kinetics. At a triphosphate concentration of $10^{-3}M$, little or no interference was observed.

Most metals can be eliminated as possible interferences by precipitating them as the hydroxide or by running the solution to be determined through a cation-exchange column to remove these metals (*14*).

Polyamines, complexed by nickel, are known to react very rapidly with cyanide and will catalyze reactions of nickel(II)-aminopolycarboxylates (*28*). Such catalytic agents could be used to speed the rate of reaction of the slower aminopolycarboxylates.

The formation of mixed cyano complexes can be a source of interference in these kinetic methods. For example, the mixed $NiEDTA(CN)^{3-}$ complex which absorbs at 267 nm had to be taken into account in the determination of NTA in commercial EDTA. However, the kinetics data can be treated to minimize the interference of mixed complex formation.

The reported results demonstrate the applicability of this reaction rate method for the determination of aminopolycarboxylate mixtures and single components. The method can be accurate for mixtures of ligands in solution to within ±5–10% at the $10^{-6}M$ level. The sensitivity of this method is on the order of $4 \times 10^{-8}M$ or 10 ppb of Na_3NTA in water.

A previous article (*8*) discussed the variation of rate constants and ϵb (molar absorptivity × cell length) on the determination of a two-component mixture using the REDKAN method. From that discussion, it seems unlikely that a 5–10% error could result from an error in the observed rate constants and ϵb. A larger source of error in these determinations is probably due to instrumental noise. For the Durrum-Gibson stopped-flow, the noise ranges from ±0.1 to 0.3% transmittance in the UV region. For reactions with small changes in per cent transmittance, this can be a significant source of error.

With the selection of optimum experimental conditions, the rapid analysis of a large number of aminopolycarboxylic acids can be determined using a kinetic method. An exception to this would be the MIDA, IDA mixture if these ligands were present as their mono nickel complexes. However, the procedure in which excess nickel hydroxide is precipitated will convert these ligands to their bis complexes with nickel. These bis complexes are easily distinguished kinetically (*11*).

Table V. Determination of NTA in Water in ppb of Na_3NTA

Conditions: Stopped-flow, 267 nm, 25.0 °C, pH 11, 0.1M $NaClO_4$, 0.1M NH_3, $5 \times 10^{-4}M$ Cyanide

	NTA		
	Added	Found	Std dev
O.R.[a]	13	26	10
Cinn.[b]	13	19	3
O. R.	25	38	2
Cinn.	25	30	11
O. R.	101	90	5
Cinn.	101	86	6

[a] O. R. = Ohio River water treated by ion-exchange method (*14*).

[b] Cinn. = City of Cincinnati water treated by ion-exchange method (*14*).

ACKNOWLEDGMENT

We would like to thank John Haberman for providing us with the water samples.

RECEIVED for review April 7, 1972. Accepted July 25, 1972. This research was sponsored by the Air Force Office of Scientific Research under AFOSR Grant 71-1988 and 1212-67.

(28) G. B. Kolski and D. W. Margerum, *Inorg. Chem.*, **8**, 1211 (1969).

Computerized Pattern Classification of Strongly Overlapped Peaks in Stationary Electrode Polarography

L. B. Sybrandt and S. P. Perone

Department of Chemistry, Purdue University, Lafayette, Ind. 47907

A computerized pattern classifier has been evaluated for the detection of doublet peaks in stationary electrode polarography. Parameters were devised that could be obtained from the zero-, first-, and second-derivative polarograms. Only 22 parameters of the 133 investigated were necessary for the successful classification of the theoretical, reversible, uncomplicated systems used in this study. Overlap conditions were so severe that visual, subjective, human interpretation of the derivative data was not possible. Prediction accuracy of about 90% or better was realized for polarograms in the range of 6–14 mV peak separation, 1–3 *n*-value, and 20:1–1:1 peak height ratios.

INSTRUMENTAL ANALYSIS of a sample usually results in the generation of a two-dimensional pattern of data. Essential to interpretation of the data is knowledge of the number of chemical species contributing to the pattern. When dealing with mixtures, overlap of individual data patterns may be so severe that the multiplicity cannot be readily obtained. Common measurement techniques where this problem may exist are mass spectrometry, spectroscopy, chromatography, and electrochemistry. The need for chemical separation or different analytical measurements can be determined if the chemical purity of an observed signal is known. Speed of decision may be an important factor. The work to be presented here demonstrates the usefulness of computerized pattern classification for the distinction between singlet and doublet peaks in stationary electrode polarography (SEP). The principles discussed in this application should be applicable to data patterns obtained from other measurement techniques.

The existence of more than one species contributing to an observed signal can be determined in many ways. A coordinate transformation is usually used to make the data pattern more amenable to human interpretation. For a continuous positive-going peak-shaped curve, one conventionally counts

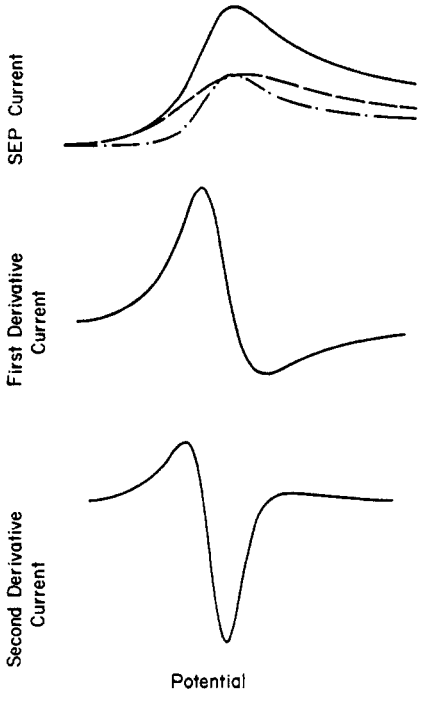


Figure 1. Theoretical stationary electrode polarogram and its derivatives for a two-component system. Conditions: 1 : 1 peak heights, ΔE_p = 12 mV, 1 and 2 *n*-values

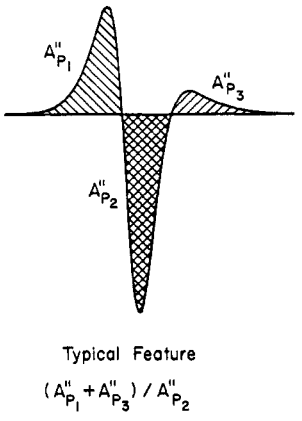


Figure 2. Typical area features extracted from second-derivative polarogram

the number of negative peaks in the second derivative. A trained observer can detect distortions in the data or its derivatives for more severe overlap conditions. Beyond the limit of visual interpretation, it may be possible to use simultaneous equations (*1–4*), curve-fitting (*5–7*), calibration plots of second-derivative peak parameters (*8*), width parameters (*9, 10*), moment analysis (*11*), self-modeling curve resolution (*12*), and characteristic vector analysis (*13*). Learning machine pattern classification methods have been reported for qualitative (*14*) and semiquantitative (*15*) analysis of mixtures. While these methods provide more information than multiplicity, they are not applicable to a problem having the following conditions: observed peak location, peak height ratio, peak separation, and component curve shapes can vary within predefined but generous limits; a mathematical description of all possible singlet curve shapes is unknown; only a single unknown curve is available for multiplicity determination. Although the theoretical SEP data used in this study are generated from a known mathematical function, this knowledge is not essential to the solution presented.

In stationary electrode polarography, an observed peak may be due to a single species, or may be the result of overlapping reduction currents from two or more components in solution. The peak potential has been found to vary as much as 8 mV under supposedly identical conditions (*14*). This factor can cause errors in a simultaneous equation approach even if the data were otherwise ideal. Also, many curve shapes have to be considered since the number of electrons transferred (*n*-value) is rarely measured to be an exact integer, even for reversible uncomplicated systems. The previously reported learning machine approach to qualitative analysis (*14*) is not practical if a multitude of possible components must be considered because it is necessary to train a weight vector for each component. Second-derivative peak detection is limited to about 10:1 concentration ratios at a peak separation of 40 mV (*5*). In the solution proposed here, peak overlap conditions too severe for visual subjective detection (second-derivative) are considered. Curve parameters, *i.e.*, *features*, are defined relative to the observed peak potential and peak current. Theoretical SEP curves are generated for one- or two-component reversible uncomplicated systems for varying conditions of peak height ratio, peak separation, and *n*-value. The defined features are extracted from the zero-, first-, and second-derivative curves for each condition. Pattern recognition methods, with these features constituting a pattern, are used to discern the doublet from singlet peaks. Those features found to be least useful to the classification problem are eliminated.

There are four advantages of the pattern recognition approach to this problem. First, many parameters can be investigated without excessive human effort. The unknown relationships among parameters are empirically determined by the classifier. Second, multidimensional relationships can be easily handled. Unlike human interpretation, the pattern classifier is not limited to two or three dimensions. Third, a pattern classifier has the ability to predict. Variations which were not considered in the formulation of the problem may be interpretable by the classifier. Finally, it may be possible to discover fundamental parametric correlations empirically, once features important to the classification problem are known.

(1) H. A. Barnett and A. Bartoli, ANAL. CHEM., **32,** 1153 (1960).
(2) Y. Israel, *Tantala*, **13,** 1113 (1966).
(3) D. F. DeTar, ANAL. CHEM., **38,** 1794 (1966).
(4) R. A. Nesbitt, R. L. Litle, and D. Isaacs, Paper presented at the Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy, Cleveland, Ohio, March 1970.
(5) W. F. Gutknecht and S. P. Perone, ANAL. CHEM., **42,** 906 (1970).
(6) A. H. Anderson, T. C. Gibb, and A. B. Littlewood, *J. Chromatogr. Sci.*, **8,** 640 (1970).
(7) A. H. Anderson, T. C. Gibb, and A. B. Littlewood, ANAL. CHEM., **42,** 434 (1970).
(8) E. Grushka and G. C. Monacelli, *ibid.*, in press.
(9) P. D. Klein, *Separ. Sci.*, **1,** 511 (1966).
(10) V. Cejka, M. H. Dipert, S. A. Tyler, and P. D. Klein, ANAL. CHEM., **40,** 1614 (1968).
(11) E. Grushka, M. N. Meyers, P. D. Schettler, and J. C. Giddings, *ibid.*, **41,** 889 (1969).
(12) W. H. Lawton and E. A. Sylvestre, *Tech. Conf. Trans. ASQC*, 1969, pp 149–160.
(13) J. L. Simonds, *J. Opt. Soc. Amer.* **53,** 968 (1963).
(14) L. B. Sybrandt and S. P. Perone, ANAL. CHEM., **43,** 382 (1971).
(15) L. E. Wangen and T. L. Isenhour, *ibid.*, **42,** 737 (1970).

PATTERN CLASSIFIER CONCEPT

Trainable pattern classifiers, a subclass of learning machines (*16*), are geometrical in concept. The *d* pieces of information, *i.e.*, *d features*, constituting a pattern can be plotted as a point in *d*-dimensional pattern space. It is assumed that points having a similar property occupy a similar region of pattern space. The object of pattern classification is to find boundaries which discriminate among these spacial regions. [Background information and previous applications of pattern recognition in chemistry can be found in references (*14–36*).]

Assuming a linear *d*-dimensional boundary separates pattern points into two categories, the side of the decision plane where a given point, *J*, resides is given by the sign of s_J in the equation

$$s_J = \sum_{I=1}^{d} w_I x_{I,J} + w_{d+1} \quad (1)$$

where $x_{I,J}$ is the magnitude of the *I*th feature of pattern *J* and w_I is the weight associated with the *I*th feature. The set of *d* features are the same for all patterns. The $d + 1$ weights represent a $(d + 1)$-dimensional weight vector and are determined by an error-correction procedure discussed previously (*14*). The convention used here for classification of patterns is

$$s_J > 0 \quad \text{two-component peak}$$
$$s_J < 0 \quad \text{one-component peak} \quad (2)$$

THE SEP DATA

Theoretical stationary electrode polarograms for uncomplicated reversible systems were calculated utilizing the $\chi(at)$ current function (*37*). The SEP current at a planar electrode is given by

$$i = nFAC_o^* \sqrt{\pi D_o a} \chi(at) \quad (3)$$

where $a = nFV/RT$, n is the charge transfer, A is the electrode area, C_o^* is the bulk concentration of the oxidized species O having a diffusion coefficient D_o, V is the rate of potential scan, t is time, and R, T, and F have their usual significance. The term at is dimensionless and proportional to potential (*37*). The $\chi(at)$ current function used here was calculated in this laboratory over a wide potential range (*38*). The nature of the calculation provides values for $\chi[(E - E_{1/2})n]$ at nonuniform increments of $(E - E_{1/2})n$, where E is the electrode potential and $E_{1/2}$ is the polarographic half-wave potential. Since peak potential, E_p, is of interest here, these data were converted to $\chi[(E - E_p)n]$ (*37*). These data were condensed because of computer core-size limitations. Linear extrapolation between nearest points provided values for $\chi[(E - E_p)n]$ at 2-mV intervals from +162 mV to −420 mV. These data were stored as a table in computer memory after normalizing to 20,000 at $(E - E_p)n = 0$ and rounding all values to the nearest integer. Linear extrapolation was again used to obtain intermediate points. Values for $\chi[(E-E_p)n]$ more anodic than $(E-E_p)n$ = 162 mV were calculated by the exponential approximation

$$\chi[(E - E_p)n] = 99702.4e^{-0.036942(E-E_p)n} \quad (4)$$

The constants in Equation 4 were determined by a six-point linear least squares fit to the tabulated data from +154 mV to +164 mV.

The SEP current was calculated according to the equation

$$i = H_1\chi[(E - E_{p_1})n_1] + H_2\chi[(E - E_{p_2})n_2] \quad (5)$$

where H_1 and H_2 are the peak heights and E_{p_1} and E_{p_2} are the peak potentials of components 1 and 2. The zero-derivative polarograms were digitized at 2-mV intervals relative to the observed peak potential and normalized to a peak height of 1.0. The observed peak potential and peak current were determined at 0.1-mV resolution by digitizing at this higher data density near the peak.

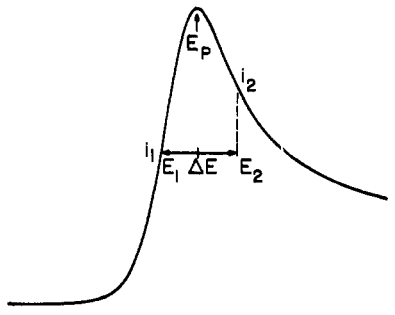
The first- and second-derivative currents, i' and i'', respectively, were obtained by convoluting the SEP either once or twice with the 11-point least squares cubic-quartic first-derivative convoluting function of Savitzky and Golay (*39*). The 11-point convolute was used to avoid spurious peaks occurring at the junction between tabulated $\chi(at)$ values and those calculated according to Equation 4. Of the first- and second-derivative convolutes given in reference (*39*), only the cubic-quartic and quartic-quintic convolutes fit the theoretical data of Perone and Mueller (*40*). Because of the method of digitization used here, the derivative peaks were located at 2-mV accuracy.

A typical two-component SEP and its derivatives are shown in Figure 1. The individual peaks are separated by 12 mV and have equal peak heights with *n*-values of 1 and 2 electrons. Visual detection of the multiplicity is not possible.

THE FEATURES

The features as used in this work are *d* real single-valued functions of the data. They are applicable to any continuous analytical signal where each mixture component contributes a single peak-shaped curve. Feature *I*, whose magnitude is $x_{I,J}$ in Equation 1 for polarogram *J*, is obtained from either the zero-, first-, or second-derivative data, with the exception

(16) N. J. Nilsson, "Learning Machines," McGraw-Hill, New York, N. Y., 1965.
(17) "Computer and Information Sciences-II," J. T. Tou, Ed., Academic Press, New York, N. Y., 1967.
(18) J. R. Slagle, "Artificial Intelligence: The Heuristic Programming Approach," McGraw-Hill, New York, N. Y., 1971.
(19) R. G. Casey and G. Nagy, *Sci. Amer.*, **224** (4), 56 (1971).
(20) T. L. Isenhour and P. C. Jurs, ANAL. CHEM., **43** (10), 20A (1971).
(21) P. C. Jurs, *ibid.*, **42**, 1633 (1970).
(22) *Ibid.*, **43**, 22 (1971).
(23) *Ibid.*, p 1812.
(24) P. C. Jurs, *Appl. Spectrosc.*, **25**, 483 (1971).
(25) P. C. Jurs, B. R. Kowalski, and T. L. Isenhour, ANAL. CHEM., **41**, 21 (1969).
(26) P. C. Jurs, B. R. Kowalski, T. L. Isenhour, and C. N. Reilley, *ibid.*, p 690.
(27) *Ibid.*, p 1949.
(28) *Ibid.*, **42**, 1387 (1970).
(29) B. R. Kowalski, P. C. Jurs, T. L. Isenhour, and C. N. Reilley, *ibid.*, **41**, 695 (1969).
(30) *Ibid.*, p 1945.
(31) B. R. Kowalski and C. N. Reilley, *J. Phys. Chem.*, **75**, 1402 (1971).
(32) B. R. Kowalski and C. F. Bender, Lawrence Radiation Lab., Livermore, Calif., *Preprint* **UCRL-73590**, Nov. 1971.
(33) L. E. Wangen, N. M. Frew, T. L. Isenhour, and P. C. Jurs, *Appl. Spectrosc.*, **25**, 203 (1971).
(34) L. E. Wangen, N. M. Frew, and T. L Isenhour, ANAL. CHEM., **43**, 845 (1971).
(35) G. Schroll, A. M. Duffield, C. Djerassi, B. G. Buchanan, G. L. Sutherland, E. A. Feigenbaum, and J. Lederberg, *J. Amer. Chem. Soc.*, **91**, 7440 (1969).
(36) P. W. Neurath, B. L. Bablouzian, T. H. Warms, R. C. Serbagi, and A. Falek, *Ann. N. Y. Acad. Sci.*, **128**, 1013 (1966).
(37) R. S. Nicholson and I. Shain, ANAL. CHEM., **36**, 706 (1964).
(38) S. P. Perone and P. E. Reinbold, Purdue University, Lafayette, Ind., unpublished work, 1967.
(39) A. Savitzky and M. J. E. Golay, ANAL. CHEM., **36**, 1627 (1964).
(40) S. P. Perone and T. R. Mueller, *ibid.*, **37**, 2 (1965).

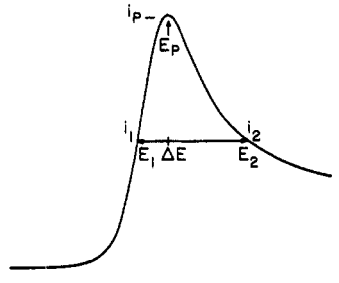


Typical Features

$\Delta i = i_2 - i_1$ at $\Delta E = 8,16,24,32$ mV about E_P

Ifunction = $i_2 * i_1 / (i_2 - i_1)^2$

Figure 3. Typical current features extracted from stationary electrode polarogram



Typical Features

$\Delta E = E_2 - E_1$ at 70,75,80,85 % i_P

Efunction = $(E_2 - E_P) * (E_P - E_1) / [(E_2 - E_P) - (E_P - E_1)]^2$

Figure 4. Typical potential features extracted from stationary electrode polarogram

of cross-derivative features to be discussed below. For other signal types, different features can be envisioned. The features chosen here are arbitrary and not meant to be all-inclusive. The thought was to include enough possibilities so that the classifier has a good chance for success.

Typical features extracted from the data are shown in Figures 2–4. Figure 2 demonstrates four area features obtained from the second-derivative. Note the peak numbering scheme used. When no subscript number appears, the single zero-derivative peak is indicated. The use of primes (none, single, double) will indicate the derivative order. Current and potential features are shown in Figures 3 and 4 for zero-derivative data. The terms Ifunction and Efunction, defined in the figures, were devised simply because they seemed related to skew. None of the peaks were found to be perfectly symmetrical.

Table I provides a general description of the types of features extracted from the data. These include shape, peak current, and peak potential features. There are a total of six peaks to be considered, and when applicable (*e.g.*, Efunction), a feature type was usually applied to each peak. Cross-derivative parameters, such as i'_{p_1}/i''_{p_2}, were included. All features were designed such that an absolute potential axis was avoided. The arbitrary nature of feature definition should be emphasized. A total of 133 features was devised for this work.

With no more consideration than the above, many features could be investigated by the pattern recognition approach without significant human effort. It was assumed that the data were smooth in order to simplify programming requirements. Linear extrapolation between data points was performed when necessary to obtain the value for a feature.

FEATURE ADJUSTMENT

Each feature has its own range and distribution of values, and these can vary greatly from feature to feature. Useful spacial information in coordinates that have narrow distributions may be masked by the relatively larger distributions of other features. This characteristic has been indicated in earlier work with combined infrared and mass spectral data (*27*). Therefore, some feature adjustment prior to training is desirable.

Feature adjustment techniques have been reported previously. The Fourier transformation has been used (*23*, *31*, *33*), but it does not appear consistent with the goal discussed above. Furthermore, some information from all of the original features is contained in each Fourier transform feature (*33*). Thus, the utility of using fewer features is diminished since values for all features have to be obtained to get the same transformed pattern. Jurs (*22*) has investigated the effects of transformations such as the logarithm on the training of mass spectral data. These transformations require that a feature contain all positive or all negative values, but this was not the case for features used here. Jurs *et al.* (*27*), adjusted the sum of infrared intensities to equal the sum of mass spectral intensities in the combined study, but this does not lessen the distribution dissimilarity between features. Kowalski and Bender (*32*) scaled their data by *autoscaling*, to be discussed below, which proved very valuable for the work presented here.

Two feature adjustment procedures were investigated in this work. They both are applied independently to each feature over the whole training set for all features. The first method used was *range-fixing*. The value, $x_{I,J}$, of feature I for pattern J is adjusted to $x'_{I,J}$ by the equation

$$x'_{I,J} = a_I x_{I,J} + b_I \tag{6}$$

where a_I and b_I are constants based on the largest and smallest values of $x_{I,J}$ in the training set, and can be calculated to fix an arbitrary range for feature I. The second adjustment scheme was *autoscaling* (*32*), which converts each coordinate into sigma units deviation from the mean, as given by the equation

$$x'_{I,J} = (x_{I,J} - \bar{x}_I)/\sigma_I \tag{7}$$

where $\bar{x}_I$ is the mean and σ_I is the standard deviation of $x_{I,J}$ for feature I in the training set. Autoscaling results in zero mean and unit variance for each feature. Autoscaling was used for the bulk of this work and will be assumed unless stated otherwise.

FEATURE ELIMINATION

Feature elimination is an important aspect of this work. Features were chosen arbitrarily with little appreciation of their worth. It is necessary to have a means by which features with little utility can be discarded. A combination of two previously reported techniques (*21*,*26*) was used. The first technique required the use of two weight vectors. One weight vector, termed WV+1, was initialized with all weights equal to +1. The other vector, WV−1, was initialized to

−1. Each weight vector was trained independently. Features were then eliminated whose trained weights had opposite signs (*21*). This process, here termed *weight-sign*, was repeated (retraining with initialization) until no more features were eliminated. The second method, termed *weight-magnitude*, was then implemented. Weights for trained vectors were ordered according to increasing absolute magnitude for **WV** + 1 and **WV** − 1 independently. The five features having the lowest combined weight order were discarded (*26*), and **WV** + 1 and **WV** − 1 were retrained in the lower dimension after initialization to +1 and −1. This second method was repeated until classification performance became degraded.

Weight-sign and weight-magnitude feature elimination may not be wholly independent. Most features discarded on the basis of opposite signs had smaller weights than those retained. Weight-sign elimination was used first because features are eliminated rapidly without training degradation and the features retained are not influenced by the subjective choice of the number of features to be discarded, as occurs with weight-magnitude elimination. However, features are retained by the weight-sign method that are not necessary for successful classification, thus necessitating the use of weight-magnitude elimination.

LEARNING MACHINE PROCEDURE

The learning machine experiments to be discussed were performed with a Hewlett-Packard 2116A 16-bit digital computer with 8K-word core memory. Peripheral devices included an ASR/35 Teletype, high speed paper tape punch, photoreader, computer controlled storage oscilloscope, and a Hewlett-Packard Model 2020 magnetic tape unit. All programs were written in Hewlett-Packard FORTRAN and BASIC with a few assembly language I/O subroutines. These programs are available upon request.

Three separate FORTRAN programs were used to implement the pattern classifier. The first program constructed theoretical SEP data and extracted the features. The tabulated $\chi(at)$ current function was loaded *via* paper tape. The parameters for a polarogram, previously punched on paper tape with a BASIC program, were then loaded *via* the reader. These parameters were peak height, *n*-value, and peak location information for each of the two components. The SEP and derivatives were displayed on the storage scope for visual inspection. The extracted features were then stored on magnetic tape without feature adjustment, followed by the six parameters. This procedure was repeated for the desired number of patterns. A training program was then used to determine scaling factors, to adjust each pattern as it was read from the magnetic tape, and to train **WV** + 1 and **WV** − 1 concurrently. The number of error corrections which occurred during an iteration was printed on the Teletype before the next iteration commenced. Weight vectors were punched at five-iteration intervals. Training was terminated when: the vectors were perfectly trained; a preset number of iterations was reached; convergence was so slow that little could be gained from further iterations.

A third program utilized the weight vectors for recognition and prediction. For this work, *recognition* will designate the ability to classify training set patterns, and *prediction* will refer to the classification of patterns with all parameters (peak height ratio, peak locations, peak separation, *n*-value) different than those used in training. A pass through the training set is an *iteration;* and each time the weight vectors are trained for feature elimination purposes constitutes a *cycle*. An *undecided* response occurs when **WV** + 1 and **WV** − 1 classify a pattern in opposite categories. An *error* results when both vectors incorrectly classify the pattern. Per cent *accuracy* is calculated on the basis of correct decisions relative to the number of decisions made. The response to any pattern is undecided for **WV** + 1 and **WV** − 1 before the vectors are trained. The combined use of both weight vectors for classification amounts to a type of committee machine (*16*) where the only correct response is a unanimous vote.

Table I. Types of Features Obtained from 0, 1st and 2nd Derivative Polarograms (133 Features Total).

Shape features	Peak current features[a]	Peak potential features[a]
Δi at ΔE	Peak height ratios	Relative E_p, $E_{1/2}$ ratios
ΔE at $N\%$ i_p		
Efunction	Peak height differences	E_p, $E_{1/2}$ differences
Ifunction		
Peak area ratios		

[a] These categories include cross-derivative features.

RESULTS AND DISCUSSION

Arbitrary limits were imposed on the parameters used to construct the theoretical SEP data. These limits were 20:1 peak height ratios, 8 to 12 mV peak separations, and *n*-values of 1 to 3. Visibly discernible distortions in the second-derivative data occurred frequently with separations greater than 12 mV. All training sets contained 20:1, 10:1, 1:1, 1:10, 1:20 peak height ratios and exactly 8, 10, 12 mV peak separations. The observed peak maximum for doublets and singlets was normalized to 1.0 for the zero-derivative. The variation in training sets resulted from the number of different electron transfer characteristics allowed. Assuming a relative potential axis, the maximum number of two-component polarograms is given by RST^2, where R is the number of peak ratios, S is the number of separations, and T is the number of *n*-values allowed. All possible two-component combinations of the stated conditions ($15T^2$) were used in training except for a few mixtures where a current in the Δi feature (Figure 3) crossed zero for the derivative peaks.

In initial work, certain training sets were used that were not sufficiently representative to provide good prediction. The simplest training set consisted of *n*-values of 1, 2, and 3 for both one- and two-component polarograms. Features were discarded on weight-sign basis only. With the range fixed at 1 to 1000 (Equation 6), training terminated after 3 cycles, requiring 147 **WV** + 1 and 35 **WV** − 1 iterations for complete convergence. Ninety-four features were retained. With the range fixed at −500 to 500, convergence occurred in only 4 **WV** + 1 and 7 **WV** − 1 iterations, terminating after 6 cycles with the retention of 24 features. After autoscaling these 24 features according to Equation 7, weight-sign elimination continued to discard features until 16 remained. Training required only 3 **WV** + 1 and 4 **WV** − 1 iterations with these features. Prediction was especially poor for one-component patterns. Patterns with only 0.01 *n*-value difference from the training set were incorrectly classified. This was also observed when *n*-values of 1.0, 1.4, 1.8, 2.2, 2.6, and 3.0 were used in the training set for both one- and two-component polarograms. A third training set consisted of *n*-values of 1, 2, and 3 for the two-component and 1.00, 1.02, . . ., 2.98,

Table II. Training Set I Conditions

Two-component patterns:

1.0, 1.4, 1.8, 2.2, 2.6, 3.0 n-values
8, 10, 12 mV peak separations
20:1, 10:1, 1:1, 1:10, 1:20 peak heights

One-component patterns:

1.00, 1.02, . . ., 2.98, 3.00 n-values
Random peak locations

Possible two-component patterns = 540
Two-component patterns used = 533
One-component patterns used = $3 \times 101 = 303$

Sampling conditions:

0.1 mV resolution for observed E_p
All potentials evaluated relative to observed E_p at 2-mV intervals
Observed peak height normalized to 1.0

Table III. Prediction Set I Conditions

Two-component patterns:

1.000 to 3.000 n-values, random
8.000 to 12.000 mV peak separations, random
20.000:1 to 1:1 peak heights, random (300)
1:1 to 20.000:1 peak heights, random (300)

One-component patterns:

1.01, 1.03, . . ., 2.97, 2.99 n-values (100)
1.001 to 2.999 n-values, random, with non-zero last digit (200)
Random peak locations

Two-component patterns used = 600
One-component patterns used = 300
Sampling conditions same as Table II

3.00 n-values for the one-component polarograms (236 total patterns). In this case, weight vectors did not converge after 373 iterations with range-fixing (−500 to 500). When autoscaling was used to adjust the features, the weight vectors converged. After the sixth and terminating cycle, 49 features remained (weight-sign basis). For a prediction set with different n-values than those used in training, one-component patterns were classified at better than 90% accuracy, with less than 5% undecided responses, but less than 75% of the two-component patterns were correctly predicted, with about 13% undecided responses. Based on these results, Training Set I was devised to obtain better prediction ability.

The conditions for Training Set I are shown in Table II. Six equally spaced n-values, three peak separations, and five peak height ratios were used for two-component polarograms. All but 7 (due to the incorrect current feature discussed earlier) of the 540 possible two-component parameter combinations were included in the training set. Each of the 101 one-component polarograms was generated three times. The true peak of each polarogram was located randomly within a range of 0.1 mV. Thus, the three polarograms having the same n-value were not likely to be located identically.

Conditions for the parameters used to generate Prediction Set I are described in Table III. The individual curves of two-component polarograms had randomly generated n-values between 1.000 and 3.000. These two curves also had randomly generated peak locations restricted to peak separations between 8 mV and 12 mV. Peak heights between 20 and 1 were randomly generated for the anodic component while keeping the cathodic component peak height constant, and vice versa, to assure an even distribution of peak height ratios. For one-component patterns, 100 n-values were used that were midway between those of Training Set I. Another 200 one-component patterns were generated with random n-values in a manner to prevent duplication of Training Set I patterns and the 100 non-random n-values of this prediction set. Random peak locations were also used for the one-component polarograms. Sampling conditions were the same as Training Set I.

Table IV. Recognition of Training Set I. Features Discarded on Weight-Sign Basis Only

Cycle No.	No. of features	No. of iterations	One-component Undecided, %	One-component Accuracy, %	Two-component Undecided, %	Two-component Accuracy, %
1	133	50	19.5	98.8	10.1	94.1
		100	13.2	95.4	8.8	96.1
2	75	50	14.9	97.7	6.8	95.8
3	66	50	12.5	98.9	8.1	96.3
4	59	50	5.6	90.9	3.6	96.9
5	57	5	6.6	89.7	5.1	87.1
		10	9.6	94.5	5.6	91.8
		15	8.9	96.7	5.1	92.5
		20	7.6	95.0	3.9	93.4
		25	7.3	96.8	3.0	93.6
		30	7.6	97.5	3.9	94.5
		35	9.6	97.4	4.9	95.3
		40	6.3	96.5	4.7	95.9
		45	5.0	92.0	3.0	96.3
		50	6.3	93.0	3.0	97.1
		75	5.0	91.7	2.1	96.9
		100	3.3	91.5	1.7	96.4
		125	7.3	93.2	2.4	96.1
		150	7.3	94.3	2.1	96.0

Table IV shows the recognition ability of Training Set I weight vectors. Only weight-sign elimination was used for the five cycles listed. Since complete convergence was not realized, training was arbitrarily terminated when the number of error corrections per iteration was not decreasing sufficiently to warrant further training. No features were discarded after cycle No. 5 with this method.

The detailed training results of cycle No. 5, Table IV, exemplify the variations in recognition observed during a training cycle. The error corrections per iteration, not listed, decreased steadily from about 85 corrections per vector at 40 iterations to 70 corrections at 150 iterations. Yet the sum of the missed patterns for the two vectors was 106 at 40 iterations and 107 at 150 iterations, with erratic jumps as high as 115 errors between these points. The essential difference between recognition results at 40 and 150 iterations was a transfer of 12 two-component patterns in the undecided category to 6 one-component errors at 150 iterations. Which pair of weight vectors to use, based on recognition only, is determined by the type of error that is least serious for the situation at hand. In general, it seems preferable to have an undecided response rather than an outright error, even though there are twice as many undecided answers in the trade. This amounts to choosing the pair of vectors where different SEP shapes were incorrectly categorized. For the training results that follow, the pair of vectors that showed the highest overall accuracy, with the undecided response being a lesser consideration, were chosen as the best classifier. The weight vectors after 40 iterations of cycle No. 5, Table IV, were considered best in this case. Recognition did not deteriorate with the weight-sign reduction from 133 to 57 features, remaining at about 96% accuracy with 5–6% undecided responses for the two categories.

Table V. Prediction Set I Results for Weight Vectors of Table IV, Identified by Cycle No. and Iteration

Cycle No.	Iteration	One-component Undecided, %	One-component Accuracy, %	Two-component Undecided, %	Two-component Accuracy, %
1	50	25.7	97.8	10.0	89.1
	100	15.3	94.1	8.8	91.6
2	50	16.7	96.0	6.5	90.5
3	50	13.7	96.9	4.3	89.0
4	50	11.0	92.1	2.2	89.9
5	15	11.3	97.7	5.2	88.6
	20	8.3	94.6	3.0	88.1
	25	5.0	96.1	1.5	88.3
	30	6.7	98.2	2.3	88.9
	35	9.0	97.8	2.3	89.2
	40	10.3	97.4	2.0	89.2
	45	6.7	92.9	2.2	89.8
	50	10.7	92.2	1.7	90.0
	75	6.3	87.2	2.0	90.6
	100	5.3	88.0	1.7	89.8
	125	8.0	88.4	1.7	90.0
	150	7.0	88.9	2.3	90.4

Table VI. Recognition of Training Set. Features Discarded on Weight-Magnitude Basis, Starting with 57 Features of Table IV

Cycle No.	No. of features	No. of iterations	One-component Undecided, %	One-component Accuracy, %	Two-component Undecided, %	Two-component Accuracy, %
6	52	60	10.2	95.2	3.0	95.4
7	47	100	3.0	97.3	3.9	94.9
8	42[a]	80	2.6	95.2	3.0	94.8
9	37	80	0	95.1	1.3	94.9
10	32	110	2.3	97.6	1.3	95.2
11	27	135	8.2	93.5	1.1	97.2
12	22	135	4.6	95.5	0.9	96.6
13	17	100	0	83.8	0	88.7
14	12	150	7.3	71.2	3.2	64.0

[a] Opposite weight signs occurred for two features, and these were discarded for next cycle.

The prediction ability of the trained vectors of Table IV is shown in Table V for Prediction Set I. The results are poorer for prediction than for recognition, but follow the same trend in both cases. A plot of training set *vs.* prediction set results for the four types of errors does show a rough correlation that justifies the choice of the best pair of weight vectors on the basis of training set results. The cycle No. 5, 40 iteration weight vectors showed a prediction ability for two-components of 89% accuracy and 2% no-decisions, and 97% accuracy and 10% no-decisions for one-components. It should be emphasized that these weight vectors were not chosen after-the-fact on prediction ability.

Eliminating features by weight-sign (Table IV) terminated at 57 features, but even more were discarded when weight-magnitude was used as the criterion. With these 57 features as a starting point, training with weight-magnitude elimination showed successful recognition using as few as 22 features. These recognition results are given in Table VI. Slightly better recognition resulted when 32 features were used, but cycle No. 12 with 22 features still provided better than 95% accuracy with less than 5% undecided responses.

Prediction ability for the vector pairs of Table VI is shown in Table VII for Prediction Set I. About 97% one-component accuracy and 7% undecided classifications occurred with the 32 features of cycle No. 10, while prediction with 22 features

Table VII. Prediction Set I Results for Weight Vectors of Table VI, Identified by Cycle No.

Cycle No.	One-component Undecided, %	One-component Accuracy, %	Two-component Undecided, %	Two-component Accuracy, %
6	9.3	95.2	2.0	88.8
7	5.3	98.6	2.3	90.4
8	7.7	97.1	1.3	91.2
9	2.7	96.9	1.3	91.1
10	6.7	96.8	1.2	92.2
11	5.3	90.5	0.8	92.6
12	4.7	92.7	1.8	92.4
13	1.3	81.1	0.5	82.7

Table VIII. Prediction for Conditions Exceeding Training Set Limits. Conditions are Same as Prediction Set I except for Stated Parameter

	Cycle No. 10 32 Features Undecided, %	Cycle No. 10 32 Features Accuracy, %	Cycle No. 12 22 Features Undecided, %	Cycle No. 12 22 Features Accuracy, %
a) Peak height[a]				
20–25:1	2.5	94.4	0	93.0
	2.5	89.2	3.5	92.8
25–30:1	3.5	87.0	3.5	89.6
30–35:1	1.0	84.8	4.0	88.5
35–40:1	3.0	78.9	5.0	81.6
b) Peak separation				
4–6 mV	3.5	88.1	1.0	86.9
6–8 mV	2.5	90.8	1.0	90.9
12–14 mV	1.5	89.8	0	89.0
14–16 mV	1.5	85.3	2.0	86.2
c) *n*-value: Two components				
0.98–1.00	1.0	69.7	5.0	90.5
0.92–0.98	1.5	69.5	0	99.5
3.00–3.08	5.0	60.3	0	2.5
d) *n*-value: One component				
0.98–1.00	3.0	41.2	1.0	9.1
0.92–0.98	0	99.3	4.3	57.9
3.00–3.08	0	99.3	0	98.5

[a] 1:20–25, etc. have also been included.

was at a 93% accuracy, 5% undecided level. Accuracy of two-component prediction was 92% with less than 2% undecided responses using either cycle No. 10 or cycle No. 12 vector pairs. These results compare favorably to those shown in Table V. It is expected that a classifier becomes a more reliable predictor when fewer features are required. The erratic trends observed in recognition and prediction as the dimensionality is decreased may be due to the lack of convergence. That is, a different weight vector occurs after every error-correction. Practicality necessitated the examination of a relatively small fraction of these vectors.

Prediction of Patterns Outside Training Set Limits. A truly challenging task for a pattern classifier is the prediction of patterns which exceed the bounds of the training set. This is the situation in Table VIII for peak height ratios, peak separations, and *n*-values outside the limits of Training Set I. The effect on prediction ability was investigated for only one of these parameters at a time. The remaining two parameters were varied within the limits stated for Prediction Set I.

Table IX. Twenty-Two Features Used in Cycle No. 12, Table VI, Which Provided Satisfactory Classification Performance. Listed According to Type

Ratios	Ifunction
$(E_{3/4} - E_p)/(E_{1/4} - E_p)$	ΔE_{i_p} = 8, 16, 32 mV
i''_{p_3}/i''_{p_2}	$\Delta E_{i'_{p_1}}$ = 8, 16 mV
i''_{p_1}/i''_{p_3}	$\Delta E_{i''_{p_2}}$ = 8 mV
i'_{p_1}/i'_{p_2}	$\Delta E_{i''_{p_3}}$ = 16 mV
Efunction	**ΔE or Δi**
85% i_p, i'_{p_1}, i''_{p_1}	75, 85% i''_{p_1}
0%, 70% i''_{p_2}	80, 85% i''_{p_2}
	$\Delta E_{i'_{p_1}}$ = 8, 16 mV

Table X

A. Features of cycle No. 12, Table VI, listed in order of decreasing weight-magnitude

Feature No.	Feature
1	$(E_{3/4} - E_p)/(E_{1/4} - E_p)$
2	i''_{p_3}/i''_{p_2}
3	i''_{p_1}/i''_{p_3}
4	Ifunction at ΔE_{i_p} = 16 mV
5	Efunction at 0% i''_{p_2}
6	Ifunction at $\Delta E_{i''_{p_1}}$ = 8 mV
7	Efunction at 85% i''_{p_1}
8	Ifunction at ΔE_{i_p} = 8 mV
9	ΔE at 85% i''_{p_1}
10	Efunction at 85% i'_{p_1}
11	Ifunction at $\Delta E_{i'_{p_1}}$ = 8 mV
12	Efunction at 85% i_p
13	ΔE at 85% i''_{p_2}
14	ΔE at 75% i''_{p_1}
15	i'_{p_1}/i'_{p_2}
16	ΔE at 80% i''_{p_2}
17	Ifunction at ΔE_{i_p} = 32 mV
18	Ifunction at $\Delta E_{i'_{p_1}}$ = 16 mV
19	Efunction at 70% i''_{p_2}
20	Ifunction at $\Delta E_{i''_{p_3}}$ = 16 mV
21	Δi at $\Delta E_{i'_{p_1}}$ = 16 mV
22	Δi at $\Delta E_{i'_{p_1}}$ = 8 mV

B. Additional ten features included in cycle No. 10, Table VI

Efunction at 85% i''_{p_3}	SEP Area (E_p + 140 mV to E_p − 80 mV)
A'_{p_1}/A'_{p_2} (Areas)	Efunction at 80% i_p
ΔE at 75% i''_{p_2}	$A'_{p_1} - A'_{p_2}$ (net area)
i'_{p_1}/i''_{p_1}	Δi at ΔE_{i_p} = 8 mV
ΔE at 70% i''_{p_2}	Efunction at 75% i_p

Prediction ability for the 32 features of cycle No. 10 and the 22 features of cycle No. 12 was examined.

Table VIII, part *a*, shows the effect of peak height ratios greater than 20:1 on prediction ability. The two sets of results given for the 20:1 to 25:1 peak height range represent two different randomizations of the six parameters used to generate a two-component polarogram. There were 200 patterns in each random sample. Prediction ability varied about 4–5% between samples, but remained at about the 90% level, which is comparable to Prediction Set I. Polarograms for the 25–30:1 peak height range listed were generated by adding 5 peak height units to the peak height of the second 20–25:1 random sample. Thus, parameter combinations for the polarograms were held constant to eliminate the variation between random samples. Subsequent peak height ranges were obtained in this manner. Prediction accuracy was still about 85% with 1% undecided responses for polarograms having 30:1 to 35:1 peak height ratios.

Table VIII, part *b*, shows the effect of peak separation on prediction ability. Again, different separation ranges were obtained by adding a constant to the peak location parameter of one of the component curves. Correct decisions were made about 90% of the time for polarograms having peak separations within 2 mV of the 8–12 mV training set limits. Better than 85% accuracy was observed for separations in the 4–6 mV and 14–16 mV ranges.

The effect of *n*-value on prediction is shown in Table VIII, parts *c* and *d*, for two-component and one-component polarograms. The decision planes used in Table VIII were unreliable classifiers for *n*-values outside the 1–3 training range. This indicates that these patterns are close to the decision plane and that the results of Table VIII, parts *c* and *d*, are erratic, and any apparently satisfactory results are fortuitous. Intuitively, individual component curve shapes are more important than peak-height ratios or peak separations in the location of a pattern point. The fact that prediction ability varied considerably more between cycle No. 10 and cycle No. 12 weight vectors for *n*-value variations than for either peak-height ratios or peak separations also indicates the fortuitous nature of the *n*-value results.

Features Retained. The twenty-two features used in cycle No. 12, Table VI, are listed according to type in Table IX. Features extracted from the zero-, first-, and second-derivative polarograms were retained. Efunction and Ifunction features, defined in Figures 3 and 4, occurred frequently. The potential ratio shown was the only feature of this type in the original set of 133 features. Notably absent from the list in Table IX are area features, relative potential features ($E_{3/4} - E_p$), derivative peak currents, and cross-derivative current features (i'_{p_1}/i''_{p_1}).

The twenty-two features of cycle No. 12 are listed in order of decreasing weight-magnitude in Table X, part *A*. Ratio features are predominant. The relative potential ratio feature has the largest weight. The current ratio, i''_{p_3}/i''_{p_2}, is the term used by Grushka and Monacelli (*8*) for one coordinate of a calibration plot of gas chromatographic data for the purpose of distinguishing doublet peaks from singlets. The feature corresponding to the other coordinate, i''_{p_1}/i''_{p_2}, was not retained. Efunction at 0% i''_{p_2} represents the two crossover potentials of the second-derivative polarogram. Upon closer inspection, features No. 1 and No. 3 are relatively independent of *n*-value for one-component polarograms compared to two-component polarograms, but features No. 2, 4, and 5 showed considerable variation in this respect. Investigation by pattern recognition of other skew functions should be worthwhile.

The additional ten features retained prior to cycles No. 11 and No. 12 are given in Table X, part *B*. Area features and a cross-derivative current ratio feature are now included.

Other Classification Attempts. In practical use, a pattern classifier could be employed to test the purity of a sample. Thus, a one-component classification would be the basis for terminating separation work. In this case, it would be most desirable to have 100% accuracy in classifying two-component samples so that an impure sample would not erroneously be considered pure. Incorrect classification of pure samples and undecided responses would result in additional separation work on which the final decision could be based. This situation might be achieved by the use of s', a non-zero decision limit (*14*). Unfortunately, the cycle No. 12 weight vectors trained in this work had improved one-component classification by the use of a slightly positive s', but a negative

s' did not improve two-component accuracy without serious degradation of one-component results.

CONCLUSIONS

This work demonstrates that a binary pattern classifier can be used to detect the presence of doublet peaks in stationary electrode polarography under widely varying conditions of peak height ratio, peak separation, and n-value. The overlap tested here for theoretical reversible systems was so severe that visual subjective interpretation from second-derivative data was not possible. The method greatly exceeds the limits of conventional derivative peak detection (*5*). Better than 90% prediction accuracy was obtained with overlapping reduction peaks as close as 6 mV. Good prediction could be obtained though complete convergence was not realized. A large number of representative curves were necessary in training to achieve successful prediction when any n-values between 1 and 3 electrons were allowed. By combined methods of feature elimination, only 22 of the 133 features investigated were necessary to implement the classifier.

The computerized pattern classification scheme demonstrated here for the detection of signal multiplicity is applicable to many analytical measurement techniques. Inherent in this scheme is the selection of features from the observed signal. This selection would be tailored to the technique in question. No prior information about the data is necessary other than category for training purposes, and this situation was assumed for the work presented here. The features which predominate in the empirical classification problem could be investigated for fundamental relationships to the experimental phenomenon being measured. However, this is not to say that only important relationships are retained. Since the features used here are certainly not intended to include every possibility, the results presented are only relative to the features chosen for investigation. Other features could be devised, other normalization procedures used, and different pattern classification rules applied (*32*), that might improve upon the results presented here. However, this work is a step toward that goal. The investigation of real analytical data is the important next step.

ACKNOWLEDGMENT

The authors thank F. E. Lytle and J. E. Davis for their helpful comments.

RECEIVED for review March 6, 1972. Accepted July 20, 1972. This work was presented in part by the authors at the 163rd National Meeting of the American Chemical Society, Boston, Mass. April 9, 1972. This work was supported by the National Science Foundation, Grant No. GP-21111. L. B. S. gratefully acknowledges Fellowships granted by Hercules, Inc., SOHIO, the Purdue Research Foundation, and the Analytical Division of the American Chemical Society sponsored by DuPont's Instrument Products Division.

Stability Constants of Some Polyamine and Polyaminocarboxylate Complexes in Methanol–Water Mixtures by Differential pH*-Potential Titrimetry

D. B. Rorabacher, B. J. Blencoe, and D. W. Parker

Department of Chemistry, Wayne State University, Detroit, Mich. 48202

Stability constants of Cu(II), Zn(II), and Cd(II) complexes with the polyamines triethylenetetramine (trien) and tetraethylenepentamine (tetren) were determined in methanol–water solvent mixtures containing 40, 65, 80, 90 (trien only), 95, and 99% methanol (by wt) at 25 °C, $\mu = 0.1M$ by means of a differential pH*-potential titrimetric approach employing a mercury–glass–calomel electrode system. For these complexes the stability constant values exhibit an accelerating increase with increasing methanol content in the solvent, the net gain being 1000-fold or greater on going from water to 99% methanol. By contrast, the stability constants of ethylene glycol bis(β-aminoethyl ether) *N,N,N',N'*-tetraacetate ion (EGTA) with Mg(II), Ca(II), Co(II), and Cd(II) and *trans*-1,2-diaminocyclohexane-*N,N,N',N'*-tetraacetate ion (CDTA) with Mg(II), Ca(II), and Ni(II) in 99% methanol fail to show notable increases relative to aqueous values, despite the more favorable electrostatic contribution as the solvent dielectric is decreased. The analytical implications are discussed.

THE STUDY OF COORDINATION REACTIONS in nonaqueous solvents and the development of analytical methods based on such reactions have been hindered by the lack of available thermodynamic data (*1*). Recent interest in this laboratory involving the influence of solvent on the kinetics of metal-complex formation (*2–4*) has led us to obtain information in this area.

Much of the current interest is centered on the use of alcoholic solvents—particularly methanol and methanol–water mixtures. In a recent paper we reported on the determination of ligand protonation constants for polyamine and polyaminocarboxylate species as a function of solvent composition in methanol–water solvent mixtures (*5*). In the current paper, we report the determination of stability constants for some complexes formed by these ligands utilizing, in part, the protonation constant data which are now available.

Previous work on stability constant measurements in methanol appears to be limited and the methods employed are

(1) A. Ringbom, "Complexation in Analytical Chemistry," Interscience, New York, N.Y., 1963, p 14.

(2) W. J. MacKellar and D. B. Rorabacher, *J. Amer. Chem. Soc.*, **93**, 4379 (1971).

(3) F. R. Shu and D. B. Rorabacher, *Inorg. Chem.*, **11**, 1496 (1972).

(4) D. B. Rorabacher and F. R. Shu, *Inorg. Chem.*, submitted for publication.

(5) D. B. Rorabacher, W. J. MacKellar, F. R. Shu, and M. Bonavita, ANAL. CHEM., **43**, 561 (1971).