X-560-70-291
PREPRINT

NASA TM X- 65363

SPACE SCIENCES DATA HANDLING-PRESENT CAPABILITIES AND REQUIREMENTS FOR FUTURE LARGE ASTRONOMY MISSIONS

GEORGE H. LUDWIG

AUGUST 1970

GSFC
GODDARD SPACE FLIGHT CENTER
GREENBELT, MARYLAND

This paper was presented at the Optical Telescope Technology Workshop held in Huntsville, Alabama, April 28 - May 1, 1969.

X-560-70-291

SPACE SCIENCE DATA HANDLING-PRESENT CAPABILITIES AND REQUIREMENTS FOR FUTURE LARGE ASTRONOMY MISSIONS

George H. Ludwig

August 1970

GODDARD SPACE FLIGHT CENTER
Greenbelt, Maryland

SPACE SCIENCES DATA HANDLING-PRESENT CAPABILITIES AND REQUIREMENTS FOR FUTURE LARGE ASTRONOMY MISSIONS

George H. Ludwig
Goddard Space Flight Center

ABSTRACT

All data transmitted from the spacecraft must be received and processed before they may be used to meet the objectives of the mission. The problem can be separated into four elements:

1. Acquisition of the data from the spacecraft.
2. Relay to a central processing point.
3. Processing to meet operational needs.
4. Processing to meet in-depth analysis needs.

The present state-of-the-art for processing data from space sciences missions includes the capability for handling up to 3×10^8 data points per day from a variety of approximately twenty different satellites. In addition, systems for handling large volumes of image data are being developed for future ERTS missions. In the processing facility for non-image data, plans are being made to change the data flow to provide more rapid delivery of data to the users and a still higher level of quality control. Many of these techniques will be required for a majority of the complex astronomy missions envisioned.

CONTENTS

SPACE SCIENCES DATA HANDLING - PRESENT CAPABILITIES AND REQUIREMENTS FOR FUTURE LARGE ASTRONOMY MISSIONS

INTRODUCTION

If a space mission is to be wholly or partially justified on the basis of the scientific experiments that it may carry, then it is important to realize that the objectives of the mission will be achieved not when the spacecraft is activated, or when the data have been returned, but only after the data have been processed, analyzed, and the results have been published by the experimenters. This paper addresses the problems of processing on the ground the very large volumes of data which are telemetered from modern spacecraft.

To introduce the problem, Figure 1 shows the general data flow for a space mission. It was drawn for the Explorer and Observatory classes of spacecraft, but is generally applicable. Any space experiment system consists first of a number of sensors which feed data in electronic form into various inputs of the data processing and reduction equipment onboard the spacecraft. The data are then collected by a central data collection subsystem and telemetered to the ground, where two uses are made of the data. The first is the real time and near-real time processing for operations in the control center, with feedback to initiate changes in the performance of the experiments and subsystems onboard the spacecraft. The second is the processing accomplished for the ultimate analyses by the investigators. This begins within a central data processing facility and concludes with the detailed data reduction and analysis by the individual experimenters.

This paper assumes that the data acquisition, communication, and data relay networks will exist at the time large optical telescopes are launched, so that the data transmitted from them can be assembled at a common data collection point within the necessary time constraints. Data relay in this case may include several forms, from magnetic tapes forwarded by mail (probably only as backup to faster means), to direct data relay, either from the spacecraft to ground stations and then through synchronous relay satellites to a common collection point, or directly to the common point through data collection satellites. The paper deals with two primary considerations: (1) The present state-of-the-art of the ground processing technology; and (2) The new technology needed.

PRESENT TECHNOLOGY

To illustrate the present technology, three representative areas are described. The first is the processing performed within the central data processing facility within the Goddard Space Flight Center for the majority of the space sciences earth satellite missions. The second is a brief discussion of a portion of the

NIMBUS control center dealing with image data. The third is a short discussion of some of the processing that the experimenters must perform before they can reach their conclusions.

As mentioned before, data processing at GSFC is, for most projects, divided into two types. Mission Control Centers for the various projects are responsible for the day-to-day and longer-term operation and evaluation of the spacecraft subsystems. In addition, they optimize the experiment operations to the extent possible without extensive experiment data reduction and analysis. The Central Processing Facility is responsible for assembling the data base for the in-depth reduction and analysis by the experimenters. This division of functions will continue to be valid for many types of experiments which require limited experimenter interaction on many of the astronomy missions. For this reason, it is likely that a new facility - an Experiment Operations Facility - containing provisions for in-depth, near real time data reduction and analysis will be needed to provide the information necessary for the effective conduct of the observing program.

Since the operational control will be discussed more completely in other papers at this workshop, most of this paper will deal with the data processing necessary for the extensive in-depth analysis of the experiment data. Figure 2 indicates the data flow within the present Central Data Processing Facility. Data are received at the central facility from the various data acquisition stations by two means, analog tapes and data lines of various types. The tapes pass initially through an evaluation procedure to facilitate control of the quality of the processes at the ground stations, and to provide quality information to the data users. Following this, all data undergo analog to digital conversion. This refers to the conversion of the demodulated receiver output signals into computer-compatible digital signals, and includes the removal of noise introduced in the telemetry process. The next major step involves preliminary processing, referred to as editing, within a digital computer. This editing includes checking the internal consistency of the data and measuring the data error rate. The remaining operations are performed within the large scale computer, which is a Univac 1108 multiprocessor with two Central Processing Units and 196,000 words of core memory. One of these operations is time tagging. Each data point received from space must be given an accurate standard or Universal Time reference since all linkage of the telemetered data with other correlative data is done through this time linkage. A number of the spacecraft sensors are directional. This requires that the spacecraft attitude at every moment of time be computed. In addition, it is necessary to perform some sorting, eliminating useless data, merging of orbit and other correlative data with the experimental data and, finally, decommutation (sorting) to form outputs for each of the various experimenters. In addition, a master digital data tape is generated which is retained for a period of time for reprocessing or regeneration of the experimenter's tapes if necessary.

The volume of data being handled at the present time in this central facility is quite large. Figure 3 illustrates the growth in analog tapes arriving at the facility over the past two years, and projects this input two years into the future. At the present time, about 2,400 tapes per week are arriving for 21 active satellites. This represents about 3.5×10^8 measurements per day, or about 3.5×10^9 bits per day of PCM-type data. Another way of expressing this data rate is to compute the long term average bit rate. In these terms, the central facility is receiving data at the rate of 35,000 bits per second for every second of every day. The impact of this data rate on the sizing of the central facility is quite impressive. Figure 4 is a view looking up one row of processing lines in the central facility. Three processing lines are included in the figure, two PCM lines, and a special purpose line for digitizing the signal from rubidium vapor magnetometers in several of the OGO spacecraft. The most distant line in Figure 4 is shown in Figure 5. The various racks include synchronizing and bit conditioning circuits, a time decoder and a core buffer memory. The digital tape recorder is out of the picture to the right. This line is representative of the class of lines which do not include general purpose computers. The central facility contains 16 major lines of this form and about 16 smaller lines.

Figures 6 and 7 indicate the size of the complete facility. The row of processors in Figure 4 is the third row of racks from the left in Figure 6. The second generation processing systems include CDC 3200 computers on-line with the front end processing equipment for two purposes, setting up the operation, and measuring the data quality. Three of these larger systems are shown at the right end of the central equipment bay in Figure 6. The Univac 1108 computer is shown in the center of Figure 7, surrounded by various support equipment. The central facility shown in Figures 6 and 7 represents about a $20 million capital investment for handling the data rate presently being received. It is operated and maintained by a staff of approximately 300 government and contract personnel.

The second area of activity to be discussed in this paper involves some of the image data processing being performed in the NIMBUS control center and processing facility. Figure 8 illustrates the handling of the NIMBUS Medium Resolution Infrared (MRIR) data obtained from one of the four major experiments being flown on NIMBUS II. After analog to digital conversion, the image data and orbit are merged to provide gridding directly on the images. Tapes containing the digitized data are sent to the experimenters for more detailed analysis. Strip charts are prepared for spacecraft performance evaluation, and the images are prepared for immediate viewing. Figure 9 indicates one of these image sets. It includes one complete orbit with the two poles being indicated in the gridding on the left. The second strip is the visual presentation, where it can be seen that the bottom half of the orbit was in darkness while the top half was in light. The other spectral regions covered by the instrument are indicated.

Moving from the GSFC facilities, the experimenters must complete a considerable amount of processing for their final analyses after they receive their data. The general data flow is indicated in Figure 10. The experimenter passes his data through quality checks, sensor calibrations, and data reduction which leads to outputs which can be viewed. Tabulations and data plots are the most common output forms. Other forms such as motion pictures and color presentations are coming into use. Figure 11 is a sample presentation. It is not by any means one of the most complex. It represents a compilation by Dr. Norman F. Ness of six months accumulation of magnetometer data from the IMP-1 spacecraft. This one chart presents a condensation of about 100,000 individual measurements. The magnetic field in the magnetosphere is represented by the vectors, where the length of each vector indicates the magnetic field strength at that point, and its direction indicates the field direction. From this presentation, one can identify the major features of the earth's magnetosphere. It required many weeks of computer processing, manual manipulation, and study in its presentation. And it is only one of many presentations required before the findings of this experiment could be published. It is significant to note that this type of analysis is very unlikely to be performed in any spacecraft by any on-board computer in the foreseeable future. Thus, it will still be necessary to retrieve the data rapidly, with the experimenters in the operational loop, in order to permit this type of analysis.

EMERGING TECHNOLOGY

Before moving on to a discussion of future technology, it is of value to address briefly the question of emerging technology. The data flow shown in the present large scale facility in Figure 2 results in a very large amount of materials handling. Two thousand input tapes per week require the handling of 6,000 tapes per week before all the operations have been performed. These lead eventually to about 25,000 bookkeeping transactions per week. Figure 12 indicates a data processing flow which is receiving considerable study at the present. It involves the collection of as much of the data as possible in near real time by data relay, and on-line data validation, analog to digital conversion (if required), initial processing and editing, and direct entry into a large data bank. This data bank is envisioned as a hierarchy of different types of storage, including core for fast access speeds, drums for medium access speeds, and tapes for slow access speeds. But the most important new feature will probably be a high density storage system involving the recording of digital data in optical form on film chips to provide a total on-line data volume of from 10^{10} to 10^{12} bits. With such a data bank, the raw data will be entered into active storage nearly as soon as they are received. More complete processing will then proceed as rapidly as correlative data, such as orbit, are available. The output will be available for delivery from the data bank, immediately in raw form, and later in more highly processed form. This output will be useful both for local display for operational purposes, and remotely to the users for their final analysis. Two factors leading to this data organization are the anticipated increased efficiency through the elimination of much of the manual handling of supplies, and faster data delivery. At the present time, an attempt is made to operate with a six-week backlog or less. That is, considerable effort is expended to provide the bulk of the data six weeks after it has been collected from the spacecraft. The future goal should be to make data available from the data bank within several hours or a day from the time of transmission. This will be especially important for an operationally complex orbiting telescope.

The other major area of emerging technology is that which will be required for the Earth Resources Technology Satellites (ERTS). Of course, ERTS will involve primarily image data, which requires techniques that are considerably different than for PCM data. The presently anticipated data flow is indicated in Figure 13, with the data entering image processors, and with the images becoming available for viewing within a short period. In addition, the PCM data will enter other operations to facilitate the gridding and the rectification of the pictures. It is presently expected that gridded overlays will be made available along with the images. This facility is expected to cost more than 10 million dollars, and will require about two years to build.

TECHNOLOGICAL PROBLEM AREAS

The data volume and rate problem has already been discussed. It is expected to continue to be a serious problem. Speed of data delivery has also been mentioned. Now that the experiments are becoming more complex, the experimenters need to receive the information from their instruments much more quickly in order that they may be sure that they are operating properly, and in order to modify their operation to take advantage of the variability of the instruments and the spacecraft. Therefore, in some cases it is necessary for experimenters to receive some portion of their data immediately, especially during critical periods of operation.

Third on the list of major problems is the control of data integrity and quality. The taking of a large number of pictures on a spacecraft without noting the time during which all of the pictures were taken is a threat to the <u>integrity</u> of the data. It may result in the absence of correlative information needed to determine the locations and conditions under which the pictures were taken. The question of data <u>quality</u> is concerned with the problem of verification that the entire system, including all the processing operations, is operating properly to ensure that high accuracy and resolution are maintained. This must be done continuously and thoroughly at each processing step.

Control and allocation of processing resources is a fourth major problem. For one reason, this becomes difficult because of the unpredictable nature of the load. Almost any change from expected operation of a spacecraft, other than the complete failure of the spacecraft, results in an increase in the work load in the ground processing and experimenter's facilities. This arises because partial failures onboard the spacecraft require more processing on the ground to straighten out the data or to evaluate the spacecraft. The unpredictable lifetimes of scientific missions also introduce large uncertainties in the processing work load. Selective data acquisition has been proposed as a means for leveling the processing work load, but has proven to be very difficult. If a certain spacecraft capability exists, in terms of information bandwidth for example, then there is a strong tendency for that capability to be used. The reason for this is that many of the experimenters are awaiting unpredictable events because of their value in interpreting nature. Thus, they cannot predict when their experiment must be operating. They simply must operate continuously to ensure readiness. This rule of full use of capability tends to be slightly less true as the technology evolves. To illustrate, the Orbiting Geophysical Observatory has a telemetry capability of 64,000 bps. But it is used only on an average of 7,000 to 8,000 bps by the use of a controlled data acquisition sequence. This is achieved by the use of two modes of operation: a monitoring mode at a low data rate which will always intercept the unexpected events; and a high data rate mode to provide periodic high resolution measurements. With a manned space station some additional capability will exist for governing the transmission of data to the ground, because the man can monitor the events occurring from the scientific experiments, and can eliminate some of the data transmission to ground. This will be especially true if the space station contains a data buffer to store the data until they have been scanned for significance.

Selective data processing has been found to be a fairly effective way of controlling the use of ground resources. Once the ground processing system becomes saturated, the lowest priority data are placed directly into storage. The data remain available for the exploitation of interesting periods. Selective data processing implies, however, the need for a scanning or browse file so that the experimenters will know what data exist.

A fifth major problem area is the control of the processing facility/experimenter's interfaces. If an experimenter changes his format requirements, this may require several months and several man months of effort to change the programs in the central processing facility. Therefore, considerable attention must be paid to the question of specifying and understanding the experimenter's data needs.

Finally, there are the problems associated with the data analysis by the various experimenters. These investigators usually have very limited facilities and manpower, especially at universities. Therefore, a very careful study needs to be performed to determine the tradeoffs between the steps performed in a central facility, which may be done in common for many experiments, and all the other steps, which the experimenter must perform within his own facilities.

AREAS REQUIRING ADDITIONAL TECHNOLOGICAL DEVELOPMENT

1. New work is necessary to develop a technological basis for operating onboard scientific data processing computers. We know how to build these computers—several are being built now for a number of future unmanned scientific missions. But it is not clear that we know how to control the programming of those computers so that we may be assured of full understanding after the fact of the steps performed by these computers.

2. Faster ground processing equipment is needed. The present state-of-the-art permits about one megabit per second processing rates in terms of special processing equipment such as signal conditioners, and in terms of computers. If PCM systems will operate at several megabit per second rates, then faster processing systems will be needed on the ground.

3. The third area involves the utilization of large volumes of image data. Several authors have mentioned image rates of one thousand pictures per day from a future space station. Others have predicted that these will be processed by the use of advanced digital techniques. With present technology, one picture having 4,000 by 4,000 elements of spatial resolution requires three computer tapes for its storage in digital form. Thus, the 1,000 pictures will occupy about three thousand computer tapes. If these three thousand computer tapes are read by a computer at its maximum tape reading speed (5 minutes per tape), then ten days will be required simply to read them. Therefore, it should not be expected that that many pictures will be digitized, retaining all of the resolution, by present techniques. In determining what will be done with these pictures, it will probably develop that some new technologies will be needed.

4. On-line, mass data storage for ground computers. Needs exist now for systems providing storage of 10^{11} bits, with access times of seconds or less, to provide several weeks or months of data on-line. Capacities of 10^{13} to 10^{14} will probably be required within the next six years.

5. On-line programming and processing systems and aids. There have been numerous estimates of programmer's efficiencies, in terms of

dollars per instruction. These estimates range from two to ten dollars per instruction, and many programs contain more than 100,000 instructions. In other words, computer programming is a very slow and expensive process. There is a very great need for improvement of the technology of writing programs and getting them to work properly. In addition, present programming systems are ill-suited for handling telemetry data, especially when it comes to manipulating individual bits.

6. Display and presentation technology needs stimulation. This involves the problem of the rapid comprehension of very large volumes of data by the experimenters and other users. A number of activities are currently underway in the investigation of color, motion picture displays, three-dimensional displays, etc. But much more activity is needed in this area.

CONCLUSION

The data will need to reach the experimenters from orbiting telescopes accurately and rapidly, some of it almost immediately, some of it within a day, and most of it within a week. It must be in easily usable form or it will tend not to be used because of the large volume. It is interesting to note that our experience with unmanned scientific missions has indicated that as many dollars are necessary for the data processing and analysis following a launch, as for building the experiments before launch. It appears reasonable to expect that this rule of thumb may also apply to the data processing for future astronomy missions.

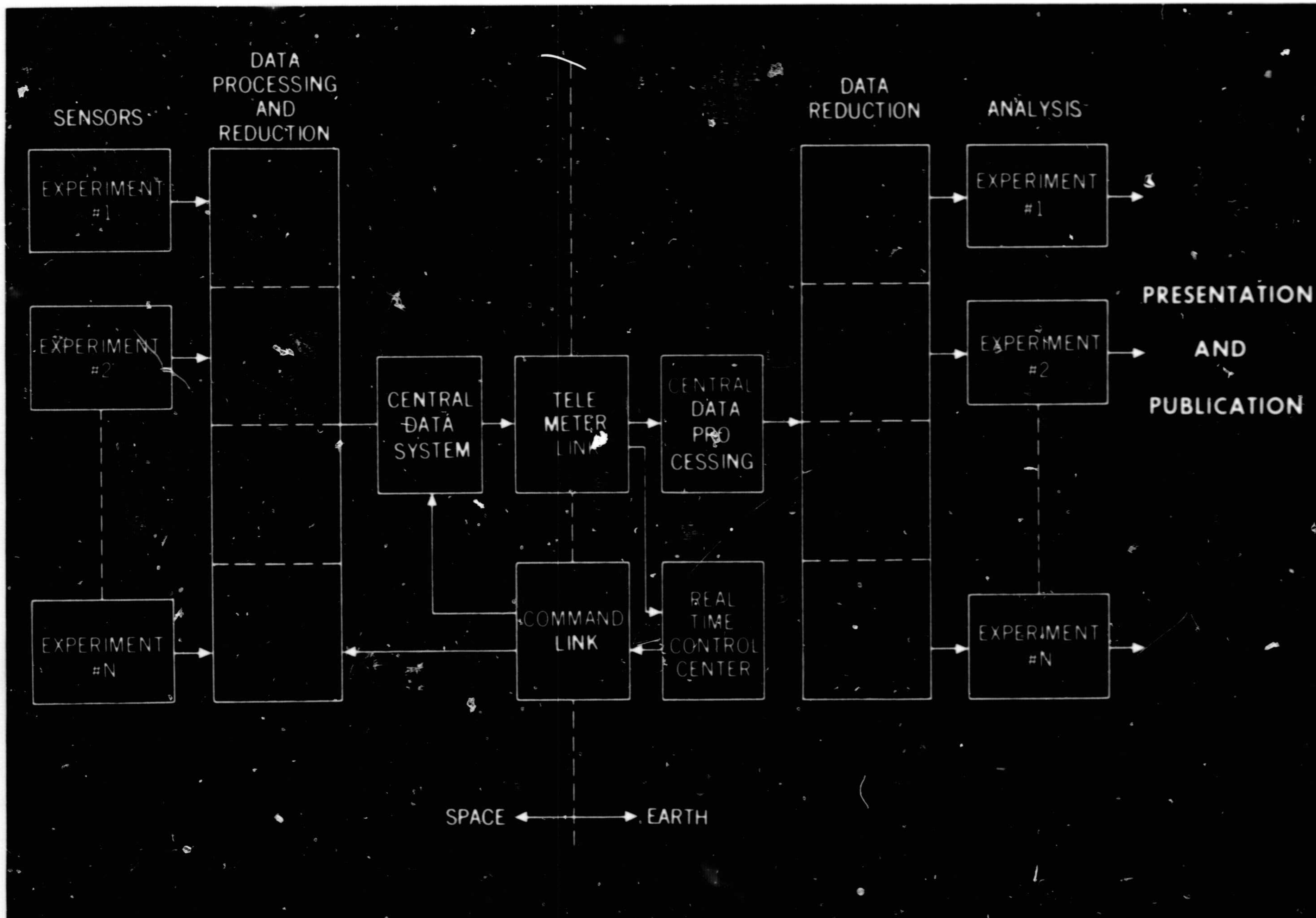
SENSORS
DATA PROCESSING AND REDUCTION
EXPERIMENT #1
EXPERIMENT #2
EXPERIMENT #N
CENTRAL DATA SYSTEM
TELE METER LINK
COMMAND LINK
CENTRAL DATA PRO CESSING
REAL TIME CONTROL CENTER
SPACE
EARTH
DATA REDUCTION
ANALYSIS
EXPERIMENT #1
EXPERIMENT #2
EXPERIMENT #N
PRESENTATION AND PUBLICATION

Figure 1

CENTRAL DATA PROCESSING FACILITY
PRESENT

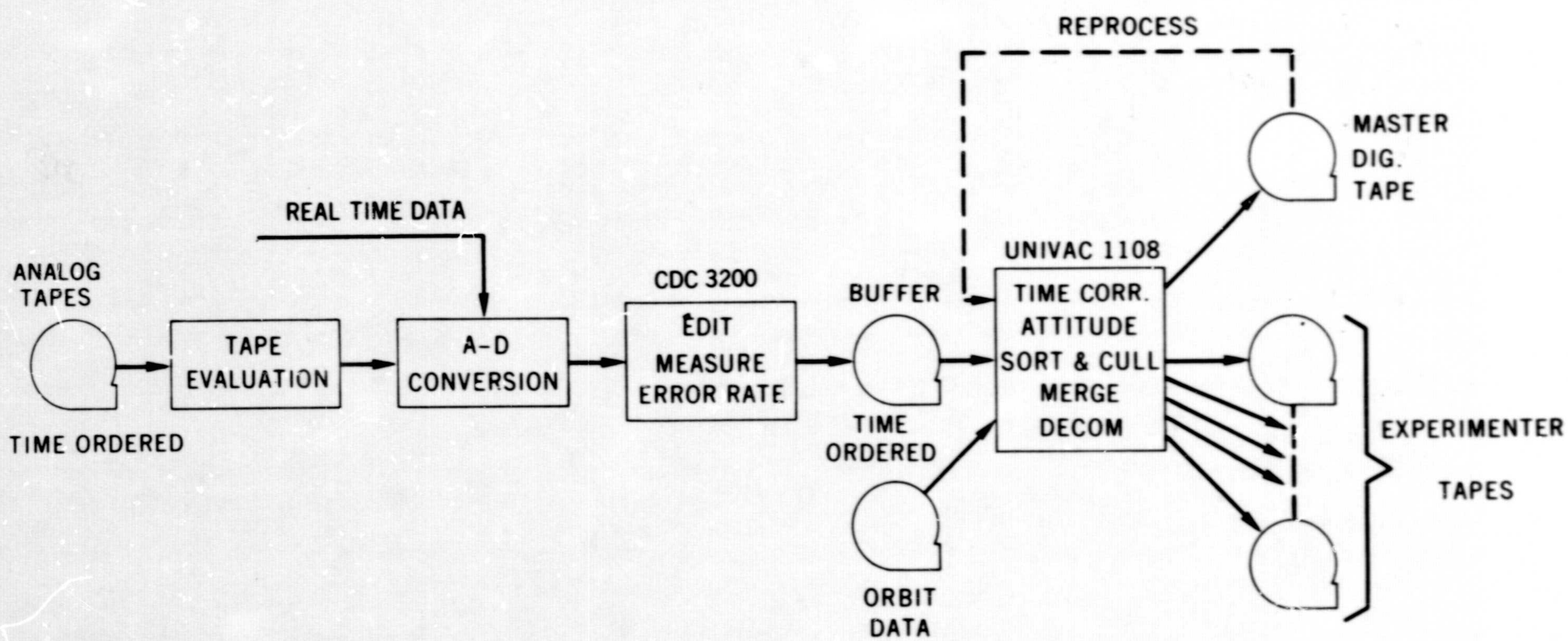

Figure 2

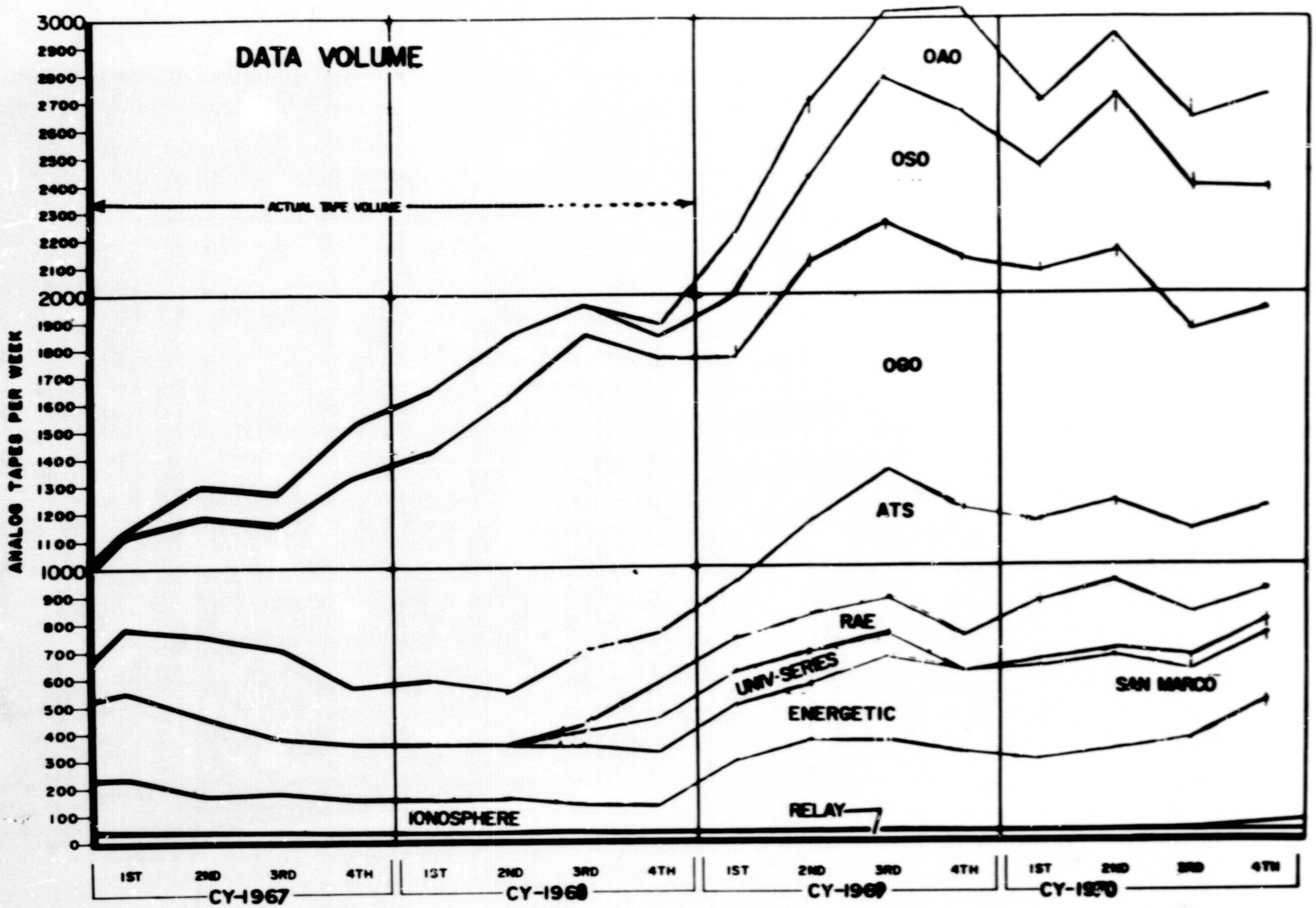
DATA VOLUME
ANALOG TAPES PER WEEK
3000
2900
2800
2700
2600
2500
2400
2300
2200
2100
2000
1900
1800
1700
1600
1500
1400
1300
1200
1100
1000
900
800
700
600
500
400
300
200
100
0
ACTUAL TAPE VOLUME
OAO
OSO
OGO
ATS
RAE
UNIV-SERIES
ENERGETIC
SAN MARCO
RELAY
IONOSPHERE
1ST
2ND
3RD
4TH
CY-1967
CY-1968
CY-1969
CY-1970

Figure 3

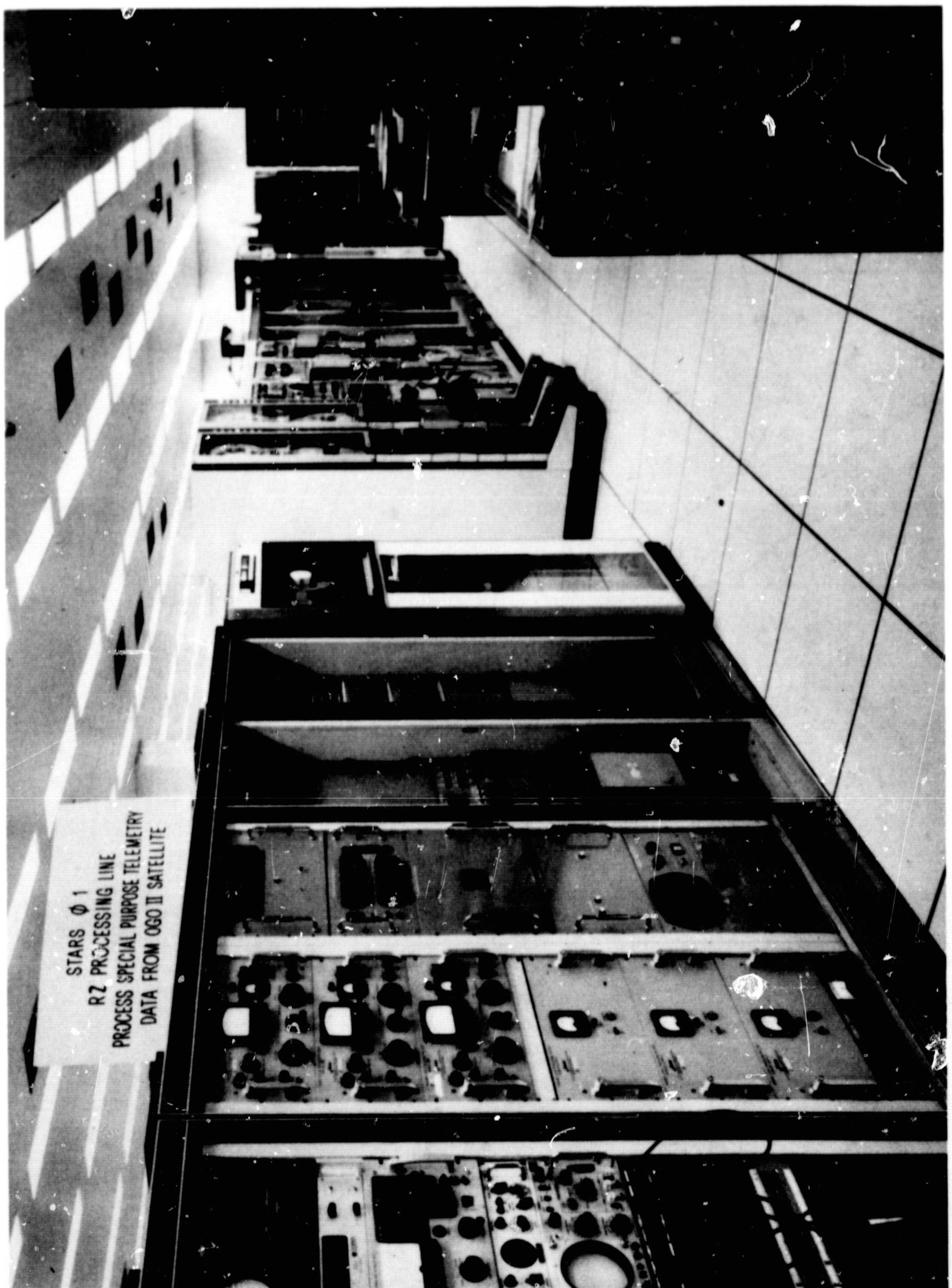
STARS Ø 1
RZ PROCESSING LINE
PROCESS SPECIAL PURPOSE TELEMETRY
DATA FROM OGO II SATELLITE

Figure 4

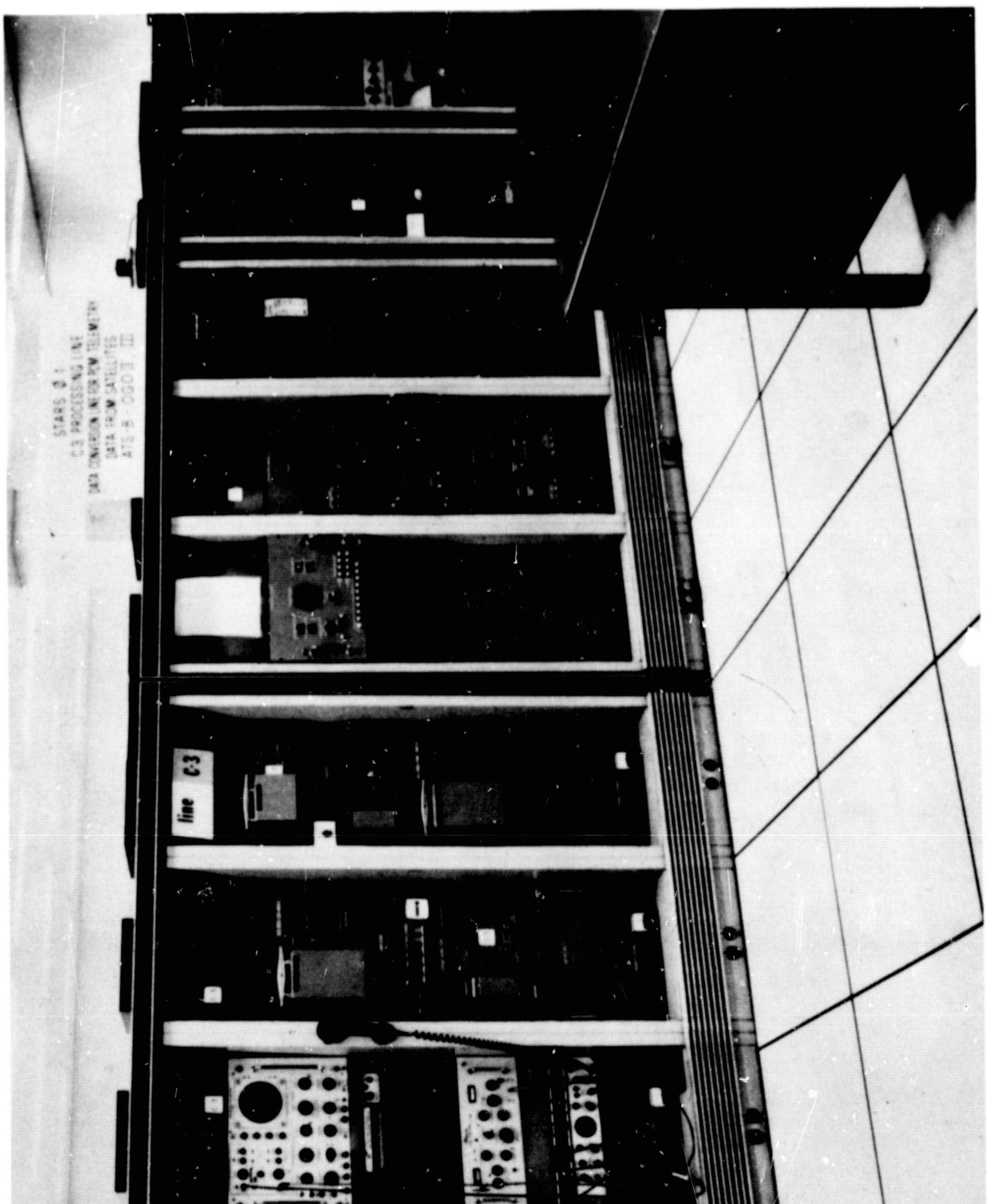
STARS Ø 1
C3 PROCESSING LINE
DATA CONVERSION LINE FOR PCM TELEMETRY
DATA FROM SATELLITES
line C-3

Figure 5

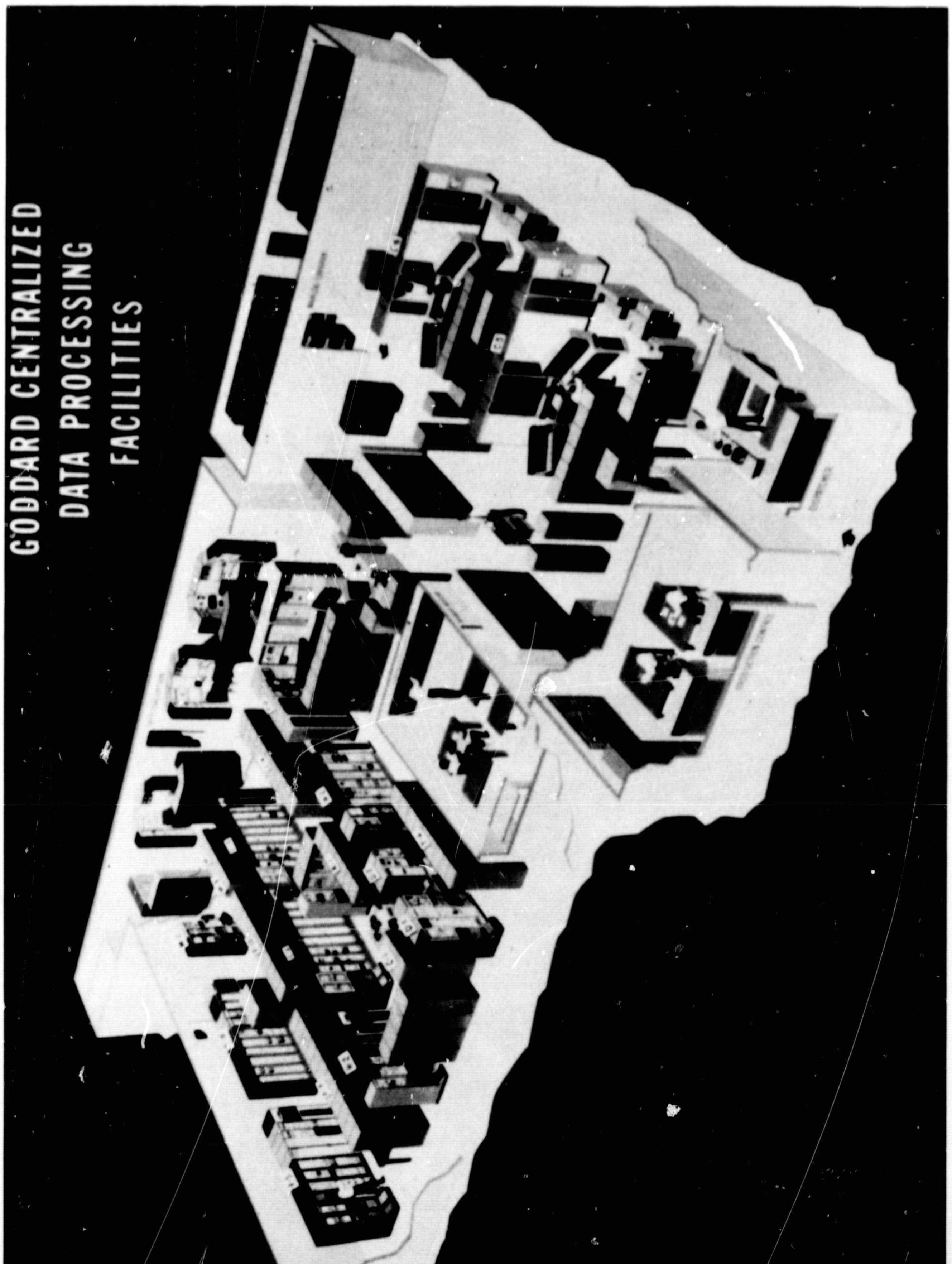

Figure 6

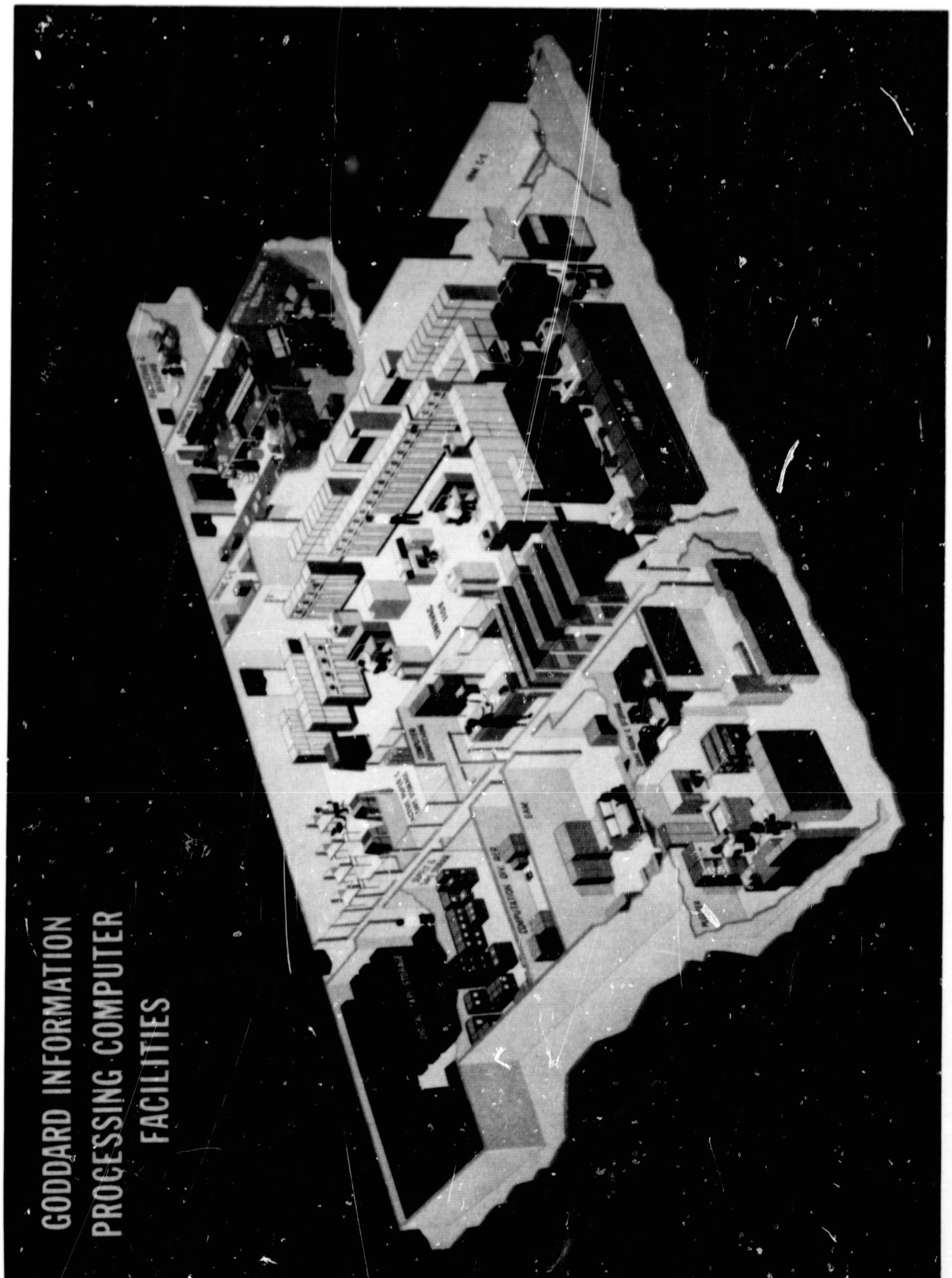

Figure 7

NIMBUS CONTROL CENTER
MRIR DATA FLOW

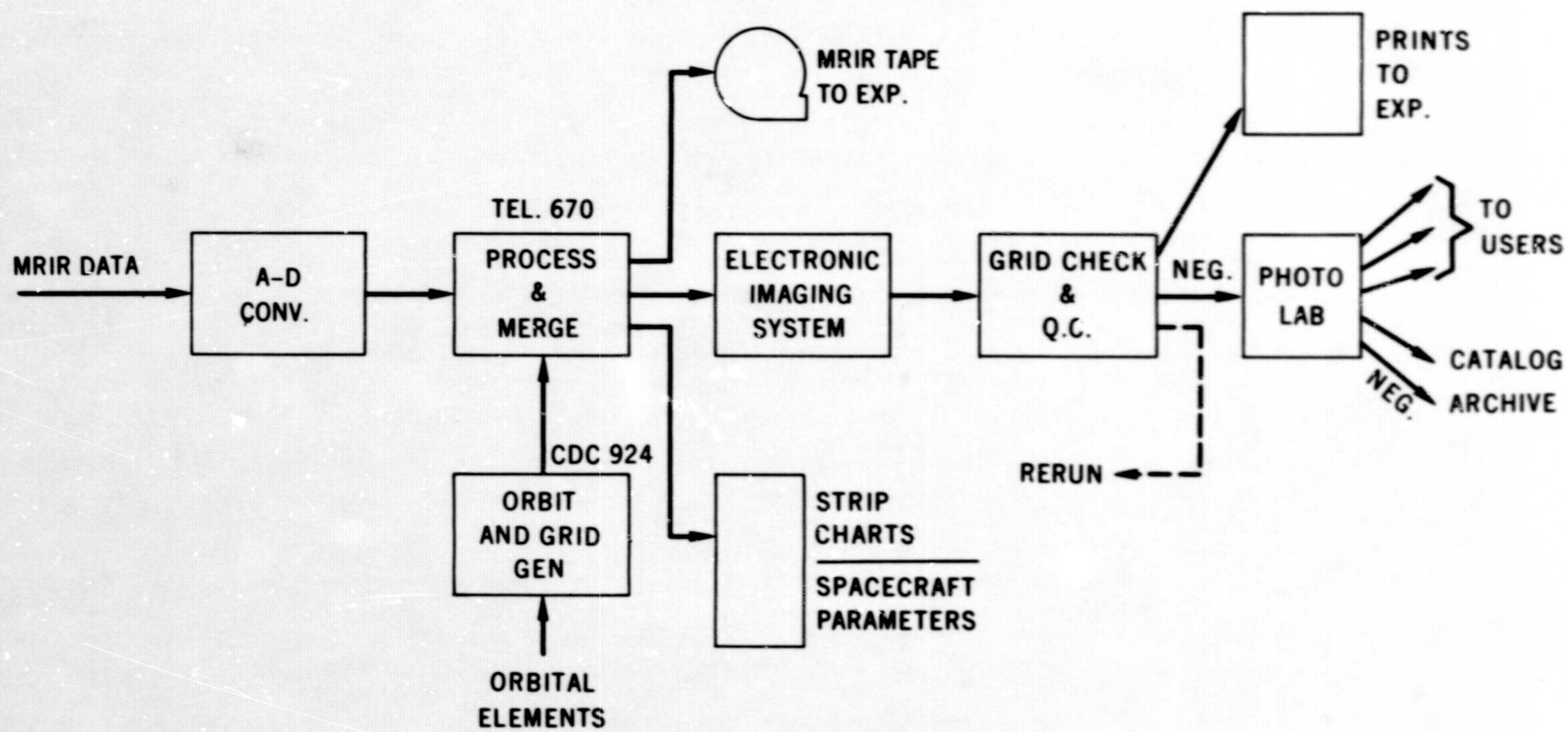

Figure 8

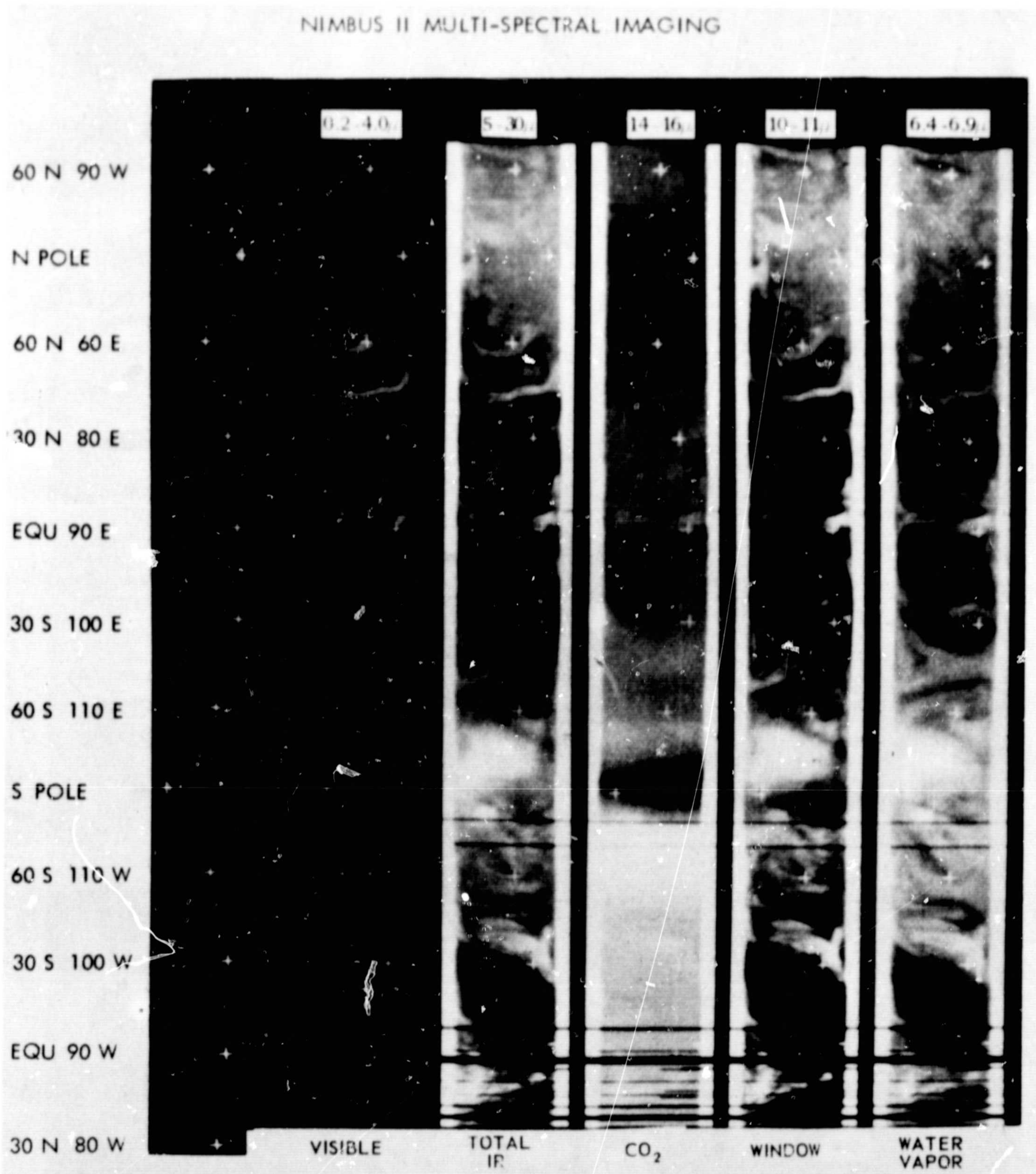
NIMBUS II MULTI-SPECTRAL IMAGING
0.2-4.0μ
5-30μ
14-16μ
10-11μ
6.4-6.9μ
60 N 90 W
N POLE
60 N 60 E
30 N 80 E
EQU 90 E
30 S 100 E
60 S 110 E
S POLE
60 S 110 W
30 S 100 W
EQU 90 W
30 N 80 W
VISIBLE
TOTAL IR
CO_2
WINDOW
WATER VAPOR

Figure 9

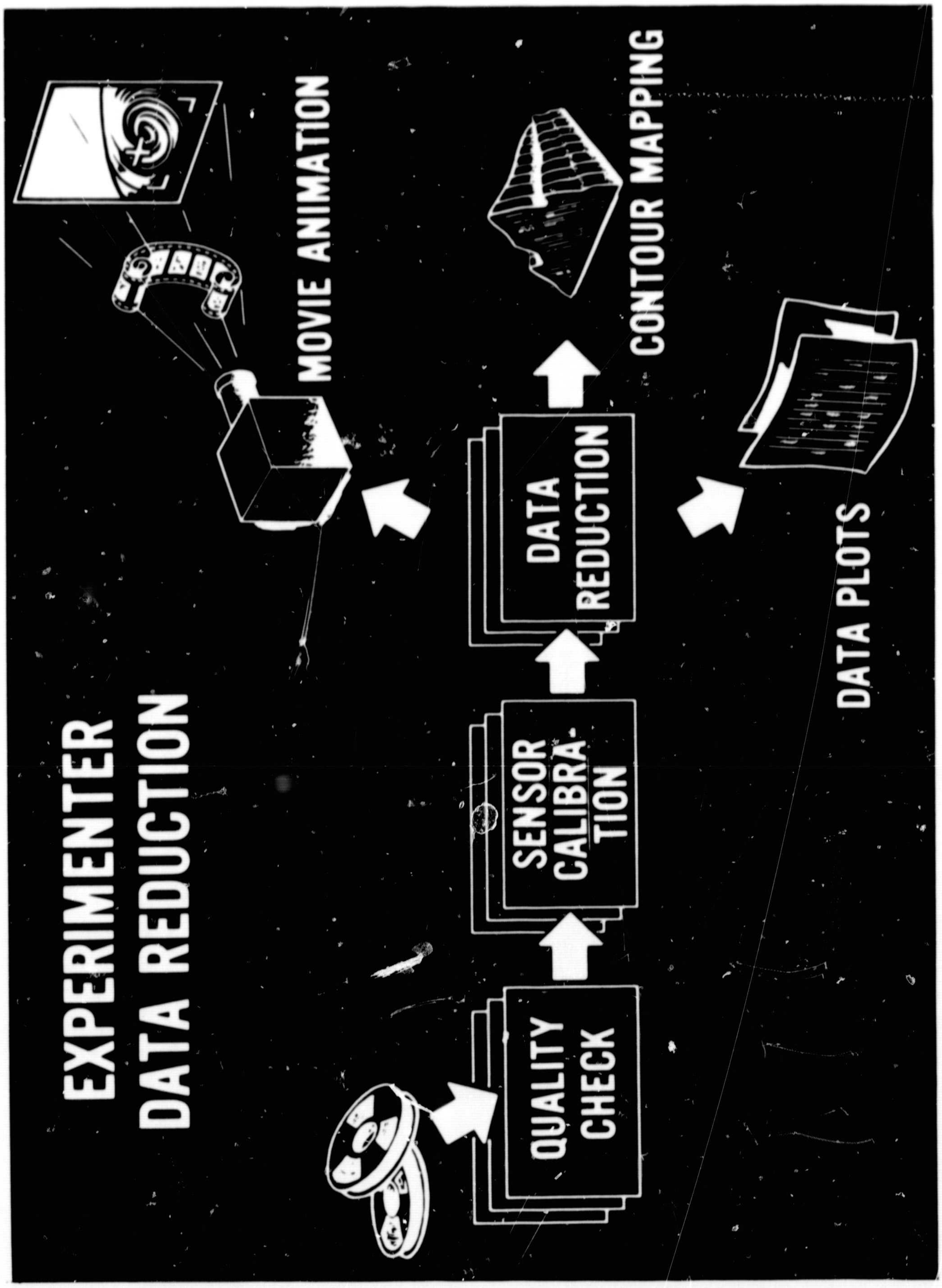
EXPERIMENTER
DATA REDUCTION
QUALITY CHECK
SENSOR CALIBRA-TION
DATA REDUCTION
MOVIE ANIMATION
CONTOUR MAPPING
DATA PLOTS

Figure 10

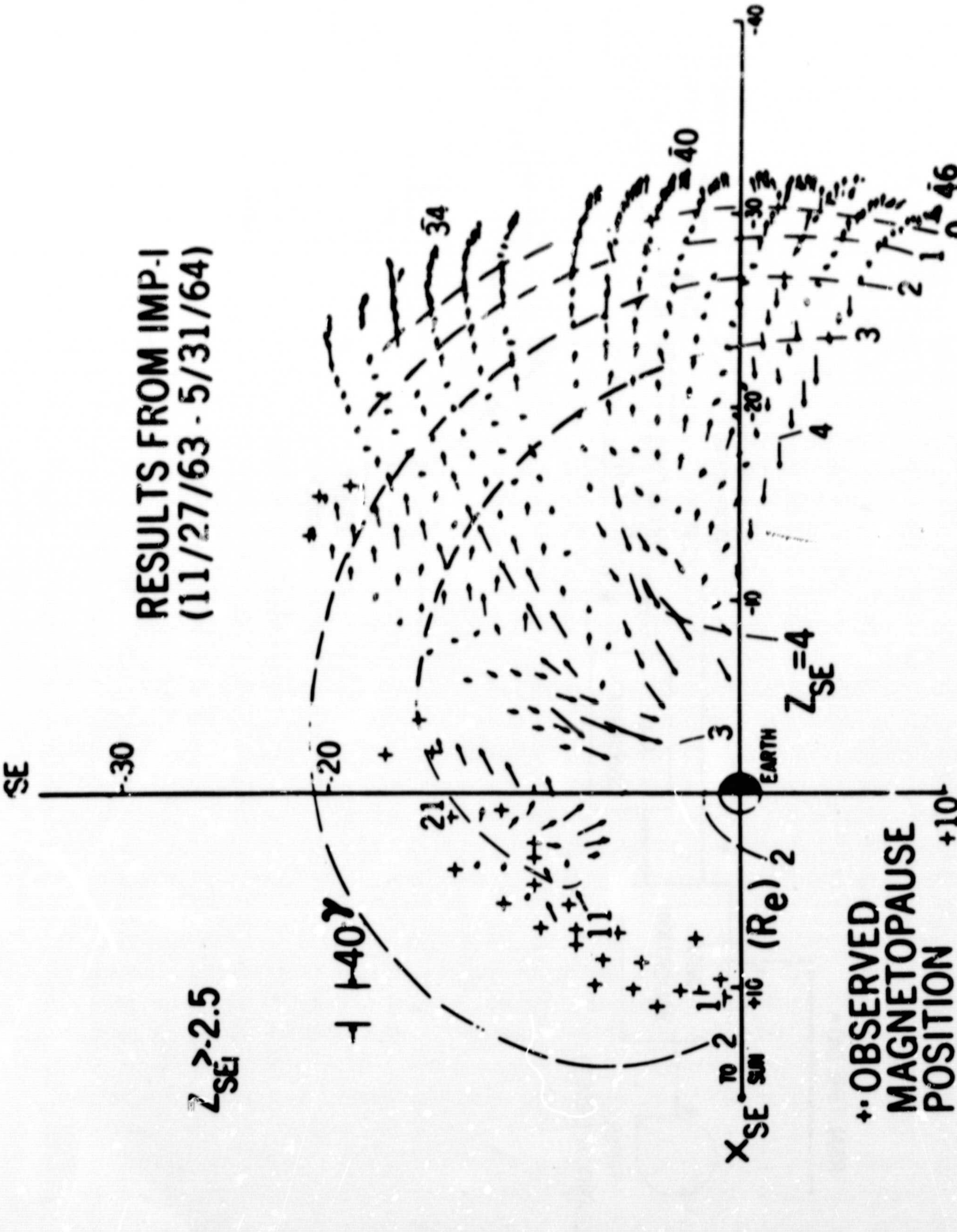
RESULTS FROM IMP-I
(11/27/63 - 5/31/64)
+· OBSERVED
MAGNETOPAUSE
POSITION
(Re)
EARTH
TO SUN
40γ

Figure 11

CENTRAL DATA PROCESSING FACILITY
PROPOSED

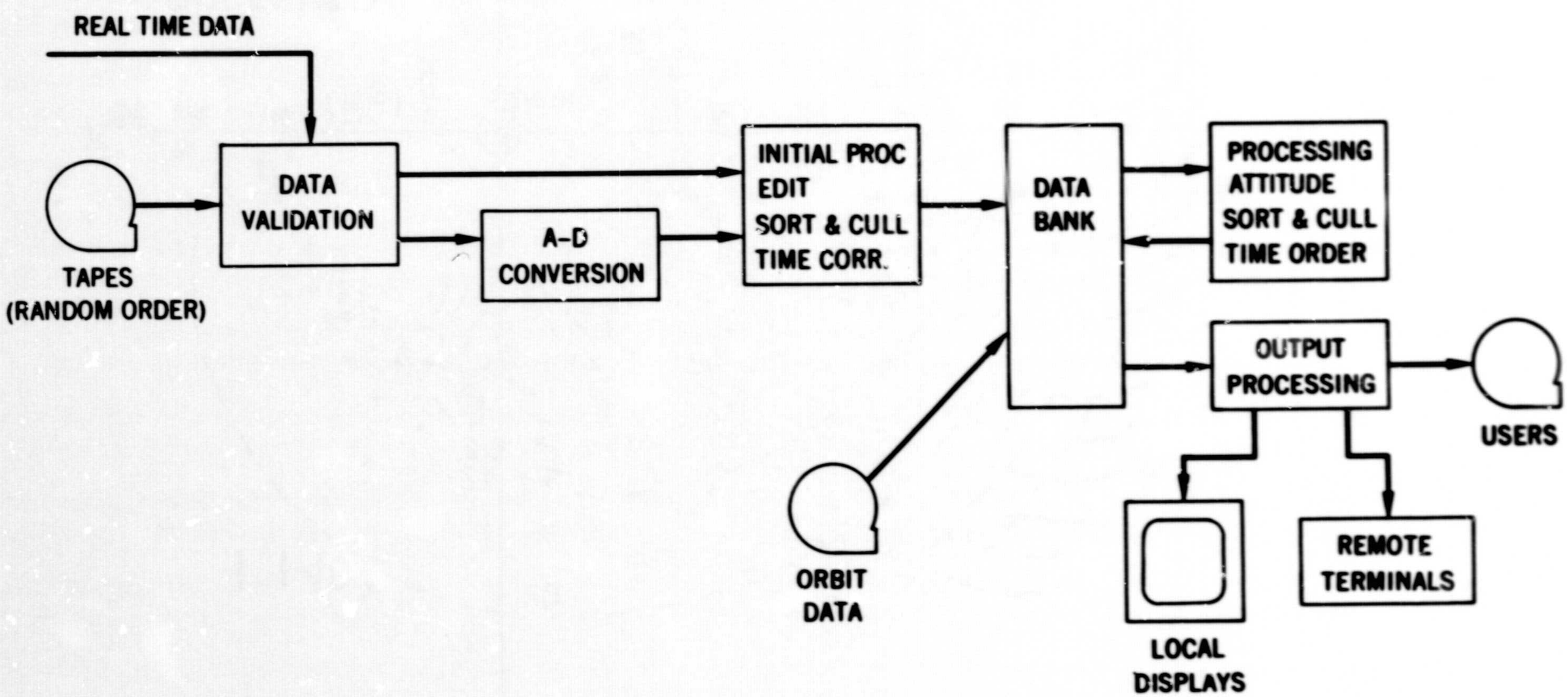

Figure 12

ERTS DATA PROCESSING SYSTEM

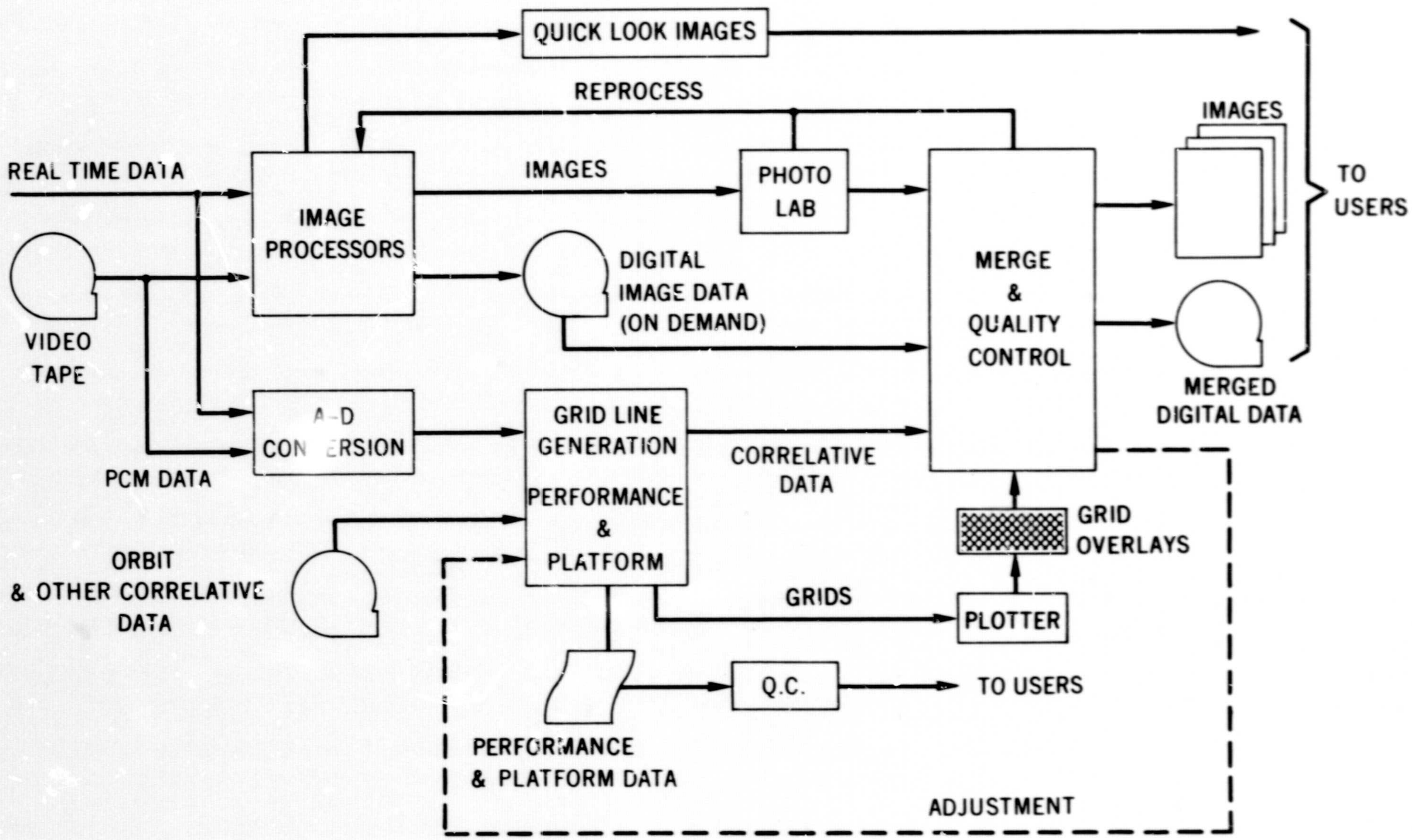

Figure 13