

A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: the APPLGRID project

Tancredi Carli^{1,a}, Dan Clements², Amanda Cooper-Sarkar³, Claire Gwenlan³, Gavin P. Salam⁴, Frank Siegert⁵, Pavel Starovoitov^{1,6}, Mark Sutton⁷

¹Department of Physics, CERN, Geneva, Switzerland

²University of Glasgow, Glasgow, UK

³University of Oxford, Oxford, UK

⁴LPTHE, UPMC Univ. Paris 6 and CNRS UMR 7589, Paris 05, France

⁵IPPP, Durham University, Durham, UK

⁶Nat. Sci. Educ. Center of Part. and HEP, Minsk, Belarus

⁷University of Sheffield, Sheffield, UK

Received: 16 November 2009 / Revised: 25 December 2009 / Published online: 23 February 2010

© Springer-Verlag / Società Italiana di Fisica 2010

Abstract A method to facilitate the consistent inclusion of cross-section measurements based on complex final-states from HERA, TEVATRON and the LHC in proton parton density function (PDF) fits has been developed. This can be used to increase the sensitivity of LHC data to deviations from Standard Model predictions. The method stores perturbative coefficients of NLO QCD calculations of final-state observables measured in hadron colliders in look-up tables. This allows the a posteriori inclusion of parton density functions (PDFs), and of the strong coupling, as well as the a posteriori variation of the renormalisation and factorisation scales in cross-section calculations. The main novelties in comparison to original work on the subject are the use of higher-order interpolation of Lagrangian form, which substantially improves the trade-off between accuracy and memory use, and a CPU and computer memory optimised way to construct and store the look-up table using modern software tools. It is demonstrated that a sufficient accuracy on the cross-section calculation can be achieved with reasonably small look-up table size by using the examples of jet production and electro-weak boson (Z , W) production in proton-proton collisions at a center-of-mass energy of 14 TeV at the LHC. The use of this technique in PDF fitting is demonstrated in a PDF-fit to HERA data and simulated LHC jet cross-sections as well as in a study of the jet cross-section uncertainties at various centre-of-mass energies.

1 Introduction

The Large Hadron Collider (LHC) at CERN will collide protons at a centre-of-mass energy of up to 14000 GeV. The combination of its high collision rate and centre-of-mass energy will make it possible to probe new interactions at very short distances. Such interactions might be revealed in the production of cross-sections of particles at very high transverse momentum (p_T) as a deviation from the Standard Model theory.

The sensitivity to new physics depends on experimental uncertainties in the measurements and on theoretical uncertainties in the Standard Model predictions. It is therefore important to work out a strategy to minimise both the experimental and theoretical uncertainties from LHC data. Residual renormalisation and factorisation scale uncertainties in next-to-leading order (NLO) QCD calculations for single inclusive jet cross-sections are typically about 5–10% and should hopefully be reduced as NNLO calculations become available. However, in some kinematic regimes, PDF uncertainties can be substantially larger than the uncertainties from higher-order corrections, for example at large p_T . One strategy to reduce such uncertainty is to use single inclusive jet or Drell-Yan cross-sections at lower p_T to constrain the proton parton density function (PDF) uncertainties at high p_T .

In order to further constrain PDF uncertainties, it would be useful to be able to include final state data such as p_T and rapidity distributions for W/Z -boson and jet production in global NLO QCD PDF fits, without recourse to inexact

^a e-mail: tancredi.carli@cern.ch

methods like the use of simple factor correcting of LO cross-sections (k -factors). We propose here a method for a consistent inclusion of final-state observables in global QCD analyses.

For inclusive data, like the proton structure function F_2 in deep-inelastic scattering (DIS) the perturbative coefficients are known analytically. During the fit the cross-section can therefore be quickly calculated from the strong coupling (α_s) and the PDFs and then be compared to the measurements. However, final state observables, where detector acceptances or jet algorithms are involved in the definition of the perturbative coefficients (called “weights” in the following), have to be calculated using NLO QCD Monte Carlo programs. Typically such programs need about one day of CPU time to accurately calculate the cross-section. It is therefore necessary to find less time consuming methods.

Any NLO QCD calculation of a final-state observable involves Monte Carlo integration over a large number of events. For deep-inelastic scattering and at hadron colliders this must usually be repeated for each new PDF set, making it impractical to consider many ‘error’ PDF sets, or carry out PDF fits. Here, the “a posteriori” inclusion of PDFs is discussed, whereby the Monte Carlo run calculates a look-up table (in momentum fraction, x , and momentum transfer, Q) of cross-section weights that can subsequently be combined with an arbitrary PDF. The procedure is numerically equivalent to using an interpolated form of the PDF.

Many methods have been proposed to solve this problem in the past [1–5]. In principle the highest efficiencies can be obtained by taking moments with respect to Bjorken- x [1, 2], because this converts convolutions into multiplications. This can have notable advantages with respect to memory consumption, especially in cases with two incoming hadrons. On the other hand, there are complications such as the need for PDFs in moment space and the associated inverse Mellin transforms.

Methods in x -space have traditionally been somewhat less efficient, both in terms of speed and in terms of memory consumption. They are, however, somewhat more transparent since they provide direct information on the x values of relevance. Furthermore they can be used with any PDF. The use of x -space methods can be further improved by using methods developed originally for PDF evolution [6–8].

Our method [9] bears a number of similarities to that of the fastNLO project [10] and the two approaches were to some extent developed in parallel. Relative to fastNLO, we take better advantage of the sparse nature of the x -dependent weights, allow for more flexibility in the scale choice by keeping explicitly the scale dependence as an additional dimension in the weighting table and provide a means to evaluate renormalisation and factorisation scale-dependence a posteriori. We also provide a broader range of processes, since in addition to di-jet production, we include W- and Z-boson production. In order to make easy use

of the large number of weight files for practically all inclusive jet p_T spectra and di-jet mass spectra made available by the fastNLO project, we provide a software interface to make use of these weight tables within the APPLGRID framework.

2 PDF-independent representation of cross-sections

2.1 Representing the PDF on a grid

We make the assumption that PDFs can be accurately represented by storing their values on a two-dimensional grid of points and using n^{th} -order interpolations between those points. Instead of using the parton momentum fraction x and the factorisation scale Q^2 , we use a variable transformation that provides good coverage of the full x and Q^2 range with uniformly spaced grid points:

$$y(x) = \ln \frac{1}{x} + a(1-x) \quad \text{and} \quad \tau(Q^2) = \ln \ln \frac{Q^2}{\Lambda^2}. \quad (1)$$

The parameter Λ should be chosen of the order of Λ_{QCD} , but need not necessarily be identical. The parameter a serves to increase the density of points in the large x region¹ and can be chosen according to the needs of the concrete application.²

The PDF $f(x, Q^2)$ is then represented by its values f_{i_y, i_τ} at the 2-dimensional grid point $(i_y \delta y, i_\tau \delta \tau)$, where δy and $\delta \tau$ denote the grid spacings, and is obtained elsewhere by interpolation using polynomials of Lagrangian form:

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

where n, n' are the interpolation orders. The interpolation function $I_i^{(n)}(u)$ is 1 for $u = i$, and otherwise is given by:

$$I_i^{(n)}(u) = \frac{(-1)^{n-i}}{i!(n-i)!} \frac{u(u-1)\dots(u-n)}{u-i}. \quad (3)$$

Defining $\text{int}(u)$ to be the largest integer such that $\text{int}(u) \leq u$, k and κ are defined as:

$$k(x) = \text{int} \left(\frac{y(x)}{\delta y} - \frac{n-1}{2} \right), \quad (4)$$

$$\kappa(Q^2) = \text{int} \left(\frac{\tau(Q^2)}{\delta \tau} - \frac{n'-1}{2} \right).$$

¹For a fixed total number of bins, as the bins at large x get finer, the low- x ones become wider.

²In case of $a = 0$ the function is analytically invertible, for $a \neq 0$ a numerical inversion has to be applied.

Given finite grids whose vertex indices range from, $0, \dots, N_y - 1$, for the y grid and, $0, \dots, N_\tau - 1$, for the τ grid, one should additionally require that Eq. 2 only uses available grid points. This can be achieved by remapping, $k \rightarrow \max(0, \min(N_y - 1 - n, k))$, and, $\kappa \rightarrow \max(0, \min(N_\tau - 1 - n', \kappa))$.

2.2 Representing the final state cross-section weights on a grid (DIS case)

To illustrate the method we take the case of a single flavour in deep-inelastic scattering (DIS).

Suppose that we have an NLO Monte Carlo program that produces events, $m = 1, \dots, N$. Each event m has an x value, x_m , a Q^2 value, Q_m^2 , as well as a weight, w_m . We define p_m as the number of powers in the strong coupling α_s in event m . Normally one would obtain the final result W of the Monte Carlo integration for one sub-process from:³

$$W = \sum_{m=1}^N w_m \left(\frac{\alpha_s(Q_m^2)}{2\pi} \right)^{p_m} f(x_m, Q_m^2), \tag{5}$$

where $f(x, Q^2)$ is the PDF of the flavour under consideration.

Instead one introduces a weight grid $W_{i_y, i_\tau}^{(p)}$ and then for each event one updates a portion of the grid with: $i = 0, \dots, n, \iota = 0, \dots, n'$:

$$W_{k+i, \kappa+\iota}^{(p_m)} \rightarrow W_{k+i, \kappa+\iota}^{(p_m)} + w_m I_i^{(n)} \left(\frac{y(x_m)}{\delta y} - k \right) \times I_\iota^{(n')} \left(\frac{\tau(Q_m^2)}{\delta \tau} - \kappa \right), \tag{6}$$

where $k \equiv k(x_m), \kappa \equiv \kappa(Q_m^2)$.

The final result for W , for an arbitrary PDF and an arbitrary α_s , can then be obtained *subsequent* to the Monte Carlo run:

$$W = \sum_p \sum_{i_y} \sum_{i_\tau} W_{i_y, i_\tau}^{(p)} \left(\frac{\alpha_s(Q^{2(i_\tau)})}{2\pi} \right)^p f(x^{(i_y)}, Q^{2(i_\tau)}), \tag{7}$$

where the sums with indices i_y and i_τ run over the number of grid points and we have explicitly introduced $x^{(i_y)}$ and $Q^{2(i_\tau)}$ such that:

$$y(x^{(i_y)}) = i_y \delta y \quad \text{and} \quad \tau(Q^{2(i_\tau)}) = i_\tau \delta \tau. \tag{8}$$

³Here, and in the following, renormalisation and factorisation scales have been set equal for simplicity.

2.3 Including renormalisation and factorisation scale dependence

If one has the weight matrix $W_{i_y, i_\tau}^{(p)}$ determined separately order by order in α_s , it is straightforward to vary the renormalisation μ_R and factorisation μ_F scales a posteriori (we assume that they were set equal in the original calculation).

It is helpful to introduce some notation related to the DGLAP evolution equation:

$$\frac{df(x, Q^2)}{d \ln Q^2} = \frac{\alpha_s(Q^2)}{2\pi} (P_0 \otimes f)(x, Q^2) + \left(\frac{\alpha_s(Q^2)}{2\pi} \right)^2 (P_1 \otimes f)(x, Q^2) + \dots, \tag{9}$$

where the P_0 and P_1 are the LO and NLO matrices of DGLAP splitting functions that operate on vectors (in flavour space) f of the PDFs. Let us now restrict our attention to the NLO case where we have just two values of p in Eq. 7. For example, in jet production in DIS, $p_{LO} = 1$ and $p_{NLO} = 2$. Introducing ξ_R and ξ_F corresponding to the factors by which one varies μ_R and μ_F respectively, for arbitrary ξ_R and ξ_F we may then write:

$$W(\xi_R, \xi_F) = \sum_{i_y} \sum_{i_\tau} \left\{ \left(\frac{\alpha_s(\xi_R^2 Q^{2(i_\tau)})}{2\pi} \right)^{p_{LO}} W_{i_y, i_\tau}^{(p_{LO})} f(x^{(i_y)}, \xi_F^2 Q^{2(i_\tau)}) + \left(\frac{\alpha_s(\xi_R^2 Q^{2(i_\tau)})}{2\pi} \right)^{p_{NLO}} \times [(W_{i_y, i_\tau}^{(p_{NLO})} + 2\pi\beta_0 p_{LO} \ln \xi_R^2 W_{i_y, i_\tau}^{(p_{LO})}) \times f(x^{(i_y)}, \xi_F^2 Q^{2(i_\tau)}) - \ln \xi_F^2 W_{i_y, i_\tau}^{(p_{LO})} (P_0 \otimes f)(x^{(i_y)}, \xi_F^2 Q^{2(i_\tau)})] \right\}, \tag{10}$$

where $\beta_0 = (11N_c - 2n_f)/(12\pi)$ and N_c (n_f) is the number of colours (flavours). The a posteriori variation of the renormalisation and factorisation scales works only, if ξ_r and ξ_f are constants. If the functional form of μ_r and μ_f is changed, for instance, from the leading jet transverse momenta to the average transverse jet momentum, the weighting grid has to be regenerated.

Though this formula is given for an x -space based approach, a similar formula applies for moment-space approaches. Furthermore it is straightforward to extend it to higher perturbative orders.

To obtain the full DIS cross-section a summation of the weights and the parton densities over the contributing sub-processes is required.

2.4 The case of two incoming hadrons

In hadron-hadron scattering one can use analogous procedures but with one more dimension. Besides Q^2 , the weight grid depends on the momentum fractions of the first (x_1) and second (x_2) hadrons.

The analogue of Eq. 7 is given by:

$$W = \sum_p \sum_{l=0}^{n_{\text{sub}}} \sum_{i_{y_1}} \sum_{i_{y_2}} \sum_{i_\tau} W_{i_{y_1}, i_{y_2}, i_\tau}^{(p)(l)} \left(\frac{\alpha_s(Q^{2(i_\tau)})}{2\pi} \right)^p \times F^{(l)}(x_1^{(i_{y_1})}, x_2^{(i_{y_1})}, Q^{2(i_\tau)}), \tag{11}$$

where n_{sub} is the number of sub-processes and the initial state parton combinations F are specified in Eqs. 12, 20 and 18.

The combinations of the incoming parton densities (defining the number of sub-processes) often can be simplified by making use of the symmetries in the weights. In the case of jet production only seven sub-processes are needed (see Sect. 2.4.1). The case of W -boson and Z -boson production is treated in Appendix A. The case of b -quark production is discussed in Ref. [11].

An automated way to find the sub-processes is discussed in Appendix B.

2.4.1 Sub-processes for jet production in hadron-hadron collisions

In the case of jet production in proton-proton collisions the weights generated by the Monte Carlo program can be organised in seven possible initial-state combinations of partons:

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

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

defined as:

$$\begin{aligned} G_H(x) &= f_{0/H}(x, Q^2), \\ Q_H(x) &= \sum_{i=1}^6 f_{i/H}(x, Q^2), \\ \bar{Q}_H(x) &= \sum_{i=-6}^{-1} f_{i/H}(x, Q^2), \\ D(x_1, x_2) &= \sum_{i=-6, i \neq 0}^6 f_{i/H_1}(x_1, Q^2)f_{i/H_2}(x_2, Q^2), \\ \bar{D}(x_1, x_2) &= \sum_{i=-6, i \neq 0}^6 f_{i/H_1}(x_1, Q^2)f_{-i/H_2}(x_2, Q^2), \end{aligned} \tag{13}$$

where $f_{i/H}$ is the PDF of flavour $i = -6, \dots, 6$ for hadron H and H_1 (H_2) denotes the first or second hadron.⁴

2.5 Including scale dependence in the case of two incoming hadrons

It is again possible to choose arbitrary renormalisation and factorisation scales. Specifically for NLO accuracy:

$$\begin{aligned} W(\xi_R, \xi_F) &= \sum_{l=0}^{n_{\text{sub}}-1} \sum_{i_{y_1}} \sum_{i_{y_2}} \sum_{i_\tau} \left\{ \left(\frac{\alpha_s(\xi_R^2 Q^{2(i_\tau)})}{2\pi} \right)^{p_{\text{LO}}} \right. \\ &\quad \times W_{i_{y_1}, i_{y_2}, i_\tau}^{(p_{\text{LO}})(l)} F^{(l)}(x_1^{(i_{y_1})}, x_2^{(i_{y_1})}, \xi_F^2 Q^{2(i_\tau)}) \\ &\quad + \left(\frac{\alpha_s(\xi_R^2 Q^{2(i_\tau)})}{2\pi} \right)^{p_{\text{NLO}}} \\ &\quad \times [(W_{i_{y_1}, i_{y_2}, i_\tau}^{(p_{\text{NLO}})(l)} + 2\pi\beta_0 p_{\text{LO}} \ln \xi_R^2 W_{i_{y_1}, i_{y_2}, i_\tau}^{(p_{\text{LO}})(l)}) \\ &\quad \times F^{(l)}(x_1^{(i_{y_1})}, x_2^{(i_{y_1})}, \xi_F^2 Q^{2(i_\tau)}) \\ &\quad - \ln \xi_F^2 W_{i_{y_1}, i_{y_2}, i_\tau}^{(p_{\text{LO}})(l)} (F_{q_1 \rightarrow P_0 \otimes q_1}^{(l)}(x_1^{(i_{y_1})}, x_2^{(i_{y_1})}, \xi_F^2 Q^{2(i_\tau)}) \\ &\quad \left. \left. + F_{q_2 \rightarrow P_0 \otimes q_2}^{(l)}(x_1^{(i_{y_1})}, x_2^{(i_{y_1})}, \xi_F^2 Q^{2(i_\tau)}) \right) \right] \Big\}, \tag{14} \end{aligned}$$

where $F_{q_1 \rightarrow P_0 \otimes q_1}^{(l)}$ is calculated as $F^{(l)}$, but with q_1 replaced with $P_0 \otimes q_1$, and analogously for $F_{q_2 \rightarrow P_0 \otimes q_2}^{(l)}$.

2.6 Reweighting to a different center-of-mass energy

From a weight grid W calculated at a particular centre-of-mass energy \sqrt{s} it is also possible to calculate a cross-section at a different centre-of-mass energy $\sqrt{s'}$ by using

⁴In the above equation and in the following we follow the standard PDG Monte Carlo numbering scheme [12], where gluons are denoted as 0, quarks have values from 1–6 and anti-quarks have the corresponding negative values.

transformed parton momentum fractions $x'_{1/2}$ and adding a flux factor in the cross-section convolution as given by Eq. 11 or Eq. 14:

$$W'(\xi_R, \xi_F) = \frac{s}{s'} W(\xi_R, \xi_F), \quad (15)$$

and the momentum fractions $x_{1/2}$ in the generalised parton densities $F(x_1, x_2, Q^2)$ are replaced by:

$$x'_{1,2} = \frac{\sqrt{s}}{\sqrt{s'}} x_{1,2}. \quad (16)$$

Using a *higher* beam energy would imply a *lower* x for the same kinematics and for the same invariant mass of the partonic system. When $\sqrt{s'} < \sqrt{s}$ it can occur that $x'_1 > 1$ or $x'_2 > 1$, in which case the parton densities should be set to zero. One should be aware that a jet transverse momentum that corresponds to moderate x values with centre-of-mass energy \sqrt{s} (and correspondingly low density of grid points in x) may correspond to large x when using a smaller $\sqrt{s'}$. In such cases, it can happen that the low density of grid points in x is no longer sufficient, given that PDFs vary more rapidly at large x than at moderate x .

Special care is also needed when taking $\sqrt{s'} > \sqrt{s}$ insofar as there will be kinematic regions accessible with the larger $\sqrt{s'}$ values that were not probed at all in the original NLO calculation at centre-of-mass energy \sqrt{s} . As a concrete example, with $\sqrt{s'} = 14$ TeV, there can be events with three jets having respectively $p_T = 6, 4, 2$ TeV. Such events contribute to the inclusive jet spectrum at $p_T = 4$ TeV. However, taking a grid calculated with $\sqrt{s} = 10$ TeV (where such events are kinematically disallowed) and using it to determine the inclusive jet spectrum with $\sqrt{s'} = 14$ TeV, this kind of contribution will be left out.

3 Technical implementation

To test the scheme discussed above, the NLO QCD Monte Carlo programs NLOJET++ [13–15] for jet production and MCFM [16, 17] for the production of W - and Z -boson are used. To illustrate the performance of the method jet and W - and Z -boson production are used as examples. However, it is worth noting that these two programs give access to many of the NLO QCD calculations presently available.

The weight grid $W_{i_{y_1}, i_{y_2}, i_\tau}^{(p)(l)}$ of Eq. 11 is filled (for each cross-section bin) in the user module of the NLO program, where one has access to the event weights and the partons' momenta. This object is called “grid” in the following. At this point the cross-section definition is specified and the physical observables that are being studied are defined (e.g. using a jet algorithm).

The weight grid for each value of the observable in question is represented as a multidimensional object with one

dimension each for x_1 , x_2 and Q^2 , one for the sub-process in question, and one for the order in α_s . The task is to store the weight grid in such a way that as little memory as possible is used and the information can be extracted in a fast way. In the following several options to reduce the necessary memory are discussed:

The simplest structure for a software implementation of the weight grid is a multidimensional array (for x_1 , x_2 and Q^2), like the TH3D-class available in the ROOT analysis framework.

The overhead of storing empty bins can be largely reduced by calculating the x_1 , x_2 and Q^2 boundaries of the weight grid using the NLO QCD program in a special run before the actual filling step. At the beginning of the filling step the adjusted boundaries of the weight grids are then read-in and an optimised weight grid is constructed.

Since the rectilinear region bounded by limits in x_1 , x_2 and Q^2 may contain many phase-space points that are unoccupied, additional memory can be saved by using methods to avoid storing elements in the weight grid that are not filled. Since the occupied regions are continuous, but irregular, grid formats for truly sparse matrices (such as the Harwell-Boeing format) are not used. Instead a custom format is favoured where the grid, lower- and upper-limits in each dimension are stored along with all the elements in between.

This is illustrated in Fig. 1 for a simple two-dimensional grid. For the three-dimensional structure, each of the row-column elements would itself be a column with its own lower and upper range delimiters. The resulting saving of memory is usually around a factor of four, even after taking into account the additional storage for the range delimiters.⁵

Since the grid itself knows the index of the first and last filled element in each row, column etc., are the indices of the first filled row, or the first filled column for a given row, the grid that contain data. Similarly when interrogating the grid for the value of an element, it is possible to ascertain whether the element is in the occupied, or unoccupied region of the grid and return the value of the filled element if filled, or 0 otherwise. This makes accessing the unfilled members of the grid much faster than otherwise.

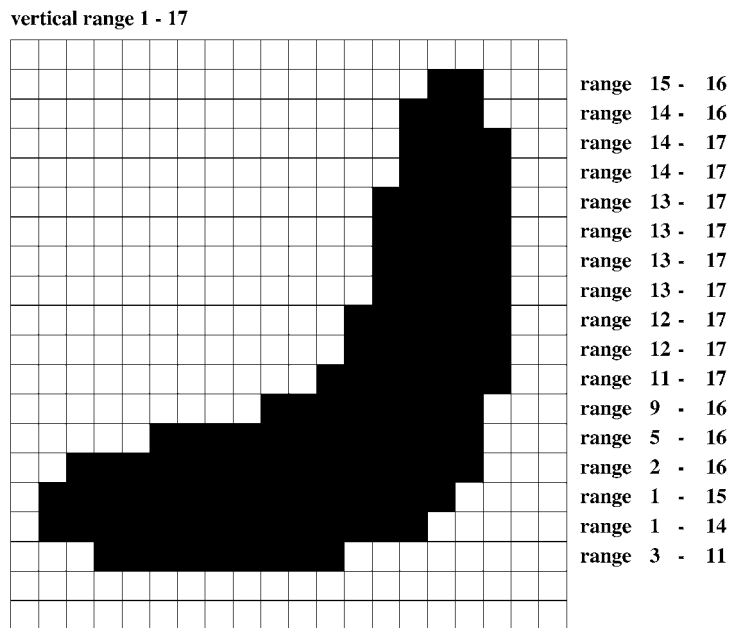
The actual implementation for the grid⁶ involves a number of related classes written in C++.⁷ The grid for a given cross-section is represented by a concrete instance of a master class, `appl::grid`. This class has a number of constructors that allow the cross-section it will calculate to be

⁵If additional savings are required in the future, packing the range delimiters for each sparse one dimensional structure into a single integer will halve this additional overhead, but will slightly increase the access time due to the unpacking.

⁶The complete code including the interfaces to NLOJET++ and MCFM is available from <http://svn.hepforge.org/applgrid>.

⁷A FORTRAN interface is also available so that the basic functionality can be accessed from within user FORTRAN code.

Fig. 1 An example of the custom two-dimensional sparse structure. Rows and columns are numbered from 0 to 20 from the top left. The elements with data members are shown filled, only rows 1 to 17 have data members, for each row the columns that have data members are shown on the right. A total of 117 elements, from the maximum of 400 elements are stored, along with the single pair of row-range delimiters, and the 17 pairs of column range delimiters for each of the individual rows



defined in terms of a fixed number of regular or variable width bins in the cross-section observable.

For each bin in the observable, the master class has a number of instances of an internal class—one for each order of α_s —so that for a cross-section with 10 bins, with contributions at leading order and next-to-leading order, the master class would contain a total of 20 instances of the internal class.

This internal class, `appl:igrd`, encodes all the information required to create the cross-section, at one particular order, for that bin. The class contains the x -to- y , y -to- x and Q^2 -to- τ , τ -to- Q^2 transform pairs, and a subclass that encodes information on how to generate the N generalised internal sub-processes for the particular interaction from the basic parton distribution functions. It also contains instances of the sparse grid class in x_1 , x_2 and Q^2 described above, for each of the N sub-processes.

When requested to perform the convolution, the master class calls the convolute method of the subclass for each order of the cross-section in each bin. The convolute method of the subclass performs the convolution over x_1 , x_2 and Q^2 for each of the sub-processes.

For each bin in the observable, the master class takes the cross-sections from the subclasses for each order from each bin and adds them to arrive at the final cross section for that bin.

The subclass for the generalised internal sub-processes are very basic classes which encode the number of sub-processes, i.e. seven in the case of jet production, and twelve in the case of Z -boson production, and simply take the 13 parton distribution values for each incoming hadron at a given scale, and generate the N internal processes from these.

When the grid is saved to a ROOT file⁸ the master class encodes the complete status of the internal grids, which transform pair, and which sub-process is required etc., so that once reading from the file, everything required to calculate the cross section (e.g. sub-process definition, CKM matrix elements etc.) is available. In this way all information to perform the cross-section calculation is available from the output file from a single function call by the user and the only additional information required is an input function for generating the PDFs and another one for calculating α_s . We use the HOPPET program [8] to calculate the DGLAP splitting functions needed for the cross-section convolution when the renormalisation and factorisation is varied (see Eq. 10).

All the various choices in the weight grid architecture and other information needed to calculate the cross-section are encoded in the output file. They are described in the following:

- The centre-of-mass energy at which the weight grid has been produced.
- The choice of the coordinate transform function. By default the form of Eq. 1 is used. However, any other function can be provided by the user.
- The interpolation order as given by Eq. 2.
- The number of grid points to be used for each dimension x_1 , x_2 and Q^2 .
- The definition of the sub-processes via a 13×13 matrix.
- The CKM matrix elements or other constants needed to calculate the cross-sections.

⁸Technically, the grid is transformed to TH3D-histograms that are stored in the output ROOT file.

- The required number of the points on the grid can be optionally reduced with the aid of reweighting factor in the filling step. This flattens out the PDF in the region where it is steeply falling.

By default the following functional form is used for the reweighting:⁹

$$w(x) = x^{a_1} (1 - 0.99x)^{a_2}. \quad (17)$$

The parameter a_1 can be adjusted to flatten out the change of the PDF at low- x while the parameter a_2 can be optimised for the high- x region. The factor 0.99 prevents the weight from being zero for $x = 1$.

Reasonable values for the parameters a_1 and a_2 have been determined by fitting the sum of the up, down and gluon PDFs. For the CTEQ6 PDFs [18], values of $a_1 = -1.5$ to -1.6 and $a_2 = 3.0$ to 3.4 have been found for the range $5 < Q < 5000$ GeV. The variation comes from a slight dependence of the a_1 and a_2 parameters on Q^2 . For other PDFs, the results of the fit can be slightly different.

The user can change the parameters or provide another functional form.

4 Accuracy of the weight grids

The choice of the weight grid architecture depends on the required accuracy, on the exact cross-section definition and on the available computer resources. For each possible application the weight grid architecture has to be carefully chosen in order to achieve the required accuracy with the available computer memory and computing time. For instance, for observables where the PDFs are steeply falling, e.g. the inclusive jet cross-section at high transverse momentum in the forward region, a fine grid in x is needed. The memory usage of weight grids for one cross-section should be kept small since, e.g. in global PDF fits, it might be necessary to read in a large number of weight grids. In addition, the convolution time depends on the number of grid nodes, and so keeping memory requirements as small as possible is in any case desirable. The number of points needed in the weight grid is kept modest by using the higher-order interpolation functions of Eq. 2, and optionally also by introducing a PDF weight, as in Eq. 17 during the filling step, or by using a sparse structure.

In the following, the influence of the grid architecture on the achievable accuracy in the cross-section calculation is discussed. The computer memory use and execution speed are also investigated. The production of jets and of W - and Z -bosons at LHC are used as examples.

In our test runs, to be independent from statistical fluctuations (which can be large, in particular in the NLO case),

⁹Such a PDF reweighting was first introduced in Ref. [10].

in addition to the weight grid, reference histograms are filled using the NLO QCD calculation without weights in the standard way. The result obtained from the weight grid is then compared to these reference histogram.

4.1 Jet production at hadron colliders

The single inclusive jet cross-section as a function of the jet transverse momentum (p_T) is calculated for jets in the central rapidity (y) region of $0 < y < 1$ and in the forward rapidity region of $2 < y < 3$. Jets are defined via the seedless cone jet algorithm as implemented in NLOJET++, which corresponds to the seedless algorithm of Ref. [19] (or SIS-Cone [20]), except for small differences in the split-merge procedure which are irrelevant at this order. The cone radius has been set to $R = 0.7$, the overlap fraction to $f = 0.5$.¹⁰ The renormalisation and factorisation scales are set to $Q^2 = p_{T,\max}^2$, where $p_{T,\max}$ is the p_T of the highest p_T jet in the required rapidity region.¹¹

To discuss the dependence of the weight grid performance on the grid architecture, a default weight grid is defined from which variations in a single parameter are studied systematically. The default weight grid consists of 30 bins in x and 10 bins in Q^2 . The points are distributed according to Eq. 1 with $a = 5$ and 5th order interpolation is used. No PDF reweighting (see Eq. 17) is used.

The ratio of the cross-section calculated with the default weight grid to the reference cross-section calculation is shown in Fig. 2 for the jet cross section in the central rapidity region ($0 < y < 1$) (a) and the forward rapidity region ($2 < y < 3$) (b). The weight grid is produced in a run where the CTEQ6mE PDF [18] has been used to calculate the jet cross-section. This PDF is used as standard in the following. To show the independence of the weight grid performance on the used PDF, Fig. 2 also includes more recent PDFs based on the analyses of a large variety of data (global analysis) like CTEQ6.6 [22] and MSTW2008 [23] or only using inclusive DIS data based on combined H1 and ZEUS data (HERAPDF01) [24]. In addition, we include a PDF that does not use a parameterised input distribution NNPDF [25]. Further comparisons of the jet cross-sections calculated with these PDFs can be found in Sect. 5.

¹⁰These choices are related to the fact that some of the NLOJET++ runs were performed some time in the past. A modern cone-algorithm (in the class of those with a split-merge procedure) would be SIS-Cone [20], and a value of $f = 0.75$ would be recommended [21].

¹¹Note that beyond LO the $p_{T,\max}$ will in general differ from the p_T of the other jets, so when binning an inclusive jet cross-section, the p_T of a given jet may not correspond to the renormalisation scale chosen for the event as a whole. For this reason separate grid dimensions for the jet p_T and for the renormalisation scale are used. This requirement has been efficiently circumvented in some moment-space approaches [2].

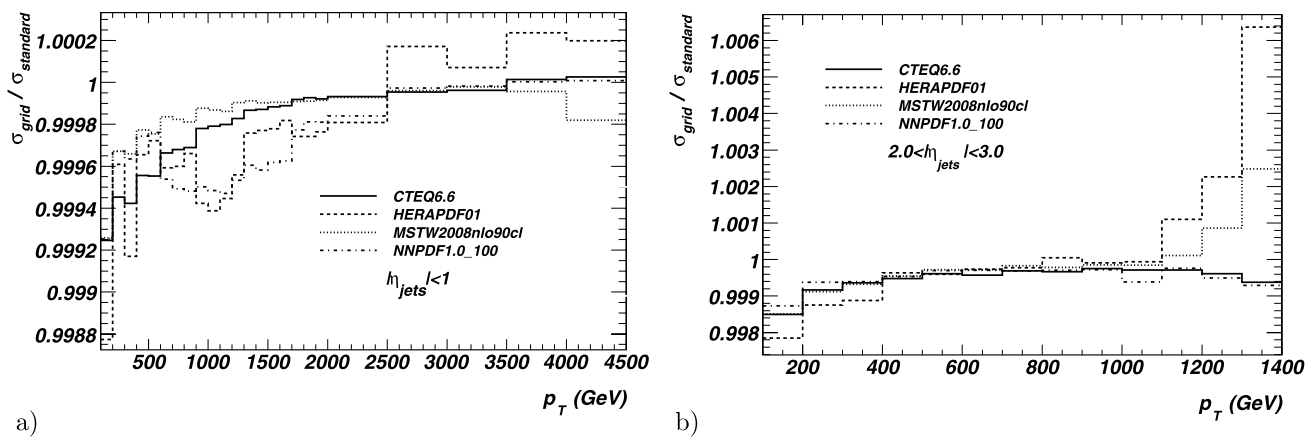


Fig. 2 Ratio of grid and standard calculations of the single inclusive jet p_T spectrum for $0 < y < 1$ (a) and for $2 < y < 3$ (b), for a variety of PDFs. The results are shown for the default weight-grid settings, i.e.

In the central region the cross-section calculated with the weight grid reaches an accuracy of about 0.1% for all transverse jet momenta and all PDFs. In the forward region a similar performance is achieved for transverse jet momenta up to 1000 GeV. For transverse jet momenta above that value the performance degrades to 0.6% and a variation with the PDF is observed.

The dependence of the accuracy on the number of x -bins is illustrated in Fig. 3. If only 25 x -bins are used, the accuracy is 0.3% in the central and 0.6% in the forward rapidity region. The accuracy decreases towards low jet transverse momenta. More accuracy is achieved by a larger number of x -bins. For 30 bins the accuracy is 0.1%. For 40 x -bins the improvement is small, but visible. A very sensitive kinematic region is the forward region with very high transverse momenta. In this region at least 30 x -bins are needed to get an accuracy of 0.1%.

Figure 4 shows the dependence of the accuracy on the number of Q^2 -bins. This dependence is rather small. When a large enough number of x -bins is chosen, no change is observed for 8 to 15 bins in Q^2 .

The dependence on the interpolation order (as defined in Eq. 3) is shown in Fig. 5. While varying the default interpolation order $n = 5$ to $n = 4$ and $n = 6$ gives similar results within 0.1%, the interpolation order $n = 3$ leads to an accuracy loss of 0.5% at low transverse jet momenta in the central regions, and by 0.4–1% at low and high transverse jet momenta in the forward region.

In conclusion, the results in Figs. 2–5 demonstrate that an accuracy of 0.1% can be reached with a reasonable weight grid size. The most critical parameter is the number of x -bins, which must be large enough to accommodate strong PDF variations in certain phase space regions. In comparison, the dependence on the number of Q^2 bins is rather

30 bins in x , 10 bins in Q^2 , a coordinate transform parameter $a = 5$ and fifth order interpolation

weak. The interpolation between the grid points is sufficiently accurate to allow the grid technique to be used and fifth order interpolation produces reasonable results. The achieved accuracy is probably sufficient for all practical applications.

In applications where a very small weight grid is needed, one can also introduce a PDF-weight to flatten out the x -dependence of the PDFs (see Eq. 17). The PDF weight is calculated using $a_1 = -1.5$ and $a_2 = 3$. This is illustrated in Fig. 6, where grids with very low number of x -bins (8, 9, 10) and eight Q^2 bins are used, the interpolation is lowered to $n = 4$. Even with the smallest weight grid an accuracy of 1% is achieved using the PDF-weight. For a somewhat larger weight grid with 10 x -bins the accuracy is 0.5% in all phase space regions.

One of the important theoretical uncertainties in NLO QCD calculations is the variation of the results with the choice of the factorisation and renormalisation scale. Equation 14 allows the calculation of the cross-section for any scale choice a posteriori from one weight grid produced at a fixed scale choice. The results from scale variations by a factor of 2 up and down is shown in Fig. 7. The renormalisation and factorisation scales are either varied together or varied individually. The weight grid result has been calculated with a single weight grid and the reference cross-sections have been calculated by repeating the standard NLO QCD calculation for each of the scale variations. The cross-section calculated with the weight grid reproduces the standard results to within about 0.1% in the central region and 0.1–0.2% in the forward region.

4.2 Reweighting jet cross-sections to a different centre-of-mass energy

As outlined in Sect. 2.6 a weight grid produced at a given centre-of-mass energy can also be used to calculate the

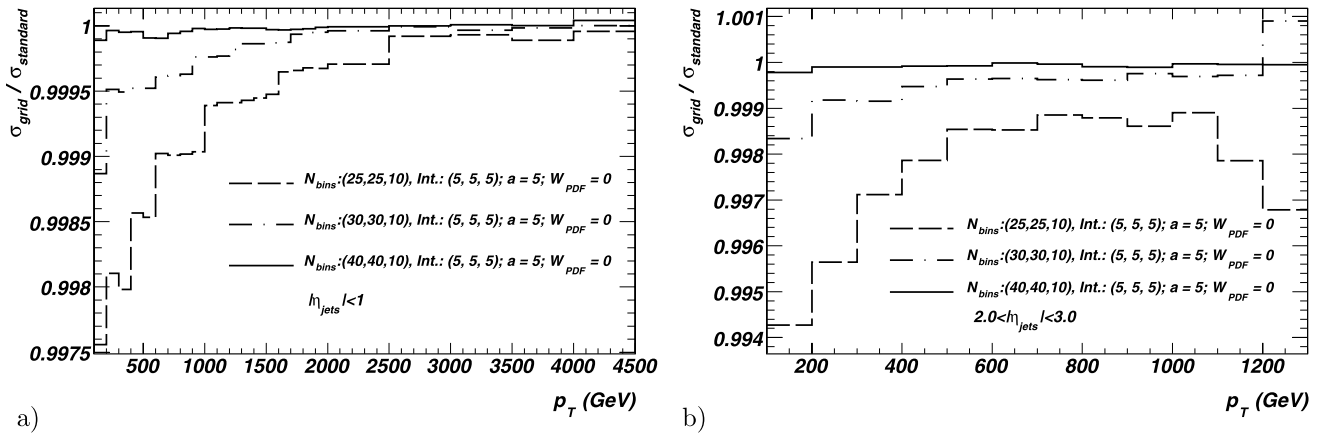


Fig. 3 Ratios of grid and standard calculations of the single inclusive jet p_T spectrum for $0 < y < 1$ (a) and for $2 < y < 3$ (b), illustrating the impact of varying the number of x -bins in the grid. All weight grids

have 10 bins in Q^2 , a coordinate transform parameter $a = 5$ and fifth order interpolation. The PDF set is CTEQ6mE

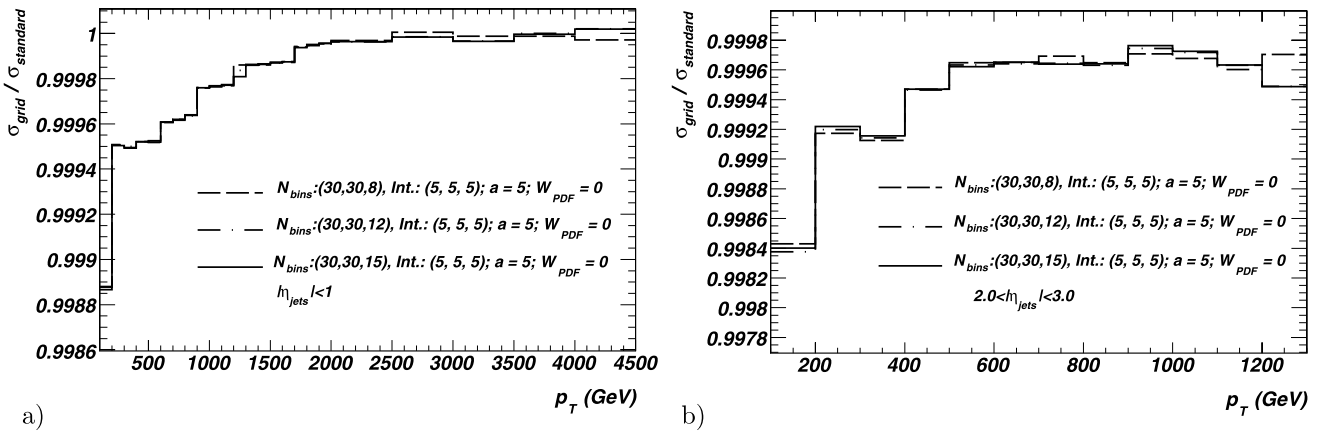


Fig. 4 Ratios of grid and standard calculations of the single inclusive jet p_T spectrum for $0 < y < 1$ (a) and for $2 < y < 3$ (b), illustrating the impact of varying the number of Q^2 -bins in the grid. All weight

grids have 30 bins in x , a coordinate transform parameter $a = 5$ and fifth order interpolation. The PDF set is CTEQ6mE

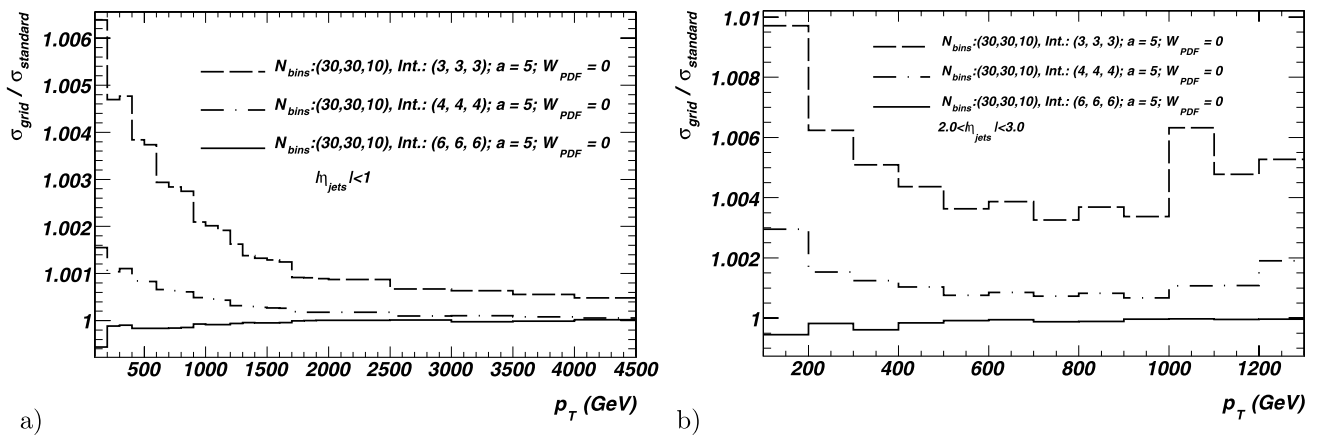


Fig. 5 Ratios of grid and standard calculations of the single inclusive jet p_T spectrum for $0 < y < 1$ (a) and for $2 < y < 3$ (b), illustrating the impact of varying the grid interpolation order. All weight grids

have 30 bins in x , 10 bins in Q^2 and a coordinate transform parameter $a = 5$. The PDF set is CTEQ6mE

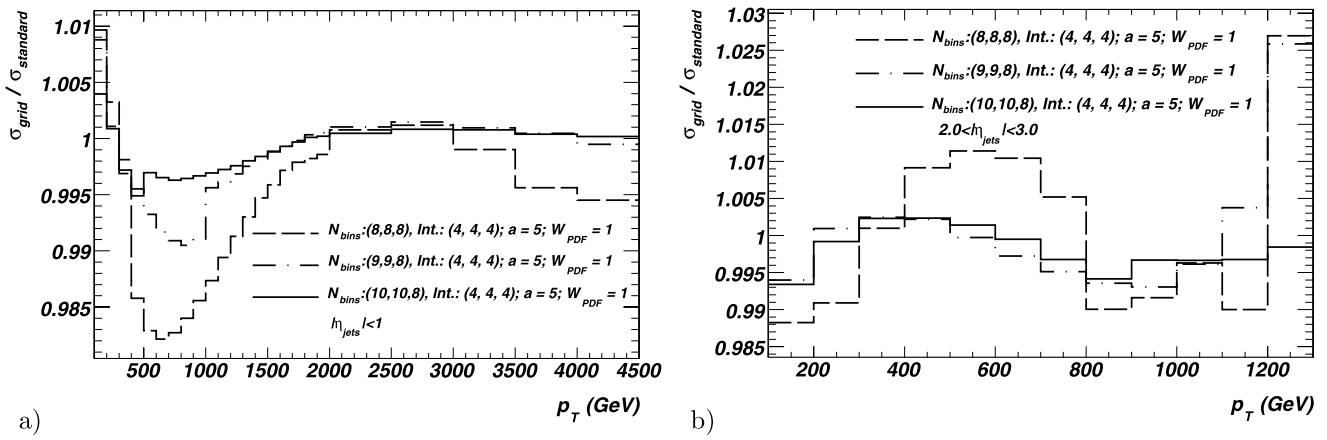


Fig. 6 Ratios of grid and standard calculations of the single inclusive jet p_T spectrum for $0 < y < 1$ (a) and for $2 < y < 3$ (b), illustrating the use of small grids with PDF reweighting. The weight grids have a

low number of x -bins (8, 9, 10), 8 bins in Q^2 , a coordinate transform parameter $a = 5$, fourth order interpolation and PDF reweighting with Eq. 17. The PDF set is CTEQ6mE

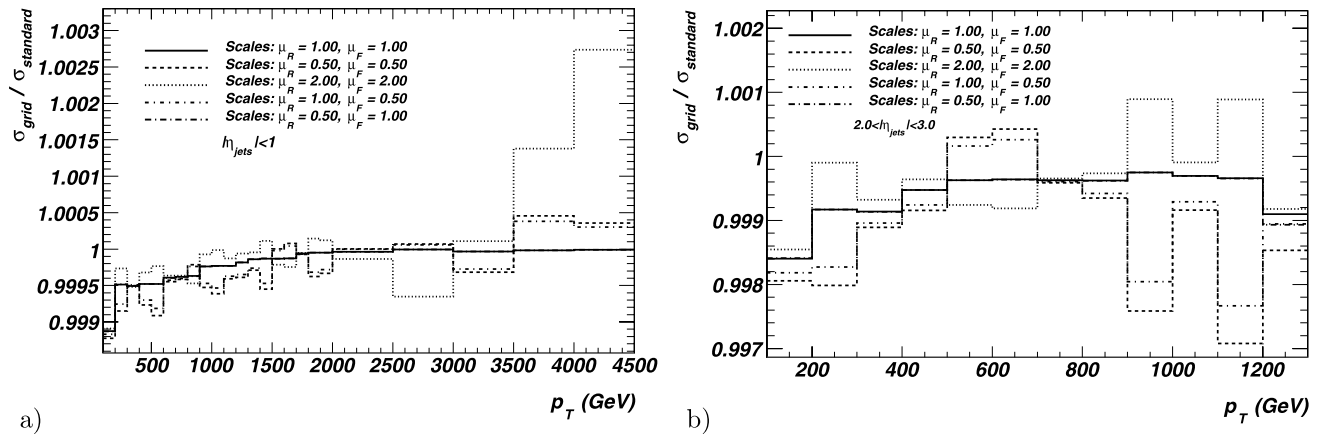


Fig. 7 Ratios of grid and standard calculations of the single inclusive jet p_T spectrum for $0 < y < 1$ (a) and for $2 < y < 3$ (b), with scale variation. The default weight grid is used, with 30 bins in x , 10 bins in Q^2 , a coordinate transform parameter $a = 5$ and fifth order interpolation.

The grid results are based on a posteriori variation of the renormalisation and factorisation scales, using Eq. 10, while the standard results have been obtained separately for each choice of renormalisation and factorisation scale. The PDF set is CTEQ6mE

cross-section at a lower or higher centre-of-mass energy. This procedure works if the coverage in x in the weight grid is large enough. For instance, when lowering the centre-of-mass energy to calculate the jet cross-section at a fixed transverse jet momentum, it might happen that the required large x values are not present in the weight grid produced at a higher centre-of-mass energy. The variation of the centre-of-mass energy has therefore to be done with care by the user.

As an example, the accuracy of the jet cross-section calculation using the default weight grid at a fixed centre-of-mass energy of $\sqrt{s} = 14000$ GeV is investigated. Reference cross-sections are calculated at various centre-of-mass energies, i.e. 1800, 5000, 7000, 10000, 14000, 16000 and 18000 GeV. Since the calculations at the various centre-of-mass energies are statistically independent, each refer-

ence cross-section as well as the default weight grid at $\sqrt{s} = 14000$ GeV needs to be calculated with large event samples. Each of the calculations is done with 50 000 000 events produced with NLOJET++.

In order to make the comparisons more meaningful the jet transverse momentum p_T is transformed to $x_T = 2p_T / \sqrt{s}$. For central jets the variable x_T gives approximately the momentum fraction of the incoming parton with respect to the proton. Fig. 8a) shows the ratio of the cross-section calculated with the standard weight grid produced at $\sqrt{s} = 14000$ GeV to the cross-section calculated at various centre-of-mass energies in the standard way as a function of x_T . For most points, the calculations agree within 2%. The observed fluctuations are statistical.

For large changes in centre-of-mass energy and large x_T values the approximation of the standard grid becomes inac-

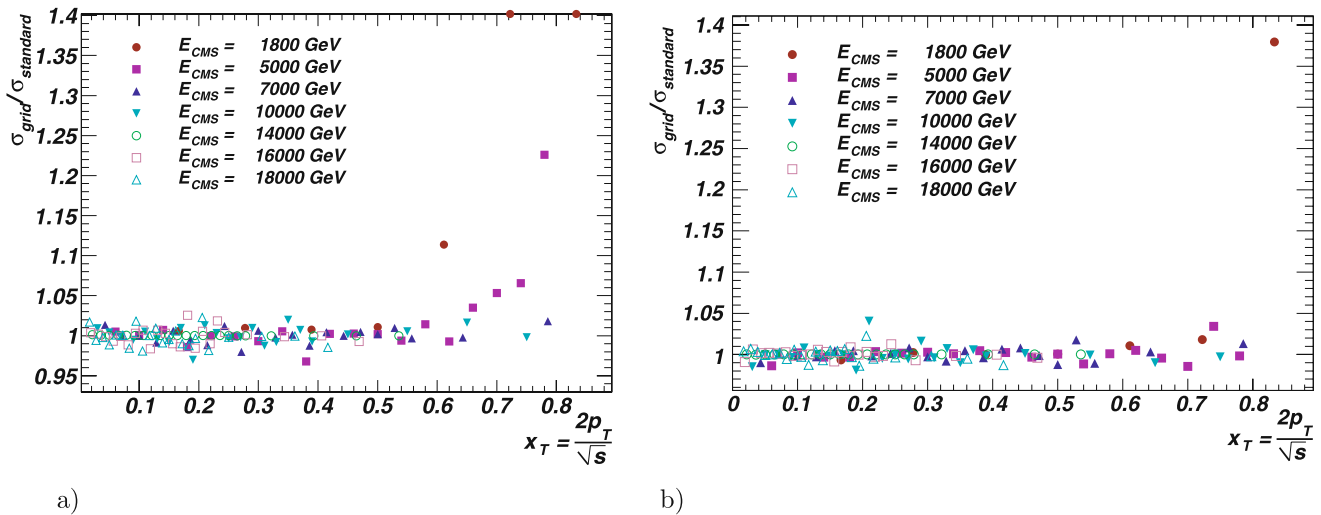


Fig. 8 Ratios of grid and standard calculations of the single inclusive jet $d\sigma/dx_T$ spectrum, with $x_T = 2p_T/\sqrt{s}$, for various centre-of-mass energies. The standard calculation has been performed separately for each centre-of-mass energy, while the grid results are all based on a common $\sqrt{s} = 14000$ GeV grid. The PDF set is CTEQ6mE. (a) The

default grid parameters are used (30 bins in x and 10 bins in Q^2). The last two points for $\sqrt{s} = 1800$ GeV are drawn at 1.4 for better visibility, but their true values are very large. (b) A larger grid with 50 bins in x and 20 bins in Q^2 is used

curate. For instance, for $\sqrt{s} = 1800$ GeV and $x_T = 0.6$ the weight grid calculation gives a result that is 10% higher than the standard calculation. This discrepancy increases further for large x_T values. The ratio of the last two x_T values becomes very large.¹²

Figure 8(b) shows the result for a larger grid using 50 bins in x and 20 bins in Q^2 . With such a grid the deviations are mostly reduced to statistical fluctuations. Only the largest x_T value for the lowest centre-of-mass energy exhibits a deviation by 30 % from the standard calculation.

A small grid with a PDF weighting leads to large discrepancies to the standard calculation and cannot be used.

In conclusion, the grid technique gives a good accuracy to compute the jet cross-section at various centre-of-mass energies. For very high transverse momenta and extreme centre-of-mass variations a large grid might be required.

4.3 W-boson production at hadron colliders

To further demonstrate the performance of the weight grid method, the production of W -bosons at LHC energies is taken as example. The observable that will be examined is the transverse-momentum distribution of the positron from W^+ -boson decays, when the positron is either central $|\eta| \leq 0.5$, or very forward, $|\eta| \geq 3.0$.

As in the previous section, a default weight grid is defined and variations in a few parameters are studied. The default weight grid consists of 25 bins in x . The points are

distributed according to Eq. 1 with $a = 5$ and a fifth order interpolation is used. No PDF weight (see Eq. 17) is used. The cross-sections are calculated with the factorisation and renormalisation scale fixed to the mass of the W -boson. Therefore, the weight grid need only be two dimensional.

The influence of the number of bins in x is shown in Fig. 9. If the number of bins in x is too small ($N_{\text{bins}} = 20$) the cross-section is reproduced to about 0.5% in the central region and 0.2% in the forward region.

For the default weight grid, lowering the interpolation order from $n = 5$ to $n = 4$ results in an accuracy loss of about 0.2% over much of the p_T range, as shown in Fig. 10. The accuracy for positrons with low transverse momenta degrades to 0.8%. The good precision for $n = 5$ can only be improved using $n = 7$.

Figure 11 shows the dependence on the grid spacing parameter corresponding to the parameter a in Eq. 1. The x -values in the cross-section calculation are not large and consequently a fine spacing at large x (corresponding to a large a parameter) is not needed and the result improves for low a values. An accuracy of better than 0.1% is achieved for all variations.

Finally, Fig. 12 shows that, if a PDF weighting is used, it is possible to use very small grid sizes. For a weight grid with only eight x -bins an accuracy of 0.1% can be achieved. In this case the gain in accuracy is small when increasing the number of x -bins. Only in the forward region and for high transverse energies the increase in the number of x -bins is beneficial.

In summary, a sufficient accuracy is achieved with about 25 x -bins and a fifth order interpolation. A grid spacing of $a = 1$ is sufficient.

¹²In Fig. 8(a) they are drawn at 1.4 for better visibility of the rest of the points.

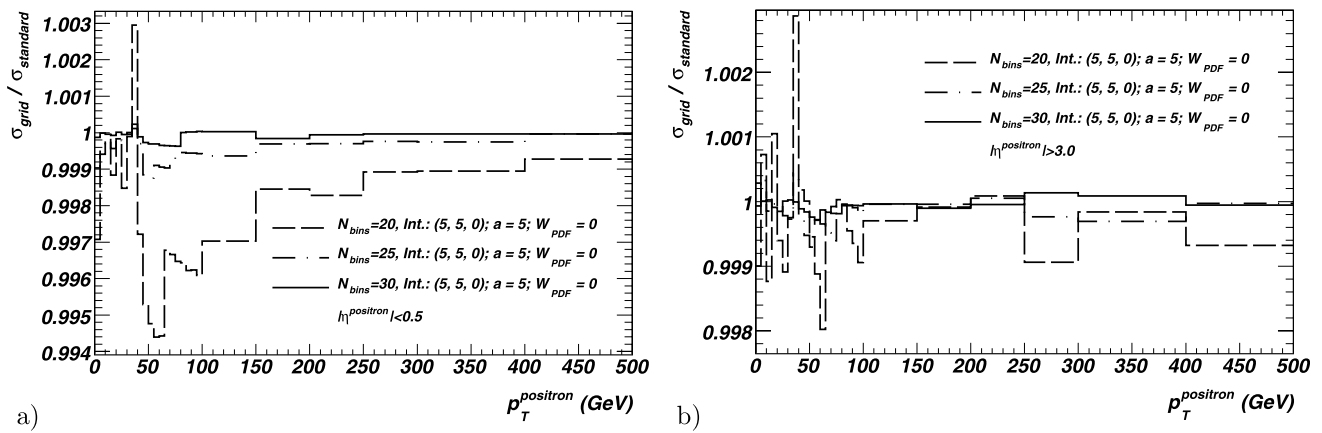


Fig. 9 Ratios of grid and standard calculations of the positron p_T spectrum in W^+ -boson production, for $|\eta_{e^+}| < 0.5$ (a) and for $|\eta_{e^+}| > 3$ (b). Results are shown for three weights grids with different

numbers of x bins. All grids use a coordinate transform parameter $a = 5$ and fifth order interpolation. The PDF set is CTEQ6mE

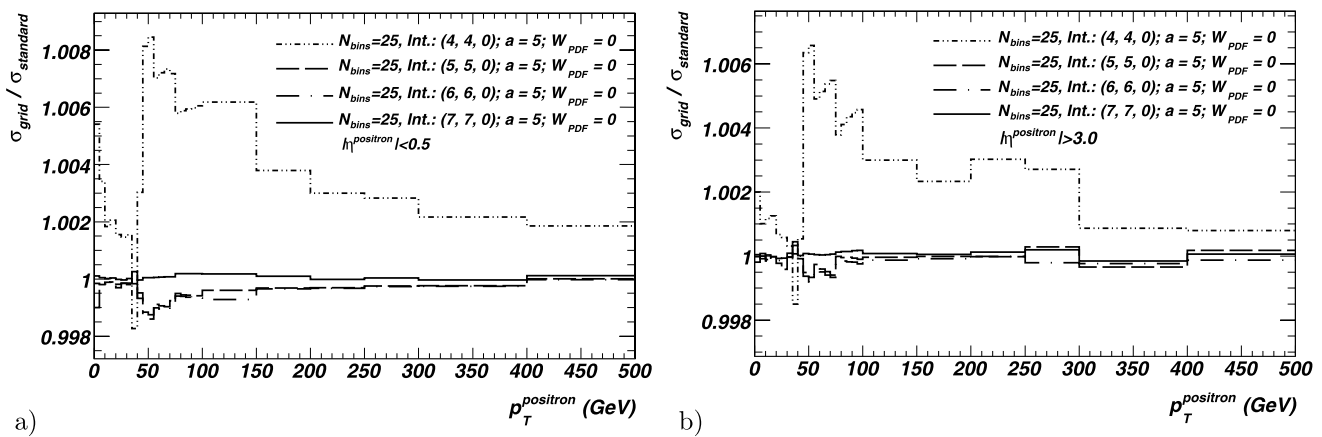


Fig. 10 Ratios of grid and standard calculations of the positron p_T spectrum in W^+ -boson production, for $|\eta_{e^+}| < 0.5$ (a) and for $|\eta_{e^+}| > 3$ (b). Results are shown for four grids, each with a different

interpolation order. All grids have 25 bins in x and a coordinate transform parameter $a = 5$. The PDF set is CTEQ6mE

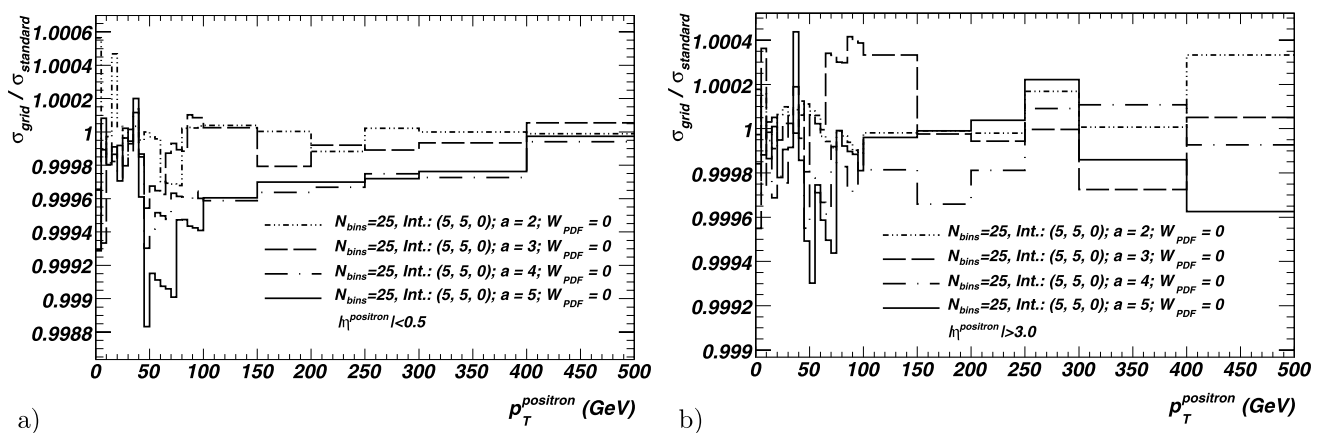


Fig. 11 Ratios of grid and standard calculations of the positron p_T spectrum in W^+ -boson production, for $|\eta_{e^+}| < 0.5$ (a) and for $|\eta_{e^+}| > 3$ (b). Results are shown for four grids, each with a different

coordinate transform parameter, a . All grids have 25 bins in x and fifth order interpolation. The PDF set is CTEQ6mE

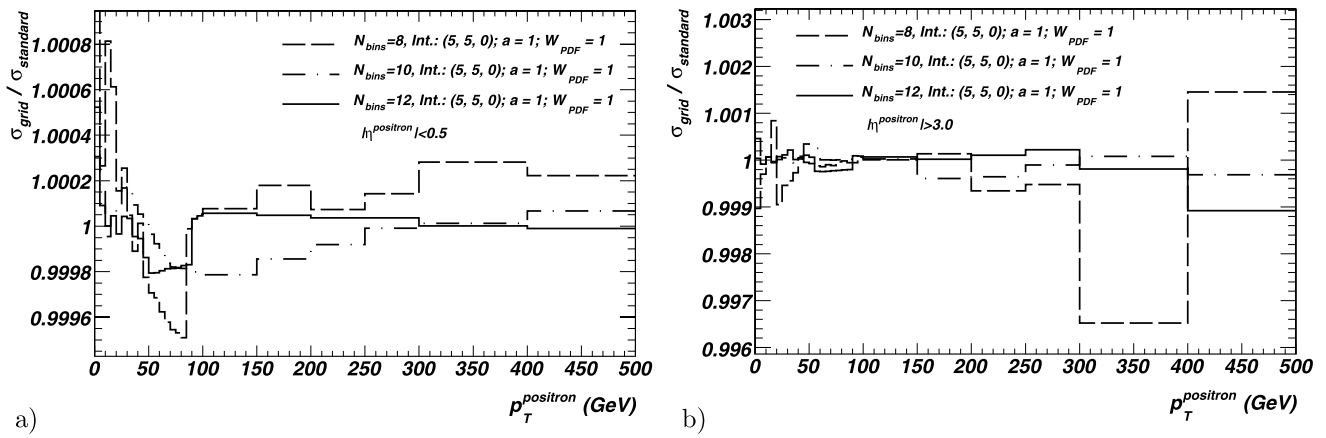


Fig. 12 Ratios of grid and standard calculations of the positron p_T spectrum in W^+ -boson production, for $|\eta_{e^+}| < 0.5$ (a) and for $|\eta_{e^+}| > 3$ (b). Results are shown for grids with a reduced number of

x bins and PDF reweighting. All the grids use fifth order interpolation and a coordinate transform parameter $a = 1$. The PDF set is CTEQ6mE

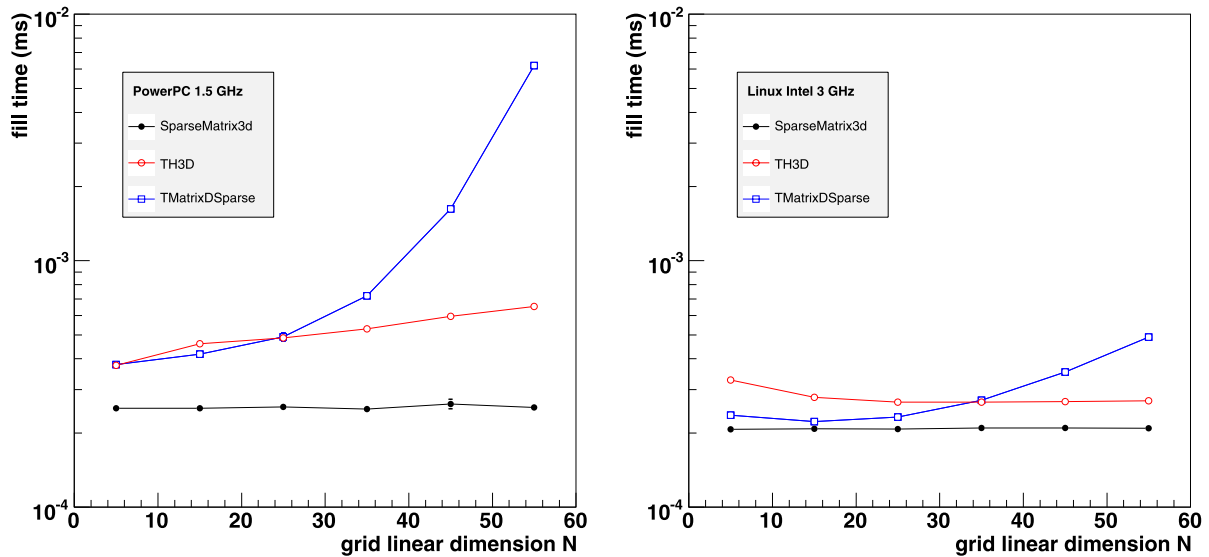


Fig. 13 The time per call for filling grid classes based on various grid architectures on a 1.5 GHz PowerPC (left) and 3 GHz Linux PC (right)

4.4 CPU and computer memory performance

The execution time for each call to the filling routine for the grid has been studied on a 1.5 GHz PowerPC and a 3 GHz Intel Xeon running Linux, using a dummy structure with N points in each dimension. Figure 13 shows the performance for various grid architectures. The grids are based on either the ROOT TH3D class, the custom sparse class (SparseMatrix3d) described in Sect. 3, or the TMatrixDSparse class which implements the 2-dimensional Harwell-Boeing matrix representation. In the latter case, a sparse 1-dimensional structure of TMatrixDSparse matrices using the classes of the SparseMatrix3d has been

used to create a sparse 3-dimensional structure. As expected the Harwell-Boeing based class is very quick for filling when the grid is small, but as the grid size becomes larger, since the occupation is reasonably large, the number of entries that must be examined becomes large and the filling time increases rapidly. For the TH3D and custom sparse structures, the filling time is largely independent of the grid size.

The reduction in memory occupied by the custom sparse grid structure after trimming away unoccupied elements is illustrated in Fig. 14. The bottom-left plot shows the absolute size of the stored elements in MBytes, both before, and after trimming away unfilled elements. The top-left plot

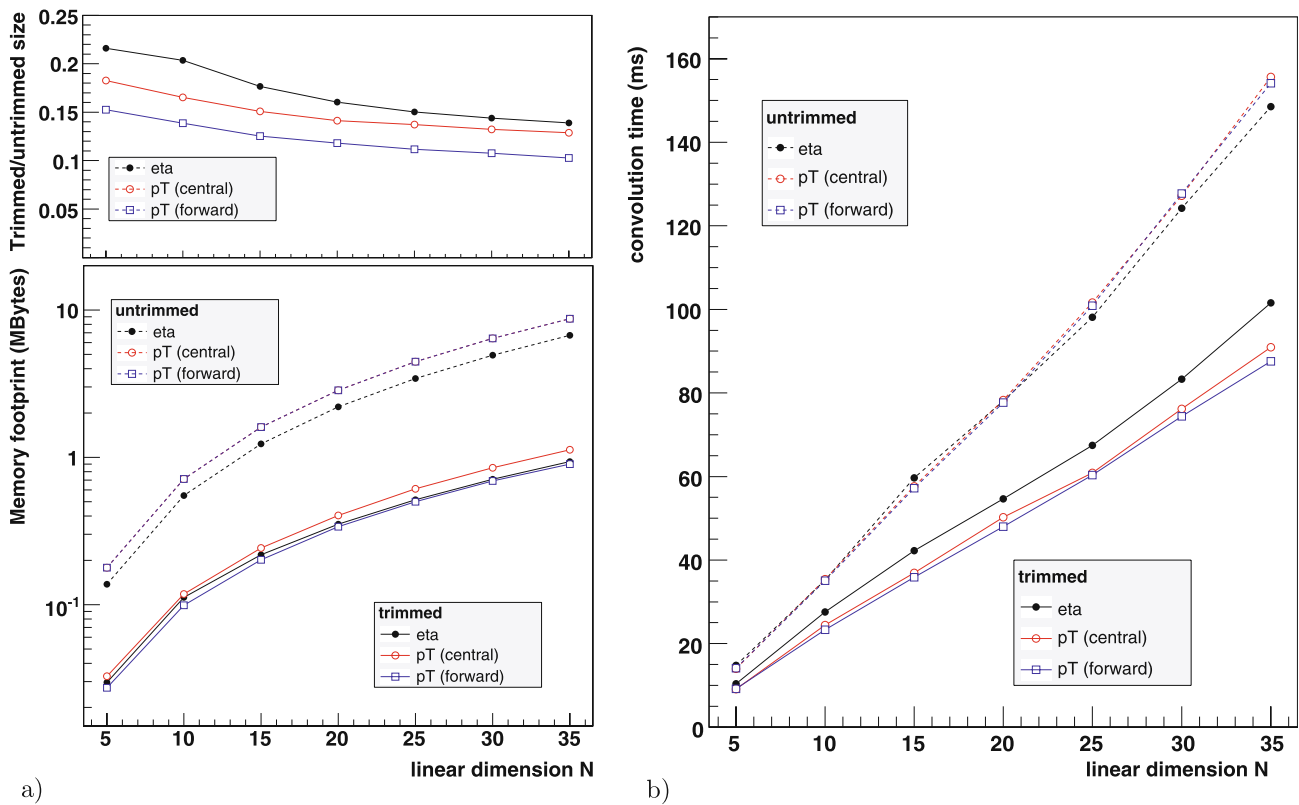


Fig. 14 (a) Memory used for the default grid architecture using a custom sparse grid (untrimmed) and after removing the unoccupied elements. The *top figure* shows the ratio of the reduced to the full case. (b) Time needed to calculate the cross-section by convoluting the co-

efficients on the grid with PDFs and α_s . The convolution times are measured on a 1.5 GHz PowerPC for a default grid. The memory and the CPU time performance is evaluated for the W -boson cross-sections as a function of the electron rapidity and transverse momentum

shows the fraction of the total, untrimmed grid size, occupied by the filled elements. As the grid spacing decreases, the overall grid size naturally increases.

The execution time using the grid to perform the final cross-section calculation including the PDF convolution has also been studied using a 1.5 GHz PowerPC. The results are based on calculations of differential cross-sections with respect to the positron pseudo-rapidity and transverse momentum in W -boson production using MCFM [16, 17], as presented in Sect. 4.3. The cross sections involve 20 and 24 bins for the lepton pseudo-rapidity and transverse energy distributions respectively. Figure 14(b) shows the convolution time for grids with N bins in dimensions x_1 and x_2 for the sparse structure. Results are given for the trimmed and untrimmed structures. In the case of the untrimmed grid, all data elements are retained in the convolution, even those with no entries.

Excluding the unfilled data elements in the convolution improves the convolution time by a factor approaching two. In addition, we see that the convolution time varies approximately linearly with the grid linear dimension. This is because the most costly part of the convolution is the calculation of the PDF at the grid nodes. With independent grid

nodes for x_1 and x_2 , there are $2N$ evaluations of the PDF for each observable bin, and so the convolution scales linearly with N .

In conclusion, the custom sparse structure using trimmed blocks gives the best performance.

5 Application example: calculation of NLO QCD uncertainty for inclusive jet cross-sections for proton proton collisions at various centre-of-mass energies

As an example in this section the uncertainties of the inclusive jet cross-section in the central region ($0 < y < 1$) are evaluated from the default grid obtained at a centre-of-mass energy of $\sqrt{s} = 14000$ GeV. The jet cross-sections are calculated at various centre-of-mass energies. The most recent PDF parameterisations along with their associated uncertainties are used, i.e. CTEQ6.6 [22], MSTW2008 [23], HERAPDF01 [24] and NNPDF [25].

Figure 15 shows the effect of the PDF uncertainty from CTEQ6.6 (a), MSTW2008 (b), HERAPDF01 (c) and NNPDF (d) on the inclusive jet cross-section with respect to the central value of the somewhat older PDF, CTEQ6mE

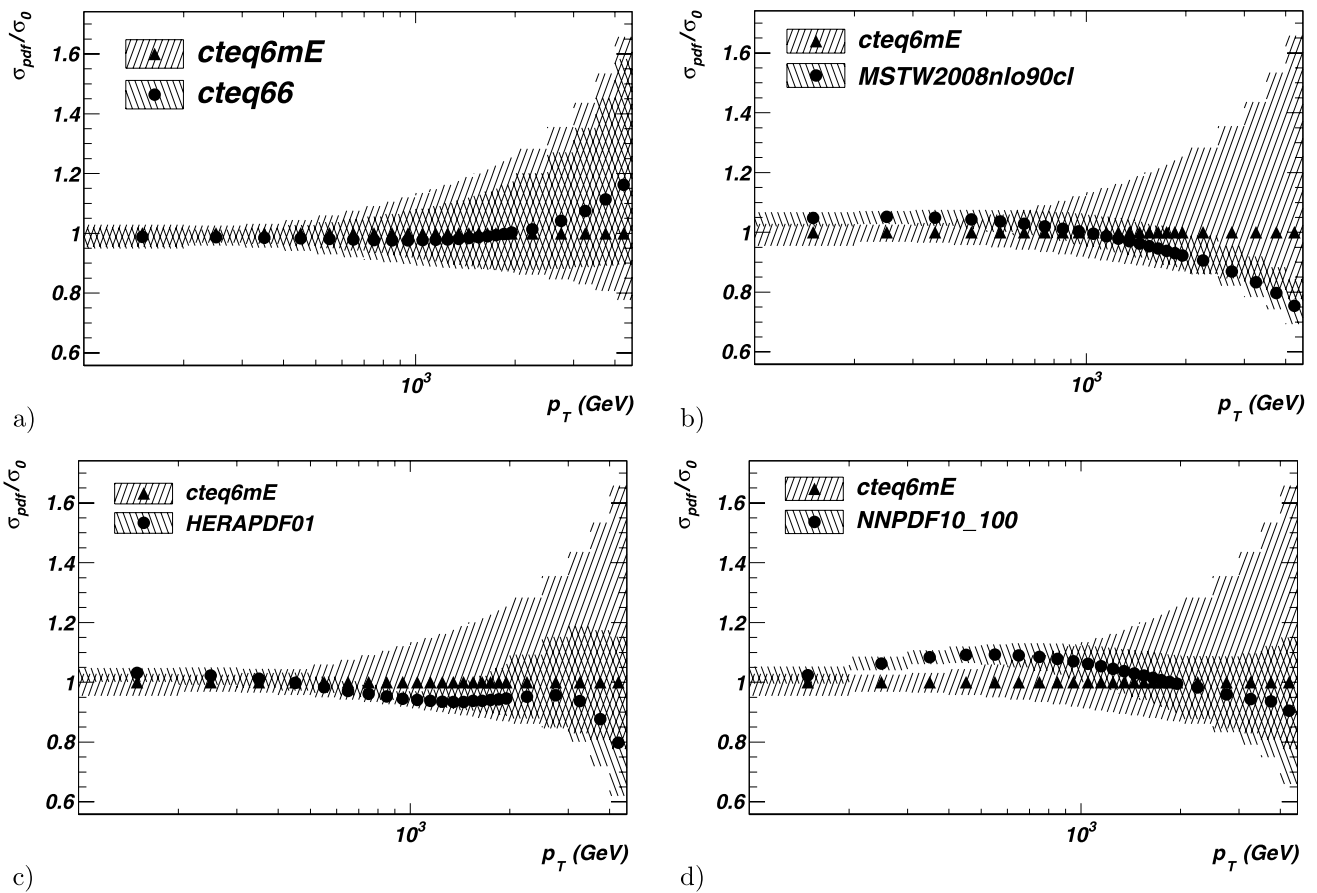


Fig. 15 PDF uncertainty of the inclusive jet cross-section for jets within $0 < y < 1$ as a function of the transverse jet momentum p_T for a centre-of-mass energy of $\sqrt{s} = 14000$ GeV. Shown is the jet cross-section uncertainty induced by the CTEQ6mE PDF and the CTEQ6.6

(a) the MSTW2008 (b), the HERAPDF (c) and the NNPDF PDF (d). The reference cross-section σ_0 is the one obtained by the central value of the CTEQ6mE PDF. The default PDF is indicated by a marker

[18]. The uncertainty from the CTEQ6mE PDF is also overlaid. The band illustrates the result of adding the jet cross-sections obtained for each of the PDF variations.¹³ The marker indicates the central value. Figure 16 shows the PDF uncertainty together with the renormalisation and factorisation scale uncertainty added in quadrature with respect to the central value for each of the PDFs.

For $p_T < 1000$ GeV the jet cross-section obtained with CTEQ6.6 is about 2% smaller than from CTEQ6mE. Above this value the CTEQ6.6 cross-section increases with respect to the one from CTEQ6mE as the jet p_T increases. At $p_T = 2000$ GeV it is about equal and at

$p_T = 4000$ GeV it is about 10% larger. The uncertainty is reduced for the CTEQ6.6 PDF. The uncertainty is about 3% for $p_T < 500$ GeV, about 8% at $p_T = 1000$ GeV and about 20% at $p_T = 3000$ GeV.

The MSTW2008 PDF gives a jet cross-section that is 5% larger than the one obtained with CTEQ6mE at $p_T < 500$ GeV and is about the same at $p_T = 1000$ GeV and then further decreases. The uncertainty is only about 2% for $p_T < 500$ GeV and then increases to about 6% at $p_T = 1000$ GeV. The MSTW2008 PDF gives a smaller uncertainty than the CTEQ6.6 PDF. It seems that the differences between the jet cross-section calculated with CTEQ6.6 and MSTW2008 are a bit larger than the individual uncertainties.

The result obtained with the HERAPDF01 is more similar to the one obtained from MSTW2008 than the one from CTEQ6.6. At low p_T the central value is about 2% higher than the one from CTEQ6mE. In the region $500 < p_T < 1500$ GeV the HERAPDF01 predicts a lower jet cross-section than the other PDFs. The uncertainty

¹³The uncertainty band is obtained using Eq. 51 and Eq. 52 in Ref. [23] for the HERA, MSTW and the CTEQ PDFs. This formula has also been suggested earlier in Ref. [26]. For the NNPDF Eq. 164 in Ref. [25] is used. The uncertainty in the NNPDF corresponds to the standard deviation of all variations, while in the case of the other PDFs it corresponds to the 90% confidence limit. For better comparison, the uncertainties of the NNPDF and the HERAPDF01 have been scaled up using Eq. 165 in Ref. [25].

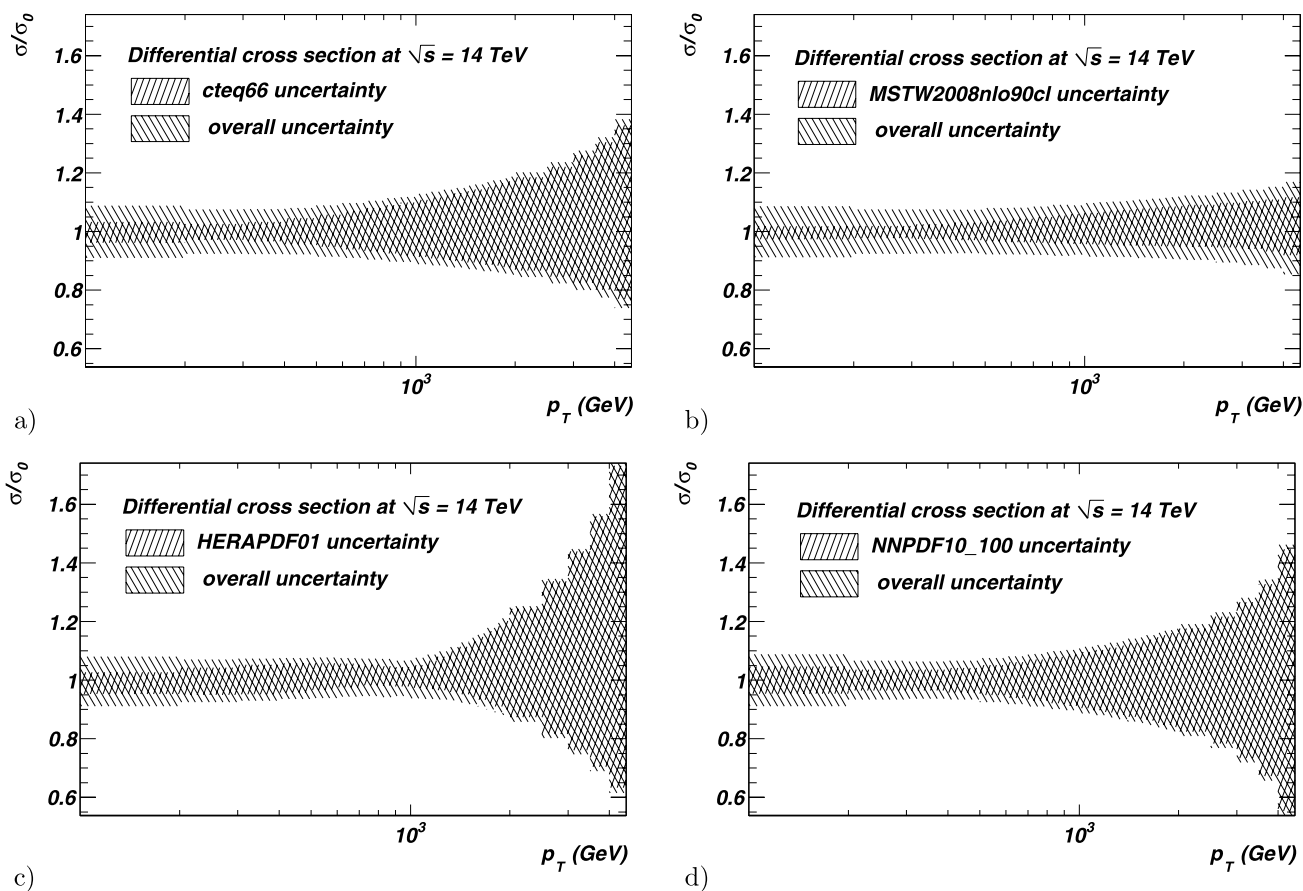


Fig. 16 Uncertainty of the inclusive jet cross-section for jets within $0 < y < 1$ as a function of the transverse jet momentum p_T at fixed centre-of-mass energy $\sqrt{s} = 14000$ GeV. Shown is the ratio of the cross-section with varied PDFs and renormalisation and factorisation scales (σ) to the cross-section calculated with the central value of each

PDF set and no scale variation, i.e. $\mu_r = \mu_f = 1 (\sigma_0)$. The inner uncertainty band shows only the PDF uncertainty. The outer band shows the PDF and the scale uncertainty added in quadrature. The uncertainty of CTEQ6.6 is shown in (a), of MSTW2008 in (b), of HERAPDF in (c) and of NNPDF in (d)

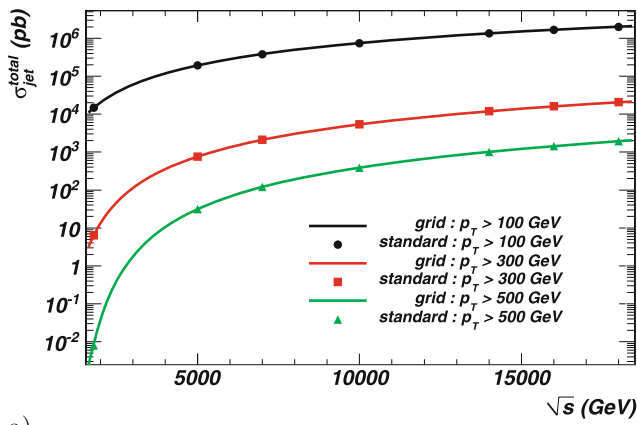
is about 5% for $p_T < 1000$ GeV and then increases to about 20–40% at $p_T = 3000$ GeV. The small uncertainty of the jet cross-section calculated with the HERAPDF01 is remarkable, since only DIS data are used. However, model and parametrization uncertainties are not included in this cross-section calculation. The MSTW and CTEQ sets do not yet include the most recent HERA data. The NNPDF predicts jet cross-sections that are 5–10% higher than the one from the other PDFs; in particular in the region $300 < p_T < 1000$ GeV. The uncertainty is about 5% at low p_T , 10% at 1000 GeV and 20–30% at 3000 GeV.

The overall uncertainty, i.e. including the PDF and the scale variation added in quadrature, is shown in Fig. 16. It is about 8% up to a p_T of about 1000 GeV and then increases towards higher p_T . It is about 20–30% at $p_T = 3000$ GeV. For very high p_T the PDF uncertainty dominates.

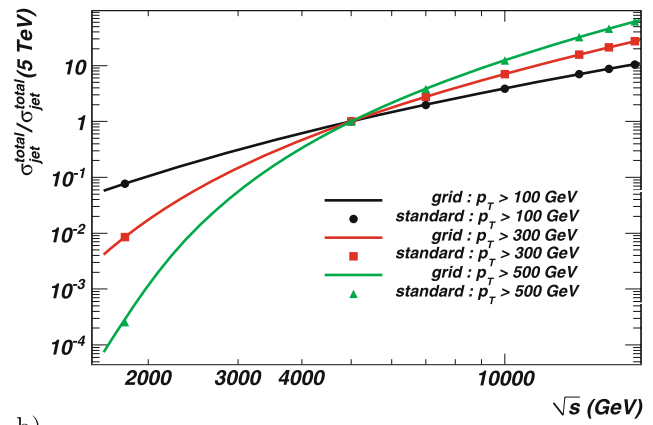
Figure 17(a) shows the total inclusive cross-section for central jets ($0 < y < 1$) integrated for $p_T > 100$ GeV, $p_T > 300$ GeV and $p_T > 500$ GeV as a function of the

centre-of-mass energy. The markers denote the reference cross-section calculated in the standard way. The lines are obtained from a weight grid produced at $\sqrt{s} = 14000$ GeV. The cross-section from the default weight grid reproduces the reference cross-sections within 1–2% (see also Sect. 4.2). For each jet transverse momentum threshold the total jet cross-section rises with increasing centre-of-mass energy. Figure 17(b) shows the centre-of-mass energy dependence of the jet cross-section normalised to the jet cross-section at 5000 GeV for each jet transverse momentum threshold. As expected the centre-of-mass energy dependence is strongest for high transverse jet momenta.

Figure 18 shows for each of the considered PDF sets the PDF uncertainty along with the renormalisation and factorisation scale uncertainty added in quadrature for jets with $p_T > 100$ GeV as a function of the centre-of-mass energy \sqrt{s} . Both the PDF and the scale uncertainties only depend weakly on the centre-of-mass energy. For high centre-of-mass energies the uncertainties are a bit smaller.



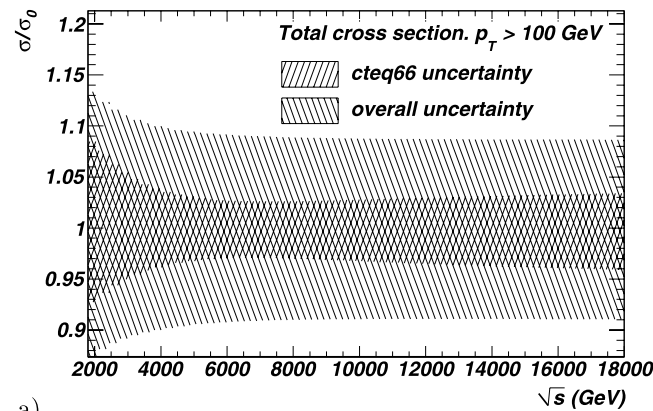
a)



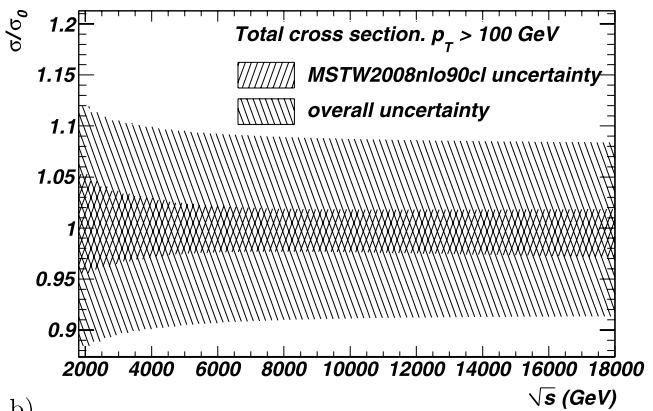
b)

Fig. 17 (a) Inclusive jet cross-section for jets within $0 < y < 1$ and with transverse jet momenta $p_T > 100$ GeV, $p_T > 300$ GeV and $p_T > 500$ GeV as a function of the centre-of-mass energy \sqrt{s} . (b) Shows the same as (a), but all results are normalised to $\sqrt{s} = 5000$ GeV.

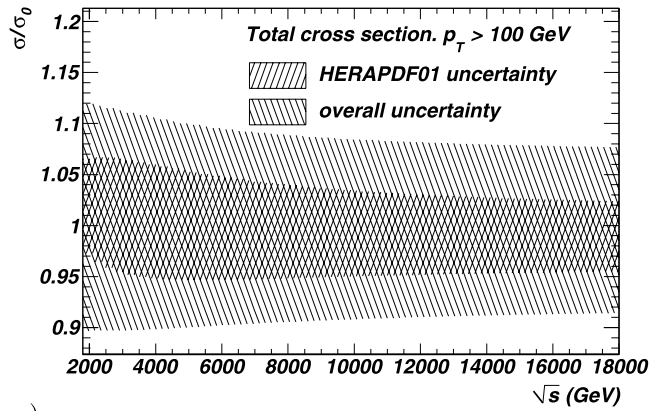
The *markers* indicate the results calculated at each centre-of-mass energy in the standard way. The *lines* indicate the results deduced from the default weight grid produced for a centre-of-mass energy at $\sqrt{s} = 14000$ GeV



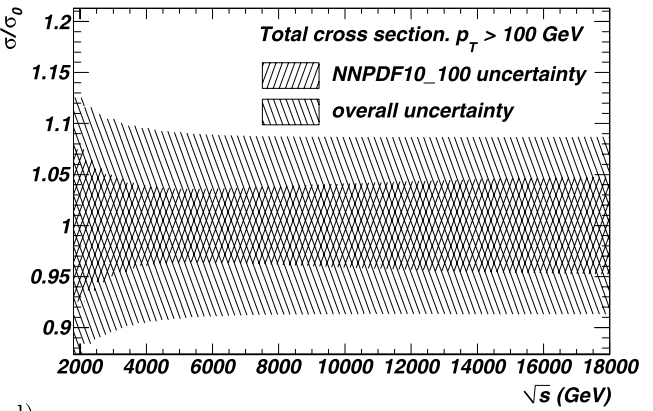
a)



b)



c)



d)

Fig. 18 Uncertainty of the inclusive jet cross-section for jets with transverse momenta $p_T > 100$ GeV and within $0 < y < 1$ as a function of the centre-of-mass energy \sqrt{s} . Shown is the ratio of the cross-section with varied PDFs and renormalisation and factorisation scale (σ) to the cross-section calculated with the central value of each PDF set and no

scale variation (σ_0). The inner uncertainty band shows only the PDF uncertainty. The outer band shows the PDF and the scale uncertainty added in quadrature. The uncertainty from CTEQ6.6 is shown in (a), from MSTW2008 in (b), from HERAPDF in (c) and from NNPDF in (d)

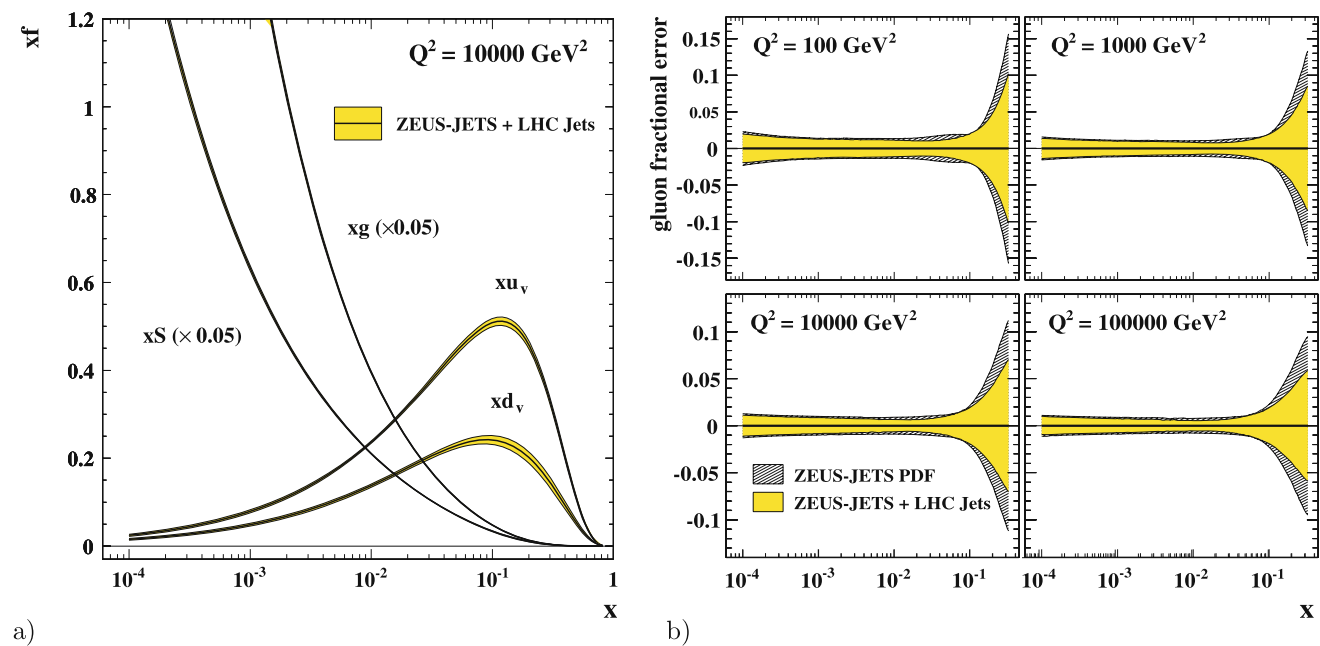


Fig. 19 The distributions of the up-valence, down-valence, total sea and gluon PDFs (a), and the fractional uncertainty on the gluon distribution at a number of Q^2 values (b) as a function of the parton momentum fraction x . The results from the fit using weight grids to include simulated LHC jet data is shown by the shaded band. For comparison,

in (b), the results of the ZEUS NLO QCD fit are also shown, indicated by the hatched band. The simulated LHC jet data included in the new fit assume a statistical uncertainty corresponding to an integrated luminosity of 10 fb^{-1} , uncorrelated systematics of 5% and a jet energy scale uncertainty of 1%

6 Application example: PDF fit including DIS data and jet production data at hadron colliders

An important application of the method outlined above, is the consistent inclusion of final state measurements from hadronic colliders into the final extraction of PDFs by NLO QCD fits. Measurements of final states—such as jet production or the production of lepton pairs via the Drell-Yan process—can provide important additional constraints on the proton PDFs, complementary to those from inclusive DIS data.

As a simple “proof-of-principle” example, the grid technique outlined in this paper has been used to include simulated LHC jet data into a NLO QCD fit. The fit framework used here is based on the recent ZEUS-JETS PDF, derived from a fit to inclusive DIS and jet data from HERA. Jet cross-sections from the TEVATRON or any other data than that from HERA are not used. Full details of the data-sets, PDF parameterisation and other assumptions are given elsewhere [5].

To represent the LHC data for inclusion in the fit, jet production from proton-proton collisions at a centre-of-mass energy of 14000 GeV was simulated using the JETRAD [27] program, using the CTEQ6.1 PDF [28]. Single inclusive jet cross sections, differential in p_T , were obtained in three regions of rapidity: $0 < |y| < 1$, $1 < |y| < 2$ and $2 < |y| < 3$. A grid with default parameters, as described in Sect. 4.1,

was produced and interfaced to the ZEUS NLO QCD fit program. Several fits were performed, using different assumptions on the statistical and systematic uncertainties on the simulated data. The PDF uncertainties were calculated using the Hessian method [29, 30], with $\Delta\chi^2 = 1$.¹⁴

A representative result is shown in Fig. 19. In this example, the statistical uncertainty on the simulated LHC jet data corresponds to an integrated luminosity of 10 fb^{-1} and uncorrelated systematic uncertainties have been assumed to be at a level of 5%. A precise jet energy scale uncertainty of 1% (corresponding to $\sim 5\text{--}15\%$ on the generated cross-sections) has also been assumed, and is included as a correlated systematic in the fit. Figure 19(a) shows the up-valence, down-valence, total sea and gluon PDF distributions as a function of x , at $Q^2 = 10000 \text{ GeV}^2$. The shaded band shows the results of the fit including the simulated LHC jet data. In Fig. 19(b), the fractional uncertainties on the gluon PDF, at a number of Q^2 values, are shown.

¹⁴Note that a using $\Delta\chi^2 = 1$ in the Hessian method is generally considered to underestimate the PDF uncertainties. However, the main aim of this study is to provide a proof-of-principle example of the use of the grids discussed in this paper, and not to provide qualitative estimates of weight expected PDF uncertainties. Furthermore, all fits shown in this section have used the same definition of the PDF uncertainties, such that any comparison should still be valid.

Comparison with the results from a fit which does not include the simulated LHC jet data indicates that some constraint on the high- x gluon could be provided by the LHC single inclusive jet data.¹⁵ However, this is reliant on a very precise knowledge of the jet energy scale.

In fact, according to this study, a precise knowledge of the jet energy scale is the key factor. Other fits, which assumed a smaller integrated luminosity (1 fb^{-1}) or larger uncorrelated systematics (10%), still indicated an improvement on the gluon uncertainties, provided the jet energy scale uncertainty was kept at a level of $\sim 1\%$. However, fits in which the latter uncertainty was assumed to be larger, indicated little or no improvement in the gluon uncertainty compared to the reference. More details can be found in Ref. [33].

Such precision on the jet energy scale is achievable, but will require a lot of experimental work on the understanding of the LHC detectors. The inclusion of TEVATRON jet cross-sections in the NLO QCD fit might provide further constraints. However, it may be the case that ratios of jet cross sections—for example, in different rapidity regions—may have substantially smaller systematic uncertainties, while retaining sensitivity to the gluon density in the proton. Further constraints on the proton PDFs are also expected from Drell-Yan data measured at LHC or any other data than those from HERA. Such data sets can now be consistently included in NLO QCD fits.

7 Conclusions

A technique has been developed to store the perturbative coefficients calculated by a NLO QCD Monte Carlo program in a look-up table (grid) allowing for a posteriori inclusion of an arbitrary parton density function (PDF) set and of alternative values of the strong coupling constant as well as for a posteriori variation of the renormalisation and factorisation scale. This extends a technique that has already been successfully used to analyse HERA data to the more demanding case of proton-proton collisions at LHC energies.

The technique can be used to constrain PDF uncertainties by allowing the consistent inclusion of final state observables in global QCD PDF fit analyses. This will help to

¹⁵Note that the fit without the simulated LHC jet data is not identical to the ZEUS-JETS fit since the standard ZEUS fit [5] uses the Offset method to determine the PDF uncertainties. The ZEUS fit shown here is a modified version of the published analysis, with uncertainties determined using the Hessian method, with $\Delta\chi^2 = 1$. Different treatments of experimental uncertainties in PDF analyses are discussed extensively elsewhere [29–32].

increase the sensitivity of the LHC to find new physics as deviations from the Standard Model predictions.

An accuracy of better than 0.1% can be reached with reasonably small look-up tables for the single inclusive jet cross-section in the central rapidity region $|y| < 1$, for jet transverse momenta (p_T) from 100 to 4500 GeV and about 0.2% for jets in the forward rapidity region $2 < y < 3$. Similar accuracy can be achieved for the differential cross-sections in rapidity and transverse momentum of electrons produced in Z and W -boson decays. This was examined in the central $y < 0.5$ and very forward $y > 3$ regions for transverse momentum up to $p_T < 500$ GeV.

The look-up tables provide a powerful tool to quickly evaluate the PDF and scale uncertainties of the cross-section at various centre-of-mass energies. The most recent PDFs predict jet cross-sections in the central rapidity region within a few percent accuracy over a large range of jet transverse momenta.

This technique has been successfully applied to a PDF fit using inclusive deep-inelastic scattering and jet data measured at the electron-proton collider HERA and using simulated LHC jet cross-sections. An improvement on the uncertainty of the gluon density can only be achieved if the jet energy scale is very precisely known. A more comprehensive analysis will be possible in the future, since the grid technique can be applied to most of the available NLO QCD calculations.

Acknowledgements We would like to thank J. Campbell, T. Kluge, Z. Nagy, K. Rabbertz, T. Schörner-Sadenius, M.H. Seymour, P. Uwer and M. Wobisch for useful discussions on the grid technique and A. Vogt for discussions on moment-space techniques. We thank Z. Nagy for help and support with NLOJET++ and J. Campbell for help and support with MCFM. C. Gwenlan and M. Sutton would like to thank the UK Science and Technology Facilities Council for support. G.P. Salam would like to thank Rutgers University for hospitality while this work was being completed. F. Siegert would like to thank CERN for the Summer Student Program. P. Starovoitov would like to thank for the support given from INTAS. We would also like to thank J. Ribassin for the work he did towards the MCFM interface.

Appendix A: Sub-processes for W - and Z -boson production

The production of W - and Z -bosons in proton-proton collisions involves flavour-dependent electro-weak couplings. Therefore, the number of sub-processes that need to be defined is larger than in the case of jet production. To reduce the number of sub-processes as much as possible, quarks are assumed to be massless and the CKM matrix elements [34, 35] to describe the contributions of the various quark flavours are used.

In the case of Z-boson production 12 combinations of initial state partons need to be distinguished:

$$\begin{aligned}
 U\bar{U}: & F^{(0)}(x_1, x_2, Q^2) = U_{12}(x_1, x_2), \\
 D\bar{D}: & F^{(1)}(x_1, x_2, Q^2) = D_{12}(x_1, x_2), \\
 \bar{U}U: & F^{(2)}(x_1, x_2, Q^2) = U_{21}(x_1, x_2), \\
 \bar{D}D: & F^{(3)}(x_1, x_2, Q^2) = D_{21}(x_1, x_2), \\
 gU: & F^{(4)}(x_1, x_2, Q^2) = G_1(x_1)U_2(x_2), \\
 g\bar{U}: & F^{(5)}(x_1, x_2, Q^2) = G_1(x_1)\bar{U}_2(x_2), \\
 gD: & F^{(6)}(x_1, x_2, Q^2) = G_1(x_1)D_2(x_2), \\
 g\bar{D}: & F^{(7)}(x_1, x_2, Q^2) = G_1(x_1)\bar{D}_2(x_2), \\
 Ug: & F^{(8)}(x_1, x_2, Q^2) = U_1(x_1)G_2(x_2), \\
 \bar{U}g: & F^{(9)}(x_1, x_2, Q^2) = \bar{U}_1(x_1)G_2(x_2), \\
 Dg: & F^{(10)}(x_1, x_2, Q^2) = D_1(x_1)G_2(x_2), \\
 \bar{D}g: & F^{(11)}(x_1, x_2, Q^2) = \bar{D}_1(x_1)G_2(x_2),
 \end{aligned}
 \tag{18}$$

where g denotes gluons and $U(D)$ denotes up (down)-type quarks. Use is made of the generalised PDFs defined as:

$$\begin{aligned}
 G_H(x) &= f_{0/H}(x, Q^2), \\
 U_H(x) &= \sum_{i=2,4,6} f_{i/H}(x, Q^2), \\
 \bar{U}_H(x) &= \sum_{i=2,4,6} f_{-i/H}(x, Q^2), \\
 D_H(x) &= \sum_{i=1,3,5} f_{i/H}(x, Q^2), \\
 \bar{D}_H(x) &= \sum_{i=1,3,5} f_{-i/H}(x, Q^2), \\
 U_{12}(x_1, x_2) &= \sum_{i=2,4,6} f_{i/H_1}(x_1, Q^2) f_{-i/H_2}(x_2, Q^2), \\
 D_{12}(x_1, x_2) &= \sum_{i=1,3,5} f_{i/H_1}(x_1, Q^2) f_{-i/H_2}(x_2, Q^2), \\
 U_{21}(x_1, x_2) &= \sum_{i=2,4,6} f_{-i/H_1}(x_1, Q^2) f_{i/H_2}(x_2, Q^2), \\
 D_{21}(x_1, x_2) &= \sum_{i=1,3,5} f_{-i/H_1}(x_1, Q^2) f_{i/H_2}(x_2, Q^2),
 \end{aligned}
 \tag{19}$$

where $f_{i/H}$ is the PDF of flavour $i = -6, \dots, 6$ for hadron H and $H_1 (H_2)$ denotes the first or second hadron.

In the case of W^+ -boson production¹⁶ 6 initial state combinations are needed:

$$\begin{aligned}
 \bar{D}U: & F^{(0)}(x_1, x_2, Q^2) = S_{12}(x_1, x_2), \\
 U\bar{D}: & F^{(1)}(x_1, x_2, Q^2) = S_{21}(x_1, x_2), \\
 \bar{D}g: & F^{(2)}(x_1, x_2, Q^2) = \bar{D}_1(x_1)G_2(x_2), \\
 Ug: & F^{(3)}(x_1, x_2, Q^2) = U_1(x_1)G_2(x_2), \\
 g\bar{D}: & F^{(4)}(x_1, x_2, Q^2) = G_1(x_1)\bar{D}_2(x_2), \\
 gU: & F^{(5)}(x_1, x_2, Q^2) = G_1(x_1)U_2(x_2),
 \end{aligned}
 \tag{20}$$

¹⁶The case of W^- -boson can be treated in an analogous way.

where the generalised PDFs are used. They are defined as:

$$\begin{aligned}
 G_H(x) &= f_{0/H}(x, Q^2), \\
 U_H(x) &= f_{2/H}(x, Q^2)(V_{ud}^2 + V_{us}^2) \\
 &\quad + f_{4/H}(x, Q^2)(V_{cd}^2 + V_{cs}^2), \\
 \bar{D}_H(x) &= f_{-1/H}(x, Q^2)(V_{ud}^2 + V_{cd}^2) \\
 &\quad + f_{-3/H}(x, Q^2)(V_{us}^2 + V_{cs}^2), \\
 S_{12}(x_1, x_2) &= f_{-3/H_1}(x_1, Q^2) f_{2/H_2}(x_2, Q^2) V_{us}^2 \\
 &\quad + f_{-3/H_1}(x_1, Q^2) f_{4/H_2}(x_2, Q^2) V_{cs}^2 \\
 &\quad + f_{-1/H_1}(x_1, Q^2) f_{2/H_2}(x_2, Q^2) V_{ud}^2 \\
 &\quad + f_{-1/H_1}(x_1, Q^2) f_{4/H_2}(x_2, Q^2) V_{cd}^2, \\
 S_{21}(x_1, x_2) &= f_{2/H_1}(x_1, Q^2) f_{-3/H_2}(x_2, Q^2) V_{us}^2 \\
 &\quad + f_{4/H_1}(x_1, Q^2) f_{-3/H_2}(x_2, Q^2) V_{cs}^2 \\
 &\quad + f_{2/H_1}(x_1, Q^2) f_{-1/H_2}(x_2, Q^2) V_{ud}^2 \\
 &\quad + f_{4/H_1}(x_1, Q^2) f_{-1/H_2}(x_2, Q^2) V_{cd}^2,
 \end{aligned}
 \tag{21}$$

where V_{ij} are the CKM matrix elements.¹⁷

For simplicity in the former equations we omitted the top contribution, since the parton densities are zero for most practical applications.

Appendix B: Automated identification of sub-processes

In general there are 169 (13×13 flavour) possible PDF combinations for proton proton collisions. In order to store only the minimal amount of information, one needs to establish which of those combinations always come with correlated weights, or equivalently one should identify the underlying physical sub-processes. So far, for each process under study, the sub-processes have been identified manually, on a case-by-case basis. However, the sub-processes can also be found in an automated way.

To simplify the discussion (without loss of generality) it will be convenient to assume that the PDFs are always evaluated at fixed values of x . For each event i and for each of the 169 PDF combination j (with PDF weight p_j), the NLO QCD program calculates matrix-element weights W_{ij} . The total weight for the event i is $\sum_j W_{ij} p_j$. The PDF combinations are called channels in the following.

To identify the sub-processes, one determines the W_{ij} weights for 169 events, giving a 169×169 matrix, whose

¹⁷The CKM matrix elements are stored together with the weight grid in the same file. This ensures that the same values are used in the NLO calculation and in the PDF combinations. This choice can be changed a posteriori according to the needs to the user. In MCFM, only four non-zero CKM matrix elements are used.

i (event) index labels the rows and whose j (PDF channel) index labels the columns. One then carries out an eigenvalue decomposition of the W_{ij} matrix. If v_n denotes the n^{th} eigenvalue and L_n and R_n the left and right eigenvectors (with components L_{ni} , etc.), then as long as there are no degenerate eigenvalues, an orthonormality relation can be written:

$$L_n \cdot R_m = \delta_{mn}, \quad (22)$$

where the normalisation is our specific choice. Then one can rewrite the W_{ij} matrix as

$$W_{ij} = \sum_n R_{ni} v_n L_{nj}, \quad (23)$$

it being straightforward, for example, to verify that both sides satisfy $\sum_i L_{ni} W_{ij} = v_n L_{nj}$ and $\sum_j W_{ij} R_{mj} = v_m R_{mi}$. Let us now assume that only N of the eigenvalues are non-zero.¹⁸ Then Eq. 23 can be interpreted as follows: there are N relevant sub-processes; each $n \leq N$ corresponds to a sub-process that multiplies a linear combination of PDF channels in which the contribution of channel j is L_{nj} . In event i , sub-process n comes with a weight $R_{ni} v_n$. It will be convenient to denote this by w_{in} . By virtue of the orthonormality condition Eq. 22, we have that $w_{in} = \sum_j W_{ij} R_{nj}$, i.e. to determine the weight of sub-process n (whose PDF channel combination is given by the left eigenvector L_{nj}) we take the right-eigenvector that corresponds to this channel and use it to right-multiply the full weight matrix, so as to eliminate all but the contribution to the n^{th} sub-process.

The next step is to observe that the sub-processes determined for the first 169 events should hold for all remaining events.¹⁹ So now for *any* event i , we can determine the weight for sub-process n , $w_{in} = \sum_j W_{ij} R_{nj}$, using the R_{nj} determined from the initial events. Having stored the w_{in} one can then subsequently reconstruct the full W_{ij} as $W_{ij} = \sum_n w_{in} L_{nj}$.

We have verified, in the context of a number of MCFM processes, that this approach is viable in practice.²⁰ However, it has yet to be fully integrated with the rest of

our grid code and the results shown above are based on the manual sub-process decompositions explicitly spelled out in Sect. 2.4.1 and in Appendix A.

One should be aware that while the automated subprocess decomposition yields a number of subprocesses that is identical to what can be found with manual decomposition, the specific linear combinations of PDF channels are usually different. To help understand why, one can take the example of jet production with the 7 subprocesses of Eq. 12. There, rather than using qq and $q\bar{q}$ channels, one might have chosen instead to store weights for the combinations of $qq + q\bar{q}$ and $qq - q\bar{q}$ channels. More generally, one would have been free to base the grid on any 7 linearly independent combinations of the channels of Eq. 12. For the automated channel decomposition process, the particular independent linear combinations that emerge depend on the random weights of the events used to identify the channels.

References

1. D. Graudenz, M. Hampel, A. Vogt, C. Berger, Z. Phys. C **70**, 77 (1996)
2. D.A. Kosower, Nucl. Phys. B **520**, 263 (1998)
3. M. Stratmann, W. Vogelsang, Phys. Rev. D **64**, 114007 (2001)
4. M. Wobisch, Internal Report PITHA 00/12 and DESY-THESIS-2000-049, PhD-thesis RWTH Aachen, 2000
5. S. Chekanov et al. (ZEUS Collaboration), Eur. Phys. J. C **42**, 1 (2005)
6. P.G. Ratcliffe, Phys. Rev. D **63**, 116004 (2001)
7. M. Dasgupta, G.P. Salam, Eur. Phys. J. C **24**, 213 (2002)
8. G.P. Salam, J. Rojo, Comput. Phys. Commun. **180**, 120 (2009)
9. T. Carli, G.P. Salam, F. Siegert, Contributed to HERA and the LHC: a Workshop on the Implications of HERA for LHC Physics, Geneva, Switzerland. hep-ph/0510324 (2005)
10. T. Kluge, K. Rabbertz, M. Wobisch, in *Proceedings of the DIS 2006 Conference*, Tsukuba, Japan, hep-ph/0609285 (2006)
11. A. Banfi, G.P. Salam, G. Zanderighi, J. High Energy Phys. **07**, 026 (2007)
12. S. Eidelman et al., Phys. Lett. B **592**, 1 (2004)
13. Z. Nagy, Phys. Rev. D **68**, 094002 (2003)
14. Z. Nagy, Phys. Rev. Lett. **88**, 122003 (2002)
15. Z. Nagy, Z. Trocsanyi, Phys. Rev. Lett. **87**, 082001 (2001)
16. J.M. Campbell, R.K. Ellis, Phys. Rev. D **60**, 113006 (1999)
17. J.M. Campbell, R.K. Ellis, Phys. Rev. D **62**, 114012 (2000)
18. J. Pumplin et al. (CTEQ Collaboration), J. High Energy Phys. **07**, 012 (2002)
19. G.C. Blazey et al., hep-ex/0005012 (2000)
20. G.P. Salam, G. Soyez, J. High Energy Phys. **0705**, 086 (2007)
21. M. Cacciari, G.P. Salam, G. Soyez, J. High Energy Phys. **0804**, 005 (2008)
22. P.M. Nadolsky et al. (CTEQ Collaboration), Phys. Rev. D **78**, 013004 (2008)
23. A.D. Martin, W. J. Stirling, R.S. Thorne, G. Watt, Eur. Phys. J. C **63**, 189–285 (2009)
24. H1 and ZEUS Collaborations, H1prelim-08-045; ZEUS-prel-08-003 (2008)
25. R. Ball et al. Nucl. Phys. B **809**, 1–63 (2009)

¹⁸This contradicts the requirement that the eigenvalues be non-degenerate. In practice the rounding errors in the original calculation of the W_{ij} cause the nominally zero eigenvalues to be slightly non-zero, thus alleviating this issue in practice.

¹⁹This is guaranteed as long as the NLO Monte Carlo weights include all sub-processes for each of the first 169 events.

²⁰For carrying out this step in practice, we found the ALGLIB library to be useful, because it provides the necessary software tools written in C++ and has the option of using the `mpfr` multiple-precision library (available on most distributions), which turned out to be necessary in order to get numerically convergent results for the full set of flavours. The ALGLIB algorithms are translations of the LAPACK ones.

26. J.M. Campbell, J.W. Huston, W.J. Stirling, *Rep. Prog. Phys.* **70**, 89 (2007)
27. W.T. Giele, E.W.N. Glover, D.A. Kosower, *Nucl. Phys. B* **403**, 633 (1993)
28. D. Stump et al. (CTEQ Collaboration), *J. High Energy Phys.* **0310**, 046 (2003)
29. M. Botje, *J. Phys. G* **28**, 779 (2002)
30. S.I. Alekhin, IHEP-2000-17 and [hep-ex/0005042](#) (2000)
31. A.M. Cooper-Sarkar, *J. Phys. G* **28**, 2609 (2002)
32. R.S. Thorne et al., *J. Phys. G* **28**, 2717 (2002)
33. A.M. Cooper-Sarkar, in *Proceedings of the DIS 2007 Conference*, Munich, Germany, [0707.1593](#) [hep-ph] (2007)
34. M. Kobayashi, T. Maskawa, *Prog. Theor. Phys.* **49**, 652 (1973)
35. N. Cabibbo, *Phys. Rev. Lett.* **10**, 531 (1963)