Journal of Chromatography, 591 (1992) 1–18
Elsevier Science Publishers B.V., Amsterdam

CHROM. 23 800

Comparison of statistical theories of spot overlap in two-dimensional separations and verification of means for estimating the number of zones

Frances J. Oros and Joe M. Davis*

Department of Chemistry and Biochemistry, Southern Illinois University at Carbondale, Carbondale, IL 62901 (USA)

(Received July 29th, 1991)

ABSTRACT

Several theories appropriate to the modeling of zone overlap in two-dimensional (2D) separations are reviewed. One of these, with specific applications to chromatography, was recently proposed by one of the authors. The others are summarized in a book published over two decades ago. These latter theories addressed various problems of overlap (none of which was chromatographic), which were once of interest in the biological and hygienic sciences. Several hundred computer simulations of 2D beds containing various numbers of randomly distributed zones were generated to evaluate these theories. One theory shows exceptional promise in correctly predicting the numbers of singlet and multiplet spots expected in 2D beds over a wide range of chromatographic saturations. This theory predicts that only one sixth of the spot capacity in a 2D bed can be used to resolve spots. A procedure based on this theory is proposed, by which one can estimate the number of detectable zones in a 2D bed from the distribution of distances between the observed spots. The procedure was tested and verified by its application to about 200 computer simulations of 2D beds containing various numbers of randomly distributed zones.

INTRODUCTION

Recently, one of us (J.M.D.) reported equations for the numbers of singlet, doublet and triplet spots expected in two-dimensional (2D) separation beds containing randomly distributed zones [1]. The separations carried out in these beds are of the sequential type [2] and commonly are implemented in a rectangular bed, through which zones migrate at right-angles in two discrete stages. Such beds are used routinely in thin-layer/thin-layer chromatography and sodium dodecyl sulfate (SDS) electrophoresis/isoelectric focusing. The elution profiles of some recently developed coupled columns, *e.g.*, capillary zone electrophoresis–liquid chromatography [3] and gas chromatography–gas chromatography [4], also closely resemble the distribution of zones in such beds. This work was the logical extension of earlier work on the surprisingly profound statistical limitations of one-dimensional (1D) separations [5–22].

The equations for these spot numbers were tested by the interpretation of computer simulations of 2D beds containing randomly distributed zones [1]. When the space available for separation, as measured by the spot capacity of the bed, exceeded the number of zones by a factor of five or more, the equations' predictions agreed fairly well with the results of the simulations. At higher saturations, the numbers of singlet spots continued to agree well with the predictions of theory, but the numbers of doublet and triplet spots exceeded those predicted by theory.

Perhaps the most surprising conclusion drawn from this study was that the spot capacity of a 2D bed is utilized for separation less efficiently than is the peak capacity of a 1D chromatogram, at least when statistical considerations apply [1]. In other

0021-9673/92/$05.00

words, per unit capacity, 2D separations actually are worse than their 1D analogues. This conclusion also was reached by Martin [23]. In essence, the only reason that 2D separations are better than their 1D analogues is that one typically has so much more capacity in two dimensions than in one dimension that one can afford to "waste" a substantial fraction of it by ineffective utilization [1].

A specific example is cited to illustrate this point [1]. If a 100-component mixture were partially resolved on a 2D bed with a spot capacity of 2000 (a realistic value), then approximately 82 singlet spots, 7 doublet spots and 1 triplet spot would be expected (these spots account for 99% of the components). In contrast, if the same mixture were partially resolved on a 1D column having a peak capacity equal to 2000 (an unrealistic value, even in capillary electrophoresis), then 90 singlet peaks, 4 doublet peaks and no triplet peaks would be expected (these peaks account for 98% of the components). For equal capacities, the number of singlets (multiplets) is smaller (larger) in a 2D separation than in a 1D separation. Perhaps the reason why this phenomenon has not been observed experimentally is that unrealistically large 1D peak capacities are required to make the comparison.

After the submission of this work [1], one of us (F.J.O.) found a short but enlightening book by Roach [24], which details several studies during the 1940s–1960s in the biological and hygienic sciences of phenomena closely related to the overlap of spots in 2D beds. This book summarizes the results of several studies on counting errors, including those due to the underestimation of the numbers of bacterial colonies grown on culture plates, the numbers of dust particles collected on sampling plates and the numbers of coal particulates (to which miners were exposed) collected on thermal precipitators. These counting errors arose from the overlap of colonies or particulates by one another, which obscured some from detection. This obscuration is identical with that faced by the separation scientist. Several theories were proposed to account quantitatively for these overlaps. As will be shown below, most of these theories are inferior to that proposed by Davis. A theory proposed by Roach, however, accounts for overlap more correctly at higher saturations (*i.e.*, smaller spot capacities) than does the theory of Davis. In fact, in terms of its completeness and utility, the theory of Roach is perhaps the best theory considered to date by which to describe overlap in 2D beds.

This paper is divided into three parts. Because the studies addressed by Roach's book are probably unknown to most chromatographers, the first part of the paper summarizes these works and contrasts them with the recent results of Davis. In the second part, the theory of Roach is extensively tested by computer simulations of 2D beds containing various numbers of randomly distributed elliptical or circular zones. In the final part, a regressional procedure based on the theory of Roach is proposed, by which one can estimate the total number of detectable zones in a single 2D bed from the distribution of distances among the observed spots.

THEORY

Comparison of theories of spot overlap in two dimensions

According to Davis [1], the numbers of singlet spots s, doublet spots d, straight-chain triplet spots t_s and interlocking triplet spots t_i expected in a 2D bed of area A, which contains m randomly distributed detectable circular zones of diameter d_0, are

$$s = \bar{m}e^{-4\alpha} \tag{1a}$$

$$d = 8\bar{m} \cdot \frac{\alpha^2 e^{-8\alpha}}{1 - e^{-4\alpha}} \tag{1b}$$

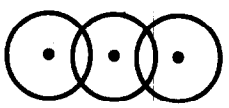
$$t_s \approx \frac{256}{5} \cdot \bar{m} \cdot \frac{\alpha^4 e^{-12\alpha}}{(1 - e^{-4\alpha})^2} \tag{1c}$$

$$\lim_{\alpha \to 0} t_i \approx \frac{512}{5} \cdot \bar{m} \cdot \frac{\alpha^6 e^{-12\alpha}}{[1 - (1 + 4\alpha)e^{-4\alpha}]^2} \tag{1d}$$

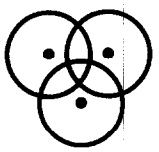
where $\bar{m}$ is a statistical approximation to m and α, the saturation of the bed, is the ratio of $\bar{m}$ to the spot capacity n_c. The spot capacity is defined as

$$n_c = A/A_0 = 4A/\pi(\beta d_0)^2 \tag{2}$$

where β is a scalar, which allows one to adjust for the degree of overlap that is acceptable in a given application (*e.g.*, zones are considered to be overlapped if they touch or overlay one another, when $\beta = 1$), and A_0 is the scaled area of a zone. This definition of spot capacity is different from that traditionally used in 2D separations [2] but is appropriate for this theory [1]. A straight-chain



a) straight-chain triplet t_s



b) interlocking triplet t_i

Fig. 1. Triplet spots (a) t_s and (b) t_i.

triplet t_s is defined by the overlap of a central zone by two other zones, which themselves do not overlap; an interlocking triplet t_i is defined by the overlap of a central zone by two other zones, which themselves overlap. These two types of triplets are illustrated in Fig. 1. The total number t of triplets is the arithmetic sum, $t_s + t_i$. Eqns. 1a–d were derived from selective interpretations of the radial distribution functions governing a 2D Poisson process.

The Taylor-series expansions of eqns. 1b–d will be useful below and are

$$d \approx \bar{m}\left(2\alpha - 12\alpha^2 + \frac{104}{3}\cdot\alpha^3 - 64\alpha^4 + \ldots\right) \quad (3a)$$

$$t_s \approx \bar{m}\left(\frac{16}{5}\cdot\alpha^2 - \frac{128}{5}\cdot\alpha^3 + \frac{1472}{15}\cdot\alpha^4 - \ldots\right) \quad (3b)$$

$$\lim_{\alpha\to 0} t_i \approx \bar{m}\left(\frac{8}{5}\cdot\alpha^2 - \frac{32}{3}\cdot\alpha^3 + \frac{608}{15}\cdot\alpha^4 - \ldots\right) \quad (3c)$$

The analysis of 500 computer simulations of 2D beds containing $m = 1000$ components showed that the number s of singlets predicted by eqn. 1a is correct to within 7%, as long as α is less than 0.50. The same analysis indicated some errors at modest levels of saturations in the equations for d, t_s and t_i; more specifically, these equations are correct to within 10% as long as α is less than about 0.20, 0.12 and 0.24, respectively. For α values less than 0.10, the simulations indicated that multiplets other than doublets and triplets are unlikely and that the number of observable spots p could be approximated by the arithmetic sum of eqns. 1a–d [1]:

$$p \approx s + d + t_s + t_i \quad \alpha \leqslant 0.10 \quad (4)$$

The other estimates of p summarized in Roach's book are reported below. The original notation has been changed to be compatible with that introduced above. For the sake of consistency, the number p in these theories is interpreted as the number of spots, although the overlapping zones in these theories actually corresponded to other phenomena, *e.g.*, bacterial colonies, dust, and coal particulates. (In passing, we note that the probabilities of forming singlet and multiplet 1D peaks from randomly distributed zones also are reported in this book, as is the mean number of 1D peaks. These results were not derived in terms of chromatography, however, but in terms of overlapping line segments.)

Bourdillon *et al.* [25], who were interested in the effects of the overcrowding of colonies on culture plates, approximated the total number p of spots on these plates as the solution to the differential equation

$$dp/d\bar{m} = 1 - p/n_c \quad (5a)$$

i.e.,

$$p = \bar{m}[1 - \exp(-\alpha)]/\alpha \quad (5b)$$

Experimental evidence was presented which confirmed the equation over a small range of α values.

Shortly thereafter, Irwin *et al.* [26], who were interested in the overlapping of dust particles on sampling plates, developed an alternative theory, in which the probability p_n that a zone forms part of a spot containing n overlapping zones was calculated from Poisson statistics as

$$p_n = \frac{(4\alpha)^{n-1}e^{-4\alpha}}{(n-1)!} \quad (6a)$$

The total number p of spots then was estimated as

$$p = \bar{m}\sum_{n=1}^{\infty} p_n/n = \bar{m}[1 - \exp(-4\alpha)]/4\alpha \quad (6b)$$

The result is identical with that of Bourdillon *et al.*, except that α is replaced by 4α. The authors recognized the existence of shortcomings in their theory, especially when α was sufficiently large that triplets and higher order multiplets would be observed.

Armitage [27], in a detailed subsequent study, approximated the number of singlet spots by eqn. 1a

and the numbers d, t_s and t_i of multiplet spots by the series

$$d \approx \overline{m}\left[2\alpha - \frac{2(4\pi + 3\sqrt{3})}{\pi}\alpha^2\right] = \overline{m}(2\alpha - 11.3080\alpha^2) \tag{7a}$$

$$t_s \approx \overline{m} \cdot \frac{6\sqrt{3}}{\pi} \cdot \alpha^2 = \overline{m}(3.3080)\alpha^2 \tag{7b}$$

$$t_i \approx \overline{m} \cdot \frac{2(4\pi - 3\sqrt{3})}{3\pi} \cdot \alpha^2 = \overline{m}(1.5640)\alpha^2 \tag{7c}$$

He further recognized that the derivation of higher order multiplets was prohibitively difficult by his approach; Roach commented on this difficulty [24]. By expanding eqn. 1a as a three-term Taylor series in α and combining this expansion with eqns. 4 and 7a–c, Armitage concluded that

$$p \approx \overline{m}\left(1 - 2\alpha + \left\{8 - \pi^{-1}\left[2(4\pi + 3\sqrt{3}) - 6\sqrt{3} - 2(4\pi - 3\sqrt{3})/3\right]\right\}\alpha^2\right) = \overline{m}(1 - 2\alpha + 1.564\alpha^2) \tag{8}$$

The theory of Roach [24] is described in some detail because, as will be shown below, its predictions closely agree with the numbers of spots found in extensive computer simulations carried out in this laboratory. Further, an understanding of Roach's theory is essential to the development of the regression procedure (see below) by which $\overline{m}$ can be estimated from the observed distribution of spots in a bed.

In his theory, Roach assumed that the centers of circular zones of diameter d_0 are randomly distributed throughout an unbounded bed (he observed that these zones could represent many phenomena, including dust particles, metal fumes, acid mists, air pollutants and airborne organisms [24]). Any zone in the bed is considered to be overlapped if its center lies within the distance d_0 of the center of an adjacent zone. In this work, for reasons that will be apparent shortly, this critical distance will be equated instead to βd_0, where β is the scaling factor previously introduced. In Roach's theory, $\beta = 1$.

The number of spots is determined in accordance with the following scheme. One arbitrarily selects any zone in the bed and finds its nearest neighbor. If the distance between these two zone centers is greater than βd_0, then the arbitrarily selected zone is a singlet spot. If the distance is less than βd_0, however, then the two zones overlap. In this latter instance, the overlapping pair forms either a doublet spot or a higher order multiplet (*e.g.*, triplet, quartet). One next finds the zone whose center lies closest to either of the centers of the two overlapping zones. In other words, if zones A and B are overlapped and lie in a bed containing zones A, B, C, D, etc., then one determines the distances between zones A and C, between zones A and D, between zones B and C, between zones B and D, etc., and then selects the shortest distance. If this distance is greater than βd_0, then the overlapping pair forms a doublet spot. If the distance is less than βd_0, however, then the selected zone overlaps with at least one of the two overlapping zones. In this latter instance, the overlapping trio forms either a triplet spot or a higher order multiplet (*e.g.*, quartet, quintet). One continues to repeat this sequence of nearest neighbor searching (exclusive of the neighbors comprising the spot) and distance comparison, until the distances between all n zones in the spot and their neighbors not in the spot are greater than βd_0. One then has isolated a spot consisting of n overlapping zones, or an n-tet spot, as shown in Fig. 2 for $n = 5$. One then repeats the above procedure by beginning with another zone and continues until all zones are addressed.

Roach recognized that the probability of these outcomes could be modeled closely (but not exactly) by the binomial distribution. The probability p_1 that the distance between a zone center and the center of its nearest neighbor is greater than βd_0 is [28]

$$p_1 = e^{-4\alpha} \tag{9}$$

The probability that the distance between a zone center and the center of its nearest neighbor is less than βd_0 is the complementary probability, $1 - p_1$. In an n-tet spot, one has $n-1$ sequences in which nearest neighbor distances are less than βd_0 and one sequence, that which breaks the spot away from other zones in the bed, in which a nearest neighbor distance is greater than βd_0. In the quintet in Fig. 2, for example, the four distances a, b, c and d are less

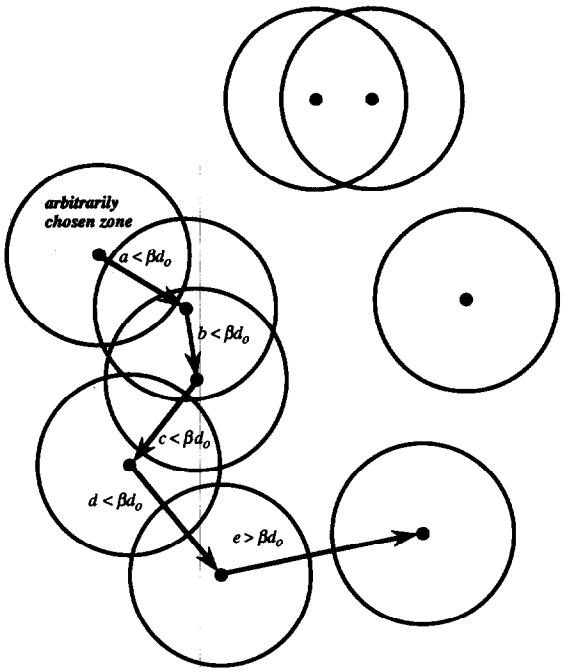


Fig. 2. Illustration of Roach's spot counting procedure. The five central zones overlap to form a quintet spot. The distance between any zone center in the spot and at least one other zone center in the spot is less than βd_0. The distances between all zone centers in the spot and all zone centers not in the spot are greater than βd_0.

than βd_0 and the one distance e is greater than βd_0. Roach recognized that these distances were *approximately* independent of each other, and it was this insight that reduced the 2D overlap problem to a tractable form. He concluded that the probability p_n that a zone forms part of a spot containing n overlapping zones is [24]

$$p_n = e^{-4\alpha}(1 - e^{-4\alpha})^{n-1} \tag{10}$$

Because each of $\bar{m}$ zones has this probability of forming an n-tet spot, the number of zones contributing to the formation of n-tet spots is $\bar{m}p_n$. But since n zones are required to form each spot, the expected number P_n of n-tet spots is $\bar{m}p_n/n$, or

$$P_n = \bar{m}e^{-4\alpha}(1 - e^{-4\alpha})^{n-1}/n \tag{11}$$

Roach then calculated the total number p of spots as the algebraic sum of all n-tet spots:

$$p = \sum_{n=1}^{\infty} P_n = \bar{m} \cdot \frac{4\alpha e^{-4\alpha}}{1 - e^{-4\alpha}} \tag{12}$$

The remainder of this paper principally will focus on the testing and use of this theory, although the other theories are also characterized.

Detailed testing of Roach's theory

To our knowledge, the only test of Roach's theory is presented in his book and is limited. In this test, Roach determined the coordinates of hundreds of points with a random number generator having six significant figures, plotted these points on graph paper, drew with a compass circles of fixed diameter about these points and manually counted the number of overlaps so generated ($\beta = 1$). More specifically, he determined the numbers of singlets and multiplets (up to septets) for m values of 100, 150, 200, 250, 300, 350, 400, 500, 600, 800, 1100, 1500 and 2220 [24]. The spot numbers so determined agreed fairly well with his theory, as will be shown below.

We have augmented substantially this testing of Roach's theory for two reasons. First, for each value of m cited above, the singlet and multiplet numbers reported by Roach were determined from only one simulation and consequently are not statistically robust. Second, and more important, zones in 2D separations actually are elliptical [2], instead of circular. In Roach's theory (indeed, in all of the theories outlined above), however, zones are assumed to be circular. The error introduced by this assumption has not been quantified, except in a cursory way [1], and is investigated here.

In general, a theory for the overlap of elliptical zones is considerably more difficult to develop than that for circular zones, because one no longer can assume that nearest neighbors overlap [1]. The overlap of randomly distributed convex laminae of arbitrary shape has been addressed by Mack [29,30] and can be adapted to elliptical zones, as shown by Roach [24]. Unfortunately, this theory is not applicable to 2D separation beds, because Mack assumed that the laminae were randomly ordered in the bed (he was interested in the random deposition of dust particles). Davis addressed the problem in a prelimi-

nary way by calculating the probability $p_1(e)$ that an elliptical zone of aspect ratio γ (*i.e.*, the ratio of the ellipse's semi-major axis to its semi-minor axis) is a singlet, which is [1]

$$p_1(e) = (2\pi)^{-1} \int_0^{2\pi} \exp\left(-4\alpha \cdot \frac{\gamma}{\cos^2\theta + \gamma^2 \sin^2\theta} \right) d\theta \tag{13}$$

He concluded that the probability of forming elliptical singlets is greater than the probability of forming circular singlets at any saturation α and that this probability increases with both α and γ.

The applicability of Roach's theory to elliptical zones, and the further testing of this theory by the overlap of circular zones, were characterized here by computer simulation. Spots p, singlets s and various multiplets were counted by visual inspection. For completeness, these simulation results also were compared with the predictions of Davis.

Procedure by which $\bar{m}$ can be estimated from a single bed

As was suggested by Davis [1], the number p of spots in a 2D bed can be counted and the spot capacity n_c of the bed can be estimated. One then can fit these data to eqn. 4 and estimate the number $\bar{m}$ of detectable components in the bed. The principal motive for this estimation is determining the extent of separation [15], the square of which equals the ratio, $s/\bar{m}$. This ratio, in turn, is a measure of the quality of the separation. The calculation of this ratio is straightforward, because once $\bar{m}$ has been determined, α can be estimated from the definition, $\bar{m}/n_c$, and s can be estimated from eqn. 1a, $\bar{m}$ and α.

An alternative procedure, by which one can estimate $\bar{m}$ from the distribution of spots in a 2D bed, is suggested here. This procedure is analogous to the "single-chromatogram method" based on the statistical model of overlap (SMO) [13], by which one can estimate the number of components in a 1D chromatogram from the distribution of distances (or times) between adjacent maxima. In essence, one determines by this procedure a series of spot numbers p, in accordance with the theory of Roach. The numbers p differ for each determination, however, because one sets the span βd_0 to different values by arbitrarily changing the scalar, β. In other words, different values of βd_0 (*i.e.*, different values of n_c^{-1}) correspond to different values of p, in accordance with eqns. 2 and 12 and the definition, $\alpha = \bar{m}/n_c$. With this scaling, one can generate a set of coordinates pairs, (n_c^{-1}, p) and then fit them to eqn. 12 with $\bar{m}$ as a least-squares parameter. One would expect this fitting usually to provide a better estimate of $\bar{m}$ than that determined from a single datum, simply because one has more information on which to draw.

It should be emphasized that the numbers p and n_c so generated have little, if any, physico-chemical meaning. They are to be distinguished from the physically meaningful spot number p, which corre-

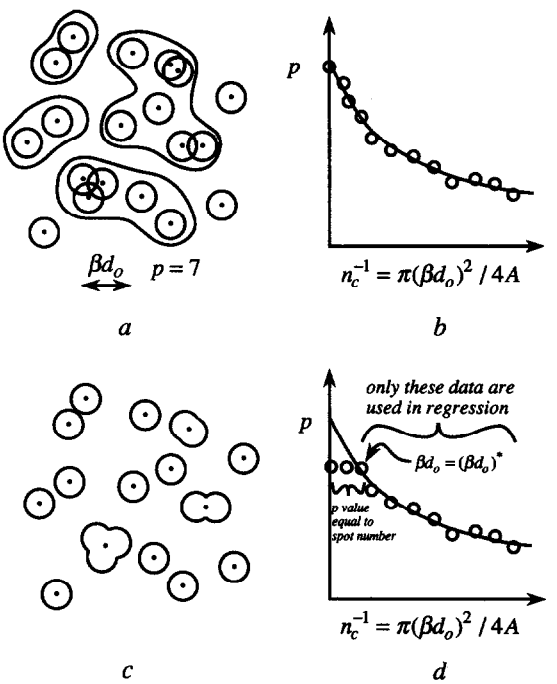


Fig. 3. Illustration of procedure by which $\bar{m}$ is estimated from a single bed. (a) Hypothetical bed in which all zone centers are visible, even when overlapped. The number $p = 7$ for the indicated value of βd_0. The groups of encircled zones each correspond to a "spot". (b) Graph of number p *vs.* reciprocal capacity n_c^{-1} constructed from the distribution of distances among zone centers in (a). Solid curve represents least-squares fit of data to eqn. 12. (c) More realistic bed in which some zone centers are obscured by overlap and spot positions are represented by a single coordinate. (d) Graph of number p *vs.* reciprocal capacity n_c^{-1} constructed from the distribution of distances among spot coordinates in (c). For βd_0 values less than $(\beta d_0)^*$, p equals the total number of spots. Solid curve represents least-squares fit of indicated data to eqn. 12.

sponds to the number of distinct clusters in the bed, and from the physically meaningful spot capacity n_c, which corresponds to the number of spots resolvable by the bed. However, we shall use the same variable to describe both quantities, because they are indistinguishable in a mathematical sense. The distinction should be apparent from context.

The generation and interpretation of these data is illustrated in Fig. 3. Fig. 3a is part of a hypothetical 2D bed in which the centers of all zones are observable, even when the zones overlap. For the βd_0 value in the figure, p equals seven and corresponds to the three singlets and the four groups of zones, which are encircled. Lest any misunderstanding arise, each zone center in an encircled group is within distance βd_0 of some other zone center in the group, and all zones in an encircled group are separated from all zones not in the group by distances greater than βd_0. A different value of βd_0 (or n_c^{-1}) would correspond to a different value of p, in accordance with the above discussion. The data are plotted in Fig. 3b as p *vs.* n_c^{-1}, the latter of which is directly proportional to βd_0. The curve represents a non-linear least-squares fitting of the data so generated to eqn. 12. Because all the zone centers in the bed are observable, the value of p at $\beta d_0 = n_c^{-1} = 0$ corresponds to the total number of zones.

In an actual bed, however, overlap obscures the positions of some zone centers, as shown in Fig. 3c. To apply the procedure to such a bed, one must approximate the positions of these centers by other easily measurable positions, such as the positions of maximum concentration or the first moments of spots. In general, the centers of closely overlapping zones are obliterated completely by overlap, and the remaining measurable positions (however they are determined) are somewhat removed from each other. Because of this obliteration, the distances between measurable positions are all greater than a critical value of βd_0, which is designated here as $(\beta d_0)^*$. For values of βd_0 less than or equal to $(\beta d_0)^*$, the number p is independent of βd_0 and equals the number of spot centers in the bed. As βd_0 is increased beyond $(\beta d_0)^*$, however, the number p decreases, as shown in Fig. 3d. Only the data for which $\beta d_0 \geqslant (\beta d_0)^*$ are fitted to theory.

At low saturations, only a few multiplets exist, and the positions of only a few zone centers are distorted by overlap. In this instance, the data plotted as p *vs.* n_c^{-1} in Fig. 3d differ little from those in Fig. 3b [except when $\beta d_0 < (\beta d_0)^*$], and a fitting of the appropriate data to eqn. 12 provides a good approximation to $\bar{m}$. As the saturation increases, however, the numbers of multiplets increase, and eventually the measurable positions of spots have little correlation with the positions of the underlying zone centers. In this instance, the least-squares fitting of the appropriate data provides a poor estimate of $\bar{m}$.

We should emphasize that this procedure will work well only if the theory to which the coordinates, (n_c^{-1}, p), are fitted correctly describes the numbers p at large values of α. This limitation arises because large α values are determined by selecting arbitrarily large values of β. As is shown below, only the theory of Roach correctly describes the number p at high saturations. Hence, only his theory was used to test the procedure proposed here. Computer simulations were analyzed to determine the saturation below which reliable estimates of $\bar{m}$ can be calculated.

The reader may find a complete 1D analogy to this procedure in ref. 13.

Methods of spot counting

One can envision at least three ways by which spots in a 2D bed can be counted. The first way, in which some degree of physical overlap is tolerated among resolved zones, corresponds to a β value of less than one [1]. The second way, in which zones that overlay one another by any amount are considered to be overlapped, corresponds to a β value equal to one. This method of counting previously was employed by both Roach [24] and Davis [1]. The final way, in which clusters of spatially resolved zones are grouped together and identified as "spots", as discussed above, corresponds to a β value greater than one. The latter two ways of counting, but not the first, are used here. More specifically, the values of p, s and multiplet numbers determined by visual inspection of 2D beds correspond to $\beta = 1$, and the values of p determined by applying the regression procedure correspond to $\beta \geqslant 1$. In a similar manner, peaks in 1D chromatograms have been counted in applications of the SMO by using various values of the resolution factor R_s^* (which is analogous to β), such as 0.5 [10–12], 1.0 [16] and 1.5 [10,11].

PROCEDURES

As observed in the second part of the Theory section, computer simulations of 2D beds were produced to augment the testing of the theory of Roach. These beds contained various numbers m of circular or elliptical zones; more specifically, m was varied among the values, 25, 50, 150, 300, 500, 750 and 1000. The coordinates of each zone center were computed by two sequential subroutine calls on an in-house random number generator adapted from ref. 31. The zones then were plotted as circles or as ellipses with aspect ratios equal to two with the Macintosh application, KaleidaGraph 2.0.2 (Synergy Software, Reading, PA, USA). These graphs then were plotted by either a dot matrix (ImageWriter II, Apple Computer, Cupertino, CA, USA) or laser printer (LaserWriter II). The various singlet and multiplet spots in these plots were counted, such that zones which touched or overlaid one another were considered to be overlapped ($\beta = 1$). A total of 240 simulations of beds containing circular zones and of 170 simulations of beds containing elliptical zones were interpreted.

The choice of an elliptical aspect ratio equal to two was completely arbitrary. Indeed, a study based on widely varying aspect ratios would be desirable. Such a study is postponed to a later date, however, to limit the data here to an amount that can be assessed in detail.

For both the circular and elliptical zones, a minimum of three and a maximum of five simulations were generated for each value of m at a particular α. The sizes of the circular and elliptical zones were limited to those available in KaleidaGraph 2.0.2. Various saturations were obtained for each m by combining various zone sizes with various square and rectangular bed shapes. The values of the bed perimeters were set in KaleidaGraph 2.0.2 (and confirmed by measuring them with a ruler), and the areas of several singlet zones were measured with calipers to three significant figures. From the measured areas, the spot capacity n_c was calculated in accordance with eqn. 2 ($\beta = 1$) and α was approximated as m/n_c. The various α values corresponding to the simulations are reported in Table I. The α values for the circular and elliptical zones typically are not equal, because of the constraints imposed by KaleidaGraph 2.0.2. Further, the α values do not form a simple numerical progression (*e.g.*, $\alpha = 0.05$, 0.10, 0.15, etc.), because of constraints on the sizes of zones and beds.

TABLE I

VALUES OF α CORRESPONDING TO COMPUTER SIMULATIONS OF 2D BEDS CONTAINING CIRCULAR AND ELLIPTICAL ZONES

m	α ($\beta = 1$)	
	Circular zones	Elliptical zones ($\gamma = 2$)
25	0.0287, 0.0470, 0.0696, 0.100, 0.130, 0.164, 0.202, 0.260, 0.298, 0.347	0.0466, 0.0976, 0.122, 0.163, 0.204, 0.248, 0.292, 0.350
50	0.0124, 0.0291, 0.0558, 0.0950, 0.139, 0.195, 0.258, 0.332, 0.403, 0.505	0.0302, 0.0573, 0.137, 0.193, 0.252, 0.331, 0.407, 0.497
150	0.0372, 0.0874, 0.172, 0.282, 0.418, 0.595	0.0325, 0.0906, 0.175, 0.294, 0.416, 0.560
300	0.0688, 0.131, 0.175, 0.259, 0.345, 0.423, 0.563	0.0651, 0.181, 0.350, 0.577
500	0.0896, 0.119, 0.219, 0.291, 0.431, 0.574	0.108, 0.137, 0.302, 0.584
750	0.134, 0.169, 0.328, 0.437, 0.647	0.163, 0.206, 0.453, 0.539
1000	0.172, 0.239, 0.438, 0.583	0.217, 0.275, 0.604, 0.719

We briefly observe that the α values calculated as described above for ellipses (but not for circles) varied with the output device. In other words, the identical graph, when plotted on the dot matrix and laser printers, resulted in slightly different α values (on average, the values varied by about 3%). Some of the slight variations reported in Table I can be attributed to this behavior. Also, some of the beds of elliptical zones, when plotted with the laser printer, contained ellipses with γ values lying between 1.9 and 2.0, instead of γs which exactly equalled 2.0. We tolerated this small error (5% or less), instead of rejecting the graph.

Fig. 4 reports five simulations of square beds containing $m = 50$ zones of different size. The sequence of random numbers defining the zone centers is the same in each simulation. The various beds span the range $0.056 \leqslant \alpha \leqslant 0.258$ ($\beta = 1$). As

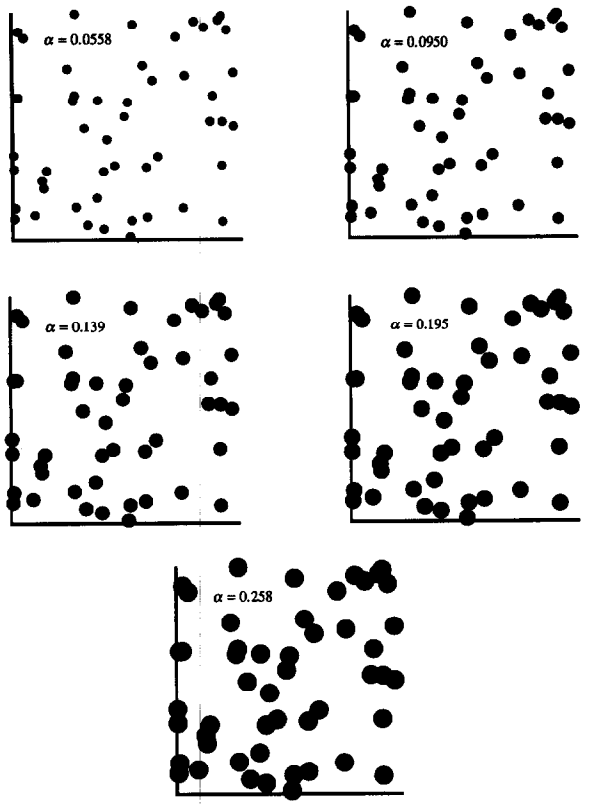


Fig. 4. Computer-simulated beds containing $m = 50$ components at various saturations α ($\beta = 1$). Coordinates of zones centers in each bed are identical.

observed above, the larger α values are not simple multiples of the smallest α, because of constraints on the generation of these plots. These beds should prove useful in gauging the severity of overlap at different saturations.

A preliminary testing of the regression procedure outlined in the third part of the Theory section was implemented by generating several computer simulations of square 2D beds containing $m = 100$, 200 and 300 zones. For each m, ten simulations were produced at each of the α values, 0, 0.05, 0.10, 0.15, 0.20 and 0.25 ($\beta = 1$). In contrast to the study described above, α could be assigned regularly spaced values here, because KaleidaGraph 2.0.2 was not used to generate beds for visual inspection. Rather, all necessary computations were implemented with a computer algorithm. In total, 180 simulations (ten simulations per α times six α values times three m values) were generated. The regression procedure was applied to each simulation.

The coordinates of the zone centers in these beds were computed by the in-house random number generator described above. For each m, a different sequence of coordinates was generated for each of the ten simulations. For each m, however, the same ten sequences of coordinates were used to generate simulations at different α values. This usage made the beds for any one value of m identical, except for the value assigned to α.

We recognized that our $\bar{m}$ estimates would not equal m exactly, even under the best of circumstances, because only a limited number of beds were sampled and because small systematic errors existed in the random number generator. Consequently, we decided to use the $\bar{m}$ estimates determined from beds for which $\alpha = 0$ as a basis by which to evaluate the quality of other $\bar{m}$ estimates. For $\alpha = 0$, the zone diameter d_0 was equated to zero. In other words, these beds contained only zone centers, *i.e.*, a sequence of randomly spaced points. (The resultant beds are analogous to the "line chromatograms" of the 1D "single-chromatogram method" [13].) The distances between each point and its neighbors were calculated. From these distances the coordinates (n_c^{-1}, p) were determined as detailed in the Theory section (here, the various non-zero spans βd_0 were interpreted as a series of arbitrary distances, as $d_0 = 0$). These coordinates then were fitted to eqn. 12, except for the coordinate $(n_c^{-1} = 0, p)$. This coordinate was specifically excluded, because its inclusion forced the $\bar{m}$ estimate to equal m. Approximately 30 to 40 data points were generated for purposes of regression. No value of p less than 20 was included in a fit to minimize the effect of small-number statistics [13].

In the other simulations, the zone diameter d_0 was determined in accordance with assigned values of $\alpha \approx m/n_c$, the spot capacity, n_c, as defined by eqn. 2 and the criterion $\beta = 1$ (no elliptical zones were considered in this part of the study). The spot number was counted in accordance with the criterion $\beta = 1$; zones separated by spans less than d_0 were considered to overlap. Once the number of spots in a given simulation had been determined, then a single representative coordinate was assigned to each spot. The representative coordinate of a singlet spot was the coordinate of the zone center, and the representative coordinate of a multiplet spot

was the arithmetic average of the coordinates of the zone centers comprising the multiplet spot. In general, the representative coordinates of multiplets determined by these means do not correspond to any physically measurable positions. They are simple to determine, however, and are sufficient for our purpose here, which is simply to verify the regression procedure (the optimization of the procedure is addressed later). The distances between these coordinates were then calculated to determine the data

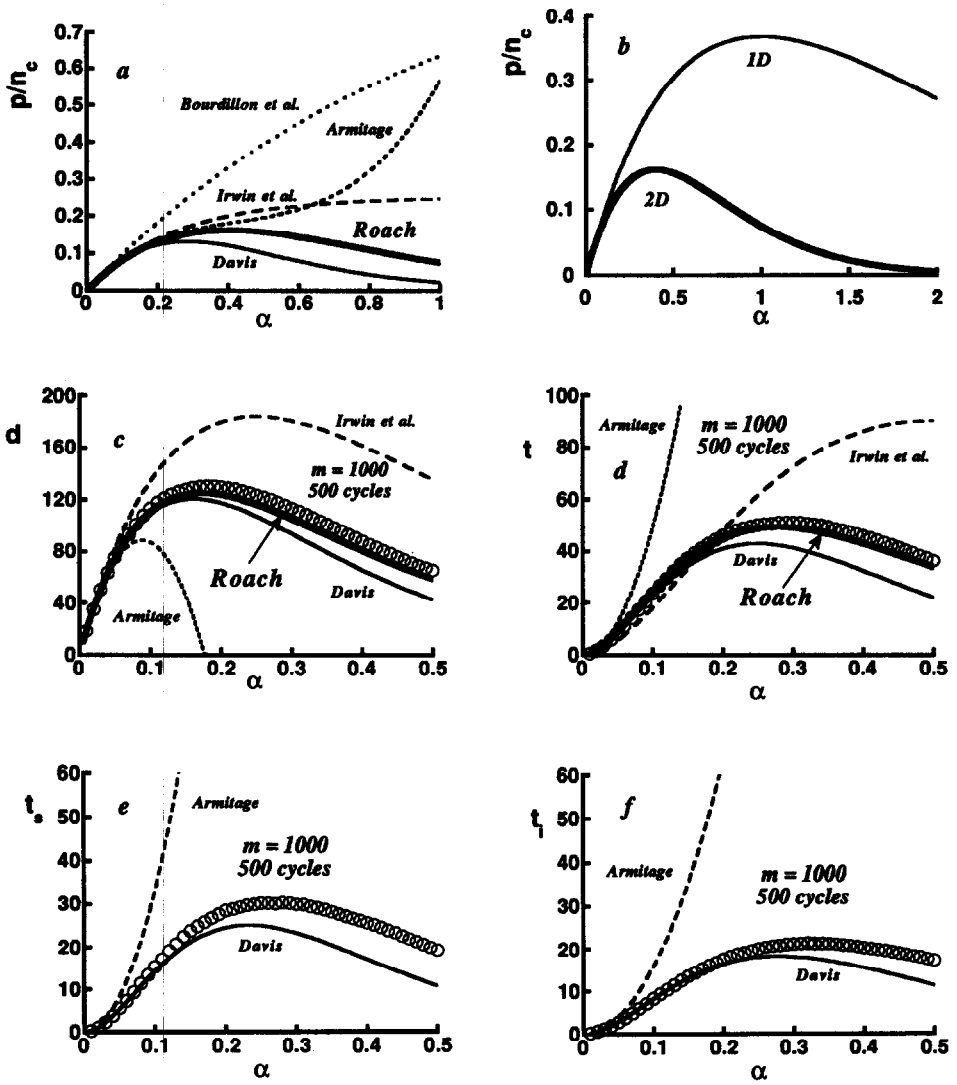


Fig. 5. (a) Graphs of p/n_c *vs.* α predicted by Davis (eqns. 1 and 4), Bourdillon *et al.* (eqn. 5b), Irwin *et al.* (eqn. 6b), Armitage (eqn. 8) and Roach (eqn. 14). All equations but the last were divided by n_c. (b) Graphs of p/n_c *vs.* α for 1D and 2D separations. The former graph is defined by $\alpha \exp(-\alpha)$ and the latter by eqn. 14. (c) Graphs of doublet number d *vs.* α ($\beta = 1$) predicted by Davis (eqn. 1b), Irwin *et al.* (eqn. 6a, with $n = 2$, multiplied by $\bar{m}$), Armitage (eqn. 7a) and Roach (eqn. 11, with $n = 2$) for $m = \bar{m} = 1000$ zones distributed randomly in a square bed. Circles represent the average numbers of doublets determined from 500 computer simulations. (d) Graphs of triplet number t *vs.* α ($\beta = 1$) predicted by Davis (sum of eqns. 1c and d), Irwin *et al.* (eqn. 6a, with $n = 3$, multiplied by $\bar{m}$), Armitage (sum of eqns. 7b and c) and Roach (eqn. 11, with $n = 3$) for $m = \bar{m} = 1000$ zones distributed randomly in a square bed. Circles represent the results of simulations, as detailed above. (e) Graphs of straight-chain triplet number t_s *vs.* α ($\beta = 1$) predicted by Davis (eqn. 1c) and Armitage (eqn. 7b). Circles represent the results of simulations, as detailed above. (f) Graphs of interlocking triplet number t_i *vs.* α ($\beta = 1$) predicted by Davis (eqn. 1d) and Armitage (eqn. 7c). Circles represent the results of simulations, as detailed above.

pairs (n_c^{-1}, p), as detailed above. Approximately 30–40 data pairs were generated for each regression. As before, no values of p less than 20 were used.

For each simulation, these coordinates were plotted in KaleidaGraph 2.0.2 and examined. The data for which $\beta d_0 < (\beta d_0)^*$ were excluded from the non-linear regression. Some subjective judgment was required in the inclusion or elimination of some of these data, as is also required in the application of the SMO.

The algorithm necessary for these computations was written in FORTRAN 77 and executed on the GX-3081 computer at Southern Illinois University. The least-squares regressions were carried out in KaleidaGraph 2.0.2.

RESULTS AND DISCUSSION

Fig. 5a is a graph of the various theoretical estimates of the ratio p/n_c *vs.* α. These ratios are obtained by dividing both sides of eqns. 1a–d and 4, 5b, 6b, 8 and 12 by n_c and then substituting the identity $\alpha = \bar{m}/n_c$. The specific result for the theory of Roach, which will be examined in more detail below, is

$$p/n_c = \frac{4\alpha^2 \exp(-4\alpha)}{1 - \exp(-4\alpha)} \tag{14}$$

and is plotted in bold. All of the theories predict essentially identical results for sufficiently small values of α (*e.g.*, $\alpha < 0.10$). Relative to the results predicted by Roach, the p/n_c estimates predicted by Bourdillon *et al.*, Irwin *et al.* and Armitage are unduly large at high saturations. Further, at large saturations, the theory of Irwin *et al.* unrealistically predicts that p/n_c approaches the constant value 1/4 (for similar reasons, Bourdillon *et al.*'s theory approaches $p/n_c = 1$ at α values larger than shown here). These findings are indicative of errors in these theories. One observes that the theory of Davis underestimates p/n_c at α values greater than *ca.* 0.20 relative to the results of Roach. This underestimation is expected, however, in part because multiplets larger than triplets are neglected in this theory.

By differentiating eqn. 14 with respect to α, equating the derivative to zero and solving the resultant equation numerically, one can show that the maximum value of p/n_c, as predicted by Roach, is 0.162 and is found at $\alpha = 0.398$. In other words, only about one sixth of the entire spot capacity of a bed can be used to resolve spots under the best of circumstances, when statistical considerations apply (as observed elsewhere [1], less than one tenth of the bed can be used to resolve singlet spots). This fraction is considerably smaller than the maximum number of peaks resolvable per unit peak capacity in 1D chromatograms, which is $e^{-1} = 0.368$ [7]. Fig. 5b is a graph of p/n_c *vs.* α, in which the ratios p/n_c for 1D and 2D separations are compared (the theoretical relationship for the 1D separation is $\alpha \exp(-\alpha)$ [7]). At all saturations, the available capacity is utilized less effectively in two dimensions than in one dimension. This graph supports the conclusion asserted in previous works [1,23] and reported in the Introduction that per unit capacity, 2D separations are actually worse than their 1D analogues.

The theories of Irwin *et al.*, Armitage, Roach and Davis all predict that the number of singlets is given by eqn. 1a. However, their predictions of the numbers of multiplet spots differ substantially. Fig. 5c is a graph of the number of doublets d *vs.* α predicted by these four theories for $\bar{m} = 1000$. The circles represent the average numbers of doublets found in 500 computer simulations of 2D beds, each of which contained $m = 1000$ zones distributed randomly in a square area (these data have been reported elsewhere [1]). The number d is correctly predicted by all four theories, when $\alpha <$ *ca.* 0.05. The theories of Irwin *et al.* and Armitage, however, clearly break down at levels of saturation greater than this. For values of $\alpha >$ *ca.* 0.15, the theory of Davis underestimates d, although the theoretical prediction more or less parallels the results of the computer simulations. The theory of Roach also slightly underestimates these results at higher saturations but agrees with them more closely than that of Davis. Similar trends are indicated in Fig. 5d, which is a graph of the total number $t = t_s + t_i$ of triplets *vs.* α predicted by all four theories and also determined by computer simulation. In particular, the theories of Irwin *et al.* and Armitage agree with the simulation results only for very small values of α.

Only the theories of Armitage and Davis distinguish between the two triplet types, t_s and t_i. Fig. 5e and f are graphs of t_s and t_i *vs.* α predicted by these theories and also determined by computer simulation. The theory of Davis has some shortcomings at large saturations but correctly describes both t_s

and t_i at modest saturations. The predictions of Armitage, however, agree with simulation only for extremely small values of α and depart drastically from these simulations at higher saturations.

It is instructive to compare the Taylor-series expansions of the equations derived by Davis to the series approximations derived by Armitage. A comparison of eqns. 3a and 7a shows that the terms linear in α are identical and that the quadratic terms in α differ by -5.76% if one arbitrarily normalizes the difference by the prediction of Davis. This difference is not unduly large, but because the theory of Armitage predicts d only to quadratic powers of α, it breaks down at even small α values. The comparisons of eqns. 3b and 7b and of eqns. 3c and 7c show that the quadratic terms in α of t_s and t_i are also close (for t_s the difference is 3.38% and for t_i it is -2.25%, if one again normalizes the differences by the predictions of Davis). In his derivation, Davis analytically determined the functional forms of t_s and t_i but had to determine the coefficients by which to multiply these forms by empirical means [1]. He suggested that someone probably would theoretically account for the values of these coefficients someday [1]. Little did he realize the work had already been done (at least for the α^2 term) some 42 years earlier!

The simulations described above were determined with a large m value (1000) to ensure that enough events were included for the statistics to be meaningful [1]. Fig. 6 is a plot of the dimensionless functions p/n_c, s/n_c and d/n_c *vs.* α ($\beta = 1$) as determined from the visual inspection of several hundred simulations of 2D beds containing a wide range of m values ($25 \leqslant m \leqslant 1000$). The various solid (dashed) curves in these figures represent the predictions of Roach (Davis), unless noted otherwise. The solid curves in Fig. 6a and b are graphs of eqn. 14, whereas the dashed curves are graphs of eqn. 4, divided by n_c, with s, d, t_s and t_i expressed by eqns. 1a–d. The solid curves in Fig. 6c and d are graphs of eqn. 1a, divided by n_c (here, the predictions of Roach and Davis are identical), whereas the dashed curve in Fig. 6d is a graph of eqn. 13, multiplied by α. The solid curves in Fig. 6e and f are graphs of eqn. 11, divided by n_c, with $n = 2$, whereas the dashed curves are graphs of eqn. 1b, divided by n_c. The numbers m of zones corresponding to the various symbols are identified in the figure caption. Not all of the available data are plotted in order to minimize the overlap of various symbols. The open diamonds in Fig. 6a, c and e represent the single determinations of Roach.

The division of the indicated equations by n_c allows one to make the substitution $\alpha = \bar{m}/n_c$, such that these functions depend only on α. Further, because eqn. 13 equals the probability $p_1(e)$ of forming elliptical singlets, one must multiply it by $\alpha = \bar{m}/n_c$ to obtain the number $s = \bar{m}p_1(e)$ of elliptical singlets per capacity n_c. For a specified value of γ, this expression also depends only on α. These various functions are represented in dimensionless form to enable us to represent spot numbers corresponding to widely varying m values in the same figure.

Except for the open diamonds, the symbols in Fig. 6 represent the means of spot numbers determined in this laboratory, and the error bars represent the standard deviations of these numbers (the standard deviations are not shown when they are smaller than the symbols). In 6a, c and e, the spots were formed from circular zones and in Fig. 6b, d and f from elliptical zones ($\gamma = 2$). For both zone shapes, the largest deviations at any α are associated with the smallest values of m. This finding, which also was observed in the 1D analogue to this study [32], is unsurprising and simply indicates that the relative standard deviation for Poisson processes increases with decreasing values of $\bar{m}$. Also, for any m, the largest deviations are associated with large α values. This observation again is unsurprising, because spot numbers decrease with increasing α and the relative deviation of any spot number from its small mean value consequently increases. This trend was also observed in the 1D analogue to this study [32]. Because of the large deviations associated with small spot numbers, no values of p, s or d less than 15 are represented in the figure.

The values of p/n_c shown in Fig. 6a agree closely with Roach's theory for αs less than *ca.* 0.30 ($\beta = 1$). At higher saturations, the values of p/n_c are slightly larger than expected, although this difference probably is not statistically significant (the theory typically lies within one standard deviation of the data). The results also agree well with the theory of Davis for α values less than *ca.* 0.20. This finding may be surprising, as Davis's theory for triplet number t_s underestimates this multiplet at α values larger than *ca.* 0.12 (see Theory section). The reason why the

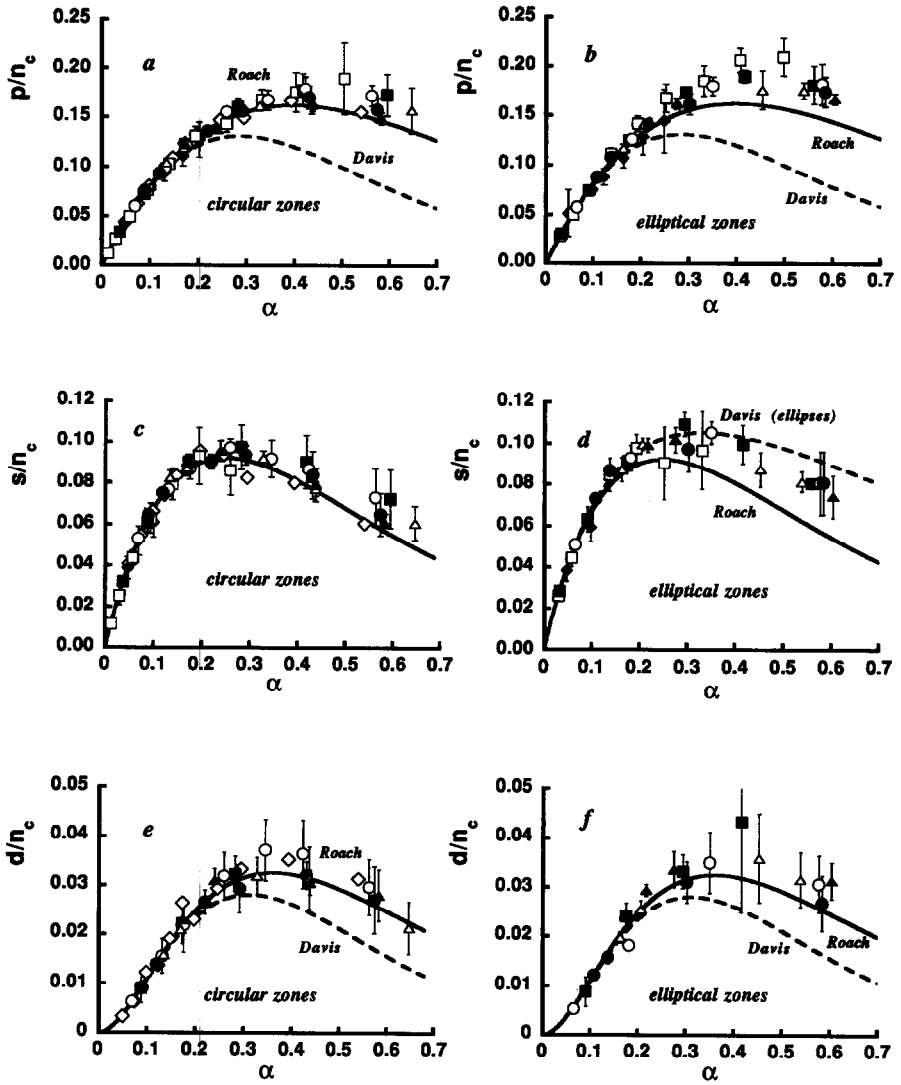


Fig. 6. Graphs of p/n_c *vs.* α ($\beta = 1$) for (a) circular and (b) elliptical zones, as determined by visual inspection. Solid curves are graphs of eqn. 14; dashed curves are graphs of eqn. 4, divided by n_c, with s, d, t_s and t_i expressed by eqns. 1a–d. Graphs of s/n_c *vs.* α ($\beta = 1$) for (c) circular and (d) elliptical zones, as determined by visual inspection. Solid curves are graphs of eqn. 1a, divided by n_c; dashed curve in Fig. 6d is a graph of eqn. 13, multiplied by α. Graphs of d/n_c *vs.* α ($\beta = 1$) for (e) circular and (f) elliptical zones, as determined by visual inspection. Solid curves are graphs of eqn. 11, divided by n_c, with $n = 2$; dashed curves are graphs of eqn. 1b, divided by n_c. Symbols corresponding to different m values are: ◆ = 25; □ = 50; ■ = 150; ○ = 300; ● = 500; △ = 750; ▲ = 1000; ◇ = Roach's data.

numbers p (or p/n_c) do not appear to be erroneous over the α range $0.12 \leqslant \alpha \leqslant 0.20$ is that the "excess" t_s triplets actually found partially compensate for the neglect of quartets and other multiplets, which become increasingly prevalent at saturations above 0.10. (A similar trend is found in the work of Armitage, whose theory severely underestimates doublets and overestimates triplets at even low α values but correctly estimates p up to $\alpha \approx 0.30$ [27], as shown in Fig. 5a.) At saturations greater than *ca.* 0.20, however, the theory of Davis underestimates p/n_c.

As shown in Fig. 6b, Roach's theory also correctly predicts the numbers p/n_c determined by counting elliptical spots ($\gamma = 2$), when α is less than *ca.* 0.25 or so. At larger α values, the numbers p/n_c (and consequently p) are greater than predicted by Roach. Unlike in Fig. 6a, this difference is statistically significant; the differences between theory and simulation are greater than one standard deviation (except for one datum) when α is greater than *ca.* 0.30. The theory of Davis adequately predicts p/n_c for α values less than *ca.* 0.20 but underestimates p/n_c at larger α values.

Fig. 6c is a plot of s/n_c *vs.* α, as determined by counting circular spots. The agreement between simulation and theory is good over the full α range, although it is best at low α values. At larger α values,

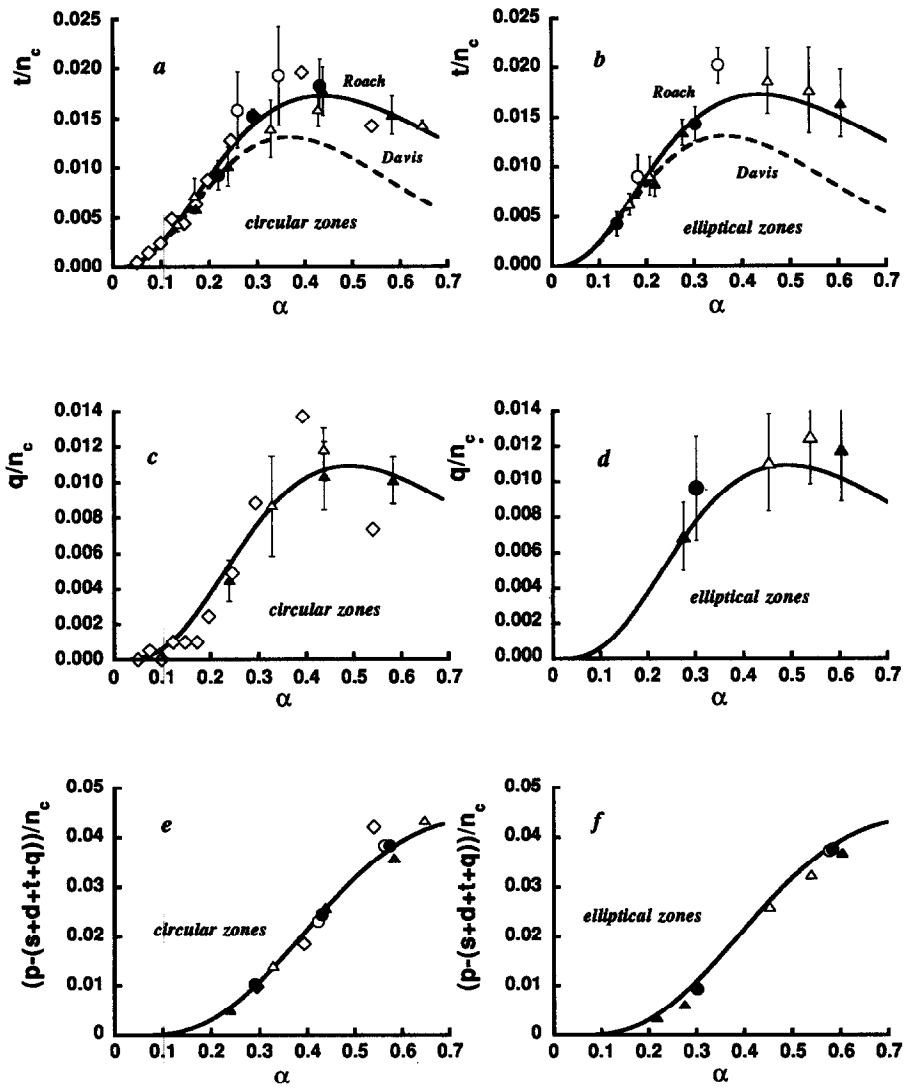


Fig. 7. Graphs of t/n_c *vs.* α ($\beta = 1$) for (a) circular and (b) elliptical zones. Solid curves are graphs of eqn. 11, divided by n_c, with $n = 3$; dashed curves are graphs of the algebraic sum of eqns. 1c and d, divided by n_c. Graphs of q/n_c *vs.* α ($\beta = 1$) for (c) circular and (d) elliptical zones. Solid curves are graphs of eqn. 11, divided by n_c, with $n = 4$. Graphs of $[p - (s + d + t + q)]/n_c$ *vs.* α ($\beta = 1$) for (e) circular and (f) elliptical zones. Solid curves are graphs of eqn. 14, less the algebraic sum of four different expressions of eqn. 11, all divided by n_c. These four expressions are defined by $n = 1, 2, 3$ and 4. Symbols corresponding to different m values are ○ = 300; ● = 500; △ = 750; ▲ = 1000; ◇ = Roach's data.

values of s/n_c slightly exceed theory but not significantly so. In part, these small discrepancies are due to "edge effects", in which zones near the border of the 2D bed are partially shielded from overlap by the absence of zones outside the bed [1]. As shown in Fig. 6d, the theory for circular singlets correctly predicts the number of elliptical singlets ($\gamma = 2$) when α is smaller than *ca.* 0.20. In contrast, this theory predicts results that are statistically smaller than the simulation results when α is larger than *ca.* 0.20. This finding is in qualitative agreement with the prediction of Davis, who argued that singlet formation is more probable for ellipses than for circles [1]. The dashed curve is the quantitative prediction of Davis ($\gamma = 2$). For α values less than *ca.* 0.40, the agreement between this prediction and simulation is good. The theory, however, does overestimate values of s/n_c at higher α values, for unknown reasons.

Fig. 6e and f are plots of d/n_c *vs.* α for circular and elliptical zones, respectively. Here, we observe trends similar to those found in previous figures. More specifically, Roach's theory correctly predicts values of d/n_c for overlapping circular zones over the full range of α values (see Fig. 6e), although the agreement is best for small α values. Roach's theory somewhat underestimates the numbers of doublets formed from overlapping elliptical zones at large α values (see Fig. 6f), but unlike in Fig. 6b and d, this underestimation does not appear to be statistically significant. At least a part of this insignificance may be attributed to the larger standard deviations of d/n_c, which result from the small numbers of doublets found at the larger α values. The theory of Davis agrees well with the simulation results for α values less than *ca.* 0.20.

Fig. 7 is a plot of the dimensionless functions t/n_c, q/n_c and $[p - (s + d + t + q)]/n_c$, *vs.* α ($\beta = 1$). Here, variable q represents the number of quartet spots, and $p - (s + d + t + q)$ represents the algebraic sum of quintets, hextets and all other higher order multiplets. The solid curves in Fig. 7a and b are graphs of eqn. 11, divided by n_c, with $n = 3$, whereas the dashed curves in these figures are the algebraic sums of eqns. 1c and d, divided by n_c. The solid curves in Fig. 7c and d are graphs of eqn. 11, divided by n_c, with $n = 4$. The solid curves in Fig. 7e and f are graphs of eqn. 14, less the algebraic sum of four different expressions of eqn. 11, all of which are divided by n_c. These four expressions are defined by the n values, 1, 2, 3 and 4. The substitution $\alpha = \bar{m}/n_c$ was used as before to scale the theoretical predictions and simulations to fit in the same graph. The numbers m of zones corresponding to the various symbols are identified in the figure caption. The open diamonds in Fig. 7a, c and e represent the single determinations of Roach. Fewer data are plotted in Fig. 7 than in Fig. 6 because spot numbers less than 15 again were excluded and this exclusion restricts the data to large m values only ($m \geqslant 300$).

The values of t/n_c plotted in Fig. 7a and b show that Roach's theory accurately predicts the number of triplets formed from either circular or elliptical ($\gamma = 2$) zones over the examined α range. The standard deviations are somewhat larger than those in Fig. 6a–d because only a few triplets were found, even for large m. Davis's theory adequately predicts t/n_c for α values less than *ca.* 0.20. Roach's theory also adequately describes q/n_c for both zone shapes, although the scatter in the data is so large that a statistical agreement between simulation and theory is hardly surprising. This scatter results from the limited number of quartets observed. Interestingly, the results of Roach's simulations are scattered about his theory for quartets (see Fig. 7c), again because so few quartets were observed in his study. This scatter reflects the lack of statistical robustness in his simulations.

Because so few multiplets for which $n \geqslant 5$ were observed, we chose to plot their algebraic sum in Fig. 7e and f. The ordinates of these plots, therefore, are the sums of the numbers of quintets, hextets, heptets, octets, *etc.*, divided by n_c. No standard deviations are reported here, because they would be meaningless; these deviations would reflect not only variations about the mean of a given multiplet but also variations among the means of different multiplets. The close agreement between simulation and theory further confirms the theory of Roach.

We now turn our attention to evaluating the utility of the regression procedure proposed in the third part of the Theory section. Table II reports the statistical component numbers $\bar{m}$ estimated by fitting the data derived from 180 simulations of 2D square beds to eqn. 12. These estimates of $\bar{m}$ are accurate to within *ca.* 10% as long as $\alpha \leqslant 0.15$ ($\beta = 1$). In particular, for α values less than or equal to 0.10, the $\bar{m}$ estimates differ from the estimates

TABLE II

MEAN AND STANDARD DEVIATION OF NUMBERS $\bar{m}$ OF COMPONENTS ESTIMATED FROM TEN COMPUTER SIMULATIONS OF SQUARE BEDS CONTAINING m = 100, 200 AND 300 ZONES

Numbers p of spots were counted in accordance with the criterion $\beta = 1$. Results in parentheses in the last column are percentage errors in $\bar{m}$, as calculated with respect to the $\bar{m}$ predictions at $\alpha = 0$.

m	α ($\beta = 1$)	p	$\bar{m}$ estimates
100	0.00	100.0 ± 0.0	100.1 ± 5.0
	0.05	91.0 ± 3.3	100.8 ± 5.6 (0.73)
	0.10	82.8 ± 3.7	98.0 ± 7.0 (−2.1)
	0.15	74.3 ± 3.1	90.6 ± 7.5 (−9.6)
	0.20	67.1 ± 4.3	83.0 ± 5.9 (−17.1)
	0.25	59.4 ± 3.5	72.0 ± 3.9 (−28.1)
200	0.00	200.0 ± 0.0	195.6 ± 8.6
	0.05	182.1 ± 5.0	196.7 ± 7.6 (0.58)
	0.10	164.7 ± 2.6	192.9 ± 6.8 (−1.4)
	0.15	148.2 ± 4.6	179.1 ± 10.6 (−8.4)
	0.20	134.1 ± 5.4	162.7 ± 10.4 (−16.8)
	0.25	118.2 ± 3.3	137.2 ± 6.5 (−29.9)
300	0.00	300.0 ± 0.0	294.4 ± 10.7
	0.05	271.8 ± 4.8	295.8 ± 10.4 (0.49)
	0.10	245.3 ± 5.7	288.3 ± 13.3 (−2.1)
	0.15	219.0 ± 7.0	267.5 ± 14.7 (−9.1)
	0.20	198.9 ± 7.4	242.4 ± 8.8 (−17.7)
	0.25	179.8 ± 6.9	211.9 ± 8.9 (−28.0)

determined at $\alpha = 0$ by only 1–2%. These are remarkably accurate estimates, given that only *ca.* 80% of the zones can be detected under these circumstances (eqn. 12 predicts that $p/\bar{m} = 0.813$ at $\alpha = 0.10$). The accuracy of these $\bar{m}$s confirms the validity of the regression procedure, although its application is probably not optimized (see below). At larger α values, the estimates of $\bar{m}$ are too small, and their accuracy decreases substantially with increasing α. By and large, the percentage errors in $\bar{m}$ are independent of m and depend only on α. For example, at $\alpha = 0.20$, the percentage errors in $\bar{m}$ are −17.1, −16.8 and −17.7 for m = 100, 200 and 300, respectively. A similar trend was found when the "single-chromatogram method" was applied to 1D chromatograms [13].

Fig. 8 is a series of superimposed plots of p *vs.* n_c^{-1} generated from six 2D beds containing m = 100 zones. The zone coordinates in these beds were identical; only zone diameter d_0 (and hence α) was varied. The symbols corresponding to various α values are identified in the figure caption. If one examines the data for which p is independent of n_c^{-1} (*i.e.*, the data on the extreme left-hand side of the plot), one sees that the position of the first usable datum in the regression is increasingly shifted to larger values of n_c^{-1} as α increases. If one now examines the data which are fitted to theory (*i.e.*, the data on the right-hand side of the plot), one sees that values of p at any n_c^{-1} are nearly identical when α is small but that these values increase with increasing α. For example, the values of p at the n_c^{-1} coordinate represented in the figure by an arrow, $n_c^{-1} = 0.00565$, are 31, 35 and 42 for α values of 0.05, 0.15 and 0.25, respectively. This increase in p with increasing α is similar to that observed in the computer simulations by which the "single-chromatogram method" was tested [13]. These increases lead to the loss of curvature in the data representing function p (*i.e.*, eqn. 12) at large αs; indeed, the variation of p with n_c^{-1} is almost linear for $\alpha = 0.20$ and 0.25. This linear variation is a warning sign that the estimates of $\bar{m}$ will be unreliable.

This warning is insufficient, however; criteria other than the qualitative shape of the p *vs.* n_c^{-1} plot are necessary to evaluate the accuracy of the estimated $\bar{m}$ values. Perhaps the most important criterion is the value of α, which can be calculated by numerical means (*e.g.*, bisection method) from the estimated $\bar{m}$, the number p of spots in the bed and eqn. 12. For example, the numerical solution to this equation for the results of one simulation, p = 92

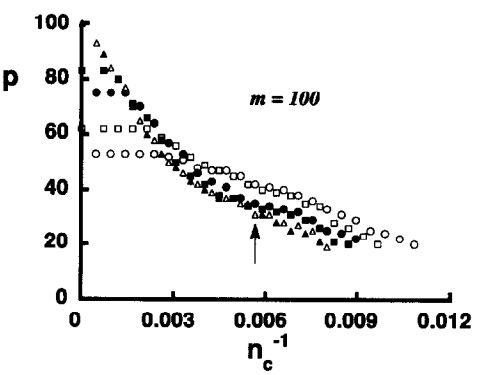


Fig. 8. Graphs of p *vs.* n_c^{-1} for m = 100 at six αs. Values of p at n_c^{-1} value represented by the arrow are discussed in the text. Symbols corresponding to different α values are ▲ = 0; △ = 0.05; ■ = 0.10; ● = 0.15; □ = 0.20; ○ = 0.25.

and $\bar{m} = 100.7$, is $\alpha = 0.044$, which agrees closely with the α value, 0.05, from which the simulation was generated. Because this estimate of α is less than 0.15, one has some confidence in the $\bar{m}$ estimate, which in fact is in error by only 0.7%. This criterion by itself is insufficient, however, because the estimates of $\bar{m}$ tend to decrease with increasing α (see Table II), leading to deceptive underestimations of α itself. Hence, one must also compare the overall appearance of the spot density in the bed to that in computer simulations, such as those reported in Fig. 4. Similar criteria are employed in evaluating the accuracy of $\bar{m}$ estimates determined by applying the "single-chromatogram method" to 1D chromatograms [13,22].

CONCLUSIONS

The theories of Roach and Davis are superior to the theories of Bourdillon *et al.*, Irwin *et al.* and Armitage in describing overlap in 2D beds. Of these two, the theory of Roach more correctly predicts the number of doublet and triplet spots over a wider range of α values than does the theory of Davis, which is correct only for α values less than *ca.* 0.20 ($\beta = 1$). Roach's theory also correctly predicts the number of higher order multiplets. When zones are elliptical, however, both Roach's and Davis's theories exhibit shortcomings in correctly predicting the total number of spots and the number of singlet spots at α values greater than *ca.* 0.25 ($\beta = 1; \gamma = 2$). This shortcoming is of practical concern, as zones in 2D separations are more elliptical than circular. In general, however, Roach's theory should be the basis of future studies of statistical limitations on 2D separations of modest saturation (*e.g.*, $\alpha < 0.25$), at least until an adequate theory for the overlap of elliptical zones is developed. Along these lines, Davis's theory for elliptical singlets shows some promise.

The regression procedure proposed here has been validated by analyses of dozens of computer simulations of 2D beds. The estimates of $\bar{m}$ so calculated are correct to within *ca.* 10% as long as α is less than *ca.* 0.15 ($\beta = 1$). This study is preliminary, however, and merely validates the proposed procedure; it does not determine the limits on it. Indeed, the findings presented here are based on the interpretation of somewhat hypothetical 2D beds, in that the representative coordinates of multiplet spots as determined here do not correspond to any physically measurable positions. By considering more realistic means for counting spots, it should be possible to extend the α range over which this procedure applies, as is now argued.

Based on the demonstrated success of the "single-chromatogram method" as a means for estimating $\bar{m}$ from the relative positions of maxima in single chromatograms [19,20,22], one anticipates that several advantages may accrue by identifying the maxima of measurable concentration pulses in a 2D bed with the representative coordinates of spots. These coordinates physically could be measured by a variety of means, *e.g.*, densitometry or fluorimetry, and experimental data could be gathered easily for purposes of regression. Because an isolated multiplet could contain several such maxima, the identification of spot coordinates with maxima would correspond to a counting procedure in which $\beta < 1$. This counting procedure could have distinct advantages. For example, if two zones slightly overlapped, it is possible that both zone centers would be observable as maxima. The distances between these centers and other similarly determined centers in the 2D bed would more accurately represent the underlying zone structure than do the distances calculated as detailed in the Procedures section. Hence it is possible (moreover, it is probable) that the range of α over which the regression is applicable can be extended by using such a counting procedure.

The reason why we did not interpret maxima as spot centers here is simple. In essence, we were not willing to delay communication of Roach's theory, particularly in the light of the increased interest in the statistical limitations on 2D separations [23], while the necessary but intricate details of this counting procedure were worked out. First, modestly sophisticated software will be required to determine from computer simulations the positions of maxima without the simultaneous detection of false maxima (we found several false maxima in our preliminary efforts in this direction). Further, we do not know yet the appropriate value to assign to β for use with such a counting procedure. This value will have to be determined empirically by counting the numbers of maxima in a series of simulations and adjusting β until simulation and theory agree, as was

done to determine the empirical resolution factor, $R_s^* \approx 0.5$, for the SMO [10–12]. These activities are best deferred to another study.

REFERENCES

1 J. M. Davis, *Anal. Chem.*, 63 (1991) 2141.
2 J. C. Giddings, *Unified Separation Science*, Wiley-Interscience, New York, 1991.
3 M. M. Bushey and J. W. Jorgenson, *Anal. Chem.*, 62 (1990) 978.
4 Z. Liu and J. Phillips, *J. Chromatogr. Sci.*, 29 (1991) 227.
5 D. Rosenthal, *Anal. Chem.*, 54 (1982) 63.
6 L. J. Nagels, W. L. Creten and P. M. Vanpeperstraete, *Anal. Chem.*, 55 (1983) 216.
7 J. M. Davis and J. C. Giddings, *Anal. Chem.*, 55 (1983) 418.
8 M. Martin, D. P. Herman and G. Guiochon, *Anal. Chem.*, 58 (1986) 2200.
9 A. Felinger, L. Pasti and F. Dondi, *Anal. Chem.*, 62 (1990) 1846.
10 J. C. Giddings, J. M. Davis and M. R. Schure, in S. Ahuja (Editor), *Ultrahigh Resolution Chromatography (ACS Symposium Series*, No. 250) American Chemical Society, Washington, DC, 1984, p. 9.
11 J. M. Davis and J. C. Giddings, *J. Chromatogr.*, 289 (1984) 277.
12 D. P. Herman, M. F. Gonnard and G. Guiochon, *Anal. Chem.*, 56 (1984) 995.
13 J. M. Davis and J. C. Giddings, *Anal. Chem.*, 57 (1985) 2168.
14 J. M. Davis and J. C. Giddings, *Anal. Chem.*, 57 (1985) 2178.
15 M. Martin and G. Guiochon, *Anal. Chem.*, 57 (1985) 289.
16 F. Dondi, Y. D. Kahie, G. Lodi, M. Remelli, P. Reschiglian and C. Bighi, *Anal. Chim. Acta*, 191 (1986) 261.
17 W. L. Creten and L. J. Nagels, *Anal. Chem.*, 59 (1987) 822.
18 S. Coppi, A. Betti and F. Dondi, *Anal. Chim. Acta*, 212 (1988) 165.
19 J. M. Davis, *J. Chromatogr.*, 449 (1988) 41.
20 S. L. Delinger and J. M. Davis, *Anal. Chem.*, 62 (1990) 436.
21 M. R. Schure, *J. Chromatogr.*, 550 (1991) 51.
22 F. J. Oros and J. M. Davis, *J. Chromatogr.*, 550 (1991) 135.
23 M. Martin, presented at the *18th International Symposium on Chromatography, Amsterdam, September 23–28, 1990.*
24 S. A. Roach, *The Theory of Random Clumping*, Methuen, London, 1968.
25 R. B. Bourdillon, O. M. Lidwell and J. E. Lovelock, *Studies in Air Hygiene (Medical Research Council Special Report Series*, No. 262), H.M. Stationery Office, London, 1948.
26 J. O. Irwin, P. Armitage and C. N. Davies, *Nature (London)*, 61 (1949) 809.
27 P. Armitage, *Biometrika*, 36 (1949) 257.
28 H. C. Tuckwell, *Elementary Applications of Probability Theory*, Chapman & Hall, London, 1988.
29 C. Mack, *Proc. Cambridge Philos. Soc.*, 50 (1954) 581.
30 C. Mack, *Proc. Cambridge Philos. Soc.*, 52 (1956) 246.
31 L. Nyhoff and S. Leestma, *FORTRAN 77 for Engineers and Scientists*, Macmillan, New York, 1985.
32 J. M. Davis, *Ph.D. Thesis*, University of Utah, Salt Lake City, UT, 1985.