Light Scattering Demystified

Theory and Practice

Øgendal, Lars Holm

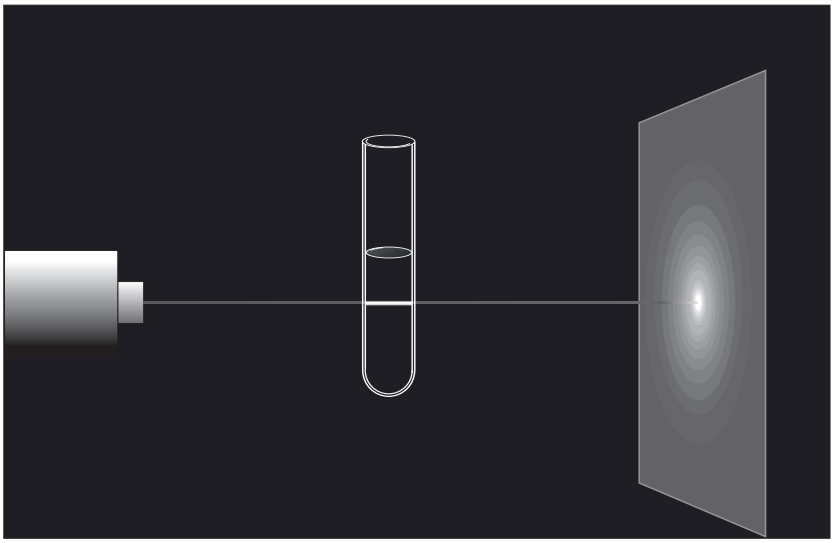
Publication date:
2017

Document version
Publisher's PDF, also known as Version of record

Light Scattering Demystified

Theory and Practice

Lars Øgendal



University of Copenhagen 12th September 2017

Contents

Preface

This is an introduction to light scattering conceived as *a measurement technique*. The actual physical phenomenon of *light scattering* is dealt with to an extent which is necessary and hopefully sufficient to understand the inner workings of the technique, to understand its strengths and limitations and which is relevant to the interpretation of data. There seems to be a lack of simple and accessible introductions to light scattering so the scientist or student who wants to get acquainted with this field has the choice of either reading a few pages in a physical chemistry textbook or embark on the the rather inaccessible specialist literature. A third source of information is the sales prospects from the companies that sell light scattering instruments but of course this material is influenced by its purpose: to sell light scattering instruments. So don't expect any scientific depth or honesty regarding the limitations, hassles and pitfalls of the method. This introduction to the field of light scattering is written with the intention of being relatively accessible, of course at the cost of mathematical rigour.

1

Introduction

1.1 What is light scattering?

In the following we shall treat light itself sometimes as photons, sometimes as electromagnetic waves whichever gives the simplest description.
The very phenomenon that light that strikes a particle (e.g. a molecule) or a collection of particles, thereby changing its direction, is called *light scattering*. This is a somewhat loose definition of light scattering and it is also not quite true, as we shall see later. For example it turns out that light can actually be scattered in the forward direction, i.e. without changing direction at all! And even the notion that light actually *strikes* anything turns out to be unfruitful.
To keep things simple one could say that light – like any other kind of electromagnetic radiation – *interacts* with matter in mainly two different ways:

1. Absorption (The photons disappear)
2. Scattering (The photons change their direction)

A third way of interaction exists: The light can change its state of polarisation. This type of interaction we will disregard here. In the case of absorption it may happen that light is being (re-)emitted at a different wavelength, i.e. as fluorescence or phosphorescence. This phenomenon is usually unwanted when using light scattering as a measurement technique because it poses the problem to the experimenter to make sure that the light measured is only scattered and not due to fluorescence. Also, scattering can be subdivided into elastic and inelastic scattering: If the scattered light has *exactly* the same wavelength as the incident light meaning that the scattered photons have *exactly* the same energy as the incident photons, the scattering is called elastic. If the photons come out of the scattering process with a changed energy the scattering process is termed *inelastic*. This means that the scattered light will have either a longer or a shorter wavelength than that of the incident light. Examples of inelastic light scattering are Raman scattering and Brillouin scattering. Neither of these type of scattering will be dealt with here. One exception however is *quasi elastic light scattering*, termed so when the scattered light has very nearly the same wavelength as the incident light. In practice this term is reserved to describe scattering of light on moving particles where the movement of the particles changes the wavelength of the scattered light by the Doppler effect. The technique called *dynamic light scattering*, DLS, is based on this Doppler shift in wavelength and is also called *quasi elastic light scattering* or QELS.

The three phenomena, reflection, refraction af diffraction are also also scattering phenomena but require the scattering particles to be highly ordered. In all of the following we will assume that the scattering particles are present in *a solution* free to move and tumble around by Brownian motion and thus inherently in an unordered state. The miracle of light scattering is that contrary to what one would think not all information about the particles is wiped out. It turns out to be possible to derive useful parameters about the dissolved particles like e.g. molecular weight, size and to some extent, shape.

Both of the above mentioned main interaction phenomena have as a result that a light beam is being attenuated when it passes through a solution[1] (see figure 1.1). The intensity of the light beam is weakened by either absorption or scattering (or both). In both cases the transmitted intensity decreases exponentially with the thickness x of the layer of material through which the light has to pass. In the context of absorption the transmitted intensity I is usually written:

$$I = I_0 \cdot 10^{-\alpha x}$$

while in the context of scattering it is frequently written as:

$$I = I_0 \cdot e^{-\tau x}$$

The quantities α and τ are called the absorption coefficient and the turbidity, respectively. The difference in the base of the exponential is just a matter of convention. Just as often the base 10 for the turbidity is used. As mentioned previously

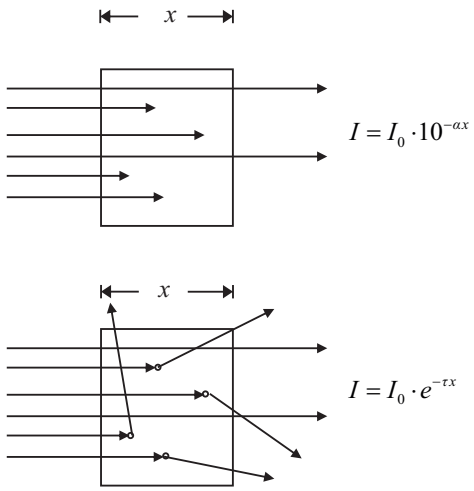


Figure 1.1: The transmitted light beam is attenuated due to absorption (top) or scattering (bottom).

we shall assume in the following that the system which scatters light is a solution

[1] This also holds true for gasses and solids

of the particles under investigation.

The principles of light scattering (e.e. the technique) is most easily discussed on the basis of a sketch of a typical setup for the measurement of static light scattering. One such is shown in figure 1.2. The sample is a solution in a *cuvette* which is usually cylindric. A monochromatic light source (i.e. a light source emitting light of only one wavelength, usually a laser) shines light on the sample. The intensity or the power of the scattered light is being measured at some known scattering angle by a detector which may be a photodiode or a photomultiplier tube.

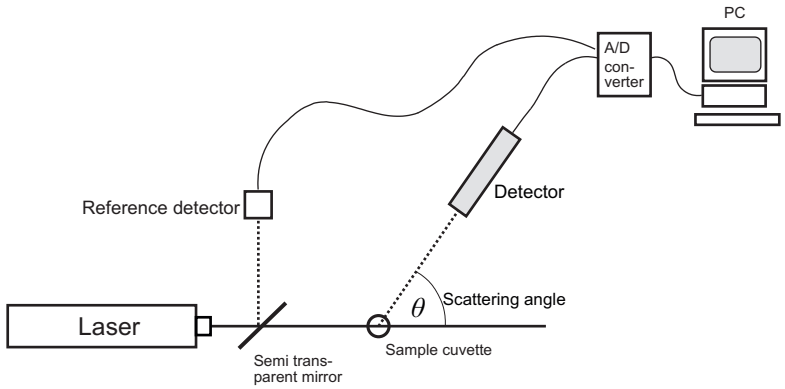


Figure 1.2: Sketch of a setup for the measurement of static light scattering, SLS. The intensity (or the power) of the scattered light is being measured as a function of the scattering angle θ, i.e. the angle of observation of the detector. The measured intensity of the scattered light is divided by the intensity of the incident light which consequently also has to be measured using a reference detector.

While absorption of light (e.g. in the detector) is most easily discussed in terms of the photon concept, scattering of light is most easily understood in terms of a wave description of light, i.e. considering light as electromagnetic radiation.

Scattering of light can explain many everyday phenomena like why the sky is blue, why clouds are white and why they are visible at all while the water vapour that creates them is invisible, why milk appears white even though it consists of

completely transparent components, etc. .... But as previously mentioned light can also be used to gain information about molecules (usually macromolecules) and particles in suspension.

On the following pages we will describe an discuss two measurement techniques, namely *static light scattering*, SLS and *dynamic light scattering*, DLS. The techniques will be described in that order because it is necessary to understand SLS in order to understand properly DLS. At this point, however, we can already reveal that the two techniques are complementary in the sense that they take advantage of two different an completely independent features of the scattered light: SLS measures the *average* intensity of the scattered light at many different scattering angles and uses this to derive information about molecular weight, particle size, particle shape and particle interactions. DLS, on the other hand, disregards the intensity of the scattered light but uses the temporal *fluctuations* in the scattered light to derive spectra of typical fluctuation times which in turn yield information about the diffusion coefficient(s) of the particles and the indirectly about their size. The difference is shown in figure 1.3.

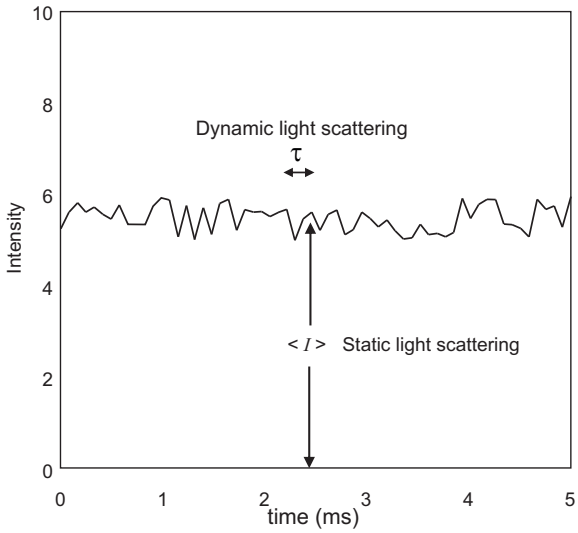


Figure 1.3: Light scattered from a solution of macromolecules has a mean intensity $\langle I \rangle$ that reflects the molecular weight of the particles while the *fluctuations* in the intensity have a characteristic fluctuation time τ that reflects the diffusion coefficient of the particles.

So, in conclusion about the difference between static and dynamic light scattering one can say:

SLS : Static light scattering employs measurement of the intensity of the scattered light at many different angles (typically 10 – 100). The intensity is usually an average over approximately one second or more. The information about the molecular weight and size of the molecule lies in the *intensities* themselves at the different scattering angles.

DLS : Dynamic light scattering employs measurement of long time series of mean intensity of the scattered light where the averaging is over very short time intervals, typically $100\,\mathrm{ns}$. Averaging over such short time intervals ensures that the magnitude of the intensity fluctuations within the time series become significant. The information about the molecules (diffusion coefficient or size) lies in the typical fluctuation times for the scattered intensity.

2

Simple theory for static light scattering, SLS

Since the late 19'th century light has been used to determine the size of particles (Tyndall 1869, Rayleigh 1871 and 1881) and later in the 20'th century to determine the molecular weight of macromolecules (Debye 1944 and 47). The technique employed is called *static light scattering* or SLS. The method is based on the fact that when light passes through a solution of molecules (particles) some portion of the light will be scattered in all directions. If all of the light is scattered the solution (or suspension) will appear completely opaque. This is the case with e.g. milk which appears white because light is scattered by microscopic fat globules and casein aggregates, called casein micelles although they are not micelles in the physical meaning of the term. In other situations the scattered fraction of the light

is so small that the solution appears completely transparent to the naked eye. This would be the case if one looked at a solution of BSA (bovine serum albumin), a protein with a molecular weight of approx. 66400 g/mol at a concentration of, say, 1 g/L. In this case it requires a very intense light source (i.e. a laser) to see the scattered light. With a 10 mW laser the scattered light shows a glowing trace in the solution (see figure 2.1.) It is this latter situation of weak scattering which

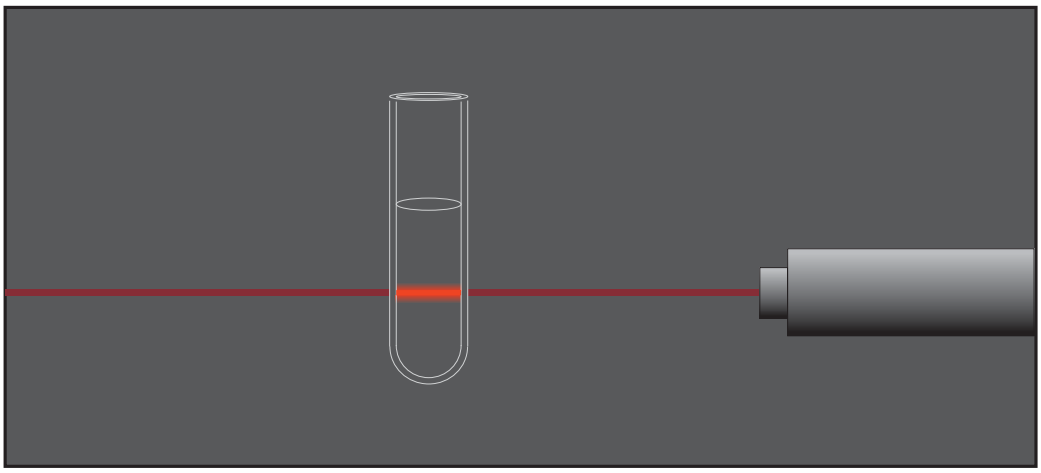


Figure 2.1: The laser to the right sends out an intense beam of light which - in principle - is invisible when seen from the side. When the laser beam passes through a solution of large molecules, like e.g. BSA, some of the light is scattered by the molecules and becomes visible from the side as a glowing trace.

is relevant when light scattering is used to determine molecular weight and size of molecules.
In order to be useful for quantitative light scattering the light itself has to fulfil two conditions:

- The light has to be *monochromatic*, i.e. to have a well-defined wavelength.
- The light has to be in a *collimated* beam, i.e. the rays have to be parallel.

Both conditions are ideal ones: In practice one will have to accept some distri-

bution of wavelengths and some angular divergence of the rays. Ordinary white light consists of all wavelengths between approximately 390 and 700 nm and is therefore highly non-monochromatic. In the early days of light scattering one had to employ colour filters, diffraction gratings or prisms to single out a narrow selection of wavelengths from the light source. This of course meant that the useful intensity of the light source was dramatically decreased. The more narrow the selected spectrum of light the weaker the (remaining) light would be. As for the collimation (i.e. the parallelity) this was achieved by having a very small light source e.g. by screening a larger source with slits or diaphragms subsequently making the rays parallel by a system of lenses. Again, the limiting factor for the parallelism is the size (the extension) of the light source which in turn means that a higher degree of collimation would necessarily be linked with a relatively small and therefore effectively weaker light source. As the precision of light scattering measurements is related to the monochromaticity and the collimation of the light used[1] it is seen that a trade-off has to be made because a weak light source in itself impairs the precision of the measurements.

This dilemma is only of historical interest today because the present day light sources are lasers which by their nature emit inherently monochromatic light at a very high intensity. And laser light is also easily collimated without loss of intensity. A third benefit from using lasers is that the beam can be made such that it has the desired *direction of polarisation*. This is a technical advantage as we shall see later but has the minor drawback that the equations used for the interpretation of light scattering data have to be used with great care if they are taken from literature prior to the advent of the laser (1960). Always check if the equations pertain to un-polarised light or polarised light. We shall briefly mention the old equations later ?? for reference and warning.

[1]The monochromaticity and collimation are much more important if the purpose of the light scattering measurement is size determination and not determination of molecular weight.

2.1 The Rayleigh ratio, R

When a light scattering experiment is done on a solution (i.e. a suspension of particles) the purpose is of course to deduce properties of *the solution* and not have this mixed up with properties of *the apparatus*. It is evident that if the laser is replaced by one of twice the power the intensity of the scattered light will be doubled. And if the distance between the scattering solution and the detector is doubled the measured intensity will be reduced by a factor of four as the observed intensity is inversely proportional to the distance between the source and the receiver. In order to obtain a measure which is *independent of apparatus characteristics* but only depends on the properties of the system one defines *the Rayleigh ratio*, R_θ, for the solution under investigation:

$$R_\theta = \frac{I_s(\theta) \cdot r^2}{I_0 \cdot V_s(\theta)} \tag{2.1}$$

where the subscript θ indicates that the Rayleigh ratio is a function of the scattering angle, $I_s(\theta)$ is the intensity of the scattered light measured at the angle of observation θ cf. figure 1.2, I_0 is the intensity of the laser used, $V_s(\theta)$ is *the scattering volume* (i.e. the volume which is illuminated by the laser while at the same time being visible for the detector (see figure 2.2), and r is the distance from the scattering volume to the detector. It turns out that the scattering angle θ, although easy to envisage, makes the equations describing light scattering cumbersome to work with. So, when we get to section 2.6 we abandon the scattering angle θ and instead introduce the so-called *scattering vector* $\vec{q}$ (see later in this chapter and figure 2.11) the length of which, q can be calculated as $q = (4\pi n/\lambda)\sin(\theta/2)$. When the scattering angle is effectively replaced by the length q of the scattering vector it then becomes more natural to conceive the Rayleigh ratio as function of q:

$$R(q) = \frac{I_s(q) \cdot r^2}{I_0 \cdot V_s(q)} \tag{2.2}$$

We shall see in the following sections that $R(q)$ is a function which depends on:

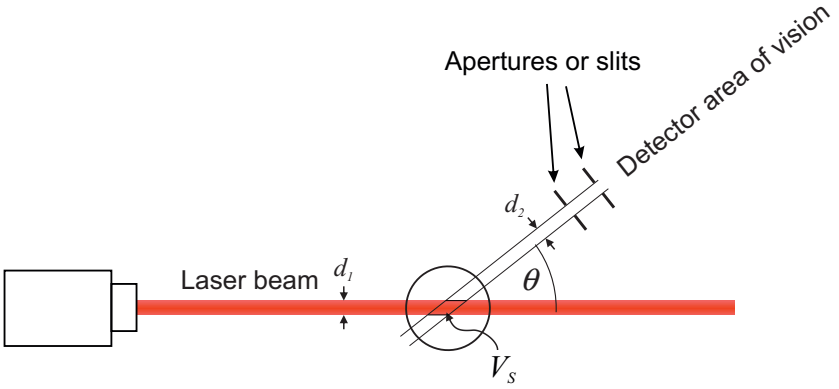


Figure 2.2: Due to the apertures usually present in a detector it only sees light that enters within a certain narrow field of observation (approximately a cylinder) with a diameter of d_2 whereas the illuminated volume is approximately a cylinder of diameter d_1. The intersection between these two volumes is called *the scattering volume*, V_s. The scattering volume attains its minimum value, V_0, when the laser beam and the angle of observation are perpendicular to each other, i.e. when $\theta = 90°$. The frequently cited relationship $V_s(\theta) = V_0/\sin(\theta)$ holds only approximately.

- the molecular weight of the suspended particles (molecules)
- the concentration of the suspended particles (molecules)
- the refractive index of the pure solvent
- the refractive index of the suspended particles (molecules)
- the size of the suspended particles (molecules)
- interaction forces between the suspended particles

2.2 What is light?

As mentioned at the beginning, the picture of light scattering as photons bouncing off the particles in the solution is not fruitful. Instead, classic electromagnetic theory (EMT) where light is treated as electromagnetic waves is much more straightforward. Light is described as an electromagnetic wave, an electric wave and a magnetic wave mutually generating each other as they propagate with the speed of light (see figure 2.3).

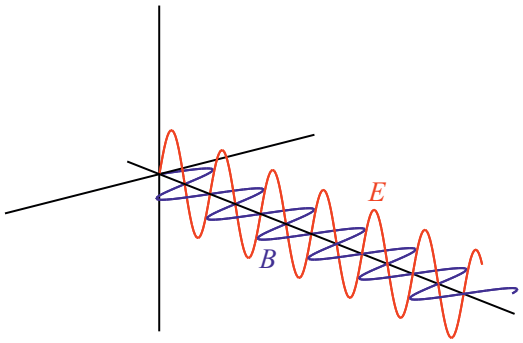


Figure 2.3: An electromagnetic wave propagates along the x-axis. The magnetic field $\vec{B}$ and the electric field $\vec{E}$ are perpendicular to each other and to the direction of propagation. The electric field is strictly vertical here, giving it only a z-component.

Light which propagates along the direction of the x-axis of some coordinate system with the electric field oscillating in the direction of the z-axis is said to be vertically polarised (i.e if the $z-$axis is in the vertical direction, of course). If the vertically polarised light is of one colour, i.e. of one wavelength λ or equivalently of one frequency ν, its electric field can be described by the following equation:

$$E(x,t) = E_z(x,t) = E_0 \cos(2\pi\nu(t - x/c) + \phi) \tag{2.3}$$

where ν is the frequency of the light, c is the speed of light and E_0 is the amplitude of the electric field component of the light. The quantity ϕ is just a constant which defines the *phase* of the cosine function at time $t = 0$ and at $x = 0$. In light

scattering we normally don't bother about the magnetic field as this normally does not interact with organic matter.

In order to simplify the appearance of equations it is customary to define the *wavenumber* k as

$$k = \frac{2\pi}{\lambda}$$

where λ is the wavelength of the light *in the medium through which the light propagates*. The equation can also be written

$$k = \frac{2\pi n}{\lambda_0}$$

where λ_0 is the wavelength of the light in vacuum and n is the refractive index of the medium. In the case of light propagating in water the refractive index is just $n = 1.33$. Also, for simplifying purposes, we define the *cyclic frequency* ω of the light as

$$\omega = 2\pi\nu$$

As $\nu/c = 1/\lambda$ we can finally rewrite equation 2.3 as:

$$E(x,t) = E_y(x,t) = E_0 \cos(\omega t - kx + \phi) \tag{2.4}$$

This equation is easily generalised to propagation in any direction and a point of observation not necessarily lying on the x-axis: First of all the wavenumber k is replaced by the *wavevector* $\vec{k}$ defined as a vector pointing in the direction of propagation and with a length $|\vec{k}| = \frac{2\pi n}{\lambda_0}$. And the point of observation is now given by a vector $\vec{r}$ instead of just an x-coordinate. The generalised version of equation 2.4 then reads:

$$E(\vec{r},t) = E_0 \cos(\omega t - \vec{k}\cdot\vec{r} + \phi) \tag{2.5}$$

According to EMT the *intensity* of the light is given by the time average of the square of the electric field:

$$I_0(x) = \epsilon_0 c \langle E^2(x,t)\rangle \tag{2.6}$$

where $\epsilon_0 = 8.85 \cdot 10^{-12}\,\mathrm{F}\cdot\mathrm{m}^{-1}$ is the so-called vacuum permittivity and $c = 3.0 \cdot 10^8\,\mathrm{m}\cdot\mathrm{s}^{-1}$ is the speed of light in vacuum. Inserting the expression for

$E(x,t)$ we get:

$$I_0(x) = \epsilon_0 c\langle E^2(x,t)\rangle = \epsilon_0 c\langle E_0^2\cos^2(\omega t - \vec{k}\cdot\vec{r})\rangle = \epsilon_0 c E_0^2\langle\cos^2(\omega t - \vec{k}\cdot\vec{r})\rangle \quad (2.7)$$

As the mean value of $\cos^2(\omega t - kx)$ calculated over one cycle equals $\frac{1}{2}$ and $\omega \approx 3\cdot 10^{15}\,\mathrm{radians}\cdot\mathrm{s}^{-1}$ for a typical HeNe laser the averaging over even some extremely short time interval like e.g. one nanosecond would mean averaging over approximately half a million cycles. This means that even if this short averaging time does not contain a whole number of cycles the result would still be extremely close to $\frac{1}{2}$. Hence we get:

$$I_0(x) = \frac{1}{2}\epsilon_0 c E_0^2 \quad (2.8)$$

2.3 Why and how do particles scatter light?

In order to understand why and how particles or molecules can scatter light it turns out that one needs to understand what an electric *dipole* is. Having touched upon this subject it turns out to be a good idea to answer the questions in several steps: First we shall look at how light is scattered from one very small (i.e. point-like) particle. Next, how light is scattered from many small particles. And this latter issue will be divided into a discussion on how light is scattered if the particles are non-interacting (an ideal solution) and a discussion on how particles in non-ideal solutions scatter light. Fundamentally, of course, light is scattered by the same laws in both cases but the equations describing the scattering quantitatively are different. After having looked at scattering from very small particles we turn towards scattering from large particles.

2.4 Dipoles

An electric dipole is an object which has an asymmetrical distribution of charge, usually such that the object has one end with positive charge and the other end with negative charge (see figure 2.4). The dipole moment μ is defined – in the simple case shown in the figure – as $\mu = d\cdot Q$, i.e. as the separation between the positive $+Q$ and the negative charge $(-Q)$ multiplied by the magnitude of the

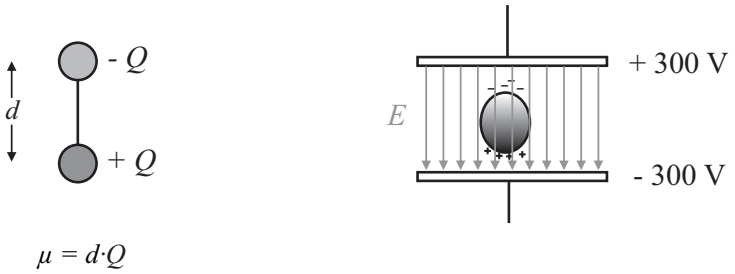


Figure 2.4: On the left side is shown an electric dipole. The dipole moment μ is defined as the product of half the charge difference (between the two ends) and their mutual distance. On the right is shown how a particle becomes a dipole when under influence of an external electric field.

charge. The dipole moment can be either *permanent* or *induced*, i.e. caused by an external electric field. When it comes to scattering of light it is only induced dipole momenta that have relevance. In the more general dipole case shown on the right in figure 2.4 the dipole moment must be defined by an integral [2]. What is interesting about the induced dipole moment is that its magnitude depends on the strength of the external field E in a linear way:

$$\mu = \alpha \cdot \epsilon_0 \cdot E$$

where α is the so-called polarisability of the particle. The vacuum permittivity, ϵ_0, is included due to tradition[3]. With this definition of the polarisabilityit has the SI-unit m^3. The linearity holds at least as long as the electric field is not *very* strong. Lasers used in light scattering do not exceed the limit of linearity even when the beam is focussed. The polarisability α of the particle is a measure of how easily an external field can move charges within the particle. The larger the value of α the further the charges will move with a given external field strength. This description of the behaviour of dipoles refers to particles in vacuum. But

[2]The dipole moment is actually a vector, $\vec{\mu}$. For a distribution of n point charges Q_i it can be calculated as $\vec{\mu} = \sum_i^n Q_i \vec{r_i}$ and for a continuous distribution as $\vec{\mu} = \iiint \vec{r}\, \rho(\vec{r}) dV$ where $\rho(\vec{r})$ is the charge density (coulomb per cubic meter). The integration is extended over the body that contains the charge.

[3]Sometimes the vacuum permittivity is excluded from the defining equation which then reads : $\mu = \alpha \cdot E$. This of course has the consequence that the polarisabilityof a given material will have a different value and a different unit

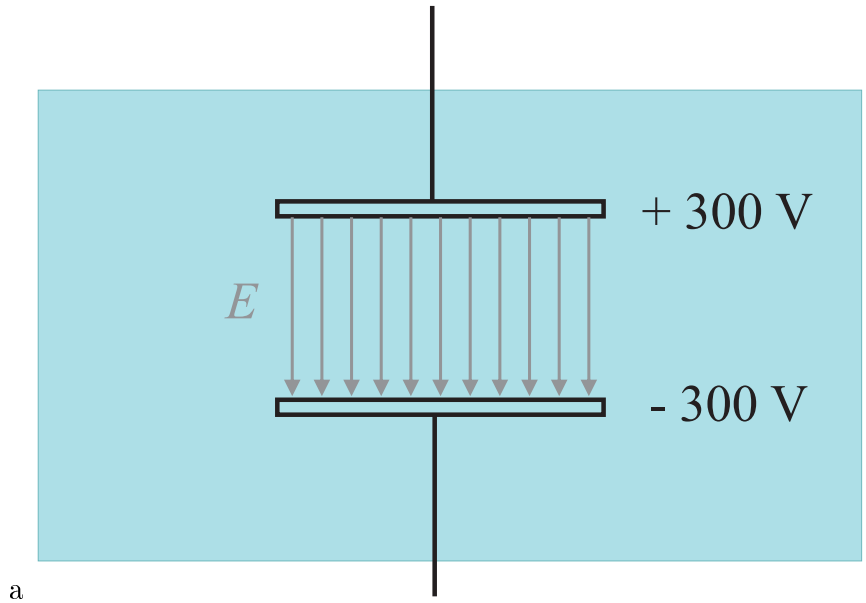
light scattering experiments are usually carried out on particles in solution, i.e. particles embedded in a medium. How to deal with this is described in figures 2.5 and 2.6. Suppose an electric field is present in a medium (figure2.5 a). Charges may move in the medium under the influence of the electric field but no net charge is generated anywhere. Now, take out a "plug" of the medium and put it in the electric field (figure2.5 b) without the rest of the medium. This plug will be polarised in the electric field with the net result that some charge will be removed from one end of the plug and put at the other end of the plug. The plug will have surface charges at the ends. If we do the opposite, i.e. remove the plug from the medium (figure2.6 a) leaving a hole in it, the hole will also be polarised but exactly opposite to the polarisation of the plug. It will have charges on the inner surface that exactly match the surface charges on the plug. Because if the plug is put back into the hole all charges will cancel as in figure figure2.5 a. Now, put a plug of the same size and shape but made of a different material than the medium into the hole in the medium. If this plug has a polarisability α_{plug} which is different from that of the medium α_{medium} the net dipole moment of the plug in the medium will be $\mu_{\text{net}} = \alpha_{\text{plug}} \cdot \epsilon_0 \cdot E - \alpha_{\text{medium}} \cdot \epsilon_0 \cdot E$ or

$$\mu_{\text{net}} = (\alpha_{\text{plug}} - \alpha_{\text{medium}}) \cdot \epsilon_0 \cdot E$$

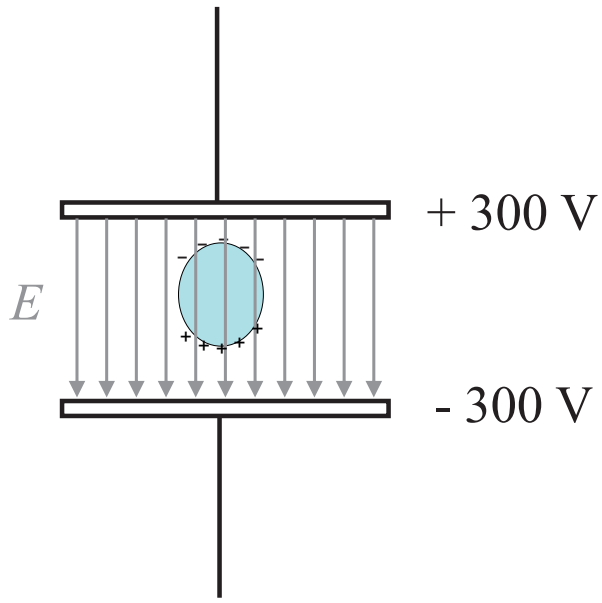
The quantity $\alpha_{\text{plug}} - \alpha_{\text{medium}}$ is called *the excess polarisability* of the particle. It is this quantity which governs the light scattering of particles in solution. When we refer to the polarisabilityof particles from now on it should be understood as the excess polarisability.

2.5 Scattering of light from one small particle

As is evident from the previous discussion of polarisabilityit is most convenient to describe light as a propagating electromagnetic wave rather than beam of individual photons. The way the scattering comes about is that the electric field of the incoming light moves the charges in the particles (molecules) thereby turning them into dipoles (see figure 2.7). The propagating electromagnetic wave consists of an electric and a magnetic field that oscillate perpendicularly to each



a



b

Figure 2.5: An electric field is present in a medium in figure a. Charges may move in the medium under the influence of the electric field but no net charge is generated anywhere. A "plug" is taken of the medium and put it in the electric field figure b while the rest of the medium is removed. This plug will be polarised in the electric field with the net result that some charge will be removed from one end of the plug and put at the other end of the plug, i.e. the plug will acquire surface charges at the ends.

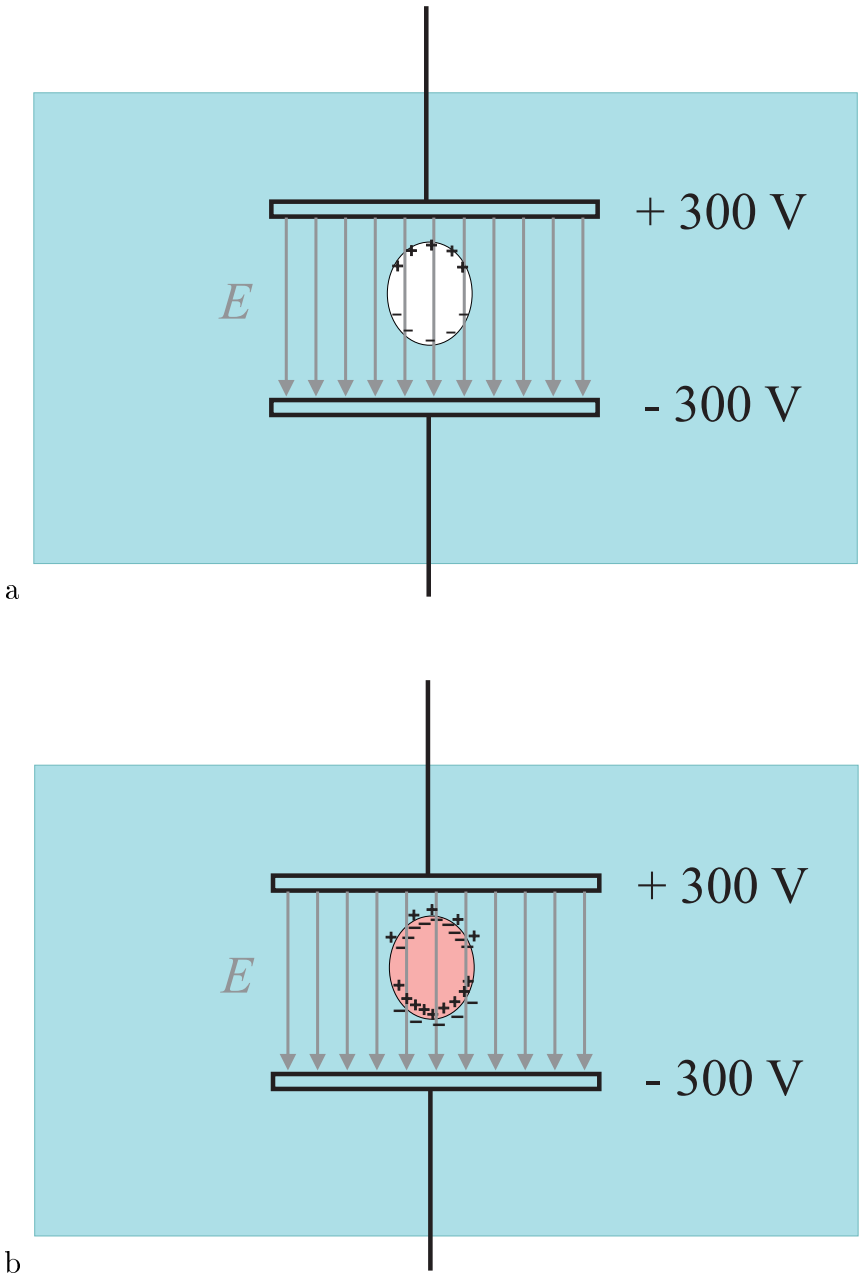


Figure 2.6: Here the plug is removed from the medium (figure a) leaving a hole in it. The hole will also be polarised but exactly opposite to the polarisation of the plug. It will have charges on the inner surface that exactly match the surface charges on the plug that was removed. Because if the plug is put back into the hole all charges will cancel as in figure 2.5 a.
In figure b a plug of the same size and shape but made of a different material than the medium is put into the hole in the medium. If this plug has a polarisabilityα_{plug} which is different from that of the medium α_{medium} the net dipole moment of the plug in the medium will be $\mu_{\text{net}} = \alpha_{\text{plug}} \cdot \epsilon_0 \cdot E - \alpha_{\text{medium}} \cdot \epsilon_0 \cdot E$ because it reflects the difference in the surface charge of the hole and of the new plug.

other and perpendicularly to the direction of propagation. The oscillations have the frequency ν (and hence the cyclic frequency $\omega = 2\pi\nu$). Only the electric field is (usually) of importance for light scattering as this is what exerts a force on the charges in the molecule (figur 2.7).

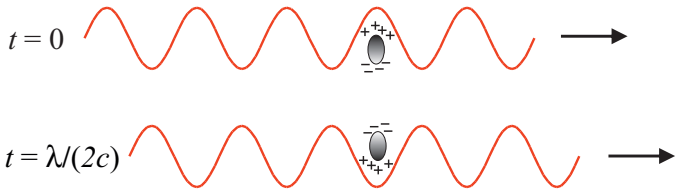


Figure 2.7: The external, oscillating electric field E displaces the electric charges in the particle rendering it negative at one end and positive at the other end. Hereby the particle becomes an "oscillating dipole" with a dipole moment that can be written as $\mu = \mu(t) = \alpha E = E_0 \cos(\omega t - kx) = \mu_0 \cos(\omega t - kx)$ where x is the position of the molecule.

The charges in the molecule move in sync with the incoming electric field turning the molecule into an oscillating dipole, oscillating with the same frequency as the light that shines on it. Not only do the charges in the molecule move in sync with the external electric field they also move simultaneously in the same direction within the molecule because the electric field has (practically) the same magnitude everywhere in the molecule when the molecule is sufficiently small compared to the wavelength of the light. Consequently, according to the theory of electromagnetic radiation (EMT), the molecule emits electromagnetic radiation into all directions (with an intensity which depends on the direction).
We shall here consider the situation where the incoming light is *polarised* [4] meaning that the electric field always oscillates in the same plane (see figure 2.8)
Now, let's make things quantitative. Remember that the molecule or particle is

[4]The reason for looking at polarised light is that in practice the light source is now always a polarised laser. In the earlier days of light scattering (up to the nineteen sixties) light sources were generally unpolarised and consequently the equations describing light scattering took this into account. This means that care should be taken if equations from older literature is used to interpret light scattering data from modern equipment.

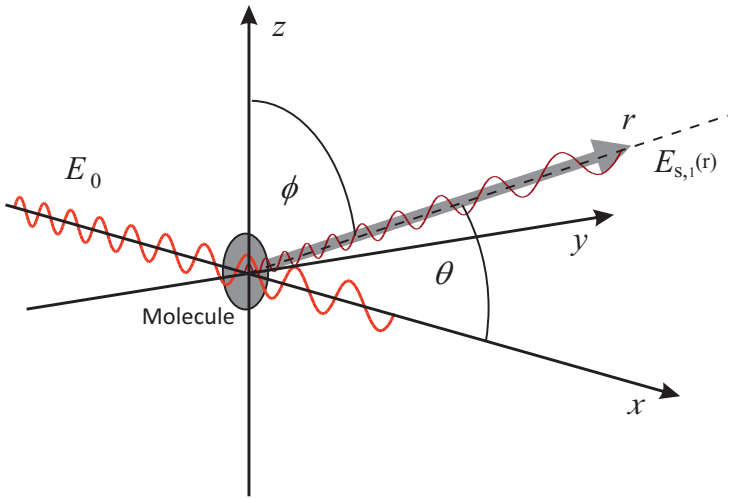


Figure 2.8: Light polarised in the $xz-$plane hits a molecule much smaller than the wavelength of the light. The electric field of the light shakes the electrons of the molecule thereby turning it into an oscillating dipole which emits radiation in all directions. The intensity of the emitted light depends on the direction of emission given by the two angles θ and ϕ.

assumed to be much smaller than the wavelength λ of the incoming light. For simplicity the particle is placed at the origo of our coordinate system. The electric field of the light that impinges on the molecule can be described as a wave propagating along the $x-$axis according to equation 2.4.

The electric field strength $E_{s,1}(r)$ of the radiation scattered from one particle (hence the number 1 in the subscript) at the distance r from the molecule (cf. figure 2.8) is according to EMT given by:

$$E_{s,1}(r) = E_0 \cdot \left(\frac{\pi \alpha \sin \phi}{r \lambda^2} \right) \cos(\omega t - kr) \tag{2.9}$$

where λ is the wavelength of the light (in the medium surrounding the molecule), α is the polarisabilityof the molecule and r is the distance from the molecule to the detector. Note that $E_{s,1}(r)$ does not depend on the angle θ relative to the x-axis.

Using equations 2.9 and 2.6 we obtain *the intensity* of the scattered light:

$$I_{s,1}(r) = \langle \epsilon_0 c E_{s,1}^2(r) \rangle = \frac{1}{2} \epsilon_0 c E_0^2 \cdot \frac{\pi^2 \alpha^2 \sin^2 \phi}{r^2 \lambda^4} \tag{2.10}$$

or, applying also equation 2.8:

$$I_{s,1}(r) = I_0 \cdot \frac{\pi^2 \alpha^2 \sin^2 \phi}{r^2 \lambda^4} \tag{2.11}$$

In practice most light scattering instruments have the detector placed in the $xy-$plane so from now on we shall assume that the angle $\phi = 90°$. Hence we get the following expression, valid for the scattered intensity from a small molecule:

$$I_{s,1}(r) = I_0 \cdot \frac{\pi^2 \alpha^2}{r^2 \lambda^4} \tag{2.12}$$

Here we see, not surprisingly, that the intensity of the scattered light at the position of the detector is proportional to the intensity I_0 of the incoming light and inversely proportional to the square of the distance from the scattering molecule to the detector. What seems perhaps more surprising is the fact that the scattered intensity is inversely proportional to the fourth power of the wavelength of the impinging light. This means that blue light ($\lambda_0 \approx 480\,\mathrm{nm}$ is scattered approximately four times more effectively than red light with a wavelength of $\lambda_0 \approx 690\,\mathrm{nm}$. This fourth power dependence on the wavelength is the well-known explanation why the sky is blue.

This kind of scattering where the scattering particles are much smaller than the wavelength of the light[5] is called *Rayleigh scattering* and is characterised by being *isotropic* i.e. equally intense in all directions relative to the direction of the incoming light (it is independent of the scattering angle θ c.f figure 1.2) and the scattered light also has the same wavelength as the incoming light. Looking at figure 2.9 it is intuitively clear that α must be proportional to the "size" of the molecule (i.e. volume or molecular weight) as putting two small particles each with the molecular polarisability α_1 next to each other will give a total dipole moment of $\mu = (2Q) \cdot d$ whereas putting the two particles end-to-end will give a total dipole moment of $\mu = Q \cdot (2d)$. In both cases the total dipole moment is simply twice the dipole moment of one particle.

Following the same line of thought is is evident that n particles have the dipole moment $\mu = n\mu_1$ whereby the compound particle gets the polarisability $\alpha = n\alpha_1$.

[5] actually also the polarisability α needs to be small. We shall return to this point later

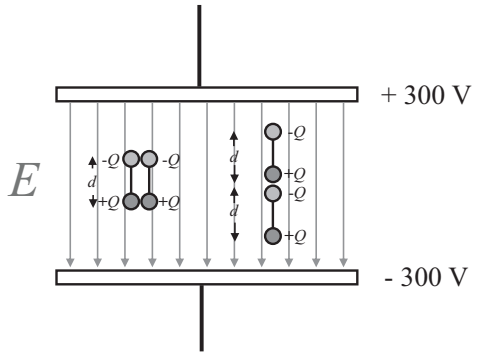


Figure 2.9: Two small particles each have the dipolemoment $\mu_0 = Q \cdot d$. Putting the two particles side by side gives the dipole moment $\mu = (2Q) \cdot d$ whereas putting them end to end gives $\mu = Q \cdot (2d)$, remembering that the charges +Q og -Q i the middle cancel out. The result is in both cases that $\mu = 2\mu_1$. The argument can easily be generalised to n particles.

As $n = M/M_1$ where M is the molecular weight of the compound particle and M_1 molecular weight of the small particle we get that $\alpha = (\alpha_1/M_1) \cdot M$. It therefore follows from equation 2.12 that the intensity of the light scattered from a single molecule is proportional to *the the square of the molecular weight*. For molecules that are large compared with the wavelength of the light this argument does no longer hold as large molecules in general don't have a well-defined state of polarisation. This situation we will now begin to look at but will only finish the discussion in section 3.3.

2.6 Scattering from one small, composite object

We shall here see that the way light is scattered from a particle that is not infinitely small is determined by the interference between scattered electromagnetic waves originating from different points of the composite particle. For those familiar with the use of complex numbers the calculations can be done quite simply and elegantly. But first we shall see how this interference comes about without the use of complex number. Then, in section 2.6, we shall see how things can be worked out more simply with the use of complex numbers.

Without the use of complex numbers

This section can be skipped if one prefers the description based on the use of complex numbers.

The terms ”particle” and ”molecule” are as usual interchangeable. What is crucial for the calculations and derivations that follow is that we can think of a particle as being composed of smaller ”point-like” particles all having the same polarisability, α.

When it comes to the description of the inner workings of light scattering it turns out that the scattering angle θ (cf. figures 1.2 and 2.2) is an inconvenient quantity to work with. Instead one uses a less intuitive quantity, the previously mentioned *scattering vector*, $\vec{q}$. The definition of the scattering vector is based on the wavevector $\vec{k}$ which, as you may recall is a vector pointing in direction of propagation of the light (the electromagnetic wave). The length of the wavevector is, as previously mentioned, $|\vec{k}| = 2\pi/\lambda = 2\pi n/\lambda_0$ (see section 2.5) and is called the wavenumber. When a photon with the wavevector $\vec{k}_{in}$ hits a particle and emerges from the collision with a different direction (i.e. it is scattered) its wavevector changes accordingly into $\vec{k}_{out}$. The situation is depicted in figure 2.10.

The *change* in the wavevector is called the scattering vector:

$$\vec{q} = \vec{k}_{out} - \vec{k}_{in} \tag{2.13}$$

The situation is illustrated in figure 2.10.

The length q of the scattering vector is easily calculated from figure 2.11. It is seen that $q = 2a$ where $a = k \cdot \sin\frac{\theta}{2}$ and the length of the wavevector is $k = \frac{2\pi}{\lambda}$. Hence we get the length of the scattering vector expressed as $q = \frac{4\pi}{\lambda}\sin\frac{\theta}{2}$. Here, as usual, λ is the wavelength of the light *in the medium surrounding the particle*. However, it is customary to express the length of the scattering vector using the wavelength of the light in vacuum λ_0 (in practise the same as in air). The connection is that $\lambda = \lambda_0/n$ where n is the refractive index of the medium. You may remember that the refractive index of a material expresses the factor by which the velocity of light in the material is reduced relative to the velocity of light in vacuum:

$$n = \frac{c_{\text{vacuum}}}{c_{\text{medium}}} \tag{2.14}$$

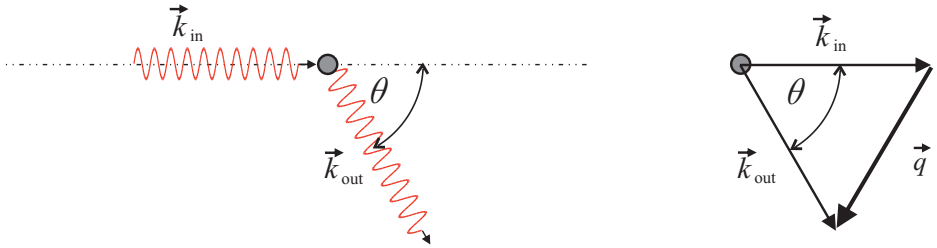


Figure 2.10: The figure on the left shows that the scattering of light on the particle changes the wavevector from $\vec{k}_{in}$ to $\vec{k}_{out}$. We assume that only the direction and not the wavelength changes as a result of the collision. Hence the length of the wavevector remains unchanged. The figure on the right shows the same situation in abstract vector form.

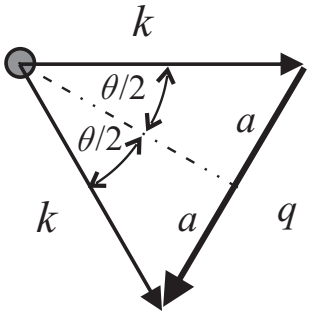


Figure 2.11: The figure shows the lengths of the vectors.It is seen that the length q of the scattering vector is $q = 2a$ where $a = k \cdot \sin\frac{\theta}{2}$. As the scattered light has the same wavelength as the incident light we know that $|\vec{k}_{out}| = |\vec{k}_{in}| \equiv k$, where the length of the wavevector is $k = \frac{2\pi}{\lambda}$. Hence the length of the scattering vector is $q = \frac{4\pi}{\lambda}\sin\frac{\theta}{2}$. Note, that λ is the wavelength of the light *in the medium embedding the particle*. Introducing the vacuum wavelength of the light λ_0 and the refractive index n of the medium the length of the scattering vector can be written $q = \frac{4\pi n}{\lambda_0}\sin\frac{\theta}{2}$

subunit

Finally we can write the length q of the scattering vector in its most common form:

$$q = \frac{4\pi n}{\lambda_0} \sin\frac{\theta}{2} \tag{2.15}$$

where θ is the scattering angle (cf. figure 1.2)

Now, let us look at two small particles being hit by the same electromagnetic wave (figure 2.12). At the position of the detector the scattered electric fields emerging

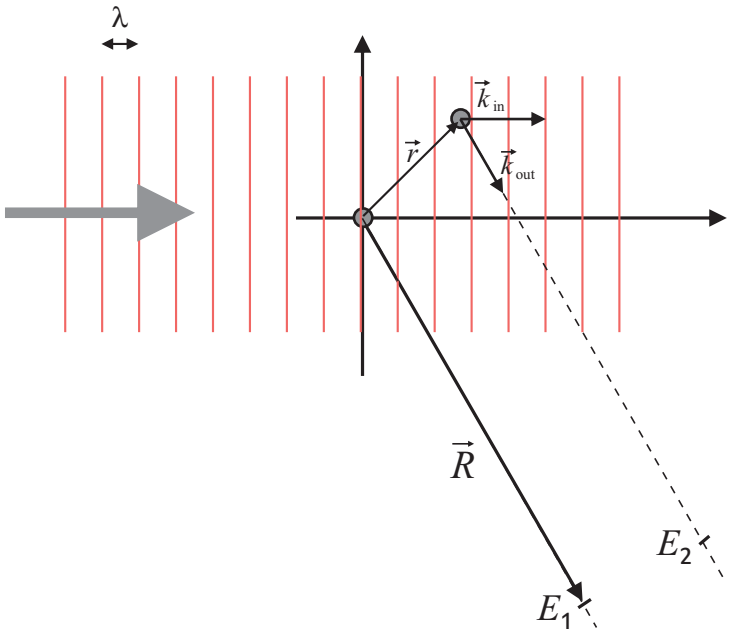


Figure 2.12: Monochromatic light with the wavelength λ and the wavevector $\vec{k}_{in}$ impinges on two small particles separated by the vector $\vec{r}$. The detector that detects the scattered light is positioned at the distance R from the origo of the coordinate system. The distance R is *very much larger* than the separation $|\vec{r}|$ between the two particles. The light scattered from the two particles has the electric field strength contributions E_1 and E_2, respectively. These field strengths have the same amplitude because the distance from the each particle to the detector is essentially the same. However, it may happen that the two fields do not have the same phase meaning that the don't attain their maximum value at the same time.

from the two particles in figure 2.12 can be described as plane waves with the same amplitude because the detector is at a very large distance from the particles.

In fact the scattered field emerging from each of the particles is a spherical wave but at a sufficiently large distance from its centre and within a relatively small area such a wave can be regarded as being plane:

$$\begin{aligned} E_1 &= E_0 \cos(\omega t - \vec{R} \cdot \vec{k}_{out}) \\ E_2 &= E_0 \cos(\omega t - \vec{R} \cdot \vec{k}_{out} + \Delta\phi) \end{aligned} \tag{2.16}$$

where the phase difference$\Delta\phi$ (not to be confused with the formerly used angle ϕ between the direction of the scattered light and the $z-$axis of the coordinate system) appears because the light scattered from the two particles travels different distances in order to reach the detector. But not only does the light travel different distances to reach the detector, the impinging light waves have also traveled different distances before they hit the particles. The total path length difference Δs can easily be shown to be $\Delta s = \vec{r} \cdot \vec{q}/k$, where k is the wavenumber and $\vec{r}$ is a vector from one of the particles to the other. This difference in pathlength of the light gives rise to the phase difference $\Delta\phi$ between waves received by the detector from the two particles:

$$\Delta\phi = \vec{r} \cdot \vec{q} \tag{2.17}$$

The intensity on the detector I_{total} of the total field $E = E_1 + E_2$ can the be calculated using equations 2.16 and 2.6:

$$\begin{aligned} I_{\text{total}} &= \epsilon_0 c \langle E^2 \rangle \\ &= \langle \Big(E_0 \cos(\omega t - \vec{R} \cdot \vec{k}_{out}) + E_0 \cos(\omega t - \vec{R} \cdot \vec{k}_{out} + \Delta\phi) \Big)^2 \rangle \\ &= \epsilon_0 c E_0^2 \langle \cos^2(\omega t - \vec{R} \cdot \vec{k}_{out}) + \cos^2(\omega t - \vec{R} \cdot \vec{k}_{out} + \Delta\phi) + \\ &\quad 2\cos(\omega t - \vec{R} \cdot \vec{k}_{out}) \cdot \cos(\omega t - \vec{R} \cdot \vec{k}_{out} + \Delta\phi) \rangle \end{aligned} \tag{2.18}$$

The two first terms in the final expression of equation 2.18 (the terms containing $\cos^2$) both have the mean value $\frac{1}{2}$ as explained earlier (page 16). The last term in equation 2.18 can be calculated using a well known trigonometric identity:

$$\cos x \cdot \cos y = \frac{1}{2} \cdot [\cos(x+y) + \cos(x-y)]$$

yielding:

$$\begin{aligned} I_{\text{total}} &= 2\cos(\omega t - \vec{R}\cdot\vec{k}_{out})\cdot\cos(\omega t - \vec{R}\cdot\vec{k}_{out} + \Delta\phi) \\ &= \cos(2\omega t - 2\vec{R}\cdot\vec{k}_{out} + \Delta\phi) + \cos(\Delta\phi) \end{aligned} \tag{2.19}$$

The mean value of the first term in the bottom line of equation 2.19 is 0. The reason for this is, again, that the mean value of $\cos(2\omega t - 2\vec{R}\cdot\vec{k}_{out} + \Delta\phi)$ calculated over one cycle equals 0 and $\omega \approx 3\cdot 10^{15}\,\text{radians}\cdot\text{s}^{-1}$ for a typical HeNe laser. So averaging over even an extremely short time interval like e.g. one nanosecond would mean averaging over approximately half a million cycles. This means that even if this short averaging time does not contain a whole number of cycles the result would still be extremely close to 0. The last term is constant and therefore simply has the mean value $\cos(\Delta\phi)$
Summing up we get:

$$I_{total} = \epsilon_0 c E_0^2\cdot(1+\cos(\Delta\phi)) = 2\cdot I_{s,\,1}\cdot(1+\cos(\Delta\phi)) \tag{2.20}$$

Equation 2.20 shows that the intensity of the light scattered from two particles can be anything from 0 to 4 times as high as the intensity scattered from one single particle. This result we are now going to generalise to an arbitrary number of particles:
Instead of assigning the field E_2 the extra phase $\Delta\phi$ in equation 2.19 one could just as well have assigned E_1 the phase ϕ_1 and E_2 the phase ϕ_2 such that $\phi_2 - \phi_1 = \Delta\phi$. We want to generalise equation 2.20 so, instead of looking at two small particles we shall consider a "molecule" (a compound particle) consisting of n subunits (e.g. amino acids or monosaccharides). For simplicity we shall assume that all subunits are identical. We number these subunits with an index i, such that $i = 1, 2, \ldots, n$, (see figure 2.13) At the detector position the total scattered electric field $E_{s,\,n}$ (from the n subunits) is then simply the sum of the contributions:

$$E_{s,\,n} = \sum_{i=1}^{n} E_0\cos(\omega t + \phi_i) \tag{2.21}$$

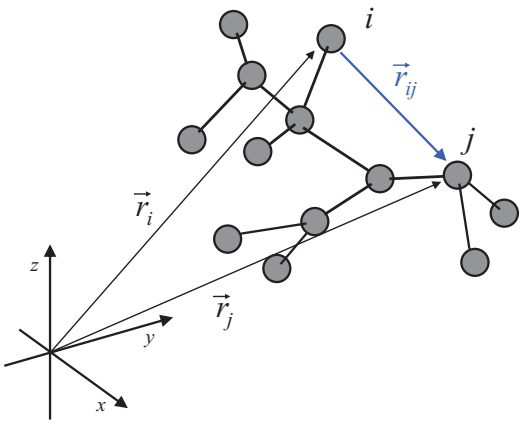


Figure 2.13: The molecule consists of n identical subunits and light is impinging on it as shown in figure ref2particle. At the detector position the field contributions from the different subunits all have the same amplitude E_0 but different phases ϕ_1, ϕ_2,..., ϕ_n. The positions of the individual subunits are given by the position vectors $\vec{r}_1$,$\vec{r}_2$ ...,$\vec{r}_n$

Using again equation 2.6 the total scattered *intensity* $I_{s,\,n}$ is then given by:

$$\begin{aligned} I_{s,\,n} &= \epsilon_0 c \langle E_{s,\,n}^2 \rangle \\ &= \epsilon_0 c \langle \left(\sum_{i=1}^{n} E_0 \cos(\omega t + \phi_i) \right) \cdot \left(\sum_{j=1}^{n} E_0 \cos(\omega t + \phi_j) \right) \rangle \\ &= \epsilon_0 c E_0^2 \langle \sum_{i=1}^{n} \sum_{j=1}^{n} \cos(\omega t + \phi_i) \cdot \cos(\omega t + \phi_j) \rangle \\ &= \epsilon_0 c E_0^2 \langle \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{1}{2} \left[\cos(2\omega t + \phi_i + \phi_j) + \cos(\phi_i - \phi_j) \right] \rangle \end{aligned} \tag{2.22}$$

Here it is seen that the terms $\cos(2\omega t + \phi_i + \phi_j)$ have a mean value of 0 (see top of page 29 for expalnation). The remaining terms are constant if the phases $\phi_1, \phi_2, \ldots$ are time independent, which we shall assume for now. This just means that the molecule lies still and doesn't rotate. Hence we get:

$$I_{s,\,n} = I_{s,\,1} \sum_{j=1}^{n} \sum_{k=1}^{n} \cos(\phi_j - \phi_k) \tag{2.23}$$

where, as usual, $I_{s,\,1} = \epsilon_0 c E_0^2$ denotes the intensity of the light scattered from one single subunit of the molecule. If the overall size, d, of the molecule is very much smaller than the wavelength of the light meaning that $d \cdot k = d \cdot 2\pi/\lambda \approx 0$,

the the pase differences are $|\phi_j - \phi_k| = |\vec{r}_j \cdot \vec{q} - \vec{r}_k \cdot \vec{q}| = |\vec{r}_{jk} \cdot \vec{q}| \leq d \cdot 2k \approx 0$. Hence for all terms in equation 2.23 we get that $\cos(\phi_j - \phi_k) \approx \cos(0) = 1$. As this double sum contains a total of n^2 terms we get:

$$I_{s,\,n} \approx I_{s,\,1} \cdot n^2 \tag{2.24}$$

and, as $n = M/M_1$ where M is the molecular mass and M_1 is the molecular mass of a subunit, we get:

$$I_{s,\,n} \approx (I_{s,\,1}/M_1^2) \cdot M^2 \tag{2.25}$$

This is basically the same result that was obtained in section 2.5, namely that the "scattering power" of a single small molecule is proportional to *the square of the molecular mass*. It is important to note that this result is valid when the molecule is so small that the phase differences between light scattered from different parts of the molecule are essentially 0. How to relax this restriction will be treated in section 3.3 dealing with scattering from large molecules.

With the use of complex numbers

This section is an alternative to section 2.6. The description of scattering is based on the complex number description of sinusoidal waves. The use of complex numbers leads to a strong simplification of the calculations leading eventually to the same equations as we saw in section 2.6. Appendix A describes very briefly how complex numbers are used. The present section has more or less the same wording as section 2.6 and can thus be read independently from the previous section.

When it comes to the description of the inner workings of light scattering it turns out that the scattering angle θ (cf. figures 1.2 and 2.2) is an inconvenient quantity to work with. Instead one uses a less intuitive quantity, the previously mentioned *scattering vector*, $\vec{q}$. The definition of the scattering vector is based on the wavevector $\vec{k}$ which, as you may recall is a vector pointing in direction of propagation of the light (the electromagnetic wave). The length of the wavevector is, as previously mentioned, $|\vec{k}| = 2\pi/\lambda = 2\pi n/\lambda_0$ (see section 2.5) and is called the wavenumber. When a photon with the wavevector $\vec{k}_{in}$ hits a particle and emerges

from the collision with a different direction (i.e. it is scattered) its wavevector changes accordingly into $\vec{k}_{out}$. The situation is depicted in figure 2.10.
The *change* in the wavevector is called the scattering vector:

$$\vec{q} = \vec{k}_{out} - \vec{k}_{in} \tag{2.26}$$

The situation is illustrated in figure 2.10.
The length q of the scattering vector is easily calculated from figure 2.11. It is seen that $q = 2a$ where $a = k \cdot \sin\frac{\theta}{2}$ and the length of the wavevector is $k = \frac{2\pi}{\lambda}$. Hence we get the length of the scattering vector expressed as $q = \frac{4\pi}{\lambda}\sin\frac{\theta}{2}$. Here, as usual, λ is the wavelength of the light *in the medium surrounding the particle*. However, it is customary to express the length of the scattering vector using the wavelength of the light in vacuum λ_0 (in practise the same as in air). The connection is that $\lambda = \lambda_0/n$ where n is the refractive index of the medium. You may remember that the refractive index of a material expresses the factor by which the velocity of light in the material is reduced relative to the velocity of light in vacuum:

$$n = \frac{c_{\text{vacuum}}}{c_{\text{medium}}} \tag{2.27}$$

Finally we can write the length q of the scattering vector in its most common form:

$$q = \frac{4\pi n}{\lambda_0}\sin\frac{\theta}{2} \tag{2.28}$$

where θ is the scattering angle (cf. figure 1.2)
Now, let us look at two small particles being hit by the same electromagnetic wave (figure 2.12). At the position of the detector the scattered electric fields emerging from the two particles in figure 2.12 can be described as plane waves with the same amplitude because the detector is at a very large distance from the particles. In fact the scattered field emerging from each of the particles is a spherical wave but at a sufficiently large distance from its centre and within a relatively small area such a wave can be regarded as being plane:

$$\begin{aligned} E_1 &= E_0 e^{i(\omega t - \vec{R}\cdot\vec{k}_{ud})} \\ E_2 &= E_0 e^{i(\omega t - \vec{R}\cdot\vec{k}_{ud} + \Delta\phi)} \end{aligned} \tag{2.29}$$

where the phase difference$\Delta\phi$ (not to be confused with the formerly used angle ϕ between the direction of the scattered light and the $z-$axis of the coordinate system) appears because the light scattered from the two particles travels different distances in order to reach the detector. But not only does the light travel different distances to reach the detector, the impinging light waves have also traveled different distances before they hit the particles. The total path length difference Δs can easily be shown to be $\Delta s = \vec{r} \cdot \vec{q}/k$, where k is the wavenumber and $\vec{r}$ is a vector from one of the particles to the other. This difference in path length of the light gives rise to the phase difference $\Delta\phi$ between waves received by the detector from the two particles:

$$\Delta\phi = \vec{r} \cdot \vec{q} \tag{2.30}$$

The intensity on the detector I_{total} of the total field $E = E_1 + E_2$ can the be calculated using equation 2.6 plus the fact that $|E|^2 = E \cdot E^*$, where E^* denotes the complex conjugate of E:

$$\begin{aligned} I_{total} &= \frac{1}{2}\epsilon_0 c|E|^2 \\ &= \frac{1}{2}\epsilon_0 cE \cdot E^* \\ &= (E_0 e^{i(\omega t - \vec{R}\cdot\vec{k}_{ud})} + E_0 e^{i(\omega t - \vec{R}\cdot\vec{k}_{ud} + \Delta\phi)}) \cdot (E_0 e^{-i(\omega t - \vec{R}\cdot\vec{k}_{ud})} + E_0 e^{-i(\omega t - \vec{R}\cdot\vec{k}_{ud} + \Delta\phi)}) \\ &= \frac{1}{2}\epsilon_0 cE_0^2(1 + 1 + e^{i\Delta\phi} + e^{-i\Delta\phi}) \\ &= \frac{1}{2}\epsilon_0 cE_0^2(2 + 2\cos(\phi)) \end{aligned} \tag{2.31}$$

Equation 2.31 can also be written:

$$I_{total} = \epsilon_0 cE_0^2 \cdot (1 + \cos\phi) = 2 \cdot I_{s,\,1} \cdot (1 + \cos(\Delta\phi)) \tag{2.32}$$

which is the same result obtained in the previous section, equation 2.20.
Equation 2.32 shows that the intensity of the light scattered from two particles can be anything from 0 to 4 times as high as the intensity scattered from one single particle. This result we are now going to generalise to an arbitrary number of particles:
Instead of assigning the field E_2 the extra phase $\Delta\phi$ in equation 2.19 one could just as well have assigned E_1 the phase ϕ_1 and E_2 the phase ϕ_2 such that $\phi_2 - \phi_1 = \Delta\phi$.

We want to generalise equation 2.20 so, instead of looking at two small particles we shall consider a "molecule" (a compound particle) consisting of n subunits (e.g. amino acids or monosaccharides). For simplicity we shall assume that all subunits are identical. We number these subunits with an index i, such that $i = 1, 2, \ldots, n$, (see figure 2.13). At the detector position the total scattered electric field $E_{s,\,n}$ (from the n subunits) is then simply the sum of the contributions:

$$E_{s,\,n} = \sum_{j=1}^{n} E_0 e^{i(\omega t + \phi_j)} \tag{2.33}$$

Using again equation 2.6 the total scattered *intensity* $I_{s,\,n}$ is then given by:

$$\begin{aligned}
I_{s,\,n} &= \frac{1}{2}\epsilon_0 c E_{s,\,n} E^*_{s,\,n} \\
&= \frac{1}{2}\epsilon_0 c \left(\sum_{j=1}^{n} E_0 e^{i(\omega t+\phi_j)}\right)\left(\sum_{k=1}^{n} E_0 e^{-i(\omega t+\phi_k)}\right) \\
&= \frac{1}{2}\epsilon_0 c E_0^2 \sum_{j=1}^{n}\sum_{k=1}^{n} e^{i(\phi_j-\phi_k)} \\
&= I_{s,\,1} \sum_{j=1}^{n}\sum_{k=1}^{n} e^{i(\phi_j-\phi_k)} \\
&= I_{s,\,1} \sum_{j=1}^{n}\sum_{k=1}^{n} (\cos(\phi_j-\phi_k) + i\sin(\phi_j-\phi_k)) \\
&= I_{s,\,1} \sum_{j=1}^{n}\sum_{k=1}^{n} \cos(\phi_j-\phi_k)
\end{aligned} \tag{2.34}$$

where $I_{s,\,1}$ denotes the (mean)intensity of the light scattered from one subunit of the molecule and where to obtain the last identity it is used that $\sin(\phi_j - \phi_k) = -\sin(\phi_k - \phi_j)$, meaning that the sine terms cancel pairwise. If the overall size, d, of the molecule is very much smaller than the wavelength of the light meaning that $d \cdot k = d \cdot 2\pi/\lambda \approx 0$, the the pase differences are $|\phi_j - \phi_k| = |\vec{r}_j \cdot \vec{q} - \vec{r}_k \cdot \vec{q}| = |\vec{r}_{jk} \cdot \vec{q}| \leq d \cdot 2k \approx 0$. Hence for all terms in equation 2.23 we get that $\cos(\phi_j - \phi_k) \approx \cos(0) = 1$. As this double sum contains a total of n^2 terms we get:

$$I_{s,\,n} \approx I_{s,\,1} \cdot n^2 \tag{2.35}$$

and, as $n = M/M_1$ where M is the molecular mass and M_1 is the molecular mass

of a subunit, we get:

$$I_{s,\,n} \approx (I_{s,\,1}/M_1^2) \cdot M^2 \tag{2.36}$$

This is basically the same result that was obtained in section 2.5, namely that the "scattering power" of a single small molecule is proportional to *the square of the molecular mass*. It is important to note that this result is valid when the molecule is so small that the phase differences between light scattered from different parts of the molecule are essentially 0. How to relax this restriction will be treated in section 3.3 dealing with scattering from large molecules.

2.7 Scattering from many identical particles

E shall now consider the situation where light is scattered from a number N of molecules that are present in a volume V which is evenly lit by a laser beam and which are at the same time visible to the detector (This is the previously mentioned scattering volume). Furthermore we assume that the molecules are all identical as discussed above (see figure 2.14). We can now reuse equation 2.23, where we now think of the individual scatterers as whole molecules whose individual scattering intensity is given by equation 2.25. The total scattering intensity $I_{s,\,total}$ from the N molecules can the be written as:

$$I_{s,\,total} = I_{s,\,n} \sum_{j=1}^{N} \sum_{k=1}^{N} \cos(\phi_j - \phi_k) \tag{2.37}$$

As the thing of interest in static light scattering (SLS) is the mean scattered intensity we shall calculate this:

$$\langle I_{s,\,total} \rangle = I_{s,\,n} \sum_{j=1}^{N} \sum_{k=1}^{N} \langle \cos(\phi_j - \phi_k) \rangle \tag{2.38}$$

For now we shall assume that the molecules move about independently of each other. This means that the phases ϕ_j and ϕ_k will change independently and randomly all the time. Consequently the phase difference $\phi_j - \phi_k$ can assume any value as long as $j \neq k$. Hence $\langle \cos(\phi_j - \phi_k) \rangle = 0$ if $j \neq k$ whereas the N led where $j = k$ have a mean value of $\langle \cos(\phi_j - \phi_k) \rangle = \langle \cos(0) \rangle = 1$. Therefore

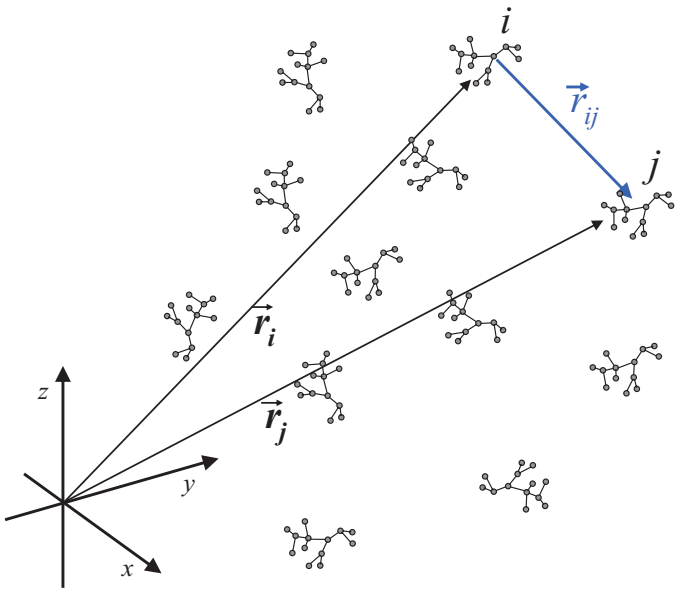


Figure 2.14: N identical molecules are in an evenly lit volume of a size so that the light scattered from all the molecules can reach the detector.

equation 2.38 can be written as:

$$\langle I_{s,\,total} \rangle = I_{s,\,n} \cdot N \tag{2.39}$$

As the number, N, of particles in the volume V is given by the molar concentration c of the molecules through the expression

$$N = V \cdot c$$

we can rewrite equation 2.39 with the aid of equation 2.25:

$$\begin{aligned} \langle I_{s,\,total} \rangle &= I_{s,\,n} \cdot N_A V c \\ &= (I_{s,\,1}/M_1^2) \cdot N_A V c \cdot M^2 \\ &= (I_{s,\,1} N_A V/M_1^2) \cdot c \cdot M^2 \\ &= (I_{s,\,1} N_A V/M_1^2) \cdot C \cdot M \end{aligned} \tag{2.40}$$

where $N_A = 6.022 \cdot 10^{23}\,\mathrm{mol}^{-1}$ er Avogadro's number and C is the weight concentration (unit $\mathrm{g} \cdot \mathrm{L}^{-1}$).

As the scattering intensity of one molecular subunit, $I_{s,\,1}$, is proportional to the intensity of the incident laser beam and inversely proportional to the square of the distance r from the scatterers to the detector we can write $I_{s,\,1}N_A/M_1^2 = K \cdot I_0/r^2$ where K is the constant of proportionality. With this we can now rewrite equation 2.40 as:

$$I_{s,\,total} = KI_0VCM/r^2 \tag{2.41}$$

or, with the definition 2.2 of the Rayleigh ratio, $R(q)$:

$$R(q) = KCM \tag{2.42}$$

Note that we have written the Rayleigh ratio as a function of q although it is in fact constant in this case (i.e. when the scattering particles are very small). The reason is of course that when the scattering particles are sufficiently large the Rayleigh ratio does in fact depend measurably on q, i.e. on the scattering angle. The constant K is called *the optical contrast constant* and can easily be shown[6] to be related to the refractive index of the solvent and the dissolved particles through the expression:

$$K = \frac{4\pi^2 n_0^2 (dn/dC)^2}{N_A \lambda_0^4} \tag{2.43}$$

where n_0 is the refractive index of the solvent, λ_0 is the wavelength of the laser beam in vacuum and dn/dC er the derivative of the refractive index of *the solution* with respect to the weight concentration of the dissolved molecules. In some older literature you may come across a slightly different definition of the optical contrast constant with a factor 2 instead of 4. For an explanation see section 2.8. The quantity dn/dC is called *the differential refractive index increment*. For most proteins in water based buffers it has a value of 0.18 – 0.20 $\mathrm{ml\ g^{-1}}$ for most polysaccharides it has a value of about 0,14 – 0.15 $\mathrm{ml\ g^{-1}}$.

In equation 2.42 the Rayleigh ratio $R(q)$ is a measured quantity. Also, the weight concentration C is also a quantity under the control of the experimenter. Hereby it seems quite straightforward to determine the molecular weight M because the constant K can be determined using the defining equation 2.43. It just takes doing

[6] See e.g. the derivation in *Physical Biochemistry*, second edition, K. E. van Holde, Prentice Hall 1985

an independent determination of $\frac{dn}{dC}$. Alternatively one can use either a tabulated value for $\frac{dn}{dC}$ or simply assume that the value is $0.18\,\mathrm{cm}^3 \cdot \mathrm{g}^{-1}$ for proteins and $0.145\,\mathrm{cm}^3 \cdot \mathrm{g}^{-1}$ for polysaccharides. This will in most cases give a molecular weight which is at most 10% off if the measurements are done otherwise correctly.

2.8 Comparisons with older literature

If you compare the equations in this set of notes with the equations found in other literature, especially older literature, you may find some strange differences. They are due to two things:

1. We have assumed the incoming light to vertically polarised. But originally light scattering experiments were done with unpolarised light, which changes both the mathematical form of the basic equations and the definition of both the optical contrast constant K and of the Rayleigh ratio $R(\theta)$.

2. The unit system used in older literature was different. We have used basically the SI-system whereas the cgs-system (also called the Gaussian unit system) was the system of choice in science. This affects the looks of even fundamental equations of physics, like the equation for the force between two charged particles. The consequence of this is that even with polarised light the old and the new equations would look different. In particular the vacuum permittivity ϵ_0 may show up in one set of equations but not in the other. Also, a factor of 4π sometimes raised to the power of two will be present in one set of equations but not in the other.

Scattering from one small particle is described by equation 2.11 which reads

$$I_{s,1}(r) = I_0 \cdot \frac{\pi^2 \alpha^2 \sin^2 \phi}{r^2 \lambda^4}$$

Note that scattering is dependent only on the angle ϕ with the z-axis, not on the angle θ with the x-axis. This will look as indicated in figure 2.15 If the incoming light is no longer polarised the z-axis looses its special meaning in relation to the

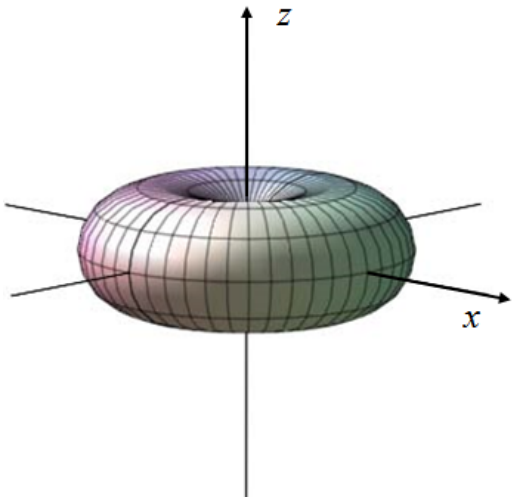


Figure 2.15: The incoming light is moving along the x-axis and is polarised in the direction of the z-axis. The surface indicates the intensity of the scattered light in different directions. Note that nothing is scattered in the direction of the z-axis, up or down

scattering. Averaging over all polarisations in the yz-plane gives the following expression for the scattered intensity:

$$I_{s,1}(r) = I_0 \cdot \frac{\pi^2\alpha^2}{r^2\lambda^4} \cdot \frac{1+\cos^2\theta}{2} \tag{2.44}$$

which is now independent of ϕ but depends on the angle θ with the x-axis. What it will look like is shown in figure 2.16. As the scattered intensity in the case of unpolarised incoming light shows an angular dependency which has nothing to do with the form factor of the scattering particle the Rayleigh ratio is in this case defined so as to factor out this dependency:
With polarised light it was defined in equation 2.1 as

$$R_\theta = \frac{I_s(\theta)\cdot r^2}{I_0 \cdot V_s(\theta)}$$

With unpolarised light the definition of the Rayleigh ratio is

$$R_\theta = \frac{I_s(\theta)\cdot r^2}{I_0 \cdot V_s(\theta)} \cdot \frac{1}{1+\cos^2\theta} \tag{2.45}$$

thus making the Rayleigh ratio of the scattering from an infinitely small particle independent of the scattering angle. This alternate definition of the Rayleigh ratio

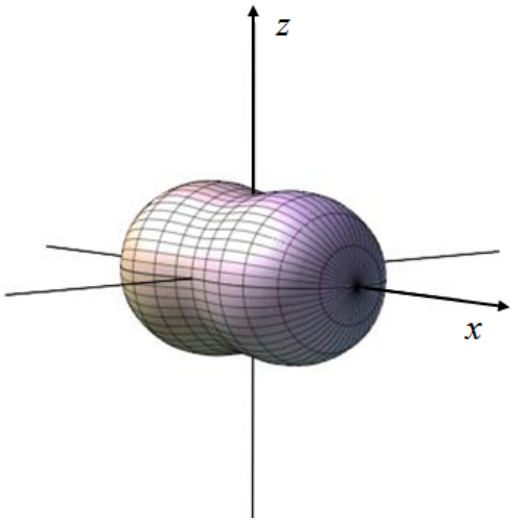


Figure 2.16: The incoming light is moving along the x-axis and is unpolarised. The surface indicates the intensity of the scattered light in different directions. Note, that now light is scattered in the direction of the z-axis, both up and down. And the scattered intensity is now dependent on the angle θ with the x-axis.

has a small impact on the definition of the optical contrast constant K which with polarised light was defined in equation 2.43 as

$$K = \frac{4\pi^2 n_0^2 (dn/dC)^2}{N_A \lambda_0^4}$$

With unpolarised light, due to the alternate definition (2.45) of the Rayleigh ratio the optical contrast constant is now defined as

$$K = \frac{2\pi^2 n_0^2 (dn/dC)^2}{N_A \lambda_0^4} \tag{2.46}$$

differing from the old definition only by a factor of 2.
With these (old) definitions of the Rayleigh ratio and of the optical contrast constant the higher level equations of light scattering (like equations 2.42, 3.2, 3.13, etc.) are unchanged.

2.9 Summing up

Here are the most important equations from this chapter:

Refractive index:

$$n = \frac{c_{\text{vacuum}}}{c_{\text{medium}}}$$

The optical contrast constant:

$$K = \frac{4\pi^2 n_0^2 (dn/dC)^2}{N_A \lambda_0^4}$$

The scattering vector:

$$\vec{q} = \vec{k}_{out} - \vec{k}_{in}$$

The length of the scattering vector:

$$q = |\vec{q}| = \frac{4\pi n}{\lambda_0} \cdot \sin\left(\frac{\theta}{2}\right)$$

The Rayleigh ratio:

$$R(q) = \frac{I_s(q) \cdot r^2}{I_0 \cdot V_s(q)}$$

3

Determination of size and molecular weight

Probably the most common reason to do light scattering measurements is that it can be used to determine molecular weights. This is what equation 2.42 shows. If two conditions are fulfilled

1. the molecules (or particles) under investigation are small enough (as a rule of thumb, smaller than $\lambda/20$, one twentieth of the wavelength of the light), and

2. their concentration is sufficiently low (difficult to set up a criterion here)

molecular weight determination is unproblematic, at least in principle because equation 2.42 actually holds in this case. However, if these two criteria are not

met equation 2.42 is often an oversimplification. We shall deal with these two possible complicating factors in turn:

3.1 Concentration effects for small molecules

The simple relationship (equation 2.39 or 2.42) that the scattered intensity being proportional to the weight concentration of the dissolved molecules turns out to be increasingly inaccurate as the concentration increases. The reason for this is obvious as equation 2.39 was based on the assumption that the molecules move *independently* of each other. This can of course not be completely true for the simple reason that two molecules cannot occupy the same portion of space. So, in this trivial sense the position of one molecule is dependent on the positions of other molecules in the solution. It is also evident that the chance of two molecules accidentally ”trying” to occupy the same portion of space is low at low concentrations and high at high concentrations. Therefore equation 2.39 is a good approximation at low concentration becoming increasingly inaccurate as concentration increases and less and less room becomes available to the individual molecules. More generally the mutual influence of the molecules positions is due to forces (repulsive or attractive or either, depending on intermolecular distances) acting between the molecules making some distances more probable than others. The aforementioned trivial effect of crowding is called ”the excluded volume effect” and can be seen as result of strong repulsive forces acting between particles at very short distances (less than their size) thus preventing the particles from penetrating into each other. The double sum in equation 2.38 or 3.1 divided by the number N of molecules in the scattering volume V_s is called *the static structure factor* and is denoted $S(q)$:

$$S(q) = N^{-1} \cdot \sum_{j=1}^{N} \sum_{k=1}^{N} \langle \cos(\phi_j - \phi_k) \rangle \tag{3.1}$$

The assumption that the molecules move independent of each other leads as shown to the simple relationship that the double sum of equation 2.38 or 3.1 assumes the value N, or in other words that $S(q) = 1$. If we don't assume independence between molecular positions or movements we can then generalise equation 2.42

relating the Rayleigh ratio $R(q)$ and the concentration and molecular weight of the molecules:

$$R(q) = KCMS(q) \tag{3.2}$$

The generalization is easy, the problem of course being how to calculate $S(q)$.
Before we proceed we rewrite equation 3.2 for comparison with equation 3.6. Rearranging gives:

$$\frac{KC}{R(q)} = \frac{1}{MS(q)} \tag{3.3}$$

For small (isotropically scattering) particles it can be shown (see e.g. *Physical Chemistry of Macromolecules*, C. Tanford, OUT-OF-PRINT-BOOKS-ON-DEMAND 1992) that the Rayleigh ratio is related to the osmotic pressure Π of the dissolved molecules through the relationship:

$$\frac{KC}{R(q)} = \frac{1}{RT}\left(\frac{\partial \Pi}{\partial C}\right)_{T,P} \tag{3.4}$$

where $R = 8.31\,\mathrm{J}\cdot\mathrm{K}^{-1}\cdot\mathrm{mol}^{-1}$ is the universal gas constant, not to be confused with the Rayleigh ratio $R(q)$. The osmotic pressure has a concentration dependence, normally written as a series expansion:

$$\frac{\Pi}{RT} = \frac{1}{M}C + A_2C^2 + A_3C^3 + \cdots \tag{3.5}$$

where the constants A_2, $A_3, \ldots$ are called *the second, the third, ... virial coefficient*,respectively. For a so-called *ideal* solution the virial coefficients are all 0. With this (i.e. equation 3.5) we can write equation 3.4 as:

$$\frac{KC}{R(q)} = \frac{1}{M} + 2A_2C + 3A_3C^2 + \cdots \tag{3.6}$$

For practical purposes it is customary only to include the second virial coefficient. This quantity can sometimes be calculated theoretically but is usually one of the things that is determined experimentally, i.e. by doing light scattering measurements.
Now we seem to have two different ways to express non-ideality of the solution, namely equation 3.3 and equation 3.6. But they just show that in the case of very

small particles the static structure factor can be written as:

$$S(q) = \frac{1}{1 + 2A_2MC + 3A_3MC^2 + \ldots} \tag{3.7}$$

which shows that the structure factor is indeed dependent on the concentration but independent of the scattering vector, q. The latter turns out to be a consequence of the assumption that the particles are sufficiently small to scatter isotropically.

3.2 Scattering from large molecules

Next we shall first disregard concentration effects but instead see how to deal with particles that are not so small that they scatter isotropically. This is also a necessary step towards developing a method to determine molecular weights. One bonus of solving this complicating circumstance is that with the solution comes a way to determine *the size*, i.e. the physical extension of the molecules.

3.3 The form factor

If a particle or molecule scatters light isotropically the reason is, as previously mentioned, that light contributions scattered from the constituent parts (subunits) of the molecule) are in phase, i.e. have no phase difference. This will be the case if the molecule is much smaller than the wavelength of the impinging light. If on the other hand the scattering particles are not much smaller than the wavelength of the light light will in general be scattered with an efficiency that decreases with increasing scattering angle θ and hence with increasing q-value. This at least hold for particles without pronounced symmetry like e.g. identical spheres and and identical cylinders. Such high symmetry particles will have scattering intensities exhibiting (local) maxima and minima but still with an overall tendency to scatter light less efficiently with increasing scattering angle. If for a solutions of particles with a definite shape and identical size we measure the intensity of the scattered light, $I_s(q)$, as a function of q then we will define a function of q called *the form factor*, $P(q)$, for these particles as the normalised scattering intensity:

$$P(q) = I_s(q)/I_s(0) \tag{3.8}$$

where $I_s(0)$, the scattering intensity at zero scattering angle (i.e. zero q value) normally has to be determined by extrapolation of $I_s(q)$ to $q = 0$, as this value is not accessible to direct measurement because a detector placed at zero scattering angle would see mainly the direct laser beam. Another way to interpret the form factor is to conceive it as the ratio between the scattering intensity of the particles and the scattering intensity of some hypothetical particles with the same mass (molecular weight) but being infinitely small. When the particles are very small the form factor is practically constant, $P(q) \approx 1$ even for the largest experimentally accessible q values.

Example: If the particles are dissolved in water (refractive index $n = 1.333$) and the light source is a HeNe laser with a vacuum wavelength of $\lambda_0 = 633\,\mathrm{nm}$ the maximum value of q corresponds to a scattering angle of $\theta_{\max} = 180°$. Hence $q_{max} = \frac{4\pi n}{\lambda_0} \cdot \sin\left(\frac{\theta_{\max}}{2}\right) = 0.026\ \mathrm{nm}^{-1}$. For larger particles one gets form factor curves that qualitatively look as shown in figure 3.1 It should be mentioned that sometimes $P(q)$ is called *the scattering function* . We can calculate the form factor using equation 2.23 and 2.24 remembering that $\phi_j - \phi_k = \vec{r}_j \cdot \vec{q} - \vec{r}_k \cdot \vec{q} = \vec{r}_{jk} \cdot \vec{q}$. As equation 2.23 holds in generel and equation 2.24 holds for infinitely small particles of the same mass we get the form factor by dividing the former equation with the latter subsequently averaging over all orientations[1] of the molecule:

$$P(q) = \langle \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} \cos(\phi_j - \phi_k) \rangle = \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} \langle \cos(\vec{r}_{jk} \cdot \vec{q}) \rangle \qquad (3.9)$$

We calculate the average value of the double sum by introducing a coordinate system with the z-axis in the direction of $\vec{q}$ and normal polar coordinates (i.e. with the angle θ relative to the z-axis and the angle ϕ in the xy-plane). Note that in this coordinate system we have $\vec{r}_{jk} \cdot \vec{q} = q\, r_{jk} \cdot \cos(\theta)$ so the mean value of one term in equation 3.9 is given by:

$$\langle \cos(\vec{r}_{jk} \cdot \vec{q}) \rangle = \frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} \cos(q\, r_{jk} \cdot \cos(\theta)) \cdot \sin(\theta)\, d\theta\, d\phi \qquad (3.10)$$

[1] As the form factor pertains to a solution of molecules with random orientations we can just do the averaging for one molecule

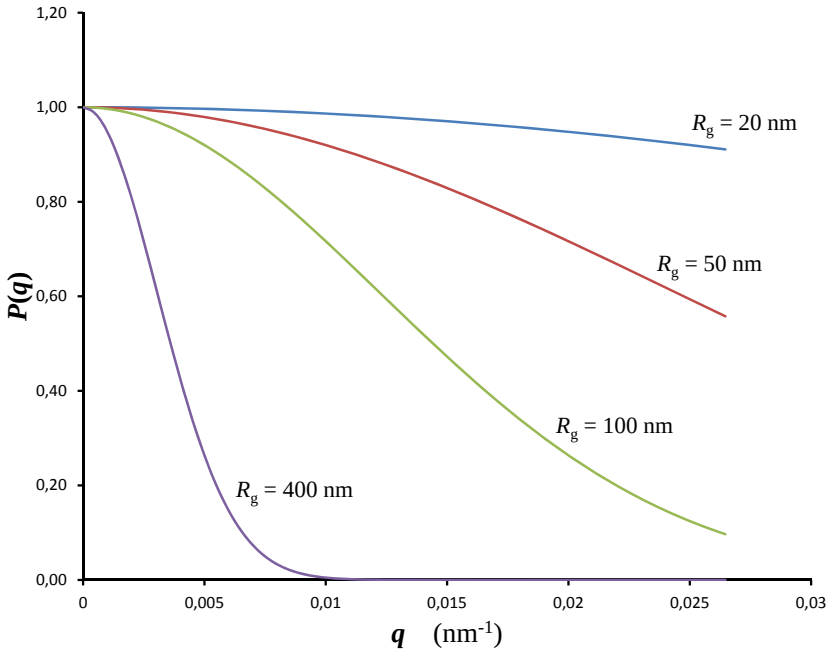


Figure 3.1: The form factor becomes more and more steeply decreasing the larger particles it represents. The particle sizes are indicated on the graphs by their radius of gyration). The wavelength of the light is assumed to be 633 nm in vacuum and the particles are suspended in water with a refractive index of $n = 1.333$. Note that the form factor always has the value 1 for $q = 0$

The calculation is easily performed using the substitution $x = q\, r_{jk}\ \cos(\theta)$, yielding:

$$\langle \cos(\vec{r}_{jk} \cdot \vec{q}) \rangle = \frac{\sin(q\, r_{jk})}{q\, r_{jk}} \tag{3.11}$$

Substituting this expression into equation 3.9 we get the general expression for the form factor of any molecule consisting of n identical point-like subunits with pair-wise distances r_{jk}:

$$P(q) = \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} \frac{\sin(q\, r_{jk})}{q\, r_{jk}} \tag{3.12}$$

This generel expression for the form factor can of course be calculated (on a computer) for a given molecule if the structure of the molecule is known. Sometimes it is useful to approximate the molecules with simple geometric bodies like spheres, ellipsoids, cylinders, boxes etc. and then calculate the form factor analytically, i.e. to arrive at a mathematical expression of the form factor based on known mathematical functions and relevant geometric parameters of the chosen geometrical bodies. Examples of analytical form factors can be found in section 3.7.

One peculiar feature of the form factor is that if one makes an approximation[2] of $P(q)$ to the lowest non-trivial order in q the approximated form factor turns out to be independent of the shape of the molecule. It only depends on some average size parameter, called *the radius of gyration*. The relevance of this we shall see in the next section.

With the introduction of the form factor it is not difficult to see that we can generalise equation 2.42 to:

$$R(q) = KCMP(q) \tag{3.13}$$

Sometimes one defines *the apparent molecular weight* as:

$$M_{\text{app}} = MP(q) \tag{3.14}$$

i.e. the apparent molecular weight is dependent on the scattering angle. For the apparent molecular weight to be the true molecular weight M it is necessary that $P(q) = 1$ which requires the scattering angle to be small enough (in fact equal to

[2] i.e. a Taylor series expansion of equation 3.12

zero degrees). This is consistent with what we have seen before: The form factor is practically 1 if q is small enough (which is very small if the particles are large and not so small if the particles are small) as seen in figure 3.1.
If we include the form factor we can generalise also equation 3.6 which was valid for small particles, noting that it can be rewritten:

$$\frac{KC}{R(q)} = \frac{1}{M} + 2A_2C + 3A_3C^2 + \cdots = \frac{1}{M}(1 + 2A_2MC + 3A_3MC^2 + \cdots) \quad (3.15)$$

We now simply substitute the expression for the apparent molecular weight for the molecular weight in this expression (3.15) thereby getting:

$$\frac{KC}{R(q)} = \frac{1}{MP(q)}(1 + 2A_2MP(q)C + 3A_3MP(q)C^2 + \cdots) \quad (3.16)$$

Now, the principle of determining the molecular weight becomes apparent: For a number of ever decreasing concentrations of the sample, C_1, C_2, C_3, ..., where $C_n \to 0$, measure the Rayleigh ratio $R(q)$ at a number of ever decreasing q values, q_1, q_2, q_3, ..., where $q_n \to 0$. Subsequently, for every q value extrapolate the values of $\frac{KC}{R(q)}$ to $C = 0$ thereby finding $1/(MP(q_1))$, $1/(MP(q_2))$, $1/(MP(q_3))$ .... Finally, extrapolate these values to $q = 0$. Hereby we obtain the true molecular weight M (or, actually $1/M$) as $P(0) = 1$. The extrapolation to $q = 0$ is less trivial than it may seem and is in practice often performed with the aid of a Guinier-plot as described in section 3.4.

3.4 Size determination

As we have just seen the form factor reflects the shape and size of the particles. If the particles are not so small that their form factor can be considered essentially constant (with the value 1) it is necessary to measure the intensity of the scattered light at several angles and subsequently perform an extrapolation of the intensity down to a scattering angle of $\theta = 0°$ (i.e. $q = 0$). This procedure may seem cumbersome which it is, especially if on e is only interested in the molecular weight of the particles. On the other hand, if the form factor is not equal to one at all scattering angles it actually contain information about the size and sometimes

the shape of the particles. In many cases one may be content with a coarse measure of the size of the particles and we shall now see how this can be accomplished. This coarse measure of size is called *the radius of gyration* R_g. We shall see how to arrive at this quantity through the calculation of some useful approximations of the form factor 3.12.

The Guinier approximation

If $q\,r_{jk} < 1$, the form factor 3.12 can be approximated by its Taylor series expansion to the smallest non-trivial order which turns out to be second order. The easiest way to do this by approximating the sine function by its Taylor series expansion: $\sin(x) = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \cdots$ leading to the approximation

$$\frac{\sin(x)}{x} \approx \frac{x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \cdots}{x} = 1 - \frac{1}{3!}x^2 + \frac{1}{5!}x^4 - \cdots$$

Keeping the first two terms only we then obtain

$$P(q) \approx \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}(1 - \frac{1}{6}\,(q\,r_{jk})^2) \tag{3.17}$$

This approximation can be written in a simpler and more useful form which is seen by expanding the right hand side of equation 3.17:

$$\begin{aligned}
\frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}(1 - \frac{1}{6}\,(q\,r_{jk})^2) &= \\
\frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}(1) + \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{1}{6}\,(q\,r_{jk})^2 &= \\
\frac{1}{n^2}\cdot n^2 + \frac{1}{6}\,q^2\cdot\frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}r_{jk}^2 &= \\
1 - \frac{1}{3}\,q^2\,R_g^2 &
\end{aligned}$$

where R_g is the aforementioned *radius of gyration* defined as:

$$R_g^2 = \frac{1}{2\,n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}r_{jk}^2 \tag{3.18}$$

The radius of gyration needs af few supplementary comments which can be found at the end of this section (page 55). We thus arrive at an expression for the form factor approximation which reads:

$$P(q) \approx 1 - \frac{1}{3}\, q^2\, R_G^2 \tag{3.19}$$

The approximation 3.19, called the *the Debye approximationen* is sometimes used "as is" but suffers from a flaw: It drops very fast down to 0 and even worse it becomes negative over a certain q value (the graph is a parabola intersecting the q-axis at $q = 3/R_g$). A negative value for the form factor is of course meaningless but even the fast drop down to zero is a bad representation of most measurements which usually drop much more gently to low values never becoming zero. This undesired feature of the Debye approximation can easily be remedied, however arbitrarily: Just choose a function which has a Taylor series expansion whose first terms coincide with the Debye approximation but which decays more gradually and stays positive. The most well-known of such approximations is *the Guinier approximation*:

$$P(q) \approx e^{-\frac{1}{3}\, q^2\, R_g^2} \tag{3.20}$$

whose Taylor series expansion is $1-\frac{1}{1!}(\frac{1}{3}q^2R_g^2)+\frac{1}{2!}(\frac{1}{3}q^2R_g^2)^2-\frac{1}{3!}(\frac{1}{3}q^2R_g^2)^3+\ldots$ and thus coincides with the Debye approximation on the two first terms. As it is based on a Taylor series expansion it requires q to be sufficiently small or rather that $q\,R_g$ being small enough. The criterion most frequently used to ensure the validity of the Guinier approximation is that $q\,R_g < 1$. How good the approximation actually is depends on the shape of the particles.

The Guinier approximation can be used to determine the size of particles but is also an intermediate step when determining the molecular weight of large particles: One measures the scattered intensity at several q values i.e. at several scattering angles. Assume that at each scattering angle the scattered intensity has been measure at a concentration C low enough that concentration effects can be neglected. Then the Rayleigh ratio can be calculated for each q value. With the Guinier approximation (3.20) we can write equation 3.13 as:

$$R(q) \approx KCM \cdot e^{-\frac{1}{3}\, q^2\, R_g^2} \tag{3.21}$$

Taking the logarithm on both sides we get:

$$\ln R(q) \approx \ln(KCM) - \frac{1}{3}\, q^2\, R_g^2 \tag{3.22}$$

It is seen that if $\ln R(q)$ is plotted against q^2 one gets a straight line with a slope coefficient a given by

$$a = -\frac{1}{3}\, R_g^2 \tag{3.23}$$

and an intercept with the vertical axis of $\ln(KCM)$ If the only information wanted from the light scattering experiment is the radius of gyration it is not necessary to calculate the Rayleigh ratio. As the Rayleigh ratio is proportional to the scattered intensity $I_s(q)$ we get, in analogy with equation 3.22:

$$\ln I_s(q) = \ln I_s(0) - \frac{1}{3}\, q^2\, R_g^2 \tag{3.24}$$

A plot of measuered Rayleigh ratios or scattered intensities, $\ln R(q)$ or $\ln(I_s(q))$ vs. q^2 is called a *Guinier plot*. In practice the slope a is found by linear regression which will also give either $\ln(KCM)$ or $\ln(I_s(0))$ from which the molecular weight may be calculated. However, before accepting the results of the linear regression one should make sure that the Guinier approximation was in fact valid. This amounts to check if the radius of gyration found R_g and the largest value of q used for the linear regression ($q_{\max}$) satisfy the condition $q_{\max}\, R_g < 1$. If the criterion is not met the linear regression should be confined to only smaller values of q, i.e. using a smaller value of $q_{\max}$. In practice there is a limit to how small scattering angles can be used in a measurement and thus to how small q values are experimentally accessible. This puts a practical upper limit to how large particles it is possible to measure. Note, that if the particles are so large that it is impossible to fulfil the criterion $q_{\max}\, R_g < 1$ then neither the radius of gyration nor the molecular weight can determined. But possibly a lower bound to both can be established (see the worked example below). There is also a limit to how small particle sizes it is possible to determine: If the particles are very small the slope in the Guinier plot $a = -\frac{1}{3}\, R_g^2$ becomes so small that the linear regression can not determine it with meaningful accuracy. Due to inevitable scatter of the measured points the slope may in practice sometimes become negative in which

case no radius of gyration satisfies equation 3.23. The lower limit is not a theoretical but only a practical one. It is a matter of the statistical scatter (noise) of the measured scattering intensities. In literature the limit of isotropic scattering is often set at a particle size of $\lambda/20$, meaning that below this size you cannot determine what the size actually is. The criterion, however, is neither a strict one nor is it very precise. It is often not specified whether the wavelength is the vacuum wavelength of the source or the ensuing wavelength in the medium surrounding the scattering particles. Likewise it is usually not specified what is meant by the *size* of the particle. In practice the smallest radius of gyration measurable today (2013) is about $R_g \approx \lambda_0/60$, where λ_0 is the wavelength of the laser employed.

Comments on the radius of gyration

Note that *any* object can be assigned a radius of gyration. This is an exact property of any object. The approximation lies in just keeping the first few terms in the Taylor series thus apparently disregarding details about the particle structure and shape and only keeping one aspect of the particle "structure": its radius of gyration.
The squared radius of gyration is thus half the average value of the squared distances between the constituent parts of the particle. The half originates from the fact that we are defining what we call a *radius* not a diameter.
As an example of this consider a "particle" consisting of two identical subunits, 1 and 2, separated by the distance d. The radius of gyration for this particle would be (with $n = 2$):

$$\begin{aligned} R_g^2 &= \frac{1}{2\,n^2}\sum_{j=1}^{n}\sum_{k=1}^{n} r_{jk}^2 \\ &= \frac{1}{2\cdot 2^2}\cdot(r_{11}^2 + r_{12}^2 + r_{21}^2 + r_{22}^2) \\ &= \frac{1}{8}\cdot(0^2 + d^2 + d^2 + 0^2) \\ &= \frac{1}{4}\cdot d^2 = \left(\frac{d}{2}\right)^2 \end{aligned}$$

which shows that the radius of gyration assumes the "natural" value of half the separation between the two constituent subunits.

Alternate definition of the radius of gyration

There is another definition of the radius of gyration which may be easier to envisage than the average of n^2 intraparticle distances: The radius of gyration is related to the (only) n distances from the particle subunits to the particle's centre of mass.

$$R_g^2 = \frac{1}{n}\sum_{j=1}^{n} r_{j,\,\mathrm{c.m.}}^2 \tag{3.25}$$

where $r_{j,\,\mathrm{c.m.}}$ denotes the distance from the j'th subunit to the particle's centre of mass. The two definitions of R_g turn out to equivalent but we shall put off the proof to section 3.6.

3.5 Size and molecular weight. How to average in mixtures

Suppose you have a solution containing a mixture of different kinds of particles. These may be entirely different molecular species or could be different oligomeric states of the same basic molecular unit, i.e. it could be monomers, dimers, trimers, tetramers, etc. of the same molecule. These will be treated as different types of particles. Suppose we have n different types of particles with molecular weights $M_1, M_2, \ldots, M_n$, radii of gyration $R_{1G}, R_{2G}, \ldots, R_{nG}$ and which are present at weight concentrations $C_1, C_2, \ldots, C_n$. Furthermore we will assume that the light scattering contributions from the different types of particles can simply be added together. This essentially means that the concentrations are so low that the particles can be considered non-interacting. Thus we can write for the total light scattering for the mixture:

$$R_{total}(q) = \sum_{i=1}^{n} R_i(q) = K \sum_{i=1}^{n} C_i M_i P_i(q) \tag{3.26}$$

How to calculate averages

Suppose you do a light scattering measurement on such a mixture in the same manner as you would do if measuring on a solution of only one species, (i.e. measure at many different angles and possibly at many different dilutions). If you then do a data analysis of the measurement results assuming that the solution only contained one species then you would arrive at an average molecular weight and possibly an average radius of gyration. In order to be able to understand, interpret or use these average values it is important to realise how the individual molecular weights and radii of gyration are weighted relative to each other. In other words, what *type of average* you get from a light scattering measurement. This we shall now look at:

First, if we want to calculate the average molecular weight $\langle M \rangle$, the corresponding weight concentration must be the total weight concentration of the species present in the solution (disregarding the solvent molecules themselves). This means that

the total light scattering can be written as:

$$R_{total}(q) = K\left(\sum_{i=1}^{n} C_i\right)\langle M\rangle\langle P(q)\rangle \tag{3.27}$$

If we set equal the two expressions for the total light scattering R_{total} from equations 3.26 and 3.27 we get

$$K\left(\sum_{i=1}^{n} C_i\right)\langle M\rangle\langle P(q)\rangle = \sum_{i=1}^{n} R_i(q) = K\sum_{i=1}^{n} C_i M_i P_i(q) \tag{3.28}$$

If, furthermore, we set $q = 0$ (extrapolate the measurements to zero scattering angle) all form factor assume the value 1, we get that the average molecular weight can be calculated as:

$$\langle M\rangle_w = \frac{\sum_{i=1}^{n} C_i\, M_i}{\sum_{i=1}^{n} C_i} = \frac{\sum_{i=1}^{n} C_i\, M_i}{C_{\text{total}}} \tag{3.29}$$

This type of average (indicated by the subscript w by the angle bracket) is called *the weight average molecular weight* as opposed to other possible averages, like e.g. the number average molecular weight where the individual molecular weights would be multiplied by the *molar* concentrations instead of the weight concentrations and the total concentration would be the total molar concentration. Note that this number average weight concentration cannot be determined by light scattering. Only the weight average molecular weight can. Note that we have tacitly assumed that the contrast constant K is the same for all molecular species. This is often the case but not always. If the different molecular species have different values of the contrast constant the simple equation 3.29 does not hold.

In a similar way we can determine an average radius of gyration for the molecules in the solution. To this end we use the Debye-approximation 3.19 for the formfaktor. Substituting into equation 3.26 we get:

$$R_{total}(q) = \sum_{i=1}^{n} R_i(q) = K\sum_{i=1}^{n} C_i M_i(1 - \frac{1}{3}R_{g\,i}^2 q^2) \tag{3.30}$$

where $R_{g\,i}$ denotes the radius of gyration of the i'th type of molecule. If again we assume that the scattering is due to some "average" particles with the average molecular weight $\langle M\rangle_w$, an average (squared) radius of gyration $\langle R_g^2\rangle$ and a

concentration $C_{\text{total}} = \sum_{i=1}^{n} C_i$ we can write the total Rayleigh ratio as:

$$R_{total}(q) = KC_{\text{total}}\langle M\rangle_w(1 - \frac{1}{3}\langle R_g^2\rangle q^2) \tag{3.31}$$

When we compare equations 3.30 and 3.31 equating the terms containing R_g we get the expression for the average radius of gyration:

$$\langle R_g^2\rangle = \frac{\sum_{i=1}^{n} C_i\, M_i\, R_{g\,i}^2}{C_{\text{total}}\langle M\rangle_w} \tag{3.32}$$

which means that the individual radii of gyration (squared) are weighted by the light scattering contribution of the same species. Using the expression for the average molecular weight 3.29 we can rewrite equation 3.32 into its more common form:

$$\langle R_g^2\rangle = \frac{\sum_{i=1}^{n} C_i\, M_i\, R_{g\,i}^2}{\sum_{i=1}^{n} C_i\, M_i} \tag{3.33}$$

This average value is called the $z-$average of $\langle R_g^2\rangle$ and the square root of this quantity is called the $z-$average radius of gyration, $\langle R_g\rangle_z$.

A word of caution: The average quantities obtained by light scattering, $\langle M \rangle_w$ and $\langle R_g \rangle_z$ can be hard to interpret. They are both strongly biased towards the influence of high molecular weight species. Imagine a solution of protein where the pure, monomeric species has a molecular weight M_1. Chemists will tell that the solution is 99% pure monomer plus 1% aggregates of the monomer. The purity may be stated in weight percent or in mole percent. Assume first that we are talking about weight percent, e.g. $9.9\,\mathrm{g}\cdot\mathrm{L}^{-1}$ of the monomeric species and $0.1\,\mathrm{g}\cdot\mathrm{L}^{-1}$ of the aggregates giving a total weight concentration of $10.0\,\mathrm{g}\cdot\mathrm{L}^{-1}$. What if the aggregates are really big consisting of, say, 100 monomeric species? The aggregates would then have a molecular weight of $100M_1$. How would light scattering judge the molecular weight? Inserting into equation 3.29 gives the average molecular weight:

$$\begin{aligned}\langle M \rangle_w &= \frac{\sum_{i=1}^{n} C_i\, M_i}{C_{\mathrm{total}}} \\ &= \frac{9.9 \cdot M_1 + 0.1 \cdot (100 M_1)}{10.0} = \frac{9.9 M_1 + 10.0 M_1}{10.0} \\ &= 1.99 M_1\end{aligned}$$

or almost twice the monomer molecular weight. Had the aggregate been made up of 1000 monomeric units but still only at a weight concentration of 1% the average molecular weight would have been almost 11 times higher than the monomer molecular weight. It is thus evident that it is extremely important to do everything possible to measure on clean well filtered samples in extremely pure buffers. Even then the usefulness of the weight average molecular weight will depend on how "well behaved" the sample is. Also one should note that in the above example specifying the purity by molar concentration would make things seem even worse: In the case where the aggregates have a molecular weight of $100M_1$ a weight fraction of 1% would correspond to a molar fraction of approximately 0.01%, i.e. the solution would be 99.99% pure by molar fraction but still the average molecular weight would be wrong by a factor of two!
Similar arguments hold regarding the $z-$average of the radius of gyration. So again these averages can be hard to interpret and could be strongly dependent on details of sample preparation.

Detailed determinations

The ambiguities of measuring in principle only average values of the molecular weight and the radius of gyration can sometimes be circumvented by combining light scattering with size exclusion chromatography. Instead of using a classic light scattering setup (see figure 1.2) another setup is used where the light scattering instrument measures the scattered intensity at several angles simultaneously. The the combined technique is called SEC-MALS (Size Exclusion - Multi Angle Light Scattering). A sketch of the setup is shown in figure 3.2.

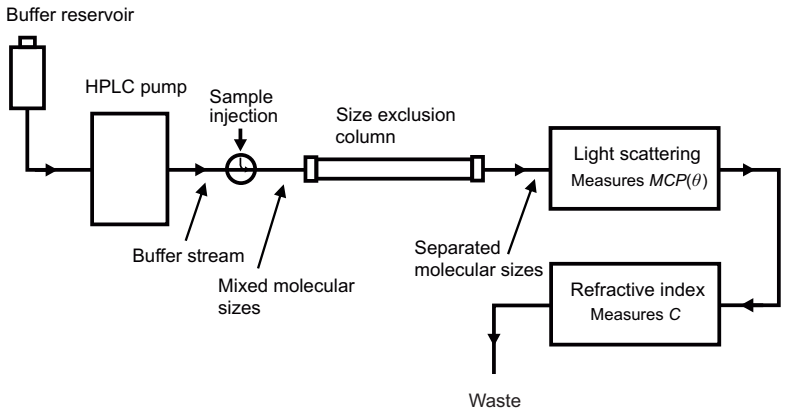


Figure 3.2: The HPLC-pump creates a constant flow of liquid (buffer) through the gel column. The sample, typically a 100 μL solution, is injected into the buffer flow through a special port (valve). As the molecules of the sample are flushed through the gel column the larger species will pass through quickly whereas the smaller species will come out later. The smaller the molecules the later they *elute*. The different size classes of molecules then pass through the light scattering instrument where the Rayleigh ratio ($KCMP(q)$) is determined. Next the molecules are carried though a differential refractometer (or, alternatively, a UV-detector) which determines the concentration C of the molecules. Having measured both $KCMP(q)$ at several angles and C the taking the ratio the molecular weight M can be calculated and also the radius of gyration if the scattering is sufficiently anisotropic. This can be done for every species eluting from the column.

The principle of the measurement is that the light scattering measurements are done on the molecules while they are flowing through a measurement cell and subsequently flowing through another instrument which measures the concentra-

tion og the molecules. The sample solution is injected into the buffer flow which continuously flows through the gel column and subsequently trough the light scattering cell and the refractometer cell. The gel column separates the molecules in the buffer flow in such a way that the larger molecules make a quick passage whereas the smaller molecules are retained longer before they eventually elute. When the molecules leave the gel column they pass through the light scattering instrument which determines the Rayleigh ratio $R(q) = KCMP(q)$ at several scattering angles (several values of q) at the same time. Remember that K denotes the optical contrast constant and $P(q)$ is the form factor for the molecular species. As the Rayleigh ratio is measure at many angles at the same time it is usually possible to perform the extrapolation to zero q so that $R(0) = KCM$ can be determined. After leaving the light scattering instrument the molecules flow through a differential refractometer which measures the difference in refractive index between the sample solution and the pure buffer, $n - n_0$. This difference can be written (neglecting higher order terms in the concentration)

$$n - n_0 = \frac{dn}{dC} \cdot C \tag{3.34}$$

As the quantity $\frac{dn}{dC}$ called the refractive index increment can be determined by separate measurements (or be found in a table) a measurement of the refractive index difference is a way to determine the weight concentration of the molecules. As the light scattering instrument in principle measures MC and the refractometer measures C the molecular weight can be found as the ratio between the two measurements.
More details about this method can be found in chapter 5.
As the light scattering instrument measures at many angles at the same time the form factor $P(q)$ is determined for each of the molecular species that flow through the instrument separated by the column. It is therefore possible to determine the radius of gyration for the different species if the $q-$values (scattering angles) of the instrument make it possible:

1. If the particles are very small the $q-$values of the instrument are not sufficiently large for the form factor to decrease measurably at even the largest

$q-$value. In this case only an upper bound can be set for the radius of gyration.

2. If the particles are very large even the smallest scattering angles of the instrument gives $q-$values that are too big for the criterion $q \cdot R_g < 1$ to be fulfilled so none of the approximations used to determine R_g are valid.

3.6 Equivalence of the two definitions of R_g

We still need to prove that the two definitions of the radius of gyration are equivalent:
First, remember that the position $\vec{r}_{\text{c.m.}}$ of the centre of mass is defined as

$$\vec{r}_{\text{c.m.}} = \frac{1}{n}\sum_{j=1}^{n}\vec{r_j} \tag{3.35}$$

where $\vec{r}_1$, $\vec{r}_2$, $\vec{r}_3$, ... are the positions of the particle's subunits.
First let us start with the new definition expanding the expression:

$$\begin{aligned}
R_g^2 &= \frac{1}{n}\sum_{j=1}^{n} r_{j,\,\text{c.m.}}^2 \\
&= \frac{1}{n}\sum_{j=1}^{n}(\vec{r_j} - \vec{r}_{\text{c.m.}}) \cdot (\vec{r_j} - \vec{r}_{\text{c.m.}}) \\
&= \frac{1}{n}\left(\sum_{j=1}^{n} r_j^2 + \sum_{j=1}^{n} r_{\text{c.m.}}^2 - 2 \cdot \sum_{i=1}^{n} \vec{r_i} \cdot \vec{r}_{\text{c.m.}}\right) \\
&= \frac{1}{n}\left(\sum_{j=1}^{n} r_j^2 + \sum_{j=1}^{n} r_{\text{c.m.}}^2 - 2\vec{r}_{\text{c.m.}} \cdot \sum_{j=1}^{n} \vec{r_j}\right) \\
&= \frac{1}{n}\left(\sum_{j=1}^{n} r_j^2 + n r_{\text{c.m.}}^2 - 2\vec{r}_{\text{c.m.}} \cdot (n\vec{r}_{\text{c.m.}})\right) \\
&= \frac{1}{n}\left(\sum_{j=1}^{n} r_j^2 - n r_{\text{c.m.}}^2\right) = \frac{1}{n}\sum_{j=1}^{n} r_j^2 - r_{\text{c.m.}}^2
\end{aligned}$$

Next, let us expand the original definition 3.18 of R_g:

$$R_g^2 = \frac{1}{2\,n^2}\sum_{j=1}^{n}\sum_{k=1}^{n} r_{jk}^2$$

$$
\begin{aligned}
&= \frac{1}{2\,n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}(\vec{r}_j-\vec{r}_k)\cdot(\vec{r}_j-\vec{r}_k)\\
&= \frac{1}{2\,n^2}\cdot\left(\sum_{j=1}^{n}\sum_{k=1}^{n}r_j^2+\sum_{j=1}^{n}\sum_{k=1}^{n}r_k^2-2\cdot\sum_{j=1}^{n}\sum_{k=1}^{n}\vec{r}_j\cdot\vec{r}_k\right)\\
&= \frac{1}{2\,n^2}\cdot\left(n\cdot\sum_{j=1}^{n}r_j^2+n\cdot\sum_{k=1}^{n}r_k^2-2\cdot(\sum_{j=1}^{n}\vec{r}_j)\cdot(\sum_{k=1}^{n}\vec{r}_k)\right)\\
&= \frac{1}{2\,n^2}\cdot\left(n\cdot\sum_{j=1}^{n}r_j^2+n\cdot\sum_{j=1}^{n}r_j^2-2\cdot(n\vec{r}_{\mathrm{c.m.}})\cdot(n\vec{r}_{\mathrm{c.m.}})\right)\\
&= \frac{1}{2\,n^2}\cdot\left(2n\sum_{j=1}^{n}r_j^2-2n^2r_{\mathrm{c.m.}}^2\right)=\frac{1}{n}\sum_{j=1}^{n}r_j^2-r_{\mathrm{c.m.}}^2
\end{aligned}
$$

It is thus seen that the two expressions for R_g can be rewritten into the same form and thus are equivalent.

3.7 Summing up

Here are the most important equations from this chapter:

The form factor definition:

$$P(q) = \frac{I_s(q)}{I_s(q=0)}$$

Apparent molecular weight:

$$M_{\text{app}} = MP(q)$$

Radius of gyration:

$$R_g^2 = \frac{1}{2\,n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} r_{jk}^2$$

or

$$R_g^2 = \frac{1}{n} \sum_{j=1}^{n} r_{j,\,\text{c.m.}}^2$$

The Debye approximation:

$$P(q) \approx 1 - \frac{1}{3}\, q^2 R_g^2 \quad \text{for} \;\; R_g \cdot q \leq 1$$

The Guinier approximation:

$$P(q) \approx e^{-\frac{1}{3} q^2 R_g^2} \quad \text{for} \;\; R_g \cdot q \leq 1$$

General equation for static light scattering:

$$\frac{KC}{R(q)} = \frac{1}{MP(q)} \left(1 + 2A_2 MP(q)C + 3A_3 MP(q)C^2 + \cdots\right)$$

Radii of gyration that can be expressed analytically

Examples:

Massive sphere of radius r:

$$R_g = \sqrt{\frac{3}{5}}\, r$$

Ellipsoid with semi-axes a, b og c:

$$R_g = \sqrt{\frac{a^2 + b^2 + c^2}{5}}$$

Spherical shell (infinitely thin) with radius r:

$$R_g = r$$

Cylinder with radius r and length L:

$$R_g = \sqrt{\frac{r^2}{2} + \frac{L^2}{12}}$$

Box with sides a, b og c :

$$R_g = \sqrt{\frac{a^2 + b^2 + c^2}{12}}$$

Note, that the last equation also gives the radius of gyration for an infinitely thin rectangular plate ($c = 0$) and for an infinitely thin rod ($b = 0$ and $c = 0$)

Special cases:

Static light scattering for **very dilute solutions**:

$$\frac{KC}{R(q)} = \frac{1}{MP(q)}$$

Static light scattering for **very small particles**:

$$\frac{KC}{R(q)} = \frac{1}{M} + 2A_2C + 3A_3C^2 + \cdots$$

Static light scattering for **very dilute solutions of very small particles**:

$$\frac{KC}{R(q)} = \frac{1}{M}$$

Form factors in closed analytical form

Many *ad hoc* form factors exist to describe empirically the "shape" of scattering data. However, only few form factors can be derived analytically and expressed in a closed mathematical form. Here we just mention the form factor of a massive, homogeneous sphere and the form factor of a random coil.

Massive sphere with radius R:

$$P(q) = \left(\frac{3 \cdot (\sin(qR) - qR\cos(qR))}{q^3} \right)^2$$

Random coil with radius of gyration R_g:

$$P(q) = \frac{2}{(qR_g)^4} \cdot (\exp(-(qR_g)^2) - 1 + (qR_g)^2)$$

How these two form factors may look is shown in figure 3.3

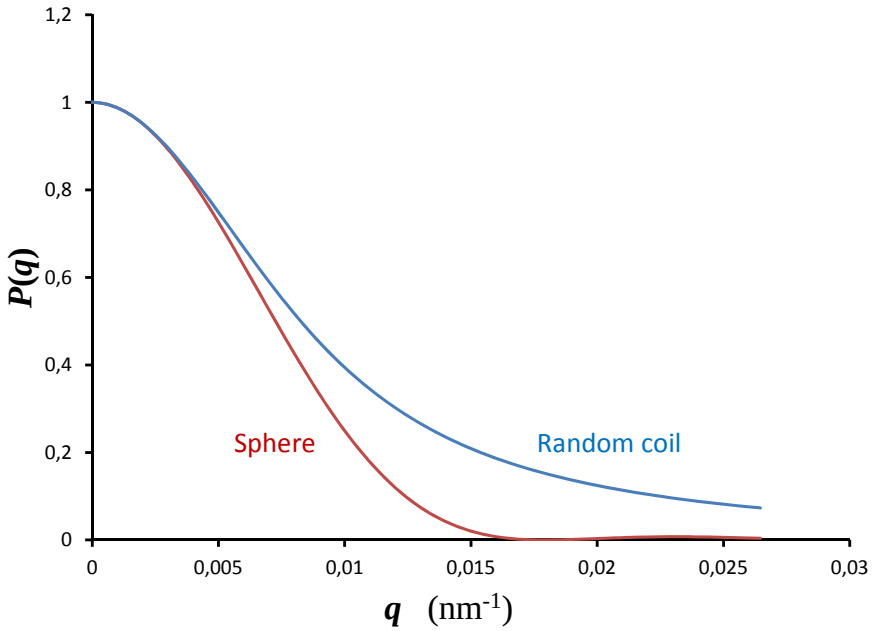


Figure 3.3: The two form factors correspond to the same radius of gyration: The sphere radius is set at 250 nm and the random coil radius of gyration is set at the value $R_g = \sqrt{\frac{3}{5}} \cdot 250\,\text{nm}$. Note that having the same radius of gyration the two form factors coincide at small q-values, as expected. The $q-$value corresponding to $q \cdot R_g = 1$ is 0.005 nm. The whole q-range corresponds to a vacuum wavelength $\lambda_0 = 633\,\text{nm}$ (a HeNe laser) and a range of scattering angles from 0° to 180°.
Note also that the sphere form factor drops to exactly zero and then first rises and later drops again. This damped oscillatory behaviour is only found with highly symmetric particles like spheres, ellipsoids and cylinders.

4

Simple theory for dynamic light scattering, DLS

Dynamic light scattering (DLS) is a relatively new technique dating back, in principle, to 1964. The technique is also known as *quasi elastic light scattering* (QELS) or *photon correlation spectroscopy* (PCS). The technique has gained increasing popularity ever since the mid-seventies. The reason why this technique came about so late is that in many ways it is technologically more demanding than static light scattering: The light source need to be very monochromatic, intense and coherent (so the light can be focussed to a very narrow bundle of rays). All of this essentially means that the light source has to be a laser (invented in 1960). Furthermore the detector system has to be very sensitive (being able to detect single photons) and relatively fast. And last, the post-processing of the detected

signals requires fast, specialised electronics with hig computing capacity.

The purpose of doing dynamic light scattering experiments is usually to measure the diffusion coefficient of molecules in solution but may also be to measure characteristic relaxation times in e.g. gel systems. We shall here only deal with measurements of diffusion processes. Through the measurement of the diffusion coefficient for particles in suspension it is possible to determine indirectly the size of the particles. The size measure obtained in this way is called *the hydrodynamic radius* or sometimes *the Stokes radius* of the particles because the determination is based on the *Stokes-Einstein* relationship which relates the diffusion coefficient of suspended spheres with radius r with their diffusion coefficient D

$$D = \frac{k_B T}{6\pi\eta r} \tag{4.1}$$

where T is the absolute temperature of the suspension, η is the viscosity of the liquid and $k_B = 1.38 \cdot 10^{-23}\ \mathrm{J \cdot K^{-1}}$ is Boltzmann's constant. Strictly speaking equation 4.1 is valid only for non-interacting particles to whose surface the surrounding liquid sticks. If the diffusion coefficient D_{exp} of some particles is experimentally determined the diffusion coefficient can be inserted into the Stokes-Einstein equation 4.1 which can then be solved for the radius. The radius thus obtained *the hydrodynamic radius* r_h and is defined as:

$$r_h = \frac{k_B T}{6\pi\eta D_{\mathrm{exp}}} \tag{4.2}$$

where D_{exp} is the measured diffusion coefficient of the suspended particles. If the particles are not spheres the hydrodynamic radius is nevertheless a rough measure of the particle size in the same way that the radius of gyration is. The two numbers are usually not exactly equal. If the particles are non-spherical their hydrodynamic radius is often approximately the radius of a sphere with the same volume as the particle. This hold as long as the particles do not have extreme geometry like being very long and thin or being wide and flat. As a rule of thumb expect the hydrodynamic radius to increase by 5 – 7% relative to the radius of a same volume sphere when the axial ratio increases by one.

The diffusion coefficient is – like the Rayleigh ratio – determined thermodynamically though the osmotic pressure Π. Equation 4.1 is a special case of *the gener-*

alised Stokes-Einstein equation:

$$D = \frac{M}{N_A f}(1-\phi)^2 \left(\frac{\partial \Pi}{\partial C}\right)_{T,P} \tag{4.3}$$

where ϕ denotes the volume fraction of the suspended (dissolved) particles and f is the so-called frictional coefficient. The frictional coefficient for spheres with the liquid sticking to their surface has been calculated by Stokes as $f = 6\pi\eta r$. Unfortunately this case is more complicated than the determination of the molecular weight when it comes to the influence of concentration. In his former case it was sufficient to make a series expansion (a virial expansion) of the osmotic pressure to arrive at the virial coefficients. In the present case not only the osmotic pressure but also also the frictional coefficient f depends on the concentration which tends to make things a lot more complicated. A theoretical exposition of the influence of concentration on the diffusion coefficient is therefore difficult. However, it is of course possible to write the concentration dependence as a Taylor series

$$D = D_0(1 + k_D C + \cdots) \tag{4.4}$$

where k_D is a coefficient for the first order dependence of the diffusion coefficient on the concentration. A number of different approaches to the calculations of this coefficient exist but we shall not be concerned with these here. It is however evident that to obtain the diffusion coefficient of some particles it is necessary to do the measurements at several concentrations that make it possible to do an extrapolation to zero concentration.

In the following we shall see hoe diffusion coefficient can be measured using dynamic light scattering.

4.1 Intensity fluctuations

When performing dynamic light scattering measurements the setup has some similarity to the setup used for static light scattering (figure 1.2). There are, however some important differences (see figure 4.1). The laser beam is focussed down to a diameter of approximately $\frac{1}{10}$ mm and the scattered photons are detected by a photomultiplier tube (PMT) or sometimes be an avalanche diode detector which are

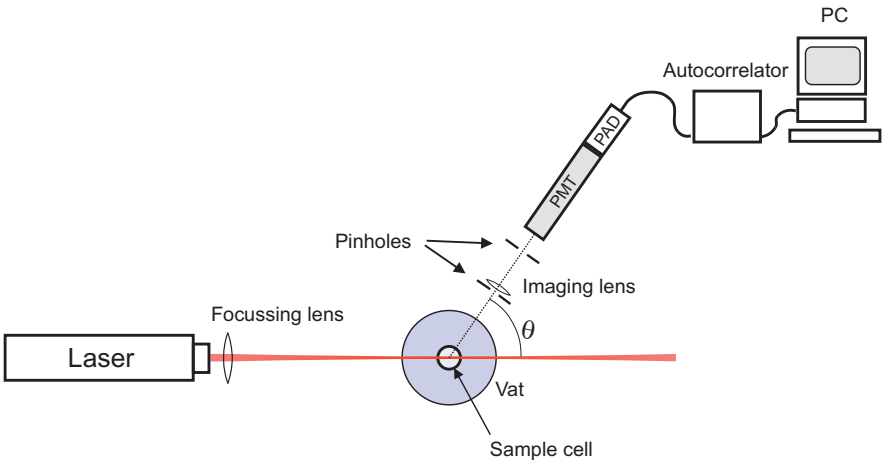


Figure 4.1: Setup for measuring dynamic light scattering (DLS). The detector is usually a photomultiplier tube (PMT) of the type that can detect single photons but can also be based on an avalanche diode. The detection is done through a number of pinholes where the smallest one is nearest to the detector. It usually has a diameter of $100-200\,\mu$m. Also the laser beam is focused in the centre of the measurement cuvette down to a diameter of about $100\,\mu$m. Furthermore the measurement cuvette is placed in a vat embedded containing a thermostatted liquid as the measured diffusion coefficient is temperature dependent.

able to detect single photons. Each detected photon creates an electrical voltage pulse which is subsequently sent to a pulse amplifier and discriminator (PAD) which only lets through pulses of a sufficient voltage (indicative of them originating from photons an not internal noise in the detector) and then shapes these pulses into a standard voltage (5 V) and a standard duration (e.g. 20 ns). These standardised pulses are the fed into a digital autocorrelator, a piece of hardware which calculates the autocorrelation function for the detected photons. The inner workings of the digital autocorrelator is discussed below and the autocorrelation function is described in section 4.2.

As we saw in section 2.7 the intensity of the light scattered from N particles can be written (see equation 2.37):

$$I_{s,\,total}(t) = I_{s,\,n} \sum_{j=1}^{N} \sum_{k=1}^{N} \cos(\phi_j(t) - \phi_k(t)) \tag{4.5}$$

Here the time dependence of the intensity is indicated explicitly with the t in the bracket. The reason why the intensity depends on time is that the particles

move about due to diffusion (*Brownian motion*). These random movements of the particles make the individual phases $\phi_j(t)$ and $\phi_k(t)$, and thereby the phase differences, change with time in a random way. The diffusion coefficient depends on two things, the size (and shape) of the particles and the viscosity of the embedding liquid. If, e.g., the particles are big or if the liquid has a high viscosity the particles will move about slowly Consequently the phases and thus $\phi_j(t) - \phi_k(t)$ will change slowly and $I_{s,\,total}(t)$ becomes a slowly fluctuating of function of time. The rapidness or the average frequency of the fluctuations thus becomes a measure of how rapidly the particles move about by diffusion or, in other words, a measure of the diffusion coefficient D.

How to measure fluctuations

If you look at equation 4.5 two things are to be noted:

1. The scattered intensity is measured at a *point*

2. If the sum contains many terms the fluctuations of the individual terms will tend to average out

In a practical DLS setup the laser beam is focussed to a very narrow diameter, typically $1/10\,\mathrm{mm}$ ensuring that not too many particles contribute to the sum. If a screen is set up next to a scattering solution the scattered light shows on the screen in the form of speckles moving rapidly around. In order for the naked eye to actually see this phenomenon one needs large, slowly diffusing particles and a very narrow laser beam. The size of the speckles is larger the smaller the diameter of the laser beam. In order to have large intensity fluctuations on the detector it needs to see only one speckle at a time, i.e. the effective detector area should match the size of the speckles: If the detector area is much smaller than the speckle size the fluctuations will still be large but little light is collected, leading to poor statistics. If the detector area is made very large in order to collect a lot of light the detector will see many speckles (but nearly the same number) all the time and the fluctuations will be weak.

4.2 Making sense of fluctuations: The autocorrelation function

Although intuitively clear that a rapidly fluctuating scattered intensity corresponds to a large diffusion coefficient (i.e. small particles) and slowly fluctuating scattered intensity corresponds to small diffusion coefficient (i.e. large particles) it is not immediately evident how to obtain a precise, quantitative measure of the fluctuations. This is of course a necessary first step before the fluctuations can be used to make precise numeric statements about the diffusion coefficient. In the early days of DLS it was customary to use the spectrum of characteristic fluctuation frequencies to quantitate the fluctuations. But soon this approach was superseded by the use of the so-called autocorrelation function $G_2(t)$. The two methods are equivalent but the electronic circuitry and the mathematical treatment of the acquired data differ. The autocorrelation function for the scattered light is defined by the equation:

$$G_2(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T I_{s,\,total}(t) \cdot I_{s,\,total}(t+\tau)\, dt \tag{4.6}$$

meaning that the integral contains scattered intensities recorded with a time difference τ. The function $G_2(\tau)$ is called *the intensity autocorrelation function for the scattered light* because it correlates *intensities*. One can define a so-called *field autocorrelation function* $G_1(\tau)$ in a similar manner substituting the intensities with the magnitude of the electric field instead:

$$G_1(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T E_{s,\,total}(t) \cdot E_{s,\,total}(t+\tau)\, dt \tag{4.7}$$

We shall come back to the relation between the two autocorrelation functions shortly. In practice the intensity is measured at discrete times. It is therefore more straightforward and precise to count how many photons are detected within a given time frame Δt called *the sample time*. Instead of the continuous quantity $I_{s,\,total}(t)$, the total scattered intensity one measures a sequence of photon counts $n(0\cdot\Delta t),\ n(1\cdot\Delta t),\ n(2\cdot\Delta t),\ n(3\cdot\Delta t), \ldots$, all recorded during the time period Δt. The sample time Δt can vary, typically from 0.1 to, say 50 μs. It is (was)

a quantity set by the experimenter. In the early days of DLS (i.e. before 1985) autocorrelators simply had a dial switch used to set the sample time. Newer correlators are so-called multiple tau correlators which are essentially an assembly of connected autocorrelators which operate at different sample times. We shall not describe the construction of such multiple tau correlators but only look at the classic linear tau correlators which are easier to describe. When the autocorrelation function is measured at discrete times we define the autocorrelation function at these times $\tau = 0 \cdot \Delta t,\ \tau = 1 \cdot \Delta t,\ \tau = 2 \cdot \Delta t,\ \tau = 3 \cdot \Delta t,\ \tau = 4 \cdot \Delta t,\ \ldots$ as:

$$\begin{aligned}
G_2(0 \cdot \Delta t) &= \frac{1}{N}\sum_{j=1}^{N} n(j \cdot \Delta t) \cdot n([j+0] \cdot \Delta t) \\
G_2(1 \cdot \Delta t) &= \frac{1}{N}\sum_{j=1}^{N} n(j \cdot \Delta t) \cdot n([j+1] \cdot \Delta t) \\
G_2(2 \cdot \Delta t) &= \frac{1}{N}\sum_{j=1}^{N} n(j \cdot \Delta t) \cdot n([j+2] \cdot \Delta t) \\
&\vdots \\
G_2(k \cdot \Delta t) &= \frac{1}{N}\sum_{j=1}^{N} n(j \cdot \Delta t) \cdot n([j+k] \cdot \Delta t) \qquad (4.8) \\
&\vdots
\end{aligned}$$

or put differently:

$$G_2(k \cdot \Delta t) = \langle n(t) \cdot n(t + k \cdot \Delta t) \rangle_t \qquad (4.9)$$

The time $k \cdot \Delta t$ is called *the correlation time* or *the lag time*. This is simply a multiple of the sample time Δt. Usually k has a maximum value of 128 or 256 effectively limiting the maximum value of the lag time. The 128 or 256 values of the autocorrelation function $G_2(k \cdot \Delta t)$ are stored in memory registers called channels. Every time the time Δt has elapsed a new term is added to each of the 256 sums defining the autocorrelation function. As these terms are products of the form $n(j \cdot \Delta t) \cdot n([j+2] \cdot \Delta t)$ 256 additions and 256 multiplications[1] have to be done during the time Δt. If $\Delta t = 0.1\,\mu$s it requires a calculation speed

[1] Alternatively the whole sequence $n(0 \cdot \Delta t)$, $n(1 \cdot \Delta t), \ldots, n(N \cdot \Delta t)$ could be store in computer memory whereupon the sums of 4.8 could be calculated once the measurements

of 2.56 billion multiplications + 2.56 billion additions per second. So far (2013) no desktop computer is fast enough to perform this task so instead specialised hardware is used to calculate the autocorrelation function a so-called *digital autocorrelator* usually comprising 256 small, primitive integrated computers running in parallel. Autocorrelators can be made as expansion boards for desktop PS's or as stand-alone boxes. Figure 4.2 shows a selection.

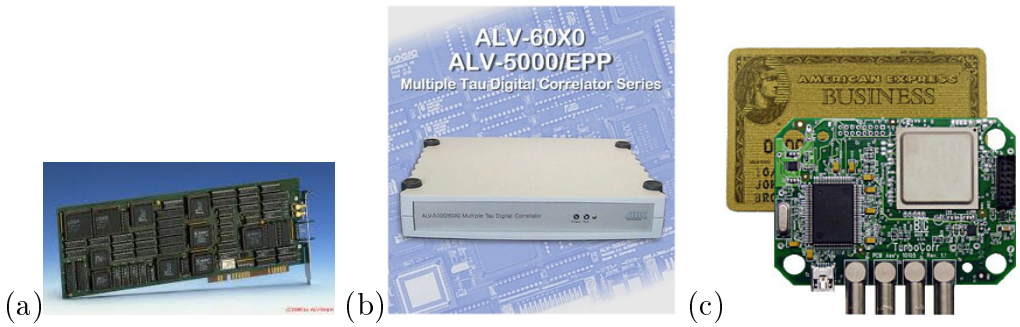


Figure 4.2: Three different digital autocorrelators from companies ALV Brookhaven Instruments: (a) ALV 5000, (b) ALV 5000/EPP, (c) TurboCorr

Returning briefly to the multiple tau correlators the general principle is simple: If a sample contains particles with only one diffusion coefficient the sample time Δt can be set byt he experimenter so that the exponential decay effectively takes place over the first half of the channels and the remaining half of the channels are used to establish the baseline. Thus it may take a few tries before a good value of Δt is found. If the sample contains particles with very different diffusion coefficient with perhaps a hundredfold ratio it is not possible to choose a sample time that will enable the capture of the two ensuing exponential decays at the same time. Also, establishing the baseline may cause some problems and at least the autocorrelation function will have to be "patched" together from different meas-

were completed. This of course would set an upper limit to the length of measurements, a limit which would be of the order minutes rather than hours even with several GB of memory to store the intermediate photon counts.

urements. With a sample of unknown composition the work involved in selecting proper sample times and the subsequent treatment of the different autocorrelation functions can be quite time consuming. But how many channels do you actually need to determine a single exponential decay? Not very many, say four + four to establish the baseline, so the 256 channels may seem like a lot. But if the sample time is chosen so that some small species produces an exponential decay within the first 8 channels (4+4) the smallest diffusion coefficient measurable with the sample time chosen would be 256/8 = 32 times smaller (corresponding to particles 32 times bigger). If the sample contains species which differ in size by more than a factor of 32 more measurements would be necessary. Now, since the determination of one diffusion coefficient can be done using 8 channels if the sample time is chosen optimally why not combine a number of correlators operating at fixed sample times that grow exponentially from one correlator to the next? So the lag times would be, e.g.:

sub-correlator 1	100 ns	200 ns	300 ns	···	···	···	···	800 ns
sub-correlator 2	1600 ns	3200 ns	4800 ns	···	···	···	···	12800 ns
sub-correlator 3	25600 ns	51200 ns	76800 ns	···	···	···	···	204800 ns
···	409600 ns			···	···	···	···	···
···				···	···	···	···	···

In this case the sample time increases by a factor of 16 from one sub-correlator to the next. Other schemes are of course possible, e.g. letting the sample time increase by a factor of only 8 enabling an overlap between the lag time of the last channel of one correlator and the first channel of the next correlator.

It should be noted that the autocorrelation function is calculated quantity which depends on statistically fluctuating photon count rates. The ”precision” or the quality of the autocorrelation function thus depends on the precision of the terms in the sums of equation 4.8 as well as on the number of terms. The individual photon counts can be assumed to have a precision (statistical spread) inversely proportional to square root of the photon count itself (as is usually the case when counting random events). Thus, the higher the photon count during the sample time Δt the higher the precision of the counts themselves. This goes to show, that

to have an accurately determined autocorrelation function it is desirable to have as many photon counts as possible during the sample time. As the mean count rate is proportional to the intensity of the scattered light it is seen (equation 2.41) that the photon counts are proportional the laser power, the molecular weight and the weight concentration. The only way to compensate for low molecular weight, low concentration or low laser intensity is to have more terms in the sums og equation 4.8, i.e. to measure for a longer time. As only the concentration and the laser power is under the control of the experimenter and as high sample concentrations are generally not desirable it is obvious that a high laser power is usually desirable (as well as high efficiency detectors).
The number of photons $n(t)$, detected during the time Δt (i.e. from time t to time $t + \Delta t$) fluctuates for a number of reasons:

1. The individual scattering processes are inherently random
2. The scattering particles move randomly relative to each other (Brownian motion) creating time varying interference (see equation 4.5)
3. The number of particles in the scattering volume may vary randomly (i.e. the number N in equation 4.5 is not constant)
4. The individual particles may exhibit varying scattering efficiency depending on their internal structure and/or their orientation both of which may change at random

Here only Brownian motion is normally of concern:
If only the statistical nature of the individual scattering processes was present the autocorrelation function would be constant as all photon counts would be independent.
The fluctuating number of scatterers in the scattering volume is usually very small in relative terms (a typical scattering volume in DLS is $10^{-3}\ \mathrm{mm}^3$ but still the average number of particles at realistic concentrations would be of the order $N \approx 10^9$. This would give number fluctuations of the order $\sqrt{10^9}$ thus giving relative number fluctuations of the order $10^{-5} - 10^{-4}$).

The effect of internal motions are usually of no concern except in gelling systems and the effect of scattering efficiency being dependent on particle orientation is only of importance if the scattering particles are very large an have a pronounced dissymmetry (like long rods) because rotational diffusion in all other cases is so fast that time dependent anisotropic scattering is averaged out during normal sampling times.

The only mechanism of concern is therefore traslational diffusion (Brownian motion). What then will the autocorrelation function look like, typically? To answer this we note that the photon count $n(t + k \cdot \Delta t)$ is close to being the same as $n(t)$, if the time separation $k \cdot \Delta t$ between the two counts is very short because the particles will not have had time to move and change the phases ϕ_i. On the other hand, if the time separation is very long then there will be no connection (correlation) between the number $n(t + k \cdot \Delta t)$ and $n(t)$. Therefore the autocorrelation function will have a value of $\langle n^2(t)\rangle_t$ at short lag times and a value of $\langle n(t)\rangle_t^2$ at very long lag times. The former is always a larger number than the latter. At intermediate lag times there will be a gradual decay from the high to the low value as the particles randomly redistribute themselves. A typical plot of G_2 will therefore look like 4.3 The decay of the autocorrelation function can be envisaged as a gradual loss of memory in the system of particles or indicating for how long the particles retain their relative positions. The decay looks like an exponential decay towards an asymptotic value but is only so when the particles responsible for the scattering all have the same diffusion coefficient D. But now we shall briefly return to the previously mentioned (p. 74) *field autocorrelation function*.

The autocorrelation function

Looking again at figure 4.1, a sketch of a typical DLS-setup, you may have wondered why the laser beam is focussed and why there are pinholes in front of the detector. Equation 4.5 seems to offer a perfect explanation why the measured scattered intensity exhibits time dependent fluctuations: The phases ϕ_j and ϕ_k change randomly due to Brownian motion of the scattering particles. But one

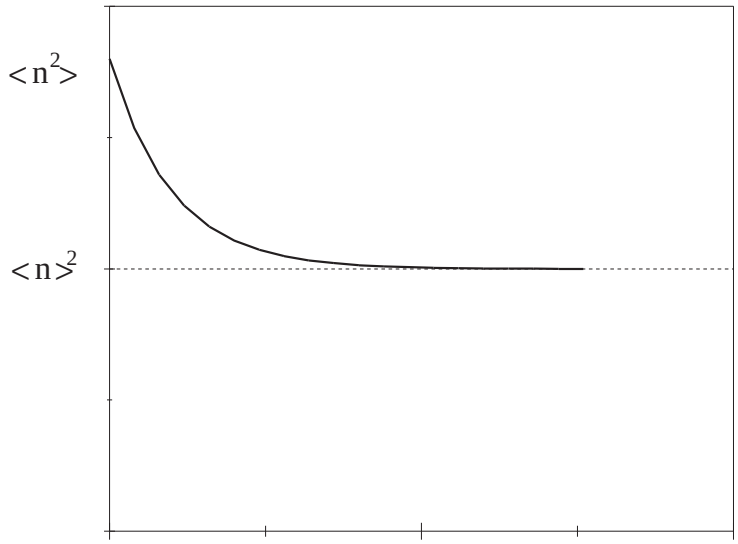


Figure 4.3: The autocorrelation function for the intensity of light scattered from a solution of particles undergoing Brownian motion always decays from a value of $\langle n^2 \rangle$ down to $\langle n \rangle^2$ at very long correlation times, n being the number of scattered photons counted within a fixed time frame Δt

should note that equation4.5 describes the intensity *at a point*. And no detector measures only at a point. The phases ϕ_j and ϕ_k are the phases of the electric fields scattered from particles number j and k as observed at a point on the detector. If the detector (including the lens and the pinholes) in figure 4.1 is replaced by a screen facing the scattering particles one will see areas of constructive interference between the electric fields alternating with areas of destructive interference (figure 4.4). The bright spots are called *speckles*. These speckles move about at random on the screen because the phase differences $\phi_j - \phi_k$ constantly change due to the Brownian motion of the scattering particles. It turns out that the diameter of the speckles depends on the diameter of the laser beam: The smaller the diameter of the laser beam the larger the diameter of the speckles. As the principle of dynamic light scattering is to utilise the fluctuations of the scattered intensity it is evident that in order to maximise the recorded *fluctuations* the detector should see only one speckle at a time: If the detector area is much larger than the area of one speckle the recorded intensity fluctuations will be small because approximately the same number of speckles will fall on the detector area all the time. Or, in

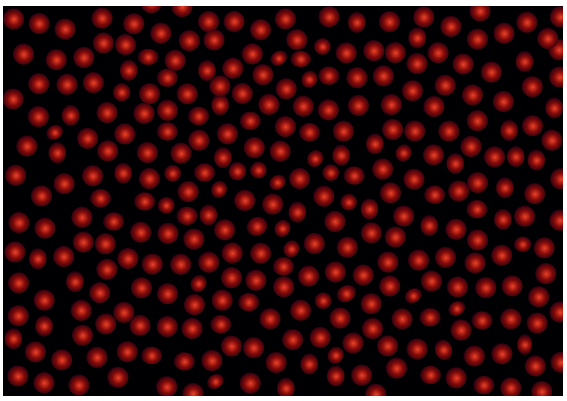


Figure 4.4: The speckles on the screen are areas where the scattered electric fields have the (approximately) the same phase and consequently interfere constructively creating a high intensity. The dark areas between the speckles are where the scattered electric fields are in counterphase thus eliminating each other. Positioning the screen further away from the scatterers will make the speckles larger. Reducing the diameter of the laser beam by focussing will also make the speckles larger.

statistical terms, the number of speckles falling on the detector area is fluctuating around a mean value N according to a Poisson distribution which has a statistical standard deviation of $\sqrt{N}$. This means that the relative magnitude of the fluctuations is $\sqrt{N}/N = 1/\sqrt{N}$, i.e. smaller the more speckles the detector sees. Thus the ideal detector will have an area which matches the area of the speckles: Either one or no speckle on the detector. Making the detector area even smaller would be a disadvantage because then less light would be detected. So why not make the speckles as large as possible by moving the detector further away from the scattering particles? The answer to this is of course that making the speckles larger in this way also makes them weaker, so it does't help. Then why not make the speckles as large as possible by focussing the laser beam as narrowly as possible? Here the answer is less straightforward. First of all there is a theoretical limit to how narrowly a laser beam can be focussed. But more importantly more narrow focussing requires a shorter focal length focussing lens. The shorter focal length the closer the lens has to be to the sample. But there are practical limits to how close the lens can be to the sample. And with a narrow laser beam optical align-

ment of the instrument becomes more critical. And finally, the more narrow the laser beam becomes the more divergent it also becomes making the initial direction of the incoming photons more uncertain. This means that narrow focussing creates an uncertainty in the magnitude of the (length of the) scattering vector q. And, as we shall see soon, an uncertainty in q directly translates into an uncertainty in the measured diffusion coefficient or equivalent hydrodynamic radii. So, in practice most DLS instruments make the same trade-offs: The laser beam is focussed down to a diameter to about $1/10\,\mathrm{mm}$ and the diameter of the pinhole in front of the detector is also about $1/10\,\mathrm{mm}$ (the diameter of the speckles).
Now, let us return to the two autocorrelations functions, the intensity autocorrelation function $G_2(\tau)$ (equation 4.6) and the field autocorrelation function $G_1(\tau)$ (equation 4.7). It is customary to normalise the functions so that the normalised intensity autocorrelation function $g_2(\tau)$ has the asymptotic value of one and the normalised field autocorrelation function $g_1(\tau)$ has the initial value of one (as the asymptotic values is zero and hence cannot be used for normalisation):

$$g_1(\tau) = G_1(\tau)/G_1(0) \tag{4.10}$$

$$g_2(\tau) = G_2(\tau)/G_2(\infty) \tag{4.11}$$

The relevance of the field autocorrelation function g_1 is that it is directly related to the fundamental physical processes creating the fluctuations: The interaction of the electric field with the particles undergoing random motions and the interference between scattered fields. The relevance of the intensity autocorrelation function g_2 is that it is related to what is directly observable (the fields are not observable as the oscillate by far too rapidly). Fortunately, and not surprisingly, there exists a relationship between the two autocorrelation functions, the so-called Siegert relationship:

$$g_2(\tau) = \beta[g_1(\tau)]^2 + 1 \tag{4.12}$$

where β is *the spatial coherence factor* a number which is always less than or equal to one. It is approximately one divided by the average number of speckles that the detector sees. Here again it is seen that the use of a large detector area does not produce better data because the informative part of g_2 which is $[g_1(\tau)]^2$

is reduced by the factor β which becomes smaller the larger the detector area, eventually wiping out all information in g_2. So, again, it is seen that the DLS setup should be optimised: The diameter of the focussed laser beam an the diameter of the pinholes should match.

So, what does the intensity autocorrelation function look like in practice? With a good deal of qualitative reasoning one can argue what the analytical form of the autocorrelation function should be: Look at figure 4.5 Let us consider diffusion in

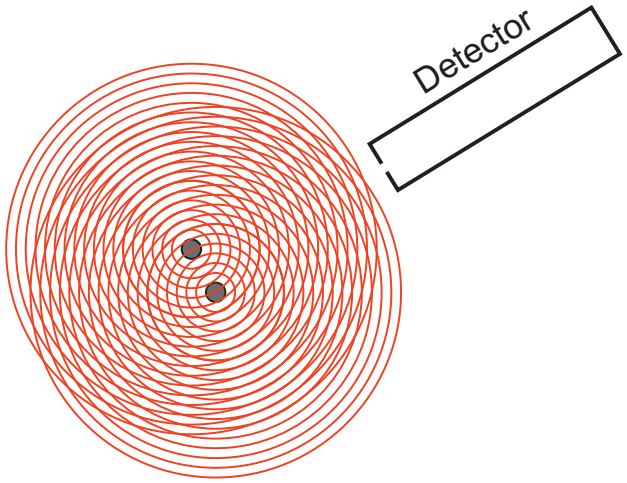


Figure 4.5: Light that reaches the small entrance to the detector will exhibit intensity fluctuations. The reason for this is that the scattering particles create constructive interference in some directions and destructive interference in other directions and these directions change when the particles move relative to each other. As a result intensity maxima and minima sweep across the detector entrance due to the Brownian motion of the scatterers. In order to have large intensity fluctuations on the detector the entrance hole must be small enough that only one interference maximum can fall on it at a time.

one dimension. The squared distance a particle will move from its initial position by diffusion during time t is given by the well-known expression:

$$\langle x^2 \rangle = 2Dt \tag{4.13}$$

An intuitive characteristic distance scale in this context would be the wavelength of the light because destructive interference between two sources of scattered light

occurs when their separation is a multiple of half wavelengths. But the interference not only depends on the distance between the scatterers but also on the scattering angle. Therefore the simplest "characteristic" distance in the context of light scattering is $1/q$ (q being the length of the scattering vector) because according to equation 2.30 light scattered from a particle separated the distance π/q from another particle will interfere destructively with light scattered from the first particle. Combining this with equation 4.13 we can "derive" a characteristic time τ_0 for the diffusion process:

$$\left(\frac{1}{q}\right)^2 = 2D\tau_0 \tag{4.14}$$

whereby $\tau_0 = 1/(2Dq^2)$. Of course this approach does not account for the precise numeric constants in equation 4.14. It might actually seem more natural to write $\frac{\pi}{q}$ instead of just $\frac{1}{q}$. But if we leave equation 4.14 as it is it does not seem unnatural to *guess* that the autocorrelation function in this simple case (of just one diffusion coefficient involved) would be an exponential decay with the characteristic time constant τ_0, i.e. G_2 would be of the form:

$$\begin{aligned} G_2(\tau) &= Ae^{-\tau/\tau_0} + B \\ &= Ae^{-2Dq^2t} + B \end{aligned} \tag{4.15}$$

Miraculously, this result turns out to be correct and we shall not give a rigorous derivation here[2]. In the more general case where n different types of particles are present each type having different diffusion coefficient D_1, D_2, D_3, ..., D_n, we can generalise equation 4.15 without proof to:

$$G_2(\tau) = (A_1e^{-D_1q^2\tau} + A_2e^{-D_2q^2\tau} + \ldots + A_ne^{-D_1q^2\tau})^2 + B \tag{4.16}$$

where the amplitude factors A_1, A_2, ..., A_n are proportional to to the light scattering contribution from each class of particles through the relationship

$$A_j \propto P_j(q)C_jM_j \tag{4.17}$$

[2]For those who do not believe in miracles a more rigorous derivation can be found in chapter 10 in *Dynamic Light Scattering*, R. Pecora, Ed., Plenum Press 1985 especially if combined with chapter 10 in *Biophysical Chemistry*, C. Cantor & P. Schimmel, Freeman and Company 1980

where P_j is the form factor for the j'th species, C_j is the weight concentration and M_j is the molecular weight.

We can normalise the measured autocorrelation function G_2 in the usual way so that it attains the asymptotic value of one for infinite lag time (equation 4.11). With the expressions (4.17) for the amplitude factors the normalised autocorrelation function can then be written:

$$g_2(\tau) = \beta(a_1 e^{-D_1 q^2 \tau} + a_2 e^{-D_2 q^2 \tau} + \ldots + a_n e^{-D_1 q^2 \tau})^2 + 1 \qquad (4.18)$$

where $a_1, a_2, \ldots, a_n$ denote the fractions
$P_1(q)C_1M_1/\sum P_i(q)C_iM_i$,$P_2(q)C_2/\sum P_i(q)C_iM_i$, $\ldots$, $P_n(q)C_n/\sum P_i(q)C_iM_i$
and β denotes *the spatial coherence factor*. The sum in the bracket is simply the field autocorrelation function. As discussed in connection with the Siegert relationship (equation 4.12) this is a number which is smaller than one and which depends on the detector geometry (and is thus an apparatus dependent constant). The more interference maxima the detector sees at the same time the smaller the coherence factor becomes. In practice the coherence factor also depends on the amount of uncorrelated light (or actually detector signal) the autocorrelator receives: Stray light from the surroundings entering the detector will diminish the coherence factor as will noise from the detector (noise being "signal" generated by the detector itself and not by actual light).

The expression 4.18 is usually the basis for the analysis of dynamic light scattering data on diffusing particles[3].

4.3 Data analysis

How to analyse autocorrelation data is a comprehensive subject as we shall see in the following, just scratching the surface.

[3]Equation 4.18 only describes diffusion of comparatively dilute suspensions of particles. And the particles should not exhibit internal motions, like bending or stretching on the length scale of $1/q$ or at least do so rapidly that the characteristic time scale is short compared to the shortest lag time of the experiment. Also, equation 4.18 does not describe internal motions in the system which are not related to diffusion processes like e.g. vibrational modes in gels.

The methods of data analysis can be divided into direct methods and indirect methods:

The problem is usually the following: Assume that a normalised autocorrelation function $g_2(t)$ has been measured. How can this measured autocorrelation function be written in the form of equation 4.18? What are the values of the different diffusion coefficient and how many (n) are there? Or stated in a slightly more general way: If we substitute the sum in equation 4.18 with an integral:

$$g_2(\tau) = \beta \left(\int_0^\infty A(D) e^{-Dq^2\tau} dD \right)^2 + 1 \tag{4.19}$$

where $A(D)$ is the distribution of diffusion coefficient which "generates" the measured intensity autocorrelation function, how do we find the distribution $A(D)$ which gives us the measured autocorrelation function when substituted into equation 4.19? Note that the integral in the bracket represents the field autocorrelation function:

$$g_1(\tau) = \int_0^\infty A(D) e^{-Dq^2\tau} dD \tag{4.20}$$

So we may equally well look for the distribution $A(D)$ which gives us the measured field autocorrelations function g_1 when substituted into equation 4.20. The field autocorrelation function g_1 can in principle be obtained from the measured intensity autocorrelation function g_2 through the Siegert relationship 4.12.

There exits a mathematical method called an inverse Laplace transformation which, at least in principle, can calculate the function $A(D)$ in equation 4.20. This would be a *direct method* of data analysis. As we shall see shortly this direct method does not work in practice. It leads to solutions that are unrealistic meaning that they usually show a mixture of features some describing the system on which the measurement was done and some features that are simply artifacts of doing a direct inversion on a measured (and therefore noise containing) autocorrelation function.

We are therefore left with only indirect methods. An indirect method means that you *assume* that you know the mathematical form of the autocorrelation function, that you can write an expression for it containing some adjustable parameters. The indirectness of the method is that you then adjust the parameters of this function so that it best fits the measures autocorrelation function. Finding the best fit means

to minimise the deviation between the measured autocorrelation function and the model function. This deviation is measured by the quantity called the chi-square, χ^2. The smaller the value of χ^2 the better the fit resembles the measured autocorrelation function. How this works is described in appendix B. But how does one choose the fitting function?

For a start one could try to simplify the problem e.g. based on an assumption that you know something about your system. The simplest assumption would be that the system contains only type of particles, i.e. there is only one diffusion coefficient involved. Such systems are called *monodisperse* whereas systems containing more than one type of particles are termed *polydisperse*. We shall look at these two possibilities in turn.

Monodisperse solutions

If the autocorrelation function is measured on a system with only one diffusion coefficient the normalised autocorrelation function g_2 can – using equations 4.15 and 4.18 – be written as:

$$g_2(\tau) = \beta \cdot e^{-2Dq^2\tau} + 1 \tag{4.21}$$

where β is the aforementioned coherence factor. If we now take the natural logarithm on both sides of equation 4.21 we get:

$$\ln(g_2(\tau) - 1) = \ln \beta - 2Dq^2\tau \tag{4.22}$$

so, if $\ln(g_2(\tau) - 1)$ is plotted versus $q^2\tau$ we get a straight line with the slope $-2D$ (which of course is determined by linear regression). If, on the other hand a plot of $\ln(g_2(\tau) - 1)$ versus $q^2\tau$ does not produce a straight line this of course means that more than one diffusion coefficient is present in the system, i.e. the system is *polydisperse*.

Polydisperse solutions

In real life it is rare that solutions only contain one type of particles so they can be represented by only one diffusion coefficient. Again one can try to simplify the

problem by assuming that the autocorrelation function has a specific form based on e.g. some knowledge or presumptions about the system. This could be that the system is composed of only two different species or it could be that the system contains particles with a narrow size distribution so that the diffusion coefficient can be described by a *Gaussian distribution* or are distributed according to some other known distribution function given by a relatively small number of parameters. In all of these cases the parameters are found through some type of regression analysis (see Appendix B). The details of the fitting procedure depend on the precise assumptions one has made about the system and thus about the mathematical form of the fitting function or about the distribution $A(D)$ of the diffusion coefficients.
We shall now look at a few of the analysis methods:

Multiexponential analysis

This is straightforward in principle: Simply assume that the measured autocorrelations function can be written as in equation 4.18:

$$g_2(\tau) = \beta(a_1 e^{-D_1 q^2 \tau} + a_2 e^{-D_2 q^2 \tau} + \ldots + a_n e^{-D_1 q^2 \tau})^2 + 1$$

Then, by non-linear regression (see Appendix B) have a computer adjust the parameters β, $a_1, a_2, \ldots, a_n$ and $D_1, D_2, \ldots, D_n$ until the calculated g_2 deviates as little as possible from the measured autocorrelation function. In fact only $2n$ and not $2n+1$ as it might seem need to be adjusted because the a's, being fractions of light scattering contribution are not independent: $a_1 + a_2 + \cdots + a_n = 1$.
This seemingly simple procedure sometimes works well but nevertheless has a number of flaws:

1. You must decide in advance how many terms to include, i.e. the value of n

2. The best fit is not necessarily a *good* fit. The goodness of the fit is determined on the basis of the value of the so-called reduced χ^2 (see Appendix B). It should ideally be close to one. Another test is to look at *a residual plot*. The residual plot is a plot of the difference between the measured autocorrelation function and the fit, i.e. a plot of $g_{2,\,\mathrm{measured}}(\tau) - g_{2,\,\mathrm{fit}}(\tau)$ vs. τ. The

ideal is that this difference looks completely random being sometimes positive, sometimes negative with no particular system. This situation indicates that the residual simply represents the ”noise” in the measured autocorrelation function. But sometimes (often) the residual plot shows a pronounced structure like an exponential or perhaps a sine-like function. This indicates that there is something apart from noise in the measured autocorrelation function that the fit can not account for: The fitting function has too few terms that you can or perhaps the wrong kind of terms. In any case, clear structure in the residual plot shows that you have the wrong model (fitting function).

3. You will always get a better fit the more terms you include in the fitting function because you get more parameters to adjust.

4. The drawback of adding more terms is that the parameters found become more and more irreproducible the more parameters you include: If you make, say, ten repeat measurements on the same system to get ten measured autocorrelation functions then the fitting procedure will not give you the same values for the fitted parameters for each of these measured autocorrelation functions. Some may not be very different but others may differ so much that it is impossible to say what set of diffusion coefficient best represents the system. The catch here is that using few parameters will sometimes generate a bad fit but the parameters will be reproducible upon repeat measurements. Increasing the number of parameters to improve the quality of the fit will produce increasingly irreproducible and thus meaningless parameter values.

How do you decide how many terms to include? Sometimes the autocorrelation function can actually be fitted well by using two terms that also give reproducible parameter values. One term rarely works well at least because there may be trace amounts of relatively large impurities present in the sample or the laser may produce light with an intensity which is not 100% constant. Random and relatively slow fluctuations of the laser intensity may be interpreted as a result of (usually)

large particles in the sample so a term to account for this is nearly always needed in the fitting function. But even with three terms reproducibility begins to be an issue and at four terms the limit is reached for the practical application of this procedure.

Cumulant analysis

We saw in section 4.3 that in the case of monodisperse solutions the analysis of DLS data was simple. If the solution is not monodisperse the relation 4.22 of course no longer holds. Still, the philosophy behind cumulant analysis (also called cumulants analysis) is that if the solution is not monodisperse but has a rather narrow Gaussian-like distribution of sizes then a plot of $\ln(g_2(\tau)-1)$ vs. $q^2\tau$ should produce if not a straight line then at least look like a polynomial. So, let us assume that diffusion coefficient (not the sizes) follow a Gaussian distribution. Then the function A in equation 4.19 can be written:

$$A(D) = A_0 e^{-\frac{(D-D_0)^2}{2\sigma^2}} \tag{4.23}$$

where D_0 is the mean diffusion coefficient, σ is the standard deviation of the distribution, and A_0 is a normalisation constant ensuring that $\int_0^\infty A(D)dD = 1$. With this equation 4.19 assumes the form:

$$g_2(\tau) = \left(\int_0^\infty A_0 e^{-\frac{(D-D_0)^2}{2\sigma^2}} e^{-Dq^2\tau} dD \right)^2 + 1 \tag{4.24}$$

which, after few integral substitutions, becomes:

$$g_2(\tau) = \left(e^{-D_0 q^2\tau + \sigma^2 q^4 \tau^2/2} A_0 \sqrt{2}\sigma \int_{-D_0/\sqrt{2\sigma}+\sigma q^2\tau/\sqrt{2}}^\infty e^{-x^2} dx \right)^2 + 1 \tag{4.25}$$

If τ is so small that $D_0/\sqrt{2\sigma} \gg \sigma q^2\tau/\sqrt{2}$ the integral in equation 4.25 is practically constant (independent of τ) and we can rewrite (approximate) equation 4.24:

$$g_2(\tau) \approx C \cdot e^{-2D_0 q^2\tau + \sigma^2 q^4 \tau^2} + 1 \tag{4.26}$$

where C is the constant given by the expression:

$$C = \left(A_0 \sqrt{2}\sigma \int_{-D_0/\sqrt{2\sigma}}^\infty e^{-x^2} dx \right)^2 \tag{4.27}$$

It is seen that $\ln(g_2(\tau)-1) = -2D_0q^2\tau+\sigma^2q^4\tau^2 = -2D_0\cdot(q^2\tau)+\sigma^2\cdot(q^2\tau)^2$ i.e. a second order polynomial in the variable $q^2\tau$. The coefficient of the first order term is the mean diffusion coefficient $2D_0$ og and the coefficient of the second order term is the variance σ^2 of the Gaussian distribution. Thus a parabolic fit to $\ln(g_2(\tau)-1)$ will thus give both the mean diffusion coefficient as the variance og the distribution. Both under the assumption that only data points corresponding to relatively small values of τ are used. This is also clear for another reason: The autocorrelation function is a decreasing function whereas the approximating parabola will eventually diverge at large values of τ.

The quantity $(\sigma/D_0)^2$ is usually called *the polydispersity* or the *the polydispersity index* of the distribution. If we set the polydispersity (or σ) equal to zero we see that equation 4.26 reduces to the earlier equation 4.21 for monodisperse solutions.

This method of making a parabolic fit to $\ln(g_2(\tau)-1)$, called second order cumulant analysis, must be used with some caution. The validity of the method requires that the distribution of diffusion coefficient has at least some resemblance with a Gaussian distribution, i.e. it should be unimodal (monomodal) and not multimodal (having some clearly separated size classes). Sometimes fitting with polynomials of higher order than two is used in which case you talk about third order (using a third order polynomial) or fourth order cumulant analysis. In all cases it is important not to extend the fitting procedure to τ values that are too large: The polynomial used will always diverge at large τ and thus become a bad representation of the autocorrelation function which is always asymptotically decreasing. On the other hand one should make sure to use enough data points to make a meaningful fit. Usually one does not go beyond third order cumulant analysis because the extra parameters are normally poorly determined, i.e. they vary a lot when the analysis is performed on autocorrelations functions obtained by simply repeating a measurement. The parameters of a second order cumulant analysis (mean diffusion coefficient and polydispersity) are usually quite reproducible which explains why this method is so popular. One should note, however, that the individual data points are weighted differently in a cumulant analysis than in an analysis employing a direct fit to the measured autocorrelation function. This means that

when employing direct fitting procedures the mean diffusion coefficient and the polydispersity may come out with slightly different values than when cumulant analysis is used.

On can wonder what it entails that the diffusion coefficient are described by a Gaussian distribution and how the diffusion coefficient of the individual species are weighted when the cumulant analysis calculates a *mean* diffusion coefficient D_0. To investigate this further note that at short correlation times the autocorrelation function 4.26 can be written as

$$g_2(\tau) \approx C(e^{-D_0 q^2 \tau})^2 + 1 \tag{4.28}$$

meaning that $\ln(g_2(\tau) - 1)$ plotted vs. $q^2\tau$ is a straight line for sufficiently small values of τ. If we now make a Taylor series expansion to first order of both equation 4.28 and equation 4.18 and subsequently equate the results we get:

$$\sum_{i=1}^{n} a_i(1 - D_i q^2 \tau) = 1 - D_0 q^2 \tau \tag{4.29}$$

leading to

$$D_0 = \sum_{i=1}^{n} a_i D_i \tag{4.30}$$

or, with the expressions for the fractions a_i written out:

$$D_0 = \frac{\sum_{i=1}^{n} P_i(q) C_i M_i D_i}{\sum_{i=1}^{n} P_i(q) C_i M_i} \tag{4.31}$$

This type of mean value is called the z-average of the diffusion coefficients:

$$D_0 = \langle D \rangle_z \tag{4.32}$$

This expression is seen to be exactly analogous to the expression for the mean radius of gyration 3.33.

Inverse Laplace transformation

The determination of the distribution of diffusion coefficient is mathematically equivalent to doing an inverse Laplace transformation of the previously mentioned

field autocorrelation function $g_1(\tau)$. To make a Laplace transformation of some function $f(\Gamma)$ means to calculate the function $g(t)$ (*the Laplace transform of* f) given by the expression:

$$g(\tau) = \int_0^\infty f(\Gamma) e^{-\Gamma t} d\Gamma \tag{4.33}$$

Note that if t is time then Γ has the dimensions of inverse time. From the measured autocorrelation function one can define the field autocorrelation function as $g_1(t) = \sqrt{|g_2(t) - 1|}$. It is necessary to take the absolute value under the square root because noise in the measured autocorrelations function g_2 may give occasional values of the baseline below one. If one then makes a change of variables, $Dq^2 = \Gamma$ and defines the function $B(\Gamma)$ describing the distribution of Γ:

$$B(\Gamma) = \frac{1}{q^2} A(D/q^2)$$

then it is seen that equation 4.20 is equivalent with:

$$g_1(\tau) = \int_0^\infty B(\Gamma) e^{-\Gamma t} d\Gamma \tag{4.34}$$

As $g_1(\tau)$ is determined experimentally (through $g_2(\tau)$) all that is needed is to perform a so-called inverse Laplace transformation of g_1 in order to find the function $B(\Gamma)$ and thereby the distribution of diffusion coefficient $A(D)$. This is a mathematically well-defined operation. But even though there exists computer algorithms that can perform inverse Laplace transformations it turns out that the result of this operation is highly ambiguous in the sense that if one repeats a measurement by simply letting the sample stay in the instrument an run a new measurement the ensuing distribution function $A(D)$ of diffusion coefficient (or, equivalently the distributions $B(\Gamma)$) can be *very* different. In fact, the results of inverse Laplace transformations can deviate arbitrarily much from the "true" distribution. The source of this ambiguity is "noise" in the experimentally determined function $g_1(\tau)$. One says that the inversion problem is "ill conditioned" or "ill-posed". It is not difficult to see how this comes about: Let us assume that for a given field autocorrelation function $g_1(\tau)$ we have found a corresponding distribution of the gammas, $B(\Gamma)$ which of course can be translated into an equivalent distribution of diffusion coefficients. Now, to this distribution function add

a term $b \cdot \sin(\Gamma\tau_1)$, where b is an amplitude constant and τ_1 is a time constant. Doing a Laplace transformation on this sum of functions will produce a result which is is $\int_0^\infty (B(\Gamma) + b \cdot \sin(\Gamma\tau_1)) \cdot e^{-\Gamma\tau} d\Gamma$. The result is (look in a table of integrals): $g_1(\tau) + \frac{b\tau_1}{\tau^2+\tau_1^2}$. The last term of this expression is the deviation from the experimentally determined field autocorrelation function g_1. This term has the maximum value of $\frac{b}{\tau_1}$ (for $\tau = 0$). This maximum value can be made as small as you wish just by choosing τ_1 large enough. This means that you can add a term to the distribution $B(\Gamma)$ with an arbitrarily large amplitude, b but with no visible consequence to the field autocorrelation function. This is so because the term added to g_1 is $\frac{b\tau_1}{\tau^2+\tau_1^2}$ which can be made arbitrarily small by choosing τ_1 sufficiently large. A similar argument shows that adding a cosine component $b \cdot \cos(\Gamma\tau_1)$ to the distribution function $B(\Gamma)$ will also have negligible impact on the ensuing Laplace transformed function. This argument shows that you can add any number of arbitrarily large terms to any found distribution $B(\Gamma)$ and still reproduce the measured field autocorrelation function within the limits of experimental precision. Adding such sine and cosine terms with different time constants $\tau_1, \tau_2, \ldots$ has the general impact on the distribution function $B(\Gamma)$ to give it more "structure", more spikes, more detail. This means that when doing an inverse Laplace transformation on a measured autocorrelation function you actually never know what features of the distribution function to trust. What features reflect properties of your system and what features are accidental artifacts due to noise in the measured autocorrelation function in combination with the inverse Laplace transformation algorithm? So, the problem can be stated in another way: Apparently there is no limit to the amount of "detail" or "information" you can get in your solution. But how much of this information is real? There are a number of different ways to cope with this problem, as we shall see.

ILT

One of the first data analysis methods to address this problem was ILT (Inverse Laplace Transformation, Ostrowsky m.fl. (1981)). The idea behind this method is to fit the experimentally measured field autocorrelation function $g_1(t)$ with an

expression of the form:

$$f(\tau) = a_1 e^{-\Gamma_1 \tau} + a_2 e^{-\Gamma_2 \tau} + \ldots + a_N e^{-\Gamma_N \tau} \tag{4.35}$$

where the *relaxation constants* Γ_i are spaced exponentially: $\Gamma_i = \varphi \Gamma_{i-1}$ where φ is a constant determining the separation between consecutive relaxation constants. This way of distributing the relaxation constants means that they will be spaced equidistantly on a logarithmic scale. What is determined when fitting method the expression 4.35 to the field autocorrelation function is only the constants a_1, a_2, $\ldots$, a_N while the "fineness of the exponential grid", φ, is determined in advance based on a knowledge of the noise in the measured g_1. Also, the the smallest and largest relaxation constants Γ_1 and Γ_N are set in advance.

CONTIN

Yet another way to limit the number of possible distribution functions $B(\Gamma)$ is to impose constraints on the solution. The idea is to somehow minimise the amount of structure in the solution without impairing too much how well the solution reproduces the measured autocorrelation function when Laplace transforming it. The data analysis method CONTIN (Provencher 1982) uses *regularisation* to minimise the amount of structure in the solution: Instead of finding a solution $B(\Gamma)$ which produces the best fit to the measured autocorrelation function, i.e. giving the smallest possible value of χ^2 the procedure attempts to minimise the expression

$$\chi^2 + \alpha^2 \int_{\Gamma_{\min}}^{\Gamma_{\max}} (B''(\Gamma))^2 \, d\Gamma \tag{4.36}$$

The term $B''(\Gamma)$ is a measure of the curvature of the function B. If the solution B has a sharp peak it will have a high curvature near the top. If the solution has several closely spaced peaks it will have high (but negative) curvature near the top of each peak but also near the bottom between peaks. So, taking the square of B'' makes sure that positive and negative curvature don't cancel each other. The integral measures the overall "spikiness" through the amount of curvature of the solution. And a lot of this type of detail is usually unwanted because it is unrealistic. Therefore the fitting procedure tries at the same time to make χ^2

small (in order to make a good fit) and to make the integral of the expression 4.36 small in order to make a the solution have a realistically small amount of structure. Which of the two terms count the most depends on the value of the regularisation parameter α. If e.g. $\alpha = 0$ only χ^2 counts and the best fit is found irrespective of its possible unrealistic amount of detail. If on the other hand α is very large χ^2 counts very little and a solution $B(\Gamma)$ with very little detail is found. It will also reproduce the measure autocorrelation function poorly, giving a large value of χ^2 but still, if α is large enough this doesn't matter. The problem which CONTIN solves is to set a sensible value for α. In terms of computing time CONTIN is somewhat slower than ILT. The difference between the solutions found by the two method is that CONTIN always produces wider and sometimes fewer peaks than ILT. One can say that CONTIN gives you *the least detailed solution (distibution of diffusion coefficients)* which is not in serious disagreement with the measured autocorrelation function.

Other methods

Within the last 30 years a number of different method to analyse autocorrelation functions have been developed. Cumulant analysis is one of the oldest but still has widespread use despite its obvious limitations, especially with modern correlators which operate with lag times that are spaced exponentially instead of linearly as was customary in the early days of dynamic light scattering and which was described in section 4.2. The newer methods are usually related to CONTIN and ILT in the sense that they use regularisation of some kind. Examples are REPES og RNONLIN which both analyse g_2 directly (without calculating g_1 first). Also MEM (maximum entropy method) is sometimes used but has not yet gained widespread usage. The principle of MEM is to assign to each of the possible solutions ($B(\Gamma)$) an "entropy" (probability) and then choose the solution with the highest entropy.

A method often used is to represent the data as a squared sum of exponentials as

shown in equation 4.18 and then do a non-linear least squares fit the the measured field autocorrelation function, g_2, to extract the parameters $a_1,\ a_2,\ \ldots$ og $D_1,\ D_2,\ \ldots$. It is necessary to limit the number of terms in the sum to e.g. 3 to avoid strong coupling (also called "cross correlation") between the parameters. If the method works it allows you to find three sharply separated diffusion coefficient (or particle sizes). But often it doesn't work because there are more than three particle classes. If they have widely separated diffusion coefficient more terms can be added but the method usually doesn't give useful and reproducible results with more than four terms, at the most. Sometimes the failure of fitting with a small sum of the type 4.18 is due to particle diffusion coefficient not falling into sharply defined classes but the classes exhibit polydispersity. It is possible to take this into account in an *ad hoc* way by fitting instead with a function of the *Kohlrausch type*:

$$g_2(\tau) = \beta(a_1 e^{-(D_1 q^2 \tau)^{\beta_1}} + a_2 e^{-(D_2 q^2 \tau)^{\beta_2}} + \ldots + a_n e^{-(D_1 q^2 \tau)^{\beta_n}})^2 + 1 \quad (4.37)$$

where $\beta_1,\ \beta_2,\ \ldots,\ \beta_n$ are all numbers lying between 0 and 1. If all β exponents are 1 we get the same type of functions as before (equation 4.18) but the closer β_i is to 0 the wider the distribution is belonging to the corresponding term. These individual terms are normally referred to as "stretched exponentials". Their main use in the analysis of autocorrelation functions for gelling systems, i.e. systems where the correlated intensity fluctuations are not necessarily related to diffusion processes in a solution but more to coupled motions in the gel.

5

Light scattering in practice

We shall here look at some of the ways in which light scattering measurements are done in practice.

5.1 Sample preparation

In light scattering sample preparation is crucial. Of course this is always so, but in light scattering special precautions need to be taken. The reason is dust. The light scattering power of a species is proportional to MC, molecular weight times weight concentration. Or put another way: The light scattering power of individual particles is proportional to M^2 (see equations 2.25 and 2.41). This means that even the tiniest amounts of dust in a sample can be devastating to measurements. The same goes for aggregated material. The precaution against this is

thought cleaning of buffers, sample cells and filtering of finished samples. Filtering the sample through a $0.2\ \mu\mathrm{m}$ syringe filter is standard procedure. Often putting the filtered sample into the syringe again filtering it once more through the same filter works wonders. The reason is that standard syringe filters often release small amounts of ”dust” from the membrane during the filtering procedure. This dust is the retained by the filter during the second round of filtration.

5.2 Apparatus

It was promised earlier, in section 3.5, that a more thorough description would follow of how detailed measurements of molecular weights can be done. The principle of the method is to separate the different molecular species in the solution so that the measurements are done on one class of molecules at a time and not on a mixture of species. One way this can be done is by passing the solutions through a size exclusion column which is usually connected in-line to the light scattering instrument and a concentration detector. The in-line setup is of course not crucial to the principle but is a major convenience. The in-line setup requires that the concentration detector and the light scattering instrument be equipped with flow cells. For concentration detector this has been the standard mode of operation for many years but for light scattering it is a mode of operation offered by only a few companies. A light scattering flow cell from Wyatt Technology is shown in figure 5.1. In this flow cell the laser beam passes trough in the direction of the sample flow. Detectors are place around the flow cell in the instrument allowing the measurement of the Rayleigh ratio at several angles simultaneously.
The separation principle of a size exclusion column is that larger molecules pass through the column faster than smaller molecules. The mechanism behind this seemingly ”backward” behaviour is that the column material is porous and spongy on many different size scales, small and large pores are interconnected. The larger molecules can only pass though the mesh of large interconnected pores in the column material whereas the small molecules can get into also some of the smaller pores where they will spend some time. Also, the speed of the carrier liquid is

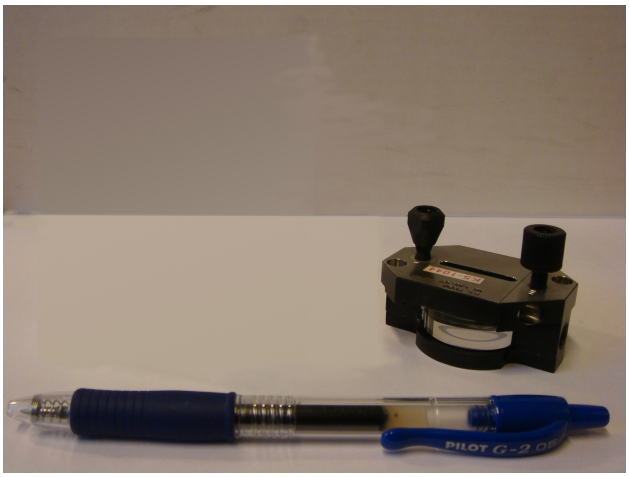


Figure 5.1: A light scattering flow cell from Wyatt Technology. Other companies use flow cells of other designs.

lower in the small pores due to viscous drag. So, consequently the molecules are sorted according to their size rather than their molecular weight.

It should be mentioned that size exclusion chromatography is not the only way to separate particles according to size. In recent years *field flow fractionation* (FFF) and *asymmetric flow field flow fractionation* (A4F) have gained widespread use especially for systems where traditional size exclusion columns have difficulties doing the separation.

In order to see how separating the different molecular species in a sample helps retrieve information about the system we can combine equations 3.16 and 3.14 to get

$$R = KCM_{\text{app.}} \tag{5.1}$$

where we have assumed that the concentration is low enough that we can disregard the virial coefficients of equation 3.16. If we measure the Rayleigh ratio R and the concentration C of the suspended, separated molecules we can calculate the apparent molecular weight:

$$M_{\text{app.}} = \frac{R}{KC} \tag{5.2}$$

where we assume the optical contrast constant K is known. In practice one often circumvents the step of knowing the contrast constant. We shall return to this issue

shortly. The Rayleigh ratio is usually measured at several angles at the same time so that a true molecular weight can be obtained by extrapolation to zero scattering angle.

How to determine concentration

One of the steps in the calculation of the molecular weight is to measure the concentration of the molecules. This is frequently done by measuring the UV absorption of the sample, but this of course only works if the suspended molecules absorb UV radiation. This will be the case if the molecules under investigation are proteins or amino acids but not if they are polysaccharides. Instead one can measure the refractive index n of the solution because its magnitude will depend on the amount of dissolved material and of course also on the refractive index of the pure solvent. In fact, up to quite high concentrations (more than 15% by volume) there is a linear relationship between the weight concentration of the dissolved material and the *difference* in the refractive index n_0 of the pure solvent and the refractive index n of the solution. Therefore the concentration can be determined by a measurement of $n - n_0$ which is conveniently done with a *differential refractometer*. This instrument contains a double flow cell where one of its compartments holds the pure solvent and the other has the solution running through it. The instrument then outputs a voltage which is proportional to the difference in refractive index $n - n_0$. If you make up a number solutions of different concentration of the protein under investigation, e.g. 0 g/L, 1.5 g/L, 2.0 g/L, 2.5 g/L, 3.0 g/L and 3.5 g/L and then send these solutions through the diffractometer sequentially the output of the instrument might look something like figure 5.2. Next, the the measured refractive index differences are plotted vs. the concentration of the solute. The result is shown in figure 5.3

What this means is that as long as the linear relationship holds we can write:

$$n - n_0 = \frac{dn}{dC} \cdot C \tag{5.3}$$

Thus, a knowledge of the value of $\frac{dn}{dC}$ enables us to determine the concentration through a measurement of $n - n_0$. It turns out that the value of $\frac{dn}{dC}$ depends

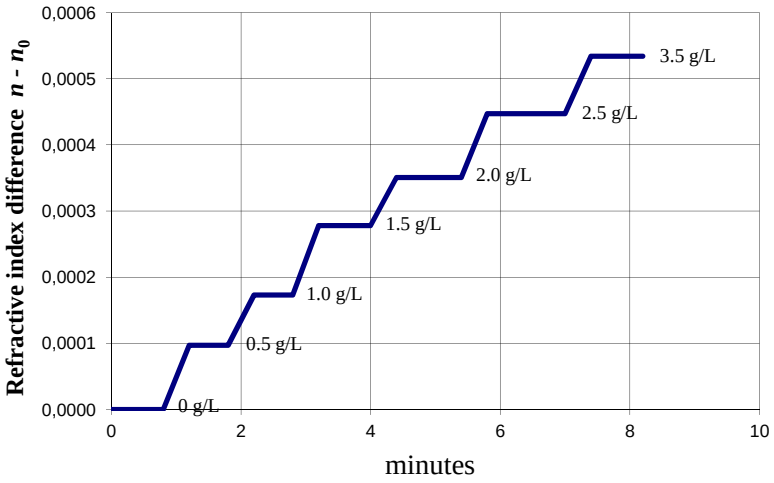


Figure 5.2: The refractive index of a solution depends on the refractive index of the solvent, on the refractive index of the pure solute and on the weight concentration C of the solute. The instrument measures the difference $n - n_0$ between the refractive index n_0 of the solvent and the refractive index n of the solution. Solutions of increasing concentration are pumped through the instrument. Solution of unchanged concentration is pumped through the instrument until the output for that concentration has reached a stable value, ie.e a plateau. Next to each plateau is shown the corresponding concentration of the solute.

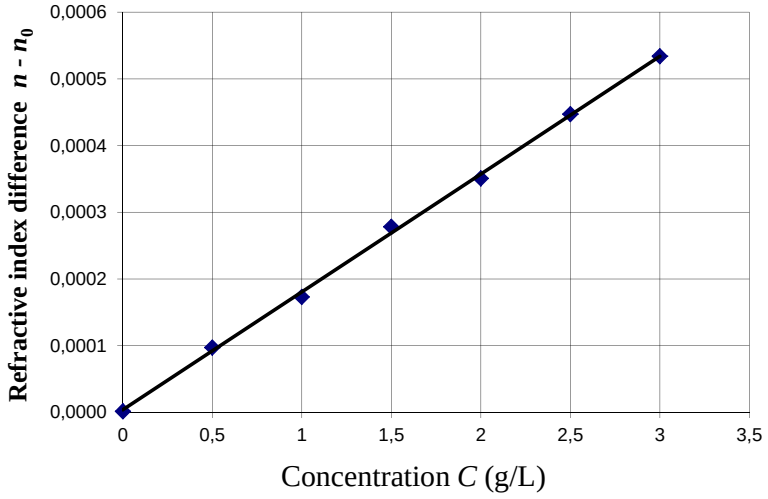


Figure 5.3: Refractive index difference as a function of the weight concentration of the solute. The points are taken fron the plateau values of figure 5.2. The relationship is linear (in most cases up to quite high concentrations). The slope of the regression line $\frac{dn}{dC}$ is characteristic of the combination of solute and solvent.

mainly on the *type* of molecule dissolved and on the solvent. For water-based solvents (buffers) the value is usually around $1.85 \cdot 10^{-4}\,\mathrm{m}^3 \cdot \mathrm{kg}^{-1}$ for proteins and approximately $1.45 \cdot 10^{-4}\,\mathrm{m}^3 \cdot \mathrm{kg}^{-1}$ for (poly)saccharides. The precise value will be somewhat dependent on the refractive index of the solvent and on other properties of the solute/solvent system (like, e.g. whether volumes are additive upon mixing). But in most cases the two quoted values will be accurate to within 5%.

Notice, that looking up $\frac{dn}{dC}$ values in the literature can be confusing as different authors often disagree on the values. If a precise value is of importance to the measurements it is necessary to do one's own determination preparing a number of different concentrations as described above. This will be especially important if the solvent is in some way "unusual", e.g. containing also apolar components, like alcohol or acetone.

Just for the sake of argument rather than for its usefulness it is possible to get a rough estimate of the $\frac{dn}{dC}$ value of a solute in a given solvent just by considering the difference in refractive index between the pure solute and the pure solvent. When the solute is present in its pure form (i.e. without being dissolved) it has a weight concentration which is the same as its density (look at the units, both mass pers unit volume). this means that we can make the estimate

$$\frac{dn}{dC} \approx \frac{\Delta n}{\Delta C} = \frac{n_{\text{rent stof}} - n_{\text{buffer}}}{\rho_{\text{rent stof}}} \tag{5.4}$$

because $\Delta C = C_{\max} - 0 = \rho_{\text{pure solute}}$. The refractive index of pure, water-free proteins is typically about 1.60 and for (poly)saccharides about 1.56 whereas the refractive index of most water-based buffers is around 1.33. The density of water-free proteins is typically $\rho \approx 1.4 \cdot 10^3\,\mathrm{kg} \cdot \mathrm{m}^{-3}$ and for polysaccharides $\rho \approx 1.58 \cdot 10^3\,\mathrm{kg} \cdot \mathrm{m}^{-3}$. Using equation 5.4 we obtain the following estimates:

$$\left(\frac{dn}{dC}\right)_{\text{protein}} \approx \frac{1.60 - 1.33}{1.4 \cdot 10^3\,\mathrm{kg} \cdot \mathrm{m}^{-3}} = 1.93 \cdot 10^{-4}\,\mathrm{m}^3 \cdot \mathrm{kg}^{-1} \tag{5.5}$$

$$\left(\frac{dn}{dC}\right)_{\text{polysac}} \approx \frac{1.56 - 1.33}{1.58 \cdot 10^3\,\mathrm{kg} \cdot \mathrm{m}^{-3}} = 1.44 \cdot 10^{-4}\,\mathrm{m}^3 \cdot \mathrm{kg}^{-1} \tag{5.6}$$

Both estimates lie within 5% of typical values.

Inline measurement of concentration in conjunction with SEC

When one makes a determination of the molecular weight of dissolved molecules using a setup as shown i figure 3.2 it is seen that the solution passes through the light scattering instrument before it passes through the concentration detector (the RI detector). Nevertheless we shall discuss the RI detector first.
In order to do a measurement a small volume (typically $100\,\mu$L) of the dissolved molecules is injected into the flow of buffer which is constantly being pumped through the column and the rest of the setup. Let us for simplicity assume that only one kind of solute molecules are present in the solution. When these molecules pass through the RI detector it will temporarily detect a higher refractive index difference $n - n_0$. If the molecules went through the column in exactly the same time the RI detector signal would suddenly jump to a higher value, the stay constant for a short while as the molecules passed, and then suddenly drop down to its original value again when all the molecules had passed (If the RI detector is "nulled" perfectly the signal will be zero when no solute molecules pass). But all molecules don't take exactly the same path through the column, so they will elute at slightly different times. So, instead of this "square" like output from the RI detector the signal will look like shown in figure 5.4. The RI detector measures the difference in refractive index between the solution being pumped through its measurement cell and a liquid in a reference cell. The other liquid is usually just the pure solvent used to dissolve the molecules. In principle the RI detector should produce a zero signal when just pure buffer is running through the measurement cell but in practice this is usually not quite so. There will be an offset for a number of reasons e.g. small differences in the optical system of the measurement cell and of the reference cell, and small differences in the composition of the solvent used to make up the solution and the solvent running through the system[1].

[1]RI measurements are extremely sensitive. Usually the running buffer which is in a reservoir is fed through the measurement cell and the reference cell at the beginning of a measurement. Then after some minutes the reference cell is shut off so that the buffer only runs through the measurement cell. A difference in refractive index between the two cells may develop if evaporation of water from the reservoir takes place thereby increasing

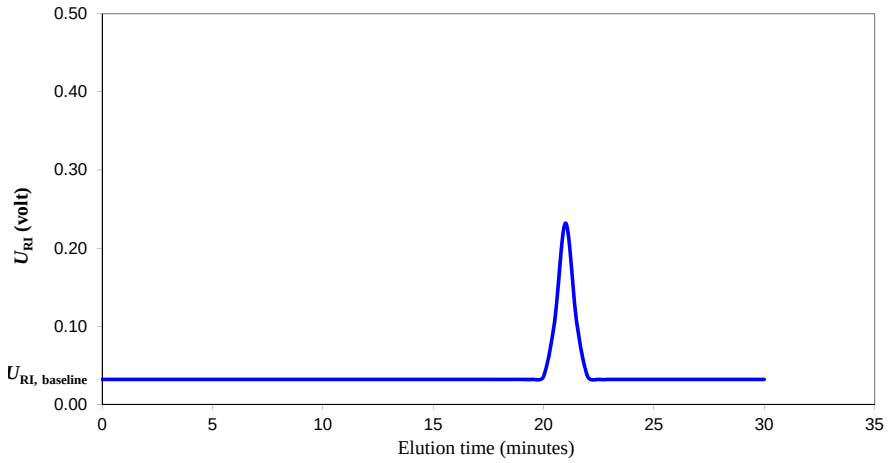


Figure 5.4: The signal from the refractive index detector (RI detector) lies at a constant value, the baseline, as long as only buffer is running through the measurement cell. The baseline value is ideally zero but often has a slight offset, which is of no importance. As the dissolved molecules pass through the measurement cell the signal rises and then falls back to its baseline value. In practice only the deviation from the baseline value is used which is exactly why the baseline offset is of no concern.

Finally, there are offset errors in the electronic system which amplifies the signals from the two cells.

All in all this means that the electrical voltage which is the output from the RI detector can be written

$$U_{\mathrm{RI}} = U_{\mathrm{RI,\ Baseline}} + k_{\mathrm{RI}} \cdot (n_{\mathrm{solution}} - n_{\mathrm{buffer}}) \tag{5.7}$$

or, applying equation 5.3:

$$U_{\mathrm{RI}} = U_{\mathrm{RI,\ Baseline}} + k_{\mathrm{RI}} \cdot \frac{dn}{dC} \cdot C \tag{5.8}$$

The constant k_{RI} depends, among other factors, on the amplification of the electronics in the RI detector but often has a value so that a 0.01 change in the refractive index produces a change in output the voltage U_{RI} of 1 volt. The precise value of k_{RI} has to be known, of course (it can be found in the instrument manual), in order to use it to determine concentrations or $\frac{dn}{dC}$ values.

the concentration of salts in the buffer. Also air may dissolve in the buffer during the day which will also change its refractive index!

Inline measurement of light scattering in conjunction with SEC

We still imagine a small amount of solution containing identical solute molecules having been injected into the buffer stream. When the molecules have passed through the size exclusion column they first pass through the flow cell of the light scattering instrument producing a signal from one of the light scattering detectors as shown shown in figure 5.5. When the instrument measures the intensity of

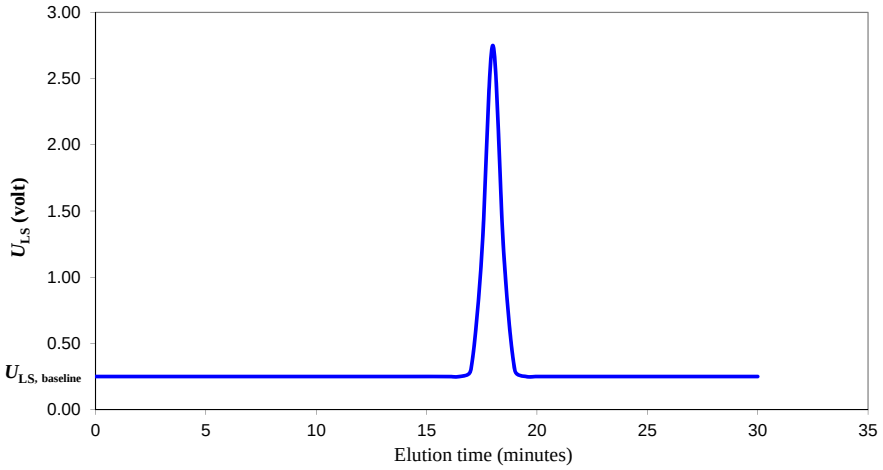


Figure 5.5: The signal from the light scattering instrument lies at a constant baseline value as long as only the pure buffer is running through the flow cell. When the dissolved molecules pass the flow cell the signal rises and subsequently subsides back to the baseline value when all the solute molecules have passed. Compare this signal with the RI signal (concentration signal) of 5.4. Note, that the RI signal is delayed approximately three minutes compared to the light scattering signal. This is simply the time it takes for the molecules to run through the tubing that connects the light scattering instrument with the RI detector. This time depends, of course, on the flow rate, the length of the tubing and the inner diameter of the tubing. And three minutes is, by the way, unrealistically long.

the scattered light the same problem arises as with the measurement of refractive index: The signal is not zero when just buffer and no solute molecules run through the measurement cell. There are a number of reasons for this: First of all the

electronics of the measurement system usually has an output voltage offset which means that the output voltage is different from zero even when the laser is turned off. Secondly, when the laser is turned on but only pure solvent runs through the flow cell some stray light will be scattered into the detector system from glass surfaces, microscopic scratches and dents and dirt particles on the flow cell and its entrance windows. And finally, even a pure liquid will scatter light to some extent due to local spontaneous statistical fluctuations in the liquids refractive index. Adding up all these factors means that the output signal of the light scattering instrument for a given scattering angle or corresponding q value can be written

$$U_{\mathrm{LS}} = U_{\mathrm{LS,\ Baseline}} + k_{\mathrm{LS}} \cdot R(q) \tag{5.9}$$

where $R(q)$ is the Rayleigh ratio at the given q value. The value of the constant k_{LS} depends on the detector efficiency, the detector area and the gain in the electronics of the instrument. If the instrument has multiple detectors each of them has its own value of the constant k_{LS}. With the help of equation 5.1 we can write equation 5.9 as

$$U_{\mathrm{LS}} = U_{\mathrm{LS,\ Baseline}} + k_{\mathrm{LS}} \cdot KCM_{\mathrm{app.}} \tag{5.10}$$

where $K = \frac{4\pi^2 n_0^2 (dn/dC)^2}{N_A \lambda_0^4}$ is the previously mentioned optical contrast constant defined in section 2.7. If, for simplicity, we write the contrast constant as $K = k_1 \cdot (dn/dC)^2$ we can rewrite equation 5.10 once more to read

$$U_{\mathrm{LS}} = U_{\mathrm{LS,\ Baseline}} + k_{\mathrm{LS}} \cdot k_1 \cdot (dn/dC)^2 \cdot CM_{\mathrm{app.}} \tag{5.11}$$

Note the similarity between this signal (equation 5.11) and the concentration signal 5.8. The RI signal is, however, delayed a certain time, in this case by three minutes. The delay is the time it takes for the molecules in the light scattering flow cell to move to the RI detector flow cell. This of course depends on the flow rate and the volume of the tubing connecting the two instruments. So, when comparing the light scattering signal and the concentration signal they should be compared with this time delay in maind, i.e. compare $U_{\mathrm{LS}}(t)$ with $U_{\mathrm{RI}}(t + \Delta t)$ where Δt is the time delay between the two signals. In this example 3 minutes. The precise value of has to be determined experimentally for a given flow rate by

comparison of positions of the peak values of the two signals. If we then combine equations 5.8 and 5.11 we can calculate $M_{\text{app.}}$ as

$$M_{\text{app.}} = \frac{(U_{\text{LS}} - U_{\text{LS, Baseline}}) \cdot k_{\text{RI}}}{(U_{\text{RI}} - U_{\text{RI, Baseline}}) \cdot k_1 \cdot k_{\text{LS}} \cdot \frac{dn}{dC}} \tag{5.12}$$

We may just as well combine the three constants k_{RI}, k_1 and k_{LS} into one constant (for each detector):

$$k_{\text{det.}} = \frac{k_{\text{RI}}}{k_1 \cdot k_{\text{LS}}}$$

With this we can write the apparent molecular weight (at some q value)

$$M_{\text{app.}} = \frac{k_{\text{det.}}}{\frac{dn}{dC}} \cdot \frac{U_{\text{LS}}(t) - U_{\text{LS, Baseline}}}{U_{\text{RI}}(t + \Delta t) - U_{\text{RI, Baseline}}} \tag{5.13}$$

where we have indicated explicitly the time delay between the two signals.
Looking at equation 5.13 it might look as if the molecular weight is a function of time. Elution time, that is. In principle this is true. But if there are no appreciable concentration effects, i.e. the second virial coefficient is sufficiently small or the concentration is sufficiently small then the light scattering is simply proportional to the concentration. In this case the numerator and the denominator of equation 5.13 will be proportional at all times and hence they will give the same value of $M_{\text{app.}}$ at all times. This, of course, is theory. In practice it is important to make sure that U_{RI} is not too close to its baseline value because then one would divide with a number close to zero in equation 5.13. The consequence of this would be to produce a very "unstable" value of $M_{\text{app.}}$. So, in practice the molecular weight is determined using values not too far from the peak values of the light scattering and RI curves. If these precautions do not produce a constant molecular weight across the central part of the peak it may be a sign that the second virial coefficient cannot be neglected. The molecular weight across the concentration peak may look "parabolic" with either upward or downward turning branches. Which of the two is the case depends on the sign of the second virial coefficient (with upward turning branches indicating a positive second virial coefficient and *vice versa*).

How to determine the molecular weight, and possibly the radius of gyration

We shall now assume that we can disregard concentration effects. Whether this is justified can be checked, as previously mentioned, by checking if the molecular weight calculated at the peak of the concentration curve is the same as the molecular weight calculated slightly off the peak (where the concentration is lower). What remains to actually determine molecular weights is to establish the value of the constant $k_{\mathrm{det.}}$ in equation 5.13, or in fact, all of the values of this constant i.e. one for each detector of the instrument. The value is determined by injecting into the apparatus a solution of a solute of known molecular weight and known value of $\frac{dn}{dC}$. It should be a solute consisting of molecules small enough that their form factor has the constant value of one at all scattering angles of the instrument. This ensures that the apparent molecular weight of the solute, $M_{\mathrm{app.}}$ is actually the true molecular weight at all scattering angles. Such a molecule could be BSA (bovine serum albumin) which in its monomeric form has a molecular weight of $M_{\mathrm{BSA}} = 66400\,\mathrm{g}\cdot\mathrm{mol}^{-1}$. When running a solution of BSA through the column a number of peaks will show up both in light scattering and in the RI detector. Usually four peaks are encountered the first, pertaining to the largest species, corresponds to aggregates of BSA and possible impurities. Then next come three more peaks corresponding to oligomers (mainly tri- and tetramers), dimers and monomers. The monomer peak is last and largest and is used to establish the value of the detector constants:

$$k_{\mathrm{det.}} = M_{\mathrm{BSA}} \cdot \left(\frac{dn}{dC}\right)_{\mathrm{BSA}} \cdot \frac{U_{\mathrm{RI,\,top}} - U_{\mathrm{RI,\,Baseline}}}{U_{\mathrm{LS,\,top}} - U_{\mathrm{LS,\,Baseline}}} \tag{5.14}$$

Note that the value of the detector constant(s) does not depend on the calibration solute used as long as its molecular weight and dn/dC value are known. But still the solute used for calibration should be a small enough molecule to scatter isotropically. This is not a major limitation as anisotropic scattering is usually not detectable for proteins with a molecular weight below $500,000\,\mathrm{g}\cdot\mathrm{mol}^{-1}$. Once the detector constants have been established they can be used in equation 5.13 to determine the apparent molecular weight of the protein one is studying. The

apparent molecular weight is calculated for each detector (q value) and a Guinier plot can then be used to determine the true molecular weight by extrapolation to $q = 0$. If the slope of the Guinier plot is sufficiently high the radius of gyration for the molecules can be determined as well. How small a radius of gyration can be determined depends on the quality of the data and on the precision with which the detector constants have been determined. In practice R_g values down to about $10\,\mathrm{nm}$ can be determined under optimal conditions.

5.3 Absolute calibration

The calibration procedure described above is often replaced by a so-called absolute calibration. The purpose is exactly the same, i.e. to establish the value of the detector constants. What is used is a liquid which *in itself* scatters light. The liquid most commonly used is toluene which scatters light due to spontaneous density fluctuations (creating fluctuations in the refractive index). The advantage of this is that you don't have to make up a fresh solution of protein every time the instrument has to be calibrated. Pure toluene is always the same, it doesn't change composition due to evaporation and it doesn't degrade with age or if kept at the wrong temperature. All liquids scatter light for the same reason, but toluene is particularly good at it. Pure toluene scatters light as efficiently as a BSA solution with a concentration of $0.65\,\mathrm{mg}\cdot\mathrm{mL}^{-1}$. For comparison pure water scatters approximately 20 times less. The details of the calibration procedure will be different from those outlined above but commercial instruments come with detailed instructions as to how calibration is done.

A

Complex numbers

Complex numbers are a very handy tool in certain types of computations that have to do with the sine and cosine functions. The reason is that with complex numbers it turns out that the sine and cosine functions are closely related to the exponential function. And it is often much more straightforward to do calculations with the exponential function than with trigonometric functions. The rules of calculation with the exponential function are few and simple whereas the rules of manipulating the trigonometric functions are intricate and seem nearly endless in number.
Complex numbers are an extension of the well-known real numbers. The were invented with the purpose of enabling solution of ordinary equations of any degree. It turns out that the only thing required to accomplish this is the introduction of a new ”number” defined as

$$i = \sqrt{-1}$$

This number is called *the imaginary unit*. With this new and strange number one can define a more general type of number called *complex numbers*. A complex number z is defined as

$$z = x + i \cdot y$$

where x and y are real numbers. If $y = 0$ z will be a real number so complex numbers are clearly a generalisation of real numbers. If, on the other hand, $x = 0$ the number z is said to be a pure imaginary number. An example of the use of the complex numbers is the solution the algebraic equation of the second degree

$$z^2 + 1 = 0$$

which has no solutions among the reals numbers but has the solutions i and $-i$ among the complex numbers. All the usual rules of algebra are valid for these "extended" numbers and no further extensions are necessary to solve any degree of algebraic equation. Not even if the coefficients of the equation are themselves complex numbers. And it is possible to calculate e.g. square roots of not only negative numbers but also of complex numbers. The result would just be another complex number.

A complex number z cannot be represented on a one-dimensional line like the reals numbers. Complex numbers require two dimensions, a coordinate system called *the complex plane*.

One defines (as for vectors) the absolute value $|z|$ of a complex number z as

$$|z| = \sqrt{x^2 + y^2}$$

One common operation on a complex number is called *complex conjugation*. What this does is the change the sign of imaginary part of the complex number. The *complex conjugate* of a complex number $z = x + i \cdot y$ is denoted z^* and is defined as

$$z^* = x - i \cdot y$$

The complex conjugate of z has the useful property (remember $i^2 = -1$):

$$zz^* = |z|^2$$

Proof: $zz^* = (x+i\cdot y)\cdot(x-i\cdot y) = x^2+i\cdot xy-i\cdot yx-i\cdot y\cdot i\cdot y = x^2-i^2\cdot y^2 = x^2+y^2 = |z|^2$

Now comes the really useful part: The exponential function with a purely imaginary argument $i\cdot y$ is defined as:

$$e^{i\cdot y} = \cos(y) + i\cdot\sin(y) \tag{A.1}$$

which shows that there is a connection between the exponential and the trigonometric functions. The most important property of the ordinary exponential is that it is its own derivative. How about this new definition? To check this we differentiate equation A.1 with respect to y:

$$\frac{d}{dy}e^{i\cdot y} = \frac{d}{dy}(\cos(y)+i\cdot\sin(y)) = -\sin(y)+i\cdot cos(y) = i(\cos(y)+i\cdot\sin(y)) = i\cdot e^{i\cdot y} \tag{A.2}$$

But note that the argument of the exponential is not y but $i\cdot y$, so in fact we should have differentiated with respect to iy, so with the above calculation we get

$$\frac{d}{d(iy)}e^{i\cdot y} = \frac{1}{i}\frac{d}{dy}e^{i\cdot y} = \frac{1}{i}i\cdot e^{i\cdot y} = e^{i\cdot y} \tag{A.3}$$

i.e. the new exponential is indeed its own derivative.

More generally the exponential of a complex number $z = x+iy$ is simply defined by the usual rules for the exponential:

$$e^{x+iy} = e^x\cdot e^{iy} = e^x\cdot(\cos(y)+i\sin(y))$$

Note that the complex conjugate of the complex exponential is

$$\begin{aligned}(e^{x+iy})^* &= \\ (e^x\cdot(\cos(y)+i\sin(y)))^* &= \\ e^x\cdot(\cos(y)-i\sin(y)) &= \\ e^x\cdot(\cos(-y)+i\sin(-y)) &= \\ e^{x-iy}&\end{aligned}$$

i.e. simply change the sign of the imaginary part in the exponent.

The usefulness of the complex exponential in physics lies in its use to describe

wave phenomena. An electromagnetic wave propagating along the $x-$axis can be described with a sine or a cosine function as e.g.

$$E(x,t) = E_0 \cos(\omega t - kx)$$

With the complex exponential this can be written in an alternative form:

$$E(x,t) = E_0 e^{i\,(\omega t - kx)}$$

In fact one gets too much here. Also an imaginary term ($E_0 i \sin(\omega t - kx)$) is included. The "trick" is simply to disregard the imaginary part of the the expression (it is "unphysical"). It is seen that using the definition of the complex conjugate (or the above expression for the complex conjugate of the exponential) that taking the complex conjugate of $E(x,t)$ gives

$$E^*(x,t) = E_0 e^{-i\,(\omega t - kx)}$$

whereby

$$\begin{aligned} E(x,t)E^*(x,t) &= E_0^2 e^{i\,(\omega t - kx)} e^{-i\,(\omega t - kx)} \\ &= E_0^2 e^{i\,(\omega t - kx) - i\,(\omega t - kx)} \\ &= E_0^2 \end{aligned}$$

which means that $E(x,t)E^*(x,t)$ or equivalently $|E(x,t)|^2$ is simply the squared amplitude of the electromagnetic wave. If $E(x,t)$ is a sum of many different cosine terms (as the total electric field originating from the scattering of several particles) it is a simple matter to compute the amplitude of this electromagnetic wave. This is the reason for the use of the complex exponential in physics.

B

Data fitting

A frequently encountered problem in experimental physics is that of describing some measured data as well as possible with a function. Assume one has done a number of measurements of some physical quantity y as a function of some other physical quantity x. It could be measurement of the pressure of an enclosed gas as a function of the volume of the gas. The ensuing set of data is $y_1,\ y_2,\ \ldots,\ y_N$ measured as a function of the quantity x at the corresponding values $x_1,\ x_2,\ \ldots,\ x_N$. The task is now somehow to determine a function f, i.e. $y(t) = f(x)$ which best fits the measured data. The question is what "best" means. What is meant is usually that the sum of squared differences between the measured values and the calculated ones is as small as possible. This quantity is

called *the chi-square* of the fit and is assigned the symbol χ^2. It is defined as

$$\chi^2 = \sum_{j=1}^{N} \frac{(y_j - f(x_j))^2}{\sigma_j^2}$$

The denominators are the statistical variance (a statistically precise measure of the uncertainty, squared actually) of the individual data points. The value of these may be established e.g. by repeating the measurements or by theoretical considerations. This is a fair definition of the goodness of the fit because data points with a large uncertainty contribute less to the sum and hence have correspondingly little influence on how to choose the function $f(x)$. Now, the function normally has to belong to a specific class of functions, e.g. fourth degree polynomials or sums of six sine functions of different frequency. What the fitting procedure comes down to is then to determine the optimum set of parameters (e.g. the five coefficients in a fourth degree polynomial) that will make χ^2 as small as possible. Of course it is always possible to make a better fit if more parameters are included in the fitting function, e.g. using a fifth degree polynomial instead of a fourth degree polynomial. Note also that the magnitude of χ^2 will grow if more data points are included since this will increase the number of terms. Therefore, in order to judge how good a fit actually is it is not sufficient to calculate χ^2 it is also necessary somehow to correct for the number of terms in the sum and for the number of parameters used. To accomplish this task one defines *the reduced chi-square* χ_r^2

$$\chi^2 = \frac{1}{N-n-1} \cdot \sum_{j=1}^{N} \frac{(y_j - f(x_j))^2}{\sigma_j^2}$$

where N is the number of data points (i.e. the number of terms in the sum) and n is the number of parameters used to describe the function $f(x)$. The sign of a good fit that $\chi_r^2 \approx 1$. If the reduced chi-square is much larger than one the fit is poor, i.e. the parameters are not optimally chosen or the choice of function for the fit is wrong so that even the minimum value of χ_r^2 is substantially larger than one. If, on the other hand the reduced chi-square is much smaller than one the reason is usually that the variances σ_j^2 have been overestimated.

As an example, say we have measured an autocorrelation function at 256 lag

times: $g_2(t_j)$ where $j = 0, \ \ldots, \ 256$ and we want to fit it with a model function of the form $y(t) = (A_1 e^{-D_1 q^2 t} + A_2 e^{-D_2 q^2 t})^2 + 1$. The parameters to adjust are $A_1, \ A_2, \ D_1, \ D_2$, i.e. we have to minimise

$$\chi^2 = \sum_{j=1}^{256} \frac{(g_2(t_j) - [(A_1 e^{-D_1 q^2 t} + A_2 e^{-D_2 q^2 t})^2 + 1])^2}{\sigma_j^2}$$

In practice this is done with a computer program which adjusts the different parameters until a minimum for χ^2 seems to be found. In order to assess the *goodness of the fit* the minimum value just found is divided by the number $256 - 4 - 1$.
A number of different programs to do this search for optimum parameters exists but none of them can guarantee that an absolute, global minimum is actually found. Nor do the programs normally guarantee that even a local minimum is found, sometimes the search simply diverges. The way most programs of this type work is by calculating the *gradient* of χ^2 which is a vector in parameter space pointing in direction where χ^2 increases the most. The program (or the user of the program) makes and initial guess about the value of the parameters, then calculates the gradient at this parameter point. Next, the program finds a new and hopefully better parameter point by moving a short distance in the exact opposite direction of the gradient. Then the gradient is calculated in this new parameter point and the procedure is repeated over and over again until the gradient is close to zero (or actually $\vec{0}$) indicating that a minimum has been found. Or a maximum or a saddle point. Further analysis may be required to ensure that the parameter point found is actually a minimum. This type of analysis is called a gradient search and has the advantage of being fast when it works. But it frequently fails if the initial parameter guess is far from a minimum point.
Sometimes more "safe" types of fitting schemes can be employed. The fit described above is so-called *non-linear fit*. An example of the opposite, a *linear fit* would be if one would fit the field autocorrelation function $g_1(\tau)$ with a model function of the form $f(t) = A_1 e^{-D_1 q^2 t} + A_2 e^{-D_2 q^2 t}$ where the diffusion coefficient D_1 and D_2 were fixed so that only the amplitude parameters A_1 and A_2 had to be optimised. In this case there is a direct method to calculate the optimum parameters with no risk of failure. Of course the problem then is what values of D_1 and D_2

to choose in advance. One way to circumvent this problem is to use not two but perhaps 100 terms with predefined diffusion coefficient $D_1,\ D_2,\ D_3, \ldots$ separated by a small factor, like e.g. 1.3, i.e. $D_{i+1} = 1.3 \cdot D_i$ for $i = 1,\ 2,\ 3, \ldots$. This, however gives rise to other problems which are discussed on pages 94 and 95.

Index