CERN-Data Handling Division
DD/83/15
M. Metcalf
September 1983

CERN LIBRARIES, GENEVA

CM-P00059842

THE ROLE OF COMPUTING IN HIGH ENERGY PHYSICS

(Submitted to IBM Perspectives in Computing)
The Role of Computing in High-Energy Physics

1. INTRODUCTION

Since computers were first introduced into the field of High-Energy Physics (HEP) in the late 1950's, the scale of their use has expanded dramatically, both quantitatively and qualitatively. This expansion has occurred in step with a similar increase in the scale and complexity of HEP experiments, and both have been made possible by the rapid advances in electronics. The same fact that electronic circuits have become smaller, faster and less power consuming, allowing the construction of both tiny computers of modest power and super-computers of immense potential, has also allowed the construction and operation of HEP equipment to develop at a phenomenal rate.

In this article we shall examine some of the more important aspects of the present and future applications of computers in HEP, but to begin with a short introduction to the research aims of HEP is called for.

2. PARTICLE PHYSICS

Since the earliest times, man has attempted to comprehend both the nature of matter, which the Greeks already decomposed into four elements, and the forces which are manifested by such diverse everyday phenomena as the motion of celestial bodies and the production of lightning. HEP is the present-day study of matter and of the forces of nature. It is perhaps better termed particle physics, as the study is effectively one of the nature of the constituent particles of matter and of the forces between them; the term high-energy derives from the incidental fact that the pursuit of this research requires the use of particle accelerators capable of imparting large amounts of energy to beams of particles.

It is now widely believed that the constituents of ordinary matter fall into two classes, hadrons and leptons. The hadrons are particles of finite size, and are composed of pairs or triplets of a more basic particle — the quark. The quarks are supposed to be bound by a force which does not permit them to be separated long enough for a free quark to be observable. Quarks carry a fractional charge of ±1/3 or ±2/3 of the charge of the electron, and have ascribed to them other attributes such as "flavour", with names like up and down, and "colour". This latter property leads to the name of the theory describing their properties and interactions - Quantum Chromodynamics (QCD). An example of the way in which hadrons are composed of quarks is given by the proton, which consists of two up quarks, each carrying +2/3 charge units, and one down quark carrying -1/3 charge units, giving a net charge of +1. Hadrons are bound to one another, for instance as protons and neutrons in the atomic nucleus, by the so-called strong force.

The second class of particles contains the leptons. These are apparently point-like in nature, and are subject to the so-called weak force. The particles in this class are the familiar electron and its heavier relations the mu and tau leptons, together with the massless, neutral particles known as neutrinos. Neutrinos permeate the whole universe, and if they possessed even a
tiny mass, their combined mass would possibly be sufficient to tip the balance between a universe which expands for ever and one which finally collapses again. For this reason, the determination of the neutrino mass, if any, is currently an active area of research.

There are now considered to be four forces in nature. They are the strong force which binds nuclei together, and which is a manifestation of the colour forces between the constituent quarks of the hadrons; the electromagnetic force, which is one hundred times weaker and which governs such familiar phenomena as electric motors and radio; the weak force, ten million times weaker still and governing certain types of radio-activity; and finally the feeble force of gravity, binding the solar system and the galaxies.

The forces are mediated by the exchange of particles, that is, the force between a pair of particles mentioned above is transmitted by a third particle passing between them. In the case of the electromagnetic force this is by the exchange of a photon, and in the case of the weak by the exchange of the W and Z particles. These latter particles are postulated by the theory of Weinberg and Salam which attempts to unify the weak and the electromagnetic forces, and the recent discovery of the W and Z at CERN is a brilliant confirmation of the theory, and comparable to the discovery by Hertz of the electromagnetic waves which confirmed Clerk Maxwell's unification of the electric and magnetic forces. Theoreticians are engaged in further attempts to unify all the forces into a single framework known as Grand Unified Theory, but are still far from their goal.

3. **HEP LABORATORIES**

The resources required to construct and operate a HEP laboratory are so considerable that the research has inevitably concentrated in a few centres - notably at Fermilab, SLAC, Cornell and Brookhaven in the USA, CERN and DESY in Western Europe, Dubna and Serpukhov in the Soviet Union, and at KEK in Japan. CERN is an institute situated near Geneva, Switzerland and funded by 13 European nations, bearing jointly the large cost of this aspect of "big science". Its annual budget is about $300M, but although the research aims are in the realm of pure rather than applied physics, the technological demands it places on industry result in a substantial spin-off from the investment. It has a staff of about 3300, plus another 1000 visiting scientists who come to use the accelerators and beams, the experimental facilities and the other services such as computing which CERN provides.

The experiments are conducted by collaborations of between 10 and 100 physicists from various institutes and universities around the world, and each such experiment undergoes an initial selection and approval procedure in which scientists from associated institutes participate.
4. ACCELERATORS

The basic instrument in HEP research is the accelerator, into which a beam of particles is injected at low energy, and which by the application of radio-frequency fields progressively accelerates the particles to speeds approaching that of light. The beams are focused and maintained in a roughly circular orbit by strong magnetic fields. The largest accelerators are several kilometres in diameter, and accelerate protons to energies of hundreds of GeV (giga-electron-volts; 1 GeV is equivalent to the energy acquired by an electron accelerated by an electric potential of $10^9$ volts). An example of such a machine is the Super Proton Synchrotron (SPS) at CERN, which can accelerate protons to 450 GeV, and extract them to impinge upon fixed targets. The resulting interactions produce other particles such as muons or neutrinos, which are used as beams by experiments. The protons can, of course, be used directly.

A second type of accelerator is the colliding beam machine, in which two beams are accelerated in opposite senses and allowed to collide with each other at certain interaction regions. These two beams often consist of a beam of particles circulating in one sense and of the corresponding anti-particles in the opposite sense. The resulting head-on collisions of particles lead to their mutual annihilation, and the decay products may be investigated. This type of machine has the property that there is no effective loss of beam energy due to kinematic considerations, as is the case in fixed target operation. The SPS can operate in a mode such that protons and anti-protons are accelerated in opposite directions simultaneously, and made to collide in two underground interaction regions, which are equipped with large detectors. One reaction which can occur is the annihilation of an up quark in a proton with an anti-down quark in an anti-proton, leading to the formation of a $W$ boson accompanied by the debris from the other quarks. The decay of the $W$ into an electron and a neutrino provides the detectable signature of the particle, which has been exploited in its recent discovery.

A variant of this type of accelerator collides counter-rotating beams of electrons and positrons. In order to prevent unacceptable losses of beam energy due to synchrotron radiation in the bending magnets, only weak magnetic fields can be used, and thus these machines have to have a relatively large diameter compared to proton accelerators. CERN is currently constructing an underground Large Electron Positron (LEP) accelerator which has a circumference of 27 km for an initial beam energy of 50 GeV.

Computers find widespread applications in accelerator design and operation. Computer aided design (CAD) is being used increasingly in the design of their components, and data base management systems are required to keep track of the huge number of electrical connections, and to help with other management services such as PERT. The extensive calculations required for the magnetic fields and the beam optics are usually performed using a few well-tried programs in batch mode; attempts to provide interactive systems for such problems have been largely unsuccessful because the high number of parameters involved makes it difficult for a human to maintain an overview of the problem in hand.

Accelerators are typically controlled through a network of computers linked to a central control room. The computers monitor the beam and the accelerator components, and can make automatic adjustments as required. The operators can inspect the state of the machine and intervene using consoles connected to the network through central control computers; the SPS system is programmed in an
interpretative special-purpose language, NODAL. The extracted beams are also monitored by computer, and the relevant information is transmitted on links to the experiments receiving them.

5. DETECTORS

The detectors used to record the passage of particles resulting from particle interactions are nowadays principally electronic detectors. These consist of a number of different parts, typically arrays of wire chambers and scintillators to record the position at which each particle traverses the component, and to measure also the energy of the particles. The arrays are sometimes such as to be able to record the passage of up to 100 separate particles, and may be used also to measure the total energy of a particle by causing it to interact with interspersed layers of target material - a so-called calorimeter.

Newer detectors are equipped with micro-processors able to process data from each of the components, and major detector parts are under the control of small mini-computers. It has been found to be desirable to have a single high-level programming language to use on these devices for testing purposes, and to this end an interpretative language PILS has been developed, giving engineers a PASCAL style tool which they can use in a simple fashion over a range of different machines. Cross-software for micro-processors is available on the central computers.

The connections between the electronics and the computers are assured by standard CAMAC interfaces - in the future the new FASTBUS standard is likely to be increasingly used.

When a physics experiment is carried out, the detector is placed under the control of one or more operation and data acquisition computers. These receive a signal from the accelerator that beams are about to pass, and activate the detector to record any event which results. As soon as signals are detected, an electronic system locks them into buffers, and the detector becomes dead, unable to record any further events until the buffers are cleared. In order to reduce the dead-time to a minimum it is vital that a fast decision be taken as to whether the contents of the buffers represent an interesting event or simply some background process, such as the passage of a cosmic ray or an interaction between the beam and the residual gas in the vacuum tube in which it circulates. The large quantities of data involved imply a multi-level trigger system, illustrated here by a scheme proposed for a LEP experiment. A fast decision (μsecs) is taken by the electronics as to whether the buffers contain background or a possible event on the basis of certain critical signals. This can reduce the primary event rate from 50 kHz to 1 kHz. A second-level trigger using hard-wired decision logic can decide in 10′s of μsec whether more complex pattern criteria are fulfilled, reducing the rate further to 20 Hz. Now the rate is low enough for special micro-processor based devices to decide in 10′s of msec whether certain criteria are fulfilled in different detector parts, operating on them in parallel, and reducing the rate to 5 Hz. This permits a more complex analysis of the signals to be carried out on the remaining events, perhaps on the on-line computer or attached emulators, giving a final rate of 2 Hz. Only now is it possible to contemplate writing the buffer contents to magnetic tape since, for these large detectors, event sizes of 250 kbytes must be envisaged, and the 2 Hz rate for such events is just manageable at current tape writing speeds.

- 4 -
The emulator just mentioned is a relatively new development, and consists of special hardware which reproduces the functionality of a commercial mainframe such as the IBM System/370, accepting pre-compiled or pre-linked FORTRAN programs for execution at much reduced cost. This permits programs to be developed in the usual way on a mainframe, and to run in real-time at an experiment, providing the on-line mini-computer with the auxiliary power of a large system. Emulators are also finding applications in relieving overloaded mainframes of some of their work, as arrays of them can be constructed for batch operation.

6. NETWORKS

The future LEP collaborations will consist of up to 300 physicists from 30 institutes, mainly in Europe but also from as far afield as Beijing. Since many of these physicists spend large fractions of their time at their home institutes, but nevertheless wish to be able to contribute to their experiment while absent from CERN, the implementation of a sophisticated data communication network connecting the computer systems at the various centres will be of vital importance. This is much more difficult to carry through in Europe than in the USA, owing to the problems stemming from the large number of national telephone administrations involved.

We can distinguish three levels of networking to which a LEP experiment might be connected. Firstly, the equipment computers themselves will be interconnected via a local area network (LAN) of about 10 Mbit/sec raw capacity, to which it will be possible also to connect terminal concentrators, emulators, personal work stations, operations computers and graphic stations. This will permit rapid access to common file bases. The whole will be linked through a gateway to the second level, a CERN-wide backbone network which for LEP will have to have a point-to-point capacity of up to 50 Mbit/sec. This will connect LANs and main computers over the whole site, allow remote job submission and retrieval, access to large common file bases and also permit back-up disc storage onto the central IBM 3850 Mass Storage System, which may possibly act as the store for a file server for the whole of the site.

The third level is a European wide network, with raw capacity on the land-based lines of up to 48 kbits/sec nationally or 9.6 kbits/sec internationally. This will, as already implied, connect institutes and allow the terminal access and transfer of small files containing programs, messages, small data sets and submitted or retrieved jobs. The use of a satellite giving potentially Mbit/sec rates for such links has been tested, but for bulk data transfer the use of lorries loaded with tape reels still remains more convenient. Satellite links may be useful for connecting computer centres in different countries.

One of the major difficulties associated with preparing software for LEP will be the distributed nature of the task, and networks will be vital in this area.
7. **OFF-LINE ANALYSIS**

Batch processing constitutes the principal load placed upon the CERN computers. The installed capacity has progressed from an IBM 709 in 1959 to a present day mixed configuration from three vendors, equivalent to about 11 IBM 168s, and growing at an annual rate of 22 percent. Computer budgets are allocated to the individual experiments according to their needs and the available resources, but in general only one-third of the data generated at CERN is analysed there, and particularly in the later phases of an experiment it is expected that the analysis be performed mainly at the participating institutes.

At any one time there may be up to 70 active experiments at CERN, and each will have at least one analysis program, which for the larger experiments will consist of 100,000 lines of FORTRAN. In addition there will be an associated collection of utility and smaller analysis programs, and the overall scheme in which they are used can be described by outlining the analysis chain of one medium size experiment.

As a first step the 1600 bpi tapes containing the raw data from the experiment are copied onto 6250 bpi tapes, in order to reduce the subsequent tape handling. Each high-density tape is then read by a 35000 line program running in one of its several modes, namely the one able to calibrate the data. In this step only special calibration events are selected and processed, in order to be able to calculate a large set of constants which describe the response of the detector components to standard particles, in this case cosmic rays. The constants are written to a permanent disc file containing the constants from a whole running period of several months. The tape can now be rewound and reprocessed in a second mode in which the physics events are read, converted from channel counts into energy units using the calibration constants, and checked as to whether the event is certainly a background event rather than a likely true physics event. Events which pass this off-line filter are written in their converted form to a second tape, but still in a format identical to the original data, for possible reprocessing through the program. Various beam monitor and other data are written with the events, but the elimination of the calibration and background events result in some data reduction.

In a third stage, the program is run in its full processing mode, in which particle trajectories are reconstructed, the physical parameters of the particles determined (space angle, vertex and momentum), and the calorimeter data evaluated to determine the characteristics of the shower of particles it contains. Finally, the whole event is classified and written to tape, still using the raw data format into which new data calculated by the program are inserted. At this stage a data expansion takes place, but these data summary tapes (DSTs) now contain all the original and reconstructed information about each real event, together with any associated monitor information. In order to ease the task of physicists who wish to analyse these data further, mini- and micro-DSTs are written, containing progressively more condensed data, but allowing physics analysis programs to run quickly to produce the histograms and fitted quantities which will be the final published result of the experiment.

As a last stage in the use of computers in data analysis, we now see the introduction of documentation and photo-composition facilities allowing the preparation of camera-ready copy for publication.
8. SOFTWARE GUIDELINES

The size of the off-line programs and the fact that they should run for 100’s or 1000’s of hours on a range of computers in many countries, and be used by many physicists, impose certain conditions on them: they should be portable, as easy to understand as possible, easy to maintain and efficient. As a consequence various standards and conventions are followed, especially in the use of FORTRAN and in the insistence that hardly any assembler code be used. Frequently, dynamic memory management systems are necessary to supplement the data structures of FORTRAN. Maintenance tools are required for the generation of program versions for different computers, to enable many users to develop a single program, and to enable code to be transported readily from machine to machine. Here, a CERN produced product, PATCHY, has found applications also outside the HEP community. Portability is also helped by the installation and availability at all European HEP computing centres of CERNLIB, a structured program library providing many facilities required by off-line programs. At the same time, the NAG library is coming into use in HEP programs.

The requirement that code be efficient, once the algorithms have been developed, means that profiling tools are applied to the codes and a major study on FORTRAN optimization has recently been published.

The need to plan the design and implementation of these large off-line systems in a manner which will allow their long-term use and development has led to an increasing interest in the software tools and methods used in other branches of science, and a recent workshop, organized to discuss just this problem, has given a new impulse to attempts to transform HEP off-line computing into HEP software engineering.

9. EVENT SIZES AND VOLUMES

We have seen how the size of LEP detectors containing up to 250,000 electronics channels will result, in spite of some data compression in the hardware, in final event sizes of up to 250Kbytes. The rates after triggering will correspond to writing a tape at full speed and, therefore, imply mounting a new tape every few minutes. The number of events involved may be as high as 10 million! Tape handling, already a major headache, threatens to become a nightmare. The CERN tape vault contains 165,000 active reels, and several times that number are archived elsewhere. The operators are assisted by a computer based book-keeping and location system, but the risk of being overwhelmed by the volume of data is a real one, and a constant watch is kept on new recording devices, in the hope that a more manageable medium will be developed to replace magnetic tape.
10. GRAPHICAL APPLICATIONS

The importance of graphics is growing at all levels of HEP computing. On-line graphics devices are used for the display of equipment status and beam conditions, enabling physicists to inspect sample events and histograms of the various parameters of the detector. This allows them to take remedial action when necessary.

Off-line applications are becoming ever more sophisticated. In the course of planning and analysing an experiment, graphics will help with the design of the detector, by allowing physicists to follow the development of simulated events inside it. By displaying the details of reconstructed real or simulated events, physicist-programmers can check the quality of the reconstruction program, and make improvements to the algorithms. For the study of rarer processes, where it is important that every event is correctly analysed and classified, powerful systems have been developed allowing physicists to check the fine details of the analysis using high-resolution vector devices which can rotate the viewed event in space, or enlarge specified regions, or wobble the image to give the impression of a third dimension. Using a pointing device, the man-machine system permits a complete evaluation of the displayed event, and a possible recuperation of a wrongly analysed section of it. Further portable packages in widespread use, such as HTV, allow semi-interactive plotting of histograms and functions, an essential part of the final physics analysis of most experiments.

Whereas until now, graphics packages have been developed in-house for specific applications, the emphasis is presently shifting towards standard products. The new international graphics standard GKS has been adopted for future work, in order to create an environment in which graphics based programs become more portable, but steps are having to be taken to overcome its limitation to 2D-graphics.

11. EVENT SIMULATION

The design and analysis of an experiment involve large-scale simulations. At the design phase, this is principally to examine the way in which changes in the detector parameters affect its ability to measure the various event parameters in a sufficiently accurate way. During the analysis phase, as many simulated as real events are generated and processed through the whole program chain, in order to be able to detect differences between the actual and the expected results, and to be able to calculate the so-called acceptance of the detector, that is the differences in its ability to detect and measure events of differing energies and topologies.

Simulation consists essentially of two steps. In the first, Monte-Carlo techniques are combined with physics theory to generate the final states of particle interactions, that is the parameters of the detectable particles. The particle spectra generated by such programs should correspond to those of known data, and allow an extrapolation into higher energy regions for the study of the likely behaviour of new accelerators and detectors.

In the second step, the particles are followed one by one through an actual detector, taking into account all known physical effects. For instance, the
program will deflect a charged particle as it passes through a simulated magnetic field, and subject it to multiple scattering and energy-loss as it passes through matter and, where appropriate, will cause the particle to decay or interact in a random way corresponding to known or supposed physical laws. The final output of such a program is a file containing the events as seen by the detector in a format identical to that used for real data, so that the analysis program operates in an identical fashion for both types of data, thereby eliminating any potential bias.

This second step can be extremely time-consuming for complex events in large detectors, especially if showers of particles are generated inside calorimeters. Whereas it is a relatively fast operation to follow, for example, a single track through a simple array of plane measuring chambers in a magnetic field, the huge number of secondary particles generated in a shower, each one of which has to be followed until it decays, interacts or has no further energy, leads to a processing time of 10's of seconds of IBM 168 equivalent for typical LEP events.

12. THEORETICAL STUDIES

The bulk of the cost of HEP is in the experimental side, and research proceeds by experiments being proposed to test theories, and by the evaluation of experimental results in terms of existing theories, or where they are found to be wanting, of new ones. Traditionally, theorists themselves have made relatively light demands on computers, using them either for numerical calculations, or for algebraic symbolic manipulation using languages such as REDUCE, or for interactive display and manipulation of functions.

In recent years, however, new approaches have resulted in the development of a Monte-Carlo technique which, working from some basic principles, allows the determination of such fundamental properties of the hadrons as their mass. This technique, deriving from lattice gauge theory, involves four-dimensional lattices of typically 250,000 points, at each of which complex arithmetic operations are performed, in order to arrive by successive iterations at a stable estimate of the required property. This has placed a sudden and very large demand for computing power on the computers available to the HEP community, estimated at 10's of CRAY-1 equivalents, and experience shows that the problem is one which is, in fact, best tackled on large vector processors, machines which until now have found little application in HEP computing.

13. TRENDS

Foretelling the future is notoriously difficult in computing, but in the current period of preparation for a new generation of accelerators and experiments, the opportunity has been taken to take stock and to plan for likely future needs. Here only a few of the broader developments are considered.

In the area of hardware we shall see a continued development of the “supercomputers” (vector and array processors), as more manufacturers enter the field with ever more powerful machines, perhaps attaining 10 Gflops (10¹⁰ floating-point operations per second) by the end of the decade. The nature of
most HEP codes is such that they are not readily vectorizable, and the emphasis has always been on the use of the largest available serial machines such as the IBM 3081K and CDC CYBER 875. The need to produce portable code running at many institutes only reinforces the existing emphasis, and vector processors will find immediate application only in special areas such as lattice gauge calculations.

At the same time, smaller machines are becoming more powerful, and the user-friendliness of their operating systems makes them increasingly attractive to physicists, who often prefer to carry out program development on a small computer belonging to their own group, rather than battling with a huge system running on a mainframe. Here we may see a move away from mainframes for development, with their use being reserved for batch number-crunching. The introduction of personal work stations attached to LANs can only accelerate that trend.

The development of new peripheral devices is kept under constant review, and new terminals, display devices and storage systems are introduced as they become available. In this area, the largest problem is that of bulk data storage, presently magnetic tape based, and the development of laser and video discs and of even higher density tapes is being awaited eagerly.

In the area of software, user-friendly operating systems have already been touched on in the context of small computers. The other main trends in software are in the area of tools and languages. Programmers will increasingly expect to be able to work in an environment tailored to their own requirements, using tools which enable them to design, construct, test and maintain their software in a coherent fashion.

14. CONCLUSIONS

As we have seen, computers in their various guises now completely permeate every area of high-energy physics research, which relies upon them totally. They do not simply relieve scientists and engineers of the need to perform mundane and repetitive tasks, but actually make possible the design, execution and analysis of experiments which would otherwise be impossible. In this particular field of human endeavour, the computer occupies a position which is not just a peripheral one of being a powerful tool, but one which is fundamental to progress in our understanding of the mysteries of nature.

15. SUGGESTIONS FOR FURTHER READING

15.1 High-energy physics


H. Georgi, A Unified Theory of Elementary Particles and Forces, Scientific American, April, 1981.

15.2 CERN


15.3 Accelerators


15.4 Special processors

Three day in-depth review on the impact of specialized processors in elementary particle physics, INFN, Padua, Italy, 1983.

15.5 Networking

Networks for High-energy Physics, ECFA/82/60, CERN, Geneva, Switzerland, 1982.

15.6 Off-line Analysis


15.7 Software Guidelines

Workshop on Software in High-energy Physics, CERN/82/12, CERN, Geneva, Switzerland, 1982.


15.8 **Graphics**


15.9 **Theoretical Studies**


15.10 **Trends**