Computers in Chemical and Biochemical Research

EDITED BY

C. E. KLOPFENSTEIN / *C. L. WILKINS*

Department of Chemistry
University of Oregon
Eugene, Oregon

Department of Chemistry
University of Nebraska
Lincoln, Nebraska

VOLUME 1

ACADEMIC PRESS Ne

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 74-187252

PRINTED IN THE UNITED STATES OF AMERICA

Contents

A Practical Computer-Based Information Retrieval System

P. F. Rusch and J. J. Lagowski

Training Chemists in Laboratory Computing*

S. P. PERONE

Department of Chemistry, Purdue University, Lafayette, Indiana

* Reprinted in part from Perone (1970) and Perone and Eagleston (1971), by permission of the copyright owner.

I. A Postgraduate Laboratory Course on Digital Computers in Chemical Instrumentation

The chemist today finds himself in the midst of a dramatic revolution in scientific instrumentation. Digital electronics and laboratory-scale digital computers appear destined soon to dominate the whole area of chemical experimentation. This revolution has come about so abruptly that the practicing scientist recognizes a very real technological gap in his background. He realizes that computerized instrumentation can provide benefits in routine and research work—but he does not know how to get started. To compound his frustration, he even finds it difficult to communicate with the experts and computer manufacturers. Not only is the *jargon* of digital computer technology quite foreign to the uninitiated, but the fundamental concepts of digital instrumentation are basically unfamiliar. The scientist is accustomed to thinking of experimental data and instrumentation from an *analog* viewpoint, and the introduction to the digital world is not without some difficulty. Not that the material is difficult—it is really quite simple for the scientist trained in logic—but mastery of the material necessitates the generation of drastically different instrumental and experimental concepts than the scientist is generally accustomed to. The principles, methodology, and jargon of digital instrumentation are so different from those of analog instrumentation that even the instrument-oriented scientist has difficulty mastering the area independently. Moreover, the scientist's ability to use the large data processing computer is of little help in understanding *on-line* applications of the digital computer.

Because scientists have recognized this technological gap, many have taken steps to overcome it. This has required herculean efforts on their part, generally, since very little detailed information is available. Most of what is available is provided by computer manufacturers, and may require thorough familiarization with a particular computer for any appreciation. The most fundamental problem, however, is that an adequate appreciation of this new technology cannot be acquired through reading or attendance at technical meetings. Moreover, it is not sufficient to have a computer available in the laboratory which the scientist can play with in his spare time. Nor is it sufficient to develop programming skills which are not oriented toward on-line experimental applications. The most efficient way to develop the technological skills necessary for the incorporation of the digital computer in laboratory experimentation is by total involvement in an intensive training program which includes substantial "hands-on" experience with digital instrumentation and the digital computer in an experimental on-line environment.

Thus, a need has existed for some time now to provide an educational program concerned with digital computer instrumentation for the chemistry laboratory. This type program is urgently sought by the practicing scientist—i.e., one who has completed his formal education. Moreover, this type program should be particularly valuable to our future scientists—the graduate and undergraduate students in our colleges and universities. Such a program has been instituted in the Chemistry Department at Purdue University under the direction of this author. It is the purpose of this article to describe the basic philosophy of this program as well as to provide the reader with many of the details of its operation. In addition, an evaluation of the results of this program will be presented.

A. On-Line Computerized Experimentation in the Laboratory

To put this entire presentation in perspective, it is necessary to provide some background regarding on-line digital computer technology. A common characteristic of the laboratory-scale digital computer is a relatively small core memory, usually providing 4 K or 8 K of random access storage of digital data and programs. A memory of this size, however, can be more than adequate for laboratory control operations, and with careful programming, some fairly sophisticated data processing.

A small digital computer may have an *instruction set* which includes the order of 50–100 *machine language* instructions. Each of these instructions corresponds to a specific binary coding which, when decoded by the computer, results in the execution of a fairly simple arithmetic or logical step. Examples of some simple machine operations are (1) binary addition of a datum in some memory location to the contents of an arithmetic register (accumulator);

(2) transfer of the contents of an accumulator to a memory location (and vice versa); (3) rotation of the binary digits ("bits") of the accumulator contents to the left or right; (4) and the application of logical tests such as determining if the accumulator is zero, non-zero, odd, even, positive, negative, etc.

Obviously, the repertoire of machine instructions includes some rather elementary operations. However, by developing appropriate "programs" composed of many of these elementary operations, the most sophisticated mathematical computations can be carried out. Since the computer can execute instructions so rapidly—the order of 10^5–10^6 instructions per second—it can complete complex computations with fantastic speed. For example, a program to multiply two n-bit integer values might require a program including 50 or 60 statements, but may require only about 200 μsec for completion. Thus, some 5000 multiplications per second could be accomplished. With hardware arithmetic units, perhaps 100,000 multiplications per second could be completed.

Thus, the digital computer is a very simple-minded device, which must be told how to accomplish even the most fundamental computations, but which can accomplish these operations with blinding speed. Moreover, it is a tireless machine that will be content to calculate endlessly and consistently. It is also a very versatile device, since it is programmable and capable of accomplishing an infinite variety of computational, logical, or control operations. Finally, it is a device that can (in fact, *must*) communicate in a variety of ways with the outside world. It is this characteristic that defines the computer as a general-purpose experimental device.

1. *Off-Line Computers*

The computer configuration with which most scientists are familiar is the *off-line* system. This configuration is diagrammed in Fig. 1. To use the computer in this configuration, the scientist typically will write a data processing program in FORTRAN or some other higher level computer language, run the experiment(s), manually tabulate the data from a strip chart recorder or oscilloscope trace, transfer the tabulated data to punched cards, add the data cards to the deck of program cards, transport the combined card deck to the computer center for processing, and then wait until the program has been executed and the results printed.

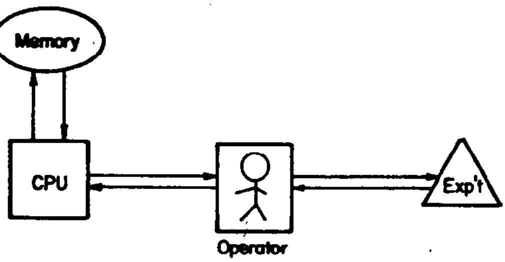

Fig. 1. Typical off-line computer configuration.

Turn-around times may vary from a few minutes to a few days, depending on the capacity of the computer facility and the number of users.

There are many variations of the above description of off-line computer usage. For example, the experimental system may include automatic data acquisition and digitizing devices which might store data on punched card, punched tape, or magnetic tape buffers which can then be transmitted more conveniently to the remote computer center. Alternatively, the laboratory may be equipped with a remote terminal (such as a Teletype) through which the investigator may enter his programs and data, as well as receive a printout of results from the computer.

The important common characteristic of all off-line computer systems, however, is that the experimental data are transmitted to the computer through some intermediate storage medium, and they are processed after some finite time delay has occurred. Depending on the modes of data acquisition and transmission to the computer, the turn-around time of the computer facility, and the speed with which the investigator can interpret the resulting print-out, the time delay for experimental modifications based on the results of previous experiments can be excessive. Should this *reaction time* be a critical factor, an off-line computer facility may be inadequate for the particular experimental studies.

2. *On-Line Computers*

For the investigator who requires very rapid or instantaneous results from his computer system—for whatever reason—the solution may be to employ an on-line computer system. Figure 2 presents a block diagram for a typical on-line computer configuration. The most important distinction of this configuration is that there

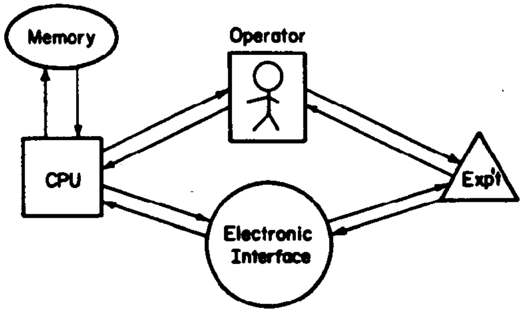

Fig. 2. Typical on-line computer configuration.

is a *direct* line of communication between the experiment and the computer. The line of communication is through an electronic interface. Data are acquired under computer control or supervision, and the program for data processing is either resident in memory or immediately readable into memory to provide for very rapid completion of the computational tasks. Results may be made available to the investigator quickly by means of teletype, line printer, oscilloscope, or other form of printout. In addition, the computer may be programmed to communicate directly with the experiment by controlling electronic or electromechanical devices—such as solid state switches, relays, stepping motors, servo motors—or any other devices which can be activated by voltage level changes.

The advantages of on-line computer operation can be summarized as the following: *Elimination of the middle man,* and the concomitant substitution of an electronic interface between the computer and the experimental system, which is much more compatible with the computer's characteristics and does not suffer from the inherent inadequacies of the human as a communication link. For example, the computer can accept input data at rates of the order of 10^5–10^6 per second, and can instantaneously transmit control information or commands back to the experimental system. If this line of communication were handled through the human investigator, as in the typical off-line configuration, the response time of the overall experimental and data processing system would be many orders of magnitude slower. Moreover, the electronic communication link is generally a more reliable, tireless, objective, and accurate information transfer medium.

The possibility of *direct computer control of the experiment* is a second advantage associated with on-line computer operation. This facility allows the complex logical and decision-making capabilities of the computer to be implemented for an infinite variety of laboratory automation or experimental design problems.

A third advantage of the on-line computer is that *real-time interaction between computer and experiment* is possible. That is, because the computer can make computations and decisions at speeds exceeding most ordinary data acquisition rates, it is possible for the computer to execute experimental control modifications before a given experiment has reached completion. These "real-time" operations allow the possibility of using the computer to monitor the progress of an experiment and modify that experiment to follow a pathway which provides more nearly optimum conditions and which could only have been arrived at as a result of some observation of the initial course of the experiment.

A *breakdown of the logistical barriers of the remote computer system* is an additional advantage of on-line computer operation. Because of the instantaneous, direct communication link between the experiment and the computer, the physical location of the computer with respect to the laboratory is relatively unimportant. Moreover, the mechanical and logistical roadblocks typically imposed by the computer facility toward the off-line introduction of data are irrelevant.

One very important aspect of using computers for on-line experimental work that should be emphasized here, however, is that some, if not all, of the programming must be done in machine (or assembly) language. That is, higher level languages, such as FORTRAN, do not have the capability for on-line operation with experimental systems. Moreover, higher level languages tend to be very inefficient in the use of memory space and in execution time. Both of these factors are critical when small computers are used for on-line work. Thus, the scientist wanting to use the small digital computer in the laboratory must become competent in programming in the fundamental machine language—a formidable task.

B. An Educational Program in Computerized Instrumentation

The widespread use of digital computers on-line in chemistry laboratories will certainly revolutionize routine laboratory methodology. Today's scientist must learn to live with the digital computer and to use it profitably for laboratory work. On this premise we have developed at Purdue an intensive 3-week summer short course entitled "Digital Computers in Chemical Instrumentation." The course is designed for practicing scientists in colleges, universities, industrial laboratories, government laboratories, and hospitals, to provide an introduction to *digital instrumentation* and the technology of digital computer implementation in the laboratory. It is the details of this summer program which will be discussed here.

1. *Objectives*

The purpose of the 3-week intensive course developed at Purdue was to afford participants an opportunity to acquire a familiarity with digital instrumentation and digital computers in the laboratory. The course was designed such that students could develop, in a short period of time, the ability to write machine-language programs, to utilize the computer for typical laboratory input/output functions, and to design and implement software and hardware for interfacing real experimental systems to the computer for online operation. An essential feature of the course had to be the availability of several digital computers and many digital logic modules to provide a large amount of hands-on laboratory experience for each participant. To make the experience most worthwhile to the participants, the lecture and laboratory material were developed within the context of applications in chemistry and related sciences. (The course material was based on the assumption that participants had no previous experience with computers, digital logic, or electronics, but were generally familiar with chemical instrumentation and laboratory methodology.)

2. *Approach*

The basic philosophy of the 3-week summer course is simply to share with the participants the background, technique, and experience in computer instrumentation developed at Purdue and elsewhere by the faculty and graduate instructors of the course. For the participants to acquire the desired skills and background in the 3-week period, it is absolutely necessary to provide for as much individual access to computing equipment and digital instrumentation as possible and practicable. To provide this "hands-on" access to the equipment in the most efficient manner, a great deal of preparation outside the laboratory is required. Most importantly, a large ratio of teaching staff to students is provided, so that students can obtain help with a minimum of delay. At least one instructor for every four students in lab is the normal working ratio at all times.

The significance of the emphasis put on the efficient use of the available computing equipment cannot be overemphasized. The natural tendency of the novice programmer is to sit down at the computer console and try to compose a program. The students in our course learn very quickly that they must have their programs fully composed and punched on tape—with perhaps one or two alternates, in case the first does not work—before they even get to the computer console. When programs do not execute properly, they are encouraged to spend a minimum of formal lab time in time-consuming debugging operations. Thus, when a program is to be edited and reexecuted by the computer, a minimum of "hands-on" time will be consumed by the student trying to reason away the logical errors in his program.

Another important aspect of the course is that it is not limited to the concepts, principles, and techniques associated with the "small" digital computer. The formal lecture material discusses large, time-shared systems as well as small, dedicated systems. Moreover, one laboratory assignment involves writing a program and designing an interface to time-share several peripheral devices. The principles learned in this laboratory assignment are extended in lecture to describe the operation of larger time-shared computer systems.

3. *Requirements*

Equipment. An essential requirement for the presentation of the short course on computers in chemical instrumentation is the availability of sufficient hardware to allow adequate hands-on access by each participant. The minimum equipment provided for a twelve-man laboratory session includes: three digital computers with 8 K-word core memories, ASR/33 teletypes, and high-speed paper tape readers. One of these computers is equipped with a high-speed paper tape punch. In addition, each computer is completely equipped for high- or low-speed computer-controlled digital data acquisition with three 10-bit analog to digital converters (ADC) (each capable of data acquisition rates up to 20 or 30 KHz), and three frequency-selectable gated clocks for precise timing and synchronization of data acquisition and other experimental events. Two computers are equipped with high-precision (16-bit), high speed (20 μsec risetime) digital to analog converters (DAC). Two computers are equipped with oscilloscope display units interfaced to the computer through dual eight-bit DAC modules.

In addition to the computer equipment, each twelve-man laboratory has available six digital logic training stations. Each training station is equipped with versatile patchboard devices for assembling and testing digital logic circuits. Systems used for these stations include the Digital Logic Laboratory, Digital Equipment Corp., Maynard, Mass., and the Elite 3-a integrated circuit device tester, El Instruments, Inc., Derbyshire, Conn. Also with each station is a dual-trace wide bandpass oscilloscope. These stations are also used for assignments requiring interface design for computer-interfaced instrumentation.

Other items necessary to completely equip the laboratory include a variety of specially designed patchboard modules and associated electronics for instrumentation interface design, which are described in detail below. Also, at least one off-line ASR/33 teletype per every four students is provided for program preparation on punched paper tape.

Personnel. One of the most subtle, and yet most critical, requirements for presentation of the course described here is the availability of qualified individuals who understand the software and hardware aspects of computerized experimentation in the laboratory, and who can teach the material. At Purdue, we have been fortunate to find such personnel among our graduate students.

The explanation for the presence of qualified personnel here is twofold. We have several staff members who have become actively involved in on-line computer applications in their research, and we offer a formal one-semester course for graduate students designed to provide first-hand familiarity with on-line computer techniques in the chemistry laboratory. (This graduate level course, first offered informally in Spring 1968, formed the basis for the present intensive 3-week summer course.) Thus, by offering both the graduate course and the 3-week summer course, we have developed a unique group of well-trained teaching assistants.

C. Detailed Course Description

1. *Schedule*

The summer course spans three weeks (or 15 working days). The first day is devoted to orientation and the first lab lecture. The next 13 weekdays are devoted to a rigorous schedule of laboratories, formal lectures, lab lectures, and open lab time. Each participant spends 1½ hours in lab lecture, 1½ hours at formal lecture, and 2½ hours in formal laboratory daily. In addition, each student has about two additional hours per day of "open" access to lab equipment so that he might work ahead or catch up. Labs also are open and supervised part-time on weekends. One day is devoted entirely to the demonstration of various on-line systems operating in the Chemistry Department at Purdue (including mass spectrometry, electrochemistry, flash photolysis, gas chromatography, and stopped-flow kinetic spectrophotometry). The last day is devoted to a "mini-symposium" on computer applications in chemistry, to which several expert outside speakers are invited.

2. *Content*

The formal lecture material covers the following three areas: digital logic and digital instrumentation; computer programming for on-line laboratory applications; and hardware and software design for instrumental interfacing, time-sharing, and computer control in the laboratory. A detailed lecture outline for the short course is given in Table I.

The formal laboratory assignments, likewise, were broken down into three categories: basic digital logic design and operation (4 days); basics of assembly language computer programming and operation (4 days); and interface design and computer programming for on-line applications (5 days). Lab lectures (1½ hours per day) were designed to prepare the student in advance for each laboratory assignment. Table II contains a detailed list of laboratory assignments.