Title
COMPUTER BASED LABORATORY-WIDE DATA ACQUISITION, ANALYSIS AND RETRIEVAL SYSTEM

Authors
Despotakis, J.A.
Fink, R.L.
Itzkowitz, M.S.
et al.

Publication Date
1970-10-01

Presented at Pacific Conference
on Chemistry and Spectroscopy
San Francisco, CA, Oct. 6, 1970

UCRL-20132
Preprint
c. 2

COMPUTER BASED LABORATORY-WIDE DATA ACQUISITION, ANALYSIS AND RETRIEVAL SYSTEM

J. A. Despotakis, R. L. Fink,
M. S. Itzkowitz, and M. P. Klein

October 7, 1970

AEC Contract No. W-7405-eng-48

LAWRENCE RADIATION LABORATORY
UNIVERSITY of CALIFORNIA BERKELEY

DISCLAIMER

Presented at Pacific Conference on
Chemistry & Spectroscopy
San Francisco, October 6, 1970

UCRL-20132

COMPUTER BASED LABORATORY-WIDE DATA ACQUISITION, ANALYSIS AND RETRIEVAL SYSTEM

J. A. Despotakis, R. L. Fink,
M. S. Itzkowitz & M. P. Klein

October 7, 1970

Lawrence Radiation Laboratory
University of California
Berkeley, California

ABSTRACT

The Laboratory of Chemical Biodynamics is a multidisciplinary laboratory whose research interests include biology, biochemistry, chemistry and physics. Spectroscopy of virtually all types covering the frequency range from the audio to gamma rays plays an important role in these activities. Experience with small dedicated computers and similar devices demonstrated their utility but also indicated their limitations. To overcome these limitations a system was designed to provide service to all analytical instruments in the laboratory. A Sigma-2 computer with 32K of core, a 1.5 Mbyte disc, 2 Magtape units, card reader and line printer comprise the center of the system. A PDP-8L with 8K of core and fast paper tape punch and reader provides output to plotter and I/O to the 17 control teletypes and storage scopes distributed through the lab. Data transmission between the individual instruments and computer is digital at a 500 KHz word rate and fully duplexed. Up to 54 instruments may be connected with 8 operating simultaneously-expandable to 32-in foreground while processing and manipulation of previously acquired data occurs in background. Long term data storage and retrieval are design goals, as is remote batch terminal operation with the major computer center in the Lawrence Radiation Laboratory.

The laboratory of Chemical Biodynamics is a research and teaching institution of the University of California, Berkeley, as well as an operating group of the Lawrence Radiation Laboratory at Berkeley. Its staff consists of about ten senior scientists, 20-25 graduate students, a like number of post-doctoral fellows and various support personnel and visitors.

Research at the laboratory is conducted in the general fields of Chemistry, Physics, Biology, and Biochemistry, with particular effort in the areas of spectroscopy, photosynthesis, genetics and molecular biology. As a natural consequence of this research, the laboratory has become populated with a great many different instruments, some commercial and many locally designed and built. Among these are: IR, visible and UV spectrometers, spectropolarimeters, mass spectrometers, CD and MCD spectrometers, fluorimeters, electron and nuclear magnetic resonance spectrometers, gas and vapor chromatographs, and a large number of specialized radiation detectors.

As time has progressed it has become clear that having the data from all or, at least many, of these instruments in digital form would be a Good Thing. There are two reasons why we felt this necessary: first, one can obtain better data by digitizing an output signal from an instrument than one can read from a chart. A cleverly, or perhaps even properly, designed digital interface to a spectrometer, for example, could use the signal from a photo-multiplier tube, and, therefore, eliminate both systematic errors and random noise associated with the rest of the electronics in the machine. Also, having the data in digital form allows rather simple repetitive scan algorithms for signal-averaging, enabling one to extract real signals from a great deal of noise.

A second reason for obtaining digital data is the ease with which such data may be massaged to yield interesting information. Various processes ranging from simple conversion to intensive units, to Fourier and Kramers-Kronig transforms, to highly sophisticated curve resolution and pattern recognition techniques may be applied by computer to digital data.

It became clear to us that we had two alternatives to pursue in attempting to convert the output of our various machines to digital form. We could buy a single small computer for each instrument, interface to it a standard output medium such as paper or magnetic tape, leave the computer dedicated to the instruments for data collection and do all processing on a separate large scale computer such as the CDC 6600's available at the main LRL computer center. Alternatively, we could invest in a medium scale computer, to which we would attach all of our instruments, and which would be capable of both taking the data and processing it in some fashion. The main advantage of the first solution is that if one of the dedicated machines goes down all the others would presumably still run; of the second, that roughly speaking you get computer power proportional to the square of the cost: for the same amount of money, we could get a much more powerful larger machine than the sum of all the little machines. In any case, for both economic and operational reasons, we opted for a medium scale machine. At this point, we tried to hack out a reasonable set of goals for the overall project before we went out to buy.

First and foremost, the machine had to be capable of collecting data from several experimental devices simultaneously. Ideally data collection would be initiated by an experimenter from a teletype adjacent to his machine. There also had to be provided some kind of monitoring scheme so that the experimenter could examine his results part way through and decide whether or not to continue. We also wanted some kind of library system whereby an experimenter could compare the experiment he just did with say, last Tuesday's, or perhaps even last year's. The data from an experiment should also be available for complicated data massaging which would be run as a low priority background job. We also thought it would be nice if some kind of interactive processing was available so that a user could compare his actual experimental results with various theoretical curves and vary parameters until he got a good fit, all from his teletype. And, as long as we were at it, we thought we wanted to be able to run certain elementary calculations from a teletype, such as a BASIC program, and do moderately sophisticated FORTRAN processing in background.

We then began looking for a machine. After some investigation, we decided to purchase a system based on a XDS Sigma-2 computer, primarily because the operating system and programs available from the manufacturer was so much superior in design and implementation to any other available in its price class. In retrospect, it is clear we made the right choice.

The machine purchased had hardware multiply/divide, eighteen priority interrupts, 32K of 16-bit core storage, a medium speed disk storage of 3/4 M words (since expanded to 1.5M words), a 200 cpm card reader, a 200 lpm line printer and a teletype with slow speed paper tape reader/punch. In order to handle a large number of teletypes without degrading the CPU performance we decided to buy a DEC PDP8/L with 8K of core, interface it to the Sigma-2, and let it do all of the line driving and line buffering necessary. We also attached to the PDP8 a high speed paper tape reader/punch, a CalComp plotter and a number of display scopes for interactive processing consoles. We also attached two IBM compatible 7-track tape drives to the Sigma-2 for large volume storage. Figure 1 shows the configuration of the system.

The most important feature of the hardware of our system is of course, the data transmission network used for controlling and taking data from all the experiments. Incidentally it is also used to drive the display scopes. All data is transmitted as three frames of ten-bit parallel information, and consists of a 6-bit address, various control and parity bits and 16 bits of data. All transmissions are acknowledged when successfully received along a separate one bit line. Automatic error detection and retransmission is performed by the network with no program knowledge or control. On the remote ends of the system, each experimental device is equipped with a box to accept commands addressed to it, check for errors and transmit data from the A to D converters in the instrument. At the local end of the network, there are sixteen experiment sub-channels, each of which may be dynamically assigned to any instrument under program control. Hardware within the sub-channels accepts incoming data and stores it within either of two program-specified core buffers, and automatically switches between buffers when one is full and informs the program so that the buffer contents may be dumped onto disk. Although all the hardware for 16 channels will be built, core limitations presently require us to limit operation to 8 concurrent sub-channels. The total data transmission rate for all devices is about 500KHz.

The software for the system breaks logically into four parts as shown in Figure 2. The monitor is responsible for all I/O on the system, including the remote I/O devices actually attached to the PDP-8. Of course, the software system on the PDP-8 is inaccessible to any user on the Sigma-2 and functions as an extension of the monitor. The I/O devices on the PDP-8 are available to both the foreground (i.e. timesharing and data collection) and the background programs, and must be reserved via the monitor to whoever wants to use them. For some devices, in particular teletypes, there really is no conflict of interest, and they are accessible on a line at a time basis to either. All teletype communication is done through a two character unique ID, so that messages may be readily identified. The background has the special ID "$$", while foreground users can specify any two character alphanumeric ID when they log onto the system. The monitor is also responsible for keeping track of ID's and routing any input messages from an ID to the appropriate program.

In order to implement the library system we alluded to earlier, it was necessary to set up a fairly sophisticated file management system to allocate the disk. About one-third of the disk is now under the aegis of the file system, and can be divided into arbitrarily sized files, both permanent and temporary, available to either the foreground or the background or both. Files are permanent in that they will survive intact over dead-starts of the machine. I/O on these files is done by a monitor service request, similar to that necessary for any other kind of I/O; in addition the file management system will schedule activity on the disk so that foreground (real-time) usage of the disk has priority over any background usage. I/O on the PDP-8 devices is also scheduled to give the foreground priority.

All foreground processing in the machine is controlled from one of the eighteen teletypes available and is initiated by a log-on request for a two character ID. All subsequent communication is, as we mentioned before, through this ID, and output messages are routed to the last teletype used for input from that ID. Thus no experimenter is tied down to a specific teletype, but may, for example, initiate his experiment from one teletype, and then reroute his output to the teletype in his office for further processing.

After logging onto the system, the user then specifies which of the two foreground sub-systems he wishes to be connected to: data collection or time-sharing. For data collection, the user would then specify the instrument he wishes to run, and type in any parameters he might need to specify the run; alternatively, on some instruments, parameters are entered by switch settings on the instrument panel. At this point the machine will inform him whether or not there are enough core, disk and/or sub-channels available, and if so, on command either from the teletype or a pushbutton, the computer will start driving his instrument and collecting data. All the while data is being taken, the user may type in various commentary about the experiment, which will be stored as part of the information on the final experiment file for that run. At the end of the experiment, which may come from passing a preset boundary in a scan, or the user's early termination or abort of the experiment, or an unrecoverable data transmission error, the user is asked if he wishes the experiment to be inserted in the permanent library. If so, the data is converted to a standard file format, merged with the commentary file, and stored on the disk as a permanent experiment; at some time later, it is copied to a tape for permanent archival storage. At any time in the future, that experiment is available to any user, although the data stored may never be changed. Experiments are kept on the disk as long as space is available; eventually they are purged on a least-used basis, but may be reloaded from tape at any time by operator command.

The other major subsystem available to foreground users is a limited type of time-sharing facility. The main feature of this system are accessibility to all current and past experimental data, including data from experiments still in progress. As the needs become obvious, various spectra simulation and comparison programs will become available under the time-sharing system. A BASIC compiler is being written to allow some limited arithmetic processing. At the moment only four users may be accommodated in the time-sharing system at one time, but this number will be expanded as the need arises. Because of core limitations, all time-sharing processing is done in the background area, with the background job in progress being rolled out to process every line input to the time-sharing system. Whenever the core is not needed by a foreground program, the background is restarted automatically.

Normal usage of the background is accomplished by card input and printer output, although all the other I/O devices are available. The background also has available to it most of the interactive facilities through its permanent ID. Both the hardware and the software in the system conspire to prevent any background job from disturbing the operation of the foreground under any circumstances. Available to the background users are a slightly restrictive dialect of FORTRAN IV, a quite sophisticated assembler and loader, a set of arithmetic and logical subroutines, and various locally written processors for handling certain kinds of data. The compiler and loader system are set up well enough so that the batch system can handle almost any problem that an IBM 7094 could handle, although the effective computation speed is about 1/10th that of a 7094. All of the background operations proceed while the foreground is running, and are delayed only when the core is required for one of the foreground processes.

As far as economics and time-scale is concerned, the selection of the main computer was begun in about January, 1969 and the machine was delivered in September of that year. Design and construction of all of the locally built hardware was begun in about the summer of 1969, and the first experiment is expected to be operational in this coming January. A good many people have been using the batch background facility of the machine almost from the time it came in the door, and to date about four man-years of programming and software-design have been expended. Another three are anticipated to complete the time-sharing and data collection subsystems, and the first more-or-less complete system is expected to be operational by the coming summer. All of the hardware design was done by John Despotakis and the total hardware cost is expected to run about three hundred thousand dollars.

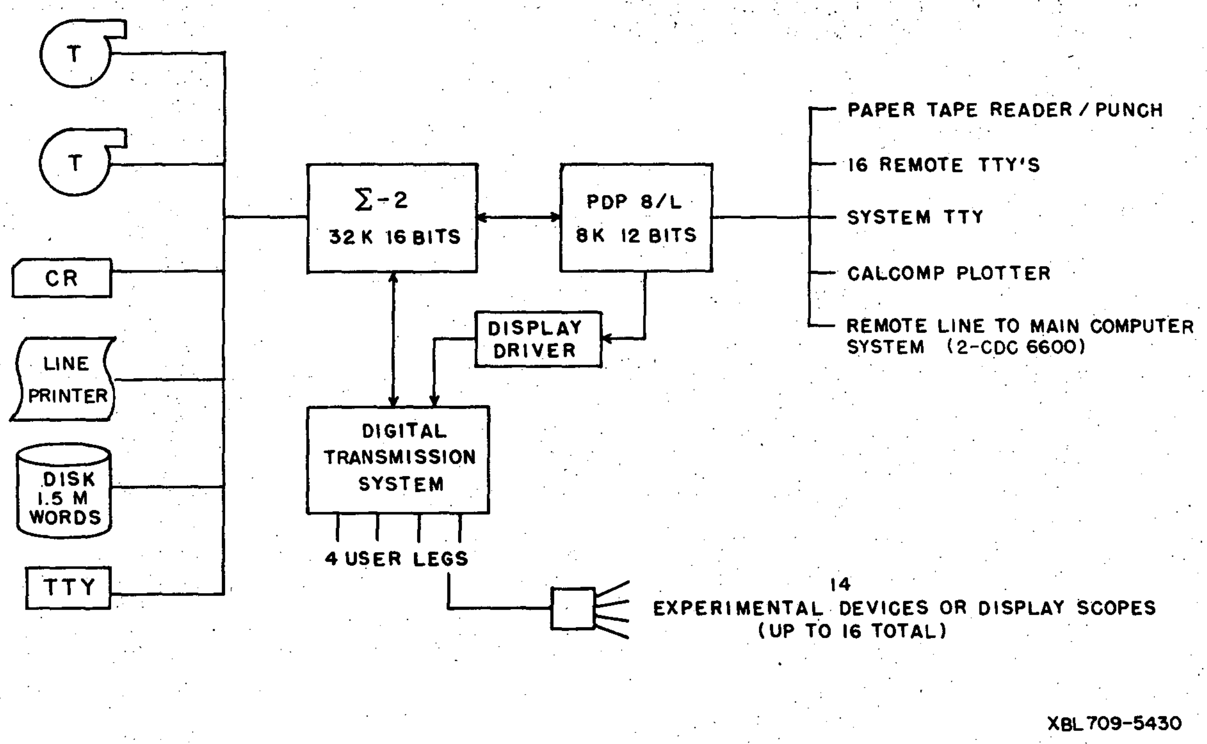
T
T
CR
LINE
PRINTER
DISK
1.5 M
WORDS
TTY
Σ-2
32 K 16 BITS
PDP 8/L
8K 12 BITS
DISPLAY
DRIVER
DIGITAL
TRANSMISSION
SYSTEM
4 USER LEGS
PAPER TAPE READER / PUNCH
16 REMOTE TTY'S
SYSTEM TTY
CALCOMP PLOTTER
REMOTE LINE TO MAIN COMPUTER
SYSTEM (2-CDC 6600)
14
EXPERIMENTAL DEVICES OR DISPLAY SCOPES
(UP TO 16 TOTAL)
XBL 709-5430

Fig. 1

SYSTEM ORGANIZATION

		DEDICATED CORE		AVAILABLE CORE
I.	MONITOR SCHEDULING & SERVICE ROUTINES	11 K		
II.	REAL TIME DATA COLLECTION	6 K	+	14 K
III.	TIME-SHARING SYSTEM	1 K	+	14 K
IV.	BATCH OPERATIONS			14 K

K = 1024 WORDS

Figure 2

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