UCLouvain
Institut de recherche en mathématique et physique
Centre de Cosmologie, Physique des Particules et Phénoménologie

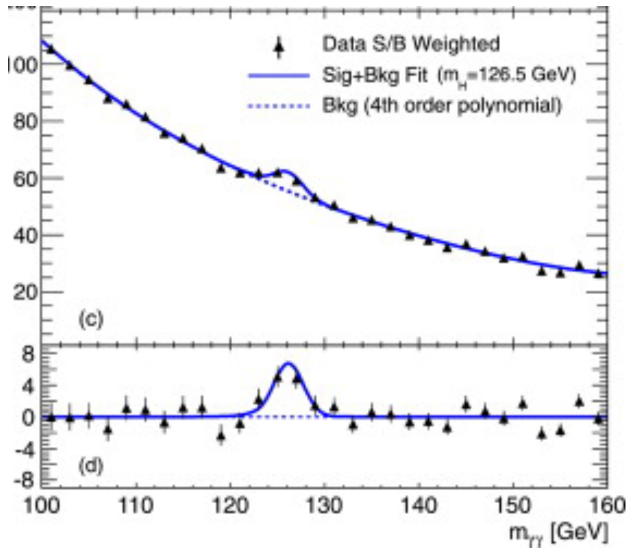
MCnet

MonteCarlo Simulation

Olivier Mattelaer

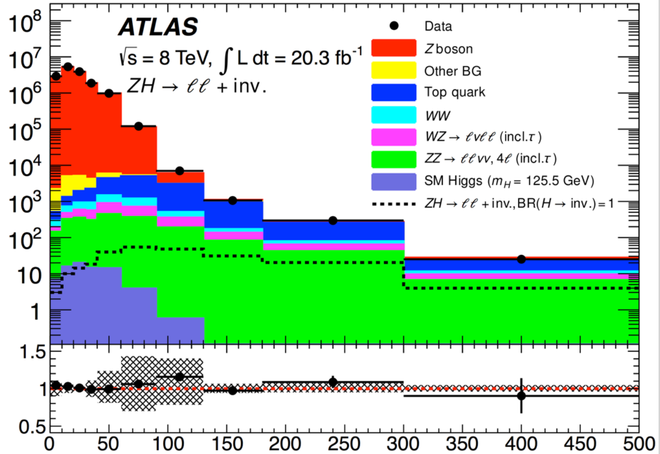
Kind of measurement

Peak



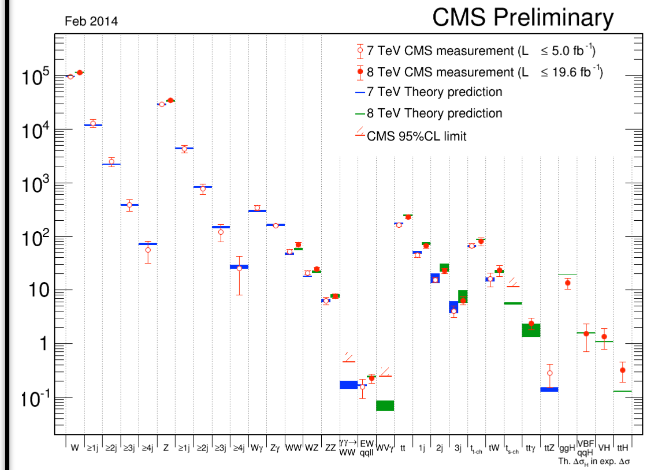
Background directly measured from **data**. Theory needed only for parameter extraction

Shape



Background **SHAPE** needed. Flexible MC for both signal and background validated and tuned to data

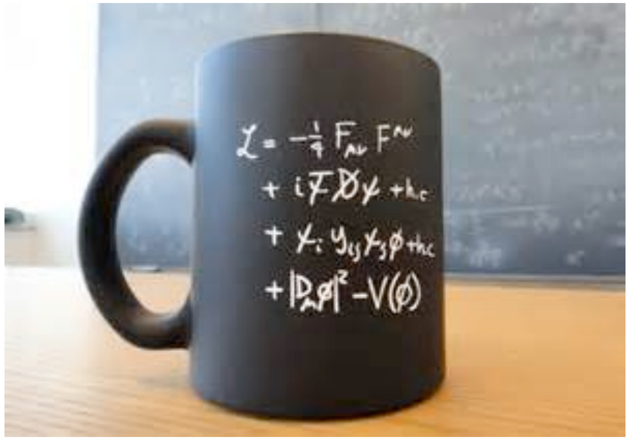
Rate



Relies on prediction for both **shape** and **normalization**. Complicated interplay of best simulations and data

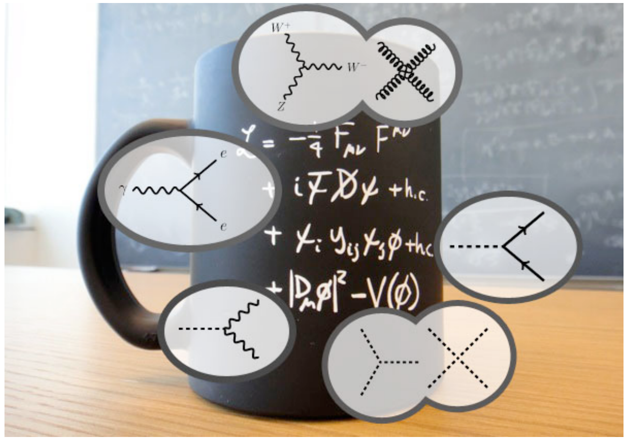
Theory side

Lagrangian



- This is Where the new idea are expressed

Feynman Rule

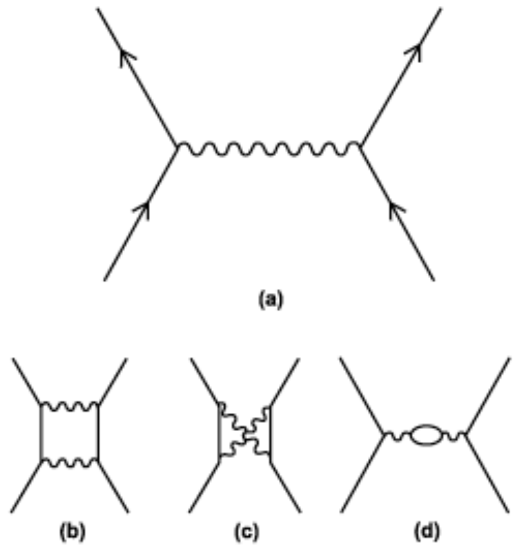


- Same information as the Lagrangian

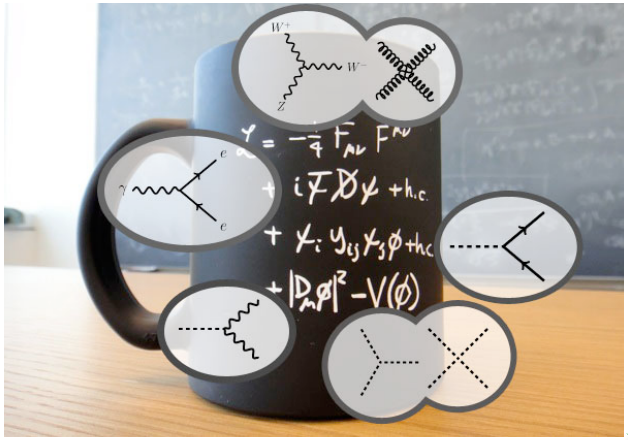
FeynRules

Cross-section

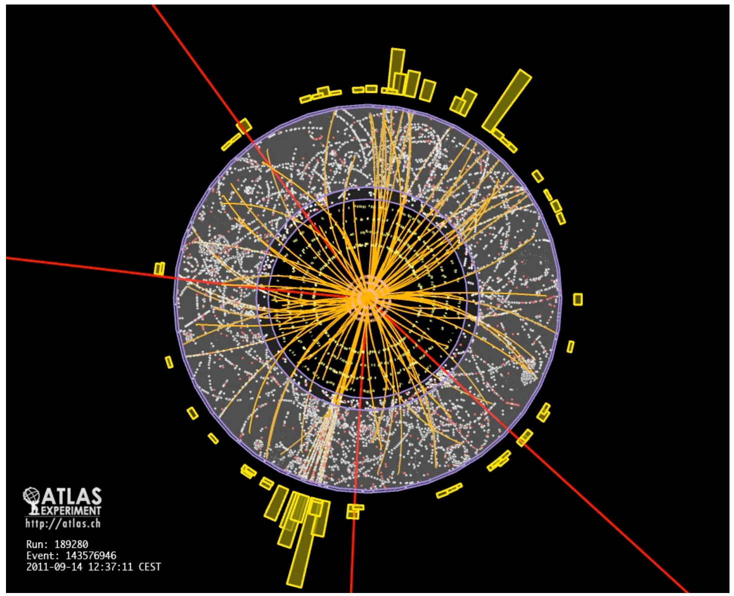
$$\frac{d\sigma}{d\ \cos\theta} = \left(\frac{d\sigma}{d\ \cos\theta}\right)_R \frac{(1+\cos\theta)/2}{\left[1+\frac{(1-\cos\theta)KE}{Mc^2}\right]}$$



- What is the precision?



Monte-Carlo Physics

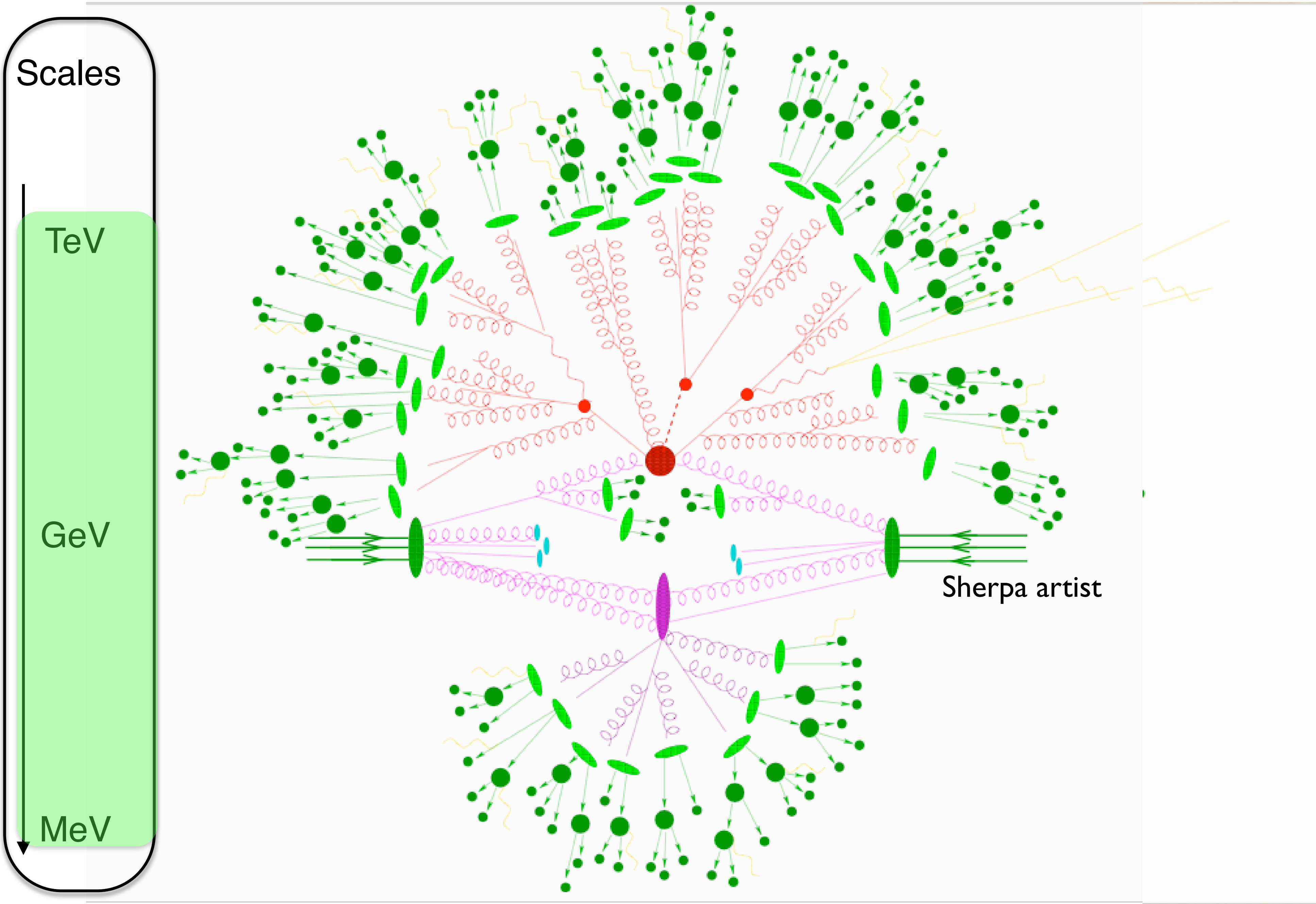


Our goal

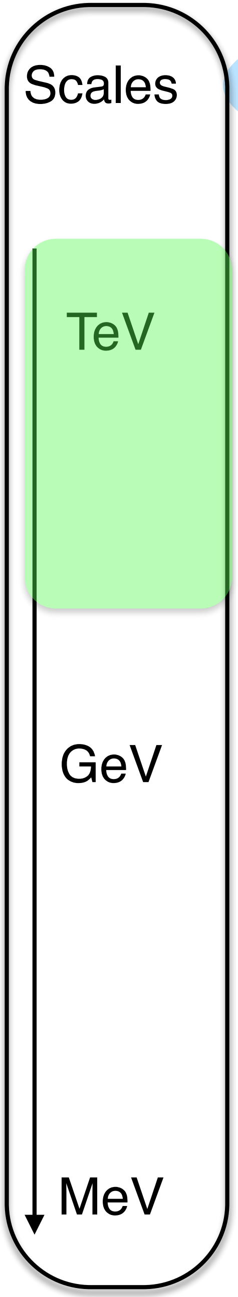
- Cross-section
- Differential cross-section
- Un-weighted events

Simulation of collider events

What are the MC for?



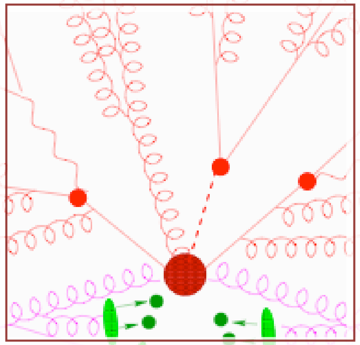
What are the MC for?



1. High-Q^2 Scattering

2. Parton Shower

☞ where BSM physics lies



☞ process dependent

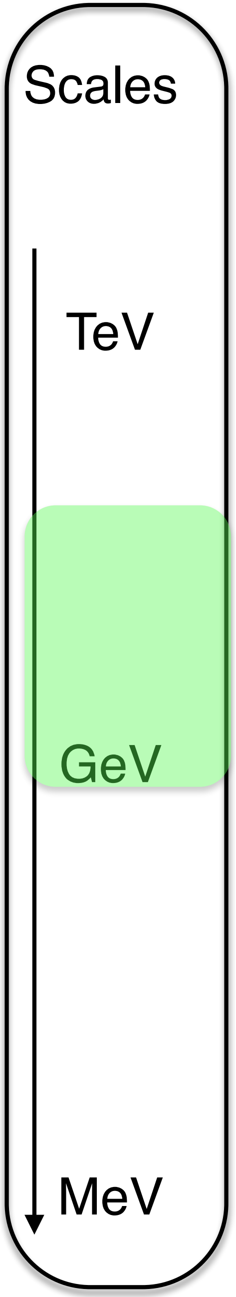
☞ first principles description

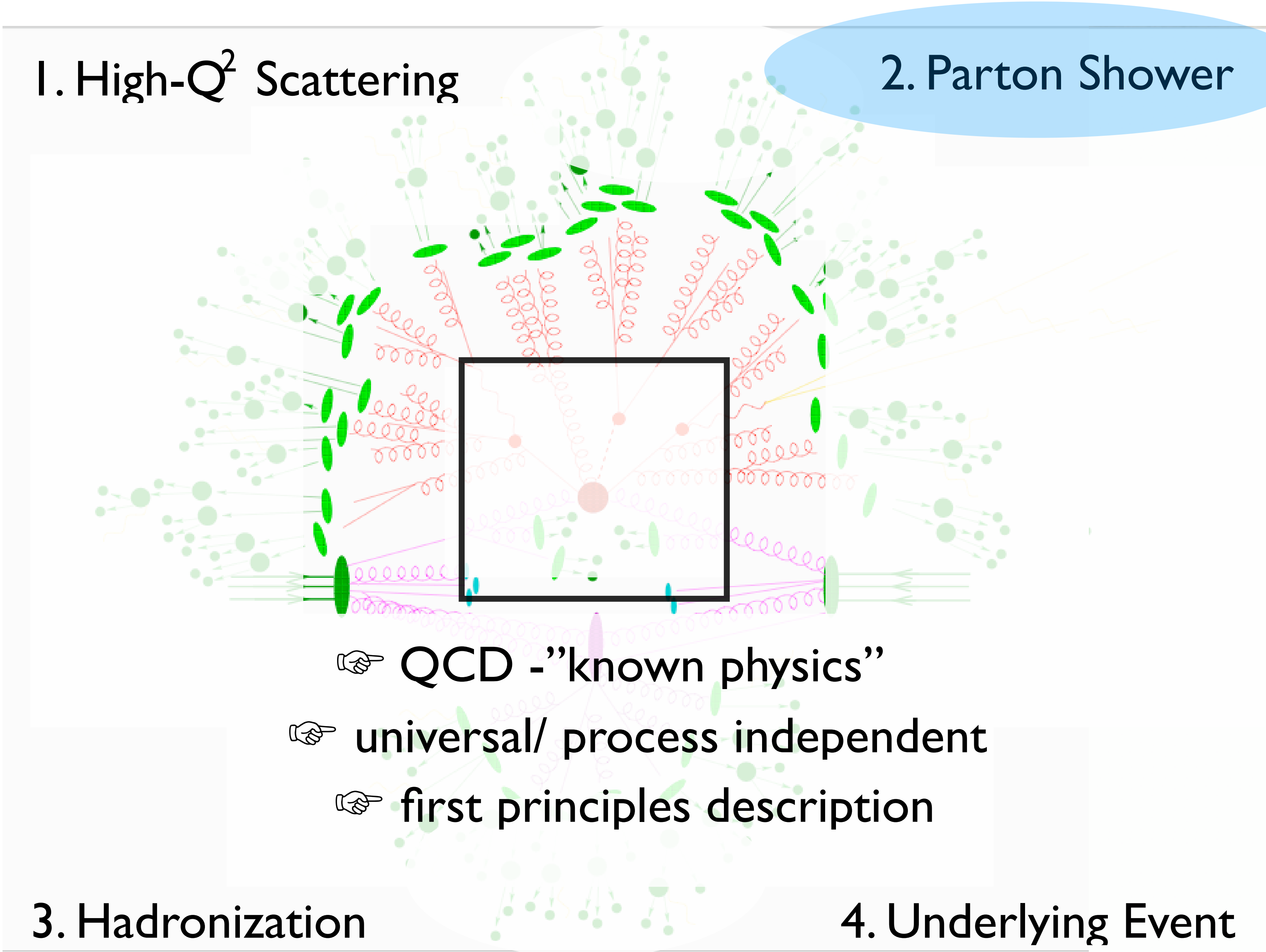
☞ it can be systematically improved

3. Hadronization

4. Underlying Event

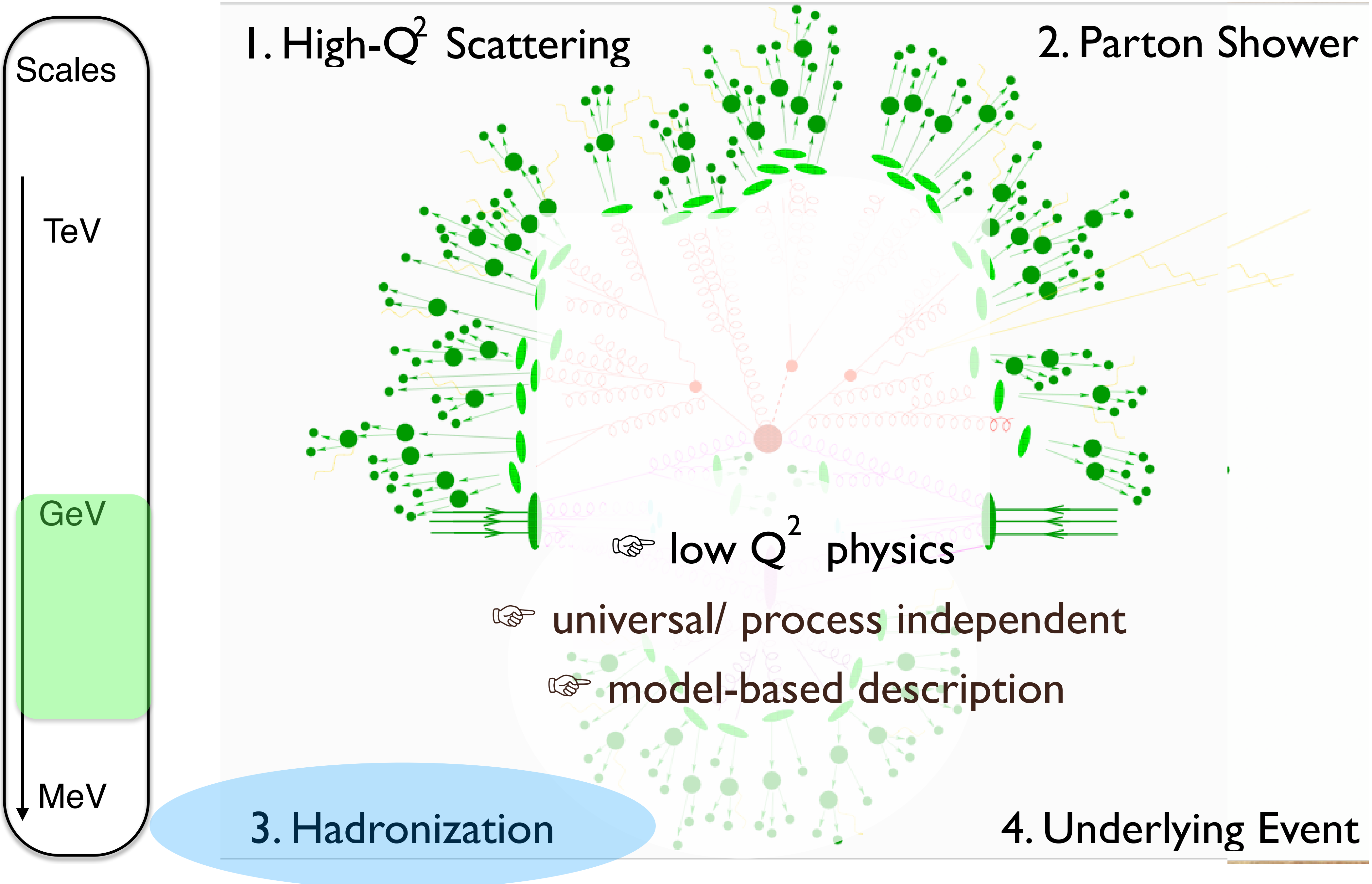
What are the MC for?





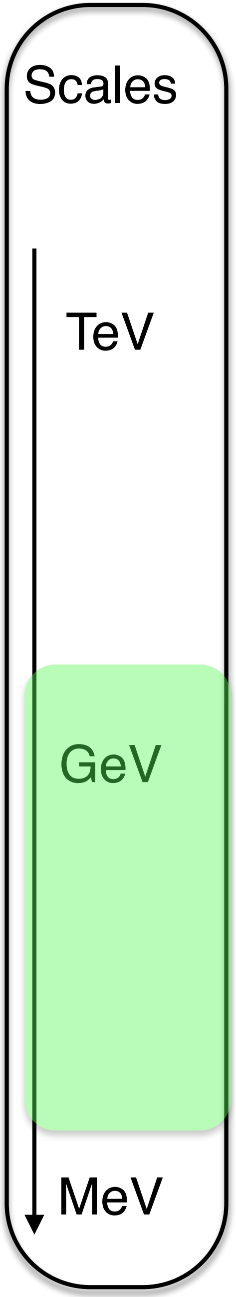
- ☞ QCD -"known physics"
- ☞ universal/ process independent
- ☞ first principles description

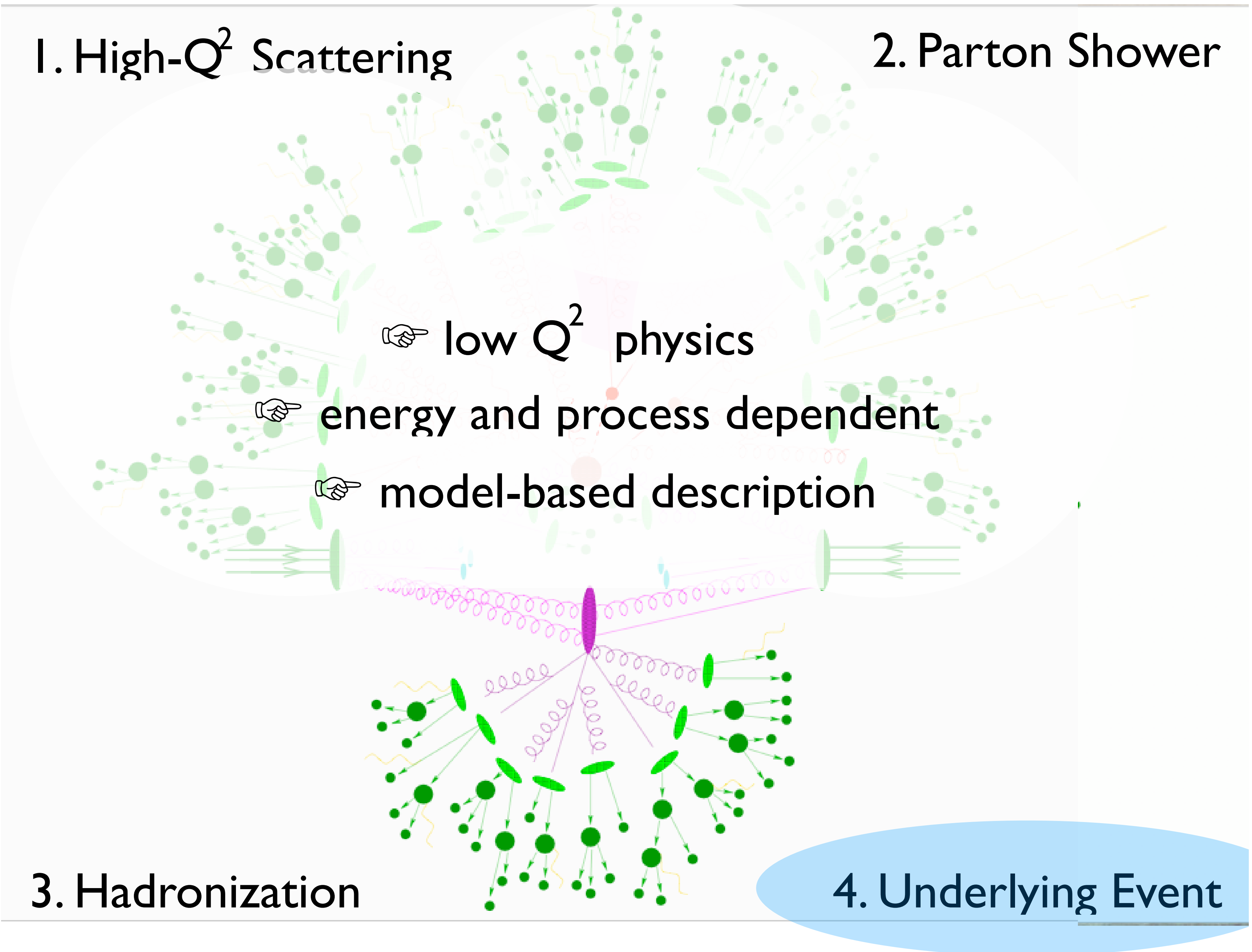
What are the MC for?



- ☞ low Q^2 physics
- ☞ universal/ process independent
- ☞ model-based description

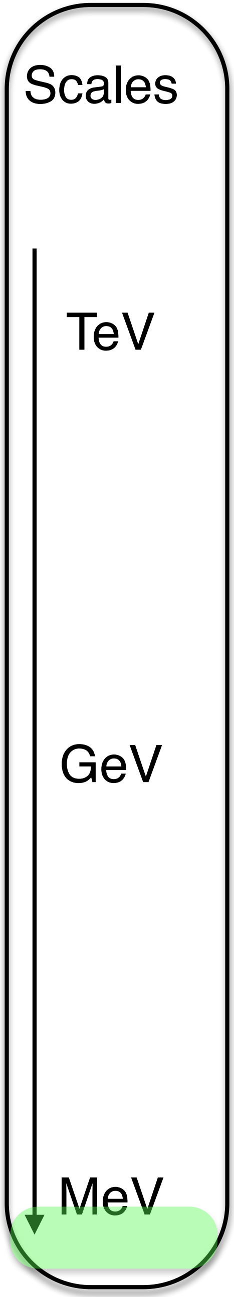
What are the MC for?

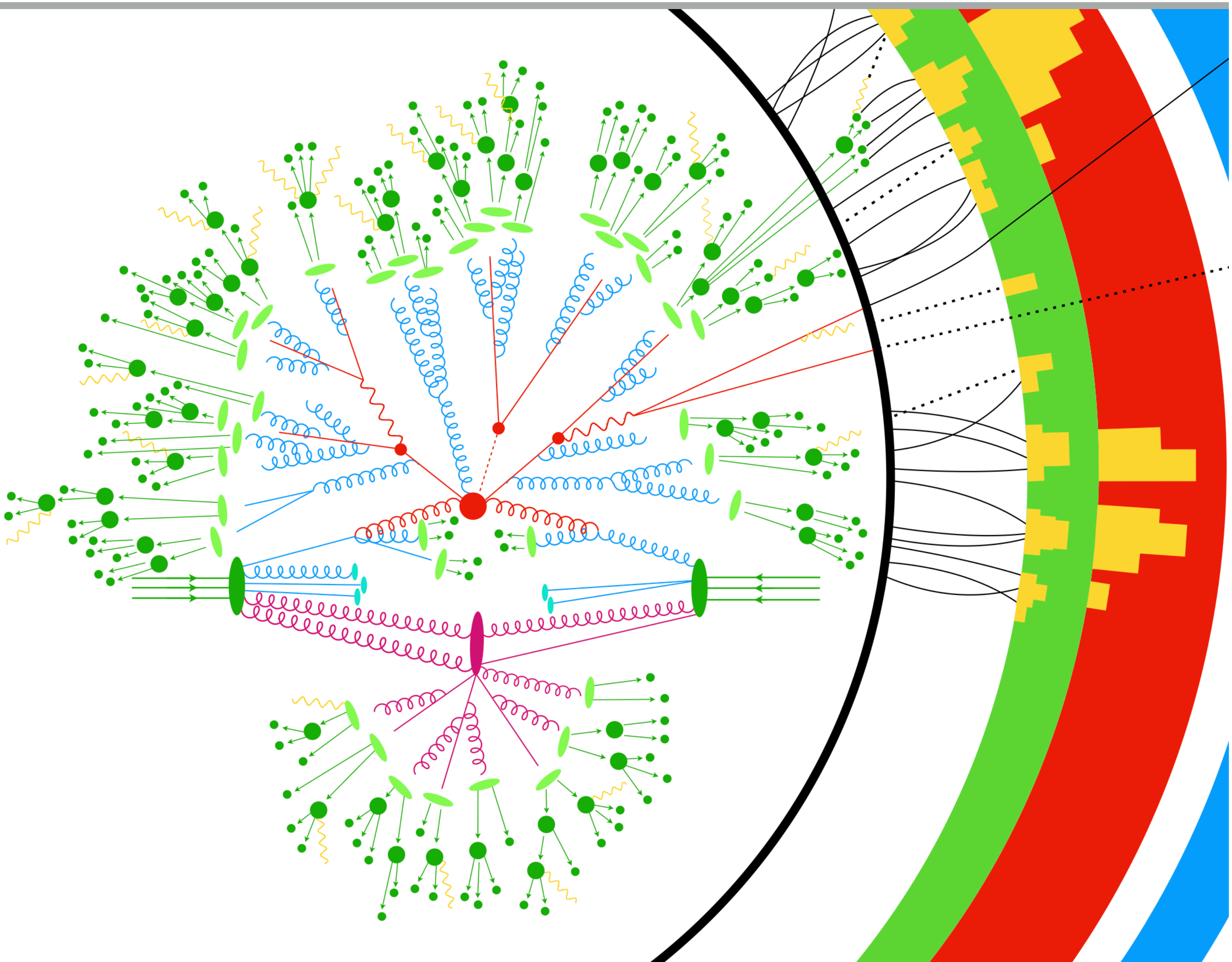




- ☞ low Q^2 physics
- ☞ energy and process dependent
- ☞ model-based description

What are the MC for?

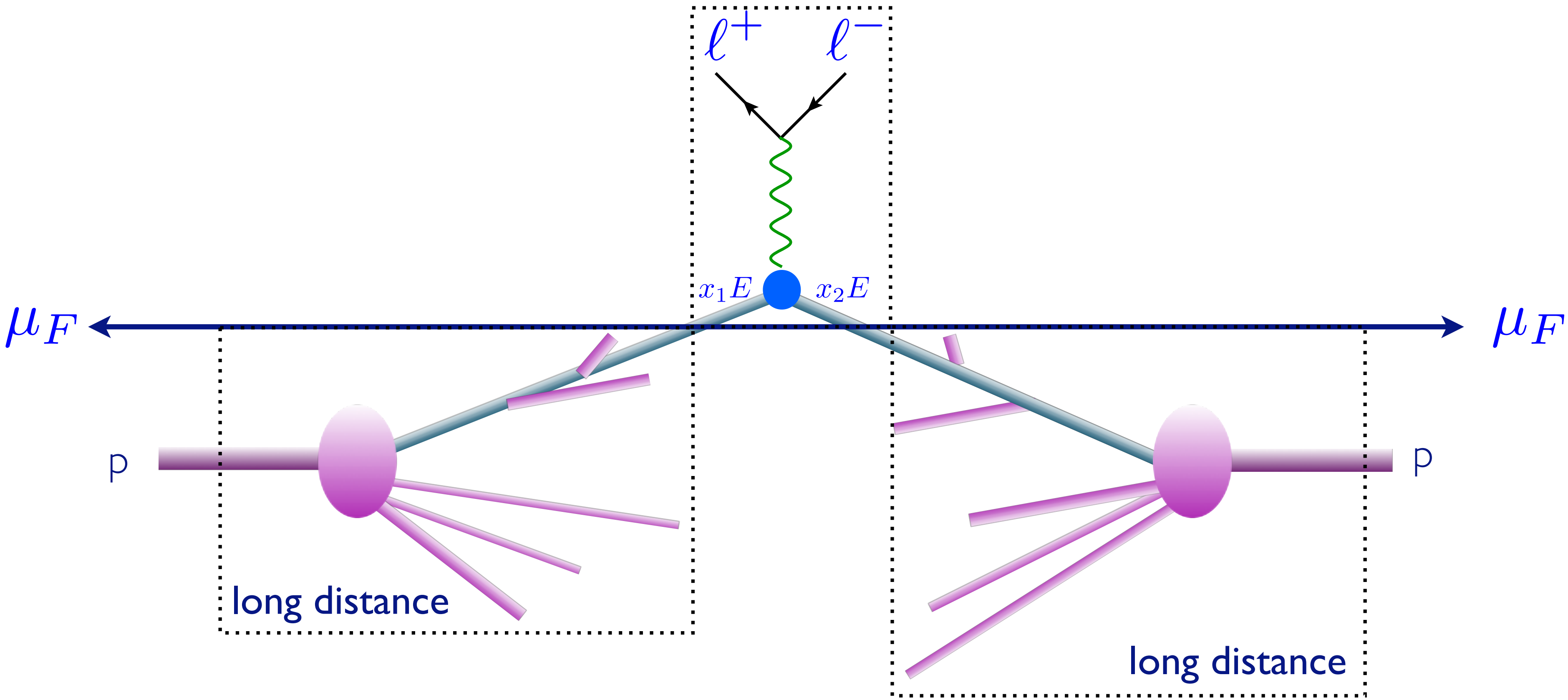




To Remember

- Multi-scale problem
  - ➡ New physics visible only at High scale
  - ➡ Problem split in different scale
    - ❑ Factorisation theorem

Master formula for the LHC



$\hat{\sigma}_{ab\to X}(\hat{s},\mu_F,\mu_R)$

Parton-level cross section

Perturbative expansion

$d\hat{\sigma}_{ab\to X}(\hat{s}, \mu_F, \mu_R)$ **Parton-level cross section**

- The parton-level cross section can be computed as a series in perturbation theory, using the coupling constant as an expansion parameter, schematically:

$$\hat{\sigma} = \sigma^{\text{Born}}\left(1 + \frac{\alpha_s}{2\pi}\sigma^{(1)} + \left(\frac{\alpha_s}{2\pi}\right)^2\sigma^{(2)} + \left(\frac{\alpha_s}{2\pi}\right)^3\sigma^{(3)} + \ldots\right)$$

LO predictions

NLO corrections

NNLO corrections

N3LO or NNNLO corrections

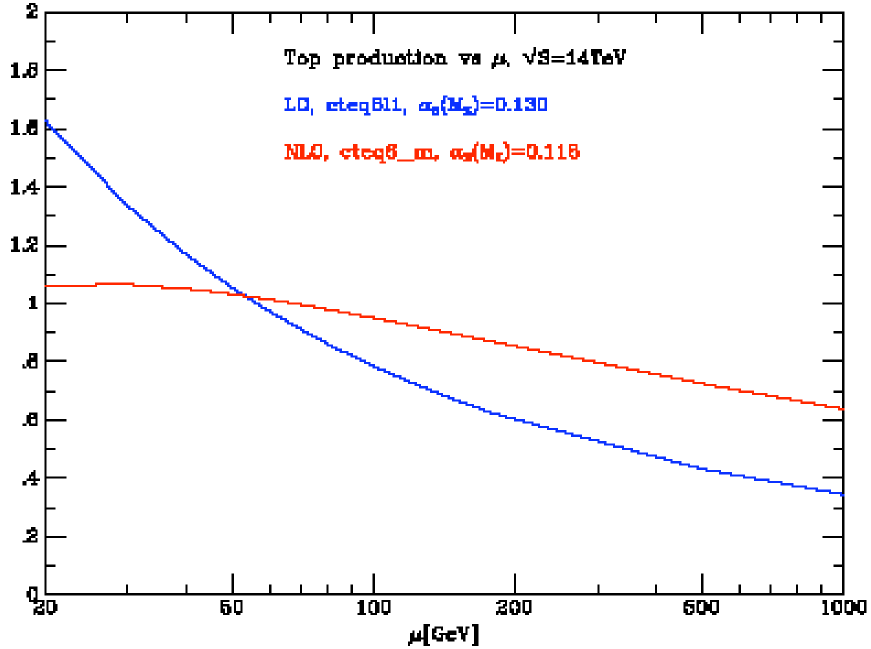
- Including higher corrections improves predictions and reduces theoretical uncertainties

Improved predictions

$$d\sigma = \sum_{a,b} \int dx_1 dx_2 \;\; f_a(x_1, \mu_F) f_b(x_2, \mu_F) \, d\hat{\sigma}_{ab \to X}(\hat{s}, \mu_F, \mu_R)$$

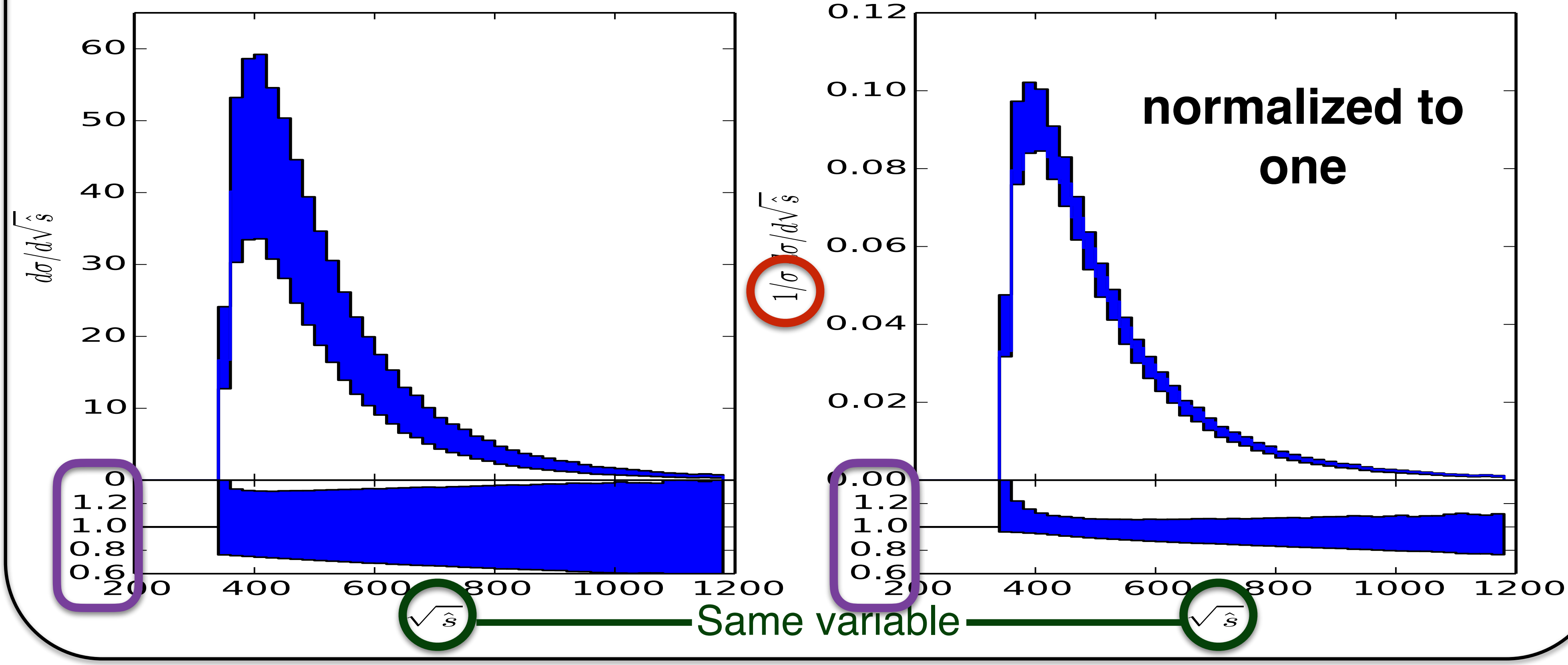
$$\hat{\sigma} = \sigma^{\text{Born}} \left(1 + \frac{\alpha_s}{2\pi} \sigma^{(1)} + \left(\frac{\alpha_s}{2\pi}\right)^2 \sigma^{(2)} + \left(\frac{\alpha_s}{2\pi}\right)^3 \sigma^{(3)} + \dots \right)$$

- Leading Order predictions can depend strongly on the renormalization and factorization scales
- Including higher order corrections reduces the dependence on these scales



LO

LO computation (top quark pair)



At LO:

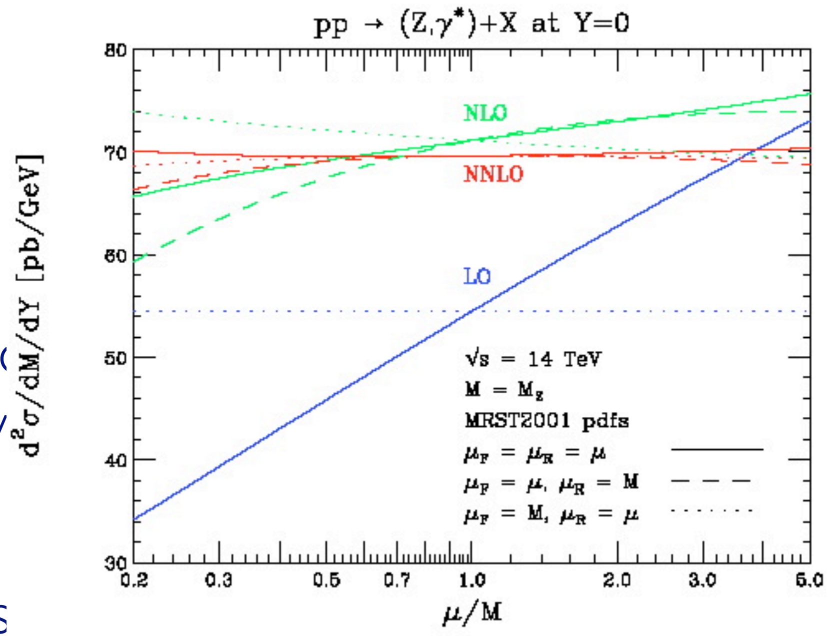
- Large scale uncertainty
- but mainly in the Normalisation
- LO is good for shape

Going NNLO...?

- NNLO is the current state-of-the-art. There are only a few results available: Higgs, Drell-Yan, ttbar
- Why do we need it?

➡ control of the uncertainties in a calculation

➡It is "mandatory" if NLO correcti are very large to check the behav of the perturbative series



➡It is needed for Standard Candles and very precise tests of perturbation theory, exploiting all the available information, e.g. for determining NNLO PDF sets

Let's focus on LO

Tevatron vs. the LHC

proto

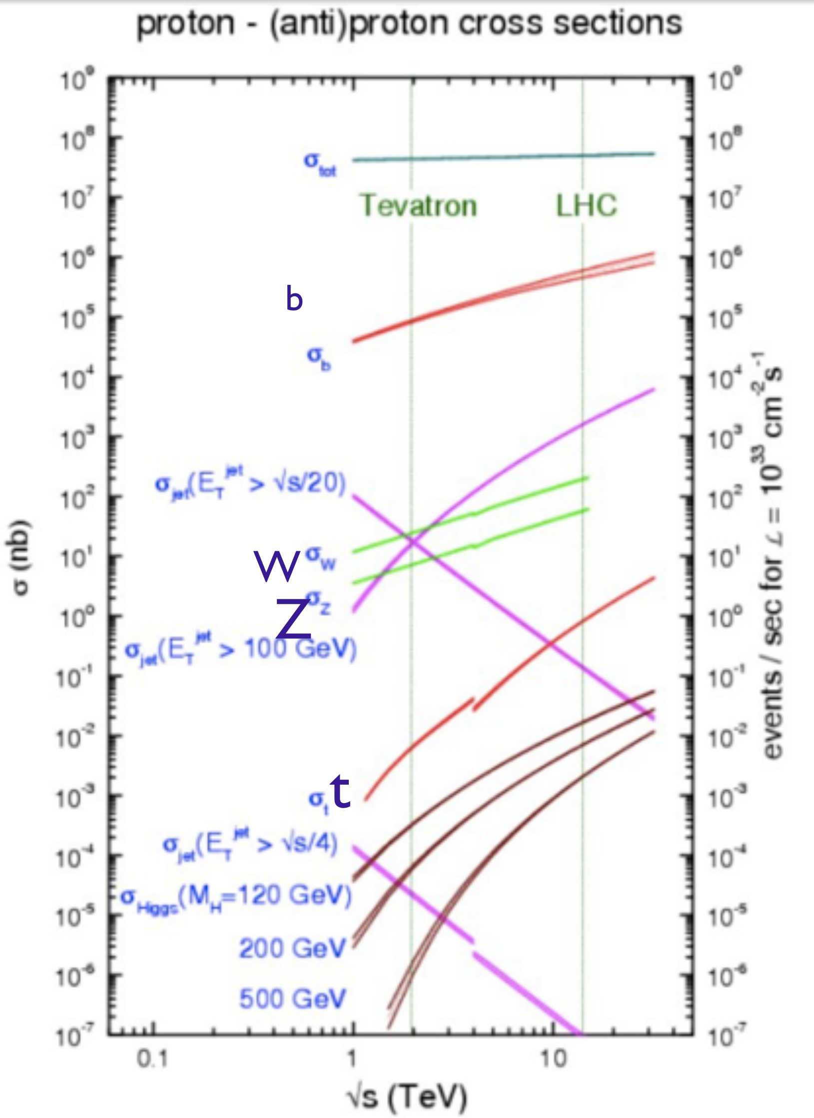
- ➡ Most important: q-q annihilation (85% of t t)
- LHC: 7-14 TeV proton-proton collider
  - ➡ Most important: g-g annihilation (90% of t t)

Hadron Colliders

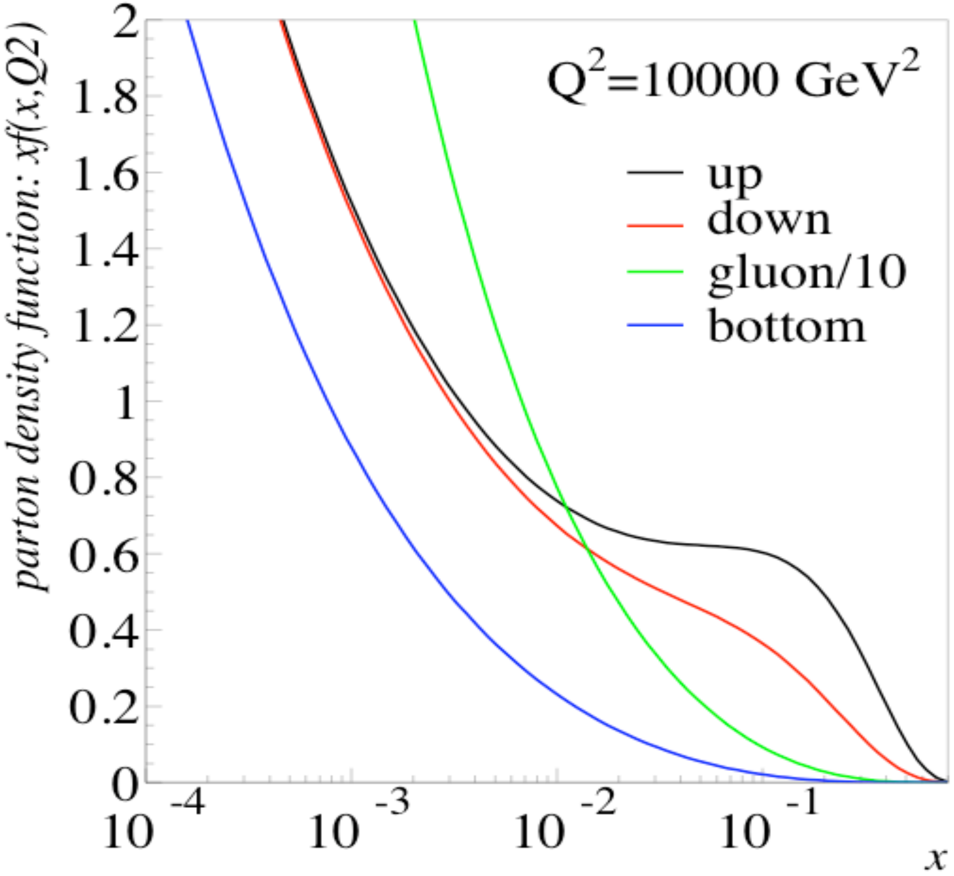
$$\sum_{a,b} \int dx_1 dx_2 d\Phi_{\text{FS}}\, f_a(x_1, \mu_F) f_b(x_2, \mu_F)\, \hat{\sigma}_{ab\to X}(\hat{s}, \mu_F, \mu_R)$$

Cross sec end on:

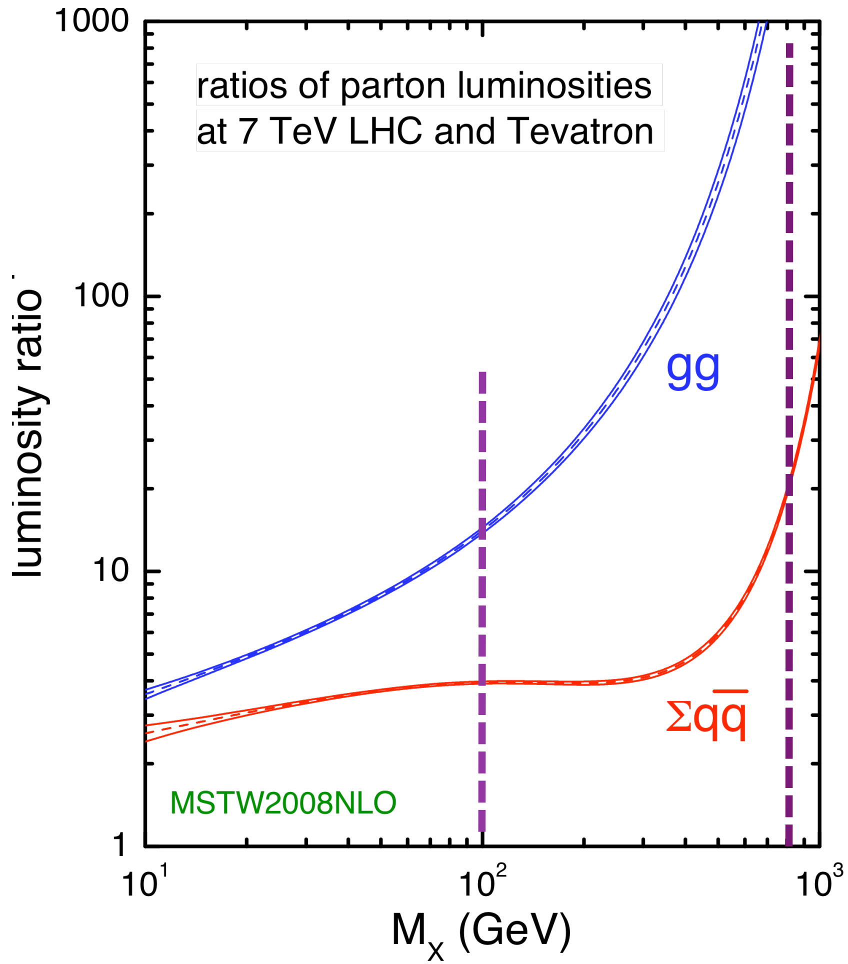
- Coupling
- Coupling (light qua rks, EW gaug
- Mass
- Single pr on



Parton densities



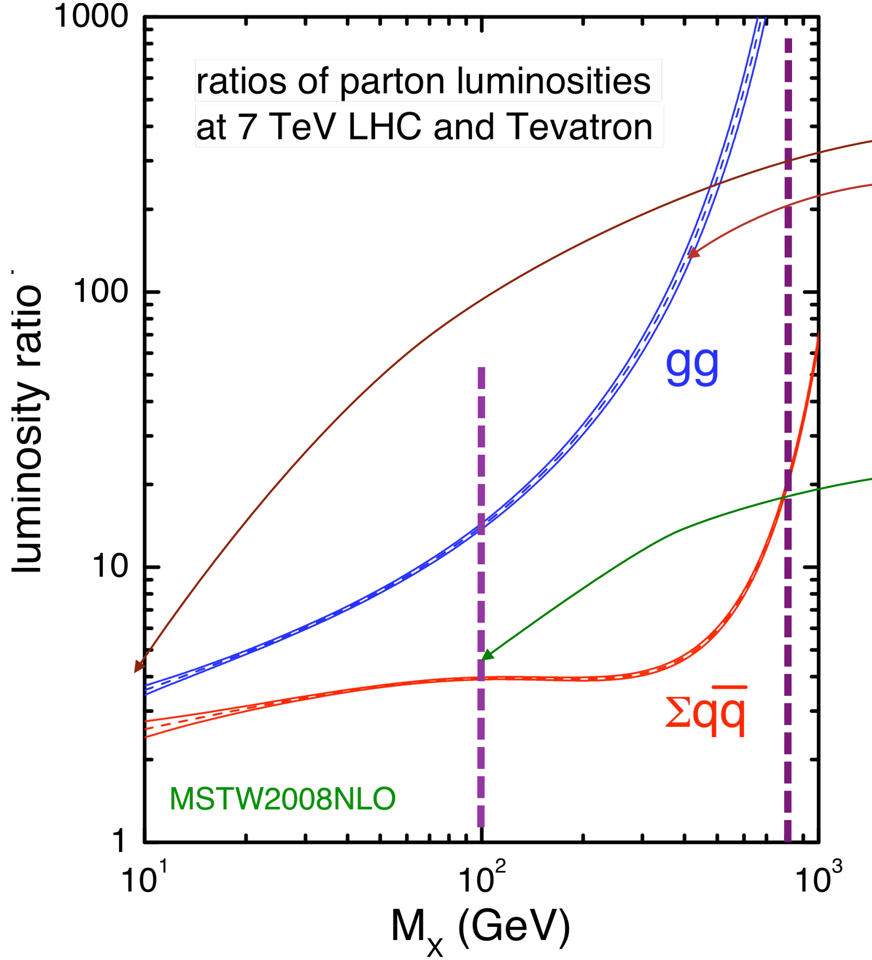
At small x (small ŝ), gluon domination.
At large x valence quarks

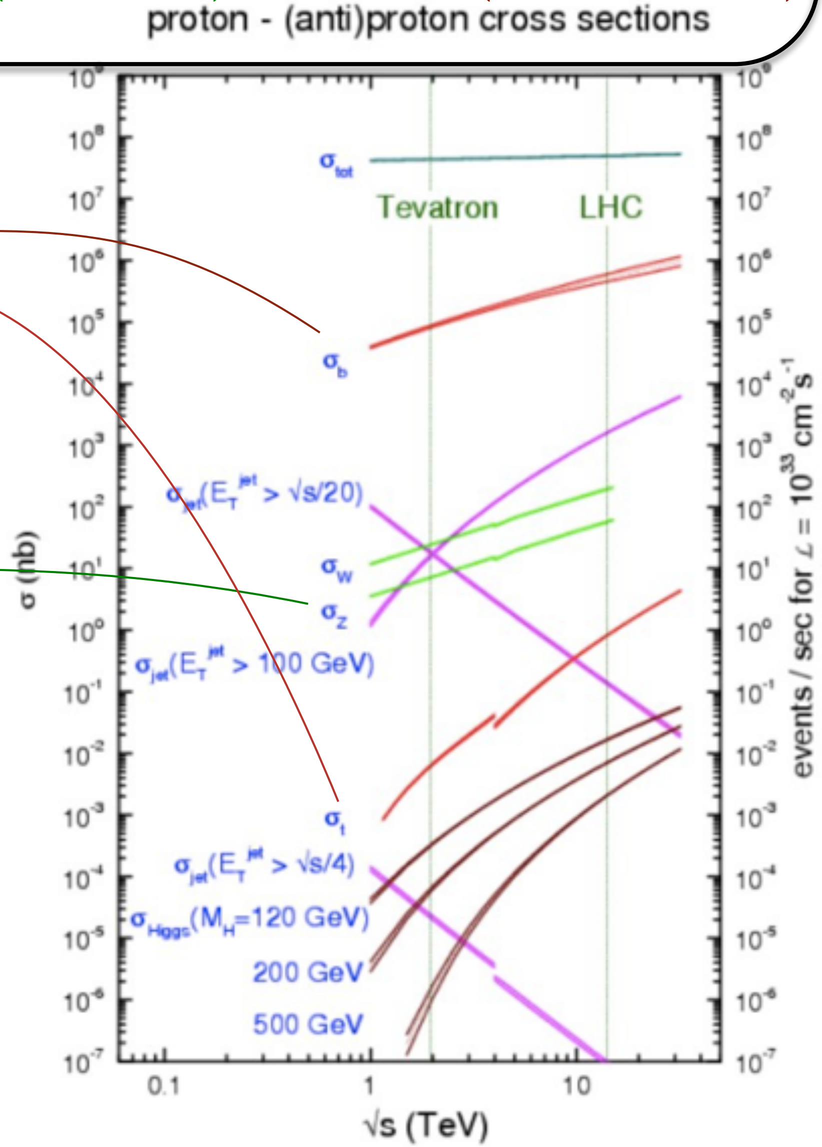


LHC formidable at large mass –
For low mass, Tevatron backgrounds smaller

Hadron colliders

$$\sum_{a,b} \int dx_1 dx_2 d\Phi_{\rm FS}\, f_a(x_1,\mu_F) f_b(x_2,\mu_F)\, \hat{\sigma}_{ab\to X}(\hat{s},\mu_F,\mu_R)$$





To Remember

$$\sum_{a,b} \int dx_1 dx_2 d\Phi_{\mathrm{FS}}\, f_a(x_1,\mu_F) f_b(x_2,\mu_F)\, \hat{\sigma}_{ab\to X}(\hat{s},\mu_F,\mu_R)$$

Phase-space integral

Parton density functions

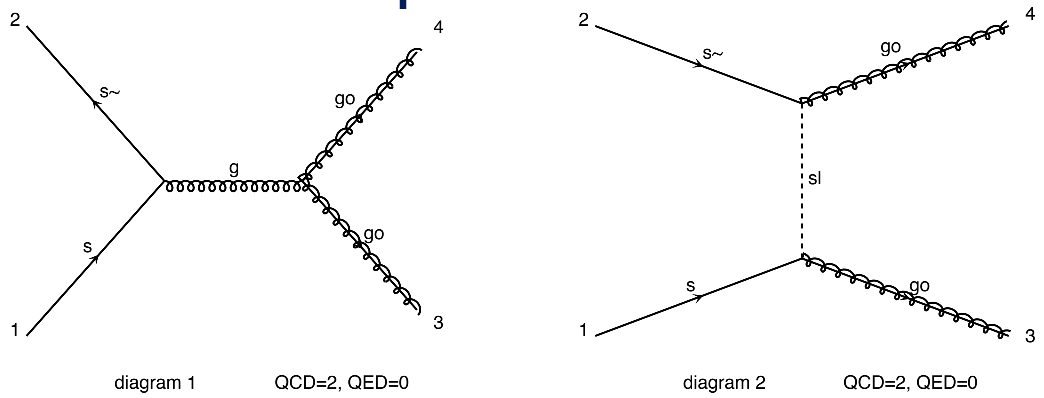
Parton-level cross section

- PDF: content of the proton
  - ➡ Define the physics/processes that will dominate on your accelerator
- LO: good for shape
- NLO/NNLO: Reduce scale uncertainty
- Computation are inclusive (+ any jet) due to renormalization/factorization scale

Matrix-Element

Calculate a given process (e.g. gluino pair)

- Determine the production mechanism



- Evaluate the matrix-element

$|\mathcal{M}|^2$ ➡Need Feynman Rules!

- Phase-Space Integration

$$\sigma = \frac{1}{2s}\int |\mathcal{M}|^2 d\Phi(n)$$

Easy enough

Hard

Tommorow

Very Hard (in general)

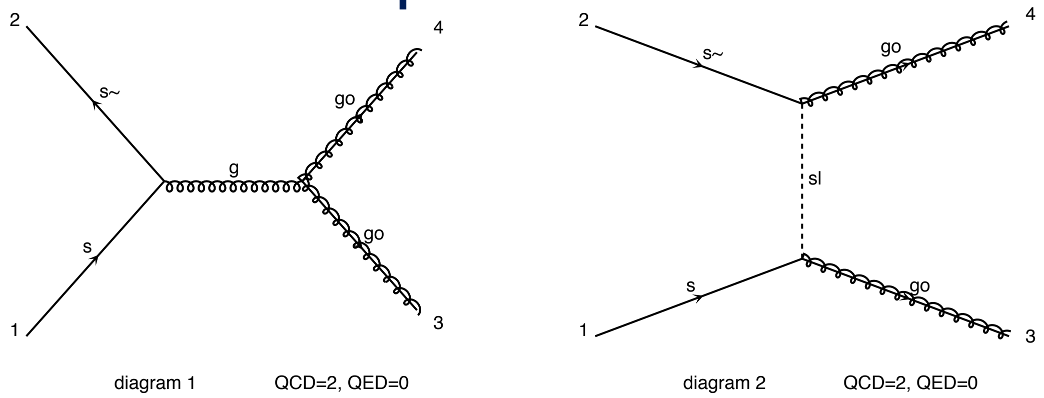
Now

Monte Carlo Integration

Matrix-Element

Calculate a given process (e.g. gluino pair)

- Determine the production mechanism



- Evaluate the matrix-element

$|\mathcal{M}|^2$ ➡Need Feynman Rules!

- Phase-Space Integration

$$\sigma = \frac{1}{2s}\int |\mathcal{M}|^2 d\Phi(n)$$

Easy enough

Hard

Very Hard (in general)

Now

Monte Carlo Integration

Calculations of cross section or decay widths involve integrations over high-dimension phase space of very peaked functions:

$$Dim[\Phi(n)] \sim 3n$$

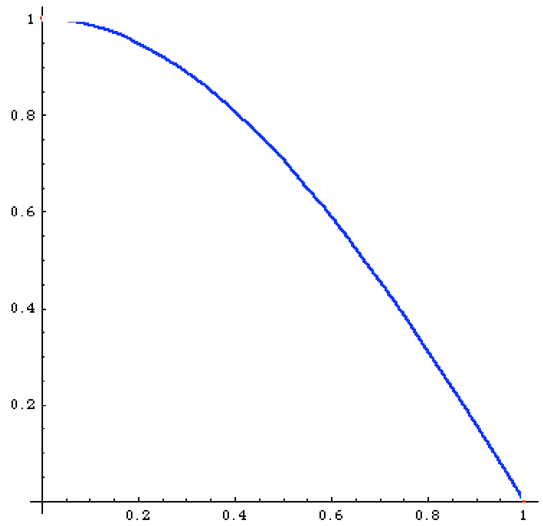
$$\sigma = \frac{1}{2s} \int |\mathcal{M}|^2 d\Phi(n)$$

General and flexible method is needed

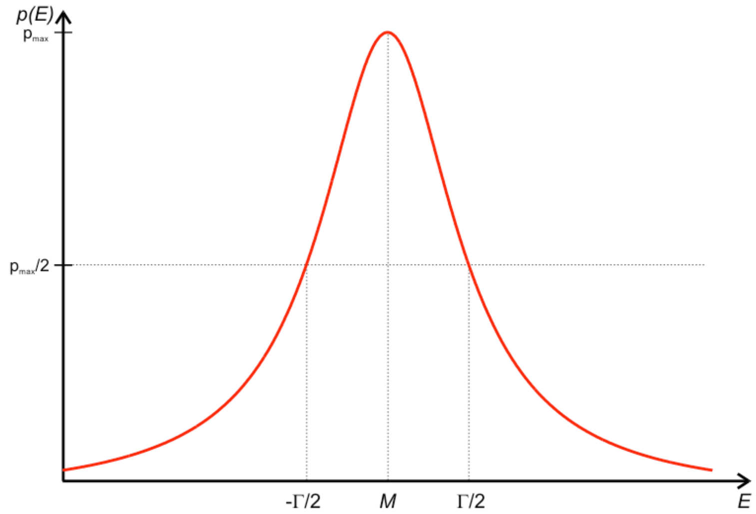
Not only integrating but also **generates events**

Integration

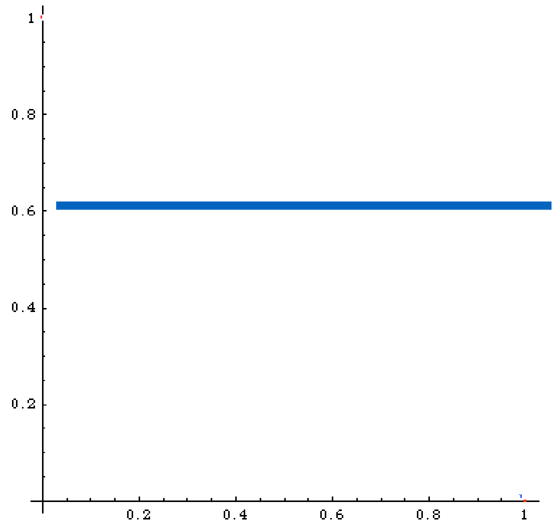
$$I = \int_0^1 dx \cos\frac{\pi}{2}x$$



$$\int \frac{dq^2}{(q^2 - M^2 + iM\Gamma)^2}$$



$$\int dx\, C$$

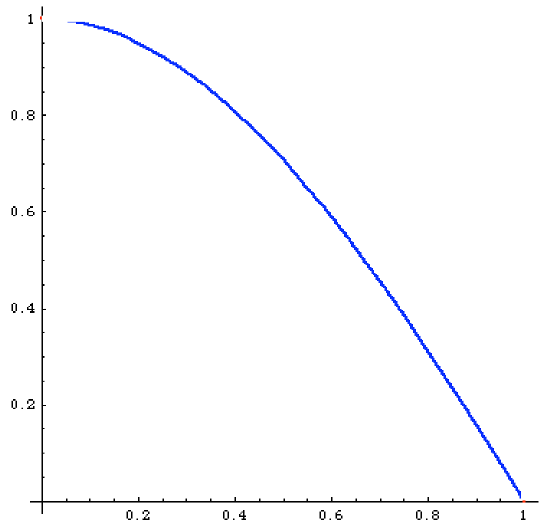


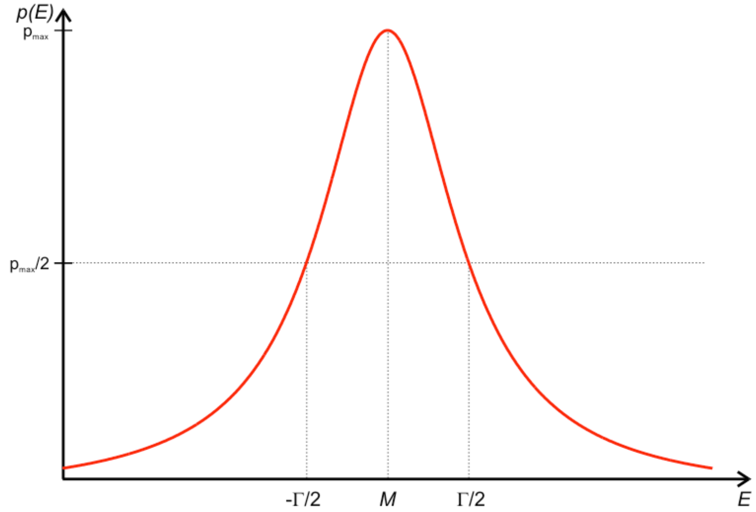
	simpson	MC
3	0,638	0,3
5	0,6367	0,8
20	0,63662	0,6
100	0,636619	0,65
1000	0,636619	0,636

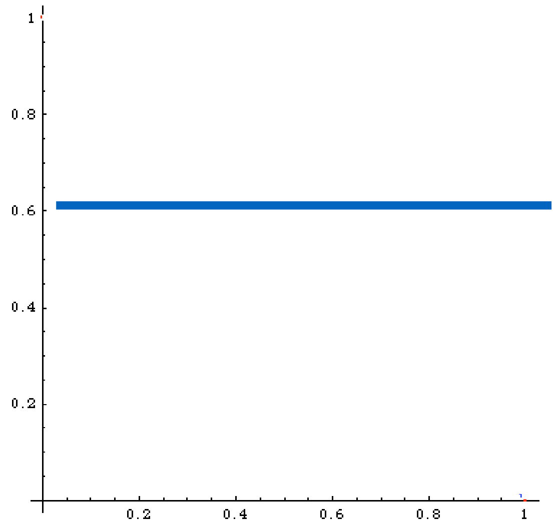
Method of evaluation	
• MonteCarlo	$1/\sqrt{N}$
• Trapezium	$1/N^2$
• Simpson	$1/N^4$

Integration

$$I = \int_0^1 dx \cos\frac{\pi}{2}x \qquad \int \frac{dq^2}{(q^2 - M^2 + iM\Gamma)^2} \qquad \int dx\, C$$







Method of evaluation

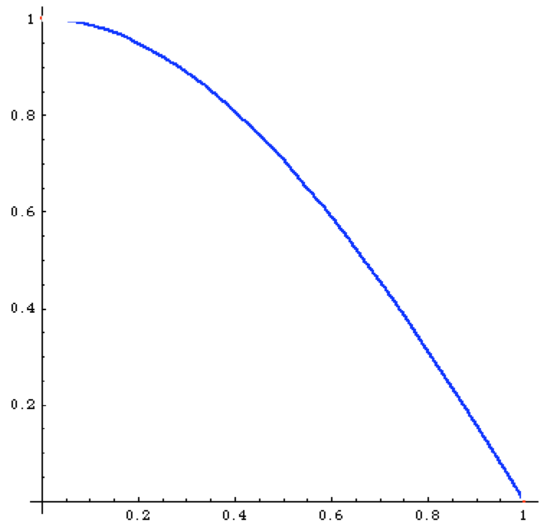
- MonteCarlo $1/\sqrt{N}$
- Trapezium $1/N^2$
- Simpson $1/N^4$

More Dimension →

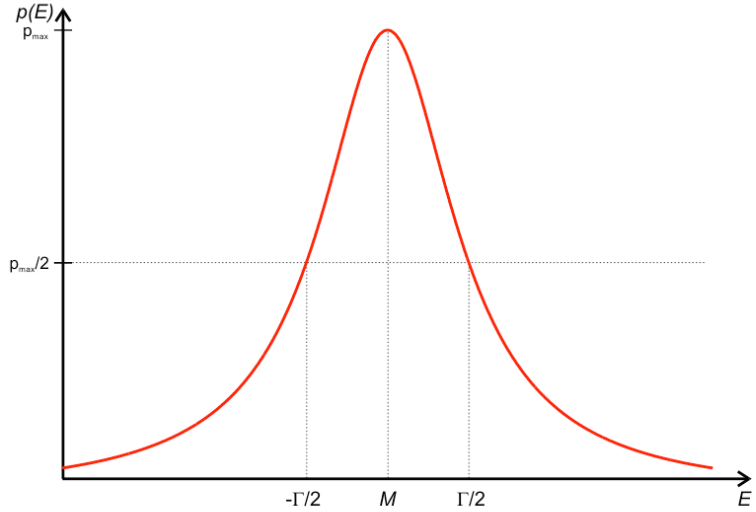
- $1/\sqrt{N}$
- $1/N^{2/d}$
- $1/N^{4/d}$

Integration

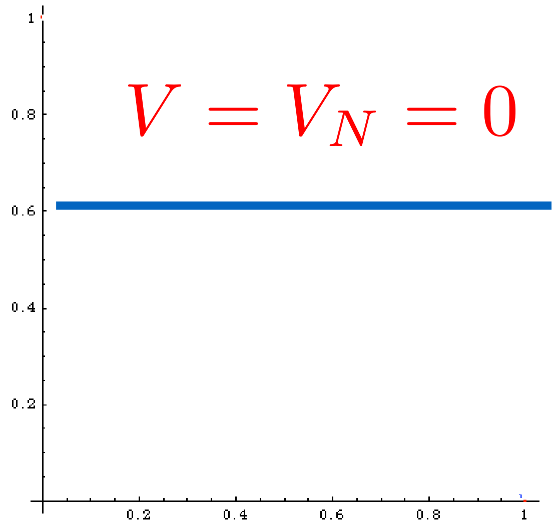
$$I = \int_0^1 dx \cos\frac{\pi}{2}x$$



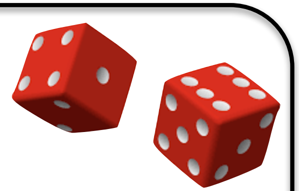
$$\int \frac{dq^2}{(q^2 - M^2 + iM\Gamma)^2}$$



$$\int dx\, C$$



$$I = \int_{x_1}^{x_2} f(x)dx \Rightarrow I_N = (x_2 - x_1)\frac{1}{N}\sum_{i=1}^{N} f(x)$$

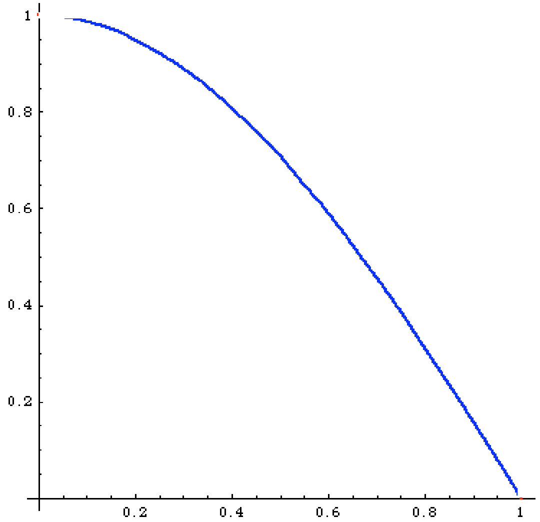


$$V = (x_2 - x_1)\int_{x_1}^{x_2} [f(x)]^2 dx - I^2 \Rightarrow V_N = (x_2 - x_1)^2 \frac{1}{N}\sum_{i=1}^{N}[f(x)]^2 - I_N^2$$

$$I = I_N \pm \sqrt{V_N/N}$$

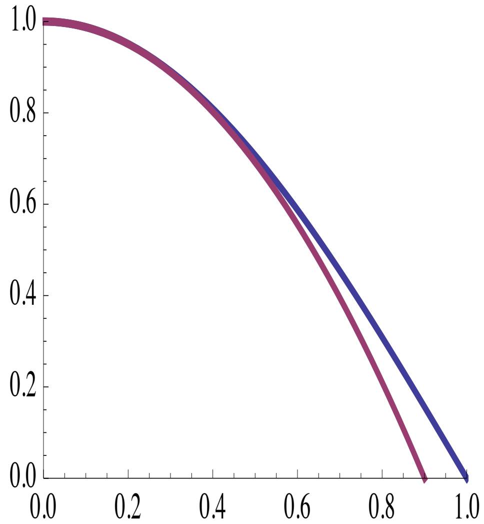
Can be minimized!

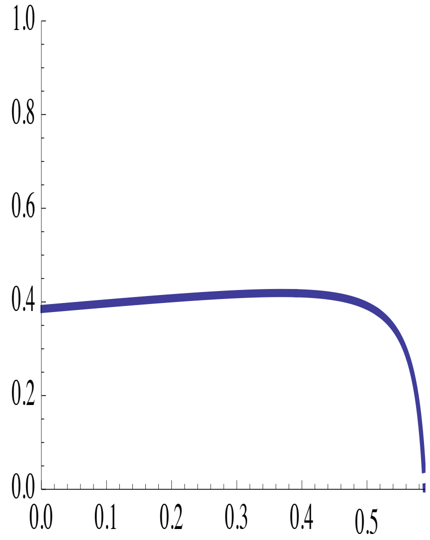
Importance Sampling



$$I = \int_0^1 dx \cos \frac{\pi}{2} x$$

$$I_N = 0.637 \pm 0.307/\sqrt{N}$$





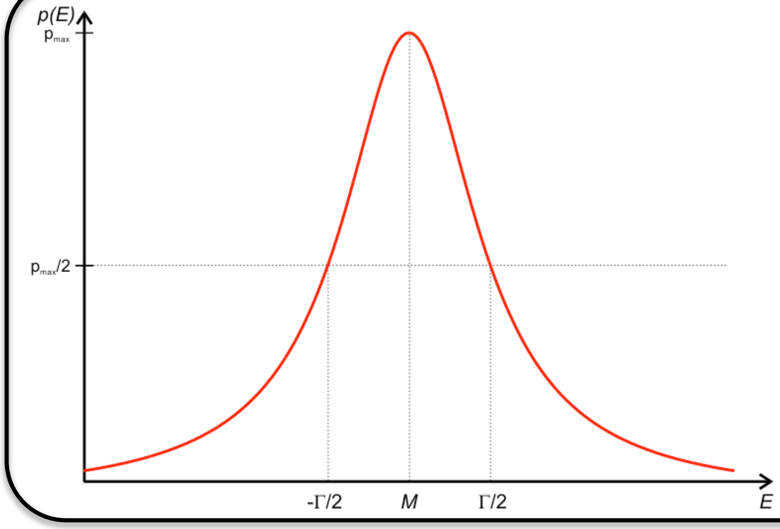
$$I = \int_0^1 dx (1 - cx^2) \frac{\cos\left(\frac{\pi}{2} x\right)}{(1 - cx^2)} = \int_{\xi_1}^{\xi_2} d\xi \frac{\cos \frac{\pi}{2} x[\xi]}{1 - x[\xi]^2 c}$$

$\simeq 1$

$$I_N = 0.637 \pm 0.031/\sqrt{N}$$

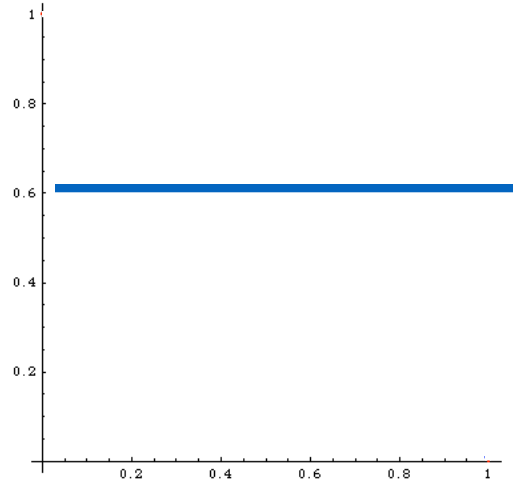
The Phase-Space parametrization is important to have an efficient computation!

Importance Sampling

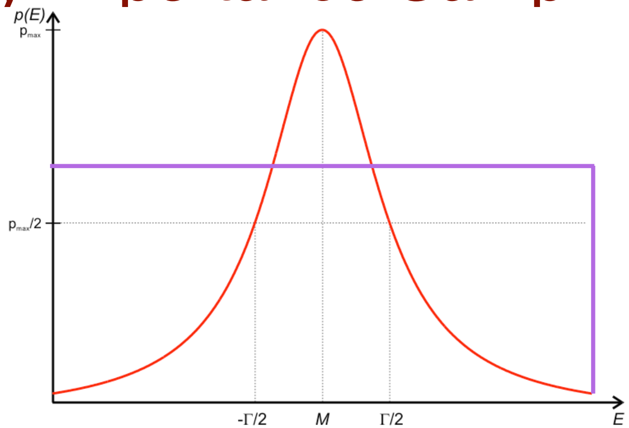


$$\int \frac{dq^2}{(q^2 - M^2 + iM\Gamma)^2}$$

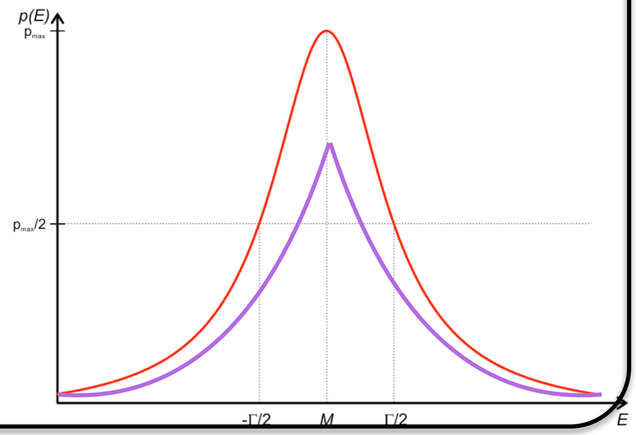
$$\xi = arctan\left(\frac{q^2 - M^2}{\Gamma M}\right)$$



Why Importance Sampling?



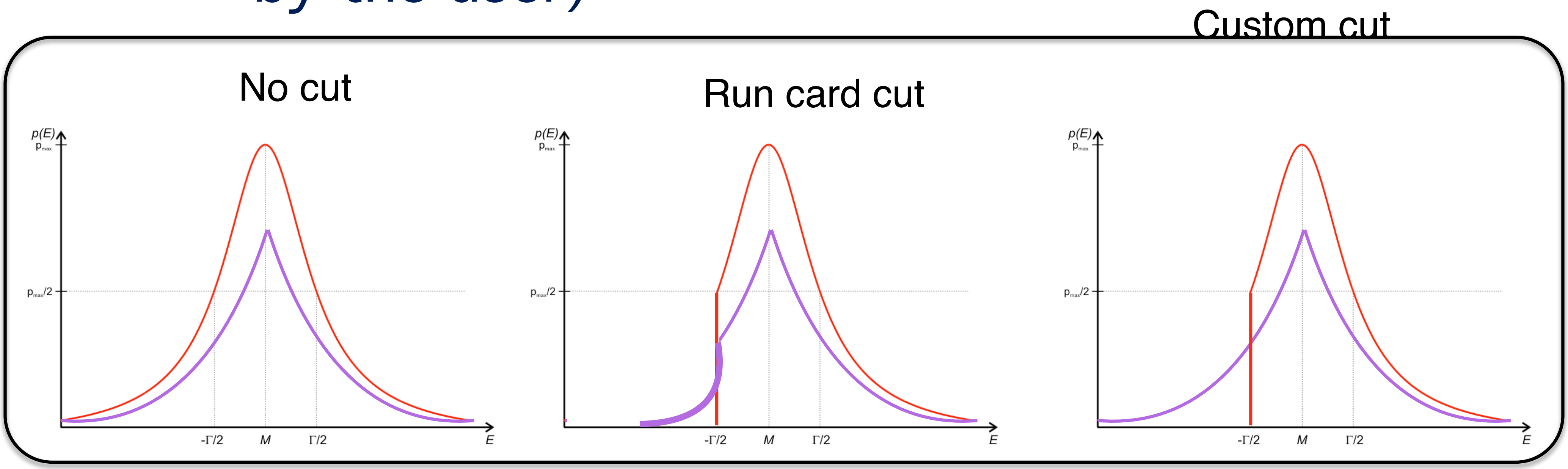
Probability of using that point p(x)



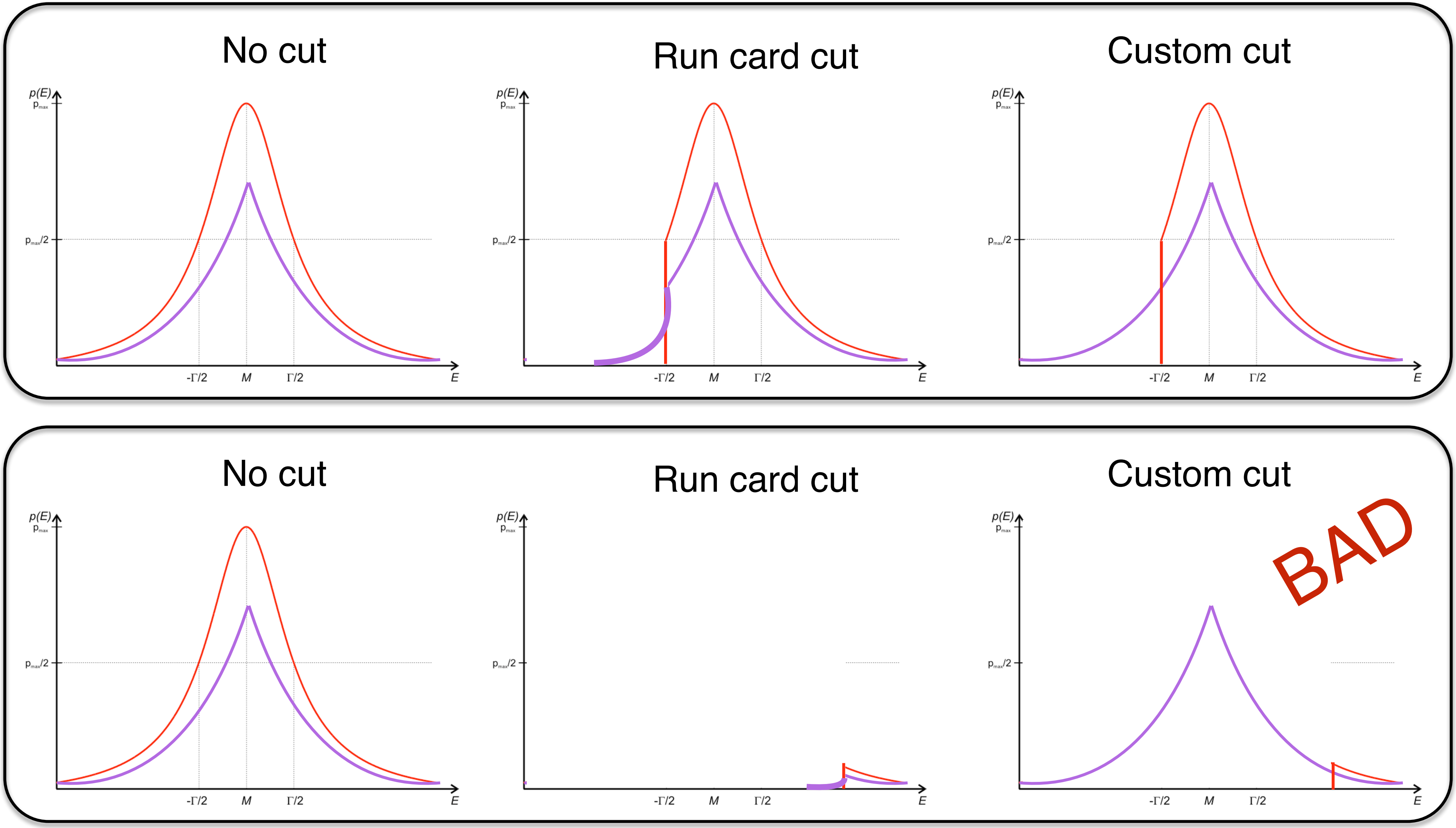
The change of variable ensure that the evaluation of the function is done where the function is the largest!

Cut Impact

- Events are generated according to our best knowledge of the function (with importance sampling
  - ➡ Adding a cut needs to modified the phase-space integrator
  - ➡ Not possible for custom cut (hardcoded by the user)



Cut Impact



Might miss the contribution and think it is just zero.

Importance Sampling

Key Point

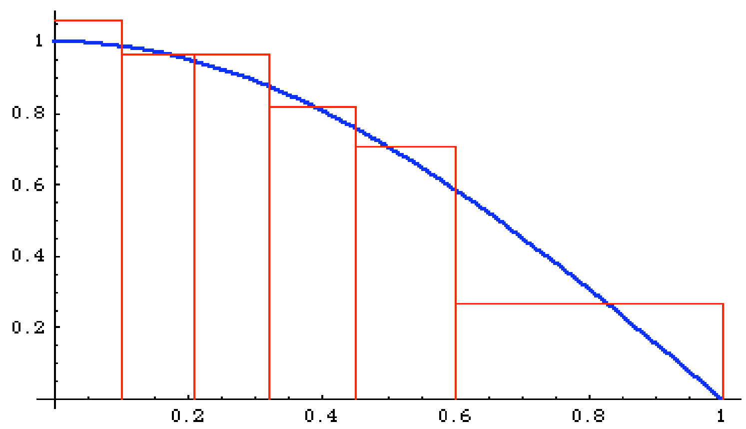
- Generate the random point in a distribution which is close to the function to integrate.
- This is a change of variable, such that the function is flatter in this new variable.
- Needs to know an approximate function.

Adaptative Monte-Carlo

- Create an approximation of the function on the flight!

Adaptative Monte-Carlo

- Create an approximation of the function on the flight!



Algorithm

1. Creates bin such that each of them have the same contribution.

   ➡Many bins where the function is large

2. Use the approximate for the importance sampling method.

VEGAS

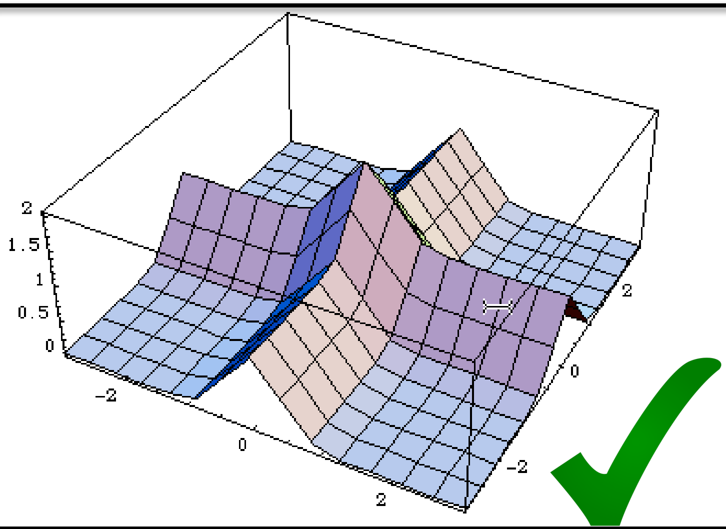
More than one Dimension

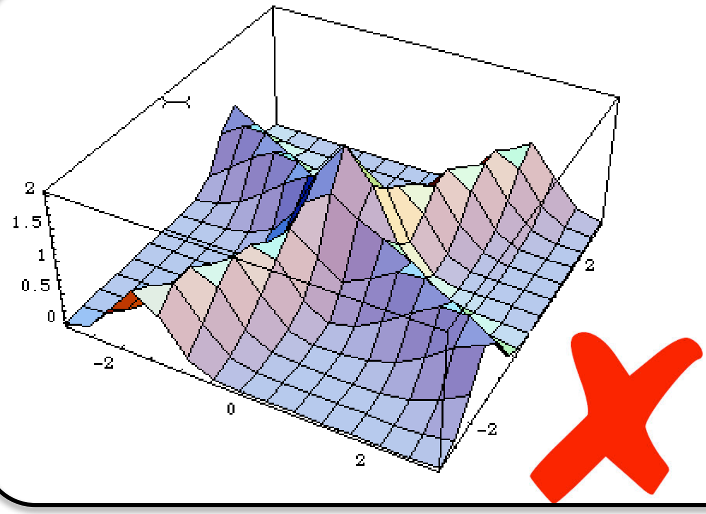
- VEGAS works only with 1(few) dimension
  - ➡memory problem

Solution

- Use projection on the axis

$$\vec{p}(x) = p(x) \cdot p(y) \cdot p(z) \ldots$$



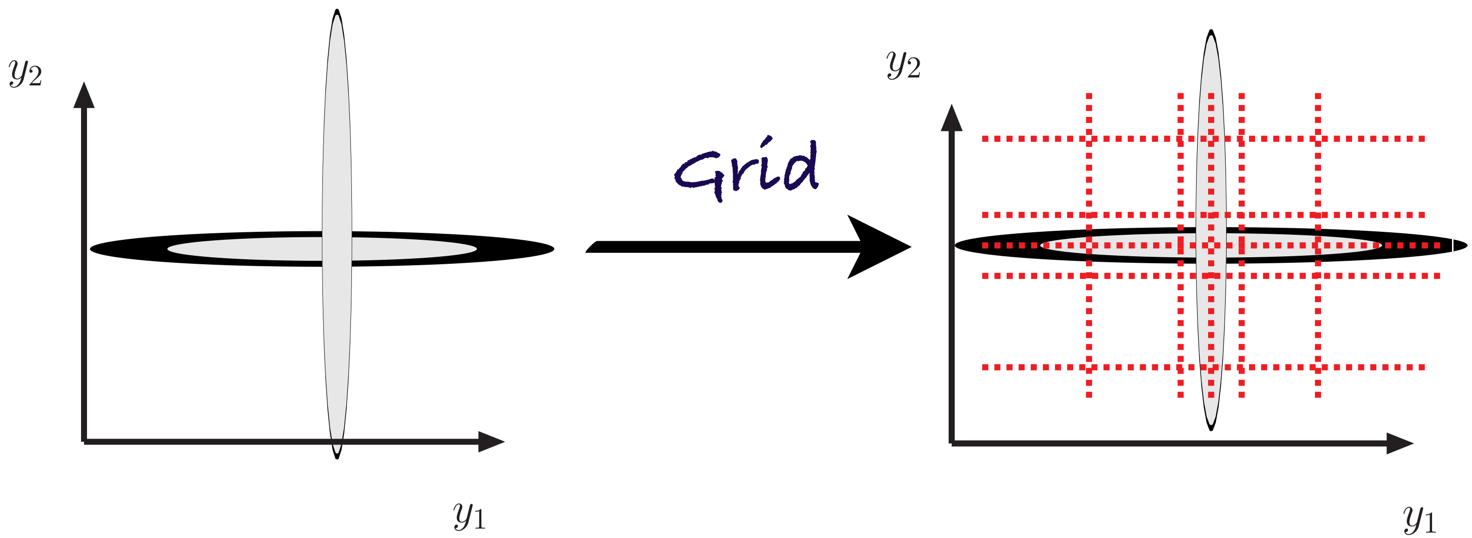


- We need to ensure the factorization !

➡Additional change of variable

Monte-Carlo Integration

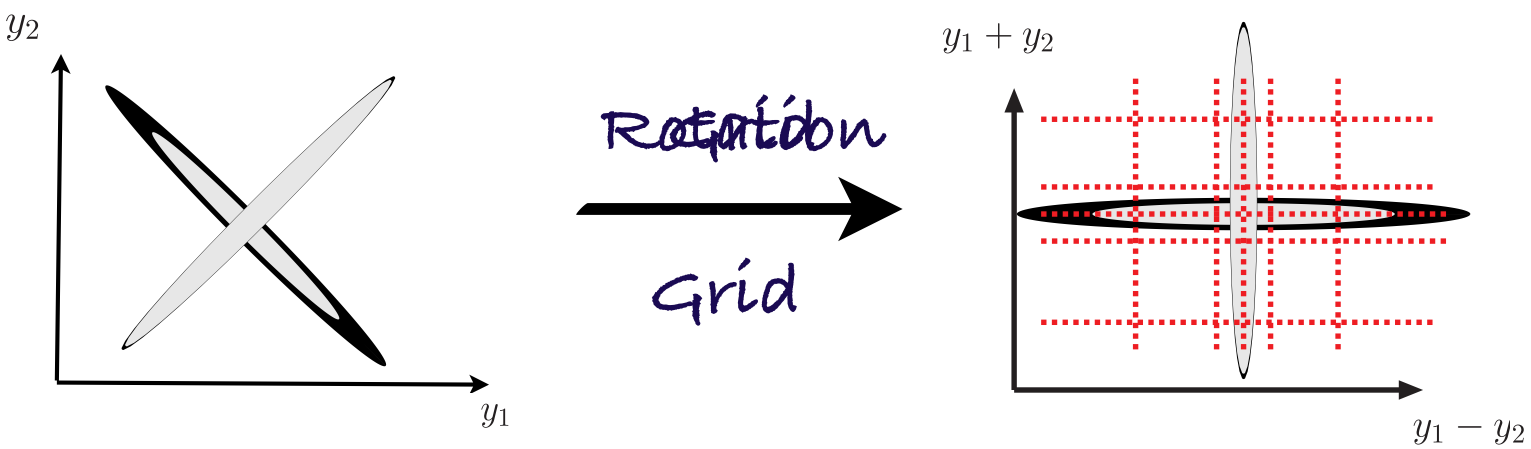
- The choice of the parameterisation has a strong impact on the efficiency



- ☐ The adaptive Monte-Carlo Technique picks point in interesting areas
  → The technique is efficient

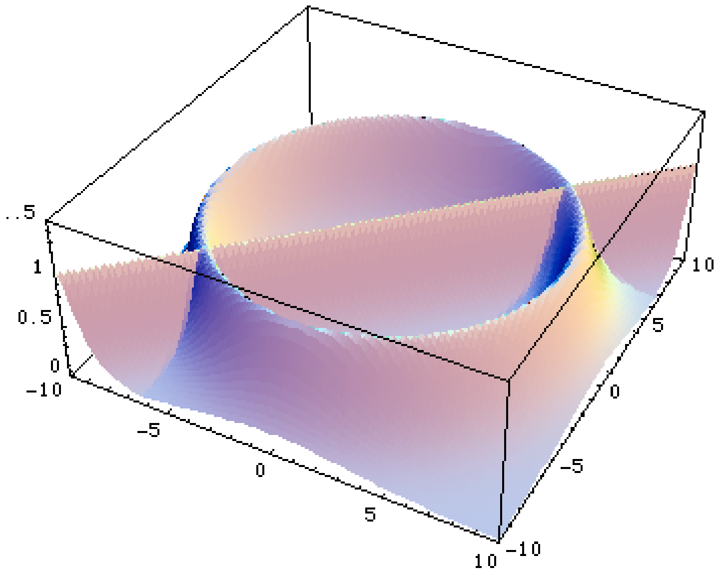
Monte-Carlo Integration

- The choice of the parametrization has a strong impact on the efficiency



- The adaptive Monte-Carlo Techniques picks point
  point interesting areas
  → The technique is very efficient / The integral converges slowly

Multi-channel



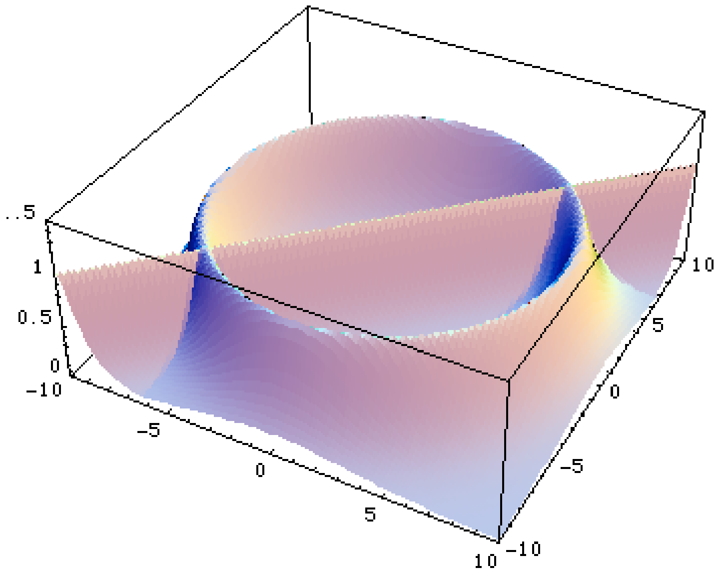
What do we do if there is no transformation that aligns all integrand peaks to the chosen axes?
Vegas is bound to fail!

Solution: use different transformations = channels

$$p(x) = \sum_{i=1}^{n} \alpha_i p_i(x) \quad \text{with} \quad \sum_{i=1}^{n} \alpha_i = 1$$

with each $p_i(x)$ taking care of one "peak" at the time

Multi-channel

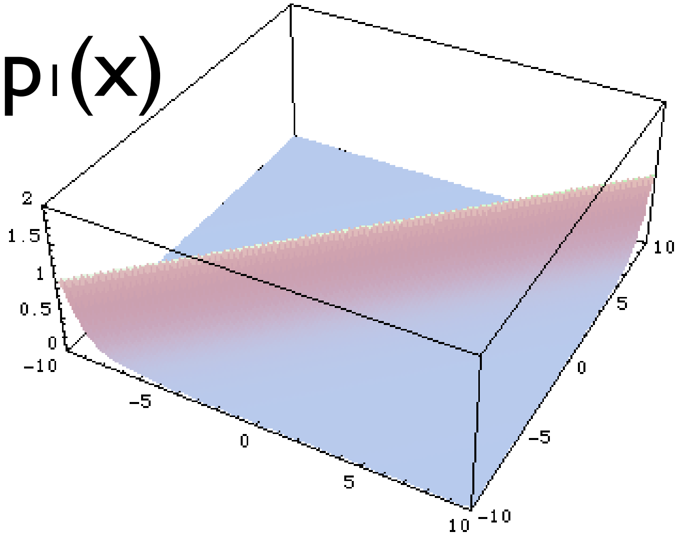


$$p(x) = \sum_{i=1}^{n} \alpha_i p_i(x)$$

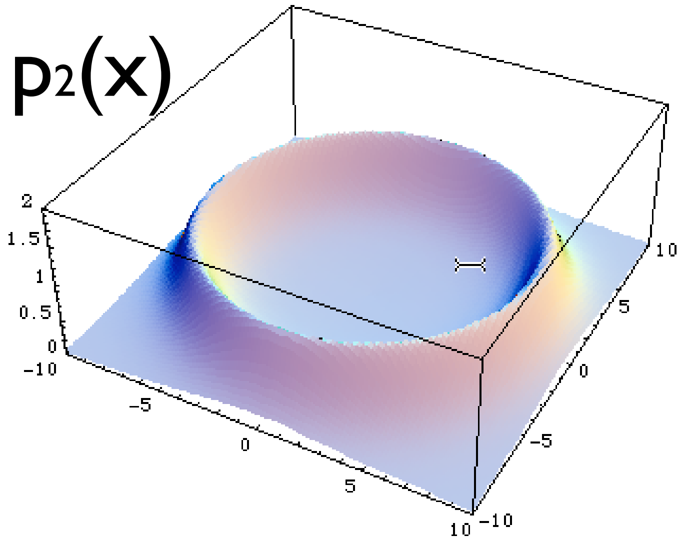
with

$$\sum_{i=1}^{n} \alpha_i = 1$$

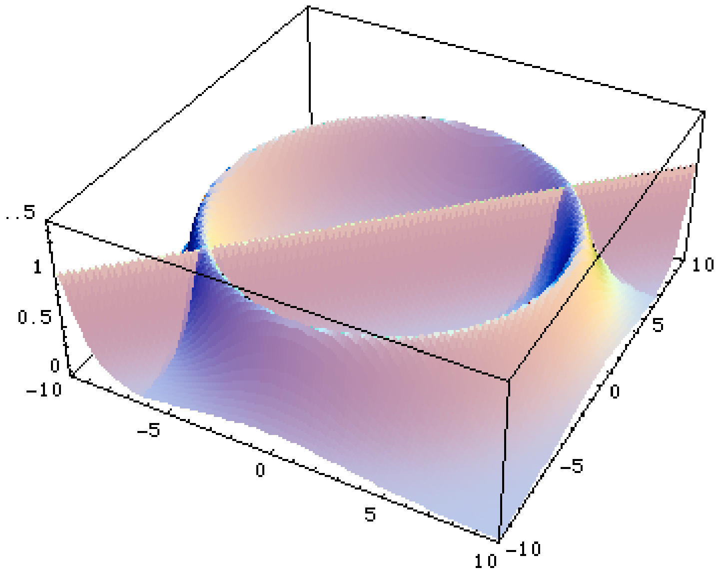
$p_1(x)$



$p_2(x)$



Multi-channel



$$p(x) = \sum_{i=1}^{n} \alpha_i p_i(x)$$

with

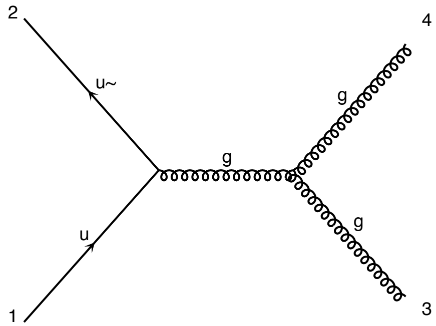
$$\sum_{i=1}^{n} \alpha_i = 1$$

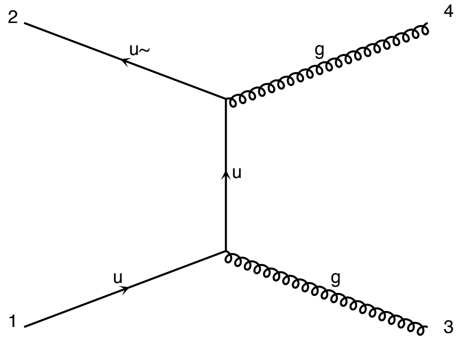
Then,

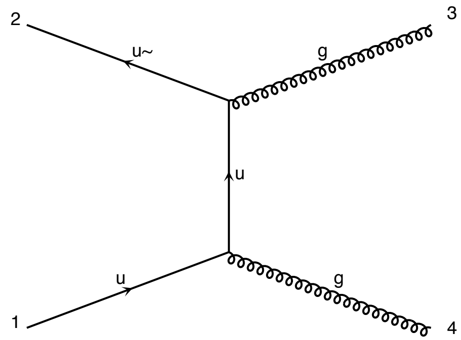
$$I = \int f(x)dx = \sum_{i=1}^{n} \alpha_i \int \frac{f(x)}{p(x)} p_i(x)dx$$

$$\approx 1$$

Example: QCD 2 → 2







$$\propto \frac{1}{\hat{s}} = \frac{1}{(p_1 + p_2)^2} \qquad \propto \frac{1}{\hat{t}} = \frac{1}{(p_1 - p_3)^2} \qquad \propto \frac{1}{\hat{u}} = \frac{1}{(p_1 - p_4)^2}$$

Three very different pole structures contributing to the same matrix element.

Multi-channel

Consider the integration of an amplitude |M|^2 at tree level which many contributing diagrams. We would like to have a basis of functions,

$$f = \sum_{i=1}^{n} f_i \quad \text{with} \quad f_i \geq 0\,, \quad \forall \;\; i\,,$$

such that:

1. we know how to integrate each one of them,
2. they describe all possible peaks,

giving us the combined integral

$$I = \int d\vec{\Phi} f(\vec{\Phi}) = \sum_{i=1}^{n} \int d\vec{\Phi}\, g_i(\vec{\Phi})\, \frac{f_i(\vec{\Phi})}{g_i(\vec{\Phi})} = \sum_{i=1}^{n} I_i\,,$$

Does such a basis exist?

*Method used in MadGraph

Trick in MadEvent: Split the complexity

$$\int |M_{tot}|^2 = \int \frac{\sum_i |M_i|^2}{\sum_j |M_j|^2} |M_{tot}|^2 = \sum_i \int \frac{|M_i|^2}{\sum_j |M_j|^2} |M_{tot}|^2$$

≈ 1

- Any single diagram is “easy” to integrate (pole structures/suitable integration variables known from the propagators)
- Divide integration into pieces, based on diagrams
- All other peaks taken care of by denominator sum

N Integral

- Errors add in quadrature so no extra cost
- “Weight” functions already calculated during $|M|^2$ calculation
- Parallel in nature

$$\int |M_{tot}|^2 = \int \frac{\sum_i |M_i|^2}{\sum_j |M_j|^2} |M_{tot}|^2 = \sum_i \int \frac{|M_i|^2}{\sum_j |M_j|^2} |M_{tot}|^2$$

P1_qq_wpwm

s= 725.73 ± 2.07 (pb)

Graph	Cross-Section ↓	Error	Events (K)	Unwgt	Luminosity
G2.2	377.6	1.67	142.285	7941.0	21
G3	239	1.16	220.04	10856.0	45.5
G1	109.1	0.378	70.88	3793.0	34.8

P1_gg_wpwm

s= 20.714 ± 0.332 (pb)

Graph	Cross-Section ↓	Error	Events (K)	Unwgt	Luminosity
G1.2	20.71	0.332	7.01	373.0	18

term of the above sum.

each term might not be gauge invariant

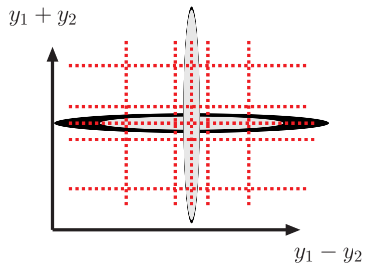
To Remember

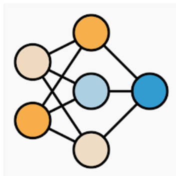
- Phase-Space integration is difficult
- We need to know the function
  - ➡ Be careful with cuts
- MadGraph split the integral in different contribution linked to the Feynman Diagram
  - ➡ Those are not the contribution of a given diagram

Can we do Better?

- Importance sampling/VEGAS is learning a function
  - ➡ HOT TOPIC: Machine Learning
  - ➡ Lot of work in progress

multi-channel VEGAS → MADNIS (Last week number)





W2j

Setup	Channel	Integral I [pb]	σ/I
VEGAS $\eta = 3.9\%$	1	2.057(4)	0.98
	3	108.4(3)	1.46
	4	31.54(7)	1.20
	7	73.4(2)	1.28
	sum	215.4(4)	1.39

→

Setup	Channel	Integral I [pb]	σ/I
VEGAS-Flow (trained α, stratified) $\eta = 14.3\%$	1	0.0059(3)	0.24
	3	100.27(6)	0.37
	4	10.86(2)	0.55
	7	104.16(8)	0.55
	sum	215.30(10)	0.47

(Preliminary)

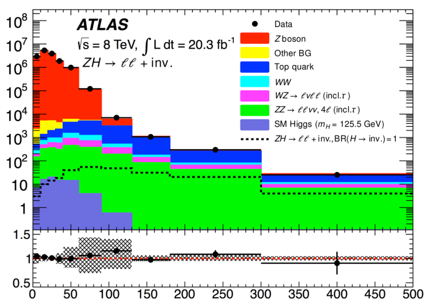
Variance reduce by a factor 3 (so convergence 9 times faster)

Event generation also three times more efficient

Event Generation

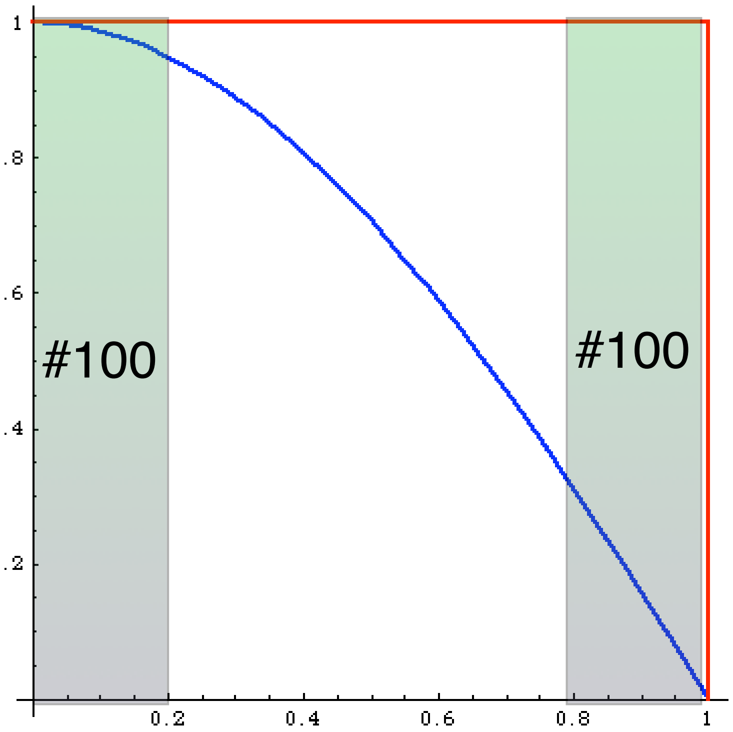
What is the goal?

- Cross-section
  - But large theoretical uncertainty

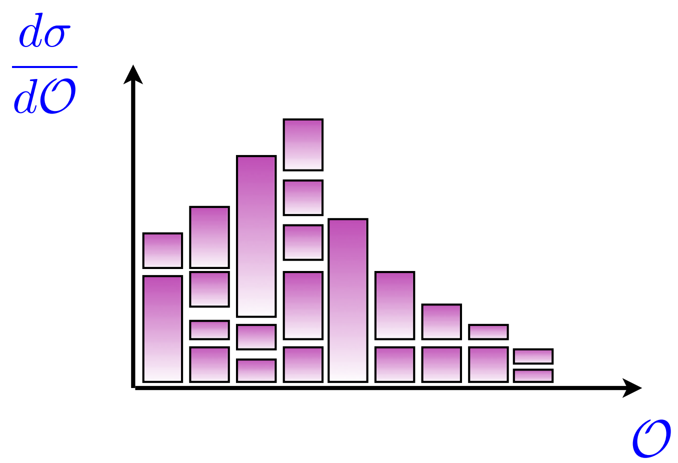


- Differential Cross-Section
  - Provided as sample of events
  - Sample size is problematic
    - Those events will need to have full detector simulation

How to get sample?

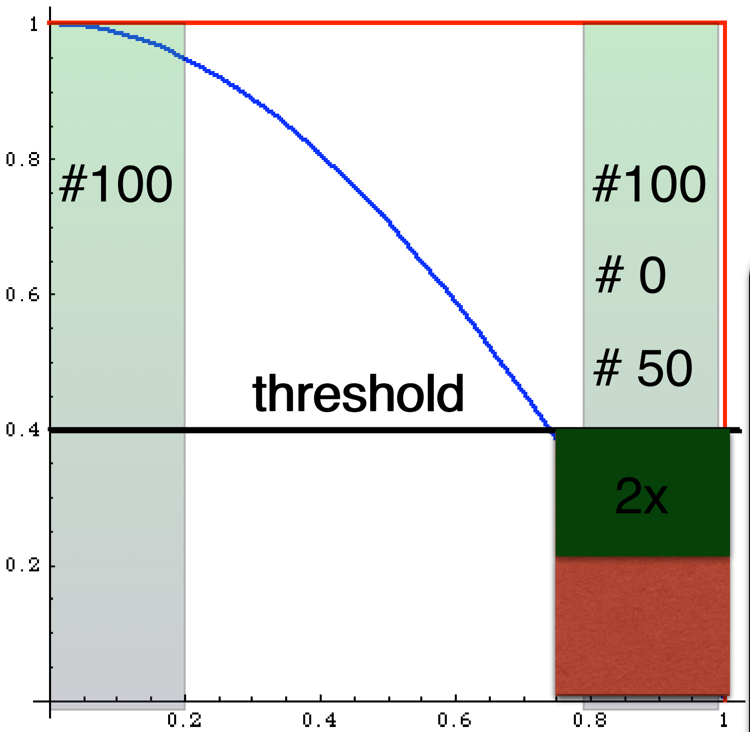


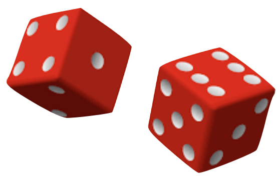
- Monte-Carlo integration use random points
  - We can keep those
    - (Uncorrelated) sample



- Points not distributed as the real function
- Need to keep track of the importance of each point (weight)
- Typically a lot of event have low information

Do we need to keep small weight?

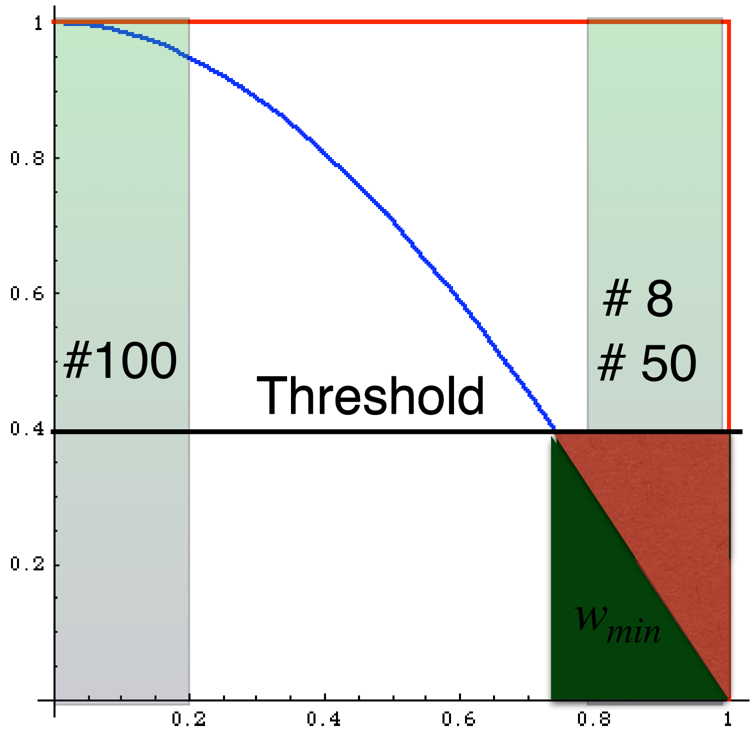




- Let's put a minimum
  - Discard events below the minimum
  - NO! We loose cross-section/ bias ourself

- Let's put a minimum
  - But keep 50% of the events below
  - Multiply the weight of each event by 2 (preserve cross-section)
  - We loose information
  - But we gain in file size

Do we need to keep small weight?

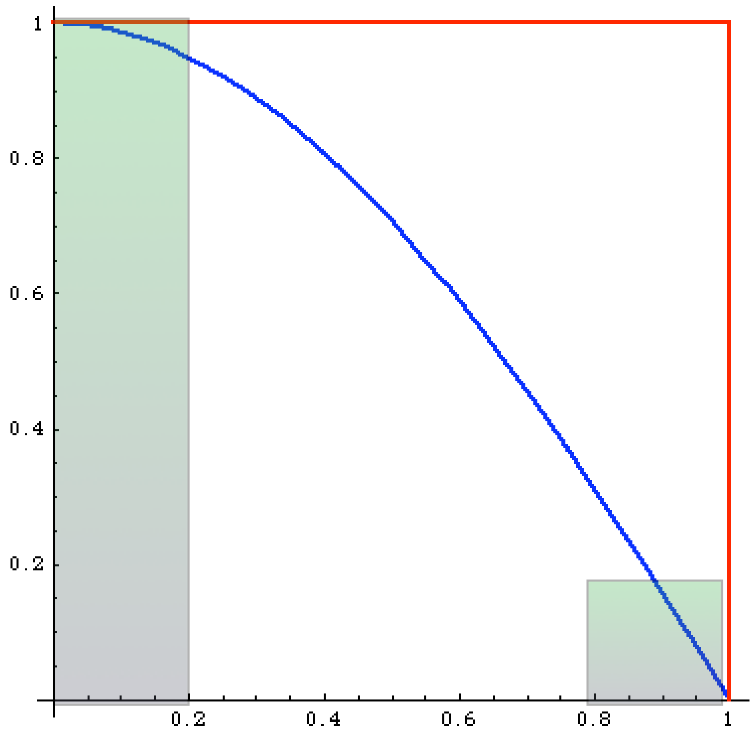


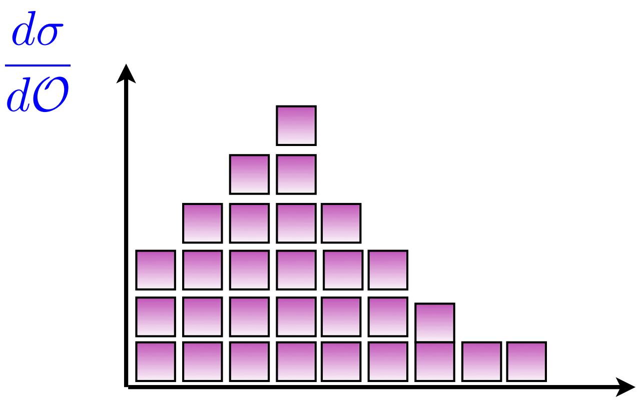
- Let's put a threshold
  - But keep 50% of the events below
  - Multiply the weight of each event by 2 (preserve cross-section)

- Let's improve
  - Let's make the threshold proportional to the weight
  - Keep each event with $\frac{100w}{w_{thres}}\%$ probability
  - If kept multiply his weight by $\frac{w_{thres}}{w}$
  - So the new weight is w_{thres}

Unweighted events

Events distributed as in nature

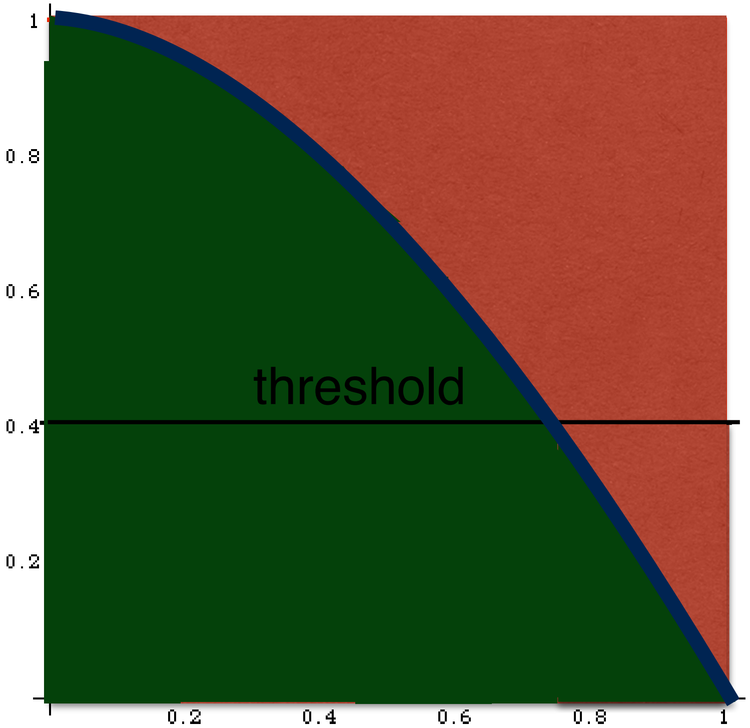




- All bins should event event proportional to their cross-section (Up to Poisson distribution)
- All events should have the same weight

- This correspond to the smallest file size or maximum compression

Do we need to keep small weight?

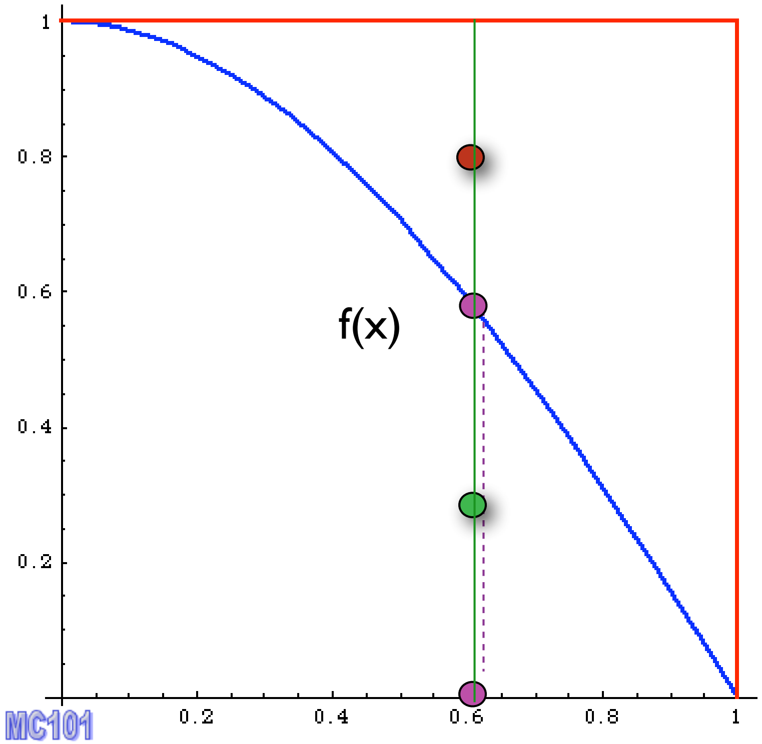


- Let's improve
  - Let's make the threshold proportional to the weight
  - So the new weight is w_{thres}

- Let's all event have the same weight
  - So set $w_{thres} > \max(w)$
- Maximal compression

Event generation

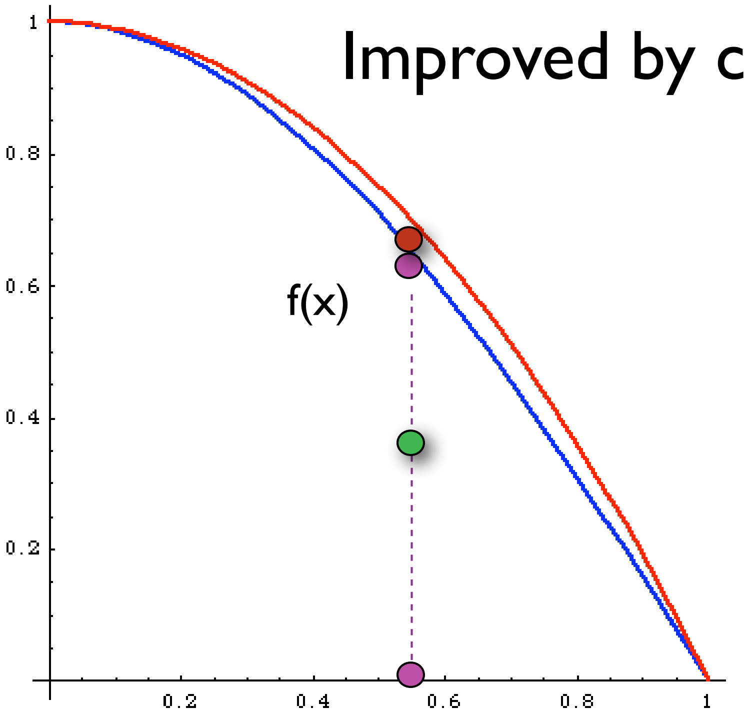
$$\int f(x)dx = \frac{1}{N}\sum_{i=1}^{N} f(x_i) = \frac{1}{N}\sum_{i=1}^{N} \frac{f(x_i)}{w_{thres}} w_{thres}$$



1. pick x_i
2. calculate $f(x_i)$
3. pick $y \in [0, max(f)]$
4. Compare:
   if $y < f(x_i)$ accept event,

   else reject it.

Event generation



Improved by combining with importance sampling:

1. pick x distributed as p(x)
2. calculate f(x) and p(x)
3. pick 0<y<1
4. Compare:
   if f(x)>y p(x) accept event,

else reject it.

much better efficiency!!!

Monte-Carlo Summary

Bad Point

- Slow Convergence (especially in low number of Dimension)
- Need to know the function
  - Impact on cut

Good Point

- Complex area of Integration
- Easy error estimate
- quick estimation of the integral
- Possibility to have unweighted events