Eur. Phys. J. C 53, 167–175 (2008)
DOI 10.1140/epjc/s10052-007-0468-3

THE EUROPEAN PHYSICAL JOURNAL C

Special Article – Tools for Experiment and Theory

QUAERO@H1: an interface to high-p_T HERA event data

S. Caron[1,a], B. Knuteson[2,b]

[1] Physikalisches Institut, Universität Freiburg, Freiburg, Germany
[2] Massachusetts Institute of Technology, Cambridge, USA

Received: 30 October 2006 / Revised version: 18 October 2007 /
Published online: 17 November 2007 –

Abstract. Distributions from high-p_T HERA event data analyzed in a general search for new physics at H1 have been incorporated into QUAERO, an algorithm designed to automate tests of specific hypotheses with high energy collider data. The use of QUAERO@H1 to search for leptoquarks, R-parity violating supersymmetry, and excited quarks provides examples to develop intuition for the algorithm's performance.

1 Introduction

The publication of most searches for new physics takes the form of an exclusion contour in a two-dimensional model parameter space, with alternative interpretations typically requiring a re-analysis of the data. Publishing the data themselves (either in raw form or as 4-vectors) is unsatisfactory, because substantial expert knowledge is required for an accurate analysis. This article continues a recently introduced paradigm of publishing frontier energy collider data via an interface that can be used to quickly test any specific hypothesis against collider event data, with the analysis performed by an algorithm that encapsulates expert knowledge of the experiment. In this way exclusion contours (or discovery regions) can be produced on demand. This type of interface can be used with well understood data to test models motivated either by new theoretical insights or from the consideration of subsequently collected data. Such an interface can also be used with freshly collected data in the exploratory process of fitting together an underlying physical interpretation for observed discrepancies while exploring in parallel more mundane experimental explanations.

The H1 detector has recorded collisions of electrons and protons at center of mass energies of 301 and 319 GeV in runs of the HERA collider between 1992 and 2000. These data have been used to understand a wide spectrum of physics, from the internal structure of the proton to tests of exotic physics at the electroweak scale. The analysis of HERA-I data has resulted in over one hundred publications in internationally recognized journals, and a general analysis of all high p_T final states has recently been performed in the H1 general search [1].

This article describes the interpretation of the distributions published by the H1 Collaboration in a general search for new physics using the QUAERO framework[1]. QUAERO was used previously by the DØ collaboration to automate the optimization of searches for new physics in the Tevatron Run I data [2]. The H1 general search is reviewed briefly in Sect. 2, with Sect. 3 covering QUAERO's knowledge of the H1 detector response. Section 4 briefly reviews the QUAERO algorithm. Section 5 contains the results of several analyses that have been performed using QUAERO@H1, allowing a comparison to previous results. A summary is given in Sect. 6.

2 General search

The H1 General Search has been published in [1]. This search, briefly described below, investigates events with high-p_T objects (electrons, muons, jets, photons, and the presence of missing transverse energy) produced in $e-p$ collisions at HERA. The histograms published by H1 (the invariant masses and the sums of the transverse momenta for high-p_T events) are used as input to the QUAERO algorithm in the studies described in this paper. This section briefly reviews the elements of this general search that have been incorporated into QUAERO@H1.

A detailed description of the H1 detector can be found in [3, 4]. The main trigger for events with high transverse momentum is provided by the liquid argon calorimeter. The trigger efficiency is close to 100% for events containing an electron or photon with transverse momentum greater than 20 GeV, 90% for events containing one or more jets with $p_T > 20$ GeV or with missing transverse momentum ($\not{p}_T$) greater than 20 GeV, and about 70% for di-muon events [1].

[a] e-mail: scaron@physik.uni-freiburg.de
[b] e-mail: knuteson@mit.edu

[1] The authors thank the H1 collaboration for providing the H1 general search histograms.

The H1 data available within QUAERO correspond to

- $36.4\,\mathrm{pb}^{-1}$ of 27.5 GeV positrons on 820 GeV protons, at a center of mass energy of 301 GeV;
- $13.8\,\mathrm{pb}^{-1}$ of 27.5 GeV electrons on 920 GeV protons, at a center of mass energy of 319 GeV; and
- $66.4\,\mathrm{pb}^{-1}$ of 27.5 GeV positrons on 920 GeV protons, at a center of mass energy of 319 GeV.

Standard object identification criteria are used to define electrons (e), muons (μ), photons (γ), and jets (j) [1]. All identified objects are required to have $p_T > 20$ GeV and $10° < \theta < 140°$, where θ is a polar angle in cylindrical coordinates with positive $\hat{z}$ in the proton direction. The charge of leptons is not distinguished. Exclusive final states containing different lepton charges are merged; only for technical reasons all leptons are assigned a positive charge in the figures. No attempt is made to identify jets containing heavy flavor. Events containing fewer than two of these objects are discarded. All objects are required to be isolated by a minimum distance of one unit in the η–ϕ plane, where η is pseudorapidity. A neutrino object ($\not{p}$) is defined for missing transverse momentum above 20 GeV. Restricting available information content to the histograms published in [1] require that missing energy is not used in the labeling of final states.

The experimental sources of systematic error affecting the modeling of these data are identical to those considered in [1]. Although QUAERO allows a full specification of correlated uncertainties, all H1 sources of systematic error are treated as uncorrelated.

Several Monte Carlo event generators are combined to estimate dominant standard model processes [1]. These generated events serve as the reference model to which hypotheses presented to QUAERO are compared. Here and below "standard model", "background", and "reference model" are used interchangeably.

3 TurboSim@H1

To keep QUAERO fast and standalone, a fast detector simulation algorithm (TURBOSIM@H1) is built in accordance with the H1 detector simulation. It is based on a large lookup table of one half million lines mapping particle-level objects to objects reconstructed in the detector. Particle-level jets are defined after hadronization using the inclusive $k_\perp$ algorithm as in [1]. Detector-level jets are also defined with this algorithm [1]. On the particle-level, only electrons, muons and photons that are not produced in the hadronization process are considered.

Sample lines in this table are shown in Fig. 1. The total table is roughly 100 MB, and as such can be read into memory and searched as a multivariate binary tree. The resulting simulation runs at roughly 10 ms per event.

Particle efficiencies are handled through lines in the TURBOSIM@H1 table that map a particle-level object to no reconstructed level object. Misidentification probabilities are handled through lines that map a particle-level object to a reconstructed-level object of a different type. The merging and overlap of particles is handled by configurations in the table that map two or three particle-level objects to zero or more reconstructed-level objects.

Validation of TURBOSIM@H1 has been performed by running an independent sample of one million events through both the H1 full simulation and TURBOSIM@H1.

#	EventType	Run.Event	zvtx(cm)	object	pT(GeV)	eta	phi(deg)		->	object	pT(GeV)	eta	phi(deg)	
1	ep->eMJX	240390.1795517551	2.98	e+	67.111	0.536	61.2	;	->	e+	72.969	0.531	61.37	;
2	ep->eMJX	268953.1873600019	18.04	e+	16.555	-0.038	47.21	;	->	j	18.955	0.003	46.05	;
3	ep->vMJX	271676.1791070467	5.46	j	86.379	1.689	50.96	;	->	j	81.975	1.686	50.95	;
4	ep->eMJX	272317.1802026211	29.48	j	46.449	1.464	-7.56	;	->	j	27.835	1.896	-21.23	
										j	23.815	1.008	10.42	;
5	ep->eMJX	278301.812519633	-5.72	j	15.606	0.225	-153.94	;	->	;				
6	ep->WX	257637.101591	-10.73	j	38.434	0.194	17.88	;	->	mu+	36.603	0.193	17.66	;
7	ep->llX	260716.176783	-3.16	e+	11.556	-0.72	-64.11							
				mu+	9.506	-0.777	-59.12	;	->	j	22.364	-0.81	-60.98	;
8	ep->eMJX	278996.533799894	-9.13	e+	30.031	0.004	12.62							
				ph	16.402	-0.055	10.59	;	->	e+	46.904	-0.028	10.96	;
9	ep->llX	253700.100976	3.13	mu+	24.717	1.161	-27.75	;	->	;				
10	ep->llX	256339.102973	-5.47	mu+	47.093	2.045	54.12	;	->	mu+	54.087	2.041	54.09	;

Fig. 1. Ten sample lines in the TURBOSIM@H1 lookup table, chosen to illustrate TURBOSIM's handling of interesting cases. Each line begins with the event's type, run and event number, and vertex position. *To the left* of the arrow ("→") is a list of nearby particle-level objects; to the right of the arrow is a list of corresponding reconstructed-level objects. The *first line* shows a positron correctly identified as a positron, while the *second line* shows a positron that has not been correctly identified. The *third line* shows a nicely reconstructed jet; in the *fourth line* the jet has been split into two; in the *fifth line* the jet is either not reconstructed or reconstructed with $p_T < 20$ GeV. The *sixth line* shows a jet identified as a muon. The *seventh line* shows a nearby positron and muon that have been reconstructed as one jet; the *eighth line* shows a positron with a photon radiated sufficiently nearby that the two are agglomerated into a single reconstructed-level positron. The *ninth* and *tenth lines* show muons that are missed and reconstructed, respectively

The event classification and the kinematic distributions of the events from the two simulation chains are compared using a Kolmogorov–Smirnov (KS) test. TURBOSIM@H1 has been found in agreement with the full simulation of H1.

4 Quaero

QUAERO provides a convenient interface to the understanding represented by high energy collider data, backgrounds, and detector response. This interface is designed to facilitate the test of any specific hypothesis against such data. The QUAERO web page, shown in Fig. 2, is available online at [5].

A physicist wishing to test her hypothesis against H1 data will provide her hypothesis in the form of commands to one of the built-in event generators [6–8]. QUAERO uses the specified event generator to generate signal events corresponding to $e^+ - p$ collisions at 301 GeV, $e^- - p$ collisions at 319 GeV, and $e^+ - p$ collisions at 319 GeV. The response of the H1 detector to these events is simulated using TURBOSIM@H1. The output of TURBOSIM@H1 is a text file containing simulated events, normalized such that the total number of expected signal events produced at H1 is equal to the sum of all event weights in the file.

Three distinct samples of events exist at this point: the data $\mathcal{D}$; the standard model prediction SM; and the hypothesis $\mathcal{H}$, which is the sum of included standard model processes and the physicist's signal. Each sample of events is partitioned into exclusive final states, categorized by reconstructed objects with $p_T > 20$ GeV. In each exclusive final state, a pre-defined list of two variables – the summed scalar transverse momentum ($\sum p_T$) and the invariant mass of all objects (m_{all}) – are ranked according to the difference between the standard model prediction and the physicist's hypothesis $\mathcal{H}$. The variable showing the most difference is used [2].

In this variable space, densities are estimated from the Monte Carlo events predicted by SM and $\mathcal{H}$. These densities are used to define a discriminant, which is binned to distinguish SM from $\mathcal{H}$.

The likelihood ratio $\mathcal{L} = p(\mathcal{D}|\mathcal{H})/p(\mathcal{D}|\text{SM})$ is determined using this binning, and systematic errors are integrated numerically.

The result returned by QUAERO is the decimal logarithm of this likelihood ratio. The measurement of model parameters using QUAERO is easily accomplished by graphing $\log_{10}\mathcal{L}$ as a function of varied parameter values, with multiple QUAERO submissions. Distributions of the standard model prediction SM, the prediction of the physicist's hypothesis $\mathcal{H}$, and the H1 data $\mathcal{D}$ in the most relevant variable in each of the most relevant final states are returned to the querying physicist in an email along with her result.

[2] Although the QUAERO algorithm is able to choose among many relevant kinematic variables, limitation to the information content of [1] requires restriction to either the single variable $\sum p_T$ or m_{all}.

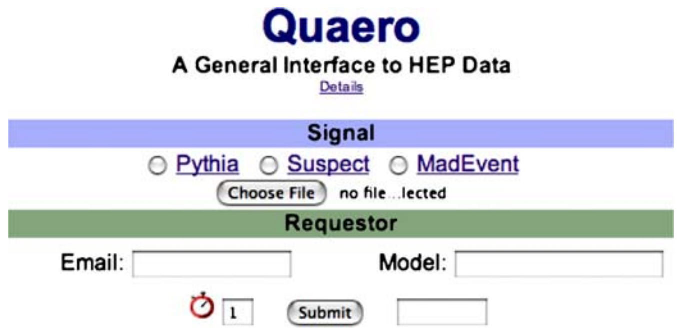

Fig. 2. The QUAERO interface, designed for HERA-I, LEP 2, Tevatron II, and the future LHC. A new hypothesis can be provided as commands to one of several event generators. The result returned quantifies the extent to which the data (dis)favors the new hypothesis relative to the standard model

Further details of the QUAERO algorithm are provided in Ref. [9].

5 Examples

QUAERO has been used to test models that have previously been considered at H1, in order to benchmark QUAERO's sensitivity, and to test models that have not yet been considered at H1, in order to see how QUAERO performs on novel searches for new physics. These examples provide intuition for the QUAERO algorithm: its strengths, and its limitations.

A rough, non-rigorous, but nonetheless useful comparison of the sensitivity of QUAERO's results (which take the form of the decimal logarithm of a likelihood ratio) with previous analyses (which typically take the form of 95% confidence level exclusion limits) can be made by comparing $\log_{10}\mathcal{L} = -1$ with the 95% confidence level exclusion limit [3]. The decimal logarithm of the likelihood returned by QUAERO can be converted into exclusion limits, measurements with errors, or (potentially) statements of discovery, perhaps with multiple QUAERO submissions.

[3] Consider a model with an overall cross section σ as its only free parameter, and a flat Bayesian prior on σ. After performing an analysis, suppose the posterior cross section distribution $p(\sigma)$ is a gaussian centered at zero, with the restriction $\sigma > 0$. Then models with σ greater than two standard deviations away from zero are excluded at a confidence level of 95%. These correspond to models with $\log_{10}\mathcal{L} < \log_{10}\exp\left(-\frac{1}{2}2^2\right) = -0.87$. Suppose instead the posterior cross section distribution $p(\sigma)$ is a decreasing exponential, with the restriction $\sigma > 0$. Then models with σ greater than three times the decay length of the exponential are excluded at a confidence level of 95%. These correspond to models with $\log_{10}\mathcal{L} < \log_{10}\exp(-3) = -1.3$. In this article $\log_{10}\mathcal{L} = -1$ is highlighted as a rough and convenient choice for the purpose of building intution when comparing with previous results.

The examples considered in this section include a search for scalar leptoquarks, R-parity violating supersymmetry, and an excited quark.

5.1 Leptoquarks

QUAERO@H1 is first used to search for leptoquarks, particles possessing both lepton and baryon quantum numbers that arise naturally in grand unified theories. Attention is restricted to a scalar leptoquark coupling to a positron and an up quark. The coupling λ of the LQ-e-u vertex and the leptoquark mass m_{LQ} are allowed to vary. The interaction Lagrangian is assumed to be of the form

$$\begin{aligned}\mathcal{L} &= \lambda \, \mathrm{LQ} \, \bar{u}_{\mathrm{R}} \, e_{\mathrm{L}} + \mathrm{h.c.} \\ &\quad + i g_s G^*_\mu \left(\mathrm{LQ}^* \overleftrightarrow{\partial}^\mu \mathrm{LQ} \right) , \end{aligned} \tag{1}$$

where LQ is a scalar leptoquark field; $\bar{u}_{\mathrm{R}}$ and e_{L} represent a right-handed anti-up quark and left-handed electron; G_μ is the gluon field; and $g_s = \sqrt{4\pi\alpha_s} \approx 1.2$. The MADEVENT [8] input to QUAERO corresponding to these Lagrangian terms is shown in Fig. 3.

The likelihood ratio QUAERO returns should be interpreted as an update of betting odds for or against the provided hypothesis. For the hypothesis of a scalar leptoquark with mass $m_{\mathrm{LQ}} = 301\,\mathrm{GeV}$, QUAERO finds $\log_{10}\mathcal{L} = -1.5$. If betting odds on this hypothesis were (say) $100:1$ against before looking at these data, then these data indicate those odds should be adjusted by an extra factor of $10^{-1.5} \approx 32:1$ against. Betting odds against this hypothesis after having run this request are now $3200:1$.

QUAERO also returns a link to plots showing how the result has been obtained. The cover page to these distributions contains the requester's name, email address, and the QUAERO request number, together with the contributions of each experiment to the final QUAERO result. One plot is returned for each of the final states contributing greater than 0.1 to the decimal logarithm of the likelihood ratio.

```
###############################################
           Name  Anti  Spin  Mass  Width  Color
           xxxx  xxxx  sfv   GeV    GeV    sto

PARTICLE    lq    lq~    s    301    0.54    t
###############################################
           Interacting particles    Coupling
INTERACTION   u        e-       lq        0.3
INTERACTION   e-       u        lq~       0.3
INTERACTION   g        lq~      lq        1.22
###############################################
```

Fig. 3. MADEVENT input given to QUAERO to specify the Lagrangian terms in (1). The line beginning with PARTICLE introduces into the theory a color triplet scalar leptoquark with mass 301 GeV. The three lines beginning with INTERACTION introduce new vertices and specify the coupling strength at each vertex. Here and elsewhere ~ denotes an anti-particle

The result of QUAERO's search for leptoquarks with interactions specified by the Lagrangian of (1) with $\lambda = 0.3$ and with mass up to 302 GeV is shown in Fig. 4. In all cases QUAERO finds $\log_{10}\mathcal{L} \le 0$, indicating the H1 data favor the standard model over the provided hypotheses.

In addition to varying the assumed leptoquark mass, the coupling strength λ, the signal production cross section, or an overall k-factor can be separately specified. An algorithm for turning this QUAERO request into a standard 95% confidence level exclusion limit involves submitting several requests assuming a range of leptoquark masses and a range of different cross sections at each mass, and computing the 95% confidence level cross section limit at each mass assuming a prior distribution for that cross section. Masses for which the 95% confidence level cross section limit is smaller than the theoretical cross section are said to be excluded at 95% confidence.

Intuitively, the 95% confidence exclusion region corresponds to a region where the signal hypothesis is moderately disfavored by the data. If the threshold is chosen to be $10:1$ against ($\log_{10}\mathcal{L} = -1$), then leptoquarks of mass less than 301 GeV are excluded. This result is in accord with the result of a previous analysis of this signal, described in [10], which derives a comparable limit for leptoquarks of this type.

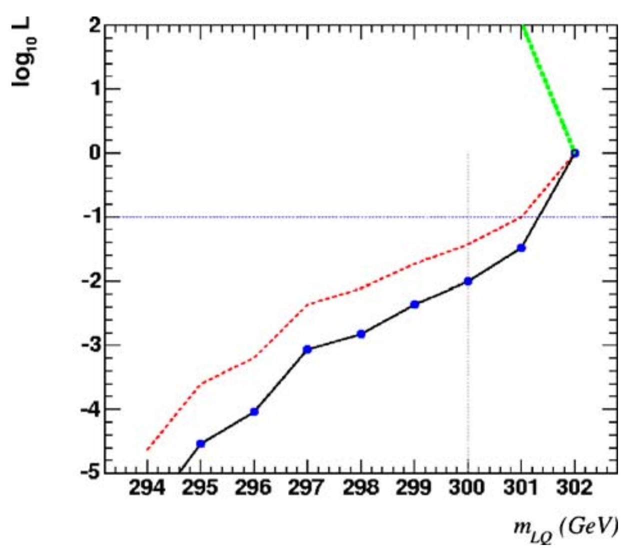

Fig. 4. QUAERO's output ($\log_{10}\mathcal{L}$) for the interaction Lagrangian of (1) with $\lambda = 0.3$ as a function of assumed leptoquark mass (m_{LQ}), shown as the solid line through filled circles. The evidence QUAERO expects to provide in favor of the standard model SM is shown as the (*dark, red*) *dashed curve* with $\log_{10}\mathcal{L} < 0$; the evidence QUAERO expects to provide in favor of the hypothesis $\mathcal{H}$ is shown as the (*light, green*) *dashed curve* with $\log_{10}\mathcal{L} > 0$. These expectations correspond to pseudo-experiments in which the pseudo-data are drawn from SM and $\mathcal{H}$, respectively. The result of a previous analysis excludes leptoquark masses $m_{\mathrm{LQ}} \lesssim 300\,\mathrm{GeV}$ at a confidence level of 95%, indicated by the vertical (*gray*) *dashed line*

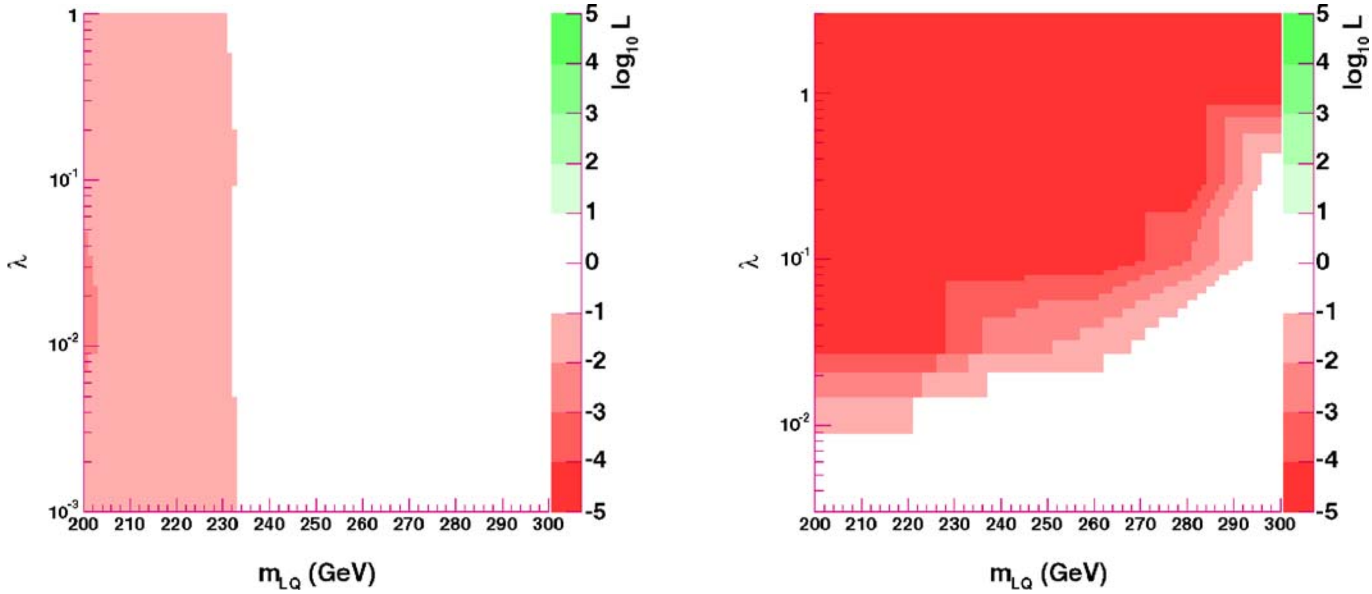

Fig. 5. QUAERO's log likelihood ratio as a function of the coupling λ and leptoquark mass m_{LQ}, in the scenario defined by the additional interaction terms of (1). Shown separately are results from QUAERO using data from DØ Tevatron Run I (*left*) and using data from H1 HERA Run I (*right*). All shaded area corresponds to $\log_{10} L < 0$

The subset of DØ Run I data made available in the first implementation of QUAERO [2] has been incorporated into the current version of QUAERO. The systematic uncertainties of these data are taken into account as described in [2].

Plots of QUAERO's result in the parameter plane of λ and m_{LQ} using the H1 and DØ data separately are shown in Fig. 5. QUAERO's result using H1 and DØ data combined is shown in Fig. 6. QUAERO is able to make use of the Tevatron's λ-independent exclusion of leptoquarks with low mass and HERA's λ-dependent exclusion at higher masses to rule out more of the parameter space than either collider is able to on its own. For a leptoquark mass of $m_{LQ} = 240\,\mathrm{GeV}$ and $\lambda = 0.01$ QUAERO finds, for example, that due to this combination $\log_{10} L$ decreases from 0.165 (using data from H1 HERA Run 1) to -0.295 (using data from H1 HERA Run 1 and D0 Tevatron Run 1). QUAERO's practical advantage in performing this type of combination lies in its speed, automation, and incorporation of information that is available but sometimes inconvenient to extract from published results.

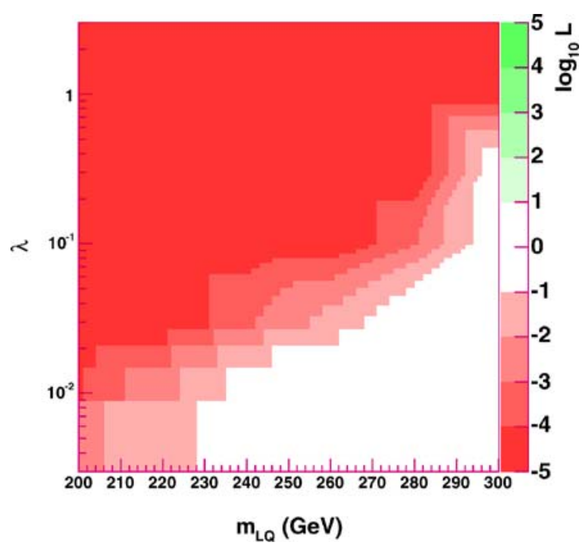

Fig. 6. QUAERO's log likelihood ratio as a function of the coupling λ and leptoquark mass m_{LQ}, in the scenario defined by the additional interaction terms of (1), combining data from DØ Tevatron Run I and H1 HERA Run I. All *shaded area* corresponds to $\log_{10} L < 0$

5.2 R-parity violating supersymmetry

An R-parity violating supersymmetry scenario is considered as a possible explanation for the $\mu j \nu$ events whose interestingness is quantified in [1]. R-parity violating supersymmetry can lead to squark production at HERA. If the muon sneutrino happens to be the lightest supersymmetric particle, then diagrams such as that shown in Fig. 7 can give rise to the final state $\mu j \nu$. This represents a scenario not yet tested by either of the HERA experiments.

Most new physics scenarios are specified "top-down", starting with the general Lagrangian and then restricting the resulting phenomenology through a judicious choice of values of the model parameters. In this subsection a "bottom-up" approach is adopted to illustrate another way in which QUAERO may be used to understand features of the data in terms of the underlying physical theory, in the spirit of BARD [12]. A new physics process is presumed, together with required masses and couplings, and rigorously tested using QUAERO. Iteration of this procedure allows a fit of parameters.

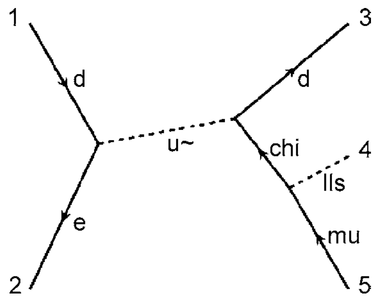

Fig. 7. A Feynman diagram generated by MADGRAPH [11] given the additional particles and interactions specified in Fig. 8. Here $u\sim$ is an up squark, *chi* is a chargino, and *lls* is a long-lived scalar, in this case a muon sneutrino

Ignoring the full sparticle spectrum, the new particles that must be postulated to allow the diagram shown in Fig. 7 are a scalar color triplet ($\tilde{u}$), an electrically charged fermion color singlet (χ^+), and a stable scalar without strong or electromagnetic interactions ($\tilde{\nu}_\mu$). Three new interactions must also be postulated, corresponding to the three vertices of Fig. 7.

These new particles and interactions can be provided to QUAERO as MADEVENT input, shown in Fig. 8, without the need to specify the complete renormalizable theory.

```
###############################################
              Name  Anti  Spin  Mass  Width  Color
              xxxx  xxxx  sfv   GeV    GeV    sto

PARTICLE      usq   usq~   s    200     1      t
PARTICLE      chi-  chi+   f    150     1      s
###############################################
              Interacting particles    Coupling
INTERACTION   d       e-       usq       0.01
INTERACTION   chi-    d        usq~      0.05
INTERACTION   mu-     chi-     lls       0.10

INTERACTION   e-      d        usq~      0.01
INTERACTION   d       chi-     usq       0.05
INTERACTION   chi-    mu-      lls       0.10
###############################################
PARAMETER  long_lived_mass = 50
###############################################
```

Fig. 8. MADEVENT input given to QUAERO to specify a signal corresponding to the new particles and interactions that must be introduced in order for a diagram such as that shown in Fig. 7 to occur. The two lines beginning with PARTICLE introduce into the theory a u squark with spin 0, mass 200 GeV, width 1 GeV, and a color triplet; and a $\chi^{\pm}$ with spin 1/2, mass 150 GeV, width 1 GeV, and a color singlet. The six lines beginning with INTERACTION introduce new vertices and specify the coupling strength at each vertex. The particle *lls* is a "long-lived scalar" that is implicitly added to the particle content of the theory; a long-lived fermion *llf* and a long-lived vector *llv* are similarly available. The line beginning with PARAMETER specifies the mass of the *lls*

Taking the widths of each new particle to be small compared to experimental resolution, this scenario contains six new parameters: three new particle masses and three new coupling strengths. The kinematics of the $\mu j \nu$ events suggest $m_{\tilde{u}} \approx 200$ GeV and $m_{\tilde{\nu}_\mu} \lesssim 50$ GeV, while bounds from LEP 1 indicate $m_{\tilde{\nu}_\mu} > M_Z/2$, leading to the choice of $m_{\tilde{u}} = 200$ GeV and $m_{\tilde{\nu}_\mu} = 50$ GeV for this example. The number of events observed by H1 in the $\mu j \nu$ final state suggest a signal production cross section of roughly 0.1 pb. Considering the single diagram of Fig. 7, the three new couplings affect the phenomenology of the postulated signal

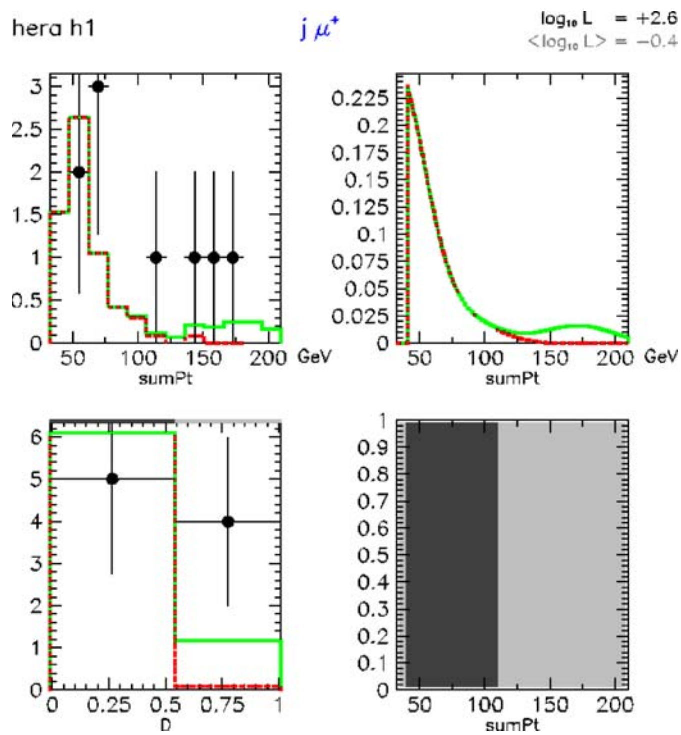

Fig. 9. Plots illustrating QUAERO's testing of the R-parity violating supersymmetry scenario described in the text, with the muon sneutrino playing the role of the lightest supersymmetric particle. The contribution of the postulated signal is seen (*upper left*) by comparing the distributions predicted by the hypothesis $\mathcal{H}$ (*light, green*) to the standard model (*dark, red*) in the final state $j\mu^+$ in QUAERO's chosen variable *sumPt* ($\sum p_T$). H1 data are shown as (*black*) *filled circles*. The *upper right* figure shows the density estimates constructed by QUAERO in this variable. Bins in the resulting discriminant are shown in the lower two figures. *Shading* in the lower two plots indicates the bin $D \lesssim 0.54$ corresponds to $40 \lesssim \sum p_T \lesssim 110$, and the bin $D \gtrsim 0.54$ corresponds to the signal-rich region $110 \lesssim \sum p_T$. The vertical axis in the lower right plot is without meaning, intended simply to give "thickness" to the shaded regions in *sumPt*. The number (*black*) at *upper right* is the log likelihood ratio $\log_{10}\mathcal{L}$ that would be found by QUAERO if that final state were analyzed alone, without integration over systematic errors; the number (*gray*) just below this is the log likelihood ratio expected if the data are drawn from the standard model distribution. Systematic errors are included in the final result returned by QUAERO

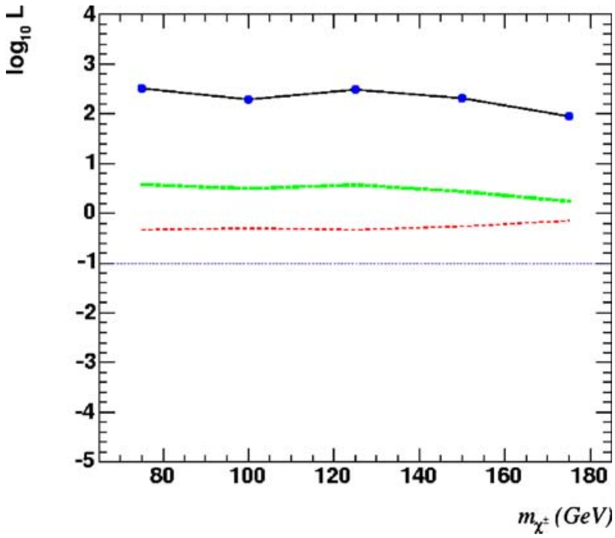

Fig. 10. QUAERO's log likelihood ratio (*solid curve*) as a function of assumed m_χ. Also shown is QUAERO's expected evidence (*dashed curves*) as a function of this parameter, if the data are drawn from the hypothesis (*light, green*) or the standard model (*dark, red*). The hypothesis provides a better fit to the data than the standard model for all values of m_χ in the range shown, corresponding to an update of betting odds of $\approx 300:1$ in favor

only by multiplying the overall cross section by the square of their product. The $\chi-d-\tilde{u}$ and $\mu-\chi-\tilde{\nu}_\mu$ couplings are of electroweak strength; the R-parity violating coupling $d-e-\tilde{u}$ is chosen so that the cross section of the process shown in Fig. 7 is 0.1 pb. The remaining parameter m_χ is allowed to vary between 75 GeV and 175 GeV.

Figure 9 shows QUAERO's analysis of this hypothesis with the choice $m_\chi = 150$ GeV. As expected, the final state $j\mu(\nu)$ contributes most to the final result. QUAERO chooses to consider the variable $\sum p_T$ in this final state. QUAERO's density estimates and choice of binning are shown in the right panes of Fig. 9. The left panes of Fig. 9 suggest that the data is better fitted by this hypothesis than by the standard model alone, a conclusion QUAERO quantifies in its returned result of $\log_{10}\mathcal{L} = 2.5$. [4]

QUAERO's result for $\log_{10}\mathcal{L}$ as a function of assumed m_χ is shown in Fig. 10. Over the range shown, this hypothesis is found to fit the data better than the standard model alone, and the decimal logarithm of the likelihood ratio is $\log_{10}\mathcal{L} \approx 2.5$. As mentioned previously, this likelihood ratio may be interpreted in terms of an "update of betting odds" for this hypothesis relative to the standard model. If betting odds against this hypothesis are (say) $10^7:1$ before the H1 data are considered, these betting odds should be revised in light of these data to $\approx 30\,000:1$ against.

[4] QUAERO's apparent insensitivity to m_χ is due to restriction to the crude variables $\sum p_T$ and m_{all}; allowed unrestricted choice of kinematic variables, QUAERO chooses observables that closely approximate m_χ. The model-independence of the variables $\sum p_T$ and m_{all} are the reason these variables are used in the model-independent analysis of [1], and the reason the variable $\sum p_T$ is used in SLEUTH [13–16].

5.3 Excited quark

If the known fermions are composite, a clear signal would be the detection of their excited states. At HERA, the hypothesis of an excited quark can be taken to correspond to new terms in an interaction Lagrangian of the form [18]

$$\mathcal{L} = \frac{1}{2\Lambda}\bar{q}^*_R\sigma^{\mu\nu}\left(g_s f_s \frac{\lambda^a}{2}G^a_{\mu\nu} + g f \frac{\tau}{2}\cdot\vec{W}_{\mu\nu} + g' f' \frac{Y}{2}B_{\mu\nu}\right) q_L + \text{h.c.}, \qquad (2)$$

where $\sigma^{\mu\nu} = (i/2)\,[\gamma^\mu, \gamma^\nu]$; g_s, g, and g' are the SU(3), SU(2), and U(1) gauge couplings; $G^a_{\mu\nu}$, $\vec{W}_{\mu\nu}$, and $B_{\mu\nu}$ are the SU(3), SU(2), and U(1) field strength tensors; Λ is the compositeness scale; q represents the first generation quark doublet; and q^* is the first generation excited quark doublet. These terms, with a specific choice of parameter values, can be specified with MADEVENT input shown in Fig. 11. The choice $f = f'$, $f_s = 0$, and $m_{u^*} = m_{d^*}$ leaves just two parameters: f/Λ and the excited quark mass m_{q^*}.

Parameters specified in Fig. 11 include the mass of the u^*, its coupling to its standard model counterpart and photon in the first two INTERACTION lines in Fig. 11,

```
################################################
           Name  Anti  Spin  Mass  Width  Color
           xxxx  xxxx  sfv   GeV    GeV    sto

PARTICLE    ux    ux~   f    200   0.04    t
################################################
       Interacting particles     Coupling
INTERACTION  u      ux      a    0.000667  dmx
INTERACTION  ux     u       a    0.000667  dmx

INTERACTION  d      ux      w-   0.00147   dmx
INTERACTION  ux     d       w+   0.00147   dmx

INTERACTION  u      ux      z    0.000825  dmx
INTERACTION  ux     u       z    0.000825  dmx

INTERACTION  ux     ux      a    0.3
INTERACTION  ux     ux      z    0.3
################################################
```

Fig. 11. MADEVENT input given to QUAERO to specify an excited quark signal. The symbols u and d denote the standard model up quark and down quark; *ux* and *dx* denote an excited up quark and an excited down quark. The line beginning with PARTICLE introduces into the theory an excited up quark with spin 1/2, mass 200 GeV, width 0.04 GeV, and a color triplet. The eight lines beginning with INTERACTION introduce new vertices and specify the coupling strength at each vertex, with *dmx* indicating a dipole moment coupling. An excited down quark and its interactions are similarly specified

its coupling to its unexcited weak isospin partner and W in the third and fourth INTERACTION lines, its coupling to its standard model counterpart and Z in the fifth and sixth INTERACTION lines, its coupling to the photon in the seventh INTERACTION line, and its coupling to the Z boson in the eighth and final INTERACTION line. An excited down quark d^* is introduced similarly.

Figure 12 summarizes QUAERO's analysis of this hypothesis with the choice $f/\Lambda = 0.01$ and $m_{q^*} = 200$ GeV. Shown are the four final states contributing most to the final result.

QUAERO's result $\log_{10} \mathcal{L}$ as a function of assumed coupling f/Λ and excited quark mass m_{q^*} is shown in Fig. 13. The current result from ZEUS [17], which makes use of $47.7\,\text{pb}^{-1}$ of $e^+ - p$ data at 300 GeV, is shown

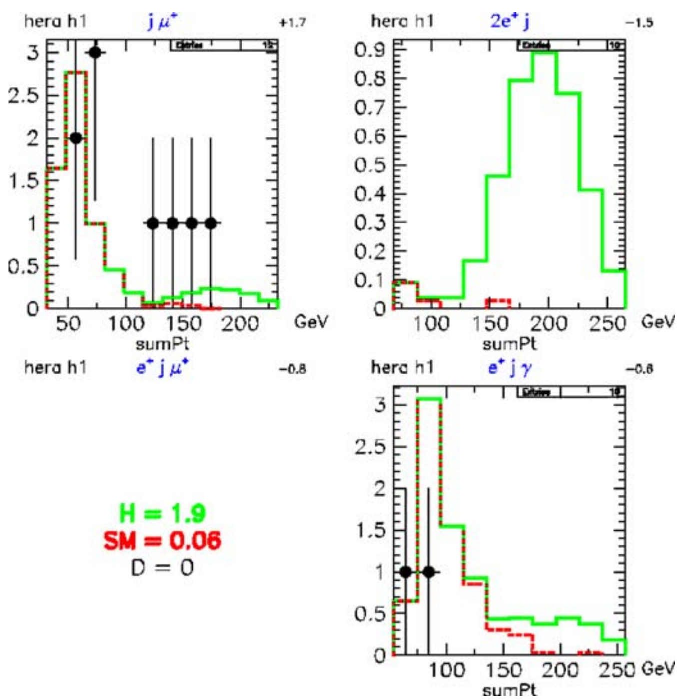

Fig. 12. Plots illustrating QUAERO's testing of the excited quark scenario described in the text, with parameters $f/\Lambda = 0.01$ and $m_{q^*} = 200$ GeV. The contribution of the excited quark signal is seen by comparing the distributions predicted by the hypothesis $\mathcal{H}$ (*light, green*) to the standard model SM (*dark, red*). H1 data are shown as (*black*) *filled circles*. The final state $2e^+ j$ contains no data events. The most useful final state for distinguishing between the excited quark hypothesis $\mathcal{H}$ and the standard model SM is found to be $j\mu^+$, followed in importance by $2e^+ j$, $e^+ j\mu^+$, and $e^+ j\gamma$. The most important variable chosen by QUAERO in each final state shown; in the final state $e^+ j\mu^+$, QUAERO decides it has insufficient Monte Carlo events to adequately populate a variable space, and simply considers the total number of events. The number (*black*) at *upper right* is the decimal log likelihood ratio $\log_{10} \mathcal{L}$ that would be found by QUAERO if that final state were analyzed alone, without integration over systematic errors. Systematic errors are incorporated into the final result returned by QUAERO

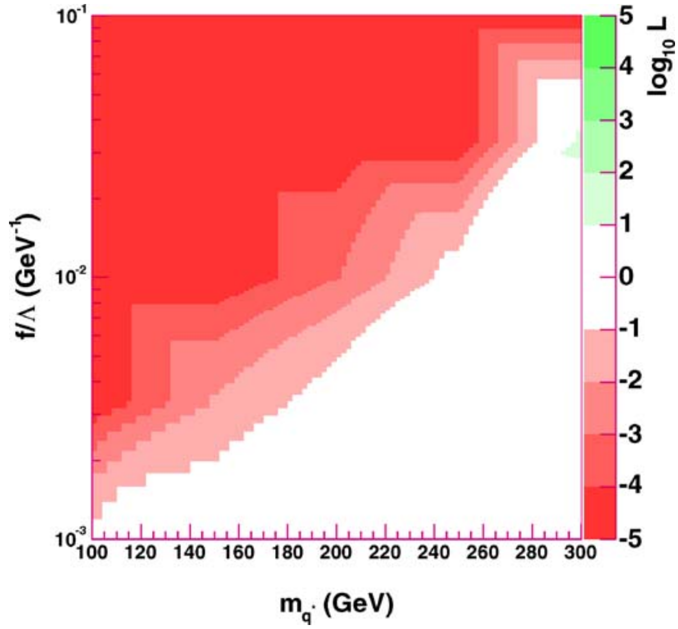

Fig. 13. QUAERO's log likelihood ratio as a function of f/Λ and m_{q^*}, under the assumptions $m_{u^*} = m_{d^*}$, $f = f'$, and $f_s = 0$. All *shaded area* corresponds to $\log_{10} L < 0$

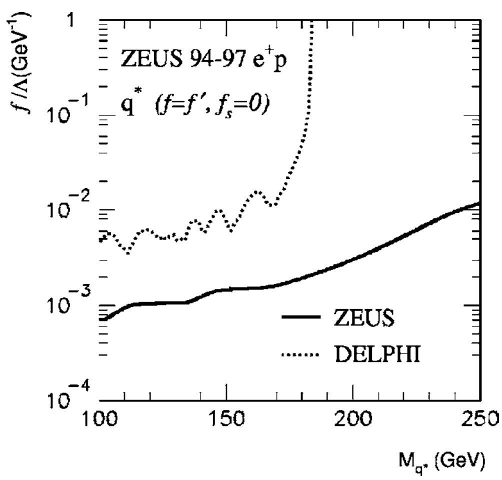

Fig. 14. A previous result from ZEUS in the parameter plane of f/Λ and m_{q^*}, under the same simplifying assumptions $f = f'$, $f_s = 0$, and $m_{u^*} = m_{d^*}$, from [17]

in Fig. 14. The differences between the QUAERO and the ZEUS limits are a result of the ZEUS analysis having a larger photon acceptance in polar angle and a larger photon identification efficiency for large excited quark masses.

QUAERO can also be used to test other parameter points within the parameter space defined by f, f', f_s, m_{u^*}, m_{d^*}, and Λ. Figure 15(left) shows the parameter plane of f/Λ and m_{q^*}, under the assumptions $m_{u^*} = m_{d^*}$, $f = -f'$, and $f_s = 0$. Figure 15(right) shows the parameter plane of f/Λ and m_{u^*}, under the assumptions $m_{q^*} \equiv m_{u^*} = 2m_{d^*}$, $f = f'$, and $f_s = 0$. All parameter points considered are found to have $\log_{10} L \leq 0$.

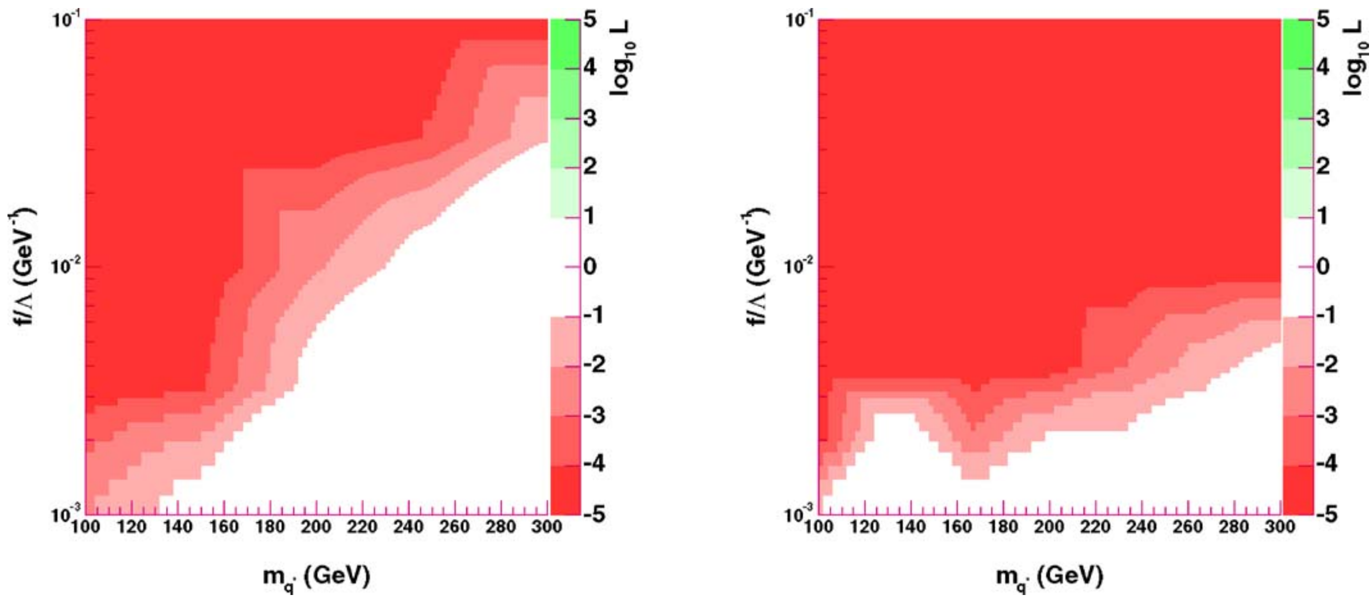

Fig. 15. QUAERO's log likelihood ratio as a function of f/Λ and m_{q^*}, under the assumptions $m_{u^*} = m_{d^*}$, $f = -f'$, and $f_s = 0$ (*left*), and under the assumptions $m_{q^*} \equiv m_{u^*} = 2m_{d^*}$, $f = f'$, and $f_s = 0$ (*right*). All *shaded area* corresponds to $\log_{10} L < 0$

6 Summary

The histograms of the invariant masses and the sum of transverse momenta from the high-p_T events selected in a general search for new physics at H1 have been incorporated into QUAERO, a framework for automating tests of hypotheses against data. The resulting interface is called QUAERO@H1. New physics scenarios can be provided to this interface in the form of commands to one of several commonly used event generators, which evaluate the short-distance consequences of each scenario. An automated analysis algorithm within QUAERO optimizes selection to distinguish between the new scenario and the standard model, returning a single number quantifying the extent to which the data (dis)favor the new hypothesis relative to the standard model alone, together with plots allowing the user to understand how the analysis has been performed. The use of QUAERO@H1 has been illustrated with searches for leptoquarks, R-parity violating supersymmetry, and excited quarks.

Acknowledgements. QUAERO is a simple framework for turning an understanding of collider data into statements about the underlying physics. Many H1 physicists are responsible for the achievement of this understanding. The authors would like to thank the Aachen and Marseille groups in particular for their roles in the H1 general search. Emmanuelle Perez, Cristi Diaconu, and André Schöning provided guidance and review within H1. Eckhard Elsen, Max Klein, Albrecht Wagner, and Rolf-Dieter Heuer have been sources of helpful guidance and comments within the wider DESY community. Stephen Mrenna and Tim Stelzer provided valuable assistance with the event generators that have been incorporated within QUAERO.

Sascha Caron acknowledges the support of a Marie Curie Intra-European Fellowship within the 6th European Community Framework Programme. Bruce Knuteson acknowledges support from a Department of Defense National Defense Science and Engineering Graduate Fellowship at the University of California, Berkeley; an International Research Fellowship from the National Science Foundation (INT-0107322); a Fermi/McCormick Fellowship at the University of Chicago; and Department of Energy grant DE-FC02-94ER40818.

References

1. H1 Collaboration, Phys. Lett. **602**, 14 (2004)
2. DØ Collaboration, Phys. Rev. Lett. **87**, 231 801 (2001)
3. H1 Collaboration, Nucl. Instrum. Methods A **386**, 310 (1997)
4. H1 Collaboration, Nucl. Instrum. Methods A **386**, 348 (1997)
5. B. Knuteson, 2006, http://mit.fnal.gov/Quaero/
6. T. Sjöstrand et al., Comput. Phys. Commun. **135**, 238 (2001)
7. A. Djouadi, J.-L. Kneur, G. Moultaka, hep-ph/0602101
8. F. Maltoni, T. Stelzer, J. High Energ. Phys. **02**, 027 (2003)
9. B. Knuteson, 2003, http://mit.fnal.gov/Quaero/quaero/doc/notes/algorithm.pdf
10. A. Aktas et al., Phys. Lett. **629**, 9 (2005)
11. T. Stelzer, W.F. Long, Comput. Phys. Commun. **81**, 357 (1994)
12. B. Knuteson, S. Mrenna, hep-ph/0602101
13. B. Knuteson, Nucl. Instrum. Methods A **534**, 7 (2004)
14. DØ Collaboration, Phys. Rev. Lett. **86**, 3712 (2001)
15. DØ Collaboration, Phys. Rev. D **62**, 092 004 (2000)
16. DØ Collaboration, Phys. Rev. D **64**, 012 004 (2001)
17. ZEUS Collaboration, Phys. Lett. **549**, 32 (2002)
18. U. Baur, M. Spira, P.M. Zerwas, Phys. Rev. D **42**, 815 (1990)