ELSEVIER

Journal of Chromatography A, 933 (2001) 1–11

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Metrics of separation in chromatography

Leonid M. Blumberg[a,*], Matthew S. Klee[b]

[a]*Fast GC Consulting, P.O. Box 585, Hockessin, DE 19707, USA*

[b]*Agilent Technologies, 2850 Centerville Road, Wilmington, DE 19808, USA*

Received 10 July 2001; received in revised form 30 August 2001; accepted 30 August 2001

Abstract

A new metric, separation measure, *S*, for chromatographic separation is proposed. Unlike other metrics such as resolution, separation number, and some versions of peak capacity, the new metric provides a consistent, additive measure of the separation of pairs of peaks as well as the separation capacities of arbitrary intervals within the analysis time. The attribute of additivity means that the separation measure of any separation interval is equal to the sum of the separation measures of its subintervals. Practical aspects of the measurement of *S* are also addressed. In addition to definition of *S*, a definition of peak capacity, *n*, that is consistent with *S*, and includes useful features of other known definitions of *n* is proposed for an arbitrary time interval.

Keywords: Separation measure; Separation capacity; Separation rate; Resolution; Peak capacity; Separation number

1. Introduction

A search for a consistent approach to the optimization of temperature programmed GC (gas chromatography) provided the main stimulus for this study. However, the metrics that we propose, were constructed to apply as well to other separation techniques such as LC (liquid chromatography) etc.

Many metrics of separation [1–19] can be used for evaluation of the results of separations in chromatography. The most widely used are resolution [1,4], R_s, separation number or Trennzahl [2–4], SN, and peak capacity [5–12], *n*. These metrics played an important role in the development of chromatography and its applications.

In studies that attempt to compare the separation-speed performance of different chromatographic techniques (GC, LC, etc.) or different modes of the same technique (isothermal GC, temperature programmed GC, etc.), a system of scalable metrics might be required. For example, it might be desirable to gradually expand the scope of a given metric from a local focus on a pair of two close peaks to a regional focus on a performance during a certain temperature plateau or its fraction. Similar need might arise in regard to a particular temperature ramp or its fraction. There might also be a need to further expand the scope of the same metric to cover the entire analysis.

Unfortunately, currently known metrics of separation, including R_s, SN, and *n*, do not constitute a compatible system and do not provide a flexibility of a scalable transition from one to another. Each of these metrics was implicitly or explicitly designed to address a certain narrow aspect of separation. Thus, R_s, intended as a measure of separation of two neighboring peaks, can be viewed as a local metric

*Corresponding author. Fax: +1-302-239-3941.

E-mail address: leon@fastgc.com (L.M. Blumberg).

0021-9673/01/$ – see front matter
PII: S0021-9673(01)01256-0

of separation. SN, that ''gives the number of well-separated peaks within any homolog pair'' [3], can be viewed as a regional metric of separation. Finally, *n*, designated to approximate ''the maximum number of peaks to be separated on a given column'' [5], can be viewed as a global metric of separation. The differences between these metrics, although rooted in the history of evolution of a separation science, are not altogether current with existing theoretical developments, and, therefore, pose some limits on advancing practice of chromatographic arts.

The construction of a single metric of separation that incorporates the useful features of R_s, SN and *n* while, unlike these three, being meaningful and useful over a wide range of techniques and conditions is the main objective of this report.

To justify the specific differences between the newly proposed metric and the existing ones, we examine the shortcomings of R_s, SN and *n*. To highlight the shortcomings, we use R_s as an example where appropriate.

The widely accepted definition of resolution, R_s, of two peaks was introduced in 1958 by a scientific committee [1] and recently re-confirmed by IUPAC [4]. It describes R_s as:

$$R_s = (t_{R2} - t_{R1})/\bar{w}_b, \quad \bar{w}_b = (w_{b1} + w_{b2})/2 \qquad (1)$$

where t_R is retention time of a peak measured as the time ''between sample injection and the emergence of the peak maximum'' [4], w_b is base width of a peak measured as the length of the ''segment of the peak base intercepted by the tangents drawn to the inflection points of either side of the peak'' [4], and $\bar{w}_b$ is average base width of the peaks. For the Gaussian peaks $w_b = 4\sigma$ where σ is the standard deviation [20,21] of the peak. Eq. (1) becomes:

$$R_s = (t_{R2} - t_{R1})/(4\bar{\sigma}), \quad \bar{\sigma} = (\sigma_1 + \sigma_2)/2 \qquad (2)$$

Some workers used this formula rather than Eq. (1) as a definition of resolution [7–9,22–24]. However, in the following analysis, we will always assume the recommended and most widely used definition of resolution described in Eq. (1).

Following is a brief review of some shortcomings of R_s, SN and *n*.

1.1. Inconsistent peak width metrics

The most obvious (but not the only) cause of incompatibility of R_s, SN and *n* with each other comes from the fact that they are based on different peak width metrics: w_b, for R_s, $2w_h$ for SN, and 4σ for *n* (w_h is half-height width [4] of a peak).

1.2. Unpredictable peak parameters (see Section 5 for more details)

To predict resolution of two peaks, one needs to predict the time coordinate of the peak apex. Unfortunately, it is not known how to theoretically predict the time coordinate of the apex of an asymmetric peak. It is unclear, for example, how to predict the time coordinates of the peak apexes in a computer simulated chromatogram, Fig. 1a, consisting of asymmetric EMG (exponentially-modified Gaussian) [25–27] peaks. A similar problem exists with the prediction of base widths, w_b, of peaks. A definition of SN [2–4] (based on the averaging of the half-height widths of the peaks) and some definitions [11] of *n* have similar shortcomings.

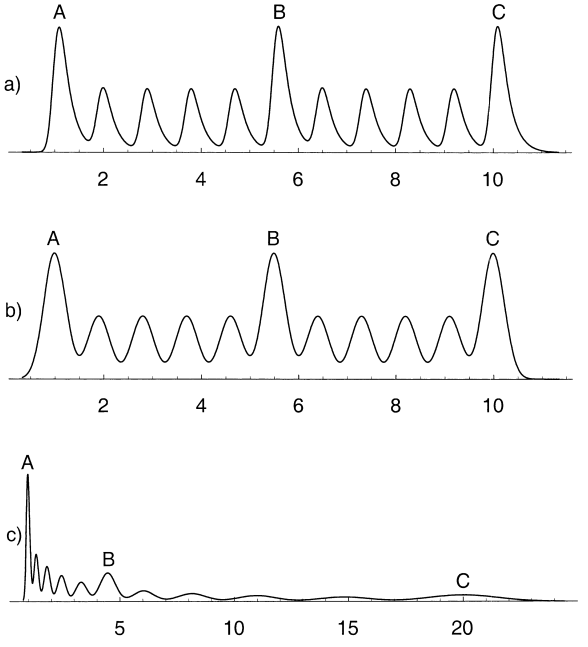

Fig. 1. Three computer generated 11-peak (EMG in (a), Gaussian in both (b) and (c)) chromatograms. The larger peaks A, B an C in all chromatograms are designated as consecutive ''homolog'' markers. In (a) and (b), all peaks have the same standard deviations, σ, and are 4σ apart from their neighbors. The standard deviations of peaks in (c) are proportional to time, as shown in Table 1.

1.3. Lack of conceptual meaning

Although the problems with the prediction of peak parameters in Eq. (1) can be solved by using Eq. (2) instead of Eq. (1), there are other potentially more problematic structural shortcomings in both expressions.

Consider a computer generated chromatogram, Fig. 1c, consisting of equally resolved Gaussian peaks having the widths that (like in isothermal GC and in isocratic LC) increase in proportion with time, t. Using Table 1 and Eq. (2) (or Eq. (1) where $w_b = 4\sigma$), one can find that, for each pair of neighboring peaks in Fig. 1c, $R_s = 1$. Assuming that $R_s = 1$ is a nominal resolution for the neighboring peaks, it is logical to expect that quantity $R_{sAC} - 1$, where R_{sAC} is the resolution of the peaks A and C, should represent the number of the nominally resolved peaks that can be placed between the peaks A and C. There are nine nominally resolved peaks between the peaks A and C in Fig. 1c. One might expect, therefore, that $R_{sAC} = 10$. However, using the data in Table 1 again, one finds that Eqs. (1) and (2) yield $R_{sAC} \approx 6.1$ which is nearly 40% below the value that one might expect from a reasonable straightforward interpretation of a concept of R_s. It is unclear, therefore, what does the value of the resolution generally mean or what kind of a concept the resolution represents. What is clear, on the other hand, is the following. Supposes that there are other peaks between the arbitrary peaks, say, A and C. The sum of the resolutions of all consecutive peak pairs starting with A and ending with C can be substantially different from the resolution of peaks A and C. Similar problem exists with SN which, in spite of its intended designation to provide ''the number of well-separated peaks within any homolog pair'' [3], can be substantially smaller than that number.

1.4. Lack of additivity

The fact that the resolution of the peaks A and C is different from the sum of the resolutions of all consecutive peak pairs starting with A and ending with C means that the resolution is not an additive quantity. Similar problem exists with SN. One can verify this fact by examining SN for the peak pairs (A, B) and (B, C) in Fig. 1c assuming that A, B and C are consecutive homolog markers.

A lack of the property of additivity in R_s and SN is a source of many theoretical problems in using these metrics. Thus, the fact that R_s and SN can substantially underestimate the actual number of peaks that a column can potentially separate within a predetermined separation region creates a substantial problem for separation-speed trade-off metrics based on these quantities. Indeed, unneeded separation power of a column can be traded for reduction in the analysis time. For example, a translatable [28,29] x-fold reduction in R_s or in SN + 1 can be traded for a x^3-fold or greater reduction in the analysis time [30–32]. This means that a 20% underestimation in the value of R_s or SN + 1 can result in missing the opportunity of a two-fold reduction in the analysis time.

The problem with underestimating the values of R_s compared to the respective actual peak count comes from peak width averaging (using $\bar{w}_b$ in Eq. (1) or $\bar{\sigma}$ in Eq. (2)) in measuring R_s for unequally wide peaks. This problem is the most pronounced in cases of columns with low plate numbers, i.e. exactly the columns that are important for reduction of analysis time.

1.5. Inconsistent definitions of peak capacity

Unlike R_s and SN, peak capacity, n, introduced by Giddings [5], is based on counting nominally separated peaks rather than on peak width averaging. This means that n is an additive quantity. Unfortunately, in introducing n, Giddings only outlined it as a general concept without reducing it to a particular formula. As a result, several different and not always compatible formulae for the calculation

Table 1
Retention times, t_R, and standard deviations, σ, (both in relative units) for the 11-peak chromatogram in Fig. 1c

	A					B					C
t_R	1.000	1.3493	1.8206	2.4565	3.3145	4.4721	6.0342	8.1418	10.9856	14.8227	20.0000
σ	0.0743	0.1003	0.1353	0.1826	0.2464	0.3325	0.4486	0.6052	0.8166	1.1019	1.4868

of n can be found in the literature [5–12]. Giddings himself used several incompatible versions of n [5,7–9].

1.6. A prototype for the new metric

For unequally wide peaks, the Giddings' idea of counting nominally separated peaks is best supported by the Lan and Jorgenson definition [11]:

$$n = \int_{t_1}^{t_2} \mathrm{d}t / w_\mathrm{b} \tag{3}$$

where t_1 and t_2 are boundaries of the entire separation space [11] in a chromatogram. This definition is structurally compatible with the definitions of n, used by Davis and Giddings in their study of statistics of peak overlap [7–9] although, unfortunately, the definitions of n used in these studies do not allow for a variation of a peak width with time. On the other hand, w_b in Eq. (3) is not a satisfactory peak width metric, as we described earlier.

For Gaussian peaks, Eq. (3) becomes:

$$n = \frac{1}{4} \int_{t_1}^{t_2} \mathrm{d}t / \sigma \tag{4}$$

This formula can be viewed as a prototype of a separation measure proposed herein.

1.7. Modifications

Our definition of the separation measure, S, is based on the following modifications to Eq. (4).

First, we allowed t_1 and t_2 to be the boundaries of an arbitrary interval within the analysis time, not only of the entire space potentially available for the peaks as in Eq. (4).

Second, we dropped the requirement that the peaks must be Gaussian.

Third, we recognized that once the standard deviation, σ, was adopted as the basic peak width measure, the quotient $1/4$ in Eq. (4) added no value and had no objective meaning. Dropping the quotient simplified the related formulae and allowed for more transparent interpretations.

As a result of these modifications, we came up with the metric, S, that, for an arbitrary interval, (t_1, t_2), always meant the same simple thing — a number of σ-wide subintervals between t_1 and t_2. This does not depend on the size of the interval (t_1, t_2), and whether the σ remains constant or varies along the interval.

In addition to the definition of separation measure, S, we also proposed a refined definition of peak capacity, n. An important component to the concept of peak capacity is its ability to accommodate different minimum acceptable separation of the neighboring peaks [7–9]. For example, if a column has $n=100$ when 6σ-separation between the neighboring peaks is required then the same column has $n=200$ if only 3σ-separation is acceptable. Many definitions [6,10–12] of n do not allow for the varying requirements for the minimum acceptable separation to be taken into account in the calculation of n. This can lead to substantial logical difficulties. Following the approach used in Davis and Giddings studies [7–9], we retained the feature in our definition of n.

2. Theory

2.1. Retention time and width of a peak

A separation is essentially a stochastic process. As a result, statistical moments [20,21] of the peaks provide the most predictable [9,22,25] peak parameters (see Section 5 for more details).

In this study, we will always assume that retention time, or elution time, t_R, of a peak is measured as the peak's first statistical moment. As a measure of the width of a peak, we will always use the peak's standard deviation (square root of the peak's second central moment).

2.2. Separation measure

We define a separation measure, S, of an arbitrary time interval (t_a, t_b) as the number of consecutive σ-wide time intervals (or, briefly, σ-intervals) within (t_a, t_b). Obviously, if σ remains the same for all retention times within (t_a, t_b) then:

$$S = \frac{\Delta t}{\sigma}, \quad \Delta t = t_\mathrm{b} - t_\mathrm{a} \tag{5}$$

Generally, however, σ can change as a function $\sigma=\sigma(t)$ of time, t. In that case, S for (t_a, t_b) can be found as a sum of infinitesimally small increments $dS = dt/\sigma$ representing separation measures of all non-overlapping infinitesimally small dt-long consecutive subintervals of (t_a, t_b). In other words, generally, separation measure, S, of an arbitrary separation interval (t_a, t_b) can be found as:

$$S = \int_{t_a}^{t_b} dS = \int_{t_a}^{t_b} dr/\sigma \quad (6)$$

Notice that the inverse, $1/\sigma$, of the standard deviation, σ, in the above expression represents a number of σ-intervals that the analysis can produce per unit of time. Hence $1/\sigma$ can be interpreted as a separation rate. For example, $\sigma=0.1$ s implies that the separation rate is $1/\sigma=10/\text{s}$, i.e. 10 σ-intervals per second.

It follows from Eq. (6) that S has the following properties.

1. It describes the separation of two peaks in units of standard deviation, σ, i.e. in the same way as the separation of two stochastic events is typically described in other scientific and technical fields.
2. It can be used with an arbitrary peak shape.
3. It is an additive quantity, i.e., the separation measure of any separation interval is equal to the sum of the separation measures of all of its non-overlapping subintervals

Indeed, it follows from Eq. (6) that, for any sequence $t_a, t_b, t_c, \ldots, t_z$ where $t_a \le t_b \le t_c \le \cdots \le t_z$, a separation measure, S, of (t_a, t_z) can be found as $S=S_a+S_b+\cdots$ where S_a is the separation measure of (t_a, t_b), S_b is the separation measure of (t_b, t_c), etc. From a broader perspective, the integral $\int_{t_0}^{t} dt/\sigma$, where t_o is void time, and t is an arbitrary time larger than t_o, can be viewed as a transformation of a time domain into a separation domain that has a uniform (the same in all locations along the separation axis) metric. From that point of view, Eq. (6) describes S as a length of an interval on the separation axis, and, therefore, as an additive metric. (A length of an interval is equal to the sum of the lengths of its consecutive non-overlapping subintervals).

Depending on its application, separation measure in Eq. (6) can have several uses. It can be used as a measure of the separation of two peaks if t_a and t_b are their retention times, and as a measure of the separation capacity of the interval (t_a, t_b) irrespective of the presence of the peaks at t_a and t_b.

2.3. Special cases

Eq. (6) for the separation measure, S, might be inconvenient for a practical use. Here we point to several factors that can simplify the evaluation of S in an everyday practice.

2.3.1. Fixed peak width

In some cases (such as temperature programmed GC, gradient elution LC, etc.), all peaks eluting within a wide time span can have nearly the same width, Fig. 1a,b. For any interval, (t_a, t_b), where σ is a fixed quantity, Eq. (6) yields Eq. (5).

For the comparison of S with the resolution, R_s, let's notice that, according to Eq. (2):

$$R_s = \Delta t/(4\sigma) \quad (7)$$

if both peaks are equally wide and Gaussian. Comparison of this expression with Eq. (5) leads to the conclusion that:

$$R_s = S/4 \text{ for equally wide Gaussian peaks} \quad (8)$$

This means, for example, that $S=6$ (6σ-separation) is equivalent to $R_s=1.5$. Similarly, $R_s=1$ is equivalent to $S=4$ (4σ-separation). It is important to stress, however, that S is valid for any peak shape while it is not clear how to predict R_s for non-Gaussian peaks, such as EMG (exponentially-modified Gaussian) [25–27] peaks in Fig. 1a, even if all the peaks have the same shape and width.

2.3.2. Peak width is a linear function of time

There is another important case where (such as in isothermal GC and isocratic LC), σ can be a nearly linear function of t within a wide region (t_A, t_B). In its general form, a linear relation between t and σ within (t_A, t_B) can be expressed as:

$$\sigma = \sigma_A + \frac{\sigma_B - \sigma_A}{t_B - t_A}(t - t_A) \quad (9)$$

where σ_A and σ_B correspond, respectively, to the

peaks at t_A and t_B. This yields for any interval (t_a, t_b) within (t_A, t_B):

$$(\sigma_b - \sigma_a)/(t_b - t_a) = (\sigma_B - \sigma_A)/(t_B - t_A) \quad (10)$$

where σ_a and σ_b are standard deviations of the peaks at t_a and t_b, respectively. Integration of Eq. (6) with σ from Eq. (9), and accounting for Eq. (10) yields:

$$S = \frac{\Delta t}{\Delta \sigma} \ln \frac{\sigma_b}{\sigma_a} = \frac{\Delta t}{\Delta \sigma} \ln\left(1 + \frac{\Delta \sigma}{\sigma_a}\right) = \frac{\Delta t}{\sigma_a} \cdot \frac{\ln(1 + \delta_\sigma)}{\delta_\sigma} \quad (11)$$

where:

$$\Delta t = t_b - t_a, \quad \Delta \sigma = \sigma_b - \sigma_a, \quad \delta_\sigma = \Delta \sigma / \sigma_a. \quad (12)$$

In a more general case where σ is not a linear function of t, it can be expressed as a piece-wise linear one. Due to the additivity of S, it can be found as a sum of local or regional separation measures calculated from Eq. (11) for each segment with linear σ.

2.4. Approximations

Suppose that the widths of the peaks ''a'' and ''b'' are nearly the same ($\sigma_a \approx \sigma_b \approx \sigma$) and, hence, a relative difference, $|\delta_\sigma|$, Eq. (12), between σ_a and σ_b is much smaller than 1, i.e.:

$$|\delta_\sigma| = |(\sigma_b - \sigma_a)/\sigma_a| \ll 1. \quad (13)$$

In that case, $\ln(1 + \delta_\sigma) \approx \delta_\sigma = \Delta\sigma/\sigma_a \approx \Delta\sigma/\sigma$. Eq. (11) becomes $S \approx \Delta t/\sigma$ indicating that Eq. (5) closely approximates separation measure, S, of peaks with similar widths.

It is tempting, to extend the simplification provided by Eq. (5) for peaks with substantially different widths. However, a direct substitution of either σ_a or σ_b for σ in Eq. (5) leads to a substantial error. A much better result comes from the replacement of σ by the average:

$$\bar{\sigma} = (\sigma_b + \sigma_a)/2 \quad (14)$$

that yields an approximation:

$$S = \tilde{S} = \frac{\Delta t}{\bar{\sigma}} = \frac{\Delta t}{(\sigma_a + \sigma_b)/2} = \frac{2\Delta t}{(\sigma_a + \sigma_b)} \quad (15)$$

that can be viewed as a result of the replacement of quantity $(\sigma_b - \sigma_a)/\ln(\sigma_b/\sigma_a)$ in Eq. (11) with the average, $\bar{\sigma}$, of σ_a and σ_b. A relative error, δ_S, Fig. 2, of this approximation can be expressed as:

$$\delta_S = \frac{\tilde{S} - S}{S} = \frac{\delta_\sigma}{(1 + \delta_\sigma/2)\ln(1 + \delta_\sigma)} - 1 \quad (16)$$

where δ_σ is defined in Eq. (12). Fig. 2 shows that δ_S can be insignificant when $|\delta_\sigma| \le 1$ (i.e. $\sigma_b \le 2\sigma_a$). Thus, even when $\delta_\sigma = 1$ (i.e. $\sigma_2 = 2\sigma_1$), $|\delta_S| \approx 4\%$ meaning that, in practice, it is probably always safe to use Eq. (15) instead of Eqs. (11) as a separation measure of the neighboring peaks. Indeed, seldom do two peaks that are close to each other have substantially different widths. However, as Fig. 2 and Example 1 below show, Eq. (15) can lead to substantial error when it is used to measure wide separation intervals between the peaks with substantially different widths.

The following strategy can be used as a practical approach to determining the separation capacity, S_{AB}, of an arbitrary interval bound by two unequally wide peaks A and B.

1. If it is known a priori (as in isothermal GC and in isocratic LC) that σ is a linear function of t, then S_{AB} can be found from Eq. (11).

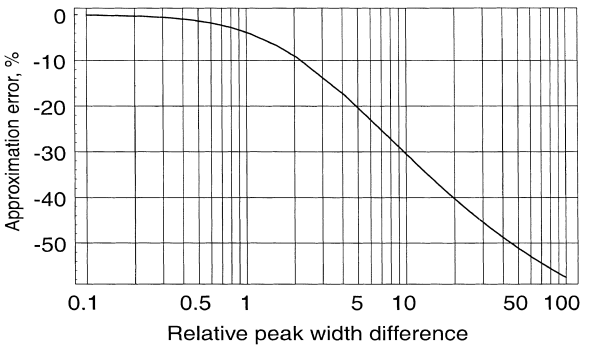

Fig. 2. Relative error, δ_S, Eq. (16), of approximation of Eq. (11) by Eq. (15) vs. relative difference, δ_σ, Eq. (12), in the widths of two peaks.

2. If the linearity of σ as a function of t is in doubt, then
   2.1. If σ_A and σ_B differ by less than a factor of 2, Eq. (15) can be used to calculate S_{AB}
   2.2. If σ_A and σ_B differ by substantially more than a factor of 2, several test components could be added to the test mixture in order to bridge the peak width gap. Step 2.1 can be used for each consecutive pair starting from A and ending at B. S_{AB} can then be found as the sum of all separation measures for the consecutive peak pairs.

Example 1. As follows from Table 1, standard deviations, σ, of the peaks in Fig. 1c, are proportional to time, t. This allows us to calculate S using Eq. (11). For the separation measures, S_o, of neighboring pairs, and the separation measures S_{AB}, S_{BC} and S_{AC}, of the separation intervals (A, B), (B, C) and (A, C), respectively, one has: $S_o = 4.03$, $S_{AB} = S_{BC} = 20.15$, and $S_{AC} = 40.3$. Suppose now that the dependence of σ on t in Fig. 1c is not known. Because, according to Table 1, $\sigma_B/\sigma_A = \sigma_C/\sigma_B \approx 4.5$ and $\sigma_C/\sigma_A = 20$, the use of Eq. (15) for calculation of $\tilde{S}_{AB}$, $\tilde{S}_{BC}$ and $\tilde{S}_{AC}$ is likely to result in substantial errors. Indeed, $\tilde{S}_{AB} = \tilde{S}_{BC} \approx 17$ and $\tilde{S}_{AC} \approx 24.3$, i.e. 15% and 40% below respective exact values. On the other hand, there are several peaks between the peaks A, B and C in Fig. 1c. The relative width increment, about 35%, of each peak compared to its immediate predecessor is not very large. Therefore, the error in calculation of $\tilde{S}_o$ using Eq. (15) should not be significant. One has: $\tilde{S}_o = 4.00$, i.e. less than 1% below exact value of $S_o = 4.03$. Due to additivity of separation measure, one can estimate $S_{AB} = S_{BC} \approx 5 \times \tilde{S}_o = 20.00$ and $S_{AC} = 10 \times \tilde{S}_o = 40.00$. The errors of these estimates are less than 1%.

Finally, as was the case for Eqs. (5) and (7), Eq. (15) is similar to Eq. (2) for Gaussian peaks. This allows us to write:

$$R_s \approx S/4 \text{ for Gaussian peaks} \quad (17)$$

However, while S is valid for any peak shape, there is no general solution for predicting R_s for non-Gaussian peaks.

2.5. Peak capacity

Let S_{min} be the lowest acceptable peak separation in a given analysis. In line with its general concept [5], we define a (benchmark) peak capacity, n, of an arbitrary separation interval (t_a, t_b) as a maximum number of S_{min}-wide non-overlapping intervals in (t_a, t_b), i.e.:

$$n = S/S_{min} \quad (18)$$

Due to Eq. (6), n can be found as:

$$n = \frac{1}{S_{min}} \int_{t_a}^{t_b} \frac{dt}{\sigma} \quad (19)$$

Similarly to the separation capacity, peak capacity in Eqs. (18) and (19) is also an additive quantity.

For the n peaks to be accommodated by the interval with the peak capacity n, they all have to be regularly separated, each S_{min} apart from its neighbors. Since it is unrealistic to expect such a regular peak packing in a real analysis, n in Eqs. (18) and (19) represents an overestimated number of peaks that can be realistically found within the interval (t_a, t_b). More realistic would be a statistical estimation of a peak capacity. Davis and Giddings have shown [7,9] that, under reasonable statistical assumptions, the number of peaks to be found within a given interval (some of them representing more than one component) can not typically exceed a statistical peak capacity, n_s, which is about 37% of n for that interval. More accurately,

$$n_s = n/e \approx n/2.72 \approx 0.368n \quad (20)$$

Statistical peak capacity seems to be a more realistic representative of the number of peaks that a given interval can contain. Unfortunately, many of these peaks can represent more than one component [7,9].

One should notice that, although peak capacity, n, provides a gross overestimation of a number of peaks that an interval can realistically contain, nevertheless, because n_s and other statistics [7,9] can be found from n, the latter is a useful benchmark.

3. Conclusion

We introduced separation measure, S — a metric of separation of peak pairs and separation capacities of arbitrary intervals within the analysis time. Quantity S has the following advantages over the widely accepted metrics of separation such as resolution, R_s, separation number, SN, and peak capacity, n.

1. Unlike R_s, SN and some versions of n, S is based on statistical moments of peaks — the most consistent and predictable peak parameters.
2. Unlike R_s, SN and some versions of n, S can be used with any shape of chromatographic peaks. It measures the separation of peaks in units of their standard deviations, σ, i.e. in the same manner as the separation of two stochastic events is measured in other scientific and technical fields. Unlike R_s and SN, the separation measure, S, of any interval (t_1, t_2) where peaks either actually exist or can potentially be found always represent the same simple thing — a number of all non-overlapping σ-wide subintervals within (t_1, t_2). This meaning of S is always the same whether or not the σ changes as a function of t. For Gaussian peaks that have roughly the same width, S is about four times larger than R_s.
3. Unlike R_s, SN and some versions of n, S is an additive metric i.e. the separation measure of any interval is always a sum of the separation measures of its all non-overlapping subintervals.
4. S is a universal scalable metric. Like R_s, S can have a local scope, i.e. it can be used to measure separation of neighboring peak pairs. Like SN, S can have a regional scope, i.e. it can be used to measure a separation potential of an arbitrary separation interval such as an interval between designated peak markers, an interval along a particular heating ramp or its fraction in a temperature programmed GC, etc. Finally, S can have a global scope, i.e. it can be used to measure a separation potential of a given column, entire analysis, etc.

Practical aspects of the evaluation of S in real chromatograms were also discussed.

In addition, we proposed a generalization of peak capacity that included useful properties of this concept known from several sources while making it applicable to an arbitrary interval within the analysis time.

4. Symbols

Symbol	Description	Measured in units of
n	Peak capacity, Eqs. (18), (19)	1
n_s	Statistical peak capacity, Eq. (20)	1
R_s	Resolution, Eqs. (1), (2)	1
S	Separation measure, Eq. (6)	1
$\tilde{S}$	Approximate separation measure, Eq. (15)	1
t	Time	time
δ_σ	Relative difference of two standard deviations, Eq. (12)	1
δ_S	Relative approximation error in separation measure, Eq. (16)	1
σ	Standard deviation of a peak	time
$\bar{\sigma}$	Average standard deviation of two peaks, Eq. (14)	time

5. Appendix: prediction of peak retention and width

This material is compiled from the 1966 paper published by Sternberg [25]. All essential elements of this compilation can also be founding in mathematical handbooks [21], as well as in the text books on linear systems, and on statistics.

A chromatographic system consists of several subsystems (a sample introduction subsystem, a column, a detector, etc.). In a linear (not overloaded, as typically is the case in analytical chromatography) chromatographic system, the contribution of each subsystem to the final properties of the peaks can be found from the impulse responses of the subsystems. Several types of impulse responses are shown in Fig. 3. A rectangular ''plug'' is sometimes used to describe a sample introduction process [25], and the effect of a TCD (thermal conductivity detector) chamber volume [33]. Ideally, a Gaussian peak represents the properties of a column. An exponential pulse can model the properties of detector electronics [25], and the effects of some types of a ''dead volume'' in a system. An EMG peak in Fig. 3 represents a class of the tailing peak shapes resulted

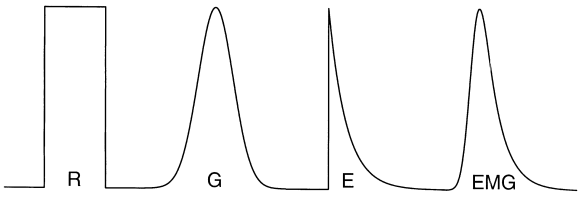

Fig. 3. Rectangular (R), Gaussian (G), Exponential (E), and EMG (exponentially-modified Gaussian) peaks.

from the combination (a convolution) of a Gaussian and exponential peaks [25–27]. A particular shape of an EMG peak depends on the relative content of the Gaussian and the exponential component.

Two parameters — the center of gravity, τ, (the first statistical moment), and the standard deviation, σ, (square root of the second central moment) — of a peak represent, respectively, its location along the abscissa, and its width.

5.1. Retention time

The center of gravity of a symmetric peak coincides with the abscissa, t_o, of its apex. This means that, whether the retention time, t_R, of a symmetric peak is defined as the center of its gravity (i.e. $t_R = \tau$), or as the time coordinate of its apex (i.e. $t_R = t_o$), it represents the same quantity. This is not the case for the exponential and the EMG peaks. Thus, for the former, $\tau = t_o + \sigma$.

Let τ_I, τ_C, and τ_D be the centers of gravity of the impulse responses of the sample introduction, the column, and the detector. Then the center of gravity, τ, of the final peak produced by the system, can be found as:

$$\tau = \tau_I + \tau_C + \tau_D \tag{21}$$

On the other hand, if at least one component of the system can have an arbitrary asymmetric impulse response then the abscissa of the apex of a peak produced by the system is generally unpredictable.

These observations provide a strong reason in favor of defining retention time of a peak as its center of gravity, and to rely on the centers of gravity of the impulse responses of the system components for the retention time predictions. For example, in a calibrated system with a priori known τ_I and τ_D in Eq. (21), one can reconstruct the actual chromatographic retention time, t_R, of a peak from the measured value, $t_{R,\,measured}$, as:

$$t_R = t_{R,\,measured} - (\tau_I + \tau_D) \tag{22}$$

5.2. Width

Let σ_I, σ, and σ_D be the standard deviations of the impulse responses of the sample introduction, the column, and the detector. Then the standard deviation, $\sigma_{measured}$, of the actual peak produced by the system, can be found as:

$$\sigma^2_{measured} = \sigma^2_I + \sigma^2 + \sigma^2_D \tag{23}$$

In a calibrated system with a priori known σ_I and σ_D in Eq. (22), the standard deviation, σ, of actual chromatographic peak can be reconstructed from the measured data as:

$$\sigma = \sqrt{\sigma^2_{measured} - (\sigma^2_I + \sigma^2_D)} \tag{24}$$

Other forms of describing the width of the peaks are widely used in chromatography. For example, the resolution [1,4] of two peaks, and some definitions [11] of the peak capacity utilize the base width, w_b, of the peaks. Other metrics, like the separation number [2–4], are based on the half-height width [4], w_h, of the peaks. Some data systems report the area-over-height widths, $w_a = a/h$, of the peaks where a and h are, respectively, the peak's area and the height. Unfortunately, neither of these metrics can be predicted from the parameters of the system components if the impulse response of at least one component can have an arbitrary shape.

For Gaussian peaks, quantities w_b, w_h and w_a can be found from σ using the relations [4,34]:

$$w_b = 4\sigma,\ w_h = \sqrt{8 \ln 2}\,\sigma \approx 2.355\sigma,$$
$$w_a = \sqrt{2\pi}\,\sigma \approx 2.507\sigma, \tag{25}$$

These relations can also be used for approximate evaluation of σ from experimentally measured w_h or

Table 2
Conversion factors, c, in the expressions $w = c\sigma$ for conversion of the standard deviation, σ, into the peak width metrics w_b, w_h and w_a

	w_b		w_h		w_a	
	$=$	$\approx$	$=$	$\approx$	$=$	$\approx$
Gaussian	4	4	$\sqrt{8\ \ln 2}$	2.355	$\sqrt{2\pi}$	2.507
Rectangular	$2\sqrt{3}$	3.464	$2\sqrt{3}$	3.464	$2\sqrt{3}$	3.464
Exponential	1	1	ln2	0.693	1	1

Column headings ''$=$'' and ''$\approx$'' indicate the exact and the approximate values, respectively.

w_a if peaks are known to be Gaussian or nearly Gaussian.

Unfortunately, for non-Gaussian peaks, the relations between σ and other peak width parameters can be substantially different from Eq. (25) as shown in Table 2. One can notice, for example, that, for the same σ, the half-height width, w_h, of a rectangular peak (3.464σ) is nearly five times larger than w_h for exponential peak (0.693σ). It is also worth noticing that, for the exponential peak $w_h < \sigma$ while, for the Gaussian peak, w_h is more than two times larger than σ. These observations demonstrate that, for an arbitrary peak, w_h can not be consistently predicted from σ. And when it comes to the base width, w_b, even the concept itself has serious problems. According to its definition, w_b is the length of the ''segment of the peak base intercepted by the tangents drawn to the inflection points of either side of the peak'' [4]. However, the very existence of the inflection points in the peak is not always guaranteed. For example, neither the rectangular nor the exponential peak has the inflection points. To generate the data for Table 2, we identified w_b of a rectangular peak with what intuitively seems to be appropriate measure of this quantity (which, in this case, is the same as w_h and w_a). In the case of an exponential peak, we identified its base width, $w_{b,e}$, with the length of the segment of the peak base intercepted by the line of the vertical rise of the peak on its left side, and by the tangent to the exponential line drawn at the peak apex. (This treatment can be justified by the fact that the properly identifiable base width of EMG peak converges to $w_{b,e}$ when σ of the Gaussian component in the EMG peak approaches zero).

All these observations provide a strong reason in favor of using the standard deviation, σ, as the peak width metric.

References

[1] A.J.P. Martin, D. Ambrose, W.W. Brandt, A.I.M. Keulemans, R. Kieselbach, C.S.G. Phillips, F.H. Stross, in: D.H. Desty (Ed.), Presented at Gas Chromatography 1958, Royal Tropical Institute, Amsterdam, 19–23 May, 1958, Academic Press, New York, 1958, p. xi.
[2] R.E. Kaiser, Z. Anal. Chem. 189 (1962) 1.
[3] R.E. Kaiser, R. Rieder, Chromatographia 8 (1975) 491.
[4] L.S. Ettre, Pure Appl. Chem. 65 (1993) 819.
[5] J.C. Giddings, Anal. Chem. 39 (1967) 1027.
[6] E. Grushka, Anal. Chem. 42 (1970) 1142.
[7] J.M. Davis, J.C. Giddings, Anal. Chem. 55 (1983) 418.
[8] J.M. Davis, J.C. Giddings, Anal. Chem. 57 (1985) 2168.
[9] J.C. Giddings, Unified Separation Science, Wiley, New York, 1991.
[10] Y. Shen, M.L. Lee, Anal. Chem. 70 (1998) 3853.
[11] K. Lan, J.W. Jorgenson, Anal. Chem. 71 (1999) 709.
[12] J.C. Medina, N. Wu, M.L. Lee, in: P. Sandra, A.J. Rackstraw (Eds.), Proceedings of 23rd International Symposium on Capillary Chromatography (CD ROM), Palazzo dei Congressi, Riva del Garda, Italy, 5–10 June, 2000, I.O.P.M.S, Kortrijk, Belgium, 2000.
[13] W.L. Jones, R. Kieselbach, Anal. Chem. 30 (1958) 1590.
[14] J.C. Giddings, Anal. Chem. 32 (1960) 1707.
[15] R.E. Kaiser, Gas Phase Chromatography, Gas Chromatography, Vol. 1, Butterworth, London, 1963.
[16] W.E. Harris, H.W. Habgood, Programmed Temperature Gas Chromatography, Wiley, New York, 1966.
[17] G. Guiochon, C.L. Guillemin, Quantitative Gas Chromatography for Laboratory Analysis and On-Line Control, Elsevier, Amsterdam, 1988.
[18] L.S. Ettre, J.V. Hinshaw, Basic Relations of Gas Chromatography, Advanstar, Cleveland, OH, 1993.
[19] V.M. Morris, J.G. Jughes, P.J. Marriott, J. Chromatogr. A 755 (1996) 235.
[20] E. Grushka, M.N. Myers, P.D. Schettler, J.C. Giddings, Anal. Chem. 41 (1969) 889.
[21] G.A. Korn, T.M. Korn, Mathematical Handbook for Scientists and Engineers, McGraw–Hill, New York, 1968.
[22] J.C. Giddings, Dynamics of Chromatography, Marcel Dekker, New York, 1965.
[23] L.M. Blumberg, Anal. Chem. 64 (1992) 2459.
[24] L.M. Blumberg, Chromatographia 39 (1994) 719.

[25] J.C. Sternberg, Adv. Chromatogr. 2 (1966) 205.
[26] J.P. Foley, J.G. Dorsey, J. Chromatogr. Sci. 22 (1984) 40.
[27] D. Hanggi, P.W. Carr, Anal. Chem. 57 (1985) 2394.
[28] W.D. Snyder, L.M. Blumberg, in: P. Sandra (Ed.), Proceedings of 14th International Symp. on Capillary Chromatography, Baltimore, 25–29 May, 1992, ISCC92, Baltimore, 1992, p. 28.
[29] L.M. Blumberg, M.S. Klee, Anal. Chem. 70 (1998) 3828.
[30] L.M. Blumberg, M.S. Klee, in: P. Sandra, A.J. Rackstraw (Eds.), Proceedings of 20th International Symposium on Capillary Chromatography (CD ROM), Palazzo dei Congressi, Riva del Garda, Italy, 26–29 May, 1998, I.O.P.M.S, Kortrijk, Belgium, 1998.
[31] L.M. Blumberg, M.S. Klee, in: P. Sandra, A.J. Rackstraw (Eds.), Proceedings of 23rd International Symposium on Capillary Chromatography (CD ROM), Palazzo dei Congressi, Riva del Garda, Italy, 5–10 June, 2000, I.O.P.M.S, Kortrijk, Belgium, 2000.
[32] L.M. Blumberg, M.S. Klee, J. Microcol. Sep. 12 (2000) 508.
[33] L.M. Blumberg, R.D. Dandeneau, J. High Resolut. Chromatogr. 18 (1995) 235.
[34] J.Å. Jönsson, in: J.Å. Jönsson (Ed.), Chromatographic Theory and Basic Principles, Marcel Dekker, New York, Basel, 1987, p. 27.