Stefano Frixione

On simulations at e^+e^- colliders

ZPW2024, Zurich, 9/1/2024

For the majority of tasks, the standard strategy at e^+e^- colliders relied on semi-analytical, "parton"-level, high-accuracy, inclusive, process-specific results (*evaluations*)

To connect these with data, the usage of numerical, "particle"-level, exclusive programs (*simulations*), either process-specific or not, is unavoidable: effects not accounted for by evalutations must be deconvoluted

- In the infinite-precision limit, this strategy cannot work: non-factorizable effects are always relevant. Thus, it may be extremely problematic also in a realistic large-accuracy scenario
- Regardless of that, a result is as good as its least good component (excellent evaluations are worthless if employed with poor simulations)

My conclusion, informed in part by our collective experience at hadron colliders, is to:

Try and embed into simulations as much parton-level perturbative information as possible

This will limit the amount of deconvolutions (if any). Evaluations may or may not be relevant as such (the underpinning computations will)

Henceforth, I'll therefore focus on simulations

My take on the current status of simulation codes, and their underlying computations

(see e.g. 2203.12557, a contributing review to the latest Snowmass)

Shortest summary:

- Status is heterogeneous
- Readiness generally low (e.g. the codes run but do not meet precision targets)

A slightly longer summary:

- QED-specific theoretical bases are well established, and mostly already employed at LEP. However, conceptual and/or technical progress is still needed, and is being pursued

- The "recycling" of QCD perturbative techniques is limited so far; potential for growth in the future (e.g. matrix element computations, use of EFTs, collinear resummation techniques)

- Dedicated high-precision tools not precise enough

- Modern multi-purpose tools, such as PSMCs, generally poorly tested in e^+e^- high-energy environments

Examples of simulations tools and tools that underpin simulations:

Multi-purpose tools:

- Pythia8
- Herwig
- Sherpa
- MadGraph5_aMC@NLO
- Whizard

Dedicated tools:

- BabaYaga
- RacoonWW, Racoon4f
- KKMC-ee, KORAL[W/Z], BH[LUMI/WIDE], YSF[WW3/ZZ]

Examples of simulations tools and tools that underpin simulations:

Multi-purpose/dedicated tools:

All tools aim to give realistic descriptions of physical observables, and thus include some form of ISR/FSR resummation

For the latter, all tools adopt collinear factorisation bar for KKMC-ee, KORAL*, BH*, YSF*, and one instance of Sherpa, that adopt YSF

Among multi-purpose tools:

- Emphasis on small angle/small energy radiation (loosely speaking: parton shower, although not really [only] such): Pythia8, Herwig, Sherpa
- Emphasis on matrix element computations: MadGraph5_aMC@NLO, Whizard

At the LHC nowadays there is a strict interplay between MEGs and PSMCs. This is essentially absent thus far in e^+e^- (as far as QED radiation is concerned)

Presently, the main difference for particles branching off ISR is:
PSMCs are exclusive in them, MEGs (semi-)inclusive

Consider a generic cross section, sufficiently inclusive:

$$\sigma = \alpha^b \sum_{n=0}^{\infty} \alpha^n \sum_{i=0}^{n} \sum_{j=0}^{n} \varsigma_{n,i,j} L^i \ell^j$$

This is symbolic, and only useful to expose the presence of:

$$\ell = \log \frac{Q^2}{\langle E_\gamma \rangle^2}\,, \qquad L = \log \frac{Q^2}{m^2}$$

Numerology: consider the production of $Z \to ll$ at:

- $\sqrt{Q^2} = m_Z$

$$\begin{aligned} L = 24.18 \quad &\Longrightarrow \quad \frac{\alpha}{\pi} L = 0.06 \\ 0 \leq m_{ll} \leq m_Z\,, \quad \ell = 6.89 \quad &\Longrightarrow \quad \frac{\alpha}{\pi} \ell = 0.017 \\ m_Z - 1\ \mathrm{GeV} \leq m_{ll} \leq m_Z\,, \quad \ell = 10.60 \quad &\Longrightarrow \quad \frac{\alpha}{\pi} \ell = 0.026 \end{aligned}$$

Consider a generic cross section, sufficiently inclusive:

$$\sigma = \alpha^b \sum_{n=0}^{\infty} \alpha^n \sum_{i=0}^{n} \sum_{j=0}^{n} \varsigma_{n,i,j} L^i \ell^j$$

This is symbolic, and only useful to expose the presence of:

$$\ell = \log \frac{Q^2}{\langle E_\gamma \rangle^2}\,, \qquad L = \log \frac{Q^2}{m^2}$$

Numerology: consider the production of $Z \to ll$ at:

- $\sqrt{Q^2} = 500$ GeV

$$L = 27.59 \quad \Longrightarrow \quad \frac{\alpha}{\pi} L = 0.069$$

$$0 \le m_{ll} \le m_Z\,, \quad \ell = 1.449 \quad \Longrightarrow \quad \frac{\alpha}{\pi} \ell = 0.0036$$

$$m_Z - 1 \text{ GeV} \le m_{ll} \le m_Z\,, \quad \ell = 1.453 \quad \Longrightarrow \quad \frac{\alpha}{\pi} \ell = 0.0036$$

It takes a lot of brute force (i.e. fixed-order results to some $\mathcal{O}(\alpha^n)$) to overcome the enhancements due to L and ℓ.

It is always convenient to first improve by means of factorisation formulae:

$$d\sigma(L,\ell) = \mathcal{K}_{soft}(\ell;L)\beta(L)d\mu \quad (1)$$

$$= \mathcal{K}_{coll}(L;\ell) \otimes d\hat{\sigma}(\ell) \quad (2)$$

Use of:

(1) YFS (resummation of ℓ)

(2) collinear factorisation (resummation of L)

Common features: $\mathcal{K}$ is an *all-order* universal factor; β and $d\hat{\sigma}$ are process-specific and computed order by order

(still brute force, but to a lesser extent)

YFS

Aim: soft resummation for:

$$\left\{ e^+(p_1) + e^-(p_2) \longrightarrow X(p_X) + \sum_{i=0}^{n} \gamma(k_n) \right\}_{n=0}^{\infty}$$

Achieved with:

$$\begin{aligned} d\sigma(L,\ell) &= \mathcal{K}_{soft}(\ell;L)\beta(L)d\mu \\ &= e^{Y(p_1,p_2,p_X)} \sum_{n=0}^{\infty} \beta_n\left(\mathcal{R}p_1, \mathcal{R}p_2, \mathcal{R}p_X; \{k_i\}_{i=0}^{n}\right) d\mu_{X+n\gamma} \end{aligned}$$

This is symbolic, and stands for both the EEX and CEEX approaches [hep-ph/0006359 Jadach, Ward, Was] that build upon the original YFS work [Ann.Phys.13(61)379]

EEX: exclusive (in the photons) exponentiation, matrix element level

CEEX: coherent exclusive (in the photons) exponentiation, amplitude level, including interference

YFS

Aim: soft resummation for:

$$\left\{ e^+(p_1) + e^-(p_2) \longrightarrow X(p_X) + \sum_{i=0}^{n} \gamma(k_n) \right\}_{n=0}^{\infty}$$

Achieved with:

$$d\sigma(L,\ell) \quad = \quad e^{Y(p_1,p_2,p_X)} \sum_{n=0}^{\infty} \beta_n \left(\mathcal{R}p_1, \mathcal{R}p_2, \mathcal{R}p_X; \{k_i\}_{i=0}^{n} \right) d\mu_{X+n\gamma}$$

- Y essentially universal (process dependence only through kinematics); resums ℓ
- The soft-finite β_n are process-specific, and are constructed by means of local subtractions involving matrix elements and eikonals (i.e. *not* BN)

$$\beta_n = \alpha^b \sum_{i=0}^{n} \alpha^i \sum_{j=0}^{i} c_{n,i,j} L^j$$

- For a given n, matrix elements have different multiplicities, hence the need for the kinematic mapping $\mathcal{R}$

Collinear factorisation

Aim: collinear resummation for:

$$\left\{ k(p_k) + l(p_l) \longrightarrow X(p_X) + \sum_{i=0}^{n} a_i(k_n) \right\}_{n=0}^{\infty} \qquad a_i = e^{\pm}, \gamma \ldots$$

with initial-state particles stemming from beams:

$$(k,l) = (e^+, e^-)\,, \quad (k,l) = (e^+, \gamma)\,, \quad (k,l) = (\gamma, e^-)\,, \quad (k,l) = (\gamma, \gamma)\,, \ldots$$

Master formula:

$$\begin{aligned} d\sigma(L,\ell) &= \mathcal{K}_{coll}(L;\ell) \otimes d\hat{\sigma}(\ell) \\ \longrightarrow \quad d\sigma_{kl} &= \sum_{ij} \int dz_+ dz_- \, \Gamma_{i/k}(z_+, \mu^2, m^2)\, \Gamma_{j/l}(z_-, \mu^2, m^2) \\ &\qquad \times d\hat{\sigma}_{ij}(z_+ p_k, z_- p_l, \mu^2; p_X, \{k_i\}_{i=0}^{n}) \end{aligned}$$

- $\Gamma_{\alpha/\beta}$ universal (the PDF); resums L
- The collinear-finite $d\hat{\sigma}_{ij}$ are process-specific, and are the standard short-distance matrix elements, constructed order by order (*with* BN). May or may not include resummation of other large logs (including ℓ)

YFS vs collinear factorisation

Both are systematically improvable in perturbation theory:
in YFS the β_n's (fixed-order), in collinear factorisation both the PDFs (logarithmic accuracy) and the $d\hat{\sigma}$'s (fixed-order, resummation)

+ YFS: very little room for systematics. Exceptions are the kinematic mapping $\mathcal{R}$, and the quark masses (when the quarks are radiators). Renormalisation schemes??

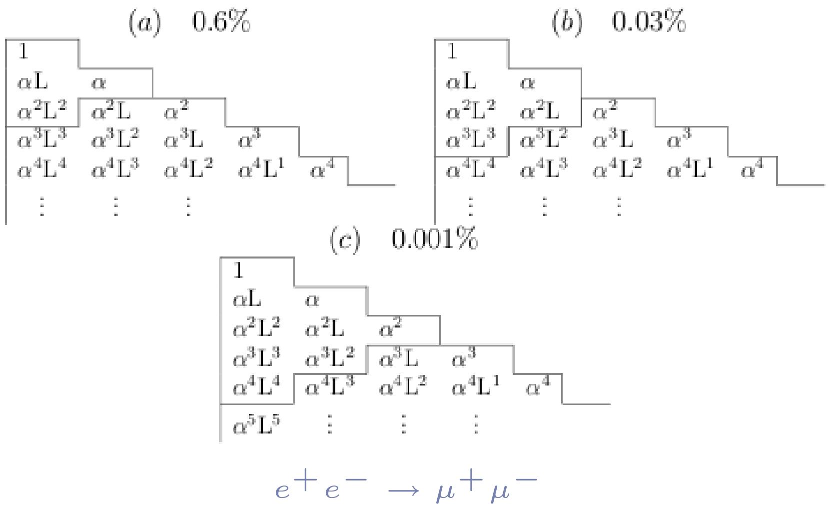
− Collinear factorisation: systematic variations much larger. At the LL (used in phenomenology so far) a rigorous definition of uncertainties is impossible (parameters are arbitrary), and comparisons with YFS are largely fine tuned

− YFS: the computations of β_n are not standard (EEX) and highly non-trivial (CEEX)

+ Collinear factorisation: the computations of $d\hat{\sigma}_{ij}$ are standard

[an aside

YFS and collinear factorisation differ on naming conventions !



1903.09895, Jadach&Skrzypek

- $\mathcal{O}(\alpha L)$ and $\mathcal{O}(\alpha^2 L)$ in YFS *typically* mean the corresponding coefficients in the β_n terms
- $\mathcal{O}(\alpha L)$ and $\mathcal{O}(\alpha^2 L)$ in collinear factorisation mean the whole LL and NLL towers, respectively

]

Folk wisdom about collinear factorisation: while YFS is exclusive in the photons (*true* in EEX and CEEX), collinear factorisation is inclusive (*not true* in general)

Folk wisdom about collinear factorisation: while YFS is exclusive in the photons (*true* in EEX and CEEX), collinear factorisation is inclusive (*not true* in general)

Firstly, one is always exclusive in the photons (possibly) emerging from the hard process ($d\hat{\sigma}_{ij}$)

Secondly, whether one is exclusive also in the photons associated with ISR depends on the implementation of the factorisation formula

- MC integration as is: inclusive (or modelled as e.g. in Whizard)
- Integration through recursion (e.g. parton shower): exclusive. Examples: Pythia8, Herwig, Sherpa, Babayaga

Folk wisdom about collinear factorisation: while YFS is exclusive in the photons (*true* in EEX and CEEX), collinear factorisation is inclusive (*not true* in general)

Open problems in *precise*, exclusive e^+e^- simulations: extend QCD matching techniques (MC@NLO, Powheg) to ISR QED*; extend logarithmic accuracy of shower to NLL; extend matching beyond NLO

Before these can be considered $\longrightarrow$

*LL solutions in pp collisions in HORACE [see e.g. hep-ph/0609170] and Powheg [see e.g. 1302.4606]

All physics simulations based on collinear factorisation done so far are based on a LL-accurate picture

This is not tenable at high energies/high statistics:

- accuracy is insufficient (see e.g. W^+W^- production)
- systematics not well defined

Fortunately, the upgrade of PDFs from LL to NLL accuracy is now established

z-space LO+LL PDFs $(\alpha \log(Q^2/m^2))^k$:

$\sim$ 1992

- $0 \leq k \leq \infty$ for $z \simeq 1$ (Gribov, Lipatov)
- $0 \leq k \leq 3$ for $z < 1$ (Skrzypek, Jadach; Cacciari, Deandrea, Montagna, Nicrosini; Skrzypek)
- matching between these two regimes
- for e^-

z-space NLO+NLL PDFs $(\alpha \log(Q^2/m^2))^k + \alpha\,(\alpha \log(Q^2/m^2))^{k-1}$:

$\longrightarrow$ 1909.03886, 1911.12040, 2105.06688, 2207.03265 (Bertone, Cacciari, Frixione, Stagnitto, Zaro, Zhao)

- $0 \leq k \leq \infty$ for $z \simeq 1$
- $0 \leq k \leq 3$ for $z < 1$ $\Longleftrightarrow$ $\mathcal{O}(\alpha^3)$
- matching between these two regimes
- for e^+, e^-, γ, and light quarks
- both numerical and analytical
- factorisation schemes: $\overline{\mathrm{MS}}$ and Δ (that has DIS-like features)

Bear in mind that PDFs are fully defined only after adopting a definite *factorisation scheme*, which is the choice of the finite terms associated with the subtraction of the collinear poles

- 1911.12040 $\longrightarrow$ $\overline{\mathrm{MS}}$
- 2105.06688 $\longrightarrow$ a DIS-like scheme (called Δ)

At variance with the QCD case, there is also an interesting *renormalisation-scheme* dependence of QED PDFs

Conclusions that have emerged since ~ 2022:

- The inclusion of NLL contributions into the electron PDF has an impact of $\mathcal{O}(1\%)$ (precise figures are observable and renormalisation-scheme dependent)
- This estimate does not include the effects of the photon PDF
- The comparison between $\overline{\mathrm{MS}}$- and Δ-based results shows differences compatible with non-zero $\mathcal{O}(\alpha^2)$ effects, as expected

  (but: these are potentially *large in the soft region*)
- Renormalisation-scheme dependence is of $\mathcal{O}(0.5\%)$

If the target is a $10^{-\mathrm{some\ large\ number}}$ relative precision, these effects must be taken into account

While they emphasise the role of different logs, even if not tuned YFS and collinear factorisation will converge towards each other by increasing the number of terms relevant to the “other” logs

- The standard strategy is to increase accuracy order by order
- However, resummations are being pursued. See e.g.
  - 2303.14260 (Jadach, Ward, Was) for collinear logs in YFS
  - 2105.06688 (SF) for soft logs in collinear factorisation

LHC legacy: the power of automation

$\sqrt{s}$ [GeV]	$\sigma(e^+e^- \to q\bar{q})$ [pb]	$\sigma(e^+e^- \to W^+W^-)$ [pb]	$\sigma(e^+e^- \to ZH)$ [pb]	$\sigma(W^+W^- \to H)$ [pb]	$\sigma(e^+e^- \to t\bar{t})$ [pb]
20.0	898.8	-	-	-	-
30.0	434.6	-	-	-	-
40.0	259.9	-	-	-	-
50.0	182.1	-	-	-	-
60.0	153.0	-	-	-	-
70.0	177.7	-	-	-	-
80.0	423.9	-	-	-	-
88.0	3891.0	-	-	-	-
91.2	29250.0	-	-	-	-
94.0	8953.0	-	-	-	-
125.0	417.9	-	-	-	-
157.5	177.4	-	-	-	-
162.5	162.0	-	-	-	-
165.0	155.2	8.773	-	0.00021	-
217.0	-	17.63	0.04278	0.004497	-
240.0	-	16.62	0.1998	0.005859	-
350.0	-	11.57	0.1306	0.024613	0.3771
360.0	-	11.22	0.1236	0.027064	0.5534

Cross-sections have been computed with MADGRAPH5_AMC@NLOv3.5.0 [1, 6], exploiting the recent developments for lepton colliders [9, 3]. In particular, ISR partonic densities with NLL-accurate evolution [7, 2, 3] have been employed, using the so-called Δ factorisation scheme [8]. All cross-sections include NLO EW and QCD corrections (the latter only when relevant), with the exception of $e^+e^- \to q\bar{q}$ where only QCD corrections are computed. NLO EW corrections are computed in the G_μ scheme; all fermions, with the exception of the top quark, are considered massless. Contributions from photons in the initial state are included whenever NLO EW corrections are computed. It is worth to note that the only processes where initial-state photons contribute at the LO are $e^+e^- \to t\bar{t}$ and $e^+e^- \to W^+W^-$. The following parameters have been employed:

$$m_t = 173.33\,\text{GeV}\,,\ \text{m}_\text{W} = 80.419\,\text{GeV}\,,\ \text{m}_\text{Z} = 91.189\,\text{GeV}\,,\ \text{m}_\text{H} = 125\,\text{GeV}\,,\ \text{G}_\mu = 1.166 \times 10^{-5}\,\text{GeV}^{-2}\,. \quad (1)$$

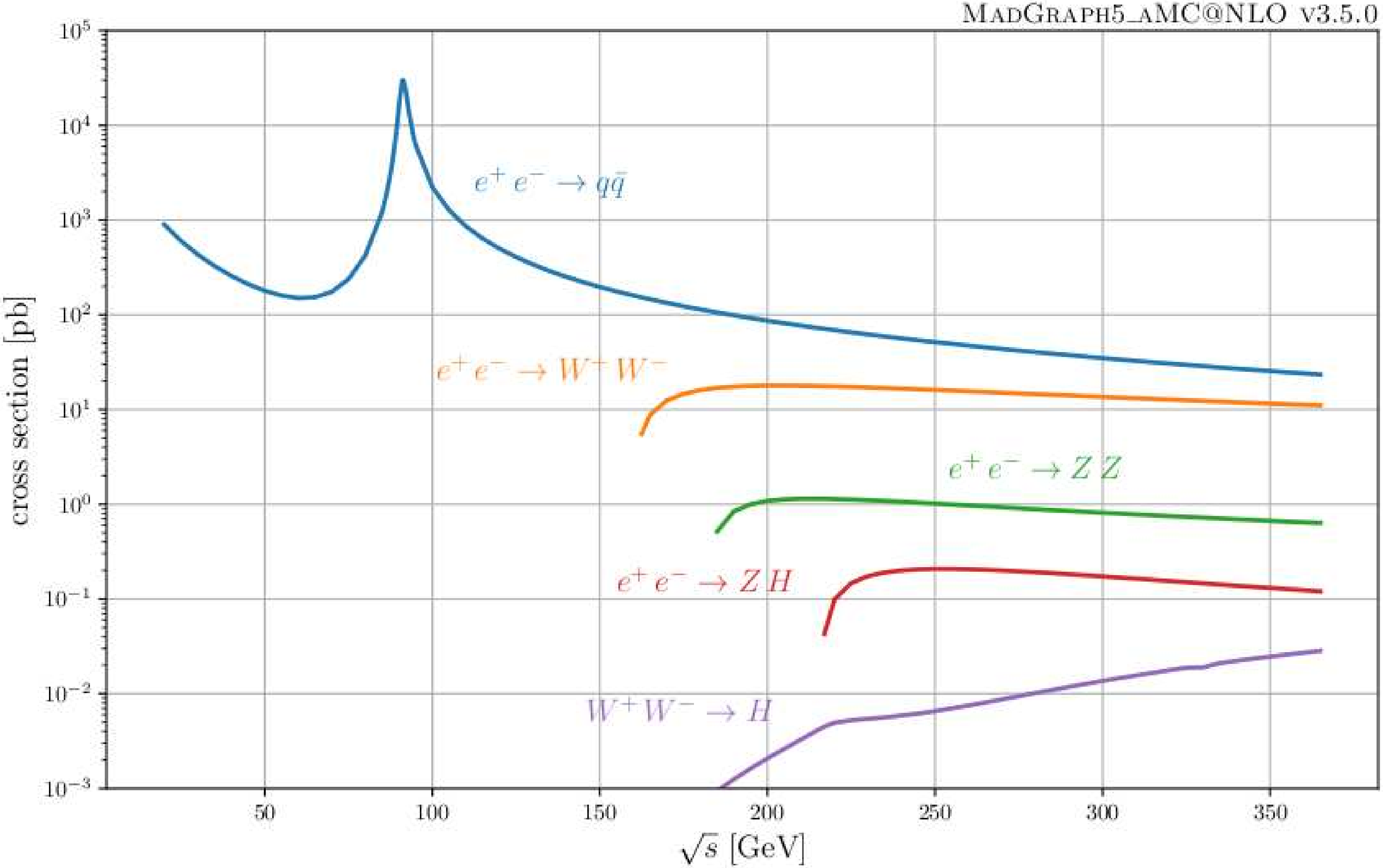
The cross section for Higgs production in VBF has been obtained as the difference of the cross sections for the processes $e^+e^- \to H\nu_e\bar{\nu}_e$ and $e^+e^- \to H\nu_\mu\bar{\nu}_\mu$, both computed at NLO EW accuracy in the complex mass scheme [4, 5]. For these cross sections, the following non-zero widths have been employed:

$$\Gamma_t = 1.3776\,\text{GeV}\,,\quad \Gamma_\text{W} = 2.093\,\text{GeV}\,,\quad \Gamma_\text{Z} = 2.499\,\text{GeV}\,, \quad (2)$$

MG5_aMC@NLO, EW(+QCD) NLO accurate results, NLL PDFs

A few days of work (Selvaggi, Zaro)

LHC legacy: the power of automation



MG5_aMC@NLO, EW(+QCD) NLO accurate results, NLL PDFs

A few days of work (Selvaggi, Zaro)

Is this sufficient in collinear factorisation?

Not quite

- What was done at the NLL gives one a blueprint to go to NNLL, if need be. Most of the ingredients are available from QCD, but one still has to figure out the $z \to 1$ behaviour analytically

- In an orthogonal direction, one must achieve an exclusive generation, at the desired logarithmic accuracy

Exclusive means the ability to retain the information on the dof's of the particles stemming from the (ISR) branchings that do not enter the hard process

- Well established within YFS; not so much within collinear factorisation

- We cannot blindly apply MC@NLO or Powheg: hadron and lepton PDFs have dramatically different behaviours

- Besides, there is currently no NLL-accurate ISR hadronic shower

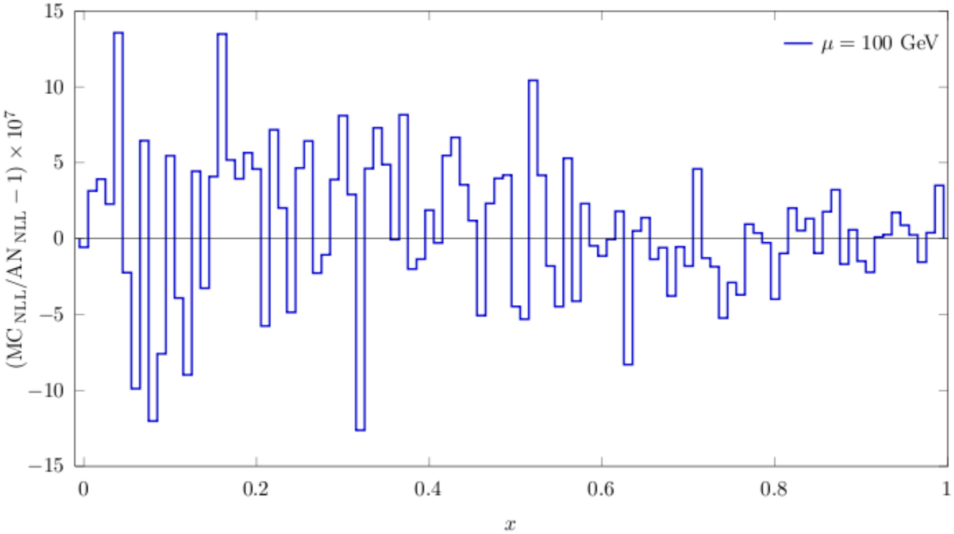
A possible approach: follow BabaYaga (Carloni Calame, Montagna, Nicrosini, Piccinini)

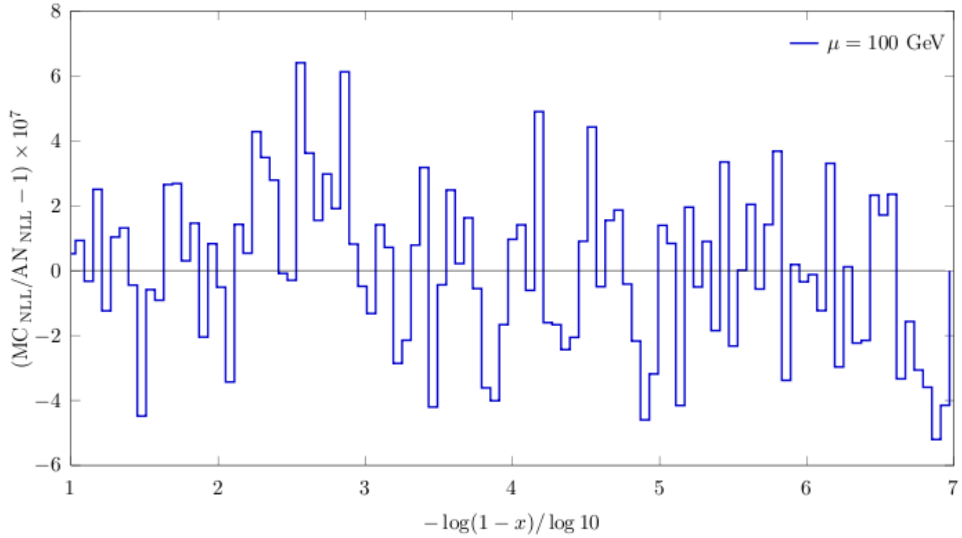
- α is small
- Thus, resumming to all orders is not that different w.r.t. to summing to a fairly high order (say, ~ 15)
- First step: write the PDFs as recursive, MC-compatible, solutions of the evolution equations, whose individual contributions can be associated with events (i.e. with given number and types of branchings)

This now works for the *non-singlet* at the *NLL* accuracy

(Carloni Calame, Frixione, Montagna, Piccinini, Stagnitto)

MC vs analytical





This is the fractional difference between the known PDFs and those generated exclusively

Agreement of $\mathcal{O}(10^{-7})$ up to $z \simeq 1 - 10^{-10}$ (cutoff $\epsilon = 10^{-14}$)

This is NLL Δ; NLL $\overline{\mathrm{MS}}$ and LL are analogous

Finally...

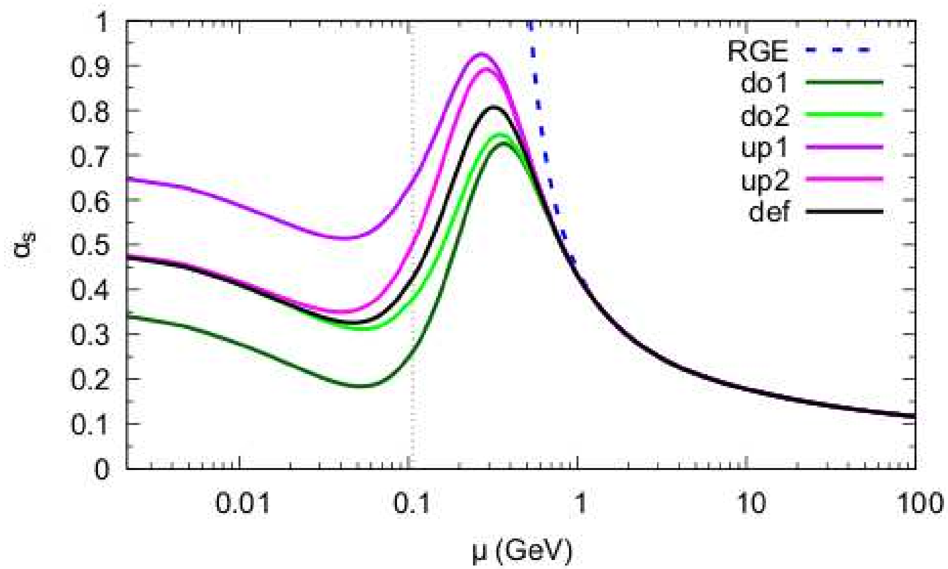
At a certain level of precision, the impact of the *strongly-interacting* partonic content of the incoming leptons cannot be ignored

In YFS, the corresponding contributions enter the β_n terms; in collinear factorisation, they entail the presence of quarks and gluon PDFs

- I'm not aware of attempts to address this issue in YFS
- In collinear factorisation there are now two different approaches, applied so far to the PDFs of the muon (the case of the electron is conceptually identical)

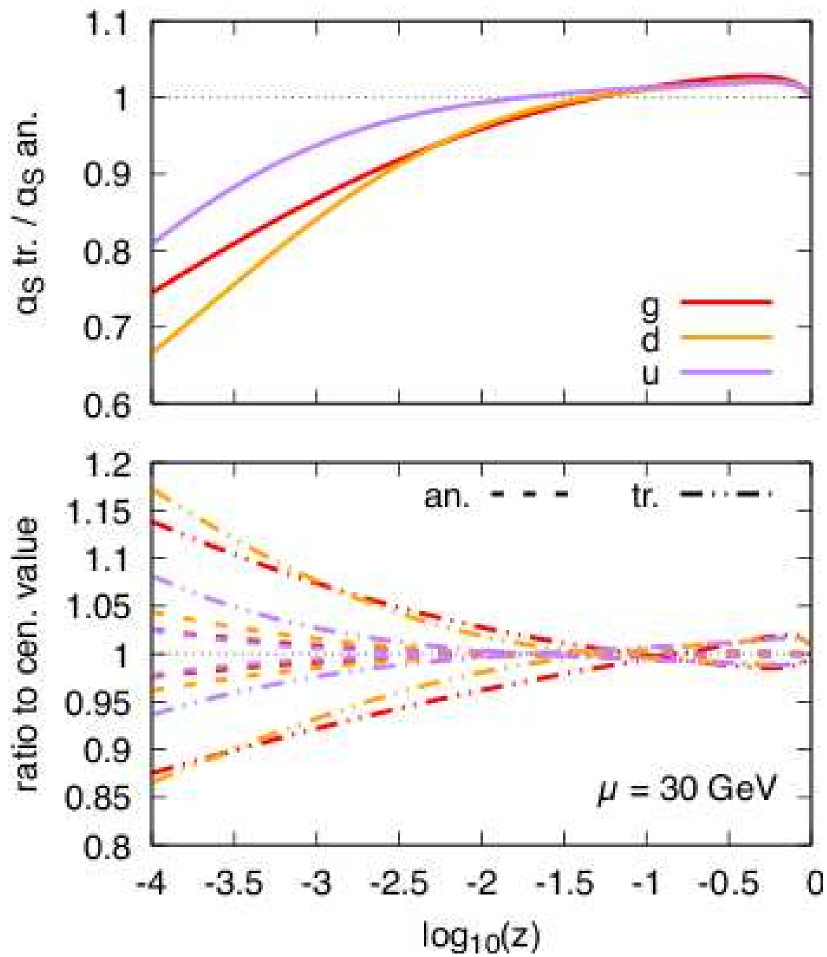
The quark and gluon PDFs force one to consider α_s in the infrared

- 2103.09844 (Han, Ma, Xie) and 2303.16964 (Garosi, Marzocca, Trifinopoulos) bypass the problem by setting $\alpha_s(\mu) = 0$ for $\mu < Q_0$, with $Q_o = \mathcal{O}(0.5 \text{ GeV})$
  ("truncated" approach)
- 2309.07516 (SF, Stagnitto) adopts a parametrisation of α_s in the infrared motivated by dispersion relations (Webber; Dokshitzer, Marchesini, Webber)
  ("analytical" approach)



In the case of the muon PDFs, there are significant differences between the truncated and the analytical approaches

I expect these to be potentially even larger for electron PDFs



PDFs ratios and their uncertainties

In the case of the muon PDFs, there are significant differences between the truncated and the analytical approaches

I expect these to be potentially even larger for electron PDFs

$\sigma(p_T^{cut} = 10$ GeV) [pb]	an.	tr.
$\mathcal{O}(\alpha_S^2)$	$18.33^{+1.30\%}_{-1.25\%}$	$15.00^{+10.23\%}_{-10.99\%}$
γ-ind.	$8.24^{+0.68\%}_{-0.91\%}$	$7.56^{+3.71\%}_{-3.75\%}$
total	$26.58^{+1.11\%}_{-1.15\%}$	$22.57^{+8.04\%}_{-8.56\%}$

Table 4: Total dijet rates for $p_T^{cut} = 10$ GeV, in pb.

$\sigma(p_T^{cut} = 100$ GeV) [fb]	an.	tr.
$\mathcal{O}(\alpha_S^2)$	$41.38^{+0.03\%}_{-0.03\%}$	$41.17^{+0.46\%}_{-0.85\%}$
γ-ind.	$90.03^{+0.01\%}_{-0.02\%}$	$89.67^{+0.24\%}_{-0.32\%}$
total	$136.91^{+0.00\%}_{-0.00\%}$	$136.35^{+0.28\%}_{-0.48\%}$

Table 5: Total dijet rates for $p_T^{cut} = 100$ GeV, in fb.

Dijet cross sections at 10 TeV

In the case of the muon PDFs, there are significant differences between the truncated and the analytical approaches

I expect these to be potentially even larger for electron PDFs

- I believe there are several issues with the truncated approach. We'll have to further the studies on this, as well as to consider the case of the electron PDFs

Conclusions

There are several interesting open problems relevant to simulations in e^+e^- collisions. To name a few:

- Automation in YFS and beyond-NLO in collinear factorisation
- Fully-exclusive N^{k}LO generation in collinear factorisation
- Resummations of collinear logs in YFS and of soft logs in collinear factorisation
- Role of strongly-interacting partons in leptons

It is difficult for me *not* to reach the conclusion that, at future colliders, simulations (as opposed to semi-analytical codes) will play an even more prominent role than at LEP