UCRL-9231
Limited distribution

UNIVERSITY OF CALIFORNIA
Lawrence Radiation Laboratory
Berkeley, California

Contract No. W-7405-eng-48

MASTER

AUTOMATIC SCANNING AND MEASURING OF BUBBLE CHAMBER PHOTOGRAPHS

Antonio Grasselli

June 1, 1960

171 01

AUTOMATIC SCANNING AND MEASURING OF BUBBLE CHAMBER PHOTOGRAPHS

Contents

AUTOMATIC SCANNING AND MEASURING OF BUBBLE CHAMBER PHOTOGRAPHS

Antonio Grasselli

Lawrence Radiation Laboratory
University of California
Berkeley, California

June 1, 1960

ABSTRACT

The development of high-energy charged particle accelerators such as the Bevatron and of improved nuclear-event detection devices such as the Berkeley 72-in. hydrogen bubble chamber has greatly increased the need for high-speed data reduction of nuclear events. Full exploitation of the potential of the 72-in. bubble chamber demands a very high-speed analysis system. This paper describes an approach to such a system.

AUTOMATIC SCANNING AND MEASURING OF BUBBLE CHAMBER PHOTOGRAPHS *

Antonio Grasselli †

Lawrence Radiation Laboratory
University of California
Berkeley, California

June 1, 1960

I. INTRODUCTION

Bubble chamber nuclear events are analyzed by a stereo reconstruction of the trajectories of the particles involved in the interactions and a computation of momentum balance and energy balance. In this study we propose a high-speed automatic analysis system for bubble chamber data. We will refer specifically to the 72-in. hydrogen bubble chamber in operation at Berkeley, although our considerations are very general and could apply to other chambers as well.

Ionizing nuclear particles which pass through the superheated liquid hydrogen of the chamber leave in their wake a trail of bubbles. These join to form a track, thereby revealing the path of the particle. Charged particles are deflected into circular arcs by a strong magnetic field which is maintained in the chamber normal to the direction of the incoming beam from the accelerator. The bubbles exist in a definitive way for only a few milliseconds. The track is recorded by high-speed photography when the bubbles have just formed and are small and the tracks are still undistorted by chamber turbulence. Figure 1 is a photograph showing in the upper left corner, a four-prong star, a two-prong star, and a decay vertex.

* This work was done under the auspices of the U. S. Atomic Energy Commission.

† Present address: Department of Electrical Engineering, Princeton University, Princeton, N. J.

Stereoscopic pairs of photographs are required to reconstruct the event in three dimensions. The 72-in. bubble chamber uses three views (stereo-triad), allowing two views with optimum stereo angle for each track. Currently, one stereo triad is photographed every 12 sec. Each triad has a data-box area defining the triad numbers, roll number, and chamber operating parameters. After exposure, film from the camera is taken on reels to an automatic film processor, where it is processed, dried, and rewound on reels. It is then presented to the scanning staff.

A. Description of the Present Analysis System

Although a complete description of the present system can be found elsewhere,[1] we will try to summarize the various phases, primarily to present data rates.

1. Scanning

In this first phase, the film is examined by scanning technicians, who record the frames where events are found, and try to specify their types. The scanner makes only a few, limited stereo decisions by comparison of two views (distinguishing, for example, interaction vertices from pseudo stars). He must identify all vertices in a stereo triad and all vertices sharing a common prong.

2. Sketching

The interactions found by scanning are re-examined by physicists or highly trained members of the scanning staff. In this step, the event is classified topologically and a sketch is drawn on a "sketch card", with the tracks numbered from each vertex. The original classification schemes were kinematic; in other words, the sketcher guessed the masses of all interacting particles. This approach was a carry-over from doing both scanning and analysis right at the scanning table. As measurement became

a more intricate and separate function, the sketcher came gradually to classify interactions by topological type only; that is, to specify the number of prongs of the vertex and to identify and list prongs shared by two vertices.

Today, the operator normally attempts no kinematic interpretations during either scanning or sketching, for this function is being absorbed by programs.

Scanning and sketching are currently done at semi-automatic projection tables (Fig. 2). The three views of the stereo-triad are projected through three lens systems onto a table to show the film images two-thirds life size. A single view is scanned for events; when one is found, two views are superimposed in order to choose the two views out of three that show most clearly a particular track's vertical displacement in the bubble chamber.

3. Measuring

Tracks are measured at a semi-automatic projection microscope called "Franckenstein" (Fig. 3). This device is equipped with a track servo following system. The operator sees on a glass screen an image about twice the chamber size. To measure, he centers a crosshair pattern on the track, with the aid of a visual display (electrical signal on a cathode-ray tube). He then turns on the servo system; a foot pedal permits him to drive the crosshair along the track. During the travel, the operator depresses a button to record x and y digital coordinates on paper tape or IBM cards. The projection-stage position is digitized by rotary analog-to-digital converters to a minimum count of about 1 micron on the film.

4. Computing

Two IBM 704 programs are currently in use (approximately 18,000 words). The track reconstruction program (PANG) calculates momentum and space angles for each track. With this information, another program (KICK) fits together the kinematics of the whole event, using hypotheses consonant with the physics of the experiment. Both programs utilize the χ^2 method of analysis; that is, at each stage all plausible hypotheses are examined, and the one is chosen that gives a minimal deviation from the ideal "fit" of all variables. In each case, all the hypotheses tried are recorded along with the one giving the best fit.

The time taken by the various phases can be summarized as follows:[2]

scanning (1 event/10 frames)	30 min/event
sketching	15 min/event
measuring	10 min/event
computing (PANG + KICK on IBM 704)	1 min/event.

At present, 20,000 events per year can be processed.

Improvements to the present analysis systems, and increases in the equipment and in the scanning and measuring staffs will permit the examination of larger number of events. However, it does not seem possible by a mere expansion of the present semi-automatic analyzing system to exploit the full potentialities of the chamber; if non-strange-particle interactions are examined, the problem is that of analyzing at least 10^6 events per year. A system is needed that can cope with the data production rate, and this means that a stereo-triad has to be analyzed in a little more than 10 sec.

B. Description of Bubble Chamber Photographs

1. Film

The film is 47 mm wide; the dimensions of the chamber image are 120 by 33 mm (Fig. 4). There is, therefore, a demagnification of approximately 15 times. The tracks appear as dark lines on a clear film base.

2. Number of Beam Tracks per Picture

Two extreme cases are presented in Fig. 5a (1.6-Bev/c $\bar{p}$ beam) and b(3-Bev/c π-meson beam). The number of beam tracks varies with the experiment but is ordinarily between 20 and 30. It is limited because dense pictures are hard to scan, and more difficult (sometimes even impossible) to interpret. The dependence of human scanning time on beam density is plotted in Fig. 6.

3. Number of Interactions and Topological Types

The number of interactions depends on the kind of beam particle and on the energy. For example, π-mesons have a total cross section of 30 mb at 1 Bev/c and therefore 30/167 = 1/6 of the incoming particles will interact in the 72-in. hydrogen bubble chamber to give two-prong stars (about 95%), zero-prong stars (about 5%), and four-prong stars (almost negligible number). Antiprotons, on the other end, will interact about half of the time, producing two-prong stars (60%), zero-prong stars (10%), four-prong stars (25%), and six-prong stars (5%). Each of the prongs may decay in the chamber; furthermore, V decays may be associated with the star.

4. Track Width

Track widths may vary from 20 to 50 microns (on the film). The width depends upon:

(a) bubble diameter. In turn, this is related primarily to chamber operating parameters and to the time at which the lights are flashed.

(b) the number of bubbles per unit length of track. In very close bubble packing, the transversal pulse shape is wider than in gappy tracks.

5. Number of Bubbles per Unit Length

The bubble density along the track depends upon the particle and its velocity. For example, the following two sets of measurements (dimensions on the film) were taken on a limited sample of 1.65-Bev/c π-meson tracks:

Set	Average gap length (μ)	Average blob length (μ)
1	58.35	315
2	71.69	176.28

6. Equations for Projected Tracks

The path described from a particle would theoretically lie on a helix. The plane of the film is almost normal to the axis of the helix. The track images on the film are, however, not circular, because of

(a) nonuniformity of magnetic field,

(b) optical distortions,

(c) energy losses,

and

(d) turbulence in the chamber.

In PANG, the set of points to be fitted is corrected for (a) and (b); then an assumption is made about the identity of the particle, and energy losses are taken into account. The degree of the fit is 4 in the horizontal projection and 3 in the vertical.

7. Noise

From the point of view of automatic scanning, various sources of noise can be distinguished on or near the track under examination:

(a) background

(i) chamber noise: frost (A, Fig. 7), "coathangers" (light reflectors on the bottom of the chamber), fiducial rakes (B, Fig. 7) thermocouple (C, Fig. 7) flare spots, randomly condensed bubbles.

(ii) film defects: scratches (D, Fig. 7) chemical smears, dust and dirt, splices, and emulsion defects.

(iii) crossing tracks and electron spirals (E, Fig. 7, and Fig. 8).

(b) track noise

(i) gaps

(ii) turbulence

(iii) multiple-Coulomb scattering.

II. AUTOMATIC SCANNING AND MEASURING

A. Principles

The underlying principle of the method proposed here is to have a completely automated system from scanning to computing. Attention is focused here on automatic scanning, and only its implementation is discussed because (a) The automation of scanning requires mechanizing a pattern-recognition function. The artificial intelligence of computers usually compares poorly with human intelligence when applied to this task; the problem is therefore highly challenging. (b) The techniques used in automatic scanning functions can be extended for measurements. The same device can be used; one need only increase the resolution in the zones in which measurements are taken. (c) It is expected that the experience gained in the study of scanning will pave the way to the implementation of measuring. The method uses a high-resolution, random-access, flying-spot scanner controlled by a digital computer. We will break down the analysis into four phases:

(1) a recognition (or filtering) phase, in which significant information is extracted from the picture. The end product is a library consisting of a track file and a vertex list.

(2) an interpretive phase, in which the library is interpreted and measuring orders are issued.

(3) a measuring phase

(4) a computing phase (PANG and KICK).

In this paper we will be mainly concerned with the recognition phase. However, in the final operating version, phases (1), (2), and (4) should be very intimately related and linked together into a unified analysis program.

The fundamental principles guiding our study have been:

(a) The system should be a completely automated, with the possibility, but not necessity, of human monitoring.

(b) Insofar as efficient, digital stored program characteristics should be implemented to provide full flexibility in eventual changes in the operating procedure.

(c) Feedback should be possible at every stage of the process. The entire program should be written as a generator assembling a set of subroutines (examples: a) detailed scan of a knockon proton on the basis of a kinematic fit; b) resolution of scanning ambiguities by stereo-decisions)

(d) There should be a very high recognition efficiency. Even 1% failures represent a number of events too large to be rescanned by humans.

(e) Elementary operations should be highly reduntant, e.g., there should be multiple possibilities to pick up any one track in the picture.

(f) Critical decisions should be delayed to the moment when all pertinent facts are known.

(g) Logical operations should be used as much as possible instead of arithmetic, to decrease running time.

(h) Bookkeeping should be simple and flexible.

B. The Flying-Spot Scanner

A cathode-ray-oscilloscope (CRT) tube provides a bright positionable light spot. Currently available commercial tubes can produce 2000-line elements in a square array (DuMont K-1725). The phosphor most suited for high-speed scanning is P-16 (maximum light energy wavelength at 3700 A).

The spot is focused and deflected magnetically. Dynamic focusing, modification of focus current with spot location, and yokes that minimize pin-cushion distortion must be used. The phosphor efficiency of these tubes

is not uniform. A light-intensity monitoring servo loop varies the control-grid potential to try to maintain constant illumination.

Light from the face of the cathode-ray tube is focused on the film by an objective lens (Fig. 9). Part of this light is diverted by a beam-splitting mirror to the spot-location digitizing system. The digitizing system in a simple form can be a vertical and a horizontal picket-fence grating upon which the spot is imaged. Photomultipliers behind these gratings detect the light modulation as the spot is moved vertically and horizontally. Reversible scalers keep account of the spot location; the contents of the scalers can be compared to the desired digital address, forming a location error indication. This digital measure of the spot distance from the desired location is converted to currents applied to the deflection yokes.

The light passing through the film is modulated by the variations in density of the film. A photomultiplier with its condensing-lens system develops a signal whose amplitude describes the density of the film in the area of the spot. By addressing the location of the spot, that location may be sampled for the presence of an exposed area or the unexposed background.

A nonambiguous spot-positioning system can be substituted for the picket fence. Such a system should be similar to the one developed at the Bell Laboratories for the Telephone System flying-spot store.

The three views of the chamber forming one stereo triad are separated on the film from each other by several other frames. If a flying-spot system to digitize the three views together were to be made, it appears difficult to position the views physically side by side. A proposed solution is to have three separate flying-spot sources and three pickups looking at the film at the three view positions (Fig. 10).

The limiting factors in the engineering design of a flying-spot scanner store are mainly:

1. Spot size and spot energy.

The tube itself is a limitation because of phosphor grain size and the light dispersion on the glass. A high-beam intensity can burn the phosphor coating because of the poor thermoconductivity of glass.

2. Spot speed.

This factor is limited by the power requirement in the deflection system, phosphor decay time, transit time through the system, and servo- and location-comparators design.

3. Digitizing of spot location.

Accurate gratings and beam-splitting design are essential.

4. Optics.

The objective optics has to be a well-designed, diffraction-limited lens.

It is now possible to make a tube that will produce a 2000- by 2000-element array with a local access time of 1 Mc. Digital flying-spot scanners with twice this resolution seem within the limits of good engineering practice in the next few years.

C. Initial Specifications

We believe that 30 microns on the film is a reasonable resolution for scanning purposes (roughly one-half to two-thirds of a track width). An example of track segments digitized at this resolution is shown in Fig. 11, together with the original pattern (the function of the black spots in Fig. 11(b) will be explained later).

The three views, when placed side by side (Fig. 4), form approximately a square. The resolution given above corresponds to a 4096×4096

array of points. For the sake of simplicity, we will assume that this array can be traced on one tube. The method proposed and the estimated running time, however, are valid even if engineering considerations suggest different arrangements (multiple scopes).

In the discussion, we assume that the random-access digital scope has the function of a fixed information memory appended to a high-speed digital computer. The random-access requirement derives from the very high data content of the pictures; at our resolution 16×10^6 bits are encoded by the scope. A sequential scanning tube would have to transfer the entire raster to a memory medium accessible to the computer. The difficulties here are both in the extremely large memory that should be available and in the time spent in loading this memory. It is therefore much better to use the film itself as the memory medium, which requires that digital positions on the scope be randomly accessible. When the X- and Y-address registers of the CRT are loaded with 12-bit X and Y geometrical coordinates, the device reads out the corresponding digital position on the film; the output is a 0 or a I depending on the light collected by the photomultiplier. It is reasonable to assume that information can be obtained (locally) at a 1-Mc rate.

As for the digital computer, we set the following tentative specifications on which our running time estimates are based:

(a) addition time: less than 1 μsec; multiplication time: approximately 4 μsec

(b) word length: 64-bits

(c) memory requirements: core memory with a 2-μsec cycle time. A large capacity is desirable (64,000 words) but not indispensable if high-speed data exchanges with back-up drums are provided.

(d) some special-purpose logical features, as discussed in one of the following paragraphs.

Note that a computer of these general characteristics is well within the current status of the art.

The basic operating procedure can be sketched as follows:

(a) The picture is searched for track candidates on a local basis.

(b) Every time a track candidate is found, a tracking mechanism explores the picture, identifies the track, and files the corresponding information in a track library.

Extensive bookkeeping is connected with (a); namely, for any search pattern, a method has to be implemented to prevent points of a track already on file from triggering the tracking mechanism again. If the picture could be stored on a changeable memory medium, a successful identification of a track could be followed by an erasure of all its points. However, we said that the very-high data content of a picture prevents its storage in a computer memory, but the film itself has to be chosen as a memory medium. Moreover, it would not be desirable to destroy permanently the information about tracks, because this would make any "second look" impossible, therefore violating our requirement on feedbacks at every stage of the process. Checking every candidate with the tracks on file is very time-consuming. The method proposed is to store in memory an image of the picture, and to modify the image as part of the bookkeeping. Because the search for track candidates involves only a sampled fraction of the total information available, this image can correspondingly be a sampled copy of the picture.

D. Sampling Grid and Transcript

1. Sampling Grid

The sampling grid on which track candidates is searched is sketched in Fig. 12. The examinations are made at every fourth point in the y direction, and at every second point in the x direction. The ratio 2:1 between sampling intervals is chosen because most of the tracks (all the beam tracks) are in the y direction. Only points on this grid can activate the searches for tracks. The identification of an initial track segment and the tracking are conducted on the scope, however, utilizing therefore the full information content of the picture.

It may be noted that only a fraction of the track points appear on the sampling grid. Every one of them, however, can initiate the identification, and high redundance is therefore still present. Tracks exactly parallel to one of the axes will very likely have few points on the grid because of scattering in the digitizing. In Fig. 11(b) the black spots are the points appearing on the sampling grid. Besides reducing considerably the problems involved in bookkeeping, a search on the sampling grid has the positive effect of eliminating a substantial amount of local noise.

2. Transcript

We will define "transcript" as an image of the picture, loaded initially into an erasable memory by interrogating the scope with the sampling grid. The transcript is continuously updated according to the result of the identifications. The transcript has

$(1024) \times (2048) = 2 \times 10^6$ bits which corresponds to

$(2 \times 10^6)/64 = 32{,}768$ core-storage positions.

Sampling the transcript with a 1-Mc CRT takes approximately 2 sec; in the data exchange within the CRT and core memory, a word is

stored in the memory every 64 μsec. This allows another program to run concurrently, which can be interrupted when needed. Total memory load time is 70 msec. If less memory space is available, only the transcript corresponding to one view can be kept in the cores at any one time.

3. Search on transcript

Every bit on the transcript has to be examined once, but the search pattern is entirely arbitrary. However, to minimize memory references, every transcript word should be extracted only once from memory. The time for memory calls in this operation is slightly more than the memory load time given above, or approximately 100 msec.

The transcript can be organized as in Fig. 13a or 13b. The second configuration minimizes memory calls in erasing.

E. Initialization

Every time a bit is found in the transcript, an initialization is started at the corresponding digital position on the scope. Initialization discriminates between track segments and local noise, and affords a rough initial orientation of the track (if any) under examination. This operation can be performed in many ways, but basically any method can be classified in one of two categories: (1) matching against a known pattern, or (2) local tracking.

1. Matching Against Known Patterns

An ($n \times n$) grid centered around the activation point is examined with m library patterns, and the number of agreements (disagreements) for every examination is counted. The track segment is identified with the pattern that gives the highest (smallest) count if this is above (below) a certain threshold. If no pattern in the set satisfies the threshold count, the candidate is rejected as noise.

This initialization method may be implemented in a number of ways. For example, the grid may be extracted into an (n×n)-bit register. An "exclusive or" operation is executed bit-by-bit between this register and every one of the masks in the library. After every matching the masks are counted; they could be stored in core memory and called sequentially, or contained in fast registers at the beginning of the operation. The masks could also be stored in circulating registers, with the grid compared in parallel with all the masks, and serially by bit, as the information is obtained from the CRT. Anyway, the basic limiting factor is the oscilloscope speed of 1 μsec/point so that the logical design of a special unit and(or) the programming of this operation are not particularly critical.

With a 7-×-7 grid, for example, the masks could be simply the four in Fig. 14. The resolution obtained is higher than four orientation, since if a pattern matches two masks above the threshold, it may be deduced that the track segment lies between the two.

Note that in the particular scheme proposed in Fig. 14, the width of the black portion of the library pattern is three counts. Since one count corresponds to one-half to two-thirds of a track width, disagreements in the black portion have to be expected, and must therefore be weighted differently. With only four masks, however, a number of simple logical schemes can be devised to take this complication into account.

If we assume 3000 initializations per view, and with 100 μsec per operation, the total initialization time is

$$(3000)\times(3)\times(100\times10^{6}) = 0.9 \text{ sec.}$$

2. Local Tracking

The spot can be programmed to follow an initial segment of the track by examining successively neighboring points. After the first few points have been obtained by looking rather blindly around the activation spot (see steps 1, 2, 3, in Fig. 15), the rough direction obtained is enough to reduce the examinations required to a few spots per step. Note that even if very simple decision rules are given, this method is intrinsically more complicated than matching against known patterns. However, a nice feature is that local tracking is not restricted to a segment of fixed length, as in the masking method, but enough information can always be acquired to define appropriately the track segment. As in the method of matching against known patterns, the examinations are positive if there is a noise-free zone at either side of the track.

Local tracking will stop when:

(a) m pulses have been observed. A straight line through the activation point and the last observation is then the simplest guess about the track direction, and the tracking proper takes over from here.

(b) a distance d from the activation point has been traveled, with less than m points observed. In this case, the candidate is rejected.

(c) a gap longer than g steps has been encountered. Again, rejection occurs.

3. Combination of the Previous Methods

A combination of the methods described could also be implemented, for example, by tracking locally with a small mask. Any one of these methods can be made to work if the parameters are carefully chosen. Masking has been successfully tried on dynamic randomly chosen examples.

It is important to note that the value of any such technique lies both in its discrimination power and implementation complexity, which is in most cases synonymous with running time. Complicated indexing of the scope and access to very large grids will therefore be avoided.

F. Tracking Phase

1. Tracking algorithm (Fig. 16)

The tracking algorithm used is essentially the same as in radar tracking.[3] At ± 45 deg to the y axis, the predicted position variable will be x, and at ± 45 deg to the x axis, the tracking will be in y.

For tracking in x, the tracking equations are

$$x^e_{n+1} = x_n + s^x_n \cdot (y_{n+1} - y_n) \qquad \text{(extrapolation)}, \qquad (1)$$

$$x_{n+1} = x^e_{n+1} + \alpha(x^*_{n+1} - x^e_{n+1}) \qquad \text{(correction)}, \qquad (2)$$

and

$$s^x_{n+1} = s^x_n + \frac{\beta}{y_{n+1} - y_n}(x^*_{n+1} - x^e_{n+1}) + \gamma(y_{n+2} - y_{n+1}) \qquad \text{(slope)} \qquad (3)$$

where

x^e_n is the extrapolated position at step n,

x^*_n is the observed position,

x_n is the corrected position,

s^x_n is the slope of the chord between x_{n+1} and x_n,

α and β are tracking parameters, and

γ is the slope correction factor.

The influence of α and β on convergence and smoothing have been computed for straight-line, constant-velocity tracking.[3] This corresponds in our case to straight-line tracking, since the velocity of the projectile is the slope of the track.

For a straight line, γ equals zero; if we use a constant extrapolation interval $dy = I$, we can rewrite the equations [by substituting Eq. (1) into Eq. (2)] as:

$$x^e_{n+1} = x^e_n + \alpha (x^*_n - x^e_n) + s_n \tag{4}$$

$$s_{n+1} = s_n + \beta (x^*_{n+1} - x^e_{n+1}) \tag{5}$$

If the system is noiseless (observed position = actual position), we obtain by solving the system:

$$x^e_n = zn - \epsilon_x + \frac{\lambda_1^n - \lambda_2^n}{\lambda_1 - \lambda_2} \left(\frac{\epsilon_x}{2} (\beta - \alpha) + \epsilon_s \right) + \frac{\epsilon_x}{2} (\lambda_1^n + \lambda_2^n) \tag{6}$$

$$s_n = z + \frac{1}{2(\lambda_1 - \lambda_2)} \left[(1-\alpha)(2\epsilon_s - \beta\epsilon_x) + z(2-\beta)(\lambda_1^n - \lambda_2^n) \right.$$

$$\left. - 2(\epsilon_s + z)(1-\alpha)(\lambda_1^{n-1} - \lambda_2^{n-1}) \right] + \frac{z}{2} (\lambda_1^n + \lambda_2^n) \tag{7}$$

We assume as initial conditions

$$x^e_0 = x_0 + \epsilon_x = 0$$

and

$$s_0 = z + \epsilon_s ,$$

where z is the actual slope, and

$$\lambda_1, \lambda_2 = 1 - \frac{\alpha + \beta}{2} \pm \frac{1}{2} \left[(\beta + \alpha)^2 - 4\beta \right]^{1/2}$$

The errors in position and slope are expressed respectively by the transient parts of Eqs. (6) and (7).

The type of convergence is overdamped for values of $(\beta + \alpha)^2 - 4\beta$ greater than zero, underdamped for values less than zero, and critically damped for zero. In order to have position smoothing if noise is present, α and β must satisfy the following conditions, for β less than 0.35:

$$\frac{2 - \beta}{4} - \frac{1}{4} (4 - 12\beta + \beta^2)^{1/2} < \alpha < \frac{2 - \beta}{4} + \frac{1}{4} (4 - 12\beta + \beta^2)^{1/2} .$$

For any given β, the value of α giving the optimum smoothing is

$$\alpha = -\frac{\beta}{2} + \sqrt{\beta} .$$

Figure 17 shows the relationship between α and β and the type of convergence. In the shaded region, smoothing occurs.

2. Choice of α and β

Further study is needed to define the influence of the values of α and β in our case. It is desirable to start the first tracking steps with fairly large parameters ($\alpha = \frac{1}{2}$, $\beta = \frac{1}{2}$), and gradually reduce them to values corresponding to optimum (or near-optimum) smoothing. Since a rigorous mathematical analysis of satisfactory variation laws appears to be cumbersome, a computer simulation on digitized pictures is suggested.

In the practical implementation, α and β can be found at every step by a table look-up operation.

A possible variation law could be:

$$\alpha = F(p, g)$$

$$\beta = G(p, g),$$

in which p is the total number of positive observation, and g is the total number of negative observations (gaps). To weight the recent information more than the past, p and g could be multiplied by a decaying factor less than one at every step. Another possibility is to have α and β depend on p, g, and a parameter k which measures the success of tracking. For example, we may define k as follows:

$$k_n = \frac{1}{n}\sqrt{\sum_{i=0}^{n} (x_i^* - x_i^e)^2}.$$

3. Computation of γ

Gamma could be computed from two successive values of the slope. However, since the curvature of most tracks is small, the values of γ would have to be sufficiently accurate to retain many significant places in the slope values. It is simpler to compute (or correct) only a few times during the tracking by taking three points P_i, P_j, and P_k reasonably far apart and calculating the slopes of the chords P_iP_j and P_jP_k.

4. Implementation of tracking

A tracking step can be broken down into the following phases:

(a) computation of the extrapolated position

(b) scope observation

(c) correction of the position

(d) computation of the slope

(e) bookkeeping.

In the tracking equations, no multiplications are required if:

$$dy = 2^i ; \alpha = 2^{-j}; \beta = 2^{-k}.$$

(where i, j, and k are positive integers) because multiplications can be substituted by shifts. There is no particular reason why α, β and dy should not range only over powers of two. We estimate therefore that the computational part of tracking (phases a, b, and d) will take approximately 20 μsec per step.

The scope observations are taken with a grid as in Fig. 18 a. Therefore, observations are classified as:

(a) normal (Fig. 18b).

(b) noisy. The pulse is abnormally large (Fig. 18c).

(c) gap. No pulse is detected in the normal range, nor in the guard bands (Fig 18d).

(d) out-of-range. The pulse lies in the guard bands (Fig. 18e). These observations will be reclassified either as normal or gaps by examining both interpretations in the succeeding tracking steps.

Note that the values of a and b can be modified during tracking. The spot control can be particularly efficient if scope indexing is done with a compact control word, as explained in one of the next paragraphs. For 10 points, at 1 μsec per point, an observation step will take 10 μsec.

The bookkeeping comprehends:

(a) table look-up of α, β, dy

(b) eventual modification of spot-control word

(c) computation of readjustment of γ (rather infrequent operation)

(d) filing of track markers (relatively few points on the track which will be used to compute the track equation).

We will assign 15 μsec (on the average) to the bookkeeping phases. For 40 tracks, three views, and an average of 400 tracking steps per track, the total tracking time will therefore be

$(40) \times (3) \times (400) \times (45 \times 10^{-6}) = 2$ sec.

Two examples, illustrating the initial choice of α, β, and γ are presented in Fig. 19.

G. Fitting, erasing, and track-file maintenance

1. Fitting of Track Equation

The markers are fitted by a least-squares method to an appropriate polynomial, whose coefficient is then filed in a track library. The order of fit should be chosen as low as possible to avoid changes in the sign of the curvature. Beam tracks are fairly well known from beam design, and they are therefore relatively easy to fit. Secondary tracks require more attention, especially at low energies. However, at the accuracy required for our purposes, the problem is a classical exercise in smoothing, and no intrinsic difficulties need be expected. In the most delicate cases, a track can be divided in sections, and fitted locally.

The number of operations involved in a fit can be easily computed, because the problem is one of solving a linear system,

$$A Z = B \qquad (8)$$

If N is the number of points used in the fit and n is the order, the computation of the coefficients of A requires $N \times (n+6)$ additions and $N \times (n+3)$ multiplications. Solving Eq. (8) for the unknowns requires $n^3/3$ additions and $n^3/3$ multiplications. For N = 30 and n = 4, 5 msec seems a reasonable time estimate for a fit and related bookkeeping.

For 40 tracks, and three views, we then have

$$(40) \times (3) \times (5 \times 10^{-3}) = 0.6 \text{ sec.}$$

2. Erasing the Track on the Transcript

It is possible to erase the track concurrently with tracking, because

the transcript has the only function of initiating a search, and therefore erasing it does not prevent a "second look" at the track (if needed) An advantage of this method is that practically no computation is involved, because tracks can be considered straight line segments on a local basis.

Another possibility is to compute relatively few points from the track equation, and to interpolate linearly between them. The computation of a point requires n multiplications and n additions, where n is the degree of the polynomial. Linear interpolation is very fast (1 addition + 1 shift per point). Because the tracks are reasonably linear over 20 counts, the maximum number of points to be computed from the track equation is 200 per track. At 25 μsec per point, 40 tracks, and three views, we have

$$(40) \times (3) \times (200) \times (25 \times 10^{-6}) = 0.6 \text{ sec}$$

The number of memory calls depends on the organization of the transcript. In the worst case, a maximum of 1024 words per track have to be retrieved and stored, and therefore the time for memory calls is

$$(1024) \times (3) \times (40) \times (4 \times 10^{-6}) = 0.5 \text{ sec}$$

Because some of this time can be overlapped with the linear interpolation and erasing, we will count only 0.25 sec for these two operations. The total erasing time for the second method is therefore 1.35 sec, with only about 0.9 sec for the first method. However, we feel that the second technique will give more precise results without increasing violently the running time.

To assure that every pertinent bit will be wiped out, the erasing must be done with a guard band around the computed points. Some points on other tracks will therefore be erased too However, the high redundancy will allowe the track to be picked up somewhere else, and points erased by accident are in the vicinty of the track being erased; therefore an

initialization would be difficult or unsuccessful.

Theoretically, it is only necessary to erase the track in that part of the transcript which has not yet been searched for track candidates. However, if the entire transcript is kept up-to-date, it represents, at every program step, a clear picture of the recognition work yet to be done. The transcript could, for example, be displayed on a scope following a console interrogation, as an immediate visual aid to interpret an eventual program failure.

3. Track File and Vertex List

For simplicity we suppose that a record in the track library contains only the track equation and the coordinates of the first and last point of the track segment (end-markers) When the last transcript item has been examined, it is assumed that all the tracks in the picture are on file. A survey of the track file is now necessary, for example, to eliminate multiple recognitions (not excluded, because they may have been triggered by an imperfect transcript erasure) The tracks are, at this point, still unrelated. In preparation for the interpretive phase, all the vertices must now be found. The patterns to be identified are m-prong starts, zero-prong stars, and decay vertices. Theoretically, the interaction vertices can be found in the set of end markers by simple comparisons End markers, however, are not accurate, and therefore further analysis is needed. All the end markers are compared; whenever two or more are found to be adjacent, areas are rescanned locally, stereo-comparisons are made, and the intersection point of the respective tracks is computed. In case of a positive identification, a common marker is assigned, with a special tag.

The list of vertices are written as a connectivity matrix. Rows and columns of this matrix correspond to end markers and to vertices,

and the entries are track identification numbers connecting the row marker with the column marker.

We estimate 0.5 sec for the reordering of the file and the construction of the connectivity matrix.

H. Critical Analysis of the Method

1. Limiting factor

A summarizing flow diagram of the recognition phase is presented in Fig. 20. Estimates running-time for the individual subphases are:

a.	load transcript	2 sec
b.	search on transcript	0.1 sec
c.	initialization	0.9 sec
d.	tracking	2 sec
e.	fitting	0.6 sec
f.	erase transcript	1.35 sec
g.	reordering of track file	0.5 sec,

or a total of approximately 7.5 sec. It is emphasized that these estimates are only guesses; however, individually considered, no one of the estimates quoted seems unrealistic for a computer of the speed assumed throughout the paper.

The interpretation of the track file can probably be carried out in about 2 sec. This would mean less than 10 sec for the first two steps of the analysis.

It is interesting to analyze the factors limiting the processing speed. The CRT rate limits both the transcript loading and the initialization. In loading, there is a serious mismatch between memory and CRT speeds because the oscilloscope furnishes information at the frequency of 1 Mc/bit, and the memory can accept information at (0.5) × (64) = 32 Mc/bit. In the

initialization, a relatively large number of points (50 to 80) have to be obtained from the scope. The logical operations and relative bookkeeping are rather simple and, although additional to the scope time, do not increase appreciably the time taken by the initialization step. In tracking, computation and bookkeeping are responsible for a large fraction of the time allotted; few points have to be retrieved on the scope in this operation. Fitting depends strongly on computational speed, and so does erasing of the transcript. In this latter operation, however, about half the time is spent in memory references. Step g has no particular limiting factor.

2. Vertex-recognition approach

Another approach to recognition could be tried. Of all the line patterns appearing on the picture, we can consider as our target set the interaction vertices. A local pattern examination can be performed by having a "window" covering the entire picture step by step. At every step the question would be asked: "is there under the window a pattern belonging to the target set?" Because every vertex is a singularity point in the maze of tracks appearing on the picture, a set of rules can be given to probe the local area under the window as in classical methods devised for character recognition. Unfortunately, however, the pattern field is too noisy to conserve locally the singularities. Figure 21 is a fair example of the local aspect of a four-prong vertex; actually, the local pattern examined would be even worse because of the digitizing. If the examination area is enlarged, techniques essentially similar to those discussed in the track-recognition approach must be used.

Because all patterns in the target set occur at the intersection of smooth segments of tracks, it seems better to focus the recognition on the tracks themselves, rather than on vertices. These can be easily identified once the track history is known in the entire pattern field.

3. Interpretation

The interpretation of the library has the function of sequencing the measurements of the interactions. It is very close to the sketching phase in the present analysis system, in which events are classified topologically and prongs numbered by stereo comparison.

Note that in the present system, we have proposed to eliminate the sketching phase. Techniques have been suggested to dispense with prong numbering.[4] The sequencing of measurements is therefore simplified, and the burden of disentangling the event will rest with the computer program. In our case, no basic difficulties are expected. The task is not trivial, and its implementation lies in the realm of clever programming; however, the necessary information is available in a suitable form. This phase will have access to the recognition subroutines, so that additional information is readily available if needed to check hypothesis or to fill in holes in the event topology.

4. Measuring

The flying-spot scanner and associated computer constitute an ideal measuring device. If the same cathode-ray tube is used (by increasing the resolution), measuring follows scanning without any time delay, so that immediate feedback is possible. Even more important, in case of a failure in the computing phase, the picture can be rescanned and remeasured at once. Interrogation mechanisms can be devised to allow the physicist to guide the analysis program if needed.

The main difficulty here is achieving the required resolution of four times that suitable for recognition. Multiple scope arrangements or mechanical movements of the film have to be used. In the first case, an array of tubes could examine the picture. The spots would be defocused in

the scanning phase, and focused during measuring, when full resolution is needed. In the second solution, only one tube is required, and the optics used in measuring would look at one zone of the stereo tried. Access to the other zones would be provided by a fast-moving stage.

I. Track Recognition Computer

Because automatic scanning is a highly repetitive production operation, it is pertinent to ask whether a special-purpose or general-purpose design should be implemented. The usual arguments against special-purpose logic are perfectly valid here, because the inflexibility introduced would jeopardize any major change in the operating procedure. Moreover, a major part of the bubble-chamber data-processing task calls for scientific computation (PANG and KICK programs), and therefore a general purpose machine is needed anyway. However, we feel that it is worth investigating to determine what special features can be efficiently imbedded in a general-purpose structure to facilitate the track recognition. The list is not exhaustive, and it is presented here only to illustrate some of the functional requirements of a track-recognition computer.

1. Oscilloscope-format control

A grid format to be transferred from the scope to the internal memory is sketched in Fig. 22a. In this particular case, the information uniquely identifying the grid is, for example,

x_i , y_i : coordinates of lower left corner point

Δx , Δy: resolution of raster

n, m: size of the raster.

An alternative is to specify the grid's center point. Formats like Fig. 22b can be specified by identifying the basic grid (enclosed in solid lines), an offset δ (in x or y), and the number of repetitions of the grid. A control word would therefore have the general format:

$$[(x_1, y_1), (\Delta x, \Delta y), (n, m), \delta, c, r, b]$$

where δ is the offset, c is the offset control (x, y on normal mode), r is the number of repetitions of the basic grid, and b is conditional (if this bit is set, termination of transfer is controlled by a sense light). The information bits from the scope will be assembled in consecutive 64-bit memory words, and therefore the internal memory organization depends on the order in which the grid is scanned by the spot. Only an initial memory location is needed to specify the destination of the transfer.

2. Logical unit

Logical operations between words or parts of words are an integral part of the recognition process. A unit could be provided which has as its output the bit-by-bit binary connectives of two input words. It is certainly useful to include all the 16 logical operations. On the result word, stored in a "logical accumulator", the following operations can be required:

(a) count zeros/ones, load result in a special counter and test. This operation is needed in masking and matching against a threshold (initialization).

(b) detect first (last) one (zero), and load position in an index register. This operation is useful in tracking observations.

It is felt that a powerful logical unit could be very useful even outside the specific requirements of automatic scanning.

3. Geometric addressing

To assure maximum ease and efficiency in indexing a geometric addressing mode could be instrumented. When in this mode, a point can be reached from memory by specifying its (x, y) coordinates in the address field of the command. Of course, memory words have to bear a fixed correspondence with (x, y) points. This arrangement would make handling the transcript easier.

4. Short loops and look-ahead facilities

The possibility of executing short loops without reference to memory is almost a necessity. How much fast storage should be available for data and what degree of complication should be provided in look-ahead features is not possible to say without a more careful examination.

5. Macro-operation set

Optimization of program running time by appropriate choice of a macro-operation set can be done by analyzing for high-frequency trains or operators. However, long-range variations in the program can easily make such a set obsolete. The conflict between operation-code optimization and long-range flexibility can be resolved by adopting a structure of the type suggested elsewhere.[5] In this solution a modifiable control memory contains microsequences, each of which corresponds to a macro-operation. Because the microsequences can be changed, the machine's order code can be optimally or near-optimally assembled, even if the requirements are modified.

III. CONCLUSION

The realization of an automatic system as discussed in the paper will require a large cooperative effort, focused on hardward design and programming. Neither the final instrument nor an operating program can be written, however, without further research and experimentation. We would like to emphasize that programming--very often underestimated--is of paramount importance in this case. In parallel with the engineering design of a flying-spot device, simulation program should be written not only to demonstrate feasibility but to pin down hidden difficulties and to experiment with a wide variety of different techniques. The design of a suitable digital computer or the use of existing equipment should be carefully considered. In the second case, some special-purpose features have to be added to existing hardware.

It is our feeling that the entire analysis problem has to be faced with a unified approach. The effort required encompasses every aspect of computer design and application, and errors of judgment could easily jeopardize its success. Therefore, it should be very carefully planned and undertaken by a team of information-processing specialists.

The purpose of this paper is not to present details, but to expose general techniques. We hope that this discussion will stimulate creative effort aimed toward the implementation of an automatic analysis system. This effort is now lacking, although badly needed. Only Hough (University of Michigan) has done early work on automatic scanning [6] and automatic measuring. [7]

ACKNOWLEDGMENT

I wish to acknowledge the contribution of Dr. B. McCormick (now at the University of Illinois) to the formulation of this study. I am particularly indebted to Mr. Jerome A.G. Russell (Lawrence Radiation Laboratory), Dr. J. Snyder (University of Illinois) and Mr. W. Clark (Lincoln Laboratory) for enlightening discussions. Members of the Hydrogen Bubble Chamber Group (Lawrence Radiation Laboratory) have been most helpful, especially Drs. M. Alston, Hugh Bradner, F. Solmitz, and L. Stevenson, and Mr. A. Natapoff.

BIBLIOGRAPHY

1. Arthur Rosenfeld, " Digital Computer Analysis of Data From Hydrogen Bubble Chambers at Berkeley, " in Proc. International Conf. High-Energy Accelerators and Instrumentation, CERN, 1959 (CERN, Geneva, 1959) pp. 533-541.
2. Margaret Alston, System Study of Data Analysis for Associated Production Experiment, Lawrence Radiation Laboratory, Alvarez Group Memo No. 122, November 1959.
3. I.S. Reed and G.P. Dinneen, An Application of Matrix Theory To The Tracking and Smoothing Equations, MIT Lincoln Laboratory Technical Report No. 14, June 1953.
4. Margaret Alston, Lawrence Radiation Laboratory, private communication.
5. B.H. McCormick, A. Grasselli, and J.A. G. Russell, A Comprehensive Data Analysis System for the 72-Inch Hydrogen Bubble Chamber, Internal Report submitted to the AEC, Division of Research, Physics and Mathematics Branch, November 23, 1959.
6. P.V.C. Hough, "Machine Analysis of Bubble Chamber Pictures," in Proc. International Conf. High-Energy Accelerators and Instrumentation, CERN, 1959 (CERN, Geneva, 1959) pp. 554-556.
7. P.V.C. Hough, Department of Physics, University of Michigan, private communication, unpublished work in progress at CERN.

FIGURE LEGENDS

Fig. 1. Bubble chamber photograph.

Fig. 2. Scanning table.

Fig. 3. Measuring projector.

Fig. 4. (a) Chamber image on the film. (b) Stereo triad.

Fig. 5. Extreme cases of number of beam tracks per picture.

Fig. 6. Dependence of scanning time on beam.

Fig. 7. Background noise, showing (A) frost, (B) fiducial rakes, (C) thermocouple, (D) scratches, and (E) electron spirals.

Fig. 8. Microphotograph of electron spiral.

Fig. 9. Block diagram of flying-spot scanner.

Fig. 10. Block diagram of proposed three-view system.

Fig. 11. Digitized track segments.

Fig. 12. Sampling grid.

Fig. 13. Organization of transcript.

Fig. 14. Library patterns.

Fig. 15. Initialization by local tracking.

Fig. 16. Tracking algorithm.

Fig. 17. Influence of α and β on straight-line tracking.

Fig. 18. Classification of observations.

Fig. 19. Examples of tracking.

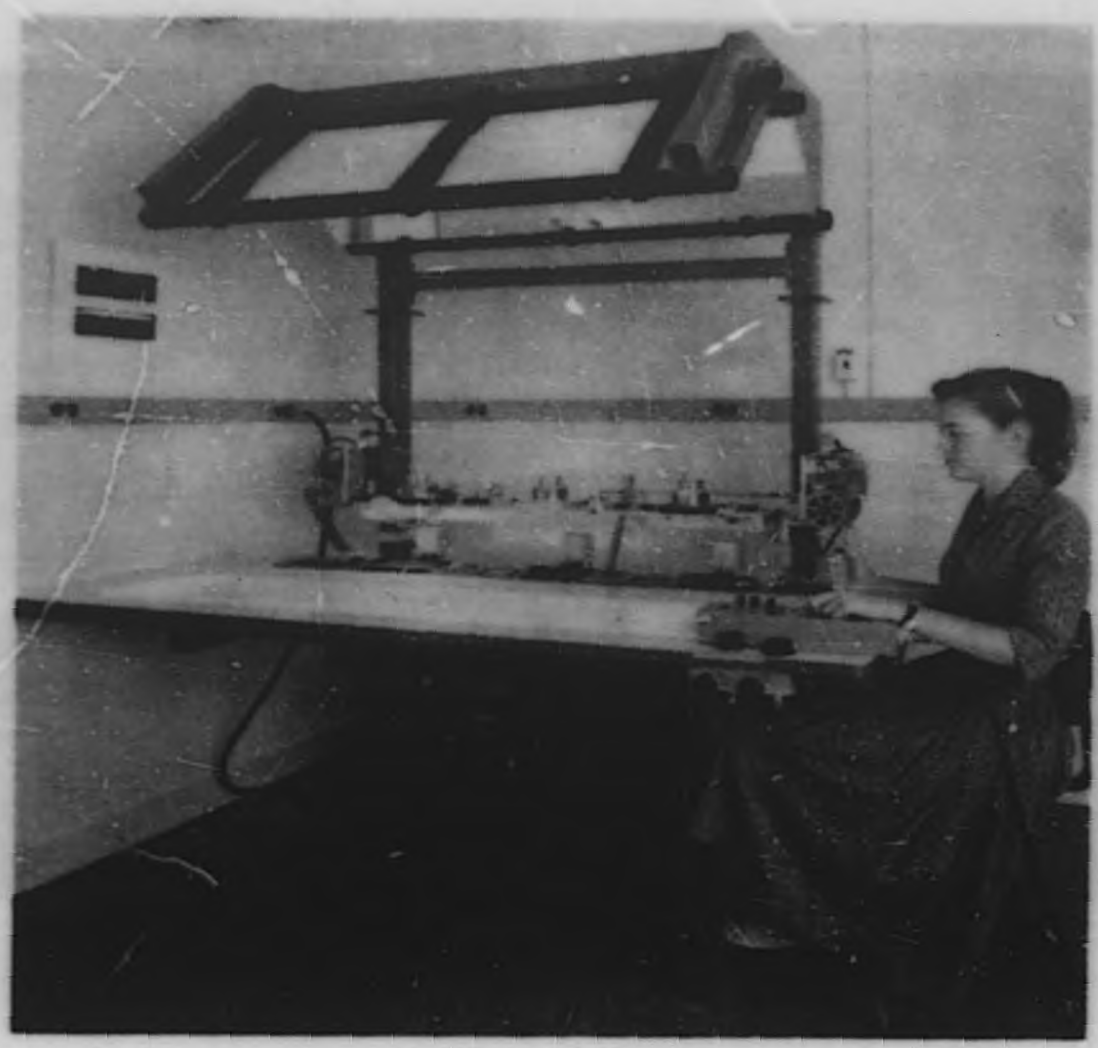
Fig. 20. Flow diagram of recognition phase.

Fig. 21. Local aspect of a four-prong vertex.

Fig. 22. Grid formats.

ZN-2671

Fig. 1.

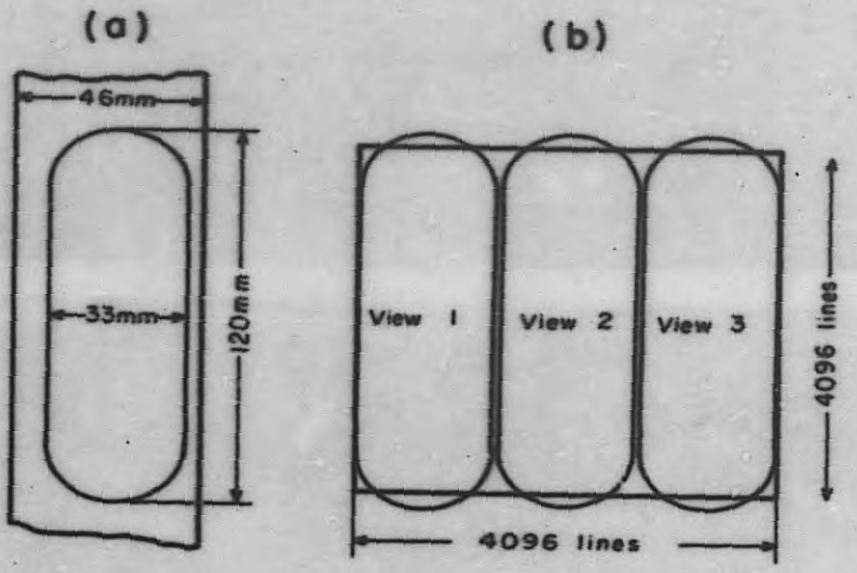


ZN-2670

Fig. 2.

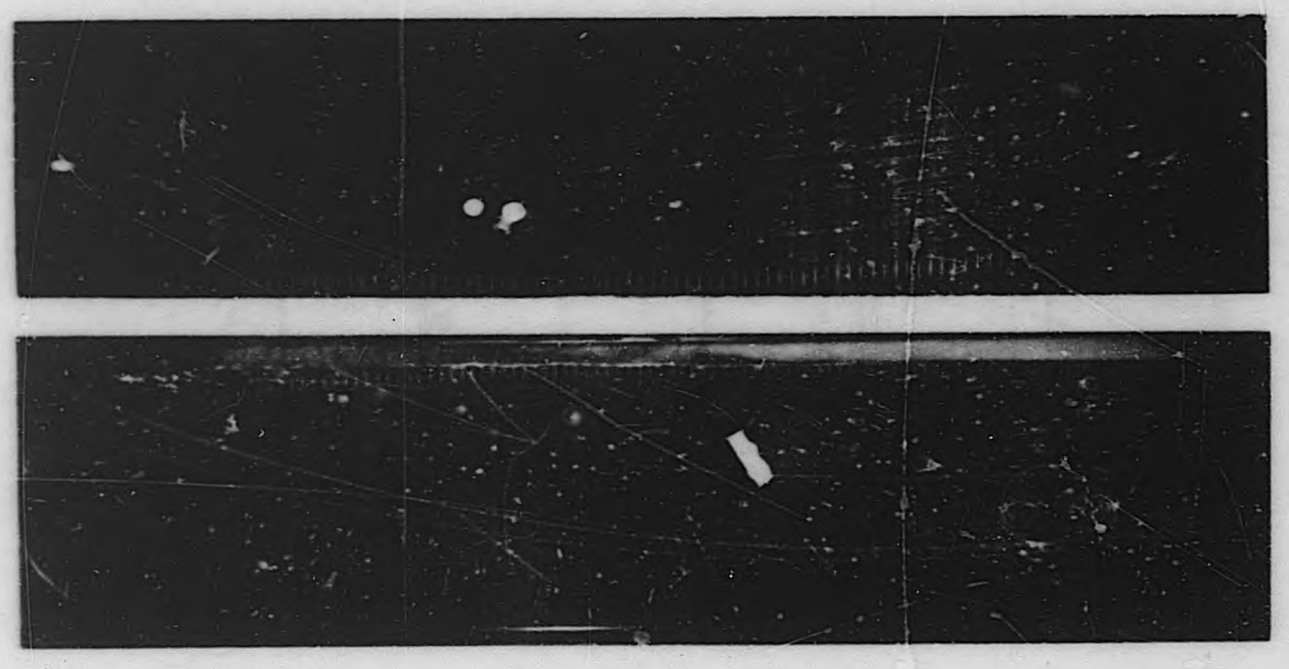
ZN-2666

Fig. 3.



MU-20651

Fig. 4.



ZN-2664

Fig. 5.

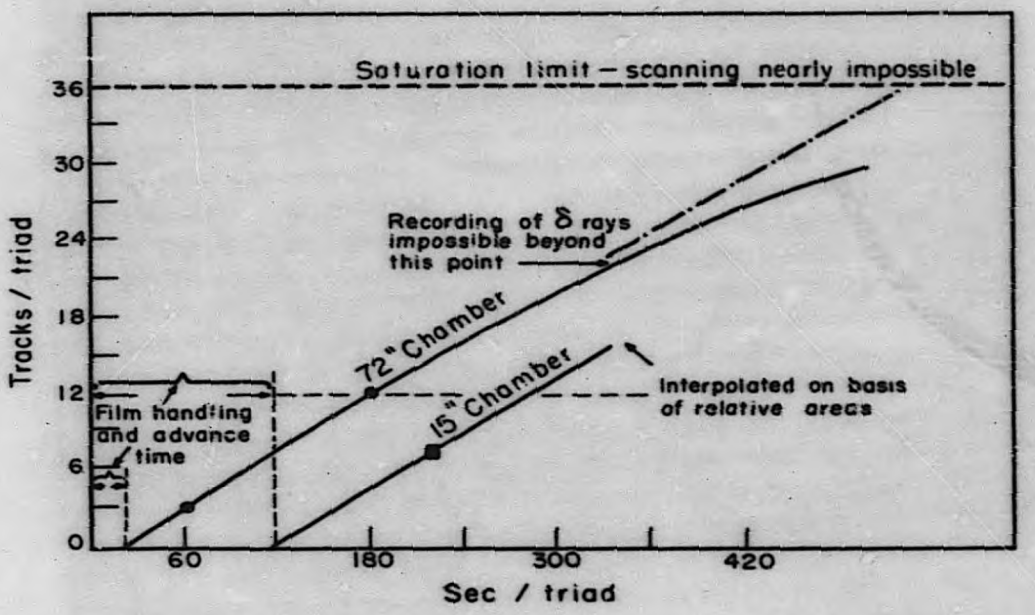

Saturation limit – scanning nearly impossible
36
30
24
18
12
6
0
Tracks / triad
Recording of δ rays impossible beyond this point
72" Chamber
15" Chamber
Interpolated on basis of relative areas
Film handling and advance time
60
180
300
420
Sec / triad
MU-20648

Fig. 6.

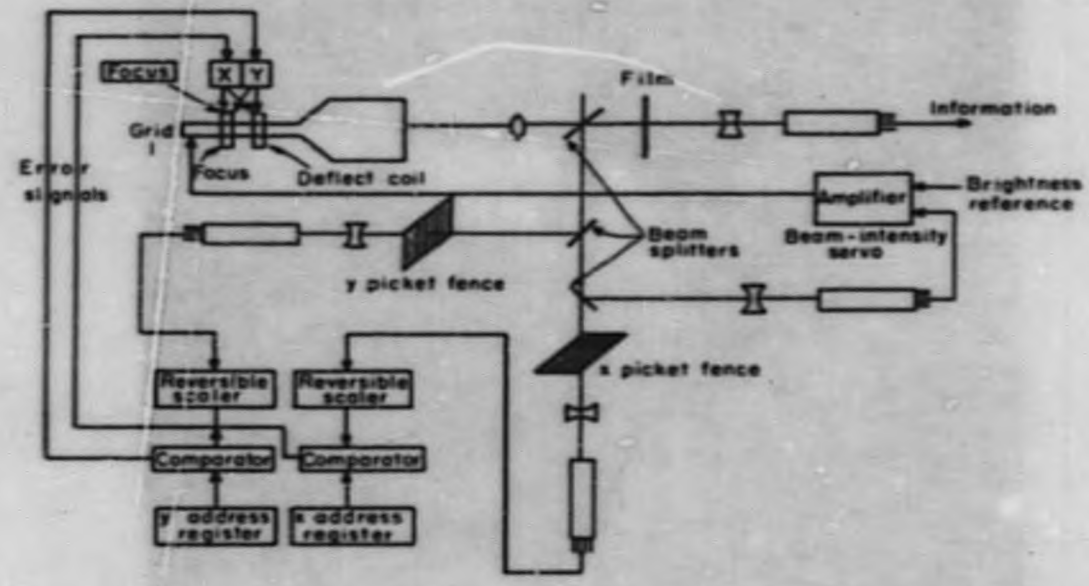
ZN-2665

Fig. 7.

171 45

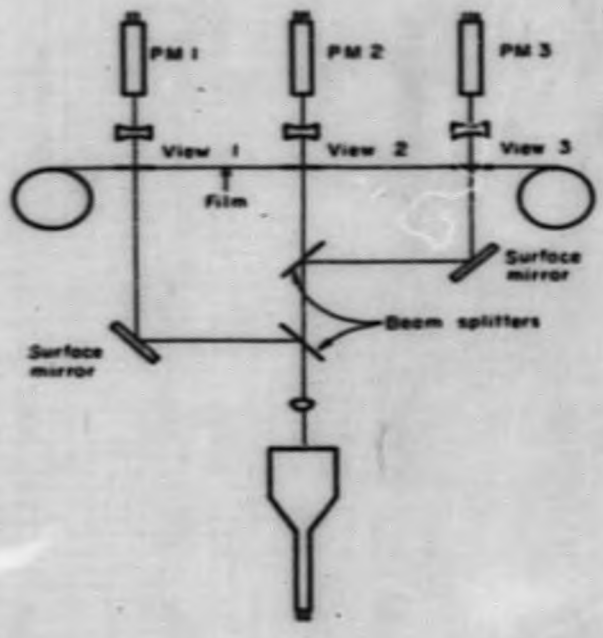
ZN-2668

Fig. 8.



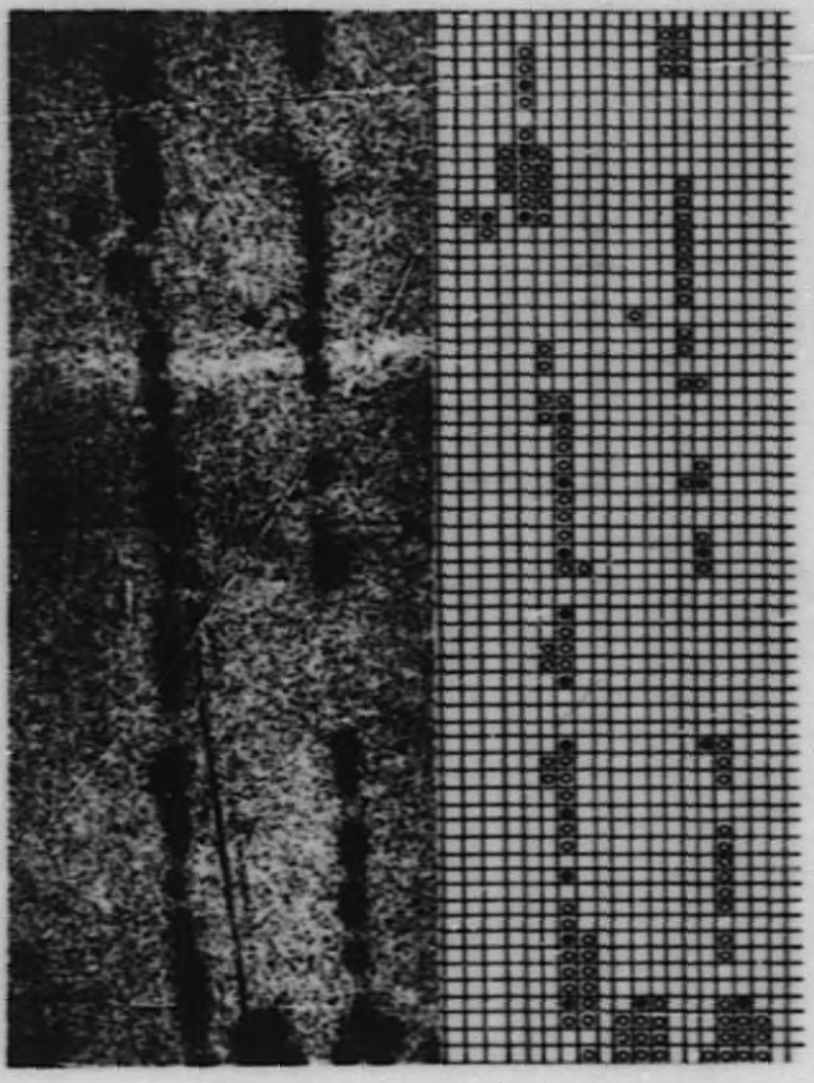
MU-20657

Fig. 9.



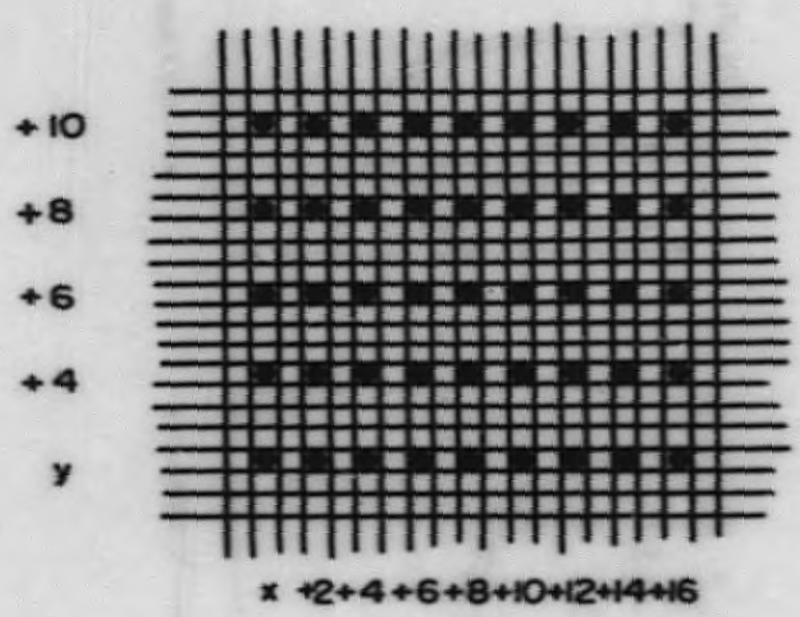
MU-20650

Fig. 10



ZN-2667

Fig. 11.



MU-20646

Fig. 12.

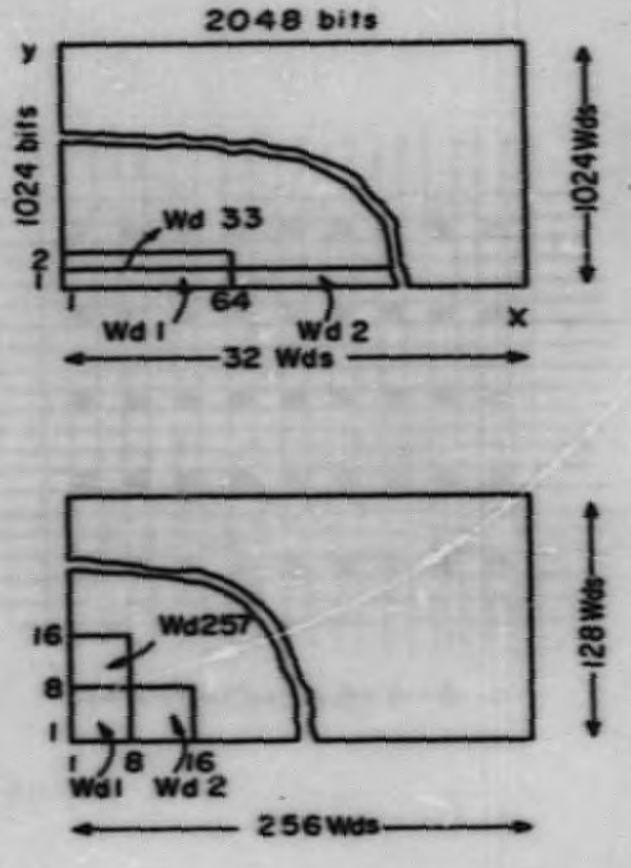
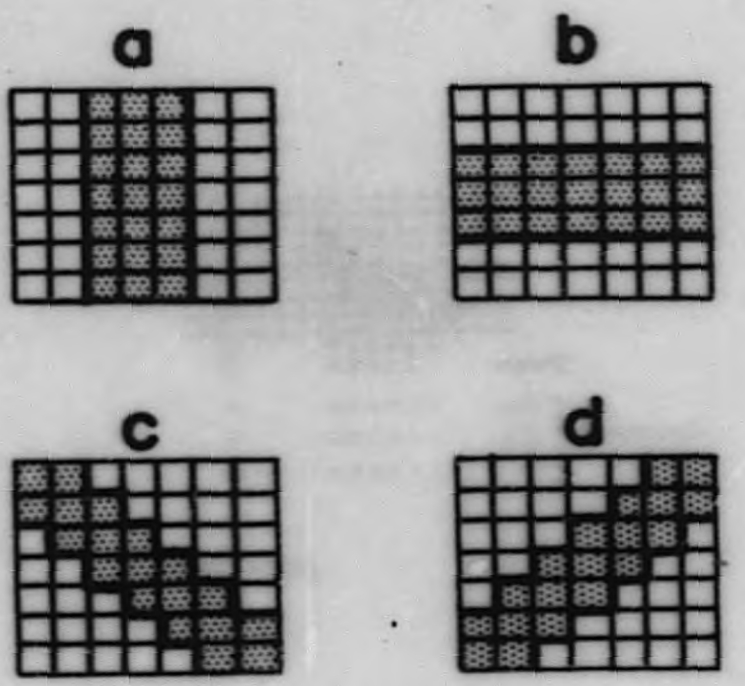
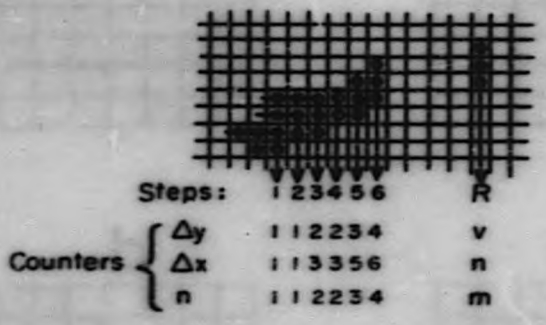

2048 bits
y
1024 bits
Wd 33
2
1
64
Wd 1
Wd 2
x
32 Wds
1024 Wds
16
8
1
Wd257
1
8
16
Wd 1
Wd 2
256 Wds
128 Wds
MU-20649

Fig. 13.



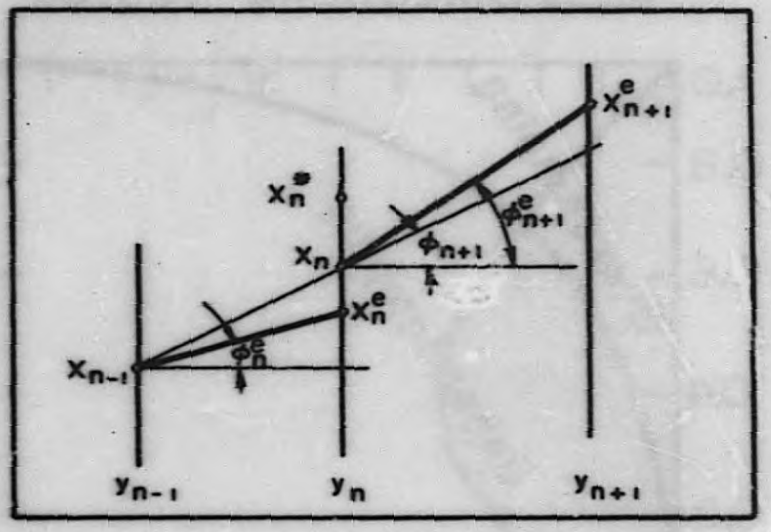
MU-20645

Fig. 14.



MU-20653

Fig. 15.



4U-20654

Fig. 16.

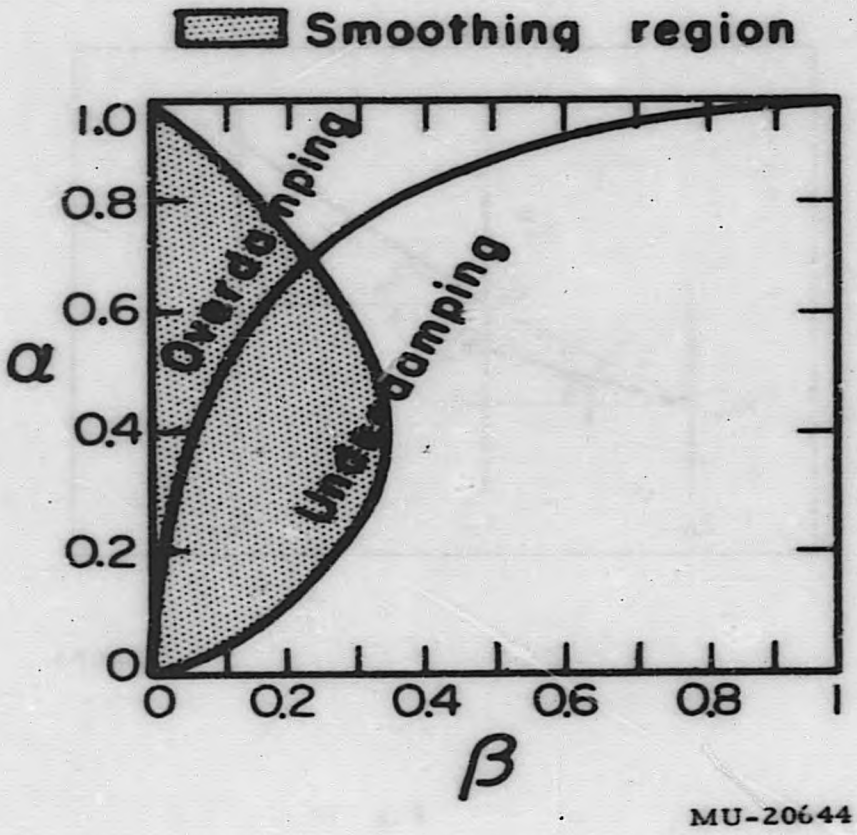

Smoothing region
Quadrupling
Unstable damping
1.0
0.8
0.6
0.4
0.2
0
α
0
0.2
0.4
0.6
0.8
1
β
MU-20644

Fig. 17.

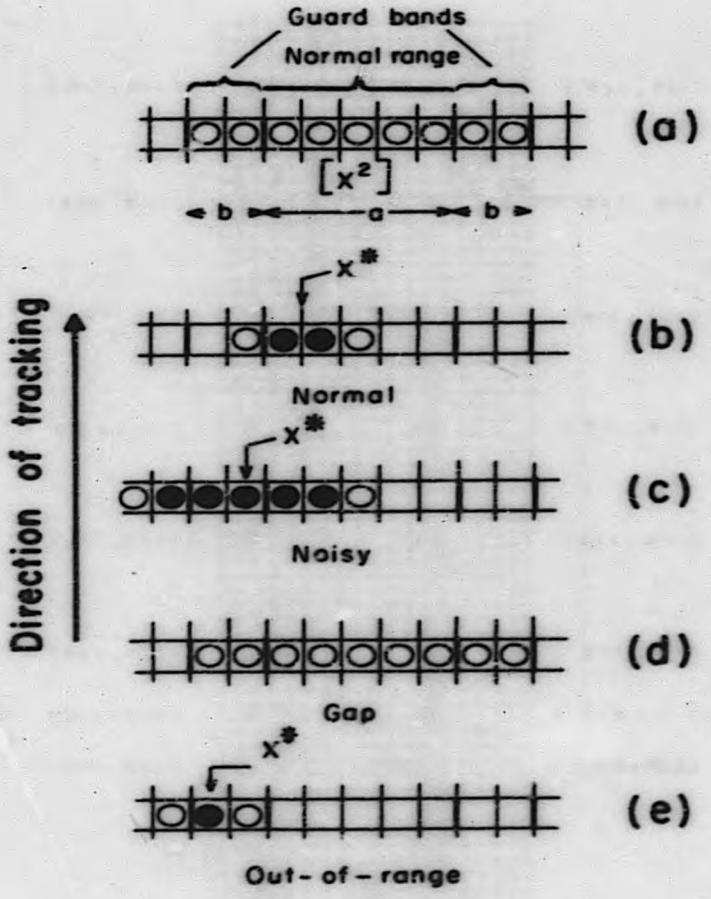
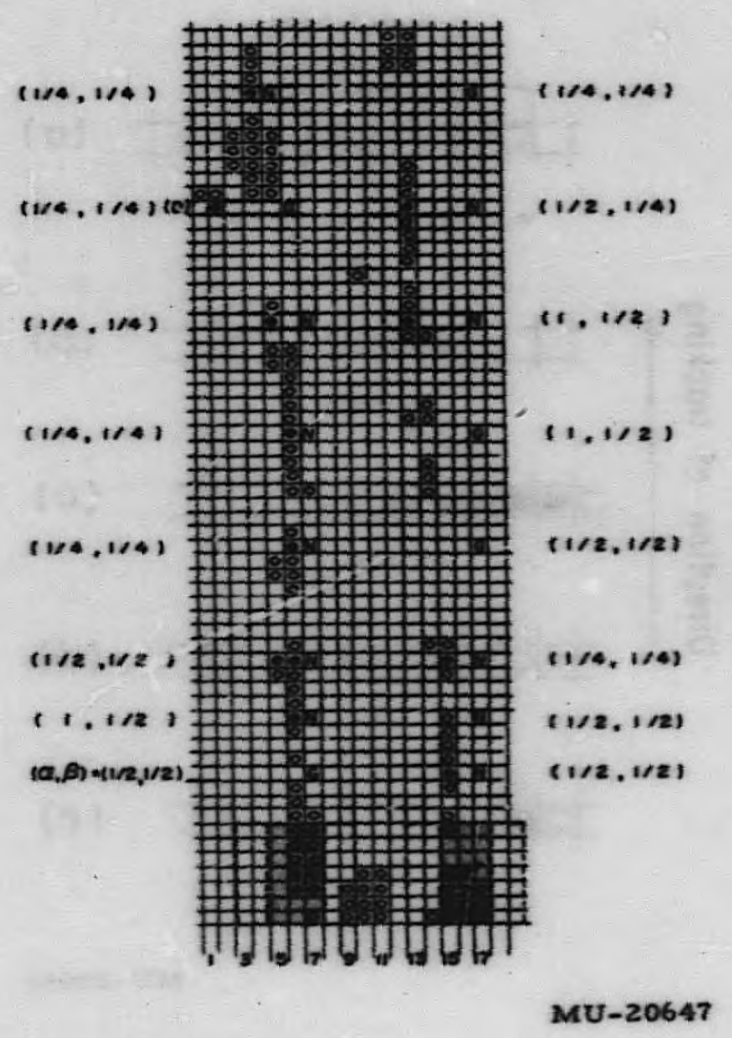

Guard bands
Normal range
(a)
[x²]
b
a
b
x*
(b)
Normal
x*
(c)
Noisy
Direction of tracking
(d)
Gap
x*
(e)
Out-of-range
MU-20656

Fig. 18.



MU-20647

Fig. 19.

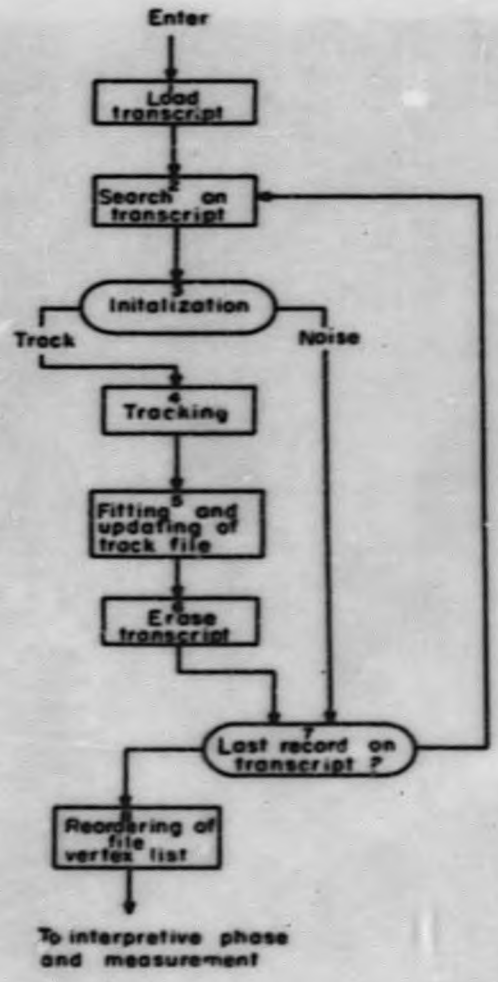
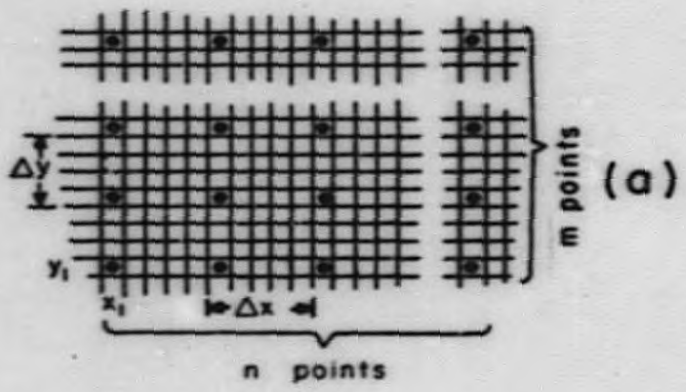

Enter
Load transcript
Search on transcript
Initialization
Track
Noise
Tracking
Fitting and updating of track file
Erase transcript
Last record on transcript ?
Reordering of file vertex list
To interpretive phase and measurement
MU-20652

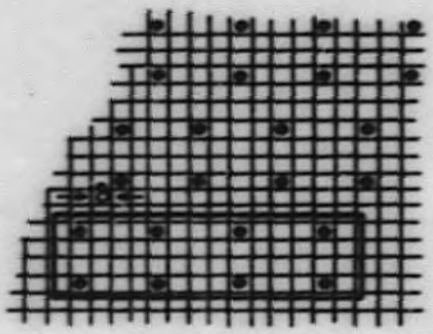
Fig. 20.

ZN-2669

Fig. 21



(a)



(b)

MU-20655

Fig. 22.