Quaero: A General Interface to TeV-Scale Event Data

Bruce Knuteson*

University of Chicago

(Dated: July 8, 2002)

We describe QUAERO, a general interface to TeV-scale event data. Data are reduced to the momenta of final state objects ($e^{\pm}$, $\mu^{\pm}$, $\tau^{\pm}$, γ, b, j, and $\not{p}$), and events are partitioned according to final state objects contained. Models are tested by running predicted events through a detector simulator and comparing the data to the new hypothesis and the standard model.

Contents

I. INTRODUCTION

It is generally recognized that the standard model, a successful description of the fundamental particles and their interactions, must be incomplete. Models that extend the standard model often predict rich phenomenology at the scale of a few hundred GeV, an energy regime accessible to the LEP collider. Due in part to the complexity of the apparatus required to test models at such large energies, experimental responses to these ideas have not kept pace. Any technique that reduces the time required to test a particular candidate theory would allow more such theories to be tested, reducing the possibility that the data contain overlooked evidence for new physics.

Once data are collected and the backgrounds have been understood, the testing of any specific model in principle follows a well-defined procedure. In practice, this process has been far from automatic. Even when the basic selection criteria and background estimates are taken from a previous analysis, the reinterpretation of the data in the context of a new model often requires a substantial length of time.

Ideally, the data should be "published" in such a way that others in the community can easily use those data to test a variety of models. The publishing of experimental distributions in journals allows this to occur at some level, but an effective publishing of a multidimensional data set by a large particle physics experiment has proven difficult. The problem appears to be that such data are context-specific, requiring detailed knowledge of the complexities of the apparatus. This knowledge must somehow be incorporated either into the data or into whatever tool the non-expert would use to analyze those data.

This article describes QUAERO, a tool that enables the analysis of high energy collider data by non-experts. The original version of QUAERO [1], developed by the DØ experiment at Fermilab, computes cross section $\times$ branching ratio limits on new phenomena. Here we extend QUAERO to allow parameter estimation, encompassing both searches for new phenomena and semi-precision measurements.

II. OVERVIEW

The QUAERO interface has been made as simple as possible. A physicist provides a model in the form of PYTHIA commands or a HEPEVT file with predicted events. These events are added to a user-defined subset of the standard model prediction to define the physicist's hypothesis. The physicist also provides his name, institution, and the email address to which results should be sent. A

---

*Electronic address: knuteson@fnal.gov; URL: http://hep.uchicago.edu/~knuteson/

click of a button submits the request.

III. PREPARATION

Incorporating data collected by a high energy physics experiment into QUAERO is straightforward. We begin with a short discussion of the QUAERO standard, the format used to store and process signal, background, and data files.

A. Standard QUAERO file

A standard QUAERO file is a text file containing the 4-vectors of final state objects. Each row of a QUAERO file represents one event, and consists of several strings and numbers separated by spaces or tabs.

For each event, each object is listed by providing the type of object — **e-** or **e+** for an electron or positron, **mu-** or **mu+** for a muon, **tau-** or **tau+** for a tau, **ph** for a photon, **j** for a non-b-tagged jet, **b** for a b-tagged jet, and **uncl** for unclustered energy — followed by the momentum 4-vector of the object — m, E, $\cos(\theta)$, and ϕ, where m and E are the mass and energy of the object in units of GeV, and θ and ϕ are the polar and azimuthal angles. All objects are assumed to be isolated. Missing energy ($\vec{\not{p}}$) is determined by energy conservation. The mass m is only given for jets, b-tagged jets, and unclustered energy, since the mass of the object is known in all other cases. We take $m_\tau = 1.78$ GeV, and $m_e = m_\mu = m_\gamma = 0$.

Associated with each event is a weight, chosen such that if all events are considered together, each with its appropriate weight, the correct distribution is obtained, with normalization equal to the total number of predicted events (in the case of signal or background) or observed events (in the case of data). Each event may have a different weight, allowing the use of Monte Carlos that produce weighted events. Each row in a QUAERO file begins with a string describing the type of event, followed by the weight of the event, followed by the center-of-mass energy of the collision, followed by the objects in the event. The end of an event is signaled by the presence of a semicolon, separated by whitespace.

The influence of systematic uncertainties are denoted in curly brackets immediately following the affected quantity. The general notation is {`err/magnitude,...`}, where `err` is an integer specifying the source of systematic error, `magnitude` is the root mean square variation in the affected quantity, and separate sources of error are separated by commas. In this way the influence of systematic errors on each individual event can be fully specified. A more detailed description is provided in Sec. III B 2.

An event in a QUAERO file containing two electrons and two b-tagged jets might therefore look like this:

`eventType`
`weight`{err/mag,...}   `sqrt(s)`{err/mag,...}
e+ E{err/mag,...} $\cos(\theta)$ ϕ
e- E{err/mag,...} $\cos(\theta)$ ϕ
b m{err/mag,...} E{err/mag,...} $\cos(\theta)$ ϕ
b m{err/mag,...} E{err/mag,...} $\cos(\theta)$ ϕ
uncl m{err/mag,...} E{err/mag,...} $\cos(\theta)$ ϕ **;**

B. Requirements

The requirements of an experiment can be broken into five parts:

1. Data
2. Backgrounds
3. Detector simulator
4. Systematic errors
5. Refinements

We consider each in turn.

1. Data

In a QUAERO data file, `eventType` is replaced by the keyword `data`, and the weight is set equal to 1. No systematic errors (curly brackets) are included in the data file. The statement that we measure an energy to be 45.2 GeV is exact; systematic errors affect our interpretation of that measurement through the modeling of our detector, and hence are included in background and signal events.

With actual numbers, an event in a data file might look like this:

`data`   1   190.0
e+   45.2   +0.11   0.21
e-   47.3   -0.05   3.56
b   4.2   46.0   -0.16   1.71
b   4.3   48.2   -0.02   4.90
uncl   0.44   3.3   +0.07   3.97 **;**

All data events are stored in one massive file.

2. Systematic Errors

The effect of systematic errors is denoted within curly brackets immediately following the affected quantity. We divide all sources of systematic uncertainty into three different types:

- Mismeasurement errors. The notation for mismeasurement errors is {`err/magnitude`}, where `err` is an integer specifying the source of systematic error and `magnitude` is the root mean square variation in the affected quantity. The uncertainty is assumed to be normally distributed. The hadronic energy scale is an example of such an error.

- Identification errors. The notation for identification errors is {`err/newID/chance`}, where `err` is an integer specifying the source of systematic error, and the identification of the object is changed to `newID` with probability `chance`. A systematic error causing more jets than estimated to be misidentified as b jets is an example of such an error.

- Choice errors. The notation for choice errors is {`err/case`}, where `err` is an integer specifying the source of systematic error, and `case` is an integer. The affected quantity is zero unless a randomly chosen integer is equal to `case`. The choice of which Monte Carlo generator to use to estimate a particular background is an example of such an error.

Uncertainties due to Monte Carlo statistics are incorporated by QUAERO automatically.

Each systematic error source is identified by a unique integer. All systematic error sources are registered in a text file called `systematicSources.txt`, which has the following format:

```
err   errorType   description
⋮
```

Here `err` is an identifying integer, `errorType` is either `m` (mismeasurement), `i` (identification), or `c_` (choice), where _ is the number of possible cases.

Correlations between these sources of error are stored in a separate text file called `systematicCorrelations.txt`, which has the following format:

```
err1   err2   correlation
⋮
```

Here `err1` and `err2` are two integers identifying two sources of error, and `correlation` is the correlation between them, satisfying $-1 \leq$ `correlation` ≤ 1. Choice errors are assumed to be uncorrelated with other sources of error.

The question of which sources of systematic error to include is left up to each experiment. An example is provided in Appendix A.

3. Backgrounds

In a QUAERO background file, `eventType` is replaced by the type of background event. Systematic errors are included.

With actual numbers, an event in a background file might look like this:

ZZ   0.0041{1/0,12/0.0002,0201/0.0001}   190.0
e+   45.2{0221/1.4}   +0.11   0.21
e-   47.3{0221/1.5}   -0.05   3.56
b{0211/j/0.001}   4.2{0222/0.61}   46.0{0222/3.6}   -0.16   1.71
b{0211/j/0.05}   4.3{0222/0.35}   48.2{0222/3.7}   -0.02   4.90
uncl   0.44{0222/0.07}   3.3{0222/0.32}   +0.07   3.97 **;**

The systematic errors appearing inside the curly brackets are explained in the example of Appendix A.

All background events are stored in one massive file.

4. Detector simulator

Signal events are provided by the physicist either in the form of PYTHIA commands, which are used to generate a HEPEVT file, or in the form of a HEPEVT file directly. The format of the HEPEVT text file is provided in Appendix B.

Signal events are converted from HEPEVT into standard QUAERO format by an executable provided by each experiment with type

```
simulate <input hepevt file> <output text
file> <generated luminosity> [<working
directory>]
```

that takes one HEPEVT file as input; simulates, reconstructs, and appropriately weights each event; and produces as output a single file in standard QUAERO format. `simulate` should assume that the input file corresponds to the stated luminosity at each of $\sqrt{s}$ = 183, 189, 192, 196, 200, 202, 205, 207, and 208 GeV and should re-weight each event accordingly.

In the special case of LEP2, it is possible for an experiment instead to provide an executable with type

```
simulate0 <cm energy> <input hepevt file>
<output text file> <generated luminosity>
[<working directory>]
```

where `<cm energy>` is 183, 189, 192, 196, 200, 202, 205, 207, or 208, and all of the events in the input file have been generated at this single center of mass energy.

In a QUAERO signal file, `eventType` is replaced by the keyword `sig`. Systematic errors are included on signal events just as they are on background events. An event in a QUAERO signal file might therefore look like this:

`sig`   0.0041{1/0,12/0.0002,0201/0.0001}   190.0
e+   45.2{0221/1.4}   +0.11   0.21
e-   47.3{0221/1.5}   -0.05   3.56
b{0211/j/0.001}   4.2{0222/0.61}   46.0{0222/3.6}   -0.16   1.71

b{0211/j/0.05} 4.3{0222/0.35} 48.2{0222/3.7} -0.02 4.90
uncl 0.44{0222/0.07} 3.3{0222/0.32} +0.07 3.97 **;**

The executable `simulate` should produce events in this form.

5. *Refinements*

Each experiment should also provide an executable `refine` with type

```
refine <input text file> <output text file>
```

that takes a standard QUAERO text file without systematic errors as input, refines the events, and produces a standard QUAERO text file as output.

The function of this executable is to refine the events after systematic offsets have been imposed. This refinement may include one or more of the following:

- There is typically an energy threshold below which objects are generally not well identified or well measured. Objects below this threshold should be removed from the event, and their energy transferred to the object `uncl`. Objects not in the fiducial region of the detector should be treated similarly.
- Experiments usually develop a set of criteria for determining whether a given event is worth analyzing. These criteria include but are not limited to the criteria imposed by the experiment's trigger. Events that fail to satisfy these criteria should be removed.
- The identification of objects may be modified to conform to the experiment's partitioning of final states.

Additional refinements may be imposed as appropriate.

IV. ALGORITHM

This section describes the QUAERO algorithm.

For a particular hypothesis $\mathcal{H}$, the quantity of interest is $\log_{10}\mathcal{L}(\mathcal{H})$, where

$$\mathcal{L}(\mathcal{H}) = \frac{p(\mathcal{D}|\mathcal{H})}{p(\mathcal{D}|\mathrm{SM})}, \tag{1}$$

$\mathcal{D}$ are the data, and SM is the standard model. The computation of this quantity requires

- a choice of variables for each final state in each experiment,
- a choice of binning,
- the calculation of the likelihood for each final state in each experiment,
- the combination of these likelihoods for all final states within an experiment,
- the combination of these likelihoods among experiments, and
- the incorporation of systematic errors.

We consider each in turn, concluding with a brief discussion regarding the interpretation of results.

A. Variables

Events with n final state objects populate a $3n-4$ dimensional space, where $2 \leq n \lesssim 6$. We are generally unable to model the full-dimensional space reliably with the limited number N_{MC} of Monte Carlo events at our disposal. We therefore restrict our attention to a d-dimensional subspace, where $d = \lfloor \log_{100} N_{MC} \rfloor$.[2]

We form this subspace by selecting variables from the following list:

- energy (E) of each object
- polar angle (θ) of each object
- distance in azimuthal angle ($\Delta\phi$) between each object pair
- distance ($\Delta\mathcal{R}$) in pseudorapidity and azimuth of each object pair
- invariant masses of all combinations of two or more objects
- topological variables sphericity ($\mathcal{S}$) and aplanarity ($\mathcal{A}$)

The variables x are ordered according to decreasing

$$\max_{x_0} \left| \int_{-\infty}^{x_0} p(x|\mathcal{H}) - \int_{-\infty}^{x_0} p(x|\mathrm{SM}) \right|. \tag{2}$$

Beginning with the first variable in this ordering and continuing until d variables have been chosen, we add the variable to those that we consider unless the smallest eigenvalue of the correlation matrix of this variable and the $q-1$ variables already chosen is smaller than $1/q$.

This is simply one of many possible prescriptions for choosing d weakly-correlated variables in which different distributions are predicted by the standard model and the hypothesis $\mathcal{H}$.

B. Choice of binning

Given N_{MC} Monte Carlo events each with weight w_i, we define an *effective* number of Monte Carlo events $N_{MC}^{\rm eff}$ by

$$N_{MC}^{\rm eff} = \frac{1}{\sum_{i=1}^{N_{MC}} w_i^2}, \tag{3}$$

In words, $N_{MC}^{\rm eff}$ is the reciprocal of the weighted average of the weights. Here and elsewhere we assume the total weight of the sample has been normalized to unity.

The Monte Carlo events are binned into $N_{\rm bins} = \left\lfloor \sqrt[2d]{N_{MC}^{\rm eff}} \right\rfloor^d$ bins, $\sqrt[d]{N_{\rm bins}}$ in each variable. An event $\vec{x}$ is assigned to bin j if

$$j = \sum_{i=1}^{d} (N_{\rm bins})^{(i-1)/d} \left\lfloor \omega_<(x_i) \sqrt[d]{N_{\rm bins}} \right\rfloor, \tag{4}$$

where $\omega_<(x_i)$ is the summed weight of all events with $i^{\rm th}$ variable less than x_i. This binning is rectangular, but forms an irregular grid.

C. Computation of the likelihood

Restricting ourselves for the moment to a particular final state (fs) within a particular experiment (exp),

$$p(\mathcal{D}_{({\rm exp})({\rm fs})}|\mathcal{H},\vec{s}) = \prod_{i=1}^{N_{\rm bins}} \frac{e^{-h_i} {h_i}^{N_i}}{N_i!}, \tag{5}$$

where $\mathcal{H}$ is the hypothesis under consideration, $\vec{s}$ is a vector of assumed systematic offsets, h_i is the number of events predicted by $\mathcal{H}$ in the $i^{\rm th}$ bin in this final state for this experiment, and N_i is the number of data events observed in that bin.

Similarly,

$$p(\mathcal{D}_{({\rm exp})({\rm fs})}|{\rm SM},\vec{s}) = \prod_{i=1}^{N_{\rm bins}} \frac{e^{-b_i} {b_i}^{N_i}}{N_i!}, \tag{6}$$

where SM is the standard model, and b_i is the number of events predicted by the standard model in the $i^{\rm th}$ bin in this final state for this experiment.

D. Combination of final states

Probabilities $p(\mathcal{D}_{({\rm exp})({\rm fs})}|\mathcal{H},\vec{s})$ from individual final states are combined into a probability $p(\mathcal{D}_{({\rm exp})}|\mathcal{H},\vec{s})$ for the experiment by multiplication:

$$p(\mathcal{D}_{({\rm exp})}|\mathcal{H},\vec{s}) = \prod_{\rm fs} p(\mathcal{D}_{({\rm exp})({\rm fs})}|\mathcal{H},\vec{s}). \tag{7}$$

Similarly,

$$p(\mathcal{D}_{({\rm exp})}|{\rm SM},\vec{s}) = \prod_{\rm fs} p(\mathcal{D}_{({\rm exp})({\rm fs})}|{\rm SM},\vec{s}). \tag{8}$$

E. Combination of experiments

Probabilities $p(\mathcal{D}_{({\rm exp})}|\mathcal{H},\vec{s})$ from individual experiments are combined into a total probability $p(\mathcal{D}|\mathcal{H},\vec{s})$ by another multiplication:

$$p(\mathcal{D}|\mathcal{H},\vec{s}) = \prod_{\rm exp} p(\mathcal{D}_{({\rm exp})}|\mathcal{H},\vec{s}). \tag{9}$$

Similarly,

$$p(\mathcal{D}|{\rm SM},\vec{s}) = \prod_{\rm exp} p(\mathcal{D}_{({\rm exp})}|{\rm SM},\vec{s}). \tag{10}$$

F. Incorporation of systematic errors

Systematic errors are incorporated by repeating the above steps many times with different systematic offsets $\vec{s}$, which allows the computation of the integrals

$$p(\mathcal{D}|\mathcal{H}) = \int p(\mathcal{D}|\mathcal{H},\vec{s})\, d\vec{s} \tag{11}$$

and

$$p(\mathcal{D}|{\rm SM}) = \int p(\mathcal{D}|{\rm SM},\vec{s})\, d\vec{s}. \tag{12}$$

The final single quantity of interest is the ratio of these quantities,

$$\mathcal{L}(\mathcal{H}) = \frac{p(\mathcal{D}|\mathcal{H})}{p(\mathcal{D}|{\rm SM})}, \tag{13}$$

although at times it is convenient to consider $\log_{10} \mathcal{L}(\mathcal{H})$.

G. Interpretation of results

For a given hypothesis $\mathcal{H}$, QUAERO's result takes the form of a single number $\mathcal{L}(\mathcal{H})$. In words, $\mathcal{L}(\mathcal{H})$ quantifies the extent to which the data support $\mathcal{H}$ in favor of the standard model. If our prior prejudice leads us to believe that the betting odds favoring $\mathcal{H}$ over the standard model are $p(\mathcal{H})/p({\rm SM})$, then QUAERO's result instructs us to modify those odds to

$$\frac{p(\mathcal{H}|\mathcal{D})}{p({\rm SM}|\mathcal{D})} = \frac{p(\mathcal{D}|\mathcal{H})}{p(\mathcal{D}|{\rm SM})} \frac{p(\mathcal{H})}{p({\rm SM})}. \tag{14}$$

The new betting odds are obtained from the old simply by multiplication by $\mathcal{L}(\mathcal{H})$.

This likelihood can be converted into more familiar forms. Results in high energy physics are often presented either in terms of a measurement of one or more parameters of a model (central value with one standard deviation errors), or in terms of an exclusion limit for one or more parameters of a model (typically at the 95% confidence level).

In either case the hypothesis $\mathcal{H}$ is taken to depend upon one or more parameters $\vec{\alpha}$. Just as in the parameter-free case treated above, if our prior prejudice leads us to believe that the betting odds favoring $\mathcal{H}(\vec{\alpha})$ over the standard model are $p(\mathcal{H}(\vec{\alpha}))/p(\mathrm{SM})$, then QUAERO's result instructs us to modify those odds to

$$\frac{p(\mathcal{H}(\vec{\alpha})|\mathcal{D})}{p(\mathrm{SM}|\mathcal{D})} = \frac{p(\mathcal{D}|\mathcal{H}(\vec{\alpha}))}{p(\mathcal{D}|\mathrm{SM})}\frac{p(\mathcal{H}(\vec{\alpha}))}{p(\mathrm{SM})}. \tag{15}$$

The new betting odds are obtained from the old simply by multiplication by $\mathcal{L}(\mathcal{H}(\vec{\alpha}))$.

The difference between making a measurement, making a discovery, and setting exclusion limits is then easily discerned: a measurement is being made if $\mathcal{L}(\mathcal{H}(\vec{\alpha}))$ shows a demonstrable peak in $\vec{\alpha}$; a discovery is being made if the hypothesis involves physics beyond the standard model and $\mathcal{L}(\mathcal{H}(\vec{\alpha})) \gg 1$; exclusion limits are set in all other cases.

In the case of a measurement, the distribution $p(\mathcal{H}(\vec{\alpha})|\mathcal{D})$ is typically fit to a multivariate gaussian in $\vec{\alpha}$ — the mean of the gaussian then corresponds to the measured central values and the covariance matrix to the errors on those values.

In the case of a discovery, the peak value of $\mathcal{L}(\mathcal{H}(\vec{\alpha}))$ should be quoted directly as a quantitative measure of the "significance" of the result.

In the case of exclusion limits, we typically introduce a cross section σ as a free parameter, ignoring for a moment that the predicted cross section $\sigma(\vec{\alpha})$ is generally a definite function of the parameters $\vec{\alpha}$. The hypothesis $\mathcal{H}(\vec{\alpha})$ is then said to be excluded at the 95% confidence level if

$$\int_0^{\sigma(\vec{\alpha})} p(\mathcal{H}(\vec{\alpha},\sigma)|\mathcal{D})\, d\sigma > 95\%, \tag{16}$$

assuming some prior $p(\sigma)$ for the cross section.

In all cases, the desired form of the result is easily obtained from the number that QUAERO provides.

V. EXAMPLES

LEP examples go here.

VI. CONCLUSIONS

QUAERO is a general interface to TeV-scale event data, enabling the testing of various hypotheses both within and beyond the standard model. Predicted events, supplied by a physicist unfamiliar with the details of the experiment, are refracted through a detector simulation and compared against the data and the standard model. This comparison is performed in a low-dimensional space of automatically chosen variables in all final states.

QUAERO is a proposed solution to a number of problems currently faced in high energy physics.

1. The archiving of data with QUAERO is straightforward, requiring little manpower relative to other schemes. By reducing events to 4-vectors of objects, tens of millions of events may be stored on a single hard drive. Data and backgrounds are stored in ASCII files, which will be readable indefinitely.

2. QUAERO's analyses are blind. The experimentalist understands the data, backgrounds, and systematic errors, and provides these to QUAERO without knowing the effect that each of his decisions will have on a final measurement of interest. Once these items have been provided, the analysis performed completely automatically by QUAERO, without any opportunity for human bias.

3. Using QUAERO, data can be made available to scientists outside the collaboration in a meaningful way. Other scientists are then able to test hypotheses against the data without requiring detailed knowledge of the experiment. A variety of policy options allow everything from the full QUAERO publication of an experiment's data to a more limited publication of those data, perhaps with internal collaboration review of QUAERO analyses.

4. QUAERO can rigorously combine the results of several experiments, with full incorporation of correlated systematic errors.

We hope that QUAERO will prove useful in testing future models against high energy collider data.

APPENDIX A: SYSTEMATIC ERROR SOURCES

Non-negligible sources of systematic error likely include the following:

- choice of Monte Carlo generator
- theoretical cross sections
- integrated luminosity
- b and τ identification
- electromagnetic and hadronic energy scale

Errors due to limited Monte Carlo statistics are automatically incorporated by QUAERO.

With these sources of error, the file `systematicSources.txt` might look like this:

systematicSources.txt		
1	c3	Monte Carlo generator
11	m	WW cross section
12	m	ZZ cross section
13	m	Z cross section
0201	m	Opal Luminosity
0211	i	Opal b-tagging
0212	i	Opal tau-tagging
0221	m	Opal EM energy scale
0222	m	Opal hadronic energy scale
0301	m	Aleph Luminosity

The left-hand column contains identifying integers for the various sources of error; the middle column provides the type of systematic error (mismeasurement, identification, or choice); and the right-hand column gives a description of the error. The notation c3 indicates that we have three Monte Carlo generators to choose among — we imagine these to be PYTHIA, HERWIG, and ISAJET.

The file `systematicCorrelations.txt` might look like this:

systematicCorrelations.txt		
11	12	0.9
12	13	0.9
0211	0212	0.3
0221	0222	0.4
0201	0301	0.75

The right-hand column gives the correlation between the sources of error specified by the left-hand and middle columns.

We can now take a closer look at the event we have been using as an example:

ZZ 0.0041{1/0,12/0.0002,0201/0.0001} 190.0
e+ 45.2{0221/1.4} +0.11 0.21
e- 47.3{0221/1.5} -0.05 3.56
b{0211/j/0.001} 4.2{0222/0.61} 46.0{0222/3.6} -0.16 1.71
b{0211/j/0.05} 4.3{0222/0.35} 48.2{0222/3.7} -0.02 4.90
uncl 0.44{0222/0.07} 3.3{0222/0.32} +0.07 3.97 **;**

The first systematic error on the weight is 1/0. By looking at `systematicSources.txt` we see that source 1 arises from the choice of Monte Carlo generator. Since source 1 is shown as type c3, a random integer between 0 and 2 will be generated. If that integer equals zero, this event will retain its weight of 0.0041; if otherwise, the weight will be set to zero. This allows us to randomly choose to use a sample of events generated with PYTHIA (to which this event belongs), HERWIG, or ISAJET, corresponding to integers 0, 1, or 2.

The second systematic error on the weight is 12/0.0002. Looking again at `systematicSources.txt` we see that source 12 arises from the theoretical uncertainty on the ZZ cross section. A gaussian random number (highly correlated with sources 11 and 13, the WW and Z cross sections, according to `systematicCorrelations.txt`) will be generated with zero mean and unit width; the weight of the event will be modified by this random number, scaled to a magnitude of 0.0002. The third systematic error on the weight is 0201/0.0001, due to an uncertainty in integrated luminosity.

Similarly, 45.2{0221/1.4} defines the effect of the uncertainty on the electromagnetic energy scale on the positron's energy in this event, and 46.0{0222/3.6} defines the effect of the uncertainty on the hadronic energy scale on one of the b quark's energy in this event. It should be emphasized that by "the uncertainty on the electromagnetic energy scale" we do not mean a sampling uncertainty on the energy measurement, which is taken into account in the background modeling without any need for an associated systematic error, but rather a possible systematic shift in the energy scale, resulting from miscalibration.

The notation **b**{0211/j/0.001} instructs QUAERO to call this object a jet (rather than a b quark) if the random number thrown for source 0211 is less than 0.001. The systematic error does not correspond to a mistag rate, but rather to an error on that rate.

This section is intended simply as an example of how systematic errors can be defined and imposed upon the quantities in a standard QUAERO file. The questions of which systematic errors to include and how each error affects these quantities is left to each individual experiment.

The assigning of appropriate correlations allows straightforward combination of results from several experiments.

APPENDIX B: HEPEVT

A template for reading a HEPEVT file into the HEPEVT common block is shown in Fig. 1.

[1] DØ Collaboration, V. Abazov *et al.*, submitted to Phys. Rev. Lett., hep-ex/0106039 (2001).

[2] $\lfloor\cdot\rfloor$ is the "floor" operator, denoting the largest integer not exceeding its argument.

```
C...HEPEVT commonblock.
      integer nmxhep
      PARAMETER (NMXHEP=4000)
      DOUBLE PRECISION PHEP,VHEP
      integer NEVHEP,NHEP,
     &        ISTHEP,IDHEP,JMOHEP,JDAHEP
      COMMON/HEPEVT/NEVHEP,NHEP,
     & ISTHEP(NMXHEP),IDHEP(NMXHEP),
     & JMOHEP(2,NMXHEP),JDAHEP(2,NMXHEP),
     & PHEP(5,NMXHEP),VHEP(4,NMXHEP)
C--------------------------------------------->

c     Open input hepevt file
      Open(Unit=90,File=inputHepevtFilename,
     & Status='old',Form='formatted',Err=110)

      READ(90,*) nEvents0
      do i=1,nEvents0
         READ(90,*) NHEP
         DO ii=1,NHEP
            READ (90,*) ISTHEP(ii),IDHEP(ii),
     &           (JMOHEP(J,ii),J=1,2),
     &           (JDAHEP(L,ii),L=1,2)
            READ (90,*) (PHEP(J,ii),J=1,5)
            READ (90,*) (VHEP(L,ii),L=1,5)
         END DO
```

FIG. 1: A template for reading a HEPEVT file into the HEPEVT common block.