Theory tools for accurate event simulation

Paolo Nason,

INFN, sez. di Milano Bicocca

Frascati, 1/10/2024

The MCWS 2006 workshop

The frontpage of this workshop web page says:

Inspired by the MCWS workshop (2006)

18 years ago ... At that time:

- LHC was getting closer
- Great expectation for new physics signals at the LHC
- Theorists were worried about being prepared to understand LHC data
- As an example: the LHC Olympics were organized, to train young and not so young theorists to understand LHC data.

Alcune considerazioni sulle LHC Olympics (LHCO)

- qual'e` il loro scopo
- in cosa consistono
- chi vi partecipa
- ricadute e sviluppi futuri

Barbara Mele

INFN Istituto Nazionale di Fisica Nucleare

Sezione di Roma

(con contributi di Gian Giudice)

LHCO: in cosa consistono (cont.)

Black boxes

➡ "unsophisticated" approach to the

LHC Inverse Problem

given a new-physics ***signal*** at LHC, how can we use it to determine the ***underlying theory*** (the TeV Lagrangian, the string/M theory vacuum, . . .)?

"black boxes" = data sets

a) generated with specified programs (mostly with Pythia) from new-physics models *(unknown to LHCO participants)*,

b) processed through a simulation of an *LHC-like* detector (PGS);

✓ participants are challenged to look at , interpret the LHC new physics blackbox signals, and find out

what underlying model has generated these data !

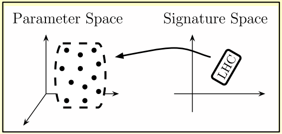
Supersymmetry and the LHC Inverse Problem

N. Arkani-Hamed, G. L. Kane, J. Thaler, and Lian-Tao Wang

arXiv:hep-ph/0512190 v1 14 Dec 2005

→ study the "inverse map" *from the space of LHC signatures to the parameter space* of theoretical models within MSSM (using 1808 LHC observables)

→ show that the *inverse map* of a point in signature space consists of *a number of isolated islands* in parameter space



→ existence of **"degeneracies"** = qualitatively different models with the same LHC signatures. (reflecting discrete ambiguities in electroweak-ino spectrum)

Top Partners at the LHC: Spin and Mass Measurement

P. Meade and M. Reece

hep-ph/0601124

→ *model independent analysis* of the phenomenology of the "*top partner*" t′ (odd under a parity which is responsible for the stability of a WIMP)

→ discover opportunities at LHC, *mass* determination, and *spin* determination of t′

Olympics

- Can be seen as an effort to alert theorists that data was coming soon
- Mostly theorists involved

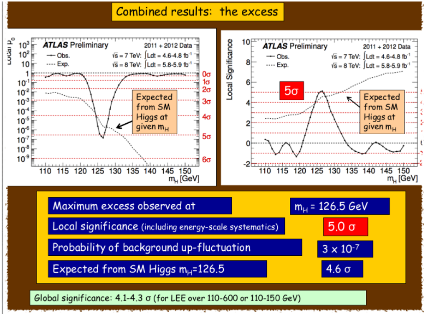
MCWS 2006 Frascati workshop:

- Emphasis on Monte Carlo's ("the natural interface between theorists and experimentalists")
- Promote cohesion and a common language in the LHC experimental and theory community.

- The Workshop led to a (two volumes) publication: https://arxiv.org/abs/0902.0293, 0902.0180
- Several introductory articles on collider physics, the LHC and its experiments, as well as introductory theory articles.
- Emphasis on multi-jet processes (expected background for new physics signals)

Higgs discovery

The first few years of running: **the LHC delivered the Higgs!**
However: no indications of new physics signals.
All particles required by the Standard Model where discovered.



CERN, 4th July 2012

Today

- Open problems (Hierarchy problem, Strong CP problem, Dark Matter, etc.) are still open
- The Higgs has been studied in great details: it looks like the SM Higgs, but important questions remain open.
- Accurate tests of the Standard Model have started, in particular (but not only) on the Higgs interactions.
- The theoretical tools and calculations have seen an unprecedented progress
- More emphasis on precision.
- New frameworks for the search of new physics have been introduced (Effective field theories).
  New physics searches $\rightarrow$ precision physics measurements.

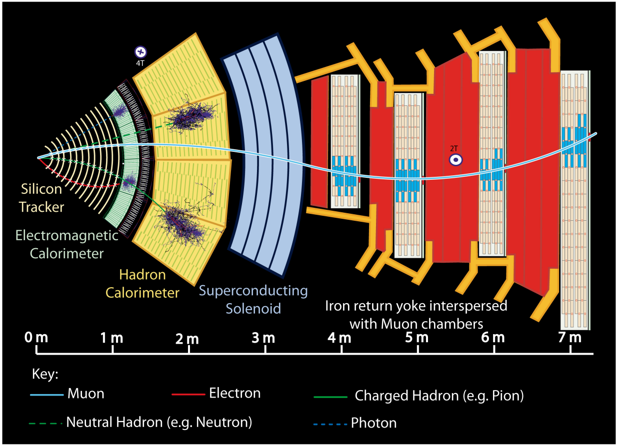
The 2020 Olympics: use of machine learning techniques to study faint BSM signals over noisy background, ...

Hadron collider physics: a very brief reminder

In most hadronic collisions nothing interesting happens. Lots of particles with small transverse momenta and a large spread in rapidity are generated. We are interested in short-distance (high transverse momentum or mass) phenomena: so called "hard interactions".

- Collision events formed by hard interaction, accompanied by an underlying event: (the remnants of the proton, multiparton interactions, etc.)
- The Hard interaction giving rise to subsequent, short distance radiation
- At the end long lived particles are formed ...

The CMS Detector



- The experiments measures tracks, calorimeter deposits, etc.

- Interaction rate: 1GHz; 1 raw event: 1Mb; cannot record everythig, keep only “interesting” events (100Hz). Complex trigger system...
- Pile up: large number of interactions per bunch crossing ($\approx$30). Must find a way to single out the interesting one.
- The experiment reconstructs Detector level objects:
    - Muons (tracker + muon chambers)
    - Electrons (tracker + Em calorimeter)
    - Photons (Em calorimeter)
    - Hadronic jets (calorimeters (+tracker))
    - ...

  that match as close as possible the corresponding particle.

The TRUE LHC INVERSE PROBLEM

From detector objects go back to the short-distance process.

Needs: a Monte Carlo generator and a Detector Simulator

To exemplify the procedure:

- Generate a large Monte Carlo sample of events.
- Feed the Monte Carlo events through the Detector Simulator
- Compute distributions in terms of Detector Level Objects, MC particles, and eventually MC primary partons.
- Compute the corrections for going from the particle (parton) level to the detector level objects (for example computing a bin migration matrix; or using Machine Learning techniques).
- Unfolding: invert the correction and apply it to real data, to obtain the particle/parton level distributions.

- “Primary parton” unfolding used often in the past. Now deprecated. Still unavoidable in certain cases (e.g. top production).

Corrected distributions are compared to calculations.

Notice that the Monte Carlo model affects unfolding. Hence the importance of accurate simulations.

Physics predictions: Fixed order calculations

Based upon the factorization theorem in QCD:

$$\mathrm{d}\sigma = \sum_{ij} \int \mathrm{d}x_1 \mathrm{d}x_2\, f_i(x_1, \mu)\, f_j(x_2, \mu)\, \mathrm{d}\hat{\sigma}(x_1 p_1, x_2 p_2, \mu)$$

where $\hat{\sigma}$ has a power expansion in terms of the strong coupling constant evaluated at the scale μ.

The parton densities $f_i(x, \mu)$ satisfy Altarelli-Parisi evolution equations, which must have sufficient accuracy not to spoil the accuracy of $\hat{\sigma}$:

LL for LO, NLL for NLO, N^nLL for N^nLO.

The cross section $\mathrm{d}\sigma$ is plagued by collinear and infrared singularities. However the integral

$$\int \mathrm{d}\sigma O(\{p_f\})$$

is finite if the observable O is insensitive (does not change abruptly) if a final state, massless, zero energy parton is added (soft safety) or a final state massless parton is split into two partons preserving its total momentum (collinear safety).

Two methods to achieve the cancellation:

- **Slicing**: slice out tiny regions of phase space around the singular one; perform the integral in the singular region analytically, and in the remaining region numerically.
- **Subtraction**: organize the integrations so that the cancellation takes place under the integral sign for soft and collinear safe observables.

Combinations of them are also possible.

NLO QCD corrections

Essentially a solved problem.

- One loop integrals all known since the work of Passarino-Veltman.
- Collinear and soft structure well understood
- Subtraction methods have proven most reliable (Catani-Seymour, Frixione-Kunsz-Signer)
- **complexity has soon become an issue**
- Methods to better handle complexity in the evaluation of real and virtual amplitudes have been found (helicity methods, unitarity, OPP (Ossola-Papadopoulos-Pittau) ...)

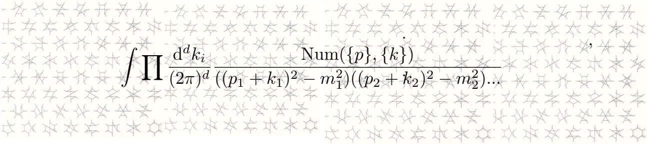
Public tools for their evaluation: Madgraph-MC@NLO, Openloops, Gosam ...

Cross sections with up to six accompanying light partons have been computed.

Beyond one Loop

LOOP AMPLITUDES

The problem of computing loop amplitudes is the problem of calculating divergent integrals of rational functions in Minkowski space.



$$\int \prod \frac{d^d k_i}{(2\pi)^d} \frac{\mathrm{Num}(\{p\},\{k\})}{((p_1+k_1)^2 - m_1^2)((p_2+k_2)^2-m_2^2)...},$$

www.edwardtufte.com

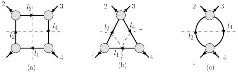
Different techniques to address this problem were developed over time, from analytic to numerical.

$$\sum c_n(s_{ij}, m_k^2) I_k = 0$$

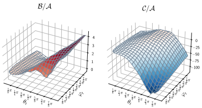
Integration-by-parts

$$s_i \frac{\partial}{\partial s_i} \vec{I} = \epsilon \hat{A}(\{s\}) \vec{I}$$

Differential equations



Generalized unitarity



Numerics: integration, solution of differential equations

$$G(\{a_n, \vec{a}_{n-1}\}, x) = \int_0^1 \frac{dt}{t-a_n} G(\{\vec{a}_{n-1}, t) \qquad K(x,a) = \int_0^x \frac{dt}{\sqrt{(1-t^2)(1-at^2)}}$$

Classes of functions, from Goncharov polylogarithmis, to elliptic integrals.

Chetyrkin, Tkachov, Laporta, Smirnov, von Manteufffel, Lee, Maierhoefer, Usovitsch, Uwer, Abreu, Cordero, Ita, Page, Zeng;, Badger, Hartano, Peraro, Sotnikov, Zola, Gehrman, Henn, Chicherin, Tancredi, Caola, Buncioni, Devoto, Chen, Czakon, Poncelet, Greiner, Heinrich, Kerner, Jones, Liu, Ma, C.Y.Wang, Moriello, Steinhauser, Schönwald, Anastasiou, Sterman, Hirschi

(From Melnikov ICHEP talk)

Beyond one Loop

Besides the 2-Loop amplitude:

- Tree graphs with two more final state partons (double real)
- One loop graphs with one real emission (real-virtual)

Subtraction-slicing methods become much more complicated, and several methods have been proposed to deal with them.

Gehrmann, Glover, Czakon, Caola, Roentsch, Melinkov, Troscanyi, Somogyi, Del Duca, Duhr, Kardos, Magnea, Bertolotti, Pelliccioli, Uccirati, Torrielli, Signorile-Signorile, Catani, Grazzini, Boughezal, Petriello, Tackmann, Gaunt, Stahlhofner,Tagliabue, Devoto ...

NNLO and N^{3}LO results

Going beyond 1-loop has been much harder; by now

- NNLO results for many 2$\to$2 processes and some 2$\to$3
- N^{3}LO results for 2$\to$1 processes; 2$\to$2 is the frontier.
- First (approximate) N^{3}LL parton densities have appeared.

Hard to predict where this will end ...

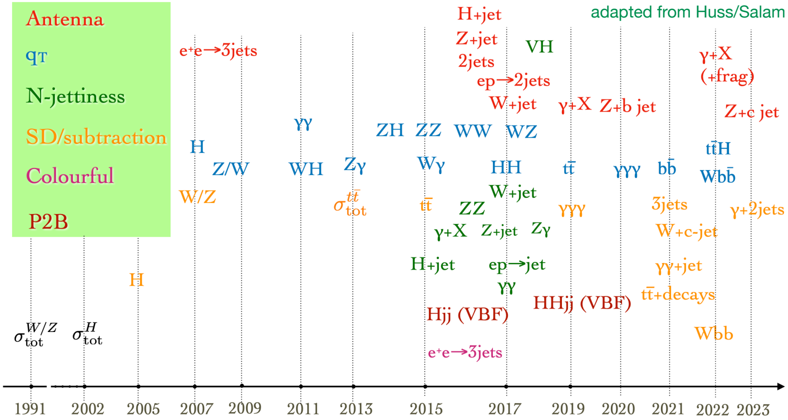
By comparison: NLO timeline

- Drell Yan: Altarelli, Ellis, Martinelli 1978
- $e^+e^- \rightarrow$3 jets, Ellis, Ross, Terrano 1980
- Prompt photon, Aurenche etal (1983)
- Heavy Flavour production: Dawson, Ellis, P.N. 1988
- Pairs of heavy bosons: Mele, Ridolfi, P.N.: 1990
- Jets HH : S.Ellis, Kunszt, Soper, 1992
- ...
- Automation of subtractions: Catani-Seymour; Frixione, Kunszt, Signer, 1996
- Complexity: OPP 2007

18 years from first calculation to automated subtraction;
$+$ 11 more years for progress on complexity.

In spite of the enormous growth in complexity, time of development is of the same order: work is split into subtasks, people become more clever, new ideas pop up.

NNLO timeline



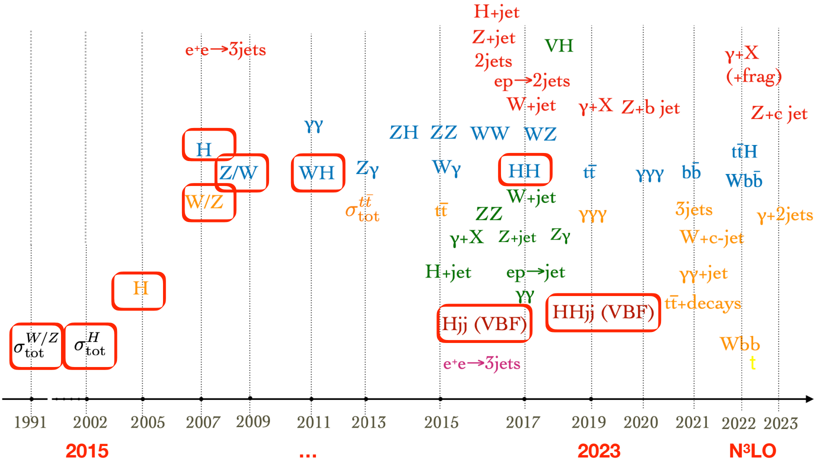
Different colour: different way to handle intermediate divergences

The dream is to have NNLO fully automated for generic processes [Sotnikov]

13

(from Zanderighi LHCP 2024)

NNLO → N3LO



H+jet
Z+jet
2jets
VH
e+e→3jets
ep→2jets
W+jet
γ+X
Z+b jet
γ+X
(+frag)
Z+c jet
γγ
ZH
ZZ
WW
WZ
tt̄H
H
Z/W
WH
Zγ
Wγ
HH
tt̄
γγγ
bb̄
Wbb̄
W/Z
$\sigma_{tot}^{t\bar{t}}$
tt̄
ZZ
W+jet
γγγ
3jets
γ+2jets
γ+X
Z+jet
Zγ
W+c-jet
H+jet
ep→jet
γγ+jet
H
γγ
tt̄+decays
Hjj (VBF)
HHjj (VBF)
$\sigma_{tot}^{W/Z}$
σ_{tot}^{H}
Wbb
t
e+e→3jets
1991
2002
2005
2007
2009
2011
2013
2015
2017
2019
2020
2021
2022
2023
2015
...
2023
N3LO

EW corrections:

- Beyond NLO: mixed QED-QCD corrections;
- Photon and lepton PDF's now available with high precision thanks to the LUX approach.

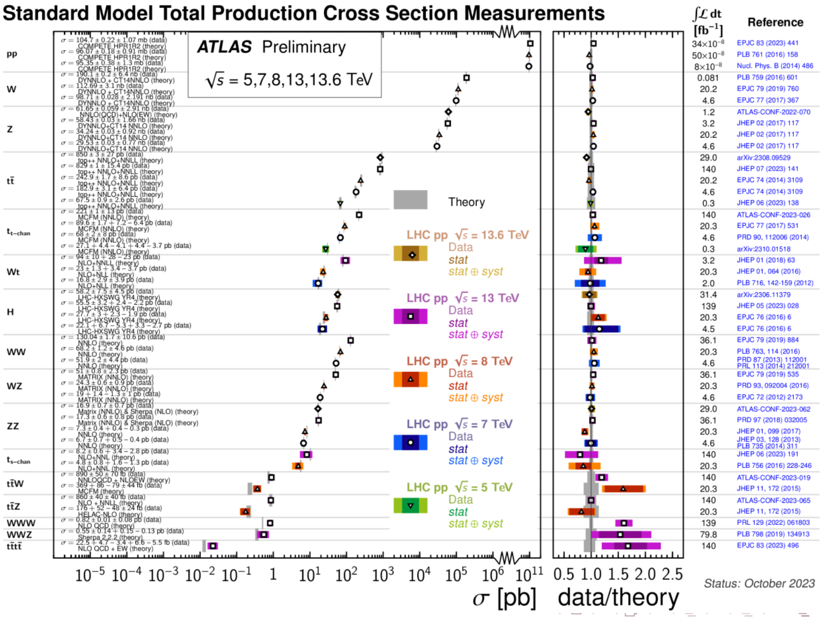
Resummation of logarithmically enhanced distributions near singular regions of the phase space have advanced in precision, and are routinely used to improve fixed order prediction:

$$\log W = \underbrace{L g_1(\alpha_s L)}_{\mathrm{LL}} + \underbrace{g_2(\alpha_s L)}_{\mathrm{NLL}} + \alpha_s \underbrace{g_3(\alpha_s L)}_{\mathrm{NNLL}} \dots$$

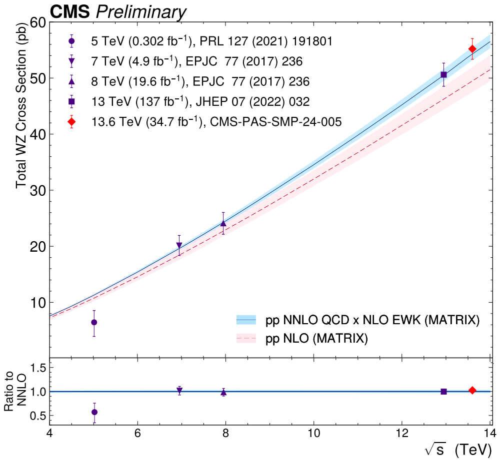
where L is a divergent log near the singular region.

Typically applied to transverse momentum distribution, but also to improve total cross section prediction in the threshold limit (assuming that the damping in parton luminosity for large values of $x_1 x_2$ is enforcing a threshold suppression).

Improved agreement with experiments



Improved agreement with experiments

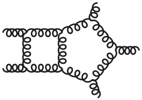


55.2±1.2(stat)±1.2(syst)±0.8(lumi)±0.1(Theo)

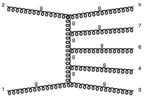
54.7+1.2-1.1(scale) NNLO QCD × NLO EW (MATRIX).

Improved agreement with experiments

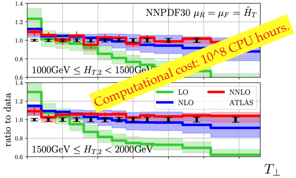
LHC experiments can measure the running of the strong coupling constant at very high energies. A useful observable is the transverse energy-energy correlator for 3j events. NLO results for this observable were known since quite some time. Pushing them to the next level — NNLO — was an enormous adventure.



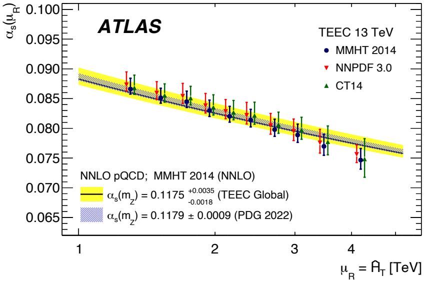
Two-loop amplitudes: Chicherin, Sotnikov, Gehrmann, Zhang, Henn, Wasser, Zola, Abreu et al.



Subtraction scheme: Czakon



Alvarez, Cantero, Czakon, Lorente, Mitov, Poncelet



(from Melnikov ICHEP talk)

Improved agreement with experiments

For a competitive measurement of the strong coupling at the LHC, one needs to find a quantity which

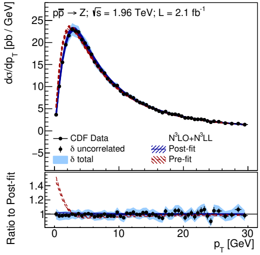
1) is proportional to the strong coupling constant;
2) can be predicted theoretically with a percent precision (NNLO and higher);
3) is independent (nearly independent) of poorly-known parton distribution functions;
4) refers to low(er) region of hard momentum region;
5) does not suffer from unknown non-perturbative effects.

Inclusive Z transverse momentum distribution seems to fit the bill.

$$\frac{d\sigma_Z}{\sigma_z dp_\perp} \sim \frac{\alpha_s(p_\perp)}{2\pi p_\perp} \ln \frac{M_Z}{p_\perp}$$

ATLAS followed up on the proposal and obtained a very precise value of the strong coupling constant which is very well-compatible with the world average.

$$\alpha_s(m_z) = 0.1183 \pm 0.0009 \qquad \text{ATLAS, 8 TeV data}$$



Camarada, Ferrera, Schott

(from Melnikov ICHEP talk)

What we learned:

- Perturbation theory still in a regime of apparent convergence (not obvious in 2006)
- Important for improved agreement with data
- Look forward to: automation? Complexity?

Shower Monte Carlo (SMC)

The Stone Guest of LHC physics.

- Hard interaction (SM lagrangian)
- Collinear and Soft gluon radiation (QCD)
- Hadronization (QCD inspired models)
- Underlying event (QCD+models)
- Decays (from data)
- All is tuned to data.

SMC are essential to correct for detector effects, estimating and subtracting backgrouds, interpreting measurements, etc.

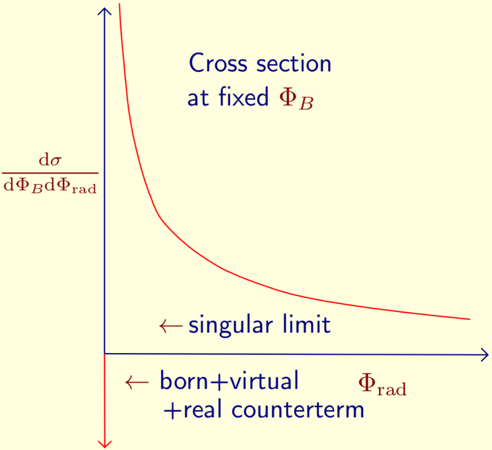
Currently used SMC:

- Hard process: LO
- Radiation:
  - Iterated Collinear approximation + soft improvement (angular ordering)
  - Iterated Soft approximation at large N_C + collinear improvement (dipole based)
  - Hardest radiation at LO (MEC=Matrix Element Corrections)
- Hadronization models: string model, cluster model
- Underlying event with a model for multiparton interactions

So far the improvement of the SMC accuracy has focused upon the top of the list:

- Correct the hardest radiation so that for inclusive quantities NLO accuracy is reached (NLO+PS). Some MC implement their own scheme. Some are available as external programs.
- Correct up to the first two hardest radiations so that NNLO accuracy is reached (NNLO+PS).

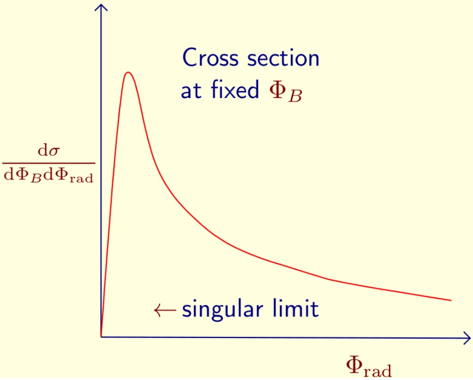
NLO calculations: a pictorial representation



At fixed Φ_B, as a function of Φ_R, the cross section is a distribution, i.e. it is divergent in the singular limit, but it has a finite integral over the singular region, i.e. mathematically it is a distribution.

NLO+PS

NLO+PS: a pictorial representation



At fixed Φ_B, as a function of Φ_R, the cross section is a smooth function.
Its integral over the singular region is the same as in the NLO cross section.
Differs with respect to the pure NLO due to NNLO and even higher order terms arising from the resummation of leading logarithmic terms.

Singularity in $\Phi_{\rm rad}$ tamed by the resummation of Sudakov enhanced contributions.

NLO+PS

Several methods available

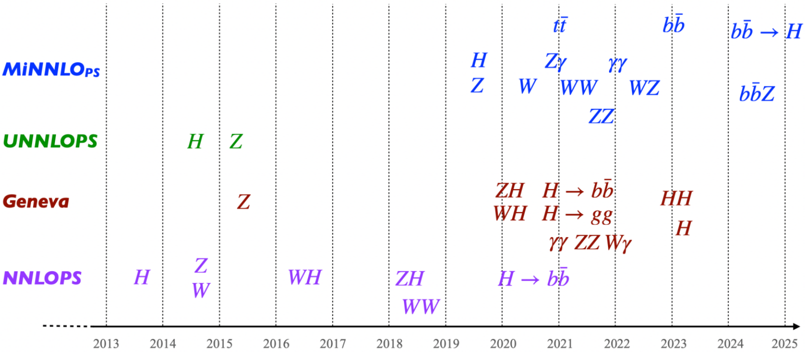
- MC@NLO: the first of the kind; fully automated in the Madgraph_MC@NLO package
- POWHEG: positive weights (an issue if large samples are needed). High degree of automation in the POWHEG BOX framework (uses matrix elements from external providers)

Widely used interfaced to Pythia and Herwig.

Several other methods have been proposed:

- KrkNLO, positive weights, restricted applicability.
- MAcNLOPS, positive weights
- UNLOPS
- ...

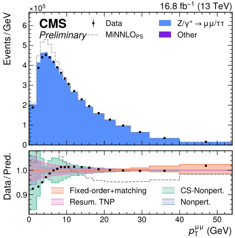
One step further: **NNLO+PS**



Several methods and processes are available.
Currently used by the experimental collaborations.

Theoretical Modeling: Technical Details

- Fully coherent theoretical treatment for W and Z (both μ and τ decays)
- Fully simulated MC samples with MiNNLOPS + Pythia 8 + Photos
  - $\mathcal{O}(\alpha_s^2)$ accuracy (also for angular coefficients), but limited logarithmic accuracy for W/Z p_T modeling from POWHEG emissions and shower



- σ^{U+L} is corrected double (triple) differentially for W (Z) production using resummed SCETLIB prediction matched to fixed order DYTurbo prediction ($N^3LL + NNLO$ for nominal predictions)
- Angular coefficients are left as-is (validated against MCFM and DYTurbo fixed order predictions)*

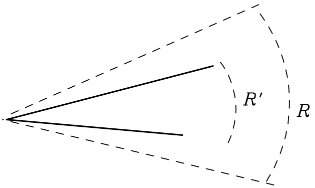
$$\frac{d^5\sigma}{dq_T^2\, dy\, dm\, d\cos\theta\, d\phi} = \frac{3}{16\pi}\frac{d^3\sigma^{U+L}}{dq_T^2\, dy\, dm}[(1+\cos^2\theta) + \frac{1}{2}A_0(1-3\cos^2\theta) + A_1\sin 2\theta\cos\phi$$
$$+\frac{1}{2}A_2\sin^2\theta\cos 2\phi + A_3\sin\theta\cos\phi + A_4\cos\theta + A_5\sin^2\theta\sin 2\phi + A_6\sin 2\theta\sin\phi + A_7\sin\theta\sin\phi]$$

Jets substructure (JSS)

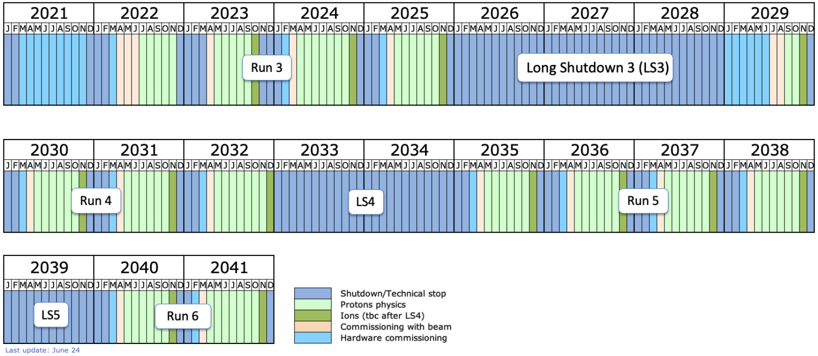
The field of JSS has seen a considerable growth during the LHC running.

- Algorithms to clean jets from underlying event and pile-up effects (grooming and trimming)
- Discriminate quarks and gluon jets
- Discriminate jets containing (hadronically decaying) heavy objects from ordinary QCD jets: boosted Higgs, W/Z, top quarks, BSM objects

Using Monte Carlo generator, resummation techniques, machine learning approaches ...



Future challenges



Future challenges

- More luminosity, better detectors, higher precision.
- Further improvements in precision in fixed order and resummed calculations
- Computational cost (handling complexity)

Areas that are left behind:

- SMC
- Non-perturbative effects

Shower accuracy

Since pre-LHC times:

- Substantial change in the implementations: Pythia8, Herwig7 (completely rewritten).
- No substantial progress in shower accuracy (angular ordered showers and dipole showers)

Shower accuracy often qualified on the same terms as resummation accuracy: LL, NLL, etc.
But, from the PDG review:

> The predictions of shower MCs, on the other hand, are cast in terms of complete sets of final-state momenta, on which one can evaluate any observable; i.e., the shower algorithm itself is normally independent of the specific observable(s) under study. Because of this, it is not easy to qualifying the accuracy of a shower MC using the same criteria adopted in resummation calculation.

We know that 2 logarithmic singularities can arise for each power of α_s (each branching in the shower), giving rise to up to $\alpha_s^2 L^{2n}$ terms for generic observables.

For a large class of observables W we have

$$\log W = \underbrace{Lg_1(\alpha_s L)}_{\text{LL}} + \underbrace{g_2(\alpha_s L)}_{\text{NLL}} + \alpha_s \underbrace{g_3(\alpha_s L)}_{\text{NNLL}} \dots$$

A possible criterion for shower accuracy is that this logarithmmic structure is respected for a large class of such observables (proposed by the PanScale collaboration)

(In the past requirements on multiplicity distributions has led to progress in shower algorithms).

- It was found that currently used showers fail the criterion for NLL accuracy (and some also for LL accuracy at subleading number of colours).
- Algorithms for NLL accurate showers have been proposed
- Work in progress towards NNLL showers
- Higher order splitting functions are being considered

Bewick, Ferrario Ravasio, Richardson, Seymour, Dasgupta, Dreyer, Hamilton, Monni, Salam, Soyez, Forshaw, Holguin, Plätzer, Nagy, Soper, van Beekveld, Soto-Ontoso, Herren, Höche, Krauss, Reichelt, Schoenherr, Karlberg, Scyboz ...

SHOWER REVOLUTION for run IV?

“best” theory framework not always successful in SMC land:

- Angular ordering: never fully adopted by PYTHIA6
- Spin correlations in splitting kernels: methods to do them right do exist, but have seldom been adopted
- CMW coupling: seldom adopted.
- truncated showers in matching: routinely neglected.

Are there areas where NLL and NNLL improvements can really make a difference?
Observables sensitive to a more detailed structure of the events, like jets substructure, but also features that can emerge from machine learning techniques, may require much more refined shower algorithms ...

Non-perturbative corrections

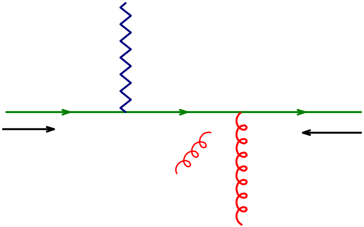
In QCD:

- Corrections like $1/Q^4$ for the simplest processes, $1/Q^2$ for DIS, $1/Q$ for processes involving jets.
- Little is understood also for 3-jet observables in e^+e^- annihilation.
- Efforts to parametrize them using SCET
- Large β_0 models have proven useful for guidelines.

SMC implement their own ideas about power corrections ... unlikely to be correct

The Z transverse momentum

An example:



Does the pattern of asymmetric radiation of very soft, nearly non-perturbative gluons affect linearly the Z transverse momentum?

Intuitive reasoning (and also SMC modeling) may lead to a positive answer...

Λ/p_T correction?

Shown not to be the case in the Large β_0 model

The calculation suggests that there are radiation-recoiling schemes where the cancellation of linear corrections is particularly transparent. Does this apply also to SMC?

Caola,Ferrario Ravasio,Limatola,Mackarov,Melnikov,Ozcelik,P.N.

Conclusions

- Change of perspective since the 2006 workshop: the LHC has quickly delivered the Higgs!
  other expectations were not met, and LHC physics has become tougher than we thought.
- It is also undeniable that both the experimental and theory community has proven to have the strength to meet the challenge
- The work done so far has paid back
- More work is at the horizon, promising to provide us with even better tools.
- We are looking where no man has looked before. Thanks to precision physics/faint signals, we may still find answers to some of the questions that have remained opened.