

KIRILL MELNIKOV (TTP KIT)

PROGRESS IN SUBTRACTION SCHEMES

Parton Showers and Resummation 2019, Vienna, June 2019

We can describe hard collisions at the LHC in terms of quark and gluon cross sections thanks to the collinear factorization theorem

$$d\sigma = \int dx_1 dx_2 f_i(x_1) f_j(x_2) d\sigma_{ij} \mathcal{F}\left(1 + \mathcal{O}\left(\frac{\Lambda_{\text{QCD}}}{Q}\right)\right) \quad \text{Collins, Soper, Sterman}$$

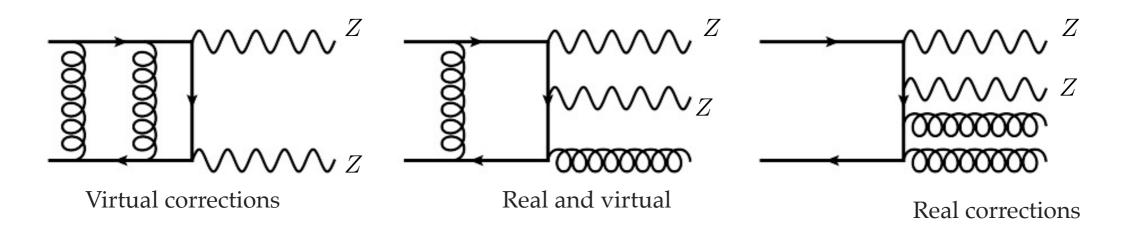
Whenever the process does not involve very large scale hierarchies, and the observable is not strongly sensitive to infrared physics, fixed order calculations provide a robust and reliable framework to obtain precision predictions at the LHC.

Non-perturbative corrections set the limit of perturbative calculation: for generic observables at the LHC scale, they can be at the percent level. Since $\alpha_s \sim 0.1$, this means that we can reliably compute up to NNLO.

The LHC is already now able to measure some standard candles at the few percent level (DY), and in the future it is expected that few-percent precision measurements could be possible for several complex final states.

Fixed order calculations also serve as an important ingredient for resummation and parton shower predictions: matching/merging, extraction of perturbative coefficients...

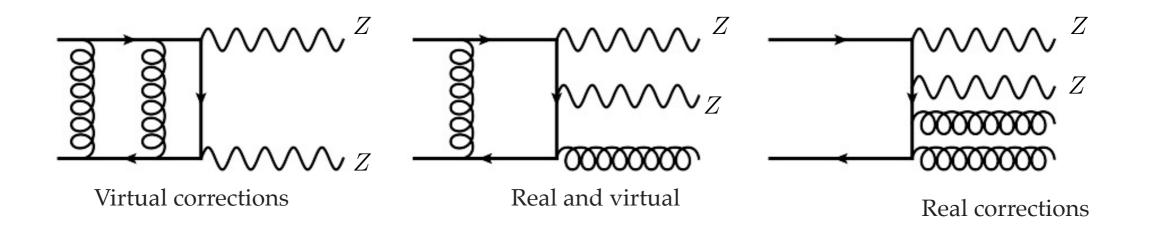
To compute NNLO corrections to $pp \rightarrow X$, one needs two-loop matrix elements for $ff \rightarrow X$, one-loop matrix elements for $ff \rightarrow X+f$, and tree-level matrix elements for $ff \rightarrow X+ff$



Over the past years, there has been a lot of progress in two-loop amplitude calculations. Both analytic and numerical approaches have been developed, and right now we know almost all relevant scattering amplitudes for $2 \rightarrow 2$ reactions. First $2 \rightarrow 3$ results are now appearing.

NNLO Corrections also require 1-loop amplitudes to be evaluated very close to degenerate kinematics. Nevertheless, it seems that at least in some cases current one-loop providers are able to cope with this situation. For example, OpenLoops results have been used for the calculations of NNLO corrections to di-boson processes

Apart from loop amplitude, at NNLO one also needs a framework to deal with extra real emission.

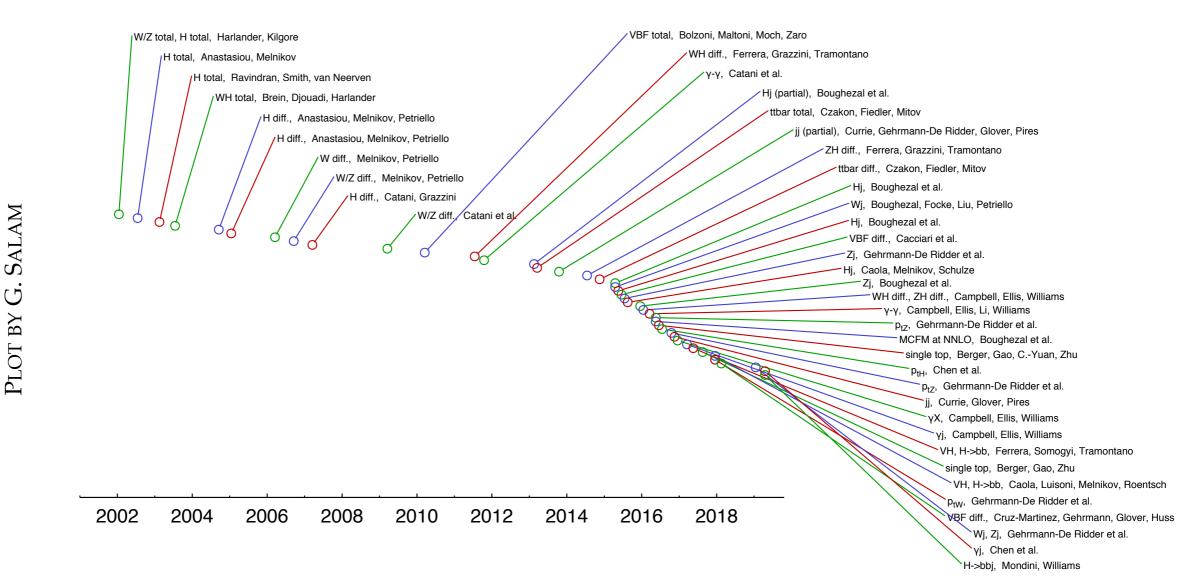


Real emission corrections are finite in the bulk of the allowed phase-space, but infra-red and collinear divergencies appear upon integration over energies and angles of the emitted partons.

To obtain fully differential results, we need to extract these singularities *without* performing the integration over the resolved phase space. The more extra emissions one has to consider, the more complicated this problem is.

For a long time, this was the main bottleneck for computing NNLO predictions. For example, di-jet amplitudes have been known for almost 20 years, but NNLO predictions for di-jet production only became available two years ago.

Over the past few years, several techniques have been developed to deal with the realemission problem. This, combined with the availability of 2-loop amplitudes, lead to a large number of NNLO predictions for important $2\rightarrow 2$ LHC processes.



Currently, going beyond $2 \rightarrow 2$ is prevented by the lack of 2-loop amplitudes. However, even when they would become available, it will require a lot of effort to obtain NNLO predictions using existing methods.

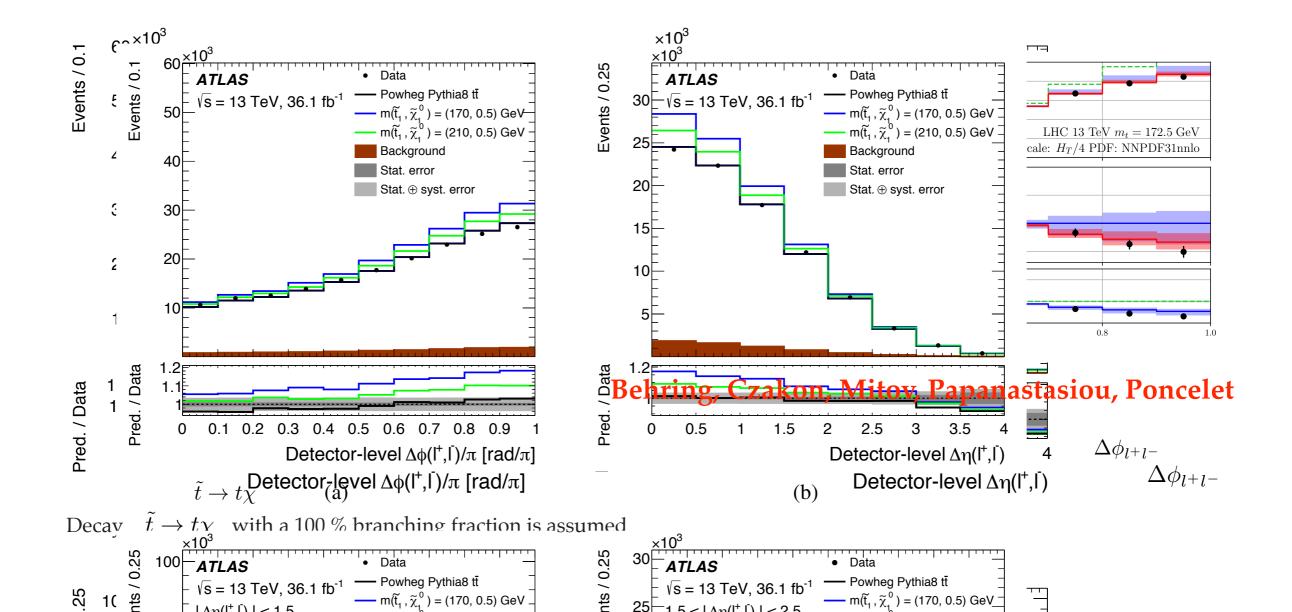
- 1. They work! In general, NNLO QCD improves the agreement between theory and data (WW, WZ, Z+j, di-photons, top pairs);
- 2.It is important to have NNLO QCD computations for fiducial cross sections measured in experiments; corrections to inclusive cross sections and to fiducial cross sections may be quite different (Higgs production in WBF, single top with decay, WW pairs); extrapolations some time lead to ``wrong'' result (WW pairs, top p_t);
- 3.NNLO QCD computations work in ``hard kinematic regions". For an object with an invariant mass O(100) GeV, ``hard" means down to transverse momenta of O(30) GeV. This requires NNLO. Resummations are important but with NNLO results available, they become relevant at low(er) transverse momenta (Z/H pt studies);
- 4. Thanks to these computations, it becomes possible to get information about physics that otherwise it is not accessible.

Thanks to their masses, spins of tops and anti-top remain entangled all the way through the moment of their decay. This effect leads to a small shape change in the $\Delta \phi_{ll}$ distribution.

The presence of stops with masses comparable to the top quark can, with some effort, be detected in that distribution.

```
\Delta \phi_{l^+l^-}
```

Accurate description of this distribution in the $SM \not r_equires$ corrections to top quark production and decay; recently it was extended to NNLO in the narrow width approximation.



Several methods have been developed to deal with real emission at NNLO:

``qt" slicing [Catani, Grazzini]

``<u>Antenna subtraction</u>" [Gehrmann-de Ridder, Gehrmann, Glover et al]

``Jettiness" slicing [Boughezal et al, Gaunt et al]

``<u>Sector decomposition & FKS</u>" subtraction [<u>Czakon</u>, Heymes, Caola, Röentsch, KM]

``Projection-to-Born" [Cacciari et al]

``Colorful'' [del Duca et al]

``Local analytic subtraction'' [Magnea et al]

``Geometric'' [Herzog]

The underlined methods can — in principle — deal with arbitrary processes.

In practice, the complexity grows very fast with the number of color particle at the Born level. Even simple $2\rightarrow 2$ calculations involving colored initial *and* final state require significant computing power.

CPU HOURS FOR FULLY DIFFERENTIAL RESULTS			
	2→1	2→2	2→3
NNLO	100-500	10⁵-10 ⁷	-
N ³ LO	-	-	-

Even with large computing farms, these requirements are likely to severely limit the breadth of high-precision phenomenological studies for complex processes.

An optimal method, able to efficiently deal with complex processes has yet to emerge

Similarly to what happened at NLO, two different strategies have been adopted to deal with real-emission singularities:

PHASE-SPACE SLICING

$$\int |\mathcal{M}|^2 \mathcal{F}_J \mathrm{d}\phi_d = \int_0^{\delta} \left[|\mathcal{M}|^2 \mathcal{F}_J \mathrm{d}\phi_d \right]_{\mathrm{s.c.}} + \int_{\delta}^1 |\mathcal{M}|^2 \mathcal{F}_J \mathrm{d}\phi_4 + \mathcal{O}(\delta)$$

- conceptually simple, straightforward implementation
- must be very careful with residual δ dependence (especially in differential distributions)
- highly non-local → severe numerical cancellations

SUBTRACTION

$$\int |\mathcal{M}|^2 \mathcal{F}_J \mathrm{d}\phi_d = \int \left[|\mathcal{M}|^2 \mathcal{F}_J - \mathcal{S} \right] \mathrm{d}\phi_4 + \int \mathcal{S} \mathrm{d}\phi_d$$

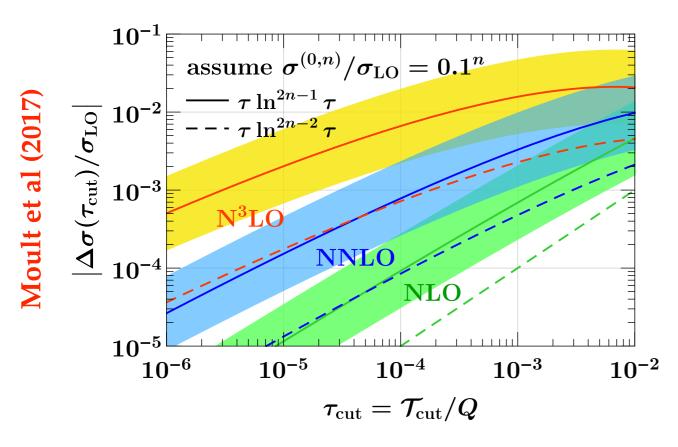
- in principle can be fully local \rightarrow better efficiency / reliability
- requires knowledge of subtraction terms, and their integration

$$\int |\mathcal{M}|^2 \mathcal{F}_J \mathrm{d}\phi_d = \int_0^{\delta} \left[|\mathcal{M}|^2 \mathcal{F}_J \mathrm{d}\phi_d \right]_{\mathrm{s.c.}} + \int_{\delta}^1 |\mathcal{M}|^2 \mathcal{F}_J \mathrm{d}\phi_4 + \mathcal{O}(\delta)$$

Starting point of a N^kLO calculation: find a variable δ that separates resolved/unresolved phase space. E.g.: q_t [Catani, Grazzini], N-jettiness [Boughezal et al, Gaunt et al].

In the unresolved region, use soft-collinear approximation to integrate over unresolved momenta. Typically, result obtained from expanding a resummation formula. Currently, this information is only known numerically for processes involving jets or massive particles (soft function).

In the resolved region, only need N^{k-1}LO corrections to X+J.



Problem: individual contributions are logarithmically divergent ~ $\ln^{2k-1} \delta$.

Power corrections of order $\delta \ln^{2k-1} \delta$. At higher order, very small δ required.

Good control at small δ difficult to achieve.

Because of this issue, eventually abandoned at NLO

$$\int |\mathcal{M}|^2 \mathcal{F}_J \mathrm{d}\phi_d = \int_0^{\delta} \left[|\mathcal{M}|^2 \mathcal{F}_J \mathrm{d}\phi_d \right]_{\mathrm{s.c.}} + \int_{\delta}^1 |\mathcal{M}|^2 \mathcal{F}_J \mathrm{d}\phi_4 + \mathcal{O}(\delta)$$

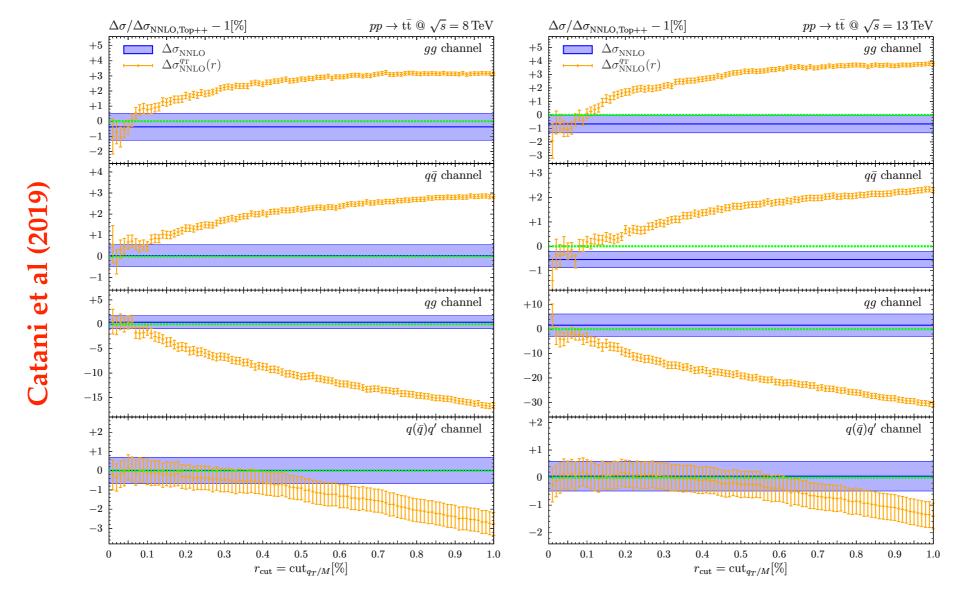
Despite being abandoned long time ago for NLO, slicing techniques have been used for NNLO calculations. This is due at least in part to the following reasons:

- 1.within this approach, one can use existing NLO results for the CPU-intensive `X+J" part of the calculation. The availability of very efficient tools for NLO calculations allowed to obtain stable enough results for several key reactions like V/H, VV, single-top, top decay, H/V+J, H→bb decay, top pair production;
- 2. despite being very CPU-intensive, we now have large computing facilities;
- 3.NNLO corrections are typically small, so often an O(20%-50%) error on the NNLO coefficient only results in a percent-level error on the total cross-section;
- 4.at NNLO, a simple-enough subtraction framework analogous to Catani-Seymour or FKS has yet to emerge.

Nevertheless, slicing techniques are very delicate and if one wants to use them it is very important to always make sure that power corrections are under control. In general, their impact becomes more difficult to control in processes with a non-trivial color structure (see e.g. Campbell et al. (2019)), and in delicate fiducial regions (e.g. isolation).

Recently, slicing the q_t formalism was extended to processes involving massive particles. This allowed for the calculation of NNLO corrections to top pair production within this approach [Catani et al (2019)].

So far, results have been obtained for the total cross section, and are in agreement with the ones obtained using a subtraction formalism [Czakon et al].



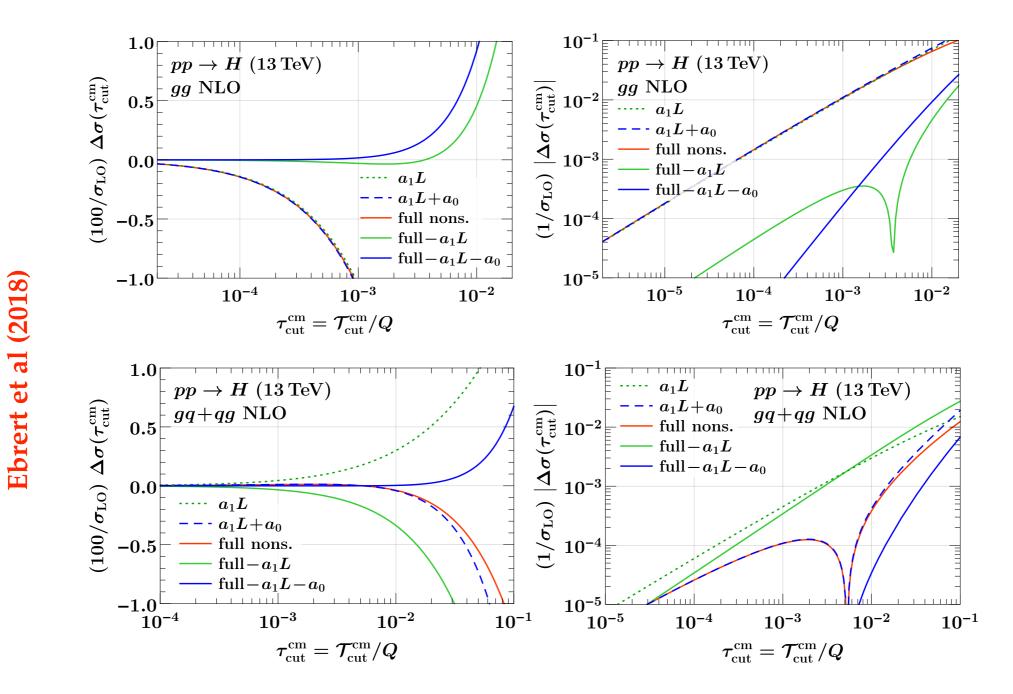
For the total cross section, a detailed study showed that power corrections are under good control.

The use of slicing techniques for complex processes or at higher orders would likely require improvements. Roughly, they can be divided in three categories:

- 1.Devise ``optimal" slicing parameters, as they can lead to better performances. For example, even within N-jettiness slicing it is well-known that different N-jettiness definitions perform very differently. In general, one would want a variable that restricts *all* radiation to be soft and collinear;
- 2. devise more differential slicing approaches. In principle, a fully differential slicing technique can be easily upgraded to a fully-fledged subtraction;
- 3. develop a better understanding of power corrections.

Recently, there has been a lot of activity on (3). Conceptually, this is non trivial because it corresponds to understanding factorization properties of QCD at next-to-leading power.

The first power corrections have been obtained using (SCET-assisted) fixed-order calculations. So far, results are only known for color singlet production [Boughezal et al, Moult et al, Ebert et al], and also in this case their structure is not completely understood. Analogous results for q_t have been obtained [Ebert et al].



For the qg channel, large cancellations between L and NL power corrections. Important to understand to which extent this is accidental.

It will be important to generalize these results to NNLO and to processes involving color in the initial and final states.

$$\int |\mathcal{M}|^2 \mathcal{F}_J \mathrm{d}\phi_d = \int \left[|\mathcal{M}|^2 \mathcal{F}_J - \mathcal{S} \right] \mathrm{d}\phi_4 + \int \mathcal{S} \mathrm{d}\phi_d$$

Slicing problems are overcome with subtraction. One devise a subtraction term *S* that *a*) reproduces the matrix element in the soft-collinear limit and *b*) is simple enough to be integrated for generic configurations of the resolved phase space.

At least in principle, subtraction are local: singularities are subtracted point-by-point in the phase space and not on average.

Because of this, there are no issues about power corrections etc., and the technique should perform much better.

Historically, subtractions outperformed slicing at NLO.

Subtraction scheme however suffer from some drawbacks:

- 1. it is non trivial to identify a good subtraction function
- 2. integration of the subtraction terms can be (prohibitively) difficult
- 3. it is non-straightforward to reuse existing NLO results for the ``X+J" part of the calculation

Currently, two different fully-fledged subtraction schemes exist for LHC processes:

- 1. antenna subtraction [Gehrmann-de Ridder, Gehrmann, Glover et al].
 - Basic idea: construct an approximation to the full matrix element valid over the whole phase space and simple enough to integrate it. In this respect, very similar to Catani-Seymour, although based on antennas and not dipoles.
 - Method is fully analytic, but there is a degree of non-locality (azimuthal averages). Although in practice this is not a problem, it may somewhat affect performances.
 - Large number of subtraction terms.
 - Led to many phenomenological results (H+J, V+J, di-jet, VBF, DIS...).
- 2. FKS+sector decomposition [Czakon, Heymes, Caola, Röntsch, KM].
 - Basic idea: partition the phase space like in FKS, in each sector parametrize in terms of energies and angles and subtract all the divergences like in FKS.
 - Method is fully general, and can deal with massive partons easily.
 - In its original formulation, some spurious singularity was subtracted, leading to unnecessary complications. This made subtraction terms quite difficult. Their integration was done numerically.
 - Led to many phenomenological results (top pairs+decay, single top, top decay, b decay, H+J...).

Apart from antenna and FKS+sector decompositions, other techniques developed and used for phenomenological applications. For example

- 1. Projection-to-Born [Cacciari et al]. Fully local and fully analytic, but requires the knowledge of NNLO corrections inclusive over QCD radiation but differential in the Born phase space. Applied to VBF, DIS@N3LO (+with antennas)
- 2. Colorful [del Duca et al]. In principle, generic but currently only developed for $e^+ e^-$ processes. Local subtraction, but at least so far semi-numerical. Applied to $e^+ e^- \rightarrow 3j$, $H \rightarrow b\bar{b},...$

Currently, no fully validated framework having the following properties exists

- 1. fully local
- 2. fully analytic
- 3. ``simple" and completely generic

In principle, such a scheme should outperform existing results.

Apart from practical considerations, developing a framework with these features is an interesting theoretical problem in QCD. Can shed more light on the structure of soft/ collinear radiation, and shed light on fixed-order/resummation/parton shower connections.

Compared to NLO, one new feature appears at NNLO: overlapping singularities.

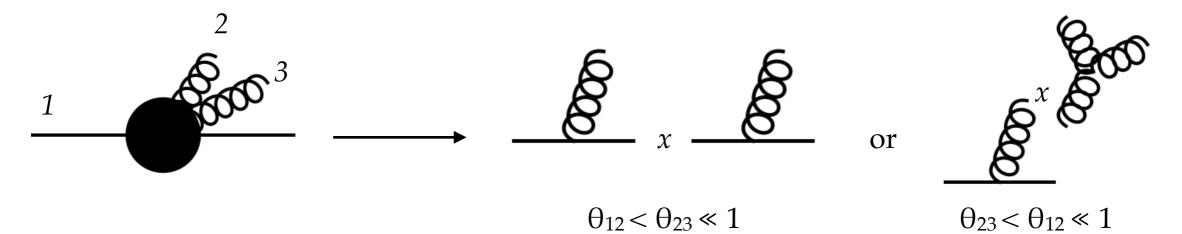
At NLO: only 2-particle invariants can vanish, $s_{ij} \sim E_j [1-\cos(\theta_{ij})] = x_E x_{\theta}$. Singularities completely factorized.

$$\int |\mathcal{M}|^2 \mathrm{d}\phi \sim \int \left[x_E^2 x_\theta |\mathcal{M}|^2 \right] \frac{\mathrm{d}x_E}{x_E^{1+2\epsilon}} \frac{\mathrm{d}x_\theta}{x_\theta^{1+\epsilon}}$$

At NNLO, 3-particle invariants can vanish, $s_{ijk} \sim \Sigma E_i E_j \cos(\theta_{ij})$. This leads to overlapping singularities. Schematically:

$$\int |\mathcal{M}|^2 \mathrm{d}\phi \sim \int \left[x_i^a x_j^b |\mathcal{M}|^2 \right] \frac{\mathrm{d}x_i}{x_i^{1+\alpha\epsilon}} \frac{\mathrm{d}x_j}{x_j^{1+\beta\epsilon}} \frac{1}{(x_i + x_j)^{\gamma\epsilon}}$$

This means that the $x_i \rightarrow 0$ limit at fixed x_j and the $x_j \rightarrow 0$ limit at fixed x_i are different. In general, this is a physical feature and not an artefact. E.g:



Overlapping singularities makes subtraction more difficult at NNLO.

In CS-like approaches (i.e. Antenna), they must be carefully reproduced by the subtraction function. It is non-trivial to devise proper simple enough subtraction function that don't contain extra spurious singularities.

In FKS-like approaches, the $x_i > x_j$ and $x_j > x_i$ cases are dealt with separately (``sector decomposition'')

Another complication of NNLO is that double real or virtual emission has in general a more complicated color structure.

Nevertheless, for double-real emission off massless particles all the non-trivial structures are still dipole-like ~ $T_i T_j$.

In (real-)virtual corrections however, 3-particle correlations f_{ijk} appear for processes involving more than three partons at the Born level.

This implies that any subtraction scheme is complete if

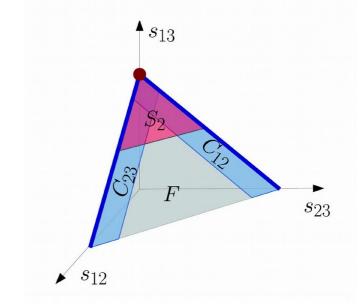
- 1. it is known for color singlet $pp \rightarrow X$ production and decay (II)
- 2. it is known for color singlet $X \rightarrow jj$ decay (FF)
- 3. it is known for DIS-like processes $p + X^* \rightarrow j + X$ (IF)
- 4. non-trivial f_{ijk} color correlations are understood

In the recent past, several new schemes have been proposed with the goal to obtain a generic, local and analytic framework for NNLO calculations.

One example is the ``geometric'' approach [F. Herzog]. It is based on the following:

- 1. identify singular regions directly in the s_{ij} space;
- 2. use this to construct a local slicing scheme in the s_{ij} space;
- 3. promote the slicing to a full subtraction.

In this framework, overlapping singularities are removed by explicitly ordering all possible limits.



A nice feature of this approach is that the integration of the countersterms is simple.

Its main drawback is that it is based on looking at the structure of individual Feynman diagrams. Every diagram is treated differently. This makes the underlying IR structure of the amplitude hidden.

Currently, this approach is at the proof-of-concept stage. So far, it has been used to reproduce the known pole-structure of the double-real correction to $H \rightarrow gg$ in pure gluodynamics ($n_f = 0$).

Another example of new schemes is the so-called ``local analytic sector subtraction'' scheme [Magnea et al].

The origina idea of this approach is to combine FKS partition with Catani-Seymour parametrization of the phase space. It is constructed in the following way:

- 1. the phase space is partitioned in different regions in a FKS-like approach. Partitions are engineered to automatically remove overlapping singularities: different orderings are treated in different partitions;
- 2. in each partition, a simple CS-like parametrization is used. This leads to very simple counterterms, whose analytic integration is straightforward.

The method is fully local and fully analytic, and combine different approaches in an interesting way.

The final result has the expected dipole-like structure, although soft radiation is not treated globally but split into the various FKS sectors.

Currently, the framework has been used to reproduce the n_f contribution to $V \rightarrow jj$ decay, although extension to the more complex initial-state case is underway.

The last new scheme I will discuss is the ``nested soft-collinear subtraction'' scheme [Caola, Röntsch, KM].

The starting point of this approach is the original FKS+sector decomposition approach. Its crucial observation is that the original approach contains overlapping singularities of the form

$$\frac{1}{E_i\theta_{ik} + E_j\theta_{jk}}$$

1

If one assumes the ordering $E_i > E_j$, such that $E_j = x_1 E_i$, and $\theta_{jk} > \theta_{ik}$, such that $\theta_{ik} = x_2 \theta_{jk}$, this becomes

$$\frac{1}{E_i \theta_{jk}} \times \frac{1}{x_1 + x_2}$$

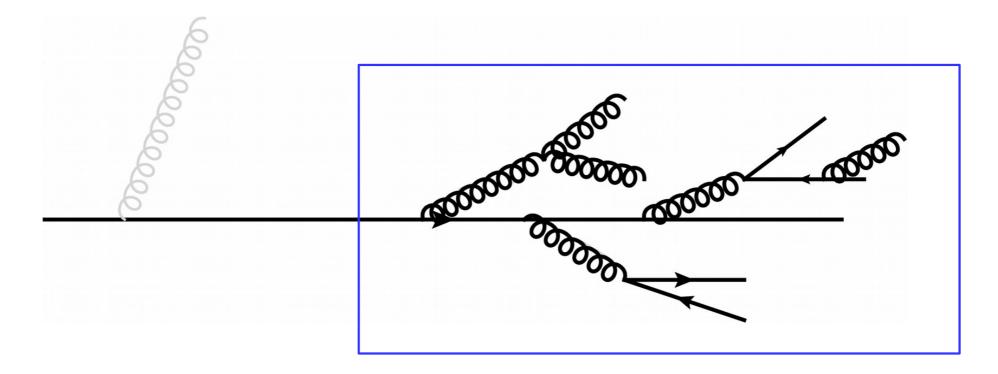
The $x_1 + x_2$ term leads to a non-trivial soft/collinear overlapping singularity.

In the original FKS+sector decomposition approach (cf. STRIPPER), this was dealt with by sector decomposition like all the other overlapping singularities.

However, this overlap is unphysical.

Indeed, such an overlap would violate color coherence, which states that for any physical quantity

- 1. soft gluons are only sensitive to the charges of the system;
- 2. if a bunch of partons become collinear, soft gluons can only resolve the total color charge of the collinear system.



It is easy to show that this implies that no non-trivial soft/collinear overlaps can appear in physical quantities.

They appear at intermediate stages in the original FKS+sector decomposition approach because singularities were identified by looking at individual propagator rather than at on-shell gauge-invariant matrix elements.

In the nested soft-collinear subtraction scheme, the non-physical soft collinear overlaps are removed.

As a consequence, soft and collinear singularities are completely factorized, and can be dealt with independently one by one, in a nested way.

Specifically, the subtraction proceeds as follows.

1. all soft singularities are subtracted globally, without introducing any sector.

The soft counterterms are simple enough that they can be analytically integrated for arbitrary (massless) processes [Caola, Delto, Frellesvig, KM (2018)];

2. FKS partitions are introduced to separate different collinear regions. In each partitions, (physical) collinear overlaps are removed using sector decomposition.

Also in this case, the counterterms are very simple and have been analytically integrated for all possible cases [Delto, Melnikov (2019)].

At the end, the singularity structure is very transparent, and relatively compact. All the singularities (double-soft/single-soft/triple-collinear/double-collinear) are subtracted independently, one after the other, in a nested way.

This scheme has several attracting features:

- it is fully analytic, fully local and it can in principle be applied to arbitrary processes;
- it is in some sense ``minimal", i.e. it contains the minimal number of subtraction needed to regulate physical QCD amplitudes. This may become relevant for high multiplicity processes;
- it is flexible. Contrary to the original FKS, is not tied to a particular parametrization. Currently, the original FKS+sector decomposition parametrization is used for convenience, but one could explore different parametrizations.

Similarly, overlapping singularities do not necessarily require sector decomposition, but could be dealt with using different approaches.

For example, they could be dealt with at the FKS-partition level, in which case this formalism would become very similar to the local analytic subtraction one.

Currently, the scheme has been fully validated for initial state production [Caola, Röntsch, KM (2019)] and final state production [Caola, Delto, Röntsch, KM, arXiv:1906.xxxx].

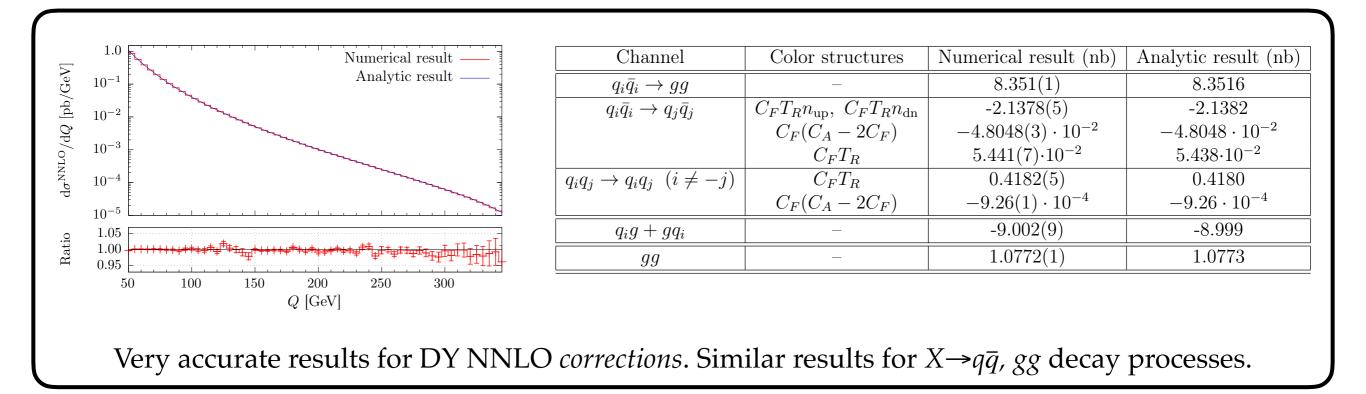
Results for the initial-final case are known, and are currently under validation [Asteriadis, Caola, Delto, Röntsch, KM, arXiv:19xx.xxxx].

Apart from theoretical considerations, from a purely practical point of view, a good NNLO framework should be:

- 1. reliable. Corrections should be computable to arbitrary accuracy;
- 2. efficient.

It is reasonable to expect that at fully local fully analytic scheme could fulfil these requirements. Within the nested soft-collinear scheme, one can test it in simple processes.

1. *reliable* → pick a process where analytic results are known, and compare to very high accuracy



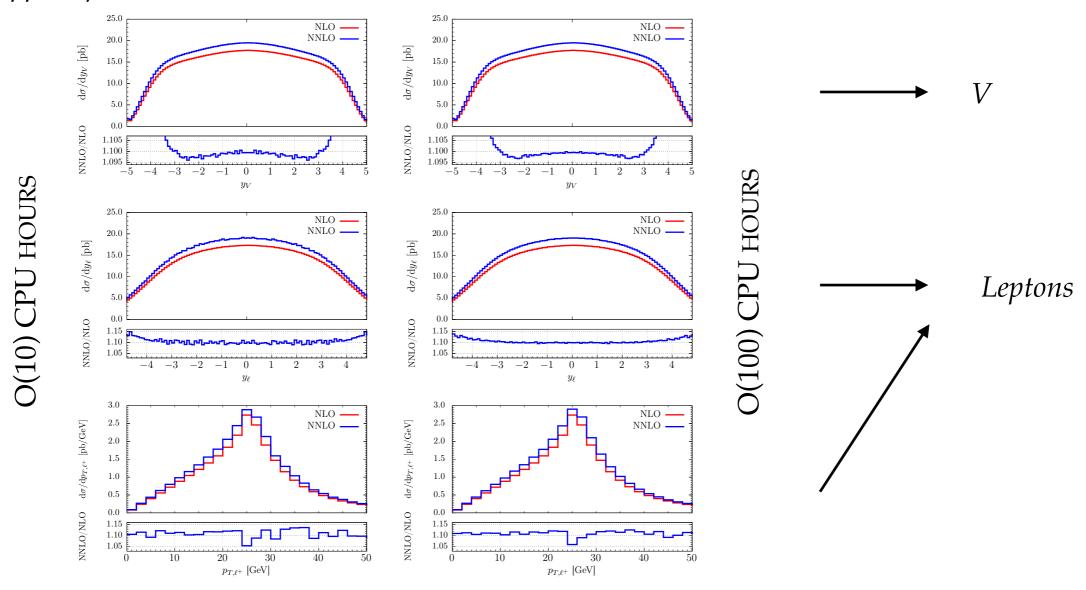
2. *efficient*: Higgs and DY production in 1 hour on a standard laptop (1 core)

Higgs total cross section:

$$\sigma_{\rm H}^{\rm LO} = 15.42(1) \text{ pb}; \quad \sigma_{\rm H}^{\rm NLO} = 30.25(1) \text{ pb}; \quad \sigma_{\rm H}^{\rm NNLO} = 39.96(2) \text{ pb}.$$

 $pp \rightarrow 2l$, symmetric cuts:
 $\sigma_{\rm DY}^{\rm LO} = 650.4 \pm 0.1 \text{ pb}; \quad \sigma_{\rm DY}^{\rm NLO} = 700.2 \pm 0.3 \text{ pb}; \quad \sigma_{\rm DY}^{\rm NNLO} = 734.8 \pm 1.4 \text{ pb}.$

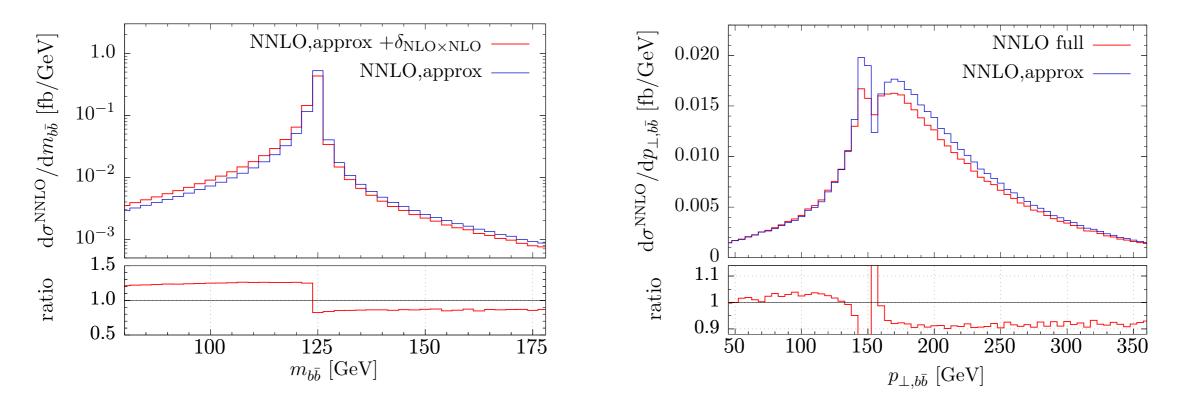
 $pp \rightarrow \gamma^* \rightarrow 2l$, differential distributions [old semi-numerical code]



Efficiency will become more and more important for complex final states. Phenomenological applications inversely proportional to the time needed to obtain results: more complex result, and richer phenomenological structure (observables...).

A ``simple'' complex problem: VH, $H \rightarrow b\bar{b}@NNLO$.

Although this process is by far simpler than a genuine process with color in the initial and final states, it can give us a rough idea of performances in more complex scenarios.

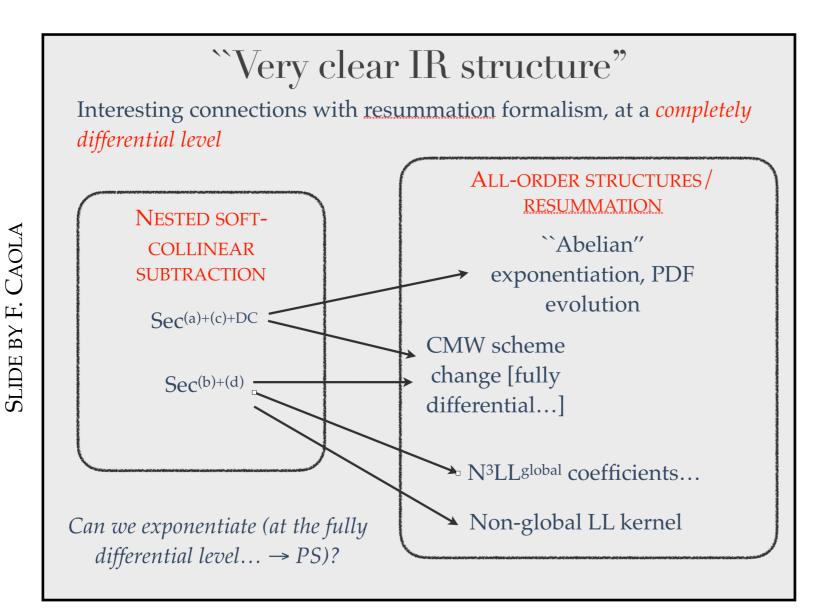


Stable result with reasonably small computing resources.

Apart from their theoretical appeal, ``simple'' subtraction schemes can also shed more light on the IR structure of perturbative QCD.

Can be informed by resummation, and provide interesting fully differential information to them.

Better understanding of subtraction schemes could also lead to more physical matching/ merging schemes and perhaps even more accurate showers.



So far, discussion mostly focused on finding ``a'' subtraction scheme, but interest in a better understanding of the underlying IR structure (see e.g. [Magnea et al (2018)]) NNLO calculations are at the core of the precision program at the LHC.

They require multi-loop amplitude, and efficient subtraction schemes.

In the recent past, there has been a lot of developments on both.

In particular, we now have several subtraction schemes for fully differential NNLO calculations. Some of them are in principle generic (*Sector decomposition+FKS, antenna, Jettiness slicing*).

As processes become more and more complex, very efficient subtraction schemes are required. So far, an optimal NNLO framework has not yet emerged.

Several proposals for better schemes in the recent past (*geometric, local analytic subtraction, nested subtraction*). They are very promising, but not yet fully developed.

Developing local and analytic subtraction schemes is also an interesting theoretical problem in QCD.

A clear organization of soft/collinear information is needed for fixed-order, resummation and parton shower approaches. ``Clean'' NNLO formalisms seem in direct correspondence with resummation approaches. They are not all-orders, but are fully local. Perhaps, one can learn from both approaches to improve on them.