```
      ∇LTRMAXCODE[□]∇
     ∇ CODE←A LTRMAXCODE B;M;K;M1;GOSU;X;HN
[1]    ΔS← 1 0 ρCODE←⍳M←K←0
[2]    AN←1↑ρB
[3]    AD←⌈/(,C←B×⍳0)/,B
[4]    HN←PARTITION1 B
[5]    NORMALISATION HN
[6]    B←,B
[7]  L1:→(ΔN<M←M+1)/EXIT
[8]    M1←ΔK[M]
[9]    →(K=M1)/NORM
[10]   GOSU←1↑ρΔS
[11]   ΔR←(GOSU,ΔCCM1])ρΔDEM1]←⍳ΔCCM1]
[12] NORM:GROW
[13]   K←M1
[14]   →(M=1)/L1
[15]   GOSU←1↑ρΔS
[16]   X←(((M-1),GOSU)ρ((ΔSC;M]-1)×ΔN))+⍳(0,¯1)↓ΔS
[17]   V←V=⌈/V←(ΔR+1)↓B[X]
[18]   ΔS←V/[1] ΔS
[19]   ΔR←V/[1] ΔR
[20]   CODE←CODE,B[X[;V⍳1]]
[21]   →L1
[22] EXIT:ΔS←(¢ΔC←ΔSC1;])[ΔS]
      ∇
```

Figure 5. APL function for unique coding and symmetry perception.

step 2 of the basic algorithm. The program GROW (see Figure 4) "grows" the partially constructed permutation vectors. PARTITION1, NORMALIZATION, and GROW are used by LTRMAXCODE (see Figure 5) which provides the maximum lower triangular code t(g^*,**B**), normalized atomic vector a ($=a^*$), and bond-atom symmetry $G^{(B^*)} \cap G^{(a^*)}$. The function SETSELECT used by NORMALIZATION is given in the preceding paper.[6]

The principal arrays used in the programs are described in Table IV.

REFERENCES AND NOTES

(1) B. Baumslag and B. Chandler, "Group Theory", McGraw-Hill, New York, 1968.
(2) W. T. Wipke and T. M. Dyott, *J. Am. Chem. Soc.*, **96**, 4834 (1974).
(3) C. Jochum and J. Gasteiger, *J. Chem. Inf. Comput. Sci.*, **17**, 113 (1977).
(4) M. Uchino, *Kagakusosetsu*, **18**, 47 (1978).
(5) J. E. Rush, *J. Chem. Inf. Comput. Sci.*, **16**, 202 (1976).
(6) M. Uchino, preceding paper in this issue.
(7) Y. Kudo and S. Sasaki, *J. Chem. Doc.*, **14**, 200 (1974).
(8) A. Nijenhuis and H. S. Wilf, "Combinatorial Algorithms", Academic Press, New York, 1975.
(9) A. L. Mackay, *J. Theor. Biol.*, **54**, 399 (1975); *J. Chem. Phys.*, **62**, 309 (1975).
(10) For more advanced algorithm, see following paper in this issue.

Algorithms for Unique and Unambiguous Coding and Symmetry Perception of Molecular Structure Diagram. III. Method of Subregion Analysis for Unique Coding and Symmetry Perception†

MASAHIRO UCHINO

Laboratory of Resources Utilisation, Tokyo Institute of Technology, 4259 Nagatsudacho, Midoriku, Yokohama, 227 Japan

Received June 4, 1979; Revised Manuscript Received January 21, 1980

A new efficient method to provide unique codes and symmetry groups is proposed. In this method, a normalized bond matrix (or adjacency matrix) is separated into several subregions, and each subregion is analyzed to avoid unnecessary generation of the permutation vectors for determination of unique codes and symmetry groups. The number of permutation vectors to be examined can be greatly reduced by this analysis of subregions and is usually smaller than (or equal to) the order of symmetry group.

INTRODUCTION

In the basic algorithm[1] to determine the right coset of the symmetry group $G^{(B)} \cap G^{(a)}$ in $G^{(s)}$, we "grow" the permutation vectors from the first interval of the permutation to the last interval (i.e., from left to right), testing at each stage to see if the partially constructed vectors can be extended to the vectors which are members of the right coset $(G^{(B)} \cap G^{(a)})g^*$. In this paper, we describe a more advanced technique to determine the right coset satisfying the criterion of "the maximum lower triangular code".[1]

In the new method, we "grow" the segments of the permutation vectors in the intervals of permutation in a certain optimal way and not mechanically from left to right. The decision for the interval in which we "grow" the segments of the permutation vectors can be made by separating the permuted bond matrix into several subregions. And for this purpose, we introduce in section I two vectors and two matrixes as functions of partially constructed vector and interval indexes. The analysis of the subregions by these functions provides six possible cases, and for each case, the action to be taken is described in section II. In section III, examples of the proposed method are presented.

I. FUNCTIONS OF PARTIALLY CONSTRUCTED PERMUTATION VECTOR

In the basic algorithm,[1] the partially constructed vector at the mth stage consists of the first m components of the permutation vector. In this paper, the partially constructed vector $\boldsymbol{g}$ consists of the constructed segments $\boldsymbol{g}(1)$, $\boldsymbol{g}(2)$, . . ., $\boldsymbol{g}(i)$ of lengths p_1, p_2, . . ., p_i which are associated with the permutation intervals of lengths c_1, c_2, . . ., c_i. When the number of permutation intervals is q, we define q-element vector $\boldsymbol{d}$ such that[1]

$$d_1 = 0, \qquad d_j = \sum_{i=1}^{j-1} c_i, \qquad j = 2, 3, \ldots, q$$

and vector $\boldsymbol{u}^j$ for $j = 1, 2, \ldots, q$ as

$$u^j{}_i = d_j + i, \qquad i = 1, 2, \ldots, c_j$$

and let $r(\boldsymbol{g}(j))$ denote the vector obtained from $\boldsymbol{u}^j$, suppressing the elements of $\boldsymbol{g}(j)$.

† Presented at the Division of Chemical Information, ACS/CSJ Congress, Honolulu, Hawaii, April 3, 1979.

0095-2338/80/1620-0124$01.00/0

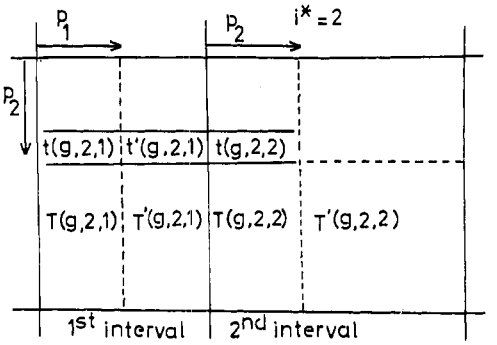

Figure 1. Subregions of bond matrix permuted by $\boldsymbol{h} = (\boldsymbol{g}(1), r(\boldsymbol{g}(1)), \boldsymbol{g}(2), r(\boldsymbol{g}(2)), \ldots)$. The functions $t(\boldsymbol{g},i,j)$, $t'(\boldsymbol{g},i,j)$, $T(\boldsymbol{g},i,j)$, and $T'(\boldsymbol{g},i,j)$ are used to describe each subregion.

We define, for a normalized[1] bond matrix **B**, the vectors $t(\boldsymbol{g},i,j)$, $t'(\boldsymbol{g},i,j)$, and the matrices $T(\boldsymbol{g},i,j)$, $T'(\boldsymbol{g},i,j)$, as the functions of partially constructed vector $\boldsymbol{g}$ and the interval indexes i and j

$$t(\boldsymbol{g},i,j) = (\mathbf{B}_{k\mathbf{g}(j)_1}, B_{k\mathbf{g}(j)_2}, \ldots, B_{k\mathbf{g}(j)_{p_j}})$$

where $i \leq j$ and $k = \boldsymbol{g}(i)_{p_i}$,

$$t'(\boldsymbol{g},i,j) = (\mathbf{B}_{kr(\mathbf{g}(j))_1}, \mathbf{B}_{kr(\mathbf{g}(j))_2}, \ldots, \mathbf{B}_{kr(\mathbf{g}(j))_{c_j-p_j}})$$

where $i < j$ and $k = \boldsymbol{g}(i)_{p_i}$,

$$T(\boldsymbol{g},i,j)_{mh} = \mathbf{B}_{r(\mathbf{g}(i))_m \mathbf{g}(j)_h}$$

$$m = 1, 2, \ldots, c_i - p_i;\ h = 1, 2, \ldots, p_j$$

$$T'(\boldsymbol{g},i,j)_{mh} = \mathbf{B}_{r(\mathbf{g}(i))_m r(\mathbf{g}(j))_h}$$

$$m = 1, 2, \ldots, c_i - p_i;\ h = 1, 2, \ldots, c_j - p_j$$

where $i \leq j$.

The subregions of the permuted bond matrix defined by these vectors and matrices are illustrated in Figure 1.

II. CONTROLLED "GROWTH" OF PARTIALLY CONSTRUCTED PERMUTATION VECTOR

The conditions determining whether we "grow" the partial vectors are given by the functions defined in section I. We describe the conditions for any member $\boldsymbol{g}$ and $\boldsymbol{g}'$ of a set S' of partially constructed vectors and for a pair of the interval indexes i and j

Condition I-a. $t(\boldsymbol{g},i,j) = t(\boldsymbol{g}',i,j)$.

Condition I-b. $t'(\boldsymbol{g},i,j) = t'(\boldsymbol{g}', i,j)$ and $t'(\boldsymbol{g},i,j)$ consists of single values.

Condition II-a. $T(\boldsymbol{g},i,j) = T(\boldsymbol{g}',i,j)$, and $T(\boldsymbol{g},i,j)$ consists of single row vectors.

Condition II-b. $T'(\boldsymbol{g},i,j) = T'(\boldsymbol{g}',i,j)$, and $T'(\boldsymbol{g},i,j)$ consists of single values.

Condition II-c. $T'(\boldsymbol{g},i,i) = T'(\boldsymbol{g}',i,i)$, and $T'(\boldsymbol{g},i,i)$ consists of single values except the principal diagonal elements.

If the set S' of the partially constructed vectors is consistent with the criteria of "the maximum lower triangular code"[1] and the conditions (I-a, I-b, II-a, II-b, and II-c) are satisfied or undefinable for all possible pairs of the interval indexes, then we do not need further "growth" of the partial vectors to determine a unique code and symmetry group. The unique bond matrix $\mathbf{B}^*$ is given by $\mathbf{B}^*_{ij} = B_{\boldsymbol{h}_i \boldsymbol{h}_j}$, $i, j = 1, 2, \ldots, n$, where vector $\boldsymbol{h}$ is constructed from $g \in S'$ as $\boldsymbol{h} = (\boldsymbol{g}(1), r(\boldsymbol{g}(1)), \boldsymbol{g}(2), r(\boldsymbol{g}(2)), \ldots, \boldsymbol{g}(q), r(\boldsymbol{g}(q)))$. Considering each $\boldsymbol{g} \in S'$ for permutation vector $\boldsymbol{h}$, we construct a set S'' of the n-element permutation vectors. The set S'' is a subset of the set S defined in the previous paper[1] and set S is obtained from set S'' by the full permutation of the components of each $r(\boldsymbol{g}(i))$ in each vector $\boldsymbol{h} \in S''$. If $\boldsymbol{g}^*$ is a member of S'', $\boldsymbol{g}^*$ is also a member of S, and the bond-atom symmetry $G^{(\mathbf{B}^*)} \cap G^{(\mathbf{a}^*)}$ is given by[1]

$$G^{(\mathbf{B}^*)} \cap G^{(\mathbf{a}^*)} = \boldsymbol{g}^{*-1} S$$

where $\boldsymbol{g}^* \in S''$. Thus, the set $\boldsymbol{g}^{*-1}S''$ provides a representation for the bond-atom symmetry, and this symmetry group is generated easily from the set $\boldsymbol{g}^{*-1}S''$.

The special case mentioned above is the goal of the present algorithm, and we proceed in the order of the pairs of the interval indexes, (1,1), (2,1), (2,2), (3,1), (3,2), (3,3), . . ., $(q,1)$, $(q,2)$, . . ., (q,q), testing at each step to see if the conditions (I-a, I-b, II-a, II-b, and II-c) hold and "growing" the segments of partially constructed vectors. In this way, the lower triangular part of the unique bond matrix $\mathbf{B}^*$ is determined from the upper row to the lower row. During the procedure, the following six cases are recognized with optimal action taken in each case.

We assume in the following that we are processing the pairs (i,j) of the indexes such that $i = i^*$ and $j = 1, 2, \ldots, i^*$, and that j^* is the smallest integer j at which one or several of the conditions (I-a, I-b, II-a, II-b, and II-c) do not hold.

Case I. $j^* \leq i^*$ and the condition I-a does not hold at j^*.

Case II. $j^* < i^*$ and the condition I-a holds but the condition I-b does not hold at j^*.

Case III. $j^* \leq i^*$ and the condition II-a does not hold at j^*.

Case IV. $j^* < i^*$ and the condition II-a holds but the condition II-b does not hold at j^*.

Case V. $j^* = i^*$ and the condition II-a holds but the condition II-c does not hold.

Case VI. j^* does not exist.

For cases I–VI, we describe the optimal actions to determine the unique code and symmetry group.

Action I for Case I. Remove from the set S' of the partially constructed vectors, the vector $\boldsymbol{g}$ such that

$$t(\boldsymbol{g},i^*,j^*) \neq \ell\text{-}\max_{\boldsymbol{g}' \in S'} t(\boldsymbol{g}',i^*,j^*)$$

Action II for Case II. Determine the maximum component value of $t'(\boldsymbol{g},i^*,j^*)$ for all $\boldsymbol{g}$ in S' and denote it by m^*. Let $t^*(\boldsymbol{g})$ be the number of times m^* appears in $t'(\boldsymbol{g},i^*,j^*)$ and let t^{**} be $t^{**} = \max_{\boldsymbol{g} \in S'} t^*(\boldsymbol{g})$. Remove from S' the vector $\boldsymbol{g}$ such that $t^*(\boldsymbol{g}) \neq t^{**}$.

For any $\boldsymbol{g}$ in new S' and for any position k at which m^* appears in $t'(\boldsymbol{g},i^*,j^*)$, "grow" the segment $\boldsymbol{g}(j^*)$ to the vector $\boldsymbol{g}(j^*)_1, \boldsymbol{g}(j^*)_2, \ldots, \boldsymbol{g}(j^*)_{p_{j^*}}, r(\boldsymbol{g}(j^*))_k$, and increment p_{j^*} by one.

Action III for Case III. Determine the lexicographically maximum row vector of $T(\boldsymbol{g},i^*,j^*)$ for all $\boldsymbol{g}$ in S' and denote it by $\boldsymbol{r}^*$. Remove from S' the vector $\boldsymbol{g}$ such that $T(\boldsymbol{g},i^*,j^*)$ does not contain $\boldsymbol{r}^*$ as row vector. For any $\boldsymbol{g}$ in new S' and for any row k at which $\boldsymbol{r}^*$ appears in $T(\boldsymbol{g},i^*,j^*)$, "grow" the segment $\boldsymbol{g}(i^*)$ to $\boldsymbol{g}(i^*)_1, \boldsymbol{g}(i^*)_2, \ldots, \boldsymbol{g}(i^*)_{p_{i^*}}, r(\boldsymbol{g}(i^*))_k$, and increment $\boldsymbol{p}_{i^*}$ by one.

Action IV for Case IV. Determine the maximum component value m^* of $T'(\boldsymbol{g},i^*,j^*)$ for all $\boldsymbol{g}$ in S', and let $t^*(\boldsymbol{g})$ be the maximum number of times m^* appears in any row vector of $T'(\boldsymbol{g},i^*,j^*)$. Let t^{**} be $t^{**} = \max_{\boldsymbol{g} \in S'} t^*(\boldsymbol{g})$. Remove from S' the vector $\boldsymbol{g}$ such that $t^*(\boldsymbol{g}) \neq t^{**}$. For any $\boldsymbol{g}$ in new S' and for any row k at which m^* appears t^{**} times in the row vector of $T'(\boldsymbol{g},i^*,j^*)$, "grow" the segment $\boldsymbol{g}(i^*)$ to $\boldsymbol{g}(i^*)_1$, $\boldsymbol{g}(i^*)_2, \ldots, \boldsymbol{g}(i^*)_{p_{i^*}}, r(\boldsymbol{g}(i^*))_k$, and increment p_{i^*} by one.

Action V for Case V. "Grow" the segment $\boldsymbol{g}(i^*)$ of each $\boldsymbol{g}$ in S' to $\boldsymbol{g}(i^*)_1, \boldsymbol{g}(i^*)_2, \ldots, \boldsymbol{g}(i^*)_{p_{i^*}}, r(\boldsymbol{g}(i^*))_k$, where position k is any position of the vector $r(\boldsymbol{g}(i^*))$, and increment $\boldsymbol{p}_{i^*}$ by one.

Action VI for Case VI. Increment i^* by one, and if i^* exceeds the number of intervals q, stop the process. Otherwise, repeat the process starting from the check of the conditions

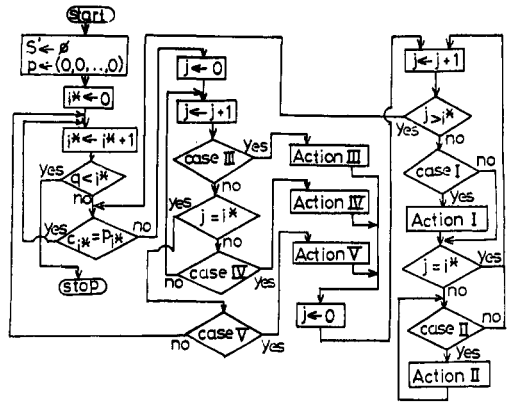

Figure 2. Flowchart of the procedure to determine the set S' of the partially constructed vectors.

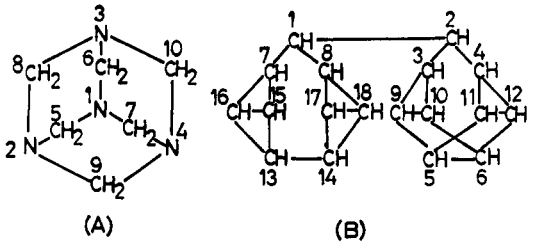

Figure 3. Examples of unique numbering determined by the method of subregion analysis. For partitioning of graph nodes, the function PARTITION1[2] was used and the numberings shown were determined by the criterion[1] of "maximum lower triangular code".

(I-a, I-b, II-a, II-b, and II-c) for new (i,j) pairs such that $i = i^*$ and $j = 1, 2, \ldots, i^*$.

In the description of actions I–V, the "grown" segment $\boldsymbol{g}(i^*)_1, \boldsymbol{g}(i^*)_2, \ldots, \boldsymbol{g}(i^*)_{p_{i^*}}$, $r(\boldsymbol{g}(i^*))_k$, or $\boldsymbol{g}(j^*)_1, \boldsymbol{g}(j^*)_2, \ldots, \boldsymbol{g}(j^*)_{p_{j^*}}$, $r(\boldsymbol{g}(j^*))_k$ must be understood as $r(\boldsymbol{g}(i^*))_k$ or $r(\boldsymbol{g}(j^*))_k$ if the segment $\boldsymbol{g}(i^*)$ or $\boldsymbol{g}(j^*)$ is empty (i.e., $p_{i^*} = 0$ or $p_{j^*} = 0$). Set S' of the partially constructed vectors is initialized by the empty set and the vector $\boldsymbol{p}$ which provides the lengths of the "grown" segments is initialized by $(0, 0, \ldots, 0)$ when we start the procedure.

The flow chart of the proposed procedure is shown in Figure 2; cases I–V and actions I–V in the flow chart correspond to those described in this section.

III. EXAMPLES OF UNIQUE NUMBERING AND SYMMETRY PERCEPTION

The normalized bond matrixes[1] obtained by the use of method I of the preceding paper[2] were processed by the algorithm of section II. The same unique bond matrix and the associated symmetry group as those of the previous paper[1] can be obtained with less computation. The number of the permutation vectors examined during the procedure has been greatly reduced in the present approach and has usually been smaller than (or equal to) the order of the symmetry group. For structure A of Figure 3, a normalized bond matrix **B** and the vector **c** are given by eq 1. If we employ the basic algorithm, we must generate at first 4! partial vectors without pruning. But in the present approach, we first generate the

$$\mathbf{B} = \begin{matrix} 0\,0\,0\,0\,1\,1\,1\,0\,0\,0 \\ 0\,0\,0\,0\,0\,0\,1\,1\,0\,1 \\ 0\,0\,0\,0\,1\,0\,0\,1\,1\,0 \\ 0\,0\,0\,0\,0\,1\,0\,0\,1\,1 \\ 1\,0\,1\,0\,0\,0\,0\,0\,0\,0 \\ 1\,0\,0\,1\,0\,0\,0\,0\,0\,0 \\ 1\,1\,0\,0\,0\,0\,0\,0\,0\,0 \\ 0\,1\,1\,0\,0\,0\,0\,0\,0\,0 \\ 0\,0\,1\,1\,0\,0\,0\,0\,0\,0 \\ 0\,1\,0\,1\,0\,0\,0\,0\,0\,0 \end{matrix} \quad \boldsymbol{c} = (4,6) \qquad (1)$$

segments of the second interval as in (I), and we "grow" the

First interval	Second interval
....	5.....
....	6.....
....	7.....
....	8.....
....	9.....
....	10.....

(I)

partial vectors to (II). In this way, we continue to "grow"

First interval	Second interval
1 3..	5.....
3 1..	5.....
1 4..	6.....
4 1..	6.....
1 2..	7.....
2 1..	7.....
2 3..	8.....
3 2..	8.....
3 4..	9.....
4 3..	9.....
2 4..	10.....
4 2..	10.....

(II)

the partial vectors, and the vector $\boldsymbol{p}$ changes during the procedure as follows: $\boldsymbol{p} = (0,0) \rightarrow \boldsymbol{p} = (0,1) \rightarrow \boldsymbol{p} = (1,1) \rightarrow \boldsymbol{p} = (2,1) \rightarrow \boldsymbol{p} = (2,2) \rightarrow \boldsymbol{p} = (3,2) \rightarrow \boldsymbol{p} = (3,3) \rightarrow \boldsymbol{p} = (4,3) \rightarrow \boldsymbol{p} = (4,3) \rightarrow \boldsymbol{p} = (4,4) \rightarrow \boldsymbol{p} = (4,5)$. The maximum number of partially constructed permutation vectors appeared in this example was equal to the order of the symmetry group (i.e., 24 compared with 144 for the basic algorithm[1]).

For structure B of Figure 3, the nodes of the graph can be partitioned into eight classes and vector $\boldsymbol{c}$ given by $\boldsymbol{c} = (1,1,2,2,2,4,2,4)$. The maximum number of the partial vectors appeared during the procedure was half the order of symmetry groups and the "growth" of the segments of the permutation intervals proceeded as follows: $\boldsymbol{p} = (0,0,0,0,0,1,0,0) \rightarrow \boldsymbol{p} = (0,0,1,0,0,1,0,0) \rightarrow \boldsymbol{p} = (0,0,1,1,0,1,0,0) \rightarrow \boldsymbol{p} = (0,0,1,1,1,1,0,0) \rightarrow \boldsymbol{p} = (0,0,1,1,1,2,0,0) \rightarrow \boldsymbol{p} = (0,0,2,1,1,2,0,0) \rightarrow \boldsymbol{p} = (0,0,2,2,1,2,0,0) \rightarrow \boldsymbol{p} = (0,0,2,2,2,2,0,0) \rightarrow \boldsymbol{p} = (0,0,2,2,2,2,3,0,0) \rightarrow \boldsymbol{p} = (0,0,2,2,2,3,0,1) \rightarrow \boldsymbol{p} = (0,0,2,2,2,4,0,1) \rightarrow \boldsymbol{p} = (0,0,2,2,2,4,1,1) \rightarrow \boldsymbol{p} = (0,0,2,2,2,4,1,2) \rightarrow \boldsymbol{p} = (0,0,2,2,2,4,2,2)$. The final set S' of the partially constructed permutation vectors in this example is given by:

. . 3 4 5 6 7 8 9 10 11 12 13 14 15 16 . .

. . 3 4 5 6 7 8 9 10 11 12 13 14 16 15 . .

. . 3 4 5 6 8 7 9 10 11 12 14 13 17 18 . .

. . 3 4 5 6 8 7 9 10 11 12 14 13 18 17 . .

. . 3 4 6 5 7 8 10 9 12 11 13 14 15 16 . .

. . 3 4 6 5 7 8 10 9 12 11 13 14 16 15 . .

. . 3 4 6 5 8 7 10 9 12 11 14 13 17 18 . .

. . 3 4 6 5 8 7 10 9 12 11 14 13 18 17 . .

. . 4 3 5 6 7 8 11 12 9 10 13 14 15 16 . .

. . 4 3 5 6 7 8 11 12 9 10 13 14 16 15 . .

. . 4 3 5 6 8 7 11 12 9 10 14 13 17 18 . .

. . 4 3 5 6 8 7 11 12 9 10 14 13 18 17 . .

. . 4 3 6 5 7 8 12 11 10 9 13 14 15 16 . .

. . 4 3 6 5 7 8 12 11 10 9 13 14 16 15 . .

. . 4 3 6 5 8 7 12 11 10 9 14 13 17 18 . .

. . 4 3 6 5 8 7 12 11 10 9 14 13 18 17 . .

From the set S' thus computed, we can obtain easily the symmetry group $G^{(\mathbf{B}^*)} \cap G^{(\mathbf{a}^*)}$ as described in section I. The unique bond matrix $\mathbf{B}^*$ can be determined by an element of

the set S' and the unique numberings thus obtained are shown in Figure 3.

ACKNOWLEDGMENT

The author wishes to thank R. Kojima and S. Takagi of Nihon Parkerizing Co. for various kinds of assistance during the preparation of the manuscript.

REFERENCES AND NOTES

(1) M. Uchino, "Basic Algorithm for Unique Coding and Computation of Symmetry Group", preceding paper in this issue.
(2) M. Uchino, "Vector Functions for Automorphism Partitioning", accompanying paper in this issue.

NEWS AND NOTES

MEETINGS

Gordon Research Conference on Scientific Information

The theme of this Conference, July 14–18, 1980, will be information systems of the '80s to support decision making on critical scientific issues.

14 July. "Biogeochemical problems requiring rapid decisions based on relevant information and data"; "Technological problems of the '80s"; "Information/data systems of the '80s".

15 July. "The ozone layer and atmospheric problems—data requirements"; "Information and data problems in determining ozone depletion rate"; "Information and data needs for environmental regulatory decisions in the '80s"; "Information systems for toxicity data".

16 July. "The search for alternate sources of energy"; "Information and data problems and tradeoffs in assessing alternate energy sources"; "Catastrophic engineering failures—causes and problems"; "Responding to industry's need for materials information"; "Materials data-needs and opportunities".

17 July. "Risk evaluation for potential carcinogens—problems that relate to having minimal data"; "Methodology for estimating direct exposure to new chemical substances"; "Technology and scientific responsibility".

Health Sciences Communications Association Meeting

HeSCA is holding its 22nd Annual Convention in Calgary, Canada, June 14–18, 1980. The Convention theme is "Images for Health". For further information contact: Stuart Winn, Coordinator, Education Services, Faculty of Health Sciences, McMaster University, 1200 Main St. West, Hamilton, Ontario, Canada L8S 4J9.

NEWS ITEMS

CAS Registry Backdating

Chemical Abstracts Service has received a grant of $137 590 from the Council on Environmental Quality, the Environmental Protection Agency, and the National Library of Medicine to study the feasibility of adding chemical substances indexed in *Chemical Abstracts* prior to 1965 to the computer files of the CAS Chemical Registry System.

CAS has registered all substances indexed in *Chemical Abstracts* since January 1965 along with substances from a number of special data collections—about 4.8 million substances in all. Many substances, possibly numbering in the millions, were indexed prior to 1965 and have not been mentioned in the literature since. Registering the substances indexed in earlier *Chemical Abstracts* indexes would enhance the Registry's usefulness by extending its coverage to the great majority of known chemical substances and by linking papers and patents cited in *Chemical Abstracts* before 1965 to the Registry System.

Under the grant, CAS will explore methods of registering the content of the pre-1965 indexes and the costs and potential problems associated with them. Various means of converting the content of the printed indexes to computer-readable form, including optical character recognition, will be evaluated. CAS also will attempt to determine whether computer programs it has developed for translating systematic chemical names into the connection table representations of chemical structures used in the Registry files can be adapted to translate the names in pre-1965 indexes.

CAS Registry Numbers have come into wide use to identify chemical substances in publications and computer files. EPA, NLM, the National Institute of Occupational Safety and Health, and the National Cancer Institute use the numbers to identify substances in computer files, as do a number of European organizations and government agencies and several large industrial and commercial information retrieval systems. CEQ recommended in a 1978 report to Congress that U.S. government agencies use the numbers as the standard means of identification for chemicals and mixtures in their files.

CA on Microform

Chemical Abstracts Service is now offering subscribers the option of receiving *Chemical Abstracts* on microfilm or microfiche without purchasing the printed form of CA.

In the past, only organizations that maintained a current subscription to the printed *Chemical Abstracts* could obtain the microform versions. Under a new policy, subscribers may opt to receive only the microform or microfiche versions of CA weekly issues in combination with semiannual indexes in printed form for $6500 per year. The fee includes a license to make single copies of abstracts for on-site use. Subscribers to current issues in microform also can obtain back issues of *Chemical Abstracts* since 1907 and the CA Collective Indexes on 16-mm microfilm or microfiche by paying a one-time fee.

A new microform licensing agreement guarantees subscribers that they will be able to keep the microform issues and indexes as long as they require them. Should the subscription to current microform issues not be renewed, the subscriber will be able to retain back issues and indexes by paying a nominal annual right-to-copy fee. The subscriber also has the option of purchasing back issues and indexes should CAS stop producing a microform version of *Chemical Abstracts*.

0095-2338/80/1620-0127$01.00/0