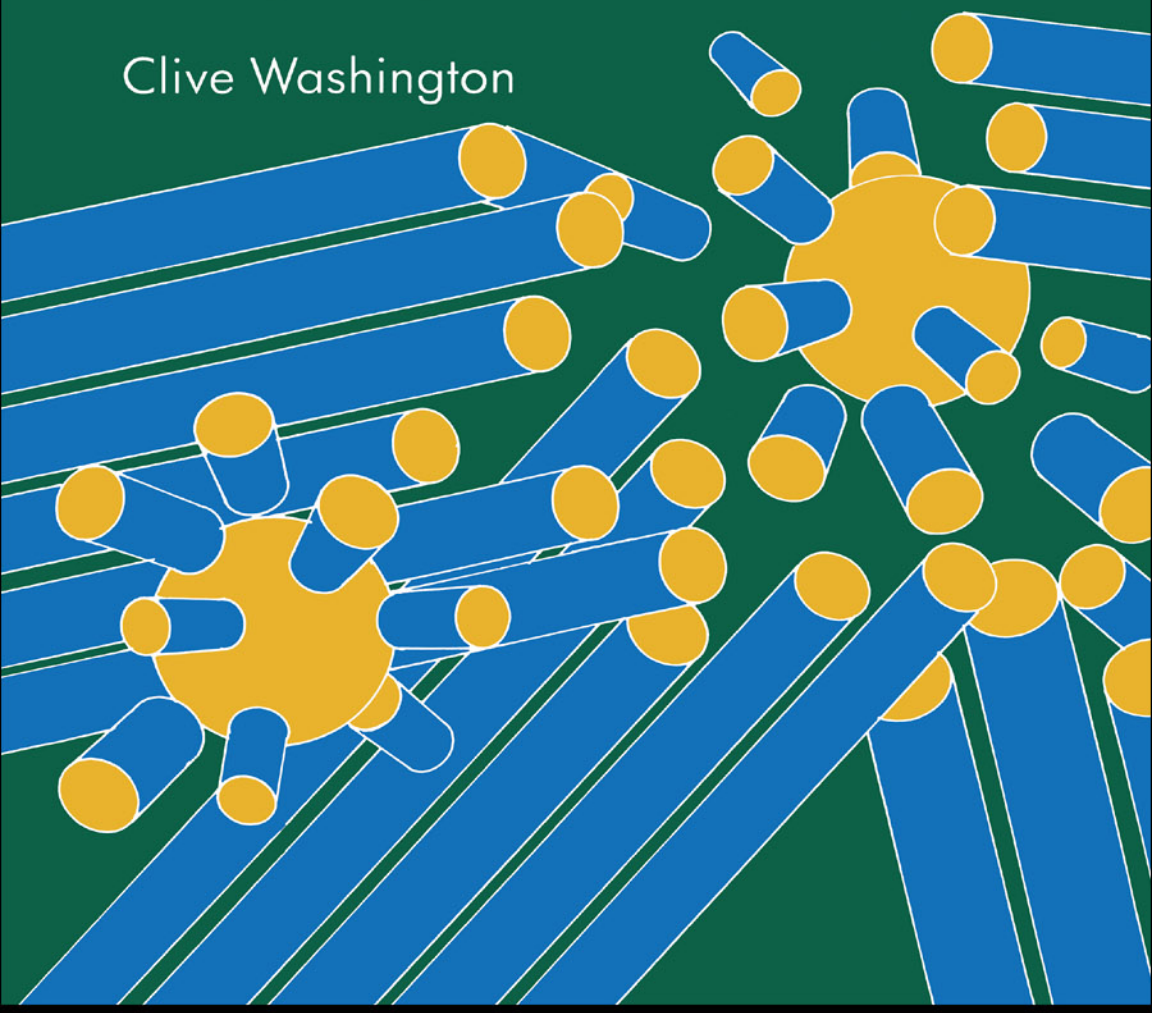
ELLIS HORWOOD SERIES IN
PHARMACEUTICAL TECHNOLOGY
incorporating Pharmacological Sciences

PARTICLE SIZE ANALYSIS IN PHARMACEUTICS AND OTHER INDUSTRIES

theory and practice

Clive Washington



PARTICLE SIZE ANALYSIS IN PHARMACEUTICS AND OTHER INDUSTRIES

Theory and Practice

ELLIS HORWOOD SERIES IN PHARMACEUTICAL TECHNOLOGY

incorporating Pharmacological Sciences

Series Editor: M H RUBINSTEIN Professor of Pharmaceutical Technology School of Health Sciences Liverpool Polytechnic

Armstrong & James **UNDERSTANDING EXPERIMENTAL DESIGN AND INTERPRETATION IN PHARMACEUTICS(Available in Cloth and Paperback)**

Bloomfield *et al* **MICROBIAL QUALITY ASSURANCE IN PHARMACEUTICALS, COSMETICS, TOILETRIES**

Broadley **AUTONOMIC PHARMACOLOGY**

Cartwright & Matthews **PHARMACEUTICAL PRODUCT LICENSING: Requirements for Europe**

Cartwright & Matthews **INTERNATIONAL PHARMACEUTICAL PRODUCT REGISTRATION: Quality, Safety, Efficacy**

Clark & Moos **DRUG DISCOVERY TECHNOLOGIES**

Cole **PHARMACEUTICAL PRODUCTION FACILITIES: Design and Application**

Cole, Hogan, Aulton **PHARMACEUTICAL TABLET COATING TECHNOLOGY**

Cook **POTASSIUM CHANNELS: Structure, Classification, Function and Therapeutic Potential**

Craig & Newton **DIELECTRIC ANALYSIS OF PHARMACEUTICAL SYSTEMS**

D'Arcy & McElnay **PHARMACY AND PHARMACOTHERAPY OF ASTHMA**

Denyer & Baird **GUIDE TO MICROBIOLOGICAL CONTROL IN PHARMACEUTICS**

Doods & Van Meel **RECEPTOR DATA FOR BIOLOGICAL EXPERIMENTS: A Guide to Drug Selectivity**

Evans *et al* **POTASSIUM CHANNEL MODULATOR DRUGS: From Synthesis to Clinical Experience**

Field & Goldthorpe **DRUG RESISTANCE IN VIRUSES: Principles, Mechanisms and Clinical Perspectives**

Ford & Timmins **PHARMACEUTICAL THERMAL ANALYSIS: Techniques and Applications**

Glasby **DICTIONARY OF ANTIBIOTIC-PRODUCING ORGANISMS**

Gould **PHYSICOCHEMICAL PROPERTIES OF DRUGS: A Handbook for Pharmaceutical Scientists**

Hardy *et al* **DRUG DELIVERY TO THE GASTROINTESTINAL TRACT**

Harvey **DRUGS FROM NATURAL PRODUCTS: Pharmaceuticals and Agrochemicals**

Hider & Barlow **POLYPEPTIDE AND PROTEIN DRUGS: Production, Characterisation, Formulation**

Izquierdo **NATURALLY OCCURING BENZODIAZEPINES: Structure, Distribution and Function**

Junginger **DRUG TARGETING AND DELIVERY: Concepts in Dosage Form Design**

Krogsgaard-Larsen & Hansen **EXCITATORY AMINO ACID RECEPTORS: Design of Agonists and Antagonists**

Kourounakis & Rekka **STEROIDS, DRUG RESPONSE AND METABOLISM: Pharmacochemical Approach to Defensive Steroids**

Labaune **HANDBOOK OF PHARMACOKINETICS: The Toxicity Asssessment of Chemicals**

Law **IMMUNOASSAY PROCEDURES: A Practical Guide**

Martinez **PEPTIDE HORMONES AS PROHORMONES**

Rainsford **ANTI-RHEUMATIC DRUGS: Actions and Side Effects**

Ramabhadran **PHARMACEUTICAL DESIGN AND DEVELOPMENT: A Molecular Biological Approach**

Ridgway Watt **TABLET MACHINE INSTRUMENTATION IN PHARMACEUTICS: Principles and Practice**

Roth *et al* **PHARMACEUTICAL CHEMISTRY Volume 1 Drug Synthesis**

Roth *et al* **PHARMACEUTICAL CHEMISTRY Volume 2 Drug Analysis**

Rubinstein **PHARMACEUTICAL TECHNOLOGY Controlled Drug Release Volume 1**

Rubinstein **PHARMACEUTICAL TECHNOLOGY Tableting Technology Volume 1**

Rubinstein **PHARMACEUTICAL TECHNOLOGY Drug Stability**

Russell & Chopra **UNDERSTANDING ANTIBACTERIAL ACTION AND RESISTANCE (Cloth & Paper)**

Rutherford **PHARMACEUTICAL SPECIFICATIONS: Standards for Drugs**

Taylor & Kennewell **MODERN MEDICINAL CHEMISTRY**

Theobald **RADIOPHARMACEUTICALS: Using Radioactive Compounds in Pharmaceutics and Medicine**

Thomas & Thursdon **CHEMISTRY FOR PHARMACY, PHARMACOLOGY AND THE HEALTH SCIENCES**

Tweed **CLINICAL TRIALS FOR THE PHARMACEUTICAL INDUSTRY**

Van Meel, Hauel, Shelley **CARDIOTONIC AGENTS FOR THE TREATMENT OF HEART FAILURE**

Vergnaud **CONTROLLED DRUG RELEASE OF ORAL DOSAGE FORMS**

Washington **PARTICLE SIZE ANALYSIS IN PHARMACEUTICS AND OTHER INDUSTRIES**

Washington *et al* **PHARMACOKINETIC MODELLING USING STELLA ON THE APPLE MACINTOSH (TM)**

Wells **PHARMACEUTICAL PREFORMULATION**

Wells & Rubinstein **PHARMACEUTICAL TECHNOLOGY Controlled Drug Release Volume 2**

Wells & Rubinstein **PHARMACEUTICAL TECHNOLOGY Tableting Technology Volume 2**

Wells & Rubinstein **PHARMACEUTICAL TECHNOLOGY Tableting Technology Volume 3**

Wilson & Washington **PHYSIOLOGICAL PHARMACEUTICS BIOLOGICAL BARRIERS TO DRUG ABSORPTION**

The above is a complete list of all Ellis Horwood titles in the pharmaceutical and pharmacological sciences, both published and in preparation. Further details can be obtained from Simon and Schuster International Group 0442–881900.

PARTICLE SIZE ANALYSIS IN PHARMACEUTICS AND OTHER INDUSTRIES

Theory and Practice

CLIVE WASHINGTON, Ph.D.
Department of Pharmaceutical Sciences,
School of Pharmacy
University of Nottingham

ELLIS HORWOOD

NEW YORK LONDON TORONTO SYDNEY
TOKYO SINGAPORE

First published in 1992 by
ELLIS HORWOOD LIMITED
Market Cross House, Cooper Street,
Chichester, West Sussex, PO19 1EB, England

This edition published in the Taylor & Francis e-Library, 2005.

A division of
Simon & Schuster International Group
A Paramount Communications Company

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the the British Library

ISBN 0-203-98417-X Master e-book ISBN

0-13-651613-0 (Print Edition)

Library of Congress Cataloging-in-Publication Data
Available from the publishers

To Neena,

For being more understanding while I wrote this book than I was while she wrote hers.

Table of contents

Preface

My intention in writing this book is to provide an introduction to the theory, and especially the practice, of particle size analysis. The idea originated from the frequent complaints of my undergraduate and postgraduate students of the absence of a suitable up to date text in this area; my series of lectures on this subject given at Nottingham formed a core for the book, which has of course been considerably expanded. I would hope that the lectures themselves have also benefited from the writing of the book! The emphasis is on the applications of particle size analysis in pharmaceutical sciences, although I hope that all users of particle characterization methods will find something of value.

In the past decade it has been notable that there has been a surge of interest in colloidal and microsphere formulations in the biomedical sciences, and this has been paralleled by a rapid development in techniques to study these systems, particularly using light scattering methods such as photon correlation and laser diffraction. I have tried to treat these areas in particular in detail, since there seems to be much confusion among their users; they are the methods most likely to be regarded as 'black boxes'. It has always struck me as ironic that the less someone understands about an instrument, the more likely they are to place blind faith in its results!

In order to keep this volume to a reasonable (affordable!) size, I have confined the discussion to methods which specifically measure particle size, and have not covered those which quantify surface area, porosity, or gas

adsorption. If my readers consider this a major omission, please let me know, so that they can be included in a later edition, if the publishers are still speaking to me.

A particular (!) problem I found in writing this book was the absence of application data in the literature. In a field like particle size analysis, a lot of useful material is placed in trade journals, which are rarely archived, rather than academic journals; much material also appears in obscure or unobtainable sources, patents, restricted company reports, etc. So of course I would be pleased to include in a later edition any further material that readers may send that has escaped my attention.

Nottingham 1991 *C. Washington*

1
Basic principles

The study of particle dimensions is of central importance in many areas of technology. Virtually all of the solid materials which are in common use are at some point in a powder or granular form. Examples can easily be found in pharmaceuticals (drugs and excipients), foods (grain, flour, sugar), materials technology (ceramics, abrasives) and building materials (sand and cement). The main reason for an interest in the dimensions of these particulates is that it influences their physical properties, and also the way that they respond to various production processes. Some of the better known influences of particle size on the behaviour of pharmaceuticals are:

1) Particle size influences dissolution. Small particles dissolve more rapidly than large ones, which is important not only in determining the behaviour of the drug *in vivo* (e.g. dissolution in the gastrointestinal fluids), but also in various manufacturing processes.

2) The flow properties of powders are strongly dependent on particle size and, in particular, particle shape. Since most powders are moved from one place to another (e.g. from storage silo to reaction vessel) by flowing, control of flow behaviour is highly important. Generally, coarse, roughly spherical particles flow much more easily than small or elongated particles.

3) The deposition of airborne particles on surfaces is dependent on size; this is of considerable significance in the study of aerosol formulations and their deposition in the various regions of the lung.

4) The stability of dispersions, such as suspensions and emulsions, depends on the size of the dispersed material. The forces between colloidal particles depend on their dimensions, and the settling rates of larger particles depend on their size and density.

No doubt many other examples will occur to the reader.

RANGE OF PARTICLE SIZES AND UNITS

We could describe a particle as a region of one phase bounded by another, a description which would fit both a liposome of dimension 10^{-8} m and a boulder of dimension 1 m. Pharmaceutical systems usually are confined to a narrower size range, and we will rarely be required to consider particles larger than a few millimetres (1 mm= 10^{-3} m) in diameter. Most process feedstocks lie in the range 1 mm to 10 micrometres (μm; 1 μm=10^{-6} m). We can usefully draw a boundary at 1 micrometre between these larger systems and smaller particulates, which are normally considered to lie in the *colloidal range.* Colloidal systems are of increasing importance in pharmacy; their normal size range is from 10 nanometres (nm; 1 nm=10^{-9} m) to 1 micrometre. Examples include the well known liposomes, parenteral fat emulsions, and some microsphere systems. One micrometre forms a useful boundary since the properties of colloidal materials are often very different to those of coarser systems, and the techniques used to study them are quite distinct from those used for larger particulates.

Some older size units which the reader may encounter are the micron (1 micron=1 micrometre), largely an engineering unit, and in the colloidal size range, the ångstrom (10 ångstroms=1 nanometre).

MEASURES OF PARTICLE SIZE

We have already been using a length to describe particle size, with the intention that it indicates the distance from one side of the particle to its opposite side. This description

is unambiguous in the case of a spherical particle, for example an emulsion droplet or a microsphere, but becomes vague if the particle is irregular, as are many materials. The problem is solved by quoting the particle size of a nonspherical particle as the diameter of a sphere which is in some way equivalent to the particle; such a sphere is termed an *equivalent sphere* and the diameter is an *equivalent diameter.*

A few examples will make this clear. Since (hypothetically) we can weigh a particle and can measure its density, we can find the particle volume. The *volume equivalent sphere* is the sphere which has the same volume as the irregular particle, and is characterized by the *volume equivalent diameter.* If we say that an irregular particle has a volume equivalent diameter of 1 micrometre, we mean that it has the same volume as a spherical particle of diameter 1 micrometre. Alternatively, if we know the surface area of the irregular particle, we can define the *surface equivalent diameter* as the diameter of a spherical particle having the same surface area as the irregular particle.

There are a large number of ways of defining equivalent diameters (Table 1.1); many of them have their origins in some practical particle sizing measurement. For example, the *Stokes' equivalent diameter* is the diameter of a sphere which has the same free-falling velocity as the particle of interest in a sedimentation experiment, and the *sieve equivalent diameter* is the diameter of a sphere which just passes through the same square-mesh sieve as the irregular particle.

In selecting a suitable equivalent diameter to describe a particulate system, we should bear in mind that all of the equivalent diameters available to us will generally be different for a given irregular particle. If we measure, for example, the volume, surface, and Stokes' equivalent diameters for a particle, they will all be numerically different (unless, of course, our particle really is a sphere!). Consequently, to obtain a good description of the particle system, we should select an equivalent diameter (and associated measurement technique) which is relevant to the

property of the particle that we are interested in. Thus, for example, if we wish to study the sedimentation properties of the material, we would select the Stokes' or free-falling diameter as a descriptor; if we were interested in the pigment covering power of the material, it would be sensible to measure the projected area diameter by microscopy. If we were mainly interested in the properties of the particles as an airborne aerosol and their deposition in the lungs, the aerodynamic diameter as measured by inertial impaction methods would be more relevant.

Table 1.1 Some equivalent diameters

Symbol	Name	Definition
d_v	Volume diameter	Diameter of a sphere with the same volume as the particle.
d_s	Surface diameter	Diameter of a sphere with the same surface area as the particle.
d_d	Drag diameter	Diameter of a sphere having the same viscous drag as a particle in a fluid at the same velocity (also termed the aerodynamic diameter in aerosol studies if the fluid is air).
d_{sk}	Stokes' diameter	Diameter of sphere of similar density having the same limiting velocity when falling under gravity in a viscous medium.
d_p	Projected area	Diameter of a circle having the same area as the projection of the particle.
d_F	Feret's diameter	The mean value between pairs of parallel tangents to the projected outline of the particle
d_M	Martin's diameter	The mean chord length of the projected outline of the particle.

DISTRIBUTIONS OF PARTICLE SIZES

It is unusual (and rarely useful) for us to characterize a particulate system simply by studying one particle selected from a batch. This would be experimentally difficult, and more importantly we would have no way of knowing if all of the particles in the batch were identical. If a batch contained a range of particle sizes, we would not only want to know about the characteristics of the 'average' particle, but have some idea of the variation between the particles. Consequently, virtually all particle measurement techniques examine a large number of particles and provide a description of the *distribution* of particle sizes.

The meaning of this is made most clear by a simple thought experiment. Suppose we have a hypothetical particle sizing method which examines a particle, reports its equivalent diameter, examines another particle, reports its diameter, and so on. The output of our instrument will be a sequence of numbers representing particle diameters. At first glance it would not be easy to draw any conclusions concerning the particles simply by examining this list of numbers.

The usual way of summarizing data of this type is by dividing it into size classes and drawing a histogram of the number of particles in each size class. A typical example is shown in Figure 1.1. The diameter (x) axis has been divided into classes, in this case 1 μm wide, and as each particle is measured, it is added into the appropriate size class. Thus, for example, if we found a particle of diameter 2.6 μm, we would add 1 to the 2–3 μm class. After a large number of particles had been measured, the histogram would reflect the distribution of particle sizes. A representation of this type is called a *frequency histogram.* Because we have only measured a relatively small number of particles, and divided them into relatively broad classes, the histogram displays very broad 'steps'. Obviously the true particle size distribution is relatively smooth (we have indicated this by the curve drawn through the histogram points in Figure 1.1), and if we needed a more detailed size

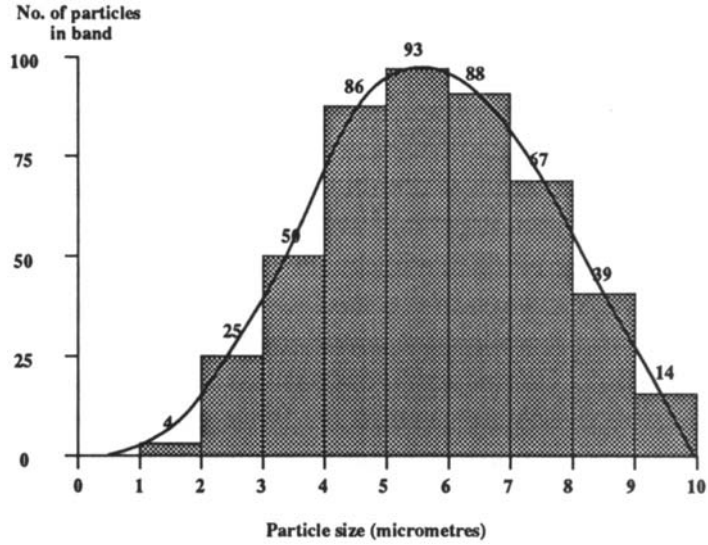


Figure 1.1 Histogram representation of particle size data.

distribution, we could use narrower classes, but would have to count more particles to fill them. If we were to measure the sizes of an arbitrarily large number of particles, we could divide them into extremely narrow size classes, and the 'jagged' shape of the histogram would approach the smooth curve which represents the size distribution in the whole sample. We rarely measure the distribution to such a high degree of accuracy, since the more particles are measured, the more time-consuming the measurement becomes.

Particle size data is rarely displayed in precisely the manner of Figure 1.1. The most important difference concerns the selection of the limits of the histogram classes (the terms *bins, bands* or *channels* are also used). In Figure 1.1, the histogram bins are each 1 μm wide, so that e.g. a 2.7 μm particle is counted into the 2–3 μm bin and a 5.8 μm particle is counted into the 5–6 μm bin. In practice the histogram channels are normally distributed not linearly but *geometrically,* so that each histogram bin is a constant factor larger than the previous one. This has two advantages:

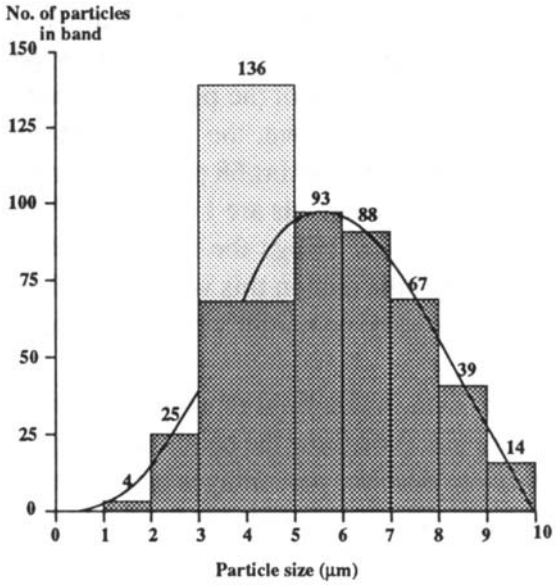


Figure 1.2. The variable bin-width problem.

1) The *resolution* of the bins is constant across the whole histogram. Resolution is defined as bin width divided by centre bin diameter. Suppose we have a linear set of bins of limit 1 μm, 2 μm, 3 μm, etc. The resolution of e.g. the 1–2 μm bin is (width/centre)= 1 μm/1.5 μm=0.67, while the resolution of the 15–16 μm bin is 1 μm/ 15.5 μm=0.065. Consequently there is less detail in the small size end of the size distribution than at the large size end. However, a geometric representation has the same resolution for each bin, so the detail is uniformly distributed over the representation.

2) The geometric representation provides the maximum amount of detail for a given number of bins. It is not desirable to try to improve the detail of the size analysis by having too many bins, since we need to measure correspondingly more particles to fill them.

Most commercial instruments present data using channels whose limits advance by the square root of two or the fourth root of two per channel. Channel limits on a root

two progression starting at 1 μm would be 1, 1.4, 2, 2.8, 4, 5.6, 8, 11.3, 16……μm.

However, now that we no longer have constant channel widths, a difficulty emerges. The shape of the distribution now becomes dependent on the particular channel sizes chosen. This is illustrated in Figure 1.2; suppose we take the two channels labelled 3–4 μm and 4–5 μm, and replace them with a single channel, 3–5 μm. The original channels contain 50 and 86 particles, so the new wider channel would contain 136 particles. If we plotted the 3–5 μm band at this height, it would rise above its neighbours, and we might think on inspecting the graph that there was an unusually large amount of material in this size band, which is untrue. The histogram would no longer follow the smooth curve of the whole distribution of particle sizes, and our representation would be misleading.

The way around this difficulty is to change the quantity plotted on the y-axis. Instead of plotting the amount, fraction, or percentage of material as the band height, we plot the percentage *per unit of band width.* Thus, in our example above, the 3–5 μm bin contains 136 particles in a 2 μm band, so it contains 68 particles per micrometre of bin width. (Since the other channels are 1 μm wide, they are already plotted correctly.) This now brings the new point back on to the smooth curve. If we plot the data in this way, we find that the shape of the distribution is the same no matter what band sizes are selected.

Typical particle size data in this final representation now resembles that shown in Figure 1.3. Note how the bands increase in width as the particle size increases. Also the percentages plotted on the y-axis are quite small; if the size distribution extends from 1 to 50 μm, there is obviously only about 2% of the particles on average in each micrometre. This form of representation is almost universally used in the representation of particle size data.

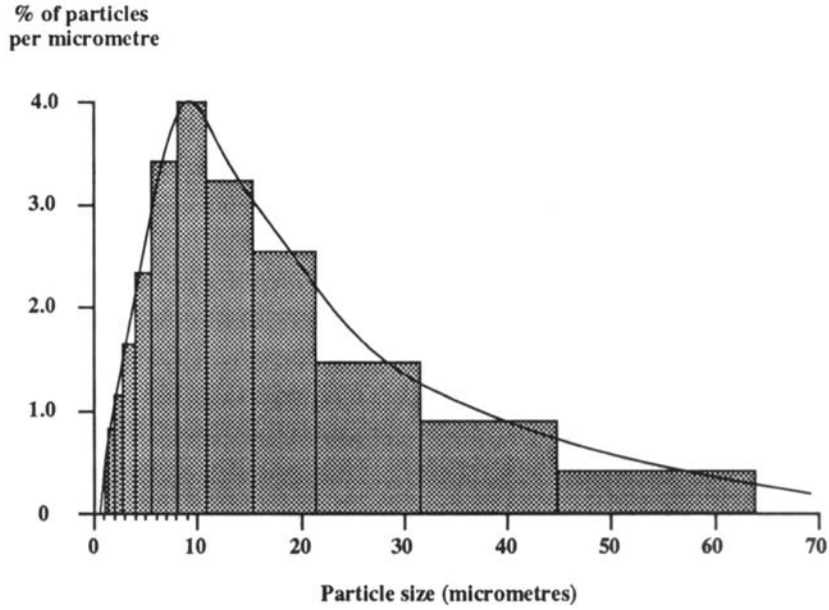


Figure 1.3. Geometric distribution of bin widths.

INCREMENTAL AND CUMULATIVE DISTRIBUTIONS

The histogram of Figure 1.3 is termed an *incremental* distribution because it shows how many particles fall within a given size increment. Although it is a convenient and intuitive representation, it is not the only representation possible. Another widely used way of depicting the data is in the form of a *cumulative* distribution, which shows how much material lies *above* or *below* a particular size.

To create a cumulative distribution from an incremental one, we proceed as follows. Firstly we must find the number of particles in each size band. This may be provided by the measuring instrument, or we may have to calculate it by multiplying the band height (number per micrometre) by the band width (micrometres). Table 1.2 demonstrates how a cumulative distribution can be found from the particle size data of the distribution in Figure 1.3. Column 1 contains the channel limits, column 2 contains their widths, i.e. the upper limit minus the lower limit, and column 3 the

Table 1.2. Calculation of cumulative distributions

Channel limits μm	Channel width μm	% per μm in channel	% in channel	cumulative % undersize
1 - 1.4	0.40	0.46	0.18	0.18
1.4 - 2	0.60	0.81	0.49	0.67
2 - 2.8	0.80	1.16	0.92	1.59
2.8 - 4	1.20	1.66	2.00	3.59
4 - 5.6	1.60	2.43	3.88	7.47
5.6 - 8	2.40	3.47	8.32	15.79
8 - 11	3.00	4.00	11.99	27.78
11 - 16	5.00	3.28	16.40	44.18
16 - 22	6.00	2.82	16.91	61.09
22 - 32	10.00	1.69	16.86	77.95
32 - 45	13.00	0.97	12.61	90.56
45 - 64	19.00	0.49	9.22	100.00

percentage per micrometre in the channel. In column 4 we have multiplied columns 2 and 3 together to obtain the total percentage in the particular channel. Finally, in column 5, we have constructed the cumulative distribution value at each channel by adding up all the material in the classes below that channel. For example, to find the amount of material smaller than 2.8 μm we would add the amounts in the 1–1.4, 1.4–2, and 2–2.8 μm channels. We can proceed in this way to find the fraction of particles below a particular size simply by adding up the fractions in all the bands smaller than that size. Ultimately all of the particles are below the largest size measured, so the last column of the cumulative distribution should add up to 100%.

In Figure 1.4, we have plotted the data of Figure 1.3 in this cumulative form. The graph shows, at any size, what fraction of the particles are smaller than that size, and so is termed a *percentage undersize* graph. It is possible to proceed in exactly the same manner but add up the total

fraction of particles in the bands *above* the size of interest to find the fraction *larger* than a given size. The resulting curve (Figure 1.5) is called a *percentage oversize* graph. It does not take a degree in mathematics to realize that these two curves are mirror images around a horizontal axis, since the total fractions above and below any given size must add up to 100%.

Cumulative and incremental distributions are both widely used, since various particle size analysis methods naturally lead to one or the other. For example, sieving (Chapter 3) sorts out the material which is sufficiently small to fall through one sieve, but too large to fall through a finer one, and so sorts the particles into increments. Plotting this data leads to an incremental distribution. Alternatively, a technique such as sedimentation (Chapter 9), where all the material larger than a certain size has sedimented at a particular time, naturally leads to a cumulative distribution. The terms differential and integral distributions are sometimes encountered as alternatives to incremental and cumulative distributions respectively.

NUMBER, AREA AND MASS DISTRIBUTIONS

The distributions we are constructing describe how a particular size-dependent property of the particles is distributed among the particles of various sizes. We are currently most familiar with distributions of particle *number* (the familiar frequency distribution), since our hypo thetical particle sizer has counted the particles of various sizes. However, we can also construct distributions which show how the surface area or mass are distributed among the various size particles. For example, suppose that after we had sorted the particles into a range of size channels, instead of counting them, we weighed them. This would tell us the *mass* of material in the size band, rather than the number of particles. We could then plot mass per micrometre, rather than number per micrometre, as the y-axis of our distribution, and the resulting graph would depict the way that the mass of material was distributed

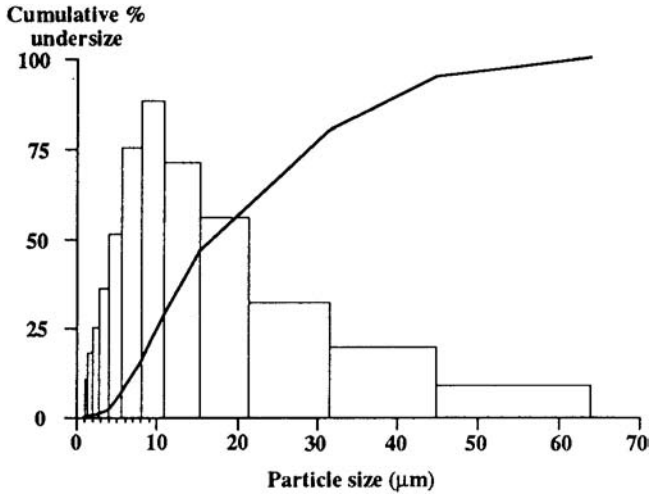


Figure 1.4. Cumulative undersize distribution.

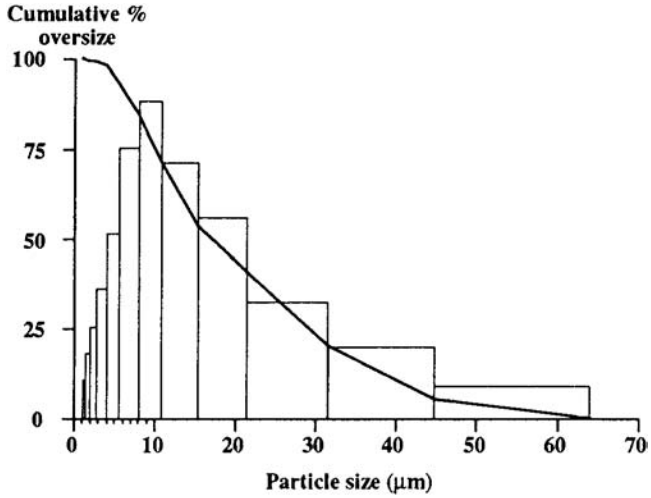


Figure 1.5. Cumulative oversize distribution.

among the various particle sizes. The same principle could also apply if we measured the surface area of the particles in each band. From this we could construct a distribution which showed how the particle surface area was distributed among the particle sizes.

It is important to realize that, for a particular sample of material, the curves describing the distribution of particle

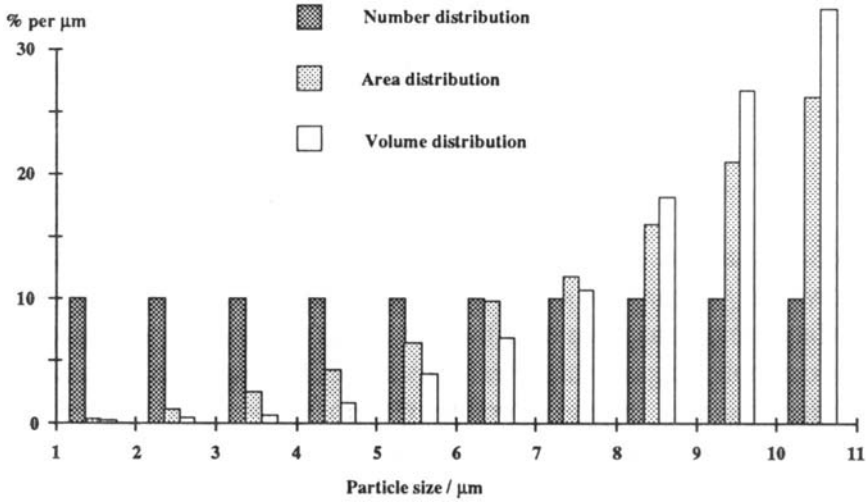


Figure 1.6. Number, area, and volume distributions.

number, area, and mass will not be of identical shape. A simple example will illustrate the differences in these distributions. Suppose we have 10 spherical particles with diameters of 1, 2, 3...10 µm, one of each size. This is of course a highly artificial distribution, but it serves to illustrate the concepts. The distribution of the number of particles (the *number distribution*) is shown in Figure 1.6, and as we would expect there is 10% of the total number of particles in each band, i.e. 1 particle per band. Now, suppose that instead of plotting the number of particles per channel, we plot the fraction of the surface area and of the mass of the material which is in each channel. The 2 µm particle has four times the surface area and weighs eight times as much as the 1 µm particle, and the 10 µm particle weighs 1000 times as much as the 1 µm particle and has 100 times the surface area. The *surface area distribution* and *mass distribution* are also plotted in Figure 1.6. (Mass distributions are also referred to as *volume distributions* since mass and volume are proportional, and so will be identically distributed.) It is evident that, while the number distribution is flat, the larger particles contribute

considerably more of the surface area and volume distributions, as we would expect.

It is evident from this example that we need to be careful to specify which type of distribution we are using. The particular type of distribution obtained depends on the sizing method used; for example, techniques such as sieving and sedimentation, in which material from a size band is weighed, provide the mass of material in a given size band, while techniques which count particles, such as the Coulter counter (Chapter 4) measure the number of particles in a given band. It is possible to convert from one distribution to the others if the shape of the particles is known; if the particles are spherical, this is straightforward, but it is difficult for irregular particles. Most modern computer-based instruments can convert and display data either as a number or mass distribution; the assumption made (that the particles are spheres) is usually hidden in a remote corner of the instruction manual!

CHARACTERIZATION OF DISTRIBUTIONS

A particle size histogram such as that of Figure 1.3 contains a considerable amount of data. In order to assist rapid comparison and examination of distributions, it is usual to extract simpler quantities from the distributions; these are usually single figures which say something meaningful about the data. Most of the following concepts will be familiar to anyone with a background in statistics.

Measures of central tendency

By 'central tendency' we mean the tendency of the particle size to cluster around a particular value. Such values are normally evident as a peak in the particle size distribution. These values are normally known as 'averages' or 'means' of a set of data. Three different quantities are in common use:

The *median* value is the size which splits the distribution into two halves, with 50% of the mass or particle number larger, and 50% of the mass or particle number smaller. For

this reason it is usually given the symbol D_{50}. In general the size with 50% of the mass on either side (the *mass median diameter*) will not be the same as the size with 50% of the number of particles on either side (the *number median diameter*). We can thus see that in constructing measures of central tendency we must be particularly careful to state whether we are using number, area or volume distributions. The easiest way to find the median value of a distribution is to construct a cumulative graph, from which the 50% point can be read off directly.

The mode of the distribution is the most common value occurring in the distribution. It is thus the value of the 'peak' of the distribution, if one exists. If the distribution has two or more peaks, it is said to be *bimodal* or *multimodal.*

The *arithmetic* mean, often just termed the *mean,* is the weighted average particle size. Suppose we have *i* channels of data, with each channel containing n_i particles, where x_i is the centre band diameter of channel *i.* The mean diameter of the number distribution is called the *number mean diameter,* NMD, and is given by:

$$\mathrm{NMD} = \frac{\sum_i x_i n_i}{\sum_i n_i} \tag{1}$$

This form is particularly convenient for particle size data, where the particles are usually divided into classes of particular sizes.

It will be realized that, since we highlighted the difference between number and mass distributions, the mean diameter based on particle number will not be the same as that based on particle mass or volume. If each of the above channels contained n_i particles, the *volume mean diameter* or VMD is given by:

$$\mathrm{VMD} = \frac{\sum_i x_i^3 n_i}{\sum_i n_i} \tag{2}$$

If the particle density is independent of size (as it often is), the terms *mass mean diameter* and *volume mean diameter* are usually used interchangeably.

Similarly the mean of an area distribution is called the *surface mean diameter* or SMD, and is given by:

$$\text{SMD} = \frac{\sum_i x_i^2 n_i}{\sum_i n_i} \tag{3}$$

Because of the cubic dependence of mass on diameter, large particles in distributions have a great effect on the volume mean diameter, but a lesser effect on the number mean diameter. If we were concerned as to the number of large particles in a sample, the volume mean diameter would provide a more useful representation of the sample than the number mean diameter. There is however a problem to watch out for here. If we are generating particle size data from an instrument which counts particles rather than weighs them, then an error of 1 count at the large end of the size axis will make a major difference to the volume mean diameter. Consequently the VMD will emphazise the noise in the data, so it is important to ensure that the data are of high quality.

These measures of central tendency are shown for a typical peaked distribution in Figure 1.7. They will all provide the same answer only if the distribution is symmetric about the mode, while if the distribution is asymmetric, the mean and median will lie on the long-tailed side of the mode.

Moments of a distribution and moment notation

A number of other measures of central tendency may be encountered, although those described above are the most useful. These are defined in terms of the so-called moment notation. If we examine equations 1–3, we can see that they all contain terms of the form:

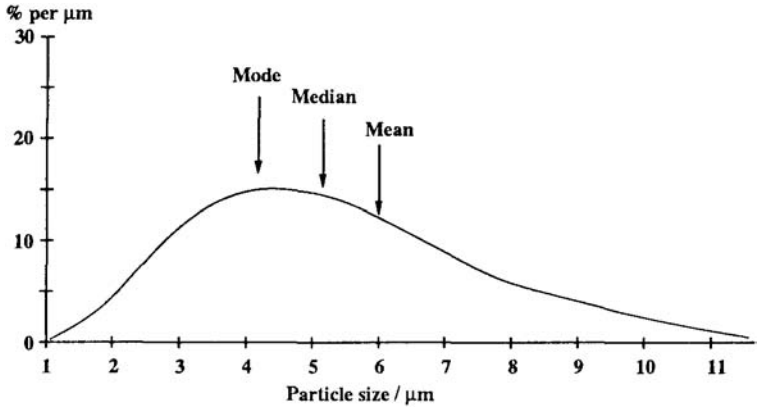


Figure 1.7. Measures of central tendency.

$$\sum_i x_i^k \, n_i$$

where k is 0, 1, 2, 3, etc. These sums have a special significance in statistics, and are called *moments* of the distribution. Again working with the hypothetical distribution of i channels of data, with n_i particles in the band of size x_i, the *k'th moment of the distribution* is defined as:

$$M_k = \sum_i x_i^k \, n_i \tag{4}$$

We can define a mean in a very general sense as a ratio of moments; thus the number mean diameter, from (1) is given by:

$$NMD = M_1/M_0 \tag{5}$$

and the volume mean diameter is given by:

$$VMD = M_3/M_0 \tag{6}$$

We can also define other means as moment ratios; for example the surface volume mean diameter, SVD, is given by:

$$SVD = M_3/M_2 \tag{7}$$

These higher order means are not often encountered, but since they are easy to calculate, they are often part of the

computer output from modern particle sizing instruments. This moment representation of means has led to the introduction of a systematic notation for these quantities. The mean corresponding to the ratio of the moments i and j is given the notation D_{ij}. Thus the number mean diameter is often termed D_{01} and the volume mean diameter is termed D_{03}.

Measures of dispersion

In a real sample of material there will exist a range of particle sizes spread about the mean. This spread may be narrow, in which case the particles are all of similar sizes, or it may be broad, if the particle sizes vary greatly. If the particles in a sample are all of the same size it is said to be *monodisperse,* while if a range of particle sizes exists, it is said to be *polydisperse.* It is often useful to be able to specify numbers which describe how much spread of particle sizes exists in a sample. The simplest way of doing this is to quote the sizes corresponding to some points on the ends of the distribution. We may quote the *quartiles,* i.e. the 25% and 75% points on the distribution; the *interquartile range* is the difference between these values. Another popular measure is to quote the 10% and 90% sizes. These quantities are illustrated in Figure 1.8.

A very widely used measure of dispersion is the *standard deviation,* σ. If we have a collection of *i* particles, where the size of particle *i* is x_i, and the mean size of the set is X, this is defined as:

$$\sigma = \sqrt{\frac{\sum_i (x_i - X)^2 \cdot x_i}{\sum_i x_i}} \qquad (8)$$

We may also encounter the *variation,* which is given by σ^2, and the *span,* which is usually defined as the 10% to 90% range divided by the mean.

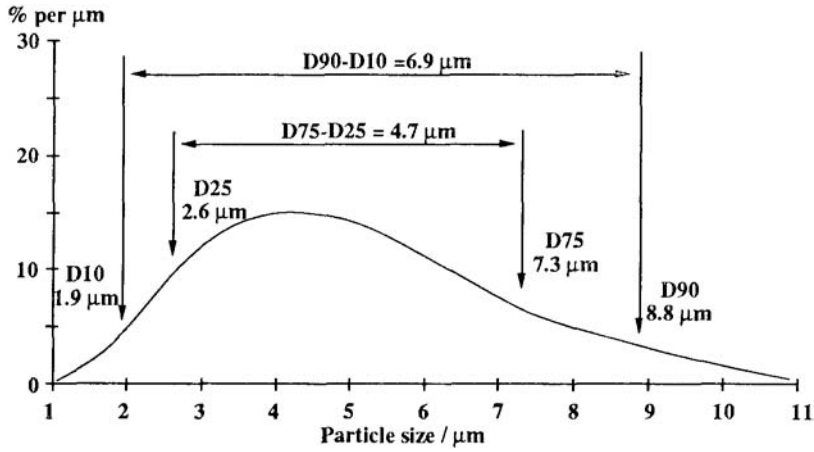


Figure 1.8. Measures of dispersion.

Measures of distribution shape

Distributions may be *symmetric* or *asymmetric*. A symmetric distribution has a vertical axis of symmetry passing through its mode, and as we noted above, in such distributions the mean, median, and mode all lie at the same point. Very few real distributions are symmetric, most having more material on one side of the mode than the other. Such distributions are said to be *skewed;* if the 'tail' of the distribution is on the large side of the mode, the distribution is said to be positively skewed or skewed to the right. If the 'tail' is on the small side of the mode, the distribution is said to be negatively skewed or skewed to the left.

Distributions may also have a rounded or pointed shape; this is quantified as the *kurtosis* of the distribution. A distribution which is pointed is termed leptokurtic; one which is flattened is termed *platykurtic.* These measures of distribution shape are shown in Figure 1.9.

MODEL DISTRIBUTIONS

Particle size distributions may take many forms, but there are a small number of model distributions which are of particular interest. These arise from the theory of statistics,

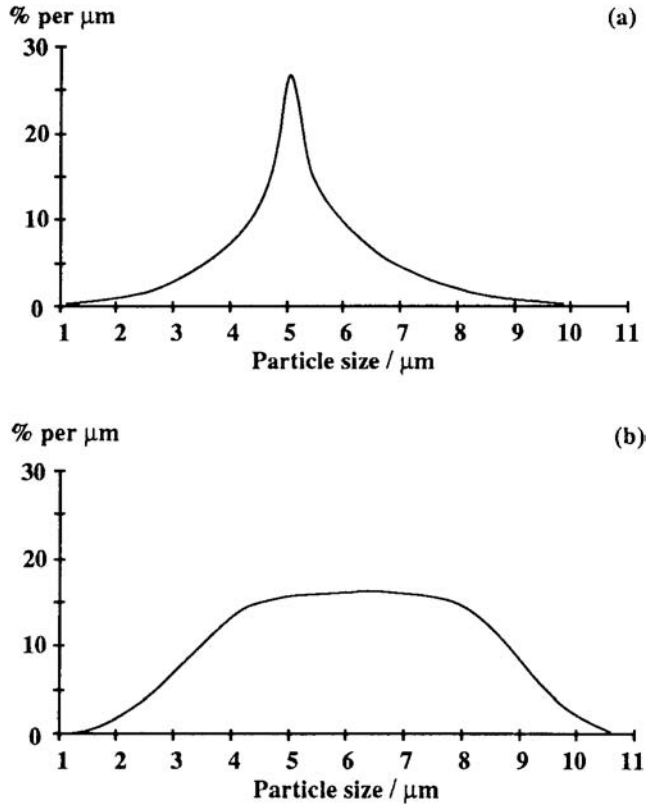


Figure 1.9. Kurtosis. (a) leptokurtic distribution, (b) platykurtic distribution.

or from studies of the behaviour of materials, and it is of value for us to study these distributions in order to find out if our experimental data conforms to them. This in turn may allow us to infer something about the material or the processes through which it has passed.

There are two theoretical distribution shapes which arise from statistical arguments which are of widespread occurrence in nature. These are the *normal* and *lognormal* distributions. Finally we will examine the *Rosin-Rammler distribution,* which has its origin in a consideration of the crushing of materials.

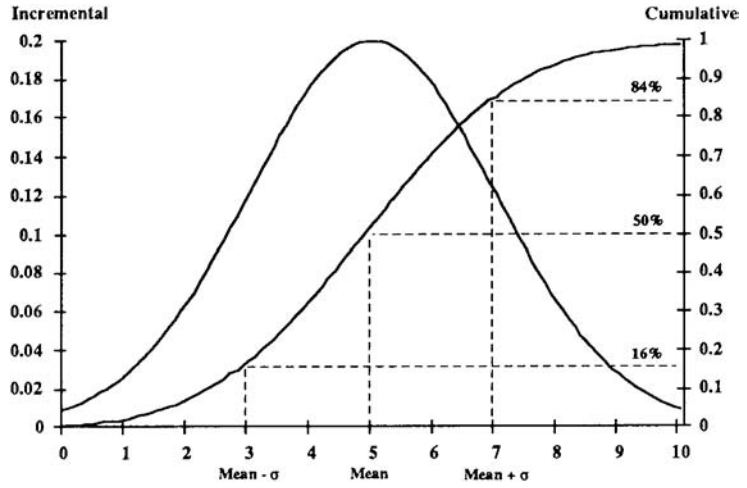


Figure 1.10. A normal distribution with X=5 and s=2.

The normal distribution

This distribution is also variously known as the *Gaussian* distribution or bell-shaped curve, and is given by the equation:

$$n(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-X)^2}{2\sigma^2}\right] \qquad (9)$$

where n(x) is the number of particles per unit size of size x, X is the mean particle size, and σ is the standard deviation of the distribution; the standard deviation is now a fundamental parameter of the distribution, and describes its width. The standard deviation can be read off approximately as the distance between the 16% and 50%, or the 50% and 84% points on the distribution. This is most easily accomplished using the cumulative form of the graph; the cumulative normal distribution is sometimes referred to as an *ogive curve.* A graph of this function, in its incremental and cumulative forms, is shown in Figure 1.10. It is symmetrical, so the mode, median and mean are all at the same point, in this case 5 μm. The standard deviation is the difference between the 16% and 50%, or the 50% and 84% points, i.e. 5 μm− 3 μm=7 μm−5 μm=2 μm. Eq. (9) shows the

function in its normalized form, i.e. the area under the curve is unity.

Normal distributions are of interest largely because they arise when an item is subject to random variation. For this reason they are commonly found in many natural systems; for example the distribution of heights of people is normal, as are many more specific biological size measurements, such as the sizes of individual organs or bones. The distribution of tablet weights in a batch is normal, as is the error in filling a liquid into a container. Obviously we strive to make both of these distributions as narrow as possible!

The lognormal distribution

This function is given by the equation:

$$n(x) = \frac{1}{\ln\sigma.\sqrt{2\pi}} \exp\left[-\frac{(\ln x - \ln X)^2}{2(\ln\sigma)^2}\right] \qquad (10)$$

where X, n(x), x, and σ have the same meanings as for the normal distribution. This function is shown in Figure 1.11 in incremental and cumulative forms. It is a positively skewed distribution ; the reason for its name becomes obvious if we plot it on a logarithmic x-axis, for it then forms a normal curve (Figure 1.12). It is of particular interest in particle technology because many particulate systems show a lognormal size distribution, particularly if they have been created by shear processes. Notable examples are emulsions made by valve homogenizer, or powders reduced by grinding. Particles which have been grown by crystallization often also show a lognormal distribution of size.

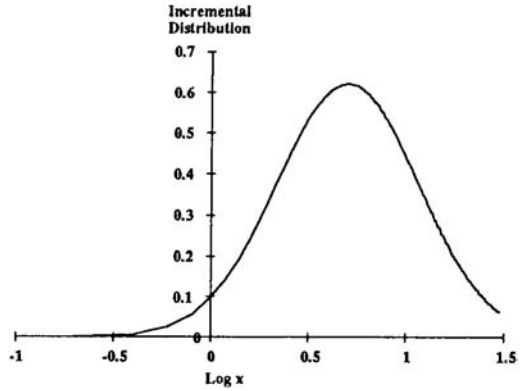


Figure 1.12. The lognormal distribution of Figure 1.11 plotted on a logarithmic x axis.

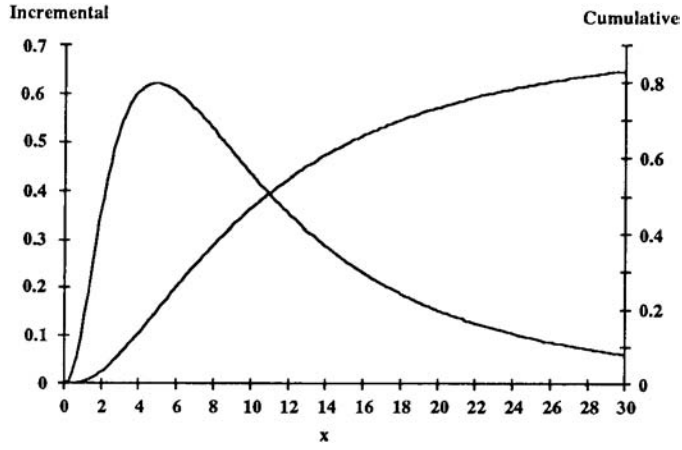


Figure 1.11. A lognormal distribution with mode=5 and σ=2.

The Rosin-Rammler distribution

This distribution arises from a consideration of the forces involved in the crushing of materials, and has been found to apply quite widely to materials whose size has been reduced by impaction. It is normally expressed in its cumulative form:

$$n(0,x) = \exp\left[- bx^m\right] \tag{11}$$

Its incremental form is:

$$n(x) = mbx^{m-1} \exp\left[- bx^m\right] \tag{12}$$

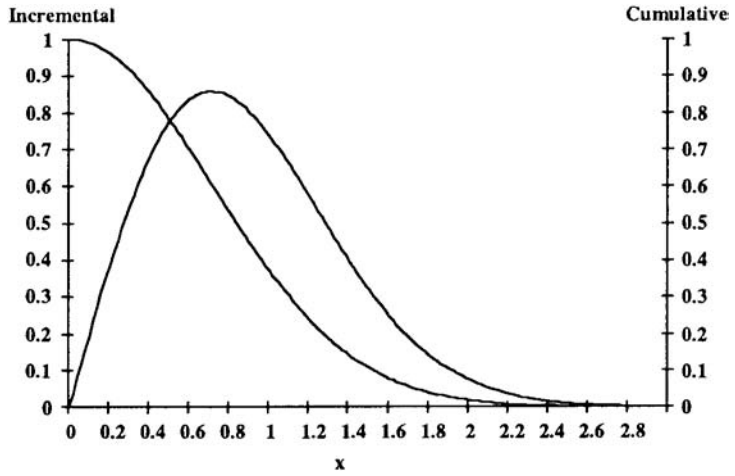


Figure 1.13. A Rosin-Rammler distribution for b=1 and m=2.

where b is a measure of the range of particle sizes present, and m is a constant (not necessarily an integer) which is approximately constant for a given material. The incremental and cumulative forms of the curve are shown in Figure 1.13 for b=1 and m=2.

Several other distribution shapes have been used but are not widespread. The Gates-Gaudin-Schumann (Gates 1915) and Gaudin-Meloy (Gaudin and Meloy 1962) are typical power-law models. There is also often a need to fit bimodal distributions, particularly in the study of microparticles and microcapsules, where the desired particles may be accompanied by unencapsulated or damaged material. In these cases it is possible to fit the distribution to sums of normal or lognormal distributions (e.g. Morris and Warburton 1984). Commercial software is available to perform the fitting operation. Perhaps a word of warning should be given; increasing the complexity of a fitting function or increasing the number of fitting parameters will always lead to an improved fit. This does not mean that the more complex function has any greater physical significance than the simpler alternatives.

IDENTIFICATION OF SPECIFIC DISTRIBUTIONS

It is desirable to be able to test an experimentally measured particle size distribution to see if it conforms to one of the distribution shapes discussed above. Fortunately the normal and lognormal distributions are so important to statisticians that they have long known how to do this. The usual approach to see if a set of data is fit by a function is to find some way of transforming the function so that it, and the data, can be plotted as straight lines on suitable axes. It is rather difficult to find a transformation which converts the incremental form of these distributions to straight lines, but it is much easier to work with the cumulative forms. If we examine the cumulative form of the normal distribution shown in Figure 1.10, we can imagine that the sigmoid form of the curve could be made into a straight line if the ends of the distribution were stretched along the y-axis. All that is needed to make this distribution into a straight line is a piece of graph paper with a suitably distorted y axis.

Because of the importance of this transformation, graph paper with this type of scale is available commercially, and is called *probability paper,* or *probit paper* for short, and the axis is called a *probability scale.* Its exact mathematical form can be derived from the equation for the normal distribution. In Figure 1.14 we can see the effect of plotting the cumulative normal distribution on a graph with this type of y axis. Note how the y axis is symmetrical about the 50% point (as is the cumulative distribution), and how the axis scale gets further apart towards the ends. The ends of the y axis never reach 0% and 100%, since the normal distribution only reaches these values at x= ±infinity. A corollary to this is that real particle size distributions cannot conform to a true normal distribution, since they would need to contain particles of negative size! This problem does not arise for the lognormal distribution, which is not defined for negative sizes.

We can use a similar device to test for a lognormal distribution if we recall that this function forms a normal curve when plotted on a logarithmic x-axis. If we plot a

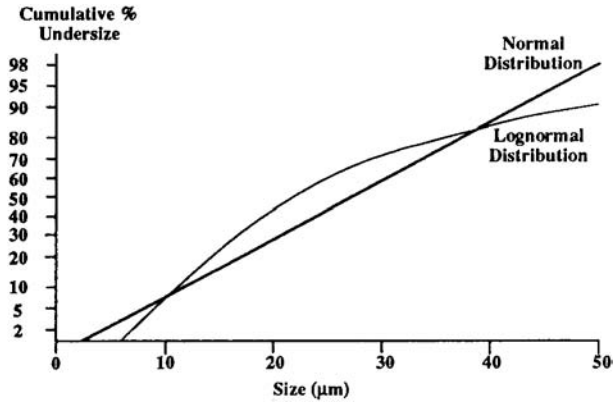


Figure 1.14. Normal and lognormal distributions on linear probability paper.

cumulative lognormal distribution on paper with a probability y-axis and a logarithmic x-axis, again a straight line will result (Figure 1.15). This paper is called *log-probability paper* and is also available commercially. In Figures 14 and 15 we have plotted both normal and lognormal distributions on each type of graph to demonstrate the power of the method to discriminate between these distributions. A normal distribution forms a straight line on the linear-probability graph (Figure 1.14), while a lognormal distribution forms a curve. Conversely on the log-probability graph (Figure 1.15), a lognormal distribution gives a straight line, while a normal distribution forms a curve. In general, any distribution other than one being specifically tested will not provide a straight line, and so we can test for normality or lognormality simply by testing the goodness of fit to the straight line on the appropriate graph. Note that we *cannot* argue that any distribution forming a curve on, for example, linear probability paper is lognormal, since there are many other distributions which will have this property. Only the distribution which forms the straight line can be identified by these graphical devices.

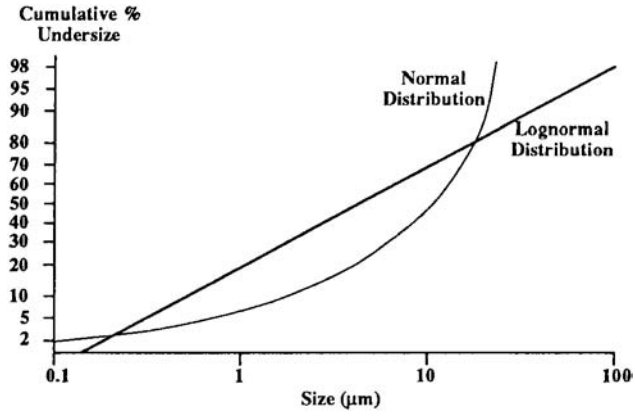


Figure 1.15. Normal and lognormal distributions on log-probability paper.

PARTICLE SHAPE

Shape is very important in determining particle properties such as adsorbing power, pigment capacity, and flow properties; however, it is probably fair to say that no completely satisfactory method of shape specification and measurement exists. Normally we wish to find some simple measure of particle shape which correlates with a property of interest in our particle system, and having established the correlation, use this for predictive purposes. Describing the shape of each particle completely would require a considerable amount of data for each particle (which would be different for each particle), and it would be difficult, if not impossible, to establish any useful correlation between the properties of the material and this vast collection of numbers. Hence we rarely specify the complete shape of each particle, but only derive one or two parameters which describe the general aspects of the shape.

A very wide number of measures of particle shape have been proposed in the past, and we will examine only a few here.

Descriptive

General aspects of particle shape can be expressed using qualitative terms; to avoid confusion, some of these have

Table 1.3. Description of particle shape

Acicular	Needle-shaped
Angular	Sharp edged; having a roughly polyhedral shape
Dendritic	Having a branched crystalline shape
Fibrous	Thread-like
Flaky	Plate-like
Granular	Irregular but of approximately spherical overall form
Irregular	Lacking any symmetry
Modular	Having a rounded, irregular shape

been standardized in BS 2955. A selection of the more commonly used ones are given in Table 1.3. These terms are widely used, but convey little quantitative information.

Descriptors based on direct measurement of dimensions

Shape data is often extracted from microscopic measurements. One of the simplest ways of defining particle shape is in terms of three mutually perpendicular measurements of the particle size, the thickness, breadth, and length (Heywood's ratios; Heywood 1963). These are defined as follows:

The *thickness* is the height of the particle when it is resting in its position of maximum stability.

The *breadth* is the minimum distance between two tangential planes which are perpendicular to the plane of maximum stability.

The *length* is the distance between two tangential planes which are perpendicular to those defining the thickness and breadth.

These definitions are highly useful for microscopy, since the particle is normally resting on its plane of maximum stability, and is viewed perpendicular to it. We can define the *flakiness* and *elongation ratios* as:

$$\text{Elongation ratio} = \text{length/breadth}$$

$$\text{Flakiness ratio} = \text{breadth/thickness}$$

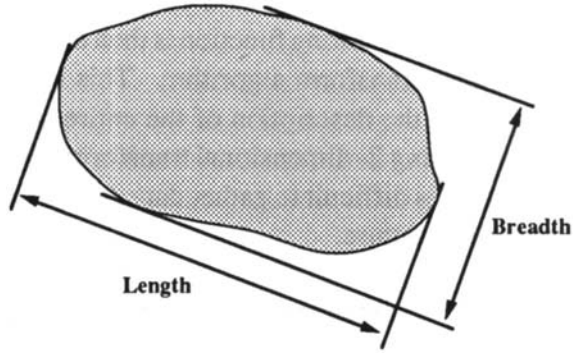


Figure 1.16. Particle shape and elongation.

In practice it may be difficult to measure the particle thickness, particularly from microscopy, and so often we are limited to using simply the elongation ratio, as shown in Figure 1.16. Some confusion can arise since this term has also been used by Hausner (1966), who defined it as the length/width ratio of an enclosing rectangle of minimum area. For certain shapes of particle, these may not be the same. Hausner also defined the bulkiness factor:

$$\text{Bulkiness factor} = \text{projected area}/(\text{length} \times \text{breadth})$$

This ratio thus describes the ratio of the particle's projected area to that of the enclosing rectangle.

Shape factors

As we noted in our discussion of equivalent diameters, measurements of different equivalent diameters for a particular particle will only yield the same result if the particle is a sphere. If the particles are nonspherical, and we measure their diameters using two different techniques, we will in general obtain different sizes. The ratio of two equivalent diameters obtained by different methods is termed a *shape factor.*

Shape factors describe the departure of the particle from a spherical form. One of the simplest is the *sphericity,* Ψ, defined by Wadell (1934) as:

$$\psi = \frac{\text{surface area of a sphere having the same volume as the particle}}{\text{surface area of the particle}} \quad (13)$$

$$= \left(\frac{d_v}{d_s}\right)^2 \quad (14)$$

Such factors are extremely useful if it is necessary to compare results obtained with different instruments. For example, suppose a powder is studied by sedimentation and yields a distribution with a mean diameter of 5.8 μm, and then measured by Coulter counter, which provides a mean diameter of 6.4 μm. The shape factor for these measurements is 5.8/6.4=0.906, suggesting that the particles do not depart too far from spheres. As long as the particle shape does not vary with particle size, we can obtain an approximate measure of the Coulter distribution by dividing all the sizes in the sedimentation distribution by this shape factor. Shape factors are thus commonly abused as means of forcing data from different instruments into agreement.

Fourier transform methods

It is possible to describe the perimeter of a 2-dimensional curve using Fourier transform techniques (Naylor and Wright 1977, Flook 1981). The curve is first transformed by measuring the perimeter distance from the particle centre of gravity as a function of angle, i.e. by 'unrolling' the curve. The resulting function is then transformed using a conventional Fourier transform algorithm. This approach can in principle be applied to the description of the entire surface of a 3-dimensional particle using 2-dimensional transforms, but in practice this is not done since it is difficult to gather data describing the full 3-dimensional particle structure.

The value of the Fourier transform approach is that firstly it may be possible to describe the profile of the particle using fewer parameters (Fourier component amplitudes and phases) than are needed to describe the curve as a series of x-y or r-θ points. Secondly, the Fourier transform identifies spatial frequencies in the profile, and it thus provides a

deeper understanding of the particle structure than can be obtained from inspection of the curve alone.

FRACTAL CHARACTERIZATION OF PARTICLES

A curve or surface is said to be fractal when its measured size depends on the particular scale at which it is examined. A fractal surface displays more detail as it is examined at increasingly smaller scales. The classic example of a fractal curve is the coastline of an island; if this is estimated from a large scale map, a particular result will be obtained. If a smaller scale map is used, the details of smaller streams and estuaries will be evident, and measuring around all of these will cause the measured perimeter to be longer than that obtained at the larger scale. If a smaller scale still is selected, even more detail will be resolved, and yet a longer perimeter obtained. Ultimately, the coastal perimeter would be measured around each grain of sand; this would be a very long distance indeed. We can conclude that for surfaces of this type, which display continuously more detail at higher magnification, any measurements of dimension will depend on the scale at which the dimension is measured. This implies that if a particle has a fractal surface, different techniques of measuring, for example, surface area may give very different results. Figure 1.17 shows this phenomenon for a typical irregular particle, while Figure 1.18 shows a micrograph of a floc (of micronized drug in a chlorofluorocarbon) displaying a strongly fractal structure.

The commonest method of characterizing a fractal curve is to state a quantity known as its *fractal dimension.* We are intuitively familiar with the concept of integral dimensions of 1, 2 and 3 to characterize a line, a surface, and a volume respectively. The fractal dimension is a similar but non-integral quantity, and can be thought of as indicating that a line or surface has some of the geometrical properties of an area or volume respectively. Thus, for example, if a line curve had a fractal dimension between 1 and 2 it would indicate that the line, owing to its infinitely convoluted character, filled a significant area, and so took on some of

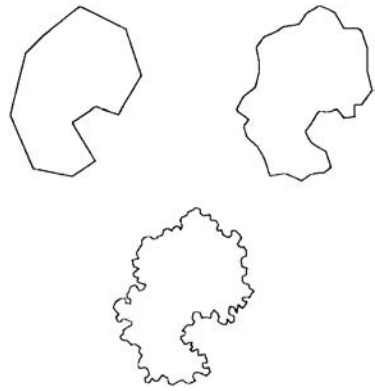


Figure 1.17. Fractal nature of particulates leads to increasing surface area as the resolution is increased.

the geometrical properties of an area. Similarly a fractal dimension between 2 and 3 indicates that a surface is so rugged that it takes on the character of a volume. Methods for measuring the fractal dimension of a particle or aggregate will be studied in Chapter 10.

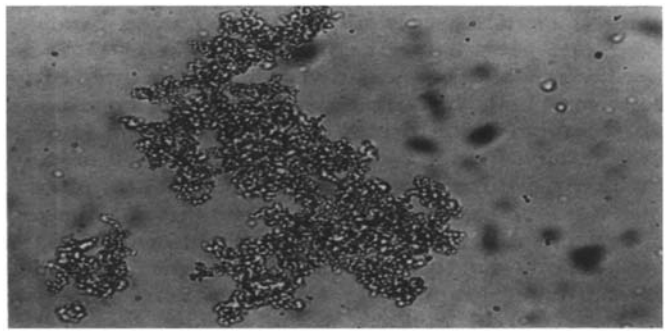


Figure 1.18. Fractal structure of flocculated drug particles in chlorofluorocarbon propellant. Magnification=100.

The study of fractals is a newly emerging and highly active one, with applications in almost every field of science. The mathematical properties of fractals are discussed by

Mandelbrot (1980), and further techniques for obtaining fractal measures and their applications to the study of particles are given by Kaye (1977, 1986) and by Avnir (1989). There have been many examples of the use of fractal measures to characterize particles, in particular irregular or aggregated materials. In particular, fractal characterization can lead to an understanding of the processes involved in the formation and structure of aggregates in flocculating systems. For example, Kaye (1986) studied the formation of carbon black agglomerates by this method, and concluded that they were formed by diffusion-limited aggregation rather than by cluster-cluster collisions. It is certain that increasing use will be made of fractal characterization of particulates in the future, and for many materials with complex structure, the fractal dimension will become as important a descriptor as the equivalent diameter.

REFERENCES

Avnir D. (ed.) (1989). *The fractal approach to heterogenous chemistry.* John Wiley, Chichester.

Flook A.G. (1981). Fourier analysis of particle shape. In: Stanley-Wood N. and Allen T. (eds), *Proc. conf. particle size analysis,* Bradford; John Wiley, Chichester. 255–262.

Gates A.O. (1915). *Trans. AIME* **52** 875–909.

Gaudin A.M. and Meloy T.P. (1962) *Trans. AIME* **237** 43–50.

Heywood H. (1963). *J. Pharm. Pharmacol.* **15** Suppl. 56T.

Hausner H.H. (1966). *Planseeber Pulvermetall.* **14** 74–84.

Kaye B.H. (1977). Characterization of the surface area of a fine-particle profile by its fractal dimension. In: Groves M.J. (ed), *Proc. conf. particle size analysis,* Salford; Heyden, London, 250–262.

Kaye B.H. (1986). *Powder Technology* **46** 245–254.

Kaye B.H. (1989). *A random walk through fractal dimensions.* VCH publishers (UK) Ltd, Cambridge.

Mandelbrot B.B. (1980). *The fractal geometry of nature.* W.H.Freeman, San Francisco,

Morris N.J. and Warburton B. (1984). *J. Pharm. Pharmacol.* **36** 73–6.

Naylor A.G. and Wright C.D. (1977). Shape analysis of particle profiles using the Fourier transform. In: Groves M.J. (ed), *Proc. conf. particle size analysis,* Salford; Heyden, London, 101–119.

Wadell H. (1934). *Physics* **5** 281–291.

2
Sample selection and preparation

Probably the most important area of any study of particle size is the selection and preparation of the sample for analysis. No particle size analysis instrument, even with the most sophisticated computer processing, can compensate for poor sample selection or preparation, and a lack of care at this stage will render all later work useless. Owing to the importance of proper sampling for any analytical procedure, a range of standard methods are specified in BS 3406.

In this chapter we will examine general methods for the selection of samples, and their preparation for analysis. In some cases specific sizing techniques require specialized methods of sample preparation, and these will be discussed in the appropriate chapters.

SAMPLE SELECTION

The object of sample selection is to ensure that the sample of material removed for study is representative of the whole batch being studied. The scale of this problem is often not fully appreciated; it is not uncommon for a process feedstock to be handled in multitonne quantities, and then characterized by a technique which only requires a few milligrams or less of material.

In order to devise methods for the selection of representative samples, it is useful to consider why a small sample taken from a larger one may not be representative of the whole. The answer is that most materials, whether in flow or storage, are continually undergoing some form of separation process. Many examples of this are known, but

probably the most commonly encountered is that of *segregation.* If a bulk material is subjected to pouring or vibration, there is a tendency for the smaller particles in the material to fall through the gaps between the larger particles. Consequently the fine material is concentrated in the bottom of the container, and the larger particles remain at the top.

Materials may 'unmix' in this way not only owing to differences in particle size, but also owing to differences in density, surface properties, electrostatic charge, etc. These phenomena are more evident in systems which contain more than one material. A striking (and inexpensive!) demonstration of this effect, which is very suitable for the lecture theatre, is commonly used by the author and can be performed as follows. Select a tall measuring cylinder of about a litre capacity, preferably one that can be sealed with a bung or glass stopper. Half fill it with ordinary granulated sugar, then add about 50 g of instant coffee (the finely powdered sort, not coarsely granulated). Pass the stoppered cylinder around the lecture theatre, and invite the audience to attempt to mix the materials thoroughly by various shaking or inversion movements.

Any attempt to shake the cylinder results in a patchy, banded pattern, usually with more brown material at the base of the cylinder. Tapping the side results in a conspicuous brown streak on that side of the vessel. The more careful observers may notice that, in addition to the difficulty in obtaining a uniform mixture of brown and white powders, the particle size of the sugar is noticeably smaller at the bottom than the top of the container. The obvious conclusion is that, given a bottle of powder to sample for analysis, the last thing you should do is give it a good shake!

Separation of this type is characteristic of most situations in which powders are moved. If a powder mixture is poured into a heap, the edges are enriched in coarser material. Powder moving through a vibratory hopper or on a conveyor may segregate into fines at the centre and larger particles near the edge. Silos of powders stored near a source of

vibration (e.g. a large generator) will be richer in large particles at the top and fines at the bottom.

Separation processes also occur in systems other than powders. Particles suspended in liquids separate as described by Stokes' law, as do liquid droplets suspended in liquids (emulsions). Finally, sampling particles in gas streams (e.g. for safety or pollution monitoring) can pose very specialized problems, since particles will tend to deposit anywhere the gas flow changes direction or velocity. Techniques for sampling of airborne particles will be examined in Chapter 8, under the discussion of aerosol systems.

The difficulties posed by segregation have led to what are termed the 'golden rules of sampling'; these are:

1) Sample from the powder when it is in motion in a stream (e.g. being poured)

2) Take several short samples at different times from the powder flow, rather than one sample for a longer time.

These rules are designed to ensure that the whole of the batch is sampled, to average out any 'unmixing' that may have taken place.

Sampling of bulk materials usually takes place in two stages. Large scale quantities are sampled to obtain a primary sample of 100 g–1 kg; this is then further divided to obtain the test sample. These steps usually require different methods, so will be examined in turn.

SELECTION OF PRIMARY SAMPLE

The technique used for sample selection broadly depends on whether the powder is in a static store or is flowing through the processing plant. It is almost always better to collect material flowing through the plant, since it then becomes possible to sample most of the particle stream at different times; the whole of the material is effectively available for analysis. Stored static materials are more difficult to sample since some parts of the bulk are often less accessible than others, and segregation may have occurred.

Sampling from flowing powders

The primary sample is usually obtained directly from process feedstock and arrangements for its collection will often be built into the processing equipment. The object in designing such apparatus is to ensure that the whole of the flowing powder is sampled, since the stream may have spatially separated itself according to particle size. This is easiest if the powder is falling through air (e.g. from a hopper or conveyor), when a sampling box can be inserted to collect material. A typical design is shown in Figure 2.1; the box is sufficiently long to sample the whole stream in one dimension, and is scanned in the opposite dimension during sampling. The smallest dimension of the box aperture should be several times larger than the largest particle in the stream. It is important to ensure that the box never gets over full, since if material collects in a heap on the top of the full box, it will be segregated.

Sampling from material moving on a flat surface (e.g. a conveyor or vibratory feeder) is more difficult owing to the need to sample the whole cross-section of the stream. If the problem is considered during apparatus design, a *divertor* can be built into the plant, which allows the entire stream to be temporarily directed to a separate container; adding this apparatus as an afterthought can be expensive and time-consuming. Some apparatus use blades or scrapers to push an entire slice of the stream off the conveyor into a sample bin; the operation of these should be checked carefully, since any material left on the belt will almost certainly be of biased size distribution.

Sampling from stored powders

The major problem of sampling stored materials is one of accessibility. Most stored materials will have been poured, and it is safest to assume that they are segregated and sample accordingly. Removing batches from stored drums does not conform to the' golden rules', and so any such sampling method is likely to be inferior to one in which the powder is sampled while in motion.

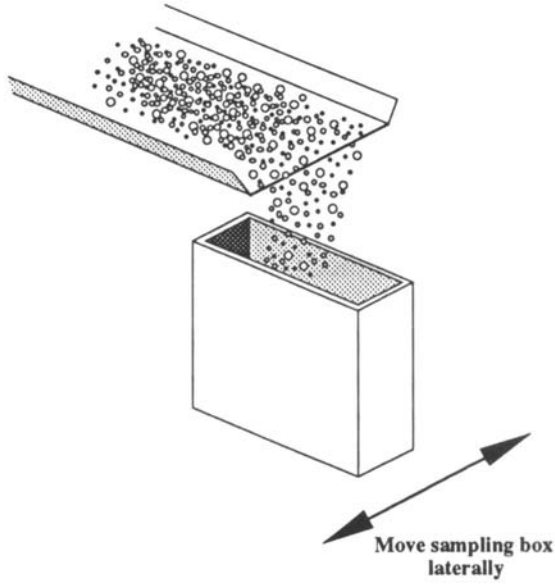


Figure 2.1. Sampling from a flowing powder.

There are two approaches to sampling from static containers; one is to use a mechanical mixer to mix the entire sample before sampling; this may not be possible if the bulk is very large. The other, more common method is to take several samples, randomly but uniformly spread throughout the bulk, and combine them. If the material is in a number of containers (e.g. drums), then a significant proportion of these should be sampled. Some workers (e.g. Allen 1981) recommend that containers for sampling should be selected randomly; this can be performed with a table of random numbers, or a sample can be taken, for example, every tenth container. For many pharmaceutical materials, in which quality control is of paramount importance, it may be considered essential to sample every container. This protocol could be relaxed somewhat if experience showed that there was very little variation in the feedstock. In general, it is better not to combine the samples from the containers before analysis in order to obtain an average, since then any information on the variability of the material would be lost.

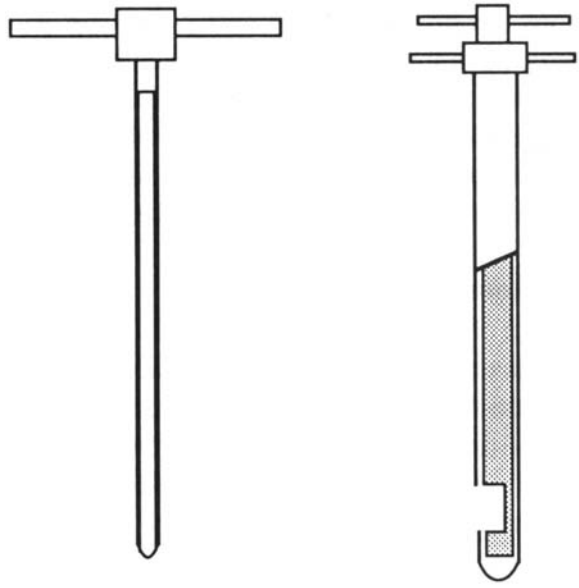


Figure 2.2. Sampling spears.

A number of devices are available to remove samples from containers. The simplest is the *scoop,* a short-handled, deep sided shovel which can be used to take a sample, normally from near the surface of the container. This is rather a poor method of sampling, since it usually selects a segregated sample, particularly if the batch cannot be mixed efficiently before sampling. To overcome this, numerous designs of *sampling spears* have been devised. A sampling spear is a long-handled device which can be thrust deep into any part of the batch, and on withdrawal returns a sample from well within the bulk; typical examples are shown in Figure 2.2. The first of these is a core cutter, which is used to cut a cylindrical core from the bulk. This works well with powders that cake, but freely flowing powders fall out as it is withdrawn. The second design overcomes this problem; it consists of a rod with a sampling chamber in the end, which can be opened or closed by rotating an outer tube. The spear is thrust into the material with the tube in the closed

position. It is then rotated into the open position, when it fills with powder. It is then closed again, and withdrawn.

Sampling spears can suffer from a number of disadvantages; they can break fragile particles, and can oversample fine materials, which may sift in through the coarser particles. This is a particular problem with the second of the spears shown in Figure 2.2. They do at least have the advantage that it is possible to obtain several samples from different parts of the batch, so that segregation within the container will provide less bias.

Since none of these methods are completely satisfactory, it may be preferable to wait until the container is emptied into the plant, and a number of samples can be taken during this process at different times so that the whole container is sampled. By this time, of course, if the particle size is unsatisfactory it may be too late to do anything about it!

Sampling from liquids

Liquid sampling is in many ways easier than solid sampling. The major problem is that sedimentation may have occurred; heavy solids will settle to the bottom of the container, and oil droplets of emulsions will rise to the top. Normally the material can be mixed with a stirrer; it is important that this is done at the lowest possible speed in order to prevent damage to the dispersion. The object is to mix the bulk rather than to apply shear, and so a paddle stirrer or propeller stirrer, which sets up a strong circulation in the material, is appropriate; devices such as blenders should never be used for this process, even at the lowest speed. This is particularly true in the case of emulsions, which may separate or coalesce if sheared. Emulsions should never be subjected to excessive shear, and all the handling equipment, such as pumps and transfer lines, should be selected with this in mind.

SPLITTING THE PRIMARY SAMPLE

At this point a sample of approximately 100 g–1 kg will have been obtained, and further division may be required to produce the test sample. Some size analysis methods (e.g sieving) require fairly large samples, and no further division may be needed. However, many modern instrumental techniques such as electrical zone sensing or light scattering require only a few milligrams of material. In these cases it is usual to separate a sample of about a gram, suspend it in a suitable medium, then take several smaller batches of this suspension for analysis, noting any variation between them.

Manual methods of sample splitting

The simplest method of sampling is shaking followed by scooping of a small number (4 or 5) of small samples. The efficiency of this method was investigated by Kaye and Naylor (1972), who found that some shaking techniques were capable of introducing a considerable bias. The worst methods are those which allow the powder to flow from one end of the container to the other, since it segregates as it forms a heap at the bottom of the vessel. Since the flow of the powder will depend on the bottle shape, shaking can never be relied upon to provide an unbiased sample.

The *quartering* method involves pouring the powder into a conical heap, flattening it into a layer, and taking a quarter 'pie slice', usually with a specially shaped dividing knife. The process can be repeated several times as required. The method relies on the fact that the heap, even if segregated, will be circularly symmetrical, and so no bias will be introduced if a slice is removed. In practice symmetry variations can arise easily, and so the method can introduce errors. In common with all manual splitting methods, it can prove operator dependent, and so is best avoided in critical analyses.

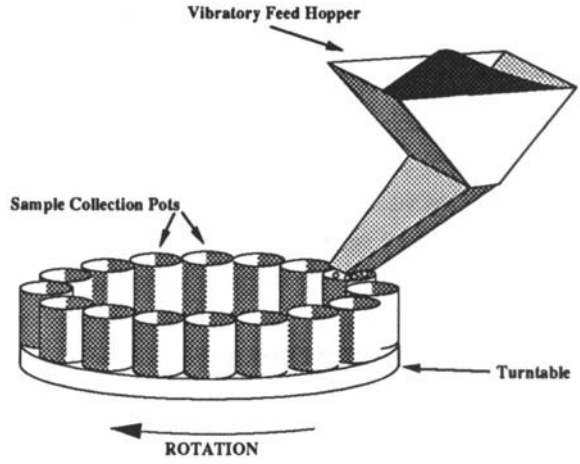


Figure 2.3. The spinning riffler.

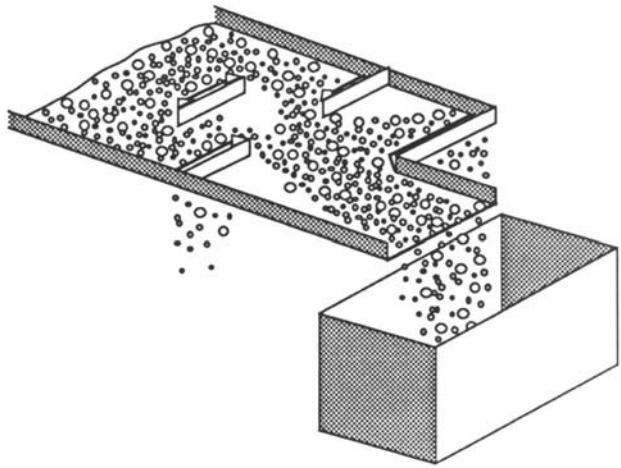


Figure 2.4. The table riffler.

Mechanical methods of sample splitting

Many devices have been described for sample splitting, the best-known being the *spinning riffler* (Figure 2.3), which is made in several sizes by most manufacturers specializing in powder analysis (e.g. Fritsch, Ladal, Pascall) and is available to the specifications recommended by BS 3406. This device consists of a vibratory feeder which streams the material on to a rotating table, which contains a ring of collecting

vessels. The sample accumulates in these vessels, and one or more can then be removed for further division or analysis. Very large versions are available for the division of bulk samples if necessary. This device is generally accepted as introducing a minimum sample bias and is very popular. Consequently they are more expensive than the other devices described below, (approximately £3000 as of 1991). It is not unusual to see makeshift devices fabricated from discarded record turntables!

The *table riffler* (Figure 2.4) consists of an inclined vibrating table, down which the sample slides; its flow is broken up and spread across the table by a series of inclined barriers before a sample is taken from a chute at the base. In many designs the chute collects half of the stream from one side of the table. This is then returned for further division until a suitably small sample is obtained. The efficiency of this device can be dependent on the initial uniformity of the powder spread on the table.

The *chute sample divider* or *box sampler* (Figure 2.5) uses a V-shaped container from whose base is a series of chutes leading to sample collectors. The powder is placed in the upper container, and emptied through the base into the sample collectors. The statistical variation introduced by this device is also larger than that obtained from rotary rifflers. Both chute and table samplers are likely to be adequate for undemanding samples (i.e. those which contain only a narrow range of particle sizes, and which flow easily), but will encounter increasing difficulties as the sample becomes more heterogeneous.

Many other devices have been described in the literature, but are not in common use. The main objection to many of these is that they have not been extensively characterized, and so, although they may work well, the confidence in the results is necessarily inferior to those produced by other tried and tested methods.

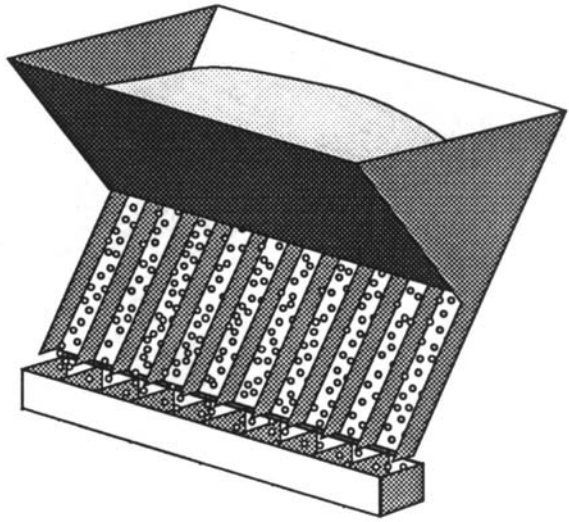


Figure 2.5. The chute sample divider.

DISPERSION OF MATERIALS

Many methods of size analysis require the sample to be dispersed in a liquid medium prior to examination. This may pose some significant problems. The powder normally consists of primary particles which may be agglomerated into masses; the usual requirement is to obtain the size distribution of the primary particles. Studying agglomerated masses can be extremely difficult, since they will tend to disintegrate in an indeterminate manner under certain conditions. The dispersion of the sample consists of a number of steps:

1) Wetting of the sample by the solvent

2) Disintegration of the agglomerates to a uniform suspension of primary particles

3) Stabilization of the suspension to ensure that agglomerates do not re-form.

If a powder is well-behaved and disperses easily, these steps occur concurrently and are often not apparent. However, a knowledge of these processes can be extremely useful in those (many!) cases in which dispersion problems are encountered.

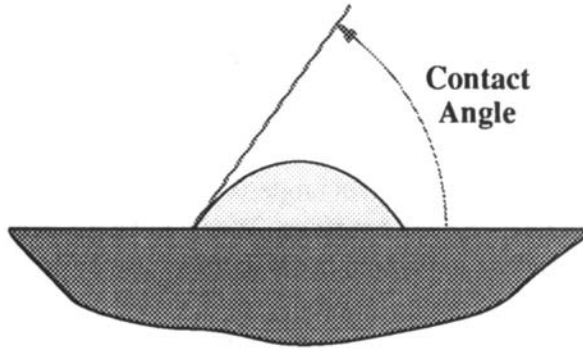


Figure 2.6. Definition of contact angle.

Wetting

Wetting of a surface is said to occur when a liquid spreads spontaneously over the surface. The theory of wetting is covered in most standard texts of applied physical chemistry (e.g. Florence and Attwood 1985). Consider the situation in Figure 2.6, in which a drop of liquid is placed on a solid surface. The drop forms a lenticular shape, the extreme edge of which lies at an angle to the solid surface. This angle is called the *contact angle* for the liquid on the material. Without delving too deeply into the mathematics, it is fairly intuitive that if the contact angle is finite, the liquid will form a stable drop and will not spread on the surface. Alternatively, if the liquid spreads on the surface, the contact angle must be zero.

Contact angles are dependent on surface tensions and interfacial free energies; in practice we would rarely consider measuring any of these quantities simply in order to resolve a powder dispersion problem. However, a more useful relation can be obtained as follows. Suppose we have a solid surface, and measure the contact angle of a range of liquids on this surface. If we then plot the measured contact angle against the surface tension of the liquid, we obtain a graph similar to that of Figure 2.7, which is called a *Zisman plot* (Zisman 1964). Liquids of high surface tension have a large contact angle, and as the surface tension is reduced, the contact angle falls until a point is reached where the contact

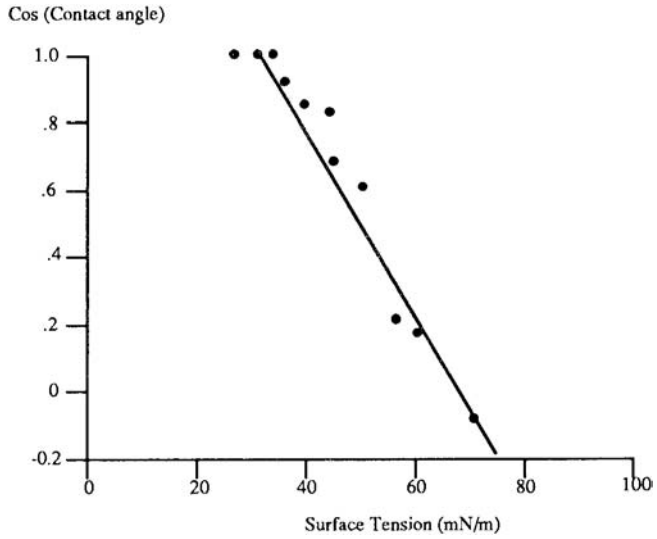


Figure 2.7. Critical surface tension of various liquids on polyethylene (redrawn from Fox and Zisman, *J. Colloid Int. Sci.* 7 (1952) 428).

angle is zero. For the majority of liquids with lower surface tensions, the contact angle on that particular solid remains zero, and these liquids will spread on the surface.

The point at which the contact angle is just zero is called the *critical surface tension* of the solid surface. We do not even need to measure this critical surface tension in order to make use of this relationship, since, if a test liquid does not wet the solid, we simply choose one with a lower surface tension. Surface tensions are tabulated in most databooks (e.g. Weast 1974) and a table of surface tensions of common liquids is given in Table 2.1.

In many cases the choice of liquid is limited by other considerations. For example, diethyl ether has a very low surface tension and wets a wide range of materials, but its flammability makes it a rather hazardous material to use. Surfactants are commonly used in order to obtain low surface tensions if a pure solvent is not appropriate; in fact most workers would use an aqueous system with a surfactant in preference to a pure organic solvent system. The only exception occurs when the powder is water-

Table 1 Surface tensions of some common liquids

Liquid	Surface tension (mN m^{-1})
Acetone	23.7
Bromoform	41.5
n–Butanol	24.6
Chloroform	27.1
Cyclohexane	25.5
Diethyl Ether	17.0
Diethyl Phthalate	37.5
Ethyl Acetate	23.9
Ethylene Glycol	47.7
Ethanol	22.7
Glycerol	63.4
n–Hexane	18.4
iso–Propanol	21.7
Methanol	22.6
Methylene Iodide	50.8
Methyl Ethyl Ketone	24.6
n–Propanol	23.8
Toluene	28.5
Water	73.0
p–Xylene	28.4

soluble, when an organic solvent must be used as a dispersant.

Disintegration

The material presented for analysis normally contains not only primary particles, but agglomerates which may be several hundred times larger. These may be aggregated by a number of mechanisms; the particles may be caked from an earlier operation, or compacted, or glued with some gummy material; in extreme cases the smaller particles may act as a

cement to stick the larger ones together. In many cases such agglomerates will disperse as soon as wetting has occurred; however, dispersion may be incomplete, and this may result in a biased or erratic sample analysis, even if a surfactant is used in the dispersant medium.

The usual method of dispersing agglomerates is by using ultrasonic shearing devices. The *ultrasonic bath,* now too well-known as a cleaning device to require description, is the standard equipment for these problems, and the wetted sample would normally be placed in such a bath for 5–10 minutes prior to analysis. If erratic results are obtained after this exposure, it is possible that harsher treatment is needed. The *ultrasonic probe* works on a similar principle to the bath, and consists of a small high power (50–200W) ultrasonic transducer which is immersed in the sample. This applies a considerable amount of shear to the sample, which will break up most aggregates. An exposure of less than 1 minute is recommended; further treatment will cause considerable sample heating and may damage the particles. In general these methods are only required for submicrometre systems which may be strongly bound by surface forces, and bath sonication will suffice for most routine samples. These methods can only be applied to solid dispersions; if emulsion systems are treated in this way, coalescence or further droplet size reduction is likely to occur on an unpredictable scale, and the final sample will bear little resemblance to that presented for analysis. Sonic probes should also be used with caution for the processing of delicate materials such as microcapsules or microparticles made from polymers or proteins, which can be easily damaged. It is better to try different surfactants than increase the shear on these systems.

Stabilization and flocculation

Even after a powder system has been wetted and its agglomerates dispersed by shearing if necessary, a stable system may not be obtained. The particles may cluster

together to form large masses called *flocs,* which may rapidly settle out of suspension.

The stabilization of particulate systems and the formation of flocculated systems is an extensive area which is beyond the scope of this book. The reader needing detailed information will find the work by Hunter (1987) useful. Briefly, the problem can be stated as follows. Particles in suspension are subject to a range of interparticle forces which can be both attractive or repulsive in nature. The most important forces are:

1) The *Van der Waals force,* which is attractive, and so causes the particles to aggregate. Van der Waals forces are dependent on the nature of the interparticle medium, suggesting that a particularly troublesome dispersion problem may be solved by selecting a different dispersing liquid.

2) *Electrostatic forces,* which are always repulsive. Electrostatic forces arise as a result of charged groups on the particle surface, such as adsorbed ionic surfactant, and are characterized by the so-called *zeta potential,* which is the potential measured by electrophoresis of the particle (Hunter 1981). In general a potential of tens of millivolts indicates a sufficiently large repulsive force to stabilize the dispersion; an adsorbed layer of most charged surfactants can provide this. Zeta potential measurement is, of course, of little value in solving a new dispersion problem, since one normally has to disperse the sample in order to measure the potential in the first place! However, if the surface properties of the material are fairly well understood, these may suggest strategies such as, for example, changing the pH of the medium in order to ionize the surface and thus cause a surface charge to develop.

Ions in solution screen the electrostatic forces, and so the repulsion is strongly influenced by the electrolyte composition of the interparticle medium; in general, the more electrolyte is present, the weaker will be the attractive forces. As the electrolyte concentration is increased the electrostatic forces will decline to a point where they cannot overcome the attractive Van der Waals forces, and the

dispersion will flocculate. This point is termed the critical flocculation concentration, or CFC. Multivalent ions are much more effective in this regard; the *Schultze Hardy rule* states that the CFC's of ions are inversely proportional to the sixth power of their charge, and so ions such as calcium and aluminium are powerful flocculants. If an electrolyte solution has to be used for dispersion (e.g. as in the Coulter counter), multivalent ions should definitely be avoided.

3) *Steric* forces are present between surfaces onto which longchain macromolecules or polymers have been absorbed. This mechanism is an alternative to electrostatic stabilization, and commonly operates between particles coated with nonionic or polymer surfactants. As two particles coated with polymer approach one another, the ends of the polymer chains interpenetrate; thermodynamic arguments show that this usually leads to repulsion between the particles. In some cases, however, the particles may be bound together by the polymer; it is then referred to as a flocculant. This can be technologically useful, but not for our purposes!

PRACTICAL TECHNIQUES FOR SAMPLE DISPERSION

The usual method of dispersing a water-insoluble sample is to try one of a small range of surfactant solutions; everyone has their own favourites, and the following list will solve most problems:

2% Sodium dodecyl sulphate
4% Pluronic F-68
4% Tween 80
4% Span 60
5% Nonidet P-40

One of these will usually work; if not, there are many thousand other surfactants available. Nonionic materials are usually preferred since they are less aggressive; some ionic surfactants can lead to solubilization of particulates. A few drops of washing-up liquid can be very effective! The

concentrations of surfactants indicated above are upper limits; significantly more should *not* be used, since this may cause a tendency to foaming. Many size analysis techniques (e.g. laser diffraction) will treat small bubbles identically to the test particles, adding them to the distribution, with consequent false results. Stock solutions of surfactants should of course be filtered, and more importantly a few drops of antibacterial preservative, such as thimerosal, should be added to prevent bacterial growth. In aqueous systems, phosphates can prove useful dispersants; sodium hexametaphosphate (Calgon) and sodium pyrophosphate are particularly useful. Before using any new surfactant, particularly for submicrometre light scattering (e.g. PCS, Chapter 7), dilute a few drops alone to check that it forms a clear solution. Many nonionics form large micelles which can be detected by light scattering and will render the results valueless.

The particle suspension should be carefully examined for the presence of flocculation prior to analysis. The simplest test is microscopy at around 400 X, which will easily distinguish most aggregated systems. A typical aggregated suspension is shown in Figure 2.8 (a flocculated fat emulsion). Flocculated suspensions can also be distinguished by their sedimentation properties; large flocs settle quickly, and form a bulky layer since they have an open structure and trap much of the suspending fluid. Alternatively, nonflocculated suspensions have a smaller particle size, and will settle slowly to form a compact sediment.

Water-soluble materials are often suspended in a nonaqueous liquid. There are many candidates, but it should be remembered that most water-soluble materials are soluble in slightly polar solvents (e.g. alcohols) to a lesser degree. Although this solubility may be small, it may be sufficient to dissolve some of the smaller particles and distort the distribution. This is especially true when very dilute suspensions are being prepared, e.g. for light scattering studies. In these cases very nonpolar materials (e.g. alkanes) are useful. Bernhardt (1984) has selected

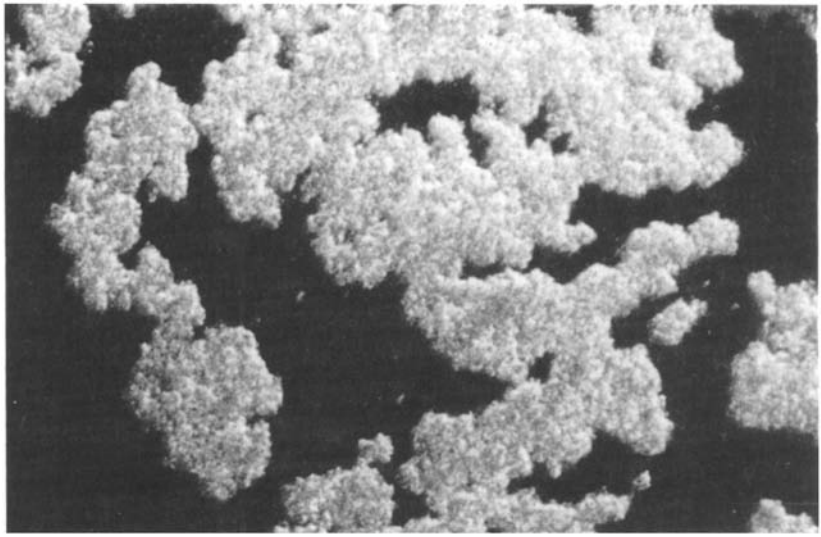


Figure 2.8. Flocculated fat emulsion.

dispersing systems for a wide range of solids, and his table is a useful starting point for a difficult system. Suspensions in nonpolar media are almost always flocculated, and a dispersing agent is essential; lecithins and oleic acid are useful in these cases.

A further common method of dispersing water-soluble materials is to suspend them in a saturated solution of the material itself. This method can be useful, but is best avoided if possible. Only a small temperature fluctuation is required in order for more material to be dissolved from the particles; usually the smallest particles dissolve first, and the distribution is correspondingly biased. Small particles can also disappear by a process called Ostwald ripening, in which the large particles grow at the expense of the smaller ones, even if the solution is saturated.

CALIBRATION AND REFERENCE STANDARDS

Although some particle measurement techniques provide an absolute size measurement, many only measure relative size and so require a calibration material to allow absolute size

to be calculated. Many materials have been used in the past as reference particles. Currently polystyrene microspheres ('polymer latices') are almost universally used. These are accurately spherical and of highly uniform size, and are calibrated by electron microscopy; they are obtainable with diameters from 50 nm to several tens of micrometres. Larger sizes (0.1 mm and larger) can be calibrated using glass beads from the National Bureau of Standards or the National Physical Laboratory. The National Physical Laboratory also can supply standard reference materials (quartz spheres) to the EEC Community Bureau of Reference (BCR) standards (Wilson 1977). The BCR also specifies a range of polydisperse test materials which are accurately characterized fractions of quartz powder.

Prior to the introduction of polymer microsphere standards, the most widely used calibration materials were various types of pollen, and these may still be encountered occasionally. Many pollens, for example that of mulberry, are nearly spherical and of uniform diameter, and so formed quite good standards, although they were often suspected of swelling in water.

REFERENCES

Allen T. (1990) *Particle size measurement.* 4th Edn., Chapman Hall, London.

Bernhardt C. (1984) *Preparation of suspensions for particle size analysis.* Academie fur wissenscaften der DDR, Forschunginstitut, Freiberg.

Florence A.T. and Attwood D. (1985) *Physicochemical principles of pharmacy.* 2nd edn. Macmillan, London.

Hunter R.J. (1981). *Zeta potential in colloid science.* Academic Press, London.

Hunter R.J. (1987). *Foundations of colloid science.* Clarendon Press, Oxford.

Kaye B.H. and Naylor A.G. (1972). *Particle Technology,* **13** 47–66.

Weast R.C. (Ed.) (1974). *Handbook of Chemistry and Physics.* 54th Edn. CRC Press. Cleveland. Ohio.

Wilson R. (1977). Some current work on particulate reference materials. In: Groves M.J. (ed), *Proc. conf. particle size analysis,* Salford; Heyden, London, 101–109.

Zisman W.A. (1964). In *Contact angle, wettability and adhesion,* Fowkes F.M. (ed.), Advances in Chemistry **43**, A.C.S., Washington, 1–51.

3
Sieving

Sieving is probably the earliest particle separation method that was devised; it is illustrated in Egyptian art, and was in common use in the 16th century, as described in *De re metallica* by Agricola (1556). Early applications were simply designed to classify powders into approximate size ranges; the use of sieving as a precision method of particle size analysis (so-called *test sieving*) was not popular until techniques for producing sieves of good precision were available in the mid 19th century.

SIEVE SIZES AND RANGES

The introduction of sieves for size analysis is often attributed to Von Rittinger (1867), who proposed that sieves be made with apertures in a geometric series, starting from 75 μm (the smallest that could then usefully be employed) and increasing in aperture by steps of √2, a so-called √2 *progression.* Many other standards have been proposed since, but most are based on a geometric progression of aperture diameters. The U.K. standard is BS 410:1969, which specifies a fourth root of two progression based on 100 μm (the progression can of course be extended to smaller as well as larger sizes), and a similar range is proposed in the American ASTM E11.81 standard. The international standard (ISO R565) uses the same range but only recommends the √2 progression, so omits every other sieve from the BS and ASTM series. The German system is similar and is specified in DIN 4187, 4188 and 4195. The French AFNOR series is slightly different in that it is based

at 1 mm and uses a tenth root of 10 progression, so that there are 10 sieves per decade of particle size. Coincidentally, consecutive sieves allow through spheres with almost exactly a 2:1 volume ratio (the error is about 0. 5%).

Sieves are produced in a range of standard diameters, the most popular for general use being 200 mm, although 300 and 450 mm are common larger sizes, and many of the smaller aperture sieves are produced in 3", 4" and 6" versions. The top and bottom rims of the sieves are usually of a size and form which allows the sieves to be stacked in a vertical column (a *nest* of sieves), so that a sample can pass through them in sequence. Well-known manufacturers include Endecotts (UK), Fritsch (Germany), Buckbee-Mears and Veco (USA), the latter two specializing in smaller microsieves.

What does sieving measure?

The *sieve equivalent diameter* is defined as the size of a sphere which will just pass through the mesh of a particular sieve. It is important to specify whether the definition refers to square-meshed or round-hole (punched or etched) sieves. A 100 μm sphere will just pass through the hole of a 100 μm square or round-hole sieve, but an irregular particle may pass through one sieve and not the other. For example, a cube of side 100 μm would pass through a similar sized square hole, but not a 100 μm diameter round hole. Consequently the particle would have different sieve equivalent diameters in round or square hole sieves.

Passage through a sieve requires that the particle has a cross-sectional area which is less than that of the aperture; a particle must have two dimensions less than the sieve hole diameter in order to pass through. Consequently sieving is often said to sort particles according to their second dimension. The third dimension of the particle can be longer than the sieve aperture (e.g. a needle-shaped particle), and the particle will still pass through the mesh. If the third dimension is very long, however, the sieving time may be

extended, since the particles will tend to lie in their stable position on their sides, rather than 'on-end'.

The material separated in the sieves is normally measured on a weight basis, either by weighing the sieves or by chemical analysis. Consequently a sieve analysis provides a mass or volume distribution. This can be converted to a distribution of diameters or surface areas if the particles are assumed to have a certain shape (usually spheres), but this is rarely performed, and the results are usually presented as percentage mass against sieve equivalent diameter.

METHODS OF SIEVE CONSTRUCTION

Woven sieves

The woven sieve is the oldest design, and it is normally made by weaving fine metal wire into a square pattern, then soldering the edges securely into a flattish cylindrical container (Figure 3.1). The apertures are not exactly square, due to the three-dimensional nature of the weave. This type of sieve was originally specified in terms of the *mesh number,* which is the number of wires to the inch of mesh cloth; e.g. a 120 mesh sieve has 120 wires per inch. This says nothing about the aperture size except that they have to be smaller than 1/120 of an inch. Since the wire thickness fills part of the sieving surface, the actual aperture is smaller than that suggested from the mesh size. Wire thicknesses are also specified in the various standards, most sieve wires being from 15% to 30% of the diameter of the aperture, to contribute sufficient strength to prevent the sieve from being distorted or damaged in use.

Sieve meshes are produced from a wide range of materials; normally the finest are woven from bronze or hard brass, while mild steel or more commonly stainless steel may be used for the larger sizes. Sieves made from silk,

Figure 3.1. Woven sieves.

nylon and other plastics are sometimes used, particularly in the food industries; monofilament nylon ('Monodur') is popular for pharmaceuticals owing to its inertness. These sieves naturally have inferior dimensional stability and poorer hole tolerances than metal sieves, owing to the flexibility of the plastic, and are more easily damaged during cleaning.

Etched Sieves

Woven sieves cannot be made with apertures smaller than around 50 μm, owing to the difficulty in handling the wires, and they would in any case be very easily damaged. Finer sieves (down to an aperture of 5 μm) are made by etching technology similar to that developed for the semiconductor industry. A thin metal foil is coated with a photoresist, and a grid pattern is exposed onto the resist through a suitable mask. The unexposed resist is washed away in a developer, leaving a pattern of holes which can be etched through the foil. The excess resist is then removed. The resulting

perforated foil is very fragile and is usually supported on a coarser mesh; they are generally only made in the smallest sizes and so are only suitable for very small-scale work. The dimensional tolerance of the holes is very high in comparison to woven sieves, where the distribution of hole sizes can be remarkably broad.

The foils used must be as thin as the aperture diameter, or the holes would resemble cylinders, which would be easy to block. Some manufacturers use profiled holes (e.g. conical holes) which are claimed to reduce blockage problems. These sieves are supplied by Veco, Endecotts, Fritsch, etc, and although expensive, are very useful. A particular application of these sieves is the separation of microparticles. Small particles of polymers (0.1–20 μm) are of considerable current interest as drug delivery systems, and extensive literature exists on their preparation and properties. It is often necessary to produce a narrow size distribution of particles from a population with much broader initial distribution, and microsieves are one of the few feasible ways of doing this.

Punched Sieves

The largest size sieves are made by punching holes in metal sheets. These sieves are available from 1 mm to tens of centimetres aperture, and have very good dimensional precision. They do not, however, find wide use in the pharmaceutical industry, which usually requires the smaller sizes.

TECHNIQUES FOR USING SIEVES

Wet sieving

There are two separate ways of performing a sieve analysis. Either the powder can be sieved in the dry state, or it can be suspended in a suitable liquid, and the suspension passed through the sieve. Both methods have their advantages and disadvantages. If the powder is ultimately required in a dry

state, it is tedious to suspend it and then have to dry the product (which may lead to caking). Wet sieving, however, is usually more rapid since the powder is carried through the sieve by the liquid flow. The only force driving a powder through a sieve in the dry state is its weight, and if the powder is very light or fine, this may lead to very slow sieving. Wet sieving is also indicated in cases of powders which form lumps or 'ball' when dry, but can be dispersed when wet. Micromesh sieves below 50 μm are usually used wet, since there is very little tendency for anything to pass through such small holes unless aided by a stream of liquid.

A wide range of wet sieving techniques have been described. Standard methods for using sieves are described in BS 1796, ASTM C136 & D452 and ISO 2591. Some authors (Colon 1970, Peterson 1969) stack the sieves in a nest which is filled with liquid and vibrated or shaken. This method uses only a small amount of suspending fluid, but much of the 'washing through' action of the liquid is lost. A faster method, but one which requires more liquid, is to place the powder on the top sieve and wash through with a liquid stream. The powder can, if necessary, be gently dispersed or spread with a camel-hair brush which is then rinsed clean. When the top sieve contains only large particles, it can be removed, and the powder on the second sieve rinsed through (much of this will already have been separated). The process is repeated until the whole sieve stack has been rinsed.

The main problem with using the sieves in a descending series in this way is that the last fraction of fine material ends up suspended in a disproportionately large amount of wash liquid. This can be avoided if the powder is first placed in the finest sieve, and the finest particles washed through and collected. The remaining powder can then be passed through the full stack of sieves, including the finest, in the usual way, but now the liquid draining through the nest can be discarded. If this technique is used, it is essential to check when the test protocol is developed that there is no remaining fine material in the wash liquid. A sizeable

fraction of the liquid should be collected and evaporated to check for traces of fine material which were not rinsed out in the first fine sieve. If these are present, the efficiency of the first rinse must be improved.

The sieves can also be used individually; the powder is dispersed in liquid and poured through the sieve into a beaker. The fine material passing through the sieve is then poured through the next sieve, and so on. Sieving in this manner using microsieves often leads to aperture blocking; Colon (1970) and Daescher (1969) discuss methods in which liquid is forced back and forth through the mesh to avoid blocking. The simplest method of doing this is to agitate the sieve vertically in a beaker of the suspending liquid.

All of these methods suffice for size classification, but for a size analysis it is necessary to find the amount of material in each sieve. If the sample is large, the sieves can be weighed, but often the sample will weigh grams or less and the sieve may weigh hundreds of grams. Consequently the precision of direct weighing is poor. It is preferable to rinse the powder into a small tared container for drying and weighing, or in some cases for chemical analysis.

Wet sieving is not restricted to aqueous suspensions; any suitable nonaqueous solvent can be used, since bronze and stainless steel sieves are inert to most solvents. They can, however, be attacked by some acids or alkalis, which should be avoided; some of the plastic sieves can prove useful at extreme pH, although the manufacturers recommendations should be followed closely.

Hand sieving

Hand sieving is the simplest method of using sieves, and is the standard by which the accuracy of other methods is evaluated. A hand sieving analysis should always be carried out when developing a new sieving size analysis protocol, even if a machine sieving device will ultimately be used for routine testing. This is because it allows greater opportunity to observe problems (such as balling and wire coating), and also because it is more gentle and hence less likely to cause

particle damage than mechanical vibratory shakers. Surprisingly, it can also be as rapid and efficient as mechanical sieving.

The standard hand sieving protocol begins by placing the powder in the finest sieve, which is placed on a catch pan and closed with a lid. The amount of material to be sieved should not overfill the pan; 200 g is a good starting load for 200 mm sieves, but much depends on the powder properties such as density. The sieve is then gently tapped sideways by hand or with a short piece of wood, gradually rotating it. The catch pan is weighed at intervals and the process continued until no detectable increase in pan weight occurs. The powder is then transferred to the largest sieve, and the process repeated to obtain the largest fraction. The sieves are then used in descending order, finally again using the finest sieve, and adding any material passing through the sieve to the first fine batch.

There are several good reasons for removing as much as possible of the fine material first. It is the most likely to be lost, for example through a poor sieve/pan seal, by adhesion to coarser particles caught in larger sieves, or by adhesion to the sieves themselves. It is also likely to slow down the sieving process, since it may assist adhesion and clumping of the larger particles.

Hand sieving is best performed in a draught-free area on a table covered with a clean smooth card, to allow any spilt material to be detected and recovered. Many workers cover the bench with white paper ("Benchkote"), but since most pharmaceuticals are white, a black surface is preferable!

Machine sieving

Sieving machines are essentially vibrators which shake a nest of sieves. A typical example is shown in Figure 3.2. A nest of 5–10 sieves, normally in a standard progression, is placed on the machine with a lid on top and a collecting pan at the bottom; the sample is placed in the top sieve and the whole shaken for a fixed period, usually 20 minutes. Sieving should then be continued in short intervals until no more

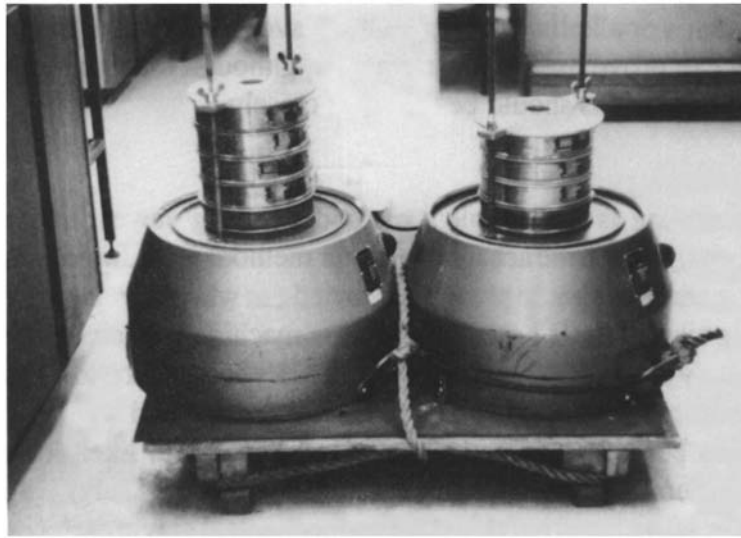


Figure 3.2. Vibratory sieve shakers in action.

than a small amount of material passes through the sieves in a given period; typical working limits are less than 0.2% in 5 minutes. The total sieving time is usually determined in an initial series of trials which can be used to produce a standard operating procedure; this can then be adhered to without the need for unnecessary intermediate weighings, unless, of course, the nature of the feedstock changes markedly.

A number of designs of sieve shaking machines are available. The simplest shake the sieves from side to side; it is a bad idea to shake the nest vertically since this may lead to particles becoming wedged in the sieve mesh, and will increase the energy of impact of the particles on the wires, increasing breakage. More complex machines impart a strong horizontal tap or jolt to the sample, which it is claimed reduces blockage and decreases sieving times. Most of these machines vibrate synchronously with the mains frequency (1500 or 3000 Hz), and the more expensive allow the vibration amplitude to be adjusted, so that separation of fragile materials can be optimized. In general, higher

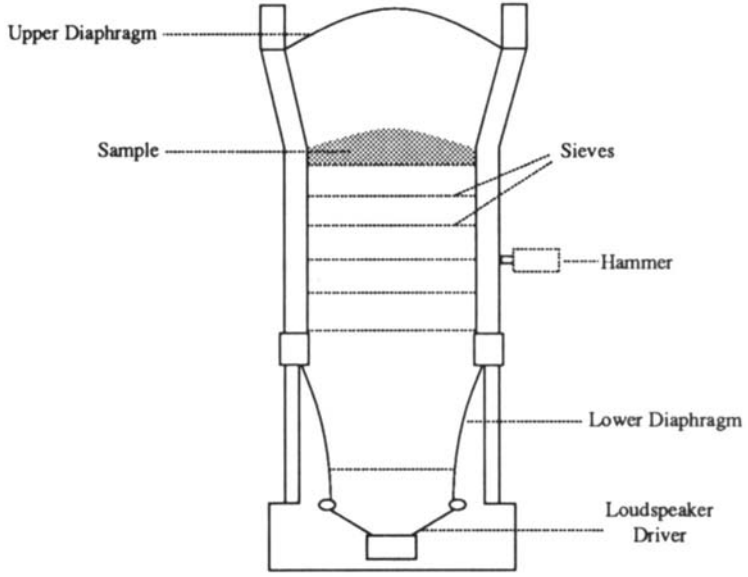


Figure 3.3. Sonic sieving.

frequencies and smaller amplitudes lead to more efficient and gentler sieving for more delicate and finer particles.

It should be borne in mind that the shaking action of sieving machines may cause loss of any fine dust in the sample. If the material is pharmacologically active, precautions should be taken to avoid inhaling this dust. The simplest is to tape the joints in the nest of sieves; a better method is to place the entire stack under an extraction hood.

Sonic and ultrasonic sieving

The sonic and ultrasonic sieving technique uses a vertically oscillating column of air to fluidize the particles in the sieve nest. The principle is shown in Figure 3.3. A nest of sieves is mounted above a loudspeaker which is driven at several kilohertz; this produces a vertical oscillation of the particles, which then sift down through the sieves. Most machines also impart a horizontal motion to the sieves at a lower frequency with a horizontal hammer action. A typical unit

costs £3–5000, which is fairly scandalous considering that a competent technician could improvise one in a day from a discarded loudspeaker.

Sonic sieving is usually used with small diameter sieves of fine aperture, and manufacturers claim that it is useful for the dry sieving of very fine particles (down to 5 μm) which would otherwise have to be wet sieved. Consequently there is much interest in the technique for the separation of microparticles. With suitable apparatus, sonic sieving can also be used for wet sieving, although it is necessary to seal the sieve stack and the loudspeaker driver.

Airjet sieving

This is a highly popular method, particularly for the sieving of fine powders. It allows most materials to be separated down to a size range of 50–75 μm, but, as with all sieving techniques, much depends on the properties of the material. In favourable cases it may allow separation down to 10–20 μm.

A cutaway view of the device is shown in Figure 3.4. The sieve is placed on the body of the instrument, and a lid covers the sieve. A vacuum pump (usually an ordinary industrial vacuum cleaner) is connected to the lower part of the instrument, which transmits the vacuum uniformly to the underside of the sieve, drawing the powder through. If this process is not to stop as soon as the pump has evacuated the apparatus, air must be let in to the top of the sieve to replace that drawn out by the vacuum pump. This air enters not from the top of the sieve, but is forced through the mesh from underneath by a series of jets, thus unblocking the mesh, which would otherwise rapidly become clogged. The jets are placed on a radial arm which rotates just under the sieve mesh, so that they sweep the entire sieve surface; they rotate at about 50 rpm. The net result is that every point on the sieve sees an alternating wave of suction followed by a reverse airjet which unblocks the mesh.

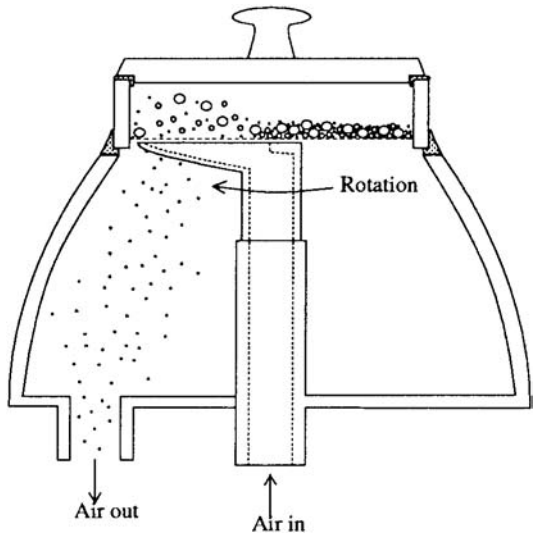


Figure 3.4. The airjet sieve.

The operation of the instrument requires that there is no appreciable leak of air from the outside; consequently it is usually supplied with sieves and lids which have rubber sealing rings; otherwise the sieves are of conventional construction. The sieve, lid, and base are held together by suction, since the inside of the device is below atmospheric pressure (which is why the air is sucked through rather than blown). The device can be used in two ways; the material which is withdrawn through the mesh can be collected on an in-line air filter and weighed, or the sieve and its lid can be removed and the loss in weight determined. The main problem with collecting material on a filter is that if there is any quantity of it, the filter may block, and degrade the performance of the instrument as a whole. The filter will also collect any airborne dust from the air drawn into the instrument, and some particles may impact inside the instrument and not reach the filter. Consequently it is better to determine the weight loss of the pan, although careful technique and accurate weighing may be required.

The airjet sieve is particularly useful for the sieving of finely ground pharmaceuticals, which are often composed of

brittle crystalline material. Used with the finer sieves, it is also useful for assessing the efficiency of micronizing devices such as the fluid energy mill. A batch of micronized powder can be sieved through, for example, a 50 or 65 μm sieve, and any large particles will be evident as the majority of the powder is removed.

Cleaning of sieves

Normally a sieve will be emptied by inversion onto a suitable surface, followed by gentle brushing; brushes for this purpose are supplied by most sieve manufacturers. The lower surface of the sieve is brushed along and across the mesh, and the occasional gentle tap will cause most of the powder to fall out. The inside of the sieve can then be brushed out to complete the powder recovery. The brush should *never* be used in a vertical, stabbing action in an attempt to remove wedged particles, as this will almost certainly distort the mesh and destroy the accuracy of the sieve.

Occasionally material remains wedged between the meshes. This presents two problems; firstly, how to effect a complete recovery of the powder for analysis, and secondly, how to ultimately clean the sieve. If the problem is serious it casts doubt on the accuracy of the entire analysis. The mass in the sieve is probably best determined by precision weighing; the sieve can then be cleaned by wet means, dried, allowed to equilibrate to room temperature and humidity, and the weight loss recorded. This whole process should be performed in a warm dry atmosphere, so that the weight loss does not include tens of milligrams of water that can easily adsorb to the sieve container surface.

Wet cleaning can normally be effected simply by running a stream of distilled water on to the underside of the sieve, perhaps with gentle brushing. If particles still remain, an ultrasonic bath containing a little detergent can be used, or if the chemistry of the material is known, a suitable solvent can be selected. Strong acids or alkalis will almost certainly damage brass or bronze sieves, but stainless steel sieves can

be subjected to more vigorous treatment. Beware, however, of attacking the solder or braze that fixes the mesh into the pan. If a sieve is still blocked we have found the careful application of a vibrating electric toothbrush (well worn, without toothpaste!) to be useful.

Cleaning problems are more severe with micromesh sieves. These can block very easily, and often the only clue to blockage is a faint haze on the polished mesh surface. The sieves can be washed with solvents but never cleaned by chemical means; it does not take a particularly active solution to remove a few micrometres of metal. Gentle and prolonged sonication is the most appropriate method, but since methods of manufacture vary, the manufacturers literature should be consulted for detailed advice; some manufacturer's recommend that their sieves are not sonicated. Often specific cleaning solutions are recommended to reduce sieve erosion by cavitation. Daescher (1969) examined a number of cleaners and cleaning solutions, and recommended that the sieves be cleaned at as high a frequency and as low a power as possible. The cleaned sieve should be examined for residual blockage and damage by microscopy.

Calibration and testing of sieves

It is rather difficult for the average user to perform a rigorous testing of his sieves, and it is certainly far cheaper to replace sieves regularly than to invest in the apparatus and expertise required for a full sieve testing program. Sieves are relatively inexpensive, and the cost-benefit ratio of testing a worn or possibly distorted sieve is high.

Sieve cloth is tested by its manufacturers by two main techniques. The first is direct metrology using travelling microscopes, for example as described in BS 410. This is usually automated to a form in which a moving microscope can count the number of apertures, and their width, in a photoelectric scanning technique. The application of image analysis techniques is also of value; if the user already has an image analysis facility, this is probably the best method

for making test measurements on sieves. Direct measurement can only be used on sieves larger than 25–50 μm; smaller sieves cannot be imaged with accuracy owing to the light diffraction by the mesh. Thus the second test method is to measure the diffraction pattern of the mesh and use optical transform techniques. This method was suggested by Konowalchuk *et al.* (1976) is very valuable for micromesh sieves, since it is capable of measuring the size distribution of all the holes at once.

Some authorities recommend that sieves are tested regularly with a two-set comparison protocol, in which two 'identical' sets of sieves are purchased and a powder is tested with both. One set is then stored as a reference, and the other used for test purposes. Every 20 or so measurements, the standard powder is tested with both sets; if a difference is found from the initial comparison, the test set is discarded, a new set is bought for reference purposes, and the old reference set becomes the new test set. A new comparison test is performed with the new sieves, which then becomes the standard until the next replacement.

Sieve sets should be compared using test materials (glass or quartz beads) with a distribution of sizes, available from National Bureau of Standards or the National Physical Laboratory, rather than with the user's own product. These materials have an accurately lognormal size distribution, so if an analysis of their size is performed with a set of sieves, and a log-probability plot constructed, the points should form a straight line. If they do not, then the best straight line can be drawn, and used to correct the sieve sizes (Figure 3.5). Be very careful, however, in using this technique near the ends of the distribution, where significant errors may be present.

A simple method of determining the cut-off diameter of sieves was described by Leschonski (1981). A standard sample is sieved using a conventional stack until the process is nearly complete and very little material is passing through the sieves. The stack is then dismantled and each sieve tapped on a piece of paper to recover a very small

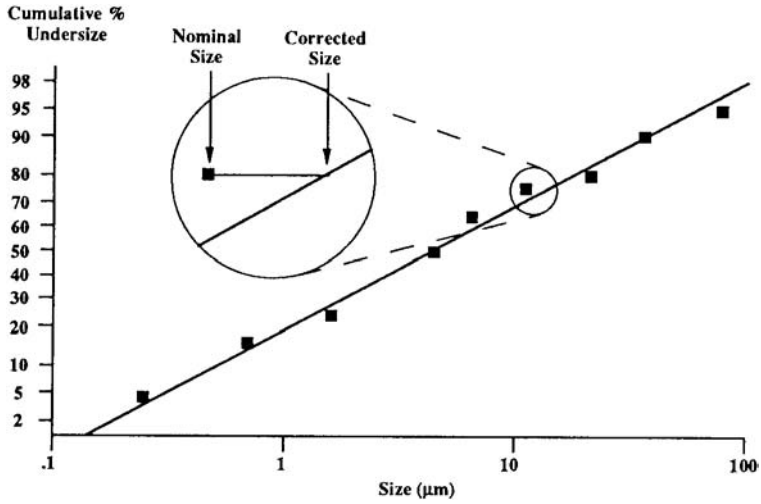


Figure 3.5. Correcting sieve sizes with a known lognormal distribution.

sample of sieved material. This sample consists of a very narrow distribution of particles which are just small enough to pass through the sieve. The size of this fraction can be determined either by hand counting and weighing, if the particles are large, or by the Coulter counter if they are too small to count by hand. In this way the volume diameter cutoff of the sieve may be estimated.

ERRORS AND PROBLEMS IN SIEVING

Errors in the sieve

Sieves are subject to many errors in manufacture which have been closely studied. Leschonski (1970) studied the range of orifice sizes in typical woven sieves and found a coefficient of variation of around 10% for the smaller sizes, falling to 3–5% for millimetre-sized sieves. This obviously sets a limit on the sharpness of cut which can be achieved. The errors are further compounded by damage and wear during use.

Sieve load

The mass of material placed in the sieve can have a considerable influence on the final result. Sieves work best when there is only a thin layer of well-separated material on the mesh; however, if there is very little material present, the error in its weighing is relatively large. A good starting point is 50–100 g for 100 mm sieves, and 200 g for 200 mm sieves.

Sieving time

The theory of sieving has been studied by many authors (e.g. Whitby 1958, Jansen and Glastonbury 1968) and a considerable degree of mathematical sophistication has been achieved. Several authors (e.g. Shergold, 1946) have investigated the time-dependence of the sieving process, and have found that most powders generally show a biphasic behaviour if the fraction of material passing through the sieve is plotted against time (Figure 3.6). The first phase is due to the passage of particles which are smaller than the sieve mesh in all dimensions, and so will pass through the mesh the first time they come into contact with it. The second phase is due to particles with two dimensions smaller than the sieve mesh, but with a larger third dimension. These particles need to be presented to the mesh in a reduced range of orientations, and so the probability of their passage is necessarily reduced.

As a result of this effect, particle shape influences sieving time. Spherical particles show a pronounced phase 1 and little phase 2, while needle-like crystals can be mainly phase 2. In this case sieving time can be very long, and more importantly, the end-point of sieving can be difficult to specify. If phase 2 is small, the phase 1/phase 2 transition can be taken as the end-point of the analysis; however, the changeover point is often ill-defined, and in addition, individual sieves in a nest will probably reach this point at different times. Consequently empirical endpoints are

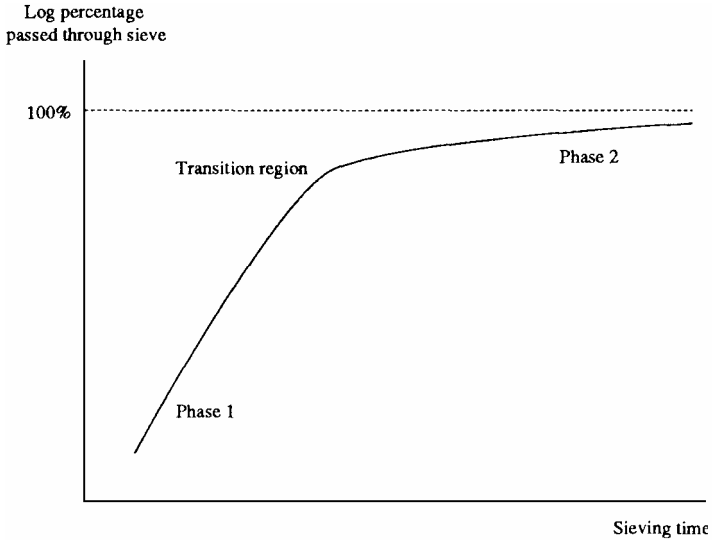


Figure 3.6. Effect of sieving time on mass passing through sieve.

usually selected, when the mass of the material in the sieves does not change after a significant time.

Properties of the feedstock

Many sieving errors are due to properties of the material. Many materials adhere to themselves or to the sieve mesh, and this can prolong sieving time considerably, and cause major losses. Sieving can often be improved in these cases by addition of a small amount of colloidal silica (e.g. Aerosil), but a better method is to sieve the powder in the wet state. Humidity can also cause powder adherence, and low-temperature drying may improve the behaviour of the powder.

Electrostatic charging can cause major difficulties, since it causes the powder to adhere to the sieves, although charged particles repel each other and so are usually well-dispersed. Charging is a problem in many powder handling applications, and flowing powders can act as Van der Graaf generators, building up thousands of volts in apparatus. All flowing powder equipment should have an efficient earth

system to prevent this, and earthing the nest of sieves may prove useful if static is troublesome. Wet sieving is a better method for powders which have noticeable electrical behaviour.

WHEN TO USE SIEVING

Sieving is a relatively straightforward technique, and should be considered whenever any of the following conditions, or a combination of them, hold:

1) It is required to produce a separated size fraction for further study.

2) The material flows easily as discrete particles.

3) The material is fairly coarse, e.g. above 100 μm, and contains few fines.

4) The powder is composed of materials with a range of different densities (which would cause difficulties for sedimentation techniques), different refractive indices (problems for light scattering) or is water-soluble or conductive (problems for Coulter counter).

5) The material must be handled dry, for example due to solubility in water or common solvents. Powders that have been granulated with a binder can also be sieved dry, as they will disintegrate on wetting.

6) The particles are large and dense, so that they settle too rapidly to be studied in suspension, for example, by light scattering.

Sieving is best avoided when:

1) Much fine material below 100 μm is present, unless specialized microsieving equipment is available.

2) The particles are fragile.

3) The particles are in the form of elongated needles.

4) The material adheres to the sieve or forms clumps.

5) The powder easily acquires an electrostatic charge.

Powders displaying these problems can be tackled using the techniques discussed, but they will make the task of sieve size analysis much more difficult, and often a more appropriate technique can be selected. The effect of variables such as sieve load and sieving time can make the

development of a rigorous test protocol rather involved, but it is essential to attend to such details if a reliable and accurate method is to be produced.

REFERENCES

Agricola (1556) *De re metallica.*

Colon F.J. (1970) *Proc. Soc. Analyt. Chem.* **7,** 163–164.

Daescher M.W. (1969) *Powder Technology* **2,** 349–355.

Jansen M.L. and Glastonbury J.R. (1968) *Powder Technology* **1,** 334– 343.

Konowalchuk H., Naylor A.G. and Kaye B.H. (1976) *Powder Technology* **13,** 97–101.

Leschonski K. (1970) Present and future states of particle size analysis. In: Groves M.J and Wyatt-Sargent J.L. (eds), *Proc. conf. particle size analysis,* Bradford; Society for Analytical Chemistry, London, 205–217.

Leschonski K. (1981) Sieve analysis: the Cinderella of particle size analysis methods? In: Groves M.J. (ed), *Proc. conf. particle size analysis,* Salford; Heyden, London, 205–217.

Peterson J.L. (1969) Method and apparatus for wet-sizing finely divided solid materials. U.S. pat. 3 438 490.

Shergold F.A. (1946) *Trans. Soc. Chem. Ind.* **65,** 245.

Von Rittinger P.R. (1867) *Aufbereit.,* no. 222 & 243.

Whitby K.T. (1958) *Symposium on particle size measurement,* ASTM publ. 234, 3–25.

4
Electrical zone sensing

The electrical zone sensing method, or 'Coulter counter' is a well-established technique for the size analysis of particles in suspension. It was originally devised for counting of blood cells (Coulter 1956), and later developed into one of the most popular and widely used size analysis techniques. It is now sufficiently well established to be used as a standard method in many ASTM specifications, and is the subject of a British Standard (BS3405 part 5, 1983). The original instrument (the Coulter model 'A') is now a much sought-after antique by those interested in particle technology, and has been replaced by a series of more advanced models. The best-known of these are the TA and TAII, which sort particles into 16 size bands. Currently the top-of-the-range instrument is the Multisizer, which provides up to 256 size bands and is under full computer control. The model ZM is essentially a particle counter, which can perform size analysis over a limited size range with the addition of a multichannel analyser ('Channelyser'); a range expander is also available. Despite the electronic sophistication, some users still prefer the earlier models!

PRINCIPLE OF OPERATION

The principle of the instrument is very simple and is illustrated in Figure 4.1. A pinhole, set into a glass tube, is immersed in a suspension of the particles, and a negative pressure applied to the inside of the tube. This causes the suspension of particles to be sucked through the pinhole. When a particle passes through, it partly closes the orifice,

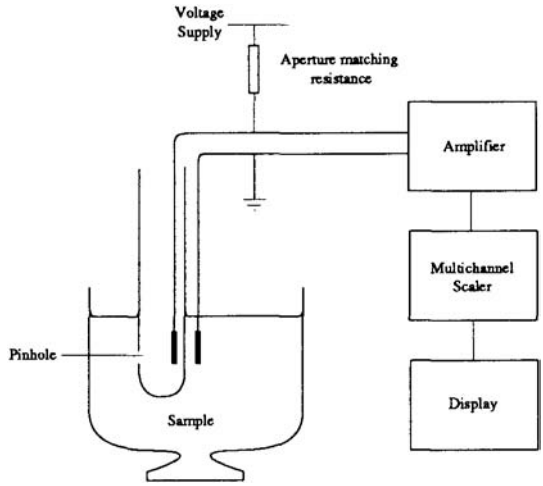


Figure 4.1. Principle of the Coulter counter.

and so the electrical resistance of the orifice rises. If a small particle passes through the hole, only a small resistance change is produced, but if a large particle is present, the hole is closed to a greater degree and a larger resistance change is produced. Consequently the size of the particle can be deduced from the change in resistance. The resistance is sensed by placing electrodes inside and outside the pinhole, so that a small current can flow between the electrodes, through the pinhole (this is the only path available). In order to allow the current to flow, the suspending medium must contain an electrolyte; this is normally a 0.9% solution of sodium chloride. The electrodes are connected to electronics which measure the resistance change and count the particle into an appropriate size band.

The central assumption behind the use of the Coulter counter is that the amplitude of the voltage pulses produced by the particles are proportional to the volume of the individual particles. This assumption has been addressed by a number of authors (e.g. De Blois and Bean 1970, Harfield and Knight 1981), and is generally held accurately for nonconducting spherical particles that are less than one

third of the aperture diameter. Since the particle volume is measured, it is a common misconception that the instrument provides volume or mass distributions; this is in fact not the case. Since the particles are counted, the fundamental measurement is one of number distribution, and if the instrument displays volume or mass distributions, these have been derived assuming spherical particles.

The instrument can be divided into two sections; the *sampling head,* which consists of the pinhole and apparatus for drawing the particle suspensions through it, and the *analyser,* which amplifies and sorts the electrical pulses to derive the size distribution.

Sampling head operation

A diagram of the sampling head is shown in Figure 4.2. It consists of the tube which contains the pinhole, and a manometer device which allows the pressure inside the tube to be reduced, thus sucking a metered volume of solution through the pinhole. Since the exact volume of solution passing through the pinhole is controlled by the manometer, the instrument can be used to count the absolute number of particles as well as derive their size distribution.

Suppose that initially no vacuum has been applied, and so the mercury level is the same on both sides of the manometer. In this position there is sufficient mercury to reach from the bottom of the bulb in the upper left-hand limb, past all the contacts, and up to the top of the right-hand limb. The remainder of the apparatus is filled with the counting electrolyte; its path extends from the fresh electrolyte reservoir, through the fill tap, filling the pinhole tube, through the count/reset tap to the vacuum pump, and from the top of the pinhole tube to the right-hand side of the manometer.

In normal operation, the fill tap is left closed; its only function is to enable the equipment to be filled with saline during the setting-up procedure, and to allow any gas

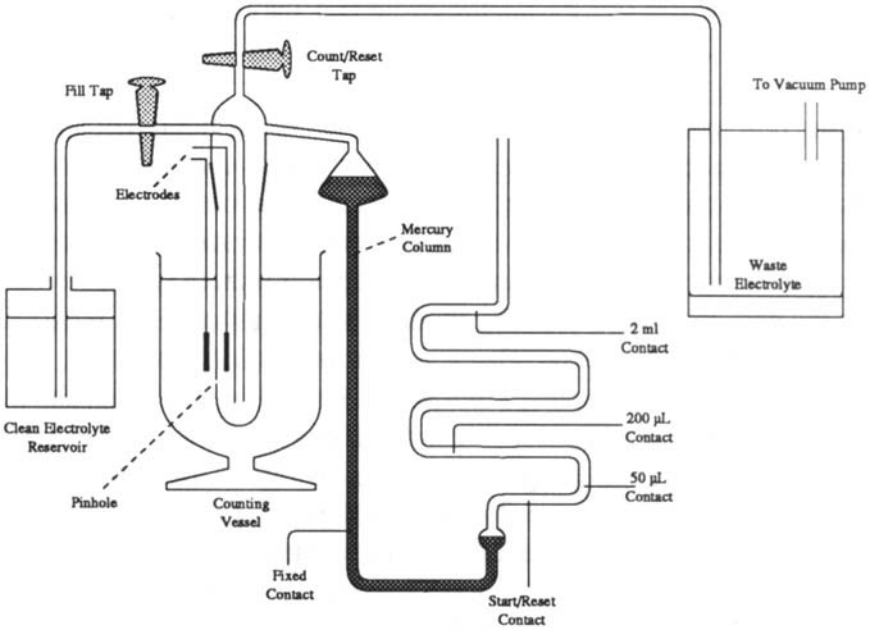


Figure 4.2. Coulter counter sampling apparatus.

bubbles which may have entered to be flushed out during use. Suppose now that the reset/count tap is opened; this connects the top of the pinhole tube to the vacuum pump, and causes the mercury to be sucked up into the bulb of the manometer. The thread of mercury in the right-hand limb will fall, uncovering all the counting contacts and the start/reset contact. The vacuum is adjusted, using a bleed valve on the pump, so that the mercury stops at this level. The sampling head is now in the 'reset' state. As the mercury thread passes the start/reset contact, the counting electronics are reset to zero.

If the reset/count tap is now closed, the mercury column will tend to fall to its original level. The only way it can do this is by drawing liquid into the pinhole tube from the sampling vessel through the pinhole. As it moves, the mercury first closes the start/reset contact, and the electronics then begin to count the pulses from the electrodes as the particles in the counting medium stream through the pinhole. The

mercury thread then closes the various volume contacts, and when the selected sampling volume contact is closed, the instrument stops counting. The mercury thread will continue to move until the manometer is balanced again. In order to repeat the process, the count/reset tap can be opened again, and the manometer will return to its unbalanced state, breaking all the contacts and resetting the electronics in the process. Successive measurements can be made simply by repeatedly closing and opening the count/reset tap. Alter natively, counting can be performed for either a fixed time or to a set total number of pulses.

During this process, a magnified image of the pinhole can be seen on a small viewing screen, and it should be confirmed that it is free of obstructions. Often, small air bubbles will obscure the pinhole; these are introduced by changing the sample when the manometer is unbalanced. They can be removed by opening both the count/reset and fill taps, when fresh electrolyte will flush the tube; this process can also be useful for removing large particles blocking the pinhole.

The electrodes are connected in series with a resistance (the *aperture matching resistance*), which is further connected to a voltage source. The external resistance is sufficiently large that the current I through the pinhole remains constant. Thus a change in aperture resistance ΔR leads to a change in voltage of $I\Delta R$. The resistance, and hence the aperture current, can be changed using a selector switch, and the manufacturers recommend appropriate values for the various apertures. The voltage pulses are then amplified before being passed to the multichannel scaler, which counts them into size bands according to their amplitude. The gain of the amplifier can be finely varied using the *size calibration adjustment control* (a calibrated multiturn potentiometer) to allow particles of a particular size to produce pulses of a desired amplitude; this is necessary for calibration of the instrument. In the latest instruments the amplifier gain is under software control and the setting of gain and current is automatic; the user need not worry about them for most purposes.

APERTURE CONSTRUCTION AND MAINTENANCE

The apertures are drilled in a small disc of synthetic sapphire mounted in the side of a glass tube. A range of apertures is available, to allow the measurement of particles of different sizes; normally an aperture can be used to measure particles from 2% to 30% of its diameter, and they are available in sizes of 15, 30, 50, 70, 100, 140, 200, 280, 400 and 560 μm. Recent work has suggested that apertures can be used with particles as large as 80% of the aperture diameter, (Harfield *et al.* 1984) but this is not common practice. The smallest apertures block in use very easily, and particularly with the 15 μm pinhole it can be useful to prepare the sample suspension, for example, by sedimentation or wet sieving, to remove any larger particles.

In practice the only cleaning that the apertures require will be the removal of blocking particles, which will be visible in the viewer. Often these can be removed simply by brushing with a sable brush, or by opening the fill tap with the count/reset tap closed. More firmly wedged particles can be removed by more vigorous brushing, or by chemical means if a suitable solvent is known. The final option is the ultrasonic bath, which should be used with caution, as the orifice may be damaged. If such vigorous treatment is used, the orifice should be recalibrated before further use.

CALIBRATION

Calibration of the orifices is normally effected using monosized particles such as polymer lattices (Chapter 2). The latex chosen for calibration should be between 5% and 20% of the aperture diameter.

The simplest procedure is to count a sample of the latex in electrolyte, and adjust the size calibration control until the peak due to the particles is split equally into two channels. The left-hand edge of the upper channel then corresponds to the number median diameter of the calibration latex. The multichannel scaler is set up in a geometric distribution, so that each channel left hand edge corresponds to a volume *twice* that of the left-hand edge of the channel immediately

below it. This corresponds to a diameter increment of a factor of 1.260 per channel. Thus, for example, if a latex of diameter 4.55 μm was split using channels 7 and 8, the left-hand edge of channel 8 would correspond to 4.55 μm; that of channel 9 to 4.55×1.26=5.73 μm, channel 10 to 4.55×1.26×1.26=7.22 μm, and so on. The diameters of the smaller channels can be found by dividing the upper channel diameters by 1.26. In general, the diameters and channel numbers are related by:

$$d(\mu m) = K\sqrt[3]{\frac{2^W}{A}}$$

where W is the channel number, A is the setting on the calibration control, and K is the *calibration constant* being derived for the pinhole under test. K is calculated for the known channel using the known latex, and the formula can then be used to derive the edge sizes of the remaining channels.

On some older instruments (e.g. the TA), the individual channel counts cannot be obtained, and the instrument displays the weight percent in each channel. A quick calibration method for such an instrument is to adjust the calibration control until the peak is split into two parts, with the upper containing twice the volume of the lower. Since each channel corresponds to particles with twice the volume of the previous channel, the two channels will then contain equal numbers of particles, and again the upper channel left-hand edge will correspond to the number median diameter of the latex. Alternatively, if the peak is split into two halves containing equal mass, then the upper channel left-hand edge will correspond to the mass median diameter of the latex. Both the number and mass median diameters of the latex are usually quoted by the supplier.

The sizes of the calibration materials are not normally under the exact control of the manufacturer. Thus, for example, it will not be possible to obtain a latex with a calibration diameter of 1.000 μm, or other round figure. However, it is highly convenient to calibrate the instrument so that the channels lie at particular sizes (a progression

based on a cube root of 2 from a base of 1 μm is popular). In order to do this, the following formula can be used to calculate a new setting for the size calibration control:

$$\frac{A_2}{A_1} = \left(\frac{d_1}{d_2}\right)^3 \cdot \left(\frac{K_2}{K_1}\right)^3 \cdot 2^{(W_2 - W_1)}$$

where A_2 is the required calibration control setting to adjust the left hand edge of channel W_2 to a diameter of d_2, when a calibration control setting of A_1 has been obtained for a calibration latex of diameter d_1 split by the left-hand edge of channel W_1. K_2 and K_1 are the pinhole calibration constants for the pinhole being adjusted and the pinhole in which the calibration latex was measured, respectively. These will normally be the same (i.e. the calibration constants for the pinhole being calibrated), but this form of the formula allows the calibration control setting to be calculated to 'line up' the calibration channels of different pinholes covering different size ranges; we will see the value of this when we consider the use of more than one pinhole to study samples with a wide range of diameters.

Calibration can also be performed using the test sample itself; the principle is to calculate from the counts the mass of material passing through the pinhole, and compare this with that added to the suspension. This is termed the *mass balance technique,* and is described more fully below. It requires no initial calibration of the instrument using another standard, and so is termed a *primary calibration;* this method is also described in BS 3406.5.

The calibration of the Multisizer is much simpler owing to the microprocessor in the instrument. The user selects a calibration latex and enters its number mode diameter from the data sheet into the appropriate instrument menu. A number differential distribution is then obtained, and the instrument's cursors placed symmetrically on either side of the peak due to the particles. The CAL button is then pressed and the instrument automatically recalibrates itself. Do not press the CAL button at any other time!

TECHNIQUES FOR THE USE OF THE COULTER COUNTER

The normal operating procedure for using the Coulter counter is described in the instrument manuals, but some practical details may prove useful here in designing an experimental procedure for a particular sample. We will assume that the reader has used the methods described in Chapter 2 to bring the sample into a well-dispersed suspension at a phase volume of around 1%. This suspension will be diluted into the counting electrolyte; the preliminary suspension need not be made in the electrolyte, since only a small volume will be added to the counting vessel, which will not significantly change the counting electrolyte concentration.

If only the size distribution of the sample is required, it is not essential to make an exact dilution of a known amount of particle suspension into a known amount of electrolyte, so a common technique is to add particle suspension until a fixed concentration index (usually 5–10 %) is reached. However, this does lose one of the main advantages of the instrument, which is to count the absolute number of particles of various sizes, and so a better technique is to work with at least approximately known volumes and dilutions. If a mark is engraved at a suitable level on the counting vessel (e.g. 150 ml) then it is easy to fill to the same volume with electrolyte. The pinhole is then immersed in the electrolyte, taking care to fully immerse the outer electrode, and the door of the counter closed. The counting vessel should always be removed with the main tap in the closed (count) position, to prevent air being continually sucked into the tube. Confirm at this point that (a) the concentration index meter reads zero, (b) that the electrolyte is visually free of suspended particles, which shine brightly in the light beam, (c) that the pinhole is clear of blockages and bubbles (check the microscope or pinhole viewer). The stirrer can be run slowly, but take care that it does not bang into the vessel as it rotates, or the resulting vibration will add to the low-channel background. Turn the tap to the

reset position; on the older instruments the concentration index needle will kick twice to indicate that the machine has reset; this can be confirmed by observing the mercury level. The tap is then turned to the count position, and the counting will commence until the selected volume of suspension has been drawn through the pinhole.

The background count so obtained should be fairly small, particularly with the larger pinholes. There should only be tens of particles in each channel, or at most hundreds in the smallest channels. It is important that there are few (preferably no) background particles in the upper channels, particularly if the weight distribution is being calculated, since a single particle in the top channel will have a volume equal to several thousand of the smallest particles.

Having recorded a suitably low background, the sample can now be added. When developing a new protocol, we normally use the following method to establish a suitable concentration, rather than relying on the concentration index. Sample is added in small aliquots until the effect of coincidences can be observed in the results. A 0– 500 μL adjustable pipette is useful for this; a 100 μL sample of the preliminary dispersion can be added to the counting vessel, and the effect on the concentration index meter observed. At this point we are aiming for a concentration level of about 1%, so if this is rapidly exceeded, the original particle suspension should be diluted further, the electrolyte replaced, and a new background recorded.

Assuming we have now reached a situation where a 100 μL sample produces approximately a 1% concentration index change, a number of counts can be taken, and the distributions recorded. A further sample is then added, and the counting repeated. This process is repeated until either (a) a 10% concentration level is recorded, or (b) the recommended maximum count rate for the particular pinhole is exceeded; this can be found by dividing the total count by the counting time.

A measure of central tendency, such as the number mean diameter, should then be obtained from each set of data; with the older instruments you will have to calculate this

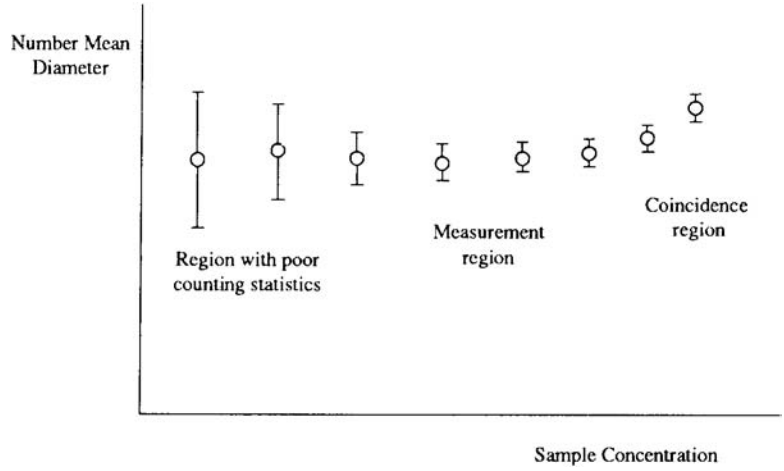


Figure 4.3. Typical number mean diameter vs. concentration of particles.

yourself, but the newer ones do it for you. The mean should be calculated from each count, so that several (preferably at least six) measurements are available at each dilution. The mean and standard deviation can then be found at each dilution; if these are plotted against dilution, a graph similar to Figure 4.3 will be obtained. It is evident that an optimum dilution exists; at low dilutions, too few counts are taken, so there is considerable random scatter in the data. Precision increases as the sample concentration is increased, but ultimately a point is reached at which coincidences begin to bias the results, increasing the mean diameter. We normally select the optimum dilution as a point between 50% and 70% of that at which coincidences start to occur, although if the samples to be measured show a wide variation in size, it may be prudent to leave more 'headroom' and work at lower concentrations. If the scatter in the measurements is too high at this dilution, a larger volume of suspension should be counted to increase the total counts. The theory of sampling errors in Coulter counter analysis has been examined by Lloyd *et al.* (1977).

Overlapping ranges

If the particle size distribution is very broad, it can be measured in sections using more than one tube. The main problem with this technique is that the larger particles will block the smaller tube. This can be prevented to some extent by passing the material through a microsieve, or by sedimentation, to remove the coarse particles, and prepare a sample for the smaller aperture tube.

The usual technique for overlapping range measurements relies on being able to line up the channel settings of the separate pinholes, so that both distribution measurements contain one (preferably more) channels defined by the same size limits. An example is shown in Table 4.1. Here, a 50 μm tube is calibrated so that its channel limits are 1.00, 1.26,... 32.0 μm, and a 200 μm tube is calibrated from 3.18 to 101. 72 μm. Both tubes would then have 11 common channels from 3.18 to 32 μm.

'Connecting up' the distributions is simply a matter of adding up the total counts for the overlap range for both tubes, and dividing one by the other to establish a 'scale factor'. The measurements from one tube (usually that containing the fewer counts) are then multiplied by the scale factor. The total counts in the whole distribution can then be obtained by adding up the scaled channel counts from both tubes (do not count the overlap channels twice!), and the whole distribution scaled to 100% using this count. As a final check that any preparation used has not biased the sample, the individual counts in the overlap channels of one tube, when multiplied by the scale factor, should be approximately equal to the counts in the same channels for the other tube. If this is not the case, it means that the sample distribution in the overlap region is different in the two tubes. This may arise if the sample has been fractionated in some way in this size range when preparing the smaller sample. A different technique of separating the large particles must then be used.

Table 4.1 Typical channel values for overlapping ranges

Tube 1	Tube 2
1.00	
1.26	
1.59	
2.00	
2.52	
3.18	3.18
4.00	4.00
5.04	5.04
6.35	6.35
8.00	8.00
10.09	10.09
12.71	12.71
16.00	16.00
20.16	20.18
25.40	25.42
32.00	32.00
	40.36
	50.85
	64.00
	80.73
	101.72

Mass balance checks

In order to check the accuracy of the analysis, it is often desirable to perform a mass balance calculation (Atkinson and Wilson 1981). This involves calculating the mass of material in the distribution, and comparing it to that added to the initial suspension. Mass balance calculations are most easily performed using a modern spreadsheet program such as 'Microsoft Excel'; the author can supply a copy of a 16-channel spreadsheet to anyone interested.

The raw counts are first averaged over any replicate samples taken, so that mean counts per size channel are obtained. The volume of a particle in each size band is next calculated, assuming a diameter of the channel centre, and spherical particles. Multiplying these together gives the volume counted in each channel, which can be multiplied by the material density and summed over the channels to obtain the total material mass which passed through the aperture. This should be equal to the mass originally taken, divided by the various dilution factors used in preparation and sampling. For example, suppose 0.1 g of material was suspended in 100 ml of electrolyte, and 0.1 ml of this added to 100 ml of electrolyte, of which 0.5 ml was counted. The mass passing through the pinhole should then be $0.1\times(0.1/100)\times(0.5/100) = 5\times10^{-7}$ g. If this does not agree with that found from the counts, the reason usually given is that some material lies below the pinhole range, and this can be quoted as a sub-measurement range fraction. Alternatively, if the operator is confident that all the material lies within the counting range of the instrument, the mass balance method can be used for calibration of the pinhole. This method has the advantage that the instrument is calibrated with the sample itself, and not with another material of different properties or shape, such as polystyrene latex. The mass balance technique is also a potentially absolute method, since it does not rely on a previous measurement of a sample using a different instrument.

It should be remembered that mass balance calculations will only be in agreement if any edit facility (q.v.), which rejects distorted pulses, is turned off, and low coincidence levels are ensured by working at low dilution (see section: Errors in electrical zone sensing, below).

Electrolyte selection and preparation

An extremely wide range of electrolytes can be used as dispersing media. The most usual solution is 0.9% saline, which will cover most requirements. A small quantity of bactericide (e.g formalin 0.5% or sodium azide 0.1%) can be

added to prevent microbial growth. The electrolyte should be passed through a 0.45 μm filter (a small pressure vessel is useful for this purpose) or a 0.22μm filter when the smallest aperture tubes are used. Even the commercial electrolytes ('Isoton') benefit from filtration. Electrolytes should be filtered immediately prior to use, as they will accumulate particles if allowed to stand for a few hours. The choice of 0. 9% as electrolyte concentration is not completely random; the instrument was originally designed for blood cell counting, so an isotonic medium must be used.

For specialized purposes this basic recipe can be adjusted in a number of ways. If very large particles are being studied with the largest orifice tubes, glycerol or sugar can be added to increase the viscosity of the medium to prevent sedimentation. However, this will increase the sampling flow time, which may partly offset any advantage gained. If very small particles are measured using the smallest apertures, and noise is a problem, a stronger electrolyte can be used (e.g. 2% or 5% saline). This will decrease the pinhole resistance, and improve the signal to noise ratio; it may be necessary to adjust the aperture matching resistance. It is important to check for flocculation if stronger electrolytes are used. Flocculation may even be a problem in some systems when using 1% saline, and in this case a different electrolyte can be tried, for example 2%–4% sodium phosphate; alternatively a surface-charging ionic surfactant such as sodium lauryl sulphate can be added to the electrolyte.

Many drugs are at least slightly water-soluble, which can lead to difficulty when sizing using aqueous electrolytes. Solubility generally manifests itself as a gradual loss of counts, and decreasing mean size. Even low solubility materials can dissolve sufficiently to show measurable changes in size distribution with time. There are three approaches to this problem:

a) Saturate the electrolyte with the sample material before use. This technique is satisfactory only if the sample solubility is low. The main disadvantage is that if the temperature rises, the sample may dissolve further, or more

sample may precipitate on the remaining particles if the temperature falls. This method can in fact be used to measure dissolution rates, as described by a number of authors (e.g Nystroem *et al.* 1985, Nystroem and Bisrat 1986)

b) Use the common-ion effect to depress the solubility of the sample; for example, a slightly soluble sulphate could be dispersed in sodium sulphate electrolyte.

c) Use a non-aqueous solvent system. This is the best option if water solubility is significant. A solvent of very low dielectric constant (e.g. hexane) cannot be used, since it will prove impossible to find an electrolyte which dissolves in it. Mixtures recommended by Coulter include 5% ammonium thiocyanate or 5% lithium chloride in methanol, acetone, or isopropanol. Some experimentation may be required to find a suitable system; addition of dimethyl sulphoxide can increase the solubility of the electrolyte, but it can also increase the solubility of the sample!

It should be noted that a change in the electrolyte will almost always require a recalibration of the instrument, since the pinhole resistance, and hence the pulse height, will no longer be the same as in the usual saline electrolyte. Since changing to nonaqueous electrolyte and back again involves cleaning of much of the glassware, it may prove cost-effective to use a second sampling stand specifically for nonaqueous systems. A two-stand switching unit is available to allow two stands to be connected to a single analyser.

ERRORS IN **ELECTRICAL ZONE SENSING**

Interference and vibration

One of the most significant problems with the zone sensing method is that small signals at the electrodes can arise for reasons other than particles passing through the pinholes. If the instrument is mounted on a bench which is subject to vibration, this may transiently change the distance between the electrodes, and this will appear as small resistance

fluctuations. Consequently the sampling head of the instrument should be mounted on a firm surface, such as a solid wooden bench. If vibration is a problem, and the instrument cannot be re-sited, a variety of antivibration mounts can be used; these vary from improvised devices such as a paving stone sitting on an inflated tyre inner-tube (cheap with average performance but usually satisfactory) to electronic vibration isolation platforms such as Newport's EVIS system (expensive, but would work through an earthquake). The biggest source of vibration is however usually the operator, (some operators are better than others in this respect) and so a hands-off procedure should be followed during counting.

Electrical interference is also a problem since the electrodes are rather sensitive to electromagnetic pickup, acting as small aerials. To reduce this problem, the sampling head is housed inside a screened box which is earthed (a Faraday cage). The door of the box has a conducting mesh on the surface to screen interference and allow the operator to see in. Earlier versions used a small steel box around the pinhole and counting vessel alone. In electrically noisy environments these precautions may not be sufficient, and some operators have resorted to electrically shielded rooms to house their instruments. This is normally not necessary, but it can be useful to place a mains interference filter in the power supply line; there is a wide range of these power supply conditioners available, mainly for the computer market, but they are equally useful for analytical instruments. Interference can also be airborne in origin; for example, we have just had to reorganize our particle sizing laboratory on receipt of a new Coulter Multisizer, in order to separate it from the PCS laser, which was causing excessive RF interference. This could be a significant problem in many sizing laboratories now that laser-based light scattering equipment is widely used.

A particularly troublesome form of interference in medical environments can be caused by inductive loop paging systems. The loop aerials should be located and the instrument sited as far as possible from them.

The effect of both vibration and electrical interference is to produce small voltage fluctuations which appear as excess counts in the lowest channels of the counter. In noisy environments this may make channel 1 unusable; if this is a problem, a smaller pinhole could be used so that the particles of interest can be moved to higher channels. It can sometimes be difficult to distinguish between interference in channel 1 and background particles in the electrolyte. In this case an experiment can be performed at night when there is less electrical noise; if this reduces the background count, then the screening should be improved.

Coincidence correction

If two particles pass through the pinhole at the same time, they will give rise to a single pulse with an amplitude larger than that which would be produced by either particle alone. This will produce a distorted size distribution. A spurious pulse of this type is called a *coincidence,* and the rate at which coincidences occur will increase as the particle concentration in the electrolyte is increased.

Coincidences fall into two types:

Type 1 coincidences, in which two particles sufficiently large to be sized by the pinhole in use are counted as a single pulse into a higher channel,

Type 2 coincidences, in which two particles below the size limit of the pinhole add to produce a pulse which is sufficiently large to be counted.

We find it useful to distinguish a third type of coincidence, which could be called type 2a or 3. In this case a particle within the pinhole range passes through at the same time as a particle below the pinhole range, thus causing the large particle to be counted into a still larger channel. This situation is particularly important in the study of parenteral nutrition emulsions, in which the Coulter counter is used to measure the distribution of the few large (> 1 μm) droplets in the presence of a vastly larger number of submicrometre droplets (Washington 1990). In these conditions we have found the normal recommendation,

to operate at a sample level leading to a 5% concentration index, to be highly misleading, and much lower sample concentrations and longer counting times are needed.

The theory of coincidence errors has been investigated by a number of authors (e.g. Wales and Wilson 1961, 1962; Edmundson 1966). The probability of more than one particle passing through the pinhole simultaneously is described by a Poisson distribution. It can be shown that the observed total count N' is always less than the true count, and is given by:

$$N' = \frac{v}{s}\left[1-\exp\left(\frac{sN}{v}\right)\right]$$

where N is the true count, s is the volume of the sensing zone, approximately equal to the pinhole volume, and v is the volume of electrolyte counted. This can be simplified to the empirical equation suggested by Coulter:

$$N = N' + 2.5 \times 10^{-9}\frac{D^3}{v} N'^2$$

where D is the aperture diameter in μm, and the counted volume v is in microlitres. It is important to note that the effect of coincidence only applies to the total count, and this expression cannot be used to correct the count in a particular channel. In general the effects of coincidences on the distribution shape will be to move the mean to a higher diameter and generally skew everything to larger sizes; an exact correction of the distribution shape is difficult and the situation is best avoided.

In order to correct some of the errors introduced by coincidences, most counters have an *edit system.* Confusion may arise at this point because this term is applied to different features on different instruments. In early counters, such as the TA, the edit system was an array of potentiometers which allowed the size distribution to be displayed after a preset number of counts were subtracted from each channel, thus allowing for the subtraction of a background distribution. In later instruments, the term is used to refer to an electronic circuit (usually called an *edit board*) which measures details of pulse shape (e.g. pulse width), and rejects mis-shapen pulses which are most likely

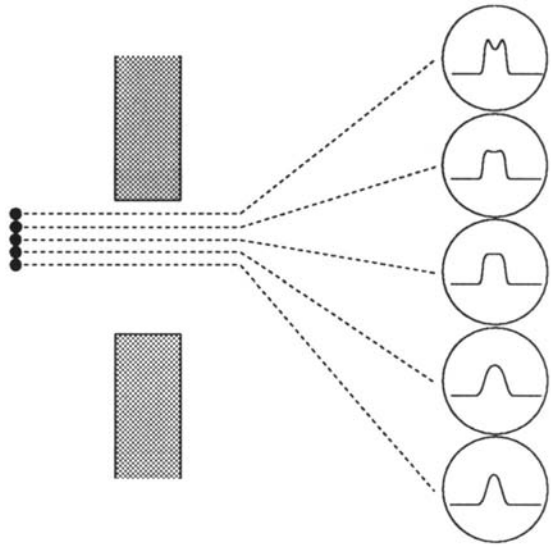


Figure 4.4. Dependence of pulse shape on particle path in the orifice.

due to coincidences. This option can be switched on or off as desired. The edit system usually rejects between 25 and 50% of the total pulses, and so it is important to turn it off if absolute counts are required.

Particle path and shape

The exact shape of the electrical pulse produced by the particle depends on the geometry of the pinhole and the exact path which the particle takes in passing through it. Early workers (e.g Kubitscheck 1960) suggested that the voltage pulses from the instrument should be flat-topped, since the particle should represent a fixed resistance for the duration of its passage through the pinhole. This was not observed experimentally, and more sophisticated studies of the electrodynamics and current flow in the orifice (Grover 1969, von Thom 1971, 1969) demonstrated that the pulse would be rounded if the particle passed through the orifice centre, but would develop an 'M' shape for paths near the orifice edge (Figure 4.4).

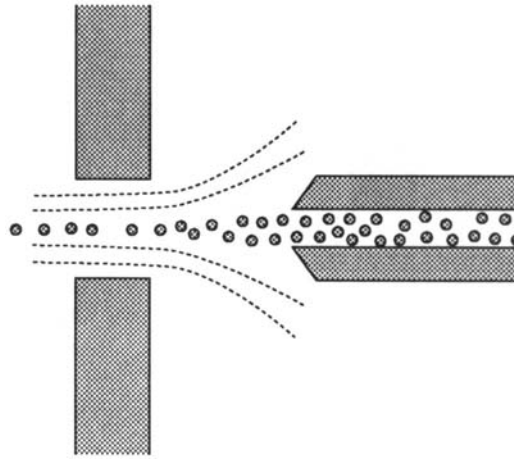


Figure 4.5. Hydrodynamic focusing of particles in the convergent stream entering the orifice.

A higher size resolution appeared possible if pulses of a particular path could be selected, and a number of approaches have been used in an attempt to realize this. They include conical and shaped orifices, which improve resolution but block very easily, and hydrodynamic focusing, which uses a convergent fluid flow to direct the particle through the orifice centre (Figure 4.5). Hydrodynamic focusing is potentially a very attractive technique, since as the particle stream is compressed in the convergent flow, its velocity increases, and so the orifice transit is more rapid. This requires faster electronics, but allows count rates to be increased without increasing coincidence effects. Although Coulter have produced research instruments based on this principle, the preferred solution at present, at least in routine instruments, is the edit board, an electronic filter which rejects misshapen pulses, including coincidences (q.v.), many of which arise from non-axial passage. The advantage of the more complex hydrodynamic scheme over the electronic solution appears to be minimal (Harfield 1987) Hydrodynamic focusing instruments are produced by Particle Data Ltd, under the tradename ‘Elzone’.

The theoretical derivation of the electrical response to the particle depends on a knowledge of the particle geometry. For spherical particles, the response depends fairly linearly on the particle volume, but this does not necessarily hold for nonspherical particles. Studies of porous and fractal particles (Lloyd 1981, Allen 1967) have suggested that electrical effects around the particle lead to the measurement of a spherical 'envelope', which averages out the surface irregularities.

Flocculation in the electrolyte

The electrolyte used for counting is approximately 0.17 M, a value which is above the critical flocculation concentration for many colloids. This may lead to slow aggregation of the diluted sample, and so the measurement should be taken soon after dilution. This potential problem can easily be checked by counting a particular sample every 15 minutes or so, to ensure that the mean size has a constant value. Fortunately, aggregation is not often encountered as a problem, since the sample is so dilute that interparticle collisions are infrequent.

Particle sedimentation

The suspension is normally kept stirred in order to prevent the larger particles settling at the bottom of the sampling vessel. This stirring cannot be too vigorous or it will induce turbulent noise in the pinhole, which increases the background count in the lower channels. At high stirring rates, vortexing may occur, introducing air bubbles, which are counted as large particles.

Sedimentation of particles is not normally a problem until the diameter exceeds around 100 μm or thereabouts, depending on the density of the material. Very dense materials, such as metal powders, may sediment appreciably even at smaller particle sizes, and it is usual to add glycerol to the electrolyte to increase the viscosity for these systems. To reduce this problem, Harfield *et al.* (1977)

described a modified beaker with a baffle to improve the mixing. This allowed glass beads up to 100 μm in diameter to be suspended without significant settling, and could be used with more viscous solutions for particles up to 350 μm. The same authors described a novel measurement head for the study of millimetre-sized samples, but this is not commercially available.

APPLICATIONS

Coulter maintain a list of currently over 1500 references to studies which have used their instruments, and a copy of this is available from them on request. Four hundred of these references are separately listed as 'pharmaceutical'; these include such diverse studies as self-emulsifying systems, microparticles, microcapsules, dissolving and swelling rate studies, liposomes, etc. For example, Ismail *et al.* (1984) studied the swelling of gelatin-acacia microcapsules and ethylcellulose capsules with the Coulter counter. The analysis illustrated not only the swelling of the capsules, but also their aggregation and ultimate degradation could be detected.

Two major areas of application are participate contamination in parenterals, and the stability of parenteral fat emulsions. These studies illustrate a number of aspects of the use of particle counting instruments.

Particle counting in parenterals

Zone sensing instruments are often employed for counting and sizing of particles in parenteral fluids. The requirement is for a minimum of particulates, and the pharmacopeias specify limits for particulate matter in injectable fluids. These are normally of a very low level (typically no particles above 5 μm), and so an instrument which can count individual particles is essential for this task. The particle size range of interest is 0.5–10 μm, so a 30 or 50 μm orifice is suitable. Since the number of particles is likely to be very small, a significant volume of fluid must be sampled. In

general it will not be sufficient, for example, to dilute 1 ml of the parenteral into 100 ml of electrolyte and count a 50 µl sample! Two methods can be used to improve the counting statistics. The first is to use the parenteral solution itself as the electrolyte. If the conductivity is too low (e.g. dextrose or water for injection) then a small volume of filtered 20% saline can be added to increase it to a suitable level. Alternatively if the conductivity is too high, it can be diluted with filtered water. The pinhole should be recalibrated by adding a calibration latex to the resulting solution, since the calibration will probably have moved slightly even if the conductivity is accurately matched to that of 1% saline.

Small volume parenterals are often supplied in glass ampoules, and it is surprising to what extent small glass particles can be introduced to the solution on opening the vial. Alexander and Veltman (1985) described a method of flame-opening the vials which did not contaminate the solutions with particles. A small flame, such as that from an ampoule sealer, is directed at the neck of the vial until the glass melts and the internal pressure causes a bubble to form. This can be burnt through and heated to enlarge the hole until the top of the vial can be flipped back.

A further useful technique for the detection of small numbers of particles is to count larger volumes than 2 ml by making repeated measurements. This can be done by adding up the separate measurements, but an easier way is to use the *reset inhibit* switch on the instrument, which stops the counter being reset when the mercury thread flows back, and so adds the new count to the previous one. If the sample volume is set to 2 ml and the count stopcock opened and closed 10 times in succession, 20 ml will have passed through the pinhole in total, and its total count will have been accumulated. The Multisizer does not have a reset inhibit since counts are automatically accumulated until the front panel 'reset' is pressed to clear the counters.

Particles can also be counted in parenterals using light-blockage instruments such as the Hiac-Royco particle counter. This instrument flows the solution through a light beam and measures the reduction in transmitted light

intensity when a particle traverses the beam, and can also measure the intensity of light scattered out of the beam. In principle the instruments are similar to the aerosol particle counters described in Chapter 8. The major problem with this type of instrument was highlighted by Haines-Nutt and Munton (1984). As we will discuss in Chapter 6, the intensity of light scattered from a particle depends on the difference in refractive index between the particle and its surroundings. Thus a particle suspended in water scatters much more light than a particle suspended in a strong sugar solution. Consequently, when Haines-Nutt and Munton studied the particulate levels in large volume parenterals using both the Coulter counter and the Hiac counter, good correlation was obtained for the dilute solutions, but those containing large volumes of dissolved sugars showed much poorer agreement.

Parenteral fat emulsions

Injectable emulsions are a special class of parenteral fluids which contain a large number of particles, with very tight constraints on their size distribution. Commercial fat emulsions such as Intralipid (Kabi-Vitrum) have a mean diameter of 200–350 nm, and a distribution sufficiently narrow that only a tiny fraction of the total oil is present as droplets larger than 1 μm. Measurement of the fraction of droplets larger than 1 μm is necessary (a) for quality control purposes during manufacture, and (b) for stability testing of the nutrition mixtures into which the emulsion is compounded. This area has been studied extensively by ourselves (Washington 1990) and others (Sherwood 1987a, b)

Coulter counter measurements are very useful in this system, but it must be remembered that even with the smallest aperture, much of the droplet distribution will remain below the measuring range of the instrument. The protocol currently used at Nottingham is to count the oil droplets larger than 1 μm using a 70 μm tube with channel 1 calibrated at 1.00 μm, or a 100 μm tube with channel 1 at

2 μm. We dilute the emulsion or TPN mixture by a factor of 10 and then add 0.1 ml of this dilution to 150 ml of electrolyte. It is particularly important to check for type 3 coincidences, since there are several orders of magnitude more droplets below the range of the instrument than within the aperture range. We normally find that satisfactory measurements are only obtained below 3% concentration index. When the droplet counts have been obtained, these can be multiplied by the dilution factors to find the number concentrations of the various sizes in the original sample, and, knowing the density of the oil, the fraction of mass above 1 or 2 μm can be calculated.

When doing an experiment of this type, in which colloid stability is studied by taking a fine cut from the tail of the size distribution, it is extremely important to maintain the instrument calibration to a high degree of accuracy, particularly if the study is being performed over an extended period of weeks or months. This is because the number population of the distribution will be changing rapidly with diameter near its tail, and so a tiny drift in calibration may admit or exclude many more droplets from the oversize count. This will correspondingly introduce a large random error into the results.

REFERENCES

Alexander D.M. and Veltman A.M. (1985) *J. Pharm. Pharmacol.* **37** 13–16.

Allen T. (1967) *Proc. Conf. Particle Size Analysis,* Loughborough, Soc. Anal. Chem.. London.

Atkinson C.M.L. and Wilson R. (1981) The mass integration method for the calibration of the electrical sensing zone technique used for the sizing and counting of fine particles. In: Stanley-Wood N.G. and Allen T. (eds), *Proc. particle size analysis conference,* Loughborough, John Wiley, Chichester, 185–197.

Coulter W.H. (1956) *Proc. Nat. Electronic Conf.* 12, 1034.

De Blois R.W. and Bean C.P. (1970) *Rev. Scient. Intrum.* **41,** 909.

Edmundson I.C. (1966) *Nature* **212,** 1450.

Grover N.B. (1969) *Biophys. J.* **9**, 1398.

Haines-Nutt R.F. and Munton T.J. (1984) *J. Pharm. Pharmacol.* **36** 534–536.

Harfield J.G., Miller B., Lines R.W and Godin T. (1977) Large particle measurement with the Coulter counter. In: Groves M.J. (ed), *Proc. conf. particle size analysis,* Salford; Heyden, London, 378– 394.

Harfield J.G and Knight P. (1981) Experimental evidence of the linear response of the Coulter counter. In: Stanley-Wood N.G. and Allen T. (eds), *Proc. particle size analysis conference,* Loughborough, John Wiley, Chichester, 151–164.

Harfield J.G (1987) *Curr. Aware. Part. Technol.* **20**, 211.

Harfield J.G, Wharton R.T and Lines R.W. (1984) *Particle Characterization* **1** 32.

Ismail N., Harris M.S. and Nixon J.R. (1984) *J. Microencapsulation* **1** 9–19.

Kubitscheck H.E. (1960) *J. Res. Nat. Bur. Stand.* A **13**, 128.

Lloyd P.J. (1981) Response of the electrical sensing zone method to non-spherical particles. In: Stanley-Wood N.G. and Allen T. (eds), *Proc. particle size analysis conference,* Loughborough, John Wiley, Chichester, 199–208.

Lloyd P.J., Stenhouse J.I.T. and Buxton R.E. (1977) Comparison of the errors due to sampling and to analysis by Coulter counter. In: Groves M.J. (ed), *Proc. conf. particle size analysis,* Salford; Heyden, London, 367–377.

Nystroem C., Majur J., Barnett M.I. and Glazer M.J. (1985) *J. Pharm. Pharmacol.* 37 217–221.

Nystroem C. and Bisrat M.J. (1986) *J. Pharm. Pharmacol.* **38** 420– 425.

Sherwood J.D. (1987a) *Anal. Proc.* **24** 277–279.

Sherwood J.D. (1987b) *Anal. Proc.* **24** 279–281.

Von. Thom R. (1971) German Patent 1 955 094.

Von Thom R., Hampe A. and Sauerbrey G. (1969) *Z. Ges. Exp. Med.,* **151,** 331.

Wales M. and Wilson J.N. (1961) *Rev. Scient. Intrum.* **32,** 1132.

Wales M. and Wilson J.N. (1962) *Rev. Scient. Intrum.* **33,** 575.

Washington C. (1990) *Int. J. Pharmaceut.* **66** 1–21.

5
Introduction to light scattering

Light scattering techniques have become both a boon and a curse to the particle sizing community. Their rapid increase in popularity over the last ten years has largely been due to the ease of use of many of the instruments, which are regarded by a large number of their users as 'black boxes', Oddly, the less intelligible the contents of the box, the more faith the users seem to have in their results!

Light scattering has been used for many decades to measure the size of small particulates, molecular weights and diffusion coefficients, and generally to obtain an enormous amount of information about macromolecular and particle systems. However, it is only in the last decade that the technique has become sufficiently straightforward to be used by non-specialists. This is largely owing to the use of lasers as light sources, and small but powerful desktop computers for the complex data reduction required.

Owing to the 'black box plus computer' appearance of many of these instruments, a great deal of confusion has arisen about what the instruments actually do; this situation is exacerbated by some of the less reputable manufacturers, whose sales pitch benefits from the user's lack of understanding. It is commonplace to see phrases in the (reputable) scientific literature such as "The particle size was measured by (insert name of favourite manufacturer) laser particle size analyser". The user is then only able to evaluate the results if he or she has a knowledge of the machine concerned.

Much of the confusion arises because there are at least three separate physical principles which can be used for the

measurement of particle size from light scattering. It is useful to briefly review these prior to the next two chapters, which treat them in depth.

LIGHT DIFFRACTION

If small particles are illuminated by a beam of light, they will scatter it in all directions. The intensity scattered in any particular direction is calculable using a number of approximate theories, or the exact theory ('Mie theory'), and depends on the size of the particles and their optical properties. This scattering is more precisely referred to as diffraction of light by the particles. Like other diffraction techniques, if the scattered light intensity is measured as a function of the angle through which it was scattered, the resulting *scattering pattern* contains information regarding the particle size, which can then be found by appropriate (complex!) calculations. The method is analogous to the measurement of the size of atomic structures (e.g. crystal lattices) by the diffraction of X-rays. We should note that, since the light beam illuminates many particles simultaneously, the instrument provides a distribution of particle sizes, but does not accurately measure the number or amount of particles. The operating size range of this technique is 1 μm to a fraction of a millimetre. Manufacturers often use complex calculations or optics to reduce the lower size range of the instrument (to about 0.1 μm), but this introduces a wide range of problems.

The main difficulty with this approach is the complexity of the calculations (which are often approximate or use iterative methods). The use of different algorithms by different manufacturers leads to differences in the results from different instruments using the same sample. Since the detailed algorithms are rarely explainable to the potential purchaser, all sales engineers insist that their software is superior to everyone else's. As we will see in the next chapter, the software can be the source of many subtle distortions of the size distribution. Our inability to standardize such factors is a major problem. Despite this

the instruments are very popular, and extremely useful if their limitations are borne in mind.

SINGLE PARTICLE SCATTERING

The size of the particle determines not only the scattering pattern, i.e. the angular distribution of scattered light, but also its intensity. Broadly speaking, the bigger the particle, the more light it scatters. Single particle sizing instruments measure this intensity at a fixed angle as single particles are passed through a light beam. The amplitude of the scattered light pulse is used to determine the size of each particle. These instruments are more accurately called single particle optical counters, and their main use is to count particles rather than size them precisely. They classify particles into rather broad size bands, since the exact size of the particle can be difficult to calculate from the scattered light intensity. The best-known of these instruments is the Hiac-Royco particle counter. Often the instruments simultaneously use the *light blockage* technique to count the larger particles, i.e. they measure the reduction in transmitted light when large particles pass through the light beam.

PHOTON CORRELATION

Photon correlation is a powerful technique for measuring the size of very small particles, from as little as 10 nm up to a few micrometres. It operates by measuring the diffusion coefficient of the particles and calculating their equivalent spherical diameter. It does this by measuring the fluctuations in scattered light intensity at a fixed angle; these fluctuations are due to diffusion of the particles in and out of the measuring volume by Brownian movement. Since large particles diffuse more slowly than small ones, the fluctuations in scattered light intensity from large particles occur over a longer timescale than the fluctuations induced by small particles. By measuring the statistical properties of the fluctuations in the scattered light (the so-called

correlation function or *noise spectrum*), the size of the particles can be determined.

Photon correlation suffers from similar difficulties to laser diffraction, in that the extraction of size distributions from the data is problematic. There are numerous software packages available for data reduction, and again the faith of the manufacturers in their individual approach to the exclusion of all others is notable.

Correlation techniques can also be used to measure the velocity of particles by the Doppler effect. If the particle is moving toward or away from the detector, the light scattered from it will be shifted in frequency by a tiny amount, and this shift can be extracted from the correlation function. Three types of instruments have arisen from this ability to measure particle velocity:

a) Electrophoretic mobility instruments. If a charged particle is placed in in electric field, its drift velocity can be measured, and hence its surface charge can be found. Examples of instruments using this approach are the Malvern Zetasizers and the Coulter Delsa. Normally the scattering is measured from many particles simultaneously, so a distribution of velocities is derived.

b) Laser Doppler velocimeters (LDV's). These instruments also measure the particle velocity in a moving mass of particles by the Doppler effect. However, they are often applied to the study of large particles in gas flows (e.g. aerosol droplets in air), which scatter a large amount of light. This enables the scattering and velocity of individual particles to be found as they pass through the beam.

c) Phase Doppler instruments. This is an extension of the LDV approach in which more than one detector is used to measure the scattered light intensity. The size of the scattering particle can be extracted from the relative phases of the signals in the two detectors. The instruments are thus capable of measuring the size and velocity of single particles in complex gas flows, an extremely powerful technique.

With such a wide range of instruments available, it is hardly surprising that new purchasers have been known to make the wrong choice, and end up with an instrument that

will not give them the data they require. Probably the best course of action, if you are considering the purchase of light scattering equipment, is to ask the manufacturer for the name of someone who has bought the instrument who could give you an impartial review. If they are unwilling to do this, look elsewhere.

6
The angular distribution of scattered light

INTRODUCTION

It has long been recognized that the scattering of light from small particles could provide information concerning their size and shape. The extraction of this information is mathematically complex, and the development of methods of size analysis based on light scattering has only become practical with the advent of two technologies; these are the use of lasers as light sources, and powerful microcomputers which make the calculations transparent to the user. The application of these techniques has made light scattering one of the most popular particle sizing techniques, and it is replacing many established methods.

The theory of the scattering of light by small particles has been developed by many authors, including Rayleigh, Mie, Fraunhofer, Gans, and Debye, and the fundamentals are treated in some detail in standard texts, probably the best of which is that by Bohren and Huffman (1983); the older works by Kerker (1969) and Van de Hulst (1957) are also valuable. The 'complete' theory describing the angular distribution of light scattered from a particle of arbitrary size is normally referred to as the Mie theory, although it is certain that many other workers contributed to its development. Unfortunately the theory is mathematically complex, and so many restrictions and simplifications to it are often made. Even with this theory, the exact angular distribution of scattered light can only be calculated for particles of very simple shape, such as spheres, concentric

spherical shells, spheroids, and infinitely long cylinders. To calculate the exact scattering pattern from, for example, a cubic particle, even in a single orientation, is an almost hopelessly complex task. Consequently, almost all practical light scattering methods assume that the particles are spheres, and provide corresponding equivalent diameters.

THEORY OF LIGHT DIFFRACTION

Optical properties of materials

In order to understand the scattering of light from small particles, it is helpful to have an understanding of their optical properties in the bulk phase. As the light enters the particle, it can undergo two processes; it can be *absorbed,* or it can be *refracted* internally. Generally a combination of both of these processes occurs. These effects are quantified by the *relative refractive index* of the material, usually given the symbol M. The relative refractive index is defined as the refractive index of the particle, N, divided by the refractive index of the surrounding medium, N' (usually water or air).

The refractive index, for most common organic materials, is a well-defined quantity which can be obtained from reference texts. However, the situation is more complex with absorbing or reflective materials such as metal oxides or metallic powders (e.g. those used in some cosmetics). To allow for the absorption and reflection, it is necessary to define the refractive index as a complex quantity, that is, a quantity having a real and imaginary part. The simplest interpretation of the complex refractive index is that the real part describes the familiar refractive properties of the material, while the imaginary part describes the light absorbing properties. We should not become confused between light absorbed by the particles and light that is apparently 'absorbed' in the bulk sample due to its being scattered out of the incident beam. If we place a particulate suspension in a conventional UV/VIS spectrometer, the instrument will indicate that the sample has an 'absorbance' since part of the light has been scattered out of the

acceptance angle of the detector. Similarly, if we look at a light source through a particle suspension, it will appear dimmed since the suspension scatters its light away from our eyes. To achieve this, the particles need not absorb light, they merely scatter it away from the light detector in both cases.

If the particles were made from a material which was completely transparent in the bulk (e.g. NBS glass beads, or a fat emulsion), then the refractive index would consist of only a real part, and the imaginary part would be zero. Alternatively, a material which was coloured (i.e. absorbed light) in the bulk would have a refractive index with both a real part (describing its refractive properties) and an imaginary part, (describing its absorbing properties). Both of these quantities influence the light scattering, and for many materials which cannot be prepared in large crystals, light scattering from particulates provides the only way of investigating their optical properties. Unfortunately data on the complex refractive index of many systems is difficult to find in the literature, since it has only been investigated in detail for a small number of well-characterized materials. You are unlikely to be able to look up the complex refractive index of ranitidine in the Rubber Handbook! Fortunately we do not often need to specify the complex refractive index for most systems, and usually make the assumption that the complex part is zero if the material is colourless.

Mie Theory

We will examine only the results from the Mie theory, since the mathematical details are too extensive to be discussed usefully here. They are treated in detail by Bohren and Huffman (1983).

The theory is derived by solving Maxwell's equations for the incidence of a plane wavefront on a particle. It calculates the induced electric field patterns in the particles (the so-called *spherical harmonic electric field modes*), then calculates the diffraction pattern from the light radiated by these modes. The Mie theory is probably best treated as a

computational black box into which the particle properties can be input, and the scattering pattern output. Bohren and Huffman provide a Fortran program which can be used for these computations; the author has similar programs written in Microsoft Quickbasic and MacFortran for the Macintosh, and can provide a copy to anyone who is interested. In order to calculate the light scattering pattern, it is only necessary to specify the complex refractive index of the sample material, the refractive index of the suspending medium, the particle size, and the wavelength of light being scattered. These are then combined into two dimensionless parameters, the relative refractive index, discussed earlier, and the *size parameter,* x, which is given by:

$$x = \frac{2\pi Na}{\lambda}$$

where a is the particle radius and λ is the wavelength of light. The size parameter thus describes the particle size in terms of the wavelength of scattered light, and since the scattering pattern is dependent only on the size parameter and not on the actual particle size, we can see that, for example, 632 nm light scattered from a 1 μm particle has exactly the same pattern as 1264 nm light scattered from a 2 μm particle with the same refractive index.

The output of the theory is shown in Figure 6.1 for spheres with M= 1.33 (e.g. water droplets in air) for a range of size parameters x from 1 to 20. The figure represents a polar plot of a section of the angular intensity scattering pattern from the spheres.

When the spheres are very small compared to the wavelength (i.e.x<<1), the scattered intensity is isotropic, i.e. the same amount of light is scattered in all directions. As the spheres become larger, the scattering in the forward direction increases; for x=3 the scattering is almost all contained in a 45° forward-pointing lobe. For larger particles the scattering converges into an even narrower forward cone; as the particles reach a size parameter of about 20, small side lobes become visible, although these are much weaker than the forward-scattered signal.

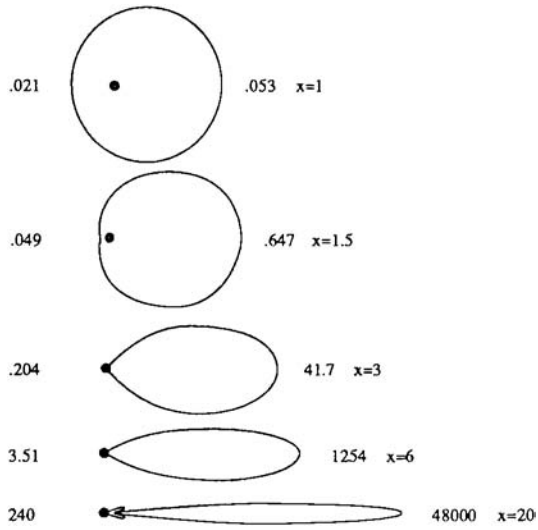


Figure 6.1. Light scattering from a spherical particle as predicted by Mie theory.

The intensity of the light scattered from a single sphere is also highly dependent on the sphere size, and increases rapidly as the spheres become larger. The numbers on the axis at 0° and 180° in Figure 6.1 represent the forward and backward scattering intensities respectively from the particles, and it is evident that the largest particle scatters many thousands of times more light than the smaller ones. The backward scattering also increases as x increases, but we do not normally notice this as most experiments only detect forward scattering.

Scattering of polarized light

The scattering patterns of Figure 6.1 do not tell a complete story, because, for the smaller particles, the angular intensity distribution is dependent on the polarization of the light illuminating the particles. Figure 6.1 shows the polar plots for light whose plane of polarization is perpendicular to the scattering plane. If the light is polarized parallel to the scattering plane, the scattering patterns of Figure 6.2 are obtained. For very small particles, no light is scattered at 90°

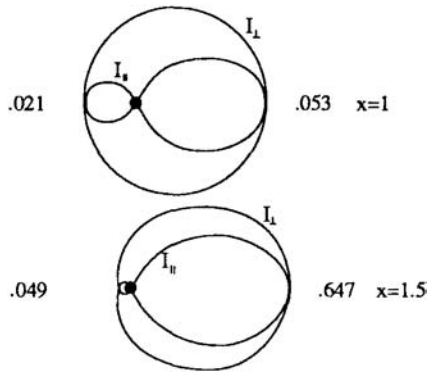


Figure 6.2. Scattering of parallel polarized light from particles.

to the incident beam, and equal intensities are scattered forward and backwards. As the particle size increases, the scattering becomes predominantly forward, until it is identical to that observed for perpendicularly polarized light.

In order to simplify data analysis from scattering experiments, a number of approximate light scattering models are available. These normally are only valid for particular ranges of refractive index and size parameter. They include *Rayleigh scattering, Rayleigh-Gans-Debye scattering, anomalous scattering,* and *Fraunhofer scattering.*

Rayleigh scattering

If the particles are extremely small compared to the wavelength of light, they can be treated as point scatterers into which the light does not penetrate. This leads to a very simple scattering pattern, which is shown in Figure 6.3. This pattern is identical to that obtained from the Mie theory for particles of very small diameter. The scattering for perpendicularly polarized light is isotropic, while that from parallel polarized light has a minimum at 90° to the illumination axis. If the light is randomly polarized, the mean of the two scattering patterns is obtained.

For randomly polarized light, the intensity at an angle θ is given by:

where I_i is the incident intensity, r is the distance from the particle to the detector, λ is the wavelength of light; the first

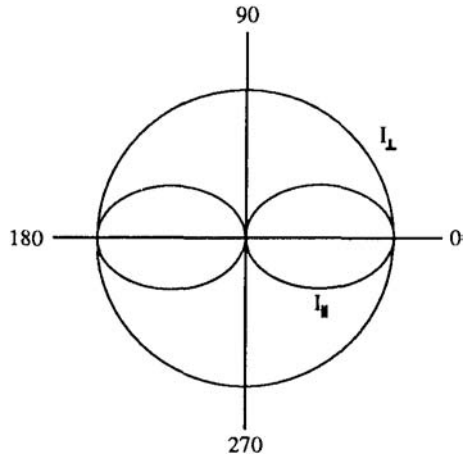


Figure 6.3. Rayleigh scattering of light.

$$I_s = I_i \alpha^2 \left(\frac{2\pi}{\lambda}\right)^4 \frac{(1+\cos^2\theta)}{2r^2}$$

$$\alpha = \frac{3(m^2-1)}{(m^2+2)} \cdot \frac{V}{4}$$

term in the brackets represents the scattering from the perpendicular polarization, and the second from the parallel polarization.

α is the particle polarizability, given by:

where V is the particle volume and m the relative refractive index.

We can thus see that the scattering intensity at any angle is proportional to the square of the volume or the sixth power of the diameter of the particle. This rather important result has two consequences. Firstly, the scattered light intensity is very sensitive to size, so small size changes lead to large changes in scattered light intensity. This gives light scattering a potentially very high resolution for studying particle size. This property is exploited in flow ultramicroscopes (Cummins *et al.* 1983) which potentially have the highest resolution of any submicron instrument, and can measure the scattering of light by single particles. The negative aspect of this result is that small particles

scatter very small amounts of light; for example, a 50 nm particle scatters only 1/64th of the light scattered by a 100 nm particle. Consequently all light scattering techniques have difficulty in studying very small particles, as they run out of sensitivity. An instrument which measures only light intensity needs an enormous dynamic range in order to measure even a modest range of particle diameters.

We should note that the Rayleigh theory suggests (as is borne out in practice) that the scattering pattern and intensity from very small particles will be independent of particle shape, since the particles are treated as point scatterers.

Rayleigh-Gans-Debye (RGD) scattering

This is an extension of the Rayleigh scattering theory which makes the assumption that the particle refractive index is close to that of the medium in which it is suspended. This allows the scattering pattern to be calculated for particles slightly larger than those for which the simple Rayleigh theory is applicable, up to a size parameter of approximately 1, depending on the relative refractive index. Like Rayleigh scattering, RGD scattering predicts that the diffraction pattern is independent of particle shape.

Fraunhofer scattering

Fraunhofer scattering is a simple diffraction theory which was used in most early laser diffraction instruments. It allows the scattering patterns to be derived for large particles, from about x=10 upwards, without reference to the optical properties of the material. The theory assumes that light that has penetrated the particle does not contribute to the scattering pattern, or alternatively that the particles are completely opaque, and so the scattering can be described as though the particles were a collection of opaque discs. Consequently it is increasingly in error for smaller particles, but is useful for rapid calculations of the scattering from particles of several micrometres and larger.

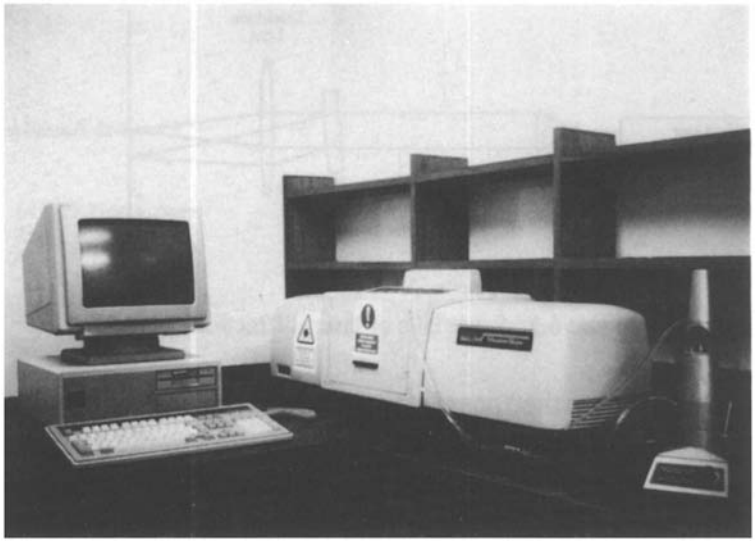


Figure 6.4. Malvern Mastersizer.

Most of the first generation of light scattering particle sizers used Fraunhofer analysis, e.g. the early Malvern 2600 series.

Anomalous scattering

The term anomalous scattering is normally used to describe effects due to light which has penetrated the particle. Consequently it requires a knowledge of the optical properties of the material from which the particle is made. As implemented on several commercial instruments, it can be considered as a modified Fraunhofer theory which requires the relative refractive index to be supplied, and hence can analyse data from particles somewhat smaller than those measurable by Fraunhofer methods alone. It does not, however, cover the whole range of sizes that can be examined by Mie theory, and generates increasing errors for particles of diameter near to, or smaller than, the wavelength of light.

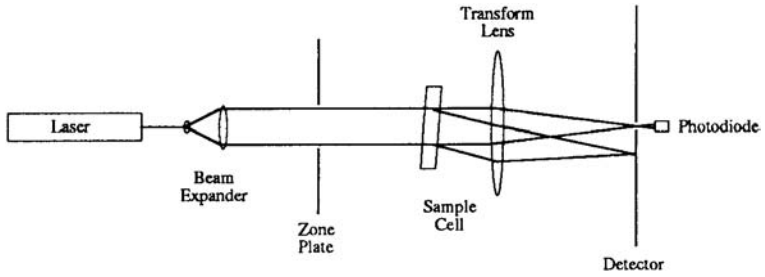


Figure 6.5. Schematic of laser diffraction sizer.

LASER DIFFRACTION PARTICLE SIZERS

To illustrate the operation of a laser diffraction particle sizer we will refer to the Malvern Mastersizer (Figure 6.4), one of the most popular and established instruments. Other manufacturers supplying diffraction sizers include Coulter, Leeds and Northrup, Sympatec, and Fritsch (see e.g. Plantz and Lowe 1982 for a short review of the application of these instruments). Figure 6.5 illustrates the principle of a diffraction sizer. The light source is almost invariably a heliumneon laser with a wavelength of 632.8 nm and a power from 2 to 10 mW. The beam diameter of these small gas lasers is normally about 1 mm, so a *beam expander* is used to produce a uniform parallel beam of 5–10 mm in diameter, to allow a useful sample volume to be illuminated. The optics often introduce some unwanted scattering and diffraction, and so are followed by a *zone plate,* whose function is to allow through only the central intensity maximum of the laser beam.

The expanded beam then passes through the sample. This is contained in a cell with optically flat windows, enclosing a sample path of 2–20 mm of suspension. The cell is normally mounted at a slight angle to the beam, so that back reflections from the cell windows are not returned to the laser to be multiply reflected; this tilt setting is a manual adjustment on some instruments (e.g. the 2600 series). Since the instrument is capable of measuring particles which are a significant fraction of a millimetre in diameter, the cell is stirred or fed from a recirculator in order to keep

the particles suspended. Malvern's 2600 series sizers, although now outdated by later models, are very useful since they have a small cell with integral stirrer. This takes a much smaller sample than most other models, and is easy to clean and dry if nonaqueous solvents are required.

The light from the sample passes through the *transform lens,* which images the scattering pattern onto the *diode array detector.* The transform lens is so called because it takes as input light which has been scattered throughout a range of angles, and produces an output image in which each scattering angle corresponds to an annular ring in the detector plane. Consequently the signal at a particular position on the detector corresponds to the intensity of light that has been scattered through a particular angle, regardless of the position of the scattering particle in the suspension; the lens has performed an angle-to-position transform.

The focal length of the lens sets the correspondence between scattering angle and image position on the detector; thus if a lens with a long focal length is used, the smallest scattering angles are imaged on to the detector, while if a shorter focal length lens is used, larger scattering angles are imaged. Consequently the larger focal length lens gathers data from larger particles, and the smaller lens from smaller particles. The manufacturers do not normally quote the range of angles measured, but specify a measurement range of particle diameters for each lens. Thus the aforementioned Mastersizer can measure from 0.1 to 80 μm with the 45 mm lens, and from 1.2 to 600 μm with the 300 mm lens. Lenses designed to image very large scattering angles require a correspondingly large numerical aperture; to avoid vignetting and to simplify the layout, Malvern use the shortest (45 mm focal length) lens in the reverse transform position, in which the lens is in front of the sample rather than behind (Figure 6.6). This arrangement only provides a true transform if the sample is thin, so the sample cell has a path length of only 2 mm.

The detector is shown in close-up in Figure 6.7. Its task is to measure the intensity distribution of the circularly

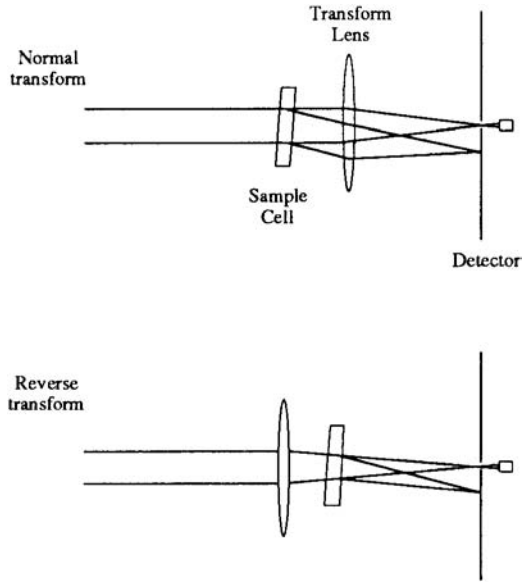


Figure 6.6. Normal and reverse transform optics.

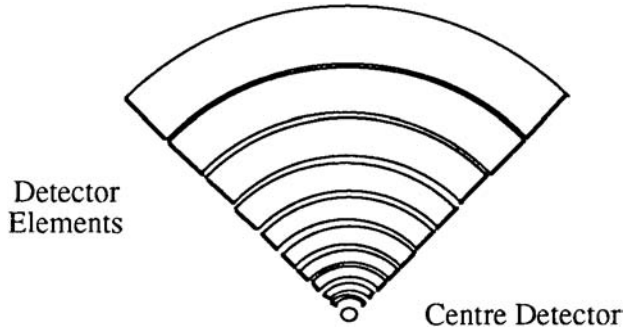


Figure 6.7. Diode array detector.

symmetrical scattering pattern, which it does with an array of 16 or 32 photodiodes, arranged outwards from the beam axis in a radial pattern. Alternate diodes may be placed on opposite sides of the detector to simplify construction.

The diodes increase in area towards the outside of the array, since the scattering intensity at large angles may be several orders of magnitude smaller than that near the centre of the diffraction pattern. The larger diodes have a correspondingly higher sensitivity. Each diode feeds its

signal to a preamplifier, which transmits an intensity signal to the computer.

The undiffracted light is focused to a point at the centre of the diode array. This is not normally imaged onto the main detector array, but passes instead through a pinhole to a separate diode behind the array. The reason for this is that there is considerable power in the centre spot, and if it were focused on to the array, the flare would be detected by the inner diode elements, increasing the background signal. Imaging it onto a separate detector avoids this problem. The signal from this detector is normally used to provide a separate 'undiffracted' light intensity, and is displayed as an indication of laser power, and also used for alignment purposes. Reduction in signal intensity at this central detector is used to calculate the sample absorbance (more properly termed *extinction* for a scattering sample), which allows the desired sample concentration to be achieved.

Data processing

We have discussed broadly the techniques that can be used to calculate the scattering pattern from a sample of known size distribution, whether or not a Mie calculation is carried out, or various approximations are made. Unfortunately we need to solve the opposite problem in order to analyse the light scattering data, i.e. we need to calculate the size distribution from the scattering pattern. This is an example of what a mathematician would call an *inverse problem,* and it is often the case that inverse problems are more difficult than the corresponding direct problems. This is very definitely the case with the analysis of light scattering data, and we will encounter a further example of this in the next chapter on photon correlation. In practice it turns out to be almost impossible to find the size distribution by carrying out simple algebraic operations on the scattering pattern, and we have to have recourse to number-crunching instead. The software usually works as follows:

1) The scattering pattern is examined, and a rough guess is made at the size distribution on the basis of any maxima

present, and the level of structure visible. We call this guess the *trial distribution.*

2) The scattering model is used to calculate the scattering pattern that would be observed from the trial distribution.

3) The trial distribution scattering pattern is compared to the experimentally observed pattern, and alterations to the trial distribution are made on the basis of the observed differences.

4) The scattering pattern is recalculated for the new trial distribution. This loop is repeated until the difference between the experimental scattering pattern and that calculated for the trial distribution is a least-squares minimum. The trial pattern is then presented as an approximation to the true size distribution. (Actually, manufacturers often state that it *is* the size distribution, but we would prefer to be pedantic at this point!)

As the trial distribution is altered, it is usually possible to constrain it in a number of ways if necessary. For example, it could be constrained so that the best normal or lognormal distribution was fitted to the data. This usually results in very rapid calculation since only a small number of adjustable parameters are used. Normally, however, the distribution is unconstrained, so that a separate particle concentration is established in each size channel. This uses a large number of variable parameters, and so the calculation is normally much slower. In order to speed up the data processing, the instruments do not usually perform repeated Mie calculations at each step through the loop. Instead, the Mie scattering patterns are precalculated for each particle size category by the manufacturers and stored in the software. In order to calculate the scattering pattern for a given size distribution, the instrument simply adds up the appropriate weights of the scattering patterns for the individual particle sizes. This process can be taken further by using simple matrix linear algebra to solve the inverse problem and split the observed scattering pattern into a linear sum of the patterns contributed by the individual sizes (Maitre and Levy 1983). This is basically a problem of solving a set of simultaneous equations, one for each

detector signal intensity, and as such requires that there are as many detector elements as there are size classes in the derived distribution; you cannot solve 10 equations for 11 unknowns. An instrument which, for example, uses a 16-element detector only has 16 equations to solve and so can only extract 16 particle size classes. If it displays a 32-channel size distribution, it has probably made some assumptions or done some interpolation somewhere without telling you!

Sample preparation

The sample should be well-dispersed in a suitable solvent using the techniques described in Chapter 2. Nonaqueous solvents can be used, although compatibility of connecting tubing, and seals in sample circulators, must be checked. We normally only use the small stirred cell on the Malvern 2600 sizer for nonaqueous work, since it is rather tedious to clean and dry a system fitted with circulating pumps, etc.

Most instruments can be fitted with a *sample processor.* These are usually modified ultrasonic baths in which the diluted sample can be placed, and are equipped with stirrers and pumps for recirculating the sample. Although they are useful for finding the best sample preparation methods, they do take rather a lot of sample. About a gram of dispersed solid is needed for the circulator supplied with the Mastersizer, which is fine if you are going to put a shovelful of soil into it, but is far too large for research work with Pharmaceuticals, in which sample sizes are usually very limited. Most instruments can be fitted with small sample circulators or stirred cells, and it is essential to check that this is the case before you purchase a particular instrument for small-sample work.

Dry powder feeders

Dry powders can be analysed by flowing the well-dispersed powder through the laser beam. Manufacturers of laser diffraction instruments normally supply accessories that

allow dry powders to be blown through the beam by compressed air, and evacuated by a vacuum cleaner. These work well if the powder is free-flowing, but if the particles are cohesive, the material may not be properly dispersed. Early dry powder feeders tended to blow the material around, but later models have completely enclosed flow paths, and so can be used without fume extraction.

Aerosol sizing accessories

One of the main phenomena that was responsible for provoking early interest in light scattering in the 19th century was the possibility of explaining the formation of atmospheric effects such as rainbows and haloes. Laser diffraction is a very useful method for aerosol studies, since most of the liquid droplets involved range from around a micrometre up to several tens of micrometres. The classical techniques for studying size distributions in aerosols are impacting devices such as the Anderson sampler; these methods will be considered in Chapter 8. An aerosol sampler head can be fitted to some laser diffraction instruments, and it allows the droplet size of aerosols to be measured.

In the simplest measurement of continuous sprays, it is simply necessary to spray the aerosol into the laser beam during the measurement. The nozzle can be held in an appropriate position so that a particular section of the spray falls in the illuminated region (Figure 6.8). It is useful to place an extractor behind the instrument, particularly if active drug aerosols are studied; this protects both the operator and the instrument optics. The signal levels produced by aerosols on early instruments such as the Malvern 2600 are rather low, but later instruments have improved sensitivity (the Mastersizer has a high gain mode which improves the situation to some extent).

More advanced methods are available for the study of metered dose inhalers or MDI's. These devices emit a plume of droplets which rapidly evaporate to leave a fine mist of particles, and only operate for a fraction of a second. In

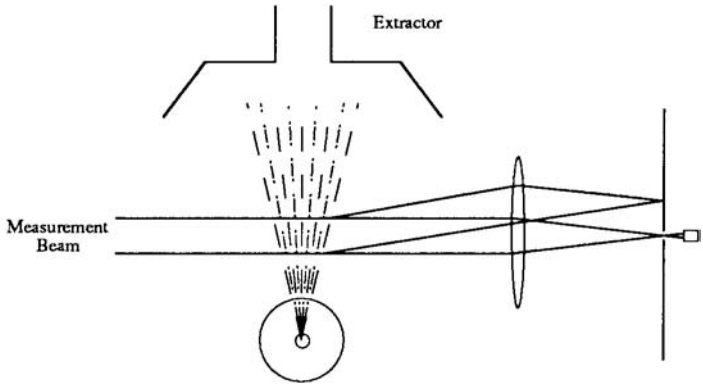


Figure 6.8. Measurement of aerosol size distribution.

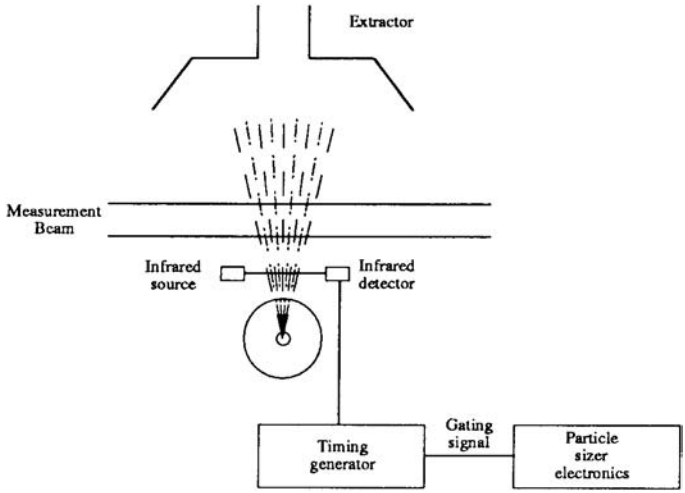


Figure 6.9. Time-resolved aerosol sampler.

order to study the aerosolization process, it is useful to measure the size distribution at various positions in the plume, and as a function of time. This can be accomplished with a *time resolved sampling accessory* (Figure 6.9). The MDI is held in a clamp which allows it to be accurately positioned, and a light sensor is positioned a few millimetres in front of the spray nozzle. This is connected to a delay timing unit which sends a detector gating signal to the particle sizer, to instruct it only to gather data for a specified time window (a few milliseconds) from a preset delay after

the MDI has been discharged. When the MDI is fired it is detected by the light sensor, and after the preset delay, the sizer gathers the diffraction pattern for a few milliseconds. Adjustment of the delay allows a snapshot of particle size to be obtained at any desired time after firing and and at any point within the plume.

CALIBRATION

Laser diffraction instruments do not need calibration in the classical sense, in that there is no knob to twiddle to bring measured sizes into line with a calibration standard. However, we would recommend a number of regular checks of instrument performance. Butters and Wheatley (1981) studied the response of a Malvern 1800 diffraction sizer to BCR reference materials in the 10–100 μm region, and found a generally good agreement. There is at present no British Standard for the use of light diffraction instruments, largely owing to difficulties in standardizing software in a field in which manufacturers are continually refining the algorithms (or so they claim). We have found the following tests to be useful indicators of instrument performance:

1) Measurement of the distribution of monodisperse standards at both ends and the centre of the range. The measured distribution width is usually narrow in the centre, but widens towards the ends, particularly for the smallest particles.

2) Measurement of a mixture of all three standards containing equal volumes at each particle size. Common effects to watch for are (a) under or overestimation of the proportion of one of the components, (b) a change in the distribution width of a component (particularly the smallest), and (c) 'peak pulling', in which particles of one size influence the mean size of one of the other peaks. All these artefacts arise as unavoidable systematic errors in the data analysis procedure.

In order to check the proper operation of the instrument, some manufacturers supply standard test objects. These consist of glass plates on which an array of small dots has

been etched or photographed. The sample cell is replaced by the test object, when a standard diffraction pattern is obtained. These are useful for rapid checks for quality control purposes and to confirm proper operation of the instrument.

There has been considerable discussion recently concerning the generation of systematic errors by laser diffraction instruments. Recently Allen and Davies (1989) compared a number of particle size instruments, including several laser diffraction instruments, using BCR quartz test materials. The results demonstrated that individual examples of the same instrument usually agree fairly closely, but instruments from different manufacturers may not show such good agreement, both with each other, and with other methods such as sedimentation or electrical zone sensing. The smallest test sample (BCR 66, 0.5–2 μm) showed considerable deviation from the calibrated distribution using laser diffraction, while sedimentation techniques were generally more accurate for this material. Larger particles (BCR 70, 1–7.5μm) also showed considerable deviation, with considerable underestimation of the smaller particles. All the instruments studied showed similar trends for BCR 67 (5–18 μm), in that the smaller particles were underestimated, and the larger ones overestimated. In general all machines showed some sort of systematic error, and it was not possible to select a 'best' instrument. To put the errors in context, it should be pointed out that all the other methods studied (including sedimentation) showed similar levels of variation, apart from the Coulter counter, which was rather more accurate. All of the laser diffraction instruments produced highly reproducible results. Most users of particle size analysers use them in a comparative or quality control sense rather than for the determination of absolute size, so these systematic errors are not often very important for most purposes.

ERRORS AND DISADVANTAGES OF DIFFRACTION SIZE ANALYSIS

Diffraction analysis unfortunately suffers from rather a large number of potential sources of systematic error and disadvantages, which most users consider to be more than offset by the extreme ease of use of the instruments. These errors can normally be kept under control by proper use of the instrument; it is important to realize the sources of error, because in our experience diffraction sizers are treated as 'black boxes' to a greater extent than any other particle size instrument.

Submicrometre resolution

Perhaps the biggest disadvantage of diffraction analysis is its poor submicrometre performance. This is expected from the theory of light scattering, and would not be such a problem if it were not for the excessive claims made by most of the instrument manufacturers. There is no doubt that several manufacturers (notably Malvern and Coulter) have made vital contributions to the development of the technique for the sizing of submicrometre particulates, but we would still regard any submicrometre capability of the instruments as something of a bonus when it works! Laser diffraction is not the method of choice for submicrometre measurement, and photon correlation (Chapter 7) is far more accurate. Unfortunately many users have purchased the newest laser diffraction instruments in the belief that they will eliminate the need for photon correlation, which, in our experience at least, they will not.

Our studies of parenteral fat emulsions (moderately polydisperse emulsions with a VMD of 250–300 nm) suggests that instrument sensitivity drops off very rapidly below approximately 0.6 μm, and the number of droplets below this size may be very severely underestimated. For example, our Malvern Mastersizer, using the 45 mm lens (range claimed 0.1 μm to 60 μm) indicates a VMD of 0.38 –0. 45 μm for a parenteral fat emulsion with PCS diameter of 0. 25 μm.

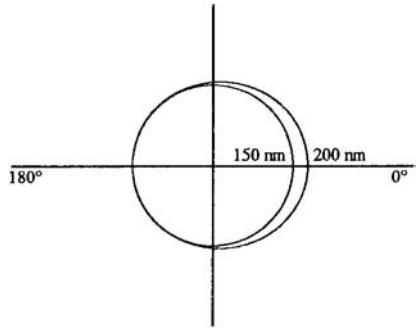


Figure 6.10. Scattering from 150 nm and 200 nm spheres.

The inaccuracies at small particle size are due to three effects. Firstly the scattering of the smallest particles is extremely weak. This is not such a problem if only small particles are present, but if they are mixed with a few larger particles, whose scattering is extremely strong, they may not be detected at all. Secondly, small particles scatter at very large angles, and so the scattered light may fall only on the outer few detector channels, which poorly sample the diffraction pattern. This is coupled to the third problem, which is that, as the particles become smaller, the scattering from them becomes almost isotropic, and so very little information is contained in it. As an example, Figure 6.10 shows the scattering from 150 and 200 nm spheres. The differences between the patterns are very small, and any instrument which tries to separate the particles on this basis is clearly in for a hard time. In order to improve this situation, the Coulter LS sizer uses auxiliary detectors to capture scattering at very large angles, and also analyses the polarization of the scattered light at different wavelengths using a system known as PIDS (polarization intensity differential scattering). This uses a second source, cell, and detectors to measure the polarization of light scattered at a range of angles near 90° at three wavelengths. This additional data provides improved resolution down to 0. 1 μm, and is capable of, for instance, resolving two separate populations of submicrometre polymer latices. On the other hand, the Leeds and Northrup instrument does not attempt

to measure below 0.7 μm by laser diffraction, and uses a wide angle detector to improve the data quality for small particles. This instrument has the advantage that its computer can control both the diffraction analyser and a photon correlation unit, thus saving a few thousand pounds if you need both instruments.

Multiple scattering

The models for light scattering by particles assume that each photon encounters only one particle. If a photon were to be scattered from a series of particles, its incidence angle at all but the first encounter would be unknown, and hence the scattering pattern could not be analysed. This situation is known as multiple scattering, and in general is highly undesirable in any light scattering experiment. In order to ensure that it does not occur, it is important to keep the sample dilute. Most of the available instruments set the sample concentration by monitoring the light absorbed by the sample, whose concentration is increased by the operator until an absorbance of 0.1–0.25 is achieved.

This leads to a minor disadvantage of the technique, since it is not possible to find out how many particles are present. The machine instructs you to add sample ad libitum until it is happy with the signal. Consequently you can find yourself in the situation in which you have two samples, one containing ten times as much material as the other, which give you identical results, since the instrument has instructed you to add ten times as much of the weaker sample. Some instruments attempt to provide a calculation of the volume fraction of disperse phase, but in our experience this is only approximate.

Flow and refractive index fluctuations

The scattering angles for the largest particles are extremely small, and any effect which deflects the light beam slightly can thus lead to errors which have the appearance of large particles. This can occur if warmer or cooler liquids are

circulated through the cell, or if the sample is not thoroughly mixed, due to local fluctuations in refractive index. Similar effects occur if the flow of sample is too rapid. This can be clearly seen by turning up the flow rate of the sample recirculator and observing the scattering pattern, when a large amount of small-angle noise will become apparent. If large particles are being measured, the flow rate should be set to the minimum value which will suspend them without sedimentation, and sufficient time should be allowed for thorough mixing prior to measurement.

Computing artefacts

We have already mentioned some of the artificial results that can be generated as a result of problems in the data fitting process. Generally, the instruments are accurate for monomodal samples in the centre of the measuring range, but artefacts may appear for samples with a more complex size distribution. Interaction of peaks in multimodal samples may occur, as may peak broadening induced by particles of a different size, or loss of intensity in a peak when particles of a different size are added.

Particle shape

As we have indicated previously, the scattering pattern can only be analysed for simple shapes such as spheres, rods, and ellipsoids. This makes the study of emulsions and microspheres straightforward, but many other feedstocks are irregular or crystalline. Even if the sample did consist of one of the other calculable shapes (e.g. long fibres, such as asbestos), the software only performs a spherical scattering calculation, so a false answer would be obtained.

Probably the easiest way to consider the problems inherent in scattering from irregular particles is as follows. Suppose we have a collection of irregular particles which are of the same size and shape, and initially are all in the same orientation. The scattering pattern produced will not match one obtained from spherical particles, and the software will

analyse the data as though it arose from a sum of spherical particles, and in general will indicate that the scattering arose from a distribution of spheres. Although it is difficult to predict the exact shape of this distribution, we might expect that it was fairly narrow if the particles were near spherical, and broader as the particles departed further from spheres. If the particles were extremely irregular, angular, or even fractal, we may obtain artefact peaks which suggest that more than one size of particle was present. If the particles are now allowed to take up .any orientation, as they would in a real measurement, the scattering pattern will be smeared out even more, and the distribution will show orientational broadening.

Consequently, the scattering from irregular particles should be interpreted with caution, since light scattering will 'mix' particle size and shape in an unpredictable manner. The results are usually adequate for ensuring constancy of size for quality control purposes, but should not be interpreted as providing more than a guideline for the true size.

A question which occasionally arises is whether or not laser diffraction can determine internal structure, for example in multiple emulsions or microcapsules. Although we have seen no hard data on these systems, we would expect the diffraction pattern to contain information describing all the dimensions of structures which could be accessed by the light. As long as the particles were not completely opaque, diffraction of light through the particle interior should be sensitive to its structure. The analysis of such data would be extremely complex. Probably the most easily analysed case would be that of the microcapsule, since the concentric sphere or shell structure has a well-studied diffraction pattern (see e.g. Bohren and Huffman 1983). We might hope to be able to measure the diameter of the interior droplets in a microemulsion, but since there may be several of these in a small space, we would expect considerable multiple scattering. Additional problems would be produced by the interaction of the spherical harmonic scattering fields on the nearby particles, which would cause

complex interference effects between them. In all cases, however, the software of existing instruments would analyse the diffraction pattern as a sum of spherical scattering patterns, and specialized software would have to be provided to extract the desired information.

LASER SAFETY

New users of laser equipment may be puzzled by the laser hazard classification, so a few words might prove useful. The hazard from a laser is determined from the so-called *accessible emission,* which is basically how much light the user can be exposed to when the instrument is used. Lasers are placed into safety classes defined by *accessible emission limits* or *AEL's.* The classes, very roughly, are defined on the basis of how much power is needed to cause a given injury. They range from class 1, which poses no real hazard, such as completely enclosed systems, to class 4, in which even diffusely scattered light can cause eye and skin damage.

The confusing thing about the system is that the rating is based on the *accessible* amount of light. Consequently if a powerful laser is built into an instrument, and the light path is completely enclosed and interlocked, so that the beam cannot be accessed in normal use, then the instrument is rated as class 1, even though it may contain a class 4 laser. If the instrument is opened, and the interlock over-ridden, the classification of the instrument reverts to that for the laser.

Most particle sizing instruments contain helium-neon lasers with a power of 5–10 mW. These are class 2 or 3 lasers, which means that the direct or reflected beam can cause eye injury, and the direct beam of class 3 lasers can cause skin injury. In a properly built system, the beam will not be visible, and it will not be possible to poke shiny objects into it (a major cause of laser accidents!). Consequently the instruments are rated class 1. People using these instruments routinely will need no specific laser training or eye tests. However, user maintenance procedures often require the beam to be viewed, and the person who

does this must conform to your organizations's full laser training requirements. Fuller details of laser safety can be found in Winburn (1990); laser safety officers should note that the EEC will probably require the U.K. safety standards to be changed in the near future.

OPTICS CLEANING

Many particle technologists have little training in optics, and manufacturers usually recommend that major optical maintenance tasks are performed by their engineers. However, a number of minor tasks need to be carried out regularly by the user. The most obvious is the cleaning of the cell and optics; any speck of dust on these surfaces diffracts light and contributes to the instrument background signal.

Optical surfaces are usually coated to reduce surface reflection, and are polished to a high finish. Any attempt to clean the surface by abrasion will destroy both the surface and the coating. Lenses should only require dust removal, and a compressed air canister is the most convenient method for this; a small sable-hair brush (kept clean for this purpose alone) will remove more stubborn specks. Fingerprints should be removed with a suitable redistilled solvent on a lens tissue; the person who applied them should also be removed, preferably permanently. We use redistilled ethyl acetate for grease removal; the golden rule is 'wipe, don't rub'.

Often a high background signal will be obtained even when the cell windows appear clean. If the cell is dismantled and the windows blotted dry, the offending deposits will be more evident. Fatty and proteinaceous materials can leave deposits that are almost invisible to the naked eye, but play havoc with the small-angle diffraction pattern. Gentle wiping with a suitable redistilled solvent will remove most materials and should leave a smear-free finish. Gradually the windows will become damaged by repeated cleaning, and should be replaced when they cannot be cleaned sufficiently to produce a suitably low background signal.

REFERENCES

Allen T and Davies R. (1989) Evaluation of instruments for particle size analysis. *Proc. 4th European Symposium of Particle Characterization,* Nurenberg, 17–46.

Bohren C.F. and Huffman D.R. (1983) *Absorption and scattering of light by small particles.* Wiley, New York.

Butters G. and Wheatley A.L. (1981) Experience with the Malvern 1800ST laser diffraction particle sizer. In: Stanley-Wood N.G. and Allen T. (eds), *Proc. particle size analysis conference,* Loughborough, John Wiley, Chichester, 425–436.

Cummins P.G., Staples E.J., Thompson L.G., Smith A.L. and Pope L. (1983) *J. Colloid. Int. Sci.* **92** 189–197.

Kerker M. (1969) *The scattering of light and other electromagnetic radiation.* Academic Press, New York.

Maitre H. and Levy A.J. (1983) *J. Optics* (Paris) **14** 4.

Plantz P.E. and Lowe C.M.T. (1982) *Drug Cosmet. Ind.* **131** 93–95**.**

Van de Hulst (1957) *Light scattering by small particles.* Wiley, New York.

Winburn D.C. (1990) *Practical Laser Safety.* Marcel Dekker, New York.

7
Photon correlation spectroscopy

Although the majority of pharmaceutical particulates are larger than 1 μm in diameter, some systems consist of considerably smaller particles. The most well-known examples of these are parenteral fat emulsions, which have mean diameters of 200–300 nm. For some years, there has been increasing interest in using very small particles to solve a range of drug delivery problems. Liposomes were among the first submicrometre systems to be investigated, and they typically have mean diameters of 20–1000 nm. It has taken nearly 30 years of development to achieve a marketable product using liposome technology (the antifungal Ambisome is being investigated in extensive trials at present, and appears certain to reach the market). At least part of this delay can be attributed to the unfamiliarity of the pharmaceutical industry with submicrometre or colloidal systems. Many investigators are now experimenting with liposomes, emulsions, microemulsions, nanoparticles made from every conceivable polymer, and even submicrometre nanocapsules, and we can only expect more colloidal and submicrometre formulations to reach the marketplace in coming years.

Photon correlation spectroscopy (PCS) is the most useful technique for particle size analysis of submicrometre particulates, having a range of application from only a few nanometres to a few micrometres. Although many other techniques have been extended into this range (e.g. laser diffraction and electrical zone sensing), they are working at the limits of their capability, and cannot operate with the precision of photon correlation. Probably the only other

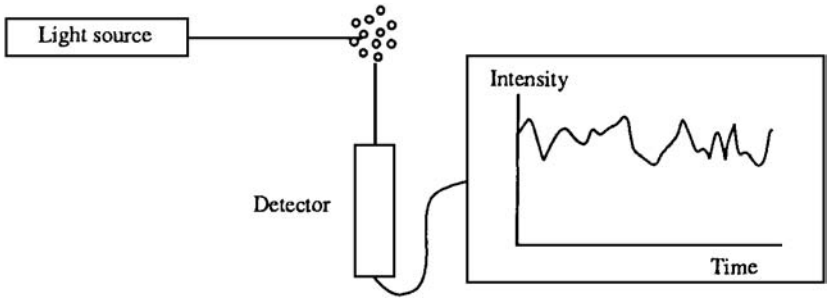


Figure 7.1. Photon correlation experiment.

technique generally applicable in this size range is centrifugal sedimentation, which is comparatively slow. The best text to date on photon correlation is that by Chu (1991) but the older work by Byme and Pecora (1976) is also useful.

PRINCIPLE OF PHOTON CORRELATION

Photon correlation is a light-scattering experiment in which the statistical intensity fluctuations in light scattered from the particles are measured. These fluctuations are due to the random Brownian motion of the particles. Consider the experimental arrangement of Figure 7.1. A focused laser beam illuminates a small volume of the sample, which consists of a dilute suspension of particles. The light scattered from these particles is collected by a lens and its intensity measured by a photomultiplier. If the sample were completely uniform, a constant light intensity would be scattered; however, the constant Brownian motion of the particles in the small illuminated region causes the intensity of the scattered light to fluctuate as particles move in and out of the beam, and so introduces a small randomly varying component (a 'noise component') into the photomultiplier signal.

If the suspended particles are small, they diffuse relatively rapidly, and so the fluctuations in the scattered light are correspondingly rapid. Alternatively, if the particles are large, their movement is slower, and the scattered light

fluctuations occur over a longer timescale. Consequently it will be appreciated that the temporal variations in scattered light intensity contain information which could allow the diffusion coefficient of the particles to be obtained. Once the diffusion coefficient is known, the equivalent diffusional spherical diameter can be found from the *Stokes-Einstein equation,* which relates the diffusion coefficient D of a spherical particle to its diameter d:

$$D = \frac{kT}{3\pi\eta d}$$

where η is the viscosity of the surrounding medium.

The problem is thus the extraction of the diffusion coefficient from the noise signal in the scattered light intensity. Since the frequency of the noise depends on the diffusion coefficient, it could be suggested that the noise frequency spectrum in the photomultiplier signal be measured with a spectrum analyser. This is perfectly feasible; in fact the frequency spectrum $P(\omega)$ of the intensity noise scattered from a collection of randomly diffusing monodisperse spheres can be shown to have a *Lorentzian distribution* (Ford, 1983):

$$P(\omega) = \frac{2Dq^2/\omega}{\omega^2 + (2Dq^2)^2}$$

where ω is the frequency, D is the diffusion coefficient and q is known as the *scattering vector,* and is given by:

$$q = \frac{4\pi n}{\lambda} \sin\left(\frac{\theta}{2}\right)$$

Thus we can see that the noise spectrum scattered from the particles depends on their diffusion coefficient, the wavelength λ of the illuminating light, the refractive index n of the suspending medium, and the angle θ through which the light is scattered.

In practice this measurement is not normally performed using frequency analysis techniques, but by measurement in the time domain, since this is accomplished using less complex equipment and leads to a more straightforward analysis. The time domain equivalent of the frequency

$$G(\tau) = \frac{1}{T}\int_0^T I(t)\, I(t+\tau)\, dt$$

spectrum is called the *intensity autocorrelation function,* G (τ), and is obtained by taking the Fourier transform of the (frequency domain) noise spectrum.

The autocorrelation function of a fluctuating quantity measured as a function of time, I(t) (e.g. a scattered light intensity) which has been measured from time t=0 to time t=T is defined formally as follows:

It is worth looking rather closely at the physical meaning of this equation; the only mathematics needed to interpret it is to recall that the integral of a graphical function represents the area under the curve. The correlation function is a function not of real time (i.e. experimental time) but rather of the *correlation time,* τ, and in order to construct it we need to consider the value of the integral as t varies from zero to some finite time.

Figure 7.2 shows a typical time-varying signal I(t) similar to that expected from the random scattering in a PCS experiment. Suppose initially that t=0; then the expression inside the integral is $I(t)^2$. If we integrate this expression over t (i.e we find the area under the curve over the duration of the experiment), then, because the peaks and troughs of the I(t) and I(t+τ) factors are aligned, they reinforce, and the resultant area is relatively large. This is the value of G(τ) at τ=0.

If we now allow τ to increase, this has the effect of shifting the I(t) and I(t+τ) curves over one another before multiplying them together. As this happens, the peaks and troughs begin to lose their alignment in the two separate parts of the integral, and so when we multiply them together and add up the areas, they begin to cancel out. Consequently the initial high value of G(τ) decreases as τ increases, until it becomes equal to some average value of $I(t)^2$ at infinitely long times. Thus the correlation function shows a slowly declining curve similar to that of Figure 7.2. The overall effect is that the 'multiplying and integrating' process measures the

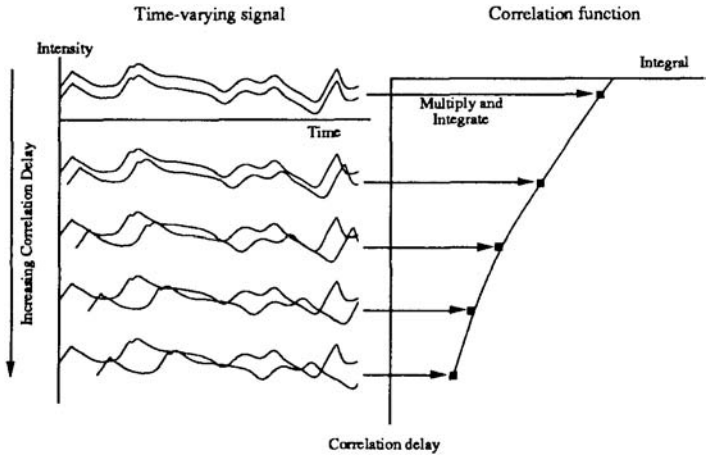


Figure 7.2. Construction of the correlation function of a randomly varying signal.

similarities, or correlations, between the curves I(t) and I(t +τ). We say that, at short times, the signal is well correlated with itself, but as time increases, the correlation is gradually lost due to random fluctuations, and the signal 'forgets' its previous state.

We stated earlier that the frequency spectrum of the noise scattered from a monodisperse collection of particles was a Lorentzian distribution, and that the autocorrelation function was the Fourier transform of the noise frequency spectrum. It is fairly straightforward to show that the Fourier transform of a Lorentzian is an exponential function; thus a collection of monodisperse particles gives rise to an autocorrelation function which decays exponentially. The exponen tial decay fits our intuitive model of an autocorrelation function as a curve with a high initial value, which declines as the correlation time increases. The exact shape of the autocorrelation function for scattering from monodisperse spheres is given by:

$$G(\tau) = 1 + \exp(-2Dq^2\tau)$$

Consequently the calculation of the diffusion coefficient from the correlation function is straightforward, and can be accomplished simply by the usual devices for analysing

exponential data, such as a logarithmic plot of G(τ) against correlation time. In a real system, the measured autocorrelation function will be given by:

$$G(\tau) = A(1 + \beta \exp(-2Dq^2\tau))$$

which is essentially the same expression, but with a scaling factor of β, to account for the fact that the intensity of the experimental data is not normalized, and a baseline offset A, which is due to the various sources of uncorrelated and background scattering from the apparatus, Rayleigh solvent scattering, dust, etc.

PARTICLE SIZE DISTRIBUTIONS FROM PCS

Photon correlation data analysis becomes much more complex when the sample particles are not monodisperse. The autocorrelation function then consists of a sum, or integral, of exponentials, one for each particle size, weighted according to the number frequency of the particles in their size distribution, and scattering intensities from each particle size. For small particles (Rayleigh scattering; Chapter 6), the intensity is proportional to the square of the polarizability, α, which is itself proportional to the third power of the particle diameter. As the particles become larger, Rayleigh scattering no longer applies, and so we write the scattering intensity I from particles of diameter d at scattering vector q as:

$$I(q,d) = N(d)\ \alpha^2 p(q,d)$$

where N(d) is the particle size frequency distribution, and p (q,d) is the *Mie factor,* which allows for the deviation of the particles from Rayleigh scattering, and can be calculated using the Mie theory discussed in Chapter 6. For very small particles the Mie factor is unity and Rayleigh scattering applies. The complete autocorrelation function is then given by:

$$G(\tau) = \int_0^\infty N(d)\ \alpha^2\ p(q,d)\ A(1 + \beta \exp(-2D_d q^2\tau))dd$$

where

$$D_d = \frac{kT}{3\pi\eta d}$$

is the diffusion coefficient of particles of diameter d. This expression is called a *Fredholm integral.*

Now that a range of particle sizes is present, it is evident that we need to know the Mie factors in order to extract a true particle size distribution from the data. In practice these are not easy to calculate, and this problem was skipped over in many early instruments, in other words, the particles were treated as Rayleigh scatterers. If we ignore the Mie factors (i.e. assume that they are all unity) and use one of the approaches described below to obtain the distribution of particle sizes, then we will obtain a size distribution which is weighted according to the scattering powers of the various particles, and a correspondingly weighted average diameter. Not surprisingly, this weight is strongly in favour of the larger particles in the distribution, owing to the d^6 dependency of scattering intensity. This average diameter is called the *z-average diameter,* d_z; this is a term borrowed from polymer chemistry. If we were measuring the diffusion coefficient of a macromolecule, the z-average diameter could be shown to be directly related to various moments of the molecular weight distribution and the radius of gyration of the molecule. This term is thus also used to denote the average diameter measured from a distribution of particle sizes, but in this context it cannot unfortunately be related simply to the moments of the distribution. It should be read as "average diameter weighted according to light scattering intensity". Modern PCS instruments are capable of applying a variety of optical models (usually Rayleigh, RGD or Mie) in order to remove the size dependence of scattering intensity and arrive at a number distribution, from which a mass distribution can be calculated if desired. Unfortunately the strong dependence of scattering intensity on diameter causes this process to magnify errors, so that the resulting distributions may not be too accurate. It is often best to simply quote the intensity-weighted distribution, which will always be more consistent, unless you are very sure of the

optical properties of the sample, and the data is of high quality. In addition, the use of optical models has led to confusion when trying to compare sizes from older instruments with newly-acquired ones, or instruments belonging to other workers. Consequently most workers use PCS data in a comparative sense, and bear in mind that there is likely to be a significant discrepancy between the measured sizes and the true number distributions. It is worth noting that in virtually all cases, $d_z > d_v > d_n$, due to the sixth power weighting of the scattered intensity for the smallest particles. Consequently if the distribution is at all polydisperse, the z-average diameter will be considerably larger than the number or mass diameter. This was demonstrated by Douglas *et al.* (1984) for distributions obtained from cyanoacrylate microparticles.

To illustrate the problems of data analysis introduced by polydisperse samples, consider a system containing two populations of monodisperse particles. The autocorrelation function will then take the form of a sum of two exponentials with different decay times and appropriate weights. The resulting function (Figure 7.3) is very close to an exponential with an 'average' decay time, and data of considerable accuracy is required to extract the original exponentials. A similar problem was described by Lanczos (1957) who showed that a sum of three exponentials could be fitted by a sum of two exponentials with distorted decay times and weights, to a high degree of accuracy. This is not simply a problem of having a suitable data analysis algorithm and data of high accuracy; the solution of the Fredholm integral is said to be *ill-conditioned,* since a small change in the input data leads to a large variation in the solution. In fact, the Fredholm integral has a unique solution only if the data is of infinite accuracy; if an infinitesimal amount of noise or error is introduced into the data, there are an infinite number of solutions!

A considerable amount of effort has been applied to this problem, with only limited success. There are two approaches; firstly to gather more experimental data in an attempt to constrain the solutions, or secondly to develop

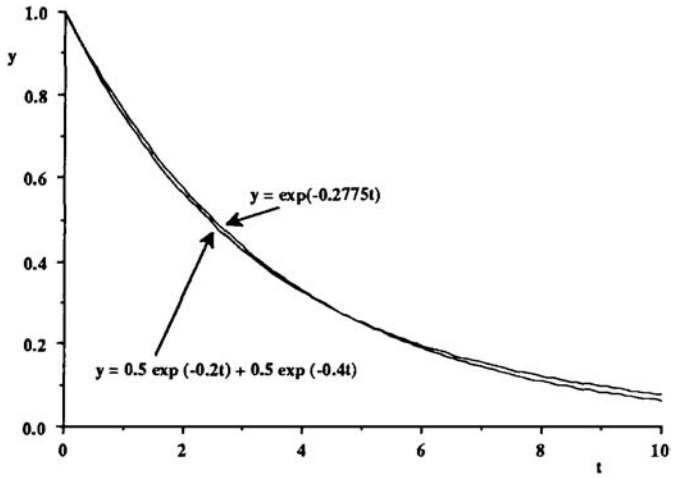


Figure 7.3. Sum of two exponentials and closest single exponential fit.

more sophisticated algorithms for data analysis. Some of the numerical approaches have been compared by Stock and Ray (1985) and Ostrowsky (1988), but not all of the methods they discuss are widely used. Most of the methods vary in their representation of the size distribution, and perform a least-squares fit by adjusting parameters of a trial distribution until the calculated correlation function matches the experimental data. The following approaches are either used in commercial equipment, or are available on mainframe computers for data analysis.

Cumulant analysis

This is the one of the oldest methods of extracting a measure of polydispersity from PCS data (Koppel 1975). The method assumes that the distribution of exponentials is unimodal, and so the mean size is described by the 'best fit average' exponential:

$$G(\tau) = \langle \exp(-2Dq^2\tau) \rangle$$

where the pointed brackets indicate an 'average' exponential, and we have stripped away any baseline or uncorrelated scattering and normalized the correlation

function magnitude to 1 at τ=0. We can then take the logarithm of this equation:

$$\ln(G(\tau)) = <-2Dq^2\tau>$$

D will now be the 'average' diffusion coefficient, and corresponds to an 'average' diameter. The data will only fit this equation exactly if the particles are monodisperse, when the true diameter will be equal to the measured average diameter. If the particles are polydisperse, the data will deviate from this expression. We can apply correction terms by expanding this equation as a power series in τ:

$$\ln(G(\tau)) = K_0 - K_1\tau + \frac{1}{2}K_2\tau^2 - \frac{1}{3!}K_3\tau^3 + \frac{1}{4}K_4\tau^4 \cdots$$

The K_n's are called the *cumulants* of the correlation function, and can be extracted by performing a polynomial fit of the log correlation function. K_1 now describes the z-average diffusion coefficient:

$$K_1 = 2q^2D_z$$

and K_2, K_3, K_4, etc, can be shown to describe the variance, skewness, and kurtosis of the distribution. Normally it is only possible to obtain K_1 and K_2, higher coefficients being buried in the experimental noise. If we assume that the data is completely fitted by K_1 and K_2, then the resultant fit of the distribution is a lognormal. The *polydispersity index* of the distribution, Q, is given by:

$$Q = \frac{-K_2}{4K_1^2}$$

Polydispersity is another term 'borrowed' from the polymer chemists, in which context it is much more rigorously defined. This description is only valid for reasonably narrow distributions (for K_2 less than about 0.25) because otherwise the power series does not converge particularly fast, and the polynomial fit may be poor.

The most significant practical problem in applying the cumulant method is the removal of the baseline scattering prior to analysis. If the data is a single exponential, or if only a narrow range of particle sizes are present, the baseline will be reached fairly quickly, but if a wide range of particle sizes are present, the larger ones will cause the correlation

function to decay for some considerable time, and it will be difficult to estimate the exact baseline position. If the data analysis algorithm misplaces the baseline, the mean diameter obtained will be too large or too small, and the poor fit will give rise to an anomalously large polydispersity in both cases. We have found by experience that the software in our Malvern 3700 PCS system normally places the baseline too low if much noise is present, and so generates spurious large sizes with correspondingly large polydispersity. These appear as pronounced outliers in series of otherwise consistent data points, and are consequently rejected.

Histogram method

In this method (Gulari *et al.,* 1979) the particle size distribution is represented as a histogram with a number of discrete size channels, usually linearly spaced (Figure 7.4 (a)). The population in each histogram channel is varied, and the corresponding correlation function calculated, until the best fit with the experimental correlation function is achieved. It thus corresponds to a minimization problem with as many variable parameters as there are channel widths, and is normally solved iteratively. In common with all minimization problems, continually closer fits can be obtained by increasing the number of variable parameters (i.e. the number of channels). However, beyond a certain point, the procedure simply begins to fit the noise in the data, and the resulting size distribution may show systematic errors or collapse into a meaningless jumble. In general it is better to restrict the analysis to a small number of size channels so that a reliable fit can be obtained.

Exponential sampling method

This method is currently one of the most popular techniques for the extraction of a size distribution, and is used on most commercial instruments; it is sometimes referred to as the Pike-Ostrowsky method. Mathematically it is rather

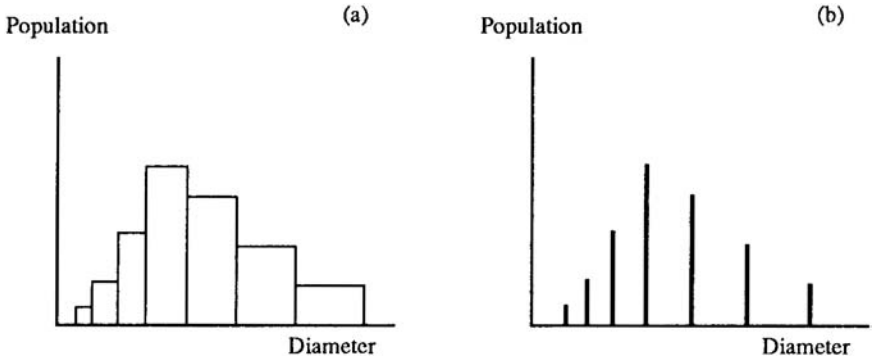


Figure 7.4. Model distributions for correlation analysis, (a) Histogram method, (b) Exponential sampling method.

complex, and can only be understood with a knowledge of Laplace transforms and eigenvalue methods. Consequently we will not discuss the finer details; these can be found in the original papers by Ostrowsky *et al.* (1981) and McWhirter and Pike (1978). The correlation function is represented as a sum of discrete exponentials, spaced logarithmically, which leads to a size distribution similar to that of Figure 7.4(b), which is a sum of a number of delta function components.

One particular problem with the exponential sampling method and its variants is the occasional generation of spurious peaks at the small end of the size distribution when polydisperse samples are studied. These appear to be an artefact of the Laplace inversion, and disappear or move randomly if the sampling parameters or scattering angle is varied. These peaks can be very misleading; for example the author has recently had a number of informal enquiries as to whether or not parenteral fat emulsions have a bimodal size distribution, and if the peak at smaller diameters could be due to scattering from liposomes of excess lecithin surfactant. At present this bimodal distribution appears to be due to artefacts of the data analysis, and can usually be removed simply by analysing the data with fewer size channels.

Constrained regularization

This method (Provencher *et al.* 1978, Provencher 1979) uses matrix techniques to extract a smoothed or 'regularized' size distribution, in which the exponential components have been constrained to take non-negative values. Again this is a mathematically complex procedure, but it is quite a popular method which is resistant to error generation from, for example, dust scattering. It is used on the Coulter N4 instrument, and is also available from its author as the program 'CONTIN' which runs on mainframe computers.

Multiple angle measurement

If correlation functions are measured at a number of angles, then it is possible to obtain a more accurate constraint on the sum of exponential fit obtained. Several manufacturers offer software which allows measurements to be taken at several angles; the value of such an approach was shown by Cummins and Staples (1987) who demonstrated the ability of the technique to resolve bimodal populations, one of the most difficult tasks for any data analysis procedure. The main reason why multiangle techniques are not more widely used is that they are time-consuming, since not only is it necessary to acquire more data, but the Mie coefficients have to be calculated at each angle.

EQUIPMENT FOR PHOTON CORRELATION

The use of photon correlation has increased considerably over the last decade, and a wide range of instruments are available from manufacturers such as Coulter, Nicomp, Malvern, and Langley-Ford. The construction of a typical instrument is illustrated in Figure 7.5. (This is in fact a block diagram of our old Malvern 3700 system, still performing as well as its Commodore 32k PET computer will allow!). The sample is contained in a cylindrical cell, normally 5–10 mm in diameter, in a thermostatically controlled water bath. The light source is a laser, focused by a lens of around 10 cm focal length to a diffraction-limited

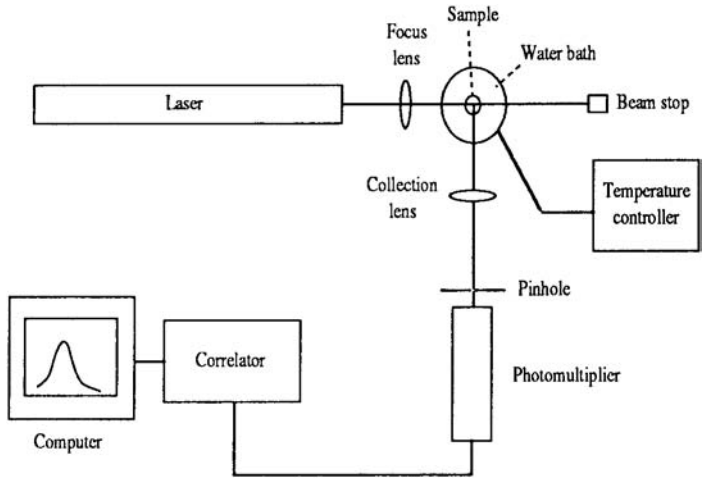


Figure 7.5. A typical PCS instrument.

beam waist. The scattered light is collected by a lens, normally at 90° to the illuminating beam for optimum control of reflections. The lens forms an image of the sample cell and beam; the light from the beam focus is selected using a pinhole, and transmitted to a photomultiplier, which provides a single electrical pulse for each photon detected. This signal passes to the correlator, which generates the autocorrelation function, and passes this to a computer, which performs the appropriate data analysis.

This classical light scattering arrangement has been replaced by one using fibre optics in the Malvern Hi-C particle sizer. This instrument (Figure 7.6) uses optical fibres to illuminate the sample. A fraction of the illuminating light is split off by the fibre splitter; the remainder passes to the probe. Light backscattered from the sample is recollected by the fibre, and passes to the detector, where it interferes with the light delayed by the splitter (a so-called *heterodyne* experiment, in contrast to the normal arrangement, which is called a *homodyne* experiment). This elegant technique allows size measurements to be obtained on concentrated samples which could not be performed using the standard geometry.

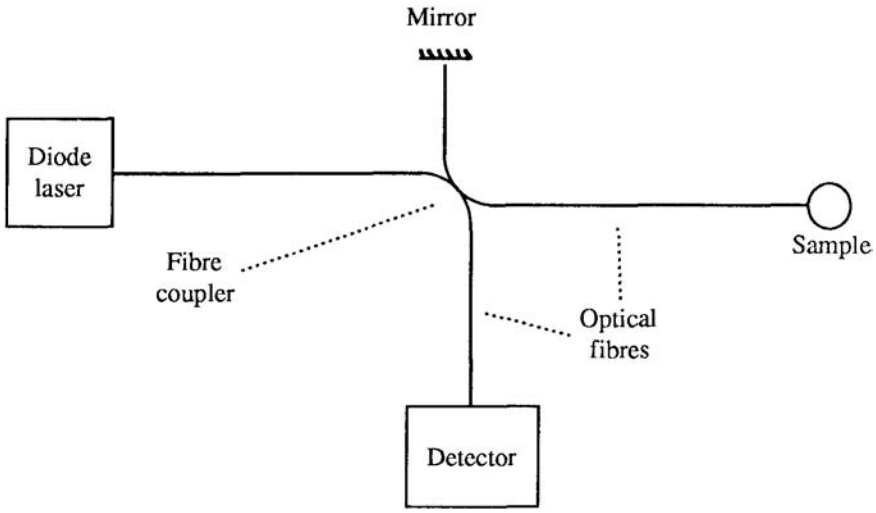


Figure 7.6. Malvern HiC correlation spectrometer.

The laser

Lasers used for PCS vary from 5–50 mW helium-neon devices, suitable for most submicrometre size analysis applications, to large (1 –5 W) argon-ion lasers, which are sufficiently powerful to generate measurable scattering from small micelles and even macromolecules. We routinely use a 40 mW helium-neon laser for analysis of particles as small as 30–40 nm; the cheapness, convenience and reliability of these instruments generally outweighs the extra sensitivity obtainable from a blue laser of similar power. Helium-cadmium or air-cooled argon-ion lasers are also popular; semiconductor lasers are used in the latest Malvern Hi-C instrument. The laser should be vertically polarized to maximize the scattered light intensity; if it is horizontally polarized the scattered intensity will be zero. A randomly polarized laser should be avoided since the intensity of the polarization components may change with time.

The sample optics

The light source is focused to a small point by the input lens, and light scattered from this focus is collected by the viewing lens. The input lens increases the intensity of illumination at the sample, but more importantly it defines the size of the area in which particle diffusion is being measured. The scattered light is made up of spherical wavefronts scattered from each particle, and the fluctuations in intensity at the detector are due to variations in the interference pattern between these wavefronts as the particles move. If the detector area is too large, it will detect many maxima and minima in this diffraction pattern, and only a small fluctuation around the mean intensity will be observed. In order to measure the true amplitude of the intensity fluctuations, the detector should only cover an area of similar size to a single diffraction maximum. Simple diffraction theory allows us to calculate the size range of the spatial fluctuations in the diffraction pattern, and we find that the detector needs to be smaller than a critical area, called the *coherence area,* given by:

$$A_{coh} = \frac{\lambda^2 r^2}{\pi a^2}$$

where λ is the wavelength of light, r is the distance from the scattering centre to the detector, and a is the radius of the illuminated scattering area. The scattering area can be calculated from the input lens properties; lenses do not focus a parallel beam to an infinitely small point, but to a finite beam waist, whose width w is given by:

$$w = \frac{4\lambda F}{\pi d_1}$$

F is the focal length of the lens and d_1 the diameter of the laser beam.

The light collected by the viewing lens is focused to an image at its focal plane, and a pinhole used to select light from the scattering centre. It is important that the pinhole detects only light from the scattering particles; if light reflected from other parts of the apparatus is allowed to reach the detector, it may further interfere with the light

scattered by the sample, generating a beat frequency, which appears on the correlation function.

The detector

Photomultipliers are almost always used for photon correlation due to their high sensitivity, stability, and speed. They are usually operated in the single-photon mode, where the arrival of a single photon at the detector surface causes the generation of an electrical pulse of only a few nanoseconds duration. There are two advantages to using photon counting detectors; firstly they are by definition highly sensitive, and allow very small amounts of scattering to be measured. Secondly, in order to measure the light intensity, one simply counts the photon pulses using digital electronics directly; no analogue to digital conversion is required. In fact there is no useful information in the pulse height, which is unrelated to the scattering intensity. The only problem with photomultipliers that is likely to be encountered is *afterpulsing,* in which a second spurious pulse is generated less than a microsecond after the main photon pulse. This can lead to distorted correlation functions if very short correlation times are used, and is usually only of importance if smaller macromolecular or micellar systems are being studied.

The correlator

The correlator has the task of calculating the autocorrelation function of the input string of photomultiplier pulses. Correlators are in fact extremely simple devices, and although early examples were bulky (the Malvern K7025 was nearly a cubic foot of cabinet, and was not the first such device!), modern spectrometers are built around single chip correlators, and some digital signal processor chips have real-time autocorrelation commands built in.

The basic function of a simple correlator is shown in Figure 7.7. The timescale over which the correlation function of the incoming pulsestream is measured is set by

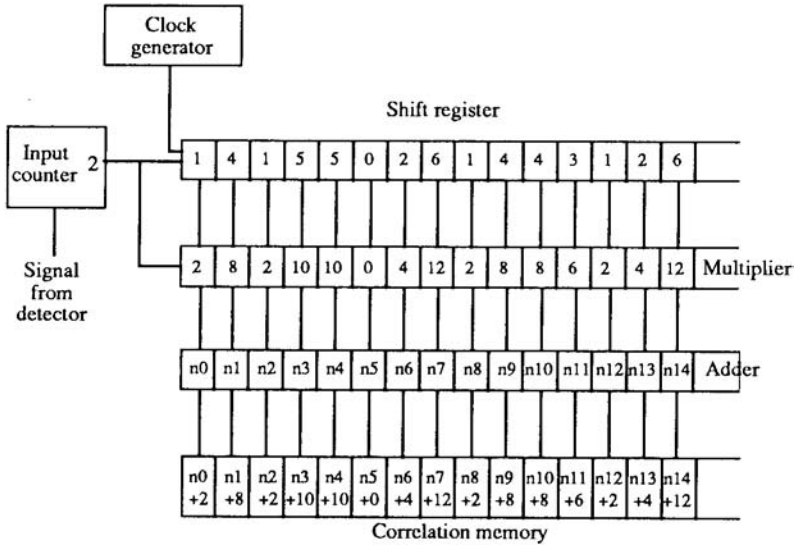


Figure 7.7. Operation of a correlator.

the timebase generator, which generates a clock pulse every few microseconds; the time between these pulses is called the *sample time.* The incoming pulses are counted by the input counter, and at the end of every sample time, the total count in that sample time is passed into the shift register, the contents of each channel of which are moved along by one channel to make room for it (the contents of the last channel 'fall off the end' and are discarded). The count in the first shift register channel is then multiplied by the count in each of the other shift register channels in turn, each product being added to the appropriate channel of the accumulating correlation function stored in the correlation function memory. The shift register and memory are typically 64,128 or 256 channels long.

Thus, for example, if we have a stream of input pulses with n(0), n(1), n(2)..n(i) pulses in each sample time interval, the first correlator channel will accumulate the sum:

$$G(1) = n(0)\,n(1) + n(1)\,n(2) + n(2)\,n(3) \ldots$$

and channel i will contain:

$$G(i) = n(0)\,n(i) + n(1)\,n(i+1) + n(2)\,n(i+2)..$$

which is a numerical approximation to the autocorrelation function of the incoming signal. Modern correlators are a good deal more sophisticated than this simple example, and can compute cross-correlation functions, probability densities, and a range of other statistical descriptions of the incoming signal. Since the analysis of correlation functions usually resolves itself into an extraction of exponentials, some correlators have features to assist this process. One feature that the user may encounter is a delay circuit, which delays the incoming pulse stream by a long period in order to measure a final point in the correlation function at extremely long times. This is effectively the uncorrelated, or background level; recall the need for an accurate baseline for data analysis.

A further attempt to gather information over a broad range of times is the variable time expansion or VTE correlator, in which the correlator consists of a number of channels with sample time τ running in parallel with a further set of channels with sample time 4τ, then another set with sample time 16τ, etc. This allows considerable detail to be gathered at short times, but also measures the correlation function for a long total duration. Modern correlators usually talk directly to computers, and the software selects ('autoranges') the sample time after a brief examination of the sample. This is useful for simple monodisperse materials, but may prove troublesome for polydisperse samples, in which the correlation function continues to decay no matter how much the computer increases the sample time. In general it is highly useful to be able to set the correlator parameters directly, if only so that results from different instruments or different laboratories can be compared.

SAMPLE PREPARATION

Preparation of PCS samples is straightforward if the particles are large or scatter strongly, but becomes increasingly difficult as smaller particles are studied. Many authors take pains to stress the importance of filtration,

etc., but we have found that routine measurement of mean sizes in the range 100–500 nm only requires dilution of the well-dispersed sample into distilled water. The sample should only scatter a small amount of light; if the room is lit by daylight, the rule we generally use is, "if you can see it, it's too strong".

Most instruments will provide some sort of intensity measurement to allow you to adjust the sample concentration to the required level. There are two reasons why the sample cannot be arbitrarily concentrated; firstly, in order to preserve the scattering geometry, each photon must only be scattered once, i.e. there must be no secondary scattering, or scattering of a single photon from several particles. If the particles are highly concentrated, the correlation functions will interfere (we say that some particles act as *local oscillators* for others) and the assumption that the scattering from separate particles is additive will not be correct. Secondary scattering can be seen as a diffuse glow around the bright line of the laser beam passing through the sample. Secondly, the correlator's input counter must not receive more photons than it can count in a single sample time, or the correlation function will be distorted. Normally a so-called *4-bit counter* is used, which can count up to 16 photons in a sample period; if more are received, an overflow signal is produced. If many overflow signals are generated, the sample should be diluted.

Specks of dust in the diluted sample scatter large amounts of light, and can be detected easily since they cause the input count to rise with a sudden jump. The light scattered from such particles is not correlated, and so it makes little contribution to the correlation function other than to raise the baseline; this is only a problem if the baseline level is high and the correlated signal is low. For measurement of mean diameters, dust causes little problem, but if accurate polydispersity or distribution measurements are required, the sample quality must be much higher. In this case, the diluent can be passed through a fine membrane filter; 10–25 nm is a convenient pore size. However, the sample itself may contain large

particles which disturb the measurement; in this case the diluted sample could be passed through a 2–5 μm filter. We prefer to seal the tube containing the diluted sample, and centrifuge it at 2000 rpm for 5–10 minutes prior to carefully transferring it to the spectrometer. This procedure does not normally bias the submicrometre size distribution, but sediments the dust to the bottom of the tube and raises air bubbles to the surface.

ERRORS IN PCS SIZE MEASUREMENT

We have already mentioned a number of sources of error in PCS measurements. These included laser beating from light scattered from other parts of the instrument, and photomultiplier afterpulsing, which are better classed as instrument faults, and artefacts generated by data analysis, such as spuriously large and polydisperse measurements from misplaced baselines, or spurious peaks from attempts to analyse at too high a resolution. The user should also be aware that the apparatus may be vibration-sensitive, and high frequency vibrations may find their way into the correlation function, and so the equipment should be mounted on a sturdy bench.

Significant errors can be caused by convection in the sample, particularly if it is several degrees cooler than the cell holder. Convection disturbs the random motion of the particles and may influence the frequency of the scattered light by the Doppler effect. It is quite common to find that the first measurement on a sample that has just been inserted into the instrument is anomalous; we normally perform a series of 6–8 repeat measurements, and often reject the first as an outlier. It is also important to ensure that accurate temperature control is maintained, since the diffusion coefficient of the particle depends on the viscosity of the medium. The viscosity of most solvents (including water) is strongly temperature-dependent; for example a shift in temperature from 25 to 26°C of a sample in water will change the apparent mean diameter of the sample particles by 3%. Similarly, if any solute is added to the

sample cell, the viscosity change must be taken into account. This is particularly important if macromolecules are added, for instance in studies of adsorbed polymer or protein layer thicknesses on microparticles.

ZETA POTENTIAL MEASUREMENT

Although it is not a description of particle size, measurement of zeta potential has become inextricably connected with the study and characterization of pharmaceutical colloids, and since it also is performed by optical correlation techniques, a brief description may be valuable here. An extensive description of this subject is given by Hunter (1981).

Overview of zeta potential

Surfaces of colloids in suspension develop a charge by virtue of the adsorption of ions or ionization of surface groups, and the charge is correspondingly dependent on both the surface chemistry and the environment of the colloid. The surface charge generates a potential around the particle, which is high near the surface, and decays with distance into the suspending medium. If the particle is placed in an electric field, it will drift with a characteristic velocity, u. The velocity per unit field strength is called the *electrophoretic mobility,* and is normally expressed in micrometres per second per volt per centimetre (electrochemists must have been the last scientists to be told about SI units!)

As the particle moves, it carries with it an ionic environment which extends a small distance into the solvent; the spherical surface separating the moving particle, ions and solvent from the stationary surroundings is called the *surface of hydrodynamic shear,* the elec trophoretic mobility is determined by the potential at this surface, which is termed the *zeta potential,* ζ. The zeta potential can be determined from the electrophoretic mobility using the *Smoluchowski equation,* which applies to large particles in weak electrolytes:

$$u = \frac{\varepsilon\zeta}{\eta}$$

where η is the viscosity of the medium and ε is the permettivity of the environment. In weak electrolytes the potential does not change rapidly with distance into the solvent, and so the zeta potential is often equated with the potential on the colloid surface, or *Stern potential,* and hence used to characterize the surface chemistry of the colloid.

Zeta potential measurement

Zeta potential is normally measured by finding the drift velocity of the colloid in a known electric field, and applying the Smoluchowski equation. In days past this was performed on a single particle basis, in which the suspended particles were observed through a microscope and their velocity across a graticule timed with a stopwatch. Not only was this highly tedious for the experimenter, but the results were subject to many sources of error. The measurement of zeta potential was revolutionized by the introduction of light scattering techniques (Ware and Flygare 1971) for the measurement of the drift velocity via the Doppler shift of scattered light.

The principle of the experiment is illustrated in Figure 7.8. Light from a source is scattered, off a particle, and into a detector. If the particle is stationary, the scattered light will have the same frequency as the incident light. However, if the particle is moving, the frequency of the scattered light will be shifted by the Doppler effect. This is the same effect that causes an ambulance siren note to change in pitch as it passes you. If the particle moves toward the detector, then the scattered waves are squeezed together, and their frequency is increased, while if the particle moves away from the detector, the frequency falls (this is the *red shift* beloved of astronomers). The new frequency f' is given by:

$$f' = f(1 + \frac{v}{c})$$

where f is the frequency of the unscattered light, v is the particle velocity toward the detector, and c is the velocity of light. Since the particle drift velocities are only a tiny fraction of the velocity of light, the frequency shifts are very small, of the order of 100–1000 Hz, while the frequency of helium-neon laser light is 5×10^{14} Hz. The only way this frequency shift can be measured is by mixing *(heterodyning)* the scattered light with light of the unshifted frequency. The difference frequency $f-f'=fv/c$ is thus generated; since this frequency is of the order of hundreds of hertz, it is detectable by conventional electronics.

In order to measure the shift frequency, we need to measure the frequency spectrum of the signal arriving at the detector. This is normally performed by measuring the correlation function of the detector signal. Since the correlation function is the Fourier transform of the frequency spectrum, we only need to transform the correlation function in order to arrive at the frequency spectrum and hence the drift velocity.

Experimental aspects

In practice the experimental arrangement used (Figure 7.9) is slightly more complex than Figure 7.8 suggests. Firstly, it is difficult to measure the light scattered in a direction exactly toward or away from the detector, both for geometric and optical reasons, and so it is normally measured at some fixed angle to the illuminating beam. This only introduces a few cosine terms into the drift velocity calculations. Two beams of light are needed; one can be considered the 'scattered' beam, and the other the 'reference' beam. One of these beams is reflected from a piezoelectric modulator prior to entering the scattering centre; this modulator is moved backwards and forwards over a few micrometres by a sawtooth driving signal. As the modulator moves towards the detector, the Doppler effect reduces the reflected beam frequency by a preset amount (usually 250 Hz or 1 kHz). The difference frequency between the two beams is now:

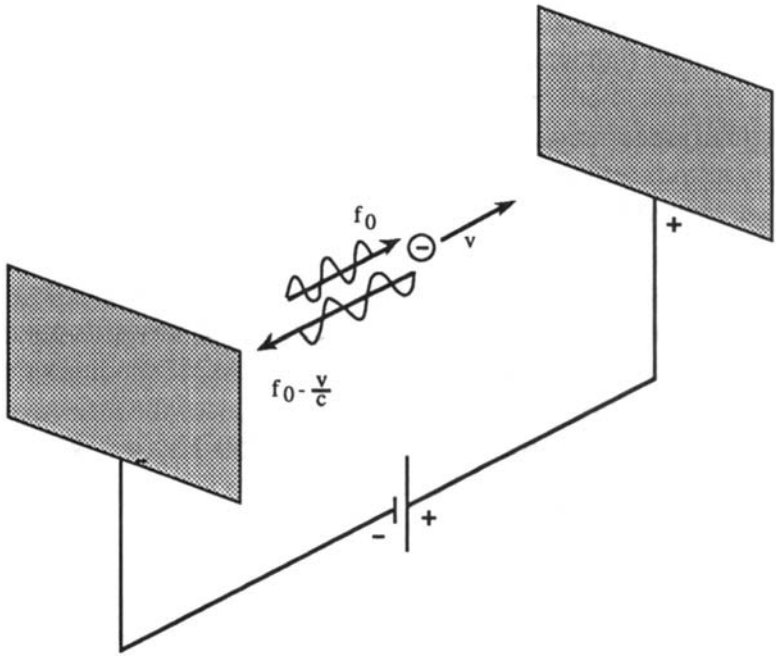


Figure 7.8. Principle of electrophoretic light scattering.

$$f' = (f + \frac{fv}{c}) - (f + 250) = \frac{fv}{c} + 250$$

The modulator has thus shifted the difference frequency; for example if the difference frequency is 100 Hz, and the modulator is driven at 250 Hz, the correlator will detect a frequency component at 350 Hz. This is of great value when studying small zeta potentials. If one beam were not frequency shifted in this manner, then zero zeta potential would correspond to zero frequency shift, and the electronics would need to have an infinitely extended low frequency response to detect it. If the modulator is driven at 250 Hz, zero mobility now corresponds to 250 Hz, which is much easier to detect than 0 Hz.

Using two well-defined beams in this manner has the additional advantage that the scattering is only analysed from the point at which the beams overlap. This is important since the apparent mobility of the particles depends on their position in the cell. The cell walls are

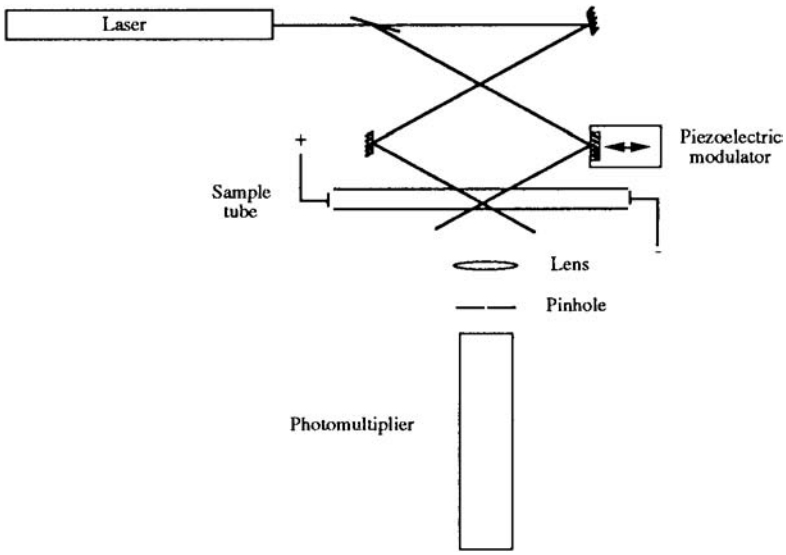


Figure 7.9. Electrophoretic light scattering apparatus.

invariably charged, and so attract an excess of negative or positive ions. These ions move in the electric field, and carry their solvent cages with them; since this causes solvent to move in one direction next to the cell wall, it must be returned in the opposite direction in the centre of the cell. This phenomenon is called *electro-osmosis,* and leads to a parabolic velocity profile of the solvent in the cell, which is independent of anything the particles might be doing. This solvent flow is superimposed on the particle movement, so that the measured particle velocity varies across the cell section (Figure 7.10); at any point the observed velocity is the vector sum of the fixed electrophoretic velocity and the electro-osmotic velocity at that point in the cell. Fortunately there is a point in the cell at which the electro-osmotic solvent flow is zero, and the measured particle velocity is then the true electrophoretic velocity. This point is called the *stationary layer,* and is always situated at a distance 14. 6% of the diameter from the cell wall in a cylindrical cell. If the two beams cross at this point, the zeta potential measured will be free of electro-osmotic errors.

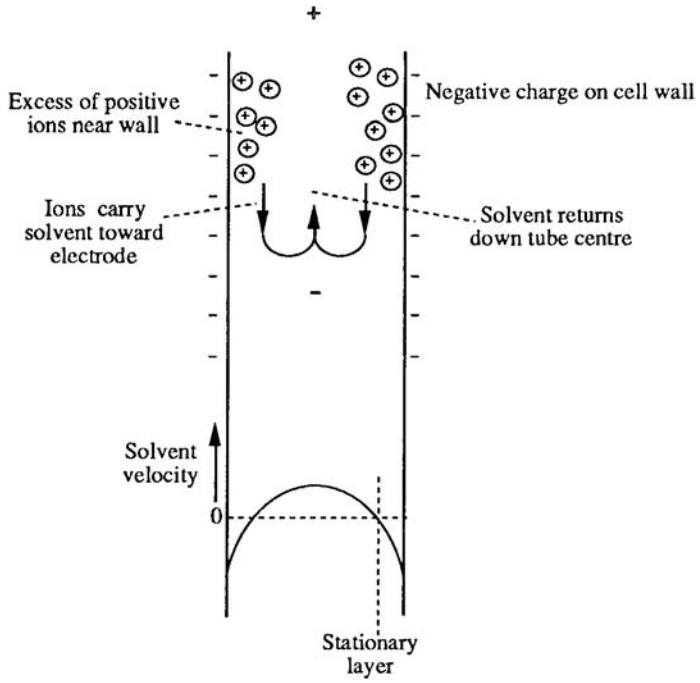


Figure 7.10. Variation in mobility in cell due to electro-osmosis.

If the velocity is measured at a number of points across the cell, the resulting velocity/position data can be fitted to a parabola; this is called a *van Gils plot,* and is a useful method of confirming the proper operation of the instrument. It also allows the magnitude of the electro-osmotic effect to be assessed. This can be important, since the contribution of electro-osmosis depends on how much charge is present on the cell walls; a clean glass or quartz cell has only a moderate charge, but if the cell is contaminated with a layer of polymer or surfactant, the charge can be many times higher. In theory the stationary layer position and electrophoretic velocity should not be affected by an increased wall charge, but if significant electroosmosis is present, a small error in beam position can lead to a large error in velocity. It can also give rise to a phenomenon known as *electro-osmotic broadening;* the scattering volume has a finite size since the laser beams

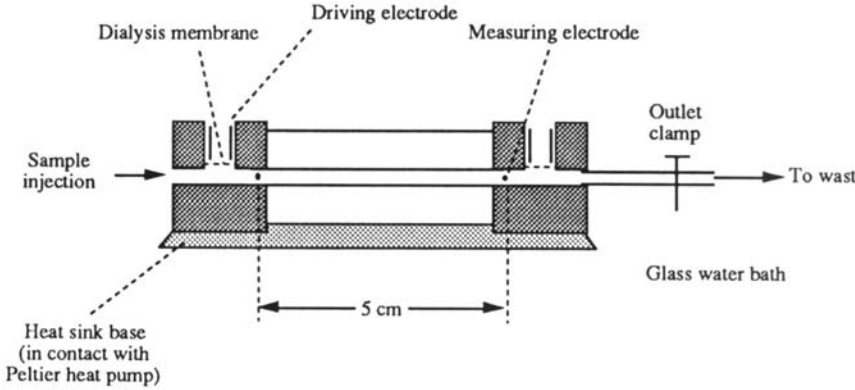


Figure 7.11. Sample cell of Zetasizer 2.

have a finite width, and so if the centre of the beam crossing point is set at the stationary layer, the front and back of the scattering volume will be slightly in front of, or behind, the stationary layer, and so have different velocities. This will cause the measured distribution of mobilities to be wider than the true distribution, to an extent that depends on the magnitude of the electroosmotic effect.

The electrical part of the experiment is also slightly more complex than the simple setup of Figure 7.9. A number of arrangements are in use, but we illustrate that used in the Malvern Zetasizers since we are most familiar with these instruments. The other main protagonist in the UK is the Coulter Delsa, which uses a different electrical cell arrangement. A plan view of the Malvern Zetasizer 2 cell is shown in Figure 7.11. The sample is contained in a horizontal cylindrical tube which is enclosed in a small water compartment for temperature control and optical matching. The two laser beams enter horizontally and cross at the front stationary layer. An electric field is generated by applying 100–150V across the two *driving electrodes,* which are platinum cylinders with about 1 cm^2 surface area; these electrodes are separated from the sample tube by *ion-permeable membranes,* which allow a current to flow, but keep the sample out of the electrode chambers (these membranes have been abandoned in the latest Zetasizer 4, a

significantly improved instrument). Since the driving electrode positions are ill-determined, the field is measured with a separate pair of *measurement electrodes,* two platinum pins 50 mm apart, one at each end of the drift tube.

If the field were continually applied in one direction, all the particles would end up at one end of the cell, and the electrodes would become badly polarized. Consequently the cell polarity is reversed every few seconds, so that the particles drift backwards and forwards up and down the tube. This would normally produce scattered light at (f_o−fv/c) when the particles were moving in one direction and (f_o+fv/c) and when they moved in the opposite direction. This would produce two Doppler shifted peaks at (250+fv/c) and (250−fv/c), i.e. symmetrically disposed about the zero mobility frequency. To prevent this, the slope of the modulator sawtooth drive is reversed as the cell polarity changes, so that the reflective surface now moves in the opposite direction and the Doppler effect *increases* the reference frequency. This changes the reference beam frequency from (f+250) to (f−250), and so the final frequencies are

$$f' = (f + \frac{fv}{c}) - (f + 250) = \frac{fv}{c} + 250$$

in one direction and

$$f' = (f - \frac{fv}{c}) - (f - 250) = -(\frac{fv}{c} + 250)$$

which are the same frequency, differing only in their phase.

A typical frequency spectrum for this experiment is shown in Figure 7.12. The scattered light contains a distribution of frequencies, from which we infer that the particles in the cell have a range of velocities. There are two possible reasons for this; the first is that a range of zeta potentials may be present in the sample. In this case the signal is said to be *hetereogenously broadened.* This is possible, but not as common as the alternative, which is that a range of velocities are present owing to the diffusional motion of the particles. This is termed *homogeneous broadening,* and this mechanism contributes to the frequency width in all

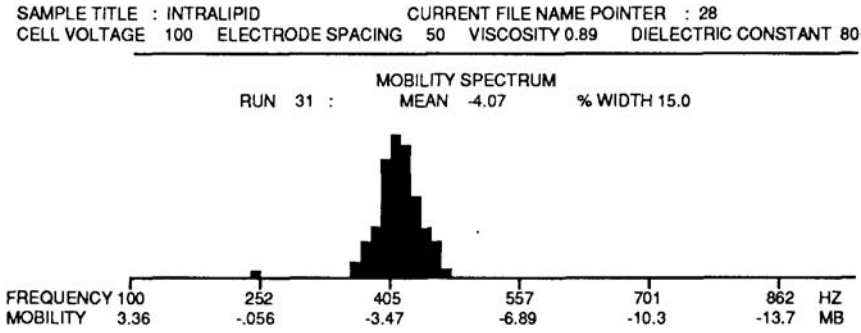


Figure 7.12. Distribution of mobilities for Intralipid 20% at pH 7.

samples. The Zetasizers 3 and 4 actually remove the majority of this diffusional broadening by measuring the correlation function in the absence of an applied field (the *zero-field measurement*) and then deconvolute the diffusional velocities from the electrophoretic velocity.

One of the first things you notice when measuring zeta potentials is that large particles provide sharp zeta potential distributions, while small ones are difficult to measure accurately since the frequency spectrum is very broad; the diffusional velocities may be as large as the electrophoretic velocity. This is due to the much larger spread of diffusional velocities for the smaller particles; in fact this is the same information in the frequency domain that PCS uses to measure the size by conventional autocorrelation; it has simply been shifted up the frequency scale by the Doppler effect. If the particles are monodisperse, they have a Lorentzian distribution of velocities, and so scatter a Lorentzian distribution of frequencies from the unshifted centre frequency. Fourier transformation of the Lorentzian lineshape yields the appropriate exponential correlation function and hence the particle size. Consequently some modern electrophoresis instruments (e.g. the Coulter Delsa) simultaneously measure zeta potential and z-average diameter. Because PCS obtains its information at a fundamental level by measuring the frequency spectrum of the Doppler shifted light, it is sometimes referred to as *quasielastic light scattering,* or QELS. (Acronyms are

apparently not subject to the requirements of the English language that U follows Q!)

Electroacoustic analysis

Recently a new approach to the measurement of zeta potential has appeared. This is the electroacoustic effect or electrosonic effect, and is used by Matec Applied Sciences in their ESA 8000 electroacoustic analyser (Babchin *et al.* 1989, O'Brien 1990). Although this is not a light scattering method, it is convenient to mention it here, since it is currently generating a lot of interest among colloid chemists. The method excites the sample with an alternating electric field of approximately 1 MHz frequency; this causes the charged particles to oscillate in a similar manner, producing an ultrasound signal in the system which can be measured with an ultrasonic transducer. Briefly, the larger the particle charge, the greater the amplitude of the ultrasound generated. Unlike laser Doppler techniques, the measurement is not absolute, and the instrument must be calibrated against a standard, normally Ludox colloidal silica. This technique actually requires a significant fraction of disperse phase (typically 1–20%) in order to provide a usable signal strength, and so can study concentrated colloidal suspensions, as well as suspensions in media of low dielectric constant. The most significant disadvantage of the method is that any ions present in the medium will also produce a resonant signal, which must be subtracted as a blank from the colloid signal. If the ion signal is large compared with the colloid signal, this subtraction may provide a significant error. We have used the method for the study of parenteral fat emulsions (Washington, in press). The electroacoustic effect has also been used for the measurement of particle size (Matec Applied Sciences Acoustosizer), although this instrument has been introduced only recently and at present little is known of its performance, although the manufacturers claim that it will cover the size range of 0.1–10 μm, and could potentially be a useful alternative to PCS.

Applications

Photon correlation has been widely applied to colloidal pharmaceuticals, particularly with the interest in drug delivery and targeting (e.g. Douglas *et al.* 1987). Numerous specific applications of correlation techniques, including the study of aerosols and the measurement of ciliary activity, are discussed by Dahneke (1983). McCracken and Sammons (1987) have compared the method (using a Coulter N4) to electron microscopy, ultracentrifugation, and size exclusion chromatography for the study of 70 nm liposomes, and concluded that, while the distributions were similar from the three techniques, PCS demonstrated a pronounced tail at larger sizes owing to the weighting of large vesicles by the sixth power law. Muller and Muller (1984) studied the oleate-cosurfactant microemulsion system with diameter 10–30 nm, and Roe and Barry (1983) studied micellar systems (deoxycholate and chlorpromazine) using an argon laser. It has also been widely used for emulsions; for example Merry and Eberth (1984) studied the production of oil in water emulsions by ultrasound using a Malvern 7025 system. The size of technetium-sulphur colloid used for nuclear medicine was studied by Pedersen and Kristensen (1981) who found good agreement with sizes measured by nucleopore filtration and electron microscopy.

The measurement of zeta potential is particularly valuable for the study of colloid stability in a predictive sense. An example of this is the study of total parenteral nutrition mixtures, which are suspensions of triglyceride emulsions in complex media including electrolytes, glucose, and amino acids. Zeta potential measurements allow the deconvolution of the various adsorption processes and allow the stability of these complex mixtures to be understood (Washington 1990)

REFERENCES

Babchin A.J., Chow R.S. and Sawatsky R.P. (1989). *Adv. Coll. Int. Sci.* **30** 111–151.

Byrne B.J. and Pecora R. (1976) *Dynamic light scattering.* John Wiley, N.Y.

Chu B. (1991) *Laser Light Scattering.* 2nd ed. Academic Press, N.Y.
Cummins P.G. and Staples E.J. (1987) *Langmuir* **3** 1109–1113.
Dahneke B.E. (ed) (1983) *Measurement of suspended particles by quasi-elastic light scattering,* John Wiley, N.Y., 32–78.
Douglas S.J., Illum L., Davis S.S. and Kreuter J. (1984). *J. Coll. Int. Sci.* **101** 149–158.
Douglas S.J., Davis S.S. and Illum L. (1987). *CRC Critical Reviews in Therapeutic Drug Carrier Systems* **3** 233–261.
Ford N.C. (1983) *Theory and practice of correlation spectroscopy.* In: Dahneke B.E. (ed), Measurement of suspended particles by quasi-elastic light scattering, John Wiley, N.Y., 32–78.
Gulari E., Gulari Er,. Tsunachima Y. and Chu B. (1979) *J. Chem Phys.* **70** 3965–3972.
Hunter R.J. (1981) *Zeta potential in colloid science.* Academic Press, London.
Koppel D.E. (1975) *J. Chem. Phys.* **57** 4814–4820.
Lanczos C. (1957) *Applied analysis,* Prentice-Hall, Englewood Cliffs, pp.272–280.
McCracken M.S. and Sammons M.C. (1987) *J. Pharm. Sci.* **76** 56–59.
McWhirter J.G. and Pike E.R. (1978) *J. Phys. A: Math. Nucl. Gen.* **11** 1729–1745.
Merry J. and Eberth K. (1984) *Int. J. Pharmaceut.* **19** 43–52.
Muller B.W. and Muller R.H. (1984). *J. Pharm. Sci.* **73** 915–918.
O'Brien R.W. (1990) *J. Fluid Mech.* **212** 81–94.
Ostrowsky N. (1988) *Particle characterization by photon correlation spectroscopy* In: Lloyd P.J., Particle size analysis 1988, J. Wiley and Sons.
Ostrowsky N., Sornette D., Parker P. and Pike E.R. (1981) *Opt. Acta* **28** 1059–1070.
Pedersen B. and Kristensen K. (1981) *Eur. J. Nucl. Med.* **6** 521–526.
Provencher S.W, Hendrix J., De Maeyer L. and Paulussen N. (1978) *J. Chem. Phys.* **69** 4273–4276.
Provencher S.W. (1979) *Makromol. Chem.* **180** 201.
Roe J.M. and Barry B.W. (1983) *Int. J. Pharmaceut.* **14** 159–171.
Stock R.S. and Ray W.H. (1985) *J. Polymer Science: Polymer Physics.* 23 1393–1447.
Ware B.W. and Flygare W.H. (1971) *Chem. Phys. Letts.* **12** 81–85.
Washington C. (1990) *Int. J. Pharmaceut.***66** 1–21.
Washington C. (1992) *Int. J. Pharmaceut.* (in press).

8
Particle size analysis of aerosols

An aerosol is a small particle or droplet suspended in air. There are two main areas in which aerosols are of interest in pharmaceutics; these are the counting and analysis of airborne particle contamination, in the validation and quality control of clean rooms and sterile facilities, and in the study of particle and droplet sprays. Although these sprays have a wide variety of uses, such as application of liquids to surfaces, their most highly developed form is the *metered dose inhaler,* a device for delivering a small amount of aerosolized drug to the pulmonary system. Much of the technology discussed in this chapter has been applied to the engineering of these devices to optimize the delivery of drugs by inhalation. Of course, these methods have a much wider sphere of application outside pharmaceutics; much of the early interest in aerosols was derived from attempts to explain atmospheric diffraction phenomena, which are now fairly well understood, (see for example Craig Bohren's recent volume *'Clouds in a glass of beer'* (Bohren 1987) for a popular exposition of this area). Other applications include studies of dust explosions in mines, industrial hygiene and pollution, and the application of paints and coatings.

A wide range of techniques have been applied to the study of aerosols. We have already discussed the application of laser diffraction instruments in Chapter 6, and will not examine these methods further here. In this chapter we will concentrate on techniques that have specific application to aerosols. Firstly, however, it is valuable to understand something about the way that moving particles in a gas stream behave under flow conditions, since this influences

our studies in a number of ways. Several particle size analysers are based on particle behaviour in flowing gases, and this is also important in achieving an unbiased sampling for analysis of suspended solids in air.

INERTIAL IMPACTION

Particles in a uniformly moving gas stream are carried with the gas, and will also settle in it under the influence of gravity. As long as the gas stream is moving uniformly, the particles will be at rest with respect to the gas flow, i.e. they will have the same velocity, apart from any small component due to gravitational settling. However, if the gas flow changes direction, for example to go round a bend in a pipe, the particle will not immediately follow the direction change. From the particle's frame of reference, it is subjected to a hydrodynamic force due to the non-uniform gas flow around it, and so will begin to accelerate in the direction of the new gas flow. The force on the particle will be dependent on the particle cross-sectional area, gas viscosity, and is in general complex, but a massive particle will accelerate more slowly than a small one, and a dense one more slowly than a light one. The result of this is that, where the gas stream changes direction, the particles tend to carry on moving in a straight line, and may impact on any object placed in their path, despite the fact that the gas flows smoothly around it. This phenomenon is called *inertial impaction.*

A common example of this effect can be seen after a motorway journey. We have all seen pictures of smoke airflow around cars in wind tunnels (usually in TV advertisements), in which smoke trails flow along the bonnet of the car, smoothly up the windscreen and over the roof. The same cannot be said of the insects in the airstream, which hit the windscreen with some force. After a period of driving in town, the car is littered with the remains of the larger insects, and after a motorway journey, even smaller ones are collected. This effect is so well-known that it has originated several popular jokes about the rearrangement of the fly's anatomy in the process. Impaction can be easily

demonstrated using a hand aerosol spray, such as the type used to apply cleaning solutions to surfaces. If a spray is directed at a dark-coloured surface with good illumination, it will be easy to see the cloud of finer droplets that avoid the surface, while the larger ones impact.

Even at this simple level we can see that the impaction process is selective of particle size. The higher the gas stream velocity, the smaller are the particles that impact on the solid surface. The impaction process has been studied in great detail, both for the design of particle size analysers, and to study the deposition of airborne particles (e.g. in the respiratory tract). Unfortunately attempts to produce an *ab initio* description of impaction rapidly lead to mathematical complexity. This is partly due to the complex geometry of even the simplest experiment, and the difficulty of describing the flow of non-uniform fluids or gases. The situation is best illustrated using an example of a common experimental impaction geometry, the jet and plate impactor. This simplified geometry is often used as a model to infer the properties of more complex geometries, often with questionable success.

Figure 8.1 shows the impaction arrangement. A gas containing suspended particles flows through a cylindrical hole (the 'jet') under laminar flow conditions. The stream of gas then encounters a smooth planar surface perpendicular to the jet axis. The larger particles impact on the plate, while the smaller ones follow the streamlines and are swept clear. In this arrangement, the probability of adherence to the plate is described by the *inertial impaction parameter,* K, given by:

where u is the velocity of the jet stream relative to the surface, D_p is the particle diameter, ρ is the particle density, μ is the gas viscosity and D_c is the diameter of the jet. C is a correction factor called the *Cunningham slip factor,* and is given by:

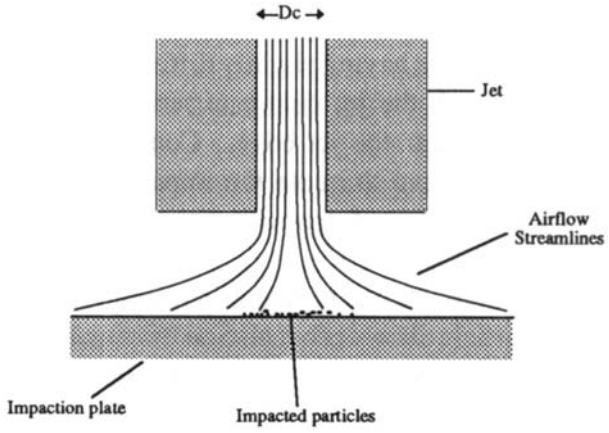


Figure 8.1. Jet and plate impactor.

$$K = \frac{C\rho u D_p^2}{18\mu D_c}$$

$$C = 1 + \left(\frac{1.6 \times 10^{-5}}{D_p}\right)$$

The slip factor is interpreted as representing the ability of the particles to slip 'between' the gas molecules as the particle size is reduced below the mean free path in the gas. The particle diameter, D_p, should more properly be referred to as the aerodynamic diameter if the particle is a sphere, or the aerodynamic equivalent diameter for an irregular particle.

In an empirical study of this experimental arrangement, an impaction parameter of greater than 0.14 leads to deposition on the plate, while a particle with a smaller impaction diameter would not be deposited. There is no clear cutoff between particles which are deposited and those which are not; this is largely due to the different conditions experienced by particles having slightly different paths, for example, those passing through the jet centre experience slightly different conditions to those near the jet edge.

Consequently, for a given arrangement, it is normal to quote the impaction parameter which leads to a 50% probability of deposition.

It is important to note that the probability of impaction is dependent not only on the particle diameter but also on its density. Consequently dense particles will impact more easily than light ones of the same diameter, which seems intuitively reasonable. Because of this effect, the cutoff diameter of the jet and plate arrangement is usually specified in terms of a particle of unit density. If a particular jet and plate have a 50% cutoff diameter of 1 μm for a particle with a density of 1 g cm^{-3}, (e.g. a water droplet) then the 50% cutoff for a particle of density 6 g cm^{-3} (for example a metal oxide dust) would be 0.408 μm. We should note that this could be of importance for the pulmonary delivery of drugs; low density particles will suffer smaller losses due to impaction in the upper airways than will high density ones.

SAMPLING AEROSOL GASES

Inertial impaction can pose significant problems for the handling of aerosols prior to measurement. This is rarely a problem for the study of inhalers or other aerosol-generating devices, since these can usually spray the aerosol directly into the measuring instrument. It is much more significant in the sampling of suspended solid particulates in an environmental situation (for example the monitoring of a clean room) in which a particle count may be required at a position in which the instrument cannot be placed, or at several positions. In these circumstances it is usual to sample the gas with a nozzle of some sort, attached to a pipe leading to the instrument.

The basic requirement for sampling particulates in gases is that the gas flow velocity in the nozzle should be equal to the flow velocity of the undisturbed gas. This is known as *isokinetic sampling.* It is most important when rapidly flowing gases are sampled, in order to ensure that the larger entrained particulates do not impact on the sampling

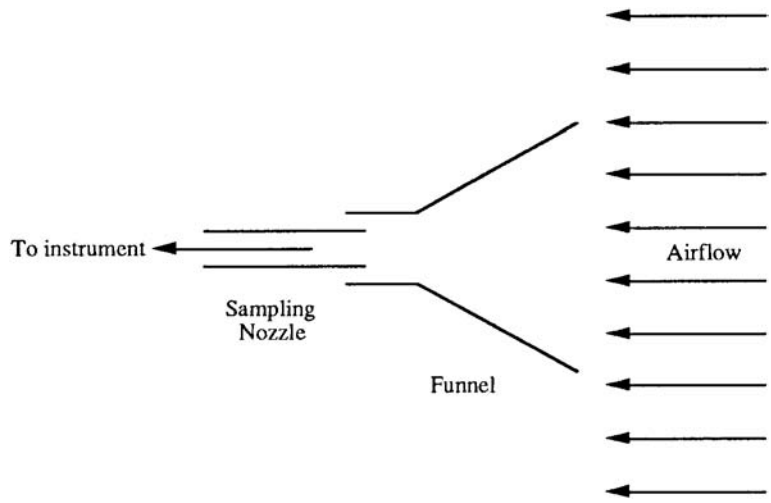


Figure 8.2. Nozzle arrangement for isokinetic sampling.

apparatus. There are a number of designs for sampling nozzles in the literature; several are described in BS 893 and BS 3405, and the area has been covered in detail by Vincent (1989). Much of the literature in this area is concerned with the analysis of flue gases, which are in rapid flow, but are equally applicable to the validation of air flows from laminar flow filters. The design and validation of clean rooms, with particular reference to microbiological quality, has been reviewed by White (1990).

In order to achieve isokinetic sampling it is necessary to know the airspeed and direction of the gas flow. Small hand-held draught meters are available for this purpose. Since the sampling instruments work at a fixed flow rate (usually 1 cu.ft/min), it is possible to calculate the orifice diameter of a circular nozzle which will entrain this volume of air. This is then passed into a funnel which slowly converges to match the inlet pipe of the particle counter. An improved arrangement assumes that this funnel will impact some particles near the edge of the flow, and so takes a larger volume of air, which is then isokinetically sampled from the end of the funnel with a further straight nozzle (Figure 8.2).

The pipework leading the particle-laden air to the instrument must be carefully selected and sited in order to achieve optimum performance. Impaction is normally minimized by ensuring that the bend radius is at least 150–300 mm; also the bore of the tubing is chosen to keep the particles in turbulent flow, so that no stagnant areas build up. The tubing must be made from a material which does not develop a static charge, which would otherwise remove a significant fraction of the particulates; radioactive sources can be used to prevent charge buildup. Sampling through long tubes is only practical if the particles are very small, since the larger ones will impact on the tubing. If correctly installed, transmissions of approximately 50% can be expected for 5 μm particles, and 90% for 1 μm, through 20–30m of tubing. This is adequate for clean room measurement, since the majority of particles are extremely small under these conditions. If larger particles are present (for instance in the study of a gas effluent from a dust extraction facility), it is usually necessary to take the instrument to the particle source. Several simplified instruments are available (e.g. filters and small impactor stacks) and are ideal for this purpose.

GENERATION OF CALIBRATION AEROSOLS

It is not straightforward to produce calibration aerosols with a precise absolutely known size. Consequently most workers are content to use the calibration that is supplied with their instruments. However, the introduction of more sophisticated devices for aerosol measurement, usually those involving light scattering and on-line computing, have provided a need for calibration and test systems, in order to diagnose faults and verify that the size calibration remains constant.

The simplest way of providing calibration particles is to use a pressure or ultrasonic nebulizer to nebulize polystyrene microspheres suspended in a suitable solvent (water, or ethanol, which evaporates more readily). The only problem with this approach is that all the solvent may not

have evaporated before sizing. This can usually be detected as a major error in the expected size distribution, and can be prevented by using a reasonable length of wide bore connecting pipe between the instrument and the nebulizer, to provide opportunity for evaporation.

An alternative method is to use a monodisperse aerosol generator. Several methods are available for the production of aerosols with well-defined sizes, the most common being the rotating disc generator and the vibrating orifice generator. The rotating disc generator (e.g. the Stag Mk. 2 from Research Engineers) has a small horizontal steel disc which is levitated and spun at several hundred revolutions per second by compressed air. The liquid to be aerosolized is slowly fed onto the disc centre by a syringe pump, where it is spun out to the disc edge and detaches as uniform droplets of a few micrometres diameter.

The vibrating orifice generator (e.g. the TSI 3450) operates by pumping a stream of liquid through a small (20 μm) orifice. Normally the stream would break up into droplets of varying sizes, but if the orifice is rapidly vibrated at several kilohertz by a piezoelectric driver, the stream breakup can be controlled and monodisperse droplets are obtained. This device also produces droplets of a few micrometres diameter. Smaller droplets can be produced in both instruments by aerosolizing mixtures of volatile and involatile liquids, or solutions of solids in volatile liquids.

Impaction particle sizers

A wide range of inertial impactors have been described in the literature, mostly concerned with the measurement of particulate pollution; much of the early work on such devices was performed by May (1945). The most important devices from the pharmaceutical viewpoint are the two-stage impinger and the Andersen sampler, which are both widely used for the study of inhalation aerosols.

The two-stage impinger

A number of simple devices are in use for the classification of aerosols into a small number of broad size channels; the intention of these methods is to separate particles on the basis of their possible behaviour in the respiratory tract. They are not really size analysis methods, since the classes of particles separated are very broad and there is usually no attempt made to correlate particle size with deposition in a particular part of the apparatus. It would be more appropriate to consider them as quality control methods.

The most important techniques from a pharmaceutical viewpoint are those which are described in the British Pharmacopaeia (1988) Appendix 17C for deposition of the dose emitted from an inhalation device (see e.g. Hallworth and Westmoreland 1987). The B.P. 'apparatus A', or two-stage classifier, is shown in Figure 8.3. The B.P. provides a rigid dimensional specification for this device, which is assembled from modified Quickfit apparatus. It contains two impaction stages, the upper one (the round bottomed flask) operates at low velocity and traps the largest particles, while the lower stage (the conical flask) operates at much higher velocity. The intention of the apparatus is that the upper flask would trap all the larger particles which would impact in the upper airways, while the lower flask collects the respirable fraction, i.e. that which would penetrate the deeper airways. The largest particles will impact on the throat; *in vivo* these would normally be swallowed. The B.P. procedure specifies analysis of the material in the conical flask to measure the deposited dose, but it is equally useful to wash out the throat and the upper impingement chamber separately to obtain a rough idea of the particle size distribution.

To use the device, a pump is connected to the lower vacuum port, and a stream of air of 60 ± 5 l min^{-1} is drawn through. This should be calibrated using a gas volume meter. The aerosol is then fired ten times into the throat, after which the drug content of the lower chamber can be assayed. The B.P. procedure makes no mention of

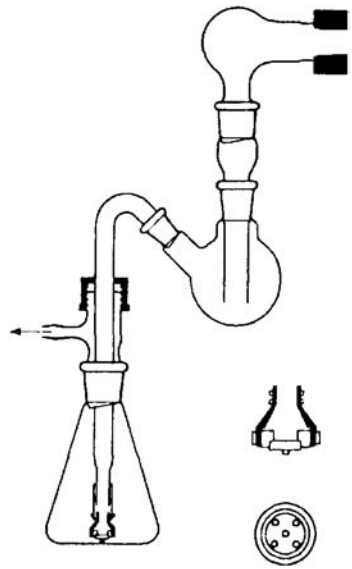


Figure 8.3. Two-stage impinger type A (from B.P. 1988).

temperature or humidity of the air, and so it is normally used at room temperature, 20–25°C. Since the classes defined by the two impactor stages are so broad, a small change in temperature is unlikely to have a major influence on the results.

Recognizing the difficulty of manufacturing glass apparatus to rigid dimensional tolerances, the B.P. has introduced a second apparatus based on similar principles, (their 'apparatus B'), which is constructed from totally machined components (Figure 8.4). This uses an impingement chamber of similar performance to the apparatus A upper chamber, in order to collect coarse material, after which all the remaining material is collected on a disc of glassfibre filter paper. This is used in exactly the same way as apparatus A, with a gas flow of 60 l min^{-1}

The Andersen sampler

The Andersen sampler is a widely used instrument for the size analysis of particles in gases; it was specifically designed to cover the size range of importance in pulmonary

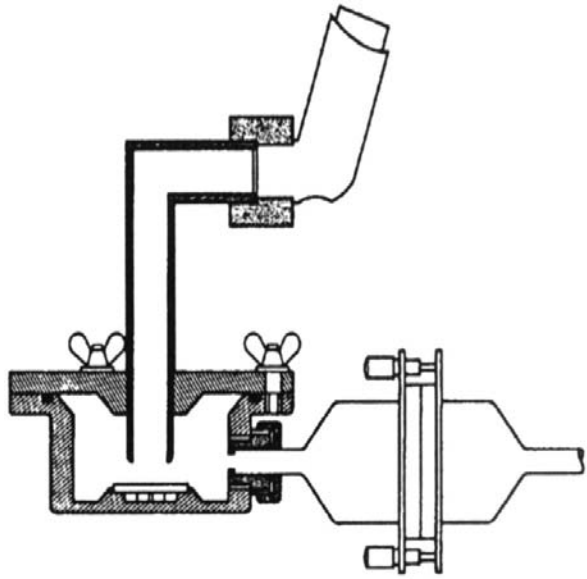


Figure 8.4. Two-stage impinger type B (from B.P. 1988).

deposition, i.e. from 10 μm to 0.1 μm. The behaviour of inhaled particulates of this size is reasonably well understood as a result of extensive studies for the development of inhalation aerosols (see, for example, Gonda 1981) (Figure 8.5).

A diagram of the instrument is shown in Figure 8.6. It consists of a stack of plates (usually 8–10) mounted in rings, somewhat akin to a stack of sieves, each plate carrying an array of finely drilled holes. The holes in the top plate are the largest, of diameter 1.5 mm, and successive plates carry holes of decreasing diameter, the lowest being 0.25 mm. The final plate is unperforated and collects the last class of particles. In each stage, below the holes in the plate, is a circular

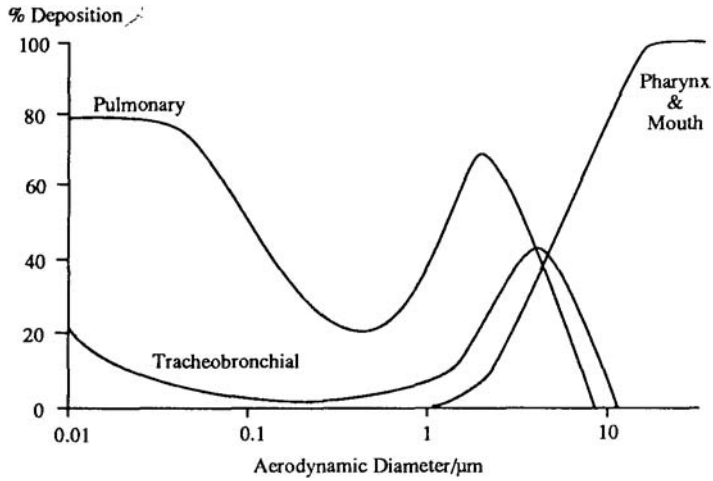


Figure 8.5. Deposition of participates in the respiratory tract.

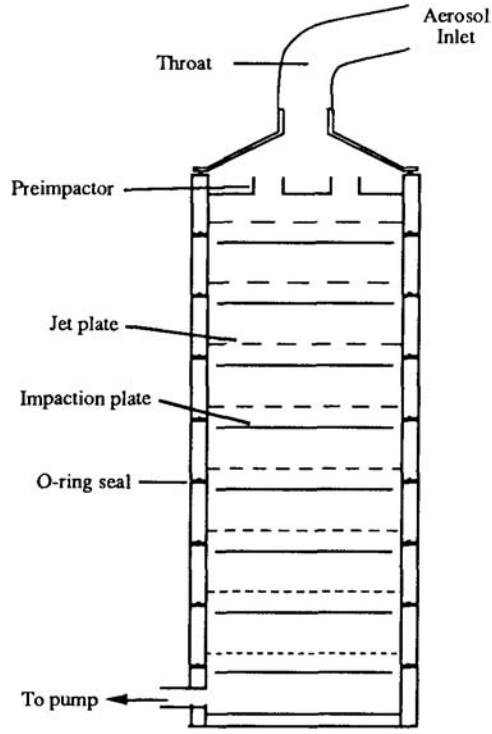


Figure 8.6. The Andersen sampler (Note that the size of the holes has been emphasized for clarity).

collection plate that serves to impact the particles from the previous set of holes. The bottom of the stack is connected through a flow valve to a vacuum pump, which draws air through the stack of plates. At the top is an inlet orifice (normally called a *throat* in inhalation studies, since that is what it is meant to simulate). The first plate contains a number of coarse holes to separate out particles above the range of the top impaction stage; it is usually called the preimpaction stage.

The aerosol is sprayed into the throat and drawn by the airflow through the holes in successive stages. The diameter and number of holes are chosen so that the airflow velocity increases down the stack, with the upper (larger) holes having the lowest velocity, and successive stages having an increasing velocity. As a particle passes down the stack it will ultimately reach a stage at which the jet provides sufficient velocity to impact it on the next lower plate. The size range collected by each plate is found from a calibration graph supplied with the instrument, which looks something like Figure 8.7. As pointed out earlier, the stages of an impactor do not provide a sharp cutoff, and so the curves in Figure 8.7 represent the probability that a particular particle size will be impacted on a particular stage. The size range of a stage is normally quoted as the 50% probability points of the stage and the one above it; thus in our example, stage 4 would collect the size range 1.05–2 μm of particles of unit density. The entire stack of the most commonly used sampler (The Andersen Mk.2) covers the size range 9–0.45 μm. Larger particles are collected on the preimpactor, and smaller ones may be collected by drawing the effluent air from the stack base through a filter.

The calibration of the instrument assumes that it is operating which particles of unit density. If this is not the case, the same calibration graph describes the relationship between the stages, but the sizes collected are different. Since the impaction diameter is inversely proportional to the square root of the density and each stage operates at a fixed impaction parameter, the correct size classes can be found simply by dividing the unit density classes by the square

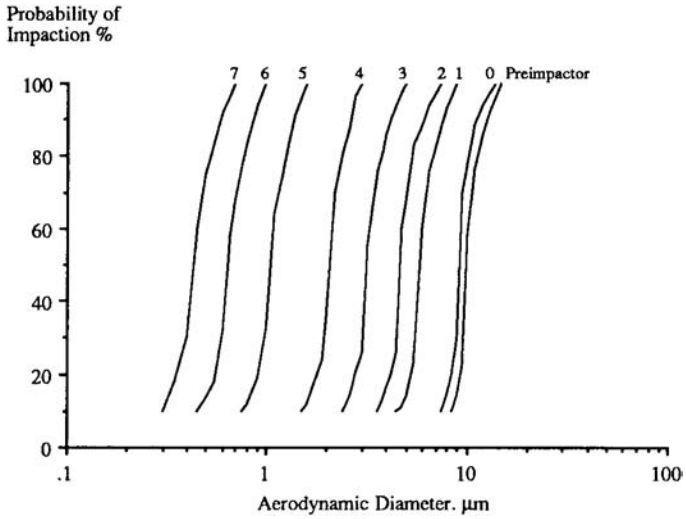


Figure 8.7. Size ranges of the Andersen Mk. 2 impactor.

root of the density. Thus the earlier example of stage 4, which collected 1.05– 2 μm of unit density particles, would for example collect from 0.74– 1.41 μm particles of density 2. 0.

Setting up and calibration

The main problem with the Andersen sampler is that it can be time-consuming to set up, and has to be cleaned between each experiment; thus it does not have the sample throughput of a light scattering instrument. Before use it should be dismantled, cleaned and dried. If the impacted material is to be measured gravimetrically, the tare of each impactor plate should be measured. The o-rings sealing the stages together should be examined for wear, and the whole stack carefully assembled, ensuring that each collection plate is properly seated on the supports provided for it. The stack is then held together by a set of spring clips (it helps to have an assistant to put these on).

Assuming that the instrument is uncalibrated, it is then necessary to adjust the pump to pass exactly 1 cubic foot per minute of air at 1 atmosphere (1 acfm) through the

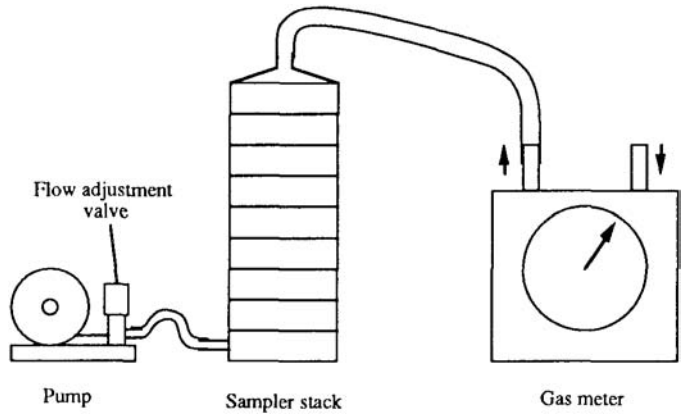


Figure 8.8. Gas flow calibration of the Andersen sampler.

stack. This is best performed using a calibrated volumetric gas meter, of the type commonly found in physiology laboratories for measuring volumes of respired gases. This is connected to the stack throat so that gas drawn through it is measured; the pump should be run for at least 3 minutes to obtain a reasonable accuracy (Figure 8.8). The flow valve is adjusted and the volume repeatedly measured; it should be possible to obtain consistent volume readings to ±2%. The gas meter can then be disconnected, and after a short pause to allow equilibrium flow to be established, the aerosol device can be fired into the throat of the sampler. The duration or number of doses depends on the sensitivity of the assay used to measure the amount of material deposited on the plates, and must be determined by trial and error.

The sampler can then be dismantled for assay. This can be performed gravimetrically, which is popular for pollution studies in which the total solids are measured, but often the amounts deposited from an aerosol device are too small to be measured in this way. If gravimetric assay is used, it is usual to place glass fibre filter discs or aluminium foil discs on the collection plates, since these have a much lower mass than the stainless steel plates. Glass plates can also be used; these are useful if the deposited material is to be examined microscopically, but it may not be possible to

correlate the optical particle size with the aerodynamic size, since the particles may have been damaged on impact.

Errors of Andersen sampler analysis.

Probably the most significant source of error is variation in the temperature of the air passing through the stack; this is a problem since the viscosity of air is strongly temperature-dependent. If the viscosity of the air changes, the pumping rate will change and the airflow valve will have to be readjusted using the gas meter. More significantly, the gas viscosity is a variable of the impaction parameter, so the calibration of the stack is only correct at a single temperature, normally 20°C. For this reason it is a good idea to install the instrument in a room with a reasonably constant temperature that does not get direct sunlight (a temperature-controlled room is ideal but expensive).

A further error can arise if too great a loading is placed on any one plate. If the plate is examined under the microscope, the particles can normally be seen as small spots on the plate, beneath the corresponding jets. If the piles become too large, they will alter the impaction characteristics of the surface, and may themselves be swept away by the airflow. This is called re-entrainment, and can be a significant source of error if the sample loading is too large. The manufacturers recommend an upper limit of 10 mg per plate, and if this is exceeded significantly, re-entrainment will occur. In order to prevent this, and thus to allow greater plate loadings, MSP have produced a uniform deposit impactor, in which the impaction plates rotate beneath the jets, thus spreading the particles uniformly on the plate. The Andersen sampler uses a regularly drilled array of holes, while the MSP impactor uses irregularly spaced holes to ensure that the entire collection plate is covered.

A significant source of error arises from so called *wall losses* or *internal losses.* As the particles pass through the instrument from one stage to the next, some of them can impact on the walls of the sampling stack and be lost from

the analysis. Early versions of the Andersen sampler suffered in this respect, but it is minimized in modern samplers; this problem was discussed by May (1975) and an impactor stack based on his design is made by Research Engineers of London.

The sampler should be cleaned with a suitable solvent immediately after use, and dried, preferably in a warm oven. If the plates become blocked, it is unwise to use ultrasonic cleaning frequently, since the jets will gradually become enlarged by cavitation, and the calibration will be lost.

A further source of error is the possibility that the impacted particles will not adhere to the surface, but will bounce off. This phenomenon is dependent on the properties of the material being sized; for example, polystyrene microspheres are particularly susceptible to loss in this manner. Rao and Whitby (1978a, b) have studied this problem and propose two possible solutions. The first is to change the nature of the collection surface, since they found that glass fibre filters had better retention than the polished metal plates. The second, more widely used approach, is to coat the plates with a thin layer of grease or oil. It is only necessary to apply a very thin film; most impactor manufacturers have a recommended method, often involving spraying a tiny amount of silicone oil on the plates. Dipping in oil which has been diluted with a volatile solvent is an alternative method.

Several other manufacturers make similar instruments; Phillips *et at.* (1990) have compared the Andersen model II with the Delron DC 16 and a 2-stage impinger, and found that the impactors were consistent with each other but that the 2-stage impinger overestimated the respirable fraction. It appeared to be most important that the aerosol had evaporated fully prior to being drawn through the impactor stack, so the throat and inlet design should be standardized. Similar results were found by Hallworth and Andrews (1976) who compared a multistage liquid impinger with the Casella Cascade Impacter and the Cascade Centripeter (Bird and Tole Ltd.).

OPTICAL COUNTING METHODS

We have already considered (Chapter 6) the use of laser diffraction methods for aerosol size analysis, and will not discuss this method further. There are at least two other technologies available for the size analysis and counting of airborne particles using optical techniques. These are scattering intensity measurements, and the supersonic time-of-flight technique used in the new Malvern Aerosizer.

Light scattering intensity measurements

Several instruments are available for the measurement of particle size via scattered light intensity. The principle of these instruments is very simple; the particles are passed in an air jet through a light source (usually a laser beam), and in doing so scatter a pulse of light whose duration depends on the time taken to pass through the beam, and whose intensity is dependent on particle size. In order to size the particles, it is necessary to detect the light pulses with a suitable detector, usually a photomultiplier, and perform a pulse height analysis in a similar manner to that used to collect and analyse the pulses from a Coulter counter. Similar instruments are used as particle counters in liquids, such as the Hiac counter.

The major drawback with this technique for precision size analysis lies in the final step, that of calculating the particle size from the pulse height. We have already seen something of the complexity of light scattering theory in Chapter 6, and in practice it is not possible to relate the particle size to the scattered intensity in a simple manner. The very smallest particles (those much smaller than the wavelength of light) scatter light in all directions isotropically, and the scattered intensity is proportional to the sixth power of the particle radius (Rayleigh's Law). As the particle becomes larger, the scattering is no longer isotropic, and so the scattered intensity at the detector will depend on the exact detector geometry, i.e. its angle from the main scattering beam, and its area or solid angle subtended from the particle.

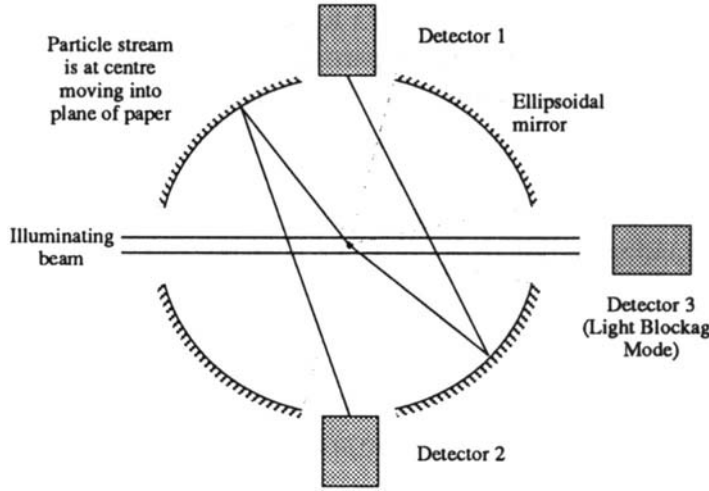


Figure 8.9 Typical light collection arrangement for optical zone counters.

From Figure 6.1, it is possible to imagine a small detector situated at some intermediate angle such as 45°, which would detect increasing amounts of light for increasing particle diameters, then for further size increases would register little or no increase (or in some cases even a decrease) in intensity as the light became scattered into a narrow forward-pointing cone. For this reason, the use of scattering intensity to measure particle size has been criticized; Kerker examined a number of early instruments operating on this principle, and found some significant discrepancies between them. The method works well for Rayleigh scatterers, which have a smooth dependence of scattered intensity on size, and whose isotropic scattering causes detector geometry to be unimportant. In practice the system is only approximately linear for particles with sizes up to about the wave-length of light. This is because the particles have a refractive index very different to their suspending medium (air) and so the Rayleigh-Gans-Debye approximation cannot be used. This is not the case for instruments which count particles in liquid by similar means, since the refractive index difference may not be so marked in this situation. As a result of these limitations, the

size analysis channels used in particle counters are rather coarse, particularly above 1 μm. Typical channel limits might be 0.2, 0.3, 0.5, 0.8, 1, 2, 5, 10 μm. The lower size limit is set by the rapidly decreasing scattering intensity from a Rayleigh scatterer; it is possible to reach 0.1 μm with high-power laser sources, but most instruments opt for a 0.2–0.3 μm limit.

In order to remove as much of the detector geometry dependence of scattering intensity as possible, the instruments gather light scattered over a wide range of angles, often with a dual concave mirror arrangement (Figure 8.9). The mirror gathers light scattered over a large solid angle and focuses it onto the detector. Some instruments (e.g. the Malvern Autocounter) have two such collection and detection systems, which are run in coincidence mode, a particle only being counted when both detectors return a signal. This allows spurious photomultiplier pulses (dark counts or cosmic ray events) to be rejected from true particle scattering signals.

AERODYNAMIC PARTICLE SIZERS

The Malvern Aerosizer

This is a recently introduced instrument for the size analysis of particles based on their behaviour in a high velocity airstream. Figure 8.10 shows the experimental arrangement. The particles in an airstream above the nozzle are sucked through the nozzle by a vacuum on the lower end. The acceleration of the particles in the nozzle depends only on their mass, as long as the transit time is sufficiently short that they do not collide during passage through the nozzle. They emerge from the nozzle with a range of velocities, the smallest particles having the fastest velocity, and the largest particles the smallest velocity.

In order to measure the particle velocity, the accelerated stream is passed through two laser beams. The point at which the particle stream passes through each beam is monitored by a separate photomultiplier. Thus an individual

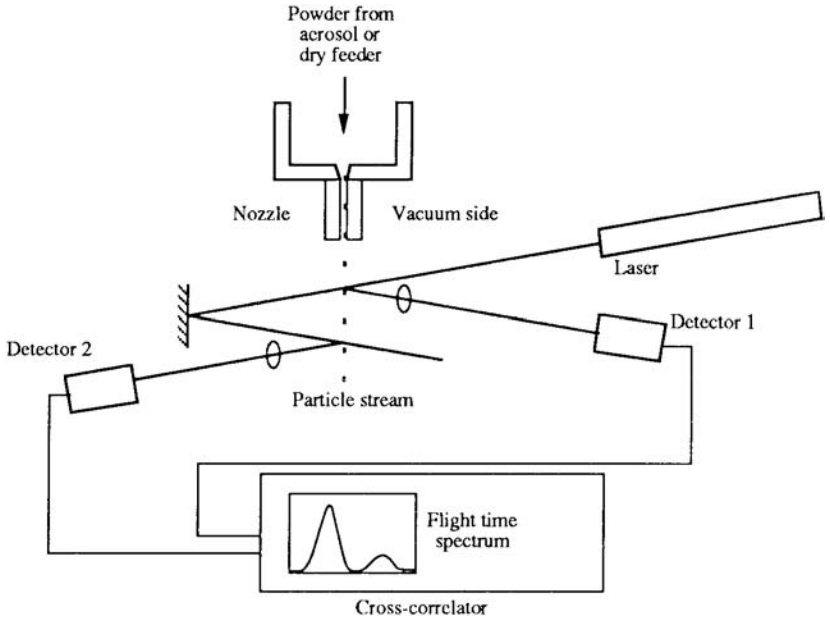


Figure 8.10. Malvern Aerosizer.

particle will produce a pulse first from one photomultiplier, then from the second, after a delay which depends on its velocity. By measuring the time delay, the velocity can be calculated, and hence the particle mass can be inferred. This can be converted to aerodynamic mean diameter using a suitable value for the density.

In practice the transit times of individual particles are not measured, since this would lead to an unacceptably low count rate; it would be necessary for each particle to clear the second detector before another could pass the first. Instead the signals from the two detectors are fed to a correlator which determines the cross-correlation function of the two signals. This is very simply related to the transit velocity; for example, if a population of particles with a velocity delay of 10 ms is present, a peak will be observed in the correlation function after 10 ms. The correlation function provides the full distribution of transit velocities, and hence a particle size distribution can be obtained.

Since the nozzle exit velocity is proportional to the mass of the particle, a small spread in particle size leads to a large velocity difference. This gives the instrument a very high resolution, which is particularly useful in the 0.5–2 μm region, where diffraction instruments have poor performance. The measured size range depends on the particle density, but 0.5–100 μm is typical.

A similar instrument is manufactured by TSI (the APS 33B Aerodynamic Particle Sizer). This instrument has a size range from 0.5 to 30 μm, but does not use correlation techniques to determine the velocity spectrum. Instead it measures particle velocity directly, so only one particle can be in the sensor at any time. Consequently it cannot measure particle concentrations as high as those measured by the Aerosizer, or coincidence effects will occur.

Phase Doppler Anemometry

The phase Doppler technique is a relatively recent method for the study of aerosols which is currently generating considerable excitement. Originally developed for the study of gas and particle flow in remote hostile environments such as combustion plants and jet engines, it is also applicable to the measurement of particle behaviour in the respiratory tract.

The basic apparatus for phase doppler anemometry resembles that used for zeta potential measurement in liquid samples (Chapter 7) and the principles of the two systems are fairly similar. A laser transmitter (usually a helium-neon or air-cooled argon-ion laser) is split into two parallel beams which are focused by a lens so that they cross at a well-defined point some tens of micrometres across, which may be as much as a metre from the laser transmitter. Where the two beams cross, interference fringes form, and a particle passing through the fringes will scatter a light signal, the overall envelope of which will be gaussian (owing to the gaussian section of the input beams), and will be modulated by a high frequency oscillation as the particle alternately crosses the peaks and troughs of the interference

pattern. As in the Zetasizer, one of the beams can be modulated by a piezoelectric drive or a Bragg cell, in order to make the fringes move in one direction or the other; this allows the direction of movement of the particle to be determined. For example, if the interference fringes were stationary, a 1 m s^{-1} particle travelling in an upward direction would scatter the same frequency as a 1 m s^{-1} particle travelling in a downward direction. If the interference fringes are, for example, moving upwards, a stationary particle will scatter at the fringe movement frequency, upward moving particles will scatter a lower frequency, and downward moving particles will scatter a higher frequency. This is exactly analogous to the use of piezoelectric frequency shifters in the Zetasizer, which allow the sign of the particle charge to be determined.

The light scattered from the particles is collected by a receiving lens, usually of large diameter for high signal-to-noise ratio, and to allow the receiver to be situated a reasonable distance from the beam crossing point. Because the particles in aerosols are generally larger than those in suspension colloids, and have a higher relative refractive index since they are suspended in air rather than water, the scattered light signals from individual particles can be detected, and the velocity of each particle measured from the high frequency component of the scattering signal.

The particle size information is obtained in a rather more complex manner. The receiver consists of not one photomultiplier but three, with prisms which send the light signals from separate parts of the receiving lens to each photomultiplier. Thus, for example, the light collected by the top part of the lens would be sent to one detector, that from the middle third to a second detector, and the lower part to a third detector. As the particle passes through the beam, it does not uniformly illuminate the receiving lens, but instead its diffraction pattern passes over the receiving lens from one side to the other. This causes the signal to be received in the photomultipliers at slightly different times, i.e. the separate detector signals are out of phase. The phase differences depend on the scattering pattern of the particle,

which is in turn dependent on its size and refractive index, as discussed in Chapter 6. The full mathematical theory is highly complex (see, for exampl, Bachalo and Houser 1984) but the size of the individual particle can be determined from the relative phases of the detector signals if the refractive index of the particle is known. This allows not only particle size and velocity to be measured, but also the correlation of velocity and size in individual particles can be studied. It is also possible to synchronize the measurement with an external signal, such as the firing of a metered dose inhaler, so that a complete profile of the size, velocity, and structure of an aerosol can be produced, allowing its operation to be studied in considerable detail. Systems are manufactured by Aerometrics (probably the most well-established) and Dantec; the only drawback is the cost, which is at present £60,000 for a basic system, and considerably more for multi-axis systems capable of measuring velocity along 3 axes simultaneously.

MISCELLANEOUS TECHNIQUES

Filtration

Filtration is widely used for the monitoring of the total mass of pollutants, but rarely to obtain particle size information. Filters can efficiently collect particles down to very small diameters (0.01 μm or less) but do not have the accurate cutoff characteristics required for size analysis in an analogy of sieving. Most filters are of the network type, and operate by impaction; even those having carefully controlled pore sizes (e.g. Nucleopore filters) do not display accurate size cutoff which is independent of filter load. It is also rarely practical to perform a size analysis of the particles collected, since they may aggregate on the filter, or be damaged or deflocculated during recovery. One of the few practical methods is the use of cellulose acetate filters, which can be dissolved in acetone and the particles transferred to a slide or electron microscopy grid for microscopic analysis. This method works well for hard particles which can be handled

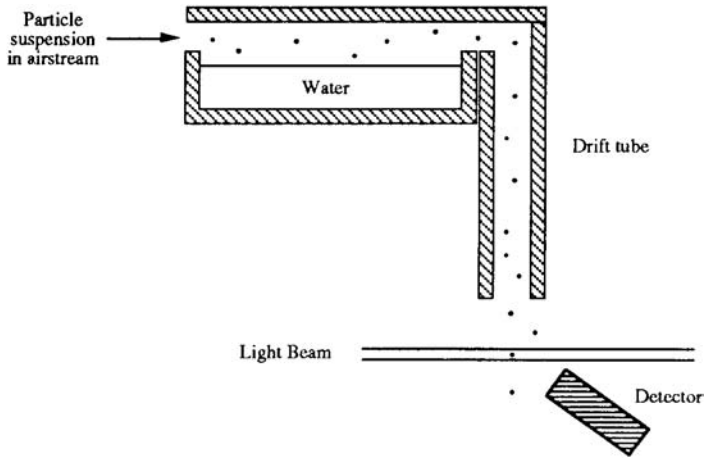


Figure 8.11. Nucleation counter.

without damage, and whose flocculation state is unambiguous.

Nucleation counters

Conventional optical counters can detect particles as small as 100 nm using powerful laser light sources. Below this size the scattering intensity from a single particle is generally too small to be detectable. However, it is possible to count particles as small as only a few nanometres using nucleation techniques. A typical apparatus is shown in Figure 8.11. The stream of air is drawn through a chamber which is saturated with water vapour; the particles act as nuclei for the condensation of water, and rapidly grow into micrometre-sized droplets which can then be counted using a conventional optical zone sensor. The major problem with this technique is that it is difficult to relate the water droplet size to the size of the particle which nucleated it, and so the instrument can only count particles, and not size them. We could note in passing that this nucleation phenomenon is of some importance in the behaviour of inhaled particles, particularly from smoke. Cigarette smoke, for example, generally consists of particles too small to impact in the

lung, and so if it is exhaled immediately, little deposition occurs. However, the majority of addicts allow the smoke to remain in their lungs for a few seconds prior to exhaling it. During this time the smoke particles grow into droplets by nucleation, and these impact and deposit. This, of course, leads to a much higher level of toxicity.

REFERENCES

Bachalo W.D. and Houser M.J. (1984) *Optical Engineering* **23** 583–590.
Bohren C.F. (1987) *Clouds in a glass of beer.* Academic Press, New York.
Gonda I. (1981) *J. Pharm. Pharmacol.* **33** 692–696.
Hallworth G.W. and Andrews U.G. (1976) *J. Pharm. Pharmacol.* **28** 898–907.
Hallworth G.W. and Westmoreland D.G. (1987) *J. Pharm. Pharmacol.* **39** 966–972.
May K.R. (1945) *J. Sci. Instr.* **22** 187–195.
May K.R. (1975) *J. Aerosol Sci.* **6** 413–419
Phillips E.M., Byron P.R., Fults K and Hickey A.J. (1990) *Pharm. Res.* **7** 1228–1233.
Rao A.K. and Whitby K.T. (1978a) *J. Aerosol Sci.* **9** 77–86.
Rao A.K. and Whitby K.T. (1978b) *J. Aerosol Sci.* **9** 87–100.
Vincent J.H. (1989) *Aerosol sampling: science and practice.* Wiley, Chichester.
White P.J.P. (1990) The design of controlled environments. In: Denyer S.P. and Baird R. (eds), *Guide to microbiological control in pharmaceuticals,* Ellis Horwood, Chichester, pp 87–124.

9
Sedimentation

Sedimentation ranks with sieving as one of the oldest methods of particle size measurement and classification, and for many years was an extremely important technique. A wide range of instruments were produced, using principles such as x-ray or optical absorbance to detect the suspended material, and centrifuges to accelerate the process or to extend its range of application. Unfortunately many of these instruments have been displaced in recent years by light scattering methods, which are often more rapid and convenient, and many older sedimentation instruments are no longer manufactured or supported. This is unfortunate because sedimentation methods are capable of providing accurate results with good size resolution. They are particularly valuable for users who perform few analyses and cannot justify the cost of the more expensive light scattering instruments. They are also among the few methods that can be used for hands-on teaching in large classes, since the cost of simple sedimentation columns is relatively low.

Although a very large number of measurement methods have been based on sedimentation, most vary only in the method used to study the gradual separation of the particles, and the fundamental physics of the system—the settling of particles in a gravitational field—is identical to all. We will briefly examine the basis of sedimentation, known as Stokes' law, before considering the application of the most popular experimental methods.

THEORY

The settling of spherical particles in dilute suspension has been fairly well understood for many years, and is described by Stokes law. If a sphere is falling through a liquid, it will be acted on by two forces. These are the accelerating force due to gravity F_g, and the viscous drag from the fluid, F_v. These are given by:

$$F_v = 3\pi\eta dv$$

and

$$F_g = \pi d^3 g(\rho_s - \rho_l)/6$$

where d is the particle diameter, v its velocity, g is the acceleration due to gravity, η is the viscosity of the fluid, ρ_s is the particle density and ρ_1 is the density of the liquid. F_g is fixed and F_v is initially zero, so the particle will accelerate downwards. As it does so, F_v increases with velocity until it is equal to F_g. The net force on the particle is then zero, and so the particle will then fall with a constant velocity (the *terminal velocity*). This velocity is given by:

$$3\pi\eta dv = \pi d^3 g(\rho_s - \rho_l)/6$$

or:

$$v = \frac{d^2 g(\rho_s - \rho_l)}{18\eta}$$

which is the conventional representation of Stokes' law. Sedimentation experiments normally assume that the particles instantaneously reach their terminal velocity. This is not the case in practice, but the time taken for the terminal velocity to be attained (the *acceleration time*) is very small and can be neglected.

Limitations of Stokes' law

Stokes' law applies only to spherical particles, and nonspherical particles are described in terms of an equivalent diameter; this is the Stokes' equivalent diameter, d_{st}. Any particle which sediments at a particular velocity is said to have a Stokes diameter which is the diameter of the sphere with the same density which sediments with the same

velocity. Recall from the discussion of equivalent diameters in Chapter 1 that this will in general be different to other equivalent diameters for nonspherical particles; the comparison of Stokes diameters with equivalent diameters from other sources (e.g. Coulter counter) would provide a shape factor for the particles.

One of the most significant limitations of Stokes' law is that the particle suspension must be dilute. The reason for this is that the derivation of the viscous drag force assumes that the fluid is infinite in extent, or that the particles are isolated. In practice this is not the case, and the particles can interact. As the particle passes through the fluid, it induces flow in it. If another particle is nearby, this local flow causes the two particles to interact, so that they sediment at a different rate than if they were widely separated. Normally the sedimentation rate is lower in the concentrated suspension (this is termed *hindered settling*) but if the forces between the particles are weakly attractive, flocculation may occur and the suspension may settle more rapidly. Consequently Stokes' law is accurate at low particle concentration, but deviations occur as the particle suspension becomes more concentrated. Scarlett and Cowlam (1977) demonstrated that the solid concentration should not exceed 1–2% in most sedimentation experiments, in order to avoid interparticle interactions.

A further source of error arises from temperature variations. Temperature enters Stokes' law through the fluid viscosity, since the viscosities of most fluids fall as the temperature rises. All sedimentation experiments should be performed at a controlled temperature if this error is to be minimized. Temperature can also have an effect if one end of the sedimentation vessel is at a slightly different temperature from the other, when convection can occur. This is only a problem with fluids of low viscosity; for example, we recently had some major problems with convection when studying the settling of micronized drugs in aerosol propellants. In this case the convection was also partly driven from cooling produced by evaporation at the top of the liquid; sealing the cell solved this problem.

One of the most significant limitations of Stokes' law is that the particles are assumed to settle under laminar flow conditions, i.e. that no turbulence is induced in the medium by their movement. This limitation has been very widely investigated, largely on a semiempirical basis. The behaviour of particles moving in viscous media is controlled by two dimensionless parameters. These are the drag coefficient, C_d, and the Reynolds number, Re, which are given by:

$$C_D = \frac{4gd(\rho_s - \rho_l)}{3\rho_s v^2}$$

and

$$Re = \frac{\rho_l v d}{\eta}$$

The drag coefficient describes the ratio of the drag force to the particle's inertia, and the Reynolds number describes the ratio of the velocity of the particle to the viscous force resisting its motion. Substituting these two expressions into Stokes' law gives:

$$C_D = 24/Re$$

which suggests that the drag coefficient should be inversely proportional to the Reynolds number. It is possible to experimentally determine the relationship between these quantities for a sedimenting sphere, and the resultant graph looks similar to that shown in Figure 9.1. At low velocities, the drag coefficient falls inversely as the particle velocity is increased. This is the region of laminar flow. The flow becomes turbulent above a Reynolds number of about 2000 but this is not a sharp transition, and is preceded by an intermediate or transition region at lower Reynolds numbers. The exact limits of this region are ill-defined, but Allen (1990) suggests that sedimentation analyses should not be made above a Reynolds number of 0.2, where a detailed analysis shows that an error of about 5% results from deviations from laminar flow.

It is straightforward to use these equations to find the corresponding particle diameter when the appropriate densities and viscosities are known. For example, Allen

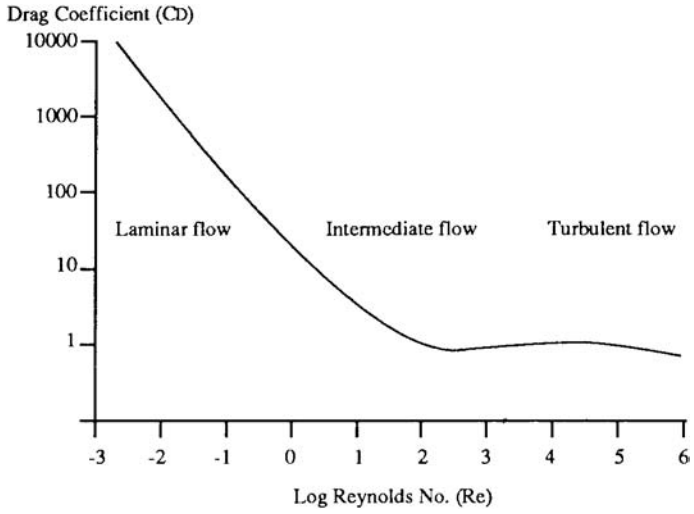


Figure 9.1. Relationship between drag coefficient and Reynolds number for a sphere

(1990) derives an upper limit of 60 μm for quartz particles settling in water. Above this size, increasingly large errors will be introduced owing to transitional flow, although fully turbulent settling will not be observed until the particles are much larger.

A wide range of methods have been devised for sedimentation analysis; these can be classified in a number of ways. They generally fall into one of two types, *incremental* or *cumulative.* Incremental methods find the size distribution by measuring the concentration of particles at a particular depth as a function of time, whereas cumulative methods find the total amount of material which has settled out as a function of time. They can also be classified according to the initial state of the suspended sample as either *homogeneous* or *line start* methods. Homogeneous methods start with the sample uniformly distributed in the sedimentation fluid, whereas line-start techniques prepare the sedimentation column and place a thin layer of particles on the top of the column. Although this simplifies the analysis, since all the particles fall through the same

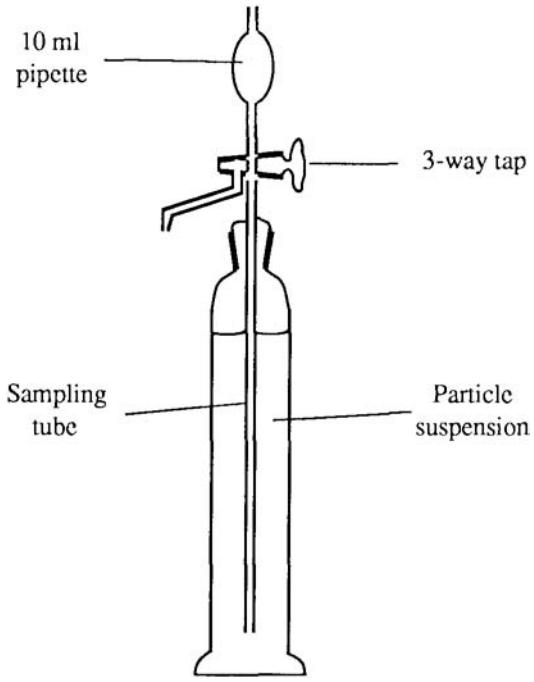


Figure 9.2. The Andreasen pipette.

distance, it can be rather difficult to set up the initial layer of particles in this class of experiments.

THE ANDREASEN PIPETTE

Although many manual apparatus have been described for sampling from sedimenting liquids, the Andreasen pipette is the most widely used. It is an homogeneous incremental method, since it starts with a uniform suspension, and analyses the concentration of material at a fixed depth (near the bottom). The apparatus (Figure 9.2) consists of a glass vessel holding about 600 ml of suspension, into which a 10 ml calibrated pipette dips, the tip being about 4 cm from the bottom of the vessel. The side of the vessel is calibrated in centimetres, with the zero at the same level as the sampling tip of the pipette. A 3-way tap is provided to allow a sample of the liquid to be withdrawn via the pipette at intervals.

The operation of the pipette is straightforward; a sample of the material is dispersed in a suitable liquid (usually water with dispersing agent) to form a suspension of 0.1–1% solid concentration, and the apparatus is filled with this suspension to the upper mark on the main vessel. The container is then agitated by inversion (not by stirring, since this causes persistent vortices) and the timer started when the agitation is complete. Because of errors in the determination of this zero time, it is rarely worth taking samples in the first minute. Samples are taken at intervals; since the Stokes' velocity is proportional to the square of the particle radius, it is sensible to use sampling times in a square of 2 or root 2 progression (2, 2.8, 4, 5.6, 8 minutes, etc.) to achieve a linear sampling of particle radius.

In order to sample, the 3-way tap is turned so that the pipette bulb can be filled from the suspension in the main vessel; a sample is withdrawn and the tap is closed. It is then turned so that the sample can be emptied into a suitable container for later analysis (gravimetric, HPLC, or whatever). Some care is needed to avoid disturbing the sedimented material below the pipette; in particular care should be taken not to overfill the pipette and allow the liquid in the pipette to flow back into the main vessel, since this can stir things up quite badly. After the sample in the pipette has been dispensed into a suitable container, a few millilitres. of clean suspending liquid are drawn into the pipette to rinse it, and this washing added to the main sample. The sample can then be made up to a known volume for analysis, or dried for weighing.

Normally water is chosen as the suspending fluid, and gives satisfactory results in the size range of 1–100 μm for most materials. However, an alternative may be needed if the material is water-soluble. The experiment should last from 30 minutes to 1 hour; faster sedimentation risks becoming turbulent, and may require a different selection of suspending fluid. The experiment can be slowed by increasing the viscosity (mixtures of water and glycerol can be used) or a denser fluid can be employed. Most dense fluids are toxic and expensive (e.g methylene iodide or

saturated solutions of thallium or mercury compounds) and this approach is rarely used.

Data analysis

The analysis method for the Andreasen pipette is similar to that for all incremental homogeneous techniques. The basis of the method is to examine the behaviour of the particles in a thin layer of the fluid at the sampling depth h (i.e. a distance h below the surface). Consider the behaviour of monodisperse particles of diameter d. These will sediment at a velocity v_d given by Stokes' law. Initially the thin layer will have particles of this diameter both above and below it, and since all the particles of this diameter are falling with the same velocity, they will enter and leave the layer from above and below at the same rate, and their concentration in the layer will not change. However, ultimately the particles from the top of the liquid column will reach the thin layer, and after this point there will then be no more particles arriving from the top, and all will have left from the bottom. Consequently the concentration of particles in the layer is constant until all the particles from the top have reached the layer, when it suddenly falls to zero. The time t_d at which this happens is then given by

$$t_d = \frac{h}{v_d}$$

We can find the behaviour of a distribution of particle sizes simply by generalizing this argument. Suppose we have a size distribution F(d), then after a time t_d all particles *larger* than d will have sedimented to a point below the sampling depth, but the concentration of all the smaller particles will be unchanged from their initial concentration. Consequently by measuring the amount of material in the sample at time t, we will find the fraction smaller than the size with Stokes' velocity v=h/t. If the initial concentration of particles is C_0 and the concentration at time t is C_t, then 100 C_t/C_o is the percentage below this diameter. If we plot this percentage against the particle diameter having a Stokes' velocity h/t,

we obtain a cumulative undersize distribution for the material.

We should note that this analysis assumes that the depth h is constant. However, we have to withdraw samples for analysis, so h decreases during the experiment. As samples are removed, the distance of the pipette tip from the top of the liquid column falls, so the smaller diameter particles in later fractions do not have to fall so far as the early samples before they reach the pipette tip. The pipette depth h should be recorded prior to the removal of each sample, so that the Stokes velocity for each sample can be calculated, and the corresponding diameter obtained. In practice this decrease of depth is useful because it speeds the analysis; the smaller particles do not have to sediment as far before sampling as the larger particles.

There are a number of errors inherent in this process, most of which are usually small. Allen (1990) discusses a number of errors, which include (a) that the sample removed is not a thin layer but is a roughly spherical mass, and (b) that a small amount of material remains in the sampling tube after each sample, leading to a systematic oversize analysis. The speed of sample removal can also affect the results, since if it is withdrawn too quickly, it will preferentially sample the smaller particles which have low inertia. BS 3406 recommends a filling time of 20 s, which should be adhered to for the whole experiment.

SEDIMENTATION BALANCES

The principle of the sedimentation balance is shown in Figure 9.3. A balance pan is immersed in a column of suspension so that the suspended particles can settle on to the pan to be weighed. The method is a cumulative homogeneous one, since it normally starts with a uniform suspension and weighs the particles as they accumulate on the balance pan. A number of instruments are available operating on this principle, from manufacturers such as Sartorius and Shimadzu. The major problem in the development of sedimentation balances is that the balance

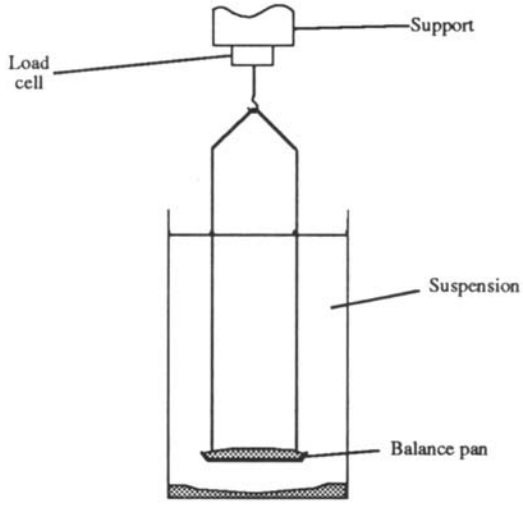


Figure 9.3. Sedimentation balance.

pan must be kept in a fixed position to avoid stirring the suspension and changing the sedimentation height. Obviously a swinging beam balance cannot be used! Early balances used electromagnetic feedback devices to change the beam restoring force as the pan moved, but modern electronic balances use low compliance load cells so that the difficulty is eliminated.

Data analysis

The method of analysing cumulative sedimentation data was originally described by Oden (1926) and an improved method was described by Bostock (1952). The material which has deposited on the balance pan by time t consists of two parts. The first is that material which has a Stokes' velocity v greater than h/t, where h is the liquid column height; we will suppose that this velocity corresponds, via Stokes' law, to a diameter d_t. All the particles with velocities greater than d_t will have reached the balance pan by this time, no matter what their starting position in the column of liquid. There will also be a second component to the sedimented material, which consists of particles smaller

than d_t, which have reached the pan because they started from intermediate positions in the liquid column.

The total amount of material in the first category, i.e. that larger than d_t, is simply the cumulative percentage oversize d_t, i.e.

$$\int_{d_t}^{d_{Max}} f(D)\, dD$$

where f(D) is the frequency size distribution of the particles. The amount of the second fraction is slightly more complex. If we consider a suspension of particles of diameter d, where only a fraction have sedimented since their Stokes' velocity v_d is smaller than h/t, then the fraction which have sedimented will be given by $v_d t/h$. Consequently the total amount of the second component will be given by:

$$\int_{d_{Min}}^{d_t} \frac{v_d t}{h} f(D)\, dD$$

So the total amount M sedimented will be given by:

$$M = \int_{d_t}^{d_{Max}} f(D)\, dD + \int_{d_{Min}}^{d_t} \frac{v_d t}{h} f(D)\, dD$$

Differentiating with respect to time:

$$\frac{dM}{dt} = \int_{d_{Min}}^{d_t} \frac{v_d}{h} f(D)\, dD$$

If we now write M_t as the mass of particles larger than d_t (i.e component 1 on the balance pan) we obtain:

$$M = M_t + t\frac{dM}{dt}$$

or

$$M_t = M - t\frac{dM}{dt}$$

M_t is simply the cumulative weight in the particle size distribution up to size d_t, and so can be calculated if M is measured at a range of times. The major error inherent in this process is the determination of the slope dM/dt. This

can be highly dependent on the accuracy and noise in the data. If data points are taken at very close time intervals, very little material will have sedimented in this time range, and so the precision of measurement will be poor. Alternatively, if the data points are taken at wider intervals, errors will result since the true slope will be approximated by a finite difference. In practice modern sedimentation balances obtain a large amount of data and employ computer analysis using curve fitting or smoothing to minimize the errors.

Sedimentation need not occur in liquid columns; air sedimentation is occasionally used in the study of aerosols, although few commercial instruments are available (e.g. the ASM 1000 from Palas). Hirst and Kaye (1971) described an air photosedimentometer for the study of dry powder sprays. It was 27 feet high and over 2 feet in diameter, constructed from PVC sheet, and was attached to the outside of the building. Not surprisingly, problems with air stability were encountered.

PHOTOSEDIMENTATION

The major problem with pipette methods of sedimentation analysis is that the withdrawal of the sample can disturb the sedimentation column. A natural solution to this problem is not to remove samples, but instead to use a remote method of estimating the amount of material in suspension at a particular point. The most obvious technique is to use a light beam passing through the suspension, and estimate the particle concentration from the attenuation of the light beam. Note that we do not call this the *absorbance* of the sample, since it includes both absorbed and scattered light contributions, and it is more correctly referred to as the *extinction* of the sample. Instruments which operate on this principle are called photosedimentometers.

The particle concentration is obtained from the extinction by assuming that the particles cause a loss of light from the beam which is proportional to their cross-sectional area.

Particles of diameter d will obscure an area which is proportional to the square of their diameter; the obscured area a_d is written as:

$$a_d = k_d d^2$$

where k_d is a constant dependent on the shape of the particles (e.g. it is $\pi/4$ for spherical particles). In a similar manner to the Beer-Lambert law for absorbing materials, the optical density D of a suspension of the particles of concentration c_d g 1^{-1} is given by:

$$D = 2.303\ c_d\ L\ n_d\ k_d\ d^2$$

where n_d is the number of the particles in 1 g of the material, and L is the path length. If a range of sizes from d_{max} to d_{min} are present, the total absorbance is given by:

$$D = 2.303\ c\ L \int_{d_{min}}^{d_{max}} n_d\ k_d\ d^2 dd$$

where c is now the total concentration of the suspended material.

In practice this approach has a severe problem. Large particles, for example, 120 μm and above, may obscure light in this geometric manner, but smaller particles will cause the diffraction phenomena that we examined in Chapter 6, which will lead to increasing errors as the size of the particles approaches that of the wavelength of the light being used. This is taken into account by several manufacturers by adding a constant of proportionality, the *extinction coefficient* K_d (a badly chosen name, since it has an alternative use), to the expression for the light obscured. The extinction coefficient is the ratio between the light that is obscured by the particle and that predicted on the basis of its cross-sectional area. Thus:

$$D = 2.303\ c\ L \int_{d_{min}}^{d_{max}} K_d n_d\ k_d\ d^2 dd$$

In principle the extinction coefficients can be calculated from diffraction theory, but in practice this is rarely performed,

since the amount of light scattered out of the beam depends on the detector geometry.

Consequently tables of K_d provided with instruments are often obtained experimentally. Since the diffraction phenomena depend on the material being studied, via the refractive index, we should not expect these correction factors to be the same for all materials.

A possible solution to this problem is to use a wide-angle detector. This collects the diffracted light scattered from the suspension over a large solid angle, and so the only light that does not reach the detector is the component which has been directly blocked by the particles. Although this is not a theoretically rigorous approach, it does provide some improvement in accuracy, and most modern instruments use this technique. Despite this, the approach is still unsatisfactory for small particles, and its uses are mainly for comparative studies rather than for the accurate determination of particle size.

Several instruments have been manufactured to perform sedimentation analysis using this method. One of the first was produced by EEL, and was a simple manual non-recording instrument. Later instruments by Palas, Ladal, Micromeritics (Sedigraph 5500L) and Fritsch (Analysette 20) provide direct recording of transmitted intensity. The first instruments contained a fixed photodetector, which measured the sedimentation at a fixed level in the cell. This caused measurement times to be unacceptably long for particles of small diameter, and so later instruments gradually moved the light and detector system towards the top of the sedimentation cell, so that the smaller particles were measured after having sedimented through a shorter distance. These instruments are termed *scanning photosedimentometers.* The same approach is used in the Sedigraph with x-ray detection (see below).

Photosedimentation is widely used for studies of micronized pharmaceuticals. Levin *et al.* (1985) studied oxytetracycline powder and found good agreement with the results of sieving or microscopy. Dissolution of the antibiotic

$$A = \mu cL$$

was prevented by using decane or isooctane as the sedimentation fluid.

X-RAY SEDIMENTATION

As we have discussed, the major problem encountered when using light transmittance to measure concentrations is that the relationship between particle concentration, size and absorbance may be complicated by diffraction phenomena. This is particularly the case if the particles have dimensions similar to the wavelength of light. This problem can be circumvented by using radiation whose wavelength is too short for appreciable diffraction to be caused by the particles. If a beam of x-rays is passed through the sample, their extremely short wavelength prevents their being diffracted by the particles, and the absorbance is proportional to the amount of material in the beam:

where μ is the x-ray absorption coefficient of the material (assuming a nonabsorbing suspending fluid), L is the path length, and c is the particle mass concentration, in direct analogy to the Beer-Lambert law.

The best-known instrument operating on this principle is Micromeritics' Sedigraph, originally described by Hendrix and Orr (1970). The principle of the instrument is shown in Figure 9.4. The dispersed sample is contained in a cell through which a beam of x-rays is passed. The beam is collimated to a cross-section of 1 cm horizontally and 0.005 cm vertically. As the sample sediments, the density of material in the beam falls, and the x-ray transmittance increases. Initially the x-ray beam is passed through the cell near the base, but as the analysis progresses the cell is moved downward through the beam to reduce the sedimentation depth as the experiment proceeds. This means that the smaller particles have to sediment a shorter distance, which reduces the duration of the experiment. The movement rate can be preset to correspond to a desired size range.

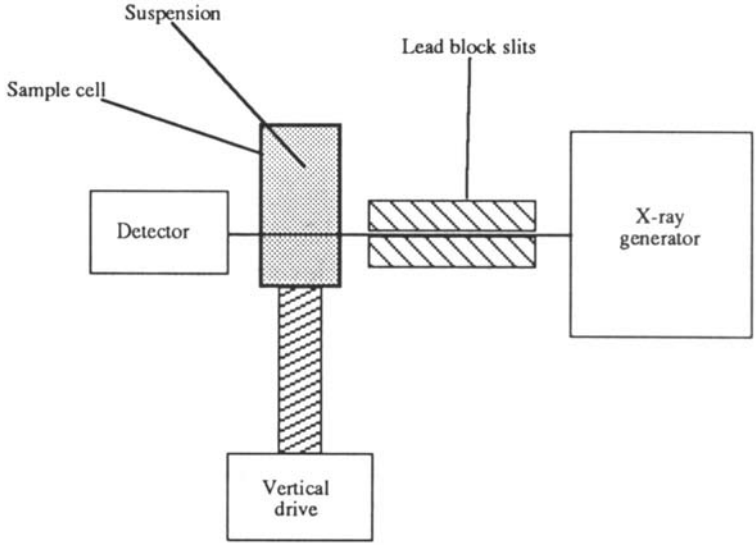


Figure 9.4. Principle of X-ray sedimentation.

Operation of the instrument is straightforward; firstly the sampling tubes are immersed in a beaker of clean suspending liquid, and the pump is operated to circulate the liquid through the cell. The recorder can then be set to zero. The sample is prepared in suspension in a suitable liquid (organic solvents are acceptable) and the sampling tubes immersed in the sample. Suspension is ensured by a magnetic stirrer. The sample suspension can now be pumped through the cell and the x-ray intensity checked; ideally the sample should have attenuated it by between 40% and 80% of its original intensity. The pump is then switched off, which isolates the suspension in the cell, and the analysis begins. The total analysis time depends on the size range present and the cell movement rate selected. A particular advantage of the instrument is that, since the zero absorbance level has been accurately set, the instrument can report the weight below the measuring range at the end of the experiment, even though this material has not sedimented, since the residual absorption can be measured.

The absorption coefficient of x-rays increases with the atomic numbers of the atoms of which the particles are made. Light elements, such as carbon and oxygen, only weakly absorb x-rays, while heavier elements absorb much more strongly. This makes x-ray sedimentation rather insensitive to organic materials such as the drugs, polymers, and many excipients used in pharmaceutical formulations. Consequently, while x-ray sedimentation is a highly useful technique, it has not been extensively applied in the pharmaceutical industry, but has found much wider use in geology, mining and metallurgy. A further problem may be encountered if the sample contains several populations of particles made of different materials. If these absorb x-rays to different extents, the instrument will weight each population in the total distribution according to its x-ray absorption coefficient. There will thus appear to be a greater proportion of dense materials and a lesser amount of the lighter materials.

CENTRIFUGAL SEDIMENTATION

The lower size limit of conventional sedimentation experiments lies in the region 1–5 μm, depending on the density of the material under study. This limit is not set by the practical experimental duration, but is reached when the sedimentation velocity becomes comparable to the random diffusion velocity of the particles. If the particles have a significant random motion, this will be added to the Stokes velocity, so that the sedimentation velocity will now be the vector sum of the Stokes velocity and the diffusion velocity. Normally the Stokes velocity is predominant, but if the particles are small, the diffusional velocity distribution becomes significant, so that the particles sediment with a distribution of velocities (Figure 9.5). It is possible to study smaller particles or less dense materials by accelerating their sedimentation in a gravitational field using centrifugal methods; under these conditions, g, the acceleration due to gravity, in Stokes' equation is replaced by $\omega^2 r$, the centripetal acceleration of the particle. This causes the

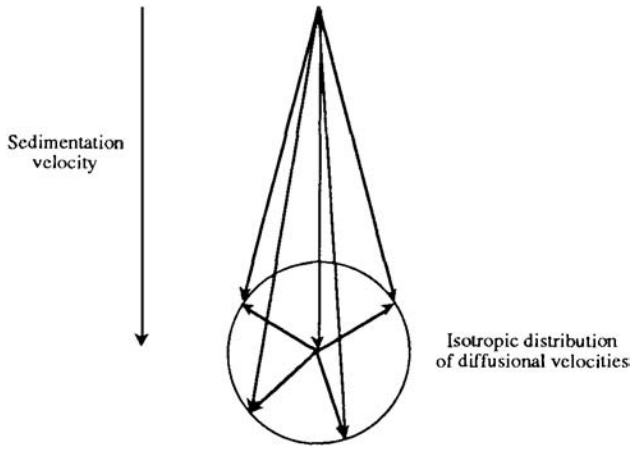


Figure 9.5. Vector addition of sedimentation and diffusional velocities.

sedimentation velocity to be higher without affecting the diffusion rate, so that the conventional sedimentation effect again predominates. Increasing the gravitational field, for example from g to $\omega^2 r$=500g, will increase the sedimentation velocity by a factor of 500. Since the velocity is proportional to the square of the diameter, this decreases the lower diameter limit of the sedimentation technique by the square root of 500, i.e. a factor of 22. This allows the method to be used for the study of submicrometre systems; for example, Groves and Yalabik (1974) detected 18 nm liposomes of excess emulsifier in a parenteral fat emulsion using a centrifugal method.

Centrifugal sedimentation instruments suffer from a complication which is not found with conventional gravitational sedimentation methods. The sedimenting particles do not all move along the axis of the cuvette or sample vessels in parallel paths, but instead move radially along paths which diverge as the particles' distance from the centre of rotation increases. This causes the particles to be diluted as they move outwards to the edge of the sedimentation cell. This effect is called radial dilution; it is most serious if the particles start close to the centre of rotation, and less serious if the distance travelled is a small

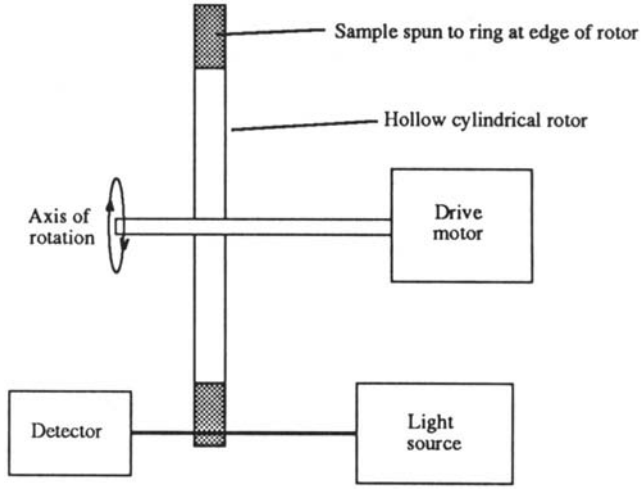


Figure 9.6. Disc centrifuge.

fraction of the radius of the cell. Most modern computerized instruments contain routines to correct for this effect.

A number of centrifugal sedimentation instruments have been designed on this basis. A typical example is the Joyce-Loebl disc centrifuge (Figure 9.6). This consists of a thin cylindrical glass rotor mounted with its axis of revolution nearly horizontal. A variable speed motor drives the rotor at speeds up to several thousand rpm. A photodetector measures the intensity of a light beam passing through the rotor near the rim; it is at this point that the sedimentation of the particles is measured in a similar manner to a conventional photosedimentation experiment. The method suffers from the same problems as conventional photosedimentation in that the transmitted light intensity is dependent on the diffraction of light by the particles. However, for very small particles below the wavelength of light, the diffraction is nearly isotropic and the extinction is proportional to the square of the particle mass, which simplifies the analysis to a certain extent.

In use the hollow rotor is partially filled with a known volume of the sedimentation fluid. This volume determines the sedimentation depth used in the experiment. The rotor is

then accelerated up to the required speed, when the sedimentation fluid spins out to form a ring at the edge of the rotor. The sample (0.5–1 ml) is injected at the centre of the rotor, and streams along the face until it reaches the inner surface of the ring of sedimentation fluid, where it is uniformly spread over the liquid surface. The particles then sediment outwards toward the rotor rim. Since the particles all start from the same position, i.e. at the inside ("top") of the fluid, this is a line-start method. The ring of fluid is about 1–2 cm deep, and the radius of the rotor is 15–20 cm, so the radial dilution effect in the instrument is small.

A significant difficulty arises with this method if the particles are suspended in a fluid of different viscosity or density to that used as the sedimentation fluid. In this case, when the sample reaches the inner fluid surface after injection, hydrodynamic instability can prevent the particles from being uniformly distributed at the interface. Instead, 'fingering' or 'streaming' occurs as masses of the sample are rapidly pulled through the sedimentation layer. The apparatus is provided with stroboscopic illumination so that the operator can easily check for the occurrence of this phenomenon. To avoid the problem, a technique called 'buffered line start' has been introduced. After the rotor has spun up to the required speed, a small sample (about 1 ml) of the sample dispersing fluid is injected onto the rotor. Immediately, the rotor is accelerated or decelerated by about 30% of its velocity, then back to the set velocity. This causes the injected dispersion fluid to mix uniformly into the first 1–2 mm of the sedimentation layer, forming a smooth gradient of density and viscosity, the inner edge of which is similar to the sample density and viscosity. The sample is then injected onto the top of this layer. This technique is claimed to largely eliminate fingering problems.

An alternative centrifugal technique is employed in the Ladal Pipette centrifuge. This is a homogeneous start technique in which samples are removed by a pipette in a similar manner to that employed in an Andreasen pipette. This method has not been widely used, and is discussed by Svarovsky and Allen (1977). Other manufacturers of

centrifugal sedimentation instruments include Brookhaven and Horiba.

To avoid the problems of measuring particle number by light absorbance discussed earlier, an x-ray version of the disc centrifuge has been developed (Hornby and Tunstall 1977). This operates in a similar manner to the Joyce-Loebl centrifuge, but the optical sensor is replaced by a narrow beam of x-rays. The instrument is used in a homogenous, rather than a line-start mode; however, there appears to be no reason why either instrument could not be operated in either mode. A commercial x-ray centrifuge is produced by Ladal and has been described by Allen (1990).

Determination of density

Calculation of particle size from sedimentation data requires an accurate value for the density of the material. This may be known from existing data on the pure material, but more commonly needs to be determined from a sample which is only available as a powder (often the research sample itself).

The most common technique for this purpose is liquid pycnometry. For this purpose it is essential to have available a liquid which wets the powder under study, but in which the powder is not soluble. The particles must not swell in the liquid, a common property of polymer or protein based microparticles in aqueous environments. Since accurate weighing is required, the liquid must also have a low volatility. Water and many higher-boiling organic solvents are suitable. It is helpful if the density of the liquid is accurately known, otherwise it will have to be determined beforehand.

Pycnometers are simply glass vessels designed to hold a precise volume. They come in a number of forms, usually u-shaped; for powder density measurements it is essential to have one with a wide entry tube, rather than the narrow entry tubes used in instruments for the measurement of pure liquid density. Specific gravity bottles can also be used, as can (at a pinch) volumetric flasks with well-fitting glass stoppers.

To measure the density, the weighed vessel (mass M_1) is about half filled with the powder and re-weighed (mass M_2). The powder weight is then (M_2-M_1). The vessel is then partly filled with the chosen fluid, and placed in a vacuum desiccator to remove air bubbles. The vacuum should be applied cautiously since no powder must be lost. When all the bubbles have been removed, the vessel can be filled to the mark with pure liquid. It is then re-weighed (mass M_3) and the weight of liquid added is then given by (M_3−M_2). This can be divided by the density of the liquid to find the volume of liquid added (V_1). The pycnometer is then cleaned, refilled with the pure liquid, and weighed (mass M_4). The mass of liquid in the pycnometer is given by (M_4–M_1), and dividing by the liquid density gives the pycnometer volume V_2. The volume of the powder is thus (V_2-V_1), and so the density is (M_2-M_1)/(V_2-V_1). It is better to determine the pycnometer volume than rely on any previous calibration, since this may have changed (e.g. due to thermal expansion of the glass). For the most accurate work, the pycnometer should be filled while its temperature is controlled in a water bath.

REFERENCES

Allen T. (1990) *Particle size measurement.* 4th ed., Chapman and Hall, London.

Bostock W. (1952) *J. Sci. Instr.* **29** 209–211.

Groves M.J. and Yalabik H.S. (1974) *J. Pharm. Pharmacol.* **26** 77P, 78P.

Hendrix W.P. and Orr C (1970) *Automatic sedimentation size analysis instrument.* In: Groves M.J. and Wyatt-Sargent J.L. (eds), Proc. conf. particle size analysis, Bradford; Society for Analytical Chemistry, London, 133–146.

Hirst C.A. and Kaye R.C. (1971) *J. Pharm. Pharmacol.* **23** 116–124.

Hornby M.R. and Tunstall D.F. (1977) *A new X-ray disc centrifuge.* In: Groves M.J. (ed.), Proc. conf. particle size analysis, Salford, Heyden, London 451–457.

Levin N.P., Khanukaev Y.A. and Mon'kin V.A. (1985) *Khim. Farm. Zh.* **19** 499–502.

Oden S. (1926) *Sedimentation analysis and its application to the physical chemistry of clay and precipitates.* In Alexander J. (ed), Colloid Chemistry, Chemical Catalog Co. N.Y. 861–909.

Scarlett B. and Cowlam S.K. (1977) *Concentration effects in particle size analysis by sedimentation.* In: Groves M.J. (ed.), Proc. conf. particle size analysis, Salford, Heyden, London 430–440.

Svarovsky L. and Allen C.J. (1977) *Errors in particle size measurement by sedimentation.* In: Groves M J. (ed.), Proc. conf. particle size analysis, Salford, Heyden, London 441–450.

10
Microscopy and image analysis

The major advantage that microscopic techniques possess over most other methods of size analysis is that the particle profile itself is measured, rather than some property which is dependent on particle size. Since the particles are visible, the operator can assess easily whether or not they are properly dispersed, their shape is visible, and an intuitive feeling for the nature of the material is obtained. Most experienced particle size analysts would examine an unfamiliar sample by microscopy in order to assess its general nature, shape, and cohesiveness, even without making any accurate measurements, prior to using one of the methods described in earlier chapters.

Microscopy has been widely used for size analysis, initially with simple eyepiece graticules, but the slowness of measurement led to the introduction of devices such as projection systems and comparators to speed the process. With all of these methods each particle must be examined on an individual basis by the operator, which is very time-consuming and subject to systematic error. The process has been revolutionized by the introduction of computerized methods of image analysis, which have become very popular, although the apparatus is expensive.

RESOLUTION

The size of particles which can be imaged by microscopy is limited by the diffraction of the light used to form the image. The resolution of the microscope is given approximately as the wavelength of the light divided by the numerical

aperture of the microscope objective; the numerical aperture depends on the lens geometry but is generally in the range 0.5–1.5. Consequently a light microscope has a resolution limit of about half a micrometre, but it is fairly obvious that as this limit is approached, the potential for error increases rapidly. In practice particles are normally increasingly oversized as their size approaches the resolution limit, because the diffraction disc image they produce is larger than the particle. Under appropriate (dark-field) illumination, particles much smaller than the wavelength of light can be detected since they scatter light; however the image produced by such particles will simply be a disc at the resolution limit of the microscope, so all such particles will seem to be about half a micrometre in diameter. This technique, which is used for counting very small particles down to a few tens of nanometres, is called *ultramicroscopy.*

STATISTICAL CONSIDERATIONS

A significant problem for microscopic analyses is that they may only examine a rather small number of particles. This is particularly so for direct manual counting, which led to the introduction of semiautomatic counting machines. Even using these, only a few thousand particles per hour can be counted. The situation is not so severe for image analysis systems, where entire frames can be rapidly scanned and analysed without significant operator intervention, particularly if automatic slide-scanning systems are used.

The statistical error in counting N particles into a channel is simply the square root of N. Consequently we have to count as many particles as possible to get good precision; unfortunately the precision only increases as the square root of N, so if for example we needed a count that was a factor of ten more accurate, we would have to increase the number of particles (and correspondingly the counting time) by a factor of a hundred.

This also means that if we are trying to derive a size distribution rather than simply a count, it pays to use only the minimum number of channels that is essential for the

task, so that the count in each channel is maximized. Consequently schemes for manual sizing operate by classifying the particles into a small number of rather coarse classes.

SAMPLE SELECTION AND PREPARATION

Selection of samples for microscopy can easily lead to systematic errors owing to segregation and selection effects. For example, transfer of a small amount of free-flowing material by spatula will select the fines as the large particles on the surface roll off the heap on the spatula. If the material is in suspension, the larger particles may settle out before a drop can be removed to be placed on the slide. As the drop is spread over the slide, smaller or larger particles may be swept non-uniformly over the surface, so that any particular spot is not representative of the whole.

It is fairly obvious from the above discussion that considerably more thought and care must be used than simply putting a bit of material onto the slide. A range of techniques have been devised and validated to varying degrees. Permanent slides can be prepared by mixing a sample of the powder with a viscous mounting medium and removing a small spot of this mixture before sedimentation can take place. The spot is transferred to a microscope slide and a cover slip slowly pressed into place. A more elegant technique is to spread a single drop of the particle suspension in medium on a water surface and allow the mounting solvent to evaporate, after which the remaining film of medium and particles can be picked up on a slide. Suitable media include collodion in amyl acetate and Canada balsam in xylene at 1%, although these may need to be diluted with solvent slightly if the water-film method is to be used.

Nonpermanent slides can be made by carefully mixing the powder into a thin paste with a suitable nonvolatile solvent in a watch glass or petri dish using a soft object such as a fine paintbrush (a spatula or glass rod may fracture the particles). A small amount of this paste can then be

transferred to a slide using the brush, and a cover slip gently applied. The solvent may need to be selected so that its refractive index is different to that of the particles, or they will be difficult to view; this may be a problem for organic drug materials which have rather low refractive indices. Useful dispersing agents are decane (many drugs have a low solubility in alkanes); monobromonapthalene is often used since it has a high refractive index, but this may make it difficult to distinguish the particles from air bubbles, since both will have lower refractive indices than the medium. Water has a very low refractive index but many materials are soluble in it; we have found that perfluorocarbons such as perfluorodecalin are very useful since they also have low refractive indices and nearly everything is insoluble in them. Surprisingly, they wet most materials adequately.

Air bubbles can be a particular problem when studying emulsions, since both the bubbles and the oil droplets are the same shape. For example, microscopy is often used to examine parenteral fat emulsions for large oil drops which are clinically undesirable. A trained operator can usually distinguish oil droplets from bubbles by slightly adjusting the focus of the microscope. As the focus is changed, air bubbles show a pronounced black ring which makes them look very 3-dimensional, while oil drops appear much flatter. This effect is due to the lensing of the droplets or bubbles; the air bubbles have a refractive index which is much lower than the surroundings, so act as strong negative lenses, while the oil drops have a slightly higher refractive index, and so act as weak positive lenses. If the droplets are still difficult to see, specialized illumination methods such as phase-contrast illumination may help.

Biologists have for many years used staining methods to reveal small structures and the same techniques may be used to render particles more easily visible. We routinely study fat emulsions by staining with 0.1% Nile red and using epifluorescence illumination, which considerably eases the problems of visualization of small oil droplets.

Particles on a slide usually fall with a preferred orientation, i.e. with a large flat face downwards. For this reason the appropriate equivalent diameters are specified for particles 'in their most stable orientation'. This may cause problems if the results are to be compared with other methods which show no orientational dependence, and so it is often necessary to prepare slides in which the particles are in random orientations. Probably the easiest method of doing this is to prepare a slide with a thin layer of very sticky mounting agent such as Canada balsam, and sprinkle the powder on to this. A little experience is necessary since the consistency must be such that the particles adhere well, but do not 'drift' after sticking. Methods have been described involving embedding the particles in resin, perspex, or epoxy, and sectioning, but these are not recommended since the sectioning will section some of the particles and rather complex stereological corrections will have to be applied; see, for example, Dullien *et al.* (1969) for details.

An elegant method of measuring particles in ointments was described by List and Groenig (1976). The ointment is placed on a filter membrane and its temperature slowly raised until the base becomes fluid (called the 'dropping point'). The liquid base is then absorbed by the filter, and the particles remain behind for counting.

INSTRUMENTATION

Although a microscope is basically a tube with a lens at each end, it is amazing how much money the manufacturers can persuade you to part with for a new instrument supposedly packed with novel features. Microscopes for particle size analysis have fairly basic requirements; they should be solidly constructed (as are most), with high quality optics. The illumination intensity should be uniform across the field, particularly if a video camera is to be attached. It is often worthwhile attaching a video camera and a high quality monitor even if computerized image analysis is not going to be used, since this will avoid eye

strain. Microscopes are usually available with dual imaging ports so that a camera can be used simultaneously to normal viewing; this does of course decrease the image brightness, which can be a problem if high magnifications are in use, and so it should be possible to switch out the camera port by moving a prism.

A wide range of illumination options are available on modern microscopes. Conventional transmission illumination is adequate for most purposes, but phase-contrast can be useful for studying colourless materials suspended in other colourless materials (e.g. oil in water emulsions). A number of more specialized schemes have evolved recently, mainly to aid biologists studying barely visible objects in cells. These include Nomarski interference microscopy and Rheinberg illumination; they are little used in size analysis but can be valuable when contrast is poor. It should be noted that it is unwise to use these methods when sizing particles close to the resolution limit, because their ability to oversize has not been thoroughly investigated. It is well—known that small particles are oversized in conventional illumination, and phase-contrast causes a bright ring to appear in some specimens, which can lead to severe overestimation. Most of these techniques are described in detail by Lacey (1989)

ELECTRON MICROSCOPY

Electron microscopy is useful since it allows particles much smaller than a micrometre to be measured. Modern scanning or transmission microscopes will resolve particles as small as a nanometre across. Preparation of samples for electron microscopy is complicated since many interesting materials (e.g. polymer or protein microspheres) do not absorb electrons, and specialized methods such as replica casting or shadowing have to be employed. We will not describe the technique in detail here, since methods can be found in standard electron microscopy texts (e.g. Agar *et al.* 1974), and since the operator of the instrument will usually be a specifically trained technician rather than the particle

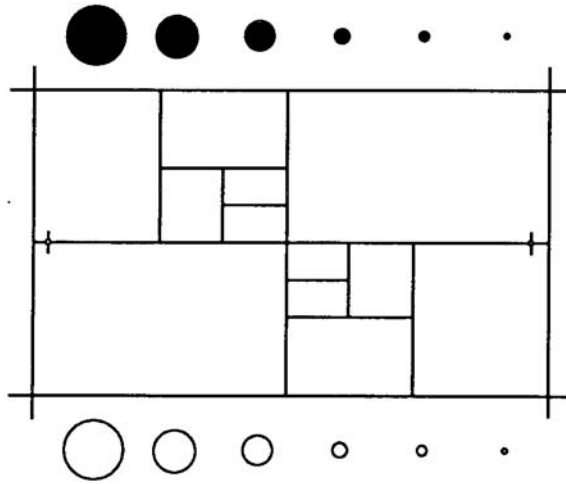


Figure 10.1. Measurement graticule as specified in BS 3406.

size analyst. In order to avoid occupying the instrument for long periods of time, measurements are usually made from photographs rather than directly from the instrument, but this problem is largely offset in modern instruments by image analysis computers which are built into the systems.

MANUAL METHODS

Direct measurement of particles manually has been almost completely superseded by more rapid techniques, although it may be used occasionally for rough measurements or studies of particle shape. Most manual methods operate by comparing the particles with a graticule, of which a number of forms have been described. Graticules with directly calibrated linear scales are not often used and usually lead to systematic overestimation of particle size; most graticules are of the 'comparison circle' type, where a range of circles of appropriate diameters is provided for sizing by comparison; thus the method classifies the particles on the basis of their projected area equivalent diameters. An example, the graticule specified in BS 3406, is shown in Figure 10.1. This type of measurement works well if the particles are roughly spherical, but it can be difficult to

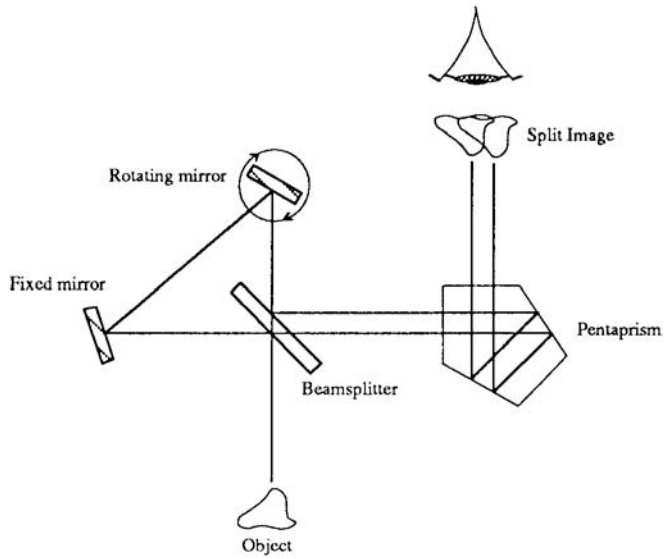


Figure 10.2. Watson image-shearing eyepiece.

estimate the equivalent circle for elongated or irregular particles. All of the graticule comparison techniques require careful operator training and experience to avoid gross systematic errors.

An alternative method of estimating the size of the particle is by an image-shearing eyepiece, which causes two slightly separated views of the field to be seen in the eyepiece. A typical example is the Watson eyepiece, shown in Figure 10.2. The image is split into two by the beamsplitter, and adjustment of the mirrors using a calibrated drum allows the two images to be separated by a variable amount. In use, one would not normally adjust the eyepiece to measure each particle. A better way of operating is to gradually increase the image separation through a range of settings and at each setting count the number of particles for which the images are completely separated.

The only significant advantage of manual measurement is that the operator can select the individual particles to be studied. This would normally lead to undesirable bias, but in heterogenous samples it may be useful to examine a

visually distinct subset of the material. This could correspond to a particular component or crystal type, or simply allow the rejection of aggregates from the analysis. An example would be the study of microspheres prepared from proteins such as albumin; these can be rather hydrophobic and difficult to disperse, and may contain a fraction of amorphous material which has not been properly dispersed. Microscopy allows the sizes of the individual microspheres to be measured even if they are still partly aggregated, as long as the individual microspheres are visible. Such methods are useful for the study of complex samples such as filter materials (Lin'kova *et al.* 1976) and porous substances such as membranes (Saltzman *et al.* 1987) where the operator can choose to measure specific parameters such as fibre lengths or void areas.

SEMIAUTOMATIC METHODS

A considerable number of devices have been described to assist the measurement and totalization of directly viewed particles. Most of these are mechanical recording devices which predate the use of computers and allow the operator to set a control of some sort to a position which describes the particle, after which the control position is recorded by pressing a switch. Most of these devices are now outdated and unsupported by manufacturers, but a few are still in use. A typical device which enjoyed some popularity is the Zeiss-Endter size analyser (Fig. 3), which uses a variable iris diaphragm to project a spot of light on to a photograph of the particle. The diaphragm can be opened or closed to adjust the light spot to the same diameter as the particle, and the iris control is linked to a set of contacts which encode its position. When the light spot has been adjusted to the particle size, a foot switch is depressed and this causes the corresponding channel counter to be incremented. The Chatfield analyser is similar but uses a 35 mm projector to project the image on to the bottom of a horizontal screen set in a table. A variable diameter spot of light is superimposed onto the image by projection from

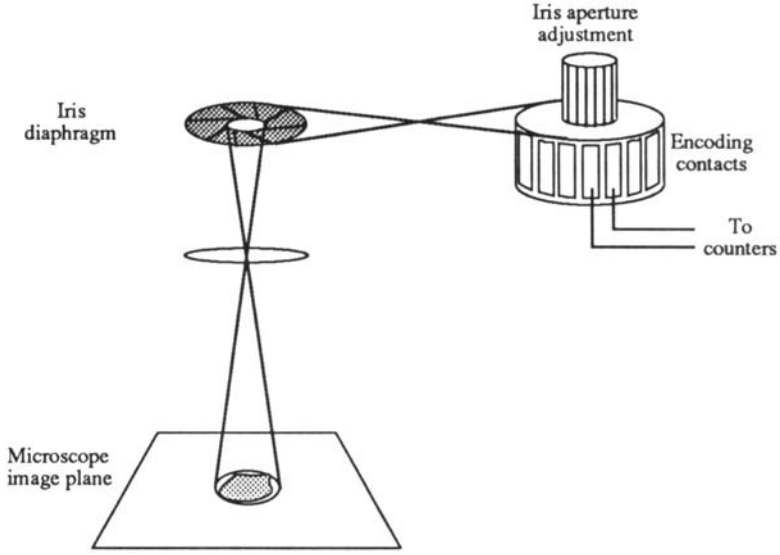


Figure 10.3. Zeiss-Endter size analyser.

above, and a similar arrangement to the Zeiss counter is used to transfer the iris diameter to a set of counters.

A semiautomatic system was constructed by Sims and Withington (1983) by using a projection microscope to project the image onto a digitizing tablet of a microcomputer for measurement. This repre sents an intermediate stage in the development of fully automatic systems and allowed the measurement of particle parameters such as length, breadth, and tracing of outlines.

AUTOMATIC IMAGE ANALYSIS

Computer analysis of microscope images has become popular since it enables the user to rapidly handle the very large amount of information in a typical picture. Particle sizing and characterization are only a small subset of the applications currently being found for computers in image processing, which include object recognition and artificial vision; the field is rapidly expanding, and probably the best current text in this area is that by Russ (1990). Most of our requirements are much more straightforward and can be

accomplished more simply than some of these very complex tasks.

Image analysis involves several separate stages. Firstly the image is acquired by hardware, and converted into electrical form which is stored in the computer memory, or on a mass storage system such as a hard disc. The image is then enhanced to emphasize and define the required details (in our case the perimeters of the particles), and finally a measurement operation is performed to extract the required numbers. These may be processed in various ways to extract further information such as size distributions or particle shapes. Normally these last three stages are all performed by software, which is convenient since it allows the operation of the system to be upgraded and tailored simply by changing the software. In some instruments, however, the earlier stages of the analysis may be performed by (expensive) dedicated hardware, the main advantage of which is speed. We rarely need this speed for operator-controlled analyses; it is essential however for on-line analysis of images at video rates, which is still too fast for most small computers.

Image processing instrumentation

A block diagram of a typical small image processing system is shown in Figure 10.4. We will briefly examine the components in turn.

The *microscope* is similar to that used for manual image analysis. It is important that the optics are selected for high quality over the entire field of view; often cheaper optics, like camera lenses, are optimized for maximum performance on-axis and can show severe aberration near the edge of the frame. This is not a problem for many microscope applications, where the object of interest would normally be placed at the centre of the field by the operator, but for image analysis the computer collects the entire field and measures it subsequently. Consequently the constraints on resolution, field flatness, astigmatism, and aberration must be applied over the whole field, and so high quality optics

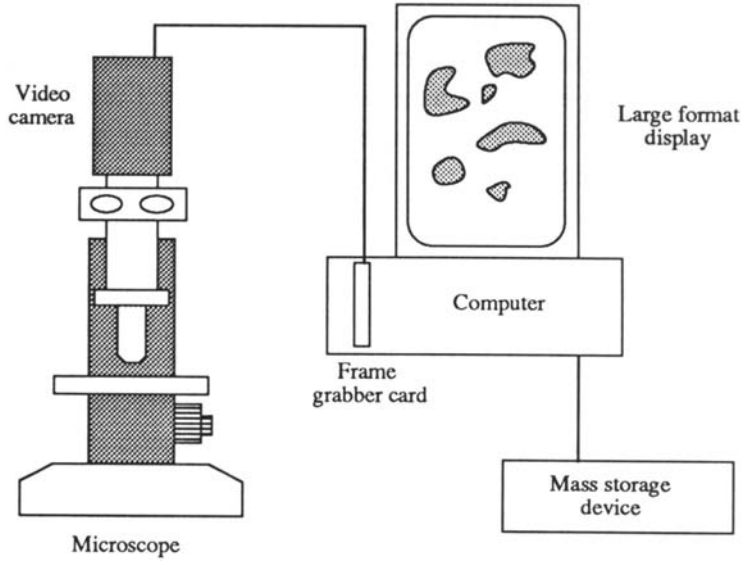


Figure 10.4. Typical image processing system.

are essential, or particles near the field edge may not be correctly measured. The illumination intensity must also be uniform over the field. Consequently it is worth purchasing the best optics available.

Concerning microscope optics, we once encountered a tricky problem which caused some considerable head-scratching. A home-made image analysis system appeared to have excellent field illumination uniformity when viewed directly, but showed hot spots when viewed through the vidicon camera. This was initially attributed to the rather contrasty nature of the camera and we put up with it for some weeks. We finally realized that the vidicon tube in the camera had a spectral response which extended a little into the infrared (at least a little further than our vision!) and it was the nonuniformity of this infrared illumination which was being detected. The problem was cured by installing an red/infrared blocking filter in the illumination optics.

Two options are available for the *camera.* These are conventional vidicon tube cameras, or the newer charge-

coupled device (CCD) cameras. Normally the older technology is the best choice; it can deliver similar performance to CCD cameras at a fraction of the cost. As an example, a small monochrome Vidicon camera currently costs £200, and delivers 600 line resolution, i.e can resolve an object 1/600th of the frame width. The corresponding 'economy' CCD camera costs £600 and has only 400 line resolution. Higher resolution versions of both technologies are available at extra cost, the practical limits being about 2000 lines (Photometrics' Star cameras, intended for astronomers, very expensive!). CCD cameras also suffer from image streaking if the camera or object is moving, an effect which is less severe in vidicon cameras. On the other hand, they use less power and have a higher low-light sensitivity than vidicons, and have considerably better field linearity and geometry. CCD cameras are rapidly being improved, and if this book ever runs to a second edition, I will probably have to report that they have displaced vidicon tubes in cost and performance terms.

Most cameras provide an analogue signal which must be converted to digital format before it can be processed by the computer. This is normally accomplished using a *frame grabber,* a small piece of hardware which consists of an analogue to digital converter and some memory to store the picture, possibly with a local processor for control purposes. Many frame grabbers now perform additional operations in hardware such as filtering, transforming, contrast control, scaling, etc. using chips called digital signal processors, which are dedicated analogue processing units. These will probably result in a significant reduction in the cost of dedicated hardware over the next few years.

The frame grabber operates by dividing the continuous analogue video signal into an array of points called *pixels,* each corresponding to a particular point in the image. The more pixels used to cover a certain area, the more detail is recorded, but this requires more memory for storage. There is no point in dividing the image into more pixels than the camera resolution provides; for example a 500-line camera signal could usefully be converted into a 500×500 pixel

image, but trying to feed the camera signal into a frame grabber capable of splitting it into a 1000×1000 image would not extract any more detail, since the detail does not exist in the original image. Also, the 1000×1000 image would require four times as much storage space, three-quarters of which would be correspondingly wasted. The frame grabber must specify the intensity (from white to black) of each pixel, which it does by allocating it a number, normally from 0 (black) to 255 (white). This is described as a *256 grey level* or *8-bit* representation. Cheaper systems with fewer grey levels are not suitable for serious work; alternatively, systems with more grey levels can rarely be justified unless the camera is of very high quality.

Frame grabbers are now available on a single board which will plug into a Mac or IBM card slot, and cost from £500. Software running on the host computer then allows the image to be displayed, processed, measured and stored to hard disc. This is convenient since the same computer can be used for other purposes, and also is very cost-effective.

The host computer needs to be fairly powerful since there is a lot of information in a 500×500 pixel image. Small desktop computers based on 80386/486 or 68030/40 processors are suitable for most simple tasks, but can be slow for computationally intensive operations such as image enhancement, using techniques such as Fourier transformation or deconvolution. In these cases a floating-point mathematics coprocessor is essential, and it may be worth considering hardware solutions for some of the processing steps; some image grabber boards have hardware for some of the more complex processing algorithms. Ultimately the decision to use such equipment lies in the time/cost tradeoff. For the most intensive operations, it is worth considering a workstation such as the Sun SPARCstation or VAXstation. Whatever system is chosen, it is essential to specify sufficient memory and disc storage; our Mac-based systems need 8 MB for comfortable image processing, and a large hard disc (200 MB or greater) for storage. If many images must be archived, a replaceable-

media device such as a worm drive or DAT streamer is essential.

It is perfectly possible to build one's own image processing facility from commercially available components. We use systems with Hitachi and Sony cameras and Mac-compatible frame grabber boards, with the processing being performed with the IPlab package (Signal Analytics) on a Mac 2CI with a standard 8-bit colour screen. A system like this is ideal for experimental purposes, where you have no idea what will be done from one day to the next. For more routine purposes, such as simple particle counting and sizing, a complete off-the-shelf system requires less investment in time and programming, and is probably more cost-effective, although the initial cost will be higher. Well-known manufacturers include Bausch & Lomb, Joyce-Loebl, Leitz, Olympus, Quantimet, Tracor, etc.

Image display

The image is stored in the computer memory as an array of numbers, each particular number representing a particular shade of grey. If colour images are being used, the frame grabber card will have split the camera signal into three separate pictures or *colour components,* corresponding to red, blue and green colours. Each of these can be displayed separately as a greyscale image, or they can be combined by the computer for display as a colour image.

In order to display the image, the computer needs to know what shade of grey or colour to display for each particular image value from 0 to 255. It does this by using a *look-up table* or LUT, which is a list of what colour or grey tone should be allocated to each pixel value. The default values in the LUT are normally those which cause the image to be displayed normally, i.e. with 0 for black through to 255 for white. However, if the LUT values are changed, the appearance of the image on the screen will alter without having to change the original data. For example, in a colour system the LUT entries for an image acquired as a greyscale picture could be changed to those corresponding to

a spectrum of colours. The image would then be displayed as a colour picture with e.g. red corresponding to the white tones, through the spectrum to violet corresponding to black tones. This technique is called *pseudocolouring* or *false colouring* since the displayed colours do not correspond to those of the original objects. Pseudocolouring is a powerful method of emphasizing the tonal relationships in an image, and can assist visualization considerably.

Image enhancement

Images acquired directly from the camera usually need some sort of processing or enhancement before the particles can be measured. This normally consists of a series of steps telling the computer what is a particle and what is not. The contrast between particles and background may be poor, particles may be touching, or there may be areas of the image which are not suitable for processing.

The most important enhancement technique used to select regions of an image is *thresholding.* This consists of specifying interactively a range of grey levels which correspond to the object of interest. For example, everything darker than a certain shade of grey could be displayed as black. Figure 10.5 shows typical images, one with a full grey scale, and one in which all the pixels darker than a specified level are set to black by the operator. The user adjusts the threshold level until all the particles are just coloured black, and separate from the lighter tones, and the computer can then be instructed to count the separate areas that the threshold has selected, and measure their size.

Even the most careful selection of threshold level cannot separate particles which are in contact. In Figure 10.6 the two touching particles would be counted as one large particle, although it is intuitively evident that the image consists of two particles. This problem can be countered to some extent using a technique called *erosion-dilation.* The operation of erosion simply consists of removing from the image those pixels which are at the edge of the particles. This removes the fine details from the particles, and

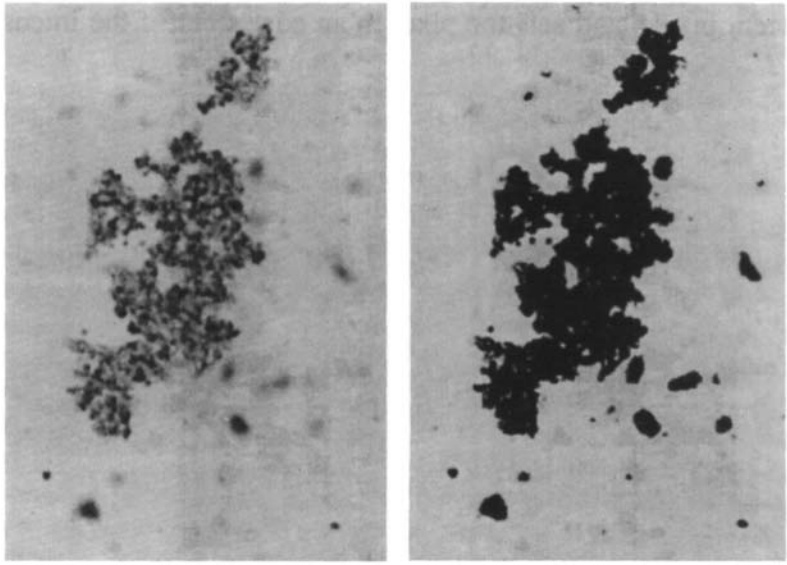


Figure 10.5. Selection of particles by thresholding.

touching particles usually separate in the process. The inverse operation of dilation adds new pixels to the particle at the edge, compensating for those removed by the erosion step. Particles can also be separated by a technique called watershed analysis (Beucher and Lantejoul 1979), which works if the touching particles are partly transparent by looking for the region of lower optical density between them.

A further technique for selecting particles is edge detection. Most software packages allow the edges of particles to be found by an algorithm which essentially measures the intensity change between adjacent pixels, and sets the pixel to an edge point if the intensity gradient exceeds a certain value. This allows particles to be automatically detected and measured, but high quality images are needed since noise may be interpreted as small particles.

Most image analysis packages allow a range of other image enhancement procedures, which are little used in particle size analysis. These include various types of filtering, which remove noise or emphasize detail. A commonly implemented filter type is an autoregressive

Figure 10.6. Separation of particles by erosion-dilation.

moving average or ARMA filter which can be tailored to provide a wide range of effects. A final technique which may be encountered is that of histogram equalization. This is a technique which enhances image contrast by effectively distorting the contrast scale so that the picture information is equally distributed among the available grey levels.

Image measurement

After the particles have been selected by edge detection or thresholding, measurement of the particle properties is normally automatic. Common options include measurement of all particles within a selected area, distribution of their areas or projected area diameters, distribution statistics such as standard deviation, skewedness and probability plots, particle asphericity or anisotropy, fraction of frame selected or thresholded, histogram distribution of intensities, etc. There is usually a option to add the particle measurements to those already in store from previous frames.

FRACTAL MEASUREMENT

Although it is possible to measure the fractal properties of aggregated or irregular particles from light scattering or neutron scattering data (Schmidt 1990), these approaches only work well for submicrometre particulates where the

scaling laws for scattering intensity are simply related to particle size. The majority of measurements of fractal properties for macroscopic systems have been performed by image analysis methods. The major disadvantage of this approach is that only a 2-dimensional silhouette of the particle is imaged. It is possible to show in some cases that the fractal dimension measured from the 2-dimensional image is the same as that of the 3-dimensional aggregate; Meakin (1990) has discussed the applicability of 2-dimensional measurements to the study of particle-particle (DLA) or cluster-cluster aggregates.

A number of techniques have been reported for extracting the fractal dimension from a 2-dimensional image. The 'classical' method of measuring the perimeter with 'rulers' of varying length (Kaye 1984) is not well-suited to computer evaluation. The most widely used are the 'box-counting' or 'mosaic amalgamation' (Kaye 1978) and erosion-dilation methods. The box counting method is especially suited to analysis of pixel-mapped images and is shown in Figure 10.7. The image is first subjected to an edge-detection algorithm which leaves only the pixels selected which lie on the edge of the particle. This image is then subdivided using an array of squares, and the number of squares which contain at least one pixel which belongs to the particle edge are counted. This is repeated using square arrays of varying edge length (in practice containing integral numbers of pixels, i.e. squares of edge 1,2,3…pixels). An estimate of the perimeter length at each measurement scale is then produced by multiplying the number of squares by the length of the side of the square. Effectively the process measures the perimeter length as a function of measurement resolution. The log of the perimeter length is then plotted against the log of measurement resolution to provide the conventional Richardson plot, and the fractal dimension is given by −(1+Islopel).

The erosion-dilation method performs a similar task; it was introduced by Flook (1978) who realized that it would be extremely rapid on image analysis systems (such as the Quantimet) which have modules to perform erosion-dilation

for particle-separating purposes. The image is dilated stepwise, a procedure which results in a gradually increasing area. A further copy of the image is similarly eroded, and the area after each erosion recorded. Subtracting the eroded image area after n steps from the corresponding n-step dilated image gives the area of the boundary region at a range of resolutions. The length of the boundary can be found by dividing the area by the width of the boundary. This is in fact not equal to the number of dilations (as I mistakenly thought when first using this method), because of odd effects associated with the geometry of the screen pixels. In order to find the boundary width for n dilations/erosions, you need to dilate and erode an isotropic Euclidean outline (a circle) whose boundary length is constant since it is not fractal. Dividing the boundary area by the width allows the length to be found, which can then be plotted against the boundary width to form a conventional Richardson plot.

REFERENCES

Agar A.W., Alderson R.H. and Chescoe D. (1974) *Principles and practice of electron microscope operation.* In: Practical methods in electron microscopy, Vol. 2., North-Holland, Amsterdam.

Beucher S. and Lantejoul C. (1979) *Proc. Int. Workshop on Image Processing,* CCETT, Rennes, France.

Dullien F.A.L., Rhodes E. and Schroeter S.R. (1969) *Powder Technology* **3** 124–135

Flook A.G. (1978) *Powder Technology* **21** 295–298.

Kaye B.H. (1978) Report 21, Institute for Fineparticle Research, Laurentian University.

Kaye B.H. (1984) *Particle Characterization* **1** 14–21.

Lacey A.J. (1989) *Light Microscopy in Biology. A practical approach.* IRL Press, Oxford.

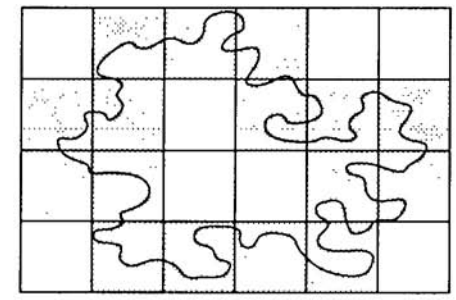
Lin'kova O.S., Zhukovskaya S.A. and Davidova N.S. (1976) *Khim. Farm.* **10** 143–148.

List P.H. and Groenig J.M. (1976) Pharm. Ind. **38** 1159–1171.

Meakin P. (1990) Simulations of aggregation processes. In Avnir D. (ed), *The fractal approach to heterogeneous chemistry,* Wiley, Chichester.

Russ J.C. (1990) *Computer-assisted microscopy.* Plenum Press, New York.

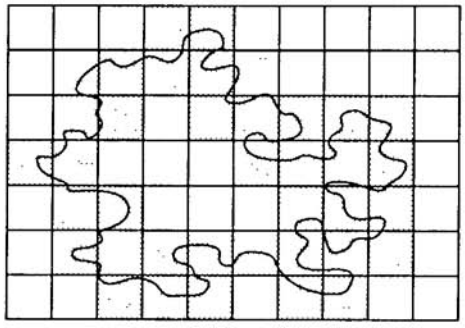
Saltzman W.M., Pasternak S.H. and Langer R. (1987) *Chem. Eng.* Sci. **42** 1989–2004.



Square 12 mm

16 squares occupied

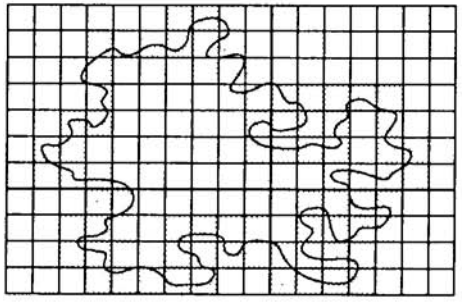
Border length 192 mm



Square 8 mm

38 squares occupied

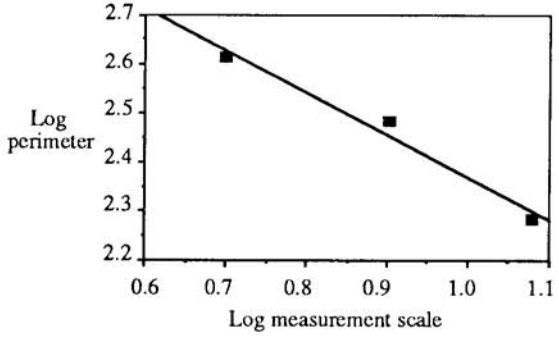
Border length 304 mm



Square 5 mm

82 squares occupied

Border length 410 mm



Gradient = -0.86
D = 1.86

Figure 10.7. Fractal dimension measurement by box counting.

Schmidt P.W. (1990) Use of scattering to determine the fractal dimension. In Avnir D. (ed), *The fractal approach to heterogeneous chemistry,* Wiley, Chichester.

Sims E.E. and Withington R. (1983) *Int. J. Pharmaceut.* **17** 167–181.

Index