programs in new publication methods. From the very beginning, Volume I was offset printed without right justification. Although this typography had general acceptance, feedback was such as to prompt a change to photocomposition for Volume II with right justification for aesthetic reasons. In researching computer-assisted photocomposition by the ACS Publications Research and Development Group, four papers in the November 1965 issue were processed via computer and offset printing. In 1966, this Journal became the first to be produced by computer-assisted photocomposition. By 1976 all ACS publications were computer-assisted photocomposed.

Author and subject indexes for the 1969 issues were computer processed by A. E. Petrarca and W. M. Lay, Department of Computer and Information Science, Ohio State University, by a modified version of the Double-KWIC Coordinate Index. In the mid-1970s, annual author and keyword subject indexes have been produced by CAS from its computer data bank of journal articles.

It has been a great privilege to serve as this Journal's editor for its first 22.5 years. From my perspective, I see a great future for it and for those who practice the art and science of the two disciplines of chemistry this Journal serves.

Computer Hardware and Software in Chemical Information Processing†

JAMES E. RUSH

James E. Rush Associates, Inc., Powell, Ohio 43065

Received March 8, 1985

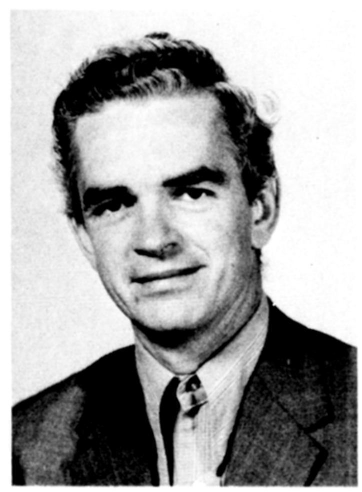
This paper examines the progress of chemists in using computer systems (hardware and software) to sustain, expand, and manage what is perhaps the world's most successful information distribution system, albeit one of many parts. Beginning with the days before stored-program computers—the "mechanization" era—this paper presents an overview of developments in computing and the application of computer and related technology to the handling of chemical information.

INTRODUCTION

The field of chemistry is blessed with perhaps the most comprehensive record of progress of any field of scientific endeavor. Chemists have always been in the van in respect to development and application of modern techniques and technologies for organization and management of chemical information. From the development of regularly published journals (e.g., *Philosophical Transactions of the Royal Society*, 1665), to present-day on-line retrieval systems, chemists have been instrumental in developing, adapting, and using a wide range of techniques and technologies to help facilitate the process of discovery and practice in chemistry and allied fields.

The purpose of this paper is to provide a brief glimpse of the progress chemists have made in handling chemical information during just the past 40–50 years. This period is noteworthy for many reasons. Nearly the whole of polymer chemistry developed during this period, as did nuclear chemistry. The wealth of discoveries in organometallic compounds, the development and application of new analytical tools (e.g., infrared, ultraviolet, and nuclear magnetic resonance spectroscopies and chromatography), important discoveries in biochemistry, great progress in space travel, and much more has occurred during the past 4 or 5 decades. I want to examine just one facet of this exciting period, namely, the development, adaptation, and use of computers in chemical information processing.

Chemical information comprises the entire record of progress in chemistry. Processing of this information has the purpose of organization, coordination, and correlation so as to facilitate further research, development, and practice in chemistry. Chemical information processing may be divided into two broad categories, that which deals with representations of chemical substances and that which deals with nonstructural or textual descriptions of substances and with processes and phenomena related thereto. Since other authors in this issue examine both of these areas in some detail, I will treat them only from the perspective of significant work related to the adaptation and use of computers. I will first review developments in the computer field and then will address some of the significant milestones in both structural and textual in-



James E. Rush received his B.S. in chemistry and mathematics from Central Missouri State University in 1957 and his Ph.D. in chemistry from the University of Missouri in 1962. He is President and a principal shareholder of James E. Rush Associates, Inc. (an Ohio Corp.). He was Director of Research and Development at OCLC, Inc. (1973–1979), and their Director of Research and Coordinator of the Office of Planning and Research (1979–1980). Prior to that, he was an Associate Professor (1968–1973) and is now Adjunct Associate Professor, in the Department of Computer and Information Science at The Ohio State University. In 1980, he was also appointed Adjunct Professor of Library Science at the University of Illinois. He was associated with Chemical Abstracts Service (1962–1968) as an Assistant Editor in the Organic Index Editing department, Assistant Head in the Chemical Information Procedures department, and Technical Liaison Officer. He is actively engaged in research and has numerous publications to his credit on chemical information handling, computing, library automation, and other topics in the field of information storage and retrieval, including two books coauthored with Charles H. Davis.

† Presented in part before the first Computer Secretariat, 189th National Meeting of the American Chemical Society, Miami Beach, FL, April 30, 1985.

0095-2338/85/1625-0140$01.50/0

Table I. Significant Milestones in the Development of Computers

8??	al-Khowarizmi	algorithms
1623	Schickard	mechanical calculator
1642	Pascal	mechanical calculator
1673	von Leibniz	mechanical calculator
1725	Bouchon	loom
1804	Lord Kelvin	harmonic analyzer
1805	Jacquard	loom (punched cards)
1820	de Colmar	arithmometer
1830+	Babbage	difference engine
1830+	Babbage	analytical engine
1853	Scheutz	difference engine
1881	Edison	vacuum tube
1883	Hollerith	punched cards
1897	Michelson/Stratton	harmonic analyzer
1904	Fleming	vacuum tube
1911	Hollerith	Tabulating Machine Co.
1923	Zworykin	cathode ray tube
1924	Hollerith/Watson	IBM
1931	Bush	differential analyzer
1937–1943	Aiken (Harvard)	MARK I (electromechanical)
1940	Bell Telephone Laboratories	complex computer (relays)
1940–1942	Atanasoff/Berry	electronic digital calculator
1941	Zuse	Z3 (electronic computer)
1943	IBM	ASCC
1943	Eckert/Shockley	mercury delay line memory
1944	von Neumann	idea of central storage
1945–1946	Moore School	EDVAC
1945–1946	Moore School	ENIAC
1947	Barden, Brattain, Shockley	transistor
1950	Eckert/Mauchly	BINAC
1951	Eckert/Mauchly	UNIVAC
1953	IBM	701
1959	Texas Instruments/ Fairchild Semiconductor	integrated circuit
1960+	IBM	360 series
1965	Digital Equipment Corp.	PDP/1
1971	Intel Corp.	4004 microprocessor chip
1975	MITS	Altair microcomputer
1976	Radio Shack	TRS-80 Model I
1982	IBM	PC

formation processing by computer. Because of the great wealth of published information to draw upon, I must necessarily be both selective and brief. Even then, the space available to me makes it difficult to do justice to my subject.

COMPUTERS AND COMPUTER TECHNOLOGY

Since computer hardware and software lies at the heart of successful, modern chemical information processing, it is important to examine the developments that have taken place in computer science and technology during the past 40–50 years. To conserve space, I have compressed most of the significant history of computing into several tables. A few brief remarks accompany these tables.

The history of the modern era of computing extends from about 1945 to the present (about 40 years). However, since computing involves the solution of problems by stepwise, sequential procedures, it is possible to trace the history of computing much farther back in time. Table I presents a number of important milestones in the development of the hardware component of computer science and technology. It goes without saying that chemists have played a significant role in much of this development (particularly with the advent of solid-state circuitry). Table II presents milestones in maxicomputer development, while Table III does the same thing for minicomputers. No comparable table is presented for microcomputers because of the great number of and differences between them.

Problems are solved by computer by casting them in the form of algorithms, detailed step-by-step procedures through

Table II. Significant Milestones in Development of Maxicomputers

1949	Renwick	EDSAC
1950	National Physical Laboratory (Turing)	Pilot ACE
1950	National Bureau of Standards	SEAC
1950	MIT (Forrester)	Whirlwind
1951	Eckert/Mauchly	UNIVAC
1953	IBM	701
1955	IBM	702
1956	IBM	704, 705
1960	Remington Rand	UNIVAC 1105
1960	IBM	709, 7090
1960+	IBM	7030 (STRETCH)
1960+	Remington Rand	UNIVAC LARC
1960–1961	IBM	1401, 1410, 7010
1962	Burroughs	D-825 multiprocessor
1962	Scientific Data Systems	910, 920 (interrupts)
1962–1963	IBM	7040, 7044
1964	IBM	System 360
1964	Control Data Corp.	6600
1968	University of Illinois	ILLIAC IV array processor
1969	Xerox Data Systems	Sigma series
1969	Control Data Corp.	7600
1970	IBM	System 370
1970	Goodyear Aerospace	STARAN array processor
1971	RCA	leaves the field
1975	Amdahl Computer Corp.	Amdahl 470 V/6
1975	Cray Research	Cray I
1976	Xerox Data Systems	leaves the field
1977	IBM	30xx series
1979	IBM	43xx series
1980	IBM	3081
1981	Honeywell	DPS 6/92
1982	Cray Research	Cray X-MP series
1983	Burroughs	B7900 series
1984	Amdahl Computer Corp.	Model 5868
1984	Sperry Corp.	1100/90 series

Table III. Significant Milestones in Development of Minicomputers

1959	Digital Equipment Corp.	PDP-1
1964	Digital Equipment Corp.	PDP-6
1974	Digital Equipment Corp.	PDP 11-70
1976	Tandem Computers	T/16 NonStop I
1977	Digital Equipment Corp.	VAX 11/780
1981	Perkin-Elmer Corp.	3210
1981	Wang Laboratories	VS100
1981	Tandem Computers	NonStop II
1982	Data General Corp.	ECLIPSE MV/4000
1982	Prime Computer	2250
1983	Hewlett-Packard	HP 3000 Series 68
1983	Perkin-Elmer Corp	3200 MPS
1983	Tandem Computers	NonStop TXP
1984	Data General Corp.	ECLIPSE MV/10000
1984	Wang Laboratories	VS300
1984	AT&T	Models 3B20A, D, S
1984	NEC	Astra 370VS

which otherwise complicated or complex problems are decomposed into simple ones. Performance of each simple step in the appropriate sequence yields the desired solution. The genesis of algorithms is attributed to a 9th century Arabic mathematician Muhammad ibn-Musa al-Khowarizmi. The term "algorithm" is a corruption of the man's name.[1]

From the 16th century onward, there were many attempts to design and build mechanical calculators. Aside from the abacus and soroban, which are mechanical calculating devices of great age and widespread use, the most significant work on mechanical calculators (or computers) was that of Charles Babbage, an English mathematician.[2–4] Both his difference engine and analytical engine represent concepts of automatic computation that are valid and useful today. But Babbage was before his time, and he failed to complete a fully working model of either machine, largely because of the difficulty of machining parts to proper tolerances and because of lack of support from the British Government (some things never change). Working machines were developed later by Scheutz[5] and still later by Vannevar Bush.[6]

The Jacquard loom was a machine that employed punched cards to control the weaving process. This machine not only revolutionized the textile industry, it created great social upheaval as well (e.g., Luddite rebellion).[7,8]

Herman Hollerith developed a punched-card sorting and tabulating machine for use in tabulating the census of 1890. This machine was the forerunner of most modern punched-card equipment. Hollerith's company became the International Business Machines Corp. (IBM) in 1924.

The history of computers continued to be that of mechanical devices until about 1940. During the period 1940–1950, there was intense work on computing devices. Two parallel paths of development were pursued until about 1945. One, culminating in the automatic sequence controlled calculator (ASCC), was electromechanical;[9] the other, which yielded the ENIAC (electronic numerical integrator and calculator), was electronic.[10] Electronic computers were, as we now know, to win out over all other types (although special-purpose computers of all types are still built and used today).

The first electronic computers were based on vacuum tubes for control of circuitry. The ENIAC contained some 18 000 vacuum tubes, 70 000 resistors, 10 000 capacitors, and 6000 switches. Physical dimensions are also impressive: 100 feet long, 10 feet high, and 3 feet deep. It consumed 140 000 W of power, and it was slow by modern standards: 200 μs for addition and subtraction and 2600 μs for multiplication. Contrast this with my microcomputer, which can add in about 1 μs and multiply in 7–8 μs.

The ENIAC, and all computers to follow, gives real meaning to the concept of quality control. As Goldstine put it: "...we should realize that the proposed machine turned out to contain over 17,000 tubes of 16 different types operating at a fundamental clock rate of 100,000 pulses per second. [That is, one clock pulse every 10 μs.] Thus, once every 10 μs an error would occur if a single one of the 17,000 tubes operated incorrectly; this means that in a single second there were 1.7 billion (1.7×10^9) chances of a failure occurring and in a day about 1.7×10^{14} chances. Put in other words, the contemplated machine had to operate with a probability of malfunction of about 1 part in 10^{14} in order for it to run for 12 hours without error. Man had never made an instrument capable of operating with this degree of fidelity or reliability, and this is why the undertaking was so risky a one and the accomplishment so great."[11]

John von Neumann is credited with making the next significant advancement in computer architecture, namely, internal memory or storage. His idea was to store both programs and data in a memory that was integral to the machine rather than external to it (as was paper tape, punched cards, or magnetic drum). The first memory was built of mercury delay lines.[12] Jay Forrester (MIT) conceived of the idea of the magnetic core,[13] which was both cheaper and much faster than mercury delay lines. It also made possible larger memories.

During this same period, a most profound invention, the transistor, was made by John Bardeen, Walter H. Brattain, and William Shockley in 1947 at Bell Telephone Laboratories. For their work, these men received the Nobel prize in physics in 1956. Shockley later moved West to found Shockley Semiconductor Laboratories (1955), and thus to spawn "Silicon Valley".[14] The development of the transistor paved the way for miniturization and cost reductions, and hence vastly improved cost/performance ratios.

Once solid-state electronic devices could be built, it was inevitable that someone would conceive the idea of fabricating entire circuits on the semiconductor material from which the transistors themselves were made. The idea is credited to a British radar expert G. W. A. Dummer, who presented the idea in a 1952 paper.[15] Integrated circuits (IC), as the concept has come to be called, were first implemented by Jack Kilby of Texas Instruments in 1958 and simultaneously by Robert Noyce of Fairchild Semiconductor (subsequently cofounder of Intel Corp.).[16] Simple integrated circuits were commercially available by 1961. The development of integrated circuits has put tremendous pressure on designers, engineers, and chemists to continuously increase the density of active elements on a single chip (a silicon substrate about 5 mm on a side). Density has doubled nearly every year since integrated circuits were introduced. IBM (among others) has developed a memory chip with over 2 million active elements and a memory capacity of 1 024 000 bits (4 times the capacity of the current 256K-bit chip).

Table IV. Approximate Share of the Market for Business-Class Microcomputers

rank	firm	market share (%) 1983	1984
1	IBM	24	28
2	Apple	30	25
3	Tandy	15	14
4	Hewlett-Packard	11	9
5	Digital Equipment	5	4
6	Compaq	3	4
7	Wang	n/a	4
8	Zenith	2	3
	other	10	9

Integrated circuits made possible the entry of Digital Equipment Corp. into the fray, and thus the birth of the minicomputer. But the truly remarkable result of the invention of integrated circuits is the fabrication of an entire central processing unit (CPU) on a single chip. Intel Corp. (Ted Hoff) marketed the first microprocessor, the Intel 4004, in 1971.[17] Thus, in about 25 years, computers had been reduced from huge collections of discrete components to incredibly small integrated circuit components.

The first commercially available microcomputer was marketed as a kit for hobbyists by MITS Corp. in 1975. Tandy Radio Shack followed quickly with its TRS-80 Model I in 1976. By 1980, the computer revolution was well under way. At present, there are over 200 companies making and marketing general-purpose microcomputers. Table IV gives some idea of the principal players.

It is not possible to overstate the importance of microprocessors and microcomputers. They already permeate virtually every thing we do, and they will become even more important as time passes and new ways are found to use them.

Let me close this discussion of computer hardware by presenting some data on sales and unit costs. Figure 1 illustrates the expected growth in sales of computers (nearly all microcomputers) through 1992. The figure excludes "home" computers. Figures 2 and 3 illustrate the trend in cost reduction of selected system components since 1976. The vertical axes are logarithmic scales. The triangles indicate halving of costs; the dots, actual costs in US dollars.

SOFTWARE

A review of the computer field would not be complete without a brief discussion of software. As already mentioned, the notion of algorithms dates as far back as the 9th century. But the development of specific algorithms for use with computers came much later. Augusta Ada Byron (Countess of Lovelace; daughter of Lord Byron) is credited with being the first "programmer". She collaborated with Charles Babbage (1833+) on documentation of his work and on the development of a program to compute the Bernouli numbers.[19]

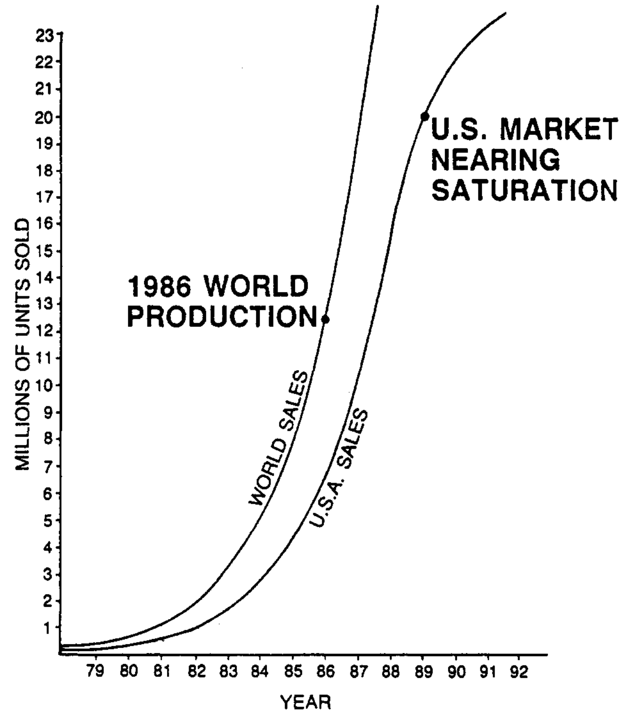


Figure 1.

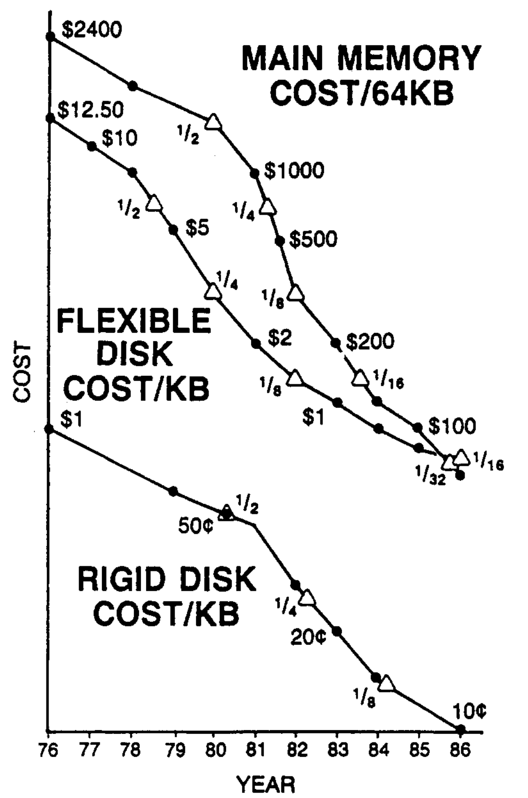


Figure 2.

All programming done prior to the early 1950s had to be done in machine language (e.g., in binary) or in some crude assembler language. Nevertheless, symbolic assembler langauges made programming much easier and less error prone than machine language. But even assembler languages were deemed too inefficient, so higher level languages began to be developed. As indicated in Table V, Grace Hopper (working at UNIVAC) developed the first compiler, called A–0.[20] Hopper also developed several other languages, including Flow-Matic (1956), the first language suitable for business data processing. Also in 1956, FORTRAN, designed by Backus and Ziller,[20a] was introduced by IBM. As Table V

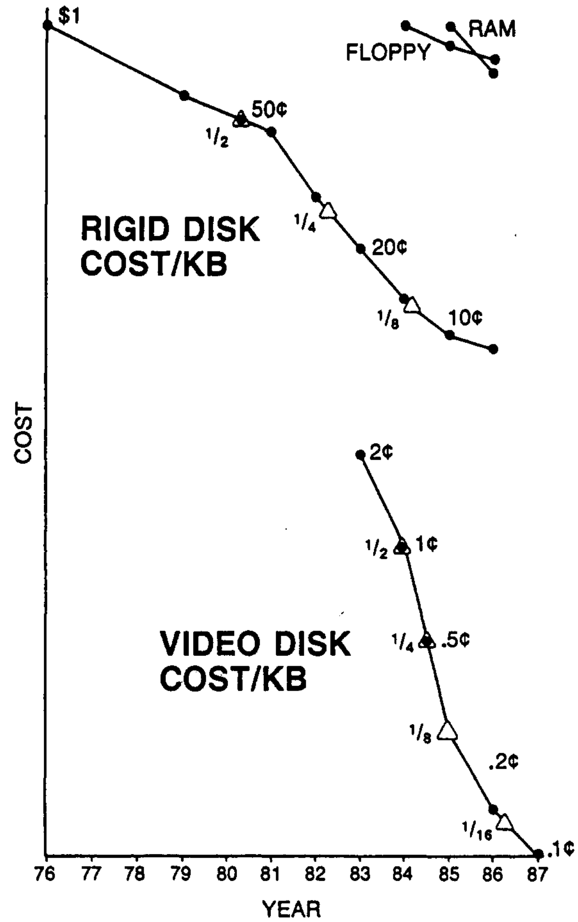


Figure 3.

indicates, a number of other important languages have been developed over the succeeding years, culminating with ADA (named for the Countess of Lovelace) in 1980. Perhaps 300 programming languages have been developed during the past 30 years, but most are simply variants of those listed in Table V.[21]

In addition to programming languages, one must consider operating systems. An operating system is a special-purpose program (or set of programs) that manages and controls all of the facilities of a computer system and makes these facilities available to programmers and users under appropriate conditions. Use and control of memory, peripheral devices, communication facilities, compilers, utilities (e.g., sort/merge), and applications programs are among the features of an operating system. Operating systems may be more or less complex. IBM introduced OS-360 with its line of 360 computers and gave new meaning to the term "job control language" or JCL. Every maxi- and minicomputer manufacturer offers a proprietary operating system. Only three generic operating systems exist for maxi- and minicomputers, PICK, UNIX, and the UCSD p-System. However, as Table VI indicates, the situation is substantially different in the microcomputer realm. There are a wide variety of operating systems available for microcomputers, the most common and widely used of which are those listed in Table VI. The existence of generic operating systems means greater portability of applications software and less dependency on a particular manufacturer's hardware.

Now let me turn to the application of computer hardware and software to the solution of problems of chemical information processing. It is convenient to consider this topic as falling in the domain of application software. Applications are divided into two broad categories, those dealing with representations of chemical structure and those dealing with textual data. Other sorts of applications, such as development

Table V. Milestones in Programming Language Development

1952	A-0	first compiler (Hopper)
1956	FORTRAN	first widely used, high-level language for scientific and engineering applications
1956	Flow-Matic (B-0)	first language suitable for business data processing and with an "English-like" syntax
1957	Comit	first realistic string-handling and pattern-matching language
1958	IPL-V	(Information Processing Language V) the first major language for list processing
1960	COBOL	first language designed by committee; probably the most widely used language today
1960	ALGOL	language developed for specifying algorithms (not a full programming language in its original presentation)
1960	LISP	(List Processor) language for list and functional processing; used in artificial intelligence work
1961	GPSS	(General Purpose System Simulator) made simulation a practical tool
1963	SIMSCRIPT 1.5	discrete simulation language
1965	BASIC	(Beginner's All Purpose Symbolic Instruction Code) replaced JOSS (1960) as interactive language
1965–1966	PL/1	(F-level) general-purpose high-level language combining features of COBOL, FORTRAN, and ALGOL
1967	APL 360	(A Programming Language) provided many high-level operators, permitted short algorithms
1967	SNOBOL 4	string-handling and pattern-matching language for text processing that succeeded Comit
1971	Pascal	introduced new ideas about data typing; provides elegant implementation of many known constructs
1976	C	programming language that serves as the system language for UNIX
1980	ADA	language designed under sponsorship of the U.S. Department of Defense in cooperation with the Defense Ministries of the U.K., France, and West Germany; named for Lady Lovelace (Augusta Ada)

Table VI. Generic Operating Systems for Microcomputers

operating system	microprocessor	word size (bits)	OS type[a]	vendor
CP/M-80	Z80, 8080, 8085	8	SU/ST	Digital Research
CP/M-86	8088, 8086	8/16	SU/ST	
Concurrent CP/M-86	8088, 8086	8/16	SU/MT MU/MT	
MP/M II	Z80, 8080, 8085	8	MU/MT	
MP/M-86	8088, 8086	8/16	MU/MT	
CP/M 68K	68000	16/32	SU/MT	
MS-DOS	8088, 8086	8/16	SU/ST	Microsoft
XENIX	68000, LSI-11, 8086, Z8000	32	MU/MT	
OASIS	Z80	8	SU/ST	Phase One Systems
OASIS	Z80	8	SU/MT	
OASIS-16	8088, 8086, LSI-11, 68000, Z8000	16/32	MU/MT	
UCSD p-System	8088, 8086, TI9900, Z8000	8/16	SU/MT	SofTech Microsystems, Western Digital Corp.
UCSD p-System	68000	16/32	MU/MT	Instrumentation Lab., Inc.
DOS	6502	8	SU/ST	Apple Computer
TRSDOS	Z80	8	SU/ST	Tandy Corp.
TRSDOS-16	68000	16/32	MU/MT	
BASIC	6502	8	SU/ST	Commodore

[a] SU, single user; MU, multiuser; ST, single task; MT, multitask.

of synthetic pathways, spectral analyses, or elucidation of structure, are dealt with by other authors in this issue.

APPLICATIONS OF COMPUTER HARDWARE AND SOFTWARE TO PROCESSING OF CHEMICAL STRUCTURE REPRESENTATIONS

The handling of chemical structure information has been extensively reviewed by Tate,[22] Rush,[23,24] Davis and Rush,[25] Wipke et al.,[26] Stobaugh,[27] Holm et al.,[28] Ash and Hyde,[29] and Lynch et al.[30] The surveys prepared by the National Academy of Sciences-National Research Council Committee on Modern Methods of Handling Chemical Information[31a,b] are also important works. In this paper, I will touch on some of the seminal work in chemical structure processing. Such processing antedates electronic computers and was, in the 1930s and 1940s, influenced by the availability of a variety of mechanical devices for punching, notching, or drilling cards of one type or another. The available technology led to the development of a variety of fragment codes, text-coding schemes, and filing, sorting, and retrieval techniques. Work during this period has continued to have considerable influence on many aspects of modern chemical information processing. The bulk of this work is admirably covered in the NAS-NRC surveys[31a,b] and in two editions of a work entitled "Punched Cards".[32a,b] This work contains much important information on coding schemes, classification, correlation, and searching. It is still valuable reading today.

Beginning in the early 1950s, electronic digital computers were turned to instead of punched-card equipment, and further advances in chemical structure handling began to be reported. This work generally pertains either to nomenclature, fragment codes, linear notations, or connection table representations of chemical structure. It is beyond the scope of this paper to do more than mention a few of the important early developments.

Among fragment codes, the Frear Code,[33a,b] the GREMAS code,[34a,b] and the SK&F code[35] are important instances. Many others are documented in earlier reviews (references 31a,b and other references at the beginning of this section). Fragment codes lend themselves to manipulation by mechanical or electronic means,[36] and they have been used for describing chemical reactions[37] and for indexing.[38]

The first well-defined linear notation system is that described by Dyson in 1946.[39] This notation subsequently was adopted by the International Union of Pure and Applied Chemistry,[40] and although it never became widely used, it initiated a surge of activity in chemical information that has endured to the present.

The most important linear notation from the point of view of adoption and use is that proposed by Wiswesser[41] in 1953. The rules of this notation are now in their third edition.[42] The notation is used by the Institute for Scientific Information for its *Chemical Substructure Index*,[43] and Chemical Abstracts Service has included Wiswesser linear notations in its *Parent Compound Handbook*.[44] Many more linear notation systems have been developed, but none has enjoyed the success of the Wiswesser linear notation.

In parallel with the development of linear notation systems, systems of structural representation called "Topological systems" have been developed. What is actually meant by this term is connection table or matrix representation of chemical structure. The earliest reference (1949) to the use of connection table representations of structure is that of Wheland

Table VII. Milestones in Computer-Based Indexing (References May Be Found in Landry and Rush)[77]

	text			nontext		
		contextual			contextual	
procedure	KWOC	KWIC	relational	KWOC	KWIC	relational
manual	classification schemes	latin squares	*Roget's Thesaurus* dictionaries			
concordances		concordances				
computer based	Luhn, 1957	Luhn, 1959	Bernier (thesaurus), 1957	Fugmann GREMAS, 1963	ISI Rotaform, 1963	Moores, 1951
	Baxendale, 1958	many variations	Doyle (semantic maps), 1962	CAS (screen generation), 1965	CAS HAIC, 1965	Meyer, 1962
	Maron, 1961		Gardin (SYNTOL), 1964		ISI Wiswesser, 1968	Gluck, 1963
	Edmundson and Wyllys, 1961		Hays, 1964			Morgan, 1965
	Salton, 1969		Quillian, 1968			
	Borko, 1970	Petrarca and Lay, 1969	Shank and Tessler, 1969			
			Armitage and Lynch, 1969			
			Rush, 1969			
			Fugmann, 1970			
			Baser et al., 1978[78]			

in his lectures in advanced organic chemistry at the University of Chicago.[45a,b] In 1951, Moores suggested that chemical structure could conveniently be represented as a tabulation of atoms and bonds,[46] apparently unaware of the earlier work of Wheland. It was not for another 6 years till the concept surfaced again, this time in a report by Ray and Kirsch.[47] In 1962, Meyer and Wenke[48] described their efforts to extend and refine the tabular representation suggested by Moores and worked out by Ray and Kirsch. The development of tabular representations of structure culminated in the work of Gluck and his colleagues at DuPont[49a,b] and of Morgan at Chemical Abstracts Service.[50] The importance of this work lies in the fact that a mathematically sound machine code for chemical structures was developed that could be produced by computer program from typed[51a–c] or hand-drawn chemical structures,[52a,b] as well as from conventional input, in an efficient and economical manner. Since 1965, extensions of the basic tabular representations described by Morgan have been developed, particularly in the areas of configuration and conformation[53a,b] and in the decription of incompletely specified structures.[54]

One of the important uses of computers in structure processing is in interconversion of structural representations. Among the types of interconversion studied are nomenclature to molecular formulas,[55a–c] notations to connection tables,[56] nomenclature to connection tables,[57a,b] connection tables to nomenclature,[58,59] connection tables to notations,[60,61] connection tables to graphical representations,[61–64] graphical representations to connection tables,[51,52] graphical representations from nomenclature,[65] graphical representations to notations,[60] and graphical representations from notations.[66] Without computers, this work could only have been done with extreme difficulty or not at all.

Yet another major use of computers in structure processing is in substructure searching. It is difficult to establish precedence in substructure searching, but perhaps the Beilstein classification[67] may be accorded this honor. In any case, by 1952–1953 many of the major chemical firms had adopted and were employing a variety of schemes for substructure searching.[31a,b] The work at Du Pont and CAS on connection table representations of structure formed the basis for considerable work on full and partial structure retrieval.[68a,b,69] Much of the work in this area concerned development and use of screens to speed the search process (see, for example, references[70–72]). A number of algorithms for identifying rings and systems of rings in chemical structures was also developed.[73–75] Work continues in this area today.

There is much more to the story of computers in handling chemical structure information, but I must leave it to the other authors in this issue to complete.

APPLICATIONS OF COMPUTER HARDWARE AND SOFTWARE TO PROCESSING OF CHEMICAL INFORMATION IN TEXTUAL FORM

To me, the problems of processing textual information so as to successfully locate wanted information are, in general, more difficult than those of processing of structure information, and less progress has been made than one would like. In this penultimate section, I will touch on some of the milestones in use of computer systems for processing textual data and will close with some predictions for the future.

Until the advent of electronic computers, various mechanical approaches to the organization, classification, codification, and indexing of textual material were developed and used. We are all familiar with indexes in the backs of books, with library card catalogs, and with the very important printed indexes such as those published by BIOSIS, CAS, and ISI. However, these sorts of indexes are produced largely by manual means, although computers are now widely used for sorting, formatting, error checking, and printing of the indexes.

Computer-based indexing began to be seriously studied in 1957 with the publication by Luhn[76a,b] of work on both keyword-out-of-context (KWOC) and keyword-in-context (KWIC) indexing. Luhn employed statistical techniques to determine which words of text to select for index terms. Table VII shows some of the landmark work in computer-based indexing. This table is drawn from an extensive review and analysis of automatic indexing by Landry and Rush.[77] Although this review is 10 years old, relatively little advancement in the state of the art of computer-based indexing has occurred during the decade just past. The most notable exception is the work of Fugmann on the graphical (or topological) representation of concept relations.[79a–c] Other workers have also developed techniques for representing textual material in the form of two-dimensional structures for indexing and retrieval.[80a,b] In order for these indexing techniques to be successful, either highly structured text or sound linguistic analysis procedures are required. Several workers have therefore devoted their attention to linguistic analysis. Good summaries of this topic are those of Hicks et al.[81] and Tennant.[82] The theory of cases, as propounded by Fillmore,[83] seems to provide the most reasonable basis for linguistic analysis. This theory considers words and phrases in a sentence to play certain well-defined roles in relation to other words or phrases in the sentence. The idea is similar is some respects to the notion of links and roles discussed by Taube[84] and others. The importance of case grammar analysis is that it allows a computer program to get at the meat of a sentence without having to know exactly what the sentence is "about". Moreover, sentences may then be categorized according to case structure, permitting additional refinements in the indexing process. Although no practical systems have been developed (the *Predicasts* scheme, though manual, may be an empirical exception), recent work by Zamora and Blower[85a,b] and by Nishida et al.[86] suggest a revitalization of interest in indexing based on case grammar analysis.

The earliest commercially produced computer-based index was the publication *Chemical Titles*.[87] It was based on the KWIC indexing method, which is still the most widely successful automatic indexing method. Although many other sorts of indexes are produced with the aid of computers, no other fully automatic methods have been reduced to routine production use.

The lack of automatic indexing systems stems, in part, from lack of a reasonable data base to which indexing procedures can be applied. Even with the power of modern computers, indexing the full text of articles, books, reports, and other publications would be prohibitive both in time and cost.

It was precisely for this reason that Rush and his students turned to automatic abstracting research. A sound abstract could serve, among other things, as a feasible basis for automatic indexing. Work toward the development of automatic abstracting systems has been reviewed by Mathis and Rush.[88] The historical development of abstracting has been reviewed by Skolnik,[89] and a good general treatment of abstracting is presented by Maizell et al.[90] The earliest work on computer generated abstracts was that of Luhn,[91] which involved statistical analysis of text. Other similar approaches to abstracting were developed in subsequent years. Several of the more important abstracting methods are discussed in Mathis and Rush,[88] in Davis and Rush,[1,24] and in the paper by Rush, Salvador and Zamora.[92] This paper introduced a new approach to abstracting based upon the notions of contextual inference and syntactic coherence. As in automatic indexing, knowledge of linguistic structure was deemed important by these authors for successful automatic abstracting.

Automatic abstracting, like indexing, suffered, at the time most of the research was done, from lack of an adequate base of machine-readable text. However, as more and more publications are generated via computer, a significant body of machine-readable text is available, and it would be worthwhile to reinitiate work in automatic abstracting and indexing.

Indexing suggests vocabulary control. Although there have been many vocabulary control devices developed over the years, few have been designed specifically for computer application to production of indexes. A notable exception is the thesaurus developed at CAS to control the index vocabulary of *Polymer Science and Technology* (POST). This thesaurus provided complete support for choice of index terms and for generation of cross-references of various kinds.[81,93] Horvath et al.[94] also describe preparation and use of a thesaurus in the B. F. Goodrich information retrieval system. The interested reader is referred to the various volumes of the *Annual Review of Information Science and Technology* for additional information. Amsler has provided a good recent summary of work related to machine-readable dictionaries.[95]

Retrieval of information from text data files, whether based upon indexes or upon free text, has been a topic of great interest to chemists for as long as there have been bodies of data that exceed our own personal memory capacities. With the advent of computers, techniques that had previously been devised for searching manual files (see, for example, reference 32), were adapted and extended for use with computers. There is a copious literature dealing with various aspects of information retrieval, the bulk of which is reviewed in *Annual Review of Information Science and Technology* (19 volumes, 1966 to date). Lancaster[96] and Meadow and Cochran[97] are also good sources of information about text retrieval systems.

MEDLARS (Medical Literature Analysis and Retrieval System, National Library of Medicine) is perhaps the first retrieval system put into commercial operation on a widespread basis.[98] Chemical Abstracts Service offered retrieval software to subscribers to its machine-readable *Chemical Titles* files beginning ca. 1962–1963. These were batch processing systems that were used both for on-demand retrieval and for selective dissemination of information (SDI).

Salton developed the SMART system[99,100] during the same period as an on-line, interactive system. Lockheed Missiles and Space Co. introduced DIALOG, also an on-line system, a year or two later,[101] and the National Aeronautics and Space Administration (NASA) adopted a version of DIALOG called RECON (remote console system) shortly thereafter.[102,103] Both of these systems are in use today, and DIALOG is now the world's largest on-line bibliographic retrieval system. System Development Corp. also entered the on-line retrieval system business in 1968 with ORBIT (on-line retrieval of bibliographic information time-shared).[104] Many other systems were also developed during the late 1960s and early 1970s, most on a smaller scale (BRS and BASIS are notable exceptions) or as experimental systems. The review by Neufeld and Cornog[105] provides a current view of information systems including their problems, successes, and prospects for the future.

There is a growing trend to provide customized, personal information retrieval services via microcomputers and appropriate telecommunication facilities. There are too many of these to mention them all here, but STAR (offered by Cuadra Associates), B-I-T-S (offered by BIOSIS), and BRS/Search (from BRS, Inc.) are representative examples of the trend. PULSAR is also a microcomputer-based retrieval system for personal use.[106] In addition, a number of software packages, called "gateways", have been developed to facilitate end-user searching of multiple on-line systems. Among these are ISI's Sci-Mate, In-Search from Menlo Corp., and KTALK from KTALK Communications. All of these and other similar software packages are described and characterized in reference 107.

Unfortunately, the systems currently available and in use for searching files of textual data have scarcely greater capabilities or better functionality than their earlier progenitors. The only really significant improvements have come about through advances in computer technology (faster, cheaper, and more capacious) and in software design (but not in functional capability). In other words, there have been almost no advances in the state of the art of textual information processing (at least on a practical scale) for 15 years. Only with a breakthrough in basic linguistic analysis will we begin to see corresponding improvements in content analysis and in information retrieval. Such a breakthrough is absolutely essential for searching of large text files (i.e., files of unstructured text).

One final area of computer-based textual information processing I wish to touch on is computer-based publishing. The National Library of Medicine was among the first to adapt computer-based photocomposition devices for production of printed products. *Index Medicus* was produced with the Photon Co. GRACE (graphic arts composition equipment) machine as early as 1964.[108] Today, very sophisticated photocomposition systems are used by BIOSIS, CAS, ISI, and many other organizations.[109] Photocomposition and other related technologies (e.g., laser printing) will continue to be important as long as printed products are produced in large quantities.

The fact that photocomposition implies machine-readable data files as precursors leads me to a final observation about computer-assisted publishing. When CAS introduced *Chemical Titles* in 1960, magnetic tapes containing the data used to produce the print product were soon made available on a subscription basis to subscribers of the printed product. This machine-readable version of CT constituted perhaps the first electronic publication, although it was not thought of in these terms. Nevertheless, organizations such as CAS, ISI, BIOSIS, Engineering Information, American Institute of Physics, American Petroleum Institute, and many more pioneered electronic publishing long before it became a popular topic, and these organizations have struggled with the difficult financial, managerial, and marketing problems that arise when one begins to offer essentially competing products to the same markets. While many problems remain to be solved, there can be no doubt that electronic publishing in its many forms is here to stay. Hills[110] and Lerner et al.[111] provide good recent reviews of electronic publishing and its affect on traditional publishing (see also reference 105).

To my mind, the single greatest advantage of electronic publishing is the capability it affords publisher and user alike to define as precisely as one could want exactly the material to be delivered to the user. In other words, the user will finally be afforded a facile means of purchasing just the data of interest (whether in textual or other form) without at the same time having to acquire a great deal of excess baggage. At the same time, the publisher should find substantially increased markets for reasonably packaged products.

There are many other topics I could discuss as part of my overall subject, but space does not permit. Let me close with the following observation. Throughout history the recording of man's endeavors and the endeavors themselves have been largely disparate enterprises. However, that separateness is disappearing. The increasing use of computers, together with software that is rapidly increasing in capabilities, even intelligence, will make possible the merging of endeavor and the record thereof. Voice input/output, graphics capabilities, and intelligent systems are now in their infancy. But if technological advances are as rapid during the next 40 years as they have been during the 40 years just past, chemists will one day be able to use computer systems, for example, to develop plausible syntheses of desired substances and then to carry out the syntheses, while the same computer system captures the experimental and observational data, records it, summarizes it, and prepares it for future use by many other individuals. Even this may be but a weak perception of the possibilities.

REFERENCES AND NOTES

(1) Davis, C. H.; Rush, J. E. "Guide to Information Science"; Greenwood Press: Westport, CT, 1979; p 176.
(2) Ibid., pp 153–158.
(3) Goldstine, H. H. "The Computer from Pascal to von Neumann"; Princeton University Press: Princeton, NJ, 1972; pp 10–26.
(4) Moseley, M. "Irascible Genius, a Life of Charles Babbage, Inventor"; Crowell: London, England, 1964.
(5) Goldstine, op. cit., pp 15–16, plate 6.
(6) Goldstine, op. cit., pp 93–96, plate 10.
(7) Davis and Rush, op. cit., p 153.
(8) "The World Book Encyclopedia"; World Book-Childcraft International: Chicago, IL, 1978; Vol. 10, p 192.
(9) "A Manual of Operations for the ASCC"; Harvard University: Camrbidge, MA, 1946.
(10) Ralston, A.; Reilly, E. D., Jr., "Encyclopedia of Computer Science and Engineering", 2nd ed.; Van Nostrand Reinhold: New York, 1983; pp 607–608.
(11) Goldstine, op. cit., p 153.
(12) Goldstine, op. cit., pp 188–191.
(13) Ralston and Reilly, op. cit., pp 946–950.
(14) Johnston, M. "High Tech, High Risk, and High Life in Silicon Valley". *Natl. Geograph.* **1982**, *162* (4), 459–477.
(15) Ralston and Reilly, op. cit., p 772.
(16) Reid, T. R. "The Chip". *Science 85* **1985**, *6* (1), 32–41.
(17) Ralston and Reilly, op. cit., p 973.
(18) Future Computing, November 1984.
(19) Lovelace, A. (Countess of) "Sketch of the Analytical Engine invented by Charles Babbage, Esq., by L. F. Menebrea, of Turin, Officer of the Military Engineers". In *Scientific Memoirs*"; Taylor: London, England, 1843; Vol. III, pp 666–731, with copious notes by translator.
(20) Hopper, G. M. "The Education of a Computer". *Proc. Assoc. Comput. Mach.* **1952**, 243–249.
(21) Sammet, J. E. "Programming Languages: History and Fundamentals"; Prentice-Hall: Englewood Cliffs, NJ, 1969.
(22) Tate, F. A. "Handling Chemical Structure Information". In *Annu. Rev. Inf. Sci. Technol.* **1967**, *2*, 285–309.
(23) Rush, J. E. "Handling Chemical Structure Information". *Annu. Rev. Inf. Sci. Technol.* **1978**, *13*, 209–262.
(24) Davis, C. H.; Rush, J. E. "Information Retrieval and Documentation in Chemistry"; Greenwood Press: Westport, CT, 1974; pp 113–269.
(25) Wipke, W. T.; Heller, S. R.; Feldmann, R. J.; Hyde, E. "Computer Representation and Manipulation of Chemical Information"; Wiley: New York, 1974.
(26) Rush, J. E. "Status of Notation and Topological Systems and Potential Future Trends". *J. Chem. Inf. Comput. Sci.* **1976**, *16* (4), 202–210.
(27) Stobaugh, R. E. "Chemical Structures: Computer Handling". In "Encyclopedia of Computer Science and Technology"; Belzer, J.; Holzman, A. G.; Kent, A., Eds.; Marcel Dekker: New York, 1976; Vol. 4, pp 478–509.
(28) Holm, B. E.; Howell, M. E.; Kennedy, H. E.; Kuney, J. H.; Rush, J. E. "The Status of Chemical Information". *J. Chem. Doc.* **1973**, *13* (4), 171–183.
(29) Ash, J. E.; Hyde, E. "Chemical Information Systems"; Horwood: Chichester, England, 1975.
(30) Lynch, M. F.; Harrison, J. M.; Town, W. G.; Ash, J. E.; Eds. "Computer Handling of Chemical Structure Information"; Macdonald: London, England, 1971.
(31) (a) National Academy of Sciences-National Research Council, Committee on Modern Methods of Handling Chemical Information "Survey of Chemical Notation Systems"; NAS-NRC: Washington, DC, 1964; NAS-NRC Publication 1150. (b) National Academy of Sciences-National Research Council, Committee on Modern Methods of Handling Chemical Information "Survey of European Non-Conventional Chemical Notation Systems"; NAS-NRC: Washington, DC, 1965; NAS-NRC Publication 1278.
(32) (a) Perry, J. W.; Casey, R. S. "Punched Cards: Their Application to Science and Industry", 1st ed.; Van Nostrand Reinhold: New York, 1951. (b) Casey, R. S.; Perry, J. W.; Berry, M. M.; Kent, A. "Punched Cards: Their Application to Science and Industry", 2nd ed.; Professional Aids: Chicago, IL, 1958.
(33) (a) Frear, D. E. H. "Punched Cards in Correlation Studies". *Chem. Eng. News* **1945**, *23*, 2077. Frear, D. E. H.; Seiferle, E. J.; King, H. L. "A New Classification System for Chemical Compounds". *Science (Washington, D.C.)* **1946**, *104*, 177–178.
(34) (a) Fugmann, R.; Braun, W.; Vaupel, W. "Zur Dokumentation chemischer forschungsergebnisse". *Angew. Chem.* **1961**, *73* (23), 745–751. (b) Rössler, S.; Kolb, A. "The GREMAS System, an Integral Part of the IDC System *73* Chemical Documentation". *J. Chem. Doc.* **1970**, *10* (2), 128–134.
(35) Craig, P. N.; Ebert, H. M. "Eleven Years of Structure Retrieval Using the SK&F Fragment Codes". *J. Chem. Doc.* **1969**, *9* (3), 141–146.
(36) Meyer, E. "Versatile Computer Techniques for Searching by Structural Formulas, Partial Structures, and Classes of Compounds". *Angew. Chem., Int. Ed. Eng.* **1970**, *9* (4), 585.
(37) Fugmann, R.; Bitterlich, W. "Reaktionen Dokumentation mit dem GREMAS System". *Chem. Z.* **1972**, *96* (6), 323–329.
(38) Skolnik, H. "A Chemical Fragment Notation Index". *J. Chem. Doc.* **1971**, *11* (3), 142–147.
(39) Dyson, G. M. "A New Notation of Organic Chemistry". "Royal Institute of Chemistry Lecture"; Chemical Society and the Society of Chemical Industry: London, 1946.
(40) International Union of Pure and Applied Chemistry, Commission on Codification, Ciphering and Punched Card Techniques "Rules for IUPAC Notation for Organic Compounds"; Wiley: New York, 1961.
(41) Wiswesser, W. J. "A Systematic Line Formula Chemical Notation". *ASLIB Proc.* **1953**, *5*, 136–147.
(42) Smith, E. G.; Baker, P. A. "The Wiswesser Line-Formula Chemical Notation (WLN)", 3rd ed.; Chemical Information Management, Inc.: Cherry Hill, NJ, 1976.
(43) Granito, C. E.; Rosenberg, M. D. "Chemical Substructure Index (CIS)—A New Research Tool". *J. Chem. Doc.* **1971**, *11* (4), 251–256.
(44) Ebe, T.; Zamora, A. "Wiswesser Line Notation Processing at Chemical Abstracts Service". *J. Chem. Doc.* **1976**, *16* (1), 33–35.
(45) (a) Wheland, G. W. "Syllabus for Advanced Organic Chemistry 321"; University of Chicago: Chicago, IL, 1946. (b) Wheland, G. W. "Advanced Organic Chemistry"; Wiley: New York, 1949; p 87.

(46) Moores, C. N. "Ciphering Structural Formulas—the Zatopleg System". *Zator Tech. Bull.* **1951**, *59*.
(47) Ray, L. C.; Kirsch, R. A. "Finding Chemical Records by Digital Computers". *Science (Washington, D.C.)* **1957**, *126* (3227), 814–819.
(48) Meyer, E.; Wenke, K. "Ein System zur topologischen Verschlüsselung organisch-chemischer Strukturformeln für die mechanisierte Dokumentation". *Nachr. Dok.* **1962**, *13* (1), 13–19.
(49) (a) "New System Identifies and Stores Formulas". *Chem. Eng. News* **1963**, *41* (49), 35–36. (b) Gluck, D. J. "A Chemical Structure Storage and Search System Developed at Du Pont". *J. Chem. Doc.* **1965**, *5* (1), 43–51.
(50) Morgan, H. L. "The Generation of a Unique Machine Description for Chemical Structures—A Technique Developed at Chemical Abstracts Service". *J. Chem. Doc.* **1965**, *5* (2), 107–113.
(51) (a) Feldman, A.; Holland, D. B.; Jacobus, D. P. "The Automatic Encoding of Chemical Structures". *J. Chem. Doc.* **1963**, *3* (4), 187–189. (b) Mullen, J. M. "Atom-by-Atom Typewriter Input for Computerized Storage and Retrieval of Chemical Structures". *J. Chem. Doc.* **1966**, *7* (2), 88–93. (c) Gottardi, R. "A Modified Dot-Bond Structural Formula Font with Improved Stereochemical Notation Abilities". *J. Chem. Doc.* **1970**, *10* (2), 75–81.
(52) (a) Meyer, E. "Ein Maschine zur Verschlüsselung chemischer Strukturformeln für die Dokumentation". *Nachr. Dok.* **1962**, *13* (3), 144–146. (b) Cossum, W. R.; Hardenbrook, M. E.; Wolfe, R. N. "Computer Generation of Atom-Bond Connection Tables from Hand-Drawn Chemical Structures". *Proc. Am. Doc. Inst.* **1964**, *1*, 269–276.
(53) (a) Petrarca, A. E.; Lynch, M. F.; Rush, J. E. "Methods for Computer Generation of Unique Structural Representations of Stereoisomers". *J. Chem. Doc.* **1967**, *7* (3), 154–165. (b) Petrarca, A. E.; Rush, J. E. "Methods for Computer Generation of Unique Configurational Descriptors for Stereoisomeric Square Planar and Octahedral Complexes". *J. Chem. Doc.* **1969**, *9* (1), 32–37.
(54) Deforeit, H.; Caric, A.; Combe, H.; Leveque, S.; Malka, A.; Valls, J. "CORA—A Semiautomatic Coding System Application to the Coding of Markush Formulas". *J. Chem. Doc.* **1972**, *12* (4), 230–233.
(55) (a) Garfield, E. "An Algorithm for Translating Chemical Names to Molecular Formulas"; Institute for Scientific Information: Philadelphia, PA, 1961. (b) Garfield, E. "Chemico-Linguistic: Computer Translation of Chemical Nomenclature". *Nature (London)* **1961**, *192* (4798), 192. (c) Garfield, E. "An Algorithm for Translating Chemical Names to Molecular Formulas". *J. Chem. Doc.* **1962**, *2* (4), 177–179.
(56) Hyde, E.; Matthews, F. W.; Thomson, L. D.; Wiswesser, W. J. "Conversion of Wiswesser Notation to a Connectivity Matrix for Organic Compounds". *J. Chem. Doc.* **1967**, *7* (4), 200–204.
(57) (a) Vander Stouw, G. G.; Naznitsky, I.; Rush, J. E. "Procedures for Converting Systematic Names of Organic Compounds into Atom-Bond Connection Tables". *J. Chem. Doc.* **1967**, *7* (3), 165–169. (b) Vander Stouw, G. G.; Elliott, P. M.; Isenberg, A. C. "Automated Conversion of Chemical Substance Names to Atom-Bond Connection Tables". *J. Chem. Doc.* **1974**, *14* (3), 185–193.
(58) Conrow, K. "Computer Generation of Baeyer System Names of Saturated, Bridged, Bicyclic, Tricyclic, and Tetracyclic Hydrocarbons". *J. Chem. Doc.* **1966**, *6* (4), 206–212.
(59) Van Binnendyk, D.; Mackay, A. C. "Computer-Assisted Generation of IUPAC Names of Polycyclic Bridged Ring Systems". *Can. J. Chem.* **1973**, *51*, 718–723.
(60) Farrell, C. D.; Chauvenet, A. R.; Koniver, D. A. "Computer Generation of Wiswesser Line Notation". *J. Chem. Doc.* **1971**, *11* (1), 52–59.
(61) Feldmann, R. J.; Heller, S. R.; Shapiro, K. P.; Heller, R. S. "An Application of Interactive Computing—A Chemical Information System". *J. Chem. Doc.* **1972**, *12* (1), 41–47.
(62) Hyde, E.; Thomson, L. D. "Structure Display". *J. Chem. Doc.* **1968**, *8* (3), 138–146.
(63) Blake, J. E.; Farmer, N. A.; Haines, R. C. "An Interactive Computer Graphics System for Processing Chemical Structure Diagrams". *J. Chem. Inf. Comput. Sci.* **1977**, *17* (4), 223–228.
(64) Crane, E. M.; Horowitz, P. "HECSAGON—System for Computer Storage and Retrieval of Chemical Structure"; Eastman-Kodak: Rochester, NY, 1963.
(65) Stillwell, R. N. "Computer Translation of Systematic Chemical Nomenclature to Structural Formulas—Steroids". *J. Chem. Doc.* **1973**, *13* (2), 107–109.
(66) Thomson, L. D.; Hyde, E.; Matthews, F. W. "Organic Search and Display Using a Connectivity Matrix Derived from Wiswesser Notation". *J. Chem. Doc.* **1967**, *7* (4), 204–209.
(67) "Beilsteins Handbuch der organischen Chemie"; Springer: Berlin, 1918.
(68) (a) Hoffman, W. S. "An Integrated Chemical Structure Storage and Search System Operating at Du Pont". *J. Chem. Doc.* **1968**, *8* (1), 3–13. (b) Schultz, J. L. "Handling Chemical Information in the Du Pont Central Report Index". *J. Chem. Doc.* **1974**, *14* (3), 171–179.
(69) (a) Krakiwsky, M. L.; Davenport, W. C.; White, R. W. "Searching for Subsets in Machine Records of Chemical Structures at Chemical Abstracts Service". Presented before the 1965 Congress of the International Federation for Documentation, Washington, DC, 1965. (b) Dubois, J. E. "French National Policy for Chemical Information and the DARC System as a Potential Tool of This Policy". *J. Chem. Doc.* **1973**, *13* (1), 8–13.
(70) Jacobus, D. P.; Davidson, D. E.; Feldman, A. P.; Schafer, J. A. "Experience with the Mechanized Chemical and Biological Information Retrieval System". *J. Chem. Doc.* **1970**, *10* (2), 135–140.
(71) Lefkovitz, D. "Substructure Search in the MCC System". *J. Chem. Doc.* **1968**, *8* (3), 166–173.
(72) Adamson, G. W.; Cowell, J.; Lynch, M. F.; McLure, A. H. W.; Town, W. G.; Yapp, A. M. "Strategic Considerations in the Design of a Screening System for Substructure Searches of Chemical Structure Files". *J. Chem. Doc.* **1973**, *13* (3), 153–157.
(73) Fugmann, R.; Dölling, U.; Nickelsen, H. "A Topological Approach to the Problem of Ring Structures". *Angew. Chem., Int. Ed. Engl.* **1967**, *6* (6), 723–733.
(74) Long, P. L.; Rush, J. E. "An Algorithm for the Identification and Characterization of Cyclic Graphs Contained in Connection Matrixes". Presented before the Division of Chemical Literature, 157th National Meeting of the American Chemical Society, Minneapolis, MN, April 13–18, 1969; CHLT 21.
(75) Corey, E. J.; Wipke, W. T. "Computer-Assisted Design of Complex Organic Syntheses". *Science (Washington, D.C.)* **1969**, *166* (3902), 178–192.
(76) (a) Luhn, H. P. "The automatic Creation of Literature Abstracts". *IBM J. Res. Dev.* **1958**, *2* (2), 159–165. (b) Luhn, H. P. "A Statistical Approach to Mechanized Encoding and Searching of Literary Information". *IBM J. Res. Dev.* **1957**, *1* (4), 309–317.
(77) Landry, B. C.; Rush, J. E. "Automatic Indexing: Progress and Prospects". "Encyclopedia of Computer Science and Technology"; Belzer, J.; Holzman, A. G.; Kent, A., Eds.; Marcel Dekker: New York, 1975; Vol. 2, pp 430–447.
(78) (a) Armitage, J.; Lynch, M. F. "Computer Generation of Articulated Subject Indexes". *Proc. Am. Soc. Inf. Sci.* **1969**, *6*, 253–257. (b) Baser, K. H.; Cohen, S. M.; Dayton, D. L.; Watkins, P. B. "On-Line Indexing Experiment at Chemical Abstracts Service: Algorithmic Generation of Articulated Index Entries from Natural Language Phrases". *J. Chem. Doc.* **1978**, *18* (1), 18–25.
(79) (a) Fugmann, R.; Nickelsen, H.; Nickelsen, I.; Winter, J. "TOSAR—A Topological Method for the Representation of Synthetic and Analytical Relations of Concepts". *Angew. Chem., Int. Ed. Engl.* **1970**, *9* (8), 589–595. (b) Fugmann, R.; Nickelsen, H.; Nickelsen, I.; Winter, J. "Representation of Concept Relations Using the TOSAR System of the IDC: Treatise III on Information Retrieval Theory". *J. Am. Soc. Inf. Sci.* **1974**, *25* (5), 287–306. (c) Fugmann, R. "Role of Theory in Chemical Information Systems". *J. Chem. Inf. Comput. Sci.* **1982**, *22* (3), 118–125.
(80) (a) Strong, S. M. "An Algorithm for Generating Structural Surrogates of English Text". M.S. Thesis, The Ohio State University, 1973. (b) Young, C. E. "Development of Language Analysis Procedures with Application to Automatic Indexing". Ph.D. Dissertation, The Ohio State University, 1973.
(81) Hicks, C. E.; Strong, S. M.; Rush, J. E. "Content Analysis". In "Encyclopedia of Computer Science and Technology"; Belzer, J.; Holzman, A. G.; Kent, A., Eds.; Marcel Dekker: New York, 1977; Vol. 6, pp 74–129.
(82) Tennant, H. "Natural Language Processing"; Petrocelli: New York, 1981.
(83) Fillmore, C. J. "The Case for Case". In "Universals in Linguistic Theory"; Bach, E.; Harms, R., Eds.; Hold, Rinehart & Winston: New York, 1968; pp 1–88.
(84) Taube, M. "Documentation, Information Retrieval and Other New Technology". *Libr. Q.* **1961**, *31* (1), 90–103.
(85) (a) Zamora, E. M.; Blower, P. E., Jr. "Extraction of Chemical Reaction Information from Primary Journal Text Using Computational Linguistics Techniques. 1. Lexical and Syntactic Phases". *J. Chem. Inf. Comput. Sci.* **1984**, *24* (3), 176–180. (b) Zamora, E. M.; Blower, Jr., P. E. "Extraction of Chemical Reaction Information from Primary Journal Text Using Computational Linguistics Techniques. 2. Semantic Phase". *J. Chem. Inf. Comput. Sci.* **1984**, *24* (3), 181–188.
(86) Nishida, F.; Takamatsu, S.; Fujita, Y. "Semiautomatic Indexing of Structured Information of Text". *J. Chem. Inf. Comput. Sci.* **1984**, *24* (1), 15–20.
(87) Dyson, G. M. "Current Research at Chemical Abstracts". *J. Chem. Doc.* **1961**, *1* (2), 26.
(88) Mathis, B. A.; Rush, J. E. "Abstracting ". In "Encyclopedia of Computer Science and Technology"; Belzer, J.; Holzman, A. G.; Kent, A. Eds.; Dekker: New York, 1975; Vol. 1, pp 102–142.
(89) Skolnik, H. "Historical Development of Abstracting". *J. Chem. Inf. Comput. Sci.* **1979**, *19* (4), 215–218.
(90) (a) Maizell, R. E.; Smith, J. F.; Singer, T. E. R. "Abstracting Scientific and Technical Literature"; Wiley-Interscience: New York, 1971. (b) Cremmins, E. T. "The Art of Abstracting"; ISI Press: Philadelphia, PA, 1982. (c) "American National Standard for Writing Abstracts, ANSI Z39.14-1979"; American National Standards Institute: New York, 1979.
(91) Luhn, H. P. "An Experiment in Auto-Abstracting". "Auto-Abstract of Area 5 Conference Papers", International Conference on Scientific Information, Washington, DC; IBM Research Center: Yorktown Heights, NY, 1958.
(92) Rush, J. E.; Salvador, R.; Zamora, A. "Indexing and Abstracting by Computer. II. Production of Indicative Abstracts from Full Text by Computer". *J. Am. Soc. Inf. Sci.* **1971**, *22* (4), 260–274.

(93) Rush, J. E. "Work at CAS on Search Guides and Thesauri". Presented at the Chemical Abstracts Service Open Forum, Miami Beach, FL, April 12, 1967.
(94) Horvath, P. J.; Chamis, A. Y.; Carroll, R. F.; Dlugos, J. "The B F Goodrich Information Retrieval System and Automatic Information Distribution Using Computer-Compiled Thesaurus and Dual Dictionary". *J. Chem. Doc.* **1967**, *7* (3), 124–130.
(95) Amsler, R. A. "Machine-Readable Dictionaries". *Annu. Rev. Inf. Sci. Technol.* **1984**, *19*, 161–209.
(96) Lancaster, F. W. "Information Retrieval Systems: Characteristics, Testing, and Evaluation"; Wiley: New York, 1968.
(97) Meadow, C. T.; Cochran (Atherton), P. "Basics of Online Searching"; Wiley: New York, 1981.
(98) Austin, C. J. "The MEDLARS System". *Datamation* **1964**, *10* (12), 28–31.
(99) Salton, G.; Lesk, M. E. "The SMART Automated Document Retrieval System—An Illutration". *Commun. Assoc. Comput. Mach.* **1965**, *8* (6), 391–398.
(100) Salton, G. Ed. "The SMART Retrieval System: Experiments in Automatic Document Processing"; Prentice-Hall: Englewood Cliffs, NJ, 1971.
(101) Summit, R. K. "DIALOG—An Operational On-Line Reference Retrieval System". In "Proceedings of the Association for Computing Machinery 22nd National Conference"; Thompson: Washington, DC, 1967; pp 51–56.
(102) National Aeronautics and Space Administration, Remote Information Retrieval Facility NASA Report CR-1318; NASA: Washington, DC, 1969.
(103) Interuniversity Communications Council, Inc. (EDUCOM) "Networks for Higher Education". "Proceedings of the Spring Conference, 1972"; EDUCOM: Princeton, NJ 1972; pp 133–142.
(104) Nance, J. W.; Lathrop, J. W. "System Design Specifications, General Purpose ORBIT"; System Development Corp.: Santa Monica, CA, Sept 15 1969; TM-DA-20/000/00.
(105) Neufeld, M. L.; Cornog, M. "Secondary Information Services and Systems". *Annu. Rev. Inf. Sci. Technol.* **1984**, *19*, 151–183.
(106) Smith, S. F.; Jorgensen, W. L.; Fuchs, P. L. "PULSAR: A Personalized Microcomputer-Based System for Keyword Search and Retrieval of Literature Information". *J. Chem. Inf. Comput. Sci.* **1981**, *21* (4), 209–213.
(107) James E. Rush Associates "Microcomputers for Libraries: Product Review and Procurement Guide"; James E. Rush Associates: Powell, OH, 1984; update service.
(108) Karel, L.; Austin, C. J.; Cummings, M. M. "Computerized Bibliographic Services for Medicine". *Science (Washington, D.C.)* **1965**, *148*, 766–772.
(109) Lerner, R. G. "SPIN". *Online* **1979**, *3* (4), 23–26.
(110) Hills, P. J. "The Scholarly Communication Process". *Annu. Rev. Inf. Sci. Technol.* **1983**, *18*, 99–125.
(111) Lerner, R. G.; Metaxas, T.; Scott, J. T.; Adams, P. D.; Judd, P. "Primary Publication Systems and Scientific Text Processing". *Annu. Rev. Inf. Sci. Technol.* **1983**, *18*, 127–149.

Publishing of Primary Information

JOSEPH H. KUNEY
Oakton, Virginia 22124

Received January 9, 1985

The American Chemical Society was a prime force in developing computer-aided composition for primary journals. The *Journal of Chemical Documentation* became the first primary journal to be "typeset" by computer-aided photocomposition on a regular production schedule. Out of that work emerged the single coding–multiple use concept that influences much of today's thinking on the application of computers to information processing and retrieval. The growing sophistication and lower costs of hardware, software, and telecommunications plus the proliferation of personal computers may produce significant changes in the traditional author–publisher–user relationship.

In 1958 the publishers of primary journals were facing problems of coping with the steadily increasing volume of manuscripts offered for publication. A problem made even more complex by the impact of inflation on costs of publishing. As concerned as they were at that time, little did publishers realize how serious the problems of inflation were to become in a short few years. Both problems were solved through a steady diet of price increases, particularly to institutional subscribers, and by launching new, more specialized journals carrying fewer pages. In those years it was apparent there was considerable elasticity in journal prices as increases were accepted by institutional subscribers with little loss in circulation. Publishers have continued to rely on these two devices, but there is now increasing evidence that the upper limits of price elasticity are being tested. Current price increases are being resisted by cuts in subscriptions and greater reliance on alternatives: secondary services, interlibrary loans, photocopying, etc.

About 1958 several technological developments created new thinking on how the problems of publishing and processing primary publications might be solved by means other than by increasing prices and by starting new journals. First, a new technology, represented by a new concept in typesetting named the Photon, made its commercial debut. The Photon was described by its makers as a "high speed, typecomposing machine which produces in positive form, on film, printed matter which can immediately and easily be used for graphic arts reproduction."

Second, publishers began to get curious about the possible use of the computer in solving their problems. A curiosity that turned into intense interest when Mike Barnett and his coworkers at MIT announced that they had successfully programmed a computer to control typesetting through the Photon.[1] Potential for significant cost reductions in the composition of journals through a merger of the computer and photocomposition were recognized immediately. About the same time a concept began to form of a system of processing and disseminating scientific literature where a single coding in machine-readable form of the content of a manuscript at the start of the primary processing stage would provide access at all further levels of information use—abstracts, subject indexes, and author indexes; in fact, any and all levels of information need from keywords to full text.

But it would be the potential for reducing costs that would get immediate attention. For most average journals (3000–10 000 circulation) in 1958, approximately half of the total cost for composition, printing, binding, paper, and distribution was for composition. Thus, a 10% saving in the cost of composition for the *Journal of Organic Chemistry*, as we pointed out in 1958, would mean that 110 additional pages could be published per year with no increase in total cost.[2]

Then, still largely true today, the printing industry was not distinguished by its support of the development of new technology. New printing technology was more likely to be the product of publisher initiative. Consistent with this pattern, the Board of Directors of the ACS, aided in no small part by the coincidence that the Photon production plant was located about 10 min from the campus of MIT, the office of the then chairman of the ACS Board, authorized the purchase of a Photon for "investigative purposes". Cost reduction was the

