(18) Corey, E. J.; Feiner, N. F. J. Org. Chem. 1980, 45, 757, 765.
(19) Wipke, W. T.; Gund, P. J. Am. Chem. Soc. 1974, 96, 299.
(20) Corey, E. J.; Feiner, N. F.; Orf, H. W.; Long, A. K.; Vinson, J. W.; Hewett, A. P. W.; Stolow, R. D., unpublished results.
(21) The steric control module is outlined in ref 17b, Chapter 5, Section VI.
(22) A complete description of the language as of 1976 appears in: H. W. Orf, Ph.D. Thesis, Harvard University, 1976, Appendix C.
(23) In order to store CHMTRN sentences as compactly as possible and to save time as they are being read by the computer, the chemical English sentences are translated into a more compact, computer-readable form by a "translator" program. The translator program, TBLTRN (TaBLe TRaNslator), was written by D. E. Barth at Harvard University in 1971. It is written in FORTRAN and is available as program 10-168 from the DECUS Program Library.
(24) Corey, E. J.; Orf, H. W.; Pensak, D. A. J. Am. Chem. Soc. 1976, 98, 210.

Computer-Aided Spectral Identification of Laser-Induced Plasma Emission

YEDIDYAH LANGSAM,* AARON TENENBAUM, and JOSEPH THURM

Department of Chemistry and Department of Computer & Information Science, Brooklyn College of the City University of New York, Brooklyn, New York 11210

Received February 7, 1980

The deconvolution of the electronic emission spectra produced by a laser-induced plasma may be performed by computer analysis. This paper presents two algorithms for matching such spectra against standard tables. The relative efficiencies of the two algorithms are analyzed.

I. INTRODUCTION

Since its very inception in the early 1960s, laser technology has been applied to a diverse number of fields. Yet it was only recently that the laser's inherent monochromaticity, high intensity, and coherence was utilized in order to directly manipulate the course of chemical reactions.[1-3] In recent years, much effort has been spent in the initiation of chemical reactions by the absorption of resonant laser energy. The technique of Laser-Induced Dielectric Breakdown (LIDB), on the other hand, creates a plasma and subsequent chemical reaction even where a resonant condition does not exist.[2,4-7] Of particular interest is the investigation of the identities and lifetimes of various species found within the laser-induced plasma and their dependence on such parameters as pressure, laser fluence, and companion gases which may be present.

Recent work in our laboratory focused on the LIDB of the metal carbonyls.[8] When a sample of metal carbonyl is irradiated by a CO_2 laser of sufficient fluence (J/cm^2), visible breakdown occurs. A number of features of the laser-induced plasma are easily apparent to the eye. The plasma is an extremely intense blue-white isotropic source of light containing distinct regions. At the center there is an extremely intense core of indeterminably small volume. In this high-energy region only highly ionized atoms and molecules may exist. Surrounding this core is a region of lesser intensity consisting of excited species rather than ions.[8] Electron energies may range from 20 keV in the core to tens of electron volts in the surrounding cloud.[9] As is to be expected, measurements of the spectral distributions of the plasma emission indicate a strong line spectrum superimposed on a continuum with peaks at characteristic wavelengths depending again on the plasma medium as well as the laser energy.

An optical multichannel analyzer was used in order to identify the species produced within a laser-induced dielectric breakdown plasma. This device records the plasma emission, after suitable background subtraction and enhancement by multipulse integration, into 500 channels as a function of wavelength. The channels may be calibrated vs. known Hg-lamp emission lines, and thus the wavelengths corresponding to each of the 500 channels may be calculated. In these experiments, the visible region was scanned in two parts, a "high" region roughly between 4164 and 7164 Å calibrated by means of the Hg lines at 5461 and 5770 Å, and a "low" region corresponding to 3093–6001-Å emission calibrated against the 4350- and 5461-Å Hg lines. Typical emission spectra obtained by using the OMA are presented in Figure 1.

As mentioned earlier, the plasma state is richly populated by a wide range of ionic and neutral, ground and excited, atomic and molecular species. The identification of these plasma states may be accomplished by comparison with all the tabulated emission lines of any species which may conceivably be present. Examination of the thousands of lines tabulated in ref 10–15 reveals that the identification is complicated not only by the many overlapping emission lines but by the line width of the OMA detector (~6 Å/channel). It was therefore thought that the nature of the emitting species could be elucidated by means of a statistical matchup of OMA spectra with known emission lines. While a manual search of this type could be described as tedious if not overwhelming at best, it could easily be done, for the many emission spectra investigated, by means of a computer.

II. METHODOLOGY

The analysis of the OMA output proceeds in three phases. The first phase analyzes the OMA spectra and calculates the wavelength (in angstroms) of each peak observed, by comparison with the mercury-line calibration. These wavelengths, from both the "high" and "low" regions, are sorted in ascending order and stored on a disk.

In each computer run, the OMA spectra are compared with up to five elements or compounds (e.g., O, O_2, CO, etc.). Each of these elements, of course, has a group of subcategories (e.g., OI, OII, $B^1\Sigma^+$, $X^3\Sigma^-$, etc.), each of which has its own characteristic spectrum. Up to five "elemental input decks" containing these spectra are input in the second phase of the processing. Each input deck consists of a card containing the name of the major element or compound (identified by a 1 in column 1), followed by one or more groups of cards representing the subcategories and their spectra. Each of these groups consists of a card containing the name of the subcategory (identified by a 2 in column 3), followed by cards containing the known wavelengths of the spectrum for that subcategory (identified by a 3 in column 5). A sample ele-

0095-2338/80/1620-0230$01.00/0

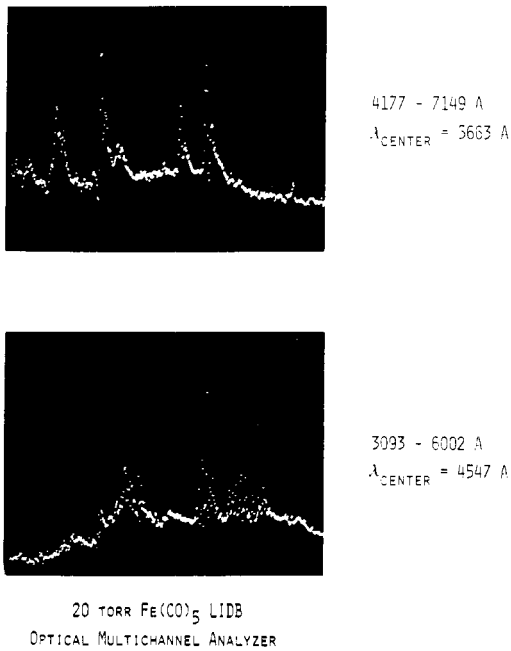

Figure 1. Optical multichannel analyzer typical spectrum.

mental input deck appears in Figure 2. In the second phase of the analysis, all of these wavelengths (for all of the subcategories in all the elemental input decks) are identified by subcategory and are sorted in ascending order of wavelength.

At this point in the processing, there exist two ordered files of wavelengths. One contains the wavelengths observed by the OMA from the experiment, and the other contains the known wavelengths of all subcategories of up to five elements or compounds. However, even if a subcategory is present in the OMA, the wavelengths may not match exactly because the calibration of the OMA and the standard tables may not have been identical. Thus, when the experimental data are matched against the standard data, the match can be made if two wavelengths match each other to within a given tolerance. The tolerance used was approximately ±6 Å and is input to the third phase of the analysis.

The third phase merges the two ordered files and identifies which standard wavelengths match each observed OMA wavelength. While many standard wavelengths are allowed to match a single OMA peak (since a peak may indicate the presence of many subcategories), only the closest OMA peak is considered to match a single standard wavelength.

An algorithm (in PL/I style) for this merge follows. Note that on OMA wavelength and a known wavelength are considered to match if they are within the tolerance and there is no other OMA wavelength closer to the known wavelength. *omaout* is the ordered file of OMA wavelength and *waves* is the ordered file of known wavelengths. *toler* is the given tolerance.

```
read file (omaout) into (inwavel);
read file (omaout) into (inwave2);
low = inwavel - toler;
high = inwavel + toler;
do while (there is more data in the waves file);
  read file (waves) into (wave);
  do while (wave < high |
          abs(wave-inwave2) < abs(wave-inwavel));
    inwavel = inwave2;
    read file (omaout) into (inwave2);
    low = inwavel - toler;
    high = inwavel + toler;
  end;
  do while (wave < low);
    read file (waves) into (wave);
  end;
  /* at this point, inwavel matches wave */
  process the match between inwavel and wave;
end;
```

```
1 '02'
 2 'B1S+ -> X3S-'
  3 7706.41
  3 7619.33
  3 6968.63
  3 6882.47
  3 6369.80
  3 6286.61
  3 5795.13
  3 5385.49
  3 7987
  3 7879.17
  3 7779.03
  3 7683.85
  3 7593.73
  3 7240
  3 7141
  3 7043
  3 6953
  3 6867.2
  3 6276.6
 2 'C1S- -> X3S-'
  3 4791.4
  3 4491.5
 2 '3D? -> 1D?'
  3 4378
  3 4326
  3 4317
  3 4244
  3 4240
  3 4221
  3 4215
  3 4135
  3 4127
  3 4114
  3 4107
  3 4090
  3 4066
  3 4071
  3 4031
  3 4009
  3 3985
  3 3887
  3 3866
  3 3861
  3 3844
  3 3813
  3 3792
  3 3771
  3 3698
 2 'A  ->  X'
  3 4880
  3 4577
    .
    .
    .
```

Figure 2. Sample OMA elemental input deck for molecular oxygen.

This method was implemented and run on an IBM 370/168 to produce the desired matches. However, the method can be implemented on a far smaller machine without change, since all large data collections are kept in external files rather than in core.

An alternate method involves the use of a binary search to locate each successive OMA wave in the ordered file of standard waves. In order to use this method, we must be able to directly access any element in the file of standard waves. Assume that there are n elements in that file. Then the statement

```
read file (waves) into (wave) key (k);
```

reads the kth element, where $1 < k < n$.

To present this method, let us assume the existence of a routine *binsrch* (*w*, *from*, *to*, *toler*, *p*, *wave*) with input parameters *w*, a wavelength, *from* and *to*, integers representing the lower and upper bounds of the section of the standard table being searched, and *toler*, a specification of how close an OMA and a standard wave must be to match. The routine sets two output parameters *p* and *wave* such that *p* is a position between

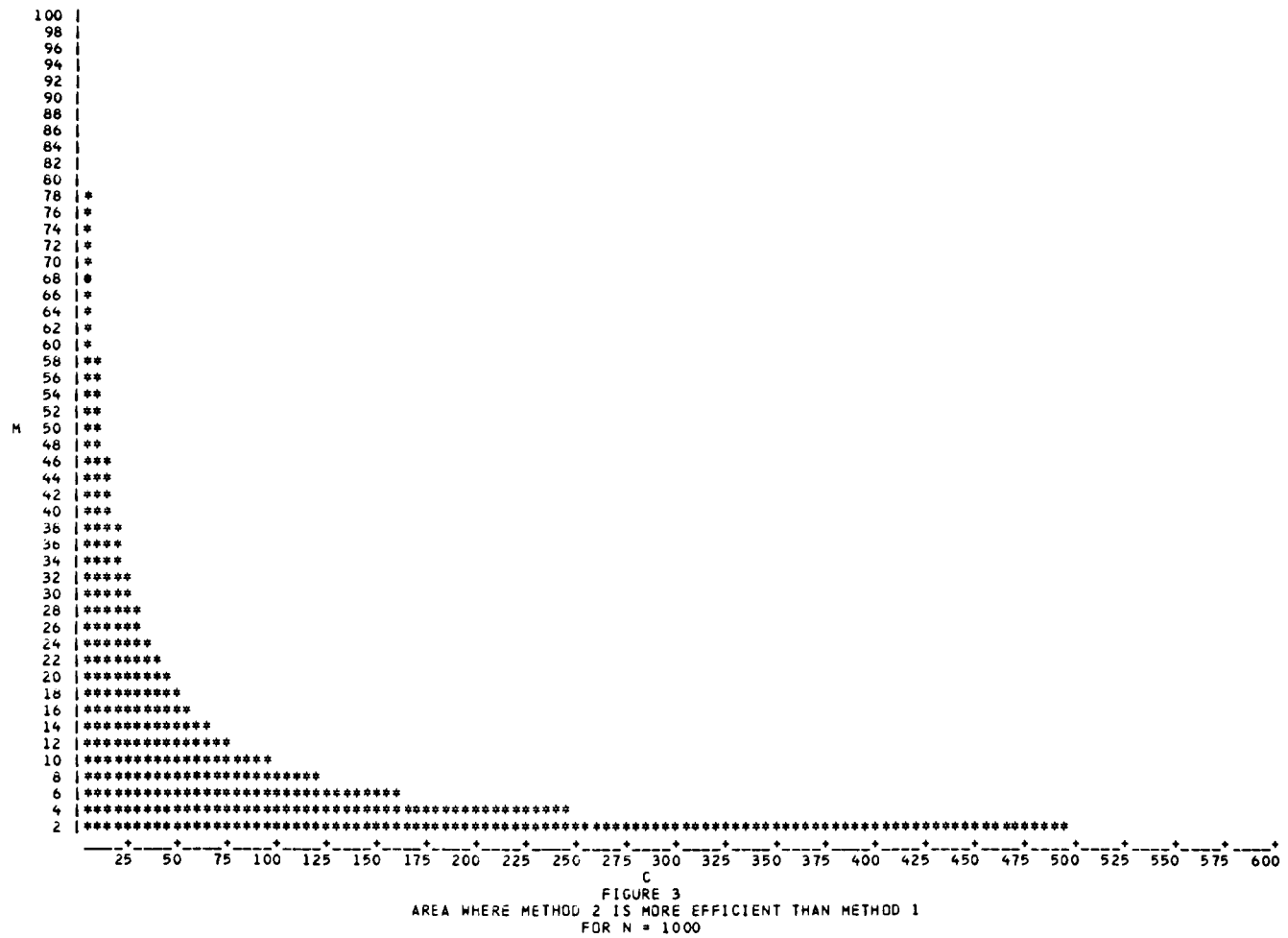

Figure 3. Area where method 2 is more efficient than method 1 for $n = 1000$.

from and *to*, in the standard table, of a wavelength *wave* which is within *toler* of *w*. We will shortly present an algorithm for *binsrch*, but first we present the method which uses *binsrch* to locate matches. We assume also that at the end of the OMA file, there is added one absurdly large value. This is needed for control purposes.

```
read file (omaout) into (inwave1);
from = 1;
to = n;
do while (there is more data in the omaout file);
  read file (omaout) into (inwave2);
  call binsrch (inwave1, from, to, toler, p, wave);
  pp = p;
  do while (abs(inwave1-wave) < toler);
    if abs(inwave1-wave) < abs(inwave2-wave) then
      process the match between inwave1 and wave;
    pp = pp + 1;
    read file (waves) into (wave) key (pp);
  end;
  pp = p - 1;
  read file (waves) into (wave) key (pp);
  do while (abs(inwave1-wave) < toler);
    if abs(inwave1-wave) < abs(inwave2-wave) then
      process the match between inwave1 and wave;
    pp = pp - 1;
    read file (waves) into (wave) key (pp);
  end;
  inwave1 = inwave2:
  from = p;
end;
```

The routine *binsrch* (*w*, *from*, *to*, *toler*, *p*, *wave*) may be outlined as follows:

```
lo = from; hi = to;
mid = (lo + hi)/2;
read file (waves) into (wave) key (mid)
do while ((abs(wave-w) > toler) & (hi > lo));
  if wave > w then hi = mid - 1;
              else lo = mid + 1;
  mid = (lo + hi)/2;
  read file (waves) into (wave) key (mid);
end;
p = mid;
```

Note that both methods first involve sorting the standard data and the OMA data. The first method sorts both files in order to be able to perform a merge. The second method uses a binary search in the standard file, so that file must be sorted. The only reason that the OMA file must be sorted in the second method is because the problem specification stipulates that only one OMA wavelength can match a specific standard wavelength. Thus, when matching an OMA wavelength against a standard wavelength, it is necessary to keep track of the next higher OMA wavelength to ensure that it does not match the standard wavelength more closely. If this restriction were elminated, it would be unnecessary for the OMA file to be sorted. Instead, each OMA wavelength could be located in the standard file in random order.

For analysis of these two methods to determine which is more effective, note that the speed of both algorithms is determined by the number of elements read from the standard file. This is because the OMA file must be processed once under both methods, and the speed of computation is insignificant when compared to the speed of I/O. Also, as we have mentioned, both files must be sorted in both methods. (Even if the standard file were small enough to be implemented as

20 TORR FE(CO)5 LIDB (12 PULSES) 5/6/78 PAGE 12
TOLER FACTOR 1

WAVE SEEN	IRON	CO	COPLUS	OXYGEN	O2
4273.00					
4273.00		4270.80 E3S- -> A3P			
4273.00	4271.16 FE I				
4273.00	4271.76 FE I				
4273.00			4271.95 A2P -> X2S+		
4273.00				4272.30 O II	
4273.00			4274.38 A2P -> X2S+		
4273.00					4274.40 A -> X
4273.00				4275.52 O II	
4273.00				4276.71 O II	
4273.00				4276.71 O II	
4273.00				4277.40 O II	
4273.00				4277.90 O II	

Figure 4. Partial output of the OMA identification computer program (for the data of Figure 1).

an in-core table, the number of accesses to this table would be a major factor in determining the efficiency of the two algorithms.)

Assume that the standard file *waves* contains n items and that the OMA file *omaout* contains m items, where n is larger than m. Obviously, the first method reads the *waves* file n times, once for each standard wave. The analysis of the second method is somewhat more complex. Reads of the *waves* file take place in two places: first, in the binary search process where we seek a standard wave which may match the OMA wave and, second, in the main algorithm in which we sequentially scan the *waves* file in a forward and backward direction until we are outside the tolerance of a match.

Let us first examine the sequential scan. Assume that the n standard wavelengths in the *waves* file are uniformly distributed between the lowest and highest wavelength. Then within any span of wavelengths of a given distance, there are a constant number of elements of the *waves* file. In particular, there are c elements in *waves* whose wavelengths are between *inwavel−toler* and *inwavel+toler*. One of these c elements is read in the binary search and the remaining $c - 1$ elements are read by the sequential search. In addition, two more elements are read from *waves* in determining the boundaries of possible matches. Thus, the sequential search performs $c + 1$ reads of *waves* for each OMA wave, yielding a total of $m*(c + 1)$ reads.

Now let us examine the reads performed in the binary search. In searching for the first OMA wave, the entire standard file of n waves is being searched. However, the search terminates as soon as the section of the file being considered spans a distance of less than 2* *toler*. The number of elements in such a section equals c. Therefore, the number of reads is approximately $[log_2 (n/c)]$, where $[x]$ denotes the smallest integer greater than or equal to x. If a binary search in the entire *waves* file were performed for each OMA reading, then the total number of reads performed in the binary search would be $m* [log_2 (n/c)]$. However, the section of *waves* being searched is decreased after each search since the OMA waves are ordered so that only the section of *waves* past the previous match must be searched.

Let us assume that the m OMA waves are also uniformly distributed in the range of the n standard waves. Then each successive binary search takes place in a file reduced by n/m. That is to say, the first search is in a file of size n, the second search is in a file of size $(m - 1)*n/m$, and the last search is in a file of size n/m. Thus, the first search requires $[\log_2 (n/c)]$ reads, the second requires $[\log_2 ((m - 1)n/(m*c))]$ reads, and the last requires $[\log_2 (n/(m*c))]$ reads. Thus, the total number of reads performed by the binary search is

$$\sum_{i=1}^{m} [\log_2 ((i*n)/(m*c))]$$

For example, suppose $n = 1000$, $m = 50$, and $c = 2$, and all the uniformity assumptions hold. Then the first method performs $n = 1000$ reads. The second method performs $m*(c + 1) = 150$ reads in the sequential process while the binary search process performs approximately

$$\sum_{i=1}^{50} [\log_2 ((i*n)/(m*c))] = 4 + 2(5) + 3(6) + 6(7) + 13(8) + 22\ (9) = 373 \text{ reads}$$

Thus the second method performs a total of 523 reads and is superior. However, if $m = 200$ and $c = 4$, the second method performs 1000 reads in the sequential process alone and so is inferior to the first method. Figure 3 illustrates those values of c and m for which the second method is superior, for $n = 1000$.

Note, however, that it is usually more efficient to perform a single sequential pass through a file (as in the first method) because of blocking and buffering capabilities than to directly retrieve elements of a file (as in the second method). Thus,

20 TORR FE(CO)5 LIDB (12 PULSES) 5/6/78

DETAILED SUMMARY OF MATCHES

FE I	MATCHED	77	TIMES	
FE II	MATCHED	0	TIMES	
O I	MATCHED	6	TIMES	
O I F	MATCHED	0	TIMES	
O II	MATCHED	85	TIMES	
O II F	MATCHED	0	TIMES	
O III	MATCHED	5	TIMES	
O III F	MATCHED	2	TIMES	
O IV	MATCHED	9	TIMES	
O V	MATCHED	5	TIMES	
O VI	MATCHED	0	TIMES	
O VII	MATCHED	0	TIMES	
B1S+ -> X3S-	MATCHED	0	TIMES	O_2
C1S- -> X3S-	MATCHED	0	TIMES	
3D? -> 1D?	MATCHED	10	TIMES	
A -> X	MATCHED	4	TIMES	
B3S- -> X3S-	MATCHED	9	TIMES	
A2P -> X2P	MATCHED	29	TIMES	O_2^+
B4S- -> A4P	MATCHED	9	TIMES	
MISC ??	MATCHED	9	TIMES	
C I	MATCHED	12	TIMES	
C I F	MATCHED	0	TIMES	
C II	MATCHED	18	TIMES	
C III	MATCHED	16	TIMES	
C III F	MATCHED	0	TIMES	
C IV	MATCHED	6	TIMES	
C V	MATCHED	0	TIMES	
B1S+ -> A1P	MATCHED	5	TIMES	CO
C1S+ -> A1P	MATCHED	5	TIMES	
B3S+ -> A3P	MATCHED	0	TIMES	
A3S+ -> A3P	MATCHED	24	TIMES	
D3D -> A3P	MATCHED	73	TIMES	
C1S+ -> A3S+	MATCHED	0	TIMES	
E3S- -> A3P	MATCHED	8	TIMES	
B2S+ -> X2S+	MATCHED	0	TIMES	CO^+
A2P -> X2S+	MATCHED	24	TIMES	
B2S+ -> A2P	MATCHED	6	TIMES	
MISC	MATCHED	4	TIMES	
A3P->X3P	MATCHED	9	TIMES	C_2
C1P->B1P	MATCHED	1	TIMES	
B3P->X3P	MATCHED	0	TIMES	
COMETS	MATCHED	7	TIMES	C_3

Figure 5. Results of the OMA identification program.

in most practical situations, the first method is to be preferred.

III. RESULTS AND DISCUSSION

The results of the third phase consist of two printouts. The first printout contains one page for every OMA peak (see Figure 4). This page contains each standard wavelength and subcategory that was found to match the given OMA peak. The second printout contains, for each subcategory, the number of OMA peaks that matched (see Figure 5).

Analysis of this information leads to conclusions regarding the identities of the species which are contained within a laser-induced plasma. Since the small dense plasma core is dominated by high-energy short-lived ionic species, and the region outside of the core is dominated by radical and excited-state species of transient existence, their identities cannot be determined by using the steady-state forms of spectroscopy. By use of the optical multichannel analyzer, one is able to acquire data over a wide wavelength range simultaneously with or delayed by a predetermined amount from the initiating laser pulse. The large number of species leads to a complex spectrum which precludes identification on a line by line basis. By use of our technique, however, most species may be identified by comparison with known spectroscopic constants of all potential species.

Although the technique presented here was described in terms of the analysis of relatively complex spectra obtained by means of an optical multichannel analyzer, it may readily be applied to the deconvolution of multicomponent spectra using more traditional forms of spectroscopy.

ACKNOWLEDGMENT

We thank Professor A. M. Ronn, A. Schwebel, and R. Sausa for many helpful discussions.

REFERENCES AND NOTES

(1) C. P. Robinson, *Ann. N.Y. Acad. Sci.*, **267**, 81, (1976).
(2) A. M. Ronn, *Sci. Am.*, **240** (5), 114 (1979), and references cited therein.
(3) E. Fuss and T. P. Colter, *Appl. Phys.* **12**, 265 and references cited therein (1977).
(4) P. D. Maker, R. W. Terhune, and C. M. Savage, *Quantum Electron., Proc. Int. Congr., 3rd, 1963*, 155 (1964).
(5) A. M. Ronn, *Chem. Phys. Lett.*, **42** (2), 207 (1976).
(6) S. T. Lin and A. M. Ronn, *Chem. Phys. Lett.*, **56** (3), 414 (1978).
(7) S. T. Lin and A. M. Ronn, *Chem. Phys. Lett.*, **49**, 255 (1977).
(8) Y. Langsam and A. M. Ronn, *Chem. Phys.*, in press.
(9) E. Yablonovitch and H. S. Kwok, *Opt. Commun.*, **18**, (1), 103 (1976).
(10) W. F. Meggers, C. H. Corliss, and B. F. Scribner, "Tables of Spectral Line Intensities Part I—Arranged by Elements", *NBS Monogr.* (U.S.) **No. 145** (1975).
(11) W. L. Wiese, M. W. Smith, and B. M. Glennon, "Atom Transition Probabilities—Hydrogen Through Neon" *Natl. Stand. Ref. Data Ser. (U.S. Natl. Bur. Stand.)*, **NSRDS-NBS 4** (1966).
(12) M. W. Smith and W. L. Wiese, *J. Phys. Chem. Ref. Data*, **2**, 85 (1973).
(13) P. H. Krupenie, "The Band Spectrum of Carbon Monoxide", *Natl. Stand. Ref. Data Ser. (U.S. Natl. Bur. Stand.)*, **NSRDS-NBS 5** (1966).
(14) P. H. Krupenie, "The Spectrum of Molecular Oxygen", *J. Phys. Chem. Ref. Data*, **1** (2), 423 (1972).
(15) R. W. B. Pearse and A. G. Gaydon, "The Identification of Molecular Spectra", Wiley, New York, 1963.

Theory of Correlation Tables. 1

T. VESZPRÉMI* and G. CSONKA

Department of Inorganic Chemistry, Budapest Technical University, 1521-H, Budapest, Hungary

Received October 29, 1979

A possible mathematical model is presented for correlation tables used in spectroscopy. A process based on information theory is demonstrated through an example for the optimum construction of correlation tables. In this example the construction of an ^1H NMR correlation table is investigated; the method is general, can be used for other spectroscopies, and is suitable for the construction of correlation tables used in computerized evaluation of spectra.

The accumulation of spectroscopic experience led to the recognition that some fragments of molecules can be observed irrespective of their chemical environment. The structural elements cause absorption signals observable whenever the fragments are present in the molecules. These rules can be expressed in two simple forms, i.e., in correlation tables and in so-called additivity rules. Both are successfully used in practical spectroscopy.

The horizontal axis of a correlation table represents the observed characteristic (e.g., frequency), and in the vertical column different structural elements can be found. As correlation tables are used primarily for gaining immediate information, their setup largely depends on the researchers themselves.

Shift limits are not exact, and researchers' definitions of fragments vary as well. As often as not, conditions (e.g., solvent) essential for using the tables successfully are excluded.

Under the given circumstances, the use of correlation tables in computational technology is far less effective than it could be.

In what follows, a mathematical model of correlation tables is specified, its structural principles are explained, and a procedure for the optimum use of correlation tables is described. Although the results are of a general nature, simple correlation tables related to ^1H NMR spectroscopy are taken as examples for better understanding.

CORRELATION TABLES IN THE LITERATURE

As an introductory illustration some ^1H NMR correlation tables have been selected. Table I contains data of some structural elements whose ^1H NMR shifts can all be clearly interpreted.

The methoxy group lucidly exemplifies the nature of the problem. The first investigated correlation table[1] makes a distinction between methyl ester and methyl ether, although the given intervals greatly overlap. Sasaki and his co-workers[2] consider the aromatic and aliphatic methyl ethers entirely distinguishable on the basis of chemical shifts. According to their correlation table the chemical shifts of methyl esters and aromatic methoxy groups are indistinguishable. In the correlation table[3] given in the third column of Table I, the methoxy group is not subdivided. On the other hand, Bible[4] gives different chemical shifts for the three methoxy groups as he regards methoxy groups as distinguishable when found in three differing environments. According to the last column

0095-2338/80/1620-0234$01.00/0