Spectral classification of uranium I energy levels using pattern-recognition techniques

R. V. Lewis and Keith L. Peterson
Department of Chemistry, East Texas State University, Commerce, Texas 75428
(Received 21 April 1986)

Previously unknown uranium I energy levels have been classified according to configuration using pattern-recognition techniques. Four features, the energy level, Landé g factor, quantum number J, and ^{235}U isotope shift, have been used to describe each level. Additionally, a fifth feature, the ^{234}U isotope shift, has been employed to describe most levels. Thirty-seven levels have been assigned with high certainty. In addition, four-feature classification versus five-feature classification is discussed.

I. INTRODUCTION

The use of pattern-recognition techniques as a tool in classifying energy levels has been demonstrated by the work of Peterson *et al.*[1] In this paper we apply similar techniques to the classification of uranium I energy levels. A uranium line list obtained from high-resolution grating measurements between 3100 and 9000 Å contains 92 000 lines of U I and U II.[2] At present there are 360 odd levels and 1240 even levels; the lowest levels of the configurations have been identified. Of these levels, 109 odd levels and 83 even levels have been identified according to their configuration.

In this study we apply pattern-recognition techniques to known energy levels, and use the resulting training to classify according to configuration 24 odd-parity and 13 even-parity levels. This represents an increase of 22.0% for odd levels and 15.7% for even levels. As in previous works,[1,3] the computer package of pattern-recognition techniques, collectively known as ARTHUR, is used.[4]

For pattern-recognition terminology used in this paper, see Appendixes A and B. Three books about pattern-recognition techniques used in this work are listed in the reference section.[5–7]

II. PROCEDURE

The observed energy levels of uranium I are described in Table I. The U I data of Blaise and Radziemski,[2] as well as the isotope-shift data of Engleman and Palmer,[8] are used. Table II lists unknown odd-parity levels with the five-features energy level (cm^{-1}), quantum number J, ^{235}U isotope shift (10^{-3} cm^{-1}), Landé g factor, and ^{234}U isotope shift (10^{-3} cm^{-1}). Table III lists unknown even-parity levels with the same five features.[2] These tables do not list all of the unclassified levels of U I which are given in Ref. 2. This is because of two reasons. First, not all unknowns have experimental values for the four (or five) features which have been found to be most useful in the classification process (see Sec. III). Second, many of the unknowns are at relatively high energy. There are very few classified levels at these high energies, and configurations which lie in these high-energy regions are not well represented. In such a situation training and testing results are not expected to be as good as when configurations are well represented (low-energy regions). This, in turn, leads to decreased confidence in subsequent classifications of unknown levels. The only high-energy unknown levels which have been included in this study lie very close in energy to classified levels. For a further discussion of these matters, see Ref. 1, especially Sec. III.

The present work is accomplished via a three-step procedure. Step 1 includes the training, testing, and classifying of unknown U I energy levels using the first four features included in Tables II and III. These four features were chosen based on the work of Peterson *et al.*,[1] who applied pattern-recognition techniques to Cm I. In that paper it was clearly demonstrated that use of *all* four features was necessary in order to realize the full capabilities of the pattern-recognition technique. It was also shown that for those levels which could not be classified on the basis of the isotope shift alone, the remaining three features often enabled a classification to be made. This illustrates the utility of using all four features simultane-

TABLE I. Even and odd configurations of U I.

Index No.	Configurations	Lowest level (cm^{-1})	Number of levels in interval
Odd			
1	$5f^36d7s^2$	0.000	10
2	$5f^36d^27s$	6249.029	186
3	$5f^47s7p$	22 792.372	140
4	$5f^36d7s8s$	32 857.449	23
5	$5f^46d7p$	34 160.569	23
Even			
1	$5f^47s^2$	7020.710	2
2	$5f^26d^27s^2$	11 502.624	6
3	$5f^37s^27p$	13 463.392	4
4	$5f^36d7s7p$	14 643.867	2
5	$5f^46d7s$	14 839.736	426
6	$5f^36d^27p$	27 886.992	544
7	$5f^36d7s8p$	33 639.562	259

TABLE II. Unknown odd-parity U I levels and data used for classification in this study.

Index No.	Level (cm^{-1})	J	Isotope shift ^{235}U (10^{-3} cm^{-1})	Landé g	Isotope shift ^{234}U (10^{-3} cm^{-1})
1	14 344.522	5	−15	1.01	−56.6
2	14 562.354	5	−275	1.060	34.6
3	19 907.079	4	−310	1.040	−350.4
4	20 195.797	4	−230	0.985	−275.7
5	22 492.342	4	−655	0.92	−655.2
6	22 678.338	3	−495	0.830	−586.6
7	23 753.142	3	−415	0.935	−663.2
8	24 196.267	4	−460	1.050	−542.3
9	24 267.659	3	−585	0.890	−661.6
10	24 401.945	5	−530	0.990	−687.8
11	24 433.473	4	−575	1.010	−664.1
12	24 445.762	5	−465	1.035	−478.9
13	24 539.157	5	−430	1.030	−530.9
14	24 562.002	3	−490	1.080	
15	24 784.732	5	−350	1.020	−427.2
16	25 391.230	4	−310	1.010	
17	25 596.412	4	−310	1.045	−36.7
18	26 489.761	5	−235	1.085	
19	27 020.619	5	−530	1.065	−629.0
20	27 093.227	6	−405	1.090	
21	27 262.210	4	−320	0.905	
22	27 920.942	6	125	0.835	48.5
23	28 122.458	5	−490	1.080	
24	28 168.352	5	−500	1.050	−577.6
25	29 531.499	5	−360	1.025	
26	30 224.832	5	−465	1.040	
27	32 788.523	4	−445	1.040	
28	32 872.074	3	−550	1.030	
29	32 885.596	5	−355	0.875	−456.9
30	33 001.993	3	−490	0.995	
31	33 444.961	4	−380	0.870	−452.6
32	33 564.602	3	−530	1.050	
33	35 559.387	4	−515	1.045	

ously. For a further discussion of these points, the reader is referred to Sec. III of Ref. 1, especially pages 274 and 275. In the case of odd parity, the configurations $5f^36d7s^2$, $5f^36d^27s$, $5f^47s7p$, and $5f^36d7s8s$ are used as categories. For even levels, the configurations $5f^47s^2$, $5f^26d^27s^2$, $5f^37s^27p$, $5f^36d7s7p$, $5f^45d7s$, and $5f^36d^27p$ are used as categories. Good training as well as successful testing is accomplished in both odd and even levels [see Table IV(a)]. Based on these positive results, pattern-recognition methods are then applied to the unknown levels; predictions of these unknowns are listed in Tables V and VI.

Step 2 includes the use of the ^{234}U isotope-shift data of Engleman and Palmer[8] as a fifth feature in the training and classification of both even- and odd-parity unknown energy levels. As in the four-feature case, good training and successful testing and classification of unknown levels is accomplished. Pattern-recognition methods are then applied to the unknown levels for which ^{234}U isotope-shift data are available. Category predictions of these unknowns are listed in Tables VII and VIII.

Table I includes two electron configurations which are not included in the training sets of step 1 and step 2. These configurations are $5f^46d7p$ (odd parity) and $5f^36d7s8p$ (even parity). Each of these configurations are represented in the data sets[2,8] by one member. These configurations are not included in the first two steps because of their one-member status. (Several of the routines of ARTHUR require categories of two members or more.) Step 3 incorporates these configurations into the training sets by using unsupervised learning techniques as well as those techniques used in steps 1 and 2 which do not require categories with two or more members. The results show that no unknown energy levels are classified in the $5f^46d7p$ configuration, and that no unknown energy levels are classified in the $5f^36d7s8p$ configuration. In addition, the inclusion of these singly represented categories in the training sets provides no change in the predictions of the unknowns in either the odd or even levels. On this basis it can be said that there is strong evidence that no odd or even levels belong to a configuration not represented in the training sets.

Plotting techniques are used in order to obtain a visual representation of the clustering of the electron configurations. Although these techniques are not heavily relied upon to determine the electron configuration of the unknown levels, the plots are very useful in indicating both how well a data set is trained, and how well unknown configurations are classified. Figure 1 is a category plot of the training set for the case of odd parity, four-feature

TABLE III. Unknown even-parity U I levels and data used for classification in this study.

Index No.	Level (cm^{-1})	J	Isotope shift ^{235}U (10^{-3} cm^{-1})	Landé g	Isotope shift ^{234}U (10^{-3} cm^{-1})
1	17 070.469	6	−55	0.890	−89.7
2	17 559.322	5	0	1.075	10.2
3	17 893.878	4	10	0.935	17.6
4	18 185.999	4	40	0.830	51.7
5	18 299.500	4	100	0.875	116.0
6	18 383.245	4	−190	0.860	−225.3
7	18 530.851	3	−15	0.560	−42.5
8	18 607.798	4	−145	0.700	−168.1
9	18 749.844	3	300	0.860	338.0
10	18 759.179	6	140	0.915	159.6
11	18 794.831	4	−45	0.855	−46.4
12	19 115.468	3	25	0.895	103.7
13	19 119.780	2	240	0.785	284.6
14	19 668.424	3	0	0.920	−52.8
15	19 783.336	6	−35	0.950	−46.6
16	19 826.672	6	−20	0.940	−26.2
17	19 828.486	2	185	0.780	227.5
18	19 864.520	3	−250	0.625	139.3
19	20 218.830	6	−390	0.830	−465.9
20	20 258.144	3	−50	0.910	−40.5
21	20 306.860	4	190	1.000	219.6
22	20 311.554	5	−210	0.890	−269.1
23	20 391.512	3	−145	0.900	−157.7
24	20 452.805	2	80	0.505	118.5
25	20 464.525	7	−235	0.980	−326.2
26	20 525.394	5	−360	0.970	−407.9
27	20 661.514	6	−500	0.835	−599.5
28	20 943.428	6	−295	0.945	−356.4
29	21 265.094	6	5	0.990	7.5
30	21 407.865	3	−300	0.470	−447.6
31	21 426.485	7	−25	1.030	−38.5
32	22 854.911	5	−400	1.075	−457.6
33	22 862.451	6	−320	0.960	−385.9
34	22 940.652	2	−225	0.630	−247.4
35	23 825.363	4	−170	0.865	−190.1
36	23 841.978	4	30	1.055	47.0
37	25 906.148	5	−235	1.120	−237.6
38	27 829.925	4	−300	1.020	−327.1
39	30 504.894	7	−465	1.080	−533.9
40	30 702.853	8	−300	1.110	
41	31 006.021	10	−390	1.220	−499.1
42	31 296.208	6	−440	1.065	
43	31 442.083	9	−590	1.130	−691.8
44	33 580.727	5	−300	1.035	
45	34 881.927	9	−660	1.115	
46	35 883.271	7	−460	1.09	
47	37 804.907	8	−540	1.055	
48	37 827.073	7	−565	1.085	

TABLE IV. Training and test results for even- and odd-level (a) four-feature training and (b) five-feature training.

Method	Odd levels Training data % correct	Odd levels Test data % correct	Even levels Training data % correct	Even levels Test data % correct
		(a)		
LEAST	95	100	86	80
PLANE	98	98	96	96
KNN	95	94	80	90
PNN	95	94	84	90
STEP	93	88	81	90
		(b)		
LEAST	97	94		
PLANE	100	100	97	100
KNN	96	88	78	80
PNN	96	88	81	90
STEP	93	94	73	90
SICL	98	94	87	80
BACLASS	97	94		

training. Figure 1 includes the unknown energy levels represented as open squares. It is obvious from Fig. 1 that the unknown levels which are grouped into clusters of a known category are most probably members of that category. Furthermore, the complete separation of the four categories in the training set indicates a successful training, and provides for a high confidence in unknown category prediction. [There is no apparent correspondence between the numerical values and units for the energy levels and isotope shifts used in Fig. 1 and those used in the tables. This is because the data in Fig. 1 has been scaled. This process transforms each feature to a mean of zero and a standard deviation of one without destroying separating information. Scaling is necessary to avoid biasing results by data which may be expressed in small (large) units and therefore be numerically quite large (small). For more information see Refs. 1 and 3, and any of Refs. 5–7 and 9.]

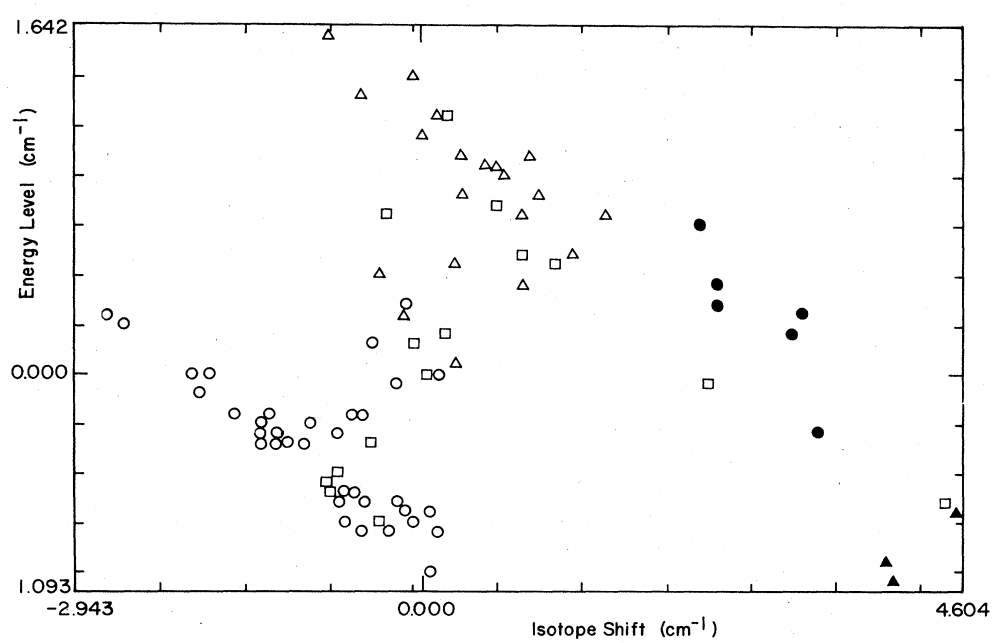

FIG. 1. Plot of the energy level feature (cm^{-1}) vs the ^{235}U isotope shift feature (cm^{-1}) for the training set and unknown configurations. Key to symbols: ○, $5f^36d7s^2$—ground state; △, $5f^36d^27s$; ●, $5f^47s7p$; ▲, $5f^36d7s8s$; □, unclassified (unknown) configurations.

III. RESULTS AND DISCUSSION

In step 1, 24 of the unknown odd levels and 10 of the unknown even levels are classified with high certainty. This means that each pattern-recognition method used in the training procedure predicted the same electron configuration; i.e., there was a unanimous committee vote. Table IX summarizes these results.

In step 2, 11 of the unknown odd levels and seven of the unknown even levels were classified with high certainty. These results are summarized in Table X.

The careful reader may have noticed from Tables VII and VIII that some pattern-recognition methods were used for five-feature training that were not used for four-feature training and vice versa. This arises from the fact that subroutines which are successful in categorizing training data as well as categorizing test data are used to predict unknown configurations. Subroutines which are unsuccessful with the training set and/or unsuccessful in predicting configurations of the test set are not used; i.e., some methods work better for four features and some work better for five features.

TABLE V. Classification results for odd levels based on four-feature training.

Index No.	Classification method[a] LEAST	PLANE	KNN	PNN	STEP
1	1	1	1	1	1
2	2	2	2	2	2
3	2	2	2	2	2
4	2	2	2	2	2
5	3	3	3	3	3
6	3	3	3	3	3
7	3	3	3	3	3
8	3	3	3	3	3
9	3	3	3	3	3
10	3	3	3	3	3
11	3	3	3	3	3
12	3	3	3	3	3
13	3	3	3	3	3
14	3	3	3	3	3
15	3	3	3	3	3
16	3	3	3	3	3
17	3	3	3	3	3
18	3	3	3	3	3
19	3	3	3	3	3
20	3	3	3	3	3
21	3	3	3	3	3
22	3	3	3	3	2
23	3	3	3	3	3
24	3	3	3	3	3
25	3	3	3	3	3
26	3	3	3	3	3
27	3	4	4	4	3
28	3	3	4	4	4
29	3	4	4	4	3
30	3	3	4	4	3
31	3	4	4	4	3
32	3	3	4	4	4
33	4	4	4	4	4

[a]Classification methods are explained in Appendix A. Integers refer to the odd-parity electron configurations of Table I.

TABLE VI. Classification results for even levels based on four-feature training.

Index No.	Classification method[a] LEAST	PLANE	KNN	PNN	STEP
1	3	1	1	1	3
2	2	2	3	3	3
3	2	4	4	4	3
4	3	4	4	4	3
5	2	2	3	3	3
6	3	4	4	4	4
7	4	4	4	4	3
8	4	4	4	4	3
9	2	2	2	2	2
10	3	2	3	3	2
11	3	1	1	1	3
12	2	4	4	4	3
13	2	2	2	2	2
14	2	4	4	4	3
15	3	3	1	1	3
16	3	3	1	1	3
17	3	2	2	2	2
18	4	4	4	4	4
19	4	4	4	4	4
20	3	3	1	1	3
21	2	2	2	2	2
22	3	4	4	4	4
23	3	4	1	1	4
24	4	3	3	3	3
25	3	4	4	4	4
26	3	4	4	4	4
27	4	5	5	5	5
28	4	4	4	4	4
29	4	3	3	3	3
30	5	5	5	5	4
31	3	3	3	3	3
32	3	4	4	4	4
33	4	4	4	4	4
34	4	4	1	1	4
35	4	4	1	1	4
36	3	2	2	2	3
37	3	4	4	4	4
38	4	4	4	4	4
39	5	6	6	6	5
40	4	4	4	4	4
41	4	4	4	4	4
42	5	6	6	6	5
43	5	6	6	6	5
44	5	6	6	6	4
45	6	6	6	6	5
46	5	6	6	6	5
47	6	6	6	6	5
48	6	6	6	6	5

[a]Classification methods are explained in Appendix A. Integers refer to the even-parity electron configurations of Table I.

TABLE VII. Classification results for odd levels based on five-feature training.

Index No.	Classification method[a]						
	LEAST	PLANE	KNN	PNN	STEP	SICL	BACLASS
1	1	1	1	1	1	1	1
2	1	1	2	2	2	1	1
3	2	3	2	2	2	3	2
4	2	2	2	2	2	1	1
5	3	3	3	3	3	3	3
6	3	3	3	3	3	3	3
7	3	3	3	3	3	3	3
8	3	3	3	3	3	3	3
9	3	3	3	3	3	3	3
10	3	3	3	3	3	3	3
11	3	3	3	3	3	3	3
12	3	3	3	3	3	3	3
13	3	3	3	3	3	3	3
15	3	3	3	3	3	3	3
17	2	4	3	3	3	4	3
19	3	3	3	3	3	3	3
22	3	4	4	4	2	4	4
24	3	3	3	3	3	3	3
29	3	4	4	4	3	4	4
31	4	4	4	4	3	4	4

[a]Classification methods are explained in Appendix A. Integers refer to the odd-parity electron configurations of Table I.

TABLE VIII. Classification results for even levels based on five-feature training.

Index No.	Classification method[a]										
	PLANE	KNN	PNN	STEP	SICL	Index No.	PLANE	KNN	PNN	STEP	SICL
1	4	4	3	3	3	22	4	4	4	4	1
2	4	4	4	3	3	23	1	4	4	3	1
3	4	4	4	3	3	24	3	3	3	3	3
4	4	3	4	3	3	25	4	4	4	4	3
5	3	3	3	3	3	26	4	4	4	4	5
6	1	4	4	4	1	27	4	4	5	5	5
7	4	4	4	3	4	28	4	4	5	4	4
8	4	4	4	3	4	29	3	3	3	3	3
9	2	2	2	2	3	30	4	4	5	4	5
10	2	2	3	2	3	31	3	3	3	3	3
11	4	4	3	3	1	32	4	4	4	4	5
12	3	4	3	3	3	33	4	4	4	4	4
13	2	2	2	2	2	34	4	4	1	4	4
14	1	4	3	3	1	35	4	4	1	3	1
15	1	1	1	3	3	36	2	2	3	3	1
16	1	1	1	3	3	37	4	4	4	4	1
17	2	2	2	2	2	38	4	4	4	4	1
18	4	4	4	3	2	39	6	4	5	4	6
19	4	4	4	4	5	41	4	4	4	4	4
20	1	4	3	3	1	43	6	6	5	5	6
21	2	2	2	2	2						

[a]Classification methods are explained in Appendix A. Integers refer to the even-parity electron configurations of Table I.

TABLE IX. Net configuration predictions for unknown levels—four-feature training.

Index No.	Level (cm^{-1})	Configuration	Index No.	Level (cm^{-1})	Configuration
	Odd parity			Even parity	
1	14 344.522	$5f^36d7s^2$	9	18 749.844	$5f^26d^27s^2$
2	14 562.354	$5f^36d^27s$	13	19 119.780	$5f^26d^27s^2$
3	19 907.079	$5f^36d^27s$	18	19 864.520	$5f^36d7s7p$
4	20 195.797	$5f^36d^27s$	19	20 218.830	$5f^36d7s7p$
7	23 753.142	$5f^47s7p$	21	20 306.860	$5f^26d^27s^2$
8	24 196.267	$5f^47s7p$	28	20 943.428	$5f^36d7s7p$
9	24 267.659	$5f^47s7p$	33	22 862.451	$5f^36d7s7p$
10	24 401.945	$5f^47s7p$	38	27 829.925	$5f^36d7s7p$
11	24 433.473	$5f^47s7p$	40	30 702.853	$5f^36d7s7p$
12	24 445.762	$5f^47s7p$	41	31 006.021	$5f^36d7s7p$
13	24 539.157	$5f^47s7p$			
14	24 562.002	$5f^47s7p$			
15	24 784.732	$5f^47s7p$			
16	25 391.230	$5f^47s7p$			
17	25 596.412	$5f^47s7p$			
18	26 489.761	$5f^47s7p$			
19	27 020.619	$5f^47s7p$			
20	27 093.227	$5f^47s7p$			
21	27 262.210	$5f^47s7p$			
23	28 122.458	$5f^47s7p$			
24	28 168.352	$5f^47s7p$			
25	29 531.499	$5f^47s7p$			
26	30 224.832	$5f^47s7p$			
33	35 559.387	$5f^36d7s8s$			

A comparison of four-feature classification versus five-feature classification produces interesting results. Step 1 (four-feature classification) classifies 73% of the unknown odd levels and 21% of the unknown even levels with high confidence. Step 2 (five-feature classification) classifies 33% of the unknown odd levels and 15% of the unknown even levels with high confidence. Step 1 and step 2 in no case predict conflicting configurations for unknown levels. In the case of odd parity, all 11 unknowns classified in step 2 were also classified in step 1 to the same electron configuration. In the case of even parity, four of the seven unknowns classified by step 2 to a particular configuration are classified to that same configuration by step 1. Three unknown energy levels are classified to configurations in step 2 which were not classified in step 1 with high confidence. It should be noted that in this work as well as in all previous work,[1,3] even-parity levels are more difficult to train successfully; also, assignment of an unknown energy level to a particular electron configuration by unanimous committee vote occurs less often in the even-parity case than in the odd-parity case. Most interesting is the fact that four-feature training classifica-

TABLE X. Netconfiguration predictions for unknown levels—five-feature training.

Index No.	Level (cm^{-1})	Configuration	Index No.	Level (cm^{-1})	Configuration
	Odd parity			Even parity	
1	14 344.522	$5f^36d7s^2$	5	18 299.500	$5f^36s^27p$
7	23 753.142	$5f^47s7p$	13	19 119.780	$5f^25d^27s^2$
8	24 196.267	$5f^47s7p$	17	19 828.486	$5f^26d^27s^2$
9	24 267.659	$5f^47s7p$	21	20 306.860	$5f^26d^27s^2$
10	24 401.945	$5f^47s7p$	24	20 452.805	$5f^36s^27p$
11	24 433.473	$5f^47s7p$	33	22 862.451	$5f^36d7s7p$
12	24 445.762	$5f^47s7p$	41	31 006.021	$5f^36d7s7p$
13	24 539.157	$5f^47s7p$			
15	24 784.732	$5f^47s7p$			
19	27 020.619	$5f^47s7p$			
24	28 168.352	$5f^47s7p$			

tions and five-feature classifications are very similar, leading to higher confidence levels in the classification of atomic energy levels.

IV. CONCLUSIONS

Pattern-recognition techniques have been used successfully in the classification of unknown uranium I energy with high confidence. This classification can be quite helpful in the further analysis of the neutral uranium spectrum. Pattern-recognition techniques may be used successfully with other complex atoms as noted by this paper and others.[1,3] The existence of representative data sets is essential for the proper use of pattern-recognition methods; hence conventional classification methods will continue to be important.

ACKNOWLEDGMENTS

We thank Bradley Hurst for his efforts in implementing ARTHUR on our computer system. We thank Mike Cagle, Computer Center Director, for his technical assistance. This work was partially supported by a grant from the Organized Research Fund of East Texas State University.

APPENDIX A: PATTERN-RECOGNITION TECHNIQUES

Method	Description
LEAST	Least-squares regression—classifies by performing a least-squares multilinear regression using all features, utilizing the generalized inverse method.[3]
PLANE	Hyperplane separation—classifies on the basis of a generated linear discriminant function.[9]
KNN	K nearest neighbor—predicts categories on the basis of K nearest neighbors, where K is 1, 3–10. A pattern belongs to that category which is represented most often among its K nearest neighbors.[3]
PNN	Percentage nearest neighbors—predicts categories on the basis of a given percentage of nearest neighbors. The routine is very similar to K nearest neighbors.[3]
STEP	Stepwise multilinear regression—classifies by performing a stepwise multilinear regression. Features used in the regression are determined by their contribution to the total variance of the data.[3]
SICL	Statistical isolinear multicategory analysis (SIMCA) classification routine—classifies both the training set and unknown data vectors on the basis of how well a model for a category fits the data.[9]
BACLASS	Bayes classification rule—classifies by performing an approximate multivariate Bayes-rule classification.[3]

APPENDIX B: PATTERN-RECOGNITION TERMINOLOGY (REF. 9)

Term	Definition
Category	One of the groups of objects studied in the classification analysis algorithms; i.e., an electron configuration.
Classification	Assignment of an energy level to a particular electron configuration via a particular classification algorithm or algorithms.
Pattern	A member of the data set which may be known or an unknown; i.e., an energy level.
Training	Introduction of the data set to the classification algorithms of ARTHUR, thereby creating the classification rules.
Training set	The data set patterns used to develop the classification rules.
Test	Introduction of an evaluation data set to test the classification rules. Use of a particular algorithm is dependent on the successful evaluation of the test set.
Test set	The data patterns for which the electron configuration is known. The test set is an evaluation data set.
Unknown	An energy level whose electron configuration is unknown.
Supervised learning	Training in which samples are known to be composed of specified categories.
Unsupervised learning	Training via cluster analysis where the goal is the discovery of systematic behavior (i.e., categories) in the data.

[1]K. L. Peterson, D. L. Anderson, and M. L. Parsons, Phys. Rev. A **17**, 270 (1978).
[2]J. Blaise and L. J. Radziemski, Jr., J. Opt. Soc. Am. **66**, 644 (1976).
[3]K. L. Peterson and M. L. Parsons, Phys. Rev. A **17**, 261 (1978).
[4]D. L. Duewar, J. R. Koskinen, and B. R. Kowalski, computer package ARTHUR, obtained from B&B Associates, P. O. Box 85505, Seattle, Washington 98105.
[5]K. Varmuza, *Pattern Recognition in Chemistry* (Springer-Verlag, Berlin, 1980).
[6]J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles* (Addison-Wesley, London, 1974).
[7]H. C. Andrews, *Introduction to Mathematical Techniques in Pattern Recognition* (Wiley, New York, 1972).
[8]R. Engleman, Jr. and B. A. Palmer, J. Opt. Soc. Am. **70**, 308 (1980).
[9]ARTHUR81 *User's Manual for the 1981 Release Version,* distributed by B&B Associates (Infometrix, Seattle, 1981).