CERN-TH.6365/92

PYTHIA at HERA

Torbjörn Sjöstrand

Theory Division, CERN
CH–1211 Geneva 23
Switzerland

Abstract

The PYTHIA program, in conjunction with the closely related JETSET program, provides a general-purpose Monte Carlo event generator for particle production in e^+e^-, ep and pp interactions. In the current paper a summary is given of the physics and programming issues of interest to a HERA user.

---

*To appear in the Proceedings of the Workshop on Physics at HERA, Hamburg, October 1991, eds. W. Buchmüller, G. Ingelman

CERN-TH.6365/92
January 1992

Basic Facts:

Program name:	PYTHIA [1]
Version:	PYTHIA 5.6, from December 1991
Author:	Torbjörn Sjöstrand Theory Division, CERN CH-1211 Geneva 23, Switzerland Phone (Switzerland) - 22 - 767 28 20 BITNET TORSJO @ CERNVM
Program size:	15324 lines
Needs:	JETSET 7.3 [2], from December 1991, by same author, size 10424 lines
Manuals:	manuals are distributed together with programs

1 Introduction

Originally, Pythia was a program intended for hadron-hadron collisions. However, as time went by, the program was generalized, such that any process implemented can be run for any lepton-lepton, lepton-hadron, or hadron-hadron collision, provided only that the appropriate structure functions are available. In the current report, we will not emphasize this generality, but instead concentrate on the information of interest to potential users from the HERA community.

For the production of a HERA event, an event generator should contain a simulation of several physics aspects, as follows.

1. The two beam particles, the e and the p, are characterized by structure functions, which describe the partonic substructure. Note that also an e and a γ inside an e are counted as partons in this sense.
2. One shower initiator parton from each beam starts off a sequence of branchings, such as $e \to e\gamma$ or $q \to qg$, 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. Also the outgoing partons may branch, to build up final state showers.
5. When a shower initiator is taken out of a beam particle, a beam remnant is left behind. The remnant may have an internal structure, and a net colour charge which relates it to the rest of the final state.
6. The QCD confinement mechanism ensures that quarks and gluons are not observable, but instead fragment to colour neutral hadrons. Some of these hadrons are unstable and decay further.

In the following sections, we will survey these aspects, not in the same order as above, but rather in the order they appear in the program execution, i.e. starting with the hard process.

Of the tasks above, actually only points 1, 2, 3 and 5 are handled by Pythia, while tasks 4 and 6 are done by the closely related Jetset program. It is also Jetset that contains a lot of the necesssary utility routines, for everything from event listing to jet-finding algorithms. The joint Pythia/Jetset package is self-contained (but it is possible to use an external structure function library).

The programs described here are Pythia version 5.6 and Jetset version 7.3. Sub-subversions are not denoted by numbers but only by a ‘last date of change’. The present report is not a full-fledged program description, but should only be considered as an appetizer, plus an update on a few developments that have not yet been properly documented elsewhere. The main references remain [1] for Pythia and [2] for Jetset. Although a few years old, these publications still correctly represent the philosophy and much of the physics that goes into the programs. However, the manual parts have been superseded by further developments. Therefore separate manual files are regularly updated and are distributed together with any new release of the program.

It is not difficult to find deficiencies in the current PYTHIA/JETSET program, and several will be discussed as we go along. However, it should be realized that the perfect event generator for HERA physics does not yet exist. It is here tempting to quote the 'Famous last words' of the LEP QCD event generators group [3]:

Due to the large uncertainties present in any realistic QCD Monte Carlo, physics studies must be based on the use of at least two complete and independent programs.

In view of the fact that HERA physics is so much more complex and uncertain, and therefore also more challenging, it is not unreasonable to increase the requirements to a minimum of *three* programs. The hope is that PYTHIA/JETSET may be a useful addition to that list.

2 Hard Processes

The first step in the PYTHIA event generation chain is to select the hard scattering. The total cross-section for a process is given by a convolution of a differential cross-section with the appropriate e and p structure functions. For a $2 \to 2$ process, $ij \to kl$, this reads

$$\begin{aligned} \sigma_{ij\to kl} &= \int\int\int dx_1\, dx_2\, d\hat{t}\, f_i^e(x_1, Q^2)\, f_j^p(x_2, Q^2)\, \frac{d\hat{\sigma}_{ij\to kl}}{d\hat{t}} \\ &= \int\int\int \frac{d\tau}{\tau}\, dy\, d\hat{t}\, x_1 f_i^e(x_1, Q^2)\, x_2 f_j^p(x_2, Q^2)\, \frac{d\hat{\sigma}_{ij\to kl}}{d\hat{t}}. \end{aligned} \quad (1)$$

Here x_1 and x_2 are the fractions of the e and p energies taken by the two incoming partons i and j, $\tau = x_1 x_2$ and $y = 0.5 \ln(x_1/x_2)$. For a $2 \to 2$ process, the differential cross-section $d\hat{\sigma}/d\hat{t}$ is a function of the Mandelstam variables $\hat{s}$, $\hat{t}$ and $\hat{u}$, and of the masses of the outgoing partons k and l — the incoming are always assumed massless. For a $2 \to 1$ process, like leptoquark production, the $\hat{t}$ integral is absent, while a $2 \to 3$ process is characterized by additional kinematics variables.

The standard Deep Inelastic Scattering (DIS) variable set (x, Q^2) has no preferred position in PYTHIA, since it does not easily generalize to other processes. One recovers the DIS language if the electron structure function is collapsed to $f_i^e(x) = \delta_{ei}\delta(x-1)$, x_2 is replaced by the Bjorken x, and $-\hat{t}$ by the DIS Q^2. Note, however, that the structure function Q^2 scale in eq. (1) need not be associated with the DIS Q^2; in fact, in PYTHIA the default structure function scale is $Q^2 = \hat{p}_\perp^2 = \hat{t}\hat{u}/\hat{s}$ (for massless outgoing particles).

A survey of PYTHIA hard scattering processes is found in Table 1. Process 10 is standard deep inelastic scattering, including $\gamma/Z/W$ exchange, but it is possible to restrict to only neutral or charged current, or only γ or Z pieces (see MSTP(21)). In process 10, as in most of the other processes, outgoing quarks are assumed to be massless. Should a massive quark (c, b, t) be produced, the allowed phase space is changed accordingly, but the massless matrix elements are still used. Cross-sections are therefore not finite when $p_\perp \to 0$, and it is necessary to make use of a $p_{\perp min}$ cutoff of order the quark mass. In process 83, on the other hand, the full massive matrix elements are used, and therefore no cuts are necessary. The price to be paid is that only one heavy flavour can be generated at a time, while process 10 covers the simultaneous generation of all allowed flavours.

The photoproduction processes are subdivided into unresolved and resolved ones, depending on whether the photon acts as a point-like particle or the partonic structure inside it is resolved. Again, outgoing quarks are assumed to be massless, except in processes 81, 82 and 84, with the same relative merits as above.

Many other processes can be simulated with PYTHIA, but most of these are of only marginal interest for HERA. In Table 1 are listed some processes for single Z, W or H production that belong to this category. Process 145, for leptoquark production, may be

Table 1: Main HERA physics processes available in PYTHIA.

ISUB	Process	Comment
	Deep Inelastic Scattering	
10	$eq \to eq, \nu q'$	
83	$eq \to \nu Q$	massive matrix elements
	Unresolved Photoproduction	
33	$q\gamma \to qg$	
34	$q\gamma \to q\gamma$	prompt γ
54	$g\gamma \to q\overline{q}$	
84	$g\gamma \to Q\overline{Q}$	massive matrix elements
	Resolved Photoproduction	
11	$qq \to qq$	
12	$q\overline{q} \to q'\overline{q}'$	
13	$q\overline{q} \to gg$	
14	$q\overline{q} \to g\gamma$	prompt γ
18	$q\overline{q} \to \gamma\gamma$	prompt γ
28	$qg \to qg$	
29	$qg \to q\gamma$	prompt γ
53	$gg \to q\overline{q}$	
68	$gg \to gg$	
81	$q\overline{q} \to Q\overline{Q}$	massive matrix elements
82	$gg \to Q\overline{Q}$	massive matrix elements
	Other Processes	
35	$q\gamma \to qZ^0$	no spin info
36	$q\gamma \to q'W^{\pm}$	no spin info
123	$eq \to eqH^0$	Higgs by Z^0Z^0 fusion
124	$eq \to \nu q'H^0$	Higgs by W^+W^- fusion
145	$eu \to LQ$	leptoquark production

of more interest. Among the many possible leptoquark models, the one programmed here is a spin 0 particle with a well-defined flavour content, i.e. the only production and decay channel is $eu \to LQ \to eu$. The flavour content of the leptoquark could easily be changed to ed, say, but currently it is not possible to allow for a variety of different decay modes of one and the same leptoquark.

In addition to hard scattering matrix elements, it is also necessary to make use of structure functions for the hard scattering selection. The structure functions in PYTHIA come in different shapes.

For protons, many sets exist on the market. These are obtained by fits to the data, constrained so that the Q^2 dependence is in accordance with the standard QCD evolution equations. The default in PYTHIA is EHLQ set 1 [4], which today is a bit outdated, but still works fine when high accuracy is not a necessity. Nine other sets are found in PYTHIA. In addition, it is possible to make use of all the sets available in PDFLIB [5] and PAKPDF [6] — the proper interfaces are included in the PYTHIA code, although commented out, so as to avoid problems with unresolved external references for people who do not have these libraries. (Warning: PYTHIA does not police external libraries, but assumes that sensible answers will be returned, also outside the nominal (x, Q^2) range of a set.) Many of the current sets go beyond leading order, in the $\overline{\rm MS}$ or DIS schemes. This offers no advantages here, since the PYTHIA matrix elements are only to leading order.

For DIS and some other processes, it is useful to specify an electron-inside-electron structure function, which parametrizes how much of the original energy is retained by the electron that actually interacts. The formulae in PYTHIA are here based on a next-to-leading order exponentiated description, see [7], p. 34. The approximate behaviour is

$$f_e^e(x, Q^2) \approx \frac{\beta}{2}(1-x)^{\frac{\beta}{2}-1}; \qquad \beta = \frac{2\alpha_{em}}{\pi}\left(\ln\frac{Q^2}{m_e^2} - 1\right). \tag{2}$$

The form is divergent but integrable for $x \to 1$, i.e. the electron likes to keep most of the energy. To handle the numerical precision problems for x very close to unity, the structure function is set, by hand, to zero for $x > 0.999999$, and is rescaled upwards in the range $0.9999 < x < 0.999999$, in such a way that the total area under the structure function is preserved.

In the photoproduction processes, an equivalent flow of photons is assumed, based on first-order formulae. There is some ambiguity in the choice of Q^2 range over which emissions should be included; in the probably most appropriate alternative (MSTP(13)=2) the form is [8]

$$f_\gamma^e(x, Q^2) = \frac{\alpha_{em}}{2\pi}\,\frac{1+(1-x)^2}{x}\,\ln\left(\frac{Q^2_{max}(1-x)}{m_e^2\,x^2}\right). \tag{3}$$

Here Q^2_{max} (PARP(13)) is a user-defined cut for the range of scattered electron kinematics that is counted as photoproduction. Note that we now deal with two different Q^2 scales, one related to the hard subprocess itself, which appears as the argument of the structure function, and the other related to the scattering of the electron, which is reflected in Q^2_{max}. In the default alternative (MSTP(13)=1) only one scale is assumed, i.e. $Q^2_{max}(1-x)/x^2$ is replaced by Q^2 in eq. (3).

Resolved photoproduction also involves the structure functions for quarks and gluons inside the photon inside the electron. Sets of photon structure functions have been obtained as for the proton; an additional complication comes from the necessity to handle the matching of the VDM and the perturbative pieces in a consistent manner. PYTHIA contains the Drees–Grassie structure functions [9]; in addition there is an interface to the PHOPDF library [10], again commented out by default. In no published set have the results been directly presented in terms of quark and gluon structure functions inside the electron. These are therefore obtained by numerical convolution according to

$$f_{q,g}^e(x, Q^2) = \int_x^1 \frac{dz}{z}\, f_\gamma^e(z, Q^2)\, f_{q,g}^\gamma\left(\frac{x}{z}, Q^2\right), \tag{4}$$

with f_γ^e as discussed above. The necessity for numerical convolution makes this structure function evaluation rather slow compared with the other ones, therefore one should only have it switched on (MSTP(12)=1) for resolved photoproduction studies.

3 Initial and Final State Radiation

To go, beyond simple $2 \to 2$ processes, to $2 \to n$ ones ($n > 2$), one may either adopt a matrix element or a parton shower approach. The two are complementary. The former accurately predicts the rate of well-separated jets, but it is computationally difficult and not suited for a study of the internal jet structure. In PYTHIA, the parton shower approach is adopted, in which the $2 \to n$ process is subdivided into initial state showers, a hard scattering (usually $2 \to 2$), and final state showers. The showers are of a universal nature, and can therefore easily be attached to any hard process. Since the subdivision is neither exact nor unique, the price to be paid is a large uncertainty in the rate of well separated jets, while internal jet structures should be well described.

In both initial and final state showers, the structure is given in terms of branchings $a \to bc$, specifically $e \to e\gamma$, $q \to qg$, $q \to q\gamma$, $g \to gg$, and $g \to q\overline{q}$. Each of these processes is characterized by a splitting kernel $P_{a\to bc}(z)$. The branching rate is proportional to the integral $\int P_{a\to bc}(z)\,dz$. The z value picked for a branching describes the energy sharing, with daughter b taking a fraction z and daughter c the remaining $1-z$ of the a energy. Once formed, the daughters b and c may branch in their turn, and so on.

Each parton is characterized by some virtuality scale Q^2, further discussed below, which gives an approximate sense of time ordering to the cascade. In the initial state shower, Q^2 values are gradually increasing as the hard scattering is approached, while Q^2 is decreasing in the final state showers. Shower evolution is cut off at some lower scale Q_0, typically around 1 GeV for QCD branchings. The same cutoff scale is also used to regularize the soft gluon emission divergences in the splitting kernels. From above, a maximum scale Q_{max} is introduced, where the showers are matched to the hard interaction itself. The relation between Q_{max} and Q_{hard} is uncertain, and the choice made can strongly affect the amount of well separated jets. For QCD processes, double-counting arguments favour $Q_{max} \approx Q_{hard}$ (more precisely, owing to a mismatch in the way variables are conventionally defined, $Q^2_{max} \approx 4Q^2_{hard}$; see PARP(67), PARP(71)), but unfortunately those arguments are not compelling for DIS.

Despite a number of common traits, the initial and final state radiation machineries are in fact quite different, and are described separately below.

The final state radiation algorithm of JETSET is described in detail in [11]. Final state showers are time-like, i.e. partons have $m^2 = E^2 - \overline{p}^2 \geq 0$. The evolution variable Q^2 of the cascade is therefore in JETSET associated with the m^2 of the branching parton (but this choice is not unique). Starting from Q^2_{max}, an original parton is evolved downwards in Q^2 until a branching occurs. The selected Q^2 value defines the mass of the branching parton, and the z of the splitting kernel how the parton energy is split between its daughters. These daughters may now, in turn, evolve downwards, in this case with maximum virtuality already defined by kinematics, and so on down to the Q_0 cutoff.

Corrections to the leading log picture, so-called coherence effects [12], lead to an ordering of subsequent emissions in terms of decreasing angles. This does not follow automatically from the mass ordering constraint, but is implemented as an additional requirement on allowed emissions. It is also possible to obtain non-trivial correlations between azimuthal angles in the various branchings, see [13], some of which are implemented as options (see MSTJ(46)). Finally, the theoretical analysis strongly suggests the scale choice $\alpha_S = \alpha_S(p_\perp^2) = \alpha_S(z(1-z)m^2)$, and this is the default in the program.

The final state radiation machinery is applied in the c.m. frame of the hard scattering. The total energy and momentum of the hard scattering subsystem is preserved, as is the direction of the outgoing partons (in that frame). However, in a process like $eq \to eq$, the outgoing electron energy will be reduced if the outgoing quark acquires a non-vanishing mass.

Both an outgoing electron and an outgoing quark can radiate photons (see MSTJ(41)), in basically the same manner as a quark radiates gluons. The major difference is that photons do not obey the angular ordering coherence condition.

In contrast to final state showers, initial state ones are space-like. This means that, in the sequence of branchings $a \to bc$ that lead up from the shower initiator to the hard interaction, particles a and b have $m^2 = E^2 - \overline{p}^2 < 0$ (with the exception of the original beam electron, which is on the mass-shell, of course). The 'side branch' particle c, which does not participate in the hard scattering, may be on the mass-shell, or have a time-like virtuality. In the latter case a time-like shower will evolve off it, rather like the final state radiation described above. To first approximation, the evolution of the space-like main branch is characterized by the evolution variable $Q^2 = -m^2$, which is required to be strictly increasing along the shower, i.e. $Q_b^2 > Q_a^2$. Corrections to this picture have been calculated [14], but are absent in PYTHIA (with the exception of a very simple option for angular ordering, which is only part of the full story).

Initial state radiation is handled within the backwards evolution scheme, as described in [15]. In this approach, the choice of the hard scattering is based on the use of evolved structure functions, cf. eq. (1), which means that the inclusive effects of initial state radiation are already included. What remains is therefore to construct the exclusive showers. This is done starting from the two incoming partons at the hard interaction, tracing the showers 'backwards in time', back to the two shower initiators. In other words, given a parton b, one tries to find the parton a that branched into b. The evolution in the Monte Carlo is therefore in terms of a sequence of decreasing space-like virtualities Q^2 and increasing momentum fractions x. Branchings on the two sides are interleaved in a common sequence of decreasing Q^2 values.

The definition of the x variable for off-mass-shell partons is not unique; in Pythia one requires $(p_i + p_j)^2 \equiv \hat{s}_{ij} = x_i x_j s$, both for $i = a$ and $i = b$, if j represents the parton from the other side of the event that is resolved at the appropriate Q^2 scale. In other words, the $z = x_b/x_a$ variable of a branching tells how much the scattering subsystem invariant mass-squared is reduced by the branching. If originally parton b was assumed to have vanishing $p_\perp$, the reconstruction of the branching $a \to bc$ introduces a $p_\perp$ for b, which is compensated by c. This implies a rotation and boost of all partons that follow after b, including the hard process (and, as a technical aside, also the partons that j branch into).

In the formalism above, there is no real distinction between gluon and photon emission. Initial state radiation off the e leg is therefore automatically included as part of the DIS and unresolved photoproduction machineries. For resolved photoproduction, the 'branching' where a parton is picked out of the photon is of a different nature than normal branchings. The related technical problems are not yet solved in Pythia, wherefore QED initial state radiation on the electron side is switched off in this case.

As we see, the showering algorithms do not preserve the DIS x and Q^2 values, defined in terms of the scattered electron direction and energy. In [16] detailed modifications are presented that make a preservation possible when radiation off incoming and outgoing electron is neglected, but these are not included in the current version of Pythia. However, recently a simple machinery has been included as an option (see MSTP(23)), which preserves x and Q^2 for the effects of QCD radiation, and also for those of primordial $k_\perp$ and the beam remnant treatment, as follows. After the showers have been generated, the four-momentum of the scattered lepton is changed to the nominally expected one, based on the x and Q^2 values. The momentum transfer vector q is adjusted accordingly, such that the scattered quark (together with all its daughters) has to compensate the extra momentum given to the lepton. The original mass of the scattered quark is not preserved in this operation; therefore a longitudinal boost of the whole quark initial shower is imposed in such a way that the scattered quark is brought back to the correct mass. Sometimes this may not be kinematically possible; therefore some events are rejected. Further, photon radiation off the lepton leg is not fully accounted for, i.e. it is assumed that the energy of final state photons is added to that of the scattered electron for the definition of x and Q^2. The scheme presented above should not be taken too literally, but is rather intended as a contrast to the more sophisticated schemes already on the market, if one would like to understand whether the kind of conservation scheme chosen does affect the observable physics.

4 Beam Remnants, Fragmentation, and Decays

The initial state radiation algorithm reconstructs one shower initiator in each beam. Together the two initiators delineate an interaction subsystem, which contains all the partons that participate in the initial state showers, the hard interaction, and the final state showers. Left behind are two beam remnants, which just sail through, unaffected by the hard process.

In some cases a remnant is a single object, as when a γ is taken out of an e beam, leaving behind an e, or a valence quark is taken out of a proton, leaving behind a diquark state. When taking an e out of an e, a soft γ is left behind, which is then more related to the cutoff of $f_e^e(x, Q^2)$ at $x = 0.999999$ than to the ordinary beam remnant concept, but is handled with the same machinery. In other cases, a remnant consists of two objects, as when a q is taken out of an e, leaving behind $e + \overline{q}$, or a sea quark out of a proton, leaving behind a $qqq\overline{q}$ state which is split into a meson plus a diquark. Currently no provisions are made for more than two remnant objects; therefore a g taken out of an e always leaves behind $e + g$ and never $e + q + \overline{q}$.

When two objects are defined, the remnant energy (more specifically, the lightcone combination energy plus longitudinal momentum) is split between the two objects, in e beams normally based on structure functions, in p beams by some parametrized shape.

Further, a primordial $k_\perp$ is introduced between the shower initiator and the remnant. For quarks and gluons inside a p or e this distribution is by default a Gaussian, but for the e side it is also possible to pick broader shapes, which better represent the perturbative expectations in $\gamma \to q + \overline{q}$ [17]. An e or γ inside an e does not have a primordial $k_\perp$. If a remnant consists of two partons, or a parton and hadron, an additional relative $p_\perp$ will be introduced between these two.

For a number of reasons (primordial $k_\perp$, remnant masses, etc.), energy and longitudinal momentum are not fully preserved in the beam remnant treatment. Therefore the event is subdivided into three pieces — the interaction subsystem and the two remnants — and the two 'largest' of these are boosted longitudinally by some small amounts to achieve overall energy and momentum conservation.

This completes the description of the partonic state of the event. To convert to hadrons, the string fragmentation [18] routines of JETSET are used. String fragmentation hardly needs a detailed description, since it is likely to be the most sophisticated and successful fragmentation model on the market, and is already described elsewhere [18, 3, 2]. This does not mean that there are no problems: the flavour composition is not calculable from first principles, but requires a plethora of parameters; the charm and bottom fragmentation functions appear somewhat softer than predicted by the Lund symmetric fragmentation function; Bose–Einstein correlations are not yet understood; and so on. One should also keep in mind that HERA will test slightly different aspects of the model than e^+e^- machines have done, in particular in the beam remnant region.

The unique aspect of string fragmentation, compared to other fragmentation approaches, is that the complete colour topology of outgoing quarks, diquarks and gluons is reflected in the way partons are connected by strings. It is the strings that fragment to produce particles, rather than the individual partons. In general, each string is stretched from a colour triplet end (quark, antidiquark), via a number of gluons, to a colour antitriplet end (antiquark, diquark). The gluons therefore act as kinks on the string, so that the net effect of a gluon goes to zero continuously in the soft and collinear limits. The initial and final state showers add activity to an event, but even when the showers are absent the string fragmentation ensures a non-negligible amount of particle production. Therefore no alternative physics mechanism need be invoked to describe the fragmentation in the proton target region.

While colour flow, and therefore string drawing, is uniquely defined in a leading log parton shower picture, this need not be the case in a hard $2 \to 2$ partonic scattering, as encountered e.g. in resolved photoproduction. In fact, several different colour topologies are here allowed, with different kinematical structure. The separation is not perfect: interference terms exist for which the colour structure is not well defined. Therefore different recipes are possible, which differ by small amounts (see MSTP(34)).

Many of the particles produced in the string fragmentation are unstable, and decay further. Decay data are taken from the Review of Particle Properties [19], where known. However, especially for charm and bottom, this knowledge is quite patchy and rapidly changing. It is therefore necessary for a Monte Carlo author to construct complete and consistent decay tables himself. The last complete update of the decay tables in JETSET

Table 2: The PYTHIA/JETSET routines that are intended to be called by users. Those marked by '⋆' are especially important/useful.

	Routine	Usage
⋆	PYINIT	initialize the event generation chain
⋆	PYEVNT	generate an event
⋆	PYSTAT	list cross-sections, resonance partial widths
	PYFRAM	translate between HERA and c.m. frame
	PYKCUT	user-defined cuts on hard process kinematics
	PYEVWT	user-defined event weighting procedure
⋆	LULIST	list event record, particle data table
⋆	LUEDIT	remove undesired entries from event record
	LUHEPC	conversion between LUJETS and HEPEVT event records
	LUROBO	generic rotations and boosts
	LUEXEC	do fragmentations+decays, if not already done
	LUGIVE	controlled setting of commonblock variables
	LUUPDA	user-controlled update of particle decay table
	KLU	function for integer event record data, e.g. charge
	PLU	function for real event record data, e.g. rapidity
	LUSPHE	sphericity event analysis
	LUTHRU	thrust event analysis
	LUCLUS	jet/cluster finding in $p_\perp$ or mass
	LUCELL	jet/cluster finding in calorimeter grid
	LUJMAS	high and low jet mass
	LUFOWO	Fox–Wolfram moments
	LUTABU	tables of particle composition in events, etc.
	RLUGET	save state of random number generator
	RLUSET	restore earlier state of random number generator

is some years in the past, and therefore much of the recent information remains yet to be included. However, by and large, the decay treatment is likely to be as good as or better than any other general-purpose fragmentation/decay package on the market.

5 How To Run PYTHIA

PYTHIA and JETSET are subroutine libraries, i.e. it is up to the user to write the main program from which the various routines are called, and where PYTHIA/JETSET commonblock variables are set or read out. Full descriptions are found in the PYTHIA and JETSET manuals distributed with the programs. Here we just give a brief survey. A list of frequently used subroutines and functions is found in Table 2, examples of useful commonblock switches and parameters in Tables 3 and 4, and examples of event information in Table 5. Some of the structure described here may be hidden by collaboration-enforced interfaces; we will only consider the 'pure' PYTHIA/JETSET programs.

The usage of PYTHIA can be subdivided into three steps.

1. The initialization step. It is here that all the basic characteristics of the coming generation are specified. The material in this section includes the following.
   - Commonblocks, at least the following, and maybe some more:

```
COMMON/LUJETS/N,K(4000,5),P(4000,5),V(4000,5)
COMMON/LUDAT1/MSTU(200),PARU(200),MSTJ(200),PARJ(200)
COMMON/PYSUBS/MSEL,MSUB(200),KFIN(2,-40:40),CKIN(200)
```

Table 3: Some main switches and parameters in Pythia/Jetset commonblocks, with default values and examples of variations (for better or worse). Full details are found in the manuals; the list below is mainly intended for quick reference.

Variable	Default	Alternative	Comment
Hard Interaction			
MSTP(21)	1	4, 5	$\gamma/Z/W$ switch for DIS
MSTP(32)	2	5	Q^2 definition in hard process
MSTP(22)	0	2	special Q^2 choice used for DIS showers
MSTP(34)	0	1	QCD matrix element interference
MSTP(35)	0	1	threshold factor for heavy flavours
MSTP(23)	0	1	x and Q^2 conservation in DIS processes
Kinematics Cuts on Hard Interaction			
CKIN(1)	2. GeV		minimum mass of hard subsystem
CKIN(2)	$-1.$ GeV		maximum mass (inactive when negative)
CKIN(3)	0. GeV		minimum $p_\perp$
CKIN(4)	$-1.$ GeV		maximum $p_\perp$ (inactive when negative)
CKIN(35)	0. GeV^2	25. GeV^2	minimum $-\hat{t}$, i.e. Q^2 for DIS
Structure Functions			
MSTP(11)	0	1	enable resolved electron
MSTP(12)	0	1	enable q/g structure functions in e
MSTP(13)	1	2	Q^2 range of γ radiation off e
PARP(13)	25 GeV^2		Q^2_{max} for MSTP(13)=2
MSTP(51)	1	6	p structure function parametrization
MSTP(52)	1	2	p structure function library (PDFLIB)
MSTP(55)	1		γ structure function parametrization
MSTP(56)	1	3	γ structure function library (PHOPDF)
Initial State Radiation (ISR)			
MSTP(61)	1	0	master switch ISR on/off
MSTP(62)	2	1, 3	coherence level in ISR
PARP(62)	1.0 GeV	0.5 GeV	Q_0 cutoff in ISR
PARP(67)	4.	1. – 9.	Q^2_{max}/Q^2_{hard} for ISR
Final State Radiation (FSR)			
MSTP(71)	1	0	master switch FSR on/off
MSTJ(41)	1	2	also allow $q \to q\gamma$ and $e \to e\gamma$ in FSR
MSTJ(42)	1	0	coherence level in FSR
MSTJ(46)	0	3	non-isotropic azimuthal angle in FSR
PARJ(82)	1.0 GeV	0.5 GeV	QCD shower Q_0 cutoff in FSR
PARJ(83)	1.0 GeV	0.1 GeV	γ emission Q_0 cutoff in FSR
PARP(71)	4.	1. – 9.	Q^2_{max}/Q^2_{hard} for FSR
Beam Remnants			
MSTP(92)	4	3	energy sharing in beam remnant
PARP(91)	0.44 GeV	0.3 – 0.6 GeV	width primordial $k_\perp$ in p
MSTP(93)	1	5	shape primordial $k_\perp$ in γ
PARP(99)	0.44 GeV	0.1 – 0.6 GeV	width primordial $k_\perp$ in γ
PARP(100)	2.0 GeV	1.0 – 10.0 GeV	upper cut primordial $k_\perp$ in γ
For continuation, see Table 4			

Table 4: Some further switches and parameters, see Table 3 for comments.

Variable	Default	Alternative	Comment
Fragmentation (F)			
MSTP(111)	1	0	master switch F+decay on/off
MSTJ(1)	1	2	string vs. independent F
MSTJ(11)	1	3	Lund/Peterson F function
PARJ(41)	0.5	0.18	a parameter in Lund F function
PARJ(42)	0.9 GeV^{-2}	0.34 GeV^{-2}	b parameter in Lund F function
PARJ(54)	0.0	-0.02	$-\epsilon_c$ parameter in Peterson F function
PARJ(55)	0.0	-0.007	$-\epsilon_b$ parameter in Peterson F function
PARJ(1)	0.10	0.08 – 0.12	qq/q ratio, i.e. baryon rate
PARJ(2)	0.30	0.25 – 0.35	s/u ratio, i.e. strangeness production
PARJ(11)	0.50	0.3 – 0.7	light vector meson probability
PARJ(12)	0.65	0.5 – 0.75	strange vector meson probability
PARJ(21)	0.35 GeV	0.37 GeV	F transverse momentum width
MSTJ(51)	0	1	toy model for Bose–Einstein correlations
Decays			
MSTJ(22)	1	4	'detector' region where decays allowed
MSTJ(26)	0	1	B–$\overline{B}$ mixing
Miscellaneous			
MRLU(1)	19780503		random number sequence
MSTP(141)	0	1	enable early cuts on hard scattering
MSTP(142)	0	1	enable weighted events
PMAS(6,1)	120. GeV		top quark mass

Table 5: Some important event information stored in the LUJETS and PYPARS commonblocks.

Variable	Meaning
N	number of entries; entries stored in 1≤I≤N
K(I,1)	status code (present/fragmented/information)
K(I,2)	particle flavour code, PDG convention [19]
K(I,3)	line number of mother
P(I,1)	p_x momentum component
P(I,2)	p_y momentum component
P(I,3)	p_z momentum component
P(I,4)	E, i.e. particle energy
P(I,5)	m, i.e. particle mass
MSTI(1)	event process type generated, see Table 1
PARI(1)	total integrated cross-section
PARI(14)	$\hat{s}$ of hard process
PARI(15)	$\hat{t}$ of hard process
PARI(17)	$p_\perp$ of hard process
PARI(33), PARI(34)	momentum fractions x at hard process
PARI(39), PARI(40)	primordial $k_\perp$ for beam remnants
PARI(75), PARI(76)	x value of γ in resolved photoproduction

```
COMMON/PYPARS/MSTP(200),PARP(200),MSTI(200),PARI(200)
```

- Selection of required processes. Some fixed 'menus' of subprocesses can be selected with different MSEL values, but with MSEL=0 it is possible to compose 'à la carte', using the subprocess numbers in Table 1. To generate processes 14, 18 and 29, for instance, one needs
```
MSEL=0
MSUB(14)=1
MSUB(18)=1
MSUB(29)=1
```
- Selection of kinematics cuts, see Table 3. To generate hard scatterings with 5 GeV $\leq p_{\perp} \leq$ 10 GeV, for instance, use
```
CKIN(3)=5.
CKIN(4)=10.
```
Unfortunately, initial and final state radiation will shift around the kinematics of the hard scattering, making the effects of cuts less predictable. One therefore always has to be very careful that no desired event types are cut out.
- Definition of underlying physics scenario, e.g. top mass. See Tables 3 and 4.
- Selection of structure function sets, Q^2 definitions, and all other details of the generation. See Tables 3 and 4.
- Switching off of generator parts not needed for toy simulations, e.g. fragmentation for parton level studies. See Tables 3 and 4.
- Initialization of the event generation procedure. Here kinematics is set up, maxima of differential cross-sections are found for future Monte Carlo generation, and a number of other preparatory tasks carried out. Initialization is performed by PYINIT, which should be called only after the switches and parameters above have been set to their desired values. Unfortunately, there is no nice default option for the HERA event frame, so one has to specify the two beam momenta explicitly, as follows:
```
P(1,1)=0.
P(1,2)=0.
P(1,3)=30.
P(2,1)=0.
P(2,2)=0.
P(2,3)=-820.
CALL PYINIT('user','e-','p',0.)
```
The order of electron and proton, or the signs or momenta, may be flipped at will.
- Any other initial material required by the user, e.g. histogram booking.

2. The generation loop. It is here that events are generated and studied. It includes the following tasks:
   - Generation of the next event, with
```
CALL PYEVNT
```
   - Printing of a few events, to check that everything is working as planned, with
```
CALL LULIST(1)
```
   - An analysis of the event for properties of interest, see Tables 5 and 2. If one desires to use the HEPEVT standard [3] for this analysis, this can be achieved by a call to the LUHEPC routine.
   - Saving of events on tape, or interfacing to detector simulation.
3. The finishing step. Here the tasks are:
   - Printing a table of deduced cross-sections, obtained as a by-product of the Monte Carlo generation activity, with the command
```
CALL PYSTAT(1)
```

- Printing histograms and other user output.

6 Conclusions

Pythia is quite a new program in the *ep* arena, and therefore much remains to be done. However, it can boast at least two big advantages: a large repertory of hard processes accessible in a consistent manner, and the best fragmentation+decay package on the market.

A major drawback is the uncertainty, inherent in all shower approaches, on the amount of hard QCD emission that should be expected. Another is that the initial state showering algorithm is not really well suited for the study of small x physics — probably the best that can be said is that if Pythia can be tuned to provide a reasonable description of event shapes in low x events, it will be difficult to claim the presence of new, interesting physics. This does not mean that the author is hostile to this kind of physics; in fact, the first indirect evidence for 'hot spots' inside the proton comes from studies with Pythia on the structure of minimum bias $p\overline{p}$ collider events [20].

It is not likely that there will be a full-time development and maintenance effort for the *ep* parts of Pythia. However, these parts are anyway rather small, and the bulk of the code is common for e^+e^-, *ep*, and *pp*. Pythia is therefore going to continue to evolve in the future, and hopefully remain an up-to-date program.

References

[1] H.-U. Bengtsson and T. Sjöstrand, Computer Physics Commun. **46** (1987) 43

[2] T. Sjöstrand, Computer Physics Commun. **39** (1986) 347;
T. Sjöstrand and M. Bengtsson, Computer Physics Commun. **43** (1987) 367

[3] T. Sjöstrand *et al.*, *in* Z Physics at LEP 1, eds. G. Altarelli, R. Kleiss and C. Verzegnassi, CERN 89-08 (Geneva, 1989), Vol. 3, p. 143

[4] E. Eichten, I. Hinchliffe, K. Lane and C. Quigg, Rev. Mod. Phys. **56** (1984) 579 and **58** (1985) 1065

[5] H. Plothow-Besch, *in* MC91 Workshop on Detector and Event Simulation in High Energy Physics, eds. K. Bos and B. van Eijk (NIKHEF-H, Amsterdam, 1991), p. 148

[6] K. Charchuła, DESY 91-093 (1991)

[7] R. Kleiss *et al.*, *in* Z Physics at LEP 1, eds. G. Altarelli, R. Kleiss and C. Verzegnassi, CERN 89-08 (Geneva, 1989), Vol. 3, p. 143

[8] M. Bilenky and G. d'Agostini, private communication

[9] M. Drees and K. Grassie, Z. Phys. **C28** (1985) 451

[10] K. Charchuła, ZEUS note 91-73 (unpublished)

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

[12] A.H. Mueller, Phys. Lett. **104B** (1981) 161;
B.I. Ermolaev and V.S. Fadin, JETP Lett. **33** (1981) 269

[13] B.R. Webber, Annu. Rev. Nucl. Part. Sci. **36** (1986) 253

[14] M. Ciafaloni, Nucl. Phys. **B296** (1987) 249;
S. Catani, F. Fiorani and G. Marchesini, Nucl. Phys. **B336** (1990) 18;
G. Marchesini and B.R. Webber, Nucl. Phys. **B349** (1991) 617

[15] T. Sjöstrand, Phys. Lett. **157B** (1985) 321;
M. Bengtsson, T. Sjöstrand and M. van Zijl, Z. Phys. **C32** (1986) 67

[16] M. Bengtsson and T. Sjöstrand, Z. Phys. **C37** (1988) 465

[17] M. Drees, private communication

[18] B. Andersson, G. Gustafson, G. Ingelman and T. Sjöstrand, Phys. Rep. **97** (1983) 31;
T. Sjöstrand, Nucl. Phys. **B248** (1984) 469

[19] Particle Data Group, J.J. Hernández *et al.*, Phys. Lett. **B239** (1990) 1

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