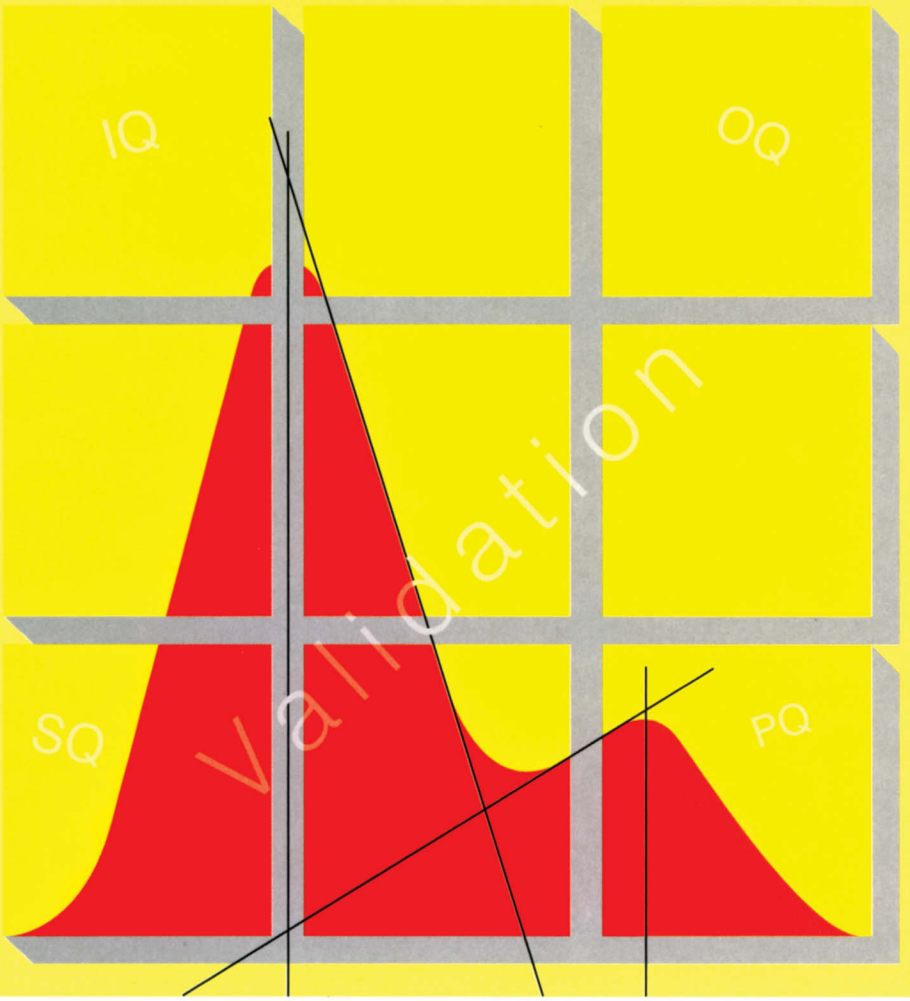
RSC CHROMATOGRAPHY MONOGRAPHS

Series Editor
Roger M. Smith

CHROMATOGRAPHIC INTEGRATION METHODS

Second Edition

by Norman Dyson



CHROMATOGRAPHIC INTEGRATION METHODS
Second Edition

RSC Chromatography Monographs

Series Editor: Roger M. Smith, *University of Technology, Loughborough, UK*

This series is designed for the individual practising chromatographer, providing guidance and advice on a wide range of chromatographic techniques with the emphasis on important practical aspects of the subject.

Supercritical Fluid Chromatography
by Roger M. Smith, *University of Technology, Loughborough, UK*

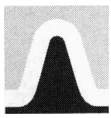
Packed Column SFC
by T.A. Berger, *Berger Instruments, Newark, Delaware, USA*

Chromatographic Integration Methods, Second Edition
by Norman Dyson, *Dyson Instruments Ltd., UK*

How to obtain future titles on publication

A standing order plan is available for this series. A standing order will bring delivery of each new volume immediately upon publication. For further information, please write to:

The Royal Society of Chemistry, Turpin Distribution Services Ltd., Blackhorse Road, Letchworth, Hertfordshire SG6 1HN, UK.
Telephone: +44(0) 1462 672555; Fax: +44(0) 1462 480947



RSC
CHROMATOGRAPHY
MONOGRAPHS

Chromatographic Integration Methods

Second Edition

Norman Dyson
Dyson Instruments Ltd., UK

THE ROYAL
SOCIETY OF
CHEMISTRY
Information
Services

ISBN 0-85404-510-4

A catalogue record for this book is available from the British Library

First Edition published 1990
Reprinted 1992, 1996

Published by The Royal Society of Chemistry,
Thomas Graham House, Science Park, Milton Road, Cambridge CB4 4WF, UK

For further information see our web site at www.rsc.org

Typeset by Keytec Typesetting, Bridport, Dorset
Printed in Great Britain by MPG Books Ltd, Bodmin, Cornwall, UK

To Jenny

Preface to First Edition

This book is about the measurement of chromatographic peaks. In particular, it describes and discusses the manual and electronic techniques used to make these measurements, and how to use integrators. The aim of the book is simply to help analysts extract more data from their chromatograms, and help them to understand how integrators work so that results are never accepted unquestioningly.

The book is written for those who use integrators most; the lab technicians and students, and the analysts who do the chromatography. It is especially for those who never got around to reading their integrator manuals. When all else fails, read the manual or this book.

There are no books on integrators (to the author's knowledge), and the subject of integration tends to merit a short chapter in books on quantitative chromatography. This text starts at the detector and works outwards. Most of it is about the integrity of peak representation, different methods of measuring peak data and validation of the results.

The electronics of data acquisition are kept to a minimum, enough to point the interested in the right direction, but not enough to redesign the front end of an integrator.

Current methods of integration measure peaks according to simple rules developed for manual measurement. Perpendiculars and tangents were shown years ago to have only limited justification yet they remain in universal employment because there is nothing better to replace them at the present time.

The goal of the 'definitive integrator' which would measure peak areas accurately, precisely, and instantly was scheduled to develop in three broad stages:

(1) Manual methods involving measurement of peaks on strip chart recorder traces. From this developed:
(2) Low cost integrator and micro-computer techniques. These were meant to be short term measures, replaced at the earliest opportunity by:
(3) Computer-curve fitting of peak shapes and deconvolution of overlapping peaks using techniques that did not require straight lines except in the few cases where they might be justified.

The move from the second to third stage has not taken place, despite much effort. Manual measurement of chromatograms brought attention to what might be called the 'problems of integration': how to define peak boundaries (the limits of integration) and draw baselines, the effects of asymmetry, separation of overlapping peaks, measurement of small peaks, signal-to-noise ratio, *etc.*

When the first commercial integrators were introduced, they eliminated the tedium of measuring peaks by hand, greatly speeded up the processing of analyses, and increased the precision and number of analyses which could be made in a working day.

What integrators did not do was to improve the methods of integration. Manufacturers simply selected the best of the manual methods and built integrators to use them.

Curve-fitting of experimental data to a good peak model was believed to be the correct way forward. Defining the peak model implicitly defines the processes and mechanisms of peak elution and these would all become known in time. With such models, there would be no need for perpendiculars, tangents, or the construction of artificial baselines as had been used for manual deconvolution. Unfortunately, finding suitable peak models has so far proved to be impossibly difficult. Peak shape is a fluid thing influenced by factors outside of the column as well as 'simple' sorption/desorption; factors such as solute polarity, quantity, the column, mobile phase composition, system dead volume, injection proficiency, detector geometry, electronics, heaters, flow controllers, other solutes, the solvent which the solutes are dissolved in, or any combination of these factors.

Defining peak shape requires all of these factors to be brought under control, but research has shown that even when this is done, and a good mathematical model created for a given peak, the model does not necessarily describe neighbouring peaks which have different properties and interact in different ways with the column; nor does it describe the same peak a week later if the column has aged in the meantime.

It has been judged that there is simply not enough information in the output of a single channel detector such as an FID or UV detector to allow measurement of a whole chromatogram. Analysts cannot even be sure how many peaks there are in a group. Under these circumstances, contemporary integrators have limited powers of data interpretation. To paint the full picture, additional data must be supplied and the quality of chromatograph manufacture must bring better experimental control of the sample. It is possible that 'three-dimensional detectors' such as the Diode Array, GC–MS or FTIR will supply the extra information to allow further progress. The 'next generation integrator' might well be based on the creation and processing of a three-dimensional field which is Gaussian in one plane and Lorentzian in the other, or approximately so. If so, it will worry the independent integrator manufacturers if they are denied information on interfacing to these complex detectors.

The purpose of this book is not 'to teach chromatographers how to use an integrator to measure peak areas correctly', because the principles on which integrators are based do not allow the accurate measurement of any peak unless it is completely resolved and stands on a flat, noise-free baseline. The purpose is to describe the rules of integration as they stand, and the implications this has on peak measurements. Only by knowing these rules and working within their limits can the integrator be used to its best.

The moral of the story can be anticipated in advance: no matter how expensive the integrator or computer, it is no substitute for good chromatography by a capable and critical analyst. The quality of results is largely determined before the integrator is ever brought into use.

Preface to Second Edition

Many friends and colleagues said they liked the diagrams in the first edition but thought that Chapter 2 was short in comparison with Chapters 1 and 3. In this second edition there are five chapters: Chapters 1 and 3 have been expanded and split into two. There are 14 additional diagrams to illustrate details in the text, a literature survey has been conducted, an Author Index has been added, some errors have been corrected and Chapter 3 remains almost unchanged.

There have been some significant papers published on the errors of peak measurement. The growing awareness of the limits of integrator performance has been coupled with the availability of tools to measure them systematically. This work has been described in Chapter 2; in particular, the work of Papas in the USA and Meyer in Switzerland has added much to the estimation of errors and helped to spread the message about the fallibility of integrators.

The other big issue since the first edition has been Regulatory demand for 'Quality' and Validation, and Chapter 4 includes directions on how to validate an integrator. Translating textbooks on Quality into 'what do I do next, here in the lab, now?' has stirred up a lot of problems and confusion. Many Standard Operating Procedures for validating chromatographic methods, accepted by Regulators, never even mention the independent calibration of the integrator as the job to be done first. If the integrator remains suspect, quantitative chromatography cannot begin; it is not only the signal measurer, it is the diagnostic tool – the chromatographer's multimeter.

A lot of good work has been published on peak deconvolution and chemometrics, though none of it has induced the major suppliers to replace the perpendicular and tangent methods for peak separation. Some elegant deconvolution software has appeared and is presently undergoing the test of time. But one task of chemometrics is still to catch and quantify the errors of perpendiculars and tangents.

Stand-alone integrators are less popular now than PC based systems. PCs take the larger market share, but the dedicated simplicity of stand-alone integrators still appeals to many analysts and where labs have standardized on a particular model, the task of revalidating on a new PC system is too daunting for casual change.

Finally, I would like to acknowledge and thank everybody who contacted me with suggestions and kind comments. The first edition clearly reached those for whom it was intended – the people at the bench, the 'infantry'. I am pleased that it helped.

N. Dyson e-mail NDYSON6650@AOL.COM

Contents

Chapter 1 Measurements and Models **1**

1 The Basic Measurements 1
Measurements and Their Use 1
2 Quantitation 1
Peak Area and Peak Height 1
Peak 'Volume' 3
Retention Time and Solute Identity 4
Column Hold-up Time 4
3 Diagnostics and System Suitability Tests 4
Baseline Noise 5
Baseline Signal Level 6
Peak Boundaries 6
Peak Width and Column Efficiency 6
Peak Asymmetry 7
Peak Resolution 9
Retention Time Stability 10
4 Results Assessment 10
Correct Peak Measurements 10
Total Peak Area 10
Correct Number of Peaks 11
Coefficient of Variation and Relative Standard Deviation 11
Comparison Against Standards 11
Trend Analysis 11
5 Performance Measurement 11
Chromatograph Utilization 11
Analyst Workload 12
Cost per Analysis 12
6 Chromatographic Peak Models 12
The Gaussian Function 13
The Function 13
Peak Height 14
Ratio of Area/Height 14
Calculation of Peak Width at Various Heights 15
Gaussian Peak Shape Tests 16
The Points of Inflection 16

Fractional Peak Area Bounded by Various Widths 17
Loss of Area from the Base of a Gaussian Peak 18
Gaussian Peak Maximum as an Approximate Parabola 18
The Exponentially Modified Gaussian Function 20
The Function 20
Practical Application of the EMG Function 22
EMG Peak Shape Tests 23
7 Statistical Moments of a Chromatographic Peak 23
Zeroeth Moment, m_0 24
First Moment, m_1 24
Second Moment, m_2 25
Third Moment, m_3 25
Fourth Moment, m_4 26
Higher Odd Moments 27
Higher Even Moments 27
Measurement of Peak Moments 27
Practical Disadvantages and Uses 27
8 Manual Peak Area Measurement 28
Height × Width at Half Height 28
The Condal-Bosch Area 29
Peak Area by Triangulation 30
Manual Measurement of Asymmetric (EMG) Peaks 31
9 References 33

Chapter 2 Errors in Peak Area Measurement 35

1 Accuracy and Precision 35
Accuracy 35
Precision 35
Inaccuracy and Uncertainty 35
Repeatability 36
Reproducibility 36
Random Errors 37
Systematic Errors 37
Spurious Errors 37
Causes of Imprecision 37
Causes of Inaccuracy 37
Precision of Integrator Measurements 37
2 Accurate Representation of the Solute Profile 39
Additivity of Signals 40
Negative Detector Signals 41
Electronic Peak Distortion 41
3 Peak Area and Solute Quantity 43
Mass and Flow Sensitive Detectors 43

Flow Sensitive Detectors 44
Mass Sensitive Detectors 45
Detector Overload 47
Linearity 48
Detector Non-linearity 49
Column Non-linearity 49
Electronic Noise and Drift 50
4 Sources of Peak Measurement Error 50
Hardware Errors 50
Noise 51
Noise and Frequency 51
Short-term Noise 51
Long-term Noise 51
Drift 51
Errors Created by Noise 53
Signal-to-noise Ratio: The Smallest Measurable Peak 54
Limits of Detection and Quantitation 55
A 'Good Baseline' 55
Errors Created by Baseline Drift 56
Drifting Baseline and Peak Measurement 56
Baseline Drift and Retention Time 58
Reduced Detector Operating Range 59
Some Preliminary Conclusions 60
Errors of Incomplete Peak Resolution 60
Height of a Fused Peak 62
Perpendicular Separation 62
Effect of Overlap on Area Measurement Accuracy 64
Error Correction Strategies 67
Error Propagation through Overlapping Groups 68
Peak Overlap and Calibration 70
Tangent Skim Errors 71
Errors of Perpendicular/Tangent Transition 72
Overlapping Peaks on Sloping Baselines 72
Multiple Fused Peaks 74
Mathematical Deconvolution of Overlapping Peaks 76
The Isomer Test 77
Errors from Peak Asymmetry 78
Asymmetry and Peak Tailing 78
Asymmetry and Base Broadening 80
Asymmetry and Manual Peak Measurement 80
Unequal Asymmetry 81
The Achievement of Accuracy in Chromatographic Analyses 82
Peak Area *vs.* Peak Height 83
User Surveys 85
5 References 86

Chapter 3 Manual Measurement of Peaks 89

1 Representation of the Detector Signal by Chart Recorder 89
Slow Response Time 89
Non-linear Signal Response 89
Pen Head Damping 90
Amplifier Noise 90
Manual Measurement of Noise 91
Dead Band 91
Chart Motor Control 92
Attenuator Accuracy 92
2 Measurement Strategies 92
Counting Squares 93
Cutting and Weighing 93
Planimeters 94
3 Measurements based on a Peak Model 94
Pencil and Rule Methods 95
Measurement of Peak Height 95
Height *vs.* Area Measurements 95
Triangulation 96
Peak Height × Width at Half Height 97
Condal-Bosch Variation 98
The Foley Variations 99
Measurement of Overlapping Asymmetric Peaks 101
Peak Shape Tests 101
4 Errors of Manual Measurement 101
Optimum Peak Shape 103
Advantages and Disadvantages of Manual Peak Measurement 105
5 References 105

Chapter 4 Digital Integrators 107

1 A Brief History of Integrators 107
Strip Chart Recorder Techniques 107
Electromechanical Counters 107
Electronic Integrators 110
TTL Instruments – The First Electronic Integrators 110
Microprocessor-based Integrators 111
Integrators with Printer Plotters 111
The Impact of the Microcomputer 112
2 Current Integrator Status 112
Standard Integrator Specification 113
Integrator Files 114
Analysis Parameters for Peak Measurement 114
Peak Width 115

Slope Sensitivity or Detection Threshold 115
Baseline Drift Tolerance 117
Time to Double 117
Minimum Area or Height Threshold 118
Analysis Duration 118
Integrate Inhibit 118
Forcing Tangent or Perpendicular Separation 118
Forcing Baseline Detection 120
Horizontal Baseline Projection 120
Inverting Negative Peaks 122
Instrument Control and Communications 123
Sample Management, Calculations and Calibrations 123
Peak and Solute Identity 123
Relative Retention Times 124
Standard Calculations 124
Area% or Normalization 124
Internal Standard 125
External Standard 125
Sample Scheduling 126
Calibration 126
Report Preparation and Output 127
Trends 128
Report Distribution 128
3 Validation of Integrators 129
The Integrator as a Diagnostic Tool 129
Specification, Installation, Operational, Performance Qualification 130
Integrator Specification and Selection (SQ) 130
Installation Checks (IQ) 131
Operational Qualification (OQ) 132
Validation and Calibration: The Difference 132
Records 132
Performance Qualification (PQ): Regular Testing of the Integrator 133
Logging Results 134
Validation of Software 134
4 Standard Chromatograms 135
Creation of Standard Chromatograms 136
Computer-based Standard Chromatograms 136
Specification of Standard Chromatograms 137
Creation of Synthetic Chromatograms 137
Baselines 137
Baseline Noise 137
Drift 138
Traceability of Synthetic Chromatograms 138
5 References 138

Chapter 5 Digital Measurement of Peak Areas 140

1 Signal Sampling 140
Analogue to Digital Conversion 140
Voltage to Frequency Conversion 140
Dual Slope, Integrating A/D Conversion 141
Resolution of A/D Converters 143
Auto-ranging of A/D Converters 143
Data Sampling Frequency 144
Sampling Frequency and Integrator Manufacture 144
Sampling Frequency and Data Processing Algorithms 144
Data Bunching 144
Baseline Bunching 145
Estimating the A/D Sampling Frequency and Bunch Size 146
Data Bunching and the Peak Width Parameter 148
Peak Width Parameter and Analysis Reprocessing 148
Peak Sampling Synchronization 149
Rounding or Truncation Errors 149
Aperture Time Jitter 149
2 Filtering and Smoothing the Chromatographic Signal 150
Filtering 150
Smoothing 150
Electronic Filters 150
Sampling Frequency and Mains Coupling 150
Smoothing Techniques 151
Digitizing and Integrating 151
Bunching 151
Moving Windows and Polynomial Curve Fitting 153
Optimum Filtering 154
Signal Subtraction 155
Chromatogram Plotting 157
Matched Filtering 157
3 Location and Measurement of Peaks 158
Finding the Peaks 159
Retention Time 160
Peak Shape Test 160
Properties of the Smoothed Data 160
Data Integrals 161
Baseline Fluctuations and Slope Sensitivity 162
Programming Slope Sensitivity 162
Slope Sensitivity and Representative Baseline 162
Updating Peak Width and Slope Sensitivity during Analyses 163
Location of the Limits of Integration 163
Small Peak Filtering 165
Location of Peak End 165

Measurement of Peak Area 166
Baseline Convention 168
Measurement of Peak Height 168
Measurement of Retention Time 168
Measurement of Two Unresolved Peaks 169
Measurement of Individual Peak Areas 170
Peak Measurement Diagnostics 170
Measurement of Fused Groups 170
Shoulders 171
Measurement of Tangent Peaks 172
Tangent Skimmed Groups 173
Small Peaks between Larger Ones 174
Tangents on Tangents 175
4 Baselines, A More Detailed Discussion 175
Baseline Drift Limit 175
False Starts 177
Formulating a Baseline Definition 177
Mis-timing 'Start' and 'End' 178
End of 'Integrate Inhibit' 178
Forcing Baseline 179
Incorrect Programming of Baseline Drift Tolerance 180
Fused Tangent Measurement 182
Single Peaks on a Rising Baseline 183
Valleys Between Fully-resolved Peaks 184
Negative Dips and Constructed Baselines 184
Assigning Baseline Beneath the Whole Chromatogram 187
Valley–Valley Skim 190
5 Conclusions 190
6 References 191

Author Index 193

Subject Index 196

Glossary of Terms

A, A_i	Area, of component i
B	Number of samples in data bunch
B/A	Asymmetry ratio
$\pm b$	Baseline gradient
$b(n)$	Coefficient used to evaluate EMG function
c, c_i	Concentration, of component i
c_i	Weighted coefficient in Savitsky–Golay smoothing
C	Capacitance
CMR	Common Mode Rejection
E	Excess, 4th statistical peak moment
F	A/D frequency
f	Bunched sample frequency
F_h	Fractional peak height ($= h_t/H$)
H	Peak height
h_1	Height at point of inflection
h_t	Detector signal at time t
h'_t	1st derivative of detector signal at time t
h''_t	2nd derivative
I_n	Latest measured integral-bunched data sample
$I_{(n-M)}$	'M'th predecessor of latest measured integral
$I(z)$	Integral term in EMG function
K	Instrument constant
K	Solute concentration/detector reponse
k	Area correction factor = A(true)/A(measured)
m	$(2m + 1)$ data points in a data smoothing window
m_n	nth statistical peak moment
N	Noise amplitude
N	Savitsky–Golay normalizing constant
N	Number of theoretical column plates
n	Number of clock pulses equal to one datum
P	Number of positive integrals to trigger peak detection
Q	Quantity of solute

R, R_i	Response factor, of component i = quantity of solute to produce unit area count
R_s	Resolution
R	Electrical resistance
S	Slope sensitivity
S	Normalization scaling factor
S/N	Signal-to-noise ratio
S	Total analyte signal
s_i	Analyte signal
t	General time variable
t_G	Retention time of Gaussian peak
t_R	Chromatographic retention time
Δt	Time width of bunched data sample
t_{mean}	Mean peak retention time (of centre of gravity)
V	Volume
V	Valley height
V	Voltage
$\dot{V}$	Volume flow rate of mobile phase $= dV/dt$
W_i, W_s	Weights, of solute i, standard s
w	Peak width
w_b	Peak base width
w_h	Peak width at fractional height h
w_t	Peak width at time t
$w_{0.5}$	Peak width at fractional height $0.5H(= w_h)$
$\pm X$	Gaussian width boundary
X	Time interval in ASTM E 685 noise measurement
y	General variable
$\overline{y}$	Smoothed central datum
Z	Long-term noise in ASTM E 685
Z	Number of zero slope integrals to end peak measurement
γ	Skew, 3rd statistical peak moment
δ	Peak separation (dimensionless units)
σ, σ_G	Standard deviation of peak, of Gaussian peak
σ/τ	Measure of peak asymmetry from EMG theory
σ^2	Variance, the 2nd statistical moment
τ	EMG time constant

CHAPTER 1

Measurements and Models

1 The Basic Measurements

The basic measurements made by an analyst on a chromatogram for purposes of quantitation are shown in Figure 1.1; they comprise all the 'Y's and 'X's of the trace, the voltages and times, their combinations, repetitions and trends. There are no other measurements to be made, but much analytical information can be derived from these measurements.

These quantities are obtained by integrators, lab computers or data processors – the names are used synonymously throughout this book. The integrator is the only window into the chromatograph to show what is happening. It is both measuring device and diagnostic tool.

Regulatory pressures from Good Laboratory Practice (GLP), Good Manufacturing Practice (GMP), ISO 9000 and bodies such as the US Food and Drug Administration (FDA) put increasing demands on the quality and credibility of analytical results. It is the aim of this book to show analysts how to use integrators to provide good quality results – provided that the chromatography is up to it.

Measurements and Their Use

Integrator measurements serve three analytical purposes:

(1) solute identification and quantitation;
(2) diagnostics, trouble-shooting and system measurement;
(3) results assessment and trend analysis;

and, increasingly, one corporate finance purpose for monitoring:

(4) performance measurement and lab resource utilization.

2 Quantitation

Peak Area and Peak Height

Solute quantity is measured from peak areas and/or heights. In a controlled analysis, peak area is the true measure of solute quantity if the solute elutes intact

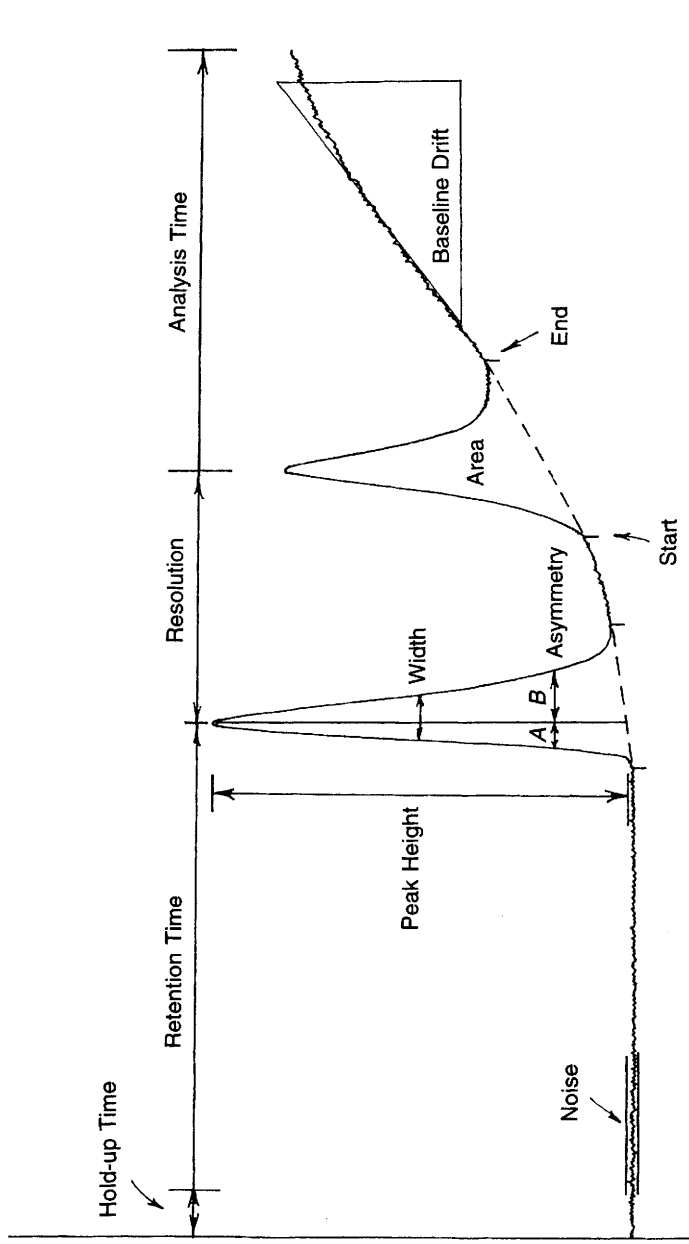


Figure 1.1 *The basic measurements*

and is detected linearly. It can be shown to be so theoretically for peaks with or without a known shape, and experimentally by plots of area against solute quantity.

Peak height is an alternative measure of solute quantity. There is no theoretical proof of this unless a specific peak shape such as Gaussian is assumed, but experimentally it is easy to show that plots based on height and quantity are also linear over a usable though smaller range (Figure 1.2), even for peaks that are not particularly symmetrical as long as their shapes do not change. Peak height is an easier measurement to make and has advantages for the measurement of small overlapping peaks, but it ignores most of the data contained in the detector signal.

The choice of which is best to use in practice is discussed at the end of Chapter 2. Area is the normal choice when signal-to-noise ratio (S/N) is large because height is susceptible to peak asymmetry while area is not, and the linear dynamic range for area measurement is greater than for height. When signal-to-noise ratio is small, height is preferred because errors of baseline placement affect height less than area.

Peak 'Volume'

The ability of the photo diode array detector (DAD) to measure peaks over a range of wavelengths allows a third measure of solute quantity: peak 'volume'. Here, all the peak areas measured at the wavelengths to which the analyte species responds are added together to create a three-dimensional measure (time, amplitude, wavelength). It is proposed that this measure is better because it contains more data and more information. For this to be true, background noise must be proportional to analyte signal otherwise the signal-to-noise ratio degrades away from λ_{max} and the benefits of the extra information become compromised by its lower quality. The usual case is that noise has origins independent of the solute signal and does degrade the signal.

Peak volume would be better than area or height if the signal-to-noise ratio is good at all measurements but this may be hard to achieve. With current diode array

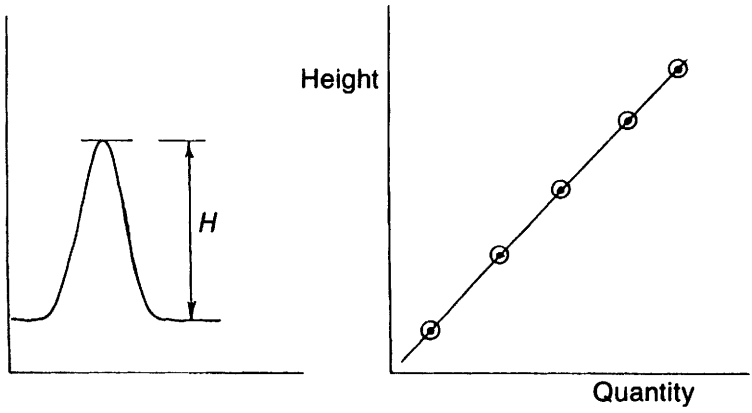


Figure 1.2 *Peak height and quantity show a linear relationship*

detectors and current integration software, peak area (measured at λ_{max}) remains the safest measure of solute quantity, especially for trace peaks.

Retention Time and Solute Identity

Once solutes in a mixture have been identified, they are subsequently recognized on a daily basis by their retention times. Integrators assign peak names and response factors to peaks which elute inside a specified time window. If another peak elutes at that time it also will be recognized as the expected peak; if two peaks co-elute they will not be uniquely identified. If retention time varies with sample size due to increasing peak asymmetry, it is quite possible for incorrect identification to be made when a peak crosses from one window into the next. When this happens, the integrator will assign the wrong name, response factor and standard concentration to that peak, and these errors will continue into the final report.

Integrators can measure relative retention time, *i.e.* the peak retention time compared with the retention time of a standard peak. Experimental variations cancel but the errors of both measurements add together so that relative retention times are more accurate but less precise than absolute retention time. There is also the usual problem of finding a suitable standard.

To an integrator, retention time is the elapsed time from the moment of injection until the peak maximum emerges, which includes the gas or solvent hold-up time. The retention time of asymmetric peaks does not coincide with the centre of gravity of the peak. Separation of the observed retention time t_R (mode), from the peak centre of gravity (mean) is one measure of asymmetry[1] (see also: First Moment and Equation 40).

The mean retention has not achieved any common use in the analytical laboratory. It is difficult to measure manually but integrators which sample the peak signal at a fixed frequency could, with a simple addition to their software, measure it very easily. Its theoretical value lies in the separation of t_R and t_{mean} being equal to the Exponentially Modified Gaussian (EMG) time constant τ (Figure 1.3).

Column Hold-up Time

The journey time of a molecule or atom of mobile phase from the beginning to the end of the column is called the column hold-up time. Since the mobile phase is the propellant of a chromatograph, no solute can emerge before this time has elapsed. The hold-up time is the shortest retention time possible, it is also equal to the total residence time of a solute in the mobile phase as it traverses the column.

Integrators measure the column hold-up time as the retention time of an unretained solute. Knowing this time allows the analyst to optimize mobile phase flow rate to achieve maximum column performance.

3 Diagnostics and System Suitability Tests

Most of the instrument checks that an analyst makes before injection to assess instrument readiness can be made automatically by standard integrator routines, or

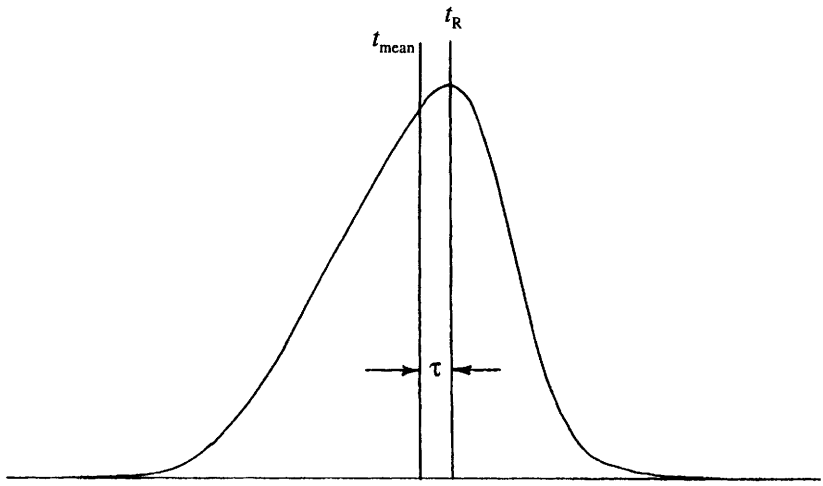


Figure 1.3 *Relationship of mean retention time* (t_{mean}) *and observed (mode) retention time* (t_R) *to EMG time constant* (τ)

by a software program running in the integrator. Absence of noise and a stable baseline in the correct place (near zero millivolts) are accepted indications that a chromatograph is ready for sample injection. Tests on peak shape, size and resolution indicate whether the autosampler, column and detector are working to specification. Injection of standard or known samples calibrates the detector response and allows comparisons with working samples ('unknowns') in order to confirm that the correct sample has been injected and is within specification or not. Finally the sheer number of analyses per day can be counted to show whether the chromatograph is fully utilized, whether the work load is building up, and other trends.

The drift away from stand-alone integrators towards computer systems linked to laboratory information management (LIM) systems has accelerated the creation of multiple reports and database storage of results. When properly introduced, this additional scrutiny of results for trends brings good improvements in quality control.

Baseline Noise

Baseline noise makes peak areas hard to measure and so reduces confidence in analysis results. Integrators assess baseline noise by means of 'Noise' or 'Slope Sensitivity' tests, and report its average value. On a daily basis this value should be recorded and the trend shown to be reasonably constant; analysts should know what noise levels are acceptable. Any worsening of noise must be addressed; it might indicate that the column needs reconditioning, the detector needs cleaning, or that some part of the electronics is about to fail.

Baseline Signal Level

Unexpected high signal levels warn the analyst that a peak which ought not to be there is eluting, or that the detector has become contaminated; it may be that temperature or flow control is drifting. Whatever the reason, some delay or corrective action is necessary before the next injection. Integrators indicate a value for the baseline level and some incorporate a 'Not Ready' warning if the level is outside a prescribed range.

Peak Boundaries

The start and end times of peaks are measured by noting the times when integration of a peak is commenced and terminated. These times are used to measure peak width and, when combined with retention time, to measure peak asymmetry (see below).

Knowing these times allows the analyst to assess column performance and judge whether the column needs conditioning or replacement.

Peak Width and Column Efficiency

The measurement of column performance in terms of efficiency, peak skewness or asymmetry, and peak resolution are key System Suitability tests.

A column is 'efficient' if the lateral diffusion of solute bands is restricted during their column residence time. The narrower the peaks are, the better the separation from neighbouring peaks and the more accurately they can be measured. Efficiency is measured as numbers of theoretical and/or effective plates per column. It is calculated from peak widths (actually, from peak variance σ^2, see Figure 1.6) and retention times.

In repeated analyses, the retention time and width of a fully-resolved peak at a fixed height above the baseline should be constant. Variations in either quantity indicate that column performance is varying or that some experiment-controlling parameter such as temperature or flow is drifting.

Peak width is related to variance and contains components of spreading added by the injection port, system dead volume, and by the detector and electronics as well as the column.

If σ is the standard deviation of a peak (see Figure 1.6), σ^2 is the peak variance[1,2] and:

$$\sigma^2_{\text{total}} = \sigma^2_{\text{inj}} + \sigma^2_{\text{col}} + \sigma^2_{\text{det}} + \sigma^2_{\text{etc.}} \quad (1)$$

Of these, the contributions to variance from injection port and column are the most important as they can become worse with instrument use. Detector broadening is important until the cause is found and removed; once removed it tends not to recur. Extra-column broadening effects are observed in isolation by connecting the injection port directly to the detector with, for example, uncoated capillary tubing.[3]

If the injection technique is not repeatable, or if the column is deteriorating, there will be a variation in peak width and apparent variation in column efficiency.

Most integrators can measure peak area and height simultaneously and this allows empirical measurement of peak width. For symmetrical peaks the ratio of area to height gives the peak width at 45.6% of the height [$= \sqrt{(2\pi)}\sigma$, see Equation 13]; for asymmetric peaks, the ratio is equal to the width at some other height. Alternatively, peak width can be measured as base width, the time interval between the start and end of peak integration ($= 6\sigma$ for a symmetrical peak). Once peak width is known, column efficiency can be calculated.

Column efficiency is generally defined as the number of theoretical plates N, calculated from:

$$N = \frac{{t_R}^2}{\sigma^2} \tag{2}$$

However, the value of σ^2 can vary with the peak model chosen[4] and t_R may be taken at the centre of gravity of the peak rather than the conventional peak top. For a Gaussian peak:[5]

$$\sigma = \frac{w_h}{2\sqrt{[-2\ln(h)]}} \qquad \text{from (16) below}$$

thus

$$\frac{N = 8{t_R}^2 \ln(h)}{{w_h}^2} \tag{3}$$

where N = number of column plates
t_R = retention time
w_h = width of peak at fractional peak height h, usually $h = 0.5H$

All are quantities the integrator can measure.

There are limitations in measuring column efficiency this way for real peaks – the theory is based on the Gaussian model and few peaks are exactly Gaussian, but monitoring the width of a well-resolved peak in the chromatogram over successive analyses will empirically indicate any column degradation or highlight imprecise control of instrument parameters. It will not necessarily identify what is causing the change or what is to be corrected. Alternative calculations of efficiency based on asymmetry or the second peak moment (*q.v.*) are not always available in integrators.

Peak Asymmetry

Asymmetry broadens the base of a peak, increases peak overlap with neighbours and makes peak areas harder to measure; an aim of method development is therefore to produce symmetrical peaks.

A peak will be symmetrical if:

(i) injection technique is good;
(ii) dead volume in the solute flow path is absent;
(iii) the solute quantity is not large enough to overload the column or detector;
(iv) the residence time of the solute in the stationary phase is long enough to achieve dynamic equilibrium;
(v) the solute adsorption isotherm is linear.

It follows that where peak asymmetry is evident, some improvements to the analysis may be desirable: change to a different column, improve the injection technique,[6] reduce the injection volume, *etc.*

Asymmetry can be described in terms of the separation of peak mean and mode times, but a more common measure of asymmetry compares the peak half widths on either side of the peak mode (see Figure 1.4). This is called the Asymmetry Ratio,[1] or sometimes just 'Tailing', and it varies with peak height.

Such measurements can in principle be made at any height but are usually made near the peak base where asymmetry is greatest, typically at 10% of peak height for which much theory has been developed[7–9] or at 5% for US Pharmacopoeia and FDA requirements. When peaks overlap, the widths at 10% peak height remain measurable for longer, but obtaining the '5% width' demands a better standard of chromatography.

Asymmetry can also be expressed as the Tailing Factor or Tailing Coefficient,[10]

$$\text{Tailing Factor} = \frac{w_{0.05}}{2A} = \frac{(B+A)}{2A} \tag{4}$$

which is the average of B/A and 1, and therefore half as variable as B/A alone.

At the top of the peak, measurements of B and A would be relatively small, more sensitive to measurement errors, and B/A is closer to unity. Nevertheless, it has been suggested[11] that it is probably more accurate, although less precise, because there is less measurement interference from neighbours and noise at the top of a peak.

Measurement of the Asymmetry Ratio is one of the System Suitability measure-

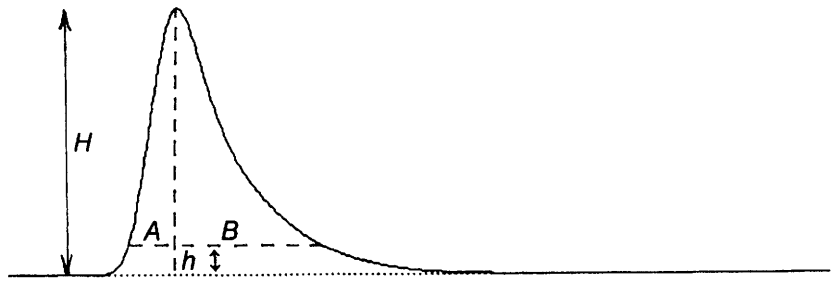


Figure 1.4 *Asymmetry ratio*[1] $= B/A.\ h = 0.05\ H$ *or* $0.10\ H$

ments and a standard routine on most integrators. Experimentally, measurements can be taken from any peak in the chromatogram, but they should only be taken from fully-resolved peaks.

The asymmetry measurement, B/A, as shown in Figure 1.4 is used in the evaluation of the mean retention time, the number of column plates and variance derived from the Exponentially Modified Gaussian peak model.[7]

Peak Resolution

Peak resolution is the degree of separation of two adjacent peaks measured in units of peak width rather than time. It compares the actual peak top separation to the average constructed peak base width and is used to judge whether the separation between two peaks is good enough for accurate quantitation and fast analysis. It is a dimensionless quantity. In Figure 1.5:

$$\text{Resolution, } R_s = \frac{t_{R2} - t_{R1}}{(w_2 + w_1)/2} \tag{5}$$

Allowing that the constructed base width of a symmetrical peak is approximately twice the width at half height $w_{0.5}$:

$$R_s \approx \frac{t_{R2} - t_{R1}}{w_{0.5,1} + w_{0.5,2}} \tag{6}$$

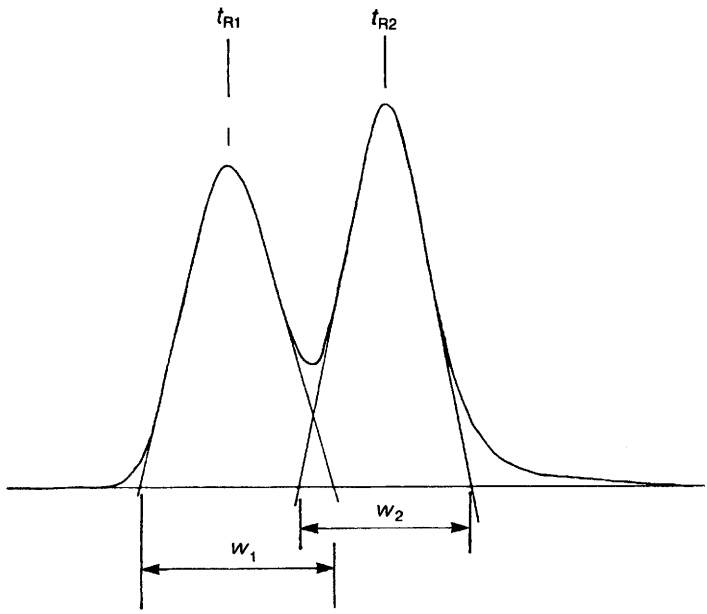


Figure 1.5 *Peak resolution*

In theory, a resolution of 1 means that the peaks are just resolved down to baseline and areas can be measured accurately. In practice, if peaks are unequal in size and asymmetric, a resolution nearer 2 is needed to obtain near baseline separation, and even that is sometimes not enough.

Equation 6 is only an approximation based on Gaussian theory and can be as much as 100% in error for asymmetric peaks, but resolution of peaks calculated from the difference in their retention times and widths (derived from Area/Height) can be monitored as an empirical test to show when the column needs reconditioning or replacement.

Most analysts seek to develop analyses which have near symmetrical peak shapes and so avoid problems of non-Gaussian theory.

Retention Time Stability

The stability of retention time values from one analysis to the next shows up the precision of the flow or temperature control of solvent composition. Standard operating procedures may impose a limit on the amount of variation that is allowed.

4 Results Assessment

No analytical report should be accepted unquestioningly. Before accepting any results from an integrator or computer, certain simple checks can be made to test the credibility and accuracy of the report.

Correct Peak Measurements

An integrator might draw a baseline in the wrong place or separate two peaks by skimming a tangent where a perpendicular would be better, or vice versa. It is important to inspect these features on the actual chromatogram and confirm their correctness and uniformity of measurement. Measurement diagnostics are printed beside peak areas on the integrator's final report describing how the peak was measured. They too should be checked and judged correct or not.

Chromatogram event marks are sometimes used as an alternative to baseline drawing in older or simpler integrators and show the start and end of peak integration. One event mark should be printed at the correct place on every wanted peak in the chromatogram. If these event marks have been switched off (or deselected as a software option), switch them back on again. They show whether Slope Sensitivity (*q.v.*) has been programmed correctly. Event marks should not litter the baseline or occur at peak maxima; this indicates that the Slope Sensitivity is too sensitive.

Total Peak Area

The total peak area count is a measure of injection volume. Its constancy over a sequence of analyses is a measure of injection proficiency and indicates good or

bad injection technique. Important analyses may set a limit on how much variation is allowed.

Correct Number of Peaks

In QC analyses, the correct number of peaks occurring at the correct retention times verifies that the correct sample has been injected. After noise peaks and other debris have been removed by the Slope Sensitivity or Minimum Peak Size filter, the remaining peak count should be what is expected.

Coefficient of Variation and Relative Standard Deviation

The scatter of results from repeated analyses of the same sample expressed either as Coefficient of Variation (CV) or Relative Standard Deviation (RSD) reflects product and analysis stability. Product specifications must include measurement uncertainty and this will be the CV or will be derived from it.

Comparison Against Standards

Results may be compared against a standard specification in order to reject or accept a sample if all other aspects of the analysis are satisfactory.

Trend Analysis

Corresponding results from a given analysis can be monitored over long periods to show any trends in product quality, production or analysis conditions. Trend analysis also shows the success, or otherwise, of changes in procedures.

5 Performance Measurement

Accountancy checks on laboratory productivity and instrument utilization are not new but the recent spread of LIMS is making them more commonplace and thorough. Integrators can be linked to corporate mainframes to provide corporate managers with information about performance measurement. Analysts, whose names are associated with these reports, should be aware of this if only to ensure that the information is not interpreted to their disadvantage.

Chromatograph Utilization

The sum of the analysis times gives the daily utilization of a chromatograph. It may be expressed as a fraction or percentage of the working day. Spare capacity can be noted and instruments used more efficiently. When little spare capacity remains it may be time to invest in another chromatograph.

Analyst Workload

When analysts are assigned to work with specific instruments the number of analyses made each day can be counted and compared with other analysts or against historic work performance.

Cost per Analysis

The total number of analytical reports produced in a set period divided by the cost of running the laboratory during that period is the cost per analysis.

The cost/analysis and workload ought to reflect the complexity of analyses which might include extraction, derivatization, *etc.* They are estimates which can be used for comparison with other, similar laboratories.

Tests like these are increasingly used in laboratories as legislation and corporate economics call for greater productivity and analytical accountability. Some of the laboratory reports generated in future will not be routinely available to the analysts involved.

6 Chromatographic Peak Models

Peak shapes are important in chromatography because they are created by the processes that take place in the chromatograph and column. When peak shapes are predictable it is because:

(1) the dynamic mechanisms inside the injector and detector are known;
(2) physical and chemical interactions between solute and column are understood;
(3) the analysis is under control.

Knowledge of column processes brings improvements and credibility to analysis results, to peak resolution and to experiment optimization, and it yields thermodynamic data.

It should be understood at this point that the use of peak models to improve resolution implicitly means time (or X) axis resolution of neighbouring peaks from each other. The separate concept of resolution of a peak from baseline noise is regarded as signal conditioning, or improvement to S/N, the signal-to-noise ratio (Y axis). These two separate ideas of resolution can become blurred.

Benefits of peak models to integrator manufacturers are less tangible. If one peak model could be applied to all peaks it could be used to resolve overlapping groups and yield more peak information, but no such universal model exists. Manufacturers cannot sell an integrator whose algorithms simulate and measure some peaks but not others.

Peak models, however, are the starting point and justification for manual measurements and are used to justify integrator measurements.

The Gaussian Function

If a solute is injected as a very narrow band into an efficient column and partitions linearly between the stationary and mobile phases, the solute distribution inside the column can be represented by a normal or Gaussian distribution curve.[12,13]

The Gaussian function applies to continuous rather than discrete data[14] and is therefore suited to zones of streaming molecules.

The Function

When applied to chromatographic peaks it takes the form,

$$h_t = \frac{A}{\sigma\sqrt{(2\pi)}} \exp\left[\frac{-(t - t_R)^2}{2\sigma^2}\right] \tag{7}$$

where h_t = height of peak at time t
A = peak area
t_R = time of peak max; the retention time
σ = standard deviation of the peak expressed in time units

The curve is bell-shaped and symmetrical about the retention time, t_R (Figure 1.6).

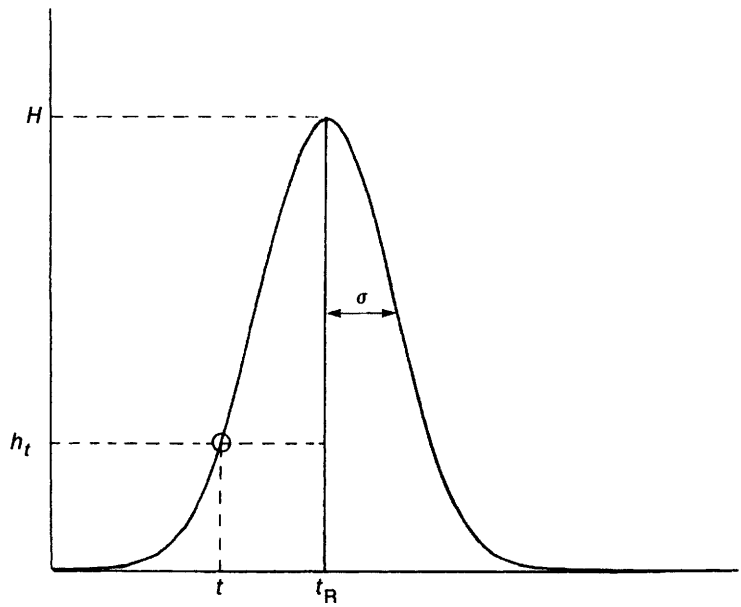


Figure 1.6 *Normal or Gaussian distribution curve*

Peak Height

The first derivative of Equation 7 is given by,

$$h_t' = \frac{-(t - t_R)}{\sigma} \cdot \frac{A}{\sigma\sqrt{(2\pi)}} \exp\left[\frac{-(t - t_R)^2}{2\sigma^2}\right] \tag{8}$$

$$= \frac{-(t - t_R)}{\sigma} h_t \tag{9}$$

where h_t' is zero at $(t - t_R) = 0$, when $t = t_R$.

At $t = t_R$ the peak is at its maximum height H:

$$H = h_{tR} = \frac{A}{\sigma\sqrt{(2\pi)}} \exp\left[\frac{-(t - t_R)^2}{2\sigma^2}\right]$$

$$\therefore H = \frac{A}{\sigma\sqrt{(2\pi)}} \tag{10}$$

$$= 0.3989A/\sigma \tag{11}$$

Equation 11 is the proof that height is a legitimate substitute for area and a measure of solute quantity for Gaussian peaks provided that σ is constant. Over a series of similar analyses, this implicitly requires that peak shape remains constant.

Calculation of Fractional Peak Height. Fractional heights F_h, are given by:

$$F_h = \frac{h}{H} 0.3989 \frac{A}{\sigma} \tag{12}$$

Example. At half height,

$$F_{0.5} = 0.5 \times 0.3989A/\sigma = 0.1995A/\sigma$$

and so on.

Ratio of Area/Height

Some integrators can compute the ratio of measured peak area to peak height. From Equation 10,

$$A/H = \sigma\sqrt{(2\pi)} = 2.5069\sigma \tag{13}$$

which is the Gaussian peak width at 45.6% of peak height. For a single Gaussian peak $(A/H)^2$ is proportional to peak variance from which can be calculated column efficiency. It will vary with those factors which affect peak width and shape, which

is nearly everything, but unless it can be maintained constant, height cannot be used as a substitute for area and column efficiencies will be measured imprecisely.

Calculation of Peak Width at Various Heights

Combining Equations 7 and 10 gives,

$$h_t = H \exp[-(t - t_R)^2]/2\sigma^2 \tag{14}$$

in which the quantity $(t - t_R)$ is equal to half the width of the peak at t.

i.e. $$(t - t_R) = w_t/2 \tag{15}$$

Substituting this into Equation 13 and rearranging gives:

$$w_t = 2\sigma\sqrt{[-2\ln(h_t/H)]} \tag{16}$$

Figure 1.7 summarizes the heights and widths at key peak locations on a Gaussian Peak. The ratio of the base width, $w_b = 6\sigma$, of a Gaussian peak to the width at half height, $w_{0.5}$, is equal to $6\sigma/2.355\sigma = 2.548$.

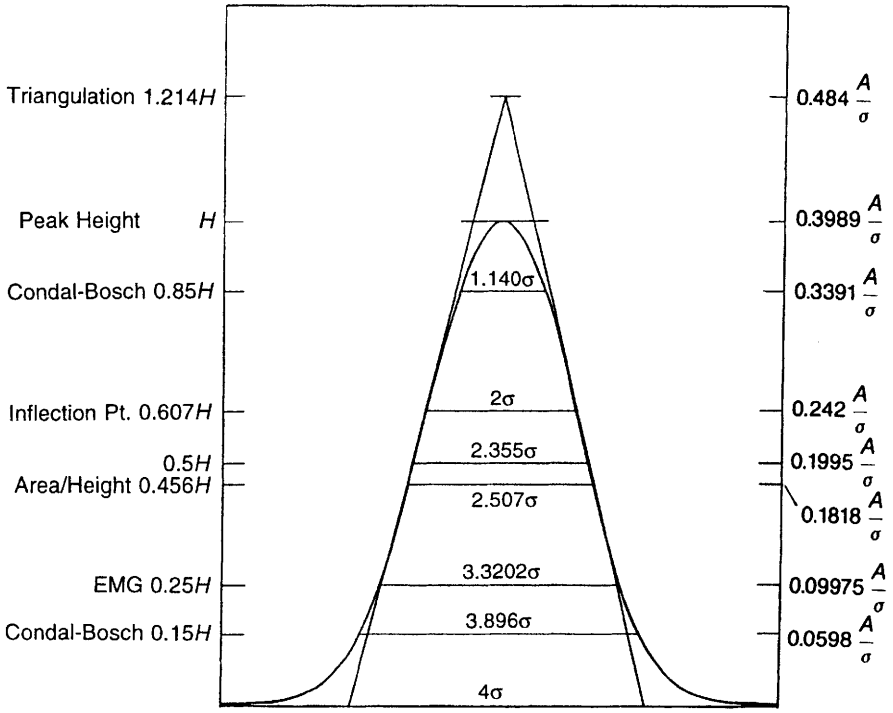


Figure 1.7 *Key dimensions of a Gaussian peak*

Gaussian Peak Shape Tests

If a peak is Gaussian, the following tests can be applied:

(i) the Asymmetry Ratio, $B/A = 1$ at all heights;
(ii) the ratios of peak widths at various heights will be known, *e.g.* at: 50%; 30%; 10% = 0.5487; 0.7231; 1, respectively.

The Points of Inflection

The points of inflection are the points of maximum slope on each side of the peak. It is through these points that tangents are drawn for the 'triangulation' method of peak area measurement.

The points of inflection are located where the second derivative $h_t'' = 0$,

$$h_t'' = \left[\frac{(t - t_R)^2}{\sigma^2} - 1\right] \frac{A}{\sigma\sqrt{(2\pi)}} \exp\left[\frac{-(t - t_R)^2}{2\sigma^2}\right] \tag{17}$$

$$= \left[\frac{(t - t_R)^2}{\sigma^2} - 1\right] h_t \tag{18}$$

and this is zero when:

$$\frac{(t - t_R)^2}{\sigma^2} - 1 = 0$$

i.e. when

$$t - t_R = \pm\sigma \tag{19}$$

Between the points of inflection the peak width is 2σ.

Height of the Peak at the Point of Inflection. This is found by substituting $(t - t_R) = \sigma$ in Equation 7:

$$h_t = \frac{A}{\sigma\sqrt{(2\pi)}} \exp\left[\frac{-\sigma^2}{2\sigma^2}\right]$$

$$= 0.242A/\sigma \tag{20}$$

from Equation 11,

$$= 0.6065H$$

Maximum Peak Slope. The maximum peak slope is determined by making the same substitution, $t - t_R = \sigma = 1$, into Equation 8 for the first derivative, h_t':

$$h_t' = \frac{\pm\sigma}{\sigma} \cdot \frac{A}{\sigma\sqrt{(2\pi)}} \exp\left[\frac{-\sigma^2}{2\sigma^2}\right]$$

$$= \pm 0.242A \tag{21}$$

At the points of inflection, the height and the gradient have the same numerical value.

Fractional Peak Area Bounded by Various Widths

The boundaries of a Gaussian peak are theoretically infinite and it is important to know how much area is lost by adopting finite dimensions compatible with observed peak widths.

The quantity of peak area between two boundaries '$-$X' and '$+$X' can be obtained by integrating Equation 7 between these limits, but it is easier to consult statistical tables[15] which give the fractional peak area from -4σ to $+4\sigma$.

A summary of the commonly quoted values is given in Table 1.1 and shown in Figure 1.8.

Table 1.1 *Area of Gaussian peak bounded by different widths*

σ	*Area %*	σ	*Area %*
0.5	69.15	2.5	99.38
1.0	84.13	3.0	99.87
1.5	93.32	3.5	99.98
2.0	97.72	4.0	100.0

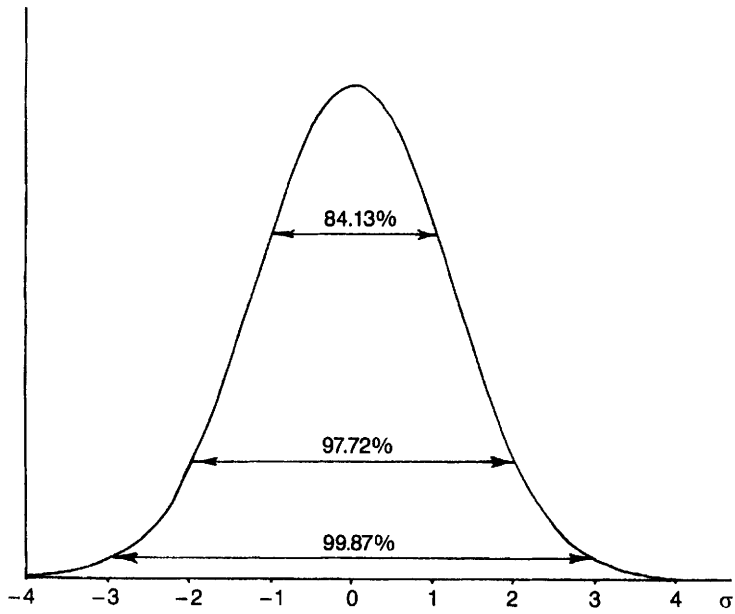


Figure 1.8 *Area of Gaussian peak bounded by different widths*

Loss of Area from the Base of a Gaussian Peak

The base of a peak will not be measured accurately if:

(1) the baseline is noisy making peak detection late, after the peak has risen above the noise;
(2) the baseline signal has drifted off the bottom of the chart recorder and cannot be followed by the pen;
(3) the integrator slope sensitivity parameter is set too large (too insensitive) so that slope is detected late and lost early;
(4) the baseline has drifted below the lower limit of the integrator's operating range, and only that part of the peak rising above it is measured.

Area lost from the base of a peak is lost from the widest part where it matters most (Figure 1.9).

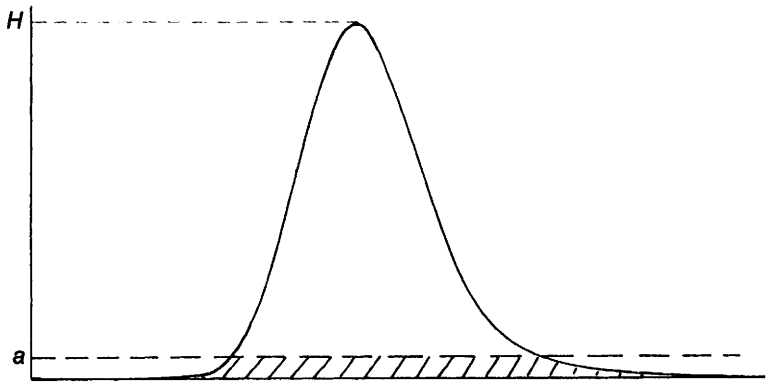


Figure 1.9 *The shaded area is lost from widest part where it matters most*

Littlewood[16] calculated the percentage area loss from the base of a Gaussian peak using Equation 22 (see Table 1.2 and Figure 1.10):

$$\text{area lost} = 200[\text{erfc}\,(\sqrt{(2 \ln H/a)} + a/H.\sqrt{(\pi^{-1} \ln H/a)})] \qquad (22)$$

(erfc = complementary error function)

Gaussian Peak Maximum as an Approximate Parabola

For a Gaussian shaped peak (Figure 1.11):

$$h_t = H \exp[-(t - t_R)^2/2\sigma^2] \qquad \text{from (14)}$$

An exponential function $\exp(x)$ can be written as the series,

$$\exp(x) = 1 + x + x^2/2! + x^3/3! + \ldots + x^n/n! \qquad (23)$$

Table 1.2 *Effect of losing peak base. More area is lost than height*

a/H %	*Area loss %*
0.5	1.42
1.0	2.66
2.0	4.98
3.0	7.10
5.0	11.2

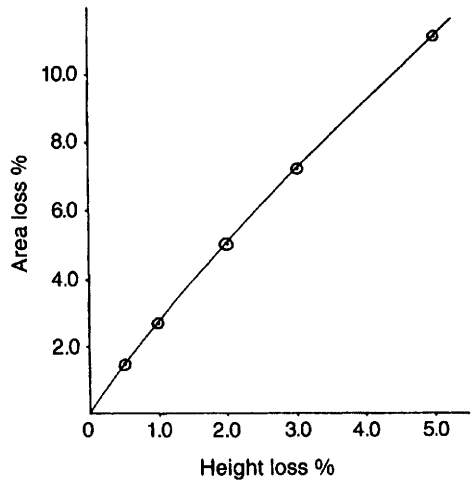


Figure 1.10 *Effect of losing peak base. More area is lost than height*

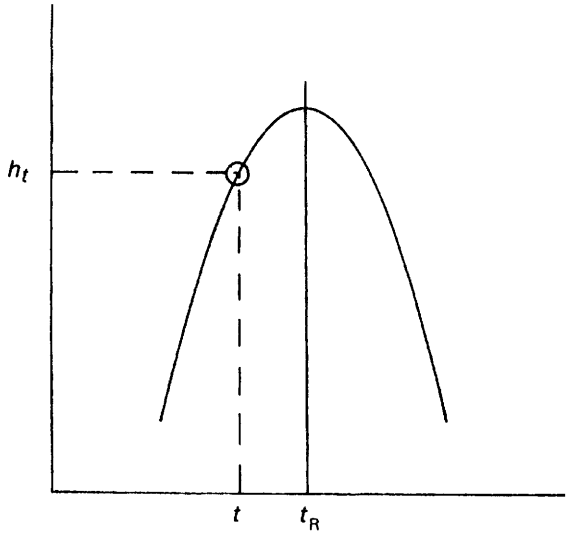


Figure 1.11 *Gaussian peak top is parabolic*

and by comparison with Equation 14,

$$x = -(t - t_R)^2/2\sigma^2 \tag{24}$$

so Equation 23 may be re-written,

$$\exp[-(t - t_R)^2/2\sigma^2] = 1 - (t - t_R)^2/2\sigma^2 + (t - t_R)^4/4 - \ldots + (t - t_R)^{2n}/2\sigma^2\, n! \tag{25}$$

Close to the peak maximum, $(t - t_R) < \sigma$ and Equation 25 rapidly converge allowing terms of the 4th order and higher to be neglected. Approximately:

$$\exp[-(t - t_R)^2/2\sigma^2] = 1 - (t - t_R)^2/2\sigma^2 \quad (26)$$

Substituting Equation 26 into Equation 14 gives,

$$h_t = H[1 - (t - t_R)^2/2\sigma^2] \quad (27)$$

and this has the form of a parabola.

When a more accurate representation over a wider range is required, the peak maximum is fitted to a polynomial (see Chapter 5). This allows more data samples to be included, and is how integrators measure retention time.

The Exponentially Modified Gaussian Function[9,17]

Not many chromatographic peaks are Gaussian and this has led to much effort since 1959 to find a better peak model.[17–23]

The exponentially modified Gaussian, or EMG function, gives good agreement between theory and experiment in many (but not all) real cases. It seems equally good for measuring peak height as well as area[24] when baseline noise is bad.[25] It is best applied to LC peaks, especially isocratic, and capillary GC peaks but not to all packed column GC peaks, in particular from solid phase columns and where there is wide dissimilarity in phase/solute polarity. There is a broad correlation between old and new column manufacturing techniques and applications: the better the quality of manufacture, the more stable the peak shape and the closer the approximation to EMG.

The EMG function is a Gaussian function convoluted (bent) on to an exponential axis of time constant τ (Figure 1.12).

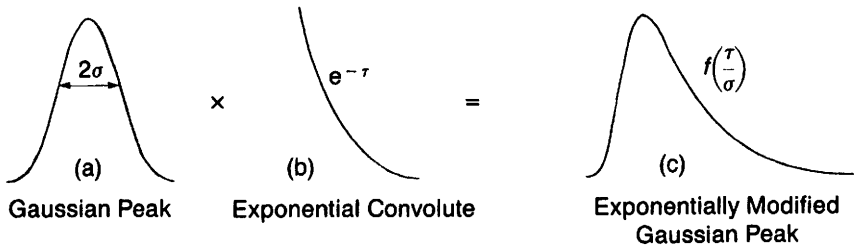


Figure 1.12 *The exponentially modified Gaussian function*

The Function

The equation for the EMG function can be expressed in several ways; a typical example is,

$$h_t = A/\tau \exp\left[(\sigma_G/\tau)^2/2 - (t - t_G)/\tau\right] \int_{-\infty}^{z} (\sqrt{2\pi})^{-1}.\exp(-y^2/2)\,dy \qquad (28)$$

where A = area of EMG peak
t_G = retention time of Gaussian peak
σ_G = standard deviation of Gaussian peak
τ = time constant of exponential axis

$$z = (t - t_G)/\sigma_G - \sigma_G/\tau \qquad (29)$$

and y = dummy variable of integration.

In order to evaluate h_t, the integral term in Equation 28 is replaced in one of two ways:

if $$I(z) = \int_{-\infty}^{z} \sqrt{(2\pi)^{-1}}.\exp(-y^2/2)\,dy$$

(a) $I(z)$ can be replaced by an error function:

$$I(z) = \text{erf}\left(\sqrt{(2\pi)^{-1}}[(t - t_G)/\sigma_G - \sigma_G/\tau]\right) \qquad (30)$$

The value of h_t is usually calculated by computer, and if the error function is or can be made a standard routine of the computer, evaluation of $I(z)$ and Equation 28 is easily carried out. Alternatively:

(b) $I(z)$ can be replaced by the approximations,

$$I(z<0) = \sqrt{(2\pi)^{-1}}.\exp(-z^2/2).\sum_{n=1}^{5} b(n)/(1 + pz)^n \qquad (31)$$

and

$$I(z>0) = 1 - I(z<0) \qquad (32)$$

in Equation 31, z is defined in Equation 29 and:

p = 0.231 6419;

$b(1)$ = 0.319 381 530;

$b(2)$ = −0.356 563 782;

$b(3)$ = 1.781 477 937;

$b(4)$ = −1.821 255 978;

$b(5)$ = 1.330 274 429

A computer program to evaluate Equations 31 or 32 and hence h_t is straightforward. For example, one written in BASIC for a microcomputer has been published.[17]

An alternative equation[21] for the exponentially modified Gaussian function in a form that does not involve an integral term is:

$$h_t/H = \sqrt{(\pi/2)}.\sigma_G/\tau \exp[-\sigma_G/2\tau((t - t_R)/\sigma_G - \sigma_G/\tau)].\text{erf}[\sqrt{2}^{-1}((t - t_R)/\sigma_G - \sigma_G/\tau)] \quad (28a)$$

As a rule of thumb, σ is a measure of on column dispersion and τ is a measure of off column and dead volume effects. The analysts' interest in τ (*i.e.* peak tailing) is as an indicator of poor experimental set-up and a need for more adjustment. However, Naish *et al.*[26] have shown a variation of τ with retention which indicates some involvement with column mechanisms, and it has long been known that perfect peak symmetry is more desirable than achievable,[27] so this rule can be an over-simplification.

Practical Application of the EMG Function

Adoption of the EMG function in analytical laboratories has been held at bay by the complexity of Equation 28. However, useful (and simple) formulae for chromatographic quantities have been derived[7,9] for asymmetric peaks where $\tau/\sigma < 3$, some involving the asymmetry ratio B/A.

These formulae provide 'chromatographic figures of merit' (Foley and Dorsey[7]), and include four alternative area equations:

Peak Area $$A = 0.753 H\, w_{0.25} \quad (33)$$

or, $$= 0.586 H\, w_{0.1}(B/A)^{-0.133} \quad (33a)$$

or, $$= 1.07 H\, w_{0.5}(B/A)^{0.235} \quad (33b)$$

or, $$= 1.64 H\, w_{0.75}(B/A)^{0.717} \quad (33c)$$

Theoretical Plates $$N = t_R^2/\sigma_G^2 + \tau^2$$

$$= 41.7[(t_R/w_{0.1})^2/(B/A + 1.25)] \quad (34)$$

Variance $$\sigma^2 = t_R^2/N = \sigma_G^2 + \tau^2 \quad (35)$$

where H = peak height
$w_{0.25}$ = peak width at 25% of height
$w_{0.75}$ = peak width at 75% of height

$w_{0.5}$ = peak width at 50% of height
$w_{0.1}$ = peak width at 10% of height

and B/A is the asymmetry ratio at 10% height in τ/σ range 1.1 to 2.8.

Where peaks fit the EMG function, Equation 33 is very convenient for manual measurement of area. It does not depend on B/A and is as easy as height $\times$ $w_{0.5}$ to calculate.

The use of these figures of merit, predominantly peak area, variance and plate count, is increasingly reported in the literature. A 1991 review[9] of the exponentially modified Gaussian function since 1983 contains 127 references.

EMG Peak Shape Tests

(i) Peaks should not be symmetrical: at 10% H, the asymmetry ratio, $B/A \neq 1$. A better test is $B/A > 1.1$.[7]

(ii) Peak areas calculated from Equations 33a, b, and c should agree in theory to within about 2% of the area calculated from Equation 33. Larger errors might perhaps be compared with the results from a Gaussian shape test before analysts decide to abandon Equation 33 and revert to Gaussian calculations.

7 Statistical Moments of a Chromatographic Peak[28–30]

Instead of defining asymmetric peak shapes by mathematical models such as Exponentially Modified Gaussian, an alternative approach is to apply measurement techniques that assume no peak shape.

Such a technique is the 'method of statistical moments' which characterizes those distributions that cannot be expressed by known curves. Application of statistical moments to chromatography requires the introduction of the concepts of finite peak boundaries, peak overlap and non-linear baseline which have no meaning in the original context of a distribution function. The results must also be compared with other models to see how they agree and differ.

There is an infinite number of peak moments but only the first five are used in connection with chromatographic peaks. The general formula for them is,

$$m_n = \frac{\int_{-\infty}^{+\infty} t^n . h_t \, dt}{\int_{-\infty}^{+\infty} h_t \, dt} \tag{36}$$

which is normalized to the zeroeth moment ($n = 0$).

Zeroeth Moment, m_0

This is the peak area. In Equation 36, $n = 0$:

$$m_0 = \frac{\int_{-\infty}^{+\infty} t^0 . h_t \, dt}{\int_{-\infty}^{+\infty} h_t \, dt} = 1 \text{ (since } t^0 = 1) \tag{37}$$

Equation 37 reflects the common practice of statisticians to normalize population totals to 1 or 100%. In chromatography the non-normalized moment is more useful:

$$m_0 = \int_{-\infty}^{+\infty} t^0 . h_t \, dt = \text{Peak Area, } A \tag{38}$$

First Moment, m_1

The first moment is the 'mean retention time', or retention time measured at the centre of gravity of the peak. It is different from the chromatographic retention time measured at the peak maximum unless the peak is symmetrical (Figure 1.13).

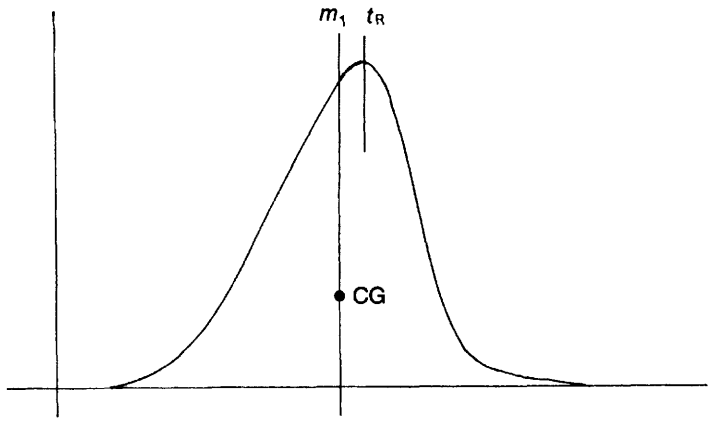


Figure 1.13 *First peak moment is the 'retention time' of peak centre of gravity*

$$m_1 = \frac{\int_{-\infty}^{+\infty} t^1 . h_t \, dt}{\int_{-\infty}^{+\infty} h_t \, dt} \tag{39}$$

in non-normalized form,

$$m_1 = \frac{1}{m_0}\int_{-\infty}^{+\infty} t^1.h_t\,dt \equiv t_R + \tau \tag{40}$$

where t_R = retention time of a Gaussian peak and τ = EMG time constant. In chromatography, m_1 is referenced to the peak retention time t_R, in which case Equation 40 modifies to:

$$m_1 = \frac{1}{m_0}\int_{-\infty}^{+\infty} (t - t_R)\,h_t\,dt \tag{41}$$

Second Moment, m_2

The second moment is the peak variance σ^2, where σ is the peak standard deviation.

$$m_2 = \frac{1}{m_0}\int_{-\infty}^{+\infty} t^2.h_t\,dt \equiv \sigma_G^2 + \tau^2 \tag{42}$$

The variance of a chromatographic peak σ^2, is a measure of lateral spreading. It is the sum of the variances contributed by different parts of the instrument system. Generally:

$$\sigma^2_{total} = \sum_i \sigma_i^2 \tag{43}$$

The chromatographic interpretation of which is:

$$\sigma^2_{peak} = \sigma^2_{inj} + \sigma^2_{col} + \sigma^2_{det} + \sigma^2_{etc.} \tag{44}$$

A comparison[4] of column efficiencies measured using peak variance calculated from equations (3), (34) and (42) concluded that the moment method, (42), gave the best results in almost all instances of noisy or skewed peaks.

Third Moment, m_3

The third moment describes vertical asymmetry, or skew. It is a measure of the departure of the peak shape from the Gaussian standard.

In normalized form:

$$m_3 = \frac{1}{m_0}\int_{-\infty}^{+\infty} t^3.h_t\,dt \equiv 2\tau^3 \tag{45}$$

Skew is sometimes expressed[1,20,17] as the dimensionless quantity γ:

$$\text{Skew, } \gamma = \frac{m_3}{m_2^{3/2}} = \frac{2\left(\dfrac{\tau}{\sigma_G}\right)^3}{\left[1 + \left(\dfrac{\tau}{\sigma_G}\right)^2\right]^{3/2}} \tag{46}$$

A symmetrical peak has a skew of zero. Peaks which tail have positive skew and their first moment is greater than the peak retention time (Figure 1.14). Peaks which front have negative skew; their first moment is less than the retention time.

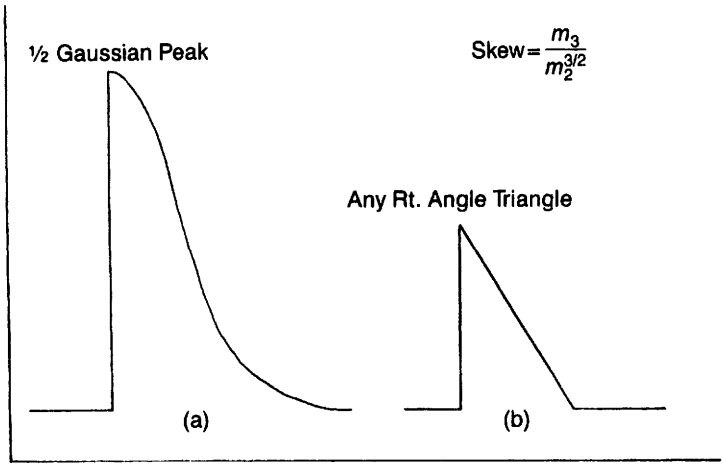


Figure 1.14 *Examples of skew:* (a) $\gamma = 0.995$; (b) $\gamma = 1.131$
(Using data from *J. Chem. Phys.*, 1963, **38**, 437)

Fourth Moment, m_4

The fourth moment or 'excess', is a measure of the compression or stretching of the peak along a vertical axis, and how this compares with a Gaussian standard for which $m_4 = 0$. It can be visualized by moving in or pulling apart the sides of a Gaussian peak while maintaining constant area.

$$m_4 = \frac{1}{m_0}\int_{-\infty}^{+\infty} t^4.h_t\,dt \equiv 3\sigma^4 + 6\sigma^2\tau^2 + 9\tau^4 \tag{47}$$

If the peak is compressed or squashed down in comparison (*i.e.* if the detector overloads) its excess is negative (Figure 1.15). If it is taller, for example, if the signal-to-noise ratio has been improved, its excess is positive.

Excess (E), also called kurtosis,[31] can be expressed in the dimensionless form:[18]

$$E = m_4/m_2^2 - 3 \tag{48}$$

For Gaussian peaks $m_4/m_2^2 = 3$ (mesokurtic) and $E = 0$; for overloaded peaks, $m_4/m_2^2 < 3$ (platykurtic) and $E < 0$.

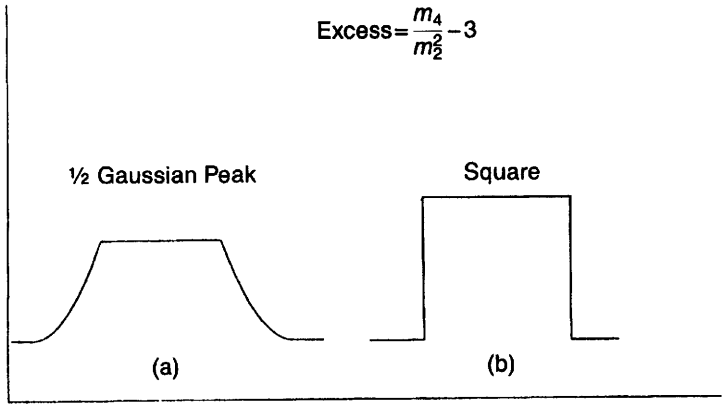


Figure 1.15 *Examples of excess:* (a) 1/2 *Gaussian peak*, E = −1.377, (*leptokurtic*); (b) *square peak*, E = −1.666; (*platykurtic*)
(Using data from *J. Chem. Phys.*, 1963, **38**, 437)

Higher Odd Moments

All higher odd moments are measures of vertical asymmetry, like skew, and for symmetrical peaks they are all zero.

Higher Even Moments

Higher even moments have no similar easy interpretation except that like the second and fourth moments they are measures of lateral dispersion.

Measurement of Peak Moments

Peak moments lend themselves very well to measurement by computer (Figure 1.16). Moments are calculated from measurements of h_i made at regular time intervals over the whole peak, which is exactly what integrators do.

Integrator measurement of peak area is the same as measurement of the zeroeth moment. Modal retention time of the peak maximum is measured in preference to the first moment, and skew, if it is measured at all, is measured as the asymmetry ratio B/A. Integrators only measure peak width and not yet excess.

Should these quantities ever be routinely wanted by analysts, integrators can provide them by simple additions to their software, and some already do (*e.g.* HP chemStation).

Practical Disadvantages and Uses

Moments are notoriously susceptible to noise.[32–34] With the exception of the zeroeth moment, S/N ratio must be better than 100:1[35] for the measurements to be meaningful. They are severely affected by baseline drift, inaccurate determination of the peak limits, tailing, incomplete resolution, and insufficient sampling frequency.

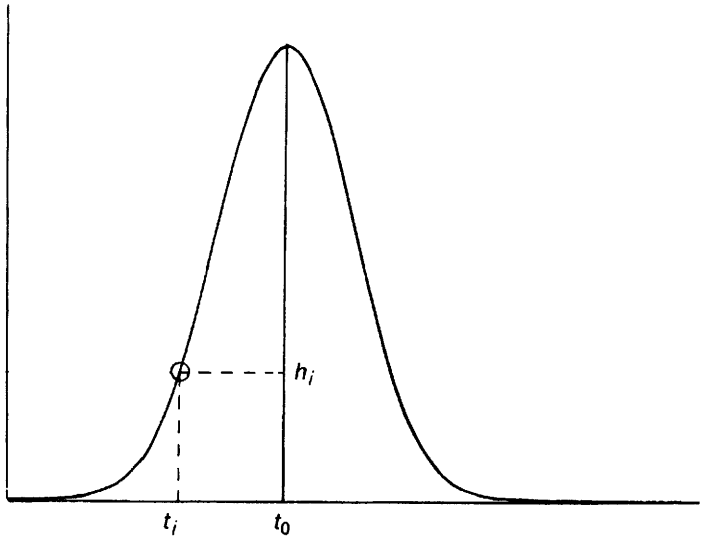


Figure 1.16 *Measurement of peak moments. Position of reference time* t_0 *is arbitrary*

If peak moments higher than zero have a future use for the analyst, it would seem to be as diagnostics in high quality chromatography. They are too sensitive for routine use. They have found no application in peak detection.

The mean retention can be compared with the chromatographic retention to measure asymmetry. Variance can be used to estimate system stability, monitor column suitability, performance and deterioration.

Skew shows column suitability also. If too many peaks are skewed, an alternative column might be a better choice. Excessive or increasing skew is an indicator of column deterioration or of system errors such as column overload or adsorption of solutes in the injector.

Excess would indicate detector overloading or saturation.

8 Manual Peak Area Measurement

Height × Width at Half Height

The area of a Gaussian peak can be calculated from the product of its height and width at half height, which is obtained by substituting $h_i/\mathrm{H} = 0.5$ into Equation 16:

$$\text{Area} = H\, w_{0.5} \tag{49}$$

$$w_{0.5} = 2\sigma\sqrt{[-2\ln(0.5)]}$$

$$= 2.3548\sigma \tag{50}$$

$$\text{Area} = 0.3989A/\sigma \times 2.3548\sigma$$

$$= 0.939A \tag{51}$$

i.e. the computed area is 93.9% of the theoretically true area, from which $k = 1/0.939$ in Figure 1.17.

While the width at half height is the most commonly used width for area measurement, it is not the most accurate. For a Gaussian peak, the optimum width to measure for best accuracy[36] is the width at 36.8% of peak height ($h_{opt} = e^1 H$).

All manual measurements of peak area computed from height × width measurements suffer when baseline noise is excessive because drawing the baseline is then subject to greater positional error and this introduces errors directly into the measurement of fractional height and the width at that height.

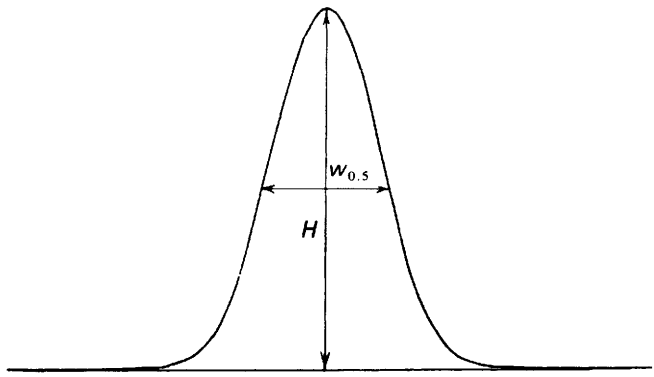


Figure 1.17 *Manual measurement of a Gaussian peak. Area* $= kw_{0.5}H$. *For a Gaussian peak* $k = 1/0.939$

The Condal-Bosch Area[37]

Instead of the width at half height, Condal-Bosch used the average of the peak widths at 15% and 85% of the peak height,

$$\text{Area} = 0.5H[w_{0.15} + w_{0.85}] \tag{52}$$

$$w_{0.15} = 2\sigma\sqrt{[-2\ln(0.15)]} = 3.896\sigma$$

$$w_{0.85} = 2\sigma\sqrt{[-2\ln(0.85)]} = 1.140\sigma$$

$$\text{PeakArea} = 0.3989\text{A}/\sigma \times (3.896\sigma + 1.140\sigma)/2$$

$$= 1.004A \tag{53}$$

which is very close to the true peak area (see Figure 1.18).

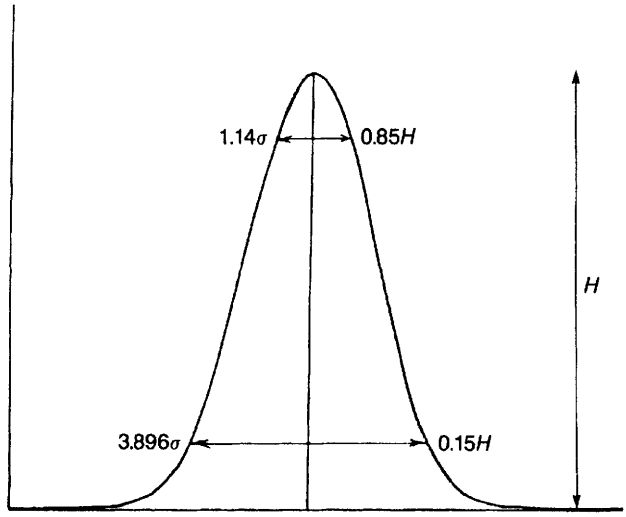


Figure 1.18 *The Condal-Bosch area*

Peak Area by Triangulation

The triangle was probably the first peak model to be used. A Gaussian peak can be related to the triangle formed when tangents drawn through the points of inflection intersect the peak baseline and each other (Figure 1.19).

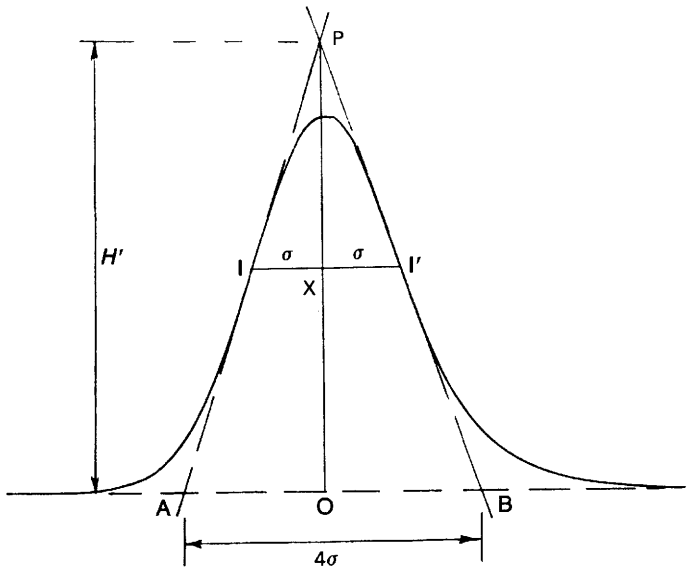


Figure 1.19 *Peak area by triangulation*

At the point of inflection I, the gradient of the peak (and therefore of the side of the triangle APB) is $\pm 0.242A$. The height of I is $0.242A/\sigma$, and the half width of the peak at I is σ (see Equations 19, 20 and 21).

In Figure 1.19, $OX = 0.242A/\sigma$ from (20)

gradient $PX/\sigma = 0.242A/\sigma$ from (21)

at I, $\sigma = 1$, $PX = OX$,

$PO = 2OX = 0.484A/\sigma = H'$

and $AO = 2\sigma = OB$

Method (1)

The peak area calculated from the area of the exterior triangle, $OP \times AO$, is,

$$\text{Area} = 0.484A/\sigma \times 2\sigma \tag{54}$$

$$= 0.968A \tag{55}$$

or 96.8% of the true peak area.

Method (2)

It is implicit in triangulation that the inflection point tangents, PA and PB, converge at an angle large enough to allow precise measurement of the intersect. If this is not the case it will be easier and more precise to measure the peak height.

Area from triangular base width and peak height, $H.AO$, is:

$$\text{Area} = 0.3989A/\sigma \times 2\sigma = 0.7978A \tag{55a}$$

The measurement of area by Equation 55a should be more precise than by method (1), although less accurate.

Manual Measurement of Asymmetric (EMG) Peaks

(1) Foley's equation (see Equation 33) for the area of a resolved exponentially modified Gaussian peak requires no overt measurements of asymmetry such as B/A yet covers a range of asymmetry up to $\tau/\sigma = 3$ (Figure 1.20). Note, it also fits a symmetrical peak:

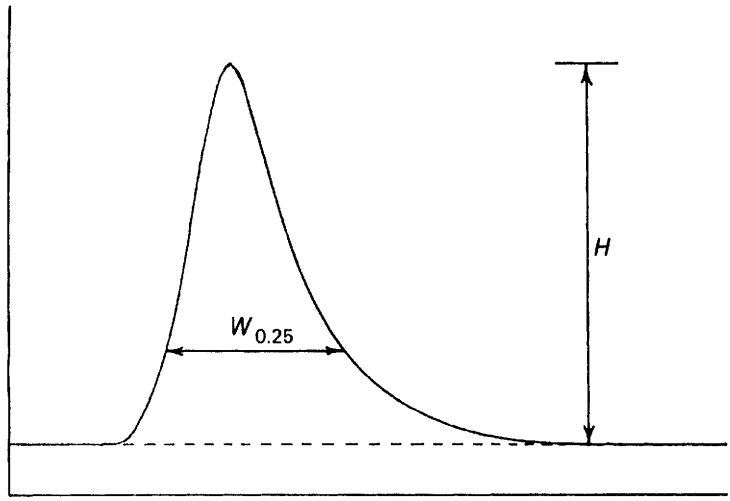


Figure 1.20 *Foley (EMG) measurement of peak area*

$$\text{Area} = 0.753 H\, w_{0.25} \tag{33}$$

if Gaussian, $$= 0.753 \times 0.3989 A/\sigma \times w_{0.25}$$

From Equation 16, $$w_{0.25} = 2\sigma\sqrt{[-2\ln(0.25)]} = 3.3302\sigma$$

$$\text{Area} = 1.0003 A \tag{56}$$

This is even closer to the true area than the Condal-Bosch calculation, requires fewer measurements, and suffers less from overlap. It is more accurate for the measure of symmetrical peaks than the more commonly used height × width at half height. Wu *et al.*,[38] in studying why some peaks were not EMG-shaped, provided a theoretical proof of Equation 33.

(2) When peak overlap prevents the accurate measurement of $w_{0.25}$, an alternative measure of EMG peak area can be made from the width at 75% of peak height, $w_{0.75}$, the peak height and the asymmetry ratio, B/A at that height:[8]

$$A = 1.64 H\, w_{0.75}\, (B/A)^{0.717} \tag{33c}$$

The accuracy is within 4% for computer simulated overlapping peaks where the valley rises to 0.45 H of the smaller peak. Equation 33c applies to the measurement of either peak for asymmetries up to $\tau/\sigma < 3$.

Applied to symmetrical peaks, $B/A = 1$, $w_{0.75} = 1.517\sigma$, and the measured area is 99.2% of the true area.

The EMG function is largely used for the study of variance, skew, and column

plate measure and the errors of these measurements have ranged up to 30%.[39] As a simple manual measure of skewed peak area it has been found to be experimentally accurate to within 10%.[8,40]

9 References

1. J.J. Kirkland, W.W. Yau, H.J. Stoklosa, and C.H. Dilks Jr., *J. Chromatogr. Sci.*, 1977, **15**, 303.
2. K. Pearson, *Methods of Statistical Analysis*, John Wiley and Sons, New York, 2nd edn., 1952, p. 33.
3. N.A. Wright, D.C. Villalanti and M.F. Burke, *Anal. Chem.*,1982, **54**, 1735.
4. J.V.H. Schudel and G. Guiochon, *J. Chromatogr.* 1988, **457**, 1.
5. A.B. Littlewood, *Gas Chromatography Principles, Techniques and Applications*, Academic Press Inc., New York, 2nd edn., 1970.
6. K. Grob, *Anal. Chem.*, 1994, **66**, 1009A.
7. J.P. Foley and J.G. Dorsey, *Anal. Chem.*, 1983, **55**, 730.
8. J.P. Foley, *Anal. Chem.*, 1987, **59**, 1984.
9. M.S. Jeansonne and J.P. Foley, *J. Chromatogr. Sci.*, 1991, **29**, 258.
10. US Pharmacopoeia, XXI, *Chromatography/Physical Tests*, p. 1230.
11. D.J. Anderson and R.W. Walters, *J. Chromatogr. Sci.*, 1984, **22**, 353.
12. E. Kreyszig, *Advanced Engineering Mathematics*, John Wiley and Sons, New York, 7th edn., 1993.
13. A. Klinkenberg and F. Sjenitzer, *Chem. Eng. Sci.*, 1956, **5**, 258.
14. C.J. Brookes, I.G. Batteley and S.M. Loxton, *Fundamentals of Mathematics and Statistics*, John Wiley and Sons, New York, 1979.
15. CRC, *Standard Mathematical Tables*, CRC Press Inc., Boca Raton, FL, 26th edn., 1992.
16. A.B. Littlewood, *Z. Anal. Chem.*, 1968, **236**, 39.
17. J.P. Foley and J.G. Dorsey, *J. Chromatogr. Sci.*, 1984, **22**, 40.
18. J.C. Sternberg, *Advances in Chromatography*, Marcel Dekker, New York, 1966, vol. 2.
19. P.T. Kissinger, L.J. Felice, D.J. Miner, C.R. Reddy and R.E. Shoup, *Contemporary Topics in Analytical and Clinical Chemistry*, Plenum Press, New York, 1978, vol. 2.
20. V. Maynard and E. Grushka, *Anal. Chem.*, 1972, **44**, 1427.
21. R.E. Pauls and L.B. Rogers, *Anal. Chem.*, 1977, **49**, 625.
22. E. Grushka, *Anal. Chem.*, 1972, **44**, 1733.
23. R.E. Pauls and L.B. Rogers, *Sep. Sci. Technol.*, 1977, **12**, 395.
24. W.A. Garland, T. Crews, S.Y. Brown and E.K. Fukuda, *J. Chromatogr.*, 1989, **472**, 250.
25. W.A. Garland, T. Crews and E.K. Fukuda, *J. Chromatogr.*, 1991, **539**, 133.
26. P.J. Naish and S. Hartwell, *Chromatographia*, 1988, **26**, 285.
27. J.C. Giddings, *Dynamics in Chromatography*, Marcel Dekker, New York, 1966, vol. 1.
28. D.A. McQuarrie, *J. Chem. Phys.*, 1963, **38**, 437.
29. O. Grubner, *Advances in Chromatography*, Marcel Dekker, New York, 1968, vol. 6.
30. O. Grubner, A. Zikanova and M. Ralik, *J. Chromatogr.*, 1967, **28**, 209.
31. E.J. Borowski and J.M. Borwein, *Reference Dictionary of Mathematics*, Collins, London, 1989.
32. E. Grushka, M.N. Myers, P.D. Schettler and J.C. Giddings, *Anal. Chem.*, 1969, **41**, 889.
33. S.N. Chesler and S.P. Cram, *Anal. Chem.*, 1971, **43**, 1922.
34. W.W. Yau, *Anal. Chem.*, 1977, **49**, 395.
35. T. Petticlerc and G. Guiochon, *J. Chromatogr. Sci.*, 1976, **41**, 531.

36. A.S. Said, *Theory and Mathematics of Chromatography*, Alfred Hüthig, Heidelberg, 1981, p. 31.
37. L. Condal-Bosch, *J. Chem. Educ.*, 1964, **41**, A235.
38. N.S. Wu, A.M. Qui and G.W. Zhao, *Chromatographia*, 1990, **29**, 248.
39. J. Doehl and T. Greibokk, *J. Chromatogr. Sci.*, 1987, **25**, 99.
40. S.D. Frans, M.L. McConnell, and J.M. Harris, *Anal. Chem.*, 1985, **57**, 1552.

CHAPTER 2

Errors in Peak Area Measurement

1 Accuracy and Precision

Modern chromatographs using stable, bonded stationary phases and automatic injection can produce a set of closely similar results from the repeated analysis of a particular sample. The relative standard deviation or coefficient of variation of these results may be very small (and therefore pleasing), while the results themselves are quite different to known answers.

This stability highlights the difference between precision and accuracy.

Accuracy

Accuracy is a measure of how close the experimental result is to the 'true result'. The difference between the 'true' and experimental result is called the bias.

Precision

Precision is a measure of how close the results from repeated experiments are to each other. A graphical interpretation of precision and accuracy is given in Figure 2.1. Because of the quality of modern manufacturing techniques, chromatographic instrumentation can achieve a precision which analysts may misinterpret as accuracy. The same result from three consecutive analyses does not mean that it is the correct result.

Method development must demonstrate accuracy as well as precision and it will do so by validation and calibration.

Inaccuracy and Uncertainty

Inaccuracy, or bias, is the difference between a single measurement and its true value. A repeated measurement may have a slightly different value and a slightly different bias. Many repeated measurements of the same quantity will have a range of values and a range of associated errors. Uncertainty is often used synonymously with inaccuracy but has been recently defined,[1] and used here, as the range of

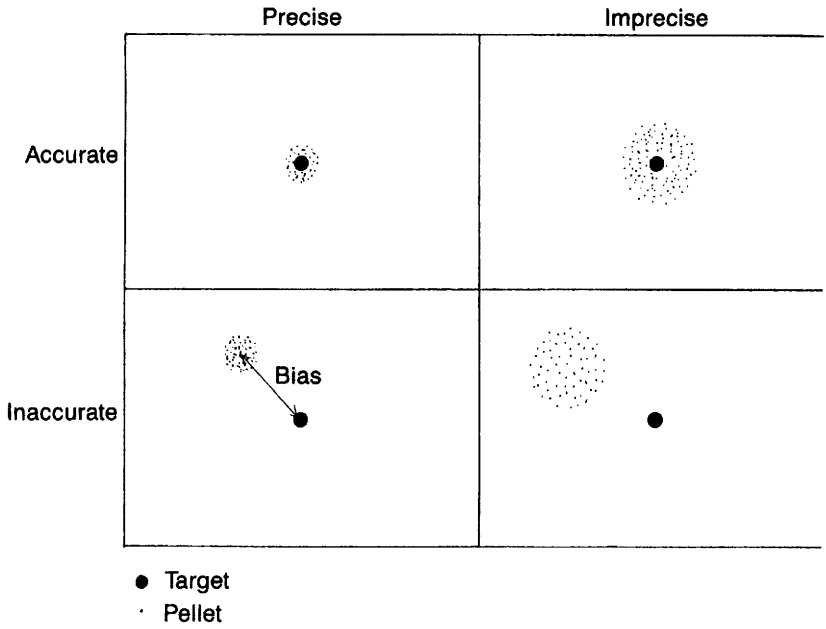


Figure 2.1 *Accuracy and precision*

measured values which includes the true value. The boundaries of this range are known as the Confidence Limits.[2]

Repeatability

Repeatability is the closeness of results between consecutive analyses carried out on the same chromatograph by the same operator working in constant conditions. It is therefore the same as precision. Published data on comparisons of the precision of various forms of peak measurement usually imply repeatability.

Reproducibility

Reproducibility is the closeness between results from analyses of the same solution by different people working with different chromatographs.

Reproducibility is also a measure of precision but it is rarely as good (as small) as the precision of repeatability. The difference between these two quantities is important when transferring methods between labs. There should be no difference at all, but where it does occur it is likely to be caused by different work styles of analysts and different suppliers of instruments and consumables.

Random Errors

These are unpredictable or indeterminate in nature and affect the precision of analysis. Coefficients of variation are a measure of random error. Statistical methods of signal improvement are used to remove random error.

Systematic Errors

Systematic errors affect the accuracy of the result. They can be fixed in size (offsets) or vary with the amount of sample analysed. Unless the true experimental result is known, or the behaviour of the systematic error is predictable, the 'right answer' and a measure of this error cannot be obtained from the experiment. Systematic error is often recognized only after a measurement is made by two independent experiments which provide different results, and the difference is too great to be explained by random experimental error.

Spurious Errors

Spurious errors are random in occurrence but systematic in cause: *e.g.* bubbles in a detector cell, result transcription errors. Results containing spurious errors are detected by outlier tests; results may have to be discarded if the error cannot be separated out.

Causes of Imprecision

Imprecision (repeatability) means that measurements differ from one analysis to the next. Reasons for this are attributed to lack of control of experiments: *i.e.* to imperfect sample preparation and injection technique, solute adsorption or decomposition, solvent leaks, columns which need conditioning, instrument instability and, generally, the kind of problems which analysts feel they should be able to solve.

Causes of Inaccuracy

There are only two causes of inaccuracy:

(1) the detector signal does not accurately represent the solute profile (it is assumed that the sample represents the original solution);
(2) peaks are measured according to incorrect rules. For example, when asymmetric peaks are measured by methods derived from Gaussian theory.

Precision of Integrator Measurements

The repeatability of chromatographic measurements where repeatability is important is typically in the range of 0.1%–0.5%. This is almost entirely due to the

variations in the chromatography. If an integrator is provided with a stable enough signal to measure, then measurement precision is very good and variations are small. For example, measurement of the areas of five computer-generated Gaussian peaks[3] gave the results shown in Table 2.1. The absolute areas are eight digit numbers of which the six most significant are identical. The relative standard deviation is 4.3×10^{-4}%. Height is not so good, RSD = 2.0×10^{-3}%.

When asymmetry is present and variable, precision falls a little further,[3] especially for height (see Table 2.2).

This level of precision is not hard to achieve when a purely electronic device measures a stable signal but it is beyond the capabilities of existing chromatographs to deliver this stable a signal. The figures should give assurance that variations in analysis results are mostly caused by instability in the analysis; used correctly, integrators contribute negligible imprecision to the results. Improvements to results will come from tightening control of the chromatograph and analysis method – and the integrator will accurately report the improvements.

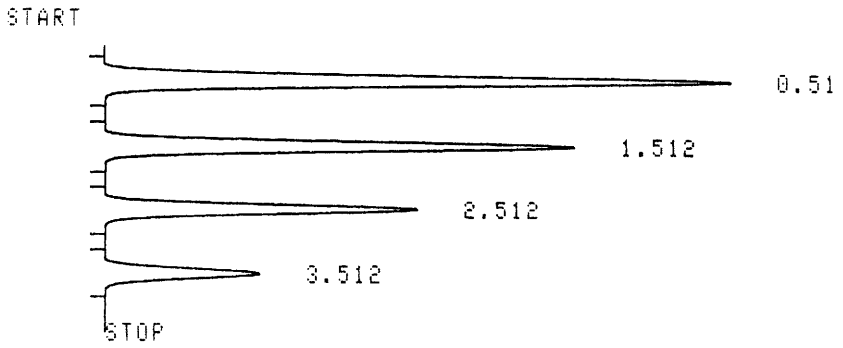
Table 2.1 *How good is an integrator?*
(Data reproduced with kind permission of ISC, see ref. 3)

Pk. no.	*Time*	*Area*	*Height*	*MK*	*Conc.*
1	1.252	13 332 980	834 593	Ref.	
2	2.997	13 332 983	834 614		1.0000
3	4.251	13 332 989	834 476		1.0000
4	5.5	13 332 932	834 380		1.0000
5	7.249	13 332 917	834 507		1.0000
	Total	66 664 792	4 172 569		4.0000

Table 2.2 *Asymmetry and measurement precision*
(Data reproduced with kind permission of ISC, see ref. 3)

Asymmetry	B/A = 5	B/A = 4	B/A = 3	B/A = 2	B/A = 1
No. of peaks	5	5	5	5	5
Average Area	8 904 419	8 905 398	8 905 593	8 905 860	8 906 099
RSD%	4.2×10^{-3}	6.4×10^{-3}	9.4×10^{-3}	8.2×10^{-3}	5.8×10^{-3}
Average Ht.	664 792	677 895	706 655	742 839	837 066
RSD%	5.60×10^{-2}	3.64×10^{-2}	3.35×10^{-2}	3.30×10^{-2}	2.13×10^{-2}
RSD (Ht /Area)	13.26	5.63	3.57	4.03	3.67

Figure 2.2 shows a simple computer-generated chromatogram of four Gaussian peaks with peak areas in the ratio of 40:30:20:10. It is measured effortlessly by a single channel integrator and the predicted ratios are shown in the CONC results column.



CHROMATOPAC C-R6A FILE 0
SAMPLE NO 0 METHOD 21
REPORT NO 2

PKNO	TIME	AREA	MK	IDNO	CONC	NAME
1	0.51	969226			39.9939	
2	1.512	727045			30.0005	
3	2.512	484787			20.0041	
4	3.512	242380			10.0015	
	TOTAL	2423437			100	

Figure 2.2 *Integrators measure stable signals easily and extremely well*

2 Accurate Representation of the Solute Profile

For measurements to have value, chromatograms must be an accurate representation of the injected analyte. Solute quantities can be measured accurately only if:

(a) solutes are eluted intact and are detected. No chemical or physical changes take place between sampling and detection. Peaks from earlier analyses are not included;
(b) the detector responds proportionally to all required solutes;
(c) solute peaks are fully-resolved on a flat and noise-free baseline;
(d) experimental parameters such as mobile phase flow rate, detector temperatures, *etc.* are constant;
(e) the detector and integrator are working correctly within their operating ranges;
(f) peaks and response factors are measured accurately.

Only two of these factors, (e) and (f), concern the integrator; the others determine the accuracy and precision of an analysis before the detector signal reaches the integrator. Data processors cannot compensate for poor chromatography – but, they *will* measure it and print plausible 'results' without a single caution to the

analyst. If these factors are allowed to persist, the detector signal does not accurately represent the profile of the eluting solute zone and the sample cannot be assayed. What the analyst sees and what the integrator measures is not what was injected on to the column.

Additivity of Signals

The detector signal is the linear combination, *i.e.* arithmetic sum, of the component signals from all detected species passing through the detector cell.

When unresolved groups of solutes pass through the detector, the output signal (S) is the sum of the component signals from each detected solute, and from stationary phase and mobile phase too if these produce a measurable response (Figure 2.3),

$$S = \sum_{0}^{n} r_i c_i \tag{1}$$

where n is the number of components contributing to the signal, r_i is the response of unit component of the solute i, and c_i is the concentration or amount of solute.

Peaks do not overlap in the sense that one covers or hides another: they sit on top of each other so that the whole signal is always visible. This does not mean that every peak can be seen – the change from packed columns to WCOTs showed how peaks can be concealed by each other; there is a limit of detection associated with resolution as well as peak size.

The absence of distinguishing features such as peak maxima has resulted in the failure of past efforts to separate overlapping peaks mathematically.

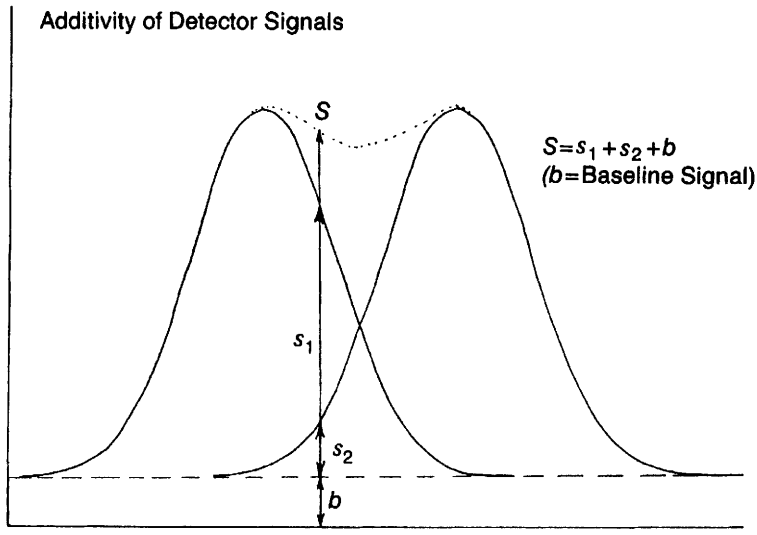


Figure 2.3 *Total signal is the sum of component signals*

Negative Detector Signals

Peak areas are reduced, or compressed,[4] by interference from overlapping negative disturbances. These disturbances are caused, for example, by valve switching, matrix peaks and bi-directional detectors such as TCD or FID where overlapping peaks create opposing detector responses which cancel each other.

Opposing peak signals must be separated in time from each other, otherwise they cannot be measured. Equation 1 means that two co-eluting peaks, equal in size and shape but opposite in signal polarity, will add up to a null response; there will be nothing for the integrator to measure.

At other times a negative baseline disturbance occurring close to a peak will ruin baseline assignment under that peak resulting in a gross under- or over-estimation of the area.

Figure 2.4(a) shows a chromatogram perturbed by negative peaks from a previous analysis.[5] The integrator constructs the baseline shown in (b) and drops perpendiculars to resolve the areas. Area measurement, especially of the smaller peaks, is inaccurate but the situation can be retrieved, as in (c), by using Integrate Inhibit during the disturbances or Forcing Baseline (see Chapter 5) before the first peak of interest. Such unforeseen events should not occur in a properly validated analysis, but, occasionally, they do.

Negative baseline disturbances are probably the worst cause of inaccurate peak measurement. If they are present in a chromatogram they must be well-separated from peaks of interest and the analyst must confirm correct baseline placement under those peaks.

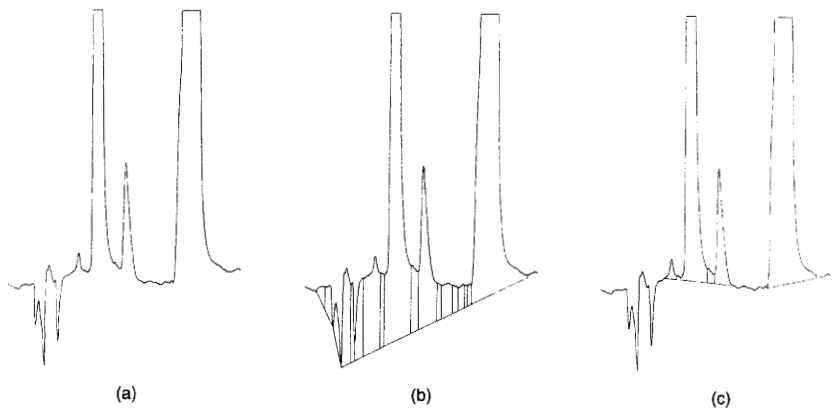


Figure 2.4 *Baseline disturbances ruin peak measurement unless actively prevented* (Reproduced with kind permission from Elsevier Science BV, see ref. 5)

Electronic Peak Distortion

The peak which appears on a strip chart recorder has passed through the recorder's electronics and been drawn by a pen of finite inertia. The peak drawn by an

integrator has been digitized, smoothed, reconstructed through a D/A converter and finally plotted. These operations distort the original solute profile.

Figure 2.5(a) shows an FID peak recorded by an oscilloscope. Figure 2.5(b) shows the same peak after it has passed through an electrometer and integrator and been filtered for noise by both. The difference in peak widths and absence of noise are obvious: the integrator peak is about 30× wider and therefore, for equal area, the height is about 30× lower. Noise has been removed by a combination of capacitive filters and software soothing in the detector electronics and the integrator, but its removal has been at the expense of peak shape, especially height.

The cosmetic appearance of the gas chromatographic peak was largely established in the 1960s by the electronics of that time. The peaks looked like Figure 2.5(b) instead of Figure 2.5(a) because the detectors and electronics represented them in that way. Modern electronics are much faster and are able to represent the peak accurately as in Figure 2.5(a), but no analyst would buy such a chromatograph, so manufacturers make chromatographs and integrators to filter noise and show peaks as icons.

Distortion by capacitive (RC) filters skews peak shape. Peaks are broadened (and this may be misinterpreted as loss of column performance), heights are reduced and the reduction may be, to some extent, a function of retention time for fast peaks if peak width increases (*i.e.* peak 'frequency' decreases) with retention. For example, using the latest WCOT columns with an old GC can result in the elution of early peaks which are narrower than the manufacturer allowed for when the amplifier and detector of the GC were designed, with the result that they are filtered

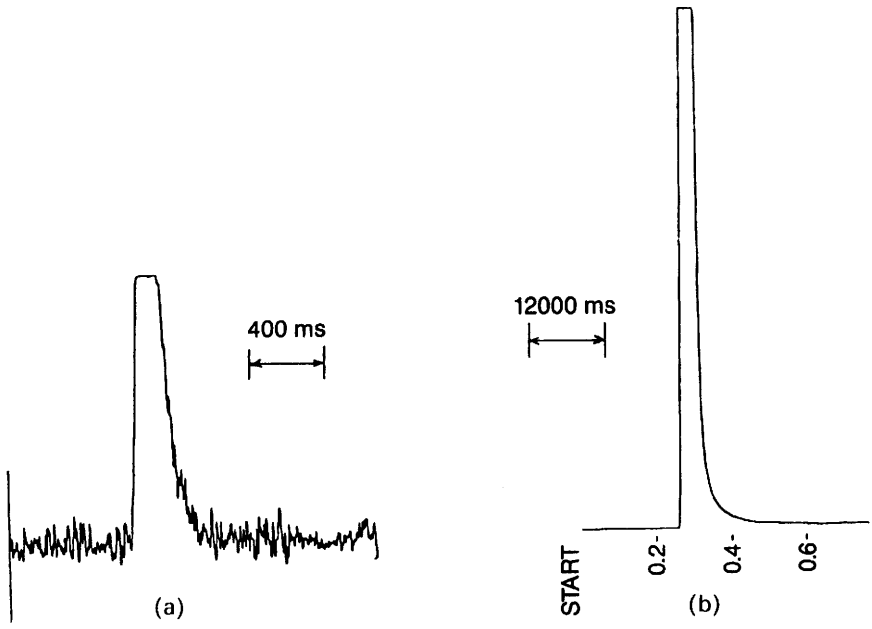


Figure 2.5 (a) *Oscilloscope trace of FID Peak;* (b) *integrator plot of same peak*

or attenuated as noise in a way that is different to later, broader peaks which are more compatible with the time constant of the system.

The redeeming feature is that peak area is not changed by RC filtering and remains a measure of solute quantity. In consequence, when very narrow peaks are being measured, integrators should be primarily regarded as area measuring devices.

UV detectors for LC are less noisy than FIDs. Their time constants are set to approximately 1/20 of peak base width.[6,7] Some UVDs have a selection of flow cells for different 'LC speeds' and their time constant can be changed by the user over a limited range.

If height is measured, which it is when solvent pulsing perturbs the area of narrow peaks, then peaks that are compared with each other should have similar widths, otherwise a specific calibration for height must be made. When integrators are used without calibration, it is safer to use area.

Strip chart recorders with fast linear response and small pen head inertia distort peak shapes less than integrators. If a chromatographer is more interested in peak shape than size, the recorder may be a more accurate instrument of display.

3 Peak Area and Solute Quantity

Mass and Flow Sensitive Detectors[8,11]

Detectors are either flow sensitive or mass sensitive. They are flow sensitive if the mere presence of solute inside the detector cell is sufficient to create a response, *e.g.* UV detector; they are mass sensitive if the solute is consumed in a chemical reaction, and it is the reaction or its products that create the response, *e.g.* FID. Solutes pass unchanged through flow sensitive detectors which can therefore be linked in series.

The difference between the two detectors is highlighted by considering the effect of stopping the (single component) mobile phase as a solute passes through the detector cell (see Figure 2.6).

In a flow sensitive detector, the detector signal is held constant and it will only change when the flow re-starts and sweeps the solute from the cell. The area of the

Figure 2.6 *Effect of stopping mobile phase on detector signal:* (a) *flow sensitive;* (b) *mass sensitive*

solute peak is increased but the height remains nearly constant (in fact, a 10% reduction in flow has been observed to produce a 1% increase in height[9]).

In contrast, the signal level from a mass sensitive detector falls to zero (at a rate determined by the detector time constant) as the solute inside the cell is consumed and no more is delivered. When flow is resumed the signal picks up again as the remaining solute is swept into the detector cell.

The peak area is split, but the sum of the two split areas is the same as the peak area if the flow had not stopped.[10] Peak height is not preserved unless the maximum has eluted before the flow is interrupted. The two types of detector are summarized in Table 2.3.

If peak area is proportional to solute quantity it must be so for both types of detector.

Table 2.3

SFC/Gas chromatography		*CZE/SFC/Liquid chromatography*	
FID	Mass	UV or UV/VIS detector	Flow
TCD	Flow	Refractive index detector	Flow
ECD	Flow (Mass)*	Fluorimetric detector	Flow
N/P FID	Mass	Electrochemical detector	Mass
FPD	Mass	Coulometric detector	Mass
Mass spectrometer	Mass	Mass spectrometer	Mass
Hall detector	Flow		
He ID	Mass		
Chemiluminescence	Mass		

*When flow conditions are not rigorously controlled an ECD can behave as mass sensitive because the electron (and so response) is a function of carrier flow rate.[11]

Flow Sensitive Detectors

In a flow sensitive detector, the output signal S, is proportional to the concentration c, of solute in a mobile phase, *i.e.*

$$c = KS \tag{2}$$

where K is a constant. Integrating this from the *start* to the *end* of the peak,

$$\int_s^e C\,\mathrm{d}t = K\int_s^e S\,\mathrm{d}t \tag{3}$$

$$= KA \tag{4}$$

where A is the peak area.

The left-hand side of Equation 3 can be transformed,

$$\int_s^e c\frac{\mathrm{d}t}{\mathrm{d}V}\mathrm{d}V = KA \tag{5}$$

or
$$\int_s^e c \frac{1}{\dot{V}} \, dV = KA \tag{6}$$

where $\dot{V}$ is the flow rate of mobile phase. As long as it is constant, it can be taken out of the integration and moved to the right-hand side of Equation 6,

$$\int_s^e c \, dV = KA\dot{V} \tag{7}$$

and what remains on the left-hand side is simply the quantity of solute, Q in the eluting zone, *i.e.*

$$Q = KA\dot{V} \tag{8}$$

For a flow sensitive detector, peak area is proportional to solute quantity provided that flow rate is constant. The majority of LC detectors and the TCD are flow sensitive.

It is therefore very important to maintain precise flow control when using flow sensitive detectors. On the other hand, peak area and sensitivity can be increased by reducing the flow rate although the analysis will take longer. If the flow rate is reduced too far, column efficiency will suffer. Bakalyar *et al.*[9,12] studied the effect of flow and solvent composition stability on area and height measurements made using a UV detector. The results are summarized in Table 2.4.

Table 2.4 *Effects of chromatographic changes on data from UV detector*[12]

	Flow rate 1%	*Composition* −1% B	*Temperature* −1% C
Area	+1%	0 to ±10%	0 to ±1%
Height	< +0.3%	−1 to −10%	0 to ±1%

Mass Sensitive Detectors

The output signal from a mass sensitive detector is proportional to the rate at which the solute is consumed by the detector reaction, or, as the consumption must be faster than the delivery for quantitation to be possible, to the rate at which solute is delivered to the detector, *i.e.*

$$\frac{dQ}{dS} = KS \tag{9}$$

Integrating this from the start to the end of the peak:

$$\int_s^e dQ = K\int_s^e S\,dt \tag{10}$$

$$\text{or } Q = KA \tag{11}$$

For a mass sensitive detector, peak area is proportional to solute quantity but independent of flow rate, which conveniently removes flow rate as a potential source of experimental error.

However, peak shape varies with flow rate. If the flow rate is high, a given solute quantity will produce a tall, slim peak. If the flow rate is low, the same quantity will produce a low, broad peak of longer retention. In both cases peak area will be the same (Figure 2.7).

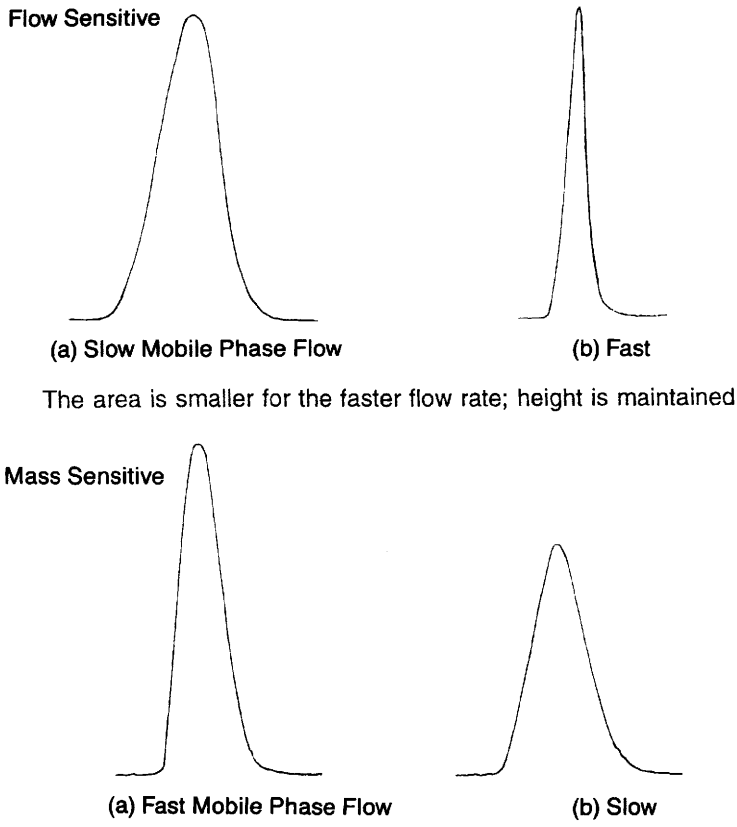


Figure 2.7 *Comparison of flow sensitive and mass sensitive detectors*

Increasing the flow rate through a mass sensitive detector will speed up an analysis without loss of sensitivity, but it will invalidate calibrations of height made at the slower flow.

Detector Overload

Detectors have a finite range of operations. FID electrometers only accept an input signal from peaks up to a certain size because the output signal is restricted to a maximum of 1 V DC. Bigger peaks are clipped at 1 V (see Figure 2.8).

As long as important peaks are kept on scale, which is usually the case when using a chart recorder, saturation can be spotted. However, integrators and computers also measure peaks which go off scale and saturation can be missed. Inspection of peak maxima which go off scale in normal chromatogram presentation ought to be part of method validation. The analyst inspects the chromatogram at a higher attenuation which brings suspect maxima on scale, and checks the peak diagnostics on the chromatogram report to see whether they warn of overload and measurement error.

Detector cells distort peak shapes through geometric or mechanical constraints by being too big or the wrong shape.[13,14] FID flames quench and vary their response considerably when certain solutes, especially halocarbons, pass through. Non-linear thermal conductivity effects cause hydrogen peaks in helium carrier gas to fold over when measured at critical concentrations by TCD.[15,16] Under certain conditions, helium ionization detectors fold peaks too.[17] The folding occurs because the linearity curve has a maximum or minimum in it.[18]

The effect of overloading or saturation is to make measured areas smaller than they should be by an unpredictable amount.

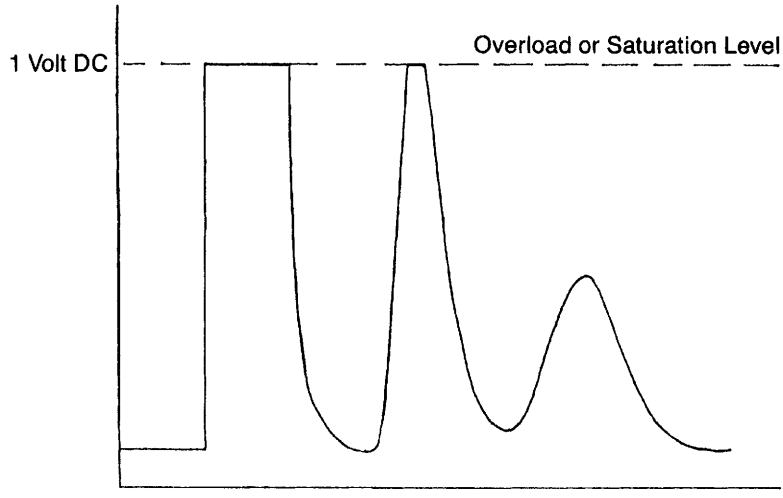


Figure 2.8 *Peak area above the overload level is lost*

Linearity

Below saturation level, there is a linear working range where detector response is proportional to solute quantity. This range is greater for peak area than peak height.[18,19] Above it there is a non-linear, but maybe useful, range where the relationship between solute and response is continuously changing. Finally there is saturation and no relationship.

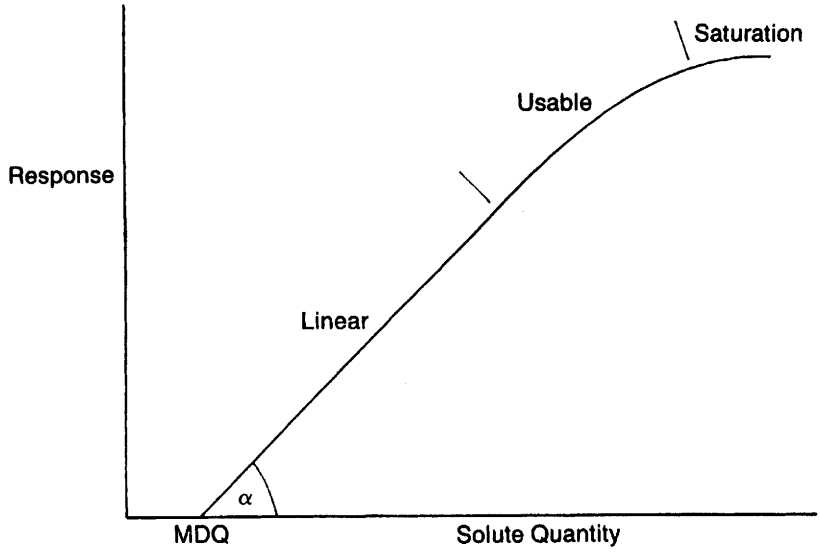


Figure 2.9 *Linear dynamic range.* MDQ = *minimum detectable quantity;* α = *sensitivity*

In the linear range the relationship between solute quantity and detector response, (output voltage) V is,

$$V = KRQ \tag{12}$$

where Q is solute quantity, R is solute response factor and K is an instrument constant. MDQ is small enough to ignore, see Figure 2.9.

Fowlis[20] has suggested that a better equation is:

$$V = KRQ^r \tag{13}$$

The index r, puts a numerical value to the meaning of linearity; the detector can be considered to be working linearly if r lies between 0.98 and 1.02.

Carr[18] generally expressed detector linearity by the polynomial:

$$V = \sum_{i=0}^{n} P_i Q^i = P_0 + P_1 Q + \sum_{i=2}^{n} P_i Q^i \tag{14}$$

In this form P_0 is the minimum detectable quantity, P_1Q is the linear term and $\sum P_iQ^i$ is the sum of all the non-linear terms among which the quadratic is dominant. It is this term which explains peak folding.

It is better to determine where the limits of linearity are by experimentation rather than to trust manufacturers' specifications. Carr also found that in UV detectors, non-linearity can occur at unexpectedly low concentrations, less than 1 millimole, and it may not be detectable by inspection.

Ultimately there is little use in judging the performance of a detector and its electronics separately except when trouble-shooting. The linear dynamic range of a detector is determined by both.

Detector Non-linearity

This normally affects the top of the peak but not the bottom which remains in the linear region, see Figure 2.10. It reduces peak height and peak area but does not affect retention time or peak skew unless the detector actually saturates or the peak is already asymmetric, in which case it might shift the peak maximum further. The fractional loss in area due to detector non-linearity is always less than fractional loss in height because area is lost from the top of the peak where it matters least.

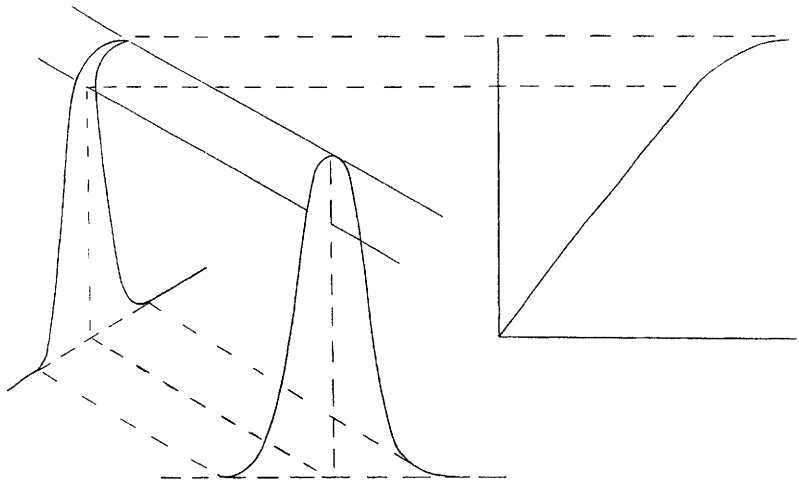


Figure 2.10 *Non-linearity occurs at peak tops and is analogous to bending over*

Column Non-linearity

This affects the peak from top to bottom. It causes peak skew which affects peak shape and therefore height and retention time, but peak area remains unchanged (Figure 2.11).

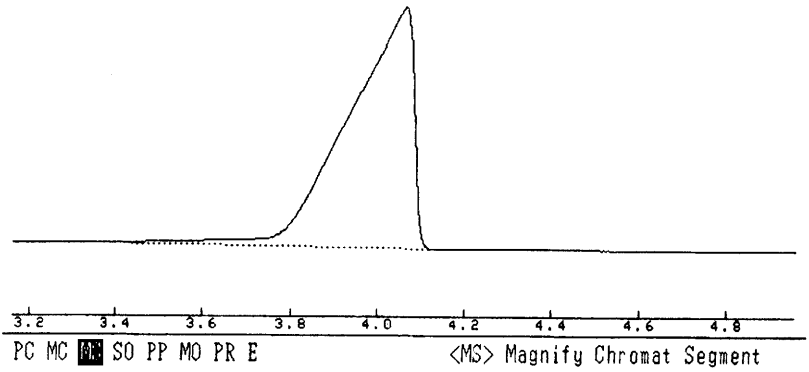


Figure 2.11 *Column asymmetry*

Electronic Noise and Drift

As technology improves, noise and drift become less, but this only encourages analysts to measure smaller quantities and so the problem never disappears. For routine analyses, where solute quantity is not a problem, electronic component noise and drift are usually much less than transducer or column noise and can be ignored except as evidence of a malfunction.

4 Sources of Peak Measurement Error

Assuming that the chromatograph is used within its operating range and the detector signal is an accurate representation of solute profile, integration errors are caused by:

(a) noise;
(b) baseline drift;
(c) separation of incompletely resolved peaks;
(d) peak asymmetry;
(e) incorrect use of the integrator.

Category (e) is deferred to Chapters 4 and 5.

Hardware Errors

Errors caused by integrator malfunction are increasingly rare. As the technology advances, more integrator functions are transferred from hardware to software/ firmware, there are fewer parts to an integrator and these are more reliably manufactured. When malfunction does arise it is usually a total breakdown, which prevents the integrator from giving misleading results.

Genuine hardware errors or malfunctions are best revealed by good validation procedures (see Chapter 4). Most problems occurring in the first year of use of an integrator are linked to operator training (finger problems).

Noise

Noise is unwanted detector signal and can be chemical or electronic in origin. The baseline noise that the analyst sees is the residue left after filtering and smoothing have removed higher frequencies. It has been allowed to remain because its frequency is too close to that of known chromatographic peaks, and filters which would remove it might also remove wanted peaks.

Noise can be measured statically or dynamically, with the mobile phase either stationary or flowing. Static measurement measures detector noise alone; dynamic measurement includes noise from the rest of the system. It should make no contribution, but if it does it warns that the detector signal is being degraded.

Noise and Frequency

Noise is characterized by its frequency and peaks by their width; one is the reciprocal of the other:

$$\text{Peak 'Frequency'} = \frac{1}{\text{Peak Base Width}} \tag{15}$$

The aim of instrument design engineers is always to minimize the amount of noise created (since what is not created does not have to be removed), filter as much of the created noise as possible, and finally to improve the signal processing techniques to measure only the peaks.

Three kinds of chromatographic noise which are recognized in ASTM E 685-79,[21] are shown in Figure 2.12.

Short-term Noise

Defined as: random variations in detector signal whose frequency is greater than 1 cycle minute^{-1}. Ideally, short-term noise is narrower than 'real peaks' and this difference can be used by integrators to recognize and disregard noise peaks. Continuing developments in capillary column technology squeeze this definition; small capillary peaks are increasingly confused with short-term noise and integrators may filter them if not programmed correctly.

Long-term Noise

Variations in detector signal whose frequency lies between 6 and 60 cycles hour^{-1}. It is the kind of noise which looks like peaks and may very well be late eluters from an earlier analysis trespassing into the current one.

Drift

Drift is change in baseline position. As noise, its frequency is even less than long-term noise. If it cannot be attributed to some known cause such as temperature or

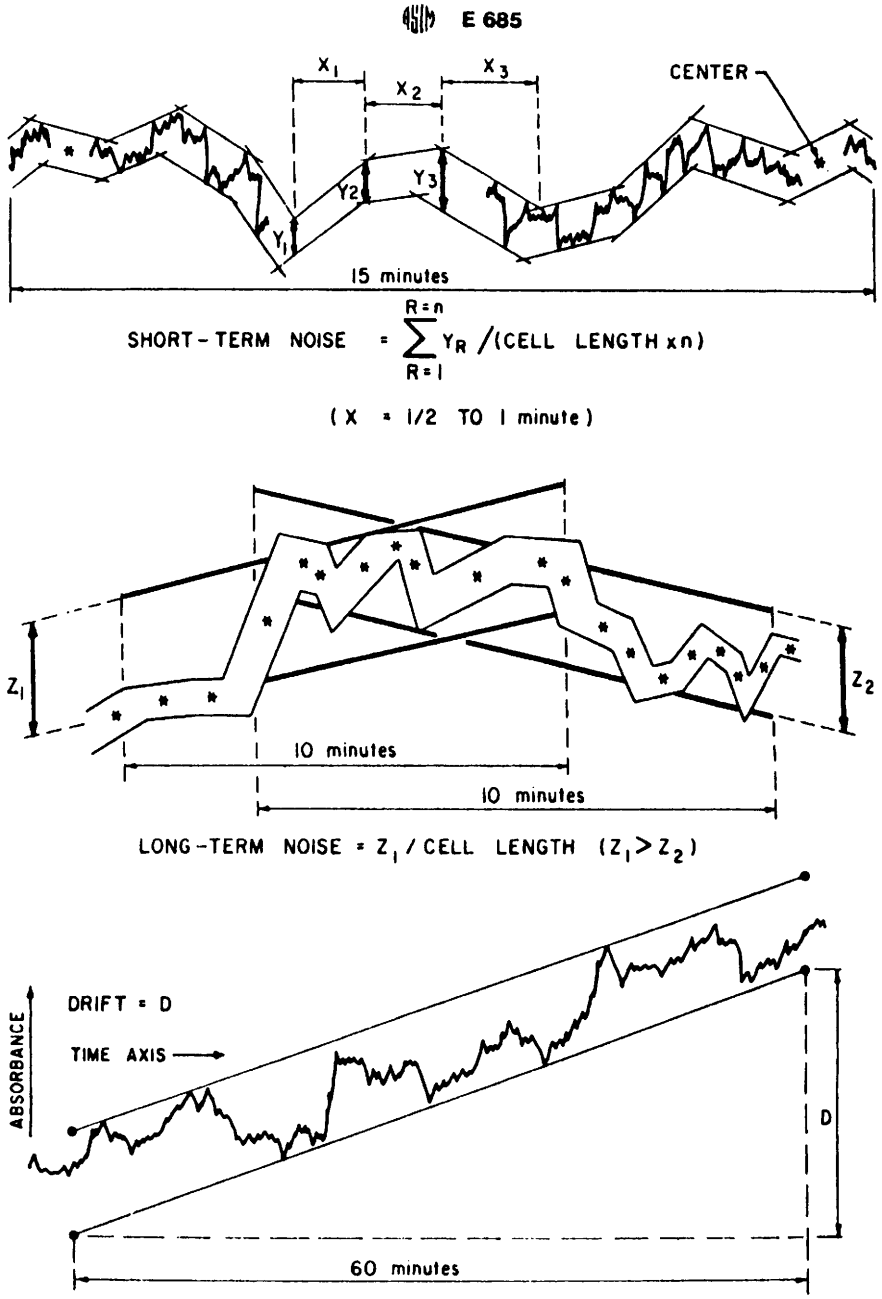


Figure 2.12 *Three types of noise recognized in* ASTM E 685-79 (Reproduced by kind permission of ASTM[21])

solvent programming, it indicates instrument instability, often due to temperature effects on the detector.

Errors Created by Noise

Noise affects integrator measurement of peaks around those regions where the peak signal is near horizontal. It blurs the base of a peak making it difficult to locate where it starts and ends and therefore where to measure the area (Figure 2.13).

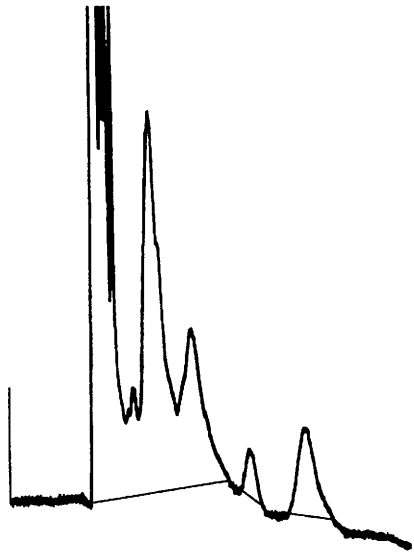


Figure 2.13 *Noise causes late detection and early termination of peak measurement*

Noise at peak tops will cause integrators to split area measurement when valley recognition is triggered by the micro-peaks. In the report these peaks are observed to have retention times within a few seconds of each other (Figure 2.14).

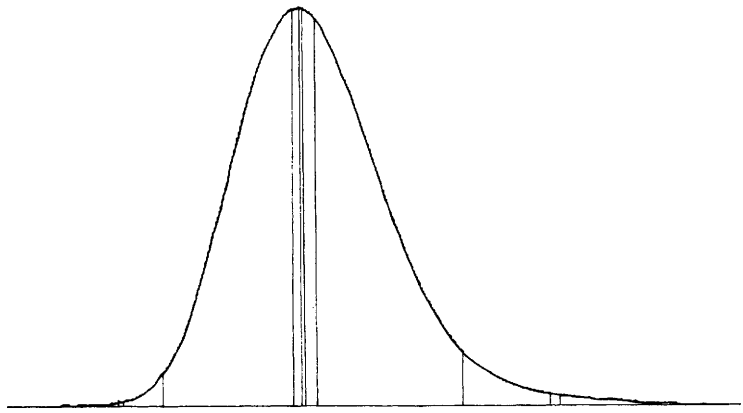


Figure 2.14 *Peak splitting results when noise valleys trigger perpendicular separation*

Noise peaks are similarly detected and measured at the beginning and end of a peak, but these usually fail a minimum area test and are not reported. They do, however, delay peak recognition and bring about early measurement termination which reduces the reported peak area. Splitting can occur in valleys or cause the integrator to locate the valley minimum in the wrong place; this either loses a slice of area (minimum area test), or transfers a slice of one peak to the other. Long-term noise is measured as peaks and included in the analysis report.

Signal-to-noise Ratio: The Smallest Measurable Peak

Noise sets a limit to the smallest quantity of solute that can be detected as a single peak. Small peaks do not simply disappear from view: electronic amplification would always bring them back. They disappear when they merge into the baseline noise and become indistinguishable from it.

The smallest fully-resolved peak that can be measured is the smallest one that can be unambiguously distinguished from baseline noise. It is described in terms of the signal-to-noise ratio S/N, which compares the height of a peak to the height of the surrounding noise (Figure 2.15).

If a solute is repeatedly analysed and diluted, it is judged that its peak can still be seen on a chromatogram when the S/N ratio is as low as 1, *i.e.* the peak is as high again as the noise amplitude, but in this kind of experiment the analyst knows the retention time of the peak and knows where to look. The observation is not unbiased.

As the S/N ratio increases to 2, the peak is more clearly visible, and when it is 3 it is unambiguous although area measurement will be imprecise.

Chromatographic noise is typically and historically measured manually as the amplitude of the baseline noise envelope. It ought, perhaps, to be computed as the standard deviation (σ) of many baseline noise amplitude measurements; good manual measurement of noise then computes a value close to 6σ.

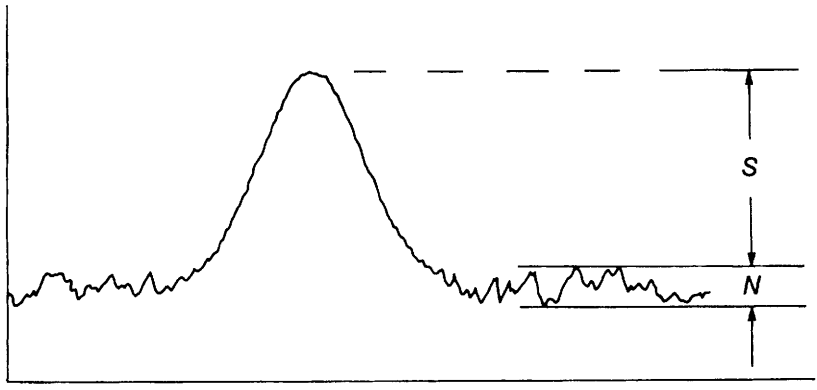


Figure 2.15 *Signal-to-noise ratio,* S/N

Limits of Detection and Quantitation

ACS guidelines[22] published in 1980 define two measures (Figure 2.16):

(1) Limit of detection LOD, at $S/N = 3$ defines the smallest peak that can confidently judged to be a peak.
(2) Limit of quantitation LOQ, at $S/N = 10$ defines the smallest peak whose area can be measured with acceptable precision.

Quantitation below $S/N = 10$ is not advised for integrators in particular because of the risk of detecting noise as peaks and of losing peaks in noise. Kaiser[23] calculates a 7% probability of this happening at $S/N = 3$.

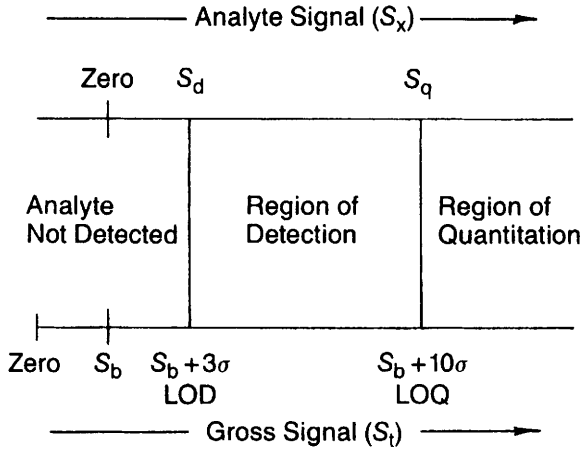


Figure 2.16 *Limits of detection and quantitation*
(Data reproduced with kind permission from *Anal. Chem.*, 1980, **52**, 2242)

In these guidelines signal is measured above the noise, so that:

$$\text{Signal/Noise ratio} = \frac{S}{N} \qquad \text{in Figure 2.15}$$

Earlier definitions have used $(S + N)/N$, *i.e.* measured signal from the bottom of the noise envelope, but when no peak is present ($S = 0$) this ratio has the value 1 implying that there is as much peak signal as noise and that solute is still there.

The S/N ratio is itself an experimentally determined quantity and as such is subject to experimental error.[24] One consequence is that when comparing detector performance or optimizing analyses for sensitivity, a factor of 2 or 3 difference in minimum detectable quantity is not meaningful.

A 'Good Baseline'

Studies of the third and fourth peak moments, skew and excess, which are very sensitive to noise around the limits of integration, have brought some workers[25,26] to judge that the accuracy of area and retention measurement is acceptable when

S/N is greater than 30:1, but a 'good baseline' is where the S/N ratio exceeds 100:1 because only at this level do the errors in measurement of variance, skew and excess become reasonably small.[25]

Errors Created By Baseline Drift

Baseline drift causes precision errors when it is not constant over a series of analyses, and inaccuracies in peak measurement even when it is repeatable.

Drifting Baseline and Peak Measurement

The baseline constructed by an integrator for peak measurement is a straight line drawn beneath the peak. If the real baseline is defined as the detector signal in the absence of the peak, then the dashed curves in Figure 2.17 are better approxima-

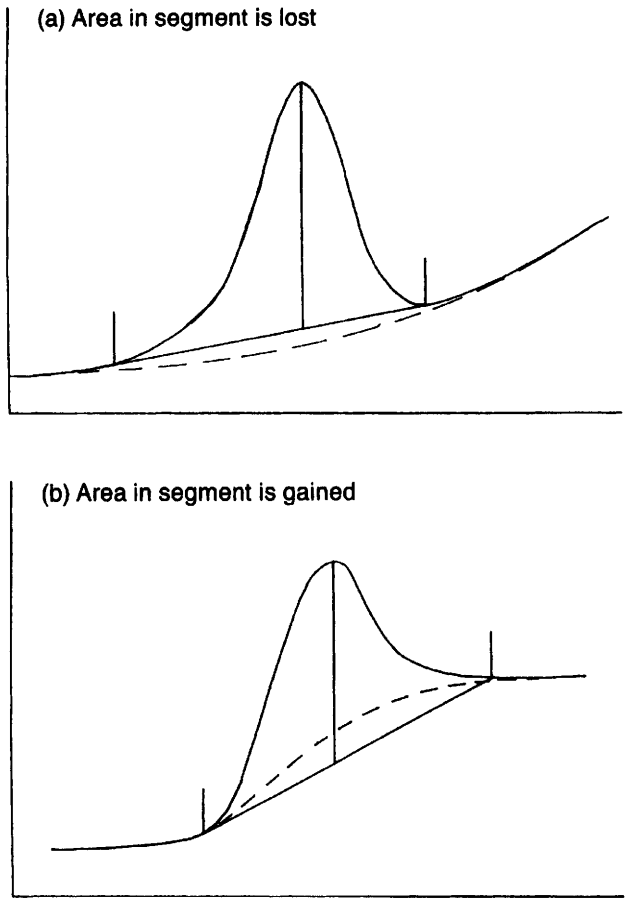
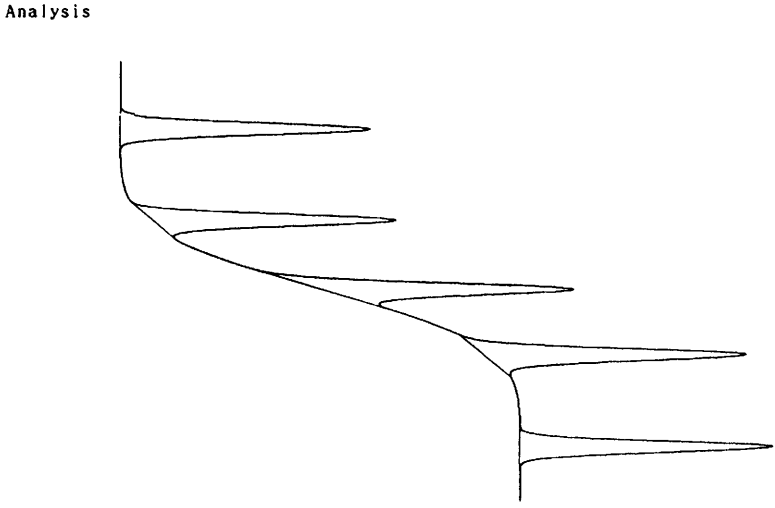


Figure 2.17 *Linear constructed baselines are not necessarily accurate*

tions. Peak areas and heights will be measured greater or smaller than they should be depending on whether the true baseline is convex or concave.

Figure 2.18 is a computer-generated chromatogram of five equally-sized peaks on a sigmoidal baseline. The first and last peaks are on flat baseline and are measured correctly, peaks 2 and 4 are located on the ankle and shoulder of the baseline (maximum curvature) and exhibit maximum measurement error. Peak 3 is on the point of inflection of the baseline where curvature error self-cancels. Referenced against peak 1, there is a 6% measurement error in the second and fourth peaks and a 0.45% error in the third.



** CALCULATION REPORT **

CH	PKNO	TIME	AREA	HEIGHT	MK	IDNO	CONC	NAME
1	1	1.255	5144737	322281	R	1		
	2	3.003	4850216	312460		2	0.9428	
	3	4.262	5167694	321998		3	1.0045	
	4	5.506	5461001	332599		4	1.0615	
	5	7.252	5157111	322675		5	1.0024	
		TOTAL	25780756	1612013			4.0111	

Figure 2.18 *Errors of linear baseline construction*
(Reproduced with kind permission of ISC, see ref. 3)

The magnitude of the error depends on the curvature of the true baseline, the position of the peak on the curve, peak width and peak overlap: overlap extends the required length of the constructed baseline. These factors can conspire at different times to make the error serious or negligible. Single capillary peaks on a solvent tail are measured with good accuracy; groups of broad peaks on a steeply-rising baseline near the end of the analysis may be measured with very poor accuracy (see Figure 2.19).

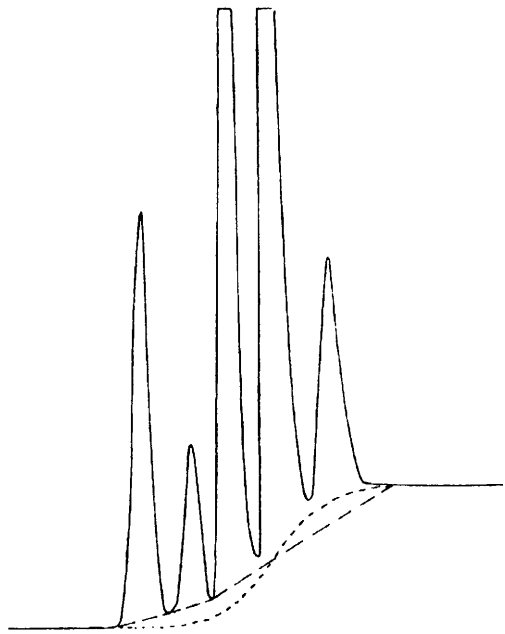


Figure 2.19 *The dotted baseline represents the true baseline; the dashed baseline is the one placed by the integrator*

Baseline Drift and Retention Time

The retention time of a peak measured on a positive baseline slope is greater than it should be because the gradient of the baseline adds to the peak top and delays the maximum. A new peak maximum is created at the point where the negative gradient of the falling peak slope is equal in (absolute) value to the positive gradient of the baseline (Figure 2.20).

Similarly, the measured retention time of a peak on a negatively-sloping baseline is always less than the true retention time.

Near the peak maximum, peaks are parabolic in shape (see Chapter 1) and the height y, of a point near the maximum is given by,

$$y = H(1 - t^2/2\sigma^2) \qquad \text{(see Eq. 27, Ch. 1)}$$

where, H is the peak height
t is time on a scale where retention time = 0
and σ^2 is the peak variance

For simplicity, in the parabolic region baseline curvature and differences in response are neglected. Positive or negative baselines are represented by,

$$y' = \pm bt \qquad (16)$$

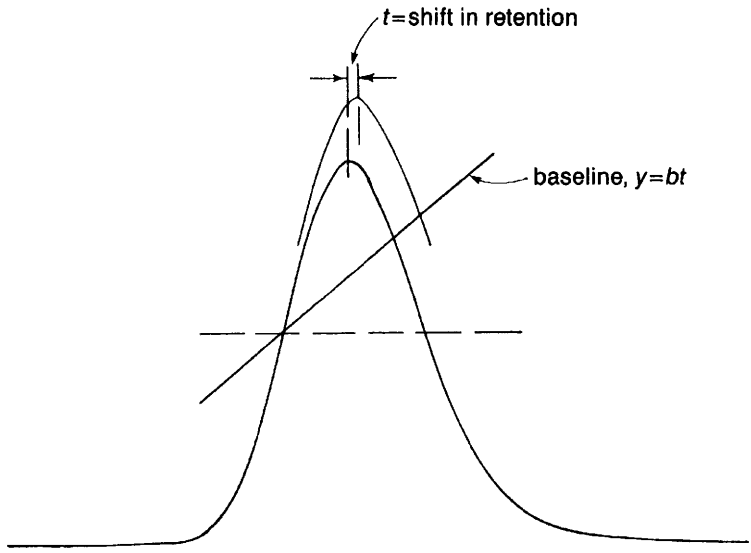


Figure 2.20 *Effect of baseline drift on retention time*

where b is the baseline gradient. The total signal S, is,

$$S = H \pm bt - Ht^2/2\sigma^2 \tag{17}$$

and this function has zero slope where:

$$t = \pm b\sigma^2/H \tag{18}$$

Equation 18 represents the shift in observed retention time. The shift is directly proportional to baseline gradient. On a flat baseline ($b = 0$) the peak maximum occurs at $t = 0$. For broad peaks, σ^2 is large; for narrow peaks it is small and the shift in retention time is proportional to peak variance. Equation 18 also shows the shift in retention time to be inversely proportional to peak height. For a peak of a given area, the taller it is, the narrower it is, and the smaller the retention shift.

Reduced Detector Operating Range

If detector drift is not checked, the baseline can drift up towards the upper limit of detector operation. This will reduce the linear working range of the detector and bring about premature saturation and loss of peak area (Figure 2.21).

A negative drifting baseline can equally drift below the lower operating limit of an integrator so that peak base area is lost, and only that part of a peak which rises into the operating range is measured. Careless use of detector back-off control can move the baseline towards either operating limit. With some integrators, pressing 'zero' does not move the baseline to zero millivolts, it moves the signal to the lower edge of the screen or chart paper and therefore does not re-establish the full input range.

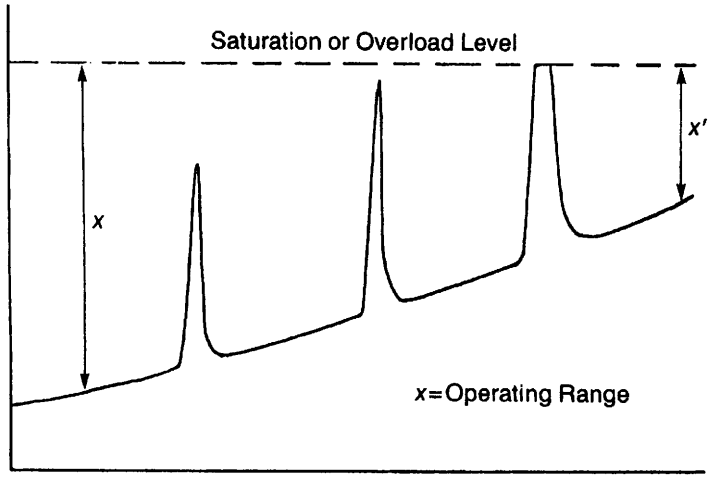


Figure 2.21 *An upwardly-drifting baseline restricts operating range*

Baselines must be kept close to ground, or virtual ground, by regular 'zeroing' of the detector signal. To do this, connect the +ve and −ve input terminals of the integrator to ground and zero the integrator using integrator controls. This calibrates the zero signal level to ground. Connect the detector signal to the integrator input and use the detector zero control (do not touch the integrator's zero) to bring the baseline back to its previous level. Thereafter, zero the signal, if necessary, by means of the detector back-off. The integrator zero should only be used to calibrate the grounded input signal to zero millivolts.

Some Preliminary Conclusions

(1) Analyses are more accurate on flat baselines than on sloping (curved) baselines, whatever the cause of the slope.
(2) Narrow peaks on curved, sloping baselines are subject to less error in area and height measurement than broad peaks because the degree of departure between the constructed baseline and the 'true' curved baseline will be less for narrow peaks.
(3) The error in measuring retention time of a peak on a sloping baseline increases with the baseline slope but is less for tall, narrow peaks than for low, broad peaks.

Errors of Incomplete Peak Resolution

Integrators and computers separate fused peaks by dropping perpendiculars from the valley between them or, if one peak is much smaller than the other and located on its tail, by skimming a tangent below it (Figure 2.22).

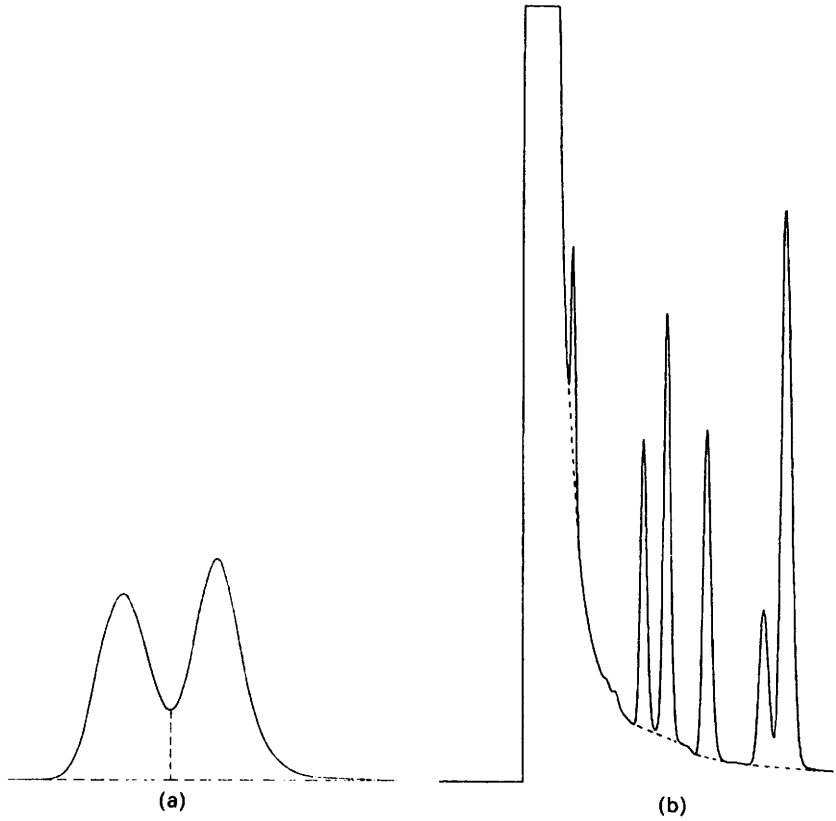


Figure 2.22 *Separation of fused peaks*

For manual measurement of overlapping peaks, triangulation is used (Figure 2.23).

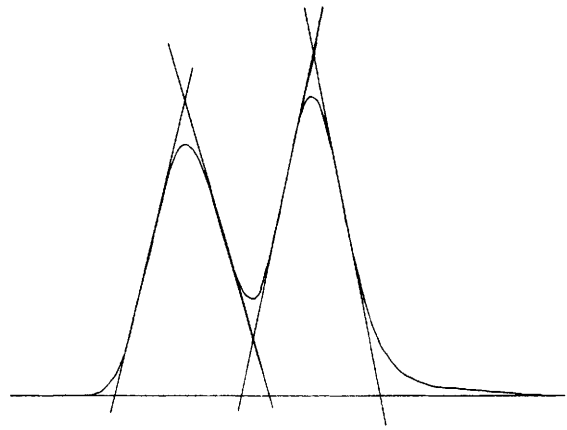


Figure 2.23 *Measurement of overlapping peaks by triangulation*

Height of a Fused Peak

If part of either peak is under the peak maximum of the other, it lifts it and creates an error in the peak height measurement. This error is greatest when a small peak sits on the tail of a larger, asymmetric one (Figure 2.24). The error in peak height that occurs when symmetrical peaks overlap is modest[32,36] until overlap becomes severe, and until then it is much less than the area measurement errors of the same peaks. There is an optical illusion in the overlap of peaks, especially symmetrical peaks: two peaks may be overlapped to such an extent that an observer is sure the peak maximum of each is lifted by the other, yet the measured peak heights show no such lift. This endurance of height when area is obviously perturbed is the main reason for choosing peak height rather than area as a solute measure, but asymmetry effects, specifically when the smaller peak is on the tail of the larger one, quickly introduce large inaccuracies to height measurements.

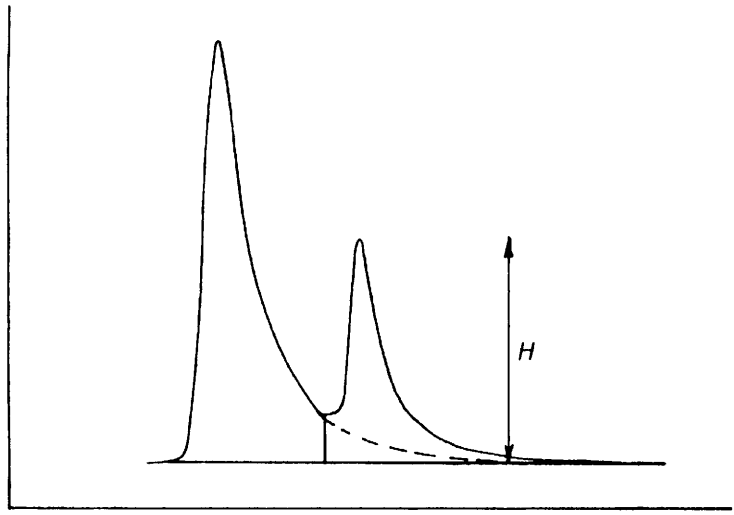


Figure 2.24 *The second peak is lifted by the tail of the first and is measured too high*

A peak between two other peaks is lifted from both sides making height measurement meaningless. However, it might still have some marginal value for quality control comparisons if the 'floating' height can be used as a diagnostic of instrument stability.

Perpendicular Separation

The use of perpendiculars or triangulation to separate two overlapping peaks will give rise to inaccurate area measurements[10,27–32,36] unless:

(1) the peaks are same height and width and are symmetrical in shape;
(2) the valley is no more than 5% peak height if response factors are different;
(3) the baseline is flat;
(4) noise does not obscure peak starts, ends and valley locations.

These are exceptional circumstances.

The measurement errors introduced by perpendicular separation of two peaks are determined by their relative sizes, their asymmetries, the elution order and the degree of overlap.[10,30–34,36] When two peaks overlap, it is the smaller peak that suffers the greater proportional error; the greater the size ratio, the greater the error of the smaller peak. If the peaks are symmetrical, perpendicular separation under-estimates the smaller area at the expense of the larger one whatever the elution order. Elution order is important when peaks are asymmetrical, especially when the larger peak elutes first. If the earlier peak is larger and asymmetrical the later peak may be unmeasurable.

As peak resolution increases, measurement errors in both peaks asymptote to zero (Figure 2.25).

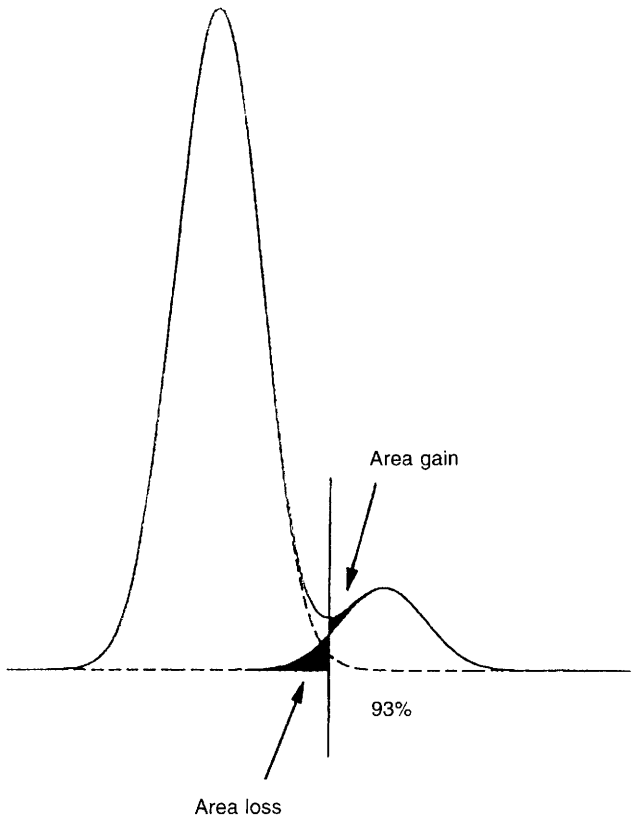


Figure 2.25 *Separation of unequal, symmetric peaks under-estimates the smaller peak* (Reproduced by kind permission from *J. Chromatogr. Sci.*, 1995, **33**, 26)

Put simply, perpendicular separation of peaks is of limited use. The valley floor between two peaks is not the correct place to locate a perpendicular for accurate measurement except in the circumstances above.

There have been attempts to devise a correct way to split two peaks: Lansdowne *et al.*[35] developed an algorithm to move the perpendicular towards the correct cleavage position, but it was based on Gaussian shape and had limited success. Kipiniak[10] suggested that the 'perpendicular' from the valley floor might be tilted in compensation, away from the smaller peak, but he did not suggest a method to calculate the angle of tilt. Little else was done because the use of perpendiculars was always regarded as a temporary measure which would be replaced in a short time by deconvolution techniques. It has yet to happen.

Effect of Overlap on Area Measurement Accuracy

The peak measurement errors introduced by separating two Gaussian peaks with a perpendicular and by triangulation were first studied quantitatively by Westerberg.[27] The error of each method was studied for two peaks of equal width (Figure 2.26) as a function of relative peak size R, and separation δ, where:

$$R = H_2/H_1 \text{ and } \delta = (t_2 - t_1)/\sqrt{2}\sigma \tag{19}$$

Comparing techniques, perpendicular separation is less sensitive to overlap errors than triangulation because the error contours on the perpendicular graph are more widely-spaced than on the triangulation graph.

The shoulder limit in Figure 2.27 is where the valley disappears and becomes a point of inflection. The detectability limit is where the two peaks appear to be one, and is defined[27] mathematically as the point where the second and third derivatives of peak height with respect to peak separation are both zero. As the peaks merge, the fact that two peaks are present remains clearly visible for the least resolution when the peaks have a relative height of $R = 0.446$. Vandeginste extended this and showed that when the peaks are allowed to have different widths, this point of detectability occurs at smaller values of R as the second peak broadens.[37]

Westerberg concluded (erroneously) that the smaller peak is always over-estimated by both methods and the measurement error increases as peak separation is reduced and the peak height ratio increases (Figure 2.26). Conversely, the error disappears when resolution is complete.

Snyder[34] pointed out that the smaller symmetrical peak is *under*-estimated. Kipiniak[10] repeated Westerberg's calculations and also showed that for symmetrical peaks, the smaller peak is under-estimated but added that when peaks are asymmetrical, the usual case, then elution order is important and the smaller peak is over-estimated when it elutes second. Foley[36] used various combinations of computer-generated symmetrical and asymmetrical peaks, separated by perpendiculars to measure areas and associated errors. Errors of up to 50% were not uncommon, and in staged examples errors up to 200% were achieved. Studies by Vandeginste[37] with symmetrical peaks of unequal width drew similar conclusions to Kipiniak.

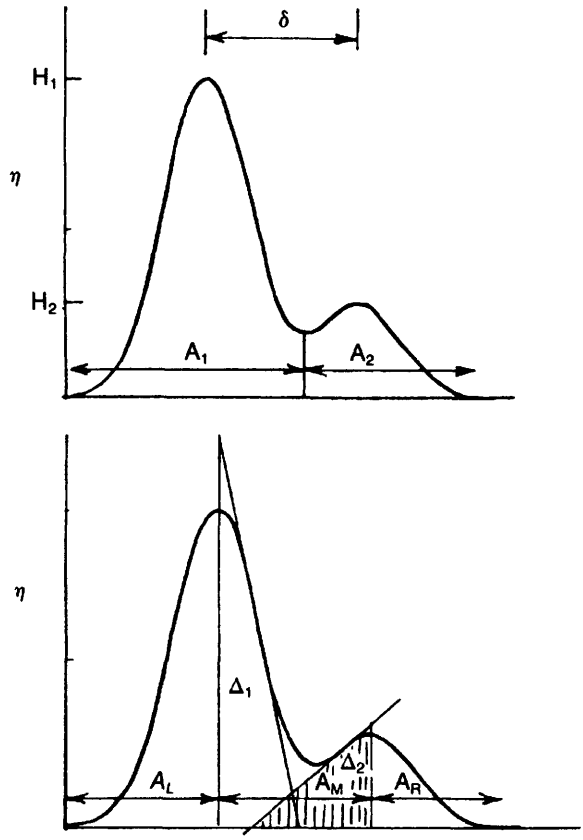


Figure 2.26 *Comparison of overlapping peak area measurement by perpendicular and triangulation*
(Re-drawn with kind permission[27] from *Anal. Chem.*, 1969, **41**, 1770)

Papas and colleagues[38–40] measured a range of simulated overlapping peaks in order to compare integrator and computer systems. They not only encountered the errors of peak overlap described here, but found that integrators from different manufacturers gave different measurements of the same signal. This was attributed to the non-uniformity of measuring algorithms and programming of integrator parameters. Chemical manufacturers whose product price or product tax depends on chromatographic analysis might inadvertantly choose the data processor that provides the least favourable assay and so penalize themselves.

Recently, Meyer[29–32] has used a more powerful chromatogram simulator for systematic and extensive measurement of the area errors by perpendicular and tangent separation of symmetrical and asymmetrical peaks, over a range of relative peak sizes (height ratios up to 1000:1 and width ratios up to 10:1), asymmetries, overlap orders and resolutions. Her results showed that for all asymmetrical pairs, area measurement could not be trusted to be accurate whether the peaks were measured by perpendicular or tangent separation; height measurement was accurate to within about 1% if the smaller peak eluted first and the pair were separated by a

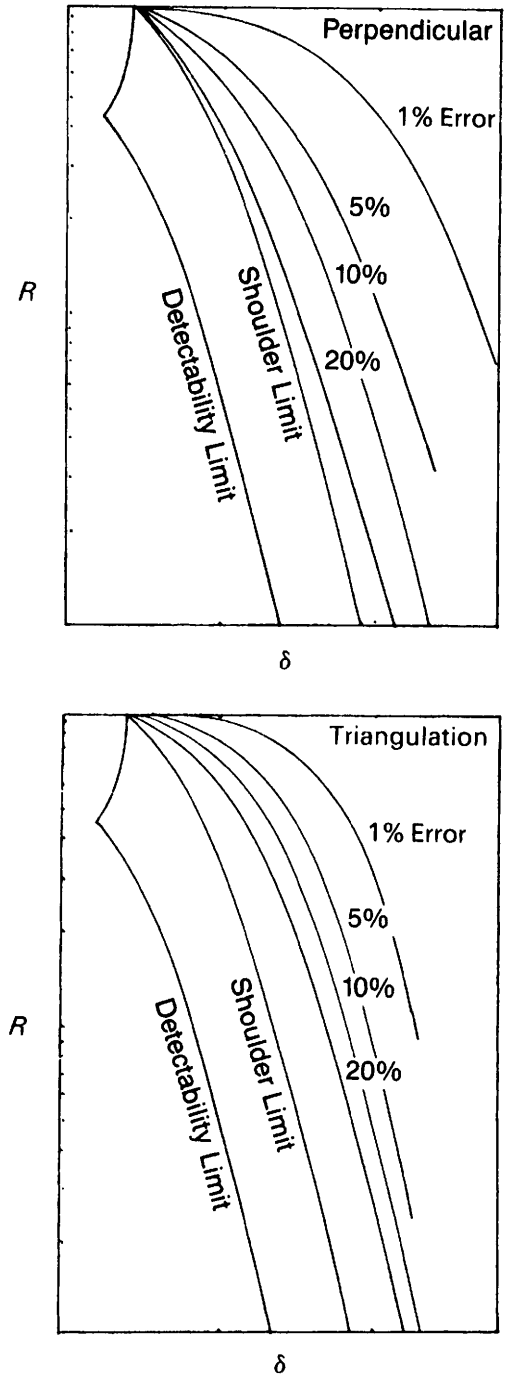


Figure 2.27 *Comparison of error using perpendicular or triangulation*
(Data reproduced with kind permission[27] from *Anal. Chem.*, 1969, **41**, 1770)

perpendicular, but height and area were unacceptably inaccurate when the smaller peak eluted later and tangent separation was used. It depends on the resolution, as resolution improves both areas asymptote to the true measure of separate peaks.

For symmetrical pairs both peaks are generally measured accurately if height is used, but the smaller peak is under-estimated if area is measured.

The accurate measurement of small peaks overlapped by others is difficult for integrators. Snyder, Foley and Meyer's papers include graphs and diagrams for visual estimation of relative peak size and associated error, but the simple fact is that there are too many diagrams and too many exceptions to the rules to allow simple conclusions. For critical analyses, and assuming that method development can do no more, the analyst must decide on and apply limits to resolution and asymmetry, and must judge carefully whether to measure area or height.

Error Correction Strategies

The best strategy to eliminate measurement error is to achieve baseline resolution of peaks before measurement, but this is not always possible. Correction of perpendicular errors by the use of 'coefficients' has been attempted by many researchers in the years since Westerberg and Snyder. Success always depended on the peak shapes being known, usually symmetrical, or at least staying constant during calibration and analysis, but real peaks seldom behave as they should. Imperceptible changes in shape can produce significant changes in measurement.

Proksch,[28] for example, prepared tables of correction coefficients to compensate for the errors when pairs of variously-sized, symmetrical peaks are separated this way, but they are only accurate when the peak shapes are constant and the smaller peak elutes first.[41]

All attempts to correct the error of perpendicular separation by coefficients limit themselves to the two peak example. With three overlapping peaks the situation is much more complex. The triplet has always been an early discard in deconvolution studies.

An alternative and reasonably accurate way to measure two overlapping asymmetric (EMG) peaks, without involving perpendiculars or tangents, is to measure the total area of both and subtract from it the area of the first peak calculated from appropriate EMG equations[36,42] (see Equations 33b and 33c, Chapter 1). Foley demonstrated inaccuracies of only 4% if τ/σ was no worse than 2 and valley height no higher than 45%. This approach could be extended in principle to peak groups provided that they were all EMG peaks and that the peaks were subtracted in time order, but the accuracy degrades with each subtraction.

Peak deconvolution software is emerging[43–45] which can resolve peaks if the number of peaks and their shapes are known.

Differences in manufacturers' instruments can be avoided by standardizing on one particular model, but the only sensible conclusion to be drawn from all of this is to acknowledge that errors arising from peak overlap are introduced by the algorithms of perpendicular and tangent separation and cannot be eliminated by anything but better chromatography. Integrators are able to generate a highly precise and totally inaccurate set of results for all the foregoing examples.

Error Propagation through Overlapping Groups

The area gained by the smaller peak in Figure 2.28 is equal to the area lost by the larger peak; there is no error in the summed measurement of the pair. In general, an integrator will measure the *total* area of groups of peaks with great accuracy provided that baseline placement is correct, but the *individual* area measurements will contain the perpendicular errors from both sides of the peak and the size of the errors will depend on how much interference is imposed by neighbours, and therefore on resolution, relative peak size and shape. If e_x is the measurement error in peak area, in area units, of a peak x, in a group of n peaks, then for the group,

$$\sum_{n} e_x = 0 \tag{20}$$

this equation can be re-arranged to extract the measurement error e_i, of any one peak in the group as:

$$e_i = -\sum_{n-1} e_x \qquad i \neq x \tag{21}$$

The measurement error in any one peak is equal and opposite in sign to the sum of all the errors in all the other peaks in the group. In other words, as a result of how integrators work, the error is determined by the group and not the peak. Even if individual errors are small, their sum can be significant if the group is large, as in complex capillary chromatograms. The fractional error in peak size, e_i/A_i may be enormous if the peak is small and errors elsewhere in the group are large.

When baseline placement is incorrect then equation (21) is incorrect too.

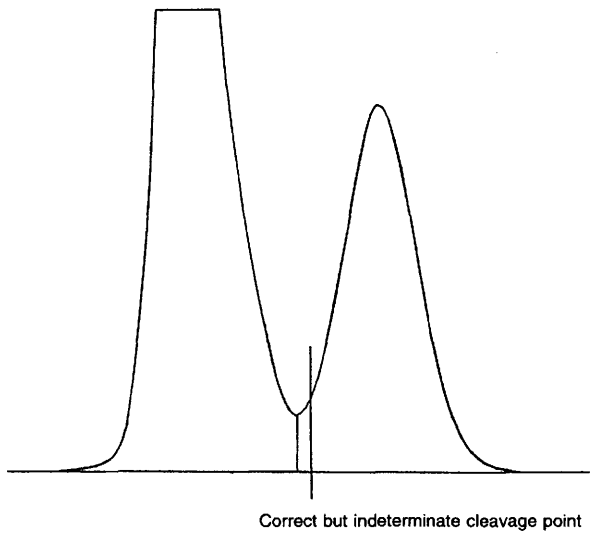


Figure 2.28 *Perpendicular separation of asymmetric peaks*, cf. *Figure* 2.25

Varying resolution across the group, tailing peaks and peaks of unequal width and size create unequal contributions of each peak to the other, either side of the perpendicular, which can result in gross over-estimates of the area of the later peak (Figure 2.29). Unequal response factors compound the errors because unit amounts of geometric area from each peak then do not equate to the same masses.

Peak tailing combines with overlap to spread the errors of perpendicular separation in a systematic manner through the group.[3] In particular, the area of the last peak in a fused group may be wildly inaccurate.

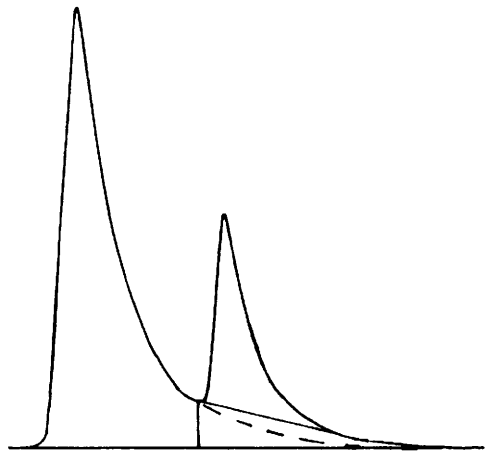
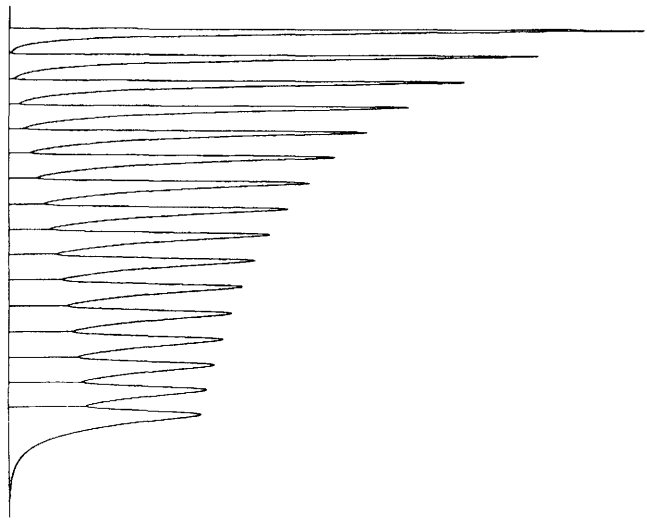


Figure 2.29 *It may be less inaccurate to skim the smaller peak*

Figure 2.30 shows a computer-generated chromatogram. The 16 peaks have the same area and asymmetry, $B/A = 4$. They are spaced at half minute intervals but their width is uniformly increased to create increasing overlap. Peak areas, measured by perpendicular separation, are referenced to the first peak and the area ratio (supposedly 1:1) reported as CONC. Except for the first and last peaks, each peak passes its tail area to its follower, and receives the tail area of its predecessor. These are not equal contributions because the peak widths gradually increase and each peak loses a little more than it gains. The last peak is over-estimated by 27% because there is no following peak to which it can donate its own tail. The sum of the area losses of the preceding peaks, $\sum(\text{Conc}_{1\text{to}15} - 1)$, is equal to -25%, in reasonable agreement with Equation 20 above.

Noting that in real groups of peaks, group and peak size, degree of overlap, asymmetry (including 'fronting') and response factors can all vary, that peak separation might include some tangents as well as perpendiculars, peak areas measured from larger groups are effectively worthless. The analyst can infer from consistency of results that the same material is measured or that composition is constant, but he should be very cautious about claiming to know what that composition is.

Therefore, measured peak groups should contain as few peaks as possible and these should be as near resolved as possible.



** CALCULATION REPORT **

CH	PKNO	TIME	AREA	HEIGHT	MK	IDNO	CONC	NAME
1	1	0.496	5305231	835362	R	1		
	2	0.996	5294473	697597	V	2	0.998	
	3	1.496	5278366	599840	V	3	0.9949	
	4	1.996	5261428	527155	V	4	0.9917	
	5	2.495	5233851	472062	V	5	0.9865	
	6	2.995	5228614	429001	V	6	0.9856	
	7	3.494	5216771	394619	V	7	0.9833	
	8	3.993	5206816	366638	V	8	0.9814	
	9	4.493	5172876	343769	V	9	0.9751	
	10	4.991	5190388	324724	V	10	0.9784	
	11	5.49	5181501	308343	V	11	0.9767	
	12	5.989	5177132	294498	V	12	0.9759	
	13	6.488	5172920	282461	V	13	0.9751	
	14	6.986	5169050	271991	V	14	0.9743	
	15	7.485	5165786	262994	V	15	0.9737	
	16	7.984	6755953	255005	V	16	1.2735	
	TOTAL		85011128	6666057			15.024	

Figure 2.30 *Error transmission through a group of peaks*

Peak Overlap and Calibration

If it is impossible to rid a chromatogram of all overlap and asymmetry, the usual case, analysts should aim to reproduce during calibration those ranges of asymmetry and overlap encountered in measurement of 'unknowns'. This may not be hard to achieve for the QC analyses of some production samples. It can be difficult when using external standard calibrations where standard solutions are made from a few pure components which give a simpler chromatogram than the 'unknowns', and which might contain not only the standard peaks but also other, unwanted peaks which overlap and interfere with the required ones. Worse deceptions arise when complex extracts of known biological samples are used as standards for calibration, and this is followed by the measurement of even more complex, unknown bio-

logical samples. In these cases there is little relation between the asymmetry and overlap of standard and unknown.

The essential goal in calibration is to produce standard chromatograms whose envelopes (*i.e.* whole chromatogram profile) match as closely as possible the overlap and asymmetry occurring in chromatograms of 'unknowns' with the hope that the errors created by asymmetry and overlap are the same in both sets and cancel.

Tangent Skim Errors

Changing from perpendicular to tangent skim reduces the extreme errors when the peak height and peak width ratios become too small (less than 10:1 approximately), but the point of transition from perpendicular to tangent skim is arbitrary and the peak width ratio is seldom considered to be important. If the skimmed peaks are in a group which is separated by perpendiculars, the errors are greater still.

Tangent skimming under-estimates the area of the smaller peak unless it is very much smaller and narrower than the peak from which it is skimmed (Figure 2.31).

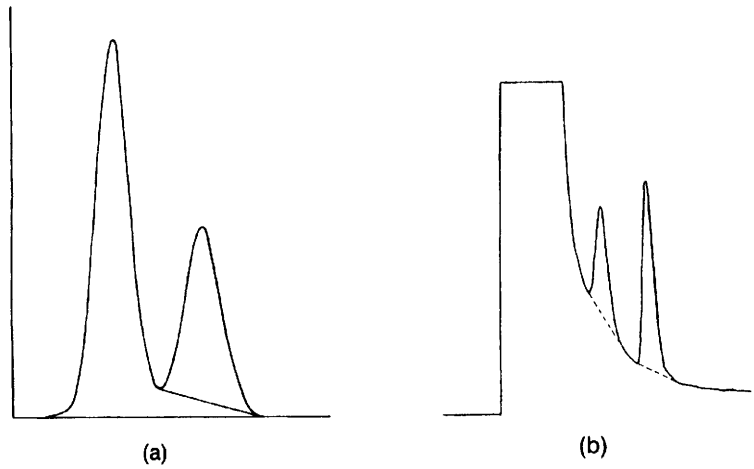


Figure 2.31 *Tangent skimming for smaller peaks:* (a) *is not very accurate;* (b) *is reasonably accurate*

This happens because the integrator draws a linear baseline under the smaller peak when in reality it should be a curve, of the host peak, and should follow a lower path. Area between the straight line and the curve is lost.

When groups of overlapping peaks are skimmed and separated by perpendiculars, the width of the group may exaggerate the departure between true and assigned baseline, and the errors of the perpendicular separation of the group are included in the final measurement.

Errors of Perpendicular/Tangent Transition

Another kind of error is introduced when integrators measure fused peaks of such marginal size that in one analysis separation is by perpendicular, but in another, tangent is skimmed beneath the smaller peak (Figure 2.32).

This creates a discontinuity in the peak measurement. To overcome it, integrators can force separation by one method exclusively. This increases precision but it does not necessarily improve the accuracy of measurement.

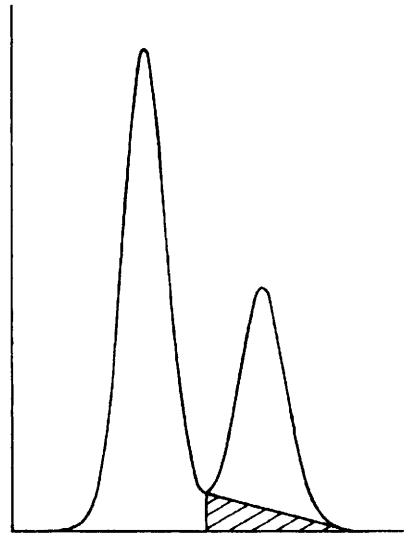


Figure 2.32 *The change from perpendicular to tangent transfers the shaded area to the large peak*

Overlapping Peaks on Sloping Baselines

Sloping baselines (which include the tails of large peaks on which smaller peaks ride) compound the errors of overlap in two specific ways:

(1) Position of skimmed peaks on the solvent tail. The curvature of a peak tail is not uniform and, in consequence, when it acts as a baseline for groups of small, skimmed peaks, the loss in area to these peak measurements depends on their position on the tail. The area loss will be greatest where the departure between tangent baseline and tail curve is worst, where the curvature of the tail is greatest.

Compounding this kind of error, Hunt showed that integrators skim small peaks from the valley position[46] not the tangent position preceding it (Figure 2.33), and this creates an additional loss of area to the skimmed peak.

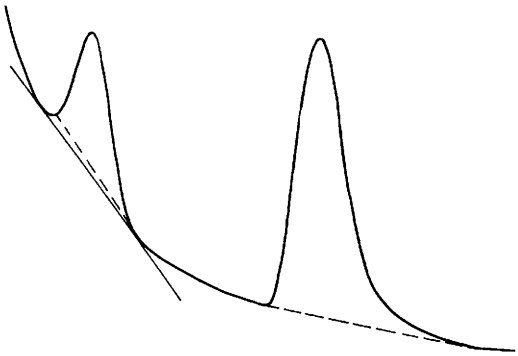


Figure 2.33 *Tangent is skimmed from previous valley, not tangent point before valley. Therefore, error varies with position on tail*

Dyson showed that the error this causes increases as the rider peak sits higher up the slope, it is greatest when the skimmed peak sits at the point of inflection of the host peak.[3] It amounted to a 27% loss when the width ratio of the solvent and skimmed peaks was 10:1. This latter error could be avoided by a simple change to the integrator software.

The 27% error which occurs when at the point of inflection ($\sigma = 1$) of the host peak (see Figure 2.34) should be contrasted with the almost 0% error of peak 3 in Figure 2.18. It too sits on the point of inflection of a sigmoidal baseline but it is measured accurately because it is measured differently. Different measuring algorithms apply to peaks on baseline and to peaks on peaks (see Chapter 5).

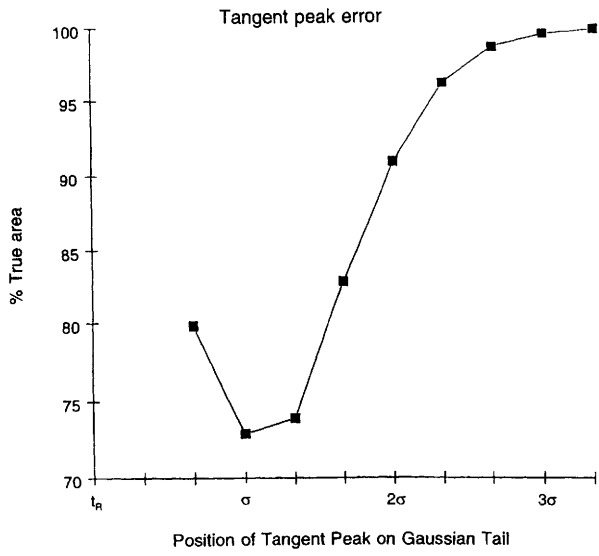


Figure 2.34 *Tangent area measurement error as a function of peak position on tail* (Reproduced with kind permission of ISC, see ref. 3)

(2) Shifting the valley position. If a valley is thought of as an inverted peak top, then just as retention time is shifted by a sloping baseline (Figure 2.20), so is the valley but in the opposite direction. A positive baseline will bring the valley forward in time, a negative baseline will delay it (Figure 2.35).

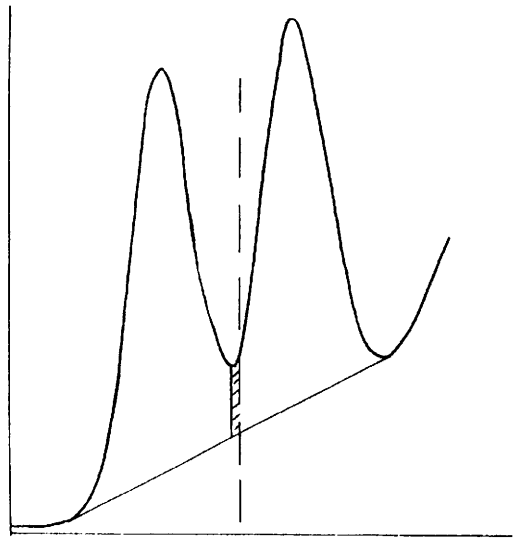


Figure 2.35 *The dashed line shows where the valley perpendicular would have been on a horizontal baseline. The shaded area is transferred from the first peak to the second*

Whether this shift introduces a serious error or not depends on baseline gradient, peak width and, additionally, on the degree of peak overlap. For narrow peaks and shallow gradients the errors can be ignored. Larger shifts will occur for broad peaks on large gradients. If resolution is poor and the valley is high the effect of the shift will be to shave a tall slice off one peak and add it to the other. With good separation the valley is low and the transferred area will be small (Figure 2.36).

The errors arising from the measurements of two same-sized, symmetrical peaks at various resolutions on various baseline gradients are shown graphically in Figure 2.37. The 'peak gradient' used as a right-hand boundary is peak height divided by half base width, *i.e.* the average gradient of the peak side.

Multiple Fused Peaks

Given the errors inherent in the measurement of fused pairs, larger groups (even three peaks) clearly offer greater opportunity to compound errors. The central peaks in particular are disturbed on both sides. It is therefore very important to configure quantitative analyses in order to minimize baseline gradients and peak overlap. If analysts can choose their analyses – as with prepared solutions – and arrange to have peaks of similar size it will reduce errors where overlap is unavoidable.

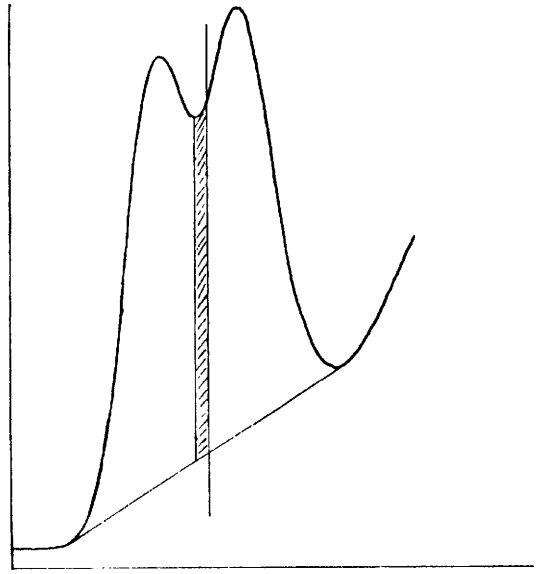


Figure 2.36 *The transferred area is significant when the valley is high and the gradient large*

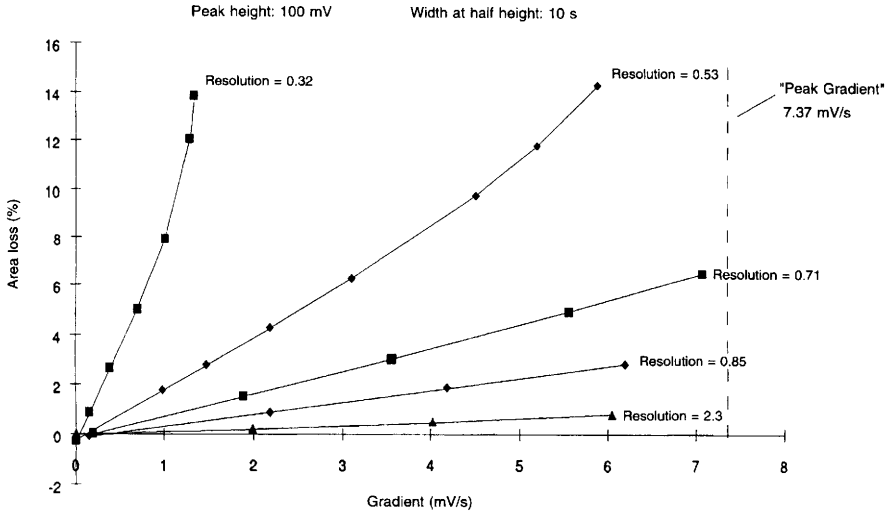


Figure 2.37 (Reproduced with kind permission of ISC, see ref. 3)

Commonly, the chromatographer is confronted by a complex chromatogram on a shifting baseline with little choice but to accept it and measure it as it is.

Calibration moderates the errors by adding compensation to the response factors (see below), but results should never be accepted unquestioningly.

Overlapping peaks as shown in Figure 2.38 are not an unusual occurrence, and with groups like these the analyst cannot even be sure of how many peaks are present[47,48] let alone measure their areas or heights. In these circumstances, the object of quantitation is to produce comparable sets of results. Two solutions that produce the same chromatograms and have the same measured composition are judged to be the same material. Repeatability and not absolute accuracy is the criterion to accept. It is a substitute that has often proved fallible but survives for want of a better alternative.

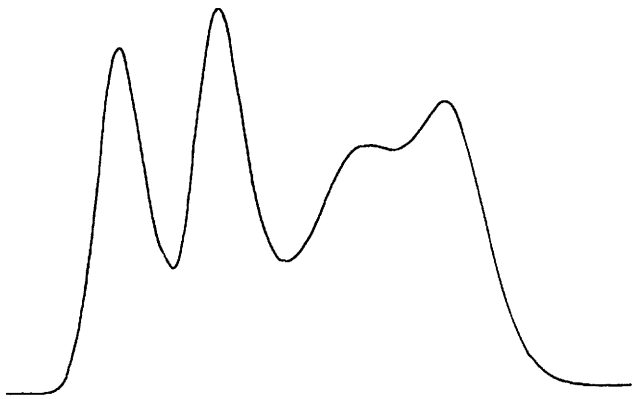


Figure 2.38 *Looks like four peaks – but it is only a guess*

Mathematical Deconvolution of Overlapping Peaks

Many attempts[49–55] have been made to deconvolute overlapping peaks by mathematical techniques and so avoid the need to use perpendiculars and tangents. All studies have required peaks to fit a mathematical model peak shape such as Gaussian, and all have failed for two basic reasons:

(1) Peaks will not conform to a single mathematical model. Even within the same chromatogram a range of models might be necessary for different peaks. One peak model can be insufficient to describe a single peak if asymmetry and overlap conspire adversely over the measured range.[56]

(2) There is simply not enough information contained within a single channel detector signal to allow numbers of peaks of indeterminate shape to be extracted from a group.[57] Noise degrades the limited information that is there. Curve-fitting procedures only work when the number of peaks is known.[37]

Dorsey *et al.* reviewed[58] developments in peak deconvolution up to 1992. They reported progress but no breakthroughs.

The Isomer Test

If a perfect deconvolution technique existed which could work on peak shape alone, it would be able to resolve a single species peak into its isomers and the correct isomer count would be proof of function. Such a technique would be able to distinguish between different species with the same retention time, *i.e.* resolve mixtures, pick out impurities and there would be no further need of a column. At the present time, only those data processors associated with GC–MS and LC–MS can pass this test provided that there are not too many isomers. As more coupled techniques and their databases are developed, the list will grow. Integrators connected to FIDs and UVDs fail the test.

Perpendiculars and tangents remain in use because FIDs and UV detectors also remain in common use and deconvolution studies have failed to deliver a better method. They are straight lines and can at least be drawn with minimal subjective judgement. Non-linear separation of tailing peaks using, for example, exponential curves has to be justified and the curve correctly located, and though an integrator could be given rules to do this, the analyst would have no easy way to test the robustness of such a technique.

'Three-dimensional' or multi-sensor detectors such as Diode Array, Coulometric Electrode Array, GC–FTIR and GC/LC–MS provide extra information, enough to allow deconvolution if the compounds have different properties (enantiomers are a problem[30]), but the mathematics of resolution remain formidable and, as yet, commercially unavailable. Sequential detectors provide extra information but it is not always possible to link appropriate detectors, and volume effects might degrade peak shape.

When the analyses of unfamiliar samples (real 'unknowns') are required, as opposed to the analysis of familiar samples of the QC type ('variations'), multi-channel detectors will take over in spite of the extra cost.

Current commercial software that claims to resolve overlapping peaks invariably invites the user to specify the number of peaks and to select a peak model. This has legitimate support in some QC labs where the analysis is well-known, and additional information can be supplied with confidence, but such software cannot be applied to unknown samples.

Synovec *et al.*[59] have shown that peak quantitation without chromatographic resolution is possible in a QC environment where the experimental conditions are stable enough for sequential comparison of chromatograms and the analyst knows what to look for; calibrations are made by spiking. Even the presence of unknown impurities can be seen but they cannot be measured and they degrade the other results. There is no claim that this technique is applicable to random unknown samples – which is what the integrator manufacturers are looking for.

Errors from Peak Asymmetry

Asymmetry by itself is not a problem to integrators because they measure peaks by the Trapezoidal Rule (see Chapter 5). In principle, an integrator can measure symmetrical and skewed peaks with equal accuracy. Asymmetry is a measurement problem because of the way it combines with peak overlap and baseline noise, and because it compounds the errors they create. It lengthens analysis times, increases (*i.e.* makes worse) the minimum detectable quantity which, ultimately, depends on peak height.

Skew is peak asymmetry about a vertical axis. It is positive if the peak tails and negative if the peak 'fronts'. It occurs round the base of a peak much more than the top and can be thought of as asymmetrical base broadening made possible by transfer of area from the top of the peak to the bottom. In fact the top of the peak often remains symmetrical while the base skews severely[60] (Figure 2.39).

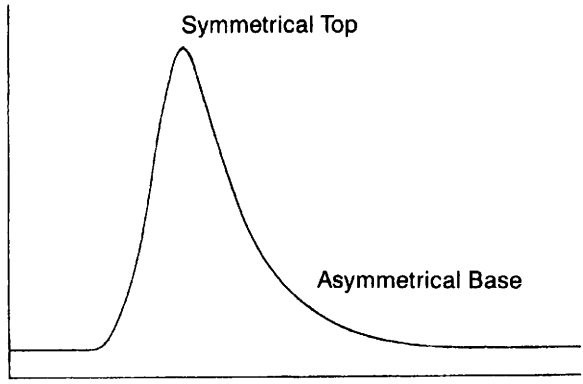


Figure 2.39 *The top of an asymmetric peak can be quite symmetrical*

Asymmetry caused by pre-column broadening and slow equilibration between solute and column will dissipate as retention increases. This is sometimes used as a visual inspection for injection proficiency (see Figure 2.40).

Asymmetry contributes to a number of direct and indirect errors in peak measurement:

Asymmetry and Peak Tailing

Tailing makes it harder for the integrator to track the end of the peak as it merges into the baseline noise. The end limit of integration is invariably located too soon, area is lost, and the measured peak area is too small (Figure 2.41).

Tailing can also persist after the recorder or plotter has ceased to show it. Figure 2.42 shows a trace of about 100 ppm of nitrogen in argon measured on a TCD using hydrogen carrier gas. Both peaks should be positive because the thermal

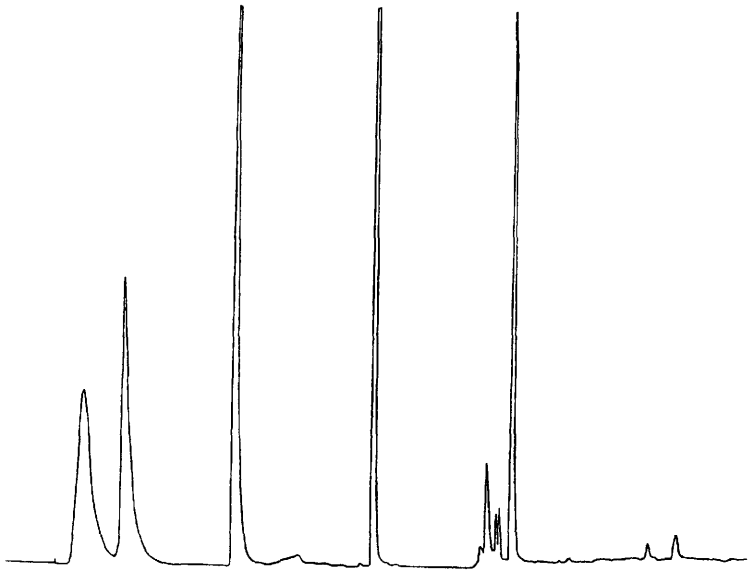


Figure 2.40 *Asymmetry caused by injection port. Note how it dissipates with time*

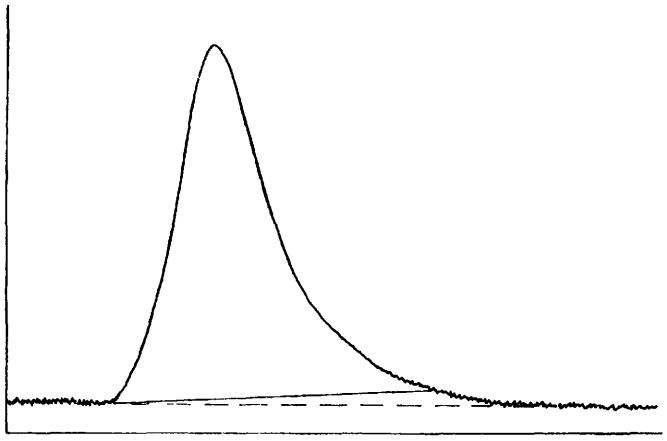


Figure 2.41 *Asymmetry causes peak tail to be lost under the noise*

conductivities of both are less than that of the carrier gas, yet the nitrogen peak is inverted. In spite of appearances, argon has not finished eluting and traces are still inside the TCD cell when the nitrogen emerges. Since nitrogen has a higher thermal conductivity than argon there is a small rise in the heat loss from the detector filament which gives rise to the negative peak. Temperature control of the resolution and therefore of the concentration of nitrogen in argon allows the tail to be monitored.

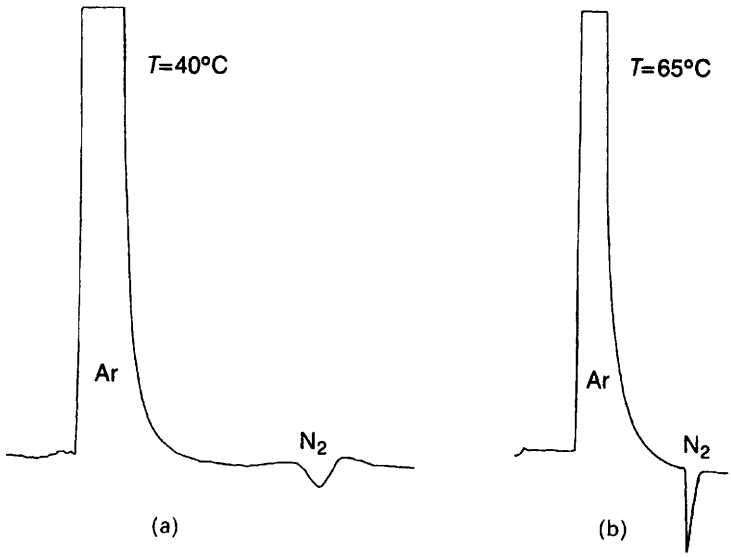


Figure 2.42 *Trace N_2 peak in argon on Molecular Sieve 5 Å columns. H_2 carrier gas and TCD*

At 40 °C any integrator would detect the end of the argon peak before the start of the nitrogen and would therefore fail to measure the whole of the argon tail, but it is clear from Figure 2.42a that the argon tail is not visible to be measured.

Jover and Juhasz[61] estimated how much area was being lost by fitting tail ordinates to an exponential curve and estimating the tail area by extrapolation of the curve. They showed that the extrapolated peak area was more than 4% larger than the conventional measurement when the peak end was judged by eye.

Asymmetry and Base Broadening

Asymmetry broadens peaks, creates overlap of neighbours and is a loss of column efficiency. At the same time it can appear that there has been an increase in resolution because peak top narrowing accompanies base broadening, in order to preserve area, and so the valley height falls (Figure 2.43). Asymmetry increases the overlap between two peaks of fixed separation and creates or increases the overlap problems described earlier.

When a peak tail disappears under the following peak, perpendicular separation transfers the tail into the second peak (Figure 2.44).

Asymmetry and Manual Peak Measurement

Commonly used manual peak measurements are based on the Gaussian peak model. Formulae for triangulation and peak height × width (50% *H*) calculations inherently assume symmetry. If this is no longer true, manual measurements of peak area (and height) cannot be assumed to be proportional to solute quantity.

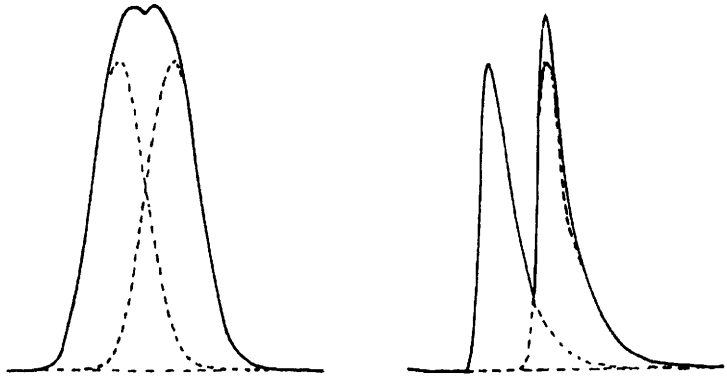


Figure 2.43 *Paradoxically, asymmetry can appear to improve resolution*
(Reproduced by kind permission from *J. Chromatogr. Sci.*, 1977, **15**, 303)

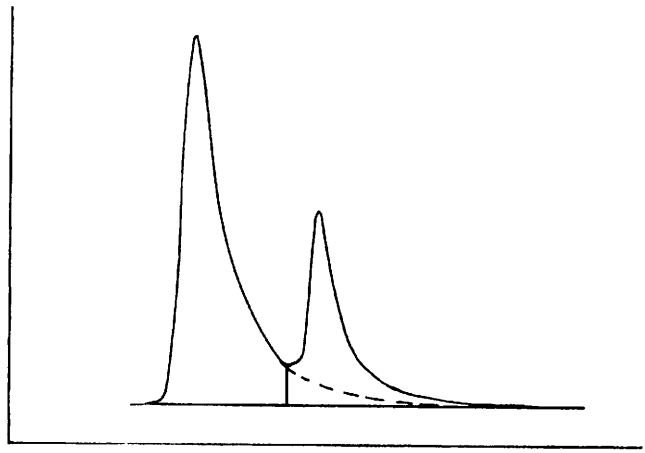


Figure 2.44 *Asymmetry causes peak tail area to be lost under, and added to, the next peak*

Other measurement techniques which tolerate asymmetry, height × width at 25% *H*, or Condal-Bosch, for example, should be adopted.

Unequal Asymmetry

Problems of asymmetry might be avoided if all peaks were asymmetrical in exactly the same way: then the various formulae could be modified by adding a 'shape factor'. For example, the area of a single symmetrical peak is,

$$\text{Area} = k_i H_i w_{0.5,i} \tag{22}$$

in which H_i is the peak height, $w_{0.5}$ is the peak width at half height and $k_i = 1/0.939$. If the peak is not symmetrical k_i will have a different constant value,

but as long as all peaks are the same shape they will all have the same k_i value which will cancel when peaks are compared,

$$\frac{A_1}{A_2} = \frac{kH_1 w_{0.5,1}}{kH_2 w_{0.5,2}} = \frac{H_1 w_{0.5,1}}{H_1 w_{0.5,2}} \tag{23}$$

or normalized:

$$A_i\% = \frac{kH_i w_{0.5,\mathrm{i}}}{\sum(kH_\mathrm{i} w_{0.5,\mathrm{i}})} \times 100 \tag{24}$$

$$= \frac{kH_\mathrm{i} w_{0.5,\mathrm{i}}}{k\sum(H_\mathrm{i} w_{0.5,\mathrm{i}})} \times 100 \tag{25}$$

$$= \frac{H_\mathrm{i} w_{0.5,\mathrm{i}}}{\sum(H_\mathrm{i} w_{0.5,\mathrm{i}})} \times 100 \tag{26}$$

The real problem with asymmetry is that it is not uniform throughout the chromatogram; it can vary with peak size and from one peak to the next. A different constant, k_i, is required for each peak shape and the above simplifications cannot be made.

The Achievement of Accuracy in Chromatographic Analyses

With so many causes of inaccuracy, it is reasonable to wonder how it is ever possible to make accurate measurements using an integrator. Accuracy is achieved through calibration, but there is a subtle feature embedded in the empirical nature of response factors.

When a response factor R_{meas}, is determined, it is a combination of the true response factor and an empirical, instrument factor:

$$R_{\mathrm{meas}} = R_{\mathrm{true}} \times R_{\mathrm{instr}} \tag{27}$$

The true response factor R_{true}, relates solute quantity to peak area. It depends only on solute quantity, detector response and mobile phase flow rate if the detector is flow sensitive. The instrument factor R_{instr}, depends on the instrument and experimental stability, *i.e.* on well-conditioned columns, good sample preparation, clean detectors and flow lines, absence of leaks, instrument maintenance and the ability of the analyst to conduct a good experiment. Provided that instrument conditions, chromatogram peak shape, retention, resolution and S/N ratio all remain constant, then R_{instr} and response factors (R_{meas}) remain constant. Peak areas or heights will change only because solute quantity changes, and the results of analyses are accurate. When conditions drift, then R_{instr} drifts and so do the measured response factors, and results have lower confidence limits.

Note that, although preferable, it is not essential that analysis conditions and parameters are at their best, it is only important that they are all constant. In this

way the measurement of response factors empirically compensates for poor chromatography.

Peak Area *vs.* Peak Height

It has been observed that peak height is sometimes a more precise measure of solute quantity than area when the accuracy of each measurement is comparable. The question has been posed why this should be so. Since precision reflects the conduct of the analysis, the question can be rephrased, 'Why should one be more precise than the other?'

Area is the true measure of solute quantity and, ignoring all other factors, is the preferred measure. Height is a substitute which in the days of manual measurement conveniently speeded up measurements and dispensed with much of the drudgery. There were good reasons for choosing height measurement. Being a simpler measure requiring fewer manual constructions to extract, height was more precise provided that good resolution and peak shape were achieved.[62] There is no need to wait for the last peak to elute when measuring height; once the peak top has eluted the analysis can be ended.

Snyder showed that if peaks are symmetrical, height is less affected by overlap than area.[63] Meyer[31] demonstrated that height is a safer choice for measuring small overlapped peaks. If manual measurement shows that height is less precise than area, it is concluded that experimental control should be improved, not that area is a better measure.

But there are shortcomings in measuring height: it ignores most of the information in the detector signal and it has a smaller linear dynamic range than area. Moreover, height is a function of peak width and this depends only on column efficiency, not solute quantity.[64]

Integrators have replaced manual measurements and the advantages of height over area have a changed perspective; ease of measurement is no longer a consideration. Integrators measure height and area in similar ways and analysts can select either without effort. Differences between measurements still mean that some aspect of the analysis procedure is not sufficiently controlled but area is a more sensitive indicator of this than peak height.

The choice of whether to measure area or height when performing quantitative analyses is a balance of several factors:

Noise. Noise affects area more than it does height. It makes the start and end of a peak more difficult to locate and peak measurement tends to be started later and finished earlier. This causes area to be lost at peak flanks, which does not interfere with height measurement. Height is only affected by inaccurate placement of the constructed baseline but in this case the fractional loss of area is usually more than twice the fractional loss in height (see Figure 1.10). When S/N ratio is poor, height is preferred.

Peak Asymmetry. Asymmetry (skew) distorts peak shape and height but not area. Area is always the more accurate measure for asymmetric peaks that are well-resolved. When asymmetry varies from peak to peak, height cannot be used

without careful calibration. It is worth reflecting that chromatograms which show no asymmetry are very rare.

Peak Overlap. No practical, accurate way to separate and measure poorly-resolved peaks has yet been developed. Many analysts therefore decide to make no measurements at all until a minimum resolution is established and they monitor every analysis to check that it has been achieved. This is a good strategy, but it puts some analyses out of bounds if the minimum resolution cannot be reached.

When overlapping peaks are separated by a perpendicular, it is the smaller peak on the tail which suffers the larger proportional error, and the error increases with size ratio, peak asymmetry and resolution until tangent skimming comes into effect and reverses the trend.

Height and area measurements are both perturbed by overlap, but height is not affected until one peak overlaps the maximum of the other. Until then, for peaks separated by a perpendicular, height is accurate and area is not. When the maximum of a peak is overlapped by another, height measurement error creeps in but it is still less than the errors in area measurements.

Tangent skimming is only accurate when the skimmed peak is very much narrower than the host peak. If peaks have similar widths, tangent skimming is less accurate than perpendicular separation; height measurement after perpendicular separation is more accurate than area, but any measurement requires careful calibration, ideally using standards that mimic the overlap and asymmetry of the 'unknown'.

Ultimately, analysts may have to measure both height and area and compare results with standard samples in order to make their choice.

Baseline Drift. Construction of a linear baseline beneath a peak instead of a more accurate, but indeterminate, curved baseline affects both area and height. Area is distorted more. The area error is approximately equal to half of that caused by inaccurate placement of the baseline due to noise – which still leaves the fractional loss in area greater than the fractional loss in height.

Baseline Disturbances. Baseline disturbances that send the signal negative near peaks of interest can cause large baseline placement errors. This affects area more than height but neither measure can be trusted – these disturbances must be separated from peaks of interest with a section of flat baseline.

Detector Non-linearity and Saturation. Height and area are lost but the fractional loss in area is always less than the fractional loss in height because the loss is made at the top of the peak where it matters less for area. Area is better than height for very large peaks, but neither is accurate when peaks are clipped.

In addition, the linear dynamic range of a detector is less for height than for area so the onset of height error starts at lower concentrations.

The best practice is to determine the limits of the detector and stay within them.

Detector Type. If mobile phase flow rate is imprecise, or pulsing, area will be the better measure of peaks produced by mass sensitive detectors like the FID, but height will be more accurate when using flow sensitive detectors such as a UVD. Most flow sensitive detectors are used for LC, however, and gradient solvent composition must be considered: if solvent composition is

not precisely controlled then even for flow sensitive detectors, area is more accurate.

The pulsing problem is reduced by better pump design. Dual headed pumps are effectively pulse-free when used with UVDs. However, pulsing can still be a problem with sensitive, specialist detectors such as the electrochemical detector.

Similar considerations apply to TCDs and temperature programming.

An ASTM study[65] in 1988 showed that LC detectors of fixed wavelength gave better reproducibility of peak areas than variable wavelength detectors. In contrast, the reproducibility of peak height measurements was unaffected by detector type.

Noise Filtration. Instrument time constants must be fast enough to remove random noise but allow the fastest eluting peaks to pass, and RC filters ought to have the same effect (*i.e.* no effect) on all fast peaks. Where time constants do distort peak shape, the effect is to broaden the peaks. This distortion lowers peak height but area is unaffected.

If sampling frequency is not high enough, integrator smoothing techniques based on Savitsky–Golay have been shown to distort both area and height,[66] but height more than area. Since the analyst has no quick way to check that smoothing distortion is absent, area must be the preferred measure.

Provided that noise is not excessive, resolution is good and that the integrator has drawn the baseline in the correct place, it is better to measure area than height. Height is more accurate when S/N ratio is poor (< 30), or what creates a similar effect, when very small peaks litter the baseline close to large peaks of interest and narrow the limits of integration. It may be preferable in HPLC to measure peak height if pump pulsing is affecting peak shape either because the pump is of older design or the detector is particularly sensitive.

If height is a more precise measure than area when the quality of the chromatography is good, it is a warning that the integrator is not programmed correctly. Adjustments should be made to any or all of the principal parameters: sampling frequency, slope sensitivity and baseline range parameters (see Chapter 4).

User Surveys

An ASTM survey of 50 laboratories in 1984[67] set out to compare the precision of height and area measurements. Similar samples were analysed by LC using integrators to measure area, and a mix of integrators and manual techniques to measure height. It was found that when the chromatography was uncomplicated, *i.e.* no small peaks, and all peaks were nearly symmetrical and well-resolved, area gave marginally better accuracy and was more precise than height.

When the analysis contained perturbations due to overlap and detector non-linearity, accuracy and precision began to suffer. If a peak was perturbed by a nearby small baseline disturbance, the accuracy of the peak measurement did not suffer but height was three times more precise than area. Detector non-linearity affected large peaks, which were under-estimated. Small rider peaks were slightly over-estimated. For both of these types of peak, area was both more accurate and more precise.

The surveyors were forced to conclude, however, that the report lacked detailed information about the different integrators and parameter values used for the various measurements. The accuracy of the results could be interpreted in terms of the factors described above, but variations in column performance were evident and it was not clear whether the chromatograms had been processed and measured in an identical fashion, or whether differences in integration had influenced precision. Comparisons were ultimately limited by the different designs and ways in which the integrators were used.

Integrators and computers should always be more precise than manual methods of peak measurement. Manual methods can be more accurate because analysts are unlikely to err in judgement. They are consistent when drawing baselines, selecting tangents or perpendiculars for measuring overlapping peaks, and are better able to see through the effects of noise and to disregard stray peaks.

Integrators and computers will measure every peak they are not instructed to ignore, and may construct highly displaced baselines if there is a signal disturbance near peaks of interest.

The great advantage of integrators and computers over manual measurement is their speed and convenience and for these reasons analysts will persevere with the shortcomings of their measuring techniques – shortcomings that can be mitigated by good chromatography, which is how it should be.

5 References

1. *Quantifying Uncertainty in Analytical Measurement*, A. Williams (Chairman), Eurochem Working Group, Pub: DTI VAM Initiative, 1995, ISBN 0-948926-08-2.
2. *Statistics for Analytical Chemistry*, J.C. Miller and J.N. Miller, Ellis Horwood Series in Analytical Chemistry, John Wiley and Sons, 1985, ISBN 0-85312-655-0.
3. N. Dyson, *Int. Lab.*, 1992, **22**(6), 38.
4. D. Westerlund, T. Fornstedt, M. Thorsteinsdottir and Z. Yu, *LC-GC International*, 1996, **9**(12), 809.
5. N. Dyson, in *Chromatography in the Petroleum Industry*, ch. 13, ed. E.R. Adlard, *J. Chromatogr. Lib.*, **56**, Elsevier Science, Amsterdam, 1995.
6. L.R. Snyder and J.J. Kirkland, *Introduction to Modern Liquid Chromatography*, Wiley-Interscience, New York, 2nd edn., 1979.
7. V.R. Meyer, *J. Chromatogr.*, 1985, **334**, 197.
8. I. Halasz, *Anal. Chem.*, 1964, **36**, 1428.
9. S.R. Bakalyar and R.A. Henry, *J. Chromatogr.*, 1976, **126**, 327.
10. W. Kipiniak, *J. Chromatogr. Sci.*, 1981, **19**, 332.
11. G. Guiochon and C.L. Guillemin, 'Quantitative Gas Chromatography for Laboratory Analyses and Process Control', *J. Chromatogr. Libr.*, **42**, Elsevier Science, Amsterdam, 1988.
12. S.R. Bakalyar and B. Spruce, Rheodyne: Technical Notes 5, Dec. 1983.
13. J.C. Sternberg, *Advances in Chromatography*, Marcel Dekker, New York, 1966, vol. 2.
14. H. Bruderreck, W. Schneider and I. Halasz, *Anal. Chem.*, 1964, **36**, 461.
15. C.J. Cowper and A.J. DeRose, *The Analysis of Gases*, Pergamon Press, Oxford, 1983, vol. 7.
16. J.G. Keppler, G. Dijkstra and J.A. Schols, *Vapour Phase Chromatography, Proc. 1st Symposium*, London, May 1956, Academic Press, London, 1957.

17. P. Deng and F. Andrews, *J. Chromatogr.*, 1985, **349**, 415.
18. P.W. Carr, *Anal. Chem.*, 1980, **52**, 1746.
19. L.M. McDowell, W.E. Barber, and P.W. Carr, *Anal. Chem.*, 1981, **53**, 1373.
20. I.A. Fowlis and R.P.W. Scott, *J. Chromatogr.*, 1963, **11**, 1.
21. ASTM E 685-79, 1st edn., Philadelphia PA 190103, USA, 1981.
22. D. MacDougall (Chairman), *Anal. Chem.*, 1980, **52**, 2242.
23. H.H. Kaiser, *Anal. Chem.*, 1970, **42**, 26A.
24. R.R. Williams, *Anal. Chem.*, 1991, **63**, 1638.
25. T. Petticlerc and G. Guiochon, *J. Chromatogr. Sci.*, 1976, **14**, 531.
26. D.T. Rossi, *J. Chromatogr. Sci.*, 1988, **26**, 101.
27. A. Westerberg, *Anal. Chem.*, 1969, **41**, 1770.
28. E. Proksch, H. Bruneder and V. Granzener, *J. Chromatogr. Sci.*, 1969, **7**, 473.
29. V.R. Meyer, *LC-GC International*, 1994, **2**, 94.
30. V.R. Meyer, *J. Chromatogr. Sci.*, 1995, **33**, 26.
31. V.R. Meyer, *Adv. Chromatogr.*, 1995, **35**, 383.
32. V.R. Meyer, *Chromatographia*, 1995, **40**, 15.
33. V.R. Meyer, *Chirality*, 1995, **7**, 567.
34. L.R. Snyder, *J. Chromatogr. Sci.*, 1972, **10**, 200.
35. R.A. Lansdowne, R.W. Morosani, R. Herrmann, R.M. King and H.G. Schmus, *Anal. Chem.*, 1972, **44**, 1961.
36. J.P. Foley, *J. Chromatogr.*, 1987, **384**, 301.
37. B.G.M. Vandeginste and L. DeGalan, *Anal. Chem.*, 1975, **47**, 2124.
38. A.N. Papas and M.F. Delaney, *Anal. Chem.*, 1987, **59**, 54A.
39. A.N. Papas and T.P. Tougas, *Anal. Chem.*, 1990, **62**, 234.
40. A.N. Papas, J. Elling and D. Noble, *Anal. Chem.*, 1995, **67**, 617A.
41. J. Novak, K. Petrovic and S. Wicar, *J. Chromatogr.*, 1971, **55**, 221.
42. N.S. Wu and W. Wei. *Chromatographia*, 1992, **34**, 450.
43. 'Peak Fit', Jandel Scientific, San Rafael, California.
44. 'Peak Calc', EDV Consulting, Spabrucken, Germany.
45. J.R. Torres-Lapasio, M.C. Garcia-Alvarez-Coque and J.J. Baeza-Baeza, *Anal. Chim. Acta*, 1997, **348**, 187.
46. R.J. Hunt, *J. High Res. Chromatogr., Chromatogr. Commun.*, 1985, **8**, 347.
47. J.M. Davis and J.C. Giddings, *Anal. Chem.*, 1983, **55**, 418.
48. M. Martin and G. Guiochon, *Anal. Chem.*, 1985, **57**, 289.
49. A.H. Anderson, T.C. Gibb and A.B. Littlewood, *Chromatographia*, 1969, **2**, 466.
50. A.H. Anderson, T.C. Gibb and A.B. Littlewood, *J.Chromatogr. Sci.*, 1970, **8**, 640.
51. J.T. Lundeen and R.S. Juvet, Jr., *Anal. Chem.*, 1981, **53**, 1369.
52. Z. Hippe, A. Bierowska and T. Pietryga, *Anal. Chim. Acta*, 1980, **122**, 279.
53. M.R. Schure, *J. Chromatogr.*, 1991, **550**, 51.
54. P. Gebaur and P. Bocek, *Anal. Chem.*, 1997, **69**, 1557.
55. F. Dondi, M.C. Pietrogrande and A. Felinger, *Chromatographia*, 1997, **47**, 435.
56. M.L. Phillips and R.L. White, *J. Chromatogr. Sci.*, 1997, **35**, 75.
57. B. Vandeginste, R. Essers, T. Bosman, J. Reijnen and G. Kateman, *Anal. Chem.*, 1985, **57**, 971.
58. J.G. Dorsey, J.P. Foley, W.T. Cooper, R.A. Barford and H.G. Barth, *Anal. Chem.*, 1992, **64**, 353R.
59. R.E. Synovec, E.L. Johnson, T.J. Bahowick and A.W. Sulya, *Anal Chem.*, 1990, **62**, 1597.
60. M. Goedert and G. Guiochon, *Chromatographia*, 1973, **6**, 39.
61. B. Jover and J. Juhasz, *J. Chromatogr.*, 1978, **154**, 226.

62. D.W. Grant and A. Clarke, *Anal. Chem.*, 1971, **43**, 1951.
63. L.R. Snyder, *J. Chromatogr. Sci.*, 1972, **10**, 200.
64. R.A. Keller, *J. Chromatogr. Sci.*, 1973, **11**, 223.
65. R.E. Pauls, R.W. McCoy, E.R. Ziegel, G.T. Fritz, D.M. Marmion and D.L. Krieger, *J. Chromatogr. Sci.*, 1988, **26**, 489.
66. S.P. Cram, S.N. Chesler and A.C. Brown, *J. Chromatogr.*, 1976, **126**, 279.
67. R.W. McCoy, R.L. Aitken, R.E. Pauls, E.R. Ziegel, T. Wolf, G.T. Fritz and D.M. Marmion, *J. Chromatogr. Sci.*, 1984, **22**, 425.

CHAPTER 3

Manual Measurement of Peaks

1 Representation of the Detector Signal by Chart Recorder

Strip Chart recorders have been discussed elsewhere.[1-3] The basic requirements of a good recorder are a fast linear response time of less than half a second full scale deflection, high input impedance, variable spans including 0–1 mV, 0–10 mV and 0–1 V, and chart speeds covering the range 0.5–10 cm min^{-1}. It should also have a pen that draws a continuous fine line and preferably one that does not run dry (this does not exist). Commonly used chart widths are 25–30 cm. Latterly, conventional laboratory strip chart recorders have been emulated by computer software.

When peaks are measured manually they are normally recorded at fast chart speeds and at attenuations selected to maximize peak height without overshooting the chart.

Strip chart recorders can distort the detector signal in a number of ways:

Slow Response Time

The time taken by the pen to deflect full scale is called the response time. It is determined by the system electronics, and the inertia of the pen head and the drive mechanism.

If the response time is slow, the pen will lag behind the detector signal as a peak emerges and 'meet it on the way down'. If the detector signal returns to baseline as fast as it departed, the pen will lag behind the signal on the way down too. The recorded peak will be smaller in height and area than it should be and the retention time will be increased (Figure 3.1).

Recorders used for chromatography have a response time of 0.5 s or less. Peak widths measured at half height must be slower than this.

Non-linear Signal Response

As the pen head draws a peak it moves through periods of acceleration, maximum speed and deceleration which creates a non-linear response over the chart width (see Figure 3.2) rather than the linear response suggested by Figure 3.1. Distortion

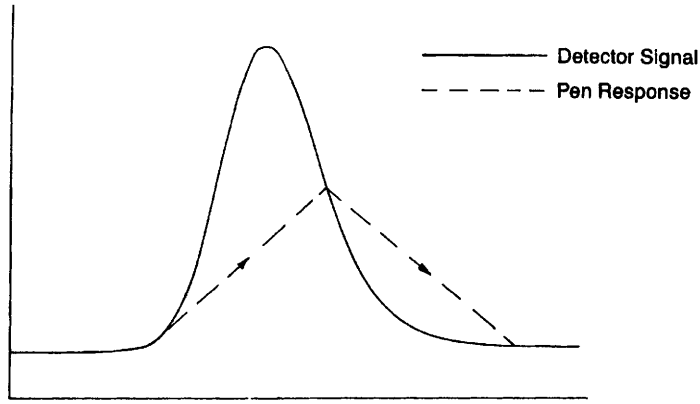


Figure 3.1 *When recorder response is slow, the pen lags behind the signal*

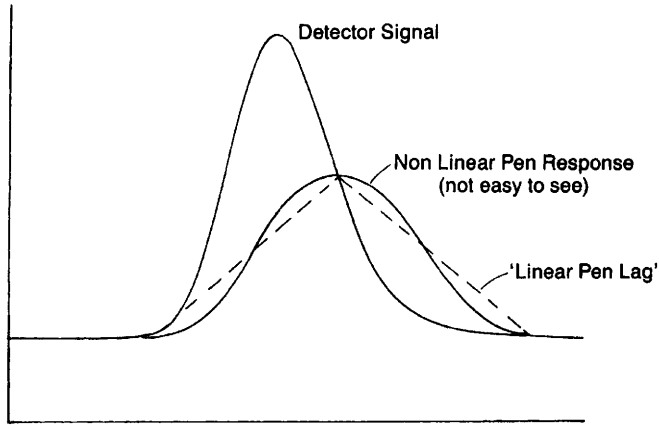


Figure 3.2 *Non-linear slow recorder response*

resulting from slow and non-linear response can escape visual detection, especially with WCOT peaks.

Pen Head Damping

Damping circuitry in the form of RC filters is built into the system electronics to prevent pen overshoot if the signal changes suddenly (when changing attenuation or recorder span for example). It adds 'extra column broadening' to fast peaks but does not change peak area (Figure 3.3).

Amplifier Noise

All electronics have some random noise which adds to the system noise superimposed on the solute signal. Such noise makes it harder to distinguish the start and end of peaks from the background noise. In a well-maintained recorder of good

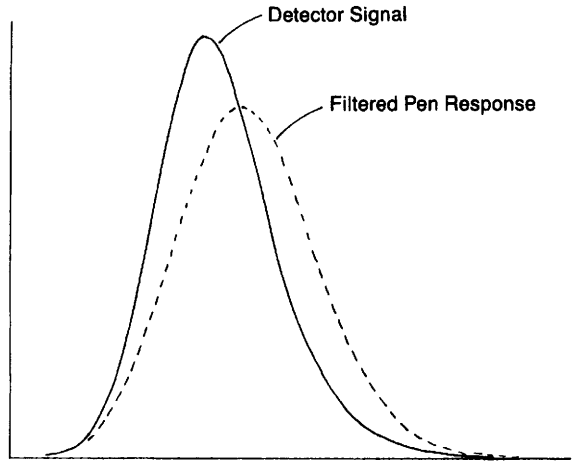


Figure 3.3 *Effect of 'RC' noise filtering. Areas are the same, shapes, heights, and retention are not*

specification, this type of noise is only a problem when measuring trace components where the highest sensitivities are required.

So far, strip chart recorders have contained little computer logic to allow software smoothing (though this will change). Noise is filtered electronically (by more RC filters) resulting in further peak broadening and skewing of shape.

Manual Measurement of Noise

Noise amplitude is measured manually on a stretch of baseline by drawing tangent lines above and below the baseline and measuring the vertical separation (see Figure 1.1). This is much easier than making many individual measurements at various baseline points and working out the average noise amplitude, or its standard deviation, σ. The envelope measurement is approximately equal to 6σ.

Dead Band

Recorder dead band is the largest change in input signal which does not produce a pen deflection. Laboratory recorders in good working order have relatively small dead bands which can be ignored in practice; worn mechanical parts make them larger. When the baseline drifts off the bottom of the chart and subsequently fails to record baseline events, it is in effect a dead band error.

Dead bands cause the onset of a positive slope to be recorded late, the loss of negative slope at baseline to be recorded early and area to be lost. The minimum detectable quantity is larger than it should be. In effect, a slice has been shaved off the bottom of the chromatogram (Figure 3.4).

Losing 1% from the base of a Gaussian peak through dead band error reduces the peak area by 2.66%.[4] Errors in small peaks are proportionally even greater as dead bands are fixed in size. Peaks may be lost altogether.

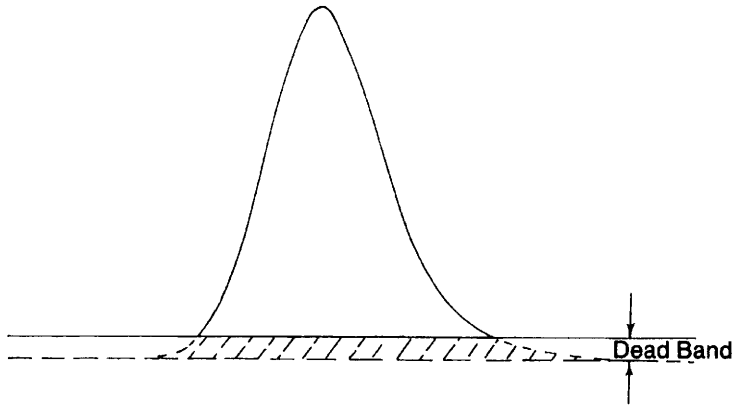


Figure 3.4 *Effect of recorder dead band. Shaded area is lost*

Chart Motor Control

The steady movement of chart paper, essential for time measurement, is maintained either by a discontinuous stepper motor or a continuous synchronous motor. Since it is common practice to improve measurement precision by recording peaks at a fast chart speed, errors of non-uniformity in chart movement directly affect peak area but not peak height. It has been found that precision of width measurement increases with peak width but only to a limiting value; no matter what the height of the peak, maximum width precision will have been achieved once the width at half height reaches 5–6 cm.[5]

Capillary chromatographers have found that the pulsed motion of stepper motors in some recorders creates apparent shoulders on fast peaks. When the number of movement pulses per motor revolution is too low, the steps can be seen.

Attenuator Accuracy

Signal attenuation to maximize peak size requires accurate attenuation. Any errors will be directly incorporated into height or area measurements and both will be affected equally.

The amount of allowed peak expansion is determined by the recorder time constant. Changing the span to increase the height of small, fast peaks requires the pen to travel much further in the same time, the time constant may not be fast enough to allow this without distortion.

2 Measurement Strategies

There are two approaches to manual measurement of chromatograms: use of methods that assume no peak shape, or use of methods based on a peak model, usually Gaussian.

Counting squares, cutting and weighing, and Planimetry are known as 'boundary

methods'. These have long been used by engineers, cartographers, *etc.*, and were quickly adopted by chromatographers. Boundary methods make no assumptions about shape and are therefore equally suitable for symmetric and asymmetric peaks. They are, however, tedious to use and are not generally employed now in working laboratories.

Counting Squares

The chromatogram is recorded on 'squared' chart paper with each peak size maximized by attenuation and chart speed. Baselines, perpendiculars and tangents are drawn by the analyst who then counts the squares enclosed within each peak boundary.

The count is made in two stages: first, the number of squares not intersected by a boundary is counted, and to this total is added the sum of part squares through which the peak boundary does intersect.

Counting part squares involves matching those parts which, in the analyst's judgement, add up to a whole square. The matched squares are then added to the sum of the whole squares. Each peak area is multiplied by the attenuation at which it was recorded, and composition of the solution is calculated.

Counting squares requires too much subjective judgement and is completely impractical when there are many peaks. It is not very accurate although practice can bring precision.

Cutting and Weighing

Peaks are carefully cut out and weighed using an accurate balance. If paper thickness and moisture content are uniform the weight is proportional to the peak area.[6] Each weight is multiplied by the attenuation at which the peak was recorded and the composition of the mixture is then calculated.

Small but important peaks riding on the tails of large peaks are a problem. If the small peak is enlarged for ease of measurement, it will be at the expense of the large peak on which it rides.

Chromatograms can be photocopied and the copy worked upon, thus preserving the original trace. Errors in cutting out peaks are not disastrous because the measurement can be repeated using another copy. If the photocopying paper is a heavier grade than the original chart paper, and its weight and moisture content are equally uniform, it will give larger weights for the peaks and so reduce weighing error.

Photocopiers with zoom facilities offer a convenient solution to the tailing peak problem, but magnifying the peak size also magnifies the width of the recorder pen line with consequent loss of measurement precision. It is prudent to check the magnification of the photocopy rather than trust the manufacturer's figure, and small peaks should not be magnified repeatedly or optical distortion will come into effect.

Planimeters[7]

A planimeter is a mechanical device for measuring the area of an irregular plane shape (Figure 3.5). Briefly, the shape of the peak is traced out by a stylus linked to a wheeled scale. When the stylus has traced the complete peak perimeter, the area is read from the scale as the difference between the initial and final readings.

Accuracy is achieved by repeating the measurements and averaging them. Planimeters measure peaks more quickly than counting squares or by cutting and weighing, and can be more precise. The technique is tiring because of the concentration and care required. Good lighting is important and working on a polished surface helps by allowing the user's arm to slide smoothly when tracing the peak.

Errors from the planimeter occur when the measuring wheel skids over the surface of the chromatogram, and this happens, in practice, when the angle between the planimeter arms becomes too small or the user's wrist or elbow slips.

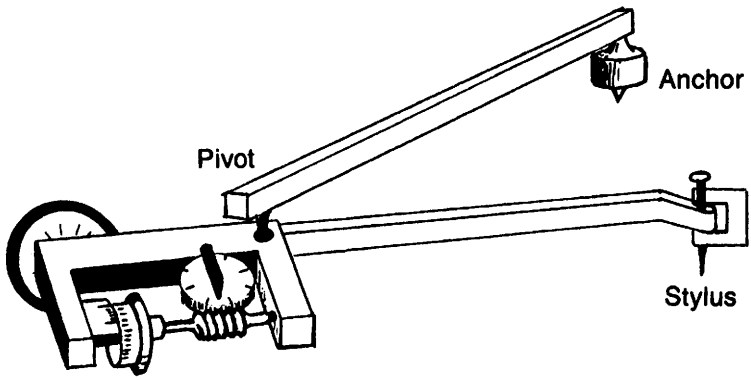


Figure 3.5 *Planimeter*

3 Measurements based on a Peak Model

If a peak can be assumed to have a known shape, its area can be calculated from peak dimensions (see Section 1). Any departure of the peak shape from the peak model introduces errors of accuracy no matter how precise the measurements. These errors can be offset by calibration, but it is always an implicit assumption of this type of calibration that both measurements (standard and unknown) will be equally wrong and the errors will cancel. It is better to improve the chromatography so that such strategies are unnecessary.

Most basic theory of peak measurement comes from an assumption of Gaussian shape, yet few real peaks are Gaussian, many are asymmetric and none have the infinite boundaries which Gaussian theory predicts.

Alternative models have been sought and of these the Exponentially Modified Gaussian (EMG) function has found the widest applicability so far, although it has its limitations too. It does not fit every peak and has infinite limits.

Pencil and Rule Methods

Pencil and rule methods measure various peak heights and widths and compute solute quantity from them. They are quicker than boundary methods but are less accurate for overlapping or asymmetric peaks. Measurements based on the Exponentially Modified Gaussian peak shape give better accuracy for asymmetric peaks once the peaks have been tested for EMG suitability. Pencil and rule methods include:

(1) peak height measurement;
(2) triangulation at peak asymptotes;
(3) peak height × half width, including Condal-Bosch variation[8] and Foley variations (EMG).[9]

Measurement of Peak Height

Peak height measurement was originally intended to be a simple and fast technique relying on the premise that for peaks of fixed shape and width, area is proportional to height (Figure 3.6). This precludes direct peak comparisons within the same chromatogram where width is related to retention time, but peaks can be calibrated by external standards of the same retention and width.

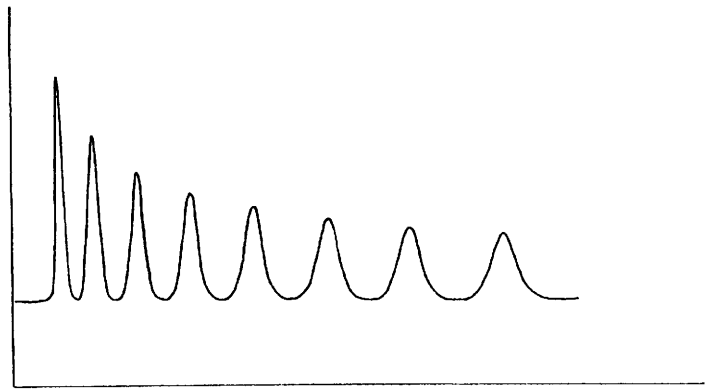


Figure 3.6 *Peaks of equal area but different height*

Height vs. Area Measurements

The great disadvantages of manual measurements compared with digital integration are slowness and imprecision. Measurement of height reduces the processing time (and tedium), making it the preferred alternative. If peaks are symmetrical and well-resolved there is no sacrifice of accuracy and height is measured with greater precision because fewer constructions and measurements are required and so fewer measurement errors are made.

Noise affects the accuracy of height measurements less than area, on chromato-

grams which are noisy ($S/N < 30$ approx.[10]), but otherwise height and area are equally good.

Peak asymmetry and detector non-linearity perturb height measurements more than area[11] and where these are major factors, peak areas should be measured.

Triangulation

Lines are drawn through the points of inflection on each side of the peak (points of maximum slope) to intersect and form a triangle with a baseline drawn beneath the peak. The tangents should not intersect at an angle less than 30° (approximately) as this adds to the imprecision of height measurement, nor greater than 120° (approximately) as this leads to imprecision in measuring base width[5] and width at half height.

The area of the triangle calculated from the base width and height is proportional to the peak area (see Section 1):

$$\text{Area} = 0.5H'w_b/k \qquad k = 0.968 \tag{1}$$

Actual peak height (H) can be measured, in which case $k = 0.798$.

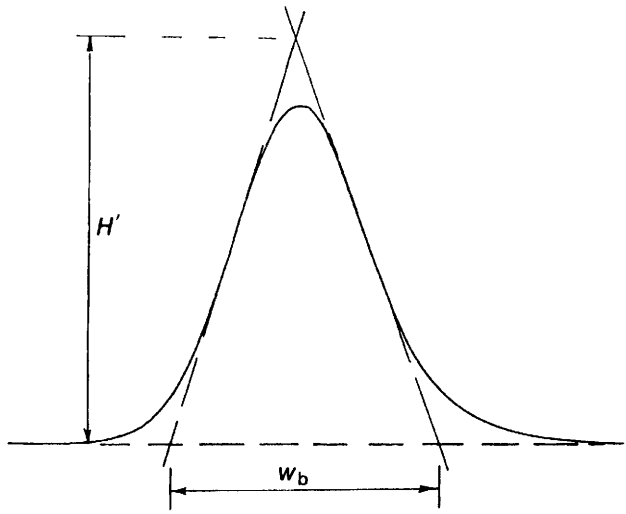


Figure 3.7 *Triangulation*

Triangulation is valid for symmetrical peaks on level baseline (Figure 3.7). Asymmetry and overlap introduce inaccuracies, the worse the peak shape or overlap, the greater the error.

When overlap and asymmetry are not severe, two overlapping peaks can be measured by triangulating the unresolved half of each peak and doubling it to obtain the whole area (Figure 3.8).

For asymmetric peaks, doubling the 'half area' does not give an accurate measure

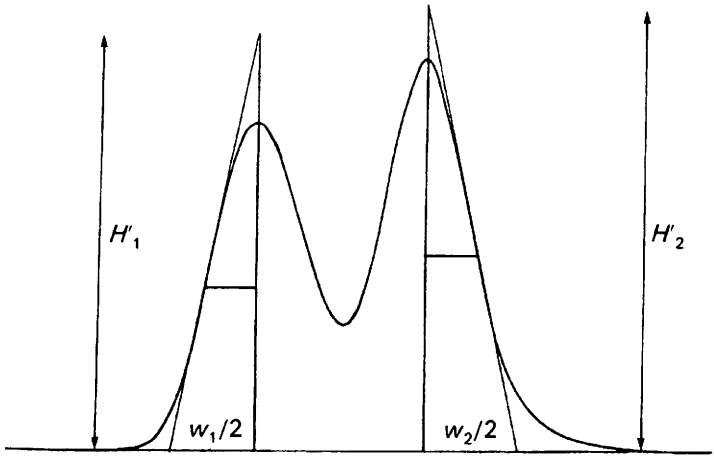


Figure 3.8 *Triangulation of overlapping peaks*

of the total area. If the asymmetry of each peak is similar, an alternative strategy is to scale each half area to the height of the other and add it to the other half.

Calibration errors occur when the peak shape changes with injected sample volume (overloading) and the relationship between constructed triangle and peak area is not linear. A multi-point calibration is required.

Peak Height × Width at Half Height

Symmetrical peak areas are computed from the product of peak height and the peak width measured at half height (Figure 3.9).

$$\text{Area} = Hw_{0.5}/k \qquad k = 0.939 \tag{2}$$

The method works well for resolved and symmetrical peaks but asymmetry and

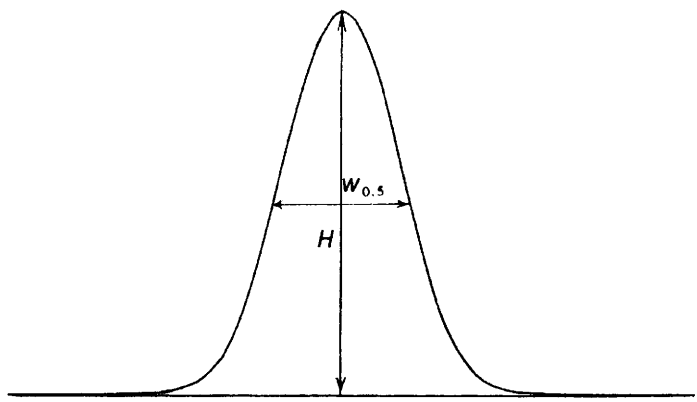


Figure 3.9 *Peak height × width at half height*

overlap reduce accuracy. The width at half height is easy to locate and measure but it is not the optimum width for maximum measurement precision.[5] Determination of this width is complicated, and depending on peak shape will occur at different heights; this makes it an impractical measure to use, and the width at half height is used instead.

Area measurement of a known peak can be speeded up by attenuating the peak by half as it elutes (Figure 3.10).

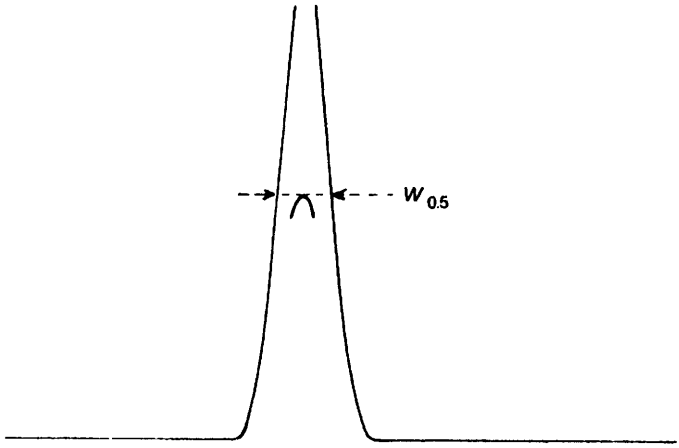


Figure 3.10 *Attenuating the peak top by half shows where to measure* $w_{0.5}$ (Reproduced with kind permission from *Techniques of Organic Chemistry*, vol. XIII, Interscience Publishers, 1968)

A horizontal tangent across the peak maximum provides the width at half height to be measured.[3] Obviously, attenuation must be accurate and must not take place until after the detector signal has passed halfway to the peak maximum.

Measurement of peak width on a sloping baseline requires some geometrical construction to ensure that the width is always measured horizontally (Figure 3.11).

Measurement of area or height on a sloping baseline is subject to error when the baseline is curved and not linear as constructed. Any attempts to estimate the curvature by eye are likely to be subjective and it is best to use the straight line.

Condal-Bosch Variation[8]

A simple method designed to compensate errors of asymmetry and measure the actual peak area ($k = 1$) was devised by Condal-Bosch. Instead of using the peak width at half height, the average of the widths at 15% and 85% height is used (see Figure 3.12):

$$\text{Area} = H.0.5[w_{0.15} + w_{0.85}]/k \tag{3}$$

since $k = 1$,

$$= H.0.5[w_{0.15} + w_{0.85}] \tag{4}$$

For any other width combinations of 'x%' and '$(100 - x)$%', the constant k is not equal to 1 (Figure 3.13).

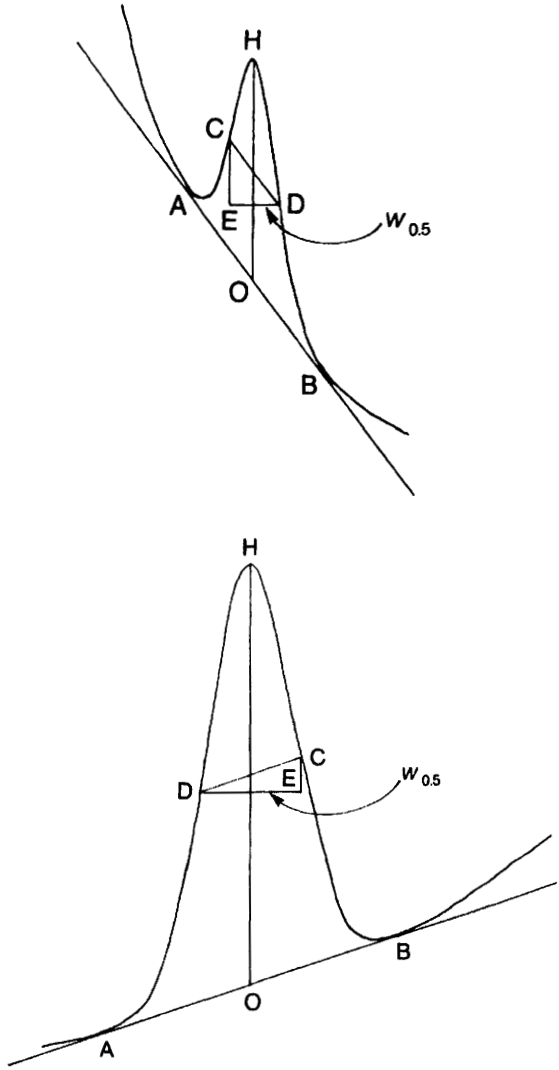


Figure 3.11 *Width at half height must be measured horizontally;* CD *is drawn parallel to* AB *through* $\frac{1}{2}$HO; ED *is horizontal*

The disadvantage of this method is that it is limited to peaks sufficiently resolved to allow the 15% width to be measured with confidence.

The Foley Variations[9]

Foley developed equations for peak area based on the EMG function. He showed that for 90% of the asymmetric (LC) peaks in his experiments, peak area could be expressed by,

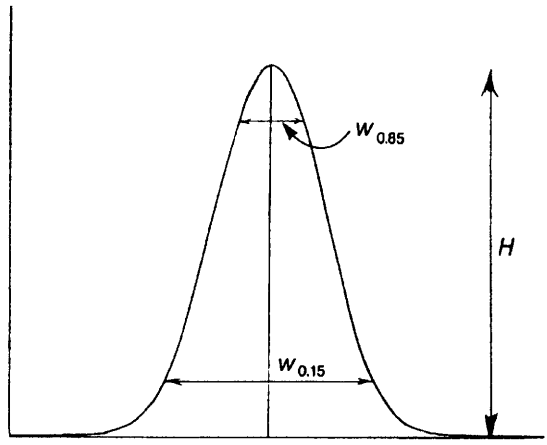


Figure 3.12 *Condal-Bosch area measurement*

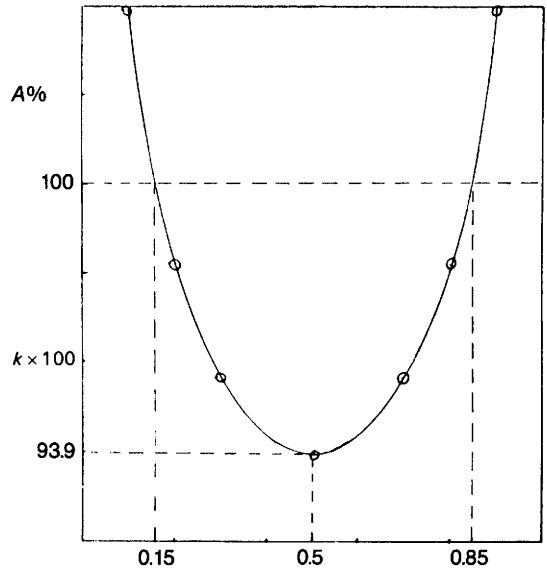


Figure 3.13 *The* 0.15 *and* 0.85 *heights are the only combination that measure the true peak area* (i.e. 100% A). *Width at half height gives* 0.939 A (*see Equation* 2)

$$\text{Area} = 0.753 H.w_{0.25} \tag{5}$$

where H is the peak height and $w_{0.25}$ is the peak width at 25% of its height (Figure 3.14).

Wu *et al.*[12] verified this equation and showed it to be accurate to within 0.6%.

Equation 5 uses only one width measurement, so the calculation is no less convenient than height × width at half height. It is quicker than Condal-Bosch, and the 25% height suffers less from peak overlap than the 15% height.

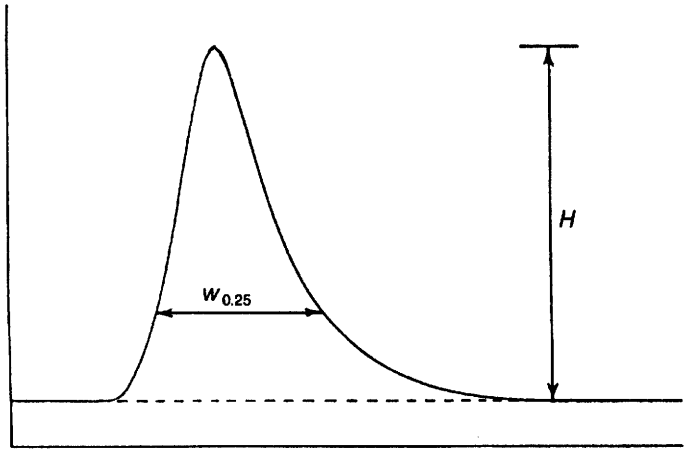


Figure 3.14 *Foley (EMG) measurement of peak area*

Measurement of Overlapping Asymmetric Peaks

Other equations measure the areas of overlapping asymmetric peaks from width measurements taken at greater peak heights to evade encroaching overlap. These require independent measurement of the asymmetry ratio B/A,[9] but they are more accurate than equations which assume Gaussian shape:

$$\text{Area} = 1.07 H w_{0.5} (B/A)^{0.235} \tag{6}$$

also

$$= 1.64 H w_{0.75} (B/A)^{0.717} \tag{7}$$

Peak Shape Tests

To prove EMG suitability, shape tests based on area are recommended. Areas calculated from the three different Equations 5, 6 and 7 should be within 2–3% of each other. Similar tests should also be applied to validate peaks before using Gaussian equations, of course, but they rarely are.

4 Errors of Manual Measurement

Delaney[13] used an EMG function to generate computer peaks of increasing asymmetry (τ/σ ranged from 0 to 4) and studied the determinate errors in using the foregoing equations to measure peak area. The results of this study are summarized in Figure 3.15.

Condal-Bosch was accurate to within 2% over the whole range. The EMG-equations are accurate because the measured peaks are EMG-based. All the Gaussian-based methods become increasingly inaccurate, reporting low areas, with height showing the most sensitivity to peak shape. Triangulation starts off being more accurate than height × width(0.5) but becomes worse at higher asymmetries ($\tau/\sigma > 2$) because the peak base broadens faster than the width at half height.

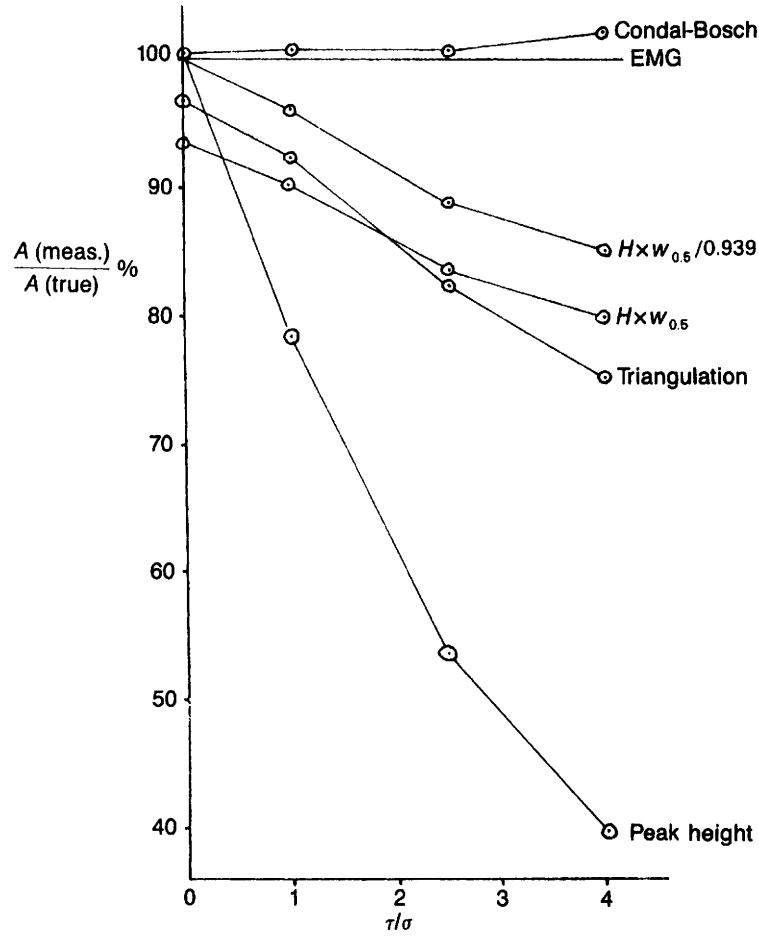


Figure 3.15 *Effect of peak asymmetry on manual measurements*
(Reproduced with kind permission from *Analyst (London)*, 1982, **107**, 606)

Figure 3.15 additionally implies that peaks should only be calibrated against others having the same shape, and this in turn implies calibration against peaks of the same size if shape varies with injection volume.

Ball *et al.*[5,14] made a systematic study of the geometrical construction errors made when peaks are measured manually. The principal constructions are shown in Figure 3.16.

With each step there is an associated measurement or placement error. The error in measuring peak area is a combination of all these errors. In general, the variance of the area measurement is the sum of the component variances and so:

$$\frac{\Delta A}{A} = \sqrt{\left(\frac{\Delta B}{A}\right)^2 + \left(\frac{\Delta H}{A}\right)^2 + \left(\frac{\Delta H_{0.5}}{A}\right)^2 + \left(\frac{\Delta w}{A}\right)^2} \qquad (8)$$

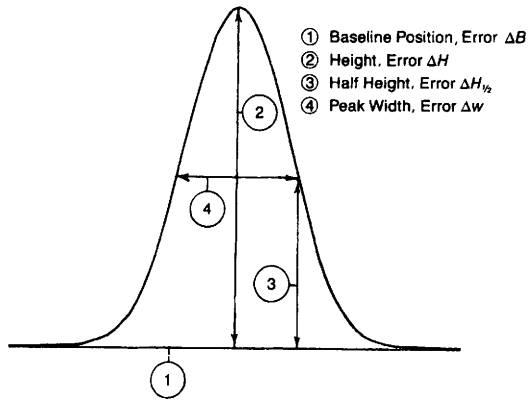


Figure 3.16 *Manual measurement errors*

Peak height only contains the errors in positioning the baseline and measuring height:

$$\frac{\Delta H}{H} = \sqrt{\left(\frac{\Delta B}{H}\right)^2 + \left(\frac{\Delta H}{H}\right)^2} \tag{9}$$

Comparing equations 8 and 9 confirms that peak height precision will be greater than that of area because fewer measurements and errors are involved.

Optimum Peak Shape

There is an optimum peak shape for maximum area precision which occurs when the ratio of peak height to width at half height is in the range 2 to 5 (Figure 3.17).

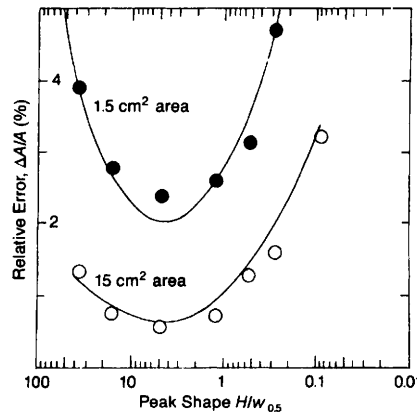


Figure 3.17 *Change in precision of peak area with change in ratio of height to width for peaks of different areas*
(Reproduced with kind permission from *Anal. Chem.*, 1968, **40**, 129)

Precision is greater for large peaks than for small (the fractional error is inversely proportional to $\sqrt{A}$), but there is a limit to the increase in precision achieved by increasing chart speed. Precision increases until the width reaches 3–5 cm but then becomes constant, independent of peak size (Figure 3.18).

There is no optimum height. Precision of height measurement increases with peak height. Provided that peaks are symmetrical the fractional error is inversely proportional to peak height, so for a given area, the narrower the peak the better.

There is, however, an optimum height at which to measure peak width[18] in order to maximize precision in height × width estimates. For a Gaussian peak this is at 36.8% of peak height ($h_{opt} = e^{-1}H$). At this height, peak width = $2\sqrt{2}\sigma$ and height × width gives 1.128 × true area. Manual measurements of peak width above 75% H and below 10% H are inaccurate because of the relative difficulties in constructing the measurement. Width accuracy measured between these heights does not vary much from the optimal.

A comparison of the precision of methods is given in Table 3.1. Height is the best and triangulation generally the worst.

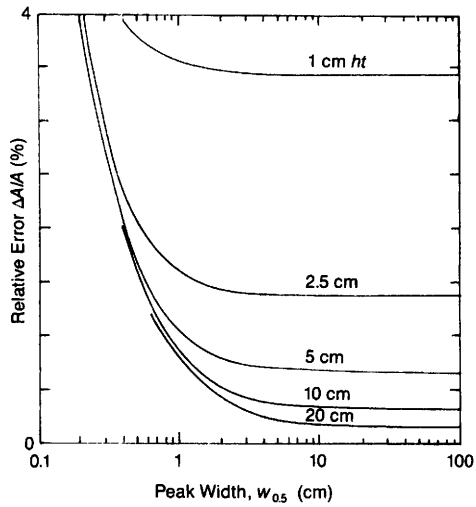


Figure 3.18 *Variation of precision with peak width for peaks of different heights* (Reproduced with kind permission from *Anal. Chem.*, 1968, **40**, 129)

Table 3.1 *Comparison of manual methods*

Method	*Precision %*	*References*
Cut & Weigh	1.7	15, 16
Planimetry	0.5–4	6, 15, 16
Height	0.25–0.9	5, 16
Triangulation	1.1–4.1	9, 15, 17
Height × $w_{0.5}$	0.8–2.5	5, 15, 16, 17
Condal-Bosch	2.0	8, 13
EMG	1.4–2.0	9

Advantages and Disadvantages of Manual Peak Measurement

All manual measurements are less convenient than automatic methods because the chromatographer is required to do some work. However, using a good chart recorder has some real benefits:

(1) The chromatogram is the most direct representation of the detector signal. It has not been digitized, processed and reconstructed as happens in integrators and computers.
(2) The analyst makes the decisions where peaks begin and end, and where baselines, perpendiculars and tangents should be drawn. The choice between perpendicular and tangent separation of marginally sized peaks can be made consistently over a series of experiments.
(3) The analyst is better than a computer in discriminating between noise and peaks, and in observing and rejecting peaks which are not part of the current analysis.
(4) Manual methods are cheap; only the planimeter need be purchased.
(5) Processing of chromatograms can be done at any convenient time, even away from the laboratory.
(6) If peaks do not go 'off scale', the analysis can be left unattended.

There are disadvantages:

(1) Manual measurements require great care to achieve good precision. The methods are time-consuming and the results are not as precise as those measured by integrator or computer.
(2) Small peaks on tails which cannot be enlarged are measured with even less precision.
(3) Overlapping peaks are not reliably measured although manual EMG-based measurements can be more accurate than integrator measurements which use perpendiculars and tangents.
(4) If important peaks do go 'off scale' the analysis must be repeated, which is not always possible if no sample remains, or it has degraded.
(5) Measuring peaks manually is always tedious.

5 References

1. R.L. Grob, *Modern Practice of Gas Chromatography*, John Wiley and Sons, New York, 2nd edn., 1985.
2. R.G. Bonsall, *J. Gas Chromatogr.*, 1964, **2**, 277.
3. O.E. Schupp, 'Gas Chromatography', *Techniques of Organic Chemistry*, vol. XIII, Interscience, New York, 1968.
4. A.B. Littlewood, *Z. Anal. Chem.*, 1968, **236**, 39.
5. D.L. Ball, W.E. Harris and H.W. Hapgood, *Anal. Chem.*, 1968, **40**, 129.
6. W.J.A. Vanden Heuval and A.G. Zacchei, *Advances in Chromatography*, Marcel Dekker, New York, 1976, vol. 14.
7. J. Janik, *J. Chromatogr.*, 1960, **3**, 308.

8. L. Condal-Bosch, *J. Chem. Educ.*, 1964, **41**, A235.
9. J. Foley, *Anal. Chem.*, 1987, **59**, 1984.
10. T. Petticlerc and G. Guiochon, *J. Chromatogr. Sci.*, 1976, **14**, 531.
11. J.J. Kirkland, W.W. Yau, H.J. Stoklosa and C.H. Dilks Jr., *J. Chromatogr. Sci.*, 1977, **15**, 303.
12. N.S. Wu, A.M. Qui and G.W. Zhao, *Chromatographia*, 1990, **29**, 248.
13. M.F. Delaney, *Analyst (London)*, 1982, **107**, 606.
14. D.L. Ball, W.E. Harris and H.W. Hapgood, *Sep. Sci.*, 1967, **2**, 81.
15. H.M. McNair and E.J. Bonelli, *Basic Gas Chromatography*, Varian Associates, Walnut Creek, CA, 1968.
16. J.M. McGill, F. Baumann and F. Tao, *Previews and Reviews*, Varian Associates, Walnut Creek, CA, 1967.
17. D.W. Grant and A. Clarke, *Anal. Chem.*, 1971, **43**, 1951.
18. A.S. Said, *Theory and Mathematics of Chromatography*, Alfred Hüthig, Heidelberg, 1981, p. 31.

CHAPTER 4

Digital Integrators

1 A Brief History of Integrators

Strip Chart Recorder Techniques

As well as the manual techniques of peak measurement described in Chapter 3, attachments to recorders were designed to allow automatic measurement. The most important of these techniques was the 'disk integrator' which went out of large scale production in the mid 1970s but which can still be made to order.

The small trace at the side of the chromatogram in Figure 4.1 is the integral of the detector signal: peak area is counted as the number of full scale deflections of this trace. The disk integrator was widely used but had the same range as the recorder pen. It suffered when the peak was attenuated or went off scale. If the chart speed was too slow the integral trace merged on itself making it difficult to count deflections, although every tenth was slightly bigger to allow quick counting. Area count resolution was limited to three significant figures, sometimes only two.

Electromechanical Counters[1,2]

The rise in signal level above a pre-set (baseline) value was used to drive a counter whose count rate was proportional to the rise. When the signal returned to the baseline, *i.e.* to the pre-set value, the counter tripped and printed the peak area.

If the signal did not return to the baseline, the analyst waited until the end of the peak emerged on the chart recorder and pressed an 'end' button to force the count to be printed. If the signal drifted below the original baseline level the counter tripped at the pre-set level and lost the peak tail. The analyst, by eye, forced area printout at valleys to obtain areas of overlapping peaks.

The disadvantages of electromechanical counters always outweighed the benefits. The disadvantages are:

(1) Peaks were only measured above a pre-set level. This projected a horizontal baseline beneath peaks. The integrator could not cope with drifting baselines (Figure 4.2).
(2) The analyst was always in attendance to terminate integration at valleys or

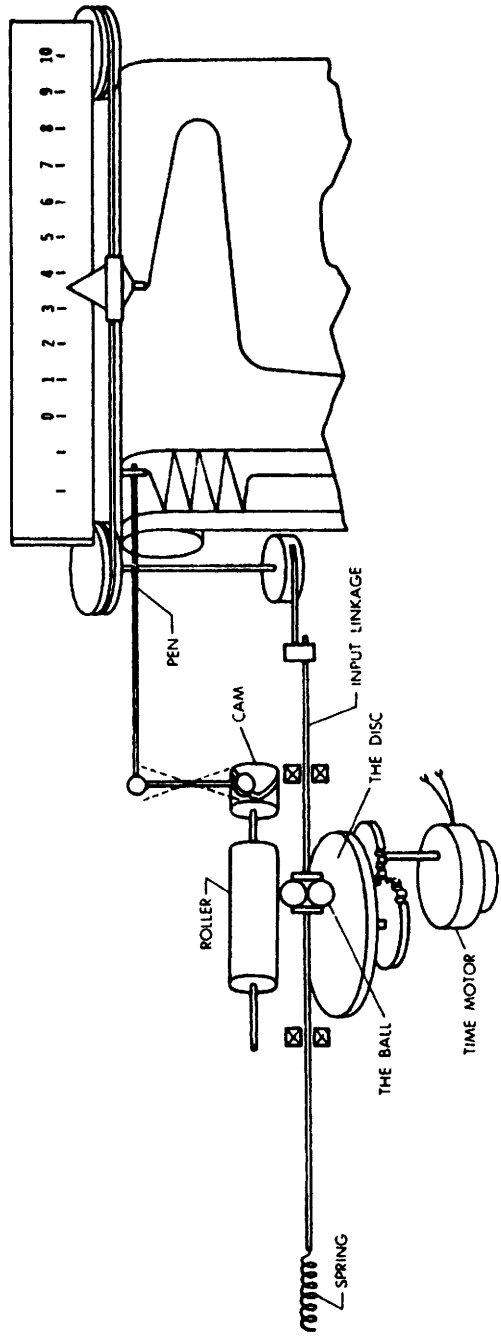


Figure 4.1 *Disk integrator*
(Reproduced with kind permission from *Modern Practice of Gas Chromatography*, ed., R.L. Grob, Wiley-Interscience, New York, 2nd edn., 1985)

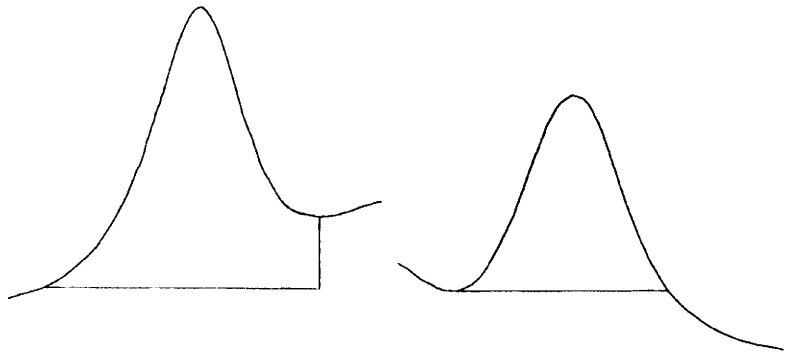


Figure 4.2 *Effect of sloping baseline on electromechanical integration of peaks*

where the baseline had drifted upwards, and to reset the baseline level just before the start of each peak.

(3) Being electromechanical, the counter had limited dynamic range and a finite count rate which, like a slow time constant, could not keep up with fast peaks and so measured low areas without any warning to the analyst (Figure 4.3).

(4) The baseline zero was adjusted to sit just above the noise level to prevent noise peaks triggering counts. It also sat above the smallest peaks and the base of wanted peaks which were not included in the count (Figure 4.4).

(5) Retention time was not measured: it had to be taken from the chromatogram by rule. The operator added his own notes to the printer report and stapled it to the recorder trace.

(6) There were no calculations. These were later made manually on area counts of limited numeric range. Count resolution was typically up to five significant figures, but measurement inaccuracies meant that they were not all believable.

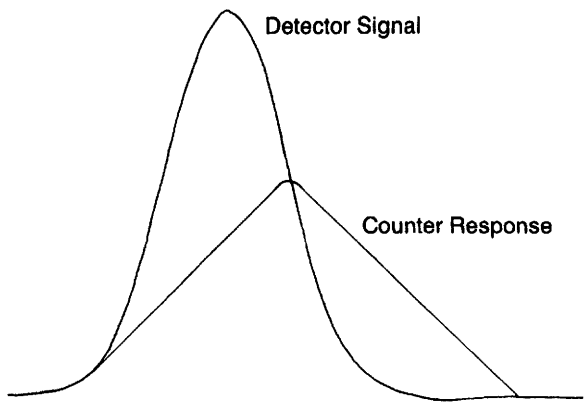


Figure 4.3 *Counter has finite rate which could lag behind the detector signal*

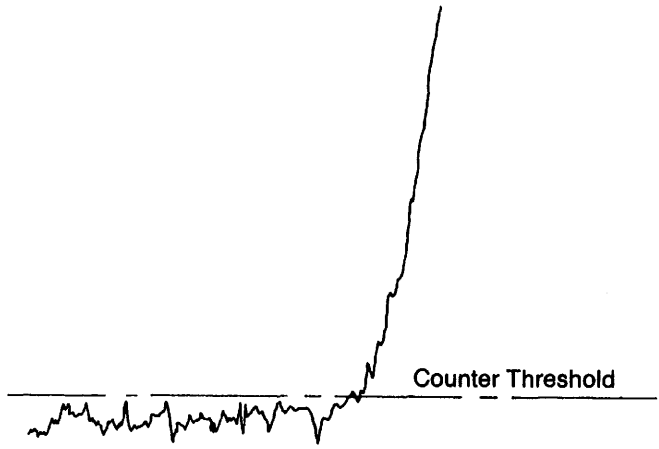


Figure 4.4 *Counter zero set just above noise level*

(7) The integrators were unreliable, and component variations between similar models introduced a repeatability error.

Analysts became distrustful of these instruments and since that time, integrators have indicated with increasing detail how they have measured peaks and from what boundaries.

Electronic Integrators

TTL Instruments – The First Electronic Integrators

In 1968 Hewlett Packard launched the 3370 series integrator, the first transistor–transistor logic (TTL) model with a wide linear dynamic range. This allowed large and small peaks in the same chromatogram to be measured without scale adjustment. It expanded the number of samples which could be processed without intervention by the analyst and could be left to operate unattended with confidence.

The HP 3370 was expensive and quickly incurred competition from similar integrators manufactured by Vidar Autolab (later Spectra Physics) and Infotronics (later Laboratory Data Control).

These integrators were totally electronic. They used voltage to frequency conversion of the analog detector signal but could not store data, which meant that their peak decisions had to be instantaneous. TTL integrators often false-tripped on noise peaks and they could not cope with baseline drift during peak elution. The baseline was simply defined as the last signal level before the start of the peak, and projected horizontally under the peak or group of peaks.

In spite of these limitations TTL integrators were the first successful integrators and a large number were sold. Many of their limitations were offset by good chromatography and isothermal analyses (it was mostly GC at that time).

Many modern integrator features were introduced by these instruments:

(1) use of V/F data conversion;
(2) automatic perpendicular separation of unresolved peaks with valley comments printed on the report;
(3) measurement of retention time;
(4) peak detection or 'event' marks to show the start and end of peaks were added to the detector signal and displayed on the chromatogram by the chart recorder;
(5) minimum peak area threshold filtered small peaks;
(6) noise discrimination based on peak width cancelled false peak starts;
(7) slope sensitivity was introduced. The integrator was able to monitor true baseline;
(8) tangent skimming was introduced but not too successfully;
(9) the first calculation, normalization or area%, was added;
(10) panel indicators showed experiment status such as Reset or Analysis, PEAK (integration), positive/negative slope, elapsed analysis time, and peak area.

Microprocessor-based Integrators[3]

In 1973 Spectra Physics launched the Minigrator, the first microprocessor-based integrator. It was followed by the Infotronics 309 series and Varian 100 series integrators.

The Minigrator had memory to store four peaks and place a trapezoidally-corrected baseline beneath them. When a larger group was measured, a horizontal baseline was projected under the first few peaks and a trapezoidal baseline drawn under the last four. It monitored trends and updated its parameters to allow for increasing peak width as the analysis progressed.

Tangent skimming of up to three rider peaks was possible (the solvent peak made up the fourth), and an inbuilt calculator allowed the analyst to make a variety of manual calculations such as area% and internal standard.

The integration report was obtained from a printer, but the chromatogram was still displayed on a separate chart recorder.

Integrators with Printer Plotters

Barely two years after the launch of the Minigrator, Hewlett Packard launched the 3380 series integrator with its own inbuilt chart recorder. The chart recorder was in fact an alphanumeric printer plotter which presented results and chromatogram on one piece of paper.

Peak memory capacity was increased to 54 peaks and calculations of area%, internal standard and external standard included response factors determined by self-calibration.

Chart paper advancement stopped at the end of the analysis and saved a lot of paper.

The Impact of the Microcomputer

The Minigrator and the HP 3380 integrators introduced most of the features which exist in contemporary integrators. The introduction of microcomputers in the late 1970s and in particular the arrival of the 16 bit IBM PC in 1982 has given rise to an alternative type of software-based integrator. Developments since then have been almost exclusively electronic or computer-based. The peak processing algorithms used by integrators have changed in detail but not much in principle; overlapping peaks are still separated by perpendiculars and tangents. What has happened is that integration software has moved into computers and networks where it can interact with spreadsheets, word processors and statistics packages.

Stand-alone integrators are dedicated data processors. Computers are inherently more flexible; they change their role with their software. The attraction of a computer-based integrator over a dedicated one is based on two features:

(1) Most scientists want the latest microcomputer. Individual purchases within large companies are often restricted because of concern over uncontrolled choice and the lack of systems analysis skills. Purchase of a microcomputer to run 'integration software' is a legitimate justification.
(2) Computers have many other uses including secondary processing by means of powerful mathematics and statistics packages, report writing and distribution along networks.

In spite of this, stand alone integrators continued to dominate the market throughout the 1980s although their share declined until the quality of integration software caught up. By 1995 they were the minority seller, bought for specific applications, analysts no longer 'equipped the lab' with them. The development of PC integration was slower than expected, reflecting the smaller number of programmers engaged in writing integration software, rather than the number of computer manufacturers or cost of hardware. To begin with, integrator designers had a considerable lead over the programmers. Integrator programs had been developed for some time, they were comprehensive, 'bug'-free and easy to use. Integration software has caught up and microcomputers have exploited their major advantages – low hardware costs and the vast range of programmers, most of whom have never heard of a chromatograph.

It is still possible that first order data processing will be built into the chromatograph, thereby saving the costs of a separate power supply, instrument case, keyboard, *etc.*, but it is just as likely that the chromatograph will become a peripheral to the computer and be accessed and controlled through a printer cable.

2 Current Integrator Status

Any description of a 'modern integrator' or 'data processor' will quickly outdate itself, but it is still worth assessing the current state of development in order to review what has been achieved and speculate on what is yet to come.

Integrators come in four guises:

(1) integrators incorporated into a chromatograph;
(2) stand-alone integrators;
(3) microcomputers with integration software;
(4) as components of larger laboratory computer (LIM) systems.

The first three are simple variants of each other, they have the same integration role and are directly comparable. LIM systems are rarely bought for their integrator specifications, but any laboratory paying the price will expect a performance at least equivalent to a much less expensive integrator.

Standard Integrator Specification

This includes:

(1) measurement of peak areas, heights and retention times with method files to store data processing parameters;
(2) standard analysis calculations of area% (normalization), internal and external standards methods, and calibration or automatic determination of response factors from standard mixtures;
(3) presentation of the chromatogram and analysis report to an alphanumeric printer plotter;
(4) storage of chromatograms, associated files and calibration data for reprocessing and regulatory audit.

All integrators offer these facilities. Computers additionally offer facilities such as:

(1) programming languages – for non-standard operations; export of data to other programs for secondary processing.
(2) Microsoft Windows™ or UNIX™ platforms;
(3) archiving of data; sophisticated 3-D and multiple display of chromatograms; report preparation involving the cutting and pasting of chromatograms, merging of files, updating of databases, *etc.*;
(4) the role of system controller with the ability to validate a system automatically, bring it to operational readiness and perform first line malfunction assessment and correction (ON ERROR, DO . . .).

An integrated system can monitor system readiness and column performance, and assess system suitability and results validation. The results of such tests are increasingly stored for regulation compliance and independent audit. Some QC chromatograph systems cannot proceed to analyse samples unless the controlling data processor has logged successful validation and calibration tests.

Computers are much more powerful and versatile than integrators and they have much greater capacity for development and expansion. It is difficult for a manufacturer to justify the development of another stand-alone integrator, and it may be that the final models are already in the market.

Integrator Files

Integrators equipped only with RAM storage typically hold ten analysis files. Computers hold files whose capacity are limited only by the memory capacity of the computer.

Each file is able to store all of the parameters and information for one type of analysis. When a file is selected it immediately configures the integrator to process the chromatogram derived from that analysis. Analysis files and stored chromatograms (data files) are linked. When a chromatogram is called for reprocessing, the analysis file is called as a default. Current audit techniques require original data, *i.e.* the chromatogram and the parameters that processed it, to be protected from alteration. Reprocessing is therefore indelibly flagged and although editing can take place, changes cannot over-write original data, and they may only be preserved as daughter files.

Files hold four types of parameters for:

(1) measurement of peaks and solute identity;
(2) instrument control and communications;
(3) sample management, calculations and calibrations;
(4) report preparation and output.

There are almost too many parameters. Extensive parameter control is supplied because experienced analysts did not trust early integrators to measure peaks in a consistent way without supervision and intervention. They asked for, and were given, the means to control the data processing. Many analysts, however, simply want to press 'start' and have the integrator program itself; for some categories of analysis (clean, simple samples) this is possible, for others it is not.

Analysis Parameters for Peak Measurement

Parameters which affect the measurement of areas, heights, retention times, *etc.*, are discussed in detail later. They include the detector signal sampling frequency or 'peak width' parameter, 'slope sensitivity' and 'baseline drift tolerance' parameters.

Analysis parameters are normally optimized during method development. They can be adjusted during an analysis by deploying a Time Program – a set of timed commands to change parameter values at prescribed times after injection. It rids the analyst of the need to attend the analysis in order to make the changes himself. At the end of the analysis the parameters return to their initial values.

Before storage and reprocessing of chromatograms became available, Time Programs included commands to start the autosampler, operate relays, switch valves, change columns, *etc.* These are now stored in a separate Event or Analysis Control Program, which is only employed during real time analysis, not during data reprocessing. Stored chromatograms can be reprocessed without re-enacting the experiment.

The role of analysis parameters is to optimize measurement of required

quantities such as peak area, and enable the integrator to ignore unwanted events such as baseline disturbances. All wanted peaks are reported, but not the injection pressure pulse on the solvent peak, or the reagents added during derivatization.

Analysis parameters are also used to override the integrator's inclination to ruin peak measurement by establishing a baseline in the wrong place, for example, at the bottom of a negative baseline excursion.

The most important analysis parameters are:

(a) Peak Width

Whatever the name given to this parameter by any individual manufacturer, it is the detector signal sampling frequency, the rate at which the detector signal is digitized for integration or storage. It is the most important integrator parameter.

For manufacturing ease, the electronic sampling frequency is fixed at a high value typically in the region of 100 to 1000 samples per second. It can be made to appear variable by adding or 'bunching' consecutive high frequency samples into a single sample whose value is the same as would have been obtained at a slower sampling frequency. The bunching is software-controlled and therefore, in contrast to a variable electronic sampling frequency, adds nothing to manufacturing costs.

The degree of bunching is related to the width of peaks being measured so that the same detection and measurement algorithms can be used throughout the chromatogram.

The analyst programs the Peak Width parameter to measure early peaks. Later (and broader) peaks are measured by an updated value from the 'Time to Double' function (see below), or by a pre-set command in the Time Program.

(b) Slope Sensitivity or Detection Threshold

Peaks are detected because the detector signal amplitude changes more rapidly when peaks elute than the baseline signal does between peaks. Integrators track the detector signal and detect the change in slope that occurs when a peak emerges, but smaller changes are allowed to pass as baseline fluctuations. There is a threshold of slope below which peaks are not detected, and this value is set by the 'slope sensitivity' parameter. Baseline drift which does not exceed this threshold is considered to have zero slope.

It is usual for the integrators to perform a test (a 'slope' or 'noise' test) to measure and self-program an appropriate value for slope sensitivity.

Effect of Peak Width and Slope Sensitivity on Measurement. Peak Width and Slope Sensitivity are the two most important parameters when programming an integrator to measure a chromatogram. But their importance is more related to the suppression of noise rather than to the accurate measurement of the peaks. Table 4.1 shows the results of measurements of the computer-generated peaks in Figure 4.5 where the S/N ratio is in excess of 500:1. Variations to peak width from 0.1 to 1.0 s and

Table 4.1 *Measured areas are independent of peak width, slope sensitivity and asymmetry (over these ranges) when* S/N *is good*

(a) Varying peak width, constant (default) slope sensitivity

Peak	B/A	*Peak area measurement*		
		Peak width (seconds)		
		0.1	0.5	1.0
1	1.00	213 713	213 692	213 710
2	1.50	212 402	212 397	212 396
3	2.00	212 556	212 523	212 579
4	3.00	213 028	213 000	213 012

(b) Varying slope sensitivity, constant (default) peak width

Peak	B/A	*Peak area measurement*		
		Slope sensitivity ($\mu V\ s^{-1}$)		
		25	225	425
1	1.00	213 738	213 715	213 733
2	1.50	212 405	212 418	212 403
3	2.00	212 538	212 550	212 535
4	3.00	213 013	213 010	212 991

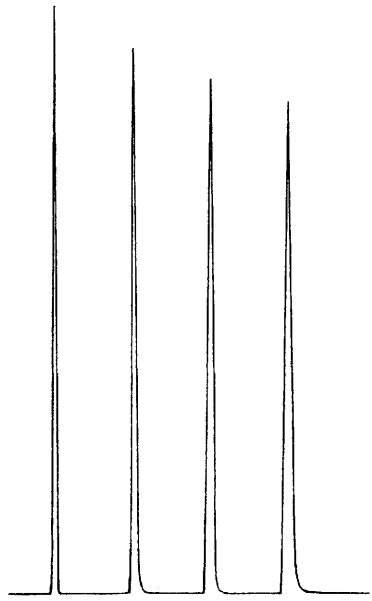


Figure 4.5 *When* S/N *is good, accurate programming of peak width and slope sensitivity is not critical, see Table* 4.1

in slope sensitivity from 25 to 425 $\mu V\,s^{-1}$ make little difference to the measured areas over a range of asymmetries from $B/A = 1$ to 3.

In real chromatograms the S/N ratio is rarely this good and the quality of the measurements deteriorates as the noise level increases. The lesson to be learned here is that improving the chromatography always takes precedence over programming the integrator. If an analyst finds it necessary to tweak the parameters continuously in order to achieve consistent measurement of standard samples, it is a clear indication that more work is needed to bring the instrument and analysis under control. Integrators do not improve bad chromatography, they just measure it.

It is worth noting that if it were not for noise, integrators could be manufactured without programmable controls and the analyst's keyboard input would be restricted to report selection.

(c) Baseline Drift Tolerance

Accurate peak sensing is only one half of the problem for an integrator: locating baseline at the end of peak is the other.

Baseline only exists at certain allowed locations, where the analyst would expect it to be. The range of these locations is determined by the 'baseline drift tolerance' parameter. The name varies from one manufacturer to another but the principle is the same. Baseline is defined as zero slope (inside the limits allowed by slope sensitivity) within a range allowed by the drift tolerance. Outside this range, zero slope is not recognized as baseline no matter how long it lasts.

The purpose of this parameter is to allow the integrator to find the new baseline after a peak, or group of peaks, has eluted, where the detector signal does not return to its original level. It prevents the integrator fixing a baseline at points of zero slope at peak tops, or in valleys.

Baseline drift tolerance is often misunderstood and confused with slope sensitivity, yet the difference should be clear: drift parameter determines how *far* the baseline can drift from its original position, slope sensitivity determines how *fast* it may drift away.

(d) Time to Double

Peaks vary in width and shape. Values of peak width and slope sensitivity that are correct for the start of an analysis may not be appropriate for peaks at the end. The 'time to double' parameter allows progressive updating of peak width and slope sensitivity at regular time intervals during the analysis and keeps each near to its optimum value for the emerging peaks.

Time to double is a relic from the days of less sophisticated integrators; it is a precursor of the time program since its function is to change parameter values during an analysis.

When integrators were first employed, isothermal GC analyses were prevalent, and in these analyses peak widths and retentions of homologues increase exponentially. At appropriate times during an analysis, the peak width parameter doubles in value and the slope sensitivity halves (*i.e.* becomes twice as sensitive). Peak width

and slope sensitivity could also be changed in a time program, but if these are the only parameters involved, it is easier to set one parameter, the 'time to double', than make up a time program. This doubling up emulates the exponential increase in peak widths and enables the integrator to track better the slower onset of broad peaks which elute late in the analysis.

In the present generation of integrators, time to double is an automatic technique. The integrator monitors peak widths during an analysis, and if the current parameter value is too small, it is increased (although not necessarily doubled) to a better value which matches eluting peaks.

Peak width and slope sensitivity parameters are interrelated. Doubling of peak width implicitly means doubling of the sensitivity: one is not changed without the other. (See Figure 5.20.)

(e) Minimum Area or Height Threshold

Long-term noise and genuine but unwanted small peaks are unavoidably measured. They can be removed from the final report by setting a minimum size criterion. This size test is only applied when peak measurement is complete, which is too late to prevent peak detection marks and retention times being added to real time chromatograms.

The same threshold applies to height or area, although a filter value appropriate to areas will be about 100 times too large when applied to the heights of the same peaks.

(f) Analysis Duration

The analysis time is programmable so that the instrument will initiate an analysis report and reset the instrument for the next injection while unattended.

Parameters which Override the Integrator's Logic. These parameters are mostly used in a Time Program to prevent measurement of unwanted peaks and incorrect placing of baselines.

(g) Integrate Inhibit

Integrate inhibit disables peak recognition for a specified period. It is used to avoid measurement of unwanted solvent peaks, or false triggering of integration by baseline disturbances (Figure 4.6).

The first signal the integrator sees after a period of integrate inhibit is defined as baseline, whatever is happening to the detector signal.

(h) Forcing Tangent or Perpendicular Separation

Integrators may separate two fused peaks by a perpendicular when the analyst would prefer tangent separation, or vice versa. This problem arises critically when the second peak is at that marginal size where small changes in concentration cause it to be separated from its neighbour by a perpendicular in one analysis but by a

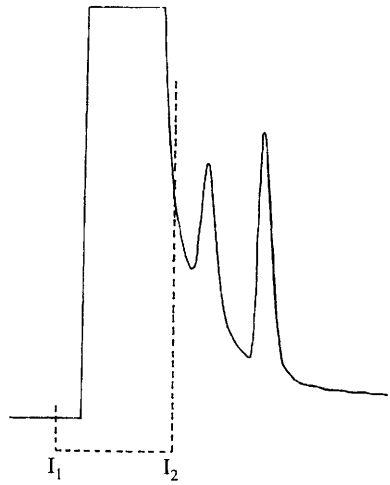


Figure 4.6 *Integrate inhibit prevents measurement of solvent peak whose tail is baseline for rider peaks. The signal at I_2 is fixed as baseline*

tangent in the next. It creates a measurement discontinuity and ruins any chance of comparing analyses. By binding the integrator to one type of separation the analyst can at least achieve consistency (Figure 4.7).

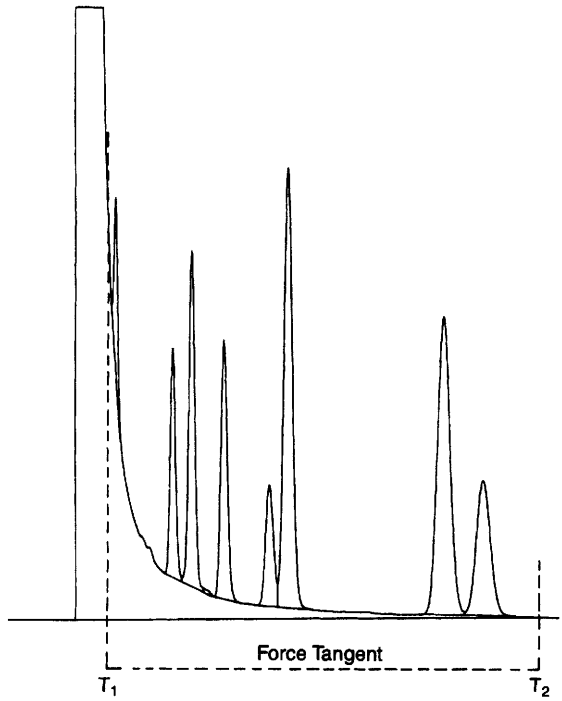


Figure 4.7 *All peaks between* T_1 *and* T_2 *are forced to tangent skim off the solvent peak*

(i) Forcing Baseline Detection

Integrators can be forced to recognize specific baseline points, implicitly to terminate a previous peak detection, or they can be forced to complete peak measurement at a moment determined by the analyst.

The integrator is typically directed to establish a baseline after some disturbance has ended, and just before a peak of interest is due to elute (Figure 4.8).

Forcing baseline recognition is used to rescue a peak from a poor environment which makes it difficult for the integrator to measure the peak correctly. The correct analytical approach to solve the problem would be to 'improve the chromatography', but even if this can be done it does not recover a peak from a completed analysis. The analyst selects the best baseline points to isolate the peak from nearby disturbances, makes up a Time Program and reprocesses the data. On many occasions, forcing baseline and integrate inhibit have the same purpose.

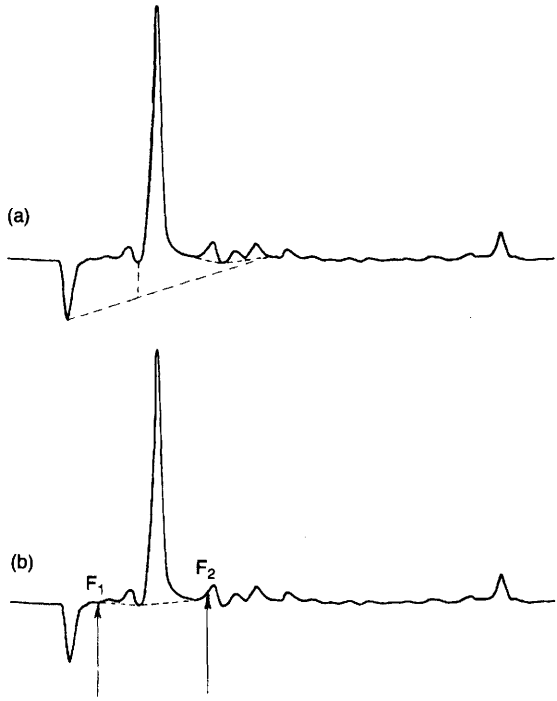


Figure 4.8 (a) *Without forcing baseline;* (b) *baseline is forced at* F_1 *and* F_2

(j) Horizontal Baseline Projection

Baseline disturbances, especially negative excursions, which spoil location of the true boundaries of a nearby peak may be ignored by projecting a horizontal baseline forwards or backwards from a neighbouring stretch of good baseline (Figure 4.9).

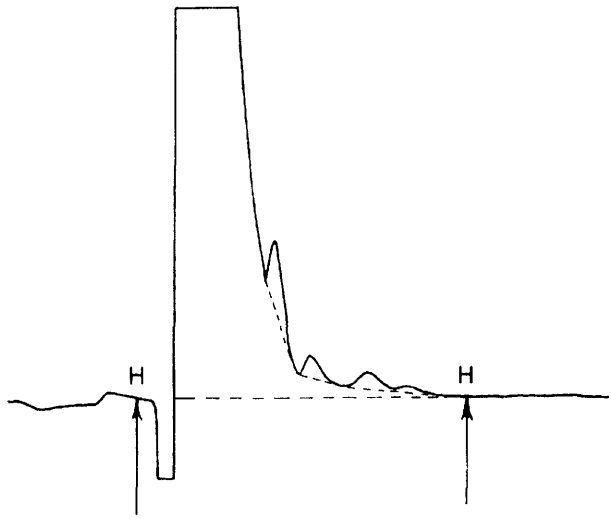


Figure 4.9 *Projecting horizontal baseline does not mean that the solvent peak is measured accurately*

This is not an accurate technique. Figure 4.10 shows a computer-generated chromatogram of two equally-sized peaks, one of which is distorted by a baseline dip, processed in three different ways:

Figure 4.10a shows how an integrator would draw the baselines and measure the peaks if left to its own methods. The baseline under the first peak is drawn from the bottom of the preceding valley and the ratio of the two peak areas, supposed to be 50:50, is 62:38. The first peak is over-estimated.

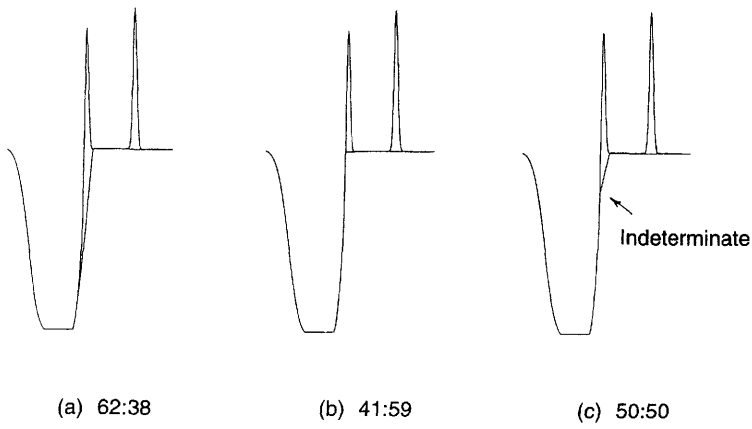


Figure 4.10 *Horizontal baseline projection does not correct this problem*

Figure 4.10b projects a horizontal baseline back under the first peak. The measured areas are now in the ratio 41:59, under-estimating the first peak.

Figure 4.10c has a baseline under the first peak which was drawn by trial and

error to give the right answer, 50:50. This baseline is otherwise indeterminate; no integrator can be given rules to draw it.

Horizontal baseline projection, like forced baseline, is intended to correct deficiencies in the chromatography, but it does not necessarily lead to accurate peak area or height measurement. The analyst believes that the integrator would produce an even worse measure if left to itself but Figure 4.10 shows that this is not necessarily true.

The technique was added to integrators before the days of high resolution columns in the hope that it would solve an otherwise unsolvable problem. It did not, but that was not known before it was available. Analysts should not use horizontal baseline projection, they should separate the peaks of interest from the baseline disturbances before measurement.

The same arguments apply to related projection techniques, which project across valleys and link stretches of real baseline.

(k) Inverting Negative Peaks

It is possible for a thermal conductivity detector (for example) to produce positive and negative peaks in the same chromatogram. Integrators will not normally measure negative peaks because positive slope is required for peak recognition. What is worse, the integrator will track a negative peak to its lowest level, establish baseline there and trigger peak detection as the signal returns to baseline.

Negative peak inversion allows measurement of negative peaks provided they do not breach the lower operation limit of the integrator, about -5 to -10 mV, below which they will be clipped. Use of this function on unresolved peaks can wreck the measurement of neighbours (see Figure 4.11).

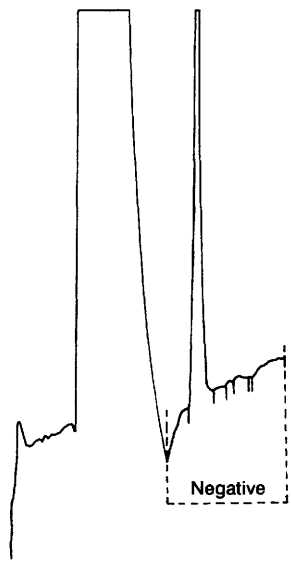


Figure 4.11 *Negative peak inversion – but the large peak cannot be measured now*

Instrument Control and Communications

Analyses involving mechanical operations such as autosampling, valve switching or column switching require a control or event program to operate the events at the correct times. Relays are a commonly-used interface between system modules; they are an uncomplicated technology and reliable.

A more sophisticated form of communication between chromatograph and controller allows status monitoring and feedback mechanisms to control the chromatograph rather than merely to synchronize operations.

Sample Management, Calculations and Calibrations

A third set of parameters, called a 'sample table' or 'solute identity file', is concerned with matching measured peak areas or heights to solute names and assigning response factors or standard concentrations to them.

The parameter table holds data on:

(1) analysis and solute names;
(2) retention times and windows (location tolerances) for component identity;
(3) response factors and standard concentrations;
(4) the type of calculation, area%, internal standard, external standard;
(5) internal standard weights and total sample weights;
(6) statistics and chemometric tests.

For a series of unattended analyses with auto-injection of solutes, this becomes a large file and it is essential that the samples are racked up in the correct analysis order. Bar code readers have been used to prevent mismatching. They read the sample name and match it to its logged data.

Peak and Solute Identity

Integrators identify peaks by their retention time. This identity will have been established by other means, but once the solute is known it is routinely identified in subsequent analyses as the peak which elutes at the expected retention time.

The integrator is a computer, and if instructed to find a peak at 5 min, it will look for a peak at 5 min but ignore the one at 4.999 or 5.001 min. Peaks are therefore identified at a specified time give or take a tolerance.

Tolerance can be defined in two ways: as a percentage or as an absolute time. A peak may be predicted to elute at 5 min (300 s) $\pm$ 10%, which means that any (or all) peaks eluting between 270 and 330 s will be identified as the expected peak. Alternatively, a peak expected at 5 min $\pm$ 0.5 min may elute at any time between 4.5 and 5.5 min and be identified.

To prevent more than one peak eluting inside the same window, the tolerance is made as narrow as possible consistent with the variation in retention time from one analysis to the next.

Relative Retention Times

If absolute retention times are not stable, relative retention times are used. One peak is assigned the role of standard, and every other peak retention is compared to it.

$$\text{Relative Retention} = \frac{t_R}{t_{std}} \tag{1}$$

Any uniform retention drift caused by solvent leaks or temperature drift, *etc.*, will affect the retention times of both peaks but not the retention ratio. Relative retention times are more stable than absolute retention times. They are less vulnerable to changing column conditions, septum leaks, and similar systematic errors.

The chosen standard must be well-separated from its neighbours so that it can be given a wide window from which it is unable to drift. If it does move out of its window it will not be identified correctly, and the whole method breaks down.

Standard Calculations

Integrators offer three standard calculations; normalization (area%), internal standard and external standard, and a variety of non-standard calculations, some of which use BASIC programs or involve exporting data to secondary software. These are well documented elsewhere[4–8] and will only be summarized here.

Area% or Normalization

The concentration, c_i, of a solute i, is calculated from,

$$c_i = \frac{A_i}{\sum A_i} \times 100\% \tag{2}$$

where A_i is the peak area of solute i, and $\sum A_i$ is the total area of all the chromatogram peaks.

If response factors of each solute are included, Equation 2 becomes,

$$c_i = \frac{A_i R_i}{\sum (A_i R_i)} \times 100\% \tag{3}$$

where R_i = absolute solute response factor, the quantity of matter to produce unit area count.

Area% as defined by Equation 3 means percentage of solutes which are eluted and detected. This might only be a fraction of what was injected. When using Area%, the analyst must be certain that nothing has been retained by the column or missed by the detector.

If solutes are known to be retained or not detected, and the measured fraction is known accurately, then instead of normalizing to 100%, the results can be scaled to the fractional percentage by means of a scaling factor, S:

$$c_i = \frac{A_i R_i}{\sum(A_i R_i)} S \times 100\% \tag{4}$$

For example, beer is typically analysed by GC using an FID which does not respond to water. If the beer is known to contain 94% water then it is more useful to normalize the analysis to 6% rather than 100%.

Internal Standard

The weight of solute in a sample is determined by comparing its area to the area of a known amount of standard added to the sample and analysed with it,

$$W_i = \frac{A_i R_{is}}{A_s} W_s \tag{5}$$

where W_i = solute quantity
W_s = quantity of internal standard
R_{is} = the relative response factor of solute to internal standard
= ratio of absolute response factors

The concentration of solute in the original sample (weight = W_{spl}) is given by:

$$\text{wt}_i\% = \frac{W_i}{W_{spl}} 100\% = \frac{A_i R_i W_s}{A_s W_{spl}} \times 100\% \tag{6}$$

If some solutes are retained by the column, or not detected, the sum of the detected solute weights will be less than the original sample weight, or their concentrations will add to less than 100%.

The internal standard method allows absolute quantities of solutes to be measured by compensating for variations in injected sample volume. It does not compensate for differences in peak shape between solute and standard if this indicates detector overloading. A good standard gives a peak that is the same size and shape as the peak it is to be compared with.

It is a weakness of the internal standard technique that one peak serves as the standard for several others where these vary in shape, size and response factors.

External Standard

The external standard method compares a solute peak area in an unknown sample with the area produced by a known amount of solute in a standard sample analysed

under identical conditions. Then:

$$W_i = \frac{A_i}{A_s} W_s \tag{7}$$

No response factor is necessary because sample and standard are the same species. If the peaks are similar in size, it will avoid peak shape/linearity problems. Equal-sized injections of standard and unknown are essential. Manual injections are not normally accurate enough; sample valves or autosamplers are to be preferred.

Standard solutions are prepared containing known amounts of all solutes required in the unknown solution. They are then analysed and calibration data are calculated over the full measuring range of each solute.

Sample Scheduling

The widespread use of autosamplers to minimize staff costs, extend system operating time and maximize injection volume accuracy requires sample details to be logged, and the samples racked into the correct position in sample holders. A sample schedule contains details about every sample, including:

Rack Position
Sample Identity
Analysis details – the analysis file to be used
Calibration details
'On Error' instructions – what to do with unexpected events or results

Sample schedules are an expanding area. Regulatory controllers are pressing for more details about failed (and often unattended) analyses. This leads to manufacturers building more diagnostics and audits into their chromatographs/data processors. Simultaneously, it provides the analyst with facilities to program alternative operations and actions as operational errors arise. A mishap during an experiment should not stop sample analysis except as a last resort, but details of what happened should be reported.

Calibration

Calibration is the determination of response factors from the analysis of prepared standards and is featured by all integrators. The maximum number of standard samples included in each calibration can vary, as can the number of repeat injections which are allowed.

The simplest calibrations involve analysis of one or two standard solutions (Figure 4.12).

When three or more standards are analysed there is the choice between multi-linear calibration, where the graph points are joined up by straight lines, or linear regression (least mean squares) which draws the best straight line through them (Figure 4.13). Data can be exported to sophisticated 'Spreadsheets' and Maths/Statistics software for further processing.

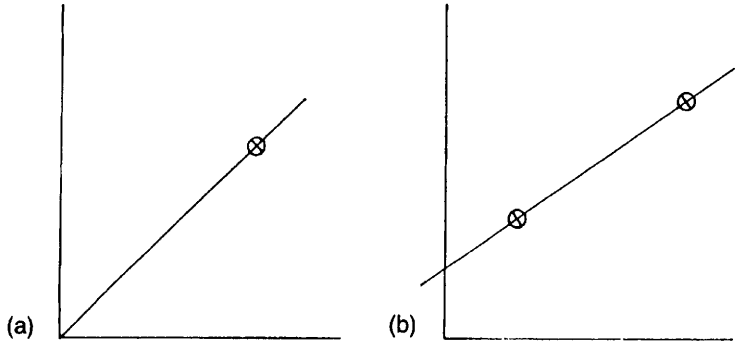


Figure 4.12 *Calibration.* (a) *Single standard,* y = mx; (b) *dual standard,* y = mx + c

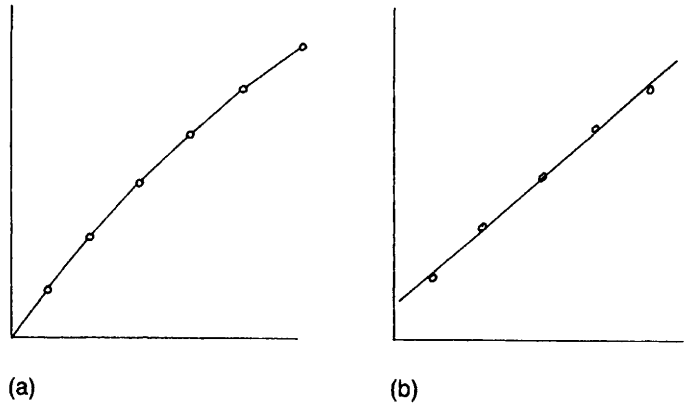


Figure 4.13 (a) *Multilinear calibration joins the 'dots'*; (b) *linear regression gives best straight line*

The more standards, the better the calibration, but the more work is involved in preparing them. Single point calibration can give rise to severe calibration errors and should not be used unless the analyst has previously performed multi-standard calibrations to demonstrate that the calibration is linear and passes through zero.

Report Preparation and Output

The last set of parameters selects from the information generated by the analysis, what is to be printed in the analysis report. The available information includes:

(1) Analysis Notes;
- date and time
- operator's name
- sample or batch name
- analysis conditions
- instrument settings
- sample preparation notes

(2) The chromatogram including peak detection marks and retention times;
(3) Analysis Results;
    peak areas, heights and retention times
    peak measurement diagnostics and error flags
    solute names and concentrations or quantities
(4) Calibration Data;
    the results of calibration experiments
    response factors
(5) Analysis Validation;
    system suitability tests
    column efficiency and peak resolution data
    peak asymmetry measurements
    error messages
    'out of spec' warnings.

Analysis reports need not include all the information listed above, but growing legislation requires this kind of information to be available to external auditors as well as analysts.

Trends

The variation in analyses results of a given product over a period of time (weeks, months) can indicate trends and drifts in the manufacturing procedure and are used for Statistical Process Control. Trends are used to monitor the effects of 'improvements' in production in order to demonstrate that the intended benefits are realized. Trends can show longer-term drift in instrument performance, especially when similar systems are compared.

Report Distribution

When an integrator or computer is linked to a corporate computer it opens up the possibility of distributing analysis information through the network to several destinations, to:

(1) The Analyst. The chromatogram and results will be presented in the usual way.
(2) Laboratory Supervisor. Confirms that the work is being carried out and that analyses are within specification, or gives early warning of which samples are failing to match a required specification.
(3) Production Manager. Wants to know when product of low quality is being made and when corrective action is required.
(4) Accountancy/Personnel. Performance measurement of laboratory staff and instruments allows both to be used optimally.
(5) Permanent Storage. The full set of results is archived to a computer or file server for trend analysis and audit reference in case of future technical problems when the product history will be reviewed.

3 Validation of Integrators

As part of a global move towards Quality and improving systems and manufacturing technique, regulators of Good Laboratory Practice (GLP) and Good Manufacturing Practice (GMP) in various countries[9] decree that analytical systems such as a chromatograph and data processor should be validated to prove functionality while in use, *i.e.* that the system works properly and is used properly. This has been considered necessary because of some disasters, especially in the pharmaceutical industry, from which it was concluded that experimental results cannot always be accepted at face value;[10] they must be proved to be credible. Rules and guidelines have been published.[9,11–14] Validation is a requirement of manufacturing processes with many consequences in Customer Protection law. The tests described below allow the analyst to validate an integrator and demonstrate that it is functioning properly. The procedures can be generalized to the rest of the system.

In a research environment, or where analyses are not directly related to product quality, validation may not be mandatory but the tests described below **are** good practice and should be carried out.

The Integrator as a Diagnostic Tool

The integrator or data processor is unique among chromatographic instruments in that it is the measuring device, it is the window into the chromatograph, and it provides diagnostic information as well as sample measurement. As such, it must be validated and calibrated separately from the chromatograph (modular validation) in order to avoid measurement bias. Then, and only then, it can be connected to the chromatograph and the system validated as a whole (holistic validation). If doubt lingers over any aspect of integrator performance, chromatographic measurement cannot proceed.

Validation means proving that an integrator is *fit for the purpose*[15] of measuring a detector signal, *i.e.* that it is capable of measuring peaks, that it has been applied correctly to the task of measuring peaks and that it was fully operational during the measurements. Validation testing need not be elaborate but it should be rigorous and provide *documented evidence* that is *sufficient to support a high degree of confidence that the* [integration] *system will consistently do what it is supposed to do.*[16] This translates in the lab to proving that the integrator is working properly, that is:

(a) it works to manufacturer's specification;
and/or (b) it works as the manufacturer intended;
(c) it can measure detector signals with quantified accuracy.

Validation *intends* to prove that the integrator will *accurately* measure real chromatographic peaks, but, because of peak overlap, asymmetry and baseline noise and the imperfect algorithms used to process them, proving an integrator is fully functional is no guarantee of measurement accuracy. Careful calibration of a specific analysis is additionally required to guarantee accuracy.

Surprisingly, peak measurement accuracy is never part of an integrator's specification, leading some to check the A/D converter for accuracy[13,17] as if this was an adequate substitute – it is not. The main errors in integration are caused by baseline noise, peak overlap and allocating baselines; errors caused by electronic components tend to be either so serious they result in instrument failure, or are small and negligible compared to chromatographic and algorithm errors. If electronic components are to be tested, it should be done by the manufacturer or by a qualified service engineer.

To meet regulatory requirements, validation requires a documented, systematic plan[11–13,18] to select, install and test the integrator and prove continuing functionality. All experimental evidence which proves functionality is retained for audit.

Specification, Installation, Operational, Performance Qualification (SQ/IQ/OQ/PQ)

The regulatory logic of validation and its relationship with total quality management can be summarized as follows:

(1) specify and purchase the data processor which has adequate specification for the designated analyses;
(2) check that the delivered data processor is the one that was specified and ordered;
(3) confirm that it has full functionality on delivery;
(4) install it correctly;
(5) calibrate it regularly and carefully;
(6) monitor that results are as expected;
(7) have regular maintenance checks and make all necessary repairs to keep it functional;
(8) keep a record of all these procedures for future audit.

There is a sequential and inevitable logic to this scheme which, if implemented correctly, will guarantee that the integrator accurately and reliably measures the detector signal presented to it.

If errors and inaccuracies do occur in spite of the scheme, the causes of the errors must be explored and any working procedures which give rise to them must be corrected so that the same errors cannot happen again. It is analogous to debugging a computer program. Systematic errors will eventually be removed, but truly random or unaccountable errors can always occur.

Integrator Specification and Selection (SQ)[19,20]

When purchasing a new, or different, integrator, the users should draw up a user specification of requirements based on the analyses to be measured. Instrument demonstrations can then be arranged with a few possible vendors. Relevant samples

are analysed to confirm that the integrator can reproduce known chromatograms and measure them as required by the user specification. The experimental details and results of this demonstration are kept in case the integrator is purchased. The general uniformity of detector output (0–1 V) and the way integrators measure peaks means that integrators tend to be selected by preference for stand-alone, LIM- or PC-based systems and for report formatting.

For re-purchases, demonstrations can be waived, but the analyst must confirm during installation that the new integrator meets the user specification and functions exactly like others of its kind when processing the same chromatogram.

Installation Checks (IQ)[19,20]

Acceptance tests applied during installation (Installation Qualification or IQ) include modular, or stand-alone, testing and holistic testing where the integrator is connected to the chromatograph and the system is tested as a whole. These tests confirm that the correct instrument was delivered and is functioning to manufacturer's specification. The analyst retains the flexibility to select appropriate tests (to begin with this means guessing what the regulators will accept). Tests include:

(1) Inspection of the instrument and packaging to confirm that:
   (a) the specified instrument was delivered as ordered;
   (b) all ordered options are there (check internally-mounted circuit boards);
   (c) nothing was broken in transit between supplier and laboratory bench;
      – report it **immediately** if damage is evident or suspected as insurance claims might have a time limit.

(2) If everything looks satisfactory, then after removing transit screws and restraints but before connecting to the chromatograph:
   (a) power up and apply manufacturer's recommended checks (see operating manual);
   (b) measure the known output from a signal generator;
   (c) power down and connect to the chromatograph;

(3) Validation:
   (a) measure a chromatogram to prove functionality:
      – this is commonly the analysis of a standard sample,
         or, the sample used during the supplier demonstration,
         or, the output from a signal generator,
      – whatever, the results must be known in advance.
   (b) integration parameters can be left at default;
         or, as they were during the demonstration,
         or, at the settings used by similar integrators measuring that sample.
   (c) the measured results must be as predicted.

These tests also prove correct connection of the integrator to the detector output.

Operational Qualification (OQ)[20–22]

There is an overlap of installation and operational qualification; installation is the first check of operation, and installation checks may have to be repeated if the analyst suspects a problem. Otherwise, operational qualification is a combination of validation (above) and calibration to determine response factors. It is always the responsibility of the user to carry out OQ tests and it is the responsibility of the analysts' employer to provide training and resources.

Validation and Calibration: The Difference

The separate meanings of validation and calibration are blurred by the common practice of analysing standards to determine response factors which, if accurate, simultaneously calibrate and confirm the functionality of the integrator. But there is a difference:

Validation proves that the integrator is functioning correctly and working as the manufacturer intended. No specific analysis details or method development are assumed; the successful measurement of **any** chromatogram using default integration parameters will suffice for validation purposes.

Calibration is the determination of response factors for a specific, known analysis. It is analysis-dependent. Method development is implicitly complete; the best integration parameters have been selected and peak retention times are known. When the analysis is changed, recalibration is necessary but re-validation is not – unless the analysis is totally new and uses parts of the software (or firmware) which have not been previously used or checked (see Validation of Software below). The purpose of calibration is therefore to tailor a general purpose integrator to the specific requirements of a quantitative analysis.

Records[23]

It is an essential part of regulation and quality control that records are kept. These should begin with the records of purchase and installation:

- the original specification of laboratory requirements and supplier's quote;
- delivery packing list;
- ROM version or software version numbers (check manuals, VDU banner screens or ask the vendor);
- a list of all installed software files.

In addition, the analyst should:

- note all installation tests and results;
- date and sign all records;
- file them in a safe place.

A common 'malfunction' of computer-based integrators following installation is the accidental deletion of an important file. Having the original list to check saves

arguments with the vendor. In their own defence, suppliers of integration software have learned to include an audit trail of user actions from which they can find the date and time of such deletions or any other unfortunate operation. These audit trails are not always available to the analyst but they can be used to justify a service bill.

Performance Qualification (PQ): Regular Testing of the Integrator

Routine maintenance ensures that the integrator continues to perform as it did when new, by checking and replacing key components if necessary. It must include independent checks to ensure that it is still capable of measuring a standard signal or chromatogram in compliance with the original specification.

Maintenance frequency can be recommended by the supplier: usually once or twice a year with the rest of the chromatograph, but it depends on usage and the cleanliness of the environment. Routine preventative maintenance does lead to a fall in the number of failures and malfunctions of a system.[24] The calibration frequency should be known to the analyst by experience; each calibration doubles in this case as re-validation.

Standard calibration procedures are used to:

- perform system suitability[25] and other tests as required;
- select best integration parameters for the analysis;
- analyse standard samples;
- determine retention times and response factors;
- show that all results meet required operational standards and specifications.

Routine maintenance tests:

- demonstrate correct functionality of the integrator;
- follow manufacturer's recommendations as stated in the operating manual;
- involve regular service visits by an approved service organization.

Maintenance records must be kept:

- to note all tests and results and their date;
- records must be signed by those responsible for producing them;
- they must be filed in a safe place and available for audit.

Following routine service by an engineer, the analyst should check that previously correct integration parameters and response factors are still programmed. It is worth noting that if an integrator is not recalibrated correctly following a maintenance visit, one regulatory body, the FDA, will blame the analyst and not the engineer who serviced it.[13] An external service engineer is required to be expert in the construction and operation of the instrument, not the details of an analysis. At best the engineer will validate an integrator with his own signal generator, not the

analyst's. Even if the engineer obligingly attempts to restore the operational parameters he may make a mistake; the FDA believe a thorough analyst will check such details when the service is finished, before the unit is returned to use.

Logging Results

All tests and their results must be systematically logged, signed and dated and kept for future audit.[11,12,26] If results are archived on computer, they should be archived in a standard format[27,28] with memory back-ups. At the time of writing it is advisable to hold key information in paper form as well as in computer files.

Validation of Software

PC-based integrators running integration software must validate hardware and software.[29] Hardware can be tested like any other electronic instrument. Validating software means dynamic testing to confirm functionality, detection of programming errors or 'bugs' and their systematic removal. This requires the co-operation of the vendor and access to source codes which the vendor will not supply unconditionally; a confidentiality agreement will be requested in return. In any case, it has long been known that it is impossible to remove all bugs from a program of any reasonable length and complexity.[30–33] Analysts, it seems here, are asked to do the impossible.

What happens in practice, and is acceptable when done systematically and logically, is that those **sections** of integration software used for an analysis are empirically validated by repeated, error-free use. Known bugs are sometimes tolerated if they can be by-passed; it may be that known bugs are unimportant, or that removing them might introduce secondary bugs. Bugs are always assumed to be present, but they may be allowed to exist in harmless inactivity in any section of software that is never used.

These used sections are the routines of data acquisition and processing used regularly for a specific analysis. They include 'Macros' and personal programs written by the analyst and data archiving/retrieval routines. Any bugs in these sections should manifest themselves during installation or calibration (better still, during demonstration prior to purchase) and the system is never allowed to measure real samples until the bugs are removed or by-passed. Dangers occur if the concentration ranges of standard samples are less than the ranges encountered in unknowns. Calibration must span the whole measurement range; interpolation is allowed, extrapolation is not.

When validated software is used in a different way (*e.g.* for a new analysis) and different sections of software come into use, the analyst must be alert to the possibility of discovering bugs in the new sections. Validation must begin again. Updated versions of operating systems or manufacturer's software do not have to be accepted and installed if this introduces bugs and invalidates sections of software that previously worked well.

Changing or repairing computer hardware can invalidate a data processing system if the software no longer works afterwards. PCs, which run office software without incident, occasionally encounter problems in the laboratory in running software that controls experiments or has sophisticated graphics. Timing problems associated with instrument control mechanisms have been traced back[34] to small changes of components on replacement circuit boards which 'should have made no difference at all', according to the manufacturer.

It is prudent to make purchase orders for upgrades conditional to the new units being compatible with the existing software and chromatographic system. When such changes are made, the system needs to be re-validated according to existing procedures and the details logged.

4 Standard Chromatograms

The appropriate way to validate and calibrate an integrator, like any other measuring device, is to provide it with an independent, standard chromatogram of known dimensions to measure; ideally this signal is traceable back to International Standards. No agreed Standard Chromatogram exists so standard samples and signal generators are used.

Analysing standard samples to validate and calibrate integrators has critical disadvantages:

(1) Chromatograph and integrator form an instrumental closed loop which conceals system errors. Precision, *i.e.* experimental stability can be measured, but accuracy remains suspect unless the 'right answer' is known independently. A carefully weighed sample prepared from pure components provides the usual independent, right answer but this is conditional on calibration being made over the whole measuring range and that the results from 'unknowns' do not lie outside this range.

(2) Repeated analyses of a standard sample never produce identical chromatograms. Imprecisions generated by sample, column and chromatograph (collectively, much larger than those of the integrator; see Tables 2.1 and 2.2) are included in the measured precision of the integrator – and carried over as calibration uncertainty into other analyses. A chromatogram from a standard sample is not really stable enough to be a candidate for a 'Standard Chromatogram'.

The perfect standard chromatogram has known dimensions and guaranteed reproducibility. Validation and calibration of the integrator then become independent of the chromatograph and accuracy can be measured by comparing measured results with prediction.

Creation of Standard Chromatograms

For many years, service engineers have used devices based on a small (AA) battery and a potentiometer to generate a signal for testing repairs. The output signals were uncalibrated and not intended to be reproducible. These devices have developed into compact signal generators which meet the requirements of regulated validation.

Historically, signals of guaranteed reproducibility have been generated in several ways:

(1) by the re-play of real chromatograms stored in fixed format;[35,36]
(2) from synthetic chromatograms[37–40] created by computer software and output through a D/A converter;
(3) from the output of electronic signal generators.[35,41]

In 1968, Johnson[35] designed a synthetic chromatogram from a real chromatogram from which peaks were cut out and removed. The remainder was mounted on a rotating cylinder whose surface was photoresistive. Light shone on this drum generated an electrical signal proportional to the area exposed by the cut-away peaks. The device acted as a flow sensitive detector, by varying the speed of rotation the apparent width of the peaks varied correspondingly. Unfortunately, the S/N ratio depended on external lighting and measurements had to be made in a dark room.

Real chromatograms stored on analogue devices such as electromechanical tape recorders suffer from fade and wear, tapes stretch and have limited frequency spectrum, playback speeds can drift and recorder components are temperature-dependent.

Fixed signal generators based on Gaussian peaks output to a digital/analogue converter (DAC) have been used to validate integrators independently of the chromatograph but their signals cannot be used for calibration. They are inexpensive, reproducible and widely used.

Real chromatograms can be stored inside an integrator for future re-play (*i.e.* validation). Some manufacturers store a 'standard' chromatogram inside their chromatograph which can be output to an integrator to validate it, but calibration is not possible.

Computer-based Standard Chromatograms

Hunt[38] and Papas and Delaney[37] used large computers to create synthetic chromatograms from peak models, baselines and simulated noise which they combined in a limited variety of ways. Dyson[40] extended this to PCs, asymmetric peaks and greater variety of outputs, which could model real chromatograms.

Computer-generated chromatograms based on Gaussian or Exponentially Modified Gaussian peaks which mimic real chromatograms are two orders of magnitude more reproducible than analytical chromatograms.[37] With development, this precision can improve and the PC can be replaced by an inexpensive commercial 'black box'.

Specification of Standard Chromatograms

Any specification for standard synthetic chromatograms is either simple or application-dependent. A survey of the literature and trade catalogues shows four categories:

(1) equal sized and baseline-resolved Gaussian peaks – to test repeatability;
(2) resolved Gaussian peaks covering a range of sizes – to test linearity;
(3) reproduction of stored chromatograms – to test accuracy;
(4) creation of variable, synthetic chromatograms – to mimic known chromatograms.

Creation of Synthetic Chromatograms

Chromatograms are created by adding peaks, baselines and noise, created separately from models. Synthetic chromatograms are constructed by varying peak size, shape and retention time. Baseline disturbances are fashioned from inverted peaks or baseline combinations. Meyer[42–44] has studied the systematic variation of peak overlap and asymmetry using such chromatograms and estimated the errors – see Chapter 2.

Baselines

Baselines are constructed from linear, sigmoidal or other functions, positive or negative in slope. Their height and range are variable and they can be superimposed to create baseline spikes or dips.

Baseline Noise

Noise can be reproduced from a stored baseline. 'White' noise can be created in software from the product of two random number generators – one to simulate amplitude, the other to simulate frequency. 'White' noise does not reflect different detector characteristics (*e.g.* the flicker of an FID).

Two kinds of synthetic white noise are required:

Truly random: for testing how robust integrator parameter settings are to noise.
Pseudo random: randomly-generated noise which is thereafter stored and output as a fixed file for the testing of integrator smoothing algorithms. Differences in the processed output must be due to the algorithms, not to variations in the signal.

Synthetic noise issuing from a D/A converter has already been smoothed by the D/A before it reaches the integrator so that the signal-to-noise ratio measured by the integrator is invariably better than intended. This makes it difficult to assess the improvement to the S/N ratio made by the integrator alone as the specification of the original noise has to be accepted on trust.

Drift

Drift can be simulated by replicating real, stored baseline drift, and this can include events such as valve switching spikes. It can also be simulated by one half of a wide synthetic peak.

Traceability of Synthetic Chromatograms

Calibration of integrators should be traceable to a Mass Standard because mass is what is eluting from the column. It should be possible (although the author has found no reports) to link the output of a D/A converter issuing a standard peak to a coulometer and so relate 'peak area' to mass deposited at a cathode. Thus calibrated in mass(coulomb)/area units, the synthetic peak can be used to calibrate an integrator in a traceable fashion.

Traceability turns chromatography from a comparative to an absolute technique. Response factors determined by coulometry provide thermodynamic and electrochemical information on analyte species.

An easier, traceable calibration and one in current use, is to relate peak height to standard voltage. This is subject to the same limitations which apply to peak height when it is used as a substitute for peak area.

5 References

1. D.T. Sawyer and J.K. Barr, *Anal. Chem.*, 1962, **34**, 1213.
2. G.L. Feldman, M. Maude and A. Windeler, *Int. Lab.*, vol. 1, March/April, 1971.
3. A.T. Leung, J.R. Hubbard and L.A. Miller, *J. Chromatogr. Sci.*, 1976, **14**, 166.
4. J. Novak, *Advances in Chromatography*, Marcel Dekker, New York, 1973, vol. 11.
5. *Modern Practice of Gas Chromatography*, ed. R.L. Grob, Wiley-Interscience, New York, 2nd edn., 1985, ch. 4.
6. L.R. Snyder and J.J. Kirkland, *Introduction to Modern Liquid Chromatography*, Wiley-Interscience, New York, 2nd edn., 1979.
7. *Quantitative Analysis using Chromatographic Techniques*, ed. E. Katz, Separation Science Series, John Wiley and Sons, New York, 1987, ch. 3.
8. G. Guiochon and C.L. Guillemin, Quantitative Gas Chromatography, *J. Chromatogr. Libr*, **42**, Elsevier, Amsterdam, 1988, ch. 14.
9. L. Huber, *Good Laboratory Practice – A Primer*, Hewlett Packard, Publication number: 12-5091-6259E, 1993.
10. R.F. Tetzlaff, R.E. Shepherd and A.J. LeBlanc, *Pharm. Technol.*, March 1993, p. 100.
11. FDA Regulations, 4-1-92 Edition, Sections 211.160 to 211.188.
12. GLP The UK Compliance Program, Department of Health, London, 1989.
13. W.B. Furman, T.P. Layloff and R.F. Tetzlaff, *J. AOAC Int.*, 1994, **77**, 1314.
14. B.A.P. Jorgan and R.D. McDowall, *LC-GC Int.*, 1997, **10**, 574.
15. *Quality Control Handbook*, ed. J.M. Juran, McGraw Hill, New York, 1988.
16. *Guide to Inspection of Computerized Systems in Drug Processing – Training Aids for Investigators*, US Dept. of Health and Human Resources, FDA, February 1983.
17. M. McConell, M. Canales and G. Lawler, *LC-GC Int.*, 1992, **5**(3), 34.
18. R.F. Tetzlaff, *Pharm. Technol.*, April 1992, p. 60.
19. L. Huber, *LC-GC Int.*, 1996, **9**, 564.

20. L. Huber, *LC-GC Int.*, 1996, **9**, 666.
21. L. Huber, *LC-GC Int.*, 1996, **9**, 794.
22. J. Elling, *et al.*, *Anal. Chem.*, 1997, **69**, 409A.
23. J.W. Dolan, *LC-GC Int.*, 1997, **10**, 216.
24. Private communication: Dyson Instruments service records .
25. US Pharmacopoeia, Vol. XXI, p. 1229.
26. R.F. Tetzlaff, *Pharm. Technol.*, January 1993, p. 80.
27. D.C. Nelson, V. Dauciunas, R. Lysakowski and M. Duff, *Int. Lab.*, 1992, vol. 22, Nov/Dec., 32.
28. R. Lysakowski, *J. Chromatogr. Sci.*, 1994, **32**, 236.
29. R.F. Tetzlaff, *Pharm. Technol.*, May 1992, p. 70.
30. R.S. Pressman, *Software Engineering – A Practitioner's Approach*, McGraw Hill International Editions, New York, 2nd ed., 1987, ISBN 0-07-100232-4.
31. T. McCabe, *A Software Complexity Measure*, IEEE Transactions, 1976, **2**, 308.
32. B.W. Boehm, Rand Inst. Report RM-6213-PR. Also Published as: Selected Rand Abstr., 1970, **8**, 52 (Jan/Dec); US Nat. Tech. Info. Service AD 703279.
33. A.N. Papas, *Chromatographic Data Systems: A Critical Review*, CRC Reviews, **20**(6), CRC Press, Boca Raton, FL, 1989.
34. N. Dyson, Information from Technical Department in Author's Company.
35. R.D. Johnson, *J. Gas Chromatogr.*, 1968, **6**, 43.
36. J.T. Shank and H.E. Persinger, *J. Gas Chromatogr*, 1967, **5**, 631.
37. A.N. Papas and M.F. Delaney, *Anal. Chem.*, 1987, **59**, 54A.
38. R.J. Hunt, *J. High Res. Chromatogr.*, 1985, **8**, 347.
39. J.P. Foley and J.G. Dorsey, *J. Chromatogr. Sci.*, 1984, **22**, 40.
40. N. Dyson, *Int. Lab.*, 1992, **22**(6), 38.
41. Product Code 900520/900521, Phase Separations, Clwyd, UK.
42. V.R. Meyer, *J. Chromatogr. Sci.*, 1995, **33**, 26.
43. V.R. Meyer, *LC-GC Int.*, 1994, **7**, 94.
44. V.R. Meyer, *LC-GC Int.*, 1994, **7**, 590.

CHAPTER 5

Digital Measurement of Peak Areas

1 Signal Sampling

The measurement of chromatographic peaks by computers or integrators is fundamentally different from the manual techniques described in Chapter 3. It is more closely related to the measurement of peak moments in which no assumption of peak shape is made.

Manual measurement of strip chart recorder peaks is made on the analogue representation of an analogue detector signal by an analyst who makes the logical decisions. Computers and integrators cannot begin to work with the detector signal until it is converted into a digital format. In this conversion, original data are lost, but if the sampling frequency is high enough, the lost information is not critical to the analysis and does not invalidate the results.

Analogue to Digital Conversion[1–3]

The two principal methods of A/D conversion used by chromatography data processors are described below.

(a) Voltage to Frequency Conversion

Voltage to frequency converters (VFCs) were specifically developed to measure small analogue signals such as the millivolt outputs of GC and LC detectors. Although relatively slow, VFCs have been widely used in single channel integrators. Their principal advantages are:

(1) they were designed to measure small voltages (and currents);
(2) a wide linear dynamic range;
(3) low drift;
(4) high resolution.

The general schematic diagram for a V/F converter is shown in Figure 5.1.

The detector signal V_{in}, charges condenser C, at a rate proportional to V_{in} until

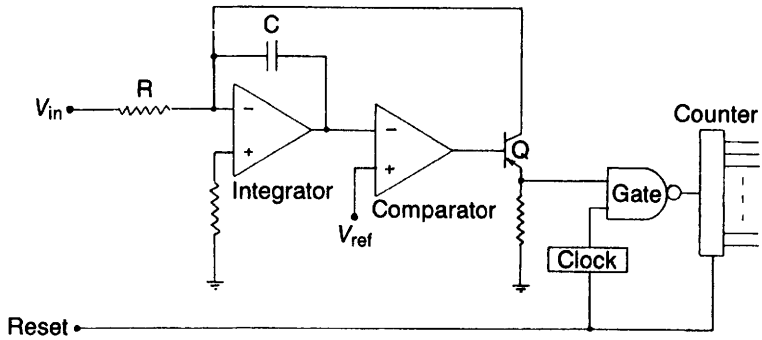


Figure 5.1 *Voltage to frequency converter*

the output of the integrator operational amplifier exceeds V_{ref} when the comparator inverts and switches on the transistor Q. This discharges the condenser creating a voltage spike at the emitter of the transistor. The discharge allows the comparator to invert back, switch off Q, and the condenser begins to charge again.

Each charge/discharge cycle generates a pulse at the emitter, and the chain of pulses is gated into a counter. The pulse count within a time window is proportional to V_{in}; by controlling the window aperture time, the integrator sensitivity can be varied.

For a typical VFC[4] working with better than 0.005% linearity, the output frequency F_{out}, is related to V_{in} by:

$$F_{out} = 10^5 V_{in} \tag{1}$$

On baseline the pulse counts are low because the detector signal is low and generally constant if the baseline is flat. When peaks elute, V_{in} increases and the pulse count increases correspondingly. When the peak has eluted, the signal returns to the baseline and the pulse count falls back to the 'baseline rate'.

The pulse count during the window aperture is the integral of the signal voltage over that time, and averages out any signal fluctuations occurring within the window.

(b) Dual Slope, Integrating A/D Conversion

Dual slope A/D converters (Figure 5.2) can sample very rapidly and have been used in conjunction with larger computers to process the signals from several chromatographs simultaneously. As column technology improves and peaks become narrower; they are gradually replacing voltage to frequency converters in single channel integrators.

Dual slope conversion is a two stage process. In the first stage an integrating operational amplifier is reset and then the detector signal V_{in}, is applied to its input for a fixed time, t_{in}. This charges condenser C at a rate proportional to V_{in}.

In the second stage, the detector signal is replaced by a reference voltage V_{ref},

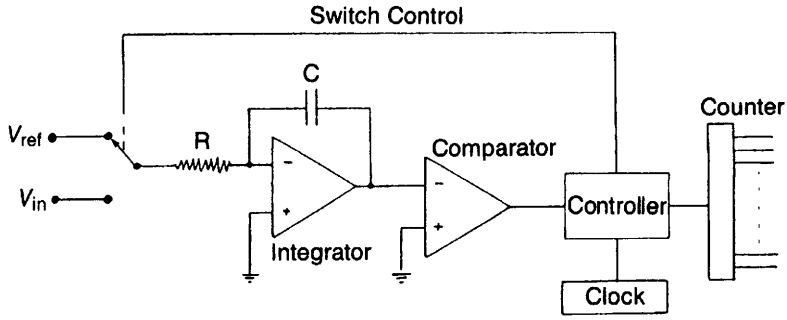


Figure 5.2 *Dual slope A/D converter*

which is opposite in polarity to the detector signal. This 'charges down' the condenser at a fixed rate, proportional to V_{ref}, and the time required to discharge it is clocked by a stream of pulses into a counter until the discharge of C is complete.

At the end of the count, the integrator and clock are reset, the input is switched back to V_{in}, and the time/pulse count proportional to V_{in} is stored (Figure 5.3).

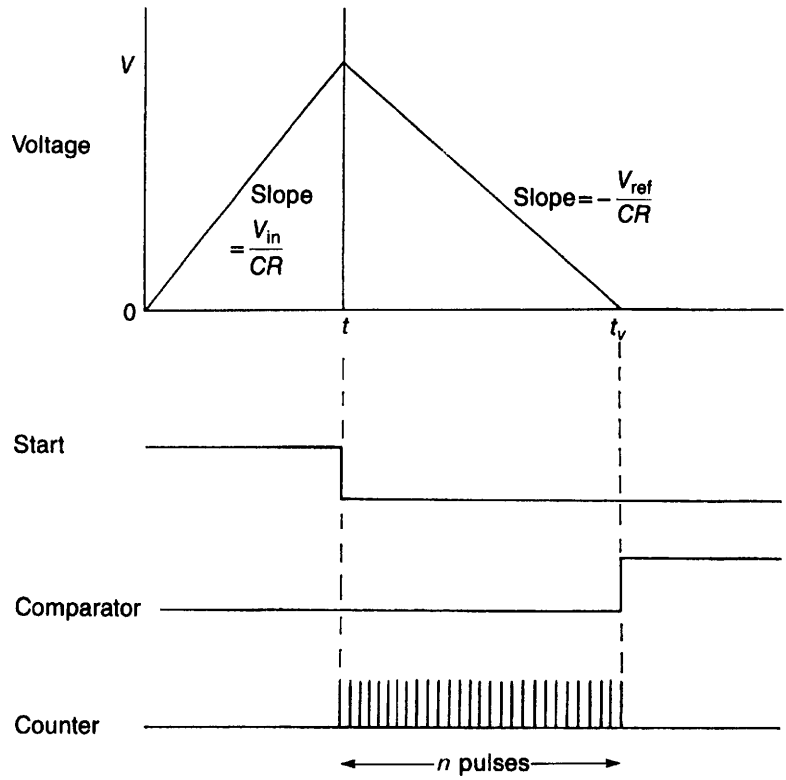


Figure 5.3 *Timing diagram*

The charge acquired by condenser C from V_{in} is discharged by V_{ref}:

$$\therefore \frac{1}{CR}\int_0^{t_{in}} V_{in}\,dt = \frac{-1}{CR}\int_{t_{in}}^{t_V} V_{ref}\,dt \tag{2}$$

If V_{in} and V_{ref} are regarded as the average values of the input and reference voltages during their period of application, the Equation 2 simplifies to:

$$\frac{V_{in}}{CR}\int_0^{t_{in}} dt = \frac{-V_{ref}}{CR}\int_{t_{in}}^{t_V} dt \tag{3}$$

The discharge time is represented by the number of pulses, n, from the clock to the counter, *i.e.* $t = nF$, where F = clock frequency;

from Equation 3, $$V_{in} = \frac{(t_V - t_{in})}{t_{in}} V_{ref}$$

or

$$V_{in} = \frac{n_t}{n_{t_{in}}} V_{ref} \tag{4}$$

Equation 4 is independent of values for R and C. It requires only that V_{ref}, t_{in} and the clock frequency F, are constant.

Resolution of A/D Converters

Resolution is the smallest change in the analogue signal that can be seen in the digital output, *i.e.* which changes the least significant bit.

A/D converter resolution is specified by the number of 'bits', *e.g.* 8 bit or 12 bit; a 12 bit A/D converter can resolve to 1 part in 2^{12}, or 1 part in 4096. If the output voltage is 1 V, the smallest change in output that the A/D can see is:

$$\frac{1\text{ V}}{4096} = 0.24\text{ mV}$$

Bearing in mind that strip chart recorders were often used in the 1960s and 70s at a span of 1 mV, and showed baseline blips of 1% FSD (10 μV), then 0.24 mV, or 24% full scale deflection of a 1 mV recorder is not very impressive.

Auto-ranging of A/D Converters

Commercial integrators and computers use auto-ranging A/D converters which divide a large detector signal into ranges, usually two or three, in order to provide the necessary resolution near the baseline where it is needed. For example, with a

12 bit converter, the range 0–10 mV is resolved to 2.4 μV, the 10–100 mV range is resolved to 24 μV and the 100–1000 mV range to 0.24 mV. The necessary calculations to cover range switching are taken care of by other circuitry or software.

Data Sampling Frequency

Many studies[5–10] have been made to determine the best sampling frequency for measuring peak areas accurately. It is invariably shown that:

(1) measurement accuracy increases with sampling frequency;
(2) errors in asymmetric peaks are greater than those of symmetric peaks sampled at the same frequency.

Theoretically there should be no 'optimum' data sampling frequency. Mathematics predicts and experiment confirms that the faster the data sampling, the more accurate is peak measurement (Figure 5.4). If the sampling rate is too slow, peak detail will be lost and measurement accuracy reduced.

There is a practical limitation, however: if the sampling frequency is too high, unwanted baseline noise peaks will be detected and measured. So the sampling frequency is set slow enough to act as a filter for baseline noise. Nyquist sampling theory[11,12] shows that the maximum sampling frequency should not sample the average noise peak more than twice if the noise is to be filtered. The maximum usable data sampling frequency is therefore determined by the baseline noise.

Sampling Frequency and Integrator Manufacture

It is simpler and cheaper to manufacture an electronic circuit of fixed sampling frequency than one whose frequency is variable. This frequency must be high enough to sample the fastest peaks, but such a rate will generate more data than necessary to process broad peaks, it will require more computer memory to store the extra data, more time to complete data processing, and it will allow detection of more unwanted noise peaks.

Sampling Frequency and Data Processing Algorithms

Peaks vary in width, generally increasing with retention. If the integrator sampling frequency is fixed, algorithms created to process the stored data must contend with increasing numbers of data samples/peaks.

Simpler processing algorithms can be used if the number of data samples/peak, or sampling density, is uniform,[13] but this is at odds with the requirement of a simple-to-manufacture, fixed sampling frequency. The problem is solved by Data Bunching.

Data Bunching

Data bunching is a software technique (with zero manufacturing cost) which collects consecutive data samples taken at high frequency into groups or bunches.

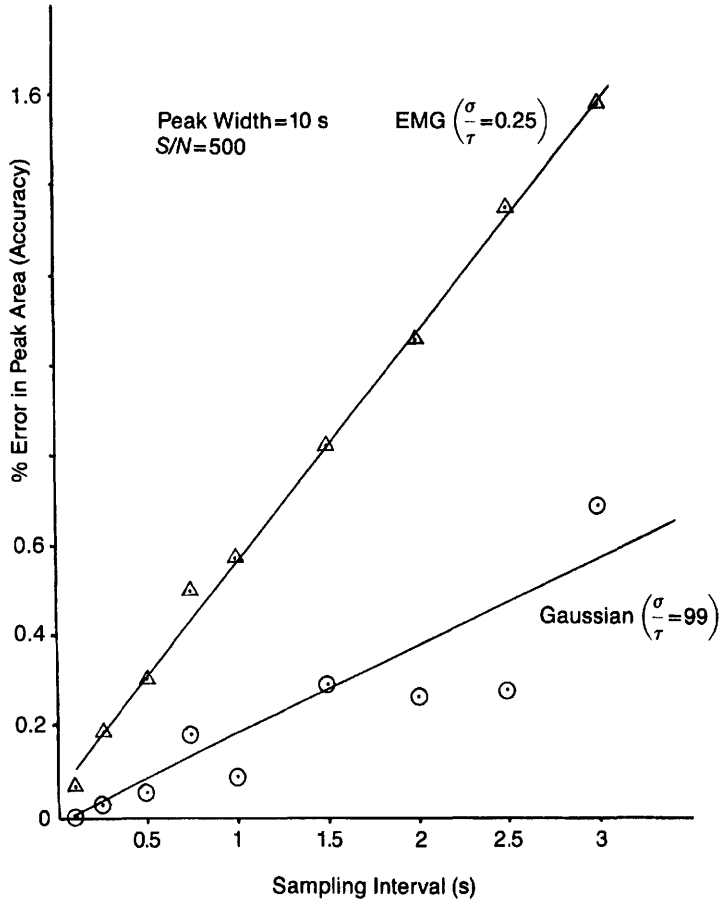


Figure 5.4 *Accuracy of area measurement improves with sampling frequency* (Data from *J. Chromatogr. Sci.*, 1988, **26**, 101)

The number of data samples in the bunch is determined by local peak width[14,15] and this creates a uniform sampling density throughout the chromatogram, or as near uniform as possible.

The effect of data bunching is similar to changing the sampling time to match the bunch width, but unlike a change in sampling frequency, bunching operations are created by software and can be reversed, or varied.

Where data storage capacity is a problem, non-linear data bunching[16] has been used to condense data files by 80% without significant information loss.

Baseline Bunching

Over long stretches of stable baseline, stored data samples will merely consist of a series of numbers having the same value. To avoid waste of memory, baseline data

are stored in one bunch represented by a single datum and the number of times it recurs.

Estimating the A/D Sampling Frequency and Bunch Size

Figure 5.4 shows that the minimum number of data samples required to measure a peak area accurately (< 0.1%) increases with peak asymmetry. For a symmetrical peak it is about 25[10], though this is conditional on the S/N ratio being not less than about 25:1. See Figure 5.5.

With greater asymmetry more samples are required. It has been shown[8] that 100 samples/peak gives near maximum peak area accuracy; faster sampling gives negligible improvement.

If a Gaussian peak width is 6σ (from -3σ to $+3\sigma$), then 100 samples/peak is approximately 17 samples/σ. The standard deviation σ, and retention time t_R, are related to column plate number, N, by,

$$N = \frac{{t_R}^2}{\sigma^2} \qquad \text{(see eq. 2, Ch. 1)}$$

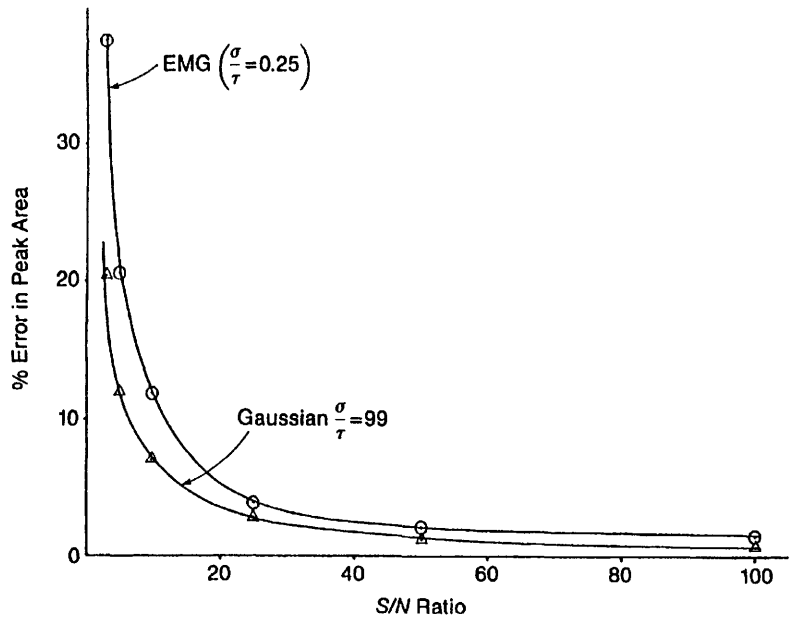


Figure 5.5 *Accuracy of area measurement increases as* S/N *ratio improves – faster for symmetrical peaks. Peaks are* 10 s *base width; sampling interval* 1 s (Reproduced[10] with kind permission from *J. Chromatogr. Sci.*, 1988, **26**, 101)

or

$$\sigma = \frac{t_R}{\sqrt{N}} \tag{5}$$

The base width of a Gaussian peak can therefore be expressed in terms of N and t_R by:

$$\text{base width} = 6\sigma = \frac{6t_R}{\sqrt{N}} \tag{6}$$

If this width is to be digitized into 100 samples, the necessary bunched sample frequency f, is given by:

$$f = 100\frac{\sqrt{N}}{6t_R} = 17\frac{\sqrt{N}}{t_R}(\text{approximately}) \tag{7}$$

The fixed sampling frequency F, of the A/D converter is designed to sample the fastest peaks, for example, early eluters from a WCOT column. If such a peak has a retention time of 250 seconds (say) in a column of 250 000 plates, the sampling frequency necessary to sample this peak 100 times will be:

$$f = 17\frac{\sqrt{250\,000}}{250} = 34 \text{ samples/s} \tag{8}$$

Higher sample densities of 100 samples/peak compensate for asymmetry.

The fixed sampling frequency F, varies with integrator manufacturer in the range of 100 to 1000 samples/s, although this may have to be shared by more than one input channel. It is sufficient for (today's) fast capillary peaks.

Bunching of samples to achieve uniform sample density through the chromatogram is estimated from:

$$\text{Bunching, } B = \frac{F}{f} \tag{9}$$

$$= \frac{Ft_R}{17\sqrt{N}} \tag{10}$$

During data collection the integrator monitors peak width and issues instructions to vary the bunching to match the peaks being eluted; usually this means increasing the number of samples in a bunch.

In order to keep peak sensing algorithms uniform (the reason for data bunching), the integrator only changes the number of data samples in a bunch between peaks, when the detector signal is on baseline. If there are no suitable stretches of baseline, a queue of bunch update commands can form, waiting to be implemented, and a uniform sampling density may not be achieved.

Data Bunching and the Peak Width Parameter

The number of integrals that constitute a 'bunch' is programmed by the width or peak width parameter. This number is estimated by the analyst when peak width is programmed and represents the initial bunching of data samples. The bunch size is continually reviewed and updated during the analysis by the integrator. If the local peak width exceeds the current bunch size by a prescribed amount the integrator will increase the bunch size accordingly.

Peak width is the most important parameter of the integrator and must be programmed carefully otherwise peaks will be measured incorrectly or even missed altogether. Its value is not actually expressed in terms of numbers of data samples per bunch; for convenience it is related to the widths at half height of early peaks. Peak width has a relatively small range of initial values; a guide to them is given in Table 5. 1.

Except for GPC, it is unlikely that an initial peak width value greater than 1000 ms will be required. If selection errors are to be made, it is better to err on the small side, *i.e.* too narrow a bunch, but then the S/N ratio should be good or too much noise will be seen and some area may be lost from the start and end of peaks.

Peak Width Parameter and Analysis Reprocessing

Reprocessing an analysis must allow the peak width or data bunching to be changed, and the data to be reprocessed with the new value. Adjustments to peak width are normally small from, say, 500 to 400 ms, and since analysis data are stored as integrals, the smallest possible change will be the addition or removal of one datum from the bunched integral. To allow the change from 500 to 400 ms means that the data have to be stored as 100 ms samples. If data are stored as 500 ms samples then two data can be added to produce a 1000 ms sample, three can be added to produce a 1500 ms sample and so on, but 500 ms samples cannot be changed to 400 ms samples.

If changing peak width is a possible requirement, data must be stored at a fast enough sampling frequency to allow small adjustment. This will generate more data which require greater storage capacity, and so the option should be considered carefully, but as computer speed and memory expand, this becomes less of a concern.

Once the best choice of peak width has been determined, data can be economic-

Table 5.1 *A guide to initial peak (bunch) width values*

Chromatogram	*Peak width (half height in* ms)
WCOTs, Megabore	100–300
Microbore, Fast LC	200–400
Packed GC	300–500
Packed LC	400–600
Dedicated amino acid analyser	700–1000
GPC	1000–2000
CE, CZE, *etc.*	10–300

ally stored at this value on the assumption that no further revision will be necessary.

Peak Sampling Synchronization

Digitizing a peak signal can be likened to laying a grid over the peak as shown in Figure 5.6. It should make no difference to the measured peak area where this grid is laid: nudging it to right or left in Figure 5.6 should still give the same area. But, it does make a difference: the measured area varies with grid position.[17] Stated another way, measurement of the peak area is not independent of the synchronization of peak start with sampling start.

Lack of uniform synchronization reduces measurement precision, since if it were ever possible deliberately to synchronize the start of one peak with sampling, the combination of regular data sampling and random peak positions would preclude synchronization of other peaks except by chance.

The loss of precision is related to the separation of the peak start from the nearest sampling start, and is therefore reduced by increasing the sampling frequency.

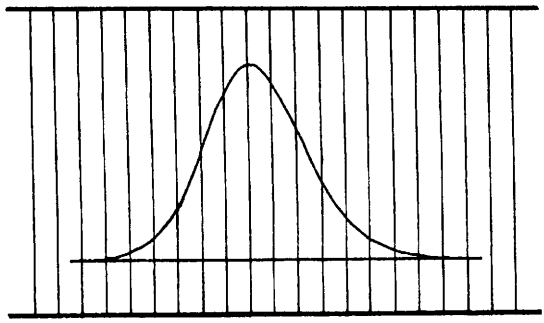


Figure 5.6 *Peak overlaid with grid*

Rounding or Truncation Errors

Sampling frequency errors have been reported[17] relating to the 'bit resolution' or smallest change in detector signal that the integrator can see, and to the limited number of decimal places used by earlier computer technology.

Errors in 'bit resolution' were reduced with the introduction of auto-ranging A/D converters, and the increased power of modern computers has solved the problem of too few decimal places. Programmers can re-introduce this error, however, by truncating calculations too severely, or at too early a stage during processing.

Aperture Time Jitter[17]

The precision of the sampling frequency is another problem that more recent electronic technology has reduced to negligible proportions.

2 Filtering and Smoothing the Chromatographic Signal

Filtering and smoothing are techniques to suppress or remove noise from the detector signal and to improve the S/N ratio.[18] Too much improves the S/N ratio but distorts peak shape[8,15,17,19] leading to a loss of measurement accuracy, height generally suffering more than area. Too little leaves residual noise to blur peaks and interfere with peak measurement, again with loss of accuracy but area suffering more than height. Improvement of the S/N ratio increasingly favours peak area measurement.

Filtering

Filtering is the real-time and irreversible suppression of noise by electronic hardware.

Smoothing

Smoothing is improvement of the S/N ratio by mathematical operations on the stored data; it is not irreversible if stored data can be reprocessed. All integrators now process data when the analysis is ended, and data occurring before and after the event can be included in the processing.

The smoothing algorithms used by commercial integrators are unknown to the analyst. He has no control over them or the assumptions they make, except perhaps for some limited selection. All that is known is that the integrator does employ them and such algorithms generally tend to preserve area measurement.

Electronic Filters

Hardware filters are based on capacitive or inductive noise suppression. Capacitive ('RC') filters are built into each signal input line. Inductive ('LC') filters remove spikes from the power supply.

Passive hardware filters are low in cost but fixed in time constant and function. Active filters do the same job but their time constants can be tuned. All hardware filters vary a little with component tolerance.

Common mode rejection (CMR) and cable screening protect integrators against the pick-up of random environmental spikes created by neighbourhood relays switching or heavy machinery powering up. Many laboratories, aware of these problems, already have spike filters built into their power lines to clean up the power supply. There is much anecdotal evidence about 'big machines in the next room switching on and off' to suggest that these filters are not always successful.

Sampling Frequency and Mains Coupling

Mains power frequency, 50 or 60 Hz, can couple (add) to the detector signal to some degree. The effect will cancel out over an integral number of cycles as the

negative half cycles cancel the positive halves. The fixed electronic sampling frequency F, is therefore selected in part to allow bunching to an integral number of mains cycles.

To avoid manufacturing different instruments for countries with 50 and 60 Hz mains, a sampling frequency of 100 Hz was an early standard because it can be bunched into samples of 10, equivalent to 10 Hz, which is the lowest common multiple of 50 and 60 Hz. Higher sampling frequencies such as 1000 Hz will also bunch into groups which cancel mains pick-up. It is a simple filter that costs nothing to manufacture.

Smoothing Techniques

Integrators use software-based smoothing techniques which work effectively if the data sampling frequency is constant for the duration of the peak(s) and fast enough to give a large number of samples per peak. The main smoothing operations are:

(1) digitizing and integrating;
(2) bunching;
(3) polynomial curve fitting (and peak location);
(4) signal subtraction.

These operations work on a single data collection. Unlike many analytical techniques, chromatography rarely generates identical data sets for repeated experiments so that ensemble averaging to remove random error has had limited value[20] when measuring big peaks. Instrument quality is improving, however, and ensemble averaging will become an increasingly feasible alternative to purge and trap, when analysis time is not critical. Ghaoui[21] obtained a three- to four-fold improvement in S/N ratio by ensemble averaging.

Digitizing and Integrating

Sampling and integrating the detector signal during A/D conversion averages the signal over the sampling interval. It removes noise (and any detailed peak information) occurring inside the interval whose frequency is greater than twice the sampling frequency.

Bunching

Bunching integrals together further attenuates residual noise by spreading its effect over the width of the bunched group (Figure 5.7). The size of the bunched group is determined by the peak width parameter, and the effect of varying it is to spread noise over greater or lesser time intervals.

Bunching of data samples is a form of boxcar averaging[2,22] in which the number of data gathered into the bunch increases during the analysis to match peak widths and maintain a uniform peak sampling density (Figure 5.8). No datum is included

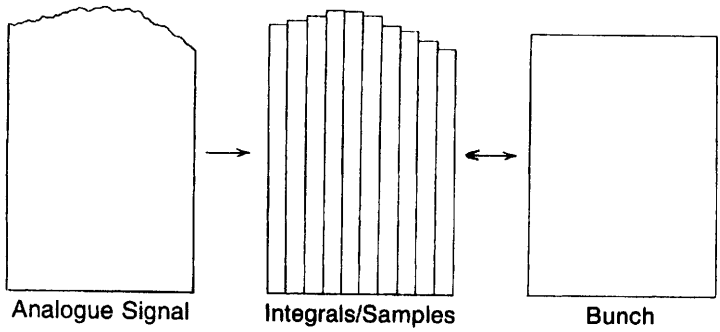


Figure 5.7 *Digitizing and bunching time averages noise, but bunching is reversible*

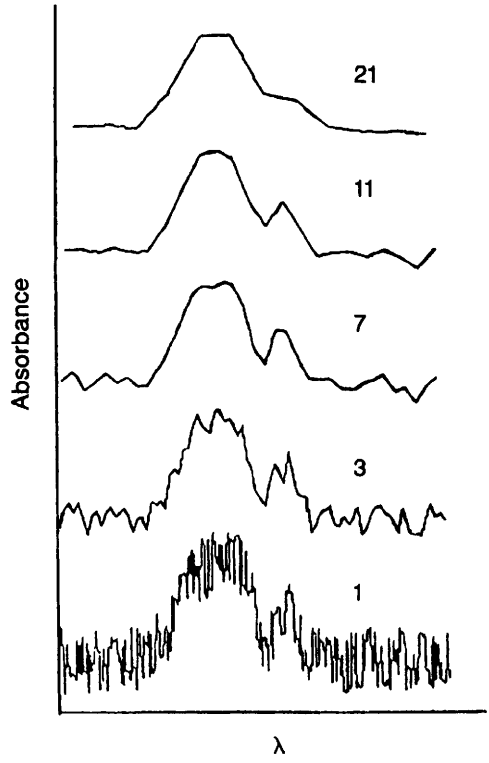


Figure 5.8 *Boxcar averaging of data. Number of samples in bunch is shown*
(Reproduced by kind permission of *J. Chem. Educ.*, 1985, **62**, 866)

in more than one bunch. The improvement to the S/N ratio increases with the square root of the number of points in the window, $2m + 1$:[22]

$$\frac{(S/N)}{(S/N)_0} = \sqrt{(2m+1)} \tag{11}$$

Moving Windows and Polynomial Curve Fitting

This third form of signal smoothing also includes simultaneous processing of data to locate peaks.

A moving window scan spanning $2m + 1$ data points, from $-m$ to $+m$ with 0 being the central point, moves along the data replacing the central datum in the window y_i, by the weighted average value of the data within the window, that is,

$$\bar{y}_i' = \sum_{-m}^{+m} C_i y_i / N \tag{12}$$

where C_i = weighted coefficient, N = normalizing coefficient, and here '=' means 'takes the value of'.

The window moves forward one point, picking up a new point at the front of the window, dropping an old one from the back, and averages a new central datum. The whole data set is processed except for the first and last m data points which are lost and therefore must not include required information.

Savitsky and Golay[23] published tables of values for C and N which generated values of $\bar{y}_i'$ identical to values which would be produced if the data had been curve fitted to an nth order polynomial by linear regression. The polynomials range in order from 2nd to 5th and the number of data points in the smoothing window ranges from 5 to 25. Additional tables allow calculation of the 1st to 5th derivatives of the curves. The 1st, 2nd and 3rd are used for peak sensing and location; higher derivatives can be similarly used if the S/N ratio is high enough.

Savitsky–Golay smoothing assumes the data sampling rate to be constant and the noise to be random,[24] ideally with a much higher frequency than the peaks. Improvement in the S/N ratio increases with the number of data in the moving window $[\sqrt{(2m + 1)}]$ but so does peak distortion measured as a loss of peak height. The tables of coefficients are symmetrical and really apply only to symmetrical peaks. If peaks are asymmetric to begin with, smoothing generates further distortion, peak broadening and loss of peak height. The worse the initial asymmetry, the greater the distortion, but it is reduced if smaller windows and higher polynomial orders are selected,[25] (see Figure 5.9).

Smoothing distortion is reduced to a negligible amount if the sampling density is high enough and the ratio of window/peak width is small, but the time required to perform the calculations increases (Figure 5.10). Peak distortion becomes apparent when the window size exceeds about 67% of the peak half width.[24]

Savitsky–Golay tables, later corrected by Steiner[26] for some numerical errors*, still form the basis of integrator smoothing and peak detection. Manufacturers and scientists have improved them, although little is published[27] as the details have commercial value.

*Not all of the errors were reported by Steiner. For example, on p. 1637 of Savitsky and Golay's paper on Repeated Convolution, d_{-3} is -21 not -33 and d_3 is $+21$.

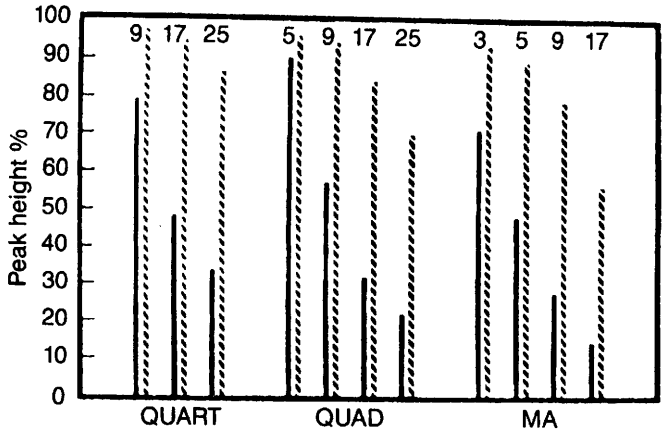


Figure 5.9 *Moving window smoothing (Savitsky–Golay) produces shape distortion if sampling frequency is not high enough. Symmetrical peaks – – –; asymmetric peaks ———*
(Reproduced with kind permission from *J. Chromatogr.*, 1976, **126**, 279)

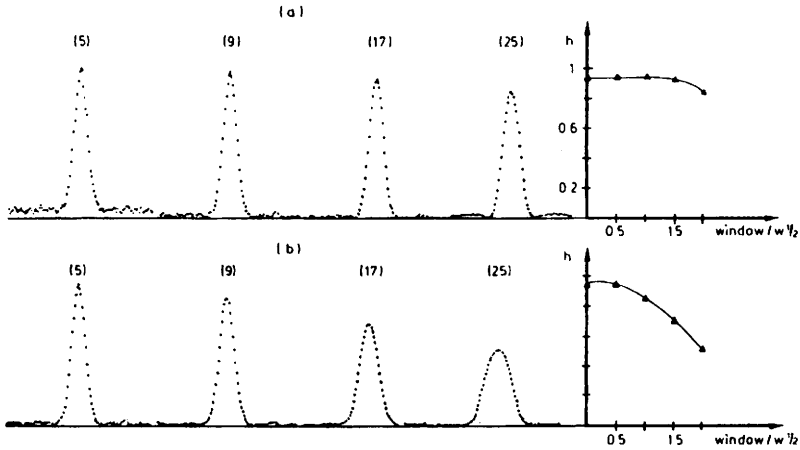


Figure 5.10 *Distortion,* h/h_0, *of a Gaussian peak as a function of the ratio between window size and half height width for* (a) *polynomial smoothing and* (b) *moving average. The window size* $(2m + 1)$ *is indicated in parentheses*
(Reproduced[24] with kind permission from *Chemometrics, a Textbook*, Elsevier, Amsterdam, 1988)

Optimum Filtering

If a narrow filter leaves too much noise, and a broad one produces peak distortion, it poses the question of what is the optimum filter width.

Van Rijswick[28] developed an equation based on a Gaussian peak, relating the S/N ratio before and after filtering. He showed that,

$$\left(\frac{S}{N}\right)_{\text{filt}} = 2.174\left(\frac{S}{N}\right)_{\text{orig}}\sqrt{\left(\frac{w_p}{\Delta t}\right)}\left[\frac{K^5}{(1+K^2)^3}\right]^{1/2} \tag{13}$$

where w_p = half width at inflection height before filtering (for Gaussian peak, $w_p = \sigma$)
N = average noise amplitude before filtering
Δt = sampling interval

$$K = \frac{\text{filter width}}{\text{peak width}} = \frac{2m+1}{w_p} \tag{14}$$

Optimum filtering is found by maximizing the 'K function', which has a maximum value at $K = \sqrt{5}$ (Figure 5.11).

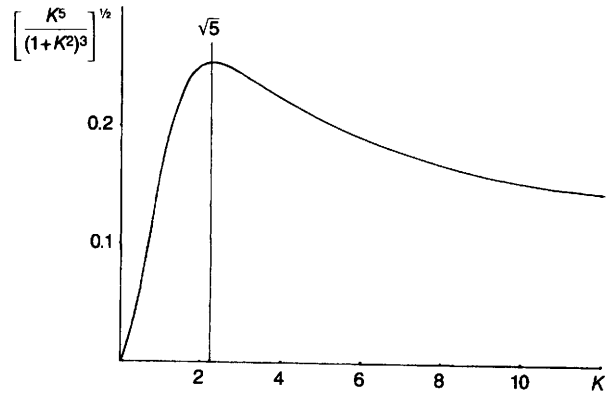


Figure 5.11 *Location of optimum filtering*

Interpreting this for a Gaussian peak, the optimum filter width is $\sqrt{5}\sigma$, and if this peak is sampled 100 times (or 17 samples/σ), the width of the filter is 38 data points, which would require considerable computation. Fortunately, Figure 5.11 has a broad enough maximum and a smaller value of K can be used without much penalty. An additional consequence of filtering is an increase in peak width to $\sqrt{[w_p^2 + (2m+1)^2]}$.

Cram *et al.*[29] showed that where peak asymmetry is present, it is necessary to use narrow windows with 4th and 5th order polynomial smoothing rather than broad windows with 2nd and 3rd order polynomials or moving average to minimize the additional distortion from the smoothing operation.

All chromatographic peak smoothing ends up as a compromise between residual noise and peak distortion.

Signal Subtraction

A recent form of signal processing made possible by the continuing developments in microcomputer technology involves the removal by subtraction of recurring events or systematic error, such as drifting baselines, valve switching spikes, *etc.*

Essentially, two chromatograms are stored, the analysis and the baseline, *i.e.* the analysis without solute injection. Provided that the analysis is repeatable, subtraction of the second from the first removes baseline disturbances common to both, and gives a chromatogram with the flat baseline.

There is a price: the difference chromatogram combines the inaccuracies and imprecision of both originals (Figure 5.12), but this penalty is no different in principle to the old practice of using dual column GCs to cancel the drift of temperature programming.

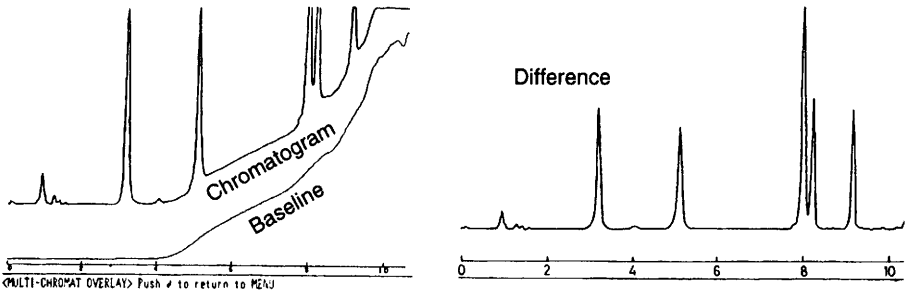


Figure 5.12 *Chromatogram and baseline are stored separately. The difference chromatogram removes baseline drift*
(Reproduced by kind permission of Shimadzu)

Subtracting a wandering baseline in this manner is one justifiable way to remove a curved baseline from beneath a group of peaks; at other times integrators construct straight baselines beneath peaks. This principle is extended to allow comparison of chromatograms: one chromatogram can be subtracted from another to highlight differences between the two (Figure 5.13).

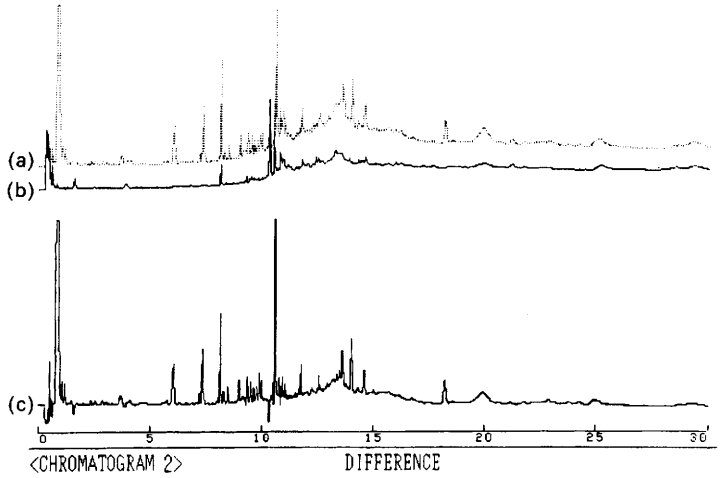


Figure 5.13 *Subtraction of one chromatogram* (a) *from another* (b) *to give difference* (c)
(Reproduced by kind permission of Shimadzu)

Chromatogram Plotting

The stored signal data are passed through a D/A converter in order to reconstruct and plot the 'original' chromatogram. If the plotter is more than a mere dot matrix printer, the reconstruction will give no hint of the digital processing that has taken place. If the integrator or computer offers screen expansion, however, it is possible to magnify the peaks and see the individual data samples represented as simple ordinates joined together for screen reconstruction (Figure 5.14 and Figure 5.15).

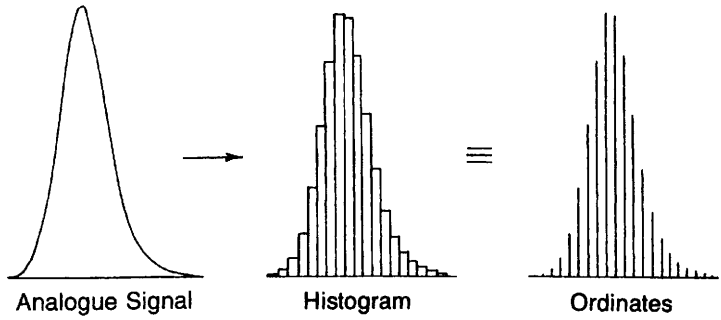


Figure 5.14 *Digitization of an analogue signal to give a series of ordinate values*

Matched Filtering[30,31]

Matched filtering is a software technique that identifies peaks in a poor S/N environment and measures them. Digitized chromatogram data are passed through or processed by this filter which contains details (models) of the peak and noise characteristics. As peak data pass through they are recognized as 'fitting the description' and are selectively measured; the filter behaves rather like a highly-selective detector.

If the peak and noise characteristics, including retention times, are accurately defined, matched filtering works well. Original data are not irreversibly changed, the S/N ratio is improved and detection limits are lowered. The improvement in S/N ratio is proportional to $1/\sqrt{(\text{peak width})}$ and so works well for capillary peaks.[32,33] It is simpler to apply than the other integration methods described in this chapter. The filter does not need a complete data set (whole chromatogram) to work on, it can be used on-line and for this reason may find use as an on-line monitor in a production site.

Matched filtering is adversely affected by peak overlap, although overlap can be part of the model. If the peak and noise characteristics are not accurately defined, if the wrong peak model is selected for example, then, depending on the degree of inaccuracy, significant errors may be introduced into the peak measurement. Worse, the required peak may not be identified at all, or worse still, the wrong peak may be identified as the required peak.

Because of this, matched filtering has been used only in commercial data processors dedicated to specific applications (for an example see reference 34) or by individuals and groups writing their own software for their own specific – and

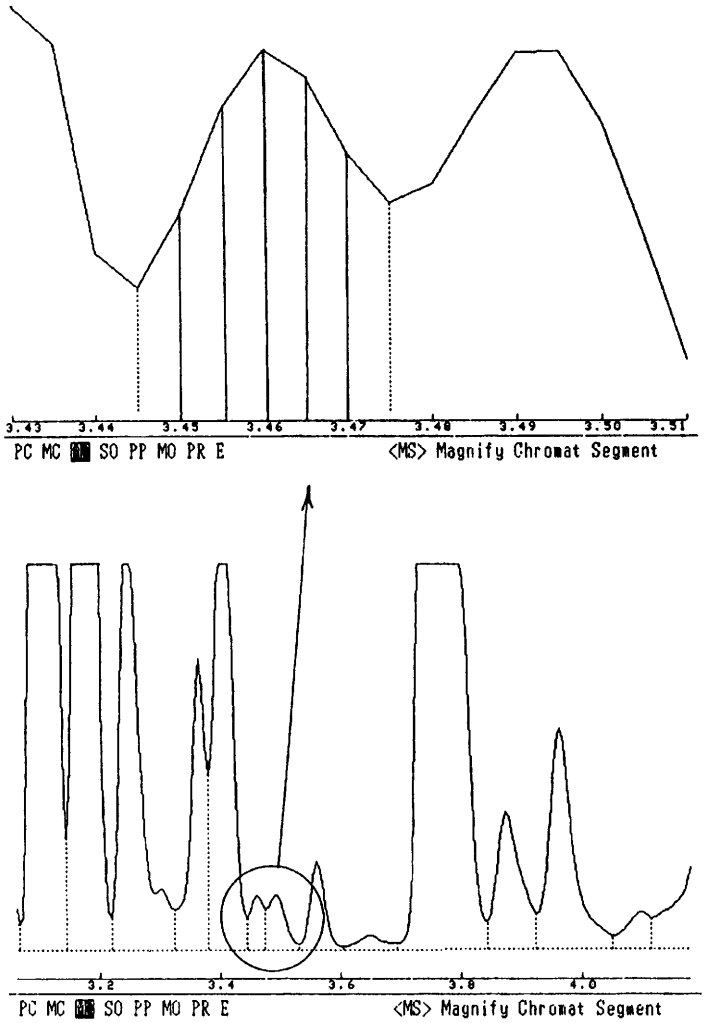


Figure 5.15 *Expansion of digitized signal shows ordinates joined to display a peak*

known – applications. The shortcomings of matched filtering, like all other deconvolution techniques, lies not in the theory but in the lack of exact repeatability of chromatogram shape from repeated injections of the 'same' sample under 'constant' conditions.

3 Location and Measurement of Peaks

Accurate peak location in commercial integrators is made by coarse and fine scans where:

(1) approximate peak location to detect every peak is obtained from Savitsky–Golay first and second derivatives;
(2) precise location of each peak start and end point (including valleys), the limits of integration, is obtained by a local search.

Finding the Peaks

Any peak can be divided into four quadrants by lines drawn through the peak maximum and points of inflection (Figure 5.16).

Each quadrant is uniquely characterized by its first and second derivatives and this is used as a coarse peak locator and shape test (Table 5.2).

The first and second derivatives are obtained from Savitsky–Golay convolution of the stored data. Peaks are located where the first derivative departs from zero by more than a prescribed threshold, the slope sensitivity S (Figure 5.17).

The non-zero threshold distinguishes between the relatively small baseline gradients and the larger peak gradients. Slope sensitivity reduces false peak detection, but inhibits precise location of the limits of integration. Nevertheless, approximate peak location shows where the peak boundaries are to be found and later prevents the integrator making detailed searches on stretches of empty baseline, and so reduces the processing time. If the S/N ratio is large and the baseline is flat, slope sensitivity can be given a small value, and the peak limits will be located with good precision even in the 'coarse' search.

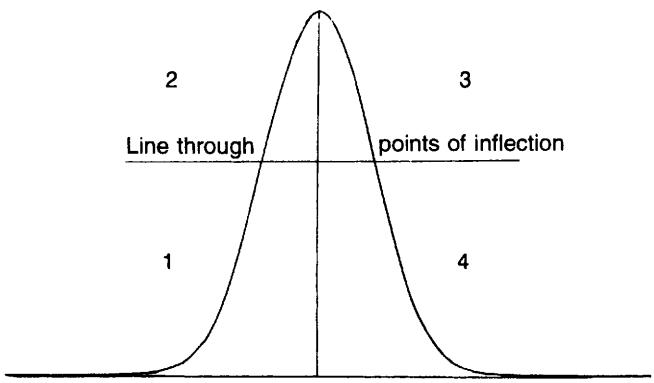


Figure 5.16 *Symmetrical Gaussian peak*

Table 5.2 *First and second derivatives of each quadrant of a peak*

Quadrant	dy/dt	d^2y/dt^2
1	>0	>0
2	>0	<0
3	<0	>0
4	<0	<0

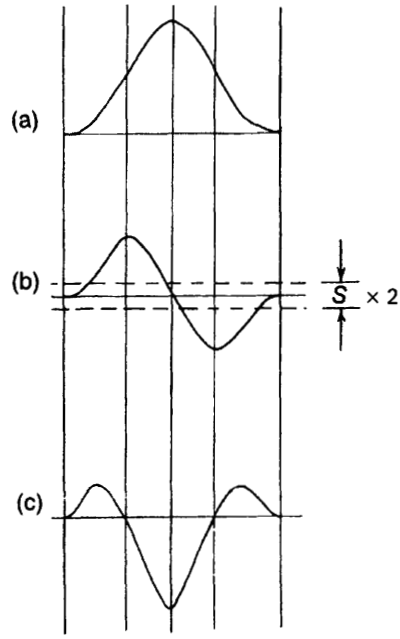


Figure 5.17 *Peak, first and second derivatives*

Retention Time

The peak maximum is located where the first derivative is zero between peak boundaries (Figure 5.17) and from this the retention time is measured. Without curve fitting, retention time accuracy and precision would be determined by the bunched sample interval.

Peak Shape Test

The combination of values given in Table 5.2 allows the integrator to examine the peak for shape and enables the integrator to distinguish between peaks and baseline ramps caused by solvent gradients or temperature programs. It is a measure against false peak detection.

Properties of the Smoothed Data

Savitsky–Golay processing to locate the first derivative is equivalent to the integrator comparing each smoothed datum to its Mth predecessor where M is the width of the convolution window ($= 2m + 1$).

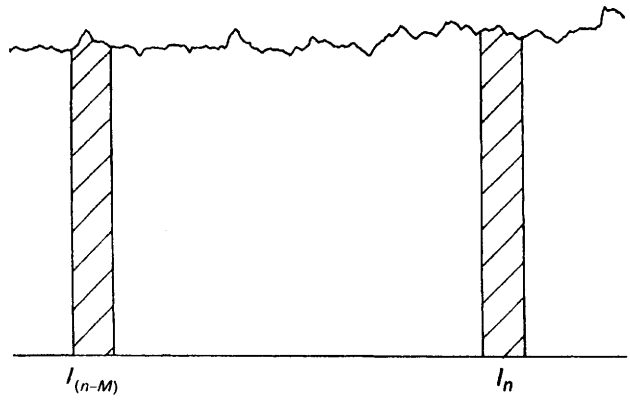


Figure 5.18 *The integrator constantly computes* $\Delta I_n = I_n - I_{(n-M)}$

Data Integrals

If I_n is the most recent smoothed datum or integral, the integrator computes,

$$\Delta I_n = I_n - I_{(n-M)}$$

and the value of ΔI_n is positive, negative or 'zero' (as defined by slope sensitivity) depending on whether the baseline is rising, falling or is flat. ΔI_n is therefore monitoring slope, and its value, which is a measure of the first derivative, can be used to show when a peak is emerging ($\Delta I_n > S$), or when the signal is on baseline ($-S < \Delta I_n < S$). Integrators which have slope indicators, keyboard LEDs for example, have them controlled by this quantity and they switch on and off according to its value.

The integrals are used in three ways as described below.

(1) As an Area Integrals are measured in units of μV sec. When those within peak limits are summed and corrected for baseline, they will compute the peak area.

(2) As a Height Since the integral's value is determined for a fixed time interval against a constant reference level, −5 or −10 mV, it can only vary if the baseline rises or falls. It is therefore entirely height-dependent.

To maintain this, integrals must not change their bunch size during peak measurement, and 'time to double' instructions are only carried out after the peak end-point has been detected, when the signal is back on baseline.

(3) As a Gradient ΔI_n is not only a difference in baseline height, it is a difference in height measured over the fixed time interval of M data slices, and so ΔI_n is also a gradient (Figure 5.19). This gradient has a discontinuity when the bunching is changed on baseline, but the integrator is programmed to account for it.

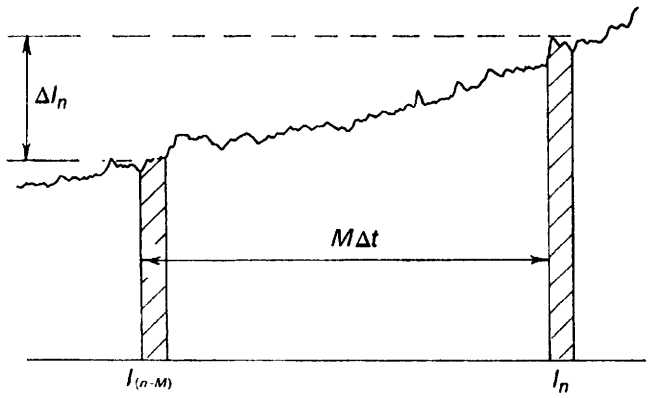


Figure 5.19 ΔI_n *is measured over a fixed time interval of* M *samples*

Baseline Fluctuations and Slope Sensitivity

Slope sensitivity is determined by giving ΔI_n a threshold value. It is the smallest shift in baseline position which the integrator sees as a positive or negative slope.

Programming Slope Sensitivity

Judging what value to give to slope sensitivity is not intuitive. Integrators therefore provide a test facility to monitor the baseline for a while, measure the average value of signal fluctuations, and self-program a value. The length of the test is linked to the peak width setting and lasts for approximately 100 bunched integrals.

Daily use of the test should make the analyst familiar with the empirical range of values to be expected; different columns, detectors and sample types all contribute to the range. With experience, the analyst may even choose to over-rule the integrator and enter an alternative value for slope sensitivity, for example, to allow more micro-peaks to be seen if these are important.

Slope Sensitivity and Representative Baseline

Automatic tests to determine baseline noise levels are performed when the signal is on 'representative' baseline.

If an integrator is asked to measure baseline noise following a thorough reconditioning of the column, it will measure a different value to that produced after samples have been analysed and the baseline is littered with chromatographic rubble. If this latter baseline is the normal one, it will be more appropriate to determine a slope sensitivity value which includes the baseline rubble and considers it to be 'zero slope'.

If programming errors are to be made, it is better to make slope sensitivity too sensitive. The integrator will see more noise and measure too many small peaks, but these can be removed from the final report by setting a minimum area threshold.

Updating Peak Width and Slope Sensitivity during Analyses

As analyses progress peaks become broader and flatter. The initial values for peak width and slope sensitivity require revision and updating. Integrators do this automatically by monitoring peak widths and comparing them with the programmed value. If there is a significant difference, the peak width parameter is increased (*i.e.* the sample bunching is increased).

Alternatively, the parameters may be programmed to increase on a timed basis via a time program.

When bunching is changed it changes the values of both the peak width parameter and the slope sensitivity (Figure 5.20).

If the bunching is doubled the effective width of each data integral is doubled, and ΔI_n is measured over twice the time interval while still recognizing the same shift in baseline level as the slope threshold. Expressed as a gradient, slope sensitivity is halved and becomes twice as sensitive.

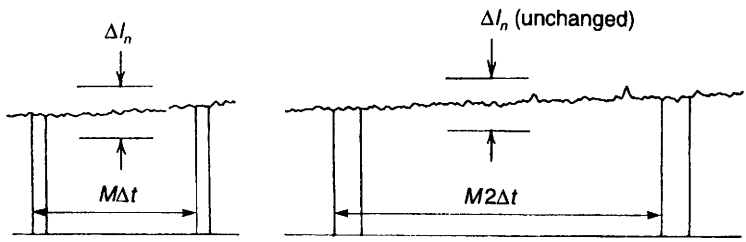


Figure 5.20 *If peak width parameter is doubled* ΔI_n *is measured over twice the time interval and slope sensitivity is halved*

Location of the Limits of Integration

Figure 5.21 shows a single peak on a flat baseline, sliced into bunched integrals as determined by the peak width parameter, and a reference level above which the detector signal is measured. The reference level is the lower limit of the integrator's operating range (-5 or -10 mV). If the detector signal drifts lower, or is inadvertently sent lower by careless use of detector back-off control, it will create a dead band error and only that part of the detector signal which projects above the reference level will be measured.

The approximate peak start is located where the detector signal first exceeded the programmed slope sensitivity value. The integrator proceeds to conduct a fine search for a more accurate location among the preceding data samples. The range of its search is limited by the time interval between first peak location S, and the peak maximum. This is no more than 2σ, where σ is the peak standard deviation; the integrator looks back over the range 4σ from the peak retention time. The 'correct' peak start is located at the lowest baseline level in this region.

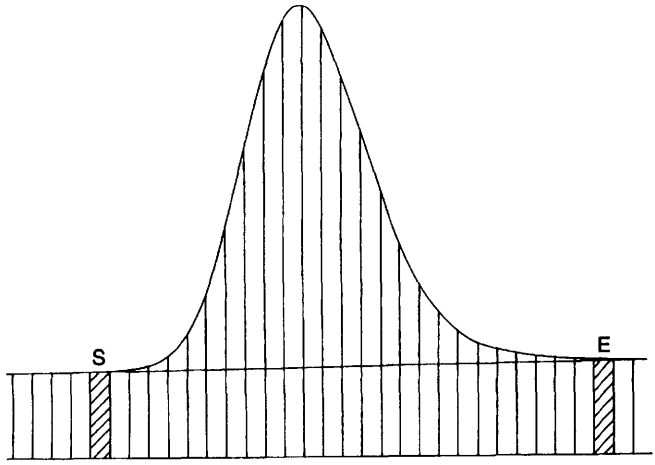


Figure 5.21 *Peak start,* S, *and end,* E

As the peak emerges, the integrals increase in value compared with their predecessors (Figure 5.22).

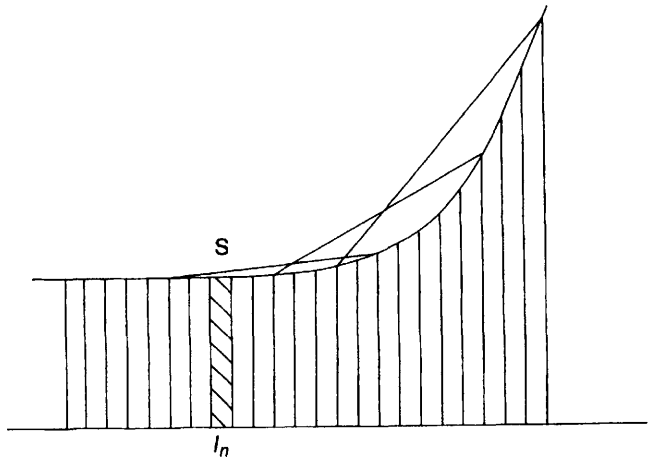


Figure 5.22 *The peak starts at* S *when* $\Delta I_n >$ *slope sensitivity*

To be recognized as the start of a peak and to trigger integration, ΔI_n must be positive and increasing in value for P consecutive integrals, where P has a value which depends on the moving window size (and manufacturer) but ranges from about 5 to 10. If it is too large the integrator will be slow to make decisions. If it is too small the integrator will measure residual noise peaks.

When the integrator observes P positive values of ΔI_n in succession, such that,

$$\Delta I_n > S \tag{15}$$

and

$$\Delta I_{n+1} > \Delta I_n \tag{16}$$

and

$$\Delta I_{n+2} > \Delta I_{n+1} \tag{17}$$

$$\vdots$$

and

$$\Delta I_{n+P} > \Delta I_{n+P-1} \tag{18}$$

and it satisfies the conditions of the peak shape test that $\mathrm{d}y/\mathrm{d}t$ and $\mathrm{d}^2y/\mathrm{d}t^2$ are both positive. The integrator triggers integrate mode which initiates several operations:

(1) data samples are accumulated from the peak start;
(2) a number of baseline integrals immediately before the peak start are averaged and stored;
(3) screen or keyboard indicators display when a peak is confirmed;
(4) a peak event mark (if available) will be added to the chromatogram.

When integrators and strip chart recorders were separate instruments, the event mark added by the integrator was drawn late by the recorder, at the point where the peak start was confirmed, not where peak onset was first noted. Just how late was determined by the peak width value. Now that intelligent plotters are part of the integrator, printing of the chromatogram is delayed slightly so that the event mark can be added to the correct place, but this has sacrificed a useful indicator of whether the peak width parameter has been programmed correctly.

Small Peak Filtering

Unless small peaks show positive slope for the correct length of time, $P \times \Delta t$, they will not trigger integration and will not be measured. Varying the bunch width (Δt) varies the qualifying time, allowing selective filtration of small peaks. This defines the fastest measurable peak.

Location of Peak End

The integrator continues to accumulate peak integrals until it locates the baseline again (Figure 5.23). The approximate location of peak end is determined when the

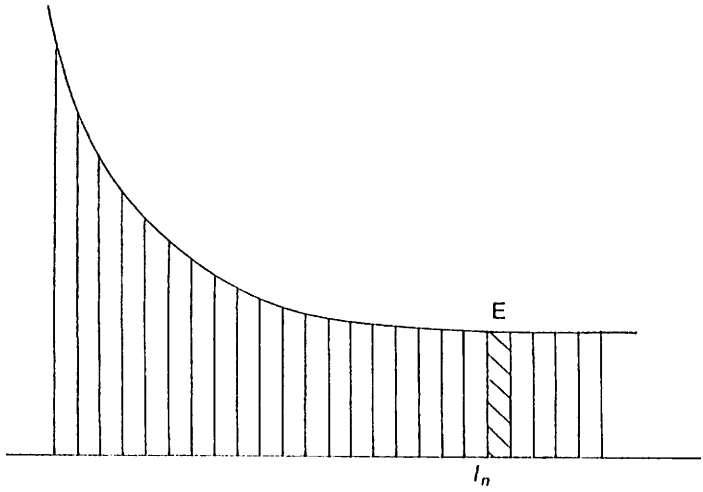


Figure 5.23 *Location of peak end*

detector signal gradient ΔI_n, becomes zero for a qualifying time of Z consecutive samples. Z is larger than P because of peak tailing which has to be monitored longer*. The integrator conducts a fine search among the successive data samples to find a more accurate location. The search range is determined by the time interval from the peak maximum to the approximate end, which will be large if there is tailing. The correct end point of the peak is taken to be the lowest baseline point in the search region.

When baseline is confirmed, more operations come into play:

(1) Integrate mode is terminated and accumulation of peak integrals stops at the peak end. The integrator now holds in its memory the sum of the data points from peak start to end (Figure 5.24).
(2) Several of the new baseline integrals immediately after the peak are averaged and stored.
(3) A peak event mark is recorded on the chromatogram, and keyboard peak indicators switch off.

Measurement of Peak Area

Peak area measurement is completed by subtracting the 'baseline area' from the accumulated integral count (Figure 5.25). The baseline area is calculated from the average of the stored baseline integrals before and after the peak, and the peak width (from the number of integrals accumulated within the peak limits). The area is calculated as a trapezium, and so the method is commonly called 'trapezoidal baseline correction'. There are three points to note:

*On at least one integrator $Z = 2P$

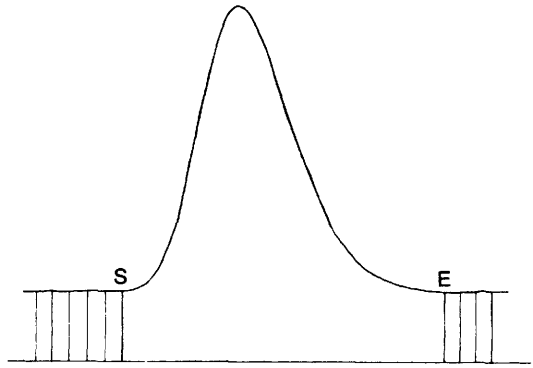


Figure 5.24 *Peak area between start and end*

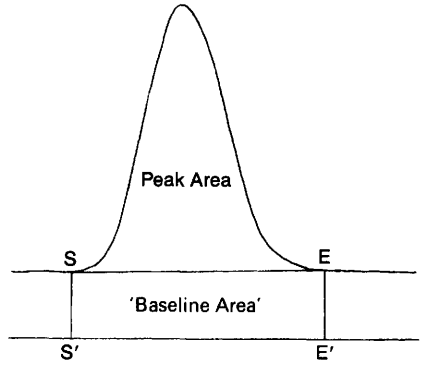


Figure 5.25 *The baseline area*

(1) Sloping baseline. It makes no difference whether the baseline is sloping or not. By calculating the baseline area beneath the peak as a trapezium, the integrator constructs a linear baseline which joins the beginning and the end of the peak and there is no assumption that these points are at the same level (Figure 5.26).

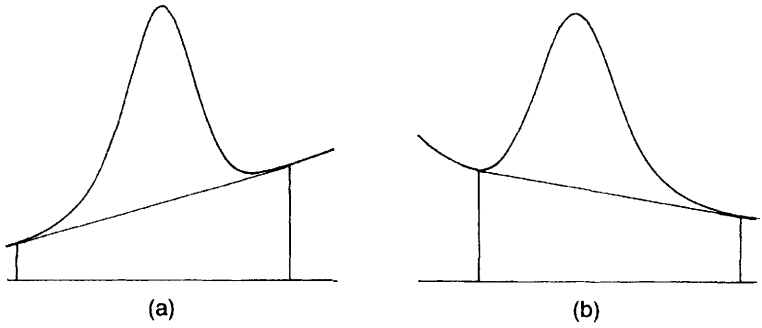


Figure 5.26 *Trapezoidal baseline correction compensates for a shift in baseline level*

(2) The interpolated baseline is a straight line. The integrator draws the same baseline that the chromatographer would draw with a pencil and ruler.
(3) The only assumptions made about peak shape are that the signal has gone up and come down again within the time limits allowed by P and Z.

Baseline Convention

Integrators construct linear baselines under peaks, although in many cases the real baseline is known to be curved.

It is a convention from the days of manual measurement that a reproducible straight line is better than a subjective curve simulation; no two chromatographers would draw exactly the same curve. After more than two decades of search, there is still no easy formula to construct the true baseline beneath a peak. There is the single exception of storing and subtracting non-linear baselines from a blank run.

The use of logarithmic or exponential models of baseline can be given general theoretical justification but it is difficult to justify fitting a curve to a specific peak if that peak can vary in shape from one analysis to the next.

Measurement of Peak Height

The procedures for measuring peak height are very similar to those for peak area. The difference is that, instead of accumulating peak integrals the integrator selects, as an approximation, the largest value (Figure 5.27) and fits the data in that region to a curve from which the height may be calculated. Trapezoidal baseline correction of this height gives the reported peak height.

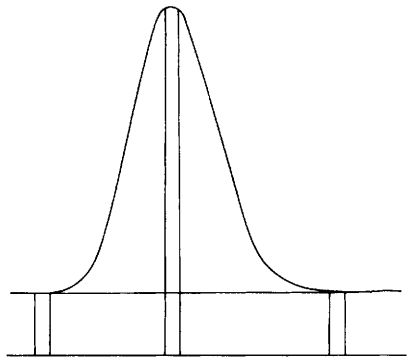


Figure 5.27 *For peak height measurement the integrator selects the largest smoothed datum*

Measurement of Retention Time

Retention time is measured as the time of the computed peak height. This is equivalent to determining when ΔI_n changes in value from positive to negative while in integrate mode.

The integrator monitors the trend in slope before confirming the peak maximum so that the signal has moved past the maximum before the change is noted. On real-time chromatograms the integrator prints the retention time after the peak maximum has eluted, when the detector signal is falling.

Measurement of Two Unresolved Peaks

As far as the integrator is concerned, the only difference between a fused pair of peaks and a single peak is the valley between them (Figure 5.28).

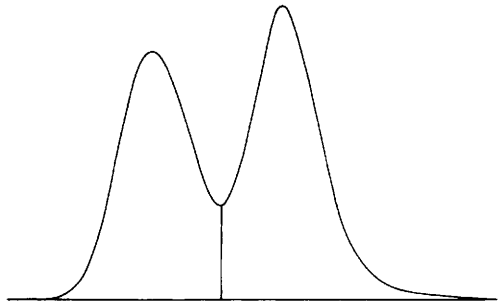


Figure 5.28 *Fused peaks*

The integrator locates the correct start of the first peak by means of a fine scan of the data and begins there to accumulate data until it locates the valley.

A **valley** is defined where the detector signal changes slope from negative to positive for P consecutive integrals while in integrate mode. It is approximately located where the Savitsky–Golay first derivative is zero between negative and positive slope values, and precisely located in this region by finding the curve minimum.

At the valley, accumulation of data samples from the first peak is ended and the total stored. Accumulation of the data samples of the second peak begins. The valley position determined from the curve fit does not in general coincide with the edge of a data sample, so the data sample that straddles the curve minimum is partitioned to each peak (Figure 5.29).

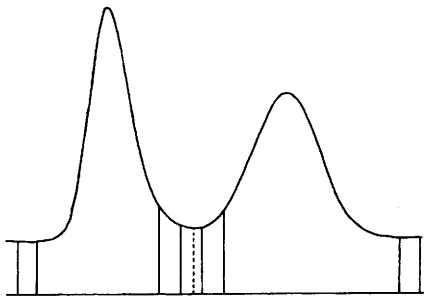


Figure 5.29 *Partitioning of data sample at curve minimum*

The peak end is located, as for a single peak, in a fine search beyond the approximate location, and accumulation of data for the second peak is terminated.

During this time, peak event marks and retention times have been added to the chromatogram and the keyboard flags will have indicated the peaks' presence.

Measurement of Individual Peak Areas

The trapezoidal 'baseline area' beneath both peaks is computed from the average of the stored baseline integrals before and after the peaks and the base width of the peaks. The area is divided at the valley into the two parts corresponding to the areas below each peak, and each peak area is calculated by subtracting each baseline area from the peak total (Figure 5.30).

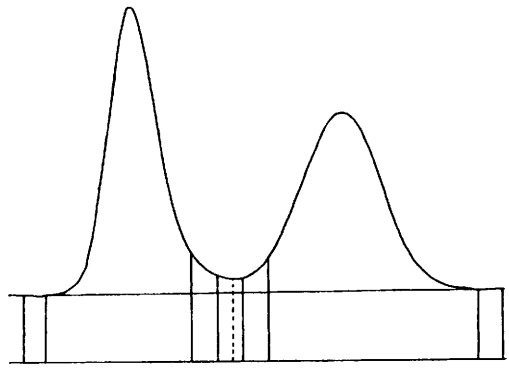


Figure 5.30 *'Baseline area' for fused peaks*

Peak Measurement Diagnostics

In the Final Report the integrator highlights peaks which have been separated by perpendiculars by printing a diagnostic code alongside the peak area. The code varies with manufacturer but is usually a letter like P (perpendicular) or V (valley); sometimes it is a numeric code.

Measurement of Fused Groups

Larger groups of unresolved peaks simply have more valleys. The start of the first peak and the end of the last are measured as in the case of a single peak. Peak integrals up to each valley are summed and stored and the peak retention times are measured.

The trapezoidal baseline area beneath the whole group is measured, and divided at the valleys into the baseline areas below each peak; these individual baseline areas are subtracted from the peak integrals to yield each peak area (Figure 5.31).

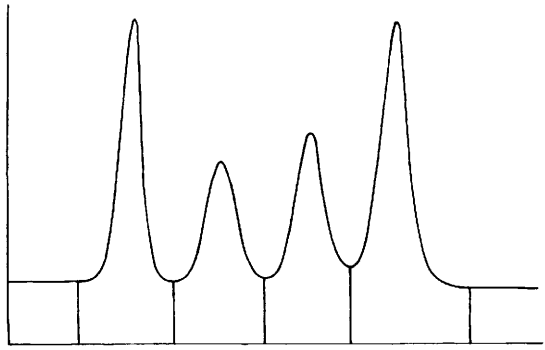


Figure 5.31 *A group of fused peaks*

None of the peak area measurements can be completed until the baseline is located at the end of the group and trapezoidal baseline correction is made. In WCOT 'forests' the whole of the chromatogram may be in temporary storage until the signal returns to baseline at the end of the analysis (Figure 5.32).

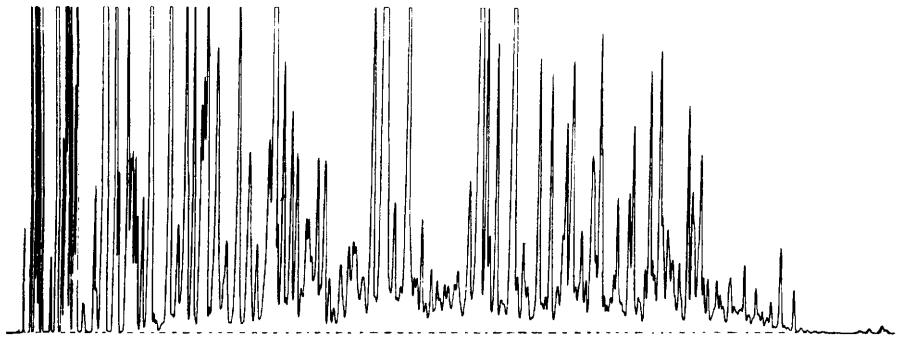


Figure 5.32 *In a WCOT forest the area of the first peak cannot be measured before the baseline is found at the end of the chromatogram*

Shoulders

A shoulder is an unresolved peak on the leading or trailing edge of a larger peak. There is no true valley in the sense of negative slope followed by positive slope.

Integrators define shoulders as regions of zero slope between two positive or two negative slopes, while in integrate mode.

What actually counts as 'zero' slope is determined by slope sensitivity, and relatively steep shoulder slopes can be made to be 'zero' by temporarily increasing the value of slope sensitivity in a Time Program to force shoulder recognition and peak detection (Figure 5.33).

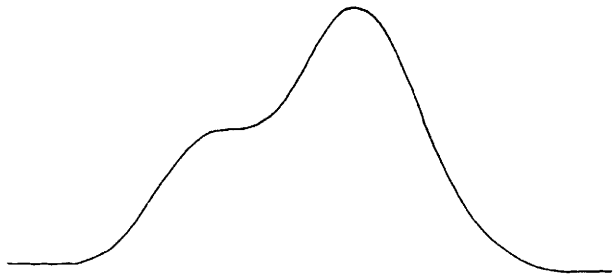


Figure 5.33 *Varying slope sensitivity can force shoulder recognition*

When it finds a shoulder, the integrator will process it as a valley and drop a perpendicular from the point of minimum slope.

Shoulder measurements are not accurate. The area of the shoulder, or smaller peak, is over-estimated[35] and includes part of the larger peak. Both areas are inaccurate. No important peak should be a shoulder; separation must be improved. If this is not possible, conclusions drawn from shoulder measurements should be used with caution.

Measurement of Tangent Peaks

At each valley, integrators perform an additional size test to see whether the next peak should be separated by a perpendicular or whether it should be skimmed by a tangent.

Historically and commonly, tangent skimming implies small peaks on the tails of big peaks, but some integrators offer the facility to skim from the leading edge too.

The decision to skim or not is based on relative peak size and the height of the valley between the peaks. A size ratio or threshold is set above which peaks are separated by a perpendicular, and below which a tangent is used. Some integrators can vary this threshold, others have it fixed with provision to override it and force separation exclusively by perpendicular or by tangent (Figure 5.34).

A tangent is skimmed if,

$$\frac{H_1 - V}{H_2 - V} > \text{Size ratio} \tag{19}$$

If not, the peaks will be separated by perpendicular cleavage.

The tangent is drawn from the valley before the small peak to that point after it where the detector signal gradient (ΔI_n) is equal to the tangent gradient. The size ratio is usually about 10:1 and not as numerically critical as might first be thought. It only determines whether a particular peak should be skimmed: it has no influence on the peak measurement once this decision is made. Provided that the integrator skims those peaks which the chromatographer expects to see skimmed, the value of the size ratio will not be questioned. (See also Figure 2.22.)

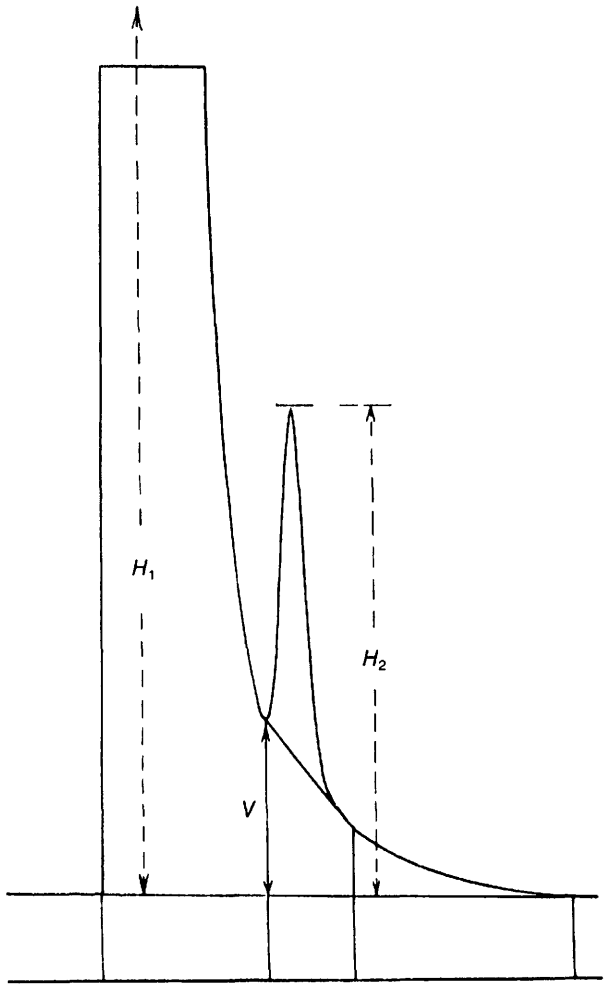


Figure 5.34 *Tangent skimming*

Tangent Diagnostics. Any peak which has been tangent skimmed off another peak will have a diagnostic code, *e.g.* 'T', printed against it in the final report.

Tangent Skimmed Groups

If a group of fused peaks is situated on the tail of a larger peak, the size ratio test may decree that the whole group is skimmed from the larger peak but separated from each other by perpendiculars dropped to the tangent drawn beneath the group (Figure 5.35).

In the final printed report, each peak in the group will have a diagnostic code indicating that it was skimmed from the larger peak and separated from the other skimmed peaks by perpendicular.

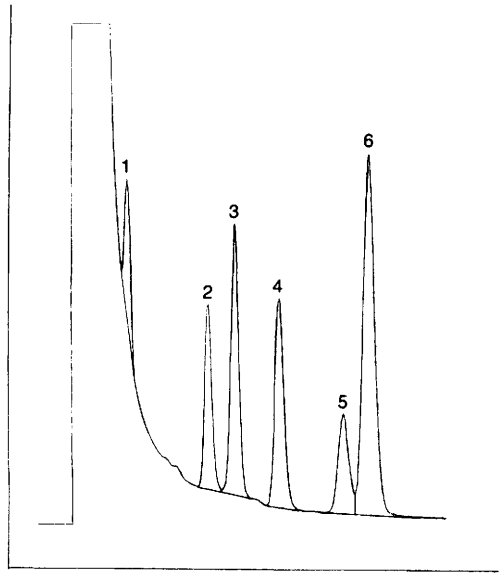


Figure 5.35 *Peaks 1 to 6 are skimmed on size comparison with the solvent. Peak pairs 2 and 3, and 5 and 6 are separated by perpendiculars on size comparison with each other*

Small Peaks between Larger Ones

If there are small peaks between larger ones it is perfectly possible that some of them will be selected for skimming while others are separated by perpendiculars. Where the tangent peak straddles the valley position between two larger peaks, it prevents location of the valley which is then shifted to the nearer edge of the smaller peak (Figure 5.36).

There is no integrator solution to this problem; the only solution is improved peak resolution.

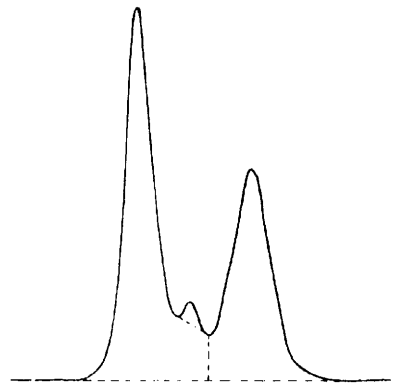


Figure 5.36 *A small peak between larger ones*

Tangents on Tangents

It is possible, though in practice unlikely, that three consecutive peaks might have the appropriate relative sizes to cause the second peak to be skimmed from the first, and the third peak to be skimmed from the second (Figure 5.37).

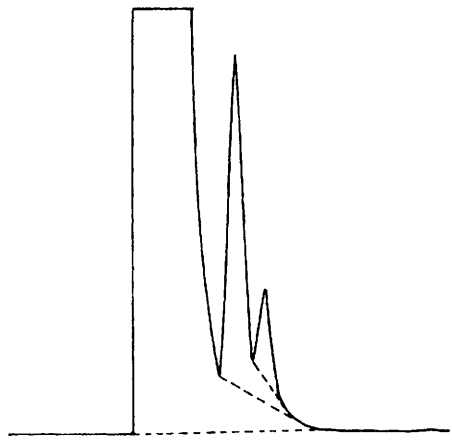


Figure 5.37 *Tangents on tangents*

4 Baselines, A More Detailed Discussion

An integrator monitoring an analysis and not subject to an integrate inhibit instruction is either 'on baseline' or in 'integrate' mode. When it is on baseline it will continue to recognize baseline until it identifies the start of a peak and enters integrate mode. It cannot leave integrate mode until it relocates, or is instructed to locate baseline.

Until now, baselines have been considered to be the detector signal in the absence of the peaks and described as 'zero slope' within the tolerance allowed by slope sensitivity, but this is an inadequate definition. Real baselines can slope as a result of solvent or temperature programming, and there are many parts of a chromatogram which show zero slope but are not baseline, for example, valleys and the tops of peaks. The tail of a solvent peak is baseline for the small peaks riding it, yet it may be falling almost vertically. Figure 5.38 shows some examples.

Positive sloping baselines are a greater problem than negative ones since unwanted or mis-timed peak detection will occur. Some positive slope can be tolerated within the definition of slope sensitivity but this does not prevent the integrator from finding 'zero' slope at peak tops or in valleys if they last long enough to qualify as baseline, *i.e.* for Z consecutive samples.

Baseline Drift Limit

A better definition of baseline limits the regions where it may be established to where the analyst expects to find it. A second parameter is required to define the

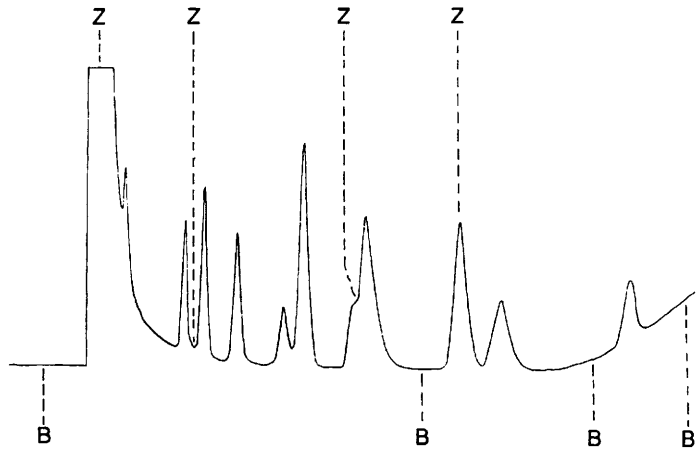


Figure 5.38 *Not all baseline is zero slope, not all zero slope is baseline.* Z = *zero slope;* B = *baseline*

boundaries of baseline position and say how far the baseline position may shift during a peak or group. The parameter has a variety of names from different manufacturers but is a 'baseline drift limit' which comes into play only when a peak is detected. Above the limit, zero slope will not be recognized as baseline no matter how long it lasts.

There are two possible ways to define limit: as a fixed millivolt threshold (Figure 5.39), or as a gradient which increases the allowed range with time (Figure 5.40). The fixed threshold cannot account for unexpected shifts in baseline which will lock the integrator into integrate mode until the end of the analysis and measure peak areas down to the pre-shift baseline. A rising limit eventually allows the correct baseline to be established. Peaks measured before the return to baseline may be measured incorrectly, but later peaks will be preserved.

Neither concept is perfect. Stored analyses can be reprocessed with corrections provided that the analyst is aware of what has happened. Integrators which display the baseline on chart or VDU are a great help in this respect.

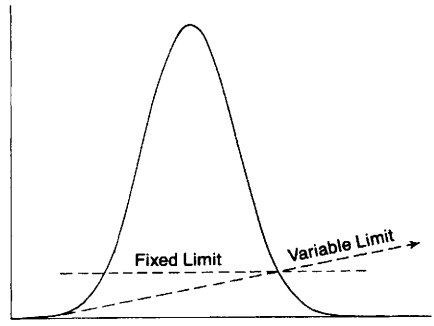


Figure 5.39 *Limiting the range of baseline drift to below the dashed line*

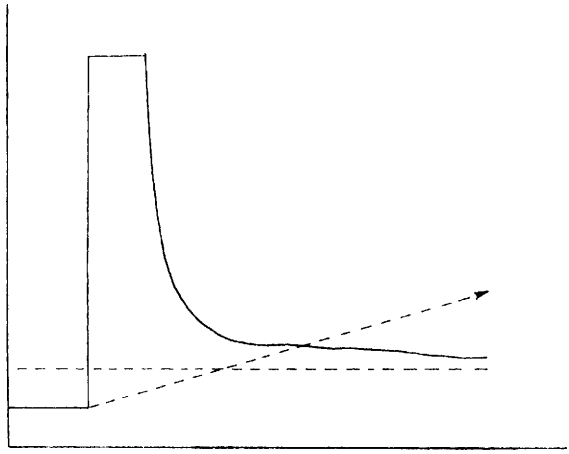


Figure 5.40 *A fixed baseline limit cannot cope with excessive tailing*

False Starts

One function of a baseline drift limit is to prevent integration of positive sloping baselines as peaks. Integration is automatically terminated by the integrator when the signal levels off within the allowed baseline range for the qualifying time.

False starts are common on noisy baselines; associated event marks litter the baseline and confuse interpretation (Figure 5.41).

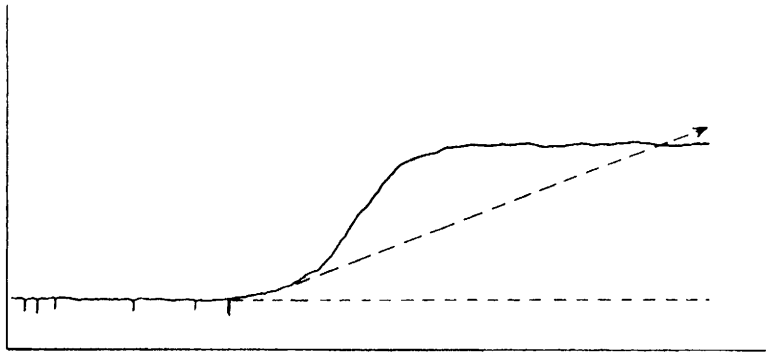


Figure 5.41 *Shifts in baseline can be 'seen' as peaks*

Formulating a Baseline Definition

There is no simple neat definition of a baseline. Integrators use a selection of definitions which can be invoked collectively or individually. To an integrator, baseline is defined as:

(1) the detector signal encountered immediately after pressing 'start' (of analysis);
(2) the detector signal encountered immediately after the end of an 'integrate inhibit' instruction;
(3) after a peak, zero slope within slope sensitivity, lasting for Z consecutive data samples below the baseline drift limit;
(4) wherever the analyst forces the integrator to ignore its own logic and establish baseline;
(5) the detector signal at the end of analysis, *i.e.* the last stored datum.

The consequences and short-comings of these definitions are best illustrated by example as shown below.

Mis-timing 'Start' and 'End'

The shaded areas of Figure 5.42 show the effect of starting the analysis too late or ending too early. The measured areas are incomplete. This type of error happens if automated analyses lose synchronization between injection time and integrator start (Figure 5.43).

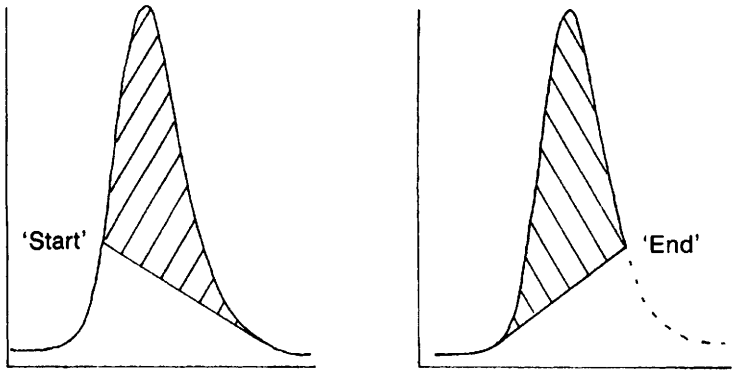


Figure 5.42 *Only the shaded area is measured*

End of 'Integrate Inhibit'

Integrate inhibit is a programmable function used to prevent the measurement of a real but unwanted peak such as the solvent, or to prevent baseline disturbances from interfering with peak measurement.

The end of integrate inhibit is timed to establish a section of true baseline before a peak of interest emerges (Figure 5.44).

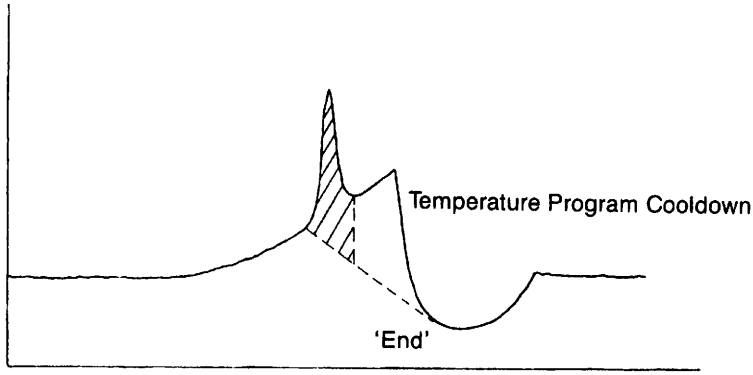


Figure 5.43 *Analysis ends late, after cooldown has started*

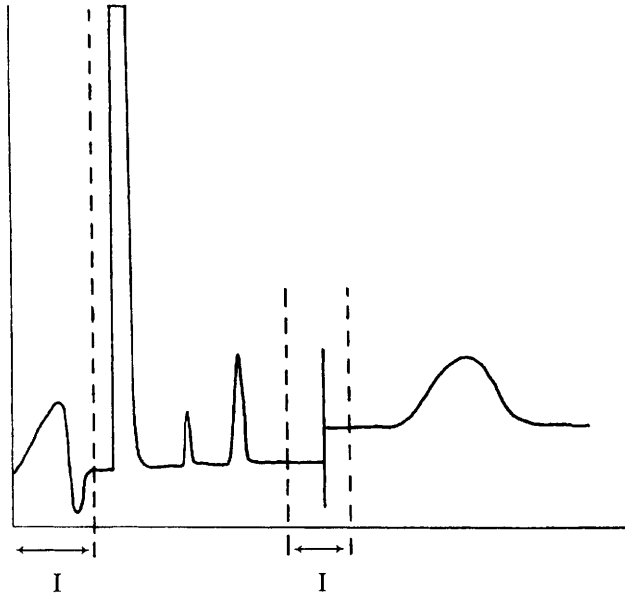


Figure 5.44 *Integrate inhibit suspends peak recognition*

Forcing Baseline

The purpose of forcing baseline is to rescue a peak from incorrect measurement resulting from baseline disturbances. If the baseline near a peak is disturbed by a negative dip or a pressure pulse it will cause at least one integration limit to be established in the wrong place, cause the trapezoidal baseline beneath the peak to be wrongly placed and the peak area to be measured incorrectly.

Baseline disturbances too close to a peak will distort the peak shape and preclude accurate measurement. Forcing baseline is only sensible when there is a small

stretch of real baseline on either side of the peak to latch on to. If there is none, further method development to distance the peak from the disturbance is essential.

Forced baseline can also direct the integrator to draw a baseline between points on either side of the peak, which, in the analyst's judgement, will form an accurate trapezoidal baseline under the peak and allow correct measurement. (See Figure 5.45.)

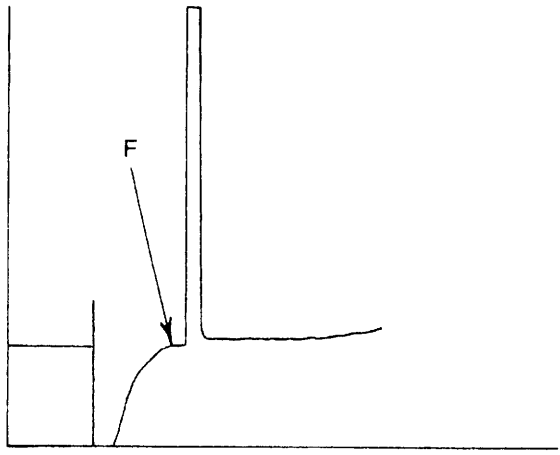


Figure 5.45 *Force baseline at* F

Integrate Inhibit and Forced Baseline often have the same role.

Figure 5.46 shows a computer-generated chromatogram of four equally-sized Gaussian peaks preceded by a baseline disturbance which interferes with baseline placement. In chromatogram (a) the peak measurements are inaccurate, there is even a 'phantom' peak at 0.685 min, where the integrator has been triggered into measurement by the rising baseline, and this measurement has continued to the 'valley' where the first real peak begins. In chromatogram (b) baseline has been forced at 1 min and peak measurement is thereafter accurate. It would have been equally as effective to program Integrate Inhibit for 1 min.

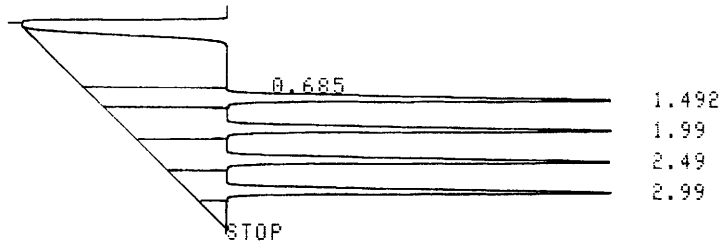
Incorrect Programming of Baseline Drift Tolerance

Low valleys between nearly-resolved peaks are always a problem. The drift tolerance must be programmed to recognize the lowest valley in the chromatogram (Figure 5.47).

The analyst should check that the integrator has correctly found a valley by referring to the VDU display or diagnostic codes printed alongside the peak areas in the final report. When a valley is found, a valley diagnostic will be printed. If no valley code is there, it is because the integrator has considered the valley to be a baseline point and so has no valley to report.

Unless the integrator draws the baseline on the chromatogram or shows it on a VDU, the analyst has only the diagnostic codes allocated to these peaks to confirm

(a)

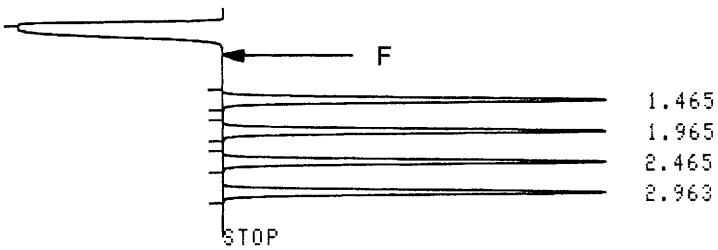



```
CHROMATOPAC    C-R6A                    FILE          0
SAMPLE NO    0                          METHOD         21
REPORT NO    13

PKNO       TIME       AREA   MK   IDNO      CONC        NAME

   1      0.685    1243553                  34.2536
   2      1.492     694225   V               19.1224
   3      1.99      724739   V               19.9629
   4      2.49      570863   V               15.7244
   5      2.99      397055   V               10.9368
                  ---------              -------------
         TOTAL     3630434                    100
```

(b)




```
CHROMATOPAC    C-R6A                    FILE          0
SAMPLE NO    0                          METHOD         21
REPORT NO    14

PKNO       TIME       AREA   MK   IDNO      CONC        NAME

   1      1.465     316840                  24.9991
   2      1.965     316845                  24.9994
   3      2.465     316879                  25.0022
   4      2.963     316844   V              24.9994
                  ---------              -------------
         TOTAL     1267408                    100
```

Figure 5.46 *Effect on measured areas of a baseline disturbance*

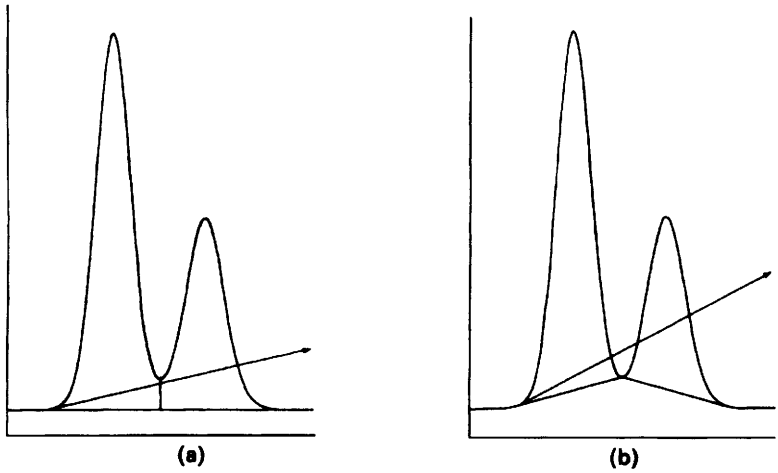


Figure 5.47 (a) *The valley is outside of baseline range and remains a valley*: (b) *The integrator locates zero slope within baseline range inside the valley and establishes baseline*

that a valley has been correctly identified between them. The ratio of the peak areas will not differ much between Figure 5.47(a) and (b) and cannot be used as a diagnostic. Even with this low valley, the area of the smaller peak is changed by approximately 10%.

Measurement discontinuities occur if the integrator measures some peak pairs by method (a) and some by method (b) in Figure 5.47 over a series of analyses. This problem can be solved, *i.e.* peaks can be measured consistently, as in (a), by programming the drift parameter for the lowest valley and programming slope sensitivity for the smallest peaks, but this may bring about measurement of other small and unwanted peaks elsewhere in the chromatogram. In this case use a Time Program to set different drift tolerances at different points on the chromatogram.

A related problem arises when overlap is caused by asymmetry in the earlier peak, which either varies with the size of the peak or with the condition of the column. The measurement transition can occur again. When peak asymmetry is an indication of column condition, the interval between conditioning or replacing columns may be defined in terms of preventing such measurement transitions.

Fused Tangent Measurement

A solvent tail is baseline to the peaks which ride on it. The valley between fused riders may lie below the start of the first rider peak, and require a *negative* baseline limit to ensure that the valley is not recognized as a baseline point (Figure 5.48a). This is normally done automatically by the integrator but if the baseline drift tolerance has been programmed by the analyst, it may not be automatic (Figure 5.48b).

The analyst must check the VDU or the diagnostics allocated to these peaks to confirm that a valley has been correctly identified between them.

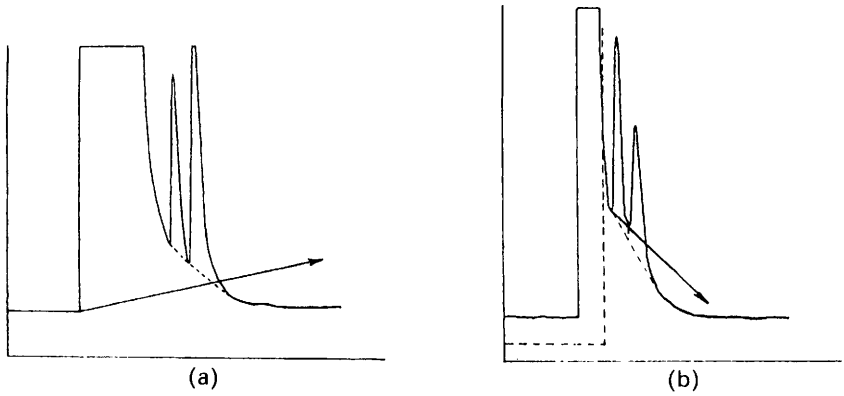


Figure 5.48 (a) *Fused tangent measurement*: (b) *If the solvent peak is ignored by integrate inhibit, a negative drift limit is required for rider peaks*

Single Peaks on a Rising Baseline

Single peaks on a rising baseline have additional problems with false early start and false end of peak measurement (Figure 5.49).

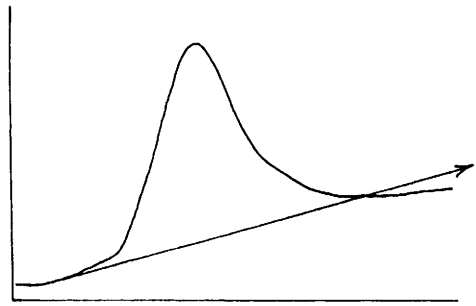


Figure 5.49 *A single peak on a rising baseline*

Slope sensitivity must be tolerant of the baseline gradient but pick up the increased peak gradient as positive slope. The baseline drift tolerance must also allow the 'valley' at the end of the peak to establish itself as a baseline point – which is the complete opposite of what is required for fused peaks.

Peaks on fast-rising baselines are difficult for any integrator to deal with, especially if the analysis is not easy to repeat, and baseline points forced on a time basis cannot be relied on.

The analyst should attempt to improve the chromatography by reducing the baseline gradient or improving retention time stability. If this cannot be done, there is no alternative but to store and reprocess each analysis in turn for the best results.

Valleys Between Fully-resolved Peaks

It is possible for integrators to drop a perpendicular of zero length (Figure 5.50). The peaks in Figure 5.50 are only just fully-resolved but there is no qualifying baseline between them. If baseline drift is restricted so that a valley is forced, the integrator will drop a perpendicular of zero length. With or without baseline detection the area is the same.

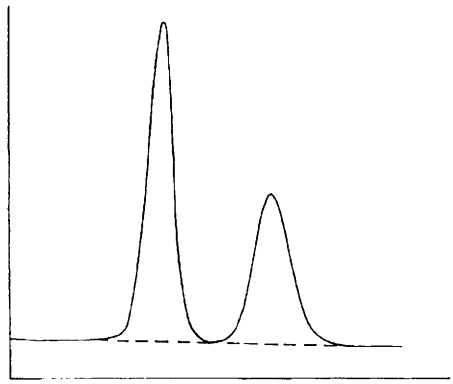


Figure 5.50 *Fully-resolved peaks with a valley between*

Negative Dips and Constructed Baselines

Negative dips in the detector signal, arising from negative peaks, injection pulses, *etc.*, are major sources of error in baseline assignment. When the signal dives into a trough the integrator will track it, at least as far as the lower limit of its input range. When the signal returns towards its original level, peak detection is triggered on the positive slope and original baseline position is measured from the bottom of the trough or the edge of the operating range. True baseline is only relocated when the baseline drift parameter returns slowly from the trough. The constructed baseline and the drift vector are one.

The depth of the trough and the relatively slow rise of the baseline boundary create an artificial baseline above which peaks are measured. Very small peaks are given incongruously large areas because of the area below the true baseline (Figure 5.51).

If a series of small peaks, all roughly the same size, have reported areas decreasing in size away from a baseline dip it is a sure sign that the baseline has been taken from the bottom of the trough (Figure 5.52, and see also Figure 5.46a).

Negative baseline excursions often create a 'phantom' peak (Figure 5.46) which is reported with retention time and area in the final report.

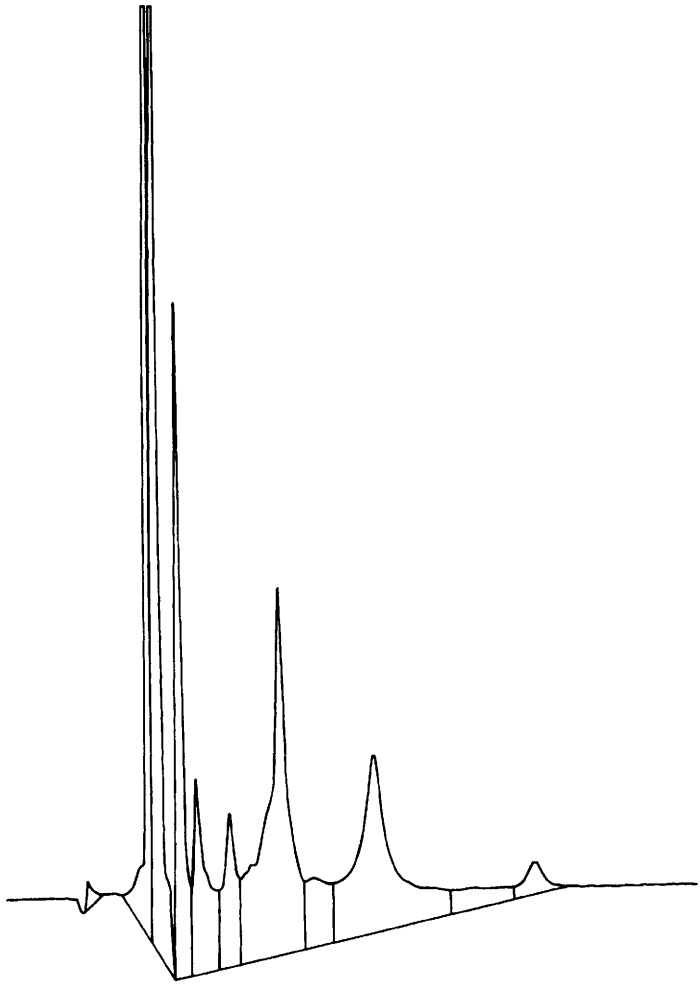


Figure 5.51 *Negative baseline dips distort baseline allocation*

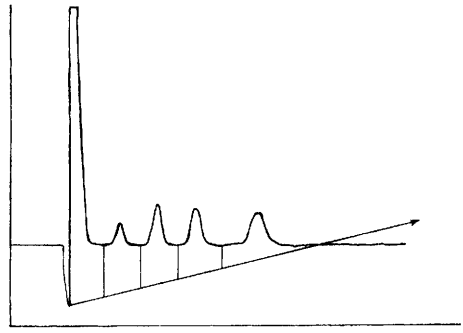
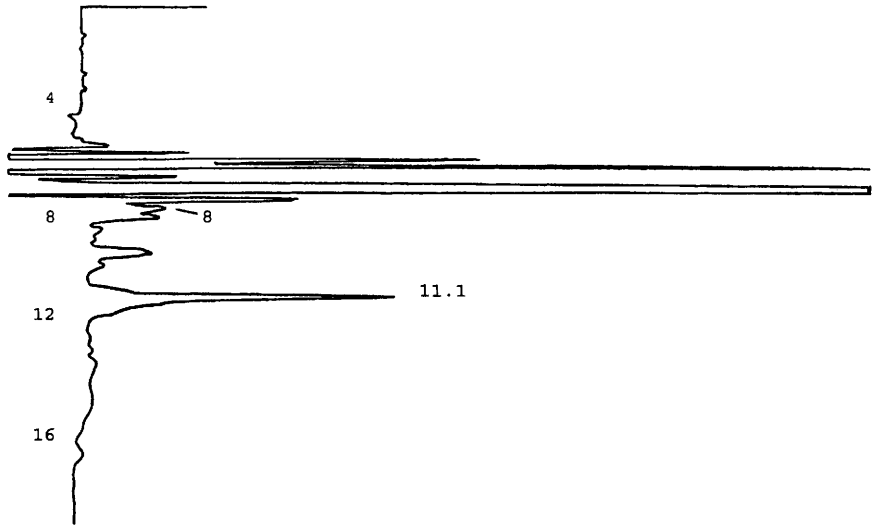


Figure 5.52 *Baseline taken from the bottom of a trough. Baseline drift vector* is the baseline

Worked Example. Figure 5.53 is a real chromatogram of a clinical sample. The report shows that the peaks at 8 and 11.1 min very nearly have the same height, 944 and 968 μV, respectively. This is obviously wrong but the cause of the error, baseline dips, is not immediately obvious because no baseline is drawn. The major peaks at 6 and 7 min dip off scale and force the integrator to draw the baseline indicated in Figure 5.54. When the heights of the two peaks are measured above this sloping baseline, they are approximately equal – but only because it is the wrong baseline.




```
CHROMATOPAC        C-R3A                    FILE            2
SAMPLE NO          0                        METHOD          3443
REPORT NO          1386                     SAMPLE WT       100
IS  WT             1
```

PKNO	TIME		HEIGHT	MK	IDNO	CONC
2	8	*	944	V	2	513.2987
			23865 ar			
3	8.333		871	V	2	473.916
			15441 ar			
5	9.567		639	V	3	350.4407
			30799 ar			
7	11.1	*	968	V	1	
			32252 ar			
			----------			----------
	TOTAL		3422			1337.6553

Figure 5.53 *The peaks at* 8 *and* 11.1 min *have the same height (!)*

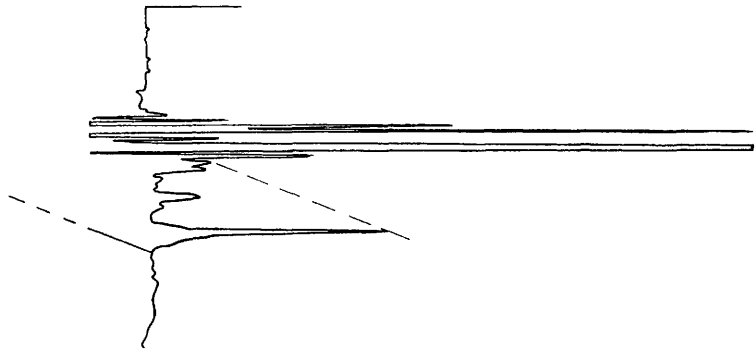


Figure 5.54 *Above the integrator's baseline the peaks are the same height. But it is the wrong baseline*

The constructed baseline is drawn from the bottom of the lowest valley, or from −5 mV if the detector signal drops below the measuring range of this integrator, to the end of the tail of the peak at 11.1 min – not earlier because the diagnostic marks in the report would have indicated baseline and not valleys (V).

As the analyst only wanted to measure the peak at 11.1 min, integrate inhibit was applied for 9.5 min. The baseline beneath the 11.1 min peak was constructed correctly and the peak measured accurately.

This example shows why integrators that draw baselines on chromatograms are preferred to those which do not.

Assigning Baseline Beneath the Whole Chromatogram

On many isothermal or isocratic chromatograms a line drawn from the start to the end of analysis will underscore the trace so that no part of the chromatogram descends below. This line can be regarded as the true baseline (Figure 5.55).

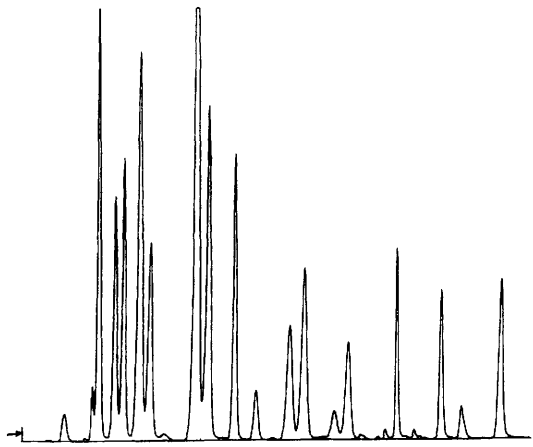


Figure 5.55 *Level baseline*

Where the baseline is sloping, the integrator will assign straight baselines under single peaks and under each fused group. If the shift in baseline across a group is large, some valleys within the group may project below this construction.

An integrator draws the baseline below a group of peaks in stages. The first construction will join the start of the first peak to the end of the last peak with a straight line (Figure 5.56) provided that the end point is within the range allowed by the baseline drift parameter. If no valley points intersect this baseline no amendments are necessary and it will be used in the measurement of peak areas.

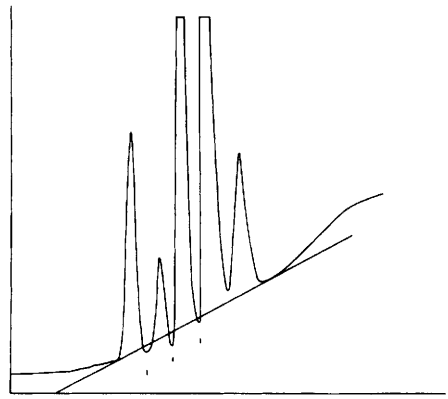


Figure 5.56 *First baseline attempt – valleys descend below it*

If valleys do intersect the baseline, the integrator registers the first of the valleys as a baseline point and draws a baseline section from the start of the first peak to this valley. A second section is drawn from the valley to the end of the group (Figure 5.57). The situation is re-assessed to determine whether any valleys intersect the reconstruction. If none does, no further amendments are required and the peaks can be measured.

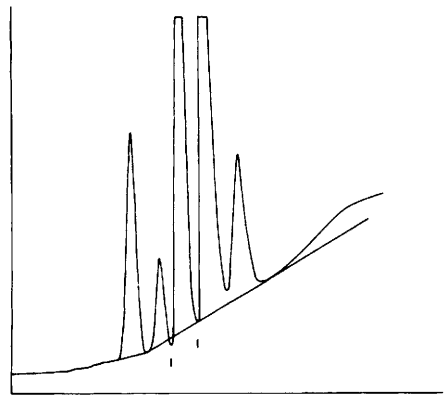


Figure 5.57 *Second attempt; baseline updated to first valley which descended, but a valley still descends below second attempt*

If intersecting valleys still exist, a new section of baseline is drawn from the first baseline valley to the next intersecting valley which is confirmed as a baseline point. A third section of baseline is drawn from the second 'baseline valley' to the end of the group, and this new section is inspected for intersecting valleys (Figure 5.58).

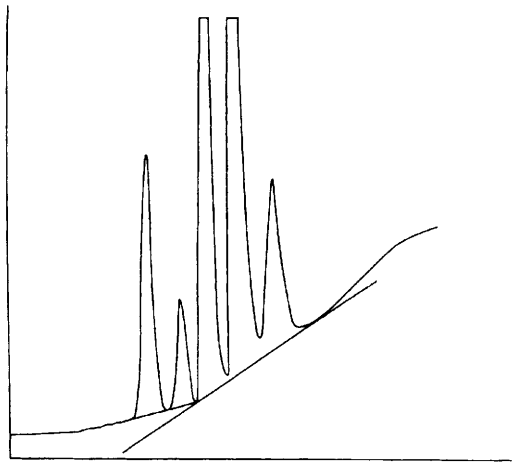


Figure 5.58 *Third attempt; baseline updated to second descending valley. This time there are no remaining valleys which cut through the baseline*

The process continues, moving forward in time, until a baseline is constructed which skirts round the lowest valleys to the tail of the last peak. Perpendiculars are dropped from the other valleys to the completed baseline. This has been likened to stretching a rubber band round the underside of the peaks touching the start and end points of the group and all the low valleys (Figure 5.59).

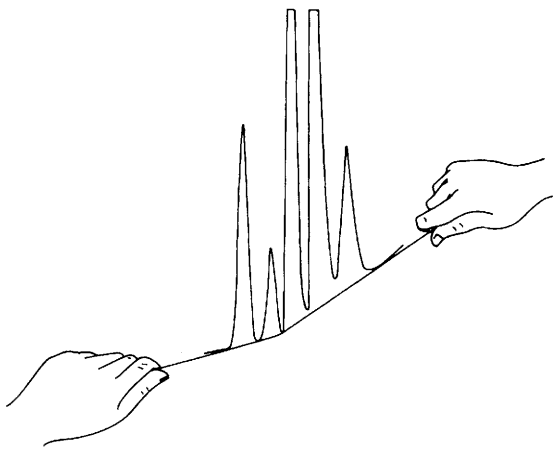


Figure 5.59 *The Elastic Band Technique. Integrators fit a baseline under a group of peaks in a manner analogous to stretching an elastic band around the underside*

As valleys become confirmed baseline points, other valleys which might have been designated as baseline points in earlier constructions become re-designated as valleys by the updates.

Valley–Valley Skim

Some analysts have in the past preferred to process a complex, unresolved bunch of peaks by skimming a baseline around the valley bottoms rather than to draw one baseline across the whole group and drop perpendiculars to it. This can be done by giving the baseline drift a high tolerance so that each valley (or at least some) comes within baseline range as the baseline updates itself to the preceding valley (Figure 5.60).

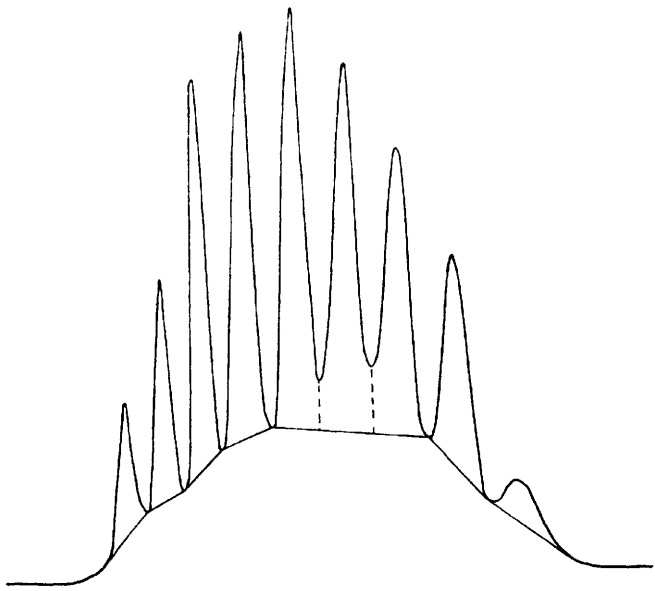


Figure 5.60 *Valley–valley skim. It ignores the mass beneath the peaks*

This technique is completely inaccurate, it ignores the mass below the valley–valley baseline. Improvements in WCOT technology to increase peak capacity and reduce column bleed are reducing the need for it.

5 Conclusions

It should be clear now that integrators are like any other tool – an excellent thing in the right hands. What they do best is measure peaks which are suitable for measuring, rapidly and without tedium. If these measurements are worth making then all subsequent calculations are worth noting and perhaps acting upon.

As long as integrators use perpendiculars and tangents and draw straight baselines beneath peaks, they are of use only in controlled circumstances, when the chromatography is good. Even then, the use of integrators requires vigilance from the operator and skill in assessing and assigning parameters. Integrators cannot improve bad chromatography, only the analysts can do that – and at the end of the day that is what they are paid for.

6 References

1. G.B. Clayton, *Data Converters*, Macmillan Press, London, 1982.
2. H.H. Willard, L.L. Merritt, Jr., J.A. Dean and F.A. Settle, Jr., *Instrumental Methods of Analysis*, Wadsworth Pub. Co., CA, 7th edn., 1988.
3. D.A. Skoog, *Principles of Instrumental Analysis*, Saunders College Pub., Philadelphia, PA, 3rd edn., 1985.
4. Burr Brown Application Note PDS 372 **B**, 'VFC32 V/F and F/V Converter', Tucson, AZ, 1982.
5. Z. Hippe, A. Bierrowska and T. Pietryga, *Anal. Chim. Acta*, 1980, **122**, 279.
6. A.H. Anderson, T.C. Gibb and A.B. Littlewood, *Anal. Chem.*, 1970, **42**, 434.
7. J. Novak, K. Petrovic and S. Wicar, *J. Chromatogr. Sci.*, 1971, **55**, 221.
8. S.N. Chesler and S.P. Cram, *Anal. Chem.*, 1971, **43**, 1922.
9. M. Goedert and G. Guiochon, *Chromatographia*, 1973, **6**, 76.
10. D.T. Rossi, *J. Chromatogr. Sci.*, 1988, **26**, 101.
11. R. Bracewell, *The Fourier Transform and its Applications*, McGraw-Hill, New York, 1965.
12. J.T. Tou, *Digital and Sampled Data Control Systems*, McGraw-Hill, New York, 1959.
13. K. Kishimoto and S. Musha, *J. Chromatogr. Sci.*, 1971, **9**, 608.
14. J.D. Hettinger, J.R. Hubbard, J.M. Gill and L.A. Miller, *J. Chromatogr. Sci.*, 1971, **9**, 710.
15. A. Fozard, J.J. Frances and A. Wyatt, *Chromatographia*, 1972, **5**, 377.
16. D. Kantoci, *J. Liquid Chromatogr.*, 1997, **20**, 1049.
17. P.C. Kelly and G. Horlick, *Anal. Chem.*, 1973, **45**, 518.
18. K.D. Jackson, S.J. Walton and D. Campbell, *J. Auto. Chem.*, 1997, **19**(5), 145.
19. C.R. Mittermayr, H. Frischenschlager, E. Rosenberg and M. Grasserbauer, *Fresenius' J. Anal. Chem.*, 1997, **358**, 456.
20. L.J. Lorenz, R.A. Culp and L.B. Rodgers, *Anal. Chem.*, 1970, **42**, 979.
21. L. Ghaoui and L.D. Rothman, *J. High Resolut. Chromatogr.*, 1992, **15**, 36.
22. R.Q. Thompson, *J. Chem. Educ.*, 1985, **62**, 866.
23. A. Savitsky and M.J.E. Golay, *Anal. Chem.*, 1964, **36**, 1627.
24. D.L. Massart, B.G.M. Vandeginste, S.N. Deeming, Y. Michotte, and L. Kaufman, *Chemometrics: a Textbook*, Elsevier, Amsterdam, 1988, ch. 15.
25. S.P. Cram, S.N. Chesler and A.C. Brown, *J. Chromatogr.*, 1976, **126**, 279.
26. J. Steiner, Y. Termonia and J. Deltour, *Anal. Chem.*, 1972, **44**, 1906.
27. M.U.A. Bromba and H. Zeigler, *Anal. Chem.*, 1983, **55**, 1299.
28. M.H.J. van Rijswick, 'Philips Res. Report Sup.', Centrex Pub. Co., Eindhoven, No. 7, 1974.
29. S.P. Cram, S.N. Chesler and A.C. Brown, *J. Chromatogr.*, 1972, **44**, 1906.
30. E.J. Van den Heuval, K.F. Van Malssen and H.C. Smit, *Anal. Chim. Acta*, 1990, **235**, 343.

31. E.J. Van den Heuval, K.F. Van Malssen and H.C. Smit, *Anal. Chim. Acta*, 1990, **235**, 355.
32. Th. Noij, J.A. Rijks, A.J. Van Es and C.A. Cramers, *J. High Resolut. Chromatogr.*, 1988, **11** (December), 862.
33. B.H. Vassos and G.W. Ewing, *Analog and Digital Electronics for Scientists*, John Wiley and Sons, New York, 1972.
34. European Patent Application 0296781A2, Date 20.6.88. Applied Biosystems Inc.
35. A.W. Westerberg, *Anal. Chem.*, 1969, **41**, 1770.

Author Index

In this index the number in parentheses is the Chapter number of the citation and this is followed by the reference number or numbers of the relevant citations within that Chapter.

Aitken, R.L., (2) 67
Anderson, A.H., (2) 49, 50; (5) 6
Anderson, D.J., (1) 11
Andrews, F., (2) 17

Baeza-Baeza, J.J., (2) 45
Bahowick, T.J., (2) 59
Bakalyar, S.R., (2) 9, 12
Ball, D.L., (3) 5, 14
Barber, W.E., (2) 19
Barford, R.A., (2) 58
Barr, J.K., (4) 1
Barth, H.G., (2) 58
Batteley, I.G., (1) 14
Baumann, F., (3) 16
Bierowska, A., (2) 52; (5) 5
Bocek, P., (2) 54
Boehm, B.W., (4) 32
Bonelli, E.J., (3) 15
Bonsall, R.G., (3) 2
Borowski, E.J., (1) 31
Borwein, J.M., (1) 31
Bosman, T., (2) 57
Bracewell, R., (5) 11
Bromba, M.U.A., (5) 27
Brookes, C.J., (1) 14
Brown, A.C., (2) 66; (5) 25, 29
Brown, S.Y., (1) 24
Bruderreck, H., (2) 14
Bruneder, H., (2) 28
Burke, M.F., (1) 3

Campbell, D., (5) 18
Canales, M., (4) 17
Carr, P.W., (2) 18, 19
Chesler, S.N., (1) 33; (2) 63; (5) 8, 25, 29
Clarke, A., (2) 62; (3) 17
Clayton, G.B., (5) 1
Condal-Bosch, L., (1) 37; (3) 8
Cooper, W.T., (2) 58
Cowper, C.J., (2) 15
Cram, S.P., (1) 33; (2) 66; (5) 8, 25, 29
Cramers, C.A., (5) 32
Crews, T., (1) 24, 25
Culp, R.A., (5) 20

Dauciunas, V., (4) 27
Davis, J.M., (2) 47
Dean, J.A., (5) 2
Deeming, S.N., (5) 24
DeGalan, L., (2) 37
Delaney, M.F., (2) 38; (3) 13; (4) 37
Deltour, J., (5) 26
Deng, P., (2) 17
DeRose, A.J., (2) 15
Dijkstra, G., (2) 16
Dilks, Jr., C.H., (1) 1; (3) 11
Doehl, J., (1) 39
Dolan, J.W., (4) 23
Dondi, F., (2) 55

Dorsey, J.G., (1) 7, 17; (2) 58; (4) 39
Duff, M., (4) 27
Dyson, N., (2) 3, 5; (4) 34, 40

Elling, J., (2) 40; (4) 22
Essers, R., (2) 57
Ewing, G.W., (5) 33

Feldman, G.L., (4) 2
Felice, L.J., (1) 19
Felinger, A., (2) 55
Foley, J.P., (1) 7–9, 17; (2) 36, 58; (3) 9; (4) 39
Fornstedt, T., (2) 4
Fowlis, I.A., (2) 20
Fozard, A., (5) 15
Frances, J.J., (5) 15
Frans, S.D., (1) 40
Frischenschlager, H., (5) 19
Fritz, G.T., (2) 65, 67
Fukuda, E.K., (1) 24, 25
Furman, W.B., (4) 13

Garcia-Alvarez-Coque, M.C., (2) 45
Garland, W.A., (1) 24, 25
Gebaur, P., (2) 54
Ghaoui, L., (5) 21
Gibb, T.C., (2) 49, 50; (5) 6

Giddings, J.C., (1) 27, 32; (2) 47
Gill, J.M., (5) 14
Goedert, M., (2) 60; (5) 9
Golay, M.J.E., (5) 23
Grant, D.W., (2) 62; (3) 17
Granzener, V., (2) 28
Grasserbauer, M., (5) 19
Greibokk, T., (1) 39
Grob, K., (1) 6
Grob, R.L., (3) 1; (4) 5
Grubner, O., (1) 29, 30
Grushka, E., (1) 20, 22, 32
Guillemin, C.L., (2) 11; (4) 8
Guiochon, G., (1) 4, 35; (2) 11, 25, 48, 60; (3) 10; (4) 8; (5) 9

Halasz, I., (2) 8, 14
Hapgood, H.W., (3) 5, 15
Harris, J.M., (1) 40
Harris, W.E., (3) 5, 14
Hartwell, S., (1) 26
Henry, R.A., (2) 9
Herrmann, R., (2) 35
Hettinger, J.D., (5) 14
Hippe, Z., (2) 52; (5) 5
Horlick, G., (5) 17
Hubbard, J.R., (4) 3; (5) 14
Huber, L., (4) 9, 19–21
Hunt, R.J., (2) 46; (4) 38

Jackson, K.D., (5) 18
Janik, J., (3) 7
Jeansonne, M.S., (1) 9
Johnson, E.L., (2) 59
Johnson, R.D., (4) 35
Jorgan, B.A.P., (4) 14
Jover, B., (2) 61
Juhasz, J., (2) 61
Juran, J.M., (4) 15
Juvet, Jr., R.S., (2) 51

Kaiser, H.H., (2) 23
Kantoci, D., (5) 16
Kateman, G., (2) 57
Kaufman, L., (5) 24
Keller, R.A., (2) 64
Kelly, P.C., (5) 17
Keppler, J.G., (2) 16
King, R.M., (2) 35
Kipiniak, W., (2) 10
Kirkland, J.J., (1) 1; (2) 6; (3) 11; (4) 6
Kishimoto, K., (5) 13
Kissinger, P.T., (1) 19
Klinkenberg, A., (1) 13
Kreyszig, E., (1) 12
Krieger, D.L., (2) 65

Lansdowne, R.A., (2) 35
Lawler, G., (4) 17
Layloff, T.P., (4) 13
LeBlanc, A.J., (4) 10
Leung, A.T., (4) 3
Littlewood, A.B., (1) 5, 16; (2) 49, 50; (3) 4; (5) 6
Lorenz, L.J., (5) 20
Loxton, S.M., (1) 14
Lundeen, J.T., (2) 51
Lysakowski, R., (4) 27, 28

MacDougall, D., (2) 22
Marmion, D.M., (2) 65, 67
Martin, M., (2) 48
Massart, D.L., (5) 24
Maude, M., (4) 2
Maynard, V., (1) 20
McCabe, T., (4) 31
McConell, M., (4) 17
McConnell, M.L., (1) 40
McCoy, R.W., (2) 65, 67
McDowell, L.M., (2) 19
McDowall, R.D., (4) 14
McGill, J.M., (3) 16
McNair, H.M., (3) 15
McQuarrie, D.A., (1) 28
Merritt, Jr., L.L., (5) 2
Meyer, V.R., (2) 7, 29–33; (4) 42–44
Michotte, Y., (5) 24
Miller, J.C., (2) 2
Miller, J.N., (2) 2
Miller, L.A., (4) 3; (5) 14
Miner, D.J., (1) 19
Mittermayr, C.R., (5) 19
Morosani, R.W., (2) 35
Musha, S., (5) 13
Myers, M.N., (1) 32

Naish, P.J., (1) 26
Nelson, D.C., (4) 27
Noble, D., (2) 40
Noij, Th., (5) 32
Novak, J., (2) 41; (4) 4; (5) 7

Papas, A.N., (2) 38–40; (4) 33, 37
Pauls, R.E., (1) 21, 23; (2) 65, 67
Pearson, K., (1) 2
Persinger, H.E., (4) 36
Petrovic, K., (2) 41; (5) 7
Petticlerc, T., (1) 35; (2) 25; (3) 10
Phillips, M.L., (2) 56
Pietrogrande, M.C., (2) 55
Pietryga, T., (2) 52; (5) 5
Pressman, R.S., (4) 30
Proksch, E., (2) 28

Qui, A.M., (1) 38; (3) 12

Ralik, M., (1) 30
Reddy, C.R., (1) 19
Reijnen, J., (2) 57
Rijks, J.A., (5) 32
Rodgers, L.B., (5) 20
Rogers, L.B., (1) 21, 23
Rosenberg, E., (5) 19
Rossi, D.T., (2) 26; (5) 10
Rothman, L.D., (5) 21

Said, A.S., (1) 36; (3) 18
Savitsky, A., (5) 23
Sawyer, D.T., (4) 1
Schettler, P.D., (1) 32
Schmus, H.G., (2) 35
Schneider, W., (2) 14
Schols, J.A., (2) 16
Schudel, J.V.H., (1) 4

Schupp, O.E., (3) 3
Schure, M.R., (2) 53
Scott, R.P.W., (2) 20
Settle, Jr., F.A., (5) 2
Shank, J.T., (4) 36
Shepherd, R.E., (4) 10
Shoup, R.E., (1) 19
Sjenitzer, F., (1) 13
Skoog, D.A., (5) 3
Smit, H.C., (5) 30, 31
Snyder, L.R., (2) 6, 34, 60; (4) 6
Spruce, B., (2) 12
Steiner, J., (5) 26
Sternberg, J.C., (1) 18; (2) 13
Stoklosa, H.J., (1) 1; (3) 11
Sulya, A.W., (2) 59
Synovec, R.E., (2) 59

Tao, F., (3) 16
Termonia, Y., (5) 26
Tetzlaff, R.F., (4) 10, 13, 18, 26, 29
Thompson, R.Q., (5) 22
Thorsteinsdottir, M., (2) 4
Torres-Lapasio, R., (2) 45
Tou, J.T., (5) 12
Tougas, T.P., (2) 39

Vandeginste, B.G.M., (2) 37, 57; (5) 24
Van den Heuval, E.J., (5) 30, 31
Vanden Heuval, W.J.A., (3) 6
Van Es, A.J., (5) 32
Van Malssen, K.F., (5) 30, 31
van Rijswick, M.H.J., (5) 28
Vassos, B.H., (5) 33
Villalanti, D.C., (1) 3

Walters, R.W., (1) 11
Walton, S.J., (5) 18
Wei, W., (2) 42
Westerberg, A.W., (2) 27; (5) 35
Westerlund, D., (2) 4
White, R.L., (2) 56
Wicar, S., (2) 41; (5) 7
Willard, H.H., (5) 2
Williams, A., (2) 1
Williams, R.R., (2) 24
Windeler, A., (4) 2
Wolf, T., (2) 67
Wright, N.A., (1) 3
Wu, N.S., (1) 38; (2) 42; (3) 12
Wyatt, A., (5) 15

Yau, W.W., (1) 1, 34; (3) 11
Yu, Z., (2) 4

Zacchei, A.G., (3) 6
Zeigler, H., (5) 27
Zhao, G.W., (1) 38; (3) 12
Ziegel, E.R., (2) 65, 67
Zikanova, A., (1) 30

Subject Index

Accountability, 11, 12, 129, 130
Accuracy, 35, 39, 64, 82, 137
Accurate representation of solute profile, 39–50, 89–92
Adsorption isotherm, 8
Amplifier saturation, 47, 60, 84
Analogue to digital conversion, 140–144, 149, 151
  auto-ranging of, 143
  dual slope integrating A/D conversion, 141–143
  resolution of A/D converters, 143
  voltage to frequency conversion, 140–141
Analysis parameters, 114–122
  analysis duration, 118
  baseline drift tolerance, 114, 117, 175–178, 180–182, 183, 190
  forced baseline, 41, 120, 179–180
  forcing perpendicular or tangent, 118
  horizontal baseline projection, 120–122
  integrate inhibit, 41, 118, 178–179
  inverting negative peaks, 122
  minimum area or height threshold, 118
  parameter override, 118
  peak width, 114, 115, 118, 148, 163
  slope sensitivity, detection threshold, 10, 114, 115, 118, 159, 162–163
  time to double, 117
Analysis report, 10, 12, 126, 127–128, 132–134
Analyst workload, 11
Aperture time jitter, 149
Area/Height ratio, 7, 10, 14
Area%, 124
ASTM E 685-79, 51
Asymmetry, *see also* Excess and Skew, 6, 78–81, 83
  and base broadening, 80
  and manual peak measurement, 80
  ratio, 8, 16, 22, 23, 31, 32, 69
  and sampling frequency, 146
  unequal, 81
Audits, 11, 12, 128, 130, 133, 134

Baselines, 175–189
  construction of, 166–168, 184, 187–190
  definitions of, 177–178
  drifting, 50, 51, 56–60, 84, 175–177, 183
  elastic band technique, 189
  'A Good Baseline', 55
  placement errors, 56, 58, 84, 179–185
  representative baseline, 162
  straight line convention, 167
  subtraction from chromatogram, 155
  zero slope, 109, 159, 175
Baseline simulation, 137
BASIC programming, 22
Bias, 35, 36
Boxcar averaging, 151
'Bugs' in software, 134

Calculations, Area%, Int. Std., Ext. Std., 109, 111, 124–126
Calibration, 70, 126–127
  achievement of accuracy in, 83
  errors of, 71
  response factors, 39, 63, 83, 126, 128
  time band and window, 123
  tolerance on, 123
Chromatogram reconstruction, 156–158
  reprocessing, 148, 152

storage, 144–146
subtraction, 155
Chromatographic Figures of Merit, 22
Chromatograph utilization, 11
Coefficient of variation, 11
Column hold-up time, 4
Column non-linearity, 49
Common mode rejection, 150
Computer simulation of peaks, 137
Condal-Bosch measurement, 29, 32, 98
Cost per analysis, 12
Coulometer, 138
Counting squares, 93
Curve fitting, 153
Cutting and weighing, 93

Data bunching, 144–151
of baseline, 145
Data sampling, 140–144
bunch size, 146–147
data processing algorithms, 144
frequency, 140, 144
and integrator manufacture, 144
mains coupling with, 150
non-linear, 145
Nyquist sampling frequency, 144
optimum, 144
and peak width parameter, 148
storage, 144–146
Detection threshold, *see* Slope sensitivity
Detectors,
accurate representation of solute profile, 39–50, 89–92
additivity of detector signals, 40
flow sensitive, 43–45
linearity, 48–49, 84, 134
mass sensitive, 43–45
noise, 50
overload, 47
reduced operating range, 39, 59
saturation, 47, 60, 84
'three-dimensional', vii, 77
Detector signal,
back-off control, 60
diagnostics, 4
drift, 50, 51
negative, 41
zeroing, 60
Digital measurement of peak areas, 158–175
Digital to analogue conversion, 136–138
chromatogram plotting, 157, 158
Disk integrator, 107
Distortion, *see* Peak shape distortion
Dual slope integrating A/D conversion, 141–143

Efficiency, 6
from Area/Height ratio, 10, 14
Elastic band technique, 189
Electromechanical counters, 107–110
disadvantages of, 107
EMG, *see* Exponentially modified Gaussian
Ensemble averaging, 151
Error function, erf, 21
complementary error function, erfc, 18
Errors,
accuracy, 35
bias, 35, 36
confidence limits, 36
precision, 36
random, 37
repeatability, 36
reproducibility, 37
spurious, 37
systematic, 37, 155
uncertainty, 35
Errors in baseline construction, 41, 56–58, 184–187
created by sloping baselines, 72
interference by negative dips, 41, 84, 122, 184–187
Errors created by noise, 53
loss of area, 78–81
peak splitting, 44, 53
Errors of incomplete peak resolution, 60, 62–70
correction strategies for, 67
elution order of a group, 63
perpendicular separation, 60, 62–70
propagation of errors through groups, 68
tangent skim separation, 71–73
Errors of peak asymmetry, 78–80
Errors of programming integrators, 162
Errors of tangent skim, 71, 72, 84
perpendicular/tangent transition, 72
position on solvent peak tail, 72
Event marks, 10

Event program, 114
Excess, *see* Statistical moments
Exponentially modified Gaussian (EMG) function, 20–23, 67
  Figures of Merit, 22
  formula for EMG peaks, 22
  EMG peak shape tests, 23
External standard, 125
Extra column broadening, 22

False peak starts, 177, 183
Fastest measurable peak, 165
Files, 114
Flow sensitive detectors, 43–45
Forcing baseline detection, 120, 179
Forcing tangent or perpendicular separation, 119
Fractional peak area, 17
Fractional peak height, 14
Fused groups of peaks, 170

Gaussian function, 13–20, 154
  area loss from base, 18
  boundaries, 17
  derivatives, 16
  function, 13
  key dimensions, 15
  parabolic maximum, 18
  peak height, 14
  peak shape tests, 15
  peak width, 15, 17
  points of inflection, 16
  ratio of Area/Height, 14
Glossary of Terms, xvi
Good Laboratory Practice, GLP, 129
Good Manufacturing Practice, GMP, 129

Higher statistical moments, 27
History of integrators, 107
Holistic validation of integrators, 129
Horizontal baseline projection, 120–122

Imprecision, causes of, 37
Inaccuracy, 35
  causes of, 37
Inflection points of Gaussian peak, 16
Installation qualification, IQ, 131
Instrument control, 123, 178
'Instrument' factor within response factor, 82

Integrate inhibit, 178
Integrators,
  comparisons of, 65, 86
  as a diagnostic tool, 129
  disk, 107
  electromechanical, 107
  electronic (TTL), 110
  microprocessor-based, 111–112
  operating range, 39, 163
  'standard specification', 113
  as system controller, 123, 178
Internal standard, 125
Inverting negative peaks, 122
Isomer test of deconvolution, 77

Keyboard indicators, 111
Kurtosis, *see* Statistical moments

Laboratory information systems, LIMs, 11, 131
Limit of detection, LOD, 55
Limits of integration, *see also* Peak start and Peak end, 163
Limit of quantitation, LOQ, 55
Linear regression, 127
Location of peaks, 158–166

Maintenance, 133–134
Manual peak area measurement, 28–33, 94–101
  advantages and disadvantages, 105
  boundary methods, 92–94
  Condal-Bosch area, 29, 32, 98
  counting squares, 93
  cutting and weighing, 93
  errors of manual measurement, 101–103
  Foley variations, 99
  height × width at half height, 29, 97
  measurement of asymmetric peaks, 32, 99
  measurement of overlapping peaks, 97, 101
  optimum shape, 103–104
  peak area by triangulation, 16, 30, 61, 96
  planimetry, 94
  strategies, 67, 92
  width at half height, 97–98, 99
Manual peak height measurement, 95

Mass sensitive detectors, 43–45
Matched filtering, 157
Mathematical deconvolution of peaks, 67, 76–77, 157
Mean retention time, 5, 28
Measurement diagnostics, 10, 170, 173
Method development, 7, 35, 114, 180
Microcomputers, 111
  impact of microcomputer, 112
Microprocessor-based integrators, 111
Minimum detectable quantity, MDQ, 48
Minimum peak Area/Height, 118
Mode retention time, 5
Modular validation of integrators, 129
Moments, *see* Statistical moments
Moving window smoothing, 152, 153
Multilinear calibration, 126, 127

Negative baseline dips, 41, 184–187
  phantom peaks caused by, 184
Negative peaks, 41, 84, 122
Noise,
  drift, 51
  errors created by, 53, 83
  filtering, 42–43, 91, 111, 137, 144, 150
  frequency, 51
  long-term noise, 51
  manual measurement, 91
  pseudo random, 137
  short-term noise, 51
  test, 162
  truly random, 137
Normalization, *see* Area%
Nyquist sampling frequency, 144

Operational qualification, OQ, 132
Optimum data sampling frequency, 144
Optimum filtering, 154–155
Optimum peak shape, 103
  peak height, 104
Overlapping peaks, 60–77, 169–171
  and calibration, 70
  measurement of areas, 97, 101, 169–171
  measurement of heights, 62
  on sloping baselines, 62, 74
Overload, *see* Detector overload

Performance qualification, PQ, 133
Parameter override, 118–122
Peak area, measurement accuracy, 85, 166–168, 183
  area and solute quantity, 43
  area total and injection volume, 11, 124, 125
  area *vs.* height, 3, 83–85, 95
  asymmetry, 7, 78–81
  base broadening, *see also* Skew and Excess, 80, 81
  loss from base, 18, 19, 53
  loss from flanks, 53, 78, 79
  loss from peak top, 47–49
Peak boundaries, 6, 23
Peak deconvolution, 67, 76–77, 157
Peak end, location of, 165–166
Peak generators, 131, 135
Peak 'gradient', 74
Peak height, 1, 14, 62
  loss due to non-linearity, 49
  measurement, 168
Peak identity, 123
Peak resolution, 9, 12, 60, 71
Peak shape distortion, 41–44, 47, 153–157
Peak shape tests, 15, 23, 101, 160
Peak splitting, 44, 53
Peak start, location of, 163–165
Peak 'volume', 3
Peak width, 6, 15
Peak width parameter, 148, 163
Performance measurement, 11
Perpendicular separation, 61, 62–64
Perpendicular/tangent transition errors, 72
Planimeter, 94
Points of inflection, Gaussian, 16
Polynomial curve fitting, 150
Precision, 35, 36
  of integrator measurements, 37, 117
Printer plotters, 111
Propagation of errors through groups, 68

Quadrants of a peak, 159
Quality, 129

Random errors, 37
Recorders, *see* Strip chart recorders
Records, 127, 132–134
Relative retention times, 124
Relative standard deviation, RSD, 11, 38
Repeatability, 36

Report distribution, 12, 128, 129
Reprocessing data, 148
Reproducibility, 36
Resolution, 9, 60
  of A/D converters, 143
Response factors, 39, 63, 126, 128
  empiricism of, 83
Results assessment, 10–12
Retention time, 4, 10, 58, 123, 160, 168
Retention time shift, 58–59
Rounding errors, 149

Specification Qualification, SQ, 131
Sample management, 123, 126
Sampling frequency of data, 140, 144
  bunch size, 146–147
  data processing algorithms, 144
  and integrator manufacture, 144
  and mains coupling, 150
  optimum, 144
Savitsky–Golay smoothing, 85, 153, 160
Scaling factor, 125
Shoulders, 171–172
Signal generators, 136
  traceable calibration, 138
Signal subtraction, 155–156
Signal-to-noise ratio (*S/N*), 12, 54–55, 85, 137, 146, 148, 150–156
Skew, *see* Statistical moments
Slope sensitivity, 162
Smallest measurable peak, 54
Small peak filtering, 150, 165
Smoothing of detector signal, 137, 150, 151–155
  boxcar averaging, 151, 152
  common mode rejection, 150
  data sampling as smoothing, 151–152
  matched filtering, 157
  polynomial smoothing, 153–155
  signal subtraction, 155–156
Solute identity, 4, 123
Solute profile, 39–50, 89–92
Spurious errors, 37
Standard chromatograms, 135–138
  creation of, 136, 137
  simulating baselines, 137
  simulating drift, 138
  simulating peaks, 137
  simulating random noise, 137
    pseudo random, 137
  specifications for, 137
  traceability, 138
Standard integrator specification, 113
Standard Operating Procedures, SOPs, 130, 133
Standards, comparison of results with, 11
Statistical moments (0th to 4th), 23–27
  excess, 26, 28, 56
  higher moments, 27
  kurtosis, 26
  measurement of peak moments, 27
  practical disadvantages, 27
  skew, 25, 28, 32, 56
  variance, 6, 32, 56, 58
Statistical Process Control, 128
Strip chart recorder, 43, 89–92, 111
  amplifier noise, 90
  attenuator accuracy, 92
  chart motor control, 92
  dead band, 91
  non-linear signal response, 89
  pen head damping, 90
  slow response time, 89
Surveys (round robins), 85
Synchronization,
  of instrument operations, 123
  of peak sampling and peak start, 149
System suitability tests, 4, 8
Systematic errors, 37, 155

Tailing factor, or tailing coefficient, 8
Tangent peaks, 172–175, 182
Tangent skim errors, 71, 72, 84
Tangent skim ratio, 172
Time band and window, 123
  tolerance on, 123
Time constants, 43, 44, 150
Time program, 114
Time to double, 117
Total peak area, 10
Traceability, 138
Transistor–Transistor Logic (TTL) integrators, 110
Trapezoidal baseline correction, 78, 111, 166–168
Trend analysis, 11, 128
Triangulation, 16, 30, 61, 62, 96
Truncation errors, 149

Uncertainty, 35

Unresolved peaks, *see* Overlapping peaks and Fused groups of peaks
US Pharmacopoeia, 8
Utilization of chromatographs, 11

Validation, 129–135
  and calibration, the difference, 132
  Installation Qualification, IQ, 131
  Operational Qualification, OQ, 132
  Performance Qualification, PQ, 133
  records of, 132, 134
  of software, 134
  Specification Qualification, SQ, 130
Valleys, 74, 169, 184
  shifted by sloping baseline, 74
  valley–valley skim, 190
Variance, *see also* Statistical moments, 6, 32, 56, 58
  from Area/Height ratio, 14
Voltage to frequency conversion, 140–141

Zero slope, *see* Baselines

This popular series is designed to provide a collection of thorough and informative bench-top guides for the practising chromatographer.

Each monograph in the series provides an up-to-date review of a specific topic in chromatography, including details on established and developing techniques. While including all the practical details that chromatographers need, these books also explain central facts and concepts in easily readable language which the non-specialist will be able to follow.

The second edition of the popular **Chromatographic Integration Methods** has been completely revised and updated. Written by an expert with many years' experience with two of the world's largest manufacturers of computing integrators, it has been expanded to include a new section on validation of integrators in response to regulatory requirements for quality and validation. A new literature survey, additional diagrams and Author Index have also been added.

Well illustrated and easily read, this is an excellent source book for those who wish to increase their understanding of integrators. **Chromatographic Integration Methods** describes and discusses both manual and electronic techniques used, with the aim of aiding analysts to obtain more data from their chromatograms, and assisting them with understanding how integrators work so that results are never accepted unquestioningly. As with the first edition, this book will be welcomed by all those in the chromatography field, particularly those at the bench.

Norman Dyson, after 7 years of study at Newcastle University, spent time in research and marketing with multinational companies before founding Dyson Instruments in 1974. As well as a wealth of experience with chromotographic techniques, Dr Dyson has lectured extensively to researchers and industrialists.

ISBN 0-85404-510-4

THE ROYAL SOCIETY OF CHEMISTRY
Information Services