Lawrence Berkeley National Laboratory
Recent Work

Title
THE USE OF A LARGE PHOTODIGITAL MASS STORE FOR BUBBLE CHAMBER ANALYSIS

Permalink
https://escholarship.org/uc/item/65r864f6

Authors
Alston, Margaret H.
Penny, Samuel J.

Publication Date
1965-05-01

UNIVERSITY OF CALIFORNIA

Lawrence Radiation Laboratory
Berkeley, California

AEC Contract No. W-7405-eng-48

THE USE OF A LARGE PHOTODIGITAL MASS STORE
FOR BUBBLE CHAMBER ANALYSIS

Margaret H. Alston and Samuel J. Penny

May 1965

DISCLAIMER

THE USE OF A LARGE PHOTODIGITAL MASS STORE FOR BUBBLE CHAMBER ANALYSIS*

Margaret H. Alston and Samuel J. Penny

May 1965

The Lawrence Radiation Laboratory at Berkeley is planning to purchase

a photodigital store with a capacity of about 3.3×10^{11} bits. This

store, which is a "read only" memory, will be connected online to the

CDC-6600 computer and will serve as a repository for many types of

information. The data in the store can be accessed directly by programs

running in the central computer or by "priority" requests initiated

at online consoles. A proposed system for the data manipulation and

retrieval required to handle the analysis of over one million bubble

chamber events per year will be presented.

The Use of a Large Photodigital Mass Store for Bubble Chamber Analysis

Margaret H. Alston and Samuel J. Penny

INTRODUCTION

The physics analysis of bubble chamber pictures creates a large volume of digital data which is currently kept on magnetic tape. This storage medium is already proving awkward for data manipulation because of tape errors and handling problems. Within the next year or two, the number of events being measured per year will increase by a factor of three or four. From our present experience it is clear that the expense in manpower and computer time with a magnetic-tape-based system of data manipulation for bubble chamber physics will become intolerable under this load. Consequently, an alternative to magnetic tape is becoming essential.

The Lawrence Radiation Laboratory at Berkeley is planning to purchase a large photodigital storage device with a capacity of about 3.3×10^{11} bits. The store will be connected online to a CDC-6600 computer such that the data in the mass store can be accessed directly by production codes running in the computer, or by priority requests initiated by physicists sitting at online remote consoles.

In addition to the data storage and manipulation required for the physics analysis of digital data obtained from bubble chamber photographs, the mass store will be used for spark chamber analysis and rapid random retrieval by the computer of files of data for general use.

This talk outlines a proposal for using the mass store in the analysis of about one million bubble chamber events per year.

HARDWARE

In the mass store the digital information is written as black and white spots on photographic emulsion. Consequently, the memory is permanent once it is written, and outdated information must be physically purged from the system. The data are arranged in physical units of 1.6×10^8 bits (equivalent to more than one full 800 BPI magnetic tape). A physical unit may be accessed randomly within about three seconds. The data within a physical unit are subdivided into fixed-length blocks with about 1.5×10^5 bits per block. Each block may be accessed randomly or many blocks can be read serially.

BUBBLE CHAMBER DATA

The bubble chamber data will be divided into three types of files in the mass store:

Main Physics File: This file will be made up of records containing general physics information consisting of coordinate measurements, and track reconstruction and kinematic analysis data from each event measurement. In addition, there may be a preliminary physics interpretation of the event. This data will produce a record for each event of about 4×10^4 bits. Each main file will contain data from one experiment and may contain as many as 5×10^5 events. Records in this file will be batched into groups of events of similar type in order to obtain faster access when reading.

These files are accumulative, with a daily addition of 100 to 3000 event measurements to each file depending on the measurement rate for each experiment. A main file will be referenced frequently during a period of two to three years from its inception. After this time the file can probably be retired from the online system to shelf storage. However, during the next five years, it may be necessary to reload the data into the online system for further processing.

Master List File: A master list file contains about 10^3 bits of information for each event describing the current status and giving the location of the physics data for the event in the main file. Each master list file will probably contain a subexperiment of about 10^5 events. Before reading events from the main file, a pass will be made on the master list file to determine which of the events are required. A list of storage locations of the selected events will be made and ordered so that access to the main file is optimized. The master list will also be used to supply the physicist with status reports, lists, and tallies of the events in his experiment.

Periodically, master lists will be rewritten to include addenda created by additions to the main files. These addenda will reside on a magnetic disc. The obsolete master lists will be purged from the system.

Summary Files: The summary files are used by the physicists in further analysis of the bubble chamber data. They contain information abstracted from the main physics files. Each summary file can contain up to 10^9 bits with variable-length records averaging less than 10^4 bits each. This data will not be indexed by the master list. Physicists will produce new summary files periodically, often purging the older files at the same time.

ANALYSIS OF BUBBLE CHAMBER EVENTS USING THE MASS STORE AND ALVAREZ GROUP PROGRAMS

Figure 1 shows the overall flow of data when using the mass store and the Alvarez Group analysis system. The physics analysis system will consist of the following programs:

TVGP   - Geometric reconstruction of tracks

FIT   - Kinematic analysis of each vertex

EXAM I - (First Phase) - Determination of the most probable

physical interpretation of the whole event, and probably

a calculation of a few quantities in the initial center-

of-mass system for each track, e.g., 4-vectors.

EXAM II - (Second Phase) - Further calculations of physics

quantities.

SUMX   - Producing histograms, scatter plots, etc.

The seven principal types of jobs shown in the Figure are as follows:

1. Processing new event measurements and entering the data

into the main file (using programs TVGP, FIT, and EXAM I).

The master list or its addenda will be updated at this time.

2. Reprocessing old measurements and entering the data into the

main file. This requires selecting the events from the main

file and processing them through TVGP, FIT, and EXAM I. Again

the master list requires updating.

3. Making summary files from the main files. This requires

selection from the main files and possibly passing the data

through EXAM II.

4. Summarizing various physical quantities from events in the main

file in the form of histograms, scatter diagrams, etc., by

selecting events from the main file and processing them

through EXAM II and SUMX.

5. Summarizing data from a summary file using SUMX

6. Producing status and control reports from the master list files

7. Updating and remaking the master list file to include its addenda

Also shown in the Figure is an online console. This could be used
to initiate and control jobs of types 3, 4, 5, and 6. In this case,
the output might be obtained online at the console.

SIMULATION STUDIES ON THE MASS STORE

Based on times supplied by the manufacturers, the mass store and
its usage were simulated using the IBM program GPSS II. This simula-
tion investigated the response times for random and serial requests to
the mass store, occurring both separately and concurrently.

The simulation models were run under varying request patterns
assumed to represent the Laboratory's usage of the mass store. The
requests fall into two categories:

1. Batched, serial requests representing access to large data
   files, primarily required by the bubble and spark chamber
   users;

2. Random priority requests for files of less than 5×10^6 bits.

It was assumed that the major use of the mass store will be bubble
chamber data analysis involving very large files of data. When this
data is written into the mass store it will be possible to group large
batches of "similar" records so that they appear in contiguous locations
in a main physics file. Requests for reading this type of file are
characterized by batch length, record length, and selection factor (the
percent of the records to be read from a given batch). A prerun will
be made through the appropriate master list file to obtain the addresses
of required events. These addresses are then ordered so that retrieval
of the data from the mass store will be in a sequential or serial manner.
The priority requests are characterized by the size of the request
and by their frequency of arrival. A further assumption was made that

priority requests, whether initiated by a main CPU program or from an online console, must be serviced as soon as possible and are, therefore, superimposed on the serial requests.

The results of the simulation showed that a random request can be serviced in about three seconds if no serial request is being read when the random request is initiated. This time increases to between five and six seconds when a serial request is in progress. Figure 2 shows the results for serial requests when no priority requests are superimposed. The reading of serial requests is degraded by about 8% per priority request per minute.

From data obtained in the simulation study, we can estimate the time required per day by a group of approximately 20 physicists and 20 graduate students to analyze 10^6 events per year, (using the mass store and a fast multiprogrammed computer (Table I)). This shows a usage of the main CPU of three hours and forty minutes spread over a five-hour period.
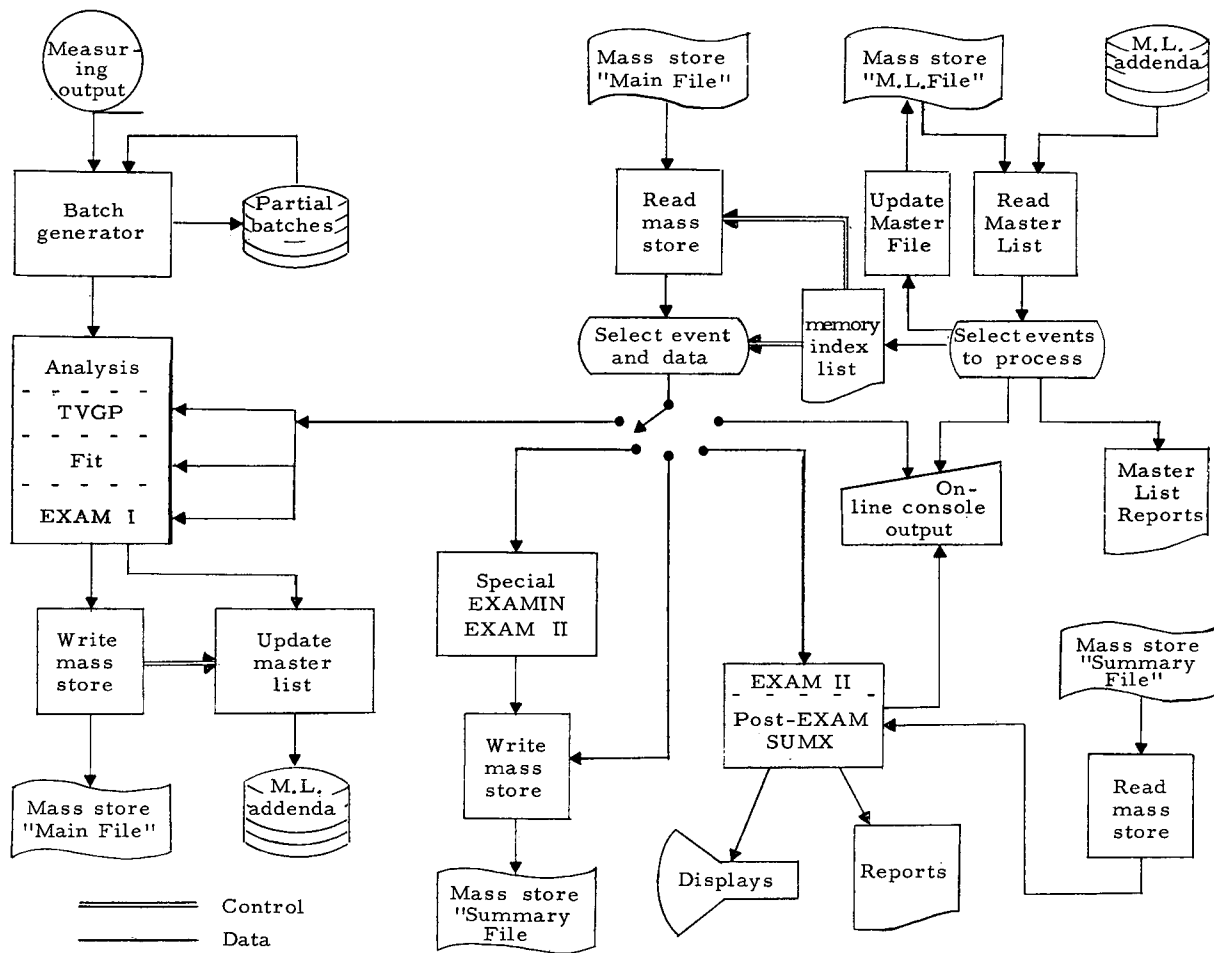
CONCLUSIONS

The mass store will be a very valuable tool for processing large numbers of bubble chamber events. The manufacturers assure us that the photographic and error-correction-detection processes should make the device much more reliable than magnetic tape. Considering our past troubles with reading and writing data on magnetic tape, the mass store may be the only way of successfully analyzing large numbers of events. In addition the mass store provides a very large online read-only memory which should prove very useful for many other types of data analysis. Unfortunately, the hardware will not be available for about two years.
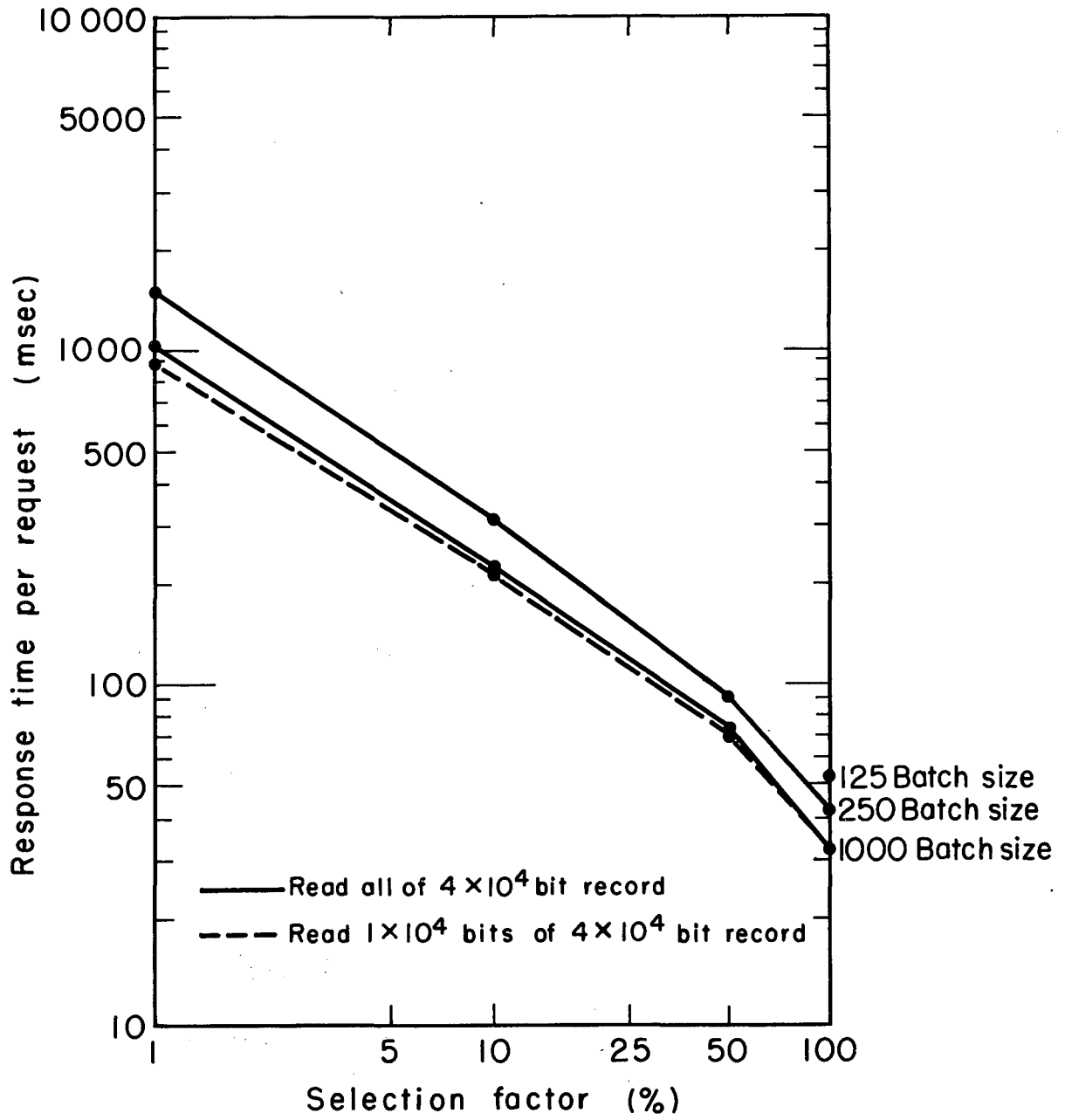
TABLE I

Type of Job	Read Time (mins.)	Write Time (mins.)	CPU Time (mins.)	Total Time for Jobs (mins.)
1) Process 3000 new events/day. Includes check read and update Master Lists.	10	24	58	60
2) Reprocessing of events. Ave. 1000 per day. Includes check read and update of Master Lists.	7	8	23	25
3) Making Summary files includes using Master List to obtain an index for access to Main File (3 jobs).	30	27	20	60
4) Exam II and Sumx on Main File includes using Master List to obtain index (8 jobs).	70	--	60	75
5) Sumx on Summary Files (20 jobs) (Some jobs might be ON-LINE)	16	--	30	30
6) Make reports, lists, etc., from Master Lists (20 jobs).	35	--	40	40
7) Update (2) and remake (1) Master List.	5	20	9	25
TOTAL	173 mins	80 mins	235 mins	315 mins = 5 1/4 hrs.

NOTES: a) Times based on expected performance of the mass store and a CPU speed 10 times that of a 7094.

b) Channel utilization in reading mode is 70%. Data bursts 78 ms.

c) Channel utilization in writing mode is 30%. Data burst 500 ms.

d) Reading and writing cannot be performed simultaneously.

MUB-6101

Fig. 1

Fig. 2

MUB-6085