# ON-LINE COMPUTERS IN DATA ANALYSIS SYSTEMS FOR HIGH-ENERGY PHYSICS EXPERIMENTS

G.R. MACLEOD

CERN, Geneva

The two techniques which we are going to discuss during this meeting - spark chambers and on-line computers - first became news in high-energy physics, in quite separate contexts, at the Berkeley Instrumentation Conference in 1960[1]. At that time, all the spark chambers which people were getting excited about were exceedingly photogenic, and the on-line use of computers in high-energy physics was foreseen specifically for the analysis of bubble chamber data. The on-line use was envisaged in the sense that one would build a special purpose box-of-tricks, which at that time was the Hough-Powell flying-spot digitizer, and that this device would run on-line with the computer to analyse bubble chamber film; that is to do the analysis <u>after</u> the experimental run at the accelerator.

Since then a large number of experiments have been done using spark chambers and the vast majority have used photography. The particular problems associated with the analysis of large numbers of spark chamber pictures are now on their way to being solved, also by the use of on-line computers. For instance, at CERN during the last fortnight or so the Hough-Powell flying-spot digitizer on-line with the IBM 7090 has processed completely automatically a hundred-thousand spark chamber pictures from an elastic scattering experiment at the rate of a thousand pictures per hour[4]. The kind of programming techniques which have been used for this automatic processing are quite valid for data from other kinds of spark chambers in the sense that once the digitizings are in the computer memory, then the programme does not really care whether they came off film or from some other device. In particular, very similar programming techniques to the ones used for the flying-spot digitizer here could, for instance, be applied to vidicon systems.

At the Instrumentation Conference held at CERN in 1962[2], there were first reports of efforts to bring spark chambers and computers together, doing away with the intermediate stage of film. Spark chamber techniques have been developed which digitize spark positions directly and provide these digitizings in a form which is suitable either to be recorded on magnetic tape, or to be fed directly to a computer. The on-line idea has also evolved; now one envisages building a special purpose box-of-tricks, which in this case is counter or spark chamber equipment in an accelerator beam, and this equipment is used on-line with the computer to analyse the recorded experimental data; that is to do the analysis during the accelerator run.

The use of on-line computers is a fairly widely used technique now in such fields as telemetry and guidance work with missiles and satellites, advanced radar systems, the oil and chemical industries and the communications industry. Getting nearer to our own field of activity, they are also being used now in low-energy nuclear physics where they masquerade under the name of multi-parametric analysers[3]. As it is only comparatively recently that high-energy physics has begun to profit from the use of on-line computers in experiments, I thought it might be useful at the start of this meeting to make a few introductory remarks illustrating the main characteristics of this kind of experimentation for those of you who are not yet familiar with it.

One can distinguish four functions which an on-line computer in an high-energy physics experiment can fulfil. These are data acquisition, checking and control, sample computation and data display. All of these things are very much interlinked and, although for the purposes of this talk I will discuss them separately, in any particular experiment it will be very difficult to draw clear dividing lines between them.

By data acquisition I mean the process whereby, on receipt of a signal from the counter trigger system, the computer reads into its memory certain numbers defining the event which has caused the trigger. Typically these would be some binary numbers from scalers or from some external buffer memory. The kind of data rates which one can reach now with available small computers are typically to transfer 18 or 24 bits in parallel in a time of the order of 2 to 5 μsec. The technique most convenient for this is to use a so-called interrupt feature of a computer. This is a property of the computer whereby the presence of an external signal, in this case the trigger signal, causes the computer to stop doing whatever it was doing and, in a time which maybe typically a few tens of μsec, to give its entire attention to handling the data which is presented. The mode of operation might be that during the accelerator burst, the computer records information from successive events and then in the interval between accelerator bursts, or when sufficient information has been collected, this raw data is read out on to some permanent storage, such as magnetic tape. With the present data rates experimenters are speaking of and using, it is quite feasible to store the raw data as it comes from the experimental equipment. When, as will surely happen in the next few years, the volume of raw data increases, it may well be necessary to do some kind of pre-processing before storing the experimental data simply to reduce the volume of magnetic tape and so on. On the other hand, it is, I think, most important that one should make a permanent storage as early in the data analysis chain as possible, simply to give oneself an insurance against later troubles in the analysis where one may want to get back as near as possible to the original experimental data; this also to some extent protects one against failures in the equipment in the later stages of the analysis.

Under the heading of "Checking and Control" the computer can do two things. Firstly it can do simple checking on the format and internal consistency of the data. It is important in designing these experiments to make sure that the data describing the event has sufficient redundancy in it that the computer can, in fact, carry out these kind of format checks on the data as it is read in. It is then possible to have a continuous check on what is going on and the computer can signal gross errors in malfunctioning of the equipment as soon as this occurs. The other kind of checking and control which has not been, I think, sufficiently emphasized in previous discussions on this subject, is that the computer can rather easily monitor all sorts of experimental parameters such as magnet currents, counting rates, voltage levels, beam intensities and so on. Taking this automatization one stage further, if these parameters are in any case susceptible to measurement they are equally susceptible to control, and it is quite practicable to think in terms of computer programmes together with analogue-to-digital and digital-to-analogue convertors on the input/output of the computer, such that the computer could, following the instructions of the experimentalist, maintain the values of certain experimental parameters, within certain preset limits. This could be done quite automatically and checked at rather frequent intervals. The computer, having this kind of information available, could also prepare automatically a rather detailed log of the experiment; the sort of things which nowadays tend to be written (or worse, not written) in note books in illegible pencil writing at three o'clock in the morning by harassed physicists, could in fact be done automatically by the computer with rather more consistency and reliability.

The sample computations which the computer may carry out are essentially concerned with analysing in a more fundamental way the data as it is recorded. Ideally one would like to analyse all the events as soon as they are recorded; restrictions on the amount of computing capacity which one can have available, the data rates involved and the sophistication of the computations one may like to make etc., all act to limit the possibilities of full computation on-line. However, one might take a sample of the events as they are recorded and, for instance, choose one in ten or one in five of the events for a rather more thorough investigation than the preliminary checking which I mentioned above. The kind of computing one would do for these samples might be geometrical calculations for instance with sonic chambers to compute the spatial coordinates of sparks, or simple kinematic calculations estimating momenta of particles, or they might be quite complicated kinematic calculations even going as far as kinematic least squares fitting. The choice in any specific case will depend very much on the computer available and the experiment one is trying to do. But this leads, I think, to one of the most important features of this kind of experimentation, which is that on the basis of these sample calculations the computer itself can act as a variable logic element in the overall detecting system of the experiment. One could, for example, run an experiment with a reduced stringency in the triggering system, thus increasing the counting rate, and then use the computer (for which one can write one or several programmes to vary the kind of selection made) to analyse the data and make the final selection of the events one wants. The computer programme could sort out the events recorded into various classes "good" or "bad", "elastic" or "inelastic", and could classify them according to criteria such as range or scattering angle, or scattered momentum, etc. All this kind of thing can be built into an overall system which includes, via the computer and its programme, this element of variable logic in the selection.

Under the function of data display, I mean the ability of the computer to feed back information to the experimentalist. A very simple sort of display is, I suppose, an electric typewriter, which is all right for printing short messages but is rather hopeless when you want to see rapidly what is going on with a lot of data. The kind of device which is most suitable at the moment for this kind of display is the cathode ray tube whereby the computer can display not only information for a given event but also histograms etc., summarising information from whole series of classes of events which have been recorded previously. It is equally important that associated with the display device there should be means by which the experimenter can communicate back to the computer. This is conventionally done through a console with switches and buttons and lights. This is a rather laborious way and requires very often some cool thinking about which button to press next; if one presses the wrong one an awful lot of information may be lost. A much more flexible way of talking back to the computer is to use a light-pen with a cathode ray tube; this kind of display enables the physicist to communicate also through a visual technique with the computer. I think this interaction between computer and the physicist is the second really fundamental gain which experimentation using an on-line computer offers - the physicist can get at information concerning events which have been computed just a few minutes or so after the events have been recorded. One can see rather rapidly the effect of changing experimental parameters; one can make the setting up on an experiment a rather less laborious and less random affair; one can actually see that the data being recorded is the data one wants to record; and one can plan the experiment on the basis of analysed data rather than having to make intuitive guesses based on what we did last time. All this I think makes for the physicist having a much better control over what is going on and gives him the information on which to base considered decisions on changes to be made during the experiment.

For all this, of course, one has to pay something and the main load which falls on the experimentalist is that he has to spend a lot of his time doing computer programming. The sample computations can in fact be handled quite readily with some automatic language of the FORTRAN type, but unfortunately the programming concerned with the input/output, with the control operations and handling interrupt situations has to be done in the basic language of the machine. About ten years ago any self respecting nuclear physicist spent a lot of his time building his own fancy coincidence circuits, and I think now we have reached a time where computer programming is a more or less essential skill for experimental physicists. This is not a terrible price to have to pay for what the experimental physicist can gain in recording and analysing a large amount of data during the course of the experiment. The other cost which does not worry the physicist so much is that one has to buy small computers. The kind of small computer ideally suited to this sort of work has as fast a memory as possible with perhaps 2000 to 4000 words of 18 to 24 bits, fixed point arithmetic and 1 or 2 magnetic tape units. For an outlay of something like half a million to a million Swiss Francs one has a fair choice of these machines. There is a very good selection of machines available in the United States; the European computer market is sadly lacking in this kind of computer and it would certainly help us if computer manufacturers, in planning their next products, would think about making available more of this kind of computer in Europe.

There are several kinds of development work which need doing to aid this kind of experimentation. Firstly, I think the most important thing is to be able to have some kind of cheap visual display. If one is prepared to pay two hundred thousand or three hundred thousand francs there are some very fancy cathode ray tube display consoles available and with which one can do all sorts of nice things, but they are a bit on the expensive side to be handed out to every experimental group. However, this kind of experimentation does require a means of displaying, either locally or remotely from the computer, a large amount of data in visual form and a comparatively cheap means of doing this is one thing which is lacking at the moment. Naturally one wants smaller, faster computers and one wants them to cost less. So far this tendency of increased speed and reduced price has been born out and one hopes it will continue to be so. The emphasis in computer design, I think, has to go very much in the direction of very flexible input/output, multichannel computers with buffered input/output and with rather sophisticated interrupt facilities. Another point on which I think development would be rather useful would be for computer manufacturers to produce the equivalent of the FORTRAN language to handle all these control functions and input/output operations simply, in order to ease the programming load which must necessarily fall on the experimentalist using this kind of technique.

There is a conflict between the requirements of a computer to be used on-line to do the data acquisition and the control functions, and the kind of computer one ideally would like to do the sample computations. The kind of small computer I described a few moments ago is ideal for the data acquisition and control, but for the sample computations one would like to do more general calculations, for example, least squares fits, for which floating point arithmetic and a larger memory are desirable. It would be very useful to have the more powerful order code of a machine like the IBM 7090 available. All these things tend to imply that for the sample computations one does not want a small computer but one wants a large computer. In the course of the next few years, I am sure that in high-energy physics the requests for small on-line computers will certainly grow (and this is a phenomena which has been observed already in other fields) as the advantages of this kind of experimentation are seen. As the amount of data experimentalists want to process increases then the requests will come for more computing capacity on-line and there will be a definite pressure to make small computers become middle-size computers become big computers. Just from the economics of this game there is some limit - a laboratory working on a fixed budget cannot afford to provide a 7090 for each experimental group. The direction which we intend to go in CERN to try and resolve this problem is to have a number of small computers on the experimental floors to carry out the data acquisition and control functions and to back these up by a large central computer which must obviously have time sharing properties. This will be connected by a data-link network to the small computers on the experimental floor. The small computers when they have enough data accumulated for the sort of computations which they themselves cannot do, can then signal this to the central computer which can take the data, do the computation and send information back to the small machine after computing. We are hoping to take delivery of a CDC 6600 towards the end of this year and we are already planning to use it in this kind of way. The communications industry regards it more or less as routine work to hook sizeable chunks of electronics together over distances of thousands of miles, so I hope that we can benefit from some of their experience and that it will not prove terribly difficult to bring some kind of scheme like this into operation.

In conclusion there are two points I would like to emphasize about experimentation with on-line computers. Firstly, it is very important that the whole experimental set-up including the detection equipment, the beam equipment, the computer and its programmes should be regarded as an entire system. One can no longer afford, as has regrettably been the case in the past, for an experimental group first to set up a beam, then to set up some spark chambers, then to fill a room with half a million pictures and only then to worry what to do about them. The whole advantage of using an on-line computer implies that one must see from the very early stages of planning an experiment that the data recording and the data analysis aspects are intimately related and must be planned together. Essentially this means, in rather crude terms, that one has to do an awful lot of planning and programming before one gets near an accelerator. Secondly, I would like to disarm one criticism which I have heard made on several occasions against using on-line computers in experimental physics of this kind. The criticism is that when you use a computer you make a completely automatic experiment which is perfectly well defined from the time you start. This is about as different as it possibly could be from the real situation. In using an on-line computer as part of his experimental equipment the physicist's role of decision making during an experiment is enhanced by having the possibility to make judgements on the basis of real live information, not just on intuition. He is in a much better position to be really aware of what is going on at any given time, what has gone on, and what he should do in the future in running his experiment.

I can perhaps finish with a somewhat outrageous remark which may contain a grain of truth. I think it is clear that counter and spark chamber experiments in the last few years have been rather limited by their inadequate data handling facilities. The well known fact that all elementary particle physics in the last few years has been done in bubble chambers may not be unconnected with the fact that the bubble chamber physicists have had well developed data handling facilities. Now we should listen to the reports from four experimental groups who have been working very hard to try and change this lamentable situation. Thank you.

References

1. International conference on Instrumentation for High-Energy Physics, Lawrence Radiation Laboratory, Berkeley (1960), Proceedings, Interscience Publishers, New York, 1961.

2. International conference on Instrumentation for High-Energy Physics, CERN, Geneva, 1962, Proceedings, Nucl.Instrum.Meth., 20 (1963).

3. Conference on the Utilization of Multi-parameter Analysis in Nuclear Physics, Grossinger, New York, Proceedings, Columbia University, N.Y., CU (PNPL) 227, 1963.

4. P.M. Blackall, G.R. Macleod, P. Zanella; Proceedings of the 1963 International Symposium on Nuclear Electronics, p. 357.