Standard Model Handles and Candles Working Group: Tools and Jets Summary Report

Convenors: C. Buttar[1], J. D'Hondt[2], M. Krämer[3], G. Salam[4], M. Wobisch[5],
Contributing authors: N.E. Adam[6], V. Adler[7], A. Arbuzov[8], D. Bardin[8], U. Baur[9], A.A. Bhatti[10], S. Bondarenko[8], V. Büge[11,12], J.M. Butterworth[13], M. Cacciari[4], M. Campanelli[14], Q.-H. Cao[15], C.M. Carloni Calame[16,17], P. Christova[8], D. D'Enterria[18], J. D'Hondt[2], S. Ferrag[1], K. Geerlings[14], V. Halyo[6], M. Heinrich[11], J. Huston[14], J. Jackson[19,20], B. Jantzen[21], L. Kalinovskaya[8], D. Kcira[22], B. Klein[11], A. Kulesza[23], P. Loch[24], G. Montagna[25,26], S. Moretti[17], D. Newbold[19], O. Nicrosini[26], H. Nilsen[27], A.A. Penin[28,29], F. Piccinini[26], S. Pozzorini[30], K. Rabbertz[11], J. Rojo Chacon[4], R. Sadykov[8], M. Schulze[31], C. Shepherd-Themistocleous[19], A. Sherstnev[32,33], P.Z. Skands[34,35], L. Sonnenschein[18,36], G. Soyez[37], R.S. Thorne[13], M. Tytgat[38], P. Van Mulders[2], M. Vazquez Acosta[18], A. Vicini[39], I. Villella[2], D. Wackeroth[9], C.-P. Yuan[14]

[1] Department of Physics and Astronomy, University of Glasgow, G12 8QQ, Glasgow, UK
[2] Vrije Universiteit Brussel, Pleinlaan 2 B-1050 Brussel, Belgium
[3] Institut für Theoretische Physik, RWTH Aachen, D-52056 Aachen, Germany
[4] LPTHE, UPMC Univ. Paris 6; Univ. Paris Diderot (Paris 7); CNRS UMR 7589; F-75252 Paris Cedex 05, France
[5] Louisiana Tech University, Ruston, Louisiana, USA
[6] Department of Physics, Princeton University, Princeton, NJ 08544, USA
[7] Université Libre de Bruxelles (ULB-IIHE), boulevard du Triomphe, 1050 Bruxelles, Belgium
[8] JINR Dubna, 141980 Russia
[9] Department of Physics, University at Buffalo, Buffalo, NY 14260, USA
[10] Rockefeller University 1230 York Ave New York NY 10065
[11] Institut für Experimentelle Kernphysik, Universität Karlsruhe, Wolfgang-Gaede-Str. 1, D-76131 Karlsruhe, Germany
[12] Institut für Wissenschaftliches Rechnen, Forschungszentrum Karlsruhe, Hermann-von-Helmholtz-Platz 1, D-76344 Eggenstein-Leopoldshafen, Germany
[13] Department of Physics and Astronomy, University College London, WC1E 6BT, UK
[14] Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824, USA
[15] Department of Physics and Astronomy, UC Riverside, Riverside, CA 92521, USA
[16] INFN, Frascati, Italy
[17] School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK
[18] CERN PH-EP, CH-1211 Geneva 23, Switzerland
[19] Rutherford Appleton Laboratory, Oxon OX11 0QX, UK
[20] H. H. Wills Physics Laboratory, University of Bristol, UK
[21] Paul Scherrer Institut, 5232 Villigen PSI, Switzerland
[22] Université Catholique de Louvain (UCL), Belgium
[23] Deutsches Elektronen-Synchrotron DESY, Notkestrasse 85, D-22607 Hamburg, Germany
[24] Department of Physics, University of Arizona, Tucson, Arizona 85721, USA
[25] Dipartimento di Fisica Nucleare e Teorica, Università di Pavia, Via Bassi 6, Pavia, Italy
[26] INFN, Sezione di Pavia, Via Bassi 6, Pavia, Italy
[27] Physikalisches Institut, Albert-Ludwigs Universitẗ Freiburg, Hermann-Herder-Strasse 3a, 790104 Freiburg, Germany
[28] Department of Physics, University of Alberta, Edmonton, AB T6G 2J1, Canada
[29] Institute for Nuclear Research of Russian Academy of Sciences, 117312 Moscow, Russia
[30] Max-Planck-Institut für Physik, Föhringer Ring 6, D–80805 Munich, Germany
[31] Institut für Theoretische Teilchenphysik, Universität Karlsruhe, D-76128 Karlsruhe, Germany
[32] Cavendish Laboratory, University of Cambridge, JJ Thomson Avenue, Cambridge, CB3 0HE, UK
[33] on leave from Moscow State University, Moscow, Russia
[34] CERN PH-TH, CH-1211 Geneva 23, Switzerland
[35] Theoretical Physics, Fermilab MS106, Box 500, Batavia, IL–60510, USA
[36] LPNHE, UPMC Univ. Paris 6; Univ. Paris Diderot (Paris 7); IN2P3/CNRS ; F-75252 Paris Cedex 05, France
[37] Brookhaven National Laboratory, Upton, NY 11973, USA

[38] University of Gent, Belgium
[39] Dipartimento di Fisica, Università di Milano and INFN, Sezione di Milano, Milano, Italy

This report summarises the activity on comparisons of existings tools for the standard model and on issues in jet physics by the Standard Model Handles and Candles working group during and subsequent to the Workshop "Physics at TeV Colliders", Les Houches, France, 11–29 June, 2007.

Contents

1 FOREWORD

The construction of the LHC and its detectors is nearing completion, and first collisions are to be expected in 2008. While in essence built to discover new physics phenomena, the proton collisions at the LHC will provide a huge number of Standard Model events including jet, W, Z and top quark processes. These events can be used to further scrutinize the Standard Model as a theory, but are essential Handles and Candles for the broad physics commissioning of the experiments. Prior to any discovery of new phenomena a deep understanding of these background events has to be obtained. A solid knowledge of the Standard Model is crucial is estimating the diverse backgrounds in the signal regions and is a pre-requisite for the correct interpretation of the observed phenomena.

The primary aim of the Standard Model Handles and Candles working group, which has been set up in the framework of the Les Houches workshop is to address issues relevant in the programme described above. Several topics relevant for the Standard Model processes considered as a background or signal are discussed. Examples are electroweak and QCD processes like Z and W boson production and the high mass tail of the Drell-Yan spectrum. The prediction and understanding of the min-bias events and the parton density distributions are other topics.

The production of jets in the proton collisions at the LHC is abundant. Therefore a thorough understanding of jet physics is primordial, including for example a common nomenclature or accord when we speak about a generic jet of particles. Along this line it becomes relevant to compare the performance of several jet algorithms. A complete chapter is devoted to this domain, resulting in a list of recommendations for the physics analyses at the LHC.

Part I
COMPARISON OF EXISTING TOOLS FOR THE STANDARD MODEL

2 A TUNED COMPARISON OF ELECTROWEAK PREDICTIONS FOR Z BOSON OBSERVABLES WITH HORACE, SANC AND ZGRAD2 [1]

2.1 Introduction

W and Z bosons will be produced copiously at the LHC and high-precision measurements of cross sections and their properties will be used for detector calibration, to understand the background to many physics analysis, and last but not least, to explore a new electroweak high-energy regime in tails of Z and W distributions. In view of the importance of single W and Z production as 'standard candles' and for searches of signals of new physics, it is crucial to control the theoretical predictions for production cross section and kinematic distributions. For a review of available calculations and tools, see Refs. [1], for instance. Good theoretical control of the predicitions requires a good understanding of the residual theoretical uncertainties. As a first step, we perform a tuned numerical comparison of the following publicly available codes that provide precise predictions for Z observables, including electroweak (EW) O( ) corrections: HORACE [2, 3], SANC [4–6], and ZGRAD2 [7]. First results of a tuned comparison of Z production cross sections can be found in Ref. [8], and predictions for single W production including QCD and electroweak corrections have been recently discussed in Ref. [1]. A study of combined effects of QCD and electroweak corrections to the neutral-current process in the high invariant-mass region can be found in these procceedings.

2.2 Results of a tuned comparison of HORACE, SANC and ZGRAD2

Setup for the tuned comparison

[1] Contributed by: A. Arbuzov, D. Bardin, U. Baur, S. Bondarenko, C.M. Carloni Calame, P. Christova, L. Kalinovskaya, G. Montagna, O. Nicrosini, R. Sadykov, A. Vicini, D. Wackeroth

For the numerical evaluation of the cross sections at the LHC ($\sqrt{s} = 14$ TeV) we chose the following set of Standard Model input parameters:

$$
\begin{aligned}
G &= 1.16637 \cdot 10^{-5}\,\mathrm{GeV}^{-2}; & &= 1/137.03599911; & \alpha_s &\quad \alpha_s(M_Z^2) = 0.1176 \\
M_Z &= 91.1876\,\mathrm{GeV}; & \Gamma_Z &= 2.4924\,\mathrm{GeV} \\
M_W &= 80.37399\,\mathrm{GeV}; & \Gamma_W &= 2.0836\,\mathrm{GeV} \\
M_H &= 115\,\mathrm{GeV}; & & \\
m_e &= 0.51099892\,\mathrm{keV}; & m_\mu &= 0.105658369\,\mathrm{GeV}; & m_\tau &= 1.77699\,\mathrm{GeV} \\
m_u &= 0.06983\,\mathrm{GeV}; & m_c &= 1.2\,\mathrm{GeV}; & m_t &= 174\,\mathrm{GeV} \\
m_d &= 0.06984\,\mathrm{GeV}; & m_s &= 0.15\,\mathrm{GeV}; & m_b &= 4.6\,\mathrm{GeV} \\
|V_{ud}| &= 0.975; & |V_{us}| &= 0.222 \\
|V_{cd}| &= 0.222; & |V_{cs}| &= 0.975 \\
|V_{cb}| &= |V_{ts}| = |V_{ub}| = & |V_{td}| &= |V_{tb}| = 0
\end{aligned}
\tag{1}
$$

The W and Higgs boson masses, M_W and M_H, are related via loop corrections. To determine M_W we use a parametrization which, for $100\,\mathrm{GeV} < M_H < 1$ TeV, deviates by at most 0.2 MeV from the theoretical value including the full two-loop contributions [9] (using Eqs. (6,7,9)). Additional parametrizations can also be found in [10, 11].

We work in the constant width scheme and fix the weak mixing angle by $c_w = M_W/M_Z$, $s_w^2 = 1 - c_w^2$. The Z and W -boson decay widths given above are calculated including QCD and electroweak corrections, and are used in both the LO and NLO evaluations of the cross sections. The fermion masses only enter through loop contributions to the vector boson self energies and as regulators of the collinear singularities which arise in the calculation of the QED contribution. The light quark masses are chosen in such a way, that the value for the hadronic five-flavor contribution to the photon vacuum polarization, $\Delta\alpha^{(5)}_{\mathrm{had}}(M_Z^2) = 0.027572$ [12], is recovered, which is derived from low-energy e^+e^- data with the help of dispersion relations. The finestructure constant, $\alpha(0)$, is used throughout in both the LO and NLO calculations of the Z production cross sections.

In the course of the calculation of Z observables the Kobayashi-Maskawa-mixing has been neglected.

To compute the hadronic cross section we use the MRST2004QED set of parton distribution functions [13], and take the renormalization scale, μ_r, and the QED and QCD factorization scales, μ_{QED} and μ_{QCD}, to be $\mu_r^2 = \mu_{QED}^2 = \mu_{QCD}^2 = M_Z^2$. In the MRST2004QED structure functions, the factorization of the photonic initial state quark mass singularities is done in the QED DIS scheme which we therefore use in all calculations reported here. It is defined analogously to the usual DIS [14] schemes used in QCD calculations, i.e. by requiring the same expression for the leading and next-to-leading order structure function F_2 in deep inelastic scattering, which is given by the sum of the quark distributions. Since F_2 data are an important ingredient in extracting PDFs, the effect of the $O(\alpha)$ QED corrections on the PDFs should be reduced in the QED DIS scheme.

The detector acceptance is simulated by imposing the following transverse momentum (p_T) and pseudo-rapidity (η) cuts:

$$
p_T^{\ell} > 20\,\mathrm{GeV}; \qquad |\eta_\ell| < 2.5; \qquad \ell = e, \mu \tag{2}
$$

These cuts approximately model the acceptance of the ATLAS and CMS detectors at the LHC. Uncertainties in the energy measurements of the charged leptons in the detector are simulated in the calculation by Gaussian smearing of the particle four-momentum vector with standard deviation Δ which depends on the particle type and the detector. The numerical results presented here were calculated using Δ values based on the ATLAS specifications. In addition to the separation cuts of Eq. 2, we apply a cut on the invariant mass of the final-state lepton pair of $M_{ll} > 50$ GeV.

electrons	muons
combine e and momentum four vectors, if $R(e,\ ) < 0.1$	reject events with $E\ > 2$ GeV for $R(\ ,\ ) < 0.1$
reject events with $E\ > 0.1\,E_e$ for $0.1 < R(e,\ ) < 0.4$	reject events with $E\ > 0.1\,E$ for $0.1 < R(\ ,\ ) < 0.4$

Table 1: Summary of lepton identification requirements.

The granularity of the detectors and the size of the electromagnetic showers in the calorimeter make it difficult to discriminate between electrons and photons with a small opening angle. In such cases we recombine the four-momentum vectors of the electron and photon to an effective electron four-momentum vector. We require that the electron and photon momentum four-vectors are combined into an effective electron momentum four-vector if their separation in the pseudorapidity – azimuthal angle plane,

$$R(e,\ ) = \sqrt{(\ (e,\ ))^2 + (\ (e,\ ))^2}, \tag{3}$$

is $R(e,\ ) < 0.1$. For $0.1 < R(e,\ ) < 0.4$ events are rejected if $E\ > 0.1\,E_e$. Here E (E_e) is the energy of the photon (electron) in the laboratory frame.

Muons are identified by hits in the muon chambers and the requirement that the associated track is consistent with a minimum ionizing particle. This limits the photon energy for small muon – photon opening angles. For muons, we require that the energy of the photon is $E\ < 2$ GeV for $R(\ ,\ ) < 0.1$, and $E\ < 0.1E$ GeV for $0.1 < R(\ ,\ ) < 0.4$. We summarize the lepton identification requirements in Table 1. For each observable we will provide "bare" results, i.e. without smearing and recombination (only lepton separation cuts are applied) and "calo" results, i.e. including smearing and recombination. We will show results for kinematic distributions and total cross sections, at LO and NLO, and the corresponding relative corrections, $(\%) = d_{NLO}/d_{LO}\ \ 1$, at the LHC. We consider the following neutral current processes: $pp \to Z/\ \to l^+l^-$ with $l = e,\ $.

Z *boson observables*

$_Z$: total inclusive cross section of Z boson production.
The results for $_Z$ at LO and EW NLO and the corresponding relative corrections are provided in Table 2.

$\frac{d}{dM(l^+l^-)}$: invariant mass distribution of the final-state lepton-pair.
The relative corrections for different $M(l^+l^-)$ ranges are shown for bare and calo cuts in Figs. 1,2.

$\frac{d}{dp_T^l}$: transverse lepton momentum distribution.
The relative corrections are shown in Fig. 3 for bare and calo cuts.

$\frac{d}{d\ l}$: pseudo rapidity distribution of the lepton.
The relative corrections are shown in Fig. 4 for bare and calo cuts.

A_{FB}: forward-backward asymmetries (as a function of $M_{l^+l^-}$).
For pp collisions at Tevatron energies, A_{FB} usually is defined by [7]

$$A_{FB} = \frac{F\ \ B}{F + B}\ ; \tag{4}$$

where

$$F = \int_0^1 \frac{d}{d\cos}\,d\cos\ \ ; \qquad B = \int_1^0 \frac{d}{d\cos}\,d\cos\ \ : \tag{5}$$

LHC, $pp \to Z/\gamma^* \to e^+e^-$						
	bare cuts			calo cuts		
	LO [pb]	NLO [pb]	[%]	LO [pb]	NLO [pb]	[%]
HORACE	739.34(3)	742.29(4)	0.40(1)	737.51(3)	755.67(6)	2.46(1)
SANC	739.3408(3)	743.072(7)	0.504(1)	737.857(2)	756.54(1)	2.532(2)
ZGRAD2	737.8(7)	743.0(7)	0.71(9)	737.8(7)	756.9(7)	2.59(9)
LHC, $pp \to Z/\gamma^* \to \mu^+\mu^-$						
	bare cuts			calo cuts		
	LO [pb]	NLO [pb]	[%]	LO [pb]	NLO [pb]	[%]
HORACE	739.33(3)	762.20(3)	3.09(1)	738.28(3)	702.87(5)	-4.79(1)
SANC	739.3355(3)	762.645(3)	3.1527(4)	738.5331(3)	703.078(3)	-4.8006(3)
ZGRAD2	740(1)	764(1)	3.2(2)	740(1)	705(1)	-4.7(2)

Table 2: Tuned comparison of LO and EW NLO predictions for $_z$ from HORACE, SANC, and ZGRAD2. The statistical error of the Monte Carlo integration is given in parentheses.

Here, $\cos$ is given by

$$\cos = \frac{2}{m(l^+l^-)\sqrt{m^2(l^+l^-)+p_T^2(l^+l^-)}}\,\big(p^+(l^-)p^-(l^+) - p^-(l^-)p^+(l^+)\big) \qquad (6)$$

with

$$p^\pm = \frac{1}{2}(E \pm p_z); \qquad (7)$$

where E is the energy and p_z is the longitudinal component of the momentum vector. In this definition of $\cos$, the polar axis is taken to be the bisector of the proton beam momentum and the negative of the anti-proton beam momentum when they are boosted into the l^+l^- rest frame. In pp collisions at Tevatron energies, the flight direction of the incoming quark coincides with the proton beam direction for a large fraction of the events. The definition of $\cos$ in Eq. (6) has the advantage of minimizing the effects of the QCD corrections (see below). In the limit of vanishing di-lepton p_T, $\cos$ coincides with the angle between the lepton and the incoming proton in the l^+l^- rest frame.

For the definition of $\cos$ given in Eq. (6), $A_{FB} = 0$ for pp collisions. The easiest way to obtain a non-zero forward-backward asymmetry at the LHC is to extract the quark direction in the initial state from the boost direction of the di-lepton system with respect to the beam axis. The cosine of the angle between the lepton and the quark in the l^+l^- rest frame is then approximated by [7]

$$\cos = \frac{|p_z(l^+l^-)|}{p_z(l^+l^-)}\,\frac{2}{m(l^+l^-)\sqrt{m^2(l^+l^-)+p_T^2(l^+l^-)}}\,\big(p^+(l^-)p^-(l^+) - p^-(l^-)p^+(l^+)\big)\,: \qquad (8)$$

In Fig. 5 (resonance region) and Fig. 6 (tail region) we show the difference A_{FB} between the NLO EW and LO predictions for the forward-backward asymmetries for bare and calo cuts at the LHC.

The predictions of HORACE, SANC and ZGRAD2 show a satisfactory level of agreement. The effect of the EW NLO corrections, calculated for the total cross sections within the specified cuts, agrees within the statistical uncertainties of the MC integration, differs for the three codes at most by two per mille and in general by few tenth of per mille. Some discrepancies are present in specific observables. This requires further investigation, which is left to a future publication.

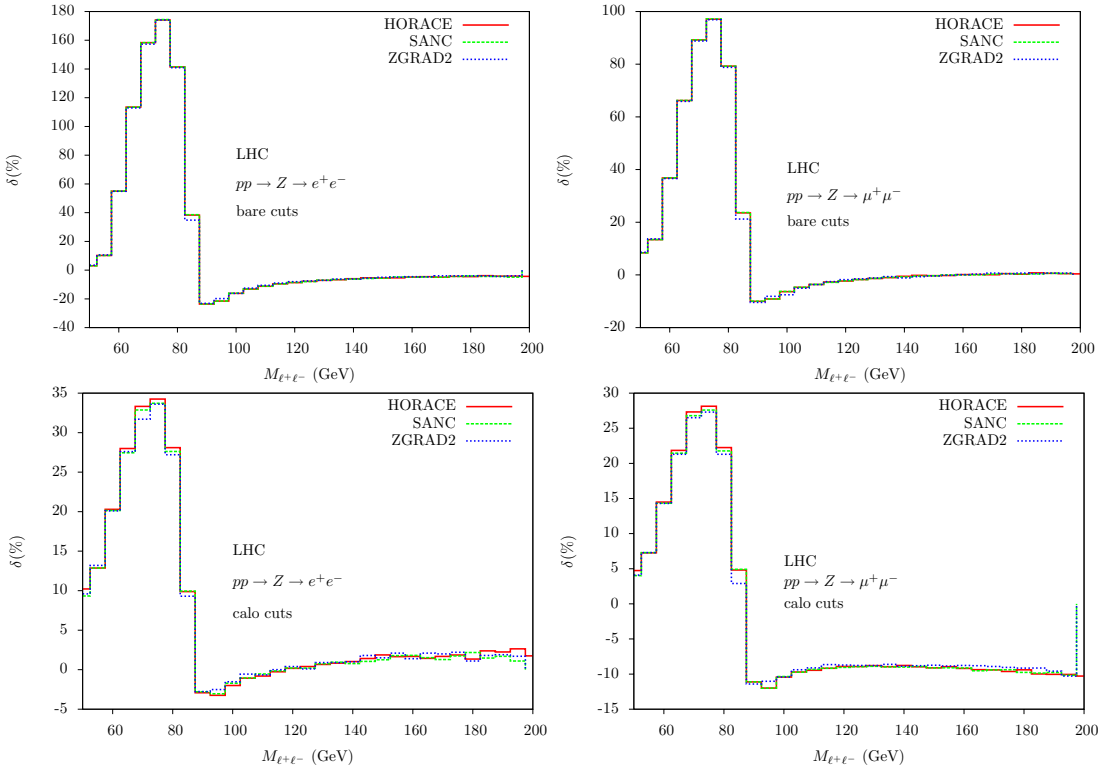

Fig. 1: The relative correction due to electroweak O( ) corrections to the M (l$^+$ l) distribution for Z production with bare and calo cuts at the LHC.

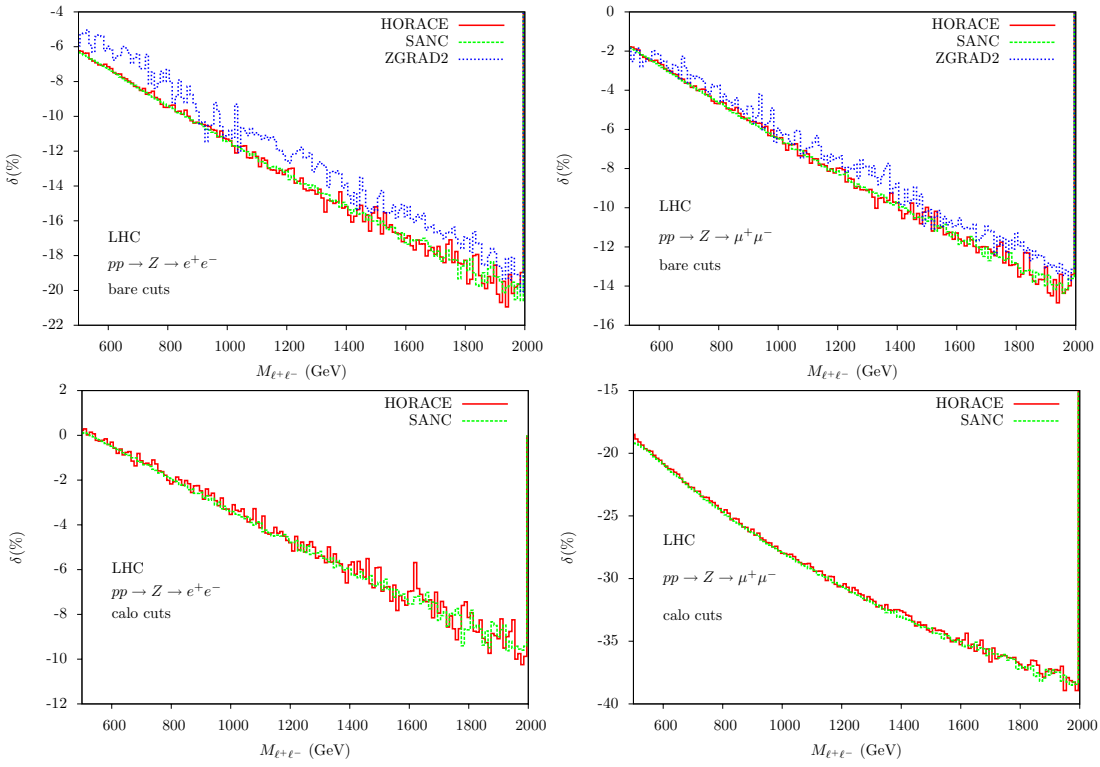

Fig. 2: The relative correction due to electroweak O( ) corrections to the M (l$^+$ l) distribution for Z production with bare and calo cuts at the LHC.

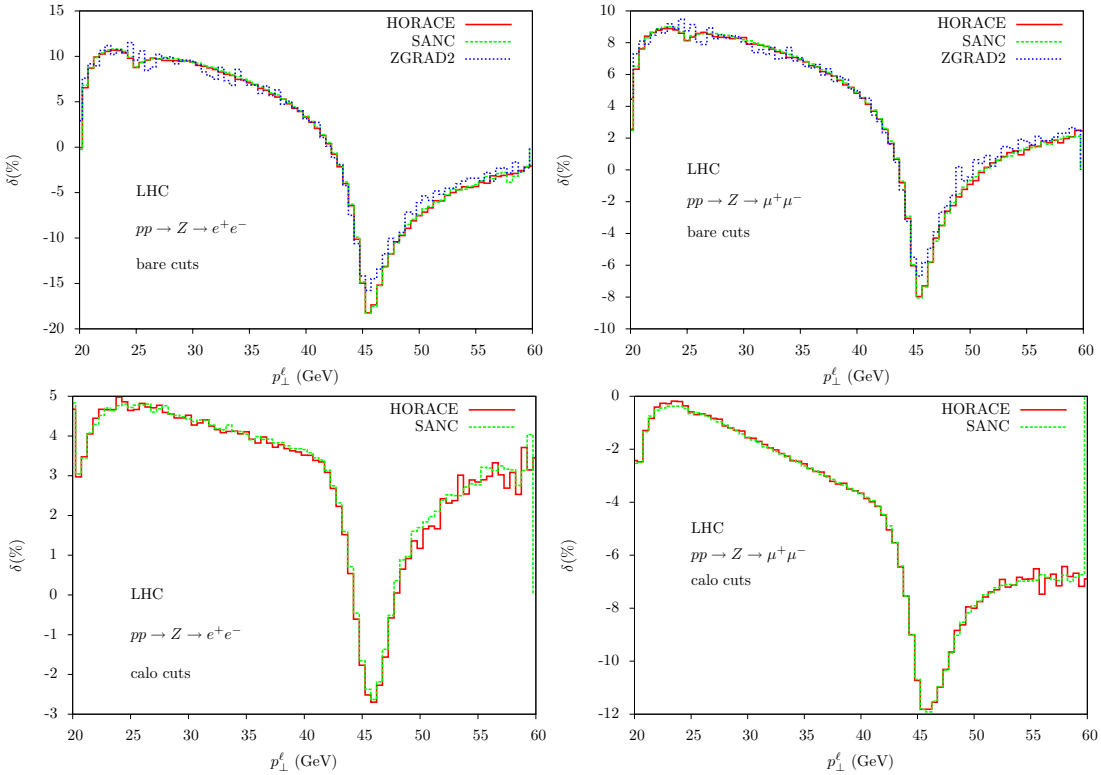

Fig. 3: The relative correction due to electroweak O( ) corrections to the p_T^l distribution for Z production with bare and calo cuts at the LHC.

Conclusions

In this report we performed a tuned comparison of the Monte Carlo programs HORACE, SANC and ZGRAD2, taking into account realistic lepton identification requirements. We found good numerical agreement of the predictions for the total Z production cross section, the M (ll), p_T^l and $_l$ distributions and the forward-backward asymmetry at the LHC. To find agreement between the available electroweak tools is only a first, albeit important step towards controlling the predictions for the neutral-current Drell-Yan process at the required precision level. More detailed studies of the residual uncertainties of predictions obtained with the available tools are needed, in particular of the impact of multiple photon radiation, higher-order electroweak Sudakov logarithms and combined QCD and EW effects (see contribution to these proceedings). Moreover, such a study should include PDF uncertainties, EW input scheme and QED/QCD scale uncertainties.

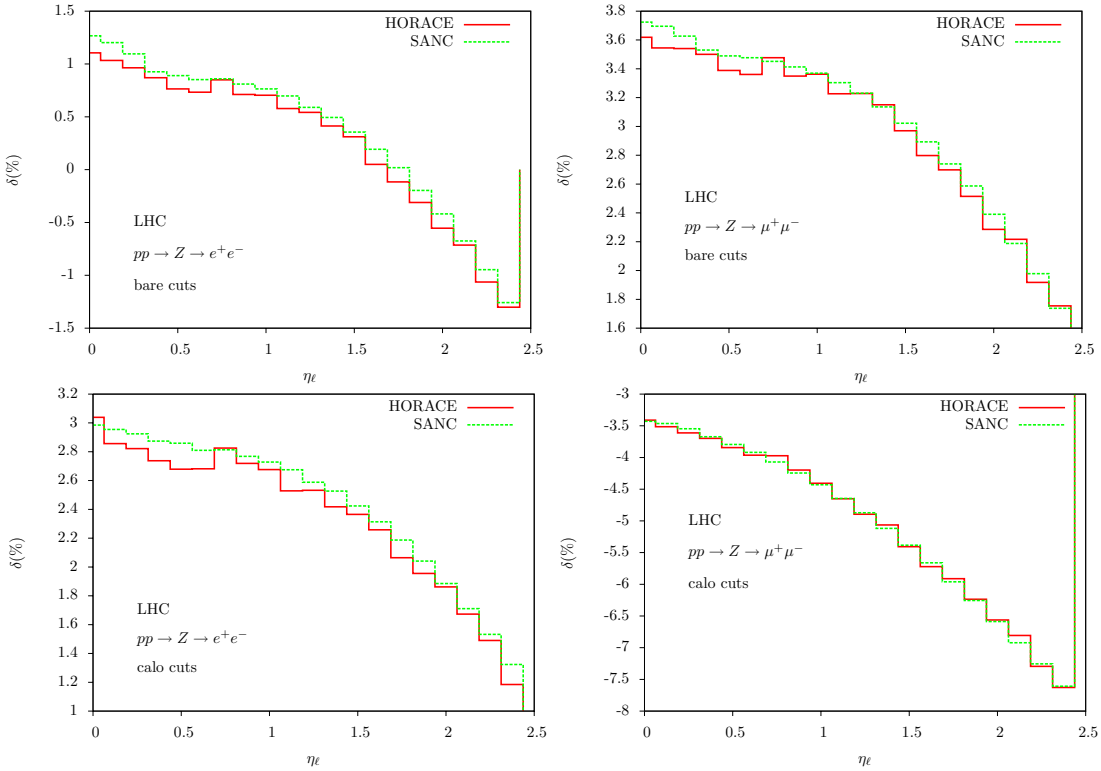

Fig. 4: The relative correction due to electroweak O( ) corrections to the $_1$ distribution for Z production with bare and calo cuts at the LHC.

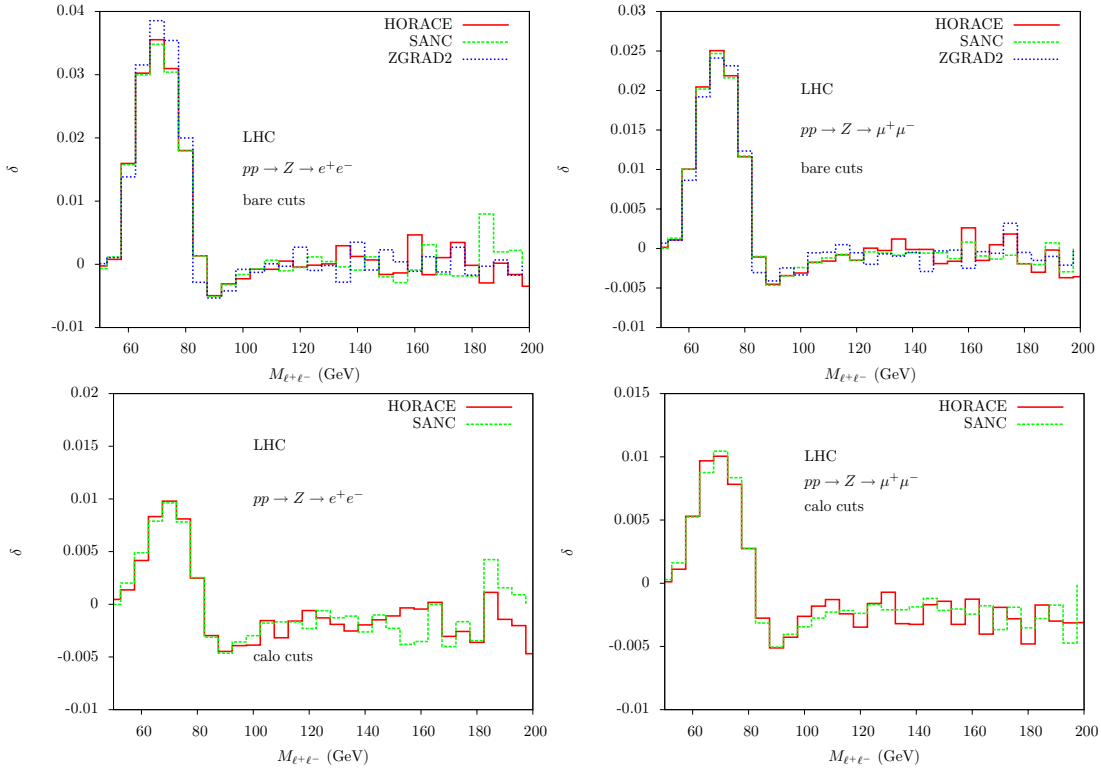

Fig. 5: The difference between the NLO and LO predictions for A_{FB} due to electroweak O( ) corrections for Z production with bare and calo cuts at the LHC.

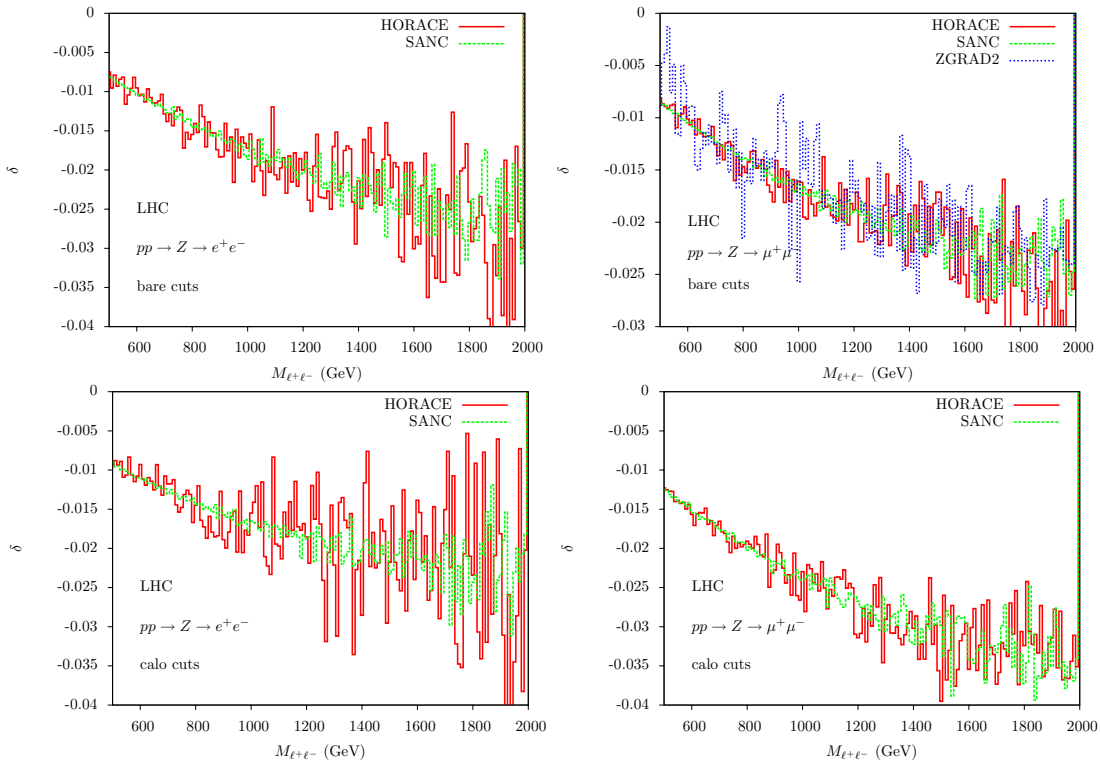

Fig. 6: The difference between the NLO and LO predictions for A_{FB} due to electroweak $O(\alpha)$ corrections for Z production with bare and calo cuts at the LHC.

3 THE NEUTRAL-CURRENT DRELL-YAN PROCESS IN THE HIGH INVARIANT-MASS REGION [2]

3.1 Introduction

The Neutral-Current (NC) Drell-Yan (DY) process, which can give rise to a high invariant-mass lepton pair, is a background to searches for new phenomena. Examples of these are new heavy resonances Z' and G* or possible excess resulting from the exchange of new particles such as the leptoquarks. These searches are an important part of the LHC physics program and require a precise knowledge of the Standard Model (SM) background in order to enable the observation of new physics signatures, which may only give rise to small deviations from the SM cross section.

The DY process has been studied in great detail (cf. [15,16] for a review), but independently in the strong (QCD) and electroweak (EW) sectors. In the high invariant-mass region QCD effects are known to be large and positive. These must be studied including both fixed order results and, for some classes of results, resummation to all orders of the contributions. The EW corrections tend to increase in size with energy, because of the virtual Sudakov EW logarithms. In the high invariant-mass region, these can be of the same order of magnitude as the QCD corrections, but have opposite sign. In addition, multiple photon radiation plays a non-negligible role in the determination of the invariant-mass distribution and induces negative corrections of the order of a few percent. In the light of this, it is a worthwhile and non-trivial exercise to combine all of these different sets of corrections, with the ultimate objective of determining the DY NC cross section, in the high invariant-mass region, to a precision of a few percent. The results presented in this contribution represent the first stage of a longer term project, with the objective of systematically investigating all of the various sources of theoretical uncertainty, which can induce effects of the order of a few percent.

3.2 Available calculations and codes

QCD corrections have been very well studied and a variety of calculations and Monte Carlo (MC) generators exist. These include, next-to-leading-order (NLO) and next-to-next-to-leading-order (NNLO) corrections to the W/Z total production rate [17, 18], NLO calculations for $W, Z + 1, 2$ jets signatures [19, 20] (available in the codes DYRAD and MCFM), resummation of leading and next-to-leading logarithms due to soft gluon radiation [21, 22] (implemented in the MC ResBos), NLO corrections merged with QCD Parton Shower (PS) evolution (for instance in the event generators MC@NLO [23] and POWHEG [24]), NNLO corrections to neutral- and charged-current DY in fully differential form [25–28] (available in the MC program FEWZ), as well as leading-order multi-parton matrix element generators matched with PS, such as, for instance, ALPGEN [29], MADEVENT [30, 31], SHERPA [32] and HELAC [33–35].

Complete $O(\alpha)$ EW corrections to DY processes have been computed independently by various authors in [3, 6, 7, 36] for NC production. The EW tools which implement exact NLO corrections to NC production are ZGRAD2 [7], HORACE [3] and SANC [6]. In HORACE the effect of multiple photon radiation to all orders via PS is matched with the exact NLO-EW calculation.

3.3 Electroweak Sudakov logarithms

At high invariant masses $Q^2 \gg M_W^2$, the EW corrections are enhanced by Sudakov logarithms of the form $\ln(Q^2/M_W^2)$, which originate from the exchange of soft and collinear virtual EW gauge bosons as well as from the running of the EW couplings. At the LHC, these corrections can reach tens of percent at the one-loop level and several percent at the two-loop level [37–39]. The EW Sudakov corrections to

[2] Contributed by: U. Baur, Q.-H. Cao, C.M. Carloni Calame, S. Ferrag, J. Jackson, B. Jantzen, G. Montagna, S. Moretti, D. Newbold, O. Nicrosini, A.A. Penin, F. Piccinini, S. Pozzorini, C. Shepherd-Themistocleous, A. Vicini, D. Wackeroth, C.-P. Yuan

the NC four-fermion scattering amplitude [40–43] can schematically be written as

$$\mathcal{A} = \mathcal{A}_B(Q^2)\left[1 + \sum_{n=1}^{\infty}\left(\frac{\alpha}{4\pi}\right)^n \sum_{k=0}^{2n} C_{n,k}\ln^k\frac{Q^2}{M_W^2}\right], \tag{9}$$

where $\mathcal{A}_B(Q^2)$ is the Born amplitude with running EW couplings at the scale Q^2. The logarithmic corrections are known to next-to-next-to-next-to-leading-logarithmic (NNNLL) accuracy at the two-loop level [42, 43], i.e. $C_{2,k}$ with $4 \geq k \geq 1$ are known. Due to very strong cancellations between dominant and subdominant logarithmic terms, the two-loop corrections to the $e^+e^- \to \mu^+\mu^-$ and $e^+e^- \to q\bar{q}$ total cross sections are much smaller than what might naively be expected and do not exceed a few per mil in the TeV region.

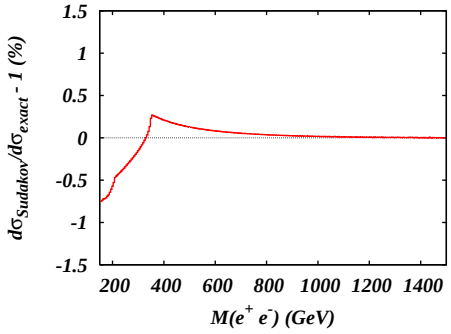

Fig. 7: Relative precision (in percent) of the Sudakov approximation: the one-loop predictions for the e^+e^- invariant mass at the LHC are compared with ZGRAD2. The results have been obtained with the following separation cuts: $p_T(l) > 20$ GeV and $|\eta(l)| < 2.5$.

Nevertheless, for the DY process, kinematic cuts and differential distributions might partially destroy the cancellations and thus lead to much bigger corrections. It is therefore important to investigate higher-order Sudakov EW corrections to differential DY distributions at the LHC. To this end we have written a FORTRAN code that implements the results of Ref. [43] in fully differential form and permits the interfacing of these to the programs ZGRAD2 [7] and HORACE [3]. The one-loop Sudakov expansion has been validated and agrees with the weak corrections of ZGRAD2 with a precision at the few per mil level or better for $Q \gtrsim 200$ GeV (see Fig. 7). The small deviations, at low invariant mass, are of the order of the mass-suppressed terms neglected in the Sudakov approximation. Fig. 8 shows the Sudakov expansion up to two loops, wherein virtual photonic contributions are subtracted as in Ref. [43] and real photon emission is not included. At the one-loop level, the Sudakov approximation (solid curve) is in good agreement with the HORACE prediction (dashed-dotted curve), which was obtained by using the set of input parameters appearing in Section 3.4.1, from the full EW correction by subtracting $\mathcal{O}(\alpha)$ photon emission in the leading-logarithmic (LL) approximation.[3] The subtraction of the QED-LL correction makes the results presented in Fig. 8 independent, up to terms of order $\mathcal{O}(m_l^2/M_W^2)$, of the final state lepton flavour. The one-loop Sudakov correction yields a negative contribution that reaches -7% at 1.5 TeV. The combination of one- and two-loop Sudakov corrections is shown by the dashed line. The two-loop effects are positive, reach 1–2% in the plotted invariant-mass range and tend to reduce the one-loop contributions.

3.4 Combining QCD and EW corrections

In the high invariant-mass region both QCD and EW effects are large and therefore, in view of the high accuracy needed by new physics searches, it is important to combine both corrections consistently, at

[3] Electromagnetic matching corrections will be addressed in a forthcoming publication, but the good agreement suggests that they should be quite small.

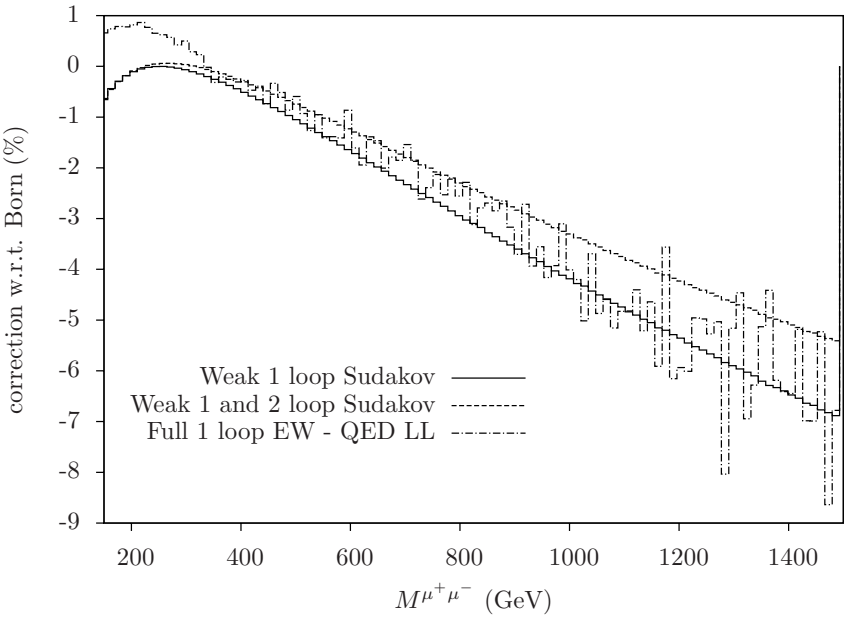

Fig. 8: EW corrections to the $^+$ invariant mass at the LHC: one-loop predictions of HORACE (dashed-dotted, see text); one-loop (solid) and two-loop (dashed) Sudakov approximation.

the event generator level, to perform a realistic simulation of this process. A first attempt to combine QED and QCD corrections can be found in [44] and results for the high invariant-mass distribution of charged lepton pairs are shown in Section 3.4.2. The combination of QCD and EW effects presented in Section 3.4.1 follows the approach first devised in [45–47].

3.4.1 Combined QCD and EW effects with MC@NLO *and* HORACE

The formula for the combination of QCD and EW effects is given by [45–47]:

$$\frac{d}{dO}_{QCD\ EW} = \frac{d}{dO}_{best\,QCD} + \frac{d}{dO}_{best\,EW} - \frac{d}{dO}_{born\ HERWIGPS} \tag{10}$$

where the differential cross-section, with respect to any observable O, is given by two terms: i) the results of a code which describes at best the effect of QCD corrections; ii) the effects due to NLO-EW corrections and to higher-order QED effects of multiple photon radiation computed with HORACE. In the EW calculation, the effect of the Born distribution is subtracted to avoid double counting since this is included in the QCD generator. In addition, the EW corrections are convoluted with a QCD PS and include, in the collinear approximation, the bulk of the $O(\ _s)$ corrections.

Preliminary numerical results have been obtained, for an e^+e final state, with the following set of input parameters:

$$G = 1{:}16639\ \ 10^{-5}\,GeV^{-2}; \qquad = 1=137{:}03599911; \qquad _s\ \ _s(M_Z^2) = 0{:}118;$$
$$M_W = 80{:}419\,GeV; \qquad M_Z = 91{:}188\,GeV; \qquad _Z = 2{:}4952\,GeV;$$
$$m_e = 0{:}51099892\,MeV; \qquad m\ = 0{:}105658369\,GeV; \quad m_t = 174{:}3\,GeV:$$

The parton distribution function (PDF) set MRST2004QED [13] has been used to describe the proton partonic content. The PDF factorization scale has been set equal to $_F = \sqrt{p_?^{Z\,2} + M^2_{e^+e}}$, where M_{e^+e} is the invariant mass of the lepton pair. The following cuts have been imposed to select the events:

$$p_?^e > 25\,GeV; \quad j^e\ j < 2{:}5; \quad M_{e^+e} > 200\,GeV: \tag{11}$$

The percentage corrections shown in the right panels of Figs. 9 and 10 have been defined as $= (_{NLO} - _{Born+PS})/_{Born+PS}$. The granularity of the detectors and the size of the electromagnetic showers in the calorimeter make it difficult to discriminate between electrons and photons with a small opening angle. We adopt the following procedure to select the event: we recombine the four-momentum vectors of the electron and photon into an effective electron four-momentum vector if, defining

$$R(e;) = \sqrt{(e;)^2 + (e;)^2}; \qquad (12)$$

$R(e;) < 0.1$ (with $;$ the distances of electrons and photons along the longitudinal and azimuthal directions). We do not recombine electrons and photons if > 2.5 (with the photon pseudo-rapidity). We apply the event selection cuts only after the recombination procedure.

We have used MC@NLO as the best QCD generator and have tuned it with MCFM/FEWZ at NLO. With the same settings, the two codes, when run at LO, give the same results as HORACE. The tuning procedure validates the interpretation of the various relative effects as due to the radiative corrections and not to a mismatch in the setups of the two codes. The results presented have been obtained using HORACE where the exact NLO-EW corrections are included, but no higher-order effects due to QED multiple emissions. Fig. 9 shows the interplay between the QCD and EW corrections for the di-lepton

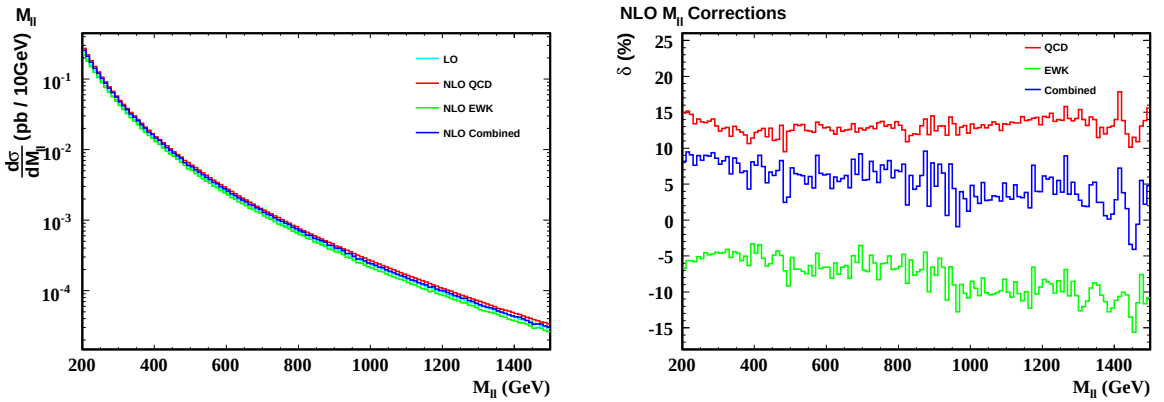

Fig. 9: QCD and EW corrections to the di-electron invariant mass.

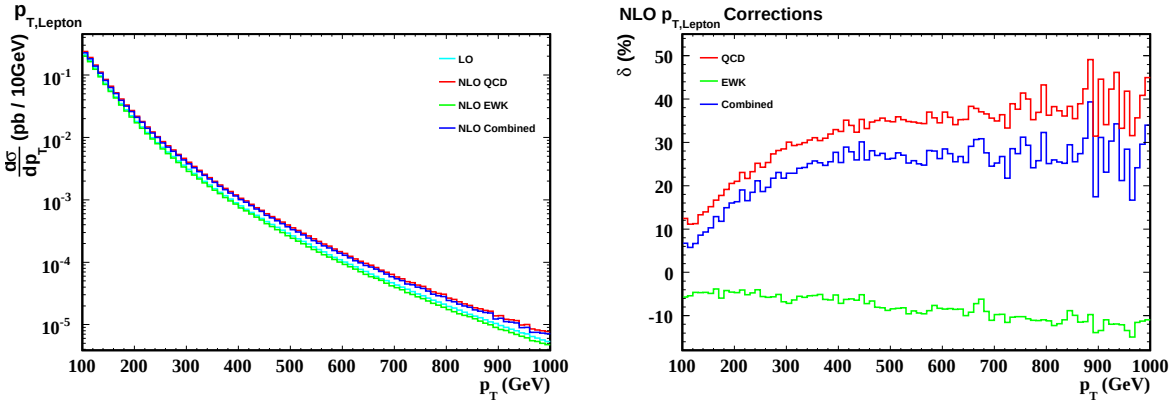

Fig. 10: QCD and EW corrections to the electron transverse momentum.

invariant mass. The QCD corrections are quite flat and positive with a value of about 15% over the mass range 200–1500 GeV. The EW corrections are negative and vary from about 5% to 10% and thus partially cancel the NLO-QCD effect. The 2-loop Sudakov logarithms (absent in this plot) would give an additional positive contribution to the cross-section. In Fig. 10 the lepton transverse-momentum

distribution is shown. The NLO-QCD corrections rise from 10% to 35% in the interval considered (100–1000 GeV). The NLO-EW corrections are negative and fall from 5% to 10% over the same range.

3.4.2 Combined QCD and EW effects with ResBos

In this work we also examine the effects of the initial-state multiple soft-gluon emission and the dominant final-state EW correction (via box diagrams) on the high invariant-mass distribution of the charged lepton pairs produced at the LHC. We shall focus on the region of $200\,\mathrm{GeV} < m_{\ell\ell} < 1500\,\mathrm{GeV}$, where $m_{\ell\ell}$ denotes the invariant mass of the two final-state charged leptons. The fully differential cross section including the contributions from the initial-state multiple soft-gluon emission is given by the resummation formula presented in Refs. [21, 44, 48, 49]. Furthermore, it has been shown that, above the Z pole region, the EW correction contributed from the box diagrams involving Z and W exchange is no longer negligible [7]. It increases strongly with energy and contributes significantly at high invariant mass of the lepton pair. Hence, we will also include the dominant EW correction via box diagrams in this study.

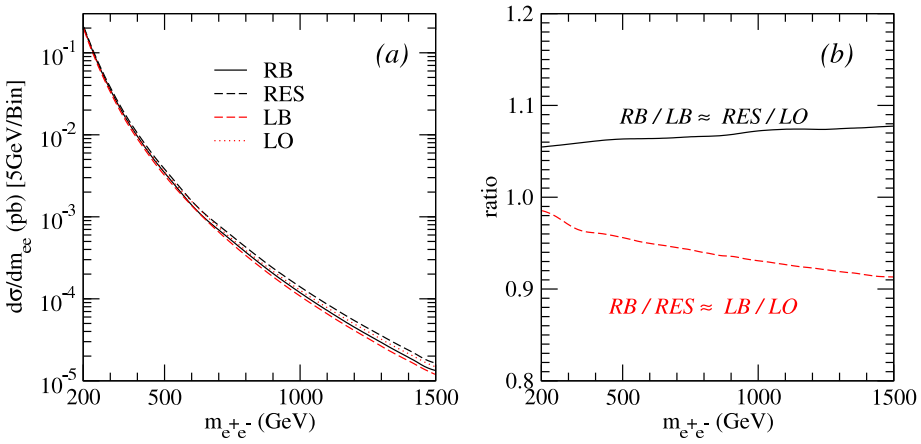

Fig. 11: (a) Invariant-mass distributions of the charged lepton pair; (b) ratios of various contributions.

For clarity, we introduce below the four shorthand notations:

LO: leading-order initial state,

LO+BOX (LB): leading-order initial state plus the ZZ/WW box diagram contribution,

RES: initial-state QCD resummation effects,

RES+BOX (RB): initial-state QCD resummation effects plus the ZZ/WW box-diagram contribution.

For this exercise, we consider the electron lepton pairs only and adopt the CTEQ6.1M PDFs [50]. Fig. 11(a) shows the distributions of the invariant mass $m_{e^+e^-}$ for RES+BOX (RB) (black solid line), RES only (black dashed line), LO+BOX (LB) (red dashed line) and LO only (red dotted line). It is instructive to also examine the ratios of various contributions, as shown in in Fig. 11(b). We note that the initial-state QCD resummation effect and the EW correction via box diagrams are almost factorized in the high invariant-mass region, e.g.

$$\frac{d_{RB}}{dm_{\ell\ell}} \Big/ \frac{d_{LB}}{dm_{\ell\ell}} \quad , \quad \frac{d_{RES}}{dm_{\ell\ell}} \Big/ \frac{d_{LO}}{dm_{\ell\ell}} ; \tag{13}$$

$$\frac{d_{RB}}{dm_{\ell\ell}} \Big/ \frac{d_{RES}}{dm_{\ell\ell}} \quad , \quad \frac{d_{LB}}{dm_{\ell\ell}} \Big/ \frac{d_{LO}}{dm_{\ell\ell}} : \tag{14}$$

The EW correction from the box diagrams reduces the invariant-mass distribution slightly around $m_{e^+e^-} \quad 200\,\mathrm{GeV}$ and largely (9%) around $m_{e^+e^-} \quad 1500\,\mathrm{GeV}$. On the other hand, the initial

state soft-gluon resummation effect increases the invariant-mass distribution by an amount of 5% at 200 GeV and 8% at 1500 GeV. Therefore, the QCD resummation effect dominates over the EW correction induced by the ZZ=WW box diagrams in the relatively low invariant-mass region, and they become comparable in the high invariant-mass region. The cancellation between both contributions in the high invariant-mass region causes the net contribution to be close to the leading order prediction. Finally, we note that the final state QED correction should also be included for predicting precision measurements. A detailed study including the soft-gluon resummation effect and the full EW correction will be presented elsewhere.

3.5 Outlook and conclusions

The preliminary results of this contribution show the non-trivial interplay between EW and QCD corrections in the high invariant-mass region of the NC DY process. For most of the observables, the NLO EW corrections are negative and partially cancel the QCD ones.

The NC DY process has been studied in great detail in the literature. This contribution is a first step towards collecting these different results and augmenting them with further studies to obtain an accurate prediction of this process. We have shown a preliminary investigation which includes, separately, results on the EW 2-loop Sudakov logarithms, QCD resummation, and combination of QCD and EW NLO corrections. The ongoing investigation aims to combine the effects above in the simulation and complete them with multiple photon emission and photon-induced partonic subprocesses. All these effects induce corrections of the order of a few percent. In addition, the di-electron and di-muon final states will be studied separately in more detail. We also aim to include the effect of real W and Z boson emission. This could result in the partial cancellation of virtual EW corrections, but it is dependent upon the definition of the observables and the experimental analysis. For completeness, we will include the systematic uncertainties from the PDFs, energy scale, choice of calculation scheme, higher-order contributions, showering model and the EW-QCD combination.

Acknowledgment

The authors would like to thank the Scottish Universities Physics Alliance (SUPA) and the Department of Physics and Astronomy of the University of Glasgow for their financial support. SM thanks the Royal Society (London, UK) for financial support in the form of a conference grant to attend this workshop.

4 COMPARISON OF HORACE AND PHOTOS IN THE Z ! ‘+ ‘ PEAK REGION [4]

4.1 Introduction

Precise measurement of gauge boson production cross-sections for pp scattering will be crucial at the LHC. W =Z bosons will be produced copiously, and a careful measurement of their production cross-sections will be important in testing the Standard Model (SM) more rigorously than ever before to potentially uncover signs of new physics.

Currently, no Monte Carlo (MC) event generators exist that include *both* higher order QCD and electroweak corrections. In what follows therefore, we evaluate whether it is possible to accurately describe the Z production cross-section under the Z peak with an event-level generator that includes only Final State QED Radiation (FSR) corrections (in the leading-log approximation) instead of the complete electroweak corrections included in the HORACE generator. In addition, we estimate the error that results if one chooses to use this MC event generator scheme.

4.2 Impact of Electroweak Corrections on Z Production Cross-Section.

The lack of a MC event generator that incorporates beyond leading order corrections in both the electroweak and QCD calculations, leads us to study which of the corrections contribute dominantly under the Z peak. By far the largest correction comes from inclusion of NLO QCD calculations. These produce a change in the cross-section of 20% or more [51], depending on the Z kinematic region considered. What we wish to determine then is the error imposed through including only the leading-log FSR contributions instead of the exact O( ) corrections matched with higher-order QED radiation that exist in HORACE. (since these are currently all that can be incorporated in addition to the NLO QCD corrections).

In order to study this error we used HORACE [52–55], a MC event generator that includes exact O( ) electroweak radiative corrections matched to a leading-log QED parton shower, and compared it to a Born-level calculation with final-state QED corrections added. The latter QED corrections were calculated by the program PHOTOS [56–58], a process-independent module for adding multi-photon emission to events created by a host generator.

In the following we compared pp ! Z= ! l+ l events generated by HORACE with the full 1-loop corrections (as described above) and parton-showered with HERWIG, to these events generated again by HORACE, but with only the Born-level calculation, and showered with HERWIG+PHOTOS. The results are shown in Figs. 12–20. In addition, the total production cross-sections of Z ! ‘+ ‘ with and without a mass cut around the Z peak and kinematic acceptance cuts are provided in Table 3.

The histograms of the Z boson distributions (Figs. 12–14) show that the HORACE Born-level calculation and Born-level with PHOTOS FSR are the same. This is expected, since PHOTOS does not modify the properties of the parent Z. The higher order calculation gives a visible difference in cross-section for M Z > 100 GeV/c^2, as is shown in Fig. 21. For the invariant mass of the lepton pair (in Fig. 15 we show this for muons), however, the two calculations agree nicely. The much better agreement (from the PHOTOS corrections) is highlighted in Fig. 22. Similarly there is good shape agreement for the other lepton kinematic quantities shown in Figs. 16 and 17. In terms of the acceptance, this agreement is quantitatively demonstrated to be better than 1%, as shown in Table 3. A reasonable agreement in the number of FSR photons emitted, and their transverse momentum spectra, between PHOTOS and HORACE is also shown in Figs. 18–20.

We conclude that the errors due to not including the complete electroweak one-loop corrections are below the 1% in the region of the Z peak as far as integrated cross sections are considered.

[4] Contributed by: N.E. Adam, C.M. Carloni Calame, V. Halyo, C. Shepherd-Themistocleous

$Z \to \ell^+\ell^-$ Production Cross-Section

	(No PS)		(Cuts Loose)		(Cuts Tight)	
HORACE Born	1984.2	2.0	1984.2	2.0	612.5	1.1
HORACE Born+PHOTOS	1984.2	2.0	1964.6	2.0	597.6	1.1
HORACE EWK Corr.	1995.7	2.0	1961.4	2.0	595.3	1.1
Error	0.58	0.14%	0.16	0.14%	0.38	0.26 %

Table 3: Calculation of the $Z/\gamma^* \to \ell^+\ell^-$ cross-section at various orders of electroweak corrections using HORACE 3.1 [52–55]. The first column gives the generator level cross-section with no QCD parton showering (No PS). This cross-section is the same for the Born calculation, and the Born calculation with PHOTOS corrections, since PHOTOS does not modify the inital cross-section. The PDF calculations are from CTEQ6.5M and the loose cut region is defined as $M_{\ell\ell} > 40$ GeV/c^2, $p_T' > 5$ GeV/c, and $|\eta_\ell| < 50.0$, while the tight cut region is defined as $40 < M_{\ell\ell} < 140$ GeV/c^2, $p_T' > 20$ GeV/c, and $|\eta_\ell| < 2.0$. In the first column we show the total generator-level cross-section before parton showering. The events are generated in the kinematic region defined by $M_Z > 40$ GeV/c^2, $p_T' > 5$ GeV/c, and $|\eta_\ell| < 50.0$.

Acknowledgments

We thank the organizers and conveners of the Les Houches workshop where this work originated. We also thank Fulvio Piccinni for useful discussions. This work was also supported in part by USDOE grant DE-FG02-91ER40671

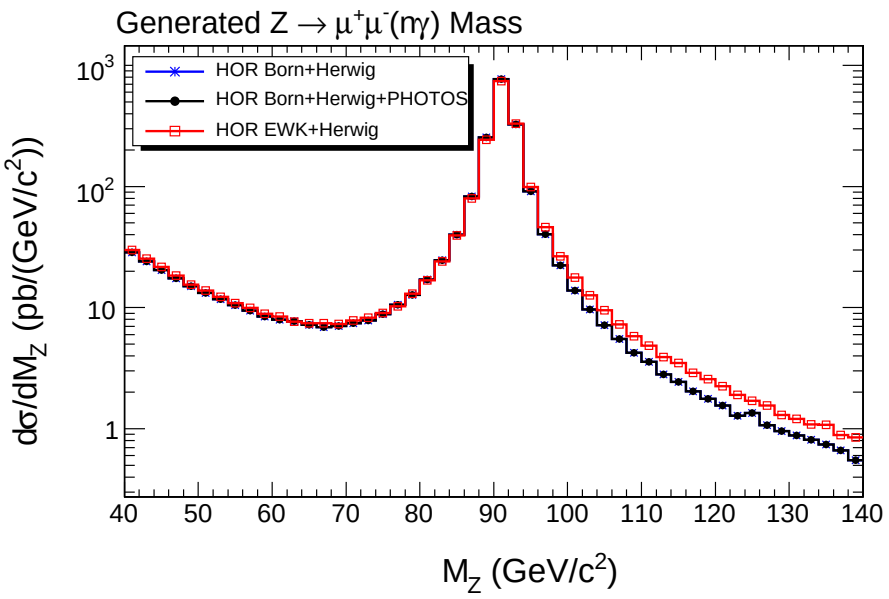

Fig. 12: Comparison of Z boson invariant mass distributions for the process Z = ! ′+ ′ (n) in HORACE 3.1 including electroweak and QED corrections showered with HERWIG (open red squares), HORACE Born-level showered with HERWIG plus PHOTOS (black circles), and HORACE Born-level (blue stars).

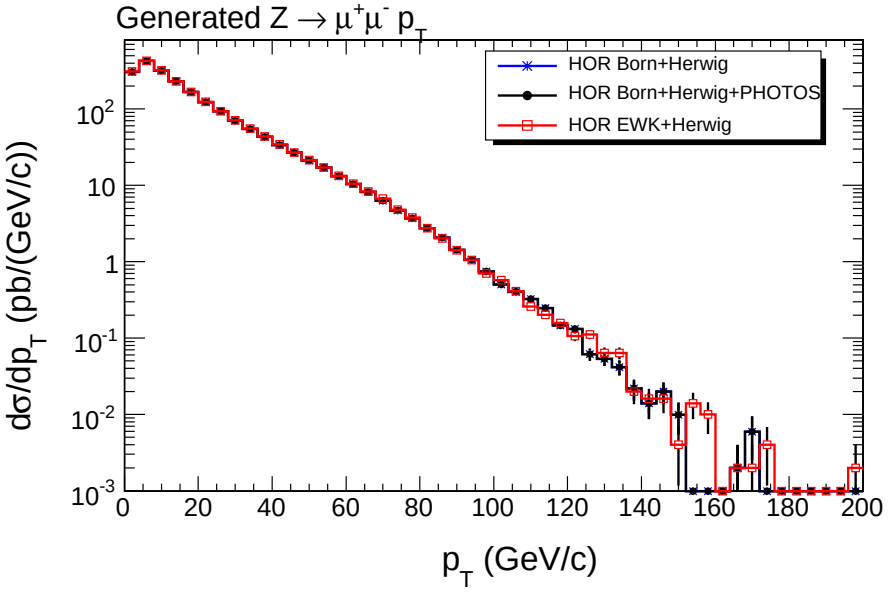

Fig. 13: Comparison of Z boson transverse momentum distributions for the process Z = ! ′+ ′ (n) in HORACE 3.1 including electroweak and QED corrections showered with HERWIG (open red squares), HORACE Born-level showered with HERWIG plus PHOTOS (black circles), and HORACE Born-level (blue stars).

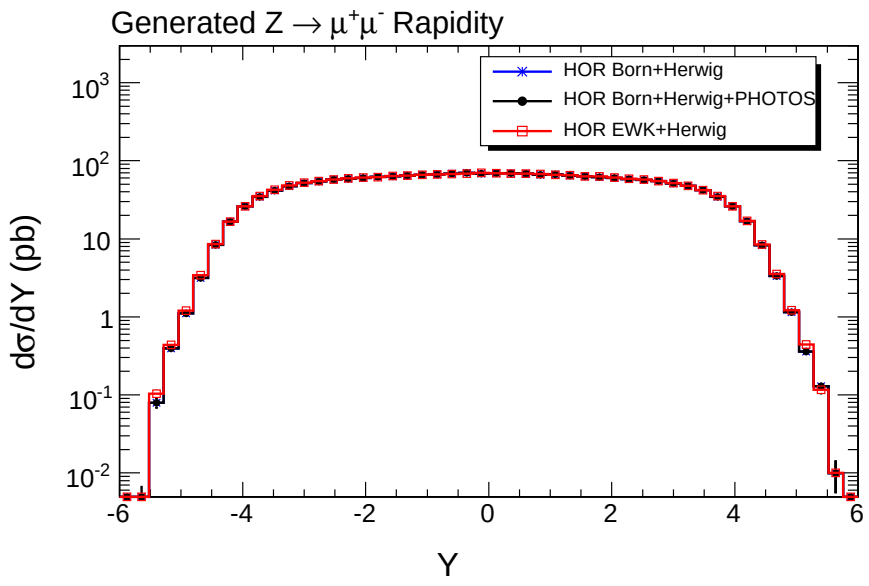

Fig. 14: Comparison of Z boson rapidity distributions for the process Z = ! ′+ ′ (n) in HORACE 3.1 including electroweak and QED corrections showered with HERWIG (open red squares), HORACE Born-level showered with HERWIG plus PHOTOS (black circles), and HORACE Born-level (blue stars).

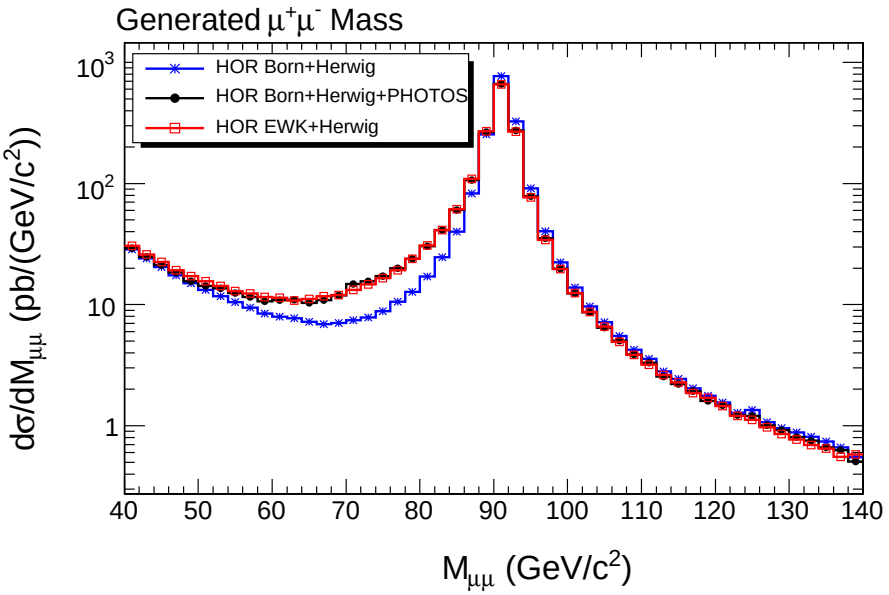

Fig. 15: Comparison of ′+ ′ invariant mass distributions for the process Z = ! ′+ ′ (n) in HORACE 3.1 including electroweak and QED corrections showered with HERWIG (open red squares), HORACE Born-level showered with HERWIG plus PHOTOS (black circles), and HORACE Born-level (blue stars).

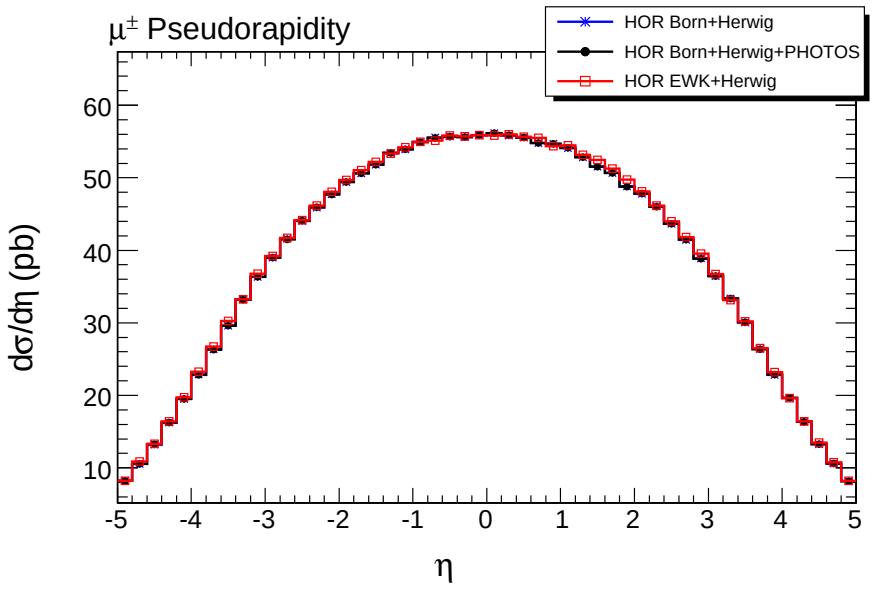

Fig. 16: Comparison of ‘+ ‘ lepton pseudo-rapidity distributions for the process Z= ! ‘+ ‘ (n) in HORACE 3.1 including electroweak and QED corrections showered with HERWIG (open red squares), HORACE Born-level showered with HERWIG plus PHOTOS (black circles), and HORACE Born-level (blue stars).

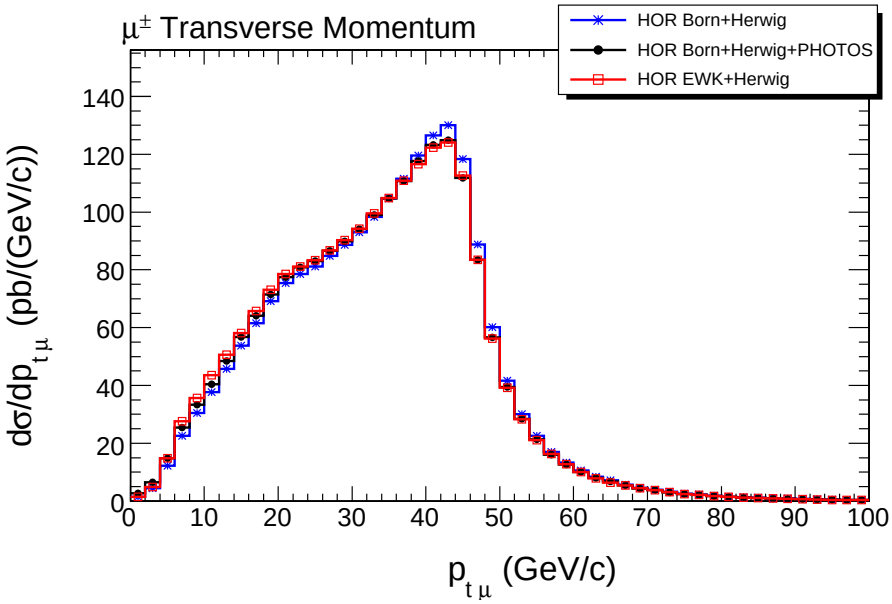

Fig. 17: Comparison of ‘+ ‘ lepton transverse momentum distributions for the process Z= ! ‘+ ‘ (n) in HORACE 3.1 including electroweak and QED corrections showered with HERWIG (open red squares), HORACE Born-level showered with HERWIG plus PHOTOS (black circles), and HORACE Born-level (blue stars).

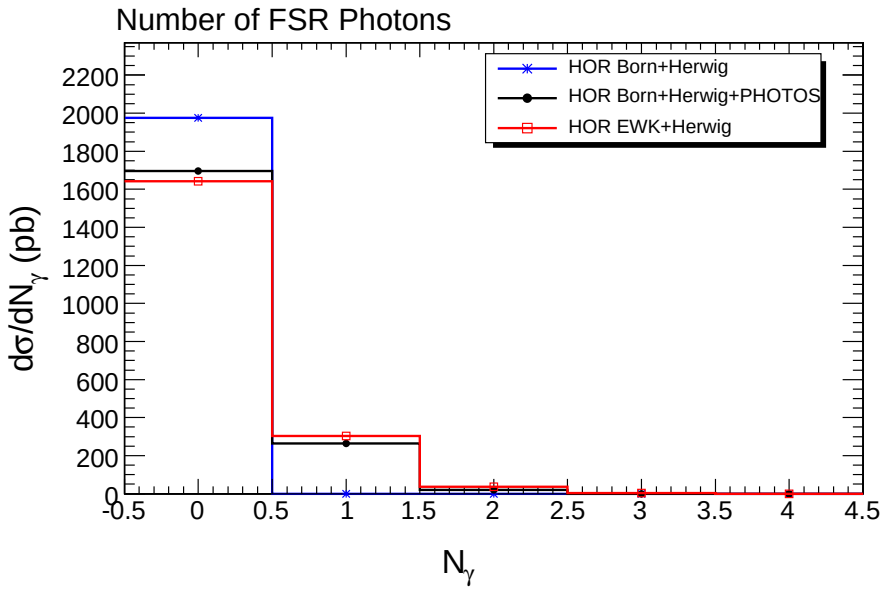

Fig. 18: Comparison of the number n of final state radiation (FSR) photons in Z = ! ‘+ ‘ (n) for HORACE 3.1 including electroweak and QED corrections showered with HERWIG (open red squares), HORACE Born-level showered with HERWIG plus PHOTOS (black circles), and HORACE Born-level (blue stars).

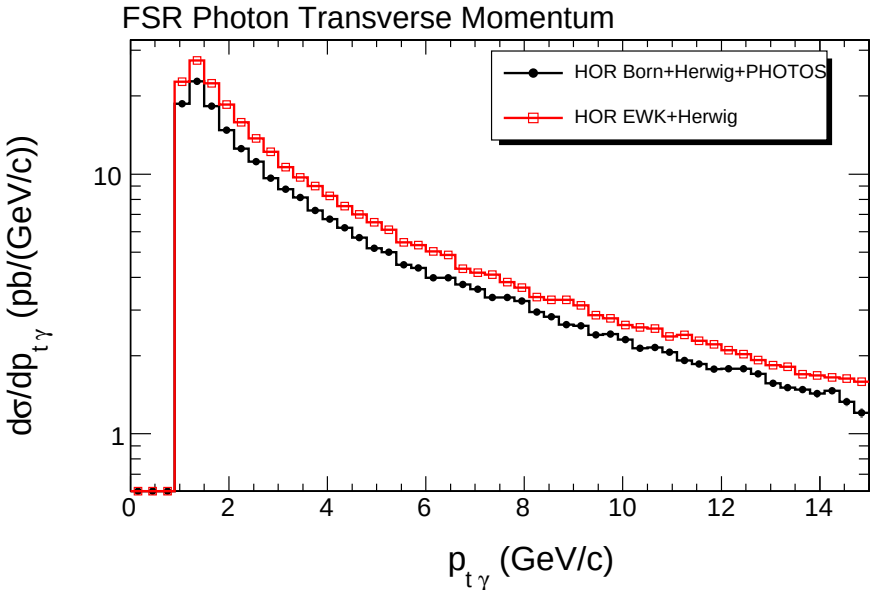

Fig. 19: Comparison of Z = ! ‘+ ‘ (n) final state radiation (FSR) transverse momentum distributions for HORACE 3.1 including electroweak and QED corrections showered with HERWIG (open red squares) and HORACE Born-level showered with HERWIG plus PHOTOS (black circles).

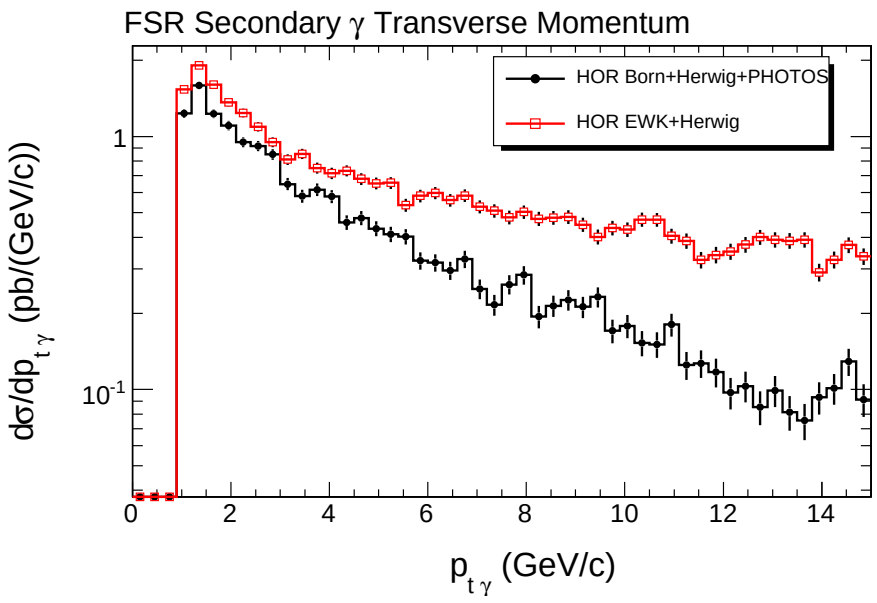

Fig. 20: Comparison of Z = ! ‘+ ‘ (n) secondary final state radiation (FSR) transverse momentum distributions for HORACE 3.1, including electroweak and QED corrections showered with HERWIG (open red squares), and HORACE Born-level showered with HERWIG plus PHOTOS (black circles). Secondary FSR includes any FSR photons other than the first hard photon.

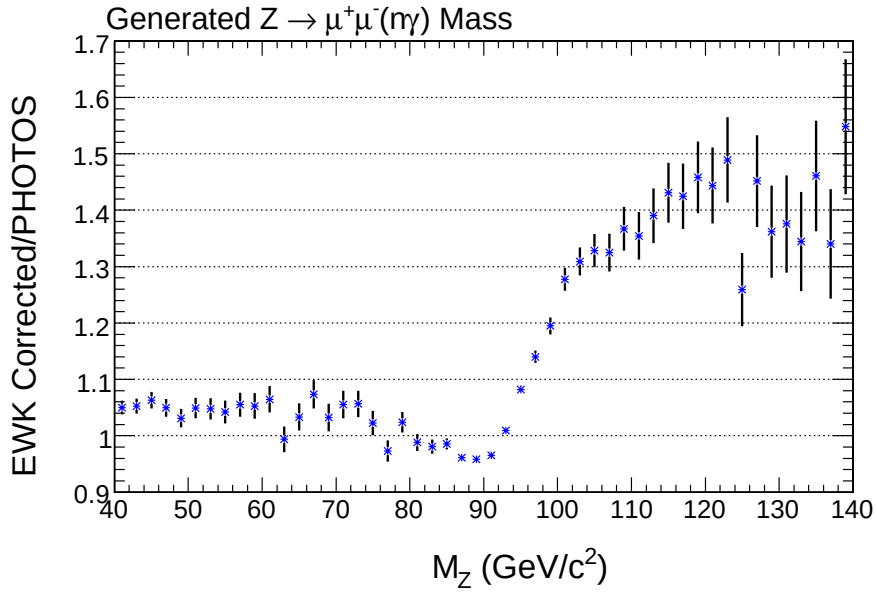

Fig. 21: Ratio of HORACE Z = ! ‘+ ‘ (n) differential cross-section with full EWK corrections, to HORACE with PHOTOS corrections, for the generated Z mass. In this case PHOTOS corrections do not contribute.

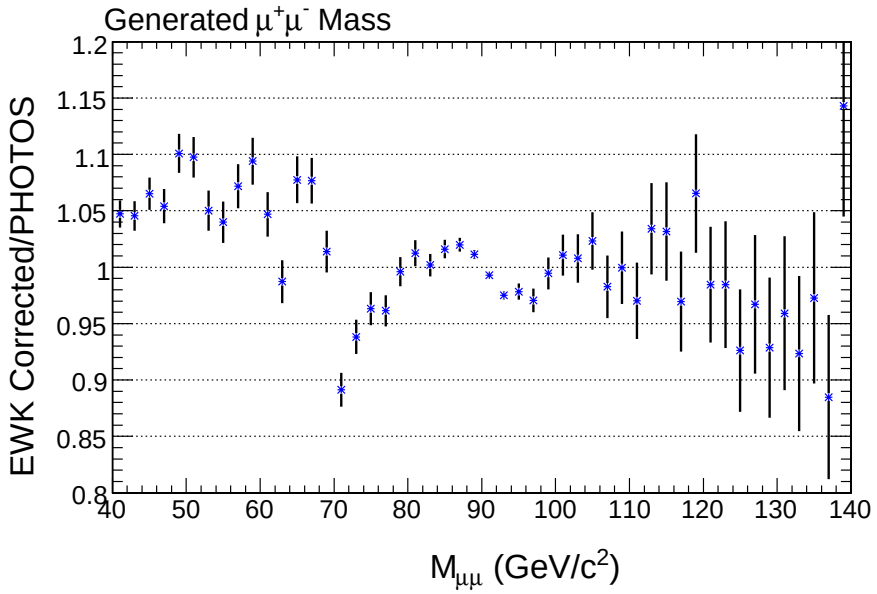

Fig. 22: Ratio of HORACE Z = ! ‘+ ‘ (n) differential cross-section with full EWK corrections, to HORACE with PHOTOS corrections, for the generated + invariant mass after parton and QED showering.

5 ELECTROWEAK CORRECTIONS TO $pp \to W j$[5]

5.1 Introduction

At the LHC, electroweak gauge bosons can recoil against hard jets reaching very high transverse momenta, up to 2 TeV or even beyond. These reactions represent an important background for new-physics searches. Moreover they can be used to determine the parton distribution functions or to measure α_s at the TeV scale. In this kinematic region, the electroweak corrections are strongly enhanced by Sudakov logarithms of the form $\ln(\hat{s}/M_W^2)$ and may amount to tens of percent at one loop and several percent at two loops.[6] The electroweak corrections to $pp \to Z j$ and $pp \to \gamma j$ were studied in Refs. [37, 38, 60, 61]. The electroweak corrections to $pp \to W j$ have been recently completed by two groups [39, 62, 63]. Besides the full set of quark- and gluon-induced $O(\alpha_s)$ reactions, these two calculations include different additional contributions that turn out to be important at high transverse momenta: two-loop Sudakov logarithms [39, 62] and photon-induced processes [63]. We also observe that, while the calculation of Ref. [63] is completely inclusive with respect to photon emission, the definition of the $W j$ cross section adopted in Refs. [39, 62] is more exclusive: $W\gamma$ final states are rejected requiring that the final-state jet has a minimum transverse momentum. However, the numerical results indicate that this difference in the definition of the observable has a quite small impact on the size of the corrections. In the following we present the results of Refs. [39, 62]. In Sect. 5.2 we define the exclusive $pp \to W j$ cross section and discuss the treatment of final-state collinear singularities using quark fragmentation functions. Compact analytic formulae for the high-energy behaviour of the one- and two-loop virtual corrections are presented in Sect. 5.3. Real-photon bremsstrahlung is briefly discussed in Sect. 5.4 and the numerical results are given in Sect. 5.5. For a discussion of QCD corrections we refer to Refs. [19, 64–67].

5.2 Observable definition

The hadronic reaction $pp \to W j(\gamma)$ receives contributions from various partonic subprocesses of the type $q\bar{q}' \to W g(\gamma)$, $gq \to W q'(\gamma)$, and $q\bar{g} \to W \bar{q}'(\gamma)$. Details concerning the implementation of PDFs and quark-mixing effects can be found in Ref. [39]. In the following we focus on the transverse momentum (p_T) distribution[7] for a generic partonic subprocess $ab \to W k(\gamma)$,

$$\frac{d\hat{\sigma}^{ab\to W k(\gamma)}}{dp_T} = \frac{1}{2\hat{s}} \left[\int d\Phi_2 \, |\mathcal{M}^{ab\to W k}|^2 \, F_{O,2}(\Phi_2) + \int d\Phi_3 \, |\mathcal{M}^{ab\to W k}|^2 \, F_{O,3}(\Phi_3) \right]. \quad (15)$$

Here $d\Phi_N$ and $F_{O,N}(\Phi_N)$ denote the phase-space measure and the observable function in the N-particle final-state phase space. The soft and collinear divergences arizing from virtual and real photons need to be extracted in analytic form and, after factorization of initial-state collinear singularities, the singular parts of virtual and real corrections must cancel. Since we are interested in W-boson production in association with a hard jet, we define

$$F_{O,N}(\Phi_N) = \Theta(p_{T,W} > p_{T,W}^{min}) \, \Theta(p_{T,k} > p_{T,j}^{min}); \quad (16)$$

requiring a minimum transverse momentum $p_{T,j}^{min}$ for the final-state parton $k = g, q, \bar{q}$. This observable is free from singularities associated with soft and collinear QCD partons. However, for partonic channels involving final-state quarks (or anti-quarks), the cut on $p_{T,W}$ restricts the emission of collinear photons off quarks and gives rise to collinear singularities. These singularities can be factorized into quark fragmentation functions [68, 69]. Let us consider the quark-photon collinear region,

$$R_q = \sqrt{(\Delta y_{q\gamma})^2 + (\Delta\phi_{q\gamma})^2} < R_{sep}; \quad (17)$$

[5] Contributed by: A. Kulesza, S. Pozzorini, M. Schulze

[6] For a recent survey of the literature on electroweak Sudakov logarithms and their impact at the LHC see Refs. [39, 59].

[7] Summing and averaging over colour and polarization is implicitly understood.

where the rapidity and azimuthal-angle separation between photon and quark becomes small. In practice one can split the 3-particle phase space according to $F_{O,3}(\ _3) = F^{rec}_{O,3}(\ _3) \quad F_{O,3}(\ _3)$, where in

$$F^{rec}_{O,3}(\ _3) = \quad (p_T \quad p_{T,W}\ ) \quad (R_q \quad R_{sep})\ (p_{T,q} \quad p^{min}_{T,j}) + \quad (R_{sep} \quad R_q\ ) \tag{18}$$

the $p_{T,q}$-cut is imposed only outside the collinear region. This contribution is collinear safe and corresponds to the case where collinear photon-quark pairs with $R_q < R_{sep}$ are recombined. The remainder,

$$F_{O,3}(\ _3) = \quad (p_T \quad p_{T,W}\ )\ (R_{sep} \quad R_q\ )\ (p^{min}_{T,j} \quad p_{T,q}); \tag{19}$$

describes the effect of the $p_{T,q}$-cut inside the collinear region. This contribution can be described by means of quark fragmentation functions $D_q(z)$ as[8]

$$\frac{1}{2\hat{s}} \int d\ _3 \overline{|M^{q^0 g! W\ q}|^2}\ F_{O,3}(\ _3) = \frac{d\hat{\ }q^0 g! W\ q}{dp_T} \int_{z_{min}}^1 dz D_q(z); \tag{20}$$

where $z = p_T\ ; = p_{T,W}$ and $z_{min} = 1 \quad p^{min}_{T,j} = p_{T,W}$. The collinear singularities were factorized into the fragmentation function, and using a parametrization derived from measurements of isolated hard photons in hadronic Z decays [69] we obtained $D_q(z) = \frac{Q_q^2}{2}[P_q(z)\ln(zR_{sep}p_{T,W}=0{:}14\,GeV)^2 + z \quad 13{:}26]$. For $R_{sep} < O(1)$ and a wide range of transverse momenta, $2p^{min}_{T,j} \quad p_{T,W} \quad 2\,TeV$, we found that the $F_{O,3}$-contribution (20) does not exceed two permille of the cross section. Therefore we could safely neglect this contribution and perform the calculation using $F_{O,3}(\ _3) \prime F^{rec}_{O,3}(\ _3)$ for final-state (anti-) quarks. We also checked that this approximation is very stable against variations of R_{sep} [39].

5.3 Virtual corrections

The electroweak couplings were renormalized in the G -scheme, where $= \sqrt{2}G_\mu M_W^2 s_w^2 =$ and $s_w^2 = 1 \quad c_w^2 = 1 \quad M_W^2 = M_Z^2$. For transverse momenta of $O(100\,GeV)$ or beyond, the virtual corrections are dominated by logarithms of the type $\ln(\hat{s}=M_W^2)$. In addition, the virtual corrections involve divergent logarithms of electromagnetic origin. The logarithms resulting from photons with virtuality smaller than M_W have been subtracted from the virtual corrections and combined with real-photon emission. As a result, the (subtracted) virtual and real corrections are free from large logarithms involving light-fermion masses, and the bulk of the electroweak effects is isolated in the virtual part (see Sect. 5.5). At one loop, the double and single electroweak logarithms (NLL approximation) can be derived from the general results of Ref. [70]. For the $ud \to W^+g$ subprocess,

$$\overline{|M_1^{ud! W^+g}|^2}^{NLL} = \overline{|M_0^{ud! W^+g}|^2}\left(1 + \frac{}{2}\left(C^{ew}_{q_L}\ln^2\frac{|\hat{s}|}{M_W^2} \quad 3\ln\frac{|\hat{s}|}{M_W^2}\right)\right)$$
$$\frac{C_A}{2s_w^2}\left(\ln^2\frac{|\hat{t}|}{M_W^2} + \ln^2\frac{|\hat{u}|}{M_W^2} \quad \ln^2\frac{|\hat{s}|}{M_W^2}\right); \tag{21}$$

where $\hat{s} = (p_u + p_d)^2$, $\hat{t} = (p_u \quad p_W)^2$, $\hat{u} = (p_d \quad p_W)^2$, $C^{ew}_{q_L} = C_F=s_w^2 + 1=(36c_w^2)$, $C_F = 3=4$, $C_A = 2$ and $\overline{|M_0^{ud! W^+g}|^2} = 32\ ^2{}_S(\ =s_w^2)(\hat{t}^2 + \hat{u}^2 + 2M_W^2\hat{s})=(\hat{t}\hat{u})$. This result is easily extended to all relevant partonic reactions by means of CP and crossing symmetries.

The exact one-loop expression for the (subtracted) virtual corrections has the general form

$$\overline{|M_1^{ud! W^+g}|^2} = 1 + 2Re\ C^A + C^N\ \overline{|M_0^{ud! W^+g}|^2} + \frac{16\ ^2{}_S}{s_w^2}Re\left(\frac{1}{2s_w^2}H_1^X(M_W^2) + \sum_{V=A,Z}\left(\frac{3}{4s_w^2}{}^{SU(2)}_{VV} + \frac{}{36c_w^2}{}^{U(1)}_{VV}\right)H_1^A(M_V^2) + \frac{}{2s_w^2}{}^{SU(2)}_{VV}\left(2H_1^N(M_V^2) \quad H_1^X(M_V^2) \quad \frac{X_V}{6}H_1^Y(M_V^2)\right)\right); \tag{22}$$

[8] For a detailed discussion we refer to App. A of Ref. [39].

where $^{SU(2)}_{AA} = s_w^2$, $X_A = -1$, $^{U(1)}_{AA} = c_w^2$, $^{SU(2)}_{ZZ} = c_w^2$, $X_Z = 1$ and $^{U(1)}_{ZZ} = s_w^2$. Explicit expressions for the functions $H_1^I(M_V^2)$ and the counterterms C^A, C^N can be found in Ref. [39]. Here we present compact NNLL expressions in the high-energy limit. This approximation includes all terms that are not suppressed by powers of $M_W^2/\hat{s}$. The NNLL expansion of the loop diagrams involving massive gauge bosons ($M_V = M_Z$; M_W) yields

$$
\begin{aligned}
H_1^A(M_V^2) &= \frac{\hat{t}^2+\hat{u}^2}{\hat{t}\hat{u}}\left(\Delta_{UV} + \ln\frac{M_Z^2}{M_V^2}\ln\frac{\hat{s}}{M_V^2} + 3\ln\frac{\hat{s}}{M_V^2} + \frac{3}{2}\ln^2\frac{\hat{t}}{\hat{s}}\right.\\
&\quad + \ln^2\frac{\hat{u}}{\hat{s}} + \ln\frac{\hat{t}}{\hat{s}} + \ln\frac{\hat{u}}{\hat{s}} + \frac{7\pi^2}{3} - 3 + \frac{\hat{t}^2-\hat{u}^2}{2\hat{t}\hat{u}}\left[\ln^2\frac{\hat{t}}{\hat{s}} - \ln^2\frac{\hat{u}}{\hat{s}} + 3\ln\frac{\hat{u}}{\hat{s}}\right.\\
&\quad \left.\left. - 3\ln\frac{\hat{t}}{\hat{s}} + 2\ln^2\frac{\hat{t}}{\hat{s}} + \ln^2\frac{\hat{u}}{\hat{s}} + \ln\frac{\hat{t}}{\hat{s}} + \ln\frac{\hat{u}}{\hat{s}} - 2\ln\frac{M_V^2}{M_W^2} + 4\pi^2\right];\right.\\
H_1^N(M_V^2) &= \frac{\hat{t}^2+\hat{u}^2}{\hat{t}\hat{u}}\left(2\Delta_{UV} + \ln\frac{M_Z^2}{M_W^2} + \ln\frac{M_V^2}{M_W^2} + \ln^2\frac{\hat{s}}{M_V^2} - \frac{1}{2}\ln^2\frac{\hat{t}}{M_V^2}\right.\\
&\quad + \ln^2\frac{\hat{t}}{M_W^2} + \ln^2\frac{\hat{u}}{M_V^2} + \ln^2\frac{\hat{u}}{M_W^2} + \ln^2\frac{\hat{t}}{\hat{u}} - \frac{3}{2}\ln^2\frac{\hat{t}}{\hat{s}} + \ln^2\frac{\hat{u}}{\hat{s}}\\
&\quad - \frac{20\pi^2}{9} - \frac{2\pi}{\sqrt{3}} + 4 + \frac{\hat{t}^2-\hat{u}^2}{2\hat{t}\hat{u}}\left[\ln^2\frac{\hat{u}}{\hat{s}} - \ln^2\frac{\hat{t}}{\hat{s}} - 2\ln^2\frac{\hat{t}}{\hat{s}} + \ln^2\frac{\hat{u}}{\hat{s}} + \ln\frac{\hat{t}}{\hat{s}}\right.\\
&\quad \left.\left. + \ln\frac{\hat{u}}{\hat{s}} + 2\ln\frac{M_V^2}{M_W^2} - 4\pi^2\right];\right.\\
H_1^Y(M_V^2) &= \frac{\hat{t}^2+\hat{u}^2}{\hat{t}\hat{u}}\left(\ln^2\frac{\hat{t}}{M_W^2} - \ln^2\frac{\hat{t}}{M_V^2} - \ln^2\frac{\hat{u}}{M_W^2} + \ln^2\frac{\hat{u}}{M_V^2} + 2\ln\frac{\hat{t}}{\hat{u}}\right);\\
H_1^X(M_V^2) &= -2 - 2\ln\frac{\hat{s}}{M_V^2} + \ln\frac{\hat{t}}{\hat{s}} + \ln\frac{\hat{u}}{\hat{s}} - 3;
\end{aligned}
\tag{23}
$$

where $\Delta_{UV} = 1/\varepsilon_E + \ln(4\pi) + \ln\mu^2/M_Z^2$. For the loop functions associated with photons we obtain $H_1^I(M_A^2) = H_1^I(M_W^2) + \frac{\hat{t}^2+\hat{u}^2}{\hat{t}\hat{u}}K^I$ with $K^A = \pi^2$, $K^N = 2\pi/\sqrt{3} - 7\pi^2/9$, and $K^X = K^Y = 0$. The functions describing the photonic and the W-boson contributions differ only by non-logarithmic terms, since the logarithms from photons with virtuality smaller than M_W have been subtracted.

At two loops , using the general results for leading- and next-to-leading electroweak logarithms in Refs. [71, 72] and subtracting logarithms from photons with virtuality smaller than M_W, we obtain $|M_2^{u\bar{d}\to W^+g}|^2 = |M_1^{u\bar{d}\to W^+g}|^2 + (\frac{\alpha}{2\pi})^2 A^{(2)}|M_0^{u\bar{d}\to W^+g}|^2$ with

$$
\begin{aligned}
A^{(2)} &= \frac{1}{2}\left(C^{ew}_{q_L} + \frac{C_A}{2s_w^2}\left(C^{ew}_{q_L} - \ln^4\frac{|\hat{s}|}{M_W^2} - 6\ln^3\frac{|\hat{s}|}{M_W^2} + \frac{C_A}{2s_w^2}\left[\ln^4\frac{|\hat{t}|}{M_W^2}\right.\right.\right.\\
&\quad \left.\left.\left. + \ln^4\frac{|\hat{u}|}{M_W^2} - \ln^4\frac{|\hat{s}|}{M_W^2}\right] + \frac{1}{6}\left(\frac{b_1}{c_w^2}\left(\frac{Y_{q_L}}{2}\right)^2 + \frac{b_2}{s_w^2}\right)\left(C_F + \frac{C_A}{2}\right)\ln^3\frac{|\hat{s}|}{M_W^2}\right);
\end{aligned}
\tag{24}
$$

where $b_1 = -41/(6c_w^2)$ and $b_2 = 19/(6s_w^2)$.

5.4 Real photon radiation

We performed two independent calculations of real photon bremsstrahlung using the dipole subtraction method [73–75]. In the first calculation, we used the subtraction method for massive fermions [73] regularizing soft and collinear singularities by means of small photon and fermion masses. In the second

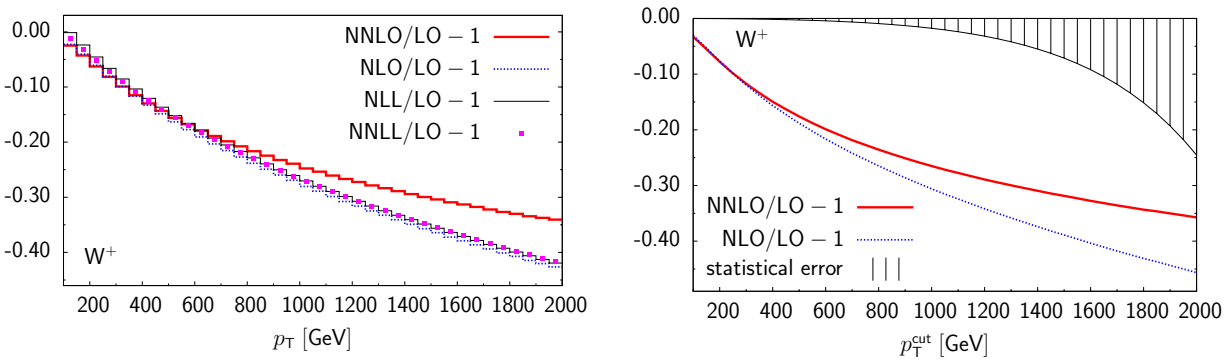

Fig. 23: Electroweak correction to $pp \to W^+ j$ at $\sqrt{s} = 14\,TeV$: (a) relative NLO (dotted), NLL (thin solid), NNLL (squares) and NNLO (thick solid) correction wrt. the LO p_T-distribution; (b) NLO (dotted) and NNLO (solid) corrections to the integrated cross section and estimated statistical error (shaded area).

calculation we used massless fermions and we subtracted the singularities in the framework of dimensional regularization [74, 75]. The initial-state collinear singularities were factorized in the $\overline{MS}$ scheme. This procedure introduces a logarithmic dependence on the QED factorization scale $_{QED}$, which must be compensated by the QED evolution of the PDFs. Since our calculation is of LO in $_S$, for consistency we should use LO QCD parton distributions including NLO QED effects. However, such a PDF set is not available.[9] Thus we used a LO QCD PDF set without QED corrections [76], and we chose the value of $_{QED}$ in such a way that the neglected QED effects are small. In Ref. [77] it was shown that the QED corrections to the quark distribution functions grow with $_{QED}$ but do not exceed one percent for $_{QED} < 100\,GeV$. Thus we set $_{QED} = M_W$. Photon-induced processes were not included in our calculation. These contributions are parametrically suppressed by a factor $= _S$. However in Ref. [63] it was found that, at very large p_T, these photon-induced effects can amount to several percent.

5.5 Numerical results

The hadronic cross section was obtained using LO MRST2001 PDFs [76] at the factorization and renormalization scale $^2_{QCD} = p_T^2$. For the jet we required a minimum transverse momentum $p_{T;j}^{min} = 100\,GeV$, and the value of the separation parameter in (17) was set to $R_{sep} = 0{:}4$. The input parameters are specified in Ref. [39]. Here we present the electroweak corrections to $pp \to W^+ j$ at $\sqrt{s} = 14\,TeV$. The corrections to W production are almost identical [39]. In Fig. 23a we plot the relative size of the electroweak corrections wrt. the LO W-boson p_T-distribution. The exact $O(\ )$ correction (NLO curve) increases significantly with p_T and ranges from -15% at $p_T = 500\,GeV$ to -43% at $p_T = 2\,TeV$. This enhancement is clearly due to the Sudakov logarithms that are present in the virtual corrections. Indeed the one-loop NLL and NNLL approximations, which describe the virtual part of the corrections in the Sudakov regime, are in very good agreement with the full NLO result. The difference between the NLO and NNLO curves corresponds to the two-loop Sudakov logarithms. Their contribution is positive and becomes significant at high p_T. It amounts to $+3\%$ at $p_T = 1\,TeV$ and $+9\%$ at $p_T = 2\,TeV$. In Fig. 23b we consider the integrated cross section for $p_T > p_T^{cut}$ and, to underline the relevance of the large electroweak corrections, we compare the relative NLO and NNLO corrections with the statistical accuracy at the LHC. This latter is estimated using the integrated luminosity $L = 300\,fb^{-1}$ and the branching ratio BR($W \to e_e + \ $) $= 2/9$. The size of the NLO corrections is clearly much bigger than the statistical error. Also the two-loop logarithmic effects are significant. In terms of the estimated statistical error they amount to 1–3 standard deviations for p_T of $O(1\,TeV)$. The relative importance of the virtual (NLO_{virt}) and real (NLO_{real}) contributions is shown in Fig. 24a. The electromagnetic logarithms have been subtracted from the virtual part and added to the real one as explained in Sect. 5.3 As a consequence, the

[9] The currently available PDFs incorporating NLO QED corrections (MRST2004QED) include QCD effects at the NLO.

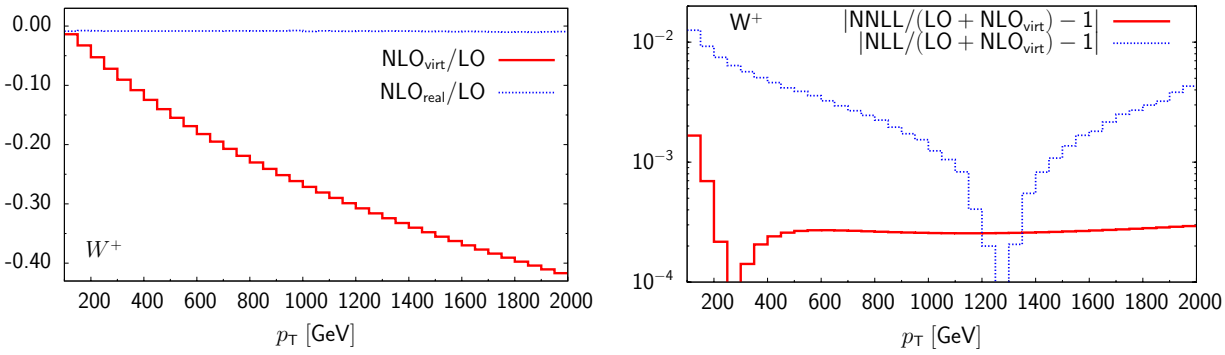

Fig. 24: p_T-distribution of W bosons in the process pp ! W + j at $\sqrt{s}$ = 14 TeV: (a) relative importance of the virtual (NLO virt) and real (NLO real) corrections; (b) precision of the NNLL (solid) and NLL (dashed) one-loop approximations.

bulk of the corrections is isolated in the virtual part, which grows with p_T and amounts up to 42% at p_T = 2 TeV. In contrast, the real part represents a small and nearly constant corrections of about 1%. In presence of additional cuts on hard photons, NLO real becomes more negative and can amount up to 5% for p_T ' 1 TeV [39]. As illustrated in Fig. 24b, the NLL and NNLL one-loop approximations provide a very precise description of the high-energy behaviour of the NLO virt part. For p_T 200 GeV, the precision of the NLL and NNLL approximations is better than 10 2 and 10 3, respectively.

Conclusions

We evaluated the electroweak corrections to large transverse momentum production of W bosons at the LHC, including the contributions from virtual and real photons. The singularities resulting from photons with virtuality smaller than M W have been subtracted from the virtual contributions and combined with real-photon bremsstrahlung. As a result, the bulk of the electroweak effects is isolated in the virtual contributions, which are enhanced by Sudakov logarithms and give rise to corrections of tens of percent at high p_T. We presented compact analytic approximations that describe these virtual effects with high precision. The complete O( ) corrections range between -15% and -40% for 500 GeV p_T 2 TeV. Considering the large event rate at the LHC, leading to a fairly good statistical precision even at transverse momenta up to 2 TeV, we evaluated also the dominant two-loop Sudakov logarithms. In the high-p_T region, these two-loop effects increase the cross section by 5-10% and thus become of importance in precision studies.

Acknowledgements

We would like to thank S. Dittmaier, B. Jäger and P. Uwer for helpful discussions.

6 SOME INTERESTING MIN-BIAS DISTRIBUTIONS FOR EARLY LHC RUNS [10]

6.1 Introduction

At first glance, the confined nature of both the initial and final state implies that there are no perturbatively calculable observables in inelastic hadron-hadron collisions. Under ordinary circumstances, however, two powerful tools are used to circumvent this problem, factorisation and infrared safety. The trouble with minimum-bias and underlying-event (MB/UE) physics is that the applicability of both of these tools is, at best, questionable for a wide range of interesting observables.

To understand why the main perturbative tools are ineffective, let us begin with factorisation. When applicable, factorisation allows us to subdivide the calculation of an observable (regardless of whether it is infrared safe or not) into a perturbatively calculable short-distance part and a universal long-distance part, the latter of which may be modeled and constrained by fits to data. However, in the context of hadron collisions the oft made separation into "hard scattering" and "underlying event" components is not necessarily equivalent to a clean separation in terms of formation/fluctuation time, since the underlying event may contain short-distance physics of its own. Regardless of which definition is more correct, any breakdown of the assumed factorisation could introduce a process-dependence of the long-distance part, leading to an unknown systematic uncertainty in the procedure of measuring the corrections in one process and applying them to another.

The second tool, infrared safety, provides us with a class of observables which are insensitive to the details of the long-distance physics. This works up to corrections of order the long-distance scale divided by the short-distance scale, Q_{IR}^n/Q_{UV}^n, where the power n depends on the observable in question and $Q_{IR,UV}$ denote generic infrared and ultraviolet scales in the problem. Since $Q_{IR}/Q_{UV} \to 0$ for large Q_{UV}, such observables "decouple" from the infrared physics as long as all relevant scales are Q_{IR}. Infrared sensitive quantities, on the other hand, contain logarithms $\log^n(Q_{UV}^2/Q_{IR}^2)$ which grow increasingly large as $Q_{IR}/Q_{UV} \to 0$. In MB/UE studies, many of the important measured distributions are not infrared safe in the perturbative sense. Take particle multiplicities, for instance; in the absence of non-trivial infrared effects, the number of partons that would be mapped to hadrons in a naïve local-parton-hadron-duality [78] picture depends logarithmically on the infrared cutoff.

We may thus classify collider observables in four categories: least intimidating are the factorisable infrared safe quantities, such as the R ratio in e^+e annihilation, which are only problematic at low scales (where the above-mentioned power corrections can be large). Then come the factorisable infrared sensitive quantities, with the long-distance part parametrised by process-independent non-perturbative functions, such as parton distributions. Somewhat nastier are non-factorised infrared safe observables. An example could here be the energy flow into one of Rick Field's "transverse regions" [79]. The energy flow is nominally infrared safe, but in these regions where bremsstrahlung is suppressed there can be large contributions from pairwise balancing minijets which are correlated to the hard scattering and hence do not factorise according to at least one of the definitions outlined above (see also [80, 81]). The nastiest beasts by all accounts are non-factorised infrared sensitive quantities, such as the particle multiplicity in the transverse region.

The trouble, then, is that MB/UE physics is full of distributions of the very nastiest kinds imaginable. Phenomenologically, the implication is that the theoretical treatment of non-factorised and non-perturbative effects becomes more important and the interpretation of experimental distributions correspondingly more involved. The problem may also be turned around, noting that MB/UE offers an ideal lab for studying these theoretically poorly understood phenomena; the most interesting observables and cuts, then, are those which minimise the "backgrounds" from better-known physics.

As part of the effort to spur more interplay between theorists and experimentalists in this field, we here present a collection of simple min-bias distributions that carry interesting and complementary information about the underlying physics, both perturbative and non-perturbative. The main point is

[10] Contributed by: P. Z. Skands

Model	Showers off MPI	MPI $p_\perp$ cutoff at $1.96\ \to\ 14$ TeV	FS Colour Correlations	Shower Ordering	Proton Lumpiness	Tevatron Constraints
A	No	$2.04 \overset{\text{fast}}{\to} 3.34$	Strong	Q^2	More	MB, UE
DW	No	$1.94 \overset{\text{fast}}{\to} 3.17$	Strong	Q^2	More	MB, UE, DY
DWT	No	$1.94 \overset{\text{slow}}{\to} 2.66$	Strong	Q^2	More	MB, UE, DY
S0	Yes	$1.88 \overset{\text{slow}}{\to} 2.57$	Strong	$p_\perp^2$	Less	MB, DY
S0A	Yes	$1.89 \overset{\text{fast}}{\to} 3.09$	Strong	$p_\perp^2$	Less	MB, DY
ATLAS	No	$2.00 \overset{\text{slow}}{\to} 2.75$	Weak	Q^2	More	UE

Table 4: Brief overview of models. Note that the IR cutoff in these models is not imposed as a step function, but rather as a smooth dampening, see [88, 89]. The labels $\overset{\text{fast}}{\to}$ and $\overset{\text{slow}}{\to}$ refer to the pace of the scaling of the cutoff with collider energy.

that, while each plot represents a complicated cocktail of physics effects, such that most models could probably be tuned to give an acceptable description observable by observable, it is very difficult to simultaneously describe the entire set. It should therefore be possible to carry out systematic physics studies beyond simple tunings. For brevity, this text only includes a representative selection, with more results available on the web [82]. Note also that we have here left out several important ingredients which are touched on elsewhere in these proceedings, such as observables involving explicit jet reconstruction and observables in leading-jet, dijet, jet + photon, and Drell-Yan events. See also the underlying-event sections in the HERA-and-the-LHC [83] and Tevatron-for-LHC [84] writeups.

6.2 Models

We have chosen to consider a set of six different tunes of the PYTHIA event generator [85], called A, DW, and DWT [79, 84], S0 and S0A [86], and ATLAS-DC2 / Rome [87]. For min-bias, all of these start from leading order QCD $2 \to 2$ matrix elements, augmented by initial- and final-state showers (ISR and FSR, respectively) and perturbative multiple parton interactions (MPI) [88, 89], folded with CTEQ5L parton distributions [90] on the initial-state side and the Lund string fragmentation model [91] on the final-state side. In addition, the initial state is characterised by a transverse mass distribution roughly representing the degree of lumpiness in the proton[11] and by correlated multi-parton densities derived from the standard ones by imposing elementary sum rules such as momentum conservation [88] and flavour conservation [94]. The final state, likewise, is subject to several effects unique to hadronic collisions, such as the treatment of beam remnants (e.g., affecting the flow of baryon number) and colour (re-)connection effects between the MPI final states [86, 88, 95].

Although not perfectly orthogonal in "model space", these tunes are still reasonably complementary on a number of important points, as illustrated in tab. 4. Column by column in tab. 4, these differences are as follows: 1) showers off the MPI are only included in S0(A). 2) the MPI infrared cutoff scale evolves faster with collision energy in tunes A, DW, and S0A than in S0 and DWT. 3) all models except the ATLAS tune have very strong final-state colour correlations. 4) tunes A, DW(T), and ATLAS use Q^2-ordered showers and the old MPI framework, whereas tunes S0(A) use the new interleaved $p_\perp$-ordered model. 5) tunes A and DW(T) have transverse mass distributions which are significantly more peaked than Gaussians, with ATLAS following close behind, and S0(A) having the smoothest distribution. 6) the models were tuned to describe one or more of min-bias (MB), underlying-event (UE), and/or Drell-Yan (DY) data at the Tevatron.

Tunes DW and DWT only differ in the energy extrapolation away from the Tevatron and hence are

[11] Note that the impact-parameter dependence is still assumed factorised from the x dependence in these models, $f(x;b) = f(x)g(b)$, where b denotes impact parameter, a simplifying assumption that by no means should be treated as inviolate, see e.g. [81, 92, 93].

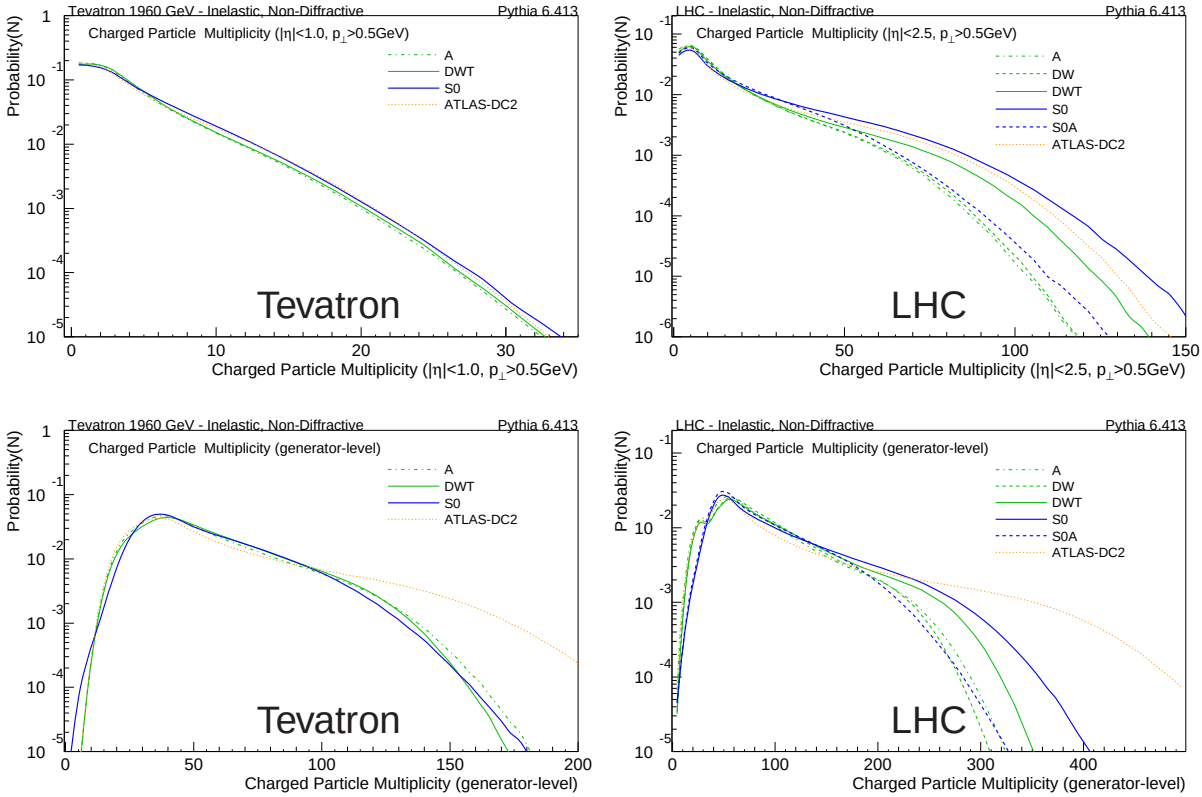

Fig. 25: Charged particle multiplicity distributions, at fiducial (top) and generator (bottom) levels, for the Tevatron (left) and LHC (right). The fiducial averages range from $3.3 < \langle N_{ch} \rangle < 3.6$ at the Tevatron to $13.0 < \langle N_{ch} \rangle < 19.3$ at the LHC.

only shown separately at the LHC. Likewise for S0 and S0A. We regret not including a comparison to other MB/UE Monte Carlo generators, but note that the S0(A) models are very similar to PYTHIA 8 [96], apart from the colour (re-)connection model and some subtleties connected with the parton shower, and that the SHERPA [32] model closely resembles the Q^2-ordered models considered here, with the addition of showers off the MPI. The JIMMY add-on to HERWIG [97, 98] is currently only applicable to underlying-event and not to min-bias.

6.3 Results

In this section we focus on the following distributions for inelastic non-diffractive events at the Tevatron and LHC: charged particle multiplicity $P(N_{ch})$, dN_{ch}/dp_T, $dN_{ch}/d\eta$, the average p_T vs. N_{ch} correlation, the forward-backward N_{ch} and E_T correlations vs. η, as well as a few plots of theoretical interest showing the multiplicity distribution of multiple interactions $P(N_{int})$. On most of the plots we include the effects of fiducial cuts, which are represented by the cuts $p_T > 0.5$ GeV and $|\eta| < 1.0$ ($|\eta| < 2.5$) at the Tevatron (LHC).

The charged particle multiplicity is shown in fig. 25, both including fiducial cuts (top row) and at generator-level (bottom row). Tevatron results are shown to the left and LHC ones to the right. Given the amount of tuning that went into all of these models, it is not surprising that there is general agreement on the charged track multiplicity in the fiducial region at the Tevatron (top left plot). In the top right plot, however, it is clear that this near-degeneracy is broken at the LHC, due to the different energy extrapolations, and hence even a small amount of data on the charged track multiplicity will yield important constraints. The bottom row of plots shows how things look at the generator-level, i.e., without fiducial cuts. An important difference between the ATLAS tune and the other models emerges. The ATLAS tune has a significantly higher component of unobserved charged multiplicity. This highlights the fact that

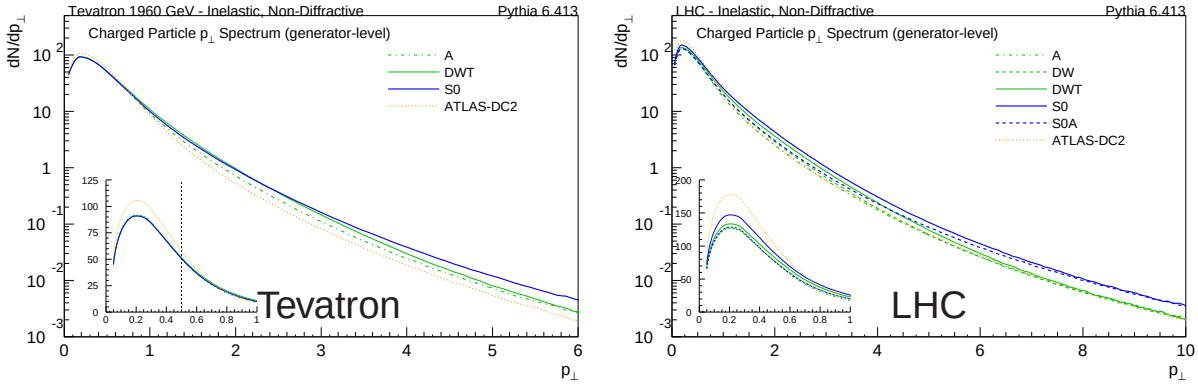

Fig. 26: Charged particle p_T spectrum, generator-level only. Insets show the region below 1 GeV on a linear scale. The fiducial distributions [82] are very similar, apart from an overall normalisation and the cut at $p_T = 0.5$ GeV.

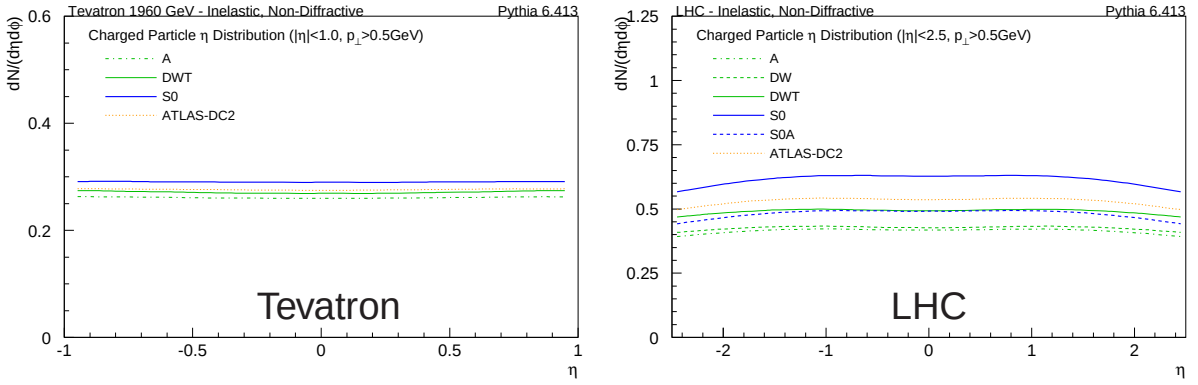

Fig. 27: Charged particle density vs. pseudorapidity, fiducial distribution only. The generator-level ones can be found at [82].

extrapolations from the measured distribution to the generator-level one are model-dependent.

The cause for the difference in unobserved multiplicity can be readily identified by considering the generator-level p_T spectra of charged particles, fig. 26. The small insets show the region below 1 GeV on a linear scale, with the cut at $p_T = 0.5$ GeV shown as a dashed line. Below the fiducial cut, the ATLAS tune has a significantly larger soft peak than the other models. The S0 model, on the other hand, has a harder distribution in the tail, which also causes S0 to have a slightly larger overall multiplicity in the central region, as illustrated in the fiducial pseudorapidity distributions, fig. 27. Apart from the overall normalisation, however, the pseudorapidity distribution is almost featureless except for the tapering off towards large $|\eta|$ at the LHC. Nonetheless, we note that to study possible non-perturbative fragmentation differences between LEP and hadron colliders, quantities that would be interesting to plot vs. this axis would be strangeness and baryon fractions, such as $N_{K^0_S}/N_{ch}$ and $N_{\Lambda^0}/(N_{\Lambda^0} + N_{\bar{\Lambda}^0})$, as well as the the p_T spectra of these particles. With good statistics, also multi-strange baryons would carry interesting information, as has been studied in pp collisions in particular by the STAR experiment [99, 100].

Before going on to correlations, let us briefly consider how the multiplicity is built up in the various models. Fig. 28 shows the probability distribution of the number of multiple interactions. This distribution essentially represents a folding of the multiple-interactions cross section above the infrared cutoff with the assumed transverse matter distribution. Firstly, the ATLAS and Rick Field tunes have almost identical infrared cutoffs and transverse mass profiles and hence look very similar. (Since ATLAS and DWT have the same energy extrapolation, these are the most similar at LHC.) On the other hand, the S0(A) models exhibit a significantly smaller tail towards large numbers of interactions caused by a combination of the smoother mass profile and the fact that the MPI are associated with ISR showers of

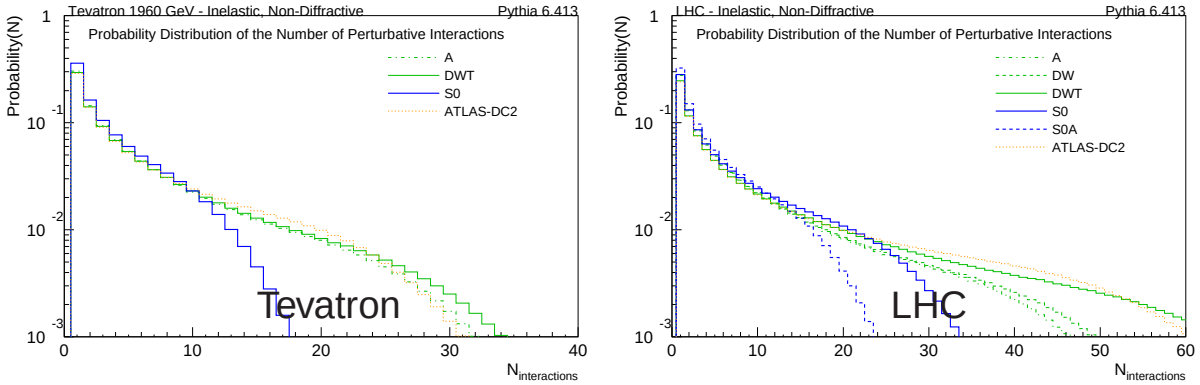

Fig. 28: Probability distribution of the number of multiple interactions. The averages range from $3{:}7 < \langle N_{\mathrm{int}} \rangle < 6{:}1$ at the Tevatron to $4{:}7 < \langle N_{\mathrm{int}} \rangle < 11{:}2$ at the LHC.

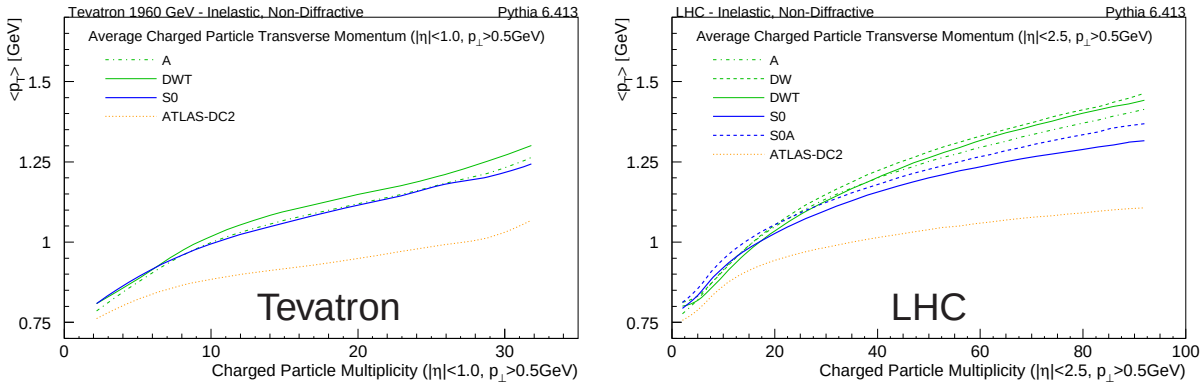

Fig. 29: The average track transverse momentum vs. the number of tracks, counting fiducial tracks only, for events with at least one fiducial track.

their own, hence each takes a bigger x fraction.

Fig. 29 shows the first non-trivial correlation, the average track momentum (counting fiducial tracks only) vs. multiplicity for events with at least one charged particle passing the fiducial cuts. The general trend is that the tracks in high-multiplicity events are harder on average than in low-multiplicity ones. This agrees with collider data and is an interesting observation in itself. We also see that the tunes roughly agree for low-multiplicity events, while the ATLAS tune falls below at high multiplicities. In the models here considered, this is tightly linked to the weak final-state colour correlations in the ATLAS tune; the naive expectation from an uncorrelated system of strings decaying to hadrons would be that $\langle p_\perp \rangle$ should be independent of N_{ch}. To make the average $p_\perp$ rise sufficiently to agree with Tevatron data, tunes A, DW(T), and S0(A) incorporate strong colour correlations between final-state partons from different interactions, chosen in such a way as to minimise the resulting string length. An alternative possible explanation could be Cronin-effect-type rescatterings of the outgoing partons, a preliminary study of which is in progress [101].

An additional important correlation, which carries information on local vs. long-distance fluctuations, is the forward-backward correlation strength, b, defined as [88, 102, 103]

$$b = \frac{\langle n_F n_B \rangle \quad \langle n_F \rangle^2}{n_F^2 \qquad \langle n_F \rangle^2} ; \tag{25}$$

where n_F (n_B) is the number of charged particles in a forward (backward) pseudorapidity bin of fixed size, separated by a central interval centred at zero. The UA5 study [102] used pseudorapidity

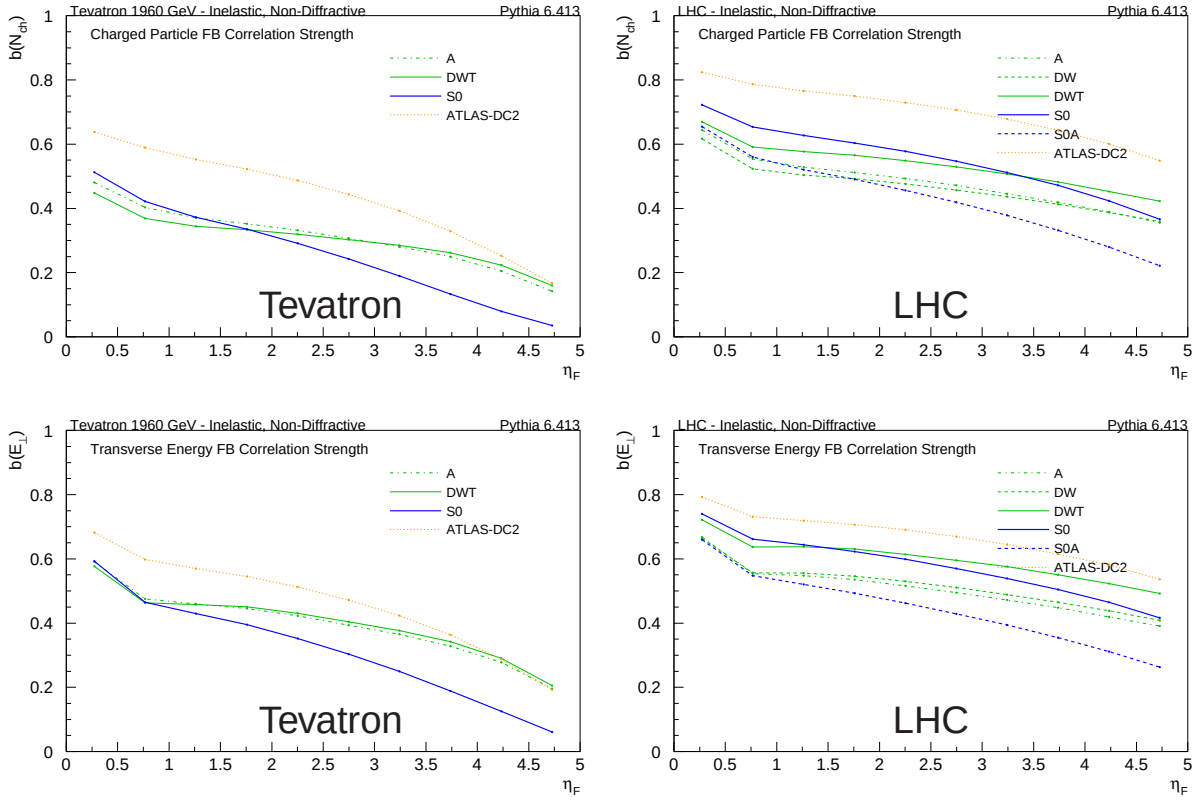

Fig. 30: Generator-level forward-backward correlation strength, b, for charged particles (top) and transverse energy (bottom).

bins one unit wide and plotted the correlation vs. the rapidity difference, . For comparison, STAR, which has a much smaller coverage, uses 0.2-unit wide bins [104]. However, as shown in a recent study [105], small bins increase the relative importance of statistical fluctuations, washing out the genuine correlations. For the Tevatron and LHC detectors, which also have small coverages, we therefore settle on a compromise of 0.5-unit wide bins. We also choose to plot the result vs. the pseudorapidity of the forward bin, F =2, such that the x axis corresponds directly to a pseudorapidity in the detector (the backward bin is then situated symmetrically on the other side of zero). Fig. 30 shows the generator-level correlations, both for charged particles (top row) and for a measure of transverse energy (bottom row), here defined as the p? sum of all neutral and charged particles inside the relevant rapidity bins. Note that we let the x axis extend to pseudorapidities of 5, outside the measurable region, in order to get a more comprehensive view of the behaviour of the distribution. The fact that the ATLAS and S0(A) distributions have a more steeply falling tail than A and DW(T) again reflects the qualitatively different physics cocktails represented by these models. Our tentative conclusions are as follows: Rick Field's tunes A, DW, and DWT have a large number of multiple interactions, cf. fig. 28, but due to the strong final-state colour correlations in these tunes, the main effect of each additional interaction is to add "wrinkles" and energy to already existing string topologies. Their effects on short-distance correlations are therefore suppressed relative to the ATLAS tune, which exhibits similar long-distance correlations but stronger short-distance ones. S0(A) has a smaller total number of MPI, cf. fig. 28, which leads to smaller long-distance correlations, but it still has strong short-distance ones. In summary, the b distributions are clearly sensitive to the relative mix of MPI and shower activity. They also depend on the detailed shape of fig. 28, which in turn is partly controlled by the transverse matter density profile. Measurements of these distributions, both at present and future colliders, would therefore add another highly interesting and complementary piece of information on the physics cocktail.

6.4 Conclusion and outlook

We have illustrated some elementary distributions in inelastic, non-diffractive events at the Tevatron and LHC, as they look with various tunes of the two underlying-event models in the PYTHIA event generator. In particular, taking the charged particle multiplicity distribution to set the overall level of the MB/UE physics, the p_T spectrum of charged particles and the $\langle p_T \rangle(N_{ch})$ correlations then add important information on aspects such as final-state colour correlations. Identified-particle spectra would yield further insight on beam remnants and hadronization in a hadron-collider environment. Finally, correlations in multiplicity and energy vs. pseudorapidity can be used to extract information on the importance of short-distance vs. long-distance correlations, which (very) roughly correspond to the type of fluctuations produced by shower- and multiple-interaction-activity, respectively.

By comparing the multiplicity distributions with and without fiducial cuts, we note that the extrapolation from observed to generator-level distributions can be highly model-dependent. It is therefore important to extend the measured region as far as possible in both and p_T.

On the phenomenological side, several remaining issues could still be addressed without requiring a more formal footing (see below). These include parton rescattering effects (Cronin effect) [101], correlations between x- and impact-parameter-dependence in the multi-parton PDFs [80, 92, 93], saturation and small-x effects [106], improved modeling of baryon production [94, 107, 108], possible breakdowns of jet universality between LEP, HERA, and hadron colliders, and closer studies of the correspondence between coherent phenomena, such as diffraction and elastic scattering, and inelastic non-diffractive processes [81, 109].

Further progress would seem to require a systematic way of improving on the phenomenological models, both on the perturbative and non-perturbative sides, which necessitates some degree of formal developments in addition to more advanced model building. The correspondence with fixed-order QCD is already being elucidated by parton-shower / matrix-element matching methods, already a well-developed field. Though these methods are currently applied mostly to X +jet-type topologies, there is no reason they should not be brought to bear on MB/UE physics as well. Systematic inclusion of higher-order effects in showers (beyond that offered by "clever choices" of ordering, renormalisation, and kinematic variables) would also provide a more solid foundation for the perturbative side of the calculation, though this is a field still in its infancy [110, 111]. To go further, however, factorisation in the context of hadron collisions needs to be better understood, probably including by now well-established short-distance phenomena such as multiple perturbative interactions on the "short-distance" side and, correspondingly, correlated multi-parton PDFs on the "long-distance" side. It is also interesting to note that current multiple-interactions models effectively amount to a resummation of scattering cross sections, in much the same way as parton showers represent a resummation of emission cross sections. However, whereas a wealth of higher-order analytical results exist for emission-type corrections, which can be used as useful cross-checks and tuning benchmarks for parton showers, corresponding results for multiple-interactions corrections are almost entirely absent. This is intimately linked to the absence of a satisfactory formulation of factorisation.

On the experimental side, it should be emphasised that there is much more than Monte Carlo tuning to be done in MB/UE studies, and that data is vital to guide us in both the phenomenological and formal directions discussed above. Dedicated Tevatron studies have already had a large impact on our understanding of hadron collisions, but much remains uncertain. Results of future measurements are likely to keep challenging that understanding and could provide for a very fruitful interplay between experiment and theory.

7 PARTON DISTRIBUTIONS FOR LO GENERATORS [12]

7.1 Introduction

It has long been known that for certain regions of x there can be large differences between PDFs extracted at different orders of perturbative QCD. It happens due to missing higher order corrections both in the parton evolution and in the MEs, which govern their extraction by comparison to experimental data. In particular, use of PDFs of the wrong order can lead to wrong conclusions for the small-x gluon. Traditionally, LO PDFs are usually thought to be the best choice for use with LO ME, usually available in Monte-Carlo programs, though it has been recognised that all such results should be treated with care. However, recently another viewpoint has appeared, namely it has been suggested that NLO PDFs may be more appropriate [112]. The argument is that NLO corrections to MEs are often small, and the main change in the total cross-section in going from LO to NLO is due to the PDFs.

In this paper we present another approach, which is based on advantages of both the LO and NLO PDF approximations, and compare all three predications for several processes with the *truth* – NLO PDFs combined with NLO MEs[13]. We interpret the features of the results noting that there are significant faults if one uses exclusively either LO or NLO PDFs. We hence attempt to minimise this problem, and investigate how a best set of PDFs for use with LO matrix elements may be obtained.

7.2 Parton Distributions at Different Orders

Let us briefly explain the reasons for the origins of the differences between the PDFs at different perturbative orders. The LO gluon is much larger at small x than any NLO gluon at low Q^2. The evolution of the gluon at LO and NLO is quite similar, so at larger Q^2 the relative difference is smaller, but always remains significant. This difference in the gluon PDF is a consequence of quark evolution, rather than gluon evolution. The small-x gluon is determined by $dF_2/d\ln Q^2$, which is directly related to the Q^2 evolution of the quark distributions. The quark-gluon splitting function P_{qg} is finite at small x at LO, but develops a small-x divergence at NLO (and further $\ln(1/x)$ enhancements at higher orders), so the small x gluon needs to be much bigger at LO in order to fit structure function evolution. There are also significant differences between the LO and NLO quark distributions. Most particularly the quark coefficient functions for structure functions in $\overline{MS}$ scheme have $\ln(1-x)$ enhancements at higher perturbative order, and the high-x quarks are smaller as the order increases. Hence, the LO gluon is much bigger at small x, and the LO valence quarks are much bigger at high-x. This is then accompanied by a significant depletion of the quark distribution for $x \sim 0.01$, despite the fact this leads to a poor fit to data.

Let us examine these differences using concrete examples. In the right of Fig. 31 we show the ratio of rapidity distributions for W -boson production at the LHC for several combination of PDF and ME to *the truth*. In this case the quark distributions are probed. Clearly we are generally nearer to the *truth* with the LO ME and NLO PDF [113] than with the LO ME and LO PDF [76]. However, this is always too small, since the NLO correction to the ME is large and positive. The depletion of the LO quark distributions for $x \sim 0.006$ (corresponding to the central y) leads to the extra suppression in the PDF[LO]-ME[LO] calculation. However, when probing the high x quarks the increase in the LO parton compensates for the increase in NLO matrix element, and for $y > 2$ this gives the more accurate result. However, overall the shape as a function of y is much worse using the LO parton distributions than the NLO distributions. The general conclusion is the NLO PDFs provide a better normalization and a better shape.

This example suggests that the opinion in [112] is correct. However, let us consider a counter-example, the production of charm in DIS, i.e. $F_2^{c\bar{c}}(x,Q^2)$. In this case the NLO coefficient function, $C_{2,g}^{c\bar{c},(2)}(x,Q^2,m_c^2)$ has a divergence at small x not presented at LO, in the same way that the quark-gluon

[12] Contributed by: A. Sherstnev, R.S. Thorne

[13] Since NLO matrix elements are most readily available in $\overline{MS}$ scheme, we will take this as the default, and henceforth NLO is intended to mean NLO in $\overline{MS}$ scheme.

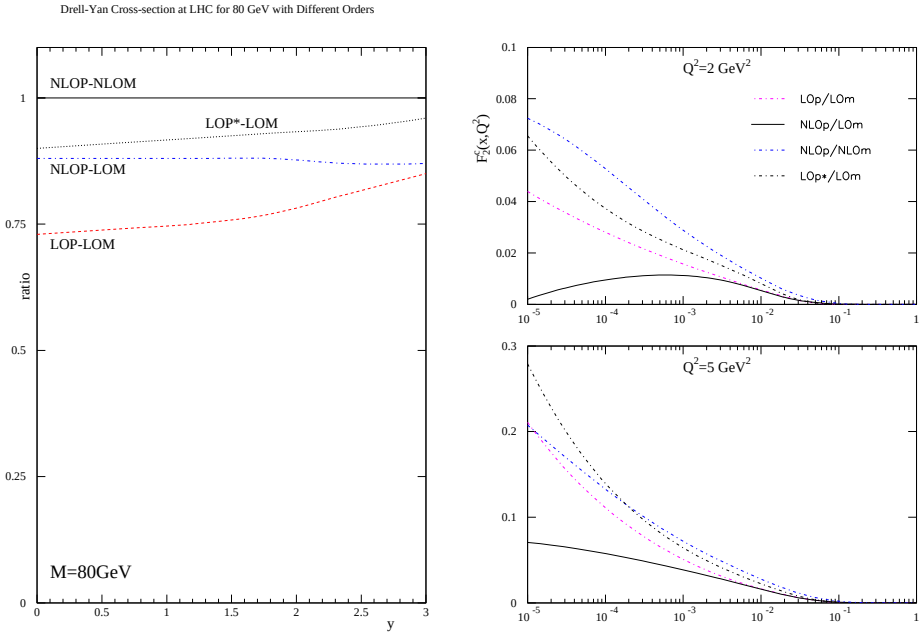

Fig. 31: Comparison of boson production at the LHC and charm production at HERA using combinations of different orders of ME and PDF.

splitting function does, the latter being responsible for the large difference between the LO and NLO gluons at small x. In the right of Fig.31 we see the large effect of the NLO coefficient functions. When using NLO partons the LO ME result is well below the *truth* at low scales. In this case the distribution is suppressed due to a lack of the divergence in both the NLO gluon and the LO coefficient function. While the LO PDFs combined with LO coefficient functions is not a perfect match to the *truth*, after all the small-x divergences are not exactly the same in matrix element and splitting function, it is better. In particular, in this case the NLO PDFs together with the LO matrix elements fail badly.

Hence, from these two simple examples alone we can conclude that both the NLO partons and the LO partons can give incorrect results in some processes. Let us try to find some *optimal* set of PDFs for use with LO matrix elements. Due to missing terms in $\ln(1 \quad x)$ and $\ln(1=x)$ in coefficient functions and/or evolution the LO gluon is much bigger as $x \; ! \; 0$ and valence quarks are much larger as $x \; ! \; 1$. From the momentum sum rule there are then not enough partons to go around, hence the depletion in the quark distributions at moderate to small x. This depletion leads to a bad global fit at LO, particularly for HERA structure function data, which is very sensitive to quark distributions at moderate x. In practice the lack of partons at LO is partially compensated by a LO extraction of much larger $_S(M_Z^2) \quad 0{:}130$. So, the first obvious modification is to use $_S$ at NLO in a LO fit to parton distributions. Indeed the NLO coupling with $_S(M_Z^2) = 0{:}120$ does a better job of fitting the low-Q^2 structure function data.

However, even with this modification the LO fit is still poor compared with NLO. The problems caused due to the depletion of partons has led to a suggestion by T. Sjöstrand[14] that relaxing the momentum sum rule for the input parton distributions could make LO partons rather more like NLO partons where they are normally too small, while allowing the resulting partons still to be bigger than NLO where necessary, i.e the small-x gluon and high-x quarks. Relaxing the momentum sum rule at input and using the NLO definition of the strong coupling does improve the quality of the LO global fit. The $^2 = 3066=2235$ for the standard LO fit, and becomes $^2 = 2691=2235$ for the modified fit with the same data set as in [113] and using $_S(M_Z^2) = 0{:}120$ at NLO. The momentum carried by input partons goes up to 113% . We denote the partons resulting from this fit as the LO* parton distribution functions.

We can make a simple test of the potential of these LO* partons by repeating the previous com-

[14] private comments at ATLAS Generators meeting, CERN, December 2006.

PDF	ME	$(pp \to Z/\gamma^*)$	K	$(pp \to tq)$	K	$(pp \to b\bar{b})$	K	$(pp \to t\bar{t})$	K
NLO	NLO	2.40 pb		259.4 pb		2.76 b		812.8 pb	
LO	LO	1.85 pb	1.30	238.1 pb	1.09	1.85 b	1.49	561.4 pb	1.45
NLO	LO	1.98 pb	1.26	270.0 pb	0.96	1.56 b	1.77	531.0 pb	1.53
LO*	LO	2.19 pb	1.09	297.5 pb	0.87	2.63 b	1.05	699.4 pb	1.16

Table 5: The total cross sections for $pp \to tq$, $pp \to b\bar{b}$, $pp \to t\bar{t}$, and $(pp \to Z/\gamma^* \to)$ at the LHC. Applied cuts: for $b\bar{b}$ ($p_T > 20$ GeV, $|\eta(b)| < 5.0$, $R(b,b) > 0.5$); for Z/γ^* ($p_T(\ ) > 10$ GeV, $|\eta| < 5.0$); no cuts for $t\bar{t}$ and single t. K-factor is defined according to $K = \sigma_{NLO}/\sigma_{LO}$.

parisons. For the W-boson production we are indeed nearer to the *truth* with the LO ME and LO* PDF than with either LO or NLO PDF. Moreover, the shape using the LO* PDF is of similar quality to that using the NLO partons with the LO ME. So in this case LO* PDF and NLO PDF are comparably successful. The exercise is also repeated for the charm structure function at HERA. When using the LO coefficient function the LO* PDF result is indeed nearest to the *truth* at low scales, being generally a slight improvement on the result using LO PDF, and clearly much better than that using NLO PDF.

These simple examples suggest that the LO* PDFs may well be a useful tool for use with Monte Carlo generators at LO, combining much of the advantage of using the NLO PDF while avoiding the major pitfalls. However, the examples so far are rather unsophisticated. In order to determine the best set of PDFs to use it is necessary to work a little harder. We need to examine a wide variety of contributing parton distributions, both in type of distribution and range of x. Also, the above examples are both fully inclusive, they have not taken into account cuts on the data. Nor have they taken account of any of the possible effects of parton showering, which is one of the most important features of Monte Carlo generators. Hence, before drawing any conclusions we will make a wide variety of comparisons for different processes at the LHC, using Monte Carlo generators to produce the details of the final state.

7.3 More examples at the LHC.

We consider a variety of final states for pp collisions at LHC energies. In each case we compare the total with LO MEs and full parton showering for the three cases of LO, LO* and NLO parton distributions. As the *truth* we use the results obtained with MC@NLO [23], which combines NLO QCD corrections and parton showers. As the main LO generator we use CompHEP [114], interfaced to HERWIG [98], but $pp \to b\bar{b}$ was calculated by HERWIG only.

The first example is the production of Z/γ^* bosons, decaying to muons. In order to exclude the dangerous region $m \to 0$, where the ME at LO has a singularity, we apply some experimentally reasonable cuts cuts $p_T > 10$ GeV and $|\eta| < 5.0$. These cuts are more or less appropriate for most analyses in CMS/ATLAS. The process is dominated by the Z peak. The mechanism is rather similar to that for W production, but now the initial quarks are the same flavour and the x at zero rapidity is slightly higher, i.e. $x_0 = 0.0065$. The similarity is confirmed in the results. Again all the total cross-sections using the LO generators are lower than the *truth*, as seen in Table 5, but that using the LO* partons is easily closest. The distributions in terms of the final state boson or the highest-p_T muon are shown in the upper and bottom plots of Fig. 32 respectively. For the boson the LO* partons gives comparable, perhaps marginally better, quality of shapes as the NLO partons, but better normalization. The LO partons have the worst suppression at central rapidity, and all partons give an underestimate of the high-p_T tail. For the muon the LO* partons give an excellent result for the rapidity distribution until $|\eta| > 4$, better in shape and normalization that the NLO partons whilst the LO partons struggle at central . Again, as in W production, the p_T distribution of the muon is better than for the boson, and in normalization is best described by the LO* PDFs.

Now we consider a somewhat different process, i.e. the single top production in the t-channel. At the partonic level the dominant process is $qb(\bar{q}\bar{b}) \to qt(\bar{q}t)$, where the b-quark has been emitted from gluon. Since the b-quark PDF is calculated based on gluon PDFs, this cross-section probes both

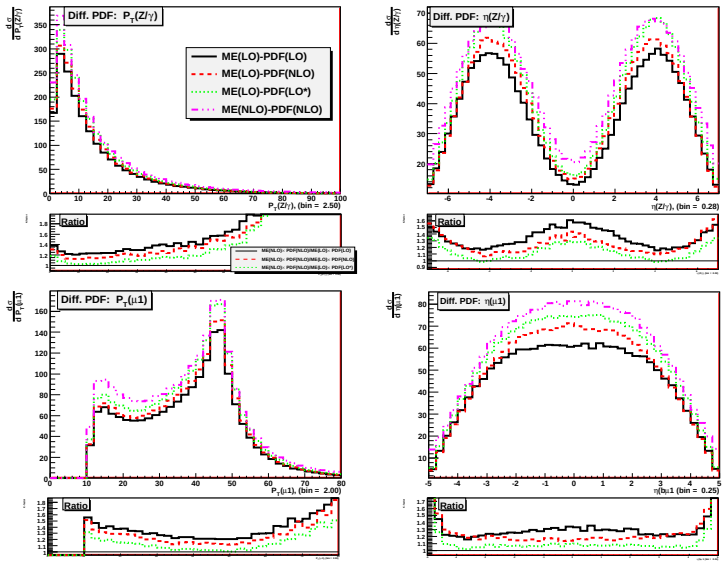

Fig. 32: The comparison between the competing predictions for the differential cross-section for Z = -boson production at the LHC (upper plots) and for the resulting highest p_t muon (bottom plots).

the gluon distribution and the quark distributions for invariant masses of above about 200 GeV, i.e. at central rapidity x_0 0:05. The t-channel nature of this process makes the invariant mass of the final state and the probed x values less precise than the the W boson production. The total cross-section for the various methods of calculation are seen in Table 5. In this case the result using the LO ME and the LO PDFs is suppressed, but that using the LO* PDFs is now larger than the *truth*. This is due to the large enhancement of the LO* gluon distribution. The NLO PDFs give the closest normalization. The distributions in terms of p_T and of the final state top and originated from the top are shown in the left of Fig. 33. For the top distribution the result using the LO generator and the LO* and NLO PDFs give a very similar result, being better than the LO PDF result both for normalization and for shape due to the suppression of the LO quarks at central rapidities. In the case of the (from the top) the distributions calculated with the LO generator look better then for the top, since the real NLO correction (irradiation if an extra parton) plays lesser role for the top decay products. In this process there is a particular NLO enhancement at central rapidity, so it gives a total cross section larger than *the truth*.

We now consider the bb production at the LHC. At LO the process consists of three contributions: gg=qq ! bb (Flavour Creation, or FCR), qb ! qb, where the second b-quark is simulated by initial parton showers (Flavour Excitation, or FEX), and the QCD 2 ! 2 process with massless partons, where the b-quarks arise from parton showers[15] (Gluon Splitting, or GSP). The 2nd and 3rd subprocesses have massless partons and, thus, soft and collinear singularities. In order to exclude the dangerous regions, we apply some reasonable cuts: p_T (b) > 20 GeV, j (b)j < 5:0, R (b; b) > 0:5. At NLO we can not separate the subprocesses, so only the FCR process exists at NLO [115]. In bb we probe rather low x 10^{-3} 10^{-2} and the gluon-gluon initial state, so the process is sensitive to the small-x divergence in the NLO MEs, and the NLO correction is very large. The total cross-sections are shown in Table 5. All the LO calculations are below the *truth*, but the reduced NLO gluon means that the NLO PDF gives by far the worst result. The best absolute prediction is obtained using the LO* partons. The differential distributions in terms of p_T and of a single b quark are shown on the upper plots and for the pseudo-rapidity and p_T of a bb pair on the bottom plots in right of Fig. 33. The LO* PDFs do well for the single b rapidity distribution, but underestimate a little at high rapidity. The LO and NLO PDFs are similar

[15]For example, the total cross-section for the improved LO PDFs from Table 5 has three terms: $_{tot} = _{FCR} + _{FEX} + _{GSL}$, where $_{FCR} = 1:6$ b, $_{FEX} = 0:57$ b, and $_{GSP} = 0:46$ b – the total cross sections for the FCR, FEX, and GSP processes respectively.

in shape, but the normalisation is worse for NLO and it fails particularly at low p_T, i.e small x. All PDFs obtain roughly the right shape for the $(b\bar{b})$, except small underestimation at very high rapidity. However, for all partons there is a problem with the shape as a function of p_T. Obviously, all the ratio curves become higher as p_T goes up. As for other processes this happens due to the different behaviour of the additional parton generated in the NLO matrix element compared to those generated by parton showers. In general, we conclude the LO* PDFs give the best results in the comparison.

Another interesting heavy quark production process is the double top quark production. The total cross sections are reported in Table 5. At the LHC this process is dominated by the gluon contribution $gg \to t\bar{t}$. For example, $_{ME[LO]\,PDF[LO]} = {}_{gg\to t\bar{t}} + {}_{q\bar{q}\to t\bar{t}} = 486{:}9\ pb + 74{:}5\ pb$. The LO* PDFs appreciably enlarge the gluonic cross section, namely, $_{ME[LO]\,PDF[LO^*]} = {}_{gg\to t\bar{t}} + {}_{q\bar{q}\to t\bar{t}} = 622{:}1\ pb + 77{:}3\ pb$. Again the LO* PDFs gives the best prediction.

7.4 Conclusions

We have examined the effects of varying both the order of the MEs and the PDFs when calculating cross-section for hadron colliders. The intention is to find the best set of PDFs to use in current Monte Carlo generators. A fixed prescription of either LO or NLO PDFs with LO matrix elements is unsuccessful, with each significantly wrong in some cases. For LO PDFs this is mainly due to the depletion of quarks for $x \quad 0{:}1 \quad 0{:}001$ and the large LO gluon above $x \quad 0{:}01$, while for NLO partons the smallness in some regions compared to LO PDFs is a major problem if the large NLO matrix element is absent. To this end we have suggested an optimal set of partons for Monte Carlos, which is essentially LO but with modifications to make results more NLO-like, and are called LO* PDFs. The NLO coupling is used, which is larger at low scales, and helps give a good fit to the data used when extracting partons from a global fit. The momentum sum rule is also relaxed for the input parton distributions. This allows LO PDFs to be large where it is required for them to compensate for missing higher order corrections, but not correspondingly depleted elsewhere.

We have compared the LO, NLO and LO* PDFs in LO calculations to the *truth*, i.e. full NLO, for a wide variety of processes which probe different types of PDF, ranges of x and QCD scales (more examples are available in [116]). In general, the results are very positive. The LO* PDFs nearly always provide the best description compared to the *truth*, especially for the s-channel processes. This is particularly the case in terms of the normalization, but the shape is usually at least as good, and sometimes much better, than when using NLO PDFs. It should be stressed that no modification of the PDFs can hope to successfully reproduce all the features of genuine NLO corrections. In particular we noticed the repeating feature that the high-p_T distributions are underestimated using the LO generators, and this can only be corrected by the inclusion of the emission of a relatively hard additional parton which occurs in the NLO matrix element correction. A preliminary version of the LO* PDFs, based on fitting the same data as in [113], is available on request. A more up-to-date version, based on a fit to all recent data, and with uncertainty bands for the PDFs, will be provided in the MSTW08 PDF set.

ACKNOWLEDGEMENTS

We would like to thank J. Butterworth, S. Moch, C. Gwenlan, P. Bartalini, M. Cooper-Sarkar, J. Huston, A. Martin, S. Mrenna, T. Sjöstrand, J. Stirling, G. Watt and B. Webber for helpful discussions. RST would like to thank the Royal Society for the award of a University Research Fellowship. AS would like to thank the Science and Technology Facilities Council for the award of a Responsive RA position.

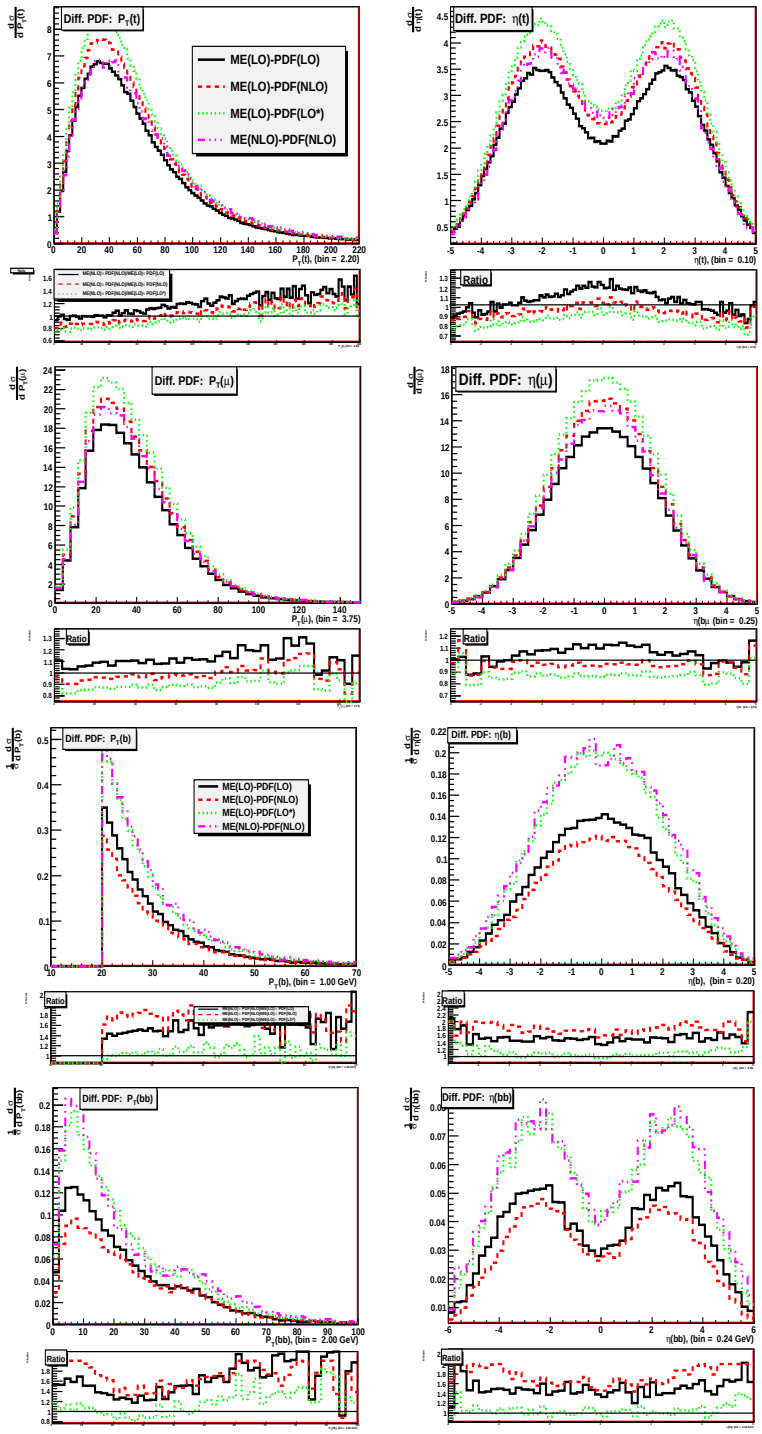

Fig. 33: The comparison between the competing predictions for the differential cross-section for single top production at the LHC (left upper plots) and for the resulting p_t muon (left bottom plots). Differential cross-sections for $b$production at the LHC (right upper plots) and for a $b\bar{b}$pair (right bottom plots).

Part II
ISSUES IN JET PHYSICS

8 JET PHYSICS INTRODUCTION[16]

This introductory section is intended to help provide the reader with some background to the current jet-related panorama at the LHC, in particular as concerns the basic principles and properties of the main jet algorithms currently in use within the Tevatron and LHC experiments and in phenomenological and theoretical discussions. Part of what is described here formed the basis of discussions during the course of the workshop and subsequent work, but for completeness additional material is also included.

Several other jet-related sections are present in these proceedings. Section 9 outlines two proposals for accords reached during the workshop, one concerning general nomenclature for jet finding, the other about the definition of the hadronic final-state that should be adopted when quoting experimental measurements. Section 10 examines how to measure the performance of jet algorithms at hadron level and determine optimal choices in two physics cases, a fictional narrow Z^0 over a range of Z^0 masses, and in top production, providing examples of simple and complex quark-jet samples. Section 11 examines the performance of jet algorithms at hadron level in inclusive jet and Z+jet production, and in $H \to gg$ decays for a range of Higgs masses, which provides examples of gluon-jet samples. Section 12 instead examines the performance of jet algorithms at detector level, using calibrated calorimetric clusters as input four-vectors, also examining the influence on jet reconstruction of the presence of a moderate pileup, as expected in the first years of LHC running. Other jet-related work that was discussed in part during the workshop, but was not the focus of workshop-specific investigation includes studies of non-perturbative effects in jets [117] and the use of jet substructure in the discovery of new particles [118], as well as methods for dealing with the problem of soft contamination of jets in the presence of pileup or in heavy-ion collisions [119–122]. We note also related discussion of jet-finding in the context of the Tev4LHC workshop [84], as well as the recent review [123]. For a review of jet algorithms for ep and e^+e colliders, see [124].

8.1 Jet algorithms

As per the accord in section 9.1, by *jet algorithm* we refer to a generic "recipe" for taking a set of particles (or other objects with four-vector like properties) and obtaining jets. That recipe will usually involve a set of parameters (a common example being the jet-radius R). The recipe plus specific values for the parameters provides a fully specified *jet definition*.

Many hadron-collider jet algorithms are currently being discussed and used in the literature. This section provides an overview of the basic principles underlying the jet algorithms for which we are aware of experimental or theoretical use in the past couple of years. There are two broad groups of jet algorithms, those based in one form or another on cones and those that involve repeated recombination of particles that are nearby in some distance measure. The nomenclature used to distinguish the flavours of jet algorithm is currently not always uniform across the field — that used here follows the lines set out in [125].

8.1.1 Cone algorithms

There are many different cone algorithms in use. Most are "iterative cones" (IC). In such algorithms, a seed particle i sets some initial direction, and one sums the momenta of all particles j within a cone of

[16] Convenors: G.P. Salam and M. Wobisch; Contributing authors: V. Adler, A. A. Bhatti, J. M. Butterworth, V. Büge, M. Cacciari, M. Campanelli, D. D'Enterria, J. D'Hondt, J. Huston, D. Kcira, P. Loch, K. Rabbertz, J. Rojo Chacon, L. Sonnenschein, G. Soyez, M. Tytgat, P. Van Mulders, M. Vazquez Acosta, I. Villella

radius R around i in azimuthal angle ϕ and rapidity y (or pseudorapidity η), i.e. taking all j such that

$$\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 < R^2; \tag{26}$$

where y_i and ϕ_i are respectively the rapidity and azimuth of particle i. The direction of the resulting sum is then used as a new seed direction, and one iterates the procedure until the direction of the resulting cone is stable.

Such a procedure, if applied to an ensemble of many particles can lead to multiple stable cones that have particles in common (overlapping cones). Cone algorithms fall into two groups, depending on how they resolve this issue.

One approach is to start iterating from the particle (or calorimeter tower) with the largest transverse momentum. Once one has found the corresponding stable cone, one calls it a jet and removes from the event all particles contained in that jet. One then takes as a new seed the hardest particle/tower among those that remain, and uses that to find the next jet, repeating the procedure until no particles are left (above some optional threshold). A possible name for such algorithms is iterative cones with progressive removal (IC-PR) of particles. Their use of the hardest particle in an event gives them the drawback that they are collinear unsafe: the splitting of the hardest particle (say p_1) into a nearly collinear pair (p_{1a}, p_{1b}) can have the consequence that another, less hard particle, p_2 with $p_{t,1a}, p_{t,1b} < p_{t,2} < p_{t,1}$, pointing in a different direction suddenly becomes the hardest particle in the event, thus leading to a different final set of jets.

A widespread, simpler variant of IC-PR cone algorithms is one that does not iterate the cone direction, but rather identifies a fixed cone (FC) around the seed direction and calls that a jet, starting from the hardest seed and progressively removing particles as the jets are identified (thus FC-PR). It suffers from the same collinear unsafety issue as the IC-PR algorithms. Note that IC-PR and FC-PR algorithms are sometimes referred to as UA1-type cone algorithms, though the algorithm described in the original UA1 reference [126] is somewhat different.

Another approach to the issue of the same particle appearing in many cones applies if one chooses, as a first stage, to find the stable cones obtained by iterating from all particles or towers (or those for example above some threshold 1 2GeV).[17] One may then run a split–merge (SM) procedure, which merges a pair of cones if more than a fraction f of the softer cone's transverse momentum is in common with the harder cone; otherwise the shared particles are assigned to the cone to which they are closer.[18] A possible generic name for such algorithms is IC-SM. An alternative is to have a "split-drop" (SD) procedure where the non-shared particles that belong to the softer of two overlapping cones are simply dropped, i.e. are left out of jets altogether. The exact behaviour of SM and SD procedures depend on the precise ordering of split and merge steps and a now standard procedure is described in detail in [127] with the resolution of some small ambiguities given in [128].

IC-SM type algorithms have the drawback that the addition of an extra soft particle, acting as a new seed, can cause the iterative process to find a new stable cone. Once passed through the split–merge step this can lead to the modification of the final jets, thus making the algorithm infrared unsafe. A solution, widely used at Run II of the Tevatron, as recommended in [127], was to additionally search for new stable cones by iterating from midpoints between each pair of stable cones found in the initial seeded iterations (IC$_{mp}$-SM). While this reduces the set of configurations for which a soft particle modifies the final jets, it does not eliminate the problem entirely. One full solution instead avoids the use of seeds and iterations, and finds *all* stable cones through some exact procedure. This type of algorithm is often called a seedless cone (SC, thus SC-SM with a split–merge procedure). Historically, the computational complexity of seedless-cone algorithms had made their use impractical for use on events with realistic numbers of particles, however, recently a geometrically-based solution was found to this problem [128].

[17] In one variant, "ratcheting" is included, which means that during iteration of a cone, all particles included in previous iterations are retained even if they are no longer within the geometrical cone.

[18] Commonly used values for the overlap threshold parameter are f = 0.5, 0.75 (see also recommendations below).

Algorithm	Type	IRC status	Ref.	Notes
inclusive k_t	$SR_{p=1}$	OK	[130–132]	also has exclusive variant
flavour k_t	$SR_{p=1}$	OK	[133]	d_{ij} and d_{iB} modified when i or j is "flavoured"
Cambridge/Aachen	$SR_{p=0}$	OK	[134, 135]	
anti-k_t	$SR_{p=-1}$	OK	[125]	
SISCone	SC-SM	OK	[128]	multipass, with optional cut on stable cone p_t
CDF JetClu	IC_r-SM	IR_{2+1}	[136]	
CDF MidPoint cone	IC_{mp}-SM	IR_{3+1}	[127]	
CDF MidPoint searchcone	$IC_{se,mp}$-SM	IR_{2+1}	[129]	
D0 Run II cone	IC_{mp}-SM	IR_{3+1}	[127]	no seed threshold, but cut on cone p_t
ATLAS Cone	IC-SM	IR_{2+1}		
PxCone	IC_{mp}-SD	IR_{3+1}		no seed threshold, but cut on cone p_t,
CMS Iterative Cone	IC-PR	$Coll_{3+1}$	[137, 138]	
PyCell/CellJet (from Pythia)	FC-PR	$Coll_{3+1}$	[85]	
GetJet (from ISAJET)	FC-PR	$Coll_{3+1}$		

Table 6: Overview of some jet algorithms used in experimental or theoretical work in hadronic collisions in the past couple of years. $SR_{p=x}$ = sequential recombination (with $p = -1; 0; 1$ characterising the exponent of the transverse momentum scale, eq. (27)); SC = seedless cone (finds all cones); IC = iterative cone (with midpoints mp, ratcheting r, searchcone se), using either split–merge (SM), split–drop (SD) or progressive removal (PR) in order to address issues with overlapping stable cones; FC = fixed-cone. In the characterisation of infrared and collinear (IRC) safety properties (for the algorithm as applied to particles), IR_{n+1} indicates that given n hard particles in a common neighbourhood, the addition of 1 extra soft particle can modify the number of final hard jets; $Coll_{n+1}$ indicates that given n hard particles in a common neighbourhood, the collinear splitting of one of the particles can modify the number of final hard jets. Where an algorithm is labelled with the name of an experiment, this does not imply that it is the only or favoured one of the above algorithms used within that experiment. Note that certain computer codes for jet-finding first project particles onto modelled calorimeters.

Cone algorithms with split–merge or split–drop steps are subject to a phenomenon of "dark towers" [129], regions of hard energy flow that are not clustered into any jet. A solution to this proposed in [129] — referred to as the "searchcone" — works around the problem by using a smaller radius to find stable cones and then expands the cones to their full radius without further iteration before passing them to the SM procedure. It was subsequently discovered that this reintroduces IR safety issues [84], and an alternative solution is a multi-pass algorithm, one that runs the cone algorithm again on the set of all particles that do not make it into any of the "first-pass" jets (this can be repeated over and over until no particles are left unclustered).

8.1.2 $2 \to 1$ Sequential recombination

Sequential recombination (SR) algorithms introduce distances d_{ij} between entities (particles, pseudojets) i and j and d_{iB} between entity i and the beam (B). The (inclusive) clustering proceeds by identifying the smallest of the distances and if it is a d_{ij} recombining entities i and j, while if it is d_{iB} calling i a jet and removing it from the list of entities. The distances are recalculated and the procedure repeated until no entities are left.

The distance measures for several algorithms are of the form

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2}\,; \tag{27a}$$

$$d_{iB} = k_{ti}^{2p}\,; \tag{27b}$$

where Δ_{ij}^2 was defined in (26) and k_{ti} is the transverse momentum of particle i. Here R is the jet-radius parameter, while p parametrises the type of algorithm. For $p = 1$ one has the inclusive k_t algorithm as defined in [132], while with $p = 0$ one obtains the Cambridge/Aachen algorithm as defined in [135]. Both are related to corresponding "exclusive" algorithms (k_t [130,131], Cambridge [134], and also [139]) with similar or identical distance measures but additional stopping conditions. A recent addition to the SR class is the anti-k_t algorithm, with $p = -1$ [125]. Together with the PR cones, it has the property that soft radiation does not affect the boundary of the jet, leading to a high proportion of circular jets with actual radius R. This property does not hold for SM and SD cones, nor SR algorithms with $p \neq 0, \pm 1$.

Other sequential recombination algorithms, used mainly in e^+e^- and DIS collisions, include the JADE algorithm [140, 141] which simply has a different distance measure, and the ARCLUS algorithm [142] which performs $3 \to 2$ recombinations (the inverse of a dipole shower).

8.1.3 General remarks

A list of algorithms used in experimental or theoretical studies in the past couple of years is given in table 6. Where possible references are provided, but some algorithms have not been the subject of specific publications, while for others the description in the literature may only be partial. Thus in some cases, to obtain the full definition of the algorithm it may be advisable to consult the corresponding computer code.

A point to be noted is that as well as differing in the underlying recipe for choosing which particles to combine, jet algorithms can also differ in the scheme used to recombine particles, for example direct 4-momentum addition (known as the E-scheme), or E_T weighted averaging of and . In the past decade recommendations have converged on the E-scheme (see especially the Tevatron Run-II workshop recommendations [127]), though this is not used by default in all algorithms of table 6.

As discussed in section 8.1.1 many of the algorithms currently in used are either infrared or collinear unsafe. For an algorithm labeled IR_{n+1} or Coll_{n+1}, jet observables that are non-zero starting with m partons in the final state (or $m-1$ partons and one W/Z boson) will be divergent in perturbation theory starting from $\text{N}^{n-m+2}\text{LO}$. Given that these are usually single-logarithmic divergences, the physics impact is that N^{n-m}LO is then the last order that can be reliably calculated in perturbation theory (as discussed for example in detail in [128]).

Because of the perturbative divergences and other non-perturbative issues that arise with non infrared and collinear safe algorithms, there have been repeated recommendations and accords, dating back to the Snowmass accord [143], to use just infrared and collinear safe jet algorithms. This recommendation takes on particular importance at the LHC, because multi-jet configurations, which will be far more widespread than at previous colliders, are particularly sensitive to infrared and collinear safety issues. Furthermore there is very significant investment by the theoretical community in multi-leg NLO computations (see for example the proceedings of the NLO Multi-leg working group of this workshop), and the benefit to be had from such calculations will largely be squandered if infrared or collinear unsafe jet algorithms are used for analyses. The set of IRC-safe algorithms that have been the subject of some degree of recent study includes k_t, Cambridge/Aachen, SISCone (which can be used as a replacement for IC-SM type algorithms) and anti-k_t (which is a candidate for replacing IC-PR type algorithms).

8.1.4 Jet algorithm packages

Given the many jet algorithms that are in use, and the interest in being able to easily compare them, two packages have emerged that provide uniform access to multiple jet algorithms. `FastJet` [144, 145], originally written to implement fast strategies for sequential recombination, also has a "plugin" mechanism to wrap external algorithms and it provides a number of cone algorithms in this manner, including SISCone [128]. `SpartyJet` [146] provides a wrapper to the `FastJet` algorithm implementations (and through it to SISCone) as well as to a number of cone algorithms, together with specific interfaces for the ATLAS and CDF environments. Both packages are under active development and include various features beyond what is described here, and so for up to date details of what they contain, readers are referred to the corresponding web pages.

8.2 Validation of jet-finding

During the Les Houches workshop, a validation protocol was defined in order to ensure that all participants were using identical jet algorithms and in the same way. For this purpose, a sample of 1000 events was simulated with Pythia 6.4 [85], for the production and subsequent hadronic decay of a Z^0, $Z^0 \to q\bar{q}$ with $M_{Z^0} = 1000$ GeV. This was run through the different participants' jet software for each of the relevant jet definitions, and it was checked that they obtained identical sets of jets.[19]

The following jet algorithms were used in the jet validation

- k_t
- Cambridge/Aachen
- Anti-k_t (added subsequent to the workshop)
- SISCone
- CDF Midpoint cone

For each, one uses values of R from $R_{min} = 0.3$ to $R_{max} = 1.0$ in steps of $R = 0.1$. In the two SM-type cone algorithms, the SM overlap threshold f was set to 0.75. This choice is recommended more generally because smaller values (including the quite common $f = 0.50$) have been seen to lead to successive merging of cones, leading to "monster-jets" (see e.g. [147]).

Readers who wish to carry out the validation themselves may obtain the event sample and further details from

`http://www.lpthe.jussieu.fr/~salam/les-houches-07/validation.php`

together with reference results files and related tools.

[19] This statement holds for comparisons carried out with double-precision inputs; where, for data-storage efficiency reasons, inputs were converted to single precision, slight differences occasionally arose.

9 ACCORDS RELATED TO THE HADRONIC FINAL STATE[20]

9.1 Jet nomenclature

In this section we aim to establish a common and non-ambiguous nomenclature to be used when discussing jet physics. Such a basis is needed for the communication of experimental results, in order to ensure that they can be reproduced exactly, or that matching theory predictions can be made. We propose that the following elements should always be specified in experimental publications:

The jet definition which specifies all details of the procedure by which an arbitrary set of four-momenta from physical objects is mapped into a set of jets. The jet definition is composed of a **jet algorithm** (e.g. the inclusive longitudinally boost-invariant k_T algorithm), together with *all* its **parameters** (e.g. the jet-radius parameter `R`, the split–merge overlap threshold `f`, the seed-threshold p_T cut, etc.) and the **recombination scheme** (e.g. the four-vector recombination scheme or "E-scheme") according to which the four-momenta are recombined during the clustering procedure. We recommend that a reference to a *full* specification of the jet algorithm is given. If this is not available, the jet algorithm should be described in detail.

The final state ("truth-level") specification. Consistent comparisons between experimental results, or between experimental results and Monte Carlo simulations, are only possible if the jet definition is supplemented with an exact specification of the set of the physical objects to which it was applied, or to which a quoted jet measurement has been corrected. This could e.g. be the set of momenta of all hadrons with a lifetime above some threshold. Discussions and recommendations of possible final state choices are given below in section 9.2.

This nomenclature proposal is summarised graphically in Fig. 34.

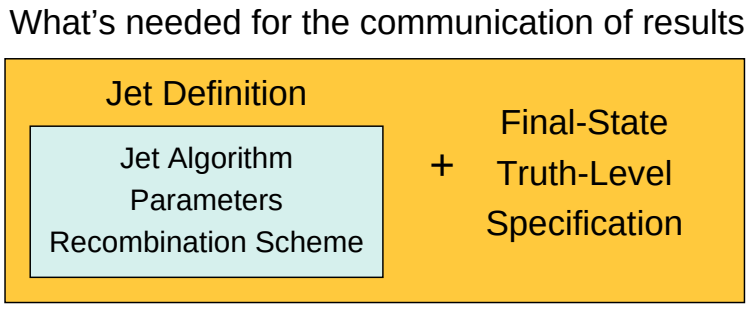

Fig. 34: A summary of the elements needed to communicate jet observables in a non-ambiguous way.

9.2 Final state truth level

Whenever experiments present "corrected" results for given jet observables, the question arises *"What exactly have these results been corrected for?"*, or in other words *"On which set of four-vectors are the quoted results of this jet measurement defined?"*. These questions address the "truth-level" to which experimental results correspond to. A detailed answer to this question is relevant since supposedly minor differences can be significant, and they certainly are for precision jet measurements[21]. In the history of jet physics at particle colliders, many different choices have been made on how jet results were presented. Experiments have corrected their jet results

back to the leading order matrix-elements in a Monte Carlo. The jets are supposed to correspond to the partons from the `2 ! 2` scattering process.

[20] Convenors: G.P. Salam and M. Wobisch; Contributing authors: V. Adler, A. Bhatti, J. M. Butterworth, V. Büge, M. Cacciari, D. D'Enterria, J. D'Hondt, J. Huston, D. Kcira, P. Loch, H. Nilsen, K. Rabbertz, J. Rojo-Chacon, L. Sonnenschein, G. Soyez, M. Tytgat, P. Van Mulders, M. Vazquez Acosta, I. Villella

[21] Note that the ambiguity addressed here does not include the jet definition, which is supposed to have already been agreed upon and fully specified.

back to the level after the parton shower in a Monte Carlo. The jets are supposed to correspond to the result of the purely perturbative phase of the hadronic reaction.

back to the level of stable particles in a Monte Carlo, but excluding the particles from the "underlying event".

for all detector effects and, in addition, also for the energies observed in interactions triggered by "minimum bias" triggers. The latter contribution is supposed to correspond to the "underlying event".

for all detector effects and nothing else. The corrected jet results correspond to jets defined on all (stable) particles from the hadronic interaction.

It would be useful for the LHC and the Tevatron experiments to have a common definition of what they call the "truth" final-state particle level (specifically for jets). While we cannot enforce any agreement, we can provide a set of recommendations, and make the following proposals:

The truth input to the jet clustering should always be physical (i.e. observable) final-state particles, not any kind of model-dependent partons (neither from a matrix-element nor from a parton-shower).

For similar reasons, the final-state particles should include everything from the main hadronic scatter. Therefore the underlying event (defined as additional partonic interactions from the same hadron-hadron interaction plus interactions of the hadron remnants) is included. This is part of the hadronic interaction and cannot be unambiguously separated from the hard subprocess (see, however, next subsection).

The contributions from pile-up due to additional hadronic collisions in the same bunch crossing, recorded in the same event, should not be included. In other words, the jet observable should be corrected for contributions from multiple hadron interactions.

A standard lifetime cut on what is considered to be "final state" should be agreed upon. A lifetime of 10 ps is used elsewhere, and we also recommend this value: only hadrons with a shorter lifetime will be allowed to decay in the Monte Carlo generators. All other particles will be considered to be stable.

Neutrinos, muons and electrons from hadronic decays should be included as part of the final state.

However, prompt muons, electrons (and radiated photons), neutrinos and photons are excluded from the definition of the final state. The same applies to the decay products of prompt taus.

The jet algorithm should be given as input the full physical four-vectors. How it treats them is part of the jet definition and the recombination scheme.

We acknowledge that these recommendations may not be useful in all circumstances. During the process of understanding and calibrating detectors, other definitions (e.g. including only visible energy in the calorimeter) may be needed. But whenever a jet measurement is presented or a jet observable is quoted, we suggest that the jets it refers to are based on a specific (and clearly stated) jet definition and the final-state truth particle definition recommended above.

9.3 A level behind the truth: Partons

It should be noted that the above definitions about the final state truth level also apply to theoretical calculations. Some theoretical calculations are implemented in Monte Carlo event generators, including the modelling of non-perturbative processes (hadronization and underlying event). These can directly be compared to experimental results that are obtained according to the recommendations from the previous section.

Other calculations provide purely perturbative results (typically at next-to-leading order in the strong coupling constant, sometimes accompanied by resummations of leading logarithms). These results correspond to the "parton level" of the jet observable. When trying to compare a perturbative

calculation to an experimental result, one needs to at least estimate the size of the non-perturbative corrections (consisting of underlying event and hadronization corrections). Typically, these are obtained using Monte Carlo event generators. We strongly recommend that each experiment should determine and publish its best estimate of non-perturbative corrections together with the data. It should be kept in mind that these corrections should always be quoted separately and not be applied to the data, but only to the perturbative calculations. Experiment and theory should meet at the level of an observable. This seems to be an established procedure, which is used in most jet analyses at LEP, HERA, and also in Run II of the Tevatron.

10 QUANTIFYING THE PERFORMANCE OF JET ALGORITHMS AT THE LHC [22]

10.1 General strategy

The performance of a given jet algorithm depends on its parameters, like the radius R, but it also depends on the specific process under consideration. For example, a jet algorithm that gives good results in a simple dijet environment might perform less well in a more complex multi-jet situation. In this contribution we wish to quantify the extent to which this is the case in the context of a couple of illustrative reconstruction tasks. This is intended to help cast light on the following question: should the LHC experiments devote the majority of their effort to calibrating as best as possible just one or two jet definitions? Or should they instead devote effort towards flexibility in their choice of jet definition, so as to be able to adapt it to each specific analysis?

One of the main issues addressed in examining this question is that of how, simply but generally, to quantify the relative performance of different jet algorithms. This physics analyses used as examples will be the reconstruction of massive particles, because such tasks are central both to Standard Model and to discovery physics at the LHC. As quality measures we shall use the mass resolution, and the signal size for fixed detector mass resolution, both defined in such a way as to be insensitive to the exact signal shape (which depends significantly on the jet definition). As test cases we will take a hypothetical Z^0 for different values of its mass, and the W boson and top quark in fully hadronic decays of $t\bar{t}$ events.

A point that we wish to emphasise is that we have purposefully avoided quality measures, used in the past, that consider the relation between jets and the hard partons produced at matrix-element level in a parton-shower Monte Carlo. This is because the relation between those two concepts depends as much on approximations used in the parton showering, as on the jet definition. Indeed in modern tools such as MC@NLO [23] or POWHEG [148] it becomes impossible, even programmatically, to identify the single parton to which one would want to relate the jet. Note however that addressing the issue of the performance of jet algorithms in contexts other than kinematic reconstructions (e.g. for the inclusive jet spectrum) would require rather different strategies than those we use here (see for example [117] and section 11). A strategy related to ours, to assess the performance of jet algorithms based on the Higgs mass reconstruction from the invariant mass of gluon jets in $H \to gg$ can be found in Sect. 11.

We note that we do not address issues of experimental relevance like the reconstruction efficiency of different jet algorithms after detector simulation, which however are discussed in the contribution of section 12.

10.2 Figures of merit

We start by defining the figures of merit that quantify the quality of the heavy object mass reconstruction through jet clustering algorithms.

We wish to avoid assumptions on the underlying shape of the invariant mass distribution that we are reconstructing, such as whether it is Gaussian, asymmetric or has a pedestal, since in general the reconstructed mass distributions cannot be described by simple functional forms. This is illustrated in Fig. 35, where different functions are fitted to two reconstructed mass spectra from the $Z^0 \to q\bar{q}$ samples for two different values of R. One sees that even in the more symmetric situation, it is difficult to reproduce it properly with any simple functional form.

Instead we shall use figures of merit that relate to the maximisation of the signal over background ratio (more precisely, $S/\sqrt{B}$), for the simplifying assumption that the background is flat and is not affected by the jet clustering procedure. Specifically, we propose the following two measures:

1. $Q^{w}_{f=z}(R)$: The width of the smallest (reconstructed) mass window that contains a fraction $f = z$

[22] Contributed by: M. Cacciari, J. Rojo-Chacon, G. P. Salam, G. Soyez

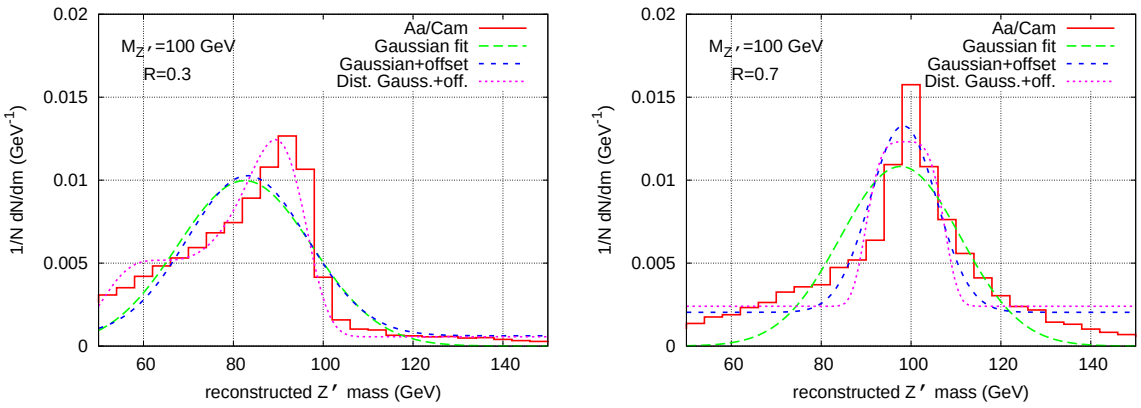

Fig. 35: The mass of the reconstructed Z^0 boson in the $M_{Z^0} = 100$ case with the Cambridge/Aachen algorithm for $R = 0.7$ (left) and $R = 0.3$ (right), together with various fits of simple probability distributions.

of the generated massive objects,[23] that is

$$f = \frac{\#\ \text{reconstructed massive objects in window of width } w}{\text{Total}\#\ \text{generated massive objects}} \quad : \tag{28}$$

A jet definition that is more effective in reconstructing the majority of massive objects within a narrow mass peak gives a lower value for $Q^w_{f=z}(R)$, and is therefore a "better" definition. The value that we will use for the fraction f will be adjusted in order to have around 25% of the reconstructed objects inside the window.[24]

2. $Q^f_{w=x\sqrt{M}}(R)$: To compute this quality measure, first we displace over the mass distribution a window of fixed width given by $w = x\sqrt{M}$, where M is the nominal heavy object mass that is being reconstructed[25] until we find the maximum number of events of the mass distribution contained in it. Then the figure of merit is given in terms of the ratio of this number of events with respect to the total number of generated events,

$$Q^f_{w=x\sqrt{M}}(R) \quad \frac{\text{Max}\#\ \text{reconstructed massive objects in window of width } w = x\sqrt{M}}{\text{Total}\#\ \text{generated massive objects}}^{-1} \quad ; \tag{29}$$

where we take the inverse so that the optimal result is a minimum of $Q^f_{w=x\sqrt{M}}(R)$, as in the previous case.

The default choice that will be used is $x = 1.25$, that is $w = 1.25\sqrt{M}$ (for compactness we omit the dimensions on x, which are to be understood as $(\text{GeV})^{1=2}$). This particular choice is motivated by experimental considerations of the CMS and ATLAS experiments, in particular the default value corresponds to the jet resolution in CMS. This means that the default values that will be used through this contribution will be $w = 1.25\sqrt{M_{Z^0}}$ for the Z^0 samples, $w = 1.25\sqrt{M_W}$ 10 GeV for the W mass distributions and $w = 1.25\sqrt{M_W}$ 15 GeV for the top quark mass distributions.

In tests of a range of possible quality measures for mass reconstructions (including Gaussian fits, and the width at half peak height), the above two choices have been found to be the least sensitive to

[23] Note that in general the number of generated massive objects differs from the total number of events, for example if in the $t\bar{t}$ samples we have $N_{ev} = 10^5$, the number of generated W bosons (and top quarks) is $N_W = 2 \cdot 10^5$.

[24] The approximate fraction of events that pass the event selection cuts for each physical process can be seen in Table 7, together with the value for the fraction z ensuring that approximately one quarter of the successfully reconstructed heavy objects are inside the window.

[25] Note that we avoid using the reconstructed mass M_{reco}, obtained from the mean of the distribution for example, since in general it depends strongly on the jet definition.

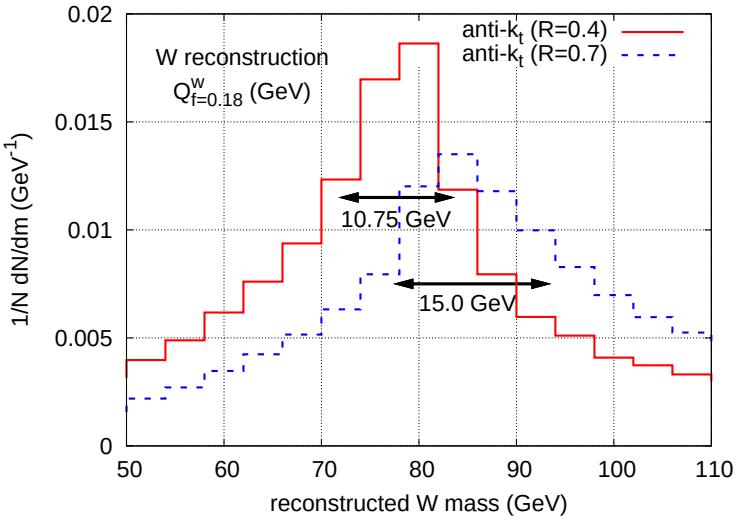

Fig. 36: The quality measure $Q^{W}_{f=0.18}(R)$ in the case of W mass reconstruction for hadronic $t\bar{t}$ production.

the precise shape of the reconstructed mass distribution, as well as to any kind of binning. Another encouraging feature, which will be seen below, is that the two measures both lead to similar conclusions on the optimal algorithms and R values.

As an example of the behaviour of these quality measures in an actual mass distribution, we show in Fig. 36 the quality measure $Q^{W}_{f=0.18}(R)$ in the case of W mass reconstruction for hadronic $t\bar{t}$ production. We observe that indeed in the case where the mass reconstruction is clearly poorer (blue dashed histogram), the value of $Q^{W}_{f=0.18}(R)$ is sizably larger.

With the aim of better comparing the performances of different jet definitions, we can establish a mapping between variations of these quality measures and variations in effective luminosity needed to achieve constant signal-over-background ratio for the mass peak reconstruction, working with the assumption that the background is flat and constant, and not affected by the jet clustering. We define the effective power to discriminate the signal with respect to the background ε for a given jet definition (JA,R) as

$$\varepsilon(JA;R) \propto \frac{N_{signal}}{\sqrt{N_{back}}}; \tag{30}$$

where N_{signal} and N_{back} are respectively the number of signal and background events. We can establish the following matching between variations in quality measures and in the effective luminosity ratios $\mathcal{L}$ as follows. Suppose that a quality measure calculated with (JA_2,R_2) gives a worse (i.e. larger) result than with (JA_1,R_1).

In the case of $Q^{W}_{f=z}(R)$ a larger value of this quality measure (i.e. a larger window width) will correspond to a larger number of background events for a given, fixed number of signal events. The jet definition $(JA_1;R_1)$ will then need a lower luminosity to deliver the same effective discriminating power as $(JA_2;R_2)$, since it deals with a smaller number of background events. So if we define

$$r_W \quad \frac{Q^{W}_{f=z}(JA_2;R_2)}{Q^{W}_{f=z}(JA_1;R_1)} = \frac{N_{back}(JA_2;R_2)}{N_{back}(JA_1;R_1)} > 1; \tag{31}$$

then at equal luminosity the discriminating power for $(JA_1;R_1)$ will be better by a factor

$$\frac{\varepsilon(JA_1;R_1)}{\varepsilon(JA_2;R_2)} = \sqrt{r_W}; \tag{32}$$

or equivalently the same discriminating power as $(JA_2;R_2)$ can be obtained with a smaller luminosity $L_1 = {}_L L_2$, where ${}_L$ is given by the inverse square of the ratio eq. (32).

$$ {}_L = \frac{1}{r_w} : \tag{33} $$

In the case of $Q^f_{w=x^p\overline{M}}(R)$ it is instead the number of signal events that varies when the quality measure changes. Defining

$$ r_f \quad \frac{Q^f_{w=x^p\overline{M}}(JA_2;R_2)}{Q^f_{w=x^p\overline{M}}(JA_1;R_1)} = \frac{N_{signal}(JA_1;R_1)}{N_{signal}(JA_2;R_2)} > 1; \tag{34} $$

then at equal luminosity the discriminating power for $(JA_1;R_1)$ will be better by a factor

$$ \frac{{}^e(JA_1;R_1)}{{}^e(JA_2;R_2)} = r_f ; \tag{35} $$

or equivalently the same discriminating power as $(JA_2;R_2)$ can be obtained with a smaller luminosity $L_1 = {}_L L_2$, where ${}_L$ is now given by the inverse square of the ratio eq. (35)

$$ {}_L = \frac{1}{r_f^2} : \tag{36} $$

In the remainder of this study we shall see that for the processes under consideration, the two quality measures indicate similar effective luminosity improvements to be gained by going from $(JA_2;R_2)$ to $(JA_1;R_1)$, once one takes into account the different functional dependence indicated above (e.g. a gain (i.e. *smaller*) by a factor of 2 in $Q^f_{w=x^p\overline{M}}(R)$ should correspond with good approximation to a gain of a factor of $2^2 = 4$ in $Q^w_{f=z}(R)$).

10.3 Jet algorithms

With the help of the quality measures defined in the previous section, we will study the performance of the following jet algorithms:

1. longitudinally invariant inclusive k_t algorithm [130–132].
2. Cambridge/Aachen algorithm [134, 135].
3. Anti-k_t algorithm [125].
4. SISCone [128] with split–merge overlap threshold $f = 0.75$, an infinite number of passes and no p_T cut on stable cones.
5. The Midpoint cone algorithm in CDF's implementation [127] with an area fraction of 1 and a maximum number of iterations of 100, split–merge overlap threshold $f = 0.75$ and seed threshold of 1 GeV.

In every case, we will add four-momenta using a E-scheme (4-vector) recombination. Each jet algorithm will be run with several values of R varying by steps of 0.1 within a range $[R_{min};R_{max}]$ adapted to observe a well defined preferred R_{best} value. Practically, we will have $R_{min} = 0.3$ and $R_{max} = 1.3$ for the Z^0 analysis and $R_{min} = 0.1$ and $R_{max} = 1.0$ for the $t\bar{t}$ samples.

Note that we have fixed the value of the overlap parameter of the cone algorithms to $f = 0.75$. This rather large value is motivated (see *e.g.* [147]) by the fact that "monster jets" can appear for smaller values of f. For sequential recombination clustering algorithms we use their inclusive longitudinally-invariant versions, suited for hadronic collisions. The jet algorithms have been obtained via the implementations and/or plugins in the `FastJet` package [144].

The infrared-unsafe CDF midpoint algorithm is only included here for legacy comparison purposes.

Process	# Gen. events	# Acc. events	Fraction acc. vs. gen.	Fraction f in Eq. 28
$Z^0 \to q\bar{q}$	50 000	23 000	0.46	0.12
Hadronic $t\bar{t}$	100 000	75 000	0.75	0.18

Table 7: Number of generated and accepted events for each process, the corresponding approximate fraction of accepted events and the fraction f of the total number of generated events which correspond to a 25% of the selected events.

10.4 Physical processes

We consider the following physical processes: $Z^0 \to q\bar{q}$ for various values of M_{Z^0} and fully hadronic $t\bar{t}$ production, and we reconstruct the mass of the Z^0 boson and that of the W boson and the top quark to assess the performance of the jet algorithms described in Sect. 10.3. We should emphasise again that the performance of a given jet definition depends on the process under consideration, thus it is important to study different jet algorithms for diverse processes with different mass scales, kinematics and jet structure.

All the samples have been generated with `Pythia` 6.410 [85] with the DWT tune [84]. For the $t\bar{t}$ samples the B mesons have been kept stable to avoid the need of B decay reconstruction for B tagging[26]. The top quark mass used in the generation is $M_t = 175$ GeV while the W mass is $M_W = 80.4$ GeV.

Now we describe for each process the main motivations to examine it and the mass reconstruction techniques employed, while results are discussed in the next section. The fraction of events that pass the selection cuts discussed above is to a good approximation independent of the particular jet definition, and their values can be seen in Table 7.

$Z^0 \to q\bar{q}$ for various values of M_{Z^0}.

This process serves as a physically well-defined source of monochromatic quarks. By reconstructing the dijet invariant mass one effectively obtains a measure of the p_T resolution and offset for each jet definition. The range of Z^0 masses is: 100, 150, 200, 300, 500, 700, 1000, 2000 and 4000 GeV. Many of these values are already excluded, but are useful to study as measures of resolution at different energies. Note also that the generated Z^0 particles have a narrow width ($_{Z^0}$ 1 GeV). This is not very physical but useful from the point of view of providing monochromatic jet sources. For each event, the reconstruction procedure is the following:

1. Carry out the jet clustering based on the list of all final-state particles
2. Keep the two hardest jets with p_T 10 GeV. If no such two jets exist, reject the event.
3. Check that the two hard jets have rapidities $|y|$ 5, and that the rapidity difference between them satisfies $|y|$ 1. If not, reject the event.
4. The Z^0 is reconstructed by summing the two jets' 4-momenta.

Fully hadronic $t\bar{t}$ decay.

This process provides a complex environment involving many jets in which one can test a jet definition's balance between quality of energy reconstruction and ability to separate multiple jets. The reconstruction of M_W and M_t is obtained as follows:

1. Carry out the jet clustering based on the list of all final-state particles
2. Keep the 6 hardest jets with p_T 10 GeV and $|y|$ 5. If fewer than 6 jets pass these cuts, reject the event.
3. Among those 6 jets, identify the b and the $\bar{b}$ jets. If the number of b=$\bar{b}$ jets is not two, then reject the event.
4. Using the four remaining jets, form two pairs to reconstruct the two W bosons. Among the

[26] The effects of imperfect B tagging should be addressed in the context of detector simulation studies.

3 possible pairings, choose the one that gives masses as close as possible to the nominal W mass.

5. Reconstruct the two top quarks by pairing the b and W jets. Pairing is done by minimising the mass difference between the two candidate t jets.

10.5 Results

Now we discuss the results for the mass reconstruction of the processes described in section 10.4 with the jet algorithms of section 10.3. We quantify the comparison between different jet definitions using the quality measures defined in section 10.2. We note that in the various histograms of this section, the lines corresponding to different jet algorithms have been slightly shifted in order to improve legibility.

10.6 Analysis of the Z^0 samples

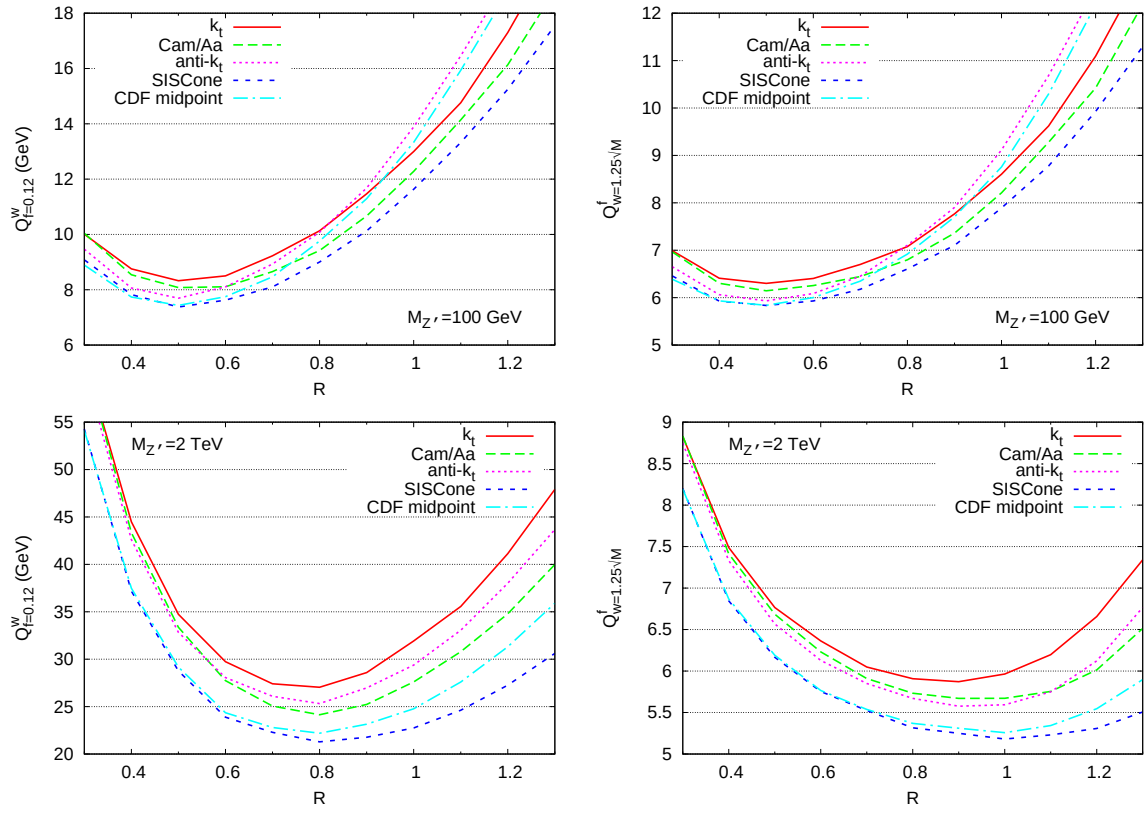

Fig. 37: The figures of merit $Q^{w}_{f=0.12}(R)$ and $Q^{f}_{w=1.25\sqrt{M}}(R)$ for the Z^0 samples corresponding to $M_{Z^0} = 100$ GeV (upper plots) and $M_{Z^0} = 2$ TeV (lower plots).

The figures of merit for $Q^{w}_{f=0.12}(R)$ and $Q^{f}_{w=1.25\sqrt{M}}(R)$ are plotted in Fig. 37, as a function of the radius R for a Z^0 of 100 GeV and 2 TeV. Each plot includes the results for the five jet algorithms under consideration. There are two lessons we can learn from this figure. Firstly, even though some algorithms give better quality results than others (we will come back on this later), the main source of quality differences does not come from the choice of algorithm but rather from the adopted value for R. Secondly, the minimum of the quality measures gives, for each jet algorithm, a preferred value $R^{M_{Z^0}}_{best}$ for R.

That preferred value over the whole range of Z^0 masses is shown[27] in Fig. 38. We observe that

[27] Varying R continuously between 0.3 and 1.3 would probably result in a smoother curve for R_{best} as a function of M_{Z^0}. However, there is no real interest in determining an R parameter with more than one decimal figure.

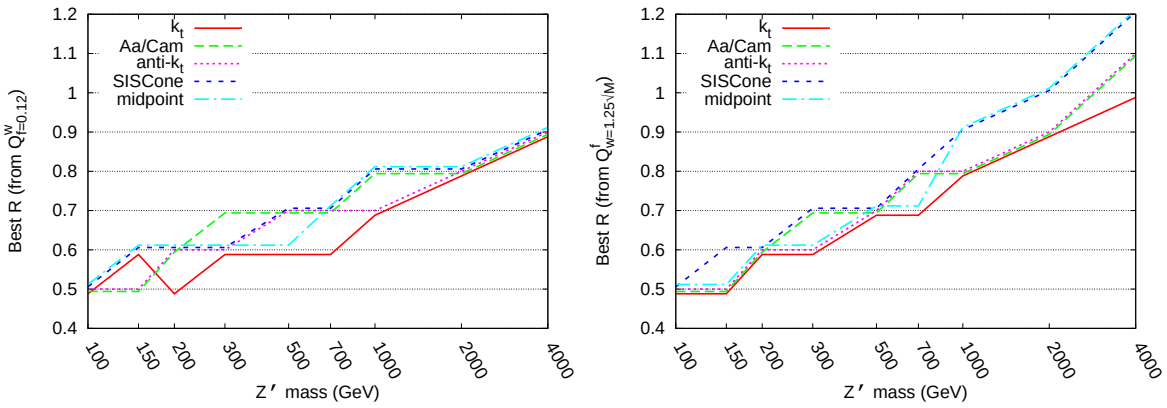

Fig. 38: The best value of the jet radius R_{best} (defined as the minimum of the corresponding figure of merit) as determined from $Q^{w}_{f=0.12}(R)$ (left plot) and $Q^{f}_{w=1.25\sqrt{M}}(R)$ (right plot) as a function of $M_{Z'}$.

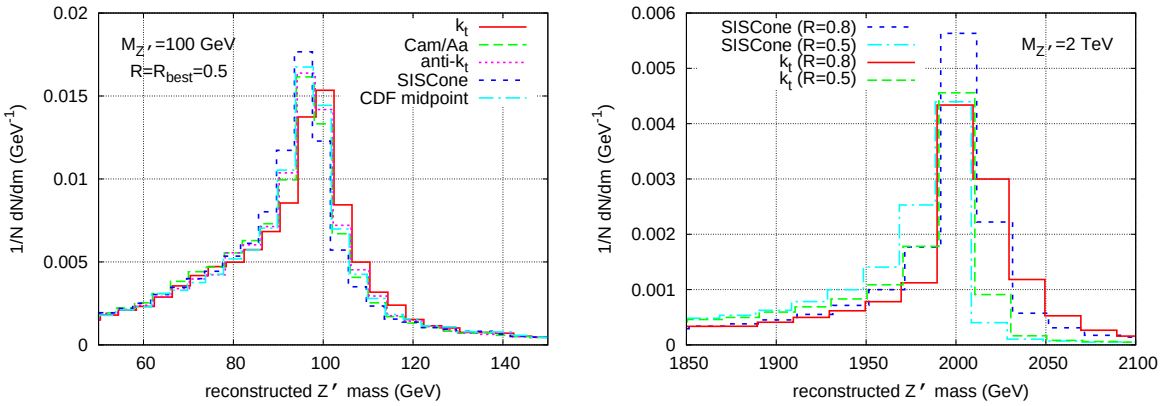

Fig. 39: The invariant mass distribution in the Z' samples for two different values of $M_{Z'}$.

the two quality measures roughly agree on the extracted preferred value, with the possible exception of the largest values of $M_{Z'}$ for which we observe small differences. Furthermore, when the mass of the Z' becomes larger, the best quality is also achieved using larger values for R: R_{best} goes from 0.5 for low Z' masses, to R_{best} 0.9 for high Z' masses.

This behaviour can be explained by the fact the as $M_{Z'}$ increases, perturbative radiation (which favours larger R) grows larger (roughly as M) while the underlying event contribution (which favours smaller R) stays fixed, thus resulting in an overall larger value for the optimal R [117]. Another relevant point is that Z' decays are mostly dijet events, so the invariant mass reconstruction is in general not affected by the accidental merging of hard partons that takes place for larger values of R in multi-jet environments like hadronic $t\bar{t}$ decays.

Given our method to quantitatively analyse the performance of jet algorithms and to extract a preferred value for R, there are a few more interesting figures we want to look at. The first one, Fig. 39, is simply the histogram of the reconstructed Z' mass. The left plot shows the reconstructed Z' peaks for the five algorithms at $R = R_{best}$ and though some slight differences exist all algorithms give quite similar results. In the right plot we show the reconstructed Z' histogram for the k_t and the SISCone algorithms using either $R = R^{2\,\mathrm{TeV}}_{best} = 0.8$, as extracted from the quality measures at 2 TeV, or $R = R^{100\,\mathrm{GeV}}_{best} = 0.5$, extracted at 100 GeV. The behaviour is again what one expects from Fig. 37, namely that SISCone with $R = 0.8$ performs a bit better than SISCone with $R = 0.5$ and k_t with $R = 0.8$, which themselves give a better peak than the k_t algorithm with $R = 0.5$.

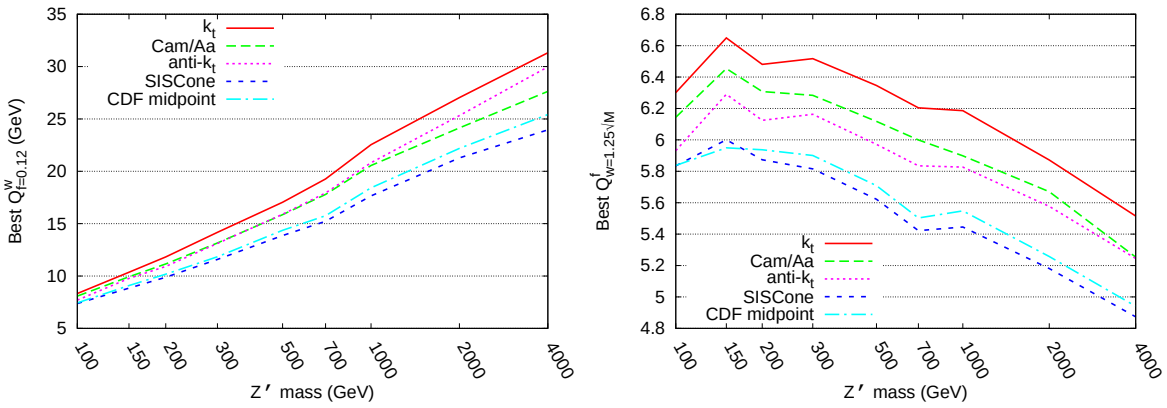

Fig. 40: The best value of the figures of merit $Q^w_{f=0.12}(R)$ and $Q^f_{w=1.25\sqrt{M}}(R)$ within all possible values of R as a function of M_{Z^0}.

Let us now consider again the whole range of Z^0 masses and discuss the initial point of interest which is finding the best algorithm to be used in jet analysis, at least from the point of view of Z^0 reconstruction. To that aim, we look at the quality measure at R_{best} as a function of the Z^0 mass and for each jet algorithm. The results are presented in Fig. 40 for $Q^w_{f=0.12}(R)$ (left plot) and $Q^f_{w=1.25\sqrt{M}}(R)$ (right plot). Note that $Q^w_{f=0.12}(R)$ has an approximately linear increase with $\ln M_{Z^0}$, while $Q^f_{w=1.25\sqrt{M}}(R)$ has a similar behaviour but in the opposite direction.

The generic conclusion is that cone algorithms split–merge (SM) steps perform better than the recombination-type algorithms, though we again emphasise that the difference is rather small and, in particular, smaller than the dependence on the parameter R. This conclusion is valid for all Z^0 masses and for both quality measures. In general, among the cone algorithms, SISCone produces results slightly better than CDF-Midpoint while, among the recombination-type algorithms, k_t is a bit worse than Cambridge/Aachen and anti-k_t, the ordering between those two depending on the mass and quality measure under consideration.

This can be understood due to the fact that SISCone has a reduced sensitivity to the underlying event (smaller effective area [147]) while stretching out up to larger distances[28], thus is able to merge emitted partons even at relatively large angles. Note that this feature, which is advantageous in a clean environment like $Z^0 \to q\bar{q}$, essentially a dijet event, is on the other hand something that degrades jet clustering with SISCone on denser environments like $t\bar{t}$.

We can quantify the differences between jet algorithms at R_{best} using the mapping between quality measures and effective luminosity ratios introduced in Sect. 10.2. For $M_{Z^0} = 100$ GeV, both quality measures coincide in that when comparing the best jet algorithm (SISCone) with the worst (k_t) one finds $_L \quad 0.85$, while for the $M_{Z^0} = 2$ TeV case, one finds that the effective luminosity ratio is $_L \quad 0.8$.

An important consequence that can be drawn for this analysis is that optimising the value of R for a given jet algorithm is crucial to optimise the potential of a physics analysis. For example, in the $M_{Z^0} = 2$ TeV case, if one chooses $R = 0.5$ (based e.g. on considerations for the $M_{Z^0} = 100$ GeV process) instead of the optimal value $R_{\text{best}} \simeq 0.8$, it is equivalent to losing a factor $_L \quad 0.75$ in luminosity (for all algorithms and both quality measures). We note that the optimal value of R at high masses is somewhat larger than what is being considered currently in many studies by the LHC experiments.

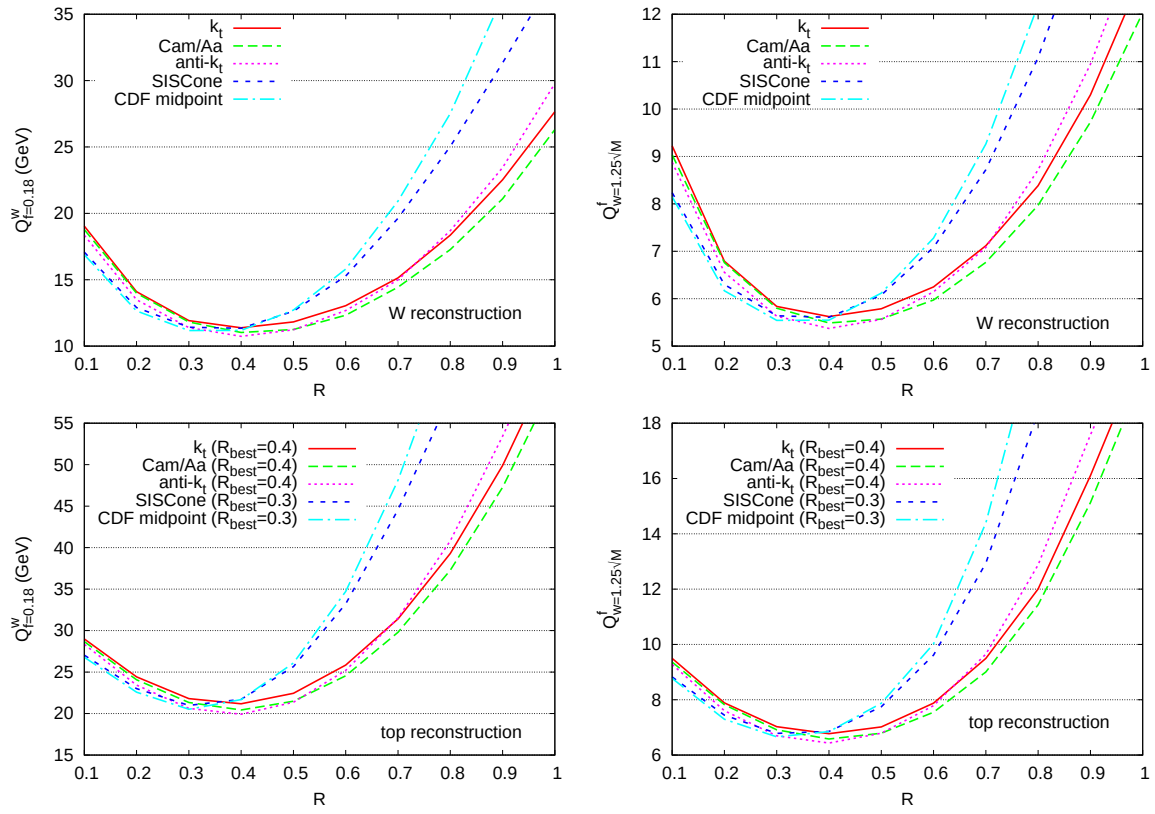

Fig. 41: The figures of merit $Q^{w}_{f=0.18}(R)$ and $Q^{f}_{w=1.25\sqrt{M}}(R)$ for the invariant mass distributions of the hadronic $t\bar{t}$ samples.

10.7 Analysis of the hadronically decaying $t\bar{t}$ sample

Hadronic $t\bar{t}$ production is a challenging environment since the jet algorithm has to reconstruct at least 6 hard jets. In this process one can test a jet definition's balance between quality of energy reconstruction and ability to separate multiple jets.

For each of the mass distributions that we reconstruct in this case (that of the W boson and that of the top quark), we show the plots of the corresponding figures of merit $Q^{w}_{f=0.18}(R)$ and $Q^{f}_{w=1.25\sqrt{M}}(R)$ in Fig. 41. Although all jet algorithms perform rather similarly at R_{best}, there is a slight preference for the anti-k_t algorithm. The resulting effective luminosity ratio computed for the top reconstruction between the two limiting algorithms is $_{L}$ 0.9. Note that at larger values of R the cone algorithms perform visibly worse than the sequential recombination ones, probably because they tend to accidentally cluster hard partons which should belong to different jets. In the same spirit, the preferred radius is $R_{best} = 0.4$ for sequential recombination algorithms, while cone algorithms tend to prefer a somewhat smaller optimal value $R_{best} = 0.3$.

For the hadronic $t\bar{t}$ samples, we show the invariant mass distributions at R_{best} in each case for M_W and M_t in Fig. 42. We observe that all algorithms lead to rather similar results at the optimal value of the jet radius.

Then, in Fig. 43, we compare the W and t invariant mass distributions for the hadronic $t\bar{t}$ samples for the best overall algorithm anti-k_t and for SISCone, both with $R = R_{best}$, compared to their counterparts for $R = 0.7$. We observe that, as indicated by the figures of merit, the choice $R = R_{best}$ for the anti-k_t algorithm leads to a somewhat larger number of events in the peak than for SISCone, but in

[28] In the limiting case it can merge two equally hard partons separated by a angular distance $2R$.

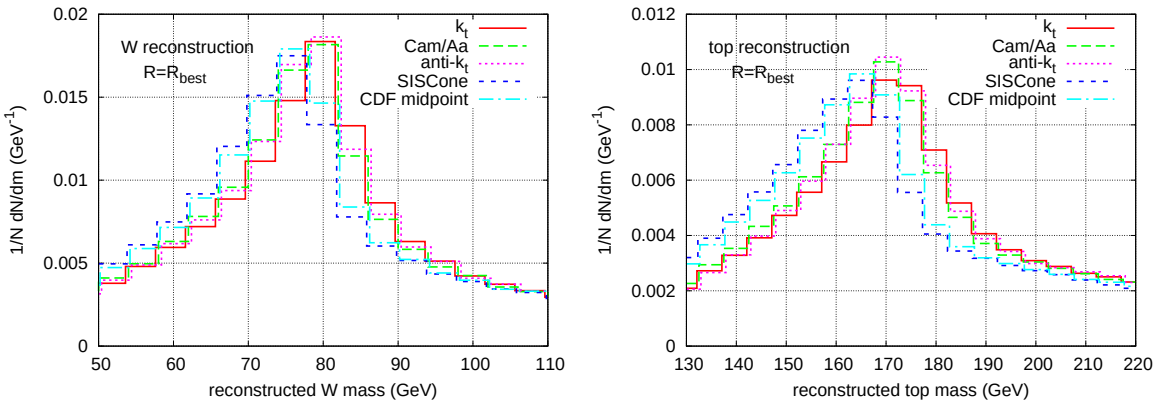

Fig. 42: The W and t invariant mass distributions for the hadronic $t\bar{t}$ samples for $R_{best} = 0.4$.

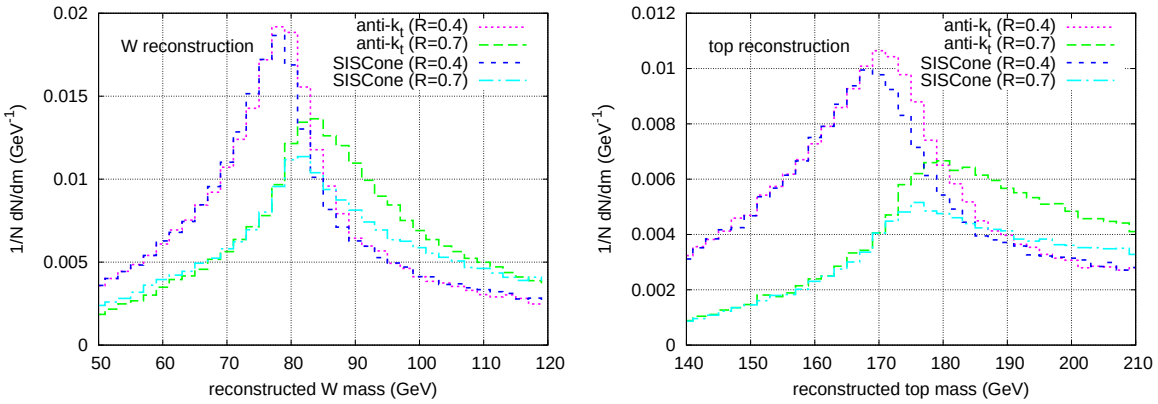

Fig. 43: The W and t invariant mass distributions for the hadronic $t\bar{t}$ samples for $R_{best} = 0.4$ for the best overall algorithm anti-k_t and SISCone compared to their counterparts for $R = 0.7$.

any case this difference is small compared with the difference between $R = R_{best}$ and $R = 0.7$. The degradation of the mass peak at large R is both due to contamination from the UE and to the fact that hard partons are sometimes accidentally merged (more often in cone algorithms with SM steps).

As in the Z^0 case, one of the main results of this study is that choosing a non-optimal value of R can result in a severe degradation of the quality of the reconstructed mass peaks. For example, comparing in Fig. 43 the results for $R = R_{best}$ and $R = 0.7$, we observe that the degradation of the mass peak can be of the order of 40 50%, confirmed by the quality measures, for which we obtain $_L$ 0.3 0.6. Thus our analysis confirms that the relatively small values of R currently being used by the LHC experiments in top reconstruction are appropriate. Specific care is needed with cone algorithms with split-merge stages, for which one should make sure that R is not larger than 0.4.

As a final remark we note that we have also examined semi-leptonic $t\bar{t}$ decays. Though there are fewer jets there, the results are rather similar (with slightly larger differences between algorithms), mainly because the semileptonic case resembles a single hemisphere of the fully hadronic case.

10.8 Summary

We have presented in this contribution a general technique to quantify the performance of jet algorithms at the LHC in the case of the mass reconstruction of heavy objects.

One result is that for simple events, as modelled by a fake Z^0 decay at a range of mass scales, SISCone and the midpoint algorithm behave slightly better than others, presumably because they reach

furthest for hard perturbative radiation, but without taking in extra of underlying event contamination. Quantitatively, our performance measures suggests that one can obtain equivalent signal$/\sqrt{\text{background}}$ with a factor $_L \simeq 0.8 - 0.9$ less luminosity than for (say) the k_t algorithm. The Cambridge/Aachen and anti-k_t algorithms are intermediate.

An effect of sometimes greater significance is the dependence of the results on the choice of R parameter. In particular we find that the optimal R increases significantly with mass scale M_{Z^0}, most probably for the reasons outlined in [117], namely an interplay between perturbative effects (which scale M_{Z^0} and prefer a larger R) and non-perturbative effects (independent of M_{Z^0} and favouring smaller R). If one takes $R = 0.5$, which is optimal at $M_{Z^0} = 100$ GeV, and uses it at $M_{Z^0} = 2$ TeV, it's equivalent to a loss of luminosity of a factor of $_L \simeq 0.75$ compared to the optimal $R \simeq 0.9$. The need for large R is likely to be even more significant for resonances that decay to gluons, as suggested by the study in section 11.

We have also examined more complex events, hadronic decays of $t\bar{t}$ events. Here the need to resolve many different jets modifies the hierarchy between algorithms, with anti-k_t performing best. Overall the differences between algorithms are however fairly small, with an effective luminosity reduction from best to worst of $_L \simeq 0.9$. The choice of the correct R is even more important here than in the Z^0 case, with small values $R \simeq 0.4$ being optimal.

Let us emphasise that our results should be taken with some care, since in general the jet clustering procedure will affect the background as well as the signal, and our measures ignore this effect. Nevertheless, while our analysis cannot replace a proper experimental $S/\sqrt{B}$ study, it does provides an indication of the typical variations that might be found in different jet definition choices at the LHC, and points towards the need for flexibility in jet finding at the LHC.

The strategy presented in this contribution can be readily applied to quantify the performance of different ideas and strategies for improving jet finding at the LHC. One possibility is the use of subjet capabilities of sequential clustering algorithms, similar to what was done in [118], but extended beyond that context. This potential for future progress in jet-finding methods is yet another reason for encouraging flexibility in LHC jet-finding.

Finally, all the MC data samples used in this contribution, together with the results of mass reconstruction using different jet algorithms can be found at the following webpage:

`http://www.lpthe.jussieu.fr/~salam/les-houches-07/`

.

Acknowledgements

This work has been supported in part by the grant ANR-05-JCJC-0046-01 from the French Research Agency and under Contract No. DE-AC02-98CH10886 with the U.S. Department of Energy.

11 INFLUENCE OF JET ALGORITHMS AND JET SIZES ON THE RECONSTRUCTION OF THE HARD PROCESS FROM STABLE PARTICLES AT LHC ENERGIES [29]

11.1 Introduction

With the advent of the LHC, a new regime in center-of-mass energy for hadron-hadron collisions will be accessed and the by far dominant feature of the events to be measured is the abundant production of jets, i.e. collimated streams of hadrons that are supposed to originate from a common initiator. In theory, these initiators are usually the outgoing partons of a hard interaction calculable in perturbative QCD (pQCD). Limitations of QCD perturbation theory, however, make it impossible to unambiguously assign a bunch of observed hadrons to such a hard parton. To achieve nevertheless the comparability of our best theoretical knowledge with experimental results, jet algorithms are employed that define a distance measure between objects and uniquely determine which of them are sufficiently close to each other to be considered to come from the same origin and hence to combine them into a jet. This same procedure is applied equally to the partons of theory calculations, the final state particles of Monte-Carlo generators, that serve as input to experiment simulations, as well as measured deposits of energy in calorimeters or tracks of charged particles. Provided the jet algorithms are well behaved, i.e. they are especially collinear- and infrared-safe (CIS), the measured jets can now be related to jets constructed of the theory objects.

However, a number of residual effects of either experimental origin or of theoretical nature, the latter comprising perturbative radiation, hadronization and the underlying event (UE), still have to be taken into account. Recent overviews showing how these have been dealt with in the past, especially at Tevatron, can be found in e.g. Refs. [84, 123]. Since energies reachable at the LHC are much larger though than everything investigated so far, the best choices of jet algorithms and parameters to delimit and/or control these residual effects have to be reevaluated. In this work we contribute to this effort by examining the influence of different jet algorithms and jet sizes on the reconstruction of characteristics of a hard process. More precisely, we have varied the respective jet size parameters, usually labelled as R or D and generically denoted as R further on, from 0.3 to 1.0 in steps of 0.1 for the following four algorithms:

The Midpoint cone algorithm, Ref. [127] (with split-merge overlap threshold f of 0.75 and a seed threshold of $1\,GeV$)

The SISCone algorithm, Ref. [128] (with split-merge overlap threshold f of 0.75, an infinite number of passes and no transverse momentum cut on stable cones)

The k_t algorithm, Refs. [131, 132, 149], in the implementation of Ref. [144]

The Cambridge/Aachen algorithm, Refs. [134, 135]

In all cases the four-vector recombination scheme or E scheme was used. We note that Midpoint cone is not collinear and infrared-safe and is included primarily for comparison.

In this first step, we restrict the analysis to examine the transition from leading-order (LO) pQCD events to fully hadronized ones using Pythia version 6.4, Ref. [85], as event generator. The parameter set of tune DWT, Ref. [150], has been chosen to represent a possible extrapolation of the underlying event to LHC energies. On occasion we have employed the S0 tune, Refs. [86, 151],[30] as an alternative. A more complete study is foreseen including further models as given by Herwig plus JIMMY, Refs. [97, 152], or Herwig++, Ref. [153, 154].

With this set-up, three primary types of reactions have been considered representing typical analysis goals:

Inclusive jet production for comparison with higher-order perturbative calculations and fits of parton density functions,

[29] Contributed by: V. Büge, M. Heinrich, B. Klein, K. Rabbertz

[30] In addition to the settings given in table I of Ref. [86], the parameters MSTP(88) and PARP(80) have been set to the non-default values of 0 and 0.01 resp. as they would be set by a call to the corresponding PYTUNE routine.

Z boson production in association with a balancing jet for a similar usage but in addition for jet calibration purposes and

production of heavy resonances with the aim of finding new particles and measuring their masses.

The choice of resonance produced, $H \to gg$, has been made so as to serve as well-defined source of monochromatic gluons and less as a realistic analysis scenario. Finally, we adopt a final state truth definition for the jet finding taking all stable[31] particles as input apart from prompt leptons or leptons from decays of heavy resonances.

Additional requirements imposed by the experimental set-up and e.g. the jet energy calibration or pile-up have to be investigated in further studies.

11.2 Inclusive jets

For inclusive jet transverse momentum spectra one emphasis is on the comparison of measured data with QCD perturbation theory to higher order, see for example Refs. [155–158]. Currently, calculations up to NLO are at disposal in the form of JETRAD, Ref. [159], or NLOJET++, Refs. [160, 161], which, like most programs of the cross section integrator type, remain at the parton level and do not allow to attach perturbative parton showers with subsequent hadronization so that a full simulation of these events is excluded.[32] As a consequence, when referring calibrated experimental data unfolded for detector effects to the NLO calculation, the required corrections cannot be determined in a completely consistent way. The theoretical "truth", i.e. NLO in this case, lies inbetween the LO matrix element (ME) cross section and the LO cross section with attached parton showers. Therefore we present in the following ratios of the inclusive jet p_T spectra of fully hadronized events with respect to a LO matrix element calculation. To focus on the hadronization step alone, the same was performed with respect to the spectrum derived from events including parton showers but without fragmentation. In the latter case one should note that the parton radiation has been performed for the hard interaction as well as for the underlying event so that this corresponds only to one part of the desired correction. Most interesting would be a comparison to the correction achievable with a NLO program with matched parton showers like MC@NLO, Refs. [23,162], for which unfortunately the inclusive jets have not yet been implemented. A theoretical study going into more detail on the subject of the composition of perturbative (parton showers) and non-perturbative (underlying event, hadronization) corrections to hard interactions can be found in Ref. [117].

In this section, the jets have been required to have a minimal transverse momentum p_T larger than $50\,GeV$. No cut on the jet rapidity or polar angle was imposed. Figure 44 shows the ratio of inclusive jet cross sections of fully hadronized events by Pythia DWT tune over Pythia LO ME for jet sizes R of 0.3 up to 1.0 for the investigated jet algorithms. For the latter, the respective parameters of the Pythia program controlling the parton shower, initial and final state radiation, multiple parton interactions (MPI) and the fragmentation have been switched off. It becomes obvious, that the effects increasing the jet p_T, initial state radiation and multiple parton interactions, and the effects reducing the jet p_T are relatively well balanced for R around 0.5 to 0.6 for Midpoint cone and SISCone as well as for k_T and Cambridge-Aachen. For smaller R, the jets tend to lose p_T due to out-of-cone effects during the evolution from LO ME to hadronized events, while larger R result in an increase of p_T due to the jets collecting particles from other sources. Corrections to derive the LO ME jet cross section from the hadronized final state will have to take these effects into account.

In Figure 45 the jet p_T distribution of fully hadronized events has been divided by the spectrum after parton showers (including the underlying event) for the same range of jet sizes R as above. This shows predominantly the influence of the hadronization model, Lund string fragmentation in the case of Pythia, on the jets, usually leading to a loss in p_T especially for cone-type algorithms and more pro-

[31] Particles with lifetimes such that $c\,10\,mm$.

[32] Additionally, it would be necessary to perform an unweighting step in order to avoid simulating huge amounts of events with positive and negative weights.

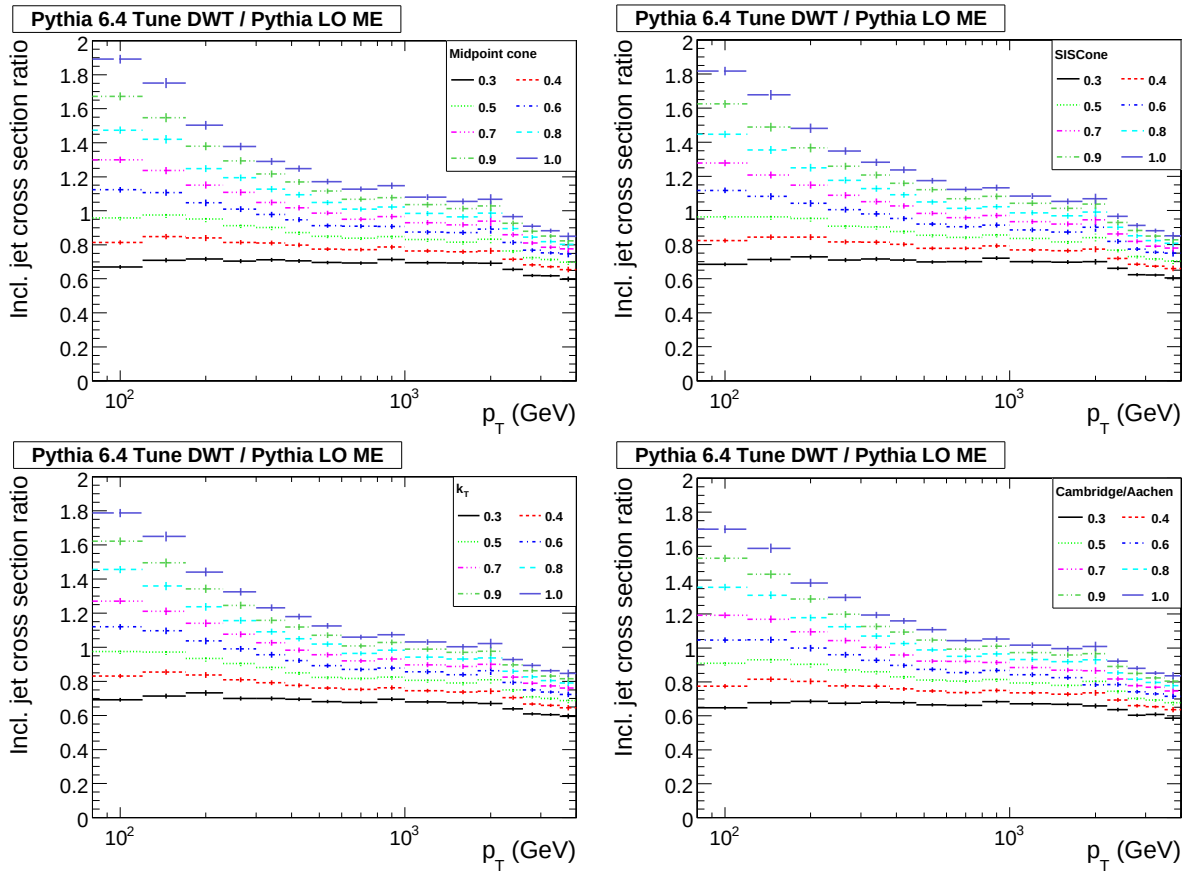

Fig. 44: Ratio of inclusive jet cross sections of fully hadronized events by Pythia DWT tune over Pythia LO ME for jet sizes R of 0.3 up to 1.0 for the Midpoint cone (upper left), SISCone (upper right), k_T (lower left) and Cambridge/Aachen algorithm (lower right).

nounced for smaller cone sizes due to out-of-cone effects. The sequential recombination type algorithms like k_T and Cambridge/Aachen are almost unaffected for all choices of R.

Finally, to emphasize the importance of the underlying event we present in Fig. 46 the same ratios as in Fig. 44 but for the alternative tune S0 employing a completely new model for both, parton shower and multiple parton interactions. Events produced with this tune contain a small fraction of jets with p_T significantly higher than it would be expected from the imposed phase space restrictions on the event generation. These events had to be removed manually to avoid artefacts in the inclusive jet cross sections due to their high weights and the procedure to combine event samples generated separately in bins of the hard momentum scale. The number of discarded events is well below one percent for all algorithms and jet sizes R.

As can be seen, the fully hadronized tune S0 events generally contain jets with higher p_T than the events produced with tune DWT, which is mainly due to an increased amount of energy spread into the event by the new MPI model. This yields the somewhat surprising consequence that an R of 0.4 delivers a ratio that is very close to unity for all applied jet algorithms over the whole p_T range.

11.3 Z plus jets

At LHC energies, events with Z bosons and jets will be much more abundant than at the Tevatron. Therefore the aspect of calibrating jet energies using the balancing transverse momentum of a reconstructed Z boson will become more important. In addition, Z plus jet reconstruction suffers much less

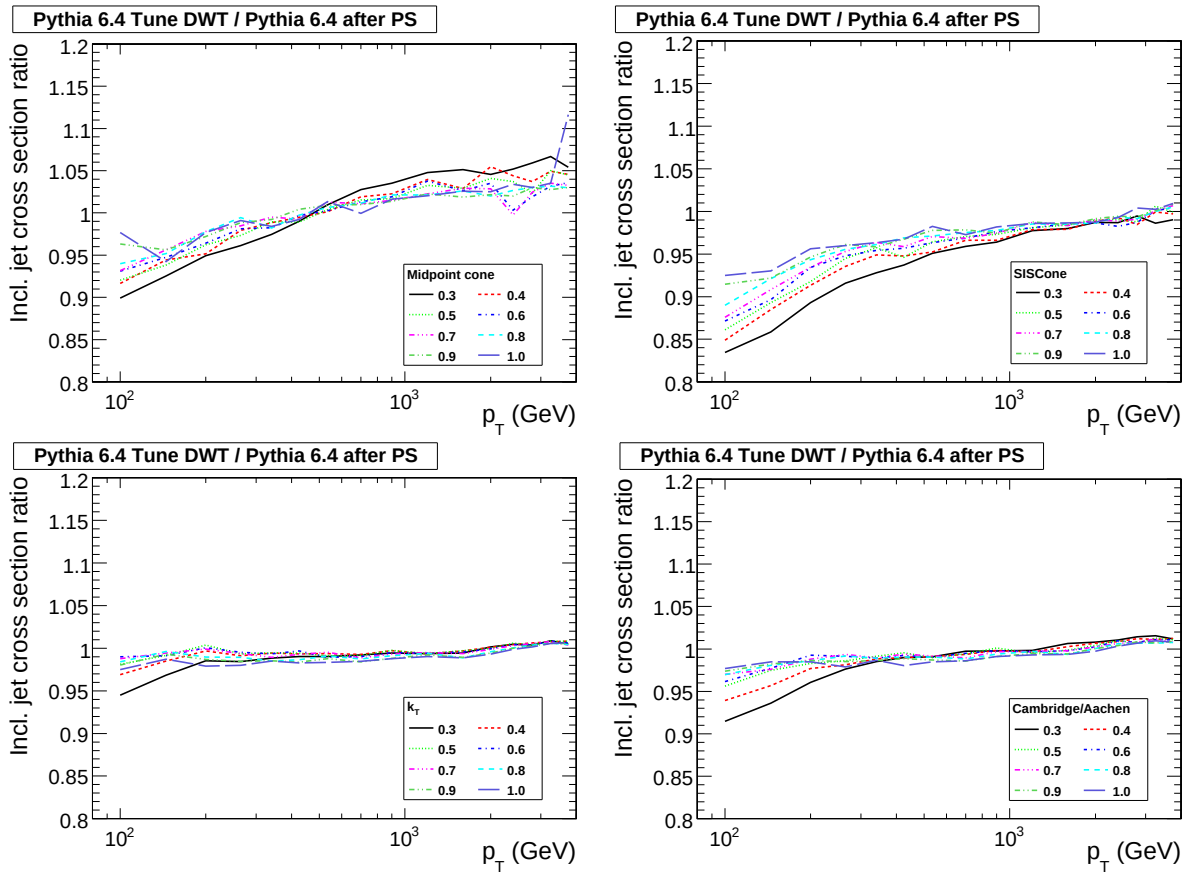

Fig. 45: Ratio of inclusive jet cross sections as in Fig. 44 but divided by Pythia tune DWT after parton showers (including the underlying event). This shows predominantly the influence of the hadronization model.

from backgrounds than the similarly useful photon plus jets process, where the huge cross section for di-jet production requires, due to misidentified jets, to impose strong isolation criteria on the photons. Restricting the analysis to decays of the Z boson into two muons, as done here, has the further advantage to decouple completely the jet energy scale from calorimetric measurements and to relate it to the muon track reconstruction instead.[33]

In the following, events will be selected with respect to the best possible jet calibration. The quantity we will be looking at is the average relative deviation of the reconstructed jet p_T from the transverse momentum of the balancing Z boson $(p_{T,\text{jet}} \quad p_{T,Z})=p_{T,Z}$. As this is only valid for events, in which the Z boson is exactly balanced by one jet of the hard process, one has to extract a clean sample of Z plus one jet events. Additional selection criteria are imposed due to geometrical and triggering limitations of a typical LHC detector.

A precise measurement of the muon kinematics with a tracking system is assumed to be feasible in the region in pseudo-rapidity[34] $j\,j$ of up to 2.4. Due to possible trigger constraints, only events are considered where both muons have transverse momenta larger than $15\,\text{GeV}$. Having identified two or more muons in an event, the pair of muons with opposite charge and an invariant mass closest to the Z mass is chosen. The event is accepted if the invariant mass of this di-muon system is closer to the Z mass than $20\,\text{GeV}$. Likewise, from the jet collection only jets in the central region with $j\,j< 1.3$ are

[33] Nevertheless, Z decays into electron-positron pairs are very useful, since already the electromagnetic energy scale is known more precisely than the hadronic one and also here track information can be exploited.

[34] $=\quad \ln(\tan\frac{}{2})$

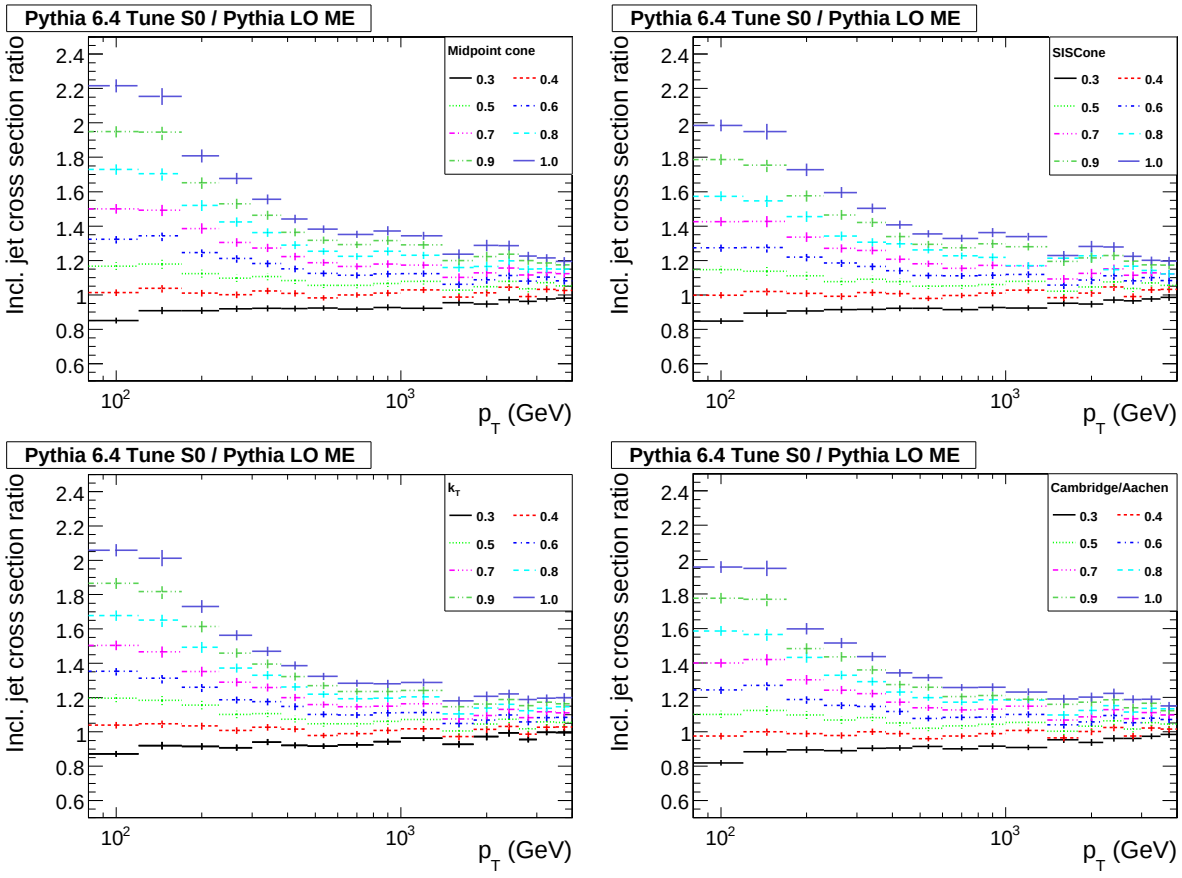

Fig. 46: Ratio of inclusive jet cross sections as in Fig. 44 but for events with Pythia tune S0.

selected where uncalibrated but otherwise reliable jet energy measurements are expected. In addition, the jets are required to have a minimal transverse momentum of $20\,\mathrm{GeV}$.

In the current implementation of the analysis, all stable particles are selected as input objects to the jet algorithm, including the two muons from the decay of the Z boson. This leads to two fake jets in the event which have to be removed manually from the jet collection. This is done by discarding jets which lie inside a cone of $R < 0.5$ around the directions of the two muons. [35] As the Z-jet system is balanced in azimuth , the muon fake jets are in the opposite hemisphere and therefore do not interfere with the determination of the properties of the jet balancing the Z boson so that the final state truth definition given in the introduction still holds.

In order to ensure a clean sample of events in which the Z boson is exactly balanced against one jet of the hard process, the second leading jet in transverse momentum is required to have less than 20% of the transverse momentum of the Z boson. In addition, the leading jet in p_T is required to be opposite in azimuthal angle by complying with $j(\mathrm{jet}, Z) < 0.15$.

The relative deviation of the reconstructed jet p_T from the transverse momentum of the balancing Z boson $(p_{T,\mathrm{jet}} - p_{T,Z})/p_{T,Z}$ is determined independently for each range in the hard transverse momentum scale set for the event generation. The mean and width of the relative difference of jet and boson p_T is performed in a two step procedure employing Gaussian fits where the first one is seeded with the mean and root mean squared (RMS) of the corresponding histogram. The second fit then uses the result of the first step as input.

[35] $R = \sqrt{(\ )^2 + (\ )^2}$

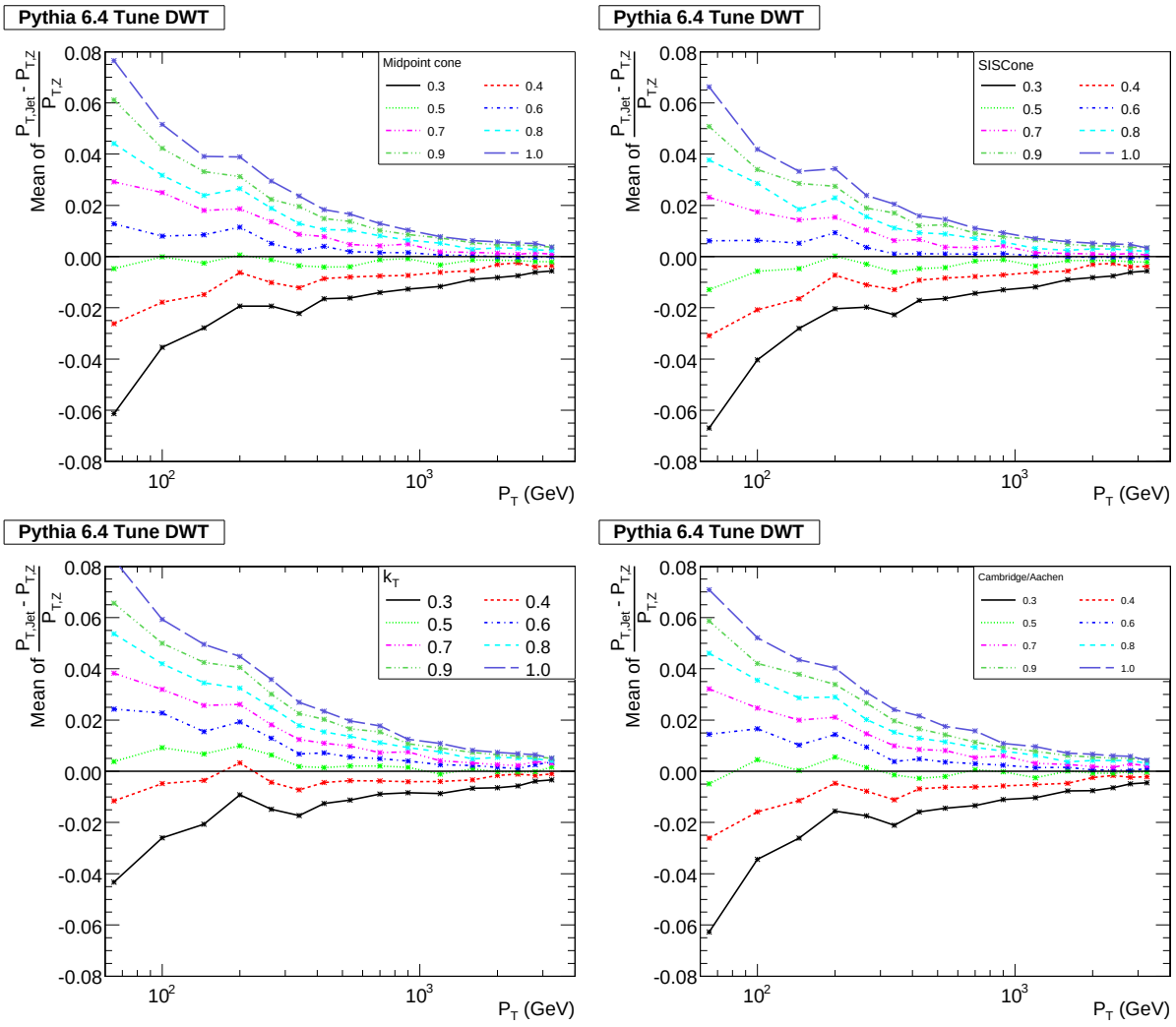

Fig. 47: Relative deviation between the transverse momentum of the jet and the balancing Z boson from a Gaussian fit of fully hadronized Pythia tune DWT events versus p_T for jet sizes R of 0.3 up to 1.0 for the Midpoint cone (upper left), SISCone (upper right), k_T (lower left) and Cambridge/Aachen algorithm (lower right).

Figure 47 presents this observable for fully hadronized Pythia tune DWT events versus p_T for jet sizes R of 0.3 up to 1.0 of the investigated algorithms. All four exhibit a very similar behaviour that small jet sizes on average under- and large jet sizes overbalance the transverse momentum of the Z. Above 500 GeV this difference remains well below 2%. To smaller transverse momenta the balance gets increasingly worse. No particular advantage can be identified for any of the four algorithms and it is always possible to choose a suitable jet size to minimize the deviations. But any such choice depends, of course, heavily on the interplay of jet energy loss due to parton showers and hadronization and energy gain because of the underlying event.

To give an estimate of the influence of the underlying event, the same quantity is shown for comparison in Fig. 48 for the alternative Pythia tune S0 for the four algorithms. The smaller jet sizes show nearly the same behaviour for both tunes. For the larger jet sizes a slight loss in energy for the tune S0 compared to DWT is exhibited. The effect decreases for larger transverse momenta.

In order to examine the influence of the underlying event on the jet energy in dependence of the jet size, the mean of the relative deviation between the transverse momentum of the jet and the balancing

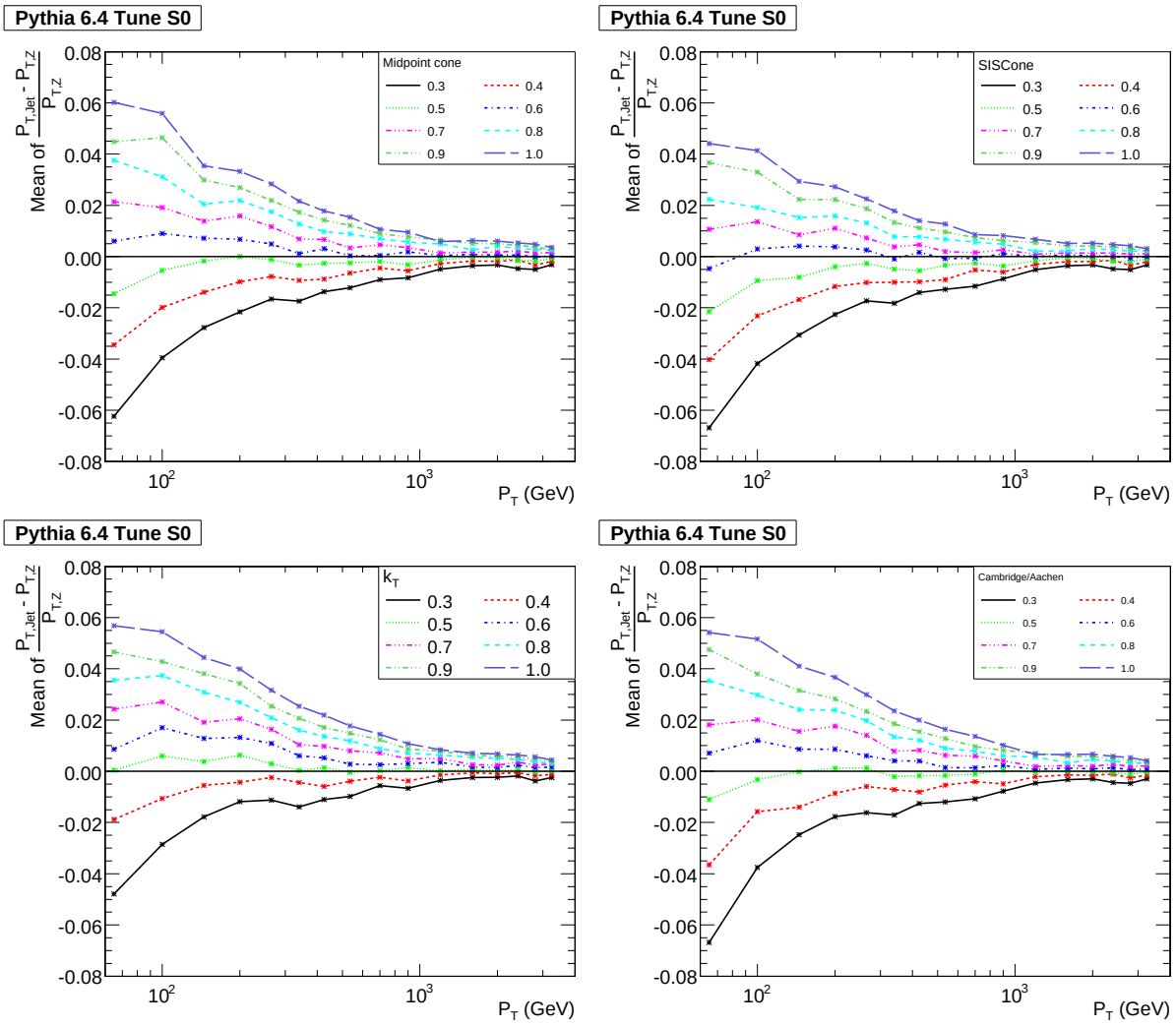

Fig. 48: Same as Fig. 47 but for Pythia tune S0.

Z boson is shown in Fig. 49 for fully hadronized Pythia tune DWT events with and without multiple interactions. Having disabled multiple interactions, the transverse momentum of the jet is systematically underestimated compared to the Z boson. This effect decreases for larger R parameters but remains visible which indicates that the jet algorithms do not accumulate the whole energy of the parton into the jet. So without the MPI even the largest employed jet size hardly suffices to collect all energy to balance the boson p_T. This feature is compensated by acquiring additional energy from the underlying event into the jet. Enabling multiple interactions, the larger jet sizes now overestimate the transverse momentum as shown in Fig. 47.

Concluding, no particular advantage of any jet algorithm can be derived with respect to the jet and Z boson momentum balance. Preferred jet sizes depend heavily on the multiple parton interactions and can only be selected once the underlying event has been determined more precisely at the LHC.

11.4 H ! gg ! jets

In the last section, we evaluate the impact of the jet algorithms and jet sizes on the mass reconstruction of a heavy resonance. More specifically, we look at the process H ! gg ! jets as a "monochromatic" gluon source. In order to reduce to a large degree the effect of the finite Higgs width, on the one hand

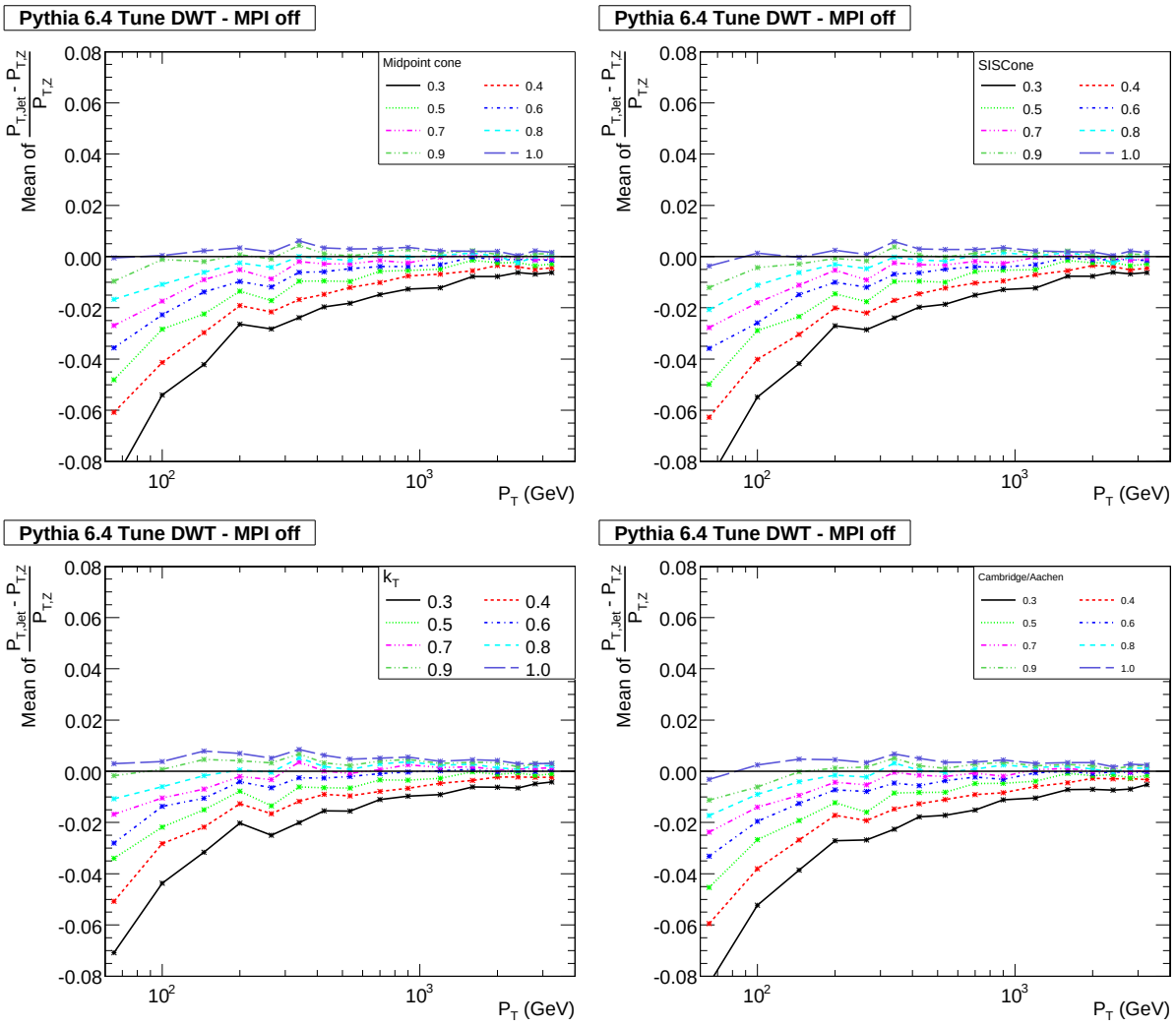

Fig. 49: Same as Fig. 47 but with multiple parton interactions switched off.

side we allow the actual Higgs mass in an event to deviate from the nominal one only by $50\,\mathrm{GeV}$, on the other hand, when comparing the mass reconstructed from the two gluon jets, the remaining difference to the nominal mass is compensated for. The two jets are required to be the leading jets in transverse momentum with a separation in rapidity[36] y of $|y_{\mathrm{jet1}} - y_{\mathrm{jet2}}|$ smaller than 1. To avoid potential problems with the gg production channel for large Higgs masses we decided to enable only the weak boson fusion, process numbers 123 and 124 in Pythia, Ref. [85].

Nevertheless it proved to be difficult to define quality observables, since Breit-Wigner as well as Gaussian fits or combinations thereof do not in general well describe the mass distributions for all jet sizes. At small R up to 0.4 the substructure of gluon jets is resolved instead of features of the hard process. At intermediate resolutions a small mass peak starts to reappear leading to asymmetric distributions which are especially awkward to deal with. The same problems arise in the reconstruction of a Z^0 mass which is investigated in more detail in chapter 10 of these proceedings. For comparison we use a similar approach here and look for the smallest mass window containing 25% of all events. As reconstructed mass value we simply chose the median, which may lie outside the location of the smallest window, since we primarily consider the width as quality measure and not the obtained mass. Figure 50

[36] $y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z}$

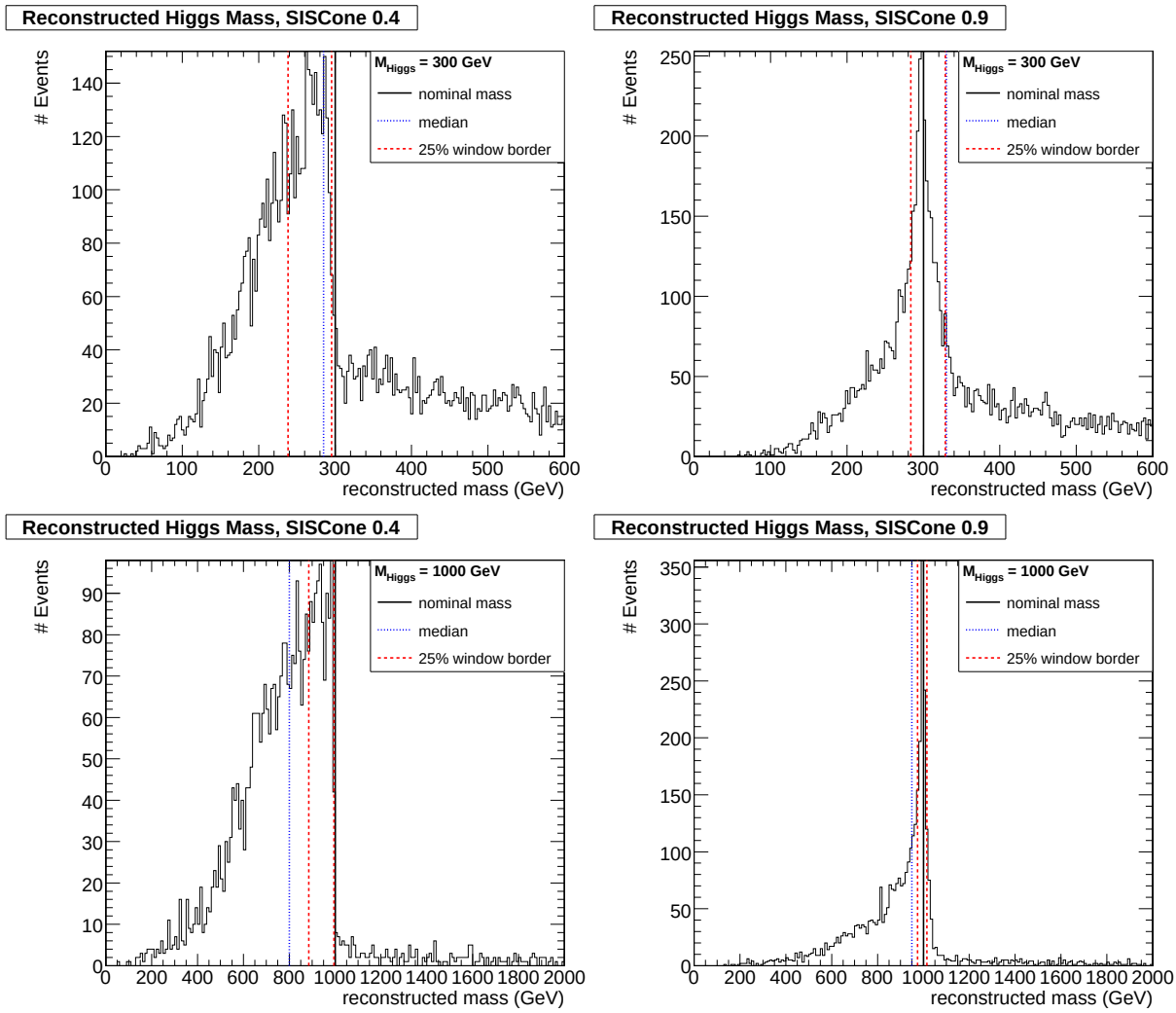

Fig. 50: Reconstructed Higgs mass distributions for the SISCone algorithm with cone sizes 0.4 (left) and 0.9 (right) for the two nominal Higgs masses of 300 (upper) and $1000\,\mathrm{GeV}$ (lower). The full black line indicates the nominal mass, the dashed red lines show the location of the determined minimal mass window and the dotted blue line corresponds to the median.

displays as example the determined mass windows and medians for the SISCone algorithm with cone sizes 0.4 and 0.9 for the two nominal Higgs masses of 300 and $1000\,\mathrm{GeV}$.

In Figure 51 the reconstructed Higgs mass and width, defined as median and the minimal mass window, is shown for all four jet algorithms versus the jet size for the four nominal resonance masses of 300, 500, 700 and $1000\,\mathrm{GeV}$. Obviously, the median systematically underestimates the nominal mass for larger Higgs masses.

Finally, in Fig. 52 the derived minimal mass window sizes are presented in dependence of the jet size R for all jet algorithms and four nominal masses of the Higgs boson. Systematically, the cone type algorithms perform somewhat better than the sequential recombination ones in the sense that they lead to smaller reconstructed widths.

Conclusions

As already observed previously, hadronization corrections for inclusive jets, especially at low transverse momenta, are smaller for jet algorithms of the sequential recombination type (k_T, Cambridge/Aachen).

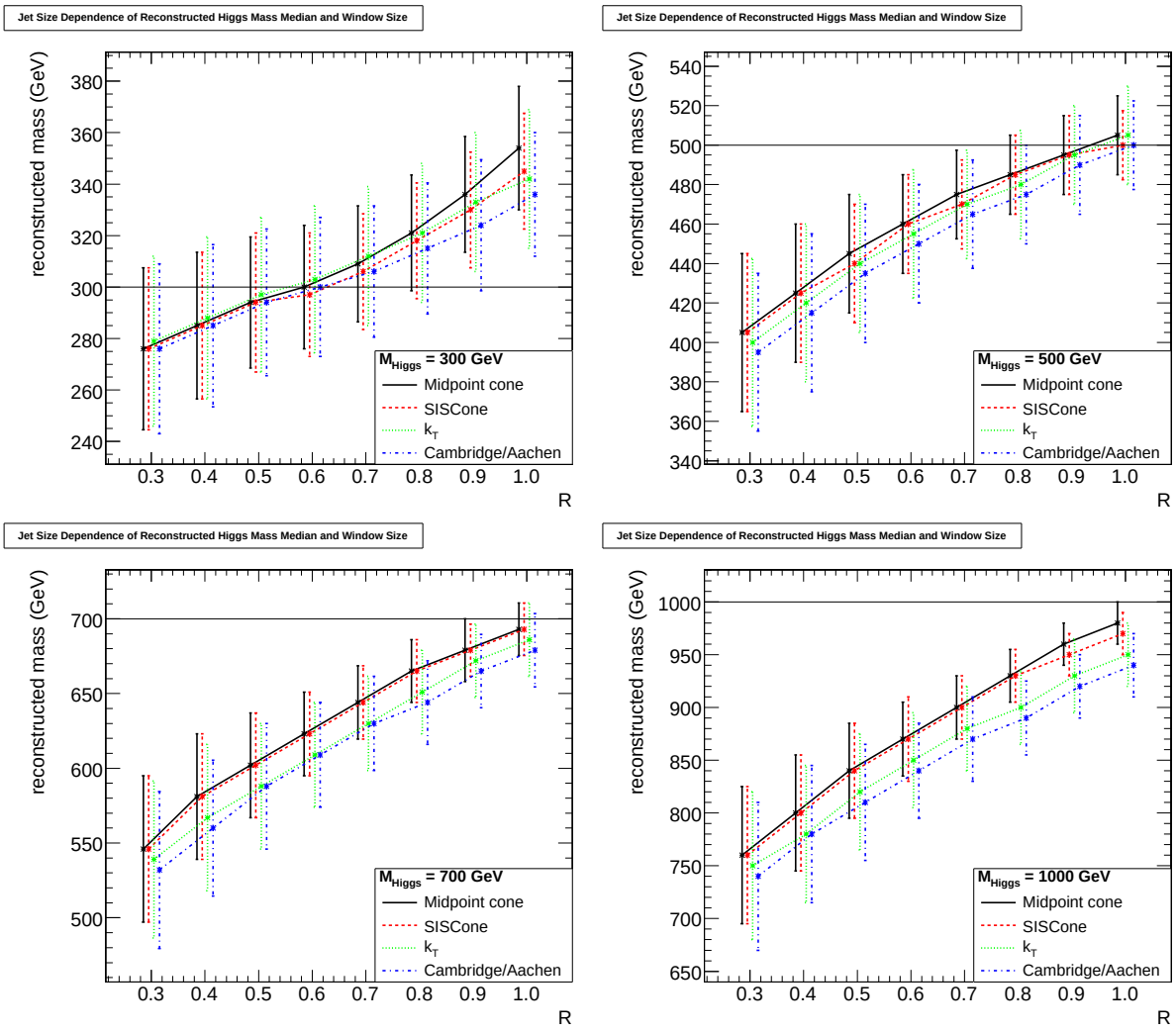

Fig. 51: Reconstructed Higgs mass and width when defined as median and minimal mass window containing 25% of all events versus the jet size. For better visibility the points have been slightly displaced in R for the different jet algorithms.

For the purpose of inclusive jet spectra, however, one is predominantly interested in the newly accessible regime of transverse momenta above 600 GeV or just below. In addition, in the complete correction a partial cancellation occurs of hadronization effects and contributions from the underlying event where no algorithm showed a distinctly better performance than the others. So provided the current extrapolations of the underlying event, one of the largest unknowns, are roughly comparable to what will be measured, all algorithms are equally well suited. For the analysis of the Z plus jets momentum balance no particular advantage of any jet algorithm was observed neither. In the case of the characterization of the reconstructed Higgs resonance via the median and the minimal mass window containing 25% of the events as proposed in chapter 10, the cone type algorithms (Midpoint cone, SISCone) exhibit smaller widths.

Concerning jet sizes, the inclusive jets analysis and the Z plus jet balance prefer medium jet sizes R of 0:4 to 0:8, i.e. somewhat smaller than the habitual value of R 1 before. This is in agreement with the expected higher jet multiplicities and larger underlying event contributions at LHC energies which require a higher jet resolution power. For the reconstruction of the Higgs resonance, especially here from two gluon jets, larger jet sizes R of 0:8 or 0:9 are required. For jet sizes below 0:5 one resolves the substructure of the gluon jets instead of recombining all decay products of the resonance.

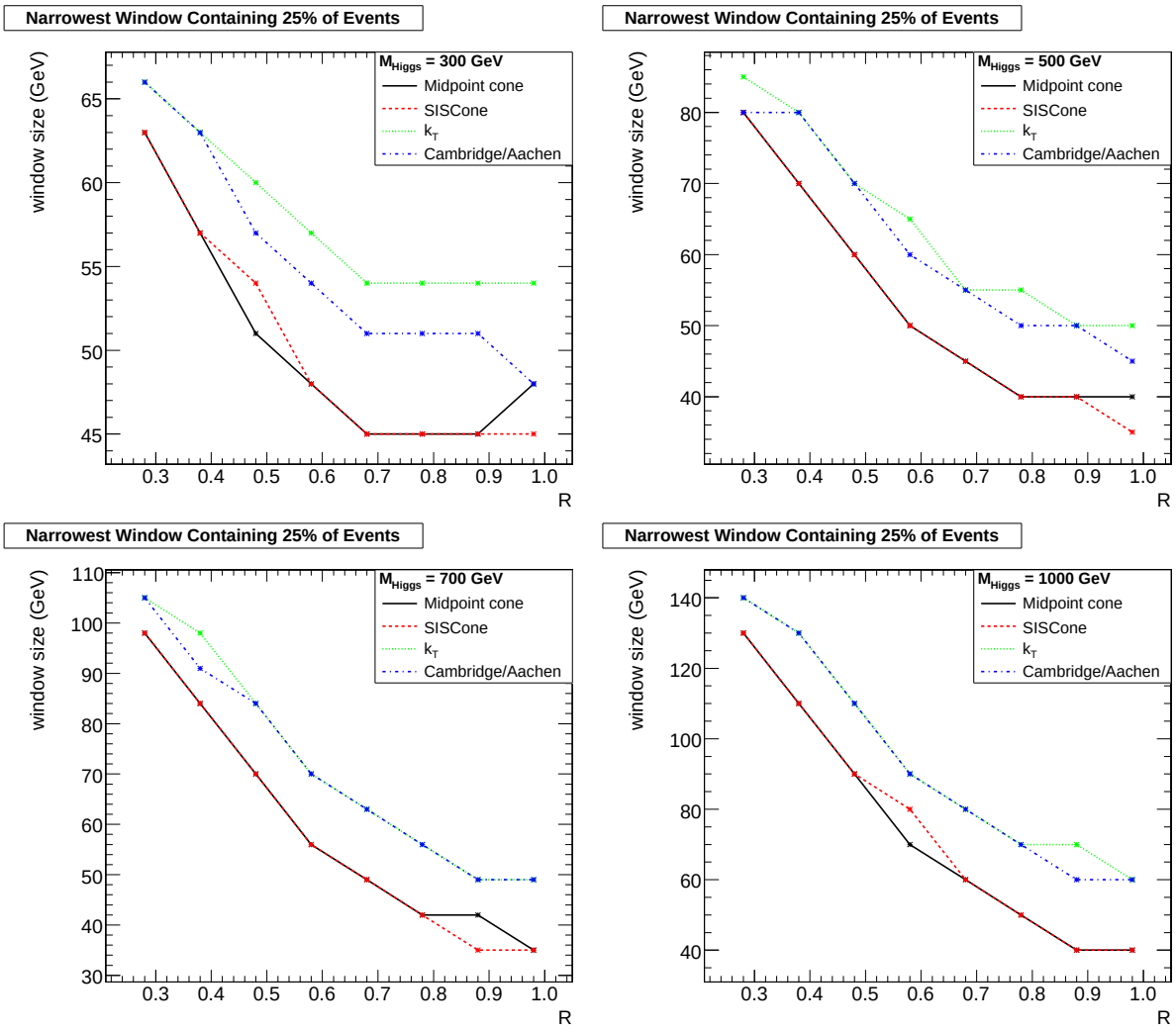

Fig. 52: Minimal mass window sizes in dependence of the jet size R for all jet algorithms and four nominal masses of the Higgs boson.

Concluding, the suitability of the considered four jet algorithms was investigated for three types of analyses and no decisive advantage for a particular one was found within the scope of this study. So apart from the fact that Midpoint cone is not collinear- and infrared-safe and was merely used for comparison, further investigations have to be performed with respect to experimental aspects. We have shown that especially the underlying event can be expected to have a significant impact on the presented analyses.

Acknowledgements

We thank Florian Bechtel, Ulrich Felzmann, Andreas Oehler, Günter Quast, Gavin Salam, Peter Skands and Markus Wobisch for discussions. We are grateful to our home institutions, the Graduiertenkolleg Hochenergiephysik und Teilchenastrophysik of the Deutsche Forschungsgemeinschaft and the Bundesministerium für Bildung und Forschung for their support.

12 A STUDY OF JET ALGORITHMS USING THE SPARTYJET TOOL [37]

12.1 Introduction

Almost all LHC physics channels will contain jets in the final state. For this reason, jet clustering algorithms deserve a great deal of attention. Even though hadron collider experiments have reconstructed jets for over 30 years, until recently the precision reached at hadron machines was not sensitive to the differences between the different jet algorithms. In addition, the available computing power often limited the choice of jet algorithms that were practical to use.

With the recent precision measurements from the Tevatron, and in light of the expectations for the LHC, it is worthwhile to re-examine the impact jet algorithms do make at hadron colliders, especially as new algorithms and ideas are being developed. Our aim in this contribution is to provide a systematic study of some characteristics of representative jet clustering algorithms and parameters, using as an input one of the closest analogues an experiment can provide to four-vectors, the ATLAS topological clusters. These are calorimeter clusters already calibrated for detector measurement effects, to effectively the hadron level. These topoclusters are passed to the clustering algorithms by the SpartyJet [146] tool, an interface to the major clustering algorithms that allows easy change and control over relevant parameters.

12.2 Algorithms considered

Jet clustering algorithms can be divided into two main classes: cones and iterative recombination (as for example the k_T algorithm). Historically, in hadron colliders, primarily cone algorithms have been used, being the only algorithm fast enough for use at trigger level, and for fear of large systematic effects in busy multi-jet environments from recombination algorithms. Fast implementations of the k_T algorithm [144], as well as the first papers performing precision measurements with it [155, 157] call for a detailed comparison of the k_T algorithm with cone-based ones.

Many implementations of cone algorithms have been developed over the years (and the experiments). Many of them have been shown to suffer from infrared safety issues, i.e. the results of the algorithm can change if very soft particles, that do not affect the overall topology of the event, are added or subtracted. Unfortunately, algorithms that have long been the default for large experiments, such as JetClu for CDF and the Atlas cone for Atlas, belong to this category. Other algorithms, such as Midpoint [127, 129] are stable under infrared correction for most (but still not all) cases. But, since they start clustering jets around energy depositions larger than a given value (seed threshold), the outcome will depend in principle on the value of this threshold. The manner in which this will affect clustering under real experimental conditions is one of the questions we will attempt to address in this study. Finally, a seedless infrared-safe cone algorithm has recently emerged [128], providing most of the desirable features a cone algorithm needs from the theoretical point of view and a similar ease of use as previous cone algorithms. Its adoption by the experimental community has been slow due to the lack of a comprehensive comparison with more traditional approaches. Most of the studies presented in the following sections will involve comparisons between the k_T algorithm (for the two different cone sizes of 0.4 and 0.6), the legacy Atlas cone and the Midpoint cone algorithm (for a cone size of 0.4), the Cambridge/Aachen algorithm (similar to the k_T algorithm, but only using the distance between clusters and not their energy) and the seedless infrared cone algorithm (SISCone; cone size of 0.4). Throughout this contribution, these algorithms will be identified by the same color, i.e. black for Kt04, red for Kt06, green for the Atlas cone(04), dark blue for SISCone(04), pink for MidPoint(04) and light blue for Cambridge/Aachen.

12.3 Datasets

To perform our studies, we have used the Monte Carlo datasets produced in the context of the Atlas CSC notes exercise. In particular, we are interested in the behavior of jet algorithms in a multi-jet

[37] Contributed by: M. Campanelli, K. Geerlings, J. Huston

environment and in the endcap region where small changes in cluster position can result in large rapidity differences. It was therefore natural to use samples from W +jets and VBF Higgs channels. The former were generated with ALPGEN [29], (interfaced to Herwig), for the case of a W boson decaying into a muon and a neutrino, produced in association with a number of partons ranging from 0 to 5; the latter are Herwig [163] samples, with a Higgs (M_H = 120 GeV) decaying into tau pairs, with each of the taus decaying into an electron or a muon and neutrinos.

Unless otherwise specified, the different algorithms are run on the same datasets; therefore, the results obtained are not statistically independent, and even small differences can be significant. Jets reconstructed with the jet axis closer than R = 0:4 with respect to the closest lepton (either from W decay or a from H !) are discarded, in to avoid biasing the jet reconstruction performances either by inclusion of those leptons in the jet, or by calling jet a lepton or a tau decay product altogether.

12.4 Jet Multiplicity

The first variable we examined is the jet multiplicity for events with a leptonically decaying W and a number of partons varying from 0 to 5.

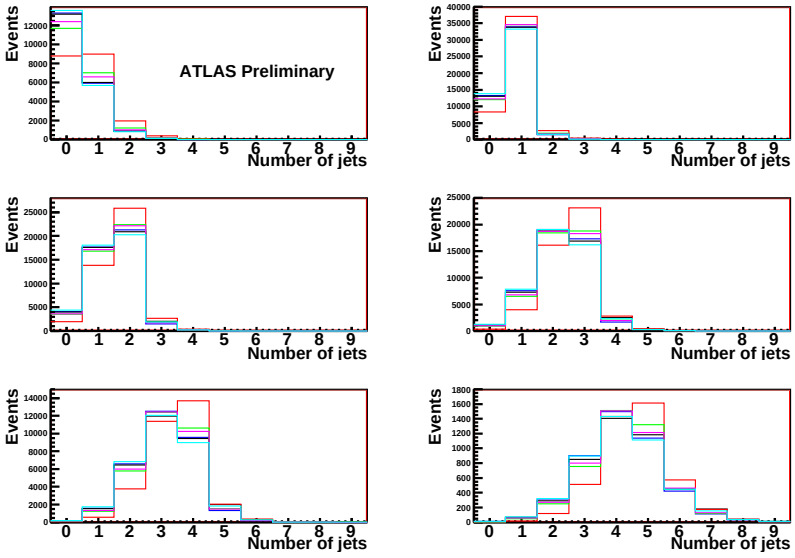

Fig. 53: Number of reconstructed jets for W + n partons Monte Carlo, with the number of partons increasing (from 0 to 5) as the plot order.

The reconstructed number of jets with p_T > 20 GeV for the various algorithms (with color code as in the end of the "Algorithms considered" session) is shown in Figure 53, where each plot represents a different number of generated partons. To understand the trends somewhat better, Figure 54 shows the difference between the mean number of reconstructed jets and the number of partons, while Figure 55 shows the RMS of this distribution. As expected, the distribution of the number of reconstructed jets broadens as the number of partons increases, both at reconstructed and generator level. Since only jets passing the 20 GeV p_T cut are included, it is understandable that the multiplicity is higher for the Kt06 than for the Kt04 algorithm. This is true as well for large jet multiplicities, where the effect of the smaller available phase space for the larger jet size is not relevant for the multiplicities considered. On the other hand, SISCone tends to reconstruct a smaller number of jets than the other algorithms.

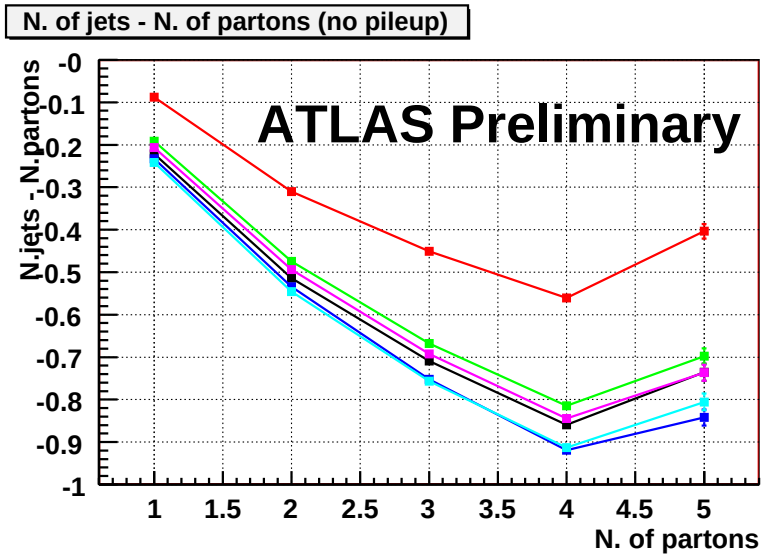

Fig. 54: Difference between the number of reconstructed jets and the number of reconstructed partons vs this latter quantity, for W + n partons Monte Carlo

12.5 Matching efficiency

One of the most important characteristics of a jet algorithm is the ability to correctly find, after detector effects, jet directions as close as possible to the generated ones. Since a parton does not have a well-defined physical meaning, we stress again here that all comparisons between generated and reconstructed quantities are done with jets reconstructed from stable particles at the hadron level, using the same algorithm as at detector level. Matching efficiencies are defined as the number of hadron level jets in a given p_T or bin that have a reconstructed jet within a given R cut.

The R distribution between the generated and the closest reconstructed jet is shown on the left side of Figure 56 for the four algorithms studied in the previous section, for a dataset of W + 2 partons Monte Carlo. We see that the Kt06 algorithm has the largest mean value of R, and therefore the worst matching, probably because of fluctuations far from the core of the jet. The same distribution for jets in VBF Higgs events shows a smaller R for all clustering algorithms, showing that, in general, matching between generated and reconstructed jets is better in VBF Higgs than in W + parton events. To better understand the properties of matching, we will study its behaviour as a function of jet kinematics. Figure 57 shows the efficiency for various p_T bins and for a range of R cuts for the algorithms considered in the previous session, on a dataset of W + 2 partons Monte Carlo. For all algorithms, an efficiency higher than 95% (in red) is reached at high jet momenta even for quite tight R cuts, while small differences among algorithms emerge at lower jet momenta. If we take the slices of this 2d plot corresponding to the cuts $R < 0.3$ and $R < 0.4$, respectively, we obtain the results in Figure 58.

These plots were produced from a W + 2 partons dataset, but all other datasets exhibit a similar behaviour, even for large parton multiplicities (see Figure 59 for W + 5 partons). SISCone does a very good job under these difficult situations, and fears of the k_T algorithm picking up too much underlying event seem justified only in the case of large jet size. The matching efficiency as a function of the jet for VBF Higgs events is shown in Figure 60. It is interesting to note how the endcap region, with $2 < |\eta| < 3$, equipped with a Liquid Argon calorimeter with good pointing capabilities, is on average more efficient than the barrel and the very forward endcap. The different distribution, as well as the harder spectrum, may explain why jets from VBF Higgs events have a better matching efficiency than those from W + parton events.

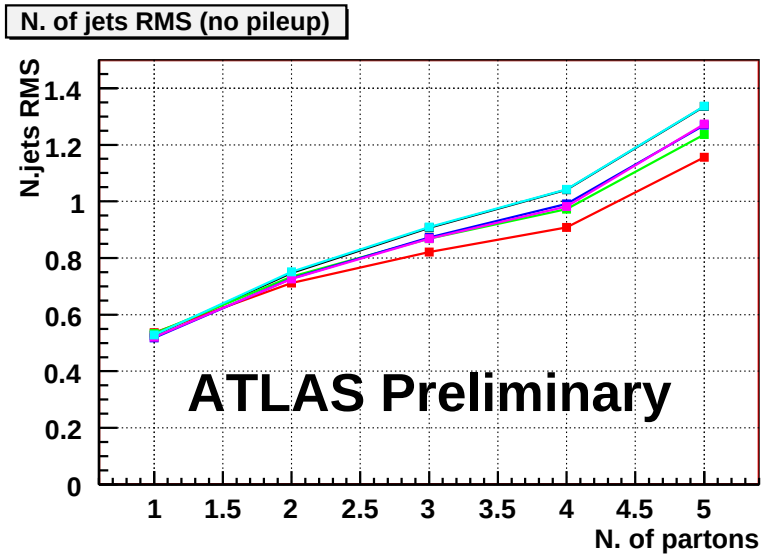

Fig. 55: RMS of the distribution of the number of reconstructed jets, as a function of the number of generated partons for W + n partons Monte Carlo

12.6 Seed threshold and split/merge parameter

An obvious argument in favour of a seedless clustering algorithm is that the seed threshold is in principle an arbitrary parameter, and the dependence of jet reconstruction on arbitrary parameters should be avoided as much as possible. On the other hand, from the experimental point of view, any seed below the calorimeter noise-suppression cut should be equivalent, and no dependence on seed threshold should be observed for reasonable values of this parameter. To test this hypothesis, we looked at W + 5 parton events, with very low jet p_T threshold (10 GeV). The number of jets reconstructed with the MidPoint algorithm with seed thresholds of 0.1, 1 and 2 GeV is shown in Figure 61. We see that no significant difference is found for the different seed values, so the claim that reasonable seed values lead to similar results seems justified, at least for inclusive distributions of the type examined here. This fact does not reduce the merits of the seedless algorithm.

To address the issue of the dependence of jet clustering on the split/merge parameter, we clustered W + 2 parton events using the Atlas cone and SISCone algorithms with this parameter set to 0.5, 0.625 and 0.75. Large differences are observed, as seen for example for the SISCone case in Figure 62. Perhaps a systematic study to fine tune this parameter could be useful. We noticed that, out of the three options considered here, the best value of this parameter is algorithm-dependent, and is in fact 0.5 for the Atlas cone and 0.75 for SISCone, which are presently the default values for these algorithms.

12.7 Energy reconstruction

Even after compensation for the different calorimeter response to electromagnetic and hadronic showers, Atlas topological clusters currently underestimate the total visible energy by about 5% due to noise-suppression thresholds, particle losses, inefficiencies etc. This effect results in a systematically higher hadron-level energy with respect to the detector-level one, and is visible as a function of jet p_T and for W + 2 parton events in Figures 63 and 64. As expected, this bias is larger for low-energy jets where the relative importance of low-energy clusters (more prone to losses etc.) is higher. Also, the behavior in regions close to the badly-instrumented parts of the detector differs considerably between the various algorithms.

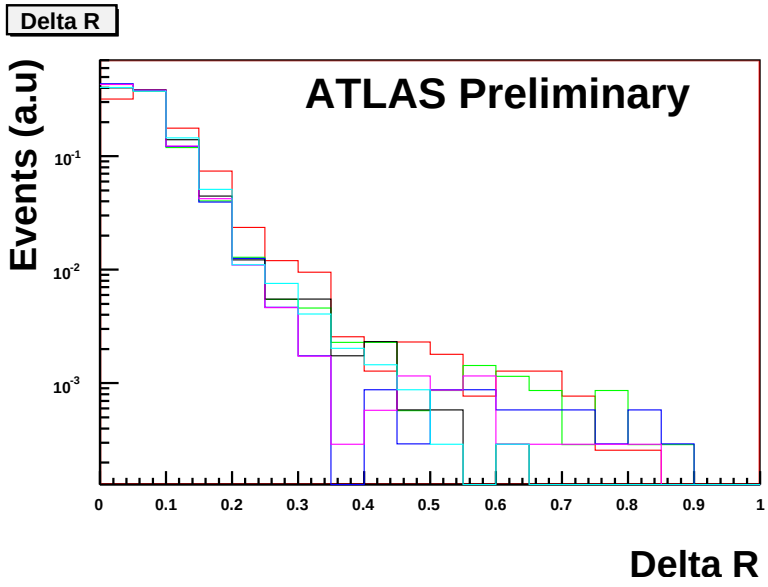

Fig. 56: Distribution of R between generated and reconstructed jets, for W + 2 partons Monte Carlo.

12.8 Cross sections

The study of W + n jet cross sections, i.e. the p_T distributions of the most energetic jet for various jet multiplicities, allows a study of the effect of jet clustering on energy distributions as well as on jet multiplicities. To select events with W boson decays into a muon and a neutrino, we require the presence in the event of a muon of at least 25 GeV in the acceptance region j j< 2:4 and missing transverse energy of at least 25 GeV. We accept jets if they have transverse momentum larger than 15 GeV, j j< 5 and R > 0:4 with respect to the muon. Events are classified according to the number of reconstructed jets. We studied the distribution of the p_T of the leading jet for W + n parton events. For space reasons, we show here only those obtained with the W + 2 parton sample, but all other distribution show similar characteristics. The reconstructed spectra of the leading jet are shown in Figures 65. We see that the different behavior observed for the jets reconstructed with the KT06 algorithm is mainly due to the very soft region. Since, with this jet size, there is the tendency of reconstructing a larger average number of jets, there are fewer events placed in the W + 1 jet category (the red histogram is always below the others for the first plot), and more in the cases where the reconstructed multiplicity is higher than the generated one (all plots from the third on). However, looking at the p_T spectra, we realize that this effect is mainly present for events with a soft leading jet, while for hard events (i.e. for higher p_T of the leading jets) all distributions tend to converge.

12.9 Pileup

We know that in the first phases of LHC operation, the proton density in the bunches will be already high enough for the events to exhibit non-negligible pileup. No study of clustering algorithms would be complete without an assessment on the behaviour under realistic running conditions. Assuming that pileup can be added linearly to the event, we overlapped three minimum-bias events to the W + partons and Higgs VBF events considered in the previous sections, and examined how the quantities considered above are modified for the various algorithms.

The first property studied here is the jet multiplicity. We see that the distribution of the number of jets for the W + partons sample (Fig. 66) is modified. The behavior of the various algorithms can be seen in the mean value and RMS of the reconstructed multiplicity as a function of the number of partons (Figures 67 and 68). A direct comparison between the no-pileup and pileup case is made in Figure 69,

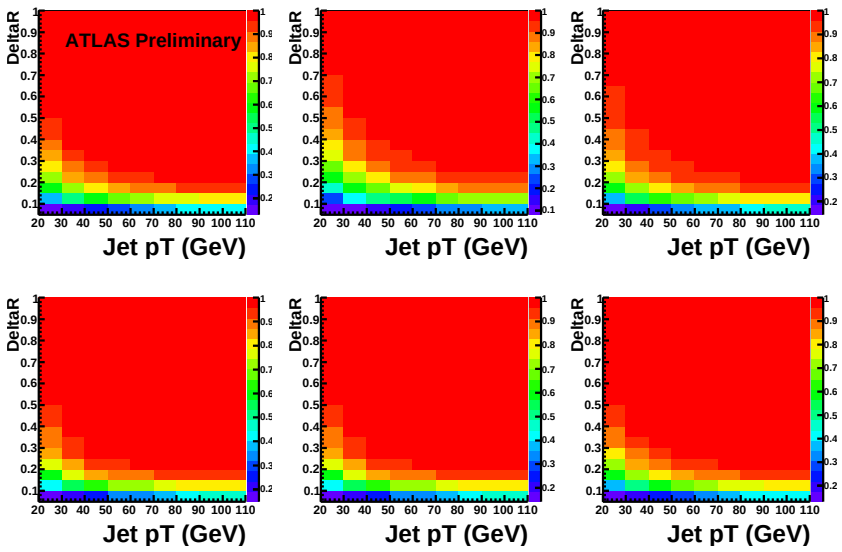

Fig. 57: Matching efficiency as a function of jet p_T and R cut for W + 2 partons.

where we show the average number of reconstructed jets for Higgs VBF events without and with pileup. Kt04 and SISCone are the two algorithms that are less sensitive to the presence of pileup.

In order to study the influence of pileup on the kinematic distributions for the reconstructed jets, Figure 70 shows the ratio of the p_T distributions with and without pileup for each reconstructed jet multiplicity, for W + 2 parton events.

The presence of pileup, leading to a modification of the jet axis direction, also influences the matching efficiency between hadron level and detector level jets. The efficiency as a function of jet p_T and , computed using the same definition as in the previous sections, is shown in Figures 71 and 72. Again, the scale of robustness of the various algorithms to the presence of pileup obtained from the other tests is confirmed.

Finally, we tested the effect of using different algorithms on a simple forward jet selection, aiming at a discrimination of VBF Higgs events from the background. The following cuts were applied to the VBF Higgs and to the W + 2 partons and the W + 3 partons Monte Carlo:

Two jets with $P_T^1 > 40$ GeV and $P_T^2 > 20$ GeV

Both jets have R > 0.4 with respect to tau decay products

$_{1,2} > 4.4$

Invariant mass between the two jets > 700 GeV

No third jet with j j< 3.2 and P > 30

The efficiencies obtained in the three samples for three of the jet algorithms under study here are summarized in Table 12.9.

While the change in efficiency for the Higgs signal is quite marginal, the same cannot be said for the difference in background rejection. Here the algorithms that have proven to be more robust under the influence of pileup exhibit a much better background rejection, and can improve the power of the analysis.

12.10 Conclusions

In this note we have systematically explored the behavior of several jet algorithms, Kt (with different jet sizes, corresponding to the choice of D parameter of 0.4 and 0.6), the Atlas Cone, SISCone, MidPoint

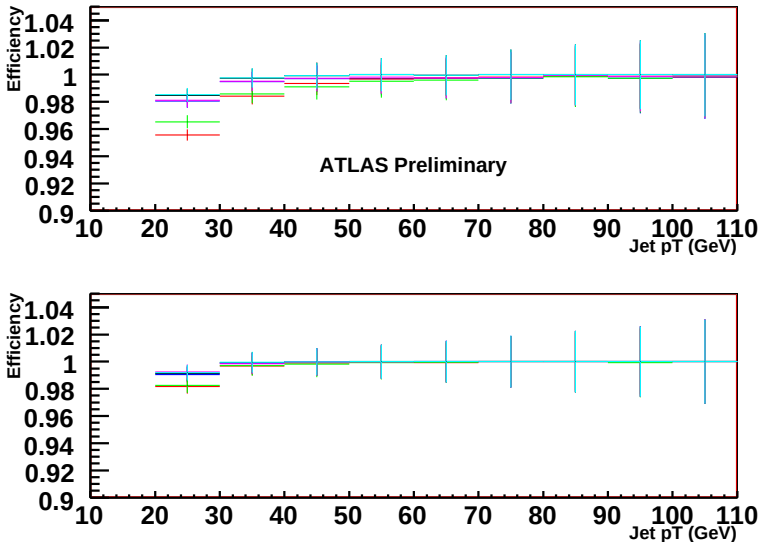

Fig. 58: Matching efficiency as a function of jet p_T for W + 2 partons. The matching requirement is that $\Delta R < 0.3$ (above) and $\Delta R < 0.4$ (below).

Algorithm	VBF Higgs		W + 2p		W + 3p	
Cone 04	15.9	0.4	0.37	0.03	1.17	0.05
KT 04	15.1	0.4	0.17	0.02	0.85	0.04
SISCOne 04	14.2	0.4	0.17	0.02	0.76	0.04

Table 8: Selection efficiency for the forward jet cuts described in the text, for the various algorithms applied to the three Monte Carlo samples of VBF Higgs, W + 2 and W + 3 partons

(all for cone size of 0.4) and Cambridge/Aachen, on several benchmarks with and without the presence of pileup. The comparison of the smaller and larger jet sizes in the k_T algorithm has shown that the use of larger jets deteriorates the resolution in jet direction, and is more vulnerable to the presence of pileup, so should be avoided for the purpose of jet finding, even if it may be more accurate in determining the jet energy.

The comparison of the different algorithms with approximately the same jet size, corresponding to a radius of 0.4, indicates that the k_T and SISCone algorithms have proven to be as good or better than algorithms more traditionally used in hadron colliders.

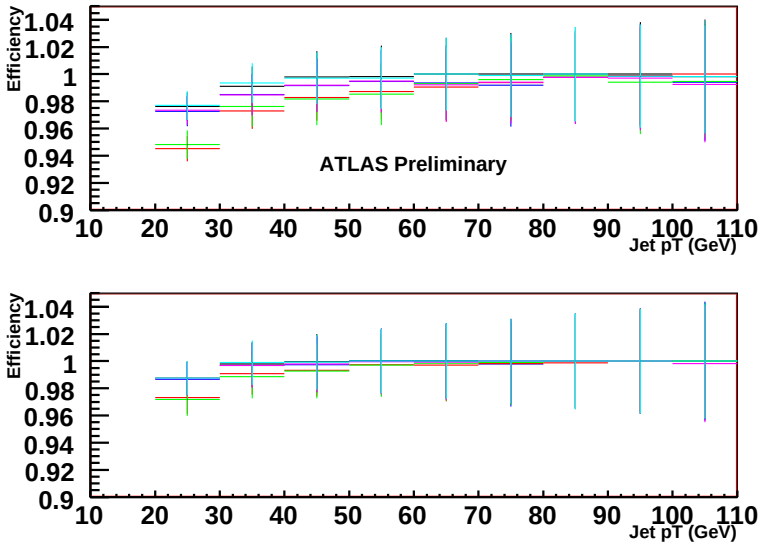

Fig. 59: Matching efficiency for W + 5 parton events. The efficiency is smaller for all algorithms with respect to the W + 1 parton case, but recombination-based algorithms show no worse behavior than the cone-based ones.

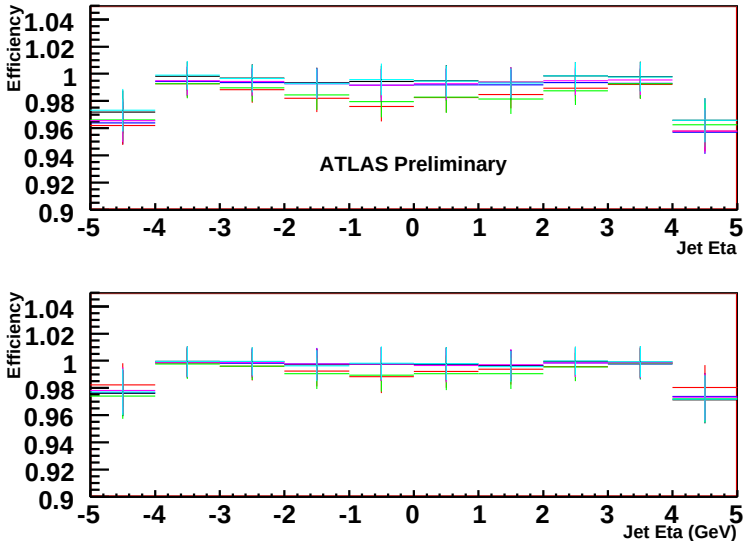

Fig. 60: Matching efficiency for a fixed R cut as a function of jet for VBF Higgs Monte Carlo. The matching requirement is R < 0.3 (above) and R < 0.4 (below).

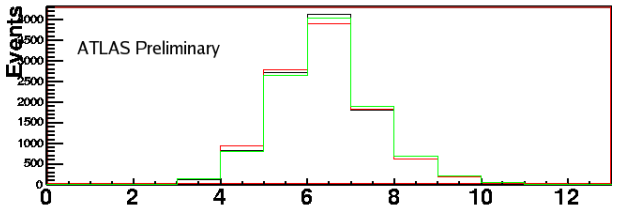

Fig. 61: Number of jets using the Midpoint algorithm with seed threshold of 0.1, 1 and 2 GeV.

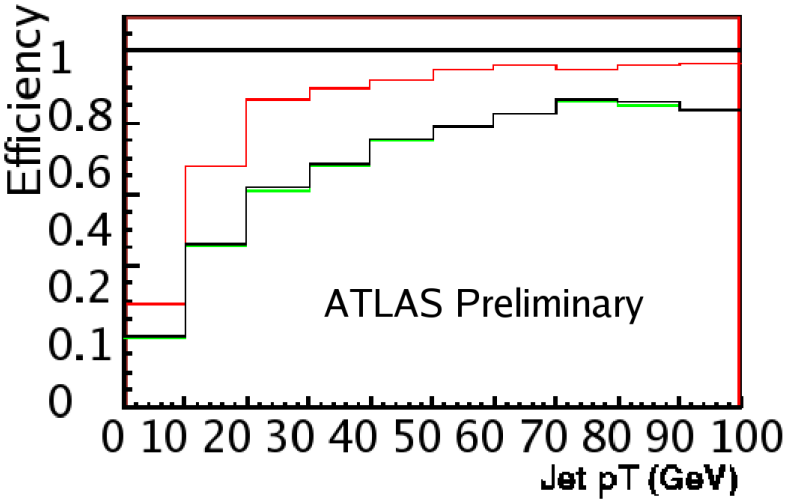

Fig. 62: Matching efficiency for R< 0:1 for SISCone, for values of the split/merge parameter of 0.5, 0.625 and 0.75.

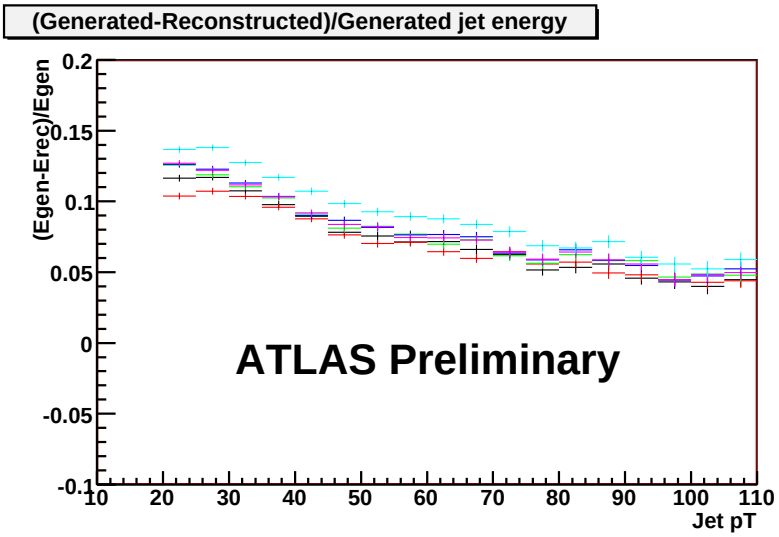

Fig. 63: Difference between hadron and detector level jet p_T, divided by the hadron level one, as a function of jet p_T. The observed bias is due to a small residual correction needed for topoclusters, especially at low energy.

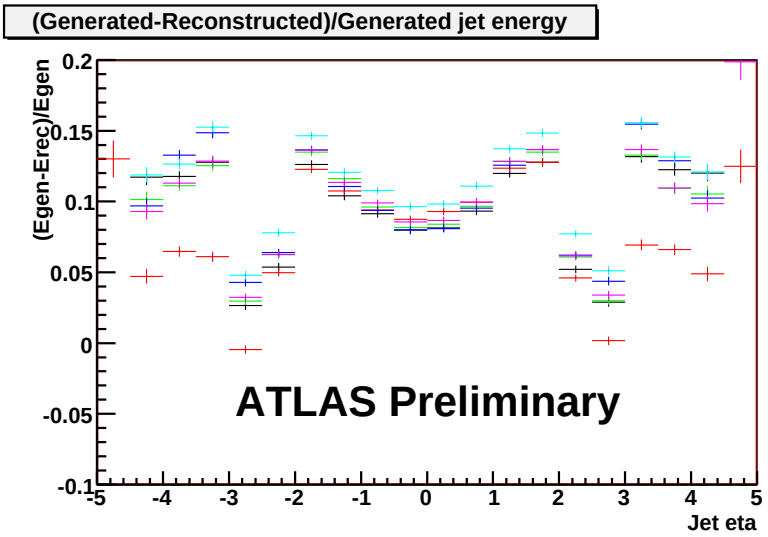

Fig. 64: Same plot as before, as a function of jet eta.

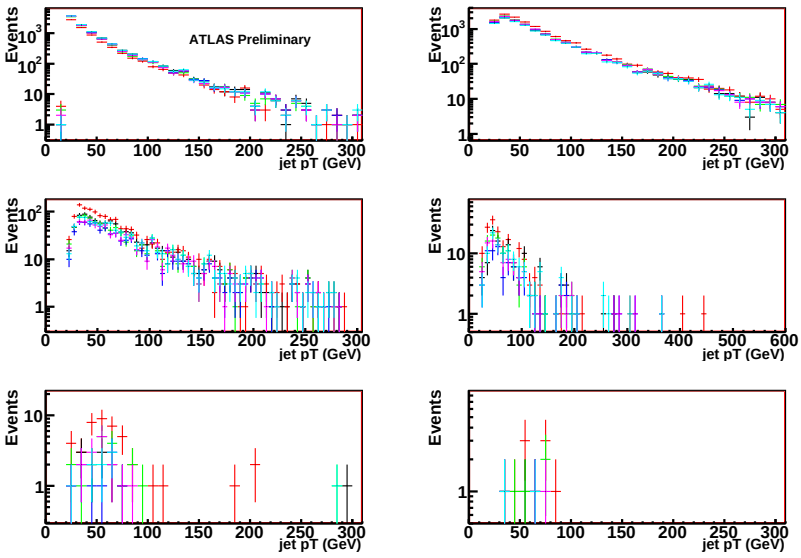

Fig. 65: Reconstructed cross sections for the W + 2 partons sample, as a function of the p_T of the leading jet, for six jet multiplicities (A.U.)

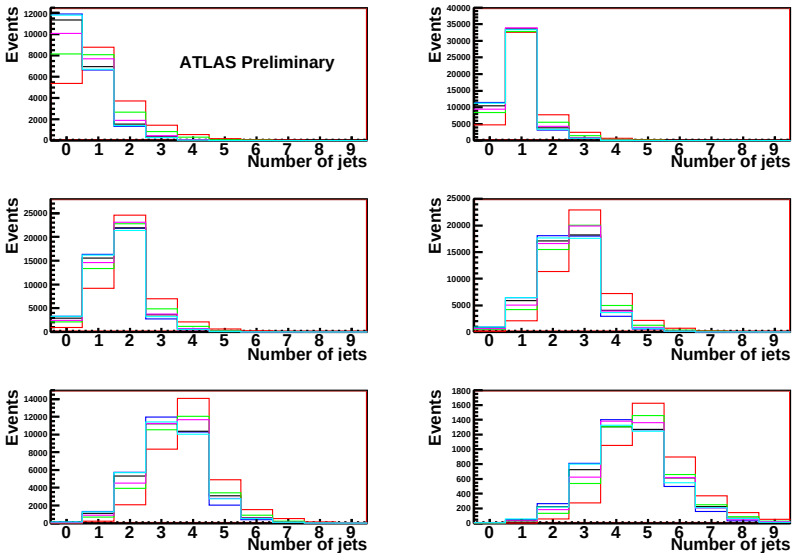

Fig. 66: Number of reconstructed jets for the various W + n parton samples, in the presence of three pileup events.

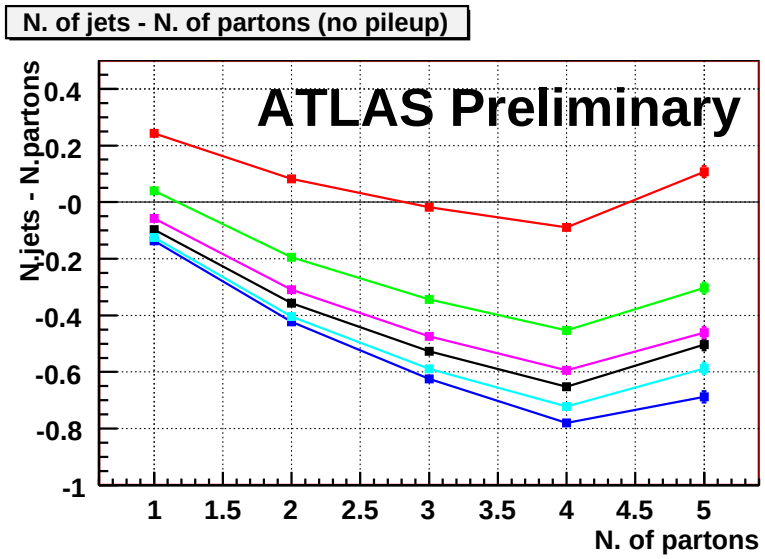

Fig. 67: Difference between number of reconstructed jets and generated partons vs number of partons (with pileup)

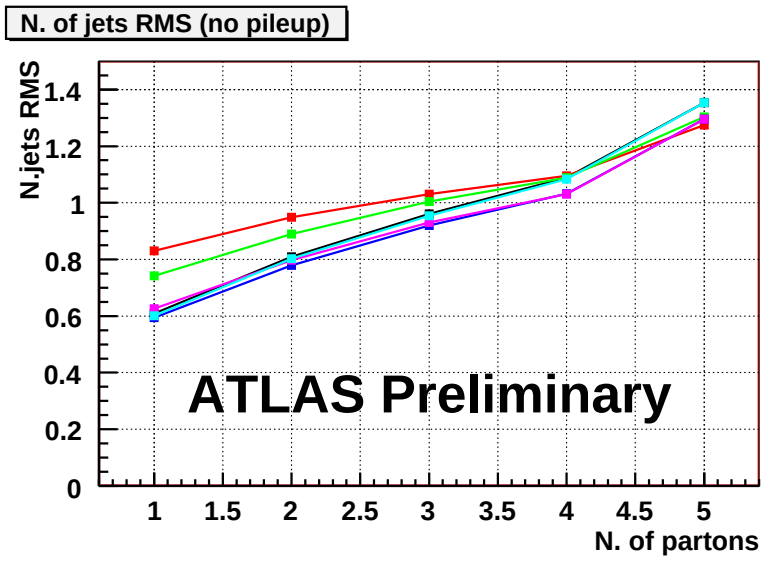

Fig. 68: RMS of the distribution of reconstructed jets vs number of partons (with pileup)

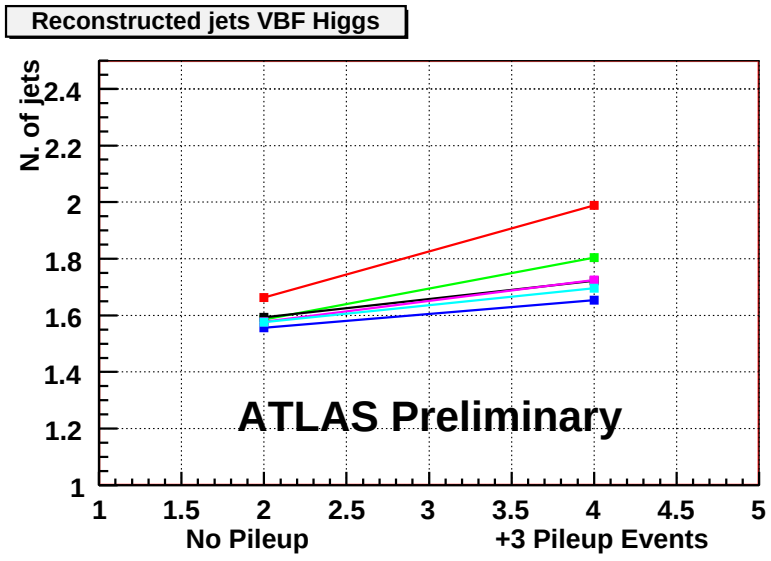

Fig. 69: Number of reconstructed jets for VBF Higgs events with/out pileup

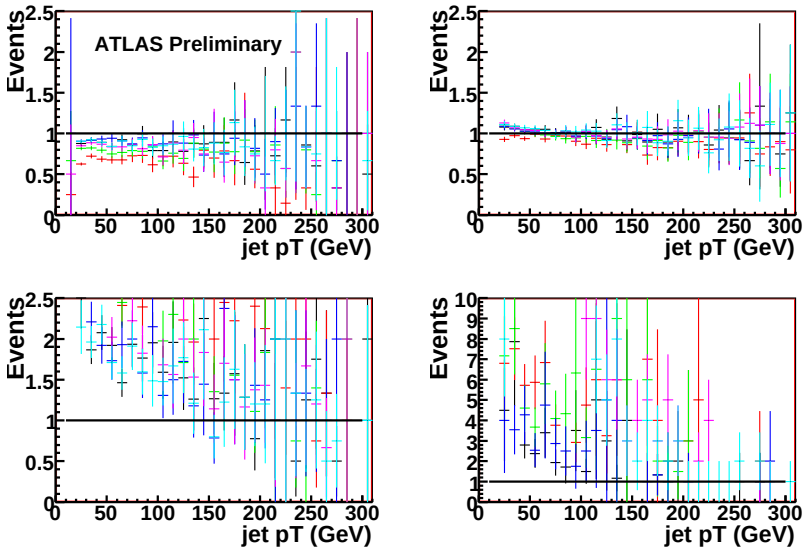

Fig. 70: Ratio between cross section with and without pileup for all algorithms (W + 2 parton sample)

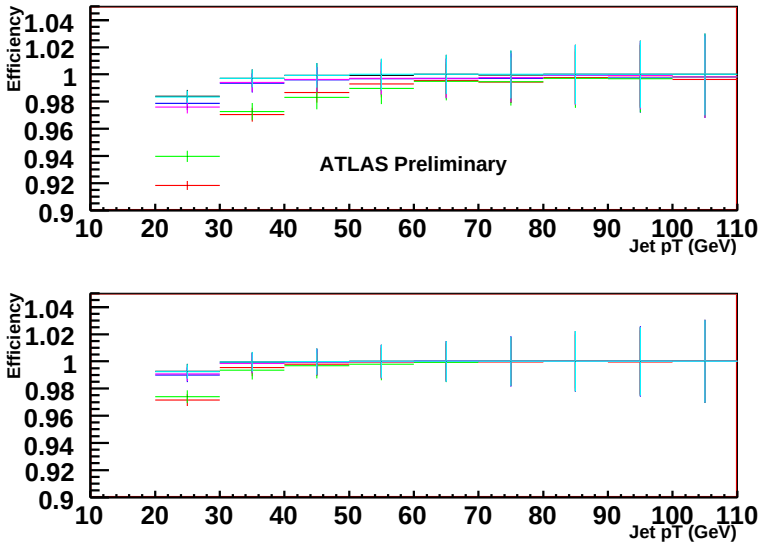

Fig. 71: Efficiency vs jet p_T with pileup (R< 0.3 and 0.4)

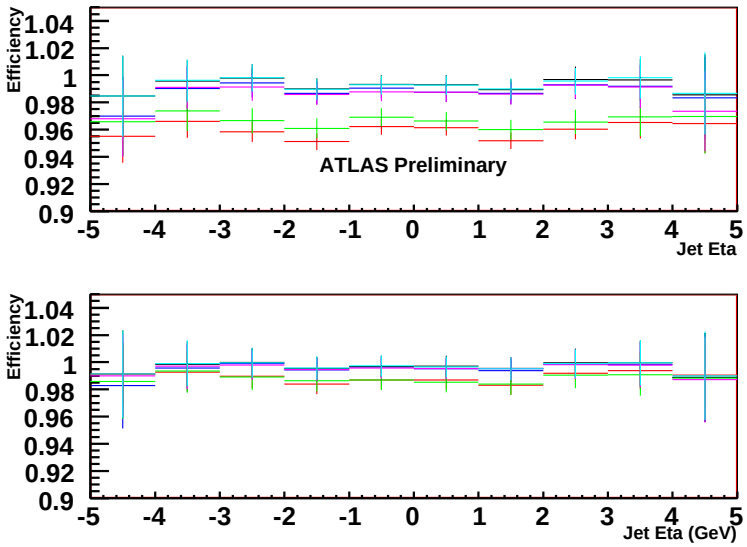

Fig. 72: Efficiency vs jet with pileup (R< 0.3 and 0.4)

References

[1] C. E. Gerber *et al.*,, **TeV4LHC-Top and Electroweak Working Group** Collaboration arXiv:0705.3251 [hep-ph].

[2] C. Carloni Calame, G. Montagna, O. Nicrosini, and M. Treccani, *JHEP* **05** (2005) 019, [hep-ph/0502218].

[3] C. M. C. Calame, G. Montagna, O. Nicrosini, and A. Vicini, *JHEP* **10** (2007) 109, [arXiv:0710.1722 [hep-ph]].

[4] A. Andonov *et al.*, *Comput. Phys. Commun.* **174** (2006) 481–517, [hep-ph/0411186].

[5] D. Bardin *et al.*, *Comput. Phys. Commun.* **177** (2007) 738–756, [hep-ph/0506120].

[6] A. Arbuzov *et al.*, arXiv:0711.0625 [hep-ph], Accepted for publication in Eur. Phys. J. C [DOI: 10.1140/epjc/s10052-008-0531-8].

[7] U. Baur, O. Brein, W. Hollik, C. Schappacher, and D. Wackeroth, *Phys. Rev.* **D65** (2002) 033007, [hep-ph/0108274].

[8] C. Buttar *et al.*, hep-ph/0604120.

[9] M. Awramik, M. Czakon, A. Freitas, and G. Weiglein, *Phys. Rev.* **D69** (2004) 053006, [hep-ph/0311148].

[10] G. Degrassi, P. Gambino, M. Passera, and A. Sirlin, *Phys. Lett.* **B418** (1998) 209–213, [hep-ph/9708311].

[11] A. Ferroglia, G. Ossola, M. Passera, and A. Sirlin, *Phys. Rev.* **D65** (2002) 113002, [hep-ph/0203224].

[12] F. Jegerlehner, *J. Phys.* **G29** (2003) 101–110, [hep-ph/0104304].

[13] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, *Eur. Phys. J.* **C39** (2005) 155–161, [hep-ph/0411040].

[14] J. F. Owens and W.-K. Tung, *Ann. Rev. Nucl. Part. Sci.* **42** (1992) 291–332.

[15] P. M. Nadolsky, *AIP Conf. Proc.* **753** (2005) 158–170, [hep-ph/0412146].

[16] U. Baur, hep-ph/0511064.

[17] G. Altarelli, R. K. Ellis, and G. Martinelli, *Nucl. Phys.* **B157** (1979) 461.

[18] R. Hamberg, W. L. van Neerven, and T. Matsuura, *Nucl. Phys.* **B359** (1991) 343–405.

[19] W. T. Giele, E. W. N. Glover, and D. A. Kosower, *Nucl. Phys.* **B403** (1993) 633–670, [hep-ph/9302225].

[20] J. Campbell and R. K. Ellis, *Phys. Rev.* **D65** (2002) 113007, [hep-ph/0202176].

[21] C. Balazs and C. P. Yuan, *Phys. Rev.* **D56** (1997) 5558–5583, [hep-ph/9704258].

[22] F. Landry, R. Brock, P. M. Nadolsky, and C. P. Yuan, *Phys. Rev.* **D67** (2003) 073016, [hep-ph/0212159].

[23] S. Frixione and B. R. Webber, *JHEP* **06** (2002) 029, [hep-ph/0204244].

[24] S. Frixione, P. Nason, and C. Oleari, *JHEP* **11** (2007) 070, [arXiv:0709.2092 [hep-ph]].

[25] C. Anastasiou, L. J. Dixon, K. Melnikov, and F. Petriello, *Phys. Rev. Lett.* **91** (2003) 182002, [hep-ph/0306192].

[26] C. Anastasiou, L. J. Dixon, K. Melnikov, and F. Petriello, *Phys. Rev.* **D69** (2004) 094008, [hep-ph/0312266].

[27] K. Melnikov and F. Petriello, *Phys. Rev. Lett.* **96** (2006) 231803, [hep-ph/0603182].

[28] K. Melnikov and F. Petriello, *Phys. Rev.* **D74** (2006) 114017, [hep-ph/0609070].

[29] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, *JHEP* **07** (2003) 001, [hep-ph/0206293].

[30] T. Stelzer and W. F. Long, *Comput. Phys. Commun.* **81** (1994) 357–371, [hep-ph/9401258].

[31] F. Maltoni and T. Stelzer, *JHEP* **02** (2003) 027, [hep-ph/0208156].

[32] T. Gleisberg *et al.*, *JHEP* **02** (2004) 056, [hep-ph/0311263].

[33] A. Kanaki and C. G. Papadopoulos, *Comput. Phys. Commun.* **132** (2000) 306–315, [hep-ph/0002082].

[34] C. G. Papadopoulos and M. Worek, *Eur. Phys. J.* **C50** (2007) 843–856, [hep-ph/0512150].

[35] C. G. Papadopoulos and M. Worek, hep-ph/0606320.

[36] V. A. Zykunov, *Phys. Rev.* **D75** (2007) 073019, [hep-ph/0509315].

[37] E. Maina, S. Moretti, and D. A. Ross, *Phys. Lett.* **B593** (2004) 143–150, [hep-ph/0403050].

[38] J. H. Kühn, A. Kulesza, S. Pozzorini, and M. Schulze, *Nucl. Phys.* **B727** (2005) 368–394, [hep-ph/0507178].

[39] J. H. Kühn, A. Kulesza, S. Pozzorini, and M. Schulze, arXiv:0708.0476 [hep-ph].

[40] J. H. Kühn, A. A. Penin, and V. A. Smirnov, *Nucl. Phys. Proc. Suppl.* **89** (2000) 94–99, [hep-ph/0005301].

[41] J. H. Kühn, S. Moch, A. A. Penin, and V. A. Smirnov, *Nucl. Phys.* **B616** (2001) 286–306, [hep-ph/0106298].

[42] B. Jantzen, J. H. Kühn, A. A. Penin, and V. A. Smirnov, *Phys. Rev.* **D72** (2005) 051301, [hep-ph/0504111].

[43] B. Jantzen, J. H. Kühn, A. A. Penin, and V. A. Smirnov, *Nucl. Phys.* **B731** (2005) 188–212, [hep-ph/0509157].

[44] Q.-H. Cao and C. P. Yuan, *Phys. Rev. Lett.* **93** (2004) 042001, [hep-ph/0401026].

[45] G. Balossini, C. M. Carloni Calame, G. Montagna, M. Moretti, O. Nicrosini, F. Piccinini, M. Treccani, and A. Vicini, in preparation.

[46] G. Balossini *et al.*, Prepared for 15th International Workshop on Deep-Inelastic Scattering and Related Subjects (DIS2007), Munich, Germany, 16-20 Apr 2007.

[47] G. Balossini *et al.*, *Acta Phys. Polon.* **B38** (2007) 2347–2355.

[48] Q.-H. Cao and C. P. Yuan, hep-ph/0401171.

[49] M. Dobbs *et al.*, hep-ph/0403100.

[50] J. Pumplin *et al.*, *JHEP* **07** (2002) 012, [hep-ph/0201195].

[51] C. Anastasiou, L. Dixon, K. Melnikov, and F. Petriello, *Phys. Rev.* **D69** (2004) 094008.

[52] C. M. Carloni Calame, G. Montagna, O. Nicrosini, and M. Treccani, *Phys. Rev.* **D69** (2004) 037301.

[53] C. M. Carloni Calame, G. Montagna, O. Nicrosini, and M. Treccani, *JHEP* **0505** (2005) 019.

[54] C. M. Carloni Calame, G. Montagna, O. Nicrosini, and A. Vicini, *JHEP* **0612** (2006) 016.

[55] C. M. Carloni Calame, G. Montagna, O. Nicrosini, and A. Vicini, *JHEP* **10** (2007) 109.

[56] E. Barberio, B. van Eijk, and Z. Wąs, *Comput. Phys. Commun.* **66** (1991) 115.

[57] E. Barberio and Z. Wąs, *Comput. Phys. Commun.* **79** (1994) 291.

[58] P. Golonka and Z. Wąs, *Eur. Phys. J.* **C45** (2006) 97.

[59] A. Denner, B. Jantzen, and S. Pozzorini, *Nucl. Phys.* **B761** (2007) 1–62, [hep-ph/0608326].

[60] J. H. Kühn, A. Kulesza, S. Pozzorini, and M. Schulze, *Phys. Lett.* **B609** (2005) 277–285, [hep-ph/0408308].

[61] J. H. Kühn, A. Kulesza, S. Pozzorini, and M. Schulze, *JHEP* **03** (2006) 059, [hep-ph/0508253].

[62] J. H. Kühn, A. Kulesza, S. Pozzorini, and M. Schulze, *Phys. Lett.* **B651** (2007) 160–165, [hep-ph/0703283].

[63] W. Hollik, T. Kasprzik, and B. A. Kniehl, *Nucl. Phys.* **B790** (2008) 138–159, [arXiv:0707.2553 [hep-ph]].

[64] R. K. Ellis, G. Martinelli, and R. Petronzio, *Nucl. Phys.* **B211** (1983) 106.

[65] P. B. Arnold and M. H. Reno, *Nucl. Phys.* **B319** (1989) 37.

[66] R. J. Gonsalves, J. Pawlowski, and C.-F. Wai, *Phys. Rev.* **D40** (1989) 2245.

[67] J. Campbell, R. K. Ellis, and D. L. Rainwater, *Phys. Rev.* **D68** (2003) 094021, [hep-ph/0308195].

[68] E. W. N. Glover and A. G. Morgan, *Z. Phys.* **C62** (1994) 311–322.

[69] D. Buskulic *et al.*,, **ALEPH** Collaboration *Z. Phys.* **C69** (1996) 365–378.

[70] A. Denner and S. Pozzorini, *Eur. Phys. J.* **C18** (2001) 461–480, [hep-ph/0010201].

[71] A. Denner, M. Melles, and S. Pozzorini, *Nucl. Phys.* **B662** (2003) 299–333, [hep-ph/0301241].

[72] M. Melles, *Phys. Rept.* **375** (2003) 219–326, [hep-ph/0104232].

[73] S. Dittmaier, *Nucl. Phys.* **B565** (2000) 69–122, [hep-ph/9904440].

[74] S. Catani and M. H. Seymour, *Nucl. Phys.* **B485** (1997) 291–419, [hep-ph/9605323].

[75] S. Catani, S. Dittmaier, M. H. Seymour, and Z. Trocsanyi, *Nucl. Phys.* **B627** (2002) 189–265, [hep-ph/0201036].

[76] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, *Phys. Lett.* **B531** (2002) 216–224, [hep-ph/0201127].

[77] M. Roth and S. Weinzierl, *Phys. Lett.* **B590** (2004) 190–198, [hep-ph/0403200].

[78] Y. I. Azimov, Y. L. Dokshitzer, V. A. Khoze, and S. I. Troyan, *Z. Phys.* **C27** (1985) 65–72.

[79] R. Field and R. C. Group,, **CDF** Collaboration hep-ph/0510198.

[80] V. L. Korotkikh and A. M. Snigirev, *Phys. Lett.* **B594** (2004) 171–176, [hep-ph/0404155].

[81] D. Treleani, *Phys. Rev.* **D76** (2007) 076006, [arXiv:0708.2603 [hep-ph]].

[82] P. Skands,, "Some interesting min-bias distributions for early LHC runs." See http://home.fnal.gov/ skands/leshouches-plots/.

[83] S. Alekhin *et al.*, hep-ph/0601012.

[84] M. G. Albrow *et al.*,, **TeV4LHC QCD Working Group** Collaboration hep-ph/0610012.

[85] T. Sjöstrand, S. Mrenna, and P. Skands, *JHEP* **05** (2006) 026, [hep-ph/0603175].

[86] P. Skands and D. Wicke, *Eur. Phys. J.* **C52** (2007) 133–140, [hep-ph/0703081].

[87] A. Moraes *et al.*, SN-ATLAS-2006-057.

[88] T. Sjöstrand and M. van Zijl, *Phys. Rev.* **D36** (1987) 2019.

[89] T. Sjöstrand and P. Z. Skands, *Eur. Phys. J.* **C39** (2005) 129–154, [hep-ph/0408302].

[90] H. L. Lai *et al.*,, **CTEQ** Collaboration *Eur. Phys. J.* **C12** (2000) 375–392, [hep-ph/9903282].

[91] B. Andersson, *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.* **7** (1997) 1–471.

[92] D. B. Renner, *J. Phys. Conf. Ser.* **9** (2005) 264–267, [hep-lat/0501005].

[93] P. Hagler *et al.*,, **LHPC** Collaboration arXiv:0705.4295 [hep-lat].

[94] T. Sjöstrand and P. Z. Skands, *JHEP* **03** (2004) 053, [hep-ph/0402078].

[95] M. Sandhoff and P. Skands, Presented at Les Houches Workshop on Physics at TeV Colliders, Les Houches, France, 2-20 May 2005, hep-ph/0604120.

[96] T. Sjöstrand, S. Mrenna, and P. Skands, arXiv:0710.3820 [hep-ph].

[97] J. M. Butterworth, J. R. Forshaw, and M. H. Seymour, *Z. Phys.* **C72** (1996) 637–646, [hep-ph/9601371].

[98] G. Corcella *et al.*, *JHEP* **01** (2001) 010, [hep-ph/0011363].

[99] B. I. Abelev *et al.*,, **STAR** Collaboration *Phys. Rev.* **C75** (2007) 064901, [nucl-ex/0607033].

[100] M. Heinz,, **STAR** Collaboration arXiv:0707.1510 [hep-ex].

[101] T. Sjöstrand and R. Corke, work in progress.

[102] R. E. Ansorge *et al.*,, **UA5** Collaboration *Z. Phys.* **C37** (1988) 191–213.

[103] A. Capella, U. Sukhatme, C.-I. Tan, and J. Tran Thanh Van, *Phys. Rept.* **236** (1994) 225–329.

[104] B. K. Srivastava,, **STAR** Collaboration nucl-ex/0702054.

[105] Y.-L. Yan, B.-G. Dong, D.-M. Zhou, X.-M. Li, and B.-H. Sa, arXiv:0710.2187 [nucl-th].

[106] E. Avsar, G. Gustafson, and L. Lönnblad, *JHEP* **01** (2007) 012, [hep-ph/0610157].

[107] T. Sjöstrand and P. Z. Skands, *Nucl. Phys.* **B659** (2003) 243, [hep-ph/0212264].

[108] R. M. Duran Delgado, G. Gustafson, and L. Lönnblad, *Eur. Phys. J.* **C52** (2007) 113–119, [hep-ph/0702241].

[109] G. Gustafson, arXiv:0712.1941 [hep-ph].

[110] K. Kato and T. Munehisa, *Comput. Phys. Commun.* **64** (1991) 67–97.

[111] Z. Nagy and D. E. Soper, *JHEP* **09** (2007) 114, [arXiv:0706.0017 [hep-ph]].

[112] J. M. Campbell, J. W. Huston, and W. J. Stirling, *Rept. Prog. Phys.* **70** (2007) 89, [hep-ph/0611148].

[113] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, *Phys. Lett.* **B604** (2004) 61–68, [hep-ph/0410230].

[114] E. Boos *et al.*,, **CompHEP** Collaboration *Nucl. Instrum. Meth.* **A534** (2004) 250–259, [hep-ph/0403113].

[115] S. Frixione, P. Nason, and B. R. Webber, *JHEP* **08** (2003) 007, [hep-ph/0305252].

[116] A. Sherstnev and R. S. Thorne, arXiv:0711.2473 [hep-ph].

[117] M. Dasgupta, L. Magnea, and G. P. Salam, arXiv:0712.3014 [hep-ph].

[118] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, arXiv:0802.2470 [hep-ph].

[119] S.-L. Blyth *et al.*, *J. Phys.* **G34** (2007) 271–281, [nucl-ex/0609023].

[120] O. Kodolova, I. Vardanian, A. Nikitenko, and A. Oulianov, *Eur. Phys. J.* **C50** (2007) 117–123.

[121] e. . d'Enterria, D. *et al.*, *J. Phys.* **G34** (2007) 2307–2455.

[122] M. Cacciari and G. P. Salam, arXiv:0707.1378 [hep-ph].

[123] S. D. Ellis, J. Huston, K. Hatakeyama, P. Loch, and M. Tonnesmann, arXiv:0712.2447 [hep-ph].

[124] S. V. Chekanov, hep-ph/0211298.

[125] M. Cacciari, G. P. Salam, and G. Soyez, arXiv:0802.1189 [hep-ph].

[126] G. Arnison *et al.*,, **UA1** Collaboration *Phys. Lett.* **B132** (1983) 214.

[127] G. C. Blazey *et al.*, hep-ex/0005012.

[128] G. P. Salam and G. Soyez, *JHEP* **05** (2007) 086, [arXiv:0704.0292 [hep-ph]].

[129] S. D. Ellis, J. Huston, and M. Tonnesmann, hep-ph/0111434.

[130] S. Catani, Y. L. Dokshitzer, M. Olsson, G. Turnock, and B. R. Webber, *Phys. Lett.* **B269** (1991) 432–438.

[131] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, *Nucl. Phys.* **B406** (1993) 187–224.

[132] S. D. Ellis and D. E. Soper, *Phys. Rev.* **D48** (1993) 3160–3166, [hep-ph/9305266].

[133] A. Banfi, G. P. Salam, and G. Zanderighi, *Eur. Phys. J.* **C47** (2006) 113–124, [hep-ph/0601139].

[134] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, *JHEP* **08** (1997) 001, [hep-ph/9707323].

[135] M. Wobisch and T. Wengler, hep-ph/9907280.

[136] F. Abe *et al.*,, **CDF** Collaboration *Phys. Rev.* **D45** (1992) 1448–1458.

[137] G. L. Bayatian *et al.*,, **CMS** Collaboration CERN-LHCC-2006-001.

[138] A. Bhatti *et al.*,, **CMS** Collaboration CMS PAS JME-07-003 in preparation (2008).

[139] A. T. Pierce and B. R. Webber, *Phys. Rev.* **D59** (1999) 034014, [hep-ph/9807532].

[140] S. Bethke *et al.*,, **JADE** Collaboration *Phys. Lett.* **B213** (1988) 235.

[141] W. Bartel *et al.*,, **JADE** Collaboration *Z. Phys.* **C33** (1986) 23.

[142] L. Lonnblad, *Z. Phys.* **C58** (1993) 471–478.

[143] J. E. Huth *et al.*, Presented at Summer Study on High Energy Physics, Reaearch Directions for the Decade, Snowmass, CO, Jun 25 - Jul 13, 1990.

[144] M. Cacciari and G. P. Salam, *Phys. Lett.* **B641** (2006) 57–61, [hep-ph/0512210].

[145] M. Cacciari, G. P. Salam, and G. Soyez,, "Fastjet." http://www.lpthe.jussieu.fr/~salam/fastjet/.

[146] P.-A. Delsart, K. Geerlings, and J. Huston,, "Spartyjet." http://www.pa.msu.edu/~huston/SpartyJet/SpartyJet.html.

[147] M. Cacciari, G. P. Salam, and G. Soyez, arXiv:0802.1188 [hep-ph].

[148] P. Nason, *JHEP* **11** (2004) 040, [hep-ph/0409146].

[149] S. Catani, Y. L. Dokshitzer, and B. R. Webber, *Phys. Lett.* **B285** (1992) 291–299.

[150] D. Acosta *et al.*, CERN-CMS-NOTE-2006-067.

[151] P. Skands,. private communication, 2008.

[152] G. Corcella *et al.*, hep-ph/0210213.

[153] S. Gieseke *et al.*, hep-ph/0609306.

[154] M. Bahr *et al.*, arXiv:0711.3137 [hep-ph].

[155] A. Abulencia *et al.*, *Phys. Rev. Lett.* **96** (2006) 122001, [hep-ex/0512062].

[156] A. Abulencia *et al.*, *Phys. Rev.* **D74** (2006) 071103, [hep-ex/0512020].

[157] A. Abulencia *et al.*, *Phys. Rev.* **D75** (2007) 092006, [hep-ex/0701051].

[158] V. M. Abazov *et al.*, arXiv:0802.2400 [hep-ex].

[159] W. T. Giele, E. W. N. Glover, and D. A. Kosower, *Phys. Rev. Lett.* **73** (1994) 2019–2022, [hep-ph/9403347].

[160] Z. Nagy, *Phys. Rev. Lett.* **88** (2002) 122003, [hep-ph/0110315].

[161] Z. Nagy, *Phys. Rev.* **D68** (2003) 094002, [hep-ph/0307268].

[162] S. Frixione and B. R. Webber, hep-ph/0612272.

[163] G. Marchesini *et al.*, *Comput. Phys. Commun.* **67** (1992) 465–508.