ELSEVIER

Journal of Chromatography A, 831 (1999) 37–49

JOURNAL OF CHROMATOGRAPHY A

Prediction of numbers of singlets, doublets, and triplets in poorly resolved separations by statistical-overlap theory

Joe M. Davis

Department of Chemistry, Southern Illinois University at Carbondale, Carbondale, IL 62901-4409, USA

Abstract

A recently improved theory for the distribution of resolution in complex separations described by statistical-overlap theory is used to predict the average numbers of singlet, doublet, and triplet peaks in such separations, even when the peak capacity is much less than the number of components requiring separation. The theory is fundamental and modifies previously published equations for these numbers by causing the average minimum resolution that defines saturation to depend on saturation itself. Interestingly, theory shows that the resolutions describing the separation and overlap of single-component peaks in singlets, doublets, and triplets actually differ. The theory predicts correctly the numbers of singlets, doublets, and triplets in computer simulations of separations, even at high saturation. Example calculations are provided to show the ease with which the theory can be used.

Keywords: Statistical-overlap theory; Peak overlap; Singlet peaks; Doublet peaks; Triplet peaks

1. Introduction

Several theories have described peak overlap by statistical means in one-dimensional separations of complex mixtures [1–6]. In these theories, the distribution of intervals between pure chemical signals, or single-component peaks (SCPs), is described by a probability distribution. The models show various degrees of success in describing overlap in experimental separations [7–14]. However, a major shortcoming of all theories is their failure to describe overlap when the peak capacity of the separation is much smaller than the number of SCPs requiring separation, i.e., when the saturation of the separation is high. Fourier models of overlap [3,12,13,15] are less sensitive to this shortcoming than are point-process models [7,11], but all models exhibit the problem in some form. Because most complex separations are saturated, the inability to describe overlap in this important case limits the application of theory.

In recent publications by the author, theory was derived [16] and generalized [17] to obtain a statistical description of peak overlap in separations having both low and high saturation. In this theory, the separation of peaks was identified with the separation of multiplets, where a multiplet is defined as a concentration profile having only one maximum but containing any number of SCPs. The average resolution separating two multiplets was shown to depend on the number of SCPs in the multiplets, the saturation, and the type of statistics governing the distribution of intervals between SCPs. The most important conclusions drawn from theory were that the average minimum resolution required for separation depends on saturation and that earlier failures of point-process overlap theory at high saturation resulted from failure to consider this dependence.

0021-9673/99/$ – see front matter
PII: S0021-9673(98)00460-9

The theory correctly predicted the number of maxima in simulations of separations, even when this number was only 12% of the number of SCPs, and also was successfully applied to saturated experimental gas chromatograms [17].

Refs. [16,17] are not the only papers extending point-process statistical theory to high saturation. In a recent paper, Dondi et al. proposed a theory for the peak amplitude distribution and related it to various statistical attributes of the separation, e.g., the number of SCPs [18]. This ''pulse-point'' statistical theory shows much promise for interpreting saturated separations. However, it cannot be used to predict the numbers of peaks, singlets, doublets, etc., in separations. In contrast, these numbers are of interest since they are simple figures of merit that tell one how ''good'' a separation is.

In this paper, extensions of the theory in Refs. [16,17] are used to predict the numbers of singlets, doublets, and triplets in separations. These numbers agree well with those found in computer simulations, as determined by mimicking a digital integrator. The motive for this work is to make statistical-overlap theory increasingly practical to encourage its use.

Due credit should be given to J. Calvin Giddings for the initial developments of statistical-overlap theory, particularly as some of the papers in this issue are a small tribute to his many contributions to separation science. It was Giddings' insight that led to the formulation of the overlap problem in the early 1980's; as a graduate student, I then was involved with mere details. While I take pride in my extensions of the work, its true intellectual father properly should be recognized.

2. Theory

2.1. Distribution of SCPs

In the statistical-overlap theory for one-dimensional separations based on point-processes, each SCP is represented by a point at its center. The intervals between adjacent SCPs (i.e., adjacent points) are described by a probability density function (pdf), $h(z)$, where z is a coordinate that parallels the separation axis and whose origin is the center of any arbitrary SCP. The number, $h(z)dz$, is the probability that the subsequent SCP lies between z and $z+dz$, and the probability p_1 that the two SCPs are separated is [1,5]

$$p_1 = p_1(\alpha) = \int_{x_0}^{\infty} h(z)\,dz \tag{1}$$

where x_0 is the average minimum interval between the SCPs required for separation. This interval is related to the average minimum resolution R_S^* required for separation by [1]

$$x_0 = 4\sigma_{ave}R_S^* \tag{2}$$

where σ_{ave} is the average standard deviation of the SCPs. By minimum resolution, one means that this resolution, or a larger one, results in separation. The reason that the minimum resolution is an average is discussed below.

Saturation α is defined as [1]

$$\alpha = \bar{m}x_0/X = 4\bar{m}\sigma_{ave}R_S^*/X \tag{3}$$

where $\bar{m}$ is the average number of SCPs in a separation of extent X. Eq. (1) shows that probability p_1 depends on α.

Various pdfs, $h(z)$, can be used in statistical-overlap theory [3,5,12,19]. If $h(z)$ does not vary over the separation, it is called homogeneous. In this paper, I will use only the homogeneous $h(z)$ of Poisson statistics, which has been justified as the most appropriate statistics for complex mixtures [3,20]. For these statistics, p_1 equals $e^{-\alpha}$. However, the theory will be presented in its most general form, i.e., in terms of $h(z)$.

Eqs. (1)–(3) are the basis of what I call ''simple overlap theory'', as developed by Giddings and the author. In simple overlap theory, statistical attributes are calculated relative to a resolution R_S^* that is independent of α. The theory developed by the author for the distribution of resolution shows that simple overlap theory is valid only at small α.

It is clear from the highly structured distribution of peaks in some comprehensive two-dimensional separations [21] that the postulate of a statistical distribution of SCPs is not always valid. It is most appropriate when the sample dimensionality greatly exceeds the dimensionality of the separation [22], and such conditions are postulated to apply to this

study of one-dimensional separations of complex mixtures.

2.2. Distribution of resolution

Fig. 1a shows two adjacent multiplets, which contain i and k overlapping SCPs having equal intensities and standard deviations σ (here, $i=3$ and $k=4$). The SCP, A, is the right-most SCP in the multiplet on the left-hand-side, and the SCP, B, is the left-most SCP in the multiplet on the right-hand-side. If SCPs A and B in these multiplets are separated, then the two multiplets themselves are separated and two maxima are observed. Thus, separation of the multiplets as two maxima depends on the resolution of A and B. On average, the two multiplets will be separated as long as the minimum resolution R_S^{ik*} of A and B is [16,17]

$$R_S^{ik*} = 0.71(1+(i+k-2)\kappa/2)^{-1} \times \left\{ \frac{\sqrt{1+(4\kappa R_S^{ik*})^2(i^2-1)/12}+\sqrt{1+(4\kappa R_S^{ik*})^2(k^2-1)/12}}{2} \right\} \quad (4)$$

where κ is a number scaling the average interval κx_0 between adjacent SCPs that are known to overlap. The interval κx_0 is shown in Fig. 1a and κ is defined below. One observes that R_S^{ik*} appears on both sides of Eq. (4), which is not explicit.

The resolution R_S^{ik*} is the average value of the pdf, $g_{ik}(R_S)$, for the minimum resolution that separates adjacent SCPs A and B in two multiplets containing i and k overlapping SCPs. Such pdfs are explained in detail elsewhere [17]. The minimum resolution of two multiplets has a distribution of values, because both SCPs and multiplets have variable intensities (or amplitudes) resulting from the

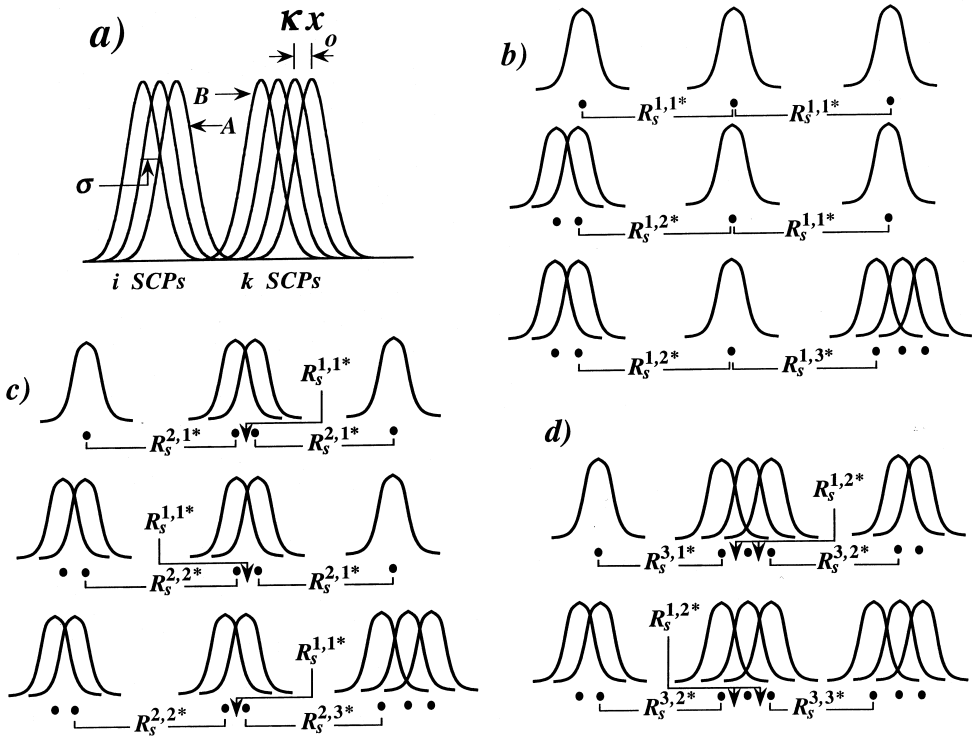

Fig. 1. (a) Two adjacent multiplets containing i and k overlapping SCPs having equal intensities and standard deviations σ (here, $i=3$ and $k=4$). SCPs A and B must be resolved to observe two maxima. Span κx_0 is the average interval between overlapping SCPs. (b–d) Various ways that (b) singlets, (c) doublets, and (d) triplets can form. Various resolutions R_S^{ik*} relevant to separation are shown.

wide range of component concentrations possible in complex mixtures. It is well known that the minimum resolution that separates two SCPs depends on the ratio of the SCP intensities (e.g., for Gaussian SCPs of equal intensity and standard deviation, the minimum resolution is 0.5), and a similar behavior is found with multiplets. Because a complex separation has many pairs of adjacent multiplets containing i and k SCPs, all of which may have different intensity ratios, the appropriate minimum resolution for describing overlap by statistical theory is an average value. The numerical value of R_S^{ik*} depends on i and k; interestingly, R_S^{ik*} can be less than 0.5 [16,17].

The number, 0.71, in Eq. (4) is the average minimum resolution, $R_S^{1,1*}$, required to separate two (and only two) SCPs [23–25], when SCP intensities follow the exponential distribution commonly observed for complex mixtures (that is, most relative intensities are small and only a few are large) [9,26]. This number is different for other intensity distributions, but otherwise Eq. (4) is derived from theory.

Eq. (4) determines only the average minimum resolution of SCPs in adjacent multiplets containing i and k SCPs. The *global* average minimum resolution, R_S^*, defining saturation α in Eq. (3) is a weighted combination of all resolutions R_S^{ik*} relevant to a particular problem. For example, if one is interested in predicting the average number p of peaks in the separation, as in Refs. [16,17], then all values of i and k are relevant because peaks can contain any number of SCPs. In this case, the global resolution R_S^* is

$$R_S^* = \sum_{i=1}^{\infty} \sum_{k=i}^{\infty} p_{ik} R_S^{ik*} \tag{5}$$

where p_{ik} is the probability that adjacent multiplets containing i and k SCPs are resolved [16,17]

$$p_{ik} = p_{ik}(\alpha) = \frac{P_{ik}}{\sum_{j=1}^{\infty} \sum_{l=j}^{\infty} P_{jl}} = \frac{(1-p_1)^{i+k-2}}{\sum_{j=1}^{\infty} \sum_{l=j}^{\infty} (1-p_1)^{j+l-2}}$$

$$= \frac{p_1(1-p_1)^{i+k-2}}{\sum_{j=0}^{\infty} (1-p_1)^{2j}}, \qquad i \geq 1; k \geq i \tag{6}$$

In Eq. (6), P_{ik} is the average number of resolved pairs of adjacent multiplets in X that contain i and k SCPs and p_1, Eq. (1), is the probability that adjacent SCPs are separated.

A general expression for scalar κ can be written for any homogeneous pdf, $h(z)$, describing the distribution of intervals between SCPs in the separation [17]

$$\kappa x_0 = \kappa(\alpha) x_0 = \int_0^{x_0} z h(z)\, dz \Big/ \int_0^{x_0} h(z)\, dz \tag{7a}$$

For the Poisson $h(z)$ [16,17]

$$\kappa = \alpha^{-1} - e^{-\alpha}/[1 - e^{-\alpha}] \tag{7b}$$

2.3. Dependence of R_S^* on α

Eqs. (1)–(6), (7a), (7b) are not independent; their simultaneous solution requires that R_S^* vary with α [16,17]. In essence, the pdf, $h(z)$, for the distribution of intervals between SCPs and the pdf, $g_{ik}(R_S)$, for the distribution of multiplet resolution are coupled by saturation α. Thus, Eq. (3) is more correctly written as

$$\alpha = \bar{m} x_0(\alpha)/X = 4\bar{m}\bar{\sigma}_{ave} R_S^*(\alpha)/X \tag{8}$$

The variation of R_S^* with α can be predicted from the above equations. It is recognition of this simple fact that enabled the author to predict correctly the average number p of peak maxima, both in computer simulations and experimental gas chromatograms [16,17], even when p was much less than $\bar{m}$.

2.4. Theory for singlets, doublets, and triplets

The above theory now is adapted to predict the average numbers s, d, and t of singlet peaks, doublet peaks, and triplet peaks, respectively, in separations subject to statistical description. Fig. 1b–d show several possible ways that singlets, doublets, and triplets can be formed. For all three peak types, the two intervals bracketing them are large enough to separate them from the SCPs that precede and follow. For the doublet, however, the interval between its two constituent SCPs is too small for separation. The triplet has three such constituent

SCPs and two such intervals. From these considerations alone, one determines from $h(z)$ that [1,5]

$$s = \bar{m}p_1^2 = \bar{m}e^{-2\alpha} \tag{9a}$$

$$d = \bar{m}p_1^2(1-p_1) = \bar{m}e^{-2\alpha}(1-e^{-\alpha}) \tag{9b}$$

$$t = \bar{m}p_1^2(1-p_1)^2 = \bar{m}e^{-2\alpha}(1-e^{-\alpha})^2 \tag{9c}$$

where the central expressions apply to any homogeneous $h(z)$ and the final expressions specifically apply to the Poisson $h(z)$.

Eqs. (9a)–(9c) are not new. However, one also must incorporate into Eqs. (9a)–(9c) the resolution distribution that determines global resolution R_S^*, Eq. (5), on which α, Eq. (8), depends. This work has not been done until now. Interestingly, theory shows that the R_S^* defining α in the p_1 (or $e^{-\alpha}$) factors of Eqs. (9a)–(9c) must differ from the R_S^* defining α in the $(1-p_1)$ (or $[1-e^{-\alpha}]$) factors. In addition, these R_S^*'s differ for singlets, doublets, and triplets.

Because p_1 is defined relative to different R_S^*'s, depending on its use, I will try to be clear in the following discussion. Recall that the definition of a factor is one of two or more numbers or algebraic expressions that, when multiplied together, produce a product. I will refer to the R_S^* determining both of the factors, p_1, of Eqs. (9a)–(9c) (i.e., the R_S^* determining $p_1 \times p_1 = p_1^2$) using the words, "R_S^* for the p_1 factor". Similarly, I will refer to the R_S^* determining p_1 in the factor, $(1-p_1)$, of Eqs. (9b) and (9c) using the words, "R_S^* for the $(1-p_1)$ factor".

2.5. Resolution distribution for p_1 factors

The derivation of an R_S^* expression for the p_1 factors of Eqs. (9a)–(9c) is identical for singlets, doublets, triplets, and all other multiplets. This factor is associated with the two intervals that span the peak of interest and the preceding and following SCPs. It is immaterial whether either the preceding or following SCP is itself a singlet or part of a doublet, triplet, etc., as shown in Fig. 1b–d. In Fig. 1b, for example, a singlet is formed in three different ways: by a singlet preceding it and a singlet following it, by a doublet preceding it and a singlet following it, and by a doublet preceding it and a triplet following it. In general, if the peak of interest contains n SCPs, only the resolution subset $R_S^{nk}*$, where k is a positive integer, is relevant to describing R_S^* for the p_1 factors. Specifically, only the resolutions $R_S^{1,k}*$ are relevant to describing intervals separating singlets from adjacent multiplets. Similarly, only the resolutions $R_S^{2,k}*$ are relevant to describing intervals separating doublets from adjacent multiplets. Various $R_S^{nk}*$'s relevant to the separation of multiplets are identified in Fig. 1b–d.

One ultimately is interested in the global resolution R_S^*, which depends on $R_S^{ik}*$ and p_{ik} (see Eq. (5)). Eq. (4) determines $R_S^{nk}*$, when i is set to n. By restricting i to the single value, n, one simplifies the double sum for p_{ik}, Eq. (6), to a single one

$$p_{nk} = p_{nk}(\alpha) = \frac{P_{nk}}{\sum_{l=1}^{\infty} P_{nl}} = \frac{(1-p_1)^{n+k-2}}{\sum_{l=1}^{\infty}(1-p_1)^{n+l-2}}$$
$$= p_1(1-p_1)^{k-1} \tag{10}$$

which is the geometric distribution and is independent of n. Thus, p_{nk} is the *same* for singlets, doublets, triplets, and higher-order multiplets.

Therefore, the global resolution R_S^* determining *all* the p_1 factors of Eqs. (9a)–(9c) is

$$R_S^* \equiv R_{S,p_1(n)}^*(\alpha) = p_1 \sum_{k=1}^{\infty}(1-p_1)^{k-1} R_S^{nk}* \tag{11}$$

where the variable $R_{S,p_1(n)}^*$ is introduced to identify the global resolution R_S^* for the p_1 factors. One only needs to specify n to apply Eq. (11) to a particular multiplet type. Because $R_S^{nk}*$ varies with n, so does Eq. (11), i.e., $R_{S,p_1(n)}^*$ differs at any α for singlets, doublets, and triplets.

Because R_S^* for the $(1-p_1)$ factors of doublets and triplets differs (see below), it is best at this stage to consider each multiplet type separately.

2.6. Singlets s

The expression for the average number s of singlets is given by Eqs. (8) and (9a), with R_S^* equal to $R_{S,p_1(1)}^*$, i.e., to Eq. (11) with $n=1$.

2.7. Doublets d

A doublet is formed when three contiguous intervals are sufficient, insufficient, and sufficient for

separation. The two intervals sufficient for separation are associated with the p_1 factors of Eq. (9b); these factors in turn are defined relative to the resolution $R_S^* = R_{S,p_1(2)}^*$ predicted by Eq. (11) with $n=2$.

However, the interval insufficient for separation and associated with the $(1-p_1)$ factor of Eq. (9b) is defined relative to a different R_S^*. By definition of a doublet, two and only two adjacent SCPs overlap. Therefore, the average minimum resolution R_S^* required for separation of the doublet is $R_S^{1,1}*$, which equals 0.71 or so and is independent of α (see Eq. (4) and following discussion) [23–25].

Thus, the p_1 factors of Eq. (9b) for the average number d of doublets contain α's, in which R_S^* equals Eq. (11) with $n=2$, and the $(1-p_1)$ factor of Eq. (9b) contains an α, in which $R_S^*=0.71$. This conclusion is nonintuitive, but one's intuition in this field has been biased by thinking of R_S^* as constant. Evidence supporting this theory is presented below.

2.8. Triplets t

A triplet is formed when four contiguous intervals are sufficient, insufficient, insufficient, and sufficient for separation. The two intervals sufficient for separation are associated with the p_1 factors of Eq. (9c); these factors in turn are defined by the resolution $R_S^* = R_{S,p_1(3)}^*$ predicted by Eq. (11) with $n=3$.

The intervals insufficient for separation result in overlap that can happen in two ways. First, three contiguous singlets can overlap to form a triplet. Secondly, an adjacent doublet and singlet can overlap to form a triplet. The first possibility is much less likely than the second, as is easily seen. Envision a separation of nearly infinite efficiency, whose efficiency then is degraded in subsequent separations. If three singlets were to overlap simultaneously to form a triplet, then the intervals between the first and second SCP, and between the second and third SCP, would have to correlate with the SCP intensities in exactly the right manner. It is much more likely that two of the three SCPs overlap at some intermediate efficiency and that the resulting doublet then overlaps with the adjacent singlet as efficiency is degraded further. The same argument is valid, if instead one starts with a separation of low efficiency and then improves it; a triplet is more likely to resolve first into a singlet and doublet than into three singlets.

In light of this, then, the R_S^* defining α in both $(1-p_1)$ factors of Eq. (11) is $R_S^{1,2}*$, the average minimum resolution separating a singlet and a doublet. Unlike $R_S^{1,1}*$, however, $R_S^{1,2}*$ varies with α (see Eq. (4)).

A possible objection to this assertion is discussed. Consider three contiguous SCPs *A*, *B*, and *C*; *A* is separated from *B* but *B* overlaps with *C*. Thus, *A* is the singlet and *BC* the doublet that can overlap to form a triplet, as discussed above. One might argue: although the $(1-p_1)$ factor associated with the interval between *A* and *B* indeed should be defined by $R_S^{1,2}*$, the $(1-p_1)$ factor associated with the interval between *B* and *C* should be defined by $R_S^{1,1}*$, since *B* and *C* form a doublet. This argument is correct, if subsequent overlap does not occur, but it is incorrect if overlap occurs and triplet *ABC* actually forms. Consider that this triplet is formed. Now consider that *B* and *C* are somehow resolved, such that singlet *C* is separated from the doublet, *AB*. Since *AB* is a doublet, the appropriate R_S^* describing the resolution of *B* and *C* must be $R_S^{1,2}*$. Thus, it also is the appropriate R_S^* for describing the overlap of *B* and *C* in a triplet.

Thus, the p_1 factors of Eq. (9c) for the average number t of triplets contain α's, in which R_S^* equals Eq. (11) with $n=3$, and the $(1-p_1)$ factors of Eq. (9c) contain α's, in which R_S^* equals $R_S^{1,2}*$, as calculated from Eq. (4).

3. Procedures

Eqs. (1), (4), (7b), (11) were solved numerically for $n=1$, 2 and 3 by specifying an arbitrary series of $R_{S,p_1(n)}^*$ values and solving iteratively for α using bisection. In each iteration of α, all parameters on which $R_{S,p_1(n)}^*$ depended were calculated. Specifically, using the α of that iteration, probability $p_1=e^{-\alpha}$ was calculated, κ was calculated from Eq. (7b), and $R_S^{nk}*$ was calculated from Eq. (4) using bisection. The convergence criterion for α was a relative change in α of less than 10^{-4} between successive iterations. For each *n*, a large number of coordinates, $(\alpha, R_{S,p_1(n)}^*)$, was so computed and then fit by a cubic

spline. Sufficient coordinates were chosen to define the splines with accuracy. The splines then were used to calculate values of $R^*_{S,p_1(n)}$ at equally spaced values of α.

In a similar manner, Eq. (4), with $i=1$ and $k=2$, and Eq. (7b) were solved using bisection by specifying a series of α values and solving iteratively for $R_S^{1,2}*$.

Theoretical predictions of s, d, and t were made at α's arbitrarily defined relative to $R^*_{S,p_1(n)}$ (i.e., $\alpha = 4\bar{m}\sigma_{ave}R^*_{S,p_1(n)}/X$). To do so, the $p_1=e^{-\alpha}$ factors of Eqs. (9a)–(9c) were evaluated straightforwardly, since these factors were defined relative to $R^*_{S,p_1(n)}$. The $(1-p_1)=(1-e^{-\alpha})$ factors, however, were evaluated at the product of α and the ratio, $R_S^{1,1}*/R^*_{S,p_1(2)}(\alpha)$, for doublets and $R_S^{1,2}*(\alpha)/R^*_{S,p_1(3)}(\alpha)$ for triplets. Since α was defined relative to $R^*_{S,p_1(n)}$, dividing α by $R^*_{S,p_1(n)}$ eliminated this resolution and multiplying α by either $R_S^{1,1}*$ or $R_S^{1,2}*$ introduced this resolution. By these actions, the $(1-e^{-\alpha})$ factors were calculated at the appropriate resolutions, $R_S^{1,1}*$ for doublets and $R_S^{1,2}*$ for triplets.

Computer simulations of multicomponent separations were generated to test theory. Each simulation contained $\bar{m}=250$ Gaussian SCPs distributed in a span X of unit extent (i.e., $X=1$), with intervals between successive SCPs determined by the Poisson $h(z)$. All SCPs had equal standard deviations σ (hence, the average standard deviation σ_{ave} of adjacent SCPs equalled σ) and exponentially distributed intensities (or amplitudes). For any α, the σ of SCPs in any simulation was chosen to satisfy Eq. (8), with $R_S^* \equiv R^*_{S,p_1(n)}$ equal to that predicted by the cubic spline. The numerical methods required in these simulations are described elsewhere [4,5].

Peaks in the simulations were determined by mimicking a simple digital integrator in a noiseless environment. The boundaries of isolated peaks were determined by the departure of the baseline slope from zero and the return of the baseline slope to zero. The boundaries of peaks partially overlapped with other peaks were the valleys between peaks. Both methods of determination were used when appropriate. The numbers of singlets, doublets, and triplets were determined by counting the numbers of peaks in which the numbers of SCPs between peak boundaries equalled one, two, and three, respectively. The numbers so determined from 100 simulations were averaged for several α's.

All computations were made on a 200 MHz Power Macintosh 4400 using Language Systems FORTRAN (Sterling, VA, USA).

4. Results and discussion

Fig. 2(a and b) are simulations of separations having low and high saturation, respectively. The vertical lines denote the boundaries of peaks, and the numbers are the number of SCPs between the peak boundaries. Several simple peaks containing one or two SCPs in Fig. 2a overlap to form more complicated peaks (e.g., a hextet) in Fig. 2b. The numbers of SCPs in various peaks were determined from simulations identical to these.

Fig. 2c, Fig. 3(a and c) are graphs of the average numbers s of singlets, d of doublets, and t of triplets in simulations containing $\bar{m}=250$ SCPs, respectively, vs. α, with α defined by $R^*_{S,p_1(n)}$ (i.e., $\alpha = 4\bar{m}\sigma_{ave}R^*_{S,p_1(n)}/X$). The solid curves are theory, and the circles are the average number of singlets, doublets, and triplets found in 100 simulations. The error bars represent one standard deviation. Fig. 2d, Fig. 3(b and d) are graphs of various average minimum resolutions R_S^* vs. α calculated from theory for singlets, doublets, and triplets, respectively. The circles represent the coordinates, $(\alpha, R^*_{S,p_1(n)})$, determined numerically; the curves passing through the coordinates are cubic splines. The graph of $R_S^{1,2}*$ vs. α in Fig. 3d is not a spline; this curve was calculated from Eqs. (4) and (7b). The resolutions $R_S^{1,2}*$ and $R^*_{S,p_1(3)}$ in Fig. 3d are graphed against different α's, i.e., $R_S^{1,2}*$ is graphed against $\alpha = 4\bar{m}\sigma_{ave}R_S^{1,2}*/X$ and $R^*_{S,p_1(3)}$ is graphed against $\alpha = 4\bar{m}\sigma_{ave}R^*_{S,p_1(3)}/X$. Finally, the inserts in Fig. 2c, Fig. 3(a and c) are graphs of the average numbers of singlets s, doublets d, and triplets t, respectively, vs. α, with α defined by the constant, $R_S^*=R_S^{1,1}*=0.71$, the average minimum resolution separating two (and only two) SCPs. The inserts consequently are similar to previously published graphs based on simple overlap theory, in which R_S^* is assumed to be constant [7,11].

In general, the agreement between simulation and

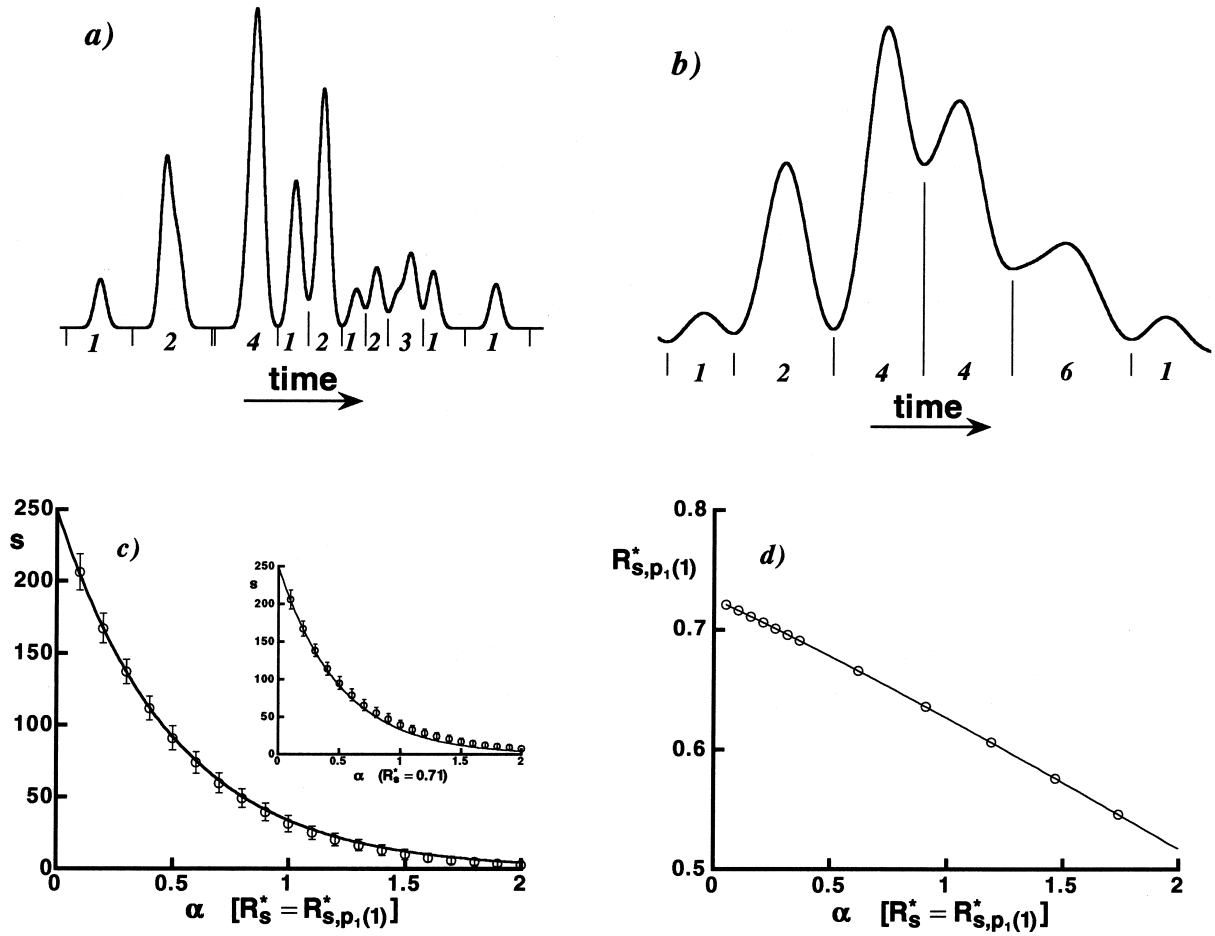

Fig. 2. Computer simulations of separations developed at (a) low and (b) high α. Vertical lines represent peak boundaries. Numbers between peak boundaries are numbers of SCPs in maxima. (c) Graph of average number s of singlets vs. α, with α defined by $R^*_{S,p_1(1)}$. Insert is graph of s vs. α, with α defined by $R^*_S = R^{1,1}_S{}^* = 0.71$. (d) Graph of $R^*_{S,p_1(1)}$ vs. α, with α defined by $R^*_{S,p_1(1)}$. Circles represent numerically determined coordinates; curve is cubic spline.

the improved theory shown in Fig. 2c, Fig. 3(a and c) is excellent, even as α approaches 2. This agreement confirms that a fairly good understanding of overlap in statistically governed separations of one dimension now exists. It is apparent from the insert in Fig. 2c that the improved theory has only a small effect on s. Indeed, the agreement between simulation and theory for s is fairly good, even when R^*_S is 0.71. The most probable reason for this behavior is that singlets found at high saturation are well resolved from other peaks and are not very sensitive to the exact magnitude of R^*_S. This situation is not true, however, for multiplets d and t, as the inserts in Fig. 3(a and c) show. In these cases, the numbers of multiplets substantially exceed theory at large α, and the improved theory is required for accurate predictions. Only at low α is overlap correctly predicted by $R^*_S = 0.71$.

The dependence on α of the various resolutions in Figs. 2 and 3 can be rationalized simply. Although κ decreases with increasing α (see Eq. (7b)), both the numerator and denominator of the expression for $R_S{}^{ik}{}^*$, Eq. (4), depend on κ. Consequently, this dependence largely cancels and $R_S{}^{ik}{}^*$ varies only slightly with α. The value of $R^*_{S,p_1(n)}$, however, depends on all $R_S{}^{nk}{}^*$'s, as shown by Eq. (11). At

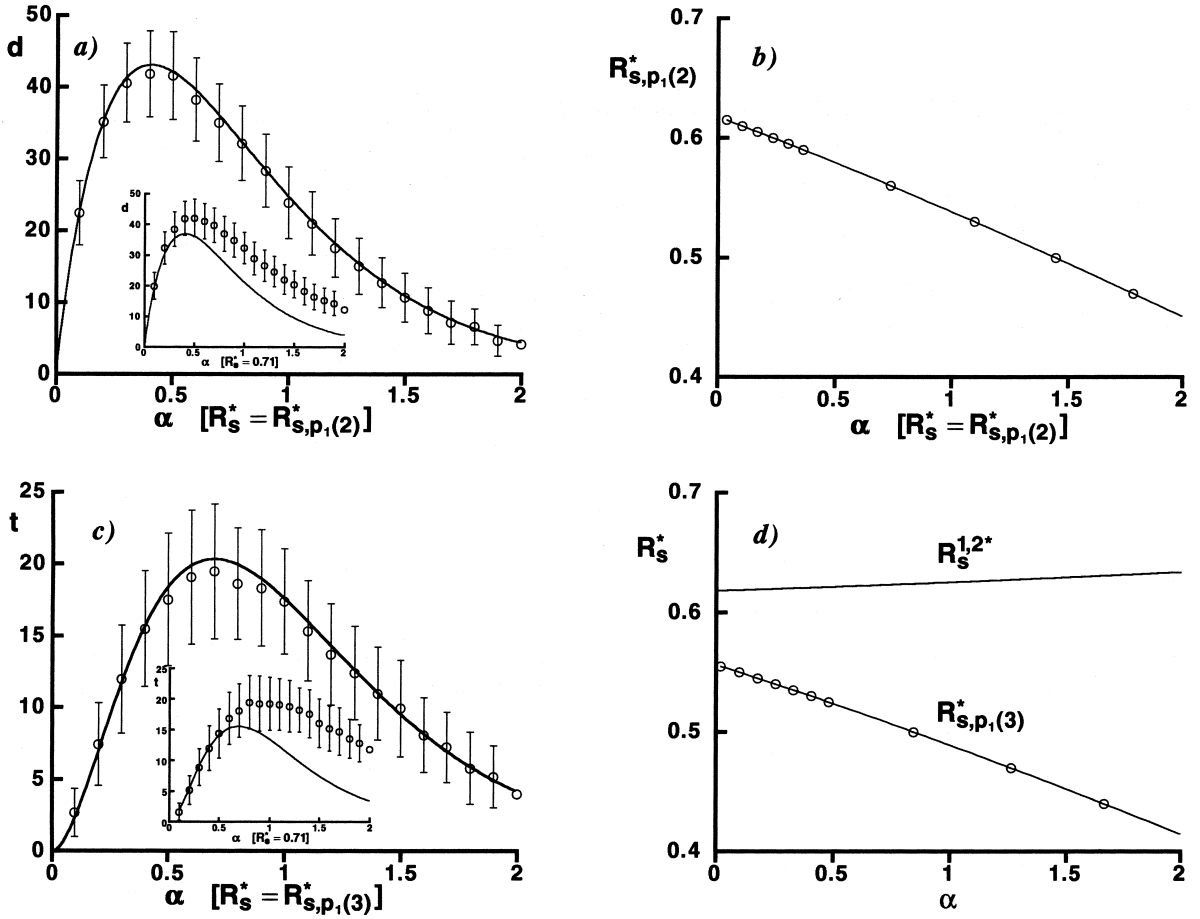

Fig. 3. (a) As in Fig. 2c, but for doublets *d*. (b) As in Fig. 2d, but for $R^*_{S,p_1(2)}$. (c) As in Fig. 2c, but for triplets *t*. (d) As in Fig. 2d, but for $R^*_{S,p_1(3)}$ and $R_S^{1,2}*$. Resolutions $R^*_{S,p_1(3)}$ and $R_S^{1,2}*$ are graphed against different α's (see Section 4).

large α's, the values of $R_S{}^{ik}*$ that are associated with large k's increasingly contribute to $R^*_{S,p_1(n)}$, since it is likely that multiplets contain large numbers of SCPs when saturation is high. Since $R_S{}^{ik}*$ decreases with increasing k at any α, $R^*_{S,p_1(n)}$ also decreases with increasing α.

Some readers may be troubled by Fig. 3(b and d), in which various resolutions R^*_S have values less than 0.5. Although 0.5 is the smallest resolution that can separate two Gaussians having equal standard deviations and intensities, the resolution of SCPs in multiplets containing more than one SCP can be less than 0.5 as is shown elsewhere [24].

It is important to realize that the dependence of the p_1 and $(1-p_1)$ factors of Eqs. (9a)–(9c) on different R^*_S's causes the expressions for *s*, *d*, and *t* to differ from Eqs. (9a)–(9c). Specifically, if α is defined relative to $R^*_{S,p_1(n)}$, these equations become

$$s = \bar{m}e^{-2\alpha} = \bar{m}\exp(-8m\bar{\sigma}_{ave}R^*_{S,p_1(1)}(\alpha)/X) \quad (12a)$$

$$d = \bar{m}e^{-2\alpha}[1-\exp[-\alpha R_S^{1,1}*/R^*_{S,p_1(2)}(\alpha)]] \quad (12b)$$

$$t = \bar{m}e^{-2\alpha}[1-\exp[-\alpha R_S^{1,2}*(\alpha)/R^*_{S,p_1(3)}(\alpha)]]^2 \quad (12c)$$

where $R_S^{1,2}*$, $R^*_{S,p_1(1)}$, $R^*_{S,p_1(2)}$, and $R^*_{S,p_1(3)}$ all depend on α. In other words, by addressing the distribution of resolution, one transforms Eqs. (9a)–(9c) into closely related but different equations. The implica-

tion is clear; a description of overlap cannot be realized by addressing the SCP distribution alone. Rather, overlap also is affected by the distribution of resolution, and the proper equations describing *s*, *d*, and *t* (and also peak number *p*) differ from their point-process counterparts.

The above ideas can be extended to predict the average numbers of other multiplets. For example, Fig. 4(a and c) are graphs of the average numbers of quartets and quintets vs. α, with α defined by $R^*_{S,p_1(n)}$, and Fig. 4(b and d) are graphs of various resolutions R^*_S vs. α calculated from theory for quartets and quintets, respectively. The symbols and curves have the same meaning as in Fig. 3. As with triplets, one considers only the overlap of multiplet pairs. One observes that quartets and quintets can be formed by two kinds of multiplet pair overlap; specifically, quartets can be formed by the overlap of either doublets and doublets or singlets and triplets, and quintets can be formed by the overlap of either doublets and triplets or singlets and quartets. For the quartet, *ABCD*, the resolutions describing overlap between the two adjacent multiplets *A* and *BCD*, *AB* and *CD*, and *A* and *BCD* are $R_S^{1,3}*$, $R_S^{2,2}*$, and $R_S^{1,3}*$, respectively; for the quintet, *ABCDE*, the resolutions describing overlap between the two adjacent multiplets *A* and *BCDE*, *AB* and *CDE*, *ABC* and *DE*, and *ABCD* and *E* are $R_S^{1,4}*$, $R_S^{2,3}*$, $R_S^{2,3}*$, and $R_S^{1,4}*$,

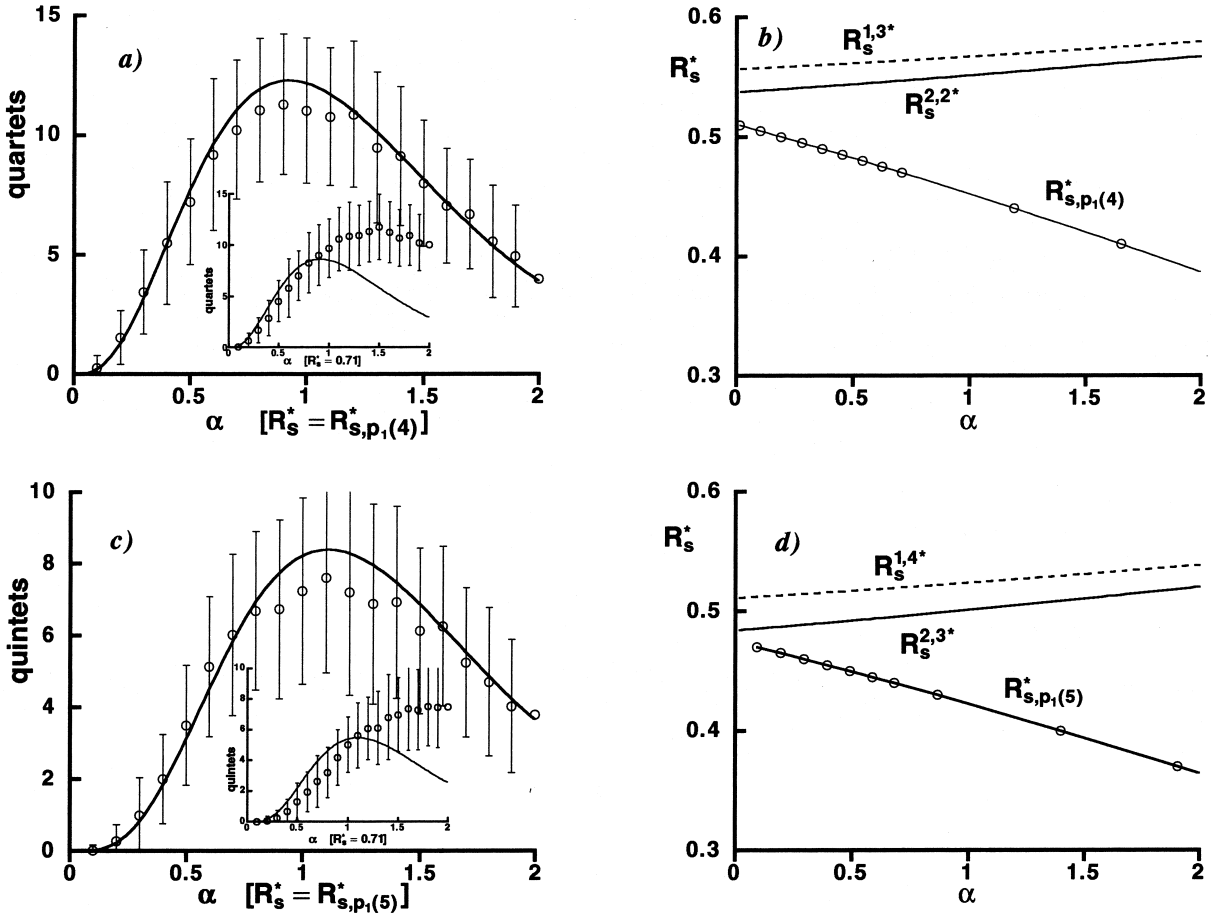

Fig. 4. (a) As in Fig. 2c, but for quartets. (b) As in Fig. 2d, but for $R^*_{S,p_1(4)}$, $R_S^{1,3}*$, and $R_S^{2,2}*$. Resolutions $R^*_{S,p_1(4)}$, $R_S^{1,3}*$, and $R_S^{2,2}*$ are graphed against different α's. (c) As in Fig. 2c, but for quintets. (d) As in Fig. 2d, but for $R^*_{S,p_1(5)}$, $R_S^{1,4}*$, and $R_S^{2,3}*$. Resolutions $R^*_{S,p_1(5)}$, $R_S^{1,4}*$, and $R_S^{2,3}*$ are graphed against different α's.

respectively. Fig. 4(b and d) are graphs of $R_S^{1,3}$*, $R_S^{2,2}$*, $R_S^{1,4}$*, and $R_S^{2,3}$* vs. α, with α defined relative to the different resolutions.

A fair agreement between simulation and theory is found in Fig. 4(a and c). It is evident that theory slightly overestimates the quartet and quintet numbers at intermediate α. Actually, if one studies Figs. 3 and 4, one sees a progressive overestimation of multiplets at intermediate α's for higher-order multiplets, and this overestimation simply is more noticeable for quartets and quintets than for doublets and triplets. Undoubtedly, a description of overlap by the fusion of only multiplet pairs is too simplistic for complex multiplets. Nevertheless, the theory appears good, if compared to the standard deviations of the multiplet numbers, and is much better than that obtained by defining α by $R_S^{1,1}$*$=0.71$, as shown by the inserts in Fig. 4(a and c). The equations based on this constant R_S^* do not agree well with simulation even as α approaches zero.

4.1. Example calculations

Because R_S^* depends on α, the theory outlined here is slightly more difficult to use than simple overlap theory, in which R_S^* is constant. The accuracy of its predictions, however, warrants the additional effort required. I conclude with a series of calculations to illustrate an application of theory.

Suppose one knew that a gas chromatographic separation spanning 12 min contained 175 SCPs distributed in accordance with Poisson statistics and having standard deviations of 2.0 s. How many of these SCPs would appear as singlets, doublets, and triplets? Furthermore, how many peaks would appear?

First, it is prudent to see if simple overlap theory can answer the questions, since one desires to use as simple a theory as possible. Previous studies have shown that simple overlap theory is adequate if α is less than 0.5 or so, when R_S^* is defined arbitrarily as 0.5 [7,27]. For $\bar{m}=175$, $\sigma_{ave}=2.0$ s, and $X=720$ s (i.e, 12 min), one calculates that $\alpha=4\bar{m}\sigma_{ave}R_S^*/X$ is $4(175)(2.0)(0.5)/720=0.97$, when $R_S^*=0.5$. It now is clear that simple overlap theory will be inadequate, so one turns to the theory developed here.

To determine s, one simply numerically solves the equation

$$\alpha = 4\bar{m}\sigma_{ave}R^*_{S,p_1(1)}/X \quad (13)$$

with $\bar{m}=175$, $\sigma_{ave}=2.0$ s, $X=720$ s, and $R^*_{S,p_1(1)}$ equal to that graphed in Fig. 2d. The solution to Eq. (13) can be approximated using spreadsheets by tabulating $(\alpha, R^*_{S,p_1(1)})$ coordinates, evaluating Eq. (13) at these coordinates, and finding the α for which both sides of Eq. (13) are identical. By doing so, one finds $\alpha=1.18$ and $R^*_{S,p_1(1)}=0.607$. From Eq. (12a), one estimates that s is $175\times\exp(-2\times1.18)$, or 16.5 singlets.

A similar approach is used to obtain d, with $R^*_{S,p_1(2)}$ in Fig. 3b replacing $R^*_{S,p_1(1)}$ in Eq. (13); by using spreadsheets, one finds $\alpha=1.04$ and $R^*_{S,p_1(2)}=0.535$. From Eq. (12b), then, one estimates that d is $175\times\exp(-2\times1.04)\times[1-\exp(-1.04\times0.71/0.535)]$, or 16.4 doublets. Here, $0.71/0.535$ represents the resolution ratio, $R_S^{1,1}$*$/R^*_{S,p_1(2)}$.

A similar approach also is used to obtain t, with $R^*_{S,p_1(3)}$ in Fig. 3d replacing $R^*_{S,p_1(1)}$ in Eq. (13); one finds $\alpha=0.96$ and $R^*_{S,p_1(3)}=0.492$. The resolution $R_S^{1,2}$* at this α has the value, 0.625, as determined from Fig. 3d. From Eq. (12c), then, one estimates that t is $175\times\exp(-2\times0.96)\times[1-\exp(-0.96\times0.625/0.492)]^2$, or 12.7 triplets. Here, $0.625/0.492$ represents the resolution ratio, $R_S^{1,2}$*$/R^*_{S,p_1(3)}$.

Finally, an identical approach is used to estimate the average number p of peaks in the separation. By introducing the coordinates, (α, R_S^*), graphed as Fig. 2b in Ref. [17] into a spreadsheet, substituting R_S^* for $R^*_{S,p_1(1)}$ in Eq. (13), and solving, one determines that $\alpha=1.06$ and $R_S^*=0.544$. From the expression for p reported in Ref. [17], $p=\bar{m}p_1=\bar{m}e^{-\alpha}$, one concludes that the separation should contain $175\times\exp(-1.06)$, or 60.6 peaks.

These predictions agree very closely with results determined by 100 computer simulations corresponding to $\bar{m}=175$, $\sigma_{ave}=2.0$ s, and $X=720$ s. From these simulations, one finds $s=14.7\pm3.5$, $d=15.4\pm3.5$, $t=12.3\pm3.2$, and $p=61.2\pm3.5$. All the theoretical estimates are correct within one standard deviation of the simulation. In contrast, if R_S^* is simply equated to $R_S^{1,1}$*$=0.71$, as in simple overlap theory, then one predicts that the separation should contain 11.1 singlets, 8.3 doublets, 6.2 triplets, and 44.0 peaks; most of these numbers clearly are wrong. Even the estimate of the average number of singlets, which is the best one, is not correct within one standard deviation.

Table 1
Determinations of *s*, *d*, *t*, and *p* by simple overlap theory, improved theory derived here, and computer simulation; $\bar{m}=175$, $\sigma_{ave}=2.0$ s, and $X=720$ s

Nature of determination	*s*	*d*	*t*	*p*
Simple overlap theory ($R_S^*=0.71$)	11.1	8.3	6.2	44.0
Improved theory	16.5	16.4	12.7	60.6
Computer simulation[a]	14.7±3.5	15.4±3.5	12.3±3.2	61.2±3.5

[a] Average and standard deviation determined from 100 simulations.

All these multiplet numbers are reported in Table 1 for easy comparison. Listings of various coordinates (α, R_S^*) for *s*, *d*, *t*, and *p* are available on request.

5. Conclusion

It is clear that a quantitative understanding of overlap in statistically limited separations of one dimension is being approached, even when overlap is severe. This understanding also is fundamental and rigorous. Only the ''0.71'' factor in Eq. (4) is an empirically determined parameter in the author's theory [16,17], and Felinger has justified this factor by additional theory [25].

The theory developed here is only slightly more awkward to use than simple overlap theory and is far more accurate. However, it is disquieting to find that the p_1 and $(1-p_1)$ factors determining *s*, *d*, *t*, *p*, etc., must be defined relative to different R_S^*'s and α's; in doing so, one loses most of the intuitive insights that motivated the development of statistical-overlap theory in the first place. Apparently, this sacrifice is the price of accuracy. From another perspective, however, this finding reaffirms something we have long understood: resolution, as defined for two and only two SCPs, is a concept useful only in simple separations; once multiplets begin to overlap, the concept of resolution is misleading. In light of this, perhaps the author's definition of multiplet resolution, $R_S{}^{ik}*$, will find some applications outside of overlap theory.

Clearly an experimental test of this theory is needed with controlled mixtures. The simplest test entails preparation of a synthetic mixture, for which the retention times of individual SCPs and observed peaks can be measured and peak multiplicities can be inferred. Such a study has been made before, but only for cases of slight overlap [11]. It also is likely that this work can be extended to SCP distributions in which the pdf $h(z)$ is not a Poisson distribution. More important, it also is likely that calculations similar to those above can be used to estimate $\bar{m}$, an important application of overlap theory.

Acknowledgements

This work was supported by the National Science Foundation (CHE-9714328).

References

[1] J.M. Davis, J.C. Giddings, Anal. Chem. 55 (1983) 418.
[2] M. Martin, D.P. Herman, G. Guiochon, Anal. Chem. 58 (1986) 2200.
[3] A. Felinger, L. Pasti, F. Dondi, Anal. Chem. 62 (1990) 1846.
[4] J.M. Davis, Anal. Chem. 66 (1994) 735.
[5] M.C. Pietrogrande, F. Dondi, A. Felinger, J.M. Davis, Chemometr. Intell. Lab. Syst. 28 (1995) 239.
[6] J.M. Davis, in: P.R. Brown, E. Grushka (Eds.), Advances in Chromatography, Marcel Dekker, New York, 1994, Ch. 3, p. 109.
[7] D.P. Herman, M.-F. Gonnord, G. Guiochon, Anal. Chem. 56 (1984) 995.
[8] J.M. Davis, J.C. Giddings, Anal. Chem. 57 (1985) 2178.
[9] F. Dondi, Y.D. Kahie, G. Lodi, M. Remelli, P. Reschiglian, C. Bighi, Anal. Chim. Acta 191 (1986) 261.
[10] J.M. Davis, J. Chromatogr. 449 (1988) 41.
[11] S.L. Delinger, J.M. Davis, Anal. Chem. 62 (1990) 436.
[12] F. Dondi, A. Betti, L. Pasti, M.C. Pietrogrande, A. Felinger, Anal. Chem. 65 (1993) 2209.
[13] M.C. Pietrogrande, L. Pasti, F. Dondi, M.H.B. Rodriguez, M.A.C. Diaz, J. High Resolut. Chromatogr. 17 (1994) 839.
[14] M. Johll, D. Bowlin, J.M. Davis, Anal. Chim. Acta 302 (1995) 133.
[15] F. Dondi, M.C. Pietrogrande, A. Felinger, Chromatographia 45 (1997) 435.
[16] J.M. Davis, Chromatographia 44 (1997) 81.
[17] J.M. Davis, Anal. Chem. 69 (1997) 3796.

[18] F. Dondi, A. Bassi, A. Cavazzini, M.C. Pietrogrande, Anal. Chem. 70 (1998) 766.
[19] A. Felinger, L. Pasti, F. Dondi, Anal. Chem. 64 (1992) 2164.
[20] A. Felinger, Anal. Chem. 34 (1995) 2078.
[21] E.B. Ledford, J.B. Phillips, J.Z. Xu, R.B. Gaines, J. Blomberg, Am. Lab. 28(9) (1996) 22.
[22] J.C. Giddings, J. Chromatogr. A 703 (1995) 3.
[23] M.Z. El Fallah, M. Martin, Chromatographia 24 (1987) 115.
[24] J.M. Davis, Chromatographia 42 (1996) 367.
[25] A. Felinger, Anal. Chem. 69 (1997) 2976.
[26] L.J. Nagels, W.L. Creten, P.M. Vanpeperstraete, Anal. Chem. 55 (1983) 216.
[27] J.M. Davis, J.C. Giddings, Anal. Chem. 57 (1985) 2168.