Status of Notation and Topological Systems and Potential Future Trends†

JAMES E. RUSH
Ohio College Library Center, Columbus, Ohio 43212

Received October 29, 1976

The language of chemistry may be divided into two broad sublanguages: one which enables us to talk about substances and one which enables us to talk about the properties, characteristics, or behaviors of these substances. This paper is concerned principally with the sublanguage of substances, although this sublanguage is not more important than the other.

NAMING

In order to talk about substances, chemists have devised many alternative languages. The most common of these languages is called "nomenclature", meaning a system of naming. Although Dr. Loening has already discussed chemical nomenclature,[1] I should like to make one or two observations concerning this topic.

One can think of language in general as a system of naming things. Man names objects, relationships between objects, and situations or events[2] with a facility directly proportional to his command of a language. In chemistry, nomenclatures of various sorts are used to name chemical substances, and each sort of nomenclature is based upon some set of rules, either informal or formal. Nomenclatures based on informal rules are sometimes called "trivial" nomenclatures,[3] although they are by no means trivial since they are in the greatest common use. Naming based upon the natural source of a substance, or upon certain of the characteristics (attributes) of a substance (such as color or cystalline form) is quite common.

Cadaverine
Cubane
Crystal Violet

Naming based upon the process or processes through which substances are obtained is also common.

Pyrocatechol
Dehydroprogesterone

The disadvantage of informal naming schemes is that those of us who are unfamiliar with a substance's antecedents will not find the language informative. This is because languages in general are products of particular cultures, and if we do not happen to be attuned to that culture, the language will have no meaning for us.[4]

For this reason, languages based upon formal systems of rules are frequently devised. Formal naming helps to broaden and strengthen a culture. The discipline of chemistry, whose adherents form a culture of chemistry, has given rise to many formal systems of naming, of which that called "systematic" nomenclature[5] is but one. Wiswesser[6] has surveyed the history of chemical nomenclature (line-formula notations) and has contrasted developments in structural chemistry with those of structure representation. Verkade[7] has more recently published a series of papers in which he reviews extensively chemical nomenclature from a historical perspective.

The convenience which formal systems of naming afford is that one need only know the rules and the basic vocabulary in order to "speak" the language, instead of having to know the name of each substance as an item. This observation applies equally to any systematized language, including those known as "natural" languages (e.g., English).

There is, in a general sense, no preferred name for a substance. Just as we find it convenient (even desirable) to give one another many alternative names, so do we find it convenient to give substances many alternative names. However, a particular name may be preferred in specific circumstances. For instance, a person may be given a nickname and a formal name in addition to his legal name, as:

Jack Kennedy
John F. Kennedy
Mr. President Kennedy

Likewise, a chemical substance may have several names:

cadaverine
1,4-diaminopentane
1,4-pentanediamine

Within the last 30 years, a great deal of emphasis has been placed upon the study of and developments in the naming of chemical substances, as evidenced by the voluminous literature on the subject.[8] The work which this literature represents has tended away from traditional, pronouncable nomenclature toward more abstract (though not more formal) systems of naming.

LINEAR NOTATION SYSTEMS

As Skolnik has pointed out,[9] many people have addressed themselves to the devising of naming systems called "linear notation systems" whose purposes are (a) to represent the structure of chemical substances either in part or in whole; (b) to reduce the length of names relative to that usual in traditional nomenclature; (c) to facilitate manipulation by mechanical or electronic means.

Linear notation systems which only partially describe the structure of chemical substances are usually referred to as "fragment codes". Perhaps the most important of the fragment codes in use is the GREMAS code,[10] although others are in use[11] and many more have been described.[8] (Fragment codes derived from "topological" representations of structure are discussed later.) Fragment coding systems represent chemical structures in a generic fashion, usually through explicit identification of functional groups, ring structures, and other "important" structural characteristics. Fragment codes are useful for naming compounds in small collections (say, to 50 000 structures) and for retrieving from the collection those compounds which exhibit specific structural characteristics (provided, of course, that the fragment code names the characteristic). Fragment codes also lend themselves to manipulation by mechanical or electronic means.[12] They serve as the basis for classification of compounds. Fragment codes have also been used as the basis for describing chemical reactions[13] (a topic I will discuss later) and for indexing.[14]

The most important linear notation systems are those devised to generate unique and unambiguous representations of chemical structures. These linear notation systems have been developed as alternatives to traditional nomenclature, although both linear notations and nomenclature serve the same basic purpose.

Conservation of storage space is a major reason for employing linear notations as representations of chemical

† Presented at the Award Symposium on "Contributions of the Division of Chemical Literature (Information) to the Chemical Society" at the ACS Centennial Meeting, New York, N.Y., April 6, 1976.

Table I. Classes of Linear Notation System

Class	Unique	Ambiguous
1	Yes	No
2	Yes	Yes
3	No	No
4	No	Yes

structure. But linear notations have the more important function of facilitating the storage and retrieval of chemical structural data so that the literature concerning such data is made more readily accessible to the chemist.

Two important characteristics of linear notation systems are uniqueness and ambiguity. These are defined as follows:

Uniqueness—The necessary and sufficient condition for *uniqueness* of a notation is that one and only one notation be derived from a chemical structure upon application of the rules of the system.

Ambiguity—The necessary and sufficient condition for a notation to be *unambiguous* is that the notation be susceptible of one and only one interpretation upon application of the rules of the system.[15]

These characteristics allow for the division of linear notation systems into four classes, as illustrated in Table I. The choice of a notation system from among these classes of systems depends upon the purpose to which the notations are to be put. A linear notation system which yields notations that are unique and unambiguous allows for precise communication about specific substances, but it may render communication about classes of substances more difficult. In contrast, a linear notation system which yields nonunique, unambiguous notations has the merit of easier application (by virtue of reduced rigor in the rules for deriving notations). However, more work is usually necessary to communicate about classes of structures, and lack of uniqueness means that files of such notations will contain redundant entries. Nevertheless, in particular situations it is conceivable that any of the four classes of linear notation system might be justifiable. The choice depends upon a variety of factors, including cost, number of structures to be described, number of chemists to be served by the system, growth in the number of structures, the expected number of specific vs. class queries to be serviced, whether manual or computer-based retrieval is to be provided, desired cost/benefit ratio, and so on.

The first well-defined linear notation system was that described by Dyson in 1946.[16] As originally conceived, this notation system provided for the derivation of notations consisting of symbols selected from the upper case Roman alphabet, the ten Arabic numerals, and ten punctuation marks (in symbolism, bearing a remarkable similarity to that employed later in the Wiswesser notation system). The symbolism used with this notation system was later expanded to include the lower case Roman alphabet and an array of special symbols.

Dyson's notation system, which subsequently was adopted by the International Union of Pure and Applied Chemistry (IUPAC) and which has become known as the IUPAC notation system,[17] used rules that parallel to a considerable degree rules for systematic chemical nomenclature. The IUPAC notation represents the earliest attempt to develop a systematic notation system for organic compounds as a successor of, or adjunct to, systematic chemical nomenclature. Despite the fact that it is not widely used (see, however, ref 18), the IUPAC notation system initiated a surge of activity in chemical information which has endured for almost 30 years.

Wiswesser proposed a new linear notation system in an article published in 1953.[19] This notation system, which now bears Wiswesser's name, has become the most widely used linear notation system largely because of the efforts of Wiswesser and of the Chemical Notation Association (which might more aptly be named the Wiswesser Notation Association). The rules of the system have been refined and extended,[20,21] and the nature and use of the system have been widely publicized through tutorials and the like. The Institute for Scientific Information has adopted the Wiswesser notation system as the basis for its Chemical Substructure Index,[22] and the Chemical Abstracts Service has included Wiswesser linear notations in its "Parent Compound Handbook".[23]

Although the Wiswesser notation system has been highly touted as the answer to structural representation and retrieval, it is not clear that this notation system or any other can be made to serve all purposes adequately. Nevertheless, the Wiswesser notation system is and will continue to be an important tool in chemical information processing.

A more recent addition to the collection of linear notation systems is that described by Skolnik[24] which places greater emphasis upon hydrogen in molecular structures. Such emphasis stems from the increased importance of nuclear magnetic resonance (NMR) spectrometry as an analytical tool.[25]

The development of new notation systems has not ceased. As long as people conceive special applications or improved techniques, notation systems will be developed. Evidence of this is afforded in the area of polymers, wherein efforts to represent polymers in connection-table form have been reported,[26] and the indexing of polymers has been described.[27] Indexing of the patent literature has also given rise to renewed activity in linear notations and fragment codes.[28] Many other applications of linear notations have been reported, and many more will no doubt be reported.

TOPOLOGICAL SYSTEMS

In parallel with the development of linear notation systems, systems of structural representation popularly called "topological systems" have been developed. Each of the following notations represents the topology of the structure shown with equal precision and accuracy, provided one is conversant with the symbology and rules of the system (language) used to generate the notation.

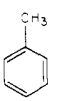
1B

B6C

6RMR5

Methylbenzene

-111235 6-7
CCCCCCC
-LL1LLL L



In the absence of such knowledge, the representations are merely nonmeaningful, not "atopological". What we actually mean when we refer to "topological" systems is connection-table or matrix representations of the topology of chemical substances as contrasted with their representation as linear sequences of symbols.

The earliest reference to the use of connection-table representations of structure is that of Wheland in his lectures in advanced organic chemistry at the University of Chicago, notes[29] for which subsequently formed the substance of a text published in 1949.[30] Wheland showed that acetaldehyde, for example, could be represented as well by a table of the form

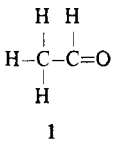
Table II

	C	C	H	H	H	H	O
C	-	1	1	1	1	0	0
C		-	0	0	0	1	2
H			-	0	0	0	0
H				-	0	0	0
H					-	0	0
H						-	0
O							-

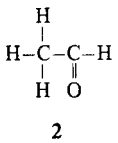
Table III

Column	Significance
1	Atom number (from original structure)
2	Atom type (numerical code for atom symbol)
3-6	Numbers of those atoms connected to the atom whose number appears in column 1 of a given row.

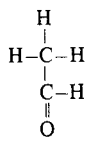
of Table II as by the diagrams **1–3** because the structural



1



2



3

diagram had no more geometrical significance than a name or a tabular representation. Wheland's interest was in the point that the above representations each adequately and uniquely described the topology of acetaldehyde.

It was not until 1951 that this tabular mode of representation appeared again in the literature, when Mooers[31] suggested that a chemical structure could conveniently be represented as a tabulation of atoms and bonds.

Remarkably, another six years passed before tabular representations again made their appearance in the literature. Ray and Kirsch[32] of the National Bureau of Standards described the use of Mooers' tabular representations as the basis for a chemical structure search system programmed for operation on the Bureau's digital computer. Much of the subsequent literature dealing with chemical structure handling techniques incorrectly marks the Ray and Kirsch paper as the point of origin of tabular representations of structure. Nevertheless, the paper by Ray and Kirsch gave impetus to the subsequent developments in tabular representation.

In 1962, Meyer and Wenke[33] described their efforts at refining and extending the tabular representation just discussed. They employed a six-column table, the columns of which had the significance shown in Table III. The tabular representations of structure developed by Meyer and Wenke are substantially similar to those developed later by Gluck (see below).

About the same time that the work of Meyer and Wenke was under way, Spialter[34] reported a "new unambiguous universal symbolism" for chemical compounds which he called the Atom Connectivity Matrix (ACM). The ACM is exactly analogous to Wheland's matrix, except that the atom symbols are placed along the principal diagonal, rather than at the heads of columns and rows. The off-diagonal elements of the ACM may contain bond symbols or other features of structure as desired. An interesting feature introduced by Spialter is the characteristic polynomial of the ACM (ACMCP). The ACMCP was supposed to be a unique representation of structure, although the point has been argued in several recent papers.[35,36] The disadvantage of the ACMCP is that it cannot be used to regenerate the structure from which it was derived, so that it serves only as a "registry number". Simpler ways of deriving (rather than assigning) "registry" numbers from structures have been developed.[37]

In 1963, Dyson, Cossum, Lynch, and Morgan described a method of forming tabular representations of structures from IUPAC notations.[38] Although this development was quickly supplanted by others, Dyson and his colleagues recognized the merit of their tabular representation of structure as an *interface language* through which communication between various systems could be effected. This idea has come to be exploited to a significant degree (as will be mentioned later).

A method of encoding chemical structures via typewriter input (punched paper tape) was described by Feldman, Holland, and Jacobus[39] in 1963. A typewriter with special character set was designed which recorded on paper tape the character struck and the position (coordinates) of the character on the page (i.e., coordinates relative to those of other characters forming the structures). These input data made it possible to produce tabular representations of structure quite easily. This input method was adopted by Gould, Gasser, and Rian[40] for the ChemSEARCH system at the Colgate-Palmolive Company.

The development of tabular representations of structure culminated in the reports of Gluck and his colleagues at DuPont[41,42] and of Morgan at Chemical Abstracts Service.[43] The details of this work may be found in the references cited. The importance of the work lies in the fact that a mathematically sound machine code for chemical structures was developed which could be produced by computer program from typed[39,44,45] or hand-drawn chemical structures[46,47] as well as from conventional input, in an efficient and economical manner. Too much emphasis was placed upon uniqueness perhaps, but for the purpose they were intended to serve, the representations had to be unique. But whether unique or not, what I shall call the Morgan representation can serve many purposes (other than the registration of chemical compounds), some of which will be mentioned later in this paper. Since 1965, extensions of the basic tabular representations described by Morgan have been developed, particularly in the areas of configuration and conformation[48,49] (where we are no longer content to treat topology but must treat geometry as well), and in the description of incompletely specified structures.[50]

One final point needs to be made concerning tabular representations of structure. The claim that a system generates unique and unambiguous structural representations is just that, a claim and nothing more, unless the claim can be substantiated through rigorous mathematical proof. The Morgan algorithm was proven to generate unique and unambiguous representations for all completely specified structures, although the proof was never published. This proof distinguishes the tabular representation of structure from any other system of representation for which the claim of uniqueness and unambiguity has been made. But for particular purposes, tabular representations may be no better or poorer than any other. I feel that the greatest merit of tabular representations lies in the interface language characteristic suggested by Dyson et al.,[38] as more recent work has begun to demonstrate.

INTERCONVERSION OF STRUCTURAL REPRESENTATIONS

The interconversion of structural representations has received growing attention during the past eight or ten years. Among the types of interconversion studied are:

- nomenclature to molecular formulas
- notations to connection tables
- nomenclature to connection tables
- connection tables to nomenclature
- connection tables to notions
- connection tables to graphical representations
- graphical representations to connection tables

- graphical representations to nomenclature

Such transformations may be reversible or not, and they are often empirically defined rather than prescribed by some formal system. I will discuss briefly some of this work. The interested reader would do well to consult the literature dealing with translation of natural languages as well.[51]

In 1961, Garfield described a method for translating chemical names to molecular formulas by computer program.[52-54] This work marked the first attempt to interconvert representations of chemical structures; it also introduced the notions of linguistics into chemical structure processing.

The work of Dyson et al.[38] also represents a significant effort in automatic translation, although this was not their main concern. Dyson[55] later made the point specifically.

Conversion from systematic nomenclature to connection tables was reported by Vander Stouw, Naznitsky, and Rush[56] in a 1967 paper. The system described by these authors was intended as an alternative to other means of input to the CAS Registry System, and as a method for editing chemical names.[57] Although the procedures described by Vander Stouw and his co-workers are effective and of broad scope, they are based upon empirically derived rules rather than on a formal system of rules such as that proposed by Elliott and Rush.[58]

The conversion of Wiswesser linear notations to connection tables has been described by Hyde and his co-workers.[59] The results of this work have been applied to the display of chemical structures via line printer[60] and to structure searching.[61]

A system designed to translate structural diagrams into Wiswisser linear notations and vice versa, has been described by Farrell, Chauvenet, and Koniver.[62] This system, which uses the connection-table representation of structure as the interface language, permits the user to obtain alternative forms of structural representation quite easily. Miller[63] has amplified certain aspects of the system, and important extensions of the system have been described by Feldmann and Koniver[64] and by Feldmann, Heller, Shapiro, and Heller.[65] I will treat other aspects of this system later.

Translation of connection-table representations of structure into systematic nomenclature was first described by Conrow in 1966.[66] Although his work was of narrow scope, Conrow pointed the way toward automatic nomenclature generation. A more recent effort in this direction is that of Van Binnendyk and Mackay[67] who developed a computer-assisted method of generating names of polycyclic bridged ring systems. No general attempt to generate nomenclature from connection tables has yet been described, although I am convinced not only of its possibility but of its feasibility.

Graphical representations of structure (structural diagrams) have been produced from nomenclature, linear notations, and connection tables. An interesting program for the translation of chemical nomenclature (of steroids) into structural diagrams has been described by Stillwell.[68] Structural diagram derivation from connection tables has been reported by Hyde and Thomson,[60] and by workers at the National Institutes of Health.[63,65] Work in this area has also been undertaken at Chemical Abstracts Service,[69] and many other workers have developed programs for display of structural diagrams.[39,45,70]

The interconversions of structural representations mentioned above all employ empirically based procedures. But now that interconvertibility has been successfully demonstrated, a more formal, general approach needs to be developed. There is some encouraging evidence of a move in this direction, such as the work of Elliott and Rush,[58] Rankin and Tauber,[71] Tauber and Rankin,[72] and Blower and Whitlock,[73] but a great deal of work is still required to define complete linguistic systems for chemical structure representation. Interestingly, this call for interconversion of representations of structure is but an echo of that issued by Opler in 1959.[74]

APPLICATIONS OF CHEMICAL STRUCTURAL REPRESENTATIONS

A great deal of work has been devoted to the development of representations of chemical structure. But in isolation these representations are of no interest to me, and I suspect they are of scant interest to most chemists. The work which has been done is fine, and there is room for yet more work in structure representation and interconversion. Yet I feel that we have been victims of one of Murphy's laws, namely, that "no matter what we want to do, we must do something else first". Happily, during the past four or five years much greater emphasis has been placed upon applications of structural representations than on the representations themselves.

I shall discuss applications in two broad subdivisions: the use of structure representations as means of access to the literature describing their properties and behavior, and the use of structure representations themselves to study chemical behavior and properties.

The justification for much of the early work on representations of structure was couched in terms of the inadequacy of nomenclature for indexing and retrieval. This argument is perhaps plausible when one considers computer applications, but it is not necessarily satisfying, and I have never accepted it at face value. Yet I cannot defend the proponents of nomenclature: the systems usually employed are developed and maintained by a cabal in which membership is difficult to attain.

I think we should eschew arguments of this sort, and focus our attention on the central issue: facile access to data related to specific substances or classes of related substances.

In the absence of direct data retrieval and manipulation capabilities, structure representations serve as means of access to documents in which such data may be found. The problem with traditional, nomenclature-based indexes is that retrieval based upon partial characterizations of structure (substructure search) is difficult to achieve. This problem has been attacked by developing alternative naming systems which make classification of chemical structures more facile, and considerable success has been achieved. Curiously, we have avoided calling substructure search "classification", although that is precisely what it is, perhaps because the classification is achieved dynamically at request time rather than statically at storage time (however, see Adamson and Bush[75]). At any rate, substructure search has not only been a concern, it is for many chemists a reality.

It is difficult to establish precedence in substructure searching, but perhaps the Beilstein classification[76] may be accorded this honor. Not until the mid-40's did any significant activity in substructure searching begin to be reported. Even then, much of the work bore on structure representations of various sorts rather than on substructure searching per se. Nevertheless, by 1952–53, many of the major chemical firms had adopted and were employing various forms of representation of structure to effect substructure searches.[8]

During the subsequent 20 years, many papers have been published describing a host of alternative schemes for substructure searching, many of which may be characterized as variations on a theme (e.g., duet for punched cards, sonata for computer and magnetic tape). I shall pay attention to just a few of the host of systems developed for effecting substructure searches.

Smith, Kline and French (SK&F) laboratories have operated a mechanized chemical information retrieval system since 1956.[11] The retrieval of chemical substances is effected through use of the SK&F fragment code, together with suitable Boolean logical operators to define desired combinations of fragments. The SK&F code numbers obtained in a search were then used to retrieve cards containing complete

structural diagrams which were edited before being sent to the laboratory chemist. More recently, chemical nomenclature has been used as an alternative or adjunct to structural diagrams, but with the development of a capability for printing structural diagrams on a line printer,[45] structural diagrams can be provided as search output without the manual effort formerly required to retrieve structure cards.

As has already been mentioned, the DuPont Company was instrumental in the development of a useful connection table representation of structure.[42] These representations form the basis for full and partial structure retrieval.[77,78] The DuPont system has the capability of searching via graph-matching techniques or through use of structural fragments generated (by computer program) at the time the connection table for a substance is entered into the file.

Although graph-matching techniques are quite precise, they are frequently very time consuming, and this fact has given rise to a great deal of work aimed at reducing search time while maintaining precision. Two basic approaches to search time reduction have been explored: (1) a priori classification of a file of structures; (2) computational techniques based on graph-theoretical considerations.

Classification of structures has been called many things, including screening, descriptor assignment, and substructure indexing, but what is involved is the partitioning of a collection of structures into a number of overlapping subsets, each subset corresponding to those structures that possess some structural feature in common.

Collection partitions may be derived in many ways. The use of fragment screens for this purpose has received considerable attention. Substructure searches conducted by Chemical Abstracts Service utilized structural fragments, derived from the structure by computer program, as screens to reduce the size of the collection of structures which must be examined in detail.[79]

Workers at the Walter Reed Army Institute of Research were among the earliest to develop a structure search system employing multiple levels of structural specificity.[80] Bit screens identifying the presence or absence of both single and complex structural features are used.

Lefkovitz has described a screening technique[81] which he subsequently developed into a Topological Screen System (TSS).[82] The screens are structural fragments derived from the Mechanical Chemical Code[83] by a relatively straightforward algorithm. The TSS appears to be an effective means of reducing search time.[82] This screening system has also been used as the basis of on-line searching of a large file of chemical structures represented as connection tables.[84]

Granito and his co-workers[85] developed a bit-screen approach to substructure searching in files of compounds represented in Wiswesser linear notations. Each notation symbol was marked as present or absent in a given linear notation by setting (or resetting) a bit in a specified location in a fixed-positional field which preceded the linear notation itself. The use of these bit screens led to substantially decreased search times relative to the times required for the same searches without use of the screens.

Lynch and his colleagues[86] have more recently addressed the problems of selecting an optional set of structural characteristics to serve as screens. The importance of the application, as well as collection size, bias, and growth rate in designing a screening system, is emphasized by these researchers. Computer programs were developed to generate screens using "bond-centered" fragments. Evaluation of this approach to screening indicates satisfactory performance.[87]

Finally, Fisanick et al.[88] have described techniques and experience with substructure searching using chemical nomenclature as the method of structure representation. Since the names, like linear notations, are comprised of linguistically significant strings of characters, text searching programs can be used to carry out substructure searches. Various types of screens were employed to reduce search time. Good search results can be obtained, although knowledge of nomenclature is required. I would also expect that considerable practice would be required to become proficient in formulating search strategies.

Many more studies related to screen development and application have been reported, but those described above are indicative of the general nature of such studies.

In contrast with the fragment-based classification approach to file partitioning for substructure search, computational approaches based upon the theory of graphs and other mathematical concepts attempt to utilize the structure representation itself as the basis of file partitioning. Although several computer programs have been developed to isolate specified structural characteristics, none have achieved the status of a full-scale substructure search system.

One approach, developed by Sussenguth[89] and called "set reduction", involves treating the search request (query) and the structures to be searched as sets of elements, each having some prescribed property, and then establishing correspondence between pairs of sets. Successive iterations partition sets into subsets with fewer elements, until node/node correspondence is established. A refined implementation of Sussenguth's technqiue has been reported by Ming and Tauber[90] which employs node degree (first- and second-order)[91], and which provides for searches involving the cooccurrence of disjoint substructures. Set reduction can be used to determine graph isomorphisms as well as for substructure search. An interesting alternative approach to set reduction has more recently been described by Figueras.[92]

Algorithms for identifying rings in chemical structures have been developed by several research groups. The early work in this area involved graph-theoretical studies on undirected graphs,[93,94] but in 1967, Fugmann, Dölling, and Nickelsen described a method of ring structure identification in chemical structure representations based upon the notion of the fundamental path.[95] Rings are formed of pair-wise combinations of fundamental paths whose endpoints coincide (excluding "envelope rings").

Long and Rush[96] described an algorithm which operated on connection tables of the form described by Morgan[43] to isolate rings and combinations of rings, which was more efficient than algorithms previously reported. Further extensions and refinements of this work have been reported.[97]

Corey and Wipke[98] became interested in ring identification techniques in connection with work on computer-aided organic synthesis. They described an algorithm based upon the random walk concept,[99] which is conceptually simple but generally quite inefficient. Wipke and Dyott[100] have more recently described a much more efficient algorithm and have presented data comparing two alternative ring-finding procedures.

Although graph-theoretical approaches to substructure identification have focused on cyclic structures, I am sure that useful procedures for handling acyclic and mixed structures will also be developed.

Structure searching has been developed to quite a high level of sophistication by a group at the National Institutes of Health.[101,102] Computer and graphics technologies have been combined with chemical structure processing algorithms to provide an interactive system which enables a user to display structural diagrams in three dimensions, manipulate the structures, plot them, and so on. The systems developed at NIH are a good example of the effective application of many of the structure-handling techniques developed during the past 15 years.

Another area of application of the many forms of structure representation is in indexing. While the screening techniques I mentioned earlier are in effect indexes, special indexes (usually printed) have been developed using fragment codes or linear notations as a basis.[14,28,103,104] Time does not permit their discussion.

Structure–activity correlation also relates to the topic of structure representation but comes more properly under the heading of inference techniques. I will touch on this matter later.

The work I have described was, until 1969, performed almost entirely by chemical information specialists or by nonchemists. Happily, during the past six years, a good many chemists have begun to evolve into a hybrid which is a cross between the information, computation, and chemistry professions. Perhaps these people do not think of themselves in this way, but that does not invalidate my thesis. What has happened, I think, is that chemists have become much more aware of the research potential of techniques developed or proposed by those in the information and computing fields. The result, if not the attitude, is evident in the work of Corey, Wipke, Lykos, Isenhour, Kowalski, Djerassi, Sasaki, and many more.

Computer-aided development of synthetic pathways is now receiving the attention that structure representation once received. Corey and Wipke[98,105] developed a computer-based system which permits the user to enter the structural diagram of a substance he wishes to synthesize and then work backward one reaction step at a time until structures representative of readily available starting materials are reached. Source and intermediate structures are stored in an on-line file in connection-table form. Corey and Wipke have been joined by many other chemists in continued research and development in computer-aided synthesis. Some of this work will be reported in the Symposium on Computer-Assisted Organic Synthesis.[106]

The storage and retrieval of known chemical reactions have also received attention, although we are not yet blessed with a generally useful technique for describing chemical reactions and providing for their retrieval.

Chemical reactions have been handled through use of the GREMAS system in Germany,[13,107] and Fugmann and his colleagues[108,109] have more recently developed a graphic approach to recording reaction information which permits storage and retrieval techniques similar to those employed for connection-table representations of structure to be used.

Osinga and Verrijn Stuart[110] have described a faceted classification scheme for use in documenting chemical reactions and have reported work on the use of Wiswesser linear notations for this purpose through the intermediary of an "atomic environment" representation.[111]

Lynch and his co-workers have studied the problem of documenting chemical reactions through use of both connection tables[112,113] and Wiswesser linear notations.[114,115] In the latter case, the approach taken differs from that described by Osinga and Verrijn-Stuart in that Lynch et al. depend entirely on the notation for description of reactions, rather than transforming the notations to a connection-table-like representation.

Other approaches to the handling of chemical reaction data have been reported. A good general treatment of the subject has been provided by Valls.[116]

Although a discussion of applications of structure representation may be continued almost indefinitely, I will end my remarks with a brief consideration of approaches to "content analysis".

In content analysis we are interested in the development of techniques which will enable us to draw inferences about the characteristics or behavior of chemical species or processes more quickly and more accurately than we are able to do unaided. For instance, can we find methods of analysis which will enable us to infer the mass spectrum of a compound from its structure? Can we infer the proper orientation of a ring system for nomenclature purposes by analysis of the connection-table representation of it? Given a collection of test data related to the efficacy of a certain group of drugs, can inferences be drawn regarding the relative behavior of the drugs, for both positive and negative effects?

There is a growing body of reports which suggests that such inferences can indeed be drawn, and it is this area to which our attention has and should continue to be directed. I must confess, however, that until recently many of us seem to have been side-tracked by issues which have proven of only secondary concern at best.

The computer-aided synthesis systems previously mentioned are essentially aimed at facilitating the inferential processes of chemists. The work of Kowalski,[117] Jurs,[118] and others in pattern recognition is aimed at enabling chemists to draw meaningful inferences from masses of analytical data which may, to the human eye, present no discernible patterns. The work in structure–activity correlation conducted by Hansch[119] is also illustrative of content analysis techniques (although not described in these terms), as is that of Heller et al. in the use of mass spectral data in structure determination.[120]

Although less mathematical in nature, the efforts of Maizell and his co-workers[121] in technological forecasting are no less inference-oriented. The use of the Delphi and similar techniques is well known in nontechnical circles.

I would be remiss if I did not conclude my remarks on content analysis techniques by mentioning DENDRAL. This system, which is still in development,[122] represents a substantial effort to adapt techniques of artificial intelligence to the problem of structure elucidation. Structure generation, based on empirical formulas, forms a central part of the system. This work has produced significant contributions to the wealth of structure handling techniques and to general problem solving. The reference cited is a good overview of DENDRAL.

CONCLUSION

This summary of work on the naming of chemical substances is necessarily only a specter of the actual work reported during the past 30 years and more. But it clearly indicates the theme of our efforts during that time and shows where I think we must go.

The process of naming will not cease. As long as anyone conceives the need for a language for some particular purpose, such languages will be developed. But the naming process must be relegated to a position of lesser importance so that the significant task of expanding our inference-making capabilities can be addressed. As I said earlier, the focus of research in chemistry is on the character and behavior of substances, not on the substances themselves. Thus, we must develop techniques which are predictive in nature rather than merely reactive. All those techniques which utilize structure to get to recorded data are reactive while those which utilize structure to derive data are predictive. A word of caution is, however, called for. Predictive techniques must not be allowed to supplant experimental techniques. The purpose of prediction is to focus and guide experimentation, and experimentation serves to prove predictions or to correct and improve the predictive technique. A theory without verification is a weak thing.

Another cautionary observation is also called for. It is heartening to note the increased interdisciplinary nature of much of the research currently being reported. It is important to take advantage of any and all available tools and techniques.

But, it is equally important to fully understand the tools and techniques to be used and to speak of them by their accepted names. There are, for instance, many parallels between the work which has been done on the naming of substances and the work done in the field of linguistics, but very little has been done to properly relate or equate this work. Similarly, notions of topology and other mathematical concepts have been mentioned in connection with the handling of structure, but it is not clear to me that these mathematical concepts are well understood by the practitioners in chemical information. Drawing upon concepts in artificial intelligence, pattern recognition, and so on must also be tempered by clear understanding of these areas of study. One must not lose sight, for instance, of the close relationship between pattern recognition and classification. And terms like "inference" and "perception" must be used with care and precision. There is an extensive body of literature related to both, of which we in chemical information have little awareness. Moreover, we often are slaves of computers rather than their masters.

Finally, I observe a lack of cohesion in chemical structure manipulation. I suppose that we are in no worse condition than was the field of chemistry itself in its infancy. But in reviewing the literature of chemical structure processing once again, in preparing this paper, I was struck by an obvious lack of historical perspective and by what I can only describe as a weak sense of direction. No one has so far bothered to devote the energy and intelligence to the task of "putting it all together". There are, to my knowledge, only four books on the subject of chemical information, that one edited by Lynch et al.,[123] the one edited by Wipke et al.,[124] the one by Davis and Rush,[5] and most recently that by Ash and Hyde.[125] While these books are useful, they do not meet the purpose I stated. What is needed, insofar as the field can provide it, is a "Fieser and Fieser" of chemical information. I hope that in addition to the flow of good research papers which I expect will continue to be published, a good, well-integrated book on the field will also be forthcoming.

REFERENCES AND NOTES

(1) K. L. Loening, "Status of Chemical Nomenclature and Future Trends", presented before the Division of Chemical Information, 171st National Meeting of the American Chemical Society, New York, N.Y., April 4–9, 1976.

(2) C. E. Young, "Development of Language Analysis Procedures with Application to Automatic Indexing", Computer and Information Science Research Center, OSU-CISRC-TR-73-2, The Ohio State University, Columbus, Ohio, 1973, pp 12–13.

(3) Chemical Abstracts Service, "The Naming and Indexing of Chemical Compounds", American Chemical Society, Washington, D. C., 1962, p 11N.

(4) J. E. Rush, "Communication, Language and Culture", paper prepared for presentation at the meeting of the International Association for Cross-Cultural Psychology, Kingston, Ontario, Aug 6–10, 1974.

(5) C. H. Davis and J. E. Rush, "Information Storage and Retrieval in Chemistry", Greenwood, Westport, Conn., 1974, Chapter 7.

(6) W. J. Wiswesser, "107 Years of Line-Formula Notations (1861–1968)", *J. Chem. Doc.*, **8**, 146–150 (1968).

(7) P. E. Verkade, "Historical Study on Nomenclature in Organic Chemistry. XI. Combes Procedure (1892) for the Nomenclature of Acyclic Compounds Containing Complex Functions", *Bull. Soc. Chim. Fr.*, 1119–26 (1975) (5–6, Part 2). See also earlier papers in this series.

(8) See, for example, Anon., "Survey of Chemical Notation Systems", prepared by the Committee on Chemical Information of the Division of Chemistry and Chemical Technology, National Research Council, and published as National Academy of Sciences Publication No. 1150, Washington, D.C., 1964; Anon., "Chemical Structure Information Handling: A Review of the Literature 1962–1968", prepared by the Committee on Chemical Information of the Division of Chemistry and Chemical Technology, National Research Council, and published as National Academy of Sciences Publication No. 1733, Washington, D.C., 1969; B. E. Holm, M. G. Howell, H. E. Kennedy, J. H. Kuney, and J. E. Rush, "The Status of Chemical Information", *J. Chem. Doc.*, **13**, 171–183 (1973).

(9) H. Skolnik, "Milestones in Chemical Information Science", presented before the Division of Chemical Information, 171st National Meeting of the American Chemical Society, New York, April 4–9, 1976.

(10) R. Fugman, W. Braun, and W. Vaupel, "Zur Dokumentation chemischer forschungsergebnisse", *Angew. Chem.*, **73**(23), 745–751 (1961); S. Rossler and A. Kolb, "The GREMAS System, an Integral Part of the IDC System for Chemical Documentation", *J. Chem. Doc.*, **10**, 128–134 (1970).

(11) P. N. Craig and H. M. Ebert, "Eleven Years of Structure Retrieval Using the SK&F Fragment Codes", *J. Chem. Doc.*, **9**, 141–146 (1969).

(12) E. Meyer, "Versatile Computer Techniques for Searching by Structural Formulas, Partial Structures, and Classes of Compounds", *Angew. Chem., Int. Ed. Engl.*, **9** (4), 585 (1970).

(13) R. Fugmann and W. Bitterlich, "Reaktionen Dokumentation mit dem GREMAS System", *Chem.-Ztg.*, **96** (6), 323–329 (1972).

(14) H. Skolnik, "A Chemical Fragment Notation Index", *J. Chem. Doc.*, **11**, 142–147 (1971).

(15) C. H. Davis and J. E. Rush, ref 5, p 146.

(16) G. M. Dyson, "A New Notation of Organic Chemistry", Royal Institute of Chemistry Lecture, published jointly with the Chemical Society and the Society of Chemical Industry, 1946.

(17) Anon., "Rules for IUPAC Notation for Organic Compounds", issued by the Commission on Codification, Ciphering and Punched Card Techniques of IUPAC, Wiley, New York, N.Y., 1961.

(18) H. F. Dammers and D. J. Polton, "Use of the IUPAC Notation in Computer Processing of Information on Chemical Structures", *J. Chem. Doc.*, **8**, 150–160 (1968).

(19) W. J. Wiswesser, "A Systematic Line Formula Chemical Notation", *ASLIB Proc.*, **5**, 136–147 (1953); see also, W. J. Wiswesser, "A Line-Formula Chemical Notation", Thomas Y. Crowell Co., New York, N.Y., 1954.

(20) E. G. Smith, "The Wiswesser Line-Formula Chemical Notation", McGraw-Hill, New York, N.Y., 1968.

(21) E. G. Smith and P. A. Baker, "The Wiswesser Line-Formula Chemical Notation (WLN)", 3rd ed, Chemical Information Management, Inc., Cherry Hill, N.J., 1976.

(22) C. E. Granito and M. D. Rosenberg, "Chemical Substructure Index (CSI)— A New Research Tool", *J. Chem. Doc.*, **11**, 251–256 (1971).

(23) T. Ebe and A. Zamora, "Wiswesser Line Notation Processing at Chemical Abstracts Service", *J. Chem. Doc.*, **16**, 33–35 (1976).

(24) H. Skolnik, "A New Linear Notation System Based on Combinations of Carbon and Hydrogen", *J. Heterocycl. Chem.*, **6**, 689–695 (1969).

(25) H. Skolnik, "A Correlative Notation System for NMR Data", *J. Chem. Doc.*, **10**, 216–220 (1970).

(26) R. E. Stobaugh, W. H. Powell, and R. J. Zalac, "Systems for Registering and Naming Polymers at Chemical Abstracts Service", Abstracts of Papers, 154th National Meeting of the American Chemical Society, Chicago, Ill., Sept 1967, No. G-016.

(27) K. L. Loening, W. Metanomski, and W. H. Powell, "Indexing of Polymers in *Chemical Abstracts*", *J. Chem. Doc.*, **9**, 248–251 (1969).

(28) M. Z. Balent and J. M. Emberga "A Unique Chemical Fragmentation System for Indexing Patent Literature", *J. Chem. Doc.*, **15**, 65–104 (1975).

(29) G. W. Wehland, "Syllabus for Advanced Organic Chemistry 321", The University of Chicago, Chicago, Ill., 1946 (Chapters 1–10 of ref 30).

(30) G. W. Wheland, "Advanced Organic Chemistry", Wiley, New York, N.Y., 1959, p 87.

(31) C. N. Mooers, "Ciphering Structural Formulas—The Zatopleg System", *Zator Tech. Bull.*, 59 (1951).

(32) L. C. Ray and R. A. Kirsch, "Finding Chemical Records by Digital Computers", *Science*, 126 (3227), 814–819 (1957).

(33) E. Meyer and K. Wenke, "Ein System zur topologischen Verschlüsselung organisch-chemischer Strukturformeln für die mechanisierte Dokumentation", *Nachr. Dok.*, **13** (1), 13–19 (1962).

(34) L. Spialter, "The Atom Connectivity Matrix (ACM) and Its Characteristic Polynomial (ACMCP). A New Computer-Oriented Chemical Nomenclature", *J. Am. Chem. Soc.*, **85**, 2012–2013 (1973).

(35) A. T. Balaban and F. Harary, "The Characteristic Polynomial Does Not Uniquely Determine the Topology of a Molecule", *J. Chem. Doc.*, **11**, 258–259 (1971).

(36) W. C. Herndon, "The Characteristic Polynomial Does Not Uniquely Determine Molecular Topology", *J. Chem. Doc.*, **14**, 150–151 (1974).

(37) M. J. Ebersole, "A General Method for Automatic Generation of Data Record Descriptors", M.S. Thesis, The Ohio State University, Columbus, Ohio, 1971.

(38) G. M. Dyson, W. E. Cossum, M. F. Lynch, and H. L. Morgan, "Mechanical Manipulation of Chemical Structure: Molform Computation and Substructure Searching of Organic Structures by the Use of Cipher-Directed, Extended and Random Matrices", *Inf. Storage Retr.*, **1** (2–3), 69–99 (1963).

(39) A. Feldman, D. B. Holland, and D. P. Jacobus, "The Automatic Encoding of Chemical Structures", *J. Chem. Doc.*, **3**, 187–189 (1963).

(40) D. Gould, E. B. Gasser, and J. F. Rian, "ChemSEARCH-An Operating Computer System for Retrieving Chemicals Selected for Equal, Analogous, or Related Character", *J. Chem. Doc.*, **5**, 24–32 (1965).

(41) Anon., "New System Identifies and Stores Formulas", *Chem. Eng. News*, 41 (49), 35–36 (1963).

(42) D. J. Gluck, "A Chemical Structure Storage and Search System Developed at DuPont", *J. Chem. Doc.*, **5**, 43–51 (1965).

(43) H. L. Morgan, "The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service", *J. Chem. Doc.*, **5**, 107–113 (1965).

(44) J. M. Mullen, "Atom-by-Atom Typewriter Input for Computerized Storage and Retrieval of Chemical Structures", *J. Chem. Doc.*, **7**, 88–93 (1966).
(45) R. Gottardi, "A Modified Dot–Bond Structural Formula Font with Improved Stereochemical Notation Abilities", *J. Chem. Doc.*, **10**, 75–81 (1970).
(46) E. Meyer, "Eine Maschine zur Verschlüsselung chemischer Strukturformeln für die Dokumentation", *Nachr. Dok.*, **13** (3), 144–146 (1962).
(47) W. E. Cossum, M. E. Hardenbrook, and R. N. Wolfe, "Computer Generation of Atom–Bond Connection Tables from Hand-Drawn Chemical Structures", *Proc. Am. Doc. Inst.*, **1**, 269–276 (1964).
(48) A. E. Petrarca, M. F. Lynch, and J. E. Rush, "A Method for Generating Unique Computer Structural Representations of Stereoisomers", *J. Chem. Doc.*, **7**, 154–165 (1967).
(49) A. E. Petrarca and J. E. Rush, "Methods for Computer Generation of Unique Configurational Descriptors for Steroisomeric Square Planar and Octahedral Complexes", *J. Chem. Doc.*, **9**, 32–37 (1969).
(50) H. Deforeit, A. Caric, H. Combe, S. Leveque, A. Malka, and J. Valls, "CORA—A Semiautomatic Coding System Application to the Coding of Markush Formulas", *J. Chem. Doc.*, **12**, 230–233 (1972).
(51) D. E. Walker, "Automated Language Processing", *Annu. Rev. Inf. Sci. Technol.*, **8**, 69–119 (1973).
(52) E. Garfield, "An Algorithm for Translating Chemical Names to Molecular Formulas", Institute for Scientific Information, Philadelphia, Pa., 1961, 68 pp.
(53) E. Garfield, "Chemico-linguistics: Computer Translation of Chemical Nomenclature", *Nature (London)*, **192** (4798), 192 (1961).
(54) E. Garfield, "An Algorithm for Translating Chemical Names to Molecular Formulas", *J. Chem. Doc.*, **2**, 177–179 (1962).
(55) G. M. Dyson, "A Cluster of Algorithms Relating the Nomenclature of Organic Compounds to Their Structure Matrices and Ciphers", *Inf. Storage Retr.* **2** (3), 159–199 (1964).
(56) G. G. Vander Stouw, I. Naznitsky and J. E. Rush, "Procedures for Converting Systematic Names of Organic Compounds into Atom-Bond Connection Tables", *J. Chem. Doc.*, **7**, 165–169 (1967).
(57) G. G. Vander Stouw, P. M. Elliott, and A. C. Isenberg, "Automated Conversion of Chemical Substance Names to Atom-Bond Connection Tables", *J. Chem. Doc.*, **14**, 185–193 (1974).
(58) P. M. Elliott and J. E. Rush, "Translation of Chemical Nomenclature by Syntax-Controlled Techniques", presented before the Chemical Documentation Program, 6th Middle Atlantic Regional Meeting of the American Chemical Society, Baltimore, Md., Feb 4–5, 1971.
(59) E. Hyde, F. W. Matthews, L. D. Thomson, and W. J. Wiswesser, "Conversion of Wiswesser Notation to a Connectivity Matrix for Organic Compounds", *J. Chem. Doc.*, **7**, 200–204 (1967).
(60) E. Hyde and L. D. Thomson, "Structure Display", *J. Chem. Doc.*, **8**, 138–146 (1968).
(61) L. D. Thomson, E. Hyde, and F. W. Matthews, "Organic Search and Display Using a Connectivity Matrix Derived from Wiswesser Notation", *J. Chem. Doc.*, **7**, 204–209 (1967).
(62) C. D. Farrell, A. R. Chauvenet, and D. A. Koniver, "Computer Generation of Wiswesser Line Notation", *J. Chem. Doc.*, **11**, 52–59 (1971).
(63) G. A. Miller, "Encoding and Decoding WLN", *J. Chem. Doc.*, **12**, 60–67 (1972).
(64) R. J. Feldmann and D. A. Koniver, "Interactive Searching of Chemical Files and Structural Diagram Generation from Wiswesser Line Notation", *J. Chem. Doc.*, **11**, 154–159 (1971).
(65) R. J. Feldmann, S. R. Heller, K. P. Shapiro, and R. S. Heller, "An Application of Interactive Computing-A Chemical Information System", *J. Chem. Doc.*, **12**, 41–47 (1972).
(66) K. Conrow, "Computer Generation of Baeyer System Names of Saturated, Bridged, Bicyclic, Tricyclic, and Tetracyclic Hydrocarbons", *J. Chem. Doc.*, **6**, 206–212 (1966).
(67) D. Van Binnendyk and A. C. Mackay, "Computer-Assisted Generation of IUPAC Names of Polycyclic Bridged Ring Systems", *Can. J. Chem.*, **51**, 718–723 (1973).
(68) R. N. Stillwell, "Computer Translation of Systematic Chemical Nomenclature to Structural Formulas–Steroids", *J. Chem. Doc.*, **13**, 107–109 (1973).
(69) N. A. Farmer and J. C. Schehr, "A Computer-Based System for Input, Storage and Photocomposition of Graphical Data", *Proc. 1974 Natl. Conf. Assoc. Comput. Mach.*, 563, (1974).
(70) E. M. Crane and P. Horowitz, "HECSAGON-System for Computer Storage and Retrieval of Chemical Structure", Eastman Kodak Co., Rochester, N.Y., 1963.
(71) K. Rankin and S. J. Tauber, "Linguistics as a Basis for Analyzing Chemical Structure Diagrams", *J. Chem. Doc.*, **11**, 139–141 (1971).
(72) S. J. Tauber and K. Rankin, "Valid Structure Diagrams and Chemical Gibberish", *J. Chem. Doc.*, **12**, 30–34 (1972).
(73) P. E. Blower, Jr., and H. W. Whitlock, Jr., "The Linguistic Approach to Organic Synthesis", presented before the Division of Chemical Information, 171st National Meeting of the American Chemical Society, New York, N.Y., April 4–9, 1976.
(74) A. Opler, "On the Automatic Manipulation of Representations of Chemical Structures", *Am. Doc.*, **10** (2), 130–134 (1959).
(75) G. W. Adamson and J. A. Bush, "A Comparison of the Performance of Some Similarity and Dissimilarity Measures in the Automatic Classification of Chemical Structures", *J. Chem. Doc.*, **15**, 55–58 (1975).
(76) "Beilsteins Handbuch der organischen Chemie", Julius Springer, Berlin, 1918.
(77) W. S. Hoffman, "An Integrated Chemical Structure Storage and Search System Operating at DuPont", *J. Chem. Doc.*, **8**, 3–13 (1963).
(78) J. L. Schultz, "Handling Chemical Information in the DuPont Central Report Index", *J. Chem. Doc.*, **14**, 171–179 (1974).
(79) M. L. Krakiwsky, W. C. Davenport, and R. W. White, "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, D.C., 1965.
(80) D. P. Jacobus, D. E. Davidson, A. P. Feldman, and J. A. Schafer, "Experience with the Mechanized Chemical and Biological Information Retrieval System", *J. Chem. Doc.*, **10**, 135–140 (1970).
(81) D. Lefkovitz, "Substructure Search in the MCC System", *J. Chem. Doc.*, **8**, 166–173 (1968).
(82) D. Lefkovitz and A. R. Gennard, "A Utility Analysis for the MCC Topological Screen System", *J. Chem. Doc.*, **10**, 86–94 (1970).
(83) D. Lefkovitz, "A Chemical Notation and Code for Computer Manipulation", *J. Chem. Doc.*, **7**, 186–192 (1967).
(84) M. Miline, D. Lefkovitz, H. Hill, and R. Powers, "Search of CA Registry (1.25 million compounds) with the Topological Screens System", *J. Chem. Doc.*, **12**, 183–189 (1972).
(85) C. E. Granito, G. T. Becker, S. Roberts, W. J. Wiswesser, and K. J. Windlinx, "Computer-Generated Substructure Codes (Bit Screens)", *J. Chem. Doc.*, **11**, 106–110 (1971).
(86) G. W. Adamson, J. Cowell, M. F. Lynch, A. H. W. McLure, W. G. Town, and A. M. Yapp, "Strategic Considerations in the Design of a Screening System for Substructure Searches of Chemical Structure Files", *J. Chem. Doc.*, **13**, 153–157 (1973).
(87) G. W. Adamson, J. A. Bush, A. H. W. McLure, and M. F. Lynch, "An Evaluation of a Substructure Search Screen System Based on Bond-Centered Fragments", *J. Chem. Doc.*, **14**, 44–48 (1974).
(88) W. Fisanick, L. D. Mitchell, J. A. Scott, and G. G. Vander Stouw, "Substructure Searching of Computer-Readable Chemical Abstracts Service Ninth Collective Index Chemical Nomenclature Files", *J. Chem. Info. Comput. Sci.*, **15**, 73–84 (1975).
(89) E. H. Sussenguth, "A Graph-Matching Algorithm for Matching Chemical Structures", *J. Chem. Doc.*, **5**, 36–43 (1965).
(90) T. K. Ming and S. J. Tauber, "Chemical Structure and Substructure Search by Set Reduction", *J. Chem. Doc.*, **11**, 47–51 (1971).
(91) R. H. Penny, "A Connectivity Code for Use in Describing Chemical Structures", *J. Chem. Doc.*, **5**, 113–117 (1976).
(92) J. Figueras, "Substructure Search by Set Reduction", *J. Chem. Doc.*, **12**, 237–244 (1972).
(93) J. T. Welch, Jr., "A Mechanical Analysis of the Cyclic Structure of Undirected Linear Graphs", *J. Assoc. Comput. Mach.*, **18** (2), 205–210 (1966).
(94) C. C. Gottlieb and D. S. Corneil, "Algorithms for Finding a Fundamental Set of Cycles for an Undirected Linear Graph", *Commun. Assoc. Comput. Mach.*, **10** (2), 780 (1967).
(95) R. Fugmann, U. Dölling, and H. Nickelsen, "A Topological Approach to the Problem of Ring Structures", *Angew. Chem., Int. Ed. Engl.*, **6**, 723–733 (1967).
(96) P. L. Long and J. E. Rush, "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, Minn., April 13–18, 1969, No. CHLT-21.
(97) P. L. Long, R. F. Phares, J. E. Rush, and L. J. White, "Fast Access to Rings in Chemical Structures", presented before the Division of Chemical Literature, 160th National Meeting of the American Chemical Society, Chicago, Ill., Sept 14–17, 1970, No. CHLT-15.
(98) E. J. Corey and W. T. Wipke, "Computer-Assisted Design of Complex Organic Syntheses", *Science*, 166 (3902), 178–192 (1969).
(99) J. C. Tiernan, "An Efficient Algorithm for Identifying Cycles in Graphs", *Commun. Assoc. Comput. Mach.*, **13** (2), 722–726 (1970).
(100) W. T. Wipke and T. M. Dyott, "Use of Ring Assemblies in a Ring Perception Algorithm", *J. Chem. Info. Comput. Sci.*, **15** (3), 140–147 (1975).
(101) R. J. Feldmann and S. R. Heller, "An Application of Interactive Graphics—The Nested Retrieval of Chemical Structures", *J. Chem. Doc.*, **12**, 48–54 (1972).
(102) R. J. Feldmann, S. R. Heller, and C. R. T. Bacon, "An Interactive, Versatile, Three-Dimensional Display, Manipulation and Plotting System for Biomedical Research", *J. Chem. Doc.*, **12**, 234–236 (1972).
(103) J. E. Dubois, "French National Policy for Chemical Information and the DARC System as a Potential Tool of This Policy", *J. Chem. Doc.*, **13**, 8–13 (1973).
(104) C. E. Granito, J. E. Schultz, G. W. Gibson, A. Gelberg, R. J. Williams, and E. A. Metcalf, "Rapid Structure Searches via Permuted Chemical Line Notations (III): A Computer-Produced Index", *J. Chem. Doc.*, **5**, 229–232 (1965).
(105) E. J. Corey, W. T. Wipke, R. D. Cramer, III, and W. J. Howe, "Computer-Assisted Synthetic Analysis. Facile Man-Machine Communication of Chemical Structure by Interactive Computer Graphics", *J. Am. Chem. Soc.*, **94**, 421–430 (1972); see also the two papers which follow in the same issue.

(106) Symposium on Computer-Assisted Organic Synthesis, Division of Chemical Information and Division of Computers in Chemistry (joint sponsors), 171st National Meeting of the American Chemical Society, New York, N.Y., April 4–9, 1976.
(107) R. Fugmann, W. Braun, and W. Vaupel, "GREMAS-ein Weg zur Klassification und Dokumentation in der organische Chemie", *Nach. Dok.*, **14** (4), 179–190 (1963).
(108) R. Fugmann, H. Nickelsen, I. Nickelsen, and J. H. Winter, "A Topological Method for the Representation of Synthetic and Analytical Relations of Concepts", *Angew. Chem., Int. Ed. Engl.*, **9**, 589–595 (1970).
(109) R. Fugmann, H. Nickelsen, I. Nickelsen, and J. H. Winter, "Representation of Concept Relations Using the TOSAR System of the IDC: Treatise III on Information Retrieval Theory", *J. Am. Soc. Inf. Sci.*, **25** (5), 287–306 (1974).
(110) M. Osinga and A. A. Verrijn Stuart, "Documentation of Chemical Reactions. I. A Faceted Classification", *J. Chem. Doc.*, **13**, 36–39 (1973).
(111) M. Osinga and A. A. Verrijn Stuart, "Documentation of Chemical Reactions. II. Analysis of the Wiswesser Line Notation", *J. Chem. Doc.*, **14**, 194–198 (1974).
(112) J. E. Armitage and M. F. Lynch, "Automatic Detection of Structural Similarities among Chemical Compounds", *J. Chem. Soc. C*, 521–528 (1967).
(113) J. E. Armitage, J. E. Crowe, P. N. Evans, M. F. Lynch, and J. A. McGuirk, "Documentation of Chemical Reactions by Computer. Analysis of Natural Changes", *J. Chem. Doc.*, **7**, 209–215 (1967).
(114) R. Clingling and M. F. Lynch, "Production of Printed Indexes of Chemical Reactions. I. Analysis of Functional Group Interconversions", *J. Chem. Doc.*, **13**, 98–102 (1973).
(115) R. Clingling and M. F. Lynch, "Production of Printed Indexes of Chemical Reactions. II. Analysis of Reactions Involving Ring Formation, Cleavage, and Interconversion", *J. Chem. Doc.*, **14**, 69–71 (1974).
(116) J. Valls, "Reaction Documentation", in "Computer Representation and Manipulation of Chemical Information", W. T. Wipke, S. R. Heller, R. J. Feldmann, and E. Hyde, Ed., Wiley, New York, N.Y., 1974, pp 83–103.
(117) J. R. Koskinen and B. R. Kowalski, "Interactive Pattern Recognition in the Chemical Laboratory", *J. Chem. Info. Comput. Sci.*, **15**, 119–123 (1975).
(118) P. D. Jurs, "Chemical Data Interpretation Using Pattern Recognition Techniques", "Computer Representation and Manipulation of Chemical Information", W. T. Wipke, S. R. Heller, R. J. Feldmann, and E. Hyde, Ed., Wiley, New York, N.Y., 1974, pp 265–285.
(119) C. Hansch, A. Leo, and D. Elkins, "Computerized Management of Structure–Activity Data. I. Multivariate Analysis of Biological Data", *J. Chem. Doc.*, **14**, 57–61 (1974).
(120) S. R. Heller, R. J. Feldmann, H. M. Fales, and G. W. A. Milne, "A Conversational Mass Spectral Search System. IV. The Evolution of a System for the Retrieval of Mass Spectral Information", *J. Chem. Doc.*, **13**, 130–133 (1973).
(121) D. J. Floto, R. A. Hagstrom, T. L. Heying, R. E. Maizell, and W. J. Mayer, "Technological Forecasting—An Experiment Relating to the Pulp and Paper Industry", *J. Chem. Doc.*, **13**, 162–165 (1973).
(122) D. H. Smith, L. M. Masintes, and N. S. Sridharan, "Heuristic DENDRAL: Analysis of Molecular Structure", "Computer Representation and Manipulation of Chemical Information", W. T. Wipke, S. R. Heller, R. J. Feldmann and E. Hyde, Ed., Wiley, New York, N.Y., 1974, pp 287–315.
(123) M. F. Lynch, J. M. Harrison, W. G. Town, and J. E. Ash, Ed., "Computer Handling of Chemical Structure Information", Macdonald, London, 1971.
(124) W. T. Wipke, S. R. Heller, R. J. Feldmann, and E. Hyde, Ed., "Computer Representation and Manipulation of Chemical Information", Wiley, New York, N.Y., 1974.
(125) J. E. Ash and E. Hyde, "Chemical Information Systems", Ellis Harwood, Ltd., Chichester, England, 1975.

Status of Computers and the Information Industry and Potential Future Trends†

CARLOS M. BOWMAN* and PATRICIA F. ROUSH

Central Research–Computation Laboratory, The Dow Chemical Company, Midland, Michigan 48640

Received April 30, 1976

Computer technology has advanced at a tremendous rate in the last three decades. Its influence has had a very significant impact on chemical information science. Large amounts of information are now available at the fingertips of the user through interactive systems. Data acquisition systems now permit the direct recording of experimental data from the laboratory. The days of the printed journal are numbered, and the use of computers in the planning of experiments will soon be a reality.

It is truly a pleasure and an honor to participate in this symposium with such distinguished speakers, and particularly in the first Herman Skolnik Award Symposium. It has been my (C.M.B.) privilege to know Herman for many years. I have come to respect his grasp of the important aspects of information science and his perceptive evaluation of work in this most important field. The naming of an award for him is truly fitting. The object of this talk is to tell you about computers in chemical information.

It is not surprising that computers and computer technology have had a tremendous impact on chemical information science as they have on almost every other aspect of our life. It is our intention to sketch briefly the development of computer technology over the last three decades and to try to relate this development to what we consider the golden years of chemical information science. We will then attempt to suggest some areas where new developments will take place and what some of the problems facing us will be.

Computer technology has developed in three major areas: processing capability, storage capacity, and real time interactive systems.

Processing capability has grown from operational speeds of a few seconds to a few nanoseconds. That is an improvement of almost nine orders of magnitude. Storage capacity has increased from being almost nonexistent, when data were stored on cards, to a very limited amount of magnetic drum and core storage, to solid-state storage which now permits immediate access to several million characters. One of the key developments in storage capacity was the introduction of magnetic media, including tapes, followed by disks. There is virtually no limit to the amount of information that can be put on-line in a computer system today.

Most computer usage in the early days was on the batch basis. It has only been seven or eight years since interactive systems began to make significant in-roads in this area. Today, a large portion of all computing is done in an interactive or

† Presented at the Award Symposium on "Contributions of the Division of Chemical Literature (Information) to the Chemical Society" at the ACS Centennial Meeting, New York, N.Y., April 6, 1976.