Computing on the Eve of LEP Data Taking: Are we ready ?

(D. O. Williams)

CERN LIBRARIES, GENEVA

CM-P00059921

Presented at the "Computing in High Energy Physics" Conference
Oxford - April 1989

COMPUTING ON THE EVE OF LEP DATA TAKING: ARE WE READY?

David O. Williams
CERN, Data Handling Division
CH-1211 Geneva 23
Switzerland

1. LEP Physics

The primary objective of LEP is to generate Z^0 particles in sufficient quantity to make a thorough study of their properties. Accordingly we expect to run LEP with the electron and positron beams at an energy close to 46 GeV/c, to give a collision energy at the Z^0 peak.

A first objective will be to measure the width of the Z^0 peak. The "standard model" makes a rather precise prediction that this should be about 2.6 MeV/c. If measurements show that it is wider than predicted, it would be a clear sign of "new physics" in the air, and people would become very excited indeed.

The next step is likely to be a rather precise measurement of the mass of the Z^0 . Then we can expect the thorough study of the Z^0 decay modes to start, perhaps with special emphasis on searches for the top quark, if it has not already been found elsewhere, and for the Higgs boson, a particle which would determine the mass scale of other elementary particles.

2. LEP Machine

At the time of writing all the signs are that the LEP ring will be closed on schedule on 14 July 1989. This day has a certain symbolic significance, being the 200th anniversary of the French revolution.

The team responsible for commissioning the machine has set up an "official" planning on the basis of days when they have access to the machine with the complete injector chain (LIL, EPA, PS and SPS) being fully operational, and with no major breakdowns. Such days have been referred to as "golden" or "CPU" days! On this planning there should be a pilot physics run for three days, 21 days after LEP start-up, but without the superconducting low β quadrupoles being activated. Physics data-taking should be possible after 35 days, and the plan is to make a serious physics run where smooth operation takes precedence over machine running-in during November and December 1989.

It has recently become clear that the lifetime of the circulating beams may be considerably better than originally expected, with a rather slow decay of the luminosity. The time between refills might be extended to almost 24 hours.

Accurate predictions of the luminosity that will be available in the early months of LEP running are, of course, difficult to make. The design target is for an average luminosity of $1.6 \times 10^{31} \text{ cm}^{-2} \text{ s}^{-1}$.

3. The Experiments

The status of the experiments was presented at the LEP Experiments Committee in mid-March.

ALEPH expect to have the detector completed in the beam position, cabled, and connected to the gas supplies by mid-May. This should allow for about two months of testing with cosmic particles before LEP start-up.

DELPHI are working hard to recover from a serious delay caused by a cold leak in their superconducting coil which developed last autumn. The initial impact was a delay of about 4.5 months, but the installation schedule was quickly rearranged, and decisions on priorities were made. The delay at present is estimated at 6-8 weeks. It is clear that DELPHI will be ready to take data when LEP starts, though not all of the sub-detectors will be fully operational.

L3 are also making good progress, with the Time Expansion Chamber currently undergoing tests before being installed towards the end of May. Unlike the other experiments, which can be moved out of the beam into a parking position, the position of L3 is fixed.

OPAL have almost all components ready and installed at Pit 6. The late delivery of that pit meant that the schedule was extremely tight, with only 12 months available for the complete installation. The planning only allows for a few days of testing with cosmics before 14 July.

The moment of truth for the **trigger electronics** and **data acquisition systems** is almost upon us. All of the experiments are now testing several sub-detectors with the "final" DAQ software.

At the LEP Experiments Committee the following worries were mentioned:

- The counting room infrastructure (racks, cooling, cabling) was rather late for several experiments.
- The Pandora boxes which provide the clocking for the whole DELPHI trigger and DAQ system are only scheduled to arrive at CERN in June.
- OPAL were waiting for software to drive the optical fibre link that brings the data from the bottom to the top of the pit.
- OPAL and ALEPH were both sometimes unable to test sub-detectors because of missing fire detection and gas safety systems.

In conclusion we can say that the installation of all of the experiments is proceeding at the "normal" hectic pace. Barring last minute hitches they will all be in the beam at LEP start-up, ready to record some data.

4. First Comment

It is rather amazing that the computing for LEP appears to be in such good shape. But much work remains ahead of us. The data volume that we are faced with is one to two orders of magnitude more than we have ever faced before. Some of the interesting physics may come from exceedingly rare decay channels. We are dealing with high complexity, high statistics physics, and it will not necessarily be the first experiment to write a tape of data that finds the most interesting physics.

5. Data Volumes

Much of the analysis of the size of the data volume that we can expect at LEP has been published in the MUSCLE Report - the Computing Needs of the LEP Experiments [1]. It is assumed that each experiment aims to accumulate about 10 million events after a few years of data taking. Using the hypotheses of [1] these 10 million events would generate:

- 4 Terabytes of raw data
- 2 Terabytes of simulated raw data
- 1-4 Terabytes of summary data.

For the summary data the smaller size (1 Terabyte) corresponds to 20 Kbytes of data per event, which many people now consider to be an unrealistically low target. Expressing the data volumes in terms of full 3480 cartridges of the present generation, which hold 200 Megabytes each, we obtain:

- 20,000 cartridges of raw data
- 10,000 cartridges of simulated raw data
- 5-20,000 cartridges of summary data

for each experiment. Note that these numbers make no explicit allowance for the copies of the data that will undoubtedly be needed.

6. Data Recording

While OPAL mainly uses **VMEbus**, the other three LEP experiments have based their read-out on **FASTBUS**. All of these systems are large by traditional standards, with up to 200 crates.

It is encouraging to see that the necessary hardware and software is coming together on time, thanks to the hard work of many people working in the experiments and in the support teams. There have been moments when it was not obvious that the struggle would be successful!

The hardware developments have included the ALEPH Event Builder (AEB) and the General Purpose Master (GPM) processors, and a set of FASTBUS interfaces culminating in the CERN Host

Interface (CHI) which is now available with an optical fibre extension running over more than 1 Km. It is encouraging to see that all of this work has been pursued in close and successful collaboration with European industry.

The software needed both for the general programming of FASTBUS, and, more specifically to initialise complex systems from a database description is available and well reviewed in [2].

The **Model** data acquisition system was designed in response to a considerable technical challenge - namely to try to provide a set of software components that could be assembled by the experiments to their own specifications, rather than to present the traditional monolithic "take it or leave it" approach to DAQ systems. The collaborations that use Model have to complement the Model packages that they select with software of their own, and they are responsible for the overall system integration.

It is encouraging to report that the technical challenge has been met successfully. Model is now rather well tested in DELPHI, L3 and OPAL, and also in many other experiments both inside and outside CERN. ALEPH have also successfully developed their own DAQ system.

In line with the recommendations of the MUSCLE report the main efforts on **recording media** at CERN have been directed towards 3480-style magnetic cartridges, and their automated handling in robotic loaders. Compared to traditional "round " tape technology these devices show clear advantages in terms of lower media cost, lower unit cost, and higher reliability. However it must be admitted that the expected increases in capacity to 400-600 Megabytes per cartridge have not appeared as quickly as many people had hoped.

This has lead to a significant pressure for the use of videotape devices such as Exabyte. These are normally defended as being especially suited for the transfer of summary data, where their large capacity (up to 2.5 Gigabytes) outweighs their low transfer speed (nominally 200 Kilobytes/second compared to 3 or even 4.5 Megabytes/second for cartridges). I am worried that some people want to push these devices, designed for the home electronics market, beyond their limits.

7. Data Transport

All of the LEP experiments make extensive use of **Ethernet**. At the pit it is used to link together workstations, terminals, intelligent microprocessors, the computers controlling the sub-detectors and the main DAQ computers. It also provides the general-purpose CERN-wide network. At least 27 Ethernet segments are linked together, using a variety of techniques, to cover distances of up to 15 Km. and to connect over 1,000 devices.

The traffic is carried over an infrastructure of G.703 time division multiplexer lines (mainly optical fibres). In addition to the general-purpose Ethernet these lines carry the Token Ring traffic for the LEP machine control, digital telephoning and TV signals for the access control system.

In addition to the general-purpose Ethernet connection to each pit, one dedicated connection will also be available, running over a separate G.703 channel. ALEPH, for example, will use this to transmit summary data generated by their Falcon system at Pit 4 to the computer centre for further processing. While all of the experiments intend to minimise the size of their summary data, it is clear that there will still be a considerable movement of magnetic media between the pit and the computer centre.

During the past six months we have seen a doubling of the average utilisation on the key backbone Ethernet segment. The huge growth of personal computers and personal workstations really seems to be showing up in increased network traffic. Given the size of the CERN site, and the fact that the user community is rather widely dispersed across it, we can reasonably predict that the bandwidth of the existing Ethernet backbone will saturate within the next 1-2 years.

This is why CERN has been pushing hard for the FDDI 100 Megabits/second fibre optic LAN standard. We have entered a joint study agreement with IBM for the evaluation of some FDDI products. We aim to attach workstations and other products from several vendors to our 3090/600E via the IBM LAN channel gateway, using FDDI standards and TCP/IP protocols. We are also discussing field tests and other forms of collaboration with several vendors. We have already set up an inter-operability lab, based on FDDI products from AMD and AMP.

In a related area we have measured transfer rates of up to 300 kilobytes/second from a Sun-4 workstation into the IBM 3090 across Ethernet, and we are hopeful that industry is starting to deal with this traditional performance bottleneck.

There is an enormous challenge at LEP to move the data processing quickly and effectively between CERN and the outside institutes. The MUSCLE report recommended that 2 Megabits/second links should be installed between CERN and the major regional processing centres in order to facilitate this work. Note that the idea is *not* to move all of the data over these links, merely to use them to keep the programs and calibration data in phase all over Europe, and to move certain key events and results.

INFN Italy, IN2P3 Lyon, CEA Saclay and ETH Zürich are all in the process of trying to obtain such lines, and the first installations might be in place by the summer. Germany also seem to be making progress with its HEP networking, and there are hopes of extending the Belwü 140 Megabytes/second wide-area LAN to CERN.

The whole community of physicists needs to be aware of the extent to which progress might be slowed down by the excessive charges of the European PTTs for leased lines, especially for so-called high-speed traffic (please recall that 2 Megabits/second is only 20% of an Ethernet!). Whether viewed in relation to American charges for equivalent distance, or in relation to back of the envelope calculations of true costs, the rates for these leased lines in Europe seem to be at least five times too

high, and perhaps considerably more! Talk to your political masters about this dangerous situation which risks penalising European science, and presumably European industry as well.

8 . Data processing

While the main worry of physicists with respect to their data processing has traditionally been to obtain enough processor power, we now need to worry at least as much about disk capacity and the problems of accessing the data.

At CERN the computer centre is currently equipped with a CRAY X-MP/48, an IBM 3090-600E with vector features, a Siemens 7890S and a VAXcluster.

The CRAY is operated mainly as a dedicated batch machine, and is obviously especially suited to simulation work. However it also runs very well on a normal "tape" production workload.

The IBM 3090 runs the baseload timesharing and batch service for 500-600 simultaneous users, and the main peripheral power is also attached to it (3480 drives, disks and cartridge loader).

After July 1 the Siemens will run a dedicated VM/HPO production service, which we expect to be of interest to groups with stable production programs.

The VAXcluster principally offers an interactive service to up to 200 simultaneous users, in addition to providing an extensive database (Oracle) and CAE (Euclid) service.

The batch capacity that is available from these machines amounts to some 70 CERN units, of which somewhat less than half can probably be reasonably scheduled for production use by the LEP experiments, so I expect no CPU crisis in 1989.

The situation with disk space is much more delicate. At CERN we currently have 30 Gbytes of staging space, which all users compete for, and about the same amount of long term "maxi disk" storage for all experiments, of which roughly 6 Gbytes are currently allocated to LEP experiments. I am personally convinced that each LEP experiment will need at least 20 Gbytes of dedicated disk space this autumn if the promise of the LEP machine is to be fulfilled.

The CPU capacity available outside CERN, in other European HEP sites, has also made rather encouraging progress in the past year. RAL, IN2P3, CEA Saclay, CINECA Bologna, University of Aachen are only some of the sites used partially or completely for HEP batch that have made substantial upgrades recently - often to powerful models of the IBM 3090 family. Once again it will be the question of peripherals (adequate disk space, sufficient cartridge drives, automated cartridge loaders) and high speed links that determines the effectiveness with which these centres are used for production processing.

At CERN we have entered a joint study agreement with IBM for the provision of an Automated Cartridge Loader. The multi-vendor nature of the CERN centre poses us some special concerns.

- We need to connect to all of our mainframes.
- We need a very high capacity (40K cartridges expandable to 80K).
- We need a high mount rate (160/hour expandable to 320).
- We need a considerable rate of cartridge entry and exit.

We hope to start the installation in May 1989, and to be operational by the autumn.

At CERN a task force led by Jamie Shiers has just completed a report dealing with several aspects of the access to the data that will be required by the experiments. We refer to this work as the FATMEN report, for File and Tape Management, Experimental Needs.

Among the topics addressed are:

- The way in which experiments will keep track of where their data is located - remember that there is at least 1Tbyte of summary data.
- The way in which CERN should keep track of cartridges as they are moved around the site.
- The way in which production data should be moved by the systems software between disk and "tape" storage.

I hope that we will all manage to collaborate successfully in this crucial area over the coming months.

9. Extracting the physics

One of the areas where the LEP experiments have managed to use common base software has been in the simulation of detector behaviour. The GEANT package (Brun *et al.*) is in heavy use in this field.

A big advantage coming from this common approach is that efforts to obtain performance gains from code or algorithmic improvements quickly feed through to the whole community. Efforts to develop a version of GEANT well adapted for vector processors have not yet reached the production stage, despite showing early promise. However, they have recently been picked up again as part of the CERN/IBM joint work on HEP supercomputing.

Of course simulation programs are not the only ones that need to be developed by the experiments. Reconstruction programs, calibration programs and physics analysis code are also needed. At the recent LEPC Delphi reported that they have now written some 480K lines of code, almost all in FORTRAN 77. While there are a few reasons to expect that Delphi may have written the largest amount of code, the other experiments are surely not far behind. Using the canonical 10 lines of

debugged and documented production quality code per day, this represents more than 150 person-years of effort!

It looks as though the LEP experiments will be the first to use workstations in a significant way for their physics analysis. While the promise has been around for several years this is the first time that sufficient numbers of VAXstations and Apollos have been available for widespread and generalised use. To give an idea of the scale, ALEPH announced at the recent LEPC that they expect to have some 60 VAXstations available at CERN when LEP starts up. In this context workstations have some important qualities:

- They offer very cheap CPU power - typically down in the area of 20-50 KSF per CERN unit. Mainframes are at about 400-500 KSF per CERN unit.
- They offer a superb graphics interface - crucial for our understanding of the physics hidden in the data.
- They offer a very superior program development environment.

May of these workstations will be used to run PAW, the Physics Analysis Workstation package developed by Brun and collaborators.

In this context I should also mention the work to support GKS graphics by Myers *et al.* It is easy to underestimate the complexity of graphics systems, both the software and the hardware. The physics community has largely stopped making extensive use of assembler instead of Fortran in order to gain speed when writing large numerical codes, but many people still seem to think that "rolling their own" graphics software is "efficient" and fun. By now more than 75 institutes (and many more licences) are supported via the CERN contract. Even if the users can always hope for more, this represents a major achievement - given the level of effort available - and it constitutes a notable success for a HEP-CCC recommendation!

Something that has been mentioned earlier, but only implicitly, is the concept of cooperative processing. We now have sufficient confidence in the stability and performance of our networks that we can think about using several computers to cooperate in carrying out the job at hand. So when carrying out physics analysis, we might use an Apollo workstation for its graphical power, the CRAY for its processing power, and the IBM to gain access to the reference data stored on disk and cartridge.

We are investing considerably at CERN to make sure that PAW can be used reliably in this way. It is also clear that if the bandwidth on international links is adequate (and reliable) there is no reason why either the workstation clients and/or the CPU or data servers cannot be located remotely. But do not expect public international X.25 circuits to provide adequate bandwidth or quality of service.

10. System Integration

The biggest problem facing the LEP experiments (in the area of computing!) is that, as always at a new machine, they have very little time to carry out real life tests between the moment when they have assembled their detectors completely and the moment when these same detectors have to be used to take data in anger.

The risk is that some points have been overlooked and that the system integration will pose unforeseen problems. There is no easy remedy for this. Effort devoted in the early days to a coherent overall design helps, but is not a panacea. At LEP some experiments (most noticeably ALEPH) have made considerable use of formal design and simulation tools, and I will be interested to see if that effort is repaid over the next few years in greater flexibility.

11. New "Cheap" Processing

It has become clear during the past year that Reduced Instruction Set Computer (RISC) architecture has made the breakthrough to commercial viability. Compiler technology has proved adequate to extract performance from simple, fast RISC processor chips. Processors such as the Apollo DN-10000, the Sun SPARC, and the R2000 and R3000 from MIPS and DEC are setting a fundamentally new level in price-performance.

In addition, work is under way at several HEP sites, including CERN, to explore the use of massively parallel architecture. As we have to face the technical and economic challenges of processing data from LEP and other future accelerators, it is good to see that the physics community is starting to acquire expertise in these areas.

12. Conclusions

I was asked to address the question of whether we are ready with our computing on the eve of LEP data-taking. If I wanted to answer **No** I would point to some problem areas:

- There has not been enough testing of the total system (trigger, DAQ, processing and physics analysis).
- There is not enough disk space in place yet for successful processing.
- We can wonder whether the workstations will really be able to access the summary data quickly and reliably enough.
- We can expect problems in managing the production processing, and gaining access to the event data, held in files and on tape, in a sufficiently organised way.
- It is not clear that we can move the data quickly enough to the computer centres outside CERN.

- It is unclear that we will find enough money to process most of the data. We will certainly need significantly extra money in 1990/1991 for additional CPU, disk and network capacity.
- The computing specialists in the experiments and the support groups are heavily overworked and it is not clear that they will all stand up to the strain of the next year.

On the other hand if I expected to answer **Yes** I would say that:

- We have never been in such a good shape a few months before starting up four major experiments at a new collider.
- The CPU power appears to be in place, at least for 1989, and any future problems *only* require money to be solved.

So finally, I must be an optimist and say that we are ready, and will process the LEP data successfully *providing* we continue to receive sufficient backing and resources.

13. References

- [1] The MUSCLE Report - The Computing Needs of the LEP Experiments. CERN DD/88/1, January 1988.
- [2] E.M. Rimmer, Software for FASTBUS. CERN DD/89/7, February 1989.