Monte Carlo Event Generators

Alan Price (Sherpa)
Midsummer school in QCD 2024
Saariselkä

JAGIELLONIAN UNIVERSITY
IN KRAKÓW

Literature

- F. James **Monte Carlo theory and practice** Rep. Prog. Phys. 43 (1980) 1145
- R. K. Ellis, W. J. Stirling, B. R. Webber, **QCD and Collider Physics**
- Campbell, Huston, Krauss, **The Black Book of Quantum Chromodynamics** (freely available at https://scoap3.org/scoap3-books/)
- MCNet School slides **https://www.montecarlonet.org/schools/**

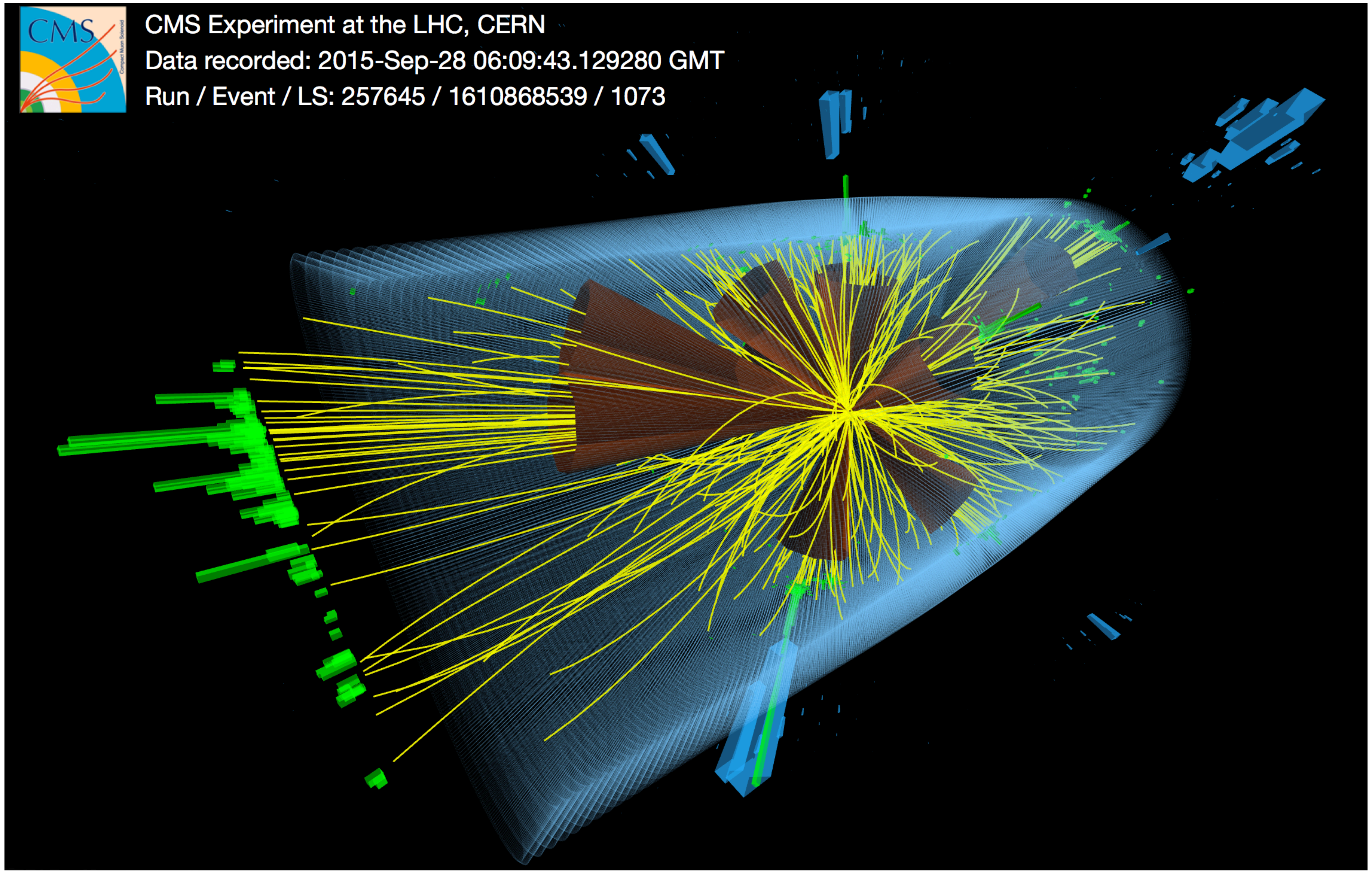
Goal

Try to falsify theoretical models by comparison with data:

- Define observables that can be measured experimentally
  - Cross section for $pp \to X$ which is defined in terms of identified particles, acceptance cuts, isolation criteria, etc.
  - Differential distributions $p_T, \eta, m_x \ldots .$
- Evaluate these observables in the SM and/or your favourite BSM
- **Compare to data**

Goal

We need to calculate the probability of events like this



Goal

We need to calculate the probability of events like this

- **Drell-Yan:** $pp \to \ell\bar{\ell} + \mathcal{O}(100)$
- **Four-Leptons:** $pp \to 4\ell + \mathcal{O}(150)$
- $t\bar{t}$ **Production:** $pp \to t\bar{t} + \mathcal{O}(700)$
- $t\bar{t}h$ **Production:** $pp \to t\bar{t}h + \mathcal{O}(1200)$

Any event at the LHC will contain large number of particles in the final state that must be modelled

Monte Carlo Event Generators

HERWIG

Traditional focus on showers, Qtilde and Dipoles shower, cluster hadronization model, NLO matching and merging.

PYTHIA

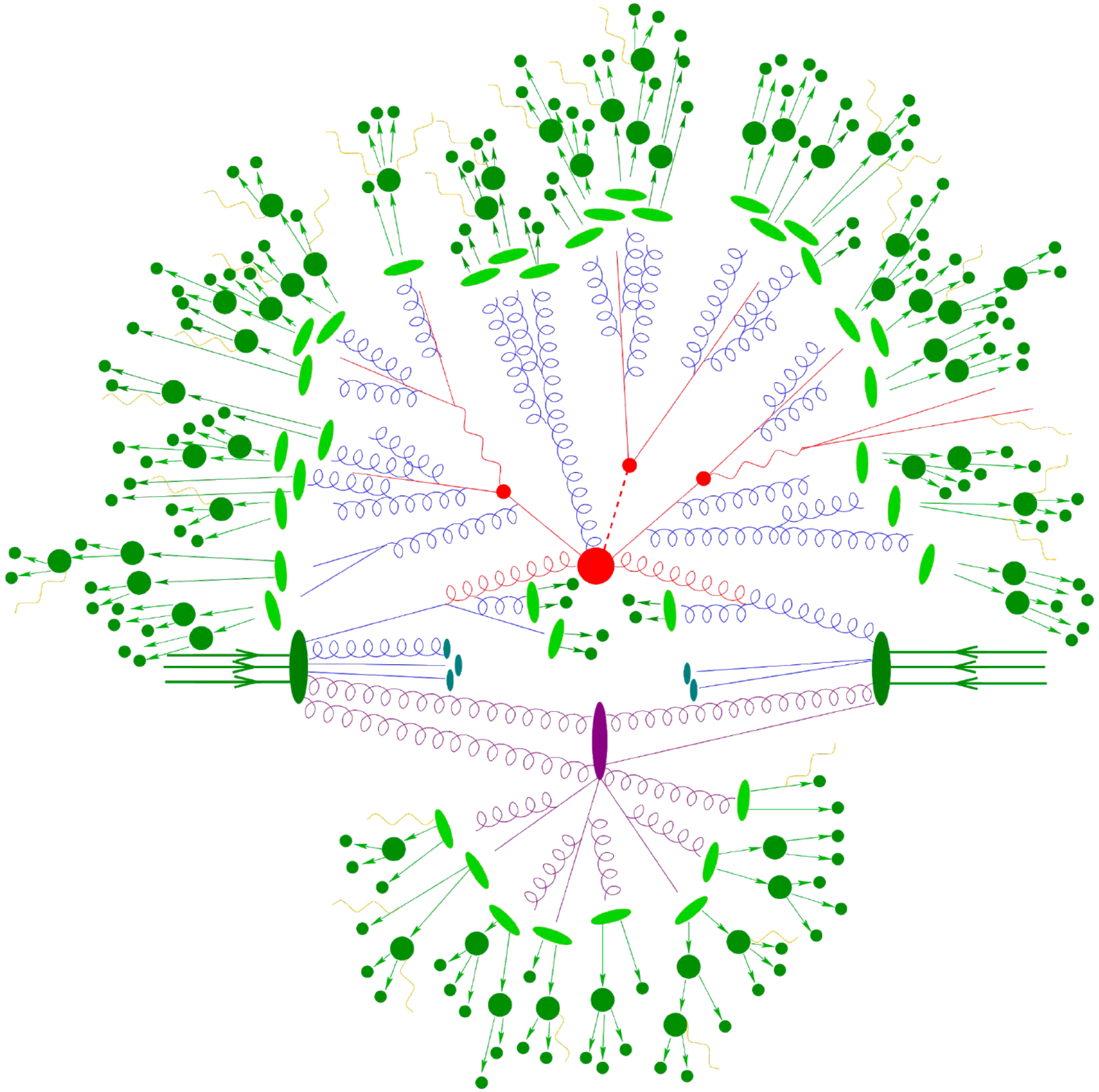
Sophisticated soft physics, pt-ordered, DIRE and Vincia shower, string hadronization, NLO merging.

SHERPA

Focus on perturbative improvements, CS and DIRE shower, cluster or string hadronization, NLO matching and merging.

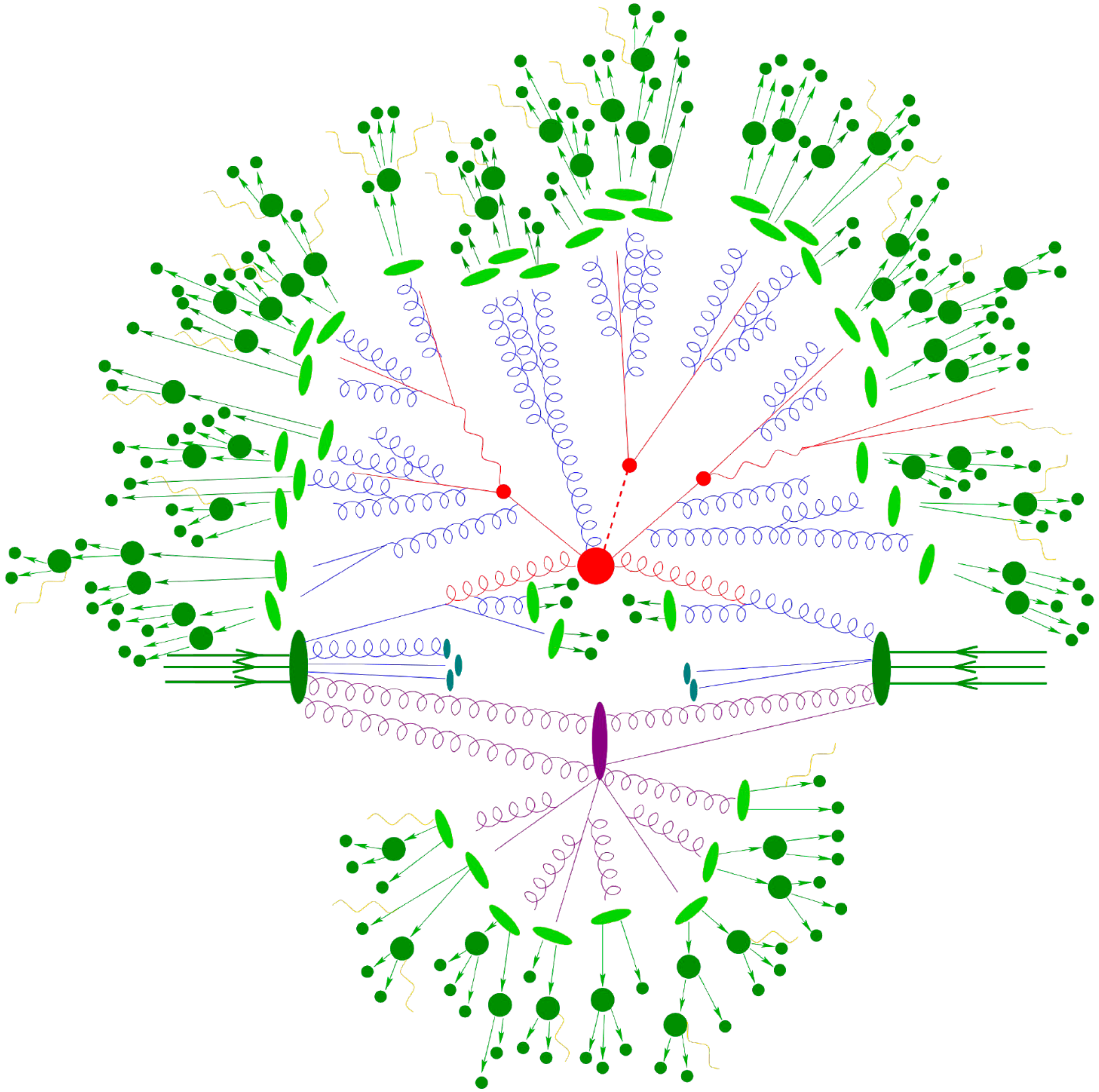
Divide and Conquer

- Hard Interaction
- Radiative Corrections
- Hadronization
- Hadron Decays
- Underlying Event



Divide and Conquer

- Hard Interaction ★
- Radiative Corrections
- Hadronization
- Hadron Decays
- Underlying Event



How do we calculate observables?

$$\langle O \rangle = \int d\Phi_n \int dx_1 \int dx_2 \, f_i(x_1, \mu_F^2) \left| \mathcal{M} \left(ab \to X; \mu_F^2, \mu_R^2 \right) \right|^2 f_j(x_2, \mu_F^2) \, O(\Phi)$$

Step 1:

Calculate the matrix element for your process of choice.

Step 2:

Preform the multidimensional integral

Next Section

Hard Scattering: Matrix Elements

How do we calculate $\left|\mathcal{M}\left(ab \to X; \mu_F^2, \mu_R^2\right)\right|^2$

- For low multiplicities we can do it by hand
- Automated Tools **Matrix Element Generators**
  - **CalcHEP**
  - **Comix/AMEGIC**
  - **HELAC**
  - **Whizard**
  - **…**
- **FeynRules** for BSM

Hard Scattering: Matrix Elements

How do we calculate $\left| \mathscr{M}\left(ab \to X; \mu_F^2, \mu_R^2\right) \right|^2$

- **Textbook:** Draw the Feynman diagrams, apply the rules, sum over the external states, find a mistake and start again.

- **Reality:** Realise that amplitudes are just complex numbers. Compute them, then sum and square

$$B = \sum_{\text{color}} |A|^2 \sum_{\text{spin}} |\mathscr{M}|^2$$

$$\langle O \rangle^{LO} = \int d\Phi_B \, B(\Phi_B) \, O(\Phi_B)$$

There is also a dependence on a scale choice which I will suppress for now

Leading Order Matrix Elements

- **Amplitudes = Complex numbers**

$$B = \sum_{\text{color}} |A|^2 \sum_{\text{spin}} |\mathscr{M}|^2$$

- With a chosen basis all components of an amplitude can be expressed explicitly

$$\langle O \rangle^{LO} = \int d\Phi_B \, B(\Phi_B) \, O(\Phi_B)$$

- Matrix multiplication is costly! Effort still grows linearly with the number of diagrams

$$\bar{u}(p_1, h_1)\Gamma(p_1, \ldots, p_{n-1}, h_1, \ldots, h_{n-1})u(p_n, h_n)$$

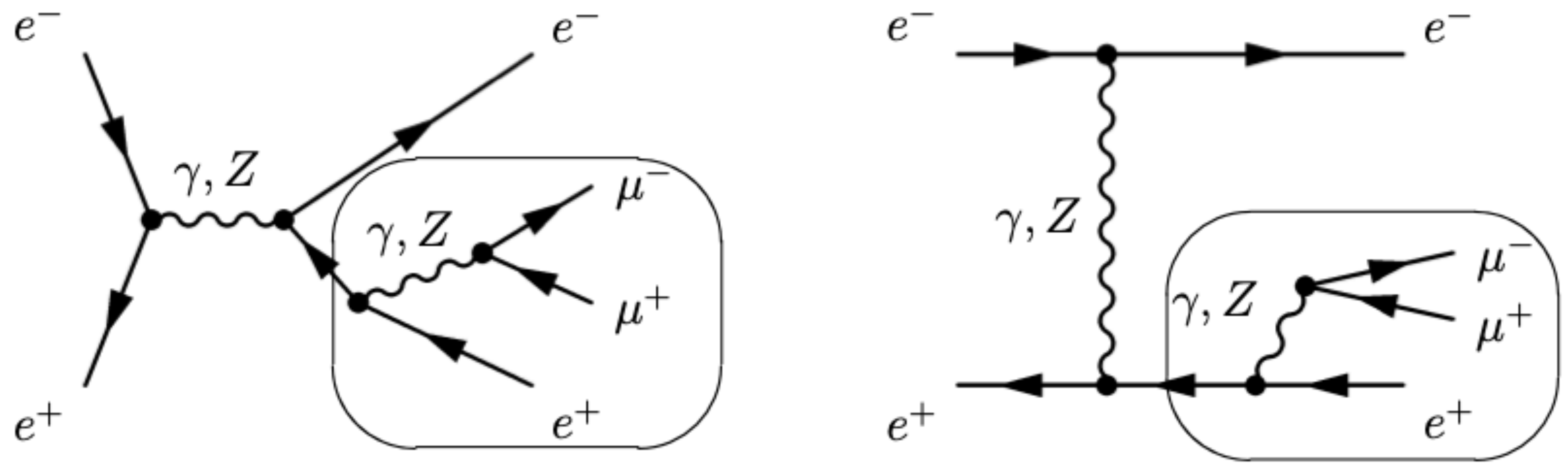
Leading Order Matrix Elements

Can we improve this?

$$B = \sum_{\text{color}} |A|^2 \sum_{\text{spin}} |\mathscr{M}|^2$$

$$\langle O \rangle^{LO} = \int d\Phi_B \, B(\Phi_B) \, O(\Phi_B)$$

- **Repeated Subgraphs:** Many diagrams will share the same subgraph



- Calculate once and reuse again.

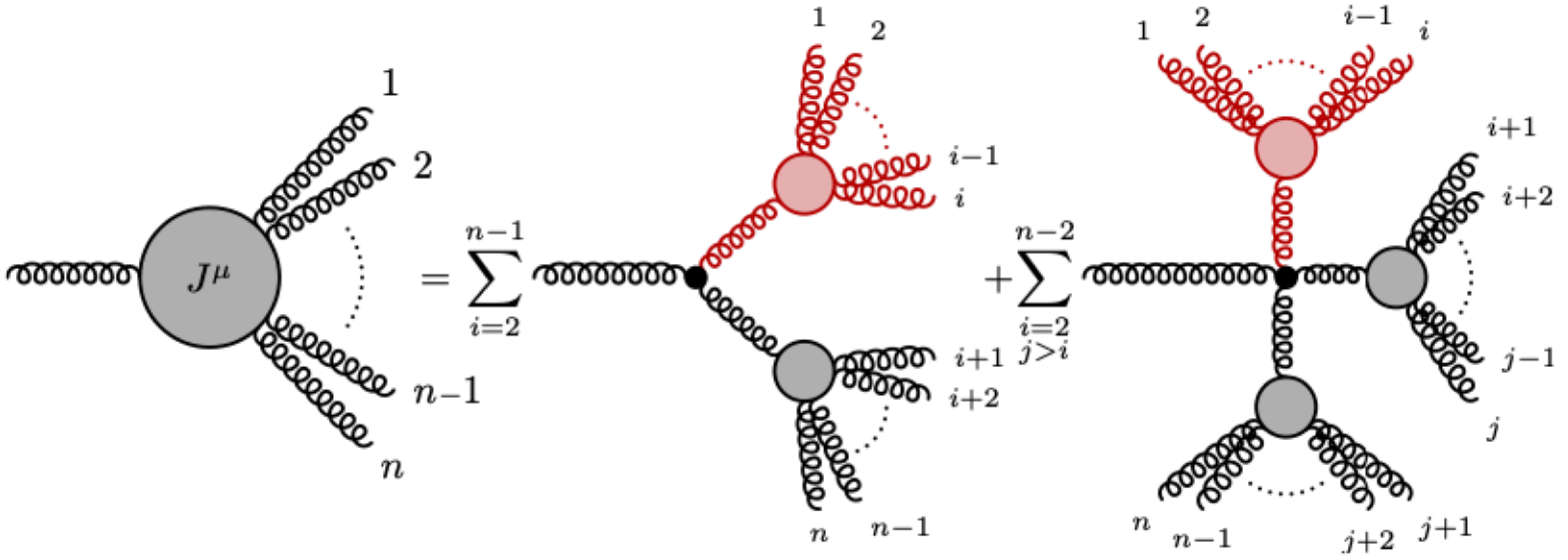
Leading Order Matrix Elements

Recurrence Relations

$$B = \sum_{\text{color}} |A|^2 \sum_{\text{spin}} |\mathcal{M}|^2$$

$$\langle O \rangle^{LO} = \int d\Phi_B \, B(\Phi_B) \, O(\Phi_B)$$

- We know that the complexity of amplitudes grows factorial with the number of external legs



Berends, Giele NPB306(1988)759

Cachazo, Svrcek, Witten JHEP09(2004)006

Britto, Cachazo, Feng NPB715(2005)499

- Use recurrence relations to reduce the overhead

Leading Order Matrix Elements

Helicity and Color Sums

$$B = \sum_{\text{color,spin}} (A\mathscr{M}) \cdot (A\mathscr{M})^\dagger$$

$$\langle O \rangle^{LO} = \int d\Phi_B \, B(\Phi_B) \, O(\Phi_B)$$

- **Helicity:** Not all helicity configurations contribute equally.
- **Solution:** Only generate amplitudes for one helicity configuration and include helicity as a dof in the Phase-space integral
- **Color:** Not all color configurations contribute equally.
- **Solution:** Only generate amplitudes for one helicity configuration and include color as a dof in the Phase-space integral

Leading Order Matrix Elements

Recurrence Relations

Duhr, Höche, Maltoni JHEP08(2006)062

Final State	BG		BCF		CSW	
	CO	CD	CO	CD	CO	CD
$2g$	0.24	0.28	0.28	0.33	0.31	0.26
$3g$	0.45	0.48	0.42	0.51	0.57	0.55
$4g$	1.20	1.04	0.84	1.32	1.63	1.75
$5g$	3.78	2.69	2.59	7.26	5.95	5.96
$6g$	14.2	7.19	11.9	59.1	27.8	30.6
$7g$	58.5	23.7	73.6	646	146	195
$8g$	276	82.1	597	8690	919	1890
$9g$	1450	270	5900	127000	6310	29700
$10g$	7960	864	64000	-	48900	-

Table 3: Computation time (s) of the $2 \to n$ gluon amplitudes for 10^4 phase space points, sampled over helicity and color. Results are given for the color-ordered (CO) and the color-dressed (CD) Berends-Giele (BG), Britto-Cachazo-Feng (BCF) and Cachazo-Svrček-Witten (CSW) relations. Numbers were generated on a 2.66 GHz Xeon$^{\mathrm{TM}}$ CPU.

NLO Matrix Elements

$$\langle O \rangle^{NLO} = \int d\Phi_B \, [B(\Phi_B) + V(\Phi_B)] \, O(\Phi_B) + \int d\Phi_R \, R(\Phi_R) \, O(\Phi_R)$$

- At **NLO** we also have to include real and virtual emissions
- $V(\Phi_B)$ virtual corrections
- $R(\Phi_R)$ real corrections
- Individually, both V and R have IR divergences but there sum is IR finite **KLN Theorem,** however they both live in separate phase spaces

NLO Matrix Elements

$$\langle O \rangle^{NLO} = \int d\Phi_B \, [B(\Phi_B) + V(\Phi_B)] \, O(\Phi_B) + \int d\Phi_R \, R(\Phi_R) \, O(\Phi_R)$$

IR Divergences

Arise in V from integrations over loop momenta

Arise in R from integrations over soft-collinear momenta

For an IR safe observable they must be removed

Subtraction Method

Create universal subtraction terms that reproduce R in the soft-collinear limit

NLO Matrix Elements: Adding Zero

Subtraction Method

Create universal subtraction terms that reproduce R in the soft-collinear limit

$$\langle O \rangle^{NLO} = \int d\Phi_B \left[B(\Phi_B) + V(\Phi_B) \qquad \right] O(\Phi_B) + \int d\Phi_R \left[R(\Phi_R) \qquad \right] O(\Phi_R)$$

NLO Matrix Elements: Adding Zero

Subtraction Method

Create universal subtraction terms that reproduce R in the soft-collinear limit

$$\langle O \rangle^{NLO} = \int d\Phi_B \left[B(\Phi_B) + V(\Phi_B) \qquad \right] O(\Phi_B) + \int d\Phi_R \left[R(\Phi_R) - S(\Phi_R) \right] O(\Phi_R)$$

- We subtract a term from R, removing IR divergences. Now we have to add it back

NLO Matrix Elements: Adding Zero

Subtraction Method

Create universal subtraction terms that reproduce R in the soft-collinear limit

$$\langle O \rangle^{NLO} = \int d\Phi_B \left[B(\Phi_B) + V(\Phi_B) + I(\Phi_B)\right] O(\Phi_B) + \int d\Phi_R \left[R(\Phi_R) - S(\Phi_R)\right] O(\Phi_R)$$

- We subtract a term from R, removing IR divergences. Now we have to add it back
- Add an integrated subtraction term to the Born phasespace
- Now both integrals are **separately IR finite** and can be treated with MC methods

Real and Virtual Corrections

Real

- These are tree-level diagrams, use the same methods as born

Virtual

- Reduce 1-loop integral into master integrals

$$\mathscr{M}^{\text{loop}} = D \times (\text{Box}) + C \times (\text{Triangle}) + B \times (\text{Bubble}) + A \times (\text{Tadpole}) + R$$

- D,C,B,A,R are coefficients that can be calculated with either tensor reduction or unitarity cuts

One-Loop corrections are automated these days with
tools like …

How do we calculate observables?

$$\langle O \rangle = \int d\Phi_n \int dx_1 \int dx_2 \, f_i(x_1, \mu_F^2) \, \left| \mathscr{M} \left(ab \to X; \mu_F^2, \mu_R^2 \right) \right|^2 f_j(x_2, \mu_F^2) \, O(\Phi)$$

Step 1:

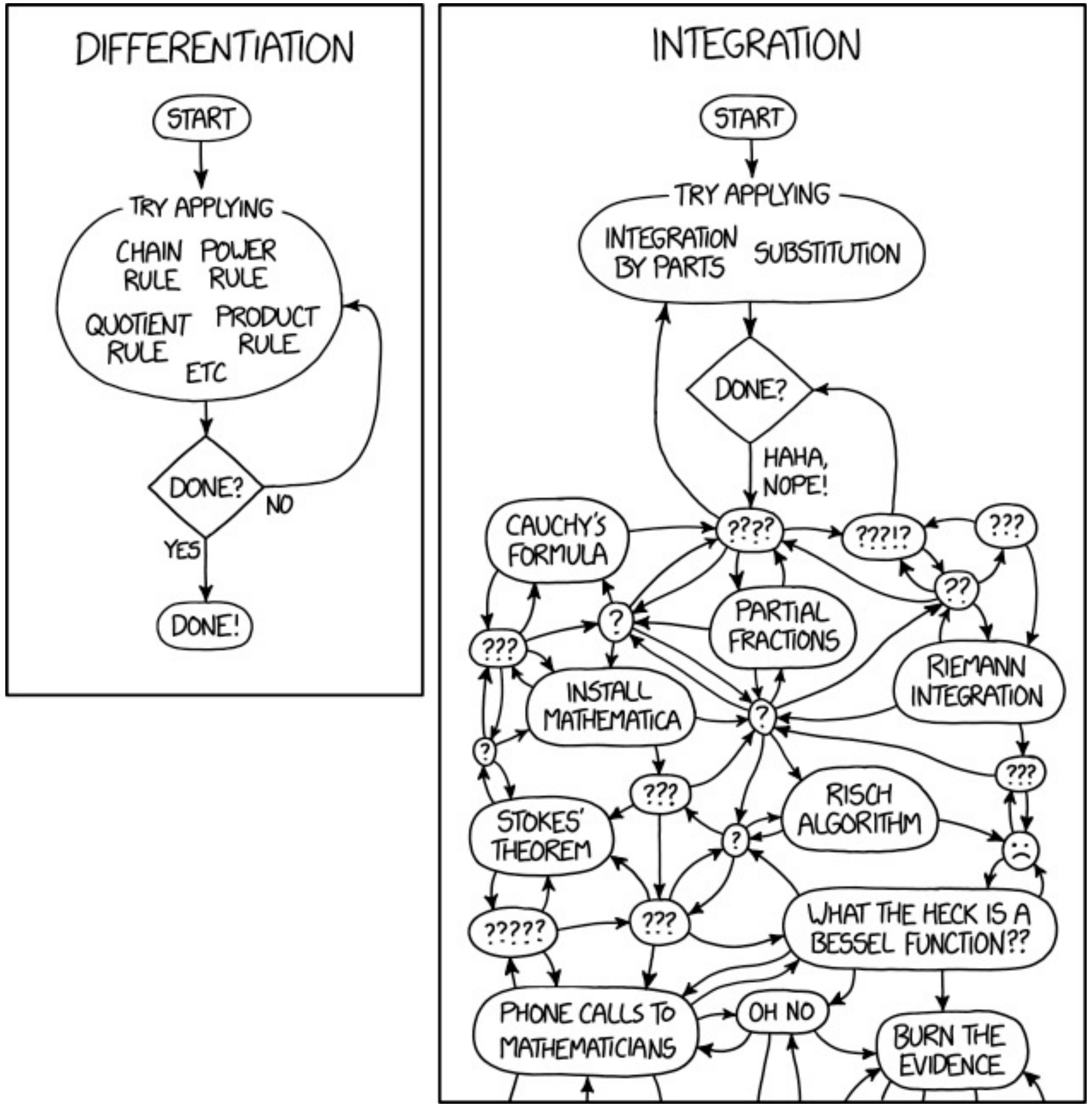
Calculate the matrix element for your process of choice.

Step 2:

Preform the multidimensional integral

~~Next Section.~~ Now

Integration



https://xkcd.com/2117/

Integration Tricks

There are some tricks to doing integrals. One is to look them up in a table of integrals. Another is to learn Mathematica

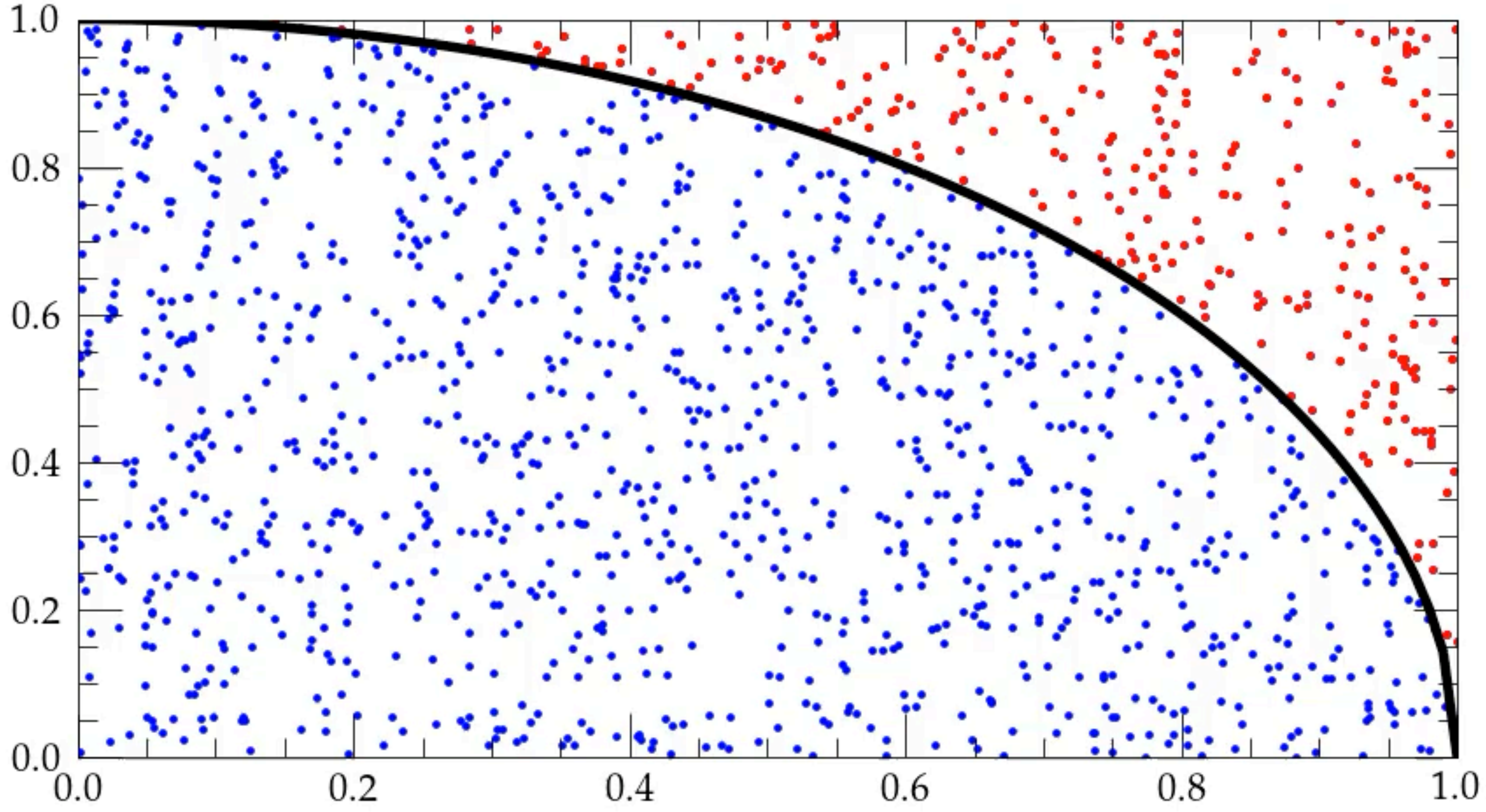
Leonard Susskind

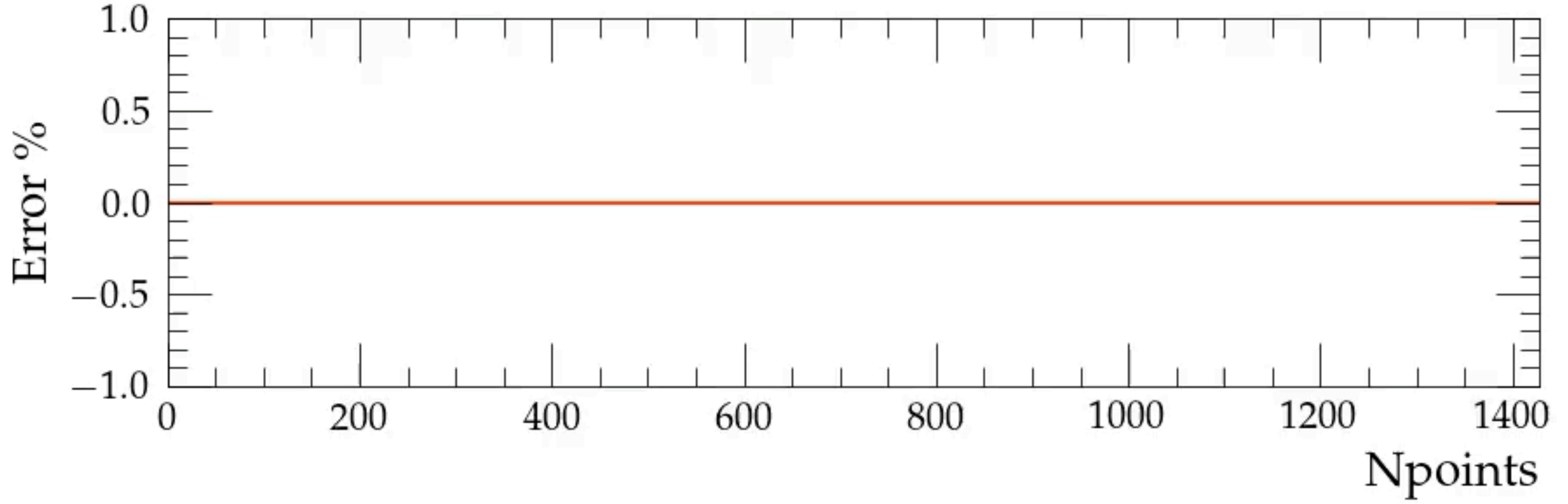
Hit or Miss: Calculating π

$$f(x) = \sqrt{1 - x^2} \quad x \in (0,1)$$

- Randomly choose $x, y \in [0,1] \otimes [0,1]$
- If $y > f(x)$ reject point

$$\pi = 4 \int_0^1 f(x)dx \approx 4 \left(\frac{\text{Accepted}}{\text{Total}} \right)$$

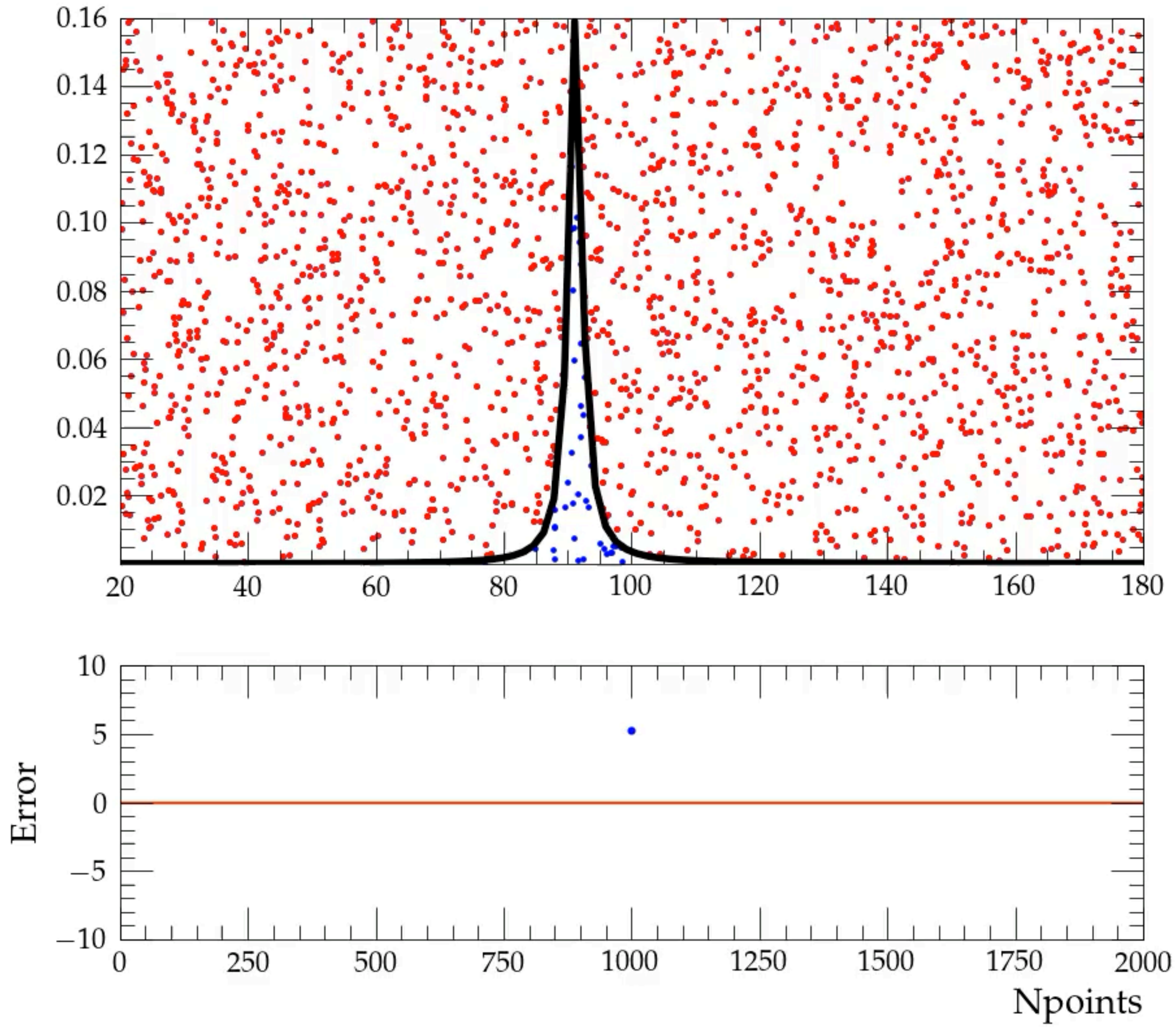




Hit or Miss: Limitations

$$f(x) = \frac{M_Z \Gamma_Z}{(s - M_Z^2)^2 + M_Z^2 \Gamma_Z^2}$$

- Not very efficient for narrow resonances
- ~100k points to get below 1% error
- In reality we will have multiple such structures



Weights, Averages, and Variance

- We can relate the average of $f(x)$ to its integral as:

$$\langle f(x) \rangle = \frac{1}{b-a} \int_a^b f(x)\, d(x)$$

- We can also estimate the average by choosing random points

$$\langle p(x) \rangle_E = \frac{1}{N} \sum_{i=1}^{N} p(x_i)$$

- Notation: We will call $p(x_i)$ ith weight w_i of the event x_i.

Weights, Averages, and Variance

- Now that we have an estimate $\langle p(x) \rangle_E$, we can also define its variance

$$\sigma^2 = \frac{1}{N-1}\left(\frac{1}{N}\sum_{i=1}^{N} w_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N} w_i\right)^2\right)$$

- So if $p(x)$ has a large variance it will require many samples, as we have seen

Weights, Averages, and Variance

- So how do we reduce this variance?

$$\sigma^2 = \frac{1}{N-1}\left(\frac{1}{N}\sum_{i=1}^{N} w_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N} w_i\right)^2\right)$$

- We know that a constant function has zero variance. How can we exploit this?

Weights, Averages, and Variance

- So how do we reduce this variance?

$$\sigma^2 = \frac{1}{N-1}\left(\frac{1}{N}\sum_{i=1}^{N} w_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N} w_i\right)^2\right)$$

- We know that a constant function has zero variance. How can we exploit this?

We need to find a function $g(x)$ such that $\frac{f(x)}{g(x)} \approx$ constant

Importance Sampling

- The problem with hit or miss is that we sample the phasespace **uniformly.** A more efficient way would be sample from a distribution that is similar to our integrand

$$\int dx\, f(x) = \int dx\, g(x) \frac{f(x)}{g(x)}$$

Importance Sampling

- The problem with hit or miss is that we sample the phasespace **uniformly.** A more efficient way would be sample from a distribution that is similar to our integrand

$$\int dx\, f(x) = \int dx\, g(x)\, \frac{f(x)}{g(x)}$$

Could have non trivial shape e.g Narrow resonances,

making it difficult to sample efficiently

Importance Sampling

- The problem with hit or miss is that we sample the phasespace **uniformly.** A more efficient way would be sample from a distribution that is similar to our integrand

$$\int dx\, f(x) = \int dx\, g(x) \frac{f(x)}{g(x)}$$

With a clever choice of $g(x)$ this can be "flattened"

Importance Sampling

- The problem with hit or miss is that we sample the phasespace **uniformly.** A more efficient way would be sample from a distribution that is similar to our integrand

$$\int dx\, f(x) = \int dx\, g(x)\, \frac{f(x)}{g(x)}$$

Cost: We no longer sample uniformly

Importance Sampling: Choosing $g(x)$

- First we need to ensure $f(x) \leq g(x)$ (at least in the range of interest)
- It must also be relatively simple to integrate as will need to sample from it
- How do you get a random number from $g(x)$?

$$\int_{x_{min}}^{y} dx\, g(x) = \# \int_{x_{min}}^{x_{max}} dx\, g(x)$$

Sampling by Inversion

$$\int_{x_{min}}^{y} dx\, g(x) = \# \int_{x_{min}}^{x_{max}} dx\, g(x)$$

$$\Downarrow$$

$$y = P^{-1}\left(P(x_{min}) + \#\big(P(x_{max}) - P(x_{min})\big)\right)$$

$$P(x) = \int dx\, p(x)$$

Sampling by Inversion: Simple Example

$$p(x) = \frac{1}{x}, \quad x \in [a, b]$$

$$P(x) = \ln(x)$$

$$y = \exp\left(\ln(a) + \#\left(\ln(b) - \ln(a)\right)\right)$$

$$y = a\left(\frac{b}{a}\right)^{\#}$$

Always good to check the limits of the ran space

$$\# \to 0,\ y = a \qquad \# \to 1,\ y = b$$

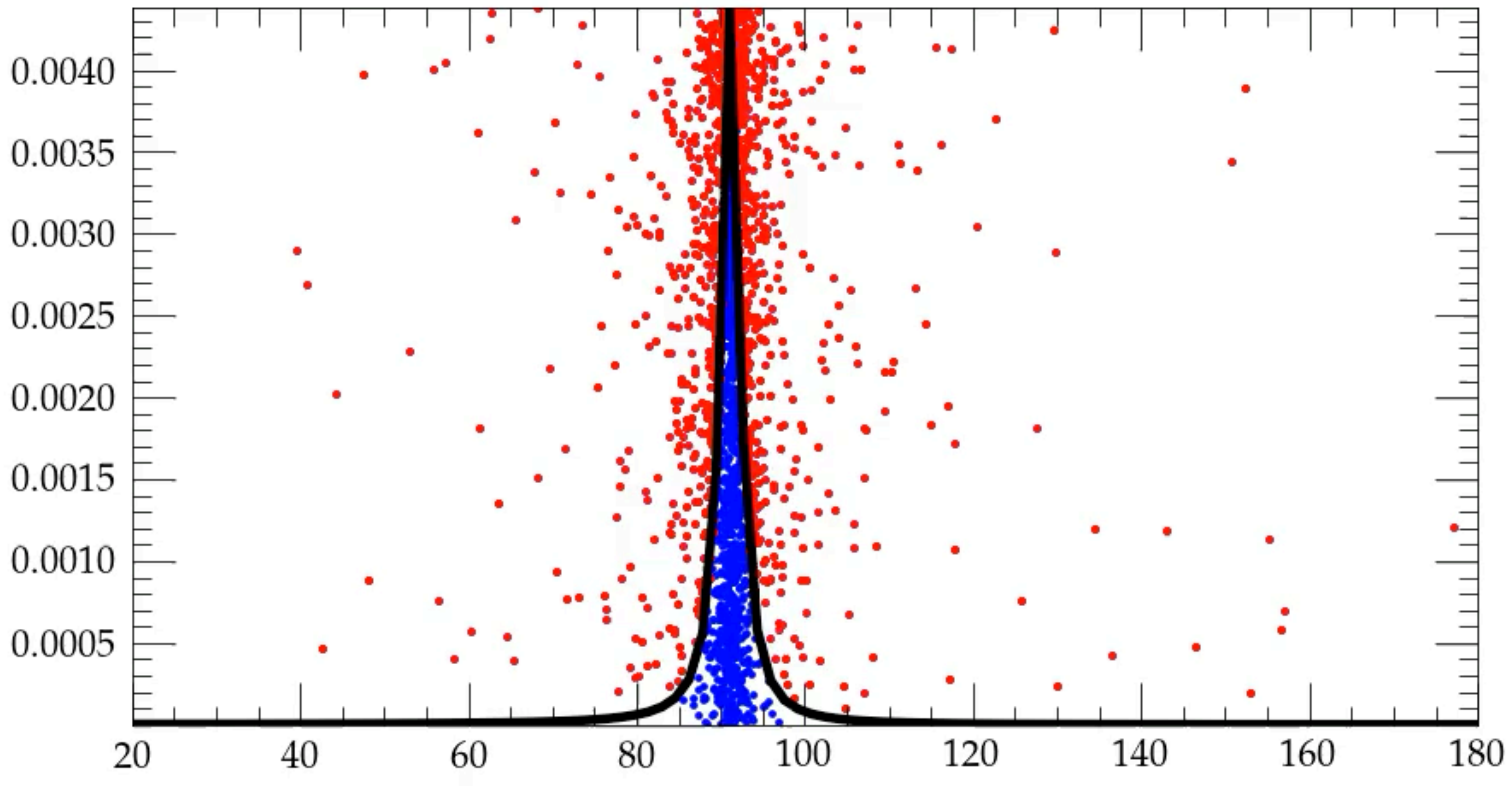
Importance Sampling

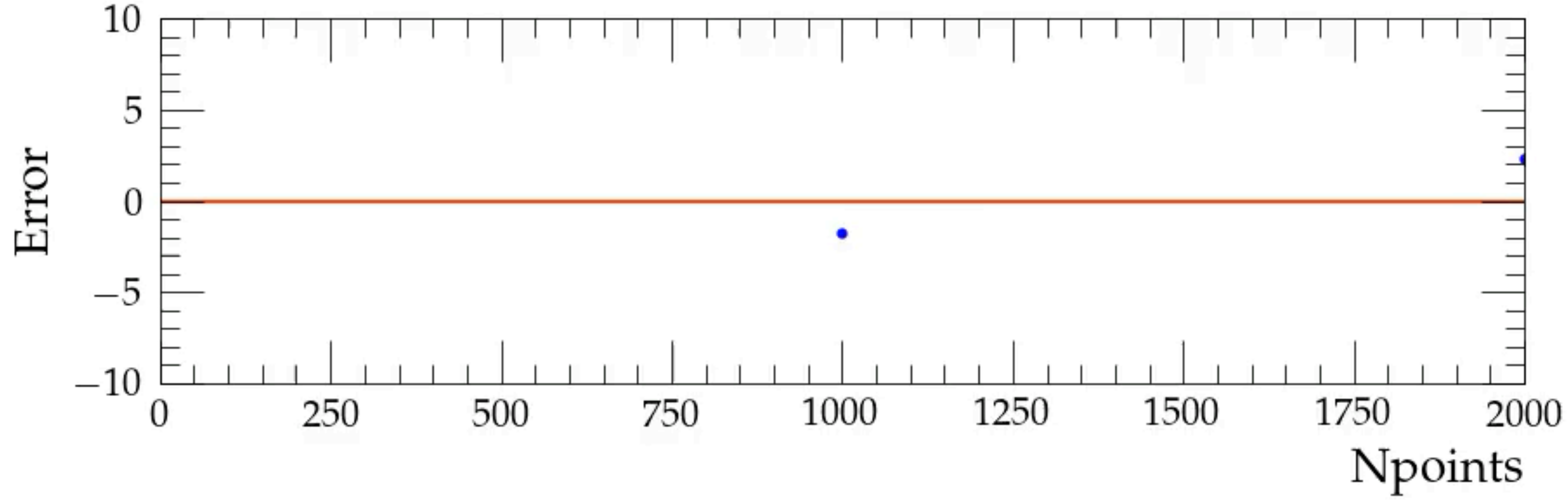
$$f(x) = \frac{M_Z \Gamma_Z}{(s - M_Z^2)^2 + M_Z^2 \Gamma_Z^2}$$

- We can integrate $f(x)$ analytically but it is useful to consider using importance sampling
- "Perfect Choice" $g(x) = f(x)$

$$x = M_Z^2 + \tan\left(z_{\text{min}} + \#(z_{\text{min}} - z_{\text{min}})\right)$$

$$y = \#_2\left(\frac{1}{M_Z G_Z}\right) \qquad z_{\text{min/max}} = \tan^{-1}\left(\frac{s_{\text{min/max}} - M_Z^2}{M_Z G_Z}\right)$$





Why is the sampling not perfect?

Importance Sampling

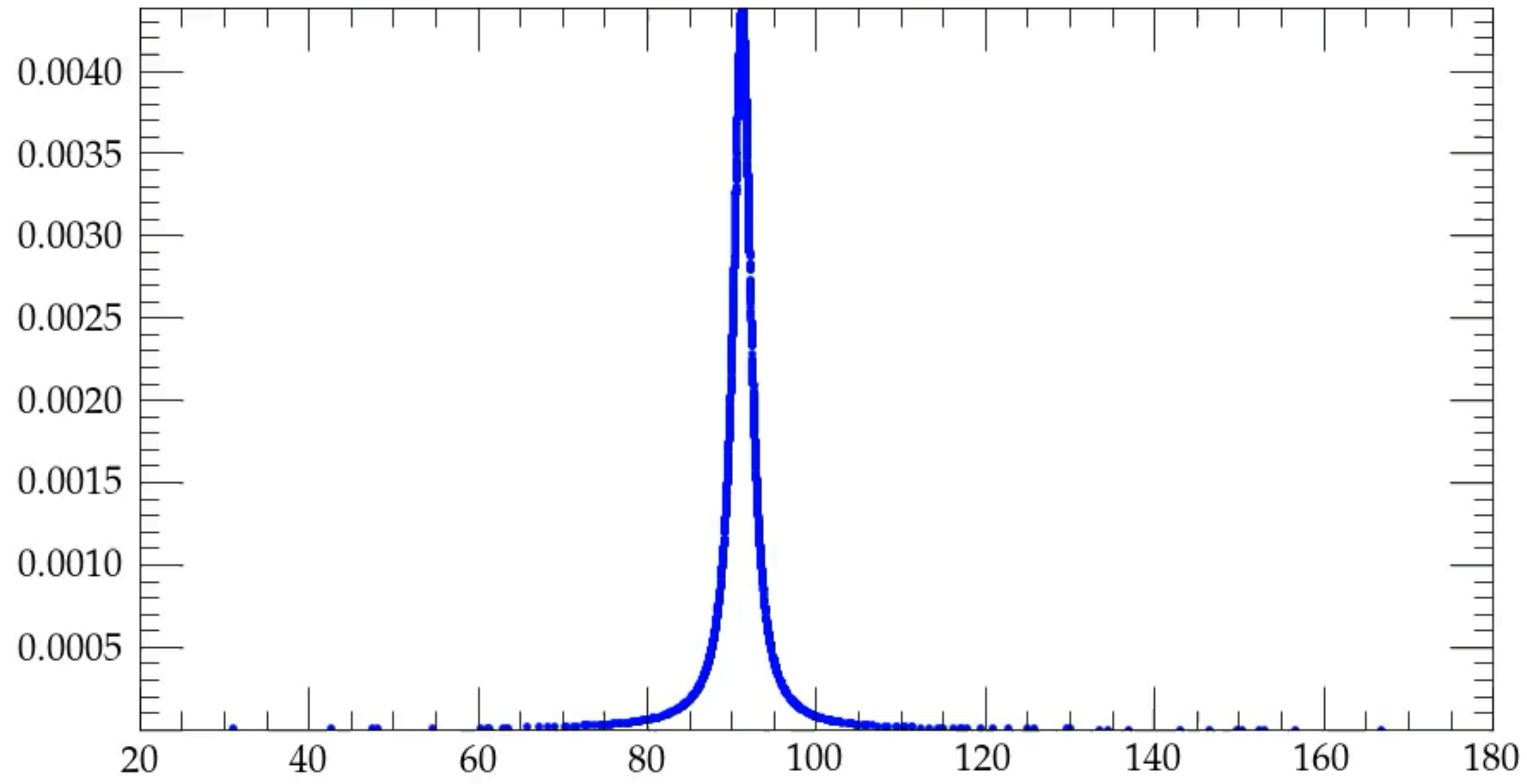
$$f(x) = \frac{M_Z \Gamma_Z}{(s - m_Z^2)^2 + m_Z^2 \Gamma_Z^2}$$

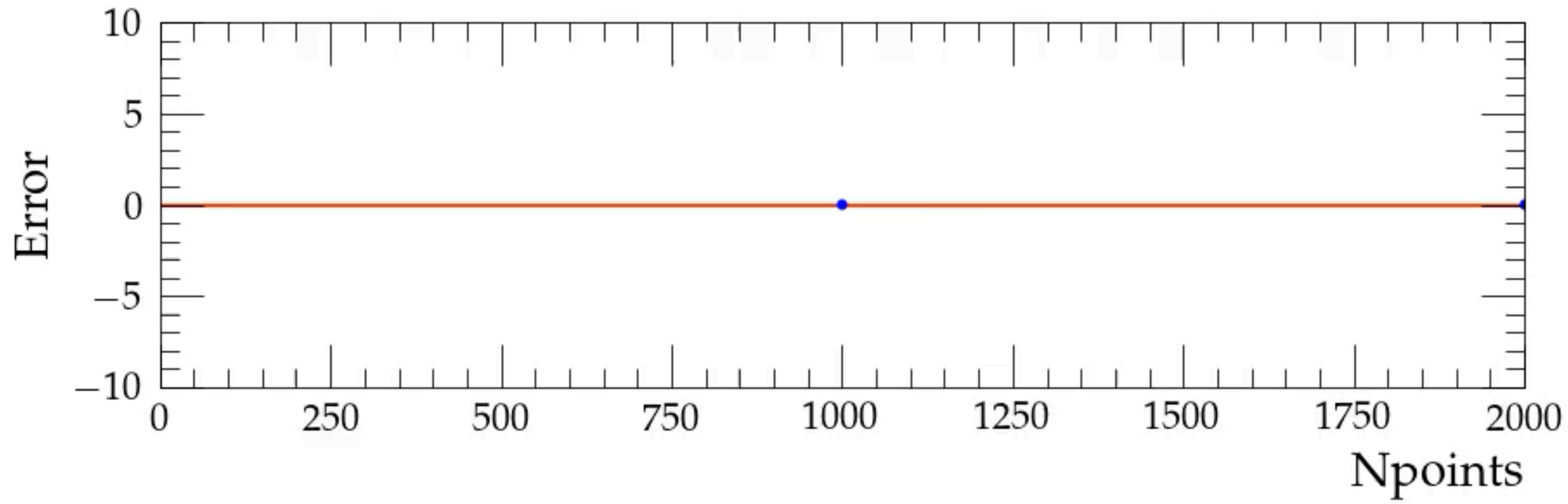
- We can integrate $f(x)$ analytically but it is useful to consider using importance sampling
- "Perfect Choice" $g(x) = f(x)$

$$x = M_Z^2 + \tan\left(z_{\text{min}} + \#(z_{\text{min}} - z_{\text{min}})\right)$$

$$y = \#_2\left(\frac{1}{M_Z G_Z}\right) \qquad z_{\text{min/max}} = \tan^{-1}\left(\frac{s_{\text{min/max}} - M_Z^2}{M_Z G_Z}\right)$$

We did not update $y(x) = f(x)$





Importance Sampling

We have seen how we can use importance sampling for a trivial integral. What about real world examples?

What if you have more than one propagator present?

Multi-Channel

Instead of having one simple estimate $g(x)$, we can use multiple separate channels.
Assuming we know how to sample from each individual $g_i(x)$

$$g(x) = \sum_{i=1}^{N} \alpha_i g_i(x), \qquad \sum_{i=1}^{N} \alpha_i = 1$$

$$I \approx E_N = \frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i)}{g(x_i)} = \frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i)}{\sum_{j=1}^{N} \alpha_j g_j(x)}$$

A channel is chosen at random according to α_i and the sampling proceeds as before.
Initially, all channels have equal probability of being picked. They are then updated based an the weight distribution of the channel Comput. Phys. Commun. 83 (1994), 141–146

Multi-Channel

Instead of having one simple estimate $g(x)$, we can use multiple separate channels.
Assuming we know how to sample from each individual $g_i(x)$

Example: $q\bar{q} \rightarrow e^+e^-$

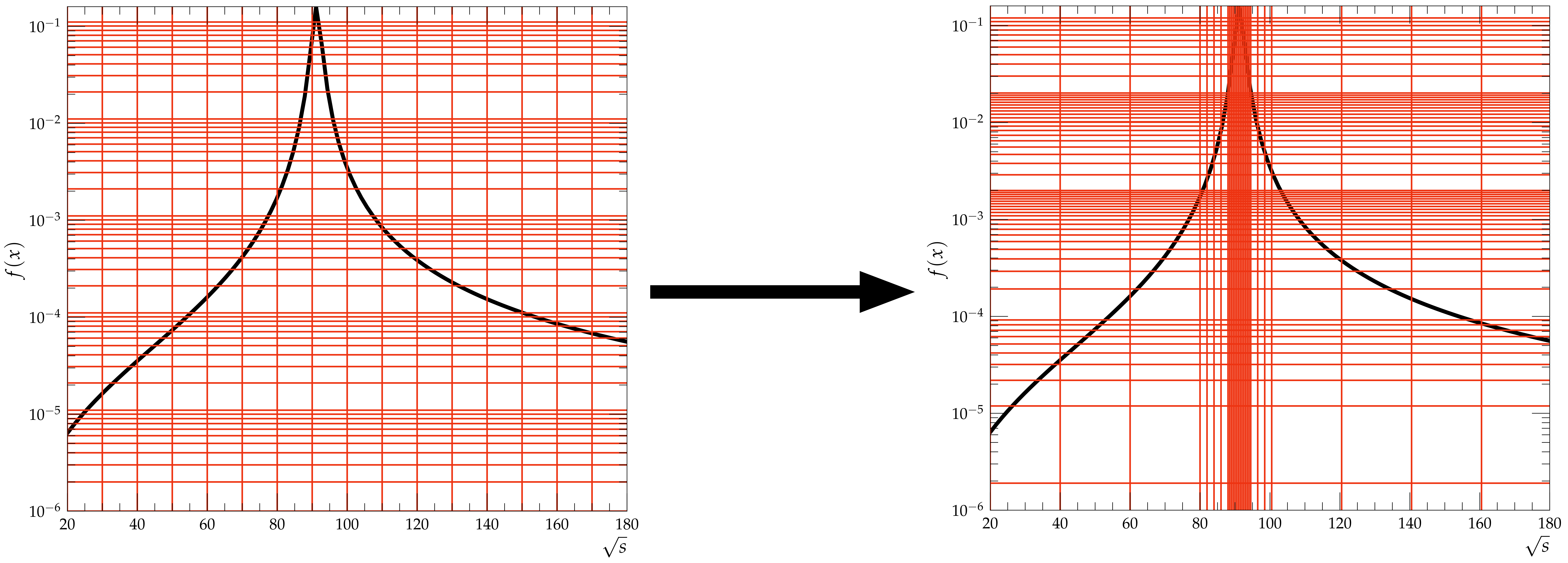
$$g(x) = \alpha_1 g_\gamma(x) + \alpha_2 g_Z(x) + \alpha_3 g_{ISR}(x)$$

Vegas Algorithm

G.P. Lepage, J.Comput.Phys.27:192,1978

- The Vegas algorithm is adaptive sampling algorithm developed by G.P Lepage
- The essential idea is to split the integration range into a number of smaller bins, where regions which narrow peaks get more thin bins and regions which are flat get wider ones
- Vegas can be used approximate the target directly, or it can be used to remap the input variables e.g uniform random numbers

Vegas Algorithm



Visual interpretation

Questions?