A Database and Retrieval System for the NBS Tables of Chemical Thermodynamic Properties

PEIMING WANG

Institute of Chemical Metallurgy, Academia Sinica, P.O. Box 353, Beijing, People's Republic of China

DAVID B. NEUMANN*

Electrolyte Data Center, National Institute of Standards and Technology, Gaithersburg, Maryland 20899

The organization and structure of a database for the *NBS Tables of Chemical Thermodynamic Properties—Selected Values for Inorganic and C_1 and C_2 Organic Substances in SI Units* is explained. The use of a commercial database management system for the storage and retrieval system is discussed. The file constructs necessary to provide rapid retrieval by elemental composition in this type of database management system are shown. Implementations both on a minicomputer and on personal computers are discussed.

INTRODUCTION

As a result of the efforts of the Chemical Thermodynamic Data Center (CTDC) at NIST,[1] the thermodynamic properties of a large number of chemical substances have been critically evaluated. The recommended values for the properties of 14 348 species have been published as a series of *NBS Technical Notes*[2] and recently as a combined volume titled *The NBS Tables of Chemical Thermodynamic Properties*.[3] These properties are the enthalpy of formation, $\Delta_f H°$, Gibbs energy of formation, $\Delta_f G°$, entropy, $S°$, and heat capacity at constant pressure, $C_p°$, all at 298.15 K, and the enthalpy difference, $[H°(298) - H°(0)]$. The values of these properties provide the basic quantities needed for the calculation or estimation of equilibrium constants and enthalpies of reaction for chemical processes and for the prediction of the thermochemical stabilities of materials.

During the past several years considerable interest has been expressed in having these fundamental chemical thermodynamic data made available as a computer-searchable database. Such a format on both minicomputers and personal computers would permit either occasional or systematic data retrieval. It would also permit programmatic access for a variety of applications such as data correlation and estimation as well as the prediction of chemical equilibria. It is believed that in such a form the *NBS Tables* can become a more convenient tool for individuals, be they scientists, engineers, or students, engaged in the study of chemistry, chemical engineering, material science, metallurgy, geochemistry, or other thermodynamically related scientific research areas.

This paper presents the results of our efforts to develop two versions of such a database for a Hewlett-Packard minicomputer and for IBM-compatible personal computers (PCs). The database in both cases is called NBSTAB and was developed by use of readily available commercial database management systems (DBMSs). In the sections that follow the minicomputer version is described in detail; as the two versions are similar, the final sections only describe the modifications that were needed to implement the system on PCs.

A DBMS called IMAGE/1000[4] was used to develop the database on the Hewlett-Packard Model HP-1000 Series A900 minicomputer.[5] A "key-indexed file management system" or DBMS known as BTRIEVE[6] was chosen for the personal computer version. The choice of the DBMS was based on several factors in addition to the usual cost and availability. They each had to provide the following: (a) rapid and flexible

access to the data in multiple files; (b) multi-key access to records; (c) support for duplicate keys and for different key types; (e) support for both fixed- and variable-length records; (f) a programmer's language interface for several programming languages (allowing individual users to write programs making use of the database for their own calculations, e.g., for simultaneous chemical equilibria); (g) a simple free-standing query system, if possible, a standard search query language (SQL) package to permit ad hoc queries and reports to be produced with a minimum of user effort.

Several additional factors and constraints were considered in moving the database to personal computers. First, as the archial version of the database is intended to remain on the minicomputer, the transfer to the PC, which would have to be repeated from time to time, should be as convenient a process as possible. Second, the personal computer based DBMS should be similar to the one on the minicomputer so that a minimum of reorganization of programs and data files would be needed. Finally, the personal computer DBMS had to permit its incorporation into executable search and retrieval programs.

The two database management systems, IMAGE/1000 and BTRIEVE, were analogous in the logical structure of their files; although their access modes are quite different, both were capable of managing keyed file access from programs written in several languages. Both have independent SQL packages permitting the user independent access to the data without computer programs or languages.

THE MINICOMPUTER DATABASE

The basic concepts and definitions of a DBMS, in this specific case, the IMAGE/1000 DBMS, have been discussed and reviewed in a number of publications.[5,7] In summary, the IMAGE/1000 system is a multifile two-level hierarchial DBMS. It consists of utility programs to create the database, to batch-load data, to perform backups and maintenance, to perform ad hoc data retrievals and update records, and to interface with several programming languages and a system to monitor usage and maintain integrity of the database in a multiuser environment. To discuss the organization of the NBSTAB database, some further information on the IMAGE/1000 file structure is given below.

Master and Detail Data Sets. The IMAGE/1000 database management system uses two types of files, master files and detail files.[5] Both are direct access files, i.e., files whose records are accessible by record number. The basic difference between the detail and master files is the mode of determining the

number of the record to be retrieved or stored.

Entries in a master data set are retrieved by key value, where a "hashing" algorithm is used to determine the record number directly from the key item's value. In a master data set the entries can have only one key item field. Multiple records with the same key item value are not possible.

Records in detail data sets may have several key items or fields. In addition, multiple records with the same value for a key item are permitted. A master data set must exist for each key item in a detail data set. The key item in the master points to the corresponding key item in the detail data set. To accomplish this, each master record contains a pointer to that detail data set record containing the first occurrence of the value of the key item corresponding to that in the master record. This establishes a chain head or the beginning of the chain of records all containing the same value for the key item. The pointers in the detail record give the record number of the next occurrence of that key item value in the detail set. Detail data set pointers are used to read in a forward or backward direction along the chain. This sequence continues until a zero pointer is encountered, which points to the end of the chain. This is shown schematically in Figure 1.

The above is a simplified view of these two types of IMAGE/1000 files, and the reader is referred to reference 5 for further information. Other modes of access based on pointers embedded by the DBMS are available. Naturally, sequential access to the detail and master data sets is also possible. Rapid access to the master and detail records based on key values is primarily accomplished via keyed access to the master data sets and via chained reads on the detail data sets.

File Categories. Three categories of files are defined in the NBSTAB database. They are defined by their use, interrelationship, and content rather than their structure. These are index files, data files, and comment files. As stated above, all IMAGE/1000 files are referred to as either master or detail data sets. The index, data, and comment files are all detail data sets.

(1) Index Files. The index files are designed to assist in the retrieval of data by chemical formula as specified through application programs by user requests. An index file contains only chemical formulas and those attributes of the formula such as its elemental composition, molecular weight, number of waters of hydration, and charge number that are derivable from the formula itself. These attributes are related purely to the content of a chemical formula and do not involve the physical state of the substance. The index files are linked to the data files through the chemical formulas. Multiple occurrences of the chemical formula will appear in the entries, i.e., records, of the data files. Each of these occurrences in the data file will be associated with a different physical state. The formula and the state together define an individual "species". The design of the index files is based on the requirements of the data retrieval algorithms discussed later in this paper.

The index files in the NBSTAB database are NBSID1 and NBSID2. They are dissimilar in structure as they are required for different retrieval techniques. Figures 2 and 3 indicate their structures.

(2) Data Files. The data files are logically connected to the index files via the chemical formula key. They contain the values of the chemical thermodynamic properties for the substances that include enthalpy of formation, Gibbs energy of formation, entropy, heat capacity at constant pressure, and the enthalpy difference, i.e., $\Delta_f H°(0)$, $\Delta_f H°(298)$, $\Delta_f G°(298)$, $S°(298)$, $C_p°(298)$ and $[H°(298) - H°(0)]$, respectively, for each substance. In addition to the property values, the data files store the number of significant decimal places in each value. Most importantly, the physical state of each substance, indicating the conditions at which the property values apply,
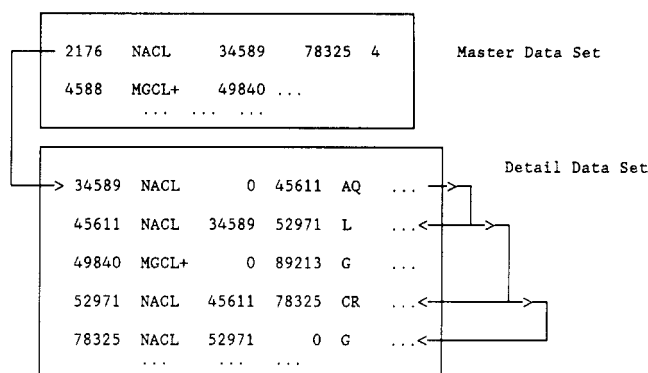

Figure 1. Schematic diagram of master and detail data sets.

FORML	ELEMS	CHAR#	COEFS	MASS	HYDR#
NA2S1O4	NAS O	0	2,1,4	142.0412	0
ZN1I+	ZNI	1	1,1	192.2744	0
S-2	S	-2	1	32.0640	0

Figure 2. Structure of the records in NBSID1.

FORML	ELEMT	COEF
NA2S1O4	NA	2
NA2S1O4	S	1
NA2S1O4	O	4
ZN1I+	ZN	1
ZN1I+	I	1
S-2	S	1

Figure 3. Structure of the records in NBSID2.

is also contained in the data files.

There are three data files in the NBSTAB database: NBSDA1, NBSDA2, and NBSDA3. NBSDA1 contains the above information primarily for pure substances, aqueous molecular species, and aqueous and gaseous ions. Dilution data for aqueous solutions are stored in data file NBSDA2. Data for nonaqueous and mixed solutions are contained in NBSDA3. Storing different types of data in separate data files provides additional flexibility and organization needed for systematic data retrieval. Figure 4 shows an example of one of these data files, NBSDA1.

(3) Comment Files. A comment file, NBSCOM, contains the supplemental descriptive information for some substances such as the name of the substance or information about configuration of the substance, e.g., α, β, cis, trans. For fully ionized aqueous species NBSCOM contains the ions from which the data for the species were measured or calculated, unless they are obvious.

Keys. Several key items are defined in the database to allow different modes of access to the substances:

ACCE# is the accession number, 1–14348, assigned to each species in the NBS tables database. It reflects the order

ACCE#	FORML	CASRN	CASFM	STATE	SCODE	NCOMM	DECIML	HO H298 G298 H298-H S298 CP
234	H1CL1O4			L	5.0	0	020000	-40.58
235	H1CL1O4			AI	10.0	0	022010	,-129.33,-8.52, ,182.0,
5659	ZN1CL2			CR	2.0	0	323322	
5660	ZN1CL2			G	7.0	0	010000	
5661	ZN1CL2			AI	10.0	0	022019	
5662	ZN1CL2			AO	11.0	0	001000	

Figure 4. Structure of the records in NBSDA1.

of the data as published in reference 3. Substances in different physical states and different conditions are considered as different species and are counted separately.

FORML is the chemical formula of a substance. Only one chemical formula is stored for each substance, although it may appear several times in conjunction with different physical states, conditions, or isomeric forms. There are 7972 distinct formulas.

CASRN and CASFM are the Chemical Abstracts Service Registry number and formula, the latter in CAS format. These key items may be useful to (a) merge the NIST-evaluated thermodynamic data with existing on-line databases and (b) obtain newly published chemical information for the corresponding substance from on-line services. The fields are not yet in use.

ELEMT is one of the elements from which a substance is composed. The possible values of ELEMT include all 103 elements in the periodic table and two isotopes of hydrogen, deuterium (D) and tritium (T). Each element has a corresponding "finding number", which follows the sequence used for the thermodynamic standard order of arrangement given in reference 3. This key item is exclusively used in the index file NBSID2.

ELEMS contains all component elements of a formula. It is designed for a specific data retrieval method discussed later and is used only in the index file NBSID1. The elements in ELEMS for each substance are arranged in order of decreasing finding number from left to right. The element of lowest finding number is rightmost. For example, the component elements in $Fe(HSO_4)_3$ are Fe, H, S, and O. Their finding numbers are 41, 2, 14, and 1, respectively. Therefore, the value of ELEMS for this substance is "FES H O".

CHAR is the signed charge number of the substance.

STATE is the physical state of the substance. The values of STATE include "CR" for crystalline solid, "L" for liquid, "G" for gas, "AM" for amorphous solid, "VIT" for vitreous or glassy, "AQ" for aqueous unspecified concentration, "AI" for aqueous fully ionized standard state, and "AO" for aqueous un-ionized standard state data. Isomers are denoted by appending a numeral to the physical state designation, e.g., CR2, CR3, G2, G3.

Eight master data sets are required for the above eight keys. These masters are associated with six detail data sets, namely, two index files, three numerical data files, and the comment file. Table I gives a summary of the files, keys, and their relationships. Figure 5 shows the general logical structure of the database. Further specifications of the fields and data sets are given in Appendix 1.

General Aspects of Data Format. The decision on the format in which the data are stored in the database is important in maintaining the integrity of the data. It can also allow easier search and retrieval programming. In the NBSTAB database numerical data are stored as ASCII character strings rather than binary numbers. Numerical values are thus made available as text and must be reinterpreted as real or integer numbers when used by application programs. This method

Table I. Summary of the Database Structure

| data set name	related key items	no. of entries
Master Data Sets		
NBSRN	ACCE#	14348
NBSF	FORML	7972
CAS1	CASRN	a
CAS2	CASFM	a
NBSEL	ELEMT	105
NBSES	ELEMS	3002
CHARGE	CHAR#	22
NBSTAT	STATE	25
Detail Data Sets		
NBSID1	FORML, ELEMS, CHAR#	7972
NBSID2	FORML, ELEMT	23485
NBSDA1	FORML, ACCE#, CASRN, CASFM, STATE	9643
NBSDA2	FORML, ACCE#, CASRN, CASFM	3904
NBSDA3	FORML, ACCE#, CASRN, CASFM	801
NBSCOM	ACCE#	1740

a Unused at present.
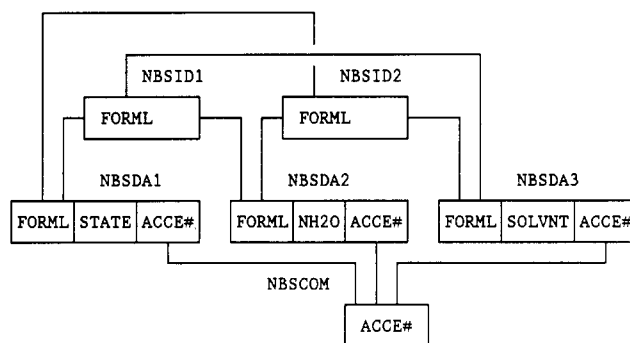
Figure 5. NBSTAB overview with some key and non-key fields shown.

of storage and transfer is at the cost of considerable extra disk space but is worthwhile due to the greater flexibility and integrity it offers.

Data Retrieval Techniques. The database is accessible by using (a) vendor-supplied general search, retrieval, and report-writing packages, (b) specialized FORTRAN 77 search and retrieval programs to be described below, or (c) user-written FORTRAN 77 programs making calls to a library of search and retrieval subroutines designed for this particular database. QUERY/1000 and ASK/1000 are vendor-supplied search, retrieval, and report-writing packages for use with IMAGE/1000 databases.[5,8] They are useful for highly specialized purposes as well as for the maintenance of the database, but they are not designed for the special problems encountered with chemical formulas.

Due to this limitation, search and retrieval programs were written in FORTRAN 77 and use calls to the IMAGE/1000 database access subroutines. The programs and their retrieval techniques are described below.

In addition, and in the process of writing the above programs, a library of subroutines for search and retrieval was constructed. These subroutines in turn call the IMAGE/1000 subroutines. This library is provided with the minicomputer database for those users needing to write specialized application programs that require the thermodynamic property values in the database. This important feature allows for the future development of thermodynamic programs for the calculation of equilibrium constants, enthalpies of reaction, heats of dilution, electrochemical potentials, etc.

Retrieval Programs. The simplest mode of retrieval of chemical thermodynamic data for substances is by their chemical formulas or, in the case of this database, by their accession numbers. However, keyed access on formulas or accession numbers causes problems. When searching with an

accession number, one must know the number of the substance in advance; and when searching using formulas, one must input the chemical formulas exactly as they are stored in the database. The latter is highly prone to error due to the many ways a chemical formula may be written. It is also inflexible, as one is frequently interested in searches for a class of substances, e.g., all substances containing a specific set of elements such as those in the Fe–O–S system.

The NBS tables database was structured to permit four retrieval options. Three of these can be used to search efficiently for the properties of a number of related substances simultaneously, while a fourth option permits retrieval of data for individual species by specifying their precise chemical formula and state. The first three techniques are all based on the situation where the user specifies a set of elements to be contained in the substances. For most chemists, searches are aimed at a specific system in which the component elements are known. An example is the study of galena. The primary component elements of galena are Pb, S, and O, which can form a variety of simple and complex compounds. Techniques for data retrieval by element sets are therefore fairly important and are discussed below.

All three retrieval modes use chained access to locate and read entries or records in detail data sets. As discussed above, a chain consists of all detail data set entries having the same value of a key item. When a request is made for all entries in which a key item takes on a particular value, the DBMS uses a keyed read on the master data set associated with the key item. That master record is then retrieved with its pointer to the entry in the detail data set that is the start of the chain. The DBMS does a direct read of that detail entry and at the same time finds within that record the pointer to the next detail entry containing the same key item value. That entry is retrieved directly. Again it has a pointer to the next entry, and so on, until the end of the chain is reached. The following discussion of the retrieval modes is based on the structure of the two index files, NBSID1 and NBSID2, shown in Figures 2 and 3.

(1) Retrieval Modes 1 and 2: Search by Restricted Set of Elements. Two of the three retrieval modes are designed to provide rapid searches for queries of the form (1) "FIND ALL substances CONTAINING: *ALL* of the elements in the set {Fe, O, S} *AND NO others*" or (2) "FIND ALL substances CONTAINING: *ANY* of the elements in the set {Fe, O, S} *AND NO others*".

The record structure of the index file NBSID1, a detail data set, is shown in Figure 2. NBSID1 is used to accelerate retrieval for the above queries. A system whose components are composed of these elements will contain substances derived from either a single combination of the elements, as in (1), or from any of the possible combinations of the set of elements, as in (2).

For query 1 the element symbols are used to construct the string to be searched within the field ELEMS in file NBSID1, e.g., "FES O". A "hashed" read is performed on the master data set for ELEMS returning the pointer to the first entry in NBSID1 containing the particular value of ELEMS. Subsequent chained reads in NBSID1 produce all the formulas containing the particular combination of elements.

For query 2 each of the element combinations is constructed. For each combination the retrieval performs, as in (1), a keyed read (i.e., via a hash algorithm) on the master associated with ELEMS. Each value of ELEMS may have more than one entry in NBSID1. For example, when ELEMS takes on the value "FES O", entries will be found for $FeSO_4$, $Fe_2(SO_4)_3$, $FeSO_4^+$, etc. Chained reads are then used to retrieve all the entries in NBSID1 with a particular value of ELEMS.

Thus, given a list of elements, the retrieval routines will find all substances containing either the entire set of elements as in query 1 or any of the possible subsets of the elements as in query 2. If the set is {Fe, S, O}, then for query 2 seven combinations of these elements are derived by the program, namely, "Fe", "S", "O", "FES", "FEO", "S O", and "FES O". A series of chained reads are then used, one for each value of the key item ELEMS. The order of the substances retrieved will be Fe, Fe^{2+}, Fe^{3+}, ..., S, S^+, S_2^{2-}, S_3^{2-}, ..., O_2, O^+, ..., FeS, FeS_2, ..., FeO, Fe_3O_4, ..., SO_2, SO_3, SO_4^{2-}, ..., $FeSO_4$, $Fe_2(SO_4)_3$, etc.

(2) Retrieval Mode 3: Search with an Open Set of Elements. In this query all substances containing a specific set of elements are requested, but in contrast to the prior queries, the set is open. Thus, all retrieved substances must contain all the specified elements, but they may also contain other elements as well. Again, a chained access technique is used. An example of this type of query would be (3) "FIND ALL substances CONTAINING: *ALL* of the elements in the set {Cl, O} *AND ANY* others".

The index file NBSID2 is designed for this case. The key item to be used is ELEMT in NBSID2. Figure 3 shows the organization of this index file. In this example all substances that contain both Cl and O as well as any other elements are to be retrieved. The algorithm will pick up one of these elements as the key item value of ELEMT and perform a chained read. All candidate substances are then on the chain. As each entry on the chain is read, a procedure is followed to determine if the corresponding substance contains the other input element(s). If it does, then the substance is accepted; if the substance does not contain the other input element(s), it is not accepted and the next entry on the chain is read. To enhance the performance of the retrieval, the element of the input set with the highest finding number is chosen as the key item value to be searched on. Due to the "standard order of arrangement" of the NBS tables, the element with the highest finding number usually appears in the least number of substances. This ensures that the chain to be searched is the shortest and therefore the least time consuming. In the present example Cl is chosen as the key item value of ELEMT to be searched on as it has the higher finding number. The retrieved substances contain both Cl and O, e.g., ClO, ClO^-, ClO_3^-, HClO, $Cd(ClO_4)_2$, $AgClO_3$, $(Cr(H_2O)_5Cl)^{2+}$.

Retrieval of the Properties Data. Once the formulas have been selected using the index files (or input by the user in the fourth retrieval mode), the property data are obtained from the data files for each of the selected formulas. The discussion above and Figures 2–5 show that the link between the index files and the data files is through the chemical formulas in the field FORML. Searches on the data files proceed by a series of chained reads, each using one of the selected values of FORML. As was shown in Figure 4, a chemical formula will usually have more than one entry in the data files due to the various conditions or physical states of substances in which that formula appears. Chained reads on the value of FORML are used to retrieve the property values for each of the selected substances in any of three data files. The files searched and the entries skipped or kept depend on the physical states requested by the user. For example, the substances in the Fe–S–O system are selected as Fe, Fe^{2+}, Fe^{3+}, ..., S, S^+, ..., O^+, O_2, ..., SO_2, FeS, FeS_2, ..., $FeSO_4$, etc. by using retrieval mode 2 above. If the user-requested physical states are CR, G, AO, and DIL, then the retrieval of the property values for these substances will be performed in data file NBSDA1 for the first three physical states requested and in NBSDA2 for the dilution data for aqueous solutions. An example of such a list of retrieved data for these substances is given in Figure 6. In the example only the enthalpy of formation data and, for dilute aqueous solutions, the heats of dilution data were requested.

Examples of the retrieval using the options described above are given in Appendix 2. The programs are interactive, and

Formula	State	DfH°(298)	
FE	CR	0	
S	CR	0	
S	CR2	0.33	
FE1.000S	CR	-100.0	
FE1S2	CR	-178.2	
FE1S2	CR2	-154.8	
FE1O	CR	-272.0	
FE1S1O4	CR	-928.4	
FE	G	416.3	
FE+	G	1181.98	
FE+2	G	2749.93	
FE+3	G	5712.8	Data retrieved from NBSDA1
S	G	278.805	
S+2	G	3543.4	
O+	G	1568.770	
O2	G	0	
S1O2	G	-296.830	
FE+2	AO	-89.1	
FE+3	AO	-48.5	
S-2	AO	33.1	
O2	AO	-11.7	
S1O2	AO	-322.980	
S1O2	in 100 H2O	-326.578	
S1O2	in 5000 H2O	-355.594	
S1O2	in 10000 H2O	-337.163	Data retrieved from NBSDA2
FE1S1O4	in 200 H2O	-992.0	
FE1S1O4	in 40000 H2O	-995.0	
....	....	.....	

Figure 6. Example of retrieval from data files.

Total Space: 16.2 Mbytes on HP-1000
Data Files: 5.1 Mbytes
Index Files: 4.7 Mbytes
Comment File: 0.34 Mbytes
Masters(Keys): 6.06 Mbytes

Figure 7. Sizes of the components of the database.

the user is given opportunities for narrowing the search results to specific physical states and to substances containing selected radicals, e.g., SO_3 and SO_4. In addition, an option is given to the user to choose output sorted either by physical states, as in Figure 5, or by the formulas for the substances, as in Appendix 2.

Disk Space Utilization. Figure 7 shows the size of the various portions of the database as installed on the CTDC HP-1000 minicomputer. It should be noted that the disk space includes extra space for the addition of new data into the database in the future as well as that required for efficient hashing algorithms. This extra space is approximately 30–40% of the total space of the database; that is, only 60–70% of the total space is in use at the present time.

THE PERSONAL COMPUTER DATABASE

The following is a description of the design, organization, and implementation of the NBS tables database for IBM or compatible PCs using the BTRIEVE system.

Each BTRIEVE file is stored in a format that contains both data and permanent indexes. Files are organized in pages. A header page contains the information about the file, such as the page size, record length, current number of records in the file, and tables describing each index. Data pages store blocks of records, with as many records as possible filling a page. Index pages store lists of each record's key fields along with the data and indexing pointers needed to maintain the indexing scheme. The index pages serve a purpose similar to that of the master files in IMAGE/1000.

BTRIEVE allows the user's program to store and retrieve data in indexed files. Access to these files can be random by key, sequential by key (forward or backward), or direct by record number. BTRIEVE uses a binary-tree,[9] a data structure with very fast access and efficient storage utilization, to keep all indexes to the data records. It maintains a separate index for each key and permits more than one index per file
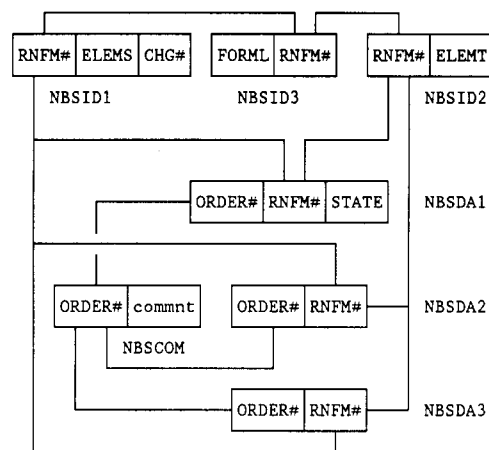

Figure 8. Relationships among files.

so that data in the file can be accessed in different ways, as required by the application. Up to 24 keys can be defined for one file.

As in IMAGE/1000, one file is linked to another file by a relation, or information common to both files. This involves using a field from one file as a key to access related records in a second file. Duplicate keys are allowed to identify a subset of records within a file that contains a single key value and to perform a chained access read within the file. A key field can be recognized and collated as either a string or an integer, as well as several other key types, e.g., float or logical. The system supports both fixed- and variable-length records in files. Unfortunately, variable-length records are unavailable on the minicomputer with IMAGE/1000, but their availability with BTRIEVE permits more efficient use of disk space in the personal computer version.

Keys and Files. Both IMAGE/1000 and BTRIEVE are multifile database management systems. The latter uses stand-alone files, while IMAGE requires the additional master files for pointers to keys in the detail data sets. As in IMAGE/1000, the relationships among files are established through common key fields. There are seven keys in the PC database:

FORML, X36, is the chemical formula.

RNFM#, I2, is the chemical formula accession number.

ORDER#, I2, is the species accession number reflecting the order of data in the *NBS Tables of Chemical Thermodynamic Properties*.

ELEMS, X20, contains all component elements of the substance.

ELEMT, X2, is a character string containing one of the component elements.

CHG#, I2, is an integer for signed charge number of the substance.

STATE, X4, identifies the physical state of the substance.

The fields FORML, ELEMS, ELEMT, CHG#, STATE, and ORDER# are the same as in the minicomputer version. CHG# corresponds to CHAR#, and ORDER# corresponds to ACCE# in the minicomputer version. The primary difference between the minicomputer and PC versions as shown in Figure 8 is the presence of the key RNFM#. This field decreases the disk space requirements of the database by using an accession number for each formula rather than repeating the formula in both the index and data files.

Figure 8 shows the interrelationship among files; only key items in each file are listed. The two items CASRN and CASFM, for CAS Registry number and CAS format of chemical formula, are not included in this version of the database. A file can be easily created for this database, establishing the connection between the CAS information and the NBS tables data when needed.

FORML	RNFM#
S1O2	108
ZN1I+	2395
NA2S1O4	7448

(a). NBSID3

RNFM#	ELEMS	CHG#	HYDR#	EL(10)
108	S O	0	0	1,2
2395	ZNI	1	0	1,1
7448	NAS O	0	0	2,1,4

(b). NBSID1

RNFM#	ELEMT	COEF
108	S	1
108	O	2
2395	ZN	1
2395	I	1
7448	NA	2
7448	S	1
7448	O	4

(c). NBSID2

Figure 9. Examples of index files.

ORDER#	RNFM#	STATE	DECIMAL	PROPT
323	108	L	010000	-320.5
324	108	G	333322	-294.286,-296.830,-300.194,10.548,248.22,39.87
325	108	AO	033010	-322.980,-300.676,161.9
12805	7448	CR	322322	-1375.653,-1387.08,-1270.16,23.225,149.58,128.20
12806	7448	CR2	000332	23.531,154.934,129.29
12807	7448	AI	022019	-1389.51,-1268.36,138.1,-201.

(a). NBSDA1

ORDER#	RNFM#	NH2O	DECIMAL	PROPT
326	108	100	3	-326.578
327	108	150	3	-327.298
328	108	200	3	-327.837
12808	7448	18	3	-1396.837
12809	7448	20	3	-1396.619
12810	7448	25	3	-1395.753
12811	7448	28.37	3	-1395.326

(b). NBSDA2

Figure 10. Examples of data files.

Files in the PC version NBS tables database are built somewhat differently in structure from those on the HP-1000 minicomputer. NBSID1 and NBSID2 are designed for the algorithms of data retrieval that were described in the discussion of the minicomputer version. NBSID3 establishes the relation between the key RNFM# and the chemical formula, FORML. The three files contain information on the attributes of substances such as the chemical formula, elemental composition, hydration number, and charge number. Molecular weights of the substances are not stored in these files as they can be calculated when needed from atomic weights. As in the minicomputer version, these attributes do not include the physical states of the substances. Examples of each of these files are given in Figure 9.

The data files that store the numerical values of the chemical thermodynamic properties are connected to the index files via the key RNFM# instead of FORML as in the minicomputer version. Numeric property values are stored as ASCII character strings to preserve their integrity. Variable-length records are used for files NBSDA1 and NBSDA3, where the standard states data and nonaqueous or mixed-solution data are contained. Compression of the field PROPT, the variable portion of the record for storing property values, is based on the decimal indicator of these properties. A 0 in the decimal

indicator DECIMAL, meaning an absent property value, will cause a 10- or 12-character compression of that field, depending on which property is absent. Such a variable-length-records compression cuts down on the use of disk space very efficiently. NBSDA2 stores heat of dilution data on aqueous species, which are in different moles of water. Figure 10 shows examples of portions of the data files NBSDA1 and NBSDA2. The relationships among these files is evident from Figures 9 and 10.

Programs and Disk Space Utilization. The algorithms and programs for search and retrieval are similar to those described for the minicomputer version. For the PC the programs were written by using the BASICA interpreter. Access to the database was accomplished by using the vendor-supplied BTRIEVE subroutines.

The modified organization of the database for PCs required far less disk space than the minicomputer version. The total space taken by all the files is about 5 MB. Each file can be stored on a single diskette of either 1.2-MB floppy or 360-KB capacity. Further compression of the files is certainly possible and will be considered for future revisions.

ACKNOWLEDGMENT

APPENDIX 1. DESCRIPTION OF DATA SETS IN THE NBS TABLES DATABASE

Master Data Sets.

data set	field	type	description
NBSRN	ACCE#	X6	accession numbers for substances reflecting order of substances in ref 3
NBSF	FORML	X40	chemical formulas of substances
CAS1	CASRN	X16	CAS Registry number
CAS2	CASFM	X40	CAS format of chemical formulas
NBSEL	ELEMT	X2	element symbols
NBSES	ELEMS	X20	component elements of formulas, 10X2, in order of decreasing finding number
CHARGE	CHAR#	I2	charge numbers of substances; only noncharacter field in database
NBSTAT	STATE	X10	standard physical states

Detail Data Sets.

NBSID1   index file containing the general information on the substances independent of their physical states; key fields are FORML, ELEMS, and CHAR#. The non-key fields are

NUMEL1	X8	contain the stoichiometric coefficient of each
NUMEL2	X8	element in the substance. NUMEL subfield 1
...	...	corresponds to the element in ELEMS subfield 1, etc.
NUMEL0	X8	...
MASS	X10	molar mass of the substance with four decimal places
HYDR#	X10	hydration number, i.e., number of waters of hydration

NBSID2   index file containing element components of the substance; key fields are FORML and ELEMT. The only non-key field is

COEF X8 stoichiometric coefficient of element in ELEMT in FORML

NBS TABLES OF CHEMICAL THERMODYNAMIC PROPERTIES

J. Chem. Inf. Comput. Sci., Vol. 29, No. 1, 1989 **37**

NBSDA1 data file containing the thermodynamic property values of substances not in solution, e.g., L, G, CR; key fields are ACCE#, FORML, CASRN, CASFM, and STATE. The non-key fields are

SCODE	X10	numerical physical state code, a decimal number indicating the physical state
NCOMM	X2	number of comment records for this entry
DECIMAL	X6	decimal indicator consisting of six integer subfields of 1 byte each, each defining the number of decimal digits in each property value below

    0   if property is absent
    1–7 number of accurate digits to the right of the decimal point whether an error value is given or not
    8   for a defined zero
    9   for an integer

H0	X12	value of $\Delta_f H°(0)$
H298	X12	value of $\Delta_f H°(298)$
G298	X12	value of $\Delta_f G°(298)$
H298H0	X10	value of $H°(298) - H°(0)$
S298	X10	value of $S°(298)$
CP	X10	value of $C_p°(298)$

NBSDA2 data file containing dilution property values of aqueous solution; key fields are ACCE#, FORML, CASRN, and CASRM. The non-key fields are

NH2O	X10	number of moles of water associated with 1 mol of substance in the solution
ISOMER	X4	isomer number of the substance, e.g., i2, i3

SCODE, NCOMM, DECIMAL, H0, H298, G298, H298H0, S298, and CP as in NBSDA1

NBSDA3 data file containing thermodynamic property values of nonaqueous and mixed-solution data; key fields are ACCE#, FORML, CASRN, and CASFM. The non-key fields are

SOLVNT	X30	description of the solvent (s)
SYSTEM	X4	designation of the physical states for mixed solution

SCODE, NCOMM, DECIMAL, H0, H298, G298, H298H0, S298, and CP as in NBSDA1

NBSCOM comment file; the key supplemental description of some species is given in this file. The key field is ACCE#; the non-key field is

COMMET X80 comment

APPENDIX 2. EXAMPLES OF RETRIEVALS FROM THE DATABASE

The following is an example of the screen input and outcome from the program NBSTAB. User input is indicated in bold. The initial display once the program is invoked is as follows.

```
+-------------------------------------------+
|                                           |
|         NIST TABLES OF CHEMICAL           |
|         THERMODYNAMIC  PROPERTIES         |
|                                           |
|                                           |
|                                           |
|         A DATA RETRIEVAL SYSTEM           |
|                                           |
|             VERSION 1.0                   |
|                                           |
|                                           |
|             PREPARED BY                   |
|                                           |
|        The Electrolyte Data Center        |
| National Institute of Standards and Technology |
|                 USA                       |
|                                           |
+-------------------------------------------+

 Type <RETURN> to continue...
```

The second screen presents the search options.

```
*****************************************************************

Four options for retrieval of data:

SEARCH BY ELEMENTS: Select those substances which contain...

   1.  ALL of a set of elements AND NO others.
   2.  ANY of a set of elements AND NO others.
   3.  ALL of a set of elements AND ANY others.

SEARCH BY FORMULAS

   4.  Retrieve data by input chemical formulas.

*****************************************************************

Enter the number of the search option desired: 1
```

Under options 1–3 the third screen prompts the user for the set of elements. The search is performed, and the number of hits is displayed.

```
    Input the set of elements. Hit return to finish:
    ===================================

       Element  1:FE
       Element  2:O
       Element  3:S
       Element  4:

    ===================================

    7 substances found in the database.
```

To provide the possibility of a more refined search, some subselection can be done. In the next screen the user is given a chance to subselect the substances based on the presence of a radical.

```
Do you want substances with a specific radical?(Y/N) Y

Options available -

  1. Keep ONLY those formulas containing a specified radical

  2. Keep ALL, but for those containing a specified element
     keep ONLY those in which the element appears in a
     single specified radical, e.g. S in S1O3.

Choose a selection option. Enter its number, (1 or 2): 1

The following radicals are available:
-------------------------------------------
  AS1O3     AS1O4     B1O2      BE1O2     BR1O3     BR1O4
  C1O3      C2O4      C1N       C1N1S     CL1O2     CL1O3
  CL1O4     CR1O4     CR2O7     I1O3      MO1O4     N1H3
  N1H4      N1O3      N1O2      P1O3      P1O4      S1O3
  S1O4      SE1O3     SE1O4     SI1O3     SI1O4     V1O3
  V1O4      W1O4      U1O2      O1H
-------------------------------------------

Choose one of the above. Enter it: S1O4

7 matches found.
```

At this point the user is prompted for a file in which the results of the search may be stored. Also, the user can preview the selected formulas.

```
  Enter an output file name: GEORGE

  Do you want to look over the selected formulas first?(Y/N) Y

     Formula           Z   Hydr_no     FE       S        O
  FE1S1O4              0     0       1.0000   1.0000   4.0000
  FE1S1O4+             1     0       1.0000   1.0000   4.0000
  FE1S1O4: H2O         0     1       1.0000   1.0000   4.0000
  FE1S1O4:4H2O         0     4       1.0000   1.0000   4.0000
  FE1S1O4:7H2O         0     7       1.0000   1.0000   4.0000
  FE(S1O4)2-          -1     0       1.0000   2.0000   8.0000
  FE2(S1O4)3           0     0       2.0000   3.0000  12.0000

  Type <RETURN> to continue...
```

At this point the user must specify the states of the substances that are to be selected.

```
The following states of the substances are available:
--------------------------------------------------------
1.  (cr)  crystalline solid
2.  (l)   liquid
3.  (g)   gaseous
4.  (am)  amorphous solid
5.  (vit) vitreous or glassy
6.  (aq)  aqueous, concentration unspecified
7.  (ai)  aqueous, fully ionized standard, m=1 mol/kg
8.  (ao)  aqueous molecules or aqueous ions
             un-ionized standard, m=1 mol/kg
9.  (dil) specified aqueous state
10. (mix) mixed or non-aqueous solution
--------------------------------------------------------

Enter the number of the state desired.
Hit <RETURN> to finish.
  1th: 1
  2th: 4
  3th: 5
  4th: 6
  5th: 7
  6th: 8
  7th: 9
  8th: 10
  9th:
```

The user must now choose the form of the output desired and the specific set of properties to be displayed and/or stored on the output file.

```
*****************************************
Two forms for the output:
  1. Formulas only;
  2. Formulas and properties.
*****************************************
Choose one. Enter its number: 2
=========================================
The following data items can be retrieved:
  1. Molar mass
  2. DHf(0)
  3. DHf(298)
  4. DGf(298)
  5. H(298)-H(0)
  6. S(298)
  7. Cp(298)
=========================================
HOW MANY items do you want to retrieve:4
Which items do you want?
Input example:
  Enter: 1,3,4,6
  for Molar mass, DHf(298), DGf(298) and S(298)

Please enter the property number(s): 3,4,6,7
```

The final output is then displayed to the user's screen and stored in the selected output file. The following is a portion of the output to the screen sorted by substances, i.e., the order of reference 3.

sort output by physical states or by substances? (P/[S]) S

formula	state	DHf(298)	DGf(298)	S(298)	Cp(298)
FE1S1O4	CR	−928.4	−820.8	107.5	100.58
FE1S1O4	AI	−998.3	−823.43	−117.6	
FE1S1O4	200	−992.0			
FE1S1O4	40000	−995.0			
FE1S1O4+	AO	−931.8	−772.7	−130	
FE1S1O4:H2O	CR	−1243.69			
FE1S1O4:4H2O	CR	−2129.2			
FE1S1O4:7H2O	CR	−3014.57	−2509.87	409.2	394.47
FE(S1O4)2−	AO		−1524.5		
FE2(S1O4)3	CR	−2581.5			
FE2(S1O4)3	AQ1	−2799			
FE2(S1O4)3	AI	−2825.0	−2242.8	−571.5	
FE2(S1O4)3	400	−2732			
FE2(S1O4)3	1000	−2733.0			
FE2(S1O4)3	2000	−2733.8			
FE2(S1O4)3	3000	−2734.7			
FE2(S1O4)3	H1CL1O4+ 53.7H2O	−2799			
FE2(S1O4)3	H2S1O4+ 220H2O	−2791			

The process can be continued for a new search or terminated.

```
Do you want to choose the output again?(Y/N) N

Output file is GEORGE

Retrieval Completed.
Welcome to NBSTAB Again!
```

REFERENCES AND NOTES

(1) Certain commercial equipment and software products are identified in this paper to adequately specify the design and use of the programs described herein. Such identification does not imply recognition or endorsement by the National Institute of Standards and Technology, nor does it imply that the material or equipment identified is necessarily the best available for the purpose.

(2) Wagman, D. D.; Evans, W. H.; Parker, V. B.; Halow, I.; Bailey, S. M.; Schumm, R. H.; Churney, K. L.; Nuttall, R. L. Selected Values of Chemical Thermodynamic Properties. *NBS Tech. Note* **1968**, 270-3, through **1981**, 270-8. Available from U.S. Government Printing Office, Washington, DC.

(3) Wagman, D. D.; Evans, W. H.; Parker, V. B.; Schumm, R. H.; Halow, I.; Bailey, S. M.; Churney, K. L.; Nuttall, R. L. The NBS tables of chemical thermodynamic properties. Selected values for inorganic and C_1 and C_2 organic substances in SI units. *J. Phys. Chem. Ref. Data, Suppl.* **1982**, *11*, No. 2.

(4) IMAGE/1000 is a product of Hewlett-Packard Corp.

(5) *IMAGE/1000-II. Data Base Management System. User Reference Manual*; Hewlett-Packard Corp.: Sunnyvale, CA, Jan 1986; Part No. 92081-90001.

(6) BTRIEVE is a product of Novell Corp.

(7) Neumann, D. B. Data Base Management Systems. Part I − A Quick Overview. In *NonBibliographic Data Banks in Science and Technology*, Proceedings of a CODATA Seminar, Stockholm, Sweden, Oct 15−22, 1983; Schwarz, S., Watson, D., Alvfeldt, O., Eds.; ICSU Press: Paris, 1985; pp 139−151. Available from UNIPUB, Ann Arbor, MI.

(8) *ASK/1000, Version C. Reference Manual*; COGELOG: Les Ulis, France, 1987; 121 pp. Distributed by Corporate Computer Systems, Inc., Holmdel, NJ.

(9) Date, C. J. *An Introduction to Database Systems*, 3rd ed.; Addison-Wesley Publishing: Reading, MA, 1982; pp 47−50.