

August 1965

COMPUTATION AND CONTROL IN COMPLEX EXPERIMENTS*

W. F. Miller

Stanford Linear Accelerator Center
Stanford University, Stanford, California

IBM Scientific Symposium

on

Man-Machine Communication

at

The Thomas J. Watson Research Center
Yorktown Heights, New York
May 5, 1965

Contents

I.	INTRODUCTION	1
II.	CURRENT SYSTEMS	3
III.	GRAPHIC DATA PROCESSING	5
IV.	CONCLUSIONS	10
	REFERENCES	12
	CAPTIONS FOR FIGURES	14

* Work supported by the U.S. Atomic Energy Commission.

I. INTRODUCTION

Most of the work presented in this Symposium has been concerned with problems that arise in the development of systems that permit a man rapid access to computational results. This paper is concerned with problems that arise in the development of systems that are employed for data analysis and control in complex experiments. Such systems might be represented schematically by the block diagram in Figure 1. Such a diagram has the same general structure as a guidance system, and, indeed, it should, since its function is to guide an experiment. Such systems, of course, give rise to many of the same problems already discussed; but, on the other hand, they introduce some additional ones.

The motivation for on-line data analysis and control systems for experiments is very much the same as the motivation for rapid access to computational results. Both developments are motivated in part by the hypothesis that rapid feedback is essential to learning. It is clearly hoped, and by now it may already have happened, that a researcher sitting at his console will make discoveries by seeing his results feed back practically instantly. In the experimental systems, it is also expected that a researcher, by having his data quickly analyzed and by having intimate control over his experiment, can make discoveries and direct the course of the experiment in a more profitable way.

Let us look a moment at the scope of the problem. As physics experiments probe deeper and deeper into the fundamental constituents of matter, the experiments become much more complex and the interpretation of the data becomes more subtle. A single experiment involving a high energy particle accelerator and a piece of detection apparatus,

such as a spark chamber or a bubble chamber, may take several weeks or even months to set up and may run continuously for equal lengths of time. The accelerators themselves are quite costly (several tens of millions of dollars) and they are expensive to keep in operation (several million dollars per year). The detection equipment is also quite expensive. A bubble chamber may cost several million dollars and a spark chamber may cost several hundred thousand dollars. Clearly, there is an economic motive to provide data analysis systems that facilitate effective use of such expensive equipment. Rapid feedback of analyzed results is particularly important during the setting up stage.

It may be worth pointing out that a typical spark chamber experiment conducted over a few weeks will generate 500,000 to 700,000 stereo views of events taking place in the spark chamber. The stereo views may be generated in pairs or triads, depending on the experiment. The manual and semi-automatic data analysis systems now in use require weeks and months to analyze this data. The Lawrence Radiation Laboratory at the University of California at Berkeley has had long experience on this problem and, by now, has good information on what it takes to analyze the data. Recent summaries [Reference 1] indicate that it takes about seventy-five people to do the scanning, measuring and maintaining the equipment in order to process 250,000 events per year from the Berkeley 72-inch bubble chamber.

It is hardly necessary to pursue this motivational discussion further. It can be simply summarized by the statement that in the current era of experimental physics a disproportionate amount of time and effort is spent on the analysis of data and the control of the experiment. As a consequence, the experimentalist is removed farther and farther from intimate contact with the essential elements of his experiment.

The goal of rapid data analysis and control systems in complex experiments is to bring the experimenter back into intimate contact with his experiments. Man-machine communication is required at a rather sophisticated level.

Over the last few years, I have been concerned with computer methods to achieve this goal. In particular, I have been concerned with the elements that would be contained in boxes 3, 4 and 5 of Figure 1.

II. CURRENT SYSTEMS

An example of a data analysis and control system of the kind depicted in Figure 1 is given in Figure 2. The system shown is a closed-loop analysis and control system involving a 3.0 MeV Van de Graaff Accelerator and its nucleonic measuring equipment. I shall not discuss this system in detail here, since it has been described elsewhere [Reference 2]. I should like to point out, however, that this system has all the essential ingredients of the general system shown in Figure 1. It has, indeed, accomplished all the objectives that the physicists who were developing it expected.

The type of control that this system provides the physicist is shown in Figure 3. This flow chart represents the first experiment programmed on the system [Reference 3]. It provided the physicist a running statistical sequential analysis, configuration control over the sample, and energy control over the accelerator. The system has evolved very rapidly and there are now several experiments programmed for the system [Reference 4].

This system is utilized in low energy nuclear physics experiments of a type that are relatively well understood. As a consequence, the data analysis codes and the experiment programmer was not required to

have the generality that would be required of such a system in high energy particle physics. Nonetheless, some quite valuable experience was gained in programmed control of experiments.

Now, how do we stand in the very complex experiments such as we encounter in high energy particle physics? We are certainly making progress in all areas indicated in Figure 1. At present, there are no high energy particle accelerators operating under computer control. [By computer, here, I mean internally-stored program computer]. At two Centers, however, computer controls are being developed. A digital control system, employing a CDC 924, has been developed for the Zero Gradient Synchrotron (ZGS) at the Argonne National Laboratory. It is expected that this system will soon be "hooked up" and that the ZGS will operate under its control. At the Stanford Linear Accelerator Center (SLAC) a computer control system is being developed for the beam switchyard. The beam switchyard will employ bending magnets to deflect the electron beam from the accelerator into different experimental areas.

The problems represented by box 3 are in somewhat better shape. Several on-line data analyzers have been employed in low energy nuclear physics [References 5, 6, 7, and 8] and one system has been developed in high energy physics with great success [Reference 9]. At Stanford we are also developing a system similar to that described in Reference 9. This system is intended to analyze on-line the data generated by the 20 BeV/C Magnetic Spectrometer.

We are also developing a computer system that will permit on-line analysis of graphic data of the kind we shall get from filmless spark chambers. There are two large general problems that have to be dealt with: (1) the control programs to permit processing of data from

devices with high burst rates, and (2) the data analysis techniques for handling the graphic data. Both of these areas are getting a great deal of attention and are important to the development of integrated systems. I shall give below a discussion of the type of programs developed to handle the data analysis part.

III. GRAPHIC DATA PROCESSING

A. General Description

In our work at Stanford, we are taking as a point of departure the Argonne work on automatic film data processing. That work was the collaborative effort of several people in the development of a film digitizer, programs for running the film digitizer on line to a large computer (CDC 3600) and the analysis programs for handling the digitized film data [Reference 2].

Although the Argonne work was done explicitly for film data, the programs maintained sufficient generality that they can be used for any graphic data from spark chambers no matter how the data is presented to the programs. Figure 4 shows the structure of these programs. The scanning and measuring program AROMA prepares, for the AIRWICK programs, the digitized information from the photographs. The AIRWICK programs identify the corresponding sparks in the two stereo views, calculate their positions in three space, and then link these sparks into tracks.

Figure 5 shows a stereo pair of photographs of an event taking place in a spark chamber and Figure 6 shows the results of two stages of processing - after AROMA and after AIRWICK.

One of the most pleasing aspects of this work is that we can give a formal description of each step. Formally, the problem is presented as (a) generation of a graph, the vertices of which are given by the three space coordinates of the sparks, and (b) the selection of the

proper tree (or forest of trees for multiple events) to represent the event. The processes physically described as (1) scanning, measuring, and image transformation, (2) pairing in the stereo views, and (3) linking into tracks have their counterparts in the formal description. The first two of these processes, i.e., scanning, etc., and pairing, etc., are concerned with the generation of the graph and the third, i.e., linking, etc., is concerned with the tree selection. Figure 7 shows the relationship between the physical processes and the formal description. The essential elements of the system are described in brief detail below.

Let me first give some data on the rates of manual and then of our automatic system. Using the manual systems and skilled human operators, one can expect to scan and measure at the rate of about 20 to 30 events per hour. We are able to scan and measure at about an order of magnitude greater speed. We can process a stereo view of complex nature in about 20 seconds and simple ones in 10-12 seconds.

B. Scanning, Measuring, and Image Transformation.

The CHLOE film digitizer is described elsewhere [References 2, 10] and will be described only briefly here. CHLOE is a hardware system for digitizing data recorded on transparent 35 mm. film. The hardware consists of (1) a controlling computer, (2) an optical scanner (CRT) operating under the control of the computer, and (3) a data link to a larger computer. A spot from the Cathode Ray Tube is projected on the film and the transmitted light is measured by a photomultiplier. From this measurement a decision is made concerning the density of any rectangular portion of a 4096x4096 raster measuring 1.25 inches on a side. In practice, one scans only a small portion of this raster.

The scanning and measuring functions are carried out in the CHLOE LIBERATOR program which resides in the controlling computer (an Advanced Scientific Instruments 210) and the AROMA program which resides in the CDC 3600. The principal feature of these codes is the CELL CONSTRUCTION ALGORITHM. The CELL CONSTRUCTION ALGORITHM may be described as follows.

Let us assume we are engaged in horizontal scans. The computer permits a left-to-right scan across a window defined on the raster by left-right limits and top-bottom limits. A point of interest is recorded when the scanner detects a change in intensity of transmitted light. As a consequence, in scanning across a spark one gets first the left and then the right end of a line segment, the left end of which is the first point in the spark and the right end is the first point out of the spark. See Figure 8. If several sparks are encountered in one horizontal scan, one obtains ordered pairs of left-right coordinates of line segments, i.e.,

$$x_{1,\text{left}} < x_{1,\text{right}} < x_{2,\text{left}} < x_{2,\text{right}} \dots$$

The CELL CONSTRUCTION ALGORITHM sorts these line segments into cells, eliminates some of the cells as clearly not being sparks, and calculates certain spark parameters such as area, centroid, and average width.

The AROMA program also provides information on the fiducial marks which are used to provide the orientation of the film coordinates to the spark chamber coordinates.

The next step is to identify the corresponding spark images in the different stereo views and to generate the three space coordinates of the sparks. Before doing this, the data is passed through a program (PREPOS) whose function is to remove optical distortions and to transform CHLOE film coordinates of spark centroids into real-space coordinates on the surface of the spark chamber.

C. Pairing In The Stereo Views.

Since the separate gaps of a spark chamber are easily discernible the pairing problem need only be concerned with pairing spark images in the same gap. The first step in the pairing is the generation of an n-dimensional relationship array - one dimension for each view.

Figure 9 shows the two-view relationship matrix that would be generated for the five sparks generated in the spark chamber. The relationship matrix contains a 1 if it is geometrically possible for the rays to have originated in the chamber and a 0 if it were not geometrically possible. The PAIRING ALGORITHM removes the ambiguities from the relationship matrix. The ambiguities are removed by application of a pair decision function

$$(1) \quad D(i,j) = M(i,j) \left[w_1 F_1(i,j) + w_2 F_2(i,j) + w_3 F_3(i,j) \right]$$

and its comatrix

$$(2) \quad A(i,j) = \sum_{n=1}^v D(n,j) + \sum_{m=1}^s D(i,m) - 2D(i,j)$$

where v is the number of rows in a block and s is the number of columns. It should be noted that the ordering of the sparks is responsible for the break up of the relationship matrix into block diagonal form. Pairings may occur only within subblocks so the decision function may be applied to each block separately. The decision function is constructed to take into account both intrinsic parameters and extrinsic or contextual parameters. The intrinsic parameter is width and enters through F_1 .

$$(3) \quad F_1(i,j) = | - |w_1 - w_2|$$

The contextual parameters are order and interference and these parameters enter $D(i,j)$ through F_2 and F_3 respectively.

$$(4) \quad F_2(i,j) = 1 - \frac{|i - j|}{P}$$

where P_{+1} is the maximum dimension of the block.

$$(5) \quad F_3(i,j) = \left(1 - \frac{S(i)-1}{R} \right) \left(1 - \frac{T(j)-1}{R} \right)$$

where R is the maximum dimension of the block, $S(i)$ is the number of geometrically possible pairings of spark i , and $T(j)$ is the number of geometrically possible pairings of spark j . The F 's are all normalized to yield values in the interval $(0,1)$. The weights w_1 , w_2 , and w_3 are experiment dependent. Large values of $D(i,j)$ mean high probability of pairing. On the other hand, low values of $A(i,j)$ mean high probability of pairing - a low value of $A(i,j)$ means that there is low probability that the i th spark of view one could be paired with any spark other than the j th spark of view two, or that the j th spark of view two could be paired with any spark other than the i th spark of view one.

By repeated application of $D(i,j)$ and $A(i,j)$ successful pairings are determined. With each determination, the relationship matrix is reduced by one row and one column until all ambiguities are resolved. It is possible that all ambiguities are not resolved and that some are left unresolved until the linking operation.

The PAIRING ALGORITHM then passes on to the LINKING ALGORITHM a relationship matrix with almost all ambiguities removed.

D. Linking Into Tracks.

The output of the PAIR program is a set of three space coordinates that form a graph. The LINK program selects the proper tree (or forest in the case of multiple events) to represent the particle tracks. The detailed tree selecting algorithm is given in Reference 2. Briefly, edges are established between the various vertices of the graph on the basis of a number of criteria, such as Euclidean distance, number of vertices, direction of the edge, linear and helical extrapolation, and geometry of the spark chamber. Finally, the subgraph selected is the minimal connector tree that contains no circuits. The determination of the minimal connector tree is accomplished by an algorithm due to Kruskal [Reference 11].

The graph data are stored in memory in a multi-word list. Each list item contains all the necessary information about a given spark. Each list item contains seventeen words in all, including the three coordinates of the spark, its gap number, its chamber number, the local degree of the vertex, pointers to as many as seven connecting sparks, distance and pointer to the closest spark, and distance and pointer to the second closest spark.

Having arrived at a set of coordinates that represent the paths of particles participating in the event the data are now sent on to fitting programs and programs that extract the physics information from the data.

IV. CONCLUSIONS

In order to develop more complete systems of the type depicted in Figure 1, it will be necessary to incorporate complex analysis

programs of the type described in Section III above. This is by no means all of the story. The kinematic analysis programs and the hypothesis testing programs that follow are also very complex. However, progress is being made and physics data are being analyzed at a very rapid rate. Moreover, independently work is proceeding on the control aspects.

One development which I feel is not getting its full share of attention in these problems is in the area of displays, control consoles, and the system control languages. That is, the command posts are not being developed as thoroughly as they should be. The splendid use of graphics that we have seen in other areas could be put to excellent use in these large analysis and control systems.

REFERENCES

1. F. Solmitz, Stanford Particle Physics Colloquium, April 5, 1965.
2. Robert Clark and W.F. Miller, "Computer Based Data Analysis Systems", Methods in Computational Physics, Vol. V., (in Press).
3. W. Miller, J. Meadows, A. Smith, and J. Whalen, "On-Line Data Acquisition with Feedback to Accelerator Control", Proc. Conference on Automatic Acquisition and Reduction of Nuclear Data, Kernforschungszentrum, Karlsruhe, Germany, July 13-16, 1964, p. 222.
4. A.B. Smith and J. Whalen - to be published.
5. Edwin Norbeck, "An MPA System That Is Flexible, Reliable and Fast", The Proc. on the Conference on the Utilization of Multiparameter Analyzers in Nuclear Physics, Grossinger, New York, November 12-15, 1962, N.Y.O. 10595, p. 56.
6. W.F. Miller, "The Role of Computers in Experimental Physics: A System for On-Line Analyzers", Proc. of Conference on Utilization of Multiparameter Analyzers in Nuclear Physics, Grossinger, N.Y., November 12-15, 1962, N.Y.O. 10595 (March 1963).
7. R.H. Vonderohe and D.S. Gemmell, "A Description of the PHYLIS On-Line Computing System", Proc. of Conference on Automatic Acquisition and Reduction of Nuclear Data, July 13-16, 1964, Session IIc, Kernforschungzentrum Karlsruhe; Gesellschaft für Kernforschung m.b.H. Karlsruhe, Germany (1964), p. 156.
8. _____ "The NBS On-Line System - Design for Flexibility", Nucleonics, Vol. 22, p. 57, December, 1964.
9. S.J. Lindenbaum, "The On-Line Computer Counter and Digitalized Spark Chamber Technique", Physics Today, Vol. 18, p. 19 (April 1965).

REFERENCES - (Cont'd.)

10. Donald Hodges, "CHLOE, An Automatic Film Scanning Equipment", ANL-AMD Technical Memorandum No. 61, November 1963 (Unpublished).
11. J. B. Kruskal, "On the Shortest Spanning Subtree of a Graph", Proc. A.M.S., Vol. 7, p 48 (1955).

CAPTIONS FOR FIGURES

Figure 1. General Schematic of Integrated Data Analysis and Control System for Complex Experiments.

Figure 2. Closed Loop Analysis and Control system Being Utilized in Low Energy Nuclear Physics Experiments.

Figure 3. Schematic Flow Chart of First Program Run on the System Depicted by Figure 2.

Figure 4. Schematic of the Flow of Data in the Argonne System.

Figure 5. Photograph of Spark Chamber Event in Two Views.

Figure 6. Output of AROMA and AIRWICK for Event Shown in Figure 5.

Figure 7. Schematic of the Relationship of Physical Processes to Formal Description.

Figure 8. Line Segments and Cells Generated by CELL CONSTRUCTION ALGORITHM.

Figure 9. Schematic of Top of Spark Chamber with Five Sparks and Corresponding Relationship Matrix.

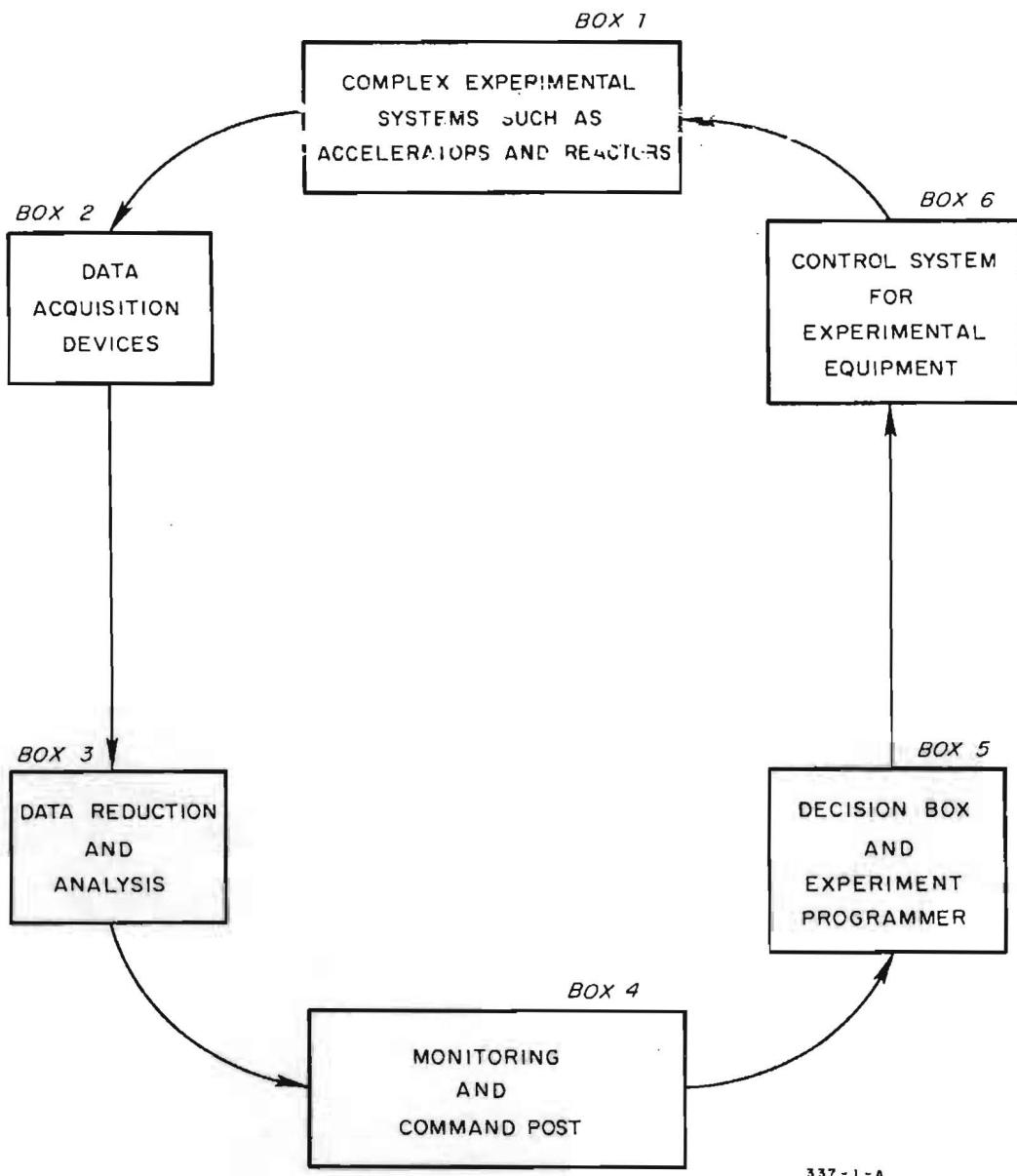


Fig. 1

337-1-A

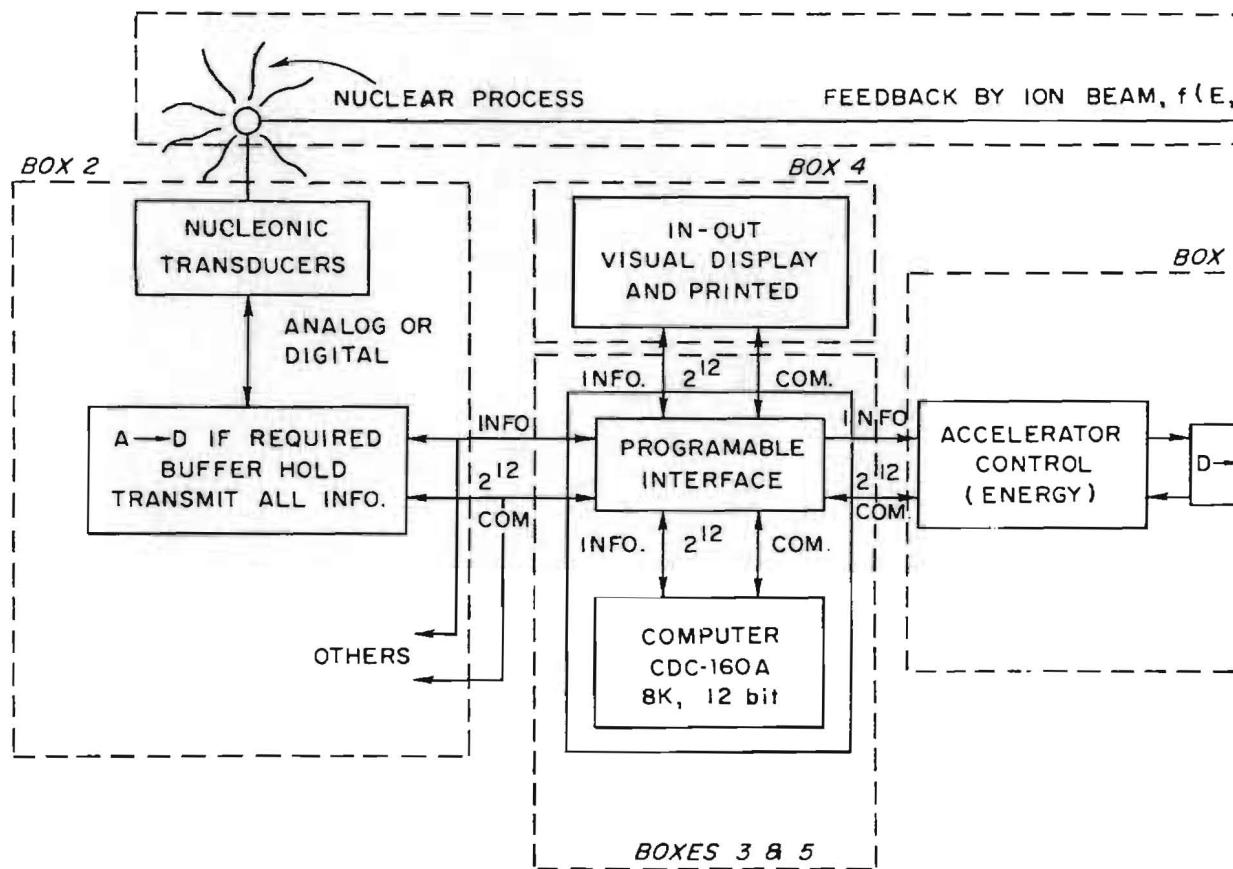


Fig. 2

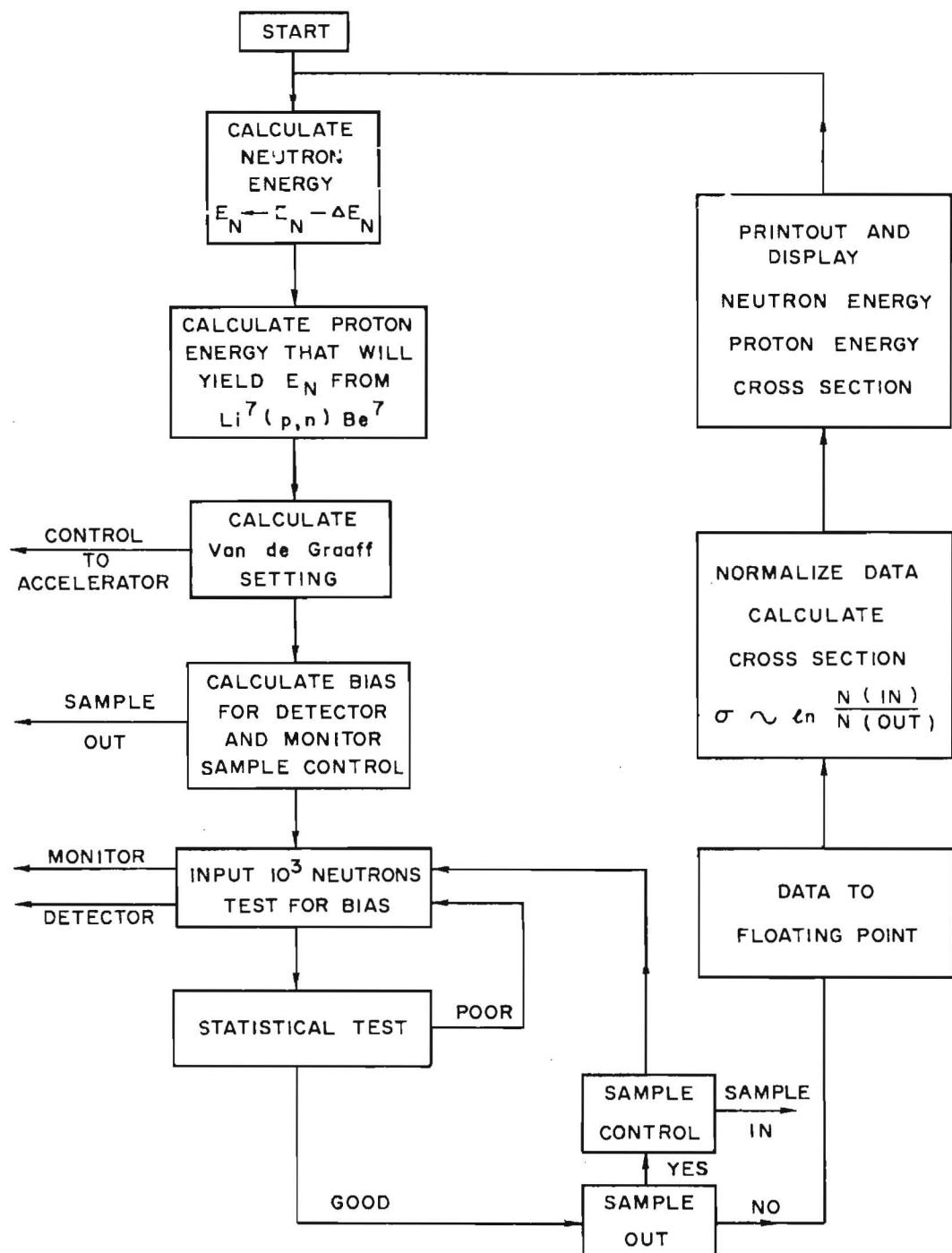


Fig. 3

337-3-A

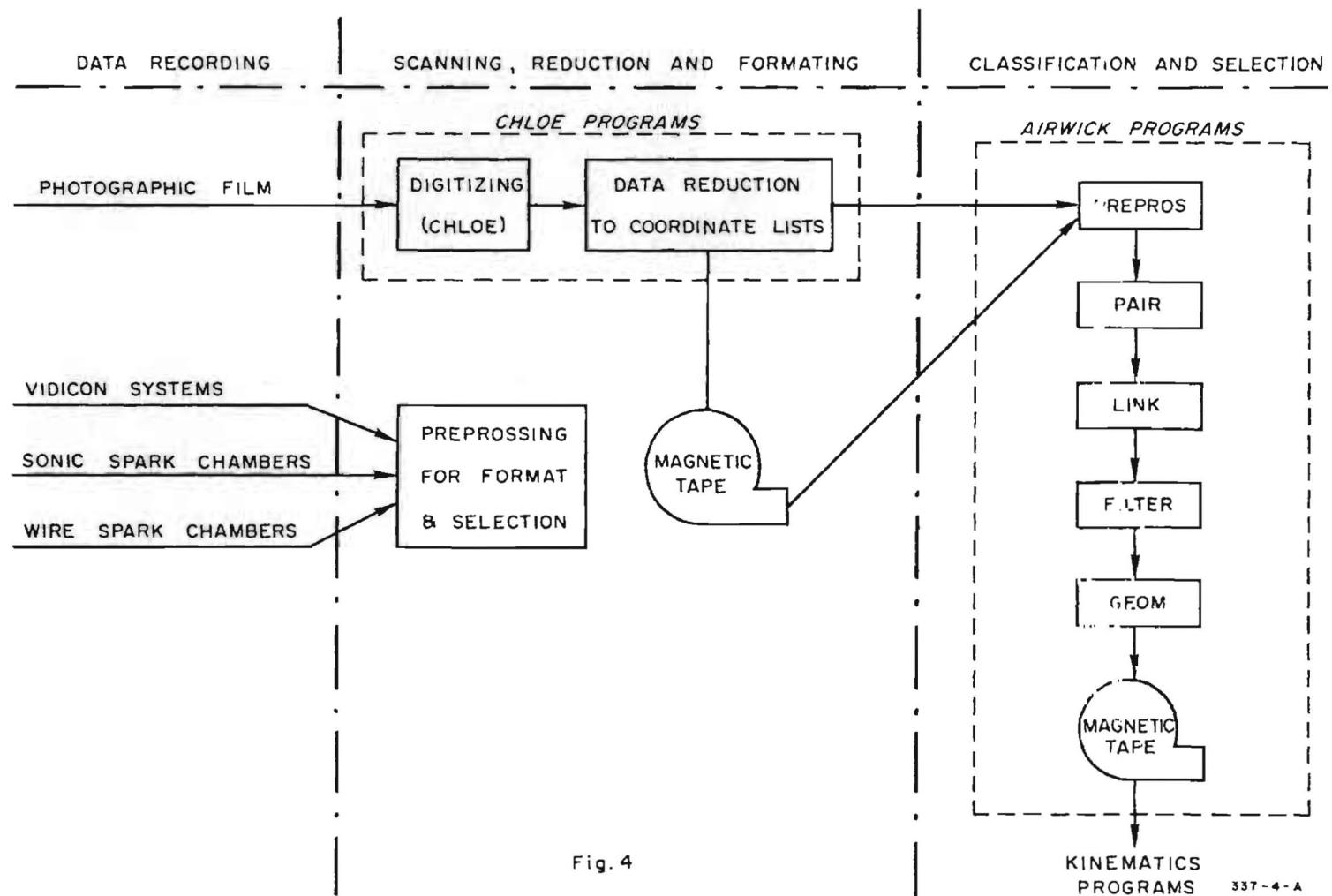
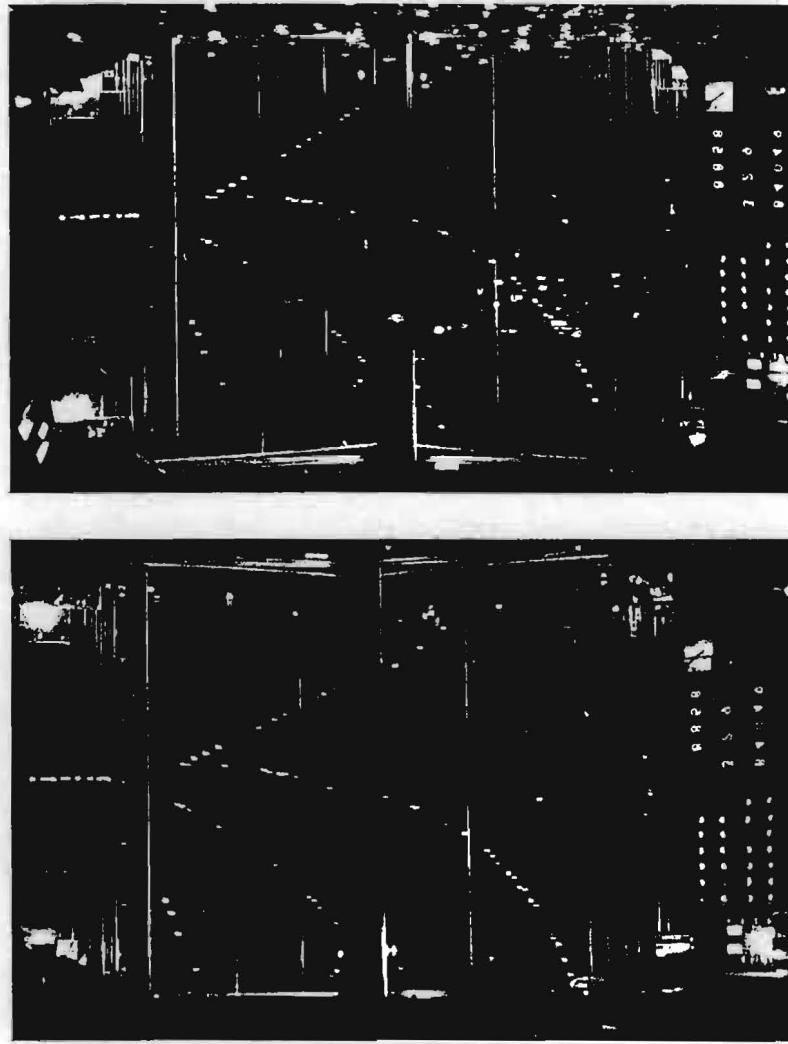


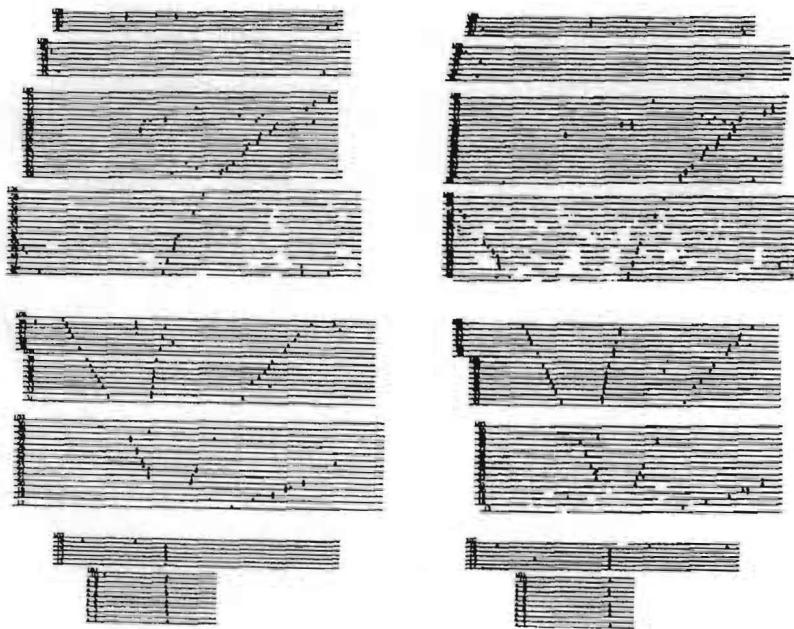
Fig. 4

KINEMATICS
PROGRAMS

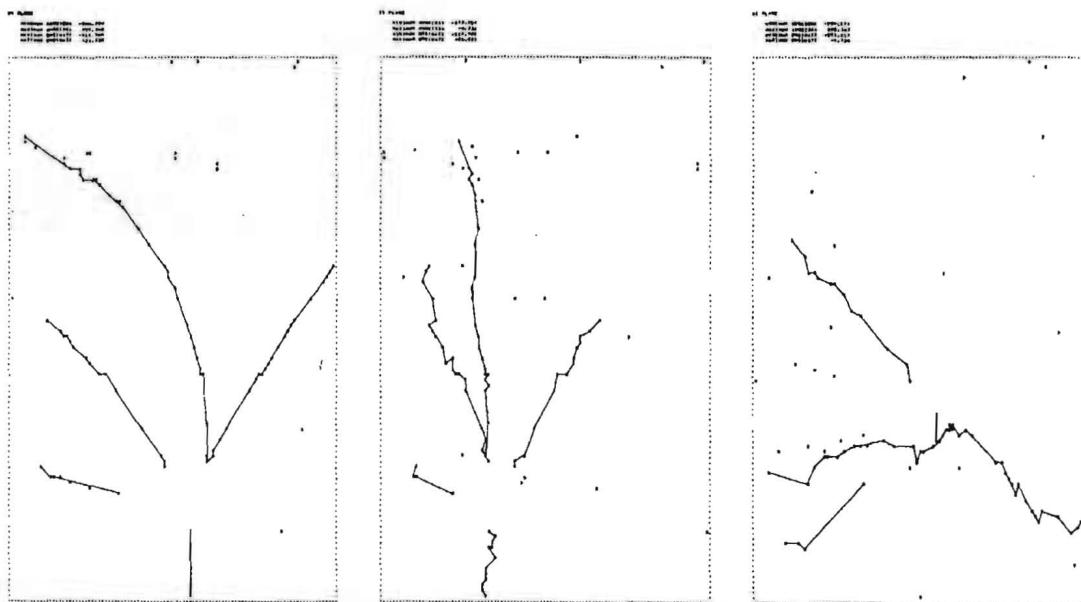
337-4-A

Fig. 5





Measurements from the Same Two Photographs after Processing
by AROMA



The Event after Processing by AIRWICK, Showing the Reconstructed
Tracks. Random Sparks Not On the Tracks Have Now Been Eliminated.

Fig. 6

PHYSICAL PROCESS

FORMAL DESCRIPTION

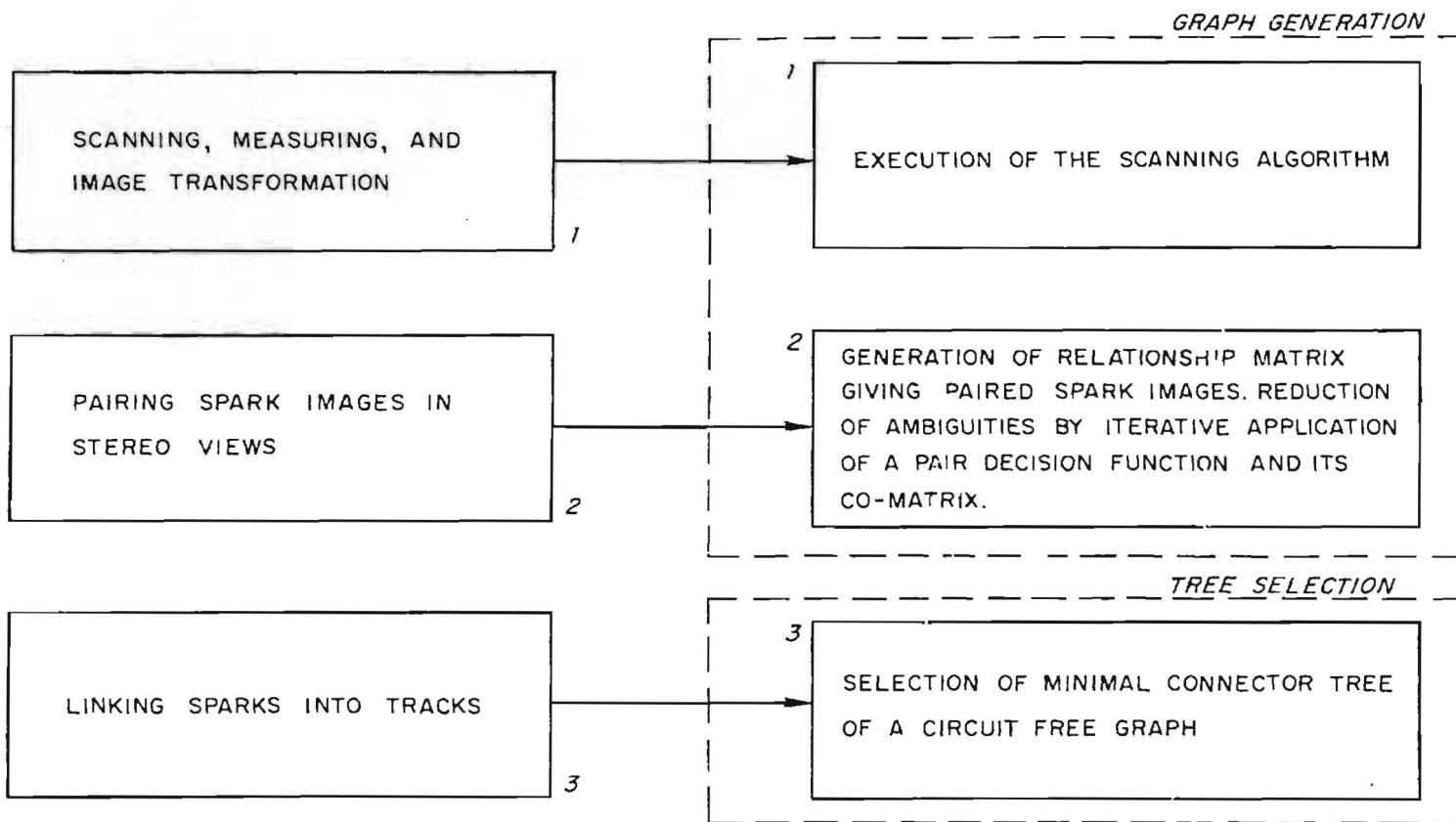
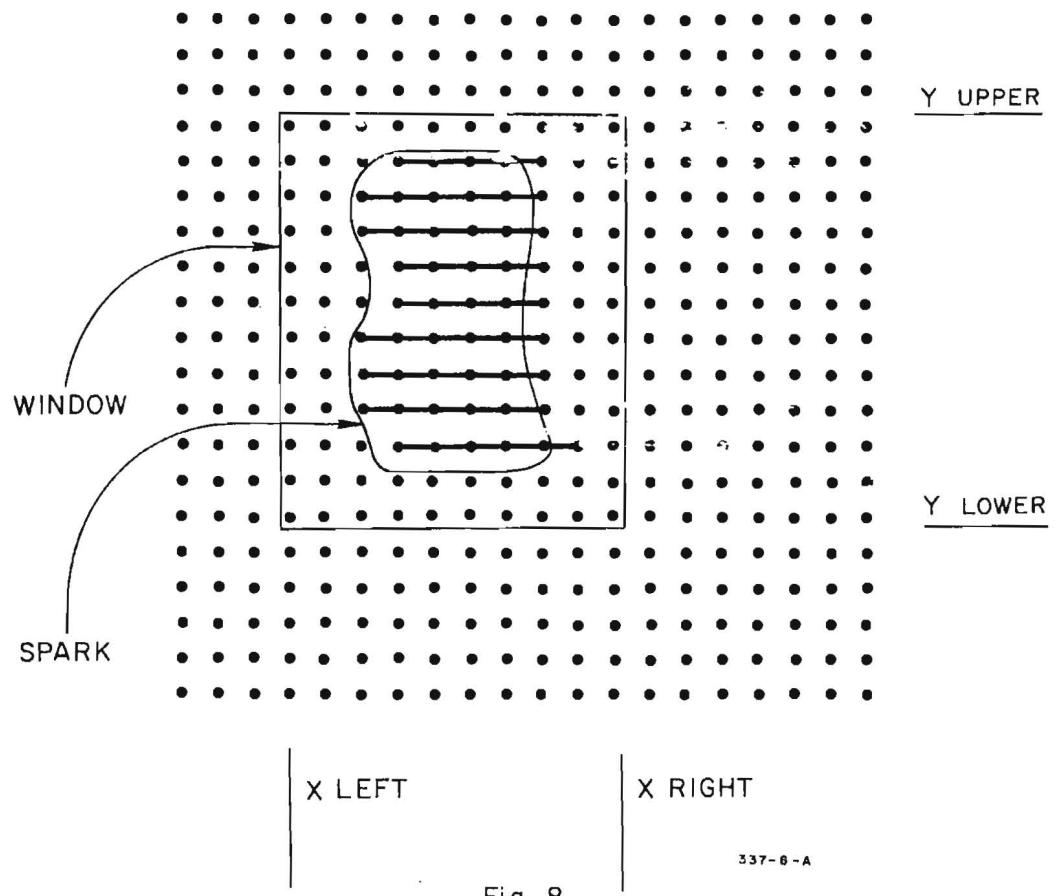
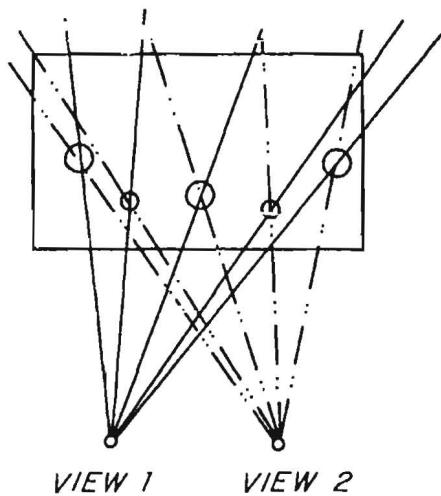


Fig. 7

337-5-A



TOP VIEW OF CHAMBER



RELATIONSHIP MATRIX

$j \rightarrow \text{VIEW } 2$

1	1	0	0	0
1	1	0	0	0
0	0	1	0	0
0	0	0	1	1
0	0	0	1	1

Fig. 9

357-7-A