Presented at Symposium on Future Extraction and Selection in Pattern Recognition, Argonne, Ill.
Oct. 5-7, 1970

UCRL-20126
Preprint

c. 2

FINDING EVENTS IN A SEA OF BUBBLES

Howard S. White

October 1970

AEC Contract No. W-7405-eng-48

UCRL-20126

c. 2

LAWRENCE RADIATION LABORATORY
UNIVERSITY of CALIFORNIA BERKELEY

DISCLAIMER

UCRL 20126

FINDING EVENTS IN A SEA OF BUBBLES

Howard S. White

Lawrence Radiation Laboratory
University of California
Berkeley, California

DAPR is a digital automatic pattern recognition system, designed to find and measure events in bubble chamber film without manual intervention. It is able to measure film which has come directly from the photographic development process, and to produce on magnetic tape a digital abstraction of the information contained in the film, from which a subsequent selection of desired events takes place by means of a digital scanning process.

The DAPR system at Lawrence Radiation Laboratory, Berkeley, achieved production status earlier this year, and already has abstracted more than fifty thousand bubble chamber picture sets. Although development continues in some areas of the programs, DAPR measurements are being used by physics experimenters in the analysis of their data. It is expected that DAPR will process an increasing fraction of the data as the measurement of current experiments is completed, and that of new experiments is begun.

DAPR SYSTEM GOALS

In planning DAPR, a number of goals were set as standards by which its performance could be rated. The HAZE system for manual scanning and automatic measurement of bubble chamber data was chosen as the basis for comparison, since extensive experience with it had been gained in a wide variety of experiments,[1)] and because of its excellent measurements at low cost. The basic information to be obtained from the bubble chamber picture is the spatial orbit for each of the particles which form tracks,

from which the particle's momentum can be determined, and the bubble density along the orbit, from which the particle's velocity can be determined.

The most important requirement was that DAPR perform as well as HAZE with regard to the accuracy of the points characterizing each track's locus, and to the measurement of the fractional digitizing which indicates the bubble density along the track. Another important goal relates to the fidelity of the event: that the correct tracks be associated with the events found. Because of the desire to obviate manual scanning, the discovery of all events within the search area is of great importance, as is also the absence of fake events caused by identifying random configurations as event vertices. Finally, DAPR must operate as cheaply as the HAZE system to be justified economically.

DAPR HARDWARE

The hardware of the DAPR system has been in use since 1963 as part of the HAZE system. The system uses an IBM 7094 II computer, which is operated in a multi-programmed mode in a manner like that of contemporary systems. Film is digitized by a Flying Spot Digitizer (FSD) of the Hough-Powell type,[2] which is operated online to the computer. The scan line of the FSD is generated mechanically, the spot being formed by the intersection of a fixed slit and a slit carried by a rotating disc. A mercury vapor arc lamp illuminates this aperature to produce a spot about 18 microns in diameter which is mechanically constrained to move at uniform velocity along a straight line. Motion of the film mounted on a precision measuring engine allows the digitizing of a raster scan of the film image. The track width in typical bubble chamber images is 25 microns, and individual hits on a well separated track yield an rms scatter of about 1.5 microns. The bubble density is measured by comparing the number of digitizings to the total number of times the spot intersects the track locus.

Since DAPR operates without manual guidance, it must conduct a scan of the entire picture area for events of interest. On the other hand, the high inertia of the mechanical stage precludes a random return to local

local areas of the image. The number of digitizings which are obtained from one raster scan of the image is too great to store in core memory, and the staging of these on the disc is relatively expensive with this computer. Therefore, DAPR was designed to operate its track following procedures in a real-time relationship with the data acquisition hardware, so that only a greatly reduced set of data need be transmitted to the disc for further processing. Since the FSD produces some 15,000 words of input per second to the real-time program, the speed of the computer places a moderately severe restriction on the complexity of the track following procedures.

TRACK FOLLOWING

The real-time program controls the FSD, and performs the track following process. Output for each track consisting of eighteen well distributed master points and a measure of bubble density, is staged to the disc for further processing by the programs which occupy the second priority level within the multiprogrammed environment.

When a few images have been accumulated on the disc, this second phase of processing is activated. Fiducials are identified, and unwanted measurements of recurrent marks in the chamber image are deleted. The point sets are combined or partitioned until ideally each actual track in the chamber image is represented by exactly one set of points in the measurement data. A track certification routine uses the rms scatter from a fitted polynomial of appropriate degree to achieve these decisions.

EVENT RECOGNITION

Many elements of pattern recognition have been used to bring the data to this point in the reduction process, but true feature extraction begins here. The primary aim of this paper is to discuss these procedures of feature extraction which find and identify event vertices, and which associate the correct tracks with them from all surrounding tracks.

Let us briefly review the characteristics of the data which are presented to this feature extraction phase of the processing. In the ideal case, three image data sets, corresponding to the three "stereo"

views of the bubble chamber, would each have its track data sets standing in one-to-one correspondence with the tracks which a person might perceive upon viewing the film images. This ideal is closely approximated when the beam tracks are well separated and the event has an open configuration, as shown in Figure 1. All too frequently the event is obscured by a bundle of closely spaced beam tracks, or by an electron spiral unrelated to the event. In some of these regions of confusion, the track images are not resolved by the digitizer, so that digitizings are made on the composite image of multiple tracks. In some regions, the digitizings are so densely distributed that no clustering along individual tracks can be discerned by the track following program, and it is not able to correctly associate digitizings with the proper point sets. Very short tracks are not found in the track following process, and some pairs of tracks which meet in the vertex at nearly 180 degrees are followed as one continuous unit. The configuration of some events is such that serious overlapping may occur from tracks of the event alone. Some examples of these images are given in Figures 2-7. Our experience shows that nearly half the events are in some degree effected by one or more of these problems.

As a basis for the vertex recognition, we therefore have point sets which generally give a highly accurate representation of the track images, but which may have a few distorted points due to confused regions. Generally, the entire track is represented, but sometimes a substantial part has not been followed, usually the part nearest the vertex. Generally the point sets represent one track, but sometimes two are joined at a kink which has not been partitioned, and sometimes two point sets redundantly represent one track. The feature extraction process would nearly be trivial if only ideal images were given it; most of its complexity comes from the processes which extract event vertices from the less-than-ideal images with which it must deal.

VERTEX FINDING

First, a list of candidate vertices is determined separately in each view by searching for a cluster of endpoints. The process is illustrated in Figure 8. A fairly large box is constructed around each endpoint in

the central region of the chamber image, and all other endpoints within the box are considered. Throughout the event recognition process, each track is represented by a circle. If the circle under consideration and one whose endpoint is within the box intersect quite close to at least one of the endpoints, that point of intersection is taken as a provisional vertex. All circles whose endpoints lie within the box are added to this provisional vertex if they pass close to it. The vertex is further screened to eliminate redundant discoveries and some common classes of fakes. A best point of intersection is calculated by a least squares procedure, which also develops a Chi-squared estimator of the probability that all tracks intersect in a common point. Survivors of this process become "view-vertices" and are collected into a table along with a list of their included tracks. When all of the endpoints within the central area of the image have been considered, the other views are searched in the same way, without reference to vertices found in previous views.

When all three views have been searched, their tables of "view-vertices" are compared to find spatial agreements. Since the optical properties of the chamber and the camera lenses are known by the program, the three "view-vertices" are highly constrained, and their constitutient tracks can be used to develop a reliable location of the vertex within the chamber space, as well as a Chi-squared estimator of probability, that all intersect in a common spatial point. This intercomparison of views is also the most powerful test available to the human scanner for answering the question of whether a vertex is a true event or an accidental configuration. All spatial vertices resulting from "view-vertices" found in two or three views and which lie within the chamber volume are retained. A vertex location is predicted in the third view if it was not otherwise detected, and various redundancies due to tracks being split between two vertices in the single view process are cleared up at this time.

Some tracks may still remain unassociated with their proper vertices for a variety of reasons. Sometimes the track-following process fails to produce a point set which covers the entire track. Sometimes two tracks intersecting at a 180 degree angle are followed as one, giving no endpoint

at the vertex. Occasionally the vertex point is poorly determined by the single-view calculation, so that tracks are not properly associated with it. And predicted vertices initially have no associated tracks. Each surviving "view-vertex" is therefore compared to all tracks not yet associated with it in a process known as the exhaustive search.

The program assumes that all tracks actually participating in the event will have been associated with their "view-vertices" at the conclusion of this process. Frequently other tracks such as nearby beam tracks, and tracks which are unassociated with the vertex, but point toward it, are included as well.

TRACK MATCH

A track matching procedure is used next to relate tracks in the different views. Many tracks accidentally passing near a "view-vertex" can be excluded because no match exists in other views. Ambiguities are flagged for further discrimination. Since the vertex point is well determined, point sets which included parts of two tracks can now be partitioned, and their separate parts matched as appropriate.

Although no ambiguities would remain in the ideal event at this point, many actual events need further work to reduce them to unambiguous status. The procedures used to simplify vertices are mostly based on probabilistic arguments, but are really justified on empirical grounds. We have adopted the philosophy that tests based on properly weighted geometric factors are compelling when no ambiguities are present, and that ambiguities can be resolved only by choices which are unlikely to be the result of accidental configurations. The cleanup procedures favor three-view track matches over two-view matches, since an accidental three-view agreement is very much less probable than is a two-view match. Preference is given to sets which define the same length of track-following coverage in the different views. The tracks are classified into several categories in terms of their resemblance to beam tracks, and preference is given to agreement of this classification between views. Surviving ambiguities are resolved by means of a Chi-squared test applied to the agreement of radius in the three views. Since only one beam track can

participate in an event, and because many are often found in the close vicinity of the vertex as a consequence of the clustering of beam tracks into a tight bundle, the program selects the beam track in each view which passes close to the vertex and yields the best fit to the true beam orbit. A final test uses the known fact of charge conservation at the vertex to further determine if the track set is completely plausible in the context of the experiment.

The best justification for these procedures is empirical: do they reduce ambiguities, and yield the same final association of tracks and vertices which one would correctly choose by manual processes? We find that they do reduce many ambiguities and almost never produce a false choice. For example, the radius Chi-squared test was evaluated by reprocessing a large number of events that had been measured in the HAZE mode, where track matching assignments are made manually. A careful study was made of the 128 tracks having disagreement between the manual HAZE choice and the automatic DAPR choice. In 127 of these, the DAPR choice was clearly the correct one, while the remaining track was entirely indeterminate, and we found no way of distinguishing which ambiguous choice was valid.

THE DATA ABSTRACT TAPE

The vertex recognition program writes a tape containing all of the tracks as point sets, together with a table giving the spatial vertices and references to their associated tracks. This Data Abstract Tape (DAT) is the digital equivalent of the film, but it contains a set of reduced measurements of each track, together with the physical vertices in perceived form, all expressed in digital format convenient for the computer. On this tape, the features of the bubble chamber pictures have been extracted, and written compactly for future use by the computer.

The process of selecting events of a certain configuration has long been achieved by visual scanning. Since scanning could proceed much more rapidly than measurement, it was first used to screen only the most suitable events for measurement. In the DAPR process, this sequence has

been reversed. It is necessary that some measurement has taken place for digital "scanning" to occur, and since in DAPR this has been a complete and highly accurate measurement, no further access to the film is required.

The DAPR scanning process compares events described on the DAT with scanning criteria which are supplied by the experimenter. The criteria may be identical to those given the manual scanner: a topological description of the event, instructions for naming participant tracks, the area to be searched, and other such selection criteria. In addition, since a very good estimate of the momentum vector has been developed for each track, much more sophisticated criteria can be given, including those which seek other tracks or vertices some distance away from the primary vertex, and identify them by means of kinematical calculations. Such calculations are not practical for manual scanners.

The DAPR scanning program edits the data from the DAT to a magnetic tape, where it is completely equivalent to data from either conventional microscope or else HAZE measurements. All vertices meeting the criteria of any desired event type are written in standard form, with the tracks named and ordered in accordance with the scanning instructions, and with the point sets and ionization information supplied from the data sets contained on the DAT. The DAPR scanning process is intentionally a very rapid one, so that one may scan for events of interest without being burdened by the requirement of searching for others of future interest because of prohibitive rescanning costs. In practice, the DAT can be scanned at the rate of about 10,000 picture sets per hour of central processor time. This compares with a maximum rate of 200 picture sets per hour for manual scanning.

DAPR RESULTS

The success of DAPR can best be measured by comparison to the basic goals previously mentioned. A very detailed comparison was made between the measurements in both HAZE and DAPR modes of about 3000 two-prong events.[3] This comparison established that the quality of measurements of events output by DAPR is fully as good as that from HAZE. Because the comparisons were made on a track-by-track basis, histograms could be

constructed of the difference between DAPR and HAZE measurements of the final track orbital parameters, scaled by their appropriate errors. Figures 9-11 show these histograms for the angles and momentum which describe the tracks of 1741 two-prong events. Not only do the central portions of the distributions support the conclusion that the majority of the tracks have statistically equivalent measurements, but the small number of tracks having disagreements greater than 4 standard deviations indicates that almost invariably the same tracks were associated with the events by DAPR as HAZE. A study of the few discrepancies showed that by far the most frequent cause for disagreement was a difference in track length over which the measurement was made. In these cases a small kink was often unknowingly included in either the DAPR or HAZE measurement. Thus, we find that the fidelity of the track association is as good in DAPR as in HAZE. Although a few fake events were found by the DAPR scanner, it was clear that most of these could have been excluded by more complete scanning instructions.

Since completion of the test experiment, considerable work has been done to improve the fraction of events which are output by the DAPR scanner. Remembering that only those events which can be reduced to unambiguous status can be handled by the DAPR scanning program, we recognize that even though most vertices are found, some are not in form to be edited to the output set. On the other hand, since their frame number and vertex location is known, a highly satisfactory finding list for events to be manually reviewed is to be had. Present results with film of reasonably good quality show that 80% of the vertices are written out in unambiguous form for final analysis.

The distribution of $\cos \theta^*$ (the recoil angle of the proton track in the center of mass system of elastic events) provides a further confirmation of the quality of the DAPR measurement. This is shown in Figure 12. The normalized HAZE data is represented by the dotted line, and the DAPR data is represented by the solid line. The depletion of DAPR data in the first two cells, and in the last cell is due to the predicted bias from the as yet incomplete vertex algorithm and short stub procedures of DAPR. When

the data in the central 37 cells were compared a χ^2 value of 8.0 was calculated for a 20 degree-of-freedom fit. Thus, except for predictable biases, the DAPR measurements are seen to be in excellent agreement with the HAZE measurements.

EXPERIMENTAL USE OF DAPR

The DAPR system has been used for measurement of experimental physics data at LRL-B since February 1970. More than 50,000 picture sets have been measured. We are presently measuring in the DAPR mode a second set of 60,000 events, with measurements proceeding at the rate of about 15,000 per week. System capacity with our existing hardware should reach 20,000 events per week for data from small chambers like the LRL 25" HBC, and will be further increased to 30,000 per week with the completion of a tandem FSD unit which is now under construction. This means that a system is now in operation which is capable of measuring 1,000,000 hydrogen bubble chamber events per year without any manual assistance.

We are continuing to develop some areas of the system in the light of experience which has been gained from early results. Three areas of track following need to be improved: the following of short tracks, the following of tracks in regions filled with beam tracks, and the earlier initialization of tracks leaving the vertices. All of these needs have long been evident, but it was felt that other parts of the system were more urgent. Similiarly, improvements are needed in track certification, so that better linking and partitioning procedures can result in better data sets being presented to the vertex recognition program. This latter needs improved procedures for separating close vertices and for associating the proper tracks with each.

THE FUTURE

We expect the use of HAZE to diminish in favor of DAPR. HAZE has been used to measure several experiments ranging in volume from 100,000 to 500,000 events, and the future appears to portend even larger experiments to be done entirely automatically with DAPR. Not only is the saving in cost of the manual scanning significant, but even more important, the

great efforts required to coordinate and maintain consistent scanning of such large experiments by manual methods are prohibitive.

As large chambers of the five-meter class come into use within the next few years, we hope that experience which is being obtained with data from the current two-meter class will lead to an easy transition. Because these new chambers will use "fish-eye lenses" in contact with the liquid, their images will suffer severe optical distortions that will make manual scanning extremely difficult, if not impossible. DAPR will not be bothered by these effects, although the large number of line elements produced by chamber features other than tracks may be a problem. If reasonable attention is given to clean operation of the chamber, we expect that DAPR's entirely automatic scanning process will be the most satisfactory means of finding and measuring events in this class of chambers.

We view DAPR as a step toward the analysis of bubble chamber data concurrently with the actual experiment. Presently, bubble chamber pictures are exposed at the rate of 4,000 per hour, while their analysis proceeds many times more slowly. With the tandem FSD, DAPR will operate within a factor of 10-15 of this exposure rate using the relatively slow IBM 7094 computer. Given the development of suitable data acquisition hardware which can look directly at the chamber, and can digitize the information in real-time, several computers exist even now with central processor speeds sufficient to perform the DAPR process in real-time synchronism with the chamber operation. Thus, a DAT could replace the photographic film as the primary store of information from the experimental run. Concurrent analysis would not only relieve the massive bottleneck in data analysis, but also could serve to feed back important information useful in guiding the course of the experiment while it is still in process.

We believe that entirely automatic feature extraction has been achieved for bubble chamber data by the DAPR system. This process is already more economical than manual systems. Furthermore, it is rapidly becoming more effective. The complete automation of this process opens new vistas for bubble chamber experimentation.

ACKNOWLEDGEMENTS

The development of DAPR has been truly the result of the efforts of many people. Dennis Hall has made invaluable contributions to all parts of the DAPR system, and is primarily responsible for the design and implementation of the track certification, matching and cleanup procedures. Programmers Joan Franz, Barbara Britton and Wen-Sue Gee have made the programs be reliable and versatile tools for feature extraction. The continued support of LRL Director Edwin McMillan has made this development possible.

This work has been supported by the U.S. Atomic Energy Commission.

REFERENCES

1. H. S. White, "Status of the LRL Flying Spot Digitizer". Proceedings of the 1966 International Conference on Instrumentation for High Energy Physics, Stanford, California. (Clearinghouse for Federal Scientific and Technical Information, NBS, Springfield, Virginia)

2. P. V. C. Hough and B. W. Powell, "A Method for Faster Analysis of Bubble Chamber Photographs". Proceedings of an International Conference on Instrumentation for High Energy Physics, Berkeley, California, September 1960. (Interscience Publishers, N.Y., 1961)

3. H. S. White and D. E. Hall, "Performance of the DAPR System." Proceedings of the 10th International Conference on Data Handling Systems in High Energy Physics, Cambridge, England, March 1970. In Press.

FIGURE CAPTIONS

1. A typical frame from the LRL 72" HBC showing the DAPR extraction of a 4-prong event. (a) The actual film image. (b) A display of the geometric information which describes the image on the Data Abstract Tape. (c) The desired event has been selected by the DAPR process.

2. An image of the LRL 25" HBC which contains two problem events. The 2-prong event at the top of the picture is overlaid by two additional beam tracks and a fiducial arm. The 2-prong event with secondary contains a short connecting track that has only a small angle deviation from another track at either vertex. Both events place a considerable demand upon the track-following procedures.

3. An electron spiral, unrelated to the 2-prong event, overlaps the recoil proton, and is likely to produce some distortion of its measured point set.

4. Two 2-prong events in the upper right corner of this picture give problems. The earlier event has a very small angle scatter with a short recoil track, and is made more difficult by being superimposed on an adjacent beam track. The other event has both outgoing tracks very forward, with a very small opening angle, so that the vertex location is subject to perturbation by small distortions of the points representing the tracks.

5. The production and decay of a Σ^+ particle is shown above and to the right of the chamber center. The short track was produced by the Σ particle before its decay. It is important to keep the DAPR program from finding a 2-prong event composed of all tracks except the short Σ track. The 2-prong event below this is ideal for DAPR.

6. The 2-prong event with secondary at the lower right of this image obscures itself. A person perceives the change in bubble density at the vertex, and infers that the track leading to the secondary 2-prong must lie beneath the recoil track. DAPR has not yet been able to do as well.

7. The 2-prong event at lower right is also a problem due to its configuration. A person would see the few bubbles of the lightly ionizing track and infer that the two tracks are mostly superimposed, but these appear no different than other noise to DAPR as it stands now.

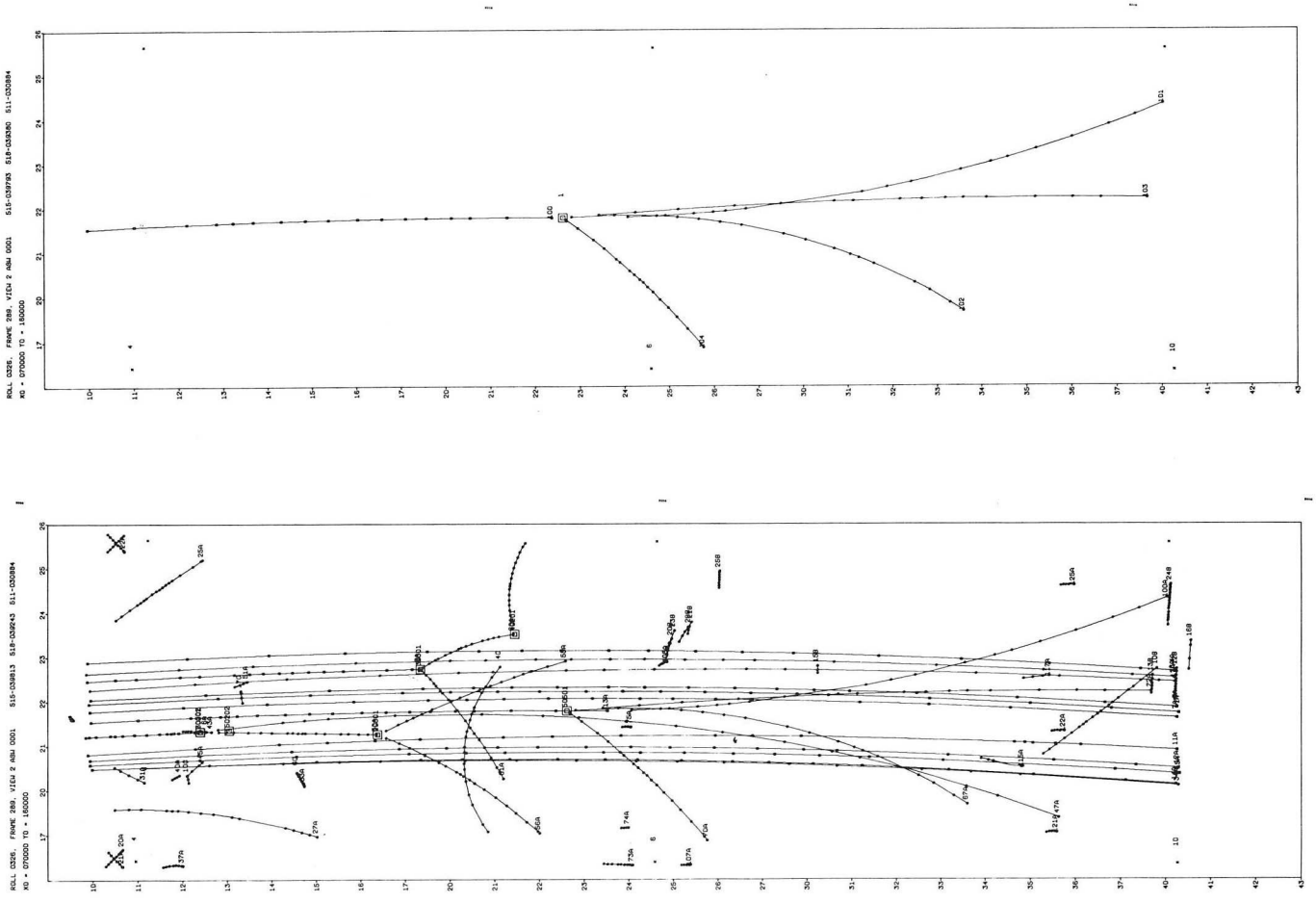
8. The vertex recognition algorithm depends upon a cluster of endpoints and the nearby intersection of track orbits to locate provisional vertices.

9. A comparison of the DAPR and HAZE measurements of the track elevation angle shows excellent agreement between the two systems. Individual track angle measurements were differenced and normalized by their stated errors.

10. A comparison similar to Figure 9, but of the azimuth angle.

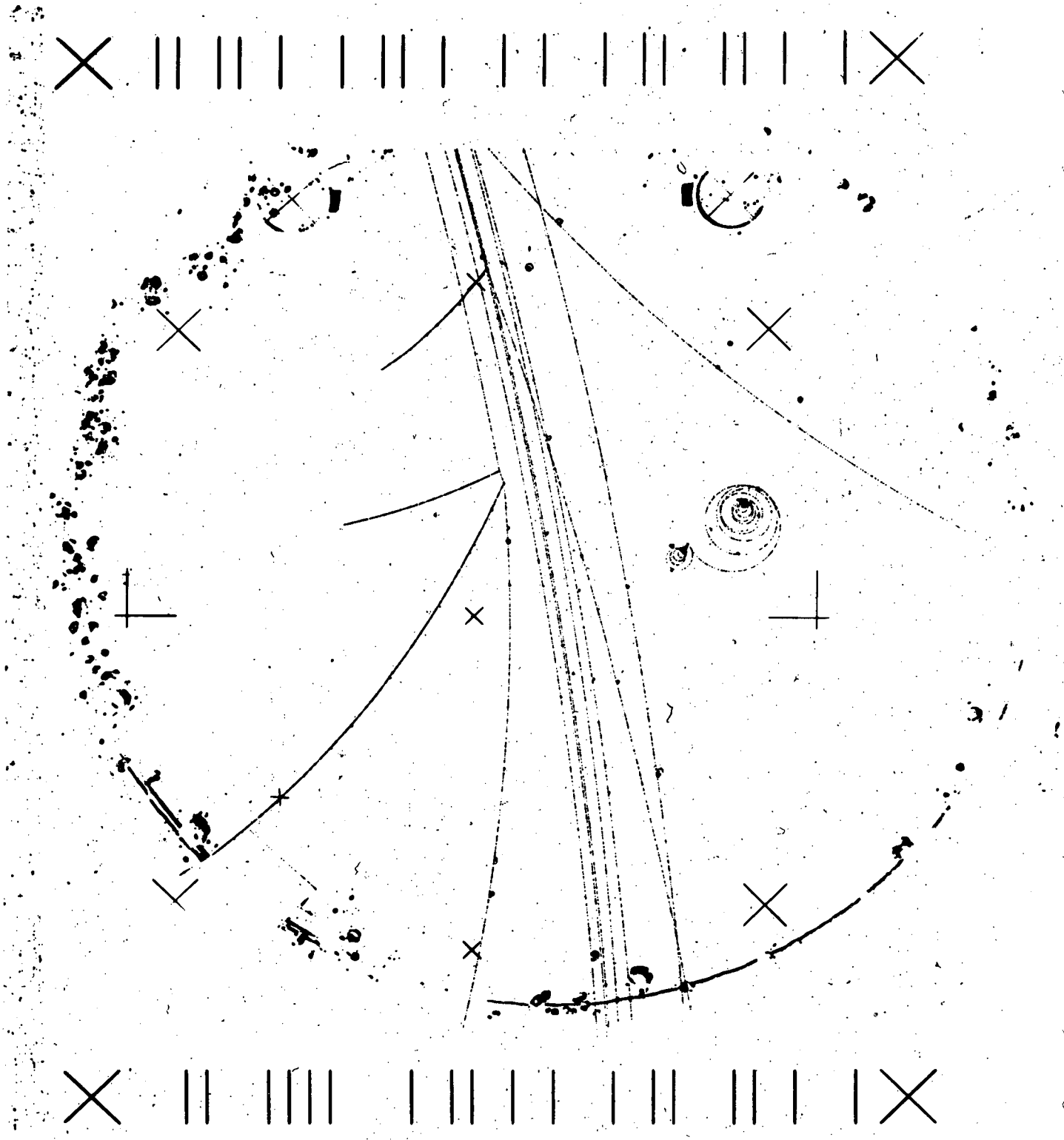
11. A comparison similar to Figures 9 and 10, but of the measured value of momentum.

12. A comparison of the proton recoil angle in the center of mass system for elastic events demonstrates lack of bias in the DAPR finding process, except for events having extremely small scattering angle. The central 37 histogram cells (omitting two at the left and one at the right of the distribution) were used for normalization, and yield a χ^2-value of 8.0 for a 20 degree-of-freedom fit.



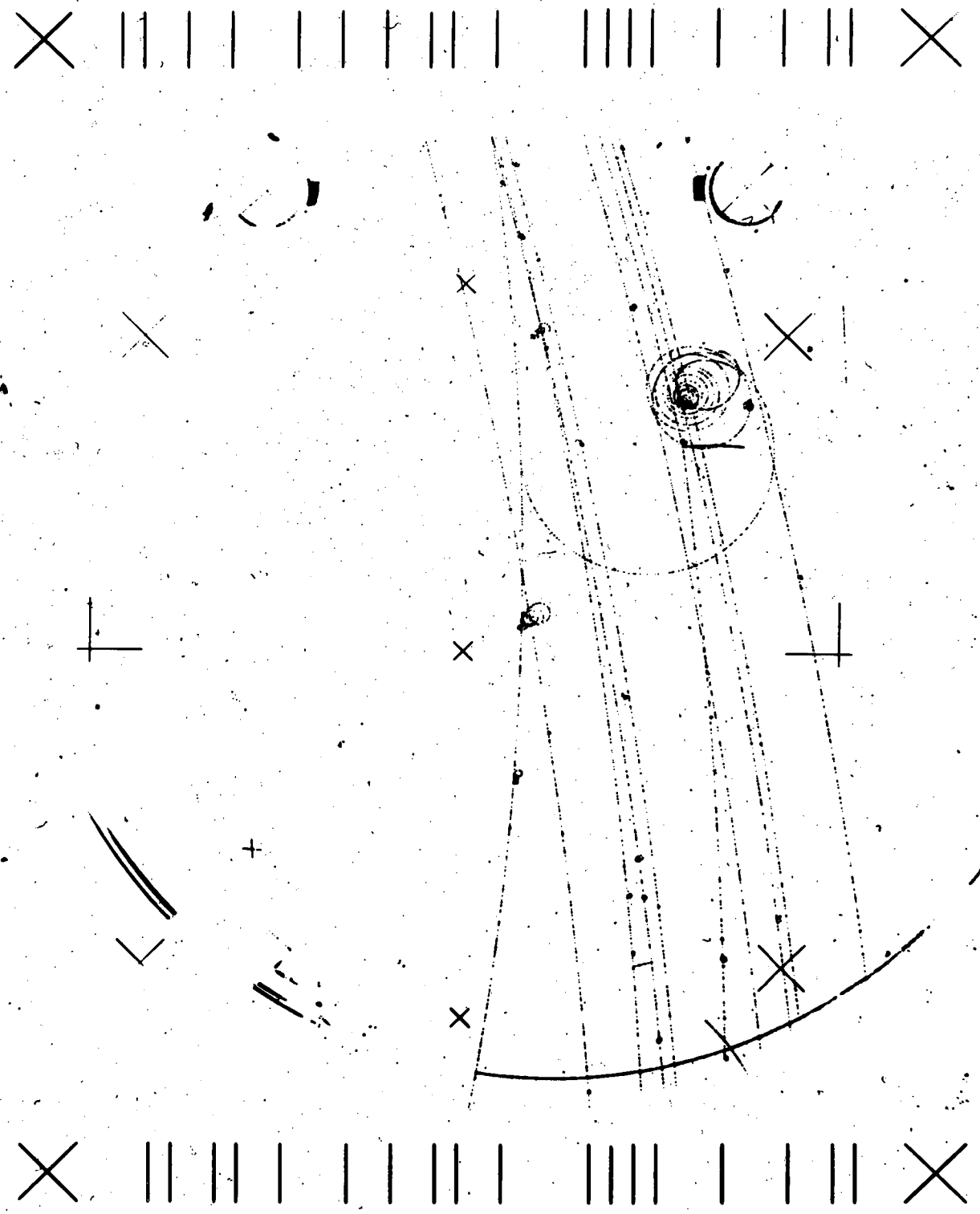
XBB 6810-6052

Fig. 1

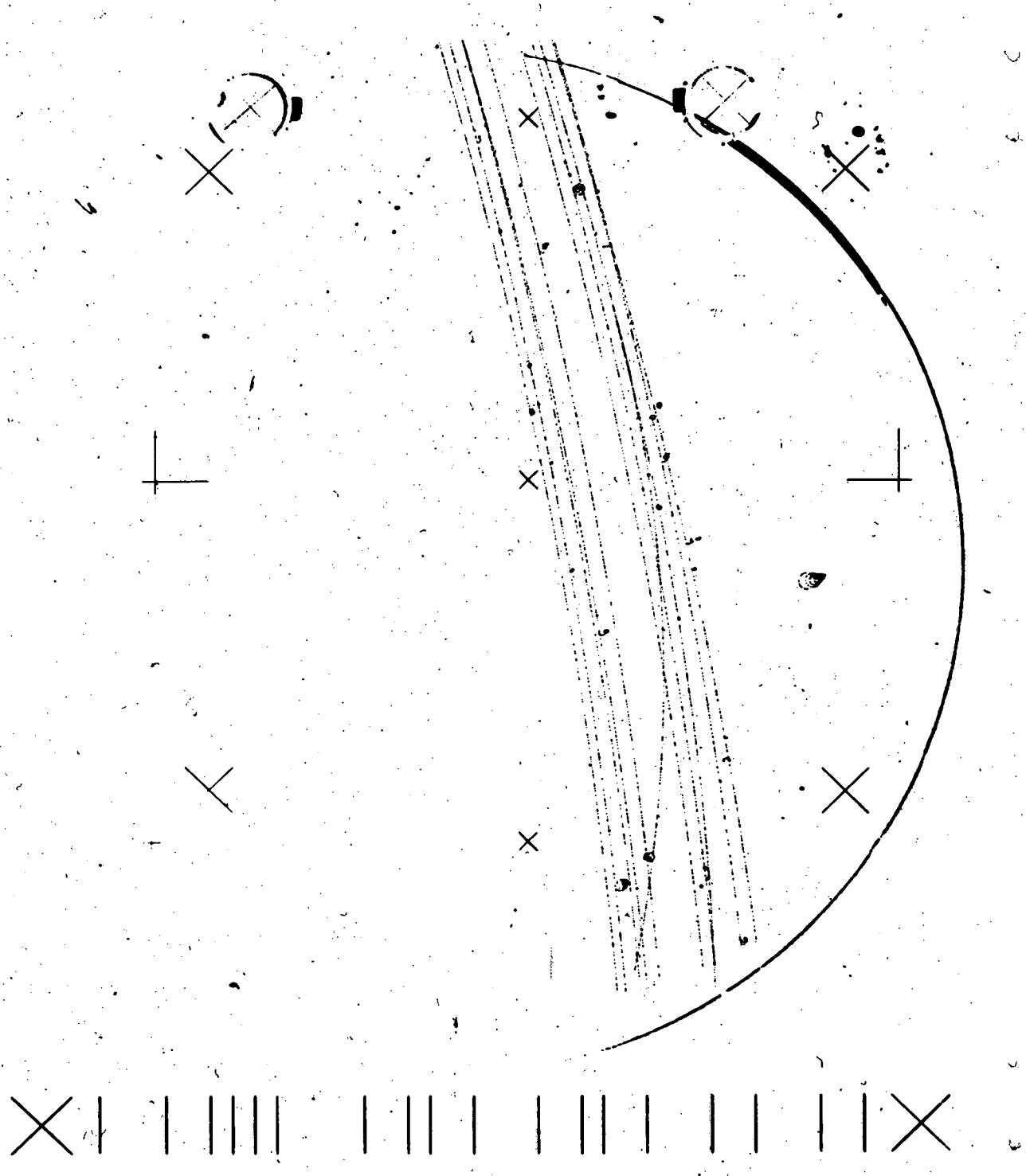


XBB 702-910

Fig. 2

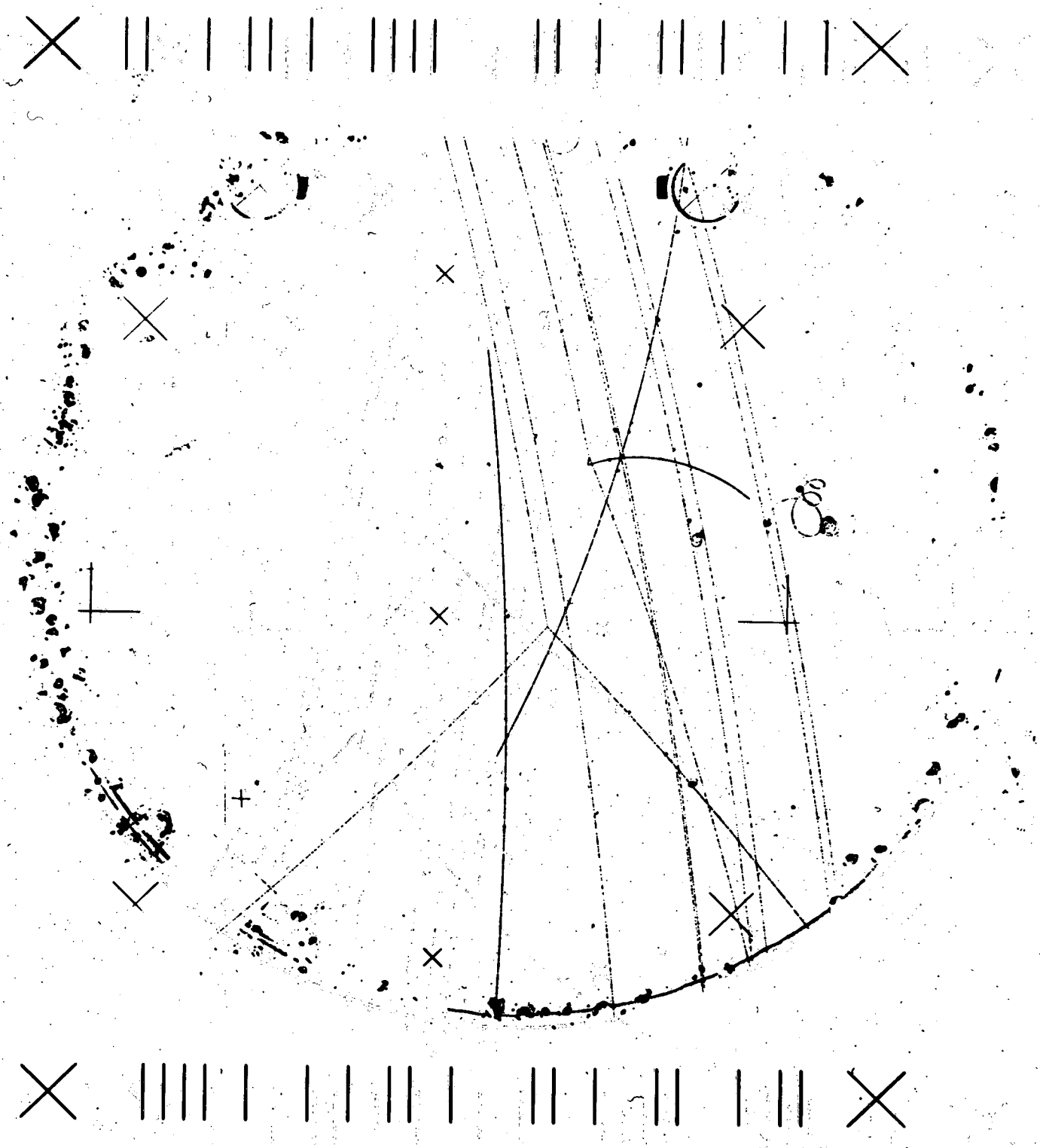


XBB 702-911



XBB 702-902

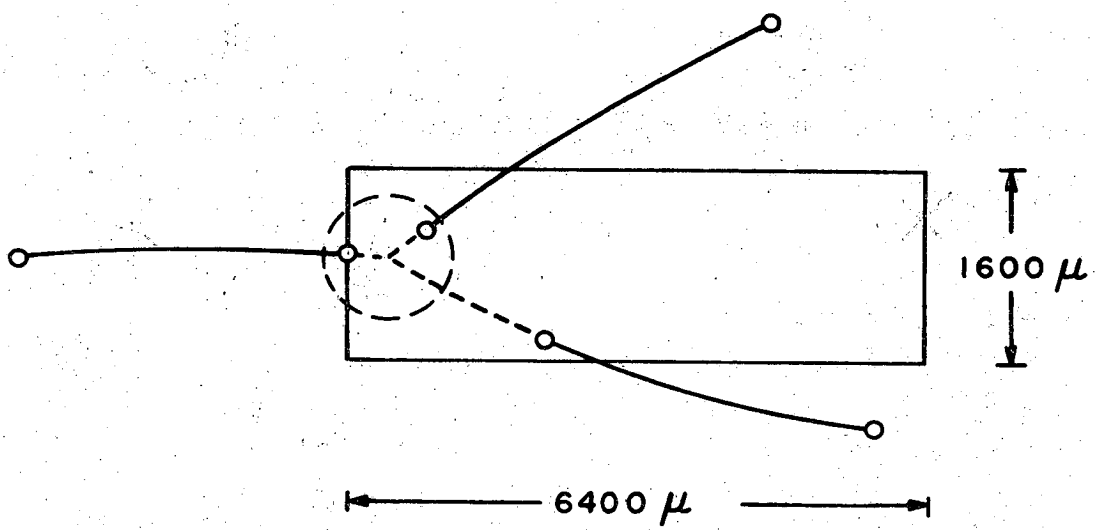
Fig. 4



XBB 702-906

Fig. 5

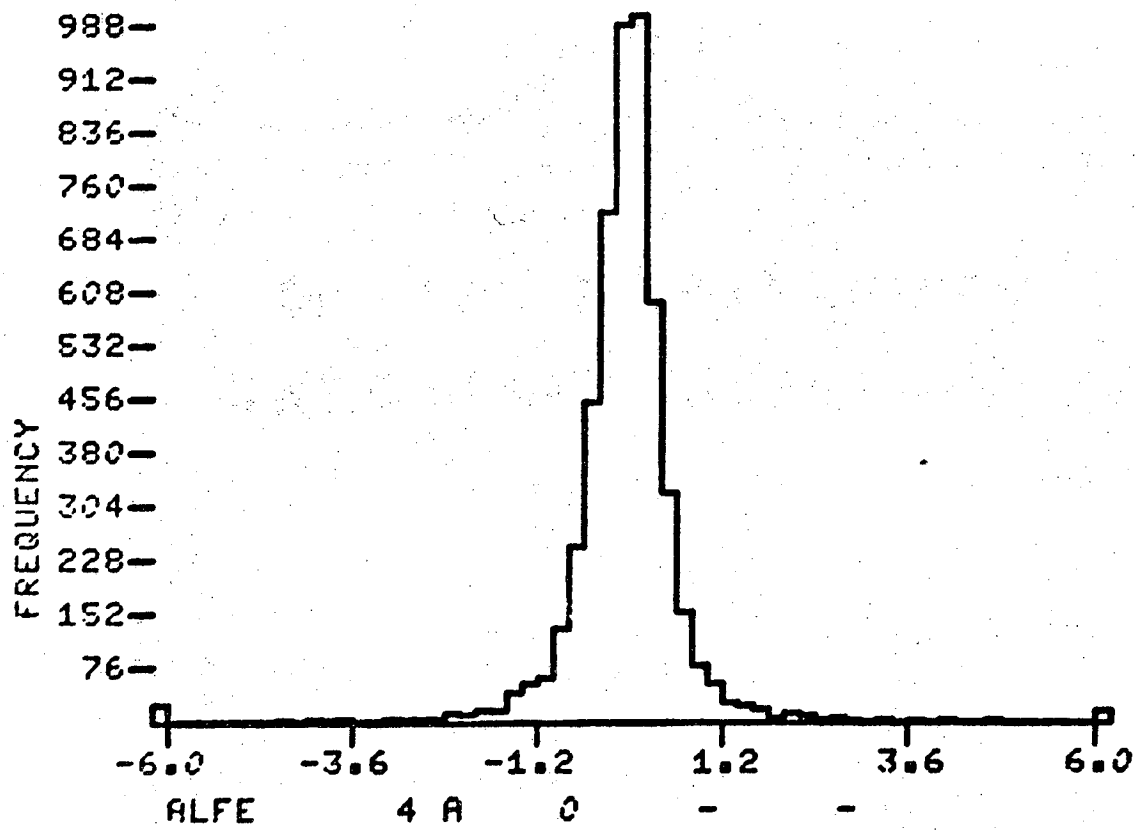
DAPR Vertex search



Primary search for view-verticies

XBL6810-6904

Fig. 8

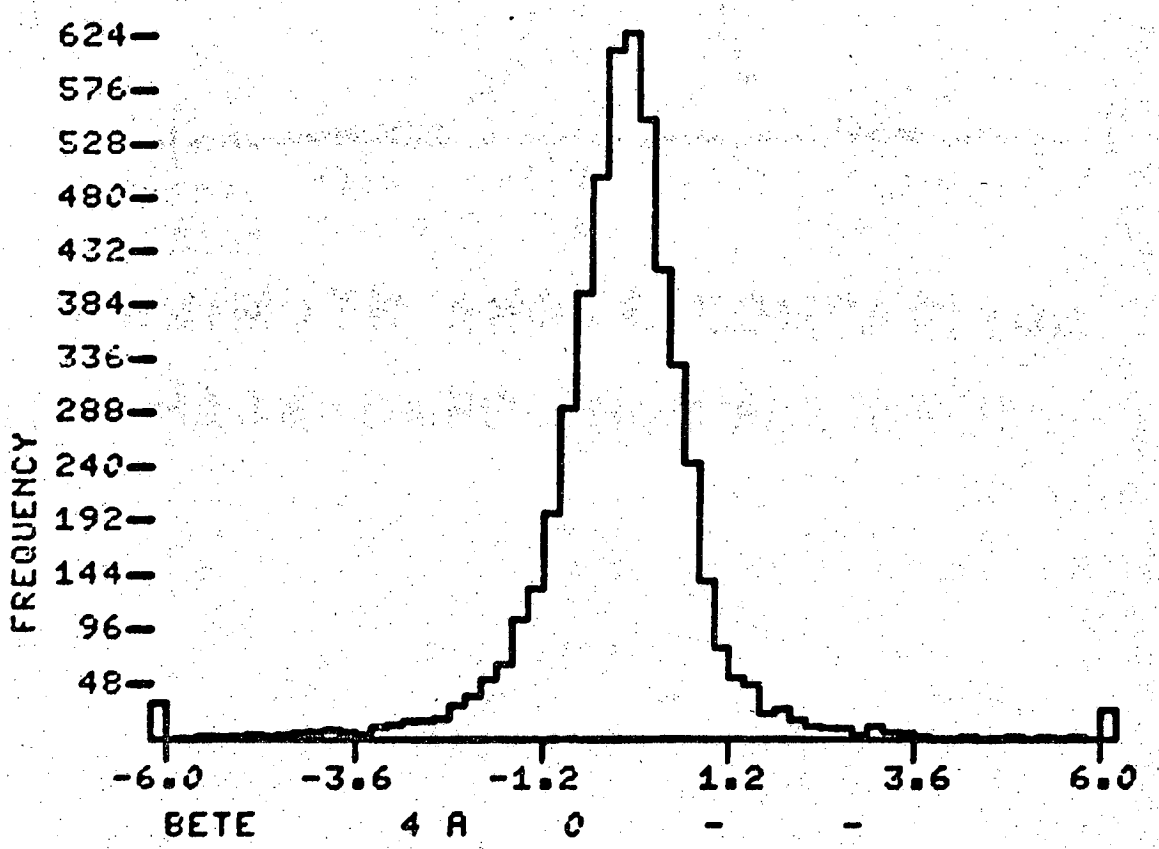


DIP ANGLE COMPARISON

HAZE-DAPR (NORMALIZED)

XBL 702-387

Fig. 9

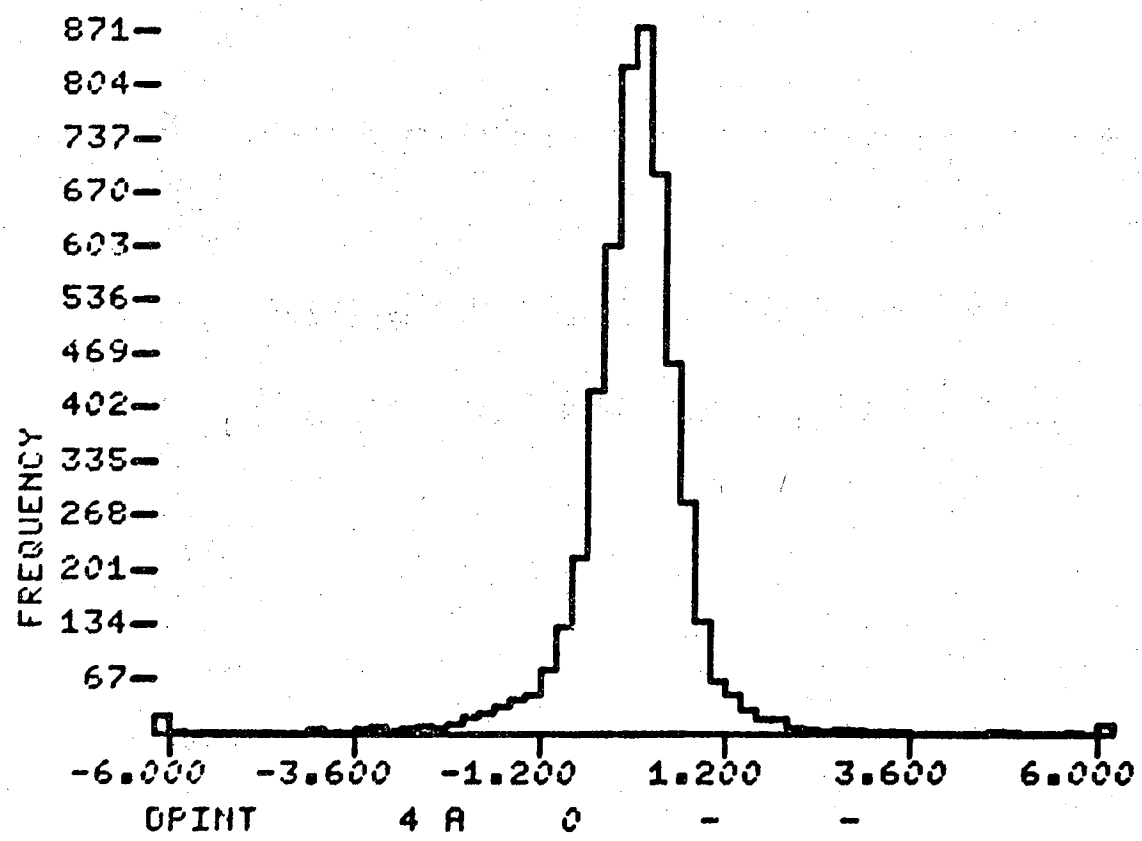


AZIMUTH ANGLE COMPARISON

HAZE-DAPR (NORMALIZED)

XBL 702-388

Fig. 10

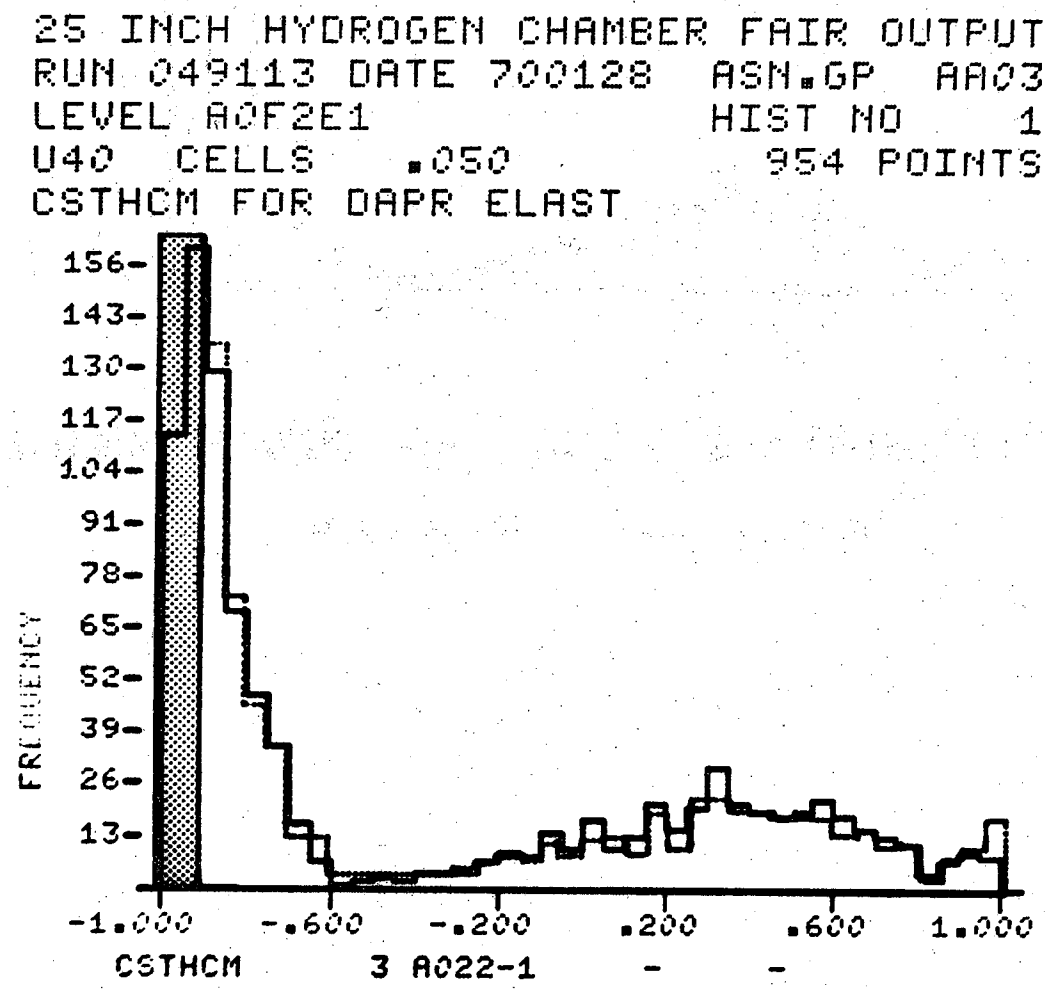


MOMENTUM COMPARISON

HAZE-DAPR (NORMALIZED)

XBL 702-389

Fig. 11



COMPARISON OF COS θ^* DISTRIBUTIONS

····· HAZE —— DAPR

XBL 702-396

Fig. 12

TECHNICAL INFORMATION DIVISION
LAWRENCE RADIATION LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720