Maximum efficiency pulse counting in computerized instrumentation

ARTICLE *in* ANALYTICAL CHEMISTRY · APRIL 1980

Impact Factor: 5.64 · DOI: 10.1021/ac50054a028

3 AUTHORS, INCLUDING:

G. Leroi
Michigan State University
112 PUBLICATIONS **2,495** CITATIONS
SEE PROFILE

Chris Enke
University of New Mexico
198 PUBLICATIONS **4,582** CITATIONS
SEE PROFILE

Maximum Efficiency Pulse Counting in Computerized Instrumentation

E. J. Darland, G. E. Leroi, and C. G. Enke*

Department of Chemistry, Michigan State University, East Lansing, Michigan 48824

A method of calculating the relative efficiency of pulse counting experiments is presented. The parameters which most affect the efficiency of multiple-point pulse counting experiments are (1) the criteria used to judge the quality of the data, (2) measurement parameters such as integration time and duty factor, (3) the strategies used to adjust the measurement parameters in response to changing experimental conditions, and (4) the range and distribution of the measured count rates. Under special circumstances, a fixed duty factor results in good efficiency. More often, however, varying the duty factor on a point-by-point basis results in a better overall efficiency by as much as a factor of two. The use of a fixed integration time almost always results in very poor efficiency if the signal-to-noise ratio of the data is used as a measure of its quality. Integration time adjustment strategies which result in near-unit efficiency under a wide variety of conditions are discussed, as are the practical aspects of implementing the improved strategies. The use of one "constant signal-to-noise" strategy is demonstrated.

The flux of a beam of particles such as photons, ions, or electrons can be measured a number of different ways. In many situations, the best method for measuring low-flux particle beams is the technique of pulse counting (also called photon counting, ion counting, etc.), in which the arrival rate of particles incident on an electron-multiplier-type transducer is measured by counting the discrete current pulses produced at the anode of the electron multiplier over a known length of time (*1*). Every pulse counting measurement involves a tradeoff between the quality of the data obtained and the duration of the experiment. When the measurement time is limited, it is important to be able to maximize the overall efficiency with which pulse counting experiments are made. The absolute minimum time in which a single pulse counting measurement of a given quality can be made depends on many factors, which have been extensively analyzed (*1–16*). Usually, however, a pulse counting experiment does not consist of a single measurement, but rather a series of measurements made as a function of wavelength, ion mass, or some other experimental variable. In this paper, we address the question of how to minimize the overall duration of such an experiment without sacrificing data quality. This question has been addressed only rarely and less extensively in the literature (*17*, *18*).

Minimization of the duration of a multiple-point experiment often requires the adjustment of the experimental parameters in response to changing experimental conditions. There are many possible approaches to making such adjustments, ("parameter adjustment strategies"), including those in which the adjustments are made (1) only once, (2) once at the beginning of each experiment, or (3) on a more frequent basis during the experiment. Until recently, practical considerations limited the options available in the last category. The flexibility and rapid decision making capability offered by computerized instruments, however, has now made it possible to implement more complex and adaptive parameter adjustment strategies.

In this paper, we demonstrate a method which can be used to predict quantitatively the efficiency of an experiment performed under stated conditions with a given adjustment strategy. We then compare several different strategies for varying (1) the integration time per point, t, and (2) the duty factor, F. This latter variable is defined as the fraction of the total time spent on a given point during which both signal and background (as opposed to only background) counts are accumulated. Through the examination of these specific cases, we find that the effectiveness of a given strategy depends on the experimental conditions and (even more dramatically) on whether the quality of the acquired data is measured by its absolute or relative precision. In the following three sections of this paper, we discuss the definitions and assumptions which are implicit in the rest of this work, present the relevant variance and signal-to-noise (SNR) expressions, and define several possible quantitative measures of efficiency. The importance of drawing a distinction between count total and count rate measurements is emphasized in the section on Variance and SNR Expressions. In the sections on Duty Factor and on Integration Time, we examine the effects of a number of possible parameter adjustment strategies on overall experimental efficiency. We then discuss some of the practical aspects of computer implementation of these strategies, and give an actual example of the advantages that can be gained with some of them.

ASSUMPTIONS AND DEFINITIONS

The Experiment. We assume that the analytical information of interest is encoded in the flux of a beam of particles and that this flux varies as a function of some controlled variable (such as wavelength, ion mass, electron energy, etc.). It is assumed that the cathode of an electron multiplier transducer (EMT) intercepts a representative and constant portion of the particle beam, and that the EMT and succeeding pulse counting electronics convert some fraction of the resulting collisions into pulses which are counted by an electronic counter. We shall use the symbol R_s for the true average rate at which these "signal" pulses arrive at the input to the counter. The counter is assumed to respond not only to the signal pulses, however, but also to "background" pulses from an independent source which need not be specified. If the true average background count rate is R_b, the average total count rate, R_t, is

$$R_t = R_s + R_b \tag{1}$$

We assume that it is possible (by closing a shutter or changing some other aspect of the instrument) to count background pulses alone. The experiment is defined to be the determination of R_s at a number of specific values of the controlled variable, and each such determination is defined to be one "point". We also assume that the computer can, under software control, clear and read the contents of the counter, control the total integration time at each point, and control the fraction of that time spent measuring signal-plus-background and background alone.

Efficiency. There are a number of quantities which can be used to measure the overall efficiency of the experiment: the number of points of a given quality that can be acquired

0003-2700/80/0352-0714$01.00/0

in a given time, the quality of a given number of points acquired in a given time, or the inverse of the time necessary to acquire a given number of points of a given quality. Which one of these related measures of efficiency is most appropriate depends on whether the quantity of data, quality of data, or experiment duration is least constrained by other considerations. We will adopt the "inverse time" method for this work, but postpone more quantitative definitions of efficiency until later.

The major factors which affect the efficiency of the experiment are (1) the criteria used to judge the quality of the data, (2) the values of R_t and R_b, (3) the measurement parameters such as integration time and duty factor, and (4) the strategies which control the adjustment of these parameters.

Data Quality. The relative precision of experimental data is often the most useful measure of its quality. An example is when each component of a mixture must be determined to within a certain percentage of its concentration. However, in some cases, such as when each mixture component must be determined to within a given percentage of the total sample, the absolute precision of the data is more useful. We will examine both the cases where the absolute precision, represented by the variance, and the relative precision, represented by the signal-to-noise ratio, are used as measures of data quality. We assume that the analyst can specify some minimum quality which the data must meet or exceed. Maximum efficiency clearly demands that no points be acquired with a greater-than-necessary quality and we assume, for most of this paper, that all points in a given experiment are to be acquired with the same quality. Instances where this may not be desirable are discussed in the section on Practical Aspects of Implementing the Chosen Strategies.

Count Rates. Equations for pulse counting variance and SNR show clearly that efficiency can be increased if R_s is increased or R_b is decreased. We will assume that all instrumental parameters (e.g., slit width in optical spectrometry) have been adjusted to provide optimum signal-to-background ratios, consistent with other experimental constraints (e.g., required optical resolution), and that pulse counting measurement system parameters (e.g., EMT voltage, discriminator setting) have also been optimized (*2*, *3*).

There are many possible sources of background pulses, and it has become fairly common to treat background pulses from each source separately. None of these results presented here, however, depend on whether single or multiple sources contribute to the overall background count rate so long as each determination of R_b is in fact a single measurement and not a computed sum of the background count rates due to separate sources determined in separate measurements. We assume that R_b may be less than, equal to, or greater than R_s, but that R_t is always low enough that negligible pulse overlap counting loss occurs.

Measurement Parameters. Each determination of the signal count rate requires subtraction of the results of two separate measurements. We define, *for each point*:

τ_b total time period during which only background pulses are counted, s

N_b total number of background counts recorded in τ_b seconds

τ_t total time period during which both background and signal pulses are counted, s

N_t total number of signal and background counts recorded in τ_t seconds

t total count integration time for the point, s $\tau_t + \tau_b$

F duty factor, or fraction of the total time spent measuring the total count rate, dimensionless $\tau_t/(\tau_t + \tau_b)$.

The observed background and total count rates, R_{ob} and R_{ot}, are calculated from the measured quantities at

$$R_{ob} = \frac{N_b}{(1 - F)t} \tag{2}$$

$$R_{ot} = \frac{N_t}{Ft} \tag{3}$$

and, finally, the observed signal count rate, R_{os}, is calculated as

$$R_{os} = R_{ot} - R_{ob} \tag{4}$$

We have chosen to adopt t and F as the measurement parameters which are to be varied in response to changing count rates in such a way as to minimize total experiment duration and to maintain the desired data quality.

Noise Sources. The accuracy with which the observed count rates R_{ot}, R_{ob}, and R_{os} reflect the true pulse arrival rates is adversely affected by the presence of noise in the measurements. We assume that both signal and background count rates suffer from random fluctuations (quantum noise) which follow Poison statistics. We also assume that if "$1/f$" noise (or "excess noise", or "drift") exist in the background count rate, its effects have been avoided by summing alternate, short measurements of the total and background count rates, rather than making a single, longer measurement of each (*9*).

Our treatment assumes all other potential noise sources are absent. This is likely to be a good assumption for signal and background flicker noise in many (though not all) experiments, and an even better assumption for other types of noise which can be virtually eliminated in high quality digital pulse counting systems (*6*).

VARIANCE AND SNR EXPRESSIONS

The Ideal Case. The discussion in following sections of this paper requires the use of variance and SNR equations which are not readily found in the pulse counting literature. Before presenting these equations, we wish to make a clear distinction between the variance of count *total* and count *rate* measurements. Let us define the "ideal case" as that in which no background pulses are present and the only noise source is quantum noise in the signal pulse arrival rate. In this case, the total number of counts expected in a measurement which has a duration of t seconds is

$$N_t \approx R_s t, \ (R_b = 0) \tag{5}$$

The variance in the count total measurement of this ideal case is

$$\sigma_N^2 = N_t \approx R_s t, \ (R_b = 0) \tag{6}$$

and the SNR of the count total measurement is then

$$\mathrm{SNR_N} = \frac{N_t}{\sigma_N} = N_t^{1/2} \approx (R_s t)^{1/2}, \ (R_b = 0) \tag{7}$$

The observed signal count rate in this case is calculated as

$$R_{os} = N_t/t, \ (R_b = 0) \tag{8}$$

The variance in the count rate measurement can be found by applying standard propagation of error techniques to Equation 8 and assuming the relative error in t to be much less than the relative error in N_t:

$$\left(\frac{\sigma_R}{R_{os}}\right)^2 = \left(\frac{\sigma_{N_t}}{N_t}\right)^2 + \left(\frac{\sigma_t}{t}\right)^2 \tag{9}$$

$$\sigma_R^2 = \frac{N_t}{t^2} \approx \frac{R_s}{t}, \ (R_b = 0) \tag{10}$$

Therefore, the count rate SNR is

$$\mathrm{SNR_R} = \frac{R_{os}}{(R_s/t)^{1/2}} \approx (R_s t)^{1/2}, \ (R_b = 0) \tag{11}$$

Table I. Count Rate Variance and SNR Expressions

case	σ^2 =		SNR =	
ideal ($R_b = 0$) ($F = 1$)	$\frac{R_s}{t}$	(10)	$(R_s t)^{1/2}$	(11)
non-ideal ($R_b = 0$) (F = arbitrary)	$\frac{\left[R_s + R_b\left(\frac{1}{1-F}\right)\right]}{Ft}$	(12)	$\frac{R_s(Ft)^{1/2}}{\left[R_s + R_b\left(\frac{1}{1-F}\right)\right]^{1/2}}$	(13)
optimum non-ideal ($R_b = 0$) ($F = F_{optimum}$)	$\frac{[(R_s + R_b)^{1/2} + R_b^{1/2}]^2}{t}$	(14)	$\frac{R_s(t)^{1/2}}{(R_s + R_b)^{1/2} + R_b^{1/2}}$	(15)

Note that the SNR expressions for count total and count rate measurements (Equations 7 and 11) are identical. Variance expressions 6 and 10 however, predict opposite changes in the variance of the count total and the count rate as the integration time is increased. Similarly, all of the SNR expressions given in this paper are valid for both count total and count rate measurements, but the variance expressions to be presented are valid for count rate measurements only.

The Non-Ideal Case. Commonly used expressions for SNR assume either that equal amounts of time are spent measuring signal and background (i.e., $F = {}^1/_2$) or that the background count rate is known with infinitely better precision than the signal count rate (a situation which can only be attained if all the time is used to measure the background, i.e., $F = 0$). More general expressions, derived elsewhere (see supplementary material), are listed in Table I as Equations 12 and 13.

The Optimum Case. The duty factor which allows data of a given quality to be obtained in the least amount of time, F_{opt}, can be found by setting the partial derivative of either the variance or SNR with respect to F equal to zero. The same results are obtained in either case, namely

$$F_{opt} = \frac{1}{1 + \left(\frac{R_b}{R_s + R_b}\right)^{1/2}} \tag{16}$$

which yields the expected result that $F_{opt} = 0.5$ when $R_s \ll R_b$. If the expression for F_{opt} is inserted into Equations 12 and 13, Equations 14 and 15 (see Table I) result. In the ideal situation where $R_b = 0$, Equations 14 and 15 reduce to Equations 10 and 11, respectively. When background is present, however, Equations 14 and 15 define the best possible tradeoff between data quality and integration time for given signal and background count rates.

QUANTITATIVE MEASURES OF EFFICIENCY

Ideal Reference. In order to provide a quantitative measure of efficiency of the usual sort (i.e., one which ranges from zero to one), it is necessary to adopt some standard which is used to define unit efficiency. It is often convenient to adopt the ideal case defined in the section on Variance and SNR Expressions as such a standard. In order to do so, we define the ideal, non-ideal, and optimum integration times (t_{ideal}, $t_{non\text{-}ideal}$, and t_{opt}) for each point in an experiment as the integration periods necessary to acquire data of just the desired quality in each of the three cases defined in the section on Variance and SNR Expressions. These integration times are obtained from Equations 10–15 in Table I.

The efficiency of a single pulse counting measurement is thus given by

$$E = \frac{t_{ideal}}{t_{non\text{-}ideal}} = \frac{F(R_s/R_b)}{(R_s/R_b) + \left(\frac{1}{1-F}\right)} \tag{17}$$

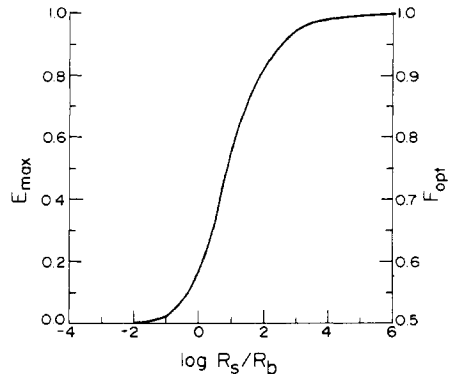

Figure 1. Optimum duty factor and maximum possible efficiency (relative to the ideal case, for single measurements) vs. signal-to-background count ratio

and the maximum possible efficiency for a measurement at the same count rates is given by

$$E_{max} = \frac{t_{ideal}}{t_{opt}} = \frac{R_s/R_b}{\left[\left(\frac{R_s}{R_b} + 1\right)^{1/2} + 1\right]^2} \tag{18}$$

The same expressions result whether variance or SNR is used to measure data quality. It can be shown that $E_{max} = 2F_{opt} - 1$. Both E_{max} and F_{opt} are plotted as a function of R_s/R_b in Figure 1.

Note that E_{max} depends neither on the desired data quality nor on the absolute signal and background count rates, but only on R_s/R_b (a quantity hereafter referred to as the signal-to-background ratio, SBR). This is a characteristic of all of the various measures of efficiency used in this paper, which greatly simplifies the discussion in following sections. It should not be forgotten, however, that the actual duration of the experiment *does* depend on these quantities. Optimization of count rates and choice of reasonable quality goals thus remain as important parts of the experiment, even though not considered here. In fact, Figure 1 demonstrates graphically the importance of reducing R_b as much as possible. For instance, when $R_s = R_b$, E_{max} is only 0.172. Thus, to reach the desired data quality under these conditions requires integration times nearly six times as long as when $R_b = 0$ (and more than three times as long as when $R_b = 0.1\ R_s$).

Optimum Reference. The use of the ideal case as the reference for quantitative measures of efficiency is thus useful when considering the effects of varying the signal and background count rates. Once the count rates have been optimized, however, it is more useful to compare the total duration of an experiment to the shortest duration possible under the circumstances, rather than to the non-attainable ideal. For the remainder of this paper, therefore, we define the *relative efficiency*, RE, of an experiment as

$$RE = \frac{\sum t_{opt}}{\sum t_{non\text{-}ideal}} \quad (19)$$

where the summations run over all points acquired in the experiment. For the special case of an experiment which consists of a single point (or of multiple points all of which have identical count rates) the relative efficiency is given by

$$RE = \frac{F\left[\left(\frac{R_s}{R_b} + 1\right)^{1/2} + 1\right]^2}{\left(\frac{R_s}{R_b}\right) + \left(\frac{1}{1-F}\right)}, \text{ (single point)} \quad (20)$$

Thus, for a single point, the relative efficiency depends only on the duty factor and the signal-to-background ratio. Figure 2 shows a plot of RE vs. SBR for several values of F.

It is generally not possible to write such a simple expression for the relative efficiency of multiple point experiments in which count rates vary, because of the number of additional factors involved. These factors, discussed further in the next two sections, include not only the strategies used to adjust F and t, but also the definition of data quality, the signal and background count rates, the number of points, and the distribution of these points within the range of count rates.

The effects of count rate could be discussed in terms of R_s and R_b, but the results are more generally applicable if the signal-to-background ratio is so used. While it is convenient to assume that R_b is constant during an experiment and variation in the SBR is due only to variation in R_s, the results presented here are equally valid if either or both R_b and R_s vary.

DUTY FACTOR

There are many possible approaches to the choice of the duty factor(s) to be used in a pulse counting experiment. In this section, we consider three different adjustment strategies: (1) A single duty factor is used for all measurements; (2) a single duty factor is used for all measurements in a given experiment, but the duty factor is varied from experiment to experiment in response to changing experimental conditions; (3) the duty factor is varied on a point-by-point basis so as to acquire each point in the optimum time.

Which of these approaches is best for a particular application depends not only on how well each minimizes the total data integration time of an experiment (i.e., on how well each maximizes the relative efficiency), but also on the extra time required to implement each strategy. Differences in implementation time will be due primarily to differences in the amount of time required to decide on the duty factor to be used. In general, this decision must be made more frequently (but will require less time) for the third of the approaches listed above. A more detailed comparison of these strategies is postponed until the end of this section.

The Best Single Duty Factor. It is clear from Figure 2 that if all points to be acquired have either very low or very high SBRs, the choice of the best single duty factor is straightforward. In all other cases, the choice of the best duty factor is more complicated and depends on the count rates, the distribution of SBRs among the points, the definition of data quality, and so forth. If all such parameters can be specified, then the relative efficiency for the experiment using a given duty factor can be calculated using Equation 19 and the appropriate expressions from Table I. The maximum of a relative efficiency vs. duty factor curve will then reveal the single best duty factor for an entire experiment under the given conditions.

Unfortunately, it is often impossible to completely specify all experimental parameters prior to the experiment. In this case, any non-arbitrary choice of duty factor must rely on the existence of a set of guidelines for choosing F_{best}, given estimates of at least some of the experimental parameters. One such set of guidelines is presented in the following paragraphs. It was developed by plotting a series of RE vs. F curves (see supplementary material) in which one parameter at a time was systematically varied. Examination of the resulting families of curves showed that under some readily defined sets of conditions (see below), one parameter exerts a controlling influence on the value of F_{best}, which can be easily calculated. In other cases, no simple calculation is possible; instead F_{best} must be chosen by examining a further set of RE vs. F curves which are carefully tailored to fit the experimental conditions likely to be encountered.

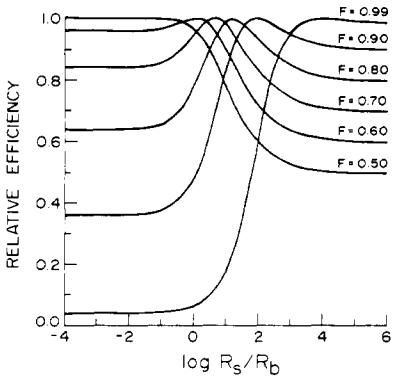

Figure 2. Relative efficiency vs. signal-to-background count ratio for single-point measurements at various duty factors

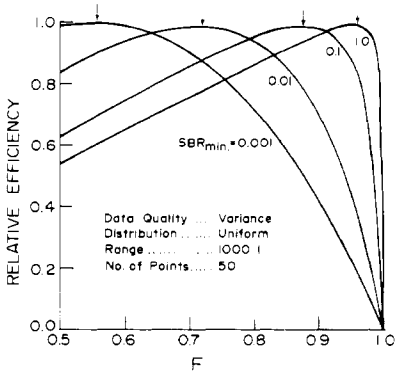

Figure 3. Relative efficiency vs. duty factor for multiple-point experiments with a fixed duty factor when data quality ≡ variance. Arrows indicate the optimum duty factor at the *mean* SBR

For much of the following discussion we assume that the points are uniformly distributed over the SBR range. Many experimental count rate distributions, however, are quite different, and have a predominance of points with lower- or higher-than-average SBRs. The effects of such distributions (referred to as "low-peaked" and "high-peaked" distributions) are also discussed. Further, the SBR range is defined to be the *ratio* (not the difference) of the maximum and minimum values of SBR encountered in an experiment (SBR_{max}/SBR_{min}). The count rates are then specified completely by the range, the form of the distribution, and either SBR_{min} or SBR_{max} (it is usually most convenient to use SBR_{min} when quality = SNR, and to use SBR_{max} when quality = variance).

Quality = Variance. Figure 3 shows the effect of varying SBR_{min} while keeping the range and all other parameters constant when data quality is measured by the variance of the data. This figure shows, as might be expected, that for a given range and distribution, an experiment with high count rates is best performed with a higher duty factor than one with lower count rates. Similar plots of RE vs. F, in which the range is varied while either SBR_{min} or SBR_{max} is held constant (see

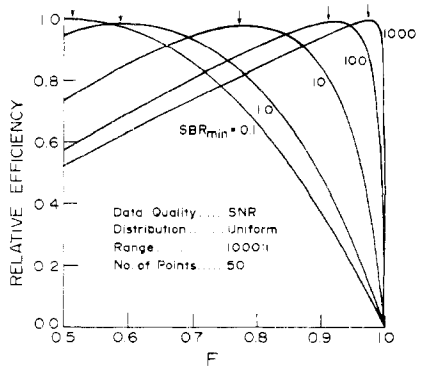

Figure 4. Relative efficiency vs. duty factor for multiple-point experiments with a fixed duty factor when data quality ≡ SNR. Arrows indicate the optimum duty factor at the *minimum* SBR

supplementary material) show that SBR_{max} has a much greater effect on the value of F_{best} than does SBR_{min}. This is consistent with the fact that points at high count rates require longer to integrate to a given variance than do points with lower count rates. Therefore, other things equal, the upper end of the SBR range exerts much more control on the value of F_{best} than does the lower end. It is interesting to note that if the SBR range is varied while keeping the *mean* SBR constant, F_{best} does not vary at all. Furthermore, F_{best} is accurately predicted by the value of F_{opt} at the mean SBR for each of the curves in Figure 3 (arrows).

The fact that F_{best} equals F_{opt} at the mean SBR is only fortuitous, and holds only for a uniform distribution. Plots of *RE* vs. *F* for various low-peaked distributions (see supplementary material) show considerable variation in F_{best} with range, no matter whether SBR_{min}, SBR_{max}, or the mean SBR is held constant. In these cases, the high SBR points will require longer integration times to reach a given quality than the low SBR points, but there are more of the latter so that the high SBR points no longer exert a controlling influence on the value of F_{best}. In fact, if the point distributions were sharply enough peaked at the low end of the range, one would expect F_{opt} at SBR_{min} to be a good predictor of F_{best}. Conversely, for any distribution which peaks above the midpoint of the SBR range, F_{best} can be easily and fairly accurately estimated by F_{opt} at SBR_{max}.

The total number of points acquired during an experiment may or may not affect the value of F_{best}. The number of points has no effect for a uniform distribution. For other distributions, it has no effect for certain combinations of range and SBR_{max}, while for other combinations it has quite a significant effect. We have not been able to find any simple rule to determine when and how the number of points will have an effect, and thus recommend that this factor be checked under the conditions which pertain.

Quality = SNR. Calculations in which SNR is used to measure data quality lead to conclusions which are similar to those discussed above in most ways: experiments with high count rates require high duty factors; the exact value of F_{best} is modified similarly by the distribution of points within the range of SBR values; and the number of points may or may not have an effect.

The differences between the two cases can be traced to the fact that when SNR is used to measure quality, the point with the *lowest* SBR takes longest to reach the desired quality. Comparison of Figures 3 and 4, for instance, reveals that merely switching definitions of quality from variance to SNR results in a much lower value of F_{best}, all other factors being equal. Similarly, F_{best} changes little with range while SBR_{min} (rather than SBR_{max}) is held constant, for a uniform distribution. This is in general true for any low-peaked distribution, and for any of these distributions, F_{opt} at SBR_{min} (arrows in Figure 4) is a good estimate of F_{best}. Distributions with higher peaks counteract the dominance of SBR_{min} to a greater or lesser extent, depending on their detailed shape.

Guidelines. As a result of these considerations we offer the following guidelines for choosing the best single duty factor for an experiment. If all points have very low SBRs (<0.3), a duty factor of 0.5 will result in a relative efficiency of at least 0.99. If all points have very high SBRs (>3000), a duty factor of 0.99 will accomplish the same result.

If data quality = variance and the data have a uniform or higher peaked distribution, F_{best} can be estimated by F_{opt} at SBR_{max}. If almost all points fall at the low end of the SBR range, F_{opt} at SBR_{min} may be used to estimate F_{best}.

If data quality = SNR and the data have a uniform or lower peaked distribution, F_{best} can be estimated by F_{opt} at SBR_{min}. If almost all points fall at the high end of the SBR range, F_{best} is given approximately by F_{opt} at SBR_{max}.

In all other cases, no simple rules can be given, and a series of calculations such as those presented above is necessary to find F_{best}. Fortunately, it is often possible to find some duty factor which will result in a relatively high efficiency for a wide range of conditions. For example, for one low-peaked distribution, a duty factor of 0.72 was found to give a relative efficiency greater than 0.80 even when the number of points, range, SBR_{min}, and SBR_{max} all varied by three orders of magnitude (see supplementary material).

Comparison of Possible Strategies. From the standpoint of minimizing the total integration time of each experiment, variation of the duty factor on a point-by-point basis is obviously the best strategy, since it guarantees that every point in every experiment will be acquired with optimum efficiency. (See the section on Practical Aspects of Implementing the Chosen Strategies for a discussion of how this may be accomplished in practice.) Simpler strategies may be just as good for particular experimental conditions, however. For instance, if all points in all experiments are known to have very low SBRs, the use of a single duty factor (0.5 in this case) will also result in a relative efficiency of nearly unity.

Often, the choice of a single duty factor will require a series of calculations similar to those described above but tailored more closely to the actual experimental conditions (all of which must then be known in advance). If the duty factor is to be varied on an experiment-by-experiment basis, the time required to choose the duty factor for each experiment must be considered.

In the case of point-by-point adjustment of duty factor, no pre-experiment calculations or knowledge of conditions is required. Instead, a calculation (albeit a very simple one) is required for every point in every experiment. This approach is thus only practical for computer-controlled instrumentation.

If one adopts the criterion that the duty factor adjustment strategy which allows a number of experiments to be performed in the least overall time is the "best" strategy, the following recommendations can be made: (1) If all experiments will be run under virtually the same conditions, and those conditions are such that F_{best} is easily chosen by the guidelines presented above, then the single *F* strategy is best. (2) If conditions differ from experiment to experiment, but F_{best} is easily determined for all the experiments, then experiment-by-experiment adjustment of *F* is probably best. (3) If the choice of F_{best} is difficult for any of the experiments to be performed, then it will probably pay to invest the programming time necessary to implement point-by-point adjustment of the duty factor. Once implemented, this method will guarantee near-optimum efficiency for all future experiments.

INTEGRATION TIME

In this section, we compare five different integration time

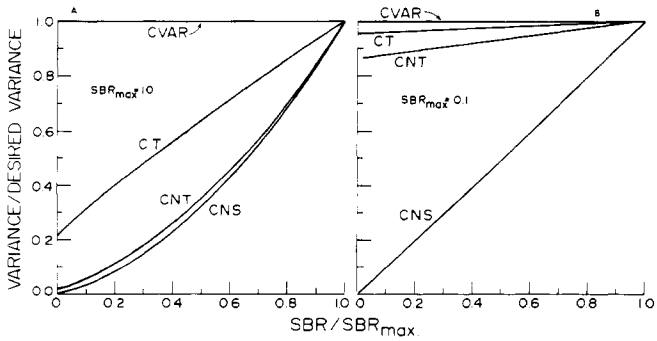

Figure 5. Actual variance produced by four integration time adjustment strategies for points at various count rates when data quality = variance

adjustment strategies. (1) The most common strategy is to use the same integration time, t, for every point in an experiment, with t chosen so as to guarantee that every point will be acquired with at least the minimum desired quality. We will refer to this as the constant integration time approach, and abbreviate it CT. At the other extreme are the two "exact" strategies which adjust the integration time on a point-by-point basis so as to ensure that each point is acquired with just the desired quality; these are (2) the constant SNR (CSNR) and (3) the constant variance (CVAR) approaches. In addition, we will examine two other approximate strategies in which each point is integrated just long enough to acquire some constant number of counts, with the number of counts specified so as to ensure that the most difficult point (i.e., that with the lowest SBR if quality = SNR, that with the highest SBR if quality = variance) is acquired with the desired quality. The counts considered may be either the signal counts only, or the total of signal and background counts; the two strategies will be referred to as (4) the constant signal count (CNS) and (5) the constant total count (CNT) strategies. For the sake of simplicity, we assume throughout this section that the optimum duty factor is used for every point.

Quality = Variance. When variance is used as a measure of the quality of the data, the CVAR, CNS, CNT, and CT strategies all measure points at SBR_{max} with exactly the desired variance, by definition. The quality with which points of lower SBR are acquired using the different approaches is shown in Figure 5 for two different sets of experimental conditions. All three approximate strategies produce lower (i.e., better than necessary) variances to varying degrees. The integration times actually used for points with various SBRs are shown in Figure 6 for conditions intermediate between those of Figures 5A and 5B. Points which have less-than-maximum SBR are seen to *require* less integration time than those at SBR_{max}, but the *actual* integration times using CNS or CNT are actually higher. Note that as the SBR approaches zero, the integration time for CNS approaches infinity, but the integration time for CNT approaches a finite limit since background counts contribute to the necessary count total.

As Figures 5 and 6 show, the points acquired at SBR_{max} take the most time when the exact measurement strategy is used, whereas these points take the least (or at best the same) time when the approximate strategies are used. Hence, the relative efficiency with which a given point is acquired increases with SBR for all three approximate strategies. It is immediately apparent that the effectiveness of these four measurement strategies declines in the order CVAR > CT > CNT > CNS. All three approximate methods are better at low SBR_{max}, but none approximates the exact method very closely except when the signal count rate is less than the background count rate for all points.

More quantitative estimates of the overall relative efficiency produced by these measurement strategies under different conditions can be readily calculated. For example, Figure 7

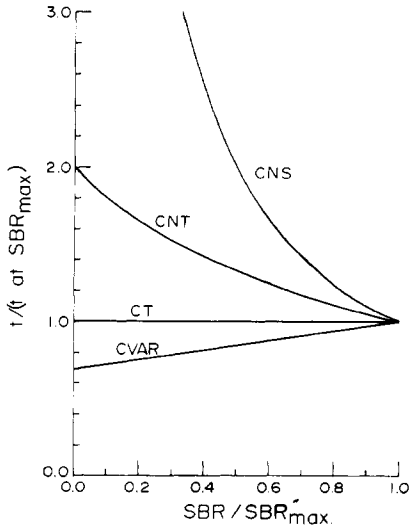

Figure 6. Integration times used by four integration time adjustment strategies for points at various count rates when data quality = variance and SBR_{max} = 1.0

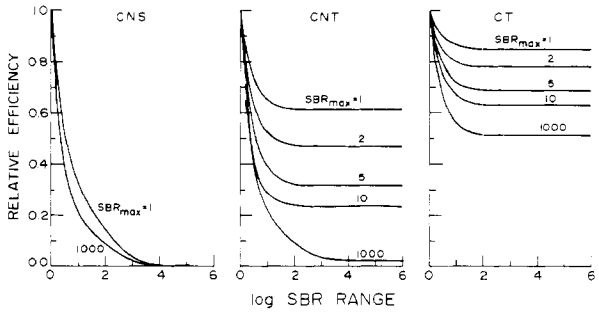

Figure 7. Relative efficiency vs. SBR range for various values of SBR_{max} for three approximate "constant variance" integration time adjustment strategies. Uniform distribution, 50 points

shows the relative efficiency as a function of SBR range for several values of SBR_{max} for each of the approximate strategies. Note that each curve represents a fixed SBR_{max}, so that an increasing range means a decreasing SBR_{min}. The decrease in relative efficiency with increasing range is due to the fact that, for a fixed number of points, as SBR_{min} decreases, Σt_{opt} also decreases while $\Sigma t_{non\text{-}ideal}$ either remains constant or increases (see Figure 6). The efficiency attained with the CNS strategy approaches zero as the range approaches infinity because this strategy requires an infinite integration time at SBR_{min} = 0; the other strategies all have non-zero limits at high ranges.

As might be expected, analogous plots for high-peaked distributions (supplementary material) look very similar except that the decrease in relative efficiency with range is more gradual and (except for CNS) approaches a higher limit at high ranges. The behavior for low-peaked distributions (supplementary material) is just the opposite.

The effect of the number of points acquired on the relative efficiency is very minor so long as this number is greater than about 50–100. For fewer points, the relative efficiency is often quite sensitive to the number of points acquired. These effects are illustrated in the supplementary material.

Quality = SNR. When the quality of the data is measured by its SNR, the points acquired at SBR_{min} have, by definition, RE = 1 for CSNR and the three approximate strategies. At all higher SBRs, the approximate strategies result in acquisition of data with better than necessary quality in varying degrees, as shown by Figure 8 (note differing ordinate scales). Integration time vs. SBR curves for one set of conditions are shown in Figure 9. These figures show that points at low

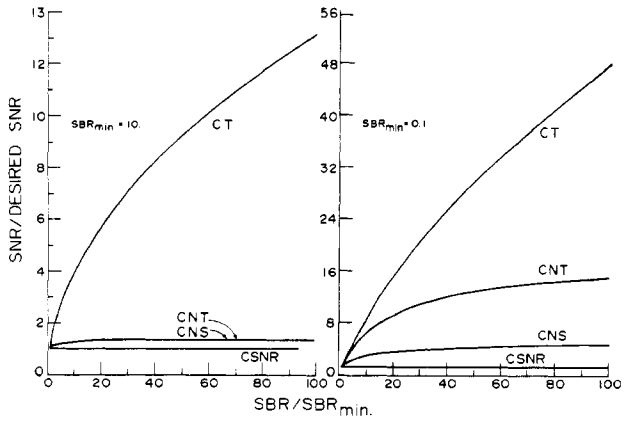

Figure 8. Actual SNR produced by four integration time adjustment strategies for points at various count rates when data quality = SNR

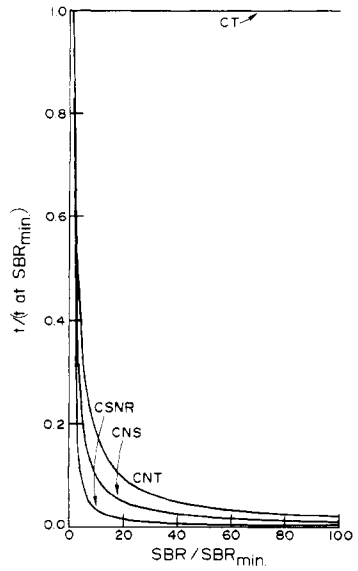

Figure 9. Integration times used by four integration time adjustment strategies for points at various count rates when data quality = SNR and SBR_{min} = 1.0

SBRs are acquired with higher relative efficiency than those at higher SBRs, no matter which measurement strategy is used. High SBR points, however, require shorter (or, for CT, the same) integration times. This is in sharp contrast to the case when quality = variance, where the points acquired with the best RE required the least time. Also by way of contrast, the effectiveness of these measurement strategies is expected to decrease in the order CSNR > CNS > CNT > CT, and all approximate strategies should be better at high values of SBR_{min}.

These expectations are borne out by the data of Figure 10, which shows plots of relative efficiency vs. range for various strategies and values of SBR_{min}. The *RE* produced by the approximate methods is seen to be more sensitive to the SBR range than when quality = variance. The curves for the CT strategy all decrease monotonically with increasing range (for a fixed number of points), eventually reaching a limit of $1/N$, where N is the number of points. This behavior is easily explained if the trends in Figures 8 and 9 are considered. In the case of the CNS and CNT strategies, however, the relative efficiency curves once again approach *RE* = 1 at high ranges. This is due to the fact that when the range is increased sufficiently, most of the (high SBR) points take very little time compared with the point at SBR_{min}, even though each is acquired with a low relative efficiency. When the background count rate is fairly high, however, (SBR_{min} is low), CNT is not

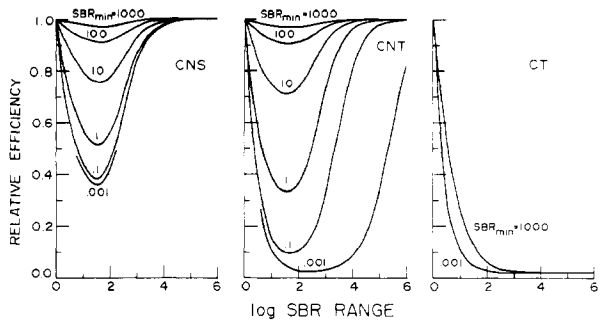

Figure 10. Relative efficiency vs. SBR range for various values of SBR_{min} for three approximate "constant SNR" integration time adjustment strategies. Uniform distribution, 50 points

nearly as good an approximation to CSNR as is CNS, and thus the relative efficiency is lower and a wider range is required to counteract the initial tendency of the *RE* to decrease with range.

Non-uniform distributions produce *RE* vs. range curves (supplementary material) which are quantitatively different than those in Figure 10, but qualitatively similar. A low-peaked distribution, for instance, results in a greater number of points with higher relative efficiencies so that the initial drop in *RE* with range is slower than with the uniform distribution. The ultimate limit, however, is still the same ($1/N$ for CT, 1 for CNS and CNT), no matter how asymmetric the distribution. High-peaked distributions create just the opposite effects: a faster initial drop and a quicker approach to the (same) high-range limits.

The relative efficiency is also influenced by the number of points acquired (supplementary material). In general, the *RE* declines as N is increased for all three approximate strategies. Eventually a limit is reached where so many points are acquired that the sums in Equation 19 become good approximations to the corresponding integrals. The number of points required to reach that limit is about 50–100 for the CT strategy, depending on other conditions. In the cases of CNT and CNS strategies, the competing effects of single point *RE* and integration times make both the number of points required to reach the limiting *RE* and the amount of decrease in *RE* much more sensitive to other conditions. For example, in the case where CNS is used with a uniform distribution in which SBR_{min} = 10 and the range is 4:1, the limiting *RE* of 0.85 is essentially achieved for $N \gtrsim 50$. If the range is increased to 1000:1, the *RE* remains higher than for the small range case as long as $N \leq 110$, but then continues to decline at higher N; the limiting *RE* ($\sim$0.65) is not reached until $N \geq 5000$.

Choice of Adjustment Strategy. Because of the large number of relevant variables, it is impractical to attempt to list all sets of conditions under which the increased efficiency of a given strategy is likely to be worth more than the costs of implementing that method. It is better to first define the experimental conditions likely to be encountered, then use graphs such as those presented here and in the supplementary material to estimate the relative efficiencies to be expected using each strategy, and finally decide whether the possible gains outweigh the likely costs at one's particular facility. We can, however, provide some general guidelines on the basis of the results of this work.

When quality = variance, the CNS, CNT, and CSNR methods can all be eliminated from consideration, since they all produce worse results than the CT method and are more complicated. If an overall relative efficiency of 0.5 is satisfactory, it can be seen from Figure 7 and the supplementary material that there are many sets of conditions when the CT strategy is likely to be quite adequate. If instrument time or sample stability is at a premium, however, then it is likely that

a near-maximum relative efficiency will be desirable, and only rarely will the CT strategy be good enough.

When quality = SNR, the CT strategy produces very poor relative efficiencies except when the range is very small. Therefore, it is likely that one of the variable strategies will be worthwhile. Any of these (CNT, CNS, CSNR) can produce dramatic increases in relative efficiency under many conditions, as demonstrated in Figure 10 and the supplementary material. Of these approaches, the CNT method requires the least programming effort and results in the least amount of time spent on calculations during the experiment, though the differences are often not great. The other two methods are more likely to produce better relative efficiencies over a wider range of conditions, however. Because of the flexibility of computer-based instrumentation, it is possible to change from one strategy to another relatively easily and quickly. Thus, the user can first implement the simplest "good" strategy (CNT) and then add others later, if and when they are found to be needed.

PRACTICAL ASPECTS OF IMPLEMENTING THE CHOSEN STRATEGIES

Thus far we have assumed implicitly that SBR_{min} and SBR_{max} represent not only the extremes encountered during an experiment, but also the extreme count rates of data about which the investigator cares. This is quite often not true, a fact which has important implications for the actual implementation of most of the adjustment strategies discussed. In this section, we illustrate the potential problems and suggest possible solutions only for the case in which the data quality is specified by its SNR; the situation when quality = variance is quite analogous.

To illustrate one potential problem, assume that a spectrum is to be acquired which consists of a series of peaks or bands superimposed on a weak, flat base line which corresponds either to zero signal or to a constant background. In this case, the analyst may be interested only in points which lie close enough to peak maxima to be of use in calculating peak height or area, or perhaps in all points which lie above the base line by some small amount which can be called the "threshold of interest". Unfortunately, the points which take the longest to acquire for a given SNR are just those points in which the analyst is not interested; thus the use of a true CSNR, CNS, or CNT strategy is likely to waste a great deal of time. Practical adjustment strategy algorithms should therefore include provisions for recognizing points which are not of interest before much time is spent on them. Such points may be acquired simply with whatever quality has been reached at the time the point is recognized as being uninteresting. This is the most common approach to this problem (*17*, *18*), but others are possible. For instance, all points below some threshold might be acquired with a lower, constant SNR. Multistep implementation might also prove useful: the simplest such algorithm would be one in which all points above a pre-selected threshold are acquired with a certain SNR, points between that threshold and a lower threshold are acquired with another SNR, and points below the second threshold are ignored. Recognition of uninteresting points can be accomplished in a number of ways, but the most straightforward approach is to simply perform brief, preliminary measurements, and to decide on the basis of those measurements whether to integrate longer, and how much longer, at a particular point. It is important to remember that quantum noise in the preliminary measurements of the signal and background count rates can lead to substantial errors in the calculation of the integration time necessary to reach the desired SNR. The adjustment strategy algorithm should therefore make allowances for the uncertainty in the preliminary measurements.

The quantum noise problem also arises if one wishes to adjust the duty factor on a point-by-point basis, since knowledge of the SBR at each point is necessary in order to calculate the optimum duty factor for that point. Figure 2 shows that the "optimum" duty factor calculated from a measured (and possible inaccurate) SBR will produce a relative efficiency greater than 0.95 as long as the true SBR is within a half an order of magnitude of the measured value. Thus, the preliminary measurement need not be terribly accurate. However, as in the case of integration time adjustment strategies, the algorithm used must take this problem into account, or the advantages of the "optimum" measurement strategy may not be realized in practice.

A PRACTICAL EXAMPLE

An illustration of the application of many of the concepts discussed above is provided by the photoionization mass spectrometer recently built at Michigan State University. The mass spectrometer is described in detail elsewhere (*19*). Briefly, the radiation from a continuum vacuum ultraviolet photon source is dispersed in a monochromator. Photons with the desired energy are then passed through an ion source region containing sample vapor at a low pressure. The vast majority of the photons are transmitted by the sample, but a few are absorbed. Depending on the photon energy, such absorptions may lead to ionization and fragmentation of the sample molecules. Any ions which are formed in the source region are extracted and mass analyzed with a quadrupole mass filter. The flux of ions, I, transmitted by the quadrupole is measured by ion counting with a continuous channel electron multiplier and with pulse counting electronics that have been described previously (*20*). The transmitted photon flux, P, (which is usually at least five orders of magnitude greater than I) is also measured.

The applications of photoionization mass spectrometry have been described elsewhere (*21–23*), but in a typical experiment, the photoionization efficiency, PIE, defined as

$$\mathrm{PIE} = I/P \tag{21}$$

is measured as a function of wavelength. Plots of PIE vs. wavelength ("PIE curves") can then be interpreted to provide much useful information about the thermodynamics and energetics of positive ions.

It is desirable to acquire PIE curves at constant SNR. From Equation 21, it can be seen that

$$\left(\frac{\sigma_{\mathrm{PIE}}}{\mathrm{PIE}}\right)^2 = \left(\frac{\sigma_I}{I}\right)^2 + \left(\frac{\sigma_P}{P}\right)^2 \tag{22}$$

The relative precision in the measurement of P is expected to be much better than that in the measurement of I. Thus, as a first approximation, it is expected that the SNR of the PIE will equal the SNR of the ion flux measurement. The data in Figure 10 and the supplementary material predict that under typical operating conditions (SBR_{min} = 10–100; range < 1000; $N \gtrsim 1000$; low-peaked distribution), the CNS measurement strategy would produce a relative efficiency of about 0.9; this is the strategy currently used. The data presented here were acquired before the work reported in the section on Duty Factor had been performed, and the duty factor was adjusted only on an experiment-by-experiment basis. We estimate that the duty factors typically used further lower the overall relative efficiency to about 0.8.

In our particular implementation of the CNS strategy, the user specifies the preliminary integration time, the desired signal count total, and a maximum integration time. At each point, the data are integrated for the prescribed preliminary time interval, and the signal count rate is calculated in two ways—assuming that the measured number of signal counts may be in error by plus or minus the square root of the number

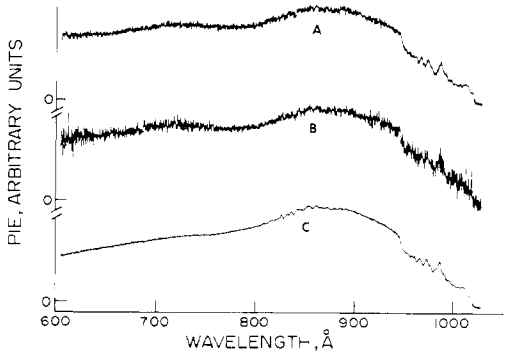

Figure 11. Comparison of data acquired with the constant time and the constant signal count strategies

of counts. If there is no question that the desired number of counts can be reached in less than the maximum integration time, any necessary further integration is performed. If one calculated count rate predicts the desired count total can be reached in the maximum time allotted but the other predicts it cannot, then the data are integrated for an additional time interval before a decision is made. If both calculations show that the desired count total cannot be reached in the allowed time, the point is simply acquired "as is", and the preliminary integration of the next point is begun.

The advantages of varying the integration time on a point-by-point basis are illustrated by the data shown in Figure 11. All three PIE curves in that figure are for the process

$$CH_3CN + h\nu \rightarrow CH_3CN^+ + e^- \qquad (23)$$

Curves A and B both contain 1680 points between 600 and 1020 Å. During these experiments the background count rate was a fairly constant 0.1 s^{-1} and the signal count rate ranged from 8500 s^{-1} at the peak of the photon continuum (800 Å) to 30 s^{-1} (at 1020 Å). Curve A was acquired using the CNS algorithm described above; actual integration times ranged from 0.3 to 150 s/point and the total experiment duration was 3.4 h. Curve B was acquired using a constant 0.3 s/point integration time. By way of reference, the much higher quality data of curve C were acquired in a 12-h CNS experiment under different conditions.

The quality of the data appears to be better in the 950–1000 Å region of curve A only because the signal and noise are both less than in other regions of the spectrum; in fact, the SNR is constant over the entire wavelength region to within a few percent. Curve A clearly reveals the two distinct steps at the ionization threshold (near 1015 Å) as well as much of the additional structure visible in curve C. This structure is interpreted elsewhere (*24*). As expected, the SNR of curves A and B are equal at the peak of the photon continuum (800 Å). The variation in SNR during the CT experiment is dramatic, however—most of the fine structure is completely obscured and even the threshold is difficult to ascertain in curve B.

It is interesting to note that nearly 73 h would have been required in order to acquire data of at least the same quality as those in curve A with the CT strategy. This is more than 20 times the duration of the CNS experiment. From data similar to those in Figure 10, the relative efficiency of a CT experiment under these circumstances is predicted to be about 0.03–0.04. Comparing these numbers with the *RE* of the CNS experiment (estimated earlier to be ~0.8), one would predict that the duration of the two experiments should differ by a factor of 20 to 26. This is entirely compatible with the experimental results.

CONCLUSIONS

All of the conclusions reached in this paper are valid for any case where the final count rate is calculated as the difference (or sum) of two separately measured count rates, such as signal and background count rates. Other cases can be treated similarly, starting with the derivation of appropriate SNR or variance expressions in a manner analogous to that in the supplementary material; they are likely to be more complicated because of an increased number of variables.

We have found that under some sets of conditions a fixed duty factor is likely to result in good (near unity) overall efficiency, and we have provided guidelines to aid the selection of the best duty factor in such cases. There are many more cases when the use of a fixed duty factor dose not result in good efficiency (though the use of $F = 0.5$ will guarantee that the efficiency will be at least 0.5). In these cases, the ability to adjust the duty factor rapidly and automatically on a point-by-point basis can guarantee an efficiency of nearly 1, and hence an improvement of as much as a factor of 2 compared with the fixed duty factor approach.

If the quality of the data is measured by its absolute precision (variance), the use of a constant integration time will often result in an efficiency exceeding 0.5, but will only rarely give an efficiency as high as 0.9. If data quality is measured by the relative precision (SNR) of the data, then the constant time approach will usually result in very poor efficiency. Two other strategies (integration of each point until some constant number of total counts or signal counts have been accumulated) will give radical improvement in efficiency (tenfold or more) compared to the constant time strategy in at least a fair number of cases. Whether relative or absolute precision is used to judge the data quality, however, best results are obtained for a wide variety of conditions when the integration time is varied on a point-by-point basis, which guarantees near-unity efficiency.

All of the discussed strategies can be readily implemented in computer-based measurement systems where the computer can also be used to recognize and avoid points of little value to the investigator and to compensate for errors in strategy based on errors in the initial estimate of SBR. The value of this approach is dramatically demonstrated in the PIMS application described.

ACKNOWLEDGMENT

We thank S. R. Crouch for many helpful conversations, comments, and suggestions during the course of this work.

Supplementary Material Available: An information data packet and nine figures will appear following these pages in the microfilm edition of this volume of the journal (13 pages). Photocopies of the supplementary material from this paper on microfiche (105 × 148 mm, 24× reduction, negatives) may be obtained from Business Operations, Books and Journals Division, American Chemical Society, Washington, D.C., 20036. Full bibliographic citation (journal, title of article, author) and prepayment, check or money order for $6.50 for photocopy ($8.00 foreign) or $3.00 for microfiche ($4.00 foreign) are required.

LITERATURE CITED

(1) Morton, G. A. *Appl. Opt.* **1968**, *7*, 1.
(2) Darland, E. J.; Leroi, G. E.; Enke, C. G. *Anal. Chem.* **1979**, *51*, 240.
(3) Niemczyk, T. M.; Ettinger, D. G.; Barnhart, S. G. *Anal. Chem.* **1979**, *51*, 2001.
(4) Hayes, J. M.; Schoeller, D. A. *Anal. Chem.* **1977**, *49*, 306.
(5) Ingle, J. D., Jr.; Crouch, S. R. *Anal. Chem.* **1972**, *44*, 777.
(6) Ingle, J. D., Jr.; Crouch, S. R. *Anal. Chem.* **1972**, *44*, 785.
(7) Davies, W. E. R. *Rev. Sci. Instrum.* **1972**, *43*, 556.
(8) Ash, K. C.; Piepmeier, E. H. *Anal. Chem.* **1971**, *43*, 26.
(9) Robben, F. *Appl. Opt.* **1971**, *10*, 776.
(10) Lakes, R. S.; Poultney, S. K. *Appl. Opt.* **1971**, *10*, 797.
(11) Jonas, M.; Alon, Y. *Appl. Opt.* **1971**, *10*, 2436.
(12) Van Gorkom, M.; Glick, R. E. *Int. J. Mass Spectrom. Ion Phys.* **1970**, *4*, 203.
(13) Franklin, M. L.; Horlick, G.; Malmstadt, H. V. *Anal. Chem.* **1969**, *41*, 2.
(14) Heroux, L. *Appl. Opt.* **1968**, *7*, 2351.
(15) Rodman, J. P.; Smith, J. H. *Appl. Opt.* **1963**, *2*, 181.
(16) Smit, C.; Alkemade, C. Th. J. *Appl. Sci. Res.* **1963–64**, *10B*, 309.
(17) Niemczyk, T. M.; Ettinger, D. G. *Appl. Spectrosc.* **1978**, *32*, 450.
(18) Wenke, D. C. Ph.D. Thesis, Michigan State University, 1972.

(19) Darland, E. J.; Rider, D. M.; Tully, F. P.; Enke, C. G.; Leroi, G. E. *Int. J. Mass Spectrom. Ion Phys.*, in press.
(20) Darland, E. J.; Hornshuh, J. E.; Enke, C. G.; Leroi, G. E. *Anal. Chem.* **1979**, *51*, 245.
(21) Reid, N. W. *Int. J. Mass Spectrom. Ion Phys.*, **1971**, *6*, 1.
(22) Rosenstock, H. M. *Int. J. Mass. Spectrom. Ion Phys.*, **1976**, *20*, 139.
(23) Chupka, W. A. In "Ion-Molecule Reactions", Vol. 1; Franklin, J. L., Ed.; Plenum Press: New York, 1972; Chapter 3.
(24) Rider, D. M. Ph.D. Thesis, Michigan State University, 1980.

RECEIVED for review October 12, 1979. Accepted December 14, 1979. The authors are pleased to acknowledge the support of the National Science Foundation (E.J.D. and G.E.L.) and the Office of Naval Research (E.J.D., G.E.L., and C.G.E.).

Factor Analysis of Trends in Selectivity in Gas–Solid Chromatography on Cation-Exchange Resins

Roland F. Hirsch* and Robert J. Gaydosh[1]

Department of Chemistry, Seton Hall University, South Orange, New Jersey 07079

Jacques R. Chrétien*

Institut de Topologie et de Dynamique des Systemes, associé au CNRS, Université Paris VII, 1, Rue Guy de la Brosse, 75005 Paris, France

Correspondence Factor Analysis (CFA) permits the identification of the major influences on behavior of saturated and unsaturated hydrocarbons in gas–solid chromatography on ion-exchangers in different cationic forms. The stationary phases are prepared from sulfonated Porapak Q ethylvinylbenzene–divinylbenzene copolymer. CFA shows that ion exchangers in the Ag^+, Ni^{2+}, Zn^{2+}, and Cd^{2+} forms offer different selectivities with the unsaturated hydrocarbons, depending on their nature, degree of substitution and π-bonding abilities. Steric hindrance of this selectivity occurs with the larger cations such as Ag^+. The H^+, K^+, Na^+, and Tl^+ forms are practically nonselective relative to the unsulfonated copolymer matrix.

Petrochemistry has attracted the constant interest of chromatographers for the systematic determination of Kovats retention indices of hydrocarbons in gas–liquid chromatography (*1–4*) and also in gas–solid chromatography (*5, 6*). The generalization of such data (*7, 8*) and elucidation of their physicochemical meaning (*9, 10*) has progressed rapidly by application of data processing methods such as topological analysis (*11*) and in connection with the establishment of data banks (*12*).

The chromatographic properties of isomeric hydrocarbons could be correlated with their reactivity if similar specific or nonspecific interactions are governing the reaction and the retention. For example, adsorption of hydrocarbons on silica or on alumina is certainly a basic aspect of both gas chromatography and catalysis of cracking processes on these materials. The reactivity of olefinic compounds in heterogeneous catalysis on Raney nickel (*13*) or homogeneous catalysis on tungsten systems (*14*) is greatly dependent on steric hindrance and strains. On the other hand, retention on graphitized carbon black is due to nonspecific interactions (*6*). Data obtained on these stationary phases permit determination of steric hindrance of alkenes and alkanes.

Electrophilic reactivity of the olefinic carbon–carbon double bond is governed by polarizability and formation of charge transfer complexes (*15, 16*). Competition between polar and steric effects has been clearly shown for alkenes (*17–19*), and the resolution of the two effects is difficult. Nevertheless, by using stationary phases of different polarities in gas–liquid chromatography, it is possible to go from practically nonspecific interactions to highly specific interactions. Unfortunately, with increasing polarity of the stationary phase, lengthening of alkyl chains or branching diminishes the specific interaction between the double bond of the alkene and the stationary phase (*11*). This trend toward a leveling effect is due to the contribution of the activity coefficient which counteracts the potential influence of vapor pressure of the solute. It can be quantified at the level of the skeleton carbon atoms by topological analysis (*10, 11*).

To avoid this limitation and to complete work currently underway in gas–solid chromatography on graphitized carbon black and in gas–liquid chromatography, it was necessary to develop fully the possibilities of specific interactions. Interest in this is evident from the analytical point of view, for example, in the analysis of higher hydrocarbons (*17*), but also for the physicochemical development of these specific interactions.

Gas–solid chromatography on macroreticular cation-exchange resins in different metal-ion forms appears to be an attractive means of reaching the goal. It offers good selectivity among hydrocarbons, depending on the ion used (*20*). The lightly sulfonated porous polymer resin has high column operating efficiency (*21*). The degree of sulfonation can be varied, and with it the extent of specificity for compounds forming complexes with the metal counterion on the packing.

In the present work we have undertaken a systematic study of the influence of the nature of the cation and of the nature and structure of the various hydrocarbons on retention of these solutes. A wide variety of hydrocarbons (alkanes, cyclanes, *n*-alkenes, branched alkenes, cyclenes, and aromatics) and a range of cations have been chosen to allow a precise delineation of the main trends in the specific interactions involved. These general trends will be determined by taking advantage of the powerful data processing technique factor analysis.

Factor analysis can be used in several different fields of chemistry (*22*); it has been used frequently in chromatography, chiefly for classification of selectivity of stationary phases (*23, 24*), but also to study and predict solute behavior (*25, 26*). When using the variant called "correspondence analysis" (*27*),

[1] Present address: American Cyanamid Corporation, Bound Brook, N.J. 08805.

0003-2700/80/0352-0723$01.00/0