

to inefficiency. If the algorithm were entirely in a more efficient language, we would expect a factor of 10 to 100 decrease in running time, a factor based on our experience in converting similar programs to more efficient languages. We are investigating an exportable version of MAXSUB which would have this desired efficiency.

Another way to reduce the computation time is to find a more efficient way to canonicalize structures or reduce the number of necessary canonicalization steps. The former is unlikely to yield much improvement because of the effort that went into studies of efficiency to begin with. The latter might yield to a more intelligent procedure for growing nuclei. The part of the algorithm which grows nuclei adds one atom at time. This procedure can generate many duplicate subgraphs by following different paths. Prospective elimination of such duplicates is desirable and we are working on this problem.

Since our structural representation is topological, information about the geometry of the resulting common subgraphs is not available. We consider this a serious limitation for studying biological structure activity relationships. Work on entering stereochemical information is in progress.⁹ As was mentioned in the discussion, associating properties with atoms and bonds is done in one module of the program and even now the set of descriptors could be enlarged in ways which might implicitly include some geometrical information, e.g., as ring membership does already. Addition of other structural properties, such as stereochemistry of certain atoms and relative locations of atoms in a structure, is possible but more difficult to implement in a general way.

REFERENCES AND NOTES

- (1) Part XXX of the series "Applications of Artificial Intelligence for

- Chemical Inference". For part XXIX see J. G. Nourse, R. E. Carhart, D. H. Smith, and C. Djerassi, *J. Am. Chem. Soc.*, **101**, 1216 (1979).
(2) We wish to thank the National Institutes of Health (RR-00612, GM20832, and GM06840) for their generous financial support for our research and for their support (RR-00785 SUMEX) of the SUMEX computer resource on which the MAXSUB program is available via nationwide computer networks.
(3) A. J. Stuper and P. C. Jurs, *J. Am. Chem. Soc.*, **97**, 182 (1975).
(4) E. J. Corey, W. T. Wipke, R. D. Cramer, III, and W. J. Howe, *J. Am. Chem. Soc.*, **94**, 421, 431 (1972).
(5) (a) J. E. Armitage and M. F. Lynch, *J. Chem. Soc. C*, 521 (1967); (b) J. E. Armitage, J. E. Crowe, P. N. Evans, M. F. Lynch, and J. A. McGuirk, *J. Chem. Soc.*, **7**, 209 (1967); (c) J. M. Harrison and M. F. Lynch, *J. Chem. Soc. C*, 2082 (1970).
(6) M. M. Cone, R. Venkataraghavan, and F. W. McLafferty, *J. Am. Chem. Soc.*, **99**, 7668 (1977).
(7) F. A. Gorin and G. R. Marshall, *Proc. Natl. Acad. Sci. U.S.A.*, **74**, 5179 (1977).
(8) (a) T. H. Varkony, R. E. Carhart, and D. H. Smith in "Computer Assisted Organic Synthesis", W. T. Wipke and W. J. Howe, Eds., American Chemical Society, Washington, D.C., 1977, p 188; (b) R. E. Carhart, T. H. Varkony, and D. H. Smith in "Computer Assisted Structure Elucidation", D. H. Smith, Ed., American Chemical Society, Washington, D.C., 1977, p 126; (c) T. H. Varkony, R. E. Carhart, D. H. Smith, and C. Djerassi, *J. Chem. Inf. Comput. Sci.*, **18**, 168 (1978).
(9) (a) J. G. Nourse, R. E. Carhart, D. H. Smith, and C. Djerassi, ref 1; (b) J. G. Nourse, *ibid.*, **101**, 1210 (1979).
(10) R. E. Carhart, MIP-R-118, Machine Intelligence Research Unit, University of Edinburgh, United Kingdom, 1977.
(11) C. Djerassi, *Pure Appl. Chem.*, **50**, 171 (1978).
(12) Using this method, chemist-selected portions of structures are converted by REACT to single "atoms", thereby losing information on the relative points of attachments with the rest of the structure. If the portion so replaced has points of attachments which are *symmetrically equivalent*, then there is no problem in interpretation of results. Otherwise, there is ambiguity in selecting points of attachment which in some instances may yield substructures which are not in fact in common.
(13) The program runs on a Digital Equipment Corp. PDP-10 computer at SUMEX computer facility, Stanford University. The program is available (to the limit of available resources) on-line to interested persons over three nationwide computer networks. For information on access to the program please contact the authors.

A Representation of π Systems for Efficient Computer Manipulation

JOHANN GASTEIGER

Institut für Organische Chemie, Technische Universität München, D-8046 Garching, West Germany

Received September 18, 1978

A data structure has been developed which is particularly well suited for the representation of π systems. This representation is based on a separation of σ , π , and n electrons. All valence electrons in molecules and reactions can be accounted for, giving algebraic properties to this description. The representation is excellently amenable to computer manipulation, and programs based on it have been developed. It can also serve as an interface to valence bond or molecular orbital calculations.

The automatic manipulation of chemical structures has become of increasing importance. Documentation and information retrieval, as well as deductive computer programs, ask for efficient data structures.¹ In particular, the selection of a representation for chemical compounds can greatly influence the performance of a system. For certain problems special data structures may offer advantages, but for a general-purpose system an unambiguous topological representation seems indispensable. Fragmentation codes and linear codes have previously been chosen because of their compactness, but for certain problems they might need decompressing and data conversion. Furthermore, some primary information on the topology of a structure might no longer be accessible.

With the advances in computer technology, storage space has lost its crucial importance. Rather, it is more desirable that in the design of a system such a representation for molecular structures is chosen which retains direct information

on atoms and bonds. This provides for an open-ended approach to the manipulation of molecules. Thus, in computer programs for chemical problems, topological representations have gained a similar importance as has the structural formula for conventional chemical communication.

In a topological representation for molecular structures information on the atoms of a molecule and their connectivities is provided. In the program systems that we are developing²⁻⁴ this information is enriched by data on bond orders and free electrons. Thus, in effect, account is taken of all valence electrons of a molecule. This allows immediate access to valence states, charges, etc. It also provides the basis for the exhaustive generation of chemical reactions which are considered as bond and electron shifting processes and are generated accordingly.²⁻⁴

This work was initiated by a mathematical model of constitutional chemistry.^{5,6} In this model molecules and ensembles of molecules are represented by so called BE

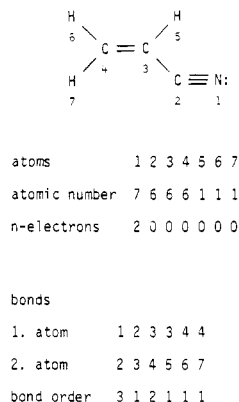
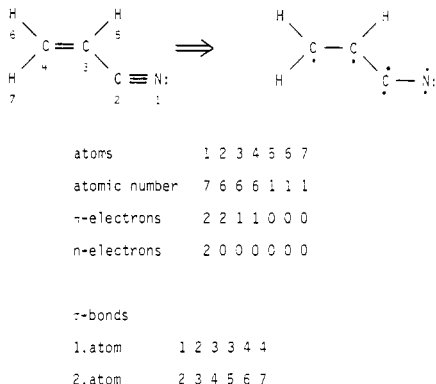


Figure 1. Internal representation of acrylonitrile.

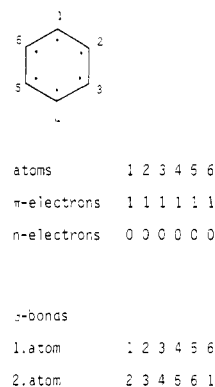
Figure 2. σ/π -separated representation of acrylonitrile.

matrices. Internally, in the program, more concise data structures are used. Figure 1 illustrates the internal representation of molecules. As can be seen, by carrying multiple bond orders and free electrons, all valence electrons are accounted for. The data structures chosen have excellently served their purpose in the programs we are developing. Only with conjugated and aromatic systems some shortcomings arise. As this representation is basically a valence bond description, different resonance structures of the same molecule receive different representations. While this may sometimes be desired, it certainly is a drawback when comparing molecules and searching files. Different representations require multiple entries for the same molecule. This is particularly inconvenient for documentation and information retrieval.

We therefore searched for a data structure which yields the same representation for different resonance structures, and retains the algebraic properties of our previous data structure. We came up with a representation which achieved this goal. Furthermore, we were rewarded with additional powerful properties of this new representation.

σ/π SEPARATION

Resonance structures of conjugated molecules differ only in the arrangements of π bonds; the σ skeleton stays untouched. It is therefore advantageous to separate σ and π electrons when representing molecules. Then, a notation with multiple bonds has to be abandoned. But we still wanted to give account of all valence electrons to retain the algebraic capabilities of our representation. Therefore, π electrons have to be carried along. Multiple bonds are now split in such a way as to assign a π electron to each center connected by the π bond. Figure 2 demonstrates this process for acrylonitrile. As the σ skeleton comprises only single bonds, entries for the bond order can be dropped and the σ bonds can now be represented only by the indices of the two atoms. This representation can be put into a pictorial form by assigning dots for the π electrons to

Figure 3. σ/π -separated representation of the carbon skeleton of benzene.

	1	2	3	4	5	6	7
	N	C	C	C	H	H	H
1	4	4	1	0	0	0	0
2	0	1	2	1	0	0	0
3	0	0	1	1	1	0	0
4	0	0	0	1	1	0	1
5	H	0	0	1	0	0	0
6	H	0	0	0	1	0	0
7	H	0	0	0	0	1	0

Figure 4. Matrix form of the representation with σ and π electrons separated for acrylonitrile.

the atoms as also shown in Figure 2. This has the immediate result that uncharged resonance structures receive the same representation as shown for benzene in Figure 3.

In charged structures the π electrons are shifted until the atom with the lowest index carries the charge. For dipolar structures the π electrons are shifted in such a manner as to annihilate the charges. Thus they get the same representation as the uncharged structures.

As with our previous representation, we also take account of the free electrons. Then we have a data structure which treats σ , π , and n electrons separately. This classification of the valence electrons has proved its usefulness in organic chemistry so frequently that it is taken as an inherent one. It is quite natural that this distinction of σ , π , and n electrons offers advantages when representing molecules in computer programs. Difficulties with the proposed representation may only arise in systems where the σ/π separation is not clear as with nonclassical ions. Then, one of the classical forms has to be taken.

Valence states of atoms can automatically be accessed directly, as σ , π , and n electrons are carried separately. Comparing the sum of the σ bonds and π and n electrons with the inherent number of valence electrons of an atom allows one to determine the charge on an atom. For atom 1 in Figure 2 the sum of σ bonds and, π and n electrons is $1 + 2 + 2 = 5$ which is equal to the valence electrons of an atom with atomic number 7 (nitrogen). Thus, this atom is uncharged. Further, it is an allowed valence state of nitrogen.

Our new representation can also be written in matrix form: π and n electrons can be carried together on the diagonal, whereas the σ skeleton is given by the off-diagonal elements (Figure 4). The numbers of the diagonal entries e_{ii} give the sum of the π and n electrons of atom i . Entries e_{ij} of 1 in the off-diagonal signify bonds between atoms i and j . Except for the diagonal entries, this matrix is an adjacency matrix, but the diagonal entries are very important. They provide these matrices with the same algebraic properties as have BE matrices.^{5,6} But different uncharged resonance structures of conjugated systems are now represented by the same matrix

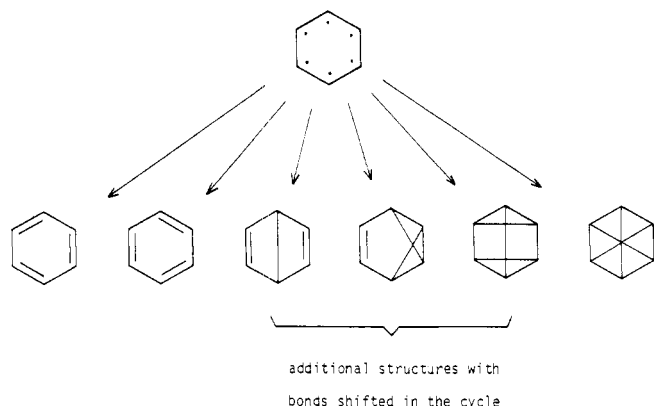


Figure 5. Generation of valence bond structures of benzene.
(indexing problems being disregarded).

ADVANTAGES OF THIS REPRESENTATION

Data Structures and Data Conversions. The initial goal, to develop a unique representation for the various resonance structures of a conjugated system, has been attained. To have a single description for a conjugated or aromatic molecules, and not have to handle various resonance structures of the same compound, is especially important in documentation and information retrieval. Conversion from a representation with multiple bond orders to the new representation is fast and straightforward.

On the other hand, from this representation with σ and π electrons being separated, all resonance structures of a given molecule can be generated easily. This is achieved by coupling two π electrons to a π bond. Permuting through all possible combinations allows exhaustive generation of all resonance structures. In Figure 5 the uncharged resonance structures of benzene are given. It should be noted that bonds across the ring in Figure 5 are π bonds and not σ bonds. Thus, the graphs do not represent valence tautomers. On the other hand, valence tautomers can also be generated by interconverting π electrons and σ bonds (see below).

Converting from the σ/π -separated representation back to a representation with multiple bonds might be desired when generating structural formulas for output. This can easily be performed by mutually coupling π electrons. In this process of coupling π electrons to multiple bonds care has to be taken to differentiate between alkynes and allenes, in particular between polyalkynes and cumulenes, as in both systems atoms with two π electrons are involved. But allenes and cumulenes have at both ends of the conjugated system an atom with only one π electron. Therefore they can be differentiated from alkynes and polyalkynes.

By shifting π electrons first, and then forming π bonds, dipolar structures can be generated, too (Figure 6). Thus, in effect, all valence structures (covalent and ionic) of a given compound can automatically be generated. For charged species and radicals the various resonance structures can be obtained by consecutively positioning the charge or free electron on the next nearest neighbor.

Aromaticity. The σ/π -separated representation allows the rapid detection of aromatic rings. First, only those rings have to be considered where each atom carries at least one π electron or a charge. For heteroatoms having at least two n electrons and not having received π electrons from multiple bonds (as has N in pyridine), two n electrons have to be counted as π electrons (e.g., for N in pyrrole). For even-membered rings the number of π electrons involved in the delocalization is obtained by summing over the π electrons of all ring atoms. In this process only one π electron must be taken from each center except when this center also carries

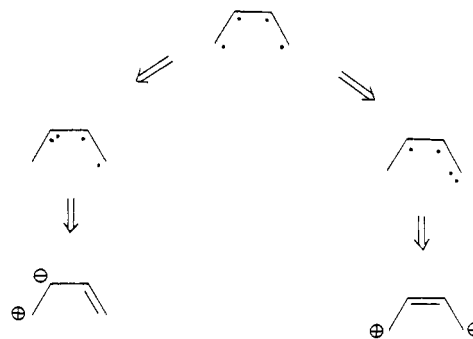


Figure 6. Generation of dipolar resonance structures.

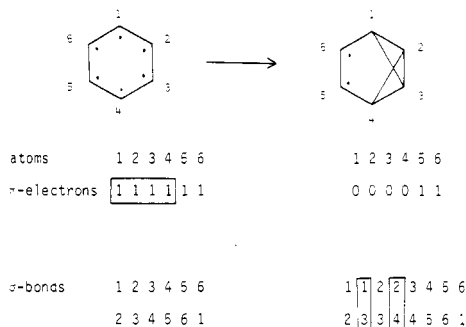


Figure 7. Generating benzvalene from benzene.

a charge. For odd-membered rings the same procedure has to be followed with the exception that two π electrons of a single center may be taken if this atom carries a negative charge or is a heteroatom. In this process not only the smallest rings⁷ but also envelopes of rings have to be considered, e.g., for azulene or dihydropyrene.

Stereochemistry. Another important application of the representation developed here can be found for the specification of stereochemistry. Cahn, Ingold, and Prelog⁸ had to use much care and pain for incorporating conjugated and aromatic molecules into their system of sequence rules. After giving rules which had to be abandoned later, they came up with rules for selecting a specific resonance structure as the reference state. These rules are artificial and cumbersome.

Our representation does not suffer from having different resonance structures for an uncharged species. For charged species or radicals simple rules can be given to specify a reference state (e.g., positioning the charge or free electron on the atom with lowest index). Then, sequence rules take a much simpler and less artificial form to define priorities of ligands. In addition, these rules are more amenable to computer manipulation.

Generating Reactions. The process of generating valence tautomers is illustrated with an example in Figure 7. By converting π electrons of atoms 1 and 3, and of atoms 2 and 4 to two σ bonds, benzvalene is obtained from benzene. If desired, the residual π electrons can be coupled to π bonds to convert to a representation with multiple bond orders. Permuting through all possible combinations of transforming two, four, or six π electrons to one, two, or three σ bonds allows the generation of all those valence tautomers where the primary cyclic structure is conserved.

The usual valence bond description of π systems has disadvantages when treating pericyclic reactions. For a given type of reaction the number of bonds that have to be rearranged is dependent on the size of the π system involved. For example, in the electrocyclic ring closure reaction of 1,3-butadiene, two (π) bonds are broken and two (one π and one σ) bonds are made. In the corresponding reaction of 1,3,5-hexatriene, on the other hand, three (π) bonds are broken and three (two π and one σ) bonds are made (Figure 8).

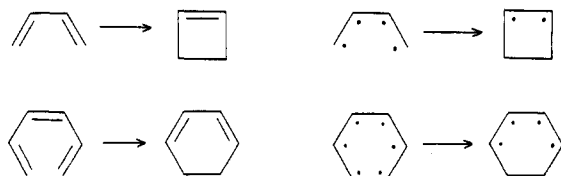


Figure 8. Electrocyclic reaction of 1,3-butadiene and 1,3,5-hexatriene.

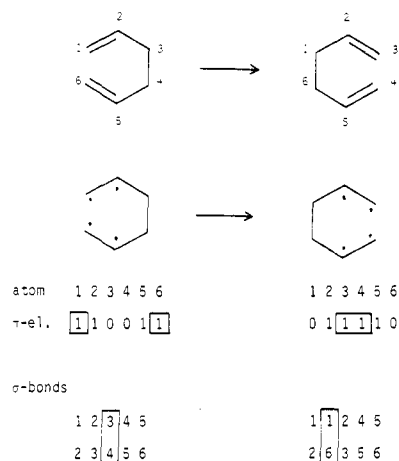


Figure 9. Cope rearrangement.

Thus, although both reactions have much in common and are classified into the same type, they have to be generated by two different bond rearrangements. With the σ/π -separated representation, on the other hand, both reactions can be generated by a single reaction operator which converts two π electrons into a σ bond (see Figure 8). Thus, the representation developed here has exposed the essential common feature of these two reactions.

Analogous situations hold for the other types of pericyclic reactions: electrocyclic ring openings, cycloadditions and cycloreversions, sigmatropic rearrangements, and cheletropic reactions. Each such type can be generated by a uniform reaction operator. In Figure 9 the necessary operations for performing a Cope rearrangement are illustrated. Changes occur only at the reaction site, at atoms 1, 3, 4, and 6: breaking one σ bond to two π electrons and making a σ bond from two π electrons. The valence configurations at atoms 2 and 5 are unaltered.

A program system for the exhaustive generation of all pericyclic reactions for a given π system has been developed⁹ which works with such uniform reaction generators on the σ/π -separated representation.

Molecular Orbital Methods. As the representation of π systems containing multiple bonds has similarities with a valence bond description of a molecule, the σ/π -separated representation is reminiscent of a molecular orbital description. We have split a molecule into a σ skeleton and into π electrons which make up the π MOs. Writing the σ/π -separated representation in matrix form has the π (and n) electrons on the diagonal and gives in the off-diagonal elements the topology of the σ skeleton (see Figure 4). This matrix has entries in the same positions as the Hückel matrix for the molecule: in the diagonal, entries of α - ϵ , and off-diagonal, entries of β . This similarity between the Hückel matrix and the topological matrix has been pointed out before.¹⁰

Now the way is paved for a direct coupling of an HMO program to any system using the σ/π -separated representation. In particular, when the molecular structures are generated by a program they can be submitted automatically to an evaluation by the HMO method as the data structure allows a direct matching. This evaluation can serve as a basis for a selection procedure.

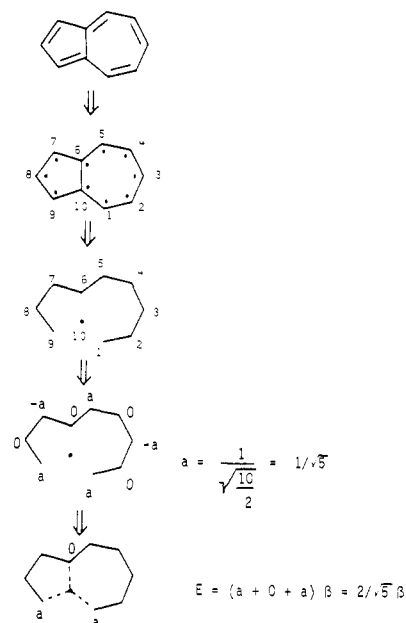


Figure 10. Calculation of the PMO energy of azulene.

An automatic evaluation by the perturbational MO (PMO) method¹¹ can easily be achieved, too. Excising the end atom of a π system (cyclic or noncyclic) or an atom which closes the ring of a π system and then treating the interactions of this atom with the residual system allows the estimation of the PMO energy of the complete system.

Figure 10 illustrates the algorithmic process with azulene. First, the atoms of the π system are ordered and indexed in a consecutive manner. In this case, the starting atom is defined as an atom adjacent to a tricoordinated center and lying on the longest path from this tricoordinated center to the other one. Removing atom 10 and its connectivities yields a π system of nine atoms. These atoms receive coefficients of a , $-a$, and 0 in a regular manner, as indicated. The magnitude of the coefficient a can be calculated from the normalization condition according to

$$\sum a_i^2 = 1$$

The perturbational MO energy of the entire system is obtained by summing over the coefficients a_i of those atoms directly connected to the atom previously removed.

$$E = \sum a_i \beta$$

In our case, an energy of $E = 2a\beta = 2/\sqrt{5}\beta$ is calculated.

PMO energies reflect, within a series, differences between aromatic, olefinic, and antiaromatic behavior. Thus, the simple procedure outlined above incorporated into a program system manipulating π systems can automatically give estimates on the reactivity of a molecule and the energy of products and thus decide among various alternatives for the outcome of a reaction. Again, a system which directs its own course can be developed.

CONCLUSION

The representation based on the separation of σ , π , and n electrons is excellently amenable to computer manipulation. Many of the properties mentioned have already been exploited in various algorithms. A program system⁹ based on this data structure has been developed. It generates all conceivable pericyclic reactions for a given molecule without having to resort to known reactions only. The discussion of valence bond, HMO, and PMO methods should have demonstrated how quantum mechanical calculations can be put to work in a completely automatic program system performing tasks of

estimating reactivities and evaluating various reaction alternatives. This evaluation can serve as a basis for a selection procedure. Programs, directing their own paths can be developed using the computer as an automatic generating and decision-making machine.

ACKNOWLEDGMENT

I am indebted to Professor Ivar Ugi for introducing me to matrix representation of chemical structures. Part of this work was supported by Deutsche Forschungsgemeinschaft.

REFERENCES AND NOTES

- (1) M. F. Lynch, J. M. Harrison, W. G. Town, and J. E. Ash, "Computer

- Handling of Chemical Structure Information", Macdonald, London, 1971.
- (2) J. Blair, J. Gasteiger, C. Gillespie, P. Gillespie, and I. Ugi in "Computer Representation and Manipulation of Chemical Information", W. T. Wipke, S. Heller, R. Feldmann, and E. Hyde, Ed., Wiley, New York, 1974.
- (3) J. Blair, J. Gasteiger, C. Gillespie, P. Gillespie, and I. Ugi, *Tetrahedron*, **30**, 1845 (1974).
- (4) J. Gasteiger and C. Jochum, *Top. Curr. Chem.*, **74**, 93 (1978).
- (5) I. Ugi and P. Gillespie, *Angew. Chem.*, **83**, 980, 982 (1971); *Angew. Chem., Int. Ed. Engl.*, **10**, 914, 915 (1971).
- (6) J. Dugundji and I. Ugi, *Top. Curr. Chem.*, **39**, 19 (1973).
- (7) J. Gasteiger and C. Jochum, *J. Chem. Inf. Comput. Sci.*, **19**, 43 (1979).
- (8) R. S. Cahn, C. Ingold, and V. Prelog, *Angew. Chem.*, **78**, 413 (1966); *Angew. Chem., Int. Ed. Engl.*, **5**, 384 (1966).
- (9) J. Gasteiger, *Z. Naturforsch.*, **34b**, 67 (1979).
- (10) Hs. H. Günthard and H. Primas, *Helv. Chim. Acta*, **39**, 1645 (1956).
- (11) M. J. S. Dewar, "The Molecular Orbital Theory of Organic Chemistry", McGraw-Hill, New York, 1969.

NEWS AND NOTES

NEWS ITEMS

New Publication

"Index to Scientific and Technical Proceedings" (ISTP), a monthly, has been introduced by the Institute for Scientific Information (ISI) to cover about 10,000 meetings that take place each year. ISTP is multidisciplinary and indexes published proceedings, i.e., as books, reports, preprints, and from the journal literature. Entries in each issue are arranged by category, permuted subject, author and editor, sponsor, meeting location, and corporate. Annual subscription price is \$500.

CAS Profiles for API

Chemical Abstracts Service is now helping scientists and engineers in the member companies of the American Petroleum Institute keep abreast of new publications and patents on petroleum chemistry, energy technology, and related chemical and chemical engineering topics. API has contracted with CAS to search the computer files of information being processed for current issues of *Chemical Abstracts* and select from them abstracts of interest to particular individuals within API member companies.

Abstracts are selected according to a keyword profile that defines an individual scientist's or engineer's specific interests. CAS mails computer printouts of the selected abstracts to participants every two weeks.

CAS agreed to perform the searches and mail selected abstracts directly to the API clients. API's Central Abstracting and Indexing Service will continue to be the principal contact for clients and will work with them in formulating and modifying search profiles.

CAS currently is processing nearly 300 profiles for API clients. CAS views the arrangement as a pilot project to assess the feasibility of making such a service available generally.

AIDIC Meetings

The Association for Information and Dissemination Centers (ASIDIC) announces the following meetings for 1979.

"The Impact of Mini and Micro Computers on Bibliographic Data Bases", Ottawa, Canada. June 3-5, 1979. Host: CAN/SDI Centers, Canada.

"Non-Bibliographic Data Bases", Boston, Massachusetts. September 16-18, 1979. Host: New England Research Application Center, Storrs, Connecticut.

For further information and/or registration material, write: ASIDIC, P.O. Box 8105, Athens, Georgia 30603.

ISI Library Grant Program

The Institute for Scientific Information (ISI) announced the continuation and expansion of its program of grants, with the introduction in 1979 of three levels of grant eligibility. The program, started in 1970 to help libraries provide improved reference services, now offers three different grant awards toward the purchase of ISI indexes to the literature of the sciences, social sciences, and arts and humanities.

Libraries that qualify for grant aid may receive awards allowing the purchase of ISI's three major indexes at 75%, 50%, or 30% of the full list price, depending on the type and size of library. ISI expects to award more than \$800,000 this year toward the purchase of the "Science Citation Index", "Social Sciences Citation Index", and "Arts & Humanities Citation Index". Over \$600,000 was awarded in new grants and in continuing assistance last year, with sixty libraries receiving new awards.

The program continues under the direction of Dr. M. V. Malin, Vice President-Professional Relations and Contract Research for ISI, who formerly served as a grant administrator at the National Science Foundation. Those wishing application forms should write to the Grant Administrator, Institute for Scientific Information, 325 Chestnut St., Philadelphia, Pennsylvania 19106.

Conferences: Technical Information—Which Way is the Future?

The Engineering Foundation will sponsor a conference on "Technical Information—Which Way Is the Future?," chaired by R. G. M. Cosgrave, Exxon Research and Engineering Co. Members of the organizing committee include C. C. Beazley, General Foods Technical Center, W. G. Stanley, Standard Oil Research Center, D. N. Streeter, IBM Corp., R. E. O'Dette, Chemical Abstracts Service, and Ellen Pearson, University of Guelph. The conference will be held at Franklin Pierce College, Rindge, New Hampshire, July 29–August 3, 1979.