Automation of fast NLO calculations for global PDF analyses

Juan Rojo
CERN, PH Division, TH Unit

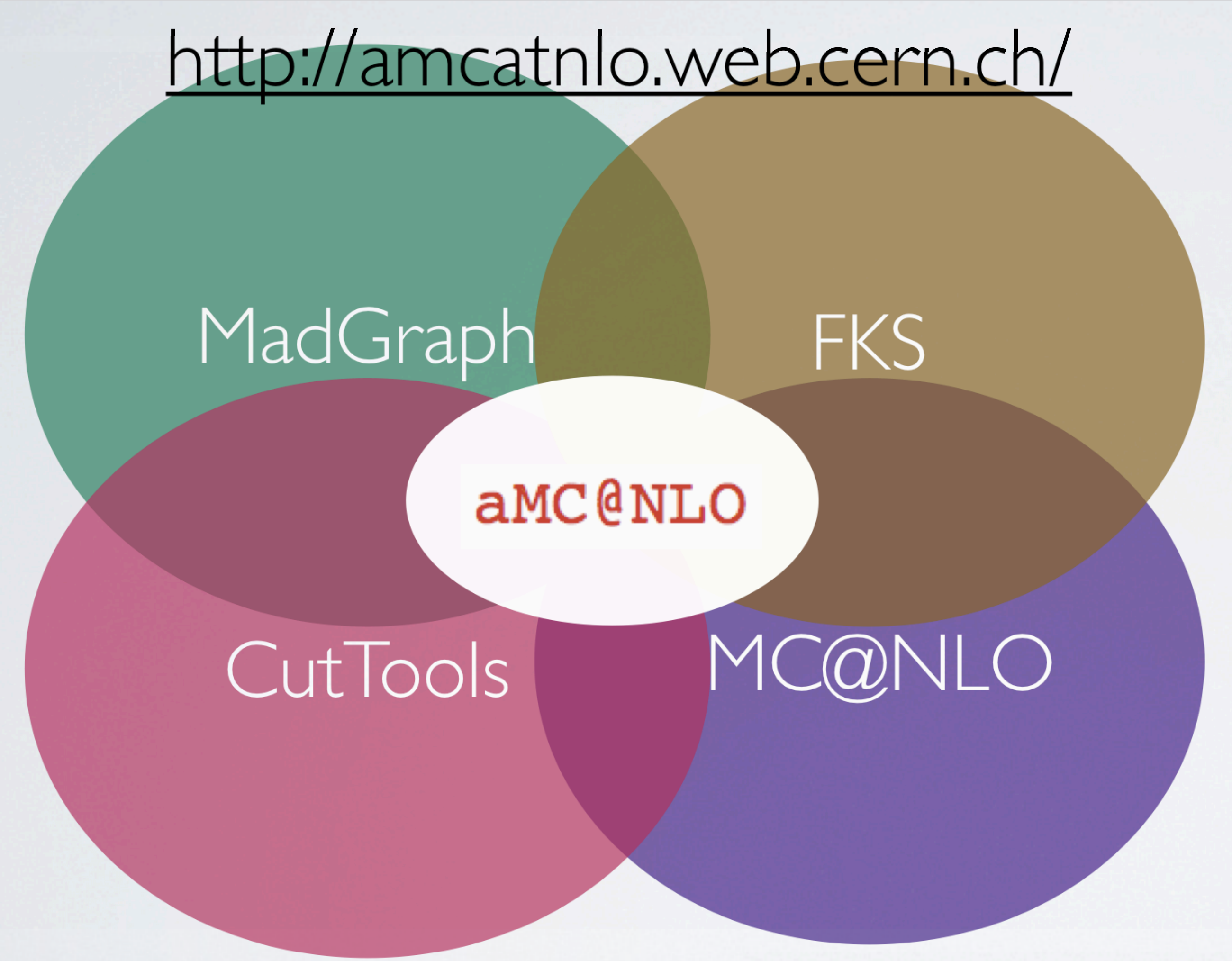
In collaboration with:
Valerio Bertone, Rikkert Frederix, Stefano Frixione, and Mark Sutton

HERAfitter User's Meeting
CERN, 10/12/2013

NLO Calculations in PDF analyses

- **(N)NLO QCD calculations** are too CPU-time intensive to be used directly into PDF analysis
- The traditional solution, **LO supplemented by iterative bin-by-bin K-factors**, is not suitable in general to match the **precision of LHC data**
- In the recent years, various approaches have been proposed to provide **fast interfaces to NLO calculations**, that can be used directly in PDF analysis, the main ones being:
  - **APPLgrid**: interfaced to **MCFM, NLOJet++** and **DYNNLO**
  - **FastNLO**: interfaced to **NLOJet++**
- Basic strategy: **interpolate PDFs** in a suitable basis, and **precompute the partonic cross-section** into a set of grids, reconstructing the final distributions via a **fast convolution**. The same ideas underlie most **x-space PDF evolution codes**: HOPPET, QCDNUM, APFEL, ...
- Main limitations of present tools:
  - Restricted to a **limited number of processes**, implementation and debugging of each new process is time consuming
  - Only QCD corrections, **no QED and electroweak corrections** available, important for many LHC processes: jets, ttbar, W,Z + jets ....
  - Only Fixed Order processes, cannot account for **Monte Carlo parton shower effects**

aMCatNLO



- **aMCatNLO** provides theory predictions with **NLO QCD accuracy for arbitrary processes** in the Standard Model, and their matching to **Parton Showers**
- Built upon the **MadGraph** framework, it uses **MadFKS** for subtraction of soft/collinear divergences, **MadLoop** (with CutTools) for the computation of virtual corrections, and the **MC@NLO** method to include parton showers
- **PDF and scale uncertainties** provided by default within a single run
- First beta version release in 2012, **first full release** this week: **MadGraph_aMCatNLO v2.0.0**

Ongoing developments:

- Complex mass scheme for **finite-width resonances**
- Automation of **QED and electroweak computations**
- Automation of **Higgs and BSM physics at NLO**

Why we want a fast interface to aMCatNLO?

- **aMCatNLO** provides theory predictions with **NLO accuracy for arbitrary processes**
  - **Available fast interfaces are restricted to a limited number of processes**
  - **The implementation of each new process is time consuming and error-prone**
  - **A fast interface to aMCatNLO would give us fast interface to all NLO LHC processes at once**
- **aMCatNLO** provides an automatic matching of NLO events with **various parton showers**
  - **Available fast interfaces allow only fixed-order computations**
  - **NLO+PS computations are not only more accurate, they also provide an exclusive events description, and allow a more direct data/theory comparisons with reduced extrapolations**
- **aMCatNLO** will soon include not only NLO QCD but also **NLO electroweak corrections**
  - **QED and electroweak corrections are important to fit TeV scale data, and are not available in form of a fast interface to PDF fits**

Therefore, a **fast interface to aMCatNLO** is of outmost important for **global PDF analysis**:

- ☑ **Increase the number of processes for which fast NLO interfaces are available**, and that thus can be used to constrain PDFs.
- ☑ Allow to perform **PDF fits with NLO+PS accuracy**, study the stability of PDF fits wrt higher order corrections, increase the number of observables that can be used in PDF fits, and eventually provide **specific PDF sets for NLO event generators**
- ☑ Include consistently **electroweak corrections in PDF fits** at the matrix element level

Fast interfaces to NLO codes

fastNLO

the APPLgrid project

- The basic idea is to **represent PDFs in a grid in x and Q²**, with some suitable interpolation

$$f(x,Q^2)=\sum_{i=0}^{n}\sum_{\iota=0}^{n'} f_{k+i,\kappa+\iota}\, I_i^{(n)}\left(\frac{y(x)}{\delta y}-k\right) I_\iota^{(n')}\left(\frac{\tau(Q^2)}{\delta\tau}-\kappa\right),$$

- Then for each **new event with weight** w_m one **updates** a portion of the interpolating grid

$$W^{(p_m)}_{k+i,\kappa+\iota}\to W^{(p_m)}_{k+i,\kappa+\iota}+w_m\, I_i^{(n)}\left(\frac{y(x_m)}{\delta y}-k\right) I_\iota^{(n')}\left(\frac{\tau(Q^2_m)}{\delta\tau}-\kappa\right)$$

- So that effectively, one is **precomputing the parton-level cross sections in a grid**, including the sum over events, and then, a posteriori, physical observables can be computed (very fast) by the **convolution of this grid with PDFs at the x, Q² points.** In the case of **DIS** we thus have

$$W=\sum_p\sum_{i_y}\sum_{i_\tau} W^{(p)}_{i_y,i_\tau}\left(\frac{\alpha_s\left(Q^{2(i_\tau)}\right)}{2\pi}\right)^p f\left(x^{(i_y)},Q^{2(i_\tau)}\right)$$

(Examples from APPLgrid paper, **arXiv:0911.2985**)

Fast interfaces to NLO codes

- In the case of **hadron-hadron collisions,** the generalization is straightforward

$$W = \sum_p \sum_{l=0}^{n_{\rm sub}} \sum_{i_{y_1}} \sum_{i_{y_2}} \sum_{i_\tau} W^{(p)(l)}_{i_{y_1},i_{y_2},i_\tau} \left(\frac{\alpha_s\left(Q^{2(i_\tau)}\right)}{2\pi} \right)^p F^{(l)}\left(x_1^{(i_{y_1})}, x_2^{(i_{y_1})}, Q^{2(i_\tau)}\right)$$

- The relevant **parton luminosities** that enter the convolution are **process dependent**
- For instance, for **jet production at NLO** we have seven independent parton luminosities

$$\begin{aligned}
\mathrm{gg}: \quad F^{(0)}(x_1,x_2;Q^2) &= G_1(x_1)G_2(x_2) \\
\mathrm{qg}: \quad F^{(1)}(x_1,x_2;Q^2) &= \left(Q_1(x_1)+\overline{Q}_1(x_1)\right)G_2(x_2) \\
\mathrm{gq}: \quad F^{(2)}(x_1,x_2;Q^2) &= G_1(x_1)\left(Q_2(x_2)+\overline{Q}_2(x_2)\right) \\
\mathrm{qr}: \quad F^{(3)}(x_1,x_2;Q^2) &= Q_1(x_1)Q_2(x_2)+\overline{Q}_1(x_1)\overline{Q}_2(x_2) - D(x_1,x_2) \\
\mathrm{qq}: \quad F^{(4)}(x_1,x_2;Q^2) &= D(x_1,x_2) \\
\mathrm{q\bar{q}}: \quad F^{(5)}(x_1,x_2;Q^2) &= \overline{D}(x_1,x_2) \\
\mathrm{q\bar{r}}: \quad F^{(6)}(x_1,x_2;Q^2) &= Q_1(x_1)\overline{Q}_2(x_2)+\overline{Q}_1(x_1)Q_2(x_2) - \overline{D}(x_1,x_2), \qquad (12)
\end{aligned}$$

where g denotes gluons, q, quarks and r, quarks of different flavour, $q \neq r$ and we have used the generalised PDFs defined as:

$$G_H(x) = f_{0/H}(x,Q^2), \qquad Q_H(x) = \sum_{i=1}^{6} f_{i/H}(x,Q^2), \quad \overline{Q}_H(x) = \sum_{i=-6}^{-1} f_{i/H}(x,Q^2),$$

$$D(x_1,x_2) = \sum_{\substack{i=-6\\ i\neq 0}}^{6} f_{i/H_1}(x_1,Q^2) f_{i/H_2}(x_2,Q^2), \qquad (13)$$

PDF reweighting in aMCatNLO

- **A key advantage of aMCatNLO** is that, once an event has been generated for a given PDF set and a given scale, it is possible to easily **recompute the event weight for any other PDF or scale** using reweighting (**arxiv:1110.4738**)
- This is efficient as compared to recomputing event weight from scratch, but still requires to **perform a sum over N events for each PDF/scale variation** -> Not practical to be used in PDF fits
- At fixed order reweighting is performed **on the fly** for each subdirectory

$$d\sigma^{(\mathrm{NLO})} = \sum_{\alpha} d\sigma^{(\mathrm{NLO},\alpha)} \,,$$

α = Resolved Events, Soft+Born, Collinear, Soft-Collinear Counterevents

$$d\sigma^{(\mathrm{NLO},\alpha)} = f_1(x_1^{(\alpha)},\mu_F^{(\alpha)}) f_2(x_2^{(\alpha)},\mu_F^{(\alpha)}) W^{(\alpha)} d\chi_{Bj} d\chi_{n+1} \,,$$

$$\mathcal{R}_i^{(\alpha)} = f_1'(x_{1;i}^{(\alpha)},\mu_F'^{(\alpha)}) f_2'(x_{2;i}^{(\alpha)},\mu_F'^{(\alpha)}) \Bigg\{ g_S^{2b+2}(\mu_R'^{(\alpha)}) \left[\widehat{W}_0^{(\alpha)}(\mathcal{K}_{n+1;i}^{(\alpha)}) + \widehat{W}_F^{(\alpha)}(\mathcal{K}_{n+1;i}^{(\alpha)}) \log\left(\frac{\mu_F'^{(\alpha)}}{Q}\right)^2 + \widehat{W}_R^{(\alpha)}(\mathcal{K}_{n+1;i}^{(\alpha)}) \log\left(\frac{\mu_R'^{(\alpha)}}{Q}\right)^2 \right] + g_S^{2b}(\mu_R'^{(\alpha)}) \widehat{W}_B(\mathcal{K}_{n+1;i}^{(\alpha)}) \delta_{\alpha S} \Bigg\} \Bigg/ \frac{d\sigma^{(\mathrm{NLO},\alpha)}}{d\chi_{Bj} d\chi_{n+1}} (\mathcal{K}_{n+1;i}^{(\alpha)}, x_{1;i}^{(\alpha)}, x_{2;i}^{(\alpha)}) \,. \quad (2.21)$$

A Fast Interface to aMCatNLO

- Once we know, for each event, which are the values of the **weights W_B, W_0, W_F, W_R,** and of the relevant **parton luminosity $q_i(x, \mu_F)$ $q_j(x, \mu_F)$,** we can reconstruct the original event weight
- Recall that in general weights, parton luminosities, scales, Bjorken-x etc depend on whether we have **resolved events**, **Born** kinematics or soft, collinear and soft-collinear **counterevents**
- The key idea to construct an interface to aMCatNLO is to **interpolate the weights W_B, W_0, W_F, W_R** in a grid in **(x, μ_F),** keeping track of the **relevant luminosities** for each event, and then use the interpolated grids to **reconstruct the original distributions** via a very fast convolution, for arbitrary PDFs and scales
- aMCatNLO determines **automatically the PDF luminosities** relevant for a particular process, only requires typing it in the **python** shell
- Suitable interpolating and convolution tools are provided by **APPLgrid** framework
- Calculations in the **APPLgrid** format are already being used by **most PDF collaborations**: NNPDF, HERAPDF, MSTW, and are integral part of **HERAfitter**, so interested parties should be able to straightforwardly use the **fast interface to aMCatNLO**

amcbridge

- The cross-talk between **MG5_aMC** and **applgrid** is performed with a **new Python/C++ library**, called **amcbridge.** It is already available from **applgrid** svn (**v1.0.0** works with **r306** 2.0.0beta4_APPLGRID branch)
- To reduce memory footprint, **LHAPDF6.0.4** (the new C++ version) is used through the code

MG5_aMC

Main calculation
Filling of topdrawer histograms
Determination of relevant PDF luminosities
Evaluation of the W0, WB, WF, WR weights for each event (reweight_xsec.f)

applgrid

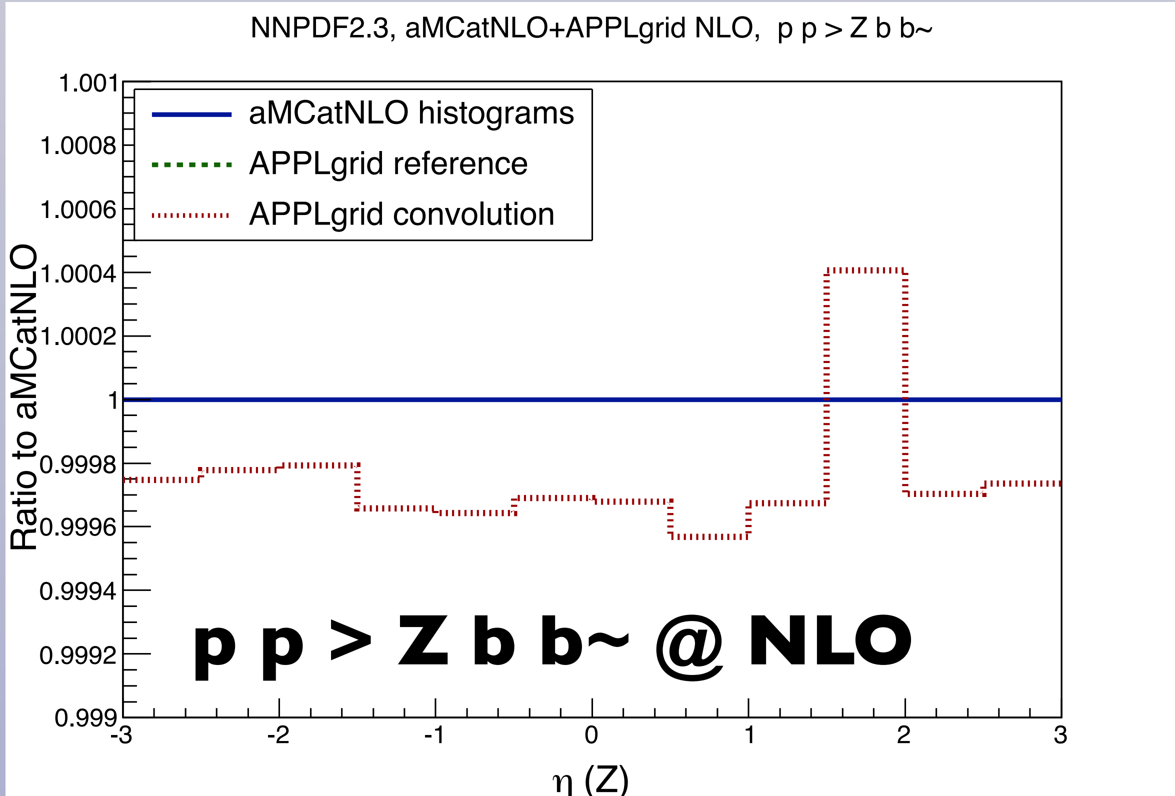
Grid initialization
A posteriori fast convolution between interpolated grids and arbitrary PDFs
A posteriori Scale variation
Grid optimization

amcbridge

Communications between MG5_aMC and applgrid
Filling of applgrids
Consistency tests (histogram filling, observables, luminosities ...)
Check reconstruction of event weights from W0, WB, WF, WR, pdf luminosities and strong coupling
Combination of applgrids from different subprocesses
Validation of the applgrids generation, comparison with topdrawer histo for central settings, and PDF/scale variations via reweighting

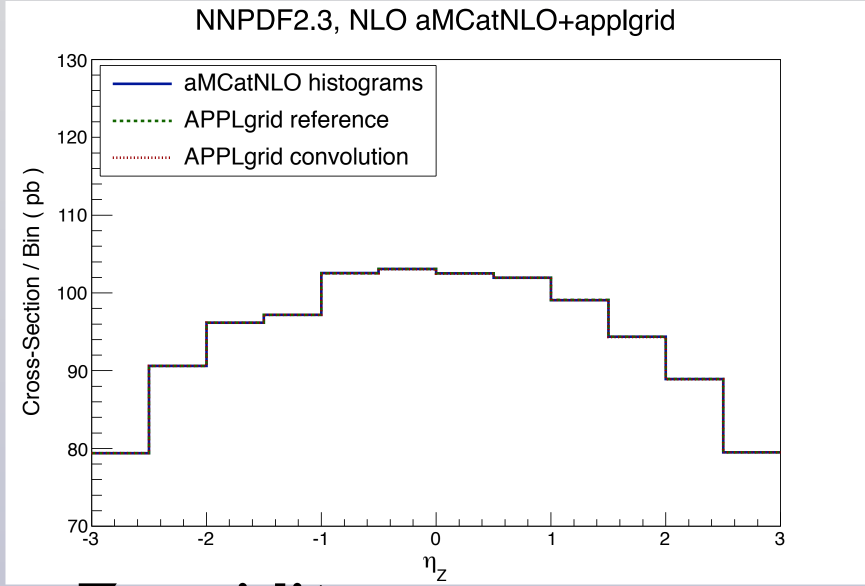
aMCatNLO+applgrid interface

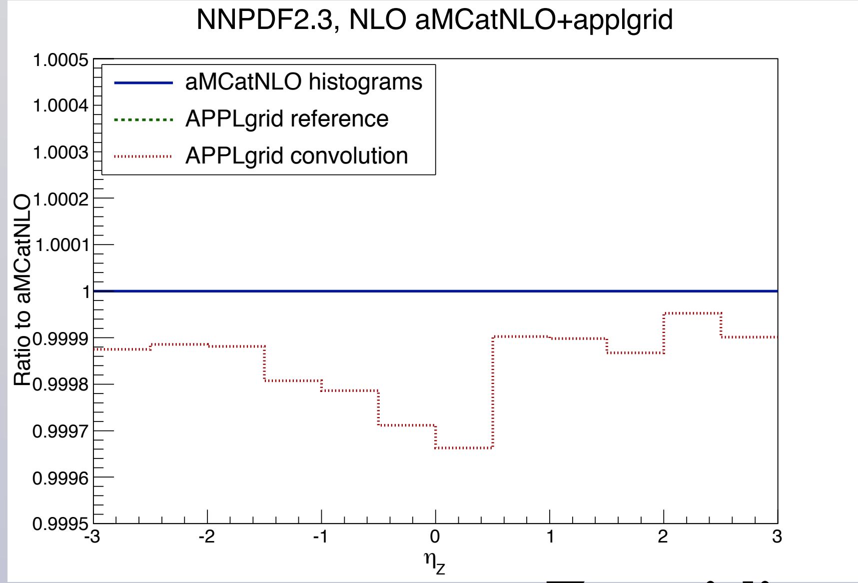
- The **applgrid/amcbridge** interface to aMCatNLO works beautifully for all the processes that we have considered
- This has required modifications in the main **applgrid** library as well, which are now being tested and will be part of **future official releases**
- **Very good agreement at NLO (well below 0.1%)** between original aMCatNLO distributions and **applgrid interpolation** for a wide variety of processes



Z rapidity

aMCatNLO+applgrid interface

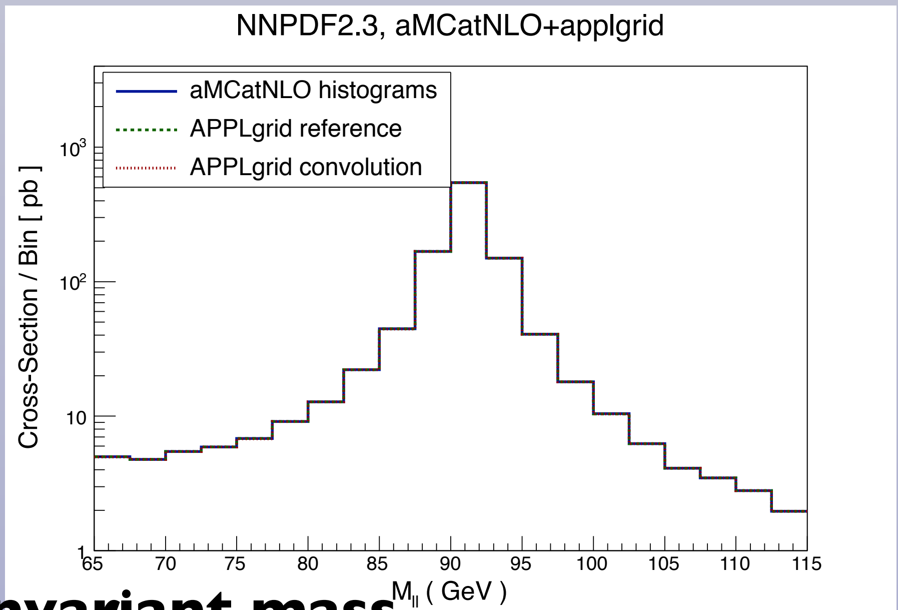


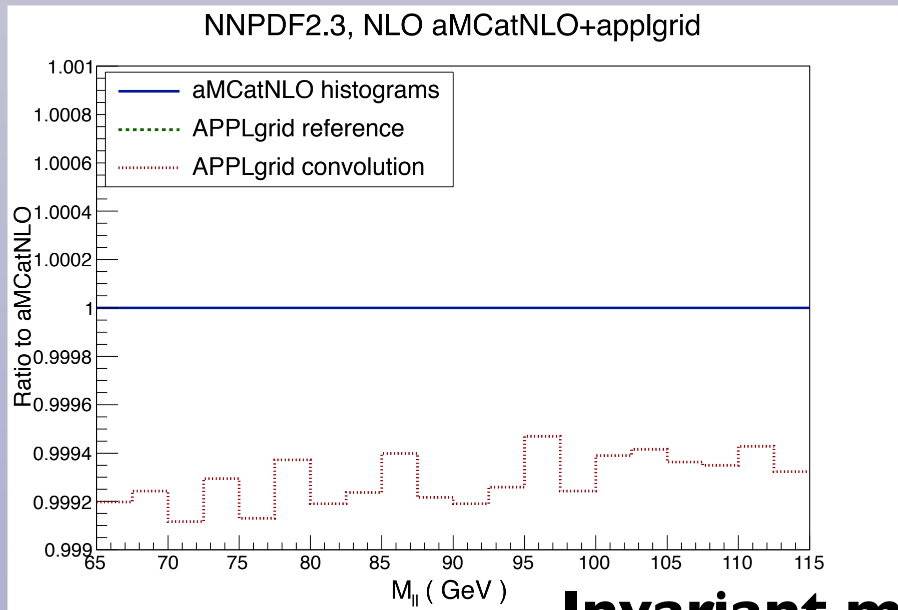


Z rapidity

Z rapidity

p p > l+ l- @ NLO





Invariant mass

Invariant mass

aMCatNLO+applgrid interface

- List of processes for which the **amcbridge** interface has been explicitly shown to work:

  - p p > t t~ [QCD]
  - p p > l+ le [QCD]
  - p p > l+ l- [QCD]
  - p p > w+ w- [QCD]
  - p p > w+ c [QCD]
  - p p > a a [QCD]
  - p p > z b b~ [QCD]
  - p p > h [QCD]
  - p p > h q q [QCD]

- The generation of the **applgrids** is remarkable efficient: even complicated processes like **p p > z b b~** can be run on a laptop in few hours (**MCFM+applgrid** typically requires rather more time to achieve smooth distributions)

- We also checked that many individual partonic processes like **u u > u u** work fine. The only missing ingredient is the implementation of the **FastJet** algorithms to define processes with final state jets

- No problems arise even if **very high statistics runs** are used

Ongoing developements

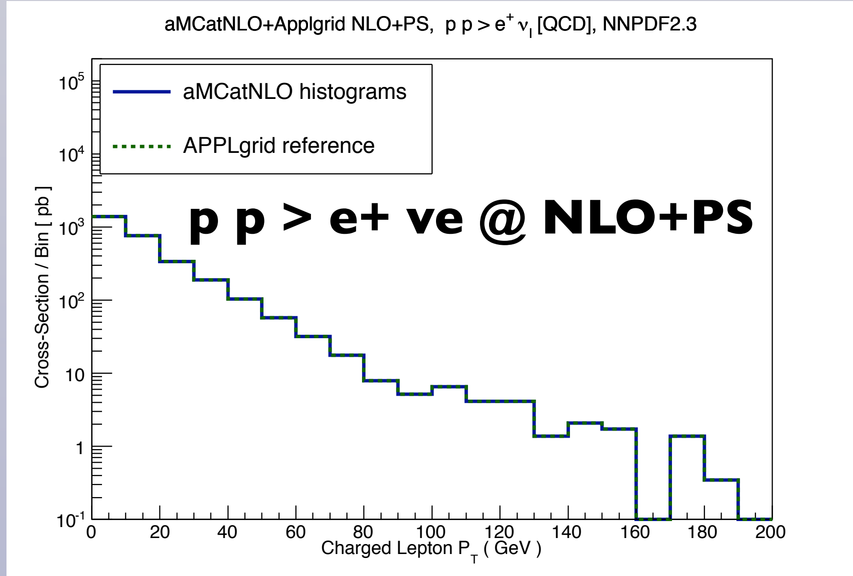
- Merged the **2.0.0beta4_APPLgrid** branch with most updated version (r350) of the main **2.0.0beta4** code, the one which will be **publicly released this week**
- A new **fixed-order structure** in aMCatnlo more suitable for **applgrid** interface, since basically the user needs to provide only the **corresponding analysis file** with the distributions that **applgrid should interpolate**
- Need to validate the **scale variation** by comparing the output of **amcbridge** with the original aMCatNLO histograms that the reweighting code produces
- Merge **2.0.0beta4_APPLgrid** into the main **MG5_aMC code,** so that it is available by default (timescale: some point in Feb). The user only needs to provide the links to **applgrid/amcbridge** and produce the corresponding analysis file
- In the meantime, **write paper** where the framework, methodology, results and user's guide are discussed (timescale: end of Jan)
- The code is already fully operative, so if you need **applgrids** with **aMCatNLO** for any particular process, just let me know!

Beyond NLO QCD

- The **automation of fast NLO QCD calculations needed in PDF fits for arbitrary processe**s can be considered as **fully completed**, and will be soon part of the **main MG5_aMC framework**

- Next step is to extend the **applgrid/amcbridge** interface to **NLO+PS calculations.** This requires some work: the optimal solution would be to **extend the NLO+PS reweighting format propagating all the information** that allow to fill applgrids starting from showered events, and modify **LHEF/HepMC** formats

- We have already checked that there are **no technical problems** in linking **applgrid**/**amcbridge** with Herwig and to fill the reference histograms after parton shower



- Once the QED/EW code is ready, will generalize **amcbridge** to processes with **both QCD and QED/EWK corrections** (different grid filling / convolution structure)

- In particular, use **amcbridge** with QCD/QED corrections to provide stringent constraints to the **photon PDF** using LHC data sensitive to **photon-initiated contributions,** and use QCD+EW corrections to fit high-ET data like jets, high mass tt, high-mass DY etc where electroweak corrections are large

Summary and outlook

- The **automation of fast NLO QCD calculations needed in PDF fits for arbitrary processe**s can be considered as **fully completed**, and will be soon part of the **default MG5_aMC framework**
- No extra work required by the user: stand-alone **MG5_aMC** (linked to **amcbridge**) will produce together with the standard distributions the corresponding **applgrids**
- Ongoing work to generalize the fast interface to **NLO+PS events**
- Also ongoing work in developing the automation of **NLO QED/EW calculations**, and generalizing **amcbridge** to be also able to easily use these in PDF fits

Summary and outlook

- The **automation of fast NLO QCD calculations needed in PDF fits for arbitrary processes** can be considered as **fully completed**, and will be soon part of the **default MG5_aMC framework**
- No extra work required by the user: stand-alone **MG5_aMC** (linked to **amcbridge**) will produce together with the standard distributions the corresponding **applgrids**
- Ongoing work to generalize the fast interface to **NLO+PS events**
- Also ongoing work in developing the automation of **NLO QED/EW calculations**, and generalizing **amcbridge** to be also able to easily use these in PDF fits

Thanks for your attention!