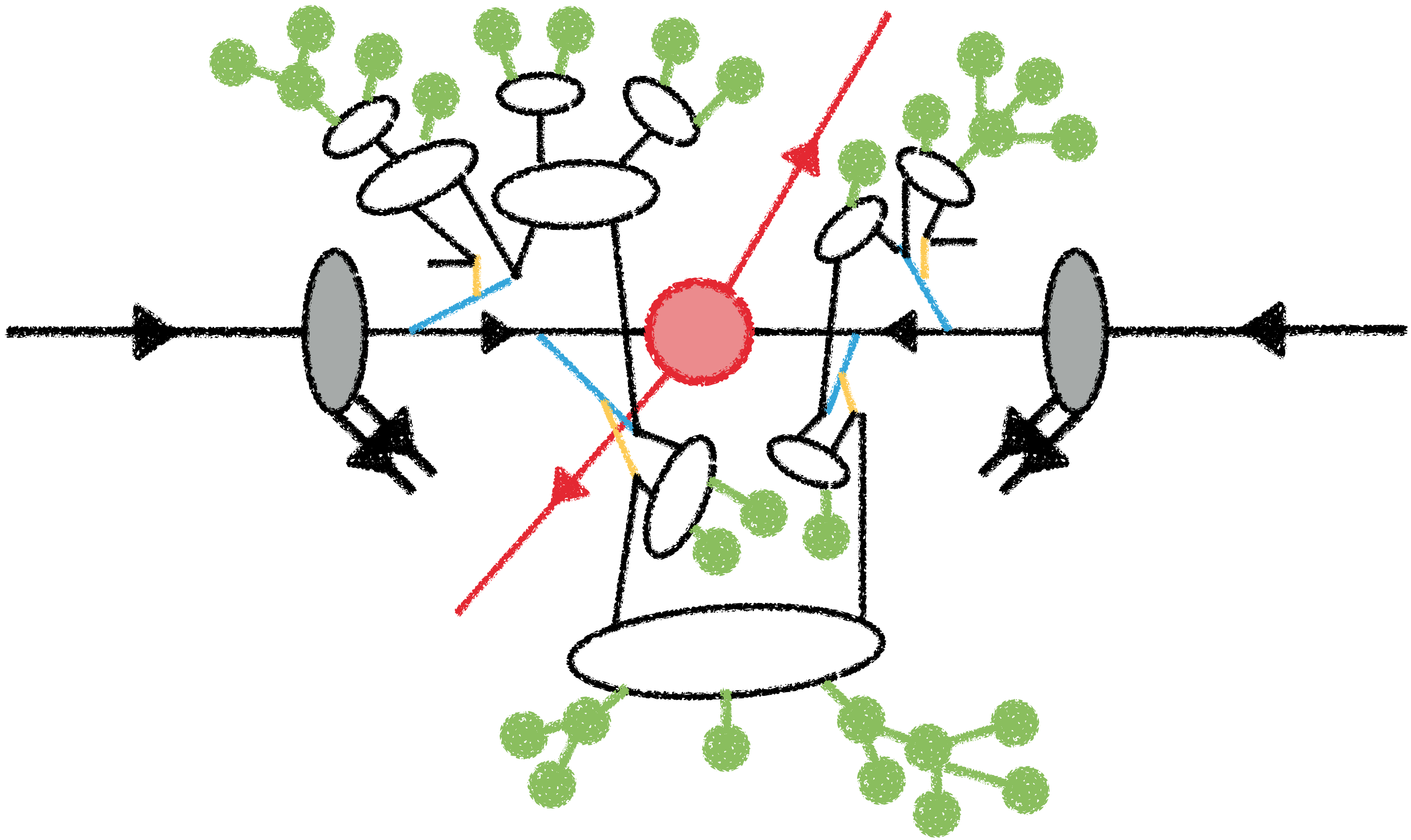
MONTE CARLO EVENT GENERATORS



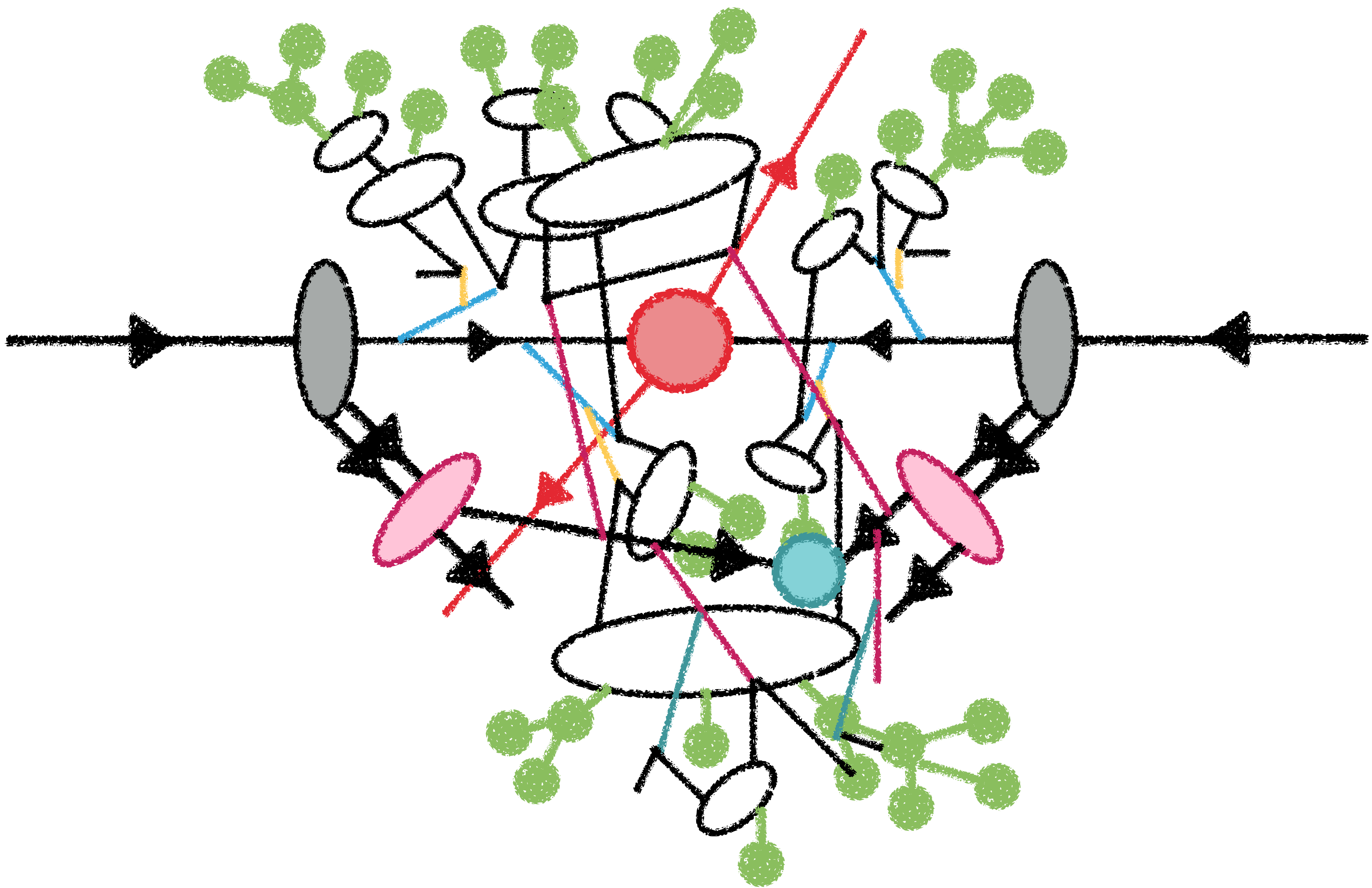
MONTE CARLO EVENT GENERATORS



MONTE CARLO EVENT GENERATORS



MONTE CARLO EVENT GENERATORS



MONTE CARLO EVENT GENERATORS

Hard interaction of partons $\mathcal{O}(\mathrm{TeV})$

Parton shower

Hadronisation

Hadron decays $\mathcal{O}(\mathrm{MeV})$

Modelling relies on:

- Factorisation of different energy scales
- Evolution from one scale to another

COURSE OUTLINE

- Lecture 1: Monte Carlo techniques and integration
- Lecture 2: Hard scattering: calculations at fixed order in perturbation theory
- Lecture 3: Parton showers, hadronisation modelling
- Lecture 4: Combining fixed order calculations with parton showers

COURSE RESOURCES

- Black Book of Quantum Chromodynamics (Campbell, Huston, Krauss)
- QCD and Collider Physics (Ellis, Stirling, Webber)
- MCnet lectures (Gieseke, Krauss)
- CERN-Fermilab lectures (Campbell)
- TASI lectures (Williams)
- Elements of QCD for hadron colliders (Salam)

MONTE CARLO TECHNIQUES

LEARNING OBJECTIVES

By the end of this lecture, you will be able to:

- List the advantages and disadvantages of Monte Carlo approaches to integration
- Analyse Monte Carlo results and evaluate performance of different strategies
- Suggest different applications of Monte Carlo methods to physical problems

WHAT ARE WE CALCULATING?

- QFT gives us squared matrix elements, which we interpret as probabilities
- We want to integrate these over phase space to produce predictions for the distribution of a certain variable, e.g. transverse momentum
- Often the integration is multi-dimensional and the integrand is very complex. We would also like to be able to place arbitrary cuts, i.e. restrict integration range

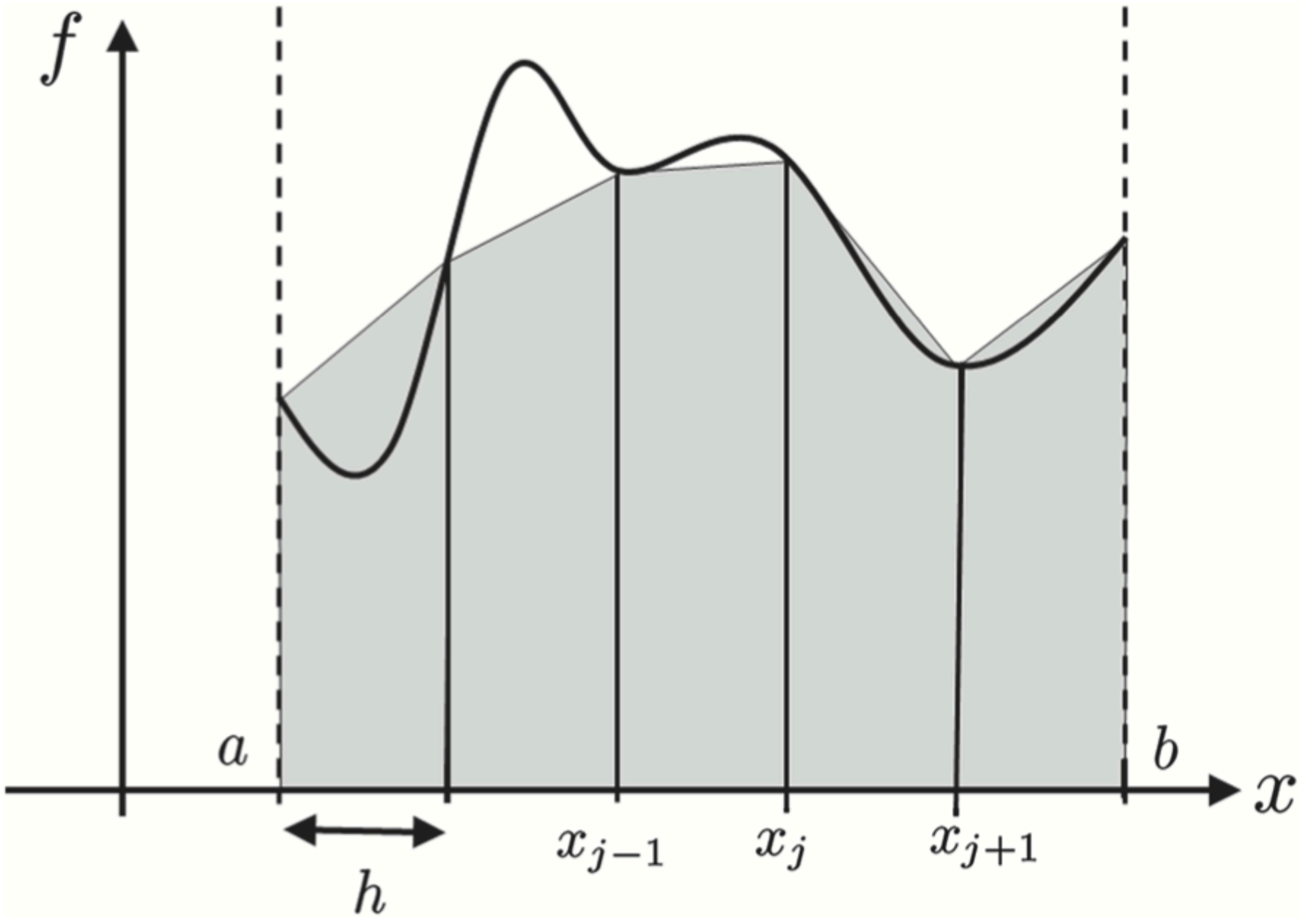
1D INTEGRATION WITH QUADRATURE METHODS

- Evaluate (for arbitrarily complicated $f(x)$)

$$I = \int_a^b f(x)\mathrm{d}x$$

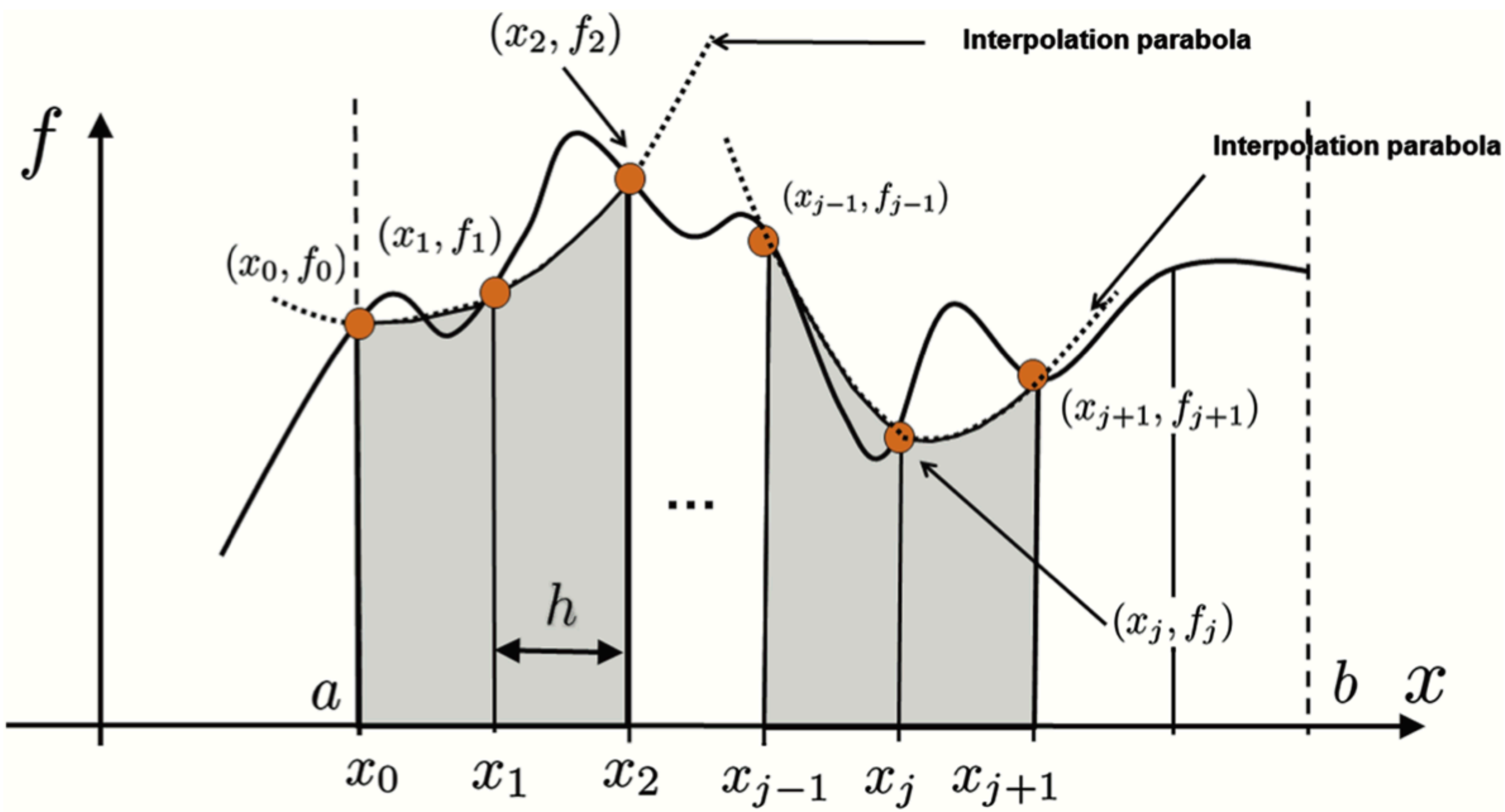
- Numerical strategy - divide $[a, b]$ into n intervals of width $h = (b - a)/n$;
- Evaluate the function in order to work out the area in each interval
- Different methods use different number of evaluations

1D INTEGRATION WITH QUADRATURE METHODS



Trapezoidal rule

1D INTEGRATION WITH QUADRATURE METHODS



Simpson's rule

1D INTEGRATION WITH QUADRATURE METHODS

- In general, the integral estimate has the form

$$\hat{I} = \sum_{i=1}^{n} w_i f(x_i)$$

- Area under the curve is approximate for a finite h - there are uncovered and overcovered areas

- Provided that $|f''(x)| \leq M$,

$$\hat{I} - I \leq \frac{(b-a)}{12} h^2 M \sim n^{-2}$$

Trapezoidal rule (exercise: show this!)

1D INTEGRATION WITH QUADRATURE METHODS

- In general, the integral estimate has the form

$$\hat{I} = \sum_{i=1}^{n} w_i f(x_i)$$

- Area under the curve is approximate for a finite h - there are uncovered and overcovered areas

- Provided that $|f''''(x)| \leq M$,

$$\hat{I} - I \leq \frac{(b-a)}{180} h^4 M \sim n^{-4}$$

Simpson's rule

HIGHER DIMENSIONAL INTEGRATION

- In higher dimensions we have

$$\hat{I} = \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_d=1}^{n} w_{i_1} w_{i_2} \ldots w_{i_d} f(x_{i_1}, x_{i_2}, \ldots, x_{i_d})$$

- Can apply same principle as before - need n samples in each dimension, so n^d samples in total
- Error is now $\mathcal{O}(N^{-r}) \sim \mathcal{O}(n^{-r/d})$ where $r = 2$ or 4
- Increasing n does not help! Curse of dimensionality

THE ROUGH IDEA

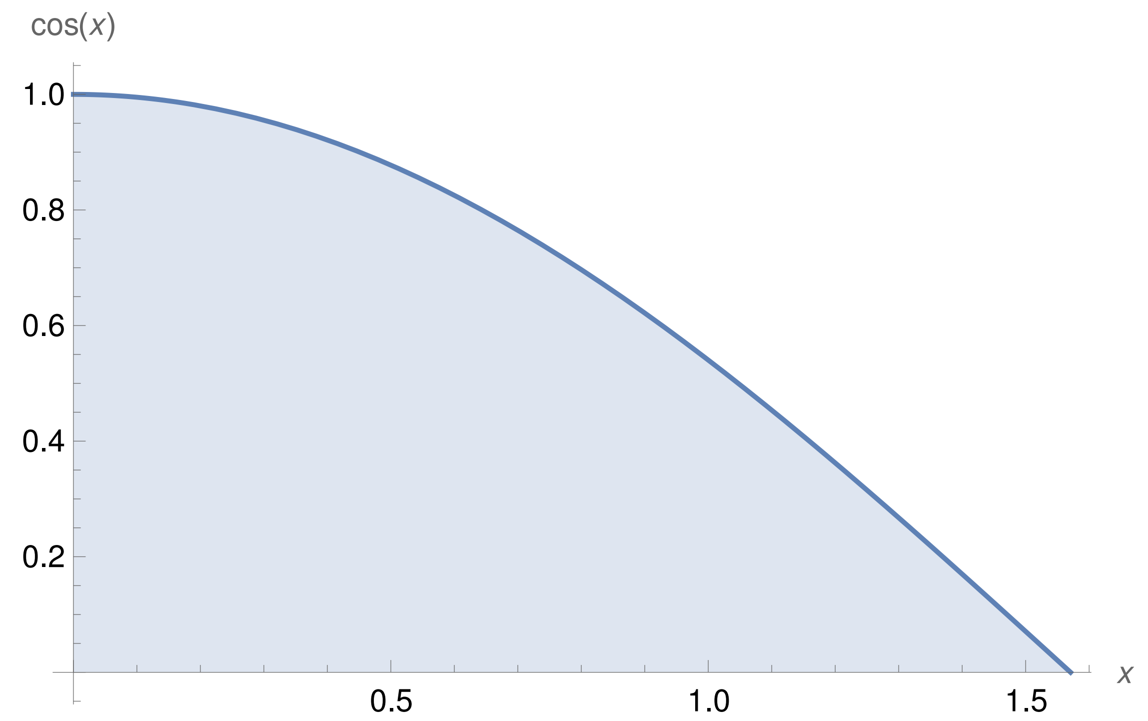
- Regard integrand as probability density

$$\mathrm{d}P = f(x)\,\mathrm{d}x$$

with probability distribution

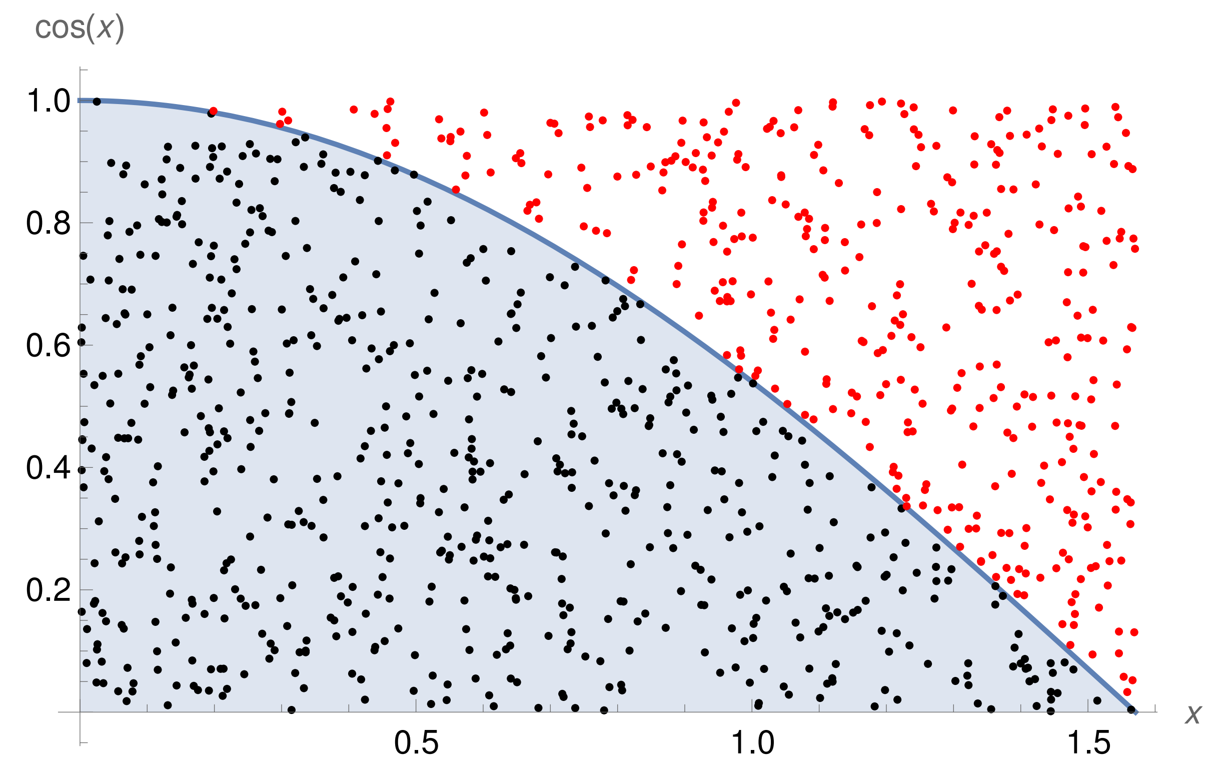
$$F(x) = \int_a^b f(x)\,\mathrm{d}x$$

so that the probability is given by the area under the curve.



HIT AND MISS APPROACH

- Throw N random points (x, y) into region volume V
- Count successes N_{win} whenever $y < f(x)$
- Estimate of integral and error are given by



$$\hat{I} = V\frac{N_{\text{win}}}{N}$$

$$\delta I \equiv I - \hat{I} \sim \frac{1}{\sqrt{N}}$$

HIT AND MISS APPROACH

- Looks good - can handle any $f(x)$, regardless of functional form, discontinuities…
- Error goes as $1/\sqrt{N}$ for any number of dimensions!
- $f(x)$ must be bounded from above
- Sampling becomes inefficient whenever $f(x)$ has large variance
- Can be improved using variance reduction

MONTE CARLO ESTIMATOR

- Let’s take a step back and define a random variable

$$F_N = \frac{(b-a)}{N}\sum_{i=1}^{N} f(X_i)$$

where X_i is also a random variable drawn from $[a, b]$.

- Average of N random function evaluations, flat sampling of X_i
- What is the expected value of this random variable?

MONTE CARLO ESTIMATOR

- What is the expected value of this random variable?

$$
\begin{aligned}
E[F_N] &= E\left[\frac{(b-a)}{N}\sum_{i=1}^{N} f(X_i)\right] \\
&= \frac{(b-a)}{N}\sum_{i=1}^{N} E[f(X_i)] \\
&= \frac{(b-a)}{N}\sum_{i=1}^{N} \int_{-\infty}^{\infty} f(x)\,p(x)\,\mathrm{d}x \\
&= \frac{(b-a)}{N}\frac{1}{(b-a)}\sum_{i=1}^{N} \int_{a}^{b} f(x)\,\mathrm{d}x \\
&= \frac{1}{N}\sum_{i=1}^{N} \int_{a}^{b} f(x)\,\mathrm{d}x = \frac{1}{N} N \cdot I = I
\end{aligned}
$$

MONTE CARLO ESTIMATOR

- What is the variance of this random variable?

$$
\begin{aligned}
\mathrm{Var}[F_N] &= \mathrm{Var}\left[\frac{(b-a)}{N}\sum_{i=1}^{N} f(X_i)\right] \\
&= \frac{(b-a)^2}{N^2}\mathrm{Var}\left[\sum_{i=1}^{N} f(X_i)\right] \\
&= \frac{(b-a)^2}{N^2}\sum_{i=1}^{N}\mathrm{Var}\left[f(X_i)\right] \\
&= \frac{(b-a)^2}{N^2} N\,\mathrm{Var}\left[f(X)\right] \\
&= \frac{1}{N}\mathrm{Var}\left[(b-a)f(X)\right]
\end{aligned}
$$

MONTE CARLO ESTIMATOR

- Chebyshev's inequality: for a random variable X, no more than $1/k^2$ of the values are more than k standard deviations from the mean

$$P\left\{\,|X-\mu|\,\geq k\sigma\right\} \leq \frac{1}{k^2}$$

$$P\left\{\,|X-E[X]|\,\geq \left(\frac{\mathrm{Var}[X]}{\delta}\right)^{1/2}\right\} \leq \delta$$

$$\delta = \frac{1}{k^2}$$

MONTE CARLO ESTIMATOR

- Plug in F_N :

$$P\left\{ |F_N - I| \geq \left(\frac{1}{N}\right)^{1/2} \left(\frac{\mathrm{Var}[(b-a)f(X)]}{\delta}\right)^{1/2} \right\} \leq \delta$$

- For a given $\delta = 1/k^2$, the error decreases as we increase the number of samples with a scaling of $1/\sqrt{N}$.
- The prefactor is the standard deviation of f.

MONTE CARLO ESTIMATOR

- For $f(x) = \cos(x), 0 \leq x \leq \pi/2$ we have that

$$E[f] = \int_0^{\pi/2} \cos(x)\,\mathrm{d}x = 1$$

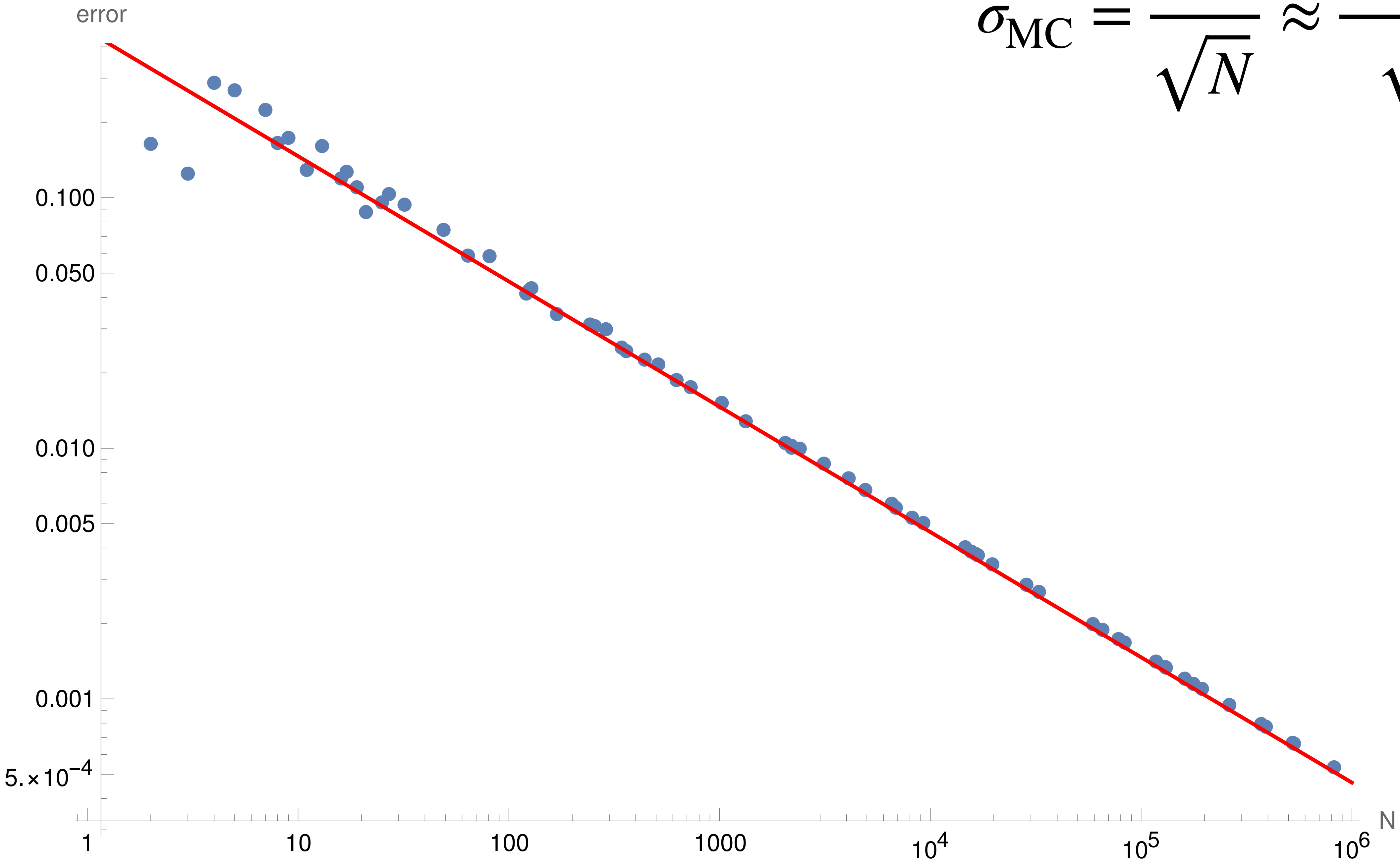
$$\mathrm{Var}[f] = E[f]^2 - \int_0^{\pi/2} \cos^2(x)\,\mathrm{d}x = 1 - \frac{\pi}{4}$$

and therefore

$$\sigma_{\mathrm{MC}} = \frac{\sigma}{\sqrt{N}} \approx \frac{0.4633}{\sqrt{N}}$$

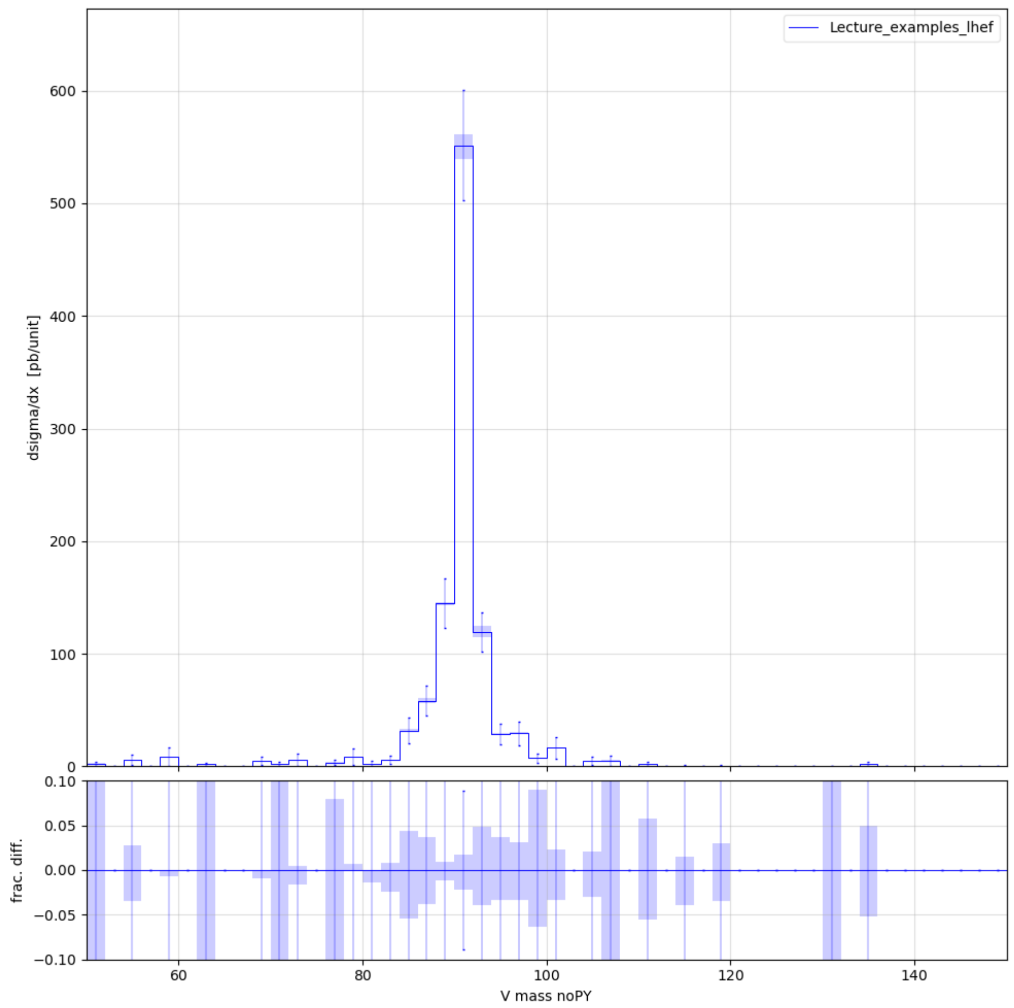
MONTE CARLO ESTIMATOR

$$\sigma_{\mathrm{MC}} = \frac{\sigma}{\sqrt{N}} \approx \frac{0.4633}{\sqrt{N}}$$

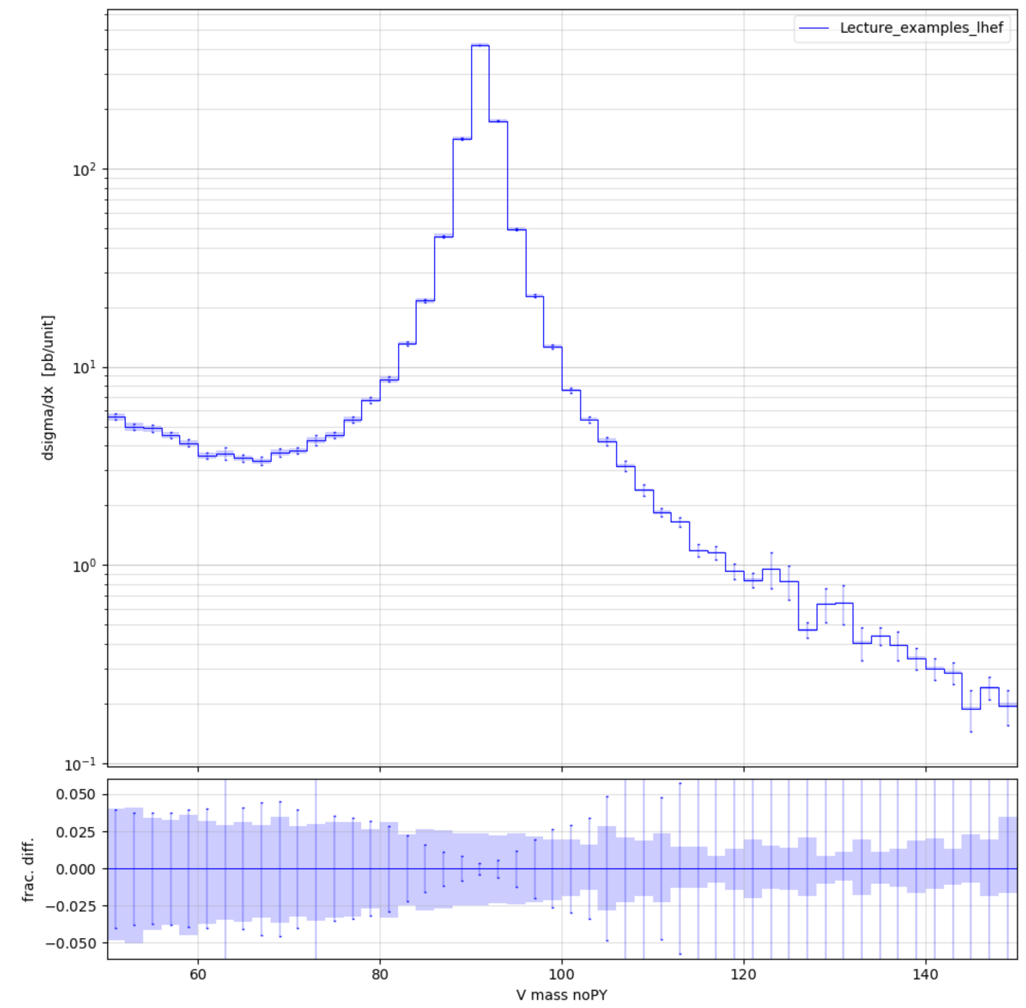


WHY DO MY RESULTS SUCK? #1

- You didn't run enough points



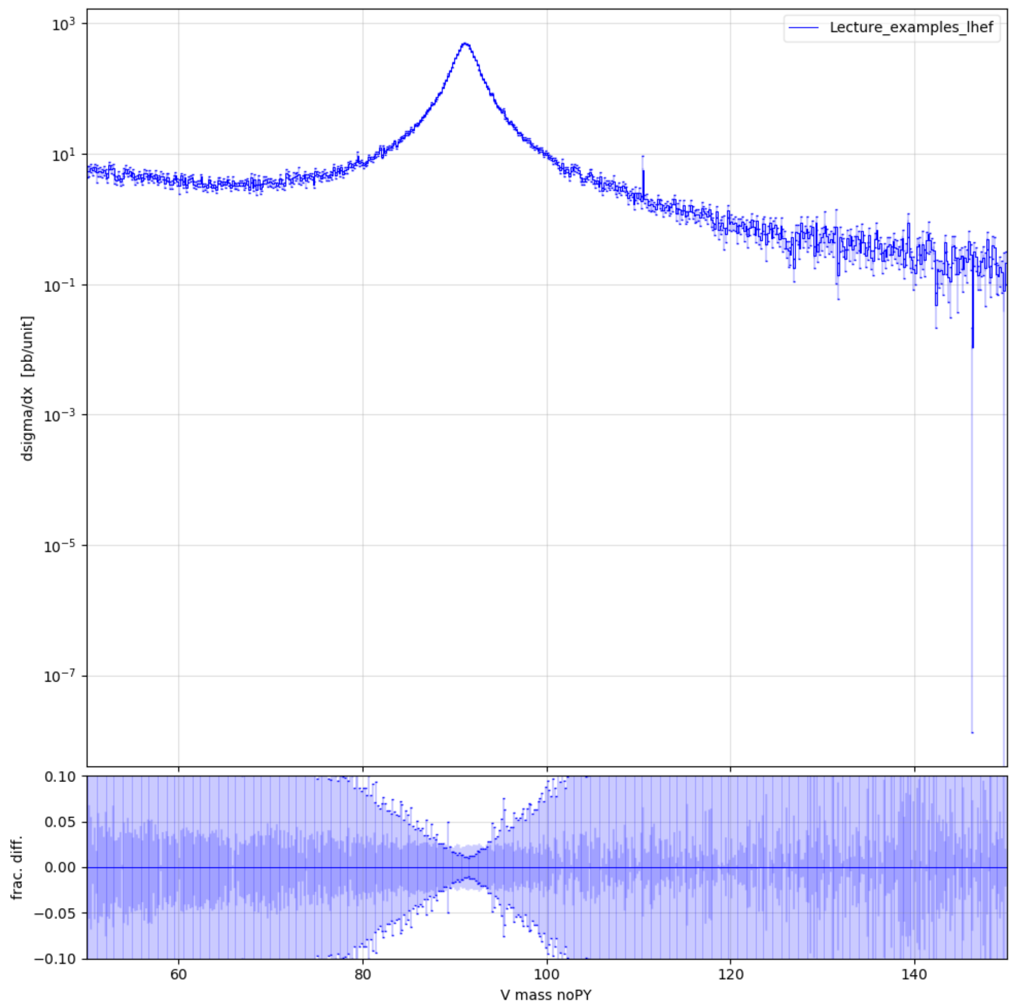
100 points



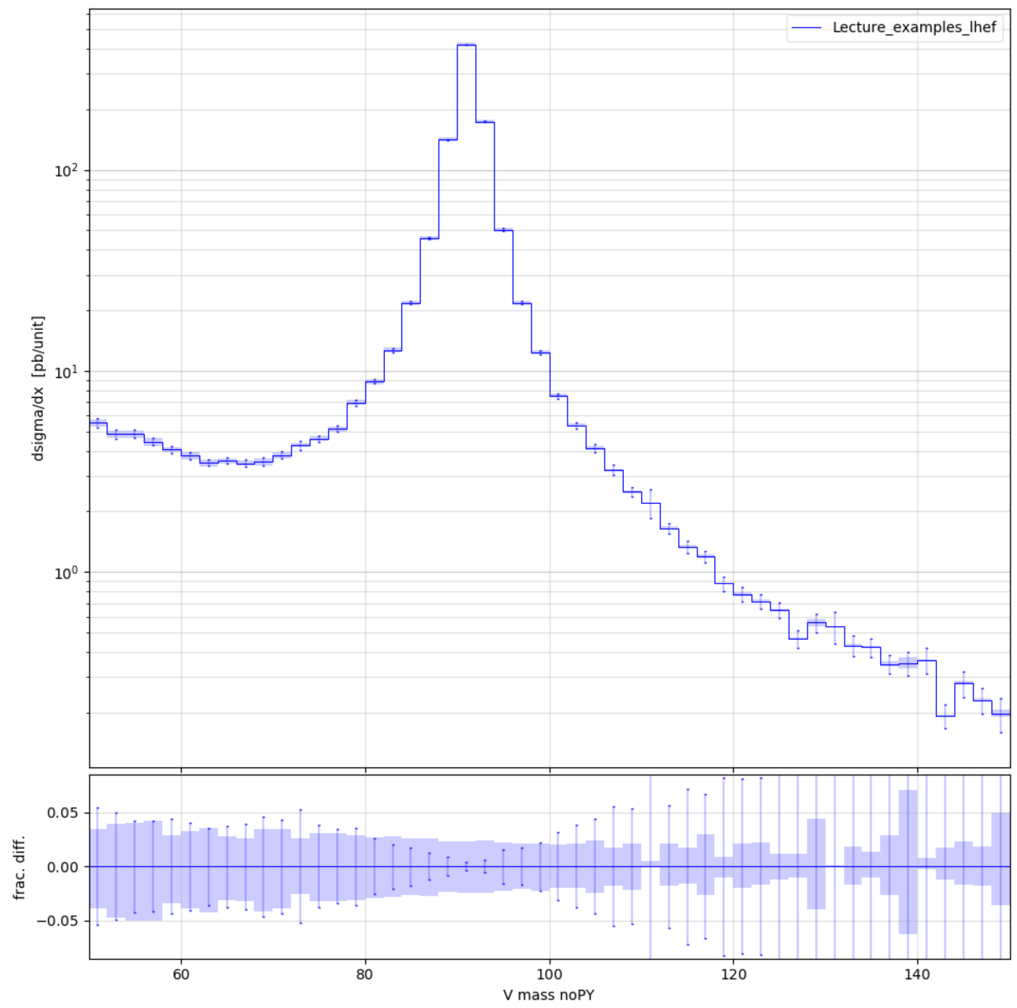
10000 points

WHY DO MY RESULTS SUCK? #2

- You didn't bin properly



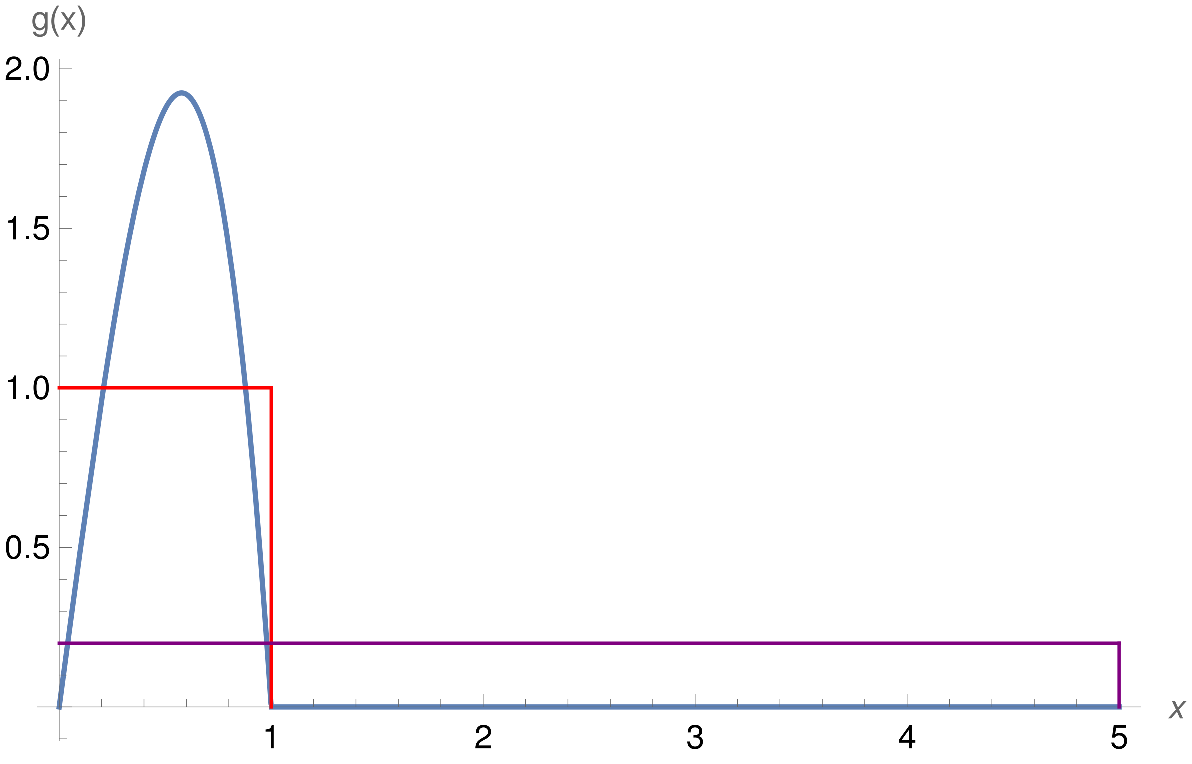
500 bins



50 bins

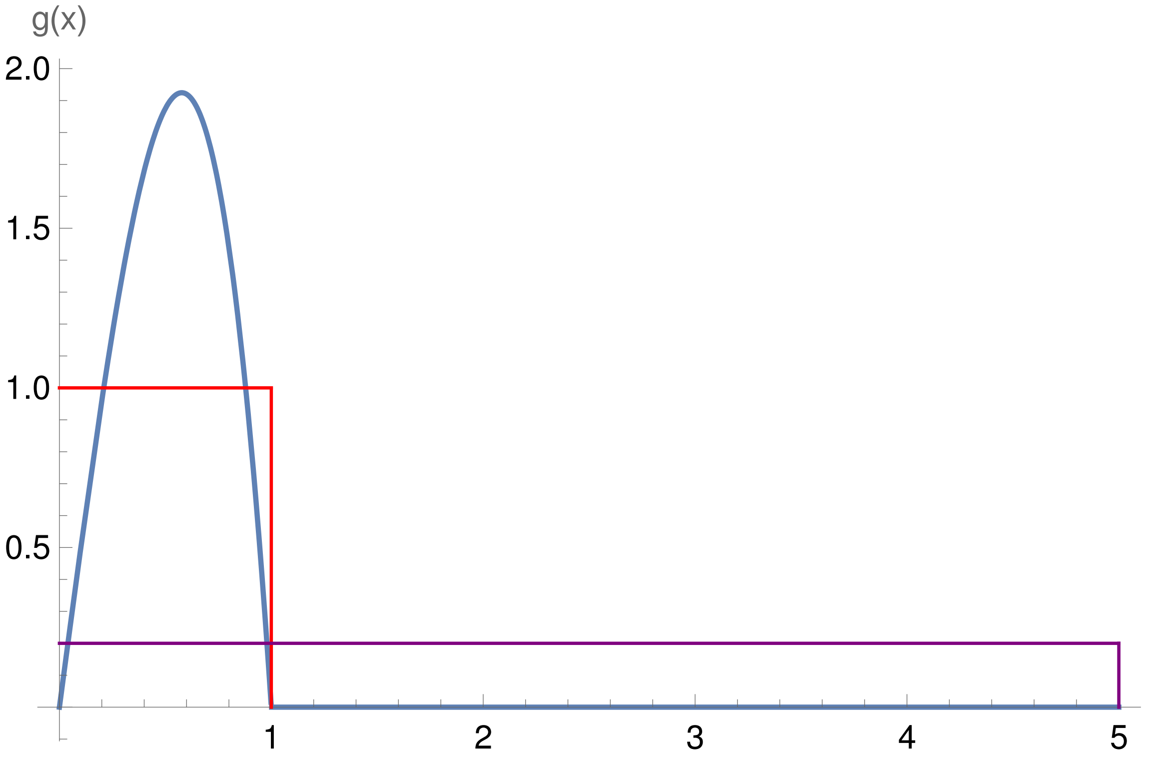
IMPORTANCE SAMPLING

- Consider an integral $\int_0^1 g(x)\,\mathrm{d}x$ where $g(x) = 0$ for $x \geq 1$
- We could try throwing points uniformly in the interval $[0,1]$
- Alternatively, we could throw points in the interval $[0,5]$ and rescale our result by 5



IMPORTANCE SAMPLING

- This would not be sensible - 80% of the time we would learn nothing about the integral, because for $x \geq 1$ the integrand is zero.
- If there are bad ways to throw points, are there good ways?



IMPORTANCE SAMPLING

- Let's revisit our random variable

$$F'_N = \frac{(b-a)}{N}\sum_{i=1}^{N} f(X_i) = \frac{1}{N}\sum_{i=1}^{N}\frac{f(X_i)}{\frac{1}{(b-a)}} = \frac{1}{N}\sum_{i=1}^{N}\frac{f(X_i)}{p(X_i)}$$

- We are not forced to throw points by sampling from a uniform distribution - can use an arbitrary distribution p
- If we draw more samples somewhere, their weights should be scaled down. If we draw fewer samples, we scale up.

IMPORTANCE SAMPLING

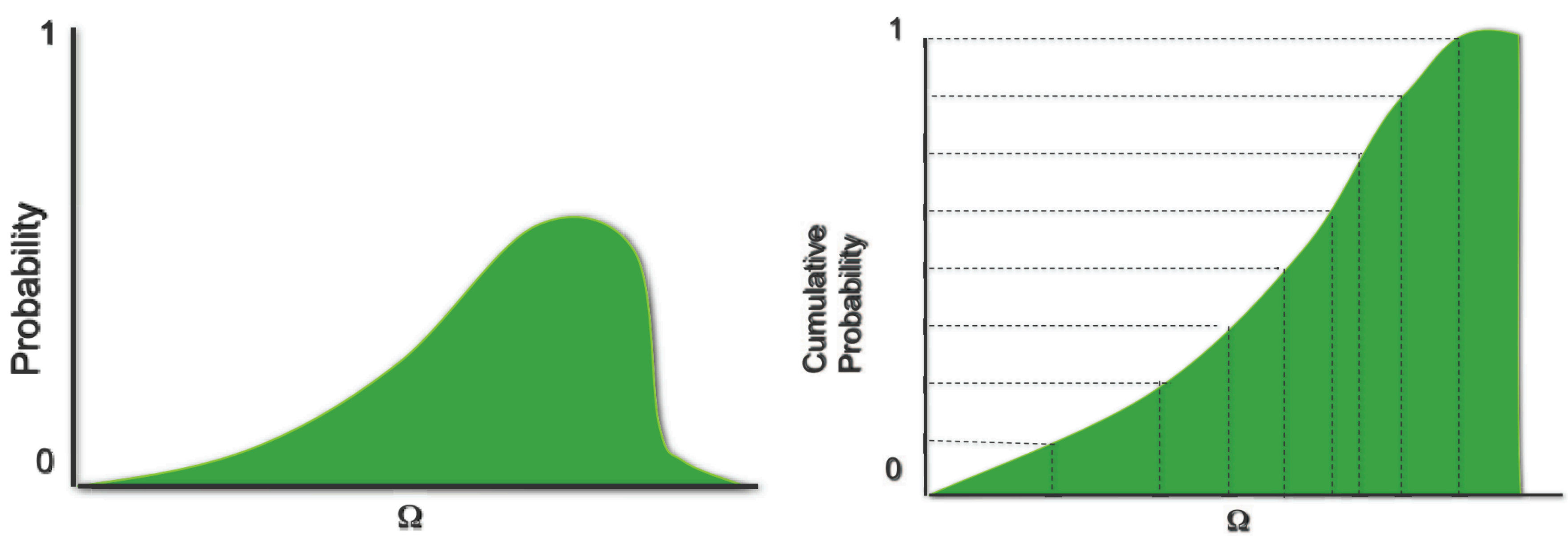
- Writing $Y_i = \dfrac{f(X_i)}{p(X_i)}$, can show that (exercise):

$$E[F'_N] = I \qquad \mathrm{Var}[F'_N] = \frac{1}{N}\mathrm{Var}[Y]$$

- How do we choose p to minimise $\mathrm{Var}[F'_N]$? If $p(X_i) = \lambda f(X_i)$ then $Y_i = \lambda$ and the variance is zero!
- If you could sample from p then you could sample from f and there would be no work to do…
- Choose p close to f to minimise error

INVERSION METHOD

- Given a uniform random number generator, how do we sample from $p(x)$?
- Answer: uniformly sample the CDF $P(x)$ and invert back.
- Gives dense samples at important regions in the domain, but need to know $P(x)$ and its inverse



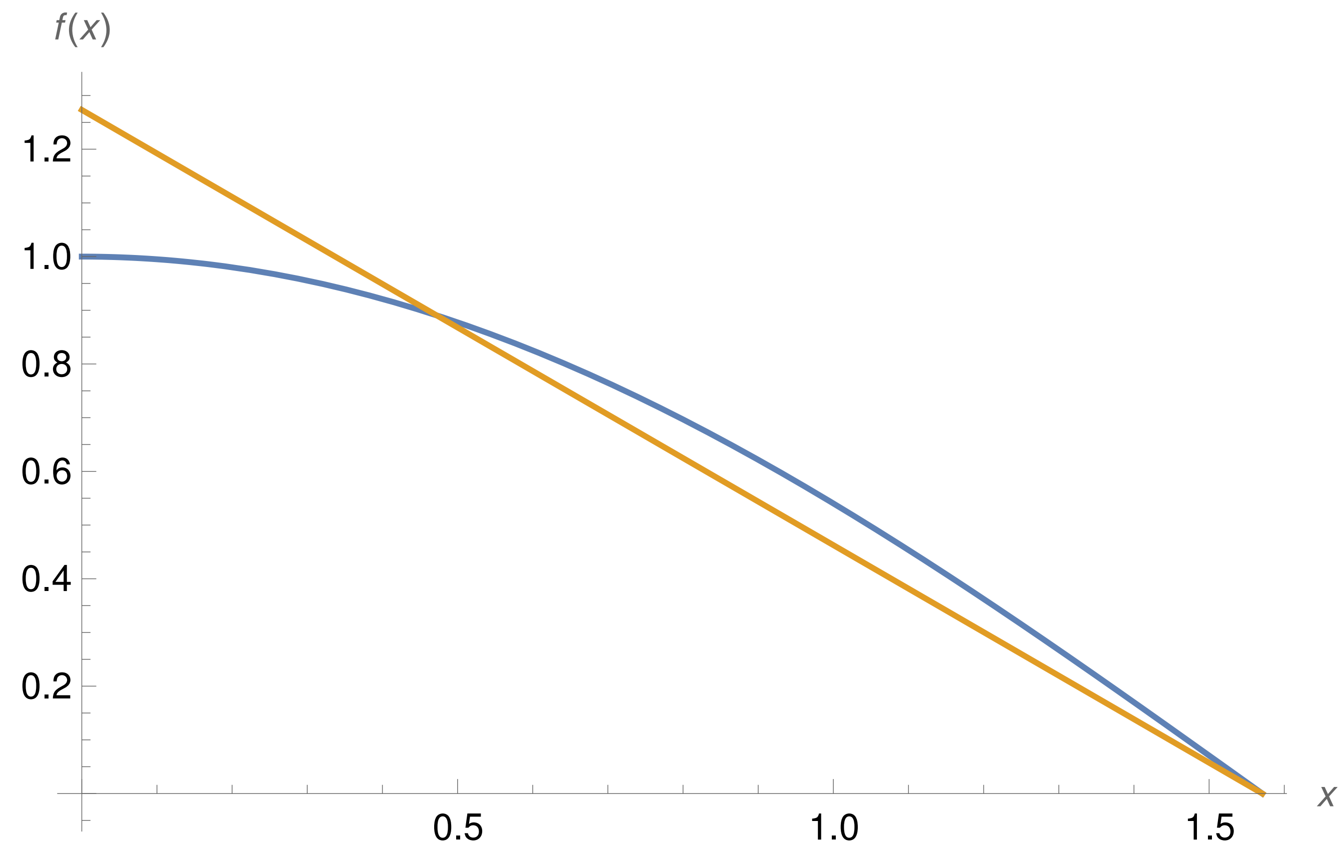
IMPORTANCE SAMPLING

- Suppose $p(x) = 4(1 - 2x/\pi)/\pi$; CDF is

$$P(x) = \frac{4}{\pi}\int_0^x (1 - 2u/\pi)\ \mathrm{d}u = \frac{4x}{\pi}\left(1 - \frac{x}{\pi}\right)$$

- For a random variable ρ

$$\rho = \frac{4x}{\pi}\left(1 - \frac{x}{\pi}\right)$$
$$x = \frac{\pi}{2}\left(1 - \sqrt{1 - \rho}\right)$$



IMPORTANCE SAMPLING

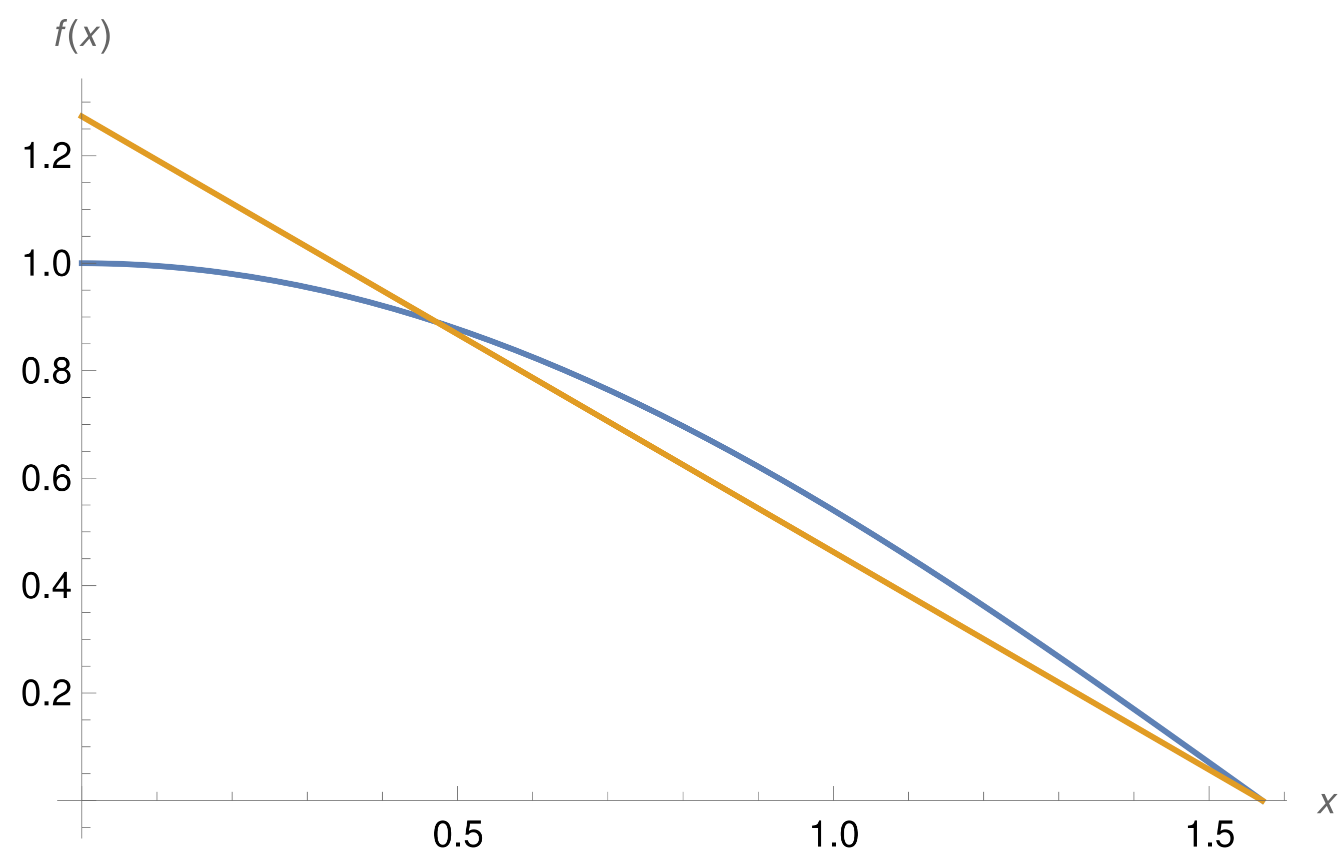
$$I = \int_0^{\pi/2} \cos(x)\,\mathrm{d}x$$

$$= \int_0^{\pi/2} \frac{\cos(x)}{1 - \frac{2}{\pi}x}\left(1 - \frac{2}{\pi}x\right)\,\mathrm{d}x$$

$$= \int_0^{1} \frac{\pi}{4}\frac{\cos(x)}{1 - \frac{2}{\pi}x}\bigg|_{x=x(\rho)}\,\mathrm{d}\rho$$

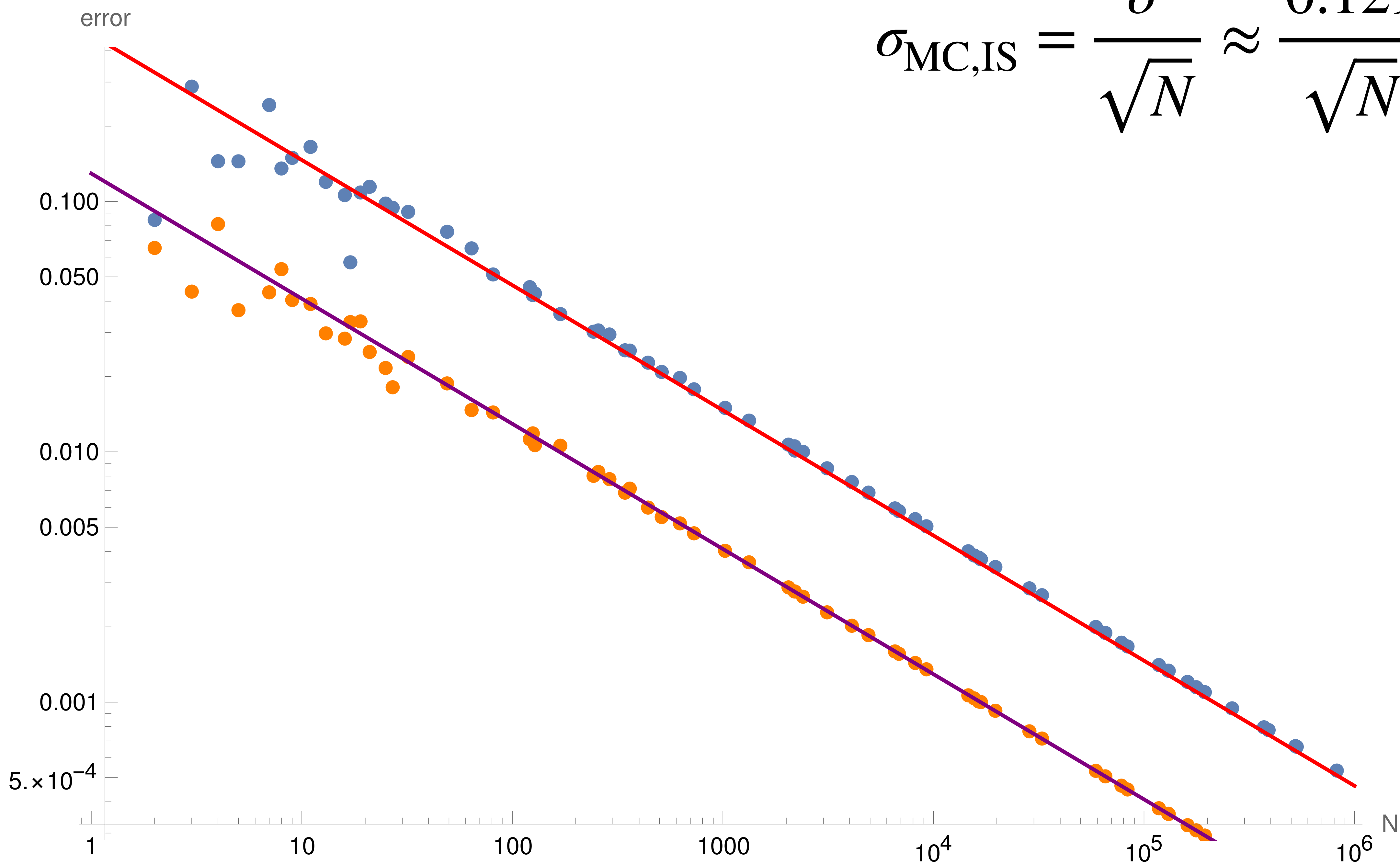
$$x = \frac{\pi}{2}\left(1 - \sqrt{1-\rho}\right)$$

$$\to x = \frac{\pi}{2}\left(1 - \sqrt{\rho}\right)$$



IMPORTANCE SAMPLING

$$\sigma_{\mathrm{MC,IS}} = \frac{\sigma'}{\sqrt{N}} \approx \frac{0.129}{\sqrt{N}}$$



IMPORTANCE SAMPLING

- Breit-Wigner peaks appear in many cross section matrix elements

$$I = \int_{s_0}^{s_1} \frac{\mathrm{d}s}{(s-m^2)^2 + m^2\Gamma^2} = \frac{1}{m\Gamma} \arctan \frac{s-m^2}{m\Gamma} \Big|_{s_0}^{s_1}$$

- Inversion method gives

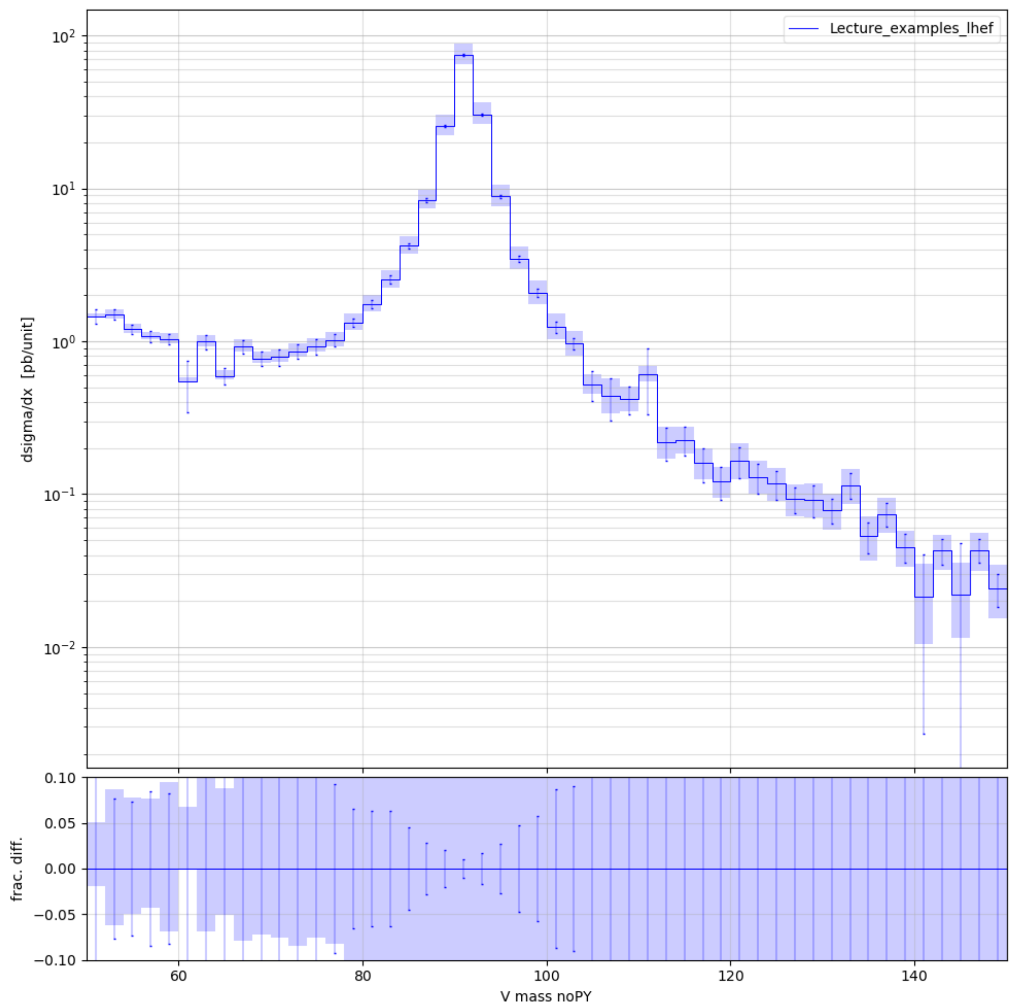
$$f(s) = \frac{m\Gamma}{(s-m^2)^2 + m^2\Gamma^2}$$

$$F(s) = \arctan \frac{s-m^2}{m\Gamma} = \rho$$

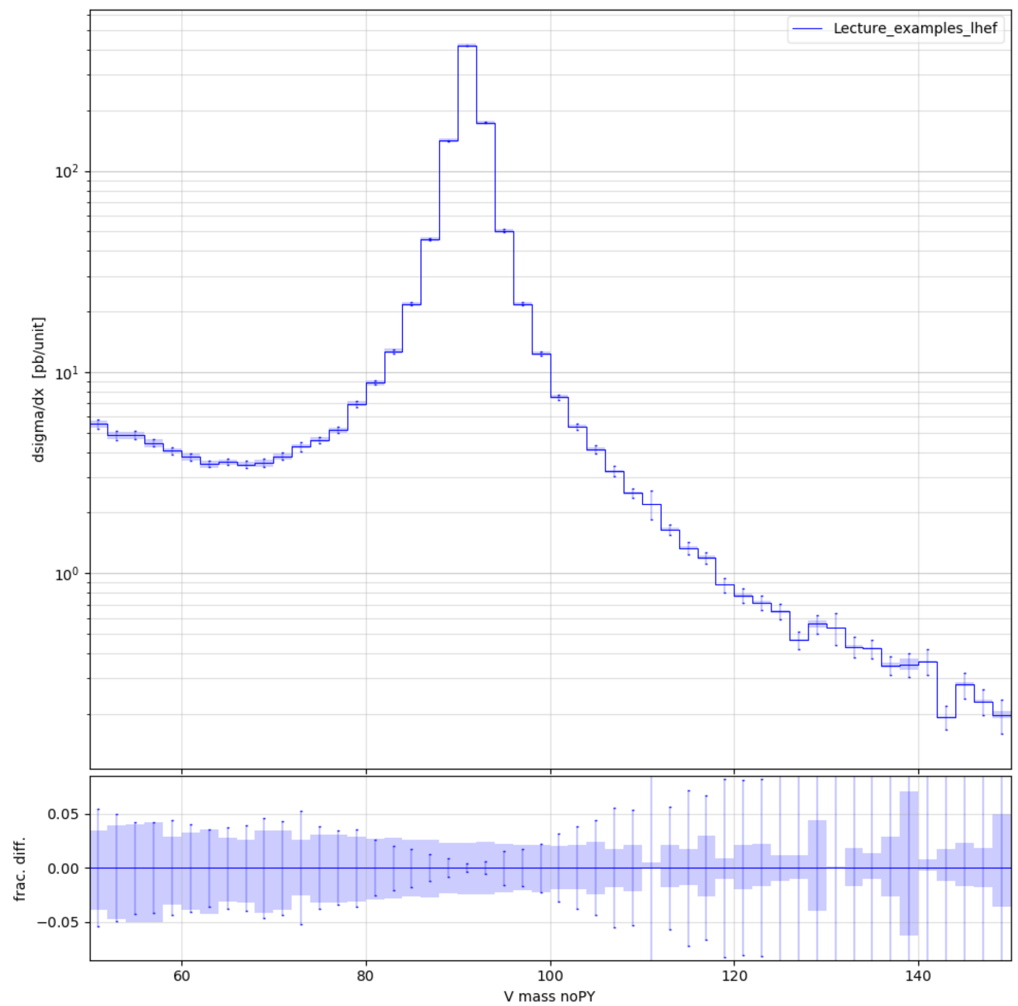
$$F^{-1}(\rho) = m^2 + m\Gamma \tan \rho$$

WHY DO MY RESULTS SUCK? #3

- Importance sampling leads to increased efficiency - fewer events needed to reach same accuracy



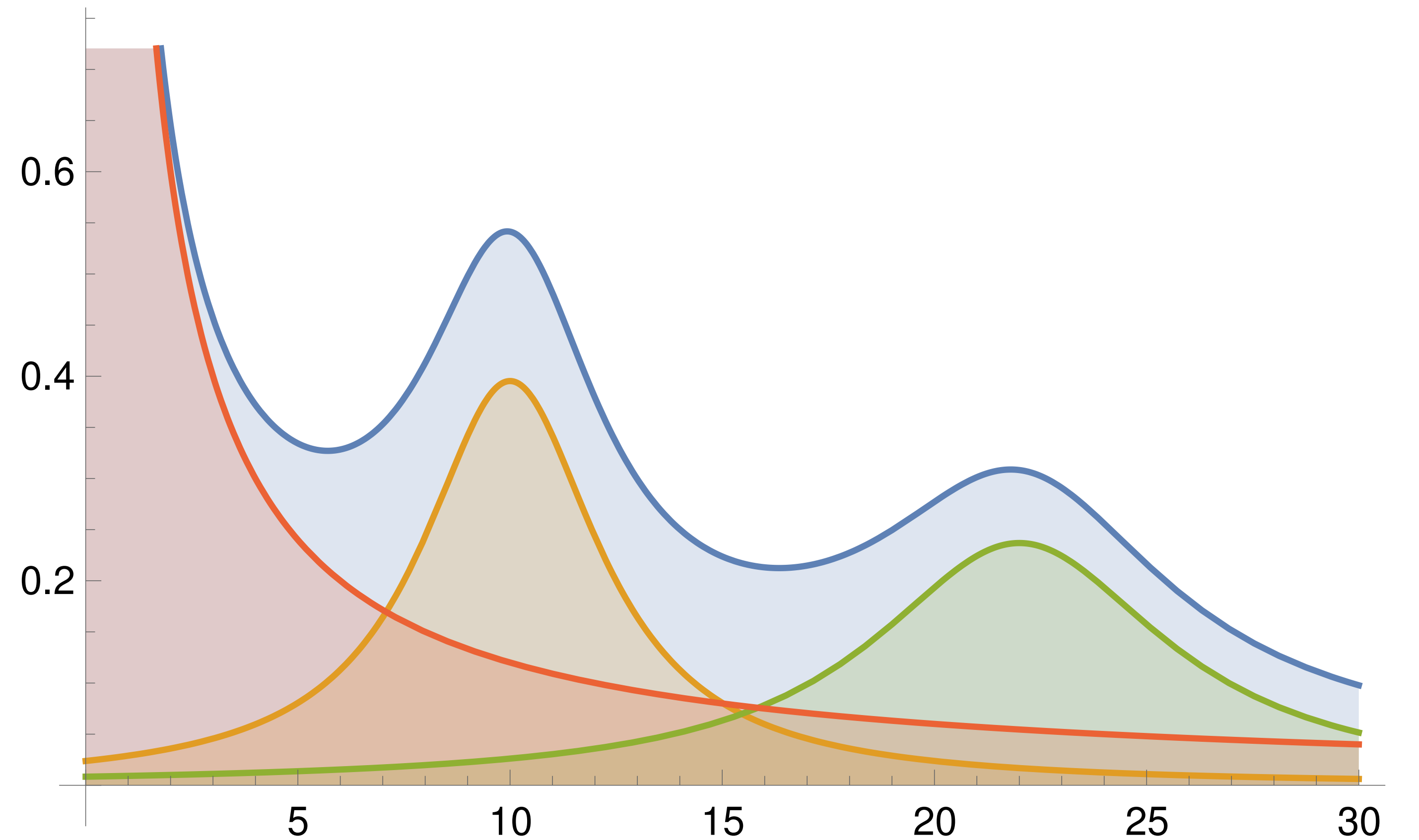
No importance sampling



Importance sampled

MULTI-CHANNEL MONTE CARLO

- Real matrix elements have complicated structures - multiple resonance peaks, s- and t-channel contributions etc.
- Can encode knowledge of peak structure in sum of sample functions with different weights



$$g(s) = \sum_i \alpha_i \, g_i(s)$$

MULTI-CHANNEL MONTE CARLO

- Rewrite

$$\int_{s_0}^{s_1} f(s)\,\mathrm{d}s = \int_{s_0}^{s_1} \frac{f(s)}{g(s)} g(s)\,\mathrm{d}s$$
$$= \int_{s_0}^{s_1} \frac{f(s)}{g(s)} \sum_i \alpha_i g_i(s)\,\mathrm{d}s$$
$$= \sum_i \alpha_i \int_{s_0}^{s_1} \frac{f(s)}{g(s)} g_i(s)\,\mathrm{d}s$$

- Select the distribution $g_i(s)$ to sample from next according to weights α_i
- Methods exist to automatically optimise α_i

SUMMARY

- Monte Carlo converges slowly in few dimensions, but quickly in many
- Functional form of integrand and domain can be arbitrarily complex
- Easy to parallelise - start many instances with different seeds on a supercomputer, combine at the end ("embarrassingly parallel")