IRIS AperTO

UNIVERSITÀ DEGLI STUDI DI TORINO

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Physics at TeV Colliders 2009:THE TOOLS AND MONTE CARLO WORKING GROUP Summary Report

This is the author's manuscript

Original Citation:

Availability:

This version is available http://hdl.handle.net/2318/122726 since

Publisher:

arXiv.org

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

08 January 2022

THE TOOLS AND MONTE CARLO WORKING GROUP
Summary Report

J. Alwall[1], *A. Arbey*[3], *L. Basso*[4,5], *S. Belov*[6], *A. Bharucha*[7], *F. Braam*[9], *A. Buckley*[8], *J. M. Butterworth*[10*], *M. Campanelli*[10], *R. Chierici*[12], *A. Djouadi*[15], *L. Dudko*[16], *C. Duhr*[7], *F. Febres Cordero*[17], *P. Francavilla*[18], *B. Fuks*[19], *L. Garren*[11], *T. Goto*[20], *M. Grazzini*[21,22] *T. Hahn*[23], *U. Haisch*[24], *K. Hamilton*[25], *S. Heinemeyer*[26], *G. Hesketh*[2], *S. Höche*[27], *H. Hoeth*[7], *J. Huston*[28], *J. Kalinowski*[29], *D. Kekelidze*[6], *S. Kraml*[30], *H. Lacker*[31], *P. Lenzi*[13], *P. Loch*[32], *L. Lönnblad*[33], *F. Mahmoudi*[34], *E. Maina*[35,36], *D. Majumder*[37], *F. Maltoni*[38*], *M. Mangano*[2], *K. Mazumdar*[37], *A. Martin*[11,39], *J. Monk*[10], *F. Moortgat*[21*], *M. Muhlleitner*[40], *C. Oleari*[41], *S. Ovyn*[38], *R. Pittau*[42], *S. Plätzer*[40], *G. Piacquadio*[2,9], *L. Reina*[43], *J. Reuter*[9], *P. Richardson*[7*], *X. Rouby*[38], *C. Robinson*[10], *T. Roy*[44], *M. D. Schwartz*[45], *H. Schulz*[31], *S. Schumann*[46*], *E. von Seggern*[31], *A. Sherstnev*[16,48] *F. Siegert*[7,10], *T. Sjöstrand* [33], *P. Skands*[2*], *P. Slavich*[49], *M. Spira*[50], *C. Taylor*[10], *M. Vesterinen*[54], *S. de Visscher*[27], *D. Wackeroth*[51], *S. Weinzierl*[52], *J. Winter*[53], *T. R. Wyatt*[54]

[*] Session converners.

[1]SLAC National Accelerator Laboratory, Theoretical Physics Group, Mail Stop 81, 2575 Sand Hill Road, Menlo Park, CA 94025, USA

[2] CERN, CH-1211 Geneva 23, Switzerland

[3]Université de Lyon, France; Université Lyon 1, F–69622; CRAL, Observatoire de Lyon, F–69561 Saint-Genis-Laval; CNRS, UMR 5574; ENS de Lyon, France

[4] School of Physics & Astronomy, University of Southampton, Highfield, Southampton SO17 1BJ, UK

[5] Particle Physics Department, Rutherford Appleton Laboratory, Chilton, Didcot, Oxon OX11 0QX, UK

[6] Joint Institute for Nuclear Research, Dubna, Moscow region, Russia, 141980

[7] IPPP, Physics Department, Durham University, DH1 3LE, UK

[8] School of Physics and Astronomy, University of Edinburgh, EH9 3JZ, UK

[9] University of Freiburg, Institute of Physics, Hermann-Herder-Str. 3, 79104 Freiburg, Germany

[10] Department of Physics and Astronomy, University College London, WC1E 6BT, UK

[11] Fermi National Accelerator Laboratory, P.O Box 500, Batavia, IL 60510, USA

[12]Institut de Physique Nucleaire de Lyon, IN2P3-CNRS, Université Claude Bernard Lyon 1, Villeurbanne, France

[13] Università degli Studi di Firenze & INFN Firenze, via Sansone 1, 50019 Sesto F.no, Firenze, Italy

[14] Department of Physics and Astronomy, UCLA, Los Angeles, CA 90095-1547, USA

[15]Laboratoire de Physique Théorique, Université Paris XI, F–91405 Orsay Cedex, France

[16]Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University, Vorob'evy Gory, Moscow 119992, Russia

[17] Universidad Simón Bolívar, Departamento de Física, Apartado 89000, Caracas 1080A, Venezuela.

[18]INFN - Pisa, Universita' di Pisa, G. Galilei Graduate School, Pisa, Italy

[19] Institut Pluridisciplinaire Hubert Curien/Département Recherche Subatomique, Université de Strasbourg/CNRS-IN2P3, 23 Rue du Loess, F-67037 Strasbourg, France

[20]KEK Theory Center, Institute of Particle and Nuclear Studies, KEK, Tsukuba, 305-0801 Japan

[21]Dept. of Physics, ETH Zürich, Zürich, Switzerland

[22] INFN, Sezione di Firenze, 50019 Firenze, Italy.

[23]Max-Planck-Institut für Physik, Föhringer Ring 6, D–80805 Munich, Germany

[24]Institut für Physik (WA THEP), Johannes Gutenberg-Universität, D–55099 Mainz, Germany

[25] INFN, Sezione di Milano-Bicocca, 20126 Milan, Italy.

[26]Instituto de Física de Cantabria (CSIC-UC), Santander, Spain

[27] Universität Zürich, CH-8057 Zürich, Switzerland

[28]Dept. of Physics and Astronomy, Michigan State University, East Lansing (MI), USA

arXiv:1003.1643v1 [hep-ph] 8 Mar 2010

[29] Instytut Fizyki Teoretycznej UW, Hoza 69, PL-00681 Warsaw, Poland
[30]Laboratoire de Physique Subatomique et de Cosmologie (LPSC), UJF Grenoble 1, CNRS/IN2P3, 53 Avenue des Martyrs, 38026 Grenoble, France
[31] Inst. f. Physik, Humboldt-Universität zu Berlin, Berlin, Germany
[32]Department of Physics, University of Arizona, Tucson, Arizona, USA
[33]Department of Theoretical Physics, Lund University, Sölvegatan14A, S-223 62, Sweden.
[34]Clermont Université, Université Blaise Pascal, CNRS/IN2P3, LPC, BP 10448, 63000 Clermont-Ferrand, France
[35]Dipartimento di Fisica Teorica, Universita' di Torino, Via Giuria 1, 10125 Torino, Italy
[36] INFN, Sezione di Torino, Via Giuria 1, 10125 Torino, Italy.
[37]Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400 005, India.
[38]Centre for Particle Physics and Phenomenology (CP3), Université Catholique de Louvain, Chemin du Cyclotron 2, B–1348 Louvain-la-Neuve, Belgium
[39] Department of Physics, Sloane Laboratory, Yale University, New Haven, CT 06520 USA
[40] Institut für Theoretische Physik, KIT, 76128 Karlsruhe, Germany
[41] Università di Milano-Bicocca, 20126 Milano, Italy.
[42]Departamento de Física Teórica y del Cosmos Campus Fuentenueva, Universidad de Granada E-18071 Granada, Spain
[43] Florida State University
[44] Department of Physics and Institute of Theoretical Science, University of Oregon, Eugene, OR 97403 USA
[45]Department of Physics, Harvard University, Cambridge, MA, USA
[46] Institut für Theoretische Physik, Universität Heidelberg, Philosophenweg 16, D-69120 Heidelberg, Germany
[48] R. Peierls Centre for Theoretical Physics, University of Oxford, OX1 3NP, UK
[49]LPTHE, 4, Place Jussieu, 75252 Paris, France
[50] Paul Scherrer Institut, CH–5232 Villigen PSI, Switzerland.
[51] University at Buffalo, SUNY
[52] Institut für Physik, Universität Mainz, D - 55099 Mainz, Germany
[53]Theoretical Physics Department, Fermi National Accelerator Laboratory, Batavia, IL 60510, USA
[54]Particle Physics Group, School of Physics and Astronomy, University of Manchester, UK.

Abstract
This is the summary and introduction to the proceedings contributions for the Les Houches 2009 "Tools and Monte Carlo" working group.

Contents

1. FOREWORD **4**

I INTERFACES **5**

2. A STANDARD FORMAT FOR LES HOUCHES EVENT FILES, VERSION 2 **5**

3. A DRAFT RUNTIME INTERFACE TO COMBINE PARTON SHOWERS AND NEXT-TO-LEADING ORDER QCD PROGRAMS **13**

4. STATUS OF THE FLAVOUR LES HOUCHES ACCORD **19**

II TUNING 27

5. STATUS OF RIVET AND PROFESSOR MC VALIDATION & TUNING TOOLS 27

6. QUANTITATIVE ERROR ESTIMATION IN MC TUNES 31

7. MATRIX ELEMENT CORRECTIONS AND PARTON SHOWER MATCHING IN INCLUSIVE Z PRODUCTION AT LHC 38

III BEYOND FIXED ORDER 44

8. MULTIPLE PARTON INTERACTIONS AS A BACKGROUND TO TOP PAIR PRODUCTION 44

9. A MATCHING SCHEME FOR $W\gamma$ NLO MATRIX ELEMENT GENERATOR AND Pythia 8 PARTON SHOWER 51

10. THEORY TESTS OF PARTON SHOWERS 55

11. HIGGS BOSON PRODUCTION VIA GLUON FUSION AT THE LHC: A COMPARATIVE STUDY 58

12. $Wb\bar{b}$ IN THE HIGH-p_T HW REGION 68

IV OBSERVABLES AND DETECTORS 80

13. DELPHES, A FRAMEWORK FOR FAST SIMULATION OF A GENERIC COLLIDER EXPERIMENT 80

14. EFFECT OF QED FSR ON MEASUREMENTS OF $Z/\gamma*$ AND W LEPTONIC FINAL STATES AT HADRON COLLIDERS 84

V JETS AND JET SUBSTRUCTURE 91

15. STATUS OF JET SUBSTRUCTURE STUDIES 91

16. HEAVY PARTICLE DECAYS IN MONTE CARLO EVENT GENERATORS 96

17. SENSITIVITY OF QCD JET MASS AND JET SUBSTRUCTURE TO PILE-UP AT LHC 108

18. A STUDY OF RADIATION BETWEEN JETS AT THE LHC 114

VI BEYOND THE STANDARD MODEL 120

19. AN UPDATE OF THE PROGRAM HDECAY 120

20. IMPLEMENTATION AND VALIDATION OF MODELS BEYOND THE STANDARD MODEL WITH FEYNRULES 122

1. FOREWORD

The working group on "Tools" and Monte Carlos for TeV-scale physics held discussions throughout the two-week period of the Les Houches meetings. The topics covered herein span both sessions. Several of the topics followed on from those discussed in the "Standard Model Handles and Candles" session of the previous workshop [1]; there has been substantial progress and several new topics were introduced.

The contributions here in fact include substantial physics results derived from the programmes, interfaces and techniques discussed, as well as status reports on existing projects and proposals for new standards and interfaces.

In Part I we have collected the more technical proposals for common standards and interfaces, mostly required because of the rapid progress in higher order calculations. Part II collects results on the tuning of MC simulations, a critical topic for understanding LHC data. Part III contains several contributions comparing various all-order calculations with fixed-order results, with and without matching between the two. Part IV reflects some key issues on the communicability of results between experiment and theory. Part V discusses recent progress and ideas in using jets and jet substructure at the LHC to study QCD and search for new physics, and finally Part VI discusses progress in some key modeling tools for beyond-the-standard-model physics.

The productivity and pleasure of this workshop is overshadowed by the dreadful loss of our friend and colleague Thomas Binoth, and we dedicate these proceedings to him.

Part I
INTERFACES

2. A STANDARD FORMAT FOR LES HOUCHES EVENT FILES, VERSION 2
[1]

2.1 INTRODUCTION

The Les Houches Accord (LHA) for user-defined processes [2] has been immensely successful. It is routinely used to pass information from matrix-element-based generators (MEGs) to general-purpose ones (here, somewhat unfairly referred to as parton shower generators — PSGs), in order to generate complete events for a multitude of processes. The original standard was in terms of two *Fortran common blocks* where information could be stored, while the actual usage has tended to be mainly in terms of *files* with parton-level events. For this purpose a new accord — the Les Houches Event File (LHEF) accord [3] — was introduced in 2006, which standardized the way such event files should be structured.

The LHEF was constructed using XML tags in order to make it flexible and easy to extend (although some additional structure is assumed inside some tags which is not formulated in XML). The format has been extremely useful, and has basically become the standard way to interface matrix element generators and parton shower programs.

As the matching and merging of tree-level matrix elements and parton showers are now becoming the state-of-art, it is reasonable to let this be reflected in an updated file format to standardize how relevant information should be given by the matrix element generators in a usable fashion for the parton shower programs. Furthermore, with the matching of next-to-leading order (NLO) calculations and parton showers becoming increasingly widespread, it is worth considering how the LHEF can be augmented in order to facilitate this.

For the CKKW-type merging algorithms [4, 5] it is convenient to allow the Sudakov-reweighting to be done in the MEG, as this will automatically regularize soft- and collinear divergencies. Hence it would be desirable if the LHEF could include information about this. This does not only mean that a weight needs to be added, but also information about which cuts has been imposed in the MEG as well as information on how the generated event was clustered to obtain the relevant scales and Sudakov form factors.

In the case the events are produced by a NLO MEG, the situation is a bit more complicated. Here a subtraction scheme is typically used to handle the cancellation between real and virtual corrections. This means that, besides loop-level events, each tree-level real event with one extra parton will need to be supplemented by *counter events* corresponding to the assumed projections of the tree-level event to born-level events with one parton less. To allow for matching or merging with a PSG, these events need to be considered together in a group of events, something that was not forseen in the original file format.

Independent of these ME-PS matching considerations, we also wish to introduce some further, minor, additions to assist the determination of errors arising from the the parton density function (PDF) parameterizations used in the MEG. Normally these error estimates are given as a set of different PDFs where the parameters have been varied around the best fit value. Hence, a given event may be associated with several weights corresponding to the different PDFs used.

Note that the scope of the format suggested here is somewhat different from the HepML schema [6] (used by eg. the MCDB project [7]). The LHEF format is specialized in the interface between matrix element generators and parton shower programs, while HepML is intended to give more general meta-information on how events have been produced by an event generator. However, there is nothing that

[1] Contributed by L. Lönnblad Email `leif.lonnblad@thep.lu.se`, J. Alwall, S. Belov, L. Dudko, L. Garren, K. Hamilton, J. Huston, D. Kekelidze, E. Maina, F. Maltoni, M. Mangano, R. Pittau, S. Plätzer, A. Sherstnev, T. Sjöstrand, P. Skands.

```
<LesHouchesEvents version="1.0">
  <!--
    # optional information in completely free format,
    # except for the reserved end tag (see next line)
  -->
  <header>
    <!-- individually designed XML tags, in fancy XML style -->
  </header>
  <init>
    compulsory initialization information
    # optional initialization information
  </init>
  <event>
    compulsory event information
    # optional event information
  </event>
  (further <event> ... </event> blocks, one for each event)
</LesHouchesEvents>
```

Fig. 1: The original structure of a Les Houches event file.

prevents the LHEF format to be included in the HepML structure in the future.

The outline of this article is as follows. In section 2.2 we review the structure of the original LHEF accord and of the Les Houches common block structure on which it is based. Then in section 2.3 we present the additional XML tags which may be used to specify additional global, and per-event information. Finally we give a brief summary and outlook.

2.2 THE ORIGINAL EVENT FILE FORMAT AND COMMON BLOCK STRUCTURE

The first version of the Les Houches event file format was a simple structure specifying how to write the Les Houches common blocks to a text file. A few XML tags were defined to simplify parsing but not much more than the information in the common blocks was formalized. The structure of a file is outlined in figure 1, where the tags are as follows.

- `LesHouchesEvents`: which contains the whole file and which mandates a `version` attribute set to `"1.0"`.
- `header`: which may contain any number of unspecified XML tags describing how the events were generated.
- `init`: This is the tag which specifies the information in the `HEPRUP` common block. The start tag must be alone on a line and the following line must contain the information which is in common for all processes in the file. The lines following this must contain the per-process information from the common block, one process per line. If there are any other lines before the end tag, they must be preceded by a `#`-sign (c.f. figure 2).
- `event`: The `init` tag may be followed by any number of `event` tags, one for each event generated. Also the `event` start tag must be alone on a line and the following line must contain the general event information from the `HEPEUP` common block. The lines following this must contain the per-particle information, one line per particle. Also here additional lines may be included before the end tag if they are preceded by a `#`-sign. (c.f. figure 3).
- Before the `init` tag one may, optionally, include arbitrary text enclosed in XML comment tags, `<!-- ... -->`, but no other text is allowed in the enclosing `LesHouchesEvents` tag.

For a more detailed description of the LHEF format we refer to [3].

```
<init>
IDBMUP(1) IDBMUP(2) EBMUP(1) EBMUP(2) PDFGUP(1) PDFSUP(1) PDFSUP(2) IDWTUP NPRUP
XSECUP(1) XERRUP(1) XMAXUP(1) LPRUP(1)
XSECUP(2) XERRUP(2) XMAXUP(2) LPRUP(2)
...
XSECUP(NPRUP) XERRUP(NPRUP) XMAXUP(NPRUP) LPRUP(NPRUP)
# Additional
# information
</init>
```

Fig. 2: The structure of the `init` tag in the original LHEF format. See [2] for the meaning of the different common block variables.

```
<event>
NUP IDPRUP XWGTUP SCALUP AQEDUP AQCDUP
IDUP(1) ISTUP(1) MOTHUP(1,1) MOTHUP(2,1) ICOLUP(1,1) ICOLUP(2,1) PUP(1,1) PUP(2,1) PUP(3,1) PUP(4,1) PUP(5,1)
IDUP(2) ISTUP(2) MOTHUP(1,2) MOTHUP(2,2) ICOLUP(1,2) ICOLUP(2,2) PUP(1,2) PUP(2,2) PUP(3,2) PUP(4,2) PUP(5,2)
...
# In total 1+NUP lines after the <event> tag
# Additional
# information
</event>
```

Fig. 3: The structure of the `event` tag in the original LHEF format. See [2] for the meaning of the different common block variables.

2.3 THE NEW FILE FORMAT

We now describe our suggestion for an updated file format which includes the additional information mentioned in the introduction. All such information is encoded in XML tags with optional attributes given in the usual way:

```
<tag attribute1="value" attribute2="value">content</tag>
```

or, for a tag without content,

```
<tag attribute1="value" attribute2="value" attribute3="value" />
```

The new tags can either be given in the `init` block, should they refer to the whole file, or in the `event` block, if they only refer to an individual event. In addition `group` tags can be inserted to group events together.

2.31 GLOBAL INFORMATION

The following tags may be included inside the `init` tag and contain additional global information about how the events in the file were produced. They must be placed after the mandatory lines containing `HEPRUP` common block information (see figure 2), but otherwise the order is unimportant. Only tags which are not marked optional below need to be supplied.

The generator **tag (optional)** This is just added to give easy access to the name of the program which has generated the file. The content of the tag is simply the name and the only allowed attribute is

- `version`: a string describing the version of the generator used.

The xsecinfo **tag (required)** The information in the `HEPRUP` common block is in principle sufficient to determine the cross sections of the processes involved. Currently, the way in which this information

is specified is a bit complicated and sometimes confusing, since it was assumed to be used to pass information between the MEG and PSG in both directions. For the event file, the communication is per definition one-way, and the information can be made more easily accessible. The tag itself has no content, and the information is given in the following attributes.

- `neve` (R)[2]: the number of events[3] in the file.
- `totxsec` (R): the total cross section (in units of pb) of all processes in the file.
- `maxweight` (D=1)[4]: the maximum weight of any event[5] in the file (in an arbitrary unit).
- `minweight` (D=`-maxweight`): if the file contains negative weights, the `minweight` is the most negative of the negative weights in the file. (Must obviously be the same unit as `maxweight`.)
- `meanweight` (D=1): The average weight of the events in the file (same unit as maxweight).
- `negweights` (D=`no`): If `yes`, then the file may contain negative weights.
- `varweights` (D=`no`): If `yes`, then the file may contain varying event weights. If `no`, all events are weighted with maxweight (or, if `negweights=yes`, with minweight).
- `eventgroups` (D=`no`): If `yes`, the events in the file may be grouped together with `group` tags, in which case the attributes above count an event group as one event rather than several separate ones.
- `maxingroupweight` (D=`maxweight`): If `eventgroups=yes`, this gives the maximum weight among the events inside groups.
- `miningroupweight` (D=`-maxingroupweight`): If `eventgroups=yes`, this gives the minimum weight among the events inside groups.

Note that it is assumed that all processes in the file are weighted with respect to a common total cross section, such that summing the weights for the events of a given process and multiplying with `totxsec/maxweight/neve` will give the cross section for that process. In this way, the per-process information in the `HEPRUP` common block can be safely ignored.

The cutsinfo **tag (optional)** This tag is used to supply information about which kinematical cuts were used to generate the events in the file. Several different cuts can be given with `cut` tags and it is possible to specify which particles should be affected by each cut using `ptype` tags.

The `cut` tag contains an actual cut made in the generation of the events. In general, all events in the file will pass this cut. The cut is defined in terms of a kinematical variable and the particles which are affected. The content of the tag is one or two numbers giving the allowed range of value of the kinematical variable according to the attribute `limit` (see below).

The variable is defined according to the following attributes of the `cut` tag:

- `p1` (D=0): Lists the particle types for which this cut applies. This can be either a number, corresponding to a given particle PDG [8] code, or a string corresponding to a group of particles previously defined with a `ptype` tag (see below). The default is zero which means *any particle type*.
- `p2`, ..., `p9`: Allows the specification of additional sets of particle types, by analogy to `p1`, in order to facilitate the application of different classes of cuts to different classes of particles.
- `type` (R): This defines the variable which is cut. The following values are predefined, but also other variables may be specified. (Where relevant, the laboratory frame is assumed, and all energy units are in GeV.)

[2](R) means the attribute is mandatory

[3]Note that if the file contains events inside `group` tags (see section 2.33 below), `neve` must refer to the number of event groups (plus the events which are outside the groups).

[4]For attributes which are not mandatory, (D=...) indicates which value is assumed it not present

[5]Note that if the file contains events inside `group` tags (see section 2.33 below), `maxweight`, `minweight` and `meanweight` must refer to the weights of the groups (and the weights of the events which are outside the groups).

- m: the invariant mass of a particle of type p1. If additional particle types are specified the cut applies to the invariant mass of the corresponding number of particles, i.e. if p1, p2 and p3 are specified the cut is on the invariant mass of any set of three matching particles.
- pt: the transverse momentum of a particle matching p1.
- eta: the pseudo-rapidity of a particle matching p1.
- y: the true rapidity of a particle matching p1.
- deltaR: the pseudo-rapidity–azimuthal-angle difference ($\sqrt{\Delta\eta^2 + \Delta\phi^2}$) between two particles matching p1 and p2 respectively.
- E: the energy of a particle matching p1.
- ETmiss: the norm of the vectorial sum of the pt of final state particles matching p1 and *not* matching p2 (Note that an empty p2 defaults to the empty set here and for HT below).
- HT: the scalar sum of the transverse momentum of final state particles matching p1 and *not* matching p2.

- limit (D=min): If set to min (max) only one number should be marked by the tag and give the minimum (maximum) for the kinematical variable, while if it is set to minmax, there should be two numbers corresponding to the minimum and maximum (in that order).

The groups of particles to be considered in the p1 and p2 attributes of the cut tag are specified by ptype tags, which simply contains the PDG codes of the particle types belonging to the group. The only allowed attribute in the ptype tag is

- name (R): the name of this group of particle types.

Here is a short example on how to specify a cut where a charged electron or muon is required to have a transverse momentum of at least 20 GeV and a minimum of 25 GeV missing transverse energy is required:

```
<cutsinfo>
  <ptype name="l+-">11 -11 13 -13</ptype>
  <ptype name="nu">12 -12 14 -14 16 -16</ptype>
  <cut type="pt" p1="l+-">20</cut>
  <cut type="ETmiss" p1="0" p2="nu">25</cut>
</cutsinfo>
```

The procinfo **tag (optional)** The procinfo tag is used to supply additional per-process information in addition to what is given in the HEPRUP common block part of the init tag. The content of the tag is simply an arbitrary string describing the process. The attributes are the following:

- iproc (D=0): The process number for which the information is given. This must correspond to the LPRUP code in the HEPRUP common block for the corresponding process. Also zero can be given, in which case it refers to all processes in the file (except those with a separate procinfo tag).
- loops (D=0): The number of loops used in calculating this process.
- qcdorder: The power of α_S used in calculating this process.
- eworder: The power of the electro-weak coupling used in calculating this process.
- rscheme (D=MSbar): The renormalization scheme used in calculating this process.
- fscheme (D=MSbar): The factorization scheme used in calculating this process.
- scheme (D=tree): Information about the scheme used to calculate the matrix elements to NLO. If absent, a pure tree-level calculation is assumed. Possible values could be CSdipole (NLO calculation with Catani–Seymour subtraction [9]), FKS [10,11], MC@NLO [12,13], POWHEG [14,15] and NLOexclusive (NLO calculation according to the exclusive cross section (see eg. [16]) within the given cuts).

The mergetype **tag (optional)** For some merging schemes (eg. for CKKW) it is possible to reweight the the events with Sudakov form factors already in the MEG. If this has been done the content of the `mergetype` tag for the corresponding process should give a name corresponding to the scheme used. The attributes are:

- `iproc`: The process number for which the information is given. A zero means all processes except those with a separate `mergeinfo` tag.
- `mergingscale` (R): The value of the merging scale in GeV.
- `maxmult` (D=`no`): If `yes`, the corresponding process is reweighted as if it is the maximum multiplicity process, i.e. the Sudakov form factor associated with evolution between the smallest clustering scale and the merging scale is not included.

2.32 PER-EVENT INFORMATION

Information about a given event may be given with XML tags after the mandatory lines containing `HEPEUP` common block information (see figure 3).

The weight **tag (optional)** An event can be associated with a number of different weights given in `weight` tags. The content of these tags is simply a sequence of weights corresponding to the cross section for the event using different PDFs, α_S values, etc. which can be used to estimate the systematic errors due to, e.g., PDF uncertainties. Each `weight` tag should be given a name for identification. Only one `weight` tag per event can be without a name and should then only contain one weight, which is the one for which the statistics in the `xsecinfo` tag is given. The attributes of the `weight` tag are as follows

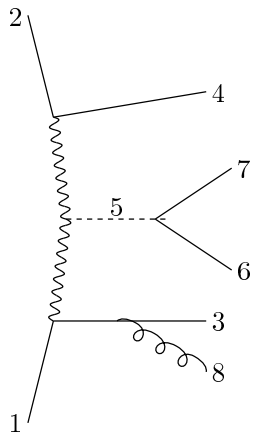
- `name`: An arbitrary string describing this set of weights. If no `name` is given this is the main weight for the event.
- `born` (D=1): If this is not a normal tree-level event but reweighted in some way (eg. by Sudakov reweighting or using loop contributions), this should be set to the relative weight of the tree-level cross section.
- `sudakov` (D=1): If this event has been reweighted by a Sudakov form factor, the size of this factor should be given here.

The last two attributes will probably only be given for the main weight. If an event has only been reweighted by a Sudakov form factor then these attributes are related by `born*sudakov=1`. The total Born cross section is obtained by summing the weights multiplied by `born` for each event of the given process, and multiplying with `totxsec/maxweight/neve` from the `xsecinfo` tag.

The clustering **tag (optional)** If an event has eg. been reweighted with Sudakov form factors, it is possible to specify how the current event has been clustered to find the scales involved. The contents of this this tag should be a series of `clus` tags. The clustering should be defined from the final state backwards in terms inverse time-like splittings, in the end defining a ”bare” ladder diagram. This is then followed by a sequence of space-like splittings.

The `clustering` tag contains a number of `clus` tags corresponding to each of the splittings. Each `clus` tag contains two or three integers. The first two numbers indicate which particles entries in the `HEPEUP` common block are clustered. If a third number is given it should correspond to an actual particle entry which corresponds to the combined object (if eg. a decayed resonance is explicitly present in the `HEPEUP` common block). If no third number is given, the clustered object is in the following referred to by the first number. The attributes of the `clus` tag are:

- `scale`: The scale (in GeV) associated with the clustering.
- `alphas`: If the event has been reweighted with an α_S at the scale of this clustering, the value of this α_S should be supplied here.




```
<clustering>
<clus>6 7 5</clus>
<clus>3 8</clus>
<clus>1 3</clus>
<clus>1 5</clus>
<clus>1 4</clus>
</clustering>
```

Fig. 4: Example of how a Feynman diagram can be encoded using a clustering tag. The numbering in the diagram corresponds to the particle entries in the `HEPEUP common block`

We also wish to draw attention to the fact that the `clustering` tag can equally be used to encode the Feynman diagram (or the most likely of the ones) used to produce the event. See figure 4 for an example.

The pdfinfo **tag (optional)** The `pdfinfo` tag contains the values of the PDFs used when generating this event, given by two numbers, $xf_1(x_1, Q^2)$ and $xf_2(x_2, Q^2)$, for the two incoming partons. The attributes are:

- `p1`: The PDG code of the first incoming parton.
- `p2`: The PDG code of the second incoming parton.
- `x1`: The momentum fraction of the first incoming parton.
- `x2`: The momentum fraction of the second incoming parton.
- `scale`: (D=`SCALUP`) The scale in GeV used in the PDFs (the default is taken from the `HEPEUP` common block).

2.33 GROUPING OF EVENTS

If we have a NLO calculation using eg. Catani–Seymour subtraction of a process with N particles in the Born level, each $N+1$ tree-level event will come with a number of counter events with N particles. For this reason there is a need to group events together. Such a group of events should be included in a `group` tag.

The group **tag (optional)** The content of this tag is a number of `event` tags which for should be considered together. If this is a $N+1$ tree-level event with a number of N-particle counter events, the first `event` must always be the $N+1$-particle event even if this event fails the cuts. The rest of the events are then the counter events. The `group` tag must also contain at least one main `weight` tag (without name attribute) which is the one for which the statistics in the `xsecinfo` tag is given. The individual weights of the events in the group should sum up to the weight of the whole group. The only allowed attribute is

- `n` (R): The number of events in the group.

Note that if event groups are present, the `neve` attribute in the `xsecinfo` tag should count an event group as a single event. Also, it is the `weight` of the event group which relates to the `maxweight` and `meanweight` attributes in the `xsecinfo` tag. To be compatible with the previous standard, where the `<event>` and `</event>` tags are required to be alone on a single line, also the `<group>` and `</group>` tags are required to be alone on a single line.

2.4 OUTLOOK

Event files which follows this new standard should have their `version` attribute in the `LesHouches-Events` (see figure 1) tag set to `2.0`. The web page `http://home.thep.lu.se/~leif/LHEF/` contains a number of `C++` classes implementing the reading and writing of files according to the new standard.

We have tried to make the new standard backward compatible with the previous one, and although all existing parsers may not be able to read new files we propose to keep the old preferred `.lhe` file name extension.

As with the previous standard, the current proposal must not be viewed as the end of the road. There may be further information exchange that ought to be standardized. It is allowed to use/promote a "private standard" of tags in the `header` block or of additional event information, and experience with such could point the way towards an extended standard at a later date.

Note that a formal description of the proposed new standard can be found at the Les Houches 09 wiki pages.[6]

[6] `http://www.lpthe.jussieu.fr/LesHouches09Wiki/images/6/60/Grammar.pdf`

3. A DRAFT RUNTIME INTERFACE TO COMBINE PARTON SHOWERS AND NEXT-TO-LEADING ORDER QCD PROGRAMS [7]

3.1 INTRODUCTION

Parton shower simulation programs like Pythia 6 and Herwig 6 [17, 18] have been the workhorses for high energy physics experiments for a long time. With the advent of the Large Hadron Collider (LHC) at CERN, the FORTRAN generators have been completely rewritten and extended as Pythia 8 and Herwig++ [19–21] and the program Sherpa has been established [22, 23].

Many new developments have been made in order to refine these simulations and to extend their applicability. Particularly, progress is now being made towards automation of schemes combining parton shower Monte Carlos and NLO QCD corrections consistently along the lines of now established schemes [12, 14]. Especially the POWHEG scheme exhibits close connections to methods correcting the hardest emission in parton shower simulations to the relevant exact real emission matrix element, [24]. Dedicated codes performing just the matching to NLO QCD are as well under development, [25].

In order to use existing and future tools for fixed-order calculations, especially in view of efforts towards automation present for these problems as well, it is desirable to specify an interface to the generic building blocks of fixed-order, particularly NLO QCD calculations carried out within the subtraction formalism.

The approach of communicating simulation results between different programs taken so far by exchanging event files of a definite format [3] is not appropriate for the task being addressed here. Rather a *runtime interface* between two distinct codes is desirable. Here, one code links to another either statically or dynamically and makes use of the implemented functionality through an interface which is now a set of definite function calls.

The interface should however not limit the particular ways in which fixed-order calculations are carried out, neither how the matching is actually performed by the parton shower Monte Carlo or a dedicated matching code. It should further not limit which programming language is actually being used and is therefore formulated in a language independent way. Any binding to a particular language is an implementational detail.

3.2 PRELIMINARIES

3.21 OBJECTS HANDLED BY THE INTERFACE

The interface aims at having each individual building block of a leading- or next-to-leading order QCD calculation at hand. In particular

- phase space generation
- tree-level matrix elements
- subtraction terms
- finite parts of one loop/Born interferences
- finite remainder terms originating from collinear factorization

are addressed as individual objects. These objects will be defined in detail in the following sections.

3.22 DIFFERENTIAL CROSS SECTIONS

The interface is formulated to handle all pieces inherent to a leading- or next-to-leading order differential cross section to be available in the following form, where α denotes any contribution as listed in the

[7] Contributed by: S. Plätzer

previous subsection:

$$\mathrm{d}\sigma_{ij}^{\alpha} = f_i(x_i,\mu_F) f_j(x_j,\mu_F) \times \\ F^{\alpha}_{\{i,j,k_1,...,k_n\}}(\{p_i,p_j,p_1,...,p_n\},\mu_R) \left| \frac{\partial\phi(\{p_i,p_j,p_1,...,p_n\},x_i,x_j)}{\partial\vec{r}} \right| \mathrm{d}^d r \ . \quad (1)$$

Here $\{i,j,k_1,...,k_n\}$ identifies a particular subprocess of which the term indexed by α contributes to, $\{p_i,p_j,p_1,...,p_n\}$ is the physical phase space point, and $\vec{r}$ is a set of d random numbers in the d-dimensional unit hypercube.

In the following the Jacobian determinant will be referred to as a phase space generator. The corresponding functionality will be addressed as an individual part of differential cross sections within the interface. Note that this does not reference the precise way in which the Monte Carlo integration or the generation of unweighted events is performed by any client code[8].

The parton density functions (PDFs) f as well as the value of renormalization scale μ_R, factorization scale μ_F and the strong coupling will be provided by the client code, such that the part of the interface representing the contribution α is completely specified as a function evaluating F^{α}. Additionally, few bookkeeping specifications will be introduced.

For tree-level matrix elements (including the real-emission part of a NLO calculation, F just represents the corresponding tree-level matrix element squared times the appropriate flux and symmetry factors.

3.23 NLO DIFFERENTIAL CROSS SECTIONS AND SUBTRACTION

The interface focuses on NLO QCD corrections being carried out within the subtraction formalism to obtain finite contributions from both virtual and real-emission corrections to any infrared safe observable.

Examples of schemes widely in use are [9, 10, 26], the interface to be specified in detail in the next section should however not limit any implementation to a particular scheme.

Few universal properties are however assumed to be common to any subtraction scheme at next-to-leading order. In particular it will be assumed that the auxiliary cross section introduced to subtract the divergent behaviour of the real emission contribution to NLO corrections for a $2 \to n$ process is of the form

$$\sum_{\alpha=1}^{k} \mathrm{d}\sigma_{sub}^{\alpha}(p_{n+1}, i_{n+1}) \mathcal{O}(p_n^{\alpha}(p_{n+1}), i_n^{\alpha}) \ . \quad (2)$$

Here, $\mathcal{O}$ denotes an infrared safe observable (sometimes also referred to as a jet defining function) and $p_n^{\alpha}(p_{n+1})$ is a unique and invertible momentum mapping from a real emission to an *underlying Born configuration*, which is identified by a flavour mapping $i_{n+1} \to i_n^{\alpha}$.

The number of individual subtraction terms k is not assumed to be fixed. Besides the unique association of a subtraction term to an underlying Born configuration each subtraction term is expected to subtract of a particular set of one or more collinear divergences, which at NLO are identified by just labelling the pair of partons becoming collinear. This information is crucial for a parton shower simulation to assign an additional parton to a unique emitter in a meaningful way thereby fixing the initial conditions for subsequent showering.

3.24 'COLLINEAR REMAINDERS'

After analytical integration of the subtraction terms, the encountered divergences cancel these present in the virtual correction and counter terms needed to renormalize the parton distribution functions. The finite remainder left from cancelling the latter divergences will then additionally depend on the factorization scale, the momentum fractions $x_{1,2}$ and convolutions in one or two variables $z \in [0,1]$ are in general

[8] Throughout this work the NLO code itself is considered to be 'client' code of the interface.

present. Further, these convolutions are assumed to be casted in a form which is suitable to be done by Monte Carlo methods as well. The interface therefore assumes a modified version of the contribution F associated to these finite remainders,

$$F^{coll}_{\{i,j,k_1,...,k_n\}} = F^{coll}_{\{i,j,k_1,...,k_n\}}(\{p_i, p_j, p_1, ..., p_n\}, \mu_R, \mu_F, x_1, x_2, z_1, z_2) \ , \tag{3}$$

where two additional random numbers $z_{1,2}$ on the unit interval are provided such that the convolution is performed on averaging over all events.

3.25 COLOUR DECOMPOSITION

In order to determine parton shower initial conditions but as well for the purpose of implementing independent subtraction schemes, the knowledge of partial amplitudes for the Born and real emission contribution is crucial.

The basis being used in this colour decomposition is however not unique. To the extent that parton shower initial conditions are usually determined by colour flows in the large-N limit, it would be desirable to have this decomposition in the fundamental representation of $SU(N = 3)$ available. Here, the basis tensors are just strings of Kronecker δ's,

$$\delta^{i_1}_{j_1} \cdots \delta^{i_n}_{j_n} \ , \tag{4}$$

where an upstairs index transforms according to the anti-fundamental, a downstairs index according to the fundamental representation, indicating outgoing colour (incoming anticolour) and outgoing anticolour (incoming colour), respectively. Note that this representation is not limited to, but well suited, for the large-N limit. The interface assumes this representation being available, but may be extended to other representations leaving it up to the client code to change basis from one to another decomposition.

3.26 NOTATION

The interface is formulated as a set of functions, taking any number of arguments and returning a result. The result may be a composite object of several types, which, depending on the language binding, may individually be passed as references to the function call.

A function of name **F** taking arguments of type A1, A2,..., returning a result of type T is denoted by T **F** (A1,A2,...), where each argument may be followed by a name identifying the meaning of the argument. The required types are defined in the following section, except for obvious simple types. Semantic definitions of each interface part are accompanied by pre- and postconditions where needed. Examples of language bindings are given to explain the relation between the abstract notation and implementational details.

3.27 TYPES USED

The purpose of this section is to introduce the more complex types used to specify the interface components. Few basic types with obvious semantics are used without further documentation.

vector – A list of objects of the same type. Elements in the list can be accessed randomly by an integer index. The type of objects stored is indicated in angle brackets, e.g. vector<double> is a vector storing doubles. The size of the vector is to be asserted in the context of a particular piece of the interface. Examples are INTEGER V(4) or std::vector<int> v(4); for FORTRAN and C++, respectively.

pair – A pair stores to values of potentially different type. This is an auxiliary concept which does not have to be supported by a particular language. A pair storing values of type T1 and T2, respectively, is denoted pair<T1,T2>.

union – A union generalizes pair to more than two entries.

processid – Defined to be vector<int>. Identifies a subprocess giving PDG ids for incoming partons in the first two entries, and PDG ids for outgoing partons in the following entries. The size of the vector is

guaranteed to be greater or equal to three. The subprocess $gg \to d\bar{d}g$, for example, would be identified by the entries $\{21, 21, 1, -1, 21\}$.

momentum – Defined to be vector<double> of length five. The first four entries contain the four-momentum components p_x, p_y, p_z, E in units of GeV. The fifth component is optional containing the invariant mass squared in units of GeV squared. The metric is agreed to be mostly-minus, $p \cdot q = E_p E_q - \vec{p} \cdot \vec{q}$.

pspoint – Defined to be union<double,pair<double,double>,vector<momentum> >. Represents a phase space point as the phase space weight (Jacobian from unit-hypercube random numbers to phase space measure), momentum fractions of incoming partons and a set of momenta. The first two momentum entries specify the momenta of incoming partons, subsequent these of outgoing partons. The size of the momentum vector is greater or equal to three. The weight is given in units of the proper power of GeV to obtain a dimensionless quantity.

colourflow – Defined to be vector<pair<int,int> >. Represents a colour flow assigned to a particular subprocess in the following convention: in association to a processid, the first and second members of an entry of a colourflow at position i contain an integer id for the colour and anticolour line, the corresponding parton at position i in the process id is connected to. By convention, a triplet always has its second colour index set to zero, an antitriplet its first index. A singlet uses the pair $\{0, 0\}$. Note that this notation is not limited to large-N flows, but may represent any basis tensor in the fundamental representation by the following identification: Each non-zero entry in the first entry of a pair in a colour flow vector identifies an index transforming according to the anti-fundamental, each entry at the second position an index transforming according to the fundamental representation of $SU(N)$ and entries of the same id are attached to a Kronecker-δ. Thus, the $1/N$ suppressed singlet contribution to a gluon (cf. the Fierz identity for the fundamental representation generators) carries the pair $\{k, k\}$, where k is not zero. Example: Consider the process $e^+e^- \to d\bar{d}g$, which would be identified by the processid $\{-11, 11, 1, -1, 21\}$. Here two colour flows are possible, the leading part $\delta^{i_d}_{j_g}\delta^{i_g}_{j_{\bar{d}}}$, and the $1/N$ suppressed $\delta^{i_d}_{j_{\bar{d}}}\delta^{i_g}_{j_g}$ contribution. The first one would be identified by the colourflow $\{\{0, 0\}, \{0, 0\}, \{k_1, 0\}, \{0, k_2\}, \{k_2, k_1\}\}$, the latter by $\{\{0, 0\}, \{0, 0\}, \{k_1, 0\}, \{0, k_1\}, \{k_2, k_2\}\}$, where $k_1 \neq k_2$ are non-zero, positive integers.

3.3 SPECIFICATION OF THE INTERFACE

Initialization and Bookkeeping

bool **initialize** () – The fixed-order code performs initialization and reads any relevant parameters from its preferred input mechanism. It returns true on success and false on failure.

pair<int,string> **alphasinfo** () – Return information on the running of the strong coupling to be used as a combination of the number of loops contributing to the QCD β-function, and a string identifying the renormalization scheme used. Values for the latter have to be agreed on.

bool **haveleadingorder** (processid) – Return true, if the process identified by the given processid can be calculated at leading order.

vector<colourflow> **colourflows** (processid) – Return the possible colourflows which could be selected for the given processid.

vector<colourflow> **largencolourflows** (processid) – Return the possible colourflows in the large-N limit which could be selected for the given processid.

bool **haveoneloop** (processid) – Return true, if one loop QCD corrections to the given process can be calculated.

string **havesubtraction** (processid) – Return a non-empty string identifying a subtraction scheme, if real emission subtraction terms are available for the given process. This does not require that the real emission process itself can be calculated. Return an empty string, if subtraction terms are not present.

Other return values have to be agreed on.

vector<processid> **realemissions** (processid) – Assuming the given processid identifies a Born process, return the real emission processes to be considered for a NLO QCD correction.

vector<int> **subtractions** (processid) – Assuming the given processid identifies a real emission process, return a list of ids for the subtraction terms to be considered. The ids need to be unique to the fixed-order code and are not used except for identifying a subtraction term. For dynamically allocated vectors, the function may return an empty vector to indicate that the process considered is non-singular, for fixed-size vectors it should indicate this by filling the vector with zeroes.

processid **underlyingborn** (processid, int) – Given a real emission process id and subtraction term id through the first and second arguments, respectively, return the underlying Born process the subtraction term maps to.

vector<pair<int,int> > **collinearlimits** (processid, int) – Given a real emission process id and subtraction term id through the first and second arguments, respectively, return the positions of the partons in the processid given, for which the identified subtraction term subtracts collinear singularities. The ordering of the returned value entries is irrelevant and by convention fixed such that the first entry is less than the second.

Kinematics and Phase Space

int **ndim** (processid) – Return the number of random numbers needed to generate a phase space point for the given process.

pspoint **phasespace** (pair<momentum,momentum>, vector<double>, processid) – Generate a phase space point given incoming particle's momenta, a list of random numbers $\in]0, 1[$ and a process id.

Dynamics

Each function call defined here represents the associated contribution F^α to differential cross section as defined in eq. 1. Results are assumed to be scaled by the appropriate power of GeV such as to return a dimensionless quantity.

double **me2** (processid, pspoint, double) – Return the helicity and colour summed matrix element squared for the given process, phase space point and value of the renormalization scale in GeV.

double **partialme2** (processid, pspoint, double, colourflow) – Return the helicity summed partial amplitude squared, identified by the given colour flow, evaluated for the given process, phase space point and value of the renormalization scale in GeV.

double **bornvirt** (processid, pspoint, double) – Return the helicity and colour summed Born-virtual interference plus the integrated subtraction terms (*i.e.* the finite part remaining after carrying out subtraction), consistent with the subtraction scheme chosen for the real emissions. Arguments in order are the process to be considered, the phase space point and the renormalization scale in GeV.

double **collinear** (processid, pspoint, double, double, pair<double,double>) – Return the finite collinear remainder contribution consistent with the subtraction scheme chosen for the process identified by the given processid and phase space point. Further arguments in order are renormalization and factorization scales in units of GeV, and two further random variables on the unit interval to perform Monte Carlo integration over convolutions present.

double **subtraction** (processid, int, pspoint, double) – Return the subtraction term identified by real emission process and subtraction term id, given a phase space point and renormalization scale in units of GeV.

CONCLUSIONS

A general runtime interface has been outlined to access the individual building blocks of a next-to-leading order (NLO) QCD calculation carried out within the subtraction formalism.

Such an interface is an indispensable tool for 'client' programs dedicated to the matching of parton showers and higher order corrections. A subset of the interface may as well be used to implement so-called matrix element corrections within parton shower Monte Carlos, correcting the hardest shower emission to the exact real emission matrix element squared.

ACKNOWLEDGEMENTS

I would like to thank Stephen Mrenna, Peter Skands, Leif Lönnblad and Nicolas Greiner for many fruitful discussions and encouraging comments.

4. STATUS OF THE FLAVOUR LES HOUCHES ACCORD[9]

4.1 INTRODUCTION

In addition to the increasing number of refined approaches in the literature for calculating flavour-related observables, advanced programs dedicated to the calculation of such quantities, e.g. Wilson coefficients, branching ratios, mixing amplitudes, renormalisation group equation (RGE) running including flavour effects have recently been developed [27–31]. Flavour-related observables are also implemented by many other non-dedicated public codes to provide additional checks for the models under investigation [32–38]. The results are often subsequently used by other codes, e.g. as constraints on the parameter space of the model under consideration [39–42].

At present, a small number of specialised interfaces exist between the various codes. Such tailor-made interfaces are not easily generalised and are time-consuming to construct and test for each specific implementation. A universal interface would clearly be an advantage here. A similar problem arose some time ago in the context of Supersymmetry (SUSY). The solution took the form of the SUSY Les Houches Accord (SLHA) [43, 44], which is nowadays frequently used to exchange information between SUSY related codes, such as soft SUSY-breaking parameters, particle masses and mixings, branching ratios etc. The SLHA is a robust solution, allowing information to be exchanged between different codes via ASCII files. The detailed structure of these input and output files is described in Refs. [43, 44].

The goal of this work is to exploit the existing organisational structure of the SLHA and use it to define an accord for the exchange of flavour related quantities, which we refer to as the "Flavour Les Houches Accord" (FLHA). In brief, the purpose of this Accord is thus to present a set of generic definitions for an input/output file structure which provides a universal framework for interfacing flavour-related programs. Furthermore, the standardised format will provide the users with a clear and well-structured result that could eventually be used for other purposes.

The structure is set up in such a way that the SLHA and the FLHA can be used together or independently. Obviously, some of the SLHA entries, such as measured parameters in the Standard Model (SM) and the Cabibbo-Kobayashi-Maskawa (CKM) matrix elements are also needed for flavour observable calculations. Therefore, a FLHA file can indeed contain a SLHA block if necessary. Also, in order to avoid any confusion, the SLHA blocks are not modified or redefined in the FLHA. If a block needs to be extended to meet the requirements of flavour physics, a new "`F`" block is defined instead.

Note that different codes may *technically* achieve the FLHA input/output in different ways. The details of how to 'switch on' the FLHA input/output for a particular program should be described in the manual of that program and are not covered here. For the SLHA, libraries have been developed to permit an easy implementation of the input/output routines [45]. In principle these programs could be extended to include the FLHA as well.

It should be noted that, while the SLHA was developed especially for the case of SUSY, the FLHA is, at least in principle, model independent. While it is possible to indicate the model used in a specific block, the general structure for the information exchange can be applied to any model.

This report summarizes the current status of the FLHA. Several issues are not defined in an unambigous way yet. This will be indicated in the text below.

4.2 CONVENTIONS AND DEFINITIONS

The structure of the Flavour Les Houches Accord input and output files is based on the existing SUSY Les Houches Accord structure and flavour quantities are defined in blocks. The general conventions for the blocks are very similar to the SLHA blocks [43] and they are not reproduced here.

Since a FLHA file can also contain SLHA blocks, to clearly identify the blocks of the FLHA, the first

[9]Contributed by: F. Mahmoudi, S. Heinemeyer, A. Arbey, A. Bharucha, T. Goto, T. Hahn, U. Haisch, S. Kraml, M. Muhlleitner, J. Reuter, P. Skands, P. Slavich

letter of the name of a block is an "F". There are two exceptions to this rule: blocks borrowed from the SLHA, which keep their original name, and blocks containing imaginary parts, which start with "IMF".

The following general structure for the FLHA file is proposed:

- BLOCK FCINFO: Information about the flavour code used to prepare the FLHA file.
- BLOCK FMODSEL: Information about the underlying model used for the calculations.
- BLOCK SMINPUTS: Measured values of SM parameters used for the calculations.
- BLOCK VCKMIN: Input parameters of the CKM matrix in the Wolfenstein parameterisation.
- BLOCK UPMNSIN: Input parameters of the PMNS neutrino mixing matrix in the PDG parameterisation.
- BLOCK VCKM: Real part of the CKM matrix elements.
- BLOCK IMVCKM: Imaginary part of the CKM matrix elements.
- BLOCK UPMNS: Real part of the PMNS matrix elements.
- BLOCK IMUPMNS: Imaginary part of the PMNS matrix elements.
- BLOCK FMASS: Masses of quarks, mesons, hadrons, etc.
- BLOCK FLIFE: Lifetime (in seconds) of flavour-related mesons, hadrons, etc.
- BLOCK FCONST: Decay constants.
- BLOCK FCONSTRATIO: Ratios of decay constants.
- BLOCK FBAG: Bag parameters.
- BLOCK FWCOEF: Real part of the Wilson coefficients.
- BLOCK IMFWCOEF: Imaginary part of the Wilson coefficients.
- BLOCK FOBS: Prediction of flavour observables.
- BLOCK FOBSERR: Theory error on the prediction of flavour observables.
- BLOCK FOBSSM: SM prediction for flavour observables.
- BLOCK FFORM: Form factors.

More details on several blocks are given in the following. The blocks SMINPUTS, VCKMIN, UPMNSIN, VCKM, IMVCKM, UPMNS, IMUPMNS are defined exactly as in the SLHA(2) and not further discussed here.

BLOCK FCINFO

Flavour code information, including the name and the version of the program:

1 : Name of the flavour calculator

2 : Version number of the flavour calculator

Optional warning or error messages can also be specified:

3 : If this entry is present, warning(s) were produced by the flavour calculator. The resulting file may still be used. The entry should contain a description of the problem (string).

4 : If this entry is present, error(s) were produced by the flavour calculator. The resulting file should not be used. The entry should contain a description of the problem (string).

This block is purely informative, and is similar to BLOCK SPINFO in the SLHA.

BLOCK MODSEL

This block provides switches and options for the model selection. The SLHA2 BLOCK MODSEL is extended to allow more flexibility.

`1` : Choice of SUSY breaking model or indication of other model. By default, a minimal type of model will always be assumed. Possible values are:

- `-1` : SM
- `0` : General MSSM simulation
- `1` : (m)SUGRA model
- `2` : (m)GMSB model
- `3` : (m)AMSB model
- `4` : ...
- `31` : THDM
- `99` : other model. This choice requires a string given in the entry `99`

`3` : (Default=0) Choice of particle content, only used for SUSY models. The defined switches are:

- `0` : MSSM
- `1` : NMSSM
- `2` : ...

`4` : (Default=`0`) R-parity violation. Switches defined are:

- `0` : R-parity conserved. This corresponds to the SLHA1.
- `1` : R-parity violated.

`5` : (Default=`0`) CP violation. Switches defined are:

- `0` : CP is conserved. No information on the CKM phase is used.
- `1` : CP is violated, but only by the standard CKM phase. All other phases are assumed zero.
- `2` : CP is violated. Completely general CP phases allowed.

`6` : (Default=`0`) Flavour violation. Switches defined are:

- `0` : No flavour violation.
- `1` : Quark flavour is violated.
- `2` : Lepton flavour is violated.
- `3` : Lepton and quark flavour is violated.

`31` : defines the type of THDM, is used only if entry `1` is given as `31`, otherwise it is ignored.

- `1` : type I
- `2` : type II
- `3` : type III
- `4` : type IV

`99` : a string that defines other models is used only if entry `1` is given as `99`, otherwise it is ignored.

BLOCK FMASS

The block `BLOCK FMASS` contains the mass spectrum for the involved particles. It is an addition to the SLHA `BLOCK MASS` which contained only pole masses and to the SLHA `BLOCK SMINPUTS` which contains quark masses. If a mass is given in two blocks the block `FMASS` overrules the other blocks. In `FMASS` we specify additional information concerning the renormalisation scheme as well as the scale at which the masses are given and thus allow for larger flexibility. The standard for each line in the block should correspond to the following FORTRAN format

```
(1x,I9,3x,1P,E16.8,0P,3x,I2,3x,1P,E16.8,0P,3x,'#',1x,A),
```

where the first nine-digit integer should be the PDG code of a particle, followed by a double precision number for its mass. The next integer corresponds to the renormalisation scheme, and finally the last double precision number points to the energy scale (0 if not relevant). An additional comment can be given after `#`.

The schemes are defined as follows:

`0` : pole
`1` : $\overline{\mathrm{MS}}$
`2` : $\overline{\mathrm{DR}}$
`3` : 1S
`4` : kin
`5` : ...

BLOCK FLIFE

The block `BLOCK FLIFE` contains the lifetimes of mesons and hadrons in seconds. The standard for each line in the block should correspond to the FORTRAN format

```
(1x,I9,3x,1P,E16.8,0P,3x,'#',1x,A),
```

where the first nine-digit integer should be the PDG code of a particle and the double precision number its lifetime.

BLOCK FCONST

The block `BLOCK FCONST` contains the decay constants in GeV. The standard for each line in the block should correspond to the FORTRAN format

```
(1x,I9,3x,I2,3x,1P,E16.8,0P,3x,'#',1x,A),
```

where the first nine-digit integer should be the PDG code of a particle, the second integer the number of the decay constant, and the double precision number its decay constant.

BLOCK FCONSTRATIO

The block `BLOCK FCONSTRATIO` contains the ratios of decay constants, which often have less uncertainty than the decay constants themselves. The ratios are specified by the two PDG codes in the form f(code1)/f(code2). The standard for each line in the block should correspond to the FORTRAN format

```
(1x,I9,3x,I9,3x,I2,3x,I2,3x,1P,E16.8,0P,3x,'#',1x,A),
```

where the two nine-digit integers should be the two PDG codes of particles, the third and fourth integers the numbers of the decay constants, which correspond to the second index of the entry in `BLOCK FCONST`, and the double precision number the ratio of the decay constants.

BLOCK FBAG

The block `BLOCK FBAG` contains the bag parameters. The standard for each line in the block should correspond to the FORTRAN format

```
(1x,I9,3x,I2,3x,1P,E16.8,0P,3x,'#',1x,A),
```

where the first nine-digit integer should be the PDG code of a particle, the second integer the number of the bag parameter, and the double precision number its bag parameter.

So far no normalisation etc. has been defined, which at this stage has to be taken care of by the user. An unambiguous definition will be given elsewhere.

BLOCK FWCOEF Q= ...

The block `BLOCK FWCOEF Q= ...` contains the real part of the Wilson coefficients at the scale `Q`.

The different orders $C_i^{(k)}$ have to be given separately according to the following convention for the perturbative expansion:

$$\begin{aligned} C_i(\mu) &= C_i^{(0)}(\mu) + \frac{\alpha_s(\mu)}{4\pi} C_{i,s}^{(1)}(\mu) + \left(\frac{\alpha_s(\mu)}{4\pi}\right)^2 C_{i,s}^{(2)}(\mu) \\ &\quad + \frac{\alpha(\mu)}{4\pi} C_{i,e}^{(1)}(\mu) + \frac{\alpha(\mu)}{4\pi}\frac{\alpha_s(\mu)}{4\pi} C_{i,es}^{(2)}(\mu) + \cdots . \end{aligned} \tag{5}$$

The couplings should therefore not be included in the Wilson coefficients.

The entries in `BLOCK FWCOEF` should consist of two integers defining the fermion structure of the operator and the operator structure itself. These two numbers are not thought to give a full representation including normalisation etc. of the operator, but merely correspond to a unique identifier for any possible Wilson coefficient. Consequently, the user has to take care that a consistent normalisation including prefactors etc. is indeed fulfilled. As an example, for the operator O_1,

$$O_1 = (\bar{s}\gamma_\mu T^a P_L c)(\bar{c}\gamma^\mu T^a P_L b) \tag{6}$$

the definition of the two numbers is given as follows. The appearing fermions are encoded by a two-digit number originating from their PDG code, where no difference is made between particles and antiparticles, as given in Table 1. Correspondingly, the first integer number defining O_1, containing the fermions $\bar{s}c\bar{c}b$, is given by 03040405. The various operators are defined in Table 2. Correspondingly, the second integer number defining O_1, containing the operators $\gamma_\mu T^a P_L\, \gamma^\mu T^a P_L$ is given by 6161.

A few more rules are needed for an unambigous definition.

- If an operators appears without fermions (as it is possible, e.g., for $F_{\mu\nu}$) it should appear right-most, so that the encoded fermions correspond to the left-most operators.
- In the case of a possible ambiguity, for instance $O_1 = (\bar{s}\gamma_\mu T^a P_L c)(\bar{c}\gamma^\mu T^a P_L b)$ corresponding to 03040405 6161 and $O_1 = (\bar{c}\gamma_\mu T^a P_L b)(\bar{s}\gamma^\mu T^a P_L c)$ corresponding to 04050304 6161 the "smaller" number, i.e. in this case 03040405 6161 should be used.

The third index corresponds to each term in Eq. (5):

`00` : $C_i^{(0)}(\mu)$

`01` : $C_{i,s}^{(1)}(\mu)$

`02` : $C_{i,s}^{(2)}(\mu)$

`10` : $C_{i,e}^{(1)}(\mu)$

`11` : $C_{i,es}^{(2)}(\mu)$

name	PDG code	two-digit number	name	PDG code	two-digit number
d	1	01	e	11	11
u	2	02	ν_e	12	12
s	3	03	μ	13	13
c	4	04	ν_μ	14	14
b	5	05	τ	15	15
t	6	06	ν_τ	16	16
$\sum_q q$		07	$\sum_l l$		17
$\sum_q qQ_q$		08	$\sum_l lQ_l$		18

Table 1: PDG codes and two-digit number identifications of quarks and leptons. The summations are over active fermions.

operator	number	operator	number	operator	number
1	30	T^a	50	δ_{ij}	70
P_L	31	$P_L T^a$	51	$P_l \delta_{ij}$	71
P_R	32	$P_R T^a$	52	$P_R \delta_{ij}$	72
γ^μ	33	$\gamma^\mu T^a$	53	$\gamma^\mu \delta_{ij}$	73
γ_5	34	$\gamma_5 T^a$	54	$\gamma_5 \delta_{ij}$	74
$\sigma^{\mu\nu}$	35	$\sigma^{\mu\nu} T^a$	55	$\sigma^{\mu\nu} \delta_{ij}$	75
$\gamma^\mu\gamma^\nu\gamma^\rho$	36	$\gamma^\mu\gamma^\nu\gamma^\rho T^a$	56	$\gamma^\mu\gamma^\nu\gamma^\rho \delta_{ij}$	76
$\gamma^\mu\gamma_5$	37	$\gamma^\mu\gamma_5 T^a$	57	$\gamma^\mu\gamma_5 \delta_{ij}$	77
$\gamma^\mu P_L$	41	$\gamma^\mu T^a P_L$	61	$\gamma^\mu \delta_{ij} P_L$	81
$\gamma^\mu P_R$	42	$\gamma^\mu T^a P_R$	62	$\gamma^\mu \delta_{ij} P_R$	82
$\sigma^{\mu\nu} P_L$	43	$\sigma^{\mu\nu} T^a P_L$	63	$\sigma^{\mu\nu} \delta_{ij} P_L$	83
$\sigma^{\mu\nu} P_R$	44	$\sigma^{\mu\nu} T^a P_R$	64	$\sigma^{\mu\nu} \delta_{ij} P_R$	84
$\gamma^\mu\gamma^\nu\gamma^\rho P_L$	45	$\gamma^\mu\gamma^\nu\gamma^\rho T^a P_L$	65	$\gamma^\mu\gamma^\nu\gamma^\rho \delta_{ij} P_L$	85
$\gamma^\mu\gamma^\nu\gamma^\rho P_R$	46	$\gamma^\mu\gamma^\nu\gamma^\rho T^a P_R$	66	$\gamma^\mu\gamma^\nu\gamma^\rho \delta_{ij} P_R$	86
$F_{\mu\nu}$	22	$G^a_{\mu\nu}$	21		

Table 2: Two-digit number definitions for the operators. T^a $(a = 1 \ldots 8)$ denote the $SU(3)_C$ generators, $P_{L,R} = \frac{1}{2}(1 \mp \gamma_5)$, and $(T^a)_{ij}(T^a)_{kl} = \frac{1}{2}(\delta_{il}\delta_{kj} - 1/N_c\, \delta_{ij}\delta_{kl})$, where i, j, k, l are colour indices.

`99` : total

The information about the order is given by a two-digit number xy, where x indicates $\mathcal{O}(\alpha^x)$ and y indicates $\mathcal{O}(\alpha_s^y)$, and 0 indicates $C_i^{(0)}$.

The Wilson coefficients can be provided either via separate new physics and SM contributions, or as a total contribution of both new physics and SM, depending on the code generating them. To avoid any confusion, the fourth entry must specify whether the given Wilson coefficients correspond to the SM contributions, new physics contributions or to the sum of them, using the following definitions:

`0` : SM

`1` : NPM

`2` : SM+NPM

The new Physics model is the model specified in the `BLOCK FMODSEL`.

The standard for each line in the block should thus correspond to the FORTRAN format

```
(1x,I8,1x,I4,3x,I2,3x,I1,3x,1P,E16.8,0P,3x,'#',1x,A),
```

where the eight-digit integer specifies the fermion content, the four-digit integer the operator structure, the two-digit integer the order at which the Wilson coefficients are calculated followed by the one-digit integer specifying the model, and finally the double precision number gives the real part of the Wilson coefficient.

Note that there can be several such blocks for different scales `Q`.

`BLOCK IMFWCOEF Q= ...`

The block `BLOCK IMFWCOEF` contains the imaginary part of the Wilson coefficients at the scale `Q`. The structure is exactly the same as for the `BLOCK FWCOEF`.

`BLOCK FOBS`

The block `BLOCK FOBS` contains the flavour observables. The structure of this block is based on the decay table in SLHA format. The decay is defined by the PDG number of the parent, the type of the observable, the value of the observable, the number of daughters and PDG IDs of the daughters.
The types of the observables are defined as follows:

- `1` : Branching ratio
- `2` : Ratio of the branching ratio to the SM value
- `3` : Asymmetry – CP
- `4` : Asymmetry – isospin
- `5` : Asymmetry – forward-backward
- `6` : Asymmetry – lepton-flavour
- `7` : Mixing
- `8` : ...

The standard for each line in the block should correspond to the FORTRAN format

```
(1x,I9,3x,I2,3x,1P,E16.8,0P,3x,I1,3x,I9,3x,I9,3x,...,3x,'#',1x,A),
```

where the first nine-digit integer should be the PDG code of the parent decaying particle, the second integer the type of the observable, the double precision number the value of the observable, the next integer the number of daughters, and the following nine-digit integers the PDG codes of the daughters. It is strongly advised to give the descriptive name of the observable as comment.

`BLOCK FOBSERR`

The block `BLOCK FOBSERR` contains the theoretical error for flavour observables, with the structure similar to `BLOCK FOBS`, where the double precision number for the value of the observable is replaced by two double precision numbers for the minus and plus uncertainties.

In a similar way, for every block, a corresponding error block with the name `BLOCK FnameERR` can be defined.

`BLOCK FOBSSM`

The block `BLOCK FOBSSM` contains the SM values of the flavour observables in the same format as in `BLOCK FOBS`. The given SM values may be very helpful as a comparison reference.

BLOCK FFORM

The block BLOCK FFORM contains the form factors for a specific decay. This decay should be defined as in BLOCK FOBS, but replacing the type of the observable by the number of the form factor. It is essential here to describe the variable in the comment area. The dependence on q^2 can be specified as a comment. A more unambiguous definition will be given elsewhere.

4.3 CONCLUSION

The interplay of collider and flavour physics is entering a new era with the start-up of the LHC. In the future more and more programs will be interfaced in order to exploit maximal information from both collider and flavour data. Towards this end, an accord will play a crucial role. The accord presented specifies a unique set of conventions in ASCII file format for most commonly investigated flavour-related observables and provides a universal framework for interfacing different programs.

The number of flavour related codes is growing constantly, while the connection between results from flavour physics and high p_T physics becomes more relevant to disentangle the underlying physics model. Using the lessons learnt from the SLHA, we hope the FLHA will prove useful for studies related to flavour physics. It is planned to update/correct the FLHA after more experience with its application will have been gathered.

ACKNOWLEDGEMENTS

The work of S.H. was partially supported by CICYT (grant FPA 2007–66387). Work supported in part by the European Community's Marie-Curie Research Training Network under contract MRTN-CT-2006-035505 'Tools and Precision Calculations for Physics Discoveries at Colliders'. The work of T.G. is supported in part by the Grant-in-Aid for Science Research, Japan Society for the Promotion of Science, No. 20244037.

Part II
TUNING

5. STATUS OF RIVET AND PROFESSOR MC VALIDATION & TUNING TOOLS [10]

The Rivet [46] package for MC generator validation and the Professor [47] system for generator tuning have become established tools for systematically verifying event simulations and optimising their parameters, where required and physically sensible. In this short report, we summarise the status and development of these tools.

5.1 Rivet

Rivet is an MC *validation* tool: it encodes MC equivalents of an increasingly comprehensive set of HEP collider analyses which are useful for testing the physics of MC generators. Rivet does not itself produce tunings, but provides a standard set of analyses by which to verify the accuracy of a given generator with a given tuning.

Several fundamental design principles have been derived from the experience on Rivet's predecessor system, HZTool [48, 49], and from iteration of the Rivet design:

- No generator steering: Rivet relies entirely on being provided, by unspecified means, with events represented by the HepMC [50] event record.
- No generator-specific analyses: all Rivet analyses are specifically not allowed to use the generator-specific portions of the supplied event records. Apart from a few limited (and deprecated) exceptions, all analyses are based solely on physical observables, i.e. those constructed from stable particles (those with status 1) and physical decayed particles (those with status 2).
- Rivet can be used either as a C++ library to be interfaced with generator author or experiment analysis frameworks, or as a command line tool (which itself makes use of the library interface). This is an example of the general philosophy to keep things simple and flexible, since we do not *a priori* know every task to which our tools will be employed.

Internally, Rivet analyses are based on a comprehensive set of calculational tools called *projections*, which perform standard computations such as jet algorithms (using FastJet [51]), event shape tensors, and a variety of other standard tasks. Use of projections makes analysis code much simpler, encapsulates any complexities arising from the ban on use of event record internal entities (the summation of photon momenta around charged leptons during Z-finding is a good example, see Section 14.), and is more efficient than just using library functions, due to a complex (but hidden) system of automatic result caching.

Users can write their own analyses using the Rivet components and use them via the Rivet API or command-line tool without re-compiling Rivet, due to use of an analysis "plugin" system. Separation between generator and Rivet on the command-line is most simply achieved by using the HepMC plain text `IO_GenEvent` format via a UNIX pipe (a.k.a. FIFO): this avoids disk access and writing of large files, and the CPU penalty in converting event objects to and from a text stream is in many cases outweighed by the general-purpose convenience. For generator-specific use of Rivet, the programmatic interface allows HepMC objects to be passed directly in code, without this computational detour. A sister tool, AGILe [52], is provided for convenience control of several Fortran-based generators, with command-line and parameter file based run-time steering of generator parameters. This is a convenient tool when exploring generator parameter space as part of a tuning.

Reference data for the standard analyses is included in the Rivet package as a set of XML files in the AIDA format. After several years of re-development as part of the CEDAR [52] project, the HepData [53]

[10] Contributed by: A. Buckley, J. M. Butterworth, H. Hoeth, H. Lacker, J. Monk, H. Schulz, J. E. von Seggern, F. Siegert

database of HEP experimental results can be used to directly export data files usable by Rivet from its Web interface at `http://hepdata.cedar.ac.uk/`. Analysis histograms are directly booked using the reference data as a binning template, ensuring that data and MC histograms are always maximally consistent.

Rivet is in use within the MC generator development community, particularly in general-purpose shower MC programs, and the LHC experimental community, for MC validation and MC analysis studies which do not require detector simulation.

5.11 Recent developments

Rivet 1.1.3 was released during the first week of this Les Houches workshop, in June 2009: this release includes many new analyses and fixes to existing analyses. Since the workshop, a huge number of extra improvements and developments have taken place in the run-up to the 1.2.0 release. Aside from many technical improvements, and the addition of a large number of QCD analyses (primarily for minimum bias and multi-jet physics) the major conceptual developments have been an emphasis on automated testing and validation of Rivet code, and the removal of hard-coded cross-section normalisations whenever possible. This latter step required more development than may be expected, due to the separation of generator and analysis: the HepMC record had to be enhanced to store cross-section information in a way which can be passed to Rivet. This has now been done, and recent versions of major generators such as Herwig++ [19, 21], Sherpa [22, 23], and Pythia 8 [20], support this HepMC feature "out of the box". AGILe's generator interfaces have also been updated to write cross-section information into their HepMC output. Determining scaling K-factors where required is now performed via post-processing scripts, which automatically support common approaches to constraining this remaining degree of freedom.

A technical development, but one worth mentioning, has been the emphasis on making Rivet analyses "self-documenting": each analysis has a structured set of metadata specifying name, authors, run conditions, a description, etc., which is used to provide interactive help, HTML documentation, and a reference section in the Rivet manual.

At the time of writing, the final stage of systematic validation of Rivet for the 1.2.0 release is underway. The validation scripts used for this checking will henceforth be included in automatic build tests, to ensure that future developments do not unexpectedly change existing analysis functionality. The final major stage of development is the upgrade of Rivet's histogramming and data analysis code, which is currently rather basic. The upgrade will enable statistically accurate combination of runs, allowing for greater parallelisation of Rivet analyses which require large event statistics.

5.2 Professor

The Professor system builds on the output of MC validation analyses such as those in Rivet, by optimising generator parameters to achieve the best possible fit to reference data. The main description of Professor's details is found in reference [47], and we will not significantly replicate it here, except in the most high-level sense.

Fundamentally, generator tuning is an example of the more general problem of optimising a very expensive function with many parameters: the volume of the space grows exponentially with the number of parameters and the CPU requirements of even a single evaluation of the function means that any attempt to scan the parameter space will fail for more than a few parameters. Here, the expensive function is running a generator with a particular parameter set to recreate a wide range of analysis observables, using a package such as Rivet. The approach adopted by Professor is to parameterise the expensive function based on a non-exhaustive scan of the space: it is therefore an approximate method, but its accuracy is systematically verifiable and it is currently the best approach that we have.

The parameterisation is generated by independently fitting a function to each of the observable bin values, approximating how they vary in response to changes in the parameter vector. One approach to

fitting the functions would be to make each function a linear combination of algebraic terms with n coefficients α_i, then to sample n points in the parameter space. A matrix inversion would then fix the values of α_i. However, use of a pseudoinverse for rectangular matrices allows a more robust coefficient definition with many more samples than are required, with an automatic least-squares fit to each of the sampled "anchor points": this is the method used by Professor. By aggregating the parameterisations of all the observable bins under a weighted goodness of fit measure – usually a heuristic χ^2– a numerical optimisation can be used to create an "optimal" tune. In practice, many different semi-independent combinations of MC runs are used to provide a systematic handle on the degree of variation expected in tunes as a result of the inputs, avoiding the problem that a single "maximum-information" tune may not be typical of the parameter space.

The first application of Professor, due to its popularity and fairly well-understood steering parameters, was the Pythia 6 MC generator [18]. This was tuned in reference [47], using both of the available parton shower and multi-parton interaction (MPI) models, to data from LEP, SLD, and Tevatron Runs I and II. It was found that the parameterisation method worked well in all cases, and a range of systematic methods and tools were developed to check the accuracy of the approximations, such as line-scans through the parameter space. It was found that a sensible maximum number of parameters to be included in a single tune was ~ 10, and hence, there being ~ 20 Pythia 6 parameters relevant to the studied observables, we separated the tune into an initial stage of final state shower and fragmentation tuning using e^+e^- observables, and then a second stage based on tuning initial state parton shower and MPI parameters to best describe hadron collider data. In these tunes, a quadratic parameterisation was used throughout, this being the simplest suitable function to account for parameter correlations.

5.21 Recent developments

The main development since the initial publication and use of Professor has been the application to more MC tunings. A first extra application was to use the same χ^2 weightings as for the Pythia 6 tune to obtain additional tunings of Pythia 6 with different PDFs. The results of this study, shown at the PDF4LHC meeting in April 2009, indicated that modified leading order PDFs, developed for use with LO MC generators and characterised by a larger than normal low-x gluon component, drive statistical generator tunings in a physically expected direction: the main effect was to increase the screening of MPI effects proportional to the size of the gluon PDF at low x values.

Moving away from Pythia 6, substantial tuning effort has been expended with the Jimmy [54, 55] MPI simulation (used with the Fortran Herwig 6 code [17]), and the Pythia 8 and Sherpa generators. In the case of Pythia 8, the default fragmentation settings are now those fixed by Professor, and use of Professor has identified a problem with description of underlying event (UE) observables in QCD events: this problem is being addressed. An extremely useful tool in this study was the `prof-I` GUI, which makes use of the Professor parameterisations not for minimisation, but for interactive mimicking of generator responses to parameter changes: this tool makes exploration of speculative tuning ideas easily testable, and helped to verify that no combination of parameters would achieve the desired effect in Pythia 8's description of UE observables.

Another major effort has been the use of Professor to tune and develop the Sherpa generator's simulation of hadronisation and soft initial-state QCD physics. This collaboration has helped to rapidly iterate model improvements and debugging, due to the fast turnaround of tune information.

Professor is additionally being used within the ATLAS, CMS, and LHCb LHC experimental collaborations for tuning studies of the main generators used for their MC simulation, and in plans for re-tuning to first LHC data at new centre of mass energies.

Development of the Professor framework and application to different implications of tuning continues: the next contribution details how ensembles of tunes created by Professor may be used for estimations of tune uncertainties in MC predictions. Other suggested extensions to optimisation of observable definitions

or parameterisation of other expensive functions, e.g. observables in SUSY parameter space, remain open to exploration.

Acknowledgements

The Rivet and Professor collaborations acknowledge support from the EU MCnet Marie Curie Research Training Network (funded under Framework Programme 6 contract MRTN-CT-2006-035606) for financial support and for many useful discussions and collaborations with its members. A. Buckley additionally acknowledges support from the Scottish Universities Physics Alliance (SUPA); H. Schulz acknowledges the support of the German Research Foundation (DFG).

6. QUANTITATIVE ERROR ESTIMATION IN MC TUNES [11]

Recent developments in Monte Carlo generator tuning have led to more robust and general-purpose "optimal" tunes to existing data, and there is a clear hope that when existing data is well-described, extrapolations to future collider energies will also be reliable. However, especially in the area of soft QCD, generator predictions are never expected to be exactly accurate: there is no single "best" tune for a given model, but rather one or more *regions* of parameter space which contains reasonable tunes. The size of these regions reflects the degree of constraint which existing data is able to place on the model – discrete choices of model are also crucial to obtain a true sense of the total uncertainty on any generator prediction. In this contribution we present studies of how the mechanisms used for systematic generator tuning can be used to quantitively estimate the contributions to the uncertainty of MC predictions from several sources.

A key example of generator uncertainty is the extrapolation of minimum bias and underlying event QCD physics to LHC design energies, i.e. $\sqrt{s} \gg 2$ TeV. The physics that drives the rise in activity is a combination of the non-perturbative total pp cross-section; the non-perturbative physics of beam remnants, diffraction, and multiple partonic scattering; and perturbative low-p_T QCD scattering. Accordingly, the most-used models are highly phenomenological and have many tweakable parameters: one interesting consequence of first LHC data in the multi-TeV energy regime will be the testing of whether these models extrapolate in agreement with nature. The expectation is that some models will fail the test!

Previous studies of prediction uncertainty have been necessarily qualitative and subjective, since MC tunes have themselves been somewhat approximate affairs. The most common approach to assigning a systematic uncertainty has been to compare predictions from two different models, such as Pythia and Phojet in the case of Minimum Bias/Underlying Event extrapolation. Discrete choices of model remain a major source of uncertainty, even when the range of historic Pythia tunes with a Tevatron-excluded energy extrapolation is excluded[12] – we encourage all extrapolated studies to make use of as many distinct (non-excluded) models as possible. In this contribution, however, we describe how to quantitively assess major sources of uncertainty arising from the tuning process itself, i.e. the reasonable scatter to be expected around a tune for a discrete model. Our baseline for this approach is the Professor tuning system, which we now summarise.

6.1 MC tuning with Professor

The "Professor" approach to MC tuning constitutes both a numerical method and a suite of tools which implement it. Fundamentally, Professor attempts to parameterise expensive functions – the bin values in a set of MC observables – by least-squares fitting of the parameterisation coefficients. The least-squares minimisation is made more approachable by use of the pseudoinverse method, implemented via a matrix singular value decomposition. Armed with a fast analytic model of how every bin of a large set of observables will respond to variations of the generator parameters, numerical optimisation of the generator's fit to reference data may be efficiently computed. A detailed description may be found in reference [47].

The benefit of this approach is clear for more than two parameters: Professor requires as input the values of observables for a moderately large number of MC runs distributed suitably in the generator parameter space, each point in the space perhaps requiring $\mathcal{O}(48)$ CPU hours to complete. A serial optimisation approach such as Markov chain sampling would hence require thousands of CPU days to have a chance of converging, if the generator itself is not batch-parallelised. Professor, given sufficiently large batch computing resources, can trivially parallelise the generation of the input MC points for any generator and thereafter complete the parameterisation and fit optimisation in negligible time, allowing for scaling to higher numbers of tune parameters than could be attempted by methods which require iteration of the

[11] Contributed by: A. Buckley, H. Hoeth, H. Lacker, H. Schulz, J. E. von Seggern

[12] This includes the default Pythia tune, the ATLAS MC08 tune, and (by construction) the Tune *T series from Rick Field. Essentially, any setup with PARP$(90) < 0.2$ is in contradiction with Tevatron data.

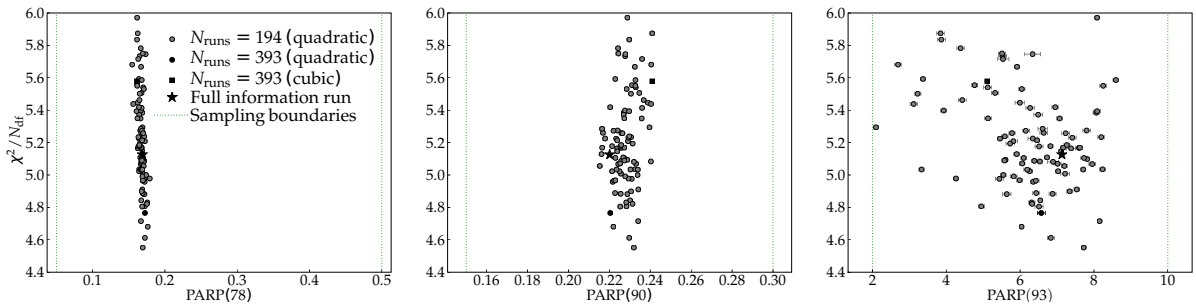


Fig. 5: A scatter plot of χ^2 vs. parameter value for a set of Professor run combinations on three Pythia6 parameters. Implicitly, this projection of tune parameter vectors on to parameter axes gives a qualitative measure of whether or not a parameter is well-constrained: these parameters become increasingly ill-defined from left to right. The different markers represent different degrees of oversampling, with the star representing the maximum information run — the points are for the same run combinations in all three scatter plots.

time-limiting step.

6.2 Qualitative error estimation in Professor

An important feature of the Professor method is that it has always allowed for *qualitative* assessment of the tune robustness. A per-bin parameterisation in p parameters will require a minimum number of MC runs, $N_{\text{min}}^{(p)}$, for the least-squares pseudoinversion to be performed. For robustness it is advisable to oversample this minimum requirement by a factor $\mathcal{O}(3)$ (or more, especially for large p) such that a tune will in fact use $N \gg N_{\text{min}}^{(p)}$ input runs. Additionally, since the parameterisation and optimisation steps are fast, we take the opportunity to make many such overconstrained tunes by in fact sampling an even greater number of runs, $N_{\text{sampled}} \gg N$. We can then randomly sample a large number of mostly-independent N-run tunes to obtain an ensemble of reasonable tunes – again, this step can be trivially batch-parallelised. The spread of this tune ensemble as projected on each parameter has been used in several Professor MC tunes as a heuristic for determining whether a parameter is well or poorly constrained, for detecting parameterisation problems, and for ensuring that the "maximum information" tune is typical of the ensemble. Several other checking methods, such as eigenvector line scans, are also used to ensure that the details of the tune, and particularly the generator parameterisation, are reliable.

To make a prediction of an observable for which there is no reference data (this may be an energy extrapolation, or simply an unmeasured feature at existing energies), the simplest approach is of course to run the generator with the obtained tune(s) and compute the observable. Using the many tunes resulting from the different run combinations will give a spread in the observable prediction, reflecting one part of tune uncertainty. In practice, since we can build a fast parameterisation of the generator behaviour (in fact many of them) on the unfitted observable as part of the main Professor tuning process, this offers a much faster turnaround than processing another large (perhaps *very* large) set of generator runs – with the proviso that the parameterisations are of course non-exact.

6.3 Sources of tune uncertainty

Before making this method more quantitive, we now consider the various sources of uncertainty in the procedure outlined so far. This will help us to understand which sources of uncertainty are computationally controllable and which will have to, for now, remain more nebulous. These main sources of tuning error are as follows:

1. Error on experimental reference data.

2. Statistical error on the MC at the anchor points from which the parameterisation is constructed.
3. Systematic limitations of the parameterising function to describe bin responses to parameter variations – pathological MC parameters with discontinuous or critical behaviour are particularly hard to generically parameterise, since a Jacobian transformation to a suitable meta-parameter is not always easily available.
4. Choice of run combinations to make the parameterisation.
5. Goodness of fit definition, including both the type of GoF measure and the choice and relative weighting of different data.
6. Reasonable minimiser scatter within the χ^2 valley containing the optimal tune point for a given parameterisation. Note that this cannot be completely disentangled from the role that error sources 1, 2 and 3 play in defining the χ^2 valley.
7. Limitations of the parameterisation(s) used to compute the parameterised MC value in extrapolated/unfitted observables. Of course, this error doesn't exist if the less convenient strategy of re-running the MC generator is used.
8. For completeness, we again highlight the systematic error associated with the discrete physics model being tuned: the total error is far from complete without considering more than one viable model. Within a given model there are also systematic uncertainties, some of which may be quantified, e.g. cross-section integration uncertainty and PDFs: the second of these is particularly quantifiable due to the existence of error or replica PDF sets, themselves expressing reasonable variations in PDF fitting.

Note that, for example, these sources of uncertainty such as the variation between members of the ensemble of run combinations are not unique errors introduced by the Professor approach: failing to test different run combinations does not *eliminate* the error associated with the choice of anchor runs used! A similar, but unquantifiable, error exists for any form of manual tuning.

6.4 Construction of tune uncertainty confidence belts

Our approach to quantifying the uncertainties from (combinations of) the sources listed in the previous section is to construct central confidence belts for observable bins from the various ensembles of tune results, parameterisations, etc. we have described. Explicitly, given a large number of reasonable and equivalent predictions for an observable bin value, we construct a band of given P-value as being the region containing fraction P of predictions, with equal fractions above and below.

There are many ways to construct such ensembles – for the purposes of this study we identify three:

Combination error: The ensemble from which we construct the confidence belt is simply the ensemble of predictions from different run combinations. This will hence represent the variation due to error sources 1, 2, and 4 – the other sources of uncertainty exist, but are not quantified by this approach[13].This approach requires that correlations between different run combinations are small, which is ensured by the $N_{\text{sampled}} \gg N$ requirement.If parameterisation (as opposed to explicit MC runs) is used for the translation of this tune ensemble into bin value predictions, then source 7 also applies. This can be included into the band construction by using many parameterisations, again constructed from run combinations. Different parameterisations should be used for minimisation and prediction to avoid reinforcing parameterisation systematics.

Statistical error: The obvious failure of the combination error approach is that error source 6 – the measure of reasonable tune variation within the χ^2 valley – is left unquantified. Hence it will be no

[13]MC error is currently not *explicitly* propagated into Professor's fit measure, due to stability problems: it enters implicitly via the statistical scatter of MC samples. This is being remedied.

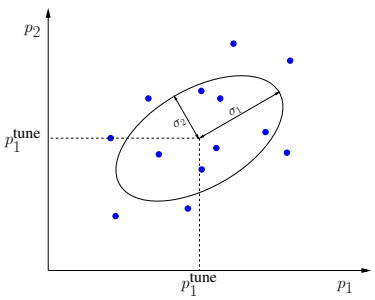


Fig. 6: Two dimensional illustration of the parameter point sampling used for the statistical uncertainty estimate. We exploit the covariance matrix returned by the minimiser for a Gaussian sampling from the corresponding p-dimensional hyper-ellipsoid. The σ_i ($i = 1 \ldots p$) are the eigenvalues obtained from an eigen-decomposition of the covariance matrix.

surprise that our "statistical" error band is constructed explicitly from an ensemble of samples from this valley. This is obtained in the simplest case by only using the maximum information Professor tune – that which is constructed from all the available MC runs. The covariance matrix of the parameters in the vicinity of the tune point is obtained – in principle directly from the parameterisation, in practice from the minimiser – and used to define a rotated hyper-Gaussian probability distribution in the parameter space. Sampling parameter points from this distribution gives us another ensemble of tune points and, as for the combination error, they can be mapped into observable predictions either by direct MC runs or by using one or many constructed parameterisations. In practice, we take advantage of the combined potential of the hyper-Gaussian sampling and MC parameterisation to build a confidence belt from $\mathcal{O}(10,000)$ samples. The quantified sources of error are hence 1, 2, 6, and 7.

These approaches to error band building will be used in the next section to construct sample error bands for underlying event predictions. Although not currently implemented in Professor, we also highlight the most complete form of quantifiable error band within the Professor approach:

Combined error: This extension is an obvious fusion of the above two ensemble/band constructions: as in the combination error, we construct an ensemble of points from run combinations, then for each run-combination point we construct a statistical ensemble. The combined ensemble of tune ensembles, and a variety of parameterisations to transform them into predictions will lead to error bands quantifying error sources 1, 2, 4, 6, and 7.

The remaining error sources are 3, 5 and 8: limitations of the parameterisation, the observable weights/goodness of fit definition, and – most importantly – the uncertainty due to different physics models. These missing systematics remain qualitative in this scheme, and reliable MC predictions should take care to include estimates of their influence, albeit in a more ad hoc fashion than for the more statistically-induced errors.

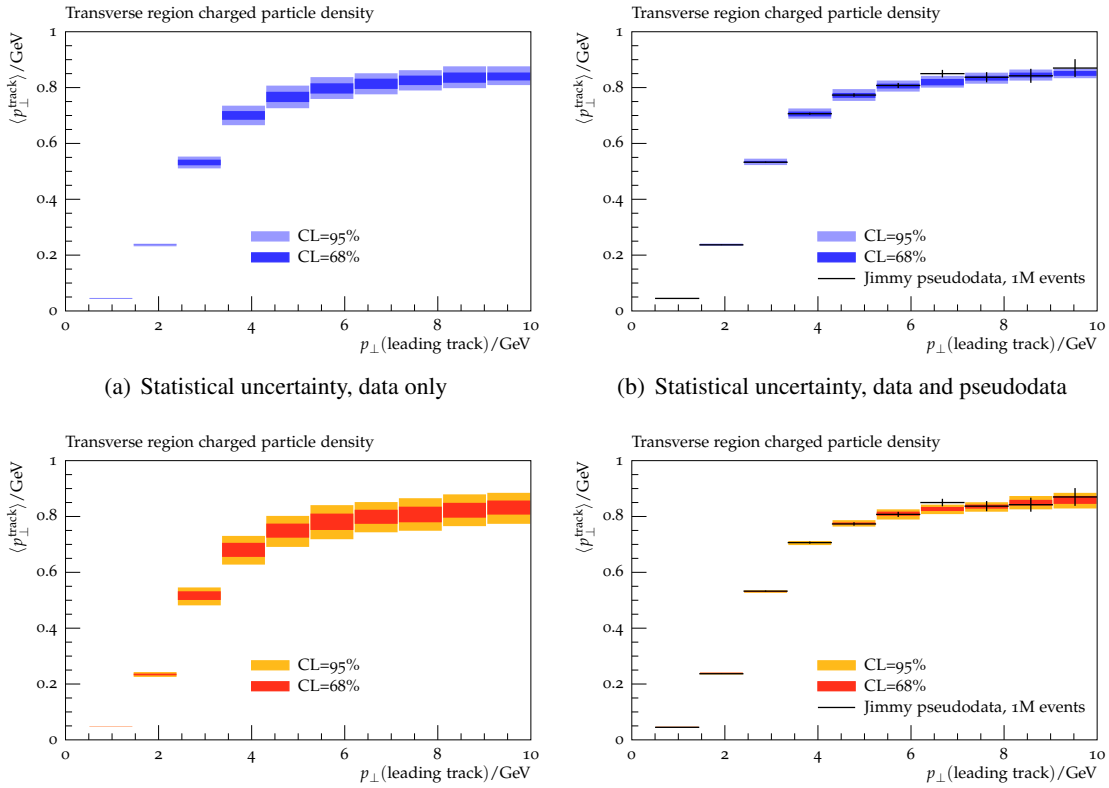
6.5 Results

We now briefly present results using the first two confidence belt definitions presented above – ongoing work is addressing the "combined" belt definition and inclusion of PDF uncertainties.

Our exploration is based on tunes of the Jimmy [54, 55] MC generator, which simulates multiple parton interactions (MPI) for Herwig 6 [17], because it has only two relevant parameters, PTJIM and JMRAD(73)[14] – the frugality with parameters makes Jimmy an ideal "toy model" testbed, while remaining phenomenologically relevant. As a Jimmy-like MPI model is ruled out by Tevatron data [56], we fix a dependence of PTJIM on the centre of mass energy with the same ansatz as used in Pythia 6 [18]:

$$\text{PTJIM} = \text{PTJIM}_{1800} \cdot \left(\frac{\sqrt{s}}{1800\ \text{GeV}} \right)^{0.274}, \tag{7}$$

[14] We treat the inverse radius-squared of the protons, JMRAD(73), to be identical to that of the anti-protons, JMRAD(91), and ignore the interplay with ISR parameters in Herwig 6 itself.



(a) Statistical uncertainty, data only

(b) Statistical uncertainty, data and pseudodata

(c) Combination uncertainty, data only

(d) Combination uncertainty, data and pseudodata

Fig. 7: "Statistical" and "combination" error bands for transverse N_{ch} flow at 7 TeV before and after adding 1M events of pseudodata of this observable (black markers) to the tuning. The error bands are calculated from the central 95 (68)% of the binvalues of an ensemble of 10000 histograms each.

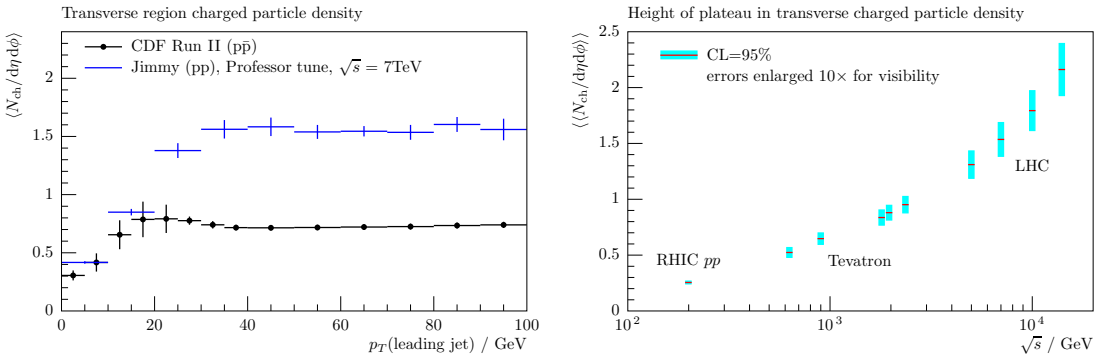
where PTJIM_{1800} is the value of PTJIM at the reference scale $\sqrt{s} = 1800$ GeV and is the $p_\perp^{\text{min}}$ parameter actually used in the tuning process. Furthermore, we use the MRST LO* PDF set [57] and use Tevatron data from CDF [58–60] and DØ [61] as a tuning reference. A more complete tune would include the exponent of the $p_\perp^{\text{min}}$ energy dependence, but for toy-study purposes we here fix it to a value consistent with known energy extrapolation fits [47, 62].

In Figures 7(a) and 7(c), the statistical and combination error band definitions are shown for a 7 TeV underlying event observable – N_{ch} flow transverse to the leading track (track with the largest transverse momentum in an event), as a function of leading track p_T– computed in Rivet [46], based on the fit to Tevatron reference data. The error due to variation of run combinations (error source 4) is notably somewhat larger than the scatter of points in the χ^2 valley (error source 6), indicating that in the Jimmy model the parameters have a strong influence on this observable, and are hence well-constrained. In Figures 7(b) and 7(d), similar band constructions are shown, but in this case the same Rivet analysis has been used to simulate the effect of adding 1M events of UE data at 7 TeV into the fit: the size of both error bands is reduced, as expected.

6.51 Effect of extrapolation

Finally, we consider systematically how error bands constructed in this way behave as extrapolations are taken further from the region of constraining data. In this case, since the computational requirements are significant, we only consider the "statistical" error band.

We use the transverse region N_{ch} density UE observable, evaluated at a range of ten centre of mass energies, $\sqrt{s}_i$, between 200 GeVand 14 TeV. We then apply the following procedure,



(a) N_{ch} vs. $p_{\perp}^{\text{lead}}$, plateau

(b) N_{ch} vs. $p_{\perp}^{\text{lead}}$, mean of plateau vs. $\sqrt{s}$

Fig. 8: Transverse region charged particle density. (a): the typical plateau observed and (b): mean-height of that plateau as a function of $\sqrt{s}$.

$\sqrt{s}$/TeV	0.2	0.63	0.9	1.8	1.96	2.36	5.0	7.0	10.0	14.0
$p_{\perp}^{\text{lead, min}}$/GeV	10	30	30	30	40	40	40	40	40	40
$p_{\perp}^{\text{lead, max}}$/GeV	30	70	80	80	90	110	110	120	150	160

Table 3: Definition of plateau regions ($p_{\perp}^{\text{lead}}$) used in the extrapolation study.

- Construct the maximum-information parameterisation of the generator response for the N_{ch} UE observable, $f(\sqrt{s_i})$
- Produce an ensemble of 10000 histograms, H_i, of the observable shown in Figure 8(a) (blue line) using the corresponding $f(\sqrt{s_i})$ and points sampled using the procedure illustrated in Figure 6
- Calculate the mean height of the N_{ch} plateau, M_i, for each of the H_i
- Construct a 95% central confidence belt, $\text{CL}(\sqrt{s_i})$ from the M_i

In Figure 8(b) the $\text{CL}(\sqrt{s_i})$ are drawn. We observe a very tight confidence belt for the energy region of the Tevatron experiments, while the confidence belt becomes wider for extrapolation to LHC energies. The definitions of the plateau regions used can be found in Table 3.

It is notable that these bands are narrow – sufficiently so that they have been visually inflated by a factor of 10 in the figure. While this reflects good stability in the tuning system, it is probably an underestimate of the true model uncertainty. A more complete study will include the exponent of energy extrapolation in equation (7) in the tune, since this may be a dominant effect in the errors for this particular observable and its inclusion will give more freedom for various features of the model to balance against each other: the toy tuning of the energy-dependent Jimmy model shown here is probably too restrictive to accurately represent the full range of variation allowed by the model, but serves as an indication of the extent to which such studies can be systematised.

6.6 Conclusions

We have catalogued a set of sources of uncertainty which either explicitly or implicitly contribute to any tune of a MC event generator, and presented a systematic approach to quantifying many of these uncertainties using the fast MC parameterisations and natural tune ensembles which arise from the Professor tuning approach. Example results have been shown, which exhibit some expected behaviours, such as the shrinking of error bands on adding new reference data in new areas of parameter space and the

blow-up of error bands as predictions venture further into unconstrained regions.

Several things should be emphasised: first and most important is that this approach does not catch all sources of error. We have presented results from two definitions and have proposed a third, more comprehensive measure, but still variations such as PDF errors and discrete model variations need to be included. However, with the increased usage of systematic tuning methods, variations between models in UE observables are not as substantial as once they were – statistical errors are a non-negligible factor in assessing the reliability of phenomenologically-based MC predictions.

The approach taken here has many obvious parallels in the world of PDF errors, with our approach having a good deal of overlap with the MC replica set approach of the NNPDF collaboration [63] as contrasted with the eigenset approach of the CTEQ and MRST/MSTW collaborations. While replica sets have the advantage of a more direct statistical uncertainty interpretation (although we do not have the option of the parameterisation-freedom exibited by the NNPDF use of neural networks), there is the pragmatic issue that $\mathcal{O}(10)$ representative error tunes would be more usable than $\mathcal{O}(10000)$ equivalent tunes. Whether such a concept can be statistically constructed remains to be seen – for now the Perugia-soft/hard tunes remain the obvious tool of choice.

Acknowledgements

Our particular thanks goes to Luigi del Debbio and Richard Ball for discussions about statistical coverage and MC error estimation. The Professor collaboration acknowledges support from the EU MCnet Marie Curie Research Training Network (funded under Framework Programme 6 contract MRTN-CT-2006-035606) for financial support and for many useful discussions and collaborations with its members. A. Buckley additionally acknowledges support from the Scottish Universities Physics Alliance (SUPA); H. Schulz acknowledges the support of the German Research Foundation (DFG).

7. MATRIX ELEMENT CORRECTIONS AND PARTON SHOWER MATCHING IN INCLUSIVE Z PRODUCTION AT LHC[15]

7.1 INTRODUCTION

In this paper we compare Pythia 6 [18], AlpGen [64] and Sherpa [23, 65] in inclusive Z production at LHC, with 10 TeV center of mass energy. A disagreement in various observables, especially the Z transverse momentum, p_T, between AlpGen and Pythia 6 was recently reported in [66]. Namely, the shape in the low Z p_T region was significantly different when comparing AlpGen showered with Pythia 6 and matrix element corrected Pythia 6. This was traced back to a change in shape in Pythia 6 when matrix element corrections are switched off (this setup is used when showering AlpGen events, because higher multiplicity AlpGen samples already provide higher order corrections). This is summarized by the Z p_T plot shown in Fig. 9. A bug in Pythia 6 causing this behavior was later corrected; here we present results obtained with the new version of Pythia 6.

We also compare different Pythia 6 tunes, checking both the virtuality ordered and the newer transverse momentum ordered shower.

Using Pythia 6 as a reference we check the performances of the matching in AlpGen and Sherpa. Since Pythia 6 is fully corrected for first order real emission, we expect agreement with AlpGen and Sherpa when those are configured to include matrix elements for Z plus up to one additional parton.

Results presented in this work are obtained using Pythia 6 versions 6.411 and 6.421 (the last contains the bug fix for the matrix element corrections problem), AlpGen 2.13 and Sherpa 1.2.0.

7.2 ANALYSIS SETUP

All generators have been processed through the same analysis, selecting events with a Z with mass between 66 GeV and 116 GeV, requiring the two leptons to have p_T greater than 20 GeV and pseudorapidity between -2.5 and 2.5. Jets are reconstructed with the $k_\perp$ algorithm, with a radius of 0.4 and a minimum p_T of 30 GeV. In order to decouple effects due to multiple interactions, hadronization and QED radiation off leptons, these have been switched off for all the generators used.

The analyses were carried out using Rivet (see Section 5.).

[15]Contributed by: P. Lenzi, P. Skands

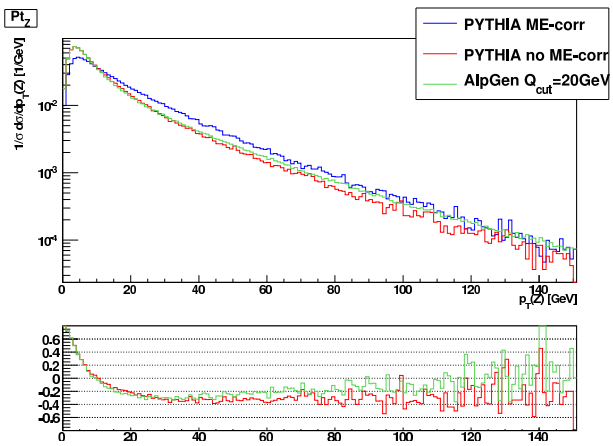


Fig. 9: Z transverse momentum spectrum for Pythia 6 (corrected and uncorrected) and for AlpGen plus Pythia 6 as in [66]. AlpGen plus Pythia 6 follows uncorrected Pythia 6 at low p_T. The change in shape of uncorrected with respect to corrected Pythia 6 was responsible for the disagreement.

7.3 MATRIX ELEMENT CORRECTIONS IN PYTHIA

Matrix element corrections in Pythia 6 are described in detail in [18]. Here we just remind the basics of the procedure. Pythia 6 modifies the shower in two steps: first the starting scale is raised to the hadronic center of mass energy (power shower configuration) so that any hard emission from the shower is kinematically possible; then the probability for the first emission (which is also the hardest in Pythia 6) is re-weighted to include the first order real emission contribution. The reweighting factor is calculated as the ratio between the exact real emission matrix element and the splitting function used in the shower.

The effect of matrix element corrections in Pythia 6.411 is summarized in Fig. 10 for Z transverse momentum spectrum. The three curves are obtained with full matrix element corrections, with power shower only, and with uncorrected shower with starting scale at the Z mass. The relative difference with respect to the ME corrected result is shown in the lower plot. The shape of the spectrum changes significantly when matrix element corrections are switched off, even at low p_T (the relative difference plot does not flatten as p_T goes to 0). This is unexpected, as the matrix element effect is supposed to go to zero at low p_T, where the description from the shower is already accurate. In fact, the splitting function used in the shower is the approximation for $p_T \to 0$ of the of the exact matrix element for emission of one parton, so no difference between the two descriptions is expected in that region.The effect of matrix element corrections is illustrated in Fig. 11 for the newer 6.421 version of Pythia 6. The same collection of curves as for version 6.411 is shown. With the newer version the deactivation of matrix element corrections has a very small effect at low p_T, as expected. A bug was found and corrected between the two versions which was causing the behavior observed in the earlier version.

7.4 PYTHIA TUNES

We compared various tunes in Pythia 6.421 on the basis of three observables, the Z transverse momentum, the jet multiplicity and the leading jet transverse momentum. In all of our comparisons we switched off the multiple interactions simulation, which means that we are comparing the effect of different tunings on the hard event simulation and on the shower.

We compared the following tunes: DW [67], D6T [67], Pro-Q0 [68], Perugia0 [69]. The virtuality

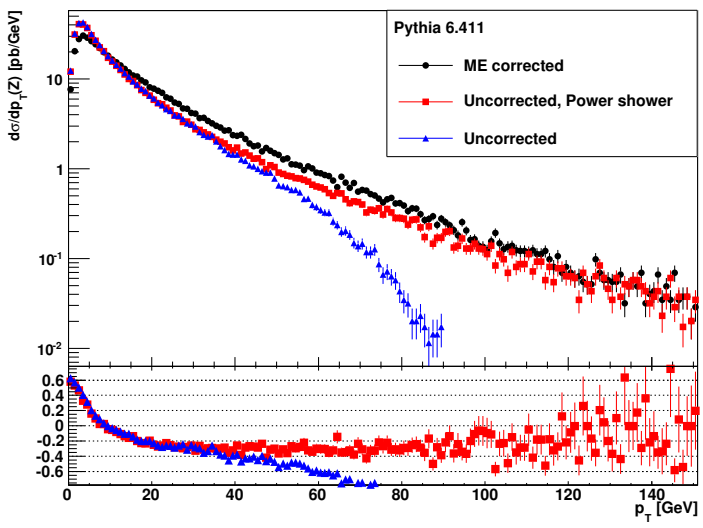


Fig. 10: Z transverse momentum spectrum for three different matrix element correction settings in Pythia 6.411: with matrix element corrections, with power shower only, without matrix element corrections. The relative difference with respect to the matrix element corrected curve is shown in the lower plot.

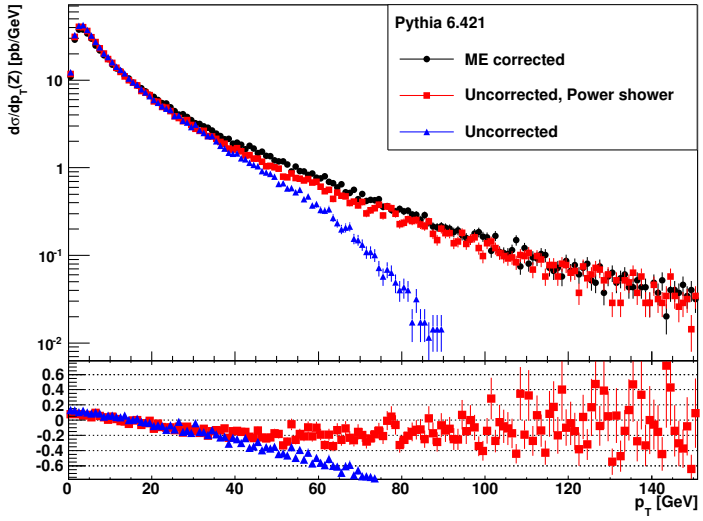


Fig. 11: Z transverse momentum spectrum for three different matrix element correction settings in Pythia 6.421: in this new version, deactivating matrix element corrections has a small effect at low p_T, as expected.

ordered shower is used in the first three, while the last uses the p_T ordered shower.

Z transverse momentum spectra obtained with the four mentioned settings are shown in Fig. 12. The agreement is generally good, tune DW and D6T agree fairly well on the whole spectrum; some

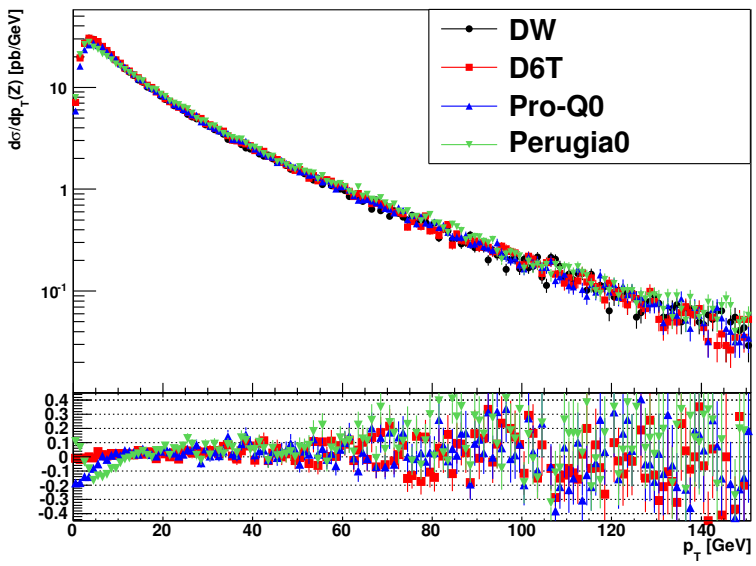
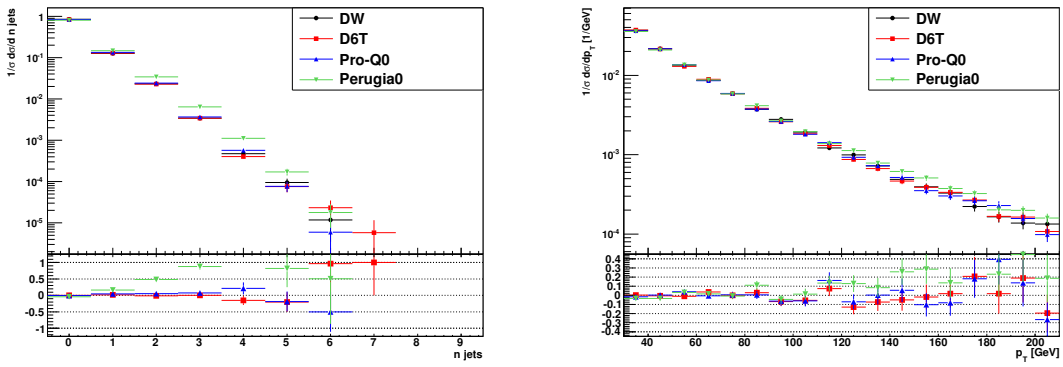


Fig. 12: Z transverse momentum spectrum for different Pythia 6 tunes. Relative difference is shown with respect to tune DW.

discrepancies are observed in the low p_T region: both tune Perugia0 and tune Pro-Q0 differ from Tune DW, and they also differ from each other. The disagreement with respect to tune DW is up to 20%.

Jet multiplicity and leading jet transverse momentum spectrum are shown in Fig. 13. The jet multiplicity

for Tune Perugia0 shows some not negligible differences with respect to the other tunes, predicting more jets. On the other hand Tune DW, D6T and Pro-Q20 predict very similar rates. Concerning the leading jet p_T spectrum in Fig. 13 (b), the shape predicted by the four tunes is very similar.



(a) (b)

Fig. 13: (a) Jet multiplicity, (b) leading jet p_T spectrum for four Pythia 6 tunes. Relative difference is shown with respect to Tune DW.

7.5 Pythia 6, Sherpa, AlpGen COMPARED

In this section we compare Pythia 6 (including full matrix element corrections) with AlpGen and Sherpa. We used Pythia 6 version 6.421, both standalone and to shower AlpGen events. Also, we used tune DW for Pythia 6.

We used matrix element corrected Pythia 6 as a reference for a consistency check of the matching prescriptions used in AlpGen and Sherpa. Both the CKKW [4] prescription used in Sherpa and MLM prescription used AlpGen [64, 70] can describe multiple parton emission corrected for the corresponding multi-parton matrix element. If just one additional parton emission from the matrix element is permitted, those prescriptions should give results compatible with Pythia 6.

While CKKW and MLM divide the phase space in a region of jet production, populated by the matrix element, and a region of jet evolution populated by the shower, the matrix element correction prescription implemented in Pythia 6 does not depend on any separation parameter, thus providing us with the "correct" reference to test the other matched calculations.

The comparison among the three generators is shown in Fig. 14 for the Z transverse momentum. The agreement in shape is generally good, well within 20% with respect to Pythia 6 for most of the spectrum. In particular, the mismatch between AlpGen and Pythia 6 observed in [66] is practically gone with the newer version of Pythia 6. This is an important result also because it establishes the re-usability of Pythia 6 tunes between standalone Pythia 6 and AlpGen plus Pythia 6 shower.

Jet multiplicity and leading jet p_T spectra are shown in Fig. 15. In (a) both AlpGen and Sherpa appear to have a longer tail for high jet multiplicity. Concerning the leading jet p_T spectrum in (b), Sherpa follows Pythia 6 all over the spectrum, AlpGen appears so have a harder tail for high p_T.

Differential jet rates for the 1→0, 2→1 and 3→2 transitions are shown in Fig. 16. Differential jet rates for the transition $n \to n-1$ are the distributions of the k_T resolution parameter in an exclusive k_T algorithm, $Q_{n\to n-1}$, that makes an n jet event turn into an $n-1$ jet event. AlpGen and Pythia 6 are similar, while Sherpa's shape shows some not negligible differences. Sherpa uses a dipole shower, while Pythia 6 uses a virtuality ordered shower (the same shower is also used with AlpGen). The source of the difference is likely in the different shower, which fills the phase space with a different pattern.

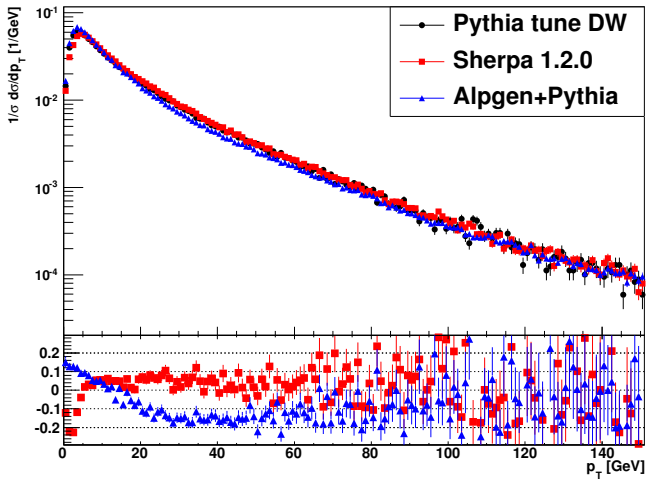
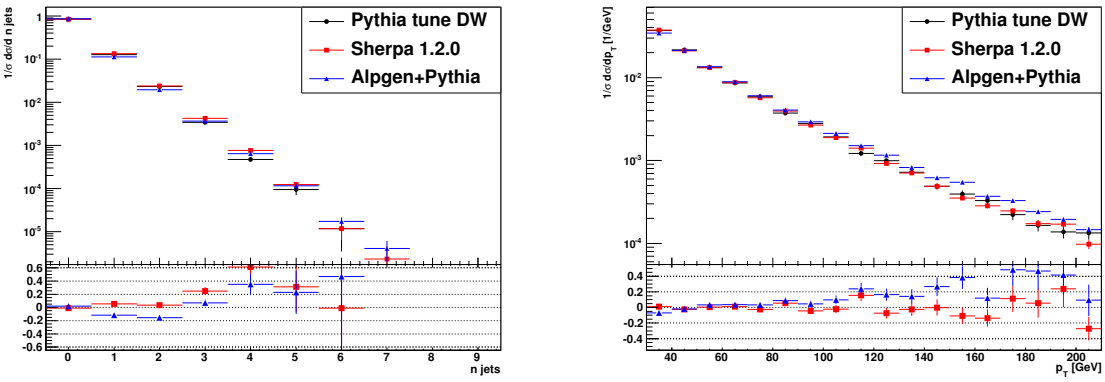
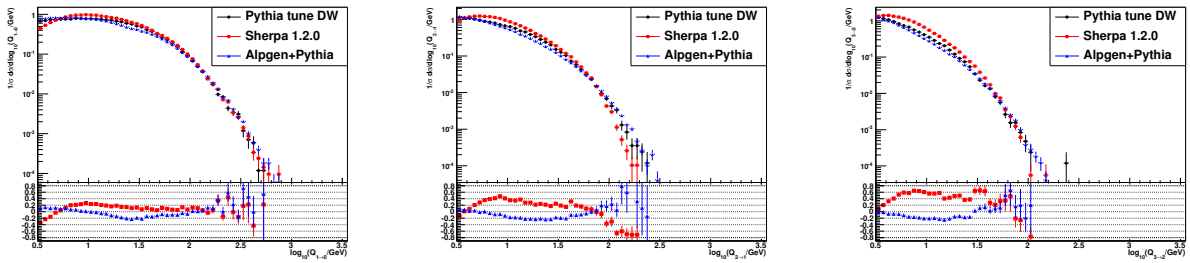


Fig. 14: Z transverse momentum spectrum for Pythia 6, Sherpa and AlpGen. The agreement is within 20%.



(a) (b)

Fig. 15: (a) Jet multiplicity, (b) leading jet p_T spectrum.



(a) (b) (c)

Fig. 16: Differential jet rates: (a) 1→0, (b) 2→1, (c) 3→2.

7.6 CONCLUSION

We studied matrix element corrections in Pythia 6 for inclusive Z production at the LHC and we compared Pythia 6, AlpGen and Sherpa on the ground of first order real emission corrections. A bug was recently

discovered and corrected in inclusive vector boson production in Pythia 6, that made the effect of matrix element corrections very big also at low boson p_T. We compared two versions of Pythia 6, before and after the fix, and observed that the low boson momentum region is correctly described in the new version, the matrix element corrected and uncorrected shower being similar in that region as expected.

We compared four different Pythia 6 tunes. For all of them we switched off the multiple interactions in order to observe the effect of the tune on the shower and on the hard event simulation. Generally good agreement was observed, with some differences especially in the low Z p_T region and in the jet multiplicity spectrum.

We showed various plots comparing Pythia 6, AlpGen and Sherpa. Differences between AlpGen plus Pythia 6 and matrix element corrected Pythia 6 observed in [66] were due to the bug affecting Pythia 6, and are now much reduced. This also assesses the re-usability of Pythia 6 tunes when Pythia 6 is used to shower AlpGen events.

7.7 PARAMETERS AND SETTINGS

In Pythia 6, the underlying event was switched off with `MSTP(81)=0` (`MSTP(81)=20` for the tune Perugia0, that uses the p_T ordered shower). Matrix element corrections were controlled with the parameter `MSTP(68)`: this was set to `=0` to switch off matrix element corrections and to set the staring scale to the Z mass, to `=2` for the power shower without matrix element corrections. The default had matrix element corrections enabled. Hadronization was switched off with `MSTJ(1)=0`. The various tunes were selected by acting on parameter `MSTP(5)`.

AlpGen samples were produced with a minimum parton p_T of 20 GeV. We used the default parameters for Sherpa, with no underlying event, no fragmentation, no QED radiation off leptons.

Part III

BEYOND FIXED ORDER

8. MULTIPLE PARTON INTERACTIONS AS A BACKGROUND TO TOP PAIR PRODUCTION [16]

8.1 INTRODUCTION

The occurrence of Multiple Parton Interactions (MPI) in hadronic collisions is by now well established experimentally [71–74]. Furthermore, they represent the basic mechanism in the description of the Underlying Event (UE) in the standard Showering Monte Carlo's (MC) like Pythia 6 [69, 75, 76] and Herwig [54, 77]. MPI rates at the LHC are expected to be large, making it necessary to estimate their contribution to the background of interesting physics reactions. On the other hand, their abundance makes it possible to study MPI experimentally in details, testing and validating their implementation in the MC.

In this note we discuss the role of MPI in top-antitop production at $\sqrt{s}$ = 10 TeV. The LHC will be a top-antitop factory and the large rate will allow accurate measurements of the top mass, one the most crucial parameters for stringent tests of the SM, and production cross section. It will therefore be of extreme importance to have full control over all potential background processes. In this study we compare the results obtained with Pythia 8 [20] with those obtained at parton level with the methods of Refs. [78, 79] which suggested that MPI could provide a significant background to top-antitop production and to other interesting processes like $W + 4j$ and $Z + 4j$ production.

8.2 SETUP

If no b–tagging is assumed, the MPI processes which provide a background to $t\,\bar{t}$ and more generally contribute to $W + 4j$ through Double Parton Interactions (DPI) are:

$$jj \otimes jjW\,, \;\; jjj \otimes jW\,, \;\; jjjj \otimes W \tag{8}$$

where the symbol $\otimes$ stands for the combination of one event for each of the two final states it connects.

Besides in W decay, isolated high–p$_T$ leptons can also be produced in the decay of heavy quarks. Therefore any reaction in which heavy quarks plus jets are produced can fake the signature of a W boson. Keeping in mind the large QCD cross-section for $b\bar{b}$ production, heavy quark and QCD DPI processes deserve attention as possible backgrounds to top pair production. This corresponds to the following extra processes:

$$b\bar{b} \otimes jj\,, \;\; b\bar{b}j \otimes jj \tag{9}$$

In this study we have decided to use Pythia 8 for simulating DPI events. Pythia 8 allows to use external Les Houches Accord [3] input for the hardest process, and to choose an internal one for the second hardest, being however limited to $2 \to 2$ processes. This set-up allows a precise study of dedicated DPI configurations.

Moreover we have generated a number of samples, most notably $W + 2j$, with `MADEVENT` [80, 81], for which DPI has been simulated by either Pythia 8 or by combining pairs of parton–level events as in Refs. [78, 79].

In summary, these are the samples that we have generated with Pythia 8:

$$t\bar{t}\,, \;\; Wj \otimes jj\,, \;\; Wjj(MG) \otimes jj\,, \;\; b\bar{b} \otimes jj\,, \;\; b\bar{b}j(MG) \otimes jj \tag{10}$$

[16] Contributed by R. Chierici, E. Maina

where with $Wjj(MG)$ and $b\bar{b}j(MG)$ we indicate that the corresponding events have been produced starting from an external Les Houches file created by MADEVENT . In all cases the second hard interaction switched on in Pythia 8 is a generic $2 \to 2$ QCD process. The default Pythia 8.130 parameter setup and tunings have been used. All the events processed with Pythia 8 have been fully showered, and jet clustering run over the events to have the possibility to apply a more realistic event selection.

We have also generated at parton level all reactions contributing to $jj \otimes jjW$, $jjj \otimes jW$ and $jjjj \otimes W$ with MADEVENT.

All samples have been generated using CTEQ6L [82] parton distribution functions and with the following parton level cuts:

$$p_{T_j} \geq 10 \text{ GeV}\,, \quad |\eta_j| \leq 5.0\,, \quad \Delta R_{jj} \geq 0.001 \tag{11}$$

where $j = u, \bar{u}, d, \bar{d}, s, \bar{s}, c, \bar{c}, b, \bar{b}, g$.

The DPI cross section for the parton level samples has been estimated as

$$\sigma = \sigma_1 \cdot \sigma_2 / \sigma_{eff} / k \tag{12}$$

where σ_1, σ_2 are the cross sections of the two contributing reactions and k is a symmetry factor, which is equal to two if the two reactions are indistinguishable and equal to one when they are different.

At the Tevatron, CDF [73] has measured $\sigma_{eff} = 14.5 \pm 1.7^{+1.7}_{-2.3}$ mb, a value confirmed by DØ which quotes $\sigma_{eff} = 15.1 \pm 1.9$ mb [74]. In Ref. [83] it is argued, on the basis of the simplest two channel eikonal model for the proton–proton cross section, that a more appropriate value at $\sqrt{s} = 1.8$ TeV is 10 mb which translates at the LHC into $\sigma_{eff}^{LHC} = 12$ mb. Treleani then estimates the effect of the removal by CDF of Triple Parton Interaction events from their sample and concludes that CDF measurement yields $\sigma_{eff} \approx 11$ mb. In the following we use $\sigma_{eff} = 12.0$ mb with the understanding that this value is affected by an experimental uncertainty of about 15% and that it agrees within about 10% with the predictions of the eikonal model.

The DPI picture in Pythia 8 assumes that interactions can occur at different p_T values independently of each other inside inelastic nondiffractive events. The expression for a DPI cross section becomes therefore:

$$\sigma = < f_{impact} > \sigma_1 \cdot \sigma_2 / \sigma_{ND} / k \tag{13}$$

where σ_{ND} is the total non-diffractive cross section and f_{impact} is an enhancement/depletion factor chosen event-by-event to account for correlations introduced by the centrality of the collision. This quantity is typically averaged during an entire run to calculate $< f_{impact} >$ in Eq. 13. Typical values at the centre of mass energy of 10 TeV are 1.33 for $< f_{impact} >$ and 51.6 mb for σ_{ND}. Comparing Eq. 13 with Eq. 12 tells us that Pythia 8 sort of predicts an effective σ_{eff}=σ_{ND}/$< f_{impact} >$ which is about a factor 3 larger than the one actually measured at the Tevatron.

Table 4 reports the cross section breakdown for the samples used in this study and processed by Pythia 8. All cross-sections are leading order as returned by the programs used. The Wjj sample generated with MADEVENT has also been combined with matrix elements QCD events, evaluating the cross section according to Eq. 12: we will call this sample as $Wjj \otimes jj$ ME, as opposed to the $Wjj \otimes jj$ P8 sample.

8.3 RESULTS

To mimic an experimental analysis, showered events are first required to have an isolated, central and energetic lepton according to the cuts:

$$p_{T_l} \geq 20 \text{ GeV}\,, \quad |\eta_l| \leq 2.5\,, \quad E_{iso} \leq 20 \text{ GeV} \tag{14}$$

with obvious notation and where E_{iso} is an isolation variable defined as the total energy flowing in a cone of opening angle ΔR=0.1 around the lepton. Later in this study we have made this cut more severe

Process	σ(nb)	$\sigma(\otimes jj)$(pb)	σ_{acc}(pb)
$t\bar{t}$	0.24	no DPI	60
$W(\to l\nu)jj$(MG)	13.7	140	23
Wj	4.8	12	2.5
$b\bar{b}$	850	2200	18
$b\bar{b}j$(MG)	688	7300	69
$b\bar{b}jj$(MG)	472	no DPI	4012
$W(\to l\nu)jj$(MG)	13.7	no DPI	627

Table 4: Total cross-sections for the processes studied in this contributions. Samples labelled with MG are made with the use of MADEVENT. The second column presents the cross sections as returned by the generators, the third column the corresponding cross sections after having added a second hard interaction to the process in Pythia 8, and the last column the accepted cross sections after the selection described in Eq. 14.

by increasing ΔR to 0.5. We will also compare the effect of different jet p_T thresholds. To be closer to an experimental selection, only electrons and muons are considered in this study. In events with a candidate lepton, at least four jets with transverse momentum above 20 GeV and $|\eta| < 5$ are requested. No b-tagging requirement is mimicked. The accepted cross sections of the top-pair signal and the DPI backgrounds are reported in Table 4. In comparison we have also reported the accepted cross sections for two equivalent non-DPI background processes.

It has to be mentioned that for the DPI events obtained by the simple combination of parton level events as in Refs. [78, 79], where no parton shower and fragmentation is present, the above cuts have been replaced by equivalent ones. The lepton isolation, for instance, becomes a cut in ΔR between the lepton and any jet.

In Figure 17 we present a few distributions at generator level after the event selection in Eq. 14. We decided to compare the top-pair signal with the Wjj process alone and with DPI $Wjj \otimes jj$ in both the ME and P8 configurations. All histograms in the plots are normalised to the number of events expected in 100/pb of integrated luminosity, with the exception of the Wjj contribution that has been scaled by a factor 10 to make the plot more readable. The plots shown represent, respectively, the highest and softest (among four) p_T of the jets in the event, the maximum pseudorapity difference among the four jets, the azimuthal difference between the jets complementary to those with the maximum azimuthal difference, the scalar sum of the transverse momenta of the four highest p_T jets, and the transverse momentum of the found lepton. The heavy quark DPI component is shown only for this last histogram. The plot is a good indication of the fact that kinematic cuts on the lepton transverse momentum, as well as a more severe cut on the lepton isolation, should be very effective in removing the heavy quark background(s), as we will also demonstrate later.

In the plots we have not superimposed several ISR components that are, as expected, largely dominating in all the corners of the phase space (for instance $Wjjjj$ with respect to $Wjj \otimes jj$ DPI or $b\bar{b}jj$ with respect to $b\bar{b} \otimes jj$ DPI). Therefore, analyses able to reach high purities in the top signal over the ISR background, should also be able to highly suppress all the analogous signatures coming from DPI events.

The DPI contributions can however reach sizeable fractions of the ISR backgrounds for analyses with non sufficiently tight cuts. In the figures, for instance, we compare Wjj(+parton shower) rescaled by a factor 10 with the $Wjj \otimes jj$ DPI component. The rate of the latter is sizeable and the two have similar kinematic characteristics: the additional jets from the second interaction have the tendency to be softer and slightly more forward in the acceptance. This suggests that harder cuts on the jet transverse momenta should also be very effective in reducing the DPI component. From the figures it also turns out that there are significant discrepancies in the normalisation, as well as in shape, for the prediction of the $Wjj \otimes jj$

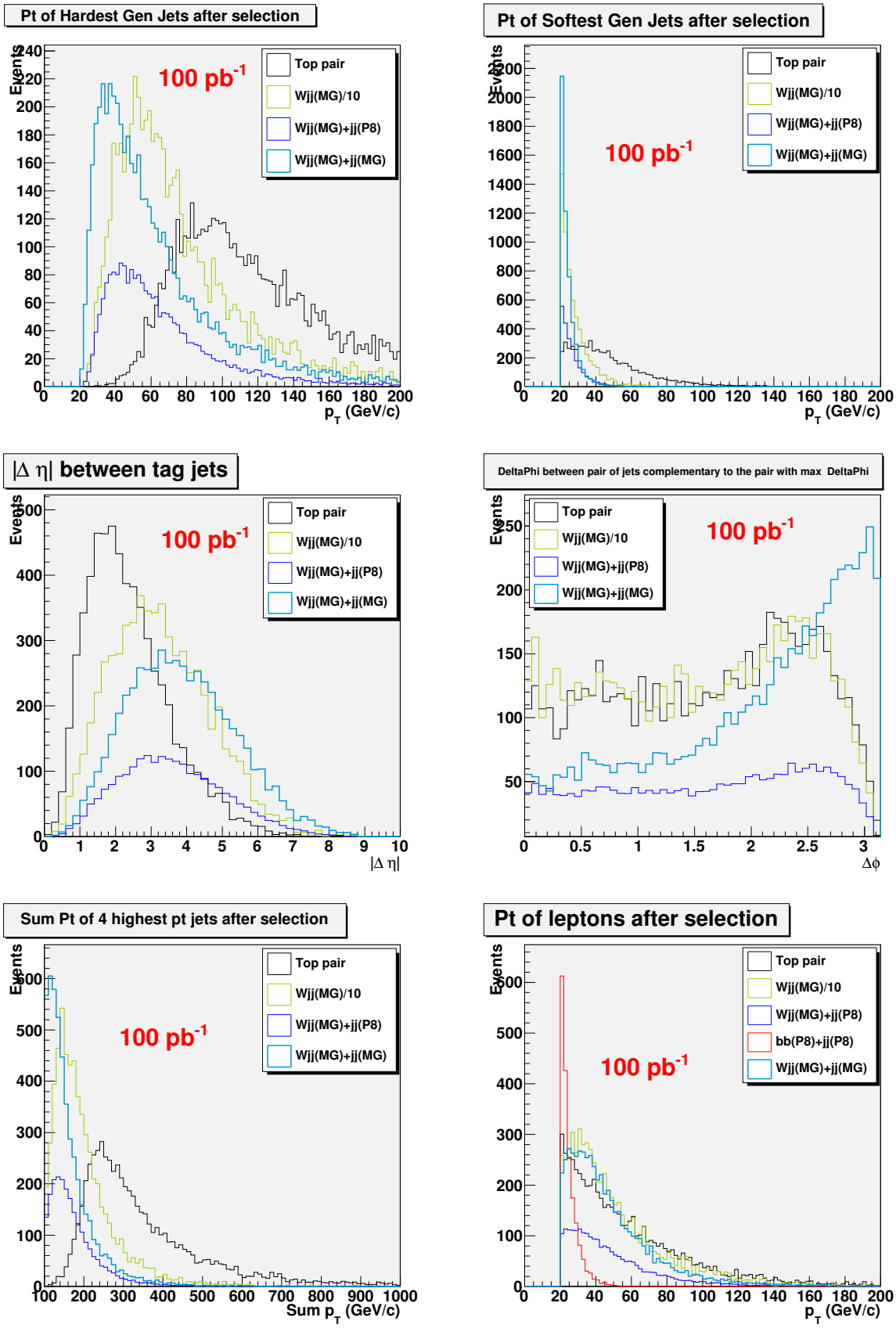


Fig. 17: Distribution of the highest p_{T_j}; the fourth highest p_{T_j}; the largest separation in $|\Delta\eta|$ among the four highest p_T jets; the $|\Delta\phi|$ of the two jets complementary to those with the largest separation in $|\Delta\phi|$ among the four highest p_T jets; the scalar sum of the p_T's of the four highest p_T jets; the p_T of the charged lepton. The plots are made with the cuts in Eq. 14 and for 100 pb^{-1} of integrated luminosity at the LHC at $\sqrt{s}$ = 10 TeV.

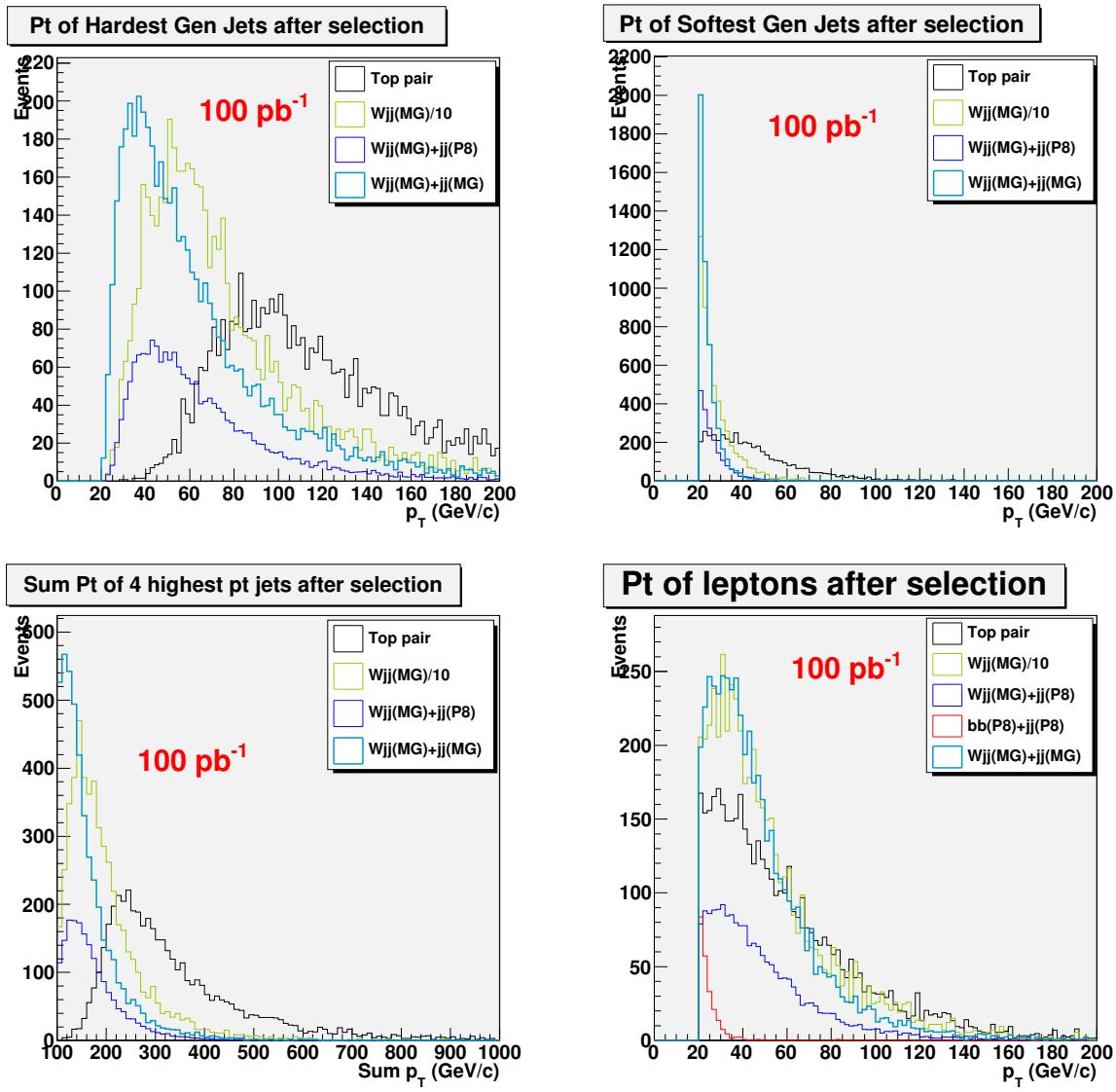


Fig. 18: Distribution of the highest p_{T_j}; the fourth highest p_{T_j}; the scalar sum of the p_T's of the four highest p_T jets; the p_T of the charged lepton. The plots are made with the cuts in Eq. 14 and a more severe lepton isolation condition, and for 100 pb^{-1} of integrated luminosity at the LHC at $\sqrt{s}$ = 10 TeV. The lepton isolation cone has now an opening of ΔR=0.5.

DPI ME and P8 components. The difference in normalisation is a direct consequence of the different normalisation cross section for the DPI component in Pythia 8 with respect to the model in [83] and the Tevatron measurements. The relatively small shape difference can everywhere be ascribed to the fact that the distributions for $Wjj \otimes jj$ ME are made at parton level prior to any showering, whereas for Pythia 8 they are all at hadron level and, therefore, also account for shower radiation.

The conclusions drawn so far have a very mild dependence on the lepton selection, with the exception of the reduction of the heavy quark component, for which a tighter isolation requirement is mandatory. The plots in Figure 18 correspond to those in Figure 17, with the exception that the isolation cone for the lepton has now an opening of ΔR=0.5. Shape and normalisations for the signal and the W+jets backgrounds do not change significantly, whereas the contribution of the heavy flavour component is significantly reduced.

Another way to tighten the analysis, and protect the signal even further from the DPI component is to increase the jet transverse momentum threshold. Figure 19 shows the scalar sum of the four highest p_T

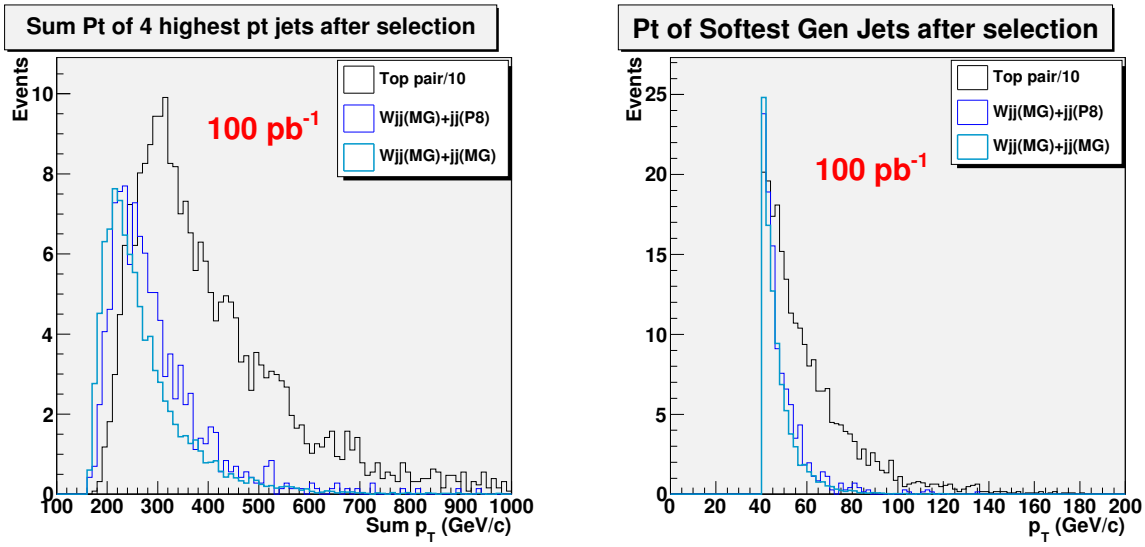


Fig. 19: Distribution of the scalar sum of the p_T's of the four highest p_T jets and the fourth highest p_{T_j}. The plots are made for 100 pb^{-1} of integrated luminosity at the LHC at $\sqrt{s}$ = 10 TeV. The lepton isolation cone has now an opening of ΔR=0.5 and $p_{T_j} \geq 10$ GeV.

jets transverse momenta and the p_T of the fourth jet in the event after that both lepton isolation conditions and jet thresholds have been tightened. The latter has been moved from 20 GeV to 40 GeV. The top pair signal component is now scaled by a factor 10, making it evident that all DPI component are significantly reduced. What is also interesting is that in this region of phase space the accepted cross section (total and differential) for $Wjj \otimes jj$ ME and P8 are essentially equivalent. This indicates that either approach is equivalent in describing DPI at relatively high transverse momentum.

8.4 FURTHER RESULTS

In Sect. 8.3 we have compared the $Wjj \otimes jj$ parton level DPI contribution to a number of contributions generated with Pythia 8. We have neglected the $jjj \otimes jW$ and $jjjj \otimes W$ components because they cannot be directly generated only with Pythia. In this section we complete our analysis in this regard. In Figure 20 we present the distribution of the largest jet transverse momentum for $jj \otimes jjW$, $jjj \otimes jW$ and $jjjj \otimes W$. We also show the Single Parton Interaction (SPI) result for $W + 4j$ and the sum of SPI and MPI contributions. We neglect the small Triple Parton Interaction component. All results in this section are purely at parton level without showering and hadronization. The selection cuts are:

$$p_{T_j} \geq 20 \text{ GeV}\,, \quad |\eta_j| \leq 5.0\,, \quad \Delta R_{jj} \geq 0.5\,, \quad p_{T_l} \geq 20 \text{ GeV}\,, \quad |\eta_l| \leq 2.5\,, \quad \Delta R_{jl} \geq 0.1 \tag{15}$$

The $W+4j/Wjj\otimes jj/jjj\otimes jW/jjjj\otimes W$ contributions are in the ratio 1/0.33/0.11/0.04 and therefore the DPI rate is about half of the SPI rate, with the $Wjj \otimes jj$ component contributing about two thirds of the DPI total. As a comparison we recall that in [78] the DPI contribution to $W + 4j$ was estimated at about 10% at $\sqrt{s}$ = 14 TeV. Only part of the discrepancy between the two results can be explained by the smaller effective cross section $\sigma_{eff} = 12.0$ mb employed here in comparison to $\sigma_{eff} = 14.0$ mb used in [78]. Therefore our preliminary analysis suggests that the DPI background to $W + 4j$ production is proportionally larger at $\sqrt{s}$ = 10 TeV than at 14 TeV.

CONCLUSIONS

We have asked ourselves the question whether MPI can be a source of unexpected background for high p_T physics at the LHC. In this study we have taken top production as a perfect use case, trying to determine

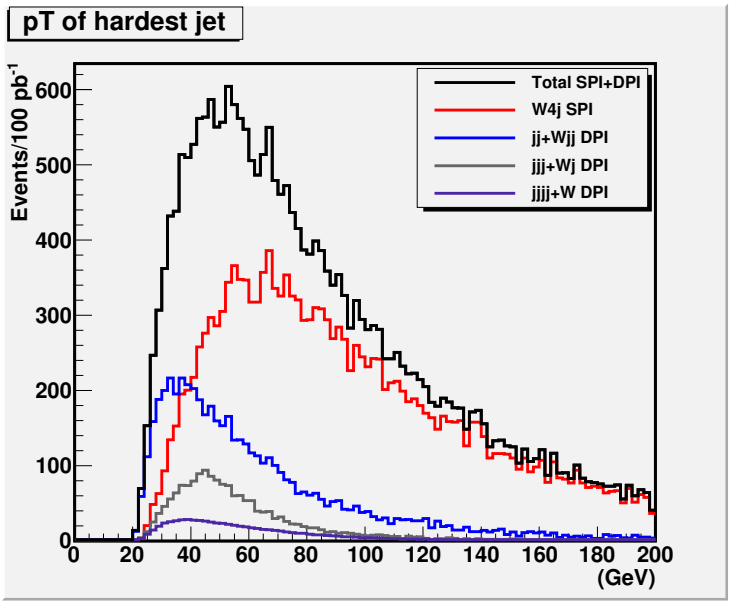


Fig. 20: Distribution of the highest p_{T_j} for $jj \otimes jjW$, $jjj \otimes jW$ and $jjjj \otimes W$. Also reported are the SPI result for $W + 4j$ and the sum of SPI and MPI contributions.

which MPI background can induce a signature compatible with $t\bar{t}$. We have shown that DPI coming from QCD processes and either W production or heavy quark production can indeed fake a top pair signature, becoming a significant fraction of the remaining background if the analyses are not sufficiently tight. As main experimental cuts for reducing drastically DPI we suggest the minimum transverse momentum threshold for jets and the lepton isolation, which are expected to be also effective for standard multi-jet backgrounds. In our study we have compared two ways of generating DPI events: one based on Pythia 8 and the other by direct combination of parton level events. The differences between the two approaches are important, especially in the softer regionss of phase space and for what concerns the global normalisation, and are ultimately due to the assumptions on the cross section to which DPI are normalised. The present results suggest that the DPI background to $W + 4j$ production is proportionally larger at 10 TeV than at 14 TeV

ACKNOWLEDGEMENTS

We are indebted to P. Skands and T. Sjöstrand for very interesting discussions on MPI and Pythia 8.

9. A MATCHING SCHEME FOR $W\gamma$ NLO MATRIX ELEMENT GENERATOR AND Pythia 8 PARTON SHOWER [17]

9.1 INTRODUCTION

Diboson production is among the early physics topics to be studied at the LHC initial phase. In particular, the $W\gamma$ process has a higher rate compared to the others, which can be pursued at moderate luminosities to probe the gauge structure of the Standard Model. The presence of anomalous couplings between the W-bosons and the photon is expected to enhance the harder part of the photon transverse momentum (p_T^γ) spectrum which is one of the experimental observables. Higher order QCD corrections will contribute significantly at the LHC, modifying the p_T^γ spectrum and hence may obscure the effect of any anomalous coupling that might exist. This calls for at least a NLO QCD event generator to accurately describe the parton level process. However, for a full event description including the initial and final state radiations (ISR and FSR) and the underlying events the parton shower approach is required. In this contribution we discuss a matching scheme for $W\gamma$ events generated by WGAMMA_NLO [84], which is a complete $\mathcal{O}(\alpha_s)$ calculation, to the Pythia [18], [20] parton showers. The proposed methodology for the matching preserves both the rate of the hard scattering process as well as various kinematic distributions of experimental interest, like the photon transverse momentum.

9.2 THE MATCHING STRATEGY

The WGAMMA_NLO package [84], [85] is a matrix element-based event generator which produces $W\gamma$ events in hadronic collisions with up to one additional parton (quark or gluon) in the final state. It has the provision for the introduction of anomalous $WW\gamma$ couplings in addition to Standard Model ones and hence interesting for experimental analysis.

The WGAMMA_NLO package produces two types of event topologies:

- 3-body final state, from $W\gamma$ +0 jet events, containing the lepton and the neutrino from W-decay and the photon;
- 4-body final state, from $W\gamma$ +1jet events, containing an additional parton (quark or gluon) forming a jet.

Figs. 21 and 22 depict the Born level diagrams and some of the higher order QCD corrections for the $W\gamma$ process respectively. Fig. 23 compares the p_T^γ spectrum at the Born level to that at NLO, as obtained from WGAMMA_NLO . The cross-section for the $W\gamma$ process is 1.390 ± 0.029 pb at NLO and 0.584 ± 0.005 pb at Born level for $p_T^\gamma > 25$ GeV$/c$ with $|\eta^\gamma| < 2.1$ and $p_T^{lepton} > 15$ GeV$/c$ with $|\eta^{lepton}| < 2.6$, η^γ and η^{lepton} being the pseudorapidity of the photon and the lepton from W-decay respectively. The selection applied here are typical of the LHC experiments. This gives a p_T^γ -dependent k-factor with a mean value of of 2.16, which is quite substantial (Fig. 23 inset).

Pythia 8 has been used for implementing the p_T-ordered parton shower to the hard-scattered part generated by WGAMMA_NLO . The ISR process in Pythia adds coloured partons to the incoming quarks of the hard scattering process which may transform a 3-body event to a 4-body event, thereby effectively changing the cross-section of the individual 3-body and 4-body processes. The idea behind the matching scheme is to generate the parton showers without changing the number of individual events; *i.e.,* the parton shower should not populate the regions of phase space for the $W\gamma j$ events, j being a parton, which are already filled by the matrix element generator.

It is to be noted that WGAMMA_NLO package produces weighted events which has to be unweighted before matching where only events with unit weight are considered. The unweighting is done using the standard "hit-and-miss" technique, where the ratio of an event weight to the highest weight is compared against a number drawn from a uniform deviate and the event is kept if the ratio is greater than the random number generated from the uniform distribution. Events with an occasional negative weight, from virtual loops, are not considered.

[17] Contributed by: D. Majumder, K. Mazumdar, T. Sjöstrand

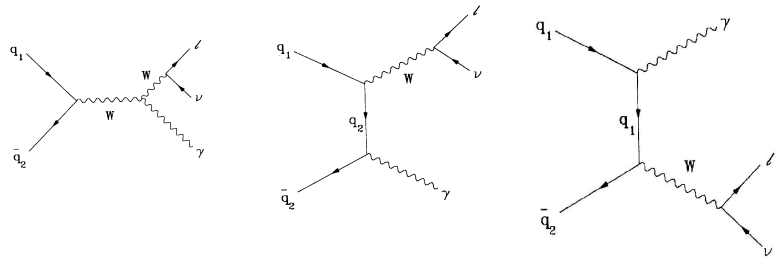


Fig. 21: Born level subprocesses for $W\gamma$ production in hadron-hadron collision : s-channel process (**left**), t-channel process (**middle**) and u-channel process (**right**).

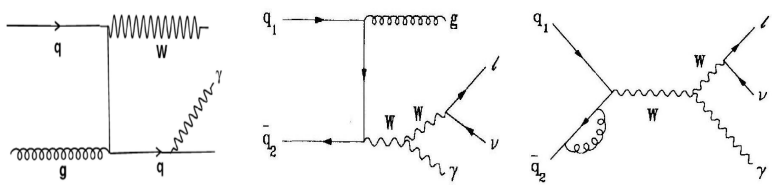


Fig. 22: Some higher order QCD diagrams for $W\gamma$ production in hadron-hadron collision.

9.21 The matching algorithm

In the matching procedure, the discrimination of 3-body vs. 4-body state is performed according to a predefined threshold value $p_T^{separate}$ (5 GeV/c for the present study), which serves as the boundary between the matrix element regime on its upper side and the parton shower regime on its lower. The matching procedure starts with projecting the 4-body system to a 3-body system, by assuming that the outgoing parton can be emitted from either of the incoming partons (flavours permitting), with relative weights determined by splitting kernels and parton densities. Effectively this corresponds to the assumption that the 4-body state never had a parton emitted and all the kinematics are recalculated based on this assumption. Subsequently, this projected 3-body final state event is treated according to the following steps:

1. Shower the projected 3-body event and compare p_T^{shower} at the first ISR branching with the p_T^{parton} in original 4-body:
   - If $p_T^{shower} > p_T^{parton}$ then the event is reclassified to 3-body; move to step 2 below.
   - If $p_T^{shower} < p_T^{parton}$ then the original 4-body event is retained as a 4-body event; move to step 3 below.

2. Shower the 3-body events:
   - Compare p_T^{shower} with $p_T^{separate}$ after first ISR branching. $p_T^{separate}$ can be considered to be the boundary between the ME calculation's regime and that of the parton shower's.
   - If $p_T^{shower} > p_T^{separate}$ then stop any further shower evolution and restart the parton shower until $p_T^{shower} < p_T^{separate}$.
   - Continue with the rest of the shower.

3. Shower the 4-body events:
   - The p_T^{shower} from the first ISR is compared with p_T^{parton}. If $p_T^{shower} > p_T^{parton}$ then restart the

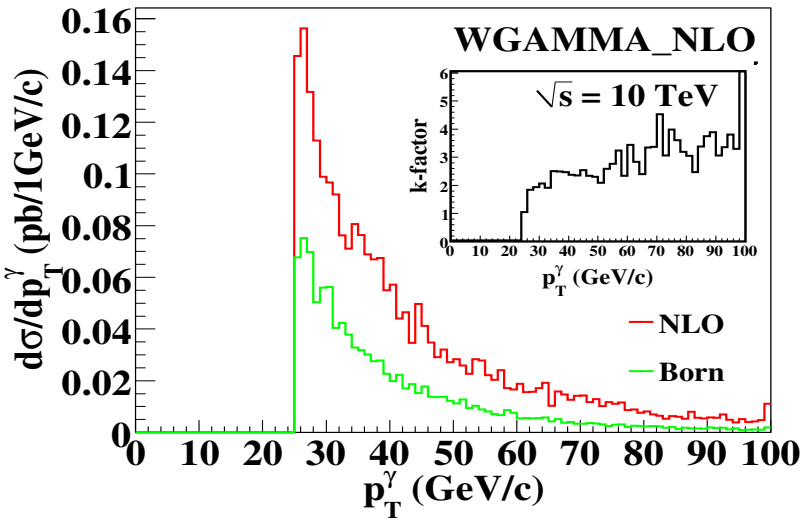


Fig. 23: Differential cross-section for the photon transverse momentum at Born level is compared with the differential cross-section for NLO $W\gamma$ process. Inset: k-factor as defined by the ratio of the NLO cross-section to the Born cross-section.

ISR branching until $p_T^{shower} < p_T^{parton}$

- Continue with rest of shower to give a complete event.

9.22 Matching results

The 4-body final states from the matrix element calculation lacked the Sudakov form factor for QCD emission which was corrected for by the above-mentioned algorithm. It is to be noted that the first ISR emission from the 3-body events are always confined below $p_T^{separate}$. But this is not so in the case where the parton showering is only required to be softer than the matrix element parton.

Fig. 24 (**left**) shows the distribution of the transverse momentum of the $W\gamma$ system obtained by WGAMMA_NLO superposed with the same after the Pythia showering. The distribution from WGAMMA_NLO shows many events with $p_T^{W\gamma}$ equal to zero which are the 3-body events. The non-zero values correspond to events with a parton in the final state. After the matching, the kink after the zeroth bin fills up due to the boost from Pythia ISR. The area under both these curves however remain the same indicating that the exclusive cross-section of the 1-jet events remain conserved after the parton shower. Finally, Fig. 24 (**right**) shows the photon p_T spectrum before and after the matching.

The matching scheme is rather simplistic and is suitable for this particular event topology which contains only one hard jet in the final state. The advantage is that it does not require any modification of the matrix element calculation of Baur for the Sudakov form factor needed for the parton emission. This is in contrast with packages based on leading order multi-parton final state calculation, where the matching scheme with parton shower effectively reduces the cross-section.

9.3 SUMMARY

The matching strategy as enunciated above preserves the cross-section of the $W\gamma$ + 1 jet events. Also, for a reasonably high cut on p_T^{γ} , the spectrum before and after the matching is identical. It remains to be seen how stable the scheme is with respect to changes in the boundary between the matrix element and parton shower, as defined by $p_T^{separate}$ and also its performance in comparison to matched events generated by other matrix element generators like Madgraph and Alpgen. The final test of the best suited method can be performed only with good-statistics real-collision data at LHC.

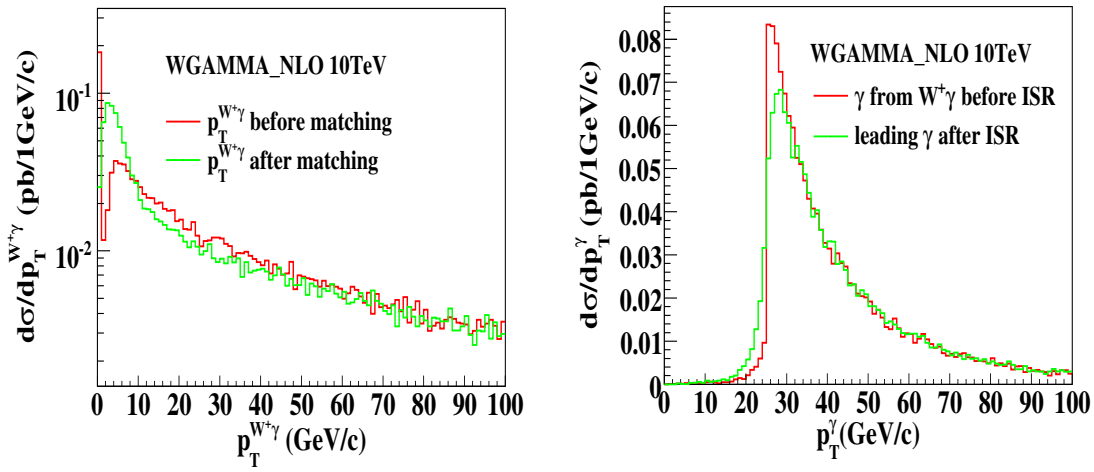


Fig. 24: The transverse momentum of the $W\gamma$ system before Pythia showering and after Pythia showering (**left**) and photon p_T spectrum before and after the showering (**right**).

ACKNOWLEDGEMENT

The main part of this work done by DM was supported by the European Union Marie Curie Research Training Network MCNet, under contract MRTN-CT-2006-035606. DM and KM are extremely grateful to U. Baur for many private communications at the initial stage of the study.

10. THEORY TESTS OF PARTON SHOWERS [18]

10.1 INTRODUCTION

The last few years have witnessed significant progress in the improvement of parton shower algorithms with the invention of new shower algorithms based on the dipole or antenna picture [86–92] known from next-to-leading order calculations. These algorithms are able to satisfy simultaneously at each step momentum conservation and the on-shell conditions. This is possible, because they are based on $2 \to 3$ splittings, where the spectator can absorb the recoil. Within the traditional $1 \to 2$ splitting algorithms it is impossible to satisfy simultaneously momentum conservation and the on-shell conditions for a splitting. In addition these new algorithms implement in a correct way simultaneously the soft and the collinear limit of the matrix elements. When both emitter and spectator are in the final state these new algorithms are very similar to the shower algorithm implemented in ARIADNE [93–97]. The new algorithms extend the dipole or antenna picture to final-initial, initial-final and initial-initial configurations.

For any shower algorithms one needs an evolution variable. A sensible choice is

$$t = \ln \frac{-k_\perp^2}{Q^2}, \tag{16}$$

where Q^2 is a fixed reference scale and $k_\perp$ is the transverse momentum of a splitting. During the shower evolution we move towards smaller (more negative) values of t. The central object of a shower algorithm is the Sudakov factor, describing the no-emission probability. For a dipole shower it is calculated from the individual dipoles. For a dipole with emitter $\tilde{i}$ and spectator $\tilde{k}$, the Sudakov factor is given by

$$\Delta_{ij,k}(t_1, t_2) = \exp\left(- \int_{t_2}^{t_1} dt \mathcal{C}_{\tilde{i},\tilde{k}} \int d\phi_{unres} \delta\left(t - T_{\tilde{i},\tilde{k}}\right) \mathcal{P}_{ij,k} \right), \tag{17}$$

where $\mathcal{C}_{\tilde{i},\tilde{k}}$ is a colour factor. $T_{\tilde{i},\tilde{k}}$ depends on the dipole invariant mass $(p_{\tilde{i}} + p_{\tilde{k}})^2$ and the phase space variables for the emission of an additional particle. The essential information is given by the function $\mathcal{P}_{ij,k}$, which encodes the singular part of the matrix elements for the emission of a particle. As an example we quote this function for the $q \to qg$ splitting:

$$\mathcal{P}_{q\to qg} = C_F \frac{8\pi\alpha_s(\mu^2)}{(p_{\tilde{i}} + p_{\tilde{k}})^2} \frac{1}{y} \left[\frac{2}{1 - z(1-y)} - (1+z) \right]. \tag{18}$$

α_s is evaluated at the scale $\mu^2 = -k_\perp^2$. The shower algorithm proceeds through the following steps: One first chooses the next scale t at which a splitting occurs according to the Sudakov factor. For this given choice of t one then chooses the momentum fraction z according to $P_{a\to bc}(z)$ and an azimuthal angle φ either uniform or according to spin-dependent splitting functions. With these three variables (t, z, φ) one can construct all four-vectors after the emission. The spectator absorbs thereby some recoil. We now have a configuration where one particles was emitted. For $t > t_{min}$ one goes back to the first step, otherwise one stops.

10.2 EVOLUTION VARIABLES

In the last year there has been some discussion whether the new shower algorithms based on dipoles or antennas [98–100] have the correct collinear limit. In particular it has been argued that the fact that the spectator takes some recoil could conflict with collinear factorisation. This question has been solved and it has been shown that the algorithms based on dipoles or antennas describe correctly the collinear limit [99, 100]. This result could have been anticipated from the fact, that the dipole splitting functions

[18] Contributed by: S. Weinzierl

$P_{a\to bc}$ reduce in the collinear limit $y \to 0$ to the Altarelli-Parisi splitting functions. For the example of a $q \to qg$-splitting we have

$$\lim_{y\to 0} \mathcal{P}_{q\to qg} \quad = \quad C_F \frac{8\pi\alpha_s(\mu^2)}{(p_{\tilde{i}} + p_{\tilde{k}})^2} \frac{1}{y} \left[\frac{2}{1-z} - (1+z) \right]. \tag{19}$$

In addition the momentum mapping

$$p_i = zp_{\tilde{i}} + k_\perp + y\,(1-z)\,p_{\tilde{k}}, \quad p_j = (1-z)p_{\tilde{i}} - k_\perp + yzp_{\tilde{k}}, \quad p_k = (1-y)p_{\tilde{k}}, \tag{20}$$

reduces in the collinear limit $y \to 0$ and $k_\perp \to 0$ to

$$p_i = zp_{\tilde{i}}, \quad p_j = (1-z)p_{\tilde{i}}, \quad p_k = p_{\tilde{k}}. \tag{21}$$

In particular, the spectator does not receive any recoil in the collinear limit. What emerged from the discussions is that the choice of the shower evolution variable is not arbitrary. This lead to the definition of an "infrared sensible" evolution variable [100]. The definition of infrared sensible is that both infinitely soft and collinear emissions should be classified as unresolved for any finite value of the evolution variable. The evolution variable based on the transverse momentum as in eq. (16) is an infrared sensible evolution variable. On the other hand, the choice of the energy of the emitted particle (in the centre-of-mass system of the emitter and the spectator) is not infrared sensible. There are configurations where the emitted particle has finite energy, but which are infinitely collinear.

10.3 MULTIPLE SOFT EMISSIONS

The shower algorithms above are derived from first-order perturbation theory. The singular functions entering the Sudakov factor are derived from tree-level matrix elements, where a single particle becomes either soft or collinear. The parton shower therefore reproduces the leading-log behaviour of the matrix elements in the limit of an infinitely large hierarchy between the scales of each successive emission. No claim is made to describe correctly the case of multiple emissions at the same scale. Nevertheless I would like to make a few comments what could be expected from future algorithms incorporating these effects. I will discuss multiple soft emissions, as derived from the singular behaviour of the matrix elements. If a single gluon becomes soft, the square of a partial tree amplitude factorises according to

$$\left| A_n^{(0)}(p_a, p_1, p_b, ...) \right|^2 \quad = \quad 4 \frac{s_{ab}}{s_{a1}s_{1b}} \left| A_{n-1}^{(0)}(p_a, p_b, ...) \right|^2. \tag{22}$$

In any strongly ordered limit the emission of r soft gluons is described by [101]

$$\left| A_n^{(0)}(p_a, p_1, ..., p_r, p_b, ...) \right|^2 \quad = \quad 4^r \frac{s_{ab}}{s_{a1}s_{12}s_{23}...s_{rb}} \left| A_{n-r}^{(0)}(p_a, p_b, ...) \right|^2. \tag{23}$$

This formula is valid for any strong ordering

$$E_{\sigma(1)} \ll E_{\sigma(2)} \ll ... \ll E_{\sigma(r)} \ll E_a, E_b, \tag{24}$$

where σ is permutation of $(1, 2, ..., r)$. However, if two gluons are emitted at the same scale $E_1 \approx E_2$ we have a more complicated formula. In the limit of two soft gluons a partial tree amplitude factorises according to [102]

$$\left| A_n^{(0)}(p_a, p_1, p_2, p_b, ...) \right|^2 \quad = \quad |\mathrm{Eik}(p_a, p_1, p_2, p_b)|^2 \left| A_{n-2}^{(0)}(p_a, p_b, ...) \right|^2, \tag{25}$$

with

$$|\mathrm{Eik}(p_a, p_1, p_2, p_b)|^2 = 16 \frac{s_{ab}}{s_{a1}s_{12}s_{2b}} + 8 \left[\frac{(s_{a12}s_{2b} + s_{a1}s_{12b} - s_{a12}s_{12b})^2}{s_{a12}^2 s_{12}^2 s_{12b}^2} \right.$$
$$\left. + \frac{s_{ab}^2}{s_{a1}s_{2b}s_{a12}s_{12b}} + \frac{s_{ab}}{s_{12}} \left(\frac{1}{s_{a1}s_{12b}} + \frac{1}{s_{a12}s_{2b}} - \frac{1}{s_{a1}s_{2b}} - \frac{4}{s_{a12}s_{12b}} \right) \right]. \tag{26}$$

The expression in the square bracket does not contribute to any strongly ordered limit ($E_1 \ll E_2 \ll E_a, E_b$ or $E_2 \ll E_1 \ll E_a, E_b$). In the strongly ordered limit it is less singular than the first term. However for $E_1, E_2 \ll E_a, E_b$ both terms scale like $1/\lambda^4$, if the momenta of the soft gluons are rescaled by λ.

11. HIGGS BOSON PRODUCTION VIA GLUON FUSION AT THE LHC: A COMPARATIVE STUDY [19]

11.1 INTRODUCTION

The primary motivation for ongoing and impending physics programmes at the Tevatron and LHC is to gain insight into the nature of electroweak symmetry breaking. The great majority of the effort in this direction is devoted to the hunt for the Higgs boson, the origin of this symmetry breaking in the Standard Model [103–106].

Of all the ways in which the Standard Model Higgs boson can be produced, the gluon fusion process [107], in which it couples to colliding gluons via a top quark loop, has the largest cross section for Higgs boson masses less than $\sim$700 GeV. This process will be extremely important in detecting and studying the Higgs boson at the LHC in the low mass region, favoured by fits of the Standard Model to electroweak precision data [108] and in part also by direct searches at the Tevatron [109, 110], for which Higgs boson decays into two photons are expected to give a clean experimental signal.

Although observing a narrow resonance in the diphoton invariant mass spectrum should be possible using only the experimental data [111], determining the quantum numbers and couplings of the resonance i.e. determining that it really is a fundamental scalar, in particular, the Standard Model Higgs boson, requires Monte Carlo simulations to predict distributions for both signals and backgrounds.

In recent years Monte Carlo event generators have been the subject of great theoretical and practical developments, most significantly in the extension of existing parton shower simulations to consistently include exact next-to-leading order (NLO) corrections [12–15, 25, 112–125] and, separately, in the consistent combination of parton shower simulations and high multiplicity tree-level matrix element generators [4, 5, 65, 70, 126–129]. The state-of-the-art in fixed order predictions has also undergone major improvement, resulting in fully differential Monte Carlo predictions at next-to-next-to-leading order (NNLO) for inclusive Higgs production and, simultaneously, NLO accuracy for production in association with a hard jet [130–133].

In this article we present predictions from most of the Monte Carlo simulations arising from this theoretical activity, for the gluon fusion process at the LHC, at the expected[20] initial hadronic centre-of-mass energy, $\sqrt{s} = 10$ TeV. As well as updating existing results, based on the originally forecast 14 TeV centre-of-mass energy, this document represents a broad comparative study among a number of fundamentally different Monte Carlo approaches, hence it also serves to gauge their relative merits and gauge the stability of our theoretical predictions with respect to the various methods.

In the following we shall concisely review the pertinent features of the simulations included in the study, prior to presenting results for a variety of simple observables concerning the gluon fusion production channel. At the end of the article we summarise the results and comment on the readiness of these theoretical tools for much anticipated Higgs analysis.

11.2 MATRIX ELEMENTS AND PARTON SHOWERS

In modern experimental particle physics, shower Monte Carlo (SMC) programs have become an indispensable tool for data analysis. From a user perspective, these programs provide an approximate but extremely detailed description of the final state in a high-energy process involving hadrons. They provide a *fully exclusive* description of the reaction, as opposed to fixed-order QCD calculations, which are only suitable for the computation of *inclusive* quantities.

After a latency period, research in SMC is once again very active, with significant advances being made in the last decade. In general these developments can be grouped into two main classes:

[19] Contributed by: M. Grazzini, K. Hamilton, S. Höche, F. Maltoni, C. Oleari, S. de Visscher, J. Winter

[20] At the time the study was started. The current plan is for running at 7 TeV and 14 TeV, which leaves this study nicely in the mod range.

1. The merging of leading-order matrix elements (ME), characterized by a high number of final-state partons, with parton showers (PS). Examples of such methods are the CKKW matching scheme [4, 126], the MLM matching procedure [70] as well as the newer merging schemes based on truncated parton showers [65, 129].
2. The interfacing of NLO calculations (that are typically available only with a small number of legs in the final state) with parton shower simulations (MC@NLO [12] and POWHEG [14]).

All of them have to face the same problems: avoiding overcounting of events as well as the presence of dead regions.

11.3 MONTE CARLO PROGRAMS FOR THE STUDY

With the exception of the HNNLO program all of the Monte Carlo programs used in this study fall into one or other of the two classes described above, implementing some form of matching/merging between fixed order calculations and parton shower simulations. In order to have a more full comparison of the available Monte Carlo tools, we also include HNNLO, which is based on a fixed order NNLO calculation of Higgs-boson production via gluon fusion.

11.31 HNNLO

The HNNLO program is based on an extension of the NLO subtraction formalism to NNLO, as described in ref. [132].

The calculation is organized in two parts. In the first part, the contribution of the regularized virtual corrections is computed up to two-loop order. In the second part, the cross section for the production of the Higgs boson in association with one jet is first evaluated up to NLO, i.e. up to $\mathcal{O}(\alpha_s^4)$), using conventional NLO subtraction methods. Now, since the $H + \mathrm{jet}$ cross section is divergent when the transverse momentum, q_{T}, of the Higgs boson becomes small, a further counterterm must be subtracted to make the result finite as $q_{\mathrm{T}} \to 0$. To this end the program uses counterterm introduced in ref. [132]. Having regularized the real and virtual parts, the two contributions can be combined to reconstruct the full cross section. Organizing the differential cross section in this way, one can construct a parton-level event generator with which arbitrary infrared safe quantities can be computed. The present version of the program includes the decay modes $H \to \gamma\gamma$ [132], $H \to WW \to l\nu l\nu$ and $H \to ZZ \to 4$ leptons [133].

The calculation is performed in the large top-mass approximation. This is known to be a good approximation provided that the Higgs boson is not too heavy and the transverse momenta of the final state jets are not too large.

11.32 MadGraph / MadEvent

Besides the possibility of generating processes in a long and extensible list of theoretical models (SM, MSSM, Higgs effective theory ...), the Monte-Carlo generator MADGRAPH/MADEVENT (MG/ME) [81] is also intended to simulate accurately the QCD radiation from initial and final states when coupled to a parton shower simulation. Such an aim can be achieved by using jet matching and a phase-space slicing technique

In MG/ME, three jet matching schemes (using Pythia 6.4 [18]) are implemented, namely the MLM [70], the $k_\perp$-MLM [134, 135] and the shower-$k_\perp$ [135] schemes. While the first two methods work with both virtuality and p_T-ordered showers, the third one only works with the p_T-ordered showers. For each of these methods, no analytic Sudakov reweighting of the events is performed, instead showered events are rejected if they are not matched to the ME-level partons. A detailed comparison of the $k_\perp$-MLM and shower-$k_\perp$ behaviours has shown that their respective outputs are very similar for the production of heavy colored particles in the SM and beyond [135]. In addition, a comparison for the production of W+jets events between the results from $k_\perp$-MLM and other simulation chains with different matching schemes has led to a similar conclusion [134].

Having the computation of the effective coupling between a scalar or pseudo-scalar to the gluons, the accurate simulation of the production of a Higgs boson accompanied by initial and final-state radiation is therefore relatively straightforward. In order to compare the MG/ME production with the results from the other generators considered in this work, we choose to simulate $H + 0, 1, 2$ partons at the ME level and apply the $k_\perp$-MLM scheme with $Q_{\rm cut}^{\rm ME} = 10$ GeV and $Q_{\rm match} = 15$ GeV. These choices were made in agreement with the smoothness of ME→PS transition regions in the $2 \to 1$ and $1 \to 0$ differential jet rates distributions.

11.33 MC@NLO

The MC@NLO method [12] was the first one to give an acceptable solution to the overcounting problem. The generality of the method has been explicitly proven by its application to processes of increasing complexity, such as heavy-flavour-pair [112] and single-top [113] production.[21] The basic idea of MC@NLO is that of avoiding the overcounting by subtracting from the exact NLO cross section its approximation, as implemented in the SMC program to which the NLO computation is matched. Such approximated cross section (which is the sum of what have been denoted in [12] as MC subtraction terms) is computed analytically, and is SMC dependent. On the other hand, the MC subtraction terms are process-independent, and thus, for a given SMC, can be computed once and for all. In the current version of the MC@NLO code, the MC subtraction terms have been computed for Herwig 6 [136], but extensions to other SMC are possible. In general, the exact NLO cross section minus the MC subtraction terms does not need to be positive. Therefore MC@NLO can generate events with negative weights. For the processes implemented so far, negative-weighted events are typically about 10–15% of the total. Their presence does not imply a negative cross section, since at the end physical distributions must turn out to be positive, but affects the overall efficiency of the simulation.

The features implemented in MC@NLO can be summarized as follows:

- Infrared-safe observables have NLO accuracy.
- Collinear emissions are summed at the leading-logarithmic level.
- The double logarithmic region (i.e. soft and collinear gluon emission) is treated correctly since Herwig 6 uses an angular-ordered shower.

11.34 POWHEG

The POWHEG (Positive Weight Hardest Emission Generator) method was proposed in ref. [14]. This method overcomes the problem of negative weighted events, and is not SMC specific. In the POWHEG method, the hardest radiation is generated first, with a technique that yields only positive-weighted events, using the exact NLO matrix elements. The POWHEG output can then be interfaced to any SMC program that is either p_T-ordered, or allows the implementation of a p_T veto.[22] However, when interfacing POWHEG to angular-ordered SMC programs, the double-log accuracy of the SMC is not sufficient to guarantee the double-log accuracy of the whole result. Some extra soft radiation (technically called vetoed-truncated shower in ref. [14]) must also be included in order to recover double-log accuracy. In fact, angular ordered SMC programs may generate soft radiation before generating the radiation with the largest p_T, while POWHEG generates it first. When POWHEG is interfaced to shower programs that use transverse-momentum ordering, the double-log accuracy should be correctly retained if the SMC is double-log accurate. The ARIADNE program [97], Pythia 6.4 [18] (when used with the new showering formalism), ADICIC++ [91] and the new parton showers based on the Catani–Seymour dipole formalism [89, 90] adopt transverse-momentum ordering, and aim to have accurate soft resummation approaches, in the limit of large number of colours.

[21] A complete list of processes implemented in MC@NLO can be found at
`http://www.hep.phy.cam.ac.uk/theory/webber/MCatNLO.`

[22] All SMC programs compatible with the *Les Houches Interface for User Processes* [2] should comply with this requirement.

Up to now, it has successfully been applied to Z pair hadroproduction [117], heavy-flavour production [118], e^+e^- annihilation into hadrons [119] and into top pairs [125], Drell-Yan vector boson production [120, 122], W' production [137], Higgs boson production via gluon fusion [121, 123], Higgs boson production associated with a vector boson (Higgs-strahlung) [121], single-top production [124] $Z + 1$ jet production [138], and, very recently, Higgs production in vector boson fusion [139].

THE POWHEG BOX The POWHEG BOX is an automated package able to construct a POWHEG implementation of a NLO process, given the following ingredients:

1. The list of all flavour structures of the Born processes.
2. The list of all the flavour structures of the real processes.
3. The Born phase space.
4. The Born squared amplitude, the color correlated and spin correlated Born amplitudes. These are common ingredients of NLO calculations regularized with a subtraction method.
5. The real matrix elements squared for all relevant partonic processes.
6. The finite part of the virtual corrections, computed in dimensional regularization or in dimensional reduction.
7. The Born color structure in the large limit of the number of colors.

The plots in this article were obtained using the POWHEG BOX, and the completion of the shower has been done both with Pythia 6 and with Herwig 6.

11.35 Herwig++

Herwig++ builds and improves upon the physics content of the parent Herwig 6 program, particularly in regard to the accurate simulation of QCD radiation. A major success of the original Herwig 6 program was in its modeling the effects of soft gluon interference, specifically the *colour coherence* phenomenon, whereby the intensity of soft gluon radiation, emitted at wide angles with respect to a bunch of collimated colour charges, is found to be proportional to the *coherent* sum of emissions from the constituents i.e. the jet parent [101, 140–148]. This effect is manifest in the perturbative series as large soft logarithms and is implemented as an angular ordering of successive emissions in the parton shower.

A further significant accomplishment of the POWHEG [14] formalism is in fully catering for such effects through a careful decomposition of the angular-ordered parton shower into a truncated shower, describing soft wide angle radiation, the hardest emission, as described above, and a vetoed shower comprised of increasingly collinear emissions. In doing so the formalism provides a means of distributing the highest p_T emission in an event according to the exact real-emission matrix element, including resummation effects, without degrading or otherwise disturbing the resummation and colour-coherence properties inherent to the parton shower.

The facility to perform truncated showers is absent from the older fortran Herwig 6 program but is implemented in the new Herwig++ program, which also includes its own native, independent, POWHEG simulation for the gluon-fusion process, the results of which are presented in section 11.5

11.36 Sherpa

Sherpa is a multi-purpose Monte-Carlo event generation framework for colliders [22,23]. The main goal of this project is a proper simulation of the perturbative aspects of the collision, although significant improvements have been made over the past years regarding the simulation of non-perturbative dynamics, like the process of hadronisation. One of the key features of the Sherpa program is a general implementation of a novel technique for combining tree-level matrix elements with parton showers, in arbitrary QCD or QCD-associated processes [65]. To this end, Sherpa makes use of its two internal tree-level matrix element generators AMEGIC++ [149] and Comix [150], which are capable of simulating both Standard Model

(AMEGIC++ and Comix) and beyond Standard Model (AMEGIC++) reactions with high-multiplicity final states. Soft and collinear parton radiation is generated in Sherpa by means of a parton shower based on Catani–Seymour dipole factorisation [89]. This formalism has apparent advantages compared to more conventional parton showers, which are based on strict $1 \to 2$ splittings and often lack the notion of a well-defined spectator parton. In the new approach, the recoil partner of a splitting parton is always a single external particle. This turned out to be an important ingredient when combining parton-shower evolution with higher-order tree-level matrix elements.

The technique for merging matrix elements and parton showers, which is employed by Sherpa, is based on phase-space slicing. A detailed description of the corresponding algorithm, and its relation with other tree-level merging techniques, can be found in [65]. The method has recently been applied to mixed QCD and electroweak processes, in particular photon and diphoton production [151] and proves to give a consistent and reliable description of data from deep-inelastic lepton-nucleon scattering [152]. Compared to the CKKW algorithm [4, 126], the new merging scheme of ref. [65] is more sophisticated and improves over CKKW by including truncated vetoed parton showers. The results are more accurate, with respect to those of the CKKW approach, which has been employed in former versions of the Sherpa event generator with already great success [134, 153–155].

11.4 Parameters for the study

In the following, we present results for the pp LHC at a centre-of-mass energy of 10 TeV. The common features of the analysis are the following:

- *Model*: We work in the Standard Model in the large top-mass limit. In this approximation the couplings between the gluons and the Higgs boson are described by a dimension-five effective interaction
$$\mathcal{L}_h = -\frac{1}{4}\, g_h\, G^a_{\mu\nu} G^a_{\mu\nu} H, \tag{27}$$
with $g_h = \alpha_s/(3\pi v)$.
- *Event samples*: The analysis was performed on the generated final state and, with the exception of HNNLO, after parton showering. Hadronization effects were included for the MC@NLO and POWHEG results only.
A Higgs boson mass of $m_H = 120$ GeV was assumed. Tree-level predictions were generated using the leading order CTEQ6l1 PDF set [156], while generators employing the POWHEG method used the next-to-leading order PDF set CTEQ6m [156]. In both cases the parametrisation of the strong coupling was chosen accordingly. All partons (excluding the top quark) were taken to be massless and their Yukawa couplings were neglected.
- *Jet definitions*: Jets were defined using the longitudinally invariant $k_\perp$-algorithm with $D = 0.7$ in the implementation of FastJet [157]. They were required to lie within a rapidity range of $|\eta| < 4.5$ and have transverse momenta of $p_T > 20$ GeV.

11.5 RESULTS

In this section we present and discuss results obtained for some key observables in the analysis of Higgs production via gluon fusion. As already mentioned, we do not expect the LO matched results to provide reliable information on total rates, so we have normalized the corresponding curves, from MADGRAPH/MADEVENT and Sherpa, to the HNNLO result. On the other hand, having NLO accuracy, Herwig++, MC@NLO and POWHEG have not been rescaled.

In fig. 25, we plot the transverse momentum of the Higgs boson and, in fig. 26, its rapidity. The Higgs boson p_T, in particular, is determined by the recoiling QCD radiation, both soft and hard, and, exactly as for Drell-Yan, it is therefore a key observable. The blow up of the small-p_T region (left panel of fig. 25) shows quite good agreement among the various MC approaches with predictions being typically

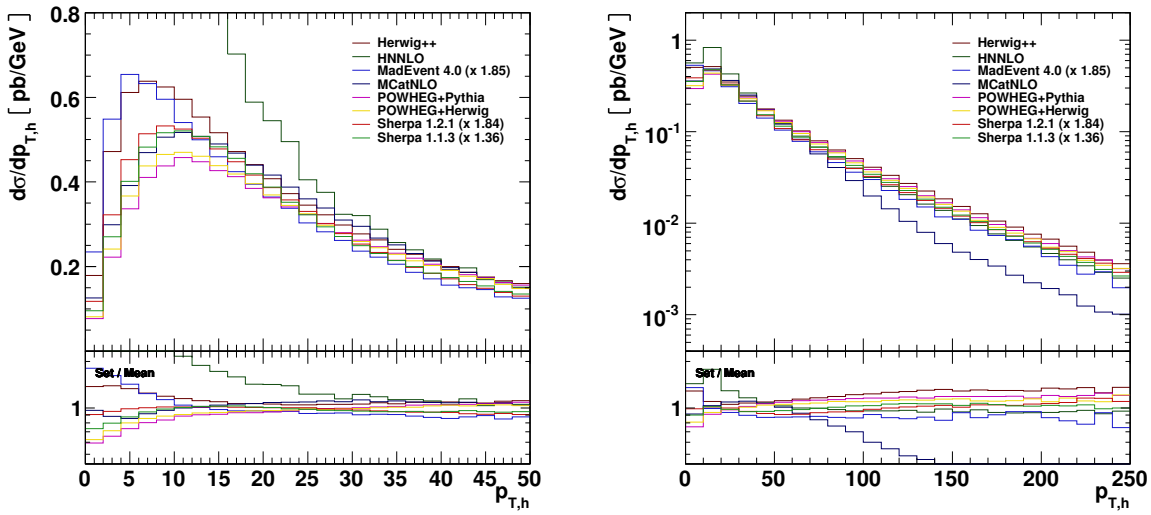


Fig. 25: The transverse momentum spectrum of the Higgs boson. Tree-level predictions have been rescaled by the global factors indicated in the legend. The lower panels display the ratio of individual results and the average of all histograms, excluding the results from HNNLO.

peaked in the range between 5 and 10 GeV. The obvious excess in the HNNLO prediction, at low-p_T, is expected on the grounds that it is based on a fixed order computation, hence, it does not resum the effects of multiple soft emissions, which are essential for a proper description of the $p_T = 0$ region. At higher values of the p_T, the agreement is also excellent, apart from MC@NLO which shows a steeper behaviour with respect to the results obtained by the POWHEG method and with the matching. As already pointed out in Refs [121, 123], this is due to NNLO terms in the POWHEG formula. It is, however, important to note that, for all the NLO codes, this particular distribution can be predicted only at LO, i.e. no $H + 2$ partons contribution and no $H + 1$ parton one-loop contributions are included. From this point of view, it is reasonable to expect the shape to be sensitive to variations in the renormalization and factorization scales, although, in practice, this sensitivity is much milder due to the resummation of higher order corrections (i.e. the shower). In any case, it is both remarkable to see that the predictions based on the POWHEG method and the ME+PS matching show such good agreement, particularly considering the fundamental differences in their methodology.

In the rapidity distributions of fig. 26, the HNNLO result shows all of its NNLO content: in fact, this is the only plot that receives contributions from the two-loop diagrams.

Figure 27 shows the jet p_T distributions for the four hardest jets (ordered in p_T). Once again the agreement among the various approaches is very good, with MC@NLO leading to significantly less events at very high p_T's; this lower number of events is in exact correspondance with that seen for the Higgs boson transverse-momentum distribution and bears the same explanation. A particularly interesting feature is the agreement found on the 3rd and 4th jet spectra. Only Sherpa has included the corresponding tree-level hard matrix elements, while all other predictions contain only one (NLO codes) or two (HNNLO and MADGRAPH/MADEVENT) hard partons. This good agreement is a mere coincidence, since in POWHEG, MC@NLO and Herwig++ these extra jets come from the shower, and are therefore only corect in the strict collinear limit. Similar comments hold for the pseudorapidity distributions of fig. 28.

Larger discrepancies are instead present for more exclusive quantities, such as the distance in the η-ϕ plane between the two leading jets, ΔR_{12}, shown in fig. 29. Indeed, we start appreciating here some interesting differences in shape: Herwig++, for example, predicts much steeper distribution than Sherpa.

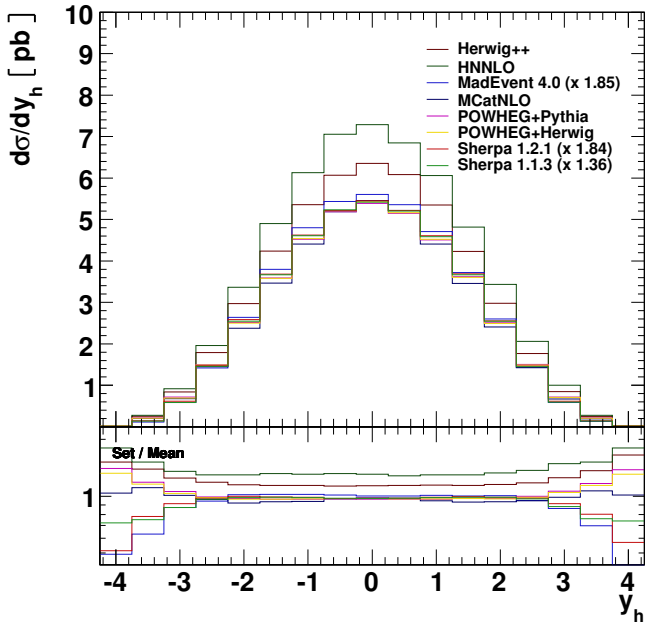


Fig. 26: The rapidity distribution of the Higgs boson. See fig. 25 for details.

Finally, in fig. 30, we plot the jets rates, which, once again, agree within 50% uncertainty even for higher multiplicities.

11.6 CONCLUSIONS

We have reported on the first comparison among several different Monte Carlo approaches to the simulation of Higgs boson production in gluon fusion. Apart from some basic choices, such as the parton distribution functions, the collider energy and our choosing to use an effective theory where the top quark mass has been taken to infinity, no detailed tuning has been performed. The main idea being that of an "out-of-the-box" comparison among various codes, all of which represent the state-of-the-art in Monte Carlo tools.

The upshot of our comparison is that, apart from the overall normalization, which is only reliable in NLO and NNLO codes, the various approaches give consistent results within their expected range of validity.

ACKNOWLEDGEMENTS

We thank Jeppe Andersen and Michelangelo Mangano for useful discussions.

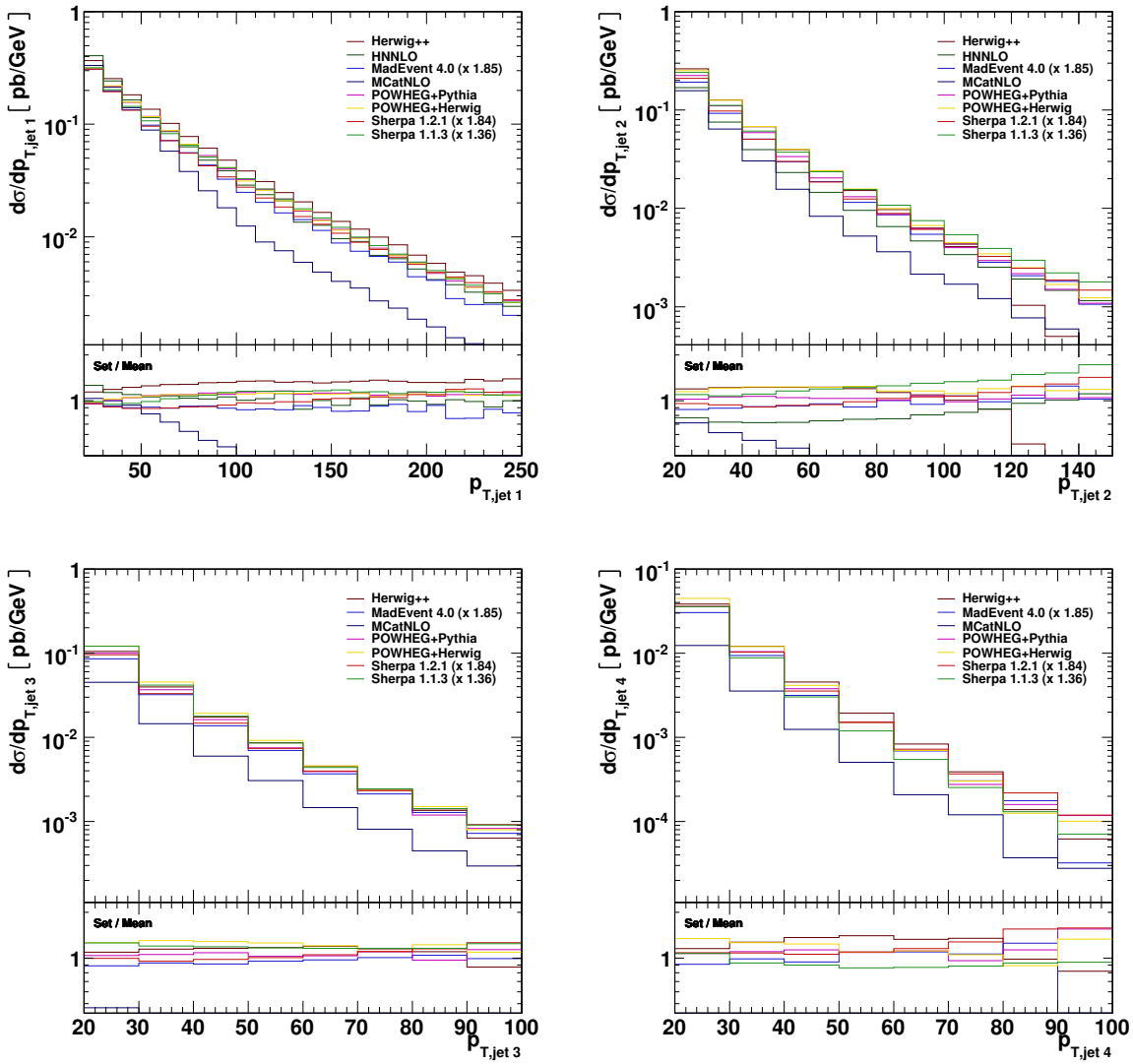


Fig. 27: The transverse momentum spectra of the first four hardest jets, ordered in p_T, accompanying the Higgs boson. See fig. 25 for details.

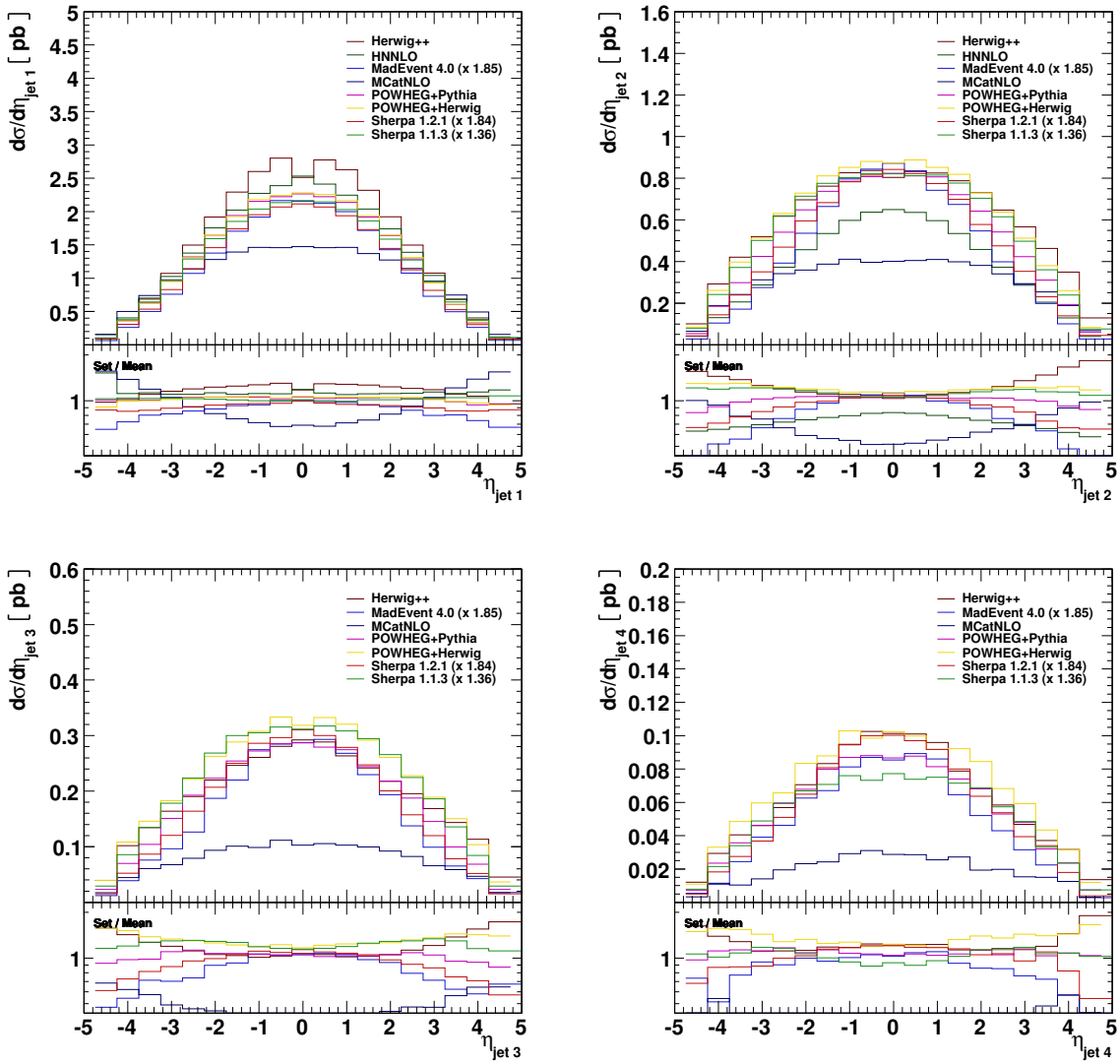


Fig. 28: The pseudorapidity distributions of the first four hardest jets, ordered in p_T, accompanying the Higgs boson. See fig. 25 for details.

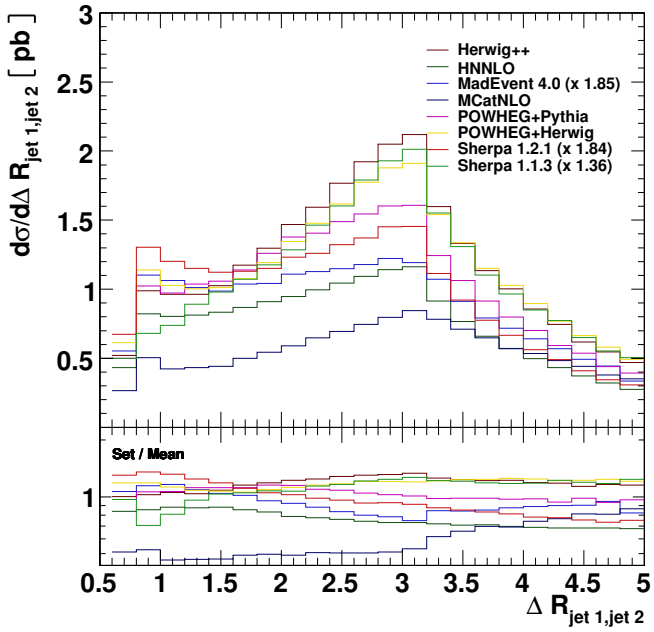


Fig. 29: The separation in the η-ϕ plane of the two leading jets accompanying the Higgs boson. See fig. 25 for details.

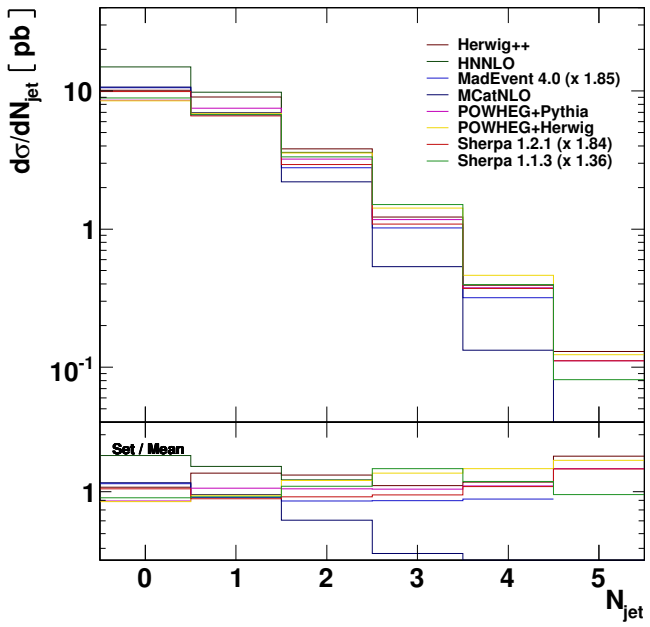


Fig. 30: The production rates for $N_{\rm jet}$ additional jets accompanying the Higgs boson. See fig. 25 for details.

12. $Wb\bar{b}$ IN THE HIGH-p_T HW REGION [23]

12.1 INTRODUCTION

The main aim of this section is to study the effect of next-to-leading order (NLO) QCD corrections to the $pp \to Wb\bar{b}$ process, in the region of phase space which is relevant for the *highly boosted* WH analysis with $H \to b\bar{b}$ [158, 159].The details of the setup for the NLO computation in this kinematic region as well as results for the scale dependence of the total cross sections are presented in Section 12.2. The application of the NLO calculation to the experimental analysis on which the mentioned Higgs search is based is not straightforward, since the latter addresses the simulation of the backgrounds through the use of a parton shower Monte Carlo (MC) algorithm which is applied on top of the QCD leading order (LO) matrix element calculation.

While for an increasing number of processes specific Monte Carlo generators have been made available to combine the NLO calculation with a parton-shower based MC generator, as in the `MC@NLO` program [12] or in the new `POWHEG` method [15, 160], no such generator is available for the $pp \to Wb\bar{b}$ process yet. It is therefore useful to define a re-weighting procedure for $pp \to Wb\bar{b}$ events generated by the use of a parton shower Monte Carlo, which for this specific case is Herwig [17], based on the distributions predicted by the NLO calculation.

An important feature of the *highly boosted* WH analysis is the application of a tight jet veto. In the cut-based analysis [159] this veto is applied at $p_T = 20$ GeV, while in a more refined likelihood based analysis [161] the maximum p_T considered for a possible additional non b-jet in the event is 60 GeV. The NLO QCD correction depends significantly on the p_T of an additional non b-jet in the event and this is addressed in this study. This situation is analogous to other backgrounds for Higgs boson signals, as is the case of the $t\bar{t}b\bar{b}$ process as a background for the search for $t\bar{t}H$ with $H \to b\bar{b}$, also considered in the present proceedings.

In the specific case of the $q\bar{q}' \to Wb\bar{b}$ process, the real corrections which enter the NLO computation include an additional gluon-induced process, $q(\bar{q})\, g \to Wb\bar{b}j$, which, given the gluon luminosity at the LHC, makes the NLO correction very large. However, by constraining the transverse momentum of the extra non-b-jet through a jet veto one can reduce the impact of this extra channel, which then improves the scale dependence of the NLO computation, as long as such extra kinematical cut is not too restrictive.

12.2 $Wb\bar{b}$ PRODUCTION AT NLO QCD WITH A TIGHT JET VETO

The calculation of the total and differential cross sections for $W^{\pm}b\bar{b}$ production at NLO QCD for this study is based on Refs. [162–164], which includes full b-quark mass effects. The input parameters are chosen as follows: $m_b = 4.79$ GeV, $m_t = 173.1$ GeV, $M_W = 80.399$ GeV, $s_W^2 = 0.223$, $G_F = 1.16639 \cdot 10^{-5}$ (the weak coupling being defined as: $g_W = (8M_W^2 G_F/\sqrt{2})^{-1/2}$). The contribution from the third generation quarks in the initial state is neglected and the non-zero CKM matrix elements are chosen to be: $V_{ud} = V_{cs} = 0.974$ and $V_{us} = V_{cd} = 0.227$.

Parton-level LO results (calculated as a cross-check) are obtained using the one-loop evolution of α_s and the CTEQ6L1 [82] set of PDFs with $\alpha_s(M_Z) = 0.130$, while parton level NLO results are obtained using the two-loop evolution of α_s and the CTEQ6M set of PDFS with $\alpha_s(M_Z) = 0.118$.

Renormalization (μ_r) and factorization (μ_f) scales are set to be equal and varied by a factor of two around a central value $\mu_0 = [M_W^2 + (p_T^b)^2 + (p_T^{\bar{b}})^2]^{1/2}$, where p_T^b and $p_T^{\bar{b}}$ are the transverse momenta of the jets generated by b and $\bar{b}$ respectively.

The jets are constructed using the k_T jet algorithm with pseudo-cone size $R = 0.3$ and the parton momenta are recombined within a jet using the so called E-invariant scheme (=sum of momenta of the constituents). We study two samples of events, i.e.

- events with just 2 b jets;

[23] Contributed by: F. Febres Cordero, G. Piacquadio, L. Reina, D. Wackeroth

- events with 2 b jets and 1 non-b jet;

where b and non-b jets are identified imposing the following cuts:

- $p_T^b > 30$ GeV, $|\eta^b| < 2.5$;
- $p_T^W > 200$ GeV, $|\eta^W| < 2.5$;
- 15 GeV $< p_T^{non-b} <$ 60 GeV, $|\eta^{non-b}| < 5$;
- $p_T^{b\bar{b}} > 200$ GeV;
- $R_{b\bar{b}} < 1.2$.

p_T^b and $|\eta^b|$ denote the transverse momentum and pseudorapidity of either one of the two b jets, p_T^W and $|\eta^W|$ are the transverse momentum and pseudorapidity of the W boson, p_T^{non-b} and $|\eta^{non-b}|$ denote the transverse momentum and pseudorapidity of the non-b jet, $R_{b\bar{b}}$ is the relative separation between the b and $\bar{b}$ jets, and $p_T^{b\bar{b}}$ the transverse momentum of the $b\bar{b}$ 2-jet system.

The renormalization and factorization scale dependence of the total cross sections for $W^{\pm}b\bar{b}$ production at the LHC ($\sqrt{s} = 14$ TeV) using this setup is studied in Table 5 and in Fig. 31. The NLO *exclusive* cross section corresponds to the *2 b-jet only* sample, while the *inclusive* one corresponds to the sum of both *2 b jet only* and *2 b jets and 1 non-b jet* samples. Thus, in Fig. 31, the curve labeled as *NLO Inc* is the sum of the curves labeled as *NLO Exc* and $2b + j$ *only*.

Table 5: LO, NLO *inclusive*, and NLO *exclusive* cross sections (in fb) for $W^+b\bar{b}$ and $W^-b\bar{b}$ production at $\sqrt{s} = 14$ TeV. Listed separately are also the cross sections (in fb) for the $Wb\bar{b} + j$ channel alone. The central values correspond to $\mu_r = \mu_f = \mu_0$, while the upper and lower bounds represent the maximal upper and lower variation obtained when varying $\mu_r = \mu_f$ between $\mu_0/2$ and $2\mu_0$.

Process	LO	NLO inclusive	NLO exclusive	$Wb\bar{b} + j$
$W^+b\bar{b}$	138^{+31}_{-24}	155^{+9}_{-11}	43^{+17}_{-34}	112^{+42}_{-28}
$W^-b\bar{b}$	76^{+17}_{-13}	90^{+7}_{-7}	26^{+9}_{-17}	64^{+24}_{-16}

It is interesting to note that, contrary to the findings in Refs. [162–164], in this kinematic regime the inclusive cross section exhibits less scale dependence than the exclusive one. Different factors may contribute to this different behavior, but mainly the kinematic cuts used in the present analysis make the jet veto much more selective and enhance the effect of unbalanced scale dependent logarithms in the exclusive cross section, while their effect tends to compensate between $Wb\bar{b}$ and $Wb\bar{b} + j$ production in the inclusive case.

12.3 COMPARISON OF FIXED-ORDER AND PARTON-SHOWER RESULTS

While the parameters used for the fixed-order LO and NLO computation were already presented in the previous section, the parameters used for the matrix element computation $pp \to Wb\bar{b}$ which serves as an input to the parton-shower algorithm are listed in the following:

- LO parton density functions CTEQ6L1;
- strong coupling constant $\alpha_s(M_Z) = 0.130$, consistent with the PDF set used;
- QCD renormalization and factorization scales both equal and set to $Q = \sqrt{M_W^2 + p_{T,W}^2}$;
- one-loop running of α_s.

The LO matrix element computation and initial three-body phase integration is based on the AcerMC generator [165] (v. 3.5). This is then passed as an external process to Herwig (v. 6.510), which applies the parton shower algorithm and produces the final state partons which are used as an input to an inclusive k_T jet clustering algorithm with parameter $R = 0.3$ and with the E-invariant parton momenta recombination scheme. The underlying event has been switched off during generation, while the effect of

hadronization has been removed by applying the jet clustering directly on top of the final state partons before hadronization. As opposed to the $Wb\bar{b}$ sample generated for the analysis presented in [159], the one-loop running of α_s was used instead of the two-loop running: this increased the LO cross section by ≈ 30 % with respect to the value used in the mentioned analysis.

The parton-level jets produced in this way can be compared to the fixed-order calculation described in Section 12.2 and will effectively include the approximative leading logarithmic resummation introduced by the parton shower algorithm. This involves a comparison of N produced jets with respectively the two or three leading jets of the LO or NLO fixed-order computation. The two b-jets are selected out of the N jets produced by the parton shower algorithm by matching them to the two b-quarks after radiation as traced in the Monte Carlo history:

$$\Delta R(\text{jet}, b - \text{quark}) < 0.4.$$

In the case where more than one jet is closer to the b-quark than $\Delta R = 0.4$, the b-quark is associated to the closest jet. As a cross-check of the eventual systematic uncertainty introduced by this choice, the same procedure is applied using the b-quarks before radiation, which are produced directly by the matrix element computation, and the difference in the results is found to be negligible. Finally, the leading additional non b-jet in the event is defined as the jet out of the N jets produced by the parton shower which, after excluding the two jets matched to the b-quarks, is the highest in transverse momentum. Events where only a single jet is matched to a b-quark are thrown away; this happens typically when the two b-quarks are closer than $\Delta R \approx 0.3$ and are therefore included into a single jet. After the jets are defined, the kinematic cuts defined in the previous section are applied: in the case of the cuts applied on the W boson, its momentum is considered after the effect of reshuffling of momenta operated by the parton shower algorithm. The distributions based on the jets obtained by using the parton shower result are labeled as LO+PS in the next section.

As a cross-check, also the result of the LO matrix element computation used as input to the parton shower algorithm is considered in the following and is labeled as LO only. In this case the two b-quarks partons resulting from the ME computation are used directly for the computation of the observables and for the application of the cuts, except for the cut on the additional non b-jet in the event, which is not applied. Both results are compared to the LO and NLO inclusive theory predictions obtained in the previous section, which are labeled as LO (theory) and NLO (theory). Note that NLO (theory) includes both the $q\bar{q}'$ initiated and $qg + \bar{q}g$-initiated processes. For comparison with the LO+PS result which is based on the $q\bar{q}' \to Wb\bar{b}$ matrix element only, we also include the NLO QCD prediction for the total cross section of the $q\bar{q}'$-initiated process separately, denoted as NLO($q\bar{q}'$) (theory).

12.4 RESULTS

The cross sections are listed in Table 6. The two LO predictions agree fairly well, even if the comparison is slightly biased by a marginally different choice of the renormalization and factorization scale between the two computations. Once the parton shower algorithm is run and the kinematic cuts are applied, including the additional jet veto, the cross section turns out to be $\approx 30\%$ lower with respect to the LO prediction. The complete NLO theory prediction, on the other side, predicts a cross section which is $\approx 14\%$ higher than the same LO prediction. As a result, the parton shower prediction (LO+PS) must be rescaled by an inclusive K factor of 1.64 in order to normalize it to the complete NLO computation. If only the $q\bar{q}'$ initiated process is included in the NLO prediction, the inclusive K factor only amounts to 1.16. The uncertainty on the complete NLO cross section estimated through variations of the factorization and renormalization scales by factors $1/2$ and 2 is $\approx 5\%$. The impact of PDF uncertainties has not been considered here.

The impact of the NLO corrections on the shapes of the distributions of most interest for the analysis are shown in the following, separately for $W^+b\bar{b}$ and $W^-b\bar{b}$, in Figs. 32-41. The crucial observable is the differential ratio between the NLO and the LO+PS predictions for the variable of interest: if this ratio is

Subprocess	LO	LO (theory)	LO+PS	NLO (theory)	NLO($q\bar{q}'$) (theory)
$W^+b\bar{b}$	144	138^{+31}_{-24}	97	155^{+9}_{-11}	111^{+6}_{-19}
$W^-b\bar{b}$	79	76^{+17}_{-13}	52	90^{+7}_{-7}	62^{+4}_{-10}
Sum	223	214^{+35}_{-27}	149	245^{+11}_{-13}	173^{+7}_{-21}

Table 6: Central value of cross sections (in fb) obtained for the $pp \to Wb\bar{b}$ process, comparing the five different methods. Details about the calculation of LO (theory) and NLO (theory) can be found in Section 12.2.

approximately flat, then it is possible to re-weight the Monte Carlo events produced by the parton shower algorithm by a simple inclusive K factor.

Also the NLO to LO+PS ratios for the distributions for the pseudorapidity of the leading (Fig. 33, top) and subleading (Fig. 33, bottom) b-jets are relatively flat, with the NLO computation predicting the b-jets to be slightly more central. On the contrary the W bosons are significantly less central (Fig. 34) in the NLO prediction, and, as a consequence, a significant correction affects the distribution of difference in pseudorapidity between the $b\bar{b}$ and W boson systems, in particular at large pseudorapidities (Fig. 35).

For the transverse momentum of the leading b-jet (Fig. 32, top), apart from some threshold effect at values close to 100 GeV, the ratio NLO to LO+PS is relatively flat. For the transverse momentum of the subleading b-jet (Fig. 32, bottom), the parton shower approximation does a perfect job and the ratio is perfectly flat.

A bit problematic is the case of the transverse momentum of the W boson system (Fig. 36) and, to a less extent, of the $b\bar{b}$ system (Fig. 37), where the NLO prediction is reliable towards higher transverse momenta, while, close to the threshold, it starts to be sensitive to the emission of multiple soft radiation (large logarithms). This is more severe for the distribution of the transverse momentum of the W boson system. In both cases, the LO+PS distribution could be re-weighted at large transverse momenta according to the NLO prediction, while the LO+PS based distribution itself could be used to extrapolate to the region close to the threshold. Correlations between the two distributions need clearly to be taken into account.

The parton shower algorithm is able to reproduce fairly well the shape of the distributions of the $b\bar{b}$ invariant mass (Fig. 38), of the $b\bar{b}W$ invariant mass (Fig. 39) and of the distance in pseudorapidity and azimuthal angle of the two b-jets (Fig. 40). In the case of the $b\bar{b}$ invariant mass, a small discrepancy is seen at very small invariant masses, again in the region where eventually large logarithmic corrections to the NLO computation can be expected.

Finally, the p_T of additional non b-jet distribution is shown in Fig. 41. The cross section corresponding to the events where no additional non b-jet above 15 GeV is found is condensed in the first bin of the distribution. As a general tendency, the NLO computation predicts significantly more radiation than the LO+PS prediction. There are two possible reasons for this:

1. The parton shower algorithm used in this study acts on top of the LO matrix element producing $Wb\bar{b}$, which means that it misses all radiation where the first splitting is of the type $g \to gg$ and the $b\bar{b}$ pair is produced later in the cascade, while the NLO prediction will include the case where the $b\bar{b}$ is produced right after the first $g \to gg$ splitting.
2. The NLO computation includes the process $q(\bar{q})g \to Wb\bar{b}j$ in the real correction term, which is strongly but not completely suppressed for small p_T of the additional non b-jet in the event. This is illustrated in Fig. 41, where we also show the theory prediction when only including the $q\bar{q}$-initiated process.

Given this overall tendency, one would expect that by moving the jet veto cut at parton level from 60 GeV to 20 GeV (as in the cut based analysis), the K factor correcting from LO+PS to the NLO prediction would move significantly from ≈ 1.6 more towards one, which would be a clear advantage for the WH

analysis. However, this requires a reliable prediction of the transverse momentum of the additional non b-jet of the event down to very low transverse momenta (if compared with the hard scale of the event which is around $p_T(b\bar{b}) \approx 200$ GeV). This is not provided by the NLO prediction, as the renormalization and factorization scale variation uncertainty of the *exclusive* 2 b-jet result reported in Table 5 of the last section demonstrates. In order to get a more reliable prediction in this kinematic region, some matching procedure between the parton shower and the NLO predictions is needed. Before being able to define such a matching procedure, further studies will be needed, based on parton shower Monte Carlo algorithms able to include the primary $g \to gg$ splittings (available in plain Herwig by using the $pp \to Wg$ matrix element), which would lead to a complete leading logarithmic prediction, or able to include the additional $q(\bar{q})g \to Wb\bar{b}j$ process, as Alpgen [64] (which however does not include the primary $g \to gg$ splittings). The use of these two algorithms will in fact allow to disentangle the two previously enumerated effects.

12.5 CONCLUSIONS

We have compared the NLO computation for the $Wb\bar{b}$ process with the predictions of the Herwig parton shower algorithm applied on top of the LO $Wb\bar{b}$ matrix element, considering for the first time the specific kinematic region explored by the high p_T WH analysis and the effect of an additional non b-jet veto (applied at parton level) of 60 GeV. Given the choice of PDFs, of strong coupling constant and renormalization and factorization scales already mentioned, we find that the cross section predicted by the NLO computation is a factor ≈ 1.6 higher than the parton shower prediction. Many of the differential distributions predicted by the parton shower algorithm agree reasonably well with the NLO computation. In general, the increase in event yield due to the NLO correction can be decreased by further lowering the jet veto: however, the NLO prediction suffers from substantial uncertainties in the kinematic region where radiation significantly softer than the scale of the process occurs and further studies are ongoing to make the prediction more reliable.

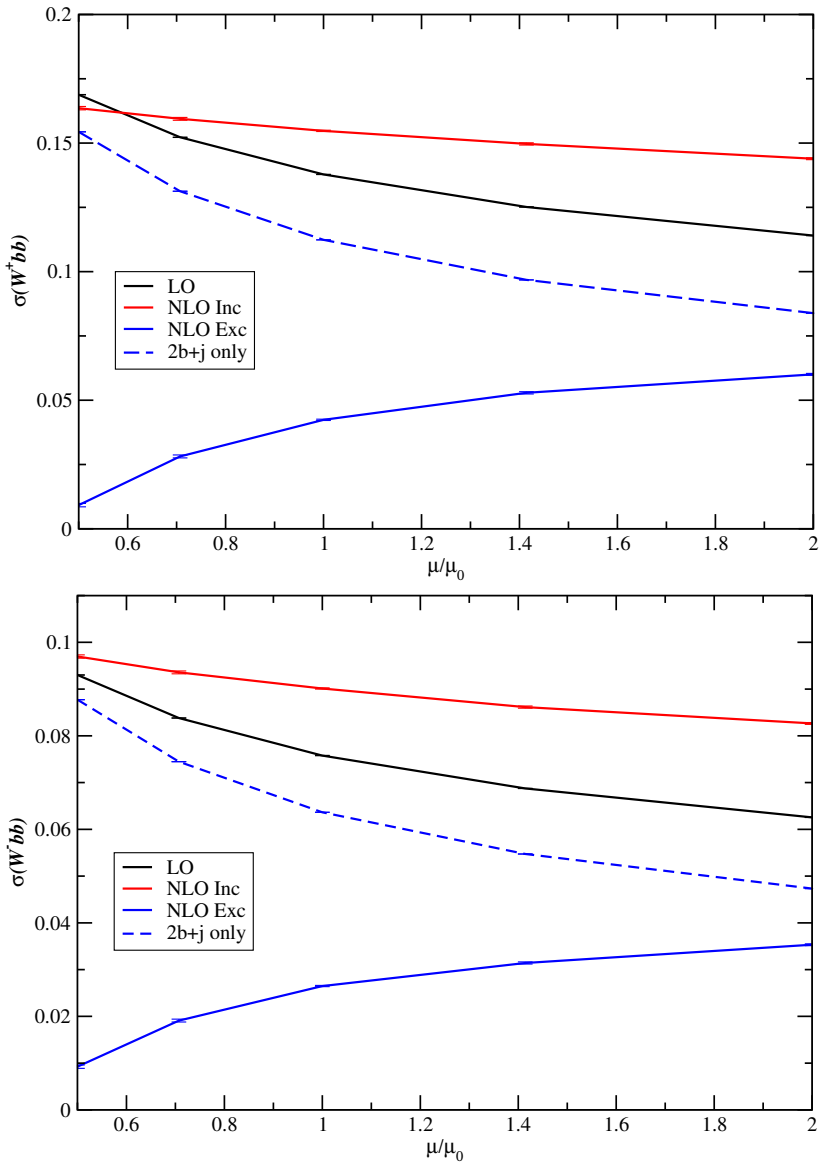


Fig. 31: Renormalization and factorization scale dependence for the parton level $W^+b\bar{b}$ (upper plot) and $W^-b\bar{b}$ (lower plot) LO (black, solid), NLO inclusive (red, solid), NLO exclusive (blue, solid) and $Wb\bar{b}+j$ (blue, dashed) production cross sections (cross sections are in pb). See Section 12.2 for details.

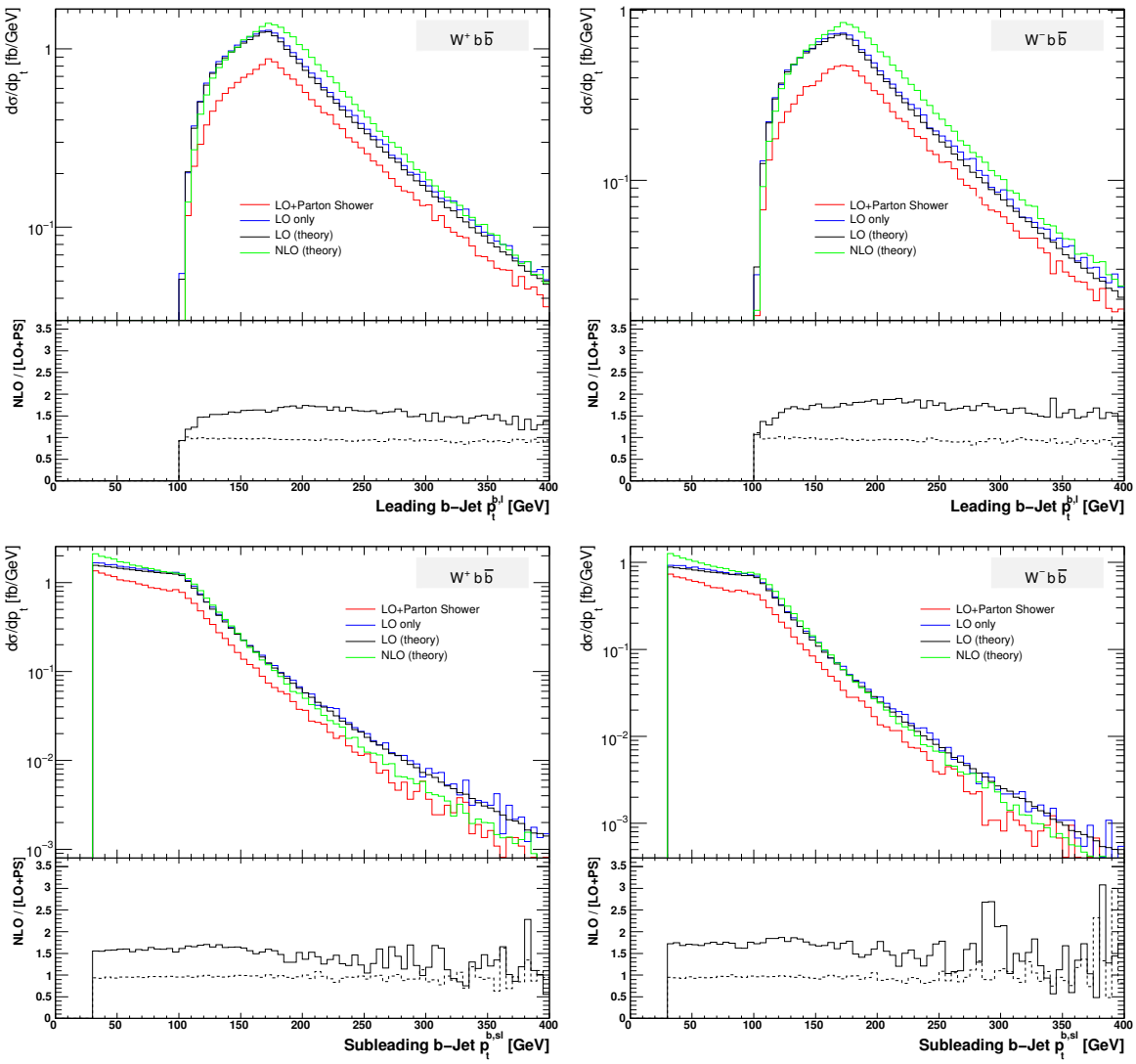


Fig. 32: Transverse momentum distribution of the *leading* and *sub-leading* b jet, for both $W^+b\bar{b}$ and $W^-b\bar{b}$ production. The lower window shows the ratio of NLO to LO+PS predictions (full line) together with the ratio between the two LO predictions (dashed line).

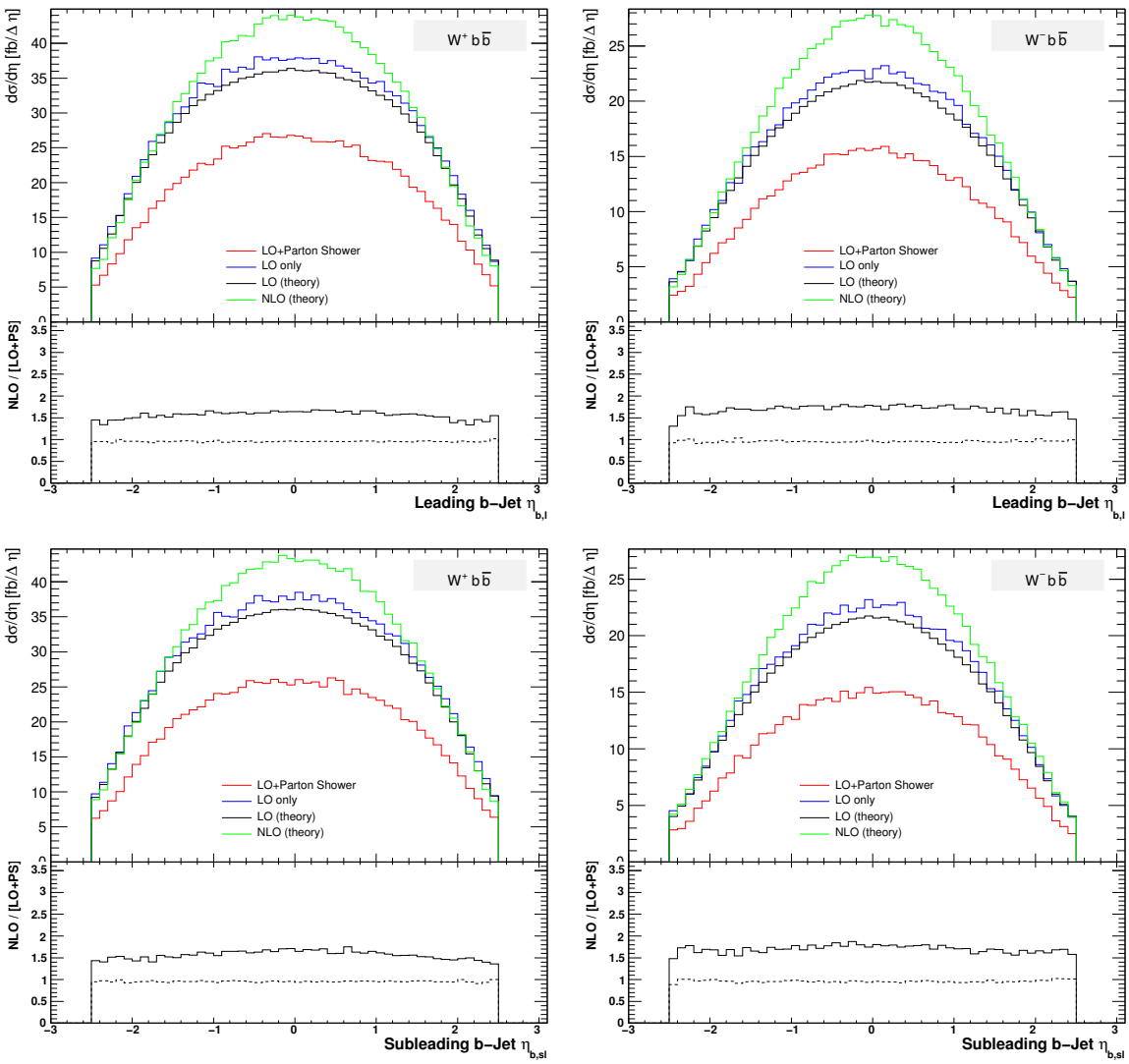


Fig. 33: Pseudorapidity distribution of the *leading* and *sub-leading* b jet, for both $W^+b\bar{b}$ and $W^-b\bar{b}$ production. The lower window shows the ratio of NLO to LO+PS predictions (full line) together with the ratio between the two LO predictions (dashed line).

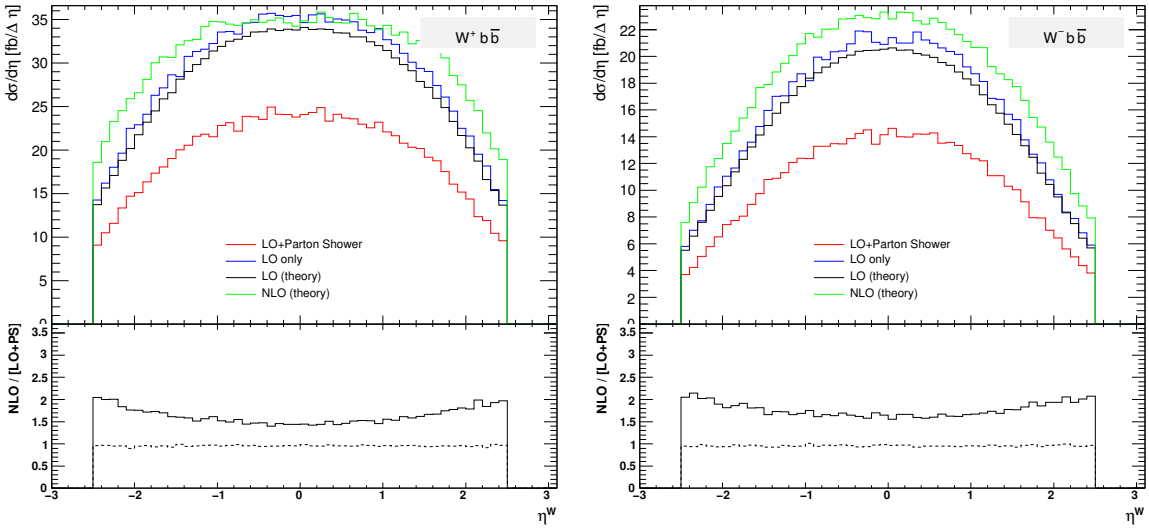


Fig. 34: Pseudorapidity distribution of the W vector boson for both $W^+b\bar{b}$ and $W^-b\bar{b}$ production. The lower window shows the ratio of NLO to LO+PS predictions (full line) together with the ratio between the two LO predictions (dashed line).

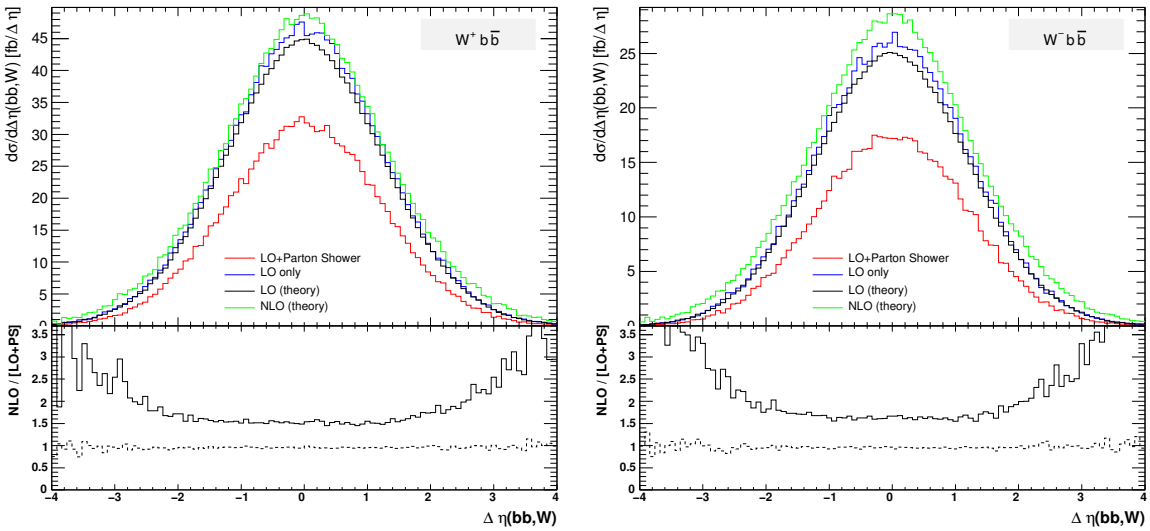


Fig. 35: Pseudorapidity difference distribution of the 2 b-jet system and W boson for both $W^+b\bar{b}$ and $W^-b\bar{b}$ production. The lower window shows the ratio of NLO to LO+PS predictions (full line) together with the ratio between the two LO predictions (dashed line).

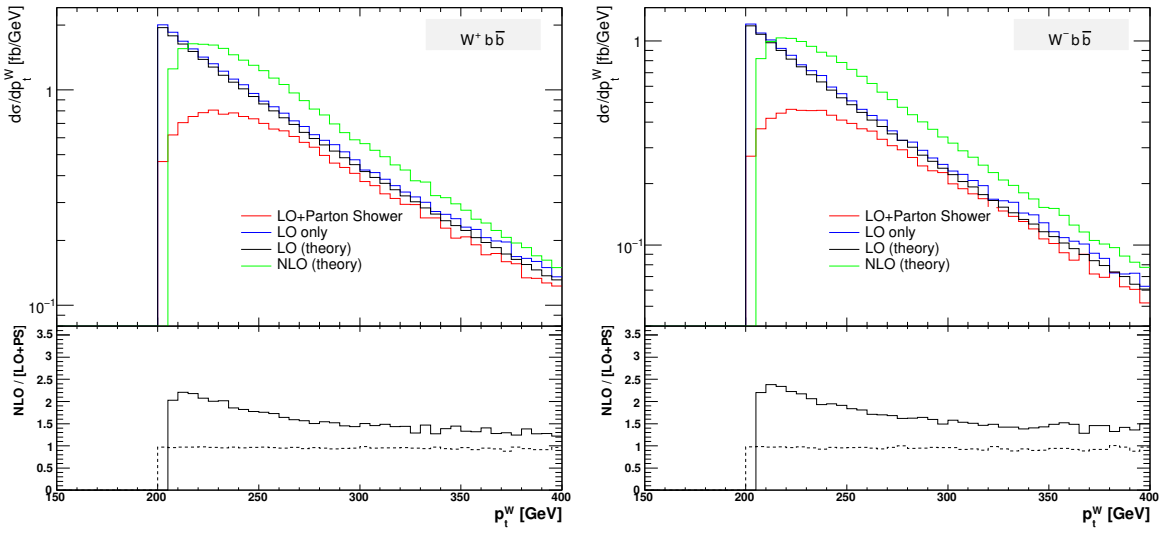


Fig. 36: Transverse momentum distribution of the W vector boson for both $W^+b\bar{b}$ and $W^-b\bar{b}$ production. The lower window shows the ratio of NLO to LO+PS predictions (full line) together with the ratio between the two LO predictions (dashed line).

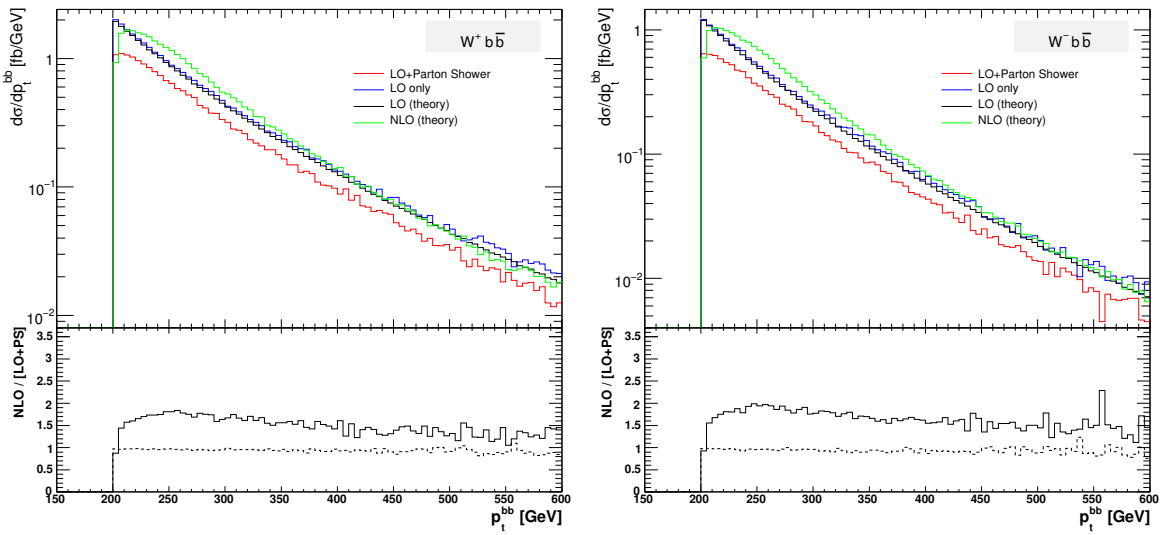


Fig. 37: Transverse momentum distribution of the 2 b-jet system for both $W^+b\bar{b}$ and $W^-b\bar{b}$ production. The lower window shows the ratio of NLO to LO+PS predictions (full line) together with the ratio between the two LO predictions (dashed line).

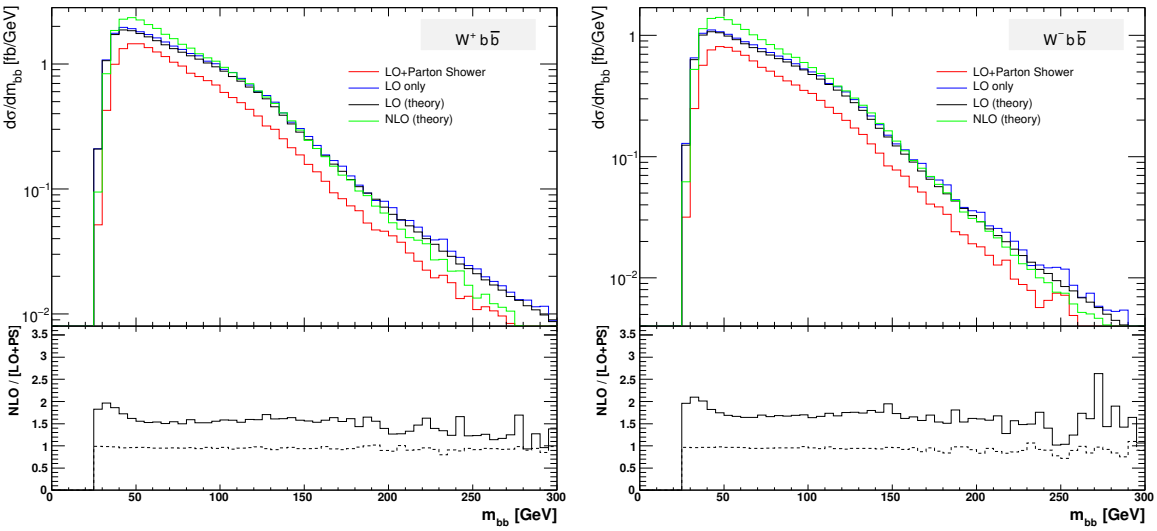


Fig. 38: Invariant mass distribution of the 2 b-jet system for both $W^+b\bar{b}$ and $W^-b\bar{b}$ production. The lower window shows the ratio of NLO to LO+PS predictions (full line), together with the ratio between the two LO predictions (dashed line).

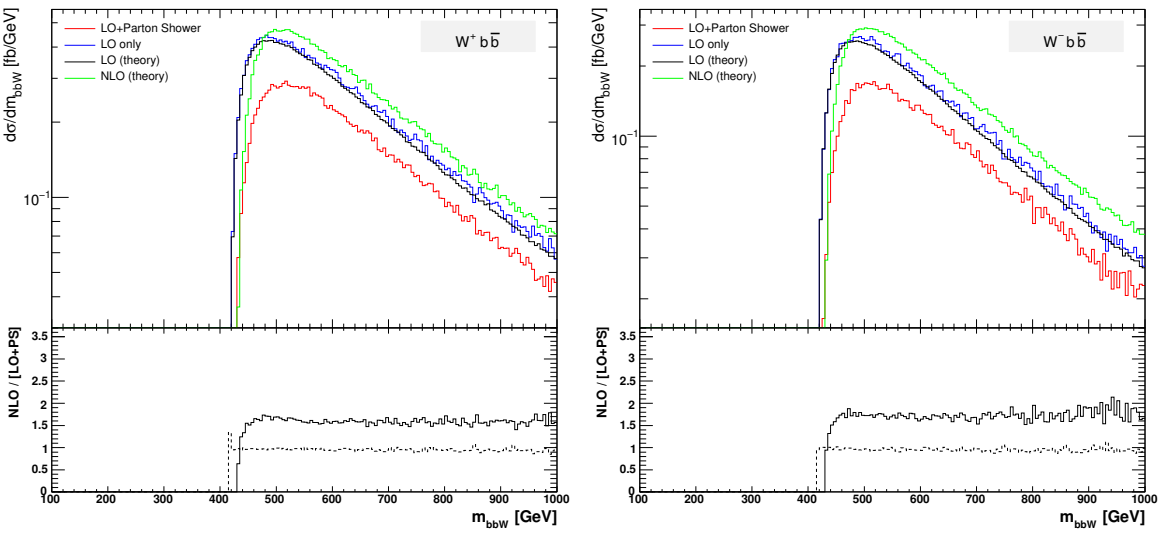


Fig. 39: Invariant mass distribution of the 2 b-jet plus W system for both $W^+b\bar{b}$ and $W^-b\bar{b}$ production. The lower window shows the ratio of NLO to LO+PS predictions (full line), together with the ratio between the two LO predictions (dashed line).

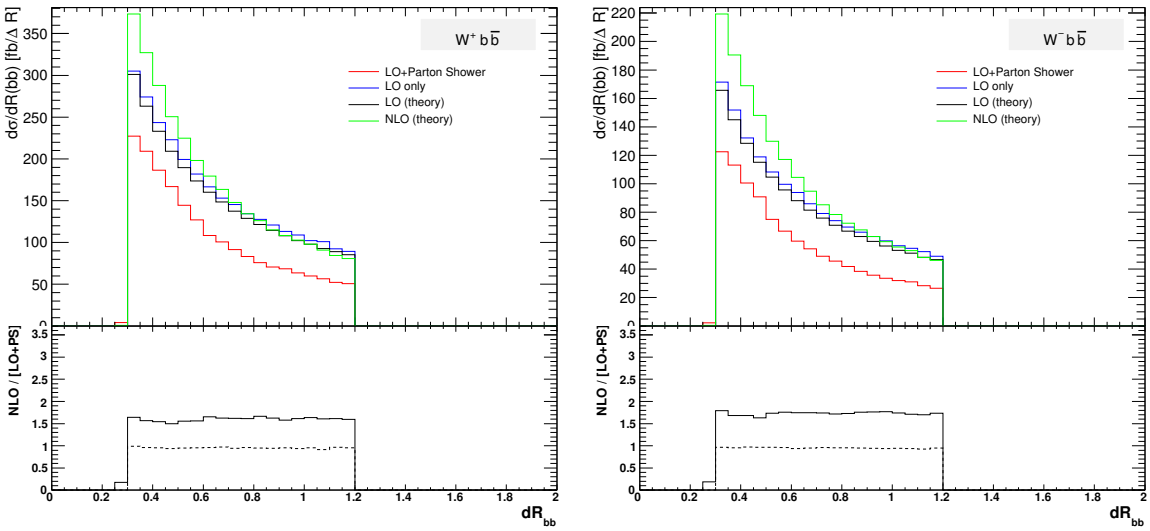


Fig. 40: Relative distance distribution of the 2 b-jet system for both $W^+b\bar{b}$ and $W^-b\bar{b}$ production. The lower window shows the ratio of NLO to LO+PS predictions (full line), together with the ratio between the two LO predictions (dashed line).

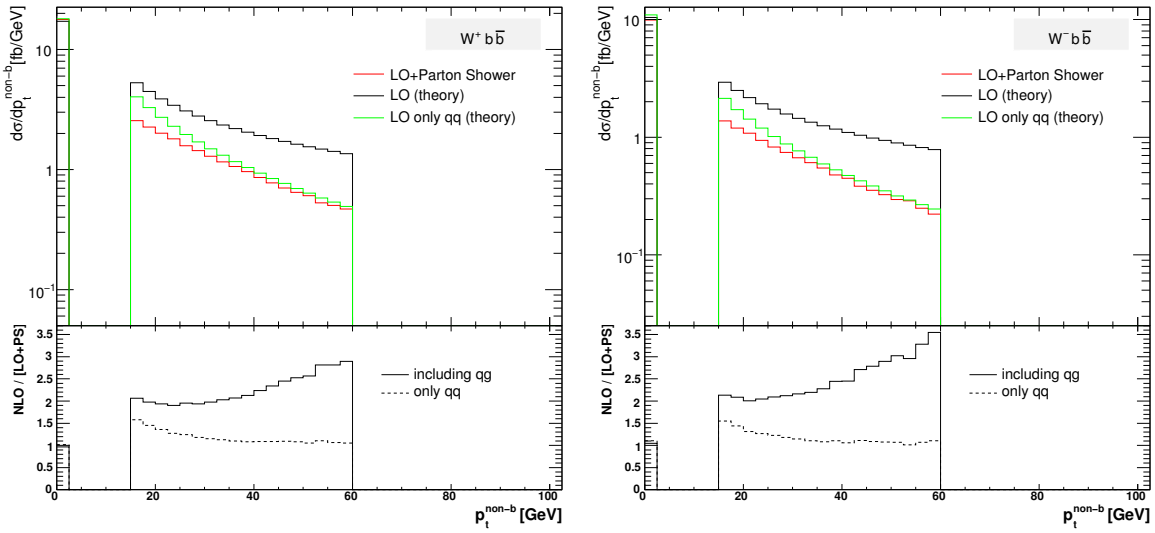


Fig. 41: Transverse momentum distribution of the non b-jet for both $W^+b\bar{b}+j$ and $W^-b\bar{b}+j$ production. The lower window shows the ratio of LO to LO+PS prediction (dashed line including only the $q\bar{q}' \to Wb\bar{b}j$ process in the LO computation, full line including also the $q(\bar{q})g \to Wb\bar{b}j$ process). Although the cross section for $Wb\bar{b}+j$ production is a LO result, it is calculated as part of the real corrections with the CTEQ6M set of PDFs and two-loop evolution of α_s.

Part IV
OBSERVABLES AND DETECTORS

13. DELPHES, A FRAMEWORK FOR FAST SIMULATION OF A GENERIC COLLIDER EXPERIMENT[24]

13.1 INTRODUCTION

Multipurpose detectors at high energy colliders are very complex systems. Their simulation is in general performed by means of the GEANT [166] package and final observables used for analyses usually require sophisticated reconstruction algorithms. This complexity is handled by large collaborations, and data and the expertise on reconstruction and simulation software are only available to their members. Precise data analyses require a full detector simulation, including transport of the primary and secondary particles through the detector material accounting for the various detector inefficiencies, the dead material, the imperfections and the geometrical details. Such a simulation is very complicated, technical and requires a large CPU power. On the other hand, phenomenological studies, looking for the observability of given signals, may require only fast but realistic estimates of the expected signal signatures and their associated backgrounds.

The DELPHES [167] framework has been designed for the fast simulation of a general-purpose collider experiment. Using this framework, observables such as cross-sections and efficiencies after event selection can be estimated for specific reactions. Starting from the output of event generators (e.g. Les Houches Event File or `HepMC`), the simulation of the detector response takes into account the subdetector resolutions, by smearing the kinematics of final-state particles.

DELPHES includes the most crucial experimental features, such as: the geometry of both central and forward detectors, the effect of magnetic field on tracks, the reconstruction of photons, leptons, jets, b-jets, τ-jets and missing transverse energy, a lepton isolation, a trigger emulation and an event display. Several common datafile formats can be used as input in DELPHES, in order to process events from many different generators.

13.2 SIMULATION OF THE DETECTOR RESPONSE

The overall layout of the multipurpose detector simulated by DELPHES consists in a central tracking system surrounded by an electromagnetic and a hadron calorimeters (ECAL and HCAL, each with a central region and two endcaps) and two forward calorimeters (FCAL, with separate electromagnetic and hadronic sections). Finally, a muon system encloses the central detector volume. In addition, possible very forward detectors are also simulated, like zero-degree calorimeters (ZDC) and very forward taggers. Even if DELPHES has been developed for the simulation of general-purpose detectors at the LHC (namely, CMS and ATLAS), it allows a flexible parametrisation for other cases, e.g. at future linear colliders.

Magnetic field, tracks and calorimetric cells

The effects of a central solenoidal magnetic field are simulated. Every stable charged particle with a transverse momentum above some threshold and lying inside the tracker has a probability to yield a reconstructed track. The calorimeters are segmented into cells in the (η, ϕ), with variable sizes. The response of calorimeters to energy deposits of incoming particles depends on their resolution and on the nature of the particles. The assumed calorimeter characteristics are not isotropic, with typically finer energy resolution and granularity in the central region [168, 169]. Their resolution is parametrised through a Gaussian smearing of the accumulated cell energy.

[24] Contributed by: S. Ovyn, X. Rouby

Electrons and photons leave their energy in the electromagnetic parts of the calorimeters (ECAL and FCAL, e.m.) only, while charged and neutral final-state hadrons interact with the hadronic parts (HCAL and FCAL, had.). Some long-living particles, such as the K_s^0 and Λ's, with lifetime $c\tau$ smaller than 10 mm are considered as stable particles by the generators although they may decay before reaching the calorimeters. The energy smearing of such particles is therefore performed using the expected fraction of the energy, determined according to their decay products, that would be deposited into the ECAL and the HCAL. Muons, neutrinos and hypothetical neutralinos are assumed not to interact with the calorimeters. At last, the muon systems provide a measurement of the true muon kinematics, with a smeared transverse momentum.

Forward detectors

Most of the recent experiments in beam colliders have additional instrumentation along the beamline. These extend the η coverage to higher values, for the detection of very forward final-state particles. In DELPHES, zero-degree calorimeters, roman pots and forward taggers have been implemented, similarly to the plans for CMS and ATLAS collaborations [168, 169].

The ZDCs allow the measurement of stable neutral particles (γ and n) coming from the interaction point (IP), with large pseudorapidities. The trajectory of the neutrals observed in the ZDCs is a straight line, while charged particles are deflected away from their acceptance window by the powerful magnets located in front of them. In their implementation in DELPHES, the ZDCs provide a measurement of the time-of-flight of the particle, from the IP to the detector location. The simulated ZDCs are composed of an electromagnetic and a hadronic sections, for the measurement of photons and neutrons, respectively. The ZDC hits do not enter in the calorimeter cell list used for reconstruction of jets and missing transverse energy. The fact that additional charged particles may enter the ZDC acceptance is neglected in the current versions of DELPHES.

Forward taggers, located very far away from the IP, are meant for the measurement of scattered particles following very closely the beam path. To be able to reach these detectors, particles must have a charge identical to the beam particles and a momentum very close to the nominal value of the beam. These taggers are near-beam detectors located a few millimetres from the true beam trajectory and this distance defines their acceptance. This fast simulation uses the HECTOR software [170], which includes the chromaticity effects and the geometrical aperture of the beamline elements of any arbitrary collider. Forward taggers are able to measure the hit positions and angles in the transverse plane at the location of the detector, as well as the time-of-flight. Out of these, the particle energy and the momentum transfer it underwent during the interaction can be reconstructed at the analysis level.

13.3 HIGH-LEVEL OBJECT RECONSTRUCTION

The results of the detector simulation are output in a file storing for each event its tracks, calorimetric cells and hits in the very forward detectors. In addition, collections of final particles ($e^{\pm}$, $\mu^{\pm}$, γ) and objects (light jets, b-jets, τ-jets, E_T^{miss}) are provided. While electrons, muons and photons are easily identified, other quantities are more difficult to evaluate as they rely on sophisticated algorithms (e.g. jets or missing energy). For most of these objects, their four-momentum and related quantities are directly accessible in DELPHES output (E, $\vec{p}$, p_T, η and ϕ).

Identification of electrons, photons and muons

Real electron ($e^{\pm}$) and photon candidates are associated to the final-state collections if they fall into the acceptance of the tracking system and have a transverse momentum above some threshold. Assuming a good measurement of the track parameters in the real experiment, the electron energy can be reasonably recovered. DELPHES assumes a perfect algorithm for clustering and bremsstrahlung recovery. Electron energy is smeared according to the resolution of the calorimetric cell where it points to, but independently

from any other deposited energy is this cell. Electrons and photons may create a candidate in the jet collection.

Generator-level muons ($\mu^{\pm}$) entering the muon system acceptance and overpassing some threshold are considered as good candidates for analyses. The application of the detector resolution on the muon momentum depends on a Gaussian smearing of the p_T. Neither η nor ϕ variables are modified beyond the calorimeters: no additional magnetic field is applied. Multiple scattering is neglected. This implies that low energy muons have in DELPHES a better resolution than in a real detector.

To improve the quality of the contents of the charged lepton collections, isolation criteria can be applied. This requires that electron or muon candidates are isolated in the detector from any other particle, within a given cone in the (η,ϕ) plane. The isolation algorithm is based on tracker data, but in addition the sum of the transverse momenta of all tracks but the lepton one within the isolation cone is provided, as well an estimate based on calorimeter data.

Jet reconstruction

A realistic analysis requires a correct treatment of partons which have hadronised. Therefore, the most widely currently used jet algorithms have been integrated into the DELPHES framework using the FASTJET tool [171]. For all jet algorithms, the calorimetric cells are used as inputs. Since several particles can leave their energy into a given calorimetric cell, which broadens the jet energy resolution, a *jet energy flow* algorithm can be switched on in DELPHES. This takes into account the measured properties of tracks pointing to the calorimetric cells of interest for the jet reconstruction.

A jet is tagged as a b-jet if its direction lies in the acceptance of the tracker and if it is associated to a parent b-quark. This identification procedure for the b-tag is based on the true generator-level identity of the most energetic parton within a cone around the (η, ϕ) region, with a radius equal to the one used to reconstruct the jet. Jets originating from τ-decays are identified using a procedure consistent with the one applied in a full detector simulation [168]. The tagging relies on the two following properties of the τ lepton. First, 77% of the τ hadronic decays contain only one charged hadron associated to a few neutrals. Secondly, the particles arisen from the τ lepton produce narrow jets in the calorimeter.

Missing transverse energy

The *true* missing transverse energy (MET), i.e. at generator-level, is calculated as the opposite of the vector sum of the transverse momenta of all visible particles. In a real experiment, calorimeters measure energy and not momentum. Any problem affecting the detector (dead channels, misalignment, noisy cells, cracks) worsens directly the measured missing transverse energy. In DELPHES, MET is based on the calorimetric cells only. Muons and neutrinos are therefore not taken into account for its evaluation. However, as muon candidates, tracks and calorimetric cells are available in the output file, the missing transverse energy can always be reprocessed a posteriori with more specialised algorithms.

Trigger emulation

Most of the usual trigger algorithms select events containing leptons, jets, and MET with an energy scale above some threshold. A trigger emulation is included in DELPHES, using a fully parametrisable trigger table, also allowing logical combinations of several conditions on the final analysis data.

13.4 ASSUMED SIMPLIFICATIONS FOR THE FAST SIMULATION

DELPHES is a fast simulation aiming at providing quickly realistic observables. It relies on several hypotheses and simplifications that allows the use of less sophisticated algorithms than in real experiments, but aims at reaching the same detection, identification and reconstruction performances. Since the framework design relies on the expectations of the real detector performances, detector geometry is

idealised: being uniform, symmetric around the beam axis, and having no cracks nor dead material. Secondary interactions, multiple scatterings, photon conversion and bremsstrahlung are also neglected.

No longitudinal segmentation is available in the simulated calorimeters. DELPHES assumes that ECAL and HCAL have the same segmentations and that the detector is symmetric with respect to the $\eta = 0$ plane and around the beam axis. A particle entering a calorimetric cell deposes all its energy, even if it enters very close to its geometrical edge. Particles other than electrons ($e^{\pm}$), photons (γ), muons ($\mu^{\pm}$), neutrinos (ν_e, ν_μ and ν_τ) and hypothetical neutralinos are simulated as hadrons for their interactions with the calorimeters. The simulation of stable particles beyond the Standard Model should therefore be handled with care.

The electron, photon and muon collections contain only real candidates. For instance, the particles which might leak out of the calorimeters into the muon system (*punch-through*) are not considered as muon candidates in DELPHES. However, fake candidates can be added into the collections at the analysis level, when processing DELPHES output data.

In current version of DELPHES, the displacement of secondary vertices is not simulated, in particular for b-jets. Extra hits coming from the beam-gas events or secondary particles hitting the beampipe in front of the forward detectors are not taken into account. At last, real triggers are intrinsically based on reconstructed data with a worse resolution than final analysis data. In DELPHES, the same data are for trigger emulation and for final analyses.

13.5 IMPROVEMENTS AND PERSPECTIVE

As a consequence of the fruitful discussions and collaboration during and after the Les Houches workshop, several improvements have been identified for the developments in DELPHES. The b-jet algorithm will be improved by adding a dependence of the tagging efficiencies with respect to the jet transverse momentum. In current versions of DELPHES, the probabilities for b-tag or mistag are uniforms, which underestimates the actual performances of the real experiments. The τ-jet collection will allow 3-prong τ's. Local detector inefficiency maps might also be implemented, as well as an improved interface to `HepMC` that would allow an easy connection to tools like Rivet [46].

CONCLUSIONS

DELPHES is a framework offering fast simulation tools allowing to investigate quickly new models and check their signatures in a realistic detection environment. It offers an intermediate step between simplified *parton-level* analysis and extensive analyses using the full-simulation power in large collaborations.

ACKNOWLEDGEMENTS

X Rouby would like to thank S Schumann, F Moortgat and F Maltoni for their support in the participation to the Les Houches workshop. S Ovyn and X Rouby are happy to thank the numerous Les Houches participants who have shown their interest in DELPHES, leading to very interesting and fruitful discussions and exchanges.

14. EFFECT OF QED FSR ON MEASUREMENTS OF $Z/\gamma*$ AND W LEPTONIC FINAL STATES AT HADRON COLLIDERS [25]

14.1 Introduction

Due to its simple quantum structure and clean experimental signature, quark-antiquark annihilation to lepton pairs (Drell-Yan), is one of the most important channels in hadron collider physics. In this paper we shall discuss some requirements that must be met, at the observable level, when using information obtained from the final-state leptons in this process. The emphasis on "observable level" is particularly important; if information about states that *are not themselves directly observable*, such as the Z boson or a "bare" lepton in a Monte Carlo event record, is used for correcting or calibrating the raw measurement, then the precision of the corrected/calibrated result will suffer from an intrinsic ambiguity which was not present in the raw measurement, i.e., the result will have been degraded.

We discuss the cause and magnitude of this degradation and argue that the experimental corrections and calibrations can equally well be performed *without* using such information, hence preserving the full precision of the raw measurement in the calibrated result.

To illustrate our conclusions and to aid future studies, we examine various possible observable definitions and present a collection of reference comparisons for the Tevatron processes $p\bar{p} \to Z/\gamma^* \to e^+e^-$ and $p\bar{p} \to Z/\gamma^* \to \mu^+\mu^-$. These same concerns apply to the equally important W signal, which we also consider.

14.2 Monte Carlo Truth

Calculations of collider physics processes rely heavily on factorizations of the full transition amplitudes (squared) into smaller, more manageable, pieces. These factorizations are typically formally correct in limits in which one single Feynman diagram dominates over all others, in which case the system has a well-defined "semi-classical" history, represented by the single dominant diagram. The individual pieces can then be treated as approximately independent since the quantum mechanical interference between them can be neglected in this limit.

Essentially, what event generators, such as Herwig, Pythia, and Sherpa, provide as intermediate particles in their event records ("Monte Carlo truth") are representations of such semi-classical histories. At the very least, these histories show how the generator approximated the process so far (which may, in turn, furnish important boundary conditions on subsequent generation steps). But even in the best case scenario, when the diagram represented by the event record really is the dominant one, it is still only exact when all other diagrams are zero, i.e., in the semi-classical limit. Therefore, any procedure (such as a correction applied to an experimental measurement) that depends on history-information is only well-defined in that limit. At the quantum level, it would suffer from an irremovable ambiguity which would set an ultimate limit to the precision that could be obtained with that procedure.

Let us consider the Drell-Yan process in more detail. For this process, e.g., the event record of the Pythia generator will contain an intermediate Z boson. This is intended to tell the user that the lepton pair comes from an s-channel diagram (which in fact includes the full Z/γ^* interference), and it also lets Pythia's parton shower know that the invariant mass of the lepton pair should be preserved in the shower. So far so good. If QED radiation was not an issue, this intermediate Z boson is identical to the sum of the final-state lepton pair, and it is therefore tempting to use this boson, even in the presence of QED effects, to define some kind of Monte-Carlo-truth / generator-level / QED-corrected / ... "Z/γ^*". However, when including QED effects, diagrams with photons emitted in the initial state will interfere with diagrams that have them emitted in the final state. Therefore, we cannot, at the amplitude-squared level, determine whether a given photon was initial- or final-state radiation, not even statistically. However, the two cases correspond to *different* semi-classical histories, with different "Z/γ^*" kinematics, and therefore any

[25] Contributed by: A. Buckley, G. Hesketh, F. Siegert, P. Skands, M. Vesterinen, T.R. Wyatt

semi-classical "Z/γ^*" definition would at least be ambiguous up to the size of the initial-final interference terms.

Another manifestation of the same basic problem is to do with the identification of which leptons came from the Z/γ^* decay, in an event with multiple leptons. Again, a correction procedure that would depend on Monte Carlo truth for this identification would be fundamentally ambiguous at least up to interference terms between diagrams with different assignments.

Finally, there is the issue of how to *define* a lepton. Especially electrons, which are close to massless, emit large amounts of near-collinear photon bremsstrahlung. The final-state electron in a Monte Carlo event record is therefore not a uniquely defined object (it depends, for instance, on if/how electron mass effects are implemented in that Monte Carlo). As we argued above, the electron "before" QED radiation is also not well-defined, since the electron interferes with other particles. And in any case it would be a mere coincidence if either object bears much resemblance to what might be called an electron in an experimental setting (to be discussed further below). It is therefore also crucial to operate with an unambiguous and collinear safe "lepton definition".

What we shall argue below is that such ambiguities are in fact both present and large for precision measurements of current interest. It is therefore important to reiterate that the full experimental precision is only preserved if the observables and correction procedures are defined entirely in terms of the final-state leptons, with these in turn defined in a way that is collinear safe against photon emission.

14.3 Definitions

In previous measurements of the kinematics of the Drell-Yan process, the reconstructed data have typically been corrected to the "Z/γ^*" boson level. This correction process can be broken into three steps:

1. Correcting the measurement leptons to the level of particles entering the detector ("unfolding" the detector resolution and efficiency).
2. Correcting from the particles entering the detector to the "Z/γ^*".
3. Correcting from the measured phase space to 4π acceptance, with no selection cuts except on the "Z/γ^*" mass.

As outlined in the previous Section, Step 2 introduces ambiguities and dependencies on the model used to simulate Drell-Yan and FSR. Step 3 involves correcting for particles that were unmeasured, so explicitly depend on a model of those particles. We therefore propose that **all measurements be made available after Step 1: corrected only for detector resolution and efficiency**. Steps 2 and 3 may of course be performed in addition, subject to problems discussed above.

In order to perform Step 1, we must define the "electron" and "muon" observables in terms of the particles that enter the detector, accounting for the different detector response in each case. We then consider how to form the "Z/γ^*" from those electrons and muons, and briefly discuss missing energy and a "W" final state. We therefore propose these definitions:

- In simulated events, we consider all generator level particles with $c\tau > 10$ mm as "stable" [1] (i.e. could possibly be detected). Only the stable particles may be used to define an observable, and no checks should be placed on the internal generator origin of any particles (no history dependence).
- Electrons: to mimic the response of a calorimeter, electrons and photons must be combined into an localized "EM cluster", and such clusters considered as particle level electrons. Here, this clustering is done with simple cones of $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$=0.2 around any electron[26]. Thus any narrow angle QED FSR is combined back into the electron, which may also pick up some additional electromagnetic energy from the underlying event. Any wider angle FSR is "lost": i.e. not associated with the electron.

[26]This simple cone clustering is sufficient to reproduce observable electrons at the Tevatron experiments. For the LHC experiments, a different cone size or more advanced clustering algorithm (e.g. anti-k_T) may better reproduce the detector level observable. Ideally, the same particle level electron definition can be agreed upon for all LHC experiments.

- Muons: the stable muons (i.e. after FSR) can be considered directly, as would be measured in the tracking system of a detector. Thus, *all* FSR is lost, and the underlying event has no effect.
- Missing transverse energy (MET): in data this is typically calculated by inverting the vector sum of calorimeter E_T and muon p_T, introducing many detector acceptance and efficiency effects which are difficult to reproduce at particle level. To initiate discussion, we define the MET simply from the vector sum of all neutrinos in the event, which is compared to the p_T of only the the W neutrino in Fig. 42.
- Z/γ^*: selections should follow the data analysis. Typically, this will mean considering all electrons or muons within a given $|\eta|$ range, then forming opposite charge pairs that lie within the required di-lepton mass window (e.g. 65-115 GeV). In the (rare) cases where two possible pairs pass these selections, the method used to select the "best" should again follow the data analysis, such as the pair closest to the Z mass.
- W: again, selections should follow data analysis. This will typically mean combining the MET with the highest p_T lepton, then placing some requirement on transverse mass.

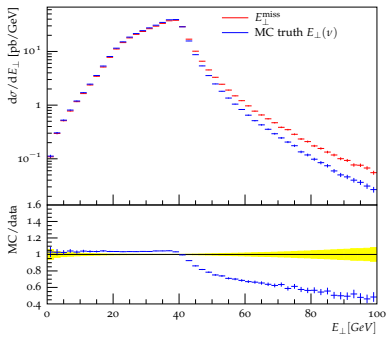


Fig. 42: Comparing the p_T of the neutrino from a W decay to the particle level MET (defined in the text).

When defining the Z/γ^* and W observables, three further selections are typically applied in data analysis that have a less clear particle level analogue. Here we provide suggestions as the basis for future discussion:

1. Lepton isolation: reproducing detector level isolation requirements at particle level is difficult, whereas determining the efficiency loss due to such requirements in data is generally rather straightforward. So we suggest the measured data should be corrected for these efficiency losses, and no isolation conditions be imposed at particle level.
2. Minimum lepton p_T (and MET) requirements: such requirements should generally be implemented at the particle level also. However, if the effect is small it may be possible to use the simulation to extrapolate into the unmeasured region.
3. Event vetoes: for example, a second lepton veto in W selection. Due to detector efficiency effects, such requirements are difficult to implement in a way that is consistent between detector and particle level, and should be treated with care.

14.4 Implications

To study the impact of the definitions outlined in the previous section, some Tevatron examples are used. Samples of $p\bar{p} \to Z/\gamma^*$, with $Z/\gamma^* \to e^+e^-$ and $Z/\gamma^* \to \mu^+ß\mu^-$, are produced using Pythia 6.421 [18], using the $Perugia$ 6 tune [62] with the CTEQ6L1 PDFs [156]. The samples are normalized to the same number of generated events. For comparison, we also consider a "generated Z/γ^*", reconstructed by searching the Pythia event record and explicitly taking the two leptons from the Z/γ^* decay, before FSR. Regardless of the definition used, the leptons used to make a Z/γ^* are required to have $|\eta| < 1.7$ and a dilepton mass between 65 and 115 GeV.

14.5 FSR Properties

First, the properties of photons from FSR are considered. Figure 43 shows, for $Z/\gamma^* \to e^+e^-$ and $Z/\gamma^* \to \mu^+ß\mu^-$, the rates at which FSR photons are emitted from the Z/γ^* decay products. Next, ΔR between FSR photons and the nearest Z/γ^* decay product. Then, the FSR photon energy and the energy of "unmeasured photons" is shown; in the case of muons, all photons are considered unmeasured; for electrons, photons outside the cone of 0.2 are considered unmeasured. Finally, the fraction of all photons within $|\eta| < 2.5$ (typical electromagnetic calorimeter coverage) in the final state which arise from FSR for different p_T requirements.

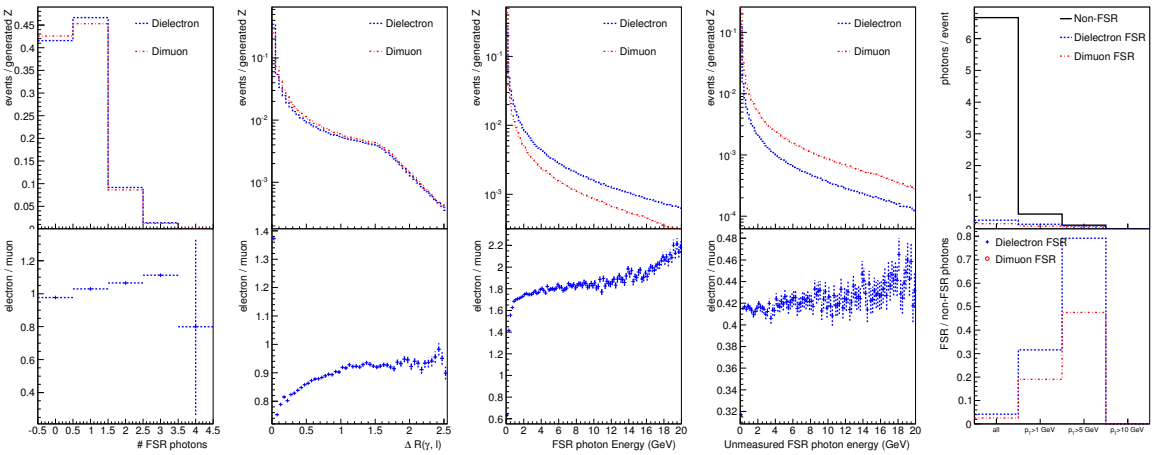


Fig. 43: Left: the rate of FSR photons from the Z decay products; centre left: $\Delta R(\gamma, l)$; centre: FSR photon energy; centre right: energy of "unmeasured" FSR photons (see text); right: the fraction of all photons that arise from FSR, at different energies.

It can be seen that the rate of FSR is significant. In $Z/\gamma^* \to e^+e^-$, the photons are typically close to the electron and will be added to the electron cluster, however (using the default Pythia settings) approximately 23% of $Z/\gamma^* \to e^+e^-$ events will still contain at least one lost FSR photon (compared to 66% of $Z/\gamma^* \to \mu^+ß\mu^-$ events). These photons are typically of very low energy, though approximately 3% of $Z/\gamma^* \to e^+e^-$ and 7% of $Z/\gamma^* \to \mu^+ß\mu^-$ events contain an unmeasured FSR photon above 5 GeV. Identifying these lost FSR photons is almost certainly experimentally impossible, and it also appears impossible at the level of stable particles (without looking into the generator event record), due to the presence of photons from other sources (primarily π^0 decays). In the following sections, we consider how these effects result in differences between the generated $Z/\gamma^*$and the measured dilepton final state, for some important kinematic properties.

14.6 Energy Scale and Resolution, and Total Cross Section

The energy scale and resolution for both the electromagnetic calorimeter and tracking system are typically derived by fitting the shape and position of the Z/γ^* peak. Here, we assess the impact of QED FSR on the measured Z/γ^* line-shape, and hence possible effects on detector calibration. Figure 44 shows the generated Z/γ^* mass distribution, and that reconstructed using the stated electron and muon definitions.

It can be seen that FSR has a drastic effect on the observable line-shape below the pole. Obviously calibrations should ensure the detector response reflects the observable leptons, and not "correct" them to the generated Z/γ^*, so using the lower side of the mass peak would clearly rely upon a model of FSR. This introduces model dependence which may complicate the extraction of model independent quantities from data analysis. However, the energy scale may possibly be determined with minimal model dependence by fitting only the peak position, which is mostly unaffected by FSR. Similarly, it may be

possible to determine lepton resolution in a model independent way from the higher edge of the mass peak, which is also mostly unaffected.

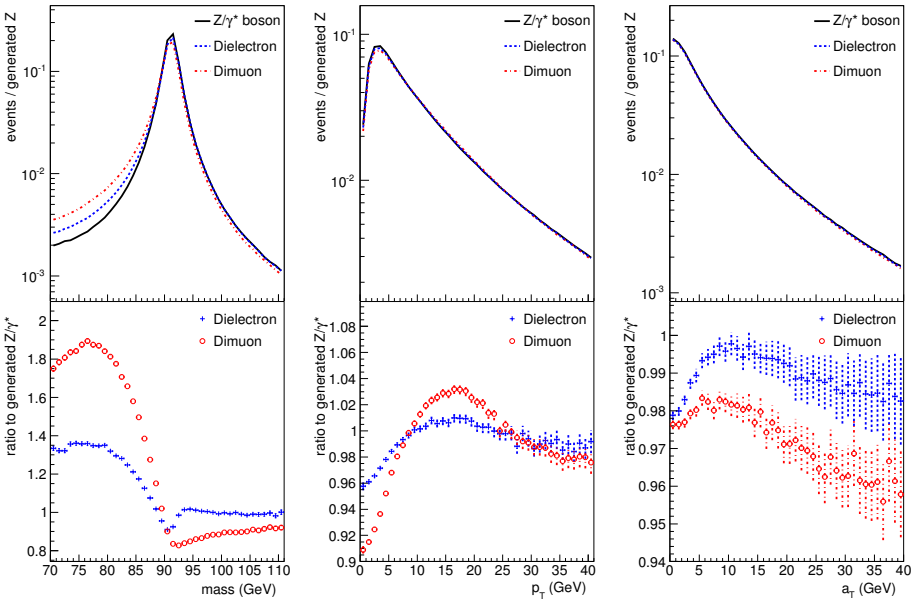


Fig. 44: Comparing the generated Z/γ^* to the observable (defined in the text). Left: the Z/γ^* mass; centre: the Z/γ^* p_T; right: the Z/γ^* a_T.

The effect of FSR on the Z/γ^* line-shape has another implication: the inclusive Drell-Yan cross section is typically measured for a given mass range, for example 65-115 GeV. Energy lost in FSR causes the dilepton mass to be below the "true" Z/γ^* mass, which means events above the mass window may migrate in, and events in the window migrate out. In the Pythia samples considered, this causes a 0.9% net loss of dielectron events, and a 2.1% net loss of dimuon events for the stated mass window.

14.7 Z/γ^* p_T and a_T

We next consider the impact on the Z/γ^* p_T and a_T [172, 173] distributions, important variables in measuring non-perturbative QCD form factors and in generator tuning. Figure 44 shows the observable Z/γ^* p_T and a_T compared to the true Z/γ^* p_T and a_T. For p_T, effects of up to 4% are seen in the dielectron channel, and up to 10% in the dimuon channel. Effects are around 2% for a_T, where the shape effect is larger in the electron channel. For both p_T and a_T, the effects are largest at low values, the crucial region for generator tuning and determining form factors.

14.8 Recovering Published Measurements?

Several Tevatron measurements of the Z/γ^* p_T have corrected back to the "generated" Z/γ^*, with full 4π acceptance [174–176]. Similarly, it is common practice in theory papers to present results at the same level (see a recent example in Ref. [177]). We perform a quick study to determine if such measurements and calculations can be approximated using only the stable final state particles, rather than having to resort to looking into the generator event history. This is done by using a wider cone around the final state electrons, to attempt to re-sum more FSR and get closer to the leptons directly from the Z/γ^* decay. Figure 45 shows the effect of using a cone of radius 0.2, 0.5 and 1.0. While the cone size of 0.5 comes closer to the "true" Z/γ^* p_T in the low p_T region, it moves further away at high p_T. Increasing the cone size to 1.0 clearly distorts the spectrum, with too many non-FSR photons being added to each electron. Similar behaviour is observed for a_T.

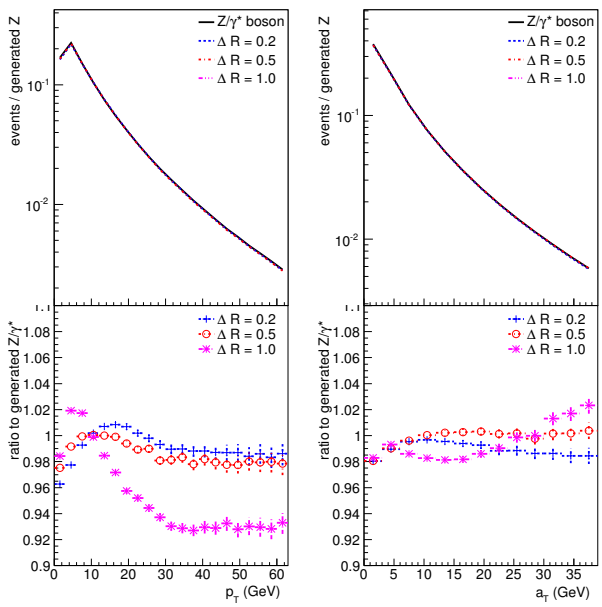


Fig. 45: The effect on the Z/γ^* p_T (left) and a_T (right) of changing the electron cluster size.

This simple study shows the difficulty in reproducing the generated Z/γ^* from only the stable particles, and the precision to which such "boson level" results can for example be implemented in Rivet (see Section 5. and used for generator tuning with Professor [47].

Fully reproducing published Z/γ^* p_T results may only be possible by mimicking the data treatment: applying a "FSR correction factor" to the spectrum reconstructed from the stable particles. Such treatment introduces similar ambiguities and model dependence as the original data analysis which corrected back to the "Z/γ^*".

14.9 Combining Electron and Muon Channels

Experimentally, combining electron and muon channel measurements is considered a simple way to increase the statistics used in a measurement, and benefit from the fact that many experimental systematics are uncorrelated between the two channels. However, we have shown that the electron and muon observables are fundamentally different in both the measured cross section and reconstructed kinematics are different in each channel. Further complications are introduced by the different acceptance of calorimeters and muon detectors. These factors degrade the value of a combination, for which the same observable must be extracted from each channel, and it is impossible to do this without relying on a model of FSR.

If a combination must be done, a "minimally model dependent" method would be:

1. Restrict electron and muon measurements to the same phase space. This will typically mean excluding some region from one of the channels, due to different coverage of muon detectors and calorimetry.
2. Apply a theoretical factor to the muon channel measurement, to correct from the observable muons to an equivalent of the electron observable: in this study, stable muons + all photons and electrons in a 0.2 cone around each muon. This factor should be published along with the combination.

In simulated events, there are then two options: apply the same theoretical factor to simulated $Z/\gamma^* \to \mu^+ß\mu^-$ as used for the muon data, then combine; or, directly treat stable muons as electrons (clustering EM energy around each muon). Figure 46 compares the Z/γ^* p_T and a_T for the muon channel calculated in this direct way to the electron channel, and good agreement is seen.

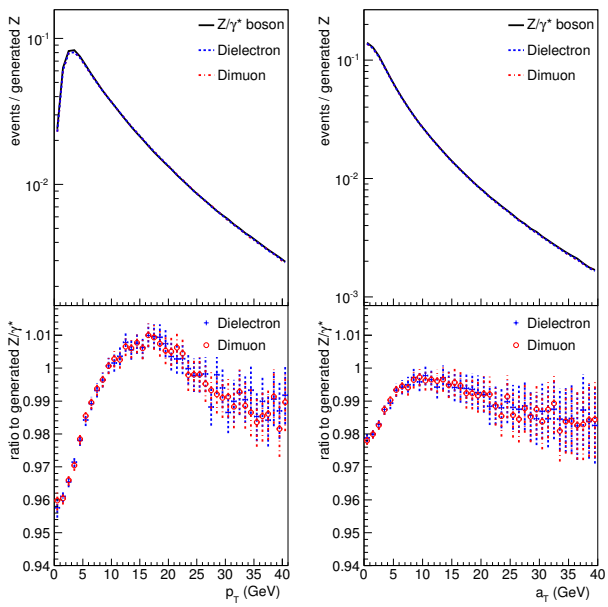


Fig. 46: The Z/γ^* p_T (left) and a_T (right) using the same treatment of electron and muon final states.

14.10 Conclusion

We advocate that all future experimental measurements and theoretical predictions of the Z/γ^* and W (and other leptonic final states) should be presented in terms of observables, and not extrapolated beyond the measured acceptance. To this end, we propose definitions of observable electrons, muons and the Z/γ^*, as well as missing transverse energy and the W boson, all at the level of particles entering a detector. These definitions mimic the quantities measured in a detector, and provide an unambiguous, model independent basis to compare data and theory.

While it is essential to provide the data in terms of observables, it may still be desirable to derive further theoretical corrections for FSR and acceptance, for comparisons to previous results. We recommend such correction factors be provided in a table, rather than being applied to the data. Using this table, (the inverse of) such corrections could also be applied to a calculation which does not include FSR or acceptance requirements, to allow direct comparisons to the data while maintaining the separation of measurement and theory.

Part V
JETS AND JET SUBSTRUCTURE

15. STATUS OF JET SUBSTRUCTURE STUDIES [27]

15.1 INTRODUCTION

In the 2007 Les Houches workshop [1] there was quite some discussion on how to define jets in ways which allow for clear comparison between experiments, and between measurements and state-of-the-art theory. The importance of such discussion has been emphasised since both by the adoption of new jet finders by ATLAS and CMS for their first data (in particular the anti-$k_\perp$ algorithm [178]), and by the rapid expansion of activity aimed at exploiting the internal properties of jets in studies and searches for new physics at the LHC.

Jet substructure has been studied at previous experiments as a means of investigating QCD. In fact jet shapes and subjet multiplicities have been used to measure the strong coupling α_s [179, 180]. What leads to a huge increase in interest at the LHC is the same fact that makes the LHC such an exciting project in general - the energy reach is substantially higher than the electroweak symmetry-breaking scale. One consequence of this which has been well appreciated for some time is that large multiplicities of electroweak-scale objects may be produced (e.g. W,Z, jets with $p_T \approx \mathcal{O}(100)$ GeV). This has been a spur to a large amount of theoretical activity on higher order calculations and parton-shower-matched Monte Carlo simulations which has been a feature of many workshops, including this one.

A consequence of the high energy of the LHC which was more slowly appreciated, but which is now quite widely understood, is that objects with masses around the electroweak scale may be produced well above threshold, and thus be highly boosted. When boosted objects decay to hadrons, the decay products will be collimated and may emerge as a single jet with distinctive substructure, even if the decay is initially to two or more quarks or gluons. It is easy to understand how such a situation will complicate the mapping from jets to partons which is the basis of almost every phenomenological study involving jets. In addition to the merging of jets due to high boosts, at the LHC, there will be unprecedented contamination from the underlying event and pileup, further deforming the naive jet-to-parton map. The basic goal of much of the work in jet substructure has been to improve the mapping from jets to partons, so that it can remain useful at the LHC, in light of new classes of signatures (like boosted object decays) and the extra contamination.

Jet substructure techniques were used first in studies of diboson production for identifying the hadronic decays of vector bosons in diboson and SUSY production [181–183], and in the 2007 workshop there was much discussion on boosted-top reconstruction, for example [184]. In the last two years, techniques have been developed and applied to several channels including top production [185–187] as well as Higgs [158, 159, 188, 189] and neutralino [190, 191] searches, and several proposals have been made to use substructure information to improve jet mass resolution in general [192–194] using substructure. Many of these techniques have been investigated by the experimental collaborations, using realistic detector simulations, and survived the test.

Another feature of the LHC relevant to jet substructure is that the calorimeters and trackers are superior to those at previous machines, allowing access to more detailed information about the jet itself. This information should be utilized in a more refined substructure analysis, but is only starting to be exploited. As the physical information from particle flow or topo-clusters becomes better understood, the power of jet substructure analyses will certainly improve. It is not surprising that a next-generation machine requires next-generation analysis tools, and the work on jet-substructure has just begun to probe the possibilities. Jet substructure is a rapidly moving field, with at least two dedicated workshops held since the meetings in Les Houches [195] and more planned. Here we give a snapshot of the current status, focused (but not exclusively) on some developments which took place in this workshop.

[27] Contributed by: J. M. Butterworth, M. D. Schwartz

15.2 GENERAL FEATURES

In QCD, the evolution from a hard parton to a jet of partons takes place in a regime where the energy scale is high enough to use perturbation theory, x is not very small, and collinear logarithms are large. This is essentially the kinematic region where DGLAP evolution should apply. This is largely understood within QCD, and can be calculated. It forms the basis of the parton-shower models implemented in the most widely used MC generators. Since the phase space between the jet p_T and the confinement scale is large, there is plenty of radiation, and parton multiplicities will be large. Confinement enters the picture at the hadronization (non-perturbative) stage, and is expected to have a small effect (at the sub-GeV level). In general, the aspects of QCD relevant for jet substructure seem to be well-modelled by tuned Monte Carlo simulations, as is borne out by the available data [179, 196–200].

There are two somewhat distinct goals of the recently developed techniques. One is to improve the single jet mass resolution, and the other is to distinguish jets originating in heavy object decays from those coming from pure QCD backgrounds. The first of these goals requires a careful definition, in line with the discussion of jets in the last workshop [1]. Jet mass must be defined in terms of infra-red and collinear safe variables on the one hand, and in terms of physical observables on the other, to avoid making completely model-dependent "measurements". However, the jet mass resolution may be optimised using models, in particular models which implement or mimic jets from the decay of a colour-singlet, by switching off initial state radiation; and models which mimic jets from parton (rather than proton) collisions by switching off underlying event. The second goal, background suppression for heavy particle decays, generally relies on the fact that in QCD the various splittings within the jet are expected to be strongly ordered in scale, with large asymmetric splittings, whereas partons resulting from a heavy particle decay will in general share the energy equally in the particle rest frame, leading to more symmetric configurations.

15.3 FILTERING

Let us begin with a discussion of what shall refer to generically as *filtering*. Filtering refers to the removal of components of a jet in an attempt to clean up the jets in a certain way. Variations on this theme have appeared, such as *pruning* [192] and *trimming* [194]. The basic idea of filtering is that one first finds a jet which is assumed to contain the radiation one is interested in as well as contaminating radiation, then one removes the contamination with the filtering step. The difference in the algorithms is more a difference in how the authors are imagining their routines used than in the way the routines themselves work. So we will organize the discussion of filtering by application. Some approaches are application-specific, optimized to find a particular signal, such as boosted Higgs or boosted top decays, while other applications are more general. We will be begin with the specific ones.

15.31 FILTERING FOR SPECIFIC APPLICATIONS

The most productive application of filtering so far has been in the search for heavy boosted objects which decay to standard model jets. Consider first a Higgs boson decaying to $b\bar{b}$ in associated production with a W. This channel is the discovery channel for a light (~ 120 GeV) Higgs at the Tevatron, but is very difficult at the LHC due to much more complicated backgrounds, in particular $t\bar{t}$ production. Going to a regime where the Higgs is boosted ($p_T > 200$ GeV) has the effect of essentially eliminating the $t\bar{t}$ background and removing a lot of the W+jets background as well. In this case, the b-jets from the Higgs decay may be within $\Delta R = 1.0$ of each other, so finding separate $R = 0.7$ jets may fail. The substructure approach pioneered in [158] was to find the jets together as a fat jet, with a cone size of $R = 1.0$ or greater, then to parse the jets to find the b-quarks as subjets.

To be specific, in [158], the authors use the Cambridge/Aachen algorithm [201, 202] to construct the fat jet. This algorithm merges constituents based on the distance $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ between them. Then they undo the clustering steps using the clustering history (conveniently stored by FastJet [203]

within the jet object). Somewhere in the declustering the b should be separated. However, the way Cambridge/Aachen works does not guarantee that the the final merger is of the two b-jets into the fat jet. Therefore, when tracing back through the declustering, one needs a different criterion, for example, looking for large mass drop – the masses of the subjets should be much smaller than the mass of the fat jet. Once a large mass drop is found, the process is stopped. All constituents of the jet which would have been split at this stage are retained, and the clustering algorithm is rerun, using the angle between the centres of the last two subjets to set a (smaller) angular scale. This effectively optimizes on an event-by-event basis the selection of final state radiation from the (color singlet) Higgs candidate, while rejecting radiation from elsewhere, including the underlying event. This is the step described as "filtering" in the analysis. The top two or three subjets are combined to reconstruct the Higgs mass. The possible inclusion of a third jet, rather than just two (representing the b and $\bar{b}$) allows for an extra hard gluon emission. This final step is critical in getting the algorithm to perform well at the Higgs mass reconstruction. The result was confirmed by an ATLAS study [159], and revived $H \to b\bar{b}$ as a discovery channel for the Higgs at the LHC.

Another application of filtering has been in hadronic boosted top decays [185]. Here, the same basic principle is applied: a fat jet is found, subjets are located through declustering, and a filter is applied to remove unwanted radiation. In this case, the three or four hardest subjets are filtered out. Three of the jets should represent the hadronic top decay products: the b-jet and the two quark jets from the W decay, the fourth may be an additional hard gluon. Using this approach, the authors found 99% background rejection efficiency at 40% signal efficiency, *per* jet. Therefore signal/background for boosted hadronic $t\bar{t}$ events over dijets is enhanced by a factor of 1000. A CMS study confirmed these efficiencies with full detector simulation [204].

Since the boosted Higgs and boosted top taggers described above rely on reversing the clustering sequence of a jet, there is naturally a dependence on the jet algorithm. We show in Fig. 47 a comparison of different algorithms for a variable used by the CMS top-tagging study [204]. Of the 3 or 4 subjets filtered out of the top jet, they looked at the minimum invariant mass of any pair. In the first panel, clustering with Cambridge/Aachen shows the peak at ~ 80 GeV corresponding to the W-mass in the signal, which is absent in the background. The second peak around ~ 20 GeV corresponds to wrongly identified subjets and mimics the QCD background. The result for the $k_\perp$ algorithm is similar. The anti-$k_\perp$ algorithm, however, does not show the W-mass peak. This is because anti-$k_\perp$ essentially works backwards to the other clustering algorithms, with the hardest particles being clustered first, and the soft stuff clustered at the end. One can see that anti-$k_\perp$ is not ideal for filtering and substructure analysis. However, the final panel shows that if the fat jet is found using anti-$k_\perp$, but the particles in the jet are reclustered using Cambridge/Aachen, the substructure is still there. We conclude that the substructure analysis and the fat-jet finder can be chosen independently.

15.32 GENERAL FILTERING TECHNIQUES

Having seen filtering successfully applied in a number of specific cases, it is natural to ask whether the approach can be generalized to cases when you do not know *a priori* what it is you are looking for. For example, in the Higgs$\to b\bar{b}$ study, the fat jet was filtered to look for 2 or 3 hard subjets, in top-tagging, the fat jet was filtered to look for 3 or 4 hard subjets. The idea behind jet *pruning* is that is is possible to filter the jet without knowing ahead of time how many hard constituents to expect. Instead, one can decompose the jet until some measure of the size of each splitting is saturated. For example, in [192], the pruning algorithm based on Cambridge/Aachen declustering stops when a splitting has $\Delta_{ij} > D_{cut}$, where $D_{\rm cut} \propto 2m_J/p_T$ is determined on a jet by jet basis. The authors demonstrated that this general method successfully finds the subjets in the specific cases of boosted top or boosted W jets.

Another general approach is jet *trimming* [194]. Here the goal is not to find hard subjet constituents in fat jets, but to clean up simple quark or gluon-initiated QCD jets. Suppose we have a massive color singlet ϕ decaying to $q\bar{q}$. Then the mass of the resonance should be reconstructable using all of the final-state

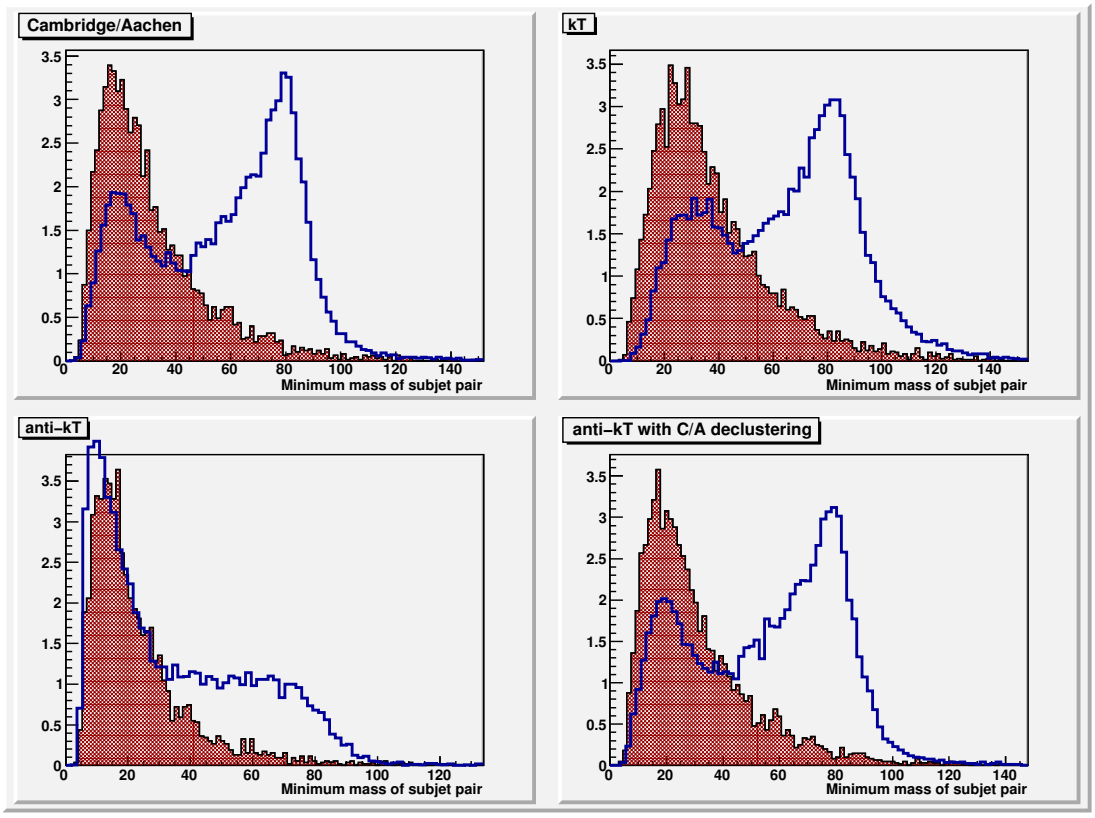


Fig. 47: Comparison of different clustering algorithms for the minimum invariant mass of two subjets. The Cambridge/Aachen and $k_\perp$ algorithms both work well and finding the subjets corresponding to the W decay, while anti-$k_\perp$ fails. The bottom right panel shows that anti-$k_\perp$ can still be used to find the fat jets, if the jet constituents are then reclustered with Cambridge/Aachen to find the subjets.

radiation from the quarks. However, in reality, it is impossible to separate the final state radiation from additional radiation originating either from the underlying event or from other hard objects that happen to also be present. By studying the radiation pattern from FSR only to FSR plus contamination, the authors observed that the contamination could be reduced with trimming.

The trimming procedure they proposed works as follows. First jets are found as usual, say with anti-$k_\perp$ with $R = 0.8$. Then the jet constituents are reclustered into many tiny jets using a much smaller clustering size, say $R = 0.2$. These tiny jets are then discarded if their p_T is below a certain threshold, say $p_T^{\text{tiny}} < 0.01 p_T^{\text{jet}}$, or only a certain number of them, say the five hardest, are kept. Figure 48 shows an example of how trimming removes contamination from underlying event. Note that in contrast to pruning, trimming does not reverse the clustering sequence to find subjets, rather it reclusters with a different scale, in a manner similar to the filtering stage in the Higgs analysis described above, but not solely focused on that final state. Thus, good results can be obtained using anti-$k_\perp$ both to find the original jets, and to find the tiny jets during the reclustering step.

15.4 SUMMARY AND FUTURE WORK

Let us briefly mention some other related ideas which have been proposed. In [205], the authors proposed that greater efficiencies could be obtained by choosing a jet radius continuously which decreases with increasing p_T. Similar scalings had been used previously, for example in the boosted top-tagging study [185], but with a discrete size change rather than a continuous one. The insight of [205] was to make such a scaling automatically part of the jet algorithm. While changing the jet size does seem to produce

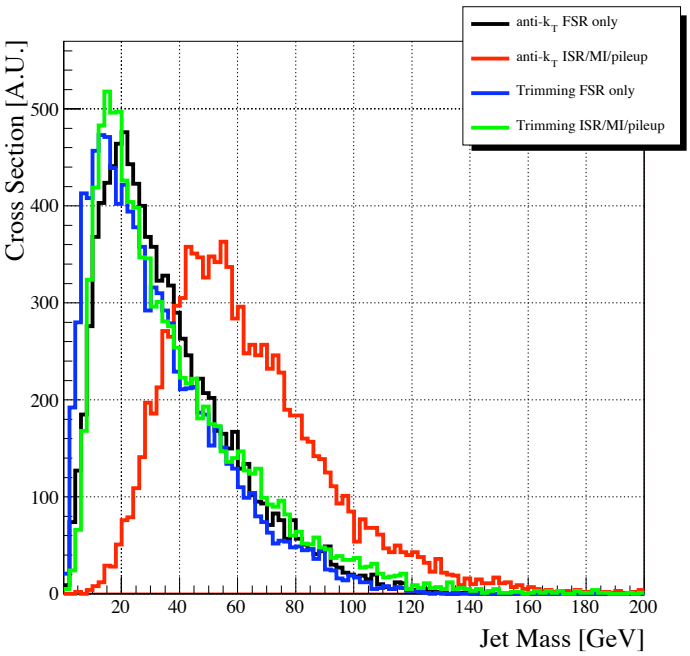


Fig. 48: Effectiveness of jet trimming at removing contamination from the underlying event. The jet mass after trimming appears similar to the jet mass with underlying event turned off. (Figure courtesy of D. Krohn.)

improvements, the benefits may not outweigh the complications in real experimental settings, such as the increased difficulty of jet energy scale calibrations. For example, CMS chose in their implementation of top-tagging [204] to keep the jet size fixed, and were able to achieve comparable efficiencies to the theoretical study with the R variation.

Another idea is to use QCD jet shapes such as planar flow or angularities to distinguish heavy particle decays from QCD jets [206, 207]. A comparison of the efficiency of this approach in top tagging, and a more complete summary of other key issues in jet substructure can be found in [208].

In addition there is clearly potential in studying QCD characteristics beyond the jets. Recently, a proposal was made of for a simple variable *pull* which can characterize the color flow of an event [209]. Color flow is something that is not contained in the 4-momentum of a jet and so may provide a nice handle on new physics which is complementary and uncorrelated with traditional techniques. For example, it may help distinguish color singlet objects (such a the Higgs) from the other color structures in the QCD background. Although color flow is sensitive to aspects of QCD that go beyond the leading-log, DGLAP evolution of basic Monte Carlos, recent progress in the development of Monte Carlos assure us that such effects must be both visible and calculable (cf. Section 10. of these proceedings). The prospects of extracting such superstructure information at the LHC are just beginning to be explored, and there is certainly much more to be done and understood.

In subsequent sections of these proceedings more details of physics relevant to jet substructure are discussed: the status of simulations for jet substructure in heavy particle decays (Section 16.), detector issues (Section 17.), radiation between jets (Section 18.) and calculations of boosted $b\bar{b}$ production in $Wb\bar{b}$ events (Section 12.), which is a significant background to Higgs searches using jet substructure.

16. HEAVY PARTICLE DECAYS IN MONTE CARLO EVENT GENERATORS [28]

In the coming years the LHC will probe a new energy regime one order of magnitude higher than that which has been accessed to date. It is therefore natural that heavy, unstable particles, such as the top-quark, will be produced in large quantities. Indeed, next-to-leading order QCD predictions for top quark pair production indicate that $t\bar{t}$ pairs will be created at a rate of around 0.3 Hz at low luminosity (10 fb^{-1} per year). Of course, it is widely anticipated that physics beyond the Standard Model (SM) will be discovered at the LHC, hence many other species of heavy particles are also expected to be produced.

It is well known that modern collider physics analyses rely upon Monte Carlo simulations for a wide variety of applications, from the modelling of detector effects through to the determination of fundamental parameters; the top-quark mass determination being a case where the dependence on simulations is particularly acute. Since new, unstable, heavy particles feature in almost all extensions of the Standard Model, it is foreseeable that many more analyses of this type will be carried out in the LHC era. Clearly, as the rôle of the event generators in the analysis becomes increasingly significant, so too does the need for them to model the underlying process precisely and, when this is not possible, it is equally important that their limitations and inaccuracies are understood.

In the following we shall critically review some of the main issues to be addressed in the simulation of heavy particle decays. The material is intended to better inform *users* of event generators regarding the nature of the methods and approximations employed, rather than to provide a comprehensive technical review of the state-of-the-art. The first part of the discussion concerns the generalities of the physics phenomena and their simulation, rather than a description of a specific event generator. In the second part we give brief, itemized, technical descriptions of the three main general-purpose event generators, Sherpa [23], Pythia 8 [18, 20] and Herwig++ [21], with reference to the background material.

16.1 SIMULATION STRATEGY

In general, for each event, all of the multi-purpose event generators begin by generating an initial 'hard' configuration of particles according to the associated tree-level matrix elements, after which parton showers are initiated from the external legs. When the showering is complete the final-state consists of a set of partons with constituent masses which are then hadronized. On the other hand, for processes involving the production and decay of unstable particles, including decay chains, rather than attempting to generate the ultimate final-state particles directly, according to high multiplicity matrix elements, each program first generates the momenta of the initial unstable heavy particles in the production phase, before going on to (quasi-)independently generate the momenta of their decay products, showering the colour charged objects in each phase. As well as being more manageable, this approach benefits from being substantially more efficient, computationally, and also from being highly versatile, in keeping with the multi-purpose paradigm that the event generators are built on; in particular, the independent generation of production and decay momenta lends itself naturally, through iteration, to the simulation of arbitrary decay chains. Of course this generality comes at the cost of varying degrees of approximation, which we will now discuss briefly below.

16.2 THE BASE APPROXIMATION

In this subsection we elaborate on the series of approximations by which the production and decay of a particle may be treated entirely independently *modulo* charge and momentum conservation. In doing so we will arrive at *the base approximation*: the common starting point from which the simulations may or may not start to model omitted physics effects and undo approximations by various means, some of which will be addressed in Section 16.3.

[28] Contributed by: K. Hamilton, L. Lönnblad, S. Schumann, T. Sjöstrand, J. Winter

16.21 RESONANT AND NON-RESONANT CONTRIBUTIONS

Experimental signals are defined in terms of the particles which actually enter the detector, leptons, photons and jets, rather than the heavy unstable objects which we typically are interested in studying, *upstream*. The corresponding theoretical predictions are then strictly required to be defined in precisely the same way, including contributions from *all* Feynman diagrams which give rise to the same external states, see e.g. [210]. Generally, however, only a limited subset of these diagrams involve the exchange of s-channel resonances which go on to decay into the stable objects comprising the signal, *i.e.* diagrams representative of heavy particle production and decay (chains). By their nature the resonant graphs give the dominant contributions to the cross section for a given signal and calculations based on this subset of Feynman diagrams are known to provide excellent approximations to the full result. Moreover, in the case where the heavy s-channel particle is that which we wish to study, experimental analyses will regularly employ cuts which further enhance the resonant contributions. In keeping with these observations the simulation of non-resonant effects in all but the simplest processes, namely $2 \to 2$ processes, is typically beyond the scope of the multi-purpose Monte Carlo event generators, furthermore, the loss of accuracy incurred through the omission of the non-resonant contributions is limited to the extent that this approximation is often also employed by tree-level event generators such as Madgraph / MadEvent [81] or Whizard [211], even though they have the facility to include them automatically, in order to increase computational efficiency.

16.22 FACTORIZATION OF PRODUCTION AND DECAY

Although there exists, in general, quantum mechanical interference between the production and decay of an unstable particle, in the limit that the resonant particle is exactly on-shell, the matrix elements corresponding to the resonant graphs may be factorized into two parts, one corresponding to the production of the heavy particle and one corresponding to its decay. Take, for example, a process $pp \to bW^+ + X$, which naturally receives resonant contributions in the form of graphs containing an s-channel top-quark; in the limit that the top-quark momentum is on-shell, one can replace the propagator numerator $i\left(\not{p}_t + m_t\right)$ with a sum over spins:

$$\lim_{p_t^2 \to m_t^2} \mathcal{M} \quad \to \quad \frac{1}{\left(p_t^2 - m_t^2\right)^2 + m_t^2 \Gamma_t^2} \mathcal{A}_\lambda^P \mathcal{A}_{\lambda'}^{P*} \mathcal{A}_\lambda^D \mathcal{A}_{\lambda'}^{D*} \tag{28}$$

where $\mathcal{A}_\lambda^P = \mathcal{A}^P u_\lambda\left(p_t\right)$ and $\mathcal{A}_\lambda^D = \bar{u}_\lambda\left(p_t\right) \mathcal{A}^D$ are the amplitudes for the on-shell production and decay processes, $pp \to t + X$ and $t \to bW^+$ respectively and where the quark helicity indices, λ, λ' are summed over. In the so-called *narrow width approximation* where the product of the unstable particle's width and mass tend to zero we have

$$\lim_{m\Gamma \to 0} \frac{1}{\left(p^2 - m^2\right)^2 + m^2 \Gamma^2} \to \frac{\pi}{m\Gamma} \delta\left(p^2 - m^2\right) , \tag{29}$$

i.e. in the narrow width approximation the replacement of Eq. (28) is valid. A completely analogous replacement to that in Eq. (28) holds for vector bosons, in that case, as with the fermion propagator numerator, the vector boson propagator numerator can be replaced by a sum over polarizations.

A further approximation commonly (but not universally) used in event generators is the neglect of *spin correlations*. This amounts to assuming that in the factorized amplitudes in Eq. (28), the graphs corresponding to different helicity states of the intermediate particle do not interfere with one another and that they are produced in equal measure *i.e.* the full matrix element is effectively replaced by the product of two spin summed matrix elements, one corresponding to the production $\mathcal{M}^P$, the other to the decay $\mathcal{M}^D$. In our pedagogical example, working in the narrow width approximation and neglecting spin correlations, the exact squared matrix is replaced according to

$$\lim_{m_t \Gamma_t \to 0} \mathcal{M} \quad \to \quad \frac{\pi}{m_t \Gamma_t} \delta\left(p_t^2 - m_t^2\right) \mathcal{M}^P \mathcal{M}^D .$$

In addition, by inserting $\mathrm{d}q_t^2\,\delta\left(q_t^2 - p_t^2\right)$ and $\mathrm{d}^4p_t\,\delta\left(p_t - p_b - p_{W^+}\right)$ into the usual Lorentz invariant phase space measure we may, without approximation, factorize the Lorentz invariant phase space, $\mathrm{d}\Phi$, for $bW^+ + X$ into parts corresponding to the same production and decay processes:

$$\mathrm{d}\Phi \;=\; \mathrm{d}\Phi_P\,\mathrm{d}\Phi_D\,\frac{1}{2\pi}\,\mathrm{d}q_t^2 \tag{30}$$

where,

$$\begin{aligned}\mathrm{d}\Phi_P &= \mathrm{d}\Phi_t\mathrm{d}\Phi_X\,(2\pi)^4\,\delta^4\left(p_a + p_b - p_t - p_X\right),\\ \mathrm{d}\Phi_D &= \mathrm{d}\Phi_b\mathrm{d}\Phi_{W^+}\,(2\pi)^4\,\delta^4\left(p_t - p_b - p_{W^+}\right),\end{aligned} \tag{31}$$

and for a given particle i, $\mathrm{d}\Phi_i$ is the usual invariant phase space measure

$$\mathrm{d}\Phi_i = \frac{\mathrm{d}^3p_i}{(2\pi)^3\,2E_i}\,. \tag{32}$$

Using this phase space factorization together with the narrow width approximation one can see that the differential cross section will factorize into two parts; denoting flux and PDF factors by $\mathcal{L}\left(x_\oplus, x_\ominus\right)$ we have,

$$\begin{aligned}\mathrm{d}\sigma &= \mathrm{d}x_\oplus\mathrm{d}x_\ominus\,\mathcal{L}\left(x_\oplus, x_\ominus\right)\,\mathrm{d}\Phi\,\mathcal{M}\\ &= \mathrm{d}x_\oplus\mathrm{d}x_\ominus\,\mathcal{L}\left(x_\oplus, x_\ominus\right)\,\mathrm{d}\Phi_P\,\mathcal{M}^P \times \frac{1}{\Gamma_t}\,\mathrm{d}\Gamma_t\\ \mathrm{d}\Gamma_t &= \frac{1}{2m_t}\,\mathcal{M}^D\,\mathrm{d}\Phi_D\,,\end{aligned} \tag{33}$$

where the first part corresponds to the spin summed production of the unstable particle and the second part to its branching ratio, differential in the decay phase space. By appealing to the narrow width approximation it is therefore possible to first generate the production and decay processes completely independently in their rest frames, then boost the decay products to the frame in which the particle was produced. The latter boost is ambiguous up to a rotation which is a direct manifestation of having neglected spin correlations.

The simulation procedure and the approximations which we have described up to this point comprise the most basic algorithm employed by Monte Carlo simulations for the treatment of heavy particle production and decay; of course, for the parton shower simulations, these tree level configurations will, by default, go on to include the effects of soft and collinear parton emissions from the coloured particles. Unless otherwise stated this represents the base accuracy and physics describing the interplay of the production and decay of unstable particles in these simulations: non-resonant contributions, spin correlations and off-shell / finite width effects are all completely neglected.

16.3 BEYOND THE BASE APPROXIMATION

Having noted the basic nature of the base approximation we will now briefly review some of the ways in which it can be improved on. Note that these enhancements can be extremely sophisticated and so in no case should they be assumed to have been implemented as a default option, for the same reason, it should be clear from the simulations' documentation whether such features are available or not.

16.31 FINITE WIDTH EFFECTS

Although the factorization of the numerator of the unstable particle propagator only holds when it is exactly on-shell, the error induced by using the replacement Eq. (28) when this is not true is $\mathcal{O}\left(\Gamma/m\right)$ *i.e.* the typical amount by which the particle is off-shell with respect to its mass. Of course, if cuts or

observables are such that the channel in question only contributes when the particle is far off-shell, *e.g.* if cuts correspond to a mass-window restricting the invariant mass to be far from the pole mass, the error is much greater and, in any case, in such circumstances the inclusion of non-resonant contributions is mandatory. Provided the signal regions receive contributions from the immediate vicinity of the resonance we may recover its line-shape up to corrections $\mathcal{O}(\Gamma/m)$, by inserting 1 into the factorized cross section formulae as,

$$1 = \int \mathrm{d}p^2 \, \frac{m\Gamma}{\pi} \, \frac{1}{(p^2 - m^2)^2 + m^2\Gamma^2} \, . \tag{34}$$

This is equivalent, of course, to not taking the exact limit $\Gamma \to 0$ in the Breit-Wigner function in Eqs. (28) and (29), we stress that this limit is nevertheless *implied* by assuming that the propagator numerator may be replaced by the sum over external wavefunctions.

In terms of the event generation procedure, the inclusion of finite width effects in this way, leads to a trivial modification of that described above for the $\Gamma = 0$ case, namely, that the first step in the algorithm is now the generation of an off-shell mass for the intermediate particles according to a Breit-Wigner function. The same procedure can be adopted using running widths.

Although this seems, in principle, a relatively straightforward means of improving the base approximation ($\Gamma = 0$), a thorough treatment of the particle width, in particular those involving next-to-leading order computations, requires that further, more subtle, issues be addressed before it is introduced anywhere at all. Of note is the question of how to ensure that the introduction of the finite width does not violate gauge invariance? The particle width term arises from the resummation of the imaginary parts of the self-energy insertions on the unstable particle's propagator *i.e.* it corresponds to the inclusion of a *subset* of higher order corrections to the process under consideration[29]. Since S-Matrix elements are only guaranteed to be gauge invariant at each order in perturbation theory when *all* relevant diagrams at the given order are summed, one should expect that naively adding a width term to the propagator will violate gauge invariance. Although this is the case in general, usually the numerical impact of the gauge breaking terms is negligible and is contained within that incurred by writing the propagator numerator as a polarization sum.

On the other hand, gauge invariance underlies the cancellation of large logarithmic corrections to inclusive quantities as laid out in the Bloch-Nordsieck [212] and KLN [213, 214] theorems, therefore, in cases where logarithmically enhanced emissions occur, the gauge variant contributions can be hugely magnified and even exceed the correct, gauge invariant, contributions. These potentially dangerous cases are confined to processes involving diagrams in which the unstable particle we are studying emits soft / collinear massless partons or photons *e.g.* gluon emission from top quarks or photon emission from W bosons. For simulations involving tree level matrix elements and / or parton showers this issue is substantially negated due to the obligatory generator level cuts, which exclude the phase space for collinear and soft regions and, also, because the splitting functions at the heart of the parton shower dynamics are derived from the *universal* factorization of matrix elements with soft / collinear emissions. Furthermore, it is clear that for the Lagrangian to be gauge invariant it is not necessary for the parameters, in particular the masses, to be real numbers. Bearing this in mind, one may adopt the so-called *complex mass scheme* [215] whereby finite width effects are included in a gauge invariant fashion by simply reinterpreting the masses, *everywhere* that they occur in the Lagrangian, as being complex numbers with the $im\Gamma$ width factor absorbed inside them. Alternatively the so-called 'fudge-factor' scheme [216] is employed whereby the total squared matrix element, neglecting finite width effects, is multiplied by the product of the Breit-Wigner function and the zero-width propagator, $p^2 - m^2$, squared for each resonant particle in the process. For further technical details concerning gauge invariance and finite width effects see Refs. [215–219].

Another pertinent issue is that of how to include threshold effects in the running width and the selection of decay modes. In the forthcoming years LHC analyses will predominantly be concerned with searches

[29] It therefore also has a dependency on the particle's mass, which we shall omit here for simplicity.

for new particles, whose masses couplings and decay modes are all unknowns. A prime example is the Standard Model Higgs boson, whose width and decay modes strongly depend on, not simply its on-shell mass but, in view of the factorization of the production and decay, also the off-shell mass generated at the beginning of the simulation, Eq. (29). This is of particular importance for particles decaying close to threshold (as the Higgs boson may do), in this case the effects from the off-shell propagator must be taken into account. This can be achieved by implementing the running width, *i.e.* including the full dependence on the particle's *off-shell* mass, in the weight factor in Eq. (34), *as well as*, in the calculation of the partial width of a selected decay mode as in Ref. [220], where this prescription is shown to reproduce well the results of some three-body decays simulated as a cascade of two two-body decays.

16.32 SPIN CORRELATIONS

As stated previously in the base approximation particle decays are performed according to the spin averaged matrix element for the decay process hence they occur isotropically in the rest frame. It follows that in order to have a reasonable description of observables sensitive to the details of *individual* decay products it will be necessary, at least, to communicate the spin information from the production to the decay parts of the simulations. Notable studies for which such correlations are particularly relevant include: the production and decay of the top quark, the production of tau leptons in Higgs decays and the spin determination of newly discovered particles *i.e.* understanding the nature of physics beyond the Standard Model.

Generalizing earlier work of Knowles and Collins [221–223], a flexible, efficient algorithm for propagating spin correlations between particle production and decay in Monte Carlo event generators, using the spin density matrix formalism, was developed by Richardson [224]. Rather than discuss the algorithm in full detail here we will describe it by considering the example of the process $h\,h \to t\,\bar{t}$ where the top quark subsequently decays, via a W boson, to a b quark and a pair of light fermions. Initially, the outgoing momenta of the $t\,\bar{t}$ pair are generated according to the partonic differential cross section

$$\mathrm{d}\widehat{\sigma}^P = \frac{1}{2\hat{s}}\,\mathcal{A}^P_{\lambda_t\lambda_{\bar{t}}}\,\mathcal{A}^{P*}_{\lambda_t\lambda_{\bar{t}}}\,\mathrm{d}\Phi_P \tag{35}$$

where $\mathcal{A}^P_{\lambda_t\lambda_{\bar{t}}}$ is the amplitude for the initial hard process, with λ_t and $\lambda_{\bar{t}}$ denoting the helicity indices of the produced t and $\bar{t}$ respectively. $\mathrm{d}\Phi_P$ is the usual phase space measure for the production process, equal to that in Eq. (31), where in this example $X = \bar{t}$. Having generated the momenta for the production process, one of the outgoing particles is then picked at random, say the top and a production spin density matrix is calculated

$$\rho^t_{\lambda_t\lambda'_t} = \frac{1}{N}\,\mathcal{A}^P_{\lambda_t\lambda_{\bar{t}}}\mathcal{A}^{P*}_{\lambda'_t\lambda_{\bar{t}}}\,, \tag{36}$$

with N defined such that ρ^t has unit trace.

The top is decayed and the momenta of the decay products distributed according to

$$\mathrm{d}\Gamma_t = \frac{1}{2m_t}\,\rho^t_{\lambda_t\lambda'_t}\,\mathcal{A}^{D_t}_{\lambda_t\lambda_{W^+}}\,\mathcal{A}^{D_t*}_{\lambda'_t\lambda_{W^+}}\,\mathrm{d}\Phi_D\,, \tag{37}$$

where the inclusion of the spin density matrix ensures the correct correlation between the top decay products and the beam. We note how the spin averaged result, Eq. (33), is recovered by replacing $\rho^t_{\lambda_t\lambda_{\bar{t}}} \to \frac{1}{2}\,\delta_{\lambda_t\lambda_{\bar{t}}}$. With the momenta for the top quark decay products now in hand we may calculate a production spin density matrix for the unstable W^+

$$\rho^{W^+}_{\lambda_{W^+}\lambda'_{W^+}} = \frac{1}{N}\,\rho^t_{\lambda_t\lambda'_t}\,\mathcal{A}^{D_t}_{\lambda_t\lambda_{W^+}}\,\mathcal{A}^{D_t*}_{\lambda'_t\lambda_{W^+}}\,, \tag{38}$$

and then generate its decay by analogy to Eq. (37). Here the use of the production matrix calculated from Eqs. (36) and (38) leads to the correct angular correlations between the light fermions, the beam and the bottom quark.

Since the children of the W^+ are light fermions the decay chain terminates and a decay matrix for the W^+ is calculated:

$$D^{W^+}_{\lambda_{W^+}\lambda'_{W^+}} = \frac{1}{N}\,\mathcal{A}^{D_t}_{\lambda_t\lambda_{W^+}}\,\mathcal{A}^{D_t*}_{\lambda_t\lambda'_{W^+}}\,, \tag{39}$$

Moving back up the decay chain, the analogous decay matrix is calculated for the top quark using the decay matrix of the W^+:

$$D^{t}_{\lambda_t\lambda'_t} = \frac{1}{N}\,\mathcal{A}^{D_t}_{\lambda_t\lambda_{W^+}}\,\mathcal{A}^{D_t*}_{\lambda'_t\lambda'_{W^+}}\,D^{W^+}_{\lambda_{W^+}\lambda'_{W^+}}\,. \tag{40}$$

Following this one contracts $D^t_{\lambda_t\lambda'_t}$ with the full $t\bar{t}$ production spin density matrix, $\mathcal{A}^P_{\lambda_t\lambda_{\bar{t}}}\,\mathcal{A}^{P*}_{\lambda'_t\lambda'_{\bar{t}}}$, to give the production spin matrix of the antitop quark, $\rho^{\bar{t}}_{\lambda_{\bar{t}}\lambda'_{\bar{t}}}$, analogous to Eq. (36), before generating its decay products in the same way as was done for the top quark. In progressing forward along the $\bar{t}$ decay chain the production spin density matrices pass information from one decay to the next leading to the correct angular correlations.

Other methods for the inclusion of spin correlations in event generators are in use, although many of these may be regarded as approximations or variants of the above [81]. The only other approach which may be regarded as distinct, involves decaying all unstable particles in the event in a single step, using matrix elements for the process under study where these particles have and have not decayed. This algorithm has been applied in Pythia, for quite some time, for a number of processes and has been recently been extended in the context of matched next-to-leading order simulations by Frixione *et al* [114]. This alternative method is based on the observation that the product of the separately spin summed matrix elements for the production and decay processes, including propagator factors, always exceeds the full spin summed matrix element. Knowing this, spin correlations may be implemented by generating production and decay momenta as in the base approximation (with the decays isotropic in their rest frame) choosing a random number $\mathcal{R}$ and then regenerating the decay kinematics if $\mathcal{R}$ is greater than the ratio of the full spin summed matrix element to the spin summed matrix element omitting spin correlations. For a given set of production momenta, isotropic decay kinematics and new values of $\mathcal{R}$ are generated repeatedly until $\mathcal{R}$ is less than the aforesaid ratio, at which point the decay kinematics can be kept.

Since the underlying Monte Carlo algorithm in this alternative approach is just the hit-or-miss method, with no spin density matrices, spinor phases *etc*, to keep track of, this method is substantially easier to implement than the one above. On the other hand this method assumes that a matrix element generator capable of providing helicity amplitudes for (arbitrarily) high multiplicity final states is available, moreover, the intention here is to simultaneously generate the momenta for all decaying particles, in 'one shot', which also impedes the event generation efficiency with respect to the recursive approach based on spin density matrices. Nevertheless, for including spin correlations in events with relatively few decays, this method offers a simple and straightforward alternative to the spin density matrix approach, where the relative inefficiency tends to add little, with respect to other elements of the simulation, to the overall event generation time. This method has been used widely in simulations based on the `MC@NLO` [12, 114] and `POWHEG` [14, 118, 124] methods.

16.33 QCD RADIATION

Having generated the decay momenta, with or without finite width effects and spin correlations, if the mother particle or any of the daughter particles carry a colour charge, one should attempt to model the effects of QCD radiation which they may emit. The vast majority of these emissions will be soft and / or collinear emissions with respect to the shower progenitors. This is straightforward to understand from the point of view of the associated amplitudes: in the limit that a massless particle is emitted with low p_T, the propagator denominator, associated to the internal line that was its mother, vanishes so the matrix element diverges for such configurations.

Such low p_T corrections are precisely those which parton shower simulations take into account, to all orders in perturbation theory. Of course, distributing emissions according to arbitrarily high multiplicity

matrix elements is not feasible, instead we appeal to the fact that, in the limit that a massless parton j is emitted collinear from an external leg i, the matrix elements factorize according to

$$\mathcal{M}_{n+1} = \frac{8\pi\alpha_{\rm S}}{q^2_{\widetilde{ij}} - m^2_{\widetilde{ij}}} \, P_{\widetilde{ij}\to ij} \, \mathcal{M}_n \, , \tag{41}$$

where $\widetilde{ij}$ denotes the mother of the branching $\widetilde{ij} \to ij$, with virtuality $q^2_{\widetilde{ij}}$, and $P_{\widetilde{ij}\to ij}$ is an Altarelli-Parisi splitting function. The form of the splitting functions does not depend on $\mathcal{M}_{n+1}$ and $\mathcal{M}_n$, where the latter is the matrix element for the n particle process, in which the branching does not occur. The matrix element factorization in Eq. (41) may be combined with a phase space factorization for the branching, leading to a factorized form of the differential cross section for the $n+1$ particle process, in a similar vein to that described in Section 16.22, in the context of heavy particle decays. Here, as with cascade decays, the factorization may be applied recursively, albeit with the caveat that using Eq. (41) implies the emissions be ordered such that they become increasingly collinear as one works from the core of the process out toward the external legs. As well as modelling the effects of enhanced collinear real emission, the parton shower method also, necessarily, incorporates their corresponding virtual corrections through *Sudakov form factors*, the effect of which is to dampen, and ultimately regulate, the diverging emission rate as $p_T \to 0$.

Depending on the accuracy of the parton shower, there is some freedom in the definition of 'increasingly collinear'. Parton shower algorithms may be formulated as an evolution in the virtualities of the branching partons, or as an evolution in the transverse momentum of the branching products. However, formal studies of colour coherence [101, 140–148] reveal that branchings involving soft gluons should be ordered in the angle between the branching products. This is a dynamical effect whereby wide-angle soft gluon emissions, from near collinear configurations of two or more partons, have insufficient transverse resolving power to be sensitive to the constituent emitters. In effect the resulting radiation pattern is determined by the colour charge and momentum of the *mother* of the emitters, rather than the emitters themselves [30].

In the LHC era we will regularly face the situation in which one cannot neglect the mass of radiating particles when modelling the distribution of their emissions, $t\bar{t}$ bar production being an obvious example. Nevertheless, the Altarelli-Parisi splitting functions governing the distribution of emissions in the older generation of shower simulations are based on the factorization of matrix elements where the mother parton, $\widetilde{ij}$, and its two daughter partons i and j are all massless. As stated above, in the limit that massless partons emit low p_T radiation the associated amplitudes diverge, however, if the emitter has a mass, the propagator denominator is proportional to $1 - \beta\cos\theta$, where the factor of β, the velocity of the emitter, screens the collinear divergence. This mass effect in the propagator denominator then greatly alters the distribution of emissions as the small angle region is approached with respect to the rather crude massless approximation, which will likely have important consequences for analyses *e.g.* the top mass determination, where a good understanding of production radiation and b quark fragmentation are key. However, in recent years, developments in the resummation of radiation from massive partons have included the introduction of *quasi-collinear splitting functions* which generalize those of Altarelli and Parisi, so as to retain the full dependence on the mass of the emitting particle in the collinear limit [227]. For example, the quasi-collinear $q \to qg$ splitting function is

$$P_{q\to qg} = C_F \left[\frac{1+z^2}{1-z} - \frac{2z\,(1-z)\,m_q^2}{p_T^2 + (1-z)^2\, m_q^2} \right] , \tag{42}$$

in which z is the light-cone momentum fraction carried by the daughter quark and p_T^2 is the transverse momentum of the branching; note how the form of the first term in Eq. (42) is the same as the usual massless $q \to qg$ Altarelli-Parisi function. The inclusion of these mass effects in parton shower simulations

[30] To avoid digressing from our stated objectives we limit ourselves to this minimal description of the parton shower approximation and refer the interested reader the following dedicated review articles [14, 18, 225, 226].

is actually non-trivial, since the full showering formalism is required to retain mass effects from the beginning *i.e.* in the shower variables, the Sudakov form factors, the cut off scales, the phase space mapping and kinematics reconstruction / momentum reshuffling.

Since, in the narrow width approximation, the production and decay processes are assumed to factorize into two independent parts, this is the sense in which we simulate the associated parton showers, effectively treating the decay as a completely independent hard sub-process. For the decay process the effects of QCD radiation are then simply modeled by initiating parton showers from the incoming and external particles, in the rest frame of the decaying particle. Naturally then, for the decay products, the same final-state showering routines are applied as are used for the hard scattering process, including those used in the initialization phase in which starting scales are assigned. Only the initial conditions for the shower evolution are different, although their choice is, nevertheless, still based on examining the colour flow in the underlying hard decay process.

The leading order decay products may not, however, be the only particles to emit radiation. Massive coloured particles, such as the top quark and particles in many models of physics beyond the Standard Model, may decay on time-scales shorter than that characteristic of the hadronization process. Consequently, subject to the available phase space, as well as undergoing time-like showers $\left(q^2 > m^2\right)$ in their production phase, these particles will also undergo a further space-like showering $\left(q^2 < m^2\right)$ of QCD radiation *prior* to their decay. In contrast to the final-state showers from the decay products, the initial-state space-like shower created by a decaying particle is quite different to that of an initial-state particle from the production process. In particular, it involves no PDFs, since the heavy parton originates from a hard scattering as opposed to a hadron. Furthermore, in the hard process it was necessary to evolve the initial-state partons backwards from the hard scattering to the incident hadrons, to efficiently sample any resonant structure in the underlying matrix elements. On the contrary, in decay processes, degrading the invariant mass of the decaying particle, via the emission of radiation, does not affect the efficiency with which any resonant structures in the decay matrix element are sampled. Hence, it is natural for the evolution of space-like decay showers to start with the unstable particle from the production process, and evolve it forward, towards the decay. In the older generation of multi-purpose event generators the inclusion of this initial-state radiation in decays was not in general possible due to, for instance, their use of non-covariant shower formalism.

16.34 QED RADIATION

In existing Monte Carlo simulations the production of QED radiation in particle decays is normally simulated using an interface to the PHOTOS program [228–230]. This program is based on the collinear approximation for the radiation of photons together with corrections to reproduce the correct soft limit [228, 229]. In recent years the program has been improved to include the full next-to-leading order QED corrections for certain decay processes [230–232].

As noted previously, in the event that the decay products are heavy, the divergences associated to multiple collinear emission are subject to a screening which increases with the mass of the emitter. On the other hand, since the propagator denominator associated to an emitter is also proportional to the energy of the emitted particle, soft emissions are always enhanced regardless of the emitting particle's mass or the emission angle. This being the case it is often preferable, depending on the application / particle masses / analysis cuts, to model the effects of photon radiation using a formalism which resums to all orders the effects of soft rather than collinear emissions.

In QED soft infrared divergences are resummed to all orders using the methods originally developed by Yennie Frautschi and Suura [233] (YFS), in doing so the soft real and virtual corrections are seen to exponentiate[31]. This resummation has been recast as photon shower simulations through a *tour de force* in Monte Carlo techniques pioneered by S. Jadach [236–241]. Such simulations have no overlap with

[31] Pedagogical examples of the exponentiation of infrared QED corrections may be found in Refs. [234, 235]

their QCD counterparts; whereas in QCD the shower forms through iterating emissions in a Markovian algorithm, the YFS QED shower algorithms are Poissonian, generating all photon radiation at the same time in 'one-shot', moreover they have always afforded the possibility to systematically include arbitrarily higher-order corrections.

Finally, we remark that *soft* resummation in QCD is fundamentally different since, unlike QED, each emitted parton is itself charged. Unfortunately such additional complexities arising from the non-Abelian character of QCD mean that applying the same Monte Carlo approach to simulating QCD radiation, despite its many attractive features, is unfortunately not appropriate.

16.35 HARD QCD RADIATION

Note that the standard QCD parton shower approach has two important drawbacks. Firstly, because the parton shower generates emissions from each leg of the hard scattering independently, each additional emission must be uniquely associated to a particular leg of the hard scattering, which can only be achieved at the price of having regions of phase space, corresponding to high p_T gluon emissions, which are unpopulated by the shower, so-called *dead-zones*. Secondly, the soft / collinear approximation to the QCD matrix elements is plainly not a good approximation all over the phase-space region populated by the parton shower.

One way to rectify these problems is through *matrix element corrections* [24] which ensure that the hardest additional radiated parton in the event is distributed *exactly* according to the corresponding real emission matrix element. In the case of *soft* matrix element corrections to decays, every emission generated in the shower which is the *hardest so far* is vetoed if the ratio of the exact differential width to the parton shower's approximation to it (based on the soft / collinear approximation to the former) is greater than a random number $\mathcal{R}$.

Where the parton shower algorithm is such that there is a dead zone in the phase space of the *first* emission, typically concentrated in the high p_T region, *hard matrix element corrections* are also necessary. Hard matrix element corrections simply use the same exact real emission matrix element to generate an event with a hard emission in the dead zone, with a probability given by the ratio of the integrated cross section in the dead zone divided by the total cross section of the leading order process. As a consequence of the different ordering variables used in the Herwig / Herwig++ and Pythia shower algorithms, the entire phase space for the first emission is covered in the latter, avoiding the need for hard matrix element corrections.

Alternatively one may use matrix element-parton shower matching techniques (ME-PS), such as the CKKW method, [4, 126] to correct the approximate soft / collinear radiation pattern of the parton shower in the regions of phase space corresponding to hard emissions, where in this case the hardness is typically measured in terms of the Durham jet measure. In general these ME-PS separate phase space into two regions according to a merging scale, defined in terms of the jet measure, y_{merge}. The region of phase space corresponding to values of $y < y_{\mathrm{merge}}$, is deemed to correspond to sufficiently soft / collinear emissions that one can expect the shower approximation to reliably distribute emissions there, this is the so-called *shower region*. Above the shower region is the *matrix element region*, where emissions are distributed according to exact fixed order matrix elements. Of course the real emission fixed order matrix elements do not include any of the virtual effects resummed in the parton shower and so the emissions in the matrix element region must be reweighted with Sudakov form factors and running coupling constants to take these important effects into account. Without these corrections distributions will exhibit unphysical discontinuities across the phase space partition at y_{merge}.

Both the matrix element correction methods and the ME-PS methods effectively take into account the real emission component of next-to-leading order corrections but not the virtual effects. Nevertheless, there are many examples in the literature (albeit for production processes) where these methods are shown to give excellent agreement with next-to-leading order calculations from the point of view of the *shapes*

of distributions [120, 121, 153–155, 242].

16.4 Herwig++

- Finite width effects
  Finite width effects are present by default for all processes in Herwig++ through the inclusion of a weight factor as in Eq. (34), retaining the dependence of the width on the off-shell mass of the particle, in *both* the production and decay stages. This approach is used in the simulation of cascade decays, where it was shown to give good agreement with results obtained using exact matrix elements in Ref. [220].
- Spin correlations
  The Herwig++ spin correlation algorithm is precisely that described in Section 16.32 and was implemented by the author of Ref. [224].
- QCD radiation
  The chief success of the older fortran Herwig 6 program was in its accounting of soft gluon interference effects, specifically, in particular colour coherence phenomena, through the angular ordering of emissions in the shower. The new Herwig++ algorithm retains angular ordering as a central feature of the showering algorithm and improves on it, most notably, in the present context, through the inclusion of the mass-dependent splitting functions and kinematics, providing a physical description of the radiation distribution emitted by massive particles in the low p_T region [21, 243, 244]. The facility to model initial-state parton showers in the decays of unstable coloured particles was envisaged at the start of the Herwig++ project and has been implemented in full generality in 2006.
- QED radiation
  The Herwig++ program includes a module which dresses the decays of light and heavy objects, involving a single electric dipole, with QED radiation [245]. This simulation uses the YFS formalism to resum the effects of multiple photon emissions, as well as a Poissonian shower algorithm following the methods of Jadach [236].
- Hard radiation
  In implementing colour coherence through angular ordering the Herwig 6 and Herwig++ shower algorithms contain dead zones in the phase space for the hardest emission, where the hardest emission occurs at wide angles with respect to the parent(s). In the context of particle decays Herwig++ includes matrix element corrections for top quark decays as described in [244]. This serves to populate the dead zone for this process and correct the radiation pattern in the live zones. Matrix element corrections to $W^{\pm}$, Z, and Higgs boson decays are not available, however, due to the neutrality of the decaying particle and the low mass of the negligible mass of their decay products, these may be implemented with quite a modest effort.

16.5 Pythia

- Finite width effects
  All particles that are defined with a width are distributed according to a Breit-Wigner, cf. eq. (34). In case of pure s-channel processes, such as $e^+e^- \to \gamma^*/Z^0 \to f\overline{f}$, a running width is used, in agreement with LEP1 conventions. For other processes normally a fixed width is used. The inclusion of this width in the matrix-element expressions, especially the interference terms, may vary from process to process — Pythia 8 does not have a matrix-element generator of its own, but encodes calculations made by many different authors.
- Spin correlations
  Pythia 8 does not come with a spin correlation algorithm, which means that the default is isotropic decays. For many processes and decays spin correlations are included with full matrix elements, however. That is, for a process such as $e^+e^- \to W^+W^- \to f_1\overline{f}_2 f_3\overline{f}_4$, first kinematics (including

masses) is selected for the $e^+e^- \to W^+W^-$ process, and thereafter the decay angles of the two W's are simultaneously sampled and weighted according to the complete $2 \to 2 \to 4$ matrix elements. For cases where spin is important and not implemented, users are recommended to use external input from dedicated matrix-element programs, such as MadGraph [81].

- QCD radiation
  The original Pythia 8 shower algorithm was mass-ordered [246], with extra angular-ordering cuts, while the current one is transverse-momentum-ordered [69], with a dipole approach to handling recoil effects. One nice feature of these algorithms is that they cover all of the phase space, at least for the first emission, thus obviating the need for several of the special HERWIG tricks described above, e.g. to handle dead zones or to separate off space-like emissions in resonance decays. The modification of the evolution variable from $\mathrm{d}m^2/m^2$ to $\mathrm{d}m^2/(m^2 - m_0^2)$ gives a consistent coverage of the soft/collinear region $m^2 \to m_0^2$ for radiation off massive particles [247]. Here m_0 is the rest mass or, for a resonance, the previously Breit-Wigner-sampled mass. This approach is also easily extended to the case of evolution in $p_\perp^2 = z(1-z)m^2$.
- QED radiation
  The same shower algorithm that does QCD emissions also can handle QED ones, in an interleaved manner. That is, QCD and QED emissions can alternate in the downwards evolution in m^2 or $p_\perp^2$, and thus compete for momentum. The QCD and QED dipoles do not need to agree.
  Optionally the PHOTOS package may be used [230–232]; by experimentalists often to handle QED corrections in hadronic resonance decays.
- Hard radiation
  The Pythia shower can easily be fixed up to overestimate radiation in the hard region, while attaching to the correct expression in the soft one (see above). A Monte Carlo rejection factor can then be implemented to match the first gluon emission to the respective QCD matrix-element expression for essentially all decays in the SM and the MSSM [247]. For QED such corrections are only included in γ^*/Z^0 and $W^\pm$ decays.

16.6 Sherpa

- Finite width effects
  In Sherpa finite fixed widths are incorporated in the propagator denominators of the internal tree-level matrix element generators AMEGIC++ [149] and COMIX [150]. The fixed width scheme is employed within these matrix element generators.
- Spin correlations
  In order to include spin correlations between production and decay Sherpa essentially employs the first algorithm described in the corresponding section above. The versatility of the implementation is enhanced by Sherpa's ability to automatically generate the matrix elements needed for the production and decay spin density matrices internally. Along this line, Sherpa allows on the level of hard matrix-element generation for the specification of the production and decay processes in a cascade-like manner. As a result contributions from non-resonant diagrams can be omitted while one preserves off-shell mass effects and spin correlations in the generation of the fully decayed final states. To incorporate spin correllations in the decays of hadrons and τ-leptons the simple rejection algorithm outlined above is employed.
- QCD radiation
  The new generation of Sherpa versions 1.2.x uses the CSS parton-shower algorithm [89], which is based on (massive) Catani–Seymour dipole subtraction kernels and a dipole-like picture of shower evolution. Emissions are ordered in transverse momentum rather than virtuality. The CSS implements truncated showering, therefore handles final-state and initial-state emissions in the production and decay of unstable coloured particles, respectively. The Sherpa parton shower of versions 1.1.x and older, APACIC++ [248], is closely related to that of the Pythia virtuality ordered shower [18].

- QED radiation
  Sherpa includes a native simulation of QED radiation in decays, PHOTONS++ [249]; like the Herwig++ QED shower this module is built upon the YFS resummation formalism and the Poissonian shower algorithms of the Krakow group [236–241].
- Hard radiation
  Here again Sherpa makes use of its internal matrix element generators to generate amplitudes for the leading order processes in the presence of additional QCD radiation. The key point is that processes of different parton multiplicity in the final state can be combined consistently with parton showers (generated by the CSS algorithm) and hadronization included according to the new strategy of matrix-element and truncated-shower merging (ME&TS) [65, 151, 152]. This approach improves over the CKKW method, implemented in versions 1.1.x and older, since it guarantees an unified treatment of local scales in the calculation of the matrix elements and parton showers. Hence, one finds that the systematic uncertainties of the ME&TS merging are sizeably reduced with respect to CKKW. Using the ME&TS matching, extra hard radiation can be reliably described in multi-jet and high-p_T scenarios. For example, Sherpa employs the new method to include exact real emission corrections to the $t\bar{t}$ production process, as well as the t and $\bar{t}$ decay processes, in the region of phase space above the ME&TS merging scale.

17. SENSITIVITY OF QCD JET MASS AND JET SUBSTRUCTURE TO PILE-UP AT LHC[32]

17.1 Introduction

New methods to measure the hadronic decays of boosted massive particles at the LHC, like W bosons [182], top quarks (e.g., [185, 186, 250]), Higgs boson [158], and BSM particles (e.g., [183]), have been proposed in recent years. Due to the boost of these massive particles, the decay products are collimated in one direction, and they are reconstructed as one single jet. The properties of these jets, such as their masses and jet sub-structure quantities, can be used to separate them from the standard QCD jets. Several strategies have been developed to distinguish these two sources and to study the properties of the corresponding jets, including the use of jet algorithms with variable size parameters [205], jet pruning, see e.g. [193], and jet trimming [194].

In this study we investigate the impact of the pile-up on the measurement of these observables. The presence of the pile-up interactions affect the low transverse momentum (p_T) spectrum. We expect that it strongly affects the jet energy calibration in the very low p_T region, becoming negligible for high p_T jets. The estimate of the effects on the jet mass and on the the jet-substructure quantities is less straightforward because these quantities are complex QCD observables.

Another important aspect is the capability to properly reconstruct and calibrate the jets. There are several approaches to assess these points. Most of them are finalized at the calibration of the jet as a single global object starting from calorimetric signals. In this case, the capability to properly calibrate every component (i.e. calorimetric signals) of the jet is not straightforward and the precision on the measurement of the substructure should be investigated in a detailed analysis of the detector performance. This analysis is beyond our goal.

In this study only the effects of the pile-up and one simple strategy to suppress it are investigated. In order to separate these effects from the detector calibration, we reconstructed jet from Monte Carlo generated particles (hadrons). The selection cuts, used to suppress the pile-up, are driven by consideration of the detector effects, especially for the reconstruction of low p_T particles, i.e. the effect of the magnetic field bending, energy losses in inactive material, and of the signal selection.

17.2 Jet reconstruction, observables and selection cuts

In this study we used a sample of QCD di-jet events in proton-proton collisions with center of mass energy $\sqrt{s} = 10$ TeV simulated with Pythia 8.120 [18, 20]. The generation was divided in different samples according to the following bins (GeV) in $\hat{p}_{\mathrm{T}}$: [17,34], [34,70], [70,140], [140,280], [280,560], [560,1120], > 1120. For the first six samples, 500,000 events were simulated, for the highest $\hat{p}_T$ sample a simulation of 200,000 events was used.

The effect of the pile-up is taken in account by adding four (Poisson distributed) additional proton-proton interactions per event. These interactions were simulated using the parameter "SoftQCD:all" in Pythia 8.120, i.e. they include elastic, single diffractive, double diffractive scattering and minimum bias interactions.

All the particles generated by the hard scatter and by the pile-up are used to build the final jets. The jet algorithm used for this purpose is the anti-$k_\perp$ [178] algorithm with distance parameter $D = 0.4$. Only the leading jet is used in this study. It is required to point into the pseudo-rapidity range $-3.2 \leq \eta \leq 3.2$.

The observables under consideration are the jet p_T, its mass m_{jet}, and the y-scale y_{scale} [182] associated with the last clustering step in the formation of the jet. For the Anti-$k_\perp$ jet algorithm, this last observable is not directly available, as it only has a meaningful definition for the $k_\perp$ [251, 252] jet algorithm. We therefore re-clustered the constituent particles of the Anti-$k_\perp$ jet using the $k_\perp$ algorithm, thus allowing us to calculate its y_{scale}.

The jet p_T, m_{jet}, and y_{scale} are then compared jet-by-jet to the corresponding observables in jets

[32]Contributed by: P. Loch, P. Francavilla

which have been reconstructed using only the particles generated by the hard scattering process. Naming $O_{\mathrm{HS+PU}}$ the observables in the presence of the pile-up and O_{HS} the observables with only the hard scatter particles, the ratio $O_{\mathrm{HS+PU}}/O_{\mathrm{HS}}$ is shown in the following plots as a function of the number of pile-up interactions (N_{vtx}) generated in the event. To properly study these ratios, we should start the jet clustering using only the particles generated by the hard scatter. In this paper we report a different way of defining the jets. We use all particles (hard scatter and pile-up) to reconstruct the jets and then we look at the constituents of the leading jet. From the list of constituents, we use the particles generated by the hard scatter to build the reference jet (and to calculate O_{HS}). This final reference jet is different from the jet reconstructed using only the hard scatter particles, but due to the peculiarities of the Anti-$k_{\perp}$ jet algorithm, we can expect a rather small difference introduced by this method.

The observables $O_{\mathrm{HS+PU}}$ depend on the number of pile-up interactions. To suppress the effect of the pile-up, which usually adds a set of low p_T particles to the hard scatter, we used a threshold on the particle p_T, thus eliminating low p_T particles. This method to suppress the effect of the pile-up is approximate for jets reconstructed in the calorimeter, but gives us an estimate of how a selection on the constituents of the jets, aimed to suppress contributions from pile-up, could change the dependence of the observables as a function of the number of pile-up interactions. We applied four different p_T thresholds, namely: none, 0.5 GeV, 1 GeV, and 2 GeV. The order of magnitude of this selection is similar to the effective cut-off that affect the charged particles in reaching the calorimeter due to the bending of the magnetic field, as in the ATLAS detector [253] at LHC, and the typical minimum energy required in the central region to generate a signal in the calorimeter at all ($p_T = 0.4$ GeV).

These cuts are applied in two ways. The goal of the first one is the estimate of the effect of a pile-up suppression after the jet reconstruction. For this purpose the cut is applied to the list of constituents of the jet, after its reconstruction. The second way emulates a signal selection prior of the jet reconstruction in that the cut is applied before the jet reconstruction. We refer to the first method as an inclusive selection and to the second as the exclusive selection. Due to the difference in the clustering, the reference is different in the two cases.

In the following section the results for the different observables are shown.

17.3 Results

17.31 Inclusive selection

In this section we show some of the results obtained using the inclusive selection discussed above. The results are divided in different plots according to the windows in $\hat{p}_{\mathrm{T}}$ for the generation of the event. In Fig.49 the dependence of $p_{\mathrm{T,HS+PU}}/p_{\mathrm{T,HS}}$ as a function of number of pile-up interaction (N_{vtx}) is shown. The different colors show the dependence for the selection cuts used in the analysis according to the following convention: red for none, blue for 0.5 GeV, green for 1 GeV and black for 2 GeV. The different plots show that the variation as a function of $N_{\mathrm{vtx}}$is smaller for bigger values of the threshold, but introduces a bias especially for low p_T (17-34 GeV). The 0.5 GeV and 1 GeV thresholds seem to be the more appropriate, because they reduce the variation on the number of pile-up interactions to less than 1% while introducing a bias smaller than 1% in the jet energy scale for jets with $p_T > 280$ GeV. The 2 GeV threshold suppresses the dependence on N_{vtx}, but the bias is bigger than 1% in some region of p_T and a corresponding correction of the jet energy scale would be needed to reach an accurate calibration.

In Fig.50 the variations of the jet mass m_{jet} are shown in various kinematic bins. These plots show similarities with respect to the figure 49. Even in this case, a selection of the constituents reduces the impact of the pile-up interactions. The variation due to pile-up in absence of any constituent p_T thresholds cuts is of the order of 20% even for the high p_T jets, indicating the need for a selection. Again, thresholds of 0.5 GeV and 1 GeV are the most appropriate, reducing the pile-up dependence to 10% and 5%, respectively, while introducing a respective bias of the order of 5% and 10%. The threshold of 2 GeV, which reduces significantly the pile-up variation from 20% to almost none, introduces a 20% bias.

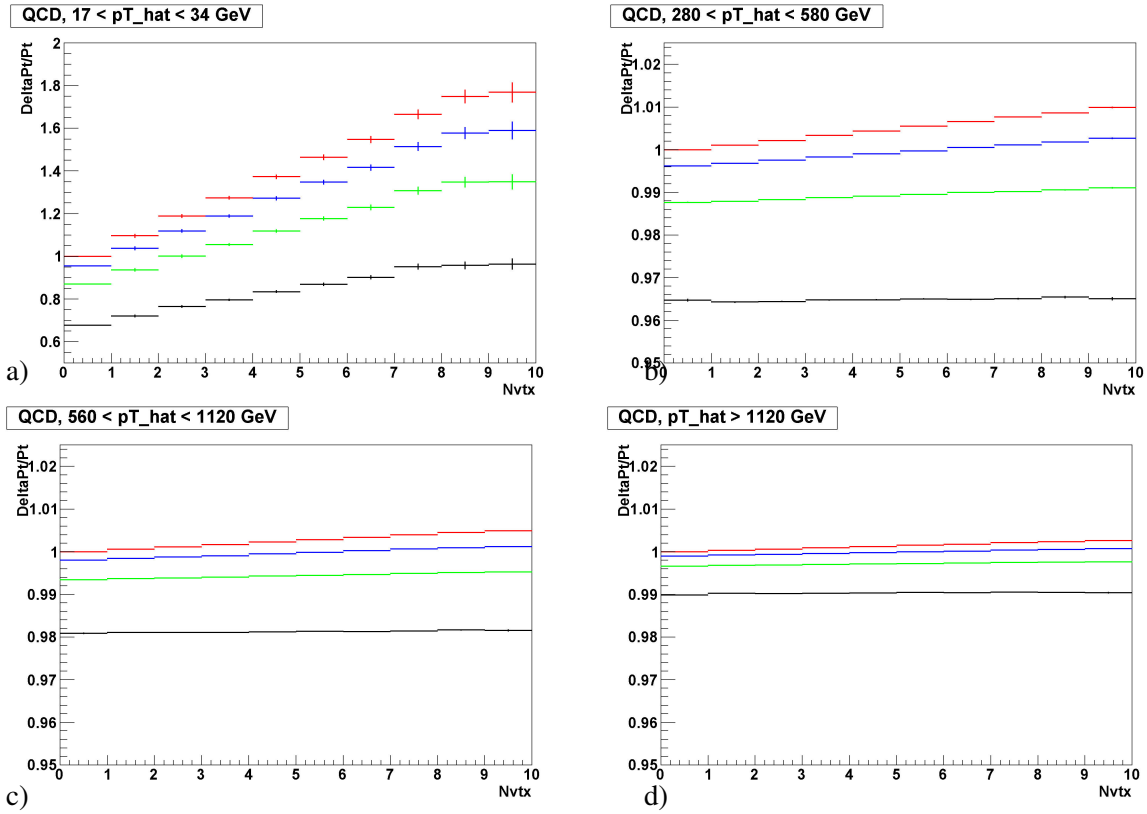


Fig. 49: Inclusive selection: comparison of the ratios of $\Delta p_T/p_T = (p_{\mathrm{T,HS+PU}} - p_{\mathrm{T,HS}})/p_{\mathrm{T,HS}}$ as a function of the number of pile-up interactions for different windows of $\hat{p}_{\mathrm{T}}$. Here red indicates no constituent p_T threshold, while blue/green/black indicate thresholds of 0.5/1.0/2.0 GeV, respectively.

A similar behavior is shown in figure 51 for y_{scale}. In this case, the 1 GeV threshold shows a good suppression of the pile-up dependence, together with a bias of less than 5%.

17.32 Exclusive selection

In the case of the exclusive selection, a particle selection cut using the p_T thresholds is applied prior to jet reconstruction. In this case, the constituents differ from the inclusive selection even for the reference. The purpose of this part is the evaluation of an intrinsic selection due to instrumental effects, such as already mentioned bending of charged particle tracks in the magnetic field and the effect of inactive material in front of the calorimeter. As for the inclusive selection, we expect a dependence of the observables on the number of pile-up interactions N_{vtx}. This dependence is reduced by applying the p_T based selections.

A comparison of the inclusive and exclusive selection is shown in Fig.52. This plot shows the ratio $p_{\mathrm{T,HS+PU}}/p_{\mathrm{T,HS}}$ for the two selections for different p_T thresholds. Suppressing the low p_T particles already in the reference jet, the bias is no more visible, and only the residual dependence on the pile-up activity is shown. This dependence is smaller than the dependence shown in the inclusive plot.

The results for the jet mass m_{jet} and the y-scale y_{scale} are shown in Fig. 53. These four plots show that the variations of m_{jet} and y_{scale} are very similar, and 1 GeV or bigger p_T threshold applied to the final state particles is useful to reduce this dependence.

17.4 Conclusion

The studies described in this paper show the dependence of some of the jet substructure observables on the pile-up activities. The variation in the jet transverse momentum is less prominent for high p_T jets, as

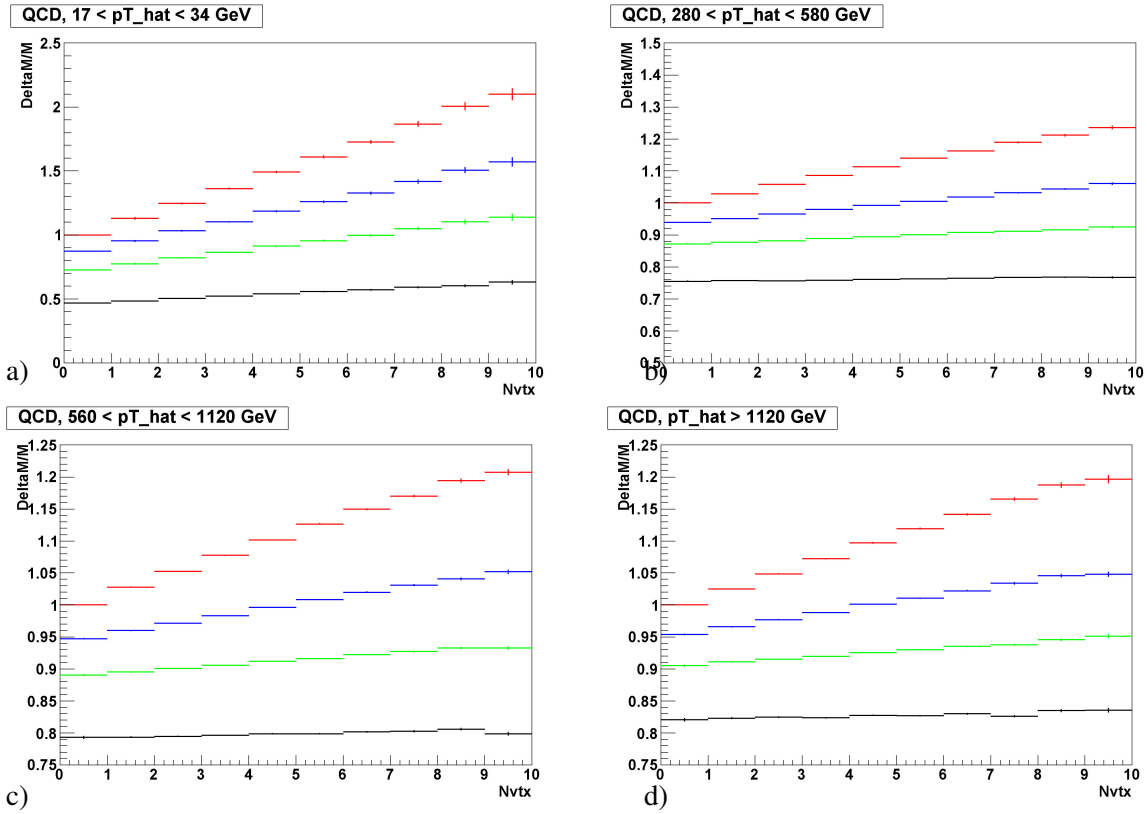


Fig. 50: Inclusive selection: comparison of the ratios of $\Delta m_{\rm jet}/m_{\rm jet} = (m_{\rm jet,HS+PU} - m_{\rm jet,HS})/m_{\rm jet,HS}$ as a function of $N_{\rm vtx}$, for different windows of $\hat{p}_{\rm T}$. Again, red indicates no threshold, while blue/green/black indicate jet constituent p_T thresholds of 0.5/1.0/2.0 GeV, respectively.

expected. In addition, a p_T threshold cut on the constituents seems to be useful to accurately calibrate the jet energy scale for high p_T jets.

For the jet mass and the y-scale, the variation introduced by the pile-up is of the order of 2% per pile-up interaction. If we consider an activity of between 0 to 10 pile-up interactions per signal event, the total variation is of the order of 20%, even for high p_T jets. In this case a pile-up suppression is needed and applying a p_T threshold on the particles seems to reduce the dependence to 5-10% while introducing a small bias of 5-10%.

The effect of the pile-up has to be compared with the experimental capability to reconstruct and calibrate the observables under investigation. This part is beyond the goal of this paper and it could be evaluated in a realistic detector simulation, which takes in account all the detector effects. The effect of the pile-up and its suppression is an important ingredient to be taken into account when using complex QCD observables like jet masses and substructure in an LHC analyses.

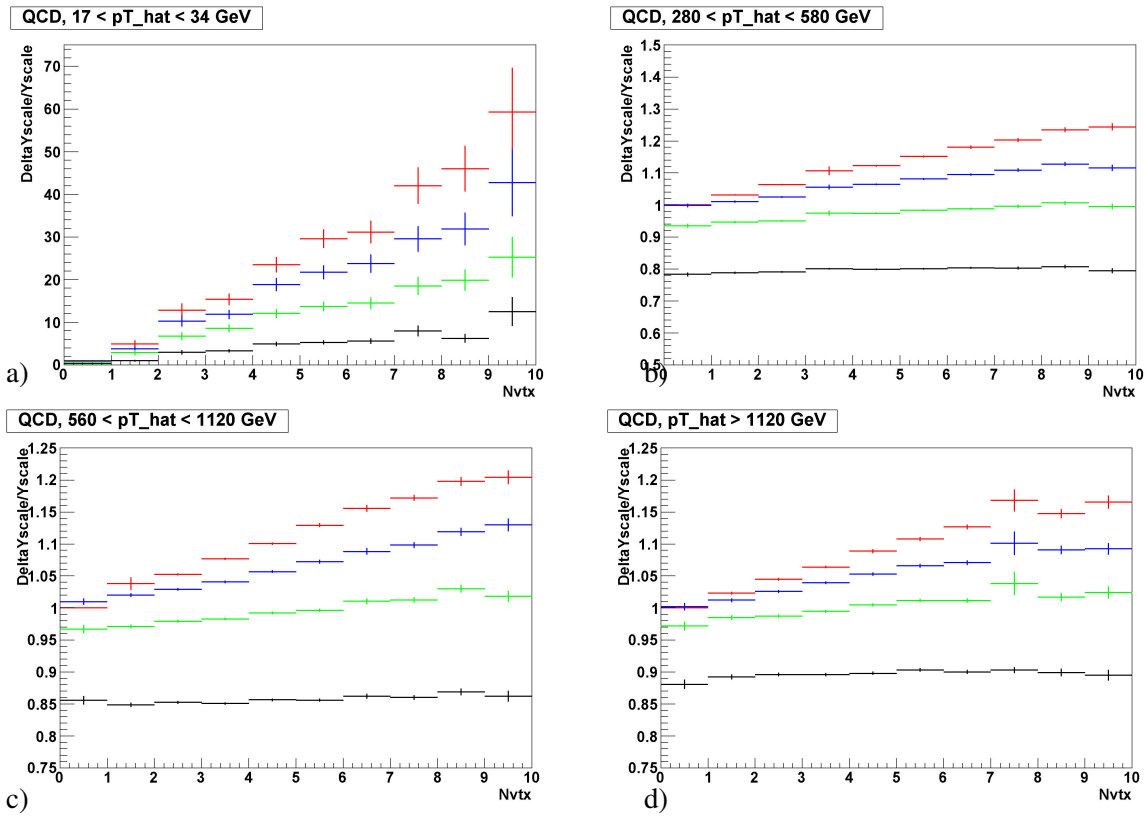


Fig. 51: Inclusive selection: comparison of the ratios of $\Delta y_{\text{scale}}/y_{\text{scale}} = (y_{\text{scale,HS+PU}} - y_{\text{scale,HS}})/y_{\text{scale,HS}}$ as a function of $N_{\text{vtx}}$for different windows of $\hat{p}_{\text{T}}$(red: no selection, blue/green/black 05/1.0/2.0 GeV thresholds).

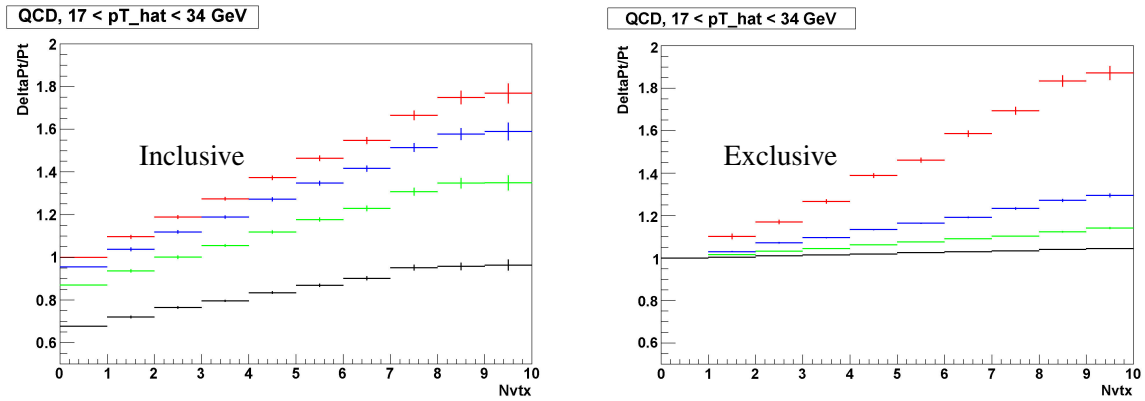


Fig. 52: Comparison of the effect of exclusive and inclusive selections: the ratios of $\Delta p_T/p_T = (p_{\text{T,HS+PU}} - p_{\text{T,HS}})/p_{\text{T,HS}}$ are shown as a function of the number of pile-up interactions N_{vtx}. The red points indicate no threshold, while blue/green/black correspond to 05/1.0/2.0 GeV thresholds, respectively.

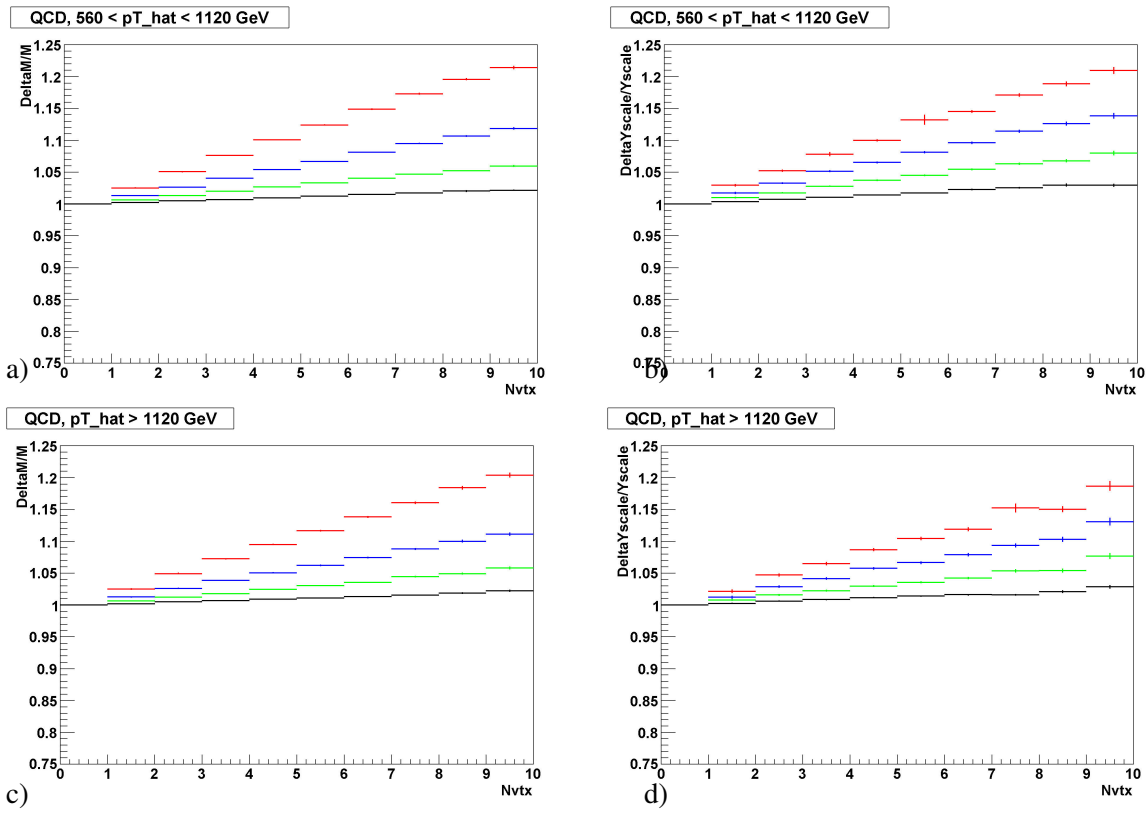


Fig. 53: Exclusive selection: comparison of the ratios of $\Delta m_{\rm jet}/m_{\rm jet} = (m_{\rm jet,HS+PU} - m_{\rm jet,HS})/m_{\rm jet,HS}$ and $\Delta y_{\rm scale}/y_{\rm scale} = (y_{\rm scale,HS+PU} - y_{\rm scale,HS}/y_{\rm scale,HS}$ as a function of $N_{\rm vtx}$, for different windows of $\hat{p}_{\rm T}$ (red: no selection, blue/green/black 0.5/1.0/2.0 GeV p_T thresholds).

18. A STUDY OF RADIATION BETWEEN JETS AT THE LHC [33]

18.1 INTRODUCTION

The pattern of radiation between jets (or between jets and the proton remnant) ought to be very different depending upon whether the exchanged state is a colour singlet or colour octet object. Particles are expected to flow differently in ϕ regions both close to and far from the jets. The typical $\eta - \phi$ scale of the radiation patterns is also relevant. Traditionally, either due to uncertainties in the theoretical predictions or because of the poor granularity of calorimeters used in hadron colliders, few of the fine details of radiation have been studied. Events generated by colour singlet or colour octet exchange have therefore been distinguished in the past by applying either a cut on the total sum of transverse momentum between the two jets [254], or on the transverse momentum of a third jet between the two [255]. The presence of Underlying Event (UE) and especially of pile-up in hadron collisions will give rise to a very small efficiency for finding events with a clean gap between the jets. The efficiency may also be very sensitive to the details of the specific models. Radiation between jets has to be studied in data, using the most powerful techniques to separate the various contributions, and possibly reduce the effect of UE or pile-up. Here we present three different classes of observables that can be used for such studies: the distribution of radiation outside jets as a function of distance from the jet axis, using different jet algorithms; the "gap grid" method, consisting of dividing the detector in several parts along η and Φ to exploit the different space distribution of QCD radiation; and the application of a one-dimensional Fourier transformation to the energy deposition in the event to highlight specific structures characterised by a given size.

18.2 RADIATION DISTRIBUTION CLOSE TO THE JET BOUNDARIES

As a first method, we studied the radiation distribution from events generated with Herwig 6 [17] close to the jet boundaries, for various jet algorithms. This shows how the various algorithms differ for average QCD events at the LHC. Figure 54 shows the Et-weighted distribution of the ΔR distance of truth-level particles from the axis of the leading jet. Using different algorithms with radius (or measure) 0.4, figure 54a shows the distribution of particles belonging to the jet, while 54b) shows particles outside of the jet. The Anti-$k_\perp$ algorithm shows a clear cut-off at the nominal value of the jet radius, as expected by its definition, while the other algorithms have a smoother behaviour. In particular the $k_\perp$ algorithm has the largest amount of radiation outside the nominal jet size.

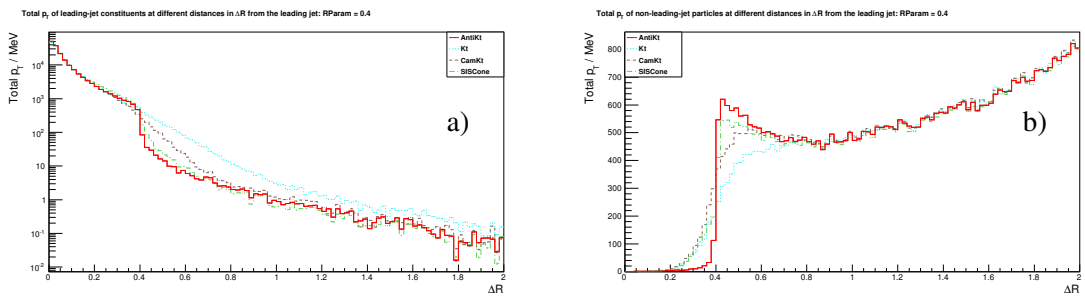


Fig. 54: Energy flow for QCD colour octet events at a distance ΔR from the jet axis using a) particles that are constituents of the jet and b) particles not part of the jet.

Figure 55 shows the same plots for events in which a colour singlet is exchanged between the two leading partons. Note that whilst 55a is broadly similar to 54a, 55b shows less than half the activity of 54b for particles that are not part of the jet. Figure 55a in fact shows a slightly less sharp jet border at R=0.4 for colour singlet exchange because the generally lower amount of radiation in colour singlet events means that, in order to pass the jet E_T cut, radiation must be drawn into the jet from a wider area. These characteristics will be exploited in the techniques described in the next sections.

[33] Contributed by: M. Campanelli, J. Monk, J. Robinson, C. Taylor

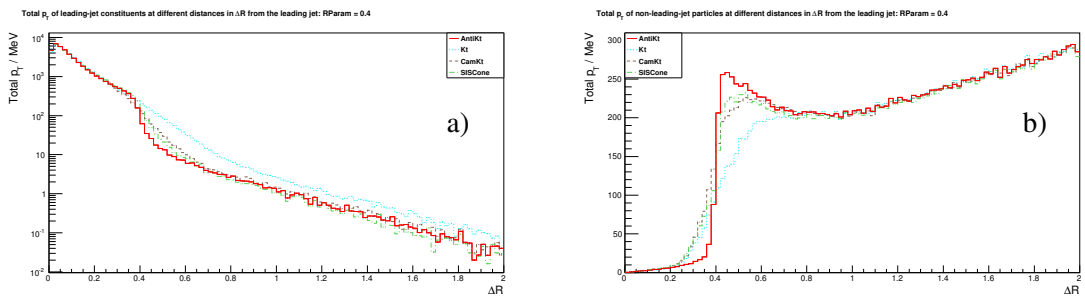


Fig. 55: Energy flow for colour singlet exchange events at a distance ΔR from the jet axis using a) particles that are constituents of the jet and b) particles not part of the jet.

18.3 GAP GRID TECHNIQUE

A method often used to distinguish events due to colour singlet and colour octet exchange is looking at the total transverse energy in the rapidity range between two jets. Events due to colour-singlet exchange (no colour flows between the jets) are characterised by almost no radiation, their golden signature being an empty rapidity gap between the two jets.

It is easy to understand why this quantity is too inclusive to be optimal. Hard radiation from the jet itself can emit particles outside the jet boundaries independently of the type of event. On the other hand, particles emitted by a colour connection between the jets tend to follow specific geometrical distributions, for instance being quite uniform in rapidity and close in ϕ to the main jets. To study these differences, we divide the detector into various regions according to the directions of the two jets with largest E_t. First, the area with $|\Delta\eta| < 0.7$ with respect to the axis of the first two jets is removed. The remaining area between the jets is divided into 16 regions of equal size in the $\eta - \phi$ plane: four intervals of equal $\Delta\eta$ (each one quarter of the eta difference between the two main jets minus $2 * 0.7$) and four of equal $\Delta\Phi$ (each of 90 degrees), with the $\Phi = 0$ angle defined as the region of the leading jet in the event, and the first region being comprised between $-45°$ and $45°$. The two detector areas between the jets and the beam pipe are also divided in four regions each, according to the same $\Delta\Phi$ convention.

The convention is that the regions between the two jets are labeled between 1 and 16, with regions 1 to 4 being the closest in eta to the main jet (and ordered such that region 1 contains $\Phi = 0$, region 3 $\Phi = \pi$, region 2 is the transverse one closer to the axis of the second jet and region 4 the remaining one) 5 to 8 for the next η region in the direction of the second jet, with the same convention with respect to $\Delta\Phi$, etc.

The average transverse energy deposition in each of the detector areas is shown in Fig. 56a) for Herwig 6 colour singlet and colour octet events, the former shown with the Jimmy UE model [54] switched both on and off. A clear difference between the various cases is visible, but what is more important is that this difference is more marked in some regions than in others. For instance, and despite the cut at $\pm$ 0.7, all three event classes show that both η regions close to the jets receive more radiation than the central regions. This is due to the splash-out from the jets themselves; the two regions with similar ϕ to the main jets also have more radiation - an effect that is more pronounced for the events with colour octet exchange. Dividing the rapidity gap between the two main jets into several smaller regions not only allows a better understanding of the radiation patterns (for instance, to compare with different Monte Carlo models), but also provides a more powerful separation between singlets and octets.

The presence of UE is a strong nuisance factor: without it some regions would almost be completely devoid of radiation in the case of colour singlet exchange, which would make the identification of such events very easy. Although there is no obvious way to identify radiation due to the UE (it is not even a fully well-defined concept), it is on average softer than the main scattering process. A possibility to exploit this fact would simply be to impose a high threshold for clusters to be included in the density calculation; however, it was shown in [256] that a better approach would be to establish this threshold

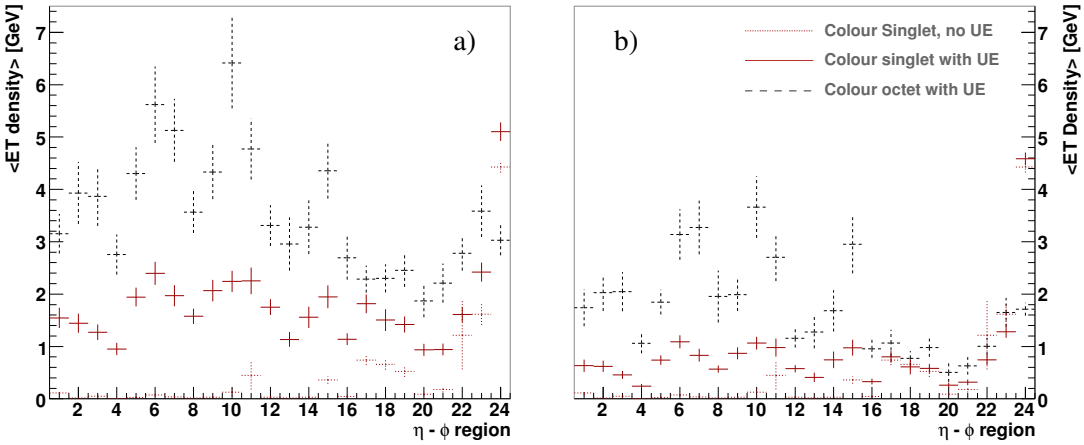


Fig. 56: The average E_T density in 24 different $\eta - \phi$ regions. Regions 1 to 16 lie between the η values of the leading two jets. Jets with colour octet exchange and UE present are shown as black dashed crosses. Jets with colour singlet exchange are shown in red; solid data-points are events with UE present and dotted data-points have UE turned off. a) on the left shows the E_T density uncorrected for UE, b) on the right shows the effect of applying a correction for UE based upon the area of a jet and the average E_T density per event.

on an event-by-event basis. Following this idea, we use the $k_\perp$ algorithm with a small Et threshold to cluster into jets all energy depositions in the events. For each jet we calculate the active area, then define an E_T density as the ratio of the jet transverse momentum and its area. All jets below 10 GeV per unit of rapidity squared are considered potential soft candidates; their average density on an event-by-event basis is taken as the value of the soft density for the event. For all jets in the event an UE contribution is calculated as the product of this soft density and the area of each jet, and subtracted from the jet transverse momentum. Finally, all jets whose transverse momentum after this subtraction is below 3 GeV are considered as coming from a soft UE-type interaction and are therefore removed. The advantage of this method with respect to applying a strong cut on all jets is that the amount of soft radiation is determined on an event-by-event basis, so coherent upward or downward fluctuations of the soft activity can be accounted for. After this procedure, the density distribution for the various detector regions is shown in Fig.56b). An improvement with respect to the previous case is visible, with the additional advantage of a smaller dependence on the MonteCarlo description of the soft interactions.

18.4 FOURIER TRANSFORMATION

The problem of separating the colour connection effect from the UE or pile-up is one of separating features of differing physical size in the event. The UE fills the whole detector with radiation at all η. Colour connection effects between jets are typically smaller than this size, i.e. are approximately the size of the jet-jet or jet-beamline interval. Hadronisation and showering effects can be expected to be of a similar to smaller size. The hard jets will be smaller still, with a radius of $R \simeq 0.5$ and there may be jets originating from softer partons with R as small as 0.1.

A Fourier-transformation of the spatial distribution of radiation in the event could separate the various scales, allowing an alternative reading of the energy flow. For the 1-dimensional case described here, we use the Φ distribution of the truth-level particles for each event, weighted by transverse energy summed in 32 bins covering the region 0-π. The axis of the hardest jet is taken as as the $\Phi = 0$ direction. We then extract and use as our variables the first 32 complex coefficients of the Fourier expansion of this distribution. Since these coefficients are linked by the relation $C_n = C^*_{N-n}$ (where N=32), there are only

32 independent real coefficients, indicating no information is lost (or gained!) on the event-by-event basis in going from the 32 bins to the 32 coefficients.

To test the effectiveness of this method, we applied it to the problem of separating colour singlet and octet exchange in hadronic collisions. Colour singlet exchanges without UE are possibly the most di-jet-like kind of events in a hadron collider. Since the $\Phi = 0$ direction is aligned to the axis of the leading jet, most of the remaining radiation will be concentrated around $\Phi = \pi$, which is the most likely location of the second jet. It is easy to show that odd coefficients can never have a peak at both 0 and π. Further, the nth coefficient corresponds to correlated activity of size $r \simeq \pi/n$. Small-n odd coefficients, corresponding as they do to large non-di-jet features, are therefore expected to be suppressed in perfect di-jet events. This is exactly what is observed in Fig. 57, where the average magnitude of the various coefficients is plotted for the two extreme cases of a set of colour-singlet exchange without the UE, and colour octet exchange with UE.

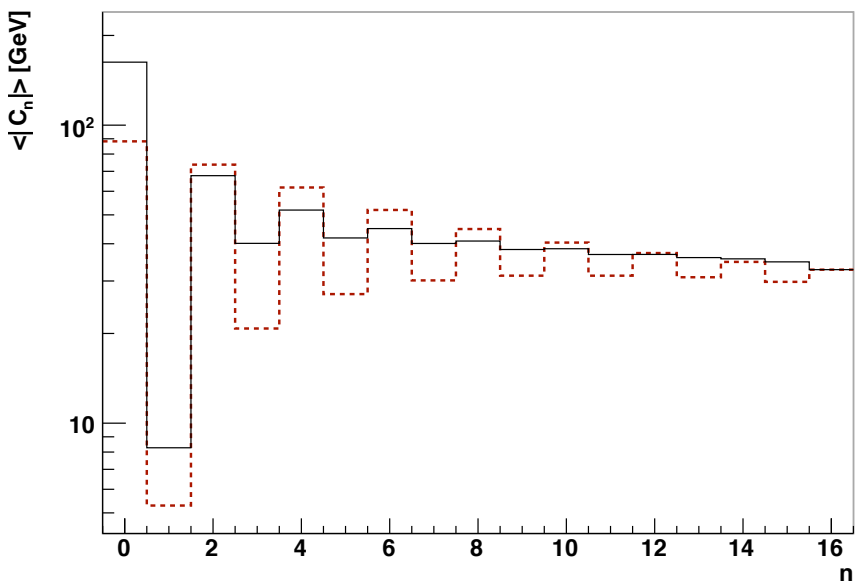


Fig. 57: The average magnitude of the first 16 Fourier coefficients for colour singlet without UE (red-dashed) and colour octet with UE (black-solid).

The suppression of the odd coefficients is clearly visible in the first case, and is much less pronounced (even if still present, since the events nevertheless contain di-jets) for the colour octet exchange case. Although there is also a difference between the even coefficients (especially at lower-n), it is much less pronounced than for the odd coefficients. In particular, we expect coefficients 3 and 5 to originate from structures of a similar size ($\pi/3 \simeq 1$) to the inter-jet radiation. Adding the UE does not change this picture significantly, since the uniform UE radiation will only influence the magnitude of coefficient zero, leaving the mean values of the others almost unchanged. This is shown more clearly in Fig.58, which shows the distribution of the magnitude of the 3rd coefficient for colour singlet and colour octet events. There is a clear difference between the two types of event and the presence or absence of UE makes little difference to the shape of the distribution from colour singlet events.

Another potentially interesting variable to look at is the number of the largest coefficient for each event, which shows the scale of the dominant feature in the event. Figure 59 shows the distribution of the largest imaginary coefficient in the event, for the two classes of colour exchange, both with UE on.

While for ordinary QCD events this distribution is peaked at low values, this is not the case for colour singlet events, where there is a peak around coefficient 8 that corresponds to radiation concentrations of size $\pi/8 \simeq 0.4$, roughly the size of a jet. Further, in the colour singlet sample there is a relative depletion

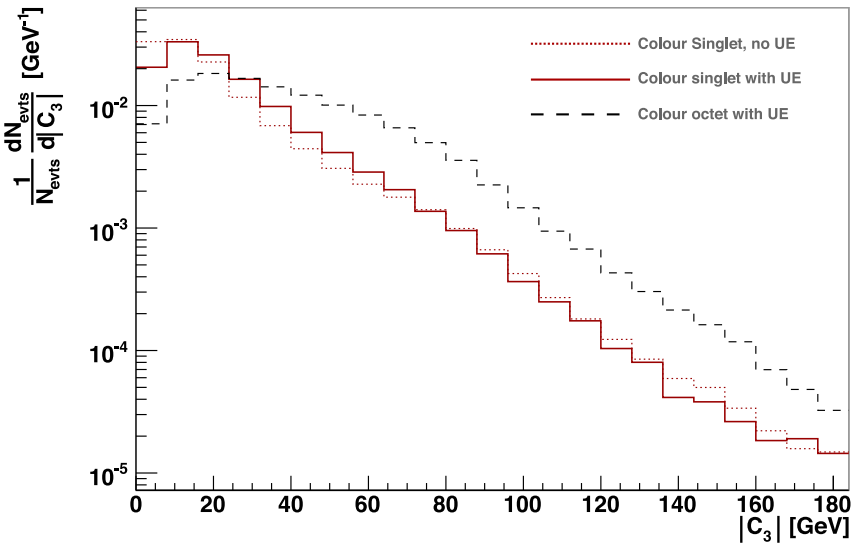


Fig. 58: Distribution of the magnitude of the 3rd coefficient for colour singlet without UE (red-dotted), colour singlet with UE (red-solid) and colour octet with UE (black-dashed). The addition of UE makes little difference to the colour singlet distributions, whilst there is a clear difference between colour octet and colour singlet.

of events in which the largest coefficient was either 3 or 5. More details on the Fourier transformation method can be found in [257].

CONCLUSIONS

Radiation between jets is a key topic for understanding the interplay between hard and soft QCD, and will be the object of many measurements with the first LHC data. Here we propose new methods to study this radiation in more detail than has been done in the past. The first measurements of these observables will offer interesting handles on the discrimination of the various theoretical models. To show the strong and weak points of the methods we applied them to the problem of separation between colour singlet and colour octet exchange events, but the techniques presented here are more general, and can be applied to most of the fully-hadronic final states produced by the first LHC data.

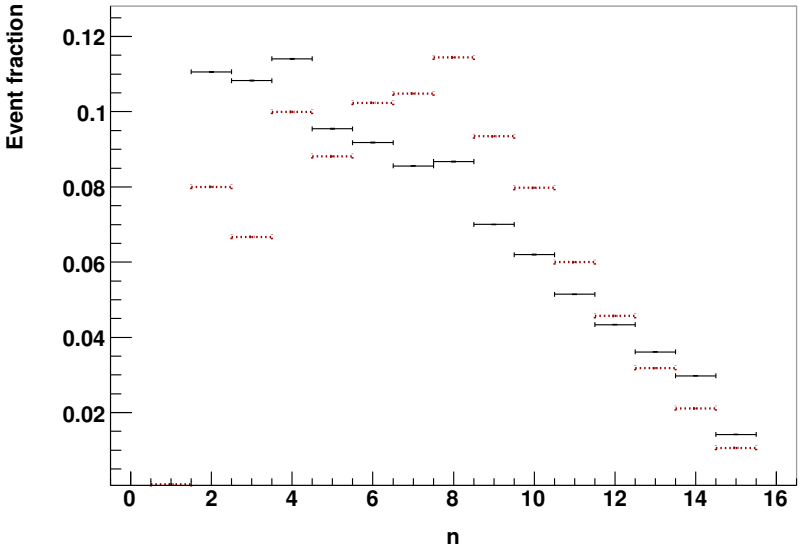


Fig. 59: The largest imaginary coefficient in the event. Solid black shows colour-octet exchange events and dotted-red shows colour singlet exchange. Both event samples have the underlying event model turned *on*. The colour octet sample shows a peak towards lower n, indicating more events have their inter-jet activity dominated by features that are large in $\eta - \phi$. The colour singlet sample, on the other hand, shows more events whose inter-jet radiation is dominated by features of size $r \simeq 8/\pi$.

Part VI
BEYOND THE STANDARD MODEL

19. AN UPDATE OF THE PROGRAM HDECAY[34]

19.1 INTRODUCTION

The search strategies for Higgs bosons at LEP, Tevatron, LHC and future e^+e^- linear colliders (LC) exploit various Higgs boson decay channels. The strategies depend not only on the experimental setup (hadron versus lepton colliders) but also on the theoretical scenarios: the Standard Model (SM) or some of its extensions such as the Minimal Supersymmetric Standard Model (MSSM). It is of vital importance to have reliable predictions for the branching ratios of the Higgs boson decays for these theoretical models.

The current version of the program HDECAY [258] can be used to calculate Higgs boson partial decay widths and branching ratios within the SM and the MSSM and includes:

– All decay channels that are kinematically allowed and which have branching ratios larger than 10^{-4}, i.e. the loop mediated, the three body decay modes and in the MSSM the cascade and the supersymmetric decay channels [111, 259, 260].

– All relevant higher-order QCD corrections to the decays into quark pairs and to the loop mediated decays into gluons are incorporated [261].

– Double off–shell decays of the CP–even Higgs bosons into massive gauge bosons which then decay into four massless fermions, and all important below–threshold three–body decays [262].

– In the MSSM, the complete radiative corrections in the effective potential approach with full mixing in the stop/sbottom sectors; it uses the renormalization group improved values of the Higgs masses and couplings and the relevant next–to–leading–order corrections are implemented [263–266].

– In the MSSM, all the decays into SUSY particles (neutralinos, charginos, sleptons and squarks including mixing in the stop, sbottom and stau sectors) when they are kinematically allowed [267–269]. The SUSY particles are also included in the loop mediated $\gamma\gamma$ and gg decay channels.

The source code of the program, written in FORTRAN, has been tested on computers running under different operating systems. The program provides a very flexible and convenient use, fitting to all options of phenomenological relevance. The basic input parameters, fermion and gauge boson masses and their total widths, coupling constants and, in the MSSM, soft SUSY-breaking parameters can be chosen from an input file. In this file several flags allow switching on/off or changing some options [*e.g.* choosing a particular Higgs boson, including/excluding the multi-body or SUSY decays, or including/excluding specific higher-order QCD corrections].

19.2 UPDATES

Since the release of the original version of the program several bugs have been fixed and a number of improvements and new theoretical calculations have been implemented. The following points summarize the most important modifications of HDECAY after its release:

– Implementation of Higgs boson decays to gravitino + gaugino [270]. A flag in the input file controls whether these decay modes are taken into account or not.

– Inclusion of SUSY–QCD corrections in neutral MSSM Higgs decays to $b\bar{b}$ [271, 272] and resummation of Δ_b effects [273, 274] up to NNLO [275, 276]. The corresponding Δ_b terms have also been included in charged Higgs decays $H^\pm \to tb$.

[34] Contributed by: A. Djouadi, J. Kalinowski, M. Mühlleitner and M. Spira

– Determination and inclusion of the RG improved two-loop contributions to the MSSM Higgs self-interactions. These two-loop corrections extend the results of Ref. [263] for the stop and sbottom contributions for arbitrary mixing parameters and mass splitting, where they have only been displayed for the scalar $\mathcal{CP}$-even Higgs mass matrix. We have checked explicitely that we reproduce the results of Ref. [263].

– An interface to the SUSY Les Houches Accord (SLHA) [43, 44] has been implemented properly. This required several transformations of the corresponding renormalization schemes to the ones used by HDECAY. This option can be switched on and off by appropriate flags in the input file. The output file in the SLHA format can also be used again as input file. Moreover, the automatic generation of the SLHA input file according to the input parameters of HDECAY is provided as an option.

– Inclusion of the full mass dependence of the NLO QCD corrections to the quark and squark loop contributions to photonic Higgs decays $\phi \to \gamma\gamma$ [277–282]. This decay width can now also be used to determine the production cross sections of Higgs bosons at photon colliders at NLO QCD.

– Inclusion of electroweak corrections to the SM Higgs boson decays $H \to W^{(*)}W^{(*)}/Z^{(*)}Z^{(*)} \to 4f$

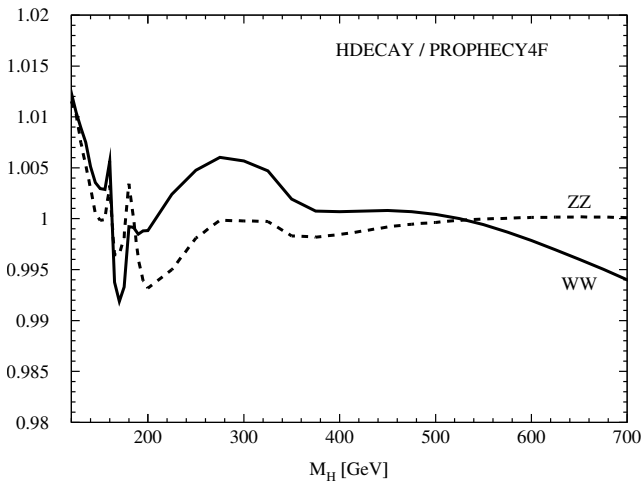


Fig. 60: Ratio of the partial widths of $H \to W^{(*)}W^{(*)}/Z^{(*)}Z^{(*)} \to 4f$ from HDECAY and PROPHECY4F [283, 284].

in approximate form which reproduces the full results of Refs. [283, 284] within 1% as can be inferred from Fig. 60, where the ratios of the partial decay widths into $W^{(*)}W^{(*)}$ and $Z^{(*)}Z^{(*)}$ are shown as a function of the Standard Model Higgs mass. The improved Born approximation of Ref. [283, 284] has been implemented with additional improvements at the WW and ZZ thresholds.

The logbook of all modifications and the most recent version of the program can be found on the web page `http://people.web.psi.ch/spira/proglist.html`.

ACKNOWLEDGEMENTS

This work is supported in part by the European Community's Marie-Curie Research Training Network HEPTOOLS under contract MRTN-CT-2006-035505.

20. IMPLEMENTATION AND VALIDATION OF MODELS BEYOND THE STANDARD MODEL WITH FEYNRULES [35]

20.1 INTRODUCTION

Monte Carlo event generators play an important role in making reliable predictions for the events to be observed at collider experiments, both to describe the backgrounds and possible candidate signals. In particular, the simulation of a hadronic collision requires not only an accurate description of the underlying hard scattering process, but also of subsequent parton showering and hadronization as efficiently provided by programs such as HERWIG [21,136], PYTHIA [18,20] and SHERPA [22,23]. As regards the generation of the hard matrix element itself, a lot of effort has gone into the development of several multipurpose matrix element generators, such as COMPHEP/CALCHEP [285–287], MADGRAPH/MADEVENT [80,81,288,289], SHERPA and WHIZARD [211, 290]. Even though these programs are in principle able to generate the (tree-level) matrix element for any process in the framework of a given renormalizable quantum field theory built on scalar, vector and fermion fields, the implementation of a full Beyond Standard Model (BSM) theory can be a tedious and error-prone task, often requiring the implementation of one vertex at the time following the conventions specific to each code.

In this section a summary report on the implementation and validation of BSM models into multipurpose matrix element generators is presented. The starting point of our approach is FEYNRULES [291], a MATHEMATICA®[36] package that allows to compute Feynman rules from a Lagrangian in an automated way. Furthermore, FEYNRULES contains a set of interfaces to various matrix element generators, allowing to implement the model into a given tool in an automated way. For the moment, interfaces to COMPHEP/CALCHEP, FEYNARTS/FORMCALC [292–297], MADGRAPH/MADEVENT and SHERPA are available[37]. A first step in the direction of deriving Feynman rules automatically starting from a model Lagrangian has been made in the context of the COMPHEP/CALCHEP event generator with the LANHEP package [298]. Our aim is to go beyond this scheme and create a general and flexible environment where communication between theorists and experimentalists in both directions is fast and robust. First, the use of MATHEMATICA as a working environment provides a powerful and user-friendly platform where BSM models can be developed and implemented. Second, the possibility to export the model to more than one matrix element generator enhances the chances that the model can be successfully dealt with by more than one code. Furthermore, since many of these matrix element generators are already embedded in the experimental softwares, the new models can easily be integrated into the experimental framework, allowing in this way for an efficient communication between theorists and the experimental community.

As the starting point of our approach is FEYNRULES, we briefly recall its basic features in Section 20.2 and discuss recent developments triggered by the activities and the discussions at the Les Houches workshop. In Section 20.3 we present a proposal of how new BSM models can be easily validated exploiting the fact that once a FEYNRULES implementation is available, the model can easily be exported to various matrix element generators. In Section 20.4 we discuss the implementation of several new models that were implemented and/or validated as an outcome of the workshop.

20.2 RECENT DEVELOPMENTS IN FEYNRULES

FEYNRULES is a MATHEMATICA package that allows to compute Feynman rules directly from a Lagrangian in an automated way. The user provides the Lagrangian for the model (written in MATHEMATICA) as well as all the information about the particle content and the parameters of the model. This information uniquely defines the model, and hence is enough to derive all the interaction vertices from the Lagrangian. FEYNRULES can in principle be used with any model which fulfills basic quantum field theoretical requirements (*e.g.*, Lorentz and gauge invariance), the only limitation coming from the kinds of fields

[35] Contributed by: L. Basso, F. Braam, C. Duhr, B. Fuks, A. Martin, J. Reuter, T. Roy, S. Schumann

[36] MATHEMATICA is a registered trademark of Wolfram Research, Inc.

[37] An interface to WHIZARD will be available in the near future.

supported by FEYNRULES. As to date, the public release of FEYNRULES supports scalar, vector, fermion (Dirac and Majorana) and spin-two fields, as well as Faddeev-Poppov ghosts, and very recently, also Weyl fermions have been included[38], as will be described at the end of this section. In a second step, the interaction vertices obtained by FEYNRULES can be exported to various matrix element generators by means of interfaces provided by the package [299]. Let us note that even though FEYNRULES itself can in principle be used to obtain the Feynman rules for any Lagrangian, the matrix element generators very often have certain information on the color and/or Lorentz structures hardcoded. In this case the interfaces check whether all the vertices are compliant with the structures supported by the corresponding matrix element generator, and if not, a warning is printed and the vertex is discarded. Each interface produces at the end a (set of) text file(s), often consistently organized in a single directory, which can be read into the matrix element generator at runtime and allows to use the new model in a way similar to all other built-in models.

As already mentioned, the most important recent development concerns the possibility to write Lagrangians in terms of two-component Weyl fermions. In the following we give a brief description how this new feature can be used inside FEYNRULES. Weyl fermions can be declared in the model file in exactly the same way as other particle classes, *e.g.*,

```
W[1] == {
        ClassName     -> chi,
        Chirality     -> Left,
        SelfConjugate -> False},

W[2] == {
        ClassName     -> xibar,
        Chirality     -> Right,
        SelfConjugate -> False}.
```

The particle class `W` refers to Weyl fermions, and we have defined in this way one left-handed and one right-handed Weyl fermions named `chi` and `xi`, respectively. Let us note that all other options allowed for fields are also available for Weyl fermions (See the FEYNRULES manual [300]). Along the same lines, there is a new option for Dirac fermions specifying their left and right-handed Weyl components, *e.g.*,

```
F[1] == {
       ClassName       -> psi,
       SelfConjugate   -> False,
       WeylComponents  ->  {chi, xibar}}.
```

More details on the use of Weyl fermions in FEYNRULES will be presented in a forthcoming publication. Here it suffices to say that since matrix element generators work with four-component spinors rather than Weyl fermions, FEYNRULES replaces at runtime all Weyl fermions by the corresponding Dirac fermion $\psi = (\chi, \bar{\xi})^T$ according to the prescription,

$$\chi \to \frac{1-\gamma^5}{2}\,\psi, \quad \xi \to \frac{1-\gamma^5}{2}\,\psi^c, \quad \bar{\xi} \to \frac{1+\gamma^5}{2}\,\psi, \quad \bar{\chi} \to \frac{1+\gamma^5}{2}\,\psi^c,$$

where $\psi^c = C\,\bar{\psi}^T = (\xi, \bar{\chi})^T$ denotes the charge conjugated Dirac spinor. After these replacements, the Lagrangian is expressed completely in terms of four-component Dirac and Majorana spinors, and the Feynman rules can be exported to the matrix element generators in the standard form.

The two-component formalism we have just described can be extremely useful when dealing with supersymmetric extensions of the SM, where Weyl fermions appear as the natural fermionic degrees

[38] This feature is still beta and will be made public with the next major release.

of freedom of chiral and gauge superfields. Let us note that this is just the first step of a more general project which aims at implementing superfields directly into FEYNRULES. In particular, we applied the Weyl-fermion formalism to rewrite the existing MSSM implementation in FEYNRULES [299, 300] completely in terms of two-component spinors, as well as to the implementation of the minimal R-symmetric supersymmetric extension of the SM described in Section 20.43. The validation of these (and other) new model implementations was done following the procedure described in the next section.

20.3 VALIDATION PROCEDURE FOR BSM MODELS

The requirements for a BSM model implementation that ensure a fast and efficient communication between all the actors involved in the simulation chain, both on the experimental and theoretical side, are manifold. The first obvious requirement is that the implementation produces reliable results when used with a matrix element generator, both from the perspective of the theoretical consistency of the model (*e.g.*, gauge invariance,...), as well as regarding the technical aspects of the matrix element generator under consideration (*e.g.*, conventions used in the code,...). A robust validation procedure must hence combine these two aspects, which however rely on two disconnected fields of expertise: if theoretical consistency is purely related to the physics content of the model, testing the technical aspects of the implementation requires for example detailed knowledge of the programming language used in the matrix element generators. Second, to ensure traceability both of the implemented models as well as of the generated event samples, every implementation of a BSM model into a matrix element generator must be clearly documented so that all the information about the physics content of the model and the choice of the parameters can be retraced at any point in the chain.

In Refs. [210, 299], several BSM implementation were validated by comparing the results obtained with different matrix element generators among themselves. In this section we propose an evolution of this strategy for the validation of a generic BSM model implementation, and we argue that the FEYNRULES approach offers a natural framework where BSM models cannot only be easily developed and implemented, but can also be validated to an unprecedented level. Since the FEYNRULES model files are built around the Lagrangian of the model, all the information about the physics content of the implementation is centered in one place and can be accessed at any point in the simulation chain, without being obscured by the conventions imposed by a specific matrix element generator. Furthermore, one can exploit the fact that FEYNRULES offers the possibility to implement the model into various generators and that different codes use different conventions and/or work in different gauges. In this way one can very easily check the sanity of an implementation by cross-checking the results obtained with different codes. For example, once a FEYNRULES implementation is available, the Feynman rules can be obtained automatically and then used to cross-check analytical results for simple processes against known results from the literature, a step that could even be automatized by using the FEYNRULES interface to FEYNARTS/FORMCALC. After the model has passed successfully all of these tests, a more detailed study of the implementation can be performed by comparing the results of different matrix element generators for a large number of (tree-level) processes. In this way, it is not only very easy to show that the implementation produces reliable results for all matrix element generators involved, but it also provides a very natural and fast way to test the theoretical consistency of the model, by computing the results in different gauges, testing the high-energy behavior of certain processes, *etc.*

To quantify the level to which a given BSM implementation has been validated, we propose in the following a four-step procedure to rate BSM model implementations regarding their validation. In this scheme, every implementation is awarded one credit for each completed step, with a maximum of four for models tested to a very high extent in all matrix element generators. More explicitly, the four steps are:

1. **Documentation [First credit] :** A first credit is awarded if the implementation is documented to ensure traceability and reproducibility. Information about the model should be included in the FEYNRULES model file via the `M$ModelInformation` variable (See the FEYNRULES manual), and should contain a twofold information:

- All references to the original publications where the model was first presented should be included, to ensure that the Lagrangian, its field content and the relations between parameters can be traced back at any time. This also includes references to other codes involved in the writing of the model file, *e.g.*, spectrum generators. In case the implemented model is incomplete or consists only in a theory fragment, this should be clearly stated.
- The operating system (Linux distribution, Mac OS, ...) and the versions of MATHEMATICA and FEYNRULES (as well as all other codes, if relevant) used to write the model file must be provided in order to ensure reproducibility at later times.

2. **Basic theory sanity checks [Second credit] :** A second credit is awarded for testing that the model satisfies the constraints imposed by quantum field theory, *e.g.*, checking the hermiticity and gauge invariance of the implemented Lagrangian. If the Feynman rules for the model are known in the literature, they should be compared to the results obtained by FEYNRULES for this model. Furthermore, simple two-to-two cross sections and/or decay rates can be easily computed either by hand or by means of FEYNARTS/FORMCALC and the corresponding FEYNRULES interface and compared to known analytic results (if available).

3. **Testing one matrix element generator [Third credit] :** After the theoretical sanity of the implementation has been checked, a more detailed validation of the model can be performed by exporting it to a matrix element generator, both to exclude mistakes in the implemented model as well as to ensure that the implementation can be used in a reliable way with a given code.

   - Basic processes like Drell-Yan and Bhabba scattering can be used to check that the matrix element generator produces reliable results for the corresponding cross-sections, followed by a more systematic study of several two-to-two and/or two-to-three key processes of interest in this model. A particular emphasis should be brought to the running of the strong and/or electromagnetic couplings, as well as to the reproduction of known SM cross section results for the sectors of the model unaffected by the BSM physics.
   - In case an independent (partial) implementation of the model is available, a comparison of the two implementations should be performed.
   - The high-energy behavior of the model can be easily checked numerically to ensure that all gauge and unitarity cancellations take place in a numerically stable and reliable way. Note that for such cancellations to take place *exactly* at tree-level, it is required that all particles have zero width, *i.e.*, all widths of all particles should be put to zero while performing these tests.
   - Some matrix element generators, like for example CALCHEP, offer the possibility to choose between different gauges. In this case gauge invariance can be easily tested by running the same processes in different gauges.
   - The outcomes of the previous steps, *i.e.*, the numerical results for the cross sections for all the tested processes, should be summarized and made public together with the model file(s) for future reference. Let us note that in order to ensure reproducibility, it is important to include the center-of-mass energy and all final state cuts, as well as the version number of the matrix element generator involved.

4. **Testing several matrix element generators [4th credit] :** Since FEYNRULES offers the possibility to obtain implementations into several matrix element generators, it is very easy to repeat the previous steps for more than one program. In this way a very sensible validation of the model can be achieved, by exploiting for example the fact that different codes may work in different gauges. Again, to ensure reproducibility, the results should be summarized and made public together with the model.

Process	MG (FR)	MG (ST)	CH (FR)	CH (ST)	SH (FR)	SH (ST)
$e^+ e^- \to Z h^0$	8.787e−3	8.788e−3	8.787e−3	8.787e−3	8.788e−3	8.788e−3
$W^+ W^- \to H^+ H^-$	3.689e−2	3.686e−2	3.685e−2	3.685e−2	3.685e−2	3.685e−2
$u \bar{u} \to \tilde{l}_3^- \tilde{l}_3^+$	2.123e−3	2.123e−3	2.123e−3	2.123e−3	2.123e−3	2.123e−3
$u \bar{u} \to \tilde{u}_5 \tilde{u}_2^*$	6.141e−1	6.142e−1	6.141e−1	6.141e−1	6.141e−1	6.141e−1
$b \bar{b} \to \tilde{\chi}_2^+ \tilde{\chi}_2^-$	9.556e−2	9.545e−2	9.556e−2	9.556e−2	9.555e−2	9.555e−2
$b \bar{t} \to \tilde{\chi}_2^- \tilde{\chi}_3^0$	3.981e−2	3.971e−2	3.977e−2	3.977e−2	3.977e−2	3.977e−2
$Z \gamma \to \tilde{l}_2^- \tilde{l}_2^+$	1.712e−2	1.711e−2	1.712e−2	1.712e−2	1.712e−2	1.713e−2
$g W^- \to \tilde{d}_5 \tilde{u}_5^*$	2.569e−1	2.566e−1	2.566e−1	2.566e−1	2.565e−1	2.565e−1
$\gamma \gamma \to \tilde{\chi}_1^+ \tilde{\chi}_1^-$	6.250e−1	6.263e−1	6.257e−1	6.257e−1	6.257e−1	6.255e−1
$W^- \gamma \to \tilde{\chi}_2^0 \tilde{\chi}_2^-$	5.235e−2	5.235e−2	5.236e−2	5.236e−2	5.238e−2	5.238e−2

Table 7: Cross sections for a selection of production processes in the MSSM scenario SPS 1a. The built-in MSSM implementation in CALCHEP, MADGRAPH and SHERPA are denoted MG (ST), CH (ST) and SH (ST), respectively, while the FEYNRULES-generated ones are MG (FR), CH (FR) and SH (FR). The center-of-mass energy is fixed to 1200 GeV.

The rating scheme we just described will be applied to the model database on the FEYNRULES website. The FEYNRULES model database offers to every user the possibility to make their model files accessible to the community. Every model is assigned a personal web page where the model can be described and all relevant files can be made public for download. By applying the described rating scheme to the model database it will be possible to progressively build a database of robust BSM model implementations that can be used in a large variety of different matrix element generators while still assuring traceability and reproducibility at any time.

Let us conclude this section by illustrating this validation procedure on the example of the FEYNRULES implementation of the MSSM. After having implemented the model in FEYNRULES, we have compared the Feynman rules computed by FEYNRULES to those which can be found in the literature, both for the general MSSM [301,302] and for a constraint MSSM where all the scalar mixings are neglected [303,304], and we have found agreement for all the vertices. Then, we have re-calculated all tree-level squark and gaugino hadroproduction helicity amplitudes in the case of general (and possibly complex) scalar mixing with the help of FEYNARTS/FORMCALC and the corresponding model file generated by FEYNRULES. The results have been compared to the analytical formulas given in Refs. [305,306] and we have found complete agreement. In order to validate the FEYNRULES-generated model files for the various matrix element generators, we have considered the very particular limit of the typical minimal supergravity point SPS 1a [307] and compared the cross section for 626 two-to-two processes in CALCHEP, MADGRAPH and SHERPA to the results obtained by the built-in MSSM implementations in CALCHEP, MADGRAPH and SHERPA. For all the tested processes we have found perfect agreement. Note that MADGRAPH and SHERPA work in unitary gauge exclusively, whereas CALCHEP employs Feynman gauge for QCD processes[39]. In this way we have demonstrated explicitly that our implementation is gauge invariant (at least in the strongly interacting sector). Furthermore, we have payed particular attention to weak boson scattering processes to ensure that all unitarity cancellation at high energy take place correctly. A very short selection of the results we obtained are shown in Table 7. The full list of tested processes, together with the results for the cross sections, is available from the FEYNRULES model database.

[39]Let us note that for the electroweak sector, CALCHEP can in principle work both in unitary or Feynman gauge. At present, our implementation is performed in unitary gauge only for the electroweak sector.

20.4 RECENTLY IMPLEMENTED MODELS

20.41 THE MINIMAL $B-L$ MODEL

In this section we discuss the implementation of the so-called "pure" or "minimal" $B-L$ model (see Ref. [308] for conventions and references) that features vanishing mixing between the two $U(1)_Y$ and $U(1)_{B-L}$ gauge groups. The classical gauge invariant Lagrangian of this model obeys the gauge symmetry,

$$SU(3)_C \times SU(2)_L \times U(1)_Y \times U(1)_{B-L}\,, \tag{43}$$

and can be decomposed as $\mathcal{L} = \mathcal{L}_{YM} + \mathcal{L}_s + \mathcal{L}_f + \mathcal{L}_Y$. The non-Abelian field strengths in $\mathcal{L}_{YM}$ are the same as in the SM whereas the Abelian ones can be written as follows,

$$\mathcal{L}_{YM}^{\text{Abel}} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} - \frac{1}{4}F'^{\mu\nu}F'_{\mu\nu}\,, \tag{44}$$

where $F_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$ and $F'_{\mu\nu} = \partial_\mu B'_\nu - \partial_\nu B'_\mu$. In this field basis, the covariant derivative reads,

$$D_\mu = \partial_\mu + ig_S T^\alpha G_\mu^{\ \alpha} + igT^a W_\mu^{\ a} + ig_1 Y B_\mu + i(\widetilde{g}Y + g'_1 Y_{B-L})B'_\mu\,. \tag{45}$$

For the "pure" or "minimal" $B-L$ model the condition $\widetilde{g} = 0$ holds, implying that there is no mixing between the $B-L$ Z' and SM Z gauge bosons.

The fermionic Lagrangian (where k is the generation index) is given by,

$$\begin{aligned}\mathcal{L}_f \;=\; & \sum_{k=1}^{3}\Big(i\,\overline{q_{kL}}\gamma_\mu D^\mu q_{kL} + i\,\overline{u_{kR}}\gamma_\mu D^\mu u_{kR} + i\,\overline{d_{kR}}\gamma_\mu D^\mu d_{kR} + \\ & + i\,\overline{l_{kL}}\gamma_\mu D^\mu l_{kL} + i\,\overline{e_{kR}}\gamma_\mu D^\mu e_{kR} + i\,\overline{\nu_{kR}}\gamma_\mu D^\mu \nu_{kR}\Big)\,,\end{aligned} \tag{46}$$

where the charges of the fields are the usual SM and $B-L$ charges (*i.e.*, $B-L = 1/3$ for quarks and -1 for leptons with no distinction between generations, hence ensuring universality). The $B-L$ charge assignments of the fields as well as the introduction of new fermionic right-handed heavy neutrinos (ν_R's) and a scalar Higgs field (χ, charged $+2$ under $B-L$) are designed to eliminate the triangular $B-L$ gauge anomalies and to ensure the gauge invariance of the theory (see Eq. (49)), respectively.

The scalar Lagrangian reads,

$$\mathcal{L}_s = (D^\mu H)^\dagger D_\mu H + (D^\mu \chi)^\dagger D_\mu \chi - V(H,\chi)\,, \tag{47}$$

with the scalar potential given by

$$V(H,\chi) = -m^2 H^\dagger H - \mu^2 \mid \chi \mid^2 + \lambda_1 (H^\dagger H)^2 + \lambda_2 \mid \chi \mid^4 + \lambda_3 H^\dagger H \mid \chi \mid^2\,, \tag{48}$$

where H and χ are the complex scalar Higgs doublet and singlet fields, respectively.

Finally, the Yukawa interactions are,

$$\mathcal{L}_Y = -y^d_{jk}\,\overline{q_{jL}}d_{kR}H - y^u_{jk}\,\overline{q_{jL}}u_{kR}\widetilde{H} - y^e_{jk}\,\overline{l_{jL}}e_{kR}H - y^\nu_{jk}\,\overline{l_{jL}}\nu_{kR}\widetilde{H} - y^M_{jk}\,\overline{(\nu_R)^c_j}\nu_{kR}\chi + \text{h.c.}\,, \tag{49}$$

where $\widetilde{H} = i\sigma^2 H^*$ and i, j, k take the values 1 to 3, where the last term is the Majorana contribution and the others the usual Dirac ones.

The implementation of the model into FEYNRULES is straightforward. Only the neutrino sector is more complicated and needs to be suitably rewritten to manifestly preserve gauge invariance. As a first step we rewrite Dirac neutrino fields in terms of Majorana ones using the following general substitution,

$$\nu^D = \frac{1-\gamma_5}{2}\nu_L + \frac{1+\gamma_5}{2}\nu_R\,, \tag{50}$$

where ν^D is a Dirac field and $\nu_{L(R)}$ are its left (right) Majorana components. If we perform the substitution of Eq. (50) in the neutrino sector of the SM, we will have an equivalent theory formulated in terms of Majorana neutrinos consistent with all experimental constraints. Furthermore, from Eq. (49) we obtain the neutrino mass matrix,

$$\mathcal{M} = \begin{pmatrix} 0 & m_D \\ m_D & M \end{pmatrix} , \tag{51}$$

with

$$m_D = \frac{y^\nu}{\sqrt{2}}\, v\,, \qquad M = \sqrt{2}\, y^M\, x\,, \tag{52}$$

where x is the Vacuum Expectation Value (VEV) of the χ field. This matrix can be diagonalized by a rotation about an angle α_ν, such that,

$$\tan 2\alpha_\nu = -\frac{2m_D}{M}\,. \tag{53}$$

For simplicity we neglect the inter-generational mixing so that neutrinos of each generation can be diagonalized independently. We also require that the neutrinos be mass degenerate. Thus, $\nu_{L,R}$ can be written as the following linear combination of Majorana mass eigenstates $\nu_{l,h}$,

$$\begin{pmatrix} \nu_L \\ \nu_R \end{pmatrix} = \begin{pmatrix} \cos\alpha_\nu & -\sin\alpha_\nu \\ \sin\alpha_\nu & \cos\alpha_\nu \end{pmatrix} \times \begin{pmatrix} \nu_l \\ \nu_h \end{pmatrix} . \tag{54}$$

Neutrino mass eigenstates will be called ν_l and ν_h, the former being SM-like. With a reasonable choice of Yukawa couplings, the heavy neutrinos can have masses $m_{\nu_h} \sim \mathcal{O}(100)$ GeV $\ll M_{Z'}$.

The last subtle point is the way the Lagrangian has to be written, in particular the Majorana-like Yukawa terms for the right-handed neutrinos (the last term in Eq. (49)). In order to explicitly preserve gauge invariance, this term has to be written, in two-component notation, as,

$$-y^M\, \overline{\nu^c}\, \frac{1+\gamma_5}{2} \nu\chi + \text{h.c.}\,, \tag{55}$$

where ν is the Dirac field of Eq. (50), whose Majorana components $\nu_{L,R}$ mix as in Eq. (54).

The implementation of the minimal $B-L$ model in FEYNRULES was validated by comparing against an independent implementation in CALCHEP performed by means of the LANHEP package [308]. We compared the matrix elements squared in MADGRAPH and CALCHEP at various randomly-generated phase space points for 284 two-to-two processes and we found perfect agreement in all cases. Let us note that the current FEYNRULES implementation was done in unitary gauge, whereas the LANHEP implementation allows for both Feynman and unitary gauges, and hence we demonstrated explicitly the gauge invariance of our implementation. Furthermore, because of mixing in the scalar sector, we paid particular attention to the unitarity cancellation in weak boson scattering to ensure that the implementation has the correct high-energy behavior. A selection of the results obtained during the validation procedure is shown in Table 8. A more extensive list, together with the corresponding FEYNRULES model file, can be obtained from the FEYNRULES model database.

20.42 THE NEXT-TO-MINIMAL SUPERSYMMETRIC STANDARD MODEL

The Next-to-Minimal Supersymmetric Standard Model (NMSSM) is a viable extension of the MSSM in the Higgs sector of the model (see, *e.g.*, the references in [309]). The main motivation for the NMSSM is a possible solution to the so-called μ problem, the fact that the supersymmetric parameter μ in the superpotential is dimensionful and should in principle be of the order of the SUSY breaking scale. A working electroweak symmetry breaking demands this parameter, however, in the ball park of a few hundred GeV. The NMSSM addresses this problem by enlarging the particle spectrum by a single chiral

Process	MG (FR)	CH (FR)	CH (LH)
$e^+e^- \to \nu_{l1}\nu_{l1}$	5.681e+01	5.687e+01	5.687e+01
$e^+e^- \to \nu_{l1}\nu_{h1}$	5.315e−12	5.307e−12	5.307e−12
$e^+e^- \to \nu_{h1}\nu_{h1}$	2.697e−02	2.691e−02	2.691e−02
$e^+e^- \to ZZ'$	1.570e−01	1.569e−01	1.569e−01
$W^+W^- \to W^+W^-$	1.420e+03	1.421e+03	1.421e+03

Table 8: Cross section (in pb) for a selection of processes in the minimal $B-L$ model corresponding to $\sqrt{s} = 4$ TeV. A p_T cut of 20 GeV was applied to all final state particles. MG (FR) and CH (FR) refer to the FEYNRULES implementations in MADGRAPH and CALCHEP respectively, whereas CH (LH) refers to the LANHEP implementation of Ref. [308].

superfield S, being a singlet under the SM gauge group. When the singlet acquires a vev, it generates an effective μ parameter. The quartic term for the Higgs-potential is generated via an F term from a cubic superpotential term.

In the following, we will use the conventions of the SUSY Les Houches Accord 2 [43, 44, 310]. This is a quite general approach, but neglecting possible CP, R-parity, or flavor violation. The superpotential and the soft-breaking terms of the NMSSM are given by

$$\begin{aligned} W_{NMSSM} &= W_{MSSM} - \epsilon_{ab}\lambda S H_1^a H_2^b + \frac{1}{3}\kappa S^3 + \mu' S^2 + \xi_F S \,, \\ V_{\text{soft}} &= V_{MSSM} + m_{\text{S}}^2 |S|^2 + \Big(- \epsilon_{ab}\lambda A_\lambda S H_1^a H_2^b \\ &\quad + \frac{1}{3}\kappa A_\kappa S^3 + m_S'^2 S^2 + \xi_S S + \text{h.c.}\Big) \,. \end{aligned} \tag{56}$$

In the WHIZARD implementation, we have ommited all non-Z_3 symmetric terms in the superpotential, as well as the corresponding soft terms, while the FEYNRULES model includes the most general superpotential. The field content of the NMSSM is almost the same as for the MSSM, except for an additional scalar and pseudoscalar Higgs boson, denoted by H_3^0 and A_2^0, as well as a fifth neutralino, $\tilde{\chi}_5^0$, coming from the additional singlino component. In addition to the conventions from Ref. [43, 44, 310] to have left-right sfermion mixing in all three generations, the FEYNRULES implementation is a bit more general, allowing for inter-generational sfermion mixings.

The tests that we have performed follow closely the strategy described in Section 20.3 We have compared the cross section for 736 processes in WHIZARD and in the FEYNRULES-generated model files for CALCHEP and MADGRAPH. A very short selection of results for processes including Higgses and neutralinos can be found in the Table 9. The full list of processes which we have investigated, together with the results for the cross sections and the selected benchmark point, will be made available in the FEYNRULES model database and on the WHIZARD home page. For the WHIZARD implementation it was checked that it yields the correct MSSM limit [311] .

20.43 THE MINIMAL R-SYMMETRIC SUPERSYMMETRIC STANDARD MODEL

The minimal R-symmetric supersymmetric standard model (MRSSM) extends the usual MSSM by introducing a global continuous $U(1)_R$ symmetry [312]. All the quark and lepton supermultiplets carry an R-charge $+1$, whereas the Higgs and gauge supermultiplets carry an R-charge 0. This implies that gaugino Majorana mass terms are forbidden, as well as the μ-term in the superpotential. Since supersymmetry must be broken, the gauginos must be massive and we need a new mechanism to introduce gaugino mass terms. This is achieved by including for each gauge supermultiplet an additional chiral supermultiplet, transforming in the adjoint representation of the gauge group and with an R-charge 0, making it possible

Process	WO (ST)	MG (FR)	CH (FR)
$e^+ e^- \to \tilde{\chi}_4^0 \tilde{\chi}_5^0$	4.882e−3	4.884e−3	4.886e−3
$\tau^- \bar{\nu}_\tau \to \tilde{\chi}_1^- \tilde{\chi}_5^0$	3.476e−4	3.470e−4	3.469e−4
$W^- Z \to H_3^0 H^-$	4.211e−3	4.213e−3	4.210e−3
$Z Z \to A_1^0 A_1^0$	6.505e−3	6.512e−3	6.513e−3
$W^+ W^- \to \tilde{\chi}_3^0 \tilde{\chi}_5^0$	2.055e−1	2.056e−1	2.056e−1

Table 9: Cross sections for a selection of production processes in the NMSSM. The built-in MSSM implementation in WHIZARD is denoted WO (ST) while the FEYNRULES-generated ones are MG (FR) and CH (FR). The center-of-mass energy is fixed to 3000 GeV.

to write down Dirac mass terms for the gauginos. As a consequence, at variance with the usual MSSM scenario where the neutralinos and gluinos are expected to be Majorana fermions, the gluinos and the weak inos in the MRSSM are four-component Dirac fermions, leading to a new phenomenology for supersymmetric models. In the following however, we only review the $SU(3)_C$ sector of the model. The electroweak sector is conceptually similar, but more involved because of the mixing in the electroweak gauge sector and we refer the interested reader to Ref. [313–316] for further details on the MRSSM and its phenomenology.

The $SU(3)_C$ sector of the MSSM describes the gauge interactions between the quark superfields $Q = (\tilde{q}_L, q_L, F_Q), U = (\tilde{u}_R^\dagger, u_L^c, F_U)$ and $D = (\tilde{d}_R^\dagger, d_L^c, F_D)$ and the gluon superfield $G = (\lambda, g_\mu, D)$. In the MRSSM this sector is augmented by an additional chiral supermultiplet $\Phi_g = (\phi, \chi, F_g)$ transforming as an octet. In order to give the gluino a mass in an R-symmetric way, the soft supersymmetry breaking part of the Lagrangian contains a Dirac mass term coupling the Weyl components of the superfields G and Φ_g,

$$\mathcal{L}_{\text{soft}} \supset M_D \left(\lambda^a \cdot \chi^a + \bar{\chi}^a \cdot \bar{\lambda}^a\right) + M_D\, D^a(\phi_a + \phi_a^\dagger)\,. \tag{57}$$

The physical gluino field then corresponds to the four-component Dirac fermion $\tilde{g} = (\chi, \bar{\lambda})^T$. The interactions between the superfields are described by the usual supersymmetric gauge interactions. Integrating out auxiliary fields and omitting all terms involving quark fields left unchanged with respect to the usual MSSM, the gauge interactions can be described by the Lagrangian,

$$\begin{aligned}
\mathcal{L}_g &= D_\mu \phi_a^\dagger\, D^\mu \phi_a + \lambda^{a\dagger}\, i\bar{\sigma}^\mu D_\mu \lambda^a + \chi^{a\dagger}\, i\bar{\sigma}^\mu D_\mu \chi^a \\
&+ \frac{1}{2} M_D^2\, (\phi_a + \phi_a^\dagger)^2 + \frac{1}{2} M_{s8}^2\, (\phi_a + \phi_a^\dagger)^2 + \frac{1}{2} M_{p8}^2\, (\phi_a - \phi_a^\dagger)^2 + M_D\, (\lambda^a \cdot \chi^a + \bar{\chi}^a \cdot \bar{\lambda}^a) \\
&+ g_s\, M_D\, (\phi_a + \phi_a^\dagger) \left(\tilde{q}_L^\dagger T^a \tilde{q}_L - \tilde{u}_R^\dagger T^a \tilde{u}_R - \tilde{d}_R^\dagger T^a \tilde{d}_R\right) \\
&+ \frac{g_s^2}{2} \left(\tilde{q}_L^\dagger T^a \tilde{q}_L - \tilde{u}_R^\dagger T^a \tilde{u}_R - \tilde{d}_R^\dagger T^a \tilde{d}_R - i f^{abc}\, \phi_b^\dagger \phi_c\right) \\
&+ i\sqrt{2}\, g_s\, f^{abc} \left(\phi_b^\dagger\, \chi^c \cdot \lambda^a + \phi_b\, \bar{\lambda}^a \cdot \bar{\chi}^c\right)\,,
\end{aligned} \tag{58}$$

where D_μ denotes the covariant derivative in the adjoint representation of $SU(3)_C$,

$$D_\mu = \partial_\mu - i g_s\, F^a\, g_\mu^a\,, \tag{59}$$

and $(F^a)_{bc} = -i f^{abc}$ denote the generators of the adjoint representation.

We implemented the full MRSSM in FEYNRULES in its most general form, *i.e.*, keeping generic and possible complex mixing between particles. The validation of the implementation is currently on-going.

20.5 CONCLUSION

In this summary report we presented the activities of the FEYNRULES working group triggered by discussion at the Les Houches workshop. We argued that the FEYNRULES framework does not only provide a natural framework where BSM models can easily be developed and implemented, but they can also be validated to an unprecedented level by exploiting the fact that the model can be exported to various matrix element generators in an automated way. We proposed a rating scheme based on a four credit points system by which FEYNRULES models can be classified with respect to their level of validation, in the perspective of building a database of robust and thoroughly validated implementations, and we illustrated the power of this proposal on several newly implemented and validated models, in particular the minimal $B - L$ model, well as several supersymmetric extensions of the Standard Model.

ACKNOWLEDGEMENTS

In addition to the invidiual acknowledgements throughout the contributions, all the authors would all like to thank the organisers for the usual great atmosphere and hospitality. The meeting was financially supported by the European Union Marie Curie Conferences and Training Courses Contract: CT-2006-046171, the Haute Savoie and Rhône Alps regions, Universite Savoie, LAPP and LAPTH Annecy, CNRS and the Ministéré des Affaires Étrangers.

References

[1] C. Buttar *et. al.*, `arXiv:0803.0678 [hep-ph]`.

[2] E. Boos *et. al.*, `hep-ph/0109068`.

[3] J. Alwall *et. al.*, *Comput. Phys. Commun.* **176** (2007) 300–304, [`arXiv:hep-ph/0609017`].

[4] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber, *JHEP* **11** (2001) 063, [`hep-ph/0109231`].

[5] L. Lönnblad, *JHEP* **05** (2002) 046, [`arXiv:hep-ph/0112284`].

[6] S. Belov, L. Dudko, D. Kekelidze, and A. Sherstnev, `arXiv:1001.2576 [hep-ph]`.

[7] S. Belov *et. al.*, *Comput. Phys. Commun.* **178** (2008) 222, [`arXiv:hep-ph/0703287`].

[8] C. Amsler *et. al.*,, **Particle Data Group** Collaboration *Phys. Lett.* **B667** (2008) 1.

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

[10] S. Frixione, Z. Kunszt, and A. Signer, *Nucl. Phys.* **B467** (1996) 399–442, [`arXiv:hep-ph/9512328`].

[11] S. Frixione, *Nucl. Phys.* **B507** (1997) 295–314, [`hep-ph/9706545`].

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

[13] S. Frixione and B. R. Webber, `arXiv:hep-ph/0612272`.

[14] P. Nason, *JHEP* **11** (2004) 040, [`arXiv:hep-ph/0409146`].

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

[16] Z. Nagy, *Phys. Rev.* **D68** (2003) 094002, [`arXiv:hep-ph/0307268`].

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

[18] T. Sjostrand, S. Mrenna, and P. Skands, *JHEP* **05** (2006) 026, [`arXiv:hep-ph/0603175`].

[19] S. Gieseke, A. Ribon, M. H. Seymour, P. Stephens, and B. Webber, *JHEP* **02** (2004) 005, [`hep-ph/0311208`].

[20] T. Sjostrand, S. Mrenna, and P. Skands, *Comput. Phys. Commun.* **178** (2008) 852–867, [`arXiv:0710.3820 [hep-ph]`].

[21] M. Bahr *et. al.*, *Eur. Phys. J.* **C58** (2008) 639–707, [`arXiv:0803.0883 [hep-ph]`].

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

[23] T. Gleisberg *et. al.*, *JHEP* **02** (2009) 007, [`arXiv:0811.4622 [hep-ph]`].

[24] M. H. Seymour, *Comp. Phys. Commun.* **90** (1995) 95–101, [hep-ph/9410414].

[25] S. Frixione, P. Nason, and G. Ridolfi, arXiv:0707.3081 [hep-ph].

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

[27] F. Mahmoudi, *Comput. Phys. Commun.* **178** (2008) 745–754, [arXiv:0710.2067 [hep-ph]].

[28] F. Mahmoudi, *Comput. Phys. Commun.* **180** (2009) 1579–1613, [arXiv:0808.3144 [hep-ph]].

[29] F. Mahmoudi, *Comput. Phys. Commun.* **180** (2009) 1718–1719.

[30] G. Degrassi, P. Gambino, and P. Slavich, *Comput. Phys. Commun.* **179** (2008) 759–771, [arXiv:0712.3265 [hep-ph]].

[31] P. Paradisi, talk given at ``Interplay of Collider and Flavour Physics, 3rd general meeting'', CERN, December 2009.

[32] G. Belanger, F. Boudjema, A. Pukhov, and A. Semenov, *Comput. Phys. Commun.* **180** (2009) 747–767, [arXiv:0803.2360 [hep-ph]].

[33] A. Arbey and F. Mahmoudi, arXiv:0906.0369 [hep-ph].

[34] S. Heinemeyer, W. Hollik, and G. Weiglein, *Comput. Phys. Commun.* **124** (2000) 76–89, [arXiv:hep-ph/9812320].

[35] S. Heinemeyer, W. Hollik, and G. Weiglein, *Eur. Phys. J.* **C9** (1999) 343–366, [arXiv:hep-ph/9812472].

[36] J. S. Lee *et. al.*, *Comput. Phys. Commun.* **156** (2004) 283–317, [arXiv:hep-ph/0307377].

[37] U. Ellwanger and C. Hugonie, *Comput. Phys. Commun.* **175** (2006) 290–303, [arXiv:hep-ph/0508022].

[38] F. E. Paige, S. D. Protopopescu, H. Baer, and X. Tata, arXiv:hep-ph/0312045.

[39] R. Lafaye, T. Plehn, and D. Zerwas, arXiv:hep-ph/0404282.

[40] P. Bechtle, K. Desch, and P. Wienemann, *Comput. Phys. Commun.* **174** (2006) 47–70, [arXiv:hep-ph/0412012].

[41] R. R. de Austri, R. Trotta, and L. Roszkowski, *JHEP* **05** (2006) 002, [arXiv:hep-ph/0602028].

[42] S. Heinemeyer, talk given at ``Interplay of Collider and Flavour Physics, 2nd general meeting'', CERN, March 2009.

[43] P. Z. Skands *et. al.*, *JHEP* **07** (2004) 036, [arXiv:hep-ph/0311123].

[44] B. Allanach *et. al.*, *Comp. Phys. Commun.* **180** (2009) 8–25, [arXiv:0801.0045 [hep-ph]].

[45] T. Hahn, *Comput. Phys. Commun.* **180** (2009) 1681–1693, [arXiv:hep-ph/0605049].

[46] A. Buckley, J. Butterworth, L. Lönnblad, H. Hoeth, J. Monk, H. Schulz, J. E. von Seggern, F. Siegert, and L. Sonnenschein, arXiv:1003.0694 [hep-ph].

[47] A. Buckley, H. Hoeth, H. Lacker, H. Schulz, and J. E. von Seggern, *Eur. Phys. J.* **C65** (2010) 331–357, [arXiv:0907.2973 [hep-ph]].

[48] J. Bromley *et. al.*,. Prepared for Workshop on Future Physics at HERA (Preceded by meetings 25-26 Sep 1995 and 7-9 Feb 1996 at DESY), Hamburg, Germany, 30-31 May 1996.

[49] B. M. Waugh *et. al.*, arXiv:hep-ph/0605034.

[50] M. Dobbs and J. B. Hansen, *Comput. Phys. Commun.* **134** (2001) 41–46.

[51] M. Cacciari, arXiv:hep-ph/0607071.

[52] A. Buckley, *PoS* **ACAT2007** (2007) 050, [arXiv:0708.2655 [hep-ph]].

[53] A. Buckley *et. al.*, arXiv:hep-ph/0605048.

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

[55] J. M. Butterworth and M. H. Seymour,, October, 2004. JIMMY4: Multiparton Interactions in Herwig for the LHC, http://projects.hepforge.org/jimmy/.

[56] M. Bahr, J. M. Butterworth, and M. H. Seymour, *JHEP* **01** (2009) 065, [arXiv:0806.2949 [hep-ph]].

[57] A. Sherstnev and R. S. Thorne, *Eur. Phys. J.* **C55** (2008) 553–575, [arXiv:0711.2473 [hep-ph]].

[58] A. A. Affolder *et. al.*,, **CDF** Collaboration *Phys. Rev.* **D65** (2002) 092002.

[59] D. E. Acosta *et. al.*,, **CDF** Collaboration *Phys. Rev.* **D70** (2004) 072002, [arXiv:hep-ex/0404004].

[60] **CDF** CollaborationR. Field, ,, "The Underlying Event and Comparisons with MC." First International Workshop on Multiple Partonic Interactions at the LHC, 2008.

[61] V. M. Abazov *et. al.*,, **D0** Collaboration *Phys. Rev. Lett.* **94** (2005) 221801, [arXiv:hep-ex/0409040].

[62] P. Z. Skands, arXiv:0905.3418 [hep-ph].

[63] R. D. Ball *et. al.*,, **NNPDF** Collaboration *Nucl. Phys.* **B809** (2009) 1–63, [arXiv:0808.1231 [hep-ph]].

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

[65] S. Hoeche, F. Krauss, S. Schumann, and F. Siegert, *JHEP* **05** (2009) 053, [arXiv:0903.1219 [hep-ph]].

[66] P. Lenzi and J. M. Butterworth, arXiv:hep-ph/0903.3918 [hep-ph].

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

[68] H. Hoeth,. Perugia MPI workshop, October, 2008.

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

[70] M. L. Mangano, M. Moretti, and R. Pittau, *Nucl. Phys.* **B632** (2002) 343–362, [hep-ph/0108069].

[71] T. Akesson *et. al.*,, **Axial Field Spectrometer** Collaboration *Z. Phys.* **C34** (1987) 163.

[72] F. Abe *et. al.*,, **CDF** Collaboration *Phys. Rev. Lett.* **79** (1997) 584–589.

[73] F. Abe *et. al.*,, **CDF** Collaboration *Phys. Rev.* **D56** (1997) 3811–3832.

[74] D. Collaboration,, **D0** Collaboration *D0 note 5910-CONF* (2009).

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

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

[77] M. Bähr, S. Gieseke, and M. H. Seymour, *JHEP* **07** (2008) 076, [arXiv:0803.3633 [hep-ph]].

[78] E. Maina, *JHEP* **09** (2009) 081, [arXiv:0909.1586 [hep-ph]].

[79] E. Maina, *JHEP* **04** (2009) 098, [arXiv:0904.2682 [hep-ph]].

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

[81] J. Alwall *et. al.*, *JHEP* **09** (2007) 028, [arXiv:0706.2334 [hep-ph]].

[82] J. Pumplin, A. Belyaev, J. Huston, D. Stump, and W. K. Tung, *JHEP* **02** (2006) 032, [arXiv:hep-ph/0512167].

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

[84] U. Baur, T. Han, and J. Ohnemus, *Phys. Rev.* **D48** (1993) 5140–5161, [arXiv:hep-ph/9305314].

[85] U. Baur, *http://ubhex.physics.buffalo.edu/ baur/wgam_nlo.tar.gz.*

[86] Z. Nagy and D. E. Soper, *JHEP* **10** (2005) 024, [arXiv:hep-ph/0503053].

[87] Z. Nagy and D. E. Soper, hep-ph/0601021.

[88] W. T. Giele, D. A. Kosower, and P. Z. Skands, *Phys. Rev.* **D78** (2008) 014026, [arXiv:0707.3652 [hep-ph]].

[89] S. Schumann and F. Krauss, *JHEP* **03** (2008) 038, [arXiv:0709.1027 [hep-ph]].

[90] M. Dinsdale, M. Ternick and S. Weinzierl, *Phys. Rev.* **D76** (2007) 094003, [arXiv:0709.1026 [hep-ph]].

[91] J.-C. Winter and F. Krauss, *JHEP* **07** (2008) 040, [arXiv:0712.3913 [hep-ph]].

[92] S. Platzer and S. Gieseke, arXiv:0909.5593 [hep-ph].

[93] G. Gustafson, *Phys. Lett.* **B175** (1986) 453.

[94] G. Gustafson and U. Pettersson, *Nucl. Phys.* **B306** (1988) 746.

[95] B. Andersson, G. Gustafson, and L. Lönnblad, *Nucl. Phys.* **B339** (1990) 393–406.

[96] B. Andersson, G. Gustafson, L. Lönnblad, and U. Pettersson, *Z. Phys.* **C43** (1989) 625.

[97] L. Lönnblad, *Comput. Phys. Commun.* **71** (1992) 15–31.

[98] Y. L. Dokshitzer and G. Marchesini, *JHEP* **03** (2009) 117, [arXiv:0809.1749 [hep-ph]].

[99] Z. Nagy and D. E. Soper, *JHEP* **05** (2009) 088, [arXiv:0901.3587 [hep-ph]].

[100] P. Skands and S. Weinzierl, arXiv:0903.2150 [hep-ph].

[101] A. Bassetto, M. Ciafaloni, and G. Marchesini, *Phys. Rept.* **100** (1983) 201–272.

[102] F. A. Berends and W. T. Giele, *Nucl. Phys.* **B313** (1989) 595.

[103] F. Englert and R. Brout, *Phys. Rev. Lett.* **13** (1964) 321–322.

[104] P. W. Higgs, *Phys. Rev. Lett.* **13** (1964) 508–509.

[105] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Phys. Rev. Lett.* **13** (1964) 585–587.

[106] T. W. B. Kibble, *Phys. Rev.* **155** (1967) 1554–1561.

[107] H. M. Georgi, S. L. Glashow, M. E. Machacek, and D. V. Nanopoulos, *Phys. Rev. Lett.* **40** (1978) 692.

[108] **ALEPH** Collaboration *et. al.*, arXiv:0811.4682 [hep-ex].

[109] G. Bernardi *et. al.*,, **Tevatron New Phenomena and Higgs Working Group** Collaboration arXiv:0808.0534 [hep-ex].

[110] T. Aaltonen *et. al.*,, **CDF and DØ** Collaboration arXiv:1001.4162 [Unknown].

[111] A. Djouadi, *Phys. Rept.* **457** (2008) 1–216, [arXiv:hep-ph/0503172].

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

[113] S. Frixione, E. Laenen, P. Motylinski, and B. R. Webber, *JHEP* **03** (2006) 092, [arXiv:hep-ph/0512250].

[114] S. Frixione, E. Laenen, P. Motylinski, and B. R. Webber, *JHEP* **04** (2007) 081, [arXiv:hep-ph/0702198].

[115] S. Frixione, E. Laenen, P. Motylinski, B. R. Webber, and C. D. White, *JHEP* **07** (2008) 029, [arXiv:0805.3067 [hep-ph]].

[116] O. Latunde-Dada, *JHEP* **11** (2007) 040, [arXiv:0708.4390 [hep-ph]].

[117] P. Nason and G. Ridolfi, *JHEP* **08** (2006) 077, [arXiv:hep-ph/0606275].

[118] S. Frixione, P. Nason, and G. Ridolfi, *JHEP* **09** (2007) 126, [arXiv:0707.3088 [hep-ph]].

[119] O. Latunde-Dada, S. Gieseke, and B. Webber, *JHEP* **02** (2007) 051, [arXiv:hep-ph/0612281].

[120] K. Hamilton, P. Richardson, and J. Tully, *JHEP* **10** (2008) 015, [arXiv:0806.0290 [hep-ph]].

[121] K. Hamilton, P. Richardson, and J. Tully, *JHEP* **04** (2009) 116, [arXiv:0903.4345 [hep-ph]].

[122] S. Alioli, P. Nason, C. Oleari, and E. Re, *JHEP* **07** (2008) 060, [arXiv:0805.4802 [hep-ph]].

[123] S. Alioli, P. Nason, C. Oleari, and E. Re, *JHEP* **04** (2009) 002, [arXiv:0812.0578 [hep-ph]].

[124] S. Alioli, P. Nason, C. Oleari, and E. Re, *JHEP* **09** (2009) 111, [arXiv:0907.4076 [hep-ph]].

[125] O. Latunde-Dada, *Eur. Phys. J.* **C58** (2008) 543–554, [arXiv:0806.4560 [hep-ph]].

[126] F. Krauss, *JHEP* **08** (2002) 015, [arXiv:hep-ph/0205283].

[127] A. Schälicke and F. Krauss, *JHEP* **07** (2005) 018, [arXiv:hep-ph/0503281].

[128] S. Mrenna and P. Richardson, *JHEP* **05** (2004) 040, [arXiv:hep-ph/0312274].

[129] K. Hamilton, P. Richardson, and J. Tully, arXiv:0905.3072 [hep-ph].

[130] C. Anastasiou, K. Melnikov, and F. Petriello, *Nucl. Phys.* **B724** (2005) 197–246, [arXiv:hep-ph/0501130].

[131] C. Anastasiou, G. Dissertori, and F. Stockli, *JHEP* **09** (2007) 018, [arXiv:0707.2373 [hep-ph]].

[132] S. Catani and M. Grazzini, *Phys. Rev. Lett.* **98** (2007) 222002, [arXiv:hep-ph/0703012].

[133] M. Grazzini, *JHEP* **02** (2008) 043, [arXiv:0801.3232 [hep-ph]].

[134] J. Alwall *et. al.*, *Eur. Phys. J.* **C53** (2008) 473–500, [arXiv:0706.2569 [hep-ph]].

[135] J. Alwall, S. de Visscher, and F. Maltoni, *JHEP* **02** (2009) 017, [arXiv:0810.5350 [hep-ph]].

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

[137] A. Papaefstathiou and O. Latunde-Dada, *JHEP* **07** (2009) 044, [arXiv:0901.3685 [hep-ph]].

[138] S. Alioli, P. Nason, C. Oleari, and E. Re,. To appear soon.

[139] P. Nason and C. Oleari, arXiv:0911.5299 [hep-ph].

[140] A. Bassetto, M. Ciafaloni, G. Marchesini, and A. H. Mueller, *Nucl. Phys.* **B207** (1982) 189.

[141] S. Catani and M. Ciafaloni, *Nucl. Phys.* **B236** (1984) 61.

[142] M. Ciafaloni, *Phys. Lett.* **B95** (1980) 113.

[143] M. Ciafaloni,, "Soft Gluon Contributions to Hard Processes." Lectures given at Summer Workshop on High Energy Physics, Trieste, Italy, Aug 1981.

[144] Y. L. Dokshitzer, V. A. Khoze, and S. I. Troian, *Adv. Ser. Direct. High Energy Phys.* **5** (1988) 241–410.

[145] G. Marchesini and B. R. Webber, *Nucl. Phys.* **B238** (1984) 1.

[146] A. H. Mueller, *Phys. Lett.* **B104** (1981) 161–164.

[147] B. I. Ermolaev and V. S. Fadin, *JETP Lett.* **33** (1981) 269–272.

[148] Y. L. Dokshitzer, V. S. Fadin, and V. A. Khoze, *Phys. Lett.* **B115** (1982) 242–246.

[149] F. Krauss, R. Kuhn, and G. Soff, *JHEP* **02** (2002) 044, [arXiv:hep-ph/0109036].

[150] T. Gleisberg and S. Hoche, *JHEP* **12** (2008) 039, [arXiv:0808.3674 [hep-ph]].

[151] S. Höche, S. Schumann, and F. Siegert, arXiv:0912.3501 [hep-ph].

[152] T. Carli, T. Gehrmann, and S. Höche, arXiv:0912.3715 [hep-ph].

[153] F. Krauss, A. Schalicke, S. Schumann, and G. Soff, *Phys. Rev.* **D70** (2004) 114009, [arXiv:hep-ph/0409106].

[154] F. Krauss, A. Schalicke, S. Schumann, and G. Soff, *Phys. Rev.* **D72** (2005) 054017, [arXiv:hep-ph/0503280].

[155] T. Gleisberg, F. Krauss, A. Schalicke, S. Schumann, and J.-C. Winter, *Phys. Rev.* **D72** (2005) 034028, [arXiv:hep-ph/0504032].

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

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

[158] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, *Phys. Rev. Lett.* **100** (2008) 242001, [arXiv:0802.2470 [hep-ph]].

[159] **ATLAS** CollaborationG. Aad *et. al.*, , Tech. Rep. ATL-PHYS-INT-2009-068, ATL-PHYS-PUB-2009-088, CERN, Geneva, June, 2009.

[160] S. Alioli, P. Nason, C. Oleari, and E. Re, arXiv:1002.2581 [hep-ph].

[161] G. Piacquadio,. Freiburg PhD thesis, CERN-THESIS-2010-027.

[162] F. Febres Cordero, L. Reina, and D. Wackeroth, *Phys. Rev.* **D74** (2006) 034007, [arXiv:0606102 [hep-ph]].

[163] F. Febres Cordero, arXiv:0809.3829 [hep-ph].

[164] F. Febres Cordero, L. Reina, and D. Wackeroth, *Phys. Rev.* **D80** (2009) 034015, [arXiv:0906.1923 [hep-ph]].

[165] B. P. Kersevan and E. Richter-Was, arXiv:hep-ph/0405247.

[166] S. Agostinelli *et. al.*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** (2003), no. 3 250 – 303.

[167] S. Ovyn, X. Rouby, and V. Lemaitre, arXiv:0903.2225 [hep-ph].

[168] D. Acosta, M. Della Negra, L. Foà, A. Hervé, and A. Petrilli,, *CMS physics: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 2006. There is an error on cover due to a technical problem for some items.

[169] G. Aad *et. al.*,, **The ATLAS** Collaboration arXiv:0901.0512 [hep-ex].

[170] X. Rouby, J. de Favereau, and K. Piotrzkowski, *Journal of Instrumentation* **2** (2007), no. 09 P09005.

[171] M. Cacciari and G. P. Salam, *Physics Letters B* **641** (2006), no. 1 57 – 61.

[172] O. C. K. Ackerstaff *et al*, *Eur. Phys. J.* **C4** (1998) 47–74.

[173] M. Vesterinen and T. R. Wyatt, *Nucl. Instrum. Meth. A* **602** (2009) 432.

[174] T. Affolder *et. al.*,, **CDF** Collaboration *Phys. Rev. Lett.* **84** (2000) 845–850, [arXiv:hep-ex/0001021].

[175] D. C. B. Abbott *et al.*, *Phys. Rev. Lett.* **84** (2000) 2792.

[176] D. C. V. M. Abazov *et al.*, *Phys. Rev. Lett.* **100** (2008) 102002.

[177] S. G. M. H. Seymour and A. Siodmok, *Acta Phys.Polon.* **B40** (2009) 2109–2118.

[178] M. Cacciari, G. Salam, and G. Soyez, *JHEP* **0804** (2008) 063, [arXiv:0802.1189 [hep-ph]].

[179] S. Chekanov *et. al.*,, **ZEUS** Collaboration *Nucl. Phys.* **B700** (2004) 3–50, [hep-ex/0405065].

[180] T. Becher and M. D. Schwartz, *JHEP* **07** (2008) 034, [arXiv:0803.0342 [hep-ph]].

[181] M. H. Seymour, *Z. Phys.* **C62** (1994) 127–138.

[182] J. M. Butterworth, B. E. Cox, and J. R. Forshaw, *Phys. Rev. D* **65** (2002) [hep-ph/0201098].

[183] J. M. Butterworth, J. R. Ellis, and A. R. Raklev, *JHEP* **05** (2007) 033, [hep-ph/0702150].

[184] G. H. Brooijmans *et. al.*, arXiv:0802.3715 [hep-ph].

[185] D. E. Kaplan, K. Rehermann, M. D. Schwartz, and B. Tweedie, *Phys. Rev. Lett.* **101** (2008) 142001, [arXiv:0806.0848 [hep-ph]].

[186] J. Thaler and L.-T. Wang, *JHEP* **07** (2008) 092, [arXiv:0806.0023 [hep-ph]].

[187] **ATLAS** CollaborationG. Aad *et. al.*, , Tech. Rep. ATL-PHYS-PUB-2009-081; ATL-COM-PHYS-2009-255, CERN, Geneva, May, 2009.

[188] T. Plehn, G. P. Salam, and M. Spannowsky, arXiv:0910.5472 [hep-ph].

[189] G. D. Kribs, A. Martin, T. S. Roy, and M. Spannowsky, arXiv:0912.4731 [hep-ph].

[190] J. M. Butterworth, J. R. Ellis, A. R. Raklev, and G. P. Salam, *Phys. Rev. Lett.* **103** (2009) 241803, [arXiv:0906.0728 [hep-ph]].

[191] **ATLAS** CollaborationG. Aad *et. al.*, , Tech. Rep. ATL-PHYS-PUB-2009-076; ATL-COM-PHYS-2009-262, CERN, Geneva, May, 2009.

[192] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, arXiv:0912.0033 [hep-ph].

[193] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, *Phys. Rev.* **D80** (2009) 051501, [arXiv:0903.5081 [hep-ph]].

[194] D. Krohn, J. Thaler, and L.-T. Wang, arXiv:0912.1342 [hep-ph].

[195] BOOST 2009 workshop, SLAC http://www-conf.slac.stanford.edu/Boost2009/ and Workshop on Jets and JetSubstructure at the LHC http://silicon.phys.washington.edu/JetsWorkshop/.

[196] V. M. Abazov *et. al.*,, **D0** Collaboration *Phys. Rev.* **D65** (2002) 052008, [hep-ex/0108054].

[197] D. Acosta *et. al.*,, **CDF** Collaboration *Phys. Rev.* **D71** (2005) 112002, [hep-ex/0505013].

[198] G. Abbiendi *et. al.*,, **OPAL** Collaboration *Eur. Phys. J.* **C37** (2004) 25–47, [hep-ex/0404026].

[199] G. Abbiendi *et. al.*,, **OPAL** Collaboration *Eur. Phys. J.* **C31** (2003) 307–325, [hep-ex/0301013].

[200] D. Buskulic *et. al.*,, **ALEPH** Collaboration *Phys. Lett.* **B384** (1996) 353–364.

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

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

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

[204] G. Giurgiu,, **for the CMS** Collaboration arXiv:0909.4894 [hep-ex].

[205] D. Krohn, J. Thaler, and L.-T. Wang, *JHEP* **06** (2009) 059, [arXiv:0903.0392 [hep-ph]].

[206] L. G. Almeida, S. J. Lee, G. Perez, I. Sung, and J. Virzi, *Phys. Rev.* **D79** (2009) 074012, [arXiv:0810.0934 [hep-ph]].

[207] L. G. Almeida *et. al.*, *Phys. Rev.* **D79** (2009) 074017, [arXiv:0807.0234 [hep-ph]].

[208] G. P. Salam, arXiv:0906.1833 [hep-ph].

[209] J. Gallicchio and M. D. Schwartz, arXiv:1001.5027 [hep-ph].

[210] K. Hagiwara *et. al.*, *Phys. Rev.* **D73** (2006) 055005, [arXiv:hep-ph/0512260].

[211] W. Kilian, T. Ohl, and J. Reuter, arXiv:0708.4233.

[212] F. Bloch and A. Nordsieck, *Phys. Rev.* **52** (1937) 54–59.

[213] T. Kinoshita, *J. Math. Phys.* **3** (1962) 650–677.

[214] T. D. Lee and M. Nauenberg, *Phys. Rev.* **133** (1964) B1549–B1562.

[215] A. Denner, S. Dittmaier, M. Roth, and D. Wackeroth, *Nucl. Phys.* **B560** (1999) 33–65, [arXiv:hep-ph/9904472].

[216] W. Beenakker *et. al.*, arXiv:hep-ph/9602351.

[217] U. Baur and D. Zeppenfeld, *Phys. Rev. Lett.* **75** (1995) 1002–1005, [arXiv:hep-ph/9503344].

[218] U. Baur and D. Wackeroth, *Phys. Rev.* **D70** (2004) 073015, [hep-ph/0405191].

[219] W. Beenakker *et. al.*, *Nucl. Phys.* **B500** (1997) 255–298, [hep-ph/9612260].

[220] M. A. Gigg and P. Richardson, arXiv:0805.3037 [hep-ph].

[221] J. C. Collins, *Nucl. Phys.* **B304** (1988) 794.

[222] I. G. Knowles, *Comput. Phys. Commun.* **58** (1990) 271–284.

[223] I. G. Knowles, *Nucl. Phys.* **B310** (1988) 571.

[224] P. Richardson, *JHEP* **11** (2001) 029, [hep-ph/0110108].

[225] F. Ambroglini *et. al.*, arXiv:0902.0293 [hep-ph].

[226] B. R. Webber, *Ann. Rev. Nucl. Part. Sci.* **36** (1986) 253–286.

[227] M. Cacciari and S. Catani, *Nucl. Phys.* **B617** (2001) 253–290, [arXiv:hep-ph/0107138].

[228] E. Barberio and Z. Was, *Comput. Phys. Commun.* **79** (1994) 291–308.

[229] E. Barberio, B. van Eijk, and Z. Was, *Comput. Phys. Commun.* **66** (1991) 115–128.

[230] P. Golonka and Z. Was, *Eur. Phys. J.* **C45** (2006) 97–107, [arXiv:hep-ph/0506026].

[231] P. Golonka and Z. Was, *Eur. Phys. J.* **C50** (2007) 53–62, [arXiv:hep-ph/0604232].

[232] G. Nanava and Z. Was, *Eur. Phys. J.* **C51** (2007) 569–583, [arXiv:hep-ph/0607019].

[233] D. R. Yennie, S. C. Frautschi, and H. Suura, *Ann. Phys.* **13** (1961) 379–452.

[234] M. Kaku,. New York, USA: Oxford Univ. Pr. (1993) 785 p.

[235] M. E. Peskin and D. V. Schroeder,. Reading, USA: Addison-Wesley (1995) 842 p.

[236] S. Jadach,. MPI-PAE/PTh 6/87.

[237] B. F. L. Ward, *Phys. Rev.* **D36** (1987) 939.

[238] S. Jadach and B. F. L. Ward, *Phys. Rev.* **D38** (1988) 2897.

[239] S. Jadach, W. Placzek, M. Skrzypek, and B. F. L. Ward, *Phys. Rev.* **D54** (1996) 5434–5442, [hep-ph/9606429].

[240] S. Jadach, B. F. L. Ward, and Z. Was, *Comput. Phys. Commun.* **130** (2000) 260–325, [hep-ph/9912214].

[241] S. Jadach, B. F. L. Ward, and Z. Was, *Phys. Rev.* **D63** (2001) 113009, [hep-ph/0006359].

[242] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, *JHEP* **01** (2007) 013, [arXiv:hep-ph/0611129].

[243] S. Gieseke, P. Stephens, and B. Webber, *JHEP* **12** (2003) 045, [hep-ph/0310083].

[244] K. Hamilton and P. Richardson, *JHEP* **02** (2007) 069, [arXiv:hep-ph/0612236].

[245] K. Hamilton and P. Richardson, *JHEP* **07** (2006) 010, [arXiv:hep-ph/0603034].

[246] M. Bengtsson and T. Sjostrand, *Nucl. Phys.* **B289** (1987) 810.

[247] E. Norrbin and T. Sjöstrand, *Nucl. Phys.* **B603** (2001) 297–342, [arXiv:hep-ph/0010012].

[248] F. Krauss, A. Schalicke, and G. Soff, *Comput. Phys. Commun.* **174** (2006) 876–902, [arXiv:hep-ph/0503087].

[249] M. Schonherr and F. Krauss, *JHEP* **12** (2008) 018, [arXiv:0810.5071 [hep-ph]].

[250] G. Brooijmans,. ATL-PHYS-CONF-2008-008.

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

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

[253] G. Aad *et. al.*,, **ATLAS collaboration** Collaboration *JINST* **3** (2008) S08003.

[254] S. Abachi *et. al.*,, **D0** Collaboration *Phys. Rev. Lett.* **72** (1994) 2332–2336.

[255] D. L. Rainwater, D. Zeppenfeld, and K. Hagiwara, *Phys. Rev.* **D59** (1999) 014037, [arXiv:hep-ph/9808468].

[256] M. Cacciari, G. Salam, and G. Soyez, *JHEP* **0804** (2008) 005, [arXiv:0802.1188 [hep-ph]].

[257] M. Campanelli and J. W. Monk, arXiv:0910.5108 [hep-ph].

[258] A. Djouadi, J. Kalinowski, and M. Spira, *Comput. Phys. Commun.* **108** (1998) 56–74, [arXiv:hep-ph/9704448].

[259] M. Spira, *Fortsch. Phys.* **46** (1998) 203–284, [arXiv:hep-ph/9705337].

[260] A. Djouadi, *Phys. Rept.* **459** (2008) 1–241, [arXiv:hep-ph/0503173].

[261] A. Djouadi, M. Spira, and P. M. Zerwas, *Z. Phys.* **C70** (1996) 427–434, [arXiv:hep-ph/9511344].

[262] A. Djouadi, J. Kalinowski, and P. M. Zerwas, *Z. Phys.* **C70** (1996) 435–448, [arXiv:hep-ph/9511342].

[263] M. S. Carena, M. Quiros, and C. E. M. Wagner, *Nucl. Phys.* **B461** (1996) 407–436, [arXiv:hep-ph/9508343].

[264] H. E. Haber, R. Hempfling, and A. H. Hoang, *Z. Phys.* **C75** (1997) 539–554, [arXiv:hep-ph/9609331].

[265] M. S. Carena *et. al.*, *Nucl. Phys.* **B580** (2000) 29–57, [arXiv:hep-ph/0001002].

[266] G. Degrassi, S. Heinemeyer, W. Hollik, P. Slavich, and G. Weiglein, *Eur. Phys. J.* **C28** (2003) 133–143, [arXiv:hep-ph/0212020].

[267] A. Djouadi, J. Kalinowski, and P. M. Zerwas, *Z. Phys.* **C57** (1993) 569–584.

[268] A. Djouadi, P. Janot, J. Kalinowski, and P. M. Zerwas, *Phys. Lett.* **B376** (1996) 220–226, [arXiv:hep-ph/9603368].

[269] A. Djouadi, J. Kalinowski, P. Ohmann, and P. M. Zerwas, *Z. Phys.* **C74** (1997) 93–111, [arXiv:hep-ph/9605339].

[270] A. Djouadi and M. Drees, *Phys. Lett.* **B407** (1997) 243–249, [arXiv:hep-ph/9703452].

[271] A. Dabelstein, *Nucl. Phys.* **B456** (1995) 25–56, [arXiv:hep-ph/9503443].

[272] J. A. Coarasa Perez, R. A. Jimenez, and J. Sola, *Phys. Lett.* **B389** (1996) 312–320, [arXiv:hep-ph/9511402].

[273] M. S. Carena, D. Garcia, U. Nierste, and C. E. M. Wagner, *Nucl. Phys.* **B577** (2000) 88–120, [arXiv:hep-ph/9912516].

[274] J. Guasch, P. Hafliger, and M. Spira, *Phys. Rev.* **D68** (2003) 115001, [arXiv:hep-ph/0305101].

[275] D. Noth and M. Spira, *Phys. Rev. Lett.* **101** (2008) 181801, [arXiv:0808.0087 [hep-ph]].

[276] D. Noth and M. Spira, arXiv:1001.1935 [hep-ph].

[277] A. Djouadi, M. Spira, J. J. van der Bij, and P. M. Zerwas, *Phys. Lett.* **B257** (1991) 187–190.

[278] A. Djouadi, M. Spira, and P. M. Zerwas, *Phys. Lett.* **B311** (1993) 255–260, [arXiv:hep-ph/9305335].

[279] K. Melnikov and O. I. Yakovlev, *Phys. Lett.* **B312** (1993) 179–183, [arXiv:hep-ph/9302281].

[280] M. Inoue, R. Najima, T. Oka, and J. Saito, *Mod. Phys. Lett.* **A9** (1994) 1189–1194.

[281] M. Spira, A. Djouadi, D. Graudenz, and P. M. Zerwas, *Nucl. Phys.* **B453** (1995) 17–82, [arXiv:hep-ph/9504378].

[282] M. Muhlleitner and M. Spira, *Nucl. Phys.* **B790** (2008) 1–27, [arXiv:hep-ph/0612254].

[283] A. Bredenstein, A. Denner, S. Dittmaier, and M. M. Weber, *Phys. Rev.* **D74** (2006) 013004, [arXiv:hep-ph/0604011].

[284] A. Bredenstein, A. Denner, S. Dittmaier, and M. M. Weber, *JHEP* **02** (2007) 080, [arXiv:hep-ph/0611234].

[285] A. Pukhov *et. al.*, arXiv:hep-ph/9908288.

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

[287] A. Pukhov, arXiv:hep-ph/0412191.

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

[289] J. Alwall *et. al.*, *AIP Conf. Proc.* **1078** (2009) 84–89, [arXiv:0809.2410 [hep-ph]].

[290] M. Moretti, T. Ohl, and J. Reuter, arXiv:hep-ph/0102195.

[291] N. D. Christensen and C. Duhr, *Comput. Phys. Commun.* **180** (2009) 1614–1641, [arXiv:0806.4194 [hep-ph]].

[292] T. Hahn and M. Perez-Victoria, *Comput. Phys. Commun.* **118** (1999) 153–165, [arXiv:hep-ph/9807565].

[293] J. A. M. Vermaseren, arXiv:math-ph/0010025.

[294] D. Fliegner, A. Retey, and J. A. M. Vermaseren, arXiv:hep-ph/9906426.

[295] D. Fliegner, A. Retey, and J. A. M. Vermaseren, arXiv:hep-ph/0007221.

[296] M. Tentyukov *et. al.*, arXiv:cs/0407066.

[297] M. Tentyukov and J. A. M. Vermaseren, arXiv:hep-ph/0702279.

[298] A. Semenov, *Comput. Phys. Commun.* **180** (2009) 431–454, [arXiv:0805.0555 [hep-ph]].

[299] N. D. Christensen *et. al.*, arXiv:0906.2474 [hep-ph].

[300] F. website http://feynrules.phys.ucl.ac.be,.

[301] J. Rosiek, *Phys. Rev.* **D41** (1990) 3464.

[302] J. Rosiek, arXiv:hep-ph/9511250.

[303] H. E. Haber and G. L. Kane, *Phys. Rept.* **117** (1985) 75–263.

[304] J. F. Gunion and H. E. Haber, *Nucl. Phys.* **B272** (1986) 1.

[305] G. Bozzi, B. Fuks, B. Herrmann, and M. Klasen, *Nucl. Phys.* **B787** (2007) 1–54, [arXiv:0704.1826 [hep-ph]].

[306] B. Fuks, B. Herrmann, and M. Klasen, *Nucl. Phys.* **B810** (2009) 266–299, [arXiv:0808.1104 [hep-ph]].

[307] B. C. Allanach *et. al.*, *Eur. Phys. J.* **C25** (2002) 113–123, [arXiv:hep-ph/0202233].

[308] L. Basso, A. Belyaev, S. Moretti, and C. H. Shepherd-Themistocleous, *Phys. Rev.* **D80** (2009) 055030, [arXiv:0812.4313 [hep-ph]].

[309] E. Accomando *et. al.*, arXiv:hep-ph/0608079.

[310] J. A. Aguilar-Saavedra *et. al.*, *Eur. Phys. J.* **C46** (2006) 43–60, [arXiv:hep-ph/0511344].

[311] J. Reuter and F. Braam, arXiv:0909.3059 [hep-ph].

[312] G. D. Kribs, E. Poppitz, and N. Weiner, *Phys. Rev.* **D78** (2008) 055010, [arXiv:0712.2039 [hep-ph]].

[313] G. D. Kribs, A. Martin, and T. S. Roy, *JHEP* **01** (2009) 023, [arXiv:0807.4936 [hep-ph]].

[314] G. D. Kribs, A. Martin, and T. S. Roy, *JHEP* **06** (2009) 042, [arXiv:0901.4105 [hep-ph]].

[315] T. Plehn and T. M. P. Tait, *J. Phys.* **G36** (2009) 075001, [arXiv:0810.3919 [hep-ph]].

[316] A. E. Blechman, *Mod. Phys. Lett.* **A24** (2009) 633–646, [arXiv:0903.2822 [hep-ph]].