ate $(-\lambda + c)$ where c is some numerical correction for electronegativity and/or σ-bond energy. The resulting polynomial is a function of λ, the usual energy eigenvalue, and when equated to zero is identical with the traditionally obtained secular equation in appropriate energy units.

SUMMARY

The physico-geometric representation of the ACMCP is described by the expression

$$\mathrm{ACMCP} = \sum_{k=0}^{N} \sum_{*} \left\{ (-1)^{\left(\sum_{i=2}^{N-k} n_i(i-1)\right)} \cdot (2)^{\left(\sum_{i=3}^{N-k} n_i\right)} \cdot \left[\prod_{j=2}^{N-k} \prod_{m=0}^{n_j} \textcircled{j}_m \right] S_k \right\}$$

where

$\sum_{*}$ = summation over all different combinations of nonadjacent rings where

$$\sum_{i=2}^{N-k} in_i = N - k$$

"combination" = collection of nonadjacent rings whose connectivities enter into evaluation of a particular coefficient in ACMCP

i = integer, $2 \leqq i \leqq N - k$

j = integer, $2 \leqq j \leqq N - k$

k = integer, $0 \leqq k \leqq N$; identifies degree of ACMCP term

m = integer, $0 \leqq m \leqq n_j$

n_i = integer, $0 \leqq n_i \leqq [(N - k)/i]$, where upper limit is the largest integer value of $(N - k)/i$; represents number of nonadjacent (nontouching) i-membered rings in a combination

$\textcircled{j}_m$ = product of bond connectivities around (defining) the mth nonadjacent j-membered ring

S_k = product of atomic symbols, k in number, which are *not* included in the rings making up a given combination

$S_0 \equiv$ 1, for any j; a definition

$\textcircled{j}_0 \equiv$ 1; a definition

Acknowledgment.—Acknowledgment is gratefully made to referee Dr. Sylvan Eisman, Pitman-Dunn Institute for Research, U. S. Army, Frankford Arsenal, Pa., who called the author's attention to a paper by F. Harary, [*J. Math. Phys.*, 38, 104 (July, 1959)] which presents a graph-theoretic study anticipating some of the relations independently found here. However, it is couched in mathematical terminology not normally understood by the average chemist and does not recognize the significance of explicit atom (vertex) labeling so important for molecular pictographs, particularly the heteroatomic ones.

Chemical Structure Fragmentation for Use in a Coordinate Index Retrieval System*

By RAY W. IHNDRIS
Cancer Chemotherapy National Service Center, National Cancer Institute,
National Institutes of Health, Bethesda, Maryland

Received January 28, 1964

This paper describes the changes that have been made in the Cancer Chemotherapy National Service Center (CCNSC) Chemical Information Retrieval System[1] established for analog searching of chemical structures.

The National Bureau of Standards Peek-a-Boo System[2] in use at the CCNSC consists of a precision-built machine which uses a microswitch to actuate the solenoid-operated punch when the plastic card is placed into punching position. The cards used for permanently recording "data" are 5 × 8 in., 8–10 mil vinyl plastic sheets. Each card is punched separately.

The Reader consists of a back-lighted opaque frame onto which the plastic cards are placed and over-layed with a grid so that the coincident holes allow light to shine through, and file numbers can be "read" from the coordinate position on the grid. The thousands and hundreds are read along the horizontal axis, and the tens and units are read on the vertical axis. The grid provides for 18,000 recording positions.

A VISIrecord file cabinet houses the plastic cards in a manner permitting easy access. The cards are shingled so that three-fourths of an inch along the top edge exposes the chemical fragment or other descriptor. The cards are notched on the right-hand edge and filed vertically so that they fit over rods in the bottom of the file cabinet. They are printed to define the space for the descriptors and also have blocks along the left-hand edge that signal the position when a card is removed. Thus it is easy to find a card for entering new data or retrieving information.

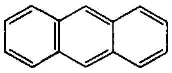
One or several descriptors are used for selection by overlaying the plastic cards, thus eliminating unwanted information. We can vary the search at any time if the desired information is not found, or conversely, if too much chaff appears.

The improvements in our system of coding chemical structures affect the ring systems, the fragment chains, and the supplier code.

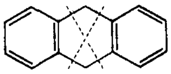
* Presented before the Division of Chemical Literature, 145th National Meeting of the American Chemical Society, New York, N. Y., Sept. 11, 1963.

Ring Systems.—The ring systems found in the chemicals of the CCNSC program are many and varied. Several hundred systems occur in each set of 18,000 compounds. Exactly how many appeared in the earlier sets is unknown because of the way the ring systems were recorded. Only aromatic rings (A) were kept as a unit (plus a few special heterocyclic rings). Partially hydrogenated aromatic rings (B and C) were split into fragments.

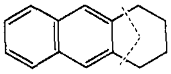
A. Kept as a unit



B. Split



C. Split



In our present system all fused and spiro rings are kept intact as indexing units.

Fragmenting a structure containing a ring system requires looking up the Revised Ring Index Number (RRI #) in "The Ring Index,"[3] which may be accomplished in several ways.

First, if one knows the name of the particular system, it can be found quickly from the alphabetic Index of Names in the back of the book. Second, if the system is not identified by name, one can search by (1) the number of rings composing the system, (2) the ring size, and (3) the ring formula. Acridine, for example, is found in the "three-ring system" in a series 6,6,6 under the ring formula C_5N–C_6–C_6. Each system has a serial number, which we use for coding into the Peek-a-Boo cards. In this case it is RRI-3523.

Our method for coding all possible ring systems uses a special group of 44 plastic cards which are numbered as follows.

0000	-000	--00	---0	.1
1000	-100	--10	---1	.2
2000	-200	--20	---2	.3
3000	-300	--30	---3	.4
4000	-400	--40	---4	
5000	-500	--50	---5	
6000	-600	--60	---6	
7000	-700	--70	---7	
8000	-800	--80	---8	
9000	-900	--90	---9	

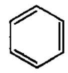
More than 10,000 ring systems can be recorded and retrieved with these cards. The four cards circled are pulled from the VISIrecord file and punched with the appropriate accession number (NSC-) in its proper coordinate position. It is necessary to punch at least four cards for each ring number.

If the particular ring system is not listed in "The Ring Index," we use the RRI # closest to where it would be placed according to *Chemical Abstracts*' rules of arranging by ring size and ring formula and add .1 to its serial number.

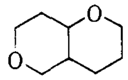
All single rings, fused rings, bridged rings, spiro rings, and steroid rings are included in this method of coding without regard to their degree of saturation. Previously, saturated rings were fragmented into chains or groups and were difficult to retrieve.

Occasionally, two or more ring systems may occur in the same structure. About 22 of the most common and frequently occurring rings have been entered on separate cards and not by their RRI #. This was done partly for easy retrieval and partly to reduce the need for punching two RRI #'s for the same compound which would result in false readings of numbers during retrieval. Substitution positions are also punched for benzene and cyclohexane rings, for example, mono, *ortho, para,* symmetrical, asymmetrical, and four or more substitutions.

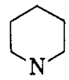
To retrieve ring systems in a more general way, the following descriptors may be used. Hetero refers to substitutions in rings of elements other than carbon. Punch "zero hetero" if there is no hetero atom as in benzene or one or more "hetero" depending on number of noncarbon atoms in each ring system.



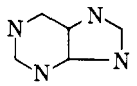
zero hetero



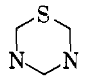
two hetero
2 O cyclic



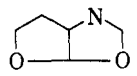
one hetero
1 N cyclic



four hetero
4 N cyclic



three hetero
1 S cyclic
2 N cyclic



three hetero
1 N cyclic
2 O cyclic

Several hetero number cards may be punched depending on how many ring systems there are in one compound. The N, O, P, or S cyclic cards are also punched according to the total number of each element in each ring system.

When a double bond occurs in a nonaromatic ring, punch $>C=C<$ cyclic, or when a ketone group is part of a ring, punch $>C=O$ cyclic, or other elements such as $>C=N-$ cyclic, the noncarbon element may be either in or out of the ring. Ring systems that contain fragment groups are coded by both the RRI # and the linear fragment system.

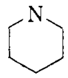
We have found this new method for coding ring systems to be very specific when using the RRI # and quite general when using the allied descriptors (*i.e.*, the hetero cards plus the cyclic designations and/or the fragment chains). Searching time is lessened and fewer false drops occur.

In the new system, two punches are required for each benzene ring, *i.e.*, the ring card plus the substitution position card, but this change solves one of our big retrieval problems. To make a survey of, say, all nitro benzenes, the $-NO_2$ card would have to be compared with each of the eight different benzene cards. With the new system we need only make one comparison, using the master benzene card and the nitro card. A more specific search can easily be made with the addition of one of the "position" cards.

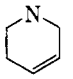
The pyridine, piperidine, and partially unsaturated piperidine rings are combined on one card. This is true also for the pyrimidine ring. The N^- cyclic can distinguish the "onium" rings from the others. This will assure completeness in the general analog type of search. The old method would split the piperidine ring into the following fragments.

1 N cyclic
1 ring, 6 members
$(\text{—CH}_2\text{—})_5$

for



or

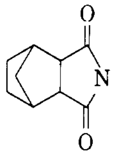
1 N cyclic
1 ring, 6 members for 
1 $-C=C-$
1 $-CH_2-$
$(-CH_2-)_2$

Many unwanted compounds would be retrieved when trying to reconstruct this ring. Each substitution on the ring requires a change in the search procedure without creating more specificity because these carbon fragments are used for chains as well as rings and may occur in other segments of the structure. The new system retains all rings as a unit without regard to degree of saturation.

For comparison, norbornane-2,3-dicarboximide would be entered into the system as follows.

Old

Fused ring 6,5 $-\overset{\overset{O}{\|}}{C}-N-\overset{\overset{O}{\|}}{C}-$
1 N cyclic $(-CH_2-)_2$
1 ring, 5 members $(-CH-)_4$
1 ring, 6 members $1-CH_2-$
Bridge



New
RRI # 2242
$>C=O$ cyclic

Nine plastic cards are required to reconstruct this formula, and this doesn't specify the exact structure or allow for substitutions on the ring. Only four RRI # cards plus a $>C=O$ cyclic for the carbonyl group are required for this specific structure. For the norborn*ene* structure, add the $>C=C<$ cyclic card. A ring system can be obscured by extraneous "drop-outs" if it is fragmented into its elements. We have eliminated that possibility by retaining it as a whole ring. A general search for similar ring systems can be made using combinations of any of the following headings as they fit the requirements of the search.

1 hetero atom
1 nitrogen cyclic
Bridge
$>C=O$ cyclic

$$-\underset{1}{\overset{\overset{O1}{\|}}{C}}-\underset{2}{N}-\underset{3}{\overset{\overset{O3}{\|}}{C}}-$$

Chain Fragmentation.—The second part of our fragmentation system to be revised is the fragment chain of atoms, which is defined as: a series of atoms linked one to the other by single or multiple bonds in which there are no singly bonded carbon-to-carbon atoms.

Following established rules, most organic radicals and functional groups attached to carbon chains or aromatic rings are indexed as units if there are less than four atoms in the radical or group. As in the past, separate cards are made for the following groups.[4]

$-NO$	$>C=O$	$-N=C<$	$-O-C-O-$
$-NO_2$	$-NCO$	$-N=N-$	$C=N-N<$
$-CN$	$-NCS$	$>N-N<$	$-COOH$
$-CNS$	$-SCN$	$-N-C-N-$	etc.

The revision of chain fragments involves groups having four atoms in a chain or three atoms in a chain plus one one or more branches on the chain. For example

$$>\underset{1}{N}-\underset{2}{N}-\underset{3}{\overset{\overset{O3}{\|}}{C}}-\underset{4}{N}< \quad \text{fragmented}$$

$$>\underset{1}{N}-\underset{2}{\overset{\overset{O2}{\|}}{C}}-\underset{3}{N}< \quad \text{fragmented}$$

$$>N-\overset{\overset{O}{\|}}{C}- \quad \text{separate card}$$

A special group of sixty plastic cards is used to code the very large number of combinations possible with nitrogen, oxygen, phosphorus, sulfur, and carbon in any one of five positions in a chain. The cards are designated as follows.

Position 1	Position 2	Position 3	Position 4	Position 5
1 N	2 N	3 N	4 N	5 N
1 N=	2 N=	3 N=	4 N=	5 N=
1 O	2 O	3 O	4 O	5 O
1 O=	2 O=	3 O=	4 O=	5 O=
1 P	2 P	3 P	4 P	5 P
1 P=	2 P=	3 P=	4 P=	5 P=
1 S	2 S	3 S	4 S	5 S
1 S=	2 S=	3 S=	4 S=	5 S=
1 C	2 C	3 C	4 C	5 C
1 C=	2 C=	3 C=	4 C=	5 C=
1 X	2 X	3 X	4 X	5 X
		3 END	4 END	5 END
				NO END

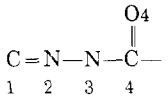
The "X" card can represent any other elements such as As (arsenic), B (boron), Cr (chromium), etc., when a part of the fragment chain. In case a single element to be designated by X is present, the card for this element is also punched. These cards will accommodate all combinations of elements up to five atoms in a chain plus their branchings. An "END" card is punched for the last atom of the chain. Should more than five atoms occur in a chain, the "NO END" card is punched, indicating a longer chain.

An order of preference has been established for coding these chains. If nitrogen is present, the preferred order is

	Example
1. $-N=N-$	$-\underset{1}{N}=\underset{2}{N}-\underset{3}{C}=\underset{4}{N}-$
2. $-N-N-$	$-\underset{1}{N}-\underset{2}{N}-\underset{3}{\overset{\overset{N3}{\|}}{C}}-\underset{4}{O}-$
3. $-N=C-$	$-\underset{1}{N}=\underset{2}{C}-\underset{3}{O}-\underset{4}{C}-\underset{5}{N}-$
4. $-N-C-$	$-\underset{1}{N}-\underset{2}{\overset{\overset{O2}{\|}}{C}}-\underset{3}{O}-$
5. $-C=N-$	$-\underset{1}{C}=\underset{2}{N}-\underset{3}{N}-\underset{4}{\overset{\overset{O4}{\|}}{C}}-\underset{5}{O}-$

After establishing ranking on the basis of nitrogen in the first two positions, further precedence is determined by oxygen, phosphorus, sulfur, and X elements. Ketone $>C=O$ and thione $>C=S$ are listed last in order of preference. Nitrogen is placed as close to the 1 position as possible. Therefore —C=N— (# 5) is listed with the nitrogen atoms. Over 300 different fragment chains have occurred in the first 5,000 organic compounds received this year.

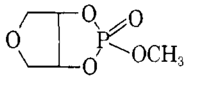
The following example illustrates the procedure for coding fragment chains. First each atom is numbered according to the established order.



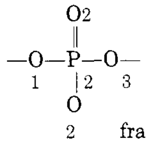
For position 1-punch # 1 C=
For position 2-punch # 2 N=
For position 3-punch # 3 N
For position 4-punch # 4 C=
For position 4-punch # 4 O=
For position 4-punch # 4 END

When double bond is indicated, the element cards for both positions concerned are punched indicating a double bond attachment.

A fragment chain may be wholly or partially included in a hetero cyclic nonaromatic ring, in which case it is coded both as a fragment chain and by the RRI #.



RRI # 0869



fragment

The ring is also coded with the following descriptors.

4 hetero
3 O cyclic
1 P cyclic

A Rolodex file card is maintained on each ring system and fragment chain for quick reference by the chemist making the structural breakdown. As new arrangements of the elements in the fragment chain are received, a Rolodex card is made and filed. This file can often supply answers to requests because only ring systems and fragment chains that have been entered into the system are recorded. If the card is not there, it is not necessary to search the coordinate index file for analogs.

the coordinate index file for analogs.

Supplier Codes.—Suppliers of the compounds screened by the CCNSC are assigned a numerical designation; for example, number 1 is the now defunct Chemical Biological Coordination Center. Number 2 is Sterling-Winthrop Research Institute, number 3 is Johns Hopkins School of Medicine, etc. As the personnel at the various companies and universities change, each new person receives a new letter designation which is placed after the number previously assigned.

It had been the practice to punch a plastic card for the individual as well as for the company or university so that a person could be distinguished from others at the same location. About 1,000 cards were required to record the suppliers and their representatives concerned with the first 54,000 compounds.

The method used for recording the RRI # has been adapted for use with the Supplier Code. At present, there are 650 numbers assigned to the suppliers plus the alphabetic designations for the individuals.

With the following plastic cards, all the Supplier Codes can be entered. The letter "P" is reserved for purchased compounds. The letters "I" and "O" are not used because they may be mistaken for numbers.

000	-00	--0	A through [✓ Z]
[✓ 100]	[✓ -10]	--1	
200	-20	--2	
300	-30	[✓ --3]	
400	-40	--4	
500	-50	--5	
600	-60	--6	
700	-70	--7	
800	-80	--8	
900	-90	--9	

By punching the three numerical cards of the Supplier Code, the company or university submitting the compounds will be recorded. The fourth card to be punched is the alphabetical designation for the individual chemist. Cards are marked in red to avoid confusion with Ring Index numbers.

To retrieve the NSC numbers of compounds from a supplier, one has only to reassemble the code number, for example

Supplier Code Number 113-Z
Ohio State University
Dr. Robert S. Brown

Number cards 100, -10, and --3 will show all compounds submitted through the university. With the addition of the alphabetic "Z" card, only those submitted by Dr. Brown will be listed.

With this system, compounds from any one of more than a thousand suppliers can be retrieved with the use of the appropriate four of the fifty-four plastic Supplier Code Cards.

SUMMARY

Two main divisions of organic structures—ring systems and the fragment chains—have been coded in a novel way that permits the recall of these structures with exactness without having a separate plastic card for each ring or fragment chain. Additional new ring systems and fragment chains can be recorded on the basic set of plastic cards so that the number of cards in the file will remain approximately the same. These improvements not only have reduced the file size from over 1,500 to less than 500, but also have resulted in an increased depth of entry into the structure. These improvements may be integrated with existing electronic data processing equipment should anyone wish to adapt them to his particular use.

Acknowledgment.—The author wishes to acknowledge the contributions of Mrs. Evelyn Wall, Mr. Frank Sordyl, and the staff of the Scientific Records Section.

REFERENCES

(1) D. F. Gamble, "A Coordinate Index of Organic Compounds," Presented before the Division of Chemical Literature, 127th National Meeting of the American Chemical Society, Cincinnati, Ohio, March 21, 1955.

(2) W. A. Wildhack, and J. Stern, "Peek-A-Boo System," in "Punched Cards—Their Applications to Science and Industry," R. S. Casey, J. W. Perry, A. Kent, and M. M. Berry, Ed., Reinhold Publishing Corp., New York, N. Y., 1958, Chapter 6.

(3) A. M. Patterson, L. T. Capell, and D. F. Walker, "The Ring Index," 2d Ed., American Chemical Society, Washington, D. C., 1960.

(4) "The Naming and Indexing of Chemical Compounds by *Chemical Abstracts*," Introduction to the Subject Index of Vol. 56, (Jan.-June, 1962) of *Chemical Abstracts*.

AUTHOR INDEX
VOLUME 4

Adams, R. M. 95
Adomaitis, V. 244
Alverson, R. A. 204
Anzalone, A. M. 169
Atherton, P. 157
Batchelor, H. W. 181
Bernier, C. L. 99, 104
Billman, R. C. 218
Caponio, J. F. 5
Chou, D. Y. P. 247
Clifford, A. F. 91
Clow, A. 221
Cohn, G. 169
Cortelyou, E. 141
Costello, Jr., J. C. 12, 26, 116
Crane, E. J. 64
Dunn, P. V. 218
Eisman, S. H. 186
Fernelius, W. C. 70
Freeman, R. R. 107
Frome, J. 33, 43
Garfield, E. 49
Gillum, T. L. 5, 29
Godfrey, J. T. 107
Gordon, M. 147
Granito, C. E. 56
Haefele, C. R. 112
Hamm, P. C. 218
Hannah, L. H. 218
Harris, T. H. 165, 229
Hirayama, K. 9
Hiz, H. 173
Hoskins, W. M. 167
Howerton, P. W. 232
Ihndris, R. W. 274
Kip, C. E. 234
Lane, J. C. 126
Leum, L. N. 256
Loening, K. L. 66
MacDougall, D. 166
Magee, R. J. 164, 195
Maizell, R. E. 107
Malkiewich, E. J. 211
Maloney, C. J. 181
McCandless, R. F. J. 147
McIlvain, J. M. 256
Mellon, M. G. 1
Miner, W. N. 20
Montague, B. A. 251, 258
O'Conner, J. 49
O'Day, P. T. 33
Ofer, K. D. 54
Jahoda, G. 153
Jones, B. A. 131
Payne, N. M. 197
Reyling, P. M. 144
Rice, C. N. 107
Scott, J. D. 63, 66
Shepherd, W. H. 107
Sher, I. H. 49
Skolnik, H. 168, 221
Skweir, E. A. 147
Sorter, P. F. 56
Southern, W. A. 237
Spencer, E. Y. 228
Spialter, L. 261, 269
Tinker, J. F. 112
Van Wazer, J. R. 84
Waldo, W. H. 218
Walker, F. 45
Wall, E. 26
Weil, B. H. 125
Wieman, J. A. 209
Yovich, J. C. 157

SUBJECT INDEX
VOLUME 4

Abstract Journal 147
Abstract Journal, Study 157
Abstract Style 9
Abstracting 26
ASTIA 5, 29
Book Reviews
The Application of Automata Theory to Problems in Information Retrieval 194
Classics in the Theory of Chemical Combination 194
Concepts of Repography. Terms and Definitions in German, English, and French 194
Crystallographic Data on Metal and Alloy Structures (2)ii
Data and Preceptions to the Early History of Documentation 193
Documentation Systems: A Structural Outline 194
Drugs in Our Society 193
Experimental Spectroscopy (2)ii
Formaldehyde 193
Mathematical Crystallography and Theory of Group Movements (2)ii
Methods of Forensic Science 194
New Methods to Pest Control and Eradication 194
New Reference Tools for Librarians 193
Organic Syntheses 194
Research Guide. A Guide to Contract Research Facilities in the Netherlands (2)ii
A Scheme for Classification of Information on Nutrition Coded for Machine Retrieval 194
Scientific and Technical Libraries. Their Organization and Administration 193
Selected Bibliography on Indexing in Science and Technology: Theory, Application and Techniques 194
Survey of Chemical Notation System 193
Charter for Information System 12
Chemical Graph 173, 186, 261, 269
Chemical Titles 107
Classification System 181
Code, Patent Office 29
Code, Pesticide 211
Code, Structural 49, 112, 211
Communication 125
Communication, Pesticidal Data 218
DDC 5
Duplicating Methods 144
EJC Course 26
Fog Index 126
Index, Analytical Chemistry 258
Index, Coordinate 26, 45, 116, 251, 256, 258, 274
Index, Correlative 99, 104
Index, Formula 49
Index, Graphical 173, 186, 261, 269
Index, Key Word 54, 107
Index, Optical Coincidence 274
Index, Ordnance 169
Index, Organic 29, 43, 112
Index, Permuted 56
Index, Ring Compounds 274
Information Service, Use 153
Information System Charter 12
Literature, Biochemical 147
Literature, Chemical 147
Literature, China 247
Literature, Electrochemistry 247
Literature, Japan 247
Literature, Lithuania 244
Literature, Pesticide 164, 165, 166, 167, 168, 195, 197, 204, 209, 211, 218, 221, 228, 229, 232, 237
Literature, Plutonium 20
Matrix—see Chemical Graph
Military Science 181
Nomenclature 64, 66
Nomenclature, Boron 95
Nomenclature, Coordination Compounds 70
Nomenclature, Halogen 91
Nomenclature, Inorganic 63
Nomenclature, Phosphorus 84
Notation, Hercules 221
Notation, Polish 186
Notation, Wiswesser, Permuted 56
Oral Presentation 131
Patent Office Code 29
Pesticide Chemical Notation 221
Pesticide Guide 228
Pesticide Patents 232
Pesticide Registration 229
Pesticide Regulation 165
Pharmaceutical Market Data 237
Phosphorus Index 29
Physics Abstracts, Study 157
Plutonium Literature 20
Retrieval System 112
Roles, EJC 116
Rotadex 49
Selective Dissemination 54, 107
Singularity Sub-Link 45
Steroid Index 29
Technical Writing 1
Thesaurus 12
Thesaurus, ASTIA 5, 29
Typist, Education in Chemistry 234
Writing 1, 125, 126
Writing, Report 141