

Measurement and Control of Information Content in Electrochemical Experiments

Sam P. Perone and Cheryl L. Ham

Lawrence Livermore National Laboratory, Livermore, CA 94550

Accepted: July 1, 1985

One of the most important problems in chemical analysis is the interpretation of analytical data. The difficulty of this task has been further compounded by the data explosion. Chemical information relevant to the particular analysis problem is hidden within excessive amounts of data. This problem could be alleviated through knowledge and control of the information content of the data. Information theory provides a means for the definition, evaluation, and manipulation of quantitative information content measurements. This paper provides a general review of some of the basic concepts in information theory, including history, terminology, entropy, and other information content measures. The application of information theory to chemical problems requires some modifications. The analyst is usually only interested in a subset of the information (data) which has been collected. Also, this relevant chemical information is dependent upon not only the informational goals of the problem, but the completely specified procedure as well. This paper reviews chemical applications of information theory which have been reported in the literature including applications to qualitative analysis, quantitative analysis, structural analysis, and analytical techniques. Measures of information and information content and figures of merit for performance evaluations are discussed. The paper concludes with a detailed discussion of the application of information theory to electrochemical experiments and the empirical determination of the information content of electroanalytical data.

Key words: chemical analysis; chemical information; data explosion; electrochemical experiments; information content measurements; information theory, chemical applications

Introduction

Data interpretation is one of the most challenging problems of chemical analysis. Both the *data-rich* and the *data-limited* cases stress the need for efficient methods to extract chemical information from the available data. Data-rich analyses result from the ability of modern chemical instrumentation to generate enormous amounts of data in short periods of time. The current trend towards more exotic

hybrid instruments buries the chemical information even deeper within the data. Alternatively, *data-limited* analyses often result from limitations in appropriate sensors, accessible techniques, time, and manpower. The need for efficient methods to extract chemical information is superseded only by the need to acquire information-rich data.

Improving accessibility of chemical information emphasizes the importance of good experimental design and requires a re-evaluation of the traditional approach to chemical analysis. The typical approach involves using foreknowledge about the samples to choose an analytical procedure. The analytical procedure, which may involve more than one analytical technique, is used to produce as much data as possible which is collected for later analysis. Data interpretation is performed by the analyst using as much intuition and background knowledge as possible. In

About the Authors, Paper: Sam P. Perone and Cheryl L. Ham are with Lawrence Livermore's Chemistry & Materials Science Department. The work they describe was supported by the Office of Naval Research and the Lawrence Livermore National Laboratory

the data-rich case, emphasis is often on data reduction. The data-limited case emphasizes information extraction.

A more efficient and desirable approach to chemical analysis would be to maximize the amount of information obtained relevant to the current problem, while minimizing the amount of analytical effort, time, and collected data. The evaluation, selection, and optimization of analytical procedures need to be investigated further. In order to study this problem, a method for the quantification of chemical information must be defined. This paper reviews some relevant information theory concepts and describes applications to chemical analysis. Illustrations have been taken from the literature, as well as from our own recent work, to demonstrate the value of applied information theory for optimized chemical analysis methods.

Information Theory Concepts

Information theory [1-4]¹ is concerned with the study of information and its transmission. The earliest information theorists studied the encoding and decoding of secret codes. Modern work on information theory can be traced back to the 1920's beginning with Carson's study of frequency modulation. Nyquist determined the minimum bandwidth required for the transmission of a definite quantity of information. Hartley established a definite bandwidth-time product required for the transmission of a definite quantity of information. However, most of the present work in information theory is based upon probabilistic models of communication developed by C.E. Shannon in 1948 [5]. Simply stated, the basic principle is that a message with a high probability of random occurrence conveys little information. The most information is conveyed by the message that is least likely to spontaneously occur.

This principle is formalized by the concept of entropy which equates information and uncertainty. Entropy is a quantitative measure of the amount of information supplied by a probabilistic experiment. It is based upon classical Boltzmann entropy from statistical physics. Shannon's formula, eq (1) defines the entropy or the average information

$$H_{av} = I_{av} = - \sum_i^N p(x_i) \log_2 p(x_i) \quad (1)$$

to be equal to the weighted average of the specific information for each event in the system under consideration. Specific information [6] is information conveyed by the occurrence of a particular event and is quantified by the $-\log_2$ of the probability of the event ($p(x_i)$). Entropy is limited by a maximum of $\log_2 N$ (Hartley's equation), where

N is the number of events, when considering a set of mutually exclusive but equally probable events [1,6,7]. For example, consider the measurement of three distinguishable intensity levels. If the probabilities of measuring these levels are 0.25, 0.30, and 0.45, respectively, the average entropy would equal 1.5 bits. The amount of specific information conveyed by the measurement of each level would equal 2.0, 1.7, and 1.1 bits, respectively. Notice that the least likely level to be measured does indeed convey the most information. The maximum entropy is equal to 1.6 bits. A thorough treatment of the mathematical basis of entropy and its properties is given in a book by Mathai and Rathie [2].

Redundancy [6] is the difference between the maximum information and the average information eq (2). Relative

$$I_{red} = \log_2 N - \sum_i^N p(x_i) \log_2 p(x_i) \quad (2)$$

redundance is the ratio of the redundancy to the maximum information [6]. Relative information content is the ratio of the actual average information to the maximum information [1]. Redundancy can then be expressed as the remaining fraction not due to relative information [1]. In the above example there is 0.1 bit of redundancy and 0.062 relative redundancy. If the actual average information is equal to 1.4 bits, the relative information content is equal to 0.88 and the redundancy equals 0.12.

Types of Information

The concept of information as used in information theory refers to the choice or uncertainty of outcomes when regarding distinguishable elements with respect to some random mechanism. Information is a system based upon elements that have been agreed upon for the representation of information (characters) and the relationships between them (codes). It is not a measure of meaning as used in the usual sense, which implies a subjective evaluation. The concept of information as used in chemical analysis encompasses the uncertainty regarding the quantity, identity, chemical structure, or properties of the analyte of interest.

Preinformation or foreknowledge [1] is the prior knowledge concerning the occurrence of events. The information conveyed by the occurrence of more than one independent event is simply the sum of the information conveyed by each event individually. However, if the occurrence of a second event is dependent upon the occurrence of the first event, the foreknowledge of the first event reduces the amount of information conveyed by the second event. Therefore, the amount of information conveyed by a series of events within the system under consideration is always less than or equal

¹Figures in brackets indicate literature references.

to the sum of the information conveyed by each of the events separately.

Preinformation is probably the most commonly used information theory concept in chemical analysis. Chemical preinformation [8] is information that is known prior to performing the current analysis. It may result from experience, preliminary analyses, etc. It is used to reduce the effort required to solve the analytical problem. Preinformation may be quantified through the use of entropy-based measures [8]. The uncertainty before the analysis for a discrete variable, such as chemical identity, is quantified by the use of the *a priori* probabilities for identification in Shannon's equation (eq 1). The uncertainty for a continuous variable such as concentration or signal intensity is expressed by integrating the *a priori* probability density function over the range of interest eq. (3).

$$H(X) = - \int_{x_1}^{x_2} p(x) \log_2 p(x) dx \quad (3)$$

Joint information [1] is information that is provided by more than one event. It can be quantified by substituting the joint probability of occurrence for the events into Shannon's equation (eq 1). In the case of independent events, the joint probability of occurrence is simply the product of the *a priori* probabilities of occurrence. For nonindependent events, it is the product of the *a priori* probability for the first event with the conditional probabilities for the other events.

Mutual information describes the amount of information in one event that determines the state of another event. It may also be thought of as the average amount of information required to distinguish members of different classes. Isenhour et al. [9] investigated the relationship between mutual information and classification in the determination of chemical functionality for 200 compounds based upon binary encoded (peak/no peak) infrared spectra. Mutual information was calculated as the difference between the total average entropy and the average conditional entropy. The total average entropy is the average amount of information required to distinguish between the 200 spectra under consideration. It is calculated as a weighted average of the probability of occurrence of a peak maximum for each spectral interval using Shannon's formula (eq 1). The average conditional entropy is the average amount of information required to distinguish between members of the same class. Average conditional entropy is calculated as the sum of the class conditional entropies weighted by class size. In the case of two separable, equally probable classes, the independent mutual information is equal to one bit. A value for mutual information greater than one bit implies the inclusion of redundant information in the data. The square root of mutual

information was shown to be linearly related to the maximum likelihood classification ability.

Figures of Merit

The application of information theory concepts to chemistry is most familiar in the evaluation of analytical methods. Figures of merit such as accuracy, precision, and detection limit have long been used to evaluate the attainment of informational goals such as concentration, resolution, and sensitivity, respectively. Figures of merit are measures of goal achievement for completely specified procedures that can be used for evaluating, selecting, and comparing analytical procedures. Other quantifiable factors that can be used to determine the applicability of analytical procedures to a particular problem include sensitivity, selectivity, speed of analysis, personnel requirements, and cost of the analytical procedure.

Grys [10] described five new functional concepts: accuracy, limit of detection, firmness, efficiency, and cost, which result in judgements of acceptability of analytical methods. Accuracy is expressed in terms of recovery and reproducibility. The contribution due to recovery can be calculated by summing the percentage of different losses throughout the whole procedure. Reproducibility is expressed by the ratio of the full range to 100 times the ideal range.

Limit of detection is the concentration of a sample that gives a reading that is equal to twice the confidence half-interval for a series of ten determinations of the blank test value determined to 99% certainty. It is measured in mg per kg, or ppm, and is given by multiplying the standard deviation by 2.17, a factor that is determined from the *t* test for $t_{0.01}$ and $n=10$.

Firmness is an index of the effects of different factors upon the results. It is equal to the total deviation from expected values caused by the presence of equimolar amounts of interfering substances or connected with 5% changes in optimum reaction conditions such as acidity or reagent concentrations.

Efficiency provides information about the time consumption during the course of the whole procedure. It is expressed as the time of effective labor for one sample in minutes divided by 100.

Cost is a measure of the expenditure of materials and equipment used for the analysis of one sample by a new method in relation to the least expensive method. The cost of any desired method by which the analysis can be performed may be substituted for that of the least expensive method. It is given by dividing the ratio of the cost of the new method to the old method by 100.

Eckschlager [11] discussed two informational variables, time-information performance and specific price of informa-

tion, which can be utilized in the evaluation and optimization of analytical methods. Time-information performance [12] can be rewritten as the ratio of information content to the time required for the analysis, including the analysis itself plus the time required to prepare the equipment for the analysis of the next sample. The time required for analysis can be partitioned into two segments, the basis time and the time required for the performance of N parallel determinations. The specific price of the information [11] is defined as the ratio of the cost of the analysis to the amount of information obtained through simultaneous determination of N components.

Danzer and Eckschlager [13] defined a general measure of information efficiency as the product of efficiency coefficients that are based upon the ratio of the value of the variable characterizing the properties required for the solution of particular analytical assignments to the actual value that the method provides. A ratio greater than one implies that more of the property is required than is provided by the method and the efficiency coefficient is assigned the value of zero. Otherwise, the efficiency coefficient is assigned the value of the ratio. They also defined a measure of information profitability as the ratio of the information efficiency to the specific price of the information. Results for the determination of manganese in low-alloy steels were tabulated for seven analytical methods: titrimetry, potentiometric titration, photometry, atomic absorption spectrophotometry, optical emission spectroscopy, optical emission spectrography, and optical emission spectrometry. The results demonstrated that information efficiency and information profitability are not always correlated. For example, although potentiometric titration has almost five times the information efficiency of titrimetry, both methods have the same information profitability when the duration of the analysis should be less than one day.

Informing Power

Informing power is a measure of the amount of information available in a given analytical procedure. The concept was developed by Kaiser [14] with respect to spectrochemical methods of analysis eq (4) as a function of the resolving

$$P_{\text{inf}} = \int_{\nu_a}^{\nu_b} R(\nu) \log_2 S(\nu) \frac{d\nu}{\nu} \quad (4)$$

power, $R(\nu)$, the maximal number of discernable steps for the amplitude, $S(\nu)$, and the spectral range, ν_a to ν_b . If the resolving power and the maximal number of steps are fairly constant over the spectral range under consideration, informing power reduces to eq (5). For example, a grating

$$P_{\text{inf}} = R_{\text{av}} \log_2 S_{\text{av}} \ln(\nu_b/\nu_a) \quad (5)$$

spectrograph system with a resolving power of 2×10^5 , a spectral range from 2000 to 8000 Å, and 100 discernable steps in measured intensity levels at each wavelength would have an informing power of 2×10^6 bits. It is obvious that here the resolving power is the most important factor in maximizing informing power. In the case of a nondispersive, monochromatic method, resolving power between peaks at different wavelengths is not applicable and informing power is simply the \log_2 of the number of discernable amplitude steps at that wavelength. For example, 100 discernable intensity steps yields an informing power of 7 bits. The informing power for the corresponding polychromatic method is that for the monochromatic method multiplied by the number of frequencies. If the number of steps is different for each of the different frequencies, then the informing power is the same as for a collection of monochromatic methods, and the \log_2 of the number of steps is summed over each of the different frequencies.

Fitzgerald and Winefordner [15] extended the application of informing power to time-resolved spectrometric systems with the addition of a second resolving parameter, R_t . If both resolving powers and number of discernable steps are nearly constant over the range, then informing power reduces to eq (5) multiplied by $R_t \ln(t_2/t_1)$. For example, an atomic fluorescence spectrometer with an average resolving power of 3000 over a spectral range from 200 to 500 nm with an average of 200 discernable intensity steps has an informing power of 5.7×10^4 bits. The informing power is increased to 9.7×10^6 bits for a range of 10^{-9} (t_1) to 10^{-6} sec with a measurement time limited by the lifetimes of the excited species of 10^{-7} sec (t_2) and a δt of 10^{-9} sec (R_t equals $(t_2 - t_1)/\delta t$). Comparisons of the informing power for a single beam molecular absorption spectrophotometer, normal molecular absorption phosphorimetry and time-resolved phosphorimetry showed an increase in the informing power by a factor of two for the normal phosphorimeter over the spectrophotometer. The addition of a time resolution element to phosphorimetry increased informing power by a factor of 450. The addition of a time resolution element to atomic fluorescence spectrometry increased the informing power by a factor of 170. Informing power was also used to compare analytical methods as well as to compare analytical instruments. A general photon counter was shown to have an informing power three times larger than that for an analog synchronous detection system.

Yost extended the application of informing power to tandem mass spectrometry [16,17], a method capable of generating enormous amounts of data. In the case of a quadrupole mass filter, the minimum resolution element, δx is constant rather than the resolving power, $R(x)$. Informing power can then be calculated as shown in eq (6). A quad-

$$P_{\text{inf}} = \frac{1}{\delta x} \log_2 S(x) [x_b - x_a] \quad (6)$$

ropole mass spectrometer with unit mass resolution, a mass range of 1000, and an ion intensity range of 2^{12} bits would have an informing power of 1.2×10^4 bits. The addition of another resolution element produces a double integral in the informing power equation that is equivalent to the the product of the informing power of the two elements. The addition of a capillary gas chromatograph with a nearly constant 10^5 theoretical plate resolution over a one hour analysis time, results in 6.6×10^6 bits of informing power. The addition of a second quadrupole mass spectrometer with the same characteristics results in 1.2×10^7 bits of informing power. The combination of a capillary gc/ms/ms system results in an informing power of 6.6×10^9 bits, an increase by a factor of 5.5×10^5 over the original quadrupole mass spectrometer. The effect of experimental parameters on informing power was also demonstrated by Yost [16,17]. The variables associated with the collisionally activated dissociation process are potential resolution elements. Energy- and pressure- resolved ms/ms has an informing power of 3.6×10^9 bits.

The informing power metric can also be applied to electrochemistry. Using a current range of -20 to $+20 \mu\text{A}$ that can be measured to within $.005 \mu\text{A}$ yields 4×10^3 discernable steps. Eq (6) can be used to calculate the informing power for a cyclic staircase voltammetry (CSCV) experiment in which each current pulse is sampled and analyzed. An experiment with a staircase step of 13.5 mV (δx) and a potential range scanned from 0.0 to -1.73 V yields 3.1×10^3 bits of information.

Boudreau and Perone [18] demonstrated quantitative resolution in programmed potential step voltammetry for overlapped peaks with 30 mV separation between half wave potentials. If only resolved peaks are analyzed and the smallest resolution is 30 mV , the informing power is 1.4×10^3 bits. The addition of a time resolution element increases the informing power for electrochemical methods. Taking 45 equally spaced current measurements on each step at a sweep rate of 1.00 V/sec for the CSCV experiment increases the amount of information to 3.5×10^7 bits. The amount of information obtained from CSCV experiments can be easily manipulated by changing or adding resolution parameters.

Informing power can be used as a figure of merit for a completely specified method or system. Although the informing power of instrumental techniques may seem excessive when compared to the maximum information as calculated by Hartley's formula, it must be remembered that informing power is simply a measure of the maximal number of bits of information available in the procedure, not necessarily the useable or necessary amount of information.

Limitations in informing power arise from differences between practical and calculated resolving power. The lack of independence between the bits of information, noise, and interference result in the reduction of the useful informing power. Informing power can be partitioned into the amount of information required for the solution of the problem and the amount of redundant information required to provide a given level of confidence.

Information Content

One of the most important concepts in information theory is that of informational gain or information content [19]. This is equal to the change in entropy due to the experiment and is quantified by the difference between the entropy using *a priori* probabilities and the entropy using *a posteriori* probabilities eq (7). The use of Shannon's

$$I(X|Y) = H(X) - H(X|Y) \quad (7)$$

formula eq (1) to calculate the entropy does not guarantee a non-negative information content. However, another informational measure eq (8) always results in non-negative

$$I(X|Y) = - \sum_i^N p(x_i|Y) \log_2 [p(x_i|Y)/p(x_i)] \quad (8)$$

values. For equal *a priori* probabilities, information content as calculated by Eqs (7,8) are equivalent. Since information content as discussed above can only be established after the analysis, these measures cannot be used as a quality criterion for selecting an analytical procedure. However, they can be used to evaluate the performance of a procedure.

Measures of information content has been applied to information theory models of structural analysis, qualitative analysis, quantitative analysis, trace analysis and instrumental analysis. Eckschlager and Stepanek have published a book [7] and a review article [8] on the application of information theory to analytical chemistry.

Structural Analysis

One of the most difficult analytical tasks is the unambiguous determination of chemical structure. However, application of information theory to structural analysis is based upon a relatively simple entropy model [8] and an informational measure introduced by Brillouin [20]. The input consists of a finite number, n_0 , of equally probable identities such as functional groups or conformational arrangements.

The output is a portion of a signal that corresponds to the identity, such as an IR band, NMR peak, or MS m/z peak, encoded only as to its presence or absence. The number of possible, but as yet undistinguished, structural arrangements is n , where $1 \leq n \leq n_0$ and $n=1$ for an unambiguous determination of the structure for the analyte.

The uncertainty prior to analysis can be expressed by substituting the *a priori* probabilities into Shannon's equation eq (1). Since the *a priori* probabilities are equal, that is $1/n_0$, then the situation reduces to the case of maximum information (Hartley's equation) and the uncertainty is equal to $\log_2 n_0$. The uncertainty after analysis is equal to $\log_2 n$. The decrease in uncertainty due to the analysis corresponds to the informational gain eq (7) and is equal to $\log_2 (n_0/n)$. It assumes its maximum value in the case of the unambiguous determination of the structure and is equal to $\log_2 n_0$.

Qualitative Analysis

The input for the model of qualitative analysis consists of a set of discrete identities, X_i , where $i=1,2,\dots,n_0$. If the output consists of a number of discrete, equally likely identities, the limiting case of Shannon's equation (Hartley's equation) can be used to calculate the entropy, and the information gained can be expressed by a ratio of the number of possible components before and after the analyses eq (9). [7]. For example, consider the case of an addition of

$$I(p, p_0) = \log_2 (n_0/n) \quad (9)$$

HCl to a solution that contains only one of a list of 25 possible cations, three of which can be precipitated by HCl. The information gained as evidenced by a precipitate would be equal to $\log_2 (25/3)$ or 3 bits. The lack of a precipitate would imply an informational gain of only 0.2 bits. To consider various combinations of components, the total number of possible combinations is given by eq 10, where

$$n_0 = \sum_m^M n_m(M) = \sum_m^M [M!/(m!(M-m)!)] \quad (10)$$

M is the total number of components, and is divided into groups of m components. In the case of a solution that contains from one to six cations, the total number of combinations is equal to 53. If two of them can be precipitated by HCl, there are 15 combinations in which neither cation is present and the information gained is 1.8 bits. For the 38 remaining cases in which either one or both cation is present, as evidenced by the appearance of a precipitate, an informational gain of 0.5 bit results.

In the case of instrumental or chromatographic qualitative or identification analyses, the output is a set of discrete signals in positions Y_j , where $j=1,2,\dots,m$ and $m \geq n_0$ [8]. The entropy can be expressed by Shannon's formula eq (1) and reaches a maximum when all of the possible identities, X_i , are equally likely. It is equal to zero when one identity is confirmed and the others are excluded, as would be the case for the *a posteriori* entropy for an unambiguous identification. The interpretation of these signals leads to an input-output relationship for the system that is represented by a set of *a posteriori* conditional probabilities, $p(x_i|y_j)$. The interpretation of these signals is also dependent upon preinformation represented by the *a priori* probabilities that may be calculated by Bayes theorem eq (11) [19].

$$p(x_i|Y) = [p(x_i)p(Y|x_i)] / \left[\sum_i p(x_i)p(Y|x_i) \right] \quad (11)$$

The information content of an analytical signal is defined as the decrease in uncertainty eq (7). In the case of unambiguous determinations, $H(X|Y)=0$ and $I(X|Y)=H(X)$, and $H(X)$ is also considered as the information required for unambiguous determination. However, most qualitative or identification procedures are chosen so as to minimize the uncertainty in identification for every possible signal. This is quantified by the informational measure of equivocation. Equivocation [8,19] is a measure of the expected or average value of the uncertainty after analysis eq (12). Equivocation

$$E = H(X|Y) = \sum_j p(y_j) H(X|y_j) \quad (12)$$

and information content are complementary quantities, their sum equaling the entropy of the identification procedure. For an "ideal" procedure or an unambiguous determination, equivocation equals zero and information equals entropy.

Cleij and Dijkstra [21] demonstrated the use of information content and equivocation in the evaluation of thin-layer chromatographic procedures. Information content and equivocation were calculated for the identification of DDT and 12 related compounds for 33 different TLC systems. Calculations of the equivocation for the combinations of two TLC systems showed that the best combinations are not produced by combinations of the best individual TLC systems. This reflects the correlation between the best individual TLC systems.

Another method for quantifying information content is from the perspective of the possible signals rather than the possible identities [19]. Signal entropy, $H(Y)$, is the uncertainty in the identity of the unknown signal and can be quantified by substituting the probabilities of measuring the signals into Shannon's equation, eq (1). The conditional

entropy, $H(Y|x_i)$, is the uncertainty in the signal if the compound is known to be x_i . It can be considered a measure of noise and is given by substituting the conditional probabilities into Shannon's equation eq (1). Expected values for entropy and information content can be expressed in a manner analogous to that shown above. Also, since entropy is strong-additive [2,19], information content can be expressed in terms of the signal entropies.

Dupuis and coworkers [22,23] applied these methods to gas-liquid chromatography. The information content for 10 stationary phases used in gas-liquid chromatography was calculated on the basis of compound identification by retrieval of retention indices from a compiled library for a set of 248 compounds [22]. The information content per column ranged from 6.5 to 7.0 bits. The information content for combinations of columns is dependent upon both the number of columns and the sequence of columns. Ten sequences of the 10 columns yielded an information content of 43.3 bits. The study was expanded to include 16 gas-liquid chromatography stationary phases [23]. The complete data set of 248 compounds, a subset of 48 aliphatic alcohols, a subset of 35 aldehydes/ketones, and a subset of 60 esters were explored. For all four sets of compounds, combinations of stationary phases that yielded the most information consisted of one nonpolar phase plus one or more polar phases.

Van Marlen and Dijkstra [24] calculated the information content for the identification of binary coded mass spectra by retrieval and determined the optimal sequence of masses which contained the most information. A set of approximately 10,000 low resolution mass spectra were binary encoded using a threshold intensity level of 1% of the intensity of the base peak. Masses greater than 300 did not yield any additional information. Individually, masses of 300 or less contained from zero ($m/z=3,4,5,6,7$) to one ($m/z=29,40,51,53,57,69,77$) bit of information. The optimal mass sequence of 120 masses contained 40.9 bits of information, demonstrating the obvious redundancy in the binary coded spectra. The optimal mass sequence is dependent upon the distribution of the peaks. A set of 200 binary coded alkane spectra yielded 9 bits of information for 25 selected masses.

Quantitative Analysis

The model for quantitative analysis is a two-stage model [8]. The input is a continuous distribution that produces a continuously variable signal. In the second stage, the signal specified by both position and intensity is decoded into results. The distribution of the results for parallel determinations is generally normal. Preinformation indicates that content of the component lies within a specified range of x_0 to x_1 so the *a priori* probability density is that of a rectangular

or uniform distribution. The information content is considered a divergence information measure that represents the error term in the measurement of the inaccuracy of the preinformation [7,8]. If the results confirm the *a priori* assumptions for the component, the information content is given by eq (13). The effect of a systematic error,

$$I(p,p_0) = \log_2 [(x_1 - x_0) / (\sigma \sqrt{2\pi e})] \quad (13)$$

δ , reduces the information content by factor of $\delta^2/2\sigma^2$ [7]. The use of parallel determinations, n_p , and the estimate of σ , s , results in eq (14), which utilizes the student's t test at

$$I(p,p_0) = \log_2 [(x_1 - x_0)n_p / (2st(v))] \quad (14)$$

the significance level of 0.038794. This level of significance is chosen so that twice the t value at infinity equals $\sqrt{2\pi e}$ as the number of degrees of freedom approaches infinity. The information content as measured by eq (14) is the practical form of eq (13) since usually only the estimate of the standard deviation is known.

Poisson distributed results can be approximated by a normal distribution with the population mean, μ , equal to the constant representing the average number of random points per unit time, λ , and the population standard deviation, σ , equal to $\sqrt{\lambda}$. This changes the equation for information content to that shown in eq (15) [25]. However, this

$$I(p,p_0) = \log_2 [(x_1 - x_0) / (\sqrt{2\pi\mu e})] \quad (15)$$

approximation is less valid for small values of lambda.

Trace Analysis

The model for trace analysis [7,8] is essentially the same as for quantitative analysis except that the output signal is often barely distinguishable from the background noise. In the first case, the information content of the component to be determined is less than or equal to the detection limit of the analytical method and the only conclusion is that the content is somewhere between zero and the detection limit. The *a posteriori* probability distribution is equal to the inverse of the detection limit of the method. The information content is given as the \log_2 of the ratio of the highest estimated content of the component to the detection limit for the component. In the second case where the content of the component to be determined is greater than the detection limit, the content can be determined quantitatively. The *a posteriori* probability distribution is a shifted log-normal distribution. This information content differs from the information content of the first case by the addition of $\log_2 [\sqrt{n_p} / k\sigma\sqrt{2\pi e}]$ term, where n_p is the number of parallel

determinations and k is the asymmetry parameter for the shifted log-normal distribution. If the mean value for the determination is close to the detection limit, a truncated Gaussian distribution is used to describe the *a posteriori* distribution. The information content can be calculated as a function of the highest estimated value, the mean value, the standard deviation, the frequency, and the distribution function. The information content for both the log-normal and the truncated Gaussian distributions converge to $\log_2 [x_1/\sigma\sqrt{2\pi e}]$.

Electrochemistry

Perone and coworkers have examined the effects of various experimental parameters on the qualitative information content of electrochemical data [26–31]. Because pattern recognition methods were used to obtain structural classification information empirically, an empirical measure of information gain was used to assess the effects of experimental parameters. This involved defining an appropriate figure of merit with which to measure the extent of informational goal achievement. Changes in information content are then determined by observing changes in attainment of the informational goal.

Byers, Freiser, and Perone [28,29] analyzed 45 compounds using cyclic staircase voltammetry. The data set consisted of 19 nitrobenzenes, 9 nitrodiphenyl ethers, and 17 ortho-hydroxy azo compounds. Of the nitrodiphenyl ethers, 4 were strong herbicides and 5 were either weak or nonherbicides. The informational goals for the problem were the classification of the 45 compounds by their structural type and the classification of the 9 nitrodiphenyl ethers according to herbicidal activity. Seven experimental variables, percent ethanol in the solvent, pH, surfactant concentration, number of cycles, scan rate, mercury drop hang time, and sampling time were varied in a fractional factorial design to generate a complete data base of collected cyclic staircase voltammograms and cyclic differential capacity curves [28] for subsequent analysis [29]. Faradaic and capacitive variable effect curves were calculated from the data. The average entropy for the three classes was 1.5 bits. The maximum entropy was 1.6 bits. The figure of merit for the informational goals was the percent correct classification as achieved by k -nearest neighbor analysis. For the structural characterization studies, the best overall percent classification ranged from 76% using capacitive variable effect features to 93% for both capacitive variable effect features and faradaic variable effect features. Overall accuracy for structural classification using the faradaic variable effect curve features ranged from 67% for percent ethanol to 93% for number of cycles. For the herbicidal prediction using variable effect curve features, the percent correct classification ranged from 78% for pH, faradaic sampling time, ca-

pacitive scan rate, and drop hang time to 100% for % ethanol, surfactant, faradaic number of cycles, and scan rate.

Barnes and Perone [30] studied the enhancement of chemical process information through experimental design. A simple model of controlled potential electrochemical processes based upon the Cottrell equation [31] was developed and implemented. The informational goal was to determine the effects of input voltage sequence, data collection fraction to analyze, and preprocessing scheme upon the determination of the diffusion coefficient. The figure of merit for goal achievement was based upon the least squares criterion function. Three input voltages, 180 mV step, pseudorandom binary sequence with peak voltage, $(E-E_0)$, of 180 mV, and white gaussian noise with mean of 90 mV and variance of 3 mV, were presented to the model. Comparisons of the identification results for the three inputs showed that the diffusion current model is fairly insensitive to the input sequence. Closer inspection of the model reveals that the anodic current is overwhelmed by the charging current. Therefore, the best input sequence for the model is that which is most easily generated.

In order to investigate the effects of timing, 3,000 data points corresponding to three milliseconds in time were generated using a step input. The least squares identifier was applied to 1 msec intervals of data both with and without the removal of charging current effects. When the charging current was present, the best results were obtained for the analysis of data taken after 10 time constants of the charging network. With the charging current removed, the best results were obtained with data taken within the first 1 msec interval when the amplitude of the anodic current and the signal to noise ratio are maximized. The best preprocessing scheme included filtering of unnecessary measurement components from the signal of interest, such as the removal of the charging current and signal averaging.

The application of information theory concepts to analytical chemistry can illuminate methods to increase the efficiency of chemical analysis. Early work shows encouraging promise for these types of applications. Optimum conditions have been established for obtaining structural, herbicidal activity, and diffusion coefficient information from voltammetric data. It has been demonstrated that the informational goal(s) will dictate the most favorable choice of experimental conditions. The use of objective systematic information enhancement methods can highlight experimental parameters that are often traditionally overlooked.

References

- [1] Young, John F., *Information Theory*, Butterworth: London, (1971).
- [2] Mathai, A.M., and P.N. Rathie, *Basic Concepts in Information Theory and Statistics*, John Wiley & Sons: New York, (1975).
- [3] Guiasu, Silviu, *Information Theory with Applications*, McGraw-Hill

- International Book Company: New York, (1977).
- [4] Pierce, John R., *An Introduction to Information Theory: Symbols, Signals, and Noise*, Second Edition, Dover Publications, Inc.: New York, (1980).
- [5] Shannon, C.E., *Bell System Technical Journal*, 27, 379 and 623 (1948).
- [6] Malissa, Conveners H., and J. Rendl, *Z. Anal. Chem.*, 272, 1 (1974) English version: I.L. Marr, *Talanta*, 22, 597 (1975).
- [7] Eckschlager, Karel, and Vladimir Stepanek, "Information Theory as Applied to Chemical Analysis", John Wiley & Sons: New York, (1979).
- [8] Eckschlager, Karel, and Vladimir Stepanek, *Anal. Chem.*, 54(11), 1115 (1982).
- [9] Ritter, S.L.; S.R. Lowery, H.B. Woodruff and T.L. Isenhour, *Anal. Chem.*, 48(7), 1027 (1976).
- [10] Grys, Stanislaw, *Z. Anal. Chem.*, 273, 177 (1975).
- [11] Eckschlager, Karel, *Anal. Chem.*, 49(8), 1265 (1977).
- [12] Danzer, K., *Z. Chem.*, 15, 326 (1975).
- [13] Danzer, Klaus, and Karel Eckschlager, *Talanta*, 25, 725 (1978).
- [14] Kaiser, H., *Anal. Chem.*, 42(2), 24A (1970).
- [15] Fitzgerald, J.J., and J.D. Winefordner, *Rev. Anal. Chem.*, 2(4), 229 (1975).
- [16] Yost, R.A., *Spectra*, 9(4), 3 (1983).
- [17] Fetterolf, D.D., and R.A. Yost, *Int. J. Mass Spectrom. Ion Processes*, 62, 33 (1984).
- [18] Boudreau, P.A., and S.P. Perone, *Anal. Chem.*, 51(7), 811 (1979).
- [19] Cleij, P., and A. Dijkstra, *Fresenius Z. Anal. Chem*, 298, 97 (1979).
- [20] Brillouin, L., "Science and Information Theory", Academic Press: New York, (1962).
- [21] Cleij, P., and A. Dijkstra, *Fresenius Z. Anal. Chem*, 294, 361 (1979).
- [22] Dupuis, Foppe, and Auke Dijkstra, *Anal. Chem.*, 47(3), 379 (1975).
- [23] Eskes Arie, Foppe Dupuis, Auke Dijkstra, Henri De Clercq, and Desire L. Massart, *Anal. Chem.*, 47(13), 2168 (1975).
- [24] van Marlen, Geert, and Auke Dijkstra, *Anal. Chem.*, 48(3), 595 (1976).
- [25] Eckschlager, K., *Coll. Czech. Chem. Commun.*, 41, 2527 (1976).
- [26] Perone, S.P., *ACS Symposium Series*, 265 (Comput. Lab.), 99 (1984).
- [27] Burgard, D., and S.P. Perone, *Anal. Chem.*, 50(9), 1366 (1978).
- [28] Byers, W. Arthur, and S.P. Perone, *Anal. Chem.*, 55(4), 615 (1983).
- [29] Byers, W. Arthur, B.S. Freiser, and S.P. Perone, *Anal. Chem.*, 55(4), 620 (1983).
- [30] Barnes, Freddie, and S.P. Perone, unpublished results, personal communication from Freddie Barnes, September 1984.
- [31] Bard, Allen J., and Larry R. Faulkner, "Electrochemical Methods, Fundamentals and Applications", John Wiley & Sons: New York, (1980).

DISCUSSION

of the Perone-Ham paper, Measurement and Control of Information Content in Electrochemical Experiments

Herman Chernoff

Statistics Center
Massachusetts Institute of Technology

The Shannon theory of information has had a profound impact in science and technology. Shannon defined information in terms of the reduction of uncertainty which, in turn, was measured by entropy. He was concerned mainly with the use of information to measure the ability to transmit data through noisy channels, i.e., channel capacity.

Statisticians have developed other, somewhat related, notions of information. In statistical theory, the major emphasis has been on how well experimental data help to achieve the goals in the classical statistical problems of estimation and hypothesis testing. These measures serve two useful functions. They serve to set a standard for methods of data analysis, methods whose efficiencies are measured in terms of the proportion of the available information that is effectively used. They also serve to design efficient experiments.

For the problem of estimation, Fisher introduced the Fisher Information which we now define. Suppose that it is desired to estimate a parameter θ using the result of an experiment which yields the data X with the density $f(x|\theta)$. The *Fisher Information* for θ corresponding to X is given by the matrix

$$J = I_X(\theta) = E_{\theta}(Y Y^T) \quad (1)$$

where Y is the *score function* defined by

$$Y = Y(X, \theta) = \frac{\partial [\log f_X(x|\theta)]}{\partial \theta} \Big|_{x=X} \quad (2)$$

If θ is a multidimensional vector, J is a nonnegative definite