LU TP 00–30
hep-ph/0010017
October 2000

# High-Energy-Physics Event Generation with PYTHIA 6.1

Torbjörn Sjöstrand[1], Patrik Edén[2], Christer Friberg[1],
Leif Lönnblad[1], Gabriela Miu[1], Stephen Mrenna[3]
and Emanuel Norrbin[1]

1) Department of Theoretical Physics, Lund University,
Sölvegatan 14A, S-223 62 Lund, Sweden

2) NORDITA, Blegdamsvej 17, DK-2100 Copenhagen, Denmark

3) Physics Department, University of California at Davis,
One Shields Avenue, Davis, CA 95616, USA

**Abstract**

PYTHIA version 6 represents a merger of the PYTHIA 5, JETSET 7 and SPYTHIA programs, with many improvements. It can be used to generate high-energy-physics 'events', i.e. sets of outgoing particles produced in the interactions between two incoming particles. The objective is to provide as accurate as possible a representation of event properties in a wide range of reactions. The underlying physics is not understood well enough to give an exact description; the programs therefore contain a combination of analytical results and various models. The emphasis in this article is on new aspects, but a few words of general introduction are included. Further documentation is available on the web.

CPC subject index: 11.2
PACS codes: 13.60.-r, 13.65.+i, 13.85.-t, 12.15.-y, 12.38.-t, 12.60.-i
Keywords: event generators, multiparticle production

## New Version Summary

*Title of program:* Pythia
*Version number:* 6.154

*Catalogue identifier:* —

*Distribution format:* uuencoded compressed tar file

*References to most recent previous versions:* Computer Physics Communications **82** (1994) 74 and Computer Physics Communications **101** (1997) 232, respectively

*Catalogue identifiers of most recent previous versions:* ACTU

*Authors of most recent previous versions:* T. Sjöstrand, and S. Mrenna, respectively

*Does the new version supersede the previous versions?:* yes

*Computers for which the new program is designed and others on which it has been tested:*
*Computer:* DELL Precision 210 and any other machine with a Fortran 77 compiler
*Installation:* Lund University

*Operating system under which the new program has been tested:* Red Hat Linux 6.2

*Programming language used:* Fortran 77; is also fully compatible with Fortran 90, i.e. does not make use of any obsolescent features of the Fortran 90 standard

*Memory required to execute with typical data:* about 800 kwords

*No. of bits in word:* 32 (double precision real uses two words)

*No. of processors used:* 1

*Has the code been vectorized or parallelized?:* no

*No. of bytes in distributed program:* about 1.8 Mb

*Keywords:* QCD, standard model, beyond standard model, hard scattering, $e^+e^-$ annihilation, leptoproduction, photoproduction, hadronic processes, high-$p_\perp$ scattering, prompt photons, gauge bosons, Higgs physics, parton distribution functions, jet production, parton showers, fragmentation, hadronization, beam remnants, multiple interactions, particle decays, event measures

*Nature of physical problem:* high-energy collisions between elementary particles normally give rise to complex final states, with large multiplicities of hadrons, leptons, neutrinos and photons. The relation between these final states and the underlying physics description is not a simple one, for two main reasons. Firstly, we do not even in principle have a complete understanding of the physics. Secondly, any analytical approach is made intractable by the large multiplicities.

*Method of solution:* complete events are generated by Monte Carlo methods. The complexity is mastered by a subdivision of the full problem into a set of simpler separate tasks. All main aspects of the events are simulated, such as hard-process selection, initial- and final-state radiation, beam remnants, fragmentation, decays, and so on. Therefore events should be directly comparable with experimentally observable ones. The programs can be used to extract physics from comparisons with existing data, or to study physics at future experiments.

*Restrictions on the complexity of the problem:* depends on the problem studied

*Typical running time:* 10–1000 events per second, depending on process studied

*Unusual features of the program:* none

# 1 Introduction

The Pythia and Jetset programs [1] are frequently used for event generation in high-energy physics. The emphasis is on multiparticle production in collisions between elementary particles. This in particular means hard interactions in $e^+e^-$, pp and ep colliders, although also other applications are envisaged. The programs can be used both to compare with existing data, for physics studies or detector corrections, and to explore possibilities at present or future experiments. The programs are intended to generate complete events, in as much detail as experimentally observable ones, within the bounds of our current understanding of the underlying physics. The quantum mechanical variability between events in nature is here replaced by Monte Carlo methods, to obtain 'correctly' both the average behaviour and the amount of fluctuations. Many of the components of the programs represent original research, in the sense that models have been developed and implemented for a number of aspects not covered by standard theory.

Although originally conceived separately, the Pythia [2] and Jetset [3] programs today are so often used together that they have here been joined under the Pythia header. To this has been added the code of SPythia [4], an extension of Pythia that also covers the generation of supersymmetric processes. The current article is not intended to give a complete survey of all the program elements or all the physics — we refer to the long manual and further documentation on the Pythia web page

`http://www.thep.lu.se/∼torbjorn/Pythia.html`

for this. Instead the objective is to give a survey of new features since the latest official publications, with some minimal amount of background material to tie the story together. Many of the advances are related to physics studies, which are further described in separate articles. Others are of a more technical nature, or have been of too limited a scope to result in individual publications (so far).

Some of the main topics are:

- An improved simulation of supersymmetric physics, with several new processes.
- Many new processes of beyond-the-standard-model physics, in areas such as technicolor and doubly-charged Higgses.
- An expanded description of QCD processes in virtual-photon interactions, combined with a new machinery for the flux of virtual photons from leptons.
- Initial-state parton showers are matched to the next-to-leading order matrix elements for gauge boson production.
- Final-state parton showers are matched to a number of different first-order matrix elements for gluon emission, including full mass dependence.
- The hadronization description of low-mass strings has been improved, with consequences especially for heavy-flavour production.
- An alternative baryon production model has been introduced.
- Colour rearrangement is included as a new option, and several alternative Bose-Einstein descriptions are added.

Many further examples will be given. In the process, the total size of the program code has almost doubled in the six years since the previous main publication.

The report is subdivided so that the physics news are highlighted in section 2 and the programming ones (plus a few more physics ones) in section 3. Section 4 contains some concluding remarks and an outlook.

# 2 Physics News

For the description of a typical high-energy event, an event generator should contain a simulation of several physics aspects. If we try to follow the evolution of an event in some semblance of a time order, one may arrange these aspects as follows:

1. Initially two beam particles are coming in towards each other. Normally each particle is characterized by a set of parton distributions, which defines the partonic substructure in terms of flavour composition and energy sharing.
2. One shower initiator parton from each beam starts off a sequence of branchings, such as $\mathrm{q} \to \mathrm{qg}$, which build up an initial-state shower.
3. One incoming parton from each of the two showers enters the hard process, where then a number of outgoing partons are produced, usually two. It is the nature of this process that determines the main characteristics of the event.
4. The hard process may produce a set of short-lived resonances, like the $\mathrm{Z}^0/\mathrm{W}^{\pm}$ gauge bosons, whose decay to normal partons has to be considered in close association with the hard process itself.
5. The outgoing partons may branch as well, to build up final-state showers.
6. In addition to the hard process considered above, further semihard interactions may occur between the other partons of two incoming hadrons.
7. When a shower initiator is taken out of a beam particle, a beam remnant is left behind. This remnant may have an internal structure, and a net colour charge that relates it to the rest of the final state.
8. The QCD confinement mechanism ensures that the outgoing quarks and gluons are not observable, but instead fragment to colour neutral hadrons.
9. Normally the fragmentation mechanism can be seen as occurring in a set of separate colour singlet subsystems, but interconnection effects such as colour rearrangement or Bose–Einstein may complicate the picture.
10. Many of the produced hadrons are unstable and decay further.

In the following subsections, we will survey updates of the above aspects, not in the same order as given here, but rather in the order in which they appear in the program execution, i.e. starting with the hard process.

## 2.1 Physics Subprocesses

### 2.1.1 Process Classifications

Pythia contains a rich selection of physics scenarios, with well above 200 different subprocesses, see Tables 1, 2 and 3. The process number space has tended to become a bit busy, so processes are not always numbered logically. Some processes are closely related variants of the same basic process (e.g. the production of a neutralino pair in processes 216–225), others are alternative formulations (e.g. $\mathrm{Z}^0\mathrm{Z}^0 \to \mathrm{h}^0$ has to be convoluted with the flux of $\mathrm{Z}^0$'s around fermions, and thus is an approximation to $\mathrm{f}_i\mathrm{f}_j \to \mathrm{f}_i\mathrm{f}_j\mathrm{h}^0$; when the latter was implemented the former was still kept). A process may also have hidden further layers of processes (e.g. $\mathrm{H}^{\pm}$ can be produced in top decays, whichever way top is produced).

One classification of the subprocesses is according to the physics scenario. The following major groups may be distinguished:

- Hard QCD processes, i.e. leading to jet production.

Table 1: Subprocesses, part 1: standard model, according to the subprocess numbering of Pythia.'f' denotes a fermion (quark or lepton), 'Q' a heavy quark and 'F' a heavy fermion.

| No. | Subprocess |
|---|---|
| Hard QCD processes: | |
| 11 | $\mathrm{f}_i\mathrm{f}_j \to \mathrm{f}_i\mathrm{f}_j$ |
| 12 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \mathrm{f}_k\overline{\mathrm{f}}_k$ |
| 13 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \mathrm{gg}$ |
| 28 | $\mathrm{f}_i\mathrm{g} \to \mathrm{f}_i\mathrm{g}$ |
| 53 | $\mathrm{gg} \to \mathrm{f}_k\overline{\mathrm{f}}_k$ |
| 68 | $\mathrm{gg} \to \mathrm{gg}$ |
| Soft QCD processes: | |
| 91 | elastic scattering |
| 92 | single diffraction ($XB$) |
| 93 | single diffraction ($AX$) |
| 94 | double diffraction |
| 95 | low-$p_\perp$ production |
| Open heavy flavour: (also fourth generation) | |
| 81 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \mathrm{Q}_k\overline{\mathrm{Q}}_k$ |
| 82 | $\mathrm{gg} \to \mathrm{Q}_k\overline{\mathrm{Q}}_k$ |
| 83 | $\mathrm{q}_i\mathrm{f}_j \to \mathrm{Q}_k\mathrm{f}_l$ |
| 84 | $\mathrm{g}\gamma \to \mathrm{Q}_k\overline{\mathrm{Q}}_k$ |
| 85 | $\gamma\gamma \to \mathrm{F}_k\overline{\mathrm{F}}_k$ |
| Closed heavy flavour: | |
| 86 | $\mathrm{gg} \to \mathrm{J}/\psi\mathrm{g}$ |
| 87 | $\mathrm{gg} \to \chi_{0\mathrm{c}}\mathrm{g}$ |
| 88 | $\mathrm{gg} \to \chi_{1\mathrm{c}}\mathrm{g}$ |
| 89 | $\mathrm{gg} \to \chi_{2\mathrm{c}}\mathrm{g}$ |
| 104 | $\mathrm{gg} \to \chi_{0\mathrm{c}}$ |
| 105 | $\mathrm{gg} \to \chi_{2\mathrm{c}}$ |
| 106 | $\mathrm{gg} \to \mathrm{J}/\psi\gamma$ |
| 107 | $\mathrm{g}\gamma \to \mathrm{J}/\psi\mathrm{g}$ |
| 108 | $\gamma\gamma \to \mathrm{J}/\psi\gamma$ |

| No. | Subprocess |
|---|---|
| W/Z production: | |
| 1 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \gamma^*/\mathrm{Z}^0$ |
| 2 | $\mathrm{f}_i\overline{\mathrm{f}}_j \to \mathrm{W}^\pm$ |
| 22 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \mathrm{Z}^0\mathrm{Z}^0$ |
| 23 | $\mathrm{f}_i\overline{\mathrm{f}}_j \to \mathrm{Z}^0\mathrm{W}^\pm$ |
| 25 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \mathrm{W}^+\mathrm{W}^-$ |
| 15 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \mathrm{gZ}^0$ |
| 16 | $\mathrm{f}_i\overline{\mathrm{f}}_j \to \mathrm{gW}^\pm$ |
| 30 | $\mathrm{f}_i\mathrm{g} \to \mathrm{f}_i\mathrm{Z}^0$ |
| 31 | $\mathrm{f}_i\mathrm{g} \to \mathrm{f}_k\mathrm{W}^\pm$ |
| 19 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \gamma\mathrm{Z}^0$ |
| 20 | $\mathrm{f}_i\overline{\mathrm{f}}_j \to \gamma\mathrm{W}^\pm$ |
| 35 | $\mathrm{f}_i\gamma \to \mathrm{f}_i\mathrm{Z}^0$ |
| 36 | $\mathrm{f}_i\gamma \to \mathrm{f}_k\mathrm{W}^\pm$ |
| 69 | $\gamma\gamma \to \mathrm{W}^+\mathrm{W}^-$ |
| 70 | $\gamma\mathrm{W}^\pm \to \mathrm{Z}^0\mathrm{W}^\pm$ |
| Prompt photons: | |
| 14 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \mathrm{g}\gamma$ |
| 18 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \gamma\gamma$ |
| 29 | $\mathrm{f}_i\mathrm{g} \to \mathrm{f}_i\gamma$ |
| 114 | $\mathrm{gg} \to \gamma\gamma$ |
| 115 | $\mathrm{gg} \to \mathrm{g}\gamma$ |
| Deep inelastic scatt.: | |
| 10 | $\mathrm{f}_i\mathrm{f}_j \to \mathrm{f}_i\mathrm{f}_j$ |
| 99 | $\gamma^*\mathrm{f}_i \to f_i$ |
| Photon-induced: | |
| 33 | $\mathrm{f}_i\gamma \to \mathrm{f}_i\mathrm{g}$ |
| 34 | $\mathrm{f}_i\gamma \to \mathrm{f}_i\gamma$ |
| 54 | $\mathrm{g}\gamma \to \mathrm{f}_k\overline{\mathrm{f}}_k$ |
| 58 | $\gamma\gamma \to \mathrm{f}_k\overline{\mathrm{f}}_k$ |

| No. | Subprocess |
|---|---|
| 131 | $\mathrm{f}_i\gamma^*_\mathrm{T} \to \mathrm{f}_i\mathrm{g}$ |
| 132 | $\mathrm{f}_i\gamma^*_\mathrm{L} \to \mathrm{f}_i\mathrm{g}$ |
| 133 | $\mathrm{f}_i\gamma^*_\mathrm{T} \to \mathrm{f}_i\gamma$ |
| 134 | $\mathrm{f}_i\gamma^*_\mathrm{L} \to \mathrm{f}_i\gamma$ |
| 135 | $\mathrm{g}\gamma^*_\mathrm{T} \to \mathrm{f}_i\overline{\mathrm{f}}_i$ |
| 136 | $\mathrm{g}\gamma^*_\mathrm{L} \to \mathrm{f}_i\overline{\mathrm{f}}_i$ |
| 137 | $\gamma^*_\mathrm{T}\gamma^*_\mathrm{T} \to \mathrm{f}_i\overline{\mathrm{f}}_i$ |
| 138 | $\gamma^*_\mathrm{T}\gamma^*_\mathrm{L} \to \mathrm{f}_i\overline{\mathrm{f}}_i$ |
| 139 | $\gamma^*_\mathrm{L}\gamma^*_\mathrm{T} \to \mathrm{f}_i\overline{\mathrm{f}}_i$ |
| 140 | $\gamma^*_\mathrm{L}\gamma^*_\mathrm{L} \to \mathrm{f}_i\overline{\mathrm{f}}_i$ |
| 80 | $\mathrm{q}_i\gamma \to \mathrm{q}_k\pi^\pm$ |
| Light SM Higgs: | |
| 3 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \mathrm{h}^0$ |
| 24 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \mathrm{Z}^0\mathrm{h}^0$ |
| 26 | $\mathrm{f}_i\overline{\mathrm{f}}_j \to \mathrm{W}^\pm\mathrm{h}^0$ |
| 102 | $\mathrm{gg} \to \mathrm{h}^0$ |
| 103 | $\gamma\gamma \to \mathrm{h}^0$ |
| 110 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \gamma\mathrm{h}^0$ |
| 121 | $\mathrm{gg} \to \mathrm{Q}_k\overline{\mathrm{Q}}_k\mathrm{h}^0$ |
| 122 | $\mathrm{q}_i\overline{\mathrm{q}}_i \to \mathrm{Q}_k\overline{\mathrm{Q}}_k\mathrm{h}^0$ |
| 123 | $\mathrm{f}_i\mathrm{f}_j \to \mathrm{f}_i\mathrm{f}_j\mathrm{h}^0$ |
| 124 | $\mathrm{f}_i\mathrm{f}_j \to \mathrm{f}_k\mathrm{f}_l\mathrm{h}^0$ |
| Heavy SM Higgs: | |
| 5 | $\mathrm{Z}^0\mathrm{Z}^0 \to \mathrm{h}^0$ |
| 8 | $\mathrm{W}^+\mathrm{W}^- \to \mathrm{h}^0$ |
| 71 | $\mathrm{Z}^0_\mathrm{L}\mathrm{Z}^0_\mathrm{L} \to \mathrm{Z}^0_\mathrm{L}\mathrm{Z}^0_\mathrm{L}$ |
| 72 | $\mathrm{Z}^0_\mathrm{L}\mathrm{Z}^0_\mathrm{L} \to \mathrm{W}^+_\mathrm{L}\mathrm{W}^-_\mathrm{L}$ |
| 73 | $\mathrm{Z}^0_\mathrm{L}\mathrm{W}^\pm_\mathrm{L} \to \mathrm{Z}^0_\mathrm{L}\mathrm{W}^\pm_\mathrm{L}$ |
| 76 | $\mathrm{W}^+_\mathrm{L}\mathrm{W}^-_\mathrm{L} \to \mathrm{Z}^0_\mathrm{L}\mathrm{Z}^0_\mathrm{L}$ |
| 77 | $\mathrm{W}^\pm_\mathrm{L}\mathrm{W}^\pm_\mathrm{L} \to \mathrm{W}^\pm_\mathrm{L}\mathrm{W}^\pm_\mathrm{L}$ |

- Soft QCD processes, such as diffractive and elastic scattering, and minimum-bias events. Hidden in this class is also process 96, which is used internally for the merging of soft and hard physics, and for the generation of multiple interactions.
- Heavy-flavour production, both open and hidden (i.e. as bound states like the $\mathrm{J}/\psi$). Hadronization of open heavy flavour will be discussed in section 2.4.1. Some new

Table 2: Subprocesses, part 2: beyond the standard model non-SUSY, with notation as above.

| No. | Subprocess |
|---|---|
| BSM Neutral Higgses: | |
| 151 | $f_i\overline{f}_i \to H^0$ |
| 152 | $gg \to H^0$ |
| 153 | $\gamma\gamma \to H^0$ |
| 171 | $f_i\overline{f}_i \to Z^0H^0$ |
| 172 | $f_i\overline{f}_j \to W^\pm H^0$ |
| 173 | $f_if_j \to f_if_jH^0$ |
| 174 | $f_if_j \to f_kf_lH^0$ |
| 181 | $gg \to Q_k\overline{Q}_kH^0$ |
| 182 | $q_i\overline{q}_i \to Q_k\overline{Q}_kH^0$ |
| 156 | $f_i\overline{f}_i \to A^0$ |
| 157 | $gg \to A^0$ |
| 158 | $\gamma\gamma \to A^0$ |
| 176 | $f_i\overline{f}_i \to Z^0A^0$ |
| 177 | $f_i\overline{f}_j \to W^\pm A^0$ |
| 178 | $f_if_j \to f_if_jA^0$ |
| 179 | $f_if_j \to f_kf_lA^0$ |
| 186 | $gg \to Q_k\overline{Q}_kA^0$ |
| 187 | $q_i\overline{q}_i \to Q_k\overline{Q}_kA^0$ |
| Charged Higgs: | |
| 143 | $f_i\overline{f}_j \to H^+$ |
| 161 | $f_ig \to f_kH^+$ |
| Higgs pairs: | |
| 297 | $f_i\overline{f}_j \to H^\pm h^0$ |
| 298 | $f_i\overline{f}_j \to H^\pm H^0$ |
| 299 | $f_i\overline{f}_i \to A^0h^0$ |
| 300 | $f_i\overline{f}_i \to A^0H^0$ |
| 301 | $f_i\overline{f}_i \to H^+H^-$ |

| No. | Subprocess |
|---|---|
| New gauge bosons: | |
| 141 | $f_i\overline{f}_i \to \gamma/Z^0/Z'^0$ |
| 142 | $f_i\overline{f}_j \to W'^+$ |
| 144 | $f_i\overline{f}_j \to R$ |
| Technicolor: | |
| 149 | $gg \to \eta_{tc}$ |
| 191 | $f_i\overline{f}_i \to \rho^0_{tc}$ |
| 192 | $f_i\overline{f}_j \to \rho^+_{tc}$ |
| 193 | $f_i\overline{f}_i \to \omega^0_{tc}$ |
| 194 | $f_i\overline{f}_i \to f_k\overline{f}_k$ |
| 195 | $f_i\overline{f}_j \to f_k\overline{f}_l$ |
| 361 | $f_i\overline{f}_i \to W^+_L W^-_L$ |
| 362 | $f_i\overline{f}_i \to W^\pm_L \pi^\mp_{tc}$ |
| 363 | $f_i\overline{f}_i \to \pi^+_{tc}\pi^-_{tc}$ |
| 364 | $f_i\overline{f}_i \to \gamma\pi^0_{tc}$ |
| 365 | $f_i\overline{f}_i \to \gamma\pi'^0_{tc}$ |
| 366 | $f_i\overline{f}_i \to Z^0\pi^0_{tc}$ |
| 367 | $f_i\overline{f}_i \to Z^0\pi'^0_{tc}$ |
| 368 | $f_i\overline{f}_i \to W^\pm\pi^\mp_{tc}$ |
| 370 | $f_i\overline{f}_j \to W^\pm_L Z^0_L$ |
| 371 | $f_i\overline{f}_j \to W^\pm_L \pi^0_{tc}$ |
| 372 | $f_i\overline{f}_j \to \pi^\pm_{tc} Z^0_L$ |
| 373 | $f_i\overline{f}_j \to \pi^\pm_{tc}\pi^0_{tc}$ |
| 374 | $f_i\overline{f}_j \to \gamma\pi^\pm_{tc}$ |
| 375 | $f_i\overline{f}_j \to Z^0\pi^\pm_{tc}$ |
| 376 | $f_i\overline{f}_j \to W^\pm\pi^0_{tc}$ |
| 377 | $f_i\overline{f}_j \to W^\pm\pi'^0_{tc}$ |

| No. | Subprocess |
|---|---|
| Compositeness: | |
| 146 | $e\gamma \to e^*$ |
| 147 | $dg \to d^*$ |
| 148 | $ug \to u^*$ |
| 167 | $q_iq_j \to d^*q_k$ |
| 168 | $q_iq_j \to u^*q_k$ |
| 169 | $q_i\overline{q}_i \to e^\pm e^{*\mp}$ |
| 165 | $f_i\overline{f}_i(\to \gamma^*/Z^0) \to f_k\overline{f}_k$ |
| 166 | $f_i\overline{f}_j(\to W^\pm) \to f_k\overline{f}_l$ |
| Doubly-charged Higgs: | |
| 341 | $\ell_i\ell_j \to H^{\pm\pm}_L$ |
| 342 | $\ell_i\ell_j \to H^{\pm\pm}_R$ |
| 343 | $\ell^\pm_i\gamma \to H^{\pm\pm}_L e^\mp$ |
| 344 | $\ell^\pm_i\gamma \to H^{\pm\pm}_R e^\mp$ |
| 345 | $\ell^\pm_i\gamma \to H^{\pm\pm}_L \mu^\mp$ |
| 346 | $\ell^\pm_i\gamma \to H^{\pm\pm}_R \mu^\mp$ |
| 347 | $\ell^\pm_i\gamma \to H^{\pm\pm}_L \tau^\mp$ |
| 348 | $\ell^\pm_i\gamma \to H^{\pm\pm}_R \tau^\mp$ |
| 349 | $f_i\overline{f}_i \to H^{++}_L H^{--}_L$ |
| 350 | $f_i\overline{f}_i \to H^{++}_R H^{--}_R$ |
| 351 | $f_if_j \to f_kf_lH^{\pm\pm}_L$ |
| 352 | $f_if_j \to f_kf_lH^{\pm\pm}_R$ |
| Leptoquarks: | |
| 145 | $q_i\ell_j \to L_Q$ |
| 162 | $qg \to \ell L_Q$ |
| 163 | $gg \to L_Q\overline{L}_Q$ |
| 164 | $q_i\overline{q}_i \to L_Q\overline{L}_Q$ |

processes have been added for closed heavy flavour, but we remind that data here are yet not fully understood, and have given rise to models extending on the more conventional Pythia treatment [5].

- W/Z production. A first-order process such as $f_i\overline{f}_j \to gW^\pm$ is now quite accurately modeled by the initial-state shower acting on $f_i\overline{f}_j \to W^\pm$, see section 2.2.1, but the former can still be useful for a dedicated study of the high-$p_\perp$ tail.
- Prompt-photon production.
- Photon-induced processes, including Deep Inelastic Scattering. A completely new

Table 3: Subprocesses, part 3: SUSY, with notation as above. A trailing + on a final state indicates that the charge-conjugated one is included as well.

| No. | Subprocess |
|---|---|
| SUSY: | |
| 201 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\mathrm{e}}_L\tilde{\mathrm{e}}_L^*$ |
| 202 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\mathrm{e}}_R\tilde{\mathrm{e}}_R^*$ |
| 203 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\mathrm{e}}_L\tilde{\mathrm{e}}_R^*+$ |
| 204 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\mu}_L\tilde{\mu}_L^*$ |
| 205 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\mu}_R\tilde{\mu}_R^*$ |
| 206 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\mu}_L\tilde{\mu}_R^*+$ |
| 207 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\tau}_1\tilde{\tau}_1^*$ |
| 208 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\tau}_2\tilde{\tau}_2^*$ |
| 209 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\tau}_1\tilde{\tau}_2^*+$ |
| 210 | $\mathrm{f}_i\overline{\mathrm{f}}_j \to \tilde{\ell}_L\tilde{\nu}_\ell^*+$ |
| 211 | $\mathrm{f}_i\overline{\mathrm{f}}_j \to \tilde{\tau}_1\tilde{\nu}_\tau^*+$ |
| 212 | $\mathrm{f}_i\overline{\mathrm{f}}_j \to \tilde{\tau}_2\tilde{\nu}_\tau{}^*+$ |
| 213 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\nu}_\ell\tilde{\nu}_\ell{}^*$ |
| 214 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\nu}_\tau\tilde{\nu}_\tau^*$ |
| 216 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\chi}_1\tilde{\chi}_1$ |
| 217 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\chi}_2\tilde{\chi}_2$ |
| 218 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\chi}_3\tilde{\chi}_3$ |
| 219 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\chi}_4\tilde{\chi}_4$ |
| 220 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\chi}_1\tilde{\chi}_2$ |
| 221 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\chi}_1\tilde{\chi}_3$ |
| 222 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\chi}_1\tilde{\chi}_4$ |
| 223 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\chi}_2\tilde{\chi}_3$ |
| 224 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\chi}_2\tilde{\chi}_4$ |
| 225 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\chi}_3\tilde{\chi}_4$ |
| 226 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\chi}_1^\pm\tilde{\chi}_1^\mp$ |
| 227 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\chi}_2^\pm\tilde{\chi}_2^\mp$ |
| 228 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\chi}_1^\pm\tilde{\chi}_2^\mp$ |
| 229 | $\mathrm{f}_i\overline{\mathrm{f}}_j \to \tilde{\chi}_1\tilde{\chi}_1^\pm$ |
| 230 | $\mathrm{f}_i\overline{\mathrm{f}}_j \to \tilde{\chi}_2\tilde{\chi}_1^\pm$ |
| 231 | $\mathrm{f}_i\overline{\mathrm{f}}_j \to \tilde{\chi}_3\tilde{\chi}_1^\pm$ |
| 232 | $\mathrm{f}_i\overline{\mathrm{f}}_j \to \tilde{\chi}_4\tilde{\chi}_1^\pm$ |
| 233 | $\mathrm{f}_i\overline{\mathrm{f}}_j \to \tilde{\chi}_1\tilde{\chi}_2^\pm$ |
| 234 | $\mathrm{f}_i\overline{\mathrm{f}}_j \to \tilde{\chi}_2\tilde{\chi}_2^\pm$ |
| 235 | $\mathrm{f}_i\overline{\mathrm{f}}_j \to \tilde{\chi}_3\tilde{\chi}_2^\pm$ |
| 236 | $\mathrm{f}_i\overline{\mathrm{f}}_j \to \tilde{\chi}_4\tilde{\chi}_2^\pm$ |
| 237 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\mathrm{g}}\tilde{\chi}_1$ |
| 238 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\mathrm{g}}\tilde{\chi}_2$ |
| 239 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\mathrm{g}}\tilde{\chi}_3$ |
| 240 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\mathrm{g}}\tilde{\chi}_4$ |
| 241 | $\mathrm{f}_i\overline{\mathrm{f}}_j \to \tilde{\mathrm{g}}\tilde{\chi}_1^\pm$ |
| 242 | $\mathrm{f}_i\overline{\mathrm{f}}_j \to \tilde{\mathrm{g}}\tilde{\chi}_2^\pm$ |
| 243 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\mathrm{g}}\tilde{\mathrm{g}}$ |
| 244 | $\mathrm{gg} \to \tilde{\mathrm{g}}\tilde{\mathrm{g}}$ |
| 246 | $\mathrm{f}_i\mathrm{g} \to \tilde{\mathrm{q}}_{iL}\tilde{\chi}_1$ |
| 247 | $\mathrm{f}_i\mathrm{g} \to \tilde{\mathrm{q}}_{iR}\tilde{\chi}_1$ |
| 248 | $\mathrm{f}_i\mathrm{g} \to \tilde{\mathrm{q}}_{iL}\tilde{\chi}_2$ |
| 249 | $\mathrm{f}_i\mathrm{g} \to \tilde{\mathrm{q}}_{iR}\tilde{\chi}_2$ |
| 250 | $\mathrm{f}_i\mathrm{g} \to \tilde{\mathrm{q}}_{iL}\tilde{\chi}_3$ |
| 251 | $\mathrm{f}_i\mathrm{g} \to \tilde{\mathrm{q}}_{iR}\tilde{\chi}_3$ |
| 252 | $\mathrm{f}_i\mathrm{g} \to \tilde{\mathrm{q}}_{iL}\tilde{\chi}_4$ |
| 253 | $\mathrm{f}_i\mathrm{g} \to \tilde{\mathrm{q}}_{iR}\tilde{\chi}_4$ |
| 254 | $\mathrm{f}_i\mathrm{g} \to \tilde{\mathrm{q}}_{jL}\tilde{\chi}_1^\pm$ |
| 256 | $\mathrm{f}_i\mathrm{g} \to \tilde{\mathrm{q}}_{jL}\tilde{\chi}_2^\pm$ |
| 258 | $\mathrm{f}_i\mathrm{g} \to \tilde{\mathrm{q}}_{iL}\tilde{\mathrm{g}}$ |
| 259 | $\mathrm{f}_i\mathrm{g} \to \tilde{\mathrm{q}}_{iR}\tilde{\mathrm{g}}$ |
| 261 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\mathrm{t}}_1\tilde{\mathrm{t}}_1^*$ |
| 262 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\mathrm{t}}_2\tilde{\mathrm{t}}_2^*$ |
| 263 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\mathrm{t}}_1\tilde{\mathrm{t}}_2^*+$ |
| 264 | $\mathrm{gg} \to \tilde{\mathrm{t}}_1\tilde{\mathrm{t}}_1^*$ |
| 265 | $\mathrm{gg} \to \tilde{\mathrm{t}}_2\tilde{\mathrm{t}}_2^*$ |
| 271 | $\mathrm{f}_i\mathrm{f}_j \to \tilde{\mathrm{q}}_{iL}\tilde{\mathrm{q}}_{jL}$ |
| 272 | $\mathrm{f}_i\mathrm{f}_j \to \tilde{\mathrm{q}}_{iR}\tilde{\mathrm{q}}_{jR}$ |
| 273 | $\mathrm{f}_i\mathrm{f}_j \to \tilde{\mathrm{q}}_{iL}\tilde{\mathrm{q}}_{jR}+$ |
| 274 | $\mathrm{f}_i\overline{\mathrm{f}}_j \to \tilde{\mathrm{q}}_{iL}\tilde{\mathrm{q}}^*_{jL}$ |
| 275 | $\mathrm{f}_i\overline{\mathrm{f}}_j \to \tilde{\mathrm{q}}_{iR}\tilde{\mathrm{q}}^*_{jR}$ |
| 276 | $\mathrm{f}_i\overline{\mathrm{f}}_j \to \tilde{\mathrm{q}}_{iL}\tilde{\mathrm{q}}^*_{jR}+$ |
| 277 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\mathrm{q}}_{jL}\tilde{\mathrm{q}}^*_{jL}$ |
| 278 | $\mathrm{f}_i\overline{\mathrm{f}}_i \to \tilde{\mathrm{q}}_{jR}\tilde{\mathrm{q}}^*_{jR}$ |
| 279 | $\mathrm{gg} \to \tilde{\mathrm{q}}_{iL}\tilde{\mathrm{q}}^*_{iL}$ |
| 280 | $\mathrm{gg} \to \tilde{\mathrm{q}}_{iR}\tilde{\mathrm{q}}^*_{iR}$ |
| 281 | $\mathrm{bq}_i \to \tilde{\mathrm{b}}_1\tilde{\mathrm{q}}_{iL}$ |
| 282 | $\mathrm{bq}_i \to \tilde{\mathrm{b}}_2\tilde{\mathrm{q}}_{iR}$ |
| 283 | $\mathrm{bq}_i \to \tilde{\mathrm{b}}_1\tilde{\mathrm{q}}_{iR} + \tilde{\mathrm{b}}_2\tilde{\mathrm{q}}_{iL}$ |
| 284 | $\mathrm{b}\overline{\mathrm{q}}_i \to \tilde{\mathrm{b}}_1\tilde{\mathrm{q}}^*_{iL}$ |
| 285 | $\mathrm{b}\overline{\mathrm{q}}_i \to \tilde{\mathrm{b}}_2\tilde{\mathrm{q}}^*_{iR}$ |
| 286 | $\mathrm{b}\overline{\mathrm{q}}_i \to \tilde{\mathrm{b}}_1\tilde{\mathrm{q}}^*_{iR} + \tilde{\mathrm{b}}_2\tilde{\mathrm{q}}^*_{iL}$ |
| 287 | $\mathrm{q}_i\overline{\mathrm{q}}_i \to \tilde{\mathrm{b}}_1\tilde{\mathrm{b}}_1^*$ |
| 288 | $\mathrm{q}_i\overline{\mathrm{q}}_i \to \tilde{\mathrm{b}}_2\tilde{\mathrm{b}}_2^*$ |
| 289 | $\mathrm{gg} \to \tilde{\mathrm{b}}_1\tilde{\mathrm{b}}_1^*$ |
| 290 | $\mathrm{gg} \to \tilde{\mathrm{b}}_2\tilde{\mathrm{b}}_2^*$ |
| 291 | $\mathrm{bb} \to \tilde{\mathrm{b}}_1\tilde{\mathrm{b}}_1$ |
| 292 | $\mathrm{bb} \to \tilde{\mathrm{b}}_2\tilde{\mathrm{b}}_2$ |
| 293 | $\mathrm{bb} \to \tilde{\mathrm{b}}_1\tilde{\mathrm{b}}_2$ |
| 294 | $\mathrm{bg} \to \tilde{\mathrm{b}}_1\tilde{\mathrm{g}}$ |
| 295 | $\mathrm{bg} \to \tilde{\mathrm{b}}_2\tilde{\mathrm{g}}$ |
| 296 | $\mathrm{b}\overline{\mathrm{b}} \to \tilde{\mathrm{b}}_1\tilde{\mathrm{b}}_2^*+$ |

machinery for $\gamma^*$p and $\gamma^*\gamma^*$ physics has been constructed here, see section 2.1.3.

- Standard model Higgs production, where the Higgs is reasonably light and narrow, and can therefore still be considered as a resonance.
- Gauge boson scattering processes, such as $\mathrm{W_LW_L} \to \mathrm{W_LW_L}$ (L = longitudinal), when the standard model Higgs is so heavy and broad that resonant and non-resonant contributions have to be considered together.
- Non-standard Higgs particle production, within the framework of a two-Higgs-

doublet scenario with three neutral ($h^0$, $H^0$ and $A^0$) and two charged ($H^\pm$) Higgs states. Normally associated with SUSY (see below), but does not have to be. The Higgs pair production processes were previously hidden in process 141, but are now included explicitly.

- Production of new gauge bosons, such as a $Z'$, $W'$ and R (a horizontal boson, coupling between generations).
- Technicolor production, as an alternative scenario to the standard picture of electroweak symmetry breaking by a fundamental Higgs. Processes 149, 191, 192 and 193 may be considered obsolete, since the other processes now include the decays to the allowed final states of the $\rho_{\mathrm{tc}}^0/\omega_{\mathrm{tc}}^0/\rho_{\mathrm{tc}}^\pm$ bosons, also including interferences with $\gamma/Z^0/W^\pm$. The default scenario is based on [6].
- Compositeness is a possibility not only in the Higgs sector, but may also apply to fermions, e.g. giving $d^*$ and $u^*$ production. At energies below the threshold for new particle production, contact interactions may still modify the standard behaviour; this is implemented not only for processes 165 and 166, but also for 11, 12 and 20.
- Left–right symmetric models give rise to doubly charged Higgs states, in fact one set belonging to the left and one to the right SU(2) gauge group. Decays involve right-handed W's and neutrinos. The existing scenario is based on [7].
- Leptoquark ($L_Q$) production is encountered in some beyond-the-standard-model scenarios.
- Supersymmetry (SUSY) is probably the favourite scenario for physics beyond the standard model. A rich set of processes are allowed, even if one obeys $R$-parity conservation. The supersymmetric machinery and process selection is inherited from SPYTHIA [4], however with many improvements in the event generation chain. Relative to the SPYTHIA process repertoire, the main new additions is sbottom production, where a classification by mass eigenstates is necessary and many Feynman graphs are related to the possibility to have incoming b quarks. Many different SUSY scenarios have been proposed, and the program is flexible enough to allow input from several of these, in addition to the ones provided internally.

Obviously the list is far from exhaustive; it is a major problem to keep up to date with all the new physics scenarios and signals that are proposed and have to be studied. One example of another physics area that has attracted much attention recently is the possibility of extra dimensions on 'macroscopic' scales. Also, a general-purpose program can not be optimized for all kinds of processes. If a generator for some kind of partonic configurations is already available, outside of PYTHIA, there exists the possibility to feed this in for subsequent treatment of showers and hadronization.

### 2.1.2 Parton Distributions

For cross section calculations, the hard partonic cross section has to be convoluted with the parton distributions of the incoming beam particles. The current default is GRV 94L for protons [8] and SaS 1D for real and virtual photons [9]. Some further parameterizations are available in PYTHIA, such as the recent CTEQ 5 proton ones [10], and a much richer repertoire if the PDFLIB library [11] is linked.

### 2.1.3 Photon Physics

Since before, a model for the interactions of real photons is available, i.e. for $\gamma$p and $\gamma\gamma$ events [12]. This has now been improved and extended also to include virtual photons, i.e. $\gamma^*$p and $\gamma^*\gamma^*$ events [13]. It is especially geared towards the transition region of rather small photon virtualities $Q^2 \lesssim 10$ GeV$^2$, where the physics picture is rather complex, while it may be overkill for large $Q^2$, where the picture again simplifies.

Photon interactions are complicated since the photon wave function contains so many components, each with its own interactions. To first approximation, it may be subdivided into a direct and a resolved part. (In higher orders, the two parts can mix, so one has to provide sensible physical separations between the two.) In the former the photon acts as a pointlike particle, while in the latter it fluctuates into hadronic states. These fluctuations are of $\mathcal{O}(\alpha_{\mathrm{em}})$, and so correspond to a small fraction of the photon wave function, but this is compensated by the bigger cross sections allowed in strong-interaction processes. For real photons therefore the resolved processes dominate the total cross section, while the pointlike ones take over for virtual photons.

The fluctuations $\gamma \to \mathrm{q\overline{q}}\,(\to \gamma)$ can be characterized by the transverse momentum $k_\perp$ of the quarks, or alternatively by some mass scale $m \simeq 2k_\perp$, with a spectrum of fluctuations $\propto \mathrm{d}k_\perp^2/k_\perp^2$. The low-$k_\perp$ part cannot be calculated perturbatively, but is instead parameterized by experimentally determined couplings to the lowest-lying vector mesons, $V = \rho^0$, $\omega^0$, $\phi^0$ and J/$\psi$, an ansatz called VMD for Vector Meson Dominance. Parton distributions are defined with a unit momentum sum rule within a fluctuation [9], giving rise to total hadronic cross sections, jet activity, multiple interactions and beam remnants as in hadronic interactions. States at larger $k_\perp$ are called GVMD or Generalized VMD, and their contributions to the parton distribution of the photon are called anomalous. Given a dividing line $k_0 \simeq 0.5$ GeV to VMD states, the parton distributions are perturbatively calculable. The total cross section of a state is not, however, since this involves aspects of soft physics and eikonalization of jet rates. Therefore an ansatz is chosen where the total cross section scales like $k_V^2/k_\perp^2$, where the adjustable parameter $k_V \approx m_\rho/2$ for light quarks. The spectrum of states is taken to extend over a range $k_0 < k_\perp < k_1$, where $k_1$ is identified with the $p_{\perp\mathrm{min}}(s)$ defined in eq. (6) below. There is some arbitrariness in that choice, and for jet rate calculations also contributions to the parton distributions from above this region are included.

If the photon is virtual, it has a reduced probability to fluctuate into a vector meson state, and this state has a reduced interaction probability. This can be modeled by a traditional dipole factor $(m_V^2/(m_V^2+Q^2))^2$ for a photon of virtuality $Q^2$, where $m_V \to 2k_\perp$ for a GVMD state. Putting it all together, the cross section of the GVMD sector then scales like

$$\int_{k_0^2}^{k_1^2} \frac{\mathrm{d}k_\perp^2}{k_\perp^2}\, \frac{k_V^2}{k_\perp^2} \left( \frac{4k_\perp^2}{4k_\perp^2 + Q^2} \right)^2 . \tag{1}$$

A real direct photon in a $\gamma$p collision can interact with the parton content of the proton: $\gamma\mathrm{q} \to \mathrm{qg}$ and $\gamma\mathrm{g} \to \mathrm{q\overline{q}}$. The $p_\perp$ in this collision is taken to exceed $k_1$, in order to avoid double-counting with the interactions of the GVMD states. For a virtual photon the DIS (deeply inelastic scattering) process $\gamma^*\mathrm{q} \to \mathrm{q}$ is also possible, but by gauge invariance its cross section must vanish in the limit $Q^2 \to 0$. At large $Q^2$, the direct processes can instead be considered as the $\mathcal{O}(\alpha_\mathrm{s})$ correction to the lowest-order DIS process. The DIS $\gamma^*$p cross section is here proportional to the structure function $F_2(x, Q^2)$ with the Bjorken $x = Q^2/(Q^2 + W^2)$. Since normal parton distribution parameterizations are frozen below some $Q_0$ scale and therefore do not obey the gauge invariance condition, an ad hoc factor $(Q^2/(Q^2+m_\rho^2))^2$ is introduced for the conversion from the parameterized $F_2(x,Q^2)=\sum e_{\mathrm{q}}^2\,q(x,Q^2)$ to a $\sigma_{\mathrm{DIS}}^{\gamma^*\mathrm{p}}$. In order to avoid double-counting between DIS and direct events, we decide to introduce a requirement $p_\perp > \max(k_1, Q)$ on direct events. In the remaining DIS ones, thus $p_\perp < Q$. The DIS rate should be reduced accordingly, by a Sudakov form factor giving the probability not to have an interaction above scale $Q$, which can be approximated by $\exp(-\sigma_{\mathrm{direct}}^{\gamma^*\mathrm{p}}/\sigma_{\mathrm{DIS}}^{\gamma^*\mathrm{p}})$.

Note that the $Q^2$ dependence of the DIS and direct processes is implemented in the matrix element expressions. This is different from VMD/GVMD, where dipole factors are used to reduce the assumed flux of partons inside a virtual photon relative to those of a real one, but the matrix elements contain no parton virtuality dependence.

After some further minor corrections for double-counting, we arrive at a picture of hadronic $\gamma^*$p events as being composed of four main components: VMD, GVMD, direct and DIS. Most of these in their turn have a complicated internal structure, as we have seen. The $\gamma^*\gamma^*$ collision between two inequivalent photons contains 13 components: four when the VMD and GVMD states interact with each other ('double-resolved'), eight with a direct or DIS photon interaction on a VMD or GVMD state on either side ('single-resolved', including the traditional DIS), and one where two direct photons interact by the process $\gamma^*\gamma^* \to \mathrm{q}\overline{\mathrm{q}}$ ('direct', not to be confused with the direct process of $\gamma^*$p).

Several further aspects can be added to the above machinery. The impact of resolved longitudinal photons is unknown, except that it has to vanish in the limit $Q^2 \to 0$, and can be approximated by some $Q^2$-dependent enhancement of the normal transverse one. For a complete description of ep events or $\mathrm{e}^+\mathrm{e}^-$ two-photon ones, a convolution with the $x$- and $Q^2$-dependent flux of virtual photons inside an electron is also now provided.

### 2.1.4 Supersymmetry

Pythia simulates the Minimal Supersymmetric Standard Model (MSSM), based on an effective Lagrangian of softly-broken Susy with parameters defined at the weak scale, which is typically between $m_{\mathrm{Z}}$ and 1 TeV. The MSSM particle spectrum is minimal in the sense that it includes only the partners of all Standard Model particles (presently without massive neutrinos), a two-Higgs doublet — one Higgs $\mathrm{H_u}$ coupling only to up-type fermions and one $\mathrm{H_d}$ coupling only to down-type fermions — and partners, and the gravitino. Once the parameters of the softly-broken Susy Lagrangian are specified, the interactions are fixed, and the sparticle masses can be calculated [14].

The masses of the scalar partners to fermions, sfermions, depend on soft scalar masses, trilinear couplings, the Higgsino mass $\mu$, and $\tan\beta$, the ratio of Higgs vacuum expectation values $\langle \mathrm{H_u}\rangle/\langle \mathrm{H_d}\rangle$. The masses of the fermion partners to the gauge and Higgs bosons, the neutralinos and charginos, depend on soft gaugino masses, $\mu$, and $\tan\beta$. Finally, the properties of the Higgs scalar sector is calculated from the input pseudoscalar Higgs boson mass $m_{\mathrm{A}}$, $\tan\beta$, $\mu$, trilinear couplings and the sparticle properties in an effective potential approach [15]. Of course, these calculations also depend on SM parameters ($m_{\mathrm{t}}, m_{\mathrm{Z}}, \alpha_{\mathrm{s}}$, etc.). Any modifications to these quantities from virtual MSSM effects are not taken into account. In principle, the sparticle masses also acquire loop corrections that depend on all MSSM masses.

$R$-parity conservation is assumed (at least on the time and distance scale of a typical collider experiment), and only lowest order, sparticle pair production processes are included. Only those processes with $\mathrm{e}^+\mathrm{e}^-$, $\mu^+\mu^-$, or quark and gluon initial states are simulated. Likewise, only $R$-parity conserving decays are allowed, so that one sparticle is stable, either the lightest neutralino, the gravitino, or a sneutrino. SUSY decays of the top quark are included, but all other SM particle decays are unaltered.

Various improvements to the simulation are being implemented in stages. Some of these can have a significant impact on the collider phenomenology. Among these are: the generalization to complex-valued soft SUSY-breaking parameters in the neutralino and chargino sector; the same in the Higgs sector, which removes the possibility of CP-even or CP-odd labels; the calculation of neutralino and chargino decay rates which are accurate for large $\tan\beta$; and matrix element weighting of particle distributions in three-body decays.

### 2.1.5 Strong Dynamics in Electroweak Symmetry Breaking

The simulation of strong dynamics associated with electroweak symmetry breaking in PYTHIA is based on an effective Lagrangian for the lightest resonances of a technicolor (TC)-like model. In TC, the breaking of a chiral symmetry in a new, strongly interacting gauge theory generates the Goldstone bosons necessary for electroweak symmetry breaking. Bound states of technifermions provide a QCD-like spectrum of technipions ($\pi_{\mathrm{tc}}$), technirhos ($\rho_{\mathrm{tc}}$), techniomegas ($\omega_{\mathrm{tc}}$), etc. The mass hierarchies, however, are unlike QCD because of the behavior of the gauge couplings in realistic models of extended TC (ETC). The difficulties of ETC in explaining the top quark mass while suppressing FCNC's is circumvented by the addition of topcolor interactions, which provide the bulk of $m_{\mathrm{t}}$.

In ETC models, hard mass contributions to technipion masses make decays like $\rho_{\mathrm{tc}} \to \pi_{\mathrm{tc}}\pi_{\mathrm{tc}}$ kinematically inaccessible. Instead, decays like $\rho_{\mathrm{tc}}^{\mathrm{ew}} \to \pi_{\mathrm{tc}}^{\mathrm{ew}}\mathrm{W}_{\mathrm{L}}$, for example, dominate, where ew denotes constituent technifermions with only electroweak quantum numbers and $\mathrm{W}_{\mathrm{L}}$ is a longitudinal W bosons. As a result, the ew technirho and techniomega tend to have small total widths.

Effective couplings are derived in the valence technifermion approximation, and the techniparticle decays can be calculated directly [6]. Technirhos and techniomegas are produced through kinematic mixing with gauge bosons, leading to final states containing Standard Model particles and/or pseudo-Goldstone bosons (technipions).

As an additional wrinkle, $\mathrm{SU}_c(3)$ non-singlet states are included along with the coloron of topcolor assisted technicolor. In this case, colored technirhos (and the coloron) can have substantial total widths and enhanced couplings to bottom and top quarks.

## 2.2 QCD Radiation

The matrix-element (ME) and parton-shower (PS) approaches to higher-order QCD corrections both have their advantages and disadvantages. The former offers a systematic expansion in orders of $\alpha_{\mathrm{s}}$, and a powerful machinery to handle multi-parton configurations on the Born level, but loop calculations are tough and lead to messy cancellations at small resolution scales. Resummed matrix elements may circumvent the latter problem for specific quantities, but then do not provide exclusive accompanying events. Parton showers are based on an improved leading-log (almost next-to-leading-log) approximation, and so cannot be accurate for well separated partons, but they offer a simple, process-independent machinery that gives a smooth blending of event classes (by Sudakov form factors) and a sensible match to hadronization. It is therefore natural to try to combine these descriptions, so that ME results are recovered for widely separated partons while the PS sets the sub-jet structure.

For final-state showers in $Z^0 \to q\overline{q}$, where q is assumed essentially massless, such solutions are the standard since long [16], e.g. by letting the shower slightly overpopulate the $q\overline{q}g$ phase space and then using a Monte Carlo veto technique to reduce down to the ME level.

### 2.2.1 Initial-State Showers

A similar technique is now available for the description of initial-state radiation in the production of a single colour-singlet resonance, such as $\gamma^*/Z^0/W^\pm$ [17]. The basic idea is to map the kinematics between the PS and ME descriptions, and to find a correction factor that can be applied to hard emissions in the shower so as to bring agreement with the matrix-element expression. The PYTHIA shower kinematics definitions are based on $Q^2$ as the spacelike virtuality of the parton produced in a branching and $z$ as the factor by which the $\hat{s}$ of the scattering subsystem is reduced by the branching. Some simple algebra then shows that the two $q\overline{q}' \to gW^\pm$ emission rates disagree by a factor

$$R_{q\overline{q}' \to gW}(\hat{s}, \hat{t}) = \frac{(\mathrm{d}\hat{\sigma}/\mathrm{d}\hat{t})_{\mathrm{ME}}}{(\mathrm{d}\hat{\sigma}/\mathrm{d}\hat{t})_{\mathrm{PS}}} = \frac{\hat{t}^2 + \hat{u}^2 + 2m_{\mathrm{W}}^2 \hat{s}}{\hat{s}^2 + m_{\mathrm{W}}^4} , \tag{2}$$

which is always between 1/2 and 1. The shower can therefore be improved in two ways, relative to the old description. Firstly, the maximum virtuality of emissions is raised from $Q^2_{\mathrm{max}} \approx m^2_{\mathrm{W}}$ to $Q^2_{\mathrm{max}} = s$, i.e. the shower is allowed to populate the full phase space. Secondly, the emission rate for the final (which normally also is the hardest) $q \to qg$ emission on each side is corrected by the factor $R(\hat{s}, \hat{t})$ above, so as to bring agreement with the matrix-element rate in the hard-emission region. In the backwards evolution shower algorithm [18], this is the first branching considered.

The other possible $\mathcal{O}(\alpha_{\mathrm{s}})$ graph is $qg \to q'W^\pm$, where the corresponding correction factor is

$$R_{qg \to q'W}(\hat{s}, \hat{t}) = \frac{(\mathrm{d}\hat{\sigma}/\mathrm{d}\hat{t})_{\mathrm{ME}}}{(\mathrm{d}\hat{\sigma}/\mathrm{d}\hat{t})_{\mathrm{PS}}} = \frac{\hat{s}^2 + \hat{u}^2 + 2m_{\mathrm{W}}^2 \hat{t}}{(\hat{s} - m_{\mathrm{W}}^2)^2 + m_{\mathrm{W}}^4} , \tag{3}$$

which lies between 1 and 3. A probable reason for the lower shower rate here is that the shower does not explicitly simulate the $s$-channel graph $qg \to q^* \to q'W$. The $g \to q\overline{q}$ branching therefore has to be preweighted by a factor of 3 in the shower, but otherwise the method works the same as above. Obviously, the shower will mix the two alternative branchings, and the correction factor for a final branching is based on the current type.

The reweighting procedure prompts some other changes in the shower. In particular, $\hat{u} < 0$ translates into a constraint on the phase space of allowed branchings, not previously implemented.

Our published comparisons with data on the $p_{\perp\mathrm{W}}$ spectrum show quite a good agreement with this improved simulation [17]. A worry was that an unexpectedly large primordial $k_\perp$, around 4 GeV, was required to match the data in the low-$p_{\perp\mathrm{W}}$ region. However, at that time we had not realized that the data were not fully unsmeared. The required primordial $k_\perp$ therefore drops by about a factor of two [19].

The method can also be used for initial-state photon emission, e.g. in the process $e^+e^- \to \gamma^*/Z^0$. There the old default $Q^2_{\mathrm{max}} = m^2_{\mathrm{Z}}$ allowed no emission at large $p_\perp$, $p_\perp \gtrsim m_{\mathrm{Z}}$ at LEP2. This is now corrected by the increased $Q^2_{\mathrm{max}} = s$, and using the $R$ of eq. (2) with $m_{\mathrm{W}} \to m_{\mathrm{Z}}$.

The above method does not address the issue of next-to-leading order corrections to the total W cross section, which instead can be studied with more sophisticated matching procedures [20]. Also extensions to other processes can be considered in the future.

There are also some other changes to the initial state radiation algorithm:

- The cut on minimum gluon energy emitted in a branching is modified by an extra factor roughly corresponding to the $1/\gamma$ factor for the boost to the hard subprocess frame. Earlier, when a subsystem was strongly boosted, the minimum energy requirement became quite stringent on the low-energy incoming side, and could cut out much radiation.
- The angular-ordering requirement is now based on ordering $p_\perp/p$ rather than $p_\perp/p_L$, i.e. replacing $\tan\theta$ by $\sin\theta$. Earlier the starting value $(\tan\theta)_{\mathrm{max}} = 10$ could actually be violated by some bona fide emissions for strongly boosted subsystems.
- The $Q^2$ value of the backwards evolution of a heavy quark like c in a proton beam is by force kept above $m_{\mathrm{c}}^2$, so as to ensure that the branching $\mathrm{g} \to \mathrm{c}\overline{\mathrm{c}}$ is not 'forgotten' by evolving $Q^2$ below $Q_0^2$. Thereby the possibility of having a c in the beam remnant proper is eliminated [21]. The procedure is not forced for a photon beam, where charm occurs as part of the valence flavour content.
- For incoming $\mu^\pm$ (or $\tau^\pm$) beams the kinematical variables are better selected to represent the differences in lepton mass, and the lepton-inside-lepton parton distributions are properly defined.

### 2.2.2 Final-State Showers

The traditional final-state shower algorithm in Pythia [16] is based on an evolution in $Q^2 = m^2$, i.e. potential branchings are considered in order of decreasing mass. A branching $a \to bc$ is then characterized by $m_a^2$ and $z = E_b/E_a$. For the process $\gamma^*/\mathrm{Z}^0 \to \mathrm{q}\overline{\mathrm{q}}$, the first gluon emission off both q and $\overline{\mathrm{q}}$ are corrected to the first-order matrix elements for $\gamma^*/\mathrm{Z}^0 \to \mathrm{q}\overline{\mathrm{q}}\mathrm{g}$. (The $\alpha_{\mathrm{s}}$ and the Sudakov form factor are omitted from the comparison, since the shower procedure here attempts to include higher-order effects absent in the first-order matrix elements.)

This matching is well-defined for massless quarks, and was originally used unchanged for massive ones. A first attempt to include massive matrix elements did not compensate for mass effects in the shower kinematics, and therefore came to exaggerate the suppression of radiation off heavy quarks [22]. Now the shower has been modified to solve this issue, and also improved and extended to cover better a host of different reactions [23].

The starting point is the calculation of processes $a \to bc$ and $a \to bc\mathrm{g}$, where the ratio

$$W_{\mathrm{ME}}(x_1, x_2) = \frac{1}{\sigma(a \to bc)} \frac{\mathrm{d}\sigma(a \to bc\mathrm{g})}{\mathrm{d}x_1\,\mathrm{d}x_2} \tag{4}$$

gives the process-dependent differential gluon-emission rate. Here the phase space variables are $x_1 = 2E_b/m_a$ and $x_2 = 2E_c/m_a$, expressed in the rest frame of parton $a$. Using the standard model and the minimal supersymmetric extension thereof as templates, a wide selection of colour and spin structures have been addressed, exemplified by $\mathrm{Z}^0 \to \mathrm{q}\overline{\mathrm{q}}$, $\mathrm{t} \to \mathrm{b}\mathrm{W}^+$, $\mathrm{H}^0 \to \mathrm{q}\overline{\mathrm{q}}$, $\mathrm{t} \to \mathrm{b}\mathrm{H}^+$, $\mathrm{Z}^0 \to \tilde{\mathrm{q}}\overline{\tilde{\mathrm{q}}}$, $\tilde{\mathrm{q}} \to \tilde{\mathrm{q}}'\mathrm{W}^+$, $\mathrm{H}^0 \to \tilde{\mathrm{q}}\overline{\tilde{\mathrm{q}}}$, $\tilde{\mathrm{q}} \to \tilde{\mathrm{q}}'\mathrm{H}^+$, $\tilde{\chi} \to \mathrm{q}\overline{\tilde{\mathrm{q}}}$, $\tilde{\mathrm{q}} \to \mathrm{q}\tilde{\chi}$, $\mathrm{t} \to \tilde{\mathrm{t}}\tilde{\chi}$, $\tilde{\mathrm{g}} \to \mathrm{q}\overline{\tilde{\mathrm{q}}}$, $\tilde{\mathrm{q}} \to \mathrm{q}\tilde{\mathrm{g}}$, and $\mathrm{t} \to \tilde{\mathrm{t}}\tilde{\mathrm{g}}$. The mass ratios $r_1 = m_b/m_a$ and $r_2 = m_c/m_a$ have been kept as free parameters. When allowed, processes have been calculated for an arbitrary mixture of "parities", i.e. without or with a $\gamma_5$ factor, like in the vector/axial vector structure of $\gamma^*/\mathrm{Z}^0$. All the matrix elements are encoded in the new function `PYMAEL(NI,X1,X2,R1,R2,ALPHA)`, where `NI` distinguishes the matrix elements and `ALPHA` is related to the $\gamma_5$ admixture.

In order to match to the singularity structure of the massive matrix elements, the evolution variable $Q^2$ is changed from $m^2$ to $m^2 - m^2_{\mathrm{on-shell}}$, i.e. $1/Q^2$ is the propagator of a massive particle. Furthermore, the $z$ variable of a branching needs to be redefined, which is achieved by reducing the three-momenta of the daughters in the rest frame of the mother. For the shower history $b \to b\mathrm{g}$ this gives a differential probability

$$W_{\mathrm{PS},1}(x_1, x_2) = \frac{\alpha_{\mathrm{s}}}{2\pi} C_F \frac{\mathrm{d}Q^2}{Q^2} \frac{2\,\mathrm{d}z}{1-z} \frac{1}{\mathrm{d}x_1\,\mathrm{d}x_2} = \frac{\alpha_{\mathrm{s}}}{2\pi} C_F \frac{2}{x_3\,(1 + r_2^2 - r_1^2 - x_2)} , \qquad (5)$$

where the numerator $1 + z^2$ of the splitting kernel for $\mathrm{q} \to \mathrm{qg}$ has been replaced by a 2 in the shower algorithm. For a process with only one radiating parton in the final state, such as $\mathrm{t} \to \mathrm{bW}^+$, the ratio $W_{\mathrm{ME}}/W_{\mathrm{PS},1}$ gives the acceptance probability for an emission in the shower. The singularity structure exactly agrees between ME and PS, giving a well-behaved ratio always below unity. If both $b$ and $c$ can radiate, there is a second possible shower history that has to be considered. The matrix element is here split in two parts, one arbitrarily associated with $b \to b\mathrm{g}$ branchings and the other with $c \to c\mathrm{g}$ ones. A convenient choice is $W_{\mathrm{ME},1} = W_{\mathrm{ME}}(1 + r_1^2 - r_2^2 - x_1)/x_3$ and $W_{\mathrm{ME},2} = W_{\mathrm{ME}}(1+r_2^2-r_1^2-x_2)/x_3$, which again gives matching singularity structures in $W_{\mathrm{ME},i}/W_{\mathrm{PS},i}$ and thus a well-behaved Monte Carlo procedure.

Also subsequent emissions of gluons off the primary particles are corrected to $W_{\mathrm{ME}}$. To this end, a reduced-energy system is constructed, which retains the kinematics of the branching under consideration but omits the gluons already emitted, so that an effective three-body shower state can be mapped to an $(x_1, x_2, r_1, r_2)$ set of variables. For light quarks this procedure is almost equivalent with the original one of using the simple universal splitting kernels after the first branching. For heavy quarks it offers an improved modelling of mass effects also in the collinear region.

Some further changes have been introduced, a few minor as default and some more significant ones as non-default options [23]. This includes the description of coherence effects and $\alpha_{\mathrm{s}}$ arguments, in general and more specifically for secondary heavy flavour production by gluon splittings.

Further issues remain to be addressed, e.g. radiation off particles with non-negligible width. In general, however, the new shower should allow an improved description of gluon radiation in many different processes.

## 2.3 Beam Remnants and Multiple Interactions

### 2.3.1 Beam Remnants

In a hadron–hadron collision, the initial-state radiation algorithm reconstructs one shower initiator in each beam, by backwards evolution from the hard scattering. This initiator only takes some fraction of the total beam energy, leaving behind a beam remnant that takes the rest. Since the initiator is coloured, so is the remnant. It is therefore colour-connected to the hard interaction, and forms part of the same fragmenting system. Often the remnant can be complicated, e.g. a g initiator would leave behind a uud proton-remnant system in a colour octet state, which can conveniently be subdivided into a colour triplet quark and a colour antitriplet diquark, each of which are colour-connected to the hard interaction. The energy sharing between these two remnant objects, and their relative transverse momentum, introduces additional nonperturbative degrees of freedom. Some of the default values have recently been updated [21].

One would expect an ep event to have only one beam remnant, and an $e^+e^-$ event none. This is not always correct, e.g. a $\gamma\gamma \to q\overline{q}$ interaction in an $e^+e^-$ event would leave behind the $e^+$ and $e^-$ as beam remnants. The photons may in their turn leave behind remnants.

It is customary to assign a primordial transverse momentum to the shower initiator, to take into account the motion of quarks inside the original hadron, basically as required by the uncertainty principle. A number of the order of $\langle k_\perp \rangle \approx m_p/3 \approx 300$ MeV could therefore be expected. However, in hadronic collisions much higher numbers than that are often required to describe data, typically of the order of 1 GeV [24, 19] if a Gaussian parameterization is used. (This number is now the default.) Thus, an interpretation as a purely nonperturbative motion inside a hadron is difficult to maintain.

Instead a likely culprit is the initial-state shower algorithm. This is set up to cover the region of hard emissions, but may miss out on some of the softer activity, which inherently borders on nonperturbative physics. By default, the shower does not evolve down to scales below $Q_0 = 1$ GeV. Any shortfall in shower activity around or below this cutoff then has to be compensated by the primordial $k_\perp$ source, which thereby largely loses its original meaning.

### 2.3.2 Multiple Interactions

Multiple parton–parton interactions is the concept that, based on the composite nature of hadrons, several parton pairs may interact in a typical hadron–hadron collision [25]. Over the years, evidence for this mechanism has accumulated, such as the recent direct observation by CDF [26]. The occurrences with two parton pairs at reasonably large $p_\perp$ just form the top of the iceberg, however. In the PYTHIA model, most interactions are at lower $p_\perp$, where they are not visible as separate jets but only contribute to the underlying event structure. As such, they are at the origin of a number of key features, like the broad multiplicity distributions, the significant forward–backward multiplicity correlations, and the pedestal effect under jets.

Since the perturbative jet cross section is divergent for $p_\perp \to 0$, it is necessary to regularize it, e.g. by a cut-off at some $p_{\perp\mathrm{min}}$ scale. That such a regularization should occur is clear from the fact that the incoming hadrons are colour singlets — unlike the coloured partons assumed in the divergent perturbative calculations — and that therefore the colour charges should screen each other in the $p_\perp \to 0$ limit. Also other damping mechanisms are possible [27]. Fits to data typically give $p_{\perp\mathrm{min}} \approx 2$ GeV, which then should be interpreted as the inverse of some colour screening length in the hadron.

One key question is the energy-dependence of $p_{\perp\mathrm{min}}$; this may be relevant e.g. for comparisons of jet rates at different Tevatron energies, and even more for any extrapolation to LHC energies. The problem actually is more pressing now than at the time of the original study [25], since nowadays parton distributions are known to be rising more steeply at small $x$ than the flat $xf(x)$ behaviour normally assumed for small $Q^2$ before HERA. This translates into a more dramatic energy dependence of the multiple-interactions rate for a fixed $p_{\perp\mathrm{min}}$.

The larger number of partons also should increase the amount of screening, however, as confirmed by toy simulations [28]. As a simple first approximation, $p_{\perp\mathrm{min}}$ is assumed to increase in the same way as the total cross section, i.e. with some power $\epsilon \approx 0.08$ [29]

that, via reggeon phenomenology, should relate to the behaviour of parton distributions at small $x$ and $Q^2$. Thus the new default in PYTHIA is

$$p_{\perp\mathrm{min}}(s) = (1.9\ \mathrm{GeV}) \left( \frac{s}{1\ \mathrm{TeV}^2} \right)^{0.08} . \tag{6}$$

## 2.4 Fragmentation and Decays

QCD perturbation theory, formulated in terms of quarks and gluons, is valid at short distances. At long distances, QCD becomes strongly interacting and perturbation theory breaks down. In this confinement régime, the coloured partons are transformed into colourless hadrons, a process called either hadronization or fragmentation.

The fragmentation process has yet to be understood from first principles, starting from the QCD Lagrangian. This has left the way clear for the development of a number of different phenomenological models. PYTHIA is intimately connected with string fragmentation, in the form of the time-honoured 'Lund model' [30]. This is the default for all applications. Improvements have been made in some areas, however.

### 2.4.1 Low-Mass Strings

A hadronic event is conventionally subdivided into sets of partons that form separate colour singlets. These sets are represented by strings, that e.g. stretch from a quark end via a number of intermediate gluons to an antiquark end. Three string mass regions may be distinguished for the hadronization.

1. *Normal string fragmentation.* In the ideal situation, each string has a large invariant mass. Then the standard iterative fragmentation scheme [30, 31] works well. In practice, this approach can be used for all strings above some cut-off mass of a few GeV.
2. *Cluster decay.* If a string is produced with a small invariant mass, maybe only two-body final states are kinematically accessible. The traditional iterative Lund scheme is then not applicable. We call such a low-mass string a cluster, and consider it separately from above. In recent program versions, the modeling has now been improved to give a smooth match on to the standard string scheme in the high-cluster-mass limit [21].
3. *Cluster collapse.* This is the extreme case of the above situation, where the string mass is so small that the cluster cannot decay into two hadrons. It is then assumed to collapse directly into a single hadron, which inherits the flavour content of the string endpoints. The original continuum of string/cluster masses is replaced by a discrete set of hadron masses. Energy and momentum then cannot be conserved inside the cluster, but must be exchanged with the rest of the event. This description has also been improved [21].

String systems below a threshold mass are handled by the cluster machinery. In it, an attempt is first made to produce two hadrons, by having the string break in the middle by the production of a new $\mathrm{q}\overline{\mathrm{q}}$ pair, with flavours and hadron spins selected according to the normal string rules. If the sum of the hadron masses is larger than the cluster mass, repeated attempts can be made to find allowed hadrons; the default is two tries. If an allowed set is found, the angular distribution of the decay products in the cluster rest framed is picked isotropically near the threshold, but then gradually more elongated along the string direction, to provide a smooth match to the string description at larger masses.

This also includes a forward–backward asymmetry, so that each hadron is preferentially in the same hemisphere as the respective original quark it inherits.

If the attempts to find two hadrons fail, one single hadron is formed from the given flavour content. The basic strategy thereafter is to exchange some minimal amount of energy and momentum between the collapsing cluster and other string pieces in the neighbourhood. The momentum transfer can be in either direction, depending on whether the hadron is lighter or heavier than the cluster it comes from. When lighter, the excess momentum is split off and put as an extra 'gluon' on the nearest string piece, where 'nearest' is defined by a space–time history-based distance measure. When the hadron is heavier, momentum is instead borrowed from the endpoints of the nearest string piece.

The free parameters of the model can be tuned to data, especially to the significant asymmetries observed between the production of D and $\overline{\mathrm{D}}$ mesons in $\pi^-$p collisions, with hadrons that share some of the $\pi^-$ flavour content very much favoured at large $x_F$ in the $\pi^-$ fragmentation region [32]. These spectra and asymmetries are closely related to the cluster collapse mechanism, and also to other effects of the colour topology of the event ('beam drag') [21]. Also other parameters enter the description, however, such as the effective charm mass and the beam remnant structure.

### 2.4.2 Baryon Production

A new advanced scheme has been introduced for baryon production with the popcorn mechanism [33], plus some minor changes to the older popcorn scheme [34]. These new features currently only appear as options, with the default unchanged, and can be separated into three parts.

Firstly, an improved implementation of SU(6) weights for baryon production. This should not be regarded as a new model, rather a more correct implementation of the old. However, in order to enable the user to see the effects of the SU(6) weighting separately, both procedures are available as different options. The main change is that, if a step $\mathrm{q} \to \mathrm{B} + \overline{\mathrm{q}\mathrm{q}}'$ is SU(6)-rejected, the new try may now instead give a $\mathrm{q} \to \mathrm{M} + \mathrm{q}'$ step (where B stands for baryon, M for meson). The old procedure leads to a slightly faster algorithm and a better interpretation of the input parameter for the diquark-to-quark production rate. However, the probability that a quark will produce a baryon and a antidiquark is then flavour independent, which is not in agreement with the model. Further, for $\mathrm{qq} \to \mathrm{M} + \mathrm{qq}'$, SU(6) symmetry is included in the weights for $\mathrm{qq}'$, while qq is kept with unit probability. The procedures for $\mathrm{qq} \to \mathrm{B} + \overline{\mathrm{q}}'$ and a final joining $\mathrm{qq} + \mathrm{q} \to \mathrm{B}$ are unchanged.

Secondly, a suppression of diquark vertices occuring at small proper times. This is based on a study of the production dynamics of the three quarks that form a baryon. The main experimental consequence is a suppression of the baryon production rate at large momentum fraction. This in particular implies a smaller rate of first-rank light baryon production, while charm and bottom baryons are less affected (since the production proper time is larger for a heavy hadron than a light one of the same momentum). It thereby substitutes and explains the older brute-force possibility to suppress the production of first-rank baryons.

Thirdly, a completely new flavour algorithm for baryons and popcorn mesons, also using the small-proper-time suppression above. While the old popcorn alternative allowed at most one meson to be produced in between the baryon and the antibaryon, the new model allows an arbitrary number. The new flavour model makes explicit use of the popcorn suppression factor $\exp(-2m_\perp M_\perp/\kappa)$, where $m_\perp$ is the transverse mass of the quark creating the colour fluctuation, $M_\perp$ is the total invariant transverse mass of the popcorn meson system, and $\kappa$ is the string tension constant. Thus two parameters, representing the mean $2m_\perp/\kappa$ for light quarks and s-quarks, respectively, govern both diquark and popcorn meson production. A corresponding parameter is introduced for the fragmentation of strings that contain diquarks already from the beginning, i.e. baryon remnants. The new procedure therefore requires far fewer parameters than the old one, and still provides a comparable quality in the description of the various baryon production rates. This was investigated in detail in [35]. (The concluding worry of an "improper treatment" was caused by an unfortunate misunderstanding and can be disregarded.) Other features, such as baryon correlations, are also modified.

Several new routines have been added, and the diquark code has been extended with information about the curtain quark flavour, i.e. the $\mathrm{q\overline{q}}$ pair that is shared between the baryon and antibaryon, but this is not visible externally. Some parameters are no longer used, while others have to be given modified values, as described in the long writeup.

### 2.4.3 Interconnection Effects

The widths of the W, Z and t are all of the order of 2 GeV. A standard model Higgs with a mass above 200 GeV, as well as many supersymmetric and other beyond the standard model particles would also have widths in the multi-GeV range. Not far from threshold, the typical decay times $\tau = 1/\Gamma \approx 0.1\,\mathrm{fm} \ll \tau_{\mathrm{had}} \approx 1\,\mathrm{fm}$. Thus hadronic decay systems overlap, between a resonance and the underlying event, or between pairs of resonances, so that the final state may not contain independent resonance decays.

So far, studies have mainly been performed in the context of W pair production at LEP2. Pragmatically, one may here distinguish three main eras for such interconnection:

1. Perturbative: this is suppressed for gluon energies $\omega > \Gamma$ by propagator/timescale effects; thus only soft gluons may contribute appreciably.
2. Nonperturbative in the hadroformation process: normally modeled by a colour rearrangement between the partons produced in the two resonance decays and in the subsequent parton showers.
3. Nonperturbative in the purely hadronic phase: best exemplified by Bose–Einstein effects.

The above topics are deeply related to the unsolved problems of strong interactions: confinement dynamics, $1/N_C^2$ effects, quantum mechanical interferences, etc. Thus they offer an opportunity to study the dynamics of unstable particles, and new ways to probe confinement dynamics in space and time [36, 37], *but* they also risk to limit or even spoil precision measurements.

The reconnection scenarios outlined in [37] are now available, plus also an option along the lines suggested in [38]. Currently they can only be invoked in process 25, $\mathrm{e^+e^- \to W^+W^- \to q_1\overline{q}_2q_3\overline{q}_4}$, which is the most interesting one for the foreseeable future. (Process 22, $\mathrm{e^+e^-} \to \gamma^*/\mathrm{Z}^0\ \gamma^*/\mathrm{Z}^0 \to \mathrm{q_1\overline{q}_2q_3\overline{q}_4}$ can also be used, but the travel distance is calculated based only on the $\mathrm{Z}^0$ propagator part.) If normally the event is considered as consisting of two separate colour singlets, $\mathrm{q_1\overline{q}_2}$ from the $\mathrm{W}^+$ and $\mathrm{q_3\overline{q}_4}$ from the $\mathrm{W}^-$, a colour rearrangement can give two new colour singlets $\mathrm{q_1\overline{q}_4}$ and $\mathrm{q_3\overline{q}_2}$. It therefore leads to a different hadronic final state, although differences usually turn out to be subtle and difficult to isolate [39]. When also gluon emission is considered, the number of potential reconnection topologies increases. Apart from the overall rate of reconnection, the scenarios in PYTHIA differ in the relative probability assigned to each of these topologies, based on their properties in momentum space and/or space–time. For instance, scenario I is based on an analogy with type I superconductors, with the colour field represented by extended flux tubes. By contrast, scenario II assumes that narrow vortex lines carry all the topological information, like in type II superconductors, even if the full energy is stored over a wider region.

Bose–Einstein effects are simulated in a simplified manner, by introducing small momentum shifts in identical final-state mesons (primarily $\pi^{\pm}$ and $\pi^0$) so as to bring them closer to each other [40]. The shifts can be chosen to reproduce a desired BE enhancement shape for small relative momentum $Q = \sqrt{m_{ij}^2 - 4m_i^2}$ between identical bosons $i$ and $j$. Typically the shape is chosen as a Gaussian, $f_2(Q) = 1 + \lambda \exp(-Q^2R^2)$, with $\lambda$ and $R$ two free parameters. The input is only exactly reproduced in the limit of an isotropic and low-density initial particle distribution; since these conditions are not completely fulfilled in reality, there are distortions [41], for better or worse. (The nontrivial three-particle correlations in data are described qualitatively, although not quantitatively.)

A major shortcoming of the algorithm is that energy is not automatically conserved, even though three-momentum is. In the original algorithm, this was solved by a uniform rescaling of all three-momenta, with undesirable side effects e.g. when studying BE effects in $\mathrm{W}^+\mathrm{W}^-$ hadronic final states. In the current version, several new options have been added that, based on different principles, instead shifts pairs apart. The default one, $\mathrm{BE}_{32}$, operates on identical particles, introducing an extra factor

$$1 + \alpha\lambda \exp(-Q^2R^2/9) \left\{1 - \exp(-Q^2R^2/4)\right\} \tag{7}$$

to $f_2(Q)$. Here $\alpha$ is a negative number adjusted event by event for overall energy conservation, with $\langle\alpha\rangle \approx -0.25$. This scenario can be viewed as a simplified version of a dampened oscillating correlation function, where only the first peak and dip has been retained. Further new options have also been introduced specifically geared towards studies of $\mathrm{W}^+\mathrm{W}^-$ hadronic events, e.g. to include the effects of the separated $\mathrm{W}^+$ and $\mathrm{W}^-$ decay vertices.

### 2.4.4 Decays

Two separate decay treatments exist in PYTHIA. One is making use of a set of tables where branching ratios and decay modes are stored, and is used e.g. for hadronic decays, where branching ratios normally cannot be calculated from first principles.

The other treatment is used for a set of fundamental resonances in or beyond the standard model, such as t, $\mathrm{Z}^0$, $\mathrm{W}^{\pm}$, $\mathrm{h}^0$, supersymmetric particles, and many more. Characteristic here is that these resonances have perturbatively calculable widths to each of their decay channels. The decay products are typically quarks, leptons, or other resonances. In decays to quarks, parton showers are automatically added to give a more realistic multijet structure, and one may also allow photon emission off leptons. If the decay products in turn are resonances, further decays are necessary. Often spin information is available in resonance decay matrix elements, leading to nonisotropic decays. This part has been improved in several processes, but is still missing in many others.

The routine used to calculate the partial and total width of resonances (now expressed in GeV throughout), has been expanded for all the new particles and decay modes introduced. Some alternative calculation schemes have also been adopted, e.g. based on a simple rescaling of the on-shell widths rather than a complete recalculation (which may at times not be feasible) based on the current mass.

The width to be used in the denominator of a resonance propagator is only well-defined near the peak. Well away from the peak, an unfortunate choice may lead to a loss of cancellation between resonant and nonresonant diagrams. A special problem exists for a massive standard model Higgs, where the width $\Gamma_{\mathrm{h}} \propto m_{\mathrm{h}}^3$ is so large that the choice of $\hat{s}$ dependence of the width significantly influences the resonance peak shape. Following [42], the default now is $\Gamma_{\mathrm{h}} \propto m_{\mathrm{h}}^2 \sqrt{\hat{s}}$.

# 3 Program News

Essentially all of the basic philosophy and framework remain from the previous PYTHIA and JETSET versions, so no user familiar with these should feel at loss with PYTHIA 6.1. Most of the changes and additions instead are under the surface, and are only visible as new options added to the existing repertoire. However, some changes are fairly obvious, and other less obvious ones still of general interest. These will be covered in this section, in fairly general terms. Again we refer to the PYTHIA web page for a detailed documentation.

## 3.1 Coding conventions

As before, the FORTRAN 77 standard is adhered to. A very few minor extensions may be used in isolated places, like the 7-character names of the PDFLIB routines [11], but are not known to cause problems on any compiler in use.

An obvious consequence of the PYTHIA/JETSET code merging is that the old JETSET routines and commonblocks have been renamed to begin with `PY` (instead of `LU` or `UL`), just like the PYTHIA ones. In most cases, the rest of the name is unchanged, but there are a few exceptions, mainly `RLU`→`PYR`, `KLU`→`PYK`, `PLU`→`PYP` and `LUXTOT`→`PYXTEE`. Three integer functions now begin with `PY`, namely `PYK`, `PYCHGE` and `PYCOMP`, and therefore have to be declared extra. The `LUDATA` block data has been merged into `PYDATA`, and the test routine `LUTEST` into `PYTEST`. For rotations and boosts, the `PYROBO` routine now requires the range of affected entries to be given, like the old `LUDBRB` but unlike `LUROBO` (but 0,0 as range arguments gives back the old `LUROBO` behaviour).

All real variables are now in `DOUBLE PRECISION`, which is assumed to mean 64 bits, and also real constants have been promoted to the higher precision. This is required to ensure proper functioning at currently studied energies, such as the LHC and beyond. To take into account this, all routines begin with the declarations

```
C...Double precision and integer declarations.
      IMPLICIT DOUBLE PRECISION(A-H, O-Z)
      IMPLICIT INTEGER(I-N)
      INTEGER PYK,PYCHGE,PYCOMP
```

and users should do the same in their main programs.

On a machine where `DOUBLE PRECISION` would give 128 bits, it may make sense to use compiler options to revert to 64 bits, since the program is anyway not constructed to make use of 128 bit precision.

The random number generator is the same as in previous versions [43], but has now been expanded to operate with a 48 bit mantissa for the real numbers.

Table 4: The current form of the main commonblock declarations.

```
COMMON/PYJETS/N,NPAD,K(4000,5),P(4000,5),V(4000,5)
COMMON/PYDAT1/MSTU(200),PARU(200),MSTJ(200),PARJ(200)
COMMON/PYDAT2/KCHG(500,4),PMAS(500,4),PARF(2000),VCKM(4,4)
COMMON/PYDAT3/MDCY(500,3),MDME(4000,2),BRAT(4000),KFDP(4000,5)
COMMON/PYDAT4/CHAF(500,2)
CHARACTER CHAF*16
COMMON/PYDATR/MRPY(6),RRPY(100)
COMMON/PYSUBS/MSEL,MSELPD,MSUB(500),KFIN(2,-40:40),CKIN(200)
COMMON/PYPARS/MSTP(200),PARP(200),MSTI(200),PARI(200)
COMMON/PYINT1/MINT(400),VINT(400)
COMMON/PYINT2/ISET(500),KFPR(500,2),COEF(500,20),ICOL(40,4,2)
COMMON/PYINT3/XSFX(2,-40:40),ISIG(1000,3),SIGH(1000)
COMMON/PYINT4/MWID(500),WIDS(500,5)
COMMON/PYINT5/NGENPD,NGEN(0:500,3),XSEC(0:500,3)
COMMON/PYINT6/PROC(0:500)
CHARACTER PROC*28
COMMON/PYMSSM/IMSS(0:99),RMSS(0:99)
COMMON/PYUPPR/NUP,KUP(20,7),NFUP,IFUP(10,2),PUP(20,5),Q2UP(0:10)
COMMON/PYBINS/IHIST(4),INDX(1000),BIN(20000)
```

FORTRAN 77 makes no provision for double-precision complex numbers, but since `COMPLEX` is used only sparingly, no problems should be expected from this omission. For the technicolor processes, some variables are declared `COMPLEX*16` in the `PYSIGH` routine. Should the compiler not accept this, that one declaration can be changed to `COMPLEX` with some drop in precision for the affected processes.

Several compilers report problems when an odd number of integers precede a double-precision variable in a commonblock. Therefore an extra integer has been introduced as padding in a few instances (`NPAD`, `MSELPD` and `NGENPD` in Table 4).

In order to cater for the increased offering of subprocesses, some arrays in commonblocks have been expanded. A few, such as `PYINT4`, have also been reorganized to represent improvements in the physics modeling. Most commonblocks and commonblock variables are easily recognizable from previous program versions, however. The current complement is given in Table 4, omitting some of the less interesting ones.

Since FORTRAN 77 provides no date-and-time routine, `PYTIME` allows a system-specific routine to be interfaced, with some commented-out examples given in the code. This routine is only used for cosmetic improvements of the output, however, so can be left at the default with time 0 given.

For a program written to run PYTHIA 5 and JETSET 7, most of the conversion required for PYTHIA 6 is fairly straightforward, and can be automatized. Both a simple FORTRAN routine and a more sophisticated PERL [44] script exists to this end. Some manual checks and interventions may still be required.

Table 5: New or modified particle codes or names.

| | |
|---|---|
| **Renamed:** | |
| 7 | $\mathrm{b}'$ |
| 8 | $\mathrm{t}'$ |
| 17 | $\tau'$ |
| 18 | $\nu'_\tau$ |
| 25 | $\mathrm{h}^0$ |
| 35 | $\mathrm{H}^0$ |
| **Moved:** | |
| 100443 | $\psi'$ |
| 100553 | $\Upsilon'$ |
| 4000001 | $\mathrm{d}^*$ |
| 4000002 | $\mathrm{u}^*$ |
| 4000011 | $\mathrm{e}^*$ |
| 4000012 | $\nu^*_\mathrm{e}$ |

| | |
|---|---|
| **Technicolor:** | |
| 51 | $\pi^0_\mathrm{tc}$ |
| 52 | $\pi^+_\mathrm{tc}$ |
| 53 | $\pi'^{\,0}_\mathrm{tc}$ |
| 54 | $\rho^0_\mathrm{tc}$ |
| 55 | $\rho^+_\mathrm{tc}$ |
| 56 | $\omega^0_\mathrm{tc}$ |
| **LR-symmetric:** | |
| 61 | $\mathrm{H}^{++}_L$ |
| 62 | $\mathrm{H}^{++}_R$ |
| 63 | $\mathrm{W}^+_R$ |
| 64 | $\nu_{R\mathrm{e}}$ |
| 65 | $\nu_{R\mu}$ |
| 66 | $\nu_{R\tau}$ |

| | |
|---|---|
| **SUSY:** | |
| 1000001 | $\tilde{\mathrm{d}}_L$ |
| 1000002 | $\tilde{\mathrm{u}}_L$ |
| 1000003 | $\tilde{\mathrm{s}}_L$ |
| 1000004 | $\tilde{\mathrm{c}}_L$ |
| 1000005 | $\tilde{\mathrm{b}}_1$ |
| 1000006 | $\tilde{\mathrm{t}}_1$ |
| 1000011 | $\tilde{\mathrm{e}}_L$ |
| 1000012 | $\tilde{\nu}_{\mathrm{e}L}$ |
| 1000013 | $\tilde{\mu}_L$ |
| 1000014 | $\tilde{\nu}_{\mu L}$ |
| 1000015 | $\tilde{\tau}_1$ |
| 1000016 | $\tilde{\nu}_{\tau L}$ |
| 1000021 | $\tilde{\mathrm{g}}$ |
| 1000022 | $\tilde{\chi}^0_1$ |
| 1000023 | $\tilde{\chi}^0_2$ |
| 1000024 | $\tilde{\chi}^+_1$ |

| | |
|---|---|
| **SUSY:** | |
| 2000001 | $\tilde{\mathrm{d}}_R$ |
| 2000002 | $\tilde{\mathrm{u}}_R$ |
| 2000003 | $\tilde{\mathrm{s}}_R$ |
| 2000004 | $\tilde{\mathrm{c}}_R$ |
| 2000005 | $\tilde{\mathrm{b}}_2$ |
| 2000006 | $\tilde{\mathrm{t}}_2$ |
| 2000011 | $\tilde{\mathrm{e}}_R$ |
| 2000012 | $\tilde{\nu}_{\mathrm{e}R}$ |
| 2000013 | $\tilde{\mu}_R$ |
| 2000014 | $\tilde{\nu}_{\mu R}$ |
| 2000015 | $\tilde{\tau}_2$ |
| 2000016 | $\tilde{\nu}_{\tau R}$ |
| 1000025 | $\tilde{\chi}^0_3$ |
| 1000035 | $\tilde{\chi}^0_4$ |
| 1000037 | $\tilde{\chi}^+_2$ |
| 1000039 | $\tilde{\mathrm{G}}$ |

## 3.2 Particle codes and data

A number of new particle codes `KF` have been introduced, or modified, see Table 5. Mostly this is based on the PDG-agreed conventions [45, 46], but some not yet standardized codes appear in the 'empty' range 41–80. Furthermore, the fourth generation fermions and neutral scalar Higgs states have been renamed. The two fermion spartners are labelled left and right, except in the third generation, where an expected larger mixing makes the two mass eigenstates a better choice of classification.

The top hadrons are gone. It is now known that top is too short-lived to form hadronic bound states, so a reasonable description is instead to have the top quarks decay before hadronization is considered. The same is now assumed about a hypothetical fourth generation. Should the need ever arise in the future to consider a new long-lived coloured object, an effective description of a hadron as a small string with an ordinary colour-matching flavour at the other end should be sufficient. One such example would be leptoquark-hadrons [47].

Bottom hadrons are now defined individually, e.g. the previous common decay scheme is gone in favour of individual branching ratios for each hadron. On the other hand, given the sketchy knowledge of many branching ratios, the default description is still fairly standardized.

Decay data is mainly based on the 1996 PDG edition [48], but with many 'educated guesses' to fill in missing information.

Since running fermion masses are used in an increasing number of processes, e.g. for Higgs couplings, a function `PYMRUN(KF,Q2)` has been introduced to give the mass as a function of $\mathrm{Q}^2$ scale.

The compressed codes, KC = PYCOMP(KF), are completely changed. We remind that KF can range up to seven-digit codes, plus a sign. They therefore cannot be used to directly access information in particle data tables. The KC codes range between 1 and 500, and give the index to the particle data arrays. Each KF code is now one-to-one associated with a KC code; the only ambiguity is that KC does not distinguish antiparticles from particles. Whereas KF codes below 100 still obey KF = KC, the mapping of codes above 100 is completely changed. It is no longer hard-coded in PYCOMP, but defined by the fourth component of the KCHG array. Therefore it can be changed or expanded during the course of a run, either by PYUPDA calls or by direct user intervention.

## 3.3 New Options

A large amount of new options have been added, related to almost all the physics changes above and more, and we here only mention some of the more significant ones.

The inclusion of SUSY processes means that all the SPYTHIA PYMSSM commonblock switches and parameters are inherited. New parameters are added also for other new physics scenarios, such as technicolor and doubly-charged Higgses.

The extensions to the physics of virtual photons, outlined in section 2.1.3, has resulted in two sets of new possibilities. One is in the description of the virtual-photon flux, where new CKIN switches has been introduced, e.g. to set the range of photon $x$ and $Q^2$. This is available when PYINIT is called with 'gamma/lepton' as beam or target, to denote that the photon flux inside the lepton has to be considered as a new administrative layer, also documented in the event record. The other is the new physics machinery. Here the main switch is MSTP(14) that sets the assumed nature of the photon or photons, e.g. 'a direct photon from the left collides with a VMD one from the right'. The default is the most general mixture, meaning 4 components for $\gamma^*\mathrm{p}$ and 13 for $\gamma^*\gamma^*$. This is the relevant approach for studies of QCD processes. There is no corresponding automatic mixing machinery for other processes, so then the relevant contributing components have to be handled separately and added afterwards. Further options are available for several of the components, e.g. the DIS process dampening in the $Q^2 \to 0$ limit, the relative normalization of the GVMD spectrum, the scale choice for parton distributions, and the possibility to add the effects of a longitudinal resolved contribution.

The matrix-element options for $\mathrm{e}^+\mathrm{e}^- \to \gamma^*/\mathrm{Z}^0 \to 2$, 3 or 4 partons have previously only been available via the LUEEVT/PYEEVT routine, that suffers from problems of its own in having a rather old-fashioned machinery for QED initial-state radiation and electroweak parameters. Now the QCD matrix-element description is accessible as an option to the shower default for $\mathrm{e}^+\mathrm{e}^-$ events generated with subprocess 1 of the standard PYTHIA machinery.

## 3.4 Interfaces

While PYTHIA contains an extensive library of subprocesses, it is far from up to all the requirements of the experimental community. Both further processes and a more detailed treatment of the existing ones is required at times. In particular, it is not uncommon with a generator dedicated to one specific process, where also higher-order electroweak corrections, absent in PYTHIA, have been included in the cross section. None of these programs are geared to handle the QCD aspects of parton showers and hadronization, however, so it makes sense to combine the individual strengths.

A generic facility to include external processes exists since long in PYTHIA. Here one can feed in partonic configurations from an external generator, together with some basic information on colour flow and which partons are allowed to radiate, and let PYTHIA construct a complete event based on this information. For the simple configurations encountered in $e^+e^-$ annihilation events, this would often be overkill, since neither the initial-state QCD radiation nor beam-remnant treatment of the generic (hadronic) collision is present.

Based on the concepts presented in the LEP2 workshop [46], a few simpler alternatives are therefore now provided for this kind of tasks:

- `CALL PY2FRM(IRAD,ITAU,ICOM)` allows a parton shower to develop and partons to hadronize from a given two-fermion starting point. `IRAD` sets whether quarks are allowed also to radiate photons or not, `ITAU` whether $\tau$ leptons should be decayed or not, and `ICOM` whether the input and output event record is `HEPEVT` or `PYJETS`. An arbitrary number of photons (e.g. from initial-state radiation) may also be stored with the input.
- `CALL PY4FRM(ATOTSQ,A1SQ,A2SQ,ISTRAT,IRAD,ITAU,ICOM)` allows parton showers to develop and partons to hadronize from a given four-fermion starting point. The extra parameters can be used to select between the two colour pairings allowed for a $q_1\overline{q}_2q_3\overline{q}_4$ state, according to some different strategies when interference terms do not allow unique probabilities to be found.
- `CALL PY6FRM(P12,P13,P21,P23,P31,P32,PTOP,IRAD,ITAU,ICOM)` allows parton showers to develop and partons to hadronize from a given six-fermion starting point. The `Pij` parameters give the relative probabilities for the six colour pairings allowed for a six-quark state, and `PTOP` the probability that the event originates from a $t\overline{t}$ pair (in which case the shower handling has to be different than e.g. in a $Z^0W^+W^-$ event).

The above routines are not set up to handle QCD four-jet events, i.e. events of the types $q\overline{q}gg$ and $q\overline{q}q'\overline{q}'$, with $q'\overline{q}'$ coming from a gluon branching. Such events are generated in normal parton showers, but not necessarily at the right rate (a problem that may be especially interesting for massive quarks like b). Therefore one would like to start a QCD parton shower from a given four-parton configuration. Some time ago, a machinery was developed to handle this kind of occurences [49]. This approach has now been adapted to the current PYTHIA version, in a somewhat modified form. In it, an imagined shower history of two branchings is (re)constructed from the four-parton state, according to relative probabilities derived in the shower language. Thereafter a normal shower is allowed to develop, with branchings chosen at random except for these two predetermined ones. The routine `CALL PY4JET(PMAX,IRAD,ICOM)` takes an original four-parton configuration stored in `HEPEVT` or `PYJETS` and lets a shower develop as described above. `PMAX` can be used to set the maximum virtuality of those parts of the shower not given from the parton configuration itself, either to a fixed value or to the lowest virtuality of the reconstructed shower.

## 3.5 Utilities

The clustering algorithm `PYCLUS` has been extended also to accept the Durham distance measure [50] as an alternative. This is $p_\perp$-based, like the original `LUCLUS` distance measure, but differs in the details.

The `GBOOK` histogramming package was written in 1979 as a lightweight substitute for `HBOOK` [51] before that program was available in FORTRAN 77. The one-dimensional histogram part now appears in the standard distribution, in order to make the sample runs offered on the web a bit more realistic. The main routines are:

- `CALL PYBOOK(ID,TITLE,NX,XL,XU)` to book a one-dimensional histogram with integer identifier `ID` (in the range $1-1000$), character title `TITLE` and `NX` bins stretching from `XL` to `XU`.
- `CALL PYFILL(ID,X,W)` to fill histogram `ID` at position `X` with weight `W`.
- `CALL PYFACT(ID,F)` to rescale the contents of histogram `ID` by a factor `F`.
- `CALL PYOPER(ID1,OPER,ID2,ID3,F1,F2)` to perform operations on several histograms, such as adding or dividing them by each other.
- `CALL PYDUMP(MDUMP,LFN,NHI,IHI)` to dump histogram contents to a file from which they could be read in for plotting in another program.
- `CALL PYHIST` to print all histograms in a simple line-printer mode, and thereafter reset histogram contents.

A commonblock of dimension 20000 is used to store the histograms; this size may need to be expanded if many histograms are to be booked.

The `PYUPDA` routine has been expanded with a new option that allows a set of particle data to be read in, in tabular form as before, as an addition to or partial replacement of the existing particle data.

# 4 Summary and Outlook

We have here given a very brief survey of news in the PYTHIA 6.1 program. A more detailed description of physics and programs is available separately [52]. Any serious user should turn to this publication, and to the original physics papers, for further information.

The treasure trove for information is the PYTHIA webpage,

`http://www.thep.lu.se/~torbjorn/Pythia.html` ,

where one may find the current and previous subversions, with documentation, sample main programs, links to related programs, etc.

The PYTHIA program is continuously being developed. We are aware of many physics shortcomings, which hopefully will be addressed in the future. It is in the nature of a program of this kind never to be finished, at least as long as it is of importance for the high-energy physics experimental community.

The main visible change in the future is the transition to C++ as the programming language for PYTHIA 7. Even if much of the physics will be carried over unchanged, none of the existing code will survive. The structure of the event record and the whole administrative apparatus is completely different from the current one, in order to allow a much more general and flexible formulation of the event generation process. Following the formulation of a strategy document [53], a first proof-of-concept version was released recently [54]. So far it only contains one reasonably complete physics module, however, namely that of string fragmentation. More realistic versions should follow, but it will take a long time to convert all important physics components from PYTHIA 6. The two versions therefore will coexist for several years, with the FORTRAN one used for physics 'production' and the C++ one for exploration of the object-oriented approach that will be standard at the LHC.

## Acknowledgements

A large number of persons should be thanked for their contributions. Bo Andersson and Gösta Gustafson are the originators of the Lund model, and strongly influenced the early development of the programs. Hans-Uno Bengtsson is the originator of the Pythia program. Mats Bengtsson is the original author of the final-state parton-shower algorithm. Several pieces of code have been donated by other persons. Further comments on the programs have been obtained from users too numerous to be mentioned here, but who are all gratefully acknowledged. To write programs of this size and complexity would be impossible without a strong user feedback.

# References

[1] T. Sjöstrand, Computer Physics Commun. **82** (1994) 74.

[2] H.-U. Bengtsson, Computer Physics Commun. **31** (1984) 323;
H.-U. Bengtsson and G. Ingelman, Computer Physics Commun. **34** (1985) 251;
H.-U. Bengtsson and T. Sjöstrand, Computer Physics Commun. **46** (1987) 43.

[3] T. Sjöstrand, Computer Physics Commun. **27** (1982) 243;
T. Sjöstrand, Computer Physics Commun. **28** (1983) 227;
T. Sjöstrand, Computer Physics Commun. **39** (1986) 347;
T. Sjöstrand and M. Bengtsson, Computer Physics Commun. **43** (1987) 367.

[4] S. Mrenna, Computer Physics Commun. **101** (1997) 232.

[5] B. Cano-Coloma and M.A. Sanchis-Lozano, Nucl. Phys. **B508** (1997) 753;
A. Edin, G. Ingelman and J. Rathsman, Phys. Rev. **D56** (1997) 7317.

[6] K. Lane, Phys.Rev. **D60** (1999) 075007;
S. Mrenna, Phys.Lett. **B461** (1999) 352.

[7] K. Huitu, J. Maalampi, A. Pietilä and M. Raidal, Nucl. Phys. **B487** (1997) 27, and private communication;
G. Barenboim, K. Huitu, J. Maalampi and M. Raidal, Phys. Lett. **B394** (1997) 132.

[8] M. Glück, E. Reya and A. Vogt, Z. Phys. **C67** (1995) 433.

[9] G.A. Schuler and T. Sjöstrand, Z. Phys. **C68** (1995) 607, Phys. Lett. **B376** (1996) 193.

[10] CTEQ Collaboration, H.L. Lai et al., Eur.Phys. J. **C12** (2000) 375.

[11] H. Plothow-Besch, Computer Physics Commun. **75** (1993) 396.

[12] G.A. Schuler and T. Sjöstrand, Nucl. Phys. **B407** (1993) 539, Z. Phys. **C73** (1997) 677.

[13] C. Friberg and T. Sjöstrand, Eur. Phys. J. **C13** (2000) 151, JHEP **09** (2000) 010 (hep-ph/0007314), LU TP 00–31 (hep-ph/0009003), to appear in Phys. Lett. **B**.

[14] H.E. Haber and G.L. Kane, Phys. Rep. **117** (1985) 75.

[15] M. Carena, J.–R. Espinosa, M. Quiros and C.E.M. Wagner, Phys. Lett. **B355** (1995) 209;
M. Carena, M. Quiros and C.E.M. Wagner, Nucl. Phys. **B461** (1996) 407.

[16] M. Bengtsson and T. Sjöstrand, Phys. Lett. **B185** (1987) 435, Nucl. Phys. **B289** (1987) 810.

[17] G. Miu and T. Sjöstrand, Phys. Lett. **B449** (1999) 313.

[18] T. Sjöstrand, Phys. Lett. **157B** (1985) 321.

[19] C. Bálazs, J. Huston and I. Puljak, hep-ph/0002032.

[20] S. Mrenna, UCD-99-4 (hep-ph/9902471).

[21] E. Norrbin and T. Sjöstrand, Phys. Lett. **B442** (1998) 407, LU TP 00–16 (hep-ph/0005110), to appear in Eur. Phys. J. **C**, DOI 10.1007/s100520000460.

[22] A Ballestrero et al., hep-ph/0006259, to appear in the Proceedings of the LEP2 Monte Carlo Workshop.

[23] E. Norrbin and T. Sjöstrand, LU TP 00–42 (hep-ph/0010012).

[24] EMC Collaboration, M. Arneodo et al., Z. Physik **C36** (1987) 527;
L. Apanasevich et al., Phys. Rev. **D59** (1999) 074007.

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

[26] CDF Collaboration, F. Abe et al., Phys. Rev. Lett. **79** (1997) 584.

[27] G. Gustafson and G. Miu, LU TP 99–43 (hep-ph/0002278).

[28] J. Dischler and T. Sjöstrand, LU TP 00–01 (hep-ph/0001237), in preparation.

[29] A. Donnachie and P.V. Landshoff, Phys. Lett. **B296** (1992) 227.

[30] B. Andersson, G. Gustafson, G. Ingelman and T. Sjöstrand, Phys. Rep. **97** (1983) 31.

[31] T. Sjöstrand, Nucl. Phys. **B248** (1984) 469.

[32] WA82 Collaboration, M. Adamovich et al., Phys. Lett. **B305** (1993) 402;
E769 Collaboration, G.A. Alves et al., Phys. Rev. Lett. **72** (1994) 812;
E791 Collaboration, E.M. Aitala et al., Phys. Lett. **B371** (1996) 157.

[33] P. Edén and G. Gustafson, Z. Phys. **C75** (1997) 41;
P. Edén, LUTP 96–29 (hep-ph/9610246).

[34] B. Andersson, G. Gustafson and T. Sjöstrand, Nucl. Phys. **B197** (1982) 45, Physica Scripta **32** (1985) 574.

[35] A. Joly, J. Gascon, P. Taras, Eur. Phys. J. **C6** (1999) 413.

[36] G. Gustafson, U. Pettersson and P. Zerwas, Phys. Lett. **B209** (1988) 90.

[37] T. Sjöstrand and V.A. Khoze, Z. Phys. **C62** (1994) 281, Phys. Rev. Lett. **72** (1994) 28.

[38] G. Gustafson and J. Häkkinen, Z. Phys. **C64** (1994) 659.

[39] E. Norrbin and T. Sjöstrand, Phys. Rev. **D55** (1997) R5;
V.A. Khoze and T. Sjöstrand, Eur. Phys. J. **C6** (1999) 271, EPJdirect **C1** (2000) 1.

[40] L. Lönnblad and T. Sjöstrand, Phys. Lett. **B351** (1995) 293, Eur. Phys. J. **C2** (1998) 165.

[41] R. Mureşan, O. Smirnova and B. Lörstad, Eur. Phys. J. **C6** (1999) 629.

[42] M. Seymour, Phys. Lett. **B354** (1995) 409.

[43] G. Marsaglia, A. Zaman and W.-W- Tsang, Stat. Prob. Lett. **9**. (1990) 35;
F. James, Computer Physics Commun. **60** (1990) 329.

[44] L. Garren, `http://www-pat.fnal.gov/mcgen/lund/convert.pl`.

[45] Particle Data Group, D.E. Groom et al., Eur. Phys. J. **C15** (2000) 1.

[46] I.G. Knowles et al., in 'Physics at LEP2', CERN 96-01, eds. G. Altarelli, T. Sjöstrand and F. Zwirner, Vol. 2, p. 103.

[47] C. Friberg, E. Norrbin and T. Sjöstrand, Phys. Lett. **B403** (1997) 329.

[48] Particle Data Group, R.M. Barnett et al., Phys. Rev. **D54** (1996) 1.

[49] J. André and T. Sjöstrand, Phys. Rev. **D57** (1998) 5767.

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

[51] R. Brun et al., CERN Program Library code Y250.

[52] T. Sjöstrand, LU TP 95–20 (hep-ph/9508391), updated version in preparation.

[53] L. Lönnblad, Computer Physics Commun. **118** (1999) 213.

[54] M. Bertini, L. Lönnblad and T. Sjöstrand, LU TP 00–23 (hep-ph/0006152), to appear in Computer Physics Commun.