OPPORTUNITIES FOR STATISTICAL SIGNAL PROCESSING IN HIGH ENERGY PHYSICS

Bruce Denby

Université Pierre et Marie Curie
UPMC/LISIF, 4 place Jussieu
75252 Paris cedex 05, France
email : denby@ieee.org

ABSTRACT

Data processing in high energy physics experiments is a multi-tiered process in which raw detector signals are first processed locally into physics objects, and then collated into event records which can be scrutinized by a fast online trigger system. The resulting selection of events are reconstructed and pass through a number of software filters before arriving at a final offline analysis where hard physical constants are extracted.

Although sophisticated statistical data analysis techniques are routinely employed high energy physics, the use of statistical signal processing in the field is has until now been rare.

Our paper will begin with an overview of a typical high energy physics data acquisition system, outlining the technologies and tradeoffs involved at each stage. We will then proceed to argue that the dominant roles of model dependence and systematic errors in final physics analyses render statistical signal processing techniques largely inapplicable at this level.

We observe, however, that at the low-level pattern recognition and event reconstruction levels, statistical signal processing techniqes have been making inroads in high energy physics for a number of years, and examples from the literature will be cited. The viability of the technique for second level triggers will be assessed. Parallels to other other approaches, such as neural networks, will also be drawn. It will be argued that the falling cost of computing hardware favors the growth of statistical signal processing methods in high energy physics.

1. INTRODUCTION

Experimental High Energy Physics (HEP) strives to study the fundamental properties of matter by observing the interactions between its fundamental constituents, the elementary particles. In order to probe the finest structures of these interactions, it is necessary to induce collisions between the particles at very high energy – of the order of 10^{12} eV, or 1 TeV – whence derives the name of this field which was, at its origin, an outgrowth of nuclear physics.

Perhaps the most fundamental equation for understanding the nature of a high energy physics experiment is

$$\mathbf{R} = s\,\mathbf{L} \quad (1)$$

where R is the interaction rate in events per second; σ, the so-called "cross section", is a measure of the intrinsic probability of the interaction to be studied; and L, the luminosity, is the number of incident particles per cm^2 per second furnished by the accelerator facility at which the experiment is to be conducted. If a team is to secure a statistically significant sample of events for a physical process on a reasonable, human timescale, it is clear that R must be at minimum several tens per year. The tiny naturally occurring cross sections of processes of interest to high energy physicist today then lead to values of L which are extraordinarily challenging from a technological standpoint, both in terms of the physical apparatus needed to produce and contain the high energy interactions, and the subsequent computational machinery used to extract the delicate information they contain.

2. THE MULTI-LEVEL HEP DATA ANALYSIS CHAIN

It is important to understand at the outset that high energy physics data possesses intrinsic time segmentation. Each particle collision, or event, as they are more frequently called, occurs at a precise time and place within the detector assembly, and – except for spurious overlaps, which for the sake of simplicity we shall ignore – is completely uncorrelated with the others. This observation precludes any talk of statistical signal processing on a time scale larger than that of the event itself – a point which is sometimes not appreciated by those outside the HEP community. In the following, we describe the multi-level data analysis chain that is relatively uniformly used by today's major HEP experiments. The interested reader may refer to [1] for more details concerning the design and operation of a large, modern experiment (ATLAS).

2.1. The trigger

The first piece of equipment encountered by the signals produced in the experimental apparatus is the so-called *trigger*. The need for it is purely technological, and presents a number of challenges. Although cross sections for the processes to be studied are very small, there are many background processes – interactions of beam particles with residual gas atoms or with a collimator, for example – whose effective cross sections are several orders of magnitude larger. The very high L values cited earlier then lead to enormous event rates for these processes. The data produced, therefore, is a mixture of huge numbers of background events – rates easily in the MHz – and a tiny number – perhaps only a few per year – of interesting, physics events. The purpose of the trigger is to reject the background events and allow only the interesting events to pass. As the final data logging stages are only able to record at a few Hz, the trigger must have both an exceedingly high rejection rate *and* very high efficiency for the events of interest. These contrasting technological goals are usually met using a 3-level trigger scheme :

- **Level-1**. With a processing time about 1 microsecond, level-1 is typically implemented with analog thresholding, and outputs an enriched sample of events at a few kHz.
- **Level-2**. The next processing level is based on pattern recognition and has a time constraint of some tens of microseconds. It is usually implemented with fast digital hardware such as FPGA's or ASIC's, which produce an output stream at several tens of Hz.
- **Level-3**. The final stage is usually a farm of standard CPU's, each working on different events, applying more sophisticated reconstruction and algorithms, which produces the final output stream of a few Hz to be transferred to permanent recording media.

2.2. Offline data analysis

Once the data acquisition has been acomplished, the offline data analysis can begin. As the data sample is now substantially enriched in interesting events, strong constraints on processing time are no longer present, and the full panoply of statistical tools can be applied. There are two distinct phases of data analysis which take place in this final, offline phase :

- **Reconstruction**. The raw hits in tracking chambers, muon hodoscope signals, calorimeter energy deposits, silicon vertex detector readouts, etc., must be transformed into physically salient variables. The reconstruction of each detector system is usually undertaken by a specialised team. The output of this stage is, for each event : lists of charged and neutral tracks; identified muons, electrons, and possibly other particles; groupings of particles into hadronic jets; labelling of missing transverse momentum and other kinematic flow variables; as well as the spatial coordinates of all of the aforementioned so-called *physics objects*.
- **Physics Analysis**. This is where the actual extraction of meaningful physical quantities is accomplished, using the physics objects of the reconstruction pass as inputs. The output could be a measurement of the mass of a particle, the branching ratio of a process into different subprocesses, or simply an estimation of the probability of existence or non-existence of a particular process. It is at this stage that the notion of *confidence levels* first emerges, enabling physicists to state the confidence with which they are able to make the various assertions that will enter into a final physics publication.

The applicability of statistical signal processing in the various processing levels – or the lack of it – will be discussed in the following sections, but first, it is interesting to examine some of the history of digital signal processing techniques in HEP.

3. THE DIGITAL SIGNAL PROCESSING REVOLUTION AND HEP

Although the foundations of digital signal processing date to the first decades of the 20th century, it did not begin to become an established field as we know it today until the 1970's and 1980's. Quite naturally, some of the first fields to profit from digital signal processing were speech and audio processing, since the low frequencies involved could easily be handled by the early Nyquist rate ADC's. Later, as sampling technology improved, radio and high frequency systems also began to profit from digital signal processing techniques.

The adoption of classical digital signal processing techniques in HEP was somewhat slower, because of the nature of the signals produced in HEP detectors. At the energies necessary to produce the hard collisions necessary to extract new physics, particles are travelling near the speed of light. Consequently, in traversing or depositing their energy in a detector measuring some meters in size, the duration of signal pulses produced will typically be only of the order of a few nanoseconds. Nuclear or atomic cascade effects can in some types of detectors extend this to microsecond or even millisecond levels, but the real time nature of HEP data acquisition systems demands that the decision-making hardware must function as quickly as possible, leading trigger designers to prefer prompt signals.

Thus, traditionally, HEP data acquisition systems simply integrated detector signals the over the duration of the event gate, applied thresholds to these, and combined the resulting flags using fast digital coincidence circuits. With the advent of flash ADC's, it became possible to store the instantaneous values of detector signals and use them in more sophisticated level-1 triggers, or as input to level-2 triggers. However, until fairly recently, it has been rare to find, in HEP, electronic signals which are sampled at regular intervals in time, as is the standard in other fields. As such, the applications to date of digital signal processing in HEP have been limited. It is probably also fair to say that these technological limitation on the usefulness of digital techniques in HEP led to "cultural limitations" on their use as well, as physicists tended not to be familiar with digital signal processing practice applied in the traditional sense. This was not part of a tradional physicists "education." This is now beginning to change, particularly as faster ADC's begin to enable new approaches to HEP data acquisition.

4. OPPORTUNITIES FOR STATISTICAL SIGNAL PROCESSING IN HIGH ENERGY PHYSICS

For simplicity purposes, in this article, we shall assume that by "statistical signal processing," or SSP, we refer to digital signal processing techniques applied to

- Detection and estimation theory, including optimal filters
- Machine learning and applications to pattern recognition.

In our, admittedly simplified, list, we have not included applications to adaptive systems, for example. Indeed, in HEP, although there are numerous dynamic aspects in accelerator physics, the detectors, calibration constants, and analysis processes are considered to be fixed systems, and ones to which, furthermore, adaptation would probably, due to the high channel counts, rapidity of execution, and experimental complexity, be very difficult to realize.

In the following, we shall examine how the techniques enumerated in our list may be used in HEP, or, in some instances, how they are already beginning to be used. We shall also point out those areas in which SSP may very well *not* be indicated. Before beginning those discussions, however, it is interesting, from a historical perspective, to take a look at an earlier "revolution" in HEP, that of neural networks.

4.1. Neural networks in HEP

From about 1988 to 1995, neural networks took HEP by storm [2, 3, 4]. In fact, it may be in part due to the lack of a "statistics culture" among high energy physicists that the promise of this new technique was embraced in such a wide range of applications in HEP. Neural nets were considered in everything from the lowest triggering levels all the way up to establishing confidence levels for final physics analyses. The traditionally conservative HEP community put up stiff resistance to those advocating neural networks, particularly as concerns their application to physics analyses. Today, although neural networks have been accepted as a tool in HEP, they have only really found their place in low-level pattern recognition tasks. One almost never sees today, for instance, a neural net used in a physics analysis. The reasons for the applicability or lack of it in the different HEP data acquisition tasks are rather clear, and, as neural networks also encompass machine learning applied to pattern recognition, the experience with neural networks will be an excellent guide in our discussion of the applicability of SSP. To do this, we shall break HEP data analysis into two broad classes : physics analysis; and low level pattern recognition.

4.2. SSP for physics analyses

The hard currency of HEP is new physics published in archival quality peer-reviewed scientific journals. It is natural for physicists armed with promising new tools, therefore, to try to use them to attack this "holy grail," since the payoffs here in improved signal to noise ratio or absolute precsion would be the biggest. Unfortunately, this level of the analysis is plagued by two problems which render SSP, as was the case for neural networks, largely inapplicable :

- **Model Dependence**. The goal of HEP is to discover the fundamental laws underlying the nature of matter. This implies that these laws are *a priori* unknown. The exact way in which the outcome of an interaction of elementary particles will decompose itself into physics objects can only be modelled according to a set of reasonable physical assumptions. The free parameters in such models typically leave considerable leeway in final state parameters. A classic example is the number of "jets" in an event, which is critically dependent on the parton hadronization model used and the details of the detector jet finding algorithm. A machine learning approach, for example, would be useless here since the *dominant* uncertainty depends not on signal to noise ratio but on the model dependence.

- **Systematic Errors.** Final results in physics analyses quote both statistical and systematic errors, and it is from the combination of these that confidence levels must be derived. Systematic errors include those due to model dependence, but also other factors such as detector acceptance, calibration, detector response functions, luminosity fluctuations, etc. In most cases, the systematic error is of comparable or greater size than the statistical one. In such a scenario, improvements to the statistical error will count, at best, in quadrature.

Detection and estimation theory require, known, accurate models of the underlying processes. This is unfortunately not the case in final physics analyses. In data driven approaches such as machine learning, classifiers or estimators obtained will be specific to the model assumptions used in creating the training data. It will then be necessary to create a family of classifiers/estimators using different guesses for the model parameters, resulting in a complicated system from which it will be difficult to untangle systematic and statistical errors. Physicists have traditionally preferred much simpler approaches for extracting confidence levels, which allow a straightforward interpretation of sources of error. For these reasons, SSP will likely not be of significant use here.

4.3. SSP for low level pattern recognition

Among the tasks classified as low level pattern recogntion, LPR, are track reconstruction from raw tracking chamber hits; vertex reconstruction; particle identification; region of interest tagging; and event topology classification. The key difference, from an SSP standpoint, between low level pattern recognition and physics analysis that the former is far less hampered by model dependence.

This is not to say that models are not present in LPR; rather, the models used here are more well defined, for two principal reasons :

- The trajectories of charged particles in known magnetic fields are reliably characterised by well understood processes in electromagnetism and multiple scattering.
- Although modelling of electromagnetic and hadronic shower formation, Cherenkov ring production, and the like, in complicated detector geometries have become very sophisticated, physicists' ability to predict the behaviour of such objects is not as reliable as in the case of track and vertex reconstruction. Nonetheless, it *is* possible to pin down the necessary model parameters by using <u>test beam data</u>. This is in contrast to physics analyses, in which in such objects as top quarks or Higgs particles are extremely rare and cannot be reliably simulated. Individual electrons, muons, photons, and hadrons, however, *can* be produced in copious quantities at test beams, and their interactions with detector elements used to produce accurate models of detector response.

For all these reasons, SSP is far more likely to be applicable in LPR than in physics analysis, and, indeed, it is in such areas that the first instances of the use of SSP in HEP have appeared. In the following, we shall outline a few of these applications, distinguishing between those occurring in the trigger and those offline.

4.4. SSP in the trigger

Once again neural networks can serve as a guide. For example, the H1 experiment at HERA, an electron-proton collider, has used hardware neural networks in their level-2 trigger for over 7 years [5, 6]. The trigger examines data from the calorimeter and track chambers which is available after level-1 and uses it to classify passed events into 12 physics categories according to the event topology – number and coordinates of found leptons, jet candidates, missing energy, vertex position, etc. An SSP-based detection theory approach would likely also be interesting to try here. The obvious caveat is that the neural network approach is only feasible because it is able to exploit fast execution in parallel hardware; any SSP applications in the trigger will also need to be able to meet the stringent trigger timing constraints.

One of the first applications of SSP in HEP, in fact, was also proposed for a trigger system, this time for a muon trigger for the ATLAS tilecal calorimeter [7]. Pulses produced in the calorimeter modules to the passage of charged particles are samples at 40 MHz. The system applies an optimal filter to 16 consecutive samples in order to distinguish muons from other types of interactions.

4.5. SSP in reconstruction

The above matched filter approach based on pulse shape was first seen in conjunction with cold liquid calorimetry, which has a relatively slow readout [8]. Recently, a rather sophisticated approach along similar lines has been presented for the offline analysis of the ATLAS tilecal calorimeters, and is destined to improve the timing an energy resolution of this device [9].

Optimal filtering has also taken hold in HEP within the framework of track reconstruction, and a number of applications have appeared recently. The use of Gaussian sum filters, which are optimal linear filters, in this field was pioneered by R. Frühwirth in 1997 [10]. More recently the technique has been extended to vertex fitting [11], and deterministic annealing filters for track finding [12] have also been proposed. These approaches to track and vertex reconstruction in the presence of non-gaussian noise appear to be among the most promising avenues for future developments of SSP within HEP at present.

5. SUMMARY AND CONCLUSIONS

The fast-pulse nature of HEP signals has, for technological reasons, delayed the penetration of SSP techniques into the field. The scenario is changing as faster ADC's become available and as high energy physicists begin to educate themselves about the benefits of SSP. The event-based structure of HEP data precludes the use of SSP on a time scale exceeding the duration of an event, and adaptive system SSP approaches are excluded since adaptation is neither necessary nor easily realisable in HEP instrumentation.

Inroads for SSP exist both at the trigger level and in the offline analysis. HEP experience with neural networks, at both trigger and offline levels, can help guide the insertion of SSP techniques, which also include data-driven learning approaches. In trigger applications, care must be taken that chosen algorithms can be mapped onto parallel hardware, in order to meet the stringent timing requirements of level-1 or level-2 triggers. Offline, the pre-eminence of model dependence and systematic error concerns render SSP less applicable for final physics analyses leading to confidence level estimates; however for low level pattern recognition, a great number of applications of SSP seem possible, and some are already beginning to establish themselves, notably the use of optimal filtering in calorimetry and track reconstruction.

6. REFERENCES

[1] Online document, The Aleph Experiment, http://aleph.web.cern.ch/aleph/alpub/draft/AlephHistory.pdf, Claus Grupen, Ian Hughes, Jim Lynch, Ron Settles, Editors.

[2] B. Denby, "Neural Networks and Cellular Automata in Experimental High Energy Physics," *Computer Physics Communications*, **49**, p. 429, 1988.

[3] P. Abreu et al., DELPHI Collaboration, "Classification of the Hadronic Decays of the Z0 into b and c Quark Pairs using a Neural Network," *Physics Letters* **B295**, pp. 383-395, 1992.

[4] S. Abachi et al., "Direct Measurement of the Top Quark Mass," *Phys. Rev. Lett.*, **79**, pp. 1197–1202, 1997.

[5] J.H. Köhne et al., "Realization of a second level neural network trigger for the H1 experiment at HERA," *Nuclear Instruments and Methods in Physics Research,* **A 389**, p. 128, 1997.

[6] B. Denby, C. Kiesling, J.-C. Prévotet, P. Garda, B. Granado, A. Wassatch, "Fast Triggering in High Energy Physics Experiments Using Hardware Neural Networks," *IEEE Transactions on Neural Networks*, **Vol. 14**, pp. 1010 – 1027, 2003.

[7] R. R. Ramos, J. M. de Seixas and A. S. Cerqueira, "A Matched Filter System for Muon Detection with Tilecal," *Nuclear Instruments and Methods in Physics Research*, **534**, p. 445, 2004.

[8] W. E. Cleland, "Signal processing considerations for liquid ionization clorimeters in a high rate environment," *Nuclear Instruments and Methods in Physics Research,* **A338**, pp. 467-497, 1994.

[9] Cristóbal Cuenca Almenar, "Implementation of an Optimal Filter for Tilecal Energy Resonstruction in the ATHENA Framework," Masters thesis, University of Valencia, Spain, 2004.

[10] R. Frühwirth, "Track fitting with non-Gaussian noise," *Computer Physics Communications,* 100, p. 1, 1997.

[11] R. Frühwirth, T. Speer, "A Gaussian-sum Filter for Vertex Reconstruction," *Nuclear Instruments and Methods in Physics Research,* **A534**, pp. 217-221, 2004.

[12] A. Strandlie, R; Frühwirth, "Adaptive filters for track finding," *Proceedings of Advanced Computing and Analysis Techniques*, Berlin, Germany, 23-27 May, 2005, to appear in the proceedings in *Nuclear Instruments and Methods in Physics Research.*