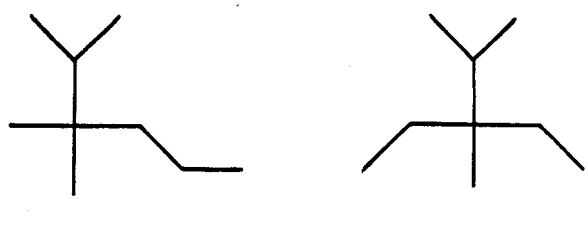

Figure 1. A pair of nonane trees with identical TI's. Beneath each tree its N-tuple code is also given.

(altogether 469 766 alkanes) in our search for nonisomorphic alkane trees with identical TI's. Our target was achieved in the family of alkane trees with nine vertices. A pair of nonanes (2-methyl-2-isopropylpentane and 3-methyl-3-isopropylpentane) possesses the identical value of TI (17 584). These nonane trees, which represent the smallest pair of alkane trees with degenerate TI values, are given in Figure 1. The nonane trees in the figure are also identified by their N-tuple codes[2] because we used these codes to generate the trees. Thus, the figure contains answers to both questions asked previously. Or, in other words, the TI is not an unique descriptor for characterizing alkanes, and the smallest pair of nonisomorphic alkane trees is diagrammed in Figure 1.

There is another interesting result in the paper by Schultz et al.[1] worth commenting on. They have noticed in their numerical analysis that the distance matrix determinants, det|**D**|, for each family of isomeric alkane trees have the same values. However, it can be proved[4] that the determinant of the distance matrix for all (alkane) trees with the same number of vertices N is given by

$$\det|\mathbf{D}| = -(-2)^{(N-2)}(N-1)$$

The above formula also immediately explains the observation of the authors[1] that det|**D**| alternates from negative values for all alkane isomers with *even* carbon content to positive values for all isomers of *odd* carbon atom content. The above formula allows quick calculations of det|**D**|. Thus, one can immediately obtain the value of det|**D**| (114 688) for 2,6-dimethyl-5-ethyl-3-isopropyloctane. Schultz et al.[1] obtained a slightly different value (114 687) for this structure. They suspected that their value is not correct because of the value of det|**D**| (114 688) for isomeric structures, pentadecane and 2,3-dimethyl-6-ethyl-5-isopropyloctane, and they stated so.

ACKNOWLEDGMENT

We thank the referees for their comments.

REFERENCES AND NOTES

(1) Schultz, H. P.; Schultz, E. B.; Schultz, T. P. Topological Organic Chemistry. 2. Graph Theory, Matrix Determinants and Eigenvalues, and Topological Indices of Alkanes. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 27-29.
(2) Knop, J. V.; Müller, W. R.; Jeričević, Ž.; Trinajstić, N. Computer Enumeration and Generation of Trees and Rooted Trees. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 91-99.
(3) Müller, W. R.; Szymanski, K.; Knop, J. V.; Trinajstić, N. An Algorithm for Construction of the Molecular Distance Matrix. *J. Comput. Chem.* **1987**, *8*, 170-173.
(4) Graham, R. L.; Pollak, H. O. On the Addressing Problem for Loop Switching. *Bell Syst. Tech. J.* **1971**, *50*, 2495-2519.

Online Search Strategies for Semiconductor or Superconductor Materials

HOWARD M. DESS

Rutgers University, Library of Science and Medicine, Piscataway, New Jersey 08855-1029

Received June 4, 1990

Recent enhancements introduced by CAS ONLINE and by INSPEC provide powerful new search capabilities for online retrieval of information about nonstoichiometric compositions such as the multitude of fractional atom compounds reported in the literature that deals with, for example, semiconductors or superconductors. In CAS ONLINE, use of the "dot connected atom sequence" permits retrieval of entire families of materials containing specified atoms. In INSPEC, use of the CI field serves a similar function. Illustrative examples of productive search strategies that use these new tools are presented along with a discussion of their limits of applicability.

INTRODUCTION

First-time online searches for semiconductor or superconductor materials can pose unexpected challenges even to experienced practitioners, who are highly skilled in many other areas of chemical searching. These problems arise from a combination of circumstances:

(1) Nonstoichiometric compositions are very prevalent among these families of compounds. The following fractional atom materials are illustrative:

$Al_{0.26}Ga_{0.74}As$
$Al_{0.8-0.85}Ga_{0.15-0.2}As$
$Ba_{0.13}Cu_{0.25}Y_{0.07}O_{0.55}$

(2) Molecular formula searches and index name searches for these types of compositions in the Registry File of CAS ONLINE or in INSPEC will yield misleadingly small numbers of hits unless special search strategies are utilized that are not required with other classes of materials.

(3) Both in CAS ONLINE and INSPEC the searcher faces the added complication of working with search strategies that may be applicable only for specific time periods: In both systems the end of 1986 is the dividing line, and distinctly different search procedures are applicable before and after this date. This situation reflects the introduction of helpful new indexing modifications and innovations in these two databases around that time.

(4) Documentation about these problems and applicable search techniques is surprisingly limited. INSPEC issued a series of descriptive pamphlets introducing and explaining the use of chemical indexing (CI) in various online systems such as DIALOG, STN, ORBIT, etc.[1] and also published a superconductor search guide[2] which, however, contained only a brief reference about chemical indexing as applied to search

0095-2338/91/1631-0084$02.50/0 © 1991 American Chemical Society

strategies for this class of materials. STN International's instructional literature has not provided any guidance to the types of search techniques discussed in later sections: Information about these matters was gained via discussions with STN help desk personnel.[4] Finally, broad-ranging searches of the literature dealing with online search topics has turned up only one article that discusses retrieval strategies for superconducting oxides.[3]

The principle objectives of this study are to point out the pitfalls that may be encountered during the course of online searches for information about semiconductor or superconductor materials in CAS ONLINE or INSPEC (which cover the bulk of the world output of scientific literature dealing with these subject areas) and to demonstrate the applicability of search strategies of proven effectiveness that are readily adaptable to cover the broadest conceivable range of compositions for these classes of compounds and for nonstoichiometric materials in general. As will be demonstrated in later sections, because CAS ONLINE and INSPEC differ from each other in so many important respects, in terms of indexing conventions, methods of organization, searchable fields, etc., search strategies that are successful in one cannot be bodily transferred for use in the other. Instead, separate and distinct search strategies must be developed for each one individually.

SEARCH STRATEGIES FOR CAS ONLINE

Use of "Dot Connected Atom" Searches in the Registry File. For comprehensive retrieval of all CA-indexed compositions comprising any given set of atoms such as semiconductor materials of the type $Al_xGa_{1-x}As$ or superconducting compounds such as $Ba_wCu_xY_yO_z$, the simplest and fastest retrieval strategy is the following:

--Use the "dot connected atom" format in the /MF index of the Registry File.
---List elements in alphabetic order (in the absence of carbon and hydrogen).
---Not valid for searches earlier than CA Vol. 106, 1987.

The following example is illustrative:

```
===> file reg
 S     Ba.Cu.O.Y/MF
 L1        824  Ba.Cu.O.Y/MF

                                    Retrieved all reported
 S     Al.As.Ga/MF                  materials containing only
                                    the indicated atomic
 L2        274  Al.As.Ga/MF         constituents.
```

If the periods are omitted from the search formulations so that, in effect, a "normal" molecular formula search is conducted, the numbers of hits are vastly reduced, and in the current examples, are zero:

```
L3          0  BaCuOY/MF
L4          0  AlAsGa/MF
```

In these last two searches, the search format instructed the computer to find and retrieve the exact molecular formulas indicated (i.e., compositions containing equiatomic amounts of each element). In the dot connected atom formulation, the computer retrieved any/all compositions containing the specified atomic constituents regardless of atomic ratios. One final point to be noted is that the dot connected atom format may also retrieve some alloy materials. However, alloys, as a class, can be easily eliminated from answer sets, if desired, via use of the class identifier field (/CI):

```
L5    270  L2 not AYS/CI
```

Bibliographic Retrieval Strategies for Families of Compositions Obtained in the Registry File. Answer sets retrieved in the Registry File as described above must be crossed over to the CA File for recovery of bibliographic information. It is important to remember that the CA File search must be conducted not only in the basic index but also in the derivative field (/D), because CAS indexing policies treat solid solutions and oxygen deficient compounds as derivatives:

```
===>  File CA
===>  S  L1 or L1/D
            4980  L1
            3727  L1/D
 L6         8415  L1 or L1/D          (total Ba-Cu-Y-O hits)
```

Answer set L6 demonstrates the maximum number of bibliographic hits retrievable based solely on the family of compositions crossed over from the Registry File via answer set L1. More usually the searcher would add appropriate key words or qualifying phrases relevant to the specific search topic under investigation so that a smaller number of answers would ultimately be recovered.

Similarly for the system Al-As-Ga:

```
===>   S  L5 or L5/D
            6089  L5
               3  L5/D
 L7         6092  L5 or L5/D          (total Al-As-Ga hits)
```

Note that in the Ba-Cu-Y-O system over 44% of the total hits in answer set L6 were obtained from the derivative field.

Of the total hits recovered in answer set L6, 99% were for publications issued in the time span 1987–1990; in set L7, 92% of the answers fell in the same time period, and 99% had publication dates of 1986 or later.

Utilization of Truncation or Expansion in the /MF Index. Truncation may be used to broaden the range of compositions retrieved:

```
L8    398  Al.As.Ga?/MF                (in Reg. File)
L9    124  Al.As.Ga.?/MF
```

The difference in hits retrieved from these two search formulations is readily explained via an EXPAND of the dot connected atom formulation:

```
===>  e  Al.As.Ga/MF
E1          1        Al.As.Fe.Ho.O.Sb.Ti.V/MF
E2          1        Al.As.Fe.Ru.Si/MF _ _ _ _
E3        | 274 --->  Al.As.Ga/MF          \
E4        |  55       Al.As.Ga.In/MF   \     \
E5        |  16       Al.As.Ga.In.P/MF \      \
E6        |   3       Al.As.Ga.In.Sb/MF \     \
E7        |   2       Al.As.Ga.N/MF      \     \
E8        |   1       Al.As.Ga.O/MF       > 124   > 398
E9        |  14       Al.As.Ga.P/MF      /      /
E10       |   1       Al.As.Ga.P.Sb/MF  /      /
E11       |  32       Al.As.Ga.Sb/MF  _/ _ _ _/
E12         57        Al.As.In/MF
```

Answer set L8 retrieved all compositions containing the elements Al-As-Ga. By adding the period after Ga in search set L9, retrievals were restricted to those compositions that contained elements falling beyond gallium in the alphabet. To recover compositions with atomic constituents that lie between arsenic and gallium in the alphabet, internal truncation can be used:

```
L10   4  Al.As.!.Ga/MF or Al.As.!!.Ga/MF
```

The additional substances retrieved in answer set L10 included

compositions of aluminum boron gallium arsenide and aluminum cobalt gallium arsenide. Alternatively, the EXPAND function can be used and desired compositions selected by visual inspection.

Searches for Specific Nonstoichiometric Compositions. Suppose one wishes to find information about a single composition such as $Ba_{0.03}Cu_{0.18}Y_{0.21}O_{0.59}$. The cheapest and most effective way to find a single compound of this type is via an EXPAND (with elements listed in alphabetic order) in the /MF index of the Registry File:

```
===>  e  ba0.03cu0.18o0.59y0.21/mf

E1              .            .            .            .

E2              .            .            .            .

E3              1  ===>   Ba0.03Cu0.18O0.59Y0.21/MF

E4              .            .            .            .
```

As demonstrated in this example, it is not necessary to use first letter capitals when typing in the element symbols, or subscripts for the numerical portions. Simply enter the formula in linear fashion as indicated, with no spaces, but do include the zeros to the left of the decimal point.

Disadvantages of Chemical Name Searches vs Dot Connected Atom Searches in the Registry File. Chemical name searches in the Registry File are not recommended for the types of materials under discussion here because they are more cumbersome to enter and too dependent on nomenclatural conventions. For example, name searches for specific compositions in the system Ba-Cu-Y-O require an entry format of the type:

"barium copper yttrium oxide ($Ba_wCu_xY_yO_z$)"/CN

If any portion of this complex string is incorrectly entered or mistyped, the search will fail. An easier way to retrieve an individual nonstoichiometric compound is to use the EXPAND function in the /MF index as was shown under Utilization of Truncation or Expansion in the /MF Index.

Finally, one could use the following "trick" to retrieve an entire family of compositions in the Registry File:

```
===>  aluminum gallium arsenide "("?/CN

 L11     268
```

However, it is much simpler to use the dot connected atom strategy outlined under Use of "Dot Connected Atom" Searches in the Registry File.

Search Strategies in CAS ONLINE for the Period 1967–1987. Prior to 1987, most nonstoichiometric compositions were not assigned CA Registry Numbers. They were instead treated as solid solutions of the end members and were indexed in the print version of CA under whichever end-member compound occurred first in the alphabet. This meant that the identities of the multitude of solid solution compositions reported in the literature were submerged anonymously under the general indexing scheme shown below, and registry numbers were usually assigned only to the end members of the series:

```
Aluminum arsenide (AlAs)  [22831-42-1]
      .
      .
--Solid solns.
          .
          .
  ---With gallium arsenide
                .
                .
```

Thus, in 1986, the last year this indexing scheme was followed, Vol. 105 of the *Chemical Substance Index* contained six pages (approximately 1500 individual listings) of such AlAs–GaAs solid-solution entries and only two entries were listed under the separate index heading "aluminum gallium arsenide". In the following year, 1987, starting with Vol. 106, the *Chemical Substance Index* began indexing these semiconductor solid-solution compositions as though they were distinct compounds (nonstoichiometric), which meant that from that time forward each one of these compositions, or ranges of compositions, would be assigned its own unique registry number despite the fact that all members of the series were still labeled as aluminum gallium arsenide. Thus, CA Vol. 109, 1988, of the *Chemical Substance Index* contains over 10 pages of listings arranged in order of increasing aluminum content and declining gallium content, in the following format:

```
aluminum gallium arsenide (Al0.18Ga0.82As) [108158-92-5]
                .
                .
aluminum gallium arsenide (Al0.2-0.7Ga0.3-0.8As) [108915-41-9]
                .
                .
aluminum gallium arsenide ((Al,Ga)As) [37382-15-3]
```

This indexing scheme is analogous to the genus species hierarchical name assignment employed in the field of biology, where the chemical name is now the generic name for an entire family of substances whose individual members (or species) are distinguished from each other by molecular formula and registry number. The online search strategies described in the previous sections are based on this indexing arrangement. However, for publications that were issued prior to 1987, an entirely different search strategy must be used, one that is geared to the type of solid-solution indexing described above and with no recourse to registry number identification for individual nonstoichiometric compositions.

The case of aluminum gallium arsenide is representative. One begins in the Registry File by retrieving answer sets for the end members of the solid-solution series AlAs---GaAs:

```
===>  File Reg          In this example, one could have equally
===>  S  AlAs/MF        well searched in the chemical name index,
 L12     1  AlAs/MF     e.g., aluminum arsenide/CN, etc.,
===>  S  AsGa/MF        because the end members are
 L13     1  AsGa/MF     stoichiometric compositions.
```

Answer sets L12 and L13 are then crossed over to the CA File and searched as indicated below:

```
===>  File CA
===>  S  (L12 or L12/D) and (L13 or L13/D)
          1598  L12
           814  L12/D
         46452  L13
         13981  L13/D
 L14      9913  (L12 or L12/D) and (L13 or L13/D)
```

Use of the derivative field (/D) is essential here because entries on solid solutions are indexed as derivatives. Of the total hits retrieved, 92.1% were pre-1987.

The overlap in retrievals from the CA File answer sets L14 and L7 (the latter based on the dot connected atom strategy) was quite small: only 273 hits. Therefore, a comprehensive search for information about all possible compositions in the family Al-Ga-As over the full-time range covered by CAS ONLINE (1967 to date) must incorporate both strategies:

```
L15     15732  L7 or L14
```

Searches involving four constituents such as GaInAsP are more complex, but the same basic strategies are applicable.

The portion of the search covering the time period 1987 to date is quite straightforward:

```
===>  File Reg
 L16      275  As.Ga.In.P/MF
 L17      271  L16 not AYS/CI
===>  File CA
 L18     1840  L17 or L17/D        Total bibliographic hits retrieved
                                   for all compositions in the family
                                   As-Ga-In-P for the period 1987+.
```

The search covering the period 1967–1987 is initiated in the Registry File by retrieving answer sets for all compositions presumed to be end members of the solid-solution family, and continuing via crossover into the CA File:

```
===>  File Reg
 L19      1  AsGa/MF
 L20      1  GaP/MF
 L21      1  InP/MF
 L22      1  AsIn/MF
===>  File CA
 L23      6385  (L19 or L19/D) and (L21 or L21/D)
 L24      3905  (L20 or L20/D) and (L22 or L22/D)
 L25      6565  L23 or L24
```

As demonstrated by the merged set L25, over 97% of the final total were retrieved in set L23 (GaAs + InP). The remainder originated with answer set L24 (GaP + InAs). If one searches all four compositions simultaneously via an "and" combination, retrievals are substantially reduced:

```
L26      3725  (L19 or L19/D) and (L20 or L20/D) and (L21 or L21/D)
                                          and (L22 or L22/D)
```

A merge of set L26 with L25 shows that this last answer set is wholly contained in set L25. Therefore, when searching a four-atom composition such as GaInAsP, the binary end-member compositions should be searched in pairs in the CA File, and these answer sets "or"ed together.

Finally, to cover the full time range of CAS ONLINE (1967 to date), answer sets L18 and L25 are "or"ed, yielding a total recovery of 8065 hits.

Chemical Name Searches in CA File: Use With Caution. Chemical name searches in CA File are quick and easy to run, and may be applicable across the entire time span covered by the database. However, searchers should be fully aware of the problems associated with this approach:

- --Variability in nomenclature usage must be taken into account.
- --Chemical name searches in CA File will indiscriminately retrieve compositions that contain other elements in addition to the ones specified.

The following example is illustrative:

```
===>  File CA
 L27      7935  barium(w)copper(w)yttrium(w)oxide
 L28      5639  yttrium(w)barium(w)copper(w)oxide
 L29      8825  L27 or L28
```

Answer set L29 retrieved an additional 619 hits over the number captured in set L6 (retrievals in CA File derived from a dot connected atom search). A spot check of some of the answers contained in this incremental 619 hits revealed compositions such as the following:

- --dysprosium barium copper yttrium oxide
- --lanthanum yttrium barium copper oxide
- --silver barium copper yttrium oxide
- --barium copper yttrium oxide sulfide
- --etc., etc.

If the searcher was interested only in compositions containing the elements Ba-Cu-Y-O, then these additional hits would represent false drops.

SEARCH STRATEGIES FOR INSPEC

None of the strategies described thus far can be utilized with INSPEC. In particular, since INSPEC has not adopted the use of CAS Registry Numbers as indexing tags for chemical compositions, this convenient and unambiguous mode of searching is simply not applicable here. In all examples that follow, INSPEC searches were conducted via DIALOG.

Chemical Name Searches in INSPEC: An Ineffective Search Strategy. The following example is illustrative:

```
 S1      28  (aluminum or aluminium)()gallium()arsenide
 S2      40  gallium()(aluminum or aluminium)()arsenide
 S3       0  barium()copper()yttrium()oxide
 S4      27  yttrium()barium()copper()oxide
```

Chemical name searches for these materials recovered trivially small or zero hits. The following sections demonstrate that molecular formula searches or modified molecular formula searches are the most productive strategies to follow in INSPEC for the materials under consideration in this study.

Comprehensive Search Covering INSPEC File 13 (1977 to Date). The following steps summarize the strategy used to achieve the highest number of hits for the family of substances comprising all Al-Ga-As compositions:

```
 S5      2478  GaAlAs
 S6      4866  AlGaAs
 S7         6  GaAsAl
 S8       265  GaAl()As      Note:  Zero answer sets such as
 S9         5  Ga()AlAs             Ga()AsAl were omitted.
 S10      316  AlGa()As
 S11      207  Al()GaAs
 S12     1335  GaAs()Al
 S13     8927  (S5-S12)/OR
```

An even broader recovery of materials will be achieved via the following formulation:

```
 S14      5839  Ga(S)Al(S)As
Then:
 S15     12504  S13 or S14
```

However, inclusion of S14 in the overall search strategy will also increase the number of false drops since it provides only the loosest positional linkages among the three atomic constituents. Therefore, many of the compositions retrieved will include elements other than those belonging to the $Al_xGa_{1-x}As$ family of compounds.

INSPEC indexing protocols necessitate use of a rich menu of retrieval strategies for nonstoichiometric materials. The searcher must, therefore, remain flexible and open-minded about testing alternative approaches to ensure optimal recoveries of relevant hits. Moreover, search complexities mount rapidly as the number of elements increases: a four-element system such as the family of compositions comrising GaInAsP could conceivably require over 100 separate search statements for a comprehensive search covering all possible indexing variations.

INSPEC Search Strategies Recommended for the Period 1987 to Date. In recognition of the types of problems mentioned in the preceding sections, INSPEC, in February of 1987, introduced a new chemical indexing field (CI=) that simplifies the search for families of inorganic materials and, moreover, also provides for linkage to as many as seven different role

indicators that may be used to direct a search into specific subject areas. For example, silicon can be searched in the following ways, depending on the type of information needed:

```
S       CI=Si  (find everything on silicon; no role indicator)
S       CI=(Si el)  (find information only on elemental silicon)
S       CI=(Si bin)  (find information on Si as part of binary
                      systems such as SiO2 or SiC, etc.)
S       CI=(Si ss)  (find information on Si as part of systems
                     containing three or more elements)
S       CI=(Si dop)  (find information on Si used as a dopant)
                                                    etc., etc.
```

Continuing with the previous search on the Al-Ga-As family of compounds:

```
S16     3622  CI=AlGaAs
S17     1135  CI=GaAlAs          (zero answer sets not shown)
S18     4749  S16 or S17
```

Role indicators were omitted from the search above in order to illustrate retrieval of the maximum number of hits relevant to the indicated compositions. The nonequivalence of answer sets S16 and S17 demonstrates that one must still pay attention to the order in which the elements are listed when searching the CI field. However, in the strategy shown below, the results are independent of the order in which the elements are listed:

```
S19     5537  CI=(Al ss(S)Ga ss(S)As ss)
```

Moreover, answer set S19 incorporates all of the answers recovered in answer set S18. A sampling of the additional 788 hits retrieved in set S19 (vs S18) indicates citations that deal with materials systems containing more than the three elements Al-Ga-As. This finding conforms to the definition of the "ss" role indicator as applicable to systems containing three *or more* elements.

Similarly, for the four-element system Ba-Y-Cu-O:

```
S20     8526  CI=(Ba ss(S)Y ss(S)Cu ss(S)O ss)
```

The following test was run on the relevance of the answers retrieved in answer set S20 to the topic of superconductivity:

```
S21     8503  S20 and supercond?
```

A relevance factor of 99.73% is quite impressive.

The query shown below was entered for comparison purposes with the results obtained in answer set S20:

```
S22     3352  Ba(S)Y(S)Cu(S)O
S23     8799  (S20 or S22) and supercond?
```

Incorporation of answer set S22 into the overall search strategy netted an additional 296 hits relevant to superconductivity in systems containing Ba-Y-Cu-O (plus other elements). This example further emphasizes the advisability of testing alternative strategies when dealing with multielement systems in INSPEC.

Narrower Searches for Specific Compositions or Roles. The preceding sections have focused on comprehensive searches for information about families of substances comprised of specific elements with very broad compositional variability. However, many times a search will have more limited objectives: a specific composition will be sought or possibly a relatively narrow range of compositions. Moreover, the searcher may wish to restrict the search to specific roles or functions for these compositions, e.g., utilizing one or more of the role indicators described earlier. In all such cases, by far the easiest, fastest, and cheapest way to conduct such a search is via the EXPAND function. For example, suppose one is seeking information on interfacial systems involving YBaCuO:

```
E  CI=YBaCuO
Ref      Items   Index-term
E1           5   CI=YBaCuNiO
E2           5   CI=YBaCuNiO ss
E3        1950  *CI=YBaCuO
E4           1   CI=YBaCuO bin
E5         179   CI=YBaCuO int
E6        1949   CI=YBaCuO ss
E7          41   CI=YBaCuO sur
E8           7   CI=YBaCuO-Ag
E9           7   CI=YBaCuO-Ag int
E10          1   CI=YBaCuO-Ag-PbO-Pb
E11          1   CI=YBaCuO-Ag-PbO-Pb int
```

Inspection of the entries called up in this fashion reveals a number of possible choices for selection. Remember, also, to check alternative formulations for the atom string that you are searching:

```
E  CI=BaYCuO
Ref      Items   Index-term
E1           3   CI=BaYCuNbO
E2           3   CI=BaYCuNbO ss
E3          77  *CI=BaYCuO
E4           5   CI=BaYCuO int
E5          77   CI=BaYCuO ss
E6           1   CI=BaYCuO-BaYCuO
E7           1   CI=BaYCuO-BaYCuO int
```

Specific nonstoichiometric materials with fractional atom compositions can also be found via the EXPAND file:

```
E  CI=Ga0.47In
Ref      Items   Index-term
E1           1   CI=Ga0.47As bin
E2           1   CI=Ga0.47As int
E3           0  *CI=Ga0.47In
E4           1   CI=Ga0.47In0.5As
E5           1   CI=Ga0.47In0.5As ss
E6           1   CI=Ga0.47In0.5As:Be
```

In conducting these expansions, the searcher must keep in mind that the CI field is relevant only for the period 1987 to date. For pre-1987 publications, the following format can be used to initiate an EXPAND in a compositional region of interest:

```
E  Al/sub 0.2    (where "sub" denotes "subscript")
Ref      Items   Index-term
E1           1   Al/sub 0.18/Ga/sub 0.82/As/GaAs
E2           1   Al/sub 0.19/Ga/sub 0.82/As-GaAs
E3           0  *Al/sub 0.2
E4           8   Al/sub 0.2/Ga/sub 0.8/As
E5           1   Al/sub 0.2/Ga/sub 0.8/As cell
E6           1   Al/sub 0.2/Ga/sub 0.8/As epitaxial layers
E7           1   Al/sub 0.2/Ga/sub 0.8/As P/sup +/- N junction
E8           2   Al/sub 0.2/Ga/sub 0.8/As solar cells
E9           1   Al/sub 0.2/Ga/sub 0.8/As-Al/sub 0.4/Ga/sub 0.6
```

This procedure is by far the preferred way to locate fractional atom compositions in INSPEC and select them for retrieval. It avoids the necessity to search directly for complex expressions that may contain multiple occurrences of the term "sub": each search involving this term requires lengthy processing times because it is so highly posted.

SUMMARY

Chemical substances classified as nonstoichiometric or solid solutions confront online searchers with special problems not usually encountered with other types of materials. Search strategies for these substances must follow certain prescribed procedures or risk gross failure to capture relevant information. The searcher must also keep in mind that certain strategies will be valid only over specific time spans: both for CAS ONLINE and INSPEC this strategic dividing line occurs at the end of 1986.

In CAS ONLINE the preferred search strategy for the period 1987 to date tests on the use of the dot connected atom formulation in the /MF index of Registry File. For earlier periods in CAS ONLINE one must utilize a strategy that simultaneously searches for the co-occurrence of the end members of a solid-solution family.

Comparable searches in INSPEC over the period 1987 to date should employ the chemical index file (CI=). Searches for earlier information must use a series of molecular formula (or atom string) searches that may be quite lengthy depending on the compositional complexity of the system being searched. Even for the period 1987+ the searcher is advised to test the completeness of retrievals obtained via use of the CI file by searching some alternative query formulations.

Finally, both in CAS ONLINE and INSPEC, searches for specific compositions or narrow compositional ranges are best initiated via use of the EXPAND file. It is far easier (and cheaper) to locate and select a given item from an EXPAND list than to search long, complex expressions that are subject to a variety of entry errors, especially mistakes in typing.

ACKNOWLEDGMENT

I acknowledge the advice and insights provided by Ken Cada of Chemical Abstracts Service and Jim Ashling of INSPEC.

REFERENCES AND NOTES

(1) *Chemical & Numerical Indexing, Dialog Implementation*; INSPEC, The Institution of Electrical Engineers, Dec 1987; 12 pp.
(2) *Search Guide to Superconductors in the INSPEC Database*; INSPEC, The Institution of Electrical Engineers, Feb 1988, p 5.
(3) Coghlan, Jill. Superconducting oxides. *Database* **1989**, *12*, 122-127.
(4) Note Added in Proof: The May/June 1990 issue of *CAS ONLINE News* offered a discussion about searches for nonstoichiometric compounds in the Registry File.

Benzenoid Series Having a Constant Number of Isomers. 3. Total Resonant Sextet Benzenoids and Their Topological Characteristics

JERRY RAY DIAS

Department of Chemistry, University of Missouri, Kansas City, Missouri 64110-2499

Received July 5, 1990

Two distinct classes of benzenoid groups having a constant number of isomers have been identified for the total resonant sextet isomer subset. One class is topologically unique and the other forms a pairwise topologically equivalent class. The numbers of isomers in these two classes are the same as those reported in Part 2 of this series for strictly peri-condensed benzenoids. This correspondence clearly signals the discovery of a fundamental topological paradigm.

INTRODUCTION

Total resonant sextet benzenoid hydrocarbons have molecular formulas whose carbon numbers are divisible by 6, and they represent a subset of the most stable isomers.[1] It is shown in this paper that unlike circulenes—benzenoids with holes—sextet benzenoids have subsets which exhibit constant-isomer series.[2,3] It also will be shown that the number of sextet isomers and the topological features of the sextet benzenoid members of these constant-sextet-isomer series match one-for-one with the properties[3] of the strictly peri-condensed constant-isomer benzenoids. This spectacular correspondence clearly points out a fundamental topological paradigm.

While compounds in the one-sextet-isomer series have only one total resonant sextet benzenoid isomer each, there are other benzenoid structures associated with these formulas.[4,5] There are for example, 186 other benzenoid structures with the molecular formula $C_{42}H_{18}$ besides the one shown in Figure 1. Similarly, in addition to those in Figure 1, there are 5725 isomers with the molecular formula $C_{60}H_{22}$ and 123 789 with the molecular formula $C_{84}H_{26}$. This work has allowed the identification of the most stable isomers from large numbers of possible benzenoids. It is conjectured that when these constant-isomer benzenoids are formed under conditions of pyrolysis, they are more stable and both topologically and kinetically more likely to persist. Unlike many other benzenoids, total resonant sextet benzenoid isomers generally do not give color reactions with concentrated sulfuric acid and they tend to have the least intensely colored crystals. Chemically, total resonant sextets represent a particularly important benzenoid isomer subset.[1,6]

In a seminal paper,[7] Stein reported a theoretical thermodynamic study of the chemical equilibria in two model systems—highly condensed to strictly peri-condensed benzenoids and total resonant sextet benzenoids. A major conclusion of this study was that for a given carbon number, the most condensed benzenoid will have the greatest thermodynamic stability. Also, the higher the temperature of formation, the larger the benzenoids tend to be. Benzenoids which are smaller than this critical size will tend to degrade, while those which are larger will tend to grow. This is because there exists a critical species, whose size increases with temperature.

In a study of the hydroprocessing of high-boiling gas oil (vacuum gas oil from the *residuum* desulfurization process), it was observed[8] that hydrocracking of high-severity product was more difficult than for low-severity product because the former contained more highly condensed benzenoids, which belong to the constant-isomer series.

A recent mass spectral study[9] of fuel feedstocks tabulated suggested aromatic compounds without taking into consideration the relative stability of the various isomers corresponding to a specific molecular formula. For example, the observed molecular ion of m/z 378 ($C_{30}H_{18}$) was assigned the heptacene

0095-2338/91/1631-0089$02.50/0 © 1991 American Chemical Society