A COMPUTATIONAL PARTON FACTORY

Proposal for a Special Programme Grant

R.D. Ball, A.T. Doyle, A.D. Kennedy and M. Krämer

Proposed Starting Date: 1 October 1999

Over the next ten years, and possibly beyond, most of the effort and expenditure in UK particle physics will be on the search for new physics at the LHC. In the meantime, UK experimentalists will be involved in hadron collider physics at the Fermilab Tevatron. It is only reasonable that a certain proportion of effort in the theoretical particle physics community is similarly directed. Much of this work will happen of its own accord: many theorists are naturally drawn to the possibility of discovering new physical phenomena.

However, before one can disentangle the signals of new physics, it is first necessary to have a good understanding of standard model backgrounds. This will be particularly true at the LHC. There is thus a need to support new projects which, while not being quite so glamorous as discovery physics, are nevertheless going to play a crucial role in future discoveries. Without directed funding, some of these projects might simply not happen at all, due to their need for specific computational expertise and resources not otherwise available. Here we address the particular problem of the accurate global determination of parton distribution functions incorporating a full analysis of the asociated experimental and theoretical errors.

1. Background

Accurate information on the distribution of partons within hadrons is fundamental to all activity at hadron colliders since the rate of any process depends on the distributions of the relevant partons inside the colliding beams. Since it is not possible to calculate these distributions with any confidence, at least in the foreseeable future, they are obtained in practice by global fits to experimental data from well understood processes (which means in practice from inclusive processes at fixed target experiments, HERA and the Tevatron). These fits then form a benchmark for comparison of new data with the standard model, and their reliability is thus a crucial ingredient in searches for new physics. Recent examples of this process at work have been the inclusive jet excess at the Tevatron, and the high Q^2 excess at HERA. While probably not indications of new physics, these measurements have demonstrated rather clearly the pressing need for precise information on parton densities and their overall uncertainties, both experimental and theoretical.

Currently there are only two groups worldwide producing sets of globally fitted partons, CTEQ and MRS. The work of both these groups has in the past been invaluable in the analysis and interpretation of a wide range of experimental data. Both groups follow essentially the same procedure: they start with a certain initial parton parametrization, evolve these partons up to the data, compute a simple χ^2, and then adjust the parameters to obtain the best global fit. The NLO evolution is performed using FORTRAN evolution codes, written many years ago when there was only a small amount of relevant data, with large errors. The evolution step is consequently slow and innacurate [1]. As the quantity and quality of data improves, the need for faster more reliable evolution codes becomes ever more pressing. Both groups combine statistical and systematic experimental errors in quadrature, no attempt being made to incorporate the correlations between systematic errors so painstakingly determined by experimentalists. Some experimental groups now perform their own limited global fits (for example to all structure function data) precisely in order to examine the effects of the correlated systematics. Theoretical errors are even more difficult to determine reliably, partly because current global fits are always tied to a particular parametrization and its associated prejudices, but also because no attempt is made to ascertain the inevitably significant errors due to NNLO corrections and higher twists.

These shortcomings are the result largely of shortage of manpower, which makes it difficult to incorporate even some of the more conventional improvements that might be made: heavy quarks, threshold resummations, higher twist, etc. With present resources it is very difficult indeed to implement any of the more innovative ideas currently in circulation, particularly those regarding the development of faster more reliable evolution codes [1], the efficient incorporation of experimental systematics [2, 3, 4], and the determination of theoretical uncertainties due to choice of parametrization (see e.g. [5]). Furthermore since the early days of QCD phenomenology there has been significant progress in computer technology and in the development of high performance computing. We believe that these advances are not currently being fully exploited by the phenomenology community. The determination of parton distributions is very suited to parallel processing: the evolution to the data inevitably involves many time consuming, independent but very similar computations.

2. Personnel

An accurate and reliable determination of parton distributions and their errors from experimental data ideally requires expertise in three distinct areas: particle physics phenomenology and the application of perturbative QCD, the analysis of experimental data and its associated statistical and systematic errors, and last but not least the application of high performance computing. Edinburgh is an ideal base for such a new initiative, since here we now have the opportunity to bring together each of these different areas.

(a) On the physics side, we have recently set up a new group in particle phenomenology, with two new lecturers (**R.D. Ball** and **M. Krämer**) working in the application of perturbative QCD and the search for new physics at the LHC. While working at CERN R.D. Ball wrote an NLO evolution code to extract unpolarized parton densities from HERA structure function data at small x. These analyses were among the first to recognise the true nature of the small x rise [6], and the value of α_s determined from them [7] is still being quoted by the PDG. He was also involved in the writing of the first NLO evolution code for polarized parton distributions, and carried out a careful analysis of the errors in global fits, particular those resulting from parametrization dependences [8, 9, 10]. This code is still being used by the SMC for their own data analysis [11], and has also been used at DESY for studies of the physics obtainable from polarized colliding beams at HERA [12]. M. Krämer is currently at CERN, working on beyond the Standard Model phenomenology [13, 14] at the Tevatron and LHC. He has considerable expertise in the calculation of NLO cross-sections, in heavy quark production [15] and in the resummation of threshold logarithms [16]. R.D. Ball is currently theory organiser (with M. Dittmar) of the parton distributions subgroup of the CERN LHC working group. M. Krämer is currently working on several projects related to the heavy flavour and electroweak working groups of the CERN LHC workshop and he is coordinator of the quarkonium production subgroup. We are both members of the European EC network QCDNET.

(b) **A.T. Doyle** is a member of the nearby PPE group in Glasgow which is involved in the analysis of the structure function data from HERA and the understanding of the systematic uncertainties arising from deep inelastic scattering experiments [18]. In particular, they have considerable expertise in the proper treatment of correlated systematics whose effects have been underestimated in earlier analyses. In Edinburgh we also have two recently appointed lecturers in particle physics experiment (S. M. Playfer and F. Muheim), working mainly in B physics (at BaBar and LHC-B).

(c) On the computational side, **A.D. Kennedy** was recently appointed to a Chair of Computational Science and Directorship of the Maxwell Institute. This Institute was set up precisely to support applications of high performance computing. We also have local expertise at Edinburgh Parallel Computing Centre. The Cray T3D and T3E are already widely used for computations in lattice QCD, in particular by other members of UKQCD in Edinburgh (R.D. Kenway, K.C.Bowler and B.J. Pendleton). If this proposal is accepted, we will make use of the resources of the Maxwell Institute and EPCC for the benefit of high energy particle phenomenology as a whole. The participation of the Maxwell Institute and the availability of support from EPCC gives us the necessary experience and expertise in managing such a medium-scale software development project.

3. Proposal

We propose that a new project be set up with the sole objective of providing the resources for a new generation of reliable customized parton distributions, with associated experimental and theoretical error determinations. These will provide essential input to the prediction and eventual analysis of phenomena at future hadron colliders, in particular the upgraded Tevatron and ultimately the LHC. To achieve this we propose to make full use of state-of-the-art high performance computing. This will require a specialised programme dedicated to this one task, carried out over several years, and supported by new funding.

We will set up a series of milestones and a suite of test programs in order to keep the project on schedule, and we will make sure that the programs are written in a modular, portable, and well-documented way so that they will be easily usable by others in the future. The following steps would be implemented sequentially:

- the development of a fast and accurate evolution and fitting code, KWIKFIT, written by a professional programmer during the first year of the project. It would be robust and portable enough to run on a wide range of computing facilities. In particular it could be set up to run on parallel machines. It would be written in C^{++}, in order to be compatible with other LHC software. We believe that an optimised evolution code in Mellin space, which includes a systematic and exhaustive use of look-up tables, could be an order of magnitude faster than existing x-space codes. Parallel processing the fitting should result in a further improvement of between one and two orders of magnitude, so that eventually it will be possible to obtain fits between two and three orders of magnitude faster than with the codes currently in use by MRS and CTEQ. A standard fit to several thousand data points might then take little more than a few minutes.

• the extra speed could be used to produce parton distributions with properly determined correlated experimental error bands. For this it would be necessary to include a proper treatment of the propagation of correlated systematic errors from each dataset, in collaboration with the relevant experimentalists. This would involve a considerable amount of hard work to incorporate the correlated systematic errors into the fitting code in such a way that the time required to obtain a global fit is not increased too much.

• To control the theoretical error, we would develop techniques to produce partons free from the bias introduced by a particular parametrization. Contrary to popular belief, Mellin space codes are no handicap in this respect, if one makes judicious use of lookup tables. Current ideas would need to be developed to a point where they became practical. A fast underlying evolution code would be essential to making this happen.

• the inclusion of orthodox theoretical refinements: the correct incorporation of heavy quarks, the resummation of threshold logarithms, a phenomenological treatment of higher twist, etc. Some of these improvements are currently being incorporated in existing codes, though time (both computing and manpower)is always a limiting factor. More speculative ideas could also be investigated. Furthermore, parameters relating to new physics (for example contact interactions) could be included directly in the fits.

• a user friendly front end, WWW based, to enable phenomenologists and experimentalists to 'order' a set of customized partons to suit their own requirements (for example, based on specific subsets of available datasets, specific choices of α_s, etc). This process could be partly or perhaps eventually fully automated for simple fits. We would also make the code itself available on the web: the idea would be to create a flexible resource for use of the whole particle physics community, not simply a small set of parton parametrizations.

• at a future date it may be useful to extend the analysis to the simultaneous fitting of polarized structure functions. Fits to photon structure functions and eventually to fragmentation functions would also be considered at a later stage.

4. Resources Requested

To carry out this programme at a scale and pace optimally matched to the Tevatron and LHC experimental programme will require extra resources. To meet the computational requirements we would need to employ a physicist programmer with a background in particle physics phenomenology, who also has a working knowledge of applied statistics and recent experience in high performance computing. We expect to be able to have such a person in post by October 1999. We anticipate no problems in finding suitable candidates, given the close relation of our group with EPCC. Once the initial coding had been set up, we would also require a regular RA, working closely with the programmer, to develop some of the less purely computational aspects of the programme. It would be an advantage to be able to employ someone more senior than usual, since the work involved would involve a measure of experience with the fitting and use of parton distributions.

We request that the programme be funded initially for a period of three years, with a starting date of 1 October 1999. This should be sufficient to write the main code and generate our first fits. We would expect to submit a second grant application for a further three years in 2002, subject of course to the successful completion of the first stage, to ensure that we can make the most of the code in the build up to the first results from LHC in 2005.

(i) *Staff:* We request a physicist programmer from 1 October 1999, starting at spine point 6, for 1 year initially and then 50% for each of two further years. We further request one two-year RA post from 1 October 2000 starting at spine point 8.

(ii) *Travel and Subsistence:* We request formula funding to attend international meetings such as DIS00,01,02,QCD00,01,02. This will be an important part of the programme, both for disseminating the results and for obtaining feedback from the both the theoretical and experimental communities about the ways in which the programme might best develop. We also request £500 per annum for travel within the UK (between Edinburgh and Glasgow, and to Durham, etc.)

(iii) *Consumables:* We request formula funding for the three years of the grant period.

(iv) *Computing:* This project is dependent on a state-of-the-art computing environment. We request support according to formula for the three years of the grant period for local computing facilities. We would also require time on the Edinburgh T3E, both for code development and final runs. Complete global fits using current programmes require of the order of between 100 and 500 processor hours: we hope to gain one order of magnitude through code development, but treating experimental systematics, theoretical systematics, and theoretical refinements are estimated to cost an order of magnitude each. While it is difficult to estimate exactly just how efficient our programmes will be, a reasonable estimate for the required computer time would be 90,000 processor hours: this should be ramped up over the three years, 5,000 hours in year 1, 25,000 in year 2, and 60,000 in year 3. We therefore request this amount of T3E time. The amount of disk space required is very small. Since the project itself will be based in Edinburgh, it should be easy to obtain the necessary specialist support.

5. Summary

We request a Special Programme Grant to set up a Computational Parton Factory. We believe this proposal to be timely, useful, well defined and practical. It could become an important component in the development and exploitation of physics at the LHC. We already have the necessary experience, ideas, energy and enthusiasm. In Edinburgh we can draw together phenomenologists, experimentalists and specialists in high performance computing to create a truly effective programme. All we need are the additional resources from PPARC without which it will not happen.

References

[1] J. Blumlein, S. Riemersma, M. Botje, C. Pascaud, F. Zomer, W.L. van Neerven and A. Vogt, A Detailed Comparison of NLO Evolution Codes, `hep-ph/9609400`.

[2] D.E. Soper and J.C. Collins, Issues in the Determination of Parton Distribution Functions; `hep-ph/9411214`.

[3] D.A. Kosower, Extracting Parton Distributions from Collider Data, Nucl.Phys.B520:263-278,1998; `hep-ph/9708392`.

[4] W.T. Giele and S. Keller, Implications of Hadron Collider Observables on Parton Distribution Function Uncertainties, Phys.Rev.D58:094023,1998; `hep-ph/9803393`.

[5] William Bialek, Curtis G. Callan, Steven P. Strong, Field Theories for Learning Probability Distributions, Phys.Rev.Lett.77:4693-4697,1996.

[6] R.D. Ball and A. De Roeck, Structure Functions, in the proceedings of International Workshop on Deep Inelastic Scattering and Related Phenomena (DIS 96), Rome, Italy, 15-19 Apr 1996; `hep-ph/9609309`.

[7] R.D. Ball and S. Forte, Determination of α_s from F_2^p at HERA, Phys.Lett.B358:365-378,1995; `hep-ph/9506233`.

[8] R.D. Ball, S. Forte and G. Ridolfi, NLO Determination of the Singlet Axial Charge and Polarized Gluon Content of the Nucleon, Phys.Lett.B378:255-266,1996; `hep-ph/9510449`.

[9] G. Altarelli, R.D. Ball, S. Forte and G. Ridolfi, Determination of the Bjorken Sum and Strong Coupling from Polarized Structure Functions, Nucl.Phys.B496:337-357,1997; `hep-ph/9701289`.

[10] R.D. Ball and H.A.M. Tallini, Spin Physics, to be published in the proceedings of 3rd UK Phenomenology Workshop on HERA Physics, Durham, England, 20-25 Sep 1998; `hep-ph/9812383`.

[11] Spin Muon Collaboration, A NLO QCD Analysis of the Spin Structure Function g_1, Phys.Rev.D58:112002,1998.

[12] A. De Roeck et al, A Study of the Polarized Structure Function g_1^p and the Polarized Gluon Distribution at HERA, Eur.Phys.J.C6:121-131,1999; `hep-ph/9801300`.

[13] W. Beenakker, M. Krämer, T. Plehn, M. Spira, P.M. Zerwas, Stop Production at Hadron Colliders, Nucl. Phys. B515 (1998) 3-14.

[14] M. Krämer, T. Plehn, M. Spira, P.M. Zerwas, Pair Production of Scalar Leptoquarks at The Tevatron, Phys.Rev.Lett.79:341-344,1997.

[15] M.E. Hayes and M. Krämer, Heavy Flavor Production at HERA, summary of the working group on 'Heavy Flavour Physics' at 3rd UK Phenomenology Workshop on HERA Physics, Durham, England, 20-25 Sep 1998; `hep-ph/9902295`.

[16] M. Krämer, E. Laenen, M. Spira, Soft Gluon Radiation in Higgs production at the LHC, Nucl. Phys. B511 (1998) 523-549.

[17] ZEUS Collaboration, Measurement of the F_2 Structure Function in Deep Inelastic Scattering Using 1994 Data from the ZEUS Detector at HERA, Z.Phys.C72:399-424,1996; `hep-ex/9607002`.

[18] A. T. Doyle, Structure Functions, talk given at the 29th International Conference on High-Energy Physics (ICHEP 98), Vancouver, Canada, 23-29 Jul 1998; `hep-ex/9812029`.

Main Objectives:

The development of a fast customised evolution code to evolve parton distributions and fit them to experimental data.

The application of the code to provide parton distributions to the particle physics community well before the start of the LHC.

Abstract:

A grant is requested to set up a new initiative in QCD Phenomenology: a Computational Parton Factory. The aim is to systematise and extend global parton fits so as to provide a professional service to the whole particle physics community, in time for the startup of the the LHC. The applicants have expertise in particle phenomenology, particle experiment and high performance computing. This will thus be the first initiative of its kind. The programme will be centred on the development of a customised computer code which evolves parton distributions quickly and accurately, fits them to experimental data with proper treatment of experimental errors, particularly correlated systematics, incorporates theoretical refinements such as heavy quarks and threshold resummations, and makes every attempt to estimate theoretical uncertainties. The programme will be made as widely available as possible, probably through a web-based interface.